Conformational Control of Cofactors in Nature – The Influence of Protein Induced Macrocycle Distortion on the Biological Function of Tetrapyrroles in Photosynthesis

by

Stuart A. MacGowan B.Sc. (HONS)

Under the supervision of Prof. Dr. Mathias Senge

September 2014

Trinity College Dublin

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Stuart A. MacGowan B.Sc. (HONS)

Summary

This thesis details new research into the influence of nonplanar distortions upon the function of porphyrin based cofactors in proteins *via* analysis of structural data deposited in the Protein Data Bank (PDB) combined with quantum molecular modelling. Previous reports of conserved macrocycle conformations across related proteins (*e.g.*, twisted haems in peroxidases) together with physicochemical studies on nonplanar porphyrins led to the idea that protein induced nonplanarity could be a modulatory parameter exploited to affect efficient biological function. Whilst the phenomena has been investigated substantially in haemoproteins, photosynthetic systems have received less attention and so this work treats reaction centres (RC) and the tetrahaem cytochrome subunit (RC-cyt) from purple bacteria, and the light-harvesting Fenna-Matthews-Olson (FMO) protein from green bacteria.

Research began with the analysis of the conformations of RC bacteriochlorophylls (BChls) as revealed by X-ray crystal structures deposited in the PDB. The RC contains four identical BChl cofactors plus two bacteriopheophytins that display unique properties and perform individual roles in the RC. The main caveat at the outset of the analysis was the quality of the experimental data since today's best resolution structures peak at only 1.87 Å. As a consequence the abundant medium resolution data available were subjected to multivariate statistical analyses to test the hypothesis that the macrocycle conformations were unique. The ability to perform this task was enabled by the method known as normal-coordinate structure decomposition, which affords an intuitive numerical description of tetrapyrrole macrocycle conformation. The initial hypothesis proved correct, and thus the analysis provided the first reliable report that the individual BChls in the RC possessed distinct nonplanar conformations that in some cases broke the apparent C_2-axis of the RC. With these conformations in hand, the relative perturbations of the BChls' key properties (redox, absorption and excited-state properties) were predicted, with reference to the current state-of-the-art regarding the known effects of nonplanarity and turned out to be correlated with each BChl's individual role as well as species-specific characteristics. This represented new direct evidence for conformational control in photosynthesis.

The subsequent project applied the statistical approach developed for the RC BChls to the four haems of the RC-cyt to determine whether their conformations contributed to their individual redox potentials (E_ms). The RC-cyt is tethered to the RC proximal to the special-pair and reduces the cation radical formed during photosynthetic charge-separation. The RC-cyt contains four haems that are arranged linearly in a series of alternating high-/ low-E_ms. In this case there was no obvious correlation of the individual cofactor conformations

with their experimental E_ms and so density functional calculations (B3LYP / 3-21g to 6-311+g**) were employed to calculate the effect. A previous report in which the E_m of a haem-NO and -O_2 binding domain (H-NOX) was lowered by mutations that altered the steric environment of the haem alone, and therefore only its conformation, allowed the discovery that when all else is equal, the haem-Fe's partial atomic charge (Q_{Fe}) is highly correlated with its E_m. Subsequently, Q_{Fe} of the RC-cyt haems were calculated and the calibration curve used to quantify the conformational contribution to the inter-haem E_m differences. When these results were combined with the previous state-of-the-art calculation of their individual E_ms the experimental agreement improved significantly. The effect was found to account for up to 70 % (54 mV) of a particular inter-haem potential difference and these results suggested that accounting for conformational differences amongst haems was a missing-link in haem-E_m variation.

Lastly, on-going research combining semi-empirical molecular modelling with a statistical experimental design is beginning to unravel the influence of BChl conformations upon site-energy inhomogeneity in the FMO protein. A full factorial series of partial molecular optimisations of a BChl model, starting from the crystal structure geometry, using the well-established PM6//ZIndo/S method suggests that the major structural contribution to the intrinsic site-energy (*i.e.*, *in vacuo* excitation energy) is the macrocycle skeletal conformation and not the acetyl orientation, as is presently thought. The approach affords total delineation of complex structural effects, including their interactions, and may provide the definitive answer as to how BChl conformational parameters influence this property. At present, it has been demonstrated that differences in the macrocycle conformations alone accounts for approximately 80% of the variation in the calculated excitation energies across all pigments from all structures. An outstanding difficulty is selecting which PDB structure represents the most accurate representation of the pigment conformations and hence provides the most accurate intrinsic site-energies; aggregation of the experimental data by averaging may not be appropriate in this case but has not been entirely ruled out.

To summarise, this thesis extends the concept of conformational control to photosynthetic systems, namely the bacterial RC, RC-cyt and the FMO protein and these results serve as a general model for this phenomena in other proteins. Additionally, many of the analytical strategies (*e.g.*, cluster and principal components analysis; full factorial designs) employed are not common in the field of structural analysis or computational chemistry and thus their use here highlights their potential as powerful techniques in these fields.

Acknowledgements

During the interview for my Ph.D. studentship Prof. Mathias Senge, my future supervisor, commented on my undergraduate grades something along the lines of, "You got an 'A' in organic and that's what matters." However, unbeknown to either of us then it would transpire that the only work I ever done in the group lab was to clean some glassware in my first week! The actual course of my Ph.D. evolved out of what was originally intended to be only the first step: I was to familiarise myself with the conformations of the pigments in photosynthetic reaction centres and that would introduce me to the topic of nonplanar porphyrins and give us an idea of what to try and mimic synthetically. I think I took this first bit of direction a little more seriously than was perhaps intended and I ended up doing it for four years! In that context, I think you can understand the sincerity of my thanks when I say that I'm grateful to Mathias not only for hiring me and giving me a great idea to play with, but for allowing me the freedom to pursue it in my own way, a freedom which played a most important part in making my Ph.D. experience so rewarding and satisfying.

My friends and colleagues in the Senge group played a big part in my time at Trinity. They are a fantastic group of people and I'm thankful for all of their support and encouragement, given freely in spite of all the marathon group presentations I put them through. Outside of work I was lucky to share with them some seriously good craic, from Christmas nights out in Dublin to pool parties in Korean hotels. Here, I'd like also to mention a special thank you to Dr. Aoife Ryan who has been a great friend to me over the past years.

Beyond the Senge group, I would like to thank Prof. Graeme Watson and Prof. Richard Cogdell for their helpful and constructive comments regarding this thesis; Prof. Watson also got me off to a great start in computational chemistry with his post-graduate course and was kind enough to set aside some time to offer some very useful career advice. Here too I would like to acknowledge Science Foundation Ireland for financial support and the Trinity Centre for High Performance Computing for providing compute facilities.

Finally, thank you to my friends and family in Glasgow, and my brother in New Zealand, who provided much needed moral support over the years. And last, to my wonderful fiancé Daisy, who kept me in good spirits and believed I would suceed even when I didn't and was tearing my hair out; it seems you were right.

Abbreviations

AA	amino acid
AHC	agglomerative hierarchical clustering
ALA	alanine
ASN	asparagine
BChl	bacteriochlorophyll
BPheo	bacteriopheophytin
Chl	chlorophyll
CI	configuration interaction
comp.	complete basis
cyt	cytochrome
DFT	density functional theory
EET	excitation energy transfer
E_m	midpoint potential
EPR	electron paramagnetic resonance
EXAFS	Extended X-ray absorption fine structure
ext.	extended basis
ET	electron transfer
ETC	electron transfer chain
EU	experimental unit
FA	factor analysis
FMO	Fenna-Matthews-Olson
GLN	glutamine
H-NOX	haem-NO and O_2 binding domain
HIS	histidine
HOMO	highest occupied molecular orbital
IP	in-plane
LH1	light-harvesting complex I
KDE	kernel density estimation
LH2	light-harvesting complex II
LHCs	light-harvesting complex
LUMO	lowest unoccupied molecular orbital
min.	minimum basis
MBS	minimum basis set Mulliken analysis
MET	methionine
MM/QM	molecular mechanics / quantum mechanics
NICS	Nucleus Independent Chemical Shift
NPA	Natural Population Analysis
NMR	Nuclear magnetic resonance
NSD	normal-coordinate structural decomposition
OEC	oxygen evolving centre
OEP	β-octaethylporphyrin
OOP	out-of-plane
P870	special-pair in *R. sphaeroides* RC
P960	special-pair in *B. viridis* RC
PCA	principal components analysis
PDB	Protein Data Bank
PS	photosynthesis
PSI	photosystem I

PSI	photosystem II
Q_{Fe}	haem-Fe partial charge
QMM	quantum molecular modelling
RC	reaction centre
RC-cyt	reaction centre cytochrome
RexEP	rotational excitation energy profile
RR	resonance Raman
SD	standard deviation
SE	standard error
SP	special-pair
TM	transition metal
TP	tetrapyrrole
TPP	meso tetraphenylporphyrin
TRP	tryptophan

List of Publications

Senge, M. O.; Ryan, A. A.; Letchford, K. A.; MacGowan, S. A.; Mielke, T (2014): Chlorophylls, Symmetry, Chirality and Photosynthesis. *Symmetry*, **6**, 781-843.

MacGowan, S. A.; Senge, M. O. (2013): Computational Quantification of the Physicochemical Effects of Heme Distortion – Redox Control in the Reaction Center Cytochrome Subunit of Blastochloris viridis. *Inorg. Chem.*, **52**, 1228-37.

MacGowan, S. A.; Senge, M. O. (2011): Conformational Control of Cofactors in Nature – Functional Tetrapyrrole Conformations in the Photosynthetic Reaction Centers of Purple Bacteria. *Chem. Commun.*, **47**, 11621-23.

Senge, M. O.; MacGowan, S. A. (2010): The Structural Chemistry of Isolated Chlorophylls. In: *Handbook of Porphyrin Science*, (Kadish, K. M.; Smith, K. M., Guilard, R., eds.), Vol. XIIV, World Scientific/Imperial College Press, ISBN 978-981-4280-16-7, pp. 253-97.

Table of Contents

1 Introduction

1.1 Overview: Why, What and How?

1.1.1 Protein-Cofactor Interactions Fine-tune Function

The biochemical processes of life are underpinned at many levels by macromolecular machines called proteins. Proteins are so general and fundamental that through evolution nature has imparted upon them the capabilities to perform functions that range from those of simple structural devices such as the keratins to those that participate as highly specialised components in complex cyclical pathways or biosynthetic production lines. This triumph of nature is possible due to the potentially limitless scope afforded by the modular and hierarchical structure of proteins that proceeds from the primary to quaternary through sequence, conformation, fold (shape) and assembly.

Many proteins also incorporate biophysically active or catalytic small-molecules. These cofactors are often situated in – or even constitute – the 'active site(s)' of the protein and affect or assist in whatever function the protein is required to perform; indeed there was a time, according to Bergmann and Niemann,[1] when it was thought that proteins, *"act merely as the passive carriers of physiologically active, or prosthetic, groups"*.* The chemical structure of the cofactor grossly determines what role it may play and thus, for example, redox active transition metal (TM) containing compounds may be found in electron transfer proteins and TMs with coordinative flexibility are present in ligand transport- or sensing proteins. This additional level of structure extends the scope in which nature may construct proteins even beyond that alluded to above.

In spite of this fact, although complex the set of known proteins is fathomable and amenable to both structural and functional classification.[3-5] This may be indicative that the set of biologically useful proteins is a mere fraction of the structures that are possible. Indeed, the notion of biological utility implies the presence of functional driving forces that direct the evolution of proteins and also that the set of useful functions is finite. With this in mind, functional variation in some cases may have been restricted to adaption or *fine-tuning* since the time that the biochemical paradigms for life on earth had evolved. The

---

* Notably, an early discovery[2] of a prosthetic group that was truly essential for enzymatic function was the haem cofactor in catalase.

structural principles of proteins and cofactors that have been mentioned so far are also relevant in this context in the limit of small perturbations characterised for example by minor biosynthetic modification or a few sequence substitutions. Additionally, in conjugated proteins covalent and non-covalent interactions between the protein and cofactor may be considered as modulators of their chemical and physical properties providing an additional mechanism beyond purely structural considerations for the fine-tuning of function.

The way in which proteins have attained such a central role in the processes of life is a result of the diversity and complexity that emerges from the opportunities afforded by the many levels of structure available to affect their construction and impart function. Amongst these possible 'design principles' the fine-tuning of cofactor properties *via* its interactions with the apoprotein plays a key-role in ensuring efficient performance and proper adaption to a particular environment. Whilst thorough investigation of this principle is required to move towards a complete understanding of the structural aspects that influence protein function, basic research in this area is also a pre-requisite to developing strategies to achieve the same level of functional mastery displayed by nature through proteins in systems created for technological applications that may benefit from the so-called biomimetic approach.

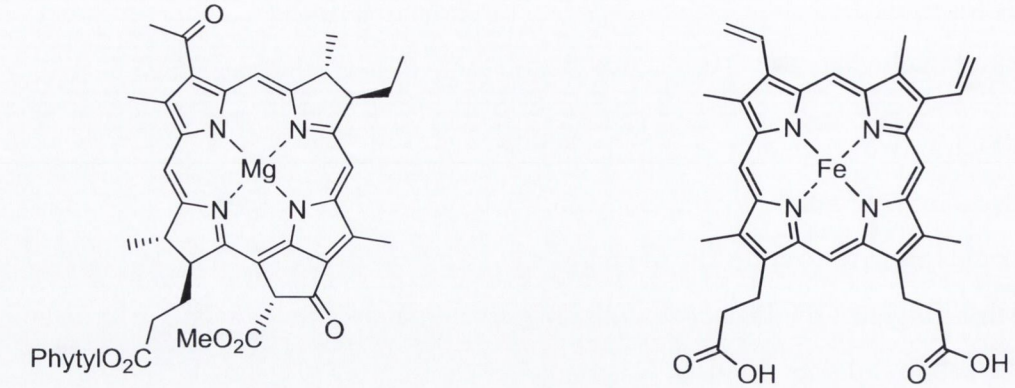
1.1.2 Tetrapyrroles are Ubiquitous in Nature

Tetrapyrrole containing proteins are a particular class of proteins that contain cofactors related to the porphin nucleus. Aside from the ubiquitous haems and chlorophylls, which have been dubbed "The Colours of Life",[6] this includes proteins that contain as prosthetic groups the vitamin B_{12}-derivatives[7, 8] (cobalamins), cofactor F430[9-16] or sirohaem, to name only a few of their perhaps less commonly thought of cousins. Additionally, there are many other proteins whose substrates are one or a group of these cofactors or else their immediate precursors such as the haem degrading protein IsdI[17] or ferrochelatase, as examples of each, respectively.

The diversity of these proteins is impressive since they encompass whole swathes of biochemistry including respiration, photosynthesis, ligand-sensing and -transport, metabolism, metal-sequestration and much more. Of great significance is the fact that many of these processes rely upon fundamentally different chemistries and thus properties that range from ligand-binding constants, redox potentials, absorption maxima, excited-state lifetimes, ion-selectivity and chemical stability are required, where relevant, to be regulated. This begs the question as to how nature has succeeded in managing these

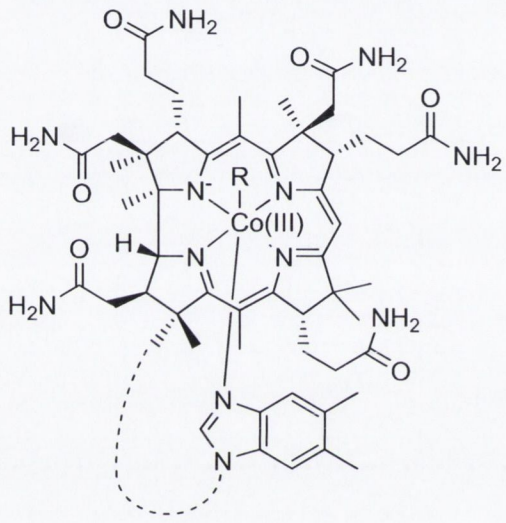
properties for its desired ends, especially when it was noted that the chemical identity of the cofactor largely dictates what it is capable of doing as well as what can be done to it.

Across all tetrapyrrole containing proteins the relevance of the latter point of this question is perhaps moot; nature has provided enough significant variation within the chemical structures of tetrapyrrolic cofactors to endow them with properties suitable for their intended task (Figure 1-1). However, within a particular class, the variation in structure is often quite subtle and does not usually account for the entirety of their functional diversity. In many cases too, chemically identical species are used to perform disparate functions. In these situations, differences between protein-cofactor interactions may be invoked to explain the functional variation that minor differences in chemical structure, or the complete lack thereof, does not. Amongst the possible interactions, hydrogen bonding, electrostatic effects, axial ligation and covalent binding are known to affect both the kinetic and mechanistic properties of catalysis and the energetics of biophysical processes like electron transfer.[18, 19] Higher-order architectural organisation is also critical in determining the role of the protein through controlling its interactions[20] with others and access to the cofactor as well as maintaining defined spatial orientations[21, 22] in multi-cofactor proteins (hence the term, 'protein scaffold'). Whilst these features of protein design are relevant to all or at least most prosthetic groups, perhaps there is something else unique to tetrapyrroles that would provide an additional rationale for their apparent ubiquity? Indeed, the suggestion that there is will form the central argument of this thesis (*vide infra*).

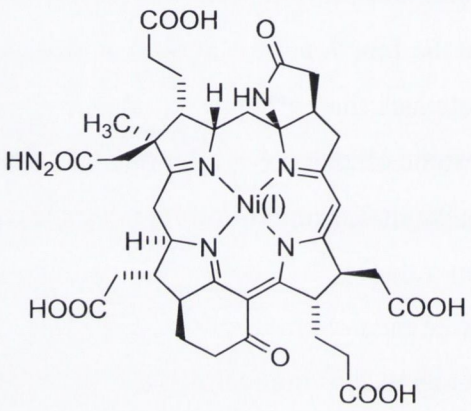


Bacteriochlorophyll a (BChl a)

Haem b



Cobalamins



F_{430}

Figure 1-1: Examples of tetrapyrrole cofactors illustrating structural diversity. Bacteriochlorophyll a is found in the reaction centres and light-harvesting complexes of purple bacteria; haem b is most well-known for its function in haemoproteins; particular cobalamins (vitamin B_{12}) are cofactors in methionine synthase and methylmalonyl coenzyme A mutase and cofactor F_{430} is essential for methanogenisis by methyl coenzyme M reductase.

1.1.3 Tetrapyrroles Have Flexible Macrocycles

Tetrapyrrole cofactors have found such multifaceted use in nature as a result of the properties conferred to them by the aromatic core in tandem with their ability to complex with and stabilise biologically useful metals. Focussing on the former, the delocalisation of electrons across the relatively large surface area of the macrocycle is present also in the charged species generated consequent to electron transfer events allowing the delocalisation of charge so that reorganisation energies are minimised.[23-26] The extended conjugation also results in strong absorptions in the visible region[27-29] that when combined with the substitution patterns observed in the chlorophylls are quite befitting for the harvesting of light energy from the sun.

A superficial consideration of the aromatic conjugation of the macrocycle would perhaps demand the conclusion that it must remain planar. However, this is not necessarily the case

(Figure 1-2) and indeed appears to be the exception rather than the rule in protein complexes.[30-34] Concurrently, a basic consideration of the effect of macrocycle nonplanarity on the p-orbital overlaps that contribute to the aromaticity suggests that they may become disrupted and thus produce some alteration of the detailed electronic structure and properties of the macrocycle, whilst the symmetry of the macrocycle is also perturbed by distortion. Perhaps unsurprisingly then, many empirically observed correlations between the extent of nonplanarity and numerous physicochemical properties,[33-40] including all of those mentioned above, have been reported. This knowledge allows one to posit that protein-induced macrocycle deformation of tetrapyrrole cofactors could serve as a mechanism by which a protein may modulate its biophysical or biochemical function.

The concept is not as farfetched as it may sound. Indeed, in a sense there is a relationship between the idea of conformational control and the notions of the "Induced Fit"[41, 42] (the antecedent of Fischer's[43] "Lock-and-Key" theory) and Pauling's concept of the "Activated Complex".[44, 45] In the former, binding of substrate to enzyme effects a mutual conformational change that is necessary for catalytic function and specificity,[41, 42] whilst the latter deals with the stabilisation of particular transition states – manifest in tight binding-constants for their analogues* – is implicated as the origin of enzymatic catalytic acceleration.[44, 45] Similarly, the main goal of this thesis is to present new evidence in support of the idea that the protein-induced deformation of the tetrapyrrole macrocycle of haems and chlorophylls is a mechanism exploited by proteins to impart desired physicochemical properties (Figure 1-3).

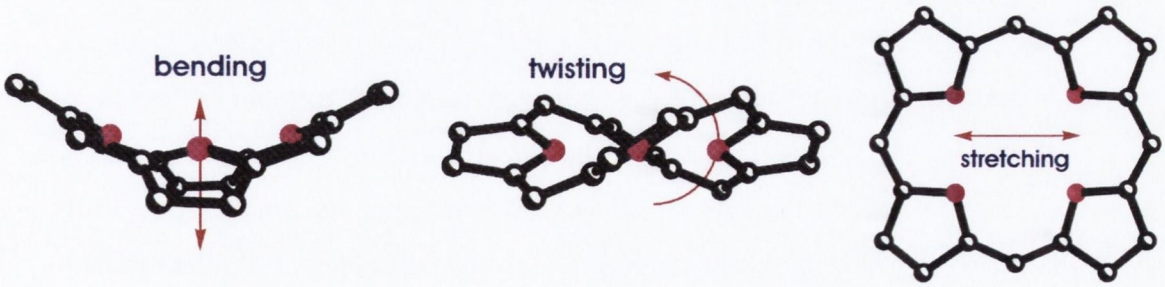


Figure 1-2: Schematic illustrations of some of the characteristic distortions of the porphyrin macrocycle. Figure provided by M. O. Senge.

* An example of conformational control acting in this way is observed in ferrochelatase and its related enzymes (described in Chapter 1.5.3).

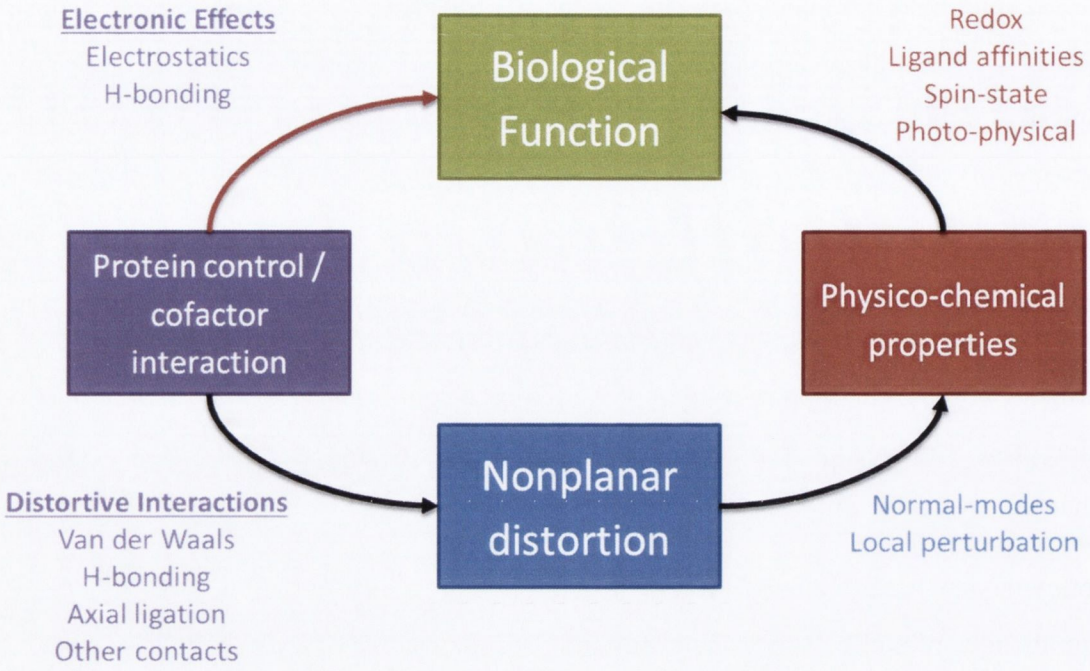


Figure 1-3: Schematic illustration of the proposed mechanism of 'conformational control'. The protein exerts a distortive influence upon the cofactor that induces a specific nonplanar conformation; this in turn modifies the cofactors physicochemical properties and thus impacts its biological function. Direct electronic influences are also utilised by the protein to influence function (*e.g.*, axial ligands, H-bonding and electrostatic effects) that are not a part of conformational control and these too are represented in the diagram.

1.1.4 Photosynthetic Complexity

Tetrapyrroles play many critical roles in photosynthetic systems. They are the primary donors and acceptors in initial photosynthetic charge-separation,[21, 22, 46, 47] the chromophores of the of the light-harvesting complexes[48-51] and the redox cofactors responsible for the re-reduction of the primary donor[21, 22, 52] once their energetic excess has driven subsequent stages of the photosynthetic machinery. The simplicity of this description belies the beautifully intricate and complicated processes involved. However, at this point it suffices to say that these individual events demand drastically different behaviours from the cofactors and although both haems and chlorophylls are involved, the complexes that perform these tasks can do so only by exploiting their architectural utility as scaffolds and their ability to tune cofactor physicochemical properties through protein-cofactor interactions.

Given the significance of photosynthesis to the biosphere of the planet on which we live as well as the importance of understanding the detailed mechanisms of its operation for potential energy applications,[53] there is good reason to study this process in detail. Moreover, because it includes representatives of one of the key-roles that tetrapyrrole (TP) containing proteins play, namely (dark) electron transfer, as well as the unique processes of light-harvesting and photo-induced electron transfer, it presents an ideal system for this

study. An extra incentive for this focus is the relative abundance of structural data, which will be shown to be critical in the following section, coupled with the underrepresentation of the application of the concept of conformational control to this system.

1.1.5 Approaches: Models, Simulations and Measurement

There are several options that present themselves when considering how to go about testing the hypothesis that macrocycle flexibility is in some way related to TP cofactors' diversity of function. One way would be the design and synthesis of model compounds that exhibit different macrocycle conformations and attempt to correlate these conformations with measurements of relevant physicochemical properties. This approach was in fact instrumental in the development of the concept[33, 37] but it is not without its limitations where, on the one hand, conformational effects are confounded with the varying chemical structures and on the other, the compounds often deviate chemically from the cofactors found in nature. Nevertheless, this method is exceedingly useful thanks to the many clever designs that have been utilised to mitigate its drawbacks and although was not used in the present study, publications from such approaches supported many of the interpretations of conformational effects that will be suggested here (Chapter 1.5.2).

A second option is the use of quantum chemical techniques and this method can surmount both of these problems as within the limits of the computing facilities available any conceivable structural model may be implemented. However, this advantage has proved also to be a weakness since the freedom runs the risk of introducing an intractable level of arbitrariness that has no parallel in reality.* This is perhaps an unfairly severe criticism since even studies that have been subsequently shown to be lacking in their structural models were indeed judiciously designed and by virtue of their discordance with prevailing thought ended up highlighting the need for deeper understanding (see Chapter 1.5.4). Another feature that renders such studies imperfect is simply that for many electron wave functions, all contemporary methods yield only approximate solutions to the Schrödinger wave equation so that the calculated electronic structure is only accurate within the context of the model and method employed.[54] Fortunately, modern methods are often sufficiently accurate for many purposes and importantly numerous benchmarks and comprehensive comparisons to experiment are available to assess relevance so that the issue is often

---

* Whilst one could argue that the same could be said for *in vitro* synthetic models of *in vivo* macrocycle nonplanarity, a synthesised model compound always exists in the physical world (providing its structure has been assigned correctly) and so always bears relevance to at least *something* that can occur in reality.

pragmatic in nature when deciding how to balance computational complexity against accuracy and available resources (Chapter 3.4).

During the course of the following research, increasing amounts of quantum molecular modelling was exploited as a core method. All of the structural models employed were derived from published experimental evaluations of the conformations of TP cofactors in protein-complexes so that the first caveat mentioned above was entirely avoided. However, this process itself introduces its own structural bias* into the electronic structure calculations and a unique aspect of the approach presented here was its explicit consideration and mitigation. As for model and method selection, calculations at both ends of the complexity spectrum were performed, from PM6//ZIndo/S to B3LYP/6-311+g**, in different studies.

The last and the most important method to be discussed is the direct measurement of the cofactor conformations in intact protein complexes. To this end X-ray crystallography is often viewed as the definitive experimental technique for the observation of chemical structure and conformation and this view lends weight to the relevance and importance of this thesis. Although capable of resolutions on the order of a few tenths of an Ångstrom (~0.4 Å) for small-molecules, proteins are typically more complex and so 'atomic detail' is not always achieved (notably however, a small protein structure was recently solved to a resolution of 0.48 Å, PDB ID: 3NIR[56]). Indeed, the first effective macromolecular structure determination of sperm whale myoglobin[57, 58] was deemed an achievement worthy of the award of a Nobel Prize as was the structure of the reaction centre,[21, 22] which represented the first integral membrane protein to be analysed by X-ray crystallography.

Since atomic detail is not often attainable (particularly relevant for reaction centres as will be highlighted later), reliance upon good chemical sense in the form of refinement constraints and restraints is an obligation.[59] This leads to difficulties when one wishes to aggregate measurements by averaging because differences in the implementation of these artificial observations introduces significant bias into the final recorded structure. The conundrum that then ensues is how to select the structures that will yield the best representation of the actual cofactor conformations. Fortunately, as X-ray crystallography

---

* For example, Linnanto and Korppi-Tommola[55] found ~100 nm variation in the calculated lowest excited state energy for the same BChl a pigment using the coordinates from different crystal structures.

provides in the end a numerical description of the structure (*i.e.* atomic coordinates) the results of many studies are particularly amenable to statistical analysis.*

The statistical approach provides an unbiased logical rigour for the study as any stated hypothesis must be objectively supported by the data in order to remain upheld. In more practical terms, multivariate analysis allows one to cope with large numbers of observations of many variables and extract the salient features quickly and efficiently whilst simultaneously drawing attention to individual cases that do not fit expectations so that suitable measures may be taken to assess their importance. Another important facet of the statistical approach is that it provides a way to establish the confidence that can be assumed with respect to the uncertainty of an observation such as the variability of a particular conformational parameter of a TP in a protein.

1.2 Scope and Organisation of Thesis

The remaining sections of this introduction outline some of the basic and key features of tetrapyrrole chemistry such as their structure and the diversity of the compound class as well as electronic and physical properties. This is followed by a general overview of some exemplary roles of TPs in biology that is focused on haemoproteins and chlorophyll complexes and the mechanisms for the *in vivo* modulation of physicochemical properties. A critical evaluation of the historical development of the general concept of conformational control through to some recent observations illustrating its role in proteins is followed by a detailed introduction to the major photosynthetic systems that were explicitly examined in this work. This is followed by a short chapter containing the detailed aims, objectives and a summary of the methods and original contributions to scientific knowledge that are presented in the thesis.

Chapter 3 details how the data that were used in this study were obtained and describes the technical and practical details of the normal-coordinate structural decomposition (NSD) procedure that was the starting point for the analysis. The statistical methods that were employed throughout the study and the chemical interpretations of the results derived from these techniques are discussed. The applicability of the computational methods applied in

---

* Note, however, that in this work the coordinates themselves were not directly assessed as they were first transformed by the method of normal-coordinate structural decomposition. This process not only handles the necessary translation and rotation of the coordinate system of the crystal structure to make individual macrocycle conformations directly comparable, but also imparts upon the quantitative comparisons an immediate qualitative interpretation.

chapters 6 and 7 are surveyed with particular emphasis on their proven utility in related systems.

Chapter 4 presents a brief analysis of the numerical stability of NSD to verify its suitability and robustness with respect to error propagation as the foundation for the research presented in this thesis.

Chapter 5 provides an account of the first attempt to perform a statistically sound analysis of the cofactor conformations in bacterial reaction centres (RCs) and the first serious interpretation of the potential biophysical effects. This chapter contains the research published by the author in *Chemical Communications* in 2011[60] with additional details and follow-up research.

Chapter 6 was the first 'routine' application of the method developed for the RC and attempted to answer the specific question of redox potential control in the tetrahaem cytochrome anchored to the RC of *B. viridis*. Quantum chemical simulations were used to quantify and delineate the precise contribution. This chapter is based upon the author's publication in *Inorganic Chemistry* in 2013[61] together with an expansion of the discussion of the relationship of specific macrocycle distortions with haem redox potentials.

Chapter 7 took the previous developments and extended the computational approach using a statistically designed of experiment. This was intended to delineate the conformational contribution to site-energy inhomogeneity in light-harvesting complexes using the FMO protein as the primary model system. Whilst the study is not yet complete, the experimental design and presently available results are described in detail.

Chapter 8 contains discussions of appropriate follow-up studies, the potential impact of the research and re-iterates the major findings.

1.3 Characteristic Aspects of Porphyrin Chemistry

1.3.1 Structure

The discussion of the systematics of porphyrin structure is most easily accomplished by first considering unsubstituted porphyrin. Porphyrin (**1**) is a heteroaromatic compound composed of four pyrrole subunits fused together by methine bridges adjoining adjacent pyrrole α-carbons (*i.e.* it is a cyclic tetrapyrrole). Although all carbons are essentially sp^2-hybridised the macrocycle possesses three distinct types, namely the α- and β-carbons of the pyrrole subunits (C_a and C_b, respectively) and the bridging meso-carbons (C_m) (**1**;

including the IUPAC recommended nomenclature[62]). Conventionally, the aromatic pathway of porphyrin is considered to extend around the periphery and imine nitrogen atoms, excluding the β-carbons of rings B and D as reflected by longer C_b-C_b bonds,[63] and therefore to contain only 18 of the 22 available π-electrons.

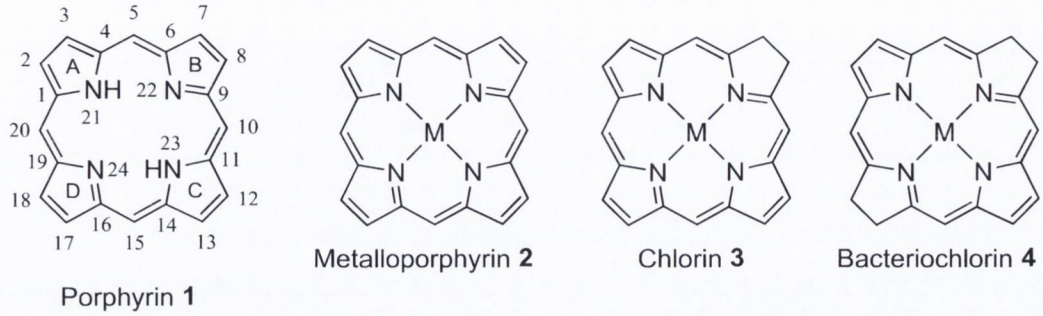


Figure 1-4: Structures of most common porphyrin based macrocycles; **1** including IUPAC numbering.

The cavity at the centre of the core provides a tetradentate square-planar coordination environment upon deprotonation that is capable of coordinating metals from the main group and transition series as well as the lanthanides and actinides.[64] The generalised metalloporphyrin (**2**) belongs to the D_{4h} molecular point group and exhibits the highest-symmetry possible for a porphyrin derivative. Metallated porphyrins may be either unligated or possess one or more axial ligands (typically 1-2) depending on the identity of the metal.

A crucial facet of porphyrin reactivity in biology is that they have multiple stable oxidation states corresponding to sequential reductions of the exo bonds (C_b-C_b) of the pyrrole rings. This allows derivation of, amongst other variants, the biologically important chlorins (**3**) that are saturated solely at ring B and the bacteriochlorins (**4**) where ring D is additionally reduced. These reductions do not disrupt the macrocycle's aromaticity although the conversion of β-carbons to sp^3-hybridisation has obvious stereochemical consequences for the affected rings. Additionally, these reductions decrease the molecular symmetry so that *both* free-base and metallated forms of chlorins are of C_{2v} symmetry and likewise bacteriochlorins are D_{2h} in either form.

Another form of the macrocycle observed in nature is represented by the chemically different corrin class. These compounds exhibit a high degree of reduction and are easily recognised by the absence of a single meso carbon so that two of the pyrrole rings are directly connected *via* a C_a-C_a bond. Vitamin B_{12} is characterised by the presence of the corrin macrocycle (Figure 1-1).

1.3.2 Aromaticity

As aromatic compounds, porphyrins undergo many of the associated electro- and nucleophilic substitution reactions including halogenation, nitration, alkyl lithiation and acylation at the exposed β- and meso-positions. Adding to their structural non-equivalence, these positions have different reactivities that have long been known to be tuned by appropriate selection of the central ion.[6] Of course, this reactivity is not strictly relevant to naturally occurring derivatives where substituents or appropriate precursors are usually incorporated before cyclisation and is mentioned only to highlight the 'typical' aromaticity displayed by porphyrins and the subtle electronic tuning that is possible.

The conventional description of the delocalisation pathway (Figure 1-5) is based on that of a bridged diaza[18]annulene system.[65] However, there are competing hypotheses and an interesting example considers the inner-cross representation coupled with the internal aromaticity of the protonated pyrrole rings A and C (Figure 1-5).[65, 66] This model was derived by quantum chemical calculation of the nucleus independent chemical shift (NCIS)* at various points within the structure of the macrocycle. It was found that the geometric centres of rings A and C displayed NICS constants that were practically identical to those of fully aromatic pyrrole whereas those of rings B and D were comparable to pyrroles with conjugated substituents (*i.e.*, of decreased aromaticity†) and contained short, localised C-C bonds.[66] Furthermore, this description is not static since the relative difference in aromaticity of the rings changes upon perturbation, for example, for symmetry and electronic reasons, metallation reduces the difference between them (Figure 1-5).[66] This latter suggestion finds support in another theoretical study that concluded that the macrocycle ring-current is shifted towards the periphery with increasing charge on the metal.[67]

The above idea is concordant with recent studies although it is worth mentioning that the precise details of porphyrin aromaticity have not yet reached full consensus.[68] However, the conclusion from Wu *et al.*[69] that, *"The "bridged annulene" view of porphyrins does not describe their aromaticity sufficiently... the appended 6π aromatic sextets confer aromaticity much more effectively than macrocyclic 4n + 2 π electron conjugations."* confirms the notion that the local aromaticities of the pyrrole rings are essential. However, it is important to emphasise that it is the origin of the aromatic stabilisation of porphyrins

---

* The NICS is the magnetic shielding present at any point in the electron density of a molecule but is not defined for space occupied for nuclei.

† As in lower aromatic stabilisation energies and greater bond-length alternation.

that has been discussed; the spectroscopic properties of the macrocycle are largely the result of the extended conjugation around the ring (*vide infra*).

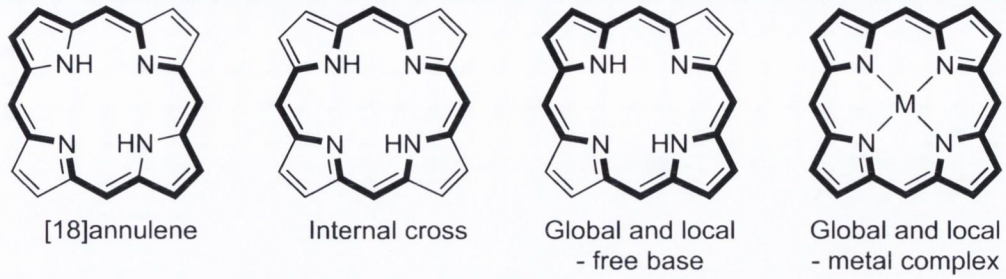


Figure 1-5: Models of macrocycle aromatic pathways. The conventional [18]annulene description (far left); the 'internal cross' used to define the *macrocycle* conjugation pathway in a recent model[5] (left); the aromatic system in free base (right) and of metalloporphyrins (far right) including the 'local' aromaticity of each ring.

1.3.3 Electronic Spectroscopy

The typical UV/Vis spectra of porphyrins[27, 29] exhibit absorptions in two regions termed the Soret or B-band (~380-420 nm) and the Q-bands (~500-600 nm) (Figure 1-6). The B-band corresponds to a strongly allowed excitation of low angular momentum whilst the Q-bands are formally forbidden due to large ΔL_z during the transition. Important vibronic overtones are also present resulting from concurrent vibrational excitation so that the lowest energy transition is more accurately denoted Q(0,0) and the next-lowest Q(1,0), with an analogous convention for the B-bands when resolved. A key-difference between the spectra of metalloporphyrins and their free-base counterparts is the number of Q-bands (Figure 1-6). This results from the reduced symmetry of the free-base (D_{2h}) so that the two axes defined by opposite pyrrole nitrogen atoms (N_{opp}-N_{opp}) are no longer equivalent resulting in further splitting of each vibronic transition. These aspects of the electronic structure of porphyrins and their derivatives are well described by the semi-quantitative MO-based "four-orbital model" of Gouterman,[27-29] which also qualitatively predicts the observed consequences of other structural perturbations, as described below.

The four-orbital model considers the two HOMOs and LUMOs of porphyrin derived from Hückel theory and mixes the four possible excitations between them using a simplified treatment of configuration interaction (CI) to account for electron interaction.[27-29] Whist the e_g LUMOs are found to be quantitatively degenerate, comparison of the nodal patterns of the Hückel a_{2u} and a_{1u} HOMOs to the $n = \pm 4$ orbitals* from both the free electron and cyclic polyene models suggested that they may similarly be considered "accidently"

---

* That is, the pair of degenerate HOMOs defined by both of these models' sole quantum number (n). For example, in the free electron model n is the angular momentum quantum number and the energy of an orbital is given by $\hbar^2 n^2/2mR^2$.

degenerate. The result of this modification* and subsequent application of a simplified form of CI to the two pairs of now degenerate one-electron excitations between the four frontier orbitals (*i.e.*, construction of the excited states from linear combinations of the single excitations) gives rise to two sets of x- and y-polarised degenerate excited states, one of which is strongly allowed and at higher energy than the other forbidden set, accounting for these aspects of the experimental spectra (Figure 1-6).[27, 29] As for the multiplicities of the Q-bands (two in metal complexes and four in the free base; Figure 1-6) each degenerate excitation has a transition dipole orientated with one or the other N_{opp}-N_{opp} axis and are so termed the Q_x- and Q_y-excitations. Therefore in metal complexes, where these axes are equivalent, the Q_x and Q_y pair is degenerate but in free base porphyrins, the axes are not equivalent and the degeneracy is lifted.

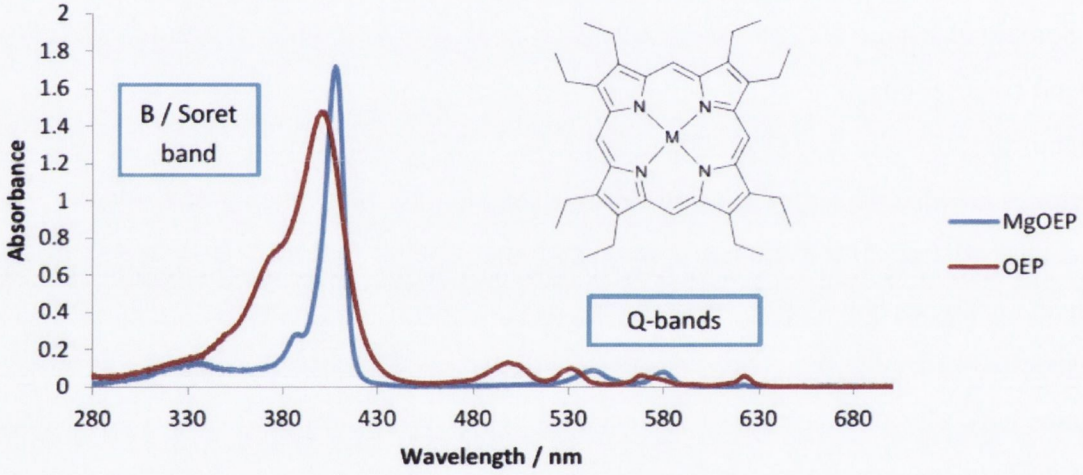


Figure 1-6: The characteristic spectral absorptions of porphyrins. The UV/Vis spectra of 2,3,7,8,12,13,17,18-octaethlyporphyrin in benzene (red; M = 2H) and its magnesium(II) complex in dichloromethane (blue; M = Mg); spectra obtained from PhotochemCAD.[70, 71]

1.4 The Diversity of Tetrapyrroles in Nature

Porphyrins and their derivatives are an important class of compounds in nature where they perform a multitude of biological functions acting as prosthetic groups in proteins. The best-known examples are the haems and chlorophylls. The haems are essential to the function of cytochromes, P450s, haemo- and myoglobins and the peroxidases[18, 72, 73] in which they provide the active-site where the chemistry takes place. Likewise, chlorophylls

* Here referring to treatment of the HOMO and HOMO – 1 as formally degenerate. This is not actually the case in Hückel calculations so that before this modification to the energies of the Hückel MOs was considered the Q- and B-bands were assigned to the lower and higher energy transitions, respectively, which resulted in the failure to account for the different oscillator strengths of the two transitions.

affect the fundamental processes in the light-reactions of photosynthesis by acting as light-harvesting chromophores and the primary agents of charge-separation.[74]

The employment of haem as O_2-transporters, electron transfer agents and the providers of high-valent Fe-oxo intermediates is dependent upon the coordination and redox chemistry of the haem-Fe central substituent. However, the utility of the porphyrin ligand in haem is clearly beyond that of the formation of a stable metal-chelate complex rendering potentially deleterious free Fe less prevalent *in vivo*, or just to provide a 'handle' with which the protein may bind Fe efficiently. In reality, haem is like all coordination complexes where ligands are crucial modulators of the properties of the metal centres.

In contrast, the macrocycle is the unambiguous site of the photochemical and photophysical processes that are required to affect photosynthesis so that in the chlorophylls the presence of the centrally coordinated Mg is intended to influence the properties of the ligand. This is clear not only because of, for example, the knowledge that charge-separation results in π-cations and -anions or the related fact of that the Mg^{2+} ion is redox inert (*i.e.*, does not change oxidation state during Chl redox reactions) but is especially emphasised by the photosynthetic roles of the pheophytins, the free-base derivatives of chlorophylls. Thus, photosynthesis is dependent upon the intense $\pi^* \leftarrow \pi$ transitions of the macrocycle to yield broad action spectra* (tuned during biosynthetic modification to suit the environment of the particular organism) whilst initial charge-separation is possible due to their capacity for photochemically induced electron transfer.

1.4.1 Haems and Haemoproteins

Structure of Haem

There is considerable diversity in the structures of haem cofactors but the most commonly observed in natural systems are haem b, the Fe-chelate of protoporphyrin IX from which all naturally occurring haems are derived, and haem c, which differs from haem b by covalent attachment to its protein *via* thioether linkages formed between the α-vinyl carbons and (typically) cysteine residues in the common CXXCH binding motif (Figure 1-7).[73] Haem b is found in the oxygen binding haemo- and myoglobins, the catalases and peroxidases, b-type cytochromes and all P450 enzymes[75-77] whilst haem c is most prevalent in their namesake the c-type cytochromes.[78, 79]

---

* The spectral range which activates photochemical processes; in this case photosynthesis.

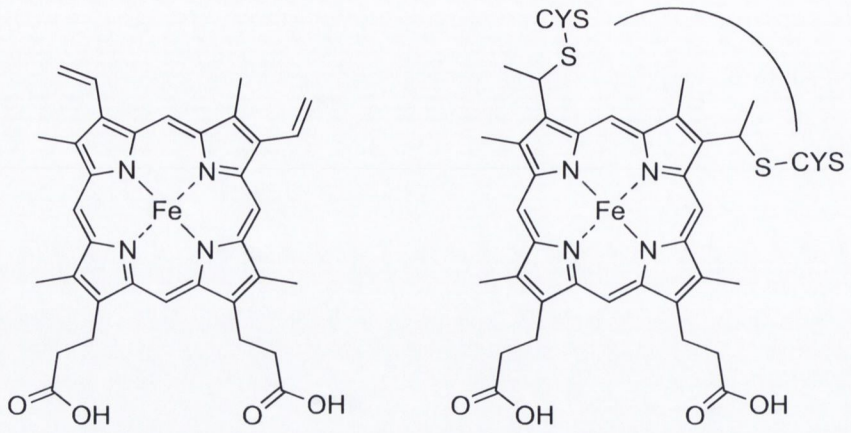


Figure 1-7: Haem b (left) and schematic illustration of protein-bound haem c (right).

Bound haems are always at least pentacoordinate and commonly an additional residue provides an octahedral environment for the haem-Fe. There are examples of haems with relatively exotic axial ligands such as N-termini, proline or cysteinate residues[73] but for haem c at least, usually bis-HIS, HIS-MET and (less often) MET-MET axial ligation is observed,[18] whilst additional aquo ligands may lead to octahedral coordination in otherwise pentacoordinate HIS ligated haems. A detailed analysis of the axial ligand preferences of haems b and -c revealed key differences including the greater prevalence of exchangeable small-ligands in haems b and stereochemical constraints imposed on HIS orientation in haems c.[80]

Haemoprotein Functions

Haem containing proteins are well-known for their involvement in respiration, metabolism and small-molecule sensing across the bacteria, plants and higher organisms. Yet despite their early discovery, new functions continue to be uncovered including roles in gene regulation, cholesterol homeostasis and antibiotic synthesis.[73]

Haemo- and myoglobin are perhaps the best-known haemoproteins and function as the transport and storage vehicles (respectively) for the supply of O_2. The quaternary structure of haemoglobin consists of four subunits, each of which contains a single HIS-ligated haem that is capable of reversibly binding O_2 at the haem-Fe. In an impressive achievement for structural biology, the mechanism of haem co-operativity,[81-83] wherein initial binding of O_2 at one site increases the affinity of the rest, was uncovered to be the result of a substantial decrease in the Fe-HIS bond length upon O_2 binding associated with the altered oxidation level and spin-state of the haem-Fe. This results in the haem 'pulling' on the protein and induces a global conformational change that is communicated to the other sub-units and is responsible for the increased O_2 affinities of the remaining deoxyhaems.

In eukaryotic bacteria and mitochondrial membranes the respiratory electron transport chain generates the proton gradient that drives metabolic oxidative phosphorylation and produces adenosine triphosphate (ATP).[84] Four of the five protein complexes in the sequence between the primary electron donors (NADH and FADH) and the terminal acceptor (O_2) utilise haems as redox centres. In Complex II, the role of the haem (when present) is not fully understood and has been ascribed a structural function in maintaining the fold of the protein.[85] Mechanistic detail is available for cyt bc_1,[86] which accepts electrons from a soluble ubiquinol into haem-B_L of the cyt b subunit directly whilst cyt c_1 is reduced via the 2Fe/2S centre simultaneously as electron transfer occurs from haem-B_L to haem-B_H. Subsequently, haem-B_H reduces a bound ubiquinone, which is released after a second reduction, and cyt c_1 reduces soluble cyt c. Cyt c oxidase (or complex IV), the terminal enzyme in the ETC, accepts four electrons from four molecules of reduced cyt c and uses them to reduce molecular oxygen to water.

Although the precise mechanism of O_2 reduction is not closed to debate, the following (very abridged) account adapted from the review by Michel *et al.* [87] is likely to contain the major features. When fully reduced, cyt c oxidase binds O_2 at the haem a_3 / Cu_b heterobinuclear site where a series of intraprotein electron transfer events and the uptake of two protons results in loss of H_2O and the production of an oxyferryl intermediate ($Fe(IV)=O^{2-}$). Subsequently, another electron and proton are delivered to the haem a_3 – Cu_b site producing a hydroxyferryl species (Fe(III)OH) that produces a second H_2O molecule after receiving an additional proton. This leaves the a_3-Cu_b site in the fully oxidised state that can be re-reduced by cyt c *via* Cu_a and then haem a.

Modulation of Properties

The modulation of haem reduction potentials is of utmost importance for establishing the position of cytochromes in inter-protein electron transfer chains (ETCs) as well as the direction of ET in multihaem cytochromes. In general, haemoprotein reduction potentials exhibit an impressive range spanning 1 V from -550 mV to +450 mV *versus* SHE[88] and are a key-determinant of their function.[89] Considerable success has been achieved with respect to delineating the effect of the axial ligand on key-properties such as reduction potentials. For instance, the horse heart mitochondrial cyt c M80H[90] and *Desulfovibrio vulgaris* cyt c_3 H70M[91] variants, both of which interconvert HIS-MET and bis-HIS haem coordination, reported a 200-220 mV lower potential for the bis-HIS ligated form. This is in reasonable concordance with isolated haem model compounds[92] which show a 150 mV decrease to result from bis-HIS coordination relative to HIS-MET, which is itself remarkably close to

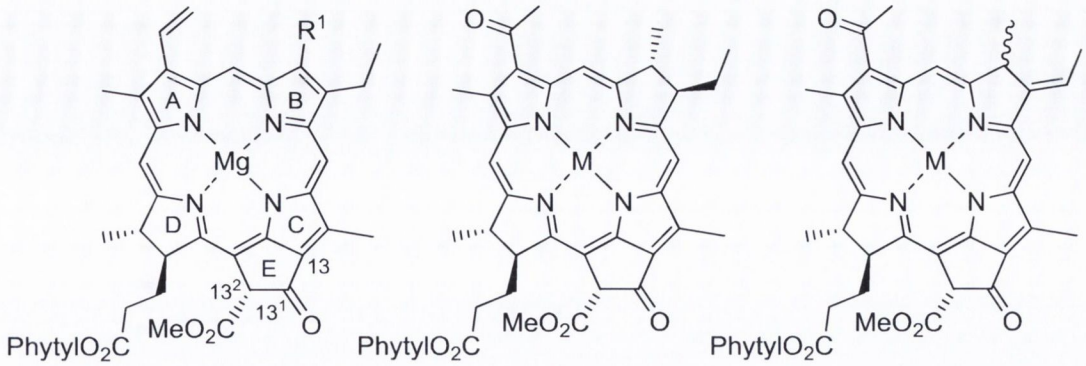
the value of 160 mV found to best describe the difference in a study of 96 haemoproteins containing 141 haems.[89]

There is also a significant influence of the peripheral substituents on the porphyrin ligand.[72] For example, a 50 mV decrease in the midpoint potential of bis-imidazole ligated haem b is observed simply by reduction of the 3,8-vinyl groups, which serves as a basic model for haem c. Moreover, through the *cis*-effect the donation characteristics of the porphyrin may alter the Fe's affinity[93, 94] for axial ligands as well as redox potentials.[72] Reconstitution experiments of native haemoproteins[95] with exogenous model porphyrins show clearly that these effects are relevant *in vitro*, too. For instance, a study that investigated the oxygen affinities of a series of reconstituted haemoglobins with haem derivatives bearing altered substituents at the 3,8 positions (vinyl in haem b) found that O_2 affinity was inversely proportional to the substituents inductive effects.[96] There are many more examples.[95]

In addition to the effects of the porphyrin and axial ligands noted above, many other factors conspire to produce the large potential range, that is equivalent to a ~100 kJ mol^{-1} shift in the free-energy of reduction, and at this point they are quite well understood.[89] These include the protonation state and solvent exposure of the haem proprionates, electrostatic effects of charged-residues in the binding-site and the interhaem redox coupling (*i.e.* the dependence of individual haem potentials on the redox state of nearby haems in multi-haem complexes).[89, 97-100] Thus, whilst variations in chemical structure are important when present, it is most commonly the particular construction of the binding site that serves as the most flexible modulator of haem properties *in vivo*.

1.4.2 Chlorophylls

One of the most important biological processes for life on Earth is that by which plants, bacteria and algae harness the energy of the sun and store it in energy rich reduced carbon compounds; the fundamental building blocks for all the naturally occurring organic matter that surrounds us. Furthermore, oxygenic photosynthesis (PS) was and is responsible for the appearance and continual regeneration of oxygen in the atmosphere that allows the efficient release of this stored energy by life *via* aerobic respiration and allowed the appearance of surface life by eliminating much of the harmful UV radiation antithetical to its mode of existence. The main photosynthetic cofactors (Figure 1-8) are the chlorophylls (Chls) found in oxygenic PS plants, and the bacteriochlorophylls (BChls) present in purple bacteria such as *Rhodobacter sphaeroides* (BChl a; **9**) and *Blastochloris viridis* (BChl b; **10**).



7 R^1= Me
8 R^1= Formyl

9 M = Mg(II)
10 M = 2H

11 M = Mg(II)
12 M = 2H

Figure 1-8: Major photosynthetic cofactors; Chl a (**7**), Chl b (**8**), BChl a (**9**), BPheo a (**10**), BChl b (**11**) and BPheo b (**12**).

At the heart of the initial stages of photosynthesis are the light-harvesting complexes (LHCs) and the reaction centres (RCs), both of which contain numerous Chls or BChls that are crucial to their functions.[46, 47] In LHCs,[50, 51] light is captured by these cofactors in $\pi^* \leftarrow \pi$ excitations and is 'funnelled' *via* non-radiative energy transfer*,[21, 22, 46, 101] toward the RC complex where electron transfer is initiated. Indeed, that the processes of light-harvesting, charge-separation and photo-protection all involve and depend upon numerous tetrapyrrole cofactors is a true testament to their versatility (see section 1.6 for further details).

The high-precision structure of the BChl *a* derivative methyl bacteriopheophorbide *a* provides interesting details regarding the pigment's conformation.[102] The core bond lengths are largely symmetric around the N_A-N_C axis and are well suited to the specific bond-order pattern indicated for BChl *a* above. The C3-acetyl is orientated significantly out-of-plane with respect to ring A (~23°), with the carbonyl oxygen closer to ring B, and the keto oxygen is evidently conjugated to the macrocycle on the basis of bond-lengths in ring E.[102]

The question, "Why Chlorophyll?" has been the subject of two titled reviews.[103, 104] In Mauzerall's 1973 discussion[102] the transition from the requirements of a prototypical

---

* When inter-chromophore distance is greater than ~10 Å, energy transfer dynamics are described by Förster's theory of Fluorescence Resonance Energy Transfer (FRET) for weakly-coupled chromophores wherein non-radiative energy transfer occurs when the oscillating electronic dipole of the excited state of the donor induces an oscillation in the acceptor that corresponds to the separation energy of the latter's ground- and excited-states. However, within LHCs where inter-chromophore separations are small, modifications to this theory are required to allow the excitation energy to be coherently delocalised over groups of strongly interacting pigments. Energy transfer between such strongly interacting systems may then be modelled as FRET between donor and acceptor clusters e.g., LHCs and RCs, respectively. See Sener's review[51] for a succinct description.

photo-oxidising chromophore toward those of a photoreductant are laid out alongside their fulfilment by uroporphyrin and the Chls, respectively. For Chls, the general properties conferred by the macrocycle's excited states are said to be augmented by the inclusion of Mg, which substantially improves its abilities as a photoreductant, and the absorption characteristics enhanced by reduction to the chlorin and the presence of the cyclopentanone ring.* The phytol chain is noted to improve lipid solubility whilst the additional peripheral substituents are considered a mixture of remnants from the biosynthetic pathway and to impart stability to potentially reactive β-positions.

Bjorn's enquiry[104] is more specific and he attempts to explain why Chl a (**7**) above all other Chls has been selected to perform the dual roles of light-harvesting and electron transfer in oxygenic photosynthesis, the latter requiring Chl a to act as the powerful oxidant and reductant in photosystem I and II, respectively. The argument is largely phenomenological and is based on the fact that there are no RCs in oxygenic phototrophs that contain any other type of Chl (with the rare exception of Chl d) and the conclusion is similarly so in the suggestion that its dominance is the result of an earlier appearance than Chl b (**8**). However, another well-known point raised is that despite the many Chls that are present in the photosynthetic proteins, only the few in the RCs are involved in the oxidation-reduction photochemistry. Thus, Bjorn emphasises the prevailing view that the details of the architecture of the protein environment is responsible for determining the precise role of the individual Chls and says that, *"the major tasks of a protein may be to exclude or position functionally compatible electron donors or acceptors, depending upon whether the complex will act as an antenna or a reaction center".*[104]

Chlorophyll-Protein Interactions

Aside from the architectural considerations noted above a number of specific protein-Chl interactions are known to influence their properties. Unlike haem, the axial ligand is not always a particularly dominant factor in determining the key-properties of protein-bound Chls as in the majority of cases the usual 5-coordination is fulfilled by HIS, the major function of which is to affect stable attachment to the protein. In detail, around 50% of protein-bound Chls are believed to be coordinated by HIS, although examples of coordination by GLN, ASN, backbone carbonyls and water are known, and the substitution of such a well conserved feature is likely to be of importance.[105] However, the most

---

* The cyclopentanone ring was also implicated in a unique and interesting mechanism for oxygen generation,[103] which despite the novelty, has not withstood the test of time.

exceptionally unique coordination yet observed appears in photosystem I (PSI) where both second accessory Chls are ligated by the sulphur atoms of MET residues, a feature that has been suggested to allow these Chls to act as the low potential electron acceptors in PSI.[106]

All Chls, and BChls in particular, possess carbonyl groups that are capable of acting as H-bond donors. In Chls, this is most important with respect to the keto oxygen on the cyclopentanone ring whilst additionally in BChls the acetyl provides another H-bond donor that is also strongly conjugated to the aromatic system. The presence of suitable acceptors in binding-sites is therefore one way in which a protein may regulate cofactor properties. Site-directed mutagenesis has been exploited to add and remove H-bonds to the C3-acetyl and C13^1-keto groups of the BChls comprising the special-pair dimer in the *Rhodobacter sphaeroides* reaction centre and have shown that their individual effects on the BChl dimer's midpoint potential are additive at about +90 mV per new H-bond.[107-109]

Another important feature of binding-sites is the presence of any charged residues in the vicinity of the pigment that induce an electrostatic perturbation upon the cofactor. Positively charged residues stabilise the reduced relative to the oxidised form whilst the converse is true for negatively charged amino acids.[19, 110] Slightly more complex is the effect upon excitation energies, where the charge and its placement relative to the transition induced charge-redistribution of the cofactor, may affect either blue-shifted or red-shifted absorptions.[19, 111] Likewise, the charged cofactors associated with intermediate states of charge-separations in RCs have a similar effect on neighbouring cofactors.[112] Additionally, the energetic influence of dielectric relaxation of the protein occurring due to Coulomb interactions with the charged chromophores has been shown to stabilise the charge-separation intermediates.[113]

1.5 Nonplanar Porphyrins and the Historical Development of Conformational Control

Whilst the aromatic π-system of porphyrins may at first imply planar compounds, it has long been known that the macrocycle possesses considerable conformational flexibility. A number of factors conspire to induce these distortions such as coordinated metals with small ionic radii (*e.g.*, Ni); modification of the core by substituents on the pyrrole nitrogen atoms (protonation or methylation) and steric crowding at the periphery.[37] In the case of the so-called highly substituted porphyrins that are characterised by a high-degree of peripheral substitution with sterically demanding substituents,[37] distortion occurs to alleviate the strain imposed by the substituent interactions whereas in metal complexes the

M-N bonds place restrictions on the core size of the macrocycle and can lead to distortion in order to maintain optimal bond lengths. After many structural studies, it was noticed that symmetrically substituted porphyrins often exhibited characteristic, symmetric distortions including the saddled, ruffled and domed conformations.[37, 114, 115] Physicochemical associations of these distortions also had become apparent and included the bathochromic shift of the Q-bands, decreased reduction potentials (easier oxidation, hindered reduction) and a shortening of the fluorescence lifetime of the molecule's excited states.[37] However, the assignment of these effects as arising from the distortions, was debated for some time,[36, 116-118] and settled, in favour of the distortive origin, only quite recently.[36]

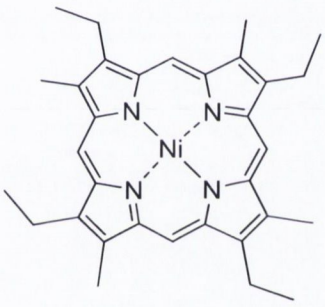
The idea that macrocycle distortion is a mediator of biological function is not new,[30, 31, 33, 119, 120] and considering the associated physicochemical effects of distortion summarised above, in this author's opinion neither should it be surprising. Indeed, that chemically identical species are able to perform such distinct functions *in vivo* participating in numerous biochemical processes *via* fundamentally distinct chemical reactions is quite unique. However, whilst model-compound studies proved the validity of the concept,[30] the first confirmation of these ideas began with the large-scale NSD analysis of haems,[32, 33] which demonstrated the existence of functionally conserved conformations. More recently, successful attempts have been made to use this knowledge to modify a cofactor's physicochemical properties *via* conformational modification induced by site-specific mutation *in vitro*.[121]

Even so, whilst the last large-scale analysis of the phenomenon was successful in obtaining recognition for the phenomenon in haemoproteins,[32, 33] the time lapse between those studies and the present has left the concept of conformational control lacking compared to other protein-cofactor modulatory methods (*e.g.*, H-bonds). Furthermore, whilst there has been significant follow-up in recent times for haemoproteins (see section 1.5.5) there has been very little equivalent consideration of the effect in photosynthetic proteins. The purpose of the following brief review then is to highlight the development of the concept of conformational control, by charting a selection of the advances in the structural and physical chemistry of nonplanar and the explicit application of these results to structural observations in biology.

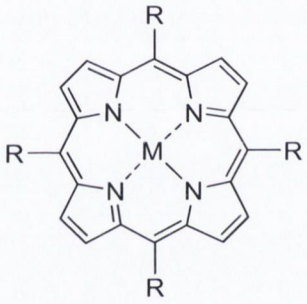
1.5.1 Early Crystallography and Stereochemical Description

Early crystal structure determinations of porphyrins published in 1963 by Fleischer of (2,7,12,17-methyl-3,8,13,18-ethylporphyrinato)nickel(II) (**13**; Ni(II)Etioporphyrin-I)[122] and (5,10,15,20-tetraphenylporphyrinato)copper(II) (**14**; Cu(II)TPP)[123] revealed that both

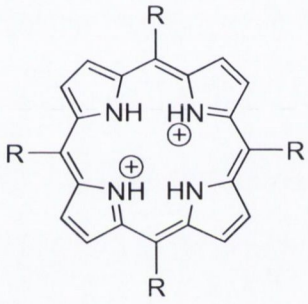
macrocycles were considerably nonplanar. In follow-up work, Fleischer[124] compared the conformations of Ni(II)Etioporphyrin-I, Cu(II)TPP, Pd(II)TPP, Zn(II)TPP and Fe(III)(OH)TPP and concluded that differences in crystal packing and chemical structure were responsible for the dispositions of porphyrins to adopt planar or nonplanar conformations. This work is notable as it may be considered the first comparative analysis of tetrapyrrole conformations in general. Shortly after Fleischer's initial observations, Hoard and coworkers[118, 119] presented their own similar results with the crystallographically isomorphous 5,10,15,20-tetraphenylporphyrin (**15**; H_2TPP) together with a theoretical rationale for the observations that perhaps could be said to mark the beginning of the concept of conformational control.[125, 126] The ideas that were put forward in this work were seminal and are worth summarising as despite the intervening 50 years the crux would appear to remain relevant today.



13 Ni(II)etioporphyrin I



14 M = Cu, R = Ph, Cu(II)TPP
15 M = 2H, R = Ph, H_2TPP
16 M = 2H, R = Pr, H_2TPrP



17 R = Ph, H_4TPP^{2+}
18 R = 4-pyridyl

Until Fleischer and Hoard's observations in the 1960s,[122, 123118, 119] Robertson and Woodward's mid-1930s landmark structure of phthalocyanine,[127-129] indicating a planar macrocycle, had been taken to indicate that the related porphyrins would too be "naturally planar".[125] However, the crystal structure of H_2TPP demonstrated the presence of large displacements of the meso carbons above and below the mean-plane. These were shown not to disrupt the local planarity of several key regions, indicating that the geometric requirements for local π-bonding were retained, but that each pyrrole was not coplanar with the C_a-C_m-C_{ipso}-C_a, dihedral was identified as the, *"principal circumstances adverse to ideal π-bonding"*.[125] Ruffling was suggested to allow puckering of the chelate ring between the metal, two adjacent N-atoms and their connecting skeletal carbons to reduce the strain present in the σ-bonding framework resulting from certain bond angles exceeding 120°. Although the affected reduction of such bond-angles was small in magnitude and appreciable only at meso carbons (as these are not part of a ring), the

observed angle of around 125° was said to be sufficiently removed from the ideal for sp^2 σ-bonding so that any small reduction would be energetically significant.[125]

A crucial observation presented later was that an S_4 ruffled conformation would not necessarily be the only nonplanar conformation likely to be found, *"A convex form (C_4 or C_{4v} symmetry) of the skeleton becomes plausible with longer M-N bonds, by allowing the metal atom to lie far enough outside the plane of the four nitrogens to take care of the added dimensional constraint. This configuration, moreover, could be induced by the additional coordination of an externally constrained ligand to M. The high spin Fe^{+2} and Fe^{+3} derivatives are outstanding candidates for such configurations. Thus Kendrew's observation that, in myoglobin, the iron atom lies more than 0.25 Å. out of the mean plane of the haem group is interpretable in simple fashion."*[126*] This foresight predicts the presence of macrocycle doming in 5-coordinate complexes, which had not yet been directly observed, and moreover suggests that it should accompany the vertical displacement of the haem-Fe that had been observed in myoglobin. In their follow up paper, they also identify the deeper meaning of their suggestion by stating that, in the context of the myoglobin structure, *"...the easy deformability of the porphine skeleton is one of its useful biological characteristics"*.[126]

Later, Silvers and Tulinsky's[130] detailed structure of triclinic H_2TPP revealed that the same compound could exhibit different conformations since their crystal form possessed a considerably flatter core, where the largest deviation of a macrocycle atom from the mean-plane was only 0.10 Å. In this case, the pyrrole protons were located on fixed and opposite rings and the slight deformation that remained was ascribed to N-H repulsion in the core (*i.e.*, the N-H...H-N distance was longer by 0.14 Å than the other N...N diagonal). Tulinsky subsequently presented an average structure for the (planar) free-base macrocycle[63, 131] by combining data from porphyrin (H_2P),[132] 5,10,15,20-tetrapropylporphyrin (**16**; H_2TPrP)[131] and triclinic H_2TPP. The substituted derivatives were noted to have slightly longer C_a-C_m bonds and it was suggested that the spectral differences between H_2TPrP and the other two studied porphyrins could be linked to altered π-electron properties manifest in additional minor structural differences. In discussion however,[63] Hoard suggested packing differences were responsible for the structural differences in H_2TPrP whilst Ader noted that *"increased chlorin character"* was an observed tendency of many meso tetraalkylporphyrins. Six years later, Lauher and Ibers[133] presented the first

---

* By "Kendrew's observation" Hoard is of course referring to the structure of sperm whale myoglobin.[58]

β-substituted porphyrin structure, namely that of (planar) 2,3,7,8,12,13,17,18-octaethylporphyrin (**19**; H_2OEP) that at the time was useful to assess differences between meso- and β-substitution, with the latter more relevant to natural systems. Their analysis also found lengthened C-C bonds at the substituted positions compared to unsubstituted H_2P and the meso-substituted H_2TPrP and H_2TPP.

An important discovery was the structural origin of the red-shifted absorption spectra of porphyrin dications.[134] In particular, crystal structures of some diacid species revealed severe nonplanar distortions of the macrocycle that were due to steric interactions of the crowded core as well as electrostatic repulsion of the partially positive pyrrole nitrogen atoms.[135, 136] Moreover, the enhanced red-shifts observed in H_4TPP^{2+} (**17**) and the dication of 5,10,15,20-tetrakis(4-pyridyl)porphyrin (**18**) compared to other dications (these named compounds are green whereas other dications are usually violet) were attributed to the rotation of the meso phenyl rings into the macrocycle plane, allowed by the nonplanarity of the compounds.[134-136]

Although M-N bond-lengths are the most variable bond-lengths in metalloporphyrins (~1.95 – 2.10 Å), it was noticed early that they are constrained relative to metal preferences in monodentate complexes by the influence of the macrocycle (*e.g.*, an increase of only 0.04 Å is observed in palladium(II) compared to nickel(II) complexes).[134] Furthermore, despite this variability and the associated (yet reduced) variation in the radius of the central hole, other skeletal parameters were observed[137, 138] to be even less flexible such as the in-plane centre to meso distance (Ct.-C_m) and the C_a-C_m-C_a bond-angle.[131, 132] These observations together with structural data for high-spin Fe(III) complexes led to the idea that in these species the coordinated Fe was *necessarily* displaced out of the 4N-plane because of size restrictions.[137] Moreover, on this basis it was suggested,[137] and later observed,[139, 140] that low-spin Fe(III) could be accommodated within the 4N-plane. This proved also to be the case in proteins, emphasised for example by the haems in low-spin ferric cyanomethaemoglobin[141, 142], whilst in contrast, the Fe(II) atom in high-spin deoxyhaemoglobin[81] was displaced by ~0.75 Å out of the 4N-plane toward the axial HIS residue.[143]

Following further advances in the synthesis and characterisation of a great many conformationally distorted porphyrins, Scheidt and Lee[115] offered a systematic evaluation of the contemporary state-of-the-art in 1987 that provided the modern stereochemical classification of nonplanarity used today. Here, the terms ruffled (*ruf*) and saddled (*sad*) conformations were formally defined with *ruf* indicating a twisted macrocycle with meso

carbons alternately displaced above and below the mean-plane and *sad*, obtained by a 45° rotation of the *ruf* conformer around the major C_2-axis, describing a similar alternate displacement of the pyrrole rings relative to the mean-plane. Additionally, the by then well-known domed (*dom*), waved (*wav*; called "stepping" by Scheidt) and "roof" type conformations were also noted.

Figure 1-9: Idealisations of the commonly observed macrocycle conformations described by Scheidt and Lee.[115] In the *sad* conformations (top left), note that the pyrrole rings are alternately tilted up/ down along their C_a-C_a axes producing a saddle shape. In the *ruf* conformation (top right), the meso carbons are alternately displaced above and below the mean-plane that could be affected by twisting a planar macrocycle. In the *dom* conformation (bottom left) all of the pyrrole rings are tilted in the same direction with respect to their C_a-C_a axes resulting in the characteristic dome shape. Finally, the *wav* conformation has one set of opposite pyrrole

Another key idea developed by Scheidt over the previous decade[144, 145] related the geometric parameters of metalloporphyrin cores (*e.g.*, M-N bond lengths) to the electronic configuration of the central ion. Specifically, any complex in which the "stereochemically active" metal d_{z2} or d_{x2-y2} orbitals were populated would exhibit greater M-N bond lengths than a complex with the same or a similarly sized metal where these orbitals were empty. This principle could thus be applied to determine spin-state given structure in ambiguous cases, or else to predict stereochemistry given spin-state, the latter bearing relevance to oxidation state changes during the functioning of haemoproteins.

Additional constraints on the coordination geometry to be discussed were that axial ligand bonds were generally larger than analogous monodentate complexes, especially with larger ligands, because of nonbonding repulsion between the ligand and the porphyrin macrocycle. Likewise, the importance in N-heterocyclic ligands (*e.g.*, imidazole and pyridine) of the dihedral angle between the axial ligand plane and the N_{opp}-N_{opp} axis (first considered by Hoard[143]) and trends in 5- and 6-coordinate complexes were described.

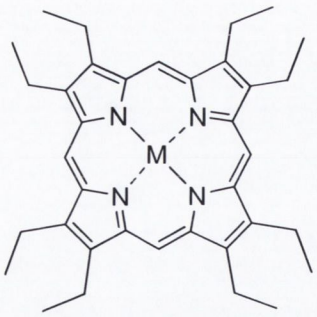
1.5.2 Model Compounds and Structural Correlations

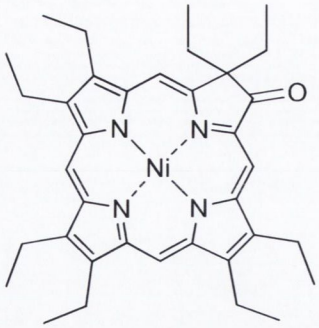
Senge[37] has previously reviewed the numerous ways in which conformationally distorted porphyrins are achieved *via* synthetic design as well as the associated physicochemical consequences. Throughout the following sections many examples of the various approaches that have proved successful will be highlighted, although here they have been presented in an order to systematically illustrate the influence of nonplanarity on physicochemical properties. The reader is referred to Senge's review[37] for a complementary and more comprehensive coverage.

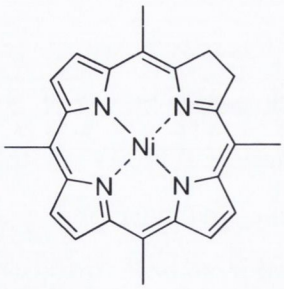
Conformational flexibility

Numerous investigations of Ni(II) derivatives indicated a tendency to *ruf* distortion resulting from core-strain induced by short Ni-N bonds exacerbated in strained or inherently more flexible derivatives and thus Ni(II) porphyrins are commonly found with substantially distorted macrocycles.[114, 146-150] For instance, nonplanar tetragonal[146] and planar triclinic A[147] forms of Ni(II)OEP (**20**) were known in the 1970s, the latter with longer Ni-N bonds, and in contrast to the analogous H_2TPP (see previous section) the tetragonal form was suggested to represent the *unconstrained* structure. A third planar form (triclinic B) was discovered later and was distinguished by the presence of significant π-π interactions that appeared to affect two different sets of Ni-N distances, with the shorter pair in-line with the stack.[150] The π-stacking was noted to be absent in the other crystal forms and least in tetragonal Ni(II)OEP perhaps lending credence to the latter representing the intrinsically preferred geometry. Another relevant result was the observation of ruffled and (relatively) planar forms of Ni(II)2,3,12,13,17,18-hexaethyl-7,7-diethyl-8-oxoporphyrin (**21**) in the same lattice.[151]

Ni(II) hydroporphyrins such as (5,10,15,20-tetramethylchlorinato)nickel(II) (**22**), provided an early illustration of the increased flexibility of reduced macrocycles emphasised by considerable deformation compared to the planar porphyrin.[148] This characteristic was also present in Fe chlorin derivatives and was suggested to be the result of intrinsically greater core-size in the hydroporphyrins that therefore encourages ruffling to shorten M-N lengths.[152] The few other studies of conformationally distorted chlorins, including Kalisch and Senge's[153] series of increasingly β-ethyl substituted tetraphenylchlorin derivatives (see following for the analogous porphyrin series), and compounds with both meso and β-substituents,[154-156] are in agreement with this observation, as it appears that chlorins have a tendency towards more *ruf* distortion.[157]







19 M = 2H, H_2OEP
20 M = Ni, Ni(II)OEP

21

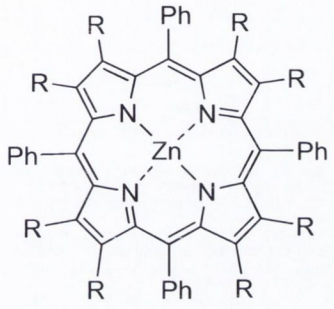
22 Ni(II)TMC

Many other reports of peripherally crowded Ni porphyrins highlight their status as one of the original models of macrocycle nonplanarity.[120, 158, 159] However, the discovery *via* resonance Raman (RR) spectroscopy that the conformational flexibility of Ni(II)OEP observed amongst the various crystal forms was also present in solution deserves special-mention, in particular that the nonplanar form displayed red-shifted absorptions, which is shown to be a characteristic marker of nonplanarity in the following section.[160] This was not the first proof that nonplanar conformations were not the privilege of compounds in the solid-state. For instance, variable temperature NMR[119, 161, 162] experiments and EXAFS[163] measurements had already demonstrated that this could be the case and many other examples[37] are known.

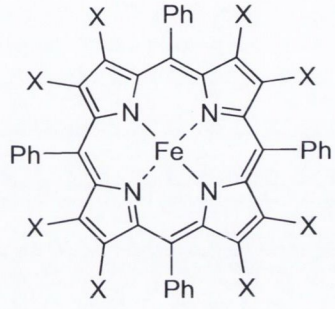
Physicochemical Associations

One of the first direct correlations between nonplanarity with both red-shifted absorption spectra and increasing ease of oxidation was made in connection with compounds of photosynthetic importance and immediately associated with the biological function of such chromophores.[119] In particular, experimental observations and INDO/S calculations upon Zn(II)2,3,7,8,12,13,17,18-octaethyl-5,10,15,20-tetraphenylporphyrin (**23**; Zn(II)OETPP), Zn(II)OEP and Zn(II)TPP demonstrated that nonplanarity destabilises the HOMO and causes the observed differences of the series' absorptions and oxidation potentials and is consistent with the insensitivity of their reduction potentials. These results served as a prelude to assessing intrinsic differences caused by conformational variation of the cofactors of the *B. viridis* reaction centre[119] and later the pigments of the Fenna-Matthews-Olson (FMO) protein.[111] Expansion of this with the synthesis and characterisation of Zn(II)OMTPP (**24**)[161] demonstrated that in these severely *sad* distorted "peripherally crowded" porphyrins the distortion arose from internal steric strain as opposed to crystal-packing effects occurring in dimeric H_2TPP π-cation radicals[115] and on the basis of ^{1}H-

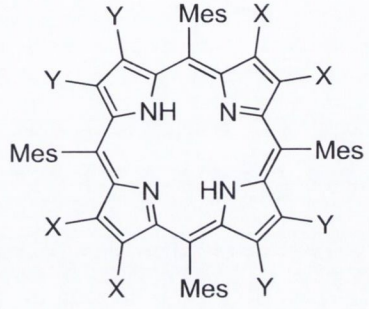
NMR remained present in solution. Interestingly too, in contrast to the conclusion for dications, the concomitant in-plane rotation of the meso-phenyl groups in these species were suggested *not* to affect increased conjugation since meso-phenyl bond-lengths remained unchanged relative to planar compounds.[161] These compounds also exhibited remarkable N-basicity as a result of the exposure of the pyrrole nitrogen atoms due to their *sad* conformations.



23 R = Et, Zn(II)OETPP
24 R = Me, Zn(II)OMTPP



25 X = H_{8-0}, Br_{0-8}, Fe(III)(Cl)(Br_xTPP)



26 X = H, Y = H
27 X = Cl, Y = H
28 X = Br, Y = H
29 X = Y = Cl
30 X = Y = Br

The metal's influence upon macrocycle conformation was investigated in detail revealing that the nonplanarity in strained 2,3,7,8,12,13,17,18-octaalkyl-5,10,15,20-tetraphenylporphyrins (*e.g.*, **23**, **24**; OATPPs) can be relaxed (*i.e.*, the opposite effect of small metals) in order to expand the core to accommodate larger metals and that the detailed influence of specific metals are related between the OATPP and the less encumbered OEP series.[159]

Oxidation Potentials

An intriguing example of the interplay between structural and electronic effects in porphyrins is the trend in oxidation potentials of the sequentially brominated M(Br_xTPP) series.[35, 164-166] Systematic studies have been performed with the free-base, Fe(III)Cl, Cu(II) and Co(II) derivatives showing that in some cases the expected increase of the oxidation potentials with increasing bromination due to inductive effects is superseded by the effect of the increasing nonplanarity of the macrocycle resulting from the increasing peripheral strain. For example, the Fe(III)(Cl)(Br_xTPP) series (**25**) reported by Kadish and coworkers[35] revealed that upon addition of a third bromine the oxidation potential *decreased below* that of the disubstituted product. This trend continued nonlinearly and resulted in the perbrominated compound exhibiting a potential 10 mV *lower* than Fe(III)(Cl)TPP. Estimates of the precise conformational effect were given as 50, 93, 118, 186, 256 and 309 mV for Fe(III)(Cl)($Br_{(3-8)}$TPP) and although not indicated by the authors,

this corresponds to a linear relationship itself, suggesting an average potential shift of ~53 mV per substituent for the conformational effect (R^2 of 0.98). Also, the invariance of the metal-centred reduction with respect to the level of distortion was taken to support the earlier theoretical result[119] that distortion raised the HOMO energy whilst having significantly less effect on that of the LUMO (see above) and/ or to indicate a decreased effect of the conformation on metal-centred processes.

Ochsenbein *et al.*[166] investigated the structures and electrochemistry of 5,10,15,20-tetramesitylporphyins (**26**; TMesPs) with Cl and Br substituents at the "antipodal positions" (**27**, **28**; *i.e.* opposite pyrrole rings tetra-β-halogenated) and the β-octahalogenated derivatives. Both shifts in the RR spectra and UV/Vis absorptions (both Q- and B-bands) were taken to indicate that the distortions were preserved in solution. The observation that the β-octahalogenated derivatives (**29**, **30**) were both *easier* to oxidise than the corresponding β-tetrahalogenated compounds (by -0.10 and -0.13 V for the Cl and Br substituted forms, respectively) was accounted for by the excessive distortion of the former (*sad* with some *ruf*). It is notable that the shift is greater for the Br derivatives since, as could be expected, this compound possesses greater *sad* distortion than that of the analogous Cl derivative with average displacements of the β-carbons from the mean plane of 0.95(2) and 0.905(5) Å,[167] respectively.[166] Additionally, although not severely distorted, the β-tetrabrominated compound possesses the relatively less-common *wav* conformation, where substituted pyrrole rings are bent above and below the mean plane.

Photophysical Correlations

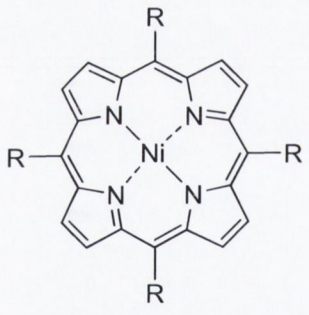
A number of studies have shown that photophysical properties are also significantly affected by the macrocycle conformation.[40, 168-177] These have revealed that nonplanarity results in decreased S_1-lifetimes (τ_{fl}) and fluorescence quantum yields (ϕ_{fl}) as a result of enhanced nonradiative decay rates (k_{IC}, k_{ISC} and k_{CT}) of the $^1(\pi,\pi^*)$ excited state. Holten and coworkers[155, 166] assessed the influence of *sad* and *ruf* distortion on free-base porphyrins using dodecasubstituted porphyrins and meso-tetraalkylporphyrins (the latter with bulky substituents), respectively, showing that both distortions affect strongly "perturbed" properties but that the *ruf* mode leads to more drastic consequences.[168, 178] These perturbations include enhanced Stokes' shifts of 850 – 975 cm^{-1}, decreased ϕ_{fl} and τ_{fl} of 0.003 – 0.009 and 400 – 800 ps, respectively, and shortened IC and ISC lifetimes of around 0.8 – 8.8 ns and 0.4 – 1.7 ns, respectively, for the *sad* compounds (cf. 106 cm^{-1} (Stokes' shift), 0.13 (ϕ_{fl}), 15 ns (τ_{fl}), 74 ns (τ_{IC}) and 22 ns (τ_{ISC}), averages for planar compounds).[168] For the *ruf* compounds, smaller Stokes' shifts of 478 – 486 cm^{-1} were

observed yet only negligible fluorescence was detected with $\phi_{fl} = 1 - 2 \times 10^{-4}$ and τ_{fl} of 8 – 46 ps.[178] Similar trends were also noted in the so-called "basket-handle" porphyrins where distortion is induced by covalently linked meso-phenyl substituents.[169, 179]

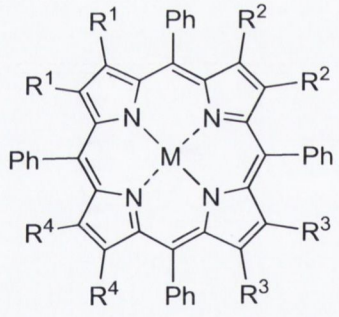
The relationships between the macrocycle conformation and the observed effects were proposed to arise because of greater conformational flexibility of the nonplanar compounds in the excited state.[168, 178] This idea was formed on the basis of the large Stokes' shifts that could not be accounted for classically as solvent reorganisation dynamics should be similar in all of the porphyrins studied. The increased flexibility would also be manifest in the observed increased k_{IC} and this was further evidenced by the fact that the *ruf* distorted compounds exhibited similar properties to planar analogues at low-temperature (78 K). These ideas were confirmed and extended in a follow-up that demonstrated that the *sad* distorted compounds also tended toward more "normal" behaviour at low-temperatures, albeit to a lesser extent than the *ruf* compounds; the trends were also shown to be present in the corresponding zinc(II) complexes.[177]

Unique excited state perturbations were found in Ni(II)T*t*-BuP (**31**) that differ considerably from the above owing to the fact that the $^1(\pi,\pi^*)$ excited state immediately decays to a (d_{z2},d_{x2-y2}) intermediate, a general feature of low-spin d^8 nickel(II) porphyrins.[180] In this case, the (d,d) lifetimes of the highly ruffled porphyrins were orders of magnitude larger than those of planar complexes (*e.g.*, 33 and 40 ns for Ni(II)T*t*-BuP and Ni(II)TAdP (**32**), respectively, and 100 – 220 ps for Ni(II)TPeP (**33**), Ni(II)T*i*-PrP (**34**) and Ni(II)TPP). Furthermore, in Ni(II)T*t*-BuP, the lifetime of the (d,d) excited state exhibited a dramatic temperature and solvent dielectric dependence, ranging over 2 ps to 50 ns in very polar and nonpolar solvents, respectively, and increased to the microsecond timescale at decreased temperatures (*e.g.*, in methylcyclohexane $\tau(d,d)$ is extended to 0.15 μs at 210 K whereas in 2-MeTHF the temperature must be reduced to 130 K to observe a 0.11μs lifetime). These results were rationalised in terms of the M-N repulsion affected by population of the d_{x2-y2} orbital that requires expansion of the core for its relief, a process that is hindered in *ruf* distorted porphyrins (see earlier). It was postulated then that a *dom* or mixed-mode conformer forms in the excited state, which therefore has a large barrier for deactivation to the pure ruf ground state, resulting in a kinetically trapped excited state. Additionally, the strong solvent dependence was taken to indicate significant polarity of the excited state nonplanar conformation and this was suggested to be another feature of some nonplanar conformations in general (*e.g.*, *dom*).

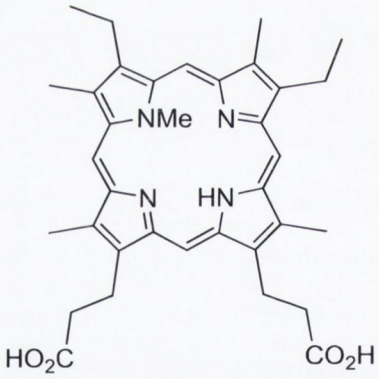
Recently, Röder *et al.*[40] have utilised the accessibility of a series of conformationally designed porphyrins exhibiting a graded degree of *sad* distortion obtained *via* the successive β-alkylation of H_2TPP[181, 182] to systematically assess the photophysical effect. This study represents an additional departure from the "all-or-nothing" approach by assessing the influence of gradually increased nonplanarity in a set of closely related compounds and additionally included the effect of metallation by comparing the free-base and zinc(II) derivatives of the $(C_2H_{5x})TPP$ (x = 2, 4, 6, 8) series. It was shown that the singlet excited state properties were more sensitive to the conformation than those of the triplet and that the onset of the conformational effect was faster for the zinc(II) derivatives. In detail, whilst the Stokes' shift and ϕ_{fl} were altered even in dialkylated Zn(II)DETPP and reached maximum perturbation in Zn(II)*cis*TETPP, the same parameters changed significantly only in H_2*trans*TETPP and levelled out only after hexaalkylation. Furthermore, both series exhibited gradually decreased S_1-liftimes, ISC quantum yields and triplet lifetimes, the latter resulting in dramatically reduced singlet oxygen quantum yields.[40] This is a particularly interesting result as it hints at a potential mechanism for photosynthetic proteins to impart photo-protective abilities on chromophores through induced nonplanarity (see Chapter 5).



31 R = *t*-Bu, Ni(II)T*t*-BuP
32 R = Adamantane, Ni(II)TAdP
33 R = Pe, Ni(II)TPeP
34 R = *i*-Pr, Ni(II)T*i*-PrP



35 R^1 = Et, $R^2 = R^3 = R^4$ = H
36 $R^1 = R^3$ = Et, $R^2 = R^4$ = H
37 $R^1 = R^2$ = Et, $R^3 = R^4$ = H
38 $R^1 = R^2 = R^3$ = Et, R^4 = H
39 $R^1 = R^2 = R^3 = R^4$ = Et



40 *N*-Methyl mesoporphyrin IX

1.5.3 Modern Protein Crystallography and (Vibrational) Spectroscopy

NSD Systematic Assessment of Haem Conformations

In 1995 Hobbs and Shelnutt[31] demonstrated that the haems in a diverse set of c-type cytochromes exhibited a conserved *ruf* conformation by a comparative structural analysis of the then available X-ray structural data. It was shown that the distortion was predominantly induced *via* the CYS thioether linkages to the macrocycle and that this

affected an asymmetric distribution of the pyrrole tilt angles with respect to the mean-plane. Additionally, on the basis of an earlier resonance Raman study that indicated a reluctance of metalloporphyrins with core-sizes ≥2 Å to ruffle,[183] it was suggested that the distortion occurred at significant energetic cost to the protein. Together with the emerging knowledge that porphyrin conformation influenced redox potentials, these results were proposed to imply a functional significance of the conserved conformation. Moreover, they suggested that the oxidation state of the haem-Fe could be communicated through the haem's conformation *via* its protein-contacts, to potentially mediate the protein's interaction with its redox-partners.

In a series of landmark publications,[32-34] the above result was expanded extensively by considering the haem conformations from a diverse set of crystal structures encompassing dozens of proteins and hundreds of crystal structures (including significant redundancy). These advances were greatly assisted by the introduction of the normal-coordinate structural decomposition (NSD) procedure for the analysis of porphyrin macrocycle conformations.[32, 184] This method quantitatively decomposes the observed macrocycle conformation into contributions from the archetypical nonplanar distortion modes (*e.g.*, *ruf*, *sad*, *dom*, *etc.*; see Chapter 3 for a thorough description). Structures of globins, cytochromes c, P450s and peroxidases, amongst many others, were included and both species conservation of haem conformations in related proteins as well as new haem-specific structural differences within multi-haem proteins were revealed.[32-34] For example, subtle differences were uncovered between the α- and β-haems in human deoxyhaemoglobin with the former possessing mostly *dom* and *ruf* deformations and the latter *sad* and *dom* distortions.[32] In comparison, the predominance of the *dom* distortion remained evident in sperm whale deoxymyoglobin although a unique contributing *wav* component was also observed.[34] The analysis of several cyts c_3 lent further credence to the notion that the segment between covalently linked CYS residues was crucial for determining haem conformation since only changes in this region were accompanied by cross-species conformational differences between equivalent haems.[33] Notably, no systematic differences between reduced and oxidised cytochromes could be detected. Another result was that the noncovalently bound haems b in peroxidases were all very strongly distorted with mostly *sad* conformations plus additional *ruf* contributions.[33, 34]

A general result for nearly all proteins was that the haem conformations were well described by only a few of the normal-coordinates of the macrocycle alone (*i.e.*, to within the experimental error) suggesting that the distortions arose from an overall energetic

perturbation from the binding-site as opposed to specific localised contacts.[32] In support of the reliability of the protein structural data for the purposes of comparing detailed macrocycle conformations, it was also reported for some examples that deviations in the NSD results amongst redundant structures resembled the coordinated error of a single atom in the resolved structures.[32]

Ferrochelatase

The gradual unravelling of the distortion mediated mechanism of metallation exploited by chelatase enzymes represents a widely accepted example of conformational control.[185-188] Aside from the fundamental nature of this research it is also biomedically relevant as mutations in ferrochelatase can lead to erythropoeitic porphyria.[189] The idea that macrocycle deformation was a key step in the mechanism of metal insertion in general was originally based on the observation that *N*-methylporphyrins underwent metallation 3-5 orders of magnitude faster than non-methylated derivatives.[190] Soon after it was discovered that *N*-MePs were potent inhibitors of ferrochelatase[191] it was suggested that the sterically imposed nonplanar conformation was similar to a reaction intermediate. This idea was subsequently supported by the generation of an efficient antibody metallation catalyst that had been raised against *N*-methylmesoporphyrin IX (**40**).[192]

The structure of the active-site of ferrochelatase was first identified by Lecerof *et al.*[193] who obtained the crystal structure of the protein bound with both an *N*-MeP substrate as well as the metallated product Cu(II)*N*-MeP after soaking in $CuSO_4$ solution. The enzyme bound the ring A methylated isomer alone, similar to the antibody bound structure although with greater distortion, indicating that tilting of pyrrole ring A contributed to the catalytic mechanism. Pyrrole rings B, C and D were each fixed by numerous contacts with the protein described as "vice-like" and remained relatively in the mean plane although a general mixed *ruf* / *sad* conformation was apparent. Ring A on the other hand was tilted substantially by ~36°, which is greater than that observed in isolated *N*-MePs. Additionally, the structural similarity of ferro- and cobaltchelatases was cited as indicating a ubiquitous mechanism of chelation for all variants.

Sigfridsson and Ryde have performed a theoretical study that specifically addresses the importance of distortions for metal insertion by ferrochelatase.[194] They assessed the energetics of macrocycle distortion of H_2P and a series of metalloporphyrins (MPs) and found that not only were such distortions energetically feasible but that the distortions also provided a release mechanism for the product. These authors confirmed the previous result that ring A of *N*-MeP was distorted further by the protein rather than as a result of the

steric crowding caused by the methyl protruding into the core. It was also shown that the level of distortion observed in the inhibitor complex was likely to be significantly greater than that induced in natural substrates, owing to the difference in the pyrrole tilt minima for that calculated for *in vacuo* compounds, specifically ~30° versus 0°. They also revealed that the dianionic deprotonated form of H_2P was even easier to distort, implying that this intermediate would be energetically stabilised by the distortion imposed by the binding-site.

Comparison of the distortion energies of various MPs established an ordering of Co > Cu > Zn > Fe that was concordant with the fact that ferrochelatase can catalyse the insertion of these other metals into the porphyrin *in vitro* (the high-specificity *in vivo* is believed to arise because ferrous iron is most probably delivered directly to the enzyme by a molecular chaperone). Consideration of the same property of CdP led to the conclusion that it was easier to distort than H_2P providing a mechanistic possibility for the mode of inhibition by Cd. Optimisation of protoporphyrin IX (*i.e.*, the free-base of haem b) in the ferrochelatase binding-site demonstrated that ring A of the natural substrate was also tilted by the protein in the same direction as the *N*-MeP structure and that rings B-D were also (slightly) tilted to form a *sad* conformation.

The proposed mechanism received definitive validation when a Raman mode sensitive to the extent of nonplanarity was shown to be directly correlated with catalytic affinity.[195] Lastly, Ferreira and coworkers[188] have taken these ideas a step further by employing directed evolution to create ferrochelatase variants that have greater selectivity for nickel over iron. These new proteins were shown to induce less saddling than the wild-type protein and thus their previous suggestion that "chelatases distort to select"[187] has proven to be correct.

1.5.4 "Origin of Red Shift" – Theoretical Debate and Controversy

A succinct account of the following debate concerning the true origins of the red-shifted spectra of nonplanar porphyrins was given by Shelnutt and coworkers[36] in the introduction to their 2003 paper that largely settled the argument. This covered the period from the initial scepticism voiced by DiMagno[116] in 1995 until their own work. However, as the course of the following debate considerably advanced understanding of the precise conformational effect they are given special consideration here. Additionally, this discussion will highlight the need for careful interpretations of studies of arbitrarily generated models of nonplanarity and suggest caution in the extrapolation of a few anecdotal results to the wider picture.

DiMagno and coworkers[116] contested the contemporary prevailing view that nonplanarity by itself induced red-shifted absorptions as a result of their findings that the highly nonplanar cobalt(II) tetrakis(heptafluoropropyl)porphyrin and corresponding free-base exhibited blue-shifted absorptions relative to Co(II)TPP and H_2TPP. In support of their conclusion, they pointed out that whilst the solid-state spectra of the MTPP derivatives were substantially red-shifted relative to their solution absorptions* the fluoroalkylated compounds showed identical spectra in both phases suggesting that, *"either the solution and solid state conformations are very similar or the conformational differences are of no consequence"*.

Additionally, in a set of AM1 and AM1/CI calculations utilising dihedral constraints to enforce varying degrees of *sad* distortion, they found that their model predicted a *blue-shift* of the Q(0,0) transition as the distortion increased. Furthermore, the geometry of their fluoroalkyl model was predicted to be significantly nonplanar by a substantial energetic margin so they assumed the solution state conformation to be similarly distorted. However, confusingly they assert that, *"These results indicate that nonplanar conformations predominate in solution and that the absence of red shifts* [compared to TPP] *in the solution and solid phase optical spectra is consistent with nonplanar structures."*[116] This appears to be out-of-line with their reasoning quoted above; had they found a planar structure in solution they would have possessed an example where the planar and nonplanar forms of *the same compound* exhibited the same absorption spectra. In reality though, the weight of their argument was only supported by the fact that *two extremely dissimilar* compounds exhibit differences in their absorptions, one of which *does* show a direct conformational dependence, supplemented by semi-empirical calculations that appear not to be critically assessed in any way.

The first direct response to this controversial publication was provided by Ghosh some five years later.[117] This study used the technically more advanced DFT/SCI method for calculation of excitation energies on geometries obtained from PW91/TZP optimisations. That they also refrained from the imposition of arbitrary restraints to delineate conformational effects, instead opting to use structures obtained from highly substituted porphyrins and replacing the substituents with hydrogens, and verified their results by comparison to experiment were two additional facets of this work that perhaps render it

---

*Taken to indicate planarity of the compound in solution with the later qualification that the red-shift in the crystalline form indirectly originates from nonplanarity only as a consequence of allowing the phenyl substituents to rotate into the porphyrin plane, not an unreasonable idea.

more reliable compared to DiMagno's report. The conclusion of this study was that nonplanar distortions were responsible for the substantial red-shifts of highly distorted porphyrins.

The position of DiMagno's follow-up[118] and response to Ghosh was that substituent induced in-plane nuclear reorganisation (IPNR) was the causative agent of red-shifted absorption spectra and that nonplanarity was related to the phenomenon by, *"collateral contiguity rather than cause-and-effect."* For the series of meso-perfluoroalkylated porphyrins the contraction of the C_a-C_m-C_a angle of the substituted meso, resulting in elongation of the transannular C_m-C_m nonbonded distance as indicated by molecular mechanics simulations, was suggested to be directly responsible for the red-shift. Additionally, uncannily similar relationships were presented demonstrating inverse proportionality between the number of meso-perfluoroalkyl or meso-phenyl substituents with the position of the Q(0,0) absorption. It is convincingly pointed out that these imply little to no influence of ruffling since the near-perfect linearity of the whole series suggests an additive effect for each substituent that does not change with the appearance of nonplanarity. At the same time, because the decrease of the internal angle at the perfluoroalkyl substituted meso is affected upon all successive substitutions this IPNR is ascribed as the causative agent.[118] Furthermore, in order to address the criticism voiced by Ghosh[117] concerning the dissimilarity of perfluoroalkyls compared with phenyl substituents, they embarked on a further computational experiment to ascertain, *"whether the negligible impact of ruffling observed in these studies is an isolated and serendipitous occurrence, or is a manifestation of a more general phenomenon".*[118] To this end, they utilised high-level DFT optimisations (again using dihedral constraints to affect ruffling) and TD-DFT calculations of electronic transitions of porphyrin, 5,10,15,20-tetramethylporphyrin (H_2TMP) and tetrakis(trifluoromethyl)porphyrin (T(CF_3)P). H_2TMP was unsurprisingly found to be natively ruffled in the unconstrained optimisation but the authors make an interesting point when they interpret the significant ease with which the ruffling is extended as indicating an *"exchange of IPNR for torsional strain... nonplanar distortion relieves IPNR".*[118] This appears to be somewhat reminiscent of Hoard's idea that ruffling occurs in order to alleviate strain on the C_a-C_m-C_a angles. However, recalling that ruffling will *decrease* the value of this angle, if IPNR has already done so one would not have expected ruffling to be energetically easier. Indeed, a cursory review of their data shows that this conclusion was not valid because the energy versus dihedral curves are

more or less identical but for the location of the minima.* This would appear to raise issue with their particular highlighting of a shallow potential energy surface for T(CF_3)P, which appears to imply that IPNR can be offset by ruffling here as well.

In their 2003 paper entitled, "Origin of the Red Shifts in the Optical Absorption Bands of Nonplanar Tetraalkylporphyrins", Shelnutt and coworkers[36] sought to ascertain *"...why the observed red shifts could not be reproduced by the constrained porphyrin structures and* [to determine] *their true origin"*. Their approach was thorough and systematic employing a meta-analysis of the data that led to the original suggestion that nonplanar porphyrins were red-shifted as well as a combined MM/INDO study focussing on the Ni and free-base forms of 5,10,15,20-tetra-*t*-butylporphyrin (Tt-BuP) and TMP and concluded with an experimental study of a series of bridled chiroporphyrins. Using an assortment of structures of a diverse variety of free-base and Ni porphyrins, the relationship between porphyrin nonplanarity and absorption spectra was explicitly demonstrated indicating correlation between the total degree of out-of-plane distortion and the Soret position. These observations were supplemented by regarding porphyrins with substituents that induced symmetric distortions (*sad* or *ruf*), with minimal differences in electronic effects, where experimental and INDO derived spectra of the *ruf* series were shown to display a quartic dependence on the degree of *ruf* obtained from MM calculations.

An exceptionally detailed computational study of the Ni and free-base derivatives of Tt-BuP and TMP was subsequently described that involved a complex sequence of structural models that were able to delineate most of the causative agents of the calculated differences in the compounds' B-bands. Electronic differences between the substituents were quantified by replacing meso *t*-Bu groups on the optimised structure with Me groups, similar to Ghosh's and DiMagno's framework, yielding only a small blue-shift leaving the majority of the difference to be accounted for. Next, again in agreement with Ghosh and DiMagno, optimisations performed with incrementally constrained ruffling dihedral angles resulted in only small red-shifts. Crucially, the concept of IPNR was then subjected to an analogous assessment by 1) imposing the bond-lengths found in Ni(II)T*t*-BuP upon Ni(II)TMeP, 2) additionally constraining the C_a-C_m-C_a angle of Ni(II)TMeP to that in Ni(II)T*t*-BuP and 3) in an attempt to maximise IPNR, by fully optimising both compounds under the constraint that they remain planar. Although it is noted that the first two models

---

*The coefficients for a harmonic approximation are 0.00614, 0.00600 and 0.00819 kcal mol^{-1} deg^{-2} to 3 s.f. for P, TMP and T(CF_3)P respectively; their mistake was to compare the energetic cost of distortion of both P and TMP to 40°, not accounting for the fact that the distortion takes P further from its minima than TMP.

did induce small red-shifts of the calculated Ni(II)TMeP absorptions, they were insignificant compared to those that resulted from the ruffle constraints. Moreover, the third model yielded blue-shifted absorptions for Ni(II)T*t*-BuP compared to Ni(II)TMeP suggesting at the very least that IPNR cannot work without access to the full conformational space available to the macrocycle (*i.e.*, in the z-direction) or as the authors consider that, *"... IPNR is unlikely to be the origin of the missing Soret red shift in T(tBu)P, although it may be important for other porphyrin systems"*.[36]

After deducing that the totality of the observed red-shift was not present in either of the artificially constrained structures, the effect of the co-planarity of the meso methyl substituents was addressed. They had noticed that in artificially ruffled Ni(II)TMeP the methyl groups tended to "droop" back into and below the plane of the macrocycle quadrant in which it is bonded. Whilst this effect was found to produce a small blue-shift that somewhat offset the red-shift induced from artificial ruffling, suitably adjusting the model to negate this yielded only a minor improvement to the overall result.

Finally, we come to what was found to be the actual origin of the red-shift. We have as yet delayed the observation that had been made regarding the correlation of the higher-order B_{1u} distortions with the lowest-energy mode of this symmetry (*i.e. ruf*). During the NSD structural analysis of tetraalkylporphyrins it was highlighted that although smaller in magnitude, the $2B_{1u}$ and $3B_{1u}$ deformations were proportional to the extent of *ruf* present. However, because of the greater energy required to distort along these modes their presence even in small amounts was considered to be significant. These distortions were also shown to affect the N-C_a-C_m-C_a torsion angle greater than simple ruffling alone so that they could reasonably be believed to substantially disrupt p-orbital overlap in these regions. Critically, the $2B_{1u}$ and $3B_{1u}$ distortions were reduced in the artificially ruffled models and so a way to include their influence was sought. Thus, two physically justifiable models were selected to uncover whether these higher-order modes would be induced by features other than bulky-substituents and if so, whether they would exhibit significant red-shifts. Specifically, Ni(II)TMeP was optimised with non-bonded atoms added close to the methyl substituents to impart steric strain and then the UV/Vis absorptions were calculated after their removal. Also, alkyl straps of varying lengths were placed across opposing meso positions to afford distorted conformations. The resulting structures, which did indeed possess significant contributions from the higher-order B_{1u} modes, were found to exhibit the majority of the red-shift that was observed experimentally.

It is insightful to end this discussion with a quote from Gouterman in response to a suggestion intended to undermine the four-orbital model, *"I think no theory at this stage of the game can explain all the facts, because they just can't be comprehended... But there are a lot of facts that we do comprehend and* [which theory is right] *depends on how many facts about them we have comprehended and on how many facts the alternative theory comprehends."*[196] It appears that DiMagno's attempt[116, 118] to debunk an idea that had been drawn based on many years of empirical observation was based on only a few isolated cases. Ghosh's concession[197] then resulted from a compelling argument afforded by DiMagno in the form of the constrained ruffling model[118] that was substantiated by his own calculations. However, it was not until Shelnutt's study[198] that the alternative hypothesis (IPNR) was subjected to the same scrutiny that its inventors had laid upon the nonplanarity hypothesis, and at this point the fair comparison, bore out the latter idea.

1.5.5 Recent Biological Studies (Perturbative and Computational)

Haemoproteins

The haem nitric oxide and oxygen binding domain (H-NOX) has become the subject of active research for its endogenous biological regulatory role in small-molecule sensing,[199] its potential as a tuneable model system for biomimetic applications[200, 201] and in particular, for the highly nonplanar conformation imposed by the protein on its single haem cofactor.[121, 202-208] This is one of the most distorted haem cofactors to be observed in natural systems[202] and this feature has been linked to its uncommonly high midpoint potential[203] and is likely crucial to its biological function.

The crystal structure of the H-NOX domain isolated from the obligate anaerobe *Thermoanaerobacter tengcongensis* (Figure 1-10) revealed an exceptionally distorted haem, whose degree of distortion was shown to be related to global structural changes on the surface of the protein, providing a potential link to haem conformation and signal transduction. This work by Kuriyan and coworkers was ground breaking as this crystal structure provided the first atomic resolution picture of a haem binding-site with significant sequence homology to the soluble guanylyl cyclase haem domain in vertebrates.[202]

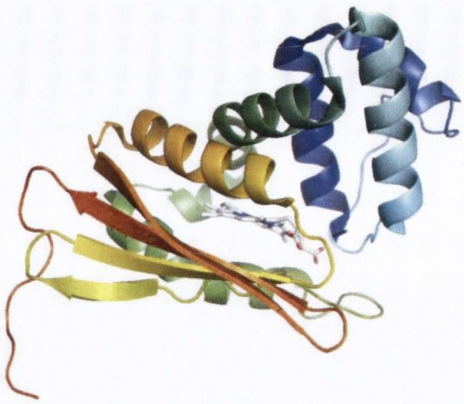


Figure 1-10: View of the H-NOX domain structure from *Thermoanaerobacter tengcongensis* (PDB ID: 1U56; image created using PyMol).[202]

Recently, Marletta's group at UC Berkeley showed that mutation of the conserved PRO115 residue present in very close proximity to the haem (van der Waals contact) to the less sterically demanding ALA residue affected haem relaxation to a more planar conformation in both the solid state *via* crystallography[203] and in solution using resonance Raman (RR) spectroscopy.[204] This structural change also led to increased oxygen affinity and lower midpoint potentials of the complex. The generation of other mutants with intermediate degrees of haem distortion demonstrated that the redox potential was systematically modulated by the haem conformation (Figure 1-11).[121] Other important developments include recent confirmation of the connection between haem flattening and signal transduction in further mutational studies[208] and an early technological application that involved the replacement of a single tyrosine residue to affect loss of the haem's oxygen binding ability for the creation of a NO sensor.[205]

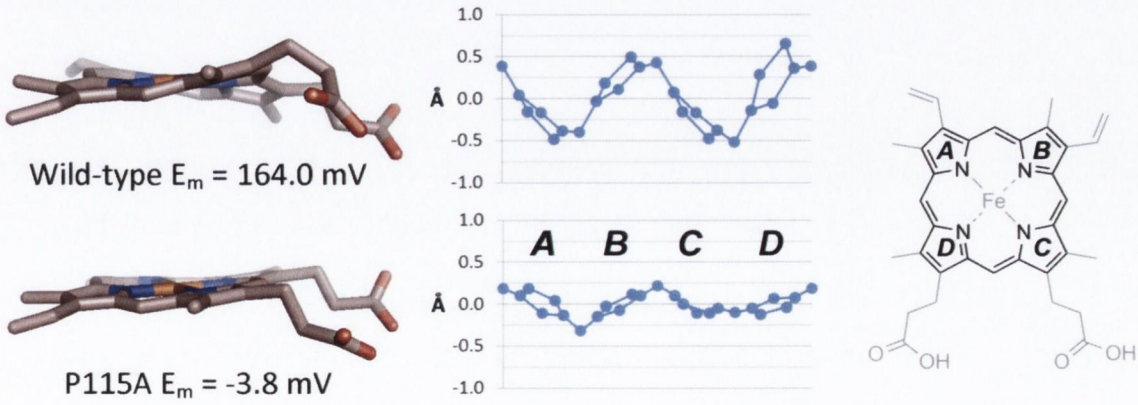


Figure 1-11: The haem conformations in the wild-type[202] and P115A[203] mutant H-NOX complexes (PDB IDs: 1U56 and 3EEE, molecules A and D, both respectively) illustrated by side-on views and skeletal deviation plots, which indicate core macrocycle atoms vertical displacements from the mean plane.

Bren and coworkers have performed a number of studies on the effects of nonplanarity in cytochromes.[209-213] One of their recent achievements in this field has been the development of NMR as a technique for assessing the impact of ruffling on the electronic structure of haem on a per atom basis.[212] This multi-faceted and impressively thorough study combined site-specific mutations* to vary the extent of haem ruffling and to control the haem MET axial ligand's binding mode in *Hydrogenobacter thermophiles* cytochrome c_{552}, with NMR, EPR and DFT to assess the resultant spin-density distributions and the haem-Fe orbital energies. They confirmed that the mutation affected decreased *ruf* distortion *via* ^{1}H NMR and were then able to identify new ^{13}C chemical-shift / ruffling correlations. The specific changes indicated that the spatial extent of the spin-density distribution was decreased at both the haem's meso and β-carbons as a result of ruffling. Notably, EPR suggested that the relative energies and contributions to the HOMO of the Fe *d*-orbitals did not change appreciably, indicating a distinct difference of the effect of *ruf* distortion in this cyt compared to other *ruf* distorted model compounds containing strong π-acceptor ligands where configurations change from $(d_{xy})^2(d_{xz},d_{yz})^3$ to $(d_{xz},d_{yz})^4(d_{xy})^1$. This is an interesting result as it highlights the complicated relationship between the effects of distortion in concert with other environmental aspects.

In a follow-up focussed upon delineating the influence of *ruf* distortion *vs.* Fe(III)-HIS bond strength on the EPR spectra of *H. thermophiles* cyt c_{552} and *Pseudomonas aeruginosa* cyt c_{551}, it was again suggested that haem ruffling was correlated with decreased reduction potentials.[213] However, here the results were confounded by the fact that the reduction potentials of the mutants could also be explained by a change in the Fe(III)-HIS bond strength. However, the effects were considered separable on the basis that the observed change in the EPR axial ligand field term was opposite to what would be expected from changes in the axial HIS bond strength. As a result, the decrease of the axial term was shown to be consistent with increased *ruf* distortion and thus it was concluded that the axial bond strength was mostly invariant across the mutants so that the observed changes in spectral features and reduction potential were attributed to the effects of haem ruffling.

The observation that the haem degrading proteins IsdI and IsdG (iron-regulated surface determinant) produce oxidation products that are very different to those from the typical haem oxygenase enzymes (HOs) has been attributed to the extremely large ruffling

---

* The particular A7F point-mutation was expected to alter the haem conformation through interaction with the Cys-X-X-Cys-His motif.

induced in the protein bound substrate haem.[214, 215] Takayama *et al.*[214] found that IsdIFe^{3+}(CN) possessed exceptionally small average methyl proton chemical shifts and large paramagnetic shifts of the meso hydrogens. This indicated greater delocalisation of the haem-Fe's unpaired electron onto the meso carbons compared to the β-pyrrole carbons and was suggested to result from the dominance of the $(d_{xy},d_{yz})^4(d_{xy})^1$ configuration over the $(d_{xy})^2(d_{xz},d_{yz})^3$ state. This was rationalised as a result of the ruffling allowed overlap of the macrocycle $2a_{2u}(\pi)$ orbital with the Fe $3d_{xy}$, the former having large values at the meso and nitrogen skeletal atoms. Such a change in the electronic configuration has been observed in ruffled FeP model compounds only when strong π-acceptor axial ligands are present although it was posited here to exist in dynamic exchange in the native protein, on the basis of the methyl proton resonances; it is highlighted that the $(d_{xy},d_{yz})^4(d_{xy})^1$ configuration may facilitate oxidative susceptibility of the haem meso positions.* Additionally, the structural consequence of the excessive ruffling appeared to present the haem β- and meso carbons to the oxygen binding site.[214] It was also suggested that ruffling works against differences in the H-bonding and dielectric properties of IsdI compared to typical HOs, which are predicted to raise the potential of IsdI relative to HOs. However, because the observed E_ms are similar, it was also suggested that the increased ruffling lowers the reduction potential, similar to the situation in *H. thermophiles* cyt c_{552} and *P. aeruginosa* cyt c_{551} noted above.

Following this study was a report of a mutant IsdI, with reduced haem ruffling and substantially diminished activity.[215] Since the E_m of the mutant was similar to wild-type the decreased activity was suggested to occur entirely due to the removal of the ability of the protein to substantially ruffle the haem. The TRP66 residue is noted to be conserved in both IsdG and IsdI and is in direct contact with the haem's β-meso carbon and therefore contributes substantially to the conformation. Mutations in this position were previously shown to reduce the catalytic activity of IsdG. In this study, the IsdI variants exhibited macrocycle distortions ranged over 1.3 – 2.3 Å, most likely a significantly greater energetic variation than in the H-NOX study (where the range was up to ~1 Å) and supposed to be enough to alter the electronic configuration. Importantly, the Trp-66 residue was shown not to be required for substrate binding although haem degradation activity was dependent on AAs with large side-chains being present in this position. The electronic

---

* This is a sensible conclusion since placement of the energetic unpaired *d*-electron in the d_{xy}-orbital, which overlaps with the macrocycle $2a_{2u}$ HOMO as a result of ruffling, renders it available to react at the meso positions.

spectra and the pK_as of the distal water ligand of two of the variants contrasted with the wild-type such that they were more similar to classic HOs. These changes were attributed to the decreased ruffling, confirmed for W66Y by X-ray crystallography. In particular, the Soret and Q-bands of the variants were blue-shifted relative to WT, in agreement with expectations but also lending more credence to the distortion / red-shift theory, as these conformation alterations were achieved without chemical modification of the haem. ^{1}H NMR indicated a significantly smaller contribution of the $(d_{xy},d_{yz})^4(d_{xy})^1$ electronic state in W66Y. W66F was neither crystallised nor were its methyl resonances assigned, however, on the basis of the nature of the mutation and the fact that all corresponding paramagnetic shifts were lower field than those in the W66Y, it was suggested that this variant possessed even less ruffling. This is consistent with the idea that ruffling is essential for degradation as the W66F mutant was kinetically slower than W66Y.

Another recent study demonstrated that O_2 affinities in protoglobin are affected differently by nonplanar and in-plane distortions so that increasing *ruf* distortion is correlated with decreased affinity (in line with the results from H-NOX studies discussed earlier), whilst core contraction also results in decreased affinity. [216] This effect is not surprising since core contraction is itself correlated with increased *ruf*, yet interestingly compression (or expansion) along a single N_{opp}-N_{opp} axis also decreases the O_2 affinity. A similar nonplanarity / E_m correlation to that found in H-NOX (*i.e.*, that is somewhat at odds with the result of the previous few conclusions) is suspected in *E. coli* succinate dehydrogenase (a complex II homolog) where its high E_m has been attributed partially to the substantial *sad* deformation of the haem b cofactor.[217] Additionally, it was speculated that the haem conformation may play a regulatory role by exhibiting sensitivity to the redox state of the quinone pool.

Photosynthetic Complexes

Zucchelli *et al.* have attempted to use NSD directly to determine the *intrinsic* site-energy distributions* of Chl pigments in light-harvesting complexes.[218, 219] In this approach, the four-orbitals of Chl that have greatest influence over the Q_y transition are considered to be perturbed in proportion to the extent of macrocycle deformations that possess the same

---

* The site-energy of a chromophore in a multi-chromophore protein is its uncoupled excitation energy and arises due to its interactions with the protein matrix. The *intrinsic* site-energy refers to the site-energy of an individual chromophore in the absence of the protein environment and thus represents the contribution of differences in pigment conformation alone (or if relevant, chemical structure). In multichromophore proteins, the individual pigment site-energies give rise to a spatial distribution throughout the protein.

symmetry. In this scheme then, both E_g (*wav*) distortions affect blue-shifts whilst the A_{2u} (*dom*) and A_{1u} (*pro*pellering) modes affect red-shifts. Although this appears to be a rather crude approximation that neglects the fact that, for example, the *ruf* and *sad* are more commonly associated with red-shifted absorptions it produces site-energy distributions in remarkable agreement with other methods.[219]

Saito *et al.* addressed the influence of differences in the PSII ETC Chls *a* conformations as revealed by the 1.9 Å crystal structure[220] using a modified NSD and a MM/QM approach.[221] They found that the dominant *dom* conformation of the P_{D1} cofactor (conserved from certain RCs) could stabilise the oxidation of HIS ligated Chls generally, but not in the RC specifically. Also, whereas doming was induced by axial ligands alone in P_{D2} and $Chls_{D1/D2}$, an interaction between the peripheral substituents of P_{D1} and the protein affected its conformation. Additionally, greater ruffling of the RC cofactors of the D1 branch (*i.e.*, the active branch) was observed to potentially contribute to electron transfer (ET) asymmetry in the RC. However, whilst all these points were noted to be correspondent with the charge-distribution in $P_{D1/D2}^{+\bullet}$ and ET asymmetry in general, it was suggested that the conformations were likely only indicative of the local steric environment of the cofactor and that the dynamics were affected by electrostatic influences of the protein. This was based on an earlier study regarding the determinants of the hole distribution of $P_{D1/D2}^{+\bullet}$ in PSII where they had found only slight charge asymmetry (favouring $P_{D1}^{+\bullet}$) when calculated without the protein environment, which was attributed in part to differing phytol conformations.[222]

1.5.6 Summary of the Consequences of Macrocycle Nonplanarity

As the conformational flexibility of a great number of porphyrins and a large variety of their physicochemical properties have been discussed, it is useful to provide a concise summary of the currently known effects of general nonplanarity and specific distortions. As has been described, increasing nonplanarity is correlated with porphyrins that have:

- red-shifted absorption spectra
- decreased oxidation potentials (*i.e.* they are easier to oxidise)
- decreased S_1-lifetimes and fluorescence quantum yields
- decreased T_1-lifetimes and ISC quantum yields (that affect reduced $^1O_2$ yields)
- altered axial ligand affinities
- different orbital interactions with coordinated metals

Additionally, in some cases the effects of particular deformations (*i.e.*, *sad*, *ruf*, *etc.*), have been distinguished. In relation to electronic absorptions, the higher-frequency deformation modes $B_{1u}(2)$ and $B_{1u}(3)$ that are correlated with the *ruf* conformation are particularly responsible for red-shifted absorption maxima. This observation may also be relevant to oxidation potentials as the red-shift has been shown to occur due to greater destabilisation of the macrocycle HOMO compared to the perturbation of the LUMO energy. One could also speculate that the higher-frequency deformation modes generally hold greater influence over properties on the basis that they introduce a larger energetic perturbation to be distributed amongst the orbitals and that the greater conformational strain may alter dynamic properties (*e.g.*, relaxation after excitation or oxidation). Another clear cut example was the observation that the perturbations to photophysical properties in nonplanar porphyrins were considerably amplified in ruffled compounds, although it must not be forgotten that they are also majorly apparent in those with the *sad* conformation.

With respect to the influence of nonplanarity in haems, different reports have concluded that nonplanarity is correlated with both increases and decreases in haem redox potentials have been described. The former are represented by the H-NOX series[121] and *E. coli* succinate dehydrogenase,[217] whilst the latter are represented by the cytochromes described in the preceding section.[212-215] It would appear that the latter studies suggest that pure *ruf* distortion affects decreased potentials, however, it must be highlighted that the effect was not directly observed; its presence was assumed on the basis that other active influences were expected to have the opposite effect so that the change in *ruf* was supposed to negate their observation. It appears that only the H-NOX study provides categorical evidence for their conformation / potential relationship. However, perhaps the most prudent position to take is that the precise details of the relationship remain an open question.

1.5.7 Conformational control in non-TP cofactors?

The concept of conformational control is in no way restricted to tetrapyrrole containing proteins. There are many established and a few more emerging manifestations of similar processes in biochemistry that are worth consideration as they help to generalise the phenomena. Broadly speaking, such effects may be classified as those that involve the control of protein conformational dynamics *via* cofactor associated parameters or else the functional control of cofactor properties *via* direct modification of its conformation. These two classifications are illustrated in TPs by contrasting, for example, haem co-operativity in haemoglobin with the maintenance of a high midpoint potential by excessive distortion in the wild-type H-NOX complex. However, as the following example shows, such a

distinction may be academic rather than reflecting any mutual exclusivity occurring in nature.

A classic example is the visual sensing protein rhodopsin, wherein the photo-isomerisation of an 11-*cis*-retinal to the all-*trans* form initiates the visual photo-transduction cascade.[223] However, whilst it has long been known that the apoprotein affects substantially increased yields of the active agonist,[223] a computational study[224] demonstrated that the structural mechanism responsible is the protein-induced distortion of the C11=C12 and C12-C13 dihedral angles in the ground state. Moreover, recent reports have indicated that deleterious mutations in the binding-site known to cause *Retinitis pigmentosa*, a condition associated with blindness, result in substantial changes in the C11=C12 dihedral and significant spectral-shifts of the chromophore.[225, 226]

In another twist, ubiquinone redox potentials have been attributed to the orientation of their methoxy substituents.[227] This study revealed that the reason that only ubiquinones with methoxy substituents may function as both Q_A and Q_B in photosynthetic reaction centres (*i.e.*, 1 and 2-electron acceptors, respectively) is because the protein distorts the Q_B's methoxy by around 20 – 25° out of the aromatic plane, which contributes ~50 mV (65 – 80%) to the observed potential difference. [227] This result is particular interesting in the context of this thesis as it extends the notion of conformational control to other cofactors in reaction centres.

1.6 Photosynthetic complexes

The initial stages of photosynthesis (PS) are naturally divided into two major events. Charge-separation is the point at which sunlight is converted into electrical energy in the form of a transmembrane electrochemical potential that is generated after transfer of an electron from the photo-excited special-pair (SP) chlorophylls along the reaction centre electron transfer chain (RC-ETC). The final acceptor of this electron in the ETC (a quinone) then links charge-separation to the generation of a transmembrane proton gradient that it used to drive the subsequent steps of PS. Of course, nature has not left the initiation of this event solely to chance that the SP is directly excited. Bacteria and plants alike make use of the large arrays of chromophores in light-harvesting complexes that serve as exceptionally efficient antennas for converting light-energy into delocalised excitations that are then directed to RCs, thereby greatly enhancing the rate at which charge-separation occurs.

Additionally, in order to repeat the cycle the oxidised SP cation radical (SP^+) must be re-reduced. This stage is mediated either directly by a sole soluble cyt c_2 or else, in many species, a bound tetrahaem cyt subunit immediately reduces SP^+ and is itself returned to the reduced state by cyt c_2. Over the few past decades tremendous progress has been made regarding our understanding of the processes (and their connections) involved in photosynthesis that has culminated in a unified description of the functioning of a photosynthetic vesicle.[228, 229]

1.6.1 The Bacterial Photosynthetic Reaction Centre

The Electron Transfer Chain

The reaction centre electron transfer chains (RC-ETC) from two species of purple photosynthetic bacteria *Rhodobacter sphaeroides* and *Blastochloris viridis* are studied in Chapter 5 of this thesis. The bacteriochlorin cofactors of the ETC are bound by one of two reaction centre proteins (L or M; synonymous with A and B, respectively), *via* numerous interactions including axial ligation by HIS residues in the case of the four bacteriochlorophylls, and are arranged in pairs of approximate C_2 symmetry (Figure 1-12).[21, 22, 230] There are a few nomenclatures for the individual pigments in the ETC and here the components of the special-pair dimer (P) will be distinguished by the addition of a sub-script indicating the protein to which it is attached (*i.e.* D_L and D_M). This formality is continued for both the monomeric 'accessory' BChls (B_A and B_B) and the BPheos (H_A and H_B) noting the distinction that in bacterial RCs it is conventional to describe the monomers of P using the L/M nomenclature and the accessories using the equivalent A/B labels.

Once P is excited to P*, an electron is transferred to the primary acceptor BPheo (H_A) in ~2 ps and ~3-5 ps, for B. viridis and *R. sphaeroides*, respectively, the beginning of the formation of the cross-membrane electron gradient that drives PS.[46, 231] This takes place through a scarcely detectable $P^+B_L^-$ intermediate that is formed rapidly after photo-excitation. Subsequent reduction of Q_A by H_A^- is followed by electron transfer from Q_A^- to Q_B. After re-reduction of P^+ either directly by cyt c_2 in *R. sphaeroides* or by the RC-cyt in *B. viridis* a second ET cycle takes place culminating in the reduction of Q_B^- to Q_B^{2-}. At this point, the fully reduced and protonated QH_2 dissociates from its binding-site in the RC and is replaced by another oxidised quinone from the cytoplasmic pool.[232] A crucial feature of the RC is that electron transfer occurs only along the L-branch despite the apparent C_2 symmetry.[233]

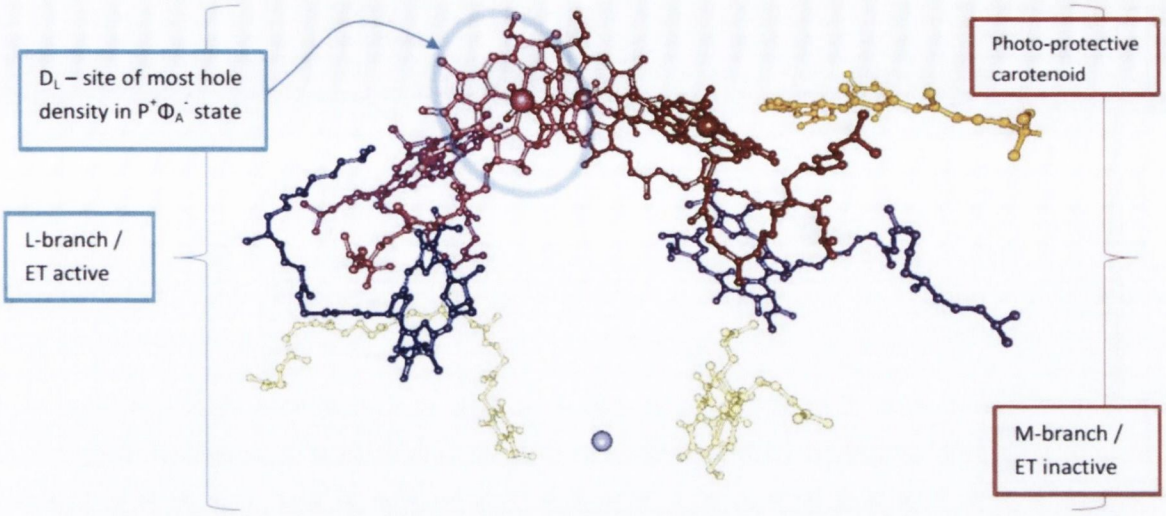


Figure 1-12: The prosthetic groups of the ETC of *R. sphaeroides*. Top left (pink; circled) D_L, top right (red) D_M, middle left (purple) B_A, middle right (dark red) B_B, bottom middle left (blue) H_A, bottom middle right (indigo) H_B, bottom left (yellow) Q_A, bottom right (yellow) Q_B bottom middle (grey) Fe^{2+} and top far right (orange) carotenoid. Note that each L, M (equivalent to A, B) pair is approximately related by a C_2 rotation about an axis which is located in the plane of the page and bisects the ETC into two halves.

There are two striking differences between the RCs from *R. sphaeroides* and *B. viridis*. Specifically, the RC from *B. viridis* utilises BChl b and is in possession of a bound tetrahaem cytochrome whilst that from *R. sphaeroides* contains BChl a and does not have a bound cyt.[46] Other differences include the identity of the carotenoid close to the ET inactive accessory BChl (B_B) and of the quinones (1,2-dihydroneurosporene, menaquinone-9 (Q_A) and ubiquinone-9 (Q_B) in *B. viridis* and spheroidene and ubiquinone-10 in *R. sphaeroides*) as well as the detailed binding interactions of the pigments with the chromophores.[46] There are also significant differences in the static physical properties of the RCs (*e.g.*, P960 in *B. viridis vs.* P865 in *R. sphaeroides*) as well as the kinetic parameters with respect to photo-induced charge-separation between the two species.[46, 231]

Reaction Centre Tetrahaem Cytochromes

Although not present in all species the absence of an RC bound tetrahaem cytochrome serving as the immediate reductant to the photo-oxidised special-pair P^+ has been described as "the exceptions rather than the rule".[234] In *B. viridis* the reaction centre cytochrome (RC-cyt) is observable in crystal structures of the RC bound close to the special-pair and consists of four haem c cofactors arranged in an approximately linear fashion extending from P to the periplasmic space outside of the membrane.[21, 22, 230] The haems are bound by two 'dihaem segments' that relate each pair *via* a local C_2-symmetry axis and the sequence homology of these segments led to the suggesting that the tetrahaem cytochrome originated from a gene duplication event.[234] Each segment binds a one low- and high-potential haem

so that a chain of haems exhibiting alternating potentials is created through to the special-pair (*i.e.*, – 60, 305, 15, 375 mV, SP; Figure 1-13).[235, 236]

Although it is still not known why the potentials alternate in this way it has been shown that the identity of the immediate reductant of P^+ is dependent on the ambient redox potential.[234] In general, and not surprisingly, re-reduction of P^+ is fastest at lower ambient potentials when more of the haems are pre-reduced prior to photo-oxidation of P. In detail, when only the high potential haems are reduced, H3 acts as the donor to P^+ affecting its re-reduction with $t_{1/2}$ of approximately 190 ns and H3 itself is then reduced by H2 with $t_{1/2}$ of 1.7 μs. In terms of the factors affecting such drastic differences in the potentials of the haems, computational studies have achieved accuracy to within ±25 mV *via* electrostatic calculations by considering effects such as the electrostatic influence of charged residues in the vicinity of the haems, the protonation state of the propionates, the identities of the axial ligands to the haem-Fe and the inter-haem redox couplings.[99]

Figure 1-13: Illustration of the RC-cyt from *B. viridis*. The image was created from the crystallographic coordinates found in PDB ID: 1PRC[237] using PyMol; from top to bottom, the special-pair BChls b (P960), haem 3 (H3), haem 4 (H4), haem 2 (H2) and haem 1 (H1).

1.6.2 Photosystems I and II

The coupling of two separate RCs with specifically poised redox potentials[110] so that one is capable of oxidising water to oxygen (PSII) and the other of generating a strong reductant to produce NADPH (PSI) results in the biological marvel that is oxygenic photosynthesis. The Z-scheme (Figure 1-14) illustrates the energetic relationships between the electron transfer events in the light-reactions of photosynthesis that result overall in the reduction of $NADP^+$ by H_2O.[238] Immediately following photo-excitation of the primary donors, the initial charge-separated states are formed in both photosystems ($P680^+/Pheo^-$ and $P700^+/A_0^-$) producing the trapped energetic electrons that participate in the 'downhill' reduction steps and, concurrently, the strong oxidant $P680^+$.[239, 240] The relatively weak-reductant produced by PSII is used to re-reduce $P700^+$ through the interconnecting electron transport chain *via* the membrane mobile quinone that, after double protonation, affects ET to the [2Fe-2S] cluster in cyt b_6f. A subsequent ET step within the cyt b_6f complex sees the reduced [2Fe-2S] cluster pass its electron to oxidised plastocyanin (PC) *via* cyt f, which consequently reduces $P700^+$, whilst the alternative ET pathway mediated by the cyt b_6 component is coupled to cyclic phosphorylation. After the accumulation of four oxidising equivalents in the Mn-containing oxygen evolving complex (OEC; *i.e.*, the S_4-state is reached), water oxidation takes place and the fully reduced state (S_0) of the OEC is recovered. Meanwhile, the ET events downstream of PSII involve the reduction of a soluble ferredoxin that in-turn reduces the flavoprotein ferredoxin-$NADP^+$ reductase (FNR).[238]

Structural similarity of both PSI and II to the RCs is maintained in terms of the overall arrangement and number of cofactors present in the ETC although in both RCs BChls are substituted for Chl *a* and derivatives. Specifically, in PSII[220, 241] there are four Chl *a* pigments, the $P_{D1/D2}$ dimer and the pair of accessories $Chl_{D1/D2}$, and two Pheos *a*, whilst in PSI[242] the special-pair is a heterodimer composed of Chl *a* and its $C13^2$ epimer Chl *a'* and the four accessories and primary acceptors are all Chls *a*. Further critical differences in the biophysical dynamics of these two RCs are apparent. Firstly, that the primary donors are believed to be 'accessory' Chls, specifically Chl_{D1} in PSII[243] and either Chl A_A or $-A_B$ in PSI,[244] contrary to the situation in RCs, even so, the photo-generated cation radicals are localised to the both P680 and P700 after initial charge-separation.[104] Additionally, as hinted by these identifications, ET is not as asymmetric in PSI when compared to RCs from the purple bacteria, although this feature is retained by PSII.[104]

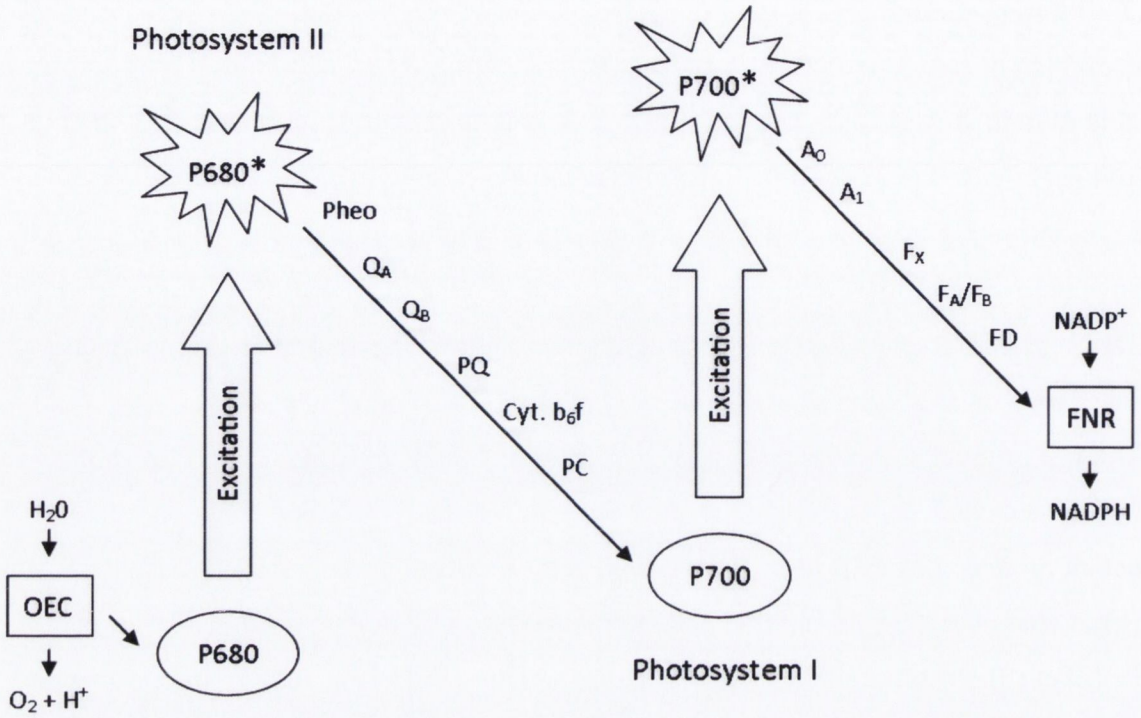


Figure 1-14: The oxygenic photosynthetic Z-scheme. Left to right: oxygen evolving complex (OEC), PSII primary donor (P680), PSII primary acceptor (Pheo), A-branch plastoquinone (Q_A), B-branch plastoquinone (Q_B), plastoquinol (PQ), cytochrome b_6f complex (cyt b_6f), plastocyanin (PC), PSI primary donor (P700), PSI primary acceptor (A_0), phylloquinone (A_1), iron-sulphur proteins (F_X and F_A/F_B), ferredoxin (FD) and ferredoxin-$NADP^+$ reductase (FNR).

1.6.3 Bacterial Photosynthetic Light-harvesting Complexes

The requirement that P is photoexcited before it may serve as the primary donor in the ETC could have been a severely rate limiting step of charge-separation impinging on overall photosynthetic efficiency had nature not devised mechanisms to maximise the rate at which P* is formed. The approach taken was the construction of dedicated light-harvesting complexes (LHCs) that typically consist of densely packed arrays of chromophores whose main task is the conversion of light into electronic excitations and their direction to the RCs. In contrast to RCs that bear comparatively close resemblances regardless of their origin, there is considerable architectural flexibility with respect to the structures of LHCs.

Light-harvesting complexes I and II

The antenna complexes of purple bacteria consist of the RC-associated light-harvesting complex I (LH1; Figure 1-15), or B875, and the peripheral antenna light-harvesting complex II (LH2; Figure 1-16), or B800/850.[245] Both proteins contain numerous BChl *a* pigments held by the apoprotein in strictly defined circular arrangements. Specifically, LH2[246-248] consists of two BChl rings denoted B800 and B850 containing 9 monomeric BChls and 9 pairs of BChl pseudo dimers, respectively, with Mg-Mg distances of ~21 and

9 Å, also respectively, whilst the RC-LH1 complex[249, 250] contains the B875 ring of 30 BChls composed of 15 BChl a dimers with similar Mg-Mg separations to B850. The energetic hierarchy of the bacterial light-harvesting complexes suggests excitation energy transfer occurs in a stepwise sequence: (LH3 →) LH2 → LH1 → RC.[251, 252] Interestingly, differences in the planarity of the BChl macrocycle conformations have been noted.[253]

Figure 1-15: Illustration of the RC-LH1 monomeric assembly; note that it is the dimeric form that is believed to be present *in vivo*. The image was created from the crystallographic coordinates found in PDB ID: 1PYH[249] using PyMol.

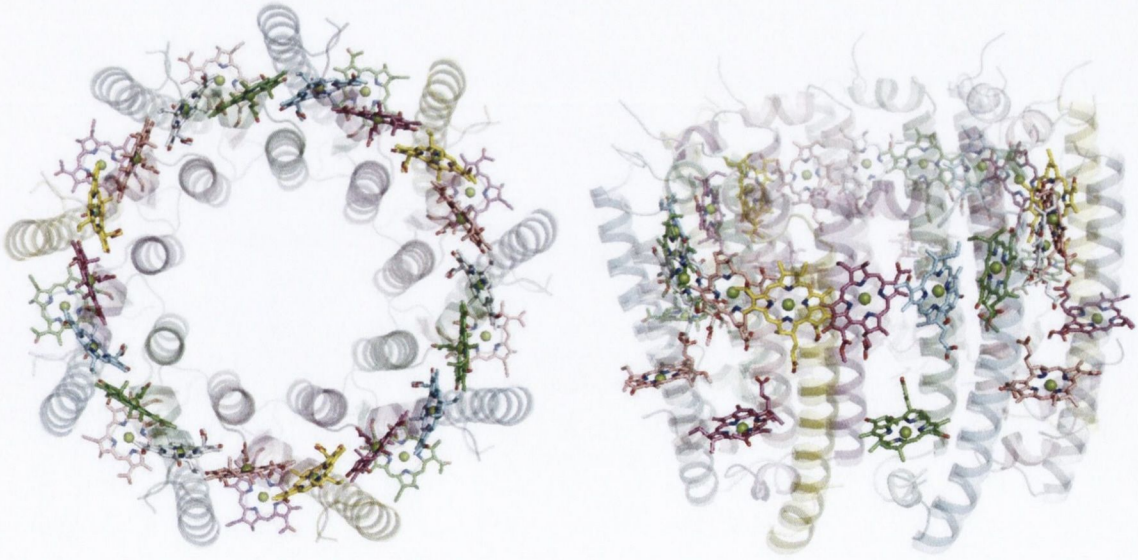


Figure 1-16: Illustration of the LH2 protein, note the presence of the face-to-face B850 and edge-to-edge B800 rings. The image was created from the crystallographic coordinates found in PDB ID: 1KZU[256] using PyMol.

In order to provide environmental adaptability, a modified form of LH2 may be expressed by some species of purple bacteria under low-light conditions known-as LH3 in which the

absorbance maximum of the B850 ring is blue-shifted to 820 nm and hence is termed B820.[245] Structural studies of LH3[254] has led to the suggestion that altered H-bonding to the C3-acetyl of BChl a in the B820 ring was responsible for the spectroscopic variation, an idea that has recently been extended[255] to include the effect that this has on the excitonic coupling in the complex.

Aside from differences in excitonic coupling arising from the various BChl packing-densities present in the B850 and B800 rings that contributes to their spectral differences,[51, 257] the detailed pigment-protein interactions experienced by each BChl may affect unique site-energies (*i.e.*, the uncoupled excitation energies; see footnote p. 44) that influence the exciton dynamics. For instance, computational studies utilising models based on crystallographic data have indicated that the protein environment red-shifts the exciton energies of the B800 ring to a greater extent than in B850.[258, 259] Moreover, direct calculation of the pigment site-energies has shown that differences caused by the pigment conformations alone echo those that are found when more of the protein environment is included. In detail, the isolated pigment models of the LH2 chromophores yielded Q_y energies of 1.59, 1.57 and 1.61 eV for the αB850, βB850 and B800 BChls, respectively, and the relative difference between the B850 BChls was maintained when the effects of neighbouring BChls, axial-ligands or H-bonding residues were explicitly included.[260, 261] An earlier study found even greater variation, suggesting the nonplanar βB850 BChl's site-energy to be ~0.04 eV lower (*i.e.* red-shifted) than its planar αB850 counterpart.[262]

The Fenna-Matthews-Olson Protein

The Fenna-Matthews-Olson (FMO) protein provided the first crystal structure of a chlorophyll-containing protein.[48] The FMO complex serves as the excitation energy transfer (EET) intermediate between the chlorosome baseplate and the RCs of the photosynthetic green sulphur bacteria.[263] The complex consists of seven BChl a pigments encapsulated between an oval-shaped β-sheet and a few α-helix stitches, in what has been referred to as the "taco shell protein", and crystallises as a trimer with (in some species) an eighth BChl situated at the interface of each monomeric unit. It is has been suggested recently that excitation energy passes through the FMO to the RC complex *via* one of two pathways: either the highest-energy (~BChl 8) or the third exciton state (~BChl 1) are initially excited followed by relaxation through BChls 2 or 4 (*i.e.*, alternate routes), and thereafter subsequent decay through BChls 7 and then 3, the latter of which serves as the exit pigment.[264] Depending on whether the initial relaxation is through BChl 2 or 4, determines calculated transfer times of 1.5 or 0.5 ps, respectively.

There are two spectral classes of FMO protein that are differentiated by the relative intensities of their absorption maxima.[265] Type I complexes show a greater intensity of the second lowest-energy excitation and include the species *Prosthecochloris aestuarii* and *Pelodictyon phaeum* whereas in type II species (*e.g., Chlorobaculum tepidium*) the intensity of this excitation is suppressed such that it is lower than the third lowest-energy optical transition.[265] Speculation as to the factors that lead to the spectral classes has included citations of the differences in the macrocycle conformations of the pigments (including nonplanarity and C3-acetyl orientation) and alternatively differences in the binding of BChl 8, specifically its ligation state.[266] This latter suggestion has since been superseded on the basis that it has been recently suggested that the FMO from *P. phaeum* does not include BChl 8 after all,[267] although it is possible that this is a preparative issue as crystal structure determinations usually ascribe this cofactor a low-occupancy and it is considered to be the most labile cofactor.

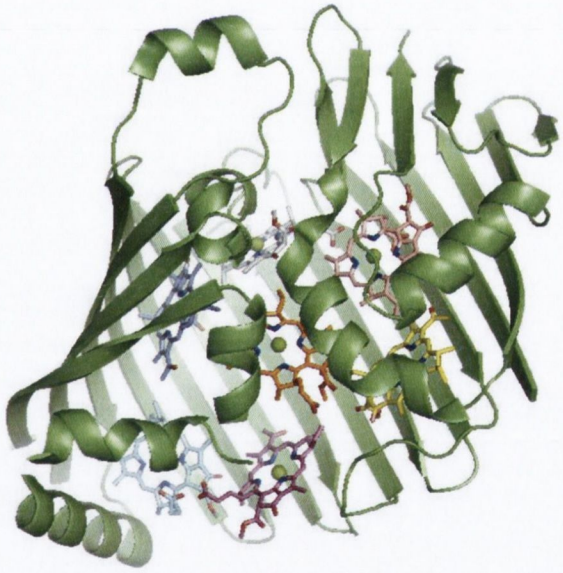


Figure 1-17: Illustration of the FMO protein monomeric asymmetric unit. From bottom left to centre *via* an approximate '9'-shaped spiral: BChl 1 (pale blue), 2 (purple), 3 (yellow), 4 (pink), 5 (grey), 6 (blue) and 7 (orange). The image was created from the crystallographic coordinates found in PDB ID: 3BCL[268] using PyMol.

The FMO protein is considered by many to be a useful model complex for studying photosynthetic energy transfer, Olson himself referred to the crystal structures as, *"...a godsend for physicists and physical chemists interested in spectral properties, excited states, and energy transfer."* As a result, numerous studies of the site-energy distributions and exciton dynamics have been reported,[265] and have included explicit investigations of the influence of variations in BChl conformation.[111] So far, the best agreement between structure-based *ab initio* calculations of the site-energies with those empirically determined

by parametric fitting to the experimental spectra, have been obtained by consideration of the effect of the electrostatic environments of the BChl binding-sites on the relative stabilities of their ground- and excited-states.[269, 270]

Chlorosomes

The chlorosomes are the main photosynthetic antenna complexes in green bacteria. In contrast to the protein-organised complexes described previously, chlorosomes contain very little protein and instead are comprised almost entirely of self-assembled chlorophyll aggregates; a small BChl a binding-protein is associated with the chlorosome may serve as an intermediate between the chlorosome and the RC apparatuses (including the FMO protein) and is referred to as the 'baseplate'.[271] Depending on the particular species, *Chlorobiaceae* (green sulphur bacteria) contain one of either BChls c, d or e (the Chlorobium chlorophylls) and their corresponding homologues, which differ in the length of the alkyl substituents attached to C8 and C12 as well as the identity of the esterifying alcohol, whereas all *Chloroflexaceae* (green nonsulphur bacteria) usually contain the BChl c derivatives.[271]

The particular BChl substitution pattern (*i.e.*, specific homologue) of the compounds isolated from chlorosomes in some species is governed by the light-conditions maintained during culture,* with reduced light leading to increased methylation.[273] This is considered to affect altered spectral properties by modulating the *in vitro* aggregation properties of the chromophores. However, structural studies of the demetallated derivatives of the homologues[119] from *Chlorobium vibrioforme* indicate substantial alterations to the macrocycle conformation, which may contribute to the red-shift of the low-light chlorosomes, although analysis of the spectral properties of the isolated chromophores[274] did not show any systematic variation corresponding to the extent of methylation. Additionally, model compounds have indicated that the C20-methyl substituent induces local nonplanarity at this position, which may contribute to the 10 nm redshift of BChl c relative to BChl d.[157, 275]

---

* An interesting application of the light-adaption process has been the analysis of the distribution of the BChl homologues in sedimentary cores, allowing the comparison of bacterial communities from different locations as well as a characterisation of their development in response to environmental conditions over time.[272]

2 Project Outline

2.1 Motivation: Outstanding Problems and their Solutions

2.1.1 Photosynthetic Systems

Whilst haemoproteins have received significant attention regarding the impact of conformational control,[34, 121, 212-217] there is a clear lack of equivalent studies of the phenomenon in non-haem driven processes. In particular, in photosynthetic systems a number of authors have noted the significance of BChl nonplanarity in photosynthetic complexes,[30, 119, 230, 276-278] although the relevance of conformational control is generally not as well accepted in relation to these proteins as compared to haemoproteins. Nevertheless, the subtle fine tuning of physicochemical properties made possible by macrocycle distortion may be applicable here as well and it is certainly worthwhile to test this hypothesis. This underrepresentation is directly addressed throughout this work by focussing on the conformations of BChls and haems in the reaction centres (RCs) and cytochrome subunits (RC-cyt) from purple bacteria and the light-harvesting Fenna-Matthews-Olson complex from green bacteria.

There are a number of potential reasons why photosynthetic systems have not been as well-studied as the haemoproteins with respect to the influence of protein-induced macrocycle distortion. To begin, as a result of their chemical instability there is a severe lack of specifically designed model compounds of conformational flexibility in chlorins,[157] to the extent that Senge[153] has provided the only substantial structural analysis to date. At present then, it is possible only to address the situation *in vivo*, by the analysis of the conformational flexibility of protein-bound chlorophyll and related cofactors directly. Therefore in this context, this study extends the concept of conformational control to the chlorin and bacteriochlorin class of tetrapyrrole compounds.

Also, analysis of the conformational variability of photosynthetic cofactors in analogy to what was performed by Shelnutt[32, 33] (Chapter 1.5.3) may have been hindered because there are relatively fewer crystallographic studies and, at first glance, less functional

variability amongst those that are available (*i.e.* fewer species and *unique*[*] proteins). Perhaps more importantly, the available structure determinations are resolved at relatively poorer resolutions owing to the experimental difficulties associated with the high molecular weight membrane proteins that are of great interest in photosynthesis.

In order to circumvent the problems posed by the limited resolutions of the available structures, the uniformity of the RC structural data was exploited explicitly to assess the experimental reliability of the measurements of BChl conformations in RCs and consequently led to their improved experimental determination. If the data could be shown to be consistent with real conformational differences amongst the cofactors (as was the case) these conformations could be correlated with their individual biophysical roles in the RC. This was achieved by applying the known-effects of macrocycle distortion to the relative differences in BChl conformations (*e.g.*, if X is more nonplanar than Y it is assumed that X is easier to oxidise, has red-shifted absorption spectra, *etc.*), a process that yields an initial estimate of the induced conformational effect isolated from other considerations of the binding-site. Additionally, considering the architecture of RCs, there is in fact significant functional variability amongst the individual BChls, so that the implications of the predicted effects may be explored in the context of the current knowledge of the biophysical dynamics of RC function. Thus, the predicted conformational influence could be correlated with suspected physicochemical differences between the RC BChls and this was extended to species specific variations.

2.1.2 Statistical Approach

The primary objective of the initial analysis of the normal-coordinate structural data (NSD) was to obtain best-estimates of the cofactor conformations within the protein in its crystalline state. This entailed the secondary aim of identifying and characterising any systematic errors present in the set of all relevant structure determinations. It will be useful to first consider the possible sources of error in the coordinates of these structures found in the Protein Data Bank (PDB), so as to describe how these may be mitigated by direct analysis of NSD results.

---

[*] For example, although there are dozens of structure determinations of the RC in the PDB the variety is in general limited to proteins with one or more site-specific mutations and the majority are from either one of only two species (*R. sphaeroides* or *B. viridis*). This is in contrast to haemoproteins where the rich diversity of different proteins with unique functions and properties gave purpose to a broad analysis of the distribution of haem conformations in general that could then be associated with functional significance.

The experimentally determined property that results from an X-ray diffraction study is an electron density map (EDM) of the unit cell, from which an atomic model is constructed to a first approximation of the density. This model is then refined to improve the agreement between the experimental and theoretical density distributions whilst adhering to sensible, chemically deduced restraints of the free parameters. Although error in the EDM itself is subject to separate contributions that arise from the precision of the measured diffraction amplitudes (data collection) as well as the estimates of the diffraction phases (phasing) and other aspects (*e.g.* scattering factors employed), it is useful here to consider these individual sources in combination to give rise to a total error in the EDM. This error will be propagated into the atomic model built from the data. However, the process of model building and refinement introduces further errors. These originate from factors such as the definitions and target values of restrained stereochemical parameters as well as their individual relative weightings to one another and the experimental observations. As one would expect then, the end result overall is that the reported atomic coordinates used in this study to assess the macrocycle conformations are not exact representations of the actual atomic configurations present in the crystal.

Fortunately, it is a simple matter in general to decrease the influence of sources of *random* error by aggregating the results from multiple measurements. However, blind averaging of all the available measurements is not necessarily appropriate since some of the sources of error described above may turn out to be *systematic* when considered across the entire dataset. For example, individual research groups may have specific protocols for refinement that they employ in all of their structure determinations that could be quite different from those used by others. As a consequence, before any attempts to obtain averages were made, the dataset as a whole was examined to determine whether any such systematic variations were present and if so, steps were taken to remove their influence.

2.1.3 Theoretical Model to Quantify Conformational Effects

In general, it has proved difficult to model the effects of nonplanarity on the physicochemical properties of porphyrins using quantum molecular modelling (QMM). Indeed, wholly different results were obtained using similar methods in the past that were dependent upon the structural models employed (see 1.5.4). Additionally, the common practice of using crystallographic coordinates obtained from protein structures in QMM studies introduces yet more difficulties regarding the reliability of the results.

However, the original development of a way to account for bias and reduce random experimental error in cofactor geometries from X-ray structural data described later in this

work, enabled the first QMM studies to be performed where this latter factor was explicitly accounted for (without geometry optimisations). Furthermore, the availability of a new experimental model (*i.e.*, the H-NOX mutants[203] described in 1.5.5) of the influence of conformational control upon haem reduction potentials, including high-resolution atomic coordinates, meant that the QM model could be calibrated and tested against these data first before applying it to a complex system.

2.1.4 Account for Non-additivity of Conformational Effects

Although the non-additivity of the effects of local and macrocyclic conformational parameters on key-physical properties has been suggested in the past,[111] no experiment has yet been able to clearly delineate their individual contributions. Whilst there is certain feasibility in using highly truncated models of haems in QMM studies of the redox influence of relative conformational differences amongst these compounds, because a key-interest in chlorophylls is their absorption properties that are necessarily very sensitive to the effects of conjugated peripheral substituents, these are not as relevant to BChls. This is in contrast to haems where truncated models are useful because the characteristic redox chemistry takes place at the haem-Fe or else involves the coupling of the aromatic pathway of the macrocycle to the metal and both of these factors are dominantly influenced by the coordination environment provided by the skeletal framework.

If it were possible to separate out the individual contributions of local and macrocyclic conformational effects (*e.g.*, the acetyl orientation from skeletal deformation), it may help to overcome another significant hurdle for the applicability of the conformational control concept to photosynthetic systems. In detail, a few studies have claimed to have categorically disproved that skeletal distortion effects key-properties of the special-pair BChls in RCs and instead attribute a dominant influence to the co-planarity of the C3-acetyl substituent. Although the model presented in this thesis for separating conformational effects is not applied directly to the special-pair, the results suggest that macrocycle deformation is usually the main conformational perturbation to BChl properties.

2.2 Aims and Objectives

In summary, there are two major objectives of the research accounted in this thesis:

1. to assess the role of conformational control in photosynthetic systems
2. to develop and apply methods to quantify and delineate the conformational contribution to physicochemical properties

Additionally, the individual aims that were required to fulfil these objectives were:

a. obtain reliable descriptions of the conformations of BChls and haems in photosynthetic proteins from the presently available structural data
b. predict the expected relative physicochemical effects, in terms of the state-of-the-art knowledge of the effects of porphyrin nonplanarity, using the relative differences in cofactor conformations or alternatively,
c. to quantify physicochemical differences arising from conformational effects using appropriate QMM techniques
d. to construct structural models and an experimental design for use in a QMM study that is capable of isolating the individual perturbations to BChl physicochemical properties arising from local and macrocyclic conformational effects

In terms of the individual projects that were conducted, Chapter 5 achieved the first two of the above aims by delivering what is argued here to be the most accurate and precise descriptions of the conformations of the individual BChls in the RCs from *R. sphaeroides* and *B. viridis* possible from the currently available experimental data (*i.e.*, aim 'a'). Additionally, the expected physicochemical outcomes of the relative conformational differences were correlated with their biophysical roles in the RC (aim 'b') and so demonstrated that the protein scaffold exerts functional conformational control on the porphyrin macrocycle in RCs (objective '1').

The following project (Chapter 6) applied the same statistical process to the structural data available for the RC-cyt from *B. viridis* and thus provided accurate determinations of the conformations of the individual haems ('a'), which were then used in subsequent theoretical calculations to directly assess the conformational impact on their midpoint potentials ('c'). The specific method employed was developed by correlating the results from a series of calculations on a recent model complex for conformational control of haem E_ms that thus provided an experimental calibration (objective '2').

The final project (Chapter 7) has not yet provided any improved model coordinates for the individual BChls from the FMO protein as the relatively high-resolution structural data available for these complexes appear to vary systematically and thus precludes reduction of error through averaging. However, this complex has been used to develop a method to delineate individual structural contributions to the intrinsic site-energy (*i.e.*, *in vacuo* excitation energy) that is proving to be quite successful ('d').

In all of the above cases, the results have extended the concept of conformational control to photosynthetic apparatuses. However, the results also serve as a model for the role of conformational regulation of tetrapyrrole cofactor biological function in other systems. Additionally (and more generally), the strategies that were employed resulted in the development of methods for the statistical analysis large volumes of protein-bound ligand crystallographic data, allowing *post hoc* assessment of the structural data's reliability. Finally, the research described in chapter 7 represents a relatively unique combination of a QMM study with a statistically designed experimental sequence, which serves as an exposition of the power and utility of this approach.

3 Methods

3.1 Data Collection

The atomic coordinates of reaction centre and light-harvesting complex BChls and haems that constitute the structural data used in this work were obtained from the Protein Data Bank Ligand Expo[279, 280] in '.ipdb' format (specific access dates are provided in the "Experimental Details" sections of each chapter). These were then processed with the normal-coordinate structural decomposition procedure (NSD)[32, 184] using the online Java application provided by Sun and Shelnutt,[281] with each structure oriented consistently (Figure 3-1), as described in the program's online documentation. The process was semi-automated using a macro written with AutoHotKey and individual results were saved in a template created using Microsoft Excel that provided immediate graphical illustration of the NSD as well as skeletal deviation plots for all individual structures.

The identity of the structure (*i.e.* PDB ID, residue ID, filename, *etc.*) was automatically associated with the results during processing but the biological identity (*e.g.* B_A, B_B, *etc.*) of the cofactor had to be added manually using either residue ID conventions for the HETATM records or associated axial ligands and was confirmed visually with reference to the full protein model, a lengthy process for the ~100 structures analysed.

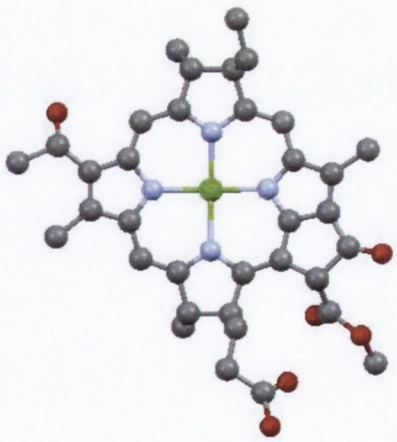


Figure 3-1: Correct orientation of the cofactors for NSD analysis (phytyl chain has been truncated). Rings A and B (indicated by the acetyl and ethyl groups, respectively) are used as the anchors to orient the structure such that reading clockwise from nine o'clock the rings follow the sequence A, B, C and D.

Once all individual structures were processed, the formatted row vectors containing NSDs and PDB annotation data (observations) were collected into '.csv' format using an MS Excel Visual Basic for Applications macro. At this stage the data files were read into R[282] where any further processing necessary (*e.g.* data reconfiguration) was accomplished using

various in-built index and sub-setting functions. Additionally, annotations relevant to all cofactors from a particular PDB structure (*e.g.* species) were added at this stage.

3.2 Structural Characterisation

The structural characterisation of nonplanar porphyrins may be achieved using a number of the macrocycle's geometric parameters that generally may be considered to fall into one of two groups. Specifically, there are those which utilise the internal angles and interatomic vectors 'as is' such as the C_a-N_{opp}-N_{opp}-C_a' dihedral angles to quantify *ruf* distortion, or alternatively, parameters derived relative to the 4N- or 24-macrocycle atom mean-plane. Although both of these approaches are valid, it is generally considered that deviations from the mean-plane best capture the nonplanar conformations of porphyrins and are most suitable for comparative analyses.[157]

Two useful quantifiers of the degree of nonplanar distortion are Δ24 and D_{oop}, which are defined as the root mean square deviation of the 24 macrocycle atoms from their least-squares plane and the related root sum-of-squares, respectively (Equation 3-1; D_{oop} is preferred in this work). Alternatively, skeletal deviation plots reveal local features of the individual z-displacements from the 24-atom mean-plane (Δz_i) by means of plotting Δz_i against an arbitrary positional parameter, assigned such that the view of each pyrrole unit is along its N-mid(C_b-C_b) bisector and reading from left to right one 'walks around' the macrocycle (Figure 3-2).[157]

Equation 3-1: The typical out-of-pane distortion of a single atom is given by Δ24, whereas a measure of the overall distortion is given by D_{oop}; in both equations Δz_i is the vertical displacement of an atom from the mean-plane.

$$\Delta 24 = \frac{1}{24}\sqrt{\sum_{i=1}^{24} \Delta z_i^2} \qquad D_{oop} = \sqrt{\sum_{i=1}^{24} \Delta z_i^2}$$

However, the advent of the method of normal-coordinate structural decomposition[32, 184] (NSD) represented a turning-point in the structural characterisation of porphyrins and it is this method that will be used extensively throughout this work. NSD is a conceptually simple procedure that employs the decomposition of the conformation of the macrocycle by a basis set composed of its various normal-modes of vibration. This technique affords clear separation of the contributing characteristic distortions to the macrocycle conformation in a quantitative yet readily interpretable fashion (Figure 3-2; *e.g.*, *ruf*, *sad*, *dom*, *etc.*). Furthermore, often a basis of only the lowest energy normal-mode of each symmetry are necessary to describe the conformation accurately, affecting reduced

dimension compared to specifying the 24 Δz_i displacements and therefore allowing the large scale analysis of porphyrinoid crystallographic data.

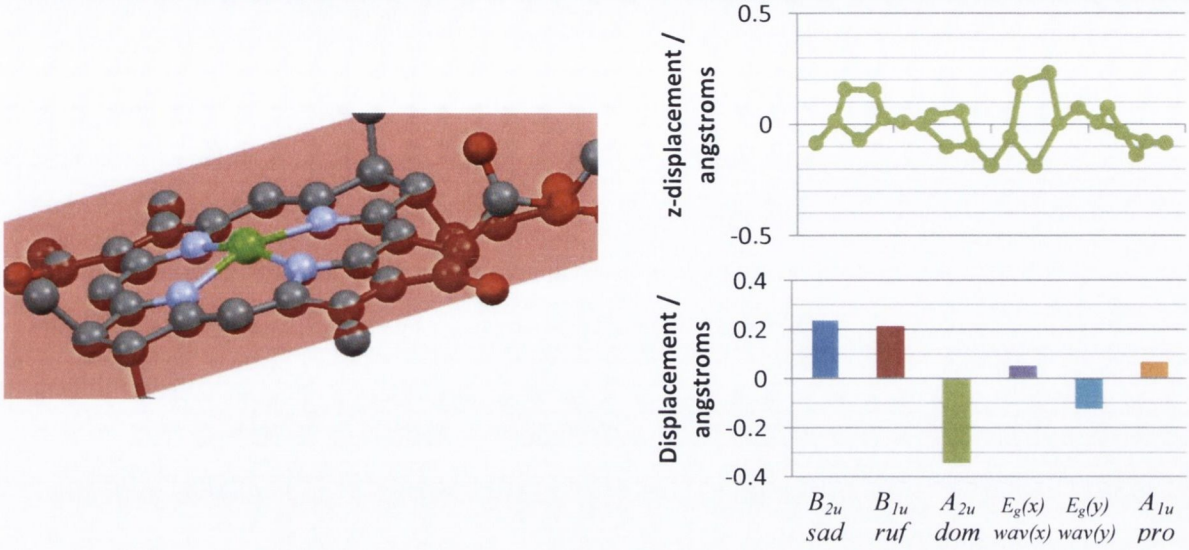


Figure 3-2: Illustrations of methods used to characterise macrocycle nonplanarity. The mean plane that is defined by the 24-macrocycle atoms (left), from which the individual skeletal deviations are derived (top right) that may be subsequently decomposed by NSD (bottom left).

3.2.1 Normal-Coordinate Structural Decomposition

The observation that the commonly observed *ruf*, *sad*, *dom* and *wav* conformers were very similar to static displacements along the lowest-energy normal-modes of vibration of the macrocycle led to the development of a new technique for the quantitative structural analysis of tetrapyrroles called normal-coordinate structural decomposition (NSD). The NSD procedure describes an observed macrocycle conformation in terms of a linear combination of displacements from a planar reference geometry along so-called normal-coordinates derived from the normal-modes of a planar porphyrin macrocycle. In Shelnutt and coworkers[32] formulation there are three major components of NSD that require attention: the calculation of the normal-modes, derivation of the normal-coordinates from the normal-modes and the selection of the 'reference macrocycle' whose normal-modes are used to construct the normal-coordinate basis.

In principle, any molecule's conformation may be described as a linear combination of deviations from a reference geometry (*e.g.*, the *in vacuo* ground state) along its 3N – 6, or 3N – 5 as appropriate, normal-modes of vibration. For porphyrins, this means that 66 normal-modes span the vector space that is required to describe any possible conformation of the 24 atom (= N) macrocycle skeleton. Furthermore, for a D_{4h} symmetric porphyrin these modes split into 2N – 6 = 45 in-plane (IP) and N – 6 = 21 out-of-plane (OOP) modes whose symmetries are given by the irreducible representations of the D_{4h} point group (*i.e.*, IP: B_{2g}, B_{1g}, E_u, A_{1g}, A_{2g}; and OOP: B_{2u}, B_{1u}, A_{2u}, E_g, A_{1u}).[32] Jentzen *et al.* calculated these

normal-modes quantitatively using a molecular mechanics force field based on DRIEDING that had been shown to accurately predict the conformations of many distorted porphyrins in the solid-state.[184] Since a model unsubstituted metalloporphyrin was used for the vibrational analysis, extraneous normal-modes involving C-H stretches and large-motions of the central metal were also obtained. Whilst the former were automatically isolated from the skeletal motions on account of their naturally high frequencies, the metal's motion was artificially decoupled by setting its mass close to zero ensuring that its modes were confined to exceptionally high frequencies. The 66 remaining normal-modes comprise an orthonormal set that span all possible conformations of the macrocycle skeleton. However, upon their conversion into ordinary coordinate space *via* division by the root-mass-matrix and re-normalisation to obtain the normal-coordinates that are used as the basis for structural decomposition they are no longer orthogonal within a given symmetry type due to the different masses of the C and N atoms.*

As for the reference macrocycle used to calculate both the normal-modes and the Cartesian displacement coordinates of porphyrins that are analysed (see later), unsubstituted Cu(II)porphyrin was chosen as its ground-state was considered to represent a least-strained geometry (*i.e.*, planar with d(M-N) close to Hoard's optimum). Thus the normal-coordinate basis employed by NSD originates from the 66 skeletal normal-modes calculated for Cu(II)P using a modified version of the DRIEDING force field with the mass of Cu set to 0.001 amu. Additionally, the Cartesian displacements of the structure under analysis, that are directly decomposed into a linear combination of the normal-coordinate basis, are determined with reference to the symmetry adapted coordinates taken from the Cu(II)P crystal structure.

The computation of a particular structure's NSD is quite straightforward and begins with the calculation of the macrocycle least-squares plane and the individual atomic deviations from this plane (Δz_i).[32] The Δz_i are sufficient for the OOP decomposition since this is achieved *via* solution of the matrix equation, $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is the normal-coordinate basis set, $\mathbf{x}$ is the vector of normal-coordinate coefficients and $\mathbf{b}$ is the vector of Δz_i.† The coordinate system is then transformed by a series of orthogonal axis rotations. The first

---

* Division by the mass-matrix is equivalent to dividing the normal-mode amplitude for each atom (*i.e.*, the corresponding entry in the normal-mode eigenvector) by its atomic mass. Since the macrocycle possesses both C and N atoms the eigenvector entries are divided by different numbers thus breaking the orthogonality of the set (*i.e.*, the orthogonality is lost in the conversion of mass-weighted to non-mass-weighted coordinates because the mass-matrix is not a scalar multiple of the unit matrix).

† Note later that if $\mathbf{A}$ is a reduced basis set (*i.e.* min. or ext.) then a least-squares solution will be required.

rotation sets the x,y-plane of the new coordinate system so that it coincides with the macrocycle mean-plane and the second rotates the z-axis so as to minimise the sum of squares of the coordinate deviations between the observed structure and the reference macrocycle.* Once these rotations are complete, the remaining x,y-displacements between the observed and reference structures give the IP deformations in an analogous manner to the linear decomposition described above. In summary, the computational steps are:

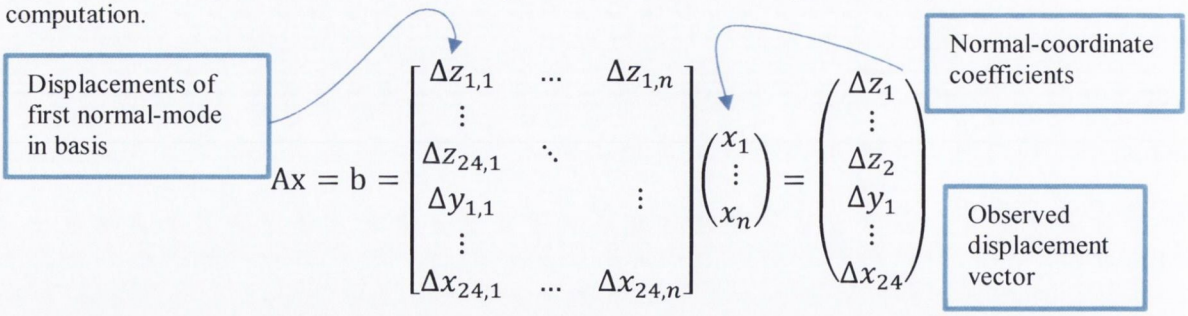
1. Determination of the macrocycle least-squares plane - the individual atomic deviations from which (Δz_i) are used to calculate the OOP normal-coordinate coefficients in the first linear decomposition.
2. Rotation of the axes so that the x,y-coordinate plane coincides with the macrocycle least-squares plane.
3. Rotation of the observed structure around the z-axis to minimise the sum-of-squares deviations between the x,y-coordinates of the macrocycle atoms under analysis from those of the reference structure.
4. The x,y atomic deviations in this orientation between the observed macrocycle and the reference structure (Δx_i, Δy_i) are used to calculate the in-plane deformations in the second linear decomposition.

Note that the above description of the linear decomposition step has been slightly simplified. The actual computation uses a 72 × n basis matrix, **A**, (where n = the number of normal-coordinates in the basis; *i.e.* n = 12, 24 or 66 for the *min.*, *ext.* or *comp.* bases respectively) where each of its columns corresponds to a normal-coordinate (*e.g.*, the column that corresponds to the lowest-energy B_{2u} normal-deformation vector contains non-zero entries, where appropriate, for the first 24 elements (Δz_i-displacements from planarity) and zero entries for the remaining 48 (Δx_i, Δy_i-displacements from reference macrocycle) and *vice versa* for any IP normal-mode. It should therefore be apparent that the observed distortion, **b**, is a 72 × 1 column vector containing the Δz_i mean-plane displacements and the Δx_i, Δy_i-displacements from reference macrocycle, that are obtained after the coordinate transformations described above. It follows then that the solution vector, **x**, is an n × 1 column vector containing the scalar coefficients for each of the normal-deformations of the basis, *i.e.* the vector of normal-coordinates (Equation 3-2).

---

* The calculation of the in-plane atomic deviations from this macrocycle can be visualised by analogy with the δ_{ip} calculation, replacing the simulated structure with the reference (Equation 3-3)

Equation 3-2: The generalised linear matrix equation that is solved during the last step of the NSD computation.



Reduced Normal-Coordinate Bases: Min. and Ext.

A crucial observation was that basis sets composed of less than the full complement of the 66 normal-modes of the macrocycle were often able to accurately reflect the macrocycle conformation and these were typically the lowest-energy modes of each symmetry type.[32] The explanation behind the adequacy of reduced bases was that displacements along normal-modes of lower energy are more accessible because they exhibit the least restoring forces, whilst the prevalence of these symmetric distortions* lent empirical weight to this idea.[32] Thus, reduced basis sets composed of only the lowest energy and both the lowest and next-to-lowest energy normal-coordinates of each symmetry type (min. and ext., respectively) are often adequate to characterise macrocycle conformations. This is particularly useful for characterising protein-bound macrocycles where in contrast to the symmetric distortions found in most synthetic compounds (*i.e.* single mode deformations), the asymmetric environment of the protein tends to induce mixed-mode conformations that are not easy to characterise visually, or using geometric parameters other than NSD. However, when using a reduced basis set the fit between the actual structure and that given by the min. or ext. basis simulations† will not be exact (Figure 3-3).[32] As a result, a measure of the accuracy of the simulated structure was required and so the root-mean-square positional error between the actual and simulated structures were defined for the OOP and IP normal-coordinates, termed δ_{oop} and δ_{ip}, respectively (Equation 3-3).

The normal-modes for the OOP distortions of the min. basis consist of the lowest energy vibration from each symmetry (Figure 3-4) and are familiar to porphyrin chemists as the

---

* The very existence of a systematic nonplanar stereochemistry of porphyrins (see Chapter 1.5.2) defined in terms of the symmetric *ruf*, *sad*, *dom*, *etc.* conformations correctly implies their ubiquitous observation in porphyrin structural studies; thus, in spite of the chemical diversity of porphyrin compounds, these conformations are inherently favoured by the macrocycle. The energetic cause of this observation became clear once it was realised that these canonical stereochemistries closely resembled the lowest-energy normal-modes of the skeletal framework.

† Since the normal-coordinates and the conformation are related in a deterministic way, it is trivial to obtain an ordinary Cartesian description of the macrocycle from the NSD normal-coordinates by projecting them back onto the reference macrocycle. The "simulated structure" is that obtained by reversing the NSD process but using only the normal-coordinate coefficients of the reduced basis.

*sad*dle (B_{2u}), *ruf*fle (B_{1u}), *dom*ed (A_{2u}), *pro*pellered (A_{1u}) and the degenerate *wav*e modes (E_g(x) and E_g(y)).[32] The IP modes that compose the min. basis (Figure 3-5) are the *meso-str*etching (B_{2g}), *N-str*etching (B_{1g}), pyr-*tra*nslation (E_u(x) and E_u(y)), *bre*athing (A_{1g}) and pyr-*rot*ation (A_{2g}).

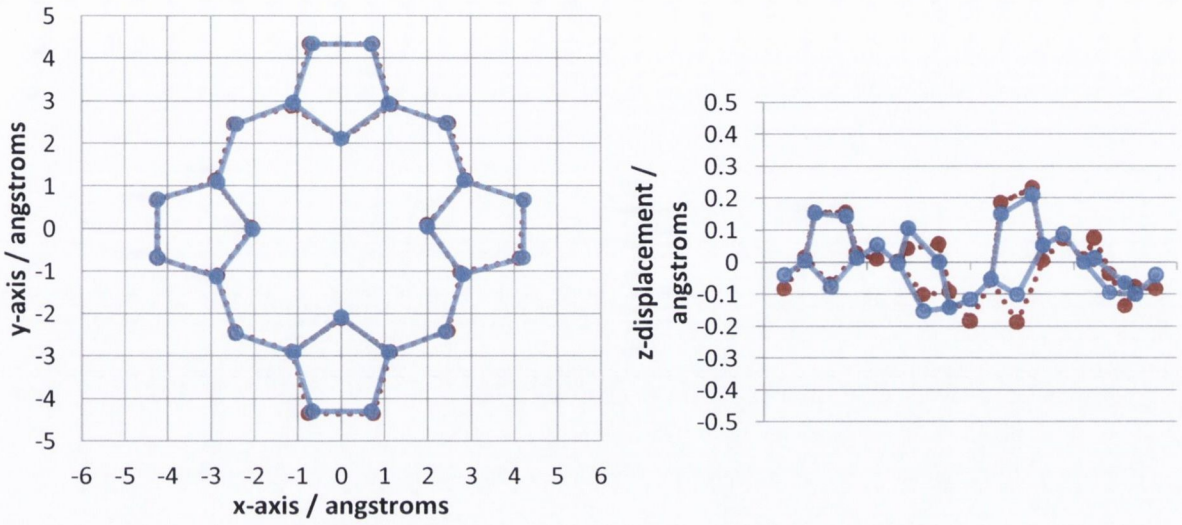


Figure 3-3: Comparison of the in- and out-of-plane (left and right, respectively) atomic displacements yield by the min. basis simulations (blue) and the observed structure (red) from the NSD analysis of BCL C302 from PDB ID: 1E6D;[283] δ_{ip} = 0.0533Å and δ_{oop} = 0.0368Å.

Equation 3-3: The goodness-of-fit between reduced basis simulations and the actual conformation.

$$\delta_{\text{oop}} = {}^{1}/_{24}\sum_{i=1}^{24}\sqrt{[(\Delta z_i)_{sim} - (\Delta z_i)_{obs}]^2}$$

$$\delta_{\text{ip}} = {}^{1}/_{24}\sum_{i=1}^{24}\sqrt{[(\Delta x_i)_{sim} - (\Delta x_i)_{obs}]^2 + [(\Delta y_i)_{sim} - (\Delta y_i)_{obs}]^2}$$

Figure 3-4: Molecular illustrations of 1Å deformations along the normal-coordinates of the minimum basis. B_{2u} (top left), B_{1u} (top right), A_{2u} (middle left), E_g(x) (middle right), E_g(y) (bottom left) and A_{1u} (bottom right).

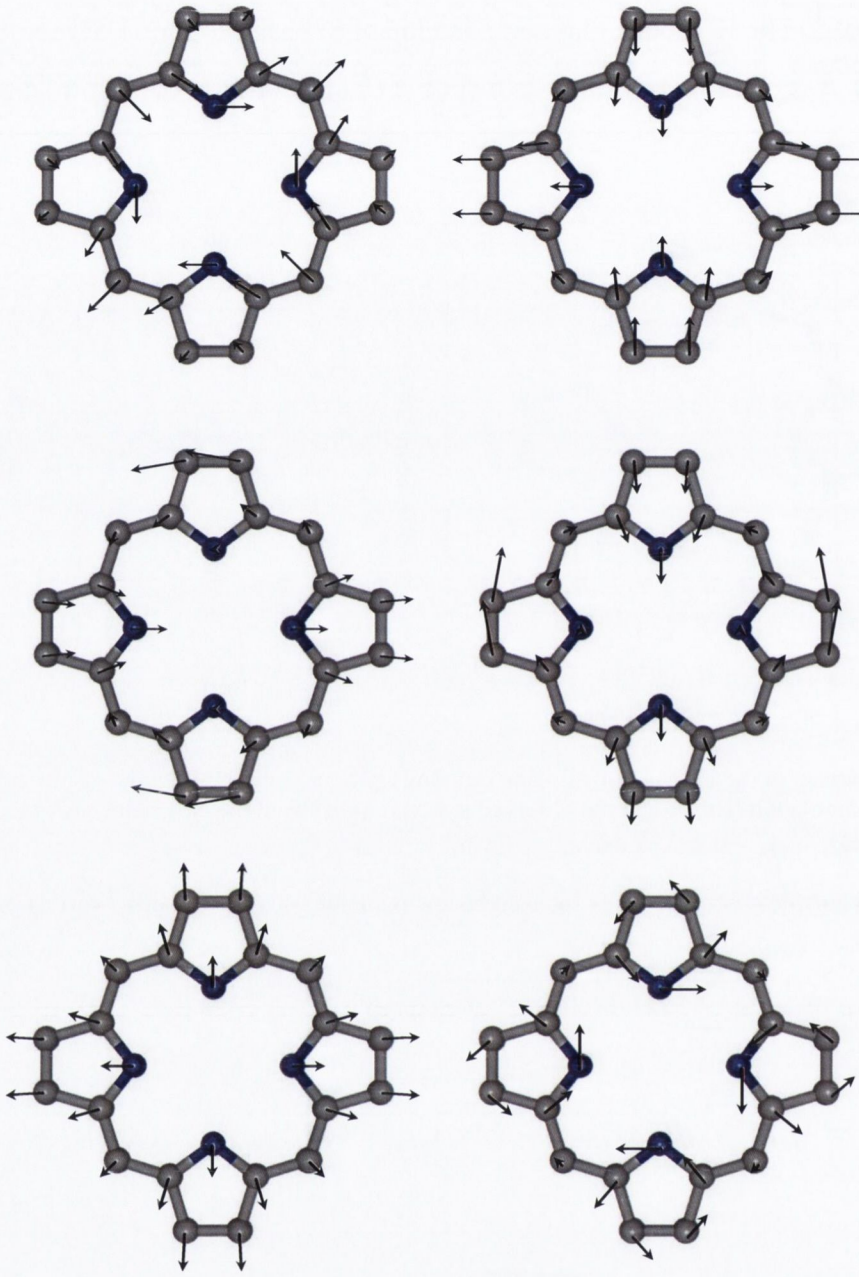


Figure 3-5: Molecular illustrations of 1Å deformations along the normal coordinates of the minimum basis. B_{2g} (top left), B_{1g} (top right), E_u(x) (middle left), E_u(y) (middle right), A_{1g} (bottom left) and A_{2g} (bottom right). Note that B_{2g} and B_{1g} modes likely alter internal H-bonding in the core and, with the addition of A_{1g}, the coordination environment of central metals. These figures were produced by generating the atomic displacement vectors in MS Excel from the normal-deformation matrices given in Jentzen *et al.*[32] and superimposing the image onto an illustration of the reference macrocycle (produced using GaussView[284] and the coordinates from Jentzen *et al.*).

3.3 Data Analysis

Although specific details regarding the conduct of any of the substantial data analyses performed are provided in later chapters when relevant, a general discussion of some of the main techniques employed is given here. Additionally, it was considered useful to include short summaries of techniques that were used extensively throughout the analyses and these may be found below in Table 3-1 for easy reference. Usually, analyses were performed with R,[282] although in some cases other software was used and this is indicated where relevant.

Table 3-1: Summaries of the main statistical techniques used throughout the analysis.

Dissimilarity analysis	Measures the multivariate 'dissimilarity' between two observations. In this work, the Euclidean distance of each pair of observations was used yielding a symmetric matrix with a leading diagonal of zeros (the dissimilarity of an object to itself). In chemical terms, it quantifies the conformational dissimilarity between two structures / observations.
Agglomerative Hierarchical Clustering	AHC groups similar observations in a hierarchical manner *via* a stepwise process. Thus, the two most similar observations are paired to form a group (cluster), then from those remaining, either the next two most similar are paired or one observation is added to the previous group and so on, until there is only one group containing all the observations. A decision must be made as to how many clusters to select. This can be achieved either by setting the number of clusters or by consideration of the *dendrogram* which shows the relationship between them, including a measure of the (dis)similarity of the clusters or a quantity proportional to an error measure at each stage. In this work the Euclidean distance was used to obtain the dissimilarity matrix and Wards' method was used for the agglomeration. In chemical terms, the process distinguishes groups of cofactors with similar conformations.
Kernel Density Estimation	KDE provides a nonparametric estimation of the distribution of a dataset by placing a symmetric function at each observation and summing the result. In this case, Gaussian kernels were used with bandwidths (*h*; *i.e.*, standard deviations) determined using the plug-in selector.

3.3.1 Multivariate Clustering

Cluster analysis is a statistical technique that may be used to reveal the relationships between observations in a multivariate dataset (*e.g.* the NSD coefficients for protein-bound cofactors explored in this thesis).[285, 286] In general, this is achieved by an algorithm that in some way iteratively groups observations together so that a particular objective function is optimised resulting in groups of observations (clusters) that have similar characteristics within the groups and distinctions between. To this end, there are two major methods that

differ fundamentally in their approach, namely k-means and hierarchical clustering, where the latter is further characterised as being either agglomerative or divisive.[285] In this work, agglomerative hierarchical clustering (AHC) of NSD data was employed to find groups of cofactors that possessed similar conformations to test the idea that the individual binding-sites imparted distinct conformations on the bound cofactor.

AHC approaches the problem of clustering by combining objects on a case-by-case basis into groups, whereupon new groups themselves are considered by the method as candidates to be merged, based on some property of the whole group, and the process continues until a single group containing all objects is formed. As an illustrative example, consider the AHC analysis in Figure 3-6, which used the Euclidean distance matrix constructed from an arbitrary 2D-dataset and Ward's method for the agglomeration. The dendrogram shows the evolution of the cluster solution. Notice that first points 6 and 7 were merged and then 8 was incorporated into this cluster, new clusters were then formed to contain points (1,2) and (3,5), the latter being quickly merged with 4. At this point the outlying observation 9 is merged with cluster (1,2) followed by the union of clusters (3,4,5) and (6,7,8).

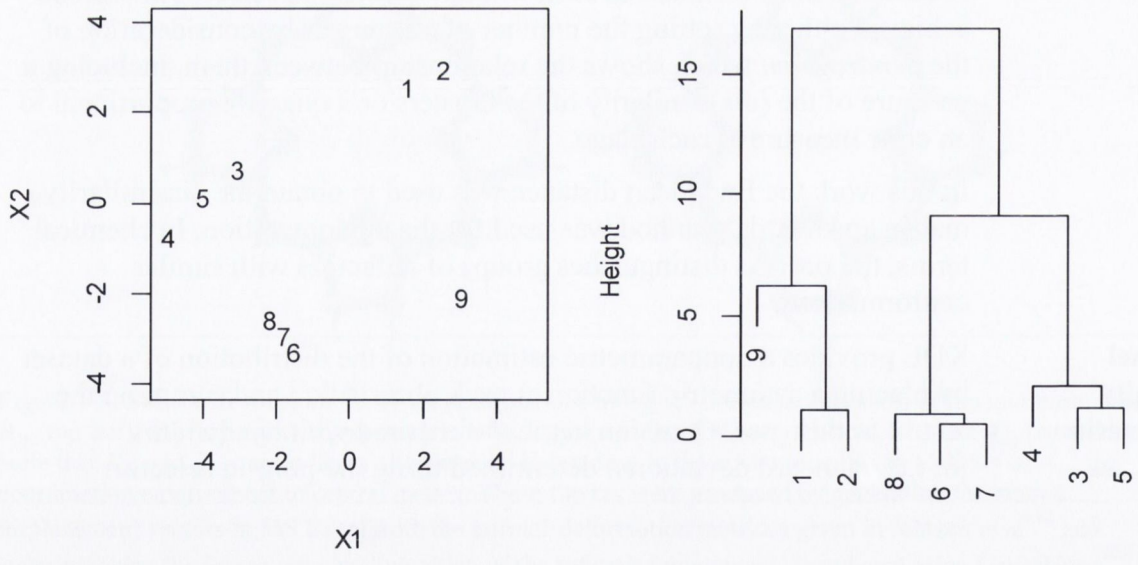


Figure 3-6: Example of an agglomerative hierarchical cluster analysis, data points (left) and dendrogram (right). The dendrogram indicates the order in which each point is merged to form a cluster and suggests the presence of four clusters that is in line with the distribution of points illustrated in the scatter-plot.

In general the height of the node representing a particular union is representative of the dissimilarity of the observations, or clusters of observations, that were merged (*i.e.*, the 'higher' the node, the less similarity exists between the observations). This provides an intuitive way to select which cluster solution is most-appropriate for a given problem, in addition to the availability of any *a priori* knowledge of what is likely to represent a

meaningful cluster. Additionally, the vertical axis of the dendrogram (Figure 3-6) may represent a specific property of the clusters, depending on the particular method employed (*vide infra*).

To initiate an AHC, an appropriate measure of the proximity of the observations is required (*i.e.*, the conformational similarity). With respect to continuous numerical data, the general form of proximity measures are dissimilarity matrices constructed from the pairwise distances defined by a particular metric.[285] This gives rise to an $n \times n$ matrix where each entry d_{ij} corresponds to the distance between observation i and j. Metrics that are in common use for clustering applications include the Manhattan, Euclidean (the l_1 and l_2 norms, respectively) and the squared Euclidean distances, amongst many others.[285] The decision as to which distance measure is most appropriate is strongly driven by the nature of the data. In the extreme, clearly categorical, ordinal, discrete and continuous data all give rise to fundamentally different notions of distance, so much so in fact that for categorical observations the adjectives similarity or uniqueness are intuitively more appropriate than 'distance' to rely the degree of proximity or isolation. However, even with continuous numerical variables the use of metrics other than the familiar Euclidean distance (see below) may be warrantied. For example, when variables are highly correlated and one wishes the distance to reflect anomalous observations within this structure the Mahalanobis distance is called for. In this work, since NSD results in a continuous numerical description of the cofactor conformations, and we are not interested in accounting for correlation in the definition of similarity, the Euclidean distance metric (Equation 3-4) was used to construct dissimilarity matrices.

Equation 3-4: The Euclidean distance between two objects i and j in n-dimensional space. The matrix is symmetric around the leading diagonal of zeros (*i.e.* the distance of an object to itself) and each off-diagonal term is the distance (or dissimilarity) of observations i and j; since the distances are all positive, naturally only the unique pairwise entries are required by the cluster algorithm.

$$d_{i,j} = \sqrt{\sum_{k=1}^{n} \left(x_{ik} - x_{kj}\right)^2}$$

Additionally it must be decided whether or not to standardise the data (*i.e.* convert to zero mean and unit variance) prior to calculation of the distance matrix. This would be essential if the observed variables were measured in different units;[285] however, as this is not the case here (all normal-coordinates are in Å) this condition clearly does not apply. The question remains relevant though and in order to address it must be recognised that scaling all observations to unit variance would have the effect of reducing the influence of normal-

deformations that varied substantially and increasing that of those whose variations were generally smaller. For example, in a clustering of all 66 normal-coordinates after scaling, one would observe the importance of the more variable low-energy distortions* (*ruf*, *sad*, *dom*, *etc.*) to be reduced because their contribution to the pairwise distances would decrease, whilst the importance of the less variable high-energy modes (*e.g.* $B_{1u}(2)$) would be inappropriately inflated. This would have a negative impact upon how well the clusters represented relationships between the actual macrocycle conformations, since all of the NSD modes are measured on the same scale (*i.e.* normalised). Thus, the NSD results were used 'as is' in all cluster analyses.

The second major aspect of an AHC is the agglomeration method, which determines how groups of observations that have already been clustered are treated in subsequent stages of the agglomeration. Simple criteria are, for example, single-linkage clustering which is also known as the nearest neighbour method that identifies the proximity of a cluster K to another L as the minimum of all pairwise distances between the objects in K to those in L, which of course is valid for 'clusters' of single objects.[285] The method used in this work, namely Ward's method, is more complex (Equation 3-5) but in essence minimises the increase of the sum total, *i.e.* over all clusters, of the within cluster sum-of-squares error.

Equation 3-5: Ward's total within cluster error sum-of-squares criterion. Where the summations are over the total number of m clusters (g), the total number of l objects in the m^{th} cluster (n_m) and the total number of k variables (p_k).[285]

$$E = \sum_{m=1}^{g} E_m; \text{where } E_m = \sum_{l=1}^{n_m} \sum_{k=1}^{p_k} \left(x_{ml,k} - \bar{x}_{m,k}\right)^2$$

In general, Ward's method has been shown to perform very well when the clusters contain a similar number of objects although is not necessarily optimal when this is not the case.[285] However, throughout this work, exploratory analyses generally involved significant trials of various clustering methods that tended to yield similar results. Thus, the results from Ward's method have been presented in all AHCs for the sake of standardisation and because the nodal structure on the dendrogram may be interpreted as related to the relative

* The lower-energy distortions exhibit greater variation (*i.e.* range of values for different cofactors) because they possess shallower potentials leading to the possibility of high, intermediate or small deformations being induced by the protein. In contrast, only slight deformations along the high-energy modes may be induced due to the steepness of their potentials so the values found for a particular chemical cofactor (*e.g.* BChl a) do not vary much as a result of different protein environments.

loss of within cluster homogeneity, or alternatively, the increase in the total within cluster sum-of-squares error.

3.3.2 Principal Components and Factor Analysis

Although not a direct alternative to cluster analysis, principal components analysis (PCA) and the related exploratory factor analysis (EFA) are well-suited as an aid to discover groups of similar observations in high-dimensional data through dimensionality reduction, which allows identification of objects that show similarity in a lower dimensional scatterplot.[285, 286] An additional benefit of using this technique as a complement to cluster analysis is that the new axes are linear combinations of the original variables (here the macrocycle normal-deformations) so that the validity of the cluster centroids (*i.e.*, within cluster NSD averages) as representatives of real *trends* in the data, may be directly assessed. Both PCA and FA are methods that effectively perform variable axis rotations to generate new 'variables' (termed principal components or factors, respectively) from linear combinations of the original set with the goal of accounting for as much variance as possible in as few PCs or factors as possible.[286] The main differences derive from the relationship of the new variables to the original set in each method, which in PCA is that the PCs are explicit linear combinations of the original variables whereas in EFA the observed variables (*i.e.*, NSD coefficients) are linear combinations of unobserved, or latent variables (*i.e.*, factors influencing the NSD coefficients). Additionally, PCs are uniquely defined by a dataset and there will usually be as many PCs as there are variables albeit that the first few PCs will account for the majority of variability in the data and are hence the most important. In contrast, in EFA the number of factors to be extracted is determined before the analysis is commenced and once obtained there correlations with the original variables (loadings) are invariant under orthogonal rotations so that they may be rotated to aid easy interpretation. For example, the Varimax method employed in Chapter 5.4 rotates the factor loadings in order to maximise the variance of the loadings each original variable across all of the factors with the intention that each variable should only be highly loaded upon a single factor making it easier to interpret the factors.[286]

3.4 Quantum Molecular Modelling

The selection of an appropriate method for the theoretical modelling of molecular systems and properties of interest is a delicate decision requiring simultaneous consideration of the required accuracy of the theoretical formalism alongside the availability of computational resources. This latter consideration is particularly relevant to porphyrins and their biological derivatives since the simplest compound (*i.e.*, unsubstituted porphyrin) contains

24 nonhydrogen atoms, whilst full models of haems and chlorophylls have around 43 (290 electrons) and 66 (418 electrons) nonhydrogen atoms, respectively. Additionally, haem is a transition metal derivative and so the presence of d-electrons necessitates further consideration.

In the research presented in this thesis, electronic properties of haems were calculated using DFT (Chapter 6), specifically the B3LYP functional[287] with various basis-sets,[288-293] and of bacteriochlorophylls using the semi-empirical models PM6 for geometry optimisations and ZIndo/S to calculate electronic excitation energies (Chapter 7). In both cases the Gaussian 09 software package[294] was used and calculations were run on the Lonsdale cluster provided and maintained by the Trinity Centre for High-performance Computing.

3.4.1 Semi-empirical NDDO and INDO/s Methods for Modelling Chls: Precedent and Benchmarks

Calculation of spectroscopic properties of porphyrins using the popular INDO/s method has a well-established precedent in porphyrins and their metal complexes.[36, 111, 119, 120] As has been discussed in Chapter 1, INDO/s correctly reproduces the nonplanarity induced red-shifts of sterically encumbered nonplanar porphyrins providing a realistic geometry is provided (*i.e.*, either from a crystallographic study or suitable *in silico* optimisation).[36, 119] For example, calculations using crystallographic model geometries produced Q_y wavelengths of 782 nm for BChl a and 734 nm for BPheo a that compare well to the experimental values of 775 and ~746 nm, respectively, slightly better than for the corresponding Q_x wavelengths of 590 *vs.* 583 nm for BChl a (calc. *vs.* expt.) and the corresponding 565 *vs.* ~526 nm for BPheo a.[111] More extensive work using geometries obtained from MM optimisations (*e.g.*, a porphyrin specific DRIEDING II), has shown that INDO/s calculations may often systematically overestimate the Q_y wavelength, whilst the Soret band is underestimated,* but crucially relative differences amongst compounds are very well modelled.[36, 120]

Linnanto and Korppi-Tommola[55, 258, 295-299] have performed extensive comparisons of the accuracy of semi-empirical and *ab initio* methods for the calculation of geometries and excitation energies of chlorophylls and bacteriochlorophylls. These studies have ranged from isolated molecules and solvent complexes all the way to full models of reaction

* It is unclear as to whether this is due to inaccuracies in the MM optimisation or the INDO/s model as the author is not aware of a relevant comparative study using experimental and optimised geometries.

centre and light-harvesting complexes.[299] In general, exceptionally good performance of the PM5//ZIndo/S scheme (*i.e.*, PM5 geometry optimisation followed by a ZIndo/S single-point energy calculation) is found for the Mg complexes of unsubstituted porphyrin, chlorin and bacteriochlorin as well as 7 Chls and 8 BChls.[297] The PM5 optimised structures of unligated and 6-coordinate complexes predicts a planar porphyrin with the Mg in the 4N-plane, whilst 5-coordinate species show a substantial out-of-plane displacement of the Mg atom, in line with both B3LYP and HF/6-31g*. Additionally, all methods correctly predict that the Mg-N bond(s) to the reduced rings are longer than those from non-reduced rings. Partial atomic charges were also found to be consistent amongst the semi-empirical, HF and DFT methods.[297]

Linear least-squares fits of the experimental Q_y, Q_x and Soret absorptions against the PM5//ZIndo/s results provided correlation coefficients greater than 0.985 for the unsubstituted complexes.[297] Likewise, the experimental and predicted absorptions of the Chl and BChl compounds were also highly correlated, although for the BChls slopes ranging ~0.7 – 1.1 were reported whilst for the Chls the slopes were all ~0.7 in the 1:1 acetone complexes; the 4-coordinate models were also highly correlated with experiment although the slopes were systematically lower. These results were rationalised in terms of the greater sensitivity of the Q_x and Soret absorptions to solvent effects, which tend to stabilise these higher-energy transitions and hence the absence of this effect reduces the slope from unity.[297]

In the above study only two Chls and two BChls were optimised using the HF and DFT methods and only ZIndo/s excitation energies were calculated.[297] However, in a companion article B3LYP and HF optimisations were performed on an additional two BChl models and excitation energies calculated using CIS and time dependent HF and B3LYP, using the 6-31g* and 6-311g* basis sets, providing more scope for comparison.[298] The main conclusion was that the best transition energies were obtained using the B3LYP/6-31g*//ZIndo/S scheme. However, PM5//ZIndo/S was also considered to perform well and was recommended for use in large calculations where the more accurate and expensive B3LYP/6-31g* optimisations were impractical. Thus in Chapter 7, where the plan is to perform thousands of geometry optimisations and single-point energy calculations upon fairly complete BChl models, a semi-empirical methodology is indicated.

3.4.2 Density Functional Theory

It has long been known that some form of treatment of the correlation energy is necessary to provide adequate description of the first-row transition metal complexes due to the near

degeneracy of the 4s and 4p orbitals as well as amongst the 3d orbitals.[300-302] Moreover, larger errors typically result from uncorrelated calculations upon metals with incompletely filled d-shells to the right of the d-block (*i.e.*, Fe – Cu).[293] Furthermore, porphyrins also display significant electronic correlation,[303] and this is hardly surprising given the near degeneracies of the frontier orbitals* as well as the many electrons that are present in the delocalised π-system.

For example, uncorrelated HF optimisations of porphyrin were shown to yield structures with severe framework bond length alternation or "frozen resonance forms" of C_{2v} symmetry, whereas inclusion of electron correlation *via* MP2 or LDF optimisation provided the correct D_{2h} structure; notably, the severity of the bond-length alternation in the HF converged structure increased with larger basis sets.[303] Additionally, the authors' suggested that the near degeneracies amongst the frontier orbitals may contribute significantly to the increased correlation energy in D_{2h} geometry as excitations from the two HOMOs were shown to increase in the D_{2h} structure indicated by the T2 cluster amplitudes. The correlation amongst the core and σ electrons were also shown to be important as MP2 treatment of the π-electrons alone lost over half of the correlation stabilisation energy.[303] The comparisons were also relevant to chlorin and moreover, vibrational calculations of chlorin implied that it is more flexible with respect to out-of-plane distortions than porphyrin, with low-energy vibrational energies for the *sad* and *ruf* modes plus an even softer mode involving a localised torsion of the saturated C_b-C_b bond.[303] Since then, Density Functional Theory (DFT) has found significant application in the modelling of biologically relevant porphyrin and related compounds,[304-308] and the efficient way in which DFT explicitly incorporates electron correlation within its theoretical formalism is certainly a major contributor to this success.

---

* Recall Gouterman's four-orbital model where an accurate description of porphyrin electronic spectra is achieved by assuming that the two HOMOs and two LUMOs are, respectively, degenerate (Chapter 1.3.3).

4 Analysis of the Robustness of NSD with Respect to Coordinate Error

4.1 Error analysis of the NSD method

With respect to the consideration of error in the NSD analysis of porphyrin X-ray structure data Jentzen *et al.*[32] laid the foundations by considering the approximation error (δ_{oop} and δ_{ip}) of the use of the truncated basis sets (*i.e.*, min. and ext.). In addition, this work also provided a derivation for the positional error in reduced dimensions that are relevant to the separate out-of-plane (1D; OOP) and in-plane (2D; IP) normal-coordinate displacements. This led to the general rule of thumb that a reduced basis approximated the nonplanar conformation of the structure to within the experimental error if δ_{oop} is less than or equal to ~0.03 Å for structures at ~2 Å resolution.

However, no analysis of the numerical stability of NSD was included in the original derivation and neither has one been published since. Therefore, whilst it was not expected that the computational procedure would be especially sensitive to experimental error, it was felt prudent to check this explicitly. In particular, the larger basis sets are not completely orthogonal with respect to deformations of the same symmetry type (see Chapter 3.2.1) and this could potentially result in a magnification of the experimental coordinate error in the NSD solution. This was noted by the original authors and creators of NSD, who commented that the affected bases (*i.e.*, min. and ext.) were almost orthogonal.[32] Here though, it is shown formally that there is no serious numerical instability of the NSD procedure *via* analysis of the basis set condition numbers.

4.2 Conditioning of NSD Bases

It is a well-known mathematical problem that in linear systems involving non-orthogonal matrices with uncertainties in any of the components,* it is possible for the relative error affected by the original uncertainty to become magnified in the solution.[309] That is to say, small changes in the input due to coordinate error can become large changes in the output, which in this case is the NSD solution vector. This problem is referred to as the matrix 'condition' such that a 'well-conditioned matrix' does not affect a large magnification of the relative error. Therefore, in order to establish the true validity of the NSD method, the

---

* In this case, the positional parameters of the macrocycle atoms are uncertain with respect to the crystallographic coordinate error.

conditioning of the basis matrices, and hence the numerical stability of the decomposition, must be established.

The final stage of the NSD computation has already been described. To summarise, the linear decomposition is simply the solution of the standard matrix equation, $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ represents the normal-deformation basis set, $\mathbf{x}$ is the vector of normal-coordinate coefficients and $\mathbf{b}$ is the vector of Δz_i displacements of macrocycle atoms from their least-squares plane. However, since the value of $\mathbf{b}$ is known only to a certain degree of precision, the matrix equation is more accurately written as shown in Equation 4-1, where $\delta\mathbf{b}$ is the coordinate uncertainty in the Δz_i displacements and $\delta\mathbf{x}$ is the error propagated into the normal-deformation coefficients.

Equation 4-1: Standard matrix equation including error terms.

$$\mathbf{A}(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b}$$

The magnitude of the error in $\mathbf{x}$, *i.e.* the norm of $\delta\mathbf{x}$, can be shown to be bounded above by the product of the norms of the inverse and the input error as in Equation 4-2, where the double brackets indicate the norm of the matrix or vector.

Equation 4-2: Upper-bound for the norm of the error in the solution vector.

$$\|\delta\mathbf{x}\| \leq \|\mathbf{A}^{-1}\|\|\delta\mathbf{b}\|$$

Additionally, the upper limit of the magnification of the relative error of $\mathbf{b}$ on $\mathbf{x}$, which here are the atomic uncertainties and NSD solution uncertainties, respectively, is given by the condition number defined in Equation 4-3, of the matrix in question.[309]

Equation 4-3: Definition of the matrix condition number.

$$\kappa(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\|$$

The calculated $\kappa(\mathbf{A})$ values are given for each of the NSD bases using the ∞-, 1- and 2-norms* (Table 4-1), whilst the 2-norms are most often used for this purpose, all are valid and are therefore given for completeness. Additionally, the square of the condition number ($\kappa(\mathbf{A})^2$) is also provided as this measure is more appropriate for systems that are solved using least-square, which most certainly is the case with the reduced bases and probably also for the complete basis.

---

* Thus providing different, yet equally valid, values for $\|\mathbf{A}\|$ and $\|\mathbf{A}^{-1}\|$. The definition of matrix norms is complex and the interested reader is referred to Cheney and Kincaid for their discussion.[309]

The condition numbers and their squares (Table 4-1) show that all of the basis sets are well conditioned (*i.e.* all $\kappa(\mathbf{A})$ are close to unity). In particular, the min. basis is found to be perfectly conditioned as expected since this basis was orthogonal by symmetry and therefore there is no magnification of the atomic positional errors in the NSD solution. The small increases in the condition numbers of the larger basis sets are due to the non-orthogonality of the modes of the same symmetry. However, since $\kappa(\mathbf{A})^2$ are still sufficiently small, these bases are considered to be well-conditioned and so only very minor magnification of the atomic positional uncertainties, at most, may be propagated into the NSD solution.

Table 4-1: ∞-, 1- and 2-norm condition numbers of the three out-of-plane basis sets. Values in parenthesis are $\kappa(\mathbf{A})^2$.

	min. OOP	**ext. OOP**	**comp. OOP**
$\kappa_\infty(\mathbf{A})$	1. (1.)	1.08815 (1.18408)	1.21039 (1.46504)
$\kappa_1(\mathbf{A})$	1. (1.)	1.08815 (1.18408)	1.23328 (1.52098)
$\kappa_2(\mathbf{A})$	1. (1.)	1.0441 (1.09014)	1.08847 (1.18476)

To conclude, the NSD procedure is a numerically stable procedure and no serious magnification of the atomic positional uncertainties occurs in the NSD solution for any basis. In particular, because the min. basis is completely orthogonal, it is perfectly conditioned and thus its use is preferable with the provision that the δ_{oop} criterion is met.

5 Functional Tetrapyrrole Conformations in Bacterial Photosynthetic Reaction Centres

Tetrapyrrole containing proteins are one of the most fundamental classes of enzymes found in nature and it is an open question to give a chemical rationale for the multitude of biological reactions that can be catalysed by these pigment-protein complexes.[30, 101, 119] A critical factor for all biological functions is the close structural interplay between bound cofactors and the respective apoprotein, in addition to other influences such as hydrogen bonding or electronic effects. In individual tetrapyrrolic systems there is considerable evidence[33, 34, 36, 37] that nonplanarity is the key geometric factor affecting their physicochemical properties and additionally the biological diversity of chemically identical tetrapyrroles has been linked with distinct, conserved conformations.[31, 33, 34, 37] However, except for studies based on individual structures, no convincing general proof has been given for this concept of conformational control thus far.

Using the tetrapyrrole cofactors found in purple bacterial photosynthetic reaction centres (RCs), the first statistically reliable test for this hypothesis in photosynthesis is presented. Here it is shown *via* analysis of all available structural data of the relevant RCs, that the various bacteriochlorophyll (BChl) chromophores possess different conformations, which are likely to impart physicochemical differences that can be correlated with their chemical role. This demonstrates that the protein scaffold exerts functional conformational control on the porphyrin macrocycle and that modulation of the macrocycle conformation is an effective and general means to fine-tune the cofactor properties *in vivo*. Additionally, these results serve as a general model for the structural factors involved in cofactor regulation in nature.

5.1 Many Roles for a Few BChls

Throughout this chapter, the focus is restricted to the six tetrapyrrole cofactors of the electron transfer chain (ETC) from the RCs of purple photosynthetic bacteria and within this restriction, further emphasis is placed upon the four BChls, D_L, D_M, B_A and B_B (Figure 5-1). The basic structure and properties of the bacterial RCs have been described in detail already (see Chapter 1.6.1). The BChl and BPheo cofactors of the ETC are anchored to the L or M protein subunits of the RC and are arranged in pairs of approximate C_2-symmetry.[22] Photo-oxidation of P occurs after excitation *via* transient reduction of B_A that in turn reduces H_A, which subsequently reduces Q_B *via* Q_A.[231] Electron transfer occurs preferentially along the L-branch, despite the apparent C_2 symmetry of the ETC.[233]

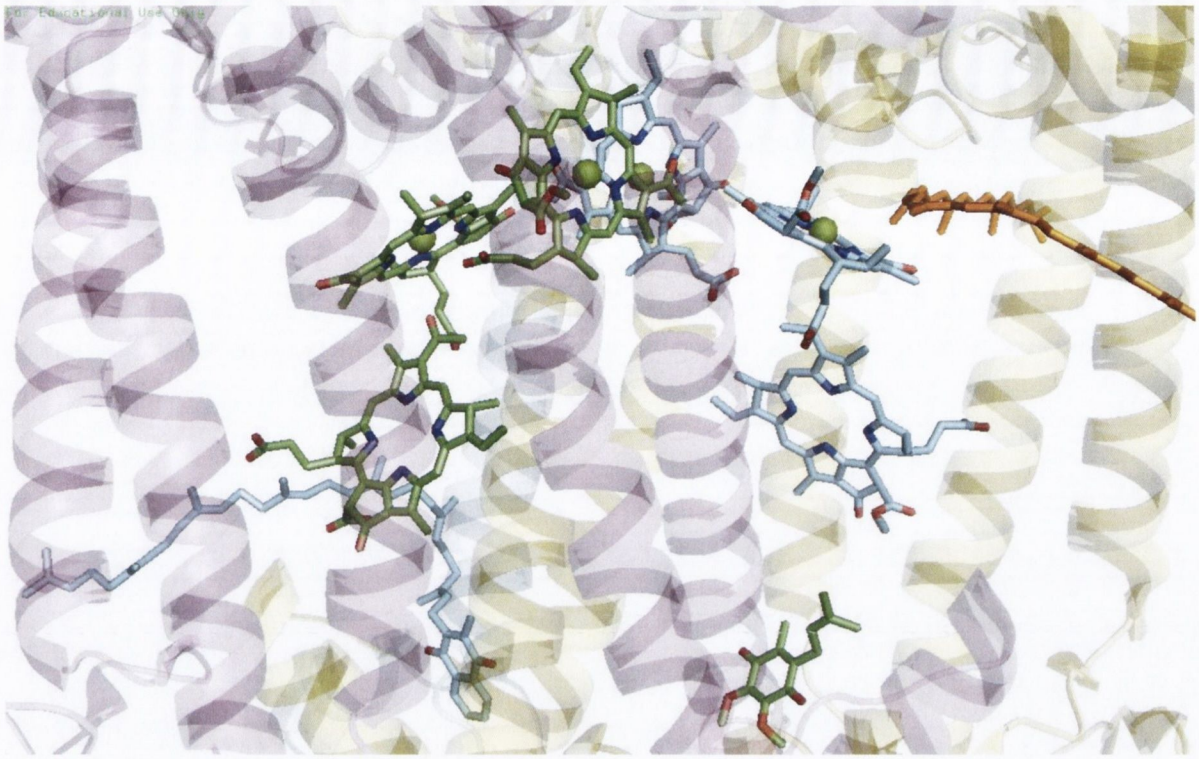


Figure 5-1: The *Blastochloris viridis* reaction centre electron transfer chain as represented by the model coordinates found in Deisenhofer *et al.*'s 2.3 Å structure (PDB ID: 1PRC).[237] Green cofactors are bound by the L-subunit and are from top to bottom, D_L, B_A, H_A and Q_B whereas blue cofactors are bound by the M-subunit and are D_M, B_B, H_B and Q_A. Figure drawn using PyMol.

These first steps of photosynthesis highlight how the processes of light-harvesting, charge-separation and photo-protection all depend upon the tetrapyrrole cofactors.[22, 310] Consequently, the RC is a true testament to their versatility, wherein chemically identical species perform one or more of these highly specialized functions within a single protein. It will be shown that amongst electrostatic effects of the binding-pockets and H-bonding mechanisms, protein modification of the cofactor conformations provides a further chemical rationale for some of the observed details of this process.

The results in this chapter were derived from the normal-coordinate structural decomposition (NSD) analysis of the BChls and BPheos from crystal structures of the RCs from *Rhodobacter sphaeroides* and *Blastochloris viridis*. The approach involved the determination of the NSD for every BChl and BPheo from relevant crystal structures obtained from the Protein Data Bank (PDB; structural summary in Table 5-1). These data were then analysed *via* a methodical statistical route that included investigation of the nature and reliability of each structure with respect to their description of the BChl and BPheo conformations. This was necessary in order to ensure that subsequent aggregation of the experimental data by averaging was appropriate, so that best estimates of the actual conformations present in the crystal were extracted.

Table 5-1: Summary of RC crystal structures included in the analysis of BChl and BPheo conformations.

Species	Crystal structures[a]	Resolution(s)[b]
B. viridis	15 (14)	1.86-3.21 (2.36)
R. sphaeroides	77 (27)	1.80-4.60 (2.66)

[a] Independent RC crystal structures assessed (retained for best estimate in parenthesis); [b] Resolution range (mean) of all structures.

The initial analyses of the NSD data showed that many RC crystal structures exhibited consistent skeletal conformations for each individual BChl in the ETC, demonstrating that the conformations provided by many of the structure determinations were consistent with the idea that the individual binding-sites imposed distinct conformation on each cofactor. Moreover, it appeared that there were both cross-species conservation of the conformations at particular sites as well as distinct differences at others. Another, and perhaps the most crucial motivation for this detailed investigation, was that many of the implied conformational differences represented a departure from the apparent C_2-symmetry of the ETC.

However, owing to the relatively low resolutions of the crystal structures with respect to the variation amongst the macrocycle conformations (Table 5-1), the experimental validity of these early observations was at first an unanswered question. The objective then became to deduce whether or not it was possible to find validated best estimates of the conformation at each site by considering every RC crystal structure that was published in the PDB; the answer to this question in the affirmative being necessary to test the hypothetical concept of the conformational control of porphyrinoid cofactors in nature, in the photosynthetic context. The analysis ultimately provided the best experimentally determined macrocycle conformations of the cofactors to date and will hopefully be of use to computational chemists and crystallographers alike in modelling RC structure and function.*

5.2 Experimental Details

5.2.1 Data-acquisition, processing and NSD

The relevant atomic coordinates were downloaded from the PDB Ligand Expo,†[280] which is a part of the RCSB PDB project.[311] Specifically, the coordinates for every ligand with the component identifier BCL, BCB, BPH and BPB, corresponding to BChls a, b and BPheos a, b, respectively, found in the directory on the 18th October 2010 were

* The coordinates are provided in Appendix A, Table A-15 to Table A-28.
† www.ligand-expo.rcsb.org; last accessed on 29th December 2013

downloaded in '.ipdb' format. Additionally, the coordinates for the BPheos a in structures 2BNP, 2BNS and 2BOZ that were labelled BH1, were downloaded later on 28th March 2011. Note that any PDB structure that contained two RCs in the asymmetric unit provided separate coordinates for the cofactors in each monomer; these PDB structures are identified in Table 5-2 and Table 5-3. The macrocycle of each cofactor was then decomposed into deformations along its normal deformations by the procedure of normal-coordinate structural decomposition[32, 34, 120] (NSD). The NSD computations were carried out using the online NSD Web Program[281] with each cofactor oriented consistently, as described in the program's online documentation (see Chapter 3.2 for further information).

Rhodobacter sphaeroides RC crystal structures

Table 5-2: PDB summary of experimental data from *R. sphaeroides* reaction centre crystal structures that were included in the analysis. Also given are the keys for Figure 5-6 and Figure 5-12 and the classifications of each crystal structure obtained from the analyses described in chapters 5.4.2 and 5.4.3.

PDB ID	Space Group	R_{Work}	Res.	T/ K	Comments[a]	K_1	K_2	RBC[b]	RBP[b]	Ref
1AIG[c]	$P4_32_12$	0.215	2.6	77	Light	1,2	1,2	4	5	312
1AIJ[c]	$P4_32_12$	0.216	2.2	77	Dark	3,4	3,4	4	5	312
1DS8[c]	$P4_32_12$	0.227	2.5	90	Dark; Cd^{2+} PT inhibitor	5,6	5,6	4	4, 5	313
1DV3[c]	$P4_32_12$	0.226	2.5	90	Light; Cd^{2+} PT inhibitor	7,8	7,8	4	4, 5	313
1DV6[c]	$P4_32_12$	0.238	2.5	90	Dark; Zn^{2+} PT inhibitor	9,10	9,10	4	4, 5	313
1E14	$P3_121$	0.226	2.7	100	FM197R/GM203D [B_A]	11	11	4	OL	314
1E6D	$P3_121$	0.174	2.3	298	WM115F/ FM197R [B_A / P]	12	12	4	4	283
1F6N	$P3_121$	0.221	2.8	277	Error in BCL L304 (B_A)	-[d]	13	-	OL	315
1FNP	$P3_121$	0.216	2.6	277	PL209Y	13	14	4	OL	315
1FNQ	$P3_121$	0.217	2.6	277	PL209E	14	15	4	OL	315
1JGW	$P3_121$	0.211	2.8	298	TM21L	15	16	2A	2	316
1JGX	$P3_121$	0.211	3.01	298	TM21D	16	17	2A	2	316
1JGY	$P3_121$	0.218	2.7	298	YM76F	17	18	2A	2	316
1JGZ	$P3_121$	0.215	2.7	298	YM76K	18	19	2A	2	316
1JH0	$P3_121$	0.225	3.5	298	EL205L	19	20	2A	2	316
1K6L	$P3_121$	0.193	3.1	293		-[e]	-[e]	1	1	317
1K6N	$P3_121$	0.203	3.1	293	EL212A/DL213A	-[e]	-[e]	1	1	317
1KBY	$P3_121$	0.195	2.5	298	HM202L [$D_M \rightarrow$BPheo]	-[d]	21	-	2	318
1L9B	P_12_11	0.22	2.4	100	Cytochrome C_2 bound	20	22	2B	2	52
1L9J[c]	P_12_11	0.248	3.25	100	Cytochrome C_2 bound	21,2	23,2	2B	2	52
1M3X	$P3_121$	0.185	2.55	298	Surface lipids [B_A / H_B]	23	25	2A	2	319
1MPS	$P3_121$	0.194	2.55	298	YM177F/FM197R [P]	24	26	4	OL	319
1OGV	$P42\ 2_12$	0.214	2.35	100		25	27	2B	2	320
1PCR	$P3_121$	0.186	2.65	-		26	28	4	OL	321
1PSS	$P2_12_12_1$	0.223	3.0	-	See ref.	27	29	OL	5	322
1PST	$P2_12_12_1$	0.218	3.0	-	See ref. [$D_M \rightarrow$BPheo]	-[d]	30	-	5	322
1QOV	$P3_121$	0.169	2.1	298	AM260W	28	-[e]	4	1	323
1RG5	$P3_121$	0.155	2.5	298	R-26.1	29	31	4	4	324

PDB ID	Space Group	R_{Work}	Res.	T/ K	Comments[a]	K_1	K_2	RBC[b]	RBP[b]	Ref
1RGN	*P3₁21*	0.189	2.8	100	R-26.1 reconstituted with spheroidene	30	32	4	4	324
1RQK	*P3₁21*	0.162	2.7	298	R-26.1 reconstituted with 3,4-dihydrospheroidene	31	33	4	4	324
1RVJ	*P3₁21*	0.218	2.75	100	DL213N/RH177H; [H_A is BPheo b]	32	-[d]	2B	-	325
1RY5	*P3₁21*	0.211	2.1	292	DL213N; [H_A is BPheo b]	33	-[d]	2B	-	325
1RZH	*P3₁21*	0.221	1.8	100	DL213N/RM233C; [H_A is BPheo b]	34	-[d]	4	-	325
1RZZ[c]	*P4₃2₁2*	0.216	2.4	100	DL213N/RM233C; [H_A is BPheo b]	35,36	-[d],34	2B	2	325
1S00[c]	*P4₃2₁2*	0.226	2.6	100	DL213N/RM233C; [H_A is BPheo b]	37,38	-[d],35	2B	2	325
1UMX	*P3₁21*	0.224	2.8	298	RM267L; [H_A/$_B$ is BPheo b]	39	-[d]	4	-	326
1YF6	*P3₁21*	0.197	2.25	100	Quintuple mutant [Q_A deficient; H_A→BChl]	40	-[d]	3	-	327
1YST	*P2₁2₁2*	0.234	3.0	-	*R. sphaeroides* Y	41	36	OL	OL	328
1Z9J	*P4₂22*	0.299	4.5	298	Redox active metal centre	42	37	2A	2	329
1Z9K	*P4₂22*	0.33	4.6	298	As above	43	38	2A	2	329
2BNP	*P4₂2₁2*	0.21	2.7	100	Dark	44	39	OL	3	330
2BNS	*P4₂2₁2*	0.209	2.5	100	Light	45	40	OL	3	330
2BOZ	*P3₁21*	0.174	2.4	100	GM203L [B_A]	46	41	4	4	331
2GMR	*P2₁2₁2₁*	0.215	2.5	100	DL210N	47	42	4	4	332
2GNU	*P4₂2₁2*	0.2	2.2	100		48	43	OL	3	333
2HG3	*P3₁21*	0.164	2.7	100	R-26.1	49	44	4	4	334
2HG9	*P3₁21*	0.179	2.45	100	R-26.1	50	45	4	4	334
2HH1	*P3₁21*	0.178	2.55	100	R-26.1	51	46	4	4	334
2HHK	*P3₁21*	0.172	2.5	100	R-26.1	52	47	4	4	334
2HIT	*P3₁21*	0.182	2.75	-	R-26.1	53	48	4	4	334
2HJ6	*P3₁21*	0.174	3.0	100	R-26.1	54	49	4	4	334
2J8C	*P3₁21*	0.178	1.87	100	Dark, pH8	55	50	3	3	335
2J8D	*P3₁21*	0.197	2.07	100	Light, pH8	56	51	3	3	335
2JIY	*P3₁21*	0.181	2.2	100	AM149W [H_B exclusion]	57	-[d]	4	-	336
2JJ0	*P3₁21*	0.203	2.8	100	AM248W [Q_A exclusion]	58	52	4	4	336
2RCR	*P2₁2₁2₁*		3.1	-		59	53	OL	5	337
2UWS	*P3₁21*	0.202	2.9	100	Light, pH6.5	60	54	3	OL	335
2UWT	*P3₁21*	0.196	2.5	100	As above	61	55	3	3	335
2UWU	*P3₁21*	0.208	2.04	100	Dark, pH6.5	62	56	3	3	335
2UWV	*P3₁21*	0.225	2.13	100	Light pH6.5	63	57	3	3	335
2UW	*P3₁21*	0.205	2.05	100	Dark, pH6.5	64	58	3	3	335
2UX3	*P3₁21*	0.185	2.5	100	Dark, pH9	65	59	3	3	335
2UX4	*P3₁21*	0.19	2.51	100	Light, pH9	66	60	3	3	335
2UX5	*P3₁21*	0.202	2.21	100	As above	67	61	3	3	335

PDB ID	Space Group	R_{Work}	Res.	T/ K	Comments[a]	K_1	K_2	RBC[b]	RBP[b]	Ref
2UXJ	$P4_32_12$	0.194	2.25	100	Dark, pH10	68	62	3	3	335
2UXK	$P4_32_12$	0.19	2.31	100	Light, pH10	69	63	3	3	335
2UXL	$P3_121$	0.176	2.88	100	Dark, pH10	70	64	3	3	335
2UXM	$P3_121$	0.186	2.7	100	Light, pH10	71	65	3	3	335
2WX5	$P3_121$	0.237	2.63	100	FL181R [B_B 6-coord.]	72	66	3	3	338
3DSY	$P3_121$	0.191	3.0	293	EL212Q	-[e]	-[e]	1	1	339
3DTA	$P3_121$	0.199	3.2	293	EL212Q, NM44D	-[e]	-[e]	1	1	339
3DTR	$P3_121$	0.195	3.1	293	EL212Q, LL227F	-[e]	-[e]	1	1	339
3DTS	$P3_121$	0.209	3.1	293	EL212Q, DL213A, RM233L	-[e]	-[e]	1	1	339
3DU2	$P3_121$	0.192	3.1	293	EL212A	-[e]	-[e]	1	1	339
3DU3	$P3_121$	0.191	2.8	273	EL212Q, DL213A, AM249Y	-[e]	-[e]	1	1	340
3DUQ	$P3_121$	0.209	2.7	273	EL212Q, DL213A, NM5D	-[e]	-[e]	1	1	340
4RCR	$P2_12_12_1$	0.227	2.8	-	R-26	73	67	OL	OL	341

[a] Provided by the author to highlight important characteristics of the structure *e.g.*, if the crystals were diffracted under illumination (charge-separated) or in the dark, the identity of point mutations, missing or modified cofactors, the presence of co-crystallised subunits, *etc.*; [b] OL denotes outlier in relevant factor analysis; [c] structures containing two RCs in asymmetric unit and therefore up to two measurements of each cofactor; [d] not included in factor analysis because of erroneous, missing or chemically different cofactor (see comments) and [e] not included in factor analysis since structures do not exhibit any reasonable experimental variation.

Blastochloris viridis RC crystal structures

Table 5-3: PDB summary of experimental data from *B. viridis* reaction centre crystal structures that were included in the analysis. Also given are the classifications of each crystal structure obtained from the analyses described in chapters 5.4.4 and 5.4.6.

PDB ID	Space Group	R_{Work}	Res.	T / K	Comments[a]	BBC[b]	BBP[b]	Ref.
1DXR	$P4_32_12$	0.194	2.00	263	HL168F [D_L lost H-bond; increased P interaction]	1	1	342
1PRC	$P4_32_12$	0.193	2.30	NA		-	1	237
1R2C	$P4_32_12$	0.202	2.86	293		1	1	343
1VRN	$P4_32_12$	0.191	2.20	100		1	1	344
2I5N	$P4_32_12$	0.173	1.96	NA		1	1	345
2JBL	$P4_32_12$	0.190	2.40	263	Stigmatellin inhibited	1	1	346
2PRC	$P4_32_12$	0.182	2.45	263	Ubiquinone-2 complex [Q_B]	1	1	232
2WJM	$P2_12_12$	0.174	1.95	100	Lipidic sponge phase	1	1	347
2WJN	$P2_12_12$	0.172	1.86	100	As above	1	1	347
3D38	$P3_121$	0.194	3.21	100	Crystallised with methyl-β-cyclodextrin	1	OL	348
3G7F	$P4_32_12$	0.177	2.50	100	HM200L, [heterodimer (D_M)]	NA	OL	349
3PRC	$P4_32_12$	0.178	2.40	263	Q_B depleted	1	1	232
5PRC	$P4_32_12$	0.190	2.35	263	Atrazine complex	1	1	350
6PRC	$P4_32_12$	0.184	2.30	263	Triazine complex	1	1	350
7PRC	$P4_32_12$	0.184	2.65	263	As above	1	1	350

[a] Provided by the author to highlight important characteristics of the structure *e.g.*, the identity of point mutations, missing or modified cofactors, the presence of co-crystallised substances, *etc.* and [b] OL denotes outlier in relevant factor analysis

5.2.2 Statistical analyses

Preliminary analysis involved the agglomerative hierarchical clustering (AHC) of each of the cofactors' min. basis NSD coefficients in order to discover whether or not the crystal structure data were consistent with the idea of distinct cofactor conformations at each site in the ETC. Thus, the AHC was performed on data matrices containing observations of the NSD min. basis for each of the cofactors of a given species' (Table 5-4; *i.e.*, species-specific 'pigment-as-observation' data configuration) using the XL Stat plug-in[351] for Microsoft Excel. The AHC results that are reported used the Euclidean distance to build the dissimilarity matrix and Ward's method for the agglomeration.

Table 5-4: Schematic representation of the data-configuration used to compare the NSD described conformations of individual pigments in chapter 5.3 ('pigment-as-observation' matrix; 4 observations per structure).

PDB ID	Pigment ID	NSD Results
1AIG	B_A	B_{2u}, B_{1u}, *etc.*
	B_B	
	D_L	
	D_M	
1AIJ	B_A	
	B_B	
...	...	...

The cofactor NSD data were then re-configured so that the results for all of the cofactors from each individual crystal structure (*i.e.*, PDB entry) were considered as a single observation (Table 5-5; *i.e.*, species-specific 'structure-as-observation' data configuration). The data were collated into $m \times n$ matrices where m was the number of included crystal structures and n was equal to the number of cofactors included multiplied by the number of NSD normal-coordinates that were assessed. For example, a single RC structure would represent an observation of the of the cofactor specific NSD min. basis distortion modes B_{2u} of B_A, B_{2u} of B_B... A_{2u} of D_M and so on. Thus, in the analyses of the experimentally correlated features, where the 12 normal-coordinates of the NSD min. basis represented the conformations of each cofactor, each observation of an RC structure consisted of 12×4 or 12×2 variables for BChl and BPheo analyses, respectively. The result of this approach was the ability to isolate so-called batch effects that are present in particular (sets of) structure determinations *via* variable-axis rotations from methods such as principal components and factor analysis (PCA and FA). All of the analyses presented in section 5.5 were performed on data matrices in this configuration.

Table 5-5: Schematic representation of the data-configuration used to compare the NSD described conformations of the BChls *afforded by each PDB structure* in chapter 5.4 ('structure-as-observation' matrix; 1 observation per structure).

PDB ID	B_A NSD	B_B NSD	D_L NSD	D_M NSD
1AIG	B_{2u} of B_A, B_{1u} of B_A, *etc.*	B_{2u} of B_B, *etc.*	B_{2u} of D_L, *etc.*	B_{2u} of D_M, *etc.*
1AIJ	...	...	...	...
...	...	...	...	...

Kernel density estimations (KDEs), Euclidean dissimilarities, FAs and the PCAs used were calculated in the R programing environment[352] using the *kde*, *daisy*, *factanal* and *prcomp* functions in the *ks*, *cluster* and *base* libraries, respectively. The bandwidth parameters of the KDEs were calculated using the plug-in selector, executed with the *hpi* (or *Hpi* for bivariate KDEs) function also in the *ks* library. The decision as to the number of factors extracted in the factor analyses was made based on initial data analysis that suggested the presence of two major systematic differences between individual structure determinations. The data matrices were untransformed, Varimax rotation was performed in the FAs and the crystal structure scores were computed using the regression method. Once the crystal structure set partitioning was complete, the mean (arithmetic) NSDs were calculated to yield the average conformation of each of the cofactors in each subset of crystal structures. Both the sample standard deviation (SD) and standard error of the mean (SE) are given as measures of the conformational variation;* the SD represents the observed single experimental precision of each of the measured parameters whereas the SE should provide appropriate confidence intervals for the precision of the mean conformations.†

5.2.3 Average Conformations and Centred NSD

In order to obtain the average conformation from the average NSD deformation coefficients the NSD procedure is effectively carried out in reverse by weighting each normal-deformation by the average value of its coefficient and projecting these

---

* These figures are provided in Appendix A, Table A-1 to Table A-14.

† Whilst it is not asserted that the samples are normally distributed, the Central Limit Theorem suggests that the sample mean will be and hence the SE provides approximate 95% confidence intervals for the means. The SDs may provide a useful test to see whether we have accounted for all of the systematic variation of the data by comparing them to the theoretical precision of a single experiment. However, we have not attempted this here as this calculation is not trivial. To do so, one would need to know the approximate experimental error in each NSD at a given resolution and would need to take into account: the theoretical mean atomic positional error, a reduced dimensionality component owing to separation of the in- and out-of-plane coordinates (performed by Jentzen *et al.*[32]), the error in the calculation of the mean plane, the effect of restraints on these errors and possibly, owing to the random distribution of positional error, its probabilistic propagation into each mode given the nodal (phase change) patterns of the different normal-deformations. In addition to this later point, should probabilistic concerns be minor, at the very least one would need to know entry-wise condition numbers for the NSD calculations.

displacements back onto the reference macrocycle used for the decomposition. In a few cases, results have been provided that refer to the 'centred NSDs'. These data were obtained by performing the NSD computation on the mean conformation obtained *via* the aforementioned procedure. The reasons obtaining these values were twofold: to obtain the D_{oop} and δ_{oop} parameters for the *averaged conformation*, as opposed to these parameters' averaged values over BChls from multiple crystal structures and additionally to ensure that the average NSD deformations were representative of the NSDs of the average conformations. This latter point was considered a prudent check since the ext. and comp. basis sets are not completely orthogonal,[32] although substantial deviations were not expected owing to the relatively small dot products of the normal-deformation vectors[32] and the analysis of the numerical stability presented in Chapter 4. In all cases, the centred normal-deformations were virtually indistinguishable from the average values (providing further verification of the numerical stability of the NSD procedure), however, the D_{oop} and δ_{oop} parameters did vary significantly and so these are provided and discussed where relevant.

5.3 Preliminary analysis: Verification of Distinct Conformations

Preliminary analysis of the cofactor conformations provided by all of the available RC crystal structures was performed using agglomerative hierarchical clustering (AHC) in order to identify structures bearing similar conformations in terms of their individual min. basis NSDs. Since in few cases were the resolutions of the available RC structures better than 2 Å, which in part determines the coordinate errors, only the lowest-energy normal-deformations of each BChl and BPheo were clustered since the smaller variations in higher-energy deformations were not likely to be meaningful experimentally significant.[32]

The *B. viridis* AHC revealed six clusters of structures bearing similar conformations, including one cluster containing only a single structure (Figure 5-2). These clusters contained predominantly only BChls from a single site in the ETC. The cluster relationships indicated by the dendrogram imply that the special-pair BChls are very distinct from the accessories and furthermore that each monomer (*i.e.*, D_M and D_L)* are conformationally distinguishable (Figure 5-2, top right). It is also apparent that the first accessory cofactors (B_A and B_B; labelled A1 in Figure 5-2) form a distinct group, but are

* Regarding the cofactor nomenclature, recall that it was noted in chapter 0.0.0 that $D_{L/M}$ is used to distinguish the monomers of the special-pair whilst the analogous $B_{A/B}$ and $H_{A/B}$ are used for the first and second accessories, respectively.

not distinguishable from each other. In contrast, the second accessories (H_A and H_B) form two groups that are each predominantly composed of one of either H_A or H_B, although this is not as complete a separation as in the D_M / D_L case, illustrated by the presence of both BPheos in each cluster. The cluster that consists of a single observation is identified as the D_M cofactor from a 'hetero-dimer' mutant structure where the D_M BChl was replaced with a BPheo through mutation of its coordinating HIS residue (PDB ID: 3G7F);[349] but note its proximity to the main D_M cluster. This is an important result as it indicates that the conformation is induced by the environment of this cofactor and is intrinsic to the site as opposed to the cofactors chemical identity.

The average distortion coefficient along each out-of-plane normal-deformation obtained for the 'characteristic' BChls of each cluster* helps to explain these features of the dendrogram (Figure 5-2, bottom left). Thus, the conformations of the BChls in the B_A and B_B sites are similar in their relative lack of the *ruf* deformation (B_{1u}) and consequent relative planarity. In contrast, D_L and D_M are by considerably nonplanar with a largely *ruf* conformation, yet differ substantially in the extent of distortion such that D_M is the most distorted of the ETC cofactors. The D_M that has been converted to a BPheo (*i.e.*, "D_M(HET)" in Figure 5-2) has a similar nonplanar conformation to the other D_Ms, but resembles the BPheos with respect to its in-plane conformation (not shown).

The AHC of the BChl NSDs from the *R. sphaeroides* RCs also suggests the presence of distinct conformations, although the inter-cofactor relationships are different (Figure 5-3). The special-pair BChls are again conformationally distinct, although in this case they are less separated from the BChl accessories and it is instead the BPheos that are most unique in this species. An important difference is the presence of a cluster that predominantly contains the B_B cofactors, meaning that in this species the first accessories are conformationally distinguishable. Additionally, although clearly unique from the BChls, the mixed composition of the two clusters that contain accessory BPheos (denoted AC2 and AC2*) suggests that they are not conformationally distinct compared to one another. Another key observation is that the cluster containing mostly B_A structures also contains a significant number of the other BChl cofactors (*i.e.* B_B, D_M, and D_L; Figure 5-3, bottom right). Furthermore, owing to the relative planarity of the B_A site (a feature of both species) it should be clear that the 'misclassified' cofactors in this cluster (*i.e.* non-B_A) must also be

---

* For example, the average is computed for only the H_B cofactors in the cluster that predominantly contains H_B so that the H_As are excluded.

relatively planar. This feature also finds correspondence in the BPheo conformations, specifically in the second cluster composed of H_As and H_Bs (AC2*) that is distinguished from the nonplanar AC2 BPheos by their relatively planarity.

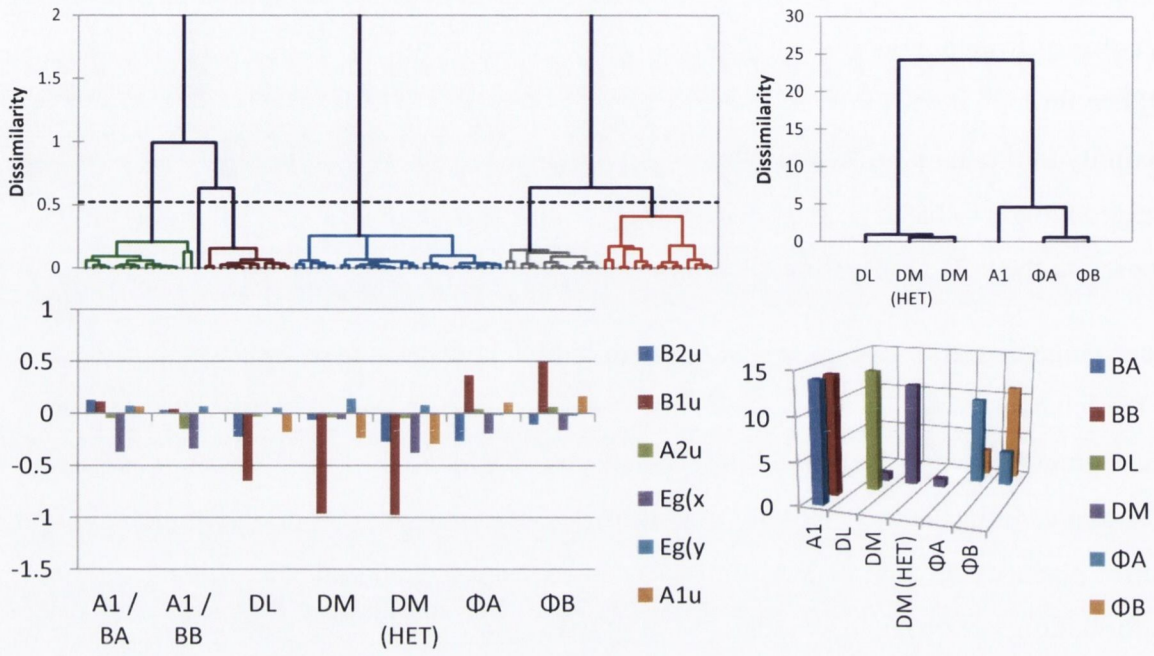


Figure 5-2: Initial AHC of *B. viridis* data. Cluster dendrograms (top) showing the clusters that predominantly contain D_L (green), D_M (brown), A1 (blue; B_A and B_B cofactors), H_A (grey; labelled Φ_A in figure) and H_B (orange; labelled Φ_B) cofactors and their relationship; dissimilarity is given by Wards' criterion. Cluster assignment by cofactor (bottom right); the axis labels indicate the cluster as seen in the dendrogram and the coloured bars enumerate how many cofactors of that functionality were present in the cluster. Average NSD min. basis distortions of the correctly classified cofactors (bottom left; *i.e.* of cofactors that were present in the cluster pertaining to their functionality). n= 84, within class variance = 0.053 (12.59%), between class variance = 0.371 (87.41%).

In contrast to in *B. viridis*, where B_A and B_B are relatively planar and indistinguishable, in *R. sphaeroides* the B_B BChls are the most nonplanar of the cofactors and tend to exhibit a predominant and substantial *ruf* distortion with additional *sad* and *wav* contributions. Additionally, here the distinguishability of the special-pair cofactors arises because the normal-mode composition of D_L is very different to that of D_M as the result of the oppositely phased *ruf* distortion as well as the predominance in D_L of a large *dom* distortion (A_{2u}). Finally, whilst in both species the BPheos H_A and H_B are quite distinct from the BChl cofactors, particularly in *R. sphaeroides*, neither species contains completely differentiable second accessories in terms of which RC protein subunit they are bound by.

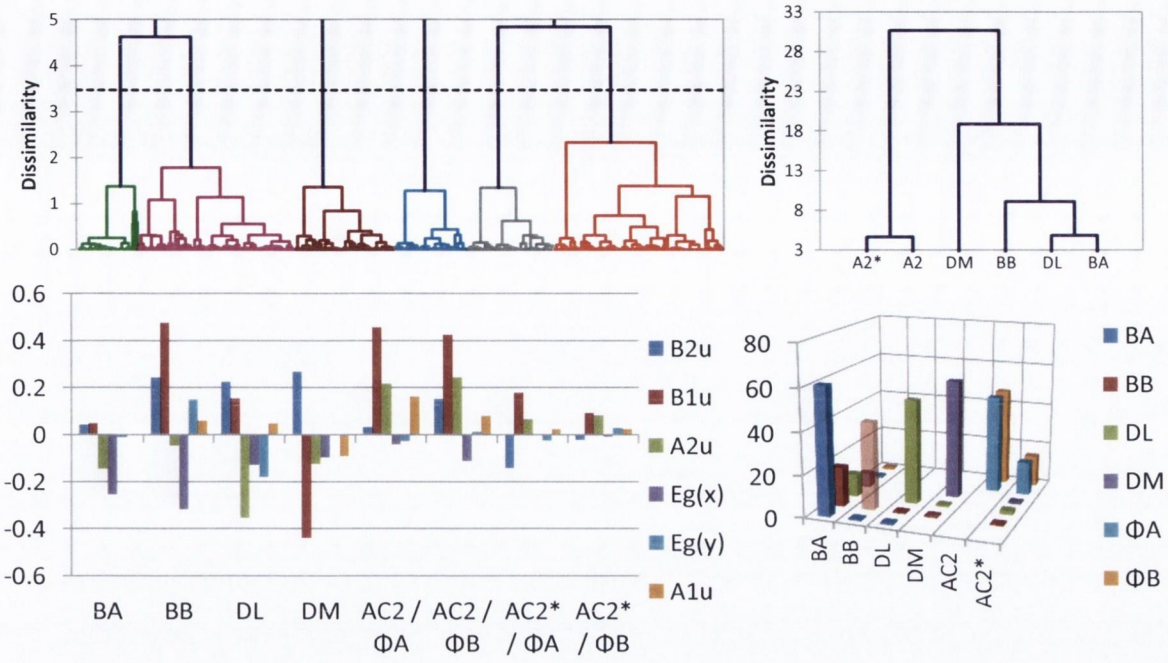


Figure 5-3: Initial AHC of *R. sphaeroides* data. Cluster dendrograms (top) showing the clusters that predominantly contain D_L (grey), D_M (brown), B_A (orange), B_B (blue), and the two clusters containing the bacteriopheophytins, AC2 (pink; primary cluster of H_A and H_B cofactors) and AC2* (green; second cluster of H_A and H_B) cofactors and their relationship; dissimilarity is given by Wards' criterion. Cluster assignment by cofactor (bottom right); the axis labels indicate the cluster as seen in the dendrogram and the coloured bars enumerate how many cofactors of that functionality were present in the cluster. Mean minimum basis distortions of the correctly classified cofactors (bottom left; *i.e.* of cofactors that were present in the cluster pertaining to their functionality). n = 383, within class variance = 0.104 (37.04%) and between class variance = 0.177 (62.96%).

Another feature of note is that in *R. sphaeroides*, meaningful clusters are obtained with a much higher total within cluster sum-of-squares error* than compared to *B. viridis*, as may be illustrated by comparing the heights of the 'tree cutting' lines that determines the final cluster solution. This is of course largely due to the considerably greater sample size afforded by the *R. sphaeroides* dataset but it is notable that in *B. viridis* a minimum of ~0.5 is necessary to achieve meaningful clusters whilst the corresponding value in *R. sphaeroides* is ~2.

These results in general afford the conclusion that at some sites the protein scaffold does impart distinct conformations upon the cofactors, as may be inferred from the overall correspondence of the cluster memberships with the cofactor identity. However, a sceptical position would require that the only definitive conclusion that may be drawn is that the *majority* of the RC structures are consistent with this hypothesis. Closer inspection revealed, for example, that the fewer crystal structures that indicate a relatively planar

* See chapter 3.3 for a detailed interpretation of the dissimilarity index in these dendrograms. Also note that the dashed lines in the dendrograms of Figure 5-2 and Figure 5-3 indicate where the dendrograms were 'cut' to afford the cluster solution presented, the actual total within cluster sum-of-squares error of the solution is the height of the first node beneath this line.

macrocycle for D_M, a BChl that was significantly nonplanar in most others, reported planar conformations for all of the cofactors. Thus, a minority of structures are actually consistent with all of the cofactors having indistinguishable and planar conformations. Fortunately, the possibility of such results was anticipated at the outset of the investigation owing to the fact that the process of crystal structure determination is not simply a direct measurement of atomic configurations. Specifically, chemical knowledge is also usually employed to improve the parameter-observation ratio and provide detail that is difficult to obtain from low resolution studies. With this in mind, the extent of any macrocycle planarity restraints, or in the extreme case, the presence of any such constraints may affect these observations of planar and/or indistinguishable macrocycles (*e.g.* the latter could arise from non-crystallographic symmetry restraints). Alternatively, it may be that the crystallographer decided that owing to the low resolution of the study it was inappropriate to suggest that the detailed conformations of the cofactors could be accurately determined and so accordingly solved this part of the structure using rigid-body refinement.

There was one other such systematic variation exhibited by a number of the *R. sphaeroides* structures that became apparent during the initial AHCs. Notice in the dendrogram (Figure 5-3, top left) that there is a systematic pattern within each of the clusters, which indicates that there is more order to the data than simply the distinct conformations of the sites. Since the complete description of this variation *via* the more detailed AHCs that were performed is inefficient, it suffices to say that by 'cutting the tree' at a lower level, additional subsets were discovered that differed in the extent of their nonplanarity and exhibited an unusual in-plane conformation and then to move on to the mode of analysis which uncovered the patterns of experimental origin.

5.4 Systematic Error: Critical Evaluation of the Crystallographic Data

5.4.1 Justification of the Partitioning of the Crystal Structure Datasets

The significance of the following classification of RC crystal structures based on the reported conformations of the tetrapyrrole cofactors rests on the premise that substantial co-variation amongst the conformations of individual cofactors between experiments is most likely an artefact resulting from different applications of stereochemical restraints. This is a reasonable assumption so long as the possibility of physically 'coupled' conformations resulting from, for example, charge-separated states, perturbations induced by point mutations or steric interaction of the special-pair, are also considered.

Excluding the special-pair, the cofactors should be sufficiently far from one another that any observed systematic differences between all of their conformations in two structures arise because of systematic differences between the experiments and not because of systematic differences between their conformations. To illustrate with an example it is useful to consider two hypothetical crystal structures X and Y. If the four BChl cofactors of X exhibit consistently and quantitatively precise increased nonplanarity over those of Y, such that the relative difference between the conformations of the cofactors within each structure is more or less maintained then two conclusions may be drawn:

1) The experiments are in agreement with respect to the relative differences of this conformational parameter between the specific cofactors.
2) In the absence of any chemical or physical explanation, the effect must arise from a systematic difference in the experiment.

With reference to potential physical explanations for such covariance amongst the cofactors' conformations there are a few possibilities. For instance, in the $P^{+}H_{A}^{-}$ state the altered conformations of the charged species relative to their neutral counterparts would be represented by co-variation in the conformations of P and H_A and if resolved, would manifest as a systematic difference between experiments conducted in the two states. Also, point mutations close to any of the cofactors could affect a change in their conformation, and were these perturbations are larger than the experimental error one would expect systematic differences to be observed between structures that were correlated with particular mutations. However, no such physically interpretable systematic differences were found; it is likely that the combination of the slightness of such conformational changes and the low resolution of the structural data renders such features presently unobservable. On the other hand, it may not be appropriate here to assert categorically that the inter-crystal structure set differences described below do not represent real conformational differences, however unlikely. Of course, even if they do represent real differences, the structure subsets should not be treated together; they are clearly not subject to random error. It is for this reason that details of all the crystal structure subsets that were uncovered have been provided.

5.4.2 *Rhodobacter sphaeroides* BChls

As described earlier, exploratory analysis indicated the presence of systematic differences between sets of whole crystal structures in addition to the inter-site conformational differences. These structure based differences are called 'batch effects' in statistical parlance as they arise from differences between crystal structure determinations each of

which delivers a 'batch' of six cofactor conformations. Knowledge of their precise identity, character, structure composition (*i.e.*, elements of each set) and origin were required to ensure a sound analysis. The systematic process that led to the partitioning of the *R. sphaeroides* crystal structures based on the bacteriochlorophyll conformations into these structure sets, as well as the material required for their explanation, is now described.

Tied structures

In protein crystallography it is common to utilise the final model obtained from a previous experiment upon a homologous structure, if one is available, in order to expedite analysis of the diffraction data (*i.e.*, molecular replacement). From the perspective of this study this procedure could introduce bias into the reported cofactor conformations since, for example, it is conceivable that a structure found in the PDB could provide BChl conformations merely replicated from an earlier structure. A simple way to test the RC crystal structures to find out if such artefacts are present is to check if the BChl NSD data display appropriate experimental variation between each structure. This was achieved by calculation of the Euclidean distance between the NSD coefficient vectors of equivalent BChls from different structures as represented in the 'structure-as-observation' data matrix. Thus, a single distance was calculated between each pair of RC structures to represent the degree of dissimilarity of their BChl conformations; any zero distance indicates that the NSDs of all four RC BChls are identical between those two structures.

The inter-structure Euclidean distance plot (Figure 5-4) shows a distinct group of outliers with near zero dissimilarity that indicates the absence of *any* experimental variability. These structures also clearly do not originate from the same statistical distribution as the rest since they are separated by a sharp discontinuity in the sorted plot. Close inspection of the sources of these outliers shows that they are themselves closely related in terms of their experimental circumstances. Specifically, they were measured at the same time, by the same researchers (Table 5-2; p. 85) and also share a common starting model. This observation confirms the interpretation that this subset contains replicas that do not represent independent measurements and therefore should not be considered suitable in any attempt to reduce random experimental error by aggregation. The structures belonging to this set will be identified as RBC1 (see Table 5-2). It is worthwhile to point out that the conformations provided by these crystal structures remain consistent with the idea of distinct cofactor conformations at relevant sites (Figure 5-5) and also provide a description of their individual conformations that is qualitatively consistent with that provided by the crystal structures that are deemed more reliable. Therefore, the exclusion of these

structures does not alter the conclusion that the sites are conformationally distinct; it is done so in concordance with the spirit of the approach to provide an unbiased, critical appraisal of the data.

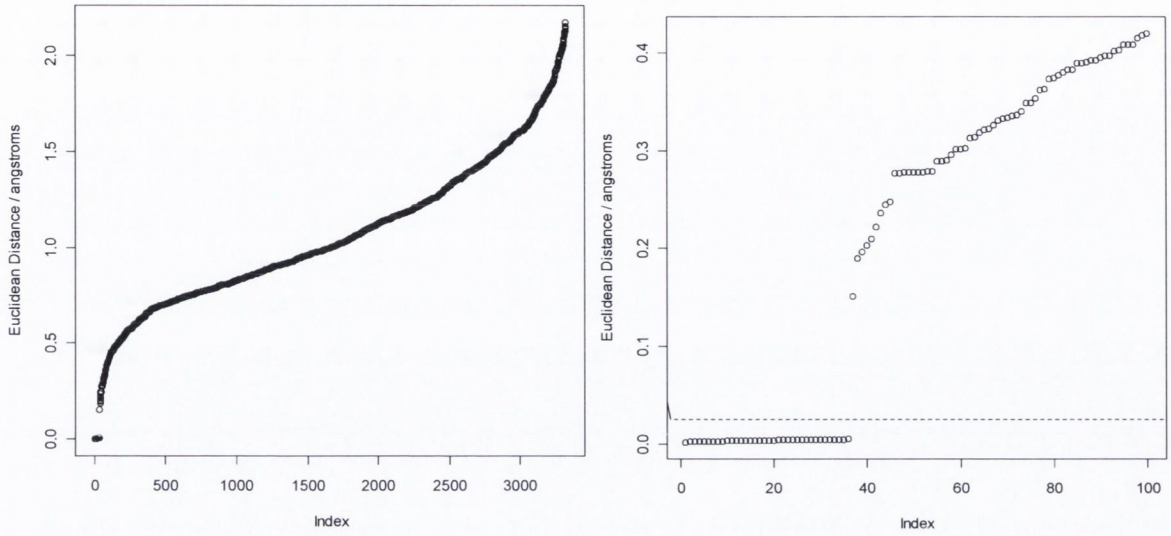


Figure 5-4: Ordered plot of inter-structure Euclidean distances of min. basis NSDs from all *R. sphaeroides* BChl cofactors obtained by constructing the Euclidean distance matrix from the 'structure-as-observation' data matrix. Each point represents the unique pairwise distance between two PDB structures, *e.g.*, d(1AIG,1AIJ) given by the distance between vectors constructed by concatenating the NSDs of the four ETC BChls (*i.e.*, each structure's [B_A NSD, B_B NSD, D_L NSD, D_M NSD] vector).

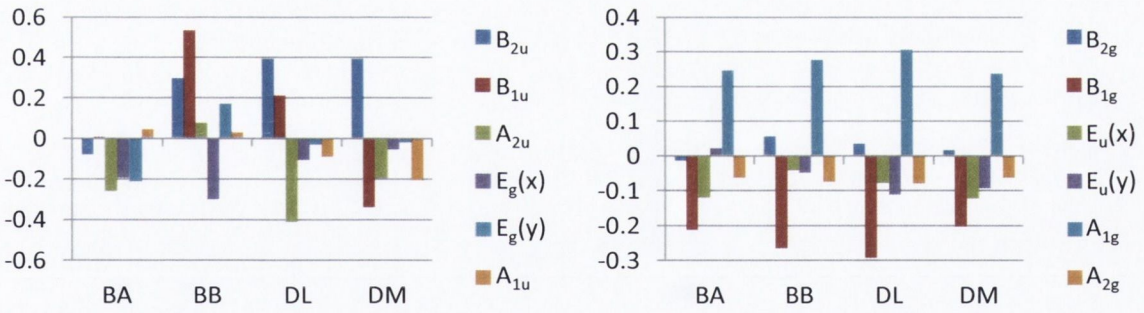


Figure 5-5: Average min. basis normal-coordinate displacements of each ETC cofactor in RBC1 ('no error set'). Out-of-plane (left) and in-plane (right). Y-axes in Ångstroms; SE < 0.00023Å; n = 9.

Factor analysis: Identification of Crystal Structure Experimental Variations

In order to identify more complex features of the experimental dependence of the reported BChl conformations in the dataset (*i.e.*, PDB structure related systematic variation of the BChl NSDs that are site independent) the high-dimensional 'structure-as-observation' BChl NSD data were analysed using principal components analysis (PCA; see Chapter 3.3.2, p. 75) or, when possible, the related exploratory factor analysis (EFA; *ibid*). An optimal EFA was found to consist of two Varimax rotated factors, accounting for 44.1% of the total variance (22.6 and 21.5% each, respectively), extracted from the 'structure as observation' NSD min. basis data matrix. The factor loadings (Table 5-6) demonstrate that Factor 1 (F1) is mostly related to the nonplanarity exhibited by the cofactors in the

structure determination whilst with the additional consideration of the factor score plot (Figure 5-6) Factor 2 (F2) is revealed to differentiate a set of structures with unique in-plane conformations. With the factor loadings in mind, the apparent clusters observable in the plot of the crystal structures' factor scores (Figure 5-6) are readily identified. Structures with large negative scores on F2 exhibit contracted macrocycles together with enhanced nonplanarity, particularly along the higher energy normal-modes (illustrated by negative δ_{oop} loading; Table 5-6) whilst those structures with large positive F1 values have resolved near planar cofactors (reflected in the negative D_{oop} loading).

The average NSD for each of these groups (Figure 5-7 to Figure 5-10) provides further clarity for these assertions. Thus, the average NSDs for the BChls in structures from the top (RBC2a, Figure 5-7) and middle right clusters (RBC2b, Figure 5-8), which are positioned in the planar range of F1, accordingly show definite planar and minimally nonplanar conformations, respectively. Conversely, those clusters with centres around F1 ~-0.7 (RBC3, Figure 5-9 and RBC4, Figure 5-10) are in possession of considerably nonplanar macrocycles. These latter structures, clustered in the lower and upper left quadrants of the factor score plot and labelled RBC3 and RBC4, respectively (Figure 5-6), are of particular interest. Whilst both of these sets exhibit nonplanar BChl conformations that are in qualitative agreement with one another with respect to the general conformations of each BChl (Figure 5-9 and Figure 5-10; *e.g.* the *ruf*, *sad*, *wav(x)* conformation of B_B) they display drastically different in-plane distortions. Indeed, the largest normal-coordinate loading on F2, which provides the distinction between these two sets, is the A_{1g} coefficient that ranges approximately 0.8 – 0.9 over the individual BChls (Table 5-6) so that the location of the RBC3 group (range F2 = -1 – -2; Figure 5-6) reflects their systematically contracted macrocycles (*i.e.*, small A_{1g} coefficients; Figure 5-9). Also, note that structures that fall significantly out with the level 25 regions of the KDE contours will be treated as outliers and omitted from further consideration (OLs; see Table 5-2, p. 85 for identification).

Table 5-6: Loadings of the original cofactor distortion variables onto the Varimax rotated factors obtained from the exploratory factor analysis of the *R. sphaeroides* BChl 'structure-as-observation' dataset. Loadings in the top 25% for each factor are marked in bold to indicate those variables that contribute most to the specified factor. The table is read such that the row label denotes the factor that the loading concerns and the major column headings identify the cofactor to which the indicated distortion in the inner table heading refers *e.g.*, the top left numeric entry indicates that the D_{oop} parameter of D_L has a correlation of -0.857 with F1.

	D_M	D_L	B_B	B_A	D_M	D_L	B_B	B_A
			D_{oop}				D_{ip}	
F1	**-0.857**	**-0.844**	**-0.823**	**-0.767**	0.301	0.082	-0.067	-0.011
F2	-0.406	-0.428	-0.411	-0.457	0.290	0.522	0.264	0.263
			δ_{oop}				δ_{ip}	
F1	-0.435	-0.543	-0.419	-0.512	**0.851**	**0.798**	**0.687**	**0.731**
F2	**-0.743**	**-0.752**	**-0.707**	**-0.733**	-0.210	-0.246	-0.013	-0.072
			B_{2u}				B_{2g}	
F1	**-0.697**	-0.526	**-0.712**	-0.421	0.496	0.243	0.084	-0.115
F2	-0.319	-0.128	-0.273	0.066	-0.297	-0.410	**-0.586**	-0.382
			B_{1u}				B_{1g}	
F1	**0.837**	**-0.618**	**-0.797**	-0.295	0.196	0.194	0.359	0.372
F2	0.341	0.225	-0.357	0.178	0.230	0.262	0.299	0.378
			A_{2u}				E_u(x)	
F1	0.389	**0.784**	-0.100	0.240	-0.382	-0.191	-0.338	-0.273
F2	0.341	0.419	-0.105	0.207	-0.100	-0.209	-0.286	-0.048
			E_g(x)				E_u(y)	
F1	0.467	0.343	**0.745**	**0.735**	0.443	0.167	0.551	-0.014
F2	0.286	**0.645**	0.496	**0.563**	0.010	0.338	0.228	-0.172
			E_g(y)				A_{1g}	
F1	0.272	0.597	-0.397	0.243	0.028	0.136	0.071	0.063
F2	-0.224	0.479	-0.244	0.390	**0.807**	**0.824**	**0.904**	**0.797**
			A_{1u}				A_{2g}	
F1	0.472	-0.477	-0.335	-0.164	0.087	-0.083	0.048	0.022
F2	0.550	**-0.584**	-0.432	0.098	**0.863**	**0.771**	**0.875**	**0.838**

Although discussion of the meaning of these whole structure systematic variations (*i.e.*, variation influencing the resolved conformations of all RC BChls provided by a crystal structure) is best postponed until after the analysis of the corresponding BPheos, it is worthwhile to provide some general comments regarding their interpretation here. The observation of a set of crystal structures that report planar or very near-planar conformations of all of the BChls (RBC2) suggests that in these structures great weights have been placed upon macrocycle planarity restraints and/or the quality of the experimental observations have been insufficient to incorporate nonplanarity into the final model. However, even the crystal structures in these sets suggest that the RC BChls possess distinct conformations and some aspects that are clearly resolved in the more nonplanar sets are observable to some extent (*e.g.*, the negative *ruf* distortion of D_M; Figure 5-7 to Figure 5-10). The interpretation of the differences between the RBC3 and RBC4 sets

is less simple as the structural differences are related to the shape of the BChls more so than to the overall extent of nonplanarity (compare Figure 5-9 and Figure 5-10). However, these differences too are most likely experimental artefacts resulting from different BChl refinement restraints rather than reflective of a real structural change as was discussed *a priori* in Chapter 5.4.1. Regarding the likelihood of which structure set provides the most accurate representation of the actual BChl conformations, two facts drawn from high resolution crystal structures of closely related isolated compounds are important: Chls are generally quite nonplanar even in the absence of a protein matrix and the central metal determines the A_{1g} distortion to a narrow range (0.23 – 0.25 Å for Mg-Chlorins).[157] Thus, one may tentatively expect the average conformations afforded by the RBC4 structures to provide the best estimate of the BChl conformations with the condition that similar patterns of whole crystal structure systematic conformational differences should be observed in the BPheos if they are truly experimental artefacts.

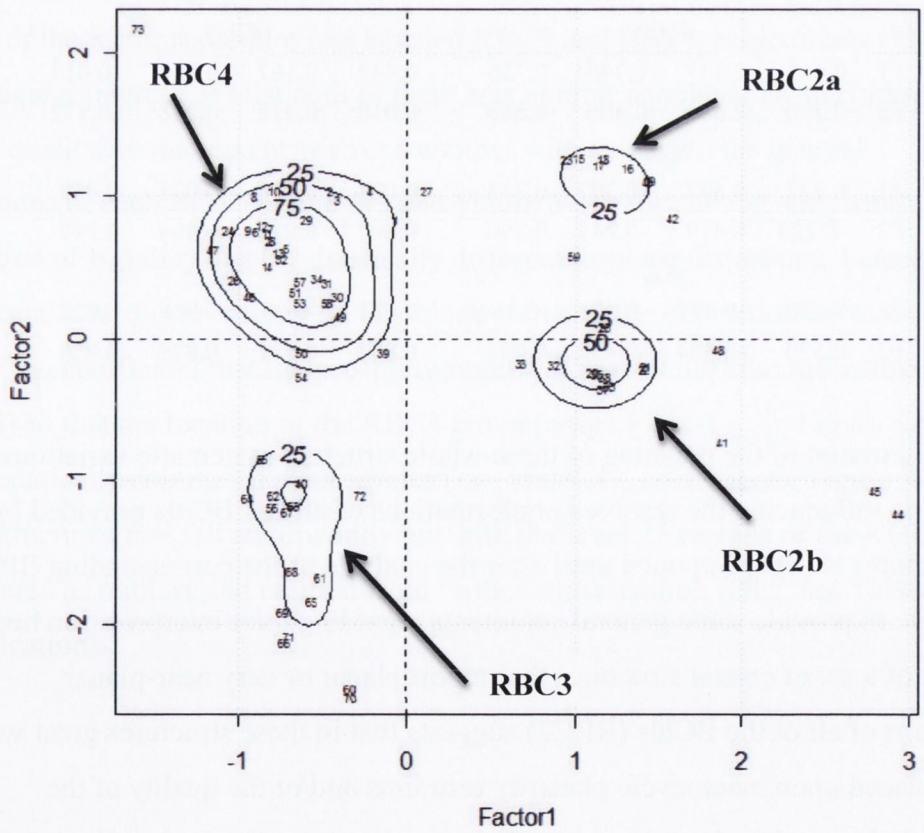


Figure 5-6: Plot of the each crystal structures' 'score' (*i.e.*, F1,F2 coordinate) on the Varimax rotated factors obtained from the exploratory factor analysis of the 'structure-as-observation' dataset. The overlaid kernel density contours provide estimations of the probability density distribution of the crystal structures on this scale. A key for the structure labels is given in Table 5-2 on p. 85 in the column headed "K1".

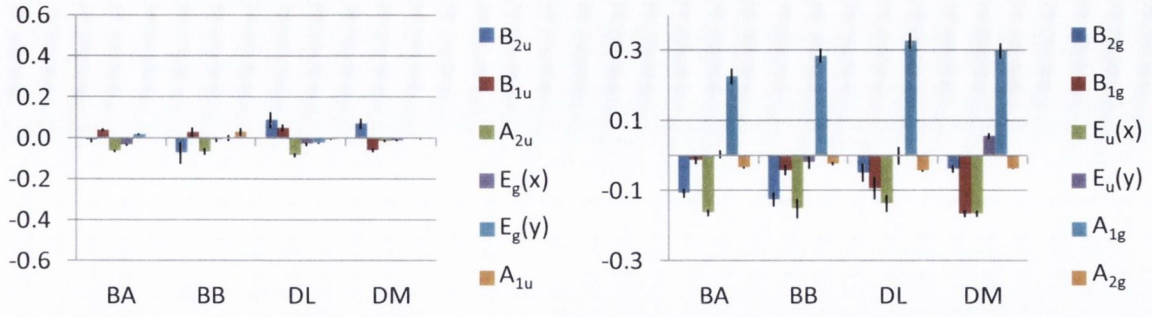


Figure 5-7: Average min. basis NSD of each ETC cofactor in RBC2a ('planar set'). Out-of-plane (left) and in-plane (right). Y-axes in Ångstroms; error bars indicate two SEs. n = 8.

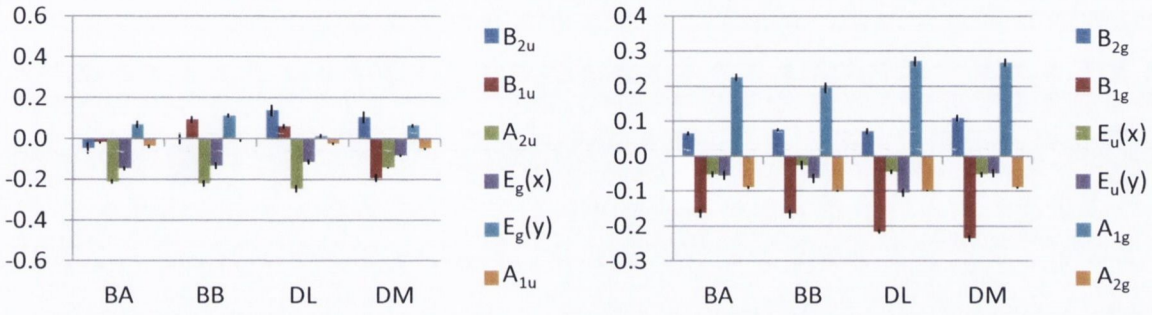


Figure 5-8: Average min. basis NSD of each ETC cofactor in RBC2b ('near planar set'). Y-axes in Ångstroms; error bars indicate two SEs. n = 10.

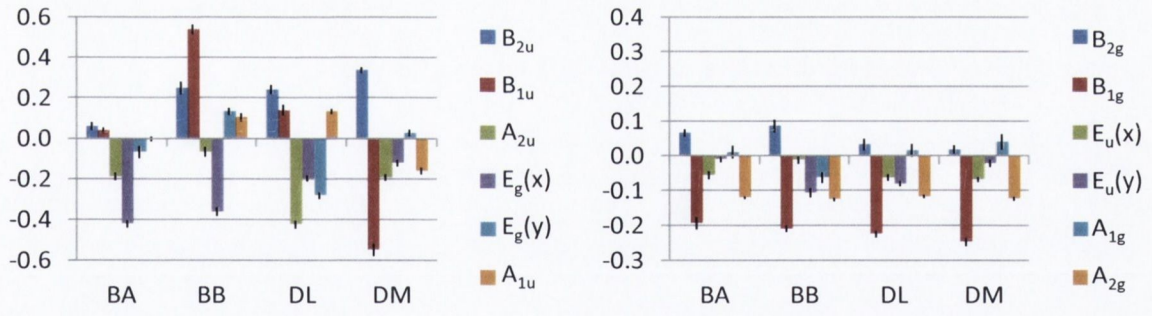


Figure 5-9: Average min. basis NSD of each ETC cofactor in RBC3 (structures exhibiting unusual IP conformation for all BChls). Y-axes in Ångstroms; error bars indicate two SE. n = 16.

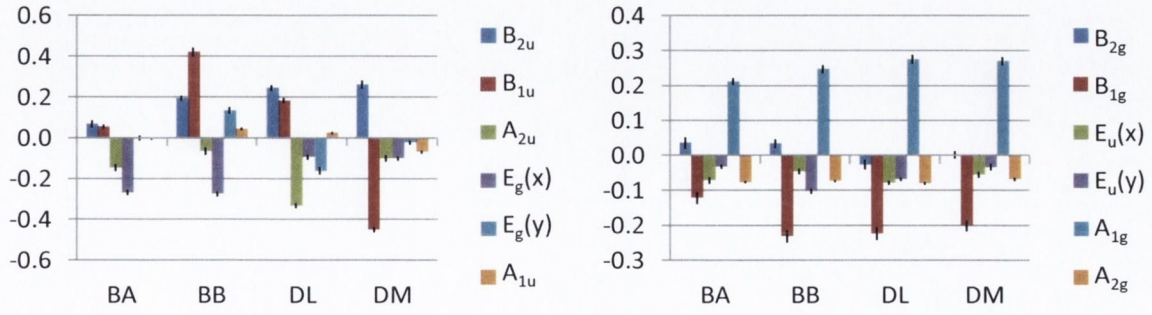


Figure 5-10: Average min. basis NSD of each ETC cofactor in RBC4 (deemed as most reliable set). Y-axes in Ångstroms; error bars indicate two standard errors. n = 32.

In summary, four distinct sets of *R. sphaeroides* RC crystal structures corresponding to the presence of qualitatively different BChl conformational features were identified:

1. Structures that did not exhibit reasonable experimental variation (RBC1; n = 9)
2. Structures that resolved the cofactors as planar or near planar (RBC2; n = 18)
3. Structures with distinct nonplanar conformations for each BChl and relatively contracted macrocycles (RBC3; n = 16)
4. Structures with distinct nonplanar conformations for each BChl and core-sizes consistent with Mg coordination (RBC4; n = 32)

5.4.3 *Rhodobacter sphaeroides* BPheos

Given the nature of the subsets identified within the *R. sphaeroides* BChl data, it was expected that similar groups should arise within that of the BPheos. Moreover, if it was the case that the crystal structures could be analogously partitioned based on the BPheo conformations *via* analysis of the relevant 'structure as observation' NSD matrix, then this would lend significant weight to the idea that these differences are experimental artefacts since such cofactor conformational coupling is even less likely to be real between the BPheos and BChls (see Chapter 5.4.1). Thus, the same analytical methodologies applied to the *R. sphaeroides* BChl data, provided by inter-structure Euclidean distance plots and two-factor EFA, were applied here to the corresponding BPheos.

Tied Structures

The inter-structure Euclidean distance plot based on the *R. sphaeroides* BPheo data (Figure 5-11) shows the presence of a subset of structures that again exhibit effectively no experimental variation (see Chapter 0.0.0 for why this is indicated). Included in this set are the structures from its counterpart in the BChl data, set RBC1 (compare RBC and RBP columns in Table 5-2, p85), which provides additional conformation that these structures do not represent independent measurements. Additionally, in this case there is one extra structure exhibiting a lack of any experimental variance, namely PDB ID: 1QOV, indicating that whilst the BChls in this structure have been refined to the reflection data, the BPheos have not.

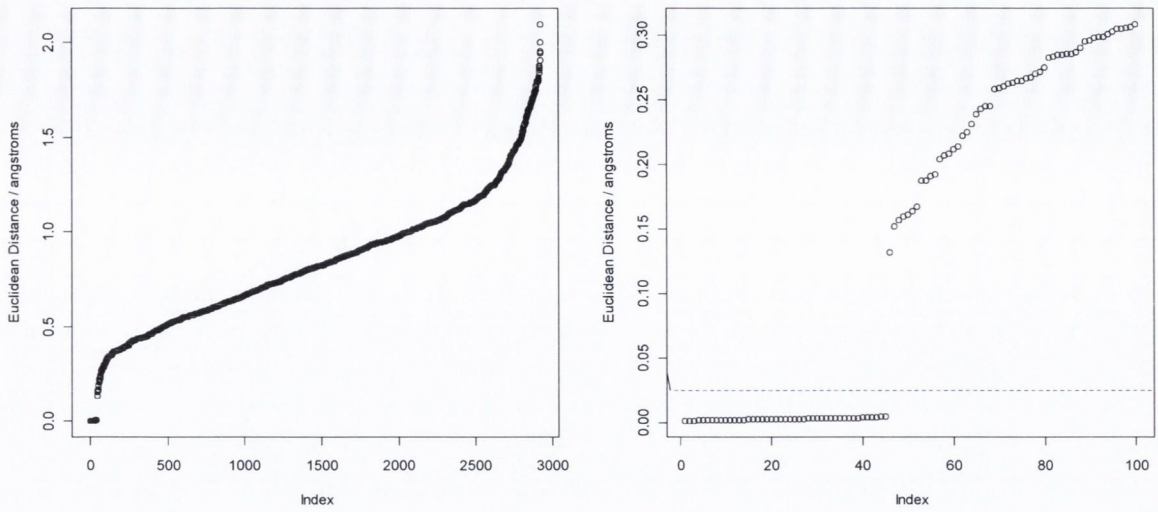


Figure 5-11: Ordered plot of inter-structure Euclidean distances of min. basis NSDs from all *R. sphaeroides* BPheo cofactors obtained by constructing the Euclidean distance matrix from the BPheo NSD 'structure-as-observation' data matrix. Each point represents the unique pairwise distance between two PDB structures, *e.g.*, d(1AIG,1AIJ) given by the distance between vectors constructed by concatenating the NSDs of the two ETC BPheos (*i.e.*, each structure's [H_A NSD, H_B NSD] vector).

Factor Analysis

Similar to the case for the BChls (Chapter 0.0.0), F1 (26.2% total variance) from the *R. sphaeroides* BPheo EFA (Figure 5-12 and Table 5-7) is seen to correlate principally with the observed nonplanarity of the BPheos and serves to differentiate, in order of increasing nonplanarity three subsets, labelled RBP2, -3 and -4. The nature of RBP2 is identified as the BPheo counterpart of RBC2 in both composition (*i.e.*, PDB structures; compared columns RBC and RBP in Table 5-2, p85) and conformational structure (*i.e.*, planar cofactors; Figure 5-13). The differentiation of RBP3 and 4, however, is more

Table 5-7: Loadings of the original cofactor distortion variables onto the Varimax rotated factors obtained from the exploratory factor analysis of the *R. sphaeroides* BPheo 'structure-as-observation' dataset. Loadings in the top 25% for each factor are marked in bold to indicate those variables that contribute most to the specified factor. The table is read such that the major column heading denotes the factor that the loading concerns and the sub-headings identify the cofactor to which the indicated distortion (row label) refers *e.g.*, the top left numeric entry indicates that the D_{oop} parameter of H_A has a correlation of 0.842 with factor 1.

	Factor 1		Factor 2			Factor 1		Factor 2	
	H_A	H_B	H_A	H_B		H_A	H_B	H_A	H_B
D_{oop}	**0.842**	**0.921**	-0.037	-0.087	D_{ip}	0.291	0.187	**0.583**	**0.869**
δ_{oop}	**0.870**	**0.830**	0.213	0.285	δ_{ip}	0.008	0.044	**0.459**	**0.670**
B_{2u}	0.220	0.596	0.032	0.155	B_{2g}	0.213	0.302	**0.681**	**0.579**
B_{1u}	**0.741**	0.663	-0.205	-0.279	B_{1g}	-0.277	-0.105	-0.392	**-0.732**
A_{2u}	**0.758**	**0.797**	0.030	-0.040	$E_u(x)$	-0.389	0.019	-0.142	-0.093
$E_g(x)$	-0.287	-0.438	-0.200	-0.241	$E_u(y)$	-0.047	-0.176	-0.374	-0.315
$E_g(y)$	-0.155	0.056	0.122	-0.234	A_{1g}	-0.018	-0.036	0.117	**0.562**
A_{1u}	0.730	0.431	0.407	0.096	A_{2g}	**-0.841**	-0.038	-0.171	-0.038

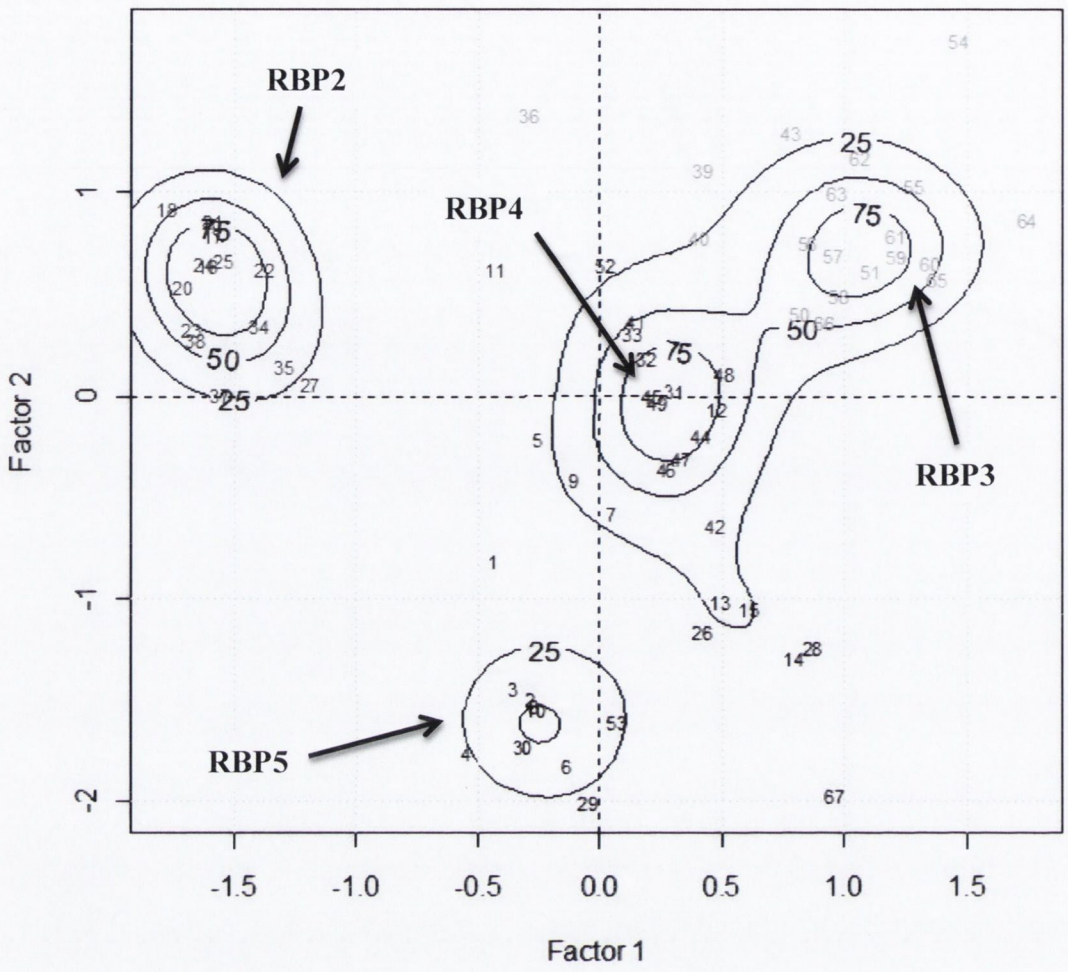


Figure 5-12: Plot of the structure scores, on the Varimax rotated factors, computed by the regression method with kernel density contours overlaid (bandwidth matrix computed using the plug-in selector). A key for the structure labels is given in Table 5-2 on p. 85 by the column headed "K2".

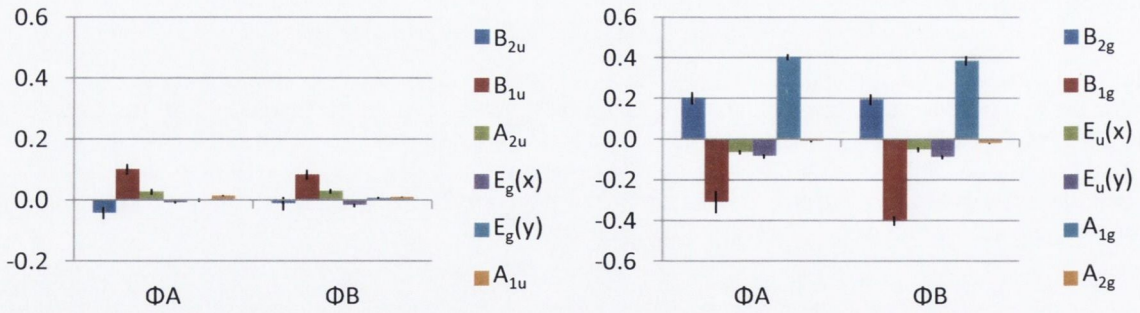


Figure 5-13: Mean minimum basis normal-coordinate displacements of each cofactor in RBP2. Out-of-plane (left) and in-plane (right). Y-axes in Ångstroms; error bars indicate two standard errors and n = 16.

interesting since these sets are also composed of the same structures that are found in the analogous RBC3 and RBC4 groups (Table 5-2, p85) and whilst in this case they are not separated on account of an in-plane A_{1g} contraction, they are separated by numerous in- and out-of-plane parameters (Table 5-7), such that the difference between RBP3 and -4 is quite similar, qualitatively, to that of RBC3 *vs.* RBC4 (Figure 5-14 and Figure 5-15; *e.g.*, RBP3 has greater A_{1u} distortion). This is further emphasised by the large positive D_{oop} and

δ_{oop} on F1 (Table 5-7). Additionally, the total variance accounted for by both factors is 40.3%, which is comparable to that accounted for by the factors in the corresponding BChl analysis (44.1%), providing further substantiation to their relatedness. That F2 in this analysis accounts for less variance than its counterpart in the BChl set is understandable, since detailed comparison of the two analyses (Figure 5-6 and Figure 5-12) reveals that here, RBP3 is differentiated from RBP4 by a *combination* of F2 and F1, as opposed to the differentiation of the corresponding BChl crystal structure sets by F2 alone.

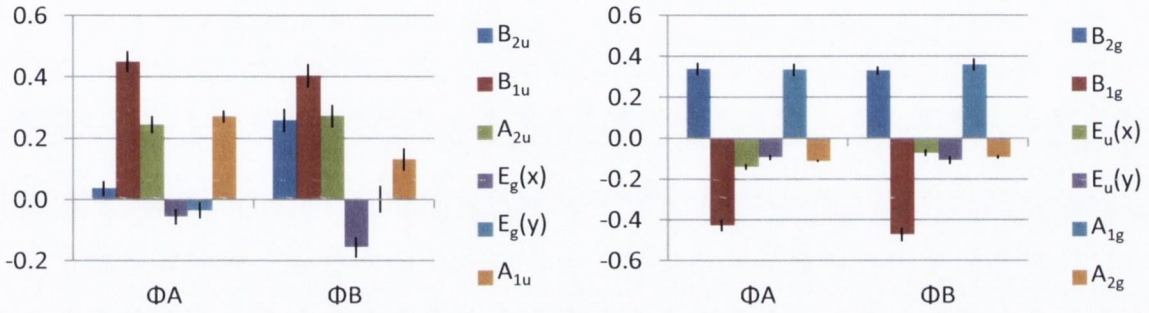


Figure 5-14: Mean minimum basis normal-coordinate displacements of each cofactor in RBP3. Out-of-plane (left) and in-plane (right). Y-axes in Ångstroms; error bars indicate two standard errors and n = 17.

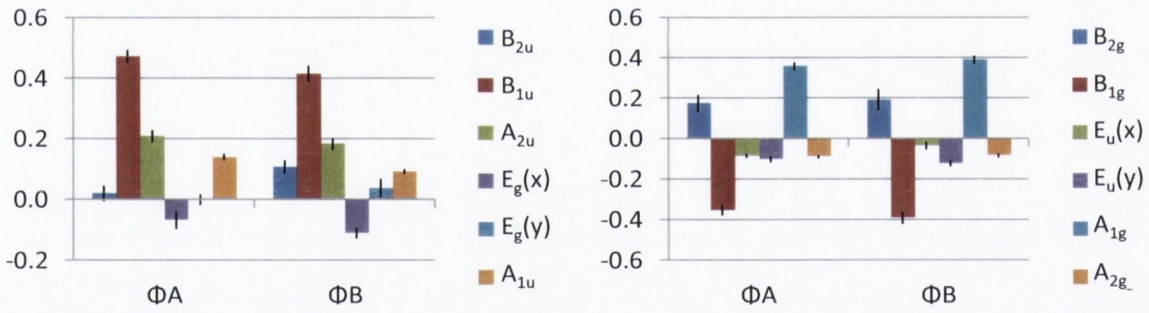


Figure 5-15: Mean minimum basis normal-coordinate displacements of each cofactor in RBP4. Out-of-plane (left) and in-plane (right). Y-axes in Ångstroms; error bars indicate two standard errors and n = 16.

The additional group of structures found in the analysis of the BPheo NSDs, labelled RBP5 in Figure 5-12, is composed of a subset of the PDB structures that were classified as RBC4 in the BChl analysis (Table 5-2, p85). This cluster is distinguished from RBP4 almost entirely by F2 (Figure 5-12), which is largely correlated with various in-plane distortions (Table 5-7). However, although comparison of the averaged NSDs of the RBP4 and RBP5 BPheos indicates core-contraction in RBP5 (Figure 5-15 and Figure 5-16), this should not be considered another analogue of RBC3 since there is no associated increase in the extent of higher energy distortions illustrated by the lack of significant F2 / δ_{oop} correlation. Notably, a number of the RBP5 structures come from crystal structures with more than one RC in the asymmetric unit (AU) wherein the other set of BPheos (*i.e.*, from the other RC in the AU) is classified as RBP4. Although the BPheos are not the focus of this chapter, this

may be a significant observation as it indicates alternative conformations in the same crystal that may quite possibly be the result of differences in crystal packing.

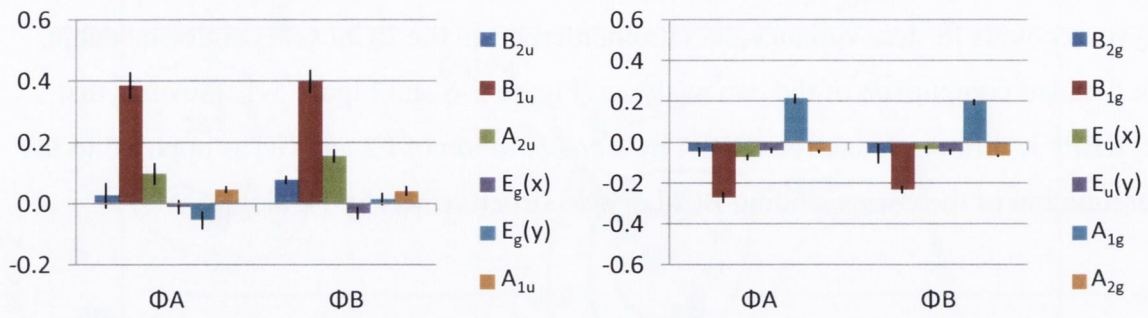


Figure 5-16: Mean minimum basis normal-coordinate displacements of each cofactor in RBP5. Out-of-plane (left) and in-plane (right). Y-axes in Ångstroms; error bars indicate two standard errors and n = 10.

In summary, statistical analysis of the 'structure as observation' NSD min. basis matrix of the *R. sphaeroides* BPheo data confirmed the presence and identity of crystal structure subsets analogous to those found in the corresponding BChl data:

1. Structures that did not exhibit reasonable experimental variation (RBP1; n = 10)
2. Structures that resolved the cofactors as planar or near planar (RBP2; n = 16)
3. A set consisting mostly of the structures found in RBC3 (RBP3; n = 17)
4. A set consisting mostly of the structures found in RBC4 (RBP4; n = 16)

In addition, in this case, the EFA defined another subset (RBP5; n = 10) that had no explicit counterpart in the *R. sphaeroides* BChl dataset but is related to RBC4.

5.4.4 The Nature of the *R. sphaeroides* Crystal Structure Sets

As noted above, the experimental classifications elucidated by the *R. sphaeroides* EFAs are on the whole correspondent by crystal structure over both BChls and BPheos This may be quantitatively confirmed by counting the PDB structures identified in each group (RBC*X* and RBC*Y*; Table 5-8), which shows that aside from outliers, the RBC and RBP groups are largely correspondent by PDB structure including the relationship between RBC4 and RBP4 and RBP5. Since this demonstrates that the systematic variation observed separately in the BChl and BPheo RC cofactor conformations (*i.e.*, variation particular to groups of PDB structures) is not restricted to one particular set of cofactors it is reasonable to infer that these observations are the result of experimental artefacts. Consequently, one set alone gives averaged cofactor conformations that are closest to the actual configurations present in the crystalline form of the RC and the goal now is to identify which one.

Table 5-8: Contingency table of PDB structure count in RBC group (columns) and RBP group (rows). This indicates the number of PDB structures classified as RBC*X* and simultaneously RBP*Y*. For example, there are nine PDB structures classified as RBC1/RBP1 and one structure as RBC4/RBP1 (1QOV, see p. 102). OL denotes outlier; NA counts PDB structures that were not included in the factor analysis and RBP group "4, 5" counts structures containing two RCs in asymmetric unit with one classed as RBP4 and the other as RBP5.

		RBC Classification					
		1	2	3	4	OL	NA
RBP Classification	1	9			1		
	2		13				1
	3			14		3	
	4				13		
	5				2	2	1
	4, 5				3		
	OL			1	5	2	1
	NA		2	1	3		

The simplest sets to identify as erroneous are the tied structure sets that display no experimental variation, namely RBC1 and RBP1. This is because these structures do not represent independent measurements of the cofactors' geometries and their inclusion would therefore overly weight their singular contribution to an average conformation.* Also easy to exclude are the planar structure sets (RBC2 and RBP2) since, as noted previously, high-resolution crystal structures of isolated and related molecules (*e.g.*, Mg-chlorins) exhibit significant nonplanar distortions even in the absence of a protein-induced perturbation as a result of intrinsic strain.[157] In addition to these arguments, a cursory glance at the resolutions of the experimental determinations in each set indicate that structures classified into these groups possessed resolutions at the lower end of the overall dataset (Table 5-9).

We must then decide between sets RBC3 and RBC4 and the corresponding BPheo groups, which both exhibit cofactors that are substantially nonplanar. The key to this rests in their conformational comparison to related small molecule crystal structures. Recall that a major feature of the RBC3 crystal structures is that the cofactors all exhibit a striking absence of the A_{1g} normal-mode or *bre*, which represents the macrocycle core size, when compared to each of the other subsets. This is inconsistent with investigations of related isolated chlorin structures (specifically, Mg-chlorophyllides) that show without exception that this deformation is almost entirely dependent on the central ion that is present in the

* These cofactor structures were most probably obtained *via* rigid-body refinement based on the cofactor coordinates from the oldest structure in the set; for this reason it would have been acceptable to allow a single structure from RBC1 to contribute to the average conformation but since the final dataset is still quite large it was unnecessary to do so.

macrocycle.[157] Intriguingly, the A_{1g} deformation exhibited by the cofactors in RBC3 is similar* to that of the reference macrocycle used to determine the in-plane normal-deformations during the NSD computation (*i.e.* copper(II) porphin)[32] whilst that exhibited by RBC4, the other subsets and the *B. viridis* bacteriochlorophylls is consistent with the required Mg-coordination. Another important factor is that the excessive higher energy distortions exhibited by the RBC3 and RBP3 cofactors are quite unprecedented in porphyrinoid cofactor conformations that are induced by protein effects alone.[32] This is because the energy required for their induction is significantly greater than that required by a combination of only lowest-energy normal-deformations. Finally, it should also be noted that the vast majority (14 of 16) of the structures in the RBC3 set were part of the same study[335] (Table 5-2, p. 85). It is possible then that at some point during their measurement or refinement a systematic error has been introduced that has gone unnoticed since the goal of their study was quite different from determination and analysis of accurate conformations of the BChl cofactors. It is for these reasons, and the appearance of the set's independence of measurement, that the RBC4 (n = 32) structures are concluded to be the best candidates to provide accurate and reliable cofactor conformations.

Table 5-9: Ranges and average resolutions of the PDB RC crystal structures present in each of the identified crystal structure sets.

Classification	Number of PDB structures in group	Resolution(s)[a]
Total in PDB	77	1.80-4.60 (2.66)
	R. sphaeroides BChl analysis	
RBC1	9	2.70-3.20 (3.02)
RBC2	15	2.10-4.60 (2.95)
RBC3	16	1.87-2.88 (2.36)
RBC4	27	1.80-3.00 (2.53)
OL	7	2.20-3.10 (2.76)
NA	3	2.50-3.00 (2.77)
	R. sphaeroides BPheo analysis	
RBP1	10	2.10-3.20 (2.93)
RBP2	14	2.35-4.60 (2.99)
RBP3	17	1.87-2.88 (2.36)
RBP4	13[b]	2.30-3.00 (2.61)
RBP5	5[b]	2.20-3.10 (2.68)
OL	9	2.55-2.90 (2.70)
NA	6	1.80-2.80 (2.32)

[a] mean in parenthesis; [b] an additional three 'half' PDB structures may be added to these entries since these possessed two RCs in the AU, one classified as RBP4 and the other as RBP5. These structures were included in the resolution statistics for both groups.

* This is by virtue of the near 0 Å displacements exhibited by the RBC3 cofactors for this mode. The in-plane normal-deformations are calculated from the Δx and Δy displacements of the structure obtained with respect to the reference macrocycle, thus these modes describe the deviation from this reference (see Chapter 3).

5.4.5 *Blastochloris viridis* BChls

The dataset comprising the conformations of the *B. viridis* BChl cofactors was neither as large nor subject to as many experimental variations as that of *R. sphaeroides* described previously. However, the exploratory analysis revealed a single structure that appeared to resolve less cofactor nonplanarity than the rest of the experiments (analogous to RBC2, albeit not as extreme). The principal component analysis* of the data was found to best illustrate the deviation of this structure, namely 1PRC, and the overall consistency of the rest. The first PC of the *B. viridis* BChl cofactors notably correlates with a number of the nonplanarity parameters (Table 5-10) such that the negative score of the outlying structure (1PRC; Figure 5-17) indicates that its cofactors are consistently less nonplanar than the rest of the data. Again, this is clearly illustrated by comparison of the BChl NSDs of structure 1PRC to the average results obtained for the remaining structures (Figure 5-18 and Figure 5-19) where it is seen that the major distortion mode of each cofactor is significantly smaller in the 1PRC BChls than the rest (*e.g.*, the *ruf* mode of D_M and D_L and *wav(x)* of the accessories). Note also, that this PC accounts for 36.72% of the total variance in the data and that the scree plot is consistent with the presences of only one significant principal component.

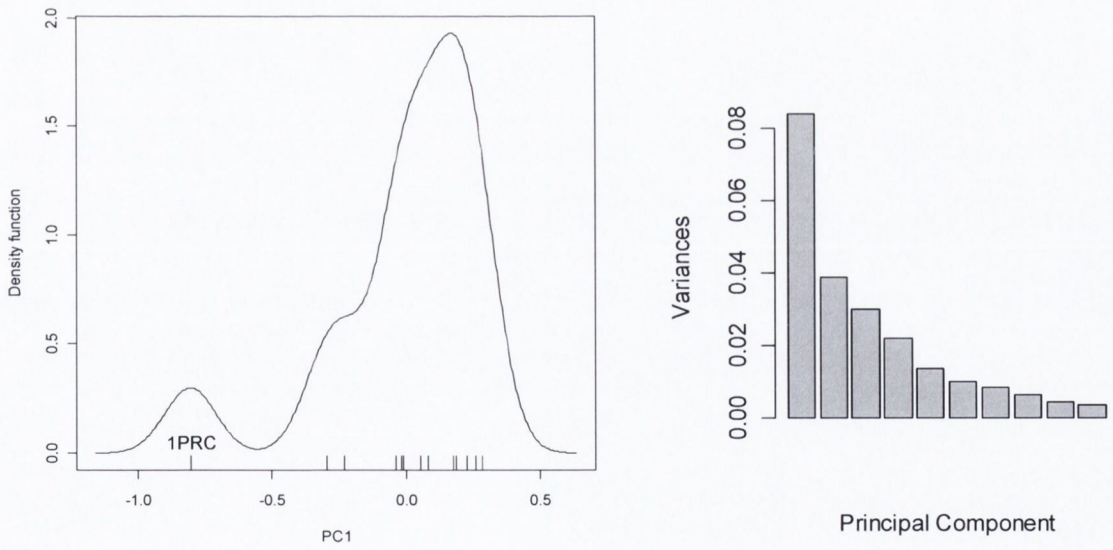


Figure 5-17: PCA of *B. viridis* bacteriochlorophyll dataset. Kernel density estimate of the distribution of the structure scores on the first principal component (right; bandwidth determined by the plug-in selector, h = 0.0957) and the PCA scree plot (left).

* Note that the preference of FA over PCA in the preceding section owed itself to the rotation of the factors allowing their ease of interpretation; FA could not be used here since this data forms an underdetermined matrix.

Table 5-10: Loadings of the NSD normal-coordinates onto the first principal component derived from the 'structure-as-observation' data matrix of the *B. viridis* BChls. Loadings in highest 20% are marked in bold.

	D_M	D_L	B_B	B_A		D_M	D_L	B_B	B_A
D_{oop}	**0.281**	**0.363**	**0.270**	**0.205**	D_{ip}	0.097	0.032	0.115	0.143
δ_{oop}	-0.015	0.001	-0.001	0.002	δ_{ip}	-0.015	-0.014	-0.011	-0.008
B_{2u}	**-0.254**	**-0.159**	-0.157	0.004	B_{2g}	-0.092	-0.016	-0.020	-0.033
B_{1u}	**-0.278**	**-0.344**	-0.048	0.039	B_{1g}	-0.147	0.019	-0.137	**-0.168**
A_{2u}	-0.039	-0.064	**-0.182**	**0.158**	$Eu(x)$	-0.088	-0.048	-0.039	-0.118
$E_g(x)$	-0.107	0.060	**-0.255**	**-0.196**	$Eu(y)$	0.107	0.012	0.010	0.031
$E_g(y)$	0.059	-0.046	-0.005	0.003	A_{1g}	0.144	0.076	0.042	0.063
A_{1u}	0.019	-0.032	0.016	0.010	A_{2g}	-0.040	0.003	-0.020	-0.11

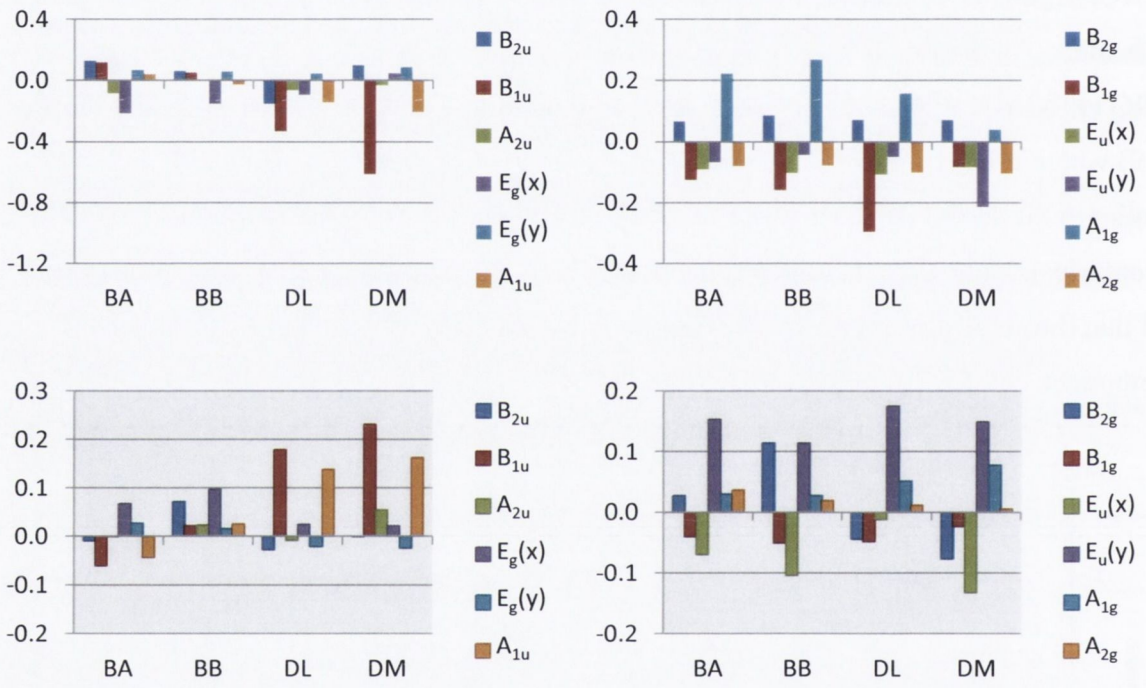


Figure 5-18: Extended basis normal-coordinate displacements of each ETC cofactor in 1PRC. Out-of-plane (left), in-plane (right), lowest energy (top) and next-to-lowest energy (bottom). Y-axes in Ångstroms.

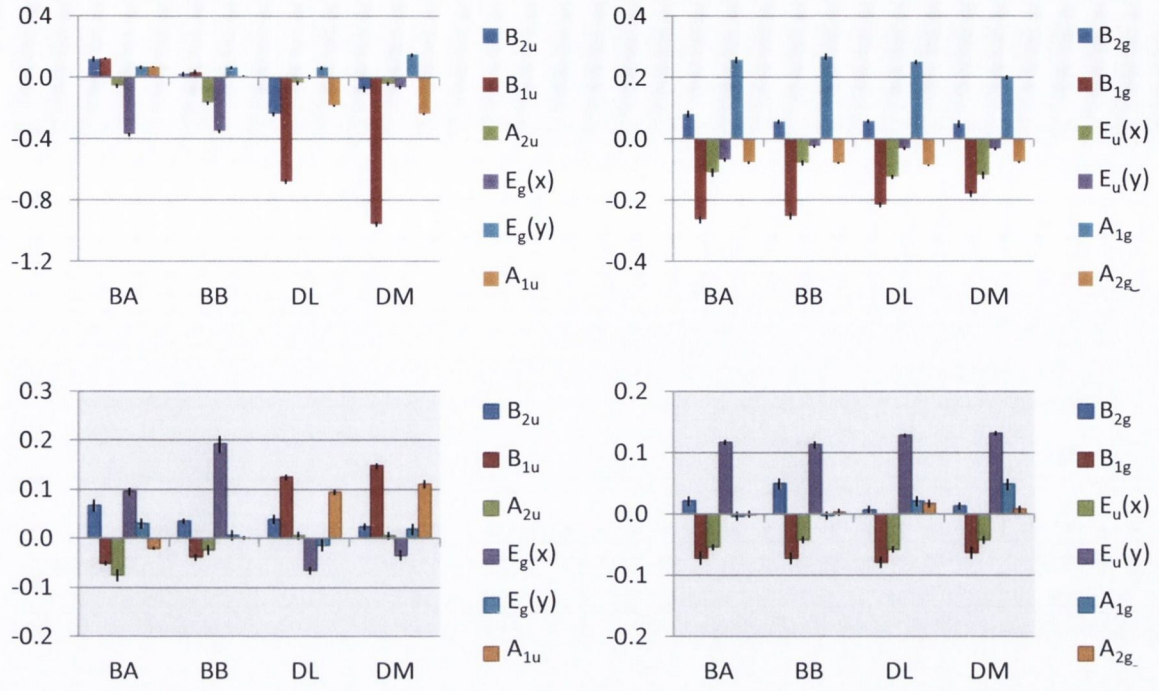


Figure 5-19: Mean extended basis normal-coordinate displacements of each ETC cofactor in BBC1. Out-of-plane (left), in-plane (right), lowest energy (top) and next-to-lowest energy (bottom). Y-axes in Ångstroms; error bars indicate two standard errors; n = 13.

5.4.6 *Blastochloris viridis* BPheos

Although no systematic variation in the *B. viridis* BPheo conformations that correlated with the crystal structure determinations was found during exploratory analysis, it is important to demonstrate this formally since this contrasts with what was found in the *R. sphaeroides* data. As confirmation, the PCA of this dataset (Figure 5-20) did not exhibit any interesting structure in either the scree plot, indicating the absence of dominant PCs; the PC loadings (Table 5-11; although note the D_{ip} and B_{1g} loads) or the distribution of the structure scores in terms of readily interpretable features. However, whilst there is no clear experimentally interpretable structure to the PCA, the two observations to the far right of the plot (Figure 5-20) flag as potential outliers in the KDE and were therefore assessed separately. It was found that these structures exhibited IP conformations that were quite elongated compared to the rest by virtue of their significantly larger B_{1g} (*N-stre*tching) distortions (Figure 5-21 to Figure 5-23; refer also to Figure 3-5, p. 70 for an illustration of the IP modes), thus accounting for the high D_{ip} and B_{1g} loadings on the first PC noted above.

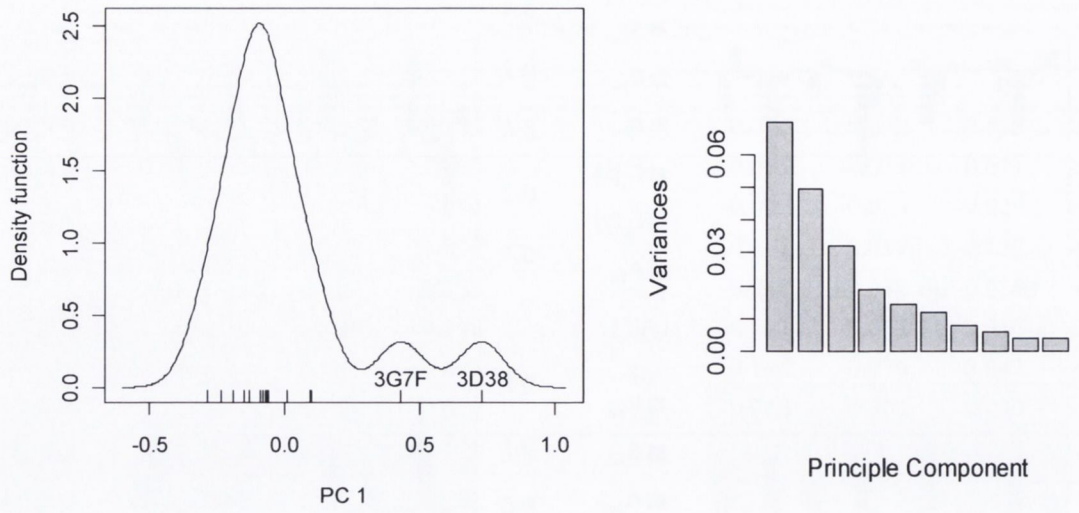


Figure 5-20: PCA of *B. viridis* bacteriopheophytin dataset. Kernel density estimate of the distribution of the structure scores on the first principal component (right; bandwidth determined by the plug-in selector, h = 0.0842 to 3 d.p.) and PCA scree plot (left).

Table 5-11: Loadings of the NSD normal-coordinates onto the first principal component derived from the 'structure-as-observation' data matrix of the *B. viridis* BPheos. Loadings in highest 20% are marked in bold.

	H_A	H_B		H_A	H_B
D_{oop}	0.017	0.020	D_{ip}	**0.310**	**0.350**
δ_{oop}	0.027	0.024	δ_{ip}	0.042	0.040
B_{2u}	0.083	**0.296**	B_{2g}	0.111	0.223
B_{1u}	-0.072	-0.073	B_{1g}	**-0.428**	**-0.309**
A_{2u}	-0.219	0.236	Eu(x)	-0.036	-0.117
E_g(x)	0.114	-0.063	Eu(y)	-0.042	0.002
E_g(y)	**-0.364**	0.170	A_{1g}	0.044	-0.014
A_{1u}	-0.020	0.085	A_{2g}	-0.101	-0.108

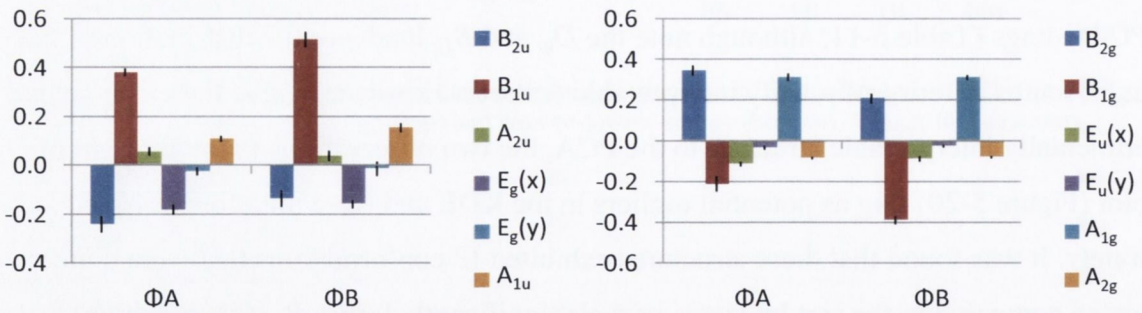


Figure 5-21: Mean minimum basis normal-coordinate displacements of each ETC cofactor in BBP1. Out-of-plane (left) and in-plane (right). Y-axes in Ångstroms; error bars indicate two standard errors and n = 13.

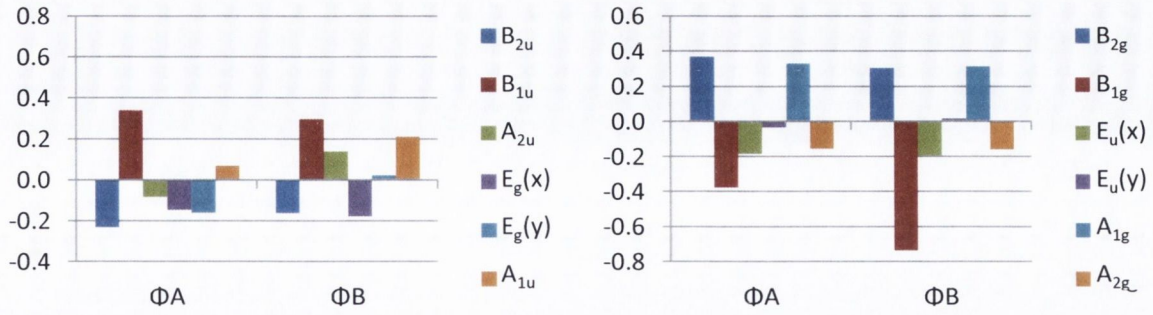


Figure 5-22: Minimum basis normal-coordinate displacements of each ETC cofactor in 3D38. Out-of-plane (left) and in-plane (right). Y-axes in Ångstroms.

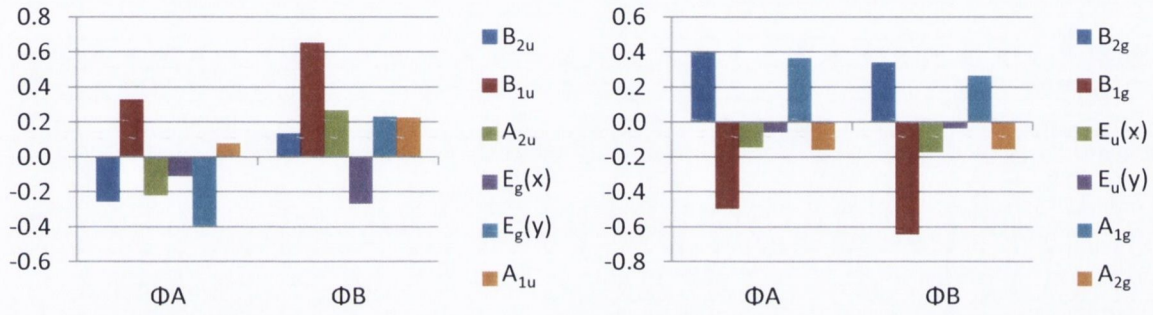


Figure 5-23: Minimum basis normal-coordinate displacements of each ETC cofactor in 3G7F. Out-of-plane (left) and in-plane (right). Y-axes in Ångstroms.

5.4.7 Quality of the *B. viridis* Crystal Structures

For the *B. viridis* BChl cofactors, the arguments for the exclusion of the planar cofactor set in the *R. sphaeroides* data apply also to the decision in proceeding with their analysis using the BBC1 set of structures and not with structure 1PRC, even though the BPheo cofactors of 1PRC do not reflect this feature. This lack of corroboration can be explained though since it is possible that different restraints have been applied to the BPheo in this structure and so, since the conformations of the BChl cofactors of 1PRC are all consistently and significantly less distorted than the rest, it appears to be reasonable to suggest that this aspect is an experimental artefact. As for the apparent outliers in the *B. viridis* BPheo PCA, although assessment of their individual conformations showed that this was due to greater IP distortions as well as slightly different OOP distortion patterns (Figure 5-21 to Figure 5-23, p. 112), their inclusion in the BBC1 set is viable since the BChls of these structures did not exhibit such systematic differences.

5.5 Conformational Control in the Reaction Centre

5.5.1 Overviews of Most Relevant Cofactor Conformations

R. sphaeroides RBC4 cofactors

The NSD min. basis was deemed to adequately represent the conformational variability of the BChls in the *R. sphaeroides* RC since the mean deviation between the simulated and actual conformations, δ_{oop} values (Table 5-12), were either within or close to Jentzen *et al.*'s criterion (*i.e.*, $\delta_{oop} \leq \sim 0.03$).[32] Additionally, the reduction in δ_{oop} of the NSD min. basis of the mean conformations (Table 5-12; *i.e.*, the centred value, see p. 89) compared to the average results over all of the BChls in RBC4 indicates that the min. basis provides a better description of the mean conformation than it does for a typical single observation. This highlights that averaging the conformations over multiple structures has resulted in the removal of apparently spurious higher-energy distortions.

The averaged RBC4 BChl NSDs (Figure 5-24; reproduced from p. 101), which include error bars indicating twice the standard error, illustrate that the cofactors assume statistically significant, unique nonplanar geometries and share a much more common IP conformation. This suggests that the IP conformation is mostly influenced by the chemical structure as this is constant amongst the cofactors whilst the OOP deformations are controlled by their individual protein environments. The key features of the BChl NSDs are that B_A is the most planar (D_{oop} = 0.36 Å; Table 5-12); B_B is the most nonplanar (D_{oop} = 0.58 Å) and displays a large B_{1u} distortion; D_L is quite nonplanar and exhibits a dominant A_{2u} distortion with contributions from B_{2u} and B_{1u}, and D_M is considerably nonplanar showing significant B_{2u} distortion and a large, negative B_{1u} deformation.

Table 5-12: Comparison of the average and 'centred' total distortions (D_{oop}) and min. basis goodness-of-fits (δ_{oop}) for the BChls in the RBC4 *R. sphaeroides* crystal structures. Note that the 'centred' values are from the NSD analysis of the mean conformation as described on p. 89.

		B_A	B_B	D_L	D_M
D_{oop}	Mean	0.3611	0.5848	0.5168	0.5694
	(S.E.)	(0.0122)	(0.0220)	(0.0120)	(0.0146)
δ_{oop}	Mean	0.0340	0.0255	0.0349	0.0205
	(S.E.)	(0.0027)	(0.0025)	(0.0027)	(0.0020)
D_{oop}	Centred	0.3198	0.5623	0.4905	0.5447
δ_{oop}	Centred	0.0268	0.0170	0.0267	0.0110

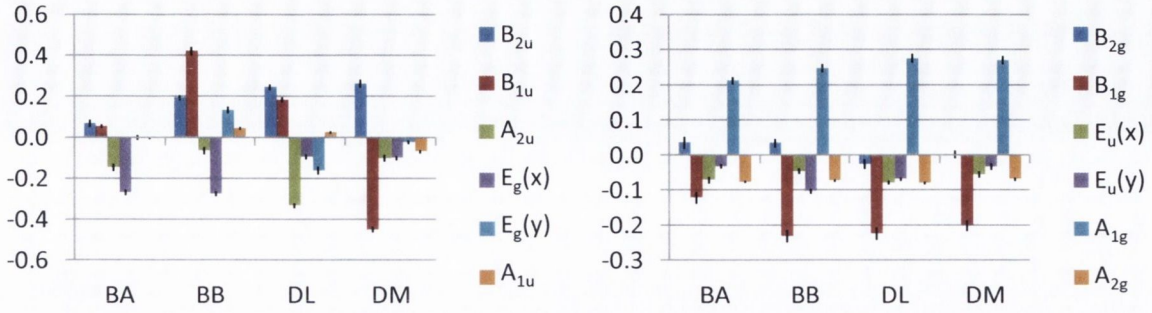


Figure 5-24: Average min. basis NSD of each ETC cofactor in RBC4 (deemed as most reliable set). Y-axes in Ångstroms; error bars indicate two standard errors. n = 32.

The effect of the oppositely phased B_{1u} deformation in D_M in comparison to the other BChls (Figure 5-24) is visible in the skeletal plots (Figure 5-25) and molecular overlay diagrams (Figure 5-26 and Figure 5-27) of the averaged RBC4 conformations where in both the clearest indication is the opposite positioning of the meso carbons with respect to the macrocycle mean-plane. Additionally, the unique doming of D_L (*i.e.*, its A_{2u} deformation; Figure 5-24) results in the tilting of rings A, B and C in one direction along their C_α-C_α axes (Figure 5-25), although the OOP orientation of ring A may be almost conserved between this BChl and B_A. It should also be observed by comparing the actual and min. basis simulated skeletal displacements that the only significant failure of the min. basis, especially for D_L and B_A, is that the sp^3-hybridisation of the β-carbons of the reduced rings B and D leading to their displacement above and below the plane is not entirely encapsulated. Considering this together with the knowledge that averaging has reduced the contributions of high-energy deformations to the BChls in their mean conformations (*vide supra*), one may conclude that the force exerted by the protein upon the cofactor is a general perturbation that induces the lowest-energy distortions of each symmetry, *i.e.*, the NSD min. deformations. This validates the use of the NSD min. basis for the conformational comparisons of these structures since these must be the distortions that are induced by protein/cofactor interactions and, in the case of the special-pair, cofactor/cofactor interactions as well, as these are the most significantly varying deformations across the BChls. To conclude this overview, illustrations of the actual averaged macrocycle conformations (*i.e.*, projections of the RBC4 averaged NSD comp. basis onto the reference macrocycle) have been provided in the context of a possible orientation of the substituents as derived from a single crystal structure (Figure 5-28; see caption for details).

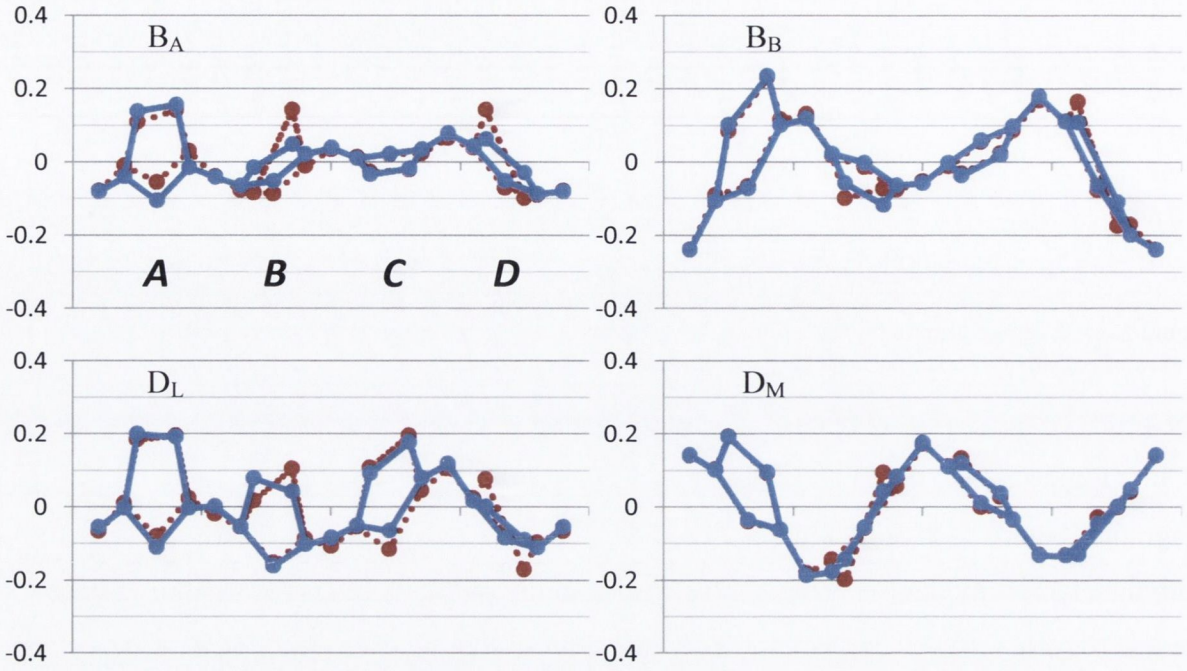


Figure 5-25: Skeletal plots of the four ETC BChl cofactors from *R. sphaeroides* as given by averaging the conformations given by each PDB crystal structure in the RBC4 set. The Δz_i-displacements afforded by the averaged NSD min. basis distortions alone (blue) are compared to the actual averaged Δz_i-displacements (red). The plots shown are of B_A (top left), B_B (top right), D_L (bottom left) and D_M (bottom right); y-axis in Ångstroms. Note that left-to-right show the displacements for rings A – D; the displacement origin indicates the mean plane.

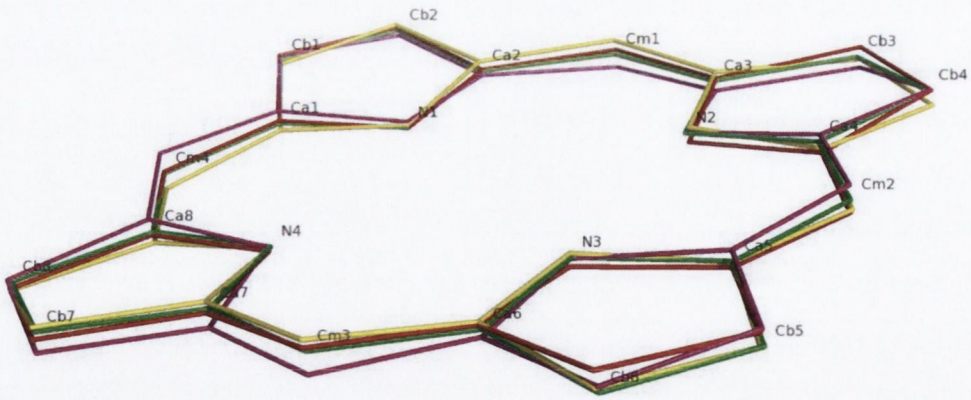


Figure 5-26: Overlay of the four ETC BChl macrocycles obtained from the average min. NSD of the *R. sphaeroides* crystal structures in RBC4. B_A, green; B_B, yellow; D_L, red and D_M, magenta. A diagram indicating the nomenclature used with respect to the BChl a structure is provided in Figure 5-27.

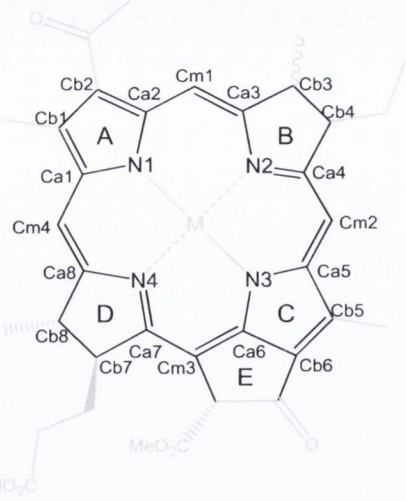


Figure 5-27: The structure of BChl *a* together with the nomenclature used to identify each atom in the overlay plot in Figure 5-26.

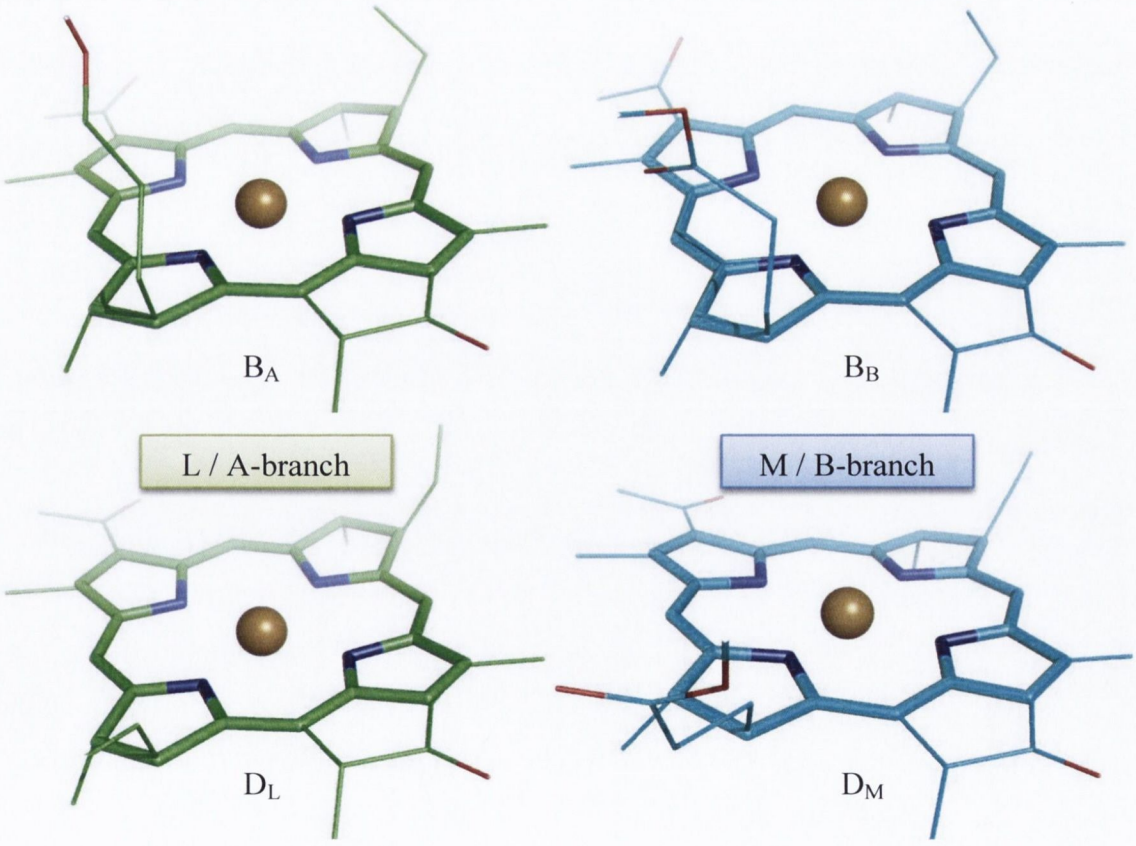


Figure 5-28: Illustrations of the average macrocycle conformations obtained from the RBC4 set of *R. sphaeroides* crystal structures. The substituents and Mg atoms are shown for illustrative purposes only and were appended to the averaged macrocycles from the coordinates in PDB structure 1AIG.

B. viridis BBC1 cofactors

In contrast to the *R. sphaeroides* RC, in *B. viridis* the NSD ext. basis is required to obtain sufficient agreement between the simulated conformations and the observed mean conformation as δ_{oop} for the min. basis is significantly greater than 0.03 Å for all of the BChls (Table 5-13). This is largely the result of strong $B_{1u}(2)^*$ and $A_{1u}(2)$ distortions in the special-pair cofactors as well as strong $E_g(x)(2)$ deformations in the accessory BChls (greatest in B_B) (Figure 5-29, reproduced from p. 111). However, similar to the situation in *R. sphaeroides*, comparison of the 'centred' δ_{oop} to the average values indicates that the ext. basis provides a better description of the mean conformation than it does on average for the BChls from an individual crystal structure (Table 5-13).

Table 5-13: Comparison of the average and 'centred' total distortions (D_{oop}) and min. and ext. basis goodness-of-fits (δ_{oop}) for the BChls in the BBC1 *B. viridis* crystal structures (n = 13). Note that the 'centred' values are from the NSD analysis of the mean conformation as described on p. 89.

		B_A	B_B	D_L	D_M
D_{oop}	Mean	0.4852	0.4897	0.7864	1.0385
	(S.E.)	(0.0132)	(0.0117)	(0.0139)	(0.0136)
min. δ_{oop}	Mean	0.0493	0.0528	0.0383	0.0415
	SE	0.0015	0.0012	0.0012	0.0018
ext. δ_{oop}	Mean	0.0375	0.0367	0.0220	0.0205
	(S.E.)	(0.0019)	(0.0017)	(0.0013)	(0.0015)
D_{oop}	Centred	0.4525	0.4505	0.7672	1.0212
δ_{oop}	Centred	0.0341	0.0340	0.0192	0.0158

In contrast to *R. sphaeroides*, the L/M-branch asymmetry of the special-pair originates from the extent rather than the make-up of the conformation, thus D_L exhibits a large negative B_{1u} distortion, albeit significantly less compared to D_M, and does not possess the A_{2u} distortion that contributed to the uniqueness of D_L in *R. sphaeroides* (Figure 5-29). The similar phase of the B_{1u} mode of D_L and D_M is reflected in directionally correlated displacements of the meso carbons above and below the macrocycle plane (Figure 5-30 and Figure 5-31). Additionally, as was noted in the discussion of the preliminary conformations, the accessory BChls of this species do not exhibit the A/B-branch asymmetry that was found in *R. sphaeroides*, although in this respect, it is worth noting the larger $E_g(x)(2)$ distortion of B_B. Analogous to the case for the *R. sphaeroides* BChls and the min. basis, in *B. viridis* the ext. basis captures the majority of the cofactors' conformations, with the exclusion of the localised perturbation of the sp^3-hybridisation of the β-carbon atoms on rings B and D, which places some of these atoms further than 0.2 Å

* The parenthetic "2" following the symmetry of the distortion denotes that it is next-to-lowest energy distortion of that symmetry that is referred to.

above the mean plane (Figure 5-30). Illustrations of the actual averaged macrocycle conformations obtained from the BBC1 set of crystal structures have been provided (Figure 5-32).

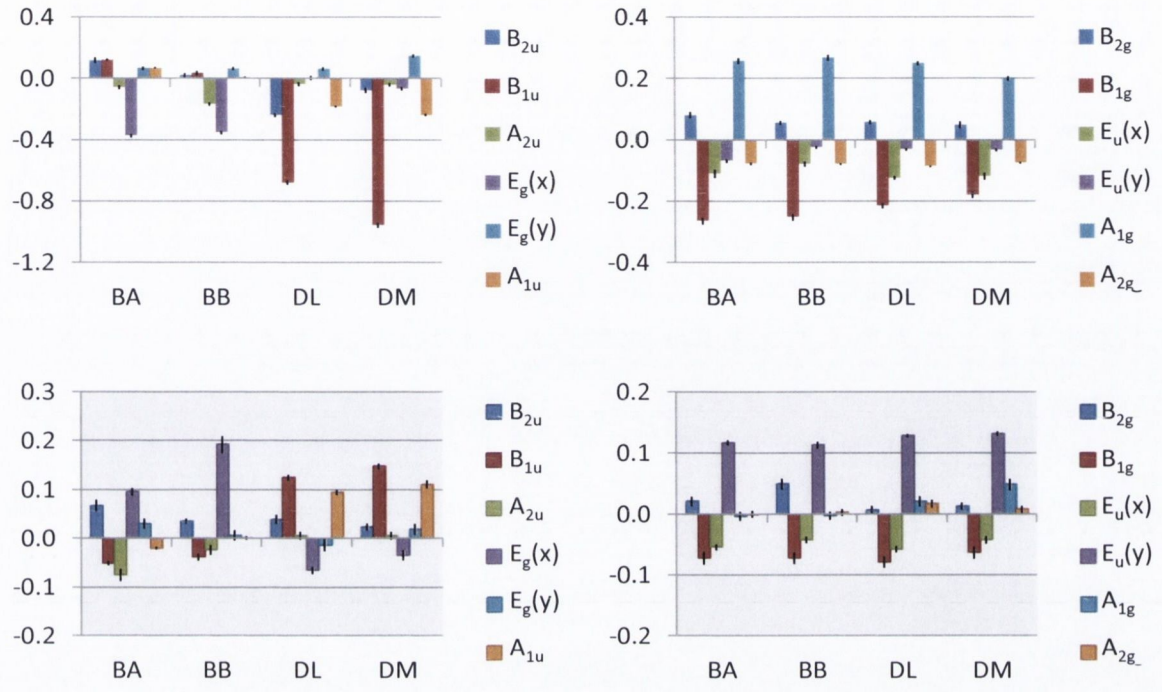


Figure 5-29: Mean extended basis normal-coordinate displacements of each ETC cofactor in BBC1. Out-of-plane (left), in-plane (right), lowest energy (top) and next-to-lowest energy (bottom). Y-axes in Ångstroms; error bars indicate two standard errors; n = 13.

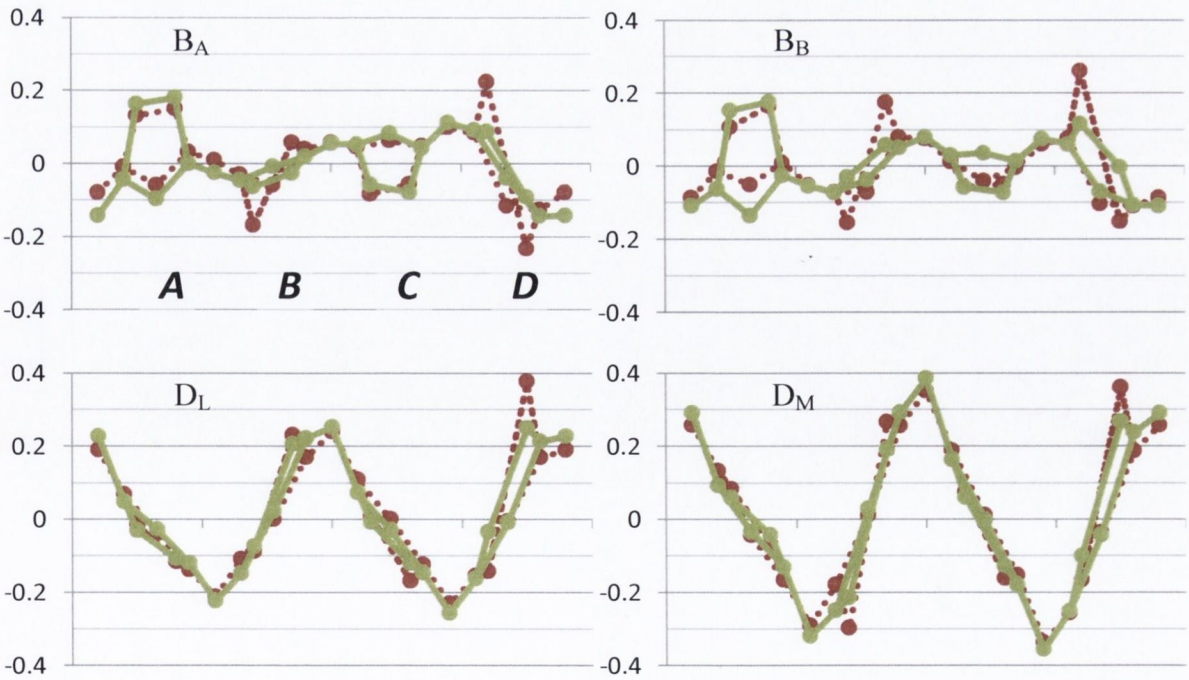


Figure 5-30: Comparison of the extended basis simulated (green) Z-axis distortions with those that are observed (red) for: BA (top left), BB (top right), DL (bottom left) and DM (bottom right); y-axis in Ångstroms. Note that left-to-right show the displacements for rings A – D.

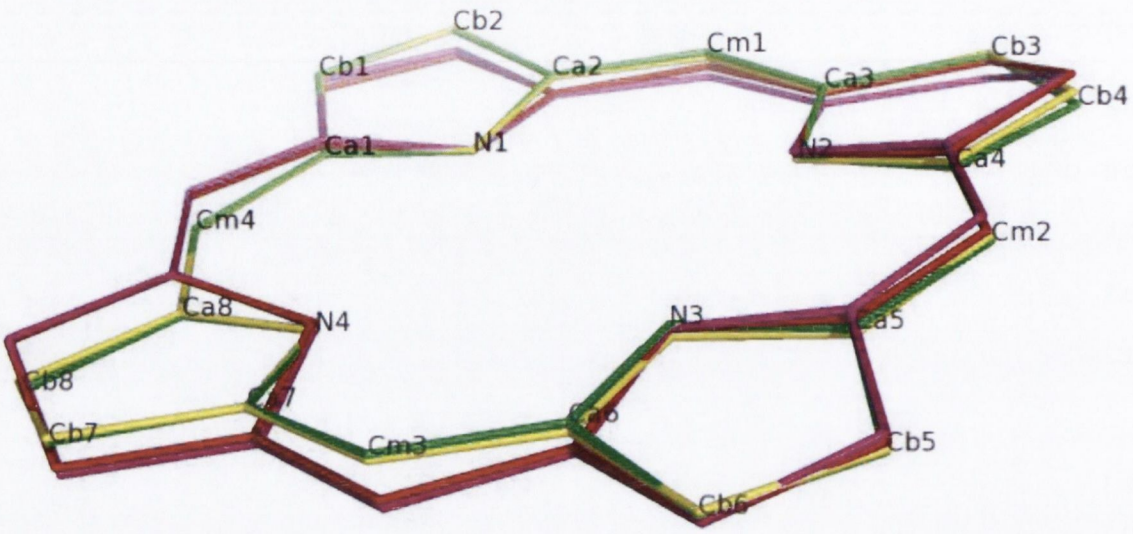


Figure 5-31: Overlay of the four ETC BChl macrocycles obtained from the average ext. NSD of the *B. viridis* crystal structures in BBC1. B_A, green; B_B, yellow; D_L, red and D_M, magenta. See Figure 5-27, p. 117 for key.

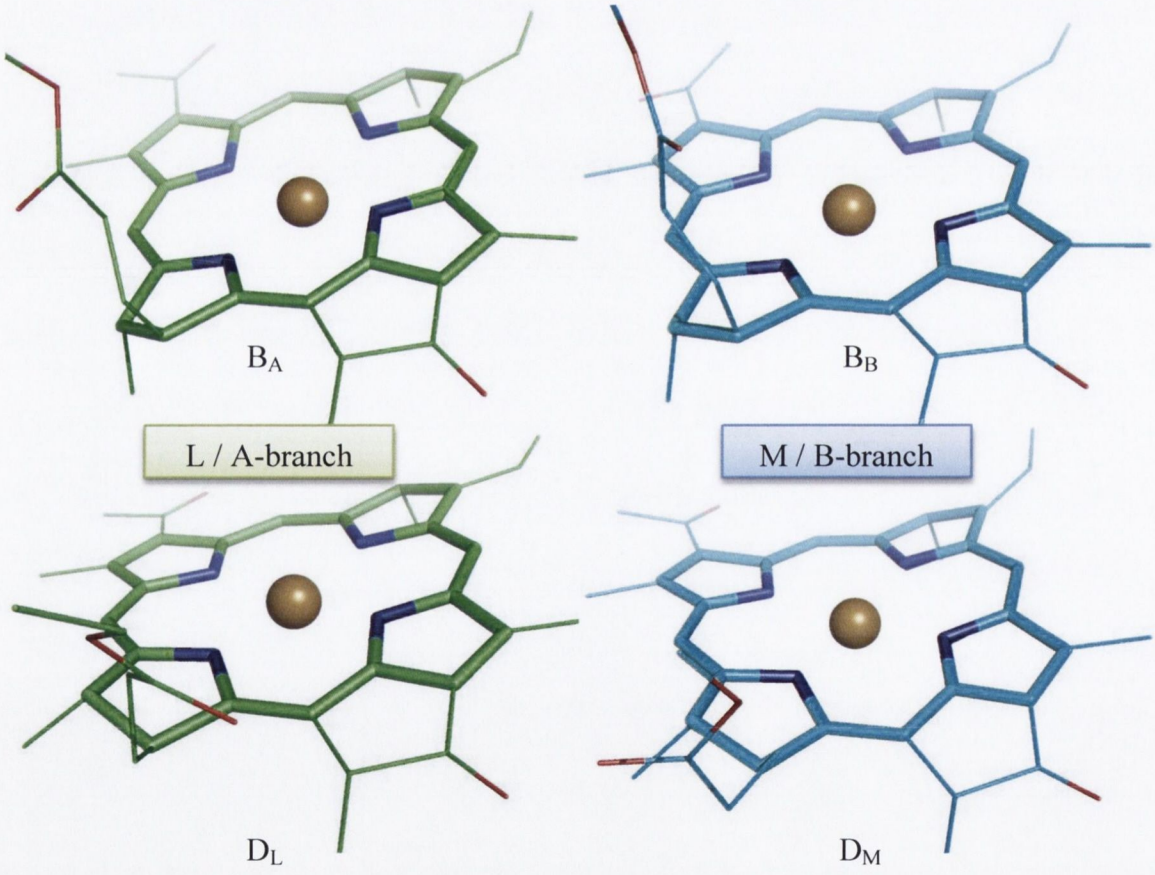


Figure 5-32: Illustrations of the average macrocycle conformations obtained from the BBC1 set of *B. viridis* crystal structures. The substituents and Mg atoms are shown for illustrative purposes only and were appended to the averaged macrocycles from the coordinates in PDB structure 7PRC.

Summary

In general, the deformation of the special-pair (SP) cofactors is always the greatest (of at least the bacteriochlorophylls (BChls)) with a degree of asymmetry existing between the

two, specifically that the M-branch half (D_M) exhibits greater distortion than that of the L-branch (D_L). With respect to the accessory pigments, the BChls are usually the least distorted of all the cofactors and the bacteriopheophytins are considerably distorted. An important deviation from this pattern exists for the M-branch accessory BChl (denoted B_B) of *R. sphaeroides*, in that this cofactor displays a far greater distortion than its L-branch counterpart (B_A) such that its relative distortion is comparable to that of the M-branch special pair cofactor. In both species, the conformations of the accessory BPheos are dominated by a 0.3-0.5Å B_{1u} distortion, with noticeable contributions from other modes. The special-pairs of both species are generally the most distorted cofactors indicating that their mutual interaction is a crucial factor in their nonplanarity, whereas the differences within each accessory pair must be mediated by the apoprotein alone. This indication that the increased distortion of the special-pair cofactors is possibly a consequence of their dimerization points to a new paradigm for its original formation (in addition to their excitonic coupling), since this feature may be crucial to the overall ET process. Note that this is similar to previously described effects in small molecule bischlorin crystal structures.[353]

5.5.2 Physicochemical Effects

In light of the fact that increasing nonplanarity of tetrapyrroles is linked with easier oxidation and more difficult reduction[38] and that photosynthetic electron transfer is subject to Marcus theory,[354, 355] the relative conformations of the BChl cofactors (above; in particular that the SPs are always the most distorted and B_A is always relatively planar) may assist in charge-separation on the whole. The conformations may also contribute to the preference for L-branch electron transfer in both species. Mechanistically, the ET asymmetry is dominated by the initial charge transfer asymmetry of the special-pair in *B. viridis*, which is echoed by the asymmetry in the extent of the nonplanar distortion exhibited by each half of the special-pair dimer in this species (Figure 5-33 and Figure 5-34). In *R. sphaeroides* the asymmetry of ET may begin with the conformational differences in the special-pair but could also be enhanced by stabilization of the $SP^+B_A^-$ intermediate[231] compared to the alternative $SP^+B_B^-$ intermediate, a mechanism that has been suggested to influence unidirectionality[356], recalling the distortion of B_B and that this will affect its hindered reduction. Note, that the rate of electron transfer is greater in *B. viridis* than in *R. sphaeroides*,[46] which again correlates with the greater distortion of the SP of the former (Figure 5-34).

The idea of the conformationally induced initial charge-transfer asymmetry (*i.e.* electron hole asymmetry at the special-pair) is supported by a recent study by Daviso *et al.* who concluded that the preferential localization of electron spin density to D_L of the special-pair radical cation is intrinsic to the dimer and *"... [is] predominantly attributable to local conformational conservation... "*.[357] Thus, the hypothesis is consistent with, and aids in the understanding of, the observed properties of the ETC and provides a rationale for these properties that would be observed as intrinsic in keeping with these recent computational findings. However, to this end, the idea of the easier oxidation of the more nonplanar macrocycle would in fact imply that it should be D_M that exhibits the greater hole density.

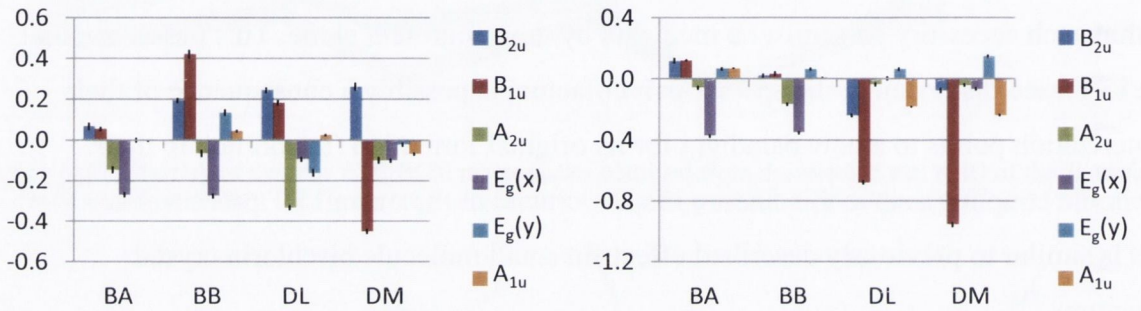


Figure 5-33: Mean out-of-plane minimum basis NSD of each ETC bacteriochlorin cofactor in *R. sphaeroides* derived from RBC4 crystal structures (left, $n = 32$) and *B. viridis* from the BBC1 set (right; $n = 13$). Y-axes in Ångstroms; error bars indicate two standard errors.

On the other hand, not only is the understanding of ET aided by conformational considerations but also that of photo-protection since it has recently been shown that increasing macrocycle distortion affects lower yields of singlet oxygen generation from the triplet state.[40] This implies that the distortion of all but the advantageously planar cofactors (*i.e.* B_A; see above) may be beneficial to the RC in reducing the risk of oxidative damage. In addition, the apparently apoprotein modulated distortion of B_B in *R. sphaeroides* may be a specific enhancement of its photo-protective role whilst the planar B_B in *B. viridis* is concordant with the incapability of efficient triplet quenching in this species (on account of BChl b's low triplet state energy).[358] This argument is also valid in terms of the triplet transfer to the carotenoid since the distortion induced red-shift may be what allows B_B in *R. sphaeroides* to absorb the low energy triplet state of the SP in the first place.

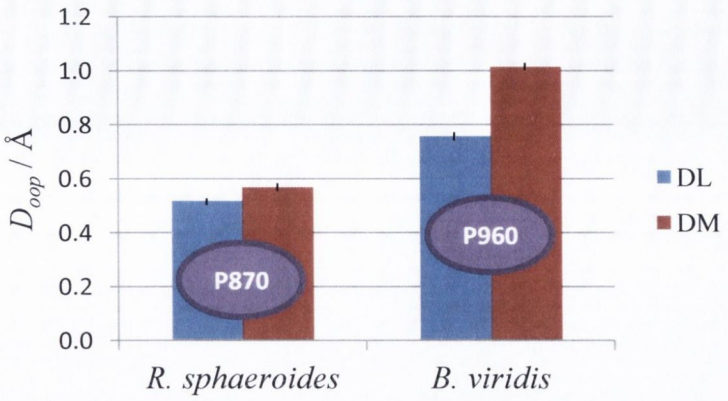


Figure 5-34: Comparison of the total out-of-plane distortion (D_{oop}) of the special-pairs from each species. *R. sphaeroides*, n = 32 (right) and *B. viridis*, n = 14 (left) for each cofactor.

Further experimental observations and reaction centre properties can also be explained with reference to the uncovered conformations. Resonance Raman (RR) studies of RCs yield conformational information on the tetrapyrroles in solution. One such study of *R. sphaeroides*[359] indicated that of the four ETC BChls, one is conformationally unique and suggested that it is D_L that is different. The analysis shows that this cofactor is unique in that its conformation is dominated by a relatively large A_{2u} distortion. Also, RR spectral studies indicate a more symmetric binding environment for the cofactors of P960 (the special-pair of *B. viridis*), yet electron paramagnetic resonance and electron nuclear double resonance studies show that CT is more asymmetric in P960 than in P870.[46] However, we can now see that the difference in overall nonplanarity is greater in the *B. viridis* special-pair. With this in mind the optical and near-infrared absorption spectra[46] of the bacteriochlorin cofactors in RCs of *B. viridis* and *R. sphaeroides* may also be conformationally modulated since the considerable red shift of the Q_y band of P960 compared to P870 (95 nm shift) can only in part be attributed to the different C8 substituents since the isolated compounds absorb at only 770 and 795 nm, respectively, in solution[360] (25 nm shift). However, the NSDs show P960 to be far more distorted than P870; this type of red-shift is another known effect of porphyrin nonplanarity and more specifically, the large $B_{1u(2)}$ distortion exhibited by P960 has been directly liked to such spectral shifts.[36]

5.5.3 Importance of the Cofactor Conformations

Although the results of this study do not allow clear assertions regarding the quantitative effects of the BChl conformations, there are a few techniques that may be employed to provide useful estimates in lieu of a follow-up work. One way to assess the magnitude of the physicochemical impact of the observed conformational variation in the RC is to approximate the distortions' perturbation energy (deformation energy). These energies can

then be compared with those that have been determined for the typical influence of other factors such as hydrogen bonding to the cofactor or electrostatics influences in the binding-site. The deformation energies can be obtained directly from the NSD analysis since, as shown by Jentzen *et al.*,[32] the total macrocycle distortion energy (E_d) is obtained from the linear sum of the contributions of each of the normal-deformations (δE_i), in turn given by the square of the normal-coordinate displacement (d_i^2) multiplied by the respective force constant of the mode (K_i). These energies give reasonable bounds for the potential magnitude of the effect that can be induced by distortion on energetically determined quantities such as reduction potentials or absorption maxima. The force constants used to provide the following estimates of the distortion energies were calculated by Jentzen *et al.*[32]

Equation 5-1: Estimation of macrocycle deformation energy.

$$E_d = \sum_{i=1}^{n} \delta E_i = \sum_{i=1}^{n} K_i \, d_i^2$$

Comparison of the estimated deformation energies (Table 5-14) to experimentally derived estimates of physicochemical changes brought about by other protein interactions from RC mutagenesis studies (Table 5-15) reveals that the nonplanar distortions represent energetic perturbations on a par with those brought about by H-bonding and electrostatic influences. For example, the approximate 100 meV stabilisation induced by each additional H-bond to the acetyl-oxygen of the BChl a monomers constituting the special-pair is certainly matched by the range of distortion energies exhibited by the cofactors of 61 – 107 meV and 105 – 317 meV across the *R. sphaeroides* and *B. viridis* RC BChls, respectively.

Table 5-14: Total macrocycle distortion energies derived from minimum basis NSD of each cofactors mean conformation.

	R. sphaeroides				***B. viridis***			
	BA	BB	DL	DM	BA	BB	DL	DM
E_d / kJ mol^{-1}	5.8	10.3	8.2	6.2	11.4	10.1	16.4	30.5
E_d / meV molecule^{-1}	61	107	85	65	119	105	170	317

Table 5-15: Influences of both H-bonding to- and the introduction of polarised residues near P870 (adapted from Allen and Williams).[19]

	H-bonding	**Polarised residues**
Q_y-band	-8 to -22 nm shift (blue-shift) of Q_y upon removal of DL acetyl-HIS H-bond	Unavailable
$P/P^{\bullet+}$ potential	420 to 765 mV (+60-125 mV / H-bond; 100 meV Hückel model stabilisation)	(Δ) -60 to +50 mV (-ve charge / +ve charge 10 Å from donor; exponential dependence on ε)

Concerning the validity of this approach in terms of simple MO theory, distortion of the macrocycle from its preferred geometry raises the total molecular energy and thus may affect different MOs in different ways (depending on the particular conformation) so that some may be destabilised or stabilised and others may remain unchanged. Of course, the total effect must be a destabilisation but not all of the perturbation will necessarily be localised to the HOMO, the orbital that influences the ionisation energy. Furthermore, the force-constants do not possess any information regarding the energetic disturbance of unoccupied orbitals, including the LUMO, so that the effect on the electron affinity and lowest excitation energy need not be related to the distortion energy.

Another facile analysis of the perturbation energy is afforded by Zucchelli *et al.*,[218, 219] that a distortion of a given symmetry modifies only the energy-level of an orbital with the same idealised symmetry. This allows specific estimation of the effect of the cofactor conformations on the Gouterman frontier orbitals of the BChls enabling a, potentially, more balanced approach in which to directly assess the conformationally induced spectral shifts of the Q_y-band as well as redox potentials (Table 5-16 and Table 5-17).

Table 5-16: Normal-deformation energies of the modes of the minimum basis associated with the two highest and lowest energy molecular orbitals of the BChl cofactors by symmetry.

Orbital	D_{2h} orbital symmetry	Matching (D_{2h}) normal-mode(s)	Perturbation / meV							
			R. sphaeroides				*B. viridis*			
			B_A	B_B	D_L	D_M	B_A	B_B	D_L	D_M
LUMO+1	B_{3g}	*sad* / *wav(y)* [reduced axis]	0	16	24	7	5	3	8	16
LUMO	B_{2g}	*wav(x)*	51	53	6	7	98	90	0	3
HOMO	A_u	*ruf*	1	31	6	35	3	0	79	157
HOMO-1	B_{1u}	*dom*	9	2	48	4	1	12	1	1

Table 5-17: Distortion induced shift of the LUMO←HOMO and LUMO+1←HOMO-1 transitions predicted by like-symmetry assumptions of Zucchelli *et al.*[218]

	R. sphaeroides				*B. viridis*			
	B_A	B_B	D_L	D_M	B_A	B_B	D_L	D_M
$\Delta E_{B2g \leftarrow Au}$ / meV	50	22	0	-28	95	90	-79	-154
$\Delta E_{B3g \leftarrow B1u}$ / meV	-9	14	-24	3	4	-9	7	15

Since it has been shown that the $B_{2g} \leftarrow A_u$ transition contributes greater than 90% of the Q_y-band transition,[218] the first row of entries in Table 5-17 provides an interesting estimate of the spectral shift due to conformational distortion (relative to a planar macrocycle). Thus, a blue-shift is predicted for all of the BChl accessory monomers, a slight red-shit for D_M of *R. sphaeroides* and a considerable red-shift of the components of the *B. viridis* special-pair (P960). Thus, taking the solution Q_y-absorption maximum of 795 nm for BChl b as

reference (in this model, assumed to be planar), $\Delta E_{B2g\leftarrow Au}$ for D_L predicts a shift to 837 nm whilst in D_M the estimated maximum would be shifted to 882 nm. Within the limits of the preceding approximations then, the red-shift of P960 relative to P870 may be assisted as much by conformational control as it is by the greater excitonic coupling of the *B. viridis* special-pair.

Another approach to the question of the magnitude of the physicochemical changes associated with the varying degrees of distortion of the cofactors is to draw on experimental results from conformationally designed synthetic porphyrins. The relevant results from two such studies concerning the effect of a gradual, sterically induced *ruf* distortion of nickel porphyrins on redox and absorption properties,[39] and that of a graded saddle distortion of free-base and zinc porphyrins[40] on their photophysical behaviour have been extracted and are illustrated below (Figure 5-35 and Figure 5-36). These studies also demonstrate that the potential physicochemical effects of macrocycle distortion of the extent observed in the RC are significant in the context of the other protein modulated influences.

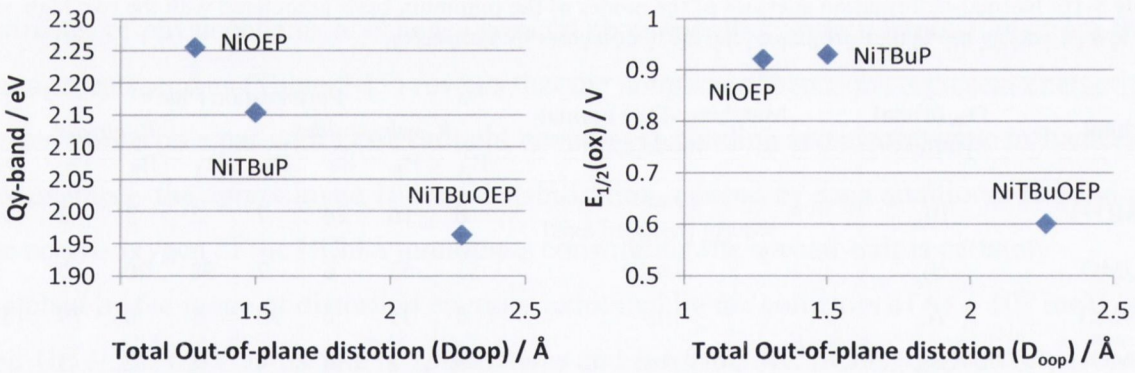


Figure 5-35: Illustration of the relation between the total out-of-plane distortion (D_{oop}) and the energy of the Q_y-band (left) and the first half-wave oxidation potential (right) a series of conformationally designed, sterically crowded *ruf* distorted Ni-porphyrins; data from Senge *et al.*[39]

For example, the approximate one Ångstrom increase in the total distortion of (5,10,15,20-tetrabutyl-2,3,7,8,12,13,17,18-octaethylporphyrinato)nickel(II) (NiTBuOEP) relative to (2,3,7,8,12,13,17,18-octaethylporphyrinato)nickel(II) (NiOEP) leads to a 0.3 V reduction in the first half-wave oxidation potential and a 0.29 eV reduction in the Q_y-band of the former relative to the latter (Figure 5-35); this level of distortion is similar to that exhibited by the special-pair cofactors of *B. viridis*. On the other hand, in a recent study by Olea *et al.*[121] mutations to the haem binding pocket of H-NOX, decreasing the total degree of distortion from 2.180 Å in the wild-type to 0.870 Å in the mutant affected a decrease of the haem midpoint potential by 171 mV. Note that this reversal in the trend between

macrocycle distortion and redox potential is of course due the fact that the haem oxidation is a metal centred process. The relevance of this result to the BChl redox processes (macrocycle centred) is only apparent when one considers that reduction potentials are related to electron density; applying a pseudo-conservation principle coupled with the knowledge that the distortion of the wild-type H-NOX cofactor was confirmed to affect a measured decrease in electron density at the iron-centre, renders the proposal acceptable that potentials of macrocycle and metal-centred redox process, when affected predominantly by conformational effects, are inversely proportional.

As another example, the decreased singlet oxygen yields affected by the *sad* distortion (Figure 5-36) as a consequence of the distortions' effects on numerous photo-physical parameters lends weight to the idea of a photo-protective role of the distortion of the B_B cofactor in *R. sphaeroides*.

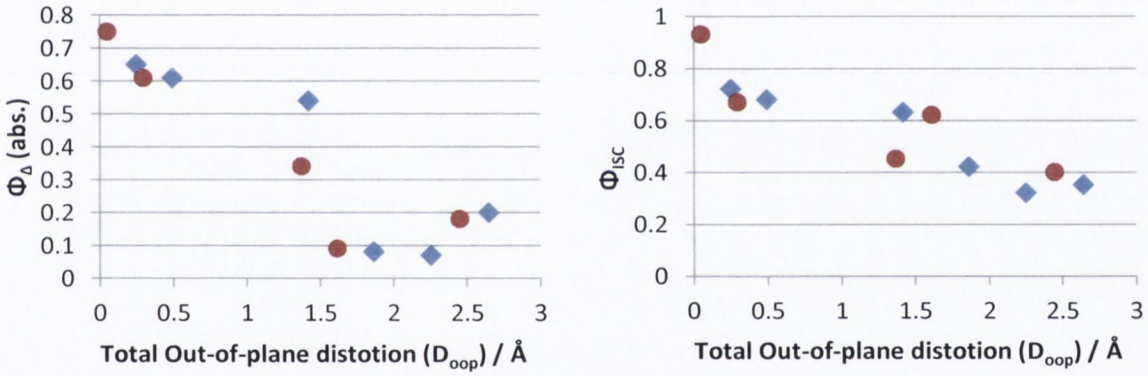


Figure 5-36: Illustrations of the association between the total out-of-plane distortion (D_{oop}) and the singlet-oxygen (left) and the intersystem crossing quantum yields of a series of conformationally designed free-base- (blue diamonds) and Zn(II) (red circles) saddled porphyrins; data from Röder *et al.*[40]

5.6 Structural Comparison to PSI and II

As there have been only a few reported crystal structures of PSI (Table 5-18), a large-scale statistical analysis analogous to that performed for the bacterial RCs is unlikely to prove fruitful. Two of the eight available structures are of such low resolution as to be immediately excluded from any attempt to glean insight into the Chl macrocycle conformations (PDB IDs: 3PCQ and 4FE1), * whilst five others with resolutions between 3.3 – 3.5 Å are unlikely to be particularly useful in this regard. However, the structure of PSI from the cyanobacteria *Synechococcus elongatus* may prove valuable for a tentative

* As a note of interest, these structures are at the forefront of methods development in macromolecular crystallography. Specifically, one reports the use of a free electron laser in the diffraction of nanocrystals[361] and the other is concerned with the analysis of very low resolution data.[362]

discussion as its resolution of 2.5 Å is similar to the mean resolution of the structures in the RBC4 subset (2.53 Å; see Table 5-9, p. 107).

In contrast to the above, there are a total of 16 structure determinations available for PSII (Table 5-19). However, although this figure is one greater than for the representatives of the *B. viridis* RC, excluding a few specific cases these structures are of significantly lower resolution with nine in the range 3 – 4 Å and a further three at 5.7 – 6.7 Å. Amongst the four that remain there are two particularly interesting candidates, namely structures 3ARC and 4IL6 at 1.9 and 2.1 Å resolution, respectively. The remaining two PDB entries are actually from a single crystal and each represents an individual monomer from the dimeric form of PSII; these 2.9 Å structures were processed but unsurprisingly they did not contribute any significant detail.

Table 5-18: Structure determinations of PSI presently deposited in the PDB.

PDB ID	Space Group	Res.	Year	Source	Structure title	Ref.
1JB0	$P6_3$	2.50	2001	*Synechococcus elongatus*	Crystal Structure of Photosystem I: a Photosynthetic Reaction Centre and Core Antenna System from Cyanobacteria	106
2O01	$P2_1$	3.40	2007	*Arabidopsis thaliana*	The Structure of a plant photosystem I supercomplex at 3.4 Angstrom resolution	363
2WSC	$P2_1$	3.30	2010	*Arabidopsis thaliana*	Improved Model of Plant Photosystem I	364
2WSE	$P2_1$	3.49	2010			364
2WSF	$P2_1$	3.48	2010			364
3LW5	$P2_1$	3.30	2010	*Pisum sativum*		364
3PCQ	$P6_3$	8.98	2011	*Thermosynechococcus elongatus*	Femtosecond X-ray protein Nanocrystallography	361
4FE1	$P6_3$	4.92	2012	*Thermosynechococcus elongatus*	Improving the Accuracy of Macromolecular Structure Refinement at 7 Å Resolution	362

Table 5-19: Structure determinations of PSII presently deposited in the PDB.

PDB ID	Space Group	Res.	Year	Source	Structure title	Ref.
1IZL	$P2_12_12_1$	3.70	2003	*Thermosynechococcus vulcanus*	Crystal Structure of Photosystem II	365
1S5L	$P2_12_12_1$	3.50	2004	*Thermosynechococcus elongatus*	Architecture of the photosynthetic oxygen evolving centre	366
1W5C	$P2_12_12_1$	3.20	2004	*Thermosynechococcus elongatus*	Photosystem II from Thermosynechococcus elongatus	367
2AXT	$P2_12_12_1$	3.00	2005	*Thermosynechococcus elongatus*	Crystal Structure of Photosystem II from Thermosynechococcus elongatus	368
3A0B	$P2_12_12_1$	3.70	2009	*Thermosynechococcus vulcanus*	Crystal structure of Br-substituted Photosystem II complex	369
3A0H	$P2_12_12_1$	4.00	2009	*Thermosynechococcus vulcanus*	Crystal structure of I-substituted Photosystem II complex	369
3ARC	$P2_12_12_1$	1.90	2011	*Thermosynechococcus vulcanus*	Crystal structure of oxygen-evolving Photosystem II at 1.9 angstrom resolution	220
3BZ1	$P2_12_12_1$	2.90	2009	*Thermosynechococcus elongatus*	Crystal Structure of cyanobacterial Photosystem II (part 1 of 2). This file contains first monomer of PSII dimer	370
3BZ2	$P2_12_12_1$	2.90	2009	*Thermosynechococcus elongatus*	Crystal Structure of cyanobacterial Photosystem II (part 2 of 2). This file contains second monomer of PSII dimer	370
3KZI	$C222_1$	3.60	2010	*Thermosynechococcus elongatus*	Crystal Structure of Monomeric Form of Cyanobacterial Photosystem II	371
3PRQ	$P2_12_12_1$	3.20		*Thermosynechococcus elongatus*	Crystal structure of cyanobacterial Photosystem II in complex with terbutryn (PART 1 OF 2). This file contains first monomer of PSII dimer	372
3PRR	$P2_12_12_1$	3.20		*Thermosynechococcus elongatus*	Crystal structure of cyanobacterial Photosystem II in complex with terbutryn (PART 2 OF 2). This file contains second monomer of PSII dimer	372
4FBY	$P2_12_12_1$	6.56	2012	*Thermosynechococcus elongatus*	Fs X-ray diffraction of Photosystem II	373
4IL6	$P2_12_12_1$	2.10	2013	*Thermosynechococcus vulcanus*	Structure of Sr-substituted photosystem II	241
4IXQ	$P2_12_12_1$	5.70	2013	*Thermosynechococcus elongatus*	RT fs X-ray diffraction of Photosystem II, dark state	374
4IXR	$P2_12_12_1$	5.90	2013	*Thermosynechococcus elongatus*	RT fs X-ray diffraction of Photosystem II, first illuminated state	374

5.6.1 Photosystem I

The NSDs of the cofactors from the highest-resolution crystal structure of PSI[106] (PDB ID: 1JB0) are shown below (Figure 5-37). Although it is not possible to construct an estimate

of uncertainty based on empirical data for a single structure, the ESDs given by the authors that were estimated using the Luzzati method (0.29 Å) represent an informative alternative and so are included on the NSD plots. With respect to their relation to the OOP normal-deformations, the 3-dimensional coordinate error is reduced by a factor of 0.5 due to the lower dimensionality (1D) of this component of the macrocycle conformation.[32] In light of these error estimates, there are almost no significant conformational differences between C_2-related pairs of cofactors aside from the greater *sad* deformation of P_B (Chl a) compared to P_A (Chl a'). This is consistent with the observation that only the special-pair binding-site shows significant differences between the PsaA and PsaB subunits.[106] However, it should be noted that the small deviations between other pseudo-C_2 related pairs indicate that all cofactors were refined individually.

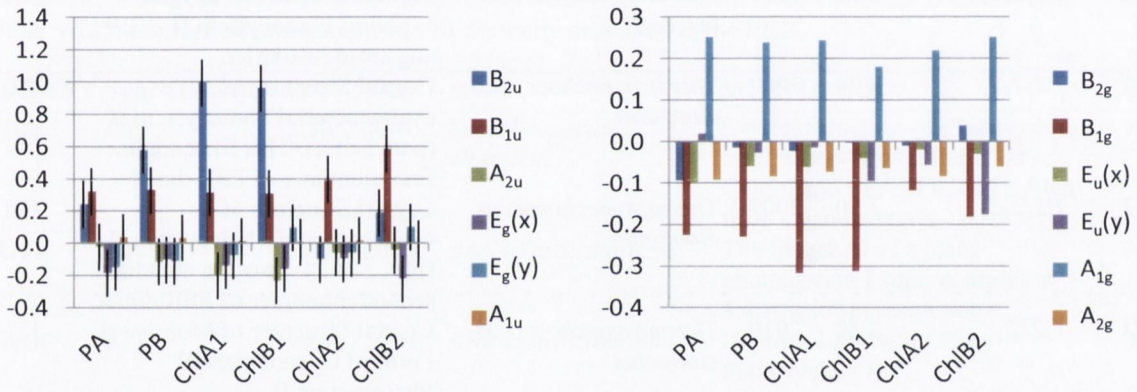


Figure 5-37: Out-of-plane min. basis NSDs (left) of the ETC cofactors in PSI (PDB ID: 1JB0); error bars show ± one half of the Luzzati error; δ_{oop} = 0.0203, 0.0247, 0.0422, 0.0389, 0.0120 and 0.0149 (left-to-right). In-plane minimum basis NSDs (right); error bars are not shown as the required ± π/4 times Luzzati error (0.23 Å) overwhelms the plot; δ_{ip} = 0.0484, 0.0443, 0.0548, 0.0571, 0.0492 and 0.0513 (left-to-right).

In general contrast to the bacterial RCs (and also PSII, described later), the cofactors with the largest degree of nonplanarity in the PSI ETC are the first accessories, Chl_{A1} and Chl_{B1}. Quite uniquely, both cofactors possess a very large *sad* distortion, which is one of the largest single-mode deformations of all the ETC (B)Chls and approached only by the ~1 Å *ruf* of the *B. viridis* D_M. This is complemented with a comparatively minor *ruf* component that is identical to that of the special-pair Chls and is also comparable to B_B, D_L or D_M from *R. sphaeroides* (*i.e.* 0.4, 0.2 or −0.4 Å), within the indicated margin of error. These Chls also exhibit a small, yet likely significant degree of the *dom* mode in the direction of their HIS ligands, whilst in terms of their in-plane conformations they have a larger *N-str* (B_{1g}) than the other PSI ETC Chls. Additionally, Chl_{B1} exhibits a moderate core contraction relative to the rest as indicated by the decreased *bre* deformation (A_{1g}) and, although technically not significant with respect to the estimated error for the IP normal-

deformations (0.23 Å), is worth mentioning because of the relative consistency of the other Chls' *bre* distortions as well as this modes usual association with increased nonplanarity.

As noted above, the P_A and P_B of the special-pair heterodimer are the only C_2-related Chls of PSI that exhibit strongly supported conformational asymmetry in the form of a dominant *sad* conformation for P_B. This is somewhat reminiscent of the situation in both the RCs where the M-branch BChl was generally more nonplanar that the other. Aside from this feature, both Chls are characterised by the same *ruf* distortion as the first accessories. Finally, the second accessory Chls may possess a slightly greater *ruf* conformation and very nearly display significant differences across the C_2-axis, with Chl_{B2} exhibiting larger *ruf* and *sad* deformations.

Taken together the analysis suggests that two PsaB Chls may be only slightly more distorted than their PsaA counterparts, if at all, which is in stark contrast to the purple bacterial RCs where the M-branch BChls were consistently more nonplanar. However, this result could be significant re-calling that in PSI, both branches participate in electron transfer.[375] It is also quite interesting to note the greater distortion of the accessory chlorophylls of this RC compared to both RCs and PSII in the light of the fact that these Chls have been implicated as the primary donors.[244]

5.6.2 Photosystem II

Saito *et al.*[221] have already performed an NSD analysis for the Chls in the 3ARC crystal structure using a modified NSD in which the basis was composed of normal-modes calculated for Mg-chlorin at the B3LYP/LACVP* level of theory. However, the NSD results that they obtained using this approach are practically identical to those shown below, except that the degree of *dom* in their analysis is larger as they include the Mg in their procedure (see Chapter 1.5.5).

The normal-deformations for the cofactors obtained from the two highest resolution structures of PSII are rather similar (Figure 5-38 and Figure 5-39). However, there are small differences between the resolved conformations from each that indicates that atomic configurations in structure 4IL6 were allowed to deviate from the starting model 3ARC. The magnitude of these differences may be judged from the observed standard deviations of the NSD parameters between the two structures. The OOP standard deviations are in the range of 0.004 – 0.085 Å (mean = 0.034 Å) and compare reasonably with an estimate of the theoretical e.s.u. for OOP modes of 0.045 – 0.073 Å for these structures, obtained by halving the e.s.u given by the authors. However, the standard deviations for the IP modes

of 0.007 – 0.074 Å (mean = 0.034 Å; same as above) are quite a bit lower than the theoretical estimate of 0.071 – 0.121 Å (*i.e.* π / 4 times e.s.u). This is not an unusual occurrence however, and so may be concluded that there is somewhat more confidence in these results as compared to the situation in PSI, but still they are discussed only tentatively and in lieu of further experimental determinations.

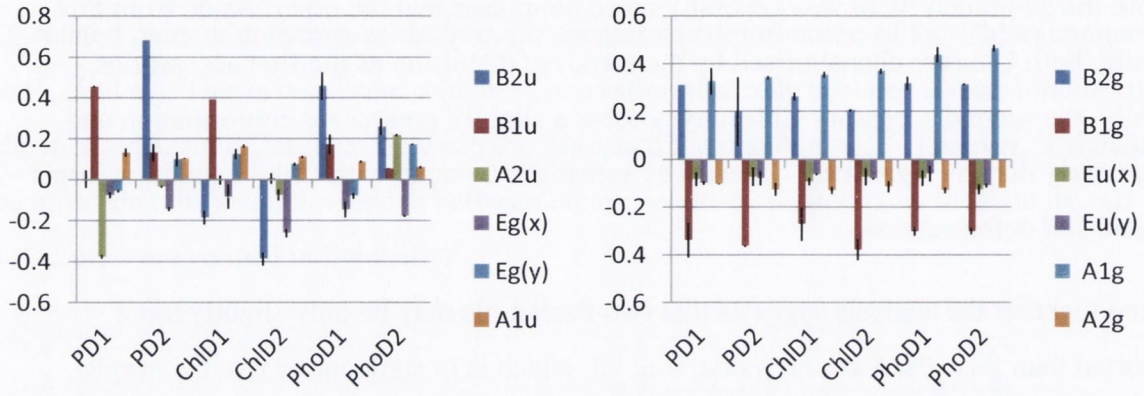


Figure 5-38: Mean out-of- (left) and in-plane (right) minimum basis NSDs of the ETC cofactors in PSII (PDB ID: 3ARC); error bars show ± s.d. between the two corresponding cofactors in the asymmetric unit. Mean δ_{oop} = 0.0410, 0.0300, 0.0364, 0.0332, 0.0294 and 0.0354 (left-to-right); mean δ_{ip} = 0.0648, 0.0644, 0.0637, 0.0670, 0.0679 and 0.0698 (left-to-right); e.s.u. based on R_{free} = 0.090 Å.

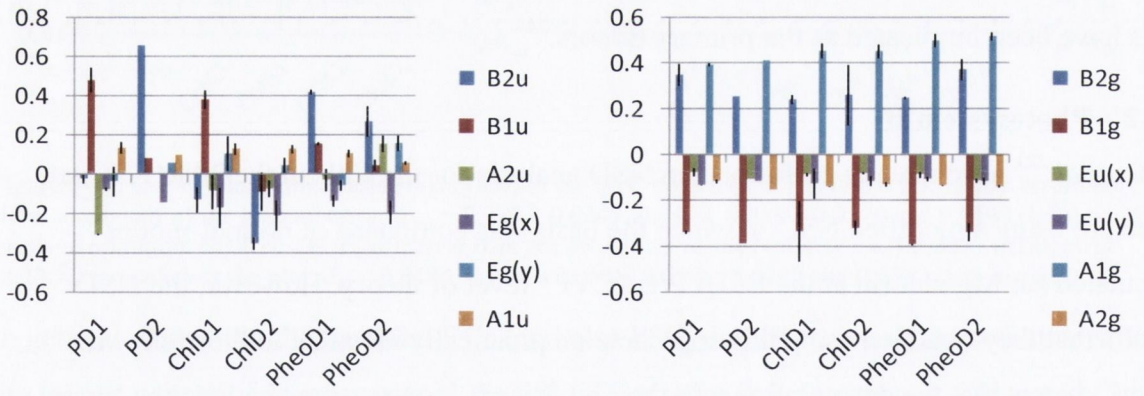


Figure 5-39: Mean out-of- (left) and in-plane (right) minimum basis NSDs of the ETC cofactors in PSII (PDB ID: 4IL6); error bars show ± s.d. between the two corresponding cofactors in the asymmetric unit. Mean δ_{oop} = 0.0409, 0.0308, 0.0328, 0.0439, 0.0249 and 0.0341 (left-to-right); mean δ_{ip} = 0.0653, 0.0685, 0.0677, 0.0677, 0.0700 and 0.0738 (left-to-right); e.s.u. based on R_{free} = 0.154 Å.

In contrast to PSI, the high resolution structure of PSII[220] indicates considerable asymmetry of the cofactor conformations across the C_2-axis. P_{D2} possesses the most distortion with a well-defined *sad* conformation.* A striking feature is that P_{D1} displays a similar degree of the *dom* distortion compared to D_L of the *R. sphaeroides* RC, although

*Note also that the previous 2.9 Å structure (PDB ID: 3BZ1)[370] also recovered the large saddling of P_{D2} as well as some of the doming of P_{D1}; these were the only differences observed in that structure.

P_{D1} has a greater contribution from the *ruf* mode that this BChl. The C_2-asymmetry continues with Chl_{D1} displaying a *ruf* conformation quite dissimilar to Chl_{D2}'s mixed *sad* and *wav* deformations. The Pheos are also suggested to be distinguishable with PheoD1 possessing substantial *sad* and some *ruf* contributions to its conformations whilst PheoD2 has about half the degree of *sad* and also, similar contributions from *dom* and *wav*. In common with the bacterial RCs, PSII does display considerable branch asymmetry with respect to electron transfer. It is perhaps significant that the active branch cofactors exhibit greater *ruf* distortion than their counterparts.

Table 5-20: Std. deviations of the OOP min. basis NSD parameters of the assessed PSII structures (Å; n = 4).

	D_{oop}	δ_{oop}	B_{2u}	B_{1u}	A_{2u}	$E_g(x)$	$E_g(y)$	A_{1u}
P_{D1}	0.027	0.002	0.032	0.039	0.041	0.009	0.045	0.021
P_{D2}	0.021	0.001	0.012	0.038	0.020	0.005	0.032	0.004
Chl_{D1}	0.022	0.004	0.051	0.028	0.073	0.085	0.054	0.027
Chl_{D2}	0.019	0.010	0.032	0.081	0.028	0.053	0.029	0.015
$Pheo_{D1}$	0.042	0.002	0.045	0.030	0.040	0.027	0.020	0.014
$Pheo_{D2}$	0.019	0.001	0.051	0.017	0.045	0.036	0.024	0.007

Table 5-21: Std. deviations of the IP min. basis NSD parameters of the assessed PSII structures (Å; n = 4).

	D_{ip}	δ_{ip}	B_{2g}	B_{1g}	$E_u(x)$	$E_u(y)$	A_{1g}	A_{2g}
P_{D1}	0.038	0.001	0.034	0.043	0.019	0.021	0.050	0.009
P_{D2}	0.056	0.002	0.088	0.015	0.030	0.028	0.040	0.023
Chl_{D1}	0.085	0.003	0.020	0.094	0.010	0.053	0.061	0.010
Chl_{D2}	0.041	0.002	0.082	0.045	0.022	0.025	0.051	0.018
$Pheo_{D1}$	0.048	0.001	0.044	0.053	0.020	0.034	0.043	0.007
$Pheo_{D2}$	0.051	0.002	0.042	0.032	0.015	0.011	0.033	0.011

5.7 Summary: Symmetry Breaking, Species Adaption and RC Efficiency

To summarise, agglomerative hierarchical clustering of the BChl conformations revealed that their conformations were associated with their positions in the ETC. However, closer inspection indicated secondary features that were related to entire crystal structures and so further analyses were performed to assess the extent of any systematic error present in the structural data. In the *R. sphaeroides* data, the inter-crystal structure Euclidean dissimilarities highlighted that a number of structures did not exhibit *any* degree of experimental variation and these were therefore removed them further consideration (RBC1); it is likely that they originate from the practice of rigid-body refinement. Next, a two-factor exploratory factor analysis revealed that structures exhibited either: planar cofactors (RBC2), functionally distinct nonplanar cofactors (RBC4) or functionally distinct

nonplanar cofactors with unusual in-plane contractions (RBC3). Amongst other evidence, a comparison to high-resolution chlorin analogues from the Cambridge Structural Database[157] allowed the conclusion that the RBC4 set of crystal structures is likely to provide the most reliable cofactor conformations. The NSD coefficients of these structures' cofactors show clearly the distinct conformations present at each position of the ETC. The *B. viridis* RC structural data were subjected to the same analysis pipeline and turned out to be considerably simpler such that only one crystal structure was excluded from the aggregation step. Like *R.* sphaeroides, each BChl possessed a unique conformation except that the accessory BChls in this species were qualitatively indistinguishable.

The BChl conformations, when interpreted in terms of the known effects of macrocycle distortion, appear to tie in with the observed physical properties of electron transfer in photosynthetic systems such as the species variant extent of the bathochromic shift of the special-pair, the preference for L-branch ET or the photo-protective action of B_B in *R. sphaeroides* amongst others and therefore provide a new chemical rationale for understanding these processes; here noting that an energetic consideration of the distortions indicates that the conformational effect are likely on a par with both H-bonding and electrostatic effects. It is thus valid to suppose that the conformation contributes to some of these observables, possibly in a deliberate rather than coincidental way; that nature has not only selected the primary structure of proteins to affect the desired protein conformations and cofactor spatial arrangements but also to induce particular cofactor conformations. The concept of conformational control of tetrapyrrole cofactors can clearly be used to address questions of functional significance in the bacterial photosynthetic reaction centre in a statistically reliable manner. This indicates the general validity of this concept and indicates that it can be extended to haems and other systems and serve as a model for structural cofactor modulation in general. Lastly, the bacterial RC BChls were compared to their Chl a counterparts in PSI and II by consideration of the highest resolution crystal structures available for these proteins. Whilst without a thorough analysis of these data it is inappropriate to draw definite conclusions, it appears that some conformational features are conserved between the PSII and *R. sphaeroides* RC proteins, such as the A_{2u} distortion of D_L/P_{DL}, there is also considerable divergence. The most striking difference is that the Chl accessories in PSI are the most distorted cofactors in this RC, which may contribute to their capability to act as the primary electron donors.

6 Conformational Modulation of Haem Redox Potentials in the *B. viridis* Reaction Centre Cytochrome Subunit

6.1 Quantifying Conformational Control

In the previous chapter it was demonstrated that the conformational variation of the bacteriochlorophylls (BChls) in the reaction centres (RCs) of purple bacteria provide an additional mechanism by which both individual BChl and inter-species physicochemical differences may be explained.[60] This result was possible due to the development of a systematic method that allowed an unbiased critical analysis of the reliability of entire RC crystal structures, and sets thereof, with respect to their appropriateness for the detailed comparison of the cofactor conformations displayed at different sites. Here, this work is taken further by assessing the role of conformational control in modulating the reduction potentials of the haems in the reaction centre tetrahaem cytochrome subunit (RC-cyt) of *Blastochloris viridis*. This also required a statistical analysis of the haem's conformations afforded by the available crystal structure data although in this case, the effect on reduction potentials was not inferred; instead it was explicitly calculated using an experimentally calibrated computational procedure.

As has been discussed in Chapter 1, the physicochemical consequences of nonplanarity were deduced primarily from model compound studies which correlated increasing macrocycle distortion with bathochromic shifts of the UV/Vis Q-bands, easier oxidation and hindered reduction (for macrocycle centred redox processes) and decreasing S_1-lifetimes.[38-40] In contrast however, computational investigations of macrocycle nonplanarity experienced a period of inconsistency, sparking lively debate and giving rise to such controversies as the short-lived concept of In-Plane Nuclear Rearrangement.[36, 116-118, 197] This debate continued until specific flaws were identified in the structural models used and it was once again widely accepted that it is the distortions of the porphyrinoid macrocycle that give rise to these effects.[36] Considering this historical development, an experimentally verifiable computational method for evaluating the physicochemical effects of macrocycle structural perturbation is a desirable goal which has not yet been entirely realised.

In this respect, the approach was enabled by the recent work of Olea *et al.* where conformational control was used to impart redox modification through mutagenesis of the haem-NO and $-O_2$ binding domain (H-NOX) from *Thermoanaerobacter tengcongensis*.[121]

They reported that the reduction potential of the H-NOX complex could be reduced *via* site-specific mutations that allowed the highly distorted haem cofactor to relax into a less nonplanar conformation. They also found that the decreased potentials were due to a measurable decrease of electron density at the haem-Fe as indicated by UV/Vis spectroscopy and the pK_as of the bound aqua ligands. Since this study controlled for non-conformational influences by selecting residue substitutions that changed neither the H-bonding of the cofactor nor the local dielectric environment, it presented an ideal reference from which to establish a quantitative relationship between the macrocycle conformation and distortion induced potential shifts through calculable quantum mechanical properties of the isolated haems.

This relationship was then used to assess the conformational contribution to the redox potential variation amongst the individual haems of the RC-cyt. This subunit is tethered to the reaction centre in the periplasmic space above the membrane and serves the purpose of re-reducing the oxidized special-pair of the electron transfer chain (Figure 6-1).[21, 22] The four haems are grouped into pairs of low- and high-potential cofactors and these pairs are arranged such that a chain of alternating low/high redox potentials is created through to the special-pair (*i.e.*, low, high, low, high, SP; Figure 6-1).[235, 236] In this order, the individual haems will be referred to here as H1, H2, H4 and H3, respectively (in order of connection to the protein backbone). As was highlighted earlier, *ab initio* electrostatic calculations of the haems' *in situ* potentials have achieved accuracy to within ±25 mV incorporating many influences of the binding-site, but here it is also noted that only minor consideration was given to the haems' varying conformations.[99]

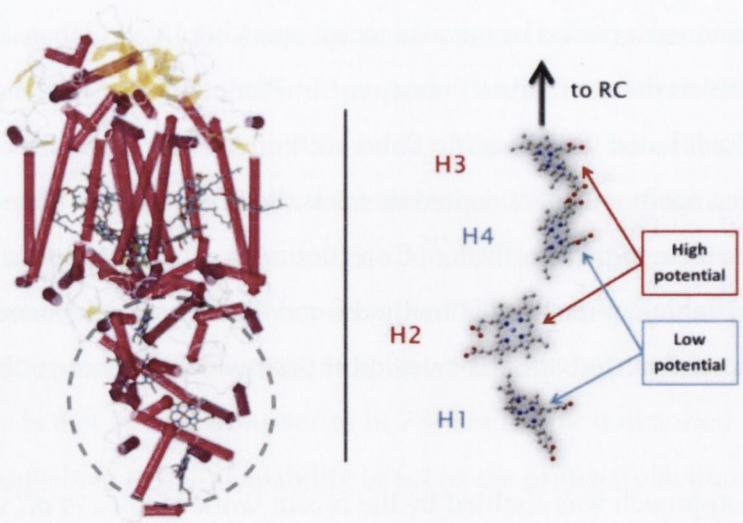


Figure 6-1: Illustration of the tetrahaem cytochrome subunit in the reaction centre of *B. viridis*; image adapted from the coordinates of PDB ID: 1PRC.[237]

Since there are 15 crystal structures available from the PDB of the *B. viridis* reaction centre, the question arises as to which one to select to perform the final part of the analysis? The results from the previous analysis of the RC structural data suggests that standard quality metrics (*e.g.*, resolutions, *R*-factors) for these large biomolecular crystal structures are not necessarily the decisive factor when one is interested solely in the detailed conformation of the cofactors.[60] Instead therefore, a statistical analysis of the resolved conformations from all of the crystal structures was performed and in so doing uncovered latent experimental biases, which resulted in three sets of mean cofactor conformations (*i.e.*, 12 haem structures) for the calculations.

To summarize, the focus of this chapter is the proposal of a facile technique to predict and delineate the redox influence of macrocycle distortion in haems and its application to further elucidate the role conformational control plays in the tetrahaem cytochrome subunit of the *B. viridis* reaction centre.

6.2 Experimental Details

6.2.1 Raw Data and NSD

The crystallographic atomic coordinates of the relevant H-NOX and RC-cyt haems were obtained from the PDB Ligand Expo[280] and analysed using the procedure of normal-coordinate structure decomposition (NSD)[32] in which the macrocycle conformation is described by a linear decomposition into a basis composed of the macrocycle's internal normal-modes of vibration. This basis may consist of either the full set of 66 (3N – 6) *normal-coordinates* or any given subset. In their original formulation, Jentzen *et al.* termed the full set the *complete basis* and suggested two other useful bases: the *minimum* and *extended basis*.[32] These reduced bases are composed of either the lowest-energy (min.) or the lowest- plus next-to-lowest-energy (ext.) normal-modes of each symmetry type. The utility of these reduced bases is that they yield a simplified analysis whilst their appropriateness is both allowed theoretically by the fact that the largest contributions to the observed conformations are expected to be distortions of relatively low energy and assessed practically by considering the root mean square deviation between the *simulated structure* (*i.e.* the conformation represented by the reduced basis) and the actual structure. Thus, NSD yields the magnitudes of each of the component normal-modes present in the chosen basis that contribute to the analysed structure; and provides a quantitative interpretation of the conformation. These data were used to determine the estimates of the RC-cyt haem conformations *via* a statistical analysis of the available crystal structures in a

manor described in detail previously[60] (see below) and to construct the various cofactor models used for the calculations.

6.2.2 Theoretical Method

The model structures used as inputs for the DFT calculations were selected so that the haem cofactor's structural parameters (*i.e.* substituent orientation, Fe-position, macrocycle conformation [including NSD reduced bases]) would be treated individually. The "diacid" model structures were formed by adding hydrogen atoms to the crystallographically resolved cofactor coordinates and adding the propionic acid proton to the oxygen with the longest C-O bond. The first Fe-porphin (FeP) model was also obtained directly from the crystal structure coordinates by replacing the macrocycle side-chains with hydrogen atoms. The "FeP-min., -ext. and -comp." models were generated by projecting the relevant weighted normal-deformations (minimum, extended or complete, respectively) back onto the reference macrocycle in 3D-Cartesian space and by positioning the Fe-atom at the coordinate-system's origin (since the reference macrocycle is centrosymmetric about this point) thereby providing a reduced model that accounts only for the macrocycle normal-mode distortions alone, eliminating any possible macrocycle conformation/Fe-position interaction.

The quantum calculations were performed with the Gaussian 09[294] package using the facilities of the Trinity Centre for High-Performance Computing. The Fe-atom Mulliken[376] and minimal basis set Mulliken[377, 378] atomic charges ($Q_{Fe\text{-}Mulk}$ and $Q_{Fe\text{-}MBS}$, respectively) of the haem models were obtained from single-point energy calculations using the B3LYP functional[287] and various basis sets. Natural atomic charges and orbital occupancies were obtained from Natural Population Analysis[379] using the NBO 3.1 implementation in Gaussian 09. In particular, each calculation was performed using the 3-21g,[288, 289] 6-31g*, 6-311g**[290-293] and 6-311+g**[290-293] basis sets in order to roughly ascertain the parameters' basis set dependency and so to find an appropriate balance between accuracy and expense. H-atom optimizations were not performed (to save on computational expense) as earlier trial computations (Appendix A; Figure B-1) indicated that they had no bearing upon the resultant Mulliken charges compared to the default hydrogen atom addition parameters used by the GaussView[284] software for SP^2-carbons (*i.e.* d(C-H) = 1.07 Å and ∠(R1-C-H) = ∠(R2-C-H) = [360° - ∢(R1-C-R2)]/2 such that the added hydrogen is coplanar with the sp^2 centre and its other directly bonded atoms).

6.2.3 Statistical Analysis of Cofactor Conformations

Analysis of the conformations of the RC-cyt haems began with agglomerative hierarchical clustering (AHC) of each cofactor's minimum basis NSD in order to discover whether or not the crystal structure data were consistent with the hypothesis of distinct cofactor conformations. Thus, the AHC was performed on data matrices containing observations of the NSD minimum basis for each cofactor within the R statistical environment[352] using the Euclidean distance measure to build the dissimilarity matrix and Ward's method for the agglomeration.

The method of AHC was ideal for this purpose because it provides information relating to the similarity of observations based on any number of numerical variables (Chapter 3.3). To recapitulate, observations are algorithmically grouped together (clustered) based on their mutual (dis)similarity, calculated using a suitable distance metric; here, the Euclidean distance ($d_{i,j} = \sqrt{[\sum(x_i - x_j)^2]}$). The decision as to how many clusters to select, between the extremes of one cluster containing all observations and as many clusters as there are observations, was achieved by consideration of the cluster dendrogram, which shows the interrelationships between possible clusters. For the agglomeration step, Ward's method was used, which can be considered to agglomerate with respect to reducing the information loss at each step. Therefore in chemical terms, since the NSD deformations of the cofactors were clustered, groups of cofactors were obtained that had similar resolved conformations.

In order to identify and isolate any systematic discrepancies between sets of structure determinations, the collected NSD data were then treated so that each PDB entry was considered as an individual observation and the minimum basis NSD deformations of its cofactors the observables. In detail, the data was formed into an $m \times n$ matrix where m is equal to the number of crystal structures included and n, the number of variables, equals the number of cofactors included (*i.e.* four haems) multiplied by the number of NSD basis parameters used (*e.g.*, B_{2u} of H1 is a distinct variable from B_{2u} of H2 and there is no categorical variable for cofactor identity). In the analyses of the experimental effects that followed, the 12 normal-coordinates of the minimum basis were used for each cofactor leading to a dataset of 15 observations of 48 variables. Application of both AHC and principal components analysis (PCA) to this data matrix allowed the crystal structures to be classified into groups exhibiting similar systematic errors (leaving only random fluctuations within each group) from which the mean conformations could be derived in the usual way.

Note that PCA, which here was used to complement AHC, is used to reduce the effective dimensionality of the data's variable space by forming linear combinations of the original variables termed the principal components (PCs) which account for as much variance, with as few PCs, as possible (*cf.* AHC which combines *observations*). Thus, here the PCs represent distortions along vectors made up of combinations of the normal-modes from the NSD analysis of the *four* cofactors in each structure and hence illustrate the influence of the structure determination *classification* (from the AHC) upon the resolved conformations.

6.3 Calibration of a Computational Procedure for Evaluating Haem Conformational Control

6.3.1 First Attempts

Initially, a number of computable parameters were tested in order to identify one which would correlate with the relative reduction potential shifts due to conformational control in the H-NOX WT, P115A, I5L and P115A/I5L complexes. The first attempts involved calculating standard estimates of the ionization potentials of the haems in the form of HOMO (Koopmans' theorem) and ΔDFT ($E_{cation} - E_{neutral}$) energies. Unfortunately, in these experiments no simple correlation with the observed trend in reduction potentials was found. Next, re-calling the electron deficiency that had been observed by Olea *et al.*,[121] it was thought that this may be manifested in the Fe-atom partial atomic charges derived from atomic population analysis.

Thus, first the Fe-atom Mulliken atomic charges ($Q_{Fe\text{-}Mulk}$) were tested against the experimentally determined potentials (E_ms) of the H-NOX mutants where they provided, under certain conditions, strong, positive correlations. In particular, it was found that $Q_{Fe\text{-}Mulk}$ from the singlet state B3LYP/3-21g wave functions of the diacid, FeP and FeP-comp. models correlated well with the H-NOX potentials but that the corresponding triplet state calculations resulted in $Q_{Fe\text{-}Mulk}$'s that showed only slight correlation (Figure 6-2). Also, $Q_{Fe\text{-}Mulk}$s obtained from any of the other basis-sets tested, in either state, provided poor correlations that often showed non-monotonic relations. Even though the strong correlation of the singlet state $Q_{Fe\text{-}Mulk}$ with the H-NOX E_ms appeared to satisfy the requirements, the lack of a good correlation with the equivalent triplet state partial-charges was worrisome since this state represented the ground state of the models (Appendix B; Table B-8 to Table B-12) in accord with previous work on unligated Fe-porphins.[380] Furthermore, the failure to reproduce the correlation with the larger, more appropriate basis

sets could lead to doubts regarding the causal interpretation of the result. These problems necessitated further analysis and given the well-known numerical instability of Mulliken population analysis with respect to basis set size,[381, 382] alternative approaches were investigated.

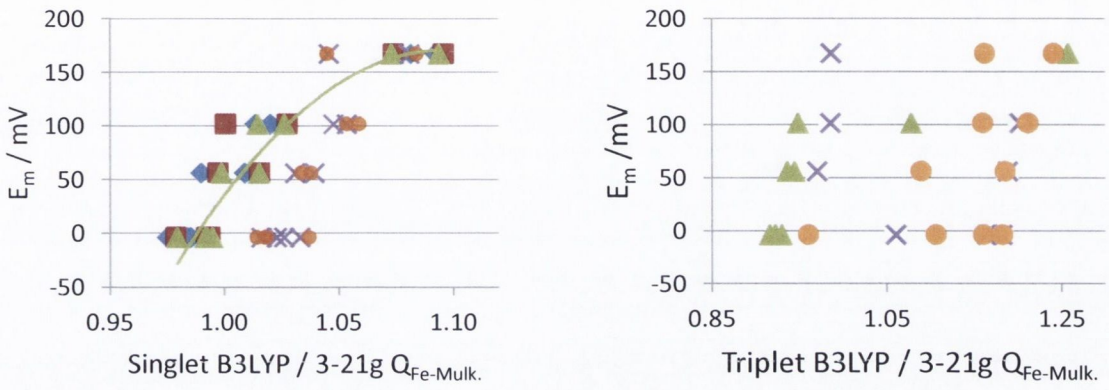


Figure 6-2: Plot of the experimental $Fe^{2+/3+}$ midpoint potentials of the H-NOX variants *vs.* the calculated partial atomic charges of the haem-Fe from the diacid (blue), FeP (red), FeP-comp. (green), FeP-ext. (crosses) and FeP-min. (orange) models, in the singlet- (left) triplet state (right). The line shows the quadratic least-squares fit to the FeP-comp. model ($R^2 = 0.92$)

6.3.2 Minimum Basis Set Mulliken Atomic Charges

The first of these was chosen to circumvent the basis set dependency of the Mulliken analysis directly, whilst retaining the remainder of the formalism. The method, known as minimum basis set Mulliken analysis (MBS) proceeds by projecting the MO-LCAO coefficients of the original basis onto those of a minimal basis and only thereafter performing the standard Mulliken analysis on these new MOs.[377, 378] As hoped, the Fe-atom partial atomic charges obtained in this way ($Q_{Fe\text{-}MBS}$) provided remarkably improved correlations with the triplet state wave functions found using basis sets larger than 3-21g; confirming that the failures there were due to the basis sensitivity of Mulliken analysis and suggesting that the 3-21g basis was insufficient to reasonably describe the triplet structure (Figure 6-3).

However, a few anomalous results were present. Specifically, for the triplet calculations, irregular drops of either one of the 1U56 cofactors occurred using the 6-31g* basis with the diacid and FeP-ext. models as well as with the 6-311g** result for the FeP model. Also, non-monotonicity of $Q_{Fe\text{-}MBS}$ from the singlet calculations using the 6-31g* and 6-311g** bases was observed such that the former provided an I5L (3NVR) haem with a partial charge that was practically the same as the I5L/P115A (3NVU) haems, whilst in the latter, $Q_{Fe\text{-}MBS}$ for the I5L/P115A haems were lower than those from P115A (3EEE) (Appendix B; Figure B-2).

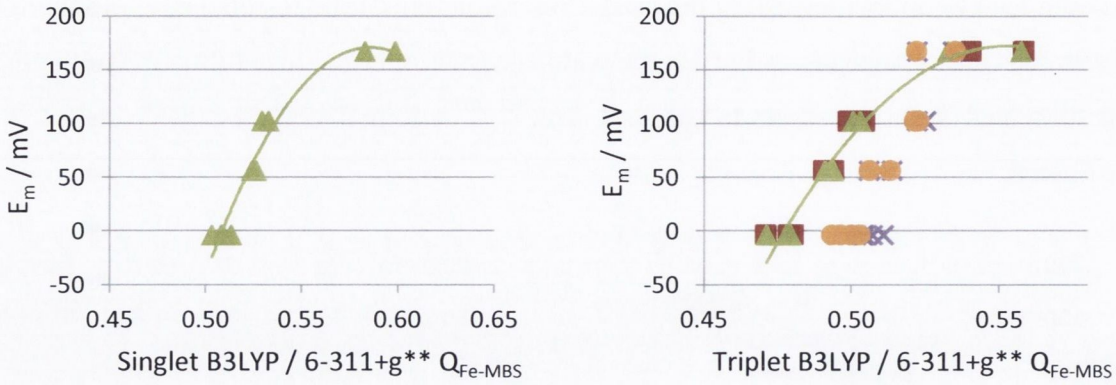


Figure 6-3: Plots of the H-NOX complex midpoint potential *vs.* $Q_{Fe\text{-}MBS}$ obtained from B3LYP/ 6-311+g** singlet (left) and triplet (right) state calculations of the models; haem diacid (blue), FeP (red), FeP-comp. (green), FeP-ext. (crosses) and FeP-min. (orange). The lines show the quadratic least-squares fit to the FeP-comp. model; $R^2 = 0.956$ and 0.967 for singlet and triplet state calculations, respectively.

On the other hand, progressively better correlations were obtained with increasing basis flexibility as may be expected for a truly causal, electronic relationship; gradual improvements observed for the triplet calculations and a discrete jump in agreement for the singlet state model with the 6-311+g** basis so much so that their respective calibration curves are statistically identical. With this in mind, the anomalies cited above should not represent any serious problem for the purpose of calibration since we are concerned primarily with the ground state (triplet) of the FeP-comp. model (as this represents the best way to isolate fully the conformational influence) and are satisfied with the conclusion that the most flexible basis set employed here should be used to effect the most reliable results. Even so, for additional confirmation, an entirely different form of population analysis was assessed, namely the Natural Population Analysis (NPA).[379]

6.3.3 NPA Fe electron populations

Surprisingly, the Fe-atom partial charges obtained *via* NPA ($Q_{Fe\text{-}NPA}$) exhibited an *inverse* correlation with the H-NOX E_ms, this being in contrast to the previous $Q_{Fe\text{-}MBS}$ / E_m trends, which were consistent with the experimental results from the H-NOX mutants (not shown). Closer inspection indicated that this corresponded to an increase of both total valence and Rydberg Fe-atom natural orbital occupancies that overshadowed a decrease in the corresponding core orbital occupancies. Although this has not yet been explained and no other report of such inverse proportionality between these two methods could be found in the literature, the strong empirical relationship, lack of anomalies for the 1U56 cofactors and reduced basis- and spin-state dependencies encouraged pursuit of NPA as a complementary approach. In particular, the Fe-core electron occupancy ($NPA_{Fe\text{-}CORE}$) may be particularly suitable not only because this parameter remained consistent with the

previous findings and expectations, but also because it provided the best correlations (Figure 6-4 and Appendix B; Table B-14).

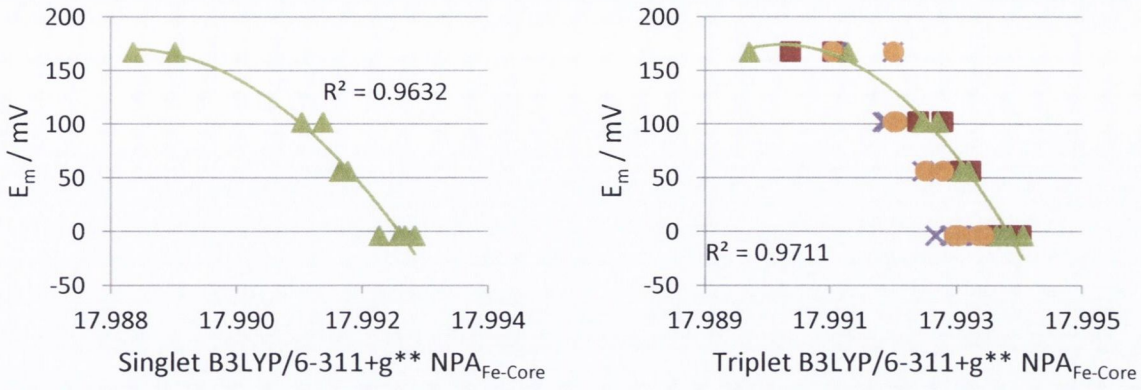


Figure 6-4: Plots of the H-NOX complex midpoint potential vs. NPA Fe core electron count obtained from B3LYP/ 6-311+g** singlet (left) and triplet (right) state calculations of the models; haem diacid (blue), FeP (red), FeP-comp. (green), FeP-ext. (purple) and FeP-min. (orange); R^2 = 0.963 and 0.971 for singlet and triplet state calculations, respectively

6.3.4 Structural Models and Overall Correlations

Turning now to the influence of the structural model used, the haem-diacid,* FeP and FeP-comp. triplet models provided very similar results (Figure 6-5), regardless of basis, emphasized by linear relations with gradients, intercepts and R^2 values within the ranges of 0.99-1.06; (-0.04)-0.00 and 0.98-1.00 for $Q_{Fe\text{-}MBS}$ (excluding the 1U56-B501 charge when outlying) and 0.89-1.07; (-1.22)-1.98 and 0.96-1.00 for $NPA_{Fe\text{-}CORE}$ (all results), respectively. These high values of R^2 confirm beyond doubt that it is purely the induced macrocycle conformation which is the cause of the potential shifts in these H-NOX mutants and that a complex interaction between the macrocycle conformation and its side-chains, Fe-centre, axial ligands and/or protein environment is not involved. Additionally, the same behaviour holds true for all of the structural models in the singlet state using the 3-21g basis, but the full set of calculations for the remaining bases were not performed (NB. only the FeP-comp. structures were assessed).

Additionally, the 2nd order polynomial fits for these models' $Q_{Fe\text{-}MBS}$ and NPA core populations as explanatory variables for the midpoint shifts resulted in correlation coefficients all greater than 0.93 for the former (with the same exclusions outlined above) and 0.95 for the latter implying their suitability for interpolating the intrinsic potential shifts brought about by conformational control here and in other systems.

---

* Not obtained for B3LYP/6-311+g**

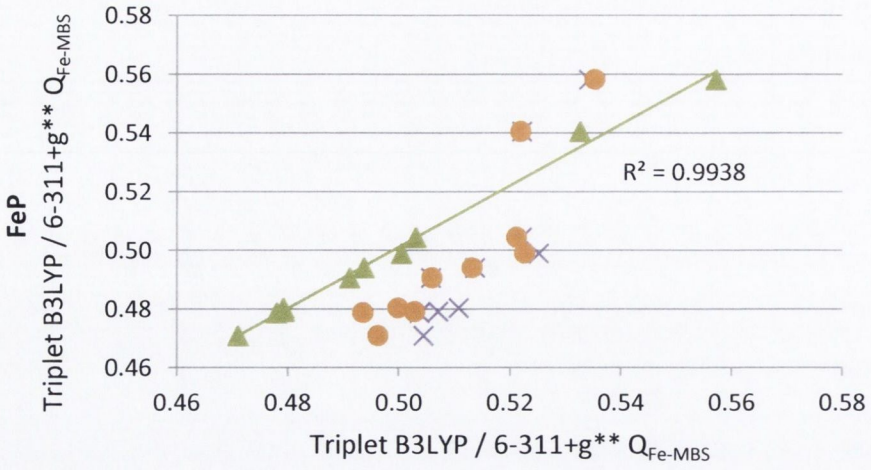


Figure 6-5: Correlation between Q_{Fe-MBS} of the FeP model *vs.* Q_{Fe-MBS} from the other structural models obtained from B3LYP/ 6-311+g** triplet state calculations; FeP-comp. (green), FeP-ext. (purple) and FeP-min. (orange). The line shows the least-squares fit between Q_{Fe-MBS} FeP *vs.* FeP-comp.; $R^2 = 0.994$.

With respect to the effect of the structural model, the oversimplified FeP-min. and -ext. structures were only able to somewhat reproduce the relative experimental trends, especially with respect to the wild-type structure where the largest discrepancy, arising from haem 1U56 - B501, is the result of its resolved conformation, in which the minimum basis distortions are very much smaller than its counterpart in the asymmetric unit (Figure 6-6; note large error bars for wild-type). However, the reasonable degree of linearity found within the results from the mutant complexes (*e.g.*, $R^2 = 0.90$ for the linear relation between H-NOX E_m and the FeP-min. $NPA_{Fe-CORE}$ from the triplet calculations with the 6-311+g**) suggests that knowledge of the reduced basis NSDs may allow extraction of trend information, in the limit of low total distortion and with caution. This could be advantageous in situations where resolutions are poor, since lower-energy distortions are often larger than their higher energy counterparts and therefore easier to resolve, although for this to be possible it is clear that further work regarding the precise sensitivity of the calculated parameters upon higher energy distortions is necessary before their application in quantitative work.

Another feature of the FeP-min. and -ext. models is that the more planar mutant structures exhibit a systematically greater calculated partial charge at the haem-Fe. This may be either a consequence of the lack of higher-energy distortions that correlate with those of lower-energies of the same symmetry or else the result of a subtler aspect of the NSD process itself, specifically that the reference macrocycle used for the in-plane deformation decomposition is a Cu(II) porphin,[32] which may mean that some higher-energy in-plane distortions are necessary to describe particular characteristic differences between this

reference and an Fe-porphin. On the other hand, the systematically lower partial charges of the reduced basis models of the wild-type cofactors is most likely the result of the increasing dominance of the higher-energy distortion modes in affecting the molecular wave function, and thus the relationships between the distortions are suggested to be the cause of this behaviour.

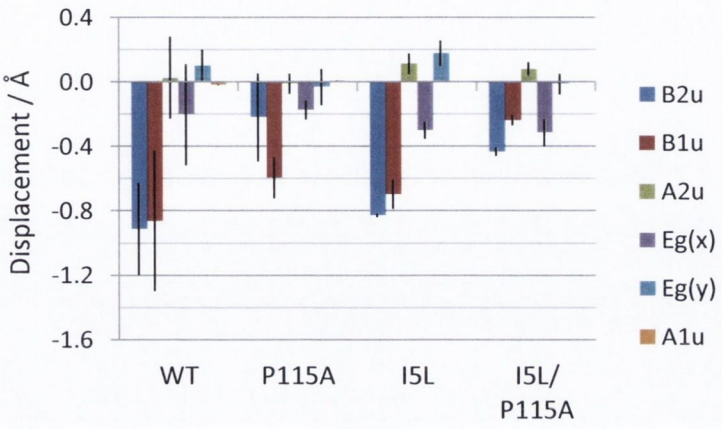


Figure 6-6: Mean out-of-plane minimum basis NSD of each haem from the H-NOX crystal structures used in the calibration. Error bars indicate two standard errors; PDBs 1U56, 3NVR and 3NVU, n = 2; 3EEE, n = 4 (haems in asymmetric unit).

6.4 Application to RC-cyt

6.4.1 Estimating Haem Conformations in *B. viridis* Cytochrome Subunit

The first task in elucidating the extent of conformational control in the RC-cyt subunit was to confirm that they possess distinct conformations with respect to one another. As previously shown,[60] and discussed in the experimental section, this can be achieved by agglomerative hierarchical clustering (AHC) of their NSDs. In this case, the clustering of the cofactors' NSDs was predominantly dependent upon the cofactor's identity which shows that the crystallographic experiments supported the hypothesis that each cofactor adopted a unique conformation (Figure 6-7).

However, further consideration as to the composition of the clusters indicated that the H1 cofactors exhibited two distinct possible distortion patterns, a feature that was either related to two actual conformers or else to systematic differences between experiments (*i.e.* different systematic errors), in which the conformation of H1 is most affected. Since the previous experience has been that systematic differences can and do occur between multiple crystal structures of this type of macromolecule, which can be attributed to restraints during refinement,[60] and because cluster membership appeared heavily

dependent on the structure authors (Table 6-1), the possibility of the latter was investigated *via* analysis of the "structure-as-observation" data-matrix.

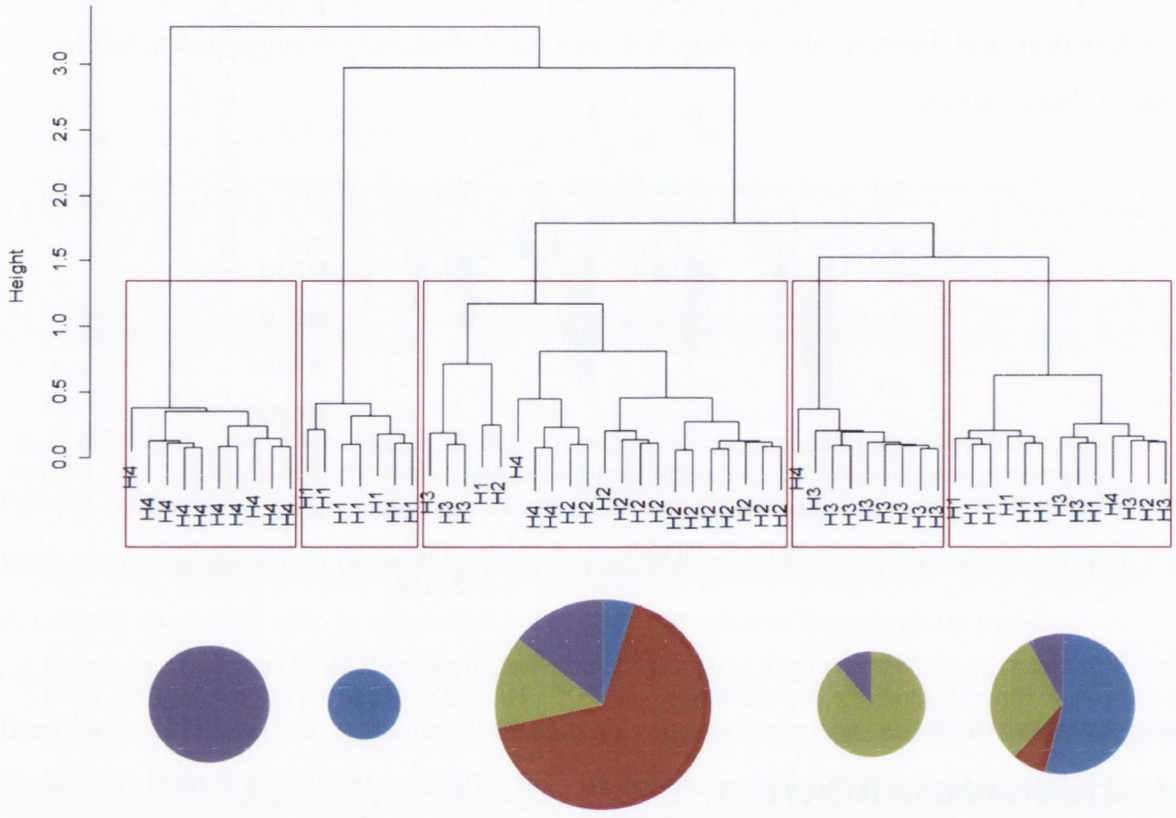


Figure 6-7: Dendrogram from the AHC of the cofactors' minimum basis, out-of-plane NSDs using the Euclidean metric and Ward's agglomeration. Pie-charts show the proportions of each cofactor in the rectangled cluster above it; blue = H1, red = H2, green = H3 and purple = H4. Note presence of two distinct clusters predominantly composed of cofactor H1.

The dendrogram and evolution of the within cluster sum of squares from the AHC of the structure-as-observation matrix suggests that there are 2-4 meaningful clusters, within which the structures show similar trends for the conformations of the cofactors and between, they exhibit systematic variations (Figure 6-8, left). Furthermore, the correlation bi-plot from the PCA analysis of the systematic variations (Figure 6-8) shows that the greatest contributing variable to the systematic differences is the B_{2u} distortion of H1 and hence the bi-modal conformation of H1 noted above is in fact the greatest discriminator between the structure classes.

The next key variables are the B_{1u} distortions of the H1, H2 and H4 cofactors and it is these correlations that all but confirm that the differences between the structures are artefactual since there is no known reason for the conformations of the cofactors to be coupled in this way. Also notable is the significant dependence of PC1 on the A_{2u} mode which becomes increasingly apparent when the mean conformations of the clusters are considered (Figure

6-9; note relatively large A_{2u} distortion of SC2 cofactors and correlate with their positions in the bi-plot Figure 6-8 [structures 5, 10 and 11]).

Table 6-1: Basic PDB structure information regarding the *B. viridis* RC crystal structures organised according to structure-cluster membership.

Cluster	PDB	Structure Authors	Res.	Fig. 4	Ref.
SC2	3D38	Li, L., Nachtergaele, S.H.M., Seddon, A.M., Tereshko, V., Ponomarenko, N., Ismagilov, R.F.	3.21	10	348
	2I5N	Li, L., Mustafi, D., Fu, Q., Tereshko, V., Chen, D.L., Tice, J.D., Ismagilov, R.F.	1.96	5	345
	3G7F	Ponomarenko , N.S., Li, L., Tereshko, V., Ismagilov, R.F., Norris Jr., J.R.	2.50	11	349
SC3	7PRC	Lancaster, C.R.D., Michel, H.	2.65	15	350
	2JBL	Lancaster, C.R.D.	2.40	6	346
	1DXR	Lancaster, C.R.D., Bibikova, M., Sabatino, P., Oesterhelt, D., Michel, H.	2.00	1	342
	5PRC	Lancaster, C.R.D., Michel, H.	2.35	13	350
	6PRC	Lancaster, C.R.D., Michel, H.	2.30	14	350
	2PRC	Lancaster, C.R.D., Michel, H.	2.45	7	232
	3PRC	Lancaster, C.R.D., Michel, H.	2.40	12	232
SC1	1PRC	Deisenhofer, J., Epp, O., Miki, K., Huber, R., Michel, H.	2.30	2	
	1R2C	Baxter, R.H., Ponomarenko, N., Pahl, R., Srajer, V., Moffat, K., Norris, J.R.	2.86	3	343
	1VRN	Baxter, R.H.G., Seagle, B.-L., Norris, J.R.	2.20	4	344
	2WJM	Woehri, A.B., Wahlgren, W.Y., Malmerberg, E., Johansson, L.C., Neutze, R., Katona, G.	1.95	8	347
	2WJN	Wohri, A.B., Wahlgren, W.Y., Malmerberg, E., Johansson, L.C., Neutze, R., Katona, G.	1.86	9	347

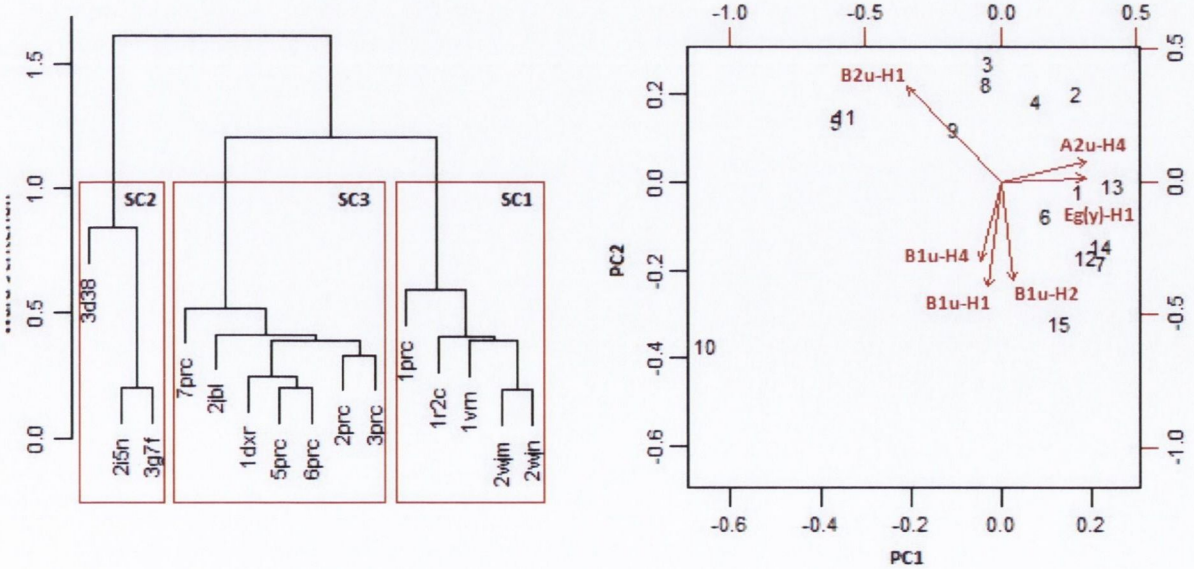


Figure 6-8: Dendrogram from the AHC of each PDB structures resolved minimum basis, out-of-plane NSDs of the four cofactors using the Euclidean metric and Ward's agglomeration (left; see S.I. Table 1 for structure key). Bi-plot from the PCA of the PDB structures' resolved minimum basis out-of-plane conformations of the four cofactors on the first two PCs and the top 25% correlating variables (right).

On the basis of this information a three cluster solution was pursued giving rise to three sets of mean conformations (Figure 6-9) highlighting the particular conformations of each cofactor and also confirming that the two apparent conformers of the H1 cofactor are largely characterized by the extent of the saddling (B_{2u} mode) of the macrocycle. In addition to this it is seen that, on the whole, the minimum basis distortions of the cofactors indicated by the means of SC1 (structure-cluster 1) and SC2 are in some ways more similar than those of SC3 (in contrast to the relationship indicated by the cluster hierarchy; Figure 6-8). Specifically, the saddling of H1 in SC1 and -2 is relatively large, whereas in SC3 it is similar across H1-H3 and the relative amount of saddling to ruffling (B_{1u}) across H1-H3 is more consistent between SC1 and -2 compared to SC3. However, the main feature that differentiates the SC2 cluster is the high-degree of doming (A_{2u}) compared to structure sets SC1 and SC3. In SC3 also, the conformations of the cofactors are more similar across the haems possibly indicating a greater influence of refinement restraints which could be responsible for the lesser nonplanarity of its conformation for H1 (Figure 6-9).

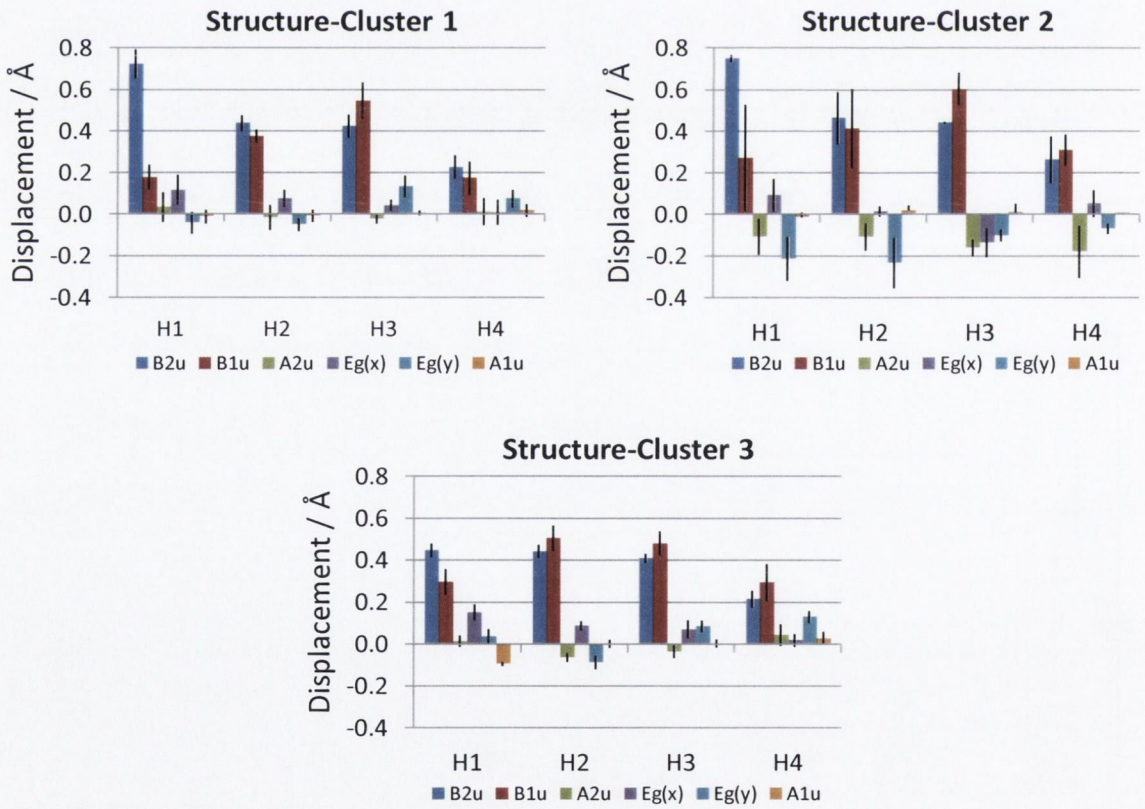


Figure 6-9: Mean out-of-plane minimum basis NSD of each haem cofactor from each cluster of crystal structures. Error bars indicate two standard errors. SC1 n = 5, SC2 n = 3 and SC3 n = 7.

6.4.2 Conformational Control in the *B. viridis* RC Cytochrome Subunit

Having established both a quantitative empirical relationship between the FeP-comp. models' triplet state B3LYP/6-311+g** $Q_{Fe\text{-}MBS}$ atomic charges and the conformationally

induced redox shifts, alongside the best estimate of the RC-cyt cofactors' conformations, the role of macrocycle mediated redox control in the RC-cyt was estimated. Providing first a qualitative consideration of the particular structure-cluster results it is seen that the SC1 structures' Fe-atom MBS Mulliken charges (Figure 6-10) lie in the order H1 < H2 ≈ H3 << H4 implying that naturally evolved conformational control contributes to the observed low-potential of the H1 cofactor, the H2 > H1 potential difference, is consistent with the similar potentials of the H2 and H3 cofactors and in some way works to increase the potential of the H4 cofactor. Alternatively, SC2 yields $Q_{Fe\text{-}MBS}$ that vary so that H1 ≈ H2 < H3 ≈ H4 which corresponds with the fact that the H4/H3 low- and high-potential pair cofactors have systematically higher potentials than the corresponding H1/H2 pair. Finally, the SC3 conformations give $Q_{Fe\text{-}MBS}$ ordered H1 > H2 << H3 ≈ H4 which again corresponds with the observed larger potentials of the H4 and H3 pair but uniquely suggests that the conformational differences exerts influence to lower the H2 potential relative to H1. Whilst these results show clear differences between the inferred conformational effect, an emergent trend is clear; it seems that conformational control operates to raise the potentials of the H3/H4 pair relative to their counterparts (or conversely, to lower those of the H2/H1 pair). Before discussing this further however, the quantitative estimations of the distortion/redox influence will be considered.

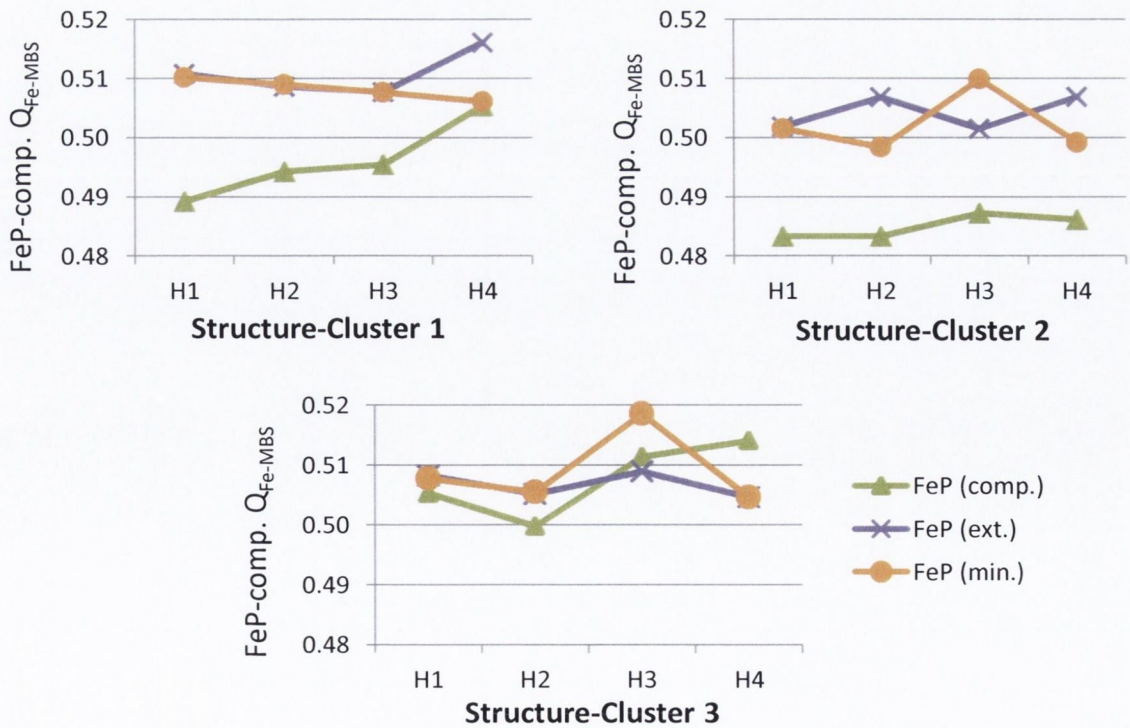


Figure 6-10: Fe-atom MBS Mulliken partial atomic charges (QFe-MBS) calculated from mean FeP-comp., -ext. and -min. models of the three structure-cluster B. viridis RC-cyt haems mean conformations in the triplet state with the B3LYP/6-311+g** method.

Noting that the RC-cyt $Q_{Fe\text{-}MBS}$ (Figure 6-10), all fall within the range of interpolation from the H-NOX calibration (H-NOX FeP-comp 0.471037 – 0.557198e; RC-cyt H1-4 FeP-comp: SC1 0.489164 – 0.505313e; SC2 - 0.483259 – 0.487204e and SC3 - 0.499779 – 0.514082e) (FeP-comp: $E_m \approx -30440\ Q_{Fe}^2 + 33634\ Q_{Fe} - 9118.6$), the "H-NOX potentials" are obtained (*vide infra*), allowing calculation of the relative potential shifts due to conformational control (Table 5-2).

In terms of a formal interpretation, if one considers that the reduction potential of a haem in a protein complex arises from a combination of perturbations effected by the environment relative to a free haem (reference) in solution, then the potential may be written as a sum of the reference potential and the perturbation as in Equation 6-1.

Equation 6-1: Simple model for the observed E_m shift of a protein-bound haem.

$$E_m^{complex} = E_m^{solution} + \Delta E_m^{perturbation}$$

To elaborate, the last term in the above is a sum of the contributions from any environmental perturbation such as H-bonding, electrostatics, axial-ligands (if different from the reference), solvent accessibility and conformational control. In this case, since we are interested solely in delineating the effect of conformational control we consider an additive separation of the environmental effects and the intrinsic conformational potential shift as shown below.

Equation 6-2: Expanded model for the observed E_m shift of a protein-bound haem.

$$E_m^{H-NOX} = E_m^{solution} + \Delta E_m^{environment} + \Delta E_m^{conformation}$$

Finally, since the H-NOX series of complexes investigated varies solely in the last term, taking the difference of any two of their potentials yields exclusively the *change* in the conformational component (Equation 6-3).

Equation 6-3: The interpretation of the E_m difference between either the H-NOX mutants described by Olea *et al.*[121] or of any of the calculated 'H-NOX E_ms' in this work.

$$\Delta E_m^{H\text{-}NOx} = \Delta(\Delta E_m^{conformation})$$

Consequently, taking the difference of any two interpolated "H-NOX potentials" (the potential that the H-NOX complex would exhibit if the haem within adopted the conformation of interest), yields the potential difference effected by their conformational difference.

A possible caveat to this approach is that the extension of the H-NOX relationship to other proteins is dependent on whether or not the magnitude of the intrinsic potential shift brought about by macrocycle distortion is sensitive to the *specific* environment of the cofactor. Since there is no suggestion in the literature to the contrary, this line of investigation was continued to assess the impact of conformational control on the *B. viridis* haems under the assumption that there is no such effect, although it must be noted that it is certainly possible that there are such differences.

The data in Table 6-2 show the experimental and most recent, state-of-the-art, theoretically calculated cofactor potential shifts together with the estimated influence of conformational control obtained from the B3LYP/6-311+g** calculation. The interpretation of the entries in the middle of the non-experimental data correspond to the percentage of the difference accounted for by the quoted method, *e.g.*, Voigt and Knapp's method accounted for 48% (36 mV) of the H1-H4 difference whilst the method reported here suggests that 72% (54 mV) of the difference is due to conformational control.

In most cases, the calculated perturbation of conformational control affects a change in potential that acts in concert with the other modulatory methods of the binding-site to increase the potential difference between cofactors (*i.e.*, has the same sign as the experimental difference), except for H4 relative to both H2 and H3. Indeed, it is notable that the H4-H2 difference is the only one which Voigt and Knapp's method overestimated, whilst that of H4-H3 is the only one that was not underestimated (the near exact agreement of their value for this difference is due in fact to a systematic error of around -30 mV in both the H3 and H4 absolute potentials). Furthermore, the two largest potential differences due to conformational control as inferred from any one of the three sets of mean conformations are found to involve the H3 or H4 potentials with those of H2 or H1, respectively, such that the potentials of the former are raised relative to the latter (Table 6-2).

It is convenient to assess the quantitative implications of this work in the context of improving the current theoretical understanding of the factors affecting the haem potentials by combining the Voigt and Knapp estimates of the inter-haem potential differences with any one of the estimates of the influence of conformational control presented here. This should be deemed acceptable under the assumption that those previous theoretical estimates were precisely lacking this effect and because the isolation of it is total. Importantly, doing so provides significantly improved experimental agreement (RMSDs = 27 mV for Voigt and Knapp,[99] and 19 mV, 17 mV and 20 mV after including the estimated

conformational effect from mean conformations SC1, SC2 and SC3, respectively). Furthermore, considering the individual corrections, it is clear that the SC1 structures' derived shifts may well provide an even greater improvement to the absolute potentials than is implied by the reduction of the RMSD experimental agreement of the inter-haem relative differences because these results break the effect of Voigt and Knapp's systematic error for the H3 and H4 cofactors, if it was possible to determine these soundly.

Whilst it has already been suggested that the H1 conformation of SC3 may be less reliable from the structure-determination point of view, a pragmatic view of these results also suggests that this may be an erroneous result. This point is taken simply from the observed experimental disagreement of the predicted relative H1-H2 difference, which was already underestimated by Voigt and Knapp, that is further reduced by SC3's contribution and there is no systematic error in the theoretical potentials for these as there was with the H3-H4 cofactors (see above).

6.5 Conformational Origin of Haem Redox Modulation

A key feature that has not yet been discussed is that in the *B. viridis* cytochrome subunit, it is the cofactor with the *lowest* total nonplanar distortion (H4, Figure 6-6, p. 145 and Figure 6-9, p. 148) that exhibits a consistently high relative partial charge and that the cofactor with the *greatest* distortion exhibits one of the *lowest* charges (H1), which is in stark contrast to the situation in the H-NOX complex. Thus, it appears that the redox influence of the cofactor conformations is inverted between the two complexes. In order to begin to explain the origin of this phenomenon it is required to investigate the correlations between the macrocycle modes of distortion and the calculated population parameters. The following discussion relates to preliminary results only as, in general, the derivation of a logical relationship between haem conformation and electronic properties is a complicated endeavour and indeed a long term goal.

Table 6-2: Absolute differences, $d_{i,j} = H_j - H_i$, and relative differences to experiment, $d_{i,j} = (H_j - H_i)$calc. / $(H_j - H_i)$expt., of the haem midpoint potentials derived from the average of the two published experimental measurements; Voigt and Kanpp's potentials of the *B. viridis* haems and of calculated E_m(H-NOX) of the *B. viridis* haems derived from $Q_{Fe\text{-}MBS}$ obtained from the triplet state B3LYP/6-311+g** wave functions of the FeP-comp. models. The third main column provides the experimental errors of the Voigt and Knapp model and of this model combined with the various structure-cluster estimates, whilst the fourth shows the change in the error relative to Voigt and Knapp that inclusion of the calculated potential shifts brings about.

		Potential difference / mV				Expt. correlation				Expt. error of model + Voigt & Knapp				Change in absolute expt. error of model			
		H1	H2	H3	H4	H1	H2	H3	H4	H1	H2	H3	H4	H1	H2	H3	H4
Expt. average[235, 236]	H1	0	365	435	75												
	H2	-365	0	70	-290												
	H3	-435	-70	0	-360												
	H4	-75	290	360	0												
Voigt & Knapp[99]	H1	0	348	395	36		0.95	0.91	0.48		-17	-40	-39				
	H2	-348	0	47	-312			0.67	1.08			-23	-22				
	H3	-395	-47	0	-359				1.00				1				
	H4	-36	312	359	0												
SC1	H1	0	19	23	54		0.05	0.05	0.72		2	-17	15		-15	-23	-24
	H2	-19	0	4	36			0.06	-0.12			-19	14			-4	-8
	H3	-23	-4	0	32				-0.09				33				32
	H4	-54	-36	-32	0												
SC2	H1	0	0	16	11		0.00	0.04	0.15		-17	-24	-28		0	-16	-11
	H2	0	0	16	11			0.23	-0.04			-7	-11			-16	-11
	H3	-16	-16	0	-5				0.01				-4				3
	H4	-11	-11	5	0												
SC3	H1	0	-17	16	23		-0.05	0.04	0.31		-34	-24	-16		17	-16	-23
	H2	17	0	33	40			0.47	-0.14			10	18			-13	-4
	H3	-16	-33	0	7				-0.02				8				7
	H4	-23	-40	-7	0												

Combining the H-NOX and cytochrome data and regressing the haem-Fe MBS Mulliken charges and NPA core populations against the individual NSD parameters revealed one candidate above all others as the main indicator of the electronic population at the metal centre, namely the A_{1g} or *breathing* mode. The extent of this distortion provided statistically significant correlation coefficients of 0.87 and 0.94 with the minimum basis cofactor projections and the corresponding $Q_{Fe\text{-}MBS}$ and $NPA_{Fe\text{-}CORE}$ values and, 0.78 and 0.83 for the complete basis projections (observed macrocycle conformation, centred Fe) and corresponding populations. Furthermore, because this normal-mode represents most closely the actual macrocycle core size, and thus the Fe-N bond lengths, this behaviour is perfectly understandable. It is important to highlight here that this result does not render the nonplanar macrocycle distortions as superfluous or secondary next to the in-plane conformation, as the outmoded concept of In-plane Nuclear Rearrangement, alluded to in the introduction, attempted to. Rather, multicollinearities are present between the various normal-coordinates that suggest that the A_{1g} distortion is *determined* by the nonplanar distortions and thus provides a conduit with which the nonplanar conformation exerts the real influence. Indeed, this structural relationship has been known for many years[36] and is supported by the author's own semi-empirical calculations that find non-zero minima along the A_{1g} coordinate for arbitrarily imposed nonplanar distortions upon a previously optimized structure (not shown).

However, this analysis provides only explicit investigation of the global situation, *i.e.* the source of conformationally induced potential shifts over a range of 171mV and two distinct complexes (for one of which we had three sets of possible, systematically different structure determinations) whilst the maximum estimated effect of natural conformational control in the cytochrome was only 54 mV. Also, a perusal of the in-plane distortions of the *B. viridis* haems (Appendix B; Table B-6) suggested that whilst some of the variation appeared to correspond with the A_{1g} coordinate (in particular, the systematic lowering of the SC2 charges), this was by no means a conclusive resolution. Therefore, although at the expense of statistical sample size, in order to assess the situation in more detail the NSD/QM-population relations were assessed arising from various subsets directly. Specifically, regressions were performed on subsets that comprised only the results from: 1) all H-NOX structures alone, 2) the H-NOX mutants only (105 mV range), 3) the H-NOX P115A and I5L/P115A mutants (60 mV range), 4) the set of *B. viridis* mean conformations and 5) the *B. viridis* structure-clusters.

Not unexpectedly, all but the largest of these subsets were too small to produce any *strictly* statistically significant correlations after alphas were adjusted for multiple comparisons. On the other hand, the results do suggest that the full set of A_{1g} modes may contribute to a considerable portion of the within group population variations. Nevertheless, even if these correlated in-plane distortions do provide the connection between the observed effects in each complex, it would then suggest that the nonplanarity induced contraction is reversed for the case of H4 which exhibits a *contracted* macrocycle even though it is the most planar.

There is however, a potential explanation for some of these observations. The first step is to understand the *ruf* distortion is in general recognised to be associated with substantial core contraction, as it is known to do so for Ni-porphyrins,[36] and on this basis the H1-H3 Fe populations can be somewhat accounted for. Next, having only to account for the high-partial charge of H4 it may be that the bis-HIS coordination of this haem, which effects a large drop in its reduction potential and a corresponding increased Fe atom electron density, results in macrocycle contraction in order to stabilize this extra density.

At this point however, all that may be concluded with certainty is that the *specific* macrocycle conformation determines the Fe's electronic population and, as a corollary, the consequent conformational influence on the redox potential. For this reason, it may be not be appropriate to consider that the influence of nonplanarity upon haem redox processes may be described by the simple maxims that are generally relevant to macrocycle centred oxidation and reduction events without full consideration of the total stereochemistry. However, it is satisfactory to note that the results described in this study would suggest that the approach used here is capable of accounting for this fact.

6.6 The Missing-link of the Potential Variation?

By applying chemical intuition in an attempt to obtain an experimentally calibrated computational procedure with which to predict and delineate the influence of protein-induced macrocycle distortion upon haem reduction potentials, a strong correlation between haem-Fe MBS Mulliken atomic charges, obtained using B3LYP wave functions with minimal- to triple-ζ, doubly polarized diffuse basis sets, with the redox potentials of intact complexes was established in the circumstance that the haem conformation was the only substantially varying property. This relationship will assist future research in the realm of naturally occurring conformational control and with further development may also prove useful for the design and tuning of haem enzymes where conformational change

from residue mutations could be predicted using standard geometry optimizations and, from the resultant cofactor conformation, the mutation's consequent physicochemical effect which could potentially reduce the cost and effort required to engineer proteins with desired redox properties.

Using this relationship together with the derivation of the best currently available estimates of the conformations' of the haem cofactors in the reaction centre tetrahaem cytochrome subunit of *Blastochloris viridis*, it was found that conformational control may account for up to 70% (54 mV) of a particular potential difference and correlates with the greater potentials of each haem in the pair proximal to the special-pair with their respective distal counterparts. The estimated influence upon the reduction potentials of H1-3 appears to work concertedly with other protein influences to enhance the differences between them, whilst for H4 it significantly reduces the effects of bis-HIS coordination[21, 22] and the partial ionization of its propionates.[99] This shows that explicit consideration of the conformational contribution to haem reduction potential modulation *in situ* may provide a missing link with respect to understanding haem potential variability in general.

7 Contribution of Bacteriochlorophyll Conformation to the Distribution of Site-energies in the FMO Protein*

7.1 Introduction

Photosynthetic light-harvesting complexes (LHCs) are responsible for the collection of photonic energy from the sun and its delivery to the reaction centres. For this reason it is vital that the pathways of intra- and inter-protein excitation energy transfer (EET) are finely controlled by LHCs to ensure optimum photosynthetic efficiency.[50, 51] The process of EET is mediated primarily by the electronic coupling between chromophores as well as the site-energies of the individual pigments (*i.e.*, uncoupled excitation energies). The distance dependence of inter-chromophore electronic coupling is reflected by the overall architecture of LHC proteins,[51] through which the chromophore spatial distributions (*i.e.*, distances and relative orientations) are precisely controlled. In contrast, site-energies are modulated by the local interactions between each pigment and its protein binding-site. The site-energies represent the excitation energies of the individual pigments within the protein in the absence of electronic coupling and so may be envisaged as the absorption spectra of a hypothetical system consisting of the pigment in question embedded in the apoprotein in the holoprotein's conformation (*i.e.*, without the other chromophores). These site-energies together with the chromophore spatial distribution then determine the character of the excitonic coupling in the complex, which results in excited-states delocalised over groups of pigments rather than isolated excitations at individual sites. Therefore the measured spectroscopic properties of the complex that provide experimental access to the EET dynamics, such as the absorption spectra, arise from the exciton structure of the complex, which is governed by the chromophore spatial distribution and the pigment-protein interactions.

There are many ways in which proteins may influence the absorption characteristics and site-energies of bound chlorophylls (Chls) and consequently affect the directionality of EET. These means include the choice of axial ligand, the H-bonding environment

---

* The writer would like to emphasize that the main approach employed in this study is entirely novel and that the analysis and interpretation of the results is not yet complete. The reader is therefore encouraged to review the methods section closely and to follow the associated elaborations of the general interpretations of the results in the results sections. The inclusion of more data than is minimally necessary to justify the explicit conclusions drawn in this section is intended to allow one to formulate their own ideas regarding the importance of the various BChl constituents.

(particularly to donors that are conjugated to the aromatic system) and the electrostatic properties of the binding-site.[19, 107, 109, 112, 383] In addition, because of the well-known correlation between macrocycle conformation and the absorption and photophysical properties of porphyrins,[30, 36, 119] it is possible that protein-induced deformation of bound Chls also affords a significant contribution to the site-energy distribution in LHC proteins.

The Fenna-Matthews-Olson (FMO) protein is a water-soluble pigment protein complex that mediates EET between the chlorosomes and reaction centres (RCs) of green bacteria (Figure 7-1).[263] The FMO protein crystallises as a C_3-symmetric trimer, that is thought to resemble the *in vivo* form, where each monomer contains seven bacteriochlorophyll a (BChl a) pigments that are packed very closely together within the encapsulating protein; notably, there is no apparent symmetry in the monomer.[48, 49, 268, 276, 384] Additionally, an eighth BChl has been observed in crystal structures from some species, although which species are in possession of BChl 8 is an open question.[266] FMO proteins are distinguished into two groups based on their absorption spectra (type I or II) although the structural reasons for the differences are as yet unknown. The nature of BChl 8 has been ascribed a role in their differentiation,[266] however this remains inconclusive as it is especially confounded by the lack of knowledge regarding its species distribution.[267] The dynamics of exciton transfer in the FMO have recently been modelled and suggest that BChls 1 or 8 receives excitation energy from the chlorosome baseplate that is then directed through either BChls 2 or 4 and onto BChl 3, through which the FMO excites the RC BChls.[264]

As it is considered an excellent model system for the study of EET in photosynthesis, there has been fairly prolific theoretical analysis of the optical spectra of the Fenna-Matthews-Olson (FMO) protein.[111, 264, 269, 270, 385, 386] In particular, because of the availability of relatively high-resolution structural studies of the FMO complex,[48, 49, 266-268, 276, 277, 384, 387-389] a number of these theoretical studies have focussed on predicting the spectral characteristics and EET dynamics directly from the atomic structure. These structure-based studies were initiated by Fajer and coworkers[111] in 1990 using the atomic coordinates provided by the 1.9 Å resolution structure of the FMO complex from *P. aestuarii* (PDB ID: 3BCL).[276] This work utilised semi-empirical ZIndo/S calculations upon various structural models derived from the experimental coordinates consisting of the BChl at various levels of truncation, with and without an imidazole ligand (as a model for HIS) and additional models including nearby charged- and aromatic residues. The procedure of sequentially truncating the structural model used in each calculation appears to have been performed in the spirit of delineating their individual contributions. However, whilst the

results they obtained provide an idea of the overall effect of a particular perturbation (*e.g.*, removal of the acetyl introduces an almost systematic blue-shift) they were unable to obtain a simple relationship between the conformational parameters and highlighted that this implied the presence of, *"a nonadditive relationship between the influences of framework distortion and acetyl orientations"*.[111]

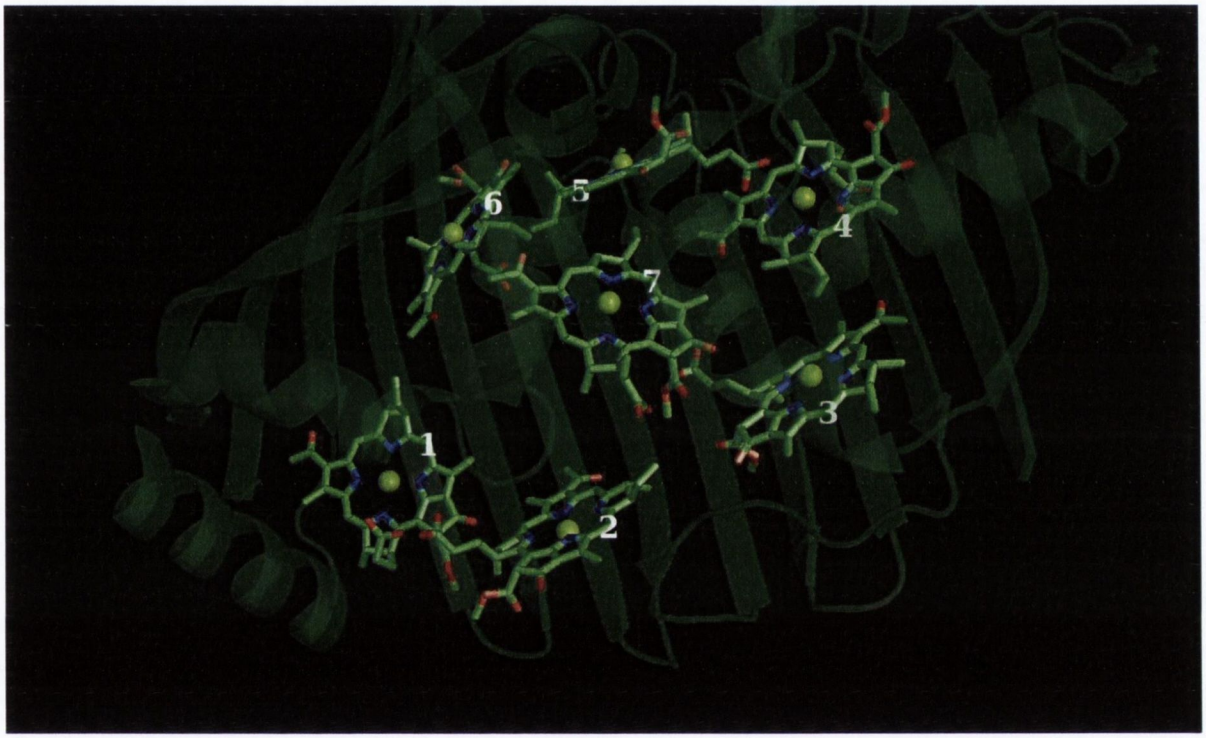


Figure 7-1: The BChl arrangement in the Fenna-Matthews-Olson protein (phytyl chains have been truncated for clarity). Drawn using PyMol and the coordinates from PDB ID: 3BCL. BChl 8 it not present in this structure although when present is situated outside of the encapsulating β-sheets, close to BChl 1.

Since then, modern work has culminated in numerous approaches with greater computational sophistication that provide more or less *ab initio* site-energies that yield simulated spectra in good agreement with experiment, yet the individual site-energies and the dominant influences of the protein are still contentious.[385] Of particular note is the research by the Renger group that culminated in the development of a method they called "Poisson-Boltzmann quantum chemistry" (PBQC) wherein site-energies were calculated by perturbing the solvent absorption with the interaction of the ground and excited state BChl charge distribution (TD-B3LYP/6–31G*) with the protein's.[264, 269, 270] However, the progression of Renger's research has culminated in their exclusion of significant conformational influences upon the site-energies stating that, *"...the different conformations of the BChls were found to be least important* [with respect to the site-energy shifts]", and found the dominant factor to be short- and long-range electrostatic field effects induced by the protein.[269, 270] In spite of the agreement between these site-

energies with those from fitting algorithms,[269, 270] the view that α-helices possess dipoles of sufficient size to influence site-energies has been challenged by Konig and Neugebauer,[385] who indicate possible flaws in Renger and coworkers' calculations.

It seems then that there is still considerable debate regarding the BChl site-energies in the FMO with respect to their values as well as to the details of the factors that affect their inhomogeneity, suggesting that there may still be cause to suspect a functional role for the variations in macrocycle conformations. Moreover, given the status of the FMO as a model complex for how LHCs control the dynamics of EET in photosynthesis, it would be interesting to uncover how the conformational properties of BChls contribute to this process. However, Fajer and coworkers' results indicate that it is no trivial matter to separate individual contributions such as the functional group conformations (*e.g.*, the C3-acetyl orientation) from the effect of the overall macrocycle conformation, highlighting that the simple approach of successive truncation is not enough to assess the situation.* It is clear then that in order to address the question of conformational control in the FMO; a more inclusive experimental design is needed to determine precisely how the site-energies of the BChls are directly influenced by conformational considerations.

Fortunately, an experimental design was found that is capable of delineating the individual contributions of a complex set of variables (*i.e.*, conformational parameters) from their mutual interactions, namely a 'full factorial' experiment.[390] In this design, all possible permutations of the experiment are conducted and simple linear models may then be used to extract the main influence of single variables as well as all orders of their interactions, quantitatively. In order to implement this design to model the conformational effects of the BChl macrocycle and its substituents on their *intrinsic* site-energies, the BChls were partitioned into molecular fragments and then each fragment was either frozen in the crystal structure geometry or allowed to relax *via* a (partial) geometry optimisation *in vacuo* at a given level of theory. Once all possible permutations of this experiment were performed the results were subject to a simple statistical analysis to obtain the separated conformational contributions to the site-energy.

---

* This contrasts with the result in Chapter 6 where the effect of conformational variability upon the redox potentials of haem cofactors was acceptably modelled by truncated models. This is most likely because the parameter used to model the haem potentials (Q_{Fe}) is largely dictated by the local coordination environment of the central ion, at the core of the macrocycle. In contrast, the low-lying excited states are modulated by orbitals that are fully delocalised over the pigment's π-system including conjugated substituents and are additionally influenced by the inductive effects of non-conjugated functional groups.

7.2 Experimental Details

7.2.1 C3-Acetyl Rotational Excitation Energy Profiles

It was considered possible that local features of the acetyl conformation, besides its coplanarity with the macrocycle, could have been responsible for the observations that led to the conclusion that there was a significant interaction between the effect of the acetyl orientation and the macrocycle conformation. In order to ensure then that this was not the case, a series of increasingly relaxed rotational excitation energy profiles (RexEPs) were calculated to see if they would coalesce when various parts of the acetyl group were relaxed at each dihedral angle. Specifically, in addition to rigid RexEPs, three relaxed RexEPs were calculated corresponding to:

1) methyl hydrogen dihedral relaxation
2) all internal parameters of the acetyl group plus the C3-C3^1 bond-length
3) as above, with the addition that all H-atoms were optimised at each scan point

Since the ZIndo/S[391] method is not suitable for geometry optimisations, the PM6[392] method was used to provide relaxed geometries for subsequent ZIndo/S calculation. As discussed in Chapter 3.4, the NDDO PM5 model has been shown to provide excellent geometries for Chls and BChls that are comparable to those contained using *ab initio* and DFT methods.[55, 258, 295-299] PM6 is likewise expected to perform well in this regard and has been shown, in general, to perform better than PM5 with respect to C, H, and N bond-lengths.[392]

The calculations were performed using the Gaussian 09 software on a Windows PC. The 'SCAN' keyword was used to affect geometry optimisations at O-C3^1-C3-C2 dihedral angles over the range -180° to 180° in 5° increments. The atoms to be optimised were specified using the OPT=ReadFreeze option. Subsequently, ZIndo/S calculations were performed with an active space of 15 HOMOs and 15 LUMOs; this method has also been shown to afford accurate first excitation energies for Chls and BChls and in particular, predicts the relative energies of the main absorption bands of compounds in this class.[55, 258, 295-299]

For simple numerical comparison, the RexEPs were fit to a 3 parameter function of the dihedral (Θ, radians) of the form shown in Equation 7-1. Where c acts as a relative baseline of the RexEP, $a + c$ gives λ at $\Theta = 0$ and the magnitude of b is an indicator of the difference (asymmetry) between the s-*cis* and s-*trans* conformations (Figure 7-2). Other functions also provide reasonable approximations and may even be considered to have

better properties, *e.g.*, cosine series, although a preliminary test indicated poorer agreement with the calculated data. Whilst the cosine term is a direct indicator of the degree of π-overlap of the acetyl with the macrocycle, it may also be possible to interpret the quadratic term as an electrostatic repulsive (steric) interaction of this group. As an alternative approach, the set of RexEPs were also clustered using the Euclidean distance between each RexEP vector to construct the dissimilarity matrix and Ward's method for the agglomeration. This is possible because the RexEPs are essentially vectors where each entry is the excitation energy at a particular dihedral angle.

Equation 7-1: The model to which each of the calculated RexEPs was fit to provide numerical parameters for direct comparison.

$$\lambda = a\cos 2\theta + b\,\theta^2 + c \;;\; -\pi < \theta < \pi$$

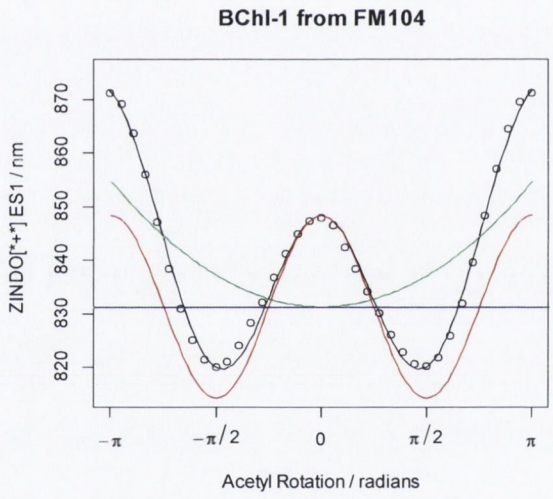


Figure 7-2: Graphical illustration showing the decomposition of the terms in the RexEP fit (Equation 7-1; constant (blue), a cos(2θ) (red) and $b\,\theta^2$ (green).

7.2.2 Fragment-based Partial Optimisations

When Fajer and coworkers[111] attempted to quantitatively model the effect of protein-induced structural perturbations of the FMO BChl pigments on their site-energies, they approached the problem of delineating the specific structural influences *via* the progressive truncation of the chromophores' structure. Whilst this was a highly informative approach, the possibility of observing interactions between the various structural features in a single pigment was rendered impossible by the experimental design. However, this sort of experiment can be improved by implementing what is known in the field of statistical experimental design as a full-factorial experiment. In this design, all possible permutations of the levels of the variables are characterised. For example, if one considers that whether or not a specific peripheral substituent of BChl a is truncated or not as a binary variable,

then the set of all 2^n possible structures defined in this way (n = the number of substituents) constitutes the full set of experimental units (EUs) in the full factorial design. The results from such a process allow the main effects of each structural feature, as well as all orders of their interactions, to be readily extracted.

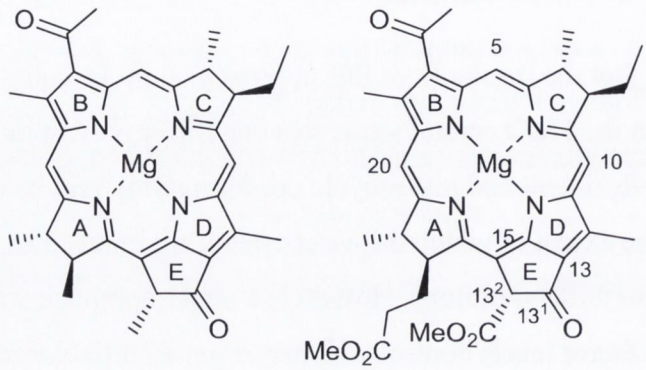
It was considered then that the results from this approach could be compared across a series of pigments from the FMO crystal structures that differ only in their conformational properties (*i.e.*, local substituent and macrocycle conformation), and would provide quantitative detail of the extent to which different effects are induced by the same functional groups due to these variations. However, a better approach was to run the calculations where the factor levels control whether or not a particular molecular component is optimised or not as this allows the conformational influence to be assessed directly. In this case, the results describe the influence of the individual conformational perturbations as well as their interactions.

Structure selection for factorial experiment

The first step in the implementation of this experiment was to define groups of atoms that represent molecular fragments of specific interest that would be treated as a single fragment, bearing in mind that this definition determines what level of detail the results provide as well as the size of the experiment. Whilst the number of single-point calculations required scales at 2^n (n = number of substituents considered), semi-empirical quantum calculations are feasible even for a highly partitioned molecular fragment scheme (*i.e.* approx. 1 s per calculation). In order to balance the demand for 'molecular resolution' *vs.* computational tractability two different schemes were used to construct the partial optimisation series and are presented below (Figure 7-3 and Table 7-1). Model B (n = 12; 4096 EUs) is feasible only for a semi-empirical approach (*e.g.*, PM6//ZIndo/S), whilst Model A (n = 5; 32 EUs) may be suitable for high-level *ab initio* calculations and is particular expedient with the semi-empirical methods.

A key point in model B was the separation of framework into individual rings (Table 7-1). This was done so that the effects of locally induced framework distortion could be differentiated from the effects of the general perturbative strain energy imposed by the binding-site that results in the common occurrence of the low-energy distortion modes in protein-bound cofactors. In order to allow sufficient freedom for each ring to relax in EUs where its neighbouring rings were fixed in the crystal structure geometry, the adjacent meso carbon atoms were included in the definitions of each ring 'group' or 'fragment'. Additionally, this model also separates the effects of specific skeletal distortion of ring E

from its substituents, afforded by the explicit specification of the ring E carbons, the $C13^1$-keto-oxygen and the $C13^2$-methylester each in their own groups. This model also separates the aliphatic substituents according to which N_{opp}-N_{opp} axis they lie upon.



Model A Model B

Figure 7-3: Definition of the structural models used in the partial optimisation calculations. Note that the ring nomenclature used throughout this chapter, indicated in the above structures, differs slightly from that used elsewhere in this thesis as here the PDB conventions have been employed.

Table 7-1: Molecular partitioning scheme of the BChls used to construct the partial optimisations. Each group of atoms (molecular fragment) is either frozen or optimised in each experimental unit (permutation) and thus its contribution to the BChl's *intrinsic* site-energy is isolated from the other BChl components.

Fragment	Model A Group[a]	Abbreviation[b]	Model B Group	Abbreviation[c]
Mg	1	Mg	1	Mg
Ring A + C_ms 5,20	2	Mac.	2	Ring | A
Ring B + C_ms 5,10			3	Ring | B
Ring C + C_ms 10,15			4	Ring | C
Ring D + C_ms			5	Ring | D
$C13^1$ + $C13^2$	3	CP	6	Ring | (E)
O=($C13^1$)			7	[C=]O
C3-acetyl	4	Ace.	8	$C(O)CH_3$
C2,C12-methyls	5	Alks.	9	Alkyls | Q_y
C7,C18-methyls + C8-ethyl			10	Alkyls | Q_x
$C13^2$-methyl ester	(3)[d]		11	Acid | $C13^2$
C17-methyl propionate	(5)[d]		12	Acid | C17

[a] Blocks of fragments assigned the same number are in the same atom group fragment; [b] these abbreviations refer to each 'Model A Group' number in the preceding column; [c] these abbreviations are used extensively throughout the tables in section 7.6.3 and are of the format "Class | Group"; [d] truncated to a methyl group (see Figure 7-3).

Model A represents a far less complicated design that was intended to retain the ability to characterise and differentiate the major conformational aspects of interest, specifically, the macrocycle framework, acetyl conformation, cyclopentanone ring and alkyl peripheral substituents. The truncation of the C17 and $C13^2$ substituents to methyl groups was done in an effort to correct an issue discovered in the model B run (model B was trialled first) in that different conformations were obtained in the fully optimised structures for this model

depending on the starting geometry (*i.e.*, crystal structure conformation). This confounded comparative results and so model A was developed to avoid this complication. Additionally, the model A design could be implemented at a fraction, specifically 0.88%, of the computational expense and was therefore useful to perform a comparative analysis of the results obtained by using the geometries found in different crystal structures.

Computational Implementation

All calculations were performed with the Gaussian 09[294] software package on the Lonsdale cluster provided by the Trinity Centre for High-performance Computing. The PM6 method and the OPT=ReadFreeze keyword to specify the atoms that were to be optimised and those which were to be frozen in the crystal geometry; in all partial optimisations hydrogen atoms were relaxed. For both models, the hydrogen atoms were added to the BChl crystal structure coordinates and optimised using PM6 and the optimised geometry was saved as a Z-matrix. A script, written in R,[282] was then used to generate all Gaussian 09 input files with the 'ReadFreeze' atom list corresponding to the EU (*i.e.*, the permutation of the partial optimisation series). ZIndo/S(15,15) calculations were then performed at each optimised geometry.

For model B, a few optimisation schemes were tested to find the most computationally efficient procedure (*i.e.* in terms of CPU time). The optimal approach involved performing a single frequency calculation on the starting geometry from the PDB coordinates and then using the analytical force constants to start each (partial) optimisation using the OPT=ReadFC option. Any EUs that were not successfully optimised using this method were restarted from the last available geometry and new analytic force constants were calculated to initiate the optimisation using the OPT=CalcFC option. EUs that were not converged at this stage, were started from the beginning with analytic force constants calculated at each stage in the optimisation using the OPT=CalcAll keyword. As a last resort, particularly when optimisations were oscillating between geometries, the step size was decrease using the OPT=MaxStep option; when necessary, this afforded a converged optimisation. The optimisations were performed using the default convergence criteria in Gaussian 09.[294]

For model A, all calculations were initiated with individual analytic force constants using OPT=CalcFC and a similar procedure was followed to that above in order to converge troublesome optimisations. These optimisations were performed using the 'tight' convergence criteria in Gaussian 09.[294] Additionally, a few of the BChl models failed to achieve SCF convergence using only the default algorithm, these were successfully

converged by allowing quadratic convergence steps when necessary using the SCF=XQC option (Table 7-2).

Table 7-2: List of structures requiring extra quadratic steps to achieve SCF convergence.

PDB ID	Residue No.
1KSA	A370
	A372
3ENI	A376[a]
	A377[b]
3OEG	A403
3VDI	A406

[a] all conformers excluding AD; [b] all conformers

Analysis of Results

Once all EUs were completed for a given structure, the first excitation energies (E_{ES1}) were compiled together with their associated atom group partial optimisation permutation vectors in effect coding. The permutation vector was an ordered *n*-tuple where each entry corresponded to one of the group definitions (*i.e.*, $n = 5$ and 12 for models A and B, respectively; Table 7-1) and was set to '1', if the atoms in the group definition were optimised or, '−1' if they were frozen in the crystal geometry.

Subsequently linear least-squares fits were obtained with the *lm* function in the *base* package of the R statistical environment,[282] using functions constructed from sets of the 1 to *n*-order variable and variable interaction terms with the latter represented as products of the elements in the permutation vectors. To illustrate, the first order terms for Mg and the macrocycle in model A were represented by the first and second entries in the associated permutation vector, respectively, whilst the second order interaction for these groups was defined as the product of these elements. As will be shown in detail later, the coefficients for the first order terms represent the 'main effect' of optimisation from the observed geometry or, conversely, the protein-induced perturbation, of the individual atom groups. The interaction terms represent the modulatory effect that one atom group has on another.

7.3 Conformational Flexibility in the FMO Complex

7.3.1 Summary of Relevant Crystal Structures

In all there are 12 crystallographic models of FMO proteins from three distinct species deposited in the Protein Data Bank (PDB) all of which are of high-resolution (up to 1.3 Å; Table 7-3 and Table 7-4; see also Table 7-5 for map of PDB BChl residue IDs to FMO BChl identity) in comparison to other photosynthetic proteins. Despite the fact that the stereochemical restraints of the macrocycle portion of the BChls were weighted by 1000 times more than the coordinates obtained by fitting to the electron density map,[384] even the

earliest structure (PDB ID: 1BCL) indicated the presence of core conformational variability and differences in the positions of the central-Mg atoms. Differences in the individual pigment-protein interactions began to be identified when a tentative "X-ray sequence" and trace of the protein α-carbon backbone were provided using the same dataset, but the BChl conformations were not updated (PDB ID: 2BCL).[49]

Table 7-3: Current crystal structures of FMO complexes deposited in the PDB; key experimental parameters are provided together with each species' spectral classification ("Type").

	Species	Type	Res.	Year	Notes	Ref.
3EOJ	*P. aestuarii*	1	1.30	2009	"Weighted full matrix least squares procedure. The bond lengths and angles of all BCL molecules, except for 378, were restrained to the group averages calculated from these seven molecules. No external library was used. The bond lengths and angles of 378 were restrained to the average values of the other seven BCL molecules" – from 3EOJ PDB remarks	266
3ENI	*C. tepidium*	2	2.20	2009	Two FMO complexes in asymmetric unit, with slightly different conformations of corresponding BChls in each. Problem with BChl 8.	266
3VDI	*P. phaeum*	1	1.99	2012	No BChl 8; starting model = 3EOJ.	267
3BSD	*C. tepidium*	2	2.30	2004	First observation of BChl 8	388

Table 7-4: Obsolete crystal structures of FMO complexes accessible in the PDB; key experimental parameters are provided together with each species' spectral classification ("Type").

	Species	Class	Res.	Year	Antecedent	Notes	Ref.
1M50	*C. tepidium*	2	2.20	2003	3ENI	Earlier model [SM: of 3ENI] refined against the same diffraction data set.	387
1KSA			2.20	1997	3ENI		277
3OEG	*P. phaeum*	1	1.99	2011	3VDI	Erroneous BChl 8	389
4BCL	*P. aestuarii*		1.90	1993	3EOJ	Updated BChl conformations	268
3BCL			1.90	1986	3EOJ	Includes structural classification of BChls into two classes by macrocycle conformation.	276
2BCL			2.80	1979	3EOJ	Trace of protein; same BChl coordinates as 1BCL	49
1BCL			2.80	1977	3EOJ	Preliminary coordinates of BChls only	48, 384

Table 7-5: PDB file residue ID numbers for the BChls corresponding to their position in the FMO.

BChl	1BCL	2BCL	3BCL	4BCL	1KSA	1M50	3BSD	3EOJ	3ENI	3OEG	3VDI
1	1	1	1	367	366	757	367	371	371	401	401
2	2	2	2	372	367	758	368	372	372	402	402
3	3	3	3	371	368	759	369	373	373	403	403
4	4	4	4	369	369	760	370	374	374	404	404
5	5	5	5	373	370	761	371	375	375	405	405
6	6	6	6	368	371	762	372	376	376	406	406
7	7	7	7	370	372	763	373	377	377	407	407

Substantial variation amongst the BChl conformations were reported some years later when the structure was refined to a resolution of 1.9 Å (PDB ID: 3BCL).[276] As a result, the BChls were classified into two groups (class I: BChls 1, 2 and 3; class II: BChls 4, 5, 6 and 7) based on their specific nonplanarity using pairwise correlation coefficients calculated from the atomic z-displacements from the mean-plane. Differences in the dihedral angles measuring the co-planarity of the acetyl group with the macrocycle plane were also reported. Additionally, the previously suggested Mg-coordination geometries were confirmed and specific ligand parameters were provided and differences amongst the C8-ethyl, C13^2-carbomethoxy and phytyl substituent conformations were observed.[276]

The NSDs of the BChls from this structure show clearly that there is substantial variation amongst the pigments (Figure 7-4). However, there is no clear distinction between the class I and II pigments, as defined by Tronrud *et al.*,[276] which confirms the suggestion by Fajer and coworkers,[111] who assessed the pigment conformations using a few more conventional parameters before using the structures in their ZIndo/S study.

The next development came after the chemically determined amino acid sequence[393] was incorporated into a new structural model (PDB ID: 4BCL).[268] Although the previously inferred sequence used to construct the earlier models was quite accurate and the major conclusions from the previous studies were upheld, the segregation of the BChls into classes determined by their apparently correlated nonplanarity was uncorroborated by the newly refined BChl parameters. Consideration of the NSD results from this structure (Figure 7-5) shows clear differences from the previous result although the overall distribution of conformations is the same; a large systematic difference is the *bre* displacement (A_{1g}) that indicates substantially contracted BChls in 4BCL relative to 3BCL.

The second species to be characterised crystallographically was *Chlorobium tepidium* and the structure was solved using the *P. aestuarii* structure (PDB ID: 4BCL) with molecular replacement to a resolution of 2.2 Å (PDB ID: 1KSA).[277] Representative of the high

sequence homology between the two species (78% identical and 88% similar plus identical residues),[394] structure alignments provided low, yet significant, r.m.s. deviations of 0.86 Å and 1.50 Å between the α-carbons and all atoms of conserved residues, respectively. Likewise, alignment of the BChls showed that their overall arrangement was similar and this was further emphasised by the lack of any significant difference in the inter-pigment distances (with respect to conjugated atoms) between the two structures. In contrast, significant differences between the conformations of corresponding BChls were observed and these were pointed out to arise mostly due to differences in the planarity of the macrocycles and additionally the acetyls of BChls 3-6 were also noted to be less coplanar than in *P. aestuarii*. Note also, that the authors considered BChl 7 to be most conserved and BChls 2 and 6 to be most divergent. Whilst the axial ligands of all BChls were conserved between the species, some differences in H-bonding to the acetyl and keto groups were observed, particularly in BChl 7, although overall these are considered to be minor and thus the authors suggested that the differences in the macrocycle conformations are the main structural reason for the spectral variation. An important systematic difference was that all Mg atoms were found to be substantially more in-plane such that the mean displacement of Mg was 0.09 Å in *C. tepidium* compared to 0.48 Å in *P. aestuarii*. The effect of the refinement protocol on this feature was investigated and it was shown that refinement of the BChls with TNT[395] (the software used to refine the *P. aestuarii* structure) led to an increase of the mean Mg displacement to 0.25 Å although this change was considered to be within the estimated positional error of the model.

The NSDs for the remaining structures, 1KSA, 1M5O and 3BSD are illustrated in Figure 7-6 to Figure 7-9, respectively. Substantial variation is apparent but it should be noted that there are many relative differences common to all. However, rather than describing these structures in detail, we will move on to a detailed comparison of the conformations as provided from the high-resolution structure 3EOJ.

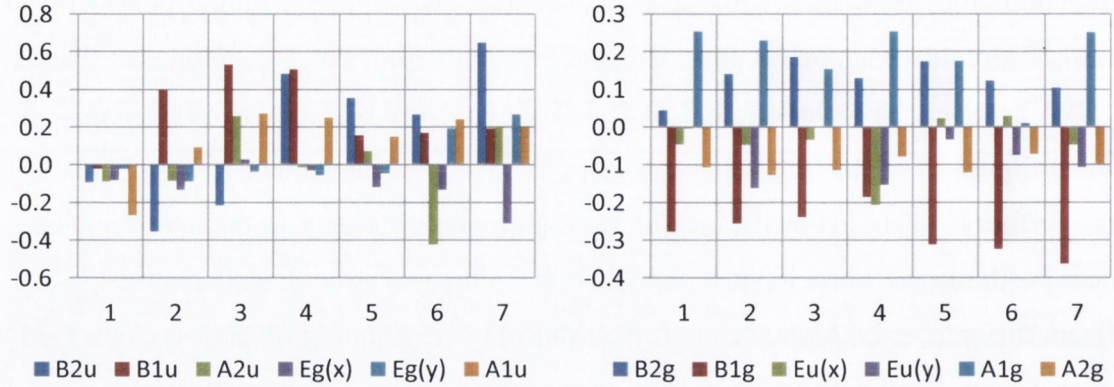


Figure 7-4: Minimum basis normal-coordinate displacements of each BChl pigment in the *P. aestuarii* FMO protein as described by PDB ID: 3BCL. Out-of-plane (top) and in-plane (bottom). Y-axes in Ångstroms. NB. No BChl 8 is present in this structure.

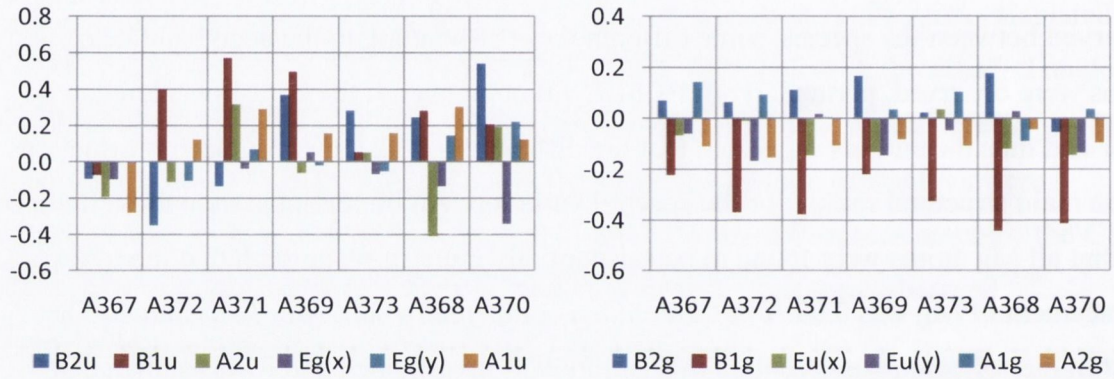


Figure 7-5: Minimum basis normal-coordinate displacements of each BChl pigment in the *P. aestuarii* FMO protein as described by PDB ID: 4BCL. Out-of-plane (top) and in-plane (bottom). Y-axes in Ångstroms. NB. No BChl 8 is present in this structure.

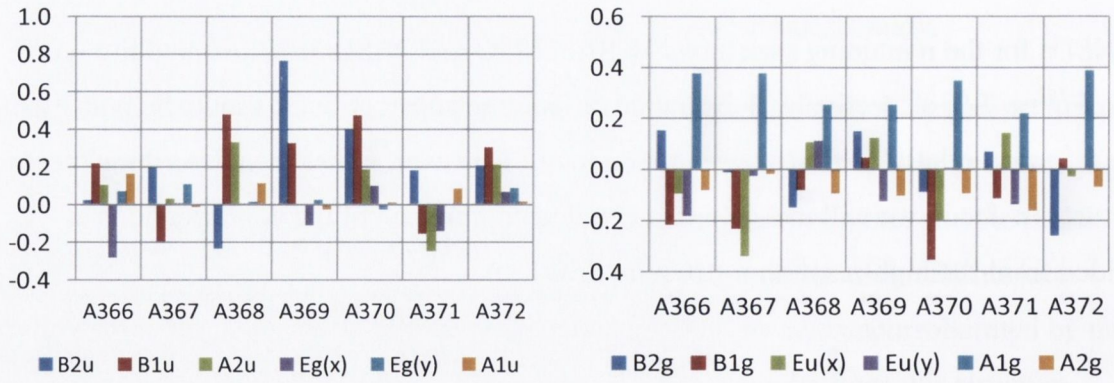


Figure 7-6: Minimum basis normal-coordinate displacements of each BChl pigment in the *C. tepidium* FMO protein as described by PDB ID: 1KSA. Out-of-plane (top) and in-plane (bottom). Y-axes in Ångstroms. NB. No BChl 8 is present in this structure.

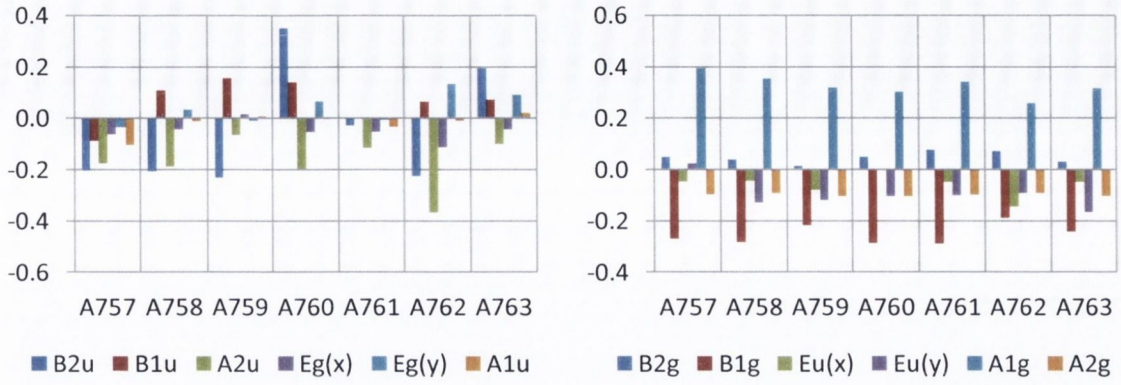


Figure 7-7: Minimum basis normal-coordinate displacements of each BChl pigment in the *C. tepidium* FMO protein as described by PDB ID: 1M50. Out-of-plane (top) and in-plane (bottom). Y-axes in Ångstroms.

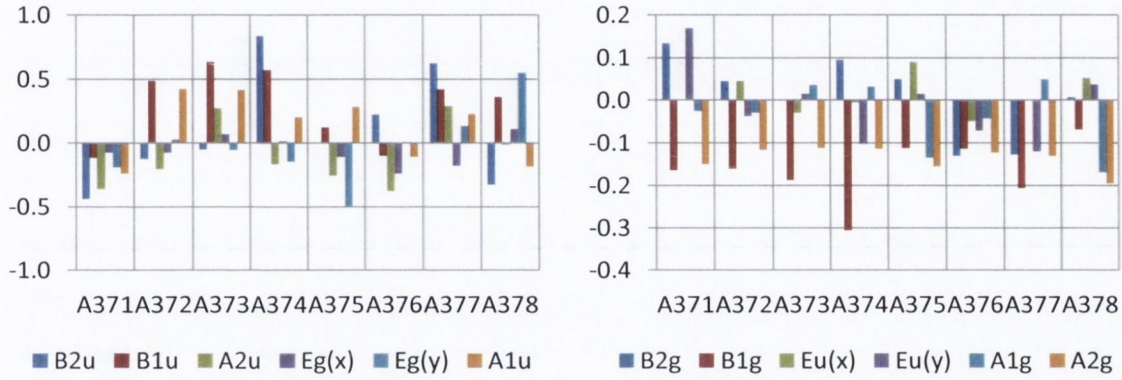


Figure 7-8: Minimum basis normal-coordinate displacements of each BChl pigment in the *C. tepidium* FMO protein as described by PDB ID: 3BSD. Out-of-plane (top) and in-plane (bottom). Y-axes in Ångstroms.

7.3.2 NSD Analysis of Pigment Conformations

In light of the high-resolutions attained by the modern crystal structures of the FMO, the consequent low coordinate ESDs and the well-defined BChl restraint procedure that was employed (see above) it seems plausible that these structures will turn out to represent the best atomic estimate of the BChl conformations. As this is the case, it is useful to perform a descriptive NSD-based conformational assessment of the BChls from these structures in lieu of a thorough statistical analysis, which would be an appropriate start for follow-up work; in any case, it is likely that the following will remain qualitatively valid if not entirely quantitatively. The following references the NSDs obtained from the crystal structure of the FMO from *Prosthecochloris Aestuarii*[266] (PDB ID: 3EOJ).

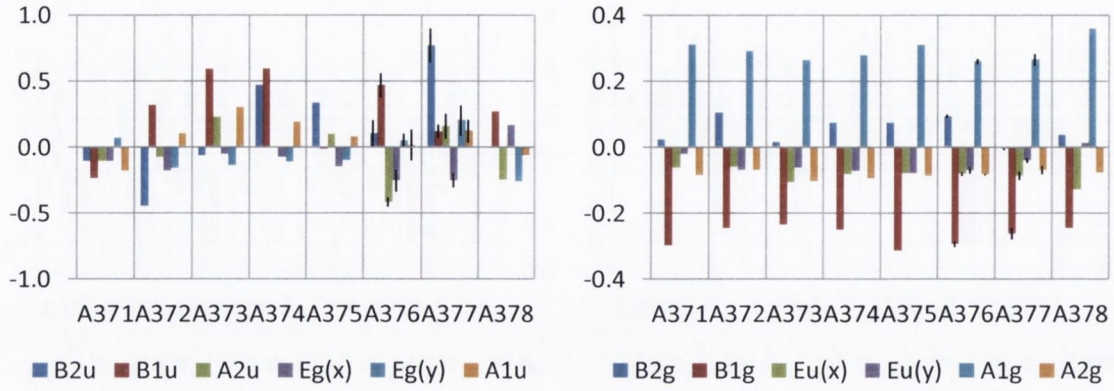


Figure 7-9: Minimum basis normal-coordinate displacements of each BChl pigment in the *P. aestuarii* FMO protein as described by PDB ID: 3EOJ. Out-of-plane (top) and in-plane (bottom). Y-axes in Ångstroms; the average values for the four alternate conformers of BChls 6 and 7 are given with the error bars indicating the twice the standard deviation.

Inspection of the minimum basis NSDs shows that the individual pigments possess a variety of conformations ranging from relatively planar (BChl 1) to exhibition of a single dominant distortion (*e.g.*, the saddled BChl 7) or a mixed-mode conformation of two strong distortions (*e.g.*, the ruffled and domed BChl 6). The in-plane conformations also reflect the relative extent of distortion since the more distorted BChls 3, -4, -6 and -7 are significantly contracted compared to the other more planar BChls. Additionally, there are significant displacements and variances along the next-to-lowest energy normal-coordinates (not shown) and whilst these appear to correlate with the larger min. basis distortions (*e.g.*, BChls 3 and -4 have the highest degree of both first- and second order B_{1u} distortion) this is not an absolute rule.

Conformation of the EET exit pigment

The pigment that has been assigned to have the lowest-energy excitation (BChl 3) exhibits the joint-largest degree of ruffling (B_{1u}; cf. BChl 4) of all the chromophores. However, in contrast to BChl 4, the extent of the ruffling of BChl 3 is not approached by any other distortion mode exhibited by this pigment. The EET exit pigment also shows the largest degree of propellering (A_{1u}), which is the highest-energy distortion of the minimum basis, and as well has the largest positive degree of doming. Notably, of the seven coupled pigments, BChl 3 shows the least contribution of saddling (B_{2u}) to its conformation although the precise amount is close to that of BChls 1 and -6. As noted above, BChl 3 exhibits a significant A_{1g} contraction relative to the more planar cofactors in the system (typically around − 0.5 Å) and this pigment shows the least meso- and *N-stre*tching, indicating a relatively 'squarer' geometry (*i.e.*, D_{4h} for the planar projection). Notably, this pigment possess the largest $B_{1u}(2)$ and $A_{1u}(2)$ distortions (not shown), of which the former

has been directly linked to red-shifted Q-bands and the latter requires the largest deformation energy of the second-order modes.

Conformation of the eighth BChl

It has recently been suggested that the BChl that is found on the surface of the FMO protein provides the primary EET pathway from the chlorosome baseplate. In terms of its conformation, this pigment is particularly unique in that although it is relatively planar, it exhibits a highly mixed distortion pattern of approximately equal amounts of ruffling (B_{1u}), doming (A_{2u}) and both orientations of waving (approximately 0.2 Å of each). The overall relative planarity of BChl 8 is supported and emphasised by the fact that it exhibits the largest A_{1g} mode of any of the chromophores but unexpectedly it shares the relatively lesser elongation of BChl 3. In terms of its next-to-lowest energy distortions this pigment shows the largest contribution of the $A_{2u}(2)$ mode to its conformation but in spite of this shows only minor second-order deformations, comparable in magnitude though not shape, to BChl 2.

Conformation of the core EET entry pigment

Although the supposed proximity of BChl 8 to the baseplate together with its dominance of the highest energy exciton band of the FMO presupposes its role as the primary EET acceptor of the complex (see above), BChl 1 is believed to be strongly coupled to BChl 8 and furthermore, the BChl 1 dominated exciton band 3 has been calculated to have strong overlap with the baseplate's emission and so this cofactor has been suggested to offer an additional linkage for EET to the RC. Like the eighth pigment, the minimum basis NSD of BChl 1 suggests a relatively planar macrocycle with a similar magnitude of ruffling. However, BChl 1 is slightly, though significantly, more contracted than BChl 8 and this turns out to be in correspondence with the second-order NSD in which BChl 1 shows an interestingly large, in consideration of its first-order distortions, $B_{1u}(2)$ and $A_{1u}(2)$ deformations.

Conformations of the other FMO pigments

Having discussed the main features of the first- and second-order NSDs of the special pigments, BChls 3, -8 and -1, it remains to impress the rest of the conformational variability of the macrocycles present in this system. Of these remaining chromophores, BChl 5 appears to be the most planar overall as emphasised by the presence of only moderate saddling and the most expanded macrocycle, but in opposition to this description, is a very large degree of $E_g(x)(2)$. Next, although both the conformations of BChls 2 and -4 are dominated by saddling and ruffling, the relative magnitude of the distortion along each

are inverted between the two with *sad* > *ruf* in BChl 2 and vice-versa in BChl 3 and furthermore the phase inversion of the *sad* mode gives them very different stereochemistries. These two pigments are further differentiated by their next-to-lowest energy distortions with BChl 4 showing a far greater extent of $B_{2u}(2)$, $B_{1u}(2)$ and $A_{1u}(2)$. Finally, BChl 7 is considerably saddled, to the extent that it is the largest symmetric distortion possessed by any of the pigments (almost 0.8 Å) and is accompanied by a contraction comparable to any of the three highly ruffled BChls. It's second-order NSD further exaggerates this feature with the largest degree of $B_{2u}(2)$ distortion but this pigments conformation is also complemented by considerable $B_{1u}(2)$.

7.4 Initial Observations

7.4.1 Correlation of NSDs and Literature Site-energies

At the beginning of this project, it was considered useful to see if any of the NSD results obtained for the BChls from the highest resolution structure were correlated with state-of-the-art values of the site-energies published in the literature. The most striking observation to result from this simple analysis was the presence of very strong correlations between the $B_{1u}(2)$ normal-deformations of the BChls from the most recent *P. aestuarii* structure (PDB ID: 3EOJ) and the fitted site-energies from Adolphs *et al.*[270] (Figure 7-10). In particular, the potential significance of the correlation of the $B_{1u}(2)$ deformation is substantiated by the fact that this mode has been shown to be one of the primary causative agents of the red-shift found in *ruf* distorted porphyrins.[36] Moreover, the correlation behaves in the theoretically expected manner, as illustrated by the apparent quadratic relationship of the reference site-energy on the extent of distortion, in which a red-shift is observed with increasing nonplanarity with the maximum of the curve at approximately zero distortion (*i.e.*, the coefficient for linear term is not statistically significant). Notably, the $A_{1u}(2)$ (*i.e.*, the next-to-lowest energy A_{1u} distortion) mode was also highly correlated with these site-energies, and likewise, with the $B_{1u}(2)$ mode. This may be representative either of an inherent structural constraint or else an artefact from the refinement procedure.

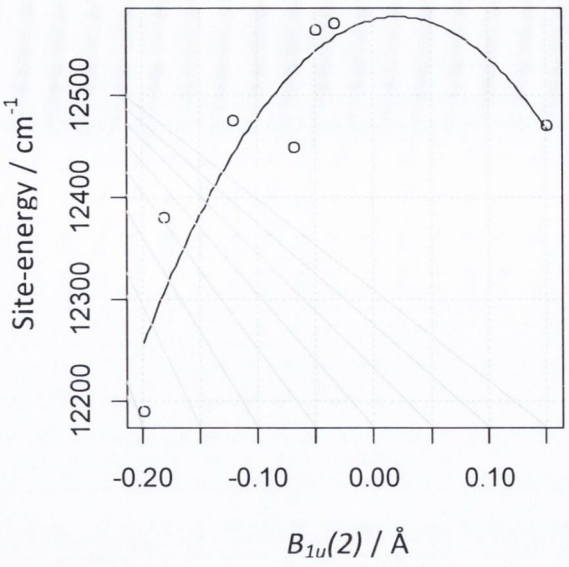


Figure 7-10: Correlation of the fitted site-energies from Adolphs *et al.*[270] with the $B_{1u}(2)$ deformation of the BChls as resolved in PDB ID: 3EOJ (averaged over the multiple conformers of BChls 6 and 7); the least-squares curve is -6575.1 x^2 + 290.5 x + 12574 with p-values of 0.025, 0.031 and <0.001, respectively.

7.4.2 Revisiting Gudowska-Nowak *et al.*

In light of the above observation that suggests a role for the macrocycle conformation in the modulation of the FMO site-energies, it was considered worthwhile to reassess the original results of Fajer and coworkers,[111] which model the effects of the BChl conformations directly, in terms of a comparison to some recent structure based calculations. Initially, it appeared that these results were in stark disagreement with Renger and coworkers'[269, 270] recent site-energies, that model the electrostatic influence of the protein at various levels and include a parameterised fit (*i.e.*, pseudo-experimental) (Table 7-6). However, it was noticed that their main point of contention was the site-energy of BChl 3, which the *in vacuo* ZIndo/S calculations by Fajer had found to have the highest site-energy whereas in modern studies, including spectral fitting techniques, it is found to possess the lowest energy.*

Interestingly, if this BChl is excluded from the comparison, the most complete BChl models used in the *in vacuo* calculations are brought into substantial agreement with recent results from the literature (Table 7-7; in particular, 'a_L' and 'b_L'). Moreover, these models give improved correlations with the fitted energies than the rougher approximations of the protein electrostatic effects afforded by the point charge-dipole (PCD) and charge density coupling (CDC) methods and only the Poisson-Boltzmann quantum chemistry technique (PBQC), which does include some explicit consideration of conformational effects,

* There is a high degree of consensus on this result and it is not questioned here.

maintains a very high correlation with the fit. This result suggests that it is perhaps only the site-energy of BChl 3 that is dominantly controlled by the pigment-protein electrostatic interaction and that those of the others are, potentially, significantly modulated by conformational differences.

Table 7-6: Correlation of all BChl site-energies (cm^{-1}) from Adolphs *et al.* and Gudowska-Nowak *et al.*[111].

	Fit[a]	PCD[a]	PBQC[a]	CDC[a]	a_L[b]	b_L[b]	b[b]	c[b]	d[b]	c_{H2}[b]	d_{H2}[b]
Fit		0.46	0.94	0.85	-0.36	-0.3	-0.35	-0.28	-0.22	-0.23	-0.17
PCD			0.24	0.68	-0.53	-0.42	-0.35	-0.25	-0.18	-0.2	-0.13
PBQC				0.79	-0.3	-0.27	-0.37	-0.31	-0.21	-0.19	-0.12
CDC					-0.28	-0.2	-0.3	-0.19	-0.13	-0.06	0.00
a_L						0.99	0.92	0.89	0.78	0.82	0.72
b_L							0.96	0.95	0.85	0.89	0.80
b								0.99	0.91	0.91	0.84
c									0.95	0.94	0.89
d										0.96	0.97
c_{H2}											0.98
d_{H2}											

[a] Referring to site-energies from Adolphs *et al.*[270] a parameterised fit to the experimental spectra (Fit), the point charge-dipole (PCD), the Possion-Boltzmann quantum chemistry (PBQC) and charge density coupling (CDC) calculations; [b] These codes denote the structural models of the FMO BChls used in Gudowska-Nowak *et al.*'s[111] ZIndo/S calculations; a = full BChl a excluding phytyl, b = model 'a' excluding C8-ethyl and C17-propionic acid, c = model 'b' excluding β-methyl groups and $C13^2$-methylester, d = model 'c' excluding C3-acetyl, subscript L = axial imidazole, subscript H2 = free-base.

Table 7-7: Correlation of BChl site-energies (cm^{-1}) from Adolphs *et al.* and Gudowska-Nowak *et al.*[111] after excluding BChl 3.

	Fit[a]	PCD[a]	PBQC[a]	CDC[a]	a_L[b]	b_L[b]	b[b]	c[b]	d[b]	c_{H2}[b]	d_{H2}[b]
Fit		-0.28	0.80	0.49	0.64	0.59	0.49	0.42	0.21	0.21	0.11
PCD			-0.73	0.31	-0.13	-0.01	0.11	0.17	0.09	0.07	0.07
PBQC				0.34	0.65	0.54	0.34	0.28	0.19	0.24	0.17
CDC					0.74	0.72	0.53	0.54	0.36	0.51	0.41
a_L						0.98	0.87	0.86	0.77	0.83	0.77
b_L							0.94	0.93	0.85	0.89	0.84
b								1.00	0.94	0.92	0.9
c									0.96	0.94	0.92
d										0.96	0.97
c_{H2}											0.99
d_{H2}											

*See footnote in Table 7-6.

Additionally, the correlations between the site-energies calculated by Gudowska-Nowak *et al.*[111] with the corresponding NSDs of their models indicates a complex interaction between the structural model employed and the effects of particular aspects of the BChl conformations (Table 7-8 and Table 7-9). For example, the correlation of the *bre* mode (A_{1g}) in the most complete structural models is such that a red-shifted excitation is associated with an expansion of the core, but this correlation gradually decreases upon sequential truncation of the model until it is inverted so that expansion of the core induces

a blue-shifted Q_y energy. Other examples are apparent, though not as drastic, the influence of the *sad* distortion (B_{2u}) is greatest in the fully truncated model that includes the Mg whereas the *ruf* mode's effect (B_{1u}) increases as axial ligands are removed and the periphery is truncated; both of these modes display the expected negative correlations with the site-energies, which were correlated as wavenumbers. The parameters representing the total in- and out-of-plane distortions (D_{ip} and D_{oop}, respectively) are also revealing, these indicate that in-plane distortion is more important in the models that retain significant substitution whereas the converse is true for the for the total nonplanarity (Table 7-9). Taken together, these results show that the conformational effect on the excitation energy is modulated by the presence of the substituents; this suggests that truncated models are not appropriate to assess the conformational influence upon this property.

Table 7-8: Correlation matrix of the min. basis NSD deformations from the BChls in PDB ID: 3BCL with the excitation energies (in cm^{-1}) for the various models from Gudowska-Nowak *et al.*[111]

	B_{2g}	B_{1g}	$E_u(x)$	$E_u(y)$	A_{1g}	A_{2g}	B_{2u}	B_{1u}	A_{2u}	$E_g(x)$	$E_g(y)$	A_{1u}
a_L	0.47	0.11	0.5	0.79	-0.28	-0.56	-0.57	0.04	0.28	0.59	-0.46	-0.09
b_L	0.33	0.08	0.53	0.82	-0.25	-0.58	-0.62	-0.08	0.20	0.56	-0.46	-0.24
b	0.17	0.05	0.52	0.8	-0.20	-0.58	-0.74	-0.12	0.16	0.51	-0.4	-0.36
c	0.08	0.08	0.50	0.79	-0.17	-0.57	-0.76	-0.19	0.09	0.52	-0.45	-0.46
d	0.01	0.19	0.38	0.58	0.00	-0.67	-0.87	-0.16	0.02	0.50	-0.61	-0.59
c_{H2}	0.04	0.16	0.34	0.67	0.13	-0.75	-0.74	-0.19	0.21	0.48	-0.63	-0.59
d_{H2}	-0.04	0.2	0.28	0.54	0.21	-0.76	-0.78	-0.21	0.14	0.44	-0.68	-0.67

*See footnote in Table 7-6.

Table 7-9: Correlation matrix of the min. basis NSD overall parameters from the BChls in PDB ID: 3BCL with the excitation energies (in cm^{-1}) for the various models from Gudowska-Nowak *et al.*[111]

	D_{ip}	δ_{ip}	D_{oop}	δ_{oop}
a_L	-0.67	-0.42	-0.48	0.09
b_L	-0.69	-0.34	-0.60	-0.04
b	-0.69	-0.30	-0.59	-0.18
c	-0.68	-0.20	-0.69	-0.26
d	-0.56	-0.06	-0.77	-0.28
c_{H2}	-0.44	-0.21	-0.78	-0.14
d_{H2}	-0.37	-0.10	-0.81	-0.20

There is yet another, more fundamental, aspect to these observations with respect to how the absorption properties of Chls and BChls have been tailored for use in photosynthesis. It is known that the incorporation of Mg(II) into porphyrin compounds in general affords a red-shifted absorption relative to similar metalloporphyrins and furthermore, substitution of π-extending substituents acts similarly;[29] these are some of the reasons why these compounds have been selected to function as chromophores.[103, 104] However, in light of

these observations, these structural features also modulate the effect of the macrocycle conformation on these properties, which in turn may allow the protein to exert an enhanced *fine-tuning* of the absorptions *via* conformational control.

7.5 C3-Acetyl Relaxed RexEPs

The PM6//ZIndo/S acetyl dihedral rotational excitation energy profiles (RexEPs) calculated for the BChls from the same structure used in the original calculations of Gudowska-Nowak et al. are illustrated in Figure 7 14. This approach was performed in a sequence of increasing complexity so that if coalescence of the RexEPs did appear, it could be attributed to a specific set of parameters. Specifically, first the acetyl methyl H-atom dihedrals were relaxed to account for the steric influence of nearest neighbours which had only a very minor effect. This second step involved relaxation of all internal coordinates of the acetyl FG, including the carbonyl carbon-macrocycle beta carbon bond length. Additionally, as a final test, all H-atoms were allowed to relax.

As most of the RexEPs from the different BChls remain vertically separated at all levels of relaxation, it may be concluded that differences in the internal conformation of the acetyl group are not, on the whole, responsible for the apparent 'nonadditivity' of the conformational differences. In detail, the RexEPs appear to form a few groups so that in terms of their wavelengths BChl $4 \approx 7 > 6 > 5 \geq 2 \geq 3 > 1$, where the "$\geq$" sign indicates the positional variation of BChl 2 across the relaxation series. Thus, the BChl conformations given in the 3BCL structure, suggest that differences between the acetyl dihedrals account for the BChl 4 and 7 site-energies from this model, whereas the BChl 2 difference from BChls 5 and 3 is accounted for by both the acetyl dihedral angle and its other conformational features.

There was still the possibility it was not the macrocycle conformation but the conformations of the other peripheral substituents included in the model that were the origin of the remaining variation. Rather than calculating RexEPs with further increased relaxation or directly assessing the influence of other conformational coordinates (*e.g.*, the planarity of the $C3^1$-carbonyl) it was decided at this point to check whether the BChl NSDs were correlated with the properties of the individual RexEPs. Initially, this was achieved by clustering the RexEPs and comparing those clusters to those found by clustering the NSDs. The excellent agreement between the RexEP cluster hierarchy and that afforded by the BChl NSDs, suggests that the macrocycle deformations are indeed responsible for the remaining variation (Figure 7-12). In this case, the good agreement was obtained when at

least the first three lowest-energy modes of each symmetry were clustered, indicating that these modes contribute to the calculated profile.*

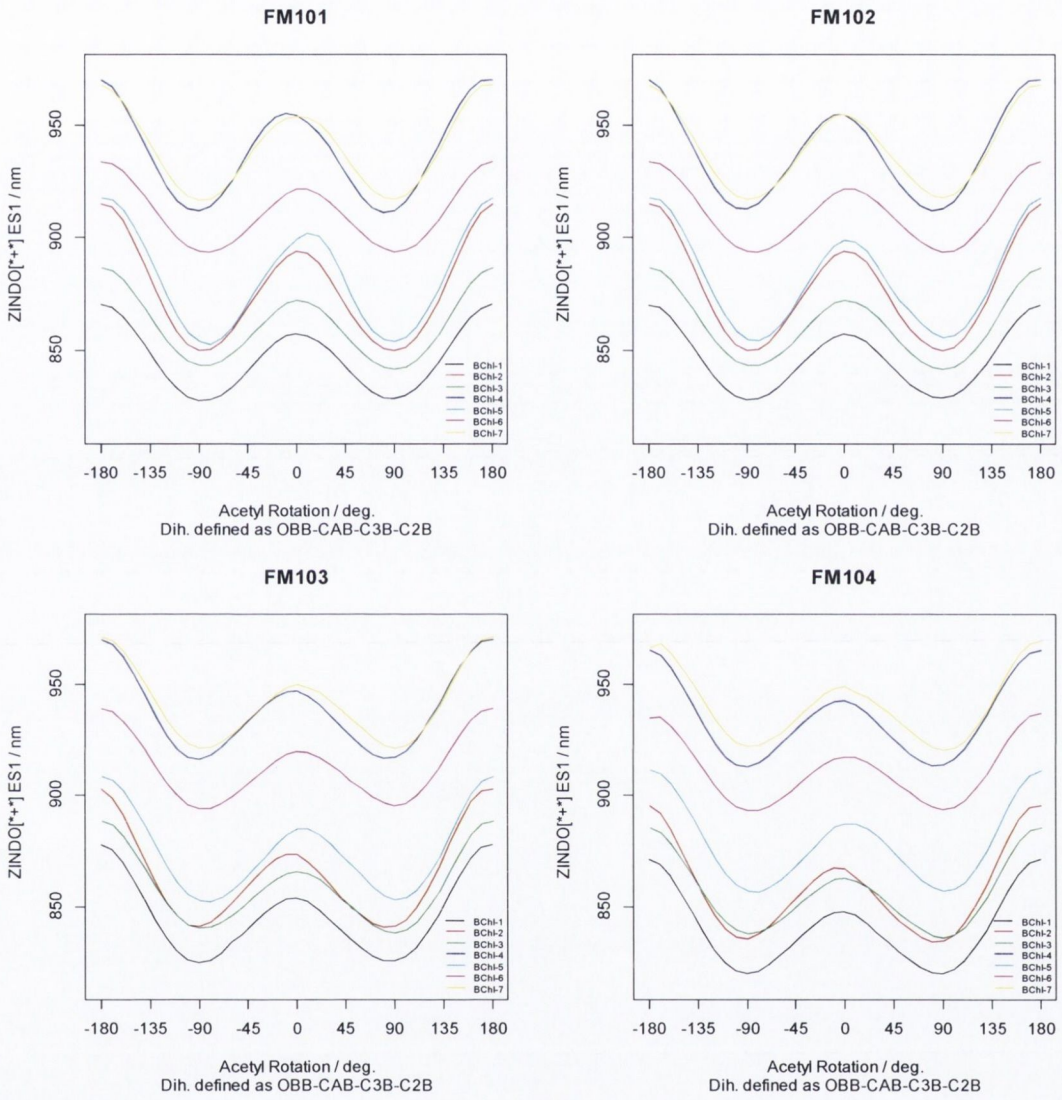


Figure 7-11: PM6//ZIndo/S RexEPs of the BChls from PDB ID: 3BCL with various levels of relaxation; rigid rotation (top left), acetyl-Me hydrogen dihedral relaxation (top right), all acetyl relaxation including the $C3^1$-C3 bond-length (bottom left) and as previously stated with the addition that all macrocycle H-atoms were relaxed. BChls 1 to 7 are represented by black, red, green, blue, light blue, pink and yellow, respectively.

Lastly, in an attempt to uncover the precise features of the macrocycle conformation that determine specific aspects of the RexEPs, the correlations between the BChl NSDs and the individual parameters of the corresponding RexEPs, as given by a parameterised fit (Table 7-10), were assessed. An interesting revelation from this approach was that the normal-deformations that are symmetric with respect to the (idealised) macrocycle *C4*-axis (*i.e.*,

* Note that this is not an assertion that these modes were reliably determined by the crystal structure. In this context, however, the NSDs are exact as the structural models for the calculations were taken exactly as they appear in the crystal structures.

the modes of '*A*' symmetry) were the modes most highly correlated with the form of the RexEP, whereas distortions of lower-symmetry were associated with the position of the RexEP (Table 7-11).

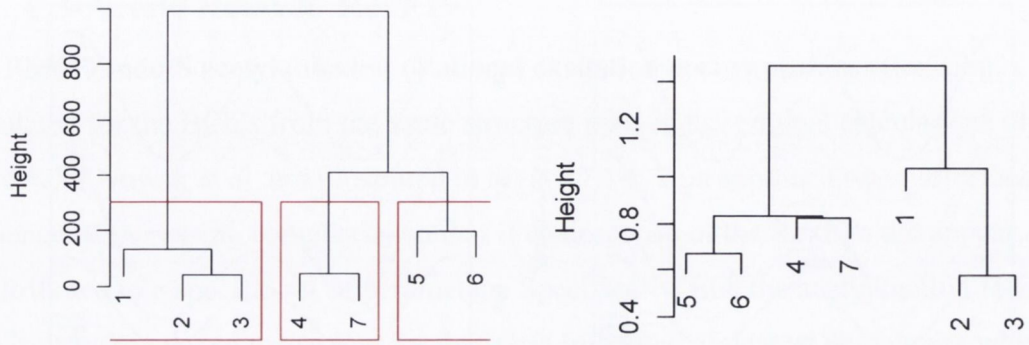


Figure 7-12: Cluster dendrogram of the ZIndo/S RexEPs from series FM104 (left) and the NSD data from the 3BCL pigments (right); the AHC was performed using the squared Euclidean distances and Ward's method.

Table 7-10: Coefficients for the ZIndo/S RexEPs calculated for the BChl pigments as given by PDB ID: 3EOJ; fitted to the equation, $\lambda = a\cos 2\theta + b\,\theta^2 + c$.

BChl	Intercept	cos 2θ	θ^2
1	831	17.0	2.4
2	847	19.6	2.9
3	847	16.1	2.3
4	925	17.6	2.4
5	869	18.6	2.5
6	904	14.7	1.9
7	933	16.6	2.1

Table 7-11: Correlation coefficients of total distortions of each symmetry with fit parameters of FM104 ZIndo/S RexEPs.

	$B_{2g}{}^c$	$B_{1g}{}^c$	$E_u(x)^c$	$E_u(y)^c$	$A_{1g}{}^c$	$A_{2g}{}^c$	$B_{2u}{}^c$	$B_{1u}{}^c$	$A_{2u}{}^c$	$E_g(x)^c$	$E_g(y)^c$	$A_{1u}{}^c$
c	0.078	0.301	0.439	-0.064	-0.036	-0.277	0.794	0.039	0.399	0.46	0.738	-0.006
cos2θ	0.154	-0.357	0.036	0.391	0.608	0.905	0.174	0.084	-0.821	-0.006	-0.164	-0.591
θ^2	0.143	-0.491	-0.06	0.553	0.484	0.906	-0.053	0.227	-0.835	-0.257	-0.212	-0.447

7.6 Full-factorial Partial Optimisations

7.6.1 General Interpretation

In general, a single-point energy (SPE) calculation upon the crystal geometry of the cofactor $E(BChl)_{Cryst}$ may be decomposed as a sum of contributions from the energy of the unperturbed BChl and the strain energy induced by the protein environment encapsulated in the crystal structure conformation (Equation 7-2).

Equation 7-2: Representation of the SPE of a pigment frozen in the geometry obtained from the crystal structure, where E(BChl) signifies the unperturbed BChl energy and E(Strain) the conformational strain.

$$E(BChl)_{Cryst} = E(BChl) + E(Strain)$$

On the other hand, since the optimisations are carried out in the absence of the protein environment, the SPE of the fully optimised structure, $E(BChl)_{Opt}$, is equivalent solely to the unperturbed BChl energy, at the given level of theory (Equation 7-3).

Equation 7-3: Identification of the SPE of the optimised pigment geometry as the unperturbed BChl energy.

$$E(BChl)_{Opt} = E(BChl)$$

Clearly then, the difference between these calculations yields the total strain energy induced by the protein that is encapsulated within the pigment's structure (Equation 7-4).

Equation 7-4: Equivalence of the protein-induced conformational strain energy with the difference of the SPE of a pigment frozen in the geometry obtained from the crystal structure and that of the optimised geometry.

$$E(BChl)_{Cryst} - E(BChl)_{Opt} = E(Strain)$$

Furthermore, this line of thought may be extended to the results of the individual *partial optimisations* and reveals that the strain energy of the optimised molecular fragment is similarly found (Equation 7-5 and Equation 7-6).

Equation 7-5: Explicit interpretation of the result of a SPE on a partially optimised structure.

$$E(BChl)_{Opt\ Frag\ X} = E(BChl) + E(Strain) - E(Strain)_{Frag\ X}$$

Equation 7-6: Equivalence of the protein-induced conformational strain energy imparted *via* perturbation of a single molecular fragment with the difference of the SPE of a pigment frozen in the geometry obtained from the crystal structure and that of the partially optimised geometry.

$$E(BChl)_{Cryst} - E(BChl)_{Opt\ Frag\ X} = E(Strain)_{Frag\ X}$$

Likewise, these arguments are valid for the excited state energies (E_{ES1}) calculated with the ZIndo/S method except that the strain energies are interpreted as the E_{ES1} perturbation of the (fragment) conformation.

Effects of Interactions

Given that the conformations of the molecular fragments of the BChl are very likely to interact with each other (see below; this is the very reason for using this design) the difference method to obtain the "intrinsic strain energy" requires more than just the two calculations (or three for a percentage) described above to get an accurate determination. For instance, in a Mg-porphyrin model if the centrally coordinated Mg is frozen in the crystal geometry and is above the 4N-plane, then a partial optimisation of the macrocycle experiences an indirect perturbation from the protein environment propagated by the Mg (*e.g.*, it may still display *dom* distortion after optimisation). This means whilst Equation 7-6 still holds formally, the *identification* of the r.h.s. term as the protein-induced

conformational strain energy imparted *via* perturbation of the optimised molecular fragment, is not strictly true as there is in fact a remnant of the protein strain propagated by the Mg atoms position. This effect could be included specifically as is done so in Equation 7-7 and points towards it solution; two more calculations should be performed where the Mg is also optimised alongside the macrocycle (Table 7-12).

Equation 7-7: Equivalence of the protein-induced conformational strain energy imparted *via* perturbation of a single molecular fragment *including the indirect effect of the fixed geometry of another molecular fragment*, with the difference of the SPE of a pigment frozen in the geometry obtained from the crystal structure and that of the partially optimised geometry.

$$E(BChl)_{Cryst} - E(BChl)_{Opt\ Frag\ X} = E(Strain)_{Frag\ X} + E(Strain)_{Frag\ Y\ on\ X}$$

The series of calculations presented in Table 7-12 in fact corresponds to the full-factorial design for a two fragment (*e.g.*, Mg and macrocycle) by two level (*i.e.*, frozen or optimised) experiment. Whilst the direct physical interpretation of each result shows clearly that the individual contributions of the fragments or each '$E(Strain)_{Frag\ X}$' are obtainable from this series since essentially the results provide four equations with four unknowns, intuitively it may be observed that *the effect of each fragment is assessed at each state of the other*. Notably, in order for this system to be determined the interaction terms must be allowed to be equivalent so that it is not possible to discern the 'effect of the experimental position of the Mg on the macrocycle', from the converse 'effect of the macrocycle on the Mg', even though they could be physically different.

Table 7-12: Interpretations of the results of the experimental units in the full-factorial design to determine the separate effects of two fragments on the strain energy; E.U. denotes experimental unit.

E.U.	Frag. X	Frag. Y	Result
0,0	Frozen	Frozen	$E(BChl) + E(Strain)$
1,0	Optimised	Frozen	$E(BChl) + E(Strain) - E(Strain)_{Frag\ X} + E(Strain)_{Frag\ Y\ on\ X}$
0,1	Frozen	Optimised	$E(BChl) + E(Strain) - E(Strain)_{Frag\ Y} + E(Strain)_{Frag\ X\ on\ Y}$
1,1	Optimised	Optimised	$E(BChl)$

The parameter of interest, aptly named the *main effect*, is found by the comparing the means of all runs with the fragment of interest frozen with those where it is relaxed (the grand mean may also be used as a reference). Thus it is these main effects that may be interpreted as the *protein induced conformational effect of the specified fragment(s)* on the SCF energy or equivalently E_{ES1} (*i.e.*, the unperturbed site-energy).

Statistically significant interactions signify a physical interaction between the geometries of the relevant substituents and/or the electronic structure. With respect to the macrocycle

conformation, geometric detail is provided by the NSDs of the structures that have (at least) some or all core macrocycle skeleton atoms relaxed. An 'exotic' example could be that the change in the degree of waving may be greater for those in which the acetyl FG is allowed to rotate whilst something more conventional would predict a dependence on whether or not the Mg and/or ring E are allowed to relax too.

7.6.2 Model A

Variation across All Crystal Structures

As it has not yet been determined if a single crystal structure best represents the actual conformations of the BChl pigments in the FMO protein, or whether it is appropriate to average the conformations over multiple structures, the full factorial series of partial optimisations defined by model A was conducted for all of the available structures. The calculated Q_y excitation energies for the crystal structure geometries, where only the hydrogen atoms were optimised, together with those calculated for the fully optimised structures are shown in Figure 7-13. There is substantial variation between the Q_y energies for the different BChls frozen in the experimental geometry from different structures. In detail, the within structure variation is greatest for 3BSD where the calculated energies range over ~180 nm and least for 3ENI, 3EOJ and 3VDI where this range is ~14 – 35 nm (Table 7-13). In contrast and as expected, the calculated Q_y energies for the fully optimised structures are on the whole homogeneous such that the range over all BChls from all structures is 667 – 692 nm with a standard deviation of 5 nm; this homogeneity serves to validate the comparisons of the substituent effects that are discussed next. Note also, that it is not surprising that the optimised model energies are substantially blue-shifted compared to those afforded by the crystal geometry, indicating a stabilised ground state relative to the excited state, which is the result of performing the ground state optimisations.

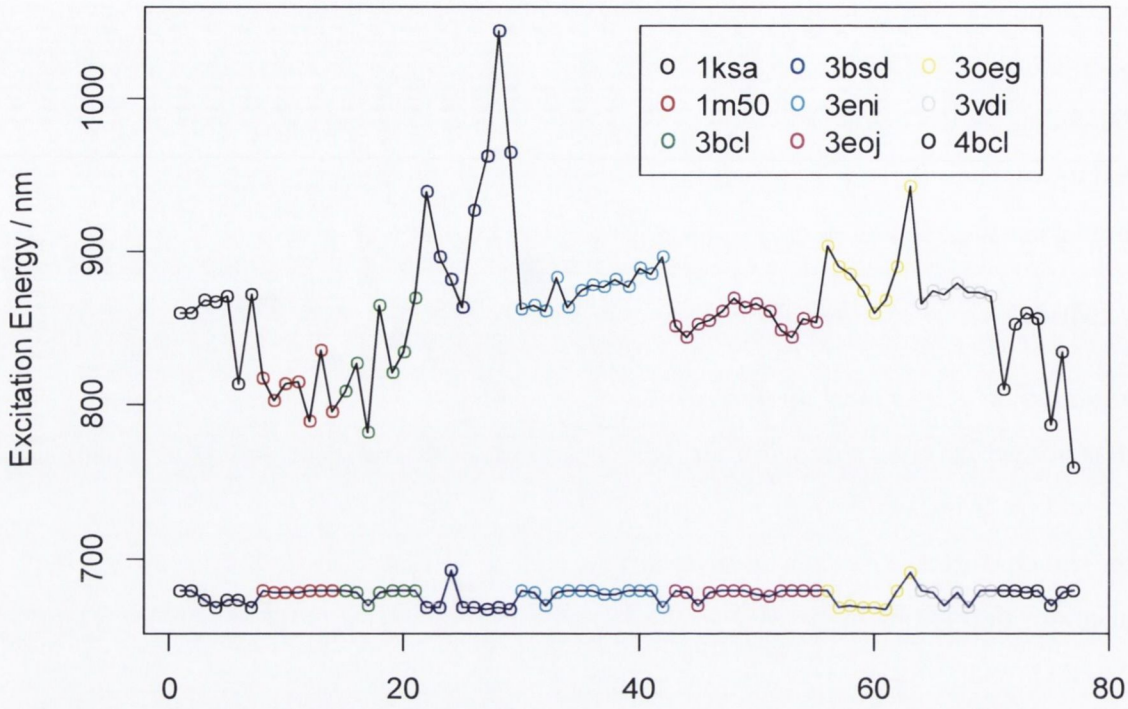


Figure 7-13: The excitation energies for each FMO pigment calculated from the crystal structure geometry (black line, top) and after full optimisation of this starting geometry (blue line, bottom). The results for each individual pigment are given for nine separate PDB structures (indicated by the colour of the data points) and include multiple conformers where present.

Table 7-13: Range and standard deviations illustrating the variation of the excitation energies of the seven FMO BChls calculated from the experimental geometries provided by each crystal structure (*i.e.*, E_{cryst}, n = 77; nm).

PDB ID	E_{cryst} Min.	E_{cryst} Max	E_{cryst} SD
1KSA	813.8	871.7	20.5
1M50	789.9	835.3	15.2
3BCL	782.7	869.1	30.3
3BSD	863.3	1043.8	57.3
3ENI	860.6	895.8	11.0
3EOJ	844.1	869.4	7.9
3OEG	860.4	942.3	25.3
3VDI	865.6	879.6	4.1
4BCL	759.0	859.5	38.7

The summary statistics for the substituent main effects, calculated for each BChl from each crystal structure, over the whole series are shown in Table 7-14. Clearly, the general result of optimising any single substituent is to blue-shift the calculated Q_y energy relative to the crystal geometries', as indicated by the averages of the main effects as well as by the previously discussed results. However, the maximum values obtained from the full set of calculations indicate that it is possible for the crystal structure geometry, which within the limits of experimental error describes the effect of the protein environment, to impart a blue-shifted Q_y energy, corresponding to an observed red-shift upon optimisation (*i.e.*, positive main effects; Table 7-14). This latter observation is most applicable to the effects

of the Mg position and acetyl (Ace.) conformation, which show the largest positive maximums. This aspect of the effect of the Mg atom's position may be interpretable on chemical grounds since it is known that the spectra of Mg porphyrins and related compounds may be red-shifted in coordinating solvents,[29] and these results suggest that part of that red-shift could be due to the out-of-plane displacement of the Mg in the 5-coordinate species.* With respect to the potential for the acetyl to induce a blue-shift, this too has been observed and results from the protein-environment forcing this substituents out-of-the plane of the macrocycle and additionally the same observation has been made regarding the effect of the cyclopentanone ring.[396]

Table 7-14: Descriptive statistics illustrating the spread and central tendencies of the excitation energies obtained from the BChls as found in the crystal structure and after full optimisation (E_{Cryst} and E_{Opt}, respectively) and the calculated fragment (atom group) main effects for all calculations (*i.e.*, all BChls from all FMO crystal structures listed in Table 7-13, n = 77; all in nm).

	E_{Cryst}	E_{Opt}	Intercept	Mac.	CP	Ace.	Alks.	Mg
Min.	759	667	754	-311.4	-51.7	-37.5	-26.7	-10.3
1st Quartile	849	670	853	-180.6	-4.4	-16	-11.4	-6.9
Median	865	678	867	-167.3	-0.8	-8.9	-6	-3.7
Mean	863	676	869	-162.4	-7.35	-9.34	-7.56	-3.41
s.d.	42.5	5.36	45.0	37.7	14.7	8.97	6.59	4.09
3rd Quartile	877	679	885	-144.3	0.6	-3.3	-2.8	-0.7
Max.	1044	692	1058	-66.2	4.1	9.9	3.9	15.2

A striking feature of these results is that the optimisation of the macrocycle conformation single-handedly swamps the effects of the other molecular fragments, with this value averaging at -162 nm whilst the next highest average main effect is for the acetyl at -9.34 nm (see Mac. column in Table 7-14). However, it would be inappropriate to conclude from this that the influence of the macrocycle conformation has been proven to render superfluous all others, however tempting it may be. One reason for this is that there is an inherent bias in the experimental design; there are far more atoms in the macrocycle group than any other. Indeed, on a per atom basis, the main effects for the macrocycle group are ~6.8 nm per atom and the acetyl is ~3.1 nm, which although not as drastic, is still suggestive of a greater influence of the macrocycle. The meaning of this comparison comes from the fact that the number of atoms in the group specification is directly proportional to the number of free parameters afforded in their optimisation. Another factor that encourages restraint regarding over interpretation of this result is that all 24

* The geometric parameters from the optimised geometries have not yet been compiled so this is somewhat speculative. However, it would be highly unlikely that the crystal geometry would have the Mg resting in the 4N-plane and for the subsequent optimisation to produce an out-of-plane displacement.

atoms either are, or in the case of saturated C_bs are bonded to, the sp^2 hybridised framework of the π-system. This means that any inherent bias in the PM6 method towards the geometries of such bonding schemes will result in a particular sensitivity of this value to the corresponding target molecular specifications that were used in the structure determination. Conversely, however, the very same argument that these atoms represent the π-system also lends credence to the result since it is perfectly reasonable then that perturbations to their geometry would have such large effects. In any case, it is important to recall what these individual experiments do: they measure the contributing conformational factors that result in differences between the experimental geometry and the PM6 *in vacuo* geometry at the ZIndo/S (15,15) level of theory. At this level, the relative differences between the main effects are of primary interest and although this is best done explicitly, if one looks at the variation of the fragment main effects it is apparent that the effect of the macrocycle also varies considerably more than those of the substituents, but not so much as the difference in the means.

With respect to how well the main effects of the substituents account for the variation of the calculated Q_y energies of the experimental geometries across all BChls form all structures, linear models that contained only these effects provided very good agreement with the full range of calculated energies (Table 7-15). Ordinary least-squares (OLS) fitting and robust linear modelling (RLM) were used to regress the individual BChls' substituent effects (Table 7-15) against the Q_y energies calculated for the experimental geometry (Figure 7-13 and E_{cryst} in Table 7-14). In this model, E_{cryst} is modelled as a weighted sum of the atom group main effects plus a constant so that if the effect is a strong determinant of the total variation then a coefficient close to negative one is expected.* Additionally, since the atom group effects describe the protein induced perturbation of the unstrained geometry, the intercept should be close to the average excitation energies calculated for the fully optimised BChls (E_{Opt} in Table 7-14).

In general, the high R^2 value indicates that these main effects alone account for ~97% of the total variation, implying that from this view, substituent interactions are relatively

---

* The sign should be negative because the regression is of the experimental geometry against the main effects, *i.e.*, a mapping from the optimised Qy energy to that calculated for the experimental geometry. In contrast, to get the main effects, the experimental geometries had effect code of '-1' so that the main effects of optimisation (effect code +1) map the experimental geometries to the optimised geometries and hence the sign reversal.

minor.* As for individual effects, since the macrocycle main effects vary the most (Table 7-14) it is not surprising that its main effect appears to be the strongest and most consistent conformational influence upon E_{cryst} (*i.e.*, $\beta_{Mac.}$ is closest to -1; Table 7-15). This is followed by either by the cyclopentanone and its methyl substituent (CP) or the acetyl group depending on whether OLS or RLM are used to perform the regression. In this respect, preference should be given to the RLM result as this is resilient with respect to outliers, which the minimum value of -51.7 nm for CP atom group clearly qualifies as (Table 7-14). These values are almost matched by the aliphatic substituents (Alks.), which indicates that non-conjugated substituents may also have considerable effect on the site-energies, in line with the recent conclusion of Matysik and coworkers,[397] who suggest that differential alkyl inductive effects in each monomer of an RC special-pair to contribute to charge-separation asymmetry.

In terms of the most unusual result, the coefficient for the Mg main effect is positive in the OLM and negative in the RLM, the latter being closer to what it should be, indicating a severe problem with outliers or non-normality. The offending structure was identified to be the BChl 8 from 3OEG. This is an interesting result as the presence of this pigment has since been retracted by the authors after they reinterpreted the density features as belonging to a disordered PEG molecule. Removal of this structure from the analysis significantly improves the agreement between the OLS and RLM models, but regression diagnostics further indicate that there is still room for improvement (not shown).

Table 7-15: Resulting parameters of linear models obtained from ordinary least-squares (OLS) and iteratively reweighted least-squares (IWLS; less sensitive to outliers) relating E_{ES1} from the crystal geometry to the calculated main effects of the molecular fragments ($E_{cryst} = \beta_{Mac}x_{Mac} + \cdots + \beta_{Mg}x_{Mg} + c$ where $x_{Fragment}$ denotes the calculated main effect of the named fragment; n = 77).

	Intercept	$\beta_{Mac.}$	β_{CP}	$\beta_{Ace.}$	$\beta_{Alks.}$	β_{Mg}	R^2
OLS	695	-0.933	-0.855	-0.780	-0.680	0.531	0.973
IWLS	687	-0.956	-0.709	-0.897	-0.731	-0.417	NA
OLS[a]	687	-0.952	-0.788	-0.918	-0.734	-0.337	0.984
IWLS[a]	686	-0.958	-0.685	-0.927	-0.737	-0.611	NA

[a] Excluding BChl 3OEG A408, which has an Mg main effect of 15.2 nm whilst the next highest is 2.1 nm.

Before moving on to the direct comparison of some individual BChls, there is one more general point worth highlighting, namely, that the variation in the macrocycle main effect term alone accounts for 86% of the total variation over all BChls and all structures (Figure 7-14). Also apparent from this plot is that a lot of the within structure variation is

* That is to say that the moderating effects of their relative conformations are minor, their presence in the structural model of BChl a is likely essential.

also mainly due to the different macrocycle conformations. This is most obvious for the high-variance structures such as 3BSD and 3OEG, but on closer inspection, is true also for the lower variance structures such as 3EOJ and 3ENI. This suggests that in terms of the PM6//ZIndo/S method and these structural coordinates, macrocycle conformational flexibility will be found to significantly affect the site-energy distribution.

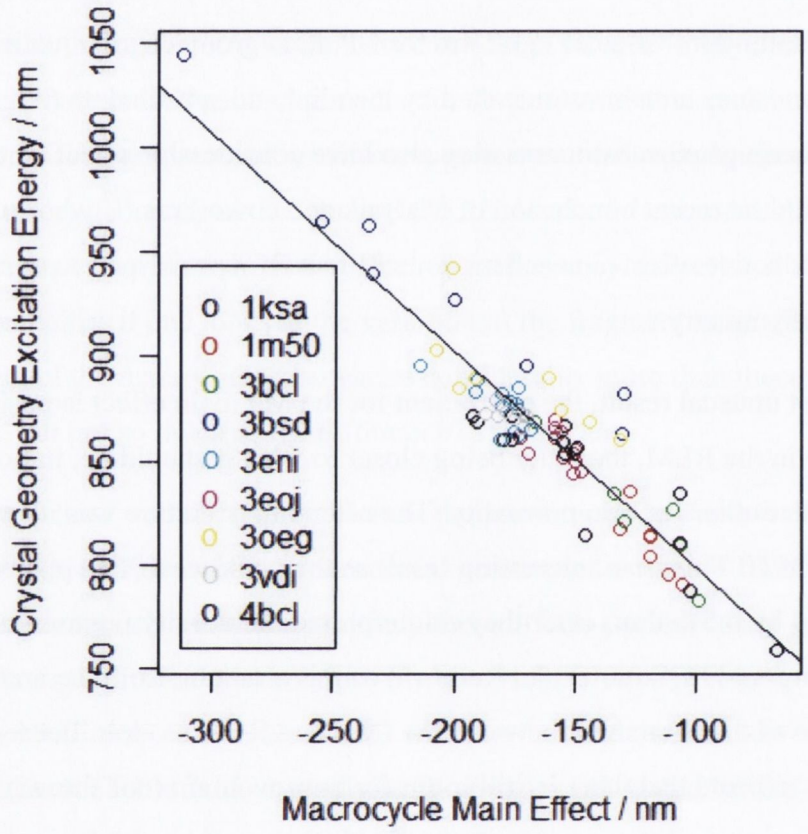


Figure 7-14: The calculated site-energies for all pigments from all structures in the crystal geometries *vs.* the estimated main effect of the macrocycle conformation. Fitted line: $y = 694 - 1.05x$, $R^2 = 0.86$.

Inter-site Differences: The Entry and Exit Pigments

In order to contextualise the calculated atom group effects they are now employed to characterise the conformational contributions to one of the important inter-BChl site-energy differences, namely that between BChls 3 and 1, which are considered to be the EET exit and entry pigments, respectively.[264] Recent estimates suggest that these BChls differ in site-energy by ~14 – 20 nm (*i.e.*, for BChl 3 – 1) for *P. aestuarii*,[264, 269, 270] and this is similar to the result explicitly calculated for *C. tepidium* of 25 nm,[398] therefore the midpoint of 17 nm for the BChl 3 *vs.* 1 difference will be considered a reference by which to contextualise the following results and is considered to represent the combined effects of all site-energy perturbations. Additionally, since it has not yet been ascertained what structures best represent the BChl conformations, the results as given by the models

derived from the four most recent high-resolution determinations, afforded by structures 3EOJ, 3ENI, 3VDI and 3BSD, will be presented. These structures represent all of species currently available, with a redundancy of two structures for *C. tepidium* (Table 7-3).

The differences between the BChl 3 and BChl 1 Qy energies calculated for the experimental and fully optimised geometries are shown in Table 7-16. It is immediately apparent that the 3BSD BChl conformations are quite unique by displaying a very large blue-shift of BChl 3 relative to BChl 1 of 57 nm, suggesting an anti-synergism of the conformational effect with respect to the reference site-energies. In contrast, the remaining structures show relatively small differences of -1.5 to 6.1 nm. Moreover, the results obtained from 3VDI and 3EOJ, the most recent and highest resolution structures respectively, suggest a synergistic effect of the conformational influences, with the site-energy as given by the reference. Another striking feature is that for structures 3EOJ, 3ENI, and 3VDI, the difference between the fully optimised BChl 3 and the equivalent BChl 1 structure is larger in magnitude than in the experimental conformations; moreover, these differences are identical across the three cases, suggesting that BChls 3 and 1 are optimised to corresponding local minima.* This suggests that there is a degree of similarity with respect to the corresponding BChls experimental geometries on the PM6 potential energy surface. The 3BSD difference on the other hand is less than half of that in the fully optimised structures in terms of magnitude, but has become positive.

Table 7-16: BChl 3 minus BChl 1for E_{cryst} and E_{Opt} together with the difference in E_{cryst} expressed as a percentage of the difference of the reference 'experimental' site-energies for these BChls (see text for discussion of the reference value) for selected PDB structures.

	3EOJ	3ENI	3VDI	3BSD
E_{cryst}	2.0	-1.5	6.1	-57.0
% Ref[a]	12	-9	36	-335
E_{opt}	-9.1	-9.1	-9.1	24.2

[a] E_{cryst} as a percentage of the reference value; 17 nm.

The next consideration is how well the substituent effect coefficients simulate these differences. Since in a single structure, the substituent effects model the difference between the experimental and fully optimised geometries, or E_{strain} (see Equation 7-4, p. 181), the same difference is modelled by the differences in the substituent effects between the two BChls. The results in Table 7-17 thus compare the summation of the BChl 3 minus BChl 1 substituent effect differences, to the difference between the calculated E_{strain}s (*i.e.*, E_{cryst} -

* Note that this is assumed, the structural data have not yet been assessed, although given the uniformity this assumption is probably accurate.

E_{opt}) for BChl 3 and BChl 1. In contrast to the global situation where the majority of the all BChl all structure data was well modelled using only the main substituent effects, it is evident that higher-order conformational and stereo-electronic interactions of the atom groups account for a significant proportion of this particular inter-site difference. This is illustrated by the fact that 3-way interactions* are necessary to get reasonable simulations in terms of the relative error, whilst 5-way interactions are necessary to get very accurate simulations (Table 7-17).† However, it would be very difficult to interpret such high-order interactions at this stage, therefore, the discussion will be largely confined to the main effects and interaction coefficients obtained from the 2-way models.

Table 7-17: Difference between the calculated E_{strain} for BChl 3 and BChl 1(*i.e.*, [BChl 3: $E_{cryst} - E_{opt}$] – [BChl 1: $E_{cryst} - E_{opt}$]) compared to the value given by the simulation from the effect coefficients.

		$\Delta(E_{Cryst} - E_{Opt})$	Mains[a]	3-way[b]	5-way[c]
3EOJ		-11.0	-15.0	-10.2	-11.0
	Error %		-36.1	7.4	0.2
3ENI		-7.6	-14.5	-7.1	-7.6
	Error %		-91.8	6.1	-0.5
3VDI		-15.2	-19.5	-15.0	-15.2
	Error %		-28.5	1.1	-0.2
3BSD		81.2	84.9	86.4	81.2
	Error %		-4.6	-6.4	0.0

Sum of all BChl 3 and BChl 1 effect differences calculated in models including only [a] main effects or up to 2-way interactions; [b] up to 3- or 4-way interactions; [c] up to 5- or 6-way interactions.

Between the substituent *main* effect differences from structures 3EOJ and 3VDI, which were noted above to display a synergistic intrinsic site-energy difference with the reference values (Table 7-16), those of the Mg, ring E, acetyl and alkyl substituents show the most consistency (Table 7-18). These results imply that the conformations of the acetyl group make the largest individual contribution of ~10 nm, and offset the opposing influences of the other substituents. However, both of these structures also suggest that the second largest individual contribution arises from differences in the macrocycle conformations of these BChls, although they differ in sign. The 3ENI structure is in moderate agreement

---

* 3-way interactions refer to interactions amongst three sets of atom groups simultaneously; they are to the two atom group interactions (2-way interactions; discussed in section 7.6.1) as the 2-way interactions are to the main effects. Mathematically, they are defined as the model coefficients for three atom optimisation state product variables.

† Note that empirically, it appears that the agreement between the simulation and the exact difference is the same for the 1- and 2-way models, the 3- and 4-way models, and so on when modelled in this way. The author has not yet uncovered the theoretical explanation for this. Additionally noteworthy is that the agreement does not have to monotonically increase with the size of the model (*i.e.*, the order of interactions included), this is because the OLS fits that determine the effect coefficients minimise the sum-of-square error of all permutation residuals.

with some of these results too, although in this structure, the conjugated groups are suggested to have almost no effect at all.

Interestingly, the differences between these BChls' interaction coefficients appear to be markedly conserved between the structures mentioned above (Table 7-19). In particular, the macrocycle-Mg interaction ranges over approximately -12 to -17 nm and thus dominates the apparent main effects of these substituents. Moreover, this interaction effect *is* the greatest contributor to the site-energy difference in the 2-way model, and not as was suggested by consideration of the main effects alone, the acetyl conformation. In addition, moderately large, conserved differences in the macrocycle-alkyl and Mg-alkyl interactions suggest that these three atom groups, in combination, are very strong moderators of the intrinsic site-energy.

Table 7-18: Differences between the atom group main effects for BChl 3 minus BChl 1 from a selection of PDB structures (nm).

Structure	Intercept	Mg	Mac.	CP	Ace.	Alks.
3EOJ	4.0	-4.1	-8.9	-2.5	11.9	-4.2
3ENI	1.2	-3.4	0.9	-0.1	0.2	-3.9
3VDI	7.7	-4.2	5.7	-2.3	7.9	-2.3
3BSD	-58.9	1.9	110.9	-3.7	12.0	-5.2

Table 7-19: Differences between the atom group interaction effects for BChl 3 minus BChl 1 from a selection of PDB structures (nm).

Structure		Mg	Mac.	CP	Ace.	Alks.
3EOJ	Mg		-12.3	2.2	1.3	8.2
	Mac.			3.6	3.4	-4.7
	CP				0.4	-2.1
	Ace.					-7.2
	Alks.					
3ENI	Mg		-17.3	0.5	3.7	8.8
	Mac.			1.7	7.6	-10.0
	CP				1.5	-4.1
	Ace.					-0.5
	Alks.					
3VDI	Mg		-14.5	1.9	3.0	6.4
	Mac.			-1.3	-2.0	-14.3
	CP				0.4	-0.9
	Ace.					-3.1
	Alks.					
3BSD	Mg		10.2	-5.0	0.5	1.5
	Mac.			-23.6	10.4	-11.6
	CP				-19.0	12.9
	Ace.					-7.5
	Alks.					

Whilst the forgoing analysis has been brief, it appears that the macrocycle skeletal conformation exerts a large moderating influence upon the intrinsic site-energy shift between these BChls, in addition to similar contributions from the Mg and alkyl groups. Moreover, the influence of the acetyl group represents the largest main effect and thus may contribute to the potentially mistaken attribution of its domination of the BChl conformational contribution to the site-energy differences. Indeed, the above results show that there are many ways that the protein* can affect a conformational perturbation to the site-energy (Table 7-18 and Table 7-19). This observation also suggests why less complicated experimental designs place great weight upon the local influence of individual substituents; it is because they *do* possess a large effect, but they are *not* necessarily dominant. The overall difference is a sum over all of the interaction effects of the substituents, in addition to their main effects. It is expected that the conformational mechanism underlying these main effects and interactions will be further characterised once linear models are constructed that relate the conformational parameters to the optimisation states of the atom groups directly. Once this analysis has been performed, it is likely that the abstract notion of a conformational interaction can be defined more precisely.

7.6.3 Illustrative Analyses from Model B

Accuracy of Linear Models

A detailed analysis of the results from the full-factorial series of partial optimisations conducted for model A upon BChl 1 from PDB structure 3BCL is now presented. Since there are 12 structural factors defined in this model, up to 12-way interactions may be specified. There are a few good reasons, however, not to include higher-order interactions one of the most compelling being the difficulty in chemically interpreting such high-order behaviour. Indeed, it would be 'unusual' to include interactions above the fourth order as typically the sparsity-of-effects principle is invoked to dismiss their significance *a priori*.[390] Another major reason to not investigate interactions of high-order is that in the strictest statistical sense, main-effects and lower order interactions become less meaningful when they are considered alongside high-order effects, if the high-order interactions are large in comparison. There are three straightforward criteria that can be used to determine the linear model of least complexity that provides a suitable description of the experiment, specifically the residuals plot, R^2 and the model p-value. These are compared for models

* Or equivalently, they can be artificially modified in a computational experiment.

that consist of main effects only and up to and including 4-way interactions in Table 7-20 and Figure 7-15 and Figure 7-16.

Table 7-20: Goodness-of-fit and significance parameters for the linear models connecting the Model B atom group effects (Table 7-1, p. 164) to the calculated E_{strain} for BChl 1 from PDB structure 3BCL; models are specified by their highest-order interactions.

Model	R^2	p-value
Main effects	0.9580	$< 2.2e^{-16}$
2-way	0.9935	$< 2.2e^{-16}$
3-way	0.9981	$< 2.2e^{-16}$
4-way	0.9993	$< 2.2e^{-16}$

The residuals plots (Figure 7-15) indicate that neglecting to consider any interactions leads to a serious systematic error, showing a quadratic dependence of the residuals on the predicted response. Such relationships are indicative that the model requires quadratic terms (*i.e.*, 2-way interactions) and the addition of these to the model corrects this violation and renders the error distribution more or less random (Figure 7-15, top right). However, even in the simplest model where only main effects of the atom groups are considered, the models' description of the variation of calculated Q_y energies is excellent and accounts for > 95 % of the variation (Table 7-20). Further consideration of these results suggests that although the error range is significantly reduced by including 3rd- and 4th-order effects, their moderate benefits are greatly outweighed by the substantially greater deviation from normality that they introduce to the error distribution (Figure 7-16). In any case, the main goal of this study is the determination of the effect coefficients as opposed to the prediction of the results of the PM6//ZIndo/S calculations from each individual permutation, although clearly this is a critical test of the model and the reliability of these parameters. Thus, only main effects and second order interactions will be discussed in-depth in the rest of this work although results are provided for models containing up to 4-way interactions for completeness and to illustrate the sensitivity of the atom group effect sizes to the number of interactions included in the model.

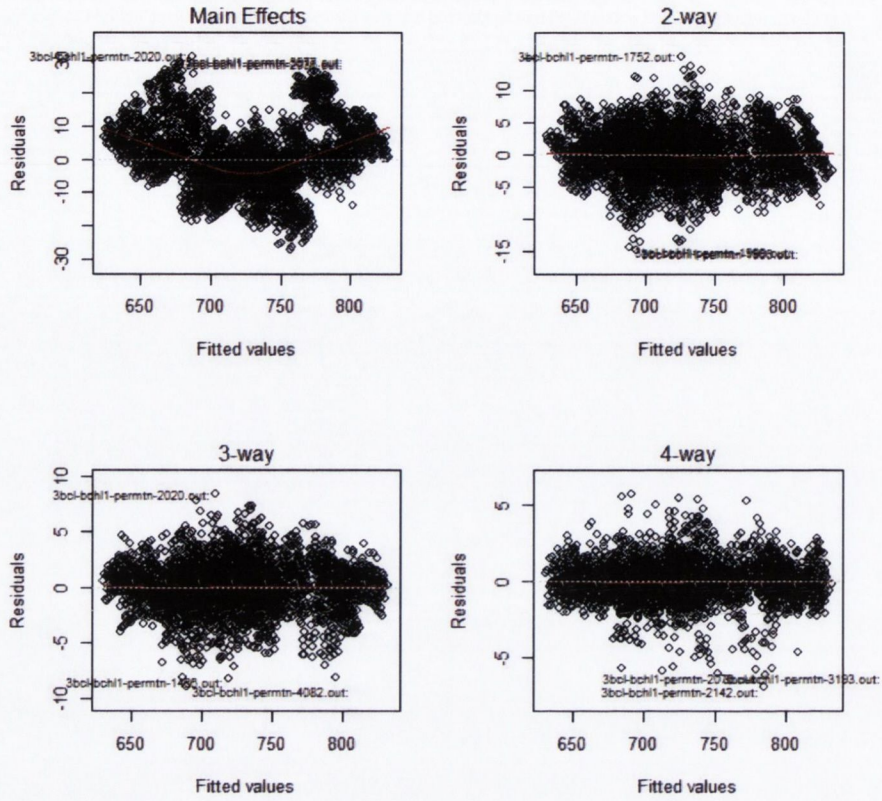


Figure 7-15: Residual plots from the linear models constructed for BChl 1 from 3BCL including atom group interactions up to and including the level indicated above each figure; note the various scales and the clear quadratic dependence of the residuals on the response when only main-effects are considered.

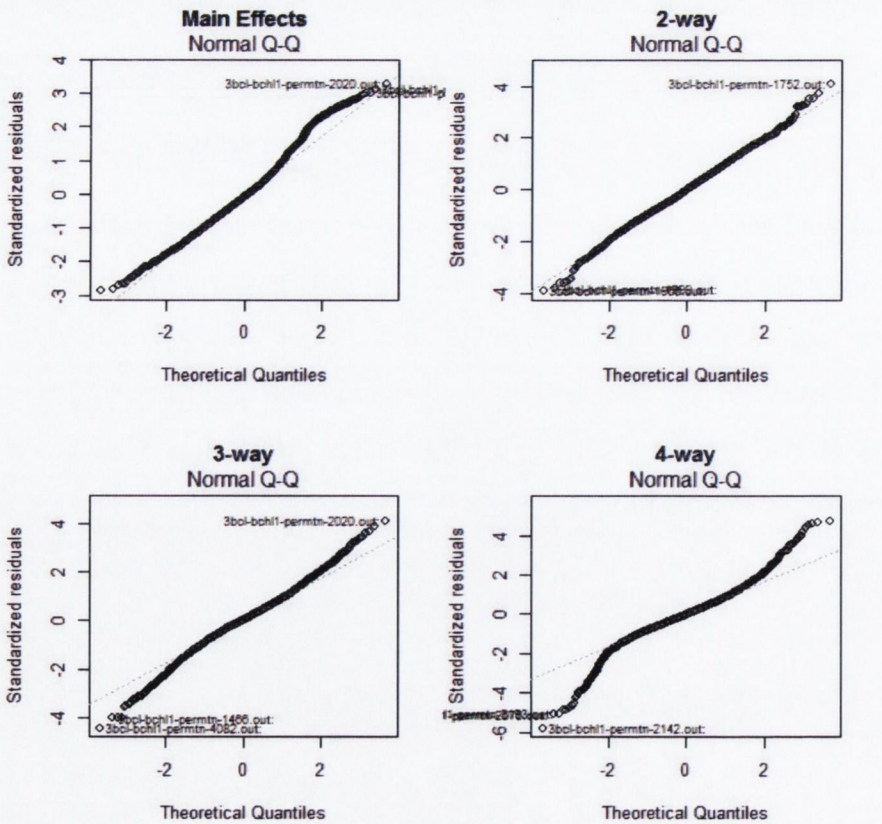


Figure 7-16: q-q plots illustrating the extent to which the residuals from the linear models constructed for BChl 1 from 3BCL including atom group interactions up to and including the level indicated above each figure are normally distributed by comparing how well the observed distribution (points) follows the theoretical distribution (dashed line).

Main effects and strong interactions

Two features above all stand out from an analysis of the model B atom group main effects (Table 7-21). Specifically, the protein-induced conformations of all but two of the defined fragments result in red-shifts of the cofactors intrinsic site-energy relative to the optimised, *in vacuo*, geometry. The two fragments for which the crystallographically resolved conformation causes a blue-shift are rings D and E (Table 7-21). The second dominant pattern is that the conformations of the macrocycle rings A-D have the greatest influence on the calculated site-energy by far. Although this later point is strongly supportive of the conformational control hypothesis, one must bear in mind that, as discussed previously, there is a bias due to the number of (conjugated) atoms contained in the fragment. However, it should be considered significant that the ring B and D alkyl substituents (Alkyls | Q_x) contain more atoms, albeit they are not conjugated.

Closer inspection of the atom group main effects (Table 7-21) indicates a number of chemically interesting points. The acetyl has the largest substituent effect only if interactions are not considered (*i.e.* excluding macrocycle or Mg fragments). This is in line with the observations regarding the BChl 3 *vs.* BChl 1 model A results and illustrates further that failure to consider atom group interactions could bias the conclusions drawn from less complete experiments (*e.g.*, sequential truncation). Additionally, the relatively large main effects of the alkyls substituents are in contrast to the *status quo* that only conjugated substituents induce significant site-energy shifts (cf. acetyl main effect). This feature may be observed either because these groups act as handles for the protein to affect core distortion, in which case they should be accompanied by large interactions, or alternatively, the effect could be the direct result of hyperconjugative interactions with the macrocycle. Lastly, the cyclopentanone carbonyl oxygen has only a very small effect if interactions are not considered and in this model appears to exert influence *via* the ring E (cf. entries E and [C=]O and the relevant interaction term in Table 7-22). This feature is not particularly interesting chemically although it points to a deficiency of the fragment definitions used here; assigning the keto oxygen to its own fragment does not allow enough freedom of movement for either this atom or the ring E skeletal framework; in hindsight it would have been wiser to have considered these fragments together as a single group.

Table 7-21:Atom group main effects from the linear model relating the conformations of the atom groups to the total perturbation of the intrinsic site-energy for BChl 1 from structure 3BCL; all units nm.

			Rings					Conjugated		Acids		Alkyls	
Model	'c'	Mg	A	B	C	D	(E)	$C(O)CH_3$	[C=]O	$C13^2$	C17	Q_x	Q_y
Mains	795.1	10.6	20.7	54.5	59.8	-30.5	-2.3	9.3	0.8[d]	3.7	1.5	8.2	2.0
2-way	811.8	16.0	36.6	76.0	77.4	-23.1	-0.4[a]	4.0	3.2	5.0	3.2	6.4	0.7[b]
3-way	809.1	9.7	35.5	73.4	76.7	-23.1	0.2[a]	1.1[c]	4.0	2.7	1.6[d]	6.3	1.1[c]
4-way	808.9	10.6	33.9	71.4	76.9	-23.7	-2.8	0.8[a]	2.0[d]	2.4	2.3	9.0	4.4

Unless otherwise indicated, p-values are significant at the 0.001 level, indeed many are $< 2.2e^{-16}$. a) Not statistically significant b) 0.1 c) 0.05 d) 0.01

The 2-way interactions from the 2-level interaction linear model unsurprisingly reflect the connectivity between the fragments (Table 7-22). Thus, the large interactions between adjacent ring pairs A-B, A-D, B-C and C-D are most likely the result of a mutual conformational and electronic dependence. These interactions may be exaggerated however due to the overlapping fragment definitions that associated each macrocycle ring with *both* adjacent meso-carbons. Other interactions that are structurally expected include those between the skeletal ring D-E and the E-keto oxygen, the latter of which is the largest interaction involving a non-core fragment. Interesting cases include the significantly stronger interplay between rings B and D and the Mg-atom, which may arise because these rings have shorter (stronger) Mg-N bonds compared to the other donors' N-atoms, which belong to the reduced rings A and C, and also since they lie on the Q_y-axis. On the other hand it is surprising that amongst the acetyl cross-terms, ring C's are far greater than ring B's, the ring to which it is attached; notably again however is that this is the third largest non-core/core interacting term. Supporting the statement made above regarding the alkyl substituents, their interactions with the core system (Mg and rings A-D) are the largest in general, although again it is important to consider that they are substantially larger fragments in terms of atom count.

Table 7-22: 2-way interaction effects from the 2-way linear model relating the conformations of the atom groups to the total perturbation of the *intrinsic* site-energy for BChl 1 from structure 3BCL; all units nm.

		Rings					Conjugated		Acids		Alkyls	
	Mg	A	B	C	D	(E)	$C(O)CH_3$	C=O	$C13^2$	C17	Q_x	Q_y
Mg		0.4	-5.3	-0.2	-5.7	-0.5	1.1	1.2	-1.0	0.7	2.6	-4.2
A			-12.2	-9.1	-8.4	-1.2	1.3	0.1	-0.8	-1.5	-2.5	2.2
B				-24.3	-0.6	0.5	1.4	0.4	0.7	-0.9	-0.5	-2.2
C					-10.4	0.0	3.8	-0.9	0.7	0.7	3.2	1.1
D						3.7	1.2	0.1	0.4	-0.2	3.2	2.1
(E)							0.2	-4.6	-1.2	-1.6	0.6	0.3
$C(O)CH_3$								-0.1	-0.2	0.1	-0.6	2.3
C=O									-0.8	0.5	-0.6	-0.2
$C^{13\wedge 2}$-										-1.1	-0.1	0.6
C^{17}-											-1.0	1.0
Q_X												-0.6
Q_Y												

Factor combinations

Whilst it is not possible to divide fragments *post hoc*, it is relatively simple to combine them after a larger analysis* to determine, for example, the gross structural effect of the macrocycle as a whole (either 24- or 26-carbon skeleton) as opposed to individual ring contributions. This simplifies the chemical interpretation significantly as ring interactions are the largest present. Another useful combination is that of $C13^1$ and $C13^2$ carbons of ring E with the keto-oxygen. The only deleterious effect of such combinations is the increasing standard errors of the factor coefficients as a result of the fact that combining any two factors halves the number of experimental runs (*e.g.*, after combining ring E and the keto-oxygen, any structures where those individual factors are in different states become meaningless to the new experimental design). However, that said the only requirement is that these remain low enough to significantly separate key-effects across a structure series. The results of the analysis after taking the approach of uniting rings A-D into a single macrocycle factor are illustrated in Table 7-23 to Table 7-25. The most notable effect is that the macrocycle coefficients for different models are all ~100 nm and that the acetyl and Q_y-alkyl effects increase quite dramatically. Thus, the results obtained from this procedure differ only from model A because more of the periphery was retained.

* This was why this model was trialled first.

Table 7-23: Goodness-of-fit and significance parameters for the linear models including the indicated highest-order factor interactions after combining rings A-D into a single macrocycle factor for BChl 1 from structure 3BCL.

Model	R^2	p-value
Main effects	0.9895	$< 2.2e^{-16}$
2-way	0.9962	$< 2.2e^{-16}$
3-way	0.9993	$< 2.2e^{-16}$
4-way	0.9998	$< 2.2e^{-16}$

Table 7-24: Main effects from the linear models relating the conformations of the atom groups to the total perturbation of the intrinsic site-energy for BChl 1 from structure 3BCL; all units nm.

			Rings		Conjugated		Acids		Alkyls	
Model	'c'	Mg	24-macrocycle	(E)	$C(O)CH_3$	[C=]O	$C13^2$	C17	Q_x	Q_y
Main effects	701.6	7.1	108.1	-2.7	6.6	0.9	2.7	0.8	9.0	3.1
2-way	706.1	4.6	104.2	0.9	10.0	2.4	5.3	0.0	5.6	12.9
3-way	707.8	8.4	100.9	1.2	10.8	2.0	5.7	1.0	11.6	7.6
4-way	708.8	8.9	99.8	3.4	12.2	3.9	6.5	1.2	11.8	7.5

Table 7-25: 2-way interaction effects from the 2-way linear model relating the conformations of the atom groups to the total perturbation of the intrinsic site-energy for BChl 1 from structure 3BCL; all units nm.

			Rings		Conjugated		Acids		Alkyls	
		Mg	24-macrocycle	(E)	$C(O)CH_3$	C=O	$C13^2$	C17	Q_x	Q_y
	Mg		-8.0	1.3	1.5	1.8	-0.6	1.2	0.9	-8.9
Rings				3.2	9.1	0.3	0.8	3.0	4.5	1.0
	(E)				0.4	-3.7	-1.7	-2.6	1.0	1.1
Conj.	$C(O)CH_3$					0.0	-0.4	0.3	-1.3	1.7
	C=O						-0.4	0.7	-1.4	0.2
Esters	$C13^2$							-2.0	-0.1	0.8
	C17								-1.2	2.4
Alkyls	Q_X									-1.2
	Q_Y									

Inter-cofactor Differences: 3EOJ BChl 1 vs. BChl 5

This section discusses the main effects for models including up to fourth order interactions for BChl 1 from structure 3BCL (Table 7-21) and BChls 1 and -5 from structure 3EOJ (Table 7-26 and Table 7-27, respectively). Generally, the magnitudes of the main effects vary considerably between each structure, in particular for the rings and the acetyl group. However, it does appear that the two representatives of BChl 1 are more similar to each other than to the BChl 5 pigment. This may be significant as it is suggestive that some aspects of BChl 1's conformation are similar between these two structures and may be indicative of the reliability of the crystallographic determinations. The *relative* effects of the individual rings from the 3EOJ structure are consistent with 3BCL differing mostly in the ordering of rings B and C. The 3BCL structure gives an ordering for the signed ring

main effect sizes of the rings as C ≈ B > A >> D whilst both of the 3EOJ structures yield B >> C > A >> D. This may be indicative of the relative importance that the conformation of each individual ring has with respect to the Q_y energy modulation. Notably, the effect of ring D takes both a negative and near zero sign (*i.e.*, suggesting a slight protein affected blue-shift). Additionally, in these structures the acetyl group continues to retain significant importance possessing effects corresponding to a 26 nm red-shift in BChl 1 and a 9 nm red-shift in BChl 5. Lastly, as observed previously, the largest interactions tend to be between connected rings and also the acetyl appears to interact most strongly with ring C, lending weight to the idea that this is a general property of the macrocycle (Table 7-28 and Table 7-29 for 3EOJ structures and Table 7-22 for the 3BCL structure).

Table 7-26:Atom group main effects from the linear models relating the conformations of the atom groups to the total perturbation of the intrinsic site-energy for BChl 1 from structure 3EOJ; all units nm.

			Rings					Conjugated		Acids		Alkyls	
Model	'c'	Mg	A	B	C	D	(E)	$C(O)CH_3$	[C=]O	$C^{13\hat{}2}$-	C^{17}-	Q_x	Q_y
Main effects	844.1	0.7	27.9	95.5	44.0	-16.7	-0.7	25.7	-0.3	0.6	-1.2	0.0	0.2
2-way	849.3	2.3	26.2	106.4	51.3	-11	-0.9	21.5	0.6	1.4	0.2	0.0	-1.4
3-way	846.0	1.5	21.5	100.3	46.8	-8.8	0.0	14.5	0.9	0.2	0.9	1.1	-2.0
4-way	844.5	0.3	18.8	98.9	44.2	-10.6	-1.6	16.2	-0.2	-0.6	0.6	0.1	-1.2

Table 7-27: Atom group main effects from the linear models relating the conformations of the atom groups to the total perturbation of the intrinsic site-energy for BChl 5 from structure 3EOJ; all units nm.

			Rings					Conjugated		Acids		Alkyls	
Model	'c'	Mg	A	B	C	D	(E)	$C(O)CH_3$	[C=]O	$C^{13\hat{}2}$-	C^{17}-	Q_x	Q_y
Main effects	853.9	3.8	40.1	85.2	53.9	0.2	0	8.8	-0.5	0.0	-0.4	1.4	0.9
2-way	854.7	6.1	36.5	89.2	54.6	8.1	-0.3	5.0	-0.6	0.0	-0.3	-1.3	-0.5
3-way	850.8	0.4	32.4	81.8	53.9	8.1	0.2	0.4	0.4	-1.6	-0.1	-2.0	-0.9
4-way	852.0	1.2	30.7	80.3	52.8	6.6	0.2	5.1	1.8	-0.8	1.1	2.4	1.1

Table 7-28: 2-way interaction effects from the 2-way linear model relating the conformations of the atom groups to the total perturbation of the intrinsic site-energy for BChl 1 from structure 3EOJ; all units nm.

		Rings					Conjugated		Acids		Alkyls	
	Mg	A	B	C	D	(E)	C=0	$C(O)CH_3$	$C13^2$	C17	Q_y	Q_x
Mg		0.5	3.4	0.4	1.7	-0.1	-0.1	0.0	0.6	-1.2	0.2	-2.2
A			0.8	-5.3	0.7	0.8	0.1	-2.5	0.4	0.5	-1.3	1.8
B				14.1	1.9	0.8	0.1	0.5	-0.2	0.6	-1.2	1.1
C					12.9	0.2	0.2	-6.2	-0.4	0.6	-0.6	-1.4
D						-3.2	0.3	-1.2	-0.4	0.3	-0.6	-1.1
(E)							0.9	-0.1	0.2	0.1	-0.1	-0.1
C=O								0.0	0.4	-0.2	0.1	0.2
$C(O)CH_3$									0.1	0.0	0.7	0.5
$C13^2$										0.9	-0.2	0.1
C17											-0.1	1.4
Q_y												-0.1
Q_x												

Table 7-29: 2-way interaction effects from the 2-way linear model relating the conformations of the atom groups to the total perturbation of the intrinsic site-energy for BChl 5 from structure 3EOJ; all units nm.

		Rings					Conjugated		Acids		Alkyls	
	Mg	A	B	C	D	(E)	C=0	$C(O)CH_3$	$C13^2$	C17	Q_y	Q_x
Mg		-4.3	2.8	-1.3	2.8	0.7	0.3	2.0	1.0	-0.5	1.1	0.1
A			2.6	-5.9	5.0	0.3	0.5	-2.0	-1.0	-0.1	-0.8	-1.4
B				3.3	1.4	0.7	0.4	-1.9	-1.0	-0.7	-0.2	0.7
C					12.0	0.5	0.2	-3.3	-1.0	0.2	-0.5	-2.6
D						-3.1	0.3	-0.2	0.0	-0.6	-0.5	-1.2
(E)							-0.9	0.1	0.5	0.4	0.3	-0.3
C=O								0.0	-0.3	-0.3	0.1	-0.3
$C(O)CH_3$									-0.1	-0.1	-2.1	0.0
$C13^2$										1.7	0.0	0.1
C17											0.2	-0.1
Q_y												-0.3
Q_x												

The following comparison is likely to become the prototype for further analyses once the remaining structures have been processed. At the outset, it is worth pointing out that the individual conformational effects appear to cancel between pigments BChl 1 and BChl 5 in structure 3EOJ. However, as the crystallographic structure based calculated intrinsic site-energies (E_{cryst}) were very similar for this pair (845 and 851 nm for BChls 1 and -5, respectively), it is a strength of the analysis that this is the case.

The crystallographic structure based site-energies for the 3EOJ pigments noted above correspond to an intrinsic shift of -5.55 nm of the electronic transition wavelength of BChl 1 relative to that of BChl 5. Similar to the procedure for the BChl 3 and BChl 1 comparison for model A (p. 188), this may be partitioned into differences caused by the conformations of each pigments molecular components, although in this case the results are significantly more detailed. To re-iterate, the difference of each main effect (Table 7-30) may be taken to represent the effect of the conformational differences between each defined fragment whilst the total difference given by the sum of the model coefficients (*e.g.*, the sum of the first row, excluding the intercept, in Table 7-30) should approximate the actual difference within the model's error margin. For the models containing interaction terms, the differences in these effects must also be summed in order to attain the estimate (see Table 7-31 for 2-way interaction differences).

Table 7-30: Difference between the main effects of the relaxation state of the molecular fragments for BChls 1 and -5 from 3EOJ; all nm. The row sums give the contribution to ΔE_{cryst} between these pigments from the main effects and are -17.7, 0.1, 3.9 and -17.6 nm for the 1-, 2-, 3- and 4-way models, respectively.

			Rings					Conjugated		Acids		Alkyls	
Model	'c'	Mg	A	B	C	D	(E)	$C(O)CH_3$	[C=]O	C^{13^2}-	C^{17}-	Q_x	Q_y
1-way	-9.8	-3.1	-12.2	10.3	-9.9	-16.9	-0.7	16.9	0.2	0.6	-0.8	-1.4	-0.7
2-way	-5.4	-3.8	-10.3	17.2	-3.3	-19.1	-0.6	16.5	1.2	1.4	0.5	1.3	-0.9
3-way	-4.8	1.1	-10.9	18.5	-7.1	-16.9	-0.2	14.1	0.5	1.8	1	3.1	-1.1
4-way	-7.5	-0.9	-11.9	18.6	-8.6	-17.2	-1.8	11.1	-2	0.2	-0.5	-2.3	-2.3

Table 7-31: Differences between the 2-way interaction effects of the relaxation state of the molecular fragments BChls 1 and -5 from 3EOJ; all nm. NB. The total sum of the 2-way interaction differences (17.6) is added to the sum of the differences of the 2-way model's main effects (0.1; Table 7-31) in order to reach the 2-way model's simulation of this inter-site difference.

		Rings					Conjugated		Acids		Alkyls		
	Mg	A	B	C	D	(E)	C=0	$C(O)CH_3$	C^{13^2}-	C^{17}-	Q_y	Q_x	Sum
Mg		4.8	0.6	1.7	-1.1	-0.8	-0.4	-2.0	-0.4	-0.7	-0.9	-2.3	-1.5
A			-1.8	0.6	-4.3	0.5	-0.4	-0.5	1.4	0.6	-0.5	3.2	-1.2
B				10.8	0.5	0.1	-0.3	2.4	0.8	1.3	-1.0	0.4	15
C					0.9	-0.3	0.0	-2.9	0.6	0.4	-0.1	1.2	-0.2
D						-0.1	0.0	-1.0	-0.4	0.9	-0.1	0.1	-0.6
(E)							1.8	-0.2	-0.3	-0.3	-0.4	0.2	0.8
C=O								0.0	0.7	0.1	0.0	0.5	1.3
$C(O)CH_3$									0.2	0.1	2.8	0.5	3.6
$C13^2$										-0.8	-0.2	0.0	-0.1
C17											-0.3	1.5	1.2
Q_y												0.2	0.2
Q_x													
Sum		4.8	-1.2	13.1	-4	-0.6	0.7	-4.2	2.4	2.8	-0.2	4	17.6

However, the sum of the differences in the coefficients from neither the model that contains no interactions nor that with second order interactions provides a good approximation of the explicitly calculated intrinsic difference (*i.e.*, -17.7 nm from both models compared to -5.55 nm explicit). In order to come to terms with this feature, one need only look at the fully relaxed optimisation, permutation 4096. It was originally expected that all full optimisations would find the same minimum and thus yield the same excitation energy corresponding to the PM6//ZIndo/S *in vacuo* approximation. This would have provided an ideal base-line from which the protein induced perturbations would be assessed. However, in this case that has turned out not to be the case. The calculated excitation energies and indeed the conformations of the full optimisations of BChls 1 and -5 differ significantly (Table 7-32, Table 7-33 and Figure 7-17 to Figure 7-19). The fully optimised transition energies for BChls 1 and -5 are 676 and 665 nm, respectively, corresponding to a difference of 11 nm, surprisingly (and worryingly) greater than the intrinsic site-energy difference. Moreover, as indicated by the NSDs (Table 7-32 and Table 7-33), the macrocycles have been optimised to distinct conformations. In detail, BChl 5 is slightly more distorted from planarity than BChl 1 by 0.02 Å and in composition has significantly more *ruf* distortion (by ~1 Å), although reassuringly the in-plane conformations are virtually indistinguishable.

Table 7-32: The excitation energies of BChls 1 and 5 from PDB structure 3EOJ calculated from the crystal structure geometry and after full optimisation of this starting geometry together with the NSD out-of-plane parameters of these geometries.

		E_{ES1} / nm	D_{oop}	δ_{oop}	B_{2u}	B_{1u}	A_{2u}	E_g(x)	E_g(y)	A_{1u}
BC1	Crystal	845.12	0.3561	0.0423	-0.1075	-0.2356	-0.1057	-0.1073	0.0707	-0.1792
	Optimised	676.44	0.2299	0.0157	-0.2157	0.0304	0.0111	-0.0674	-0.0022	0.0268
BC5	Crystal	850.67	0.4016	0.0503	0.3351	-0.0117	0.1008	-0.1493	-0.0991	0.0812
	Optimised	665.46	0.2500	0.0187	-0.2105	-0.1210	-0.0496	-0.0155	0.0127	0.0265

Table 7-33: The excitation energies of BChls 1 and 5 from PDB structure 3EOJ calculated from the crystal structure geometry and after full optimisation of this starting geometry together with the NSD in-plane parameters of these geometries.

		E_{ES1} / nm	D_{ip}	δ_{ip}	B_{2g}	B_{1g}	Eu(x)	Eu(y)	A_{1g}	A_{2g}
BC1	Crystal	845.12	0.4443	0.0545	0.0236	-0.2971	-0.0622	-0.0194	0.3116	-0.0850
	Optimised	676.44	0.5238	0.0591	0.0763	-0.3214	-0.0677	-0.0399	0.3850	-0.1041
BC5	Crystal	850.67	0.4694	0.0450	0.0743	-0.314	-0.0787	-0.0786	0.3104	-0.0863
	Optimised	665.46	0.5226	0.0589	0.0811	-0.3196	-0.0702	-0.0421	0.3836	-0.1026

At first, it was thought that this may be the result of one or more incomplete geometry optimisations, however this was discounted as frequency calculations showed the structures to be true local minima. The origin of the discrepancy was revealed after the inspection of the optimised structures to be the result of highly dissimilar starting

geometries as illustrated by superimposing both crystallographic and fully optimised structures with their equivalent counterpart (Figure 7-17). Inspection of this figure provides further reassuring that the structures have indeed been optimised to a point where they are in many ways visually indistinguishable (excluding the $C13^2$- and C17-esters) and is highlighted by RMS deviations of 0.843 and 0.171 Å for the crystal geometries and the optimised structures, respectively.

The rotation of the $C13^2$-ester group in particular appears to have left a slight kink in the cyclopentanone ring in the optimised structures (Figure 7-17, right), that is reminiscent of the usual zig-zag conformation observed for this ring in phytochlorins.[157] In terms of the starting structures, the orientation of the propionate moiety has created an even clearer difference between the conformation of ring-D and is best seen in the skeletal plots (Figure 7-18; see Figure 7-3, p. 164 for ring legend) where it is highlighted that the corresponding β-carbons on each pigment are distinctly on opposite sides of the plane with respect to each other.

It is unfortunate that such small fluctuations have created such a large difference in the calculated excitation energies, although it would seem reasonable to suggest that it is the increased ruffling of the BChl 5 pigment that has created the red-shift. This is only apparent in the NSD though (Table 7-32) as the optimised skeletal plots indicate no clear qualitative difference (Figure 7-19). Another potential candidate would be the kink in ring-E that cannot be seen on the skeletal plots as it is not included but is visible on the superimposition (Figure 7-17).

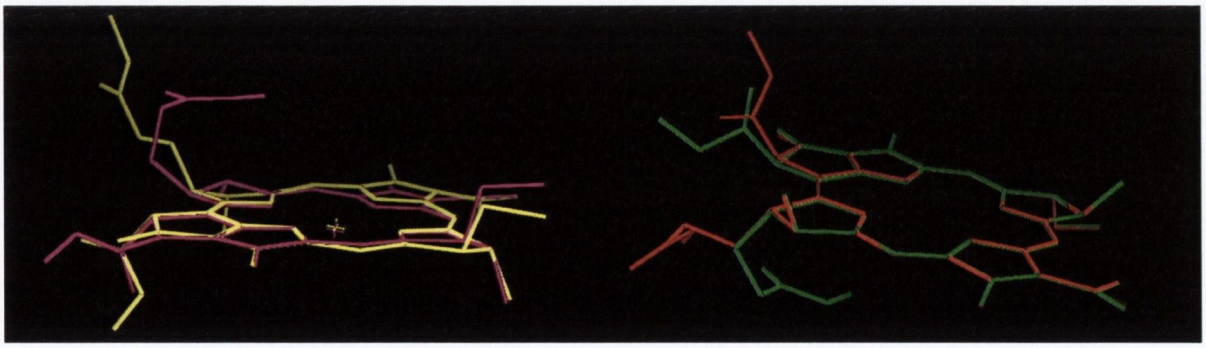


Figure 7-17: Crystallographically resolved BChl 1 (magenta) aligned to BChl 5 (yellow); all atom RMS = 1.99 and 64 atom RMS = 0.843. Fully optimised BChl 1 (green) aligned to BChl 5 (red). All atom RMS=1.77; 53 atom RMS=0.171.

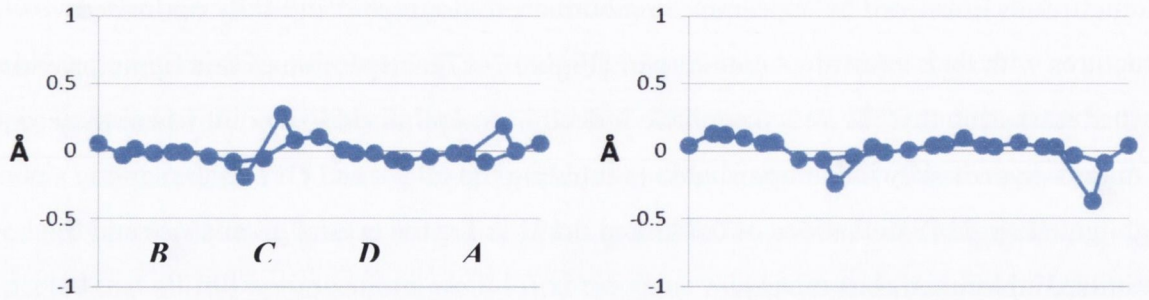


Figure 7-18: Skeletal plots of the crystallographically resolved conformations for BChl 1 (left) and BChl 5 (right) from PDB structure 3EOJ. Note that the PDB ring nomenclature is indicated (Figure 7-3, p. 164).

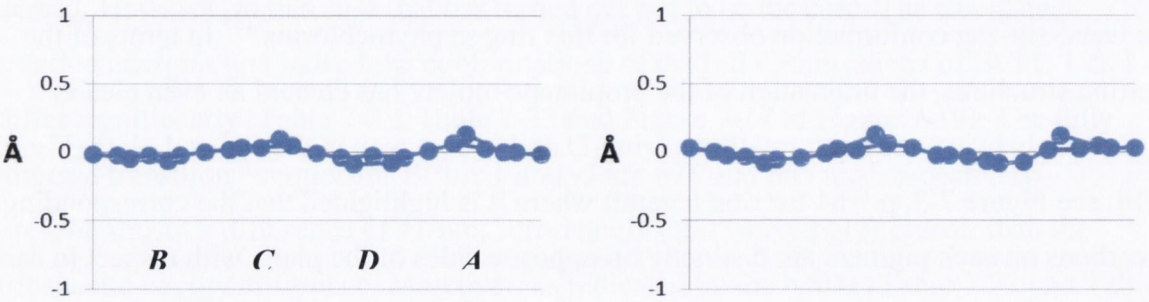


Figure 7-19: Skeletal plots of the fully optimised conformations for BChl 1 (left) and BChl 5 (right) using the coordinates from PDB structure 3EOJ as the starting geometry. Note that the PDB ring nomenclature is indicated (Figure 7-3, p. 164).

In a typical optimisation, one way to 'fix' this would be to manually perturb the conformer of higher energy so that it could re-optimise to the lower energy minimum. In this experiment however, such an intervention would defeat the purpose of the experimental design since in order to provide meaningful comparisons across different pigments they should all have been treated identically. Moreover, a significant technical challenge to manual perturbation is that the offending substituent would have to be perturbed in all of the partial optimisation permutations where it is relaxed (*i.e.* in $N/2$ experimental units), which here amounts to 2048 optimisations and this would have to be automated. Another alternative that is more consistent with the identical treatment of each structure would be to exclude the offending substituent from the structural model in the first place and this indeed proved a successful action and was one of the motivations for the implementing "Model A" as described previously.

A simple correction that can be performed at present is simply to level the base-lines and although only an approximate solution, in this case at least provides an acceptable result. Removing the difference of 10.98 nm in the excitation energies of the fully-optimised structures yields an estimated intrinsic shift of -6.72 nm, a substantial improvement with respect to agreement with the explicitly calculated shift of -5.55 nm.

This agreement is the first verification of the approach in general and suggests that the fragment effects in model A may also be considered as local geometric perturbations of the excitation energies (at the given level of theory). The inter-site differences (Table 7-30) suggest that there are numerous cancellations that result in the relatively small shift of -5.55 nm for this pair. Indeed, the difference in the acetyl conformation alone swamps this, which is again consistent with the fact that the acetyl can have a major effect of the site-energy However, these results suggest that the effect of interchanging the conformations of any of the individual tetrapyrrolic rings would achieve a similar effect.

7.7 Conclusion

7.7.1 Summary

Preliminary results regarding the conformational analysis of the BChl pigments in the FMO protein together with the application of a computational approach designed to delineate the individual contributions of the conformation of the macrocycle and peripheral substituents, as well as their interactions, on the unperturbed site-energies have been reported and discussed. After a brief account of a series of acetyl relaxed rotational excitation energy profiles that demonstrated this functional group's local conformation was not responsible for the entirety of the intrinsic site-energy variation in a particular crystal structure, it was first shown that the variation across all BChls over all structures was largely explained by the macrocycle conformations.

Whilst no conclusive statement can be made at this time with respect to a definite biological effect, the initial results suggest that the exact conformational contribution to these intrinsic site-energy shifts is the result of more than a simple summation of the individual effects of the molecular components of BChl. However, such a summation does do well in approximating the conformational effect and therefore it remains to be seen to what extent substituent interactions effect previous conclusions regarding Q_y modulation in BChls in general.

The results obtained from the reanalysis of Gudowska-Nowak *et al*'s[111] study using the NSD method suggests that whilst the method of truncation can provide information regarding differences in substituent effects, never before has the effect of the macrocycle conformation been adequately treated because different effects are produced by different distortions depending on which substituents are included. Moreover, in studies that have used the truncation method,[396] it is possible that effects that have been attributed to differences in substituent conformations (*e.g.*, of the acetyl or cyclopentanone) are as much

a product of the macrocycle conformation as they are of differences in these local parameters.

7.7.2 Validity and Improvements

The validity of the approach in terms of computational accuracy could be tested with a restrained optimisation of the BChl pigment in its binding pocket with the relevant method (*i.e.* here, PM6). This could be important because the methods "preference" for specific bond parameters may bias the factorial analysis to provide a large contribution for runs with more unrestrained atoms, or else particular types for which the crystal structure model (*e.g.*, restraints) deviates from the preferred significantly (*e.g.*, aromatic bond-lengths). Apart from simply performing the optimisation as a control for the strain energy described above, it could also be used as a starting point for another full-factorial partial optimisation experiment. Another method would be to use a higher-level of theory for the SPE calculation (*e.g.*, PM6//TD-DFT) although an even smaller experiment than model B could be designed by comparing only the alkyls, acetyl, cyclopentanone, and the macrocycle (16 EUs) could be performed at the DFT//TD-DFT level.

A final note on the reliability of the computational approach: numerous recent benchmarks for the related PM5//ZIndo/S for chlorophylls and bacteriochlorophylls have shown that this method can compete with high-level *ab initio* approaches such as DFT/TD-DFT providing reasonable geometries and excitation energies (particular for the Q-bands), discussed in detail in Chapter 3.4.[55, 258, 295-299] The only potentially major geometric problem found with the optimisations here, is that the acetyl dihedral angles in the model A calculations for the 3EOJ structures, have so far been consistently 37-38° out-of-the-plane compared to high-level DFT calculations that find an angle of ~7°. This may explain why the experiment calculates a structurally induced red-shift to result from the protein's interaction with the acetyl, which if would possibly have made more sense if a blue-shift was observed as a result of, for example, out-of-the plane rotation by an H-bond. Even so, this is not a major problem, provided it is given due consideration, if the main focus is actually the inter-site differences. As a general note, the relative difference will benefit significantly from cancellation of errors; therefore a discussion based on these should be reasonably reliable.

8 Conclusions and Future Work

8.1 Follow-up Studies

Following the completion of some of the individual research projects presented in this thesis, relevant data became available in the literature that could not be incorporated due to other priorities and the constraints of time. This was also the case with some potentially good ideas with respect to possible follow-up studies. In order to prevent these ideas from disappearing into the abyss of forgotten memories, a few of the points on the author's wish list of what could be done next are described in brief below.

8.1.1 NSD Analysis of New RC Crystal Structures

A few new *B. viridis*[399-401] and *R. sphaeroides*[402-404] RC crystal structures have been published since the original analysis[60] of the RC pigment conformations was performed (Table 8-1 and Table 8-2). In particular, there are four new structures available for the *B. viridis* RC, two of which are at high-resolution that, could be combined with the previous structures in a repeat of the original formal analysis, whilst the 12 new moderate resolution *R. sphaeroides* structures could be treated both separately as well as combined with the original data. The analysis of these structures NSDs using the statistical approaches applied to the original crystal structure set, would serve as a useful confirmatory test of the conclusions that were provided earlier regarding the precise cofactor conformations.

Table 8-1: Crystal structures of the *B. viridis* RC available in the PDB that have not been analysed *via* NSD.

PDB ID	Space Group	Res.	year	Structure Title	Ref.
2X5U	$P2_12_12$	3.00		80 microsecond laue diffraction snapshot from crystals of a photosynthetic reaction centre without illumination.	399
2X5V	$P2_12_12$	3.00		80 microsecond laue diffraction snapshot from crystals of a photosynthetic reaction centre 3 millisecond following photoactivation.	399
3T6D	$P4_32_12$	1.95		Crystal Structure of the Reaction Centre from Blastochloris viridis strain DSM 133 (ATCC 19567) substrain-08	400
3T6E	$P4_32_12$	1.92		Crystal Structure of the Reaction Centre from Blastochloris viridis strain DSM 133 (ATCC 19567) substrain-94	400
4AC5	$P2_12_12_1$	8.20		Lipidic sponge phase crystal structure of the Bl. viridis reaction centre solved using serial femtosecond crystallography	401

Table 8-2: Crystal structures of the *R. sphaeroides* RC in the PDB that have not been analysed *via* NSD.

PDB ID	Space Group	Res.	Year	Structure title	Ref.
3I4D	$P3_121$	2.01	2010	Photosynthetic reaction centre from rhodobacter sphaeroides 2.4.1	NA
3V3Y	$P3_121$	2.80	2012	Photosynthetic Reaction Centre From Rhodobacter Sphaeroides strain RV	403
3V3Z	$P3_121$	2.90	2012	I(L177)H mutant structure of photosynthetic reaction centre from Rhodobacter sphaeroides	403
3ZUM	$P3_121$	2.50	2011	Photosynthetic reaction centre mutant with phe l146 replaced with ala	402
3ZUW	$P3_121$	2.31	2011	Photosynthetic reaction centre mutant with tyr l128 replaced with his	402
4H99	$P3_121$	2.97	2013	Bacterial Photosynthetic Reaction Centre from Rhodobacter sphaeroides with ILE M265 replaced with THR	NA
4H9L	$P3_121$	2.77	2013	Bacterial Photosynthetic Reaction Centre from Rhodobacter sphaeroides with ILE M265 replaced with SER	NA
4HBH	$P3_121$	2.93	2013	Bacterial Photosynthetic Reaction Centre from Rhodobacter sphaeroides with ILE M265 replaced with ASN	NA
4HBJ	$P3_121$	2.74	2013	Bacterial Photosynthetic Reaction Centre from Rhodobacter sphaeroides with ILE M265 replaced with GLN	NA
4IN5	$P3_121$	2.20	2013	(M)L214G mutant of the Rhodobacter sphaeroides Reaction Centre	404
4IN6	$P3_121$	2.70	2013	(M)L214A mutant of the Rhodobacter sphaeroides Reaction Centre	404
4IN7	$P3_121$	2.85	2013	(M)L214N mutant of the Rhodobacter sphaeroides Reaction Centre	404

8.1.2 Recalibration of H-NOX E_m / Q_{Fe} Relation for Pure *in silico* Implementation

It was noted in the conclusion of Chapter 6 that the relationship between calculated haem-Fe partial charges (Q_{Fe}) and the conformationally induced potential shift could be used to predict the outcome of point-mutations whose major effects were alterations to the steric environment of the haem. For this to be possible, it would probably be necessary to first re-calibrate the H-NOX midpoint potential *vs.* Q_{Fe} relationship after performing *in situ* optimisations of the H-NOX haems in their respective binding-pockets. This would ensure validity in using Q_{Fe} calculated for a haem in a hypothetical binding pocket, where the conformational effects of a potential residue mutation was predicted *via* geometry optimisation, to predict the "H-NOX potential" and hence to estimate the potential shift of the mutant relative to a specified reference. If the above steps were performed and the method further validated and developed with respect to factors such as spin-state dependencies (*e.g.*, the quintuplet was not assessed) and potential interactions between conformational effects and different axial ligands were explored, then it may be possible to use this procedure, confidently, for *any* haem containing protein.

8.1.3 Absolute Spectral Shifts Induced by Protein Influence on FMO Pigments

In a similar vein to the above, but for different reasons, the optimisation of the FMO BChls in their native environments prior to performing the series of partial optimisations would also be an interesting experiment. In this case, the results for a single BChl could be confidently interpreted as the explicit protein perturbation of a particular molecular component relative to the *in vacuo* absorption, instead of being only suitable for comparative studies. The reason that this is necessitated is that the PM6 method, or any other theoretical formalism used to perform the partial optimisations, will most probably have an intrinsic disposition towards certain conformational parameters that are quite different from the bias of the crystal structure determination. In detail, although the overall conformations of tetrapyrroles change readily as the result of external perturbations, most of the framework bond-lengths are relatively similar between both nonplanar and planar structures. However, a particular theoretical method will have a preference for particular bond-lengths so that an optimisation in the binding-site will allow the structure to attain these values before the partial optimisations and thereby eliminate the effect from the results for individual BChls. In contrast, the comparative study does not require this step because this feature is equivalent in any two experiments and is therefore cancelled out upon taking differences. An additional benefit from this approach is that differences amongst the crystal structures with respect to their intrinsic bias (*i.e.*, systematic differences) should be somewhat mitigated.

8.1.4 Optimising the Experimental Design

The necessity of the full-factorial design could be assessed by 'boot-strapping' fractional versions from the dataset. These fractional designs would consist of specified subsets of the original optimisations that retain the property of disentangling main and low-order interactions at the expense of confounding the high-order interactions that were shown in the full-factorial design to be relatively insignificant. This would allow the development of an optimal trade-off between the number of permutations required to ascertain the desired effects and the computational expense of the theoretical method used for geometry optimisation and the calculation of the excitation energies and allow, for example, a DFT//TD-DFT analogue of the original experiment to be performed.

8.1.5 Exploiting the datasets: From NSD to Electronic Perturbation

Finally, given the fact that a large-number of properties have been calculated for a large number of haems and BChls with varying skeletal conformations, it would be useful to attempt to construct a statistical model relating the model's NSDs directly to the calculated

property. This approach would yield fundamental insight into the effects of specific structural deformations, and crucially their combinations, on important properties such as redox potentials and electronic absorptions. Additionally, if such a model could be reliably constructed, it would allow direct prediction of the conformational perturbation of these properties, at the employed level of theory, from only NSD data.

8.2 Applications

8.2.1 Methodological Impact: Validation of Ligand Conformations in the PDB

The data-driven critique of the tetrapyrrole conformations afforded by particular crystal structures could serve as a prototype for the validation of small-molecule structures found in the PDB. This would be especially useful as an alternative method for ligand validation in older structures for which structure factors have not been deposited. Additionally, the use of crystallographic co-ordinates in modern theoretical studies of protein function is increasing and should one wish to assess the electronic structure of a molecule in its experimental conformation (*i.e.* without QMM optimisation) structure selection is a key consideration. A common approach is to choose the highest resolution, most recent or in some cases the most well-known structure. However, these criteria are not necessarily relevant to the intended use of the structure especially in cases when the purpose of the original crystallographic study was different from that which the theoretical study intends to address. Another important point is that small-changes in bond-lengths and angles can have a disproportionate impact on the outcome of a theoretical calculation. It is important therefore to give adequate thought to the relevance of the structure to the study in hand and if possible to perform an unbiased (quantitative) critique and comparison of structures to ensure that this source or error is minimised. In summary, first systematic differences should be accounted for and then aggregation of related structures by averaging should be performed in order to minimise the random error of the experiments.

8.2.2 Conformational Design in Bioengineering and Other Applications

The endeavours of chemists who make new molecules are often driven by the need to design a compound with specific properties tailored to suit an intended application and a useful strategy in this regard may be the enhancement or impediment of an already existing property of an existing molecule (*i.e.*, fine-tuning). Often, when a particular application is in mind, the interest lies in a subset of physicochemical properties that are of critical importance because they are linked to the desired action, but there may also be additional constraints on other properties. Thus, the development of compounds with improved

performance in applications is, in essence, a problem of multiple optimisation where desired and unwanted properties are both linked to structure. This is the challenge: how does one optimise a few properties simultaneously whilst ameliorating unwanted effects when the methods used in their control are usually closely entangled with both? To illustrate, consider that the target application is a photosensitiser for use in photodynamic therapy (PDT). Amongst many others, two useful properties of a good PDT agent are the possession of red-shifted absorptions and a high yield of $^1O_2$. In a hypothetical universe where the degree of nonplanarity was the only way to modify a PDT agent's properties, one would certainly be able to induce red-shifted absorptions but most likely this would occur with a decrease in the $^1O_2$ yield (see Chapter 1.5.2). As a result it is a general necessity that as many means of control are available so as to provide a chemical basis that spans as much 'property space' as possible. Thus, the expansion of the canon of methods for controlling porphyrin properties is naturally beneficial to any applications that benefit from the use of this class of compounds. Indeed, this is also an *a priori* argument for the exploitation of conformational control in nature.

This idea also extends to the rational design or modification of proteins, if not even more so. In the introduction to this thesis, three conclusive reports of the successful modification of protein-bound porphyrin cofactor conformations to affect altered properties were mentioned; the redox modulation of the H-NOX complex *via* site-specific mutations chosen to allow the haem to relax into a less nonplanar conformation[121], the retardation of the haem degrading activity of IsdI by a mutation that decreased *ruf*[215] and the directed evolution of ferrochelatase Ni-variants to show further specificity for Ni that was associated with decreased *sad* distortion.[188] Whilst the latter is not strictly an example of a designed conformational change, even though it was intended, the H-NOX and IsdI mutants certainly are and both show that, in a controlled experimental situation, modification of the conformation of a protein-bound porphyrin effects substantially altered properties. It is considered that with further development, the research presented in this thesis could contribute to the provision of a framework in which the effects of conformational control could be efficiently exploited in protein engineering.

8.3 Summary and Concluding Remarks

This thesis has attempted to address some specific questions concerning the biological importance of protein-induced variation in the macrocycle conformations of haem, chlorophyll and bacteriochlorophyll cofactors in protein-complexes involved in photosynthesis. In order to develop this idea in context, a brief highlight of the biological

diversity of these cofactors was presented to illustrate some of the ways that proteins may modulate their intrinsic properties in order to affect desired function. Subsequently, a review of the many studies that have associated the unique physicochemical properties of substituted nonplanar porphyrins explicitly with their conformations was presented since these compounds serve as models for the conformational diversity exhibited by protein-bound cofactors. Additionally, consideration of these results alongside the observation that haem conformations were conserved across related haemoproteins was suggested to imply a functional role for macrocycle deformation in proteins and some contemporary studies in concordance with this view were discussed. A selection of photosynthetic proteins were then described in some detail to demonstrate that the fine-tuning of haem and chlorophyll physicochemical properties *via* protein-interactions is especially important here so that they represent ideal systems to look for conformational effects.

The first investigation involved the analysis of the reaction centres (RCs) from the purple bacteria *R. sphaeroides* and *B. viridis*. It was first shown that the majority of the available crystal structure determinations were consistent with the presence of distinct macrocycle conformations for some of the RC BChls in both species, but that there was systematic variation between experiments that had to be accounted for in order to obtain reliable results. This led to the development of an approach to identify such systematic variation and the subsequent empirical determination of a few underlying factors that had significant influence on the conformations reported by each experiment. Specifically, some of the crystal structures of the *R. sphaeroides* RC were shown to possess conformations that were simply replicated from other structures (*i.e.*, rigid-body refinement). Other structures were found to possess planar conformations for all BChls and BPheos and this was suggested to be the result of heavily-weighted planarity restraints. The final distinction between structures was that a small group implied unusually contracted macrocycle cores for the RC cofactors that were considered inconsistent with the Mg coordination of BChl. The remaining structures, that were not considered outliers, were then used to provide averaged conformations for each cofactor.

These averaged conformations provided the first statistically reliable evidence that some of the RC BChl cofactors did possess distinct conformations and some of these conformations were conserved across both species. In detail, the special-pair cofactors were conformationally distinct from one another and were typically the most nonplanar BChls, which was attributed to the close interaction of these chromophores. The general feature of their nonplanarity was suggested to assist in their ready photo-oxidation and the

asymmetry considered a possible contributor to the RC electron transfer uni-directionality. Amongst the first accessories, the active branch BChl (B_A) was relatively planar in both species, a factor that was suggested to be useful for its ready reduction, whilst the inactive branch BChl (B_B) in *R. sphaeroides* was highly *ruf* distorted and that of *B. viridis* was relatively planar. The distortion of this BChl in *R. sphaeroides* was suggested to contribute to efficient photo-protection in this species whilst its absence in *B. viridis* was consistent with its inability to grow photosynthetically under anaerobic conditions. The conformations of the BPheos did not appear to be distinguishable and were not ascribed functional significance.

Next, the statistical methodology that was developed during the course of the above research was applied to determine best estimates of the haem conformations in the RC-cyt from *B. viridis*. These conformations were then employed to quantitatively calculate the conformational contribution to the interhaem redox differences *via* a calibrated computational approach that was developed using the data from the recent H-NOX model[121] for conformational control in haemoproteins. The result was that an improved agreement between the state-of-the-art *ab initio* calculations of Voigt and Knapp[99] and the experimental interhaem potential differences was achieved after incorporation of the explicit conformational effect. Furthermore, the effect of conformational control was found to account for 54 mV of a particular interhaem potential difference. As a corollary, the successful calibration using the H-NOX data proved definitively that the observed correlation between macrocycle deformation and the mutant complexes' midpoint potentials was a truly causal relationship.

The final study then employed a relatively unique experimental design in order to attempt to address the question of potential interactions between the macrocycle conformation and the local conformations of the peripheral and conjugated substituents of BChls with respect to their influences on the Q_y excitation energies of these chromophores. This study was conducted in the context of the observation that the BChl pigments of the FMO protein possess different conformations that are likely to contribute to the site-energy distribution in this protein. The initial results demonstrated the validity of the concept and the power of the experimental design and it is expected that once this project is completed, a definitive answer to the questions of what conformational factors contribute to BChl absorption shifts in general, as well as how these effect the site-energy distribution in the FMO complex.

Thus, possible roles for conformational control have been suggested for three important stages of photosynthesis; namely, charge-separation, re-reduction of the special-pair and

light-harvesting. In general, these results are relevant to all biological processes that involve tetrapyrroles, as they serve to illustrate the variety of properties and functions that may be modulated by means of the protein-induced conformational perturbation of cofactor geometry. Additionally, the procedures developed herein, with respect to the quantum molecular modelling studies, as well as statistical approach to afford unbiased analysis of structural data, highlight the advantages of combining tools from different fields and the rigour provided by such numerical techniques. It is possible that these methods could be transferred to other problems of chemical and biological significance. However, to return to the main point, earlier on in this work, it was highlighted how the various photosynthetic processes, *"all involve and depend upon numerous tetrapyrrole cofactors"*, it is hoped that the author has at least intrigued the reader to the notion that this is, in part, due to the inherent, no pun intended, flexibility of the porphyrin macrocycle.

References

1. M. Bergmann and C. Niemann, *Annu. Rev. Biochem.*, 1938, **7**, 99-124.
2. K. Zeile and H. Hellstrom, *Hoppe-Seyler's Z. Physiol. Chem.*, 1930, **192**, 171-92.
3. A. G. Murzin, S. E. Brenner, T. Hubbard and C. Chothia, *J. Mol. Biol.*, 1995, **247**, 536-40.
4. A. Andreeva, D. Howorth, J. M. Chandonia, S. E. Brenner, T. J. Hubbard, C. Chothia and A. G. Murzin, *Nucleic Acids Res.*, 2008, **36**, D419-25.
5. B. H. Dessailly, N. L. Dawson, K. Mizuguchi and C. A. Orengo, *Biochim. Biophys. Acta*, 2013, **1834**, 874-89.
6. L. R. Milgrom, *The Colours of Life: An Introduction to the Chemistry of Porphyrins and Related Compounds*, Oxford University Press, Oxford, 1997.
7. K. Gruber, B. Puffer and B. Kräutler, *Chemical Society Reviews*, 2011, **40**, 4346-63.
8. B. Kräutler and B. Puffer, in *Handbook of Porphyrin Science*, eds. K. M. Kadish, K. M. Smith and R. Guilard, World Scientific, Singapore, 2012, vol. 25, pp. 131-263.
9. L. N. Todd and M. Zimmer, *Inorg. Chem.*, 2002, **41**, 6831-7.
10. Y. Song, R. E. Haddad, S. L. Jia, S. Hok, M. M. Olmstead, D. J. Nurco, N. E. Schore, J. Zhang, J. G. Ma, K. M. Smith, S. Gazeau, J. Pecaut, J. C. Marchon, C. J. Medforth and J. A. Shelnutt, *J. Am. Chem. Soc.*, 2005, **127**, 1179-92.
11. C. Mbofana and M. Zimmer, *Inorg. Chem.*, 2006, **45**, 2598-602.
12. E. C. Duin and M. L. McKee, *J. Phys. Chem. B*, 2008, **112**, 2466-82.
13. R. Sarangi, M. Dey and S. W. Ragsdale, *Biochemistry*, 2009, **48**, 3146-56.
14. P. E. Cedervall, M. Dey, A. R. Pearson, S. W. Ragsdale and C. M. Wilmot, *Biochemistry*, 2010, **49**, 7683-93.
15. S. Ebner, B. Jaun, M. Goenrich, R. K. Thauer and J. Harmer, *J. Am. Chem. Soc.*, 2010, **132**, 567-75.
16. P. E. Cedervall, M. Dey, X. Li, R. Sarangi, B. Hedman, S. W. Ragsdale and C. M. Wilmot, *J. Am. Chem. Soc.*, 2011, **133**, 5626-8.
17. T. Matsui, S. Nambu, Y. Ono, C. W. Goulding, K. Tsumoto and M. Ikeda-Saito, *Biochemistry*, 2013, **52**, 3025-7.
18. S. K. Chapman, S. Daff and A. W. Munro, in *Metal Sites in Proteins and Models Iron Centres*, eds. H. A. O. Hill, P. J. Sadler and A. J. Thomson, Springer, Berlin, 1997, vol. 88, pp. 39-70.
19. J. P. Allen and J. C. Williams, in *Chlorophylls and Bacteriochlorophylls*, eds. B. Grimm, R. J. Porra, W. Rüdiger and H. Scheer, Springer, Dordrecht, 2006, vol., pp. 283-95.
20. S. Jones and J. M. Thornton, *Proc. Natl. Acad. Sci. U.S.A.*, 1996, **93**, 13-20.
21. J. Deisenhofer, O. Epp, K. Miki, R. Huber and H. Michel, *J. Mol. Biol.*, 1984, **180**, 385-98.
22. J. Deisenhofer, O. Epp, K. Miki, R. Huber and H. Michel, *Nature*, 1985, **318**, 618-24.
23. R. H. Felton, in *The Porphyrins*, ed. D. Dolphin, Academic Press, New York, 1978, vol. 5, pp. 53-125.
24. C. E. Castro, in *The Porphyrins*, ed. D. Dolphin, Academic Press, New York, 1978, vol. 5, pp. 1-27.
25. D. Mauzerall, in *The Porphyrins*, ed. D. Dolphin, Academic Press, New York, 1978, vol. 5, pp. 29-52.
26. M. P. Johansson, M. R. A. Blomberg, D. Sundholm and M. Wikström, *Biochim. Biophys. Acta.*, 2002, **1553**, 183-7.
27. M. Gouterman, *Journal of Molecular Spectroscopy*, 1961, **6**, 138-63.
28. M. Gouterman, G. H. Wagnière and L. C. Snyder, *Journal of Molecular Spectroscopy*, 1963, **11**, 108-27.
29. M. Gouterman, in *The Porphyrins*, ed. D. Dolphin, Academic Press, New York, 1978, vol. 3, pp. 1-165.
30. M. O. Senge, *Journal of Photochemistry and Photobiology B: Biology*, 1992, **16**, 3-36.
31. J. D. Hobbs and J. A. Shelnutt, *J. Protein Chem.*, 1995, **14**, 19-25.
32. W. Jentzen, X.-Z. Song and J. A. Shelnutt, *J. Phys. Chem. B*, 1997, **101**, 1684-99.
33. J. A. Shelnutt, X.-Z. Song, J.-G. Ma, S.-L. Jia, W. Jentzen and C. J. Medforth, *Chemical Society Reviews*, 1998, **27**, 31.
34. W. Jentzen, J.-G. Ma and J. A. Shelnutt, *Biophys. J.*, 1998, **74**, 753-63.
35. K. M. Kadish, F. D'Souza, A. Villard, M. Autret, E. Van Caemelbecke, P. Bianco, A. Antonini and P. Tagliatesta, *Inorg. Chem.*, 1994, **33**, 5169-70.
36. R. E. Haddad, S. Gazeau, J. Pecaut, J. C. Marchon, C. J. Medforth and J. A. Shelnutt, *J. Am. Chem. Soc.*, 2003, **125**, 1253-68.
37. M. O. Senge, in *The Porphyrin Handbook*, eds. K. M. Kadish, K. M. Smith and R. Guilard, Academic Press, San Diego, 2000, vol. 1, pp. 239-347.
38. M. O. Senge, *Chem. Commun.*, 2006, 243-56.
39. M. O. Senge, M. W. Renner, W. W. Kalisch and J. Fajer, *J. Chem. Soc., Dalton Trans.*, 2000, 381-5.

40. B. Roder, M. Buchner, I. Ruckmann and M. O. Senge, *Photochemical & Photobiological Sciences*, 2010, **9**, 1152-8.
41. D. E. Koshland, *Proc. Natl. Acad. Sci. U.S.A.*, 1958 **44**, 98-104.
42. D. E. Koshland, *Angew. Chem. Int. Ed. Engl.*, 1995, **33**, 2375-8.
43. E. Fischer, *Berichte der Deutschen Chemischen Gesellschaft*, 1894, **27**, 2985-93.
44. L. Pauling, *Nature*, 1948, **161**, 707-9.
45. T. L. Amyes and J. P. Richard, *Biochemistry*, 2013, **52**, 2021-35.
46. A. J. Hoff and J. Deisenhofer, *Physics Reports*, 1997, **287**, 1-247.
47. J. Deisenhofer and H. Michel, *Biosci. Rep.*, 2004, **24**, 323-61.
48. R. E. Fenna and B. W. Matthews, *Nature*, 1975, **258**, 573-7.
49. B. W. Matthews, R. E. Fenna, M. C. Bolognesi, M. F. Schmid and J. M. Olson, *J. Mol. Biol.*, 1979, **131**, 259-85.
50. T. Renger, *Photosynth. Res.*, 2009, **102**, 471-85.
51. M. Sener, J. Strumpfer, J. Hsin, D. Chandler, S. Scheuring, C. N. Hunter and K. Schulten, *ChemPhysChem*, 2011, **12**, 518-31.
52. H. L. Axelrod, E. C. Abresch, M. Y. Okamura, A. P. Yeh, D. C. Rees and G. Feher, *J. Mol. Biol.*, 2002, **319**, 501-15.
53. D. Gust, T. A. Moore and A. L. Moore, *Acc. Chem. Res.*, 2009, **42**, 1890-8.
54. C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, Wiley, West Sussex, England, 2004.
55. J. Linnanto and J. Korppi-Tommola, *PCCP*, 2006, **8**, 663-87.
56. A. Schmidt, M. Teeter, E. Weckert and V. S. Lamzin, *Acta Crystallographica Section F: Structural Biology and Crystallization* 2011, **67**, 424-8.
57. J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff and D. C. Phillips, *Nature*, 1958, **181**, 662-6.
58. J. C. Kendrew, *Science*, 1963, **139**, 1259-66.
59. G. J. Kleywegt, *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, 2007, **63**, 94-100.
60. S. A. MacGowan and M. O. Senge, *Chem. Commun.*, 2011, **47**, 11621-3.
61. S. A. MacGowan and M. O. Senge, *Inorg. Chem.*, 2013, **52**, 1228-37.
62. G. P. Moss, *Pure Appl. Chem.*, 1987, **59**, 779-832.
63. A. Tulinsky, *Annals of the New York Academy of Sciences*, 1973, **206**, 47-69.
64. W. R. Scheidt, 2012, **25**, 1-179.
65. T. M. Krygowski and M. K. Cyrański, *Chem. Rev.*, 2001, **101**, 1385-420.
66. M. K. Cyrañski, T. M. Krygowski, M. Wisiorowski, N. J. R. van Eikema Hommes and P. v. R. Schleyer, *Angew. Chem. Int. Ed. Engl.*, 1998, **37**, 177-80.
67. E. Steiner, A. Soncini and P. W. Fowler, *Organic & biomolecular chemistry*, 2005, **3**, 4053-9.
68. M. Bröring, *Angew. Chem. Int. Ed. Engl.*, 2011, **50**, 2436-8.
69. J. I. Wu, I. Fernandez and P. Schleyer, *J. Am. Chem. Soc.*, 2013, **135**, 315-21.
70. H. Du, R.-C. A. Fuh, J. Li, L. A. Corkan and J. S. Lindsey, *Photochem. Photobiol.*, 1998, **68**, 141-2.
71. J. M. Dixon, M. Taniguchi and J. S. Lindsey, *Photochem. Photobiol.*, 2005, **81**, 212-3.
72. C. J. Reedy and B. R. Gibney, *Chem. Rev.*, 2004, **104**, 617-49.
73. A. W. Munro, H. M. Girvan, K. J. McLean, M. R. Cheesman and D. Leys, in *Tetrapyrroles - Birth, Life and Death*, eds. M. Warren and A. Smith, Springer, New York, 2009, vol., pp. 160-83.
74. B. Grimm, R. J. Porra, W. Rüdiger and H. Scheer, eds., *Chlorophylls and Bacteriochlorophylls*, Springer, Dordrecht, The Netherlands, 2006.
75. N. Mochizuki, R. Tanaka, B. Grimm, T. Masuda, M. Moulin, A. G. Smith, A. Tanaka and M. J. Terry, *Trends Plant Sci.*, 2010, **15**, 488-98.
76. S. Schneider, J. Marles-Wright, K. H. Sharp and M. Paoli, *Natural product reports*, 2007, **24**, 621-30.
77. I. G. Denisov and S. G. Sligar, 2010, **5**, 165-201.
78. I. Bertini, G. Cavallaro and A. Rosato, *Chem. Rev.*, 2006, **106**, 90-115.
79. J. M. Stevens and S. J. Ferguson, 2012, **20**, 371-401.
80. C. Fufezan, J. Zhang and M. R. Gunner, *Proteins*, 2008, **73**, 690-704.
81. M. F. Perutz, *Nature*, 1970, **228**, 726-34.
82. M. F. Perutz, *Nature*, 1970, **228**, 734-9.
83. M. F. Perutz, A. J. Wilkinson, M. Paoli and G. G. Dodson, *Annu. Rev. Biophys. Biomol. Struct.*, 1998, **27**, 1-34.
84. Y. Hatefi, *Annu. Rev. Biochem.*, 1985, **54**, 1015-69.
85. C. Hägerhäll, *Biochim. Biophys. Acta.*, 1997, **1320**, 107-41.
86. A. R. Crofts, J. T. Holland, D. Victoria, D. R. Kolling, S. A. Dikanov, R. Gilbreth, S. Lhee, R. Kuras and M. G. Kuras, *Biochim. Biophys. Acta.*, 2008, **1777**, 1001-19.
87. H. Michel, J. Behr, A. Harrenga and A. Kannt, *Annu. Rev. Biophys. Biomol. Struct.*, 1998, **27**, 329-56.

88. C. J. Reedy, M. M. Elvekrog and B. R. Gibney, *Nucleic Acids Res.*, 2008, **36**, D307-13.
89. Z. Zheng and M. R. Gunner, *Proteins*, 2009, **75**, 719-34.
90. A. L. Raphael and H. B. Gray, *Proteins*, 1989, **6**, 338-40.
91. I. Mus-Veteau, A. Dolla, F. Guerlesquin, F. Payan, M. Czjzek, R. Haser, P. Bianco, J. Haladjian, B. J. Rapp-Giles and J. D. Wall, *J. Biol. Chem.*, 1992, **267**, 16851-8.
92. G. S. Wilson, *Bioelectrochem. Bioenerg.*, 1974, **1**, 172-9.
93. W. S. Caughey, W. Y. Fujimoto and B. P. Johnson, *Biochemistry*, 1966, **5**, 3830-43.
94. F. A. Walker, M.-W. Lo and M. T. Ree, *J. Am. Chem. Soc.*, 1976, **98**, 5552-60.
95. T. Hayashi, 2010, **5**, 1-69.
96. Y. Sugita and Y. Yoneyama, *The Journal of biological chemistry*, 1971, **246**, 389-94.
97. M. R. Gunner and B. Honig, *Proc. Natl. Acad. Sci. U.S.A.*, 1991, **88**, 9151-5.
98. J. Mao, K. Hauser and M. R. Gunner, *Biochemistry*, 2003, **42**, 9829-40.
99. P. Voigt and E. W. Knapp, *J. Biol. Chem.*, 2003, **278**, 51993-2001.
100. G. Tollin, L. K. Hanson, M. Caffrey, T. E. Meyer and M. A. Cusanovich, *Proc. Natl. Acad. Sci. U.S.A.*, 1986, **83**, 3693-7.
101. R. Huber, *Eur. J. Biochem.*, 1990, **187**, 283-305.
102. K. M. Barkigia, D. S. Gottfried, S. G. Boxer and J. Fajer, *J. Am. Chem. Soc.*, 1989, **111**, 6444-6.
103. D. Mauzerall, *Annals of the New York Academy of Sciences*, 1973, **206**, 483-94.
104. L. O. Bjorn, G. C. Papageorgiou, R. E. Blankenship and Govindjee, *Photosynth. Res.*, 2009, **99**, 85-98.
105. H. Scheer, in *Chlorophylls and Bacteriochlorophylls*, eds. B. Grimm, R. J. Porra, W. Rüdiger and H. Scheer, Springer, Dordrecht, 2006, vol.
106. P. Jordan, P. Fromme, H. T. Witt, O. Klukas, W. Saenger and N. Krauss, *Nature*, 2001, **411**, 909-17.
107. X. Lin, H. A. Murchison, V. Nagarajan, W. W. Parson, J. P. Allen and J. C. Williams, *Proc. Natl. Acad. Sci. U.S.A.*, 1994, **91**, 10265-9.
108. J. Rautter, F. Lendzian, C. Schulz, A. Fetsch, M. Kuhn, X. Lin, J. C. Williams, J. P. Allen and W. Lubitz, *Biochemistry*, 1995, **34**, 8130-43.
109. J. P. Allen, K. Artz, X. Lin, J. C. Williams, A. Ivancich, D. Albouy, T. A. Mattioli, A. Fetsch, M. Kuhn and W. Lubitz, *Biochemistry*, 1996, **35**, 6612-9.
110. H. Ishikita, W. Saenger, J. Biesiadka, B. Loll and E. W. Knapp, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 9855-60.
111. E. Gudowska-Nowak, M. D. Newton and J. Fajer, *J. Phys. Chem.*, 1990, **94**, 5795-801.
112. L. K. Hanson, J. Fajer, M. A. Thompson and M. C. Zerner, *J. Am. Chem. Soc.*, 1987, **109**, 4728-30.
113. H. Treutlein, K. Schulten, A. T. Brünger, M. Karplus, J. Deisenhofer and H. Michel, *Proc. Natl. Acad. Sci. U.S.A.*, 1992, **89**, 75-9.
114. C. Kratky, R. Waditschatka, C. Angst, J. E. Johansen, J. C. Plaquevent, J. Schreiber and A. Eschenmoser, *Helvetica Chimica Acta*, 1985, **68**, 1312-37.
115. W. R. Scheidt and Y. J. Lee, in *Metal Complexes with Tetrapyrrole Ligands I*, ed. J. W. Buchler, Springer, Berlin, 1987, vol. 64, pp. 1-70.
116. S. G. DiMagno, A. K. Wertsching and C. R. Ross, *J. Am. Chem. Soc.*, 1995, **117**, 8279-80.
117. A. B. J. Parusel, T. Wondimagegn and A. Ghosh, *J. Am. Chem. Soc.*, 2000, **122**, 6371-4.
118. A. K. Wertsching, A. S. Koch and S. G. DiMagno, *J. Am. Chem. Soc.*, 2001, **123**, 3932-9.
119. K. M. Barkigia, L. Chantranupong, K. M. Smith and J. Fajer, *J. Am. Chem. Soc.*, 1988, **110**, 7566-7.
120. W. Jentzen, M. C. Simpson, J. D. Hobbs, X. Song, T. Ema, N. Y. Nelson, C. J. Medforth, K. M. Smith, M. Veyrat, M. Mazzanti, R. Ramasseul, J.-C. Marchon, T. Takeuchi, W. A. Goddard and J. A. Shelnutt, *J. Am. Chem. Soc.*, 1995, **117**, 11085-97.
121. C. Olea, Jr., J. Kuriyan and M. A. Marletta, *J. Am. Chem. Soc.*, 2010, **132**, 12794-5.
122. E. B. Fleischer, *J. Am. Chem. Soc.*, 1963, **85**, 146-8.
123. E. B. Fleischer, *J. Am. Chem. Soc.*, 1963, **85**, 1353-4.
124. E. B. Fleischer, C. K. Miller and L. E. Webb, *J. Am. Chem. Soc.*, 1964, **86**, 2342-7.
125. J. L. Hoard, M. J. Hamor and T. A. Hamor, *J. Am. Chem. Soc.*, 1963, **85**, 2334-5.
126. M. J. Hamor, T. A. Hamor and J. L. Hoard, *J. Am. Chem. Soc.*, 1964, **86**, 1938-42.
127. J. M. Robertson, *Journal of the Chemical Society (Resumed)*, 1935, 615.
128. J. M. Robertson, *Journal of the Chemical Society (Resumed)*, 1936, 1195.
129. J. M. Robertson and I. Woodward, *Journal of the Chemical Society (Resumed)*, 1937, 219.
130. S. J. Silvers and A. Tulinsky, *J. Am. Chem. Soc.*, 1967, **89**, 3331-7.
131. P. W. Codding and A. Tulinsky, *J. Am. Chem. Soc.*, 1972, **94**, 4151-7.
132. B. M. L. Chen and A. Tulinsky, *J. Am. Chem. Soc.*, 1972, **94**, 4144-51.
133. J. W. Lauher and J. A. Ibers, *J. Am. Chem. Soc.*, 1973, **95**, 5148-52.
134. E. B. Fleischer, *Acc. Chem. Res.*, 1970, **3**, 105-12.
135. E. B. Fleischer and A. L. Stone, *Chem. Commun.*, 1967, 332.
136. A. Stone and E. B. Fleischer, *J. Am. Chem. Soc.*, 1968, **90**, 2735-48.
137. J. L. Hoard, M. J. Hamor, T. A. Hamor and W. S. Caughey, *J. Am. Chem. Soc.*, 1965, **87**, 2312-9.

138. T. A. Hamor, W. S. Caughey and J. L. Hoard, *J. Am. Chem. Soc.*, 1965, **87**, 2305-12.
139. R. P. Countryman, D. M. Collins and J. L. Hoard, *J. Am. Chem. Soc.*, 1969, **91**, 5166-7.
140. L. J. Radonovich, A. Bloom and J. L. Hoard, *J. Am. Chem. Soc.*, 1972, **94**, 2073-8.
141. W. A. Hendrickson and W. E. Love, *Nature New Biology*, 1971, **232**, 197-203.
142. W. A. Hendrickson, W. E. Love and J. Karle, *J. Mol. Biol.*, 1973, **74**, 331-61.
143. D. M. Collins, R. Countryman and J. L. Hoard, *J. Am. Chem. Soc.*, 1972, **94**, 2066-72.
144. W. R. Scheidt, *Acc. Chem. Res.*, 1977, **10**, 339-45.
145. W. R. Scheidt and C. A. Reed, *Chem. Rev.*, 1981, **81**, 543-55.
146. E. F. Meyer Jnr, *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*, 1972, **28**, 2162-7.
147. D. L. Cullen and E. F. Meyer, *J. Am. Chem. Soc.*, 1974, **96**, 2095-102.
148. A. Ulman, J. Gallucci, D. Fisher and J. A. Ibers, *J. Am. Chem. Soc.*, 1980, **102**, 6852-4.
149. C. Kratky, C. Angst and J. E. Johansen, *Angew. Chem. Int. Ed. Engl.*, 1981, **20**, 211-2.
150. T. D. Brennan, W. R. Scheidt and J. A. Shelnutt, *J. Am. Chem. Soc.*, 1988, **110**, 3919-24.
151. A. M. Stolzenberg, P. A. Glazer and B. M. Foxman, *Inorg. Chem.*, 1986, **25**, 983-91.
152. S. H. Strauss, M. E. Silver, K. M. Long, R. G. Thompson, R. A. Hudgens, K. Spartalian and J. A. Ibers, *J. Am. Chem. Soc.*, 1985, **107**, 4207-15.
153. M. O. Senge, W. W. Kalisch and S. Runge, *Tetrahedron*, 1998, **54**, 3781-98.
154. K. M. Barkigia, M. A. Thompson, J. Fajer, R. K. Pandey, K. M. Smith and M. G. H. Vicente, *New J. Chem.*, 1992, **16**.
155. W. W. Kalisch, M. O. Senge and K. Ruhlandt-Senge, *Photochem. Photobiol.*, 1998, **67**, 312-23.
156. K. M. Shea, L. Jaquinod, K. G. Richard and K. M. Smith, *Chem. Commun.*, 1998, 759-60.
157. M. O. Senge and S. A. MacGowan, in *Synthesis and Structural Studies*, eds. K. M. Kadish, K. M. Smith and R. Guilard, World Scientific, Singapore, 2011, vol. 13, pp. 253-97.
158. C. J. Medforth, M. O. Senge, K. M. Smith, L. D. Sparks and J. A. Shelnutt, *J. Am. Chem. Soc.*, 1992, **114**, 9859-69.
159. L. D. Sparks, C. J. Medforth, M. S. Park, J. R. Chamberlain, M. R. Ondrias, M. O. Senge, K. M. Smith and J. A. Shelnutt, *J. Am. Chem. Soc.*, 1993, **115**, 581-92.
160. W. Jentzen, E. Unger, G. Karvounis, J. A. Shelnutt, W. Dreybrodt and R. Schweitzer-Stenner, *J. Phys. Chem.*, 1996, **100**, 14184-91.
161. K. M. Barkigia, M. D. Berber, J. Fajer, C. J. Medforth, M. W. Renner and K. M. Smith, *J. Am. Chem. Soc.*, 1990, **112**, 8851-7.
162. C. J. Medforth, C. M. Muzzi, K. M. Shea, K. M. Smith, R. J. Abraham, S. Jia and J. A. Shelnutt, *Journal of the Chemical Society, Perkin Transactions 2*, 1997, 839-44.
163. K. M. Barkigia, M. W. Renner, L. R. Furenlid, C. J. Medforth, K. M. Smith and J. Fajer, *J. Am. Chem. Soc.*, 1993, **115**, 3627-35.
164. K. M. Kadish, J. Li, E. Van Caemelbecke, Z. Ou, N. Guo, M. Autret, F. D'Souza and P. Tagliatesta, *Inorg. Chem.*, 1997, **36**, 6292-8.
165. K. M. Kadish, E. Van Caemelbecke, F. D'Souza, M. Lin, D. J. Nurco, C. J. Medforth, T. P. Forsyth, B. Krattinger, K. M. Smith, S. Fukuzumi, I. Nakanishi and J. A. Shelnutt, *Inorg. Chem.*, 1999, **38**, 2188-98.
166. P. Ochsenbein, K. Ayougou, D. Mandon, J. Fischer, R. Weiss, R. N. Austin, K. Jayaraj, A. Gold, J. Terner and J. Fajer, *Angew. Chem. Int. Ed. Engl.*, 1994, **33**, 348-50.
167. D. Mandon, P. Ochenbein, J. Fischer, R. Weiss, K. Jayaraj, R. N. Austin, A. Gold, P. S. White and O. Brigaud, *Inorg. Chem.*, 1992, **31**, 2044-9.
168. S. Gentemann, C. J. Medforth, T. P. Forsyth, D. J. Nurco, K. M. Smith, J. Fajer and D. Holten, *J. Am. Chem. Soc.*, 1994, **116**, 7363-8.
169. N. C. Maiti and M. Ravikanth, *Journal of the Chemical Society, Faraday Transactions*, 1995, **91**, 4369.
170. C. M. Drain, C. Kirmaier, C. J. Medforth, D. J. Nurco, K. M. Smith and D. Holten, *J. Phys. Chem.*, 1996, **100**, 11984-93.
171. J. L. Retsek, C. J. Medforth, D. J. Nurco, S. Gentemann, V. S. Chirvony, K. M. Smith and D. Holten, *J. Phys. Chem. B*, 2001, **105**, 6396-411.
172. J. L. Retsek, C. M. Drain, C. Kirmaier, D. J. Nurco, C. J. Medforth, K. M. Smith, I. V. Sazanovich, V. S. Chirvony, J. Fajer and D. Holten, *J. Am. Chem. Soc.*, 2003, **125**, 9787-800.
173. A. Y. Lebedev, M. A. Filatov, A. V. Cheprakov and S. A. Vinogradov, *J. Phys. Chem. A*, 2008, **112**, 7723-33.
174. W. Chen, M. E. El-Khouly and S. Fukuzumi, *Inorg. Chem.*, 2010.
175. F. Nifiatis, W. Su, J. E. Haley, J. E. Slagle and T. M. Cooper, *J. Phys. Chem. A*, 2011, **115**, 13764-72.
176. M. Zawadzka, J. Wang, W. J. Blau and M. O. Senge, *J. Phys. Chem. A*, 2013, **117**, 15-26.
177. S. Gentemann, N. Y. Nelson, L. Jaquinod, D. J. Nurco, S. H. Leung, C. J. Medforth, K. M. Smith, J. Fajer and D. Holten, *J. Phys. Chem. B*, 1997, **101**, 1247-54.

178. S. Gentemann, C. J. Medforth, T. Ema, N. Y. Nelson, K. M. Smith, J. Fajer and D. Holten, *Chemical Physics Letters*, 1995, **245**, 441-7.
179. N. C. Maiti and M. Ravikanth, *Journal of the Chemical Society, Faraday Transactions*, 1996, **92**, 1095.
180. C. M. Drain, S. Gentemann, J. A. Roberts, N. Y. Nelson, C. J. Medforth, S. Jia, M. C. Simpson, K. M. Smith, J. Fajer, J. A. Shelnutt and D. Holten, *J. Am. Chem. Soc.*, 1998, **120**, 3781-91.
181. W. W. Kalisch and M. O. Senge, *Tetrahedron Lett.*, 1996, **37**, 1183-6.
182. M. O. Senge and W. W. Kalisch, *Inorg. Chem.*, 1997, **36**, 6103-16.
183. K. K. Anderson, J. D. Hobbs, L. Luo, K. D. Stanley, J. M. E. Quirke and J. A. Shelnutt, *J. Am. Chem. Soc.*, 1993, **115**, 12346-52.
184. X.-Z. Song, W. Jentzen, S.-L. Jia, L. Jaquinod, D. J. Nurco, C. J. Medforth, K. M. Smith and J. A. Shelnutt, *J. Am. Chem. Soc.*, 1996, **118**, 12975-88.
185. R. Franco, J.-G. Ma, Y. Lu, G. C. Ferreira and J. A. Shelnutt, *Biochemistry*, 2000, **39**, 2517-29.
186. Z. Shi, R. Franco, R. Haddad, J. A. Shelnutt and G. C. Ferreira, *Biochemistry*, 2006, **45**, 2904-12.
187. S. Al-Karadaghi, R. Franco, M. Hansson, J. A. Shelnutt, G. Isaya and G. C. Ferreira, *Trends Biochem. Sci*, 2006, **31**, 135-42.
188. N. R. McIntyre, R. Franco, J. A. Shelnutt and G. C. Ferreira, *Biochemistry*, 2011, **50**, 1535-44.
189. T. Tahara, M. Yamamoto, R. Akagi, H. Harigae and S. Taketani, *Int. J. Hematol.*, 2010, **92**, 769-71.
190. M. J. Bain-Ackerman and D. K. Lavallee, *Inorg. Chem.*, 1979, **18**, 3358-64.
191. H. A. Dailey and J. E. Fleming, *The Journal of biological chemistry*, 1983, **258**, 11453-9.
192. A. G. Cochran and P. G. Schultz, *Science*, 1990, **249**, 781-3.
193. D. Lecerof, M. Fodje, A. Hansson, M. Hansson and S. Al-Karadaghi, *J. Mol. Biol.*, 2000, **297**, 221-32.
194. E. Sigfridsson and U. Ryde, *Journal of Biological Inorganic Chemistry*, 2003, **8**, 273-82.
195. S. Venkateshrao, J. Yin, A. A. Jarzecki, P. G. Schultz and T. G. Spiro, *J. Am. Chem. Soc.*, 2004, **126**, 16361-7.
196. A. H. Corwin, *Annals of the New York Academy of Sciences*, 1973, **206**, 201-9.
197. H. Ryeng and A. Ghosh, *J. Am. Chem. Soc.*, 2002, **124**, 8099-103.
198. C. J. Medforth, R. E. Haddad, C. M. Muzzi, N. R. Dooley, L. Jaquinod, D. C. Shyr, D. J. Nurco, M. M. Olmstead, K. M. Smith, J. G. Ma and J. A. Shelnutt, *Inorg. Chem.*, 2003, **42**, 2227-41.
199. L. Plate and M. A. Marletta, *Trends Biochem. Sci*, 2013, **38**, 566-75.
200. M. B. Winter, E. J. McLaurin, S. Y. Reece, C. Olea, Jr., D. G. Nocera and M. A. Marletta, *J. Am. Chem. Soc.*, 2010, **132**, 5582-3.
201. M. B. Winter, P. J. Klemm, C. M. Phillips-Piro, K. N. Raymond and M. A. Marletta, *Inorg. Chem.*, 2013, **52**, 2277-9.
202. P. Pellicena, D. S. Karow, E. M. Boon, M. A. Marletta and J. Kuriyan, *Proc. Natl. Acad. Sci. U.S.A.*, 2004, **101**, 12854-9.
203. C. Olea, E. M. Boon, P. Pellicena, J. Kuriyan and M. A. Marletta, *ACS chemical biology*, 2008, **3**, 703-10.
204. R. Tran, E. M. Boon, M. A. Marletta and R. A. Mathies, *Biochemistry*, 2009, **48**, 8568-77.
205. E. M. Boon and M. A. Marletta, *J. Am. Chem. Soc.*, 2006, **128**, 10022-3.
206. Z. Dai and E. M. Boon, *J. Am. Chem. Soc.*, 2010, **132**, 11496-503.
207. Z. Dai and E. M. Boon, *J. Inorg. Biochem.*, 2011, **105**, 784-92.
208. S. Muralidharan and E. M. Boon, *J. Am. Chem. Soc.*, 2012, **134**, 2044-6.
209. S. E. Bowman and K. L. Bren, *Inorg. Chem.*, 2010, **49**, 7890-7.
210. M. D. Liptak, R. D. Fagerlund, E. C. Ledgerwood, S. M. Wilbanks and K. L. Bren, *J. Am. Chem. Soc.*, 2011, **133**, 1153-5.
211. U. Bren and C. Oostenbrink, *Journal of chemical information and modeling*, 2012, **52**, 1573-82.
212. M. D. Liptak, X. Wen and K. L. Bren, *J. Am. Chem. Soc.*, 2010, **132**, 9753-63.
213. M. Can, G. Zoppellaro, K. K. Andersson and K. L. Bren, *Inorg. Chem.*, 2011, **50**, 12018-24.
214. S. J. Takayama, G. Ukpabi, M. E. Murphy and A. G. Mauk, *Proc. Natl. Acad. Sci. U.S.A.*, 2011, **108**, 13071-6.
215. G. Ukpabi, S. J. Takayama, A. G. Mauk and M. E. Murphy, *The Journal of biological chemistry*, 2012, **287**, 34179-88.
216. D. E. Bikiel, F. Forti, L. Boechi, M. Nardini, F. J. Luque, M. A. Marti and D. A. Estrin, *J. Phys. Chem. B*, 2010, **114**, 8536-43.
217. Q. M. Tran, C. Fong, R. A. Rothery, E. Maklashina, G. Cecchini and J. H. Weiner, *PloS one*, 2012, **7**, e32641.
218. G. Zucchelli, D. Brogioli, A. P. Casazza, F. M. Garlaschi and R. C. Jennings, *Biophys. J.*, 2007, **93**, 2240-54.
219. G. Zucchelli, S. Santabarbara and R. C. Jennings, *Biochemistry*, 2012, **51**, 2717-36.
220. Y. Umena, K. Kawakami, J. R. Shen and N. Kamiya, *Nature*, 2011, **473**, 55-60.

221. K. Saito, Y. Umena, K. Kawakami, J. R. Shen, N. Kamiya and H. Ishikita, *Biochemistry*, 2012, **51**, 4290-9.
222. K. Saito, T. Ishida, M. Sugiura, K. Kawakami, Y. Umena, N. Kamiya, J. R. Shen and H. Ishikita, *J. Am. Chem. Soc.*, 2011, **133**, 14379-88.
223. Y. Shichida and T. Matsuyama, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2009, **364**, 2881-95.
224. O. Weingart, *J. Am. Chem. Soc.*, 2007, **129**, 10618-9.
225. E. W. Hernandez-Rodriguez, E. Sanchez-Garcia, R. Crespo-Otero, A. L. Montero-Alejo, L. A. Montero and W. Thiel, *J. Phys. Chem. B*, 2012, **116**, 1060-76.
226. E. W. Hernandez-Rodriguez, A. L. Montero-Alejo, R. Lopez, E. Sanchez-Garcia, L. A. Montero-Cabrera and J. M. de la Vega, *J. Comput. Chem.*, 2013, **34**, 2460-71.
227. A. T. Taguchi, P. J. O'Malley, C. A. Wraight and S. A. Dikanov, *Biochemistry*, 2013.
228. M. K. Sener, J. D. Olsen, C. N. Hunter and K. Schulten, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 15723-8.
229. M. Sener, J. Strumpfer, J. A. Timney, A. Freiberg, C. N. Hunter and K. Schulten, *Biophys. J.*, 2010, **99**, 67-75.
230. J. Deisenhofer and H. Michel, *Science*, 1989, **245**, 1463-73.
231. W. W. Parson and A. Warshel, in *The Purple Phototrophic Bacteria*, eds. C. N. Hunter, F. Daldal, M. C. Thurnauer and J. T. Beatty, Springer, Dordrecht, 2009, vol., pp. 355-77.
232. C. R. Lancaster and H. Michel, *Structure*, 1997, **5**, 1339-59.
233. M. E. Michel-Beyerle, M. Plato, J. Deisenhofer, H. Michel, M. Bixon and J. Jortner, *Biochim. Biophys. Acta.*, 1988, **932**, 52-70.
234. W. Nitschke and S. M. Dracheva, in *Anoxygenic Photosynthetic Bacteria*, eds. R. E. Blankenship, M. T. Madigan and C. E. Bauer, Kluwer, Dordrecht, 1995, vol., pp. 775-805.
235. S. M. Dracheva, L. A. Drachev, A. A. Konstantinov, A. Y. Semenov, V. P. Skulachev, A. M. Arutjunjan, V. A. Shuvalov and S. M. Zaberezhnaya, *Eur. J. Biochem.*, 1988, **171**, 253-64.
236. G. Fritzsch, S. Buchanan and H. Michel, *Biochim. Biophys. Acta.*, 1989, **977**, 157-62.
237. J. Deisenhofer, O. Epp, I. Sinning and H. Michel, *J. Mol. Biol.*, 1995, **246**, 429-57.
238. Govindjee, L. O. Bjorn and K. Nickelsen, in *Photosynthesis Research for Food, Fuel and the Future*, eds. T. Kuang, C. Lu and L. Zhang, Springer, Berlin, 2012, vol., pp. 827-33.
239. J. H. Golbeck, *Annual Review of Plant Physiology and Plant Molecular Biology*, 1992, **43**, 293-324.
240. J. Barber, *Quarterly Reviews of Biophysics*, 2003, **36**, 71-89.
241. F. H. Koua, Y. Umena, K. Kawakami and J. R. Shen, *Proc. Natl. Acad. Sci. U.S.A.*, 2013, **110**, 3889-94.
242. I. Grotjohann and P. Fromme, *Photosynth. Res.*, 2005, **85**, 51-72.
243. A. R. Holzwarth, M. G. Muller, M. Reus, M. Nowaczyk, J. Sander and M. Rogner, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 6895-900.
244. A. R. Holzwarth, M. G. Muller, J. Niklas and W. Lubitz, *Biophys. J.*, 2006, **90**, 552-65.
245. H. Zuber and R. J. Cogdell, in *Anoxygenic Photosynthetic Bacteria*, eds. R. E. Blankenship, M. T. Madigan and C. E. Bauer, Kluwer, Dordrecht, 1995, vol., pp. 315-48.
246. G. McDermott, S. M. Prince, A. A. Freer, A. M. Hawthornthwaite-Lawless, M. Z. Papiz, R. J. Cogdell and N. W. Isaacs, *Nature*, 1995, **374**, 517-21.
247. M. Z. Papiz, S. M. Prince, T. Howard, R. J. Cogdell and N. W. Isaacs, *J. Mol. Biol.*, 2003, **326**, 1523-38.
248. J. Koepke, X. Hu, C. Muenke, K. Schulten and H. Michel, *Structure*, 1996, **4**, 581-97.
249. A. W. Roszak, T. D. Howard, J. Southall, A. T. Gardiner, C. J. Law, N. W. Isaacs and R. J. Cogdell, *Science*, 2003, **302**, 1969-72.
250. M. F. Richter, J. Baier, J. Southall, R. J. Cogdell, S. Oellerich and J. Kohler, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 20280-4.
251. X. Hu, T. Ritz, A. Damjanović and K. Schulten, *J. Phys. Chem. B*, 1997, **101**, 3854-71.
252. X. Hu, A. Damjanović, T. Ritz and K. Schulten, *Proc. Natl. Acad. Sci. U.S.A.*, 1998, **95**, 5935-41.
253. A. Freer, S. Prince, K. Sauer, M. Papiz, A. H. Lawless, G. McDermott, R. Cogdell and N. W. Isaacs, *Structure*, 1996, **4**, 449-62.
254. K. McLuskey, S. M. Prince, R. J. Cogdell and N. W. Isaacs, *Biochemistry*, 2001, **40**, 8783-9.
255. J. Chmeliov, E. Songaila, O. Rancova, A. Gall, B. Robert, D. Abramavicius and L. Valkunas, *J. Phys. Chem. B*, 2013.
256. S. M. Prince, M. Z. Papiz, A. A. Freer, G. McDermott, A. M. Hawthornthwaite-Lawless, R. J. Cogdell and N. W. Isaacs, *J. Mol. Biol.*, 1997, **268**, 412-23.
257. R. G. Alden, E. Johnson, V. Nagarajan, W. W. Parson, C. J. Law and R. G. Cogdell, *J. Phys. Chem. B*, 1997, **101**, 4667-80.
258. J. Linnanto, J. E. I. Korppi-Tommola and V. M. Helenius, *J. Phys. Chem. B*, 1999, **103**, 8739-50.
259. Z. He, V. Sundström and T. Pullerits, *J. Phys. Chem. B*, 2002, **106**, 11606-12.
260. J. Neugebauer, *J. Phys. Chem. B*, 2008, **112**, 2207-17.

261. J. Neugebauer, *Physics Reports*, 2010, **489**, 1-87.
262. G. D. Scholes, I. R. Gould, R. J. Cogdell and G. R. Fleming, *J. Phys. Chem. B*, 1999, **103**, 2543-53.
263. J. M. Olson, *Photosynth. Res.*, 2005, **20**, 421-7.
264. M. Schmidt am Busch, F. Müh, M. El-Amine Madjet and T. Renger, *J. Phys. Chem. Lett.*, 2011, **2**, 93-8.
265. M. T. Milder, B. Bruggemann, R. van Grondelle and J. L. Herek, *Photosynth. Res.*, 2010, **104**, 257-74.
266. D. E. Tronrud, J. Wen, L. Gay and R. E. Blankenship, *Photosynth. Res.*, 2009, **100**, 79-87.
267. D. E. Tronrud and J. P. Allen, *Photosynth. Res.*, 2012, **112**, 71-4.
268. D. E. Tronrud and B. W. Matthews, in *The Photosynthetic Reaction Center*, eds. J. Deisenhofer and J. R. Norris, Academic Press, New York, 1993, vol. 1, pp. 13-21.
269. F. Muh, A. Madjet Mel, J. Adolphs, A. Abdurahman, B. Rabenstein, H. Ishikita, E. W. Knapp and T. Renger, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 16862-7.
270. J. Adolphs, F. Muh, A. Madjet Mel and T. Renger, *Photosynth. Res.*, 2008, **95**, 197-209.
271. R. E. Blankenship, J. M. Olson and M. Miller, in *Anoxygenic Photosynthetic Bacteria*, eds. R. E. Blankenship, M. T. Madigan and C. E. Bauer, Kluwer, Dordrecht, 1995, vol., pp. 399-435.
272. M. A. Wilson, D. A. Hodgson and B. J. Keely, *Org. Geochem.*, 2004, **35**, 1299-307.
273. K. M. Smith and F. W. Bobe, *J. Chem. Soc., Chem. Commun.*, 1987, 276.
274. S. C. Otte, E. J. van de Meent, P. A. van Veelen, A. S. Pundsnes and J. Amesz, *Photosynth. Res.*, 1993, **35**, 159-69.
275. M. O. Senge and K. M. Smith, *Photochem. Photobiol.*, 1994, **60**, 139-42.
276. D. E. Tronrud, M. F. Schmid and B. W. Matthews, *J. Mol. Biol.*, 1986, **188**, 443-54.
277. Y. F. Li, W. Zhou, R. E. Blankenship and J. P. Allen, *J. Mol. Biol.*, 1997, **271**, 456-71.
278. M. Bixon, J. Fajer, G. Freed, D. Gamliel, A. J. Hoff, H. Levanon, K. Möbius, R. Nechshtai, J. R. Norris, A. Scherz, J. L. Sessler and D. Stehlik, *Isr. J. Chem.*, 1992, **32**, 363-518.
279. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235-42.
280. Z. Feng, L. Chen, H. Maddula, O. Akcan, R. Oughtred, H. M. Berman and J. Westbrook, *Bioinformatics*, 2004, **20**, 2153-5.
281. L.-S. Sun and J. A. Shelnutt, Normal-Coordinate Structural Decomposition Engine, http://jasheln.unm.edu/jasheln/content/nsd/NSDengine/nsd_index.htm.
282. R. D. C. Team, Vienna, Austria, 2010.
283. J. P. Ridge, P. K. Fyfe, K. E. McAuley, M. E. van Brederode, B. Robert, R. van Grondelle, N. W. Isaacs, R. J. Cogdell and M. R. Jones, *Biochem. J*, 2000, **351 Pt 3**, 567-78.
284. R. Dennington, T. Keith and J. Millam, Semichem Inc., Shawnee Mission KS, 2009.
285. B. S. Everitt, S. Landau, M. Leese and D. Stahl, *Cluster Analysis*, Wiley, West Sussex, England, 2011.
286. A. C. Rencher and W. F. Christensen, *Methods of Multivariate Analysis*, Wiley, Hoboken, NJ, 2012.
287. A. D. Becke, *The Journal of chemical physics*, 1993, **98**, 5648.
288. J. S. Binkley, J. A. Pople and W. J. Hehre, *J. Am. Chem. Soc.*, 1980, **102**, 939-47.
289. K. D. Dobbs and W. J. Hehre, *J. Comput. Chem.*, 1987, **8**, 861-79.
290. R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, *The Journal of chemical physics*, 1980, **72**, 650.
291. A. J. H. Wachters, *The Journal of chemical physics*, 1970, **52**, 1033.
292. P. J. Hay, *The Journal of chemical physics*, 1977, **66**, 4377.
293. K. Raghavachari and G. W. Trucks, *The Journal of chemical physics*, 1989, **91**, 1062.
294. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.
295. J. Linnanto and J. Korppi-Tommola, *PCCP*, 2000, **2**, 4962-70.
296. J. Linnanto and J. Korppi-Tommola, *J. Phys. Chem. A*, 2001, **105**, 3855-66.
297. J. Linnanto and J. Korppi-Tommola, *J. Comput. Chem.*, 2004, **25**, 123-38.
298. J. Linnanto and J. Korppi-Tommola, *J. Phys. Chem. A*, 2004, **108**, 5872-82.
299. J. Linnanto, A. Freiberg and J. Korppi-Tommola, *J. Phys. Chem. B*, 2011, **115**, 5536-44.
300. M. Guse, N. S. Ostlund and G. D. Blyholder, *Chemical Physics Letters*, 1979, **61**, 526-31.

301. T. H. Dunning, B. H. Botch and J. F. Harrison, *The Journal of chemical physics*, 1980, **72**, 3419.
302. B. H. Botch, T. H. Dunning and J. F. Harrison, *The Journal of chemical physics*, 1981, **75**, 3466.
303. J. Almlof, T. H. Fischer, P. G. Gassman, A. Ghosh and M. Haeser, *J. Phys. Chem.*, 1993, **97**, 10964-70.
304. T. Vangberg and A. Ghosh, *J. Am. Chem. Soc.*, 1999, **121**, 12154-60.
305. A. Ghosh and E. Steene, *Journal of Biological Inorganic Chemistry*, 2001, **6**, 739-52.
306. A. B. Alemayehu, E. Gonzalez, L. K. Hansen and A. Ghosh, *Inorg. Chem.*, 2009, **48**, 7794-9.
307. K. E. Thomas, A. B. Alemayehu, J. Conradie, C. M. Beavers and A. Ghosh, *Acc. Chem. Res.*, 2012, **45**, 1203-14.
308. A. Ghosh, *J. Porphyrins Phthalocyanines*, 2000, **4**, 380-1.
309. W. K. Cheney, *Numerical Mathematics and Computing*, Thomson-Brooks/Cole, Belmont, CA, 2008.
310. G. Feher, J. P. Allen, M. Y. Okamura and D. C. Rees, *Nature*, 1989, **339**, 111-6.
311. H. M. Berman, *Nucleic Acids Res.*, 2000, **28**, 235-42.
312. M. H. Stowell, T. M. McPhillips, D. C. Rees, S. M. Soltis, E. Abresch and G. Feher, *Science*, 1997, **276**, 812-6.
313. H. L. Axelrod, E. C. Abresch, M. L. Paddock, M. Y. Okamura and G. Feher, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 1542-7.
314. P. K. Fyfe, J. P. Ridge, K. E. McAuley, R. J. Cogdell, N. W. Isaacs and M. R. Jones, *Biochemistry*, 2000, **39**, 5953-60.
315. A. Kuglstatter, U. Ermler, H. Michel, L. Baciou and G. Fritzsch, *Biochemistry*, 2001, **40**, 4253-60.
316. A. Camara-Artigas, C. L. Magee, J. C. Williams and J. P. Allen, *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, 2001, **57**, 1281-6.
317. P. R. Pokkuluri, P. D. Laible, Y. L. Deng, T. N. Wong, D. K. Hanson and M. Schiffer, *Biochemistry*, 2002, **41**, 5998-6007.
318. A. Camara-Artigas, C. Magee, A. Goetsch and J. P. Allen, *Photosynth. Res.*, 2002, **74**, 87-93.
319. A. Camara-Artigas, D. Brune and J. P. Allen, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 11055-60.
320. G. Katona, U. Andreasson, E. M. Landau, L. E. Andreasson and R. Neutze, *J. Mol. Biol.*, 2003, **331**, 681-92.
321. U. Ermler, G. Fritzsch, S. K. Buchanan and H. Michel, *Structure*, 1994, **2**, 925-36.
322. A. J. Chirino, E. J. Lous, M. Huber, J. P. Allen, C. C. Schenck, M. L. Paddock, G. Feher and D. C. Rees, *Biochemistry*, 1994, **33**, 4584-93.
323. K. E. McAuley, P. K. Fyfe, J. P. Ridge, N. W. Isaacs, R. J. Cogdell and M. R. Jones, *Proc. Natl. Acad. Sci. U.S.A.*, 1999, **96**, 14706-11.
324. A. W. Roszak, K. McKendrick, A. T. Gardiner, I. A. Mitchell, N. W. Isaacs, R. J. Cogdell, H. Hashimoto and H. A. Frank, *Structure*, 2004, **12**, 765-73.
325. Q. Xu, H. L. Axelrod, E. C. Abresch, M. L. Paddock, M. Y. Okamura and G. Feher, *Structure*, 2004, **12**, 703-15.
326. P. K. Fyfe, N. W. Isaacs, R. J. Cogdell and M. R. Jones, *Biochim. Biophys. Acta.*, 2004, **1608**, 11-22.
327. M. L. Paddock, C. Chang, Q. Xu, E. C. Abresch, H. L. Axelrod, G. Feher and M. Y. Okamura, *Biochemistry*, 2005, **44**, 6920-8.
328. B. Arnoux, J. F. Gaucher, A. Ducruix and F. Reiss-Husson, *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, 1995, **51**, 368-79.
329. M. Thielges, G. Uyeda, A. Camara-Artigas, L. Kalman, J. C. Williams and J. P. Allen, *Biochemistry*, 2005, **44**, 7389-94.
330. G. Katona, A. Snijder, P. Gourdon, U. Andreasson, O. Hansson, L. E. Andreasson and R. Neutze, *Nature structural & molecular biology*, 2005, **12**, 630-1.
331. J. A. Potter, P. K. Fyfe, D. Frolov, M. C. Wakeham, R. van Grondelle, B. Robert and M. R. Jones, *The Journal of biological chemistry*, 2005, **280**, 27155-64.
332. S. Hermes, J. M. Stachnik, D. Onidas, A. Remy, E. Hofmann and K. Gerwert, *Biochemistry*, 2006, **45**, 13741-9.
333. P. Wadsten, A. B. Wohri, A. Snijder, G. Katona, A. T. Gardiner, R. J. Cogdell, R. Neutze and S. Engstrom, *J. Mol. Biol.*, 2006, **364**, 44-53.
334. A. W. Roszak, A. T. Gardiner, N. W. Isaacs and R. J. Cogdell, *Biochemistry*, 2007, **46**, 2909-16.
335. J. Koepke, E. M. Krammer, A. R. Klingen, P. Sebban, G. M. Ullmann and G. Fritzsch, *J. Mol. Biol.*, 2007, **371**, 396-409.
336. P. K. Fyfe, J. A. Potter, J. Cheng, C. M. Williams, A. J. Watson and M. R. Jones, *Biochemistry*, 2007, **46**, 10461-72.
337. C. H. Chang, O. el-Kabbani, D. Tiede, J. Norris and M. Schiffer, *Biochemistry*, 1991, **30**, 5352-60.
338. D. Frolov, M. Marsh, L. I. Crouch, P. K. Fyfe, B. Robert, R. van Grondelle, A. Hadfield and M. R. Jones, *Biochemistry*, 2010, **49**, 1882-92.
339. P. R. Pokkuluri, P. D. Laible, D. K. Hanson and M. Schiffer, 2009.

340. P. R. Pokkuluri, P. D. Laible, S. L. Ginell, D. K. Hanson and M. Schiffer, 2009.
341. T. O. Yeates, H. Komiya, A. Chirino, D. C. Rees, J. P. Allen and G. Feher, *Proc. Natl. Acad. Sci. U.S.A.*, 1988, **85**, 7993-7.
342. C. R. Lancaster, M. V. Bibikova, P. Sabatino, D. Oesterhelt and H. Michel, *The Journal of biological chemistry*, 2000, **275**, 39364-8.
343. R. H. Baxter, N. Ponomarenko, V. Srajer, R. Pahl, K. Moffat and J. R. Norris, *Proc. Natl. Acad. Sci. U.S.A.*, 2004, **101**, 5982-7.
344. R. H. Baxter, B. L. Seagle, N. Ponomarenko and J. R. Norris, *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, 2005, **61**, 605-12.
345. L. Li, D. Mustafi, Q. Fu, V. Tereshko, D. L. Chen, J. D. Tice and R. F. Ismagilov, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 19243-8.
346. C. R. Lancaster, C. Hunte, J. Kelley, 3rd, B. L. Trumpower and R. Ditchfield, *J. Mol. Biol.*, 2007, **368**, 197-208.
347. A. B. Wohri, W. Y. Wahlgren, E. Malmerberg, L. C. Johansson, R. Neutze and G. Katona, *Biochemistry*, 2009, **48**, 9831-8.
348. L. Li, S. Nachtergaele, A. M. Seddon, V. Tereshko, N. Ponomarenko and R. F. Ismagilov, *J. Am. Chem. Soc.*, 2008, **130**, 14324-8.
349. N. S. Ponomarenko, L. Li, A. R. Marino, V. Tereshko, A. Ostafin, J. A. Popova, E. J. Bylina, R. F. Ismagilov and J. R. Norris, Jr., *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2009, **1788**, 1822-31.
350. C. R. Lancaster and H. Michel, *J. Mol. Biol.*, 1999, **286**, 883-98.
351. , Addinsoft, New York, USA, 2011.
352. R. D. C. Team, R Foundation for Statistical Computing, Vienna, Austria, 2008.
353. M. O. Senge, W. W. Kalisch and K. Ruhlandt-Senge, *Chem. Commun.*, 1996, 2149.
354. R. Marcus, *Reviews of Modern Physics*, 1993, **65**, 599-610.
355. C. C. Moser, J. M. Keske, K. Warncke, R. S. Farid and P. L. Dutton, *Nature*, 1992, **355**, 796-802.
356. W. W. Parson, Z. T. Chu and A. Warshel, *Biochim. Biophys. Acta.*, 1990, **1017**, 251-72.
357. E. Daviso, S. Prakash, A. Alia, P. Gast, J. Neugebauer, G. Jeschke and J. Matysik, *Proc. Natl. Acad. Sci. U.S.A.*, 2009, **106**, 22281-6.
358. P. D. Laible, V. Chynwat, M. C. Thurnauer, M. Schiffer, D. K. Hanson and H. A. Frank, *Biophys. J.*, 1998, **74**, 2623-37.
359. V. Palaniappan, P. C. Martin, V. Chynwat, H. A. Frank and D. F. Bocian, *J. Am. Chem. Soc.*, 1993, **115**, 12035-49.
360. J. Oelze, in *Methods in Microbiology*, ed. G. Gottschank, Academic Press, London, 1985, vol. 18, pp. 257-84.
361. H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall, R. B. Doak, F. R. Maia, A. V. Martin, I. Schlichting, L. Lomb, N. Coppola, R. L. Shoeman, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmess, C. Caleman, S. Boutet, M. J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt, L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Homke, C. Reich, D. Pietschner, L. Struder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K. U. Kuhnel, M. Messerschmidt, J. D. Bozek, S. P. Hau-Riege, M. Frank, C. Y. Hampton, R. G. Sierra, D. Starodub, G. J. Williams, J. Hajdu, N. Timneanu, M. M. Seibert, J. Andreasson, A. Rocker, O. Jonsson, M. Svenda, S. Stern, K. Nass, R. Andritschke, C. D. Schroter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Wang, I. Grotjohann, J. M. Holton, T. R. Barends, R. Neutze, S. Marchesini, R. Fromme, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, I. Andersson, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson and J. C. Spence, *Nature*, 2011, **470**, 73-7.
362. A. T. Brunger, P. D. Adams, P. Fromme, R. Fromme, M. Levitt and G. F. Schroder, *Structure*, 2012, **20**, 957-66.
363. A. Amunts, O. Drory and N. Nelson, *Nature*, 2007, **447**, 58-63.
364. A. Amunts, H. Toporik, A. Borovikova and N. Nelson, *The Journal of biological chemistry*, 2010, **285**, 3478-86.
365. N. Kamiya and J. R. Shen, *Proc. Natl. Acad. Sci. U.S.A.*, 2003, **100**, 98-103.
366. K. N. Ferreira, T. M. Iverson, K. Maghlaoui, J. Barber and S. Iwata, *Science*, 2004, **303**, 1831-8.
367. J. Biesiadka, B. LollThese authors contributed equal, J. Kern, K.-D. Irrgang and A. Zouni, *PCCP*, 2004, **6**, 4733.
368. B. Loll, J. Kern, W. Saenger, A. Zouni and J. Biesiadka, *Nature*, 2005, **438**, 1040-4.
369. K. Kawakami, Y. Umena, N. Kamiya and J. R. Shen, *Proc. Natl. Acad. Sci. U.S.A.*, 2009, **106**, 8567-72.
370. A. Guskov, J. Kern, A. Gabdulkhakov, M. Broser, A. Zouni and W. Saenger, *Nature structural & molecular biology*, 2009, **16**, 334-42.
371. Z. Chen, F. Medina, M. Y. Liu, C. Thomas, S. R. Sprang and P. C. Sternweis, *The Journal of biological chemistry*, 2010, **285**, 21070-81.

372. M. Broser, C. Glockner, A. Gabdulkhakov, A. Guskov, J. Buchta, J. Kern, F. Muh, H. Dau, W. Saenger and A. Zouni, *The Journal of biological chemistry*, 2011, **286**, 15964-72.
373. J. Kern, R. Alonso-Mori, J. Hellmich, R. Tran, J. Hattne, H. Laksmono, C. Glockner, N. Echols, R. G. Sierra, J. Sellberg, B. Lassalle-Kaiser, R. J. Gildea, P. Glatzel, R. W. Grosse-Kunstleve, M. J. Latimer, T. A. McQueen, D. DiFiore, A. R. Fry, M. Messerschmidt, A. Miahnahri, D. W. Schafer, M. M. Seibert, D. Sokaras, T. C. Weng, P. H. Zwart, W. E. White, P. D. Adams, M. J. Bogan, S. Boutet, G. J. Williams, J. Messinger, N. K. Sauter, A. Zouni, U. Bergmann, J. Yano and V. K. Yachandra, *Proc. Natl. Acad. Sci. U.S.A.*, 2012, **109**, 9721-6.
374. J. Kern, R. Alonso-Mori, R. Tran, J. Hattne, R. J. Gildea, N. Echols, C. Glockner, J. Hellmich, H. Laksmono, R. G. Sierra, B. Lassalle-Kaiser, S. Koroidov, A. Lampe, G. Han, S. Gul, D. Difiore, D. Milathianaki, A. R. Fry, A. Miahnahri, D. W. Schafer, M. Messerschmidt, M. M. Seibert, J. E. Koglin, D. Sokaras, T. C. Weng, J. Sellberg, M. J. Latimer, R. W. Grosse-Kunstleve, P. H. Zwart, W. E. White, P. Glatzel, P. D. Adams, M. J. Bogan, G. J. Williams, S. Boutet, J. Messinger, A. Zouni, N. K. Sauter, V. K. Yachandra, U. Bergmann and J. Yano, *Science*, 2013, **340**, 491-5.
375. M. Guergova-Kuras, B. Boudreaux, A. Joliot, P. Joliot and K. Redding, *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **98**, 4437-42.
376. R. S. Mulliken, *The Journal of chemical physics*, 1955, **23**, 1833.
377. J. A. Montgomery, M. J. Frisch, J. W. Ochterski and G. A. Petersson, *The Journal of chemical physics*, 1999, **110**, 2822.
378. J. A. Montgomery, M. J. Frisch, J. W. Ochterski and G. A. Petersson, *The Journal of chemical physics*, 2000, **112**, 6532.
379. A. E. Reed, R. B. Weinstock and F. Weinhold, *The Journal of chemical physics*, 1985, **83**, 735.
380. C. Rovira, K. Kunc, J. Hutter, P. Ballone and M. Parrinello, *J. Phys. Chem. A*, 1997, **101**, 8914-25.
381. W. Liu and L. Li, *Theoretica Chimica Acta*, 1997, **95**, 81-95.
382. J. J. Philips, M. A. Hudspeth, P. M. Browne and J. E. Peralta, *Chemical Physics Letters*, 2010, **495**, 146-50.
383. J. Eccles and B. Honig, *Proc. Natl. Acad. Sci. U.S.A.*, 1983, **80**, 4959-44962.
384. R. E. Fenna, L. F. Eyck and B. W. Matthews, *Biochem. Biophys. Res. Commun.*, 1977, **75**, 751-6.
385. C. König and J. Neugebauer, *Journal of Chemical Theory and Computation*, 2013, **9**, 1808-20.
386. J. Gao, W. J. Shi, J. Ye, X. Wang, H. Hirao and Y. Zhao, *J. Phys. Chem. B*, 2013, **117**, 3488-95.
387. A. Camara-Artigas, R. E. Blankenship and J. P. Allen, *Photosynth. Res.*, 2003, **75**, 49-55.
388. A. Ben-Shem, F. Frolow and N. Nelson, *FEBS Letters*, 2004, **564**, 274-80.
389. C. R. Larson, C. O. Seng, L. Lauman, H. J. Matthies, J. Wen, R. E. Blankenship and J. P. Allen, *Photosynth. Res.*, 2011, **107**, 139-50.
390. X. Li, N. Sudarsanam and D. D. Frey, *Complexity*, 2006, **11**, 32-45.
391. M. C. Zerner, in *Reviews in Computational Chemistry*, eds. K. B. Lipkowitz and D. B. Boyd, Wiley, New York, 1991, vol. 2, pp. 313-65.
392. J. J. Stewart, *J. Mol. Model.*, 2007, **13**, 1173-213.
393. S. T. Daurat-Larroque, K. Brew and R. E. Fenna, *The Journal of biological chemistry*, 1986, **261**, 3607-15.
394. S. Dracheva, J. C. Williams and R. E. Blankenship, in *Research in Photosynthesis*, ed. N. Murata, Kluwer, Dordrecht, 1992, vol. 1, pp. 53-6.
395. D. E. Tronrud, L. F. Ten Eyck and B. W. Matthews, *Acta Crystallographica Section A: Foundations and Advances*, 1987, **43**, 489-501.
396. H. Yamasaki, Y. Takano and H. Nakamura, *J. Phys. Chem. B*, 2008, **112**, 13923-33.
397. K. B. Sai Sankar Gupta, A. Alia, H. J. de Groot and J. Matysik, *J. Am. Chem. Soc.*, 2013, **135**, 10382-7.
398. S. I. E. Vulto, M. A. de Baat, R. J. W. Louwe, H. P. Permentier, T. Neef, M. Miller, H. van Amerongen and T. J. Aartsma, *J. Phys. Chem. B*, 1998, **102**, 9577-82.
399. A. B. Wohri, G. Katona, L. C. Johansson, E. Fritz, E. Malmerberg, M. Andersson, J. Vincent, M. Eklund, M. Cammarata, M. Wulff, J. Davidsson, G. Groenhof and R. Neutze, *Science*, 2010, **328**, 630-3.
400. A. W. Roszak, V. Moulisova, A. D. Reksodipuro, A. T. Gardiner, R. Fujii, H. Hashimoto, N. W. Isaacs and R. J. Cogdell, *Biochem. J*, 2012, **442**, 27-37.
401. L. C. Johansson, D. Arnlund, T. A. White, G. Katona, D. P. Deponte, U. Weierstall, R. B. Doak, R. L. Shoeman, L. Lomb, E. Malmerberg, J. Davidsson, K. Nass, M. Liang, J. Andreasson, A. Aquila, S. Bajt, M. Barthelmess, A. Barty, M. J. Bogan, C. Bostedt, J. D. Bozek, C. Caleman, R. Coffee, N. Coppola, T. Ekeberg, S. W. Epp, B. Erk, H. Fleckenstein, L. Foucar, H. Graafsma, L. Gumprecht, J. Hajdu, C. Y. Hampton, R. Hartmann, A. Hartmann, G. Hauser, H. Hirsemann, P. Holl, M. S. Hunter, S. Kassemeyer, N. Kimmel, R. A. Kirian, F. R. Maia, S. Marchesini, A. V. Martin, C. Reich, D. Rolles, B. Rudek, A. Rudenko, I. Schlichting, J. Schulz, M. M. Seibert, R. G. Sierra, H. Soltau, D. Starodub, F. Stellato, S. Stern, L. Struder, N. Timneanu, J. Ullrich, W. Y. Wahlgren, X.

Wang, G. Weidenspointner, C. Wunderer, P. Fromme, H. N. Chapman, J. C. Spence and R. Neutze, *Nat. Methods*, 2012, **9**, 263-5.
402. K. Gibasiewicz, M. Pajzderska, J. A. Potter, P. K. Fyfe, A. Dobek, K. Brettel and M. R. Jones, *J. Phys. Chem. B*, 2011, **115**, 13037-50.
403. L. G. Vasilieva, T. Y. Fufina, A. G. Gabdulkhakov, M. M. Leonova, R. A. Khatypov and V. A. Shuvalov, *Biochim. Biophys. Acta.*, 2012, **1817**, 1407-17.
404. R. G. Saer, A. Hardjasa, F. I. Rosell, A. G. Mauk, M. E. Murphy and J. T. Beatty, *Biochemistry*, 2013, **52**, 2206-17.

Appendix A

NSD Descriptive statistics

Table A-1: Descriptive statistics of minimum basis out-of-plane parameters and normal-coordinate displacements for each ETC cofactor in all RBC2a experiments. n = 8.

	B_A	B_B	D_L	D_M	B_A	B_B	D_L	D_M
		D_{oop}				D_{ip}		
MEAN	0.0948	0.1851	0.1766	0.1122	0.3096	0.3686	0.4009	0.3937
SD	0.0156	0.0869	0.0372	0.0481	0.0511	0.0453	0.0478	0.0391
SE	0.0055	0.0307	0.0132	0.0170	0.0181	0.0160	0.0169	0.0138
		δ_{oop}				δ_{ip}		
MEAN	0.0061	0.0110	0.0045	0.0032	0.0588	0.0683	0.0621	0.0643
SD	0.0011	0.0111	0.0010	0.0006	0.0018	0.0157	0.0020	0.0027
SE	0.0004	0.0039	0.0004	0.0002	0.0006	0.0056	0.0007	0.0009
		B_{2u}				B_{2g}		
MEAN	-0.0102	-0.0737	0.0895	0.0707	-0.1075	-0.1260	-0.0501	-0.0392
SD	0.0376	0.1481	0.1064	0.0721	0.0367	0.0564	0.0718	0.0288
SE	0.0133	0.0524	0.0376	0.0255	0.0130	0.0199	0.0254	0.0102
		B_{1u}				B_{1g}		
MEAN	0.0393	0.0285	0.0501	-0.0613	-0.0127	-0.0419	-0.0933	-0.1664
SD	0.0181	0.0655	0.0498	0.0267	0.0350	0.0410	0.0888	0.0291
SE	0.0064	0.0232	0.0176	0.0094	0.0124	0.0145	0.0314	0.0103
		A_{2u}				$E_u(x)$		
MEAN	-0.0640	-0.0679	-0.0851	-0.0146	-0.1636	-0.1513	-0.1373	-0.1664
SD	0.0124	0.0447	0.0256	0.0139	0.0253	0.0789	0.0678	0.0232
SE	0.0044	0.0158	0.0091	0.0049	0.0089	0.0279	0.0240	0.0082
		$E_g(x)$				$E_u(y)$		
MEAN	-0.0343	-0.0088	-0.0297	-0.0126	0.0039	-0.0189	-0.0005	0.0561
SD	0.0140	0.0294	0.0206	0.0083	0.0335	0.0533	0.0729	0.0221
SE	0.0049	0.0104	0.0073	0.0029	0.0118	0.0189	0.0258	0.0078
		$E_g(y)$				A_{1g}		
MEAN	0.0187	0.0000	-0.0248	0.0020	0.2261	0.2850	0.3273	0.2989
SD	0.0109	0.0406	0.0078	0.0048	0.0641	0.0553	0.0637	0.0608
SE	0.0039	0.0144	0.0027	0.0017	0.0227	0.0196	0.0225	0.0215
		A_{1u}				A_{2g}		
MEAN	0.0003	0.0303	-0.0069	-0.0058	-0.0337	-0.0242	-0.0432	-0.0370
SD	0.0031	0.0492	0.0031	0.0052	0.0068	0.0112	0.0059	0.0027
SE	0.0011	0.0174	0.0011	0.0018	0.0024	0.0040	0.0021	0.0010

Table A-2: Descriptive statistics of minimum basis out-of-plane parameters and normal-coordinate displacements for each ETC cofactor in all RBC2b experiments. n = 10.

	B_A	B_B	D_L	D_M	B_A	B_B	D_L	D_M
		D_{oop}				D_{ip}		
MEAN	0.2903	0.3159	0.3275	0.2980	0.3143	0.3003	0.3888	0.3947
SD	0.0629	0.0604	0.0586	0.0712	0.0449	0.0432	0.0367	0.0239
SE	0.0199	0.0191	0.0185	0.0225	0.0142	0.0137	0.0116	0.0076
		δ_{oop}				δ_{ip}		
MEAN	0.0238	0.0228	0.0188	0.0142	0.0687	0.0684	0.0661	0.0701
SD	0.0070	0.0062	0.0053	0.0026	0.0046	0.0021	0.0016	0.0012
SE	0.0022	0.0020	0.0017	0.0008	0.0014	0.0007	0.0005	0.0004
		B_{2u}				B_{2g}		
MEAN	-0.0468	0.0006	0.1391	0.1049	0.0655	0.0753	0.0719	0.1090
SD	0.0878	0.0884	0.0841	0.0865	0.0163	0.0163	0.0232	0.0281
SE	0.0278	0.0280	0.0266	0.0273	0.0051	0.0052	0.0073	0.0089
		B_{1u}				B_{1g}		
MEAN	-0.0192	0.0937	0.0614	-0.1933	-0.1634	-0.1652	-0.2160	-0.2355
SD	0.0175	0.0501	0.0328	0.0594	0.0396	0.0372	0.0199	0.0286
SE	0.0055	0.0159	0.0104	0.0188	0.0125	0.0117	0.0063	0.0090

	B_A	B_B	D_L	D_M	B_A	B_B	D_L	D_M
		A_{2u}				$E_u(x)$		
MEAN	-0.2114	-0.2212	-0.2472	-0.1426	-0.0519	-0.0261	-0.0456	-0.0526
SD	0.0367	0.0443	0.0558	0.0099	0.0236	0.0382	0.0165	0.0240
SE	0.0116	0.0140	0.0176	0.0031	0.0075	0.0121	0.0052	0.0076
		$E_g(x)$				$E_u(y)$		
MEAN	-0.1473	-0.1332	-0.1148	-0.0847	-0.0552	-0.0641	-0.1046	-0.0493
SD	0.0270	0.0504	0.0339	0.0135	0.0385	0.0413	0.0370	0.0339
SE	0.0085	0.0159	0.0107	0.0043	0.0122	0.0131	0.0117	0.0107
		$E_g(y)$				A_{1g}		
MEAN	0.0713	0.1134	0.0121	0.0646	0.2266	0.1956	0.2717	0.2678
SD	0.0547	0.0296	0.0288	0.0218	0.0317	0.0455	0.0431	0.0333
SE	0.0173	0.0094	0.0091	0.0069	0.0100	0.0144	0.0136	0.0105
		A_{1u}				A_{2g}		
MEAN	-0.0378	-0.0055	-0.0260	-0.0511	-0.0897	-0.0980	-0.0989	-0.0896
SD	0.0264	0.0251	0.0165	0.0130	0.0110	0.0073	0.0018	0.0062
SE	0.0084	0.0079	0.0052	0.0041	0.0035	0.0023	0.0006	0.0020

Table A-3: Descriptive statistics of minimum basis out-of-plane parameters and normal-coordinate displacements for each ETC cofactor in all RBC3 experiments. n = 16.

	B_A	B_B	D_L	D_M	B_A	B_B	D_L	D_M
		D_{oop}				D_{ip}		
MEAN	0.4995	0.7487	0.6498	0.7200	0.2698	0.3092	0.2952	0.3108
SD	0.0900	0.0756	0.0578	0.0890	0.0322	0.0387	0.0393	0.0285
SE	0.0225	0.0189	0.0145	0.0223	0.0080	0.0097	0.0098	0.0071
		δ_{oop}				δ_{ip}		
MEAN	0.0946	0.0737	0.0979	0.0666	0.0506	0.0528	0.0539	0.0556
SD	0.0237	0.0233	0.0168	0.0249	0.0059	0.0053	0.0057	0.0050
SE	0.0059	0.0058	0.0042	0.0062	0.0015	0.0013	0.0014	0.0012
		B_{2u}				B_{2g}		
MEAN	0.0604	0.2486	0.2418	0.3382	0.0661	0.0866	0.0326	0.0180
SD	0.0760	0.1230	0.0874	0.0596	0.0403	0.0729	0.0669	0.0493
SE	0.0190	0.0307	0.0218	0.0149	0.0101	0.0182	0.0167	0.0123
		B_{1u}				B_{1g}		
MEAN	0.0387	0.5398	0.1404	-0.5483	-0.1934	-0.2099	-0.2247	-0.2478
SD	0.0579	0.0946	0.1019	0.1179	0.0698	0.0370	0.0386	0.0441
SE	0.0145	0.0237	0.0255	0.0295	0.0174	0.0093	0.0097	0.0110
		A_{2u}				$E_u(x)$		
MEAN	-0.1881	-0.0672	-0.4238	-0.1913	-0.0562	-0.0115	-0.0627	-0.0681
SD	0.0747	0.0983	0.0746	0.0675	0.0460	0.0424	0.0358	0.0288
SE	0.0187	0.0246	0.0186	0.0169	0.0115	0.0106	0.0089	0.0072
		$E_g(x)$				$E_u(y)$		
MEAN	-0.4214	-0.3613	-0.1966	-0.1219	-0.0101	-0.1067	-0.0800	-0.0217
SD	0.0761	0.0859	0.0596	0.0622	0.0298	0.0528	0.0296	0.0381
SE	0.0190	0.0215	0.0149	0.0156	0.0074	0.0132	0.0074	0.0095
		$E_g(y)$				A_{1g}		
MEAN	-0.0684	0.1348	-0.2822	0.0270	0.0109	-0.0621	0.0159	0.0411
SD	0.1190	0.0696	0.0614	0.0655	0.0727	0.0618	0.0736	0.0870
SE	0.0297	0.0174	0.0154	0.0164	0.0182	0.0154	0.0184	0.0217
		A_{1u}				A_{2g}		
MEAN	-0.0023	0.1053	0.1342	-0.1601	-0.1205	-0.1252	-0.1163	-0.1230
SD	0.0435	0.0758	0.0495	0.0708	0.0161	0.0175	0.0160	0.0211
SE	0.0109	0.0190	0.0124	0.0177	0.0040	0.0044	0.0040	0.0053

Table A-4: Descriptive statistics of minimum basis out-of-plane parameters and normal-coordinate displacements for each ETC cofactor in all RBC4 experiments. n = 32.

	B_A	B_B	D_L	D_M	B_A	B_B	D_L	D_M
		D_{oop}				D_{ip}		
MEAN	0.3611	0.5848	0.5168	0.5694	0.3052	0.3894	0.4031	0.3704
SD	0.0688	0.1244	0.0677	0.0824	0.0740	0.0692	0.0609	0.0735
SE	0.0122	0.0220	0.0120	0.0146	0.0131	0.0122	0.0108	0.0130
		δ_{oop}				δ_{ip}		
MEAN	0.0340	0.0255	0.0349	0.0205	0.0498	0.0519	0.0499	0.0546
SD	0.0153	0.0144	0.0155	0.0116	0.0050	0.0053	0.0046	0.0044
SE	0.0027	0.0025	0.0027	0.0020	0.0009	0.0009	0.0008	0.0008
		B_{2u}				B_{2g}		
MEAN	0.0691	0.1940	0.2457	0.2630	0.0358	0.0341	-0.0261	0.0021
SD	0.0961	0.0702	0.0757	0.1078	0.0973	0.0693	0.0799	0.0609
SE	0.0170	0.0124	0.0134	0.0191	0.0172	0.0122	0.0141	0.0108
		B_{1u}				B_{1g}		
MEAN	0.0551	0.4220	0.1835	-0.4504	-0.1219	-0.2318	-0.2237	-0.2019
SD	0.0451	0.1220	0.0678	0.0808	0.0956	0.1012	0.0992	0.0874
SE	0.0080	0.0216	0.0120	0.0143	0.0169	0.0179	0.0175	0.0154
		A_{2u}				$E_u(x)$		
MEAN	-0.1472	-0.0656	-0.3326	-0.1012	-0.0713	-0.0463	-0.0785	-0.0559
SD	0.1008	0.0953	0.0673	0.0858	0.0520	0.0436	0.0318	0.0450
SE	0.0178	0.0169	0.0119	0.0152	0.0092	0.0077	0.0056	0.0079
		$E_g(x)$				$E_u(y)$		
MEAN	-0.2699	-0.2761	-0.0951	-0.1004	-0.0323	-0.1035	-0.0684	-0.0335
SD	0.0739	0.0682	0.0672	0.0620	0.0370	0.0377	0.0253	0.0483
SE	0.0131	0.0121	0.0119	0.0110	0.0065	0.0067	0.0045	0.0085
		$E_g(y)$				A_{1g}		
MEAN	0.0001	0.1349	-0.1625	-0.0228	0.2122	0.2472	0.2754	0.2702
SD	0.0729	0.0842	0.1054	0.0522	0.0558	0.0642	0.0686	0.0632
SE	0.0129	0.0149	0.0186	0.0092	0.0099	0.0113	0.0121	0.0112
		A_{1u}				A_{2g}		
MEAN	-0.0022	0.0451	0.0253	-0.0689	-0.0766	-0.0729	-0.0796	-0.0685
SD	0.0383	0.0342	0.0348	0.0463	0.0164	0.0185	0.0224	0.0262
SE	0.0068	0.0060	0.0061	0.0082	0.0029	0.0033	0.0040	0.0046

Table A-5: Minimum basis out-of-plane parameters and normal-coordinate displacements for each ETC cofactor in 1PRC.

B_A	B_B	D_L	D_M	B_A	B_B	D_L	D_M
	D_{oop}				D_{ip}		
0.3019	0.1875	0.4142	0.663	0.2998	0.3533	0.3796	0.2809
	δ_{oop}				δ_{ip}		
0.0394	0.0458	0.0401	0.0518	0.0472	0.0492	0.0515	0.0521
	B_{2u}				B_{2g}		
0.1285	0.0619	-0.154	0.1016	0.0674	0.0859	0.0714	0.0714
	B_{1u}				B_{1g}		
0.1168	0.0503	-0.3335	-0.6127	-0.1271	-0.1604	-0.2989	-0.0871
	A_{2u}				$E_u(x)$		
-0.0856	0.0065	-0.0628	-0.0368	-0.0911	-0.1037	-0.1082	-0.0834
	$E_g(x)$				$E_u(y)$		
-0.2169	-0.1562	-0.0963	0.0465	-0.0692	-0.0454	-0.0525	-0.2157
	$E_g(y)$				A_{1g}		
0.0698	0.06	0.0452	0.0876	0.2218	0.2691	0.1565	0.0381
	A_{1u}				A_{2g}		
0.0412	-0.0272	-0.146	-0.2064	-0.083	-0.0804	-0.1035	-0.1063

Table A-6: Extended basis parameters and normal-coordinate displacements for each ETC cofactor in 1PRC.

	B_A	B_B	D_L	D_M	B_A	B_B	D_L	D_M
		D_{oop}				D_{ip}		
N/A	0.3201	0.2294	0.4737	0.7232	0.3503	0.4072	0.4272	0.3634
		δ_{oop}				δ_{ip}		
N/A	0.0401	0.0443	0.0176	0.0145	0.0360	0.0332	0.0345	0.0314
		B_{2u}				B_{2g}		
1st	0.1280	0.0627	-0.1544	0.1016	0.0673	0.0853	0.0711	0.0703
2nd	-0.0093	0.0711	-0.0282	-0.0004	0.0271	0.1138	-0.0451	-0.0778
		B_{1u}				B_{1g}		
1st	0.1168	0.0503	-0.3335	-0.6127	-0.1266	-0.1601	-0.2967	-0.0844
2nd	-0.0607	0.0215	0.1784	0.2313	-0.0411	-0.0514	-0.0484	-0.0237
		A_{2u}				$E_u(x)$		
1st	-0.0857	0.0078	-0.0636	-0.0341	-0.0910	-0.1040	-0.1094	-0.0847
2nd	-0.0011	0.0228	-0.0085	0.0537	-0.0701	-0.1047	-0.0125	-0.1323
		$E_g(x)$				$E_u(y)$		
1st	-0.2160	-0.1544	-0.0956	0.0480	-0.0681	-0.0442	-0.0499	-0.2148
2nd	0.0673	0.0967	0.0244	0.0206	0.1532	0.1146	0.1754	0.1494
		$E_g(y)$				A_{1g}		
1st	0.0700	0.0599	0.0454	0.0855	0.2228	0.2681	0.1576	0.0379
2nd	0.0264	0.0150	-0.0219	-0.0252	0.0294	0.0271	0.0515	0.0771
		A_{1u}				A_{2g}		
1st	0.0412	-0.0272	-0.1460	-0.2064	-0.0811	-0.0789	-0.1017	-0.1052
2nd	-0.0431	0.0254	0.1378	0.1618	0.0363	0.0192	0.0107	0.0050

Table A-7: Descriptive statistics of minimum basis out-of-plane parameters and normal-coordinate displacements for each ETC cofactor in all BBC1 experiments; n = 13.

	B_A	B_B	D_L	D_M	B_A	B_B	D_L	D_M
		D_{oop}				D_{ip}		
MEAN	0.4466	0.4269	0.7582	1.0145	0.4245	0.3956	0.3739	0.3259
SD	0.0619	0.0590	0.0794	0.0765	0.0337	0.0664	0.0399	0.0465
SE	0.0109	0.0104	0.0140	0.0135	0.0060	0.0117	0.0071	0.0082
		δ_{oop}				δ_{ip}		
MEAN	0.0493	0.0528	0.0383	0.0415	0.0450	0.0453	0.0459	0.0437
SD	0.0084	0.0071	0.0067	0.0104	0.0059	0.0067	0.0072	0.0078
SE	0.0015	0.0012	0.0012	0.0018	0.0010	0.0012	0.0013	0.0014
		B_{2u}				B_{2g}		
MEAN	0.1173	0.0225	-0.2394	-0.0766	0.0803	0.0551	0.0568	0.0482
SD	0.0956	0.0802	0.0608	0.1066	0.0597	0.0379	0.0359	0.0609
SE	0.0169	0.0142	0.0107	0.0188	0.0105	0.0067	0.0063	0.0108
		B_{1u}				B_{1g}		
MEAN	0.1230	0.0359	-0.6792	-0.9567	-0.2644	-0.2521	-0.2148	-0.1793
SD	0.0408	0.0865	0.0748	0.0837	0.0600	0.0594	0.0431	0.0549
SE	0.0072	0.0153	0.0132	0.0148	0.0106	0.0105	0.0076	0.0097
		A_{2u}				$E_u(x)$		
MEAN	-0.0514	-0.1636	-0.0365	-0.0416	-0.1096	-0.0781	-0.1234	-0.1188
SD	0.0685	0.0606	0.0693	0.0619	0.0759	0.0431	0.0331	0.0664
SE	0.0121	0.0107	0.0122	0.0109	0.0134	0.0076	0.0059	0.0117
		$E_g(x)$				$E_u(y)$		
MEAN	-0.3750	-0.3589	0.0070	-0.0661	-0.0703	-0.0260	-0.0328	-0.0347
SD	0.0563	0.0672	0.0746	0.0589	0.0292	0.0230	0.0243	0.0250
SE	0.0099	0.0119	0.0132	0.0104	0.0052	0.0041	0.0043	0.0044
		$E_g(y)$				A_{1g}		
MEAN	0.0677	0.0670	0.0655	0.1498	0.2568	0.2671	0.2493	0.1993
SD	0.0601	0.0517	0.0622	0.0554	0.0554	0.0514	0.0351	0.0432
SE	0.0106	0.0091	0.0110	0.0098	0.0098	0.0091	0.0062	0.0076
		A_{1u}				A_{2g}		
MEAN	0.0706	0.0113	-0.1827	-0.2380	-0.0782	-0.0792	-0.0877	-0.0762
SD	0.0244	0.0229	0.0348	0.0435	0.0161	0.0186	0.0141	0.0180
SE	0.0043	0.0041	0.0061	0.0077	0.0028	0.0033	0.0025	0.0032

Table A-8: Descriptive statistics of extended basis out-of-plane parameters and lowest energy normal-coordinate displacements for each ETC cofactor in all BBC1 experiments; n = 13.

	B_A	B_B	D_L	D_M	B_A	B_B	D_L	D_M
		D_{oop}				D_{ip}		
MEAN	0.4852	0.4897	0.7864	1.0385	0.4590	0.4333	0.4180	0.3760
SD	0.0749	0.0661	0.0787	0.0771	0.0378	0.0656	0.0396	0.0380
SE	0.0132	0.0117	0.0139	0.0136	0.0067	0.0116	0.0070	0.0067
		δ_{oop}				δ_{ip}		
MEAN	0.0375	0.0367	0.0220	0.0205	0.0283	0.0290	0.0275	0.0254
SD	0.0105	0.0099	0.0073	0.0084	0.0088	0.0091	0.0088	0.0084
SE	0.0019	0.0017	0.0013	0.0015	0.0016	0.0016	0.0016	0.0015
		B_{2u}				B_{2g}		
MEAN	0.1191	0.0234	-0.2389	-0.0765	0.0808	0.0560	0.0578	0.0489
SD	0.0967	0.0802	0.0611	0.1067	0.0597	0.0375	0.0360	0.0608
SE	0.0171	0.0142	0.0108	0.0189	0.0106	0.0066	0.0064	0.0108
		B_{1u}				B_{1g}		
MEAN	0.1230	0.0359	-0.6792	-0.9567	-0.2646	-0.2525	-0.2150	-0.1797
SD	0.0408	0.0865	0.0748	0.0837	0.0610	0.0600	0.0434	0.0551
SE	0.0072	0.0153	0.0132	0.0148	0.0108	0.0106	0.0077	0.0097
		A_{2u}				$E_u(x)$		
MEAN	-0.0567	-0.1664	-0.0365	-0.0421	-0.1101	-0.0787	-0.1239	-0.1193
SD	0.0682	0.0617	0.0698	0.0632	0.0761	0.0431	0.0331	0.0663
SE	0.0121	0.0109	0.0123	0.0112	0.0134	0.0076	0.0059	0.0117
		$E_g(x)$				$E_u(y)$		
MEAN	-0.3728	-0.3527	0.0054	-0.0670	-0.0686	-0.0245	-0.0312	-0.0331
SD	0.0561	0.0658	0.0751	0.0586	0.0286	0.0229	0.0240	0.0253
SE	0.0099	0.0116	0.0133	0.0104	0.0051	0.0041	0.0043	0.0045
		$E_g(y)$				A_{1g}		
MEAN	0.0686	0.0671	0.0645	0.1488	0.2576	0.2679	0.2499	0.2006
SD	0.0602	0.0515	0.0620	0.0557	0.0558	0.0518	0.0352	0.0434
SE	0.0106	0.0091	0.0110	0.0098	0.0099	0.0092	0.0062	0.0077
		A_{1u}				A_{2g}		
MEAN	0.0706	0.0113	-0.1827	-0.2380	-0.0777	-0.0788	-0.0866	-0.0752
SD	0.0244	0.0229	0.0348	0.0435	0.0160	0.0187	0.0143	0.0180
SE	0.0043	0.0041	0.0061	0.0077	0.0028	0.0033	0.0025	0.0032

Table A-9: Descriptive statistics of extended basis out-of-plane next-to-lowest energy normal-coordinate displacements for each ETC cofactor in all BBC1 experiments; n = 13.

	B_A	B_B	D_L	D_M	B_A	B_B	D_L	D_M
		B_{2u}				B_{2g}		
MEAN	0.0670	0.0354	0.0390	0.0236	0.0210	0.0493	0.0062	0.0131
SD	0.0705	0.0287	0.0539	0.0393	0.0400	0.0507	0.0423	0.0332
SE	0.0125	0.0051	0.0095	0.0069	0.0071	0.0090	0.0075	0.0059
		B_{1u}				B_{1g}		
MEAN	-0.0514	-0.0386	0.1249	0.1483	-0.0720	-0.0721	-0.0789	-0.0634
SD	0.0266	0.0426	0.0324	0.0302	0.0608	0.0543	0.0515	0.0549
SE	0.0047	0.0075	0.0057	0.0053	0.0108	0.0096	0.0091	0.0097
		A_{2u}				$E_u(x)$		
MEAN	-0.0752	-0.0243	0.0056	0.0056	-0.0541	-0.0424	-0.0580	-0.0425
SD	0.0657	0.0563	0.0435	0.0442	0.0265	0.0282	0.0280	0.0327
SE	0.0116	0.0100	0.0077	0.0078	0.0047	0.0050	0.0050	0.0058
		$E_g(x)$				$E_u(y)$		
MEAN	0.0955	0.1926	-0.0660	-0.0347	0.1171	0.1132	0.1282	0.1320
SD	0.0424	0.0964	0.0425	0.0593	0.0208	0.0320	0.0141	0.0203
SE	0.0075	0.0170	0.0075	0.0105	0.0037	0.0057	0.0025	0.0036
		$E_g(y)$				A_{1g}		
MEAN	0.0303	0.0062	-0.0135	0.0184	-0.0038	-0.0027	0.0208	0.0485
SD	0.0581	0.0603	0.0705	0.0615	0.0415	0.0378	0.0465	0.0519
SE	0.0103	0.0107	0.0125	0.0109	0.0073	0.0067	0.0082	0.0092

	B_A	B_B	D_L	D_M	B_A	B_B	D_L	D_M
		A_{1u}				A_{2g}		
MEAN	-0.0201	0.0018	0.0948	0.1111	-0.0003	0.0029	0.0170	0.0080
SD	0.0159	0.0187	0.0309	0.0449	0.0323	0.0317	0.0335	0.0326
SE	0.0028	0.0033	0.0055	0.0079	0.0057	0.0056	0.0059	0.0058

Table A-10: Descriptive statistics of minimum basis in- and out-of-plane parameters and normal-coordinate displacements for each ETC cofactor in all RBP2 experiments; n = 16.

	H_A	H_B	H_A	H_B
	D_{oop}		D_{ip}	
MEAN	0.1482	0.1301	0.6048	0.6185
SD	0.0442	0.0545	0.0630	0.0642
SE	0.0114	0.0141	0.0163	0.0166
	δ_{oop}		δ_{ip}	
MEAN	0.0039	0.0037	0.0596	0.0613
SD	0.0010	0.0010	0.0035	0.0049
SE	0.0003	0.0002	0.0009	0.0013
	B_{2u}		B_{2g}	
MEAN	-0.0426	-0.0119	0.2010	0.1948
SD	0.0741	0.0825	0.1171	0.1008
SE	0.0191	0.0213	0.0302	0.0260
	B_{1u}		B_{1g}	
MEAN	0.1017	0.0839	-0.3097	-0.4009
SD	0.0650	0.0618	0.2054	0.0861
SE	0.0168	0.0159	0.0530	0.0222
	A_{2u}		$E_u(x)$	
MEAN	0.0279	0.0295	-0.0645	-0.0524
SD	0.0329	0.0284	0.0401	0.0494
SE	0.0085	0.0073	0.0104	0.0128
	$E_g(x)$		$E_u(y)$	
MEAN	-0.0077	-0.0175	-0.0852	-0.0891
SD	0.0128	0.0198	0.0437	0.0354
SE	0.0033	0.0051	0.0113	0.0091
	$E_g(y)$		A_{1g}	
MEAN	-0.0006	0.0077	0.4041	0.3869
SD	0.0180	0.0123	0.0550	0.0873
SE	0.0046	0.0032	0.0142	0.0225
	A_{1u}		A_{2g}	
MEAN	0.0151	0.0110	-0.0115	-0.0200
SD	0.0065	0.0040	0.0108	0.0044
SE	0.0017	0.0010	0.0028	0.0011

Table A-11: Descriptive statistics of minimum basis in- and out-of-plane parameters and normal-coordinate displacements for each ETC cofactor in all RBP3 experiments; n = 17.

	H_A	H_B	H_A	H_B
	D_{oop}		D_{ip}	
MEAN	0.6118	0.6362	0.6870	0.7161
SD	0.0638	0.1280	0.0670	0.0488
SE	0.0213	0.0427	0.0223	0.0163
	δ_{oop}		δ_{ip}	
MEAN	0.0648	0.0768	0.0605	0.0653
SD	0.0178	0.0254	0.0049	0.0073
SE	0.0059	0.0085	0.0016	0.0024
	B_{2u}		B_{2g}	
MEAN	0.0365	0.2596	0.3392	0.3309
SD	0.0722	0.1128	0.0883	0.0581
SE	0.0241	0.0376	0.0294	0.0194

	H_A	H_B	H_A	H_B
		B_{1u}		B_{1g}
MEAN	0.4511	0.4048	-0.4287	-0.4732
SD	0.1022	0.1135	0.0809	0.0986
SE	0.0341	0.0378	0.0270	0.0329
	A_{2u}		$E_u(x)$	
MEAN	0.2451	0.2729	-0.1415	-0.0723
SD	0.0802	0.1109	0.0387	0.0509
SE	0.0267	0.0370	0.0129	0.0170
	$E_g(x)$		$E_u(y)$	
MEAN	-0.0564	-0.1562	-0.0955	-0.1070
SD	0.0742	0.1017	0.0358	0.0571
SE	0.0247	0.0339	0.0119	0.0190
	$E_g(y)$		A_{1g}	
MEAN	-0.0355	0.0019	0.3354	0.3609
SD	0.0798	0.1304	0.0850	0.0895
SE	0.0266	0.0435	0.0283	0.0298
	A_{1u}		A_{2g}	
MEAN	0.2711	0.1321	-0.1118	-0.0934
SD	0.0590	0.1090	0.0191	0.0250
SE	0.0197	0.0363	0.0064	0.0083

Table A-12: Descriptive statistics of minimum basis in- and out-of-plane parameters and normal-coordinate displacements for each ETC cofactor in all RBP4 experiments; n = 16.

	H_A	H_B	H_A	H_B
	D_{oop}		D_{ip}	
MEAN	0.5585	0.5080	0.5779	0.6320
SD	0.0514	0.0779	0.0498	0.0818
SE	0.0171	0.0260	0.0166	0.0273
	δ_{oop}		δ_{ip}	
MEAN	0.0260	0.0232	0.0553	0.0561
SD	0.0074	0.0060	0.0068	0.0046
SE	0.0025	0.0020	0.0023	0.0015
	B_{2u}		B_{2g}	
MEAN	0.0200	0.1073	0.1757	0.1929
SD	0.0766	0.0633	0.1255	0.1563
SE	0.0255	0.0211	0.0418	0.0521
	B_{1u}		B_{1g}	
MEAN	0.4724	0.4156	-0.3540	-0.3919
SD	0.0659	0.0821	0.0731	0.0851
SE	0.0220	0.0274	0.0244	0.0284
	A_{2u}		$E_u(x)$	
MEAN	0.2084	0.1835	-0.0855	-0.0360
SD	0.0609	0.0567	0.0348	0.0535
SE	0.0203	0.0189	0.0116	0.0178
	$E_g(x)$		$E_u(y)$	
MEAN	-0.0677	-0.1116	-0.1022	-0.1241
SD	0.0862	0.0525	0.0433	0.0438
SE	0.0287	0.0175	0.0144	0.0146
	$E_g(y)$		A_{1g}	
MEAN	-0.0004	0.0373	0.3586	0.3909
SD	0.0511	0.0894	0.0520	0.0601
SE	0.0170	0.0298		
	A_{1u}		A_{2g}	
MEAN	0.1394	0.0912	-0.0898	-0.0839
SD	0.0301	0.0241	0.0230	0.0286
SE	0.0100	0.0080	0.0077	0.0095

Table A-13: Descriptive statistics of minimum basis in- and out-of-plane parameters and normal-coordinate displacements for each ETC cofactor in all RBP5 experiments; n = 9.

	H_A	H_B	H_A	H_B
		D_{oop}		D_{ip}
MEAN	0.4511	0.4571	0.3801	0.3615
SD	0.1181	0.0972	0.0681	0.0323
SE	0.0394	0.0324	0.0227	0.0108
	δ_{oop}		δ_{ip}	
MEAN	0.0169	0.0159	0.0549	0.0541
SD	0.0069	0.0076	0.0041	0.0016
SE	0.0023	0.0025	0.0014	0.0005
	B_{2u}		B_{2g}	
MEAN	0.0265	0.0769	-0.0469	-0.0557
SD	0.1259	0.0397	0.0657	0.1468
SE	0.0420	0.0132	0.0219	0.0489
	B_{1u}		B_{1g}	
MEAN	0.3857	0.3992	-0.2718	-0.2325
SD	0.1314	0.1119	0.0769	0.0526
SE	0.0438	0.0373	0.0256	0.0175
	A_{2u}		$E_u(x)$	
MEAN	0.0981	0.1564	-0.0748	-0.0376
SD	0.1047	0.0655	0.0391	0.0521
SE	0.0349	0.0218	0.0130	0.0174
	$E_g(x)$		$E_u(y)$	
MEAN	-0.0111	-0.0329	-0.0409	-0.0483
SD	0.0706	0.0592	0.0425	0.0417
SE	0.0235	0.0197	0.0142	0.0139
	$E_g(y)$		A_{1g}	
MEAN	-0.0560	0.0141	0.2143	0.1965
SD	0.0877	0.0607	0.0711	0.0432
SE	0.0292	0.0202	0.0237	0.0144
	A_{1u}		A_{2g}	
MEAN	0.0460	0.0415	-0.0462	-0.0679
SD	0.0333	0.0460	0.0266	0.0160
SE	0.0111	0.0153	0.0089	0.0053

Table A-14: Descriptive statistics of minimum basis in- and out-of-plane parameters and normal-coordinate displacements for each ETC cofactor in all BBP1 experiments; n = 13.

	H_A	H_B	H_A	H_B
	D_{oop}		D_{ip}	
MEAN	0.5271	0.6099	0.5529	0.5665
SD	0.0650	0.1122	0.0984	0.0417
SE	0.0180	0.0311	0.0273	0.0116
		δ_{oop}		δ_{ip}
MEAN	0.0301	0.0378	0.0493	0.0485
SD	0.0066	0.0093	0.0081	0.0101
SE	0.0018	0.0026	0.0022	0.0028
	B_{2u}		B_{2g}	
MEAN	-0.2420	-0.1369	0.3464	0.2090
SD	0.1193	0.1210	0.0923	0.0882
SE	0.0331	0.0336	0.0256	0.0245
	B_{1u}		B_{1g}	
MEAN	0.3812	0.5148	-0.2115	-0.3877
SD	0.0540	0.1144	0.1323	0.0713
SE	0.0150	0.0317	0.0367	0.0198
	A_{2u}		$E_u(x)$	
MEAN	0.0571	0.0380	-0.1132	-0.0869
SD	0.0617	0.0756	0.0374	0.0479
SE	0.0171	0.0210	0.0104	0.0133

	H_A	H_B	H_A	H_B
	$E_g(x)$		$E_u(y)$	
MEAN	-0.1818	-0.1591	-0.0338	-0.0265
SD	0.0709	0.0603	0.0340	0.0268
SE	0.0197	0.0167	0.0094	0.0074
		$E_g(y)$		A_{1g}
MEAN	-0.0243	-0.0140	0.3136	0.3117
SD	0.0616	0.1088	0.0627	0.0421
SE	0.0171	0.0302	0.0174	0.0117
	A_{1u}		A_{2g}	
MEAN	0.1082	0.1547	-0.0838	-0.0801
SD	0.0435	0.0648	0.0264	0.0217
SE	0.0121	0.0180	0.0073	0.0060

RC BChl Averaged Coordinates

Table A-15: Atomic coordinates of the best estimate (RBC4) of the conformation of the *R. sphaeroides* B_A macrocycle (angstroms). n = 32.

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
N1	-2.0330	-0.0311	-0.1061	-2.0343	0.0010	-0.0364	-2.0640	0.0060	-0.0550
N2	-0.0114	2.0712	-0.0518	0.0173	2.0694	-0.0922	0.0210	2.0591	-0.0870
N3	2.0005	0.0297	0.0226	1.9808	0.0648	0.0100	2.0010	0.0630	0.0209
N4	0.0145	-2.0566	-0.0520	0.0389	-2.0811	-0.0675	0.0311	-2.0689	-0.0721
Cm1	-2.4203	2.4363	-0.0426	-2.4322	2.4073	-0.0540	-2.4150	2.4080	-0.0390
Cm2	2.4323	2.4467	0.0382	2.4712	2.4492	0.0570	2.4500	2.4411	0.0339
Cm3	2.4201	-2.4268	0.0774	2.4181	-2.3973	0.0866	2.4431	-2.3819	0.0650
Cm4	-2.4220	-2.4607	-0.0821	-2.3987	-2.4028	-0.1020	-2.4069	-2.4080	-0.0790
Ca1	-2.8697	-1.1304	-0.0406	-2.8580	-1.0843	-0.0143	-2.8569	-1.1040	-0.0100
Ca2	-2.8410	1.0898	-0.0160	-2.8698	1.0820	0.0158	-2.8770	1.1050	0.0290
Ca3	-1.0982	2.9136	-0.0658	-1.1022	2.8813	-0.0976	-1.0910	2.8481	-0.0780
Ca4	1.1050	2.8852	0.0204	1.1346	2.8982	0.0165	1.1410	2.8662	-0.0111
Ca5	2.8500	1.1140	0.0092	2.8683	1.1134	0.0127	2.8810	1.1152	0.0129
Ca6	2.8112	-1.0866	0.0334	2.7735	-1.0541	0.0228	2.7661	-1.0438	0.0229
Ca7	1.1085	-2.8912	0.0423	1.1269	-2.9201	0.0529	1.1471	-2.8930	0.0379
Ca8	-1.0859	-2.8853	-0.0879	-1.0744	-2.8832	-0.1144	-1.0869	-2.8500	-0.0910
Cb1	-4.2498	-0.6850	0.1380	-4.2539	-0.6820	0.1170	-4.2269	-0.6970	0.1080
Cb2	-4.2267	0.6769	0.1548	-4.2597	0.6713	0.1471	-4.2470	0.6790	0.1410
Cb3	-0.6441	4.3007	-0.0160	-0.6724	4.2724	-0.0208	-0.7361	4.3121	-0.0850
Cb4	0.7168	4.2755	0.0470	0.6958	4.2812	0.0737	0.7429	4.3021	0.1420
Cb5	4.2241	0.6599	-0.0353	4.2357	0.6287	-0.0362	4.2180	0.6352	-0.0271
Cb6	4.1931	-0.6963	-0.0187	4.1627	-0.7184	-0.0232	4.1471	-0.7388	-0.0221
Cb7	0.6923	-4.2786	0.0614	0.6989	-4.3134	0.0809	0.7542	-4.3459	0.1429
Cb8	-0.6661	-4.2709	-0.0299	-0.6672	-4.2836	-0.0345	-0.7358	-4.3089	-0.1000

Table A-16: Atomic coordinates of the best estimate (RBC4) of the conformation of the *R. sphaeroides* B_B macrocycle (angstroms). n = 32.

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
N1	-2.0131	-0.0312	-0.0726	-2.0167	0.0037	-0.0515	-2.0511	0.0151	-0.0651
N2	-0.0115	2.1156	-0.0030	0.0191	2.1124	-0.0180	0.0171	2.0960	-0.0150
N3	1.9921	0.0267	0.0591	1.9732	0.0622	0.0416	1.9960	0.0629	0.0540
N4	0.0135	-2.0685	-0.0675	0.0395	-2.0930	-0.0657	0.0349	-2.0819	-0.0771
Cm1	-2.4221	2.4442	0.1166	-2.4265	2.4145	0.1126	-2.4060	2.4141	0.1299
Cm2	2.4318	2.4575	-0.0584	2.4681	2.4566	-0.0397	2.4501	2.4449	-0.0530
Cm3	2.4084	-2.4438	0.1793	2.3987	-2.4141	0.1834	2.4229	-2.3981	0.1669
Cm4	-2.4116	-2.4724	-0.2415	-2.3824	-2.4077	-0.2434	-2.3921	-2.4109	-0.2401
Ca1	-2.8430	-1.1398	-0.1080	-2.8328	-1.0910	-0.1046	-2.8411	-1.1139	-0.0901

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
Ca2	-2.8271	1.0862	0.0997	-2.8566	1.0788	0.1158	-2.8610	1.1072	0.1159
Ca3	-1.1051	2.9532	0.0229	-1.1035	2.9224	0.0087	-1.0949	2.8901	0.0180
Ca4	1.1069	2.9295	-0.0640	1.1350	2.9399	-0.0599	1.1401	2.9070	-0.0760
Ca5	2.8431	1.1122	-0.0030	2.8639	1.1113	-0.0082	2.8760	1.1120	-0.0130
Ca6	2.7949	-1.1011	0.0962	2.7544	-1.0680	0.0902	2.7519	-1.0541	0.0850
Ca7	1.1004	-2.9102	0.1088	1.1145	-2.9441	0.1168	1.1348	-2.9130	0.1100
Ca8	-1.0832	-2.9010	-0.1996	-1.0678	-2.8951	-0.2044	-1.0812	-2.8639	-0.1711
Cb1	-4.2218	-0.7098	0.1015	-4.2298	-0.7133	0.0917	-4.2000	-0.7288	0.0829
Cb2	-4.2087	0.6519	0.2338	-4.2442	0.6402	0.2339	-4.2230	0.6422	0.2319
Cb3	-0.6554	4.3478	-0.0581	-0.6839	4.3225	-0.0573	-0.7559	4.3541	-0.1000
Cb4	0.7096	4.3263	-0.1191	0.6873	4.3309	-0.1174	0.7452	4.3650	-0.0730
Cb5	4.2152	0.6364	-0.0373	4.2306	0.6107	-0.0363	4.2090	0.6179	-0.0320
Cb6	4.1768	-0.7215	0.0178	4.1455	-0.7386	0.0171	4.1279	-0.7621	0.0230
Cb7	0.6833	-4.2992	0.1076	0.6888	-4.3406	0.1196	0.7368	-4.3770	0.1639
Cb8	-0.6734	-4.2891	-0.1111	-0.6747	-4.3007	-0.1249	-0.7362	-4.3250	-0.1761

Table A-17: Atomic coordinates of the best estimate (RBC4) of the conformation of the *R. sphaeroides* D_L macrocycle (angstroms). n = 32.

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
N1	-2.0275	-0.0179	-0.1108	-2.0302	0.0160	-0.0650	-2.0610	0.0169	-0.0801
N2	-0.0274	2.1116	-0.1629	0.0013	2.1080	-0.1517	0.0010	2.0979	-0.1541
N3	1.9919	0.0150	-0.0654	1.9740	0.0499	-0.1208	2.0030	0.0509	-0.1171
N4	0.0307	-2.0808	-0.0856	0.0545	-2.1041	-0.0603	0.0430	-2.0900	-0.0661
Cm1	-2.4338	2.4619	0.0024	-2.4429	2.4381	-0.0303	-2.4301	2.4399	-0.0201
Cm2	2.4209	2.4464	-0.0849	2.4600	2.4499	-0.0733	2.4399	2.4339	-0.1081
Cm3	2.4282	-2.4537	0.1170	2.4221	-2.4285	0.1153	2.4430	-2.4221	0.1040
Cm4	-2.4042	-2.4641	-0.0550	-2.3806	-2.4066	-0.0847	-2.3821	-2.4071	-0.0672
Ca1	-2.8527	-1.1299	-0.0035	-2.8421	-1.0852	0.0059	-2.8461	-1.1040	0.0108
Ca2	-2.8426	1.1034	-0.0047	-2.8708	1.0990	0.0138	-2.8780	1.1209	0.0239
Ca3	-1.1130	2.9615	-0.0537	-1.1144	2.9319	-0.0720	-1.1061	2.8969	-0.0551
Ca4	1.0977	2.9167	-0.1044	1.1265	2.9260	-0.0827	1.1330	2.8989	-0.0880
Ca5	2.8362	1.1052	-0.0511	2.8571	1.1068	-0.0762	2.8760	1.1079	-0.0561
Ca6	2.8050	-1.1066	0.0778	2.7657	-1.0766	0.0508	2.7600	-1.0641	0.0440
Ca7	1.1198	-2.9259	0.0144	1.1348	-2.9574	0.0411	1.1530	-2.9311	0.0240
Ca8	-1.0723	-2.9055	-0.1128	-1.0601	-2.9006	-0.1251	-1.0711	-2.8721	-0.0981
Cb1	-4.2356	-0.6938	0.1990	-4.2406	-0.6971	0.1939	-4.2110	-0.7121	0.1829
Cb2	-4.2250	0.6699	0.1904	-4.2568	0.6591	0.1961	-4.2410	0.6689	0.1948
Cb3	-0.6470	4.3525	0.0772	-0.6747	4.3263	0.0667	-0.7471	4.3569	0.0169
Cb4	0.7174	4.3163	0.0405	0.6958	4.3183	0.0539	0.7579	4.3469	0.1039
Cb5	4.2114	0.6439	0.0914	4.2254	0.6182	0.1057	4.2019	0.6290	0.1070
Cb6	4.1842	-0.7136	0.1755	4.1523	-0.7313	0.1911	4.1290	-0.7490	0.1950
Cb7	0.6984	-4.3148	0.0002	0.7045	-4.3521	0.0215	0.7650	-4.3841	0.0750
Cb8	-0.6606	-4.2977	-0.0909	-0.6612	-4.3079	-0.1134	-0.7320	-4.3301	-0.1721

Table A-18: Atomic coordinates of the best estimate (RBC4) of the conformation of the *R. sphaeroides* D_M macrocycle (angstroms). n = 32.

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
N1	-2.0255	-0.0205	-0.0368	-2.0277	0.0208	-0.0370	-2.0599	0.0190	-
N2	-0.0175	2.0985	-0.0570	0.0095	2.0927	-0.0571	0.0111	2.0860	-
N3	2.0002	0.0191	0.0111	1.9823	0.0644	-0.0007	2.0051	0.0650	-
N4	0.0199	-2.0833	-0.0460	0.0438	-2.1112	-0.0361	0.0361	-2.0960	-
Cm1	-2.4267	2.4539	-0.1880	-2.4396	2.4252	-0.1768	-2.4209	2.4310	-
Cm2	2.4247	2.4524	0.1747	2.4615	2.4561	0.1689	2.4370	2.4460	0.1711
Cm3	2.4251	-2.4471	-0.1331	2.4121	-2.4079	-0.1212	2.4381	-2.3960	-

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
Cm4	-2.4153	-2.4639	0.1402	-2.3800	-2.3996	0.1270	-2.3889	-2.3990	0.1390
Ca1	-2.8562	-1.1274	0.0999	-2.8363	-1.0757	0.1096	-2.8409	-1.0950	0.1030
Ca2	-2.8383	1.1011	-0.0629	-2.8754	1.0964	-0.0728	-2.8849	1.1240	-
Ca3	-1.1064	2.9430	-0.1785	-1.1097	2.9075	-0.1726	-1.0980	2.8760	-
Ca4	1.1000	2.9122	0.0828	1.1268	2.9222	0.0780	1.1301	2.8920	0.0561
Ca5	2.8422	1.1129	0.1092	2.8671	1.1164	0.1179	2.8811	1.1140	0.1101
Ca6	2.8139	-1.1004	-0.0353	2.7712	-1.0605	-0.0535	2.7651	-1.0470	-
Ca7	1.1132	-2.9227	-0.1340	1.1259	-2.9532	-0.1197	1.1471	-2.9230	-
Ca8	-1.0840	-2.9093	0.0467	-1.0669	-2.9123	0.0414	-1.0809	-2.8750	0.0390
Cb1	-4.2373	-0.6848	0.1903	-4.2369	-0.6850	0.1903	-4.2049	-0.7030	0.1920
Cb2	-4.2218	0.6779	0.0942	-4.2629	0.6693	0.0952	-4.2449	0.6740	0.0909
Cb3	-0.6538	4.3338	-0.1451	-0.6789	4.3024	-0.1677	-0.7349	4.3280	-
Cb4	0.7090	4.3080	0.0433	0.6902	4.3112	0.0636	0.7330	4.3390	0.0941
Cb5	4.2183	0.6594	0.1192	4.2371	0.6232	0.1242	4.2211	0.6290	0.1321
Cb6	4.1946	-0.6989	0.0363	4.1633	-0.7263	0.0358	4.1431	-0.7460	0.0271
Cb7	0.6908	-4.3130	-0.1311	0.6954	-4.3536	-0.1515	0.7671	-4.3880	-
Cb8	-0.6693	-4.3008	-0.0001	-0.6717	-4.3223	0.0148	-0.7559	-4.3550	-

Table A-19: Atomic coordinates of the best estimate of the conformation of the *B.viridis* B_A macrocycle (angstroms); n = 13.

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
N1	-2.0231	-0.0432	-0.0973	-2.0189	-0.0088	-0.0423	-2.0310	-0.0041	-0.0600
N2	-0.0005	2.1162	-0.0092	0.0168	2.1130	-0.0383	0.0270	2.1050	-0.0599
N3	1.9734	0.0402	0.0815	1.9564	0.0763	0.0449	1.9719	0.0800	0.0621
N4	0.0052	-2.0844	-0.0411	0.0203	-2.1090	-0.0999	0.0210	-2.1121	-0.1170
Cm1	-2.4065	2.4444	-0.0258	-2.4374	2.4342	-0.0312	-2.4191	2.4480	0.0091
Cm2	2.4387	2.4665	0.0541	2.4855	2.4760	0.1111	2.4599	2.4660	0.0571
Cm3	2.4028	-2.4314	0.1112	2.4086	-2.4075	0.1494	2.4450	-2.4021	0.0980
Cm4	-2.4195	-2.4894	-0.1426	-2.4118	-2.4485	-0.1082	-2.4120	-2.4551	-0.0800
Ca1	-2.8578	-1.1511	-0.0445	-2.8385	-1.1133	-0.0149	-2.8360	-1.1221	-0.0110
Ca2	-2.8242	1.0862	-0.0008	-2.8519	1.0866	0.0332	-2.8481	1.1009	0.0301
Ca3	-1.0885	2.9567	-0.0472	-1.1045	2.9324	-0.0723	-1.0911	2.9100	-0.0309
Ca4	1.1185	2.9319	0.0174	1.1501	2.9352	0.0331	1.1400	2.9370	0.0391
Ca5	2.8310	1.1217	0.0513	2.8511	1.1272	0.0331	2.8559	1.1220	0.0441
Ca6	2.7755	-1.0855	0.0412	2.7415	-1.0555	0.0503	2.7229	-1.0470	0.0480
Ca7	1.1014	-2.9154	0.0908	1.1204	-2.9387	0.0901	1.1380	-2.9040	0.0839
Ca8	-1.0864	-2.9249	-0.1452	-1.0872	-2.9290	-0.1775	-1.1110	-2.9031	-0.1280
Cb1	-4.2356	-0.7017	0.1627	-4.2255	-0.7054	0.1351	-4.2200	-0.7241	0.1300
Cb2	-4.2101	0.6626	0.1800	-4.2341	0.6509	0.1617	-4.2330	0.6539	0.1491
Cb3	-0.6296	4.3521	-0.0644	-0.6472	4.3240	-0.0629	-0.6991	4.3489	-0.1689
Cb4	0.7344	4.3281	-0.0274	0.7250	4.3208	-0.0336	0.7669	4.3160	0.0582
Cb5	4.1988	0.6491	-0.0590	4.2088	0.6235	-0.0836	4.2130	0.6300	-0.0839
Cb6	4.1562	-0.7065	-0.0800	4.1261	-0.7236	-0.0756	4.1100	-0.7409	-0.0569
Cb7	0.7024	-4.3097	0.0857	0.7067	-4.3358	0.1225	0.7320	-4.3501	0.2219
Cb8	-0.6561	-4.3125	-0.0914	-0.6605	-4.3249	-0.1243	-0.7030	-4.3530	-0.2341

Table A-20: Atomic coordinates of the best estimate of the conformation of the *B.viridis* B_B macrocycle (angstroms); n = 13.

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
N1	-2.0204	-0.0363	-0.1362	-2.0149	-0.0044	-0.0351	-2.0309	0.0039	-0.0510
N2	-0.0075	2.1059	-0.0380	0.0079	2.1033	-0.0548	0.0191	2.0989	-0.0720
N3	1.9849	0.0352	0.0349	1.9693	0.0712	-0.0503	1.9821	0.0839	-0.0411
N4	0.0108	-2.0940	-0.0699	0.0211	-2.1181	-0.0928	0.0210	-2.1190	-0.1040
Cm1	-2.4111	2.4485	-0.0546	-2.4410	2.4388	-0.0899	-2.4160	2.4479	-0.0521

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
Cm2	2.4308	2.4656	0.0794	2.4761	2.4724	0.1323	2.4501	2.4590	0.0748
Cm3	2.4123	-2.4375	0.0752	2.4154	-2.4114	0.1085	2.4581	-2.4120	0.0610
Cm4	-2.4210	-2.4802	-0.1101	-2.4163	-2.4418	-0.1211	-2.4189	-2.4480	-0.0870
Ca1	-2.8576	-1.1403	-0.0627	-2.8413	-1.1024	-0.0194	-2.8399	-1.1091	-0.0140
Ca2	-2.8227	1.0934	-0.0261	-2.8451	1.0947	0.0178	-2.8440	1.1099	0.0090
Ca3	-1.0943	2.9491	-0.0733	-1.1122	2.9257	-0.1116	-1.0979	2.9060	-0.0711
Ca4	1.1085	2.9232	0.0500	1.1370	2.9265	0.0789	1.1260	2.9250	0.0799
Ca5	2.8347	1.1242	0.0300	2.8583	1.1269	-0.0030	2.8591	1.1230	0.0119
Ca6	2.7917	-1.0880	0.0129	2.7565	-1.0589	-0.0145	2.7391	-1.0500	-0.0041
Ca7	1.1067	-2.9279	0.0582	1.1246	-2.9461	0.0900	1.1441	-2.9110	0.0741
Ca8	-1.0888	-2.9264	-0.1061	-1.0906	-2.9347	-0.1522	-1.1139	-2.9100	-0.1100
Cb1	-4.2353	-0.6845	0.1527	-4.2271	-0.6789	0.1222	-4.2199	-0.6960	0.1071
Cb2	-4.2080	0.6789	0.1763	-4.2271	0.6764	0.1664	-4.2229	0.6749	0.1650
Cb3	-0.6412	4.3425	-0.0301	-0.6674	4.3155	-0.0494	-0.7130	4.3430	-0.1551
Cb4	0.7214	4.3185	0.0567	0.7033	4.3119	0.0790	0.7351	4.3030	0.1738
Cb5	4.2061	0.6633	-0.0586	4.2203	0.6296	-0.0642	4.2261	0.6420	-0.0561
Cb6	4.1724	-0.6938	-0.0733	4.1423	-0.7192	-0.0540	4.1271	-0.7451	-0.0491
Cb7	0.6941	-4.3221	0.1154	0.7100	-4.3442	0.1590	0.7371	-4.3530	0.2611
Cb8	-0.6664	-4.3172	-0.0027	-0.6591	-4.3327	-0.0419	-0.7069	-4.3671	-0.1510

Table A-21: Atomic coordinates of the best estimate of the conformation of the *B.viridis* D_L macrocycle (angstroms); n = 13.

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
N1	-2.0343	-0.0394	-0.0276	-2.0301	0.0000	-0.0430	-2.0419	0.0050	-
N2	-0.0085	2.0966	0.0215	0.0104	2.0936	0.0006	0.0251	2.0861	-
N3	1.9783	0.0380	-0.0310	1.9605	0.0760	0.0175	1.9750	0.0831	0.0000
N4	0.0138	-2.0820	-0.0097	0.0295	-2.1093	-0.0173	0.0220	-2.1050	-
Cm1	-2.4113	2.4466	-0.2244	-2.4359	2.4314	-0.1772	-2.4210	2.4420	-
Cm2	2.4341	2.4579	0.2524	2.4801	2.4580	0.2049	2.4570	2.4430	0.2380
Cm3	2.4146	-2.4286	-0.2567	2.4112	-2.3979	-0.2368	2.4510	-2.3899	-
Cm4	-2.4201	-2.4804	0.2265	-2.4067	-2.4258	0.1977	-2.4090	-2.4380	0.1890
Ca1	-2.8709	-1.1436	0.0474	-2.8511	-1.0980	0.0703	-2.8450	-1.1080	0.0670
Ca2	-2.8337	1.0924	-0.1195	-2.8628	1.0898	-0.1541	-2.8640	1.1030	-
Ca3	-1.0907	2.9437	-0.1486	-1.1026	2.9220	-0.1285	-1.0900	2.8980	-
Ca4	1.1125	2.9083	0.2210	1.1450	2.9074	0.1814	1.1340	2.9090	0.1691
Ca5	2.8333	1.1201	0.0723	2.8572	1.1217	0.1253	2.8580	1.1180	0.1080
Ca6	2.7861	-1.0821	-0.1471	2.7478	-1.0459	-0.1611	2.7311	-1.0379	-
Ca7	1.1116	-2.9133	-0.1635	1.1263	-2.9430	-0.1632	1.1421	-2.9090	-
Ca8	-1.0828	-2.9165	0.2119	-1.0819	-2.9173	0.2026	-1.1029	-2.8940	0.1689
Cb1	-4.2504	-0.6864	-0.0315	-4.2437	-0.6996	-0.0211	-4.2339	-0.7120	-
Cb2	-4.2211	0.6784	-0.1074	-4.2491	0.6559	-0.1172	-4.2500	0.6600	-
Cb3	-0.6242	4.3347	-0.0748	-0.6411	4.3114	-0.1159	-0.6870	4.3421	-
Cb4	0.7380	4.3030	0.2048	0.7295	4.2939	0.2595	0.7650	4.2870	0.2301
Cb5	4.2032	0.6592	-0.0083	4.2193	0.6402	-0.0028	4.2241	0.6431	0.0079
Cb6	4.1660	-0.6957	-0.1215	4.1352	-0.7051	-0.1445	4.1210	-0.7189	-
Cb7	0.7081	-4.3063	-0.0349	0.7109	-4.3417	-0.0998	0.7231	-4.3740	-
Cb8	-0.6515	-4.3047	0.2489	-0.6578	-4.3178	0.3227	-0.6839	-4.3329	0.3759

Table A-22: Atomic coordinates of the best estimate of the conformation of the *B.viridis* D_M macrocycle (angstroms); n = 13.

	Minimum			Extended			Complete		
Atom	x	y	z	x	y	z	x	y	z
N1	-2.0301	-0.0342	-0.0348	-2.0272	0.0055	-0.0413	-2.0420	0.0100	-0.0450
N2	-0.0076	2.0803	0.0280	0.0067	2.0755	0.0286	0.0160	2.0760	0.0102
N3	1.9762	0.0327	-0.0032	1.9632	0.0725	0.0241	1.9770	0.0749	0.0102

	Minimum			Extended			Complete		
Atom	x	y	z	x	y	z	x	y	z
N4	0.0127	-2.0644	-0.0433	0.0242	-2.0907	-0.0602	0.0170	-2.0950	-0.0370
Cm1	-2.4109	2.4409	-0.3202	-2.4275	2.4232	-0.2775	-2.4140	2.4340	-0.2909
Cm2	2.4309	2.4484	0.3855	2.4626	2.4438	0.3308	2.4400	2.4350	0.3493
Cm3	2.4134	-2.4238	-0.3573	2.3968	-2.3847	-0.3198	2.4380	-2.3730	-0.3389
Cm4	-2.4168	-2.4703	0.2894	-2.3919	-2.4151	0.2540	-2.4000	-2.4240	0.2559
Ca1	-2.8646	-1.1368	0.0905	-2.8425	-1.0911	0.1280	-2.8370	-1.0970	0.1309
Ca2	-2.8335	1.0911	-0.1316	-2.8634	1.0890	-0.1746	-2.8630	1.0990	-0.1670
Ca3	-1.0894	2.9258	-0.2499	-1.0971	2.9050	-0.2241	-1.0829	2.8840	-0.1789
Ca4	1.1101	2.8928	0.2914	1.1319	2.8906	0.2638	1.1120	2.9000	0.2543
Ca5	2.8291	1.1133	0.1627	2.8552	1.1143	0.2060	2.8580	1.1060	0.1873
Ca6	2.7871	-1.0820	-0.1795	2.7511	-1.0435	-0.2014	2.7350	-1.0359	-0.1528
Ca7	1.1089	-2.8952	-0.2508	1.1144	-2.9201	-0.2452	1.1250	-2.8910	-0.2580
Ca8	-1.0809	-2.8992	0.2374	-1.0736	-2.9058	0.2161	-1.0950	-2.8800	0.1859
Cb1	-4.2438	-0.6874	0.0578	-4.2375	-0.6973	0.0762	-4.2260	-0.7100	0.0829
Cb2	-4.2192	0.6760	-0.0431	-4.2503	0.6580	-0.0611	-4.2490	0.6580	-0.0691
Cb3	-0.6271	4.3139	-0.2154	-0.6416	4.2942	-0.2757	-0.6450	4.3080	-0.2988
Cb4	0.7337	4.2846	0.1899	0.7253	4.2783	0.2564	0.7280	4.2740	0.2653
Cb5	4.1989	0.6584	0.0580	4.2211	0.6353	0.0652	4.2310	0.6370	0.0733
Cb6	4.1657	-0.6954	-0.1284	4.1415	-0.7102	-0.1446	4.1250	-0.7240	-0.1627
Cb7	0.7075	-4.2849	-0.1003	0.7117	-4.3203	-0.1704	0.7280	-4.3510	-0.1661
Cb8	-0.6505	-4.2846	0.2672	-0.6531	-4.3062	0.3467	-0.6761	-4.3151	0.3599

Table A-23: Atomic coordinates of the best estimate of the conformation of the H_A macrocycle (RBP4, angstroms); n = 16.

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
N1	-2.0206	-0.0702	0.0520	-2.0312	-0.0431	0.0126	-2.0640	-0.0411	0.0170
N2	0.0188	2.1611	0.0649	0.0639	2.1581	0.0640	0.0530	2.1450	0.0640
N3	1.9820	0.0659	0.0840	1.9612	0.0971	0.0680	1.9970	0.0960	0.0779
N4	-0.0152	-2.1149	0.0651	0.0204	-2.1351	0.0809	0.0131	-2.1430	0.0710
Cm1	-2.3942	2.4382	0.1551	-2.3825	2.4025	0.1625	-2.3620	2.3979	0.1779
Cm2	2.4618	2.5038	-0.1489	2.4985	2.5104	-0.1121	2.4750	2.5010	-0.1170
Cm3	2.3834	-2.4314	0.1852	2.3755	-2.4149	0.1784	2.3962	-2.4060	0.1871
Cm4	-2.4391	-2.5249	-0.1786	-2.4064	-2.4593	-0.1400	-2.4179	-2.4541	-0.1550
Ca1	-2.8652	-1.1795	-0.0489	-2.8702	-1.1324	-0.0822	-2.8719	-1.1461	-0.0710
Ca2	-2.8109	1.0705	0.0795	-2.8343	1.0632	0.0898	-2.8400	1.0919	0.0699
Ca3	-1.0814	2.9898	0.1347	-1.0696	2.9495	0.1336	-1.0611	2.9039	0.1129
Ca4	1.1376	2.9870	-0.1033	1.1659	3.0106	-0.0907	1.1699	2.9800	-0.0730
Ca5	2.8518	1.1470	-0.0367	2.8710	1.1465	-0.0540	2.8920	1.1501	-0.0510
Ca6	2.7657	-1.0797	0.0920	2.7198	-1.0578	0.1048	2.7071	-1.0590	0.0791
Ca7	1.0848	-2.9415	0.1618	1.1037	-2.9793	0.1584	1.1261	-2.9470	0.1590
Ca8	-1.1062	-2.9651	-0.1302	-1.0810	-2.9511	-0.1021	-1.0879	-2.9261	-0.0800
Cb1	-4.2397	-0.7181	-0.0551	-4.2632	-0.7059	-0.0603	-4.2369	-0.7162	-0.0641
Cb2	-4.2009	0.6479	0.0044	-4.2345	0.6539	0.0134	-4.2170	0.6508	0.0199
Cb3	-0.6329	4.3969	0.0315	-0.6838	4.3665	0.0581	-0.7611	4.4009	0.0719
Cb4	0.7355	4.3867	-0.1513	0.6889	4.4030	-0.1978	0.7629	4.4430	-0.2121
Cb5	4.2169	0.6445	-0.0984	4.2291	0.6083	-0.1014	4.2090	0.6232	-0.1099
Cb6	4.1532	-0.7147	-0.0387	4.1142	-0.7440	-0.0391	4.0841	-0.7639	-0.0239
Cb7	0.6874	-4.3444	0.0505	0.7049	-4.3874	0.0672	0.7632	-4.4050	0.0671
Cb8	-0.6728	-4.3549	-0.1708	-0.6602	-4.3594	-0.2119	-0.7288	-4.3761	-0.2179

Table A-24: Atomic coordinates of the best estimate of the conformation of the H_B macrocycle (RBP4, angstroms); n = 16.

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
N1	-2.0084	-0.0730	0.0400	-2.0203	-0.0430	0.0286	-2.0559	-0.0349	0.0270
N2	0.0236	2.1799	0.0596	0.0721	2.1752	0.0616	0.0671	2.1760	0.0640
N3	1.9922	0.0677	0.0931	1.9715	0.1037	0.0468	2.0081	0.1010	0.0581
N4	-0.0220	-2.1235	0.0419	0.0128	-2.1448	0.0546	0.0080	-2.1550	0.0421
Cm1	-2.3941	2.4363	0.1332	-2.3812	2.4030	0.1371	-2.3699	2.3931	0.1599
Cm2	2.4656	2.5161	-0.1110	2.5066	2.5261	-0.0707	2.4821	2.5170	-0.0780
Cm3	2.3762	-2.4392	0.1662	2.3638	-2.4231	0.1786	2.3730	-2.4140	0.1751
Cm4	-2.4427	-2.5307	-0.1771	-2.4085	-2.4635	-0.1537	-2.4160	-2.4639	-0.1660
Ca1	-2.8541	-1.1834	-0.0504	-2.8614	-1.1331	-0.0651	-2.8609	-1.1539	-0.0499
Ca2	-2.8015	1.0654	0.0928	-2.8239	1.0627	0.1072	-2.8309	1.0901	0.0900
Ca3	-1.0850	3.0011	0.0915	-1.0732	2.9567	0.0839	-1.0639	2.9141	0.0678
Ca4	1.1390	3.0118	-0.0818	1.1674	3.0373	-0.0600	1.1731	3.0080	-0.0531
Ca5	2.8612	1.1530	-0.0150	2.8874	1.1539	-0.0391	2.9061	1.1560	-0.0380
Ca6	2.7698	-1.0859	0.0984	2.7163	-1.0652	0.0944	2.7051	-1.0630	0.0751
Ca7	1.0753	-2.9527	0.1293	1.0939	-2.9909	0.1398	1.1140	-2.9610	0.1402
Ca8	-1.1146	-2.9745	-0.1333	-1.0864	-2.9629	-0.1210	-1.0950	-2.9329	-0.0908
Cb1	-4.2268	-0.7271	-0.0165	-4.2530	-0.7072	-0.0229	-4.2279	-0.7159	-0.0270
Cb2	-4.1900	0.6377	0.0625	-4.2227	0.6519	0.0641	-4.1998	0.6551	0.0660
Cb3	-0.6505	4.4130	-0.0285	-0.7150	4.3786	-0.0239	-0.7928	4.4071	-0.0291
Cb4	0.7198	4.4129	-0.1586	0.6592	4.4297	-0.1848	0.7342	4.4590	-0.1791
Cb5	4.2272	0.6400	-0.0774	4.2448	0.5925	-0.0751	4.2231	0.6050	-0.0829
Cb6	4.1601	-0.7220	-0.0197	4.1145	-0.7631	-0.0171	4.0901	-0.7790	-0.0009
Cb7	0.6701	-4.3571	0.0272	0.7002	-4.4014	0.0424	0.7590	-4.4230	0.0592
Cb8	-0.6902	-4.3658	-0.1665	-0.6647	-4.3732	-0.2058	-0.7300	-4.3850	-0.2298

Table A-25: Atomic coordinates of the best estimate of the conformation of the H_A macrocycle (RBP5, angstroms); n = 10.

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
N1	-2.0051	-0.0024	0.0305	-2.0048	0.0173	0.0458	-2.0270	0.0239	0.0440
N2	-0.0230	2.1026	0.0159	0.0140	2.1128	-0.0095	0.0169	2.0900	-0.0080
N3	1.9711	0.0007	0.0359	1.9423	0.0239	0.0251	1.9470	0.0351	0.0270
N4	0.0262	-2.0842	0.0426	0.0611	-2.1114	0.0558	0.0591	-2.0950	0.0540
Cm1	-2.4247	2.4665	0.1230	-2.4231	2.4013	0.1203	-2.4051	2.4029	0.1171
Cm2	2.4063	2.4397	-0.1449	2.4371	2.4188	-0.1213	2.4270	2.4111	-0.1250
Cm3	2.4233	-2.4571	0.1529	2.4300	-2.4128	0.1457	2.4402	-2.4079	0.1470
Cm4	-2.3944	-2.4548	-0.1250	-2.3682	-2.3739	-0.1244	-2.3759	-2.3781	-0.1190
Ca1	-2.8240	-1.1159	-0.0466	-2.8214	-1.0585	-0.0494	-2.8450	-1.0670	-0.0531
Ca2	-2.8259	1.1121	0.0693	-2.8509	1.0783	0.0867	-2.8720	1.0969	0.0930
Ca3	-1.1060	2.9516	0.0892	-1.1049	2.9178	0.0755	-1.1031	2.8880	0.0760
Ca4	1.0916	2.9137	-0.0894	1.1252	2.9393	-0.0944	1.1289	2.9010	-0.0961
Ca5	2.8045	1.0973	-0.0669	2.8079	1.0766	-0.0651	2.8260	1.0801	-0.0640
Ca6	2.7938	-1.1102	0.0939	2.7631	-1.0679	0.0859	2.7780	-1.0459	0.0821
Ca7	1.1158	-2.9259	0.1041	1.1438	-2.9584	0.1157	1.1511	-2.9390	0.1180
Ca8	-1.0707	-2.9114	-0.0836	-1.0581	-2.9003	-0.0834	-1.0699	-2.8650	-0.0840
Cb1	-4.2043	-0.6862	-0.0468	-4.2231	-0.6767	-0.0650	-4.1890	-0.7061	-0.0660
Cb2	-4.2031	0.6755	0.0207	-4.2421	0.6708	0.0304	-4.2090	0.6859	0.0300
Cb3	-0.6524	4.3407	0.0480	-0.6811	4.3204	0.0561	-0.7571	4.3570	0.0540
Cb4	0.7093	4.3115	-0.0791	0.6919	4.3360	-0.0809	0.7559	4.3710	-0.0771
Cb5	4.1782	0.6553	-0.0701	4.1820	0.6277	-0.0715	4.1380	0.6411	-0.0690
Cb6	4.1672	-0.7005	0.0295	4.1457	-0.7131	0.0300	4.1520	-0.7509	0.0290
Cb7	0.7024	-4.3148	0.0152	0.7019	-4.3549	0.0264	0.7651	-4.3920	0.0270
Cb8	-0.6560	-4.3039	-0.1183	-0.6682	-4.3130	-0.1344	-0.7318	-4.3370	-0.1370

Table A-26: Atomic coordinates of the best estimate of the conformation of the H_B macrocycle (RBP5, angstroms); n = 10.

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
N1	-2.0011	-0.0069	0.0479	-1.9969	0.0068	0.0559	-2.0150	0.0139	0.0540
N2	-0.0320	2.0930	0.0476	0.0001	2.1068	0.0519	0.0070	2.0789	0.0511
N3	1.9837	0.0048	0.0634	1.9553	0.0231	0.0511	1.9621	0.0290	0.0531
N4	0.0336	-2.0713	0.0409	0.0630	-2.0977	0.0331	0.0640	-2.0771	0.0311
Cm1	-2.4308	2.4590	0.1394	-2.4320	2.3912	0.1269	-2.4180	2.3969	0.1241
Cm2	2.4043	2.4371	-0.1283	2.4325	2.4094	-0.1130	2.4260	2.4069	-0.1170
Cm3	2.4256	-2.4561	0.1477	2.4367	-2.4085	0.1467	2.4461	-2.3931	0.1480
Cm4	-2.3938	-2.4467	-0.1491	-2.3761	-2.3698	-0.1368	-2.3849	-2.3751	-0.1340
Ca1	-2.8248	-1.1137	-0.0541	-2.8243	-1.0578	-0.0514	-2.8470	-1.0681	-0.0500
Ca2	-2.8224	1.1053	0.0978	-2.8410	1.0675	0.1034	-2.8620	1.0839	0.1060
Ca3	-1.1147	2.9427	0.0911	-1.1186	2.9130	0.0899	-1.1121	2.8859	0.0890
Ca4	1.0837	2.9014	-0.0822	1.1156	2.9269	-0.0758	1.1209	2.8900	-0.0709
Ca5	2.8191	1.0987	-0.0425	2.8203	1.0717	-0.0525	2.8350	1.0729	-0.0539
Ca6	2.8018	-1.1099	0.0981	2.7764	-1.0675	0.1002	2.7950	-1.0441	0.0980
Ca7	1.1160	-2.9203	0.1015	1.1457	-2.9457	0.1036	1.1541	-2.9241	0.1080
Ca8	-1.0694	-2.8908	-0.0978	-1.0628	-2.8833	-0.1042	-1.0720	-2.8491	-0.1060
Cb1	-4.2043	-0.6846	-0.0518	-4.2229	-0.6678	-0.0606	-4.1940	-0.6991	-0.0650
Cb2	-4.1989	0.6753	0.0420	-4.2316	0.6764	0.0461	-4.1990	0.6928	0.0490
Cb3	-0.6627	4.3300	-0.0084	-0.6926	4.3137	-0.0020	-0.7661	4.3599	0.0000
Cb4	0.6987	4.2978	-0.1306	0.6814	4.3237	-0.1410	0.7360	4.3490	-0.1429
Cb5	4.1943	0.6501	-0.0687	4.1975	0.6200	-0.0785	4.1490	0.6310	-0.0739
Cb6	4.1773	-0.7067	0.0170	4.1609	-0.7206	0.0280	4.1670	-0.7610	0.0251
Cb7	0.6871	-4.3055	0.0100	0.6919	-4.3384	0.0250	0.7540	-4.3801	0.0220
Cb8	-0.6705	-4.2827	-0.1308	-0.6786	-4.2929	-0.1462	-0.7460	-4.3201	-0.1450

Table A-27: Atomic coordinates of the best estimate of the conformation of the *B.viridis* H_A macrocycle (BBP1, angstroms); n = 13.

	Minimum			Extended			Complete		
Atom	x	y	z	x	y	z	x	y	z
N1	-2.0454	-0.1085	-0.0429	-2.0421	-0.0745	-0.0225	-2.0437	-0.0767	-0.0240
N2	0.0625	2.1091	0.0302	0.0886	2.1087	0.0301	0.1007	2.1146	0.0315
N3	1.9941	0.1070	0.0438	1.9722	0.1450	0.0087	1.9903	0.1623	0.0143
N4	-0.0576	-2.0938	0.0418	-0.0354	-2.1216	0.0609	-0.0427	-2.1263	0.0534
Cm1	-2.3597	2.3935	0.0897	-2.3848	2.3689	0.0582	-2.3590	2.3821	0.0852
Cm2	2.4986	2.5226	-0.0988	2.5481	2.5282	-0.0644	2.5203	2.5132	-0.0856
Cm3	2.3621	-2.3773	0.1815	2.3684	-2.3475	0.1627	2.4042	-2.3595	0.1469
Cm4	-2.4851	-2.5435	-0.1689	-2.4720	-2.4890	-0.1611	-2.4711	-2.4944	-0.1530
Ca1	-2.9148	-1.1986	-0.0935	-2.8990	-1.1521	-0.1124	-2.8969	-1.1563	-0.1180
Ca2	-2.8126	1.0433	0.0026	-2.8414	1.0363	0.0356	-2.8387	1.0537	0.0289
Ca3	-1.0428	2.9211	0.1061	-1.0557	2.8901	0.0820	-1.0576	2.8620	0.0772
Ca4	1.1683	2.9551	-0.0327	1.2042	2.9663	-0.0087	1.1893	2.9647	-0.0102
Ca5	2.8811	1.1762	-0.0699	2.9006	1.1769	-0.1031	2.9056	1.1697	-0.0851
Ca6	2.7683	-1.0347	0.0457	2.7291	-0.9998	0.0491	2.7172	-0.9917	0.0352
Ca7	1.0613	-2.8919	0.1835	1.0851	-2.9203	0.1885	1.0941	-2.8890	0.1834
Ca8	-1.1405	-2.9611	-0.1011	-1.1370	-2.9621	-0.0911	-1.1543	-2.9326	-0.0664
Cb1	-4.2845	-0.7062	-0.0555	-4.2817	-0.7038	-0.0654	-4.2768	-0.7227	-0.0588
Cb2	-4.2141	0.6600	-0.0122	-4.2433	0.6522	-0.0038	-4.2508	0.6580	-0.0140
Cb3	-0.6110	4.3268	0.1252	-0.6382	4.2961	0.1550	-0.7033	4.3089	0.1232
Cb4	0.7530	4.3398	0.0162	0.7354	4.3409	-0.0150	0.8011	4.3412	0.0176
Cb5	4.2418	0.6784	-0.1792	4.2520	0.6459	-0.1745	4.2569	0.6574	-0.1890
Cb6	4.1631	-0.6787	-0.1219	4.1294	-0.7007	-0.1121	4.1067	-0.7162	-0.0911
Cb7	0.6876	-4.2990	0.1614	0.6946	-4.3315	0.2071	0.7284	-4.3440	0.2462
Cb8	-0.6737	-4.3396	-0.0513	-0.6770	-4.3526	-0.1037	-0.7198	-4.3785	-0.1477

Table A-28: Atomic coordinates of the best estimate of the conformation of the *B.viridis* H_B macrocycle (BBP1, angstroms); n = 13.

	Min.			Ext.			Comp.		
Atom	x	y	z	x	y	z	x	y	z
N1	-2.0050	-0.0740	-0.0359	-2.0019	-0.0388	-0.0574	-2.0019	-0.0265	-0.0588
N2	0.0297	2.1411	0.0189	0.0561	2.1398	0.0136	0.0640	2.1552	0.0107
N3	1.9656	0.0728	0.0400	1.9436	0.1131	0.0289	1.9511	0.1295	0.0303
N4	-0.0260	-2.1291	0.0255	-0.0039	-2.1572	0.0479	-0.0130	-2.1668	0.0412
Cm1	-2.3736	2.4329	0.1440	-2.4017	2.4105	0.1140	-2.3722	2.4168	0.1330
Cm2	2.4557	2.5107	-0.1490	2.5069	2.5208	-0.0954	2.4760	2.5135	-0.1212
Cm3	2.3753	-2.4205	0.2210	2.3828	-2.3914	0.1776	2.4206	-2.4024	0.1793
Cm4	-2.4453	-2.5268	-0.2136	-2.4318	-2.4752	-0.1663	-2.4307	-2.4838	-0.1703
Ca1	-2.8573	-1.1747	-0.0969	-2.8400	-1.1287	-0.1396	-2.8423	-1.1321	-0.1331
Ca2	-2.7868	1.0716	0.0362	-2.8169	1.0672	0.0642	-2.8124	1.0826	0.0525
Ca3	-1.0651	2.9693	0.1344	-1.0796	2.9356	0.1236	-1.0808	2.9058	0.1177
Ca4	1.1365	2.9771	-0.0919	1.1725	2.9900	-0.0802	1.1544	2.9923	-0.0676
Ca5	2.8316	1.1574	-0.0733	2.8518	1.1604	-0.1011	2.8571	1.1576	-0.0721
Ca6	2.7527	-1.0652	0.0710	2.7122	-1.0310	0.0923	2.7034	-1.0261	0.0707
Ca7	1.0792	-2.9467	0.2008	1.1041	-2.9736	0.1970	1.1131	-2.9424	0.1797
Ca8	-1.1151	-2.9815	-0.1531	-1.1108	-2.9853	-0.1265	-1.1315	-2.9561	-0.0935
Cb1	-4.2267	-0.6948	-0.0323	-4.2217	-0.6882	-0.0564	-4.2233	-0.6999	-0.0592
Cb2	-4.1765	0.6699	0.0273	-4.2060	0.6667	0.0572	-4.2074	0.6608	0.0578
Cb3	-0.6292	4.3746	0.1267	-0.6587	4.3403	0.1883	-0.7068	4.3497	0.1720
Cb4	0.7349	4.3718	-0.0495	0.7149	4.3732	-0.1186	0.7684	4.3651	-0.1035
Cb5	4.1940	0.6731	-0.1385	4.2036	0.6349	-0.1550	4.2115	0.6393	-0.1716
Cb6	4.1373	-0.6847	-0.0709	4.1019	-0.7126	-0.0519	4.0859	-0.7272	-0.0330
Cb7	0.6880	-4.3531	0.1631	0.6974	-4.3838	0.2174	0.7255	-4.3982	0.2501
Cb8	-0.6740	-4.3714	-0.1039	-0.6747	-4.3865	-0.1737	-0.7089	-4.4068	-0.2113

Appendix B

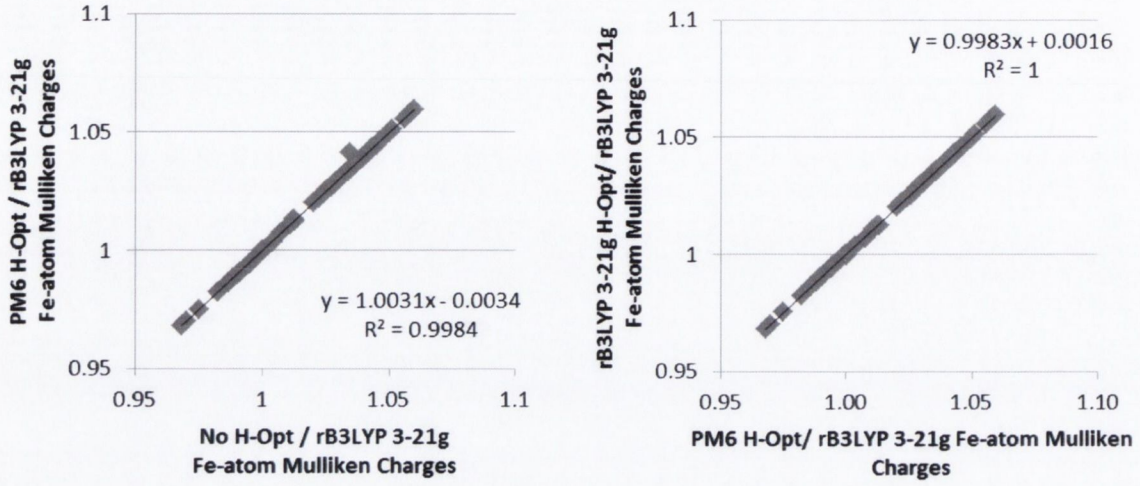


Figure B-1: Comparison of Fe-atom Mulliken charges obtained from B3LYP/3-21g single-point energies with different methods of H-atom optimisation for the input structures.

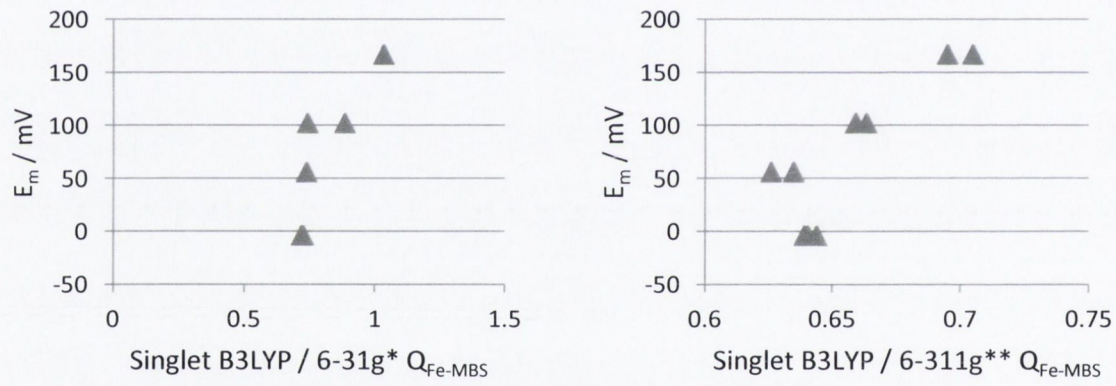


Figure B-2: Plots of the H-NOX complex midpoint potential *vs.* $Q_{Fe\text{-}MBS}$ obtained from B3LYP/ 6-31g* (left) and 6-311g** (right) singlet state calculations of the FeP-comp models.

Table B-1: Triplet B3LYP/6-31g* Q_{Fe} of H-NOX haem macrocycle-only models.

	Fe-atom Mulliken Atomic Charge / AU									
	Haem Diacid		**FeP**		**FeP-comp.**		**FeP-ext.**		**FeP-min.**	
	Std.	*MBS*	*Std.*	*MBS*	*Std.*	*MBS*	*Std.*	*MBS*	*Std.*	*MBS*
1U56 - A500	1.085	0.839	1.089	0.840	1.087	0.839	0.914	0.703	1.089	0.832
1U56 - B501	0.925	0.707	1.064	0.812	1.063	0.812	1.084	0.799	1.086	0.800
3NVR - A200	1.080	0.764	1.079	0.774	1.077	0.776	1.064	0.796	1.067	0.795
3NVR - B200	1.078	0.779	1.082	0.782	1.082	0.780	1.089	0.810	1.065	0.793
3NVU - A200	1.075	0.764	1.079	0.767	1.078	0.768	1.067	0.779	1.068	0.778
3NVU - B200	1.081	0.769	1.084	0.772	1.085	0.772	1.092	0.790	1.092	0.789
3EEE - A200	1.071	0.747	1.077	0.751	1.078	0.751	1.090	0.783	1.092	0.785
3EEE - B200	1.075	0.739	1.076	0.740	1.076	0.740	1.090	0.775	1.067	0.761
3EEE - C200	1.079	0.747	1.083	0.750	1.083	0.749	1.093	0.775	1.065	0.758
3EEE - D200	1.082	0.748	1.083	0.751	1.082	0.751	1.093	0.779	1.065	0.761

Table B-2: Triplet B3LYP/6-311g** Q_{Fe} of H-NOX haem macrocycle-only models.

	Fe-atom Mulliken Atomic Charge / AU									
	Haem Diacid		**FeP**		**FeP-comp.**		**FeP-ext.**		**FeP-min.**	
	Std.	*MBS*	*Std.*	*MBS*	*Std.*	*MBS*	*Std.*	*MBS*	*Std.*	*MBS*
1U56 - A500	1.279	0.664	1.276	0.668	1.275	0.667	1.242	0.642	1.245	0.642
1U56 - B501	1.283	0.652	1.126	0.560	1.254	0.646	1.277	0.637	1.278	0.637
3NVR - A200	1.279	0.615	1.276	0.616	1.276	0.618	1.252	0.635	1.257	0.633
3NVR - B200	1.280	0.617	1.279	0.622	1.279	0.621	1.250	0.631	1.255	0.631
3NVU - A200	1.280	0.606	1.279	0.610	1.279	0.611	1.259	0.621	1.261	0.621
3NVU - B200	1.286	0.610	1.285	0.614	1.285	0.614	1.285	0.630	1.285	0.629
3EEE - A200	1.280	0.595	1.280	0.601	1.281	0.599	1.285	0.626	1.288	0.627
3EEE - B200	1.285	0.590	1.283	0.593	1.283	0.593	1.288	0.622	1.264	0.610
3EEE - C200	1.289	0.596	1.289	0.601	1.289	0.600	1.292	0.622	1.263	0.607
3EEE - D200	1.291	0.598	1.288	0.601	1.288	0.601	1.291	0.625	1.261	0.609

Table B-3: Triplet B3LYP/6-311+g** Q_{Fe} of H-NOX haem macrocycle-only models.

	Fe-atom Mulliken Atomic Charge / AU									
	Haem Diacid		**FeP**		**FeP-comp.**		**FeP-ext.**		**FeP-min.**	
	Std.	*MBS*	*Std.*	*MBS*	*Std.*	*MBS*	*Std.*	*MBS*	*Std.*	*MBS*
1U56 - A500			0.386	0.558	0.389	0.557	0.428	0.534	0.424	0.535
1U56 - B501			0.436	0.540	0.376	0.533	0.387	0.522	0.384	0.522
3NVR - A200			0.323	0.499	0.347	0.501	0.358	0.525	0.393	0.523
3NVR - B200			0.385	0.504	0.373	0.503	0.384	0.522	0.424	0.521
3NVU - A200			0.358	0.491	0.340	0.491	0.338	0.506	0.360	0.506
3NVU - B200			0.420	0.494	0.376	0.494	0.367	0.514	0.390	0.513
3EEE - A200			0.440	0.480	0.440	0.479	0.432	0.511	0.434	0.500
3EEE - B200			0.416	0.471	0.414	0.471	0.401	0.504	0.419	0.496
3EEE - C200			0.420	0.479	0.421	0.478	0.405	0.504	0.401	0.493
3EEE - D200			0.384	0.479	0.393	0.479	0.386	0.507	0.389	0.503

Table B-4: Structure-cluster means of the min. basis out-of-plane distortions from *B. viridis* RC-cyt haems.

		Normal-mode displacement / Å					
		B2u	B1u	A2u	Eg(x)	Eg(y)	A1u
Cluster 1	H1	0.7232	0.1785	0.0337	0.1176	-0.0417	-0.0120
	H2	0.4395	0.3771	-0.0169	0.0765	-0.0500	-0.0088
	H3	0.4243	0.5478	-0.0235	0.0427	0.1342	0.0067
	H4	0.2267	0.1736	0.0112	0.0093	0.0759	0.0193
Cluster 2	H1	0.7504	0.2723	-0.1078	0.0922	-0.2141	-0.0148
	H2	0.4649	0.414	-0.1081	0.0149	-0.2328	0.0206
	H3	0.4430	0.6042	-0.1610	-0.1370	-0.1016	0.0107
	H4	0.2638	0.3107	-0.1797	0.0516	-0.0706	0.0105
Cluster 3	H1	0.4482	0.2969	0.0151	0.1519	0.0371	-0.0944
	H2	0.4428	0.5053	-0.0643	0.0878	-0.0864	0.0019
	H3	0.4102	0.4816	-0.0356	0.0717	0.0855	0.0045
	H4	0.2161	0.2937	0.0438	0.0185	0.1311	0.0277

Table B-5: Structure-cluster standard errors of the min. basis out-of-plane distortions from *B. viridis* RC-cyt haems.

		Twice the standard error of the mean normal-mode displacement / Å					
		B2u	B1u	A2u	Eg(x)	Eg(y)	A1u
Cluster 1	H1	0.0692	0.0595	0.0714	0.0702	0.0522	0.0334
	H2	0.0384	0.0305	0.0600	0.0403	0.0308	0.0302
	H3	0.0545	0.0877	0.0210	0.0254	0.0524	0.0103
	H4	0.0559	0.0785	0.0648	0.0586	0.0428	0.0304
Cluster 2	H1	0.0159	0.2567	0.0849	0.0793	0.1040	0.0234
	H2	0.1251	0.1898	0.0659	0.0249	0.1213	0.0228
	H3	0.0032	0.0765	0.0404	0.0695	0.0288	0.0394
	H4	0.1121	0.0751	0.1250	0.0639	0.0241	0.0045
Cluster 3	H1	0.0319	0.0602	0.0272	0.0381	0.0350	0.0091
	H2	0.0306	0.0599	0.0202	0.0213	0.0320	0.0191
	H3	0.0217	0.0556	0.0308	0.0426	0.0267	0.0182
	H4	0.0394	0.0882	0.0505	0.0299	0.0276	0.0308

Table B-6: Structure-cluster means of the min. basis in-plane distortions from *B. viridis* RC-cyt haems.

		Normal-mode displacement / Å					
		B2g	B1g	Eu(x)	Eu(y)	A1g	A2g
Cluster 1	H1	-0.1336	0.0116	0.0319	0.0503	0.0836	-0.0029
	H2	-0.1290	0.0792	-0.0104	-0.0292	0.0885	-0.0024
	H3	-0.0608	0.0073	0.0751	-0.0144	0.0753	0.0149
	H4	-0.0056	0.0192	0.0090	0.0096	0.0777	0.0053
Cluster 2	H1	-0.0649	0.0405	-0.0130	0.0001	0.1590	0.0131
	H2	-0.0124	-0.0757	-0.0028	-0.0100	0.1690	0.0170
	H3	0.0418	-0.0240	-0.0014	-0.0001	0.1506	0.0196
	H4	0.0134	-0.0553	0.0145	0.0088	0.1568	0.0159
Cluster 3	H1	-0.0801	0.0833	0.0623	0.0105	0.0800	0.0051
	H2	-0.1139	-0.0433	-0.0097	-0.0049	0.1200	0.0008
	H3	-0.0093	0.0313	0.0095	-0.0485	0.0517	-0.0015
	H4	-0.0444	-0.0784	0.0149	0.0141	0.1001	0.0019

Table B-7: Structure-cluster standard errors of the min. basis in distortions from *B. viridis* RC cytochrome subunit haems.

		Twice the standard error of the mean normal-mode displacement / Å					
		B2g	B1g	Eu(x)	Eu(y)	A1g	A2g
Cluster 1	H1	0.0986	0.0452	0.0207	0.0579	0.1405	0.0089
	H2	0.1301	0.0736	0.0412	0.0335	0.1517	0.0120
	H3	0.0675	0.0656	0.0399	0.0123	0.1130	0.0242
	H4	0.0813	0.0992	0.0193	0.0235	0.0350	0.0123
Cluster 2	H1	0.0406	0.0293	0.0326	0.0155	0.0674	0.0098
	H2	0.0434	0.0219	0.0182	0.0239	0.0640	0.0146
	H3	0.0365	0.0849	0.0237	0.0394	0.0485	0.0118
	H4	0.0575	0.0190	0.0235	0.0296	0.0761	0.0059
Cluster 3	H1	0.0295	0.0247	0.0160	0.0125	0.0292	0.0013
	H2	0.0641	0.0249	0.0158	0.0148	0.0381	0.0041
	H3	0.0462	0.0262	0.0121	0.0189	0.0319	0.0032
	H4	0.0529	0.0174	0.0217	0.0117	0.0155	0.0054

Table B-8: B3LYP SCF energies (in Hartrees) of the diacid models of the H-NOx cofactors (NC = no convergence).

	3-21g		6-31g*		6-311g**		6-311+g**	
	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet
1u56 - A500	-3082.400899	NC	-	-3098.448233	-3098.946382	-3098.997255	-	-
1u56 - B501	-3082.372683	-3082.437892	-	-3098.437527	-3098.923322	-3098.973694	-	-
3eee - A200	-3082.365046	-3082.394042	-	-3098.418729	-3098.920138	-3098.971398	-	-
3eee - B200	-3082.337188	-3082.366305	-	-3098.398347	-3098.899396	-3098.950563	-	-
3eee - C200	-3082.369827	-3082.398315	-	-3098.428308	-3098.929537	-3098.980726	-	-
3eee - D200	-3082.337504	-3082.403095	-	-3098.397199	-3098.898443	-3098.949497	-	-
3nvr - A200	-3082.322035	-3082.360480	-	-3098.377846	-3098.879249	-3098.930470	-	-
3nvr - B200	-3082.370640	-3082.398901	-	-3098.418034	-3098.918546	-3098.969399	-	-
3nvu - A200	-3082.348914	-3082.415678	-	-3098.421754	-3098.924038	-3098.974755	-	-
3nvu - B200	-3082.364679	-3082.392029	-	-3098.421590	-3098.922335	-3098.973373	-	-

Table B-9: B3LYP SCF energies (in Hartrees) of the FeP models of the H-NOx cofactors.

	3-21g		6-31g*		6-311g**		6-311+g**	
	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet
1u56 - A500	-2240.638386	-2240.679288	-	-2252.046553	-	-2252.361629	-	-2252.389330
1u56 - B501	-2240.639394	-2240.680241	-	-2252.054542	-	-2252.373995	-	-2252.392642
3eee - A200	-2240.618413	-2240.656071	-	-2252.022277	-	-2252.340734	-	-2252.370789
3eee - B200	-2240.615772	-2240.682565	-	-2252.021579	-	-2252.340290	-	-2252.370699
3eee - C200	-2240.619875	-2240.672387	-	-2252.023946	-	-2252.342440	-	-2252.372614
3eee - D200	-2240.618845	-2240.684468	-	-2252.023207	-	-2252.341734	-	-2252.371696
3nvr - A200	-2240.595886	-2240.677263	-	-2252.014535	-	-2252.332708	-	-2252.361630
3nvr - B200	-2240.621401	-2240.650529	-	-2252.021366	-	-2252.339053	-	-2252.367970
3nvu - A200	-2240.607033	-2240.688744	-	-2252.024973	-	-2252.343210	-	-2252.372811
3nvu - B200	-2240.629186	-2240.660292	-	-2252.029685	-	-2252.347779	-	-2252.377298

Table B-10: B3LYP SCF energies (in Hartrees) of the FeP-comp. models of the H-NOx cofactors.

	3-21g		6-31g*		6-311g**		6-311+g**	
	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet
1u56 - A500	-2240.632879	-2240.672832	-2251.988080	-2252.040936	-2252.306438	-2252.356418	-2252.336949	-2252.384214
1u56 - B501	-2240.636691	-2240.677492	-2251.993401	-2252.051633	-2252.311828	-2252.367018	-2252.343281	-2252.395080
3eee - A200	-2240.615036	-2240.681583	-2251.964924	-2252.018645	-2252.286682	-2252.337753	-2252.319949	-2252.367733
3eee - B200	-2240.613168	-2240.679820	-2251.964896	-2252.018657	-2252.286836	-2252.337911	-2252.320579	-2252.368281
3eee - C200	-2240.618125	-2240.683885	-2251.968066	-2252.021961	-2252.289989	-2252.340968	-2252.323614	-2252.371115
3eee - D200	-2240.615821	-2240.681473	-2251.954209	-2252.019884	-2252.288030	-2252.338903	-2252.321431	-2252.368871
3nvr - A200	-2240.607104	-2240.659521	-2251.962269	-2252.008872	-2252.276104	-2252.327610	-2252.308511	-2252.356603
3nvr - B200	-2240.616993	-2240.681487	-2251.969340	-2252.016719	-2252.283429	-2252.334982	-2252.315830	-2252.363939
3nvu - A200	-2240.606224	-2240.687748	-2251.975397	-2252.023327	-2252.290970	-2252.342029	-2252.323644	-2252.371622
3nvu - B200	-2240.629224	-2240.694401	-2251.988080	-2252.029490	-2252.296939	-2252.348068	-2252.329626	-2252.377489

Table B-11: B3LYP SCF energies (in Hartrees) of the FeP-ext. models of the H-NOx cofactors.

	3-21g		6-31g*		6-311g**		6-311+g**	
	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet
1u56 - A500	-2240.637086	NC	-	-2252.057804	-	-2252.366062	-	-2252.393479
1u56 - B501	-2240.635064	-2240.714804	-	NC	-	-2252.366656	-	-2252.395378
3eee - A200	-2240.645039	-2240.673961	-	-2252.048335	-	-2252.365430	-	-2252.394575
3eee - B200	-2240.627234	-2240.670898	-	-2252.046738	-	-2252.363897	-	-2252.393409
3eee - C200	-2240.646826	-2240.675277	-	-2252.051735	-	-2252.369051	-	-2252.398458
3eee - D200	-2240.644858	-2240.682272	-	-2252.049279	-	-2252.366493	-	-2252.395762
3nvr - A200	-2240.621035	-2240.701523	-	-2252.041249	-	-2252.357672	-	-2252.385832
3nvr - B200	-2240.630456	-2240.671750	-	-2252.043616	-	-2252.365499	-	-2252.393588
3nvu - A200	-2240.632483	-2240.713775	-	-2252.055413	-	-2252.372369	-	-2252.400963
3nvu - B200	-2240.653694	-2240.681527	-	-2252.056404	-	-2252.373584	-	-2252.402447

Table B-12: B3LYP SCF energies (in Hartrees) of the FeP-min. models of the H-NOx cofactors.

	3-21g		6-31g*		6-311g**		6-311+g**	
	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet
1u56 - A500	-2240.636304	-2240.678368	-	-2252.042223	-	-2252.363262	-	-2252.390846
1u56 - B501	-2240.635673	-2240.680695	-	-2252.050408	-	-2252.367281	-	-2252.395982
3eee - A200	-2240.652073	-2240.681445	-	-2252.054196	-	-2252.371105	-	-2252.405681
3eee - B200	-2240.633636	-2240.682090	-	-2252.058347	-	-2252.375352	-	-2252.404467
3eee - C200	-2240.635279	-2240.678649	-	-2252.060655	-	-2252.377811	-	-2252.406984
3eee - D200	-2240.634318	-2240.715800	-	-2252.058858	-	-2252.375945	-	-2252.399841
3nvr - A200	-2240.632756	-2240.679974	-	-2252.053038	-	-2252.369582	-	-2252.397705
3nvr - B200	-2240.640757	-2240.682348	-	-2252.059180	-	-2252.375611	-	-2252.403601
3nvu - A200	-2240.635221	-2240.683447	-	-2252.059262	-	-2252.376379	-	-2252.404881
3nvu - B200	-2240.655169	-2240.683767	-	-2252.059297	-	-2252.376602	-	-2252.405413

Table B-13: Parameters of 2nd order polynomial obtained by least-squares fitting of H-NOx E_m data to calculated Q_{Fe}: - $E_m = a*Q_{Fe}^2 + b*Q_{Fe} + c$. a, b and c given as multiples of 10^3.

		Heme Diacid				**FeP**				**FeP-comp**				**FeP-ext**				**FeP-min**			
		a	*b*	*c*	R^2	*a*	*b*	*c*	R^2	*a*	*b*	*c*	R^2	*a*	*b*	*c*	R^2	*a*	*b*	*c*	R^2
Singlet	3-21g	-19.64	34.79	-15.23	0.96	-19.62	34.91	-15.35	0.90	-21.46	38.06	-16.70	0.96	-84.89	147.03	-63.50	0.81	-57.73	100.96	-43.97	0.79
	6-31g*									1.70	-1.31	0.00	0.74								
	6-311g**									13.40	-15.64	4.55	0.75								
	6-311+g**									-28.87	33.81	-9.73	0.96								
Triplet	3-21g	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	-1.27	2.69	-1.27	0.80	-0.29	0.68	0.17	0.48	3.40	-5.53	2.24	0.35
	6-31g*	25.16	-38.28	14.59	0.52	-17.99	30.48	-12.74	0.95	-18.13	30.70	-12.82	0.95	59.02	-89.38	33.83	0.72	-9.94	18.41	-8.26	0.77
	6-311g**	-40.49	53.42	-17.45	0.97	42.28	-51.37	15.64	0.60	-40.69	54.03	-17.76	0.96	191.47	-232.90	70.79	0.81	165.88	-201.81	61.38	0.84
	6-311+g**	N/A	N/A	N/A	N/A	-29.05	32.17	-8.74	0.96	-30.44	33.63	-9.12	0.97	-41.46	48.80	-14.07	0.80	2.86	1.78	-1.60	0.88

Table B-14: Parameters of 2nd order polynomial obtained by least-squares fitting of H-NOx Em data to calculated NPA $Core_{Fe}$: - Em = a* $Core_{Fe}^2$ + b* $Core_{Fe}$ + c.

		Heme Diacid				FeP				FeP-comp				FeP-ext				FeP-min			
		a /	*b* /	*c* /	R^2	*a* /	*b* /	*c* /	R^2	*a* / 10^7	*b* /	*c* /	R^2	*a* / 10^7	*b* /	*c* /	R^2	*a* / 10^7	*b* /	*c* /	R^2
Singlet	3-21g	-1.72	6.20	-5.58	0.95	-1.87	6.74	-6.06	0.92	-1.72	6.20	-5.58	0.97	-2.44	8.79	-7.91	0.78	-1.38	4.96	-4.46	0.77
	6-311g**									-1.6	5.72	-5.1	0.96								
	6-311+g**									-0.92	3.30	-2.97	0.95								
Triplet	3-21g																				
	6-31g*	-1.98	7.11	-6.39	0.96	-1.83	6.59	-5.93	0.96	-1.85	6.64	-5.97	0.97	-6.08	21.9	-19.7	0.81	-1.80	6.47	-5.82	0.81
	6-311g**	-1.99	7.17	-6.45	0.97	-2.41	8.66	-7.79	0.95	-2.07	7.44	-6.69	0.97	-3.99	14.4	-12.9	0.84	0.39	-1.40	1.26	0.81
	6-311+g**	N/A	N/A	N/A	N/A	-1.33	4.79	-4.31	0.96	-1.36	4.89	-4.40	0.97	-2.45	8.82	-7.93	0.84	-1.20	4.30	-3.87	0.87

Table B-15: Parameters of linear fits obtained by the method of least-squares of H-NOx Em data (excluding 1U56 structures) to calculated NPA $Core_{Fe}$: - Em = a* $Core_{Fe}$ + b.

		Heme Diacid			FeP			FeP-comp			FeP-ext			FeP-min		
		a / 10^4	*b* / 10^6	R^2	*a* / 10^4	*b* / 10^6	R^2	*a* / 10^4	*b* / 10^6	R^2	*a* / 10^4	*b* / 10^6	R^2	*a* / 10^4	*b* / 10^6	R^2
Singlet	3-21g	-9.22	1.66	0.88	-9.24	1.66	0.79	-9.38	1.69	0.93	-9.45	1.70	0.87	-8.97	1.61	0.82
	6-311g**							-8.39	1.51	0.88						
	6-311+g**							-6.92	1.24	0.92						
Triplet	3-21g	-10.5	1.89	0.84				-9.78	1.76	0.89	-8.04	1.45	0.91	-10.3	1.85	0.89
	6-31g*	-6.23	1.12	0.89	-6.16	1.11	0.88	-6.19	1.11	0.90	-10.3	1.85	0.86	-7.79	1.40	0.88
	6-311g**	-9.95	1.79	0.92	-9.93	1.79	0.91	-10.0	1.80	0.95	-12.1	2.18	0.83	-9.19	1.65	0.72
	6-311+g**	N/A	N/A	N/A	-8.07	1.45	0.91	-8.06	1.45	0.94	-9.11	1.64	0.85	-8.31	1.50	0.90

Table B-8: B3LYP SCF energies (in Hartrees) of the diacid models of the H-NOx cofactors (NC = no convergence).

	3-21g		6-31g*		6-311g**		6-311+g**	
	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet
1u56 - A500	-3082.400899	NC	-	-3098.448233	-3098.946382	-3098.997255	-	-
1u56 - B501	-3082.372683	-3082.437892	-	-3098.437527	-3098.923322	-3098.973694	-	-
3eee - A200	-3082.365046	-3082.394042	-	-3098.418729	-3098.920138	-3098.971398	-	-
3eee - B200	-3082.337188	-3082.366305	-	-3098.398347	-3098.899396	-3098.950563	-	-
3eee - C200	-3082.369827	-3082.398315	-	-3098.428308	-3098.929537	-3098.980726	-	-
3eee - D200	-3082.337504	-3082.403095	-	-3098.397199	-3098.898443	-3098.949497	-	-
3nvr - A200	-3082.322035	-3082.360480	-	-3098.377846	-3098.879249	-3098.930470	-	-
3nvr - B200	-3082.370640	-3082.398901	-	-3098.418034	-3098.918546	-3098.969399	-	-
3nvu - A200	-3082.348914	-3082.415678	-	-3098.421754	-3098.924038	-3098.974755	-	-
3nvu - B200	-3082.364679	-3082.392029	-	-3098.421590	-3098.922335	-3098.973373	-	-

Table B-9: B3LYP SCF energies (in Hartrees) of the FeP models of the H-NOx cofactors.

	3-21g		6-31g*		6-311g**		6-311+g**	
	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet
1u56 - A500	-2240.638386	-2240.679288	-	-2252.046553	-	-2252.361629	-	-2252.389330
1u56 - B501	-2240.639394	-2240.680241	-	-2252.054542	-	-2252.373995	-	-2252.392642
3eee - A200	-2240.618413	-2240.656071	-	-2252.022277	-	-2252.340734	-	-2252.370789
3eee - B200	-2240.615772	-2240.682565	-	-2252.021579	-	-2252.340290	-	-2252.370699
3eee - C200	-2240.619875	-2240.672387	-	-2252.023946	-	-2252.342440	-	-2252.372614
3eee - D200	-2240.618845	-2240.684468	-	-2252.023207	-	-2252.341734	-	-2252.371696
3nvr - A200	-2240.595886	-2240.677263	-	-2252.014535	-	-2252.332708	-	-2252.361630
3nvr - B200	-2240.621401	-2240.650529	-	-2252.021366	-	-2252.339053	-	-2252.367970
3nvu - A200	-2240.607033	-2240.688744	-	-2252.024973	-	-2252.343210	-	-2252.372811
3nvu - B200	-2240.629186	-2240.660292	-	-2252.029685	-	-2252.347779	-	-2252.377298

Table B-10: B3LYP SCF energies (in Hartrees) of the FeP-comp. models of the H-NOx cofactors.

	3-21g		6-31g*		6-311g**		6-311+g**	
	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet
1u56 - A500	-2240.632879	-2240.672832	-2251.988080	-2252.040936	-2252.306438	-2252.356418	-2252.336949	-2252.384214
1u56 - B501	-2240.636691	-2240.677492	-2251.993401	-2252.051633	-2252.311828	-2252.367018	-2252.343281	-2252.395080
3eee - A200	-2240.615036	-2240.681583	-2251.964924	-2252.018645	-2252.286682	-2252.337753	-2252.319949	-2252.367733
3eee - B200	-2240.613168	-2240.679820	-2251.964896	-2252.018657	-2252.286836	-2252.337911	-2252.320579	-2252.368281
3eee - C200	-2240.618125	-2240.683885	-2251.968066	-2252.021961	-2252.289989	-2252.340968	-2252.323614	-2252.371115
3eee - D200	-2240.615821	-2240.681473	-2251.954209	-2252.019884	-2252.288030	-2252.338903	-2252.321431	-2252.368871
3nvr - A200	-2240.607104	-2240.659521	-2251.962269	-2252.008872	-2252.276104	-2252.327610	-2252.308511	-2252.356603
3nvr - B200	-2240.616993	-2240.681487	-2251.969340	-2252.016719	-2252.283429	-2252.334982	-2252.315830	-2252.363939
3nvu - A200	-2240.606224	-2240.687748	-2251.975397	-2252.023327	-2252.290970	-2252.342029	-2252.323644	-2252.371622
3nvu - B200	-2240.629224	-2240.694401	-2251.988080	-2252.029490	-2252.296939	-2252.348068	-2252.329626	-2252.377489

Table B-11: B3LYP SCF energies (in Hartrees) of the FeP-ext. models of the H-NOx cofactors.

	3-21g		6-31g*		6-311g**		6-311+g**	
	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet
1u56 - A500	-2240.637086	NC	-	-2252.057804	-	-2252.366062	-	-2252.393479
1u56 - B501	-2240.635064	-2240.714804	-	NC	-	-2252.366656	-	-2252.395378
3eee - A200	-2240.645039	-2240.673961	-	-2252.048335	-	-2252.365430	-	-2252.394575
3eee - B200	-2240.627234	-2240.670898	-	-2252.046738	-	-2252.363897	-	-2252.393409
3eee - C200	-2240.646826	-2240.675277	-	-2252.051735	-	-2252.369051	-	-2252.398458
3eee - D200	-2240.644858	-2240.682272	-	-2252.049279	-	-2252.366493	-	-2252.395762
3nvr - A200	-2240.621035	-2240.701523	-	-2252.041249	-	-2252.357672	-	-2252.385832
3nvr - B200	-2240.630456	-2240.671750	-	-2252.043616	-	-2252.365499	-	-2252.393588
3nvu - A200	-2240.632483	-2240.713775	-	-2252.055413	-	-2252.372369	-	-2252.400963
3nvu - B200	-2240.653694	-2240.681527	-	-2252.056404	-	-2252.373584	-	-2252.402447

Table B-12: B3LYP SCF energies (in Hartrees) of the FeP-min. models of the H-NOx cofactors.

	3-21g		6-31g*		6-311g**		6-311+g**	
	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet
1u56 - A500	-2240.636304	-2240.678368	-	-2252.042223	-	-2252.363262	-	-2252.390846
1u56 - B501	-2240.635673	-2240.680695	-	-2252.050408	-	-2252.367281	-	-2252.395982
3eee - A200	-2240.652073	-2240.681445	-	-2252.054196	-	-2252.371105	-	-2252.405681
3eee - B200	-2240.633636	-2240.682090	-	-2252.058347	-	-2252.375352	-	-2252.404467
3eee - C200	-2240.635279	-2240.678649	-	-2252.060655	-	-2252.377811	-	-2252.406984
3eee - D200	-2240.634318	-2240.715800	-	-2252.058858	-	-2252.375945	-	-2252.399841
3nvr - A200	-2240.632756	-2240.679974	-	-2252.053038	-	-2252.369582	-	-2252.397705
3nvr - B200	-2240.640757	-2240.682348	-	-2252.059180	-	-2252.375611	-	-2252.403601
3nvu - A200	-2240.635221	-2240.683447	-	-2252.059262	-	-2252.376379	-	-2252.404881
3nvu - B200	-2240.655169	-2240.683767	-	-2252.059297	-	-2252.376602	-	-2252.405413

Table B-13: Parameters of 2nd order polynomial obtained by least-squares fitting of H-NOx E_m data to calculated Q_{Fe}: - $E_m = a*Q_{Fe}^2 + b*Q_{Fe} + c$. a, b and c given as multiples of 10^3.

		Heme Diacid				FeP				FeP-comp				FeP-ext				FeP-min			
		a	*b*	*c*	R^2	*a*	*b*	*c*	R^2	*a*	*b*	*c*	R^2	*a*	*b*	*c*	R^2	*a*	*b*	*c*	R^2
Singlet	3-21g	-19.64	34.79	-15.23	0.96	-19.62	34.91	-15.35	0.90	-21.46	38.06	-16.70	0.96	-84.89	147.03	-63.50	0.81	-57.73	100.96	-43.97	0.79
	6-31g*									1.70	-1.31	0.00	0.74								
	6-311g**									13.40	-15.64	4.55	0.75								
	6-311+g**									-28.87	33.81	-9.73	0.96								
Triplet	3-21g	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	-1.27	2.69	-1.27	0.80	-0.29	0.68	0.17	0.48	3.40	-5.53	2.24	0.35
	6-31g*	25.16	-38.28	14.59	0.52	-17.99	30.48	-12.74	0.95	-18.13	30.70	-12.82	0.95	59.02	-89.38	33.83	0.72	-9.94	18.41	-8.26	0.77
	6-311g**	-40.49	53.42	-17.45	0.97	42.28	-51.37	15.64	0.60	-40.69	54.03	-17.76	0.96	191.47	-232.90	70.79	0.81	165.88	-201.81	61.38	0.84
	6-311+g**	N/A	N/A	N/A	N/A	-29.05	32.17	-8.74	0.96	-30.44	33.63	-9.12	0.97	-41.46	48.80	-14.07	0.80	2.86	1.78	-1.60	0.88

Table B-14: Parameters of 2nd order polynomial obtained by least-squares fitting of H-NOx Em data to calculated NPA $Core_{Fe}$: - Em = a* $Core_{Fe}^2$ + b* $Core_{Fe}$ + c.

		Heme Diacid				FeP				FeP-comp				FeP-ext				FeP-min			
		a /	*b* /	*c* /	R^2	*a* /	*b* /	*c* /	R^2	*a* / 10^7	*b* /	*c* /	R^2	*a* / 10^7	*b* /	*c* /	R^2	*a* / 10^7	*b* /	*c* /	R^2
Singlet	3-21g	-1.72	6.20	-5.58	0.95	-1.87	6.74	-6.06	0.92	-1.72	6.20	-5.58	0.97	-2.44	8.79	-7.91	0.78	-1.38	4.96	-4.46	0.77
	6-311g**									-1.6	5.72	-5.1	0.96								
	6-311+g**									-0.92	3.30	-2.97	0.95								
Triplet	3-21g																				
	6-31g*	-1.98	7.11	-6.39	0.96	-1.83	6.59	-5.93	0.96	-1.85	6.64	-5.97	0.97	-6.08	21.9	-19.7	0.81	-1.80	6.47	-5.82	0.81
	6-311g**	-1.99	7.17	-6.45	0.97	-2.41	8.66	-7.79	0.95	-2.07	7.44	-6.69	0.97	-3.99	14.4	-12.9	0.84	0.39	-1.40	1.26	0.81
	6-311+g**	N/A	N/A	N/A	N/A	-1.33	4.79	-4.31	0.96	-1.36	4.89	-4.40	0.97	-2.45	8.82	-7.93	0.84	-1.20	4.30	-3.87	0.87

Table B-15: Parameters of linear fits obtained by the method of least-squares of H-NOx Em data (excluding 1U56 structures) to calculated NPA $Core_{Fe}$: - Em = a* $Core_{Fe}$ + b.

		Heme Diacid			FeP			FeP-comp			FeP-ext			FeP-min		
		a / 10^4	*b* / 10^6	R^2	*a* / 10^4	*b* / 10^6	R^2	*a* / 10^4	*b* / 10^6	R^2	*a* / 10^4	*b* / 10^6	R^2	*a* / 10^4	*b* / 10^6	R^2
Singlet	3-21g	-9.22	1.66	0.88	-9.24	1.66	0.79	-9.38	1.69	0.93	-9.45	1.70	0.87	-8.97	1.61	0.82
	6-311g**							-8.39	1.51	0.88						
	6-311+g**							-6.92	1.24	0.92						
Triplet	3-21g	-10.5	1.89	0.84				-9.78	1.76	0.89	-8.04	1.45	0.91	-10.3	1.85	0.89
	6-31g*	-6.23	1.12	0.89	-6.16	1.11	0.88	-6.19	1.11	0.90	-10.3	1.85	0.86	-7.79	1.40	0.88
	6-311g**	-9.95	1.79	0.92	-9.93	1.79	0.91	-10.0	1.80	0.95	-12.1	2.18	0.83	-9.19	1.65	0.72
	6-311+g**	N/A	N/A	N/A	-8.07	1.45	0.91	-8.06	1.45	0.94	-9.11	1.64	0.85	-8.31	1.50	0.90