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

Stuart A. MacGowan[a] and Mathias O. Senge[a,*]

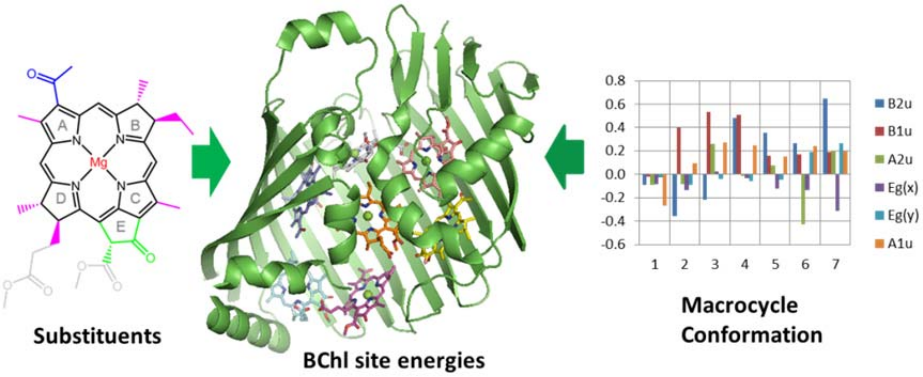
[a] School of Chemistry, SFI Tetrapyrrole Laboratory, Trinity Biomedical Sciences Institute, 152-160 Pearse Street, Trinity College Dublin, The University of Dublin, Dublin 2, Ireland.

To whom correspondence should be addressed. E-mail: sengem@tcd.ie

Abstract

The structural data for the Fenna-Matthews-Olson (FMO) protein indicate that the bacteriochlorophylls (BChls) display a significant degree of conformational heterogeneity of their peripheral substituents and the protein-induced nonplanar skeletal deformations of the tetrapyrrole macrocycle. As electronic properties of chromophores are altered by such differences, a conformational effect may influence the site-energies of specific pigments and thus play a role in mediating the excitation energy transfer dynamics, but this has not yet been established. The difficulty of assessing this question is shown to be partly the result of the inability of the sequential truncation approach usually employed to account for interactions between the conformations of the macrocycle and its substituents and an alternative approach is suggested. By assigning the BChl atoms to meaningful atom groups and performing all possible permutations of partial optimizations in a full-factorial design, where each group is either frozen in the crystal geometry or optimized *in vacuo*, followed by excited state calculations on each resulting structure (PM6//ZIndo/S), the specific effects of the conformations of each BChl component as well as mutual interactions between the molecular fragments on the site-energy can be delineated. This *factorial relaxation* procedure gives different estimates of the macrocycle conformational perturbation than the approach of sequentially truncating the BChl periphery. The results were evaluated in the context of published site-energies for the FMO pigments from three species to identify how conformational effects contribute to their distribution and instances of cross-species conservation and functional divergence of the BChl nonplanarity conformational contribution are described.

Graphical Abstract



Highlights

- The effects of the conformational perturbations of individual moieties of the FMO BChls a on the pigment site-energies are strictly delineated
- The stereoelectronic interplay of BChl molecular constituents are qualitatively described
- The effect of macrocycle distortion on a BChl a site-energy is altered by different chemical environments
- The conformational perturbations brought on by the macrocycle are more conserved amongst the structural representations of spectral type I FMO species
- A critical comparative analysis of the traditional method employed to delineate competing structural effects versus our new method is provided
- The macrocycle conformational effect may be under evolutionary pressure to afford the observed distribution of site-energies

Keywords
Chlorophylls; Factorial Design; Fenna-Matthews-Olson Protein; Light-harvesting; Nonplanar Porphyrins; Partial Optimizations; Photosynthesis; Semi-empirical Quantum Chemistry

Abbreviations
AHC, agglomerative hierarchical clustering; BChl, bacteriochlorophyll; Chl, chlorophyll; $E(BChl)_{cryst}$, *in vacuo* SPE of BChl in protein bound conformation; ΔQy[Mac], site-energy perturbation due to macrocycle distortion; d_{ip}, mean absolute deviation of NSD minimum basis model from observed structure for IP normal-modes; d_{oop}, as previous for OOP modes; D_{ip}, NSD total IP distortion; D_{oop}, NSD total OOP distortion; E_{ES}, excited-state energy; EET, excitation energy transfer; EU, experimental unit (*i.e.*, a unit element in a particular factorial experiment); FMO, Fenna-Matthews-Olson protein; HOMO, highest occupied molecular orbital; IP, in-plane; LHC, light-harvesting complex; LUMO, lowest unoccupied molecular orbital; NSD, normal-coordinate structural decomposition; OLS, ordinary least-squares; OOP, out-of-plane; PDB, Protein Data Bank; RC, reaction center; REP, rotational energy profile; RLM, robust linear modelling; SCF, self-consistent field; SPE, single-point energy.

1. Introduction

Photosynthetic light-harvesting complexes (LHCs) collect energy from sunlight and deliver it to the reaction centers. The pathways of intra- and inter-protein excitation energy transfer (EET) are finely controlled by LHCs to ensure optimum photosynthetic efficiency [1,2]. 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 [2] through which the chromophore spatial distributions are precisely controlled. In contrast, site-energies are modulated by the local interactions between each pigment and its protein binding-site.

There are many ways in which proteins may influence the absorption characteristics of bound chlorophylls (Chls) and consequently affect EET dynamics. These include the choice of axial ligand, the H-bonding environment, the electrostatic properties of the binding-site [3–6] and for bacteriochlorophylls *a* (BChls a) the degree of coplanarity of the C3-acetyl with the macrocycle plane [7,8]. Additionally, because of the well-known correlation between macrocycle conformation and photophysical properties of porphyrins [9–12], it is possible that protein-induced deformation of bound Chls and BChls also contributes significantly to the site-energy distribution in LHC proteins. The photobiological relevance of this is indicated by Zucchelli *et al.*'s analysis of Chl ring deformation in LHCs [13,14], who presented strong evidence for its role in generating the spectral profile of the main antenna complexes of PSII, and our own work relating it to functional properties of reaction centers [15,16].

The Fenna-Matthews-Olson (FMO) protein is a water-soluble pigment protein complex that mediates EET between the chlorosomes and reaction centers (RCs) of green bacteria (Fig. 1) [17]. It crystallizes as a C_3-symmetric trimer, thought to resemble the *in vivo* form, where each monomer contains seven BChls a that are packed closely together within the encapsulating protein; notably, there is no apparent symmetry within the monomer [18–22]. These BChls are known to display significant conformational variation with respect to the local geometric parameters of their substituents as well as in the nonplanar conformations of their tetrapyrrole macrocycles. Additionally, an eighth BChl is found in crystal structures from some species, although which species are in possession of BChl 8 remains an open question [23,24]. The dynamics of exciton transfer in the FMO have recently been modelled and suggest alternate pathways wherein 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 [24].

<Fig. 1>

The FMO protein is considered an excellent model system for the study of photosynthetic EET in general and its optical spectra have been the subject of substantial theoretical analysis [7,8,25–28]. Due to the availability of high-resolution structural data [18–24,29–32] a number of these studies have focused on predicting the spectral characteristics and EET dynamics directly from the atomic structure. Initially, Fajer and coworkers in 1990 [26] employed the semi-empirical ZIndo/S method. The structural models derived from the experimental coordinates consisted of the BChls *in vacuo* subject to various levels of truncation, with and without axial ligand and, where relevant, including nearby charged- and aromatic residues. However, whilst this established an idea of the overall effect of the presence of a particular perturbation (*e.g.*, removal of the acetyl introduces an almost systematic blue-shift) they did not obtain a simple relationship between individual conformational parameters and the calculated site-energy and instead implied the presence of, *"a nonadditive relationship between the influences of framework distortion and acetyl orientations"* [26]. Modern work has culminated in 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 modes of influence by the protein are still contentious [7,8,25,27].

Thus, there is still 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 heterogeneity suggesting that there is still cause to suspect a functional role for the simultaneously observed variation in macrocycle conformations that have been reported since some of the earliest crystal structures [10,26]. While it would be interesting to uncover exactly how the conformational properties of BChls contribute to its properties it is no trivial matter to separate individual contributions such as local functional group conformation from the effect of the overall macrocycle conformation, suggesting that the standard approach of sequential truncation is not necessarily adequate. Furthermore, sequential truncation is inherently flawed, as porphyrin macrocycle distortion and substituents influence each other [33]. Here we attempted to address this issue by employing a more sophisticated experimental design to model precisely how the individual BChl site-energies are influenced by protein-induced conformational variation.

In addition, previously we emphasized the necessity of critically assessing all relevant structural data when one intends to investigate the effects of conformational differences between protein-bound tetrapyrrole cofactors since these features are often close to the positional error limits of protein crystallography [15,16]. Thus, we first consider the extent of BChl macrocycle conformational variation observed in these data *via* a comparative analysis of the normal-coordinate structural decompositions (NSD) [34] of the pigment conformations afforded by each structure. Although this aspect of the study remains inconclusive owing to the high-degree of inter-structure conformational variation, its potential to bias our study is considered and mitigated by applying the approach to multiple crystal structures.

In summary, we report the development and application of a computational approach to delineate the effects of the conformations of the individual molecular components and substituents of BChl pigments upon their unperturbed excitation energies to assess the role of conformational control in modulating this EET parameter in the FMO protein. The paper is organized as follows: following the experimental details (§2) we begin by describing macrocycle conformational variation amongst the FMO BChls via normal-coordinate structural decomposition (NSD) of the BChls from available crystal structures of the complex (§3.1). Next, we present empirical evidence that demonstrates the presence of a regulatory effect of the BChl substituents upon the influence of macrocycle distortion (§3.2.1) and similarly an effect of macrocyclic nonplanarity upon the influence of the C3-acetyl dihedral angle (§3.2.2). We then discuss the theoretical justification of our computational approach (§3.3.1) and describe the results in terms of: the conformational factors most influencing the site-energies calculated for all pigments over all crystal structures (§3.3.2); general trends in the relative importance of the conformational influence of the different components of BChl (§3.3.3); how inter-pigment variation of the most important conformational factors identified may contribute to the distribution of published site-energies for the complexes from three species (§3.3.4), with detailed consideration of a particularly biologically relevant site-energy difference provided in SI §3.3.4A; and lastly, how the results of the truncation procedure differ from those obtained using our method by a direct comparison of the two approaches (§3.3.5). Finally, we discuss the performance and reliability of our computational design (§4.1) as well as the results obtained from each analysis stage as a whole (§4.2) and conclude by highlighting our most general results (§5).

2. Methods

2.1 Factorial Relaxation (Fragment-based Partial Optimizations)

It is useful to overview our experiment in juxtaposition to Fajer and coworkers' [26] wherein the protein-induced structural perturbations of the FMO BChls on their site-energies were modelled by progressively truncating the BChl structures. We argue that there are two caveats inherent to this approach: 1) Each truncation produces a new chemical entity that will behave differently under conformational perturbation (*e.g.*, the rotational energy profile of the C3-acetyl would be qualitatively different upon truncation of the C2-Me group); and 2) The sequence of truncation influences the outcome of the experiment (*e.g.*, if we sequentially truncate the C7-methyl and C8-ethyl then ΔE of the *first* removal would include the additional strain energy of their interaction regardless of which we choose to remove first). We address (2) above by first assigning binary variables X_i to represent whether or not specific substituents (*e.g.*, the C3-acetyl or phytyl) of BChl are truncated or not. This allows enumeration of all 2^n possible structures that may be defined in this way where n is the number of substituents included. This highlights the fact that a sequence of these structures that leads from the fully-substituted model BChl to unsubstituted porphyrin constitutes only a subset of all possible structures. This design is known as an *unbalanced fractional factorial design* and suffers from a convolution of the effects of each variable (*i.e.*, substituent) with their mutual interactions (*e.g.*, steric interactions). In contrast performing the study using the full set of structures in a scheme known as a *full factorial design* allows deconvolution of the main effects of adding or truncating a substituent as well as all orders of their mutual interactions (*i.e.*, where the presence of one substituent influences the effect of the presence of another) *via* statistical analysis [34]. The second modification to Fajer and coworkers' [26] protocol is to replace truncation with conformational relaxation as this allows the conformational influence to be assessed directly in the context of the actual molecular structure. This is achieved by reassigning our variables X_i to represent whether or not a particular molecular component is relaxed using PM6 optimization or frozen in the resolved protein bound conformation. In this case, we probe the influence of the individual and interacting conformational perturbations brought on by the protein environment as opposed to the perturbation induced by their presence or absence. These two modifications result in the procedure that we will refer to as *factorial relaxation* (Fig. S1).

2.2.1 Structure Selection for Factorial Experiment

The first step in our implementation was to define groups of atoms into meaningful molecular fragments. The choice of partitioning scheme is crucial as it dictates the structural resolution of the experiment as well as its computational intensity and interpretative complexity. The extreme possibilities illustrate these competing factors more clearly: if only one fragment is defined (*i.e.*, the entire BChl) then two calculations per BChl are required (2^1), namely the excited state energies of the crystal conformation and that of the fully relaxed structure, yielding a single parameter per BChl describing the effect of the overall protein induced conformation on E_{ES}. On the other hand, if each individual BChl atom was considered independently then $2^{66} - 1$ geometry optimizations would ensue and the analysis would return 66 parameters describing the main effects of the conformational perturbation of each atom, 66^2 pairwise interaction terms, 66^3 3-way terms and so on, which clearly would present interpretative difficulties. Given that the number of partial geometry optimizations and single-point calculations, together termed an experimental unit (EU), required scales in this way (2^n, n = number of fragments) only semi-empirical quantum calculations are feasible for a highly partitioned molecular fragment scheme (*e.g.,* usually < 5 min for PM6 optimization and approx. 1 s per ZIndo/S single-point on an 8 core 2.3 GHz HPC node).

The partitioning of BChl a used to construct the partial optimization series is presented below (**Error! Reference source not found.** 2). These groups were chosen as a balance between efficiency and resolution that possesses the ability to differentiate the major conformational

aspects of interest, specifically: the macrocycle framework, C3-acetyl substituent, 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 a more inclusive model that was also tried where different conformations were obtained in the fully optimized structures depending on the starting geometry afforded by the crystal structure; this confounded comparative results and so these substituents were excluded to avoid this complication. Additionally, this is a computationally expedient model and was therefore useful to perform a comparative analysis of the results obtained by using the geometries found in different crystal structures with only 32 EUs per BChl (n = 5).
<Fig. 2>

2.2.2 Computational Implementation

The PM6 method [36] was used to provide relaxed geometries for subsequent ZIndo/S calculation [3,37,38]. It's predecessor the NDDO PM5 model has been shown to provide excellent geometries for Chls and BChls that are comparable to those obtained using Hartree Fock and DFT methods [39–40]. 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 [36]. ZIndo/S excitation energies obtained using PM5 geometries of Chls and BChls have also been extensively benchmarked and shown to afford accurate excitation energies and able to predict the relative energies of the main absorption bands of compounds in this class [39–45]; an active space of 15 HOMOs and 15 LUMOs was employed. All calculations were performed with the Gaussian 09 [46] software package on the Lonsdale cluster provided by the Trinity Centre for High-performance Computing. The PM6 method was employed with the OPT=ReadFreeze keyword to specify those atoms to be optimized and those to remain frozen in the crystal geometry; in all partial optimizations hydrogen atoms were relaxed. For all BChl models, hydrogen atoms were added to the BChl coordinates found in the PDB file, the representation converted to a Z-matrix and the geometry optimization performed using PM6. ZIndo/S(15,15) calculations were then performed at each optimized geometry. An R [47] script was used to generate all Gaussian 09 input files with the 'ReadFreeze' atom list set corresponding to the permutation in the partial optimization series. All calculations were initiated with individual analytic force constants using OPT=CalcFC as preliminary benchmarking indicated this approach expedited geometry convergence. Any experimental units (EUs) that were not successfully optimized using this method were restarted from the last available geometry and new analytic force constants were calculated to initiate the optimization using the OPT=CalcFC option. EUs that were not converged at this stage, were started from the initial geometry with analytic force constants calculated at each stage in the optimization using the OPT=CalcAll keyword. As a last resort, particularly when optimizations were oscillating between geometries, the step size was decreased using the OPT=MaxStep option; when necessary, this afforded a converged optimization. The optimizations were performed using the 'tight' convergence criteria in Gaussian 09 [46]. Additionally, a few of the BChl models failed to achieve SCF convergence using the default SCF algorithm and were successfully converged by allowing quadratic convergence steps when necessary using the SCF=XQC option (Table S1).

2.2.3 Determination of Molecular Contributions to Protein Induced Site-energy Shifts

Once all EUs were completed for a given BChl, the ZIndo/S first excitation energies (E_{ES1}) were compiled together with their associated group partial optimization permutation vectors in effect coding (see Table S2 for an example). In detail, the permutation vectors are ordered n-tuples where each entry corresponds to one of the atom group definitions (*i.e.*, element 1 represents the optimization state of the Mg) and was set to '1', if the atoms in the

corresponding group were optimized or, '−1' if they were frozen in the crystal geometry (NB, the permutation vectors were stored as the *factor* type in R, not *numeric*, which ensures the correct linear regression algorithm was called by the *lm* function) (Eqn. 1). Subsequently linear regressions were performed with the *lm* function in the *base* package of the R statistical environment [47] using model equations constructed from sets of the 1 to *n* variables and variable interaction terms, where the latter are represented as products of the elements in the permutation vectors (Eqn. 2). To illustrate, the first order terms for Mg and the macrocycle were represented by the first and second entries in the associated permutation vector, respectively, whilst the second order interaction variable for these groups was defined as the product of these elements. As discussed in more detail later (§3.3.1), the β_x coefficients for the first order terms represent the 'main effect' of optimization from the observed geometry which is the negative of the protein-induced conformational perturbation of the individual molecular component *x*, whilst the interaction term coefficients $\beta_{x:y}$ represent the modulatory effect that one component has on another. The results in §3.3.2-5 were obtained using the first-order model (Eqn. 1) whilst the application and interpretation of higher order models of the form in Eqn. 2 is provided in SI §3.3.4A.

Equation 1. The model equation for calculating the main effects (β_i) of the BChl molecular components.

$$E_{Qy} = \beta_{Mg}X_{Mg} + \beta_{Mac}X_{Mac} + \beta_{Acetyl}X_{Acetyl} + \beta_{MeCP}X_{MeCP} + \beta_{Alkyls}X_{Alkyls}\,; X = 1\ or - 1$$

Equation 2. Part of the model equation that incorporates a 2nd-order interaction between the Mg and macrocycle BChl a components.

$$E_{Qy} = \beta_{Mg}X_{Mg} + \beta_{Mac}X_{Mac} + \beta_{Mg:Mac}X_{Mg}X_{Mac} + \cdots; X = 1\ or - 1$$

2.3 C3-Acetyl Rotational Energy Profiles

The C3-acetyl Qy rotational energy profiles (REPs) were obtained using the Gaussian 09 [46] software on a Windows PC. The 'SCAN' keyword was used to affect geometry optimizations at O-C3^1-C3-C2 dihedral angles over the range -180° to 180° in 5° increments. The atoms to be optimized were specified using the OPT=ReadFreeze option. Subsequently, excitation energies were calculated using ZIndo/S(15,15). In addition to the rigid acetyl dihedral Qy REP, three relaxed Qy REPs were calculated corresponding to: 1) Methyl hydrogen dihedral relaxation, 2) All internal parameters of the acetyl group plus the C3-C3^1 bond-length and 3) As 2, with the addition that all H-atoms were optimized at each scan point.

For simple numerical comparison, the Qy REPS were fit to a 3 parameter function of the dihedral angled (Θ, radians) of the form $\lambda = a\cos 2\theta + b\,\theta^2 + c\ ;\ -\pi < \theta < \pi$ in R using the *lm* function in *base*. Where *c* indicates the relative position of the profiles of the Qy REPs from different BChls, the magnitude of *b* is an indicator of the peak difference between the s-*cis* and s-*trans* conformations and *a* is related to the amplitude of the oscillatory modulation of the Qy band as a function of Θ (see Fig. S3 for graphical illustration).

2.4 Normal-coordinate Structural Decomposition

The normal-coordinate structural decomposition (NSD) method developed by Shelnutt and coworkers provides a quantitative description of the macrocycle conformation in terms of displacements along the vibrational modes of the macrocycle [35]. For most BChls the full set of 66 normal-modes is not required for an accurate representation of the structure as the largest displacements occur along the lowest-energy vibrations. The normal-modes that comprise a suitable minimum basis set representation for the out-of-plane (OOP) distortions for tetrapyrroles, including BChls, relate to the commonly observed *sad*dled (B_{2u}), *ruf*fled (B_{1u}) and *dom*ed (A_{2u}) conformations of tetrapyrroles as well as the less well-known

*pro*pellered (A_{1u}) and degenerate *wav*e modes (E_g(x) and E_g(y)) (Fig. 3). The in-plane (IP) modes that complete the min. basis 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}) vibrational modes. When employing the reduced basis set description it is important to be aware of the difference between the structure described by the decomposition and the actual structure. In the NSD procedure this is measured separately for the IP and OOP conformations by d_{oop} and d_{ip}, respectively, both of which represent the mean absolute deviation of the NSD min. basis model to the observed structure. Full details on our use of NSD for the qualitative, quantitative and statistical analysis of tetrapyrrole cofactor conformations in pigment-protein complexes have been given before [12,15,16].
<Fig. 3>

2.5 Statistical Analysis

2.5.1 Clustering

Cluster analysis is a method used to reveal the relationships between observations in a multivariate dataset (*e.g.* the NSD coefficients) [48,49]. This is achieved by iteratively grouping observations such as to optimize an objective function resulting in clusters that have similar characteristics within the groups and distinctions between. We employed agglomerative hierarchical clustering (AHC) in §3.3.2 to explore the relationships of the BChls in terms of their conformations *via* the NSD variables and, separately, their acetyl dihedral Qy REPs, which are essentially ordered vectors. AHC groups similar observations in a hierarchical manner *via* a stepwise process. The two most similar observations are paired to form a 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 if it is not known previously how many clusters ought to be found. This can be informed by consideration of the *dendrogram* which illustrates the cluster hierarchy by displaying the measure of the (dis)similarity of the clusters the value of the objective function upon the formation of each new cluster. Here the Euclidean distance was used to obtain the dissimilarity matrix and Wards' method was used for the agglomeration. The analysis was implemented with the *hclust* function from the R *stats* package.

2.5.2 Linear Regression

Throughout this study we used linear regression to assess the relationships between various variables. Unless otherwise specified these models were constructed using the *lm* function in the R *stats* package and where shown the model confidence is illustrated by the shaded area surrounding the fit calculated from the standard error of the prediction obtained using the *predict.lm* function (these functions are called by *geom_smooth* from the *ggplot2* package used to generate the figures). Where specified robust linear modelling was employed to ameliorate the influence of outliers with the *rlm* function from the *MASS* package in R, which uses an iterated re-weighted least squares strategy and was run with the default M-estimation method.

2.5.3 Bootstrap Significance Testing

In §3.3.3 we discuss correlations of particular sets of BChl pigments' constituent conformational effects with reference site-energies reported for the appropriate species. We are particularly interested in 1:1 relationships between these parameters but as they are not observed with the same intercept for all 7 BChls we must calculate the probability of observing a given linear relationship between a subset k points drawn from sample on N (= 7 Chls) by random chance. We achieved this by generating randomly sampled sets of

uncorrelated *x, y* coordinates and counting how many times we observed any *k* set of points meeting the observed criteria to be tested. This was done in reference to a suitable tolerance ε for the gradient (*m*) such that (-)1 – ε < *m* < (-1) + ε and an upper bound for the r.m.s.d. of the residuals. The largest p-value obtained when generating *x*, *y* using either a uniform or normal sampling distribution is quoted where relevant in text. Full details of the algorithm employed and the parameters used for each test are given in SI §3.3.4B.

3. Results and Discussion

3.1 Macrocycle Conformational Flexibility in the FMO Complex

In this section we begin with an overview of the available structural data for the FMO complex and compare in general terms systematic differences and/or similarities amongst the different representations of each of the BChls afforded by the different structures. Subsequently, we provide a detailed description of the BChl conformations induced by the protein at each site.

3.1.1 Summary of FMO Crystal Structures

In all there are 12 crystallographic models of FMO proteins from three distinct species deposited in the Protein Data Bank (PDB), many with high-resolution (1.3 – 2.2 Å; Tables S3, S4 and S5) in distinction to other photosynthetic proteins [23,24,31]. 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 [19], even the earliest structure (PDB ID: 1BCL) indicated a little 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) [20]. Substantial variations amongst the BChl conformations were first reported some years later when the structure was refined to a resolution of 1.9 Å (PDB ID: 3BCL) [20]. 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 were also reported, as were specific ligand and substituent parameters [21]. The NSD analyses of the BChls from this structure show clearly that there is substantial variation amongst the pigments (Fig. S8). However, there is no obvious distinction between class I and II pigments, as defined by Tronrud *et al*. [21], which is in agreement with the later conclusions of Fajer and coworkers [26].

The next development came after the chemically determined amino acid sequence [50] was incorporated into a new structural model (PDB ID: 4BCL) [22]. Although the crystallographically inferred sequence used to construct the earlier models proved 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 [22]. Consideration of the NSD results from this structure (Fig. S9) 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 characterized crystallographically was from *Chlorobaculum tepidum* and the structure was solved using the *Ptc. aestuarii* structure (PDB ID: 4BCL) with molecular replacement to a resolution of 2.2 Å (PDB ID: 1KSA) [29]. Alignment of the

BChls showed that their overall arrangement was similar while 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 observed to be less coplanar than in *Ptc. aestuarii*.
Tronrud *et al.* reported a very high-resolution (1.3 Å) structure of the *Ptc. aestuarii* FMO (3EOJ) and a re-refinement of the problematic structure 1M50 of *Cbl. tepidum* [23]. 3EOJ is a particularly interesting structure determination from our perspective because BChls 1-7 were not refined against an external library as instead the bond lengths and angles were restrained to their group averages. In their comparison of the structures from each species the authors attributed their different spectral classifications to alternate binding modes of the eighth BChl. The type of nonplanarity possessed by the 3EOJ BChls are similar to the predeceasing 4BCL structure however, the in-plane NSD coefficients illustrate a systematically greater A_{1g} displacement indicating a symmetric '*bre*athing' expansion of the core together with a smaller *B1g* distortion, which indicates decreased stretching along the N-N axis. In our experience of the conformations of photosynthetic cofactors and compared to small-molecule analogues the 3EOJ conformations are more likely to be correct. As this structure is probably the best currently available reflection of the BChl conformational diversity *in situ* a detailed discussion of the 3EOJ structure is provided in the following section. The 3ENI structure that was also reported displays a marked deviation from the original solution to the diffraction data seen in 1M50 and the overall OOP displacements are more akin to 3EOJ albeit the nonplanar distortions are typically greater in 3ENI. We did not yet comment on structure 1M50 but it suffices to say that the BChls display quite distinct NSD profiles marked by a large degree of A_{2u} '*dom*ing' distortions and are probably not good representations of the actual BChl conformations; these conformations highly diverge from its own predecessor 1KSA, which itself displays the overall characteristic NSD profile of the other structures barring a uniquely large B_{2u} displacement of BChl 4.
The third species' FMO to have its structure determined was *Pel. phaeum* and is represented in the PDB by structures 3OEG and 3VDI. Both structures display similar out-of-plane distortions that are similar in profile to the structures previously discussed. However, the earlier of the two, 3OEG, possess far smaller in-plane deformation coefficients across the board, particularly with respect to the A_{1g} mode. Structure 3VDI was a re-refinement of the diffraction data used to build 3OEG and the BChl conformations are more consistent with the NSD profiles emerging as a general consensus.

3.1.2 NSD Analysis of Pigment Conformations
Since it emerged during the forgoing discussion that the BChls possess site-specific conformations that are on the whole well-conserved in different crystal structures, even from different species, we will describe the pigment conformations in reference to a single representative structure. Additionally, in terms of resolution and BChl refinement procedures employed (see above) it seems that structure 3EOJ [23] from *Ptc. aestuarii* is a fitting choice for this purpose. Inspection of the NSD min. basis deformations (Fig. 4) shows that the individual pigments possess a variety of conformations ranging from the relatively planar BChl 1 to a conformation dominated by a single 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 NSDs behave as expected with respect to the magnitude of nonplanar distortion since the more distorted BChls 3, -4, -6 and -7 are significantly contracted (*i.e.*, smaller A_{1g}) compared to the other more planar BChls, which is a known property of nonplanar porphyrins [51]. Additionally, there are significant displacements and inter-pigment variations along the next-to-lowest energy normal-coordinates 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 (not shown).
<Fig. 4>
Conformation of the EET exit pigment: The pigment that has been assigned the lowest site-energy (BChl 3) exhibits the joint-largest degree of ruffling (*cf.* BChl 4) of all the chromophores. However, in contrast to BChl 4, the extent of the ruffling (B_{1u}) 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 (A_{2u}). 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-stretching (B_{2g} and B_{1g}, respectively), indicating a less skewed in-plane geometry. With respect to the next-to-lowest out-of-plane distortions of each symmetry type this pigment has the largest $B_{1u}(2)$ and $A_{1u}(2)$ distortions (not shown but see Fig. 6), of which the former has been directly linked to redshifted Q-bands [51] and the latter requires the largest deformation energy of the next-to-lowest energy modes [35].
Conformation of the eighth BChl: Recently it was suggested that the BChl 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, doming and both orientations of waving (approximately 0.2 Å of each). The overall relative planarity of BChl 8 is indicated by the fact that it exhibits the largest A_{1g} mode of any of the chromophores but it also 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 (not shown).
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, 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 [25]. Like the eighth pigment, the min. 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: Of the remaining chromophores, BChl 5 appears to be the most planar overall as emphasized by the presence of only moderate saddling and the most expanded macrocycle, however, there 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 pigment's conformation is also complemented by considerable $B_{1u}(2)$.
Alternate conformers of BChls 6 and 7: The observation of alternate conformers of BChls 6 and 7 in structure 3EOJ [23] warrants special mention since the differences amongst their conformations described by NSD provides an indication of the sensitivity of this experimental structure to small changes in pigment conformation, since the variability in each deformation has been resolved at a single site. For each of these pigments, two separate portions of the BChl molecule have been assigned independent positional occupancies giving rise to four potential conformers of each BChl. Specifically, BChl 6 is characterized by two possible conformations of Ring B correlated with the motion of the ethyl substituent and alternate conformations of Rings D and E that are associated with the position of the methyl ester whilst BChl 7 was also modelled as having two Ring B conformers but the other flexible component is the acetyl substituent (Fig. 5). The influence on the overall NSDs of these pigments (Fig. S4) shows that conformational differences of less than ~0.1 Å are experimentally detected in this structure, lending credibility to the significance of the much larger variations of macrocycle deformations observed at different sites that were the focus of the preceding discussion.
<Fig. 5>

3.1.3 Correlation of NSDs and Literature Site-energies
We noted previously that the BChl assigned to have the lowest site-energy (BChl 3) also possesses the greatest $B_{1u}(2)$ displacement of all the FMO BChls as resolved in 3EOJ. To explore the significance of this occurrence we determined whether any NSD/ site-energy correlations were present amongst all of the BChls. The most striking observation was the presence of a very strong correlation between the $B_{1u}(2)$ normal-deformations of the BChls and the fitted site-energies from Adolphs *et al*. [8] (Fig. 6). The potential significance of this result is substantiated by the fact that this mode has been shown to be one of the primary causative agents of the redshift found in *ruf* distorted porphyrins [51]. 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 redshift is observed with increasing nonplanarity with the minimum of the curve at approximately zero distortion (*i.e.*, the coefficient for the linear term is not statistically significant). However, whilst this is a noteworthy observation we are not at this time ascribing the $B_{1u}(2)$ displacement a causal role in the distribution of BChl site-energies and highlight this feature simply as an intriguing experimental observation that does contribute to the plausibility of a significant conformational effect; we will return to this observation in the analysis of the results from the factorial relaxation computations.
<Fig. 6>

3.2 Evidence for the Interaction of Core and Periphery Conformations
3.2.1 Revisiting Gudowska-Nowak et al.
Here we explore the data behind the apparent nonadditivity of substituent and macrocycle conformational effects first described by Fajer and co-workers [26] in the context of the influence of variable BChl nonplanarity on their results. The correlations between the site-energies calculated by Fajer and co-workers from the seven BChl pigments found in the 3BCL structure with a selection of the corresponding NSD deformations of the BChls with the corresponding NSDs of their models indicate a complex interaction between the structural model employed and the apparent effects of particular BChl deformations (Table 1). For example, the correlation of the *bre* mode (A_{1g}) in the most complete structural models is such that a redshifted excitation is associated with an expansion of the core, but the correlation

coefficient 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 (B_{1u}) monotonically increases as axial ligands are removed and the periphery is truncated; both of these modes display the expected negative correlations with the site-energies (in cm^{-1}). The behavior of the site-energy with respect to the parameters representing the total in- and out-of-plane distortions (D_{ip} and D_{oop}, respectively) are also revealing where the correlation of D_{oop} is greatest in the most truncated models with the reverse effect observed for D_{ip} (Table 1). This latter observation may be interpreted to indicate that the effect of nonplanarity is subdued by the presence of substituents whereas the effect of in-plane deformation is enhanced, although it is important to note that in-plane and out-of-pane distortions are correlated themselves and that it has been considered that in-plane deformation is induced by the protein via out-of-plane distortion [51]. Taken together, these results show that the conformational effect on the excitation energy is modulated by the presence of the substituents suggesting that truncated models are not appropriate to assess the conformational influence upon this property.
<Table 1>

3.2.2 C3-Acetyl REPs

To explore the interaction between the physicochemical consequences of macrocycle distortion and the substituents from a different perspective we calculated the acetyl dihedral Qy rotational energy profiles (REPs) for each of the BChls from structure 3BCL (Fig. 7). This was done to assess whether or not different macrocycle conformations altered the influence of the acetyl dihedral on the site-energy as well as to confirm that differences in acetyl-macrocycle coplanarity did not account for the full range of conformationally perturbed site-energy shifts reported originally by Fajer and co-workers [26]. Multiple Qy REPs were obtained in a series of increasing relaxation of the acetyl moiety in case the internal geometry of the acetyl was altered as a result of different coplanarities with the macrocycle. The extent of relaxation was increased gradually so that if coalescence of any Qy REP did occur it could be attributed to a specific set of parameters as described in §2.3.
<Fig. 7>

As most of the Qy REPs from the different BChls remain vertically separated at all levels of relaxation it is concluded that differences in the C3-acetyl dihedral angles are not entirely responsible for the conformationally induced site-energy shifts provided by these structural models. However, the Q_y REPs of BChls 4 and 7 were coalescent indicating that the acetyl dihedral was the sole feature leading to their intrinsic difference. A more complex situation was observed for BChl 2 where its Qy REP overlapped with BChl 5 in the rigid rotation scans or when only acetyl H-atoms were relaxed but when further relaxation was allowed the BChl 2 Qy REP shifted into coalescence with BChl 3. This is a particularly interesting occurrence as it indicates the subtle interplay that can occur between conformational parameters, in this case even within the same macrocycle substituent. However, there was still the possibility it was not the macrocycle conformation but the geometric properties of one or more other peripheral substituents included in the model structures that were the origin of the remaining variation. Whilst calculating new Qy REPs with additionally increased degrees of allowed relaxation or directly assessing the influence of other conformational coordinates (*e.g.*, the planarity of the $C3^1$-carbonyl) would be useful to explore this further we found that we could connect the BChl macrocycle NSD deformations to the variability in the BChl Qy REPs directly. Initially, we noticed that the cluster hierarchies obtained for the BChls from clustering their Qy REPs or NSDs displayed remarkable similarity (Fig. 8) showing that BChls with similar conformations also have similar Qy REPs. Notably, this

convergent clustering was not observed until at least the first three lowest-energy NSD modes of each symmetry were clustered indicating that these modes contribute to the calculated profile.
<Fig. 8>
It is clear from the clusters' BChl membership that the dominant feature driving the Qy REP clustering is their vertical displacement (Fig. 8). Since this feature can be adequately explained by an additive combination of the macrocycle conformation and the acetyl dihedral we cannot use the similarity of the structural clusters as an assertion that the macrocycle conformation modifies the form of the Qy REP and so we investigated whether the forms of the Qy REPs were correlated with macrocycle deformations. This was achieved by fitting an equation of the form $\lambda = a\cos 2\theta + b\,\theta^2 + c$ to each profile and comparing the parameters to the NSD coefficients (Table S6). We first note that the both the amplitude, *a*, and the peak difference between the *cis* and *trans* acetyl conformations, *b*, vary amongst the BChls in addition to the intercept demonstrating that the effect of acetyl rotation is not wholly independent of other aspects of BChl conformation. Upon direct correlation of NSD deformations with the profile parameters (Table S7) we found that deformations that are symmetric with respect to the macrocycle *C4*-axis (*i.e.*, the modes of '*A*' symmetry) such as the A_{2g} and A_{2u} were the modes most highly correlated with the amplitude and *cis*/*trans* peak difference whilst distortions of lower-symmetry, most notably B_{2u} and $E_g(y)$ were associated with the vertical displacement of the profile. This is an exceptionally interesting result since as well as confirming the presence of a complex relationship between the macrocycle conformation and the C3-acetyl dihedral it shows that this relationship may be understood at the level of individual macrocycle distortion modes.

3.3 Full-factorial Partial Optimizations

The motivation and detailed outline of the computational design employed in this section were described in methods §2.1.

3.3.1 Physical Interpretation of Main Effects and Interactions

In general, the *in vacuo* single-point energy (SPE) of a protein bound BChl, $E(BChl)_{cryst}$, may be decomposed as the sum of contributions from the energy of the unperturbed BChl and the strain energy induced by the protein environment that is encapsulated in the pigment's conformation so that, $E(BChl)_{cryst} = E(BChl) + E(Strain)$. This unperturbed BChl energy is obtained from the SPE of fully relaxed BChl (*i.e.*, optimized *in vacuo*) and so it follows that calculation of the two values $E(BChl)_{cryst}$ and $E(BChl)_{opt}$ allows the total strain energy induced by the protein to be deduced from their difference, $E(Strain) = E(BChl)_{cryst} - E(BChl)_{opt}$. Similarly, extending this argument to the energies obtained from partial optimizations where only parts of the BChl structure are relaxed suggests that the strain energies of individual molecular fragments may also be readily obtained (Eqn. 3). However, because the conformations of the molecular fragments of a BChl may interact with each other the difference method to obtain a fragment's "intrinsic strain energy" requires additional calculations to provide an accurate determination as Eqn. 3 does not account for this possibility. To illustrate the problem, in a Mg-porphyrin model if the centrally coordinated Mg is frozen in the crystal geometry and is above the 4N-plane due to axial ligation, then a partial optimization of the macrocycle experiences an indirect perturbation from the protein environment propagated by the Mg's position (*e.g.*, it may still display *dom* distortion after optimization). This means that the identification of the r.h.s. term in Eqn. 3 as the protein-induced conformational strain energy is not strictly correct. This effect should be included explicitly as is done so in Eqn. 4 and points towards its solution; we must calculate the effect of the position of the Mg atom on the apparent strain energy of the macrocycle. This can be

achieved by two further calculations where the Mg is also optimized alongside the macrocycle resulting in the series shown in Table S8, noting that it is a tautology as to whether we call the Mg-macrocycle interaction the effect of the Mg position on the macrocycle conformation or *vice versa*. This series corresponds to the full-factorial design for a two fragment (*e.g.*, Mg and macrocycle) by two level (*i.e.*, frozen or optimized) experiment. The explicit interpretation of each calculation (Table S8) shows that the individual contributions of the fragments, each '$E(Strain)_{Frag\ X}$', are provided since we have a precisely determined linear system (*i.e.*, here four equations with four unknowns); intuitively it may be observed that *the effect of each fragment is assessed at each state of the other*. These arguments hold too for excited state energies or absorption wavelengths (*e.g.* Qy) calculated with the ZIndo/S method where strain energies are interpreted as the intrinsic site-energy perturbation of the (protein induced) conformation.

Equation 3. Explicit interpretation of the result of a SPE on a partially optimized structure leading to the 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 optimized geometry.

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

Equation 4. 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 optimized geometry.

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

The determination of the overall perturbations associated with individual molecular fragments (*i.e.* $E(Strain)_{Frag\ X}$), now referred to as the *main effects*, is found *via* linear regression of the excitation energies of the partially optimised structures against the geometry state of each molecular fragment (§2.1). In terms of the qualitative physical interpretation of the derived main effect model parameters, β_{Mac}, β_{Acetyl}, etc. re-call they measure the extent to which the Qy energy changes as that component is allowed to relax to its *in vacuo* state. Therefore we interpret the negatives of these parameters to be, explicitly, *the contribution of each molecular component to the overall protein induced conformational perturbation of the pigment unperturbed site-energies*.

The presence of large interaction terms physically described above as $E(Strain)_{Frag\ Y\ on\ X}$, or in terms of the model coefficients $\beta_{Mac:Acetyl}$, signify any kind of physical dependence between the effects of the conformational perturbations of the interacting substituents. These could arise from either conformational constraints (*e.g.*, the potential for Mg induced *dom* distortion described above or steric clashes) or perturbation to the response of the electronic structure feature under perturbation from conformational change. This latter case could be partly responsible for the effect of the macrocycle conformation on the acetyl Qy REPs (§3.2.2) wherein one could envisage the magnitude of the π-extension of the BChl core aromatic system provided by rotation of the acetyl into the ring A plane being modulated by the planarity of ring A with the other components of the aromatic system and hence its conjugation.

3.3.2 Variation across All Crystal Structures

In lieu of determining whether a single crystal structure best represents the actual conformations of the BChl pigments in the FMO protein, or if it is appropriate to average the BChl conformations over multiple structures, we performed our factorial relaxation method using all available structures. The calculated Q_y excitation energies for the crystal structure geometries (E_{cryst}) together with those obtained for the fully optimized structures (E_{opt}) are shown in Fig. 9. In accord with the observed structural variation there are substantial differences amongst the Q_y energies found for the different BChls frozen in the experimental geometry from different structures. The within structure variation is greatest for 3BSD where the calculated energies range over ~180 nm and least for 3ENI, 3EOJ, 3OEG and 3VDI where this range is ~14 – 35 nm (Table 2). The unrealistically large variation of ΔE_{cryst} in structures 3BSD, 3BCL and 4BCL are most likely the result of inaccuracies in the BChl stereochemistries found in these structures. In contrast, E_{opt} is largely homogeneous such that the range over all BChl models is 667 – 692 nm (s.d. = 5 nm). This homogeneity is expected given that relaxation was performed *in vacuo* and serves to validate the comparisons of the substituent effects that are discussed next. Note also, the optimized model energies are substantially blue-shifted compared to those afforded by the crystal geometry suggesting that the general effect of the conformational perturbation is to redshift the Qy absorption. This is an unsurprising result as it arises due to the ground state experiencing a greater stabilization during optimization relative to the excited state, which may be expected in most ground state optimizations.

<Fig. 9>

<Table 2>

The summary statistics for the substituent main effects (β_{Mac}, β_{Acetyl}, …), calculated for each BChl over the whole series of crystal structures are shown in Table 3. The general result of optimizing any single substituent is to blue-shift the calculated Q_y energy relative to the crystal geometries', as indicated by the negative averages found for the main effects. This is in line with the previously noted blueshifts upon full structural relaxation. However, the maximum values of these parameters 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 blueshifted Q_y energy (corresponding to an observed redshift upon optimization, *i.e.*, positive main effects). This latter observation is most applicable to the effects of the Mg position and acetyl conformation, which show the largest positive maximums (NB. we see later that the largest positive coefficient for the Mg effect is likely due to an erroneous structural model). With respect to the potential for the acetyl to induce a blue-shift, this likely results from the protein-environment forcing it out-of-the plane of the macrocycle (Fig. 7) and a similar effect has been suggested regarding the cyclopentanone ring [52], which can also possess a positive main effect (Table 3).

<Table 3>

Strikingly, the optimization 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 (Table 3). Although this is an intriguing result, it does not automatically mean that the influence of the macrocycle conformation has been proven to be the most important as other factors may confound this result. In the first place, batch effects from different crystal structure determinations are prevalent in this comparison (*vide infra*). Another reason for caution is that the number of degrees of freedom in each geometry optimization is proportional to the number of relaxed atoms and there are far more atoms in the macrocycle atom 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 per atom, which whilst still a considerable difference, quite so drastic. Moreover, because all 24 atoms either are, or in the case of saturated beta macrocycle

carbons are bonded to, the sp^2 hybridized framework of the π-system a systematic difference between the PM6 preferred geometries of an extended π-system to that present in the experimental structure could be responsible for this observation. Conversely, the same argument also lends credence to the result since it is reasonable that perturbations to their geometry would have large effects. In any case, it is important to recall what these calculations provide explicitly: 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 of the substituents found for different FMO pigments are of primary interest since such comparison will ameliorate many of the systematic biases due to the computational protocol. However, even in this regard, the inter-pigment variation of the BChl fragment main effects implies that the macrocycle effect is the largest conformational influence (s.d. = 38 nm, Table 3).

With respect to how well the substituent and macrocycle main effects account for the variation in the calculated Q_y energies of the experimental geometries across all BChl models, linear regressions upon only these parameters provided very good agreement with the full range of E_{cryst} (Table 4). Before proceeding, notice that the regression intercepts are close to the average values of the fully optimized structures and that the macrocycle and substituent coefficients tend toward -1 (Table 3). This is expected since the individual main effects are multiplied by their respective coefficients then summed together with the intercept (see §3.3.1). Thus, if the effect is a strong and consistent determinant of the total variation of E_{cryst} then a coefficient close to -1 is expected. Effect coefficients that deviate from negative unity may be observed either because of unusual structures where a molecular component behaves qualitatively different upon relaxation (outliers, *vide infra*) or because the conformational effect of the substituent is strongly moderated by interactions with other BChl constituents. To account for outliers, we supplemented the ordinary least-squares (OLS) regressions with a robust linear modelling (RLM) technique (§2.5.2).

With OLS, the high R^2 value (0.97) indicates that main effects alone account for ~97% of the total variation, implying that across all structural models substituent interactions are relatively minor. In this regression too, the β_{Mac} regression coefficient is closest to -1 suggesting that the majority of the large macrocycle conformational effect is exerted independently of the conformational state of the other BChl components in this range, although it is also important to realize that a considerable factor here is it's size compared to the other main effects. The next most variable conformational effects amongst the pigments were the MeCP components followed by the acetyl substituents (Table 3) and these too appear to be relatively strong main effects (Table 4), with the acetyl's greater closeness to -1 perhaps implying less influence from interactions than the MeCP. Structurally, this may be explained by the constraint on the MeCP due to its relatively rigid ring structure. Notably, the pigment with the minimum main effect calculated for the MeCP group is an outlying structure with respect to this parameter, as can be seen by comparing the β_{MeCP} range [minimum,1^{st} quartile] and range [3^{rd} quartile, maximum] (Table 3), but this does not appear to have a very large effect on its behavior over all structures (compare regressions in Table 4 and *cf.* β_{Mg}, an effect that is heavily influenced by an outlier; *vide infra*). Also interesting is that the aliphatic substituents' effect (β_{Alk}) varies almost as much as the acetyls' (Table 3), indicating that non-conjugated substituents may also have considerable effect on the site-energies, in line with the recent conclusion of Matysik and coworkers [53].

The most drastic difference between the OLS and RLM regressions (Table 4) is observed for the regression coefficient for the Mg main effect (β_{Mg}) that is calculated to alter E_{cryst} by +0.53 in the first regression provided; the drastic departure from the expected value of -1 noted above indicating a unique situation. However, the RLM regression coefficient of -0.42

indicates that this is driven by one or more outliers and inspection identified BChl 8 from structure 3OEG to have an abnormally large positive β_{Mg} (15.5 nm). Removal of this structure from the analysis significantly improves the agreement between the OLS and RLM models (Table 4). This result is explained by the fact that it is most likely that this BChl 8 model is entirely erroneous as the density at this site was later found to originate from a disordered PEG molecule. The remaining departure from -1 is possibly due to a strong influence of the macrocycle conformation on the observed effect of the Mg since the orientation of the pyrrole N atoms (*e.g.*, degree of A_{2u} *dom* distortion) and core size (A_{1g} *bre*) will influence the location of the central Mg.

<Table 4>

Before moving on to a direct analysis of within structure inter-pigment variation, it is worth highlighting that variation in β_{Mac} alone accounts for a large proportion of the total variation over all BChls and all structures (R2 = 0.86; Figs. 10 and S15). A considerable proportion of this variance can be observed to arise due to systematic differences between structure determinations (Fig. 10). Yet, a significant degree of within structure variation is also accounted for by the different macrocycle conformations. This is most obvious for the high-variance structures such as 3BSD and 3OEG, but on closer inspection is true too for structures with a narrower range of β_{Mac} 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. Note that on a linear energy scale, when E_{cryst} is regressed against β_{Mac} (Fig. S14), R^2 drops significantly to 0.67 although the slope remains close to 1 (0.92). These values also drop significantly when regression is performed over a narrower range that excludes outlying structures (Figures 10 and S14) illustrating that on scales more comparable to the substituent effects on E_{cryst}, that typically corresponds to the level of E_{cryst} inter-pigment variation calculated within a single crystal structure, the other substituent conformations become more important.

<Fig. 10>

3.3.3 General Trends in the Effects of BChl Conformation

We will now consider some observations regarding the relative importance, in terms of effect size, of the conformations of individual components of the BChl molecule and their interactions with reference to the results from the highest resolution structure, PDB ID: 3EOJ. The main effects obtained from the first order models (Table 5) again illustrate that the macrocycle effect dwarves that of any other constituent of the BChl molecules and that it provides the dominant contribution to the inter-site variation [range: -173, -148 nm]. In all cases too, the next largest effect comes from the acetyl, which appears to fall into two qualitative groups given by the ranges [-8, -5 nm] for BChls 3, 5, 7(AC) and 7(AD) and [-22, -18 nm] for all other structures, with the greatest inter-site difference (17 nm) not too much smaller than that of the macrocycle effects (25 nm). The next largest effects tend to be those of the alkyl substituents. The majority of the effects of the MeCP group are insignificant, with respect to the standard errors of the model coefficients, excluding the 'C' and 'B' alternate conformers of BChls 6 and 7, respectively. This variation in β_{MeCP} amongst alternate conformers provides proof of the geometric relevance of these values when one considers that it is only the conformation of the MeCP fragment that is different between the 'C' and 'D' conformers of BChl 6 and the acetyl dihedral between conformers 'A' and 'B' of BChl 7 (Fig. 5).

<Table 5>

These results show that the macrocycle and acetyl conformations provide the greatest overall individual contributions to the total conformational effect; a reasonable conclusion given the importance of skeletal distortion and the π-extending properties of the C3-acetyl on the BChl

conjugation. However, the situation is somewhat altered when constituent interactions are modelled as in general the magnitude of β_{Acetyl} decreases significantly as more interaction terms are included (Table 5). For instance, in BChl 4 this parameter drops to ~60% of its original value once 3rd order interactions are included and in BChl 2 falls to ~70% and ~45% its original value when 2nd and 3rd order interactions are included, respectively. Although in some cases larger reductions of the macrocycle conformational effects are observed with increasing model complexity, proportionally these changes are small since the macrocycle effect is so large to begin with. This may be interpreted as an indication that the conformation of the C3-acetyl is in some way interacting with another geometric parameter of the BChl, as was indicated by the REPs in §3.2.2, since the reduction in its main conformational effect size must be accounted for by its interactions.

Analysis of the effect coefficients from the 2nd and 3rd order models confirms that there is a significant interaction between the acetyl and macrocycle conformations, so that the skeletal conformation influences the effect of the acetyl's and *vice-versa*. Furthermore, in a number of cases this interaction is almost as large as β_{Acetyl} itself (Table S11 and S12) implying that the notion of the acetyl conformation influencing the Qy energy in isolation from the macrocycle's conformation may be an oversimplification. Even so, the range of values for the acetyl main effects continues to span substantially after allowing for conformational interactions and remains divided into two qualitative groups as was the case for the 1st order models (2nd order: [-7.0, -4.7 nm] and [-22.9, -12.6 nm]; 3rd order: [-4.2 nm] and [-18.8, -10.2 nm]) upholding that its conformation does exert influence upon the site-energy distribution. Additionally, in the majority of cases the inter-site difference of $\beta_{Mac:Acetyl}$ is not statistically significant with respect to the error in their determination. Thus, these calculations suggest that the interaction is an intrinsic property of BChl and is not itself modified by the protein to affect the desired distribution of site-energies.

3.3.4 Biological Relevance: Comparison of the Conformational Perturbations to Experimental Site-energies

As there is considerable variation of each pigments' macrocycle conformation across the various crystal structures in order to meaningfully assess how the intrinsic site-energy distributions are influenced by conformational effects we must consider the within-structure variation of the calculated perturbations. By restricting ourselves to a relative comparison we ought to mitigate the influences of the specific experimental protocols that led to each crystal structure determinations (*e.g.*, restraints) as within each structure these should be relatively constant so that the observed conformational effects may be assumed to be supported by the diffraction data. We can also compare the conformational site-energy perturbations to the relevant experimentally derived site-energies for each species in order to see how the different pigment conformations contribute to the distribution resulting from the combination of all environmental interactions. The following discussion relates to results obtained from the calculations performed on the four most recent structural determinations (PDB IDs: 3ENI, 3EOJ, 3OEG and 3VDI). For all data see Figs. S15-S18 and Tables S18-S20 for significance tests; the calculation of p-values quoted in the following is described in §2.5.3 and SI see §3.3.4B.

Beginning with the overall intrinsic site-energy differences given by E_{cryst} (*i.e.* the ZIndo/S Qy energy of the the protein bound conformation), the data suggests that some inter-pigment differences are accounted for by structural variation alone. For instance, E_{cryst} in structure 3ENI varies approximately 1:1 with the corresponding *Cbl. tepidum* site-energies [54] of BChl pairs 2/4, 4/5 and 6/7. In contrast, the nearly 30 nm range of experimental site-energies spanned by BChls 1, 2, 3 and 5 cannot be due to the total conformational perturbation as for these BChls E_{cryst} is relatively constant suggesting that other influences are responsible for the

in situ variation. In *Ptc. aestuarii*, represented by structure 3EOJ, other pairwise differences amongst specific pigments' site-energies [55] are consistent with the E_{cryst}, namely BChl pairs 1/2, 4/7 and 5/6. For *Pel. phaeum* we have two high-resolution crystal structures (3OEG and 3VDI) and so are able to get a sense of the variability ZIndo/S applied to the protein bound conformation due to errors in structure determination directly and in this example it is relatively large. Specifically, the BChl intrinsic site-energies derived from structure 3OEG are consistent with the available site-energies [56,57] amongst BChls 3, 4, 5 and 6 and although the results from structure 3VDI are consistent with respect to the BChl 4/5 difference there is little other similarity between the two structures. In conclusion, direct correlation of the intrinsic site-energies with those derived from experiment does not lend great weight to the notion that conformational control is a dominant factor overall. These data suggest that it could play a role in some cases, but these results cannot confirm this conclusion on its own as all such pairwise 1:1 linearities are statistically insignificant when subjected to stringent testing (SI §3.3.4B).

The calculated site-energy perturbations due to the macrocycle component alone obtained *via* our factorial relaxation approach (β_{Mac}) show some interesting global trends with respect to the experimental site-energies (Fig. S16; re-call that the negative of β_{Mac} is the predicted shift due to the macrocycle perturbation so that BChls lying on a 1:-1 diagonal is supportive of the macrocycle exerting a dominant role in their site-energy difference). To begin, in these four structures the exit pigment BChl 3 receives the 2nd or 3rd greatest redshift due to the protein induced conformation of the macrocycle even though the overall conformational perturbation provided by $E(BChl)_{cryst}$ is middling to blue-shifting. Moreover, in structure 3ENI the isolated macrocycle contributions to the site-energies are consistent with the reference values in at least four and perhaps five of the BChls demonstrating pairwise relationships described above Specifically, the perturbations for BChls 2, 4, 5 and 6 are reasonably co-linear with a gradient close to -1 (m = -0.93, p = 0.02; Table S18) whilst the variation amongst the alternate conformers of BChl 7 falls within the region of this relationship. In structure 3EOJ a similar trend is observed where a few pairwise correlations between the total structurally induced site-energy perturbations and the experimental values described above have coalesced into a more general relationship where here the β_{Mac} parameters for BChls 1, 2, 3 and 7 are proportional to the site-energy shifts again (m = -0.90, p = 0.02). This latter result is particularly relevant when considered with the correlation between the $B_{1u}(2)$ macrocycle deformation and the experimental site-energies noted previously (§3.1.3, Fig. 6) as it indicates that even though the sum over all contributions of the protein induced conformational perturbation may not be conductive towards the observed site-energy the macrocycle component is seen to be so when it is properly resolved. BChl 5 is the stark exception in this structure as although this possess the shortest Qy absorption wavelength experimentally, and so accordingly has the smallest $B_{1u}(2)$ distortion (Fig. 6), this pigment was found to exhibit the greatest redshift due to macrocycle perturbation. We do not know why this is the case but point out that this BChl has the second largest contribution from other next-to-lowest energy out-of-plane distortions as illustrated by its high d_{oop} value (Fig. 4); these distortions may be at least partially responsible. In the spirit of our assertion that the individual conformational features of different BChl constituents manifest competing modifications to the site-energies it is possible that the low $B_{1u}(2)$ distortion contributes to the relatively low site-energy of BChl 5 but that a different distortion mode raises it; of course it is also possible that the redshifting macrocycle perturbation is a spurious observation due to either experimental variability or an inadequacy of PM6//ZIndo/S to assess this BChl's particular conformation. Whatever the cause, the macrocycle contribution is offset by the relative conformation of its acetyl resulting in a relative blueshift compared to the other

pigments (Fig. S18) and thus its possession of only the second largest redshift due to overall structural perturbation (Fig. S15).
Unfortunately the protein induced macrocycle site-energy shifts for the *Pel. phaeum* structures remain inconsistent with one another. The results also call into question the reliability of the 3OEG conformations as here β_{Mac} varies 2.5 times more than in any other (~80 nm range for β_{Mac} in 3OEG *vs.* 30-35 nm range in 3EOJ, 3ENI and 3VDI) this feature of 3OEG is perhaps explained by the presence of unusually low A_{1g} NSD deformations (Fig. S6). Notably, this large variation is less prevalent in E_{cryst} described earlier. Focusing then on 3VDI as the representative for *Pel. phaeum* we find BChls 7 and 3 may be dominantly influenced by the macrocycle conformation (m = -1.10, p = 0.51) and although this relationship is also found in 3OEG it is quite likely a chance event, especially given the pitfalls just noted. Even so, further corollary evidence is present that suggests these two pigments could be differentiated under the influence of their macrocycle perturbations as this feature is also present in the *Pr. aestuarii* structure 3EOJ, and given that these are both spectral type II FMOs [23], could be indicative of a conserved feature between the two species.
In order to elaborate on conservation and divergence between these two type II spectral class species we compared the inter-pigment site-energies directly (Fig. S17). It is striking that a substantial correlation exists between the two type II FMOs (Pearson = 0.57) whereas no clear correlation is observed between either 3EOJ and 3ENI (-0.16) or 3ENI and 3VDI (0.02). Moreover, accounting for a systematic shift of ~15 nm, the β_{Mac} values found for BChls 1, 3, 5 and 6 in the *Ptc. aestuarii* and *Pel. phaeum* FMOs are correlated to near unitary proportionality (m = 0.85, p = 0.04) suggesting that the macrocycle conformational perturbation may be conserved between these species. On the other hand BChls 2 and 4 possess quite different β_{Mac}, wherein the *Pel. phaeum* BChls experience a greater redshifting macrocycle perturbation, which is concordant with the redshifted experimental site-energy of BChl 4 in this species (Fig. S17). This could be an example of conformationally driven divergence between these species. For a more detailed discussion of an inter-site difference, exemplified by a comparison of BChls 3 and 1, which are considered to be the EET exit and entry pigments, see SI §3.3.4A.

3.3.5 Comparison of Factorial Relaxation to the Truncation Method – Implications for the Recognition of Conformational Control

Aside from the fact that the method employed in this study provides far more detail than truncation allows, our primary motivation for pursuing this approach was that there were indications that the influence of the skeletal conformation of the macrocycle was affected by the presence of peripheral substituents (§3.3.1) and that the influence of the C3-acetyl conformation was similarly affected by the skeletal conformation (§3.3.2). We now test the former idea explicitly by comparing the macrocycle main effects obtained from the partial optimizations to the estimate obtained by comparing E_{cryst} as calculated for a fully truncated, macrocycle only model of the BChl pigments.
Over the results obtained for all of the BChls from all of the assessed crystal structures there is good agreement overall between the two approaches (Fig. 11 and Fig. S19). However, the regression coefficient of -1.2 shows that although the truncation approach provides a proportionally accurate estimation of the macrocycle deformation Qy energy shift it slightly underestimates it. Additionally, the residuals from this fit are significant [range: -32.0, 44.9 nm; inter-quartile range: -7.8, 8.2 nm] illustrating that when the often subtle inter-pigment differences are of interest, the truncation method is not ideal. A few structures exhibit unusually large or small lowest-energy excitations for the truncated models, with similarly outlying macrocycle main effects and when these are excluded from the OLS fit a further

departure from a unitary relationship is obtained (Fig. 11; yellow line). The gradient (m = -1.6) indicates that in general the effect of the macrocycle conformation calculated *via* our approach is amplified by the presence of the peripheral substituents. This is further illustrated when it is observed that, for the majority of structures, our calculation of the macrocycle effect varies by ~100 nm whereas the estimate from truncation varies over only ~50 nm. Thus, the underestimate of the effect macrocycle distortion by the truncation method could have suppressed recognition of the importance of macrocycle conformation as a modulatory mechanism of site-energy distributions.

Considering the relationship between the two methods for a single structure, there can be substantial differences between the two approaches in the context of comparing the estimated conformational effects to actual site-energy distributions. For instance, BChls 1 and 2 from structure 3EOJ possess virtually identical lowest-energy excitations from the truncation model whilst there is a substantial difference between the their directly calculate macrocycle effects and conversely, 3 of the 4 alternate conformers of BChl 6 have the β_{Mac} but different energies from the truncation model (Fig. S20).

<Fig. 11>

4. Discussion

4.1 Method Reliability

There are a few questions regarding the validity of our method that we will now address. Firstly, although PM6//ZIndo/S is exceptionally well benchmarked with respect to predicting the spectral differences amongst chemically different chlorophylls and compares well with high-level approaches including DFT//TD-DFT [41-45] (see also §2.2.2), there are fewer reports of how it performs with respect to predicting energetic perturbations arising from conformational changes, especially in the diversity of conformations that have been examined here. Fortunately, there are numerous reports of ZIndo/S performing well with both experimental and molecular mechanics predicted structures of highly distorted porphyrins, [51] which goes some way to validate its application here. More recently QM/MM approaches using ZIndo/S as the QM component have been applied to the entire FMO complex and found remarkable agreement with the site-energies for all pigments except BChl 1 [58] and there has been a study comparing the use of ZIndo/S and TD-DFT in QM/MM studies of the FMO that concluded that the force-field had a greater impact than the QM approach [59]. Additionally, we would highlight that in the past it was found that it is not the quantum chemical method but how the structural models are generated that can be a strong determinant of whether a particular approach is informative or not the key factor being that the structural models should be optimized under a perturbation rather than arbitrarily conformationally distorted [51]. This is the case for our factorial relaxation method where all relaxed molecular components are optimized in the presence of the perturbation induced by the fixed conformations of the other parts of the molecule. Even so, a natural follow-up to this work would be to apply the factorial relaxation experimental design with a different method of geometry relation and/or excitation energy calculation.

A potential problem that we did observe with the PM6 optimizations was that the C3-acetyl dihedral angles in the 3EOJ structures were consistently optimized to 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 redshift to result from the protein's interaction with the acetyl. This should not have a deleterious effect on the assessment of the inter-pigment variations discussed here as if the protein induced an out-of-plane rotation that corresponds to a blue-shifting perturbation, our analysis would find a smaller redshifting perturbation of this acetyl compared to BChls with more in-plane acetyls where a smaller redshift corresponds to a blue-shift in relative terms.

One area where in hindsight we could have done better would have been to optimize the BChl pigments in the presence of their immediate protein environments with the PM6 method and use those structures as the starting point for our factorial relaxation experiment. This may have significantly ameliorated the inter-structural variation. Additionally, this may have allowed better insight regarding the relative importance of each molecular component as we can envisage that the PM6 preferences for specific bond parameters (*e.g.*, C-C bond lengths) could bias larger molecular components toward the possession of larger effect sizes since there would be more such deviations to correct. We would certainly recommend that any future application of this method be performed with this addition.

4.2 Conformational Perturbations of the BChl Site-energies

In this study we have investigated the role of protein induced conformational distortions of bacteriochlorophylls in the Fenna-Matthews-Olson protein complex in modulating their individual site-energies. We started by describing the BChl conformations in terms of deformations along the normal-modes of the tetrapyrrole macrocycle using the NSD procedure of Shelnutt and coworkers [33] that allowed the salient features of each BChls' conformation to be discovered. In addition to the reliability afforded by recent high-resolution structure determinations that employed careful refinement of the BChls we confirmed the assertion that each BChl adopts a unique and particular macrocycle conformation by qualitative comparison of the pigment conformations from different crystal structures. We also described how these conformations may influence the site-energies with reference to the known effects of macrocycle distortions obtained from previous model compound and theoretical studies [13,14,26,51] and demonstrated that a distortion known to influence the Qy absorption, B_{1u} ruffling [51], was highly correlated with the experimental site-energies. We then focused on determining if we could explicitly calculate the conformational effect *via* the semi-empirical PM6//ZIndo/S method and distinguish the influence of local perturbations such as the C3-acetyl dihedral or distortion of the MeCP component from the general conformational properties of the macrocycle such as its degree and type of nonplanarity. We first revisited an example of the classic strategy for achieving this goal, which we termed sequential truncation, and showed by quantitative comparison of the resultant calculated site-energy perturbations to the BChl NSD deformations that the effects of macrocyclic distortion and the molecular environment in which it occurs are coupled. We then demonstrated that this coupling works both ways by showing that the C3-acetyl dihedral Qy rotational energy profile behaves differently depending on the specific macrocycle conformation of the BChl. Taken together these results highlighted the necessity for a new approach to delineate the site-energy contributions of the conformations of individual BChl components.

After providing the theoretical basis of our proposed solution of partitioning BChl into its molecular constituents and systematically relaxing each *via* PM6 partial optimization for subsequent ZIndo/S calculation in a factorial screen we described a variety of analyses of these data. We first discussed considerable inter-structure variability and concluded that only the more recent, higher-resolution structure determinations proved suitable for detailed comparisons. We then looked at some general trends that were common to most of the BChls studied. Here we found that: 1) in terms of inter-pigment variability in the FMO, the magnitudes of the conformational site-energy perturbations are in the approximate order, macrocycle > acetyl > alkyls > Mg > MeCP and 2) when BChl component interactions are modelled some of the influence of the acetyl conformation is absorbed by its interactions, including that with the macrocycle conformation. These results provide strong evidence that the macrocycle conformation is a dominant influence on the site-energy distribution in terms of conformational effects but also show that there is significant interplay between all the

structural features of BChls. Moreover, in studies that have used the truncation method [52], it is possible that effects attributed to differences in particular substituent conformations (*e.g.*, of the acetyl or MeCP) are as much a product of the macrocycle conformation as they are of differences in these local parameters. Whilst for now our take-home message with respect to the intra-molecular conformational perturbation interactions is that they are present at significant magnitudes a natural follow up analysis of the data we have generated would be to identify the conformational and electronic origins of the interactions by comparison of the $\beta_{x:y}$ coefficients to the conformational and electronic parameters present in individual models within factorial relaxation series or each experimental unit as described in §2.1. As well as affording deep insight into the mechanism of macrocycle conformational perturbation the discovery of the physical origin of specific interactions would allay any concerns that they may be artifacts of the computational approach employed.

Next we looked at how the conformational perturbations may contribute to the observed site-energy distributions in the FMO complexes from *Ptc. aestuarii*, *Pel. phaeum* and *Clb. tepidum* by comparing our results to the available experimentally derived references. Here we found evidence that the overall conformational perturbation could contributes to the site-energy distributions in these species. A greater revelation though was that the isolated macrocycle contribution corresponds even more with the experimental values. Whilst we are uncertain if this is the result of improved robustness in the calculation ΔQy_{Mac}, since it is obtained using the results from all models in the factorial screen (whereas the overall perturbation is obtained from just the protein bound structure) or is a real phenomenon, the result suggests a key role for macrocycle deformation in site-energy modulation. A compelling concordance between the correlation of the calculated macrocycle induced site-energy perturbations and the experimental references 3EOJ, which possessed a remarkable B1u(2) / experimental site-energy relationship, lends further weight to this notion. In this analysis we also found evidence that the macrocycle perturbation was relatively conserved between the two spectral type II species, *Ptc. aestuarii* and *Pel. phaeum*, whilst the difference between these species' BChl 4 pigments was directly proportional to the difference in the experimental values, the latter an instance of functional divergence.

From one viewpoint, the reasonably strong correlations observed between the conformational perturbations and the reference site-energies (§3.3.4) are surprising since these variations have been previously explained in their entirety by the electrostatic environments of the pigments. For instance, the PBQC method of Muh *et al*. [7,8] predicts the Louwe site-energies to well within 2 nm ($Qy_{Louwe} = 0.93 \pm 0.01 * Qy_{PBQC} + 52 \pm 45$, $R^2 = 0.98$) and it is disconcerting that two different independent effects explain the same property. Konig and Neugebauer have contested that some of the electrostatic arguments in the PBQC calculation are flawed and highlight the numerous difficulties in predicting the FMO site-energies directly from the crystal structure (notably, advising that optimizations are performed on the BChls in their respective binding-pockets before carrying out excited-state calculations, an element our model lacks; see above) [27]. Additionally, another benchmarking study on QM/MM methods in the FMO found that *in vacuo* excited-state calculations on appropriately *in situ* optimized BChls yields energies that range over $400cm^{-1}$ [60], which encompasses a large range of the Louwe site-energies. Comparably, our simplest calculations of the total conformational perturbations vary between $180\text{-}340cm^{-1}$ within structures 3EOJ, 3ENI and 3VDI and although slightly higher we see similar variation of our ΔQy_{Mac} parameter (Fig. S14) and the correlation between this feature in some structures with the reference values is uncanny (§3.3.3). To reconcile this apparent contradiction one need consider a few potentialities. Our calculations of the conformational contributions to the site-energies could be inaccurate and the assumption that a relative comparison benefits from a cancellation of errors is misplaced; the finding that the electrostatic environment dictates the site-energies in

its entirety is an exaggeration or both forces are significant and there is a complex interplay between them when operating on the same BChl. From our perspective, the latter seems most reasonable since the idea that the conformational perturbation may behave differently in different electrostatic environments is in line with our thesis that the intramolecular environment modifies the site-energy shifts of the same conformational changes. Similarly then, we suppose that a conformationally distorted BChl would respond uniquely to electrostatic influences compared to a differently distorted BChl due to the corresponding electronic structural differences between the two structures. This could be explored in our factorial relaxation framework where toggling the presence of electrostatically important residues can be included in each experimental unit.

Our final analysis comparing the calculated macrocycle perturbations from sequential truncation to those obtained from factorial relaxation demonstrated that although large differences are modelled in proportion the two methods provide disparate results when subtle differences are considered. The BChl a peripheral substituents were seen to enhance the macrocycle effect and thus results based on truncation underestimate this feature. We suggest that our approach is better theoretically justified so that if its implementation is feasible for the problem at hand should be employed preferentially. Speculatively, this boosting effect the BChl a substituents have on the physicochemical magnitude of macrocycle distortion could mean that the structural decoration of the macrocycle allows the protein to exert improved control of the site-energies. Since the evolutionary emergence of the structures of the photosynthetic cofactors is driven by functional adaption this hints at the possibility that in analogy to the idea that the acquisition of desirable redshifted absorptions drove the selection of Mg coordination and substitution with π-extending groups like the acetyl [61–63], perhaps the various substituents of BChl a were also selected to affect efficient fine tuning by conformational control.

5. Conclusion

This paper has described a series of analyses directed towards elucidating the influence of BChl conformational variability on the pigment site-energies in the FMO protein. Our main contribution stems from the *factorial relaxation* approach we developed to accurately delineate the influence of the protein induced tetrapyrrole distortion from the peripheral substituent conformations of BChls a. We demonstrated the efficacy of factorial relaxation by quantitating the contributions of the BChl constituents to the overall conformational effect. The inaccurate *sequential truncation* approach is not restricted to the distant past [52] and we suggest that our method should be preferentially employed whenever possible. Our results provide general insights into the modulation of BChl site-energies in photosynthetic proteins. In the FMO, we found that some site-energy variations could be accounted for almost entirely by a simple estimate of the total conformational effect, but this result is statistically inconclusive. However, the influence of the isolated macrocycle and C3-acetyl are in statistically significant, quantitative concordance with the site-energy distributions provided by currently accepted reference sources. In a cross-species comparison, our method suggests that some of the macrocycle perturbations are conserved amongst BChls belonging to the same spectral classification. These results provide further evidence that the modulatory mechanism we have termed conformational control [12,15] is likely exploited by all tetrapyrrole containing proteins in the initial components of the photosynthetic apparatus.

Acknowledgements

This work was supported by Science Foundation Ireland (SFI P.I. 09/IN.1/B2650 and SFI IvP 13/IA/1894). Calculations were performed on the Lonsdale cluster maintained by the Trinity

Centre for High Performance Computing. This cluster was funded through grants from Science Foundation Ireland.

References

[1] T. Renger, Theory of excitation energy transfer: from structure to function, Photosynth. Res., 102 (2009) 471–485.
[2] M. Sener, J. Strumpfer, J. Hsin, D. Chandler, S. Scheuring, C.N. Hunter, K. Schulten, Förster energy transfer theory as reflected in the structures of photosynthetic light-harvesting systems, ChemPhysChem, 12 (2011) 518–531.
[3] L.K. Hanson, J. Fajer, M.A. Thompson, M.C. Zerner, Electrochromic effects of charge separation in bacterial photosynthesis: theoretical models, J. Am. Chem. Soc., 109 (1987) 4728–4730.
[4] X. Lin, H.A. Murchison, V. Nagarajan, W.W. Parson, J.P. Allen, J.C. Williams, Specific alteration of the oxidation potential of the electron donor in reaction centers from Rhodobacter sphaeroides, Proc. Natl. Acad. Sci. U.S.A., 91 (1994) 10265–10269.
[5] J.P. Allen, K. Artz, X. Lin, J.C. Williams, A. Ivancich, D. Albouy, T.A. Mattioli, A. Fetsch, M. Kuhn, W. Lubitz, Effects of hydrogen bonding to a bacteriochlorophyll-bacteriopheophytin dimer in reaction centers from Rhodobacter sphaeroides, Biochemistry, 35 (1996) 6612–6619.
[6] J. Eccles, B. Honig, Charged amino acids as spectroscopic determinants for chlorophyll *in vivo*, Proc. Natl. Acad. Sci. U.S.A., 80 (1983) 4959–44962.
[7] F. Muh, A. Madjet Mel, J. Adolphs, A. Abdurahman, B. Rabenstein, H. Ishikita, E.W. Knapp, T. Renger, Alpha-helices direct excitation energy flow in the Fenna Matthews Olson protein, Proc. Natl. Acad. Sci. U.S.A., 104 (2007) 16862–16867.
[8] J. Adolphs, F. Muh, A. Madjet Mel, T. Renger, Calculation of pigment transition energies in the FMO protein: from simplicity to complexity and back, Photosynth. Res., 95 (2008) 197–209.
[9] K.M. Barkigia, L. Chantranupong, K.M. Smith, J. Fajer, Structural and theoretical models of photosynthetic chromophores. Implications for redox, light-absorption properties and vectorial electron flow, J. Am. Chem. Soc., 110 (1988) 7566–7567.
[10] M.O. Senge, New trends in photobiology, J. Photochem. Photobiol. B:Biol., 16 (1992) 3–36.
[11] M.O. Senge, Excercises in molecular gymnastics – bending, stretching and twisting porphyrins, Chem. Commun., (2006) 243–256.
[12] M.O. Senge, S.A. MacGowan, J.M. O'Brien, Conformational control of cofactors in nature – The influence of protein-induced macrocycle distortion on the biological function of tetrapyrroles. Chem. Commun., 51 (2015), 17031–17083.
[13] G. Zucchelli, D. Brogioli, A.P. Casazza, F.M. Garlaschi, R.C. Jennings, Chlorophyll ring deformation modulates Qy electronic energy in chlorophyll-protein complexes and generates spectral forms, Biophys. J., 93 (2007) 2240-2254.
[14] G. Zucchelli, S. Santabarbara, R.C. Jennings, The Q(y) absorption spectrum of the light-harvesting complex II as determined by structure-based analysis of chlorophyll macrocycle deformations, Biochemistry, 51 (2012) 2717–2736.
[15] S.A. MacGowan, M.O. Senge, Conformational control of cofactors in nature--functional tetrapyrrole conformations in the photosynthetic reaction centers of purple bacteria, Chem. Commun., 47 (2011) 11621–11623.
[16] S.A. MacGowan, M.O. Senge, Computational quantification of the physicochemical effects of heme distortion: redox control in the reaction center cytochrome subunit of *Blastochloris viridis*, Inorg. Chem., 52 (2013) 1228–1237.
[17] J.M. Olson, The FMO protein, Photosynth. Res., 20 (2005) 421–427.

[18] R.E. Fenna, B.W. Matthews, Chlorophyll arrangement in a bacteriochlorophyll protein from Chlorobium limicola, Nature, 258 (1975) 573–577.
[19] R.E. Fenna, L.F. Eyck, B.W. Matthews, Atomic coordinates for the chlorophyll core of a bacteriochlorophyll a-protein from green photosynthetic bacterial, Biochem. Biophys. Res. Commun., 75 (1977) 751–756.
[20] B.W. Matthews, R.E. Fenna, M.C. Bolognesi, M.F. Schmid, J.M. Olson, Structure of a bacteriochlorophyll a-protein from the green photosynthetic bacterium Prosthecochloris aestuarii, J. Mol. Biol., 131 (1979) 259–285.
[21] D.E. Tronrud, M.F. Schmid, B.W. Matthews, Structure and X-ray amino acid sequence of a bacteriochlorophyll A protein from Prosthecochloris aestuarii refined at 1.9 A resolution, J. Mol. Biol., 188 (1986) 443–454.
[22] D.E. Tronrud, B.W. Matthews, Refinement of the Structure of a Water-Soluble Antenna Complex from Green Photosynthetic Bacteria by Incorporation of the Chemically Determined Amino Acid Sequence, in: J. Deisenhofer, J.R. Norris (Eds.) The Photosynthetic Reaction Center, Academic Press, New York, 1993, pp. 13–21.
[23] D.E. Tronrud, J. Wen, L. Gay, R.E. Blankenship, The structural basis for the difference in absorbance spectra for the FMO antenna protein from various green sulfur bacteria, Photosynth. Res., 100 (2009) 79–87.
[24] D.E. Tronrud, J.P. Allen, Reinterpretation of the electron density at the site of the eighth bacteriochlorophyll in the FMO protein from Pelodictyon phaeum, Photosynth. Res., 112 (2012) 71–74.
[25] M. Schmidt am Busch, F. Müh, M. El-Amine Madjet, T. Renger, The Eighth Bacteriochlorophyll Completes the Excitation Energy Funnel in the FMO Protein, J. Phys. Chem. Lett., 2 (2011) 93–98.
[26] E. Gudowska-Nowak, M.D. Newton, J. Fajer, Conformational and environmental effects on bacteriochlorophyll optical spectra: correlations of calculated spectra with structural results, J. Phys. Chem., 94 (1990) 5795–5801.
[27] C. König, J. Neugebauer, Protein Effects on the Optical Spectrum of the Fenna–Matthews–Olson Complex from Fully Quantum Chemical Calculations, Journal of Chemical Theory and Computation, 9 (2013) 1808–1820.
[28] J. Gao, W.J. Shi, J. Ye, X. Wang, H. Hirao, Y. Zhao, QM/MM modeling of environmental effects on electronic transitions of the FMO complex, J. Phys. Chem. B, 117 (2013) 3488–3495.
[29] Y.F. Li, W. Zhou, R.E. Blankenship, J.P. Allen, Crystal structure of the bacteriochlorophyll a protein from *Chlorobium tepidum*, J. Mol. Biol., 271 (1997) 456–471.
[30] A. Camara-Artigas, R.E. Blankenship, J.P. Allen, The structure of the FMO protein from *Chlorobium tepidum* at 2.2 A resolution, Photosynth. Res., 75 (2003) 49–55.
[31] A. Ben-Shem, F. Frolow, N. Nelson, Evolution of photosystem I – from symmetry through pseudosymmetry to asymmetry, FEBS Letters, 564 (2004) 274–280.
[32] C.R. Larson, C.O. Seng, L. Lauman, H.J. Matthies, J. Wen, R.E. Blankenship, J.P. Allen, The three-dimensional structure of the FMO protein from *Pelodictyon phaeum* and the implications for energy transfer, Photosynth. Res., 107 (2011) 139–150.
[33] C.J. Medforth, M.O. Senge, K.M. Smith, L.D. Sparks, J.A. Shelnutt, Nonplanar distortions modes for highly substituted porphyrins, J. Am. Chem. Soc., 114 (1992) 9859–9869.
[34] X. Li, N. Sudarsanam, D.D. Frey, Regularities in data from factorial experiments, Complexity, 11 (2006) 32–45.
[35] W. Jentzen, X.-Z. Song, J.A. Shelnutt, Structural Characterization of Synthetic and Protein-Bound Porphyrins in Terms of the Lowest-Frequency Normal Coordinates of the Macrocycle, J. Phys. Chem. B, 101 (1997) 1684–1699.

[36] J.J. Stewart, Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements, J. Mol. Model., 13 (2007) 1173–1213.
[37] M.C. Zerner, Semiempirical Molecular Orbital Methods, in: K.B. Lipkowitz, D.B. Boyd (Eds.) Reviews in Computational Chemistry, Wiley, New York, 1991, pp. 313–365.
[38] M.A. Thompson, M.C. Zerner, A theoretical examination of the electronic structure and spectroscopy of the photosynthetic reaction center from Rhodopseudomonas viridis, J. Am. Chem. Soc., 113 (1991) 8210–8215.
[39] J. Linnanto, J.E.I. Korppi-Tommola, V.M. Helenius, Electronic States, Absorption Spectrum and Circular Dichroism Spectrum of the Photosynthetic Bacterial LH2 Antenna of Rhodopseudomonas acidophilaas Predicted by Exciton Theory and Semiempirical Calculations, J. Phys. Chem. B, 103 (1999) 8739–8750.
[40] J. Linnanto, J. Korppi-Tommola, Spectroscopic properties of Mg-chlorin, Mg-porphin and chlorophylls a, b, c1, c2, c3 and d studied by semi-empirical and ab initio MO/CI methods, Phys. Chem. Chem. Phys., 2 (2000) 4962–4970.
[41] J. Linnanto, J. Korppi-Tommola, Spectroscopic Properties of Mg−Chlorin, Mg−Bacteriochlorin, and Bacteriochlorophyllsa,b,c,d,e,f,g, and h Studied by Semiempirical and Ab Initio MO/CI Methods, J. Phys. Chem. A, 105 (2001) 3855–3866.
[42] J. Linnanto, J. Korppi-Tommola, Semiempirical PM5 molecular orbital study on chlorophylls and bacteriochlorophylls: comparison of semiempirical, ab initio, and density functional results, J. Comput. Chem., 25 (2004) 123–138.
[43] J. Linnanto, J. Korppi-Tommola, Structural and Spectroscopic Properties of Mg−Bacteriochlorin and Methyl Bacteriochlorophyllides a,b,g, and h Studied by Semiempirical, ab Initio, and Density Functional Molecular Orbital Methods, J. Phys. Chem. A, 108 (2004) 5872–5882.
[44] J. Linnanto, J. Korppi-Tommola, Quantum chemical simulation of excited states of chlorophylls, bacteriochlorophylls and their complexes, Phys. Chem. Chem. Phys., 8 (2006) 663–687.
[45] J. Linnanto, A. Freiberg, J. Korppi-Tommola, Quantum chemical simulations of excited-state absorption spectra of photosynthetic bacterial reaction center and antenna complexes, J. Phys. Chem. B, 115 (2011) 5536–5544.
[46] 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, D.J. Fox, Gaussian 09, Revision D.01, in, Gaussian, Inc., Wallingford CT, 2009.
[47] R.D.C. Team, R: A Language and Environment for Statistical Computing, in, Vienna, Austria, 2010.
[48] B.S. Everitt, S. Landau, M. Leese and D. Stahl, *Cluster Analysis*, Wiley, West Sussex, England, 2011.
[49] A.C. Rencher and W.F. Christensen, *Methods of Multivariate Analysis*, Wiley, Hoboken, NJ, 2012.

[50] S.T. Daurat-Larroque, K. Brew, R.E. Fenna, The complete amino acid sequence of a bacteriochlorophyll a-protein from *Prosthecochloris aestuarii*, J. Biol. Chem., 261 (1986) 3607–3615.
[51] R.E. Haddad, S. Gazeau, J. Pecaut, J.C. Marchon, C.J. Medforth, J.A. Shelnutt, Origin of the red shifts in the optical absorption bands of nonplanar tetraalkylporphyrins, J. Am. Chem. Soc., 125 (2003) 1253–1268.
[52] H. Yamasaki, Y. Takano, H. Nakamura, Theoretical investigation of the electronic asymmetry of the special pair cation radical in the photosynthetic type-II reaction center, J. Phys. Chem. B, 112 (2008) 13923–13933.
[53] K.B. Sai Sankar Gupta, A. Alia, H.J. de Groot, J. Matysik, Symmetry break of special pair: photochemically induced dynamic nuclear polarization NMR confirms control by nonaromatic substituents, J. Am. Chem. Soc., 135 (2013) 10382–10387.
[54] S.I.E. Vulto, M.A. de Baat, R.J.W. Louwe, H.P. Permentier, T. Neef, M. Miller, H. van Amerongen, T.J. Aartsma, Exciton Simulations of Optical Spectra of the FMO Complex from the Green Sulfur Bacterium Chlorobium tepidum at 6 K, J. Phys. Chem. B, 102 (1998) 9577-9582.
[55] R.J.W. Louwe, J. Vrieze, A.J. Hoff, T.J. Aartsma, Toward an integral interpretation of the optical steady-state spectra of the FMO-complex of *Prosthecochloris aestuarii*. 2. exciton simulations, J. Phys. Chem. B, 101 (1997) 11280–11287.
[56] E.L. Read, G.S. Engel, T.R. Calhoun, T. Mancal, T.K. Ahn, R.E. Blankenship, G.R. Fleming, Cross-peak-specific two-dimensional electronic spectroscopy, Proc. Natl .Acad. Sci. USA, 104 (2007) 14203-8.
[57] E.L. Read, Investigations of Photosynthetic Light Harvesting by Two-Dimensional Electronic Spectroscopy, Ph.D Thesis, University of California, Berkeley, 2008.
[58] W. Jentzen, X.-Z. Song, J.A. Shelnutt, Structural Characterization of Synthetic and Protein-Bound Porphyrins in Terms of the Lowest-Frequency Normal Coordinates of the Macrocycle, J. Phys. Chem. B, 101 (1997) 1684-1699.
[59] D.E. Tronrud, J. Wen, L. Gay, R.E. Blankenship, The structural basis for the difference in absorbance spectra for the FMO antenna protein from various green sulfur bacteria, Photosynth. Res., 100 (2009) 79-87
[60] N.H. List, C. Curutchet, S. Knecht, B. Mennucci, J. Kongsted, Toward Reliable Prediction of the Energy Ladder in Multichromophoric Systems: A Benchmark Study on the FMO Light-harvesting Complex, J. Chem. Theory Comput., 9 (2013) 4928-4938.
[61] M. D. Dolphin (Ed.) The Porphyrins, Academic Press, New York, 1978, pp. 1–165.
[62] D. Mauzerall, Why chlorophyll?, Ann. N. Y. Acad. Sci., 206 (1973) 483–494.
[63] L.O. Bjorn, G.C. Papageorgiou, R.E. Blankenship, Govindjee, A viewpoint: why chlorophyll a?, Photosynth. Res., 99 (2009) 85–98.

Figure and Tables:

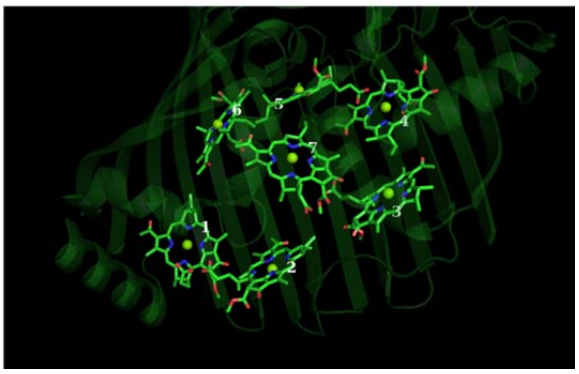
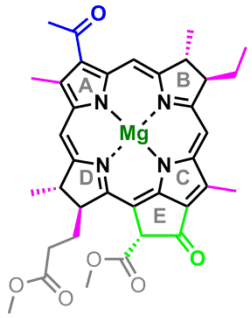


Fig. 1. The BChl arrangement in the Fenna-Matthews-Olson protein (phytyl chains have been truncated for clarity). Drawn using PyMol and the coordinates from PDB ID: 3BCL [21]. BChl 8 it not present in this structure although when present is situated outside of the encapsulating β-sheets, close to BChl 1.

a)



b)

Fragment	Group	Color	Abbrev.
Mg	1	dark green	Mg
Bacteriochlorin ring	2	black	Mac.
Methyl cyclopentanone ring E	3	green	MeCP
C3-acetyl	4	blue	Acetyl
C2,7,12,17,18-methyls + C8-ethyl	5	purple	Alkyls

Fig. 2. a) Definition of the structural model used in the partial optimization calculations where atoms of the same color are part of the same group (light-grey indicates the omission of the phytyl and Me-ester). Note that the ring nomenclature indicated in the above structures corresponds to their labelling in the PDB standard. b) Molecular partitioning scheme of the BChls used to construct the partial optimizations. Each group of atoms (molecular fragment) is either frozen or optimized in each experimental unit (permutation) and thus its contribution to the BChl's *intrinsic* site-energy is isolated from the other BChl components.

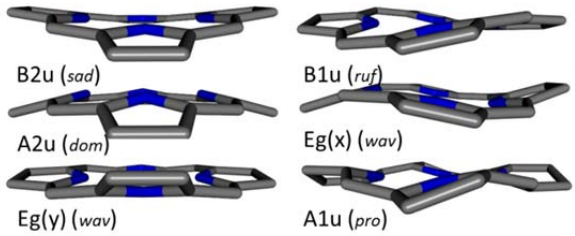


Fig. 3. Idealizations of the six commonly observed out-of-plane macrocycle conformations. In the *sad* conformations the pyrrole rings are alternately tilted up/ down along their C_a-C_a axes producing a saddle shape. In the *ruf* conformation, the meso carbons are alternately displaced above and below the mean-plane that could be affected by twisting a planar macrocycle. In the *dom* conformation 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.

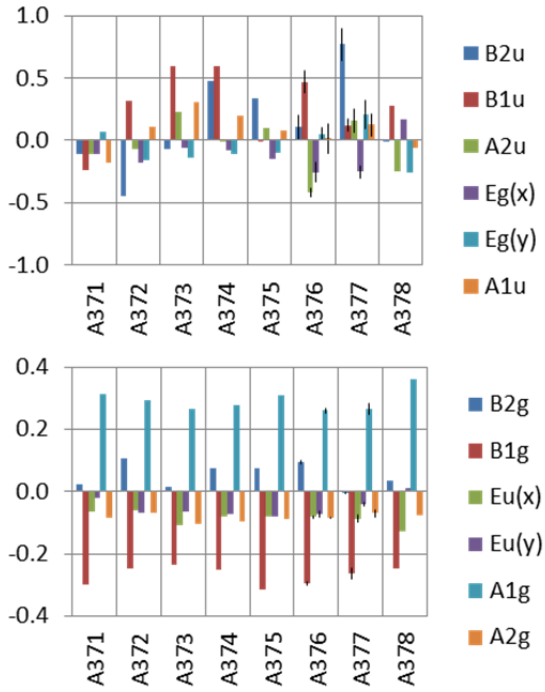


Fig. 4. Normal structural decomposition analysis showing the minimum basis normal-coordinate displacements of each BChl pigment in the *P. aestuarii* FMO protein as described by PDB ID: 3EOJ. Out-of-plane (top) and in-plane (bottom). Y-axes in Ångstroms; the average values for the four alternate conformers of BChls 6 and 7 are given with the error bars indicating the twice the standard deviation. The adequacy of the NSD min. basis to describe the macrocycle geometry is measured by d_{oop}, which for BChls 1 to -7 respectively is: 0.0422, 0.0265, 0.0495, 0.0415, 0.0502, 0.0579* and 0.0450*and d_{ip}: 0.0545, 0.0505, 0.0547, 0.0492, 0.0451, 0.0492* and 0.0533*. * Denotes that the value is the average over the four possible alternative conformations reported in the crystal structure.

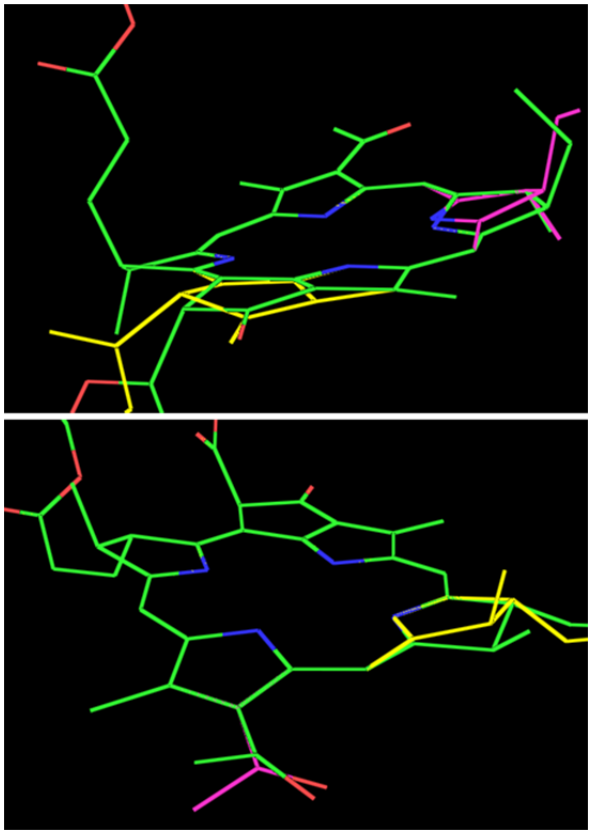


Fig. 5. 3EOJ-A376 (top) and -A377 (bottom) conformers from alternate occupancies. AC (green/green), AD (green/ yellow), BC (pink/ green), BD (pink/ yellow). For A376 the occupancies for the A, B, C and D entries are 0.42, 0.57, 0.38 and 0.62, respectively and so give the AC, AD, BC and BD conformations uncorrelated occupancies of 0.16, 0.26, 0.22 and 0.35. For A377 the A-D occupancies are 0.42, 0.57, 0.48 and 0.52 giving the AC-BD conformations occupancies of 0.20, 0.22, 0.27 and 0.30.

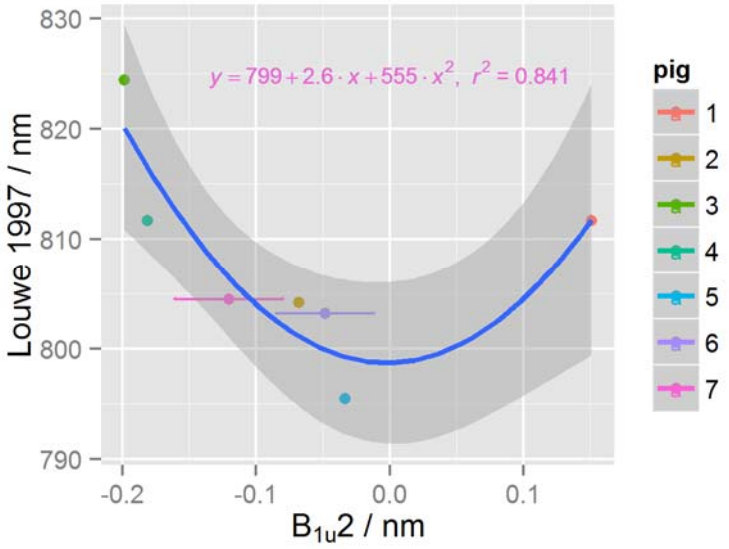


Fig. 6. Correlation of the fitted site-energies from Adolphs *et al*. [8] with the $B_{1u}(2)$ deformation of the BChls as resolved in PDB ID: 3EOJ. BChls 6 and 7 are averaged by their multiple conformers; shaded area indicates the standard error of prediction.

Table 1. 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*. [26].

Code[a]	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}	D_{ip}	D_{oop}
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	-0.67	-0.48
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	-0.69	-0.60
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	-0.69	-0.59
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	-0.68	-0.69
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	-0.56	-0.77
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	-0.44	-0.78
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	-0.37	-0.81

[a] These codes denote the structural models of the FMO BChls used in Gudowska-Nowak *et al*.'s [26] 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.

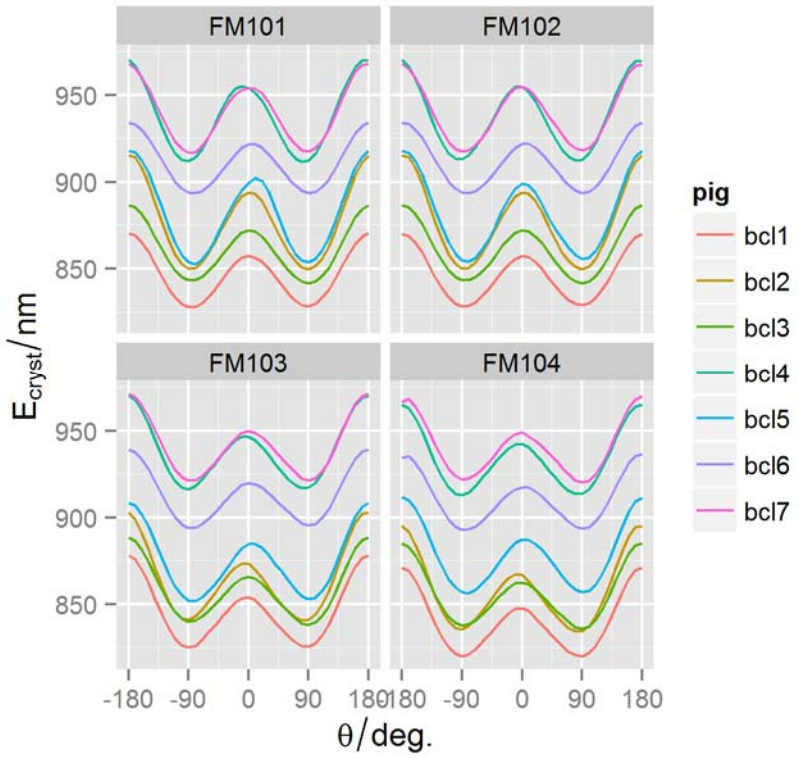


Fig. 7. PM6//ZIndo/S REPs of the BChls from PDB ID: 3BCL with various levels of relaxation; rigid rotation (top left), acetyl-Me hydrogen dihedral relaxation (top right), all acetyl relaxation including the $C3^1$-C3 bond-length (bottom left) and as previously stated with the addition that all macrocycle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

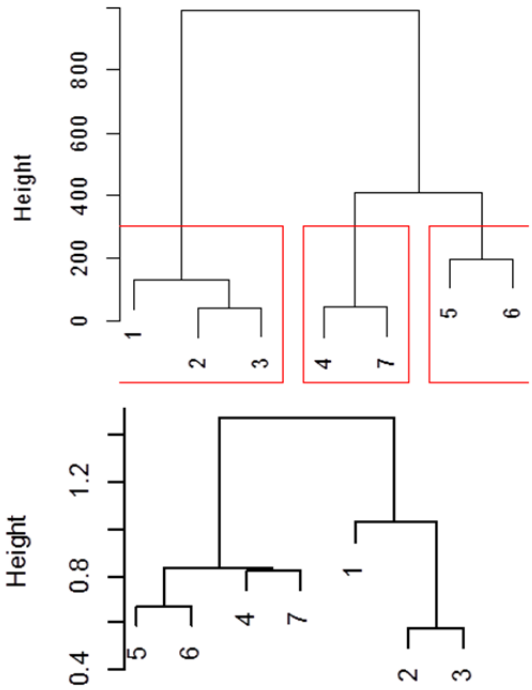


Fig. 8. Cluster dendrogram of the ZIndo/S REPs from series FM104 (top) and the NSD data from the 3BCL pigments (bottom); the AHC was performed using the squared Euclidean distances and Ward's method.

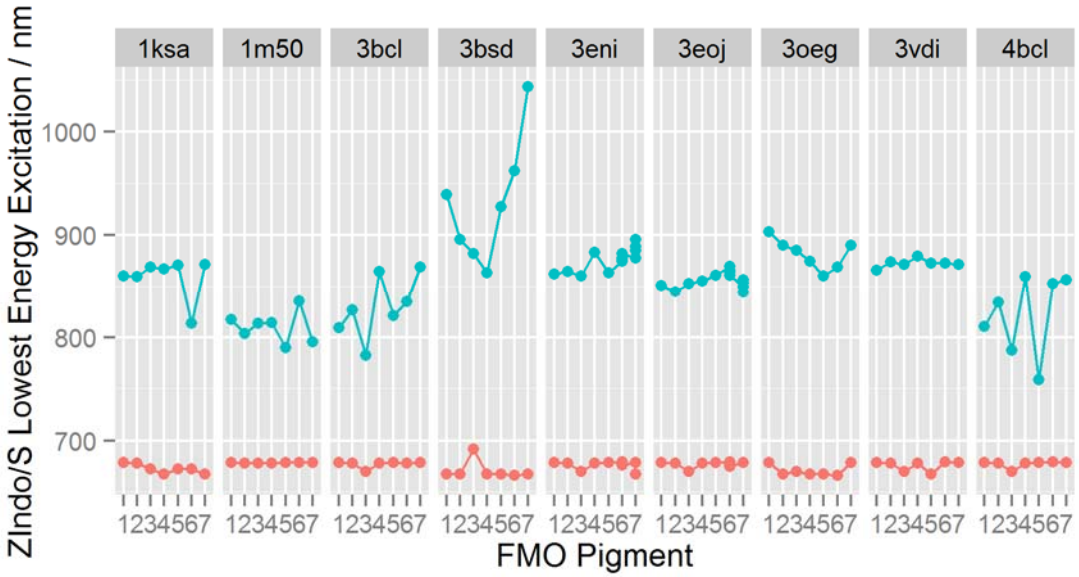


Fig. 9. The excitation energies for each FMO pigment calculated from the crystal structure geometry (cyan, top) and after full optimization of this starting geometry (red, bottom). The results for each pigment are given for nine separate PDB structures. For BChls 6 and 7 from both 3EOJ and 3ENI the data point is the average over the four multiple conformers.

Table 2. 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	Min.	Max	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[a]	860.4	942.3	25.3
3vdi	865.6	879.6	4.1
4bcl	759.0	859.5	38.7

[a] The maximum value in 3OEG belongs to the BChl 8 pigment.

Table 3. 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 optimization (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	β_{Mac}	β_{MeCP}	β_{Acetyl}	β_{Alkyls}	β_{Mg}
Mean	863	676	869	-162	-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
Min.	759	667	754	-311.4	-51.7	-37.5	-26.7	-10.3
1st Qu.	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
3rd Qu.	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

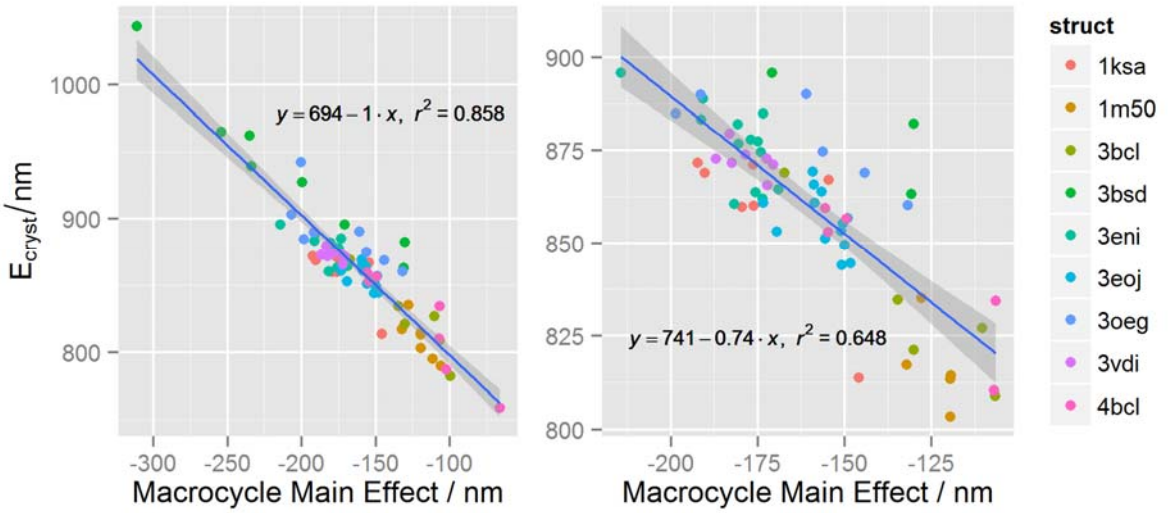


Fig. 10. The calculated unperturbed site-energies for all pigments from all structures in the crystal geometries *vs*. the estimated main effect of the macrocycle conformation, $E_{cryst} = \gamma_{Mac}\beta_{Mac}$ (right panel: E_{cryst} 800 – 900 nm, right); shaded area indicates the standard error of prediction.

Table 4. Resulting parameters of linear models obtained from ordinary least-squares (OLS) and robust linear modelling (RLM; iteratively reweighted least-squares) relating E_{cryst} to the calculated main conformation effects of the molecular fragments ($E_{cryst} = \gamma_{Mac}\beta_{Mac} + \ldots + \gamma_{Mg}\beta_{Mg} + c$) where β_{Mg} denotes the calculated main effect of the named fragment; n = 77).

	Intercept	$\beta_{Mac.}$	βMe_{CP}	$\beta_{Ace.}$	$\beta_{Alks.}$	β_{Mg}	R^2
OLS	695	-0.93	-0.86	-0.78	-0.68	0.53	0.97
RLM	687	-0.96	-0.71	-0.90	-0.73	-0.42	NA
OLS[a]	687	-0.95	-0.79	-0.92	-0.73	-0.34	0.98
RLM [a]	686	-0.96	-0.69	-0.93	-0.74	-0.61	NA

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

Table 5. The component main effects upon E_{ES1} for BChls 1 – 5 in 3EOJ from 1st to 3rd order models (data for BChls 6 and 7 are provided in Table S10–S13).

Pigment	Order	β_{Mg}	β_{Mac}	β_{MeCP}	β_{Acetyl}	β_{Alkyls}
A371	First	0.1	-155.6	-0.1	-18.8	1.2
	Second	-0.4	-149.3	1.9	-18.9	1.8
	Third	-2.3	-149.9	2.3	-15.1	0.6
A372	First	-0.8	-148.3	1.3	-17.9	-0.8
	Second	-0.2	-144.4	3.3	-12.6	3.5
	Third	-2.5	-147.4	2.8	-8	0.5
A373	First	-4.3	-169.5	-0.5	-7.9	-5.9
	Second	-4.5	-158.2	-0.6	-7	-2.4
	Third	-0.7	-154.8	0.6	-4.2	1.1
A374	First	-1.5	-150.5	-1	-21.7	-5.9
	Second	0	-139.8	1.4	-19.1	-0.2
	Third	-1	-140.2	1.6	-12.9	0.3
A375	First	-1.8	-173.4	-0.5	-5	-3.2
	Second	-4.7	-170.1	-1.4	-4.7	-0.8
	Third	-4.5	-167.7	-0.1	-3.7	0.5

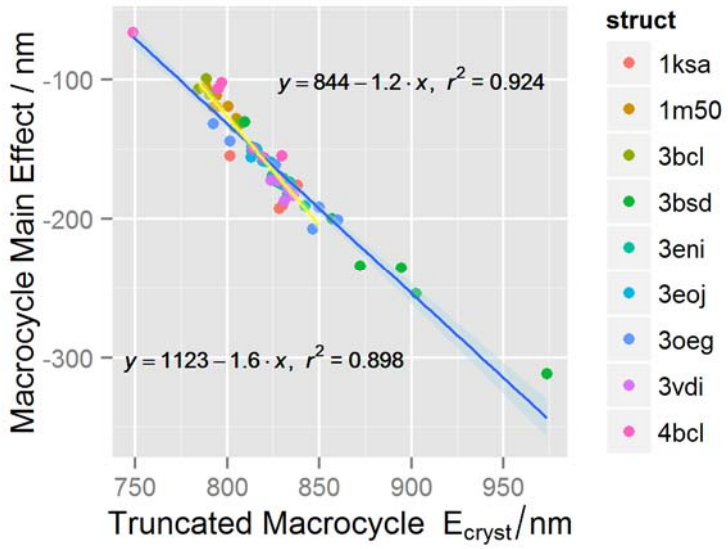


Fig. 11. Comparison of the macrocycle main effects in the BChl structures calculated *via* the factorial series of partial optimizations to the estimate afforded by a truncated model; yellow line illustrates regression of values over the range x = 775 – 850 nm (bottom regression equation), shaded areas indicates the standard error of prediction.