Biophysical Chemistry 153 (2011) 173–178

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Semiempirical configuration interaction calculations in biochemical environments Parametrization and application to γD-crystallin, an eye-lense protein

Sandra Kruse, Sebastian Krapf, Benjamin Lampe, Thorsten Koslowski *

Institut für Physikalische Chemie, Universität Freiburg, Albertstraße 23a, D-79104 Freiburg im Breisgau, Germany

ARTICLE INFO

Article history:
Received 1 October 2010
Received in revised form 11 November 2010
Accepted 12 November 2010
Available online 19 November 2010

Keywords:
Eye-lense protein
Optical absorption
Theory
Simulation
Semiempirical model
Charge transfer
Configuration interaction

ABSTRACT

We approach the problem of optical excitations in molecular aggregates in complex biochemical environments from a computational, all-atom perspective. The system is divided into a π orbital part described by a Pariser–Parr–Pople model with configuration interaction using singly excited Slater determinants (PPP-CIS). It is coupled to the protein and water charges of a classical force field. Strategies for a high-accuracy reparameterization and an efficient computational solution are presented. For γD-crystallin, a band edge consisting of charge-transfer states emerges for a coupled molecular aggregate compared to the uncoupled residues. The energies of some charge-transfer states strongly depend on the dielectric properties of the model, giving a first insight into the potential temporal evolution of these excitations. Possible biochemical implications are discussed.

1. Introduction

Important processes of life are based on the interaction of light with complex aggregates of absorbing molecules arranged in a protein matrix, followed by chemical reactions or the separation of charges. Among others, these fundamental reactions include photosynthesis [1–3], DNA damage and repair [4], visual perception [5,6] and magnetoreception [7–10]. Significant progress has been made in understanding the initial process of the absorption of light in the visible or near ultraviolet range, with contributions from both quantum chemistry and exciton theory. The problem of the optical absorption of large molecular aggregates and charge transfer in their excited states has also become of recent considerable interest in the field of light harvesting by organic solar cells [11].

It is now generally accepted that the properties of excited states cannot be directly deduced from a mean-field solution of the ground state electronic Hamiltonian – provided by Hartree–Fock or density functional theory – but require the consideration of correlation effects. From the perspective of ab initio quantum chemistry, complete active space self consistent field (CASSCF) calculations supplemented by elements of second order perturbation theory (CASPT2) have helped to gain quantitative insight into the spectra of comparatively small molecules since the 1990s. Pioneered by Roos and coworkers, biologically relevant molecules such as aromatic amino acid side chains have been studied, including the application of simple reaction fields of the Onsager type to address solvatochromic effects [12]. Ingenious methods have been developed to push the frontier of excited state ab initio theory to ever larger molecules, with a variant of coupled-cluster theory in the resolution of identity approximation (CC-RI) [13], large scale configuration interaction computations based on single excitations (CIS) [14] and time-dependent density functional theory (TDDFT) [15] tailored to biomolecules as important roads of theory. Fast, efficient and accurate by any ab initio standard, these methods are difficult to apply concerning the sheer size of proteins and their complexes or the time scales required to describe biological processes, with typical molecular dynamics simulations based on trajectories sampled over 10 ns with a 1–2 fs time step, each requiring the computation of the total energy and its gradients or, less demanding, a quantum-chemical post-processing of the simulated structures.

Large aggregates, on the other hand, lie usually in the field of exciton theory. Here, experimental or theoretical excitation energies of the constituents and their coupling are used in a simple variational scheme akin to tight-binding theory. Classical examples include Scheibe- and J-aggregates [16,17]. For biological systems, these approaches have become particularly useful in studying large light harvesting clusters [18], including the possibility of the formation of a coherent quantum state participating in energy transfer [19]. Usually, the models are restricted to a single excitation per molecule, short-range coupling and a dielectric continuum approach that mimics the protein and its aqueous environment.

* Corresponding author.
E-mail address: Thorsten.Koslowski@physchem.uni-freiburg.de (T. Koslowski).

0301-4622/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2010.11.005

In this work, we make the attempt to model the electronic excitation spectrum of a comparatively large and complex system on an atomistic basis. We parameterize the integrals of a semiempirical model making use of all low-energy $\pi-\pi$ transitions with the help of published high-level ab initio and experimental data. In this way, we hope to transfer the accuracy of experiments and ab initio methods to a system-specific semiemprical model that permits an efficient numerical solution. We focus on an eye-lense protein, γD-crystallin, as a model system. Our choice is motivated by the large content of aromatic amino acids serving as potentially interacting absorbing units, the availability of a microscopic structure model and a system size that is sufficiently small to permit comparatively fast benchmark computations.

The remaining part of this work is organized as follows. In the next section, we will describe the geometrical model underlying our computations, the methods used to generate microscopic structures, and the Hamiltonian applied to compute electronic excitations. It is followed by a description of the parametrization of the chromophores, and a section devoted to numerical results for the protein studied here, and a part devoted to computational aspects. Conclusions will be derived in the final section.

2. Methods

The crystallins form a significant fraction of the content (~90% of the protein) of the eye lenses of vertebrates [20–23], their fibers exhibit a high refractive index while maintaining transparency. Despite their intimate relation to chaperones, they exhibit a strong tendency towards aggregation, finally leading to the formation of cataracts. As they participate in the process of visual perception, have a complex evolutionary history not yet completely unrevelled and contribute to a pathological process, they have been the subject of numerous structural studies [24–30].

We have selected a protein from one of the more recent studies, which combined a two-dimensional NMR and a neutron diffraction experiment, with ten structures published. Hence, we do not only focus on a single geometry, but have a set of structures reflecting the flexibility of the protein at our disposal. Whereas similar information may be obtained from a molecular dynamics or Monte Carlo simulation using a single initial geometry, we prefer structural information that has a stronger experimental basis. Being interested in electronic phenomena involving π electrons, our choice among the crystallins was also motivated by the high content of aromatic amino acids in γD-crystallin.

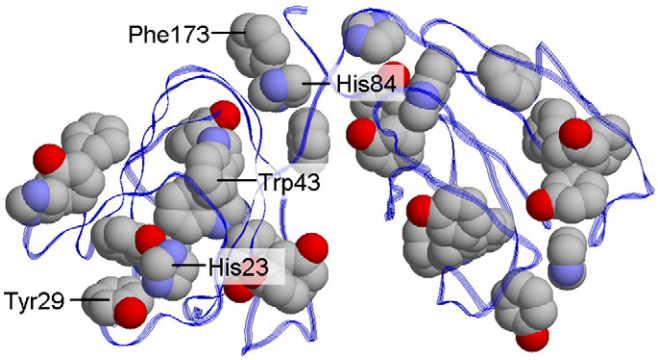

Fig. 1. Cartoon model of γD-crystallin from *H. sapiens* from a combined 2D NMR and neutron diffraction study [31], protein database entry 2KLJ. The strands indicate the protein backbone, spheres represent the nonhydrogen side chain atoms of aromatic amino acids. Carbons are plotted in gray, oxygens in red and nitrogens in blue. Selected amino acids that are discussed in the text have been labeled.

Our models are based on γD-crystallin from *H. sapiens* expressed by *E. coli* studied by Wang et al. [31], as shown in Fig. 1 as a cartoon model. The corresponding protein data base entry 2KLJ supplies ten structures compatible with the 2D NMR and neutron diffraction data. The published experimental structures include all amino acids according to the protein sequence. As with all NMR studies on proteins, degrees of freedom not obtainable from the 2D spectrum are modeled by a classical force field, e.g. of the Amber type in the Wang et al. study [31]. We have minimized these structures without constraints using the Amber 99 force field [32] implementation of the Tinker molecular modeling suite [33] utilizing the MINIMIZE program of the package. A water shell of 10 Å has been added, typically containing 2250 water molecules, and consecutively reminimized. A standard threshold for minimizations, a gradient norm of 10^{-2} kcal/mol atom Å, was always reached. 30 of 174 aromatic amino acids of the protein exhibit an aromatic character, we have six histidines, six phenylalanines, 14 tyrosines and four tryptophanes. The resulting models consist of 2825 atoms located in the protein, of which 200 contribute to the π electron Hamiltonian. With a slightly varying number of water molecules added for each structure, our models fall short of 10000 atoms by a narrow margin. This model reflects the experimentally determined composition of the eye lense [34] with a water mass fraction of ~0.6 to 0.7 and a remaining protein content.

We have verified our approach to create a set of protein models by applying the procedure described above to a γ-crystallin structure obtained in an X-ray study at a resolution of 1.47 Å performed by Najmudin et al. [28] (PDB entry 4GCR). After relaxation, the experimental and the model structure can hardly be distinguished by visual inspection, and the average root mean square deviation of the protein backbone dihedral angles amounts to 4.3°. Whereas updates of the Amber 99 force field show improvements in the balance between secondary structure elements for model peptides [35], the accuracy achieved in our study for the comparatively compact and rigid γ-crystallins using Amber 99 seems sufficient.

The total charge of the system is small (plus six elementary charges stemming from the histidines), so no counterions have been added to the water shell. We note that their use would be imperative to achieve global charge neutrality once a periodic simulation box and a k-space sampling technique such as the Ewald summation were used.

To describe the electronic structure of the π system, we make use of the semiempirical model of Pariser, Parr and Pople (PPP) [36–38]. It is based on the separability of σ and π contributions to the electronic Hamiltonian, which is exact at the mean field level for planar molecules. Hamiltonians of this type are frequently used in the field of combined classical and quantum mechanical molecular dynamics [39–41] or to describe the electronic excitations in conducting polymers [42] or chromophores [43]. As the aromatic amino acid side chains considered here interact weakly and display small fluctuations around their planar minimium structure, the $\sigma-\pi$ separation is only approximate. The PPP basis set consists of one p_z orbital per nonhydrogen atom. The zero differential overlap approximation is applied to all integrals that represent electron–electron interactions. All one- and two-electron integrals retained are treated as parameters, in the remaining part of this work they will be referred to as α_a (one-electron matrix diagonal), β_{ab} (one-electron off-diagonal) and γ_{ab} (two-electron integrals), where the linear Mataga–Nishimoto approximation is applied if $a \neq b$ [44]. Long-range one-electron interactions were assumed to decay exponentially with distance. Within a reference cartesian coordinate system, they are characterized by the parameters $V_{pp\sigma}$ and $V_{pp\pi}$, which are combined with the help of the Slater–Koster [45] rules for amino acid residues with an arbitrary mutual orientation. For this purpose,

atomic orbitals are constructed that lie orthogonal to the plane of the π system. Linear electron–phonon coupling,

$$\beta(r) = \beta_0 + 3.21(r-1.397), \tag{2.1}$$

is used for all nearest-neighbor bonds [43].

As an external potential contributing to the Hamiltonian diagonal, we consider the interaction with all charges residing in the protein and its water environment, which take the values assigned within the Amber 99 force field [32] and the TIP3P water model [46]. We have

$$V_{a,ext} = -\frac{e^2}{4\pi\varepsilon\varepsilon_0}\sum_{b\in E}\frac{z_b}{r_{ab}}. \tag{2.2}$$

The customary Amber rules apply for these interactions [32]: they are set to zero for the nearest and next-nearest neighbors and are reduced by a factor of 1.2 for neighbors separated by four bonds. In the closed shell ground state, the mean field PPP matrix elements read

$$h_{aa} = \alpha_a + V_{a,ext} - \frac{1}{2}q_a\gamma_{aa} + \sum_{b\neq a}(q_b - Z_b)\gamma_{ab} \tag{2.3}$$

for the diagonal and

$$h_{ab} = \beta_{ab} - \frac{1}{2}p_{ab}\gamma_{ab} \tag{2.4}$$

for the off-diagonal matrix elements. The charge and bond orders are denoted by q_a and p_{ab}, respectively. They are computed from a previous step of an iterative self-consistent field calculation of the Roothaan–Hall equations.

Electronic excitations are based on the converged PPP ground state, Ψ^0, taking into account all single excitations from occupied molecular orbitals i to vacant eigenstates k via a linear combination of Slater determinants,

$$\Psi = \Psi^0 + \sum_{ik} C_i^k \Psi_i^k \tag{2.5}$$

For the CIS matrix elements, we have [47]:

$$\left\langle \Psi_i^k \middle| \hat{H} \middle| \Psi_j^l \right\rangle = \sum_{ab}\left(2c_{la}c_{ja}c_{kb}c_{ib} - c_{ia}c_{ja}c_{kb}c_{lb}\right)\gamma_{ab} - (\varepsilon_k - \varepsilon_i)\delta_{ij}\delta_{kl} \tag{2.6}$$

making use of the ground state energy ε_i and LCAO-MO expansion coefficients c_{ia}. Intensities are computed in the dipole approximation [47].

With 112 occupied and 88 vacant states, the dimension of the CIS matrix equals 9856, which may look small by today's standards. Nevertheless, its diagonalization may become a serious computational bottleneck if applied for many steps of a molecular dynamics simulation. Hence, we will describe various cutoff strategies that help to reduce the size of the matrix, and we will study the performance gained by switching from a standard full matrix Givens–Householder transformation plus QL diagonalization algorithm to sparse matrix techniques.

3. Parametrization

For each amino acid residue, we vary the PPP parameters by a Monte Carlo procedure to obtain the best possible match (rms deviation minimum) between the lowest three to four simulated transitions and their experimental counterparts, as described in refs. [48,49] for the case of matching molecular orbitals in the electronic ground state. Here, published CASSCF and CASPT2 data on benzene [50], phenole [51] and indole [52] and their comparison to experiments in that work turned out to be essential to assign the correct $\pi-\pi^*$ transitions. For the imidazolium cation, we have performed a TDDFT calculation using a triple zeta plus polarization function basis set as a reference lacking reliable experimental data. All PPP parameterization calculations have been performed in the vacuum.

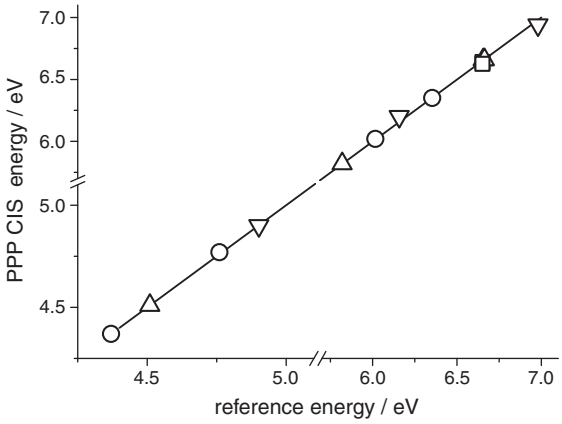

Fig. 2. Parametrization of the short-range interactions of the semiempirical electronic structure Pariser–Parr–Pople Hamiltonian. Aromatic amino acid side chain parameters have been varied to achieve the smallest energy difference between reference experimental and theoretical (configuration interaction singles, PPP CIS) $\pi-\pi^*$ optical absorption lines. The individual symbols refer to the amino acids His (□), Tyr (△), Trp (○) and Phe (▽). All energies are in electron Volts.

The quality of the parametrization fit is illustrated with the help of Fig. 2, which shows the PPP transition energies as a function of the reference excitations, with errors always less than 10^{-2} eV. As expected, intensities are less accurate, they can vary up to a factor of two. The parameters thus obtained are listed in Tables 1 and 2. We briefly touch the issue of fitting more complex spectra: here, it turned out to be preferable to find the rms minimum between the experimental spectrum and the computed broadened excitation lines. As an example, we present the near UV spectrum of buckminsterfullerene, C_{60}, in Fig. 3 [53,54].

Long-range one-electron matrix elements have been parameterized making use of the largest π molecular orbital splittings of stacked benzene dimers in the range from 3.4 to 7.5 Å using the B3LYP hybride functional and a 6–311+G basis set as a reference. Reference computations have been performed using the Gaussian 03 program, they have been fitted by varying the exponent and the prefactors. We obtain $V_{pp\sigma}(r) = 51.8exp(-1.37r)$ for the σ and $V_{pp\pi}(r) = -13.9\exp(-1.37r)$ for the π interactions, the corresponding fit is shown in Fig. 4.

4. Results

The computed UV/Vis spectrum of the protein is displayed in Fig. 5. It is based upon ten geometries, a dielectric constant $\varepsilon = 1.7$ [55] has been used, and a Gaussian broadening of the lines with an rms value of 5 nm has been applied. Only CI eigenstates up to 7.0 eV have been considered.

Table 1
Pariser–Parr–Pople atomic parameters for aromatic amino acid residues. One-electron orbital energy α_a, on-site electron–electron repulsion parameter γ_{aa} and number of π electrons per atom, n_a.

Acid	a	α_a [eV]	γ_{aa} [eV]	n_a
Phe	C	10.50	10.91	1
Tyr	C	10.50	11.30	1
Tyr	O	31.89	13.91	2
Trp	C	10.50	11.82	1
Trp	N	25.59	17.79	2
His^+	C	10.50	11.44	1
His^+	N	24.11	16.65	2

Table 2
Pariser–Parr–Pople bond parameters, β_0(ab), for aromatic amino acid residues. Nearest-neighbor one-electron hopping matrix elements at a distance of 1.307 Å.

Acid	ab	β_0(ab) [eV]
Phe	CC	2.395
Trp	CC	2.252
Trp	CN	3.098
Tyr	CC	2.204
Tyr	CO	1.930
His^+	CC	2.806
His^+	CN	2.558

To address the nature of the band, we classify the PPP Fock matrix eigenstates according to their localization behavior. All eigenstates around the HOMO-LUMO are found to be unambiguously localized on a single amino acid side chain. Hence, all single excitations entering the CI calculations can be classified as either being of a local type (the occupied and the vacant state are localized on the same amino acid) or of a charge-transfer type (the occupied and the vacant state are localized on different amino acids), to which we refer as being of an interacid or interacid charge transfer type. Using the CI expansion coefficients corresponding to an excitation energy E_α, we may sum over all coefficients characterizing charge transfer excitations,

$$Q_\alpha = \sum_{i,k\in CT} \left(C^k_{i\alpha}\right)^2. \tag{4.1}$$

With $0 \le Q_\alpha \le 1$, we have a measure for the charge transfer character of an excitation and are able to weight the computed spectrum accordingly. This charge transfer contribution is shown in Fig. 5 as a dashed line.

In addition, one may sum

$$P_\alpha = \sum_{i,k} \left(C^k_{i\alpha}\right)^4 \tag{4.2}$$

for a normalized eigenstate to obtain the so-called participation ratio, a quantity frequently used as a measure of localization in tight-binding theory [56–58]. In the context of CI calculations, its inverse gives the number of Slater determinants that participate in a CI eigenstate. It is close to unity for an excitation that is essentially based on a transition between two Hartree–Fock eigenstates, and equals the dimension of the CIS matrix if transitions between mean-field states are strongly mixed and all molecular orbitals contribute equally to a CI eigenstate.

With the help of these numbers, we may re-inspect the spectrum shown in Fig. 5. It can be divided into three regions: between 230 and 290 nm, we find excitations that can be traced back to Trp, Tyr and Phe molecular orbitals. Generally, P is larger than unity here, whereas Q is close to zero, indicating a coupling of single-particle excitations that is not accompanied by interacid charge transfer. In the high-energy region below 230 nm, we find large P values and a wide spectrum of Q values, suggesting a broad mixture of localized and charge transfer states from all four amino acids. At the absorption edge above 300 nm, we solely find states with $P = Q = 1$ for all ten realizations. Consequently, the band edge purely consists of interacid charge transfer absorption. Its intensity is small for two reasons. First, the number of CI eigenstates in that energy range is small. Second, despite a large charge separation, excitations show a small transition dipole moment due to small products of MO expansion coefficients or – more physically – strong spatial localization of molecular orbitals on single amino acids. The charge transfer character of the absorption edge can also be probed by simply setting all interactions between different amino acids to zero. As evident from Fig. 5, the resulting spectrum now lacks the edge. Estimating the absorbance of the band edge as 10^{-4} for human eye lense parameters, it is easy to resist the temptation to speculate about the possible biological effect of electronic interacid coupling in γD-crystallin.

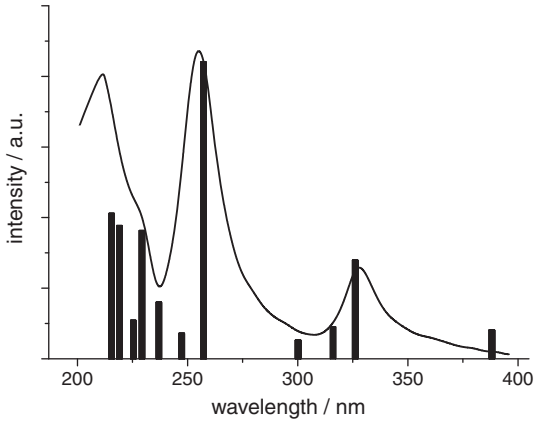

Fig. 3. Illustrating a parametrization strategy to model the electronic structures of more complex chromophores. Parameters of a semiempirical Pariser–Parr–Pople Hamiltonian have been varied to minimize the integral difference in intensity between the experimental UV absorption spectrum (solid line, [53]) and its computed counterpart on the configuration interaction singles level, here shown as an unbroadened line spectrum.

5. Computational aspects

In the following, we briefly describe the measures taken to enhance the performance of the CIS computations. As rather obvious computational bottlenecks, we have identified the computation of the CIS matrix and its diagonalization. As a reference for benchmarks, we have naively implemented Eq. (2.6) with inner loops running over atoms a and b and used the EISPACK TRED2 and TQL2 routines [59,60] to diagonalize the matrix. Execution times are given in Table 3 for a dual core Intel Xeon 5150 processor with a clock rate of 2.66 GHz, the GNU Fortran Compiler, optimization level 2, openSUSE release 10.3 Linux operating system, profiling using gprof.

Inspecting Eq. (2.6) and making use of $\Gamma = \max_{a,b,b\neq a}\gamma_{ab}$ and $0 \le c_{ia} \le 1\ \forall i,a$, it is possible to set up the following hierarchy of inequalities:

$$\begin{aligned} &\left|\sum_{a\neq b}\left(2c_{la}c_{ja}c_{ib}c_{kb} - c_{ia}c_{ja}c_{kb}c_{lb}\right)\gamma_{ab}\right| \\ \le &\Gamma\left(\sum_a 2\left|c_{la}c_{ja}\right|\sum_{b\neq a}|c_{ib}c_{kb}| + \sum_a\left|c_{ia}c_{ja}\right|\sum_{b\neq a}|c_{kb}c_{lb}|\right) \\ \le &\Gamma\left(\sum_a 2\left|c_{la}c_{ja}\right| + \sum_a\left|c_{ia}c_{ja}\right|\right) \equiv \tilde{\Gamma} \end{aligned} \tag{5.1}$$

Consequently, one obtains an upper limit to the interatomic contribution to a CIS matrix element prior the evaluation of the innermost loop. If $\tilde{\Gamma}$ is smaller than a given threshold value (taken as 10^{-4} eV), only the intraatomic contribution to the inner loop will be evaluated, here leading to a tenfold speedup in the CIS matrix element computation.

We further note that long-range charge transfer excitations do not contribute to the absorption spectrum due to their low intensity, and that their coupling to other excitations is small. Locating the donor and acceptor amino acid of a charge transfer Slater determinant, one may evaluate the edge-to-edge distance between the two amino acids, R_{DA} and discard the Slater determinant in the CIS expansion if R_{DA} is larger than a conservative distance cutoff, here set to be 8 Å. As a consequence, the number of Slater determinants is reduced from 9856 to 1302 for the first of the γD-crystallin models, and the corresponding computer times are reduced by a factor of 30 and 250 for the CIS matrix computation and diagonalization, respectively. A further reduction of the diagonalization effort is possible once an iterative eigenvalue and eigenvector Lanczos algorithm is applied

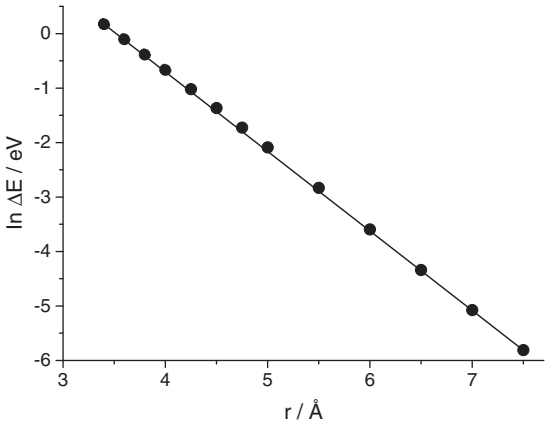

Fig. 4. Parametrization of the long-range electronic coupling parameters between different amino acids. Between 3.4 and 7.5 Å, the splittings between the two lowest π orbitals of a stacked arrangement of two phenylalanine side chains have been computed using ab initio density functional theory (●). Parameters of the semiempirical Pariser–Parr–Pople model have been varied until this model reproduces the exponential decay of the electronic coupling between the amino acids (solid line as an interpolation to the Pariser–Parr–Pople couplings).

[61–63] This statement also holds if a larger amount of Slater determinants is considered, provided that the CIS matrix is kept sparse by introducing an appropriate cutoff, here taken as 10^{-4} eV.

Enhancing the performance now permits the study of protein $\pi-\pi^*$ excitations for a variety of system realizations or parameters. As an example, we look at the excitation energy as a function of the effective dielectric constant, ε, that determines the magnitude of coupling between the π electron system to charged protein atoms and between π electrons localized on different aromatic acids. Our motivation for studying this effect is as follows. All excitations are vertical, the geometry of the protein and the water shell are fixed, whereas the electronic environment responds immediately to a redistribution of charges induced by the excitation. With time proceeding, not only the electronic degrees of freedom, but also their nuclear counterparts respond to the charge shifts induced by an excitation, these are likely to be particularly large for excitations that exhibit a strong charge transfer character. As a simple collective representative of the degrees of freedom responding to charge shifts one may consider the dielectric constant, ε. Upon excitation, the π electrons only experience the electronic contribution to ε, whereas an increasing number of degrees of freedom couples to the excitation

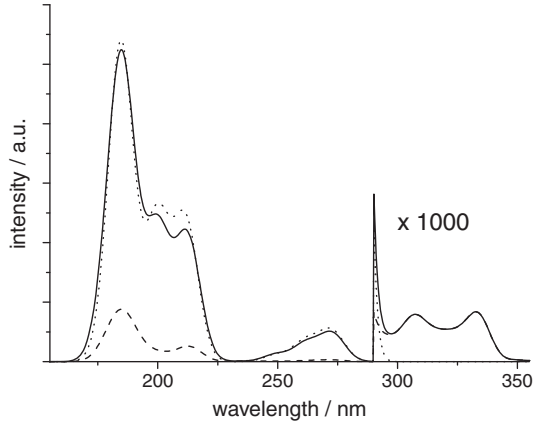

Fig. 5. Absorption spectrum, as accumulated from ten model realizations of the eye lense protein γD-crystallin (solid line) calculated using the Pariser–Parr–Pople Hamiltonian. Charge transfer between different aromatic amino acids contributes significantly to the total absorption. This contribution is shown as a dashed line. In the hypothetical absence of electronic coupling between different amino acids, the spectrum is represented by a dotted line.

Table 3
Benchmarks, computer time in seconds for the generation of the configuration interaction matrix and its diagonalization, CI matrix dimension n_{CI} and number of eigenpairs n_E. Model a — naive computation of all CI matrix elements and Givens–Householder plus QL algorithm diagonalization. Model b — as a, making use of estimates for off-diagonal CI matrix elements. Model c — as b, introduction of a radial cutoff for interchromophore charge-transfer excitations. Model d — as c, Lanczos diagonalization algorithm. Model e,f — as b, Lanczos diagonalization algorithm.

Model	CI matrix setup	CI matrix diagonalization	n_{CI}	n_E
a	3673	2792	9856	9856
b	377	2717	9856	9856
c	12.8	11.0	1302	1302
d	10.0	2.95	1302	225
e	276	72.6	9856	833
f	275	4.63	9856	14

with an increasing delay time, finally leading to the static limit of combined electronic and ionic contributions to ε. Consequently, one gets a rough idea about the evolution of selected excitations with time by increasing ε. As typical examples, we show the dependence of four charge transfer states upon ε in Fig. 6. All have His^+ as an electron acceptor, all exhibit a small intensity, but may be operative as charge transfer exciton trapping states. As a function of ε, their energy stays constant ($Tyr29^+$-His23), increases (two $Phe173^+$-His84 excitations) or decreases ($Trp43^+$-His84), with the participating amino acid residues shown in Fig. 1. Hence, PPP CIS eigenstates may exhibit avoided crossings at particular energy and ε values. The corresponding level splittings can be computed by a closer inspection of the avoided crossings, they enter transition probabilities for charge transfer theories based upon the concepts of Marcus or other Golden Rule expressions. For the four crossings shown in Fig. 6, they are all smaller than the limit given by the numerical accuracy, $\sim 10^{-8}$ eV.

6. Conclusions

With application to γD-crystallin, an eye-lense protein, we have presented a semiempirical approach that focusses on the π electron subsystem of aromatic amino acid residues. A Pariser–Parr–Pople Hamiltonian is coupled to the nuclear charges of the protein and to its aqueous environment, the resulting mean-field equations are solved self-consistently at the Hartree–Fock level. Using configuration interaction single excitation theory, it is possible to accurately recover the low-energy experimental $\pi-\pi$ excitations of isolated residues and the ab initio couplings between these entities. The strategy of a system-specific reparametrization may be straightforwardly adapted to larger building blocks of biological light harvesting complexes or

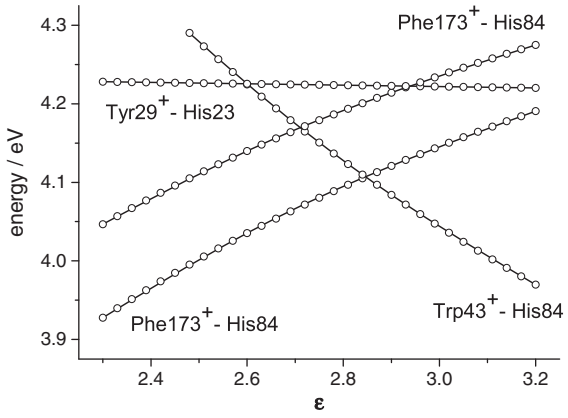

Fig. 6. According to the model computations, low-level excitation lines are characterized by charge transfer between different aromatic amino acids. Increasing the dielectric constant after the excitation mimics the onset of the protein and solvent dielectric response of these excitations with time, shown here for four sample lines.

organic solar cells, with few exceptions such as the carotenoids that are believed to require a higher level of theory, such as considering doubly-excited Slater determinants [43,64].

The use of a population analysis to identify low-intensity charge transfer processes, upper limits to CIS matrix elements and sparse matrix techniques enable an efficient computation of the spectrum of a protein–water system containing 112 π electron pairs and 10^4 atoms. As a consequence, fluctuation effects can be taken into account by the quantum mechanical post-processing of a large number of experimental or simulated structures. Orientational effects are effortlessly considered, and we are not restricted to a single excitation per residue or chromophore. As the most important physical effect, interacid electronic coupling induces the formation of a band edge consisting of charge transfer excitations.

Whereas we have gained a first glance at the evolution of excitations by varying the dielectric constant, it is highly desirable to represent the polarizable environment by a model that exhibits the same spatial resolution as the electronic Hamiltonian and the classical force field. Here, a Langevin dipole model may become useful [65]. Naturally, $n-\pi$ transitions cannot be considered within the model of Pariser, Parr and Pople, yet semiempirical treatments of all valence electrons such as the widely used ZINDO [66] method may be coupled to classical force fields in a similar fashion.

According to our computations, charge transfer states of γD-crystallin only exhibit a weak absorption intensity in the low energy ultraviolet range and do not offer a direct additional protective function beyond the absorbance of individual amino acids. However, they act as low-lying trap states, which can be populated following relaxation processes after the initial absorption of an UV photon. From the trap states, a red-shifted and less harmful photon may be reemitted with a small transition probability. As an alternative, the thus created radical pair may undergo consecutive chemical reactions with the possibility of irreversibly changing the properties of the aromatic amino acids forming the pair, which in turn may shift the delicate balance of interactions within the eye lense crystallin network towards aggregation. Mutations introduced to γD-crystallin are found to induce cataracts [67], and they are assumed to act by changing the noncovalent binding properties within the eye lense protein β/γ-crystallin network [67]. We suggest the working hypothesis that such changes, initiated by charge transfer radical states, may also underlie the formation of eye lense cataracts upon UV-B irradiation, as e.g. observed by Taylor et al. [68]. Identification of the protein defects and their simulation in a large-scale simulation of protein aggregation may become the first step towards the verification or falsification of the proposed mechanism.

In a more general context, the procedures outlined in this manuscript may help to identify photoexcited charge transfer states and paths in biochemical systems such as light harvesting complexes or blue-light receptors (DNA repair enzymes, cryptochromes), or in man-made energy-converting systems in the field of artificial photosynthesis and organic photovoltaics.

Acknowledgements

It is a pleasure to thank A. Blumen, L. Mühlbacher, O. Mülken, M. Thorwart and P.J. Walla for the fruitful discussions and helpful comments. We gratefully acknowledge the use of J. Ponder's Tinker package.

References

[1] G.R. Fleming, R. van Grondelle, Phys. Today 47 (1994) 48.
[2] T. Ritz, A. Damjanovic, K. Schulten, Chemphyschem 3 (2002) 243.
[3] R.E. Blankenship, Molecular Mechanisms of Photosynthesis, Blackwell, Oxford, 2002.
[4] S. Weber, BBA Bioenerg. 1707 (2005) 1.
[5] R.R. Birge, Annu. Rev. Phys. Chem. 41 (1990) 683.
[6] W.D. Hoff, K.-H. Jung, J.L. Spudich, Annu. Rev. Biophys. Biomol. Struct. 26 (1997) 223.
[7] T. Biskup, E. Schleicher, A. Okafuji, G. Link, K. Hitomi, E.D. Getzoff, S. Weber, Angew. Chem. Int. Ed. 48 (2009) 404.
[8] SchultenK. , Festkörperprobleme 22, in: TreuschJ. (Ed.), Vieweg, Braunschweig, 1982, p. 61.
[9] T. Ritz, S. Adem, K. Schulten, Biophys. J. 78 (2000) 707.
[10] T. Ritz, P. Thalau, J.B. Phillips, R. Wiltschko, W. Wiltschko, Nature 429 (2004) 177.
[11] S. Günes, H. Neugebauer, N.S. Sariciftci, Chem. Rev. 107 (2007) 1324.
[12] M.P. Fülscher, L. Serrano-Andres, B.O. Roos, J. Am. Chem. Soc. 119 (1997) 6168.
[13] C. Hättig, in: J. Grotendorst, S. Blügel, D. Marx (Eds.), Computational Nanoscience: Do It Yourself!, John von Neumann institute Jülich NIC Series, 31, 2006, p. 245.
[14] A. Dreuw, M. Head-Gordon, Chem. Rev. 105 (2005) 4009.
[15] J. Neugebauer, J. Phys. Chem. B 112 (2008) 2207.
[16] D. Möbius, Adv. Mater. 7 (1995) 437.
[17] T. Kobayashi, J-aggregates, World Scientific, Singapore, 1996.
[18] O. Mülken, L. Mühlbacher, T. Schmid, A. Blumen, Phys. Rev. E 81 (2010) 41114.
[19] P. Nalbach, M. Thorwart, J. Chem. Phys. 132 (2010) 194111.
[20] K.K. Sharma, P. Santhoshkumar, BBA 1790 (2009) 1095.
[21] L. Takemoto, C.M. Sorensen, Exp. Eye Res. 87 (2008) 496.
[22] H. Bloemendal, W. de Jong, R. Jaenicke, N.H. Lubsen, C. Slingsby, A. Tardieu, Progr. Biophys. Mol. Biol. 86 (2004) 407.
[23] H. Bloemendal, CRC Crit. Rev. Biochem. 12 (1982) 1.
[24] A. Laganowsky, D. Eisenberg, Protein Sci. 19 (2010) 1978.
[25] A. Laganowsky, J.L. Benesch, M. Landau, L. Ding, M.R. Sawaya, D. Cascio, Q. Huang, C.V. Robinson, J. Horwitz, D. Eisenberg, Protein Sci. 19 (2010) 1031.
[26] C. Bagneris, O.A. Bateman, C.E. Naylor, N. Cronin, W.C. Boelens, N.H. Keep, C. Slingsby, J. Mol. Biol. 392 (2009) 1242.
[27] V. Nalini, B. Bax, H.P. Driessen, D.S. Moss, P.F. Lindley, C. Slingsby, J. Mol. Biol. 236 (1994) 1250.
[28] S. Najmudin, V. Nalini, H.P. Driessen, C. Slingsby, T.L. Blundell, D.S. Moss, P.F. Lindley, Acta Crystallogr. D 49 (1993) 223.
[29] L.M. Sampaleanu, F. Vallee, C. Slingsby, P.L. Howell, Biochemistry 40 (2001) 2732.
[30] O.A. Bateman, A.G. Purkiss, R. van Montfort, C. Slingsby, C. Graham, G. Wistow, Biochemistry 42 (2003) 4349.
[31] J. Wang, X. Zuo, P. Yu, I.J. Byeon, J. Jung, X. Wang, M. Dyba, S. Seifert, C.D. Schweiters, J. Qin, A.M. Gronenborn, Y.X. Wang, J. Am. Chem. Soc. 131 (2009) 10507.
[32] J. Wang, P. Cieplak, P.A. Kollman, J. Comp. Chem. 21 (2000) 1049.
[33] P. Ren, J.W. Ponder, J. Phys. Chem. B 107 (2003) 5933.
[34] A. Huizinga, A.C.C. Bot, F.F.M. de Mul, G.F.J.M. Gijs, J. Greve, Exp. Eye Res. 48 (1989) 487, and references therein.
[35] V. Hornak, R. Abel, R. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Proteins: Struct. Funct. Bioinf. 65 (2006) 712.
[36] R. Pariser, R.G. Parr, J. Chem. Phys. 21 (1953) 466.
[37] R. Pariser, R.G. Parr, J. Chem. Phys. 21 (1953) 767.
[38] J.A. Pople, Proc. Phys. Soc. A 68 (1955) 81.
[39] A. Warshel, M. Karplus, J. Am. Chem. Soc. 94 (1972) 5612.
[40] J. Lobaugh, P.J. Rossky, J. Phys. Chem. A 103 (1999) 9432.
[41] J. Lobaugh, P.J. Rossky, J. Phys. Chem. A 104 (2000) 899.
[42] A. Takahashi, S. Mukamel, J. Chem. Phys. 100 (1994) 2366.
[43] P. Tavan, K. Schulten, Phys. Rev. B 36 (1987) 4337.
[44] K. Nishimoto, N. Mataga, Z. Phys. Chem. 12 (1957) 335.
[45] J.C. Slater, G.F. Koster, Phys. Rev. 94 (1954) 1498.
[46] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. 79 (1983) 926.
[47] J. Ladik, K. Appel, Theor. Chim. Acta 4 (1966) 132.
[48] T. Cramer, S. Krapf, T. Koslowski, J. Phys. Chem. B 108 (2004) 11812.
[49] T. Cramer, T. Steinbrecher, A. Labahn, T. Koslowski, PCCP 7 (2005) 4039.
[50] B.O. Roos, K. Andersson, M.P. Fülscher, Chem. Phys. Lett. 192 (1992) 5.
[51] J. Lorentzon, P.-Å. Malmqvist, M. Fülscher, B.O. Roos, Theor. Chem. Acc. 91 (1995) 91.
[52] L. Serrano-Andrés, B.O. Roos, J. Am. Chem. Soc. 118 (1996) 185.
[53] J.P. Hare, H.W. Kroto, R. Taylor, Chem. Phys. Lett. 177 (1991) 394.
[54] The following parameters apply: $\gamma = 10.53$ eV, $\beta_0 = -1.60$ eV (nearest neighbours), $\beta_1 = 0.45$ eV (next-nearest neighbours, no electron–phonon coupling)
[55] This corresponds to the smallest value for which the SCF procedure converged.
[56] P. Dean, R.J. Bell, Discuss. Faraday Soc. 50 (1970) 55.
[57] T. Koslowski, W. von Niessen, Phys. Rev. B 42 (1990) 10342.
[58] T. Koslowski, W. von Niessen, J. Phys. Condens. Matter 4 (1992) 6109.
[59] B.T. Smith, J.M. Boyle, J.J. Dongerra, B.S. Garbow, Y. Ikebe, V.C. Uleba, C.B. Moler, Matrix Eigensystem Routines — EISPACK Guide, Springer, Berlin-Heidelberg-New York, 1976.
[60] B.S. Garbow, J.M. Boyle, J.J. Dongerra, C.B. Moler, Matrix Eigensystem Routines — EISPACK Guide Extension, Springer, Berlin-Heidelberg-New York, 1977.
[61] C. Lanczos, J. Res. NBS B 45 (1950) 225.
[62] J.K. Cullum, R. Willoughby, Lanczos Algorithms For Large Symmetric Eigenvalue Problems I + II, Birkhäuser, Boston-Basel-Stuttgart, 1985.
[63] T. Koslowski, W. von Niessen, J. Comp. Chem. 14 (1993) 769.
[64] I. Ohmine, M. Karplus, K. Schulten, J. Chem. Phys. 68 (1978) 2298.
[65] J. Florian, A. Warshel, J. Phys. Chem. B 101 (1997) 5583.
[66] M.A. Thompson, M.C. Zerner, J. Am. Chem. Soc. 113 (1991) 8210.
[67] K. Wang, C. Cheng, L. Li, H. Liu, Q. Huang, C. Xia, K. Yao, P. Sun, J. Horwitz, X. Gong, Investig. Ophthalmol. Vis. Sci. 48 (2007) 3719.
[68] H.R. Taylor, S.K. West, F.S. Rosenthal, B. Muñoz, H.S. Newland, H. Abbey, E.A. Emmett, N. Engl. J. Med. 319 (1988) 1429.