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.

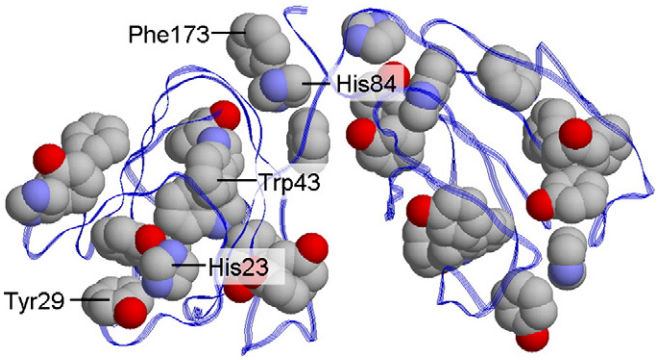

Fig. 1. Cartoon model of γD-crystallin from *H. sapiens* from a combined 2D NMR and neutron diffraction study [31], protein database entry 2KLJ. The strands indicate the protein backbone, spheres represent the nonhydrogen side chain atoms of aromatic amino acids. Carbons are plotted in gray, oxygens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