Modeling structure and absorption spectra of the bacteriophytochrome-based fluorescent protein IFP1.4

Igor V. Polyakov, [a,b] Bella L. Grigorenko [a,b] and Alexander V. Nemukhin [*,a,b]

[a.] Department of Chemistry, Lomonosov Moscow State University, Leninskie Gory 1/3, 119991 Moscow, Russian Federation

[b.] Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Kosygina 4, 119334 Moscow, Russian Federation

* Corresponding author:

Prof. Alexander V. Nemukhin, Chemistry Department, Lomonosov Moscow State University, Leninskie Gory 1/3, Moscow, 119991, Russian Federation

Phone: +7-495-939-10-96

E-mail: anemukhin@yahoo.com; anem@lcc.chem.msu.ru

Abstract

We report the results of quantum mechanics/molecular mechanics (QM/MM) simulations of structures and absorption spectra of the fluorescent protein IFP1.4 engineered from the chromophore-binding domain of *Deinococcus radiodurans* (*Dr*CBD). In this work, we focus on different protonation states of the biliverdin chromophore in the red-absorbing form of the protein. To this goal, the protein with the all-protonated chromophore as well as the structures obtained by removal of protons from the biliverdin pyrrole rings to a suitable acceptor within the system are considered. Several quantum chemistry methods to compute the $S_0 \rightarrow S_1$ excitation energies are used in the QM part to estimate shifts in the absorption band maxima upon chromophore deprotonation.

Introduction

An interest to proteins engineered from bacterial phytochromes is explained by their use as fluorescent markers for *in vivo* imaging, because their absorption and emission bands fall into the near-infrared optical transparency window of biological materials.[1,2] Detailed characterization of structures and spectral properties of the corresponding chromophore-binding protein domains, which

carry the linear tetrapyrrole biliverdin IXα (BV) as a chromophore, is a necessary step in the design of novel efficient markers.

One of the first proteins in this family IFP1.4 (infrared fluorescent protein, version 1.4) [3-5] was created from the chromophore-binding domain of *Deinococcus radiodurans* (*Dr*CBD). The parent bacteriophytochrome called *Dr*BphP, containing this domain, has attracted considerable attention as a suitable template for constructing fluorescent markers. [6-12] The IFP1.4 variant was engineered from the wild type *Dr*CBD by introducing a set of point mutations, which consequently increased the fluorescence quantum yield from ~2 % to ~7÷9 %. The reasons for these replacements (D207H/I208T/A288V/M54V/G119A/V186M/L311K/T135L/L195M/H196Q/Y307E/L314G) are explained in Refs 3-5. It seems that the key replacement, accounting for an increased quantum yield in IFP1.4, is D207H, which prevents conformational changes between the conventional phytochrome forms, the red-absorbing Pr and far-red absorbing Pfr ones.[9] The paper [10] demonstrated that the D207H replacement in *Dr*CBD did not have a considerable effect on optical spectra of the biliverdin chromophore.

Practical applications of IFP1.4 for bioimaging is described in Ref 3. This protein is monomeric, its fluorescence is stable over a wide pH range from 5 to 9. IFP1.4 has an absorption band maximum at 684 nm (the corresponding energy gap is 1.81 eV) and emission band maximum at 708 nm (1.75 eV), respectively. Presently, several crystal structures of *Dr*BphP (or *Dr*CBD) are available in the Protein Data Bank. The structure PDB ID 2O9C solved at the 1.45 Å resolution describes the wild-type *Dr*CBD, [7] while the IFP1.4 mutant is characterized by the structures PDB ID 2O8G (1.65 Å) [4] and PDB ID 5AJG (1.11 Å). [5] We also mention, for the completeness, that the structures of *Dr*BphP, containing two protein domains PHY and CBD, were recently deposited to PDB with the codes PDB ID 4O01 (3.24 Å) and PDB ID 4O0P (3.8 Å) for the wild-type phytochrome[12] and PDB ID 5C5K (3.31 Å) for the F469W mutant with the enhanced Pfr stability. [11]

Although quantum-based computational studies of phytochrome properties are fairly active in the recent years, [13-27] there are only two papers describing the IFP1.4 variant. Felix et al.[5] used classical molecular dynamics simulations to compare structural features, including protonation states of the chromophore and distributions of water molecules, in the parent *Dr*CBD protein and in the IFP1.4 variant. Polyakov et al. [23] constructed model systems referring to the Pr and the hypothetical Pfr form of IFP1.4 and estimated positions of the Q-band and the Soret bands for these structures.

A series of papers by Falklöf and Durbeej[16,18,19-22] described properties of the related protein *Dr*CBD on the base of the crystal structure PDB ID 2O9C. The computational work of Modi et al.[27]

considers a large molecular system composed of the PHY and CBD domains from *Dr*BphP on the base of the low resolution crystal structures PDB ID 4O01 and PDB ID 4O0P. In several theoretical papers, the authors addressed the important question on possible protonation states of the biliverdin chromophore in the protein. [5, 16, 27]

The goal of the present work is to provide a comprehensive description of the ground state structure and absorption spectra of IFP1.4 for different protonation states of the chromophore. Unlike previous works, we consider removal of protons from the pyrrole rings of the chromophore to a suitable proton acceptor within a model system. This allows us to range the corresponding species by energy.

Models and Methods

Our starting point was the crystal structure PDB ID 5AJG of IFP1.4 solved with the high resolution 1.11 Å.[5] In simulations, hydrogen atoms were added by assuming positive charges of the side chains Lys, Arg and negative charges of Glu and Asp. The protonation state of His was manually suggested for every residue according to its local environment, in particular, taking into account the recipes formulated in Ref 5 for His207 and His260.

At preliminary stages, the system was inserted into the TIP3P water box (resulting in the size ~94·93·93 $Å^3$) and neutralized by adding seven Na^+ ions. To carry out classical molecular dynamics (MD) simulations, the AMBER force field parameters were applied. AMBER-compatible parameters for the biliverdin molecule were taken from Ref 27. The GROMACS software package (http://www.gromacs.org) was applied to run MD trajectories. Snapshots from a MD trajectory were taken to minimize the energy and to generate staring points for QM/MM calculations.

A very large QM part (165 atoms) included the BV chromphore, the side chains of the important residues (Cys24, Ser206, His207, Thr208, Arg254, His260, Ser274, His290) and 6 water molecules. We applied the NWChem program [28] for QM/MM calculations. Energies and forces in the QM subsystem were computed using the density functional theory with PBE0 functional [29] with the D3 correction [30] and the 6-31G* basis set. The MM subsystems was modeled with the AMBER force field parameters.

Vertical excitation energies and corresponding oscillator strengths were calculated using the TD-DFT, SOS-CIS(D) [31] and XMCQDPT2//CASSCF [32] approaches. In the latter case, the following technical details should be specified: number of electrons and number of orbitals in the active space (e.g., CAS(16/12)), number of states in density averaging in CASSCF (e.g., SA10-CAS), number of

states in the effective Hamiltonian (e.g., XMC20). Accordingly, we use here notations like XMC20//SA10-CAS(16/12) to specify a calculation protocol. TD-DFT and SOS-CIS(D) calculations were performed with QChem (https://www.q-chem.com), and XMCQDPT2 calculations were carried out with the Firefly quantum chemistry package (Firefly version 8, www http://classic.chem.msu.su/gran/firefly/index.html). The cc-pVDZ basis set was used in the transition energy calculations.

Results and discussion

First, we show the results for the IFP1.4 structures in the ground electronic state S_0. Fig. 1 illustrates the configuration of the QM-part in the structure with all protonated pyrrole nitrogen atoms in BV.

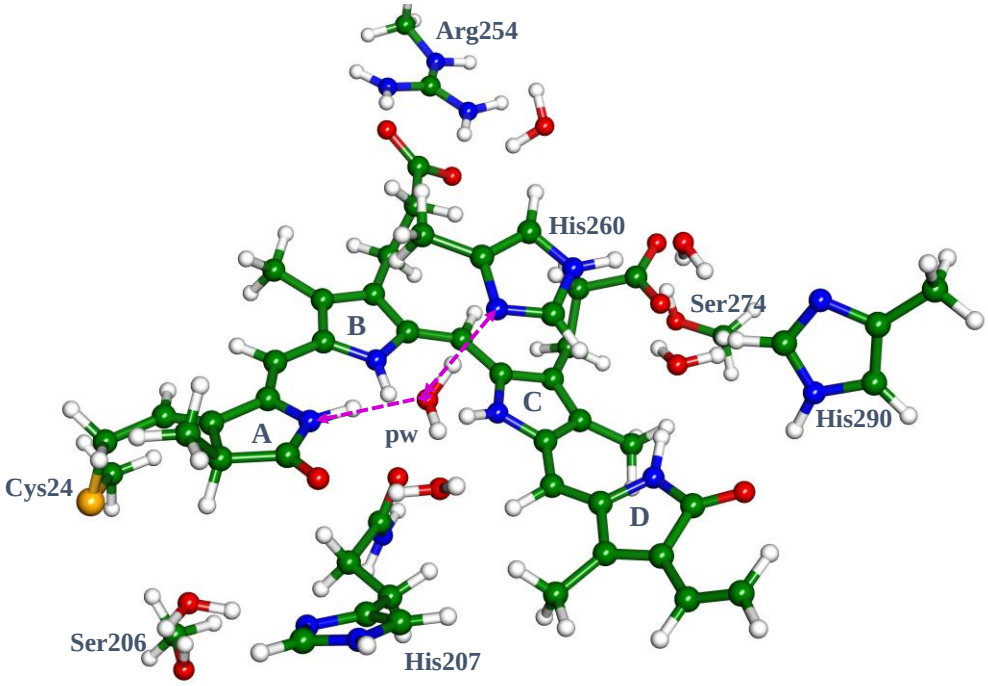

Figure 1. QM-part of the calculated structure of the model system with the all-protonated form of biliverdin. Here and in other figures, carbon atoms are shown green, oxygen – red, nitrogen – blue, hydrogen – white. Magenta arrows illustrate the proton transfer route to deprotonate the ring A, moving the proton to the Nδ atom of His260 via the pyrrole water molecule 'pw'.

To create model systems with the deprotonated pyrrole rings we moved protons from the corresponding ring to an appropriate proton acceptor. As such, we considered the imidazole ring of His260 when deprotonating A, B and C pyrrole rings, and the carboxyl group of the hydrocarbon tail

attached to the C-ring (see Fig. 1). The nearby water molecules, e.g., the pyrrole water 'pw', could serve as suitable proton shuttles. The obtained model systems with the deprotonated BV are designated here as BCD, ACD, ABD and ABC, respectively. Table 1 lists the computed energies of these structures illustrated in Fig. 2.

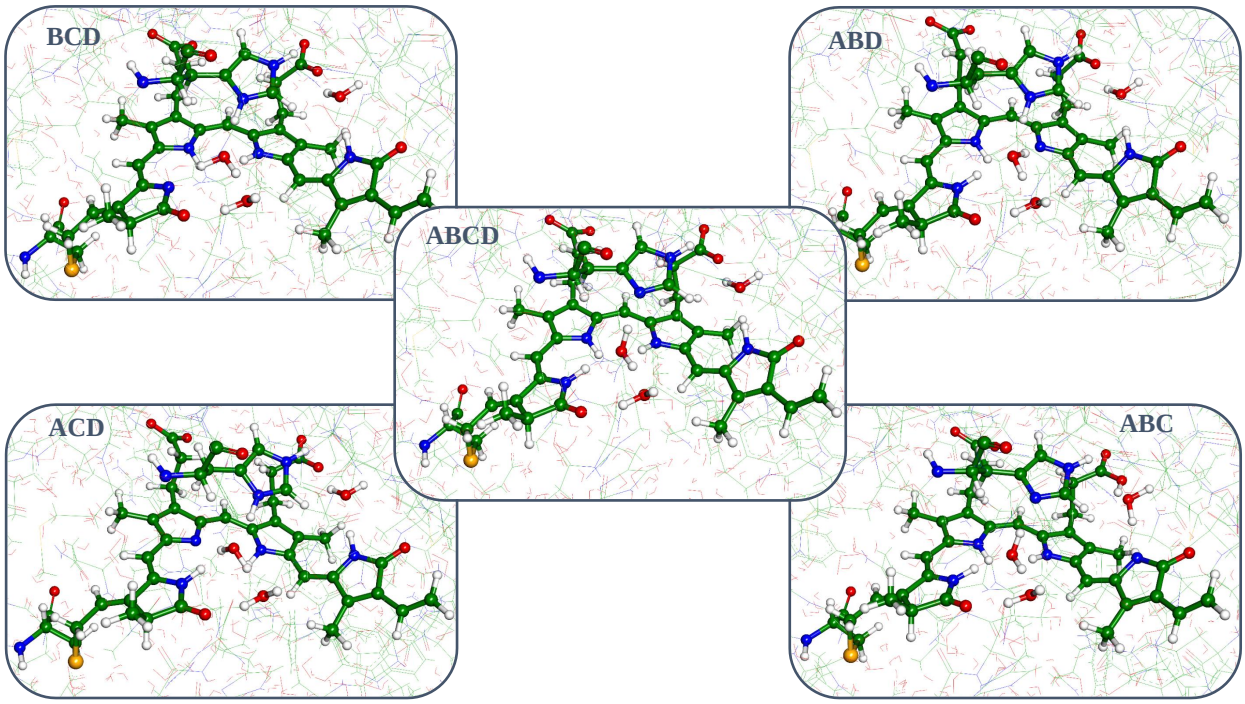

Figure 2. Structures of BV inside the protein in different protonation states.

Table 1. Energies of the IFP1.4 models for various protonation states of biliverdin (see Fig.2).

Energy	All-protonated ABCD	Deprotonated A BCD	Deprotonated B ACD	Deprotonated C ABD	Deprotonated D ABC
QM/MM energy, au	-4315.31547	-4315.30745	-4315.30574	-4315.30585	-4315.29255
Relative energy, kcal/mol	0	5	6	6	14

From these results, we confirm that the all-protonated structure of BV is the most likely candidate for the ground state structure of IFP1.4 (and, presumably, for *Dr*CBD, *Dr*BphP), in agreement with previous computational papers. [5,16,27] We emphasize that the present paper presents

the first computational study, allowing one to range energetically the systems with the different protonation BV forms. The estimated energy gaps of 5-6 kcal/mol between the all-protonated and the A-, B-, C-deprotonated forms are small enough to assume fractions of the deprotonated BV in the experiments with the *Dr*CBD-related proteins. We note that Ref 22 suggests chromophore's C-deprotonation in *Dr*CBD upon electronic excitation, while Ref 27 suggests that there is a fraction of the *Dr*BphP protein with the D-deprotonated chromophore in the ground state.

Table 2 lists the results of calculations of the vertical $S_0 \rightarrow S_1$ excitation energies by using a variety of present-day quantum chemical approaches in the QM-part of the model system. Experimental position of the band maximum in the absorption spectrum of IFP1.4 [3] is 684 nm, which corresponds to the energy gap 1.81 eV.

Table 2. Vertical excitation energies (eV) and corresponding wavelengths (nm) in parentheses for the various protonation states of biliverdin in the IFP1.4 protein.

Method	ABCD	BCD	ACD	ABD	ABC
TD-DFT(wB97X-D3)	2.22 (559)	2.24 (554)	2.38 (521)	2.33 (532)	2.10 (590)
TD-DFT(rCAM-B3LYP)	2.20 (564)	2.27 (546)	2.64 (470)	2.38 (521)	2.11 (588)
TD-DFT(B3LYP)	2.00 (620)	1.99 (623)	2.08 (596)	2.06 (602)	1.91 (649)
TD-DFT(BLYP) [a]	1.88 (660)	1.89 (656)	1.94 (639)	1.91 (649)	1.80 (689)
XMC15//SA8-CAS(16/12)	1.76 (705)	1.61 (770)	2.19 (566)	1.85 (670)	---
XMC6//SA2-CAS(16/12)	1.73 (717)	1.67 (743)	2.10 (590)	1.84 (674)	1.79 (734)[b]
SOS-CIS(D)	1.79 (693)	1.88 (680)	2.26 (549)	2.11 (588)	1.68 (738)

[a] The bright state is not the second singlet state.
[b] The CAS(6/6) expansion is used in the ABC case.

The values for the all-protonated structure ABCD computed with the SOS-CIS(D), TD-DFT(BLYP) and XMCQDPT2 approaches are within the 0.05 eV discrepancy with the experimental data. We find that performance of other functionals in TD-TDF is less successful.

Comparison of the computed shifts in absorption band maximum upon deprotonation of the pyrrole rings of biliverdin shows that the SOS-CIS(D) and TD-DFT(BLYP) variants lead to a consistent picture, predicting the blue shifts for deprotonation of A-, B- and C-rings and the red shift for D-deprotonation.

The TD-DFT(B3LYP) approach and the most advanced XMCQDPT2 method predict a slight red shift for the A- and D-deprotonation, while the B- and C-deprotonated models experience a blue shift. Overall, agreement between the results obtained with the GGA functional BLYP, the hybrid functional B3LYP, the long-ranged corrected functionals rCAM-B3LYP and wB97X-D3, and the perturbation theory based methods SOS-CIS(D) and XMCQDPT2 is modest for this system.

Conclusions

We highlight the following results of the described QM/MM-based simulations of IFP1.4. The calculations show that deprotonation of the A, B and C pyrrole rings of the biliverdin chromophore leads to the systems, which are fairly close in energy (5-6 kcal/mol) to the lowest energy all-protonated species. Practical quantum chemistry methods SOS-CIS(D) and TD-DFT(BLYP) in the QM-part used to compute the $S_0 \rightarrow S_1$ excitation energies in the system predict a consistent picture of the shifts in the absorption band maxima upon chromophore deprotonation.

Acknowledgements

This work was supported by the grant from the Russian Science Foundation (project #17-13-01051) to IVP and BLG. The research was carried out using the equipment of the shared research facilities of HPC computing resources at the Lomonosov Moscow State University supported by the project RFMEFI62117X0011. The use of supercomputer resources of the Joint Supercomputer Center of the Russian Academy of Sciences is also acknowledged.

References

1 K.G. Chernov, T.A. Redchuk, E.S. Omelina and V. V. Verkhusha, *Chem. Rev.*, 2017, **117**, 6423–6446.

2 M.E. Auldridge and K.T. Forest, *Crit. Rev. Biochem. Molec. Biol.*, 2011, **46**, 67–88.

3 X. Shu, A. Royant, M.Z. Lin, T.A. Aguilera, V. Lev-Ram, P.A. Steinbach PA and R.Y. Tsien, *Science*, 2009, **324**, 804–807.

4 S. Bhattacharya, M.E. Auldridge, H. Lehtivuori, J.A. Ihalainen and K.T. Forest, *J. Biol. Chem.*, 2014, **289**, 32144−32152.

5 M. Feliks, C. Lafaye, X. Shu, A. Royant and M. Field, *Biochemistry*, 2016, **55**, 4263–4274.

6 J. Wagner, J. Brunzelle, K. Forest and R. Vierstra, *Nature*, 2005, **438**, 325−331.

7 J. Wagner, J. Zhang, J. Brunzelle, R. Vierstra and K. Forest, *J. Biol. Chem.*, 2007, **282**, 12298−12309.

8 J.R. Wagner, J. Zhang, D. von Stetten, M. Günther, D.H. Murgida, M.A. Mroginski, J.M. Walker, K.T. Forest, P. Hildebrandt and R.D. Vierstra, *J. Biol. Chem.*, 2008, **283**,12212– 12226.

9 M.E. Auldridge, K.A. Satyshur, D.M. Anstrom and K. Forest, *J. Biol. Chem.*, 2012, **287**, 7000-7009.

10 H. Lehtivuori, I. Rissanen, H. Takala, J. Bamford, N.V. Tkachenko and J.A. Ihalainen, *J. Phys. Chem. B*, 2013, **117**, 11049-11057.

11 E.S. Burgie, J. Zhang and R.D. Vierstra, *Structure*, 2016, **24**, 448–457.

12 H. Takala, A. Björling, O. Berntsson, H. Lehtivuori, S. Niebling, M. Hoernke, I. Kosheleva,\ R. Henning, A. Menzel, J.A. Ihalainen, and S. Westenhoff, *Nature*, 2014, **509**, 245-248.

13 S. Gozem, H.L. Luk, I. Schapiro and M. Olivucci, *Chem. Rev.*, 2017, **117**, 13502–13565.

14 A.V. Nemukhin, B.L. Grigorenko, M.G. Khrenova and A.I. Krylov, *J. Phys. Chem. B*, 2019, **123**, 6133-6149.

15 J. Hasegawa, M. Isshiki, K. Fujimoto and H. Nakatsuji, *Chem. Phys. Lett.*, 2005, **410**, 90–93.

16 B. Durbeej, O.A. Borg, L.A. Eriksson, *Chem. Phys. Lett.* 2005, **416**, 83–88.

17 M.A. Mroginski, D.H. Murgida and P. Hildebrandt, *Acc. Chem. Res.* 2007, **40**, 258-266.

18 B. Durbeej, *Phys. Chem. Chem. Phys.* 2009, **11**, 1354-1361.

19 R.A. Matute, R. Contreras, G. Pérez-Hernández and L. González, *J. Phys. Chem.* B, 2008, **112**, 16253–16256.

20 O. Falklöf and B. Durbeej, *J. Comput. Chem.*, 2013, **34**, 1363–1374.

21 O. Falklöf and B. Durbeej, *Chem. Phys. Chem.*, 2016, **17**, 954–957.

22 O. Falklöf and B. Durbeej, *Chem. Photo. Chem.*, 2018, **2**, 453–457.

23 I.V. Polyakov, B.L. Grigorenko, V.A. Mironov and A.V. Nemukhin, *Chem. Phys. Lett.*, 2018, **710**, 59–63.

24 M.G. Khrenova, A.M. Kulakova and A.V Nemukhin, *Org. Biomol. Chem.*, 2018, **16**, 7518–7529.

25 C. Wiebeler, A. Gopalakrishna Rao, W. Gärtner and I. Schapiro, *Angew. Chem. Int. Ed.*, 2019, **58**, 1934–1938.

26 C. Wiebeler and I. Schapiro, *Molecules*, 2019, **24**, 1720.

27 V. Modi, S. Donnini, G. Groenhof and D. Morozov, *J. Phys. Chem. B.*, 2019, **123**, 2325-

2334.

28 M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus and W. A. de Jong, *Comput. Phys. Commun.*, 2010, **181**, 1477–1489.

29 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158-6170.

30 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104-1-154104-19.

31 Y.M. Rhee and M. Head-Gordon, *J. Phys. Chem. A*, 2007, **111**, 5314–5326.

32 A.A. Granovsky, *J. Chem. Phys.*, 2011, **134**, 214113.