PCCP

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/c9cp01533g

Polarizable embedding for simulating redox potentials of biomolecules†

Ruslan N. Tazhigulov, [a] Pradeep Kumar Gurunathan, [b] Yongbin Kim, [b] Lyudmila V. Slipchenko [b] and Ksenia B. Bravaya *[a]

Received 19th March 2019,
Accepted 15th April 2019

DOI: 10.1039/c9cp01533g

rsc.li/pccp

Redox reactions play a key role in various biological processes, including photosynthesis and respiration. Quantitative and predictive computational characterization of redox events is therefore highly desirable for enriching our knowledge on mechanistic features of biological redox-active macromolecules. Here, we present a computational protocol exploiting polarizable embedding hybrid quantum-classical approach and resulting in accurate estimates of redox potentials of biological macromolecules. A special attention is paid to fundamental aspects of the theoretical description such as the effects of environment polarization and of the long-range electrostatic interactions on the computed energetic parameters. Environment (protein and the solvent) polarization is shown to be crucial for accurate estimates of the redox potential: hybrid quantum-classical results with and without account for environment polarization differ by 1.4 V. Long-range electrostatic interactions are shown to contribute significantly to the computed redox potential value even at the distances far beyond the protein outer surface. The approach is tested on simulating reduction potential of cryptochrome 1 protein from Arabidopsis thaliana. The theoretical estimate (0.07 V) of the midpoint reduction potential is in good agreement with available experimental data (0.15 V).

I Introduction

Redox processes are ubiquitous in nature and industry. They play a crucial role in energy storage, photovoltaic devices, biological processes,including photosynthesis, respiration, DNA repair, magnetoreception, and many more.[1–10] A thorough discussion of different models for evaluating redox potentials for half-reactions can be found in a review by Marenich et al.[11] One can distinguish two classes of approaches for computation of redox potentials and relevant energies based on whether implicit[12–16] or explicit[17–26] solvent models are used.

Both classes of methods have been successfully used for calculating redox potentials for half-reactions and ionization energetics in homogeneous solvents.[12–25] In case of redox reactions in proteins or heterogeneous environment in general, the explicit solvent models, which can account for specific interactions, are needed.[27] Electronic embedding quantum mechanics/molecular mechanics (QM/MM) schemes combined with linear response approximation (LRA)[28–30] have been shown to reproduce the differences in redox potentials with a good accuracy,[29,31] yet computing the absolute values remains challenging. A similar strategy has been recently used to describe electron transfer and hemes' redox potentials in bacterial decaheme cytochromes.[32] Electronic embedding QM/MM has been also exploited in the framework of free energy perturbation simulations to evaluate the reduction potential of FAD in cholesterol oxidase:[33] the obtained accuracy with respect to the experimental reference was 0.8 V.

While the importance of polarization in determination of redox energetics in biomolecules has been recognized (see for example work by Zhang et al.[34] with implicit account for environment polarization through molecular fractionalization with conjugate caps (MFCC) charge scheme[35]), to our knowledge polarizable embedding models have not been used for calculating redox free energies and absolute redox potentials of proteins. Thus, the role of polarization on predictive computation of absolute values of redox potentials for biological macromolecules is yet to be explored.

The effects of environment polarization on excitation energies of chromophores in protein matrix have been previously investigated by Beerepoot et al.[36] The authors showed that the excitation energies of chromophores in green fluorescent protein (GFP) and rhodopsin converge with respect to the size of the polarizable shell at E20 Å. Note that the quantum part's charge density redistribution is less dramatic upon excitation in comparison to ionization or electron attachment, which are accompanied with the overall change of the charged state of the system, and one can therefore expect the redox potential values to be more sensitive to

[a] Department of Chemistry, Boston University, Boston, Massachusetts 02215, USA. E-mail: bravaya@bu.edu
[b] Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, USA

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9cp01533g

the environment polarization. Moreover, the excitation energies for the chromophores in homogeneous aqueous solutions converge faster with respect to the environment polarization than the excitation energies for the same chromophores in heterogeneous protein environment.[36] Thus, the effects of polarization on the computed values of redox potentials are expected to be even more pronounced in proteins than in homogeneous solutions,[21,23,24,26] where the differences in computed redox potentials in case of electronic (purely electrostatic) and polarizable embedding can be as large as 0.5 V.[26]

Another important aspect of theoretical description of redox potentials is the treatment of system size effects. QM/MM calculations of redox potentials are often performed on truncated systems of finite size.[29,31,34] This poses a question on the magnitude of the effects of neglected long-range electrostatic interactions. For example, our previous computational studies of redox potentials of small molecules in aqueous solutions indicate that the convergence of the redox potentials with respect to the solvent shell radius is not reached even at the radius of 30 Å.[26] The slow convergence was attributed to the long-range electrostatic interactions between the redox-active site and bulk environment.[26,38,39] Therefore, for protein–solvent media, the long-range static electrostatic interactions are expected to contribute to the computed quantities as well, and are needed to be properly taken into account.

Here, we explore the role of long-range electrostatic interactions and of the environment polarization by considering FAD cofactor reduction in cryptochrome 1 from Arabidopsis thaliana (Cry1At, Fig. 1), for which the reference experimental value of the corresponding midpoint potential is available,[40–42] using hybrid QM/MM methods within LRA framework. Cryptochromes belong to the class of flavoprotein photoreceptors that are involved in growth and development, regulate entrainment of plant and animal circadian rhythms, and are proposed as primary magnetoreceptors in migratory birds.[9] Effective Fragment Potential method for biomolecules (BioEFP)[43] is used to represent polarizable environment within a QM/MM scheme. The structure of the manuscript is as follows. The QM/MM protocol along with BioEFP approach are described in Section II. Computational details are provided in Section III. The computed energetic parameters and redox potentials of FAD in Cry1At are presented in Section IV. The conclusions are given in Section V.

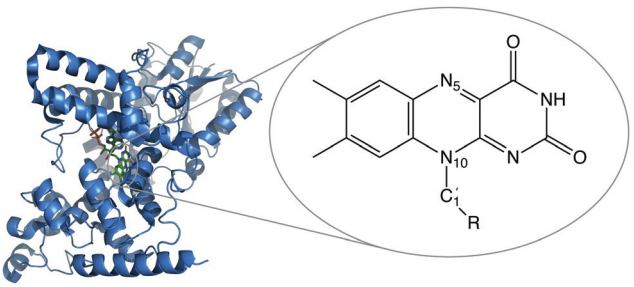

Fig. 1 A structure of Cry1At (PDB: 1U3D[37]) and chemical structure of flavin chromophore, part of FAD cofactor located inside the protein.

II Theory and methods

Redox potentials within linear response approximation LRA[17,20,21,24,26,28–32] was shown to be a powerful tool for computing reaction free energies and redox potentials and was previously validated for multiple systems, including both solvated organic molecules and biological macromolecules. In LRA,[17,28,30] oxidation free energy ($\Delta_r G^{LRA}$) and reorganization energy (λ^{LRA}) can be obtained from two quantities (eqn (1) and (2)): the ensemble-averaged vertical electron affinity ($\langle VEA \rangle_{Ox}$) of the oxidized form and the ensemble-averaged vertical ionization energy ($\langle VIE \rangle_{Red}$) of the reduced form.

$$\Delta_r G^{LRA} = \frac{1}{2}\left(\langle VEA \rangle_{Ox} + \langle VIE \rangle_{Red}\right) \quad (1)$$

$$\lambda^{LRA} = \frac{1}{2}\left(\langle VEA \rangle_{Ox} - \langle VIE \rangle_{Red}\right) \quad (2)$$

where $\langle VEA \rangle_{Ox} = \langle E(Ox) - E(Red) \rangle_{Ox}$ is computed for the ensemble of the oxidized form, $\langle VIE \rangle_{Red} = \langle E(Ox) - E(Red) \rangle_{Red}$ is computed for the ensemble of the reduced form. E(Ox) and E(Red) are electronic energies of the oxidized and reduced forms, respectively.

While eqn (1) and (2) yield formal expressions for $\Delta_r G^{LRA}$ and λ^{LRA}, in practice, QM/MM energy calculations are performed on finite systems. Therefore, the finite system $\Delta_r G^{LRA,f}$ and $\lambda^{LRA,f}$ are the quantities that are obtained directly from QM/MM calculations:

$$\Delta_r G^{LRA;f} = \frac{1}{2}\left(\langle VEA \rangle^{f}_{Ox} + \langle VIE \rangle^{f}_{Red}\right) \quad (3)$$

$$\lambda^{LRA;f} = \frac{1}{2}\left(\langle VEA \rangle^{f}_{Ox} - \langle VIE \rangle^{f}_{Red}\right) \quad (4)$$

Superscript f indicates that the calculations are performed on finite systems.

The missing long-range electrostatic interactions were shown to be crucial for quantitative estimates of vertical energy gaps (VEGs) such as VEA and VIE, and free energies,[26,38,39] and can be further accounted for following thermodynamic cycle shown in Fig. 2. The thermodynamic cycle relates the bulk reaction free energy ($\Delta_r G^{LRA}$) with the reaction free energy for the finite system ($\Delta_r G^{LRA,f}$) and the solvation free energies for the finite systems representing the oxidized ($\Delta G^{f}_{solv}(Ox)$) and reduced forms ($\Delta G^{f}_{solv}(Red)$). The final expression for $\Delta_r G^{LRA}$, therefore, will have the following form:

$$\Delta_r G^{LRA} = \Delta_r G^{LRA,f} + \Delta\Delta G^{f}_{solv} \quad (5)$$

where $\Delta\Delta G^{f}_{solv} = \Delta G^{f}_{solv}(Ox) - \Delta G^{f}_{solv}(Red)$ is differential solvation free energy (see Fig. 2).

To mitigate artifacts caused by inconsistencies in the number of particles for the structures representing ensembles of oxidized and reduced forms, $\Delta\Delta G^{f}_{solv}$ is further approximated by the average between the corresponding quantities for oxidized and reduced ensembles:

$$\Delta\Delta G^{f}_{solv} \approx \frac{1}{2}\left(\langle \Delta\Delta G_{solv} \rangle^{f}_{Ox} + \langle \Delta\Delta G_{solv} \rangle^{f}_{Red}\right)$$

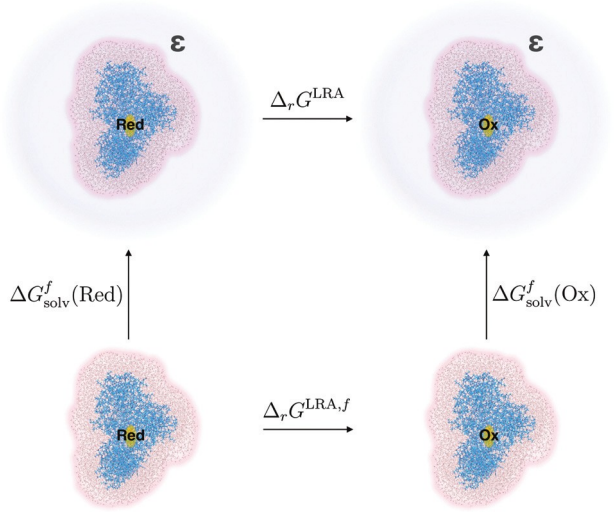

Fig. 2 Thermodynamic cycle for calculation of bulk oxidation free energy ($\Delta_r G^{\mathrm{LRA}}$) as shown in eqn (5). FAD in oxidized and reduced states is embedded in the explicit protein–solvent environment. The entire model system is immersed in the solvent (blue) with the static dielectric constant ε.

where $\langle\Delta G^{f}_{\mathrm{solv}}\rangle_{\mathrm{Ox}}$ is an ensemble-averaged difference in solvation free energy between oxidized and reduced forms for the oxidized form ensemble, and $\langle\Delta G^{f}_{\mathrm{solv}}\rangle_{\mathrm{Red}}$ is the similar quantity, computed for the reduced form ensemble (see Section III for more details).

Once the free energy of one-electron oxidation process is computed ($\Delta_r G^{\mathrm{LRA}}$), the standard reduction potential can be evaluated as follows:

$$E = \frac{\Delta_r G^{\mathrm{LRA}} + \Delta_r G[H^+ \rightarrow \frac{1}{2}H_2] + \mathrm{EC} - \mathrm{FD}}{F} \qquad (6)$$

where $\Delta_r G[H^+ \rightarrow \frac{1}{2}H_2]$ is the standard Gibbs free energy for H^+ reduction. EC − FD originates from the integrated heat capacity and the entropic contribution for electron, assuming the electron convention and Fermi–Dirac statistics (0.038 eV),[11,44,45] and F is Faraday constant. The value of 4.281 eV is used in this work for the reference free energy of H^+ reduction $\Delta_r G[H^+ \rightarrow \frac{1}{2}H_2]$, which was obtained omitting the gas–liquid interface surface potential.[45]

The reduction potential relative to normal hydrogen electrode (NHE) can be further obtained as:

$$E1^0 = E1 + \Delta E(\mathrm{SHE} - \mathrm{NHE}) \qquad (7)$$

where ΔE(SHE − NHE) corresponds to switching from the standard hydrogen electrode (SHE) to NHE reference (+0.006 V).[45]

Biomolecular/macromolecular effective fragment potential method for proteins (BioEFP)

The Effective Fragment Potential (EFP) method is an ab initio force field (FF) method that describes the interactions between solute and solvent molecules in a complex environment.[46–50] The Biomolecular Effective Fragment Potential, or BioEFP, is an extension of the EFP method for modeling the interactions in proteins and macromolecules.[43] In this work, the chromophore lumiflavin was modeled using density functional theory (QM region), while BioEFP was employed to model the protein environment and surrounding water (EFP region). The interactions between the QM and EFP regions are represented via electrostatic and polarization terms, corresponding to the polarizable embedding:

$$\hat{H} = \hat{H}_0 + \sum_{pq} |p\rangle\langle q| \langle p|\hat{V}^{\mathrm{Coul}} + \hat{V}^{\mathrm{pol}}|q\rangle; \qquad (8)$$

where $\hat{H}$ is the Hamiltonian of the combined QM/EFP system, $\hat{H}_0$ is an unperturbed Hamiltonian of the QM part, $\hat{V}^{\mathrm{Coul}}$ and $\hat{V}^{\mathrm{pol}}$ are electrostatic and polarization perturbations due to effective fragments, and |pi, |qi are the atomic orbitals of the QM region.

The electrostatic term in EFP is modeled using a multipolar expansion truncated at octupoles; multipoles are centered at all the atoms and bond midpoints. Polarization term is described via induced dipoles that are computed at the expansion points of the static anisotropic polarizability tensors located at the localized molecular orbital centroids of the effective fragments. Induced dipoles are iterated until self-consistency with induced dipoles of other fragments and a wave function of the QM region is reached.

The electrostatic and polarization terms are considered the most significant ones, as far as the effect of the EFP environment on the electronic properties of the QM solute are concerned,[51–54] even though this question requires further investigation.[55,56] The electrostatic and polarization EFP terms contribute to the QM Hamiltonian via one-electron integrals (eqn (8)), altering the quantum Hamiltonian and molecular wave function of the QM region.

In the BioEFP method, the FF parameters corresponding to the protein are obtained in preparatory first-principles calculations as follows. The polypetide chain is fragmented along Cα–C bonds. The resulting fragments, each containing a peptide group and an amino acid residue, are capped with hydrogen atoms along all the fragmented bonds. EFP parameters for the capped fragments are computed at HF/6-31G(d) level using the MAKEFP module in the GAMESS electronic structure package.[57] Following this, the parameters corresponding to the capping hydrogen atoms and bond midpoints are removed. The resulting excess charge is redistributed to the closest carbon atom, to ensure an integer charge in all the fragments. A detailed summary of this procedure is available in ref. 43.

The EFP parameters for ions were obtained using a mixed basis set simulation (6-31G(d) for electrostatics and 6-311++G(3df,2p) for polarization). Similarly, EFP parameters for water were prepared with 6-31+G(d) basis for electrostatics and 6-311++G(3df,2p) basis for polarization. Water molecules were described with simplified potentials, in which multipoles were truncated at the quadrupole level and distributed to atoms only (no bond midpoints), and polarizability was described with a single polarizability tensor located at the center of mass.

To account for charge penetration effects, electrostatic and polarization terms were augmented by short-range screening functions. The fragment–fragment electrostatic interactions were damped using an exponential screening function, while the QM–fragment interactions were damped with a Gaussian-type screening, as discussed in ref. 43 and 58.

Further, to avoid overpolarization of neighboring amino acid residues, a Gaussian-type damping was applied to screen the fragment–fragment polarization at short distances.[58] A damping parameter a = 0.3 was used for the amino acid residues and water fragments, while a more rigorous screening a = 0.1 was used for ions.

III Computational details

As evaluation of redox potentials relies on the ensemble-averaged VEGs, the first step of simulations is sampling of the configurational space for the ensembles of oxidized and reduced states. The sampling was performed with molecular mechanics molecular dynamics (MM MD) techniques. The initial structure was obtained from Protein Data Bank (PDB: 1U3D[37]). Non-hydrolyzable analog of adenosine triphosphate (ATP), adenylyl-imidodiphosphate (AMP-PNP), was replaced by ATP following the work by Cailliez et al.[59] The positions of crystallographic water molecules and magnesium ion were not altered. The protonation states of amino acid side chains were assigned following PropKa analysis, [60] pK_a calculations with Poisson–Boltzmann equation (PBEQ) solvers in CHARMM-GUI, [61–63] and following discussion by Solov'yov et al.[64] Specifically, E350 and D396 were protonated, H253 was doubly protonated (positively charged), and H255 was singly protonated at the ε position. CHARMM36 FF for the protein,[65–67] general FF bonding and van der Waals parameters (excluding charges for lumiflavin) for the non-protein residues,[68] and TIP3P model for water were used.[69] The MM FF point charges for the oxidized form of FAD were adapted from the work by Solov'yov et al.[64] The charges of the semireduced form of FAD (q_i^{SR}) were adjusted in the following way:

$$q_i^{SR} = q_i^{OX}(FF) + (q_i^{SR}(NBO) - q_i^{OX}(NBO))$$

where $q_i^{OX}(FF)$ are point charges from ref. 64, $q_i^{OX}(NBO)$ and $q_i^{SR}(NBO)$ are natural bond orbital (NBO)[70,71] charges for oxidized and semireduced forms of lumiflavin, respectively. The NBO charges were obtained for lumiflavin in the gas phase for DFT charge density evaluated with oB97X-D functional [72,73] and 6-31+G(d,p) basis. The point charge on H-atom capping C_1^0 of CH_2 group (Fig. 1) in DFT calculations was uniformly redistributed among all lumiflavin atoms. The total charge of the entire model system with the oxidized form of lumiflavin prior solvation was 9.

Protonated crystal structure was immersed in water dodecahedron box. 9 Na^+ counterions were added for neutralization of total system charge followed by MM energy minimization. Equilibration was performed in two steps: NVT-equilibration (T = 300 K, t = 500 ps, velocity rescale thermostat [74]), followed by NPT-equilibration (T = 300 K, p = 1 bar, t = 500 ps, velocity rescale thermostat[74] and Parinello–Rahman barostat[75]) with 1 ns total equilibration time. 10 ns NPT MD simulation was then performed to obtain the production run trajectory. During MM minimization, equilibration and production MD simulations, periodic boundary conditions were enforced, bonds with the H-atoms were constrained using LINCS algorithm. [76] 2 fs time step was used. All classical MD simulations were performed with GROMACS package.[77]

The first 5 ns of the production run trajectory were discarded, and 50 snapshots were selected with the equal 100 ps interval within the next 5 ns. For each MM MD snapshot, the geometry of lumiflavin was locally optimized within the fixed MM environment, eliminating possible artifacts coming from the MM MD geometries.[78] The local QM/MM optimization was performed at oB97X-D/6-31G(d) level.

Two sets of energy calculations have been performed using the selected snapshots. In one, the role of long-range electrostatic interactions on the computed $hVEAi^f$, $hVIEi^f$ and $D_rG^{LRA,f}$ was explored. In this case, the model systems for the following energy calculations were obtained from MM MD snapshots by including the entire protein molecule, all water molecules inside the protein, all counterions, and the water molecules located within the certain distance (R) from the protein outer surface (Fig. 3). R values have been varied from 3 to 15 Å. The vertical energy gaps were computed at oB97X-D/6-31G(d) level for 50 configurations for each R value, utilizing both hybrid approaches: QM/BioEFP[43] and electronic embedding QM/MM with the environment being represented by static multipoles (non-polarizable QM/BioEFP scheme referred to from here on as QM/NP-BioEFP). Quantum part included lumiflavin moiety of FAD cofactor capped at C_1^0 with H-atom. 1–4 electrostatic interactions were turned off. The rest of the FAD cofactor belonged to MM subsystem. All electronic structure calculations were performed in Q-Chem.[79]

The model system with water molecules within 10 Å from the protein outer surface (R = 10 Å) was chosen for another set of calculations, in which the convergence of computed quantities (hVEAi and hVIEi) with respect to the size of the polarizable shell has been explored. The polarizable shell (Fig. 3) was defined by the amino acid residues and water molecules located within certain distance (r) from lumiflavin center of mass. The size of

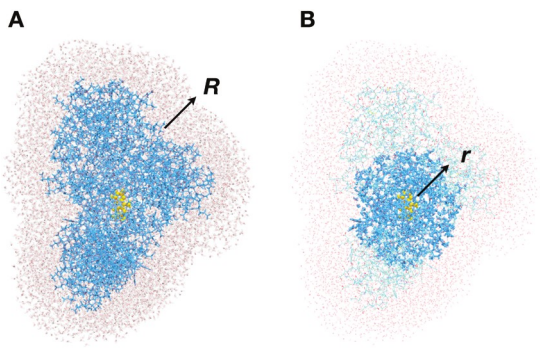

Fig. 3 (A) Water shell around the protein, defined by radius R. (B) Polarizable shell, defined by the distance r from the center of mass of the QM part (lumiflavin).

the shell (r) was varied from 10 to 50 Å. The part of the FAD cofactor that belonged to the MM subsystem and ATP were always kept polarizable. If C_a-atom or oxygen atom were within r distance from the lumiflavin center of mass, the residue or water molecule, respectively, were treated as polarizable. Otherwise only static electrostatic interactions were accounted for. For each r, the ensemble-averaged value over 50 configurations was computed.

The bulk oxidation free energy (D_rG^{LRA}) was obtained from $D_rG^{LRA,f}$ by adding the precomputed difference in solvation energies ($DD\hat{G}_{solv}$, Fig. 2 and eqn (5)). The latter was evaluated by numerically solving the PBEQ with Adaptive Poisson–Boltzmann Solver (APBS)program.[80] The cavity, specifying the boundary between solute system and continuous solvent,was defined as the outer boundary of the composite system, including protein, FAD, ATP, water (R = 10 Å), and counterions. Scaled Bondi radii[81,82] (with the factor f = 1.2 [83,84]) with adjusted radius for hydrogen to 1.1 Å [85] were chosen to form the solvation cavity, as it is commonly used in polarizable continuum models (PCM).[86] The point charges on all atoms of the finite system except the lumiflavin chromophore for both oxidized (Flv_{OX}) and semireduced form (Flv_{SR}) were zeroed out (Fig. 4). The total charge of lumiflavin was either 0 for the oxidized form or 1 for the semireduced one. The scheme was validated for phenoxyl radical and phenolate anion embedded in spherical water clusters by comparing the numerical PBEQ results with the analytical Born solvation free energies (see Section S6 of ESI†). The additive basis set correction (6-31G(d) to aug-cc-pVTZ) was applied to the resulting final bulk free energies (see Section S2 of ESI†).

Once D_rG^{LRA} is available, the standard reduction potential can be obtained from eqn (6). Note, however, that the computed reduction potential corresponds to $Flv_{OX} \overset{e}{\rightarrow} Flv_{SR}$ half-reaction in Cry1At. Yet, the quantity measured experimentally by spectroelectrochemical titration is the midpoint potential for $Flv_{OX} \overset{H^+;e}{\rightarrow} FlvH_{SR}$ half-reaction[40–42] at pH 7.4. At this pH both protonated $FlvH_{SR}$ and unprotonated Flv_{SR} semireduced forms of lumiflavin coexist, with the protonated form $FlvH_{SR}$ being the dominant one.[40,41] The direct comparison between the computational and experimental quantities is possible using Nernst equation based on available estimates of pK_a of N5-atom of flavin (Fig. 1) and pH value corresponding to the experimental conditions (see Section S7 of ESI†).

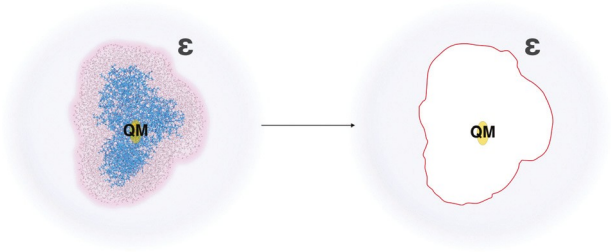

Fig. 4 Solvation of model system by polarizable continuum (blue) representing the aqueous solution with the static dielectric constant e. The charges of lumiflavin were preserved (yellow), and the charges of the MM part were zeroed-out (white).

IV Results and discussion

Below we present the computed ensemble-averaged vertical energy gaps, free energies, and redox potential, and discuss the effects of environment polarization and long-range electrostatic interactions on the energetics of FAD reduction in Cry1At.

Polarization convergence

Fig. 5 shows computed hVEAi and hVIEi for different sizes of the environment polarizable shell (r, Fig. 3). One can see that hVEAi and hVIEi converge at r E 40 Å. Expanding the polarizable shell from 10 Å to 40 Å results in 0.83 and 0.66 eV shifts in hVEAi and hVIEi, respectively. Slow convergence with respect to the environment polarization is in line with the previous studies of excitation energies of photoactive proteins (GFP and rhodopsin) by Kongsted and co-workers.[36] It was shown that the excitation energies converge with respect to the size of polarizable shell at E20 Å. [36] Ionization and electron attachment are accompanied by the change of the charge state of the chromophore, and, therefore, one would expect even more pronounced dependence of the corresponding energy gaps on the environment polarization and their slower convergence. The trends in hVEAi and hVIEi dependency on r are similar: both monotonically rise with the increase of r and exhibit steps at 18–20 Å and 28–30 Å.

Cryptochrome is a globular protein of a non-spherical shape (Fig. 1). Therefore, with the polarizable shell increasing, multiple protein–solvent boundaries are crossed and the nature of polarization interactions with the QM part changes: interactions with the protein vs. interactions with the protein–solvent environment. Thus, the steps are attributed to encountering protein–solvent interfaces accompanied by major changes in polarization interactions at the particular values of r.

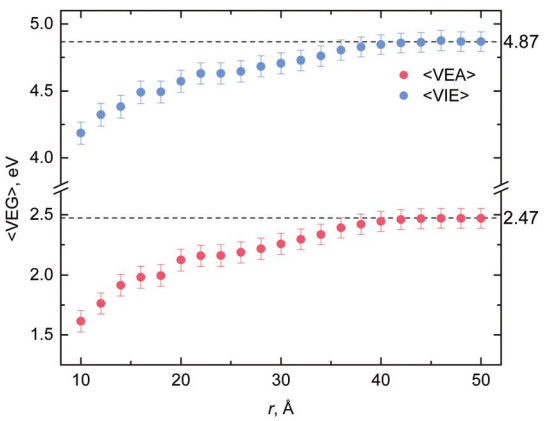

Fig. 5 Ensemble-averaged vertical electron affinity (hVEAi) and vertical ionization energy (hVIEi) computed for different sizes of the polarizable shells, r. The dashed lines represent the values obtained by QM/BioEFP (fully polarizable environment). The dots represent the values obtained from QM/BioEFP/NP-BioEFP calculations with the fragments beyond radius r described by NP-BioEFP.

QM/NP-BioEFP hVEAi and hVIEi computed without account for environment polarization are 0.80 and 3.74 eV, respectively, being shifted by 1.67 and 1.13 eV from the corresponding QM/BioEFP (fully polarizable environment) values. Thus, proper account for environment polarization is critical for accurate estimates of ionization and electron attachment energies in heterogeneous protein environment. The shifts in VEAs or VIEs for molecules in homogeneous environment were previously shown to be as large as several tenths of eV.[21,23,24,26]

System size effects

While the target quantities are D_rG^{LRA} and $E1^0$(NHE), the computed QM/MM values of hVEAi and hVIEi are obtained for the finite systems. Therefore, the dependence on the system size has to be explored. Computed hVEA^f, hVIE^f and $D_rG^{LRA,f}$ as a function of system size (R, Fig. 3) are shown in Fig. 6 and 7.

The shifts in computed QM/NP-BioEFP hVEGis between 3 Å and 15 Å water shells are significant: 0.70 eV and 0.99 eV for hVEAi and hVIEi, respectively. Same shifts become 0.88 eV and 1.09 eV once the polarization is taken into account. QM/NP-BioEFP and QM/BioEFP hVEGi dependences on R exhibit the same trends, indicating that the long-range effects are the ones of static electrostatic origin.

As follows from Fig. 6, hVEGis have not reached convergence with respect to the system size even for 15 Å water shell around the protein. Therefore, further account for long-range electrostatic interactions is necessary (see Sections II and III). Our previous studies[26] showed that sampling of the configurational space using TIP3P water model (also used here) leads to overestimated pre-factor in Born corrections to vertical energy gaps, which in turn corresponds to higher effective charge of the solute. As the slope is overestimated, the effect on absolute values becomes more pronounced, when more explicit water molecules are present in the model system. The model system should be therefore large enough to account for important polarization effects, yet, small enough not to exhibit artifacts due to configurational sampling with employed MM FFs. Here, we chose R = 10 Å to satisfy the above criteria.

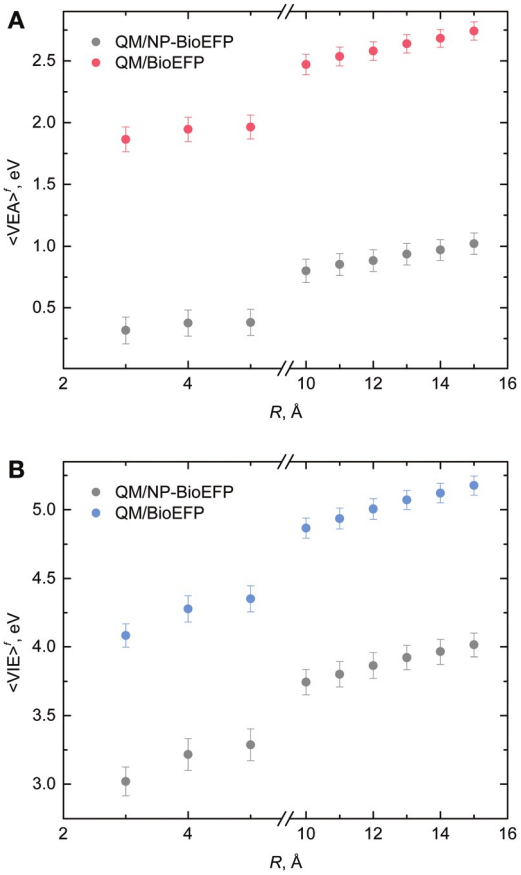

Fig. 6 Ensemble-averaged vertical electron affinities (hVEAi, A) and vertical ionization energies (hVIEi, B) computed for different sizes of water shell, R.

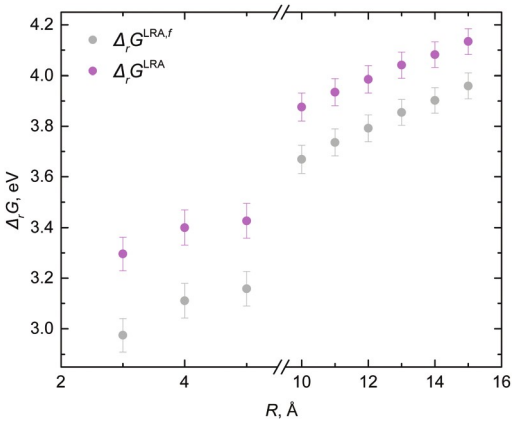

Fig. 7 QM/BioEFP D_rG^{LRA} dependence on the size of the water shell, R without and with bulk solvation correction (DDG $^f_{solv}$).

Redox potential in heterogeneous protein environment

Once the hVEGis are computed, the free energy of oxidation and reduction potential can be evaluated within LRA (eqn (3), (6) and (7)). As discussed above, the environment, including the protein, has a significant effect on the VEGs.[21,23,24,26,36] Computed ensemble-averaged hVEA^f and hVIE^f as well as linear response free energy $D_rG^{LRA,f}$ for the gas phase (configurations of lumiflavin from MM MD without protein–solvent environment), QM/NP-BioEFP and QM/BioEFP are listed in Table 1. All values are averaged over 50 configurations, and provided for 10 Å water shell configurations (R = 10 Å).

Interaction with the protein–solvent environment significantly alters hVEGis of lumiflavin as can be seen from Table 1. Taking into account only static electrostatic interactions

Table 1 Ensemble-averaged vertical electron affinity (hVEAi ^{f}), vertical ionization energy (hVIE^f), and linear response reaction free energy ($D_rG^{LRA,f}$) calculated for lumiflavin chromophore at oB97X-D/6-31G(d) level. Protein, the rest of FAD, ATP, counterions, and water molecules within 10 Å from the protein (R = 10 Å) are present in the model system. All values are averaged over 50 configurations and given in eV

	Gas-phase	QM/NP-BioEFP	QM/BioEFP
hVEA^f	1.20 0.01	0.80 0.10	2.47 0.08
hVIE^f	1.77 0.01	3.74 0.09	4.87 0.07
$D_rG^{LRA,f}$	1.48 0.01	2.27 0.07	3.67 0.06

(gas phase vs. QM/NP-BioEFP) results in the shifts of 0.40, 1.97, and 0.79 eV in $hVEA_i^f$, $hVIE_i^f$, and $D_rG^{LRA,f}$, respectively. Further account of environment polarization (QM/NP-BioEFP vs. QM/BioEFP) also causes large shifts in computed energetic parameters: 1.67, 1.13, and 1.40 eV for $hVEA_i^f$, $hVIE_i^f$, and $D_rG^{LRA,f}$, respectively. Importantly, the effects of polarization on $hVEA_i^f$ and $hVIE_i^f$ are even more dramatic for the heterogeneous protein environment than for the previously reported values on small solutes in homogeneous aqueous solutions.[26]

The bulk free energy (D_rG^{LRA}) and the redox potential (E^0(NHE)) were evaluated following eqn (6), (7) and (9). Note that as the geometry of the lumiflavin was locally optimized for each snapshot in the corresponding frozen environment, the lumiflavin nuclear degrees of freedom do not contribute to the final reaction Gibbs free energy. Therefore, the difference in zero-point vibrational energy (DZPVE) and thermochemical correction (D_rG^{thermo}) for lumiflavin approximated by their gas-phase values were added. Finally, since the hVEGis ($hVEA_i^{f,SB}$ and $hVIE_i^{f,SB}$) were computed with the relatively small basis set without diffuse basis functions, 6-31G(d), the additive basis set corrections to aug-cc-pVTZ have been added ($hDVEA_i^{BSC}$ and $hDVIE_i^{BSC}$). The results based on the final expression for D_rG^{LRA} (eqn (9)) are provided in Table 2 with more details reported in Section S1, S2 and S5 of ESI.†

$$D_rG^{LRA} = \frac{1}{2}\langle VEA_i^{f,SB} + VIE_i^{f,SB}\rangle + \frac{1}{2}\langle DVEA_i^{BSC} + DVIE_i^{BSC}\rangle + DDG_{solv}^{f} + DZPVE + D_rG^{thermo} \quad (9)$$

Presence of the environment as well as accounting for environment polarization is crucial for the quantitative estimates of redox potentials (Table 2). The differences in final values of redox potentials between the gas-phase and QM/NP-BioEFP, and between the gas-phase and QM/BioEFP values are 0.99 V and 2.39 V, respectively. Our best estimate of the midpoint potential of FAD in Cry1At (0.07 V) is in good agreement with the available experimental data.[40–42] Nevertheless, further improvements in the computational protocol are desirable with the main challenges being proper description of long-range electrostatic interactions and reliable configurational sampling. The former challenge could be addressed in a three-layered QM/polarizable MM/nonequilibrium PCM schemes. To our knowledge, there is no existing production level code enabling this type of calculations for vertical ionization and electron attachment energies. The role of solvent models on configurational sampling, in particular the possibility of using polarizable forcefields or many-body potentials, is the subject of the ongoing work.

Table 2 Free energy (D_rG^{LRA}) and midpoint potential of FAD in Cry1At w.r.t. normal hydrogen electrode (E^0(NHE)). Protein, the rest of FAD, ATP, counter-ions, and water molecules within 10 Å from the protein (R = 10 Å) were included in QM/NP-BioEFP and QM/BioEFP calculations, reflecting the magnitudes of electrostatic and polarization contributions w.r.t. to the gas-phase value. D_rG^{LRA} and E^0(NHE) values are given in eV and V, respectively

	Gas-phase	QM/NP-BioEFP	QM/BioEFP	Experiment
D_rG^{LRA}	2.00 0.01	2.99 0.07	4.39 0.06	—
E^0(NHE)	2.32 0.01	1.32 0.07	0.07 0.06	0.15 0.02 [40]

V Conclusions

Here, we have presented the results of computational studies of redox potential of FAD in Cry1At and explored the effects of long-range electrostatic interactions and environment polarization on the relevant energetic parameters. We demonstrate that the account of environment polarization is crucial for accurate estimates of redox potentials of biomolecules: the shift in the value of reduction potential between non-polarizable and polarizable QM/MM schemes is 1.4 V. We also show that proper care should be taken of long-range electrostatic interactions if absolute values of the redox potentials are the target: computed VEGs are not converged with respect to the system size even for 15 Å water shell around the protein. The theoretical estimate of the midpoint potential of FAD in Cry1At, 0.07 V, is reported for the first time and is in excellent agreement with available experimental data.[40–42]

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

The research reported in this publication was supported in part by National Institute Of General Medical Sciences of the National Institutes of Health under award R43GM126804. P. K. G., Y. K. and L. V. S. acknowledge support of the National Science Foundation (Grant CHE-1800505). This research was supported in part through computational resources provided by Boston University Shared Computing Cluster and Information Technology at Purdue, West Lafayette, Indiana.

References

1 A. Hagfeldt and M. Grätzel, Chem. Rev., 1995, 95, 49–68.
2 M. Grätzel, Nature, 2001, 414, 338–344.
3 D. Gust, T. A. Moore and A. L. Moore, Acc. Chem. Res., 2001, 34, 40–48.
4 R. P. Sinha and D.-P. Häder, Photochem. Photobiol. Sci., 2002, 1, 225–236.
5 M. Grätzel, J. Photochem. Photobiol., C, 2003, 4, 145–153.
6 B. Halliwell, Plant Physiol., 2006, 141, 312–322.
7 C. H. Foyer and G. Noctor, Antioxid. Redox Signaling, 2009, 11, 861–905.
8 D. Gust, T. Moore and A. Moore, Acc. Chem. Res., 2009, 42, 1890–1898.
9 I. Chaves, R. Pokorny, M. Byrdin, N. Hoang, T. Ritz, K. Brettel, L.-O. Essen, G. T. J. van der Horst, A. Batschauer and M. Ahmad, Annu. Rev. Plant Biol., 2011, 62, 335–364.

10 Y. Tachibana, L. Vayssieres and J. R. Durrant, Nat. Photonics, 2012, 6, 511.
11 A. V. Marenich, J. Ho, M. L. Coote, C. J. Cramer and D. G. Truhlar, Phys. Chem. Chem. Phys., 2014, 16, 15068–15106.
12 D. Close, J. Phys. Chem. A, 2011, 115, 2900–2912.
13 C. A. Schroeder, E. Pluhařová, R. Seidel, W. P. Schroeder, M. Faubel, P. Slavíček, B. Winter, P. Jungwirth and S. E. Bradforth, J. Am. Chem. Soc., 2015, 137, 201–209.
14 A. Muñoz-Losa, D. Markovitsi and R. Improta, Chem. Phys. Lett., 2015, 634, 20–24.
15 E. Pluhařová, P. Slavíček and P. Jungwirth, Acc. Chem. Res., 2015, 48, 1209–1217.
16 T. A. Pham, M. Govoni, R. Seidel, S. E. Bradforth, E. Schwegler and G. Galli, Sci. Adv., 2017, 3, e1603210.
17 J. Blumberger and M. Sprik, in Redox Free Energies from Vertical Energy Gaps: Ab Initio Molecular Dynamics Implementation, ed. M. Ferrario, G. Ciccotti and K. Binder, Springer, Berlin, Heidelberg, 2006, pp. 481–506.
18 J. VandeVondele, M. Sulpizi and M. Sprik, Angew. Chem., Int. Ed., 2006, 45, 1936–1938.
19 J. VandeVondele, R. Ayala, M. Sulpizi and M. Sprik, J. Electroanal. Chem., 2007, 607, 113–120.
20 J. Cheng, M. Sulpizi and M. Sprik, J. Chem. Phys., 2009, 131, 154504.
21 D. Ghosh, A. Roy, R. Seidel, B. Winter, S. Bradforth and A. Krylov, J. Phys. Chem. B, 2012, 116, 7269–7280.
22 J. Cheng, X. Liu, J. VandeVondele, M. Sulpizi and M. Sprik, Acc. Chem. Res., 2014, 47, 3522–3529.
23 P. R. Tentscher, R. Seidel, B. Winter, J. J. Guerard and J. S. Arey, J. Phys. Chem. B, 2015, 119, 238–256.
24 J. J. Guerard, P. R. Tentscher, M. Seijo and J. S. Arey, Phys. Chem. Chem. Phys., 2015, 17, 14811–14826.
25 R. Jono, Y. Tateyama and K. Yamashita, Phys. Chem. Chem. Phys., 2015, 17, 27103–27108.
26 R. Tazhigulov and K. Bravaya, J. Phys. Chem. Lett., 2016, 7, 2490–2495.
27 J. Blumberger, Chem. Rev., 2015, 115, 11191–11238.
28 G. King and A. Warshel, J. Chem. Phys., 1990, 93, 8682–8692.
29 M. H. M. Olsson, G. Hong and A. Warshel, J. Am. Chem. Soc., 2003, 125, 5025–5039.
30 Y. Tateyama, J. Blumberger, M. Sprik and I. Tavernelli, J. Chem. Phys., 2005, 122, 234505.
31 M. Cascella, A. Magistrato, I. Tavernelli, P. Carloni and U. Rothlisberger, Proc. Natl. Acad. Sci. U. S. A., 2006, 103, 19641–19646.
32 A. Barrozo, M. Y. El-Naggar and A. I. Krylov, Angew. Chem., Int. Ed., 2018, 57, 6805–6809.
33 G. Li, X. Zhang and Q. Cui, J. Phys. Chem. B, 2003, 107, 8643–8653.
34 C. Wei, R. Lazim and D. Zhang, Proteins, 2014, 82, 2209–2219.
35 C. Ji, Y. Mei and J. Z. Zhang, Biophys. J., 2008, 95, 1080–1088.
36 M. T. Beerepoot, A. H. Steindal, K. Ruud, J. M. H. Olsen and J. Kongsted, Comput. Theor. Chem., 2014, 1040–1041, 304–311.
37 C. A. Brautigam, B. S. Smith, Z. Ma, M. Palnitkar, D. R. Tomchick, M. Machius and J. Deisenhofer, Proc. Natl. Acad. Sci. U. S. A., 2004, 101, 12142–12147.
38 R. Barnett, U. Landman, C. Cleveland and J. Jortner, Chem. Phys. Lett., 1988, 145, 382.
39 J. V. Coe, J. Phys. Chem. A, 1997, 101, 2055–2063.
40 P. Müller, J. P. Bouly, K. Hitomi, V. Balland, E. D. Getzoff, T. Ritz and K. Brettel, Sci. Rep., 2014, 4, 5175.
41 V. Balland, M. Byrdin, A. Eker, M. Ahmad and K. Brettel, J. Am. Chem. Soc., 2009, 131, 426–427.
42 C. Lin, D. Robertson, M. Ahmad, A. Raibekas, M. Jorns, P. Dutton and A. Cashmore, Science, 1995, 269, 968–970.
43 P. Gurunathan, A. Acharya, D. Ghosh, D. Kosenkov, I. Kaliman, Y. Shao, A. Krylov and L. Slipchenko, J. Phys. Chem. B, 2016, 120, 6562–6574.
44 J. E. Bartmess, J. Phys. Chem., 1994, 98, 6420–6424.
45 A. A. Isse and A. Gennaro, J. Phys. Chem. B, 2010, 114, 7894–7899.
46 P. Day, J. Jensen, M. Gordon, S. Webb, W. Stevens, M. Krauss, D. Garmer, H. Basch and D. Cohen, J. Chem. Phys., 1996, 105, 1968–1986.
47 M. Gordon, M. Freitag, P. Bandyopadhyay, J. Jensen, V. Kairys and W. Stevens, J. Phys. Chem. A, 2001, 105, 293–307.
48 D. Ghosh, D. Kosenkov, V. Vanovschi, C. Williams, J. Herbert, M. Gordon, M. Schmidt, L. Slipchenko and A. Krylov, J. Phys. Chem. A, 2010, 114, 12739–12745.
49 M. S. Gordon, Q. A. Smith, P. Xu and L. V. Slipchenko, Annu. Rev. Phys. Chem., 2013, 64, 553–578.
50 M. Gordon, D. Fedorov, S. Pruitt and L. Slipchenko, Chem. Rev., 2012, 112, 632–672.
51 L. Slipchenko, J. Phys. Chem. A, 2010, 114, 8824–8830.
52 D. Kosenkov and L. Slipchenko, J. Phys. Chem. A, 2011, 115, 392–401.
53 A. DeFusco, N. Minezawa, L. Slipchenko, F. Zahariev and M. Gordon, J. Phys. Chem. Lett., 2011, 2, 2184–2192.
54 D. Ghosh, O. Isayev, L. Slipchenko and A. Krylov, J. Phys. Chem. A, 2011, 115, 6028–6038.
55 C. I. Viquez Rojas, J. Fine and L. V. Slipchenko, J. Chem. Phys., 2018, 149, 094103.
56 L. V. Slipchenko, M. S. Gordon and K. Ruedenberg, J. Phys. Chem. A, 2017, 121, 9495–9507.
57 M. S. Gordon and M. W. Schmidt, Theory and applications of computational chemistry, Elsevier, 2005, pp. 1167–1189.
58 L. Slipchenko and M. Gordon, Mol. Phys., 2009, 107, 999–1016.
59 F. Cailliez, P. Müller, M. Gallois and A. de la Lande, J. Am. Chem. Soc., 2014, 136, 12974–12986.
60 M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, J. Chem. Theory Comput., 2011, 7, 525–537.
61 S. Jo, T. Kim, V. G. Iyer and W. Im, J. Comput. Chem., 2008, 29, 1859–1865.
62 W. Im, D. Beglov and B. Roux, Comput. Phys. Commun., 1998, 111, 59–75.
63 S. Jo, M. Vargyas, J. Vasko-Szedlar, B. Roux and W. Im, Nucleic Acids Res., 2008, 36, W270–W275.
64 I. A. Solov'yov, T. Domratcheva, A. R. Moughal Shahi and K. Schulten, J. Am. Chem. Soc., 2012, 134, 18046–18052.
65 R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. MacKerell, J. Chem. Theory Comput., 2012, 8, 3257–3273.
66 A. D. MacKerell, M. Feig and C. L. Brooks, J. Comput. Chem., 2004, 25, 1400–1415.

67 A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, J. Phys. Chem. B, 1998, 102, 3586–3616.
68 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell, J. Comput. Chem., 2009, 31, 671–690.
69 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, J. Chem. Phys., 1983, 79, 926–935.
70 E. Glendening, J. Badenhoop, A. Reed, J. Carpenter, J. Bohmann, C. Morales and F. Weinhold, NBO 5.0, Theoretical Chemistry Institute, University of Wisconsin, Madison, WI, 2001.
71 F. Weinhold and C. R. Landis, Chem. Educ. Res. Pract., 2001, 2, 91–104.
72 J. Chai and M. Head-Gordon, J. Chem. Phys., 2008, 128, 084106.
73 J. Chai and M. Head-Gordon, Phys. Chem. Chem. Phys., 2008, 10, 6615–6620.
74 G. Bussi, D. Donadio and M. Parrinello, J. Chem. Phys., 2007, 126, 014101.
75 M. Parrinello and A. Rahman, J. Appl. Phys., 1981, 52, 7182–7190.
76 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, J. Comput. Chem., 1997, 18, 1463–1472.
77 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, SoftwareX, 2015, 1–2, 19–25.
78 M. K. Lee and D. F. Coker, J. Phys. Chem. Lett., 2016, 7, 3171–3178.
79 Y. Shao, Z. Gan, E. Epifanovsky, A. Gilbert, M. Wormit, J. Kussmann, A. Lange, A. Behn, J. Deng and X. Feng, et al., Mol. Phys., 2015, 113, 184–215.
80 E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G. W. Wei, M. J. Holst, J. A. McCammon and N. A. Baker, Protein Sci., 2018, 27, 112–128.
81 A. Bondi, J. Phys. Chem., 1964, 68, 441–451.
82 M. Mantina, A. C. Chamberlin, R. Valero, C. J. Cramer and D. G. Truhlar, J. Phys. Chem. A, 2009, 113, 5806–5812.
83 R. Bonaccorsi, P. Palla and J. Tomasi, J. Am. Chem. Soc., 1984, 106, 1945–1950.
84 J. Tomasi and M. Persico, Chem. Rev., 1994, 94, 2027–2094.
85 R. S. Rowland and R. Taylor, J. Phys. Chem., 1996, 100, 7384–7391.
86 J. Tomasi, B. Mennucci and R. Cammi, Chem. Rev., 2005, 105, 2999–3094.