Taking Water into Account with the Fragment Molecular Orbital Method

Okiyama, Yoshio
Fukuzawa, Kaori
Komeiji, Yuto
Tanaka, Shigenori

(Citation)
Quantum Mechanics in Drug Discovery:105-122

(Issue Date)
2020-02-04

(Resource Type)
book part

(Version)
Accepted Manuscript

Taking Water into Account with the Fragment Molecular Orbital Method

Yoshio Okiyama[a], Kaori Fukuzawa[b], Yuto Komeiji[c], Shigenori Tanaka[d]

[a] Division of Medicinal Safety Science, National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa 210-9501, Japan; okiyama@nihs.go.jp

[b] School of Pharmacy and Pharmaceutical Sciences, Hoshi University, 2-4-41 Ebara, Shinagawa, Tokyo 142-8501, Japan; k-fukuzawa@hoshi.ac.jp

[c] Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology, Tsukuba Central 6, Tsukuba, Ibaraki 305-8566, Japan; y-komeiji@aist.go.jp

[d] Graduate School of System Informatics, Kobe University, 1-1 Rokkodai, Nada, Kobe 657-8501, Japan; tanaka2@kobe-u.ac.jp

Keywords: Fragment molecular orbital (FMO) method; Solvent effect; Water molecule; Inter-fragment interaction energy (IFIE); Statistically corrected IFIE (SCIFIE); Ligand binding; Hydration shell; Dielectric continuum; Poisson-Boltzmann equation

ABSTRACT

This chapter describes the current status of development of the Fragment Molecular Orbital (FMO) method for analyzing the electronic state and intermolecular interactions of biomolecular systems *in solvent*. The molecular orbitals and the inter-fragment interaction energies (IFIEs) for a specific snapshot can be obtained directly by performing FMO calculations by exposing water molecules and counter-ions around biomolecular systems. Then, it is necessary to pay attention to the thickness of the water shell surrounding the biomolecules. The single-point calculation for these snapshots does not incorporate the effects of temperature and configurational fluctuation, but the SCIFIE (statistically corrected IFIE) method is proposed as a many-body correlated method that partially compensates for this deficiency. Furthermore, implicit continuous dielectric models have been developed as effective approaches to incorporating the screening effect of the solvent in thermal equilibrium, and we illustrate their usefulness for theoretical evaluation of IFIEs and ligand-binding free energy on the basis of the FMO-PBSA (Poisson-Boltzmann surface area) method and other computational methods.

1. Introduction

Fragment molecular orbital (FMO) calculations in vacuum have shown their effectiveness in various fields such as ligand--protein interaction analysis [1]. However, it has also been recognized that incorporation of solvent effect, together with thermal fluctuation and entropy effects, can further improve the agreement between calculation and experimentation of ligand binding affinity, for instance. In addition, consideration of solvent effects is also important when we study the stable structure of biomolecular complex systems composed of proteins, nucleic acids, small ligand molecules, surrounding ions, and so on. There are two major ways to introduce the solvent effects into FMO calculations, that is, explicit approach that takes water molecules into account, and implicit approach that describes solvent in terms of continuum model. This article illustrates the current status of these approaches and some future issues.

2. Explicit water

Treatment of explicit solvent water seems to be essential when considering the role of molecular water surrounding biomolecules. Biomolecules form hydrogen bonds with water at their interface, which involves charge transfer. The electrostatic potential of the surface of charged solute molecules changes dramatically depending on the presence or absence of molecular water [2]. Interfacial water gradually changes to bulk water as the distance from the solute increases. The presence of solvent water molecules also affects intramolecular

interactions of solute biomolecules. Here, the role of molecular water is described from both viewpoints of shell waters as a solvent environment and crystal waters.

2.1 Shell solvent

Effects of the explicit water solvent on FMO calculation for biological molecules have been investigated systematically with various solute molecules including a protein [3], double-stranded (ds) DNA [4, 5], single-stranded (ss) DNA, and its complex with an ssDNA binding protein (SSB) [6, 7]. As reviewed below, the overall tendency of the solvent effects was similar among the solute molecules as follows: the first solvent shell within 4 Å from the solute surface governs the solute internal energy and charge, while the second shell, up to ca. 8 Å, is also important.

The sequence of papers used the following protocol to investigate the electronic properties of the solute molecules as functions of the solvent thickness. The molecular configurations for FMO were prepared by classical molecular mechanics (MM)--molecular dynamics (MD) (MM--MD) at room temperature. Several configurations were selected from the generated trajectories, were annealed to 0 K, and were finally energy-minimized. From the structures thus optimized, the solutes (protein and/or DNA) and solvent shell (water and counter ions) were excised to construct a systematic series of configurations with varying shell thickness. The thickness of the solvent shell was determined by measuring the distance between non-hydrogen atoms of the solute and solvent. As an example, see Figure 1 for constructed series of molecular configurations of dsDNA [4, 5]. The configurations were subjected to FMO calculation.

<Figure 1 here>

The protein net charge, internal energy, and solvent--protein interaction were calculated as functions of the solvent thickness in the study of ubiquitin, the first comprehensive FMO study of the solvation on a protein [3]. The net charge converged at 4 Å, and the internal energy at 8 Å. The protein--solvent interaction showed a marginal convergence around 8 Å. Another finding was that water molecules within 4 Å of the outer surface were significantly less polarized than the bulk water.

Similar results were obtained for dsDNA [4, 5], ssDNA, and ssDNA:SSB complex [6, 7]. See Figure 2 for the graphs of solute charge, internal energy, and the solute--solvent interaction of the ssDNA:SSB. Frontier molecular orbitals (MOs) of fragments within dsDNS showed a different behavior [4, 5]; the highest occupied MOs (HOMOs) and lowest unoccupied MOs (LUMOs) converged only marginally at 5 Å and they did not converge completely even at 12 Å. The quantity of solvent needed for conversion thus depends on the kind of physical property of interest.

<Figure 2 here>

The effect of counterions on the solute properties may be another concern. In the protein solvation study [3], the ubiquitin protein was chosen because it bears no net charge at the neutral pH and hence no counterion was necessary to neutralize the simulated system. In the dsDNA study [4, 5], enough counterions were added to neutralize the molecular system in the preparative MM--MD calculation, but only those counterions present within the solvent shell were preserved for the subsequent FMO calculation, resulting in FMO

calculations of negatively charged molecular systems. The effect of counterions was minutely investigated later in the studies of the ssDNA:SSB complex [6]. As is clearly seen in Figure 2, the solute charge and internal energy were virtually uninfluenced by presence of enough counterions to neutralize the system. Even the solute-solvent interaction was only slightly influenced. Thus, neutralization by counterions did not affect much the internal properties of the solutes investigated.

This fact does not necessarily mean that counterions have no effect on the solutes. We should keep in mind that in Refs. [6, 7], FMO calculations were performed for only molecular configurations in which counterions surrounded the solutes at a distance, not in direct contact. In fact, in an earlier test FMO calculation of DNA fragments *in vacuo* [4, 5], neutralization of a fragment by a directly bound counterion affected greatly the MOs.

Taken together, the series of FMO calculations of protein and DNA have revealed that the solute net charge and internal energy are affected dominantly by the first water solvent shell within ca. 4 Å of the solutes and auxiliarily by the second shell up to ca. 8 Å and that the counterions do not influence much the solutes unless directly bound to the solutes.

Based upon these findings, minute solvation effects on the solutes were further analyzed. An important finding was modulation of interactions within dsDNA [4, 5]. Upon solvation, there is a minus charge transfer from the DNA to the solvent, which elicits charge rearrangement within the DNA molecules and eventually results in enhancement of the base--base H-bond and in weakening of the stacking. In the SSB recognition by ssDNA, desolvation was clearly seen by inter fragment interaction energy (IFIE) analysis [7]. The

nucleotide units of ssDNA interact with the solvent in the isolated state, but upon SSB binding many of them change their partners to specific amino acid residues of SSB. This desolvation is clearly illustrated in the IFIE (Figure 3) in which the interactions with the solvent are replaced with that with SSB.

<Figure 3 here>

While systematic evaluation of size of shell water has been discussed as above, there have been some studies for protein--ligand interaction analysis using a specific size of shell water. Kurita et al. suggested that the protonation state of histidine residue under water solvent environment plays an important role in the interaction between a chiral ligand and vitamin D receptor (VDR) that covered with 6 Å shell water [8]. They have also shown that in VDR--ligand interaction surrounded by 6 Å shell water, the EC_{50} activity value of the ligand with tetrazole ring was dramatically changed by the difference of nitrogen position, and such "activity cliff" was caused by the hydrogen bond which induced by CH/π interaction [9].

As another aspect of the role of molecular water, electrostatic potentials of the surface of charged solute biomolecules changes dramatically due to charge transfer with surrounding water molecules. In the case of the estrogen receptor alpha (ERα) ligand binding domain, the total charge of the solute protein was originally at −6.0*e*, but changed to about −2.3*e* by charge transfer with the surrounding 4 Å water shell. As a result, the electrostatic potential on the protein surface approached to neutral, and the characteristics of the charged amino

acid residues exposed on the surface became clearer, which helps to understand molecular recognition [2].

2.2 Crystal water

It is also important to deal with crystal water that has been solved by X-ray crystal structure analysis. Several calculations have been made specifically to treat only water molecules that are important for ligand binding, or to contain all crystal waters. For example, in a protein--ligand binding study, the presence of crystal water improved the binding property prediction [10].

The presence of special bound water, which crosslinks protein and ligand, is also known. The crystal structure of ERα--ligand complex includes one bound water molecule, which is conserved in most structures. FMO calculations revealed that the water molecule plays a key role in ligand binding as a member of the hydrogen bonding network among ERα and ligand (Figure 4) [11].

<Figure 4 here>

3. SCIFIE: Effective incorporation of many-body screening

FMO calculations, in the ground state, give electronically polarized molecular state with minimal energy for a fixed molecular configuration. In this sense, the electrostatic screening effect has already been included in the (structurally) static and electronic ways. Thus, through the FMO calculations with explicit water molecules as discussed in Section

2, the electrostatic screening effect can be described in terms of the electron polarization with fixed molecular configuration. However, it is well recognized [1] that the IFIEs thus calculated often fail to account for the electrostatic screening effect due to surrounding water or distant ions even qualitatively, while it is empirically known to be significant. This defect can then be explained physico-chemically by the fact that the calculations do not take into account the effects of configurational polarization of charged or polar molecules or functional groups at finite temperature. This important contribution to dielectric screening is predominantly attributed to polar water molecules surrounding biomolecules, whose measure (e.g., dielectric constant of about 80 in the case of water) is associated with the configurational dynamics of water molecules; its effective incorporation into FMO calculations by means of continuous dielectric model is discussed in Section 4. In this section, on the other hand, we employ a viewpoint of what is (mainly electrostatic) screening effect due to many-body correlations among (charged or polarized) classical "particles", and address an alternative scheme called statistically corrected inter-fragment interaction energy (SCIFIE) approach [12] in which a partial and effective correction for many-body screening is performed on the IFIEs obtained for a snapshot molecular configuration. This SCIFIE scheme is practically useful because FMO calculations are often carried out for a snapshot configuration due to their high computational cost, while the ensemble average over multiple configurations at finite temperature would also be possible at later stage.

Let us consider a molecular configuration for which an FMO-IFIE calculation has been performed. There are a variety of pairwise IFIEs with positive (repulsive) or negative

(attractive) sign in that system due to mixed ionic, polar, dispersion, exchange-repulsion, charge-transfer and other molecular interactions. Then, the basic idea of SCIFIE is as follows [12]. As for molecular interactions, in general, the contribution of electrostatic (Coulombic) interaction is predominantly large and retaining the charge neutrality imposes a significant constraint on system energetics to avoid substantial increase in total energy without the charge neutrality. For example, Coulombic systems always have negative charges in the near vicinity of positive charges (and vice versa) to keep the charge neutrality locally; then considering an electric charge distantly interacting with these positive and negative charges, it interacts strongly with each individual positive or negative charge, but effectively feels a much milder interaction with their mixture due to the cancellation among positive and negative contributions. This picture thus provides an underlying basis for those physical accounts such as the Ewald summation and the electrostatic screening effect. Hence, instead of considering the strong interactions with high magnitudes between "bare" electric charges, it would be relevant to account for effective, screened interactions among "dressed" particles that are alternatively placed with opposite charges. The difficulties found in "bare" IFIEs are analogous to these observations, and we may (partly) overcome them by considering the effective interactions associated with or represented by inter-particle correlations. Such a description can be given in the context of classical many-body problem [13] through the concepts of screened interaction and direct correlation function.

In the SCIFIE scheme, each fragment *i* in a molecular system is regarded as a classical particle embedded in the network of mutual interactions whose "bare" energy values are

given as u_{ij} between the fragments i and j. The effective interaction or the potential of mean force w_{ij} is then related to the pair correlation function h_{ij} as

$$h_{ij} = e^{-\beta w_{ij}} - 1 \, .$$

The parameter β is usually related to the absolute temperature T and the Boltzmann constant k_B via $\beta = 1/k_B T$, but, in the present formalism, may be regarded as an optimization parameter to control the degree of screening, which will be specified afterwards. The pair correlation function h_{ij} is related to the direct correlation function c_{ij} in terms of the Ornstein--Zernike (OZ) relation [13]:

$$h_{ij} = c_{ij} + \sum_{k \neq i,j} c_{ik} h_{kj} \, .$$

The direct correlation function is conceptually introduced as a difference between the total and indirect parts of inter-particle correlation. We may thus express it in the Percus--Yevick (PY) approximation [13] for classical many-body problem as

$$c_{ij} = e^{-\beta w_{ij}} - e^{-\beta (w_{ij} - u_{ij})},$$

providing a closure equation to determine w_{ij} for a given set of u_{ij}. Here, one may also employ other closure approximations such as the hypernetted chain (HNC) approximation [13].

The procedure above on the basis of the PY approximation may be formulated as follows. By introducing the Mayer function,

$$f_{ij} = e^{-\beta u_{ij}} - 1 \, ,$$

we can rewrite the PY equation as

$$c_{ij} = \frac{f_{ij}(h_{ij}+1)}{f_{ij}+1} .$$

Substituting this equation into the OZ relation, we find

$$h_{ij} = \frac{f_{ij}(h_{ij}+1)}{f_{ij}+1} + \sum_k \frac{f_{ik}(h_{ik}+1)}{f_{ik}+1} h_{kj} ,$$

or

$$h_{ij} = f_{ij} + (f_{ij}+1) \sum_k \frac{f_{ik}(h_{ik}+1)}{f_{ik}+1} h_{kj} .$$

For given u_{ij} and f_{ij}, we can thus obtain h_{ij} through solution to this equation, and also find the effective interaction as

$$w_{ij} = -\frac{1}{\beta} \ln\left(h_{ij}+1\right) .$$

In the formulation above, the effective, screened interactions are obtained through the inter-fragment correlations in a system with given set of bare interactions. The (physical) origin of the screening may partially be attributed to the conformational fluctuations and associated entropic effects in the system, but the parameter β should not be chosen according to room temperature, but rather as a measure of randomness and collectivity. The energy scale of IFIEs obtained by single-point FMO calculation is of electronic origin and hence much higher than that of thermal energy at room temperature. While the electronic state of the pertinent system is very sensitive to the conformational fluctuations due to

thermal energy, the relevant value of β in the present screening model at a fixed conformation should be independent of the thermal energy associated with the molecular system whose conformations could fluctuate at finite temperature. The parameter β is thus regarded as not "thermodynamic" one but "information theoretic" one, and its optimal value can be determined by a "maximum screening" ansatz under the energy constraint condition so that a quantity characterizing the degree of screening

$$\sigma = \sum_{i<j} |w_{ij}|$$

is minimized. It is here remarked that the pair correlation function h_{ij} can be related to the mutual information in information theory [14]. In addition, in order to find a more realistic evaluation for the effective interaction between fragments, the statistical averaging over the thermal motion at room temperature may be taken later via generating many conformations for FMO calculations with, e.g., MD simulations.

The SCIFIE scheme, as formulated above, has been successfully applied to some model and more realistic (biomolecular, organic and inorganic) systems [2, 15–17], in which one observes the effective screening of "bare" IFIEs.

4. Implicit water

4.1 Overview

An implicit solvent treatment is a practical way in quantum chemistry and a first choice to include solvent effects into the description of electronic state of a solute molecule. All biochemical processes take place in the environment of water solution that can strongly affect the reaction energies [18]. Therefore, such effects are conveniently introduced by treating surroundings as a dielectric continuum also in FMO. There are several FMO applications with such continuum-solvent methods as polarizable continuous model (PCM) [19], solvation model density (SMD) [20], and Poisson--Boltzmann surface area (PBSA) model [21, 22]. An implicit treatment has the advantage of avoiding costly statistical sampling of the explicit solvent configurations, while giving some sort of averaged results consistent with an explicit treatment. On the other hand, there is a discrepancy with an explicit treatment in the description of solute--solvent interactions including electron's outflow. Therefore, it would be indispensable to use the implicit or explicit treatment properly depending on the situation.

One of the essential features of the FMO application with the implicit solvent model is the ability to describe the reasonable MO on a local fragment, which is the most symbolic and critical characters defining the electronic properties of the system. HOMO is especially crucial in analyses because of its relation to the stability and reactivity, while its energy inside biomolecules often seems unstable in the gas phase. By imposing induced charges as

an implicit solvent, the energy is successfully lowered to guarantee the stable electronic state for DNA [22] and proteins [23].

Another is the ability to give the reasonable solvation free energy of whole system. For example, the solvation free energy of dsDNA in an implicit solvent is asymptotically close to that in an explicit solvent sufficiently including water and ions in the previous section (Figure 5). Based on such a solvated FMO ability, the qualification of modeled structures was performed by Simoncini et al. [24]. For eight protein targets, they assessed the quality of their predicted models by using FMO energy to extract the best model from each ensemble. The FMO energy was derived from FMO2-HF/6-31G(d) with Grimme's dispersion correction [25, 26] (HF-D3) in PCM solvent. They found FMO-based model ranking was superior to empirically derived one.

<Figure 5 here>

IFIE analysis and binding affinity prediction will be of special interest in drug discovery. The following sections present their significant analysis and application examples. Note that all FMO calculations combined with an implicit solvent model below use the FMO2-MP2/6-31G(d) level with some exceptions.

4.2 Fragment interaction energy

IFIE and its decomposition (PIEDA) analyses can reveal the electronic mechanisms leading to local stability inside biomolecules. How is then IFIE affected when a biomolecule is immersed in an implicit solvent ? Actually, an internal IFIE in solution, that is, an IFIE

purely considered between solute fragments, hardly changes compared to that in vacuo, even if one uses an implicit or explicit solvent model. Accordingly, Fedorov et al. explicitly included the solute--solvent interaction into IFIE as a natural division of total FMO energy in solution and treated its contribution as the solvation term in PIEDA [27]. The solvation term arises counteractively to the electrostatic (ES) term, and they eventually cancel out each other (see Figure 6). This is called the solvent-screening effect and *the "perfect" screening is observed for the interaction of two-point charges* under ideal conditions [28].

<Figure 6 here>

Screening effect

Okiyama et al. practically and systematically studied solvent-screening effects on fragment interactions for DNA and bioactive proteins, ubiquitin and ERα, using FMO-PBSA approach [22, 29]. Solvent screening effectively reduces the excessive electrostatic interactions between charged fragments also inside such biomolecules while hardly affecting the electron-correlated ones. For example, IFIEs between backbone fragments including charged phosphate groups for DNA or between polar-residue fragments on the molecular surface for proteins are demonstratively damped according to the distance. Screening gives clear view for significant interactions including secondary structure patterns.

On the other hand, if two fragments are both insufficiently exposed to the solvent or neutral, there is little or no influence on IFIE. This is exactly the case for proteins, whose fragments are never uniformly exposed to the solvent. It should be noted whether focused

interactions are subject to screening effects, because especially a ligand complexed with a protein is more or less buried inside the receptor.

Here, we update the story as *the screening between two fragments depends on their chargeability, solvent exposure, and distance* in the case of biomolecular systems. When focused on the interactions with a ligand, application of subsystem analysis [30] in consideration of desolvation is also promising in the future.

Energetic map

The energetic map analysis is an application of screening-included IFIE in the context of drug discovery. Śliwa et al. proposed a methodology to determine significant interactions in each protein--ligand complex by using a PIEDA-based energetic map of the orthosteric binding site [31]. They demonstrated the applicability to serotonin receptors, 5-HT_{1A}, 5-HT_{2A}, and 5-HT_7 complexed with selected long-chain arylpiperazines by using FMO-PCM approach. This methodology identifies significant attractive/repulsive areas in the binding sites that would provide hints for designing new binders to target proteins.

4.3 Binding free energy

Accurate prediction of binding free energy is the central theme in drug discovery. In the FMO-based approach, the binding energy can be estimated either by the supermolecular approach or by the summed energy of IFIEs (IFIE sum) or total interaction energy (TIE). Likewise, the solvent effect can also be incorporated either by directly coupling an FMO polarizable medium with a continuum solvent or by adding molecular mechanically (MM)

derived solvation free energy to an FMO energy obtained *in vacuo*. In short, there are four ways as the combination to estimate binding free energy depending on the computational cost or accuracy (see Table 1).

<Table 1 here>

Protein-mutative prediction

Jensen et al. proposed a methodology to predict the cleavability of target peptides for the human immunodeficiency virus (HIV)-1 protease based on Chaudhuri and Gray's algorithm [32] with applying FMO-derived binding energies [33]. They estimated the binding energies by using FMO-PCM approach. The energy-only methodology enables us to discriminate the cleavability of peptides for wildtype/mutant HIV-1 proteases and also to predict the capability of mutant HIV-1 protease to cleave an new target peptide as the substrate.

Ligand-binding prediction

Minor structural or chemical changes in ligands sometimes lead to dramatic changes in their activity, which we call the activity cliff. Because the cause of this cliff is critical for controlling the activity, its understanding is a big pharmaceutical issue. Watanabe et al. struggled with this theme on the serine/threonine kinase Pim1 and its inhibitors by using FMO [34]. The six benzofuranone-class inhibitors have only difference in the nitrogen positions of indole ring. The binding free energies were estimated by combining IFIE-sum on ligand in vacuo and solvation free energy obtained from MM-PBSA. While comparing

structural preparation between crystal geometry itself and MM- or QM-optimized one, this approach gives the good correlation with pIC50 (R^2=0.85 as the best) and suggestions on the cliff's mechanism with PIEDA.

Okimoto et al. considered different ways of deriving binding free energy to predict affinity between tankyrase 2 and its inhibitors [35]. They concluded the ability of FMO combined with MM-PBSA approach is the best to rank the binding affinity compared to those of other MM-based approaches. The correlation (R=0.856) was derived for 23 compounds including two classes of common features with different flexibility.

Morao et al. demonstrated the calculated results for several G protein-coupled receptors (GPCRs) [36]. The TIEs values calculated by FMO-DFTB3 in PCM solvent showed good correlations with experimental values in pKi of ten compounds for human β2-adrenoceptor (R^2=0.7833), in pEC50 of ten compounds for κ-opioid receptor (R^2=0.662), and in pIC50 of seven compounds for human P2Y12 receptor (R^2=0.8121). The authors realized tremendous speed up for these calculations by using semi-empirical approach correlated well with FMO2-MP2/6-31G(d).

Okiyama et al recently applied FMO-PBSA with fully polarizable media to predict ligand-binding affinities of ERα [29]. The binding free energies of five bioactive compounds correlated well with their in vitro activities (R=0.990). This study has a great impact in treating agonist/antagonist compounds with different charges.

Miscellaneous

On the structure. In the initial structure where one tries to perform FMO, its artificial strains would be often introduced when the coordinates are determined through experimental analyses or molecular simulations. Many of these causes are attributed to the accuracy of the MM force field. Such strains can be removed by QM relaxation to modify the structure into QM-preferred one. Actually, the ONIOM optimization for the pharmacophores of Pim1 [34] and tankyrase 2 [35] improved their correlation factors. Therefore, QM optimization including heavy atoms using QM/MM or ONIOM will lead to significant refinement of analysis results by mitigating strains, even if one employs an X-ray crystal structure with high resolution. In the ERα study, the positions of hydrogens making critical hydrogen bonds are optimized by taking the pharmacophore out once. If the strain energy remains in the structure, it appears as an increment of the deformation energy of a ligand.

On the dielectricity. When coupling an implicit solvent directly to FMO, there is no need to be aware of the inner (solute) dielectric constant because the solute is a polarizable medium. On the other hand, when combining the solvation free energy of MM-PBSA with FMO, one must arbitrarily set some value of the constant for the MM-PBSA calculation. One never knows the best value in advance, however, it would be recommended to employ unity from the results of several studies [34, 35].

On the nonpolar term and other contributions. While the electrostatic contribution is mainly the subject of debate in implicit solution, the nonpolar one is another key for

accurate prediction of the binding free energy. The empirical SA-based approach in MM-PBSA is one of the effective and proven ways for estimating the nonpolar energy of biomolecular systems. The PCM nonpolar energy is, however, known to have a large disagreement with such an empirically-derived one when applied to large biomolecules [35, 37]. It would be also future issue to incorporate the entropic and thermal corrections by using vibration analysis or estimation from the number of rotatable bonds.

5. Summary

This review article has comprehensively addressed the description of solvation (mainly hydration) effects in the FMO method. When we employ the explicit solvation model by water molecules, the electronic states and associated inter/intra-molecular interactions are significantly affected by water shell with counterions, whose magnitudes are varied depending on the modeling of shell thickness. Whereas one may perform full *ab initio* MD simulations based on the FMO method in the future, we are currently limited to collecting finite numbers of snapshot FMO calculations for the interaction analysis. The SCIFIE analysis then works for incorporating the many-body screening effects into the FMO-IFIE evaluation. The implicit solvation model such as the FMO-PBSA, on the other hand, provides a good compromise between cost and accuracy to obtain the screened IFIEs and the binding free energy for the solvated system in thermodynamic equilibrium, which are compared well with corresponding experimental results.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers 17H06353 and 18K03825.

References

1. Tanaka S, Mochizuki Y, Komeiji Y, Okiyama Y, Fukuzawa K (2014) Electron-correlated fragment-molecular-orbital calculations for biomolecular and nano systems. Phys Chem Chem Phys 16:10310–10344 . doi: 10.1039/c4cp00316k

2. Watanabe C, Fukuzawa K, Tanaka S, Aida-Hyugaji S (2014) Charge clamps of lysines and hydrogen bonds play key roles in the mechanism to fix helix 12 in the agonist and antagonist positions of estrogen receptor α: Intramolecular interactions studied by the ab initio fragment molecular orbital method. J Phys Chem B 118:4993–5008 . doi: 10.1021/jp411627y

3. Komeiji Y, Ishida T, Fedorov DG, Kitaura K (2007) Change in a protein's electronic structure induced by an explicit solvent: An ab initio fragment molecular orbital study of ubiquitin. J Comput Chem 28:1750–1762 . doi: 10.1002/jcc.20686

4. Fukuzawa K, Watanabe C, Kurisaki I, Taguchi N, Mochizuki Y, Nakano T, Tanaka S, Komeiji Y (2014) Accuracy of the fragment molecular orbital (FMO) calculations for DNA: Total energy, molecular orbital, and inter-fragment interaction energy. Comput Theor Chem 1034:7–16 . doi: 10.1016/j.comptc.2014.02.002

5. Fukuzawa K, Kurisaki I, Watanabe C, Okiyama Y, Mochizuki Y, Tanaka S, Komeiji Y (2015) Explicit solvation modulates intra- and inter-molecular interactions within DNA: Electronic aspects revealed by the ab initio fragment molecular orbital (FMO) method. Comput Theor Chem 1054:29–37 . doi: 10.1016/j.comptc.2014.11.020

6. Komeiji Y, Okiyama Y, Mochizuki Y, Fukuzawa K (2017) Explicit solvation of a single-stranded DNA, a binding protein, and their complex: A suitable protocol for fragment molecular orbital calculation. Chem-Bio Informatics J 17:72–84 . doi: 10.1273/cbij.17.72

7. Komeiji Y, Okiyama Y, Mochizuki Y, Fukuzawa K (2018) Interaction between a single-stranded DNA and a binding protein viewed by the fragment molecular orbital method. Bull Chem Soc Jpn 91:1596–1605 . doi: 10.1246/bcsj.20180150

8. Terauchi Y, Suzuki R, Takeda R, Kobayashi I, Kittaka A, Takimoto-Kamimura M, Kurita N (2019) Ligand chirality can affect histidine protonation of vitamin-D receptor: ab initio molecular orbital calculations in water. J Steroid Biochem Mol Biol 186:89–95 . doi: 10.1016/j.jsbmb.2018.09.020

9. Takeda R, Kobayashi I, Shimamura K, Ishimura H, Kadoya R, Kawai K, Kittaka A, Takimoto-Kamimura M, Kurita N (2017) Specific interactions between vitamin-D receptor and its ligands: Ab initio molecular orbital calculations in water. J Steroid Biochem Mol Biol 171:75–79 . doi: 10.1016/j.jsbmb.2017.02.018

10. Takeda R, Kobayashi I, Suzuki R, Kawai K, Kittaka A, Takimoto-Kamimura M, Kurita N (2018) Proposal of potent inhibitors for vitamin-D receptor based on ab initio fragment molecular orbital calculations. J Mol Graph Model 80:320–326 . doi: 10.1016/j.jmgm.2018.01.014

11. Fukuzawa K, Mochizuki Y, Tanaka S, Kitaura K, Nakano T (2006) Molecular Interactions between Estrogen Receptor and Its Ligand Studied by the ab Initio Fragment Molecular Orbital Method. J Phys Chem B 110:24276–24276 . doi: 10.1021/jp065705n

12. Tanaka S, Watanabe C, Okiyama Y (2013) Statistical correction to effective interactions in the fragment molecular orbital method. Chem Phys Lett 556:272–277 . doi: 10.1016/j.cplett.2012.11.085

13. Hansen JP, McDonald IR (2006) Theory of Simple Liquids, 3rd ed. Academic Press, London

14. Jaynes ET (2003) Probability Theory: The Logic of Science. Cambridge University Press, Cambridge

15. Okiyama Y, Tsukamoto T, Watanabe C, Fukuzawa K, Tanaka S, Mochizuki Y (2013) Modeling of peptide-silica interaction based on four-body corrected fragment molecular orbital (FMO4) calculations. Chem Phys Lett 566:25–31 . doi: 10.1016/j.cplett.2013.02.020

16. Kato K, Fukuzawa K, Mochizuki Y (2015) Modeling of hydroxyapatite--peptide interaction based on fragment molecular orbital method. Chem Phys Lett 629:58–64 . doi: 10.1016/j.cplett.2015.03.057

17. Ando H, Shigeta Y, Baba T, Watanabe C, Okiyama Y, Mochizuki Y, Nakano M (2015) Hydration effects on enzyme-substrate complex of nylon oligomer hydrolase: Inter-fragment interaction energy study by the fragment molecular orbital method. Mol Phys 113:319–326 . doi: 10.1080/00268976.2014.941311

18. Ryde U, Söderhjelm P (2016) Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods. Chem Rev 116:5520–5566 . doi: 10.1021/acs.chemrev.5b00630

19. Fedorov DG, Kitaura K, Li H, Jensen JH, Gordon MS (2006) The polarizable continuum model (PCM) interfaced with the fragment molecular orbital method (FMO). J Comput Chem 27:976–985 . doi: 10.1002/jcc.20406

20. Fedorov DG (2018) Analysis of solute--solvent interactions using the solvation model density combined with the fragment molecular orbital method. Chem Phys Lett 702:111–116 . doi: 10.1016/j.cplett.2018.05.002

21. Watanabe H, Okiyama Y, Nakano T, Tanaka S (2010) Incorporation of solvation effects into the fragment molecular orbital calculations with the Poisson--Boltzmann equation. Chem Phys Lett 500:116–119 . doi: 10.1016/j.cplett.2010.10.017

22. Okiyama Y, Nakano T, Watanabe C, Fukuzawa K, Mochizuki Y, Tanaka S (2018) Fragment molecular orbital calculations with implicit solvent based on the Poisson--Boltzmann equation: Implementation and DNA study. J Phys Chem B 122:4457–4471 . doi: 10.1021/acs.jpcb.8b01172

23. Nishimoto Y, Fedorov DG (2016) The fragment molecular orbital method combined with density-functional tight-binding and the polarizable continuum model. Phys Chem Chem Phys 18:22047–22061 . doi: 10.1039/C6CP02186G

24. Simoncini D, Nakata H, Ogata K, Nakamura S, Zhang KYJ (2015) Quality Assessment of Predicted Protein Models Using Energies Calculated by the Fragment Molecular Orbital Method. Mol Inform 34:97–104 . doi: 10.1002/minf.201400108

25. Grimme S (2006) Semiempirical GGA-type density functional constructed with a long-range dispersion correction. J Comput Chem 27:1787–1799 . doi: 10.1002/jcc.20495

26. Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys 132:154104 . doi: 10.1063/1.3382344

27. Fedorov DG, Kitaura K (2012) Energy Decomposition Analysis in Solution Based on the Fragment Molecular Orbital Method. J Phys Chem A 116:704–719 . doi: 10.1021/jp209579w

28. Mazanetz MP, Chudyk EI, Fedorov DG, Alexeev Y (2016) Applications of the fragment molecular orbital method to drug research. In: Zhang W (ed) Computer-Aided Drug Discovery. Methods in Pharmacology and Toxicology. Humana Press, New York, NY, pp 217–255

29. Okiyama Y, Watanabe C, Fukuzawa K, Mochizuki Y, Nakano T, Tanaka S (2019) Fragment Molecular Orbital Calculations with Implicit Solvent Based on the Poisson--Boltzmann Equation: II. Protein and Its Ligand-Binding System Studies. J Phys Chem B 123:957–973 . doi: 10.1021/acs.jpcb.8b09326

30. Fedorov DG, Kitaura K (2016) Subsystem Analysis for the Fragment Molecular Orbital Method and Its Application to Protein--Ligand Binding in Solution. J Phys Chem A 120:2218–2231 . doi: 10.1021/acs.jpca.6b00163

31. Śliwa P, Kurczab R, Kafel R, Drabczyk A, Jaśkowska J (2019) Recognition of repulsive and attractive regions of selected serotonin receptor binding site using FMO-EDA approach. J Mol Model 25:114 . doi: 10.1007/s00894-019-3995-6

32. Chaudhury S, Gray JJ (2009) Identification of Structural Mechanisms of HIV-1 Protease Specificity Using Computational Peptide Docking: Implications for Drug Resistance. Structure 17:1636–1648 . doi: 10.1016/j.str.2009.10.008

33. Jensen JH, Willemoës M, Winther JR, De Vico L (2014) In Silico Prediction of Mutant HIV-1 Proteases Cleaving a Target Sequence. PLoS One 9:e95833 . doi: 10.1371/journal.pone.0095833

34. Watanabe C, Watanabe H, Fukuzawa K, Parker LJ, Okiyama Y, Yuki H, Yokoyama S, Nakano H, Tanaka S, Honma T (2017) Theoretical analysis of activity cliffs among benzofuranone-class Pim1 inhibitors using the fragment molecular orbital method with molecular mechanics Poisson--Boltzmann surface area (FMO+MM-PBSA) approach. J Chem Inf Model 57:2996–3010 . doi: 10.1021/acs.jcim.7b00110

35. Okimoto N, Otsuka T, Hirano Y, Taiji M (2018) Use of the Multilayer Fragment Molecular Orbital Method to Predict the Rank Order of Protein--Ligand Binding Affinities: A Case Study Using Tankyrase 2 Inhibitors. ACS Omega 3:4475–4485 . doi: 10.1021/acsomega.8b00175

36. Morao I, Fedorov DG, Robinson R, Southey M, Townsend-Nicholson A, Bodkin MJ, Heifetz A (2017) Rapid and accurate assessment of GPCR--ligand interactions Using the fragment molecular orbital-based density-functional tight-binding method. J Comput Chem 38:1987–1990 . doi: 10.1002/jcc.24850

37. Söderhjelm P, Kongsted J, Ryde U (2010) Ligand Affinities Estimated by Quantum Chemical Calculations. J Chem Theory Comput 6:1726–1737 . doi: 10.1021/ct9006986

Figure captions

Figure 1. Thickness of water solvent for DNA [5].

Figure 2. Properties of the ssDNA/SSB complex as functions of solvent thickness with and without neutralization [6].

Figure 3. Interaction of ssDNA with the solvent and SSB [6].

Figure 4. Hydrogen bond network among estrogen receptor alpha, ligand, and water molecule (dotted line). The double arrow represents the CH/π interaction. The numbers shown along with the dotted line and double arrow indicate inter-fragment interaction energies (IFIEs) in kcal/mol between amino acid residue and ligand (blue) or water (purple) calculated at the MP2/6-31G(d)//HF/6-31G(d) level.

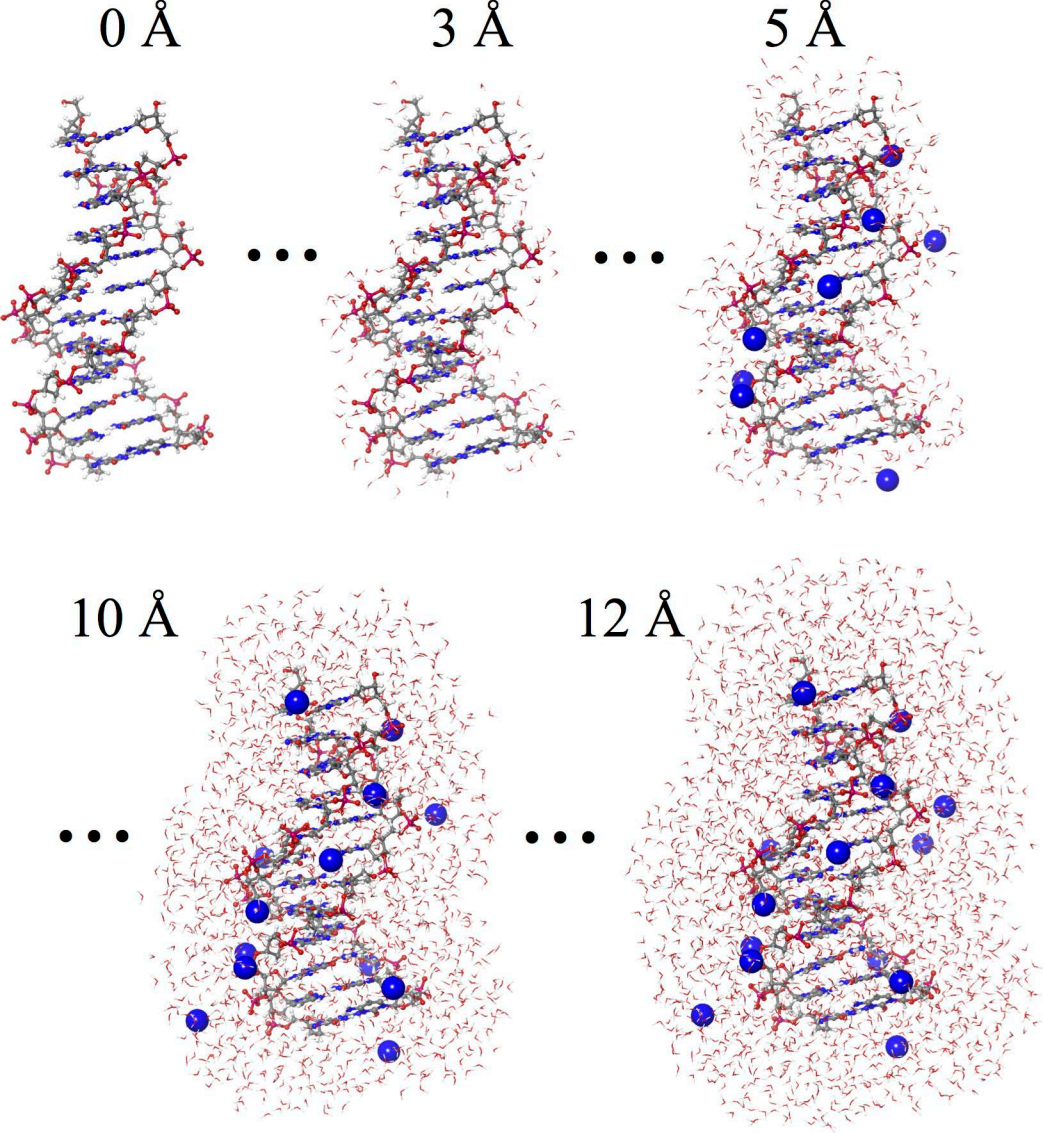
Figure 5. Solvation free energy of dsDNA in an explicit solvent as a function of the shell thickness by comparison with that in an implicit solvent. The number of counterions contained in the explicit solvent is also plotted. Reprinted with permission from [22].

Figure 6. (a) Crystal structure of superchignolin (PDB ID: 5AWL) with modeled hydrogen atoms. Both termini of the structure are in a zwitterionic state. (b) IFIEs of the structure in solution. The FMO-PBSA calculation at the MP2/6-31G(d) level yields the PIEDA components including the solvation term.

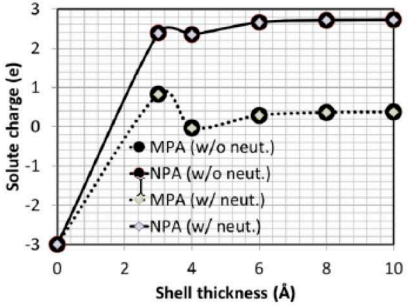
Tables

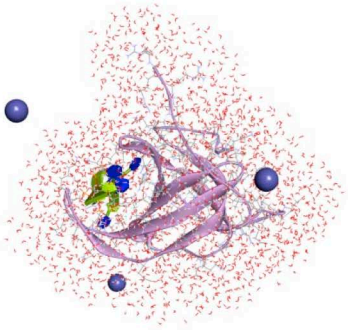
Table 1. *Variation of binding free energy estimation based on FMO.*

Binding energy \ Solvent effect	MM-energy additive	Directly coupled
Interaction-based approach	Solute unpolarized Desolvation included Low-cost Reference [34]	Solute polarized No desolvation Slightly expensive Reference [36]
Supermolecular approach	Solute unpolarized Desolvation included Moderate-cost Reference [35]	Solute polarized Desolvation included Expensive References [29, 33, 35]

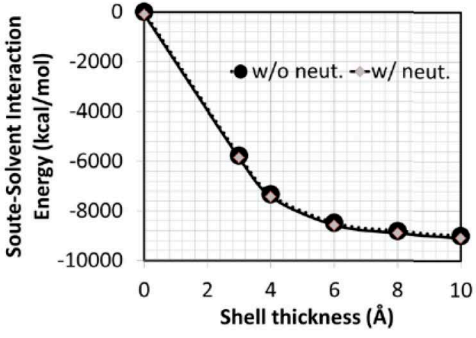

0 Å
3 Å
5 Å
10 Å
12 Å

Atomic charge

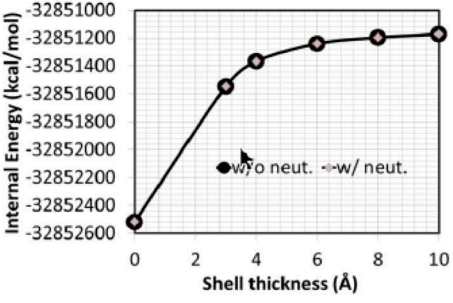




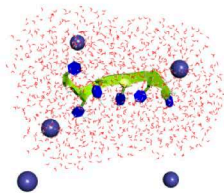
Interaction energy

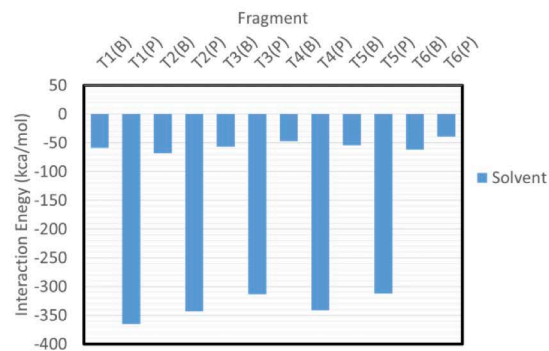


Internal energy

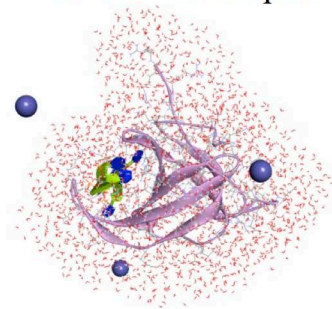


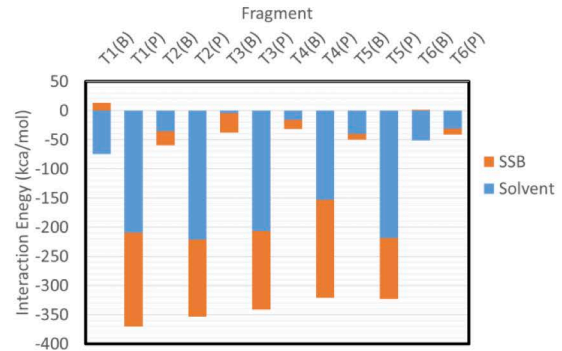
Isolated ssDNA

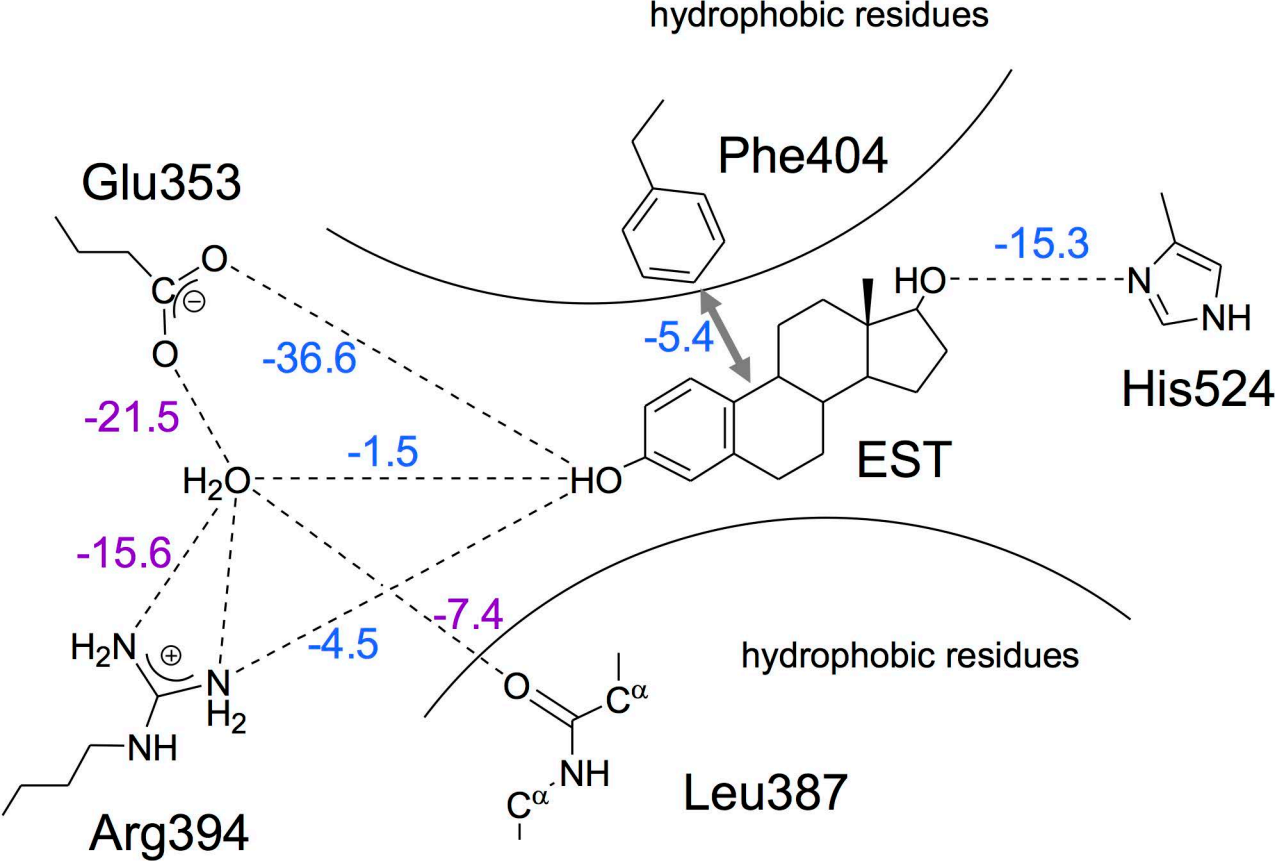


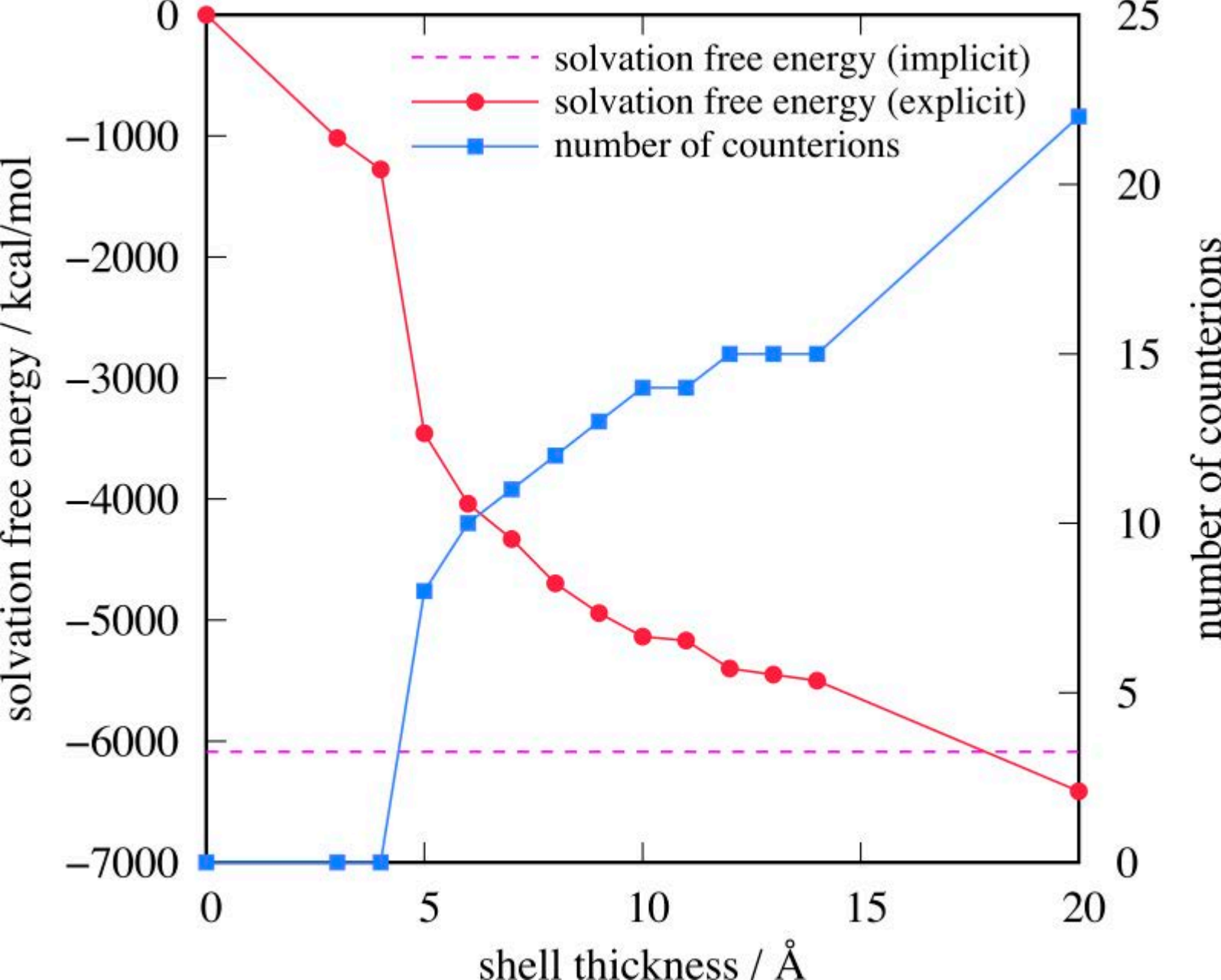


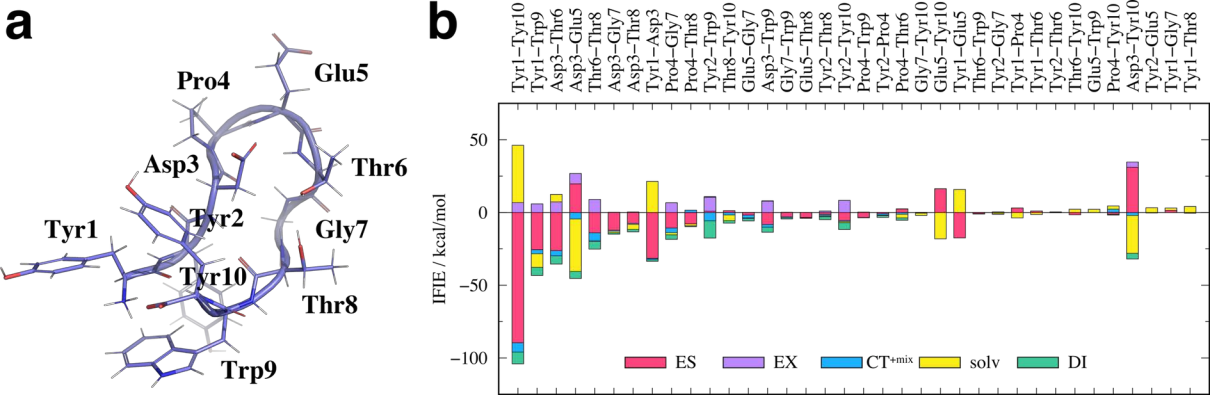
ssDNA in Complex






hydrophobic residues
Phe404
Glu353
-15.3
HO
N
NH
His524
-5.4
-36.6
-21.5
-1.5
H_2O
HO
EST
-15.6
-7.4
H_2N
-4.5
hydrophobic residues
N
H_2
NH
O
C^α
NH
C^α
Leu387
Arg394


solvation free energy (implicit)
solvation free energy (explicit)
number of counterions
solvation free energy / kcal/mol
number of counterions
shell thickness / Å
0
−1000
−2000
−3000
−4000
−5000
−6000
−7000
25
20
15
10
5
0
0
5
10
15
20


a
Pro4
Glu5
Asp3
Thr6
Tyr2
Tyr1
Gly7
Tyr10
Thr8
Trp9
b
Tyr1–Tyr10
Tyr1–Trp9
Asp3–Thr6
Asp3–Glu5
Thr6–Thr8
Asp3–Gly7
Asp3–Thr8
Tyr1–Asp3
Pro4–Gly7
Pro4–Thr8
Tyr2–Trp9
Thr8–Tyr10
Glu5–Gly7
Asp3–Trp9
Gly7–Trp9
Glu5–Thr8
Tyr2–Thr8
Tyr2–Tyr10
Pro4–Trp9
Tyr2–Pro4
Pro4–Thr6
Gly7–Tyr10
Glu5–Tyr10
Tyr1–Glu5
Thr6–Trp9
Tyr2–Gly7
Tyr1–Pro4
Tyr1–Thr6
Tyr2–Thr6
Thr6–Tyr10
Glu5–Trp9
Pro4–Tyr10
Asp3–Tyr10
Tyr2–Glu5
Tyr1–Gly7
Tyr1–Thr8
IFIE / kcal/mol
50
0
−50
−100
ES
EX
CT+mix
solv
DI