

The Effective Fragment Potential: A General Method for Predicting Intermolecular Interactions

Mark S. Gordon^{*}, Lyudmilla Slipchenko^{*}, Hui Li^{**}
and Jan H. Jensen^{***}

Contents	1. Introduction	177
	2. EFP2 Theory	178
	2.1 Contributing interaction terms	178
	2.2 Energy gradients and molecular dynamics	180
	2.3 Interface with continuum	181
	2.4 EFP-QM interactions	182
	2.5 EFP-QM across covalent bonds	182
	3. Example Applications	183
	3.1 Benzene dimer	183
	3.2 Benzene–water	186
	3.3 The prediction of the p <i>K</i> _a value of Lys55 using QM/MM	188
	4. Summary and Future Developments	190
	Acknowledgements	191
	References	191

1. INTRODUCTION

The interactions between molecules and molecular systems play key roles in many important phenomena in chemistry, the biological sciences, materials science and engineering, and chemical and mechanical engineering, among many other disciplines. Important examples include the structure and properties of weakly interacting clusters; the behavior and properties of liquids; solvent effects on ions,

^{*} Department of Chemistry, Iowa State University and Ames Laboratory, Ames, IA 50011, USA

^{**} Department of Chemistry, University of Nebraska, Lincoln, NE 68588, USA

^{***} Department of Chemistry, Universitetsparken 5, University of Copenhagen, 2100 Copenhagen, Denmark

electrolytes, amino acids and other biomolecules, and the mechanisms of chemical reactions; the structures and structure-activity relationships of large molecules like polymers, proteins and enzymes; aggregation of polymers in solution; and the nature of interfacial (e.g., gas-liquid and liquid-solid) phenomena.

Because intermolecular interactions are so important, one needs theoretical methodologies that can accurately account for the broad range of such interactions. The most desirable theoretical approach would be to use a level of quantum mechanics (QM) that can treat both intermolecular and intramolecular (typically covalent) interactions with an acceptable level of accuracy. The range of intermolecular interactions summarized above will involve many different *types* of intermolecular forces, while intramolecular interactions will include the same forces as well as strong covalent bonds. Realistically, the minimal QM levels of theory that can adequately treat all of these phenomena are second order perturbation theory (MP2) [1] and (preferably) coupled cluster (CC) theory with some accounting of triples; i.e., CCSD(T) [2]. Unfortunately, a sufficiently high level of QM comes at a significant computational cost; for example, CCSD(T) scales $\sim N^7$ with problem size, where N is the number of atomic basis functions. This places serious limitations on the sizes of accessible molecular systems.

An alternative approach to the study of intermolecular interactions is to employ a model potential. Such potentials, broadly referred to as molecular mechanics (MM), can generally not account for bond-breaking, but can, in principle, account for the range of intermolecular interactions. If one is concerned with both intermolecular interactions and breaking chemical bonds, a combined QM/MM approach can be used [3]. Ideally, model potentials should be derived from first principles and should contain all of the essential underlying physics.

A particularly promising model potential is the effective fragment potential (EFP) that has been developed by the authors and many co-workers [4,5]. The EFP method was originally developed specifically to describe aqueous solvent effects on biomolecular systems and chemical reaction mechanisms. This EFP1 method contains fitted parameters for the repulsive term and, while very successful for its original purpose [6,7], it is difficult to extend beyond water. Therefore, for the last decade, the authors and co-workers have been developing a more general (EFP2) method that includes all of the essential physics and that has no empirically fitted parameters. The remainder of this work focuses exclusively on the EFP2 method and a few illustrative examples. For simplicity the method will henceforth be referred to as EFP.

2. EFP2 THEORY

2.1 Contributing interaction terms

All of the terms in the EFP method may be thought of as truncated expansions. At present, the EFP interaction energy is a sum of five terms:

$$E = E_{\text{coul}} + E_{\text{ind}} + E_{\text{exrep}} + E_{\text{disp}} + E_{\text{ct}}. \quad (1)$$

Equation (1) specifically refers to EFP-EFP interactions. EFP-QM interactions are discussed later. E_{coul} refers to the Coulomb portion of the electrostatic interaction. This term is obtained using the distributed multipolar expansion introduced by Stone, with the expansion carried out through octopoles. The expansion centers are taken to be the atom centers and the bond midpoints. So, for water, there are five expansion points (three at the atom centers and two at the O–H bond midpoints), while in benzene there are 24 expansion points. E_{ind} is the induction or polarization part of the electrostatic interaction. This term is represented by the interaction of the induced dipole on one fragment with the permanent dipole on another fragment, expressed in terms of the dipole polarizability. Although this is just the first term of the polarizability expansion, it is robust, because the molecular polarizability is expressed as a tensor sum of localized molecular orbital (LMO) polarizabilities. That is, the number of polarizability points is equal to the number of bonds and lone pairs in the molecule. This dipole-induced dipole term is iterated to self-consistency, so some many body effects are included.

The exchange repulsion E_{exrep} is derived as an expansion in the intermolecular overlap. When this overlap expansion is expressed in terms of frozen LMOs on each fragment, the expansion can reliably be truncated at the quadratic term [8]. This term does require that each EFP carries a basis set, and the smallest recommended basis set is 6-31++G(d,p) [9] for acceptable results. Since the basis set is used only to calculate overlap integrals, the computation is very fast and quite large basis sets are realistic. The dispersion interaction can be expressed as the familiar inverse R expansion,

$$E_{\text{disp}} = \sum_n C_n R^{-n}. \quad (2)$$

The coefficients C_n may be derived from the (imaginary) frequency dependent polarizabilities summed over the entire frequency range [10,11]. If one employs only dipole polarizabilities the dispersion expansion is truncated at the leading term, with $n = 6$. In the current EFP code, an estimate is used for the $n = 8$ term, in addition to the explicitly derived $n = 6$ term. Rather than express a molecular C_6 as a sum over atomic interaction terms, the EFP dispersion is expressed in terms of LMO-LMO interactions. In order to ensure that the dispersion interaction goes to zero at short distances, the damping term proposed by Tang and Toennies [12] is employed.

The charge transfer interaction E_{ct} is derived by considering, using a supermolecule approach, the interactions between the occupied valence molecular orbitals on one fragment with the virtual orbitals on another fragment. This leads to significant energy lowering in *ab initio* calculations on ionic or highly polar species when incomplete basis sets are employed. An approximate formula [13] for the charge transfer interaction in the EFP2 method was derived and implemented using a second order perturbative treatment of the intermolecular interactions for a pair of molecules at the Hartree–Fock level of theory. The approximate formula is formulated in canonical orbitals from Hartree–Fock calculations of independent molecules and uses a multipolar expansion (through quadrupoles) of the molecular electrostatic potentials. Orthonormality between the virtual orbitals of the other

molecule to all the orbitals of the considered molecule is enforced so the charge transfer is not contaminated with induction. Implemented in the EFP method, the approximate formula gives charge transfer energies comparable to those obtained from Hartree–Fock calculations. The analytic gradients of the charge transfer energy were also derived and implemented, enabling efficient geometry optimization and molecular dynamics simulations [13].

The two electrostatic terms discussed above, E_{coul} and E_{ind} must be modulated by damping, or screening, expressions. The Coulomb point multipole model breaks down when fragments approach too closely, since then the actual electron density on the two fragments is not well approximated by point multipoles. The latter interactions become too repulsive and must be moderated by a screening term [14]. On the other hand, the induction interaction becomes too attractive if fragments approach each other too closely, so a damping term is needed here as well. To avoid this unphysical behavior, the multipole electrostatic potential is augmented by exponential damping functions $f_{\text{damp}} = 1 - \exp(-\alpha R)$, with parameters α being determined at each multipole expansion point by fitting the multipole damped potential to the Hartree–Fock one. Damping terms in the electrostatic energy are derived explicitly from the damped potential and the charge density. The damping procedure can be extended to higher electrostatic terms, such as charge–dipole, dipole–dipole, etc. Charge–dipole, dipole–dipole, and dipole–quadrupole damping is applied to the polarization energy.

EFPs currently have internally fixed geometries. Analytic gradients for all terms have been derived and implemented, so full geometry optimizations and Monte Carlo and molecular dynamics simulations [15,16] can be performed. Because the method involves no empirically fitted parameters, an EFP for any system can be generated by a “makefp” run in the GAMESS suite of programs.

2.2 Energy gradients and molecular dynamics

Because the analytic gradients of all the interaction terms in the EFP2 method have been derived and implemented, molecular dynamics simulations can be performed [16]. The EFP electrostatic and dispersion interactions have relatively simple expressions, and the corresponding analytic energy gradients (forces and torques) can be derived in a straightforward manner. The exchange repulsion and charge transfer interactions are modeled with approximate formulas that employ MOs from SCF calculations on independent molecules. The MOs and their energies are the EFP parameters for these two interaction terms. The gradients are obtained by differentiating the corresponding energy expressions with respect to molecular rotational and translational displacements. The timings for the gradient evaluation for exchange repulsion and charge transfer are only 1–2 times those for the corresponding energy evaluation [17].

All of the EFP interaction terms are pairwise additive, except the induction (polarization) energy, which is modeled with asymmetric anisotropic polarizability tensors located at the centroids of the localized molecular orbitals. The analytic energy gradients (both forces and torques) for anisotropic polarizability tensors are derived via a direct differentiation approach in the form of matrix equations.

The forces and torques on the polarizability tensors can be evaluated with the induced dipoles (as if they were permanent) and the total electric fields at the polarizability tensors. Once the induced dipoles have been determined, the exact polarization energy gradients can be evaluated analytically at a very low cost [16].

Periodic boundary conditions (PBC) with the minimum image convention (MIC) [18] have been implemented for the EFP method by using the distances between the centers of masses of the (rigid) molecules as the inter-molecular distances [15]. To ensure energy conservation in molecular dynamics simulations, switching functions are applied to modify the intermolecular potential so that the interaction potential energies and forces for molecular pairs smoothly decrease to zero within the periodic cell [19]. The application of the MIC-PBC and switching functions can be rigorously applied to both the pairwise interaction terms and the EFP polarization energy. Using the MIC-PBC and a fifth-order polynomial switching function, very good energy conservation has been realized in molecular dynamics simulations with the EFP method [16].

2.3 Interface with continuum

Low-cost continuum models are often used to assess bulk solvation effects. The polarizable continuum models (PCM) [20] are continuum solvation models in which the solvent effects are described with induced surface charges. In a PCM calculation, the solutes can be modeled with *ab initio* methods or force fields, or both. In a combined QM/EFP/PCM calculation [21], the EFP induced dipoles and PCM induced charges are iterated to self-consistency as the QM wavefunction converges.

The EFP induced dipoles and PCM induced charges may be described by a supermatrix equation [22]:

$$\mathbf{B} \cdot \mathbf{w} = \mathbf{p}. \quad (3)$$

The matrix \mathbf{p} is a combined set of the external electrostatic fields that represent the effects of the QM field on the EFP polarizability tensors and the PCM potential, while \mathbf{w} is a combined set of induced dipoles and surface charges. The physical meaning of the supermatrix equation (3) is that the EFP induced dipoles and PCM induced charges are uniquely determined by the external field and potential; therefore, the right hand side of Eq. (3) involves only the external field/potential, and the left side involves only the induced EFP dipoles and PCM charges. The interactions among the induced dipoles and charges are implicitly described with the matrix \mathbf{B} . The supermatrix Eq. (3) can be solved either with direct inversion or various iterative methods.

The gradients (both forces and torques) of the polarization energy in a combined EFP/PCM calculation have been derived and implemented [22]. It is found that all of the energy gradient terms can be formulated as simple electrostatic forces and torques on the induced dipoles and charges as if they were permanent static dipoles and charges, in accordance with the electrostatic nature of these models. Geometry optimizations can be performed efficiently with the analytic

gradients of the EFP + PCM polarization energies. Due to the intrinsic discontinuity in the molecular surface tessellation in the PCM method, the gradients are not strictly continuous. Although this non-continuity rarely affects geometry optimizations, it prevents good energy conservation in molecular dynamics simulations.

2.4 EFP-QM interactions

The discussion in the foregoing sections has focused primarily on interactions among fragments. In order to have a general method that is able to treat solvent effects on chemical reactions, the analogous EFP-QM interactions are required. The first two interaction terms in Eq. (1), E_{coul} and E_{ind} , have already been developed for the EFP-QM interface [4,5], including energy gradients. A general expression for the QM-EFP exchange repulsion interaction, E_{exch} , has been derived and coded [23], and the corresponding expressions for the energy gradients are in progress [24]. Once these gradients have been implemented, one will have an EFP-QM interface at a level of theory that is comparable to Hartree-Fock. The derivation of the most important remaining component, the dispersion interaction E_{disp} , and its energy gradient has been completed, and the implementation is in progress [25]. This will provide a correlated EFP-QM interface.

2.5 EFP-QM across covalent bonds

Since the fragments are represented by model potentials (EFPs), the method may be considered to be in the general category of QM/MM (quantum mechanics/molecular mechanics) methods. In other contexts QM/MM methods have also been very useful for describing extended systems in which the QM and MM regions are separated by covalent bonds rather than weak intermolecular forces. To make the link between the *ab initio* and MM portions a covalently bonded *ab initio*/EFP interface has been developed [26] and implemented in GAMESS [27]. The method is similar in spirit to that of Assfeld and Rivail [28]. The essential features of the approach are as follows:

- (1) A buffer region consisting of several LMOs, typically surrounding the α -carbon of a given side-chain, is defined as the *ab initio*/EFP boundary. Once the buffer region is defined, these LMOs are obtained by an *ab initio* calculation on all or a subset of the system, projected onto the buffer atom basis functions [29]. These LMOs are subsequently frozen in the EFP calculations by setting select MO Fock matrix elements to zero [30,31]. The *ab initio*/buffer region interactions are calculated by including the exact quantum mechanical Coulomb and exchange operators corresponding to the charge distribution of the buffer region, in the *ab initio* Hamiltonian. This requires calculation of two-electron integrals over basis functions in the buffer region. Since the buffer MOs are frozen, the changes in induction (polarization) contributions from the buffer region are neglected during a geometry optimization of the *ab initio* region. The effect of this approximation on the chemical reaction of interest can be systematically reduced by increasing the size of the *ab initio* region.

- (2) Variational collapse of the *ab initio* wavefunction into the buffer region is avoided by keeping the *ab initio* MOs orthogonal to the buffer LMOs by Schmidt orthogonalization. This is an approximation relative to a full *ab initio* calculation because the MOs are allowed to build up “orthogonality tails” only in the buffer region, not in the EFP region. The associated error can again be systematically reduced by increasing the size of the buffer region.
- (3) The remaining part of the system (or within a defined radius of the active region) is represented by an EFP. The presence of the buffer region provides sufficient separation between the EFP and the *ab initio* regions to ensure that the remaining interactions can be treated as non-bonded interactions via the EFP terms presented above.

3. EXAMPLE APPLICATIONS

3.1 Benzene dimer

The benzene dimer, a prototype for π - π interactions, has attracted extensive theoretical and experimental attention [32–39]. π - π interactions govern structures of proteins and DNA, self-assembly of aromatic macromolecules, and drug-intercalation into DNA. Combined theoretical and experimental studies suggest that there are two minima on the potential energy surface of the benzene dimer: The perpendicular T-shaped and parallel-slipped configurations; the transition state sandwich structure is highest in energy (Figure 10.1). A rotational experiment by Arunan and Gutowsky [35] determined a 4.96 Å separation between the benzene centers of mass in the T-shaped configuration. The binding energy of the dimer was determined to be $D_0 = 1.6 \pm 0.2$ kcal/mol by Krause et al. [37] and as 2.4 ± 0.4 kcal/mol by Grover et al. [36].

Accurate *ab initio* calculations for the benzene dimer require using both an extensive basis set with diffuse functions and a high level of dynamic correlation. Recently, several independent studies were devoted to theoretical investigation of the potential energy surface of the benzene dimer [33,34,38–40]. This work closely follows the analysis presented by Sherrill and co-workers [33,34]. They estimated

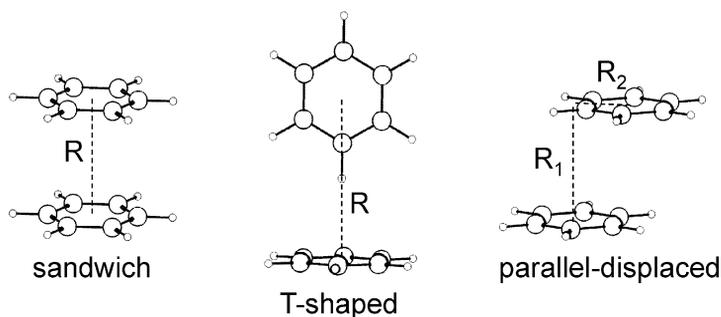


FIGURE 10.1 Sandwich, T-shaped, and parallel-displaced configurations of the benzene dimer.

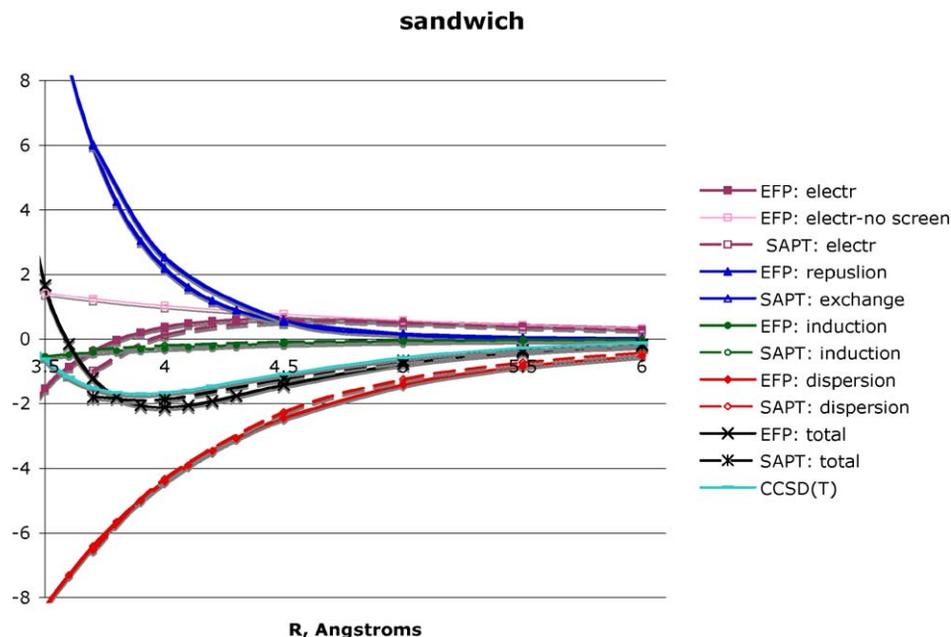


FIGURE 10.2 Comparison of the EFP and SAPT energy components (kcal/mol) in the sandwich benzene dimer. SAPT and CCSD(T) energies from Ref. [33].

potential energy curves for the dimer by the coupled-cluster method including singles and doubles with perturbative triples corrections [CCSD(T)] [2], using the aug-cc-pVQZ basis set. They also analyzed the nature of the π - π interactions by using symmetry-adapted perturbation theory (SAPT) [41]. To analyze the quality of the EFP results for the benzene dimer, the EFP and SAPT/aug-cc-pVDZ potential curves for each energy term are compared separately, as well as the EFP and CCSD(T) total binding energy curves.

The EFP potential for benzene was constructed using the 6-311++G(3df,2p) basis set, at the MP2/aug-cc-pVTZ [42] monomer geometry from Ref. [34]. Multipoles were generated using a numerical DMA, with high-order electrostatic screening [14].

Figures 10.2 and 10.3 present comparisons of the EFP and SAPT results for electrostatic, exchange-repulsion, polarization, and dispersion terms. The total EFP, SAPT, and CCSD(T) binding energies are also plotted. Figure 10.2 shows the potential energy curves for the sandwich configuration and Figure 10.3 gives the T-shaped curves. The equilibrium inter-monomer distances (defined in Figure 10.1) in the benzene dimer vary from $R = 3.7$ Å to 4.0 Å in the sandwich and from $R = 4.9$ Å to 5.1 Å in the T-shaped configurations depending on the level of theory and basis set used. The CCSD(T)/aug-cc-pVQZ values are $R = 3.9$ Å and $R = 5.0$ Å, respectively. These intermolecular separations are used as the reference values in the following discussion. Note that the sandwich and T-shaped configurations used here are not minima on the potential energy surface of the dimer;

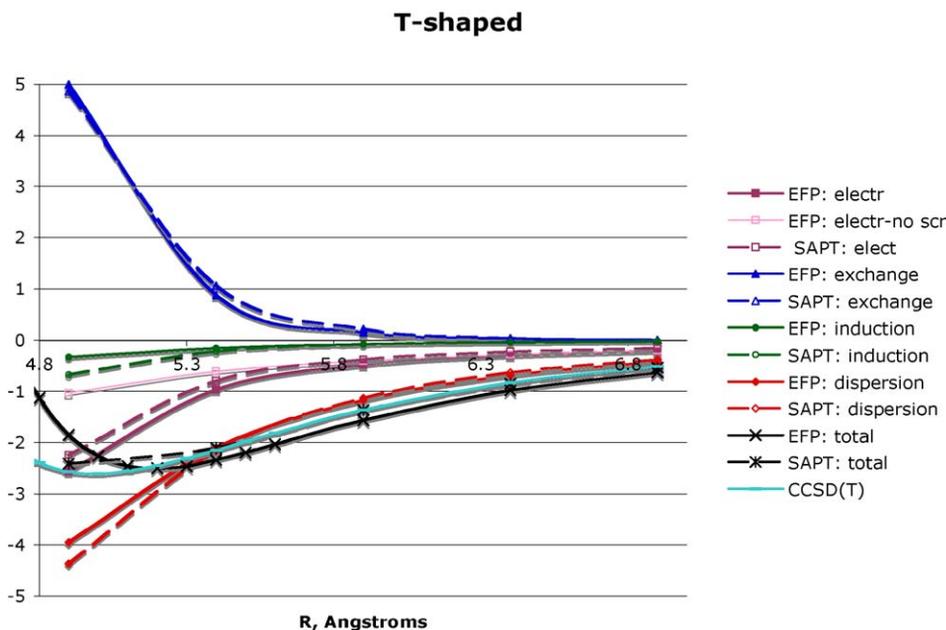


FIGURE 10.3 Comparison of the EFP and SAPT energy components (kcal/mol) in the T-shaped benzene dimer. SAPT and CCSD(T) energies from Ref. [33].

these structures were chosen as representative ones. In the real T-shaped-like minimum, the “upper” benzene is tilted, so that one C–C bond is almost parallel to the plane of the “lower” benzene (see orientation of the benzenes in Figure 10.1). The sandwich structure is a transition state between different parallel-displaced configurations [39].

The electrostatic curves are plotted both with and without damping. The EFP curves without damping underestimate the electrostatic interaction in the equilibrium region by >1 kcal/mol. The damping correction accounts for most of the charge-penetration energy, so that the damped EFP energies differ from the SAPT values by 0.2–0.3 kcal/mol at the equilibrium geometries. Electrostatic damping is very important for a system like the benzene dimer, because both equilibrium geometries and binding energies would be significantly in error without the damping term.

For each term in the interaction energy, EFP is in excellent agreement with SAPT, generally within less than 0.5 kcal/mol and often much better. Overall, EFP over-binds the sandwich dimer by about 0.4 kcal/mol and under-binds the T-shaped structure by 0.1 kcal/mol, as compared to CCSD(T). The equilibrium intermolecular separations calculated by EFP are 0.1–0.2 Å longer than those calculated by CCSD(T).

Table 10.1 summarizes the interaction energies of the three structures of the dimer calculated by MP2, CCSD(T), and EFP. Relative to CCSD(T) with the same basis set, MP2 underestimates the equilibrium distances by 0.1–0.2 Å and over-

TABLE 10.1 Equilibrium geometries (angstroms) and interaction energies (kcal/mol) for different configurations of the benzene dimer

Method	Basis	Sandwich		T-shaped		Parallel-displaced		
		R	Energy	R	Energy	R ₁	R ₂	Energy
MP2 ^a	aug-cc-pVDZ ^b	3.8	-2.83	5.0	-3.00	3.4	1.6	-4.12
	aug-cc-pVTZ	3.7	-3.25	4.9	-3.44	3.4	1.6	-4.65
	aug-cc-pVQZ ^b	3.7	-3.35	4.9	-3.48	3.4	1.6	-4.73
CCSD(T) ^a	aug-cc-pVDZ ^b	4.0	-1.33	5.1	-2.24	3.6	1.8	-2.22
	aug-cc-pVQZ ^b	3.9	-1.70	5.0	-2.61	3.6	1.6	-2.63
EFP	6-311++G (3df,2p)	4.0	-2.11	5.2	-2.50	3.8	1.2	-2.34

^a Reference [33].

^b Basis sets as described in Ref. [33].

estimates the binding energies by 0.7–2.1 kcal/mol. The best agreement between MP2 and CCSD(T) is for the T-shaped structure, while the worst is for the parallel-displaced configuration. EFP overestimates the inter-monomer separations by 0.1–0.2 Å, and inaccuracies in the interaction energies are 0.1–0.4 kcal/mol. In general, the agreement between the EFP and CCSD(T) methods is very reasonable, and EFP is in better agreement with CCSD(T) than is MP2. This is striking in view of the orders of magnitude less computer time required by EFP. For example, a single-point energy calculation in the 6-311++G(3df,2p) basis set (660 basis functions) by MP2 requires 142 minutes of CPU time on one IBM Power5 processor, whereas the analogous EFP calculation requires only 0.4 seconds.

3.2 Benzene–water

Interactions of aromatic molecules with solvent are of fundamental interest, since these interactions are common in bio-systems. The simplest systems of this type, small benzene–water complexes, have attracted both experimental and theoretical attention [43–53]. Zwier and coworkers [45–47] have presented accurate IR data on (benzene)_{1–2} (water)_{1–8} complexes. Accurate assignments would provide unambiguous insight on the structures of these clusters. However, accurate theoretical investigation of these clusters is still very challenging. Both an extensive basis set with diffuse functions and a high level of dynamic correlation are required for an accurate treatment of these systems [43]. Moreover, binding in water–benzene complexes is complicated, since both electrostatic interactions, for hydrogen-bonded water and benzene π – π interactions play a role. Thus, an accurate analysis of these systems requires a balanced description of the different types of intermolecular interactions.

In this example application, the predictions of EFP and accurate *ab initio* methods are compared with experimental values for the water–benzene dimer, with a focus on the question, what is the nature of the water–benzene interaction?

TABLE 10.2 Intermolecular distances (angstroms) and binding energies (kcal/mol) in the water–benzene dimer

Method	Ref.	R_e^a	D_e^b	D_0
MP2/aVTZ	Feller [43]	3.21	−4.01 (−3.13)	−3.01 (−2.13) ^c
CCSD(T)/aVTZ	Feller [43]		−3.85	−2.85 ^c
Est. MP2/CBS	Feller [43]		−3.9 ± 0.2	−2.9 ± 0.2 ^c
EFP	This work	3.38	−3.90	−2.87
Expt.	Gotch, Zwier [45]	3.32		−1.63 to −2.78
	Suzuki et al. [52]	3.35		
	Gutowksy et al. [51]	3.33		
	Cheng et al. [49]			2.25 ± 0.28
	Courty et al. [50]			2.44 ± 0.09

a Distance between the water and benzene centers of mass.

b Values in parentheses correspond to counterpoise-corrected (CP) binding energies.

c Using estimated zero-point vibrational energy from Ref. [43].

EFP parameters for water and benzene were obtained with the 6-311++G(3df,2p) basis set. The benzene EFP is the same as that discussed above. Electrostatic parameters for water were obtained with a numerical DMA, and electrostatic interactions were screened by the high-order electrostatic damping functions. Additionally, for both benzene and water, polarization interactions were screened as well, with damping parameters at all centers being set to 1.5.

Table 10.2 summarizes the experimental and theoretical intermolecular distances and binding energies for the lowest energy structure of the water–benzene dimer. Experimental values of the intermolecular distances (determined as the distance between the centers of mass in water and benzene) are 3.32–3.35 Å, so MP2 underestimates these values by more than 0.1 Å, whereas EFP overestimates them by about 0.05 Å. There is significant disagreement in the measured binding energies in the benzene dimer, e.g., the binding energy was determined to be 1.63–2.78 kcal/mol in Ref. [45], 2.25 ± 0.28 kcal/mol in Ref. [49] and 2.44 ± 0.09 kcal/mol in Ref. [50]. Theoretical binding energies dramatically depend on the method and basis set used.

Even though there is no accurate estimate of the CCSD(T)/complete basis set (CBS) binding energy in the literature, it seems that MP2 generally overestimates the CCSD(T) binding by about 0.2 kcal/mol. This results in an estimated 2.7 kcal/mol CCSD(T) binding energy, in reasonable agreement with the available experimental data. The EFP binding energy in the water–benzene dimer is 3.9 kcal/mol, which, when combined with the EFP ZPE value of 1.03 kcal/mol, results in a net binding of 2.87 kcal/mol. This agrees very well with the MP2/CBS limit and over-binds the experimental [CCSD(T)] values by about 0.5 [0.2] kcal/mol.

Figure 10.4 compares binding in the water dimer and benzene dimers with that in the water–benzene dimer. The energy components were calculated by EFP

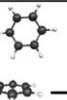
	electrost.	exch.- repulsion	induction	dispers.	total binding	total+ ZPE
	-8.59	5.33	-0.99	-0.85	-5.10	-2.57
	-3.85	2.32	-0.57	-1.81	-3.90	-2.87
	-3.91	3.21	-0.43	-1.44	-2.58	-1.59
	-0.10	2.86	-0.27	-4.85	-2.38	-2.02
	-1.82	2.39	-0.21	-3.21	-2.86	-2.44

FIGURE 10.4 Binding in water dimer, benzene dimer, and water–benzene dimer by EFP. All values are in kcal/mol.

at the EFP equilibrium geometries for each dimer. It is well known that the dominant contribution to binding in the water dimer is the electrostatic interaction (-8.6 kcal/mol), whereas the polarization and dispersion interactions are almost 10 times weaker. Contrarily, binding in the parallel-displaced benzene dimer is dominated by dispersion forces (-4.9 kcal/mol). The T-shaped benzene dimer has significant contributions from both dispersion and electrostatic forces, and, not surprisingly, this is also true for the two structures of the benzene–water dimer. It is also educational to compare the total binding energies in the dimers shown in Figure 10.4. The water dimer is the most strongly bound, the benzene dimers have the weakest interaction energies, but the water–benzene dimer is in between. However, including ZPEs makes the situation less obvious. The ZPE values are largest in the water dimer and the smallest in the benzene dimers, benzene–water dimers again being in the middle. This results in a striking observation that the ZPE-corrected binding energies of the dimers are much less spread energetically. Thus, the immiscibility of benzene in water is due to an unfavorable entropy, rather than enthalpy, contribution.

3.3 The prediction of the pK_a value of Lys55 using QM/MM

The approach taken in this work to predict the pK_a of an amine group in the protein, HA^+ , is related to the standard free energy change, ΔG , due to proton transfer to a reference compound (using Lys residues as an example),

$$pK_a = 10.60 + \{[G(A) - G(HA^+)] - [G(CH_3NH_2) - G(CH_3NH_3^+)]\} / 1.36. \quad (4)$$

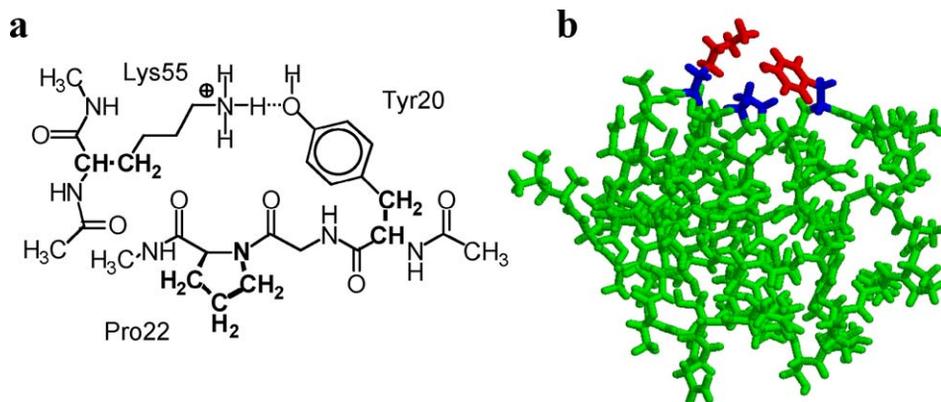


FIGURE 10.5 (a) Subsystem of OMTKY3 used to obtain the buffer region (bold) used for (b) *ab initio*/buffer/EFP regions (red/blue/green) used for the computation of the pK_a of Lys55.

Here, 10.60 is the experimentally determined pK_a of methylamine at 298 K [54] and 1.36 is $RT \ln 10$ for $T = 298$ K in kcal/mol. $G(X)$ is the total free energy (in kcal/mol) of molecule X , which is the sum of the ground state electronic energy (E_{ele}), thermochemical energy (G_{trv}), and solvation energy (G_{sol}):

$$G = E_{ele} + G_{trv} + G_{sol}. \quad (5)$$

The solution structure of OMTKY3 has been determined using NMR by Hoogstraeten et al. [55], and was obtained from the Protein Data Bank (entry 1OMU). The first of the 50 conformers is used without further refinement of the overall structure.

The electronic and geometric structures of the Lys55 and Tyr20 side chains are treated quantum mechanically at the MP2/6-31+G(2d,p)//RHF/6-31G(d) level of theory (Figure 10.5), while the rest of the protein is treated with an EFP, described in more detail below. The use of the diffuse functions on atoms near the buffer region causes SCF convergence problems due to couplings with the induced dipoles in the EFP region, so the 6-31+G(2d,p) basis set was used only for the $C^{\delta}H_2C^{\epsilon}H_2NH_3 \cdots HO-C^{\xi}(C^{\epsilon 1,2}H)_2$ atoms in the MP2 calculation.

The *ab initio* region is separated from the protein EFP by a buffer region [26] comprised of frozen LMOs corresponding to all the bond LMOs connecting the bold atoms in Figure 10.5, as well as the core and lone pair LMOs belonging to those atoms. The Pro22 buffer is needed to describe its short-range interactions with Tyr20 [56]. The buffer LMOs are generated by an RHF/6-31G(d) calculation on a subset of the system (shown in Figure 10.5), projected onto the buffer atom basis functions, and subsequently frozen in the EFP calculations by setting select off-diagonal MO Fock matrix elements to zero. The *ab initio*/buffer region interactions are calculated *ab initio*, and thus include short-range interactions.

The EFP describing the rest of the protein is generated by nine separate *ab initio* calculations on overlapping pieces of the protein truncated by methyl groups. Two different regions of overlap are used depending on whether it occurs on the

protein backbone or on a disulfide bridge, as described in Ref. [56]. The electrostatic potential of each protein piece is expanded in terms of multipoles through octupoles centered at all atomic and bond midpoint centers using Stone's Distributed Multipole Analysis [57]. The monopoles of the entire EFP are scaled to ensure a net integer charge and the dipole polarizability tensor due to each LMO in the EFP region is calculated by a perturbation expression, as described in Ref. [56].

The vibrational free energy (G^{vib}) of the optimized part of the *ab initio* region is calculated by the Partial Hessian Vibrational Analysis (PHVA) method [58]. The solvation energy (ΔG_s) is calculated with the EFP/PCM interface developed by Bandyopadhyay et al. [59] for small molecules and extended to macromolecules by Li, Pomelli, and Jensen [60,61].

The $\text{p}K_a$ of Lys55 computed using this approach is 11.4 pH units, in good agreement with the experimental value [62] of 11.1 considering the uncertainty in the experimental values is roughly ± 0.1 pH units. A more thorough summary of this work can be found in reference [63], and further application of this general approach to other residue types and proteins can be found in [64–67].

4. SUMMARY AND FUTURE DEVELOPMENTS

As illustrated in this work, the effective fragment potential is an accurate method for treating the broad range of intermolecular interactions, at a small fraction of the cost of *ab initio* calculations that produce comparable accuracy. Because no empirically fitted parameters are required, an EFP can easily be generated automatically for any closed shell species simply by running the appropriate GAMESS calculation on the isolated molecule. The long-term goal of the method is to use it both as a stand-alone method to study intermolecular interactions and as an interface with electronic structure methods to provide a sophisticated QM/MM approach to such phenomena as solvent effects on chemical reactions and solvent-induced spectroscopic shifts. To attain this goal, several new features are in progress.

While the leading R^{-6} term in the dispersion expansion generally accounts for the largest part of the dispersion interaction, higher order terms are clearly needed when polar or ionic species are present in the system of interest. The current method has an estimate for the R^{-8} term¹¹, but a more rigorous accounting for this and other terms is needed. At present, an EFP can only be generated for closed shell species, but one can imagine many instances in which an EFP for an open shell compound (e.g., radical) would be useful. A preliminary open shell EFP version is nearly completed and will be made available shortly [24]. At present, the geometry of an EFP is internally frozen. This is not unreasonable for simple molecules like water, but if one wishes to use EFPs to study biomolecules or polymers, it is desirable to at least allow the molecule to relax along torsional coordinates.

As discussed in section 2.4, several new developments are in progress in order to make the EFP-QM interface fully viable. In addition, EFP interfaces are being built with methods that can treat excited electronic states. In addition to the existing MCSCF interface, these include CI singles and time-dependent density functional theory in the short term and more sophisticated CI and coupled cluster methods in the longer term.

ACKNOWLEDGEMENTS

The development of the effective fragment potential method has been supported by both the US Air Force Office of Scientific Research and by the US Department of Energy through its Scientific Discovery through Advanced Computing (SciDAC) initiative. JHJ acknowledges past and present support from the US National Science Foundation (MCB 209941) and a Skou Fellowship to JHJ from the Danish Research Agency (Forskningsrådet for Natur og Univers).

REFERENCES

1. Møller, C., Plesset, M.S. *Phys. Rev.* 1934, 46, 618;
2. Pople, J.A., Binkley, J.S., Seeger, R. *Int. J. Quantum Chem.* 1976, S10, 1.
3. Paldus, J., in Wilson, S., editor. *Handbook of Molecular Physics and Quantum Chemistry*, vol. 2. Chichester: Wiley; 2003, p. 272–313.
4. Friesner, R.A., Guallar, V. *Ann. Rev. Phys. Chem.* 2005, 56, 389.
5. Day, P.N., Jensen, J.H., Gordon, M.S., Webb, S.P., Stevens, W.J., Krauss, M., Garmer, D., Basch, H., Cohen, D. *J. Chem. Phys.* 1996, 105, 1968.
6. Gordon, M.S., Freitag, M.A., Bandyopadhyay, P., Kairys, V., Jensen, J.H., Stevens, W.J. *J. Phys. Chem.* 2001, 105, 293.
7. Webb, S.P., Gordon, M.S. *J. Phys. Chem.* 1999, A103, 1265.
8. Adamovic, I., Gordon, M.S. *J. Phys. Chem.* 2005, 109, 1629.
9. Jensen, J.H., Gordon, M.S. *Mol. Phys.* 1996, 89, 1313.
10. Hehre, W.J., Ditchfield, R., Pople, J.A. *J. Chem. Phys.* 1972, 56, 2257.
11. Amos, R.D., Handy, N.C., Knowles, P.J., Rice, J.E., Stone, A.J. *J. Phys. Chem.* 1985, 89, 2186.
12. Adamovic, I., Gordon, M.S. *Mol. Phys.* 2005, 103, 379.
13. Tang, K.T., Toennies, J.P. *J. Chem. Phys.* 1984, 80, 3726.
14. Li, H., Gordon, M.S., Jensen, J.H. *J. Chem. Phys.* 2006, 124, 214107.
15. Slipchenko, L., Gordon, M.S. *J. Comp. Chem.* 2006, 28, 276.
16. Netzloff, H., Gordon, M.S. *J. Chem. Phys.* 2004, 121, 2711.
17. Li, H., Netzloff, H.M., Gordon, M.S. *J. Chem. Phys.* 2006, 125, 194103.
18. Li, H., Gordon, M.S. *Theor. Chem. Accts.* 2006, 115, 385.
19. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. *J. Chem. Phys.* 1953, 21, 1087.
20. Leach, A.R. *Molecular Modelling—Principles and Applications*, 2nd edn. Harlow: Prentice Hall; 2001.
21. Cancès, E., Mennucci, B., Tomasi, J. *J. Chem. Phys.* 1997, 107, 3032;
22. Barone, V., Cossi, M. *J. Phys. Chem. A* 1998, 102, 1995.
23. Bandyopadhyay, P., Mennucci, B., Tomasi, J., Gordon, M.S. *J. Chem. Phys.* 2002, 116, 5023.
24. Li, H., Gordon, M.S. *J. Chem. Phys.* 2007, 126, 124112.
25. Jensen, J.H. *J. Chem. Phys.* 2001, 114, 8775.
26. Kemp, D., Gordon, M.S., in preparation.
27. Smith, T., Gordon, M.S., in preparation.
28. Kairys, V., Jensen, J.H. *J. Phys. Chem. A* 2000, 104, 6656.
29. Schmidt, M.W., Baldridge, K.K., Boatz, J.A., Elbert, S.T., Gordon, M.S., Jensen, J.H., Koseki, S., Matsunaga, N., Nguyen, K.A., Su, S., Windus, T.L., Dupuis, M., Montgomery Jr., J.A. *J. Comput. Chem.* 1993, 14, 1347;
30. Gordon, M.S., Schmidt, M.W., in Dykstra, C.E., Frenking, G., Kim, K.S., Scuseria, G.E., editors. *Theory and Applications of Computational Chemistry*. Elsevier, 2005. Ch. 41.
31. Assfeld, X., Rivail, J.L. *Chem. Phys. Lett.* 1996, 263, 100.
32. King, H.F., Stanton, R.E., Kim, H., Wyatt, R.E., Parr, R.G. *J. Chem. Phys.* 1967, 47, 1936.
33. Bagus, P.S., Hermann, K., Bauschlicher, C.W. *J. Chem. Phys.* 1984, 80, 4378.

31. Stevens, W.J., Fink, W.H. *Chem. Phys. Lett.* 1987, 139, 15.
32. Hobza, P., Selzle, H.L., Schlag, E.W. *J. Am. Chem. Soc.* 1994, 116(8), 3500;
Hobza, P., Selzle, H.L., Schlag, E.W. *J. Phys. Chem.* 1996, 100(48), 18790;
Jaffe, R.L., Smith, G.D. *J. Chem. Phys.* 1996, 105(7), 2780;
Park, Y.C., Lee, J.S. *J. Phys. Chem. A* 2006, 110, 5091;
Spirko, V., Engvist, O., Soldan, P., Selzle, H.L., Schlag, E.W., Hobza, P. *J. Chem. Phys.* 1999, 111(2), 572;
Tsuzuki, S., Honda, K., Uchimaru, T., Mikami, M., Tanabe, K. *J. Am. Chem. Soc.* 2002, 124(1), 104;
Tsuzuki, S., Uchimaru, T., Sugawara, K., Mikami, M. *J. Chem. Phys.* 2002, 117(24), 11216;
Bornsen, K.O., Selzle, H.L., Schlag, E.W. *J. Chem. Phys.* 1986, 85(4), 1726;
Felker, P.M., Maxton, P.M., Schaeffer, M.W. *Chem. Rev.* 1994, 94(7), 1787;
Janda, K.C., Hemminger, J.C., Winn, J.S., Novick, S.E., Harris, S.J., Klemperer, W. *J. Chem. Phys.* 1975, 63(4), 1419;
Law, K.S., Schauer, M., Bernstein, E.R. *J. Chem. Phys.* 1984, 81(11), 4871;
Scherzer, W., Kratzschmar, O., Selzle, H.L., Schlag, E.W. *Zeitschrift Fur Naturforschung Section a-a Journal of Physical Sciences* 1992, 47(12), 1248;
Steed, J.M., Dixon, T.A., Klemperer, W. *Journal of Chemical Physics* 1979, 70(11), 4940;
Puzder, A., Dion, M., Langreth, D.C. *J. Chem. Phys.* 2006, 124(16);
Sato, T., Tsuneda, T., Hirao, K. *J. Chem. Phys.* 2005, 123(10).
33. Sinnokrot, M.O., Sherrill, C.D. *J. Phys. Chem. A* 2004, 108(46), 10200.
34. Sinnokrot, M.O., Valeev, E.F., Sherrill, C.D. *J. Am. Chem. Soc.* 2002, 124(36), 10887.
35. Arunan, E., Gutowsky, H.S. *J. Chem. Phys.* 1993, 98(5), 4294.
36. Grover, J.R., Walters, E.A., Hui, E.T. *J. Phys. Chem.* 1987, 91(12), 3233.
37. Krause, H., Ernstberger, B., Neusser, H.J. *Chem. Phys. Lett.* 1991, 184(5–6), 411.
38. Jung, Y., Head-Gordon, M. *Phys. Chem. Chem. Phys.* 2006, 8(24), 2831.
39. Podeszwa, R., Bukowski, R., Szalewicz, K. *J. Phys. Chem. A* 2006, 110(34), 10345.
40. Park, Y.C., Lee, J.S. *J. Phys. Chem. A* 2006, 110(15), 5091;
Piecuch, P., Kucharski, S.A., Bartlett, R.J. *J. Chem. Phys.* 1999, 110(13), 6103.
41. Jeziorski, B., Moszynski, R., Szalewicz, K. *Chem. Rev.* 1994, 94(7), 1887.
42. Dunning, T.H. *J. Chem. Phys.* 1989, 90(2), 1007;
Kendall, R.A., Dunning, T.H., Harrison, R.J. *J. Chem. Phys.* 1992, 96(9), 6796.
43. Feller, D. *J. Phys. Chem. A* 1999, 103(38), 7558.
44. Fredericks, S.Y., Pedulla, J.M., Jordan, K.D., Zwier, T.S. *Theor. Chem. Accts.* 1997, 96(1), 51.
45. Gotch, A.J., Zwier, T.S. *J. Chem. Phys.* 1992, 96(5), 3388.
46. Gruenloh, C.J., Carney, J.R., Arrington, C.A., Zwier, T.S., Fredericks, S.Y., Jordan, K.D. *Science* 1997, 276(5319), 1678;
Gruenloh, C.J., Carney, J.R., Hagemester, F.C., Arrington, C.A., Zwier, T.S., Fredericks, S.Y., Wood, J.T., Jordan, K.D. *J. Chem. Phys.* 1998, 109(16), 6601;
Pribble, R.N., Garrett, A.W., Haber, K., Zwier, T.S. *J. Chem. Phys.* 1995, 103(2), 531;
Pribble, R.N., Zwier, T.S. *Faraday Discussions* 1994, 97, 229.
47. Pribble, R.N., Zwier, T.S. *Science* 1994, 265(5168), 75.
48. Tarakeshwar, P., Choi, H.S., Lee, S.J., Lee, J.Y., Kim, K.S., Ha, T.K., Jang, J.H., Lee, J.G., Lee, H. *J. Chem. Phys.* 1999, 111(13), 5838;
Tsuzuki, S., Honda, K., Uchimaru, T., Mikami, M., Tanabe, K. *J. Am. Chem. Soc.* 2000, 122(46), 11450;
Baron, M., Kowalewski, V.J. *J. Phys. Chem. A* 2006, 110(22), 7122;
Jedlovsky, P., Kereszturi, A., Horvai, G. *Faraday Discussions* 2005, 129, 35.
49. Cheng, B.M., Grover, J.R., Walters, E.A. *Chem. Phys. Lett.* 1995, 232(4), 364.
50. Courty, A., Mons, M., Dimicoli, N., Piuze, F., Gageot, M.P., Brenner, V., de Pujo, P., Millie, P. *J. Phys. Chem. A* 1998, 102(33), 6590.
51. Gutowsky, H.S., Emilsson, T., Arunan, E. *J. Chem. Phys.* 1993, 99(7), 4883.
52. Suzuki, S., Green, P.G., Bumgarner, R.E., Dasgupta, S., Goddard, W.A., Blake, G.A. *Science* 1992, 257(5072), 942.
53. Day, P.N., Pachter, R., Gordon, M.S., Merrill, G.N. *J. Chem. Phys.* 2000, 112(5), 2063.
54. CRC Handbook of Chemistry and Physics, 83 edn. Cleveland, OH: CRC Press; 2003.
55. Hoogstraten, C.G., Choe, S., Westler, W.M., Markley, J.L. *Protein Science* 1995, 4, 2289.

56. Minikis, R.M., Kairys, V., Jensen, J.H. *J. Phys. Chem. A* 2001, 105, 3829.
57. Stone, A.J., Price, S.L. *J. Phys. Chem.* 1988, 92, 3325.
58. Li, H., Jensen, J.H. *Theor. Chem. Accts.* 2002, 107, 211.
59. Bandyopadhyay, P., Gordon, M.S., Mennucci, B., Tomasi, J.J. *Chem. Phys.* 2002, 116, 5023.
60. Li, H., Pomelli, C.S., Jensen, J.H. *Theor. Chem. Accts.* 2003, 109, 71.
61. Li, H., Jensen, J.H. *J. Comp. Chem.* 2004, 25, 1449.
62. Forsyth, W.R., Gilson, M.L., Antosiewicz, J., Jaren, O.R., Robertson, A.D. *Biochem.* 1998, 37, 8643.
63. Jensen, J.H., Li, H., Robertson, A.D., Molina, P.A. *J. Phys. Chem. A* 2005, 109, 6634.
64. Li, H., Robertson, A.D., Jensen, J.K.H. *Proteins-Structure Function and Bioinformatics* 2004, 55, 689.
65. Naor, M.M., Jensen, J.H. *Proteins-Structure Function and Bioinformatics* 2004, 57, 799.
66. Porter, M.A., Hall, J.R., Locke, J.C., Jensen, J.H., Molina, P.A. *Proteins-Structure Function and Bioinformatics* 2006, 63, 621.
67. Wang, P.F., Flynn, A.J., Naor, M.M., Jensen, J.H., Cui, G.L., Merz, K.M., Kenyon, G.L., McLeish, M.J. *Biochem.* 2006, 45, 11464.