JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

A New Tabu-Search-Based Algorithm for Solvation of Proteins

Christoph Grebner,[†] Johannes Kästner,[‡] Walter Thiel,[§] and Bernd Engels[†,*]

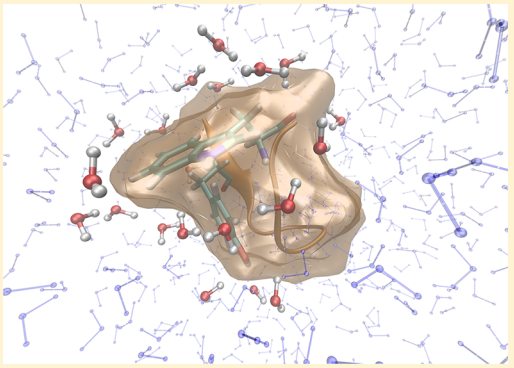
[†]Julius-Maximilians-Universität Würzburg, Institut für Physikalische und Theoretische Chemie, Emil-Fischer-Straße 42, D-97074 Würzburg, Germany

[‡]Universität Stuttgart, Institut für Theoretische Chemie, Pfaffenwaldring 55, D-70569 Stuttgart, Germany

[§]Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470 Mülheim an der Ruhr, Germany

Supporting Information

ABSTRACT: The proper description of explicit water shells is of enormous importance for all-atom calculations. We propose a new approach for the setup of water shells around proteins based on Tabu-Search global optimization and compare its efficiency with standard molecular dynamics protocols using the chignolin protein as a test case. Both algorithms generate reasonable water shells, but the new approach provides solvated systems with an increased water−enzyme interaction and offers further advantages. It enables a stepwise buildup of the solvent shell, so that the more important inner part can be prepared more carefully. It also allows the generation of solute structures which can be biased either toward the (experimental) starting structure or the underlying theoretical model, i.e., the employed force field.



1. INTRODUCTION

Water, as the most important chemical solvent for biological systems, has a huge impact on the structure of solvated biomolecules and their reactions and properties. Conversely, the structure of the solvent shell itself is mutually influenced by the solute.[1,2] The solvent environment can be divided into an inner shell, which directly interacts with the solute (microsolvation), and an outer shell, which influences the solute by long-range interactions. These latter interactions are often summarized as bulk effects.

While bulk effects are well captured by continuum solvation methods like the polarizable continuum model (PCM)[3,4] or the conductor-like screening model (COSMO),[4,5] this is not true for effects connected with microsolvation,[1,4,6,7] which may arise from, e.g., strong hydrogen bonds or charge-dipole interactions between the solute and specific solvent molecules. The strong influence of microsolvation on the geometrical structure of the solute,[8−11] on its electronic[5,7,19−23] and spectroscopic[12−23] properties, and on reaction kinetics and mechanisms[24−26] is well documented. Interesting examples are β-sheets in peptides. Going from the gas phase to more solvent-like surroundings, they twist more and more.[9] A study of histidine-adenine dimers by Leavens et al. showed that the shapes of clusters which are bonded through $\pi-\pi$ interactions (stacking vs T-shaped) also depend on solvation effects.[11] Another well-known example is the preference between zwitterionic and neutral states.[10,27−29] For instance, Cao et al.[29] studied this preference and the favored conformational structure as a function of the number of surrounding water molecules in the case of 3-fluoro-γ-aminobutyric acid. Parac et al.[12] compared the absorption spectra of guanine in the gas phase and in water (using microsolvated cluster models, continuum treatments, and quantum mechanics/molecular mechanics (QM/MM) approaches) and analyzed the dependency of excited-state properties on the water environment in much detail. Solvation also influences charge transfer in proteins as shown in theoretical studies by Ufimtsev et al.[30] and Anisimov et al.[31] In a series of systematic investigations, Ghomi et al.[13−21] compared the calculated and measured vibrational spectra of amino acids and peptides surrounded by water molecules and demonstrated the huge influence of microsolvation on these spectra.

Water molecules are also important for the reactivity of proteins as certain reactions only occur with the assistance of a water molecule in a specific position. Several studies on the reaction mechanisms in cytochrome P450 underlined that the proper orientation of particular water molecules can be essential for obtaining reliable QM/MM results.[24−26] The importance of particular water molecules is also evident from the recent development and application of empirical water placement algorithms.[32−35] The influence of solvent shells obtained from molecular dynamics (MD) simulations on subsequent QM/MM calculations was discussed in a study on the protonation states of *mt*KasA by Lee et al.[10]

All these investigations point to the strong influence of microsolvation on the properties of biomolecules and indicate that the reliability of QM/MM computations strongly depends on an adequate setup for the geometrical structure of the microsolvation shell. This represents a problem since normally

Received: October 16, 2012
Published: November 29, 2012

   dx.doi.org/10.1021/ct300898d | *J. Chem. Theory Comput.* 2013, 9, 814−821

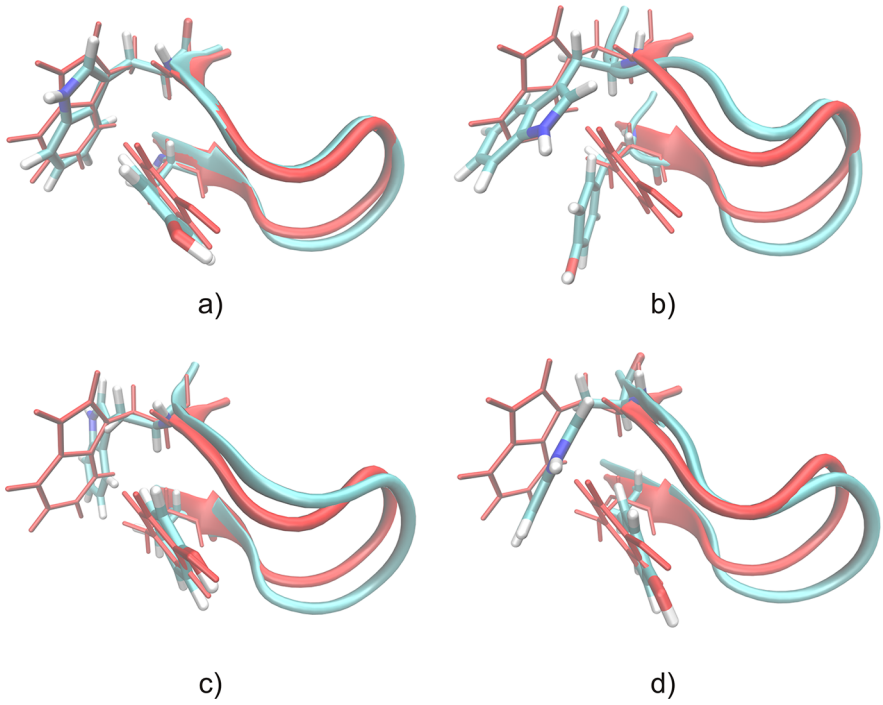


Figure 1. Energetically lowest predicted structures of the 1UAO protein (cyan). The arrangement of the Tyr-2 and Trp-9 residues is indicated. The reference NMR structure is shown in red. (a) TS-fixed, (b) TS-free, (c) MD-fixed, and (d) MD-free.

only a few of these important inner-shell water molecules are precisely located by X-ray or NMR data. Hence, a careful preparation of the solvation shell is necessary.[1,36] For very small systems, an accurate preparation of the microsolvation shell can still be done by hand as shown by Ghomi. However, for larger systems such systematic approaches become too expensive, and proper setup protocols are needed. The standard approach for building up the solvent shell in QM/MM investigations comprises the following steps. The protein is surrounded by pre-equilibrated water spheres. All water molecules within a certain range of the protein are removed to avoid clashes of the nuclei, and the complete system is equilibrated by long MD runs involving heating and cooling periods.[10,12,24−27,29−31,37−39] These MD simulations should qualitatively reproduce the physical motion of an NVT ensemble and thus cover bulk effects adequately. Even so, the positioning of the important inner-shell water molecules may miss the most stable arrangement, since conformational search MD simulations often do not find the global minimum structure.[40]

In this respect, routines developed for global optimization problems should be more efficient. One established method is Tabu-Search, which was first introduced by Glover for global optimization problems[41−44] and was also applied to the multiple minima problem by Klinowski and Cvijovic.[45] It uses a *steepest descent−modest ascent* strategy together with diversification parts. In the latter, the search is guided to new promising areas of the search space when it gets stuck. In a recent study, the combination of the Gradient Only Tabu-Search (GOTS)[46] with the well-known Monte Carlo with Minimization (MCM)[47] or Basin Hopping (BH)[48] approach was proven to be very efficient in conformational search.[40] This motivated us to apply this algorithm to the generation of realistic water solvent shells around biomolecules. A possible problem in this context is that global optimization routines do not correctly reflect the physical motions of the water as a function of the temperature or pressure. Consequently, they might lead to a more crystalline (ice-like) water phase that would not mimic the correct behavior of a solvent shell of fluid water. To investigate this possible drawback, we compare water shells generated by our approach with the results of MD simulations.

The artificial mini-protein chignolin whose structure is known from NMR spectroscopy[49] (PDB code 1UAO) represents an ideal test system to compare the efficiency of different approaches since the computed structures seem to depend strongly on the simulation protocol.[50,51] The orientation of the two residues Tyr-2 and Trp-9 is most problematic. In the experimental NMR study, an edge-to-face (T-shaped) orientation of the aromatic rings of the residues was obtained (Figure 1, in red). By contrast, MD simulations performed by Suenaga et al.[51] and Satoh et al.[50] on the basis of the Amber force field[52] mostly predicted configurations as shown in Figure 1c, while the T-shaped conformation was found rather rarely. Similarly, the T-shaped structure was not predicted as a global minimum in a study by Kusumaatmaja et al. using the Amber force field and an implicit solvent model.[53] It could not be determined whether the theoretical or the NMR structures are correct. To shed light on this important issue, we employed the OPLS-AA force field in combination with our Tabu-Search-based approach to build up an appropriate water shell and to compute geometrical structures of the protein. The latter were validated by density functional theory (DFT) computations.

2. COMPUTATIONAL DETAILS

All computations started from the β-hairpin structure of chignolin as given in the PDB database (PDB code: 1UAO, amino acid sequence: Gly-Tyr-Asp-Pro-Glu-Thr-Gly-Thr-Trp-Gly).[49] MD simulations were performed with a time step of 1 fs in an NVT ensemble at 310 K. The system was heated in 10 K

dx.doi.org/10.1021/ct300898d | J. Chem. Theory Comput. 2013, 9, 814−821

steps from an initial temperature of 10 K to the final temperature. During heating, the velocities were scaled to stay in reasonable ranges. The starting structure for MD (chignolin plus solvent shell) was prepared using the VMD program.[54] The protein molecule was placed in a pre-equilibrated sphere of 526 water molecules with a diameter of about 30 Å. During the simulations, spherical boundary potentials were applied to avoid the evaporation of water molecules. Finally, 2000 snapshots (each 1 ps) were optimized using an L-BFGS algorithm[55−59] implemented in CAST.[60]

To check whether the water shell is sufficiently large to ensure converged properties of the protein, a bigger system (diameter about 38 Å and 1307 water molecules) was also simulated using NAMD[61] and employing the Charmm27[62,63] and the OPLS-AA[64−67] force fields with a 1 fs time step in an NVT ensemble. After heating from 25 to 310 K and running a 2 ns classical equilibration with frozen solute and spherical boundary conditions, a 5 ns production run was performed. All 1000 snapshots were minimized using the conjugate gradient algorithm implemented in NAMD.[61] The results were in good agreement with those from the MD simulations carried out with CAST. Therefore, we will focus on the results obtained with the smaller water sphere.

The calculations were performed using the Tabu-Search-Dimer approach and the solvation algorithm described in the next section. Each global optimization run encompassed 1000 main iterations (*steepest descent*−*modest ascent* steps). The step size of the basin hopping diversification in Cartesian coordinates was adjusted to the system size and varied between 3.0 Å in the first to 0.7 Å in the last solvation step.[68] For the stepwise procedure, we performed five solvation steps with 30, 115, 226, 467, and 526 water molecules. All calculations were repeated several times and gave very similar results.

All calculations with the CAST program[60] were performed employing the OPLS-AA force field[64−67] in combination with the TIP3P potential[69] for the surrounding water molecules. We did not apply any cutoff for nonbonded interactions. For the placement of new solvent molecules, the program VegaZZ[70] was used.

The radial pair distribution function (radial PDF) was calculated with the corresponding tool implemented in VMD.[54] The protein atoms and the water oxygen atoms were taken as selections 1 and 2, respectively.

The obtained results were further analyzed by computations for a model system which comprises the phenol ring of Tyr-2 and the indole system of Trp-9, with one hydrogen atom added at each ring to saturate the dangling bonds (Figure 2). We

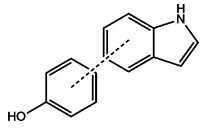


Figure 2. Schematic view of the model system for DFT calculations.

performed single-point calculations and geometry optimizations with DFT. For these calculations, we employed RI-DFT (B3LYP)[71−73] with the D3 dispersion correction[74] in combination with the 6-311G** basis set using the Turbomole 6.3.1 program.[75] COSMO[76] served as continuum solvent model (ε = 78.5).

3. ADAPTATION OF THE DIMER METHOD TO TABU-SEARCH

The Tabu-Search method can be adapted to problems in theoretical chemistry[46,77,78] and further improved by using the basin hopping approach of Wales and Doye[48] and Scheraga and Li[47] for the diversification part. For details about this implementation, we refer to the literature and the Supporting Information (SI).[40] In this work, we present a new approach in which the modest ascent part is considerably enhanced by using the well-known dimer method.[79−81] As pointed out by Henkelman and Jónsson, the Dimer method captures the most important qualities of the eigenvector following method[82−84] while using only first derivatives.[79] We focus on modifications of the original approach that were introduced to utilize the dimer approach for the Tabu-Search. For the general theory of the dimer method, we refer to the literature and SI.[79−81]

Dihedral coordinates involving single or hydrogen bonds represent the softest degrees of freedom. It is therefore quite natural to vary only these coordinates in the modest ascent part of the Tabu-Search. Since the geometrical coordinates of the neighboring local minima will differ slightly, all coordinates are relaxed in the steepest descent part. The modification and definition of dihedral angles follow the concept of main and dependent torsions proposed by Echenique and Alonso,[85] which ensures the proper rotation of whole groups. The setup is straightforward for molecules, but in large molecular clusters like solvated biomolecules, one has to guarantee that dihedral angles include atoms in close proximity only. But even with a careful definition of dihedral angles, atomic centers sometimes come too close during the variation of dihedral angles. These clashes are avoided by implementing a distance check routine, which automatically scales down the calculated step sizes in such cases.

The dimer method is applied in the space of dihedral angles. The modest ascent dimer search starts from a local minimum. The dimer direction is chosen randomly. Dimer rotation and translation steps are performed alternatingly. First, the dimer is rotated in the space of dihedral angles to minimize the torque on the dimer. This aligns the dimer with the softest vibrational mode. Then a translation step is performed by propagating the dimer downhill in energy (against the gradient) in all directions perpendicular to the dimer direction and toward higher energy (along the gradient) in the dimer direction. Rotation steps are performed until the predicted rotational angle[80,81] is smaller than a threshold (e.g., 20 degrees) or a maximum number of rotations (e.g., 20) is reached.

In contrast to mechanistic studies where it is essential to locate transition states very accurately, the Tabu-Search here only needs to leave a minimum very quickly. Experience shows that it is not necessary to follow the exact modest ascent path. Therefore, rather large step sizes can be used, and the search is stopped as soon as the dimer end point is lower in energy than the corresponding dimer start point after a translation step.

For larger systems (more than a few 100 atoms), the new Tabu-Search algorithm is much faster than the original GOTS algorithm, and in most of the cases, it converges to energetically lower-lying final minima. For example, for the protein ubiquitin comprising 1231 atoms, the new approach is about 40 times faster than the old one. Furthermore, it yields structures that are lower in energy.

4. SOLVATION ALGORITHMS

In the standard MD-based setup procedure, additional potentials at the boundary are necessary to avoid the evaporation of water molecules. Consequently, the whole solvent shell has to be added in one go, since additional potentials will distort the structure of the solute when using small solvent shells. In our new approach, an evaporation of water molecules cannot take place as only dihedral angles are modified in the modest ascent part. Hence, the solvent shell can be built up stepwise. This has the advantage that the optimization of the most important first solvation shell, which interacts directly with the solute (microsolvation), can be performed more carefully since a considerably smaller system has to be treated. After optimization of the first shell, the second, third, etc. shells are added. In each step, a global optimization is performed to adapt all water molecules to the growing solvent shell. For these outer shells that do not interact directly with the solute and thus contribute to bulk effects, less precise requirements are necessary in these optimizations. It is of course also possible to add all water molecules at once; however, it will be shown later that the stepwise approach is superior.

The stepwise buildup of the solvent shell offers some flexibility. In a first variant, the solute is fixed to the (experimental) starting structure during the buildup of the solvent shells. The optimization of the solute could already be started after the buildup of the first few solvent shells, but in the present work a full optimization was performed only after the buildup was completed. In the following, this strategy will be called *Tabu-Search fixed (TS-fixed)*. In a second variant, the solute is included in the optimization from the very beginning (*Tabu-Search free, TS-free*).

The outcome of both variants will be different if either the force field or the (experimental) starting structure contains errors. As shown later, in TS-fixed the obtained structure of the enzyme is biased toward the (experimental) starting structure since the cage formed by the solvent shells restricts geometrical changes of the solute. This results from the modest ascent strategy, which preferentially varies the relative orientation of the outer-shell solvent molecules since such variations need less energy than modifications in the more rigid inner part. In TS-fixed, smaller adaptations of the inner solvent shell and the protein are still feasible, but it is normally not possible for a whole residue to change its orientation. Hence, if an erroneous (experimental) starting structure is taken it will be normally retained in TS-fixed. By contrast, in TS-free the solute will adopt geometries that are preferred by the underlying computational method (i.e., the force field) during the build-up of the first few solvation shells since only very small hindrances are present. So, if the starting structure is wrong but the force field is sufficiently accurate, the correct structure is obtained in TS-free. On the other hand, if the (experimental) starting structure is correct but the force field is wrong, TS-fixed will still provide the correct geometry but TS-free will fail. Hence, significant differences in the outcome of both procedures point to deficiencies either in the force field or in the experimental structure or both.

A third Tabu-Search variant is to perform a global optimization immediately and only with the complete water shell. This approach will be called *TS-complete*.

The corresponding MD variants (*MD-fixed* and *MD-free*) were also applied. To obtain completely relaxed geometries, 2000 snapshots from each run were optimized. It turns out that both MD-based procedures lead to similar geometries in the protein; i.e., differences as found in the TS-based variants are absent. In this regard, the optimized and nonoptimized MD simulations behave in an analogous manner. This occurs because MD does not enforce a modest ascent but allows motions along all degrees of freedom (according to the forces acting on each atom). In the following, we will only discuss the results of the MD-free procedure.

5. RESULTS AND DISCUSSION

Table 1 compares the various approaches in terms of the required CPU time and the energy of the energetically lowest-lying geometries. While the MD-free calculation required about 174 CPU-h in total, the TS-fixed and TS-free approaches needed 150 and 93 h, respectively. Both TS-based approaches delivered arrangements that are considerably lower in energy than the MD-free approach (by 281 and 243 kJ/mol, respectively). The TS-complete strategy needed considerably more time than the other two Tabu-Search approaches and delivered similar results. Hence, this approach seems to be less efficient.

Table 1. Solvation of Chignolin with 526 Water Molecules

method	CPU time [h]	ΔE [kJ/mol][a]
MD-free	174	+281
TS-fixed	150	0
TS-free	93	+38
TS-complete	285	−33

[a]Relative energy with respect to the best solution obtained by TS-fixed. For more information, see text.

Table 2. Relative Interaction Energies between the Residues Tyr-2 and Trp-9[a]

		intermonomer orientation	
method	geometry from	T-shaped[b]	parallel[c]
DFT[d]	OPLS-AA	−0.3	1.2
OPLS-AA	OPLS-AA	−4.8	−6.5
DFT[d]	DFT[d]	−5.2	0.4
OPLS-AA	DFT[d]	−0.6	−5.4

[a]The energies are given in kJ/mol relative to the energy obtained from a single-point calculation on the NMR structure. A schematic view of the chosen model system can be found in Figure 2. [b]Predicted by NMR and TS-fixed. [c]Predicted by TS-free, MD-free, and MD-fixed. [d]B3LYP-D3/6-311G**, COSMO (ε = 78.5).

Table 3. Analysis of Energy Contributions[a]

split energy systems	MD-free (lowest structure)	TS-fixed	TS-free
E(TOT)	−28206	−28487	−28453
E(ENZ)	−1361	−1005	−1306
E(WAT)	−23852	−23848	−24183
E(TOT)−E(ENZ)−E(WAT)	−2993	−3634	−2964

[a]For the MD protocol, we analyzed the most stable structure of the whole simulation. For more information, see text. All energies are given in kJ/mol.

The root-mean-square deviations (RMSD) from the starting structure of the complete system (water + enzyme) in the MD simulation converged to a final value of about 17 Å, which

dx.doi.org/10.1021/ct300898d | J. Chem. Theory Comput. 2013, 9, 814−821

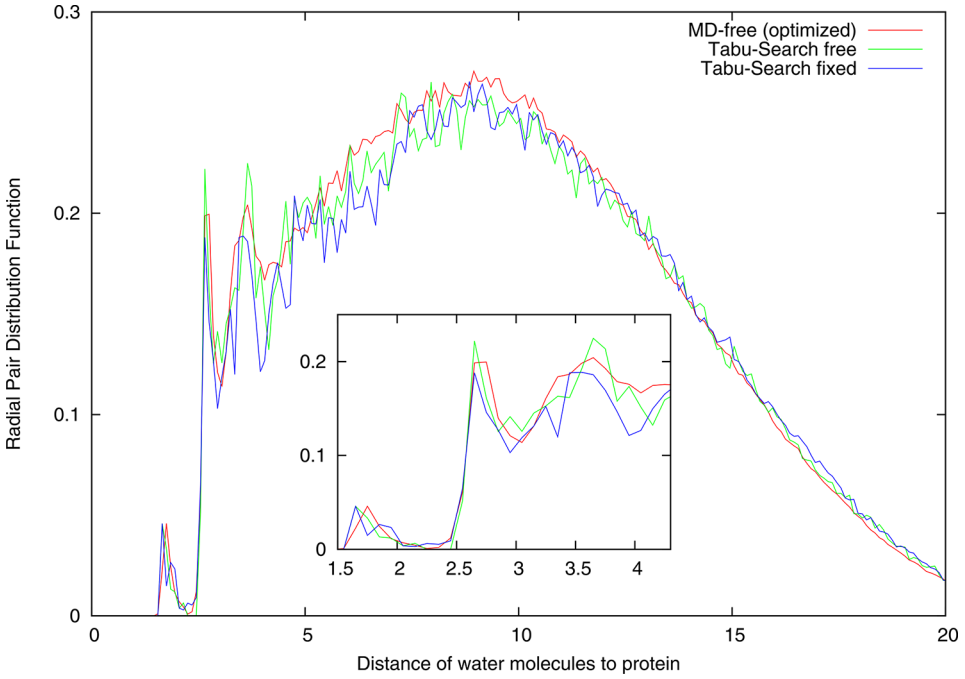


Figure 3. Radial pair distribution function (radial PDF) of water molecules around the chignolin protein. The 20 lowest structures of each simulation were included in the calculation of the radial PDF. The radial PDF was calculated with the corresponding tool implemented in VMD.[54]

corresponds to the radius of the complete system. This confirms that the simulation was run for a sufficiently long time since the water molecules have been redistributed over the whole solvent sphere. The total energy obtained from geometry optimization of selected frames has a standard deviation of 63 kJ/mol and a maximal fluctuation of 502 kJ/mol. The RMSD values of the protein and the backbone are 2.5 Å and 1.0 Å, respectively, showing only modest overall deviation of the secondary structure of the protein with respect to the starting structure. The RMSD of the TS-free simulation closely resembles that of the MD-data. As expected, the TS-fixed optimization arrives at much smaller RMSD values for the protein and the backbone (1.0 Å and 0.5 Å, respectively). This shows that the TS-fixed simulation indeed stays much closer to the starting structure. Figures showing the RMSD of all systems as well as the behavior of the total energy during the optimizations can be found in the Supporting Information.

All simulations agree with the NMR experiment in giving an overall β-hairpin structure of chignolin and also in the orientation of most residues. Only very flexible and solvent-exposed side chains deviate. Especially, and in accordance with the literature,[50,51] considerable differences are found in the relative orientations of the Tyr-2 and Trp-9 residues (Figure 1a–d). The final structure obtained from the TS-fixed simulation closely resembles the reference structure from NMR in which the Tyr-2 and Trp-9 adopt a T-shaped configuration. In all other calculations, the two residues drift apart and form more parallel arrangements. Such orientations were also found in simulations that employed the Amber force field.[50,51] The results of our two MD simulations, MD-free and MD-fixed, are very similar. As discussed above, the differences between the TS-fixed and TS-free results indicate that either the force field underestimates the stability of the T-shaped structure in comparison to the parallel one or the NMR data were erroneously interpreted.

To shed light on this issue, we compared the force field (OPLS-AA) computations with B3LYP-D3 calculations for the model system depicted in Figure 2. We started from the relative orientation of both residues as predicted by NMR, TS-free, and TS-fixed. For all structures, the DFT energies of the force field structure and the DFT-optimized structure were calculated. Likewise, we computed the OPLS-AA energies for the DFT and the force-field optimized systems. The results are shown in Table 2. The DFT approach predicts the T-shaped orientation to be more stable (by 6 kJ/mol). The force-field energies for the same structures predict the parallel orientation as being more stable (by about 5 kJ/mol). The same is true for the force-field optimized structures, although the energy differences are smaller. These results indicate deficiencies of the force field.

The different geometry predictions of MD and TS-free on the one hand and TS-fixed on the other could be caused by the latter simulations generating nonmeaningful water shells. To eliminate this possibility, Table 3 analyzes the various contributions to the total energy in more detail, while Figure 3 compares the radial pair distribution functions (radial PDF) predicted by MD-free, TS-fixed, and TS-free, respectively. The radial PDF of the MD simulation (with nonoptimized frames) shows very similar behavior (see Supporting Information). In Table 3, E(TOT) denotes the total energy of the complete system, while E(ENZ) and E(WAT) are the energies of the isolated enzyme and the water system obtained from single-point computations at the geometries obtained from the simulations. An illustration of the subsystems is given in the Supporting Information. Interesting insights are obtained from E(TOT)−E(ENZ)−E(WAT), i.e., the interaction energy between the enzyme and the water shell, which reflects differences in microsolvation.

As expected for a small protein, the energy of the water shell completely dominates the total energies (84–85%). These differ by only about 1%, and also the radial PDFs given in Figure 3 indicate high similarity in the different treatments. For

dx.doi.org/10.1021/ct300898d | J. Chem. Theory Comput. 2013, 9, 814–821

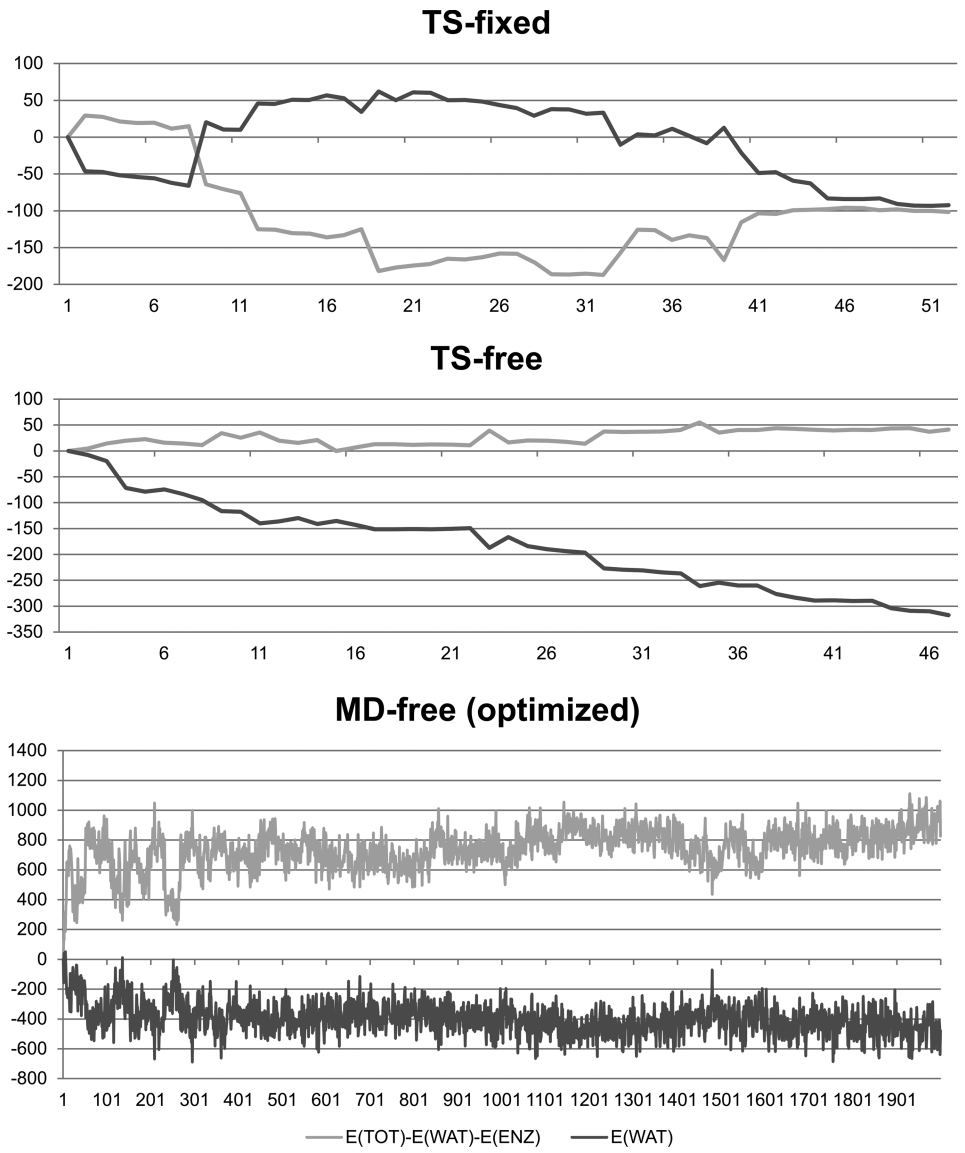


Figure 4. Evolution of the energy contributions of the solvation shell for the different simulation protocols. The abscissa gives the frame number; the ordinate gives the energy values. Gray: E(TOT)−E(WAT)−E(ENZ). Black: E(WAT). For more information, see the text. All energies are given in kJ/mol relative to the first frame.

distances larger than 4 Å, the radial PDF predicted by MD-free is more smeared out, but the overall differences from the TS-based methods are small. All simulations predict a first maximum at around 1.75 Å while the second and the third distinct maxima are found at about 2.65 Å and 3.60 Å, respectively. These comparisons confirm that the TS-based approaches yield meaningful water shells.

Nevertheless, the TS-fixed simulation computes somewhat more structured water shells than the other simulations. Its first maximum appears at R = 1.65 Å, and an additional shoulder is found at about R = 1.85 Å. An additional smaller maximum appears at R = 2.25 Å. More distinct patterns are also found for distances up to 8 Å. The TS-free simulation gives, in accordance with the MD-free simulation, only one maximum which lies at R = 1.65 Å for TS-free and R = 1.75 Å for MD-free. For distances larger than 4 Å, its radial PDF closely resembles that from the TS-fixed simulation. All of these observations indicate that the TS-fixed simulation generates a network, in which the water molecules can occupy more strongly bound positions than in the other two simulations. This may result from vacancies that remain open in the TS-fixed simulations due to the constraints on the protein atoms during the relaxation of the water molecules and that are filled in both the TS-free and MD-free setups.

A similar interpretation also holds for the interaction energies between solute and solvent, E(TOT)−E(ENZ)−E(WAT). For MD-free and TS-free, they contribute 10.6% (with a standard deviation of 0.5%) and 10.4% to the total energy, respectively. For TS-fixed, this contribution increases to 12.8%. This proves that the TS-fixed approach yields water surroundings with stronger microsolvation. The quality of the description of microsolvation interactions can be further examined by investigating the evolution of the water−enzyme and water−water interaction energies during the optimizations (Figure 4). The energies are plotted relative to the first frame of the simulations. The total energies are in the range of the values given in Table 3. In the MD-free and TS-free simulations, the energy of the water shell decreases as the simulation proceeds,

dx.doi.org/10.1021/ct300898d | J. Chem. Theory Comput. 2013, 9, 814−821

while the interaction of the enzyme with the water shell increases in the beginning and then equilibrates at a certain value. In the case of TS-free, the enzyme−water interaction stays more or less constant and does not affect the optimization. For MD-free, the enzyme−water interaction even diminishes during the simulation. The fluctuation (standard deviation) of this energy contribution is about 569 kJ/mol.

In the TS-fixed calculations, the trend is completely the opposite. During the very first frames, the water−water interaction increases by about 50 kJ/mol while the water−enzyme interaction is more or less unaffected. However, in the subsequent optimization, the water−water interaction is lowered again in favor of an increased water−enzyme interaction. In the end, the two interactions converge to similar stabilization energies of about 100 kJ/mol resulting in a well-balanced and optimized microsolvation.

In all calculations, the interactions within the enzyme itself (see Table 3, E(ENZ)) represent only about 5% of the total energy. The slightly smaller contribution found for the TS-fixed simulation results from the geometry of the protein being biased toward the NMR structure but not to the underlying force field.

6. SUMMARY AND CONCLUSIONS

We propose a new solvation procedure based on Tabu-Search optimization. The nature of the optimization algorithm allows for a stepwise build-up of the solvent shell. This enables a more careful preparation of the more important inner part of the solvation shell (microsolvation). Furthermore, by adjusting the optimization protocol, either a bias to the (experimental) starting structure or to the chosen force field can be introduced. The chignolin protein was taken as a test case. A comparison to standard molecular dynamics simulations shows that the enzyme−water interactions in the microsolvation regime are stronger and thus better described in the Tabu-Search-based approach. A detailed investigation of the evolution of the enzyme−water and the water−water interaction energies confirms these findings. Comparison between force-field and DFT interaction energies for a model system consisting of the crucial tyrosine and tryptophan residues reveals force field deficiencies in the proper description of the T-shaped configuration. In summary, the present investigation shows that the water shells generated by the Tabu-Search-based simulations provide an improved description of microsolvation. The algorithm can be applied straightforwardly, and only the most important inner solvent shell has to be treated very accurately. Thus, it should serve as a useful tool for the preparation of starting structures for subsequent QM/MM computations.

■ ASSOCIATED CONTENT

Supporting Information

Illustration of subsystems used for Table 3, figures of RMSD values and of radial PDFs, and further theoretical details. This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: bernd.engels@mail.uni-wuerzburg.de. Tel.: (+49)931-31-85394. Fax: (+49)931-31-85331.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Financial support by the Deutsche Forschungsgemeinschaft within the framework of the GRK 1221 "Control of Electronic Properties of Assemblies of π-Conjugated Molecules" and the research unit 1809 "Lichtinduzierte Dynamik in molekularen Aggregaten" is gratefully acknowledged. We also acknowledge the financial support of the Volkswagen Stiftung.

■ REFERENCES

(1) Schmidt, T. C.; Paasche, A.; Grebner, C.; Ansorg, K.; Becker, J.; Lee, W.; Engels, B. In *Topics in Current Chemistry*; Wong, C.-H., Houk, K. N., Hunter, C. A., Krische, M. J., Lehn, J.-M., Ley, S. V., Olivucci, M., Thiem, J., Venturi, M., Vogel, P., Wong, C.-H., Yamamoto, H., Eds.; Springer: Berlin/Heidelberg, 2012; pp 1−77.
(2) Marcus, Y. *J. Chem. Thermodyn.* **2007**, *39*, 1338−1345.
(3) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669−681.
(4) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999−3093.
(5) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.
(6) Lee, M. S.; Salsbury, F. R.; Olson, M. A. *J. Comput. Chem.* **2004**, *25*, 1967−1978.
(7) Joshi, K.; Semrouni, D.; Ohanessian, G.; Clavaguéra, C. *J. Phys. Chem. B* **2012**, *116*, 483−490.
(8) Hooft, R. W. W.; Sander, C.; Vriend, G. *Proteins* **1996**, *26*, 363−376.
(9) Ireta, J. *J. Chem. Theory Comput.* **2011**, *7*, 2630−2637.
(10) Lee, W.; Luckner, S. R.; Kisker, C.; Tonge, P. J.; Engels, B. *Biochemistry* **2011**, *50*, 5743−5756.
(11) Leavens, F. M. V.; Churchill, C. D. M.; Wang, S.; Wetmore, S. D. *J. Phys. Chem. B* **2011**, *115*, 10990−11003.
(12) Parac, M.; Doerr, M.; Marian, C. M.; Thiel, W. *J. Comput. Chem.* **2009**, *31*, 90−106.
(13) Derbel, N.; Hernández, B.; Pflüger, F.; Liquier, J.; Geinguenaud, F.; Jaïdane, N.; Lakhdar, Z. B.; Ghomi, M. *J. Phys. Chem. B* **2007**, *111*, 1470−1477.
(14) Guiffo-Soh, G.; Hernández, B.; Coïc, Y.-M.; Boukhalfa-Heniche, F.-Z.; Fadda, G.; Ghomi, M. *J. Phys. Chem. B* **2008**, *112*, 1282−1289.
(15) Guiffo-Soh, G.; Hernández, B.; Coïc, Y.-M.; Boukhalfa-Heniche, F.-Z.; Ghomi, M. *J. Phys. Chem. B* **2007**, *111*, 12563−12572.
(16) Hernández, B.; Carelli, C.; Coïc, Y.-M.; De Coninck, J.; Ghomi, M. *J. Phys. Chem. B* **2009**, *113*, 12796−12803.
(17) Hernández, B.; Pflüger, F.; Adenier, A.; Kruglik, S. G.; Ghomi, M. *J. Phys. Chem. B* **2010**, *114*, 15319−15330.
(18) Hernández, B.; Pflüger, F.; Adenier, A.; Nsangou, M.; Kruglik, S. G.; Ghomi, M. *J. Chem. Phys.* **2011**, *135*, 055101−055101.
(19) Hernández, B.; Pflüger, F.; Nsangou, M.; Ghomi, M. *J. Phys. Chem. B* **2009**, *113*, 3169−3178.
(20) Hernández, B.; Pflüger, F.; Derbel, N.; De Coninck, J.; Ghomi, M. *J. Phys. Chem. B* **2010**, *114*, 1077−1088.
(21) Pflüger, F.; Hernández, B.; Ghomi, M. *J. Phys. Chem. B* **2010**, *114*, 9072−9083.
(22) Zischang, J.; Lee, J. J.; Suhm, M. A. *J. Chem. Phys.* **2011**, *135*, 061102−061102.
(23) Leon, I.; Cocinero, E. J.; Millán, J.; Jaeqx, S.; Rijs, A. M.; Lesarri, A.; Castaño, F.; Fernández, J. A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 4398−4409.
(24) Altun, A.; Shaik, S.; Thiel, W. *J. Comput. Chem.* **2006**, *27*, 1324−1337.
(25) Kumar, D.; Altun, A.; Shaik, S.; Thiel, W. *Faraday Discuss.* **2011**, *148*, 373−383.
(26) Zheng, J.; Altun, A.; Thiel, W. *J. Comput. Chem.* **2007**, *28*, 2147−2158.

(27) Mladenovic, M.; Fink, R. F.; Thiel, W.; Schirmeister, T.; Engels, B. *J. Am. Chem. Soc.* **2008**, *130*, 8696−8705.
(28) Chan-Huot, M.; Sharif, S.; Tolstoy, P. M.; Toney, M. D.; Limbach, H.-H. *Biochemistry* **2010**, *49*, 10818−10830.
(29) Cao, J.; Bjornsson, R.; Bühl, M.; Thiel, W.; van Mourik, T. *Chem.—Eur. J.* **2012**, *18*, 184−195.
(30) Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. *J. Phys. Chem. Lett.* **2011**, *2*, 1789−1793.
(31) Anisimov, V. M.; Bugaenko, V. L.; Cavasotto, C. N. *ChemPhysChem* **2009**, *10*, 3194−3196.
(32) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2009**, *113*, 13337−13346.
(33) Li, Y.; Sutch, B. T.; Bui, H.-H.; Gallaher, T. K.; Haworth, I. S. *J. Chem. Inf. Model.* **2011**, *51*, 1347−1352.
(34) Rossato, G.; Ernst, B.; Vedani, A.; Smiesko, M. *J. Chem. Inf. Model.* **2011**, *51*, 1867−1881.
(35) Wallnoefer, H. G.; Liedl, K. R.; Fox, T. *J. Chem. Inf. Model.* **2011**, *51*, 2860−2867.
(36) Sindhikara, D. J.; Yoshida, N.; Hirata, F. *J. Comput. Chem.* **2012**, *33*, 1536−1543.
(37) Geerke, D. P.; Thiel, S.; Thiel, W.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2008**, *10*, 297−302.
(38) Kästner, J.; Sherwood, P. *Mol. Phys.* **2011**, *108*, 293−306.
(39) Rommel, J. B.; Kästner, J. *J. Am. Chem. Soc.* **2011**, *133*, 10195−10203.
(40) Grebner, C.; Becker, J.; Engels, B.; Stepanenko, S. *J. Comput. Chem.* **2011**, *32*, 2245−2253.
(41) Glover, F. *Comput. Oper. Res.* **1986**, *13*, 533−549.
(42) Glover, F. *INFORMS J. Comput.* **1989**, *1*, 190−206.
(43) Glover, F. *INFORMS J. Comput.* **1990**, *2*, 4−32.
(44) Glover, F.; Laguna, M. *Tabu Search*; Kluwer Academic Publisher: Norwell, MA, 1997.
(45) Cvijovic, D.; Klinowski, J. *Science* **1995**, *267*, 664−666.
(46) Stepanenko, S.; Engels, B. *J. Comput. Chem.* **2008**, *29*, 768−780.
(47) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 6611−6615.
(48) Wales, D. J.; Doye, J. P. K. *J. Phys. Chem. A* **1997**, *101*, 5111−5116.
(49) Honda, S.; Yamasaki, K.; Sawada, Y.; Morii, H. *Structure* **2004**, *12*, 1507−1518.
(50) Satoh, D.; Shimizu, K.; Nakamura, S.; Terada, T. *FEBS Lett.* **2006**, *580*, 3422−3426.
(51) Suenaga, A.; Narumi, T.; Futatsugi, N.; Yanai, R.; Ohno, Y.; Okimoto, N.; Taiji, M. *Chem.—Asian J.* **2007**, *2*, 591−598.
(52) Cornell, W. D.; Cieplak, P.; Merz, K. M.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *177*, 5179−5197.
(53) Kusumaatmaja, H.; Whittleston, C. S.; Wales, D. J. *J. Chem. Theory Comput.* **2012**, DOI: dx.doi.org/10.1021/ct3004589.
(54) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(55) Broyden, C. G. *J. Inst. Math. Its Appl.* **1970**, *6*, 76−90.
(56) Fletcher, R. *Comp. J.* **1970**, *13*, 317−322.
(57) Goldfarb, D. *Math. Comput.* **1970**, *24*, 23−26.
(58) Shanno, D. F. *Math. Comput.* **1970**, *24*, 647−656.
(59) Nocedal, J. *Math. Comput.* **1980**, *35*, 773−782.
(60) Grebner, C.; Becker, J.; Weber, D.; Engels, B. *CAST: a new program for conformational search and global optimization*; Universität Würzburg: Würzburg, Bavaria, Germany, 2012
(61) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(62) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(63) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(64) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(65) Jorgensen, W. L.; Tirado-Rives, J. *Abstr. Pap.—Am. Chem. Soc.* **1998**, *216*, 43.
(66) Tirado-Rives, J.; Jorgensen, W. L. *Abstr. Pap.—Am. Chem. Soc.* **1992**, *204*, 43.
(67) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(68) With a growing system size, the probability of atom clashes increases with a larger Cartesian step size. Therefore, the Cartesian step size within the Basin Hopping diversification is reduced.
(69) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335−340.
(70) Pedritti, A.; Vistoli, G. *VegaZZ*. http://www.vegazz.net/ (accessed February 20, 2012).
(71) Eichkorn, K.; Htiser, M.; Ahlrichs, R.; Eichkorn, K.; Treutler, O.; Marco, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *242*, 652−660.
(72) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119−124.
(73) Sierka, M.; Hogekamp, A.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *118*, 9136−9136.
(74) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104−154104.
(75) *TURBOMOLE*, V6.3.1; TURBOMOLE GmbH: Karlsruhe, Germany, 2011. http://www.turbomole.com (accessed Dec. 2012).
(76) Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187−2193.
(77) Stepanenko, S.; Engels, B. *J. Comput. Chem.* **2007**, *28*, 601−611.
(78) Stepanenko, S.; Engels, B. *J. Phys. Chem. A* **2009**, *113*, 11699−11705.
(79) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **1999**, *111*, 7010−7022.
(80) Heyden, A.; Bell, A. T.; Keil, F. J. *J. Chem. Phys.* **2005**, *123*, 224101−224101.
(81) Kästner, J.; Sherwood, P. *J. Chem. Phys.* **2008**, *128*, 014106−014106.
(82) Cerjan, C. J.; Miller, W. H. *J. Chem. Phys.* **1981**, *75*, 2800−2806.
(83) Munro, L. J.; Wales, D. J. *Phys. Rev. B* **1999**, *59*, 3969−3980.
(84) Kumeda, Y.; Wales, D. J.; Munro, L. J. *Chem. Phys. Lett.* **2001**, *341*, 185−194.
(85) Echenique, P.; Alonso, J. L. *J. Comput. Chem.* **2006**, *27*, 1076−1087.