JCTC Journal of Chemical Theory and Computation

Free Energy Barriers for the N-Terminal Asparagine to Succinimide Conversion: Quantum Molecular Dynamics Simulations for the Fully Solvated Model

Ilya Kaliman,[*,†] Alexander Nemukhin,[†,‡] and Sergei Varfolomeev[†,‡]

Department of Chemistry, M.V. Lomonosov Moscow State University, Moscow 119991, Russian Federation, and N.M. Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow 119334, Russian Federation

Abstract: Deamidation of asparagine residues represents one of the main routes for the post-translational modification of protein sequences. We computed the estimates of the free energy barriers for three stages of the deamidation process, deprotonation, cyclization, and deamination, of the conversion of asparagine to the succinimide intermediate within the fully solvated model with explicit water molecules. The Born−Oppenheimer molecular dynamics in the Gaussian and Plane Wave (GPW) approximation as implemented in the CP2K quantum chemistry package was utilized to sample the configurational space. By applying the metadynamics technique, the estimates of the free energy barriers were obtained for three separated stages of the reaction. In agreement with the experimental kinetic measurements, the estimated activation barriers do not exceed 21 kcal/mol. We demonstrate that the use of fully solvated models is the critical issue in theoretical studies of these reactions. We also conclude that more extensive sampling is necessary to obtain full free energy profiles and accurate barriers for the reaction stages.

Introduction

Spontaneous deamidation of asparagine residues leading to aspartic acid residues presents one of the pathways of protein degradation that affects structure and function of proteins.[1] The immense importance of this process is due to its relevance to the development of Alzheimer's disease.[2] The instability of glutaminyl and asparaginyl residues in proteins has been suggested to play a central biological role, and it is proposed that these residues serve as some sort of biological "clocks" for the regulation of processes in cells.[3] It was also proposed that accumulation of protein defects resulting from these processes may be one of the root causes of human aging.[4] The isomerization and racemization processes are known to play an important role not only in vivo but also during purification, storage and, transformation of proteins and polypeptides.[5]
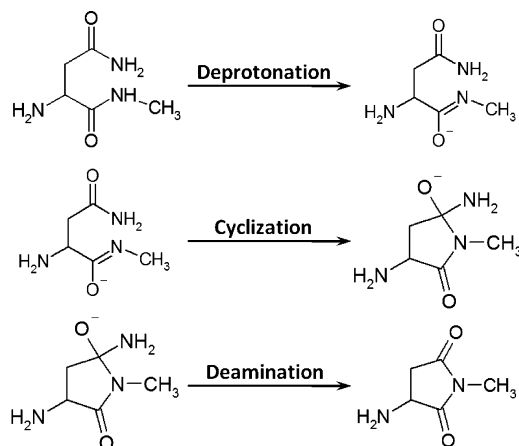
The process of isomerization proceeds through formation of a cyclic imide intermediate.[1,6] This succinimide can be later hydrolyzed leading to aspartate and isoaspartate in the proportion 1:3.[7] A competent summary of what is known about the asparagine deamidation mechanism from the experimental side is presented in the paper of Peters and Trout,[8] and there is no need to retell it in full. For the goals of the present work, it is essential to underline the following: (i) the deamidation mechanism in a protein is the same as in a model peptide;[9] (ii) the conversion of asparagine to succinimide is the rate limiting step;[10] (iii) the rates of deamidation of model peptides obey the Arrhenius equations with the activation energies near 22 kcal/mol at pH 5−7.5.[11]

The tentative mechanism of succinimide ring formation proposed by Capasso et al.[12] can be illustrated in Scheme 1. In the first stage, the deprotonation of a peptide group occurs which may be stimulated either by basic pH or by presence of a nearby molecular group that plays the role of a base. Therefore, such a reaction route can be substantiated in proteins. Next, the cyclization to the tetrahedral intermediate takes place as a result of nucleophilic attack of nitrogen from

　* Corresponding author e-mail: ilya.kaliman@gmail.com.
　† M.V. Lomonosov Moscow State University.
　‡ Russian Academy of Sciences.

Free Energy Barriers for Asparagine to Succinimide

J. Chem. Theory Comput., Vol. 6, No. 1, 2010 **185**

Scheme 1. Stages of the Succinimide Ring Formation from Asparagine

the peptide group on the carbonyl C atom. Finally, leaving of the NH_2 group occurs (the deamination stage) resulting in a cyclic imide product. This stage may be facilitated by general base catalysis.

Beyond experimental investigations, the asparagine isomerization was also studied by the methods of quantum chemistry. The reaction mechanisms for the molecular rearrangements shown in Scheme 1 were modeled by using density functional theory (DFT) calculations corrected by the polarized continuum model (PCM).[13−15] The results qualitatively agree with the mechanism proposed by Capasso et al.[12] Related theoretical studies[16,17] have been performed to rationalize the observation that aspartyl and asparagine residues racemize rapidly compared to other amino acid residues in proteins and peptides. This effect has been attributed to the increased acidity of the α-carbon atoms of succinimide residues. In a series of papers by Catak et al.,[18−20] it is shown that inclusion of an explicit aqueous environment in modeling these reactions is a very important issue. With the help of DFT-based calculations, the authors studied the water-assisted mechanisms of succinimide formation with different amounts of explicit water molecules. They concluded that the most favorable reaction mechanism corresponded to the formation of a succinimide intermediate and involved tautomerization of the asparagine amide to the corresponding imidic acid in the initial reaction step. Peters and Trout[8] simulated a network of elementary reactions for asparagine deamidation also by using the DFT-based quantum chemistry methods corrected for the effects of aqueous environment in the continuum solvent model.

Significantly, all the quantum chemical calculations[8,13−15,18−20] resulted in activation barriers that grossly overestimated, by at least 10 kcal/mol, the experimental value of 22 kcal/mol.[11] Substantial lowering of energy barriers from around 50 kcal/mol was achieved when explicit water molecules were added to the model system. It was suggested[18] that molecular dynamics simulations within the fully solvated model that also accounts for water assistance should be necessary to accurately simulate these processes. In this work, we describe the application of a quantum molecular dynamics procedure within the fully solvated model for estimates of free energy reaction

profiles for the stages illustrated in Scheme 1. This approach allows us to obtain the results for the water-assisted mechanism quantitatively consistent with the experimental data. As a model system, we consider the peptide shown in Scheme 1, which was also used in earlier works.[13−15]

Methods and Computational Details

All simulations were performed by using the CP2K program (http://cp2k.berlios.de). The Quickstep module[21] of the CP2K program was utilized for electronic structure calculations. This approach provides the $O(n)$ implementation of the density functional theory calculations allowing simulations for hundreds of atoms. It relies on the pseudopotential and plane wave methodology in addition to the conventional Gaussian basis set approximations (the Gaussian and Plane Wave method, GPW)[22] to make practical calculations for extended systems.

The method of metadynamics[23,24] was used to obtain free energy profiles for the reaction stages. At present, the available computational resources do not allow observation of such rare events as chemical reactions in model systems by using equilibrium methods of molecular dynamics due to the prohibitively large volume of the configurational space to be explored. The method of metadynamics is one of nonequilibrium-based approaches that allows one to perform simulation of chemical reactions in feasible time.[25] This method was recently used for studies of bacterial chloride channels,[26] deprotonation of formic acid,[27] flexible ligand docking,[28] and a study of decarboxylation mechanism in orotidine-5′-monophosphate decarboxylase.[29]

The method of metadynamics assumes that a small set $s_i(n)$ of relevant collective variables (CVs) can be selected allowing one to construct projections of the free energy surface $F(s)$ on these collective variables. The CVs may be specific functions of atomic coordinates including distances, angles, coordination numbers, dihedral angles, and so forth. The history-dependent Gaussian hills are added to the biasing potential during the simulation (eq 1). This potential builds up until it counterbalances the underlying free energy well, allowing the system to escape via a saddle point to a nearby local minimum, where the procedure is repeated. When all minima are "filled" with Gaussian potential hills, the system moves barrier-free among the different states. The free energy is obtained as a negative sum of added Gaussian hills (eq 2).[30]

$$V_{\text{bias}}(\mathbf{s}, t) = \sum_{t_i} H \exp\left(-\frac{|\mathbf{s} - \mathbf{s}(t_i)|^2}{2\omega^2}\right) \quad (1)$$

$$F(\mathbf{s}) = -V_{\text{bias}}(\mathbf{s}, t) \quad (2)$$

In the limit of the infinitely long trajectory, the biasing potential exactly cancels the underlying free energy surface along the collective variable.

Our model system was composed of 333 atoms including the peptide and 104 water molecules. The periodic boundary conditions were imposed to simulate the bulk. The dimensions of the model cell were $15 \times 15 \times 15$ Å^3 (Figure 1).
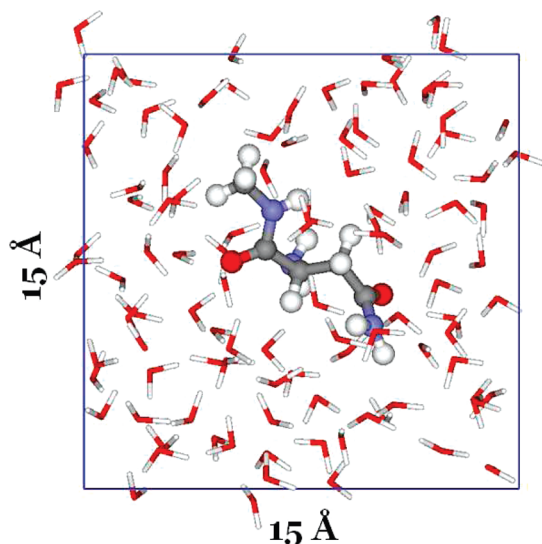
Figure 1. Schematic presentation of the model system.

We used the largest possible cell that allowed us to complete simulations in feasible time. However, the size of this cell does not allow us to pretend for description of such macroscopic properties as pH or concentration.

A dual basis set was used in which the wave functions were described by the Gaussian functions and the electronic density was described by auxiliary plane waves. The TZV2P Gaussian basis set[31,32] was used in all GPW calculations. The plane wave basis was extended to the density cutoff of 280 Ry. We utilized the pseudopotential of Goedeker, Teter, and Hutter (GTH)[33] to describe the core electrons. The BLYP gradient corrected functional was used to compute exchange-correlation contributions.[34,35] To sample the configurational space, we used the Born–Oppenheimer molecular dynamics in which the convergence criterion for the SCF procedure was set to 10^{-7} au at each step. An orbital transformation scheme[36] that is known to substantially accelerate convergence of the SCF procedure was used. The time step of 0.5 fs was utilized in all molecular dynamic calculations. The CSVR (canonical sampling through velocity rescaling)[37] scheme was applied for thermolization of the system. All simulations were performed at 300 K. The metadynamics parameters chosen were 0.001 au for the hill height and 0.5 au for the hill width. The mass of virtual particle in metadynamics simulations was selected to be 100 au.

The initial equilibration of the entire system was performed by using the empirical CHARMM force field.[38] After solvation and energy minimization, the system was equilibrated for 500 ps. Next, the equilibration with fully quantum description was performed for additional 20 ps. For modeling of the second and the third stages of the reaction (Scheme 1), the initial conformations and velocities were extracted from the metadynamics simulation of previous stage with preliminary equilibration for another 10 ps. The trajectory lengths for metadynamics simulations were 25–30 ps.

Results and Discussion

In our molecular dynamics simulations, we use the Gaussian and Plane Wave method to describe the electronic structure

of the model system. Here, we provide an additional check of this approach against the conventional all-electron density functional theory with the B3LYP functional. The latter approach, known as an adequate tool for molecular modeling, was used in previous studies of this reaction.[8,13–15,18–20] The 6-311++G(d,p) basis set was used in all DFT/B3LYP calculations with the PC GAMESS quantum chemistry package (http://classic.chem.msu.su/gran/gamess/index.html).[39] We compared the computed equilibrium geometry configurations and relative energies for the stationary points of the model peptide referred to the reagent, transition state, intermediate, and product structures shown in insets of Table 1. These calculations were performed without water environment.

The comparison of optimized geometries shows almost perfect agreement between the Gaussian and Plane Wave based method on one hand and the all-electron B3LYP approach on the other hand: the deviation between the structures obtained with these two methods does not exceed 0.1 Å for internuclear distances and 2° for angles. Comparison of the relative energies shows that the GPW method describes the potential energy surface of the model system accurately enough (Table 1) but requires much less computational resources.
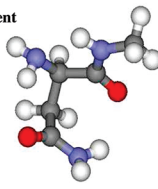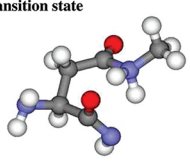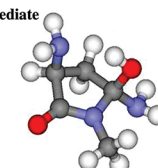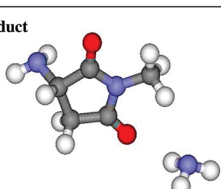
Next, we describe and discuss the resulting free energy profiles for the three stages of the mechanism of succinimide formation shown in Scheme 1: deprotonation, cyclization, and deamination, inside the shell of water molecules (Figure 1).

Deprotonation Stage. For modeling this stage, we select the distance between N and H atoms, describing cleavage of the N–H bond from the initial value of 1.08 Å, as the collective variable. The computed free energy profile along this collective variable for the deprotonation stage is presented in Figure 2 showing the activation barrier not exceeding 21 kcal/mol.

Initially, the system is located in the potential energy minimum near $r_{N-H} \approx 1$ Å. As soon as this minimum is filled with Gaussians, the N–H bond breaks, and the proton moves to the bulk phase. We notice that, upon cleavage of the N–H bond, the zwitterionic species is formed. We also notice that the leaving proton is transferred to the nearby water molecule but not directly to the amino group. The result of the subsequent hydrogen-bond reorganization is a transfer of another proton to the amino group leading to the zwitterionic conformation. We see that the proton that protonates the $-NH_2$ group is not physically the proton from the collective variable so we cannot reliably sample the corresponding conformations. An important issue is that the zwitterionic species is stable only in the presence of solvent water molecules. Attempts to optimize the corresponding structure without a water environment resulted in the back transfer of the proton to the nitrogen atom.
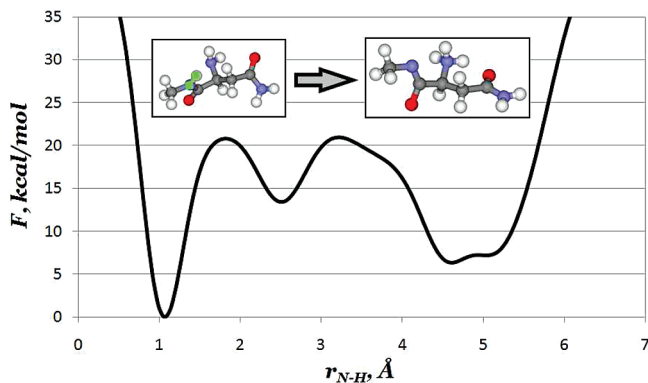
The outer minimum refers to the system in which the proton from the collective variable is in the bulk. Because of the difficulties to sample the corresponding conformations, it is hardly possible to accurately estimate the free energy of this product state. The shape and depth of the outer minimum should depend on concentration and the

Free Energy Barriers for Asparagine to Succinimide

J. Chem. Theory Comput., Vol. 6, No. 1, 2010 **187**

Table 1. Computed Relative Energies of Four Structures of the Model Peptide (kcal/mol)

| Structure	GTH/BLYP/TZV2P/280Ry	B3LYP/6-311++G(d,p)
Reagent	0.0	0.0
Transition state	55.1	55.4
Intermediate	19.9	16.7
Product	6.4	7.7

size of the cell. However, this does not affect estimates of the activation barrier heights. We think that the presence of the minimum near $r_{N-H} \approx 2.5$ Å might also be attributed to the difficulties of sampling the region with the proton in the bulk.

In Figure 3, the CV dynamics graph is shown. From this picture, it is clear that the backward reaction is not observed because the process is not reversible. From this fact, we conclude that the negative metadynamics biasing potential gives us only an estimate of the reaction barrier. On the basis

of the observed bumps in free energy profile, we conclude that the error should be within 5 kcal/mol. To obtain a full free energy profile and a more accurate value for the barrier, one needs to conduct a much longer simulation. However, it is unclear whether it is possible to see the backward reaction in a feasible time during the simulation.

Cyclization Stage. Next, we modeled the cyclization stage corresponding to the formation of the succinimide ring. It results from the attack of the peptide bond nitrogen on the carbon from the amide group. For this stage, we

Figure 2. Free energy profile for the deprotonation stage. Collective variable is marked with the green line and squares.
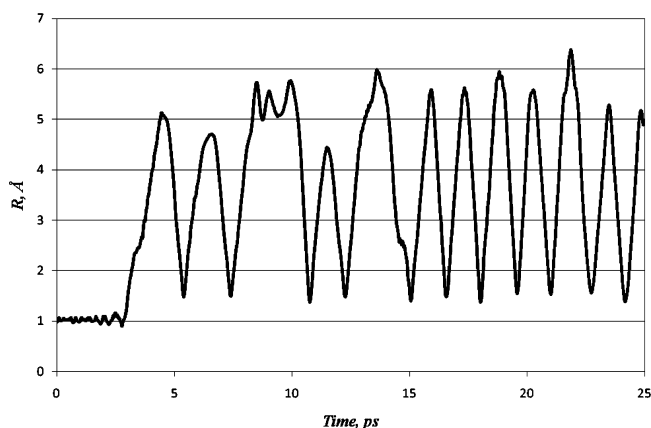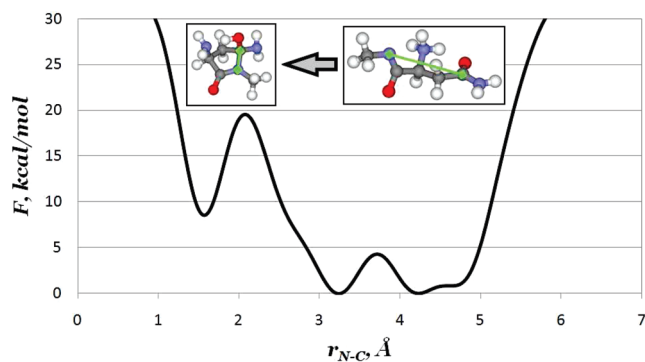

Figure 3. Dynamics of the N–H distance.

Figure 4. Free energy profile for the cyclization stage. Collective variable is marked with the green line and squares.
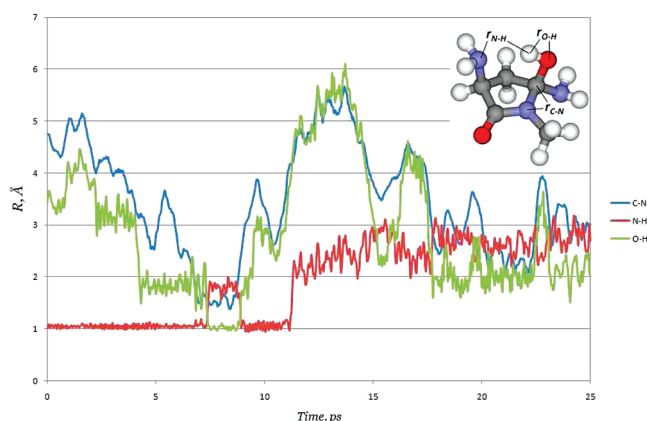

Figure 5. Dynamics of the C−N, N−H, and O−H distances.

Figure 6. Free energy profile for the deamination stage. Collective variable is marked with the green line and squares.

Figure 7. Dynamics of the C−N distance.

used the distance between N and C as the collective variable. The resulting free energy profile is shown in Figure 4. The calculated free energy barrier is about 20 kcal/mol. Initially, the system resided in a rather wide energy well (between 3.2 and 4.8 Å) that was accounted for by an unconstrained motion of the peptide chain. We think that the small maximum near $r_{N-C} \approx 3.7$ Å can be attributed to the insufficient sampling. When this well is "filled" with Gaussian functions, the system moved to another local minimum at $r_{N-C} \approx 1.5$ Å, and the chemical bond was formed between N and C atoms yielding the succinimide ring.

We notice that the excess proton from the amino group was transferred to the newly formed negatively charged oxygen leading to hydroxyl formation accompanied by the hydrogen bonding with the nearby amino group. In Figure 5, we show the dynamics of the C−N, N−H, and O−H distances. From the graph, it is clear that, as soon as the ring is formed (about 7 ps), the proton is transferred to the oxygen (it is worth noting similarities in the behavior of the C−N and O−H distances). But after the system recrosses the barrier and the ring is broken, we observe the back transfer of the proton.

Deamination Stage. Finally, the free energy profile for the deamination process was calculated. The collective variable chosen for this stage was the distance between the nitrogen atom from the leaving amino group and the adjacent carbon atom (the C−N bond distance). The resulting free
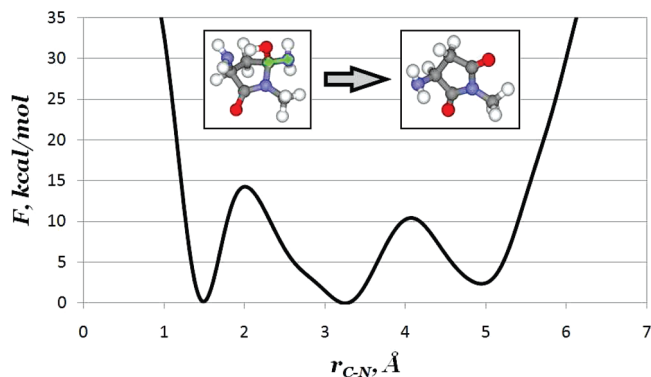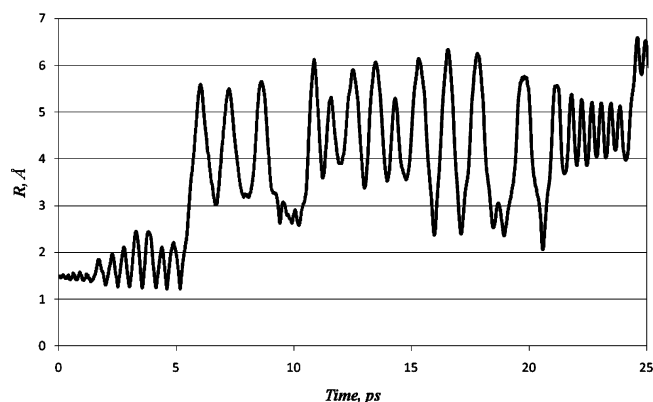
energy profile is shown in Figure 6. The computed free energy barrier for this stage is about 15 kcal/mol. The profile shows that this stage is characterized by the lowest free energy barrier that suggests that the deamination stage should proceed relatively easily compared to the first two stages. From the metadynamics trajectory, we see that the state of the living group is NH_3. The leaving -NH_2 group of the peptide is protonated by the excess proton formed in the deprotonation stage.

The peak in the profile near 4 Å on the CV axis may be attributed to insufficient sampling in the regions where the formed ammonia is far from the newly formed succinimide. As in the case with the deprotonation stage, it is difficult to sample the region of the outer minimum, and its shape should depend on concentration and the size of the cell.

In Figure 7, the dynamics of the C−N collective variable is shown. As in the case of deprotonation reaction, we are unable to see the recrossing of the barrier because the process is not reversible so the negative metadynamics biasing potential gives us only an estimate of the activation barrier for the given process.

Conclusion

Following the results of quantum simulations for separate stages of the mechanism of the cyclic imide formation from the asparagine residue in aqueous solution with an explicit treatment of water molecules, we conclude that the rate limiting stage refers either to deprotonation or to cyclization.

Free Energy Barriers for Asparagine to Succinimide

J. Chem. Theory Comput., Vol. 6, No. 1, 2010 **189**

The corresponding free energy barriers of these steps are estimated as 20−21 kcal/mol. The estimated free energy barrier for the deamination stage is 15 kcal/mol. Based on the observed bumps in free energy profiles, the error of obtained free energy barriers should be about 5 kcal/mol. We conclude that longer trajectories and better sampling are necessary to obtain full free energy profiles and more accurate barriers for the reaction stages. However, it is unclear whether it is possible to observe a backward reaction for stages 1 and 3 in feasible simulation time. Nevertheless, our estimates for the free energy barriers are consistent with the available experimental kinetic measurements showing the Arrhenius activation barrier of about 22 kcal/mol.[11]

If solvent molecules are ignored, the unrealistic energy barriers (above 50 kcal/mol) are predicted in quantum calculations. We show that one of the reaction intermediates can be stabilized only in the presence of water molecules. Therefore, the results of simulations demonstrate that the use of fully solvated models and implication of extensive sampling are the critical issues in theoretical studies of these reactions.

References

(1) Reissner, K. J.; Aswad, D. W. *Cell. Mol. Life Sci.* **2003**, *60*, 1281–1295.

(2) Payan, I. L.; Chou, S.; Fisher, G. H.; Man, E. H.; Emory, C.; Frey, W. H. *Neurochem. Res.* **1992**, *17*, 187–191.

(3) Robinson, N. E.; Robinson, A. B. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 944–949.

(4) Clarke, S. *Aging Res. Rev.* **2003**, *2*, 263–285.

(5) Kerrow, J. H.; Robinson, A. B. *Anal. Biochem.* **1971**, *42*, 565–568.

(6) Capasso, S.; Mazzarella, L.; Sica, F.; Zagari, A. *J. Peptide Res.* **1989**, *2*, 195–200.

(7) Geiger, T.; Clarke, S. *J. Biol. Chem.* **1987**, *262*, 785–794.

(8) Peters, B.; Trout, B. L. *Biochemistry* **2006**, *45*, 5384–5392.

(9) Capasso, S.; Salvadori, S. *J. Peptide Res.* **1999**, *54*, 377–382.

(10) Capasso, S.; Kirby, A.; Salvadori, S.; Sica, F.; Zagari, A. *J. Chem. Soc. Perkin Trans. 2* **1993**, 437–442.

(11) Patel, K.; Borchardt, R. T. *Pharm. Res.* **1990**, *7*, 703–711.

(12) Capasso, S.; Mazzarella, L.; Sica, F.; Zagari, A.; Salvadori, S. *J. Chem. Soc. Perkin Trans. 2* **1993**, 679–682.

(13) Konuklar, F. A.; Aviyente, V.; Sen, T. Z.; Bahar, I. *J. Mol. Model.* **2001**, *7*, 147–160.

(14) Konuklar, F. A.; Aviyente, V.; Ruiz-Lopez, M. F. *J. Phys. Chem. A* **2002**, *106*, 11205–11214.

(15) Konuklar, F. A.; Aviyente, V. *Org. Biomol. Chem.* **2003**, *1*, 2290–2297.

(16) Radkiewicz, J. L.; Zipse, H.; Clarke, S.; Houk, K. N. *J. Am. Chem. Soc.* **1996**, *118*, 9148–9155.

(17) Radkiewicz, J. L.; Zipse, H.; Clarke, S.; Houk, K. N. *J. Am. Chem. Soc.* **2001**, *123*, 3499–3506.

(18) Catak, S.; Monard, G.; Aviyente, V.; Ruiz-Lopez, M. F. *J. Phys. Chem. A* **2006**, *110*, 8354–8365.

(19) Catak, S.; Monard, G.; Aviyente, V.; Ruiz-Lopez, M. F. *J. Phys. Chem. A* **2008**, *112*, 8752–8761.

(20) Catak, S.; Monard, G.; Aviyente, V.; Ruiz-Lopez, M. F. *J. Phys. Chem. A* **2009**, *113*, 1111–1120.

(21) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103–128.

(22) Lippert, G.; Hutter, G.; Parrinello, M. *Mol. Phys.* **1997**, *92*, 477–487.

(23) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.

(24) Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302–238305.

(25) Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L. *J. Phys. Chem. B* **2005**, *109*, 6676–6687.

(26) Gervasio, F. L.; Parrinello, M.; Ceccarelli, M.; Klein, M. L. *J. Mol. Biol.* **2006**, *361*, 390–398.

(27) Lee, J. G.; Asciutto, E.; Babin, V.; Sagui, C.; Darden, T.; Roland, C. *J. Phys. Chem. B* **2006**, *110*, 2325–2331.

(28) Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600–2607.

(29) Stanton, C. L.; Kuo, I. W.; Mundy, C. J.; Laino, T.; Houk, K. N. *J. Phys. Chem. B* **2007**, *111*, 12573–12581.

(30) Ensing, B.; De Vivo, M.; Liu, Z.; Moore, P.; Klein, M. L. *Acc. Chem. Res.* **2006**, *39*, 73–81.

(31) VandeVondele, J.; Hutter, J. *J. Chem. Phys.* **2007**, *127*, 114105–114113.

(32) Schaefer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.

(33) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703–1710.

(34) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(35) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785789.

(36) VandeVondele, J.; Hutter, J. *J. Chem. Phys.* **2003**, *118*, 4365–4369.

(37) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 14101–14107.

(38) MacKerrell, A. D., Jr; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr; 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., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(39) Granovsky, A. A. PC GAMESS, version 7.1; Department of Chemistry, M. V. Lomonosov Moscow State University: Moscow, Russia, 2007.