

Thermodynamic and Kinetic Characterization of a β -Hairpin Peptide in Solution: An Extended Phase Space Sampling by Molecular Dynamics Simulations in Explicit Water

Isabella Daidone,¹ Andrea Amadei,^{2*} and Alfredo Di Nola¹

¹Department of Chemistry, University of Rome "La Sapienza," Rome, Italy

²Dipartimento di Scienze e Tecnologie Chimiche, University of Rome "Tor Vergata," Rome, Italy

ABSTRACT The folding of the amyloidogenic H1 peptide MKHMAGAAAAGAVV taken from the syrian hamster prion protein is explored in explicit aqueous solution at 300 K using long time scale all-atom molecular dynamics simulations for a total simulation time of 1.1 μ s. The system, initially modeled as an α -helix, preferentially adopts a β -hairpin structure and several unfolding/refolding events are observed, yielding a very short average β -hairpin folding time of \sim 200 ns. The long time scale accessed by our simulations and the reversibility of the folding allow to properly explore the configurational space of the peptide in solution. The free energy profile, as a function of the principal components (essential eigenvectors) of motion, describing the main conformational transitions, shows the characteristic features of a funneled landscape, with a downhill surface toward the β -hairpin folded basin. However, the analysis of the peptide thermodynamic stability, reveals that the β -hairpin in solution is rather unstable. These results are in good agreement with several experimental evidences, according to which the isolated H1 peptide adopts very rapidly in water β -sheet structure, leading to amyloid fibril precipitates [Nguyen et al., *Biochemistry* 1995;34:4186–4192; Inouye et al., *J Struct Biol* 1998;122:247–255]. Moreover, in this article we also characterize the diffusion behavior in conformational space, investigating its relations with folding/unfolding conditions. *Proteins* 2005;59:510–518.

© 2005 Wiley-Liss, Inc.

Key words: β -hairpin peptide; thermodynamic characterization; molecular dynamics simulations

INTRODUCTION

The most stable fold of a protein is determined by its amino acid composition, solvent environment (composition, pH, ionic strength) and physical state (temperature, pressure). Considering that interactions at the atomic level play a crucial role in the equilibrium between folded and unfolded conformers, molecular dynamics (MD) simulations could, in principle, be used to calculate the folded/unfolded equilibrium and could yield the kinetics of the

folding process. However, given the high computational cost required for the complete sampling of the protein-peptide configurational space, MD simulation in atomic detail of the folding/unfolding equilibrium was not, in practice, considered as a possible investigation tool. For this reason in folding studies, the molecular models used were often of a simple nature: one interaction site per residue,^{1,2} implicit solvent approximation,^{3–6} motions restricted to lattice sites,^{7,8} etc. On the other hand, other methods were developed to enhance the configurational space sampling in atomistic simulations such as essential dynamics (ED) sampling,⁹ highly parallel simulation algorithms (PRD, REMD),^{10–12} or other generalized-ensemble methods.¹³ However, the direct MD simulation of the folding/unfolding equilibrium in the canonical ensemble would be the most reliable procedure to obtain both thermodynamic and kinetic properties.

Only recently, with the aid of high power computers, all-atoms MD simulations, in explicit water, provided the folding of peptides into α -helix¹⁴ or very short β -structures.¹⁵ In a previous work¹⁶ the more complex folding of a 14-residue peptide (the prion protein H1 peptide) into a in-register β -hairpin conformation starting from an ideal α -helix has been achieved. The syrian hamster Prion protein residues 109–122 (H1 peptide) is considered to be important for the α -to- β conformational transition that leads to amyloid formation, and is responsible for prion diseases. According to several experimental evidences on the isolated H1 peptide, it adopts very rapidly in water β -sheet structure from which amyloid fibrils precipitate,^{17,18} while in 2,2,2-trifluoroethanol (TFE) or membrane-mimicking environments the H1 peptide adopts an α -helical conformation.^{19,20} These properties make the study of this peptide very interesting, and may provide a

Grant sponsor: the European Community Training and Mobility Research Network Project "Protein (mis)folding"; Grant number: HPRN-CT-2002-00241.

*Correspondence to: Andrea Amadei, Dipartimento di Scienze e Tecnologie Chimiche, University of Rome "Tor Vergata" via della Ricerca Scientifica 1, I-00133 Rome, Italy. E-mail: andrea.amadei@uniroma2.it

Received 17 June 2004; Accepted 7 November 2004

Published online 23 March 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20427

key for understanding protein folding or the cause of amyloid diseases.

In the present study, further simulations of the H1 peptide at physiological conditions have been performed to obtain a complete description of its conformational free energy landscape, including the folding/unfolding equilibrium, by means of long time scale (1.1 μ s) all-atoms MD simulations in explicit water. To our knowledge, this is one of the first attempt to simulate the thermodynamic equilibrium of a complex system, such as a β -hairpin, for more than 1 μ s using realistic models for both the peptide and the solvent and with a completely unbiased sampling.

Finally, in this article we also investigate in detail the diffusion behavior in conformational space, relating its properties with folding/unfolding conditions. Results show a characteristic dual diffusion regime, observed previously in small proteins,²¹ which can be utilized to better understand the kinetics of conformational transitions.

METHODS

MD Simulations Protocol

MD simulations, in the NVT ensemble, with fixed bond lengths²² and a time step of 2 fs for numerical integration were performed with the GROMACS software package²³ and with the GROMOS96 force field.²⁴ Water was modeled by the simple point charge (SPC) model.²⁵ A nonbond pairlist cutoff of 9.0 Å was used, and the pairlist was updated every four time steps. The long-range electrostatic interactions were treated with the particle mesh Ewald method²⁶ using a grid with a spacing of 0.12 nm combined with a fourth-order B-spline interpolation to compute the potential and forces in between grid points. The isokinetic temperature coupling²⁷ was used to keep the temperature constant at 300 K. The peptide, in its different starting conformations, was solvated with water and placed in a periodic truncated octahedron large enough to contain the peptide and ≈ 1.0 nm of solvent on all sides. In all the simulations a negative counter ion, Cl^- , was added by replacing a water molecule to achieve a neutral condition. The side chains were protonated as to reproduce a pH of about 7 and the N-terminal and C-terminal were amidated and acetylated, respectively, to reproduce the experimental conditions.¹⁷

Two all-atom MD simulations in explicit water at 300 K of the H1 peptide (MKHMAGAAAAGAVV), for a total of 1.1 $\approx \mu$ s of simulation time, were carried out: (1) 240 ns starting from the α -helix conformation obtained from the simulation in 30% (v/v) TFE/water mixture of the previous work;¹⁶ (2) 850 ns starting from the β -hairpin conformation observed in the previous simulation, using a new set of initial velocities.

ED Analysis

The principles of the ED analysis are described in detail elsewhere.^{28,29} Briefly, from all the structures of both simulations a covariance matrix of positional fluctuations (C_α only) was built and diagonalized. Eigenvectors are directions in configurational space, and the corresponding eigenvalues indicate the mean square fluctuations along

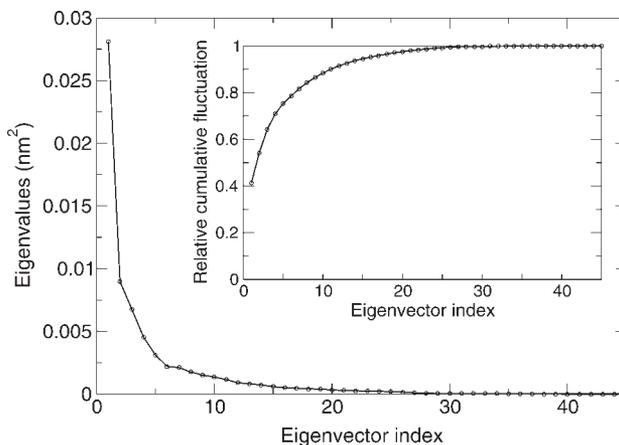


Fig. 1. Eigenvalues, in decreasing order of magnitude, obtained from the C_α coordinates covariance matrix as a function of the eigenvectors index. The corresponding relative cumulative positional fluctuation is given in the inset.

these axes. The procedure corresponds to a linear multidimensional least-squares fitting of the trajectories in configurational space. Sorting the eigenvectors by the size of the eigenvalues shows that the configurational space can be divided in a low dimensional (essential) subspace in which most of the positional fluctuations are confined, and a high dimensional (near-constraints) subspace in which merely small vibrations occur. In Figure 1, the eigenvalues obtained from C_α coordinates covariance matrix are reported as a function of eigenvectors index, and are ordered in descending order of magnitude. The corresponding relative cumulative positional fluctuation (with respect to the total positional fluctuation) is given in the inset. We used the two principal components with the highest eigenvalues, defining the first “essential plane,” for thermodynamic and kinetic calculations. This is because such a plane accounts for almost 60% of the overall positional fluctuation (inset of Fig. 1), hence describing the most relevant conformational degrees of freedom and the main conformational transitions of the peptide backbone. We performed similar calculations over planes defined by other eigenvectors. However, within such planes some of the relevant conformational transitions are not detectable, and hence the corresponding eigenvectors are not suitable as conformational coordinates to describe the large conformational fluctuations as well as the folding/unfolding transitions.

Thermodynamic Properties

Given a system in thermodynamic equilibrium, the change in free energy on going from a reference state, ref , of the system to a generic state, i (e.g., from unfolded to folded), at constant temperature and constant volume can be evaluated as

$$\Delta A_{ref \rightarrow i} = -RT \ln \frac{p_i}{p_{ref}} \quad (1)$$

where R is the ideal gas constant, T is the temperature and p_i and p_{ref} are the probabilities of finding the system in

state i and state ref , respectively. We will describe the free energy surface as a function of principal components (essential eigenvectors) from ED analysis. Structures sampled every 1 ps were projected onto the plane defined by the two first principal components. A grid 20×20 has been used to divide this plane in 400 cells and for every cell the number of points were counted and the relative probability was calculated. Finally, the $\Delta A_{ref \rightarrow i}$ was evaluated. We chose as the reference state the grid cell with the highest probability, that is, the cell corresponding to the β -hairpin folded structures ensemble. Surfaces of the total (peptide + solvent) internal energy changes, $\Delta U_{ref \rightarrow i}$, and entropy changes, $\Delta S_{ref \rightarrow i}$, were calculated as well, via the average internal energy of the simulation box in each cell and

$$\Delta S_{ref \rightarrow i} = \frac{\Delta U_{ref \rightarrow i} - \Delta A_{ref \rightarrow i}}{T} \quad (2)$$

To evaluate the local stability of the secondary structure elements we calculated for every grid cell the ratio between the number of folded structures (β -hairpin) and the number of the unfolded ones; using Equation (1) the ΔA of secondary structure formation, $\Delta A(\text{formation})$, was then evaluated for every position in the essential plane.

To check the effect of different grid spacing on the thermodynamic properties, the same type of free energy landscapes were constructed using different grids, 10×10 , 20×20 , and 30×30 (data not shown). Interestingly, all the different grids provided similar free energy landscapes with the same free energy maximum variation (≈ 14 kJ/mol), the surface being slightly more corrugated on going from the grid with a lower cell density (10×10) to the more dense one (30×30).

Kinetic Properties

For the study of diffusion properties, we chose the subspace defined by the first two essential coordinates. In particular, different regions of the essential plane, where the coordinates do not encounter a relevant free energy gradient, were analyzed separately. To generate an ensemble of independent trajectories we used all the trajectory fragments starting within one of the selected regions, and the corresponding ensemble mean square displacement, from each initial point as a function of time, was evaluated. To increase the statistics we averaged such a property over the first two essential degrees freedom, assuming for both a similar diffusion behavior. All the curve fits are obtained using the graphing tool Xmgrace (<http://plasma-gate.weizmann.ac.il/Grace/doc/UsersGuide.html>), which makes use of the Levenberg-Marquardt algorithm and provides χ^2 and correlation coefficient evaluations. Moreover, we also evaluated the noise (standard deviations, σ) for the model parameters, obtained by fitting simulation data, calculating their standard deviations over n subsets of trajectories and then extrapolating for the complete statistical sample:

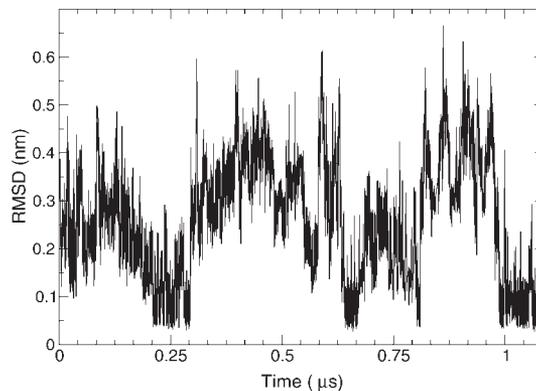


Fig. 2. Root-mean-square deviation (RMSD) of the backbone atoms with respect to the β -hairpin structure versus time.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (a_i - \bar{a})^2}{(n-1)n}} \quad (3)$$

$$\bar{a} = \frac{\sum_{i=1}^n a_i}{n} \quad (4)$$

where a_i is the generic parameter evaluated in the i th subset. Note that the previous equation is based on the approximation that the parameters obtained by the whole number of trajectories are equivalent to the ones obtained averaging the corresponding values over the n subsets. In the present case we used three independent subsets that resulted to be a good compromise between the statistics within each subset and the sample size used in the last equation, given by the number of subsets.

RESULTS

Thermodynamic Characterization of the Conformational Transitions

Two all-atom MD simulations of the syrian hamster H1 peptide, for a total simulation time of $\approx 1.1 \mu\text{s}$, were carried out. In Figure 2 and Figure 3(a) the root-mean-square deviation (RMSD), with respect to the β -hairpin structure, and the time evolution of the secondary structure are reported, respectively. Within the first $0.24 \mu\text{s}$ of simulation the α -helix structure, used as the initial simulation structure, is rapidly lost and interestingly, after $\approx 0.20 \mu\text{s}$, a β -hairpin structure is formed, with the same structural properties of the one observed in the previous work.¹⁶ Further, $0.85 \mu\text{s}$ of simulation were performed in the same conditions starting from the β -hairpin conformation. Many unfolding/refolding events of the β -hairpin are observed, with an average folding time of ≈ 200 ns, ensuring the reversibility of the folding of this peptide in the conditions used. The long time scale accessed by our simulations and the reversibility of the folding allow to properly explore the configurational space of the 14-residue peptide at physiological conditions.

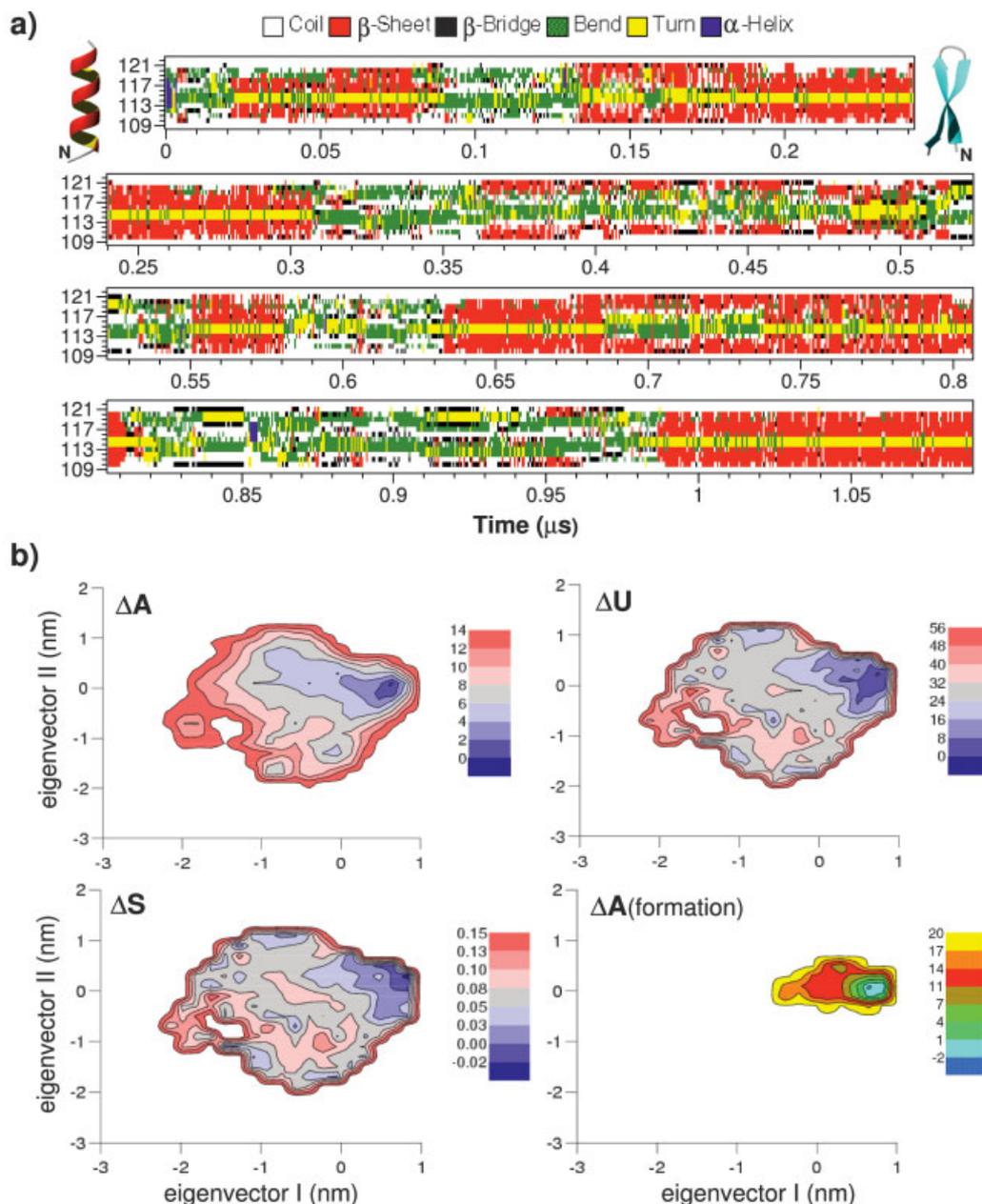


Fig. 3. (a) Time evolution of the H1 peptide secondary structure. The starting and final structures are shown on the left and right sides, respectively. The N-terminal in each snapshot is indicated with "N." In the first panel the first part of the simulation, starting from an ideal α -helix, is reported. The formation of the 2:2 β -hairpin at $t \approx 0.18 \mu\text{s}$ can be observed. In the following panels the second part of the simulation, starting from the β -hairpin structure with a new set of velocities, is reported. The analysis of the secondary structures was performed with the DSSP program.³¹ (b) Contour maps of the free energy, ΔA , internal energy, ΔU , entropy, ΔS , and free energy change associated to the β pin formation, $\Delta A(\text{formation})$, as a function of the position in the essential plane. ΔA , ΔU , and ΔS are calculated with respect to the state with the highest probability, that is, the one corresponding to the β -hairpin folded structures ensemble. Energy and entropy values are given in kJ/mol and $\text{kJmol}^{-1}\text{K}^{-1}$, respectively.

α -Helix and β -hairpin structures are populated for $\approx 5\%$ and $\approx 30\%$ of the total time, respectively. The rest is populated by partial folded β -hairpins, unfolded or "molten globule" like structures. In Figure 3(b) the free energy surface as a function of the two first essential components (see Methods) is reported. This free energy profile, ob-

tained by the probability per grid cell as described in the Methods section, shows a characteristic funneled landscape (i.e., a surface characterized by a single deep minimum) with a downhill free energy change toward the β -hairpin basin of $\approx -14 \text{ kJ/mol}$. Such a free energy surface was obtained projecting the complete 1.1- μs simu-

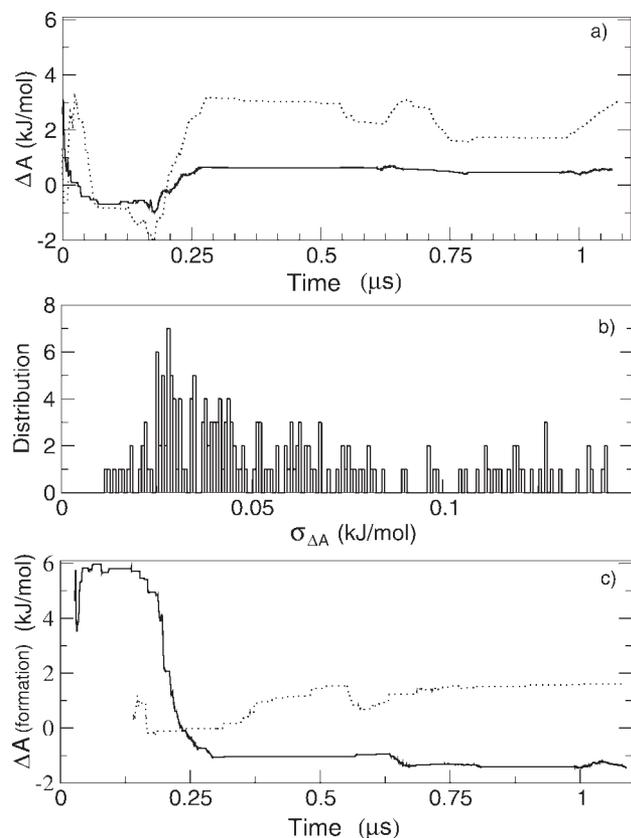


Fig. 4. Time convergence of the ΔA (a) and $\Delta A(\text{formation})$ (c) for two given essential plane positions. We chose a grid cell in the free energy minimum region, that is, within the contour line at $\Delta A = 0$ of Figure 3(b) (solid line), and a grid cell within the contour line at $\Delta A \approx 3$ of Figure 3(b) (dashed line). In (b), the probability distribution of the free energy standard deviations, $\sigma_{\Delta A}$, for all the cells are reported.

lation trajectory. To estimate its reliability we tested the convergence of the free energies within grid cells of the plane (note that free energy values are defined with respect to the grid cell corresponding to the global minimum). The results show rather stable values (within 0.1–0.2 kJ/mol) after about 0.3 μs for grid cells close to the free energy minimum. In grid cells located far from this region a worse convergence is observed, although after about 0.3 μs ΔA values are obtained within a noise of about 1.5 kJ/mol. In Figure 4(a) we report such a convergence plot for two given grid cells belonging to the previously mentioned subspaces, that is, within the contour lines $\Delta A = 0$ kJ/mol and $\Delta A \approx 3$ kJ/mol of Figure 3(b). In Figure 4(b) we report the probability distribution of the free energy standard deviations, $\sigma_{\Delta A}$, over the grid cells utilized, showing rather small statistical errors affecting the free energy values. It is worth noting that the level of convergence observed for the free energy variations considered and the very limited corresponding statistical errors, do not necessarily mean that all the possible conformational transitions are sampled with the same accuracy.

The absolute free energy minimum, as well as the adjacent region within the contour line at $\Delta A \approx 2$ kJ/mol in Figure 3(b), is mainly populated by β -hairpin structures,

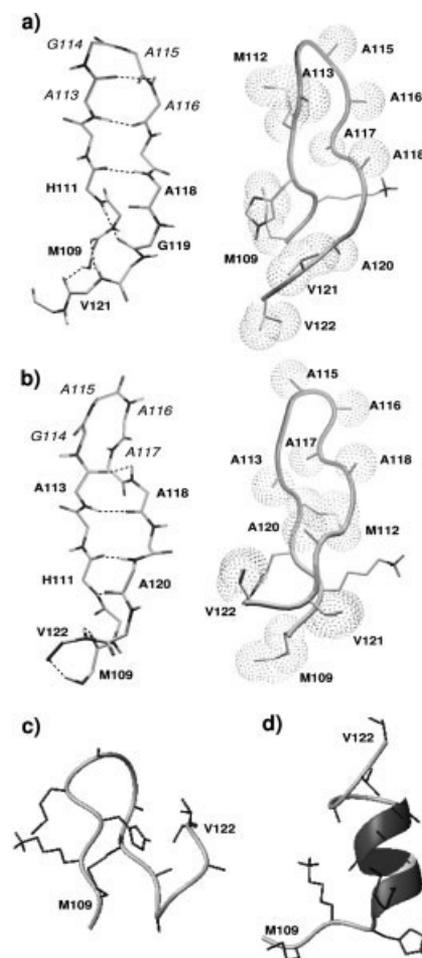


Fig. 5. Structures of the H1 peptide observed along the simulations. (a) A 2:2 β -hairpin with a type II' β -turn; (b) a 4:4 β -hairpin with a type IV β -turn. Note that in the β -hairpins the alanines, in particular in the turn regions, are exposed to the solvent; (c) a representative "molten globule"-like structure; (d) a representative α -helix structure.

including the complete β -hairpin conformation also observed in the previous work.¹⁶ Such a structure corresponds to a 2:2 β -hairpin with a type II' β -turn sequence of (A113–)G114–A115(–A116) and is characterized by six interstrand hydrogen bonds (HB), with an antiparallel bulge involving G119 [Fig. 5(a)]. A free energy plateau, within the contour line at $\Delta A \approx 6$ kJ/mol, is characterized by an ensemble of either completely unfolded or partial β -hairpin structures, which only rarely evolve into the complete β -hairpin. Such partial β -hairpin structures mainly involve two types of structured conditions: either they share the same turn of the complete β -hairpin structure, but with flanking terminals (i.e., some HB are lost), or they have a different turn type. Note that among the latter, a 4:4 β -hairpin with a type IV β -turn sequence of G114–A115–A116–A117 [Fig. 5(b)], was already observed in the simulations of the previous work.¹⁶ Three "molten globule"-like states are present with free energy local minima at ≈ 6 , 8, and 10 kJ/mol, respectively, and are characterized by bent conformations. A representative structure is given in Figure 5(c). The rest of the accessible

essential subspace corresponds basically to completely unfolded structures.

α -Helix structures [Fig. 5(d)] are sampled seven times throughout the simulations but each time for a very short period, about 500 ps. α -helix conformers do not populate any free energy minimum and are rather “disperse” through the gray plateau, within the contour line at $\Delta A \approx 8$ kJ/mol, with rather high free energies.

ΔU and ΔS profiles share the same funneled-like shape of the free energy [see Fig. 3(b)]. Note that the internal energy and the entropy values are calculated for the whole system, that is, peptide and solvent. Interestingly, the absolute free energy minimum region (the subspace inside the contour line at $\Delta A \approx 2$ kJ/mol) includes the absolute internal energy and entropy minima, thus meaning that the complete β -hairpin state is the most energetically stable, with the lowest entropy.

To evaluate the local stability of the complete β -hairpin structure, we used the same essential plane to evaluate the free energy change, $\Delta A(\text{formation})$, associated to its formation from any other possible structure [Fig. 3(b)]. This was accomplished for every position (grid cell) of the essential plane, calculating the probability for the complete β -hairpin (p_β) and for any other possible structure (p) to occur which was then used to obtain the β -hairpin formation free energy $\Delta A(\text{formation}) = -RT \ln(p_\beta/p)$. Interestingly, except for a small region corresponding to the absolute free energy minimum, β -hairpin formation free energies are always positive, thus revealing that the H1 peptide has a rather unstable secondary structure.

In Figure 4(c) we also show the convergence of such a free energy change for two grid cells, clearly showing that also for this evaluation 1.1 μs is enough to obtain reliable results. A similar evaluation for the α -helix structure is not really possible because of its rare occurrence (5% of the total simulation time). However, its very high approximate free energy shows, as expected, that the α -helix structure is very unstable (data not shown), although a relatively high number of α -helix unfolding/refolding transitions (seven times) was observed.

Finally we evaluated the global β -hairpin formation free energy, that is, over the whole accessible conformational space. This was accomplished considering the essential plane as a unique cell and evaluating the corresponding probabilities, p_β and p , from which the ΔA was then obtained. ΔU and ΔS were calculated as described in the Method section. It is interesting to note that the global β -hairpin formation free energy obtained, $\Delta A \approx +2.5$ kJ/mol, shows that the “folded structure” is not the thermodynamic most stable condition for this peptide in water. Such a feature is due to the entropy decrease (≈ -0.070 kJ mol $^{-1}$ K $^{-1}$), which overcompensates the internal energy stabilization (≈ -18.7 kJ/mol). Even in the absolute free energy minimum, where the β -hairpin structure is mostly stable, its formation free energy, $\Delta A(\text{formation})$, is only about -2 kJ/mol [Fig. 3(b)].

Our results are consistent with experimental data on a non-amyloidogenic 16 residues β -hairpin peptide using a nanosecond laser temperature-jump technique.³⁰ These

experimental data provided an apparent ΔG for the β -hairpin folding transition of ≈ -2.5 kJ/mol, a value close to our estimate for the H1 peptide, although the latter is ≈ 5 kJ/mol less stable. This relative instability could explain the amyloidogenic nature of the H1 peptide.

Kinetic Characterization of the Conformational Transitions

In the previous subsection we characterized the thermodynamics within the conformational space of the H1 peptide. In this subsection we characterize its kinetics in the essential plane used for the previous thermodynamic analysis. Note that the essential eigenvectors defining such a plane represent the most relevant conformational degrees of freedom of the peptide backbone, hence describing the main conformational transitions. In a previous article²¹ it was shown that the kinetics of the essential degrees of freedom in proteins can be described by a diffusion behavior characterized by a dual regime: a fast type of diffusion within a single energy local minimum, switching exponentially to a slower one, probably corresponding to “hopping” between multiple harmonic wells. Such a dual diffusion behavior should be determined by the relaxation of the medium (defined by all the other coordinates) associated to the hopping and resulting into an increase of viscosity.

In the previous article²¹ we characterized the diffusion using relatively short time intervals (up to 20 ps). In this study, due to the huge simulation time available, we can afford a better statistical characterization of the essential degrees of freedom diffusion, extending our investigation over longer times (up to 100 ps). However, a single-exponential relaxation of the velocity autocorrelation function, which was utilized to describe such a conformational diffusion in the previous study,²¹ is not really suitable to describe this process over longer time intervals (up to 100 ps) afforded in the present study. A more accurate model can be obtained considering two relaxation modes of the velocity autocorrelation function (corresponding to a bi-exponential switching from the fast to the slow diffusion regime), which can be considered as a generalization of the previous model (see Appendix). The equation obtained from this generalized model, for the mean square displacement, neglecting the initial fast (within a few tens of fs) relaxation, is for a given q (essential) degree of freedom

$$\langle \Delta q^2(t) \rangle \cong 2D_\infty t + 2[D_0 - A_1]\tau_1[1 - e^{-t/\tau_1}] + 2[D_0 - A_2]\tau_2[1 - e^{-t/\tau_2}] \quad (5)$$

where D_∞ is the long-time diffusion constant, D_0 the short-time diffusion constant, τ_1 , τ_2 the “relaxation times” of the two switching modes and A_1, A_2 two parameters defined by the integral of the velocity autocorrelation function (see Appendix).

Such a model was used to describe the diffusion in the configurational subspace defined by the first two C_α essential degrees of freedom, assuming at least for the “relaxation times” the same behavior (this was actually checked to be a good approximation). To increase the statistics, we

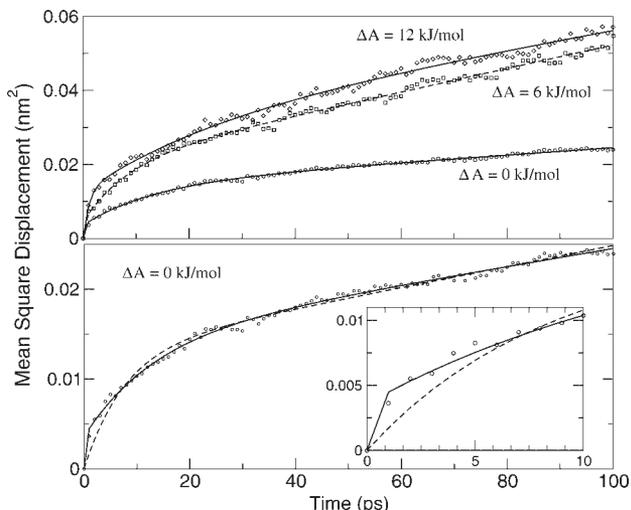


Fig. 6. (Upper panel) Mean square displacement, as a function of time, averaged over the first two principal eigenvectors. The theoretical models (solid line), derived in the Appendix, were parameterized fitting the simulation data. In particular, we report results for three selected regions of the essential plane, one in the region of the free energy minimum ($\Delta A = 0$ kJ/mol) (circles), another in the region of the free energy plateau within the contour line at $\Delta A \approx 6$ kJ/mol (squares), and the third in the completely unfolded region ($\Delta A \approx 12$ kJ/mol) (diamonds). (Lower panel) A comparison between the previous model, based on a single-exponential relaxation, and the present one, based on a bi-exponential relaxation, is shown for one of the regions ($\Delta A = 0$ kJ/mol). In the inset the first 10 ps are shown in more details.

averaged the mean square displacement over the first two essential degrees of freedom.

In Figure 6, we show the comparison between the theoretical models and the ensemble mean square displacements obtained by simulations. In particular, we report the results obtained for three selected regions of the essential plane, one in the region of the free energy minimum ($\Delta A = 0$ kJ/mol), another in the region of the free energy plateau within the contour line at $\Delta A \approx 6$ kJ/mol and the third in the completely unfolded region ($\Delta A \approx 12$ kJ/mol). The plot reported in the upper panel clearly shows the high accuracy of the model used in the whole time range. Note that for all the three theoretical models the χ^2 values are in the range 10^{-5} – 10^{-4} with correlation coefficients always higher than 0.997 and full fitting convergence was achieved within 500 steps. In the lower panel a comparison between the simpler model, as used in the previous article,²¹ and the present generalized one is shown for one of the regions ($\Delta A = 0$ kJ/mol). It is evident how the bi-exponential relaxation of the velocity autocorrelation function provides a more accurate model. In particular, a dramatic improvement is observed in the first few ps of diffusion (see inset of Fig. 6).

The diffusion constants and the “relaxation times” are reported in Table I. Although the short-time diffusion constants, D_0 , are of the same order of magnitude for all the regions, revealing a similar diffusion behavior within a single local energy well, the long-time diffusion constants, D_∞ , are similar for the less structured regions, but it is significantly lower in the β -hairpin region. Thus, when the system enters its long-time diffusion regime, hopping

between multiple energy basins, the more structured state encounters a greater viscosity of the medium defined by the other coordinates including the solvent.

For what concerns the “relaxation times,” that is, the time required to switch from the fast to the slower diffusion behavior, the most striking difference can be observed for the slower mode relaxation time, τ_2 , values which are similar for the two more structured regions, while for the completely unfolded one its value is almost double. This could be explained considering the roughness of the internal energy surface in the unfolded region [left side of the ΔU landscape in Fig. 3(b)]. The presence of deep valleys and high mountains, with internal energy differences up to ≈ 25 kJ/mol, could be the cause of the longer time required for spreading the trajectories over such a corrugated internal energy region. Interestingly, a similar trend is observed for the faster mode relaxation time, τ_1 .

CONCLUSIONS

The thermodynamic and kinetic properties of the H1 peptide MKHMAGAAAAGAVV taken from the syrian hamster Prion protein was explored in explicit aqueous solution at 300 K using long time scale all-atom MD simulations in the canonical ensemble for a total simulation time of 1.1 μ s. To our knowledge, this is one of the first attempt to simulate the thermodynamic equilibrium of a complex system, such as a β -hairpin, for more than 1 μ s using realistic models for both the peptide and the solvent and with a completely unbiased sampling of the configurational space. The peptide, initially modeled as an α -helix, preferentially adopts β -hairpin structures; furthermore, many unfolding/refolding events of the β -hairpin were observed, with an average folding time of ≈ 200 ns. The free energy profile, as a function of the first two essential eigenvectors, that represent the most relevant conformational degrees of freedom of the peptide backbone, has the characteristic features of a funneled landscape, with a downhill surface toward the bottom. ΔU and ΔS profiles share the same funnel-like shape of the free energy and their absolute minima almost correspond to the absolute free energy minimum region. Although complete β -hairpin structures mostly populate the free energy minimum, its global free energy of formation, from any other structure, is $\approx +2.5$ kJ/mol. This positive value clearly shows that the “folded structure” is not the thermodynamic most stable condition for this peptide in water. Such a feature is due to the entropy decrease (≈ -0.071 kJ mol⁻¹ K⁻¹), which overcompensates the internal energy stabilization (≈ -18.7 kJ/mol).

According to several experimental evidences, the H1 peptide adopts very rapidly in water β -sheet structure from which amyloid fibrils precipitate,^{17,18} in agreement with our results. Considering the relative instability of the β -hairpin structure in water, revealed in the present study, the interaction with other monomers could be a source of stabilization, leading to amyloid fibril formation.

Furthermore, in this article, we also characterize the diffusion behavior in conformational space, investigating its relations with folding/unfolding conditions. The results

TABLE I. Diffusion Constants and “Relaxation Times” for Three Selected Regions of the Essential Plane, One in the Region of the Free Energy Minimum ($\Delta A = 0$ kJ/mol), Another in the Region of the Free Energy Plateau within the Contour Line at $\Delta A \approx 6$ kJ/mol, and the Third in the Completely Unfolded Region ($\Delta A \approx 12$ kJ/mol)

Region	D_0^a nm ² ps ⁻¹	D_∞^a nm ² ps ⁻¹	τ_1^a ps	τ_2^a ps
$\Delta A = 0$ kJ/mol	.026 (0.001)	$5.3 \cdot 10^{-5}$ ($0.4 \cdot 10^{-5}$)	< 1	13.2 (2.4)
$\Delta A \approx 6$ kJ/mol	.032 (0.001)	$15.9 \cdot 10^{-5}$ ($1.5 \cdot 10^{-5}$)	< 1	8.5 (1.1)
$\Delta A \approx 12$ kJ/mol	.032 (0.001)	$12.8 \cdot 10^{-5}$ ($1.9 \cdot 10^{-5}$)	≈ 1.1	24.8 (2.1)

^a D_0 is the short-time diffusion constant, D_∞ the long-time diffusion constant and τ_1, τ_2 the “relaxation times” of the two switching modes (see Appendix). Standard deviations (see Methods) are given in parentheses.

show that it is possible to accurately describe the kinetics, over the same essential plane used for the thermodynamic characterization, with a dual diffusion model. A first diffusion regime, up to a few ps, probably corresponding to the diffusion of the essential coordinates in a single energy basin, is characterized by a higher diffusion constant. The second diffusion mode is probably connected with the motions from one energy well to another, and is characterized by a lower diffusion constant, resulting from an increased friction due to the solvent and the other nonessential coordinates. Moreover, in our model a bi-exponential switching (i.e., two relaxation times) from the faster to the slower diffusion mode, yields a very accurate description of the diffusion of the essential coordinates over time intervals up to 100 ps.

Different diffusion behaviors have been observed in relation to the degree of unfolding of the peptide. The more structured regions of the essential plane seem to be associated with a slower long-time diffusion (i.e., higher viscosity of the medium of the other coordinates) with respect to the less structured ones, for which the D_∞ values are almost three times larger. Interestingly, the relaxation times required to switch from the faster to the slower diffusion regime are longer for the largely unfolded conformational region, being almost double with respect to the folded or partially folded regions. This could be due to the higher roughness of the internal energy surface in the unfolded region, resulting in higher energy barriers to be crossed for spreading the trajectories from an energy well to the others.

REFERENCES

- Erman B, Dill KA. Gaussian theory of protein folding. *J Chem Phys* 2000;112:1050–1056.
- Zhou Y, Karplus M. Interpreting the folding kinetics of helical proteins. *Nature* 1999;401:400–403.
- Schaefer M, Bartels C, Karplus M. Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J Mol Biol* 1998;284:835–848.
- Wang H, Varady J, Ng L, Sung SS. Molecular dynamics simulations of β -hairpin folding. *Proteins Struct Funct Genet* 1999;37:325–333.
- Ulmschneider J, Jorgensen W. Polypeptide folding using monte carlo sampling, concerted rotation, and continuum solvation. *J Am Chem Soc* 2004;126:1849–1857.
- Snow C, Qiu L, Du D, Gai F, Hagen S, Pande VS. Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. *Proc Natl Acad Sci USA* 2004;101:4077–4082.
- Gutin AM, Abkevich VI, Shakhnovich EI. A protein engineering analysis of the transition state for protein folding: simulation in the lattice model. *Fold Des* 1998;3:183–194.
- Klimov DK, Thirumalai D. Mechanisms and kinetics of β -hairpin formation. *Proc Natl Acad Sci USA* 2000;97:2544–2549.
- Daidone I, Amadei A, Roccatano D, Di Nola A. Molecular dynamics simulation of protein folding by essential dynamics sampling: folding landscape of horse heart cytochrome *c*. *Biophys J* 2003;85:2865–2871.
- García AE, Sanbonmatsu KY. Exploring the energy landscape of β hairpin in explicit solvent. *Proteins Struct Funct Genet* 2001;42:345–354.
- Gnanakaran S, Nymeyer H, Portman J, Sanbonmatsu KY, García AE. Peptide folding simulations. *Curr Opin Struct Biol* 2003;13:168–174.
- Yang WY, Pitera JW, Swope WC, Gruebele M. Heterogeneous folding of the trpzip hairpin: full atom simulation and experiment. *J Mol Biol* 2004;336:241–251.
- Mitsutake A, Sugita Y, Okamoto Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* 2001;60:96–123.
- van Gunsteren WF, Bürgi R, Peter C, Daura X. The key to solving the protein-folding problem lies in an accurate description of the denatured state. *Angew Chemie Int Ed* 2001;40:351–355.
- Daura X, van Gunsteren WF, Mark AE. Folding-unfolding thermodynamics of a β -heptapeptide from equilibrium simulations. *Proteins Struct Funct Genet* 1999;34:269–280.
- Daidone I, Simona F, Roccatano D, Broglia R A, Tiana G, Colombo G, Di Nola A. β hairpin conformation of fibrillogenic peptides: structure and α - β transition mechanism revealed by molecular dynamics simulations. *Proteins Struct Funct Bioinf* 2004;57:198–204.
- Nguyen J, Baldwin MA, Cohen FE, Prusiner SB. Prion protein peptides induce α -helix to β -sheet conformational transitions. *Biochemistry* 1995;34:4186–4192.
- Inouye H, Kirschner DA. Polypeptide chain folding in the hydrophobic core of hamster scrapie prion: analysis by x-ray diffraction. *J Struct Biol* 1998;122:247–255.
- Heller J, Kolbert AC, Larsen R, Ernst M, Bekker T, Baldwin M, Prusiner SB, Pines A, Wemmer DE. Solid-state NMR studies of the prion protein H1 fragment. *Protein Sci* 1996;5:1655–1661.
- Jayawickrama D, Zink S, Velde DV, Effiong R, Larive C. Conformational analysis of the β -amyloid peptide fragment, β (12–28). *J Biomol Struct Dyn* 1995;13:229–244.
- Amadei A, Ceruso MA, Di Nola A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins molecular dynamics simulations. *Proteins Struct Funct Genet* 1999;36:419–424.
- Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. Lincs: a linear constraint solver for molecular simulations. *J Comp Chem* 1997;18:1463–1472.
- van der Spoel D, van Drunen R, Berendsen HJC. GROningen MACHine for Chemical Simulation. Nijenborgh, Groningen: Department of Biophysical Chemistry, BIOSON Research Institute; 1994.
- van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG. Biomolecular simula-

tion: the GROMOS96 manual and user guide. Zürich: Hochschulverlag AG an der ETH Zürich; 1996.

25. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. *J Phys Chem* 1987;91:6269–6271.
26. Darden T, York D, Pedersen L. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J Chem Phys* 1993;98:10089–10092.
27. Brown D, Clarke JHR. A comparison of constant energy, constant temperature, and constant pressure ensembles in molecular dynamics simulations of atomic liquids. *Mol Phys* 1984;51:1243–1252.
28. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. *Proteins Struct Funct Genet* 1993;17: 412–425.
29. de Groot BL, Amadei A, Scheek RM, van Nuland NA, Berendsen HJC. An extended sampling of the configurational space of HPr from *E. coli*. *Proteins Struct Funct Genet* 1996;26:314–322.
30. Muñoz VM, Thompson P, Hofrichter J, Eaton W. Folding dynamics and mechanism of β -hairpin formation. *Nature* 1997;390:196–199.
31. Kabsch W, Sander C Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* 1983;22:2577–2637.

APPENDIX

In this section the theory used to model the kinetics in the essential plane will be described in detail.

Given a coordinate q , the ensemble mean square displacement from an initial point, as a function of time, can be expressed as:

$$\langle \Delta q^2(t) \rangle = 2 \int_0^t I(t') dt' \quad (6)$$

with $\Delta q(t) = q(t) - q(0)$ and

$$I(t') = \int_0^{t'} \gamma(t'') dt'' \quad (7)$$

where $\gamma(t'') = \langle \dot{q}(0)\dot{q}(t'') \rangle$ is the velocity autocorrelation function of q . As in the previous article,²¹ the function $I(t)$ is considered rapidly converging to a positive value within t_0 , corresponding to a fast first relaxation of the order of 30–40 fs, while for $t > t_0$ a second slower, first-order relaxation is used to model the slowly converging tail in the velocity autocorrelation function. However, different from the previous model where the diffusion was studied using relatively short time intervals (up to 20 ps) and a single-exponential mode was utilized, extension over longer time intervals (up to 100 ps) afforded in this study shows that a bi-exponential relaxation of the velocity autocorrelation function is necessary to model accurately the diffusion. Hence, considering $\gamma(t) = \gamma_1(t) + \gamma_2(t)$, Equation (6) becomes

$$\langle \Delta q^2(t) \rangle = 2 \left[\int_0^{t_0} I_1(t') dt' + \int_{t_0}^t I_1(t') dt' + \int_0^{t_0} I_2(t') dt' + \int_{t_0}^t I_2(t') dt' \right] \quad (8)$$

where

$$I_1(t') = \int_0^{t'} \gamma_1(t'') dt'' \quad I_2(t') = \int_0^{t'} \gamma_2(t'') dt'' \quad (9)$$

If we assume, for $t > t_0$, a simple first-order kinetics affecting the two components, then

$$I_1(t') = [I_1(t_0) - I_1(\infty)]e^{-(t'-t_0)/\tau_1} + I_1(\infty)$$

$$I_2(t') = [I_2(t_0) - I_2(\infty)]e^{-(t'-t_0)/\tau_2} + I_2(\infty)$$

with relaxation time constants τ_1 and τ_2 . Therefore

$$\langle \Delta q^2(t) \rangle = 2\Delta + 2I_1(\infty)(t - t_0) + 2[I_1(t_0) - I_1(\infty)]\tau_1[1 - e^{-(t-t_0)/\tau_1}] + 2I_2(\infty)(t - t_0) + 2[I_2(t_0) - I_2(\infty)]\tau_2[1 - e^{-(t-t_0)/\tau_2}] \quad (10)$$

with

$$\Delta = \int_0^{t_0} I(t') dt'.$$

Finally, considering that for a time range up to 100 ps we can neglect the initial fast convergence, Δ , $t_0 \approx 0$, Equation (10) becomes

$$\langle \Delta q^2(t) \rangle \cong 2D_\infty t + 2[D_0 - A_1]\tau_1[1 - e^{-t/\tau_1}] + 2[D_0 - A_2]\tau_2[1 - e^{-t/\tau_2}] \quad (11)$$

where

$$D_0 = I_1(t_0) + I_2(t_0)$$

$$D_\infty = I_1(\infty) + I_2(\infty)$$

$$A_1 = I_1(\infty) + I_2(t_0)$$

$$A_2 = I_2(\infty) + I_1(t_0)$$

Equation (11) was used to evaluate the time behavior of $\langle \Delta q^2(t) \rangle$ in the time range of 1–100 ps. In particular, three structurally different regions of the essential plane, where the coordinates do not encounter a relevant free energy gradient, were analyzed.