Solvent Electrostriction-Driven Peptide Folding Revealed by Quasi-Gaussian Entropy Theory and Molecular Dynamics Simulation

Frank Noé,[†,‡] Isabella Daidone,[‡] Jeremy C. Smith,[‡,§] Alfredo di Nola,[⊥] and Andrea Amadei*[,¶]

DFG Research Center Matheon, Free University of Berlin, Arnimallee 6, 14159 Berlin, Germany, Interdisciplinary Center for Scientific Computing, University of Heidelberg, Im Neuenheimer Feld 368, 69120 Heidelberg, Germany, Center for Molecular Biophysics, University of Tennessee/Oak Ridge National Laboratory, One Bethel Valley Road, P.O. Box 2008, Oak Ridge, Tennessee 37831-6255, Department of Chemistry, University of Rome "La Sapienza", P. le Aldo Moro 5, 00185 Rome, Italy, and Department of Chemical Sciences and Technology, University of Rome "Tor Vergata", Via della Ricerca Scientifica 1, 00133 Rome, Italy

Received: February 16, 2008; Revised Manuscript Received: June 16, 2008

A quantitative understanding of the complex relationship between microscopic structure and the thermodynamics driving peptide and protein folding is a major goal of biophysical chemistry. Here, we present a methodology comprising the use of an extended quasi-Gaussian entropy theory parametrized using molecular dynamics simulation that provides a complete description of the thermodynamics of peptide conformational states. The strategy is applied to analyze the conformational thermodynamics of MR121-GSGSW, a peptide well characterized in experimental studies. The results demonstrate that the extended state of the peptide possesses the lowest partial molar entropy. The origin of this entropy decrease is found to be in the increase of the density and orientational order of the hydration water molecules around the peptide, induced by the "unfolding". While such a reduction of the configurational entropy is usually associated with the hydrophobic effect, it is here found to be mainly due to the interaction of the solute charges with the solvent, that is, electrostriction.

1. Introduction

The conformational thermodynamics of folding/unfolding processes of peptides and proteins has been greatly studied.[1−4] For the folded state to be stable in equilibrium, the increase of the configurational volume of a polypeptide upon unfolding needs to be compensated for by the remaining terms, which may be (i) increased solute potential energy, (ii) increased solvent potential and solute−solvent interaction energies, and (iii) decreased solvent entropy due to solvent rearragement around the solute. Terms (i)−(iii) are often combined to form the energy landscapes depicted in protein folding funnels.[5,6] A quantitative understanding of the relative contributions of these terms to unfolding is still elusive, and in particular, the question remains as to whether unfolding is avoided mainly by an increase of the energy or a decrease of the entropy of the complete solute−solvent system at fixed pressure.

Protein folding may be largely driven by the hydrophobic effect, causing an initial collapse of the extended chain.[7,8] An indication of this is the positive change of heat capacity typically observed for protein unfolding.[9] Exposure of hydrophobic solutes may cause a decrease of entropy[10] due to the necessary creation of a cavity in the solvent[2] and water depletion for hydrophobic clusters of sufficient size.[11] Hydrophobic solvation may also increase solvent potential energy as water−water hydrogen bonds may be disrupted due to contacts of water with the hydrophobic surface.[1,2] For large enough solutes, the effect of hydrophobic solvation on the solute chemical potential is approximately proportional to the solvent-accessible hydrophobic surface area,[8,12,13] and this has driven the incorporation of corresponding terms into widely used computational implicit solvent models.

Solvation of ions can also induce order in the hydration water structure, in particular, by increasing the solvent density and the number of orientationally ordered solvent molecules close to the solute, thus reducing the entropy, an effect termed electrostriction.[1] Ion solvation typically decreases the solute−solvent interaction energy, but if this effect can be compensated for by energetic favorable ion pairing in the folded state, a lower entropy in the unfolded state due to electrostriction could, in principle, drive the chain collapse. Indeed, in some cases of peptide unfolding, negative heat capacities have been found,[14,15] and in other studies, the magnitude of the peptide partial molar entropy decrease upon charge solvation has been estimated to be similar to that of hydrophobic solvation for protein unfolding.[10] However, relatively little attention has been directed to ion desolvation in peptide/protein unfolding studies, and its effect on entropy is not considered explicitly in widely used computational implicit solvent models.

In the present paper, we show a case of a mixed hydrophobic/hydrophilic peptide whose extended state is destabilized by entropy decrease that is, to a significant amount, due to charge solvation. The peptide is a small glycine−serine repeat modified with the oxazine derivative MR121 (a fluorescent dye) and tryptophan residue (a specific quencher) at the terminal ends, MR121-GSGSW (see Figure 1). In proteins, the glycine−serine motif is frequently found in hairpins and loops, and hence, glycine−serine-based peptides are commonly used as model systems for studying end-to-end contact formation in unstructured polypeptide chains.[16] Experimentally, the intrachain

* To whom correspondence should be addressed. E-mail: andrea.amadei@uniroma2.it. Phone: +39 06 7259 4905. Fax: +39 06 7259 4328.
[†] Free University of Berlin.
[‡] University of Heidelberg.
[§] University of Tennessee/Oak Ridge National Laboratory.
[⊥] University of Rome "La Sapienza".
[¶] University of Rome "Tor Vergata".

10.1021/jp801391t CCC: $40.75
Published on Web 08/12/2008

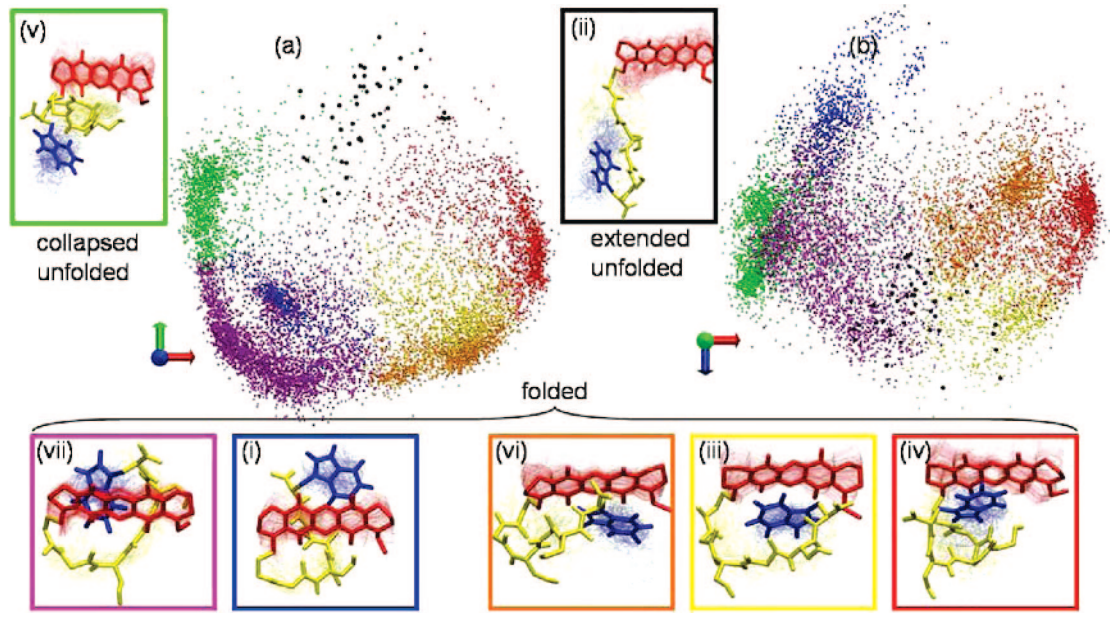

Figure 1. Visualization of the trajectory in the three-dimensional principal subspace. Each dot corresponds to one structure of the simulation projected onto the principal subspace. The colors indicate the assignment to the seven conformational states. (a) View on the $x-y$ plane of the principal subspace. (b) View on the $x-z$ plane of the principal subspace. (i)−(vii) Representative structures for the 7 conformational states, shown with an overlay of line representations of 100 structures in the same state that are randomly drawn from the trajectory. The dye MR121 is indicated in red, and the fluorescent tryptophan side chain is in blue. The frame colors correspond to the colors of the points in the principal subspace projection.

contact formation is studied by monitoring selective fluorescence quenching of the MR121 fluorophore by the Trp. End-to-end contact formation and dissociation result in "off" and "on" switching of the fluorescence signal, respectively, providing information on the peptide conformational thermodynamics and kinetics.[17]

Recent advances in computational power allow the thermodynamics of reversibly folding peptides to be computed using explicit solvent and thus to produce testable predictions.[18−20] However, a theory that, from a set of simulations, computes the complete experimentally relevant thermodynamics for each conformational state of a system has been, as yet, elusive. Here, addressing this need, we present an extension of the quasi-Gaussian entropy (QGE) theory. The theory provides a quantitative model for the full partial molecular thermodynamics, that is, changes in partial molecular energy, entropy, heat capacity, and volume, for all conformational states depending on temperature. The partial molecular thermodynamics provides the change of thermodynamic properties upon insertion of a single molecule at constant pressure, analogously to the partial molar thermodynamics. The QGE theory presented earlier[21−24] is basically an extension of the fluctuation theory.[25] In QGE theory, the fundamental expressions of statistical mechanics are reformulated in terms of the distribution function of the fluctuations of a macroscopic property, such as the potential energy of the system. By modeling the distribution of this property, an analytical solution for the thermodynamics of the system can be obtained. It has been shown that this theory reproduces experimental fluid−liquid state thermodynamics with high accuracy for a variety of physical−chemical systems over very large temperature and density ranges.[26−30] However, attempts to describe the statistical mechanics of conformational equilibria based on simulation data have been hitherto limited to simple molecules in the (ideal) gas phase.[31]

The extended QGE theory is applied here to derive the thermodynamics associated with molecular dynamics (MD) simulations of the peptide at several temperatures between 293 and 600 K. The different conformational states of the peptide are defined by the long-lived states at the experimental temperature of 293 K.[17] The probability of open conformations quantitatively agrees with the experimentally measured probability of fluorescent conformations. Using estimations of the mean potential energies and free-energy changes of the conformational states from MD simulation, an analytical QGE model is constructed that describes the complete thermodynamics of the conformational states over a large temperature range. In particular, we analyze the temperature dependence of partial molecular Gibbs and Helmholtz free energies, internal energy, and entropy and relate these properties to the structures of the corresponding conformations. Furthermore, by analyzing the QGE model parameters, the physical relationships between molecular structures, partial molecular volumes, and energetics are revealed. In this way, an extended state is found that possesses the lowest partial molecular energy and entropy of all states. An analysis of the solvent structure and density shows that this behavior is due to a decrease of the solvent volume per molecule and to an increase of the number of orientationally ordered solvating water molecules. A major part of these effects is, in turn, shown to be induced by the solvation of charges upon extension, that is, electrostriction.

2. Theory

In the QGE theory, the thermodynamic quantities of the system are expressed in terms of an excess with respect to a theoretical thermodynamic reference condition (QGE reference condition). Such a reference state is identical to the actual system conditions for chemical composition, number of molecules, volume, and temperature, but its Hamiltonian does not include any intra- and intermolecular potential energy, that is, molecules do not interact, and only the kinetic energy, the reference (quantum) vibrational energy and the reference electronic ground-state energy are considered. Therefore, for such a virtual ideal gas state, the semiclassical degrees of freedom of each

molecule move freely, and the (quantum) vibrational modes within the molecule, classically equivalent to holonomic constraints and typically associated with bond length and angle coordinates, ensure the topological stability of the molecules. The thermodynamics of this reference condition can be typically obtained by simple statistical mechanical calculations, while the excess thermodynamics can be expressed by the distribution of the system's potential energy fluctuations. Hence, for a given model distribution considered, a corresponding analytical model for the excess thermodynamics is obtained. It has been shown that the QGE Gamma state (i.e., the QGE solution based on modeling the potential energy fluctuations by a Gamma distribution) provides an accurate quantitative description of the excess thermodynamics of liquid systems made of rigid or semirigid molecules.[26−30,32−34] Here, the solute is considered to be infinitely diluted, and therefore, the complete solution thermodynamics can be obtained considering only a single solute embedded in a large number of solvent molecules,[33] that is, the simulation box termed "whole system" here.

In order to treat the thermodynamics of a flexible molecule like a peptide solvated by water molecules (rigid solvent molecules), we considered that it is always possible to partition the peptide intramolecular configurational space such that each configurational subspace (conformational state) is equivalent to a semirigid molecular species that is hence covered by the QGE Gamma state solution. Therefore, the overall thermodynamics of the solute−solvent system can be described by a set of Gamma states.

To express the solute conformational state thermodynamics in the QGE framework, the chemical potential of the solute's *i*th conformational state, μ_i, is related to the corresponding excess chemical potential, μ_i'[33,35] (the chemical potential shift between the actual and QGE reference conditions)

$$\Delta\mu_i = \Delta\mu_i' + \Delta\mu_{i,\mathrm{ref}} = \mu_i - \mu_\mathrm{r} = A_i - A_\mathrm{r} = -k_\mathrm{B}T \ln \frac{P_i}{P_\mathrm{r}} \quad (1)$$

The Δ always refers to the change of a property in a given conformation with respect to its value in the reference conformational state *r*, for example, $\Delta\mu_i' = \mu_i' - \mu_\mathrm{r}'$. A_i and A_r are the Helmholtz free energies of the whole solute−solvent system with the solute in the *i*th and reference conformations, respectively, the subscript ref indicates properties in the QGE reference condition, and *P* is the equilibrium probability of a given conformational state. The reference chemical potential change in eq 1 is readily obtained from the definition of the QGE reference condition,[33,35] assuming no significant variation of the solute quantum vibrational partition function for the different conformational states

$$\Delta\mu_{i,\mathrm{ref}} = -k_\mathrm{B}T \ln \frac{\int_i \sqrt{\det \tilde{M}}\,\mathrm{d}\boldsymbol{x}_\mathrm{in}}{\int_\mathrm{r} \sqrt{\det \tilde{M}}\,\mathrm{d}\boldsymbol{x}_\mathrm{in}} = -k_\mathrm{B}T \ln \frac{\sqrt{\det \tilde{M}_i}}{\sqrt{\det \tilde{M}_\mathrm{r}}} - k_\mathrm{B}T \ln \frac{\Omega_i}{\Omega_\mathrm{r}} \quad (2)$$

with

$$\sqrt{\det \tilde{M}_i} = \frac{\int_i \sqrt{\det \tilde{M}}\,\mathrm{d}\boldsymbol{x}_\mathrm{in}}{\int_i \mathrm{d}\boldsymbol{x}_\mathrm{in}} \qquad \Omega_i = \int_i \mathrm{d}\boldsymbol{x}_\mathrm{in}$$

and likewise for state *r*. Here, $\boldsymbol{x}_\mathrm{in}$ are the solute (classical) internal coordinates, the subscript of the integral sign indicates that integration is taken only over the intramolecular configurational subspace associated with the corresponding conformational state and $\tilde{M}$ is the (classical) solute mass tensor. Note that in the present case where only the bond length stretching vibrational modes are considered in the peptide (quantum) vibrational partition function, the assumption that vibrational energies are independent of the conformational states is an excellent approximation. In fact, for the high-frequency stretching modes, the atomic environment modifications induced by the conformational transitions, which are unaccompanied by covalent rearrangement, typically provide frequency variations within a few cm^{-1}, negligible in our calculations.

The excess chemical potential in eq 1 is provided by the Gamma state solution of the QGE theory[33]

$$\Delta\mu_i' = \Delta u_{0,i}' - \Delta c_{V0,i}' T_0 \Lambda(T) + p'(T)\Delta v_i - k_\mathrm{B}T \ln \frac{\bar{\epsilon}_i}{\bar{\epsilon}_r} \quad (3)$$

Substituting eqs 2 and 3 into eq 1 provides along an isochore

$$\Delta\mu_i = \Delta u_{0,i}' - \Delta c_{V0,i}' T_0 \Lambda(T) + p'(T)\Delta v_i - k_\mathrm{B}T \ln \gamma_i \quad (4)$$

with $\Delta u_{0,i}' = u_i'(T_0) - u_\mathrm{r}'(T_0)$, $\Delta c_{V0,i}' = c_{V,i}'(T_0) - c_{V,\mathrm{r}}'(T_0)$, $\Delta v_i = v_i - v_\mathrm{r}$, and $\gamma_i = \Omega_i \bar{\epsilon}_i \sqrt{(\det \tilde{M}_i)}\,/\Omega_\mathrm{r}\bar{\epsilon}_\mathrm{r}\sqrt{(\det \tilde{M}_\mathrm{r})}$. Here, u' is the excess partial molecular internal energy, c'_V the excess isochore partial molecular heat capacity, and v the partial molecular volume, which is temperature independent along an isochore for a Gamma state. The $\bar{\epsilon}$is the confinement fraction providing a pure entropic partial molecular term that is typically associated with the hard-body-excluded volume; $p' = p - p_\mathrm{ref}$ is the excess pressure, and T_0 is the reference temperature

$$\Lambda(T) = \frac{1}{\delta_0} + \frac{T}{T_0\delta_0^2} \ln \frac{T(1-\delta_0)}{T(1-\delta_0) + T_0\delta_0}$$

where δ_0 is a dimensionless constant. Note that within solute infinite dilution conditions, p' and Λ, being intensive properties, are fully determined by the solvent, that is, they are identical to the corresponding pure solvent Gamma state functions.[33,35]

Using general thermodynamic relations for excess partial molecular properties,[33] we may relate the partial molecular Helmholtz free energy *a* and the chemical potential with the corresponding excess properties via

$$\mu' = a' + p'v = \mu - \mu_\mathrm{ref} = a - a_\mathrm{ref} + pv - p_\mathrm{ref}v_\mathrm{ref}$$

with $a_\mathrm{ref} = \mu_\mathrm{ref} - p_\mathrm{ref}v_\mathrm{ref}$, $p_\mathrm{ref} = \rho_\mathrm{s}k_\mathrm{B}T$, and $p_\mathrm{ref}v_\mathrm{ref} = k_\mathrm{B}T$; this provides

$$a' = a - a_\mathrm{ref} + \rho_\mathrm{s}k_\mathrm{B}Tv - k_\mathrm{B}T$$

where ρ_s is the solvent molecular density. In the QGE reference condition, the partial molecular internal energy is independent of the system volume, that is, $u' = u - u_\mathrm{ref}$, where, within our approximations, u_ref is a constant identical for all of the conformational states. Using the general expressions of the Gamma state excess thermodynamic properties,[33] we obtain along an isochore for the partial molecular properties

$$\Delta a_i(T) = \Delta u_{0,i}' - \Delta c_{V0,i}' T_0 \Lambda(T) - k_\mathrm{B}T \ln \gamma_i - \rho_\mathrm{s}k_\mathrm{B}T\Delta v_i \quad (5)$$

$$\Delta s_i(T) = \frac{T_0 \Delta c_{V0,i}'}{T}\left[\frac{T - T_0}{T(1-\delta_0) + T_0\delta_0} + \Lambda(T)\right] + k_\mathrm{B} \ln \gamma_i + \rho_\mathrm{s}k_\mathrm{B}\Delta v_i \quad (6)$$

$$\Delta u_i(T) = \Delta u_i'(T) = \Delta u_{0,i}' + \frac{(T - T_0)\Delta c_{V0,i}' T_0}{T(1-\delta_0) + T_0\delta_0} \quad (7)$$

$$\Delta c_{V,i}(T) = \Delta c_{V,i}' = \Delta c_{V0,i}' \left(\frac{T_0}{T(1-\delta_0) + T_0\delta_0}\right)^2 \quad (8)$$

$$\Delta u_{V,i}(T) = \left(\frac{\partial(\beta\Delta\mu_i)}{\partial\beta}\right)_V = U_i(T) - U_r(T) \quad (9)$$

with $\beta = (k_B T)^{-1}$, U_i and U_r being the internal energies of the whole system with the solute in the ith and reference conformations, respectively, s the partial molecular entropy, and use having been made of $\Delta a_{i,\mathrm{ref}} = \Delta\mu_{i,\mathrm{ref}}$. Note that the entropy change (eq 6) is expressed by three terms. The first one, proportional to $\Delta c_{V0,i}'$, reflects the changes of the potential energy fluctuations due to the conformational change, the second one ($k_B \ln \gamma_i$) provides the effect of the different accessible configurational volumes for the rototranslationally fixed solute, and the third one ($\rho_s k_B \Delta v_i$) is due to the change of the available three-dimensional volume resulting from different solute partial molecular volumes.

The properties $\Delta u_{0,i}'$, $\Delta c_{V0,i}'$, Δv_i, and $\ln \gamma_i$ are regarded as parameters and can be estimated if simulation data are available for at least four different thermodynamic conditions (e.g., temperatures). The detailed fitting procedure described in previous papers[32,33] is performed in two stages. First, $\Delta u_0'$, $\Delta c_{V0}'$, and Δv may be obtained by fitting the mean potential energy of the whole simulation box with the corresponding Gamma state expression, for each solute conformation. Second, by fitting the chemical potential change as obtained by MD simulations via $\Delta\mu_i = A_i - A_r = -k_B T \ln(P_i/P_r)$ with eq 4, γ may be evaluated for each conformation.

The previous equations readily provide the solute thermodynamics considering all of its conformational states (Supporting Information). Of particular interest here is to distinguish between the species with ends not in contact, which is experimentally characterized as "unfolded" (U), and the species with ends in contact, or "folded" (F). The previous equations may then be formulated for each subset to provide the thermodynamics of the end-to-end contact formation process. In particular, the probability of the unfolded condition, P_U, and the unfolding free energy, $\Delta\mu_U$, are given by

$$P_U = \frac{\sum_{i\in U} \exp(-\beta\Delta\mu_i)}{\sum_j \exp(-\beta\Delta\mu_j)} \quad (10)$$

$$\Delta\mu_U = -k_B T \ln \frac{P_U}{1-P_U} = -k_B T \ln \frac{\sum_{i\in U} \exp(-\beta\Delta\mu_i)}{\sum_{j\notin U} \exp(-\beta\Delta\mu_j)} \quad (11)$$

It is worthwhile noting that this QGE model, describing the peptide thermodynamics along an isochore, might be also used to reconstruct the peptide thermodynamics along an isobar by using QGE models for different solution isochores.

3. Methods Section

A detailed description of the simulation methods is provided in the Supporting Information. A set of MD simulations of the MR121-GSGSW peptide in water over a wide temperature range (293−600 K) was performed using the GROMOS96 force field in explicit solvent using periodic boundary conditions. The C-terminal end of the peptide was modeled as COO^- to reproduce a pH of about 7 as in the experimental conditions.[17] Simulation lengths at the different temperatures were the following: 1040 ns at 293 K, 250 ns at 350 K, 300 ns at 400 K, 200 ns at 450 K, 150 ns at 500 K, and 140 ns at 600 K.

A principal component analysis was applied to the 293 K simulation data, and the three first principal components were selected to define the essential conformational subspace used for the visualization in Figure 1 and to partition the state space into conformational states (see below).

In the version of QGE used here, it is assumed that the potential energy fluctuations of the system within a given conformation are well described by a Gamma distribution. This assumption is valid if the conformational state is "simple" in the sense that it consists of a single free-energy basin. Conformational states with internal energy barriers may exhibit a more complicated potential energy distribution that may be better modeled by multiple Gamma states. For this reason, the system states (i.e., conformations) were defined in a way partitioning the principal subspace at the free-energy barriers, thus minimizing state-internal energy barriers. This was done here using the improved Perron cluster cluster analysis (PCCA),[36,37] thus providing seven long-lived conformational states which were used to parametrize a QGE model with seven Gamma states. In practice, it is likely that many different ways of partitioning provide a valid multi-Gamma-state behavior. Note that the essential subspace dimension used to define the conformational states, that is, the full dimension configurational ensembles whose projections in the essential subspace belong to the conformational regions considered, should guarantee that each conformation (free-energy basin) corresponds to a rigid or semirigid solute condition if a Gamma state basis set is to be used. When an insufficient number of essential subspace dimensions is considered (in the present case, 1 or 2), significantly different structures may be projected on top of each other, possibly leading to an unsuitable definition of states. Ideally, for a multi-Gamma-state QGE model, many conformational states defined in a high-dimensional subspace should be used so as to ensure that within each state, the energy fluctuations can be very precisely described by Gamma distributions, that is, rigid or semirigid solute condition. The more states used, however, the less statistics for each state that is available to parametrize the model. The seven conformational states defined within the three-dimensional essential subspace used here are then a compromise between accuracy and statistics, that is, the simplest accurate multi-Gamma-state model for the peptide studied, providing a good resolution for the relevant conformational transitions. A comparison of the quality of the QGE model resulting from varying the number of states and simulation lengths is provided in the Supporting Information.

After the clustering, each structure along the trajectory is assigned to one conformational state. In Figure 1, the projection of the $T = 293$ K trajectory on the principal subspace is colored according to the assignment to each of the seven conformations. Representative structures are also given. Five conformations mainly contain structures in which the MR121 dye and the Trp are stacked (**i**, **iii**, **iv**, **vi**, **vii**). In the two remaining conformations, (**ii**, **v**), the end groups are not stacked; **v** is still relatively compact due to contacts within the peptide and between the peptide and the end groups, while **ii** is a completely extended or "unfolded" conformation.

Using the above definition of the seven conformational states, a QGE model was built as described in the theory section. State **vi** was chosen as the reference state, and thus, all QGE

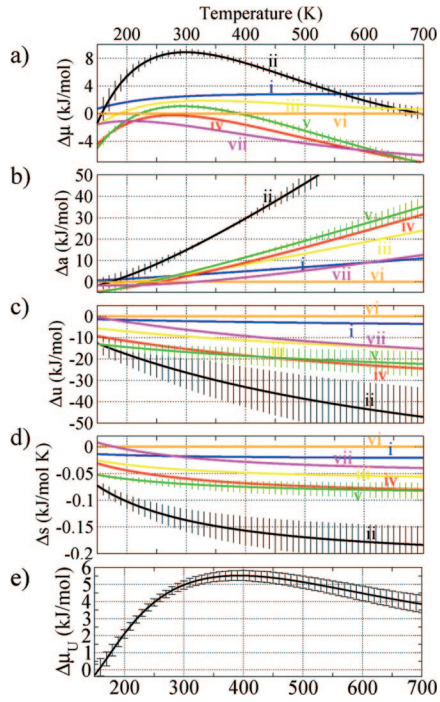

Figure 2. Predictions of the QGE model in the 150−700 K temperature range. The coloring scheme is the same as that in Figure 1. Panel e shows the unfolding free energy, that is, change of the free energy of the open species (states **ii**, **v**) with respect to the free energy of the closed species (states **i**, **iii**, **iv**, **vi**, **vii**). The vertical bars indicate the standard deviation of the properties resulting from the uncertainties of the MD input data used to obtain the QGE model. For visual clarity, standard deviations are only shown for states **ii** and **v**, characterized by the largest noises.

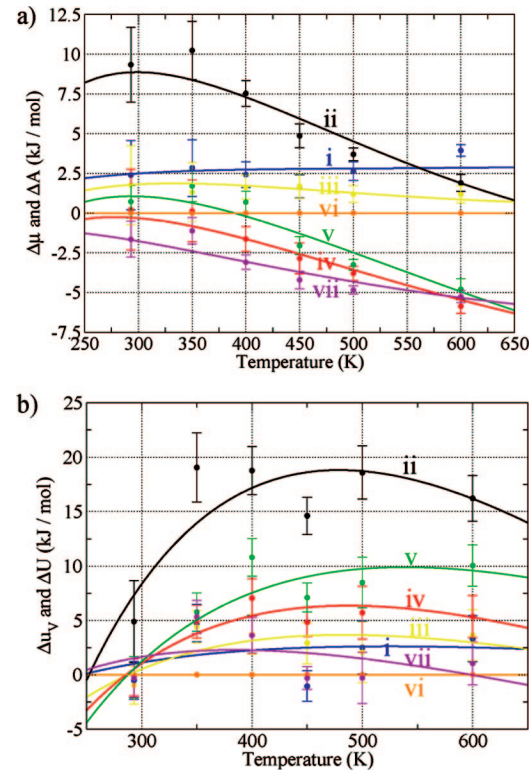

Figure 3. Validation of the QGE model by comparing the predictions of the model (solid lines) with the MD simulations (bullets with error bars). (a) The free-energy changes of the full system with respect to the blue reference state, ΔA, as obtained by MD simulations compared to the corresponding QGE prediction, $\Delta\mu$. (b) The mean potential energy changes with respect to the blue reference state, ΔU, as obtained by MD simulations compared to the corresponding QGE prediction, Δu_V. The coloring scheme is the same as that in Figure 1, and the error bars correspond to one standard error of the simulation data.

parameters and all properties computed from the QGE model are given in terms of differences with respect to this state. The use of a different conformation as the arbitrary reference state would only shift the conformational properties without changing their relative values. To validate the reliability of the QGE model, the uncertainties of the QGE parameters and model predictions resulting from finite sampling of the MD simulations were estimated as described in the Supporting Information.

4. Results

In order to assess the accuracy of the QGE model, $\Delta\mu$ and Δu_V as provided by the QGE model are compared with the corresponding values directly obtained by the simulations at different temperatures as indicated by eqs 1 and 9. The QGE model predictions and the simulation data differ by no more than 1−2 standard errors of the simulation data (see Figure 3), showing that the model is consistent (note that in all of the figures and tables, the partial molecular properties are expressed for 1 mol of solute, thus being equivalent to partial molar properties). It must be pointed out that Δu_V was not directly used to parametrize the model,[33] and hence, the accurate reproduction of this property shows that the model is consistent with the data and thus predictive. As a second test, the QGE model is used to calculate the relative fluorescence quantum yield, which corresponds to the equilibrium fraction of fluorescent conformations if the fluorescent emission can be considered to be much faster than conformational changes. Assuming, as usual, that radiative relaxation only occurs when the dye and Trp are far apart (in our model conformations, **ii** and **v**), this quantum yield is given by eq 10. At 293 K, the QGE model predicts a relative quantum yield value of 13%, reproducing exactly the experimental value (personal communication, M. Sauer, University of Bielefeld, Germany).

Figure 2 shows the temperature dependence of the QGE predictions for the partial molecular Gibbs ($\Delta\mu$) and Helmholtz (Δa) free energies, the internal energy (Δu), and the entropy (Δs) for each of the seven conformational states (note that these are changes with respect to the reference conformational state **vi**). The comparison of these curves with the corresponding model parameters (Table 1) and with the main structural features of the corresponding states (Figure 1) is instructive.

Conformations **i** and **vi** are almost mirror images with respect to the relative positions of the MR121 dye and the Trp side chain. In both conformations, the dye and Trp point in the same direction, but the stacking order of these end groups is inverted. Conformation **i** can thus be transformed into **vi** (and vice versa) by moving the Trp to the other side of the dye. Indeed, the QGE parameters and thermodynamic properties of these two conformations are similar (see Table 1 and Figure 2). State **i** has a slightly lower $\Delta u_0'$, resulting in a slightly lower partial molecular energy ($\Delta u_i(T)$; see Figure 2). At the same time, its partial molecular accessible configurational volume (expressed via ln γ) is slightly lower, resulting in a decreased partial molecular entropy $\Delta s_i(T)$. The partial molecular Helmholtz, $\Delta a(T)$, and Gibbs, $\Delta\mu(T)$, free energies indicate that **vi** is thermodynamically more stable as a result of its higher entropy overcompensating the unfavorable energy. Similarly to the pair **i** and **vi**, conformations **iii** and **vii** are also nearly mirror images with an inverse stacking order. Here, however, in both conformations, the dye and Trp point in opposite directions. Again,

TABLE 1: QGE Parameters for Each of the Seven Conformational States[a]

state	$\Delta u_0'$ in kJ/mol	$\Delta c_{V0}'$ in kJ/ (mol/K)	Δv in L/mol	ln γ
i	−2.23 &±; 5.22	−0.0046 &±; 0.019	−0.0088 &±; 0.013	−2.17 &±; 3.88
ii	−26.10 &±; 7.48	−0.075 &±; 0.024	−0.090 &±; 0.018	−20.1 &±; 5.11
iii	−9.23 &±; 4.08	−0.021 &±; 0.015	−0.026 &±; 0.010	−6.07 &±; 3.08
iv	−15.29 &±; 2.54	−0.034 &±; 0.010	−0.043 &±; 0.006	−8.62 &±; 1.94
v	−16.66 &±; 2.94	−0.020 &±; 0.11	−0.047 &±; 0.007	−8.07 &±; 2.25
vi	0	0	0	0
vii	−6.27 &±; 3.04	−0.033 &±; 0.012	−0.021 &±; 0.007	−4.95 &±; 2.35

[a] The errors indicated correspond to a standard deviation of the parameters resulting from the uncertainties in the input MD data used to obtain the QGE model.

the QGE parameters and thermodynamics of these two conformations are rather similar, with state **iii** being the one with lower $\Delta u_0'$/lower energy ($\Delta u_{\mathbf{iii}}(T) \leq \Delta u_{\mathbf{vii}}(T)$) and lower accessible configurational volume ln γ/lower entropy. The partial molecular Gibbs and Helmholtz free energies are again entropically dominated, leading to **vii** being thermodynamically more stable than **iii**.

When grouping together the states with no end-to-end contact, **ii** and **v**, to define the "open" set, the thermodynamics of opening is obtained as a function of temperature from eq 11. In Figure 2e, the opening free energy ($\Delta\mu_U$) is shown and exhibits an inverted U-shape typically observed in proteins.[9,38] The figure indicates that the closed states are predominant over almost the whole temperature range considered. They are most stable in the range of 300−500 K, although the magnitude of $\Delta\mu_U$ being about $2k_BT$ indicates that the open states are still accessible. Only for very low temperatures (below 155 K) are the open states stabilized over the closed ones. This trend may be compared to "cold denaturation", which may occur in proteins at temperatures close to the liquid−solid transition.[39−41]

The extended open state **ii** is especially interesting. This state has the lowest partial molecular internal energy and entropy for $T > 150$ K. This cannot be explained in terms of solute contributions (intrapeptide interactions and chain flexibility), which have the opposite effect. A direct computation of the solute internal mean potential energy for each conformation from MD simulations shows that this energy term increases upon extension by more than 100 kJ/mol. Furthermore, the volume of the solute conformational region corresponding to state **ii** is similar to the volumes of the other conformational regions (see Figure 1), indicating that the state **ii** negative entropy change is not due to a reduction of the solute internal configurational volume. Moreover, state **ii** has a strongly negative partial molecular volume change, $\Delta v_{\mathbf{ii}}$, which is manifested as opposing trends of the Gibbs and Helmholtz free energies with temperature (see Figure 2). The collapsed open state **v** exhibits similar thermodynamic behavior, having the second-lowest partial molecular energy, entropy, and volume. This correlation of partial molecular energy, entropy, and volume can be understood further in terms of the QGE parameters Δv, $\Delta u_0'$, ln γ, and $\Delta c_{V0}'$, which are highly positively correlated (Figure 4). Decrease of the partial molecular volume (negative Δv) is associated with an improvement of the energetic interactions (negative $\Delta u_0'$). The changes in partial molecular volume are also correlated with the changes in partial molecular heat capacity ($\Delta c_{V0}'$) and accessible configurational space (ln γ), which both determine the partial molecular entropy as given in eq 6. Here, the change of accessible configurational space is the dominant component, for example, for the extended open state **ii** at 293 K; the contribution of this term is five times larger than that of the remaining terms.

The negative values of $\Delta v_{\mathbf{ii}}$ and $\Delta v_{\mathbf{v}}$ indicate a volume reduction of the complete solute−solvent system at constant pressure upon extension. To determine whether this is due to solute or solvent volume changes, the mean solute and solvent volume changes were computed for each structure of the molecular dynamics simulation at $T = 293$ K. To avoid conclusions based on artifacts produced by the volume computation algorithm, two approaches were compared, (i) a Richards-type volume (see Supporting Information for details on the algorithm) which partitions three-dimensional space according to van der Waals radii of the atoms and (ii) the solvent-excluded volume using a probe sphere of 0.14 nm.[42] Table 2 reports the mean volume changes with respect to state **vi**, showing that the solvent volume changes are the dominant component of Δv. While the solute volume increases upon extension, this is overcompensated by the strong reduction of solvent volume, independent of the volume computation algo-

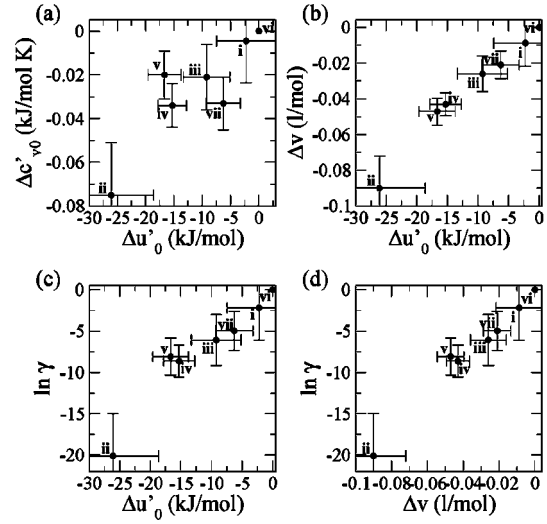

Figure 4. Correlations among the QGE parameters for the seven conformations. The error bars indicate the standard deviation of the parameters resulting from the uncertainties in the MD input data used to obtain the QGE model.

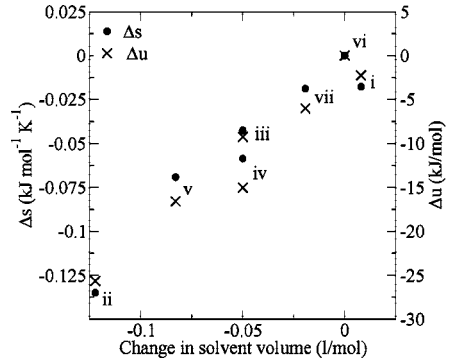

Figure 5. Correlations between changes in solvent volume, partial molecular energy, and entropy at $T = 293$ K.

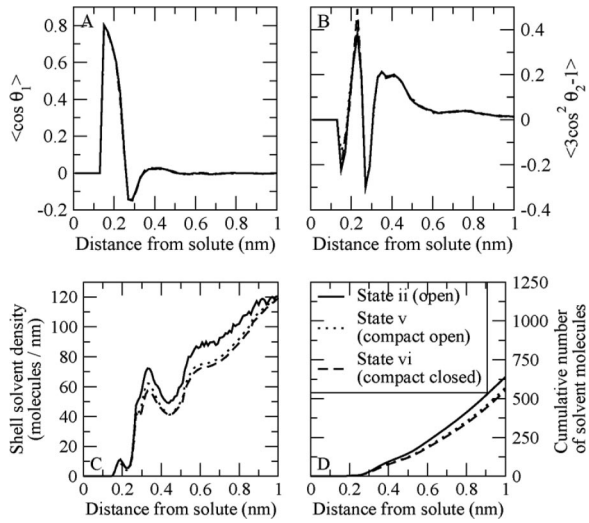

Figure 6. (A) $\langle \cos \theta_1 \rangle$, with θ_1 being the angle between the vectors (i) from closest solute atom to the solvent O and (ii) from solvent O to the midpoint of solvent H's. (B) $\langle 3 \cos^2 \theta_2 - 1 \rangle$, with θ_2 being the angle between the vectors (i) from closest solute atom to the solvent O and (ii) the normal of the solvent plane. (C) Number of solvent molecules per shell thickness at a given distance from solute. (D) Cumulative number of solvent molecules after a given distance from the solute.

TABLE 2: Changes of Solute Partial Molecular Volume and Mean Solute and Solvent Volume of the Seven Conformations As Obtained by the 293 K MD Simulation (all volumes measured in L/mol)

		Richards volume		solvent-excluded volume	
state	Δv	ΔV_{solute}	$\Delta V_{solvent}$	ΔV_{solute}	$\Delta V_{solvent}$
i	−0.0088	−0.0081	−0.0007	0.0028	−0.0116
ii	−0.090	0.0385	−0.1285	0.0272	−0.1172
iii	−0.026	0.0045	−0.0305	0.0032	−0.0292
iv	−0.043	0.0055	−0.0485	0.0014	−0.0444
v	−0.047	0.0199	−0.0669	0.0034	−0.0536
vi	0	0	0	0	0
vii	−0.021	0	−0.021	−0.0008	−0.0218

rithm. Therefore, both the partial molecular energy and entropy are highly correlated with the changes in solvent volume (Figure 5). In the extended open state **ii**, the solvent volume is reduced by 0.13 L/mol per solute molecule, corresponding to the volume of about seven bulk water molecules. Thus, the entropy reduction in the open states **ii** and **v** can be first explained by a reduction of the accessible configurational volume due to the reduction of solvent volume.

Second, the accessible configurational volume may also be reduced by restricting the orientational degrees of freedom of the solvation waters. The water molecules in the first solvation shells exhibit clear orientational preferences with respect to the bulk in all conformational states (see Figure 6A and B). Here, the per-water distribution of orientation angles near positively charged, negatively charged, and hydrophobic surface patches was not found to change significantly between different solute conformations. This is demonstrated by the water orientation correlation functions shown in Figure 6A and B, which are almost identical in different conformations. However, the absolute number of water molecules in the first solvation shells is larger in the extended open state (see Figure 6C and D). This is due to both the increase of solvent density and the sheer increase of spatial extent in the extended state (64 water molecules within 0.35 nm from the solute in state **ii** versus 50−54 water molecules in the other states). Thus, state **ii** has a larger number of oriented water molecules, providing an additional entropy reduction.

What is the structural origin of the configurational volume reduction accompanied by entropy decrease? Such an effect may occur when exposing hydrophobic surfaces[13,43] and when solvating charges (electrostriction).[1,32,44,45] Both the hydrophobic effect and electrostriction can reduce the configurational volume by inducing order in the solvent structure.[1,11] The present peptide has both hydrophobic (the Trp side chain and parts of the MR121) and hydrophilic (the COO^- terminus, the peptide backbone, and parts of the positively charged MR121) groups, each covering about 50% of the solvent-accessible surface. Both hydrophobic and hydrophilic surfaces increase by 10−20% upon extension.

Interestingly, both the hydrophobic effect and electrostriction contribute to the collapse and closure process in the present peptide (see Figure 7). In fact, the peptide collapse from the extended open state **ii** to the collapsed open state **v** is characterized by decreasing solvent density and number of oriented solvent molecules, both contributing to the increase of the peptide partial molecular entropy, as a result of a drastic reduction of the peptide mean dipole moment (from 76 to 43 Debye). The slight reduction of the mean hydrophobic surface area upon the **ii** → **v** transition clearly implies no significant hydrophobic effect in the conformational change. Thus, the **ii** → **v** transition is mainly a "polar" collapse, driven by electrostriction. The closure of state **v** into the collapsed closed states further increases the peptide partial molecular entropy, as a consequence of the reduced hydrophobic surface area (the peptide mean dipole remains constant). Therefore, direct collapse and closure **ii** → {**i**,**iii**,**iv**,**vi**,**vii**} is thus driven by a combination of the hydrophobic effect and electrostriction.

5. Conclusions

In the present work, the QGE theory is extended to provide an analytical model for the conformational thermodynamics of complex solvated molecules. This theory is applied to quantitatively describe the complete equilibrium statistical mechanics of the conformational states of the fluorescent peptide MR121-GSGSW in solution. The QGE model was parametrized by using MD simulations at several temperatures along the typical liquid−water isochore.

If a suitable definition of conformational states is used, that is, each state corresponds to a single free-energy basin in the conformational space considered, equivalent to a rigid or semirigid solute condition, a single Gamma state QGE model describes accurately the partial molecular thermodynamics of each conformational state over a large temperature range. The model shown here used seven conformational states and was shown to be reliable as it is both consistent with the MD simulations and also precisely predicts the experimentally measured fluorescence relative quantum yield. This model provides a comprehensive picture of the coupling between structures and thermodynamics in the conformations accessible to the peptide.

For the system studied here, the extended or "unfolded" state is characterized by the lowest partial molecular energy and entropy. This, at first sight, counterintuitive result cannot be explained in terms of solute intramolecular energy and flexibility, which are larger in the extended state. Rather, peptide collapse and closure is driven by solvent effects. The low energies and entropies for the extended and open states are due to solvent volume reduction and an increase of the number of

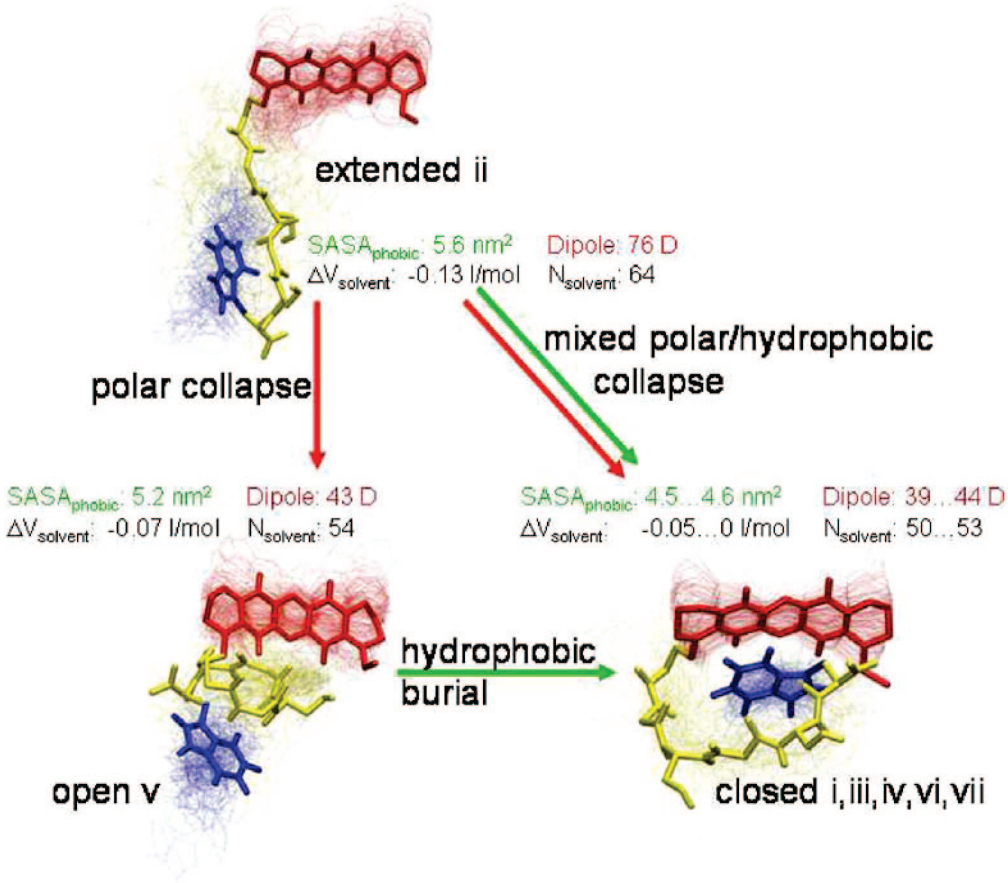

Figure 7. Hydrophobic effect versus electrostriction in the collapse and closure of the peptide. The polar collapse **ii** → **v**, driven by electrostriction, is determined by the peptide large reduction of dipole moment. The closure **v** → closed states is driven by the hydrophobic effect. The direct folding **ii** → closed states involves a mixture of polar and hydrophobic collapse. The standard deviations of the SASA are below 0.13, and the standard errors are below 10^{-4} for all states.

oriented solvent molecules around the solute. It is found that the entropy reduction is driven not only by the hydrophobic effect but also by a significant, and perhaps dominating, ordering and density increase in the first solvation shell upon exposure of charges, that is, electrostriction. Thus, the compaction of the extended open conformation **ii** into the open conformation **v** is not a hydrophobic but rather is a "polar" collapse. The present results thus illustrate the essential role of the solvent in determining peptide conformational thermodynamics. While the results confirm the usefulness of explicit solvent models in peptide MD simulations, they also indicate that a way of improving implicit solvent models might be to explicitly incorporate terms that consider the entropy change upon charge solvation.

To the knowledge of the authors, this is the first time that the thermodynamic relevance of electrostriction has been explicitly shown for a complex mixed hydrophobic/hydrophilic system. A negative heat capacity change upon extension is also observed, which is indeed regarded as a signature for polar solvation[9] and was also found for the unfolding of other polypeptide systems.[14,15] This contrasts with the typical results for protein unfolding, where a positive heat capacity change is observed[9] and where the hydrophobic effect is regarded as the dominant contribution to the initial collapse. Most peptides and proteins possess a more homogeneous charge distribution than the present system, and one may thus assume that the effect of electrostriction is reduced in such systems. Nevertheless, given the present results and the fact that proteins do have many polar and charged groups, electrostriction may turn out to be an underrated contribution to protein folding thermodynamics.

Acknowledgment. This work was carried out under the HPC-EUROPA project (RII3-CT-2003-506079), with the support of the European Community-Research Infrastructure Action under FP6 "Structuring the European Research Area" Programme.

Supporting Information Available: (i) A theory to compute the partial molecular thermodynamics of the conformationally fluctuating peptide, (ii) a detailed description of the simulation setup and the methods used for discretizing conformation space, (iii) details on the algorithm to calculate atomic volumes, and (iv) tests of the QGE fitting procedure with other definitions of conformational states. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Dill, K. A. *Biochemistry* **1990**, *29*, 7133–7155.
(2) Baldwin, R. L. *J. Mol. Biol.* **2007**, *371*, 283–301.
(3) Santiveri, C. M.; Santoro, J; Rico, M.; Jimenez, M. A. *J. Am. Chem. Soc.* **2002**, *124*, 14903–14909.
(4) Maynard, A. J.; Sharman, G. J.; Searle, M. S. *J. Am. Chem. Soc.* **1998**, *120*, 1996–2007.
(5) Onuchic, J. N.; Wolynes, P.; Luthey-Schulten, Z.; Socci, N. D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3626–3630.
(6) Shea, J.-E.; Brooks, C. L., III. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499–535.
(7) Sadqi, M.; Lapidus, L. J.; Munoz, V. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12117–12122.

(8) Raschke, T. M.; Tsai, J.; Levitt, M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5965–5969.
(9) Prabhu, N. V.; Sharp, K. A. *Annu. Rev. Phys. Chem.* **2005**, *56*, 521–548.
(10) Makhatadze, G. I.; Privalov, P. L. *Protein Sci.* **1996**, *5*, 507–510.
(11) Lum, K.; Chandler, D.; Weeks, J. D. *J. Phys. Chem. B* **1999**, *103*, 4570–4577.
(12) Chothia, C. *Nature* **1974**, *248*, 338–339.
(13) Chandler, D. *Nature* **2005**, *29*, 640–647.
(14) Richardson, J. M.; Makhatadze, G. *J. Mol. Biol.* **2004**, *335*, 1029–1037.
(15) García, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782–2787.
(16) Krieger, F.; Fierz, B.; Bieri, O.; Drewello, M.; Kiefhaber, T. *J. Mol. Biol.* **2003**, *332*, 265–274.
(17) Neuweiler, H.; Löllmann, M.; Doose, S.; Sauer, M. *J. Mol. Biol.* **2007**, *365*, 856–869.
(18) Schäfer, H.; Daura, X.; Mark, A. E.; van Gunsteren, W. F. *Proteins* **2001**, *43*, 45–56.
(19) Daidone, I.; Ulmschneider, M.; di Nola, A.; Amadei, A.; Smith, J. C. *Proc. Nat. Acad. Sci. U.S.A.* **2007**, *104*, 15230–15235.
(20) Daidone, I.; D'Abramo, M.; Di Nola, A.; Amadei, A. *J. Am. Chem. Soc.* **2005**, *127*, 14825–14832.
(21) Amadei, A.; Apol, M. E. F.; Nola, A. D.; Berendsen, H. J. C. *J. Chem. Phys.* **1996**, *104*, 1560–1574.
(22) Amadei, A.; Apol, M. E. F.; Berendsen, H. J. C. *J. Chem. Phys.* **1997**, *106*, 1893–1912.
(23) Amadei, A.; Apol, M. E. F.; Berendsen, H. J. C. *J. Chem. Phys.* **1998**, *109*, 3004–3016.
(24) Amadei, A.; Apol, M. E. F.; Berendsen, H. J. C. *J. Chem. Phys.* **1998**, *109*, 3016–3027.
(25) Landau, L. D.; Lifshitz, E. M. *Course of Theoretical Physics, Part 1 (Statistical Physics)*, 3rd ed.; (Pergamon Press: Oxford, U.K., 1980; Vol. 5, 3rdedition.
(26) Apol, M. E. F.; Amadei, A.; Berendsen, H. J. C. *J. Chem. Phys.* **1996**, *104*, 6665–6678.
(27) Apol, M. E. F.; Amadei, A.; Berendsen, H. J. C. *Chem. Phys. Lett.* **1996**, *256*, 172–178.
(28) Amadei, A.; Roccatano, D.; Apol, M. E. F.; Berendsen, H. J. C.; Di Nola, A. *J. Chem. Phys.* **1996**, *105*, 7022–7025.
(29) Roccatano, D.; Amadei, A.; Apol, M. E. F.; Di Nola, A.; Berendsen, H. J. C. *J. Chem. Phys.* **1998**, *109*, 6358–6363.
(30) Amadei, A.; Apol, M. E. F.; Chillemi, G.; Berendsen, H. J. C.; Di Nola, A. *Mol. Phys.* **1999**, *96*, 1469–1490.
(31) Amadei, A.; Iacono, B.; Grego, S.; Chillemi, G.; Apol, M. E. F.; Paci, E.; Delfini, M.; Di Nola, A. *J. Phys. Chem. B* **2001**, *105*, 183–1844.
(32) D'Alessandro, M.; D'Abramo, M.; Brancato, G.; Di Nola, A.; Amadei, A. *J. Phys. Chem. B* **2002**, *106*, 11843–11848.
(33) D'Abramo, M.; D'Alessandro, M.; Amadei, A. *J. Chem. Phys.* **2004**, *120*, 5226–5234.
(34) D'Alessandro, M.; D'Abramo, M.; Paci, M.; Amadei, A. *Phys. Scr.* **2005**, *118*, 196.
(35) Amadei, A.; Apol, M. E. F.; Berendsen, H. J. C. *J. Chem. Phys.* **1997**, *106*, 1893–1912.
(36) Deuflhard, P.; Weber, M. *ZIB Report* **2003**, *03−09*.
(37) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. *J. Chem. Phys.* **2007**, *126*, 155102.
(38) Roccatano, D.; Nola, A. D.; Amadei, A. *J. Phys. Chem. B* **2004**, *108*, 5756–5762.
(39) Privalov, P. L.; Gill, S. J. *Adv. Protein Chem.* **1988**, *39*, 191–234.
(40) Privalov, P. L. *Crit. Rev. Biochem. Mol. Biol.* **1990**, *25*, 281–305.
(41) Robertson, A. D.; Murphy, K. P. *Chem. Rev.* **1997**, *97*, 1251–1267.
(42) Eisenhaber, F.; Lijnzaad, P.; Argos, P.; Sander, C.; Scharf, M. *J. Comput. Chem.* **1995**, *16*, 273–284.
(43) Imai, T.; Hirata, F. *J. Chem. Phys.* **2005**, *122*, 094509.
(44) Murphy, K. P.; Gill, S. J. *Thermochim. Acta* **1990**, *172*, 11–20.
(45) Murphy, K. P.; Gill, S. J. *J. Mol. Biol.* **1991**, *222*, 699–709.

JP801391T