Density functional theory study of the conformational space of an infinitely long polypeptide chain

Joel Ireta[1,2,a] and Matthias Scheffler[1]
[1]*Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin-Dahlem, Germany*
[2]*Departamento de Química, División de Ciencias Básicas e Ingeniería, Universidad Autónoma Metropolitana-Iztapalapa, A.P. 55-534, México, Distrito Federal 09340, Mexico*

(Received 12 May 2009; accepted 30 July 2009; published online 28 August 2009)

The backbone conformational space of infinitely long polyalanine is investigated with density-functional theory and mapping the potential energy surface in terms of (L, θ) cylindrical coordinates. A comparison of the obtained (L, θ) Ramachandran-like plot with results from an extended set of protein structures shows excellent conformity, with the exception of the polyproline II region. It is demonstrated the usefulness of infinitely long polypeptide models for investigating the influence of hydrogen bonding and its cooperative effect on the backbone conformations. The results imply that hydrogen bonding together with long-range electrostatics is the main actuator for most of the structures assumed by protein residues. © *2009 American Institute of Physics*.
[DOI: 10.1063/1.3207815]

I. INTRODUCTION

Geometric patterns such as helices, sheets, and turns are commonly observed in protein crystals. These conformational motifs, known as the secondary structure of proteins,[1] are representative of the most populated regions of the backbone conformational space of proteins.[2–4] Such conformational space is typically mapped using the torsional angles ϕ and ψ which give the relative position of a residue along the protein backbone [Fig. 1(a)]. Current studies based on electronic structure methods for investigating the conformational space of a residue in proteins, are through the calculation of Ramachandran plots for dipeptides. The latter refers to the calculation of the potential energy surface (PES) for such dipeptides in terms of the (ϕ, ψ) coordinates. However the PES calculated with these *ab initio* electronic structure methods only roughly correlate with the distribution of (ϕ, ψ) values from protein crystals.[5–9] Such observation was noted since the first Hartree–Fock calculations of the PES of dipeptides[5] and holds true even if the map is calculated with second-order Møller–Plesset perturbation theory (MP2) or density-functional theory (DFT).[7–9] The inclusion of solvent and temperature effects improve the agreement between the predicted (by the calculated PES) and the observed backbone conformational space, nevertheless inconsistencies are still notable.[8] It is well accepted that these discrepancies are connected to the inherent inability of dipeptides for describing long-range electrostatics and hydrogen bonding (hb) networks which are crucial for stabilizing the secondary structure of proteins. Particularly hb networks are expected to largely affect the conformational space of a residue due to the hb directionality and hb cooperativity. The former refers to the preference of hb's to form angles close to 180° between the hydrogen-donor group and hydrogen-acceptor group.[10] The later refers to the strengthening of each individual hb in hydrogen bonded systems due to the alignment of hb's. Long peptides are needed to study the energetically allowed conformational regions for a residue in the presence of a cooperative network of hb's; e.g., it has been suggested that a peptide must consists of at least ten residues to be stabilized by a fully cooperative network of hb's against transformation into an extended structure.[11]

In the present work, the influence of peptide hb's and long-range electrostatic interactions on the conformational space of a residue is investigated using infinite polypeptides. The PES of an infinitely long chain of polyalanine is calculated using DFT together with the generalized gradient approximation of Perdew, Burke, and Ernzerhof (PBE) for the exchange-correlation functional.[12] DFT-PBE describes hb's with an error that may increase to as much as 1.5 kcal/mol with respect to MP2 results,[10] being the largest errors for cases in which hb's are highly bent. Although DFT-PBE does not describe van der Waals forces properly, such approximations are still adequate for the purposes of the present work as we are mainly interested here on the influence of hb on the backbone conformational space. Moreover in a previous work we have shown that DFT-PBE describes hb cooperativity within 13% error with respect to MP2 results.[11] It has been also shown that the geometry predicted for the α-helix using infinitely long chains and DFT-PBE is in agreement with experimental values.[11,13–15] These results indicate that the methodology used here captures the effect of the hb cooperative effect on the peptide backbone conformation. The use of an infinite polypeptide enables us to focus on the conformational space of a single residue fully including the hb cooperativity and long-range electrostatic interactions. We found convenient to map the PES of the infinite chain using cylindrical coordinates instead of the standard (ϕ, ψ) coordinates. Hence the PES is mapped in terms of the rotation θ and the translation L along the chain axis z [Fig. 1(b)]. It has been shown that such procedure adequately describes the

[a]Electronic mail: iret@xanum.uam.mx.

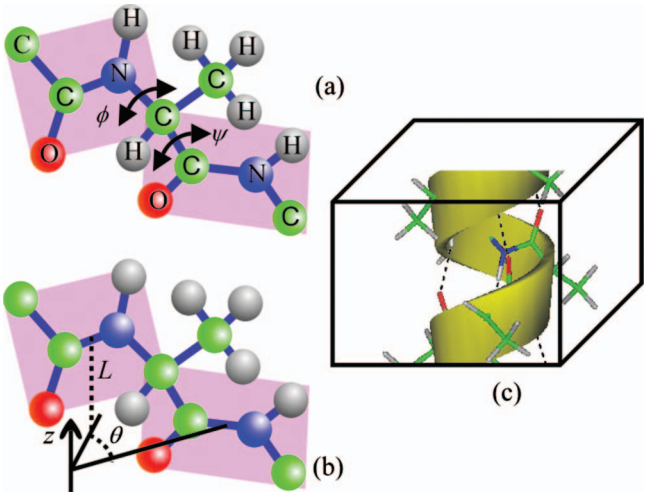

FIG. 1. (a) Torsional angles typically used for describing backbone conformations. (b) Alternative description of the backbone conformation in terms of a rotation θ and a translation L along the local chain axis z. (c) Polyalanine inside the employed supercell. Dotted lines denote hydrogen bonds.

PES of right-handed helical conformations.[15] We show here that the PES thus calculated gives a satisfactory two-dimensional representation of the protein conformational space. The predicted conformational space is compared to experimentally observed backbone conformations in an extended set of protein structures resolved by x-ray crystallography and obtained from the protein data bank.[16] The agreement between the predicted and observed conformational space is excellent except for the so-called polyproline II (PPII) conformational region. Our calculations of the PES are carried out in vacuum and at 0 K. Obviously factors like van der Waals interactions, peptide-water interactions and temperature contribute to determine backbone conformations in protein crystals. Nevertheless, we find that the values for (L, θ) observed in the analyzed protein structures cluster along the pathways of the investigated PES. This result indicates that hb's and long-range electrostatic interactions outweigh the other (here neglected) factors except for the ~10% of residues in the PPII conformation. These interactions are thus the main actuators for most of the structures assumed by protein residues. We will discuss the possible connection of these neglected factors with the discrepancies between the predicted and the observed conformational space.

II. METHOD

As we describe the infinitely long chain using the supercell approach [Fig. 1(c)], where periodic boundary conditions are imposed, then $L=c/N$ and $\theta=360°\,\mathrm{m/N}$. c is the length of the lattice side parallel to the chain axis, N the number of residues per supercell and m is the number of chain turns per supercell. The PES of the infinite chain is calculated as $\Delta E(L,\theta)=N[\mu^{\infty}(L,\theta)-\mu^{\infty}_{\mathrm{FES}}]$, with $N=1$ and μ^{∞} (L, θ) is defined as $\mu^{\infty}(L,\theta)=E^{\infty}(L,\theta)/N$, where E^{∞} (L, θ) is the total energy of a chain consisting of N residues per supercell. $\mu^{\infty}_{\mathrm{FES}}$ is the corresponding value for the unstrained infinite chain in the fully extended conformation (FES). $\mu^{\infty}_{\mathrm{FES}}$ is chosen as energy zero since FES lacks hb's.

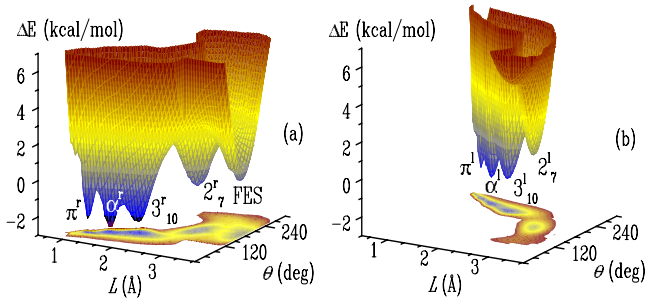

FIG. 2. Calculated potential energy surface (PES) of an infinite polyalanine chain as a function of the rotation θ and the translation L along the chain axis. (a) PES connected to structures in which $50°<\chi<245°$. (b) PES connected to structures in which $\chi\leq 50°$ or $\chi\geq 245°$. Labels indicate the conformation connected to each minimum.

E^{∞}, (L, θ) is calculated employing Troullier–Martins pseudopotentials,[17,18] as incorporated in the parallel version of the FHI98MD code.[19] The energy cutoff of the plane-wave basis set is 70 Ry. Convergence checks with respect to the number of k-points show that the Γ-point, or two k-points [(0, 0, 1/4), (0, 0, 3/4)] for structures in which $c<6.4$ Å, are sufficient for an adequate sampling of the Brillouin zone. The geometry is optimized constraining L and θ and fully relaxing the internal degrees of freedom. Since m and N must be integer numbers inside the supercell the PES is calculated on an irregular grid of points (L_i, θ_i), in which $\Delta L \sim 0.01$ Å and $\Delta\theta<15°$ around minima and saddle points. The final map is thus obtained interpolating a 650-point irregular grid using cubic splines. The direction of the rotations, i.e., the conformation handedness, is considered as counter clockwise, and the chain axis direction as pointing from the amide nitrogen to the carbonyl carbon in a residue. Thus $\theta<180°$ characterizes right-handed conformations and $\theta>180°$ the left-handed ones.

III. RESULTS

The PES associated to the folding of an infinitely long FES of polyalanine into a helical conformation is shown in Fig. 2. The PES is presented as two separate plots as ΔE (L, θ) is double valued for 2.5 Å$<L<$3.3 Å, i.e., two different conformations characterized by the same (L, θ) parameters minimize the DFT-PBE energy respect to the internal degrees of freedom. The latter is a consequence of the possibility for the peptide plane to flip between two possible conformations without altering the position of the C_α atoms flanking it, hence L and θ remains the same for both conformers. As the relative orientation of the side group respect to the peptide plane is different in these conformers, they are energetically no equivalent. Such peptide-plane flipping has been observed in several molecular dynamics simulations[20] and it is expected to play an important role during early stages of protein folding.[21] We use here the $C-N-C_\alpha-C_\beta$ (χ) dihedral angel to measure the relative orientation of the side group respect to the peptide plane. Polyalanine structures in which $50°<\chi<245°$ are connected to the PES in Fig. 2(a) (thereafter PES A), otherwise structures are connected to the PES in Fig. 2(b) (thereafter PES B). (L, θ) parameters for minimum energy structures are given in Table

TABLE I. (L, θ) parameters connected to the minima of the potential energy surface shown in Fig. 2.

	Right-handed		Left-handed	
Structure	L (Å)	θ (deg)	L (Å)	θ (deg)
π-helix	1.17	80.0	1.16	278.0
α-helix	1.50	98.2	1.47	263.5
3_{10}-helix	1.95	115.5	1.93	244.5
2_7	2.83	173.0	2.81	190.0
FES	3.57	183.5		

I. In PES A minima occurring at $L<2.5$ Å are related to right-handed helical conformations: the 4.5^r_{16}-helix (i.e., 4.5 residues per turn and 16 atoms in the ring formed by the hb's, the superscript stands for right handed) or π^r-helix, the 3.6^r_{13}-helix or α^r-helix and the 3^r_{10}-helix [Fig. 2(a)]. Minima occurring at $L>2.5$ Å are related to extended structures: the right-handed 2^r_7-conformation or γ'-turn and the FES [Fig. 2(a)]. The PES B reveals four minima connected to left-handed conformations: the π^l-helix, the α^l-helix, the 3^l_{10}-helix and the 2^l_7-conformation or γ-turn [Fig. 2(b)]. The minimum energy pathway along PES A and PES B projected on the L coordinate is presented in Fig. 3(a). Such projection shows that the α^r-helix minimum is the lowest one, as expected. It also shows that the minimum energy pathway along PES B is higher in energy by ~1 kcal/mol with respect the minimum energy pathway along PES A, owing to larger repulsive interactions in left-handed conformations. Regions with $L<0.7$ Å were excluded as they cannot be accommodated by our supercell.

To corroborate if the predicted conformational space is realistic, the latter is compared to experimentally observed backbone conformations. We analyze an extensive set of protein structures derived from x-ray crystallography. These structures correspond to those listed in the PDB as determined at 1.5 Å resolution or better, with a sequence identity of less than 50% and amount to 1562 structures to date. We have corroborated that the outcome from the comparison (see below) is independent from the x-ray resolution and the percent of sequence identity of the analyzed protein set. The total number of analyzed residues amount to $\sim 3.6\times 10^5$ residues. This set of residues does not contain glycines and prolines as the conformational space of these two residues is known to be markedly different from others, owing to the lack of a side group in glycine and the constraints imposed by the ring formation in the proline backbone.[3] Residues flanking glycines and prolines were also excluded from the analyzed set for avoiding conformations induced by the peculiarities of the structure of glycine and proline. The analyzed residues were further classified according to its χ angle: residues in which $50°<\chi<245°$ were grouped into the set A, otherwise were grouped into the set B. For the purpose of the comparison the conformation of these residues is characterized in terms of (L, θ) cylindrical coordinates instead of the standard (ϕ, ψ) parameters. The rotation θ_i and the translation L_i along the local chain axis z_i characterizing the conformation of the ith residue [Fig. 1(b)] are calculated using an orthogonal transformation that superimposes the geometry of two consecutive peptide planes, a method known as quaternion-based superposition fit.[22–24] The optimal rotation and translation that minimize the root mean square deviation of distances between the set of orthogonally transformed coordinates and target coordinates is given by the eigenvector with the smallest eigenvalue of a certain 4×4 matrix **M**. The first element of this eigenvector gives θ_i and the remaining three elements determine z_i. The elements of the matrix **M** are formed from the two sets of coordinates that will be superimposed with their centroids moved to the origin. This method characterizes adequately the secondary structure of proteins with the exception of the handedness,[23] for getting the handedness, the direction of z_i is set such that it points from the nitrogen atom to the carbonyl group in a residue. Then the handedness that produces the best fit between the transformed and the target coordinates is chosen. In Fig. 4 the distribution of (L, θ) values is plotted over the predicted conformational space. (L, θ) values from the set A of residues were plotted over the PES A and the corresponding (L, θ) from the set B over the PES B. Figure 4 shows that the clustering of (L, θ) values approximately follows the shape of the calculated PES. Regions in which backbone conformations are not allowed according to the DFT-PBE results [enclosed by dashed lines in Fig. 4] are

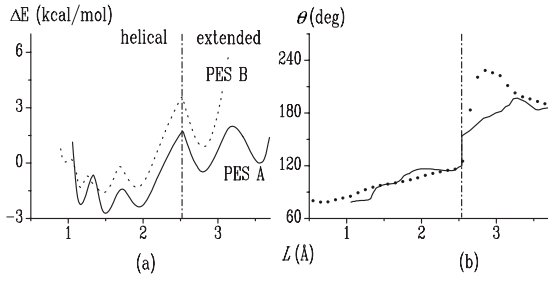

FIG. 3. (a) Minimum energy pathway projected on L. Solid line, projection corresponding to PES A. Dotted line, projection corresponding to PES B. (b) θ_i values from residues in the set A averaged over $L_i \pm \Delta L$ intervals, with $\Delta L=0.05$ Å (dots). θ_i values along the minimum energy pathway of the PES in Fig. 2(a) (solid line). Dashed line indicates the transition between the helical and extended regions.

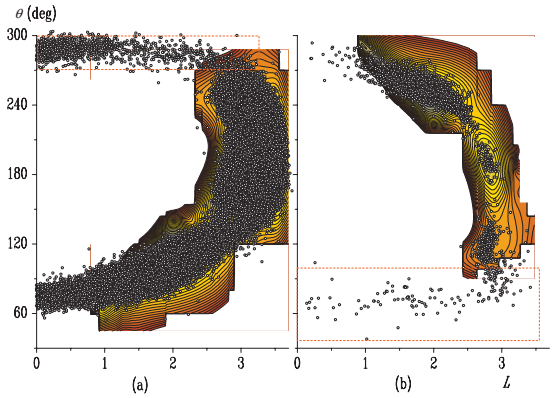

FIG. 4. Distribution of (L_i, θ_i) values obtained from an extended set of protein structures derived by x-ray crystallography. (a) Values from the set A of residues plotted over PES A. (b) Values from the set B of residues plotted over PES B. Dashed lines enclose sparsely populated regions in which DFT-PBE predicts that conformations should not be observed. These regions are populated by less than 0.5% of the analyzed residues.

also populated although sparsely, i.e., less than 0.5% of the analyzed residues. These residues may be stabilized in such conformations as a result of the tertiary structure of the proteins in which other hb's may be formed; e.g., between side groups or side group–water. Moreover we cannot rule out the possibility that some of these conformations results from artifacts of the crystallographic refinement.[2,3] A further analysis of the (L, θ) values from residues in the set A, shows that the values for θ averaged over $L \pm \Delta L$ intervals, with $\Delta L = 0.05$ Å, are essentially located along the minimum energy pathway of the calculated PES [Fig. 3(b)], except for the values corresponding to 2.5 Å $< L <$ 3.25 Å. This discrepancy results from a large population of residues adopting a PPII conformation. DFT-PBE does not predict a minimum connected to the PPII conformation and the region on the PES A connected to such conformation is aside the minimum energy pathway. Nevertheless, only ~10% of analyzed residues populate the PPII region.

IV. DISCUSSION

The action of peptide hb's and long-range electrostatic interactions on the conformational space of a residue is illuminated in greatest clarity by the PES presented above. This PES is qualitatively different from the PES obtained using dipeptides or short polypeptide chains in several respects but primarily due to the presence of well formed minima associated to the helical conformations. The PES obtained with dipeptide models in gas phase does not show a minimum associated to the α-helix conformation regardless of whether electronic structure method, DFT or MP2, is used for doing the calculations.[7,8] In a PES calculated with a short polypeptide chain (a hexapeptide in gas phase calculated using DFT and the Becke–Lee–Parr–Yang approximation to the exchange-correlation functional) a single minimum associated to both right-handed and left-handed helical conformations is revealed[25] (in the PES calculated here three minima for each handedness are revealed), and the right-handed helix minimum is not the lowest in energy. According to experimental results very short polyalanine peptide chains do not form helices neither in aqueous solvent nor in gas phase,[26–28] e.g., peptide chains with at least seven residues are needed to form helices in gas phase.[28] The fact that long peptides are needed to observe the helix formation may be indicative of the crucial role of hb cooperativity and long-range electrostatic interactions for forming the helical conformations both in solvent and in gas phase. Certainly solvent dramatically influence the formation of helices. At the moment we cannot say, based on electronic structure calculations, how solvent will affect the PES calculated here. Further investigations are needed to assess the effect of solvent in the process of the helix formation. Investigations along this line are work in progress in our group.

In the PES presented above is particularly noticeable the reduction of the conformational space due to the development of hb's between non-nearest-neighbor residues (i.e., for $L <$ 2.5 Å): it favors only one of two possible conformations that could result from peptide-plane flipping, and restricts the θ dimension to a shorter range. In proteins the permanent peptide dipole carried by each residue is fully compensated by the environment (i.e., solvent, the specific folding of the protein, charged residues, etc.). Such dipole compensation is reproduced by infinite chain models owing to the periodic boundary conditions used for calculating the DFT-PBE energy. Infinite chain models are thus a limiting case in which a residue is in the presence of a fully cooperative network of hb's. These facts may contribute for the overall good agreement between the predicted conformational space and the experimentally observed backbone conformations. Nevertheless, discrepancies exist likely due to solvent, side chain and temperature effects that were neglected in our investigation. Moreover protein tertiary structure, which is also neglected in the PES calculation, could stabilize certain conformations too. Particularly the high density population along the PPII region cannot be anticipated from the predicted conformational space. At the moment we cannot say which of the neglected factors for calculating the PES are responsible for the observed discrepancy. In literature aqueous solvent[29,30] and electronic effects[31] like the $n \rightarrow \pi^*$ interaction between the oxygen of the carbonyl group and the carbon of the next carbonyl group along the peptide backbone have been mentioned as the stabilizing factors for the PPII conformation. To investigate the role of water in the observed discrepancy we have analyzed only the set of residues from the protein crystals that are not in contact with water, i.e., we have excluded from the set of analyzed residues those in which the distance between the water oxygen and the nitrogen/oxygen from the peptide group is smaller than e.g., 4 Å. We find that the density of the population along the PPII region significantly reduces respect to the density of population in the other regions. Still a considerable number of residues adopt the PPII conformation. It seems that aqueous solvent is not the only factor stabilizing PPII conformations. Whether the $n \rightarrow \pi^*$ interaction or other factors such as the tertiary structure or van der Waals interactions are also stabilizing the PPII conformation requires further investigations which are beyond the scope of this work.

V. CONCLUSIONS

In conclusion, our results show that an infinite polyalanine chain is a good model for investigating the backbone conformational space of a residue embedded in a long chain under the influence of both long-range (hb's and electrostatics) and nearest-neighbor interactions. Thus the PES of infinite polypeptide chain represents a limiting case that can be used for, e.g., critically assess different methodologies for describing the protein structure.[32] Moreover such PES succeeds in describing the backbone conformational space of folded proteins in a simple two-dimensional representation. It is thus like a Ramachandran plot but in cylindrical coordinates in which hydrogen bonding and its cooperative effect are fully taken into account. Therefore helical and extended regions are clearly distinguished and the *ab initio* predicted conformational space agrees with the observed one.

ACKNOWLEDGMENTS

J.I. acknowledges support from "programa de repatriació 2007, No. MOD-ORD-12-08 PCI-085-04-08" of the CONACYT.

[1] T. E. Creighton, *Proteins: Structures and Molecular Properties* (Freeman, New York, 1993).
[2] G. J. Kleywegt and T. A. Jones, Structure (London) **4**, 1395 (1996).
[3] S. C. Lovell, I. W. Davis, W. B. Arendall III, P. I. W. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson, and D. C. Richardson, Proteins **50**, 437 (2003).
[4] B. K. Ho, A. Thomas, and R. Brasseur, Protein Sci. **12**, 2508 (2003).
[5] T. Head-Gordon, M. Head-Gordon, M. J. Frisch, C. L. Brooks, and J. A. Pople, J. Am. Chem. Soc. **113**, 5989 (1991).
[6] C.-H. Yu, M. A. Norman, L. Schäfer, M. Ramek, A. Peeters, and C. van Alsenoy, J. Mol. Struct. **567–568**, 361 (2001).
[7] R. Vargas, J. Garza, B. P. Hay, and D. A. Dixon, J. Phys. Chem. A **106**, 3213 (2002).
[8] Z.-X. Wang and Y. Duan, J. Comput. Chem. **25**, 1699 (2004).
[9] Z. Gáspári, I. Hudáky, A. Czajlik, and A. Perczel, J. Mol. Struct.: THEOCHEM **675**, 141 (2004).
[10] J. Ireta, J. Neugebauer, and M. Scheffler, J. Phys. Chem. A **108**, 5692 (2004).
[11] J. Ireta, J. Neugebauer, M. Scheffler, A. Rojo, and M. Galván, J. Phys. Chem. B **107**, 1432 (2003).
[12] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[13] R. Kaschner and D. Hohl, J. Phys. Chem. A **102**, 5111 (1998).
[14] R. Improta, V. Barone, K. N. Kudin, and G. E. Scuseria, J. Am. Chem. Soc. **123**, 3311 (2001).
[15] J. Ireta, J. Neugebauer, M. Scheffler, A. Rojo, and M. Galván, J. Am. Chem. Soc. **127**, 17241 (2005).
[16] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucleic Acids Res. **28**, 235 (2000).
[17] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[18] M. Fuchs and M. Scheffler, Comput. Phys. Commun. **119**, 67 (1999); see: http://www.fhi-berlin.mpg.de/th/fhi98md/fhi98pp.
[19] M. Bockstedte, A. Kley, J. Neugebauer, and M. Scheffler, Comput. Phys. Commun. **107**, 187 (1997); see: http://www.fhi-berlin.mpg.de/th/fhi98md/.
[20] A. Kitao, S. Hayward, and N. Go, Proteins **33**, 496 (1998).
[21] S. Hayward, Protein Sci. **10**, 2219 (2001).
[22] S. K. Kearsley, Acta Crystallogr., Sect. A: Found. Crystallogr. **45**, 208 (1989).
[23] G. R. Kneller and P. Calligari, Acta Crystallogr., Sect. D: Biol. Crystallogr. **62**, 302 (2006).
[24] J. R. Quine, J. Mol. Struct.: THEOCHEM **460**, 53 (1999).
[25] I. A. Solov'yov, A. V. Yakubovich, A. V. Solov'yov, and W. Greiner, Phys. Rev. E **73**, 021916 (2006).
[26] S. Marqusee, V. H. Robbins, and R. L. Baldwin, Proc. Natl. Acad. Sci. U.S.A. **86**, 5286 (1989).
[27] G. E. Job, R. J. Kennedy, B. Heitmann, J. S. Miller, S. M. Walker, and D. S. Kemp, J. Am. Chem. Soc. **128**, 8227 (2006).
[28] M. F. Jarrold, Annu. Rev. Phys. Chem. **51**, 179 (2000).
[29] C. D. Poon, E. T. Samulski, C. F. Weise, and J. C. Weisshaar, J. Am. Chem. Soc. **122**, 5642 (2000).
[30] F. Eker, K. Griebenow, and R. Schweitzer-Stenner, J. Am. Chem. Soc. **125**, 8178 (2003).
[31] J.-C. Horng and R. T. Raines, Protein Sci. **15**, 74 (2006).
[32] E. Penev, J. Ireta, and J.-E. Shea, J. Phys. Chem. B **112**, 6872 (2008).