PROTEINS: Structure, Function, and Genetics 3:262–265 (1988)

SHORT COMMUNICATION

Microfolding: Conformational Probability Map for the Alanine Dipeptide in Water From Molecular Dynamics Simulations

Amil G. Anderson and Jan Hermans
Department of Biochemistry, School of Medicine, University of North Carolina, Chapel Hill, North Carolina 27599-7260

ABSTRACT **A direct attack on the protein-folding problem has been initiated with the free energy perturbation methods of molecular dynamics. The complete conformational probability map for the alanine dipeptide is presented. This work uses the SPC model for the explicit hydration of the dipeptide. Free energy differences for the four observed minima (β,α_R,α_L,C_7^{ax}) are given, and the free energy barriers between minima are outlined.**

Key words: protein conformation, conformational equilibria, free energy stimulation, protein folding, thermodynamic perturbation

The alanine dipeptide, N-acetylalanyl-N-methylamide, serves as a paradigm for studying the thermodynamics of protein conformations and folding. The dipeptide has two major degrees of freedom, the phi (ϕ) and psi (ψ) dihedral angles, which also determine the conformation of the carbon backbone of proteins. Hydrophobic interactions, polar group hydrogen bonds, and the dipole-dipole interaction between adjoining peptide bonds compose the overall force-field that influences the conformational equilibrium of the dipeptide in a water environment, as well as that of protein molecules. Free energy calculations based on molecular dynamic simulations have allowed us to compute directly the equilibrium distribution for conformational states of the dipeptide in water. In addition, we have mapped the free energy along the pathways between conformations.

While Monte Carlo simulations[1] and integral equation theory[2] have been used to evaluate free energies of hydration for the dipeptide model, our goal has been distinctly different. Our study of *microfolding*, i.e., the establishment of conformational equilibria for the small dipeptide model, is aimed at providing insight into the larger problem of how proteins fold and maintain their specific three-dimensional structure. We have applied molecular dynamics simulations, with their explicit treatment of water and detailed trajectories, to this conformational study.

Two complementary approaches were taken to calculate the ϕ-ψ conformational probability surface for the alanine dipeptide. First, separate simulations were run in the regions of the previously identified potential energy minima conformations, α_R, C_7^{ax}, α_L, and β. The probability surfaces derived from the trajectories of these simulations allowed us to identify the most probable conformations for our model of the dipeptide in solution. The free energy differences between the four energy minima were then calculated using the thermodynamic integration method.[3,4] The dipeptide was forced from one conformation to another by the application of an extra potential to the dihedral angles.[5] The perturbation results in an applied torque causing internal rotation about a main chain bond. The ΔA_{move} values were calculated with 250,000 step (2 fs/step) simulations with duplicate simulations run in the reverse direction to show the thermodynamic reversibility of the process. The change in the free energy, ΔA_{impose}, which represents the loss in conformational freedom caused by the imposition of the perturbing potential, was calculated from unperturbed trajectories.[5] The free energy difference between two conformations A and B of the dipeptide is then given by:

$$\Delta A_{A \rightarrow B} = \Delta A^{A}_{impose} + \Delta A_{move} - \Delta A^{B}_{impose}$$

With these free energy differences we were able to scale together the regional probability distributions to produce an overall ϕ-ψ probability map.

One can see from the map in Figure 1 that the ϕ-ψ conformations representing the free energy minima are $\beta(-110, 120)$, $\alpha_R(-120, -40)$, $\alpha_L(60, 100)$, and $C_7^{ax}(70, -60)$. The α_L and C_7^{ax} conformations would not be expected to play a very frequent role in protein structures, and indeed Richardson's compilation of the crystallographic data shows this to be the case.[6,7] For the α_R region, one can actually identify two minima, $(-70, -40)$ being close to the preferred conformation for alanine in α-helices and a lower free energy minimum shifted 50° in ϕ to $(-120, -40)$. This is a clear example of how the forces involved in secondary structure, i.e., peptide hydrogen bonds, can shift the conformational equilibrium away from the free energy minimum determined by hydrogen bonding with solvent. Richardson's data clearly show that

Received March 17, 1988; accepted March 24, 1988.

Address reprint requests to Jan Hermans, Dept. of Biochemistry, University of North Carolina, FLOB, Chapel Hill, NC 27514.

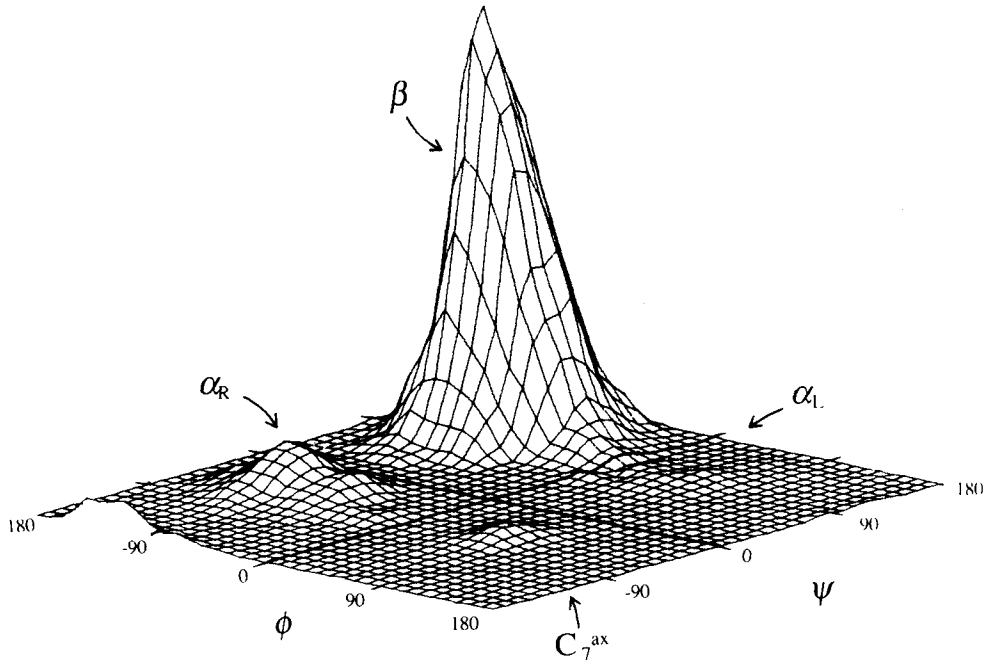
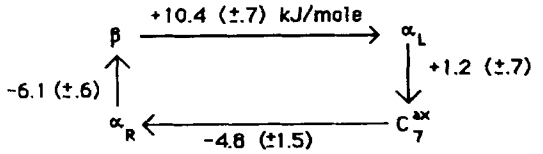


Fig. 1. Molecular dynamics simulations were run using CEDAR[13] and periodic boundary conditions with a 6 Å cutoff for nonbonded interactions applied to the dipeptide in a box of 76 SPC[14] water molecules. The dipeptide is kept centered, surrounded by 2 layers of water. Hydrogens on carbon atoms were treated as being part of a larger carbon atom, i.e., a united atom model. The length of the unperturbed simulations in the β, α_L, C_7^{ax}, and α_R regions using a time step of 2 fs were 500 ps, 44 ps, 50 ps, and 89 ps, respectively. For the free energy difference simulations, the perturbation potential was of the form: $V(\lambda,\phi) = K_\phi [1 - \cos(\phi - \phi_i - \lambda(\phi_f - \phi_i))]$ where ϕ is the value of the dihedral angle before the perturbation is applied, ϕ_i and ϕ_f are the initial and final values of the dihedral angle ϕ, and similarly for the dihedral angle ψ. As λ goes from 0 to 1, the angle for the minimum of the potential goes from the initial to the final conformation, while the free energy difference, ΔA_{move}, is accumulated.[5] The probability distributions for each region were individually normalized. The α_R, α_L and C_7^{ax} distributions were then scaled relative to the β region distribution by multiplying each of the three distributions by the factors $\exp[-\Delta A_{\beta\rightarrow X}/RT]$, where $\Delta A_{\beta\rightarrow X}$ is the free energy change going from β to one of the other regions. The scaled distributions were then combined with the β region distribution to produce a new probability map for all of ϕ-ψ space, which was then normalized. The statistics for this map favor the equilibrium regions, i.e., probability maxima, with very little sampling of the transition regions.

when only residues not part of well-defined secondary structures are considered, the distribution of conformations is more in line with the dipeptide model.[7] The β region distribution in our map is clearly compatible with the crystallographic data and does not show evidence of the multiple energy minima that have been suggested by other studies.[8] These previous studies had been based mainly on vacuum models of the dipeptide; other work has shown that the conformational distribution is markedly effected by the presence of the polar solvent water, which has a tendency to flatten the overall map.[2]

The free energy data used to scale the map in Figure 1 show how well the molecular dynamics free energy methodology can work. The thermodynamic cycle for the conformational change



has a closure error of less than 1 kJ/mole, a value that greatly exceeds the precision of the currently available experimental data for the solution equilibria of the alanine dipeptide.[9–12] We anticipate that recent advances in the areas of peptide synthesis and nuclear magnetic resonance will allow a direct experimental verification of these results in the near future.

The probable transition paths between the different conformations were examined with another set of simulations. A single-angle perturbation was performed where one of the dihedral angles was forced to move from a value characteristic of one conformation to that of another conformation (e.g., move from $\beta \rightarrow \alpha_r$ by forcing ψ from $+120°$ to $-60°$). The other dihedral angle was not restricted by a forcing potential and thus was allowed to explore a wider range of conformations. The total set of simulations for the sampling of the conformational paths took 1.6 million steps or 3.2 ns of simulated time. The data collected from these simulations give an informative view (Fig. 2) of the transition regions that were not sampled by the unperturbed simulations and poorly sampled by the earlier double-angle perturbation trajectories, as the dipeptide conformation was too tightly constrained during these moves. It is interesting to note that the free energy barriers are only 2–4 kJ/mole greater than the free energy difference be-

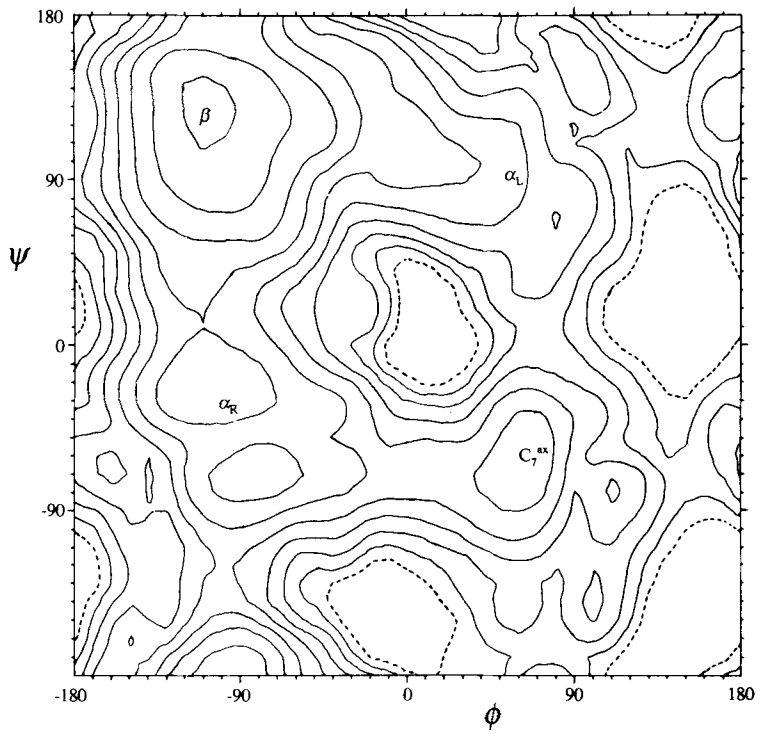


Fig. 2. Simulations were run using the same basic form as in Figure 1, except that only a single dihedral had a perturbing potential applied. This single dihedral, ρ, is one of the two dihedrals angles ϕ or ψ. For each pair of conformations A and B, the perturbed dihedral angle went from A via $\rho = 0°$ to B and back in the reverse direction in a total of 200,000 steps. Other simulations went from A via $\rho = \pm 180°$ to B and back, again in 200,000 steps. For each of the separate paths a probability map was calculated. The perturbation of the system leads to an oversampling, relative to a Boltzmann distribution, of the less favorable regions between the most probable conformations. This bias is corrected by scaling the resulting probability distribution with the free energy function $\Delta A(\rho)$ accumulated as the trajectory move from the beginning to end conformations, i.e., for a simulation where ϕ is the angle being forced, all cells in a probability map, $P(\phi,\psi)$, with a given ϕ value are multiplied by the factor $\exp[-\Delta A(\phi)/RT]$. The scaled probability maps were then added together, using the regions of overlap to adjust each map relative to the accumulating total map. Contour levels were drawn at 2 kJ/mole intervals. The lowest drawn probability contour (highest free energy) is shown with a dashed line and corresponds to 18 kJ/mole. The highest probability contour is in the center of the β region. The statistics for this map favor the transition regions; the details of the equilibrium positions are shown best in Figure 1.

tween endpoint conformations. The elementary step of the folding process, i.e., internal rotation about a main chain single bond, can apparently occur with equal ease about the two bonds on either side of a small side chain, and similarly in either direction over a full 360°.

An examination of other dipeptide models would require the expenditure of ~40 Cray 2 single processor hour equivalents. This is not a trivial cost, but knowledge of how larger side chains or sidechains with different polar properties affect backbone conformational equilibria would further our understanding. The methodology can also be applied to peptide models with more residues, in which case it becomes possible to address questions of secondary structure formation, albeit at a higher computational cost.

ACKNOWLEDGMENTS

We would like to acknowledge the financial support of NSF (equipment, personnel, and supercomputer time) and NIH (equipment). The Computer Science Department of Duke University allowed us generous use of their Convex C-1 mini-supercomputer.

REFERENCES

1. Mezei, M., Mehrotra, P.K., Beveridge, D.L. Monte Carlo Determination of the Free Energy and Internal Energy of Hydration for the Ala Dipeptide at 25C. J. Am. Chem. Soc. 107:2239–2245, 1985.
2. Pettitt, B.M., Karplus, M. The potential of mean force surface for the alanine dipeptide in aqueous solution: A theoretical approach. Chem. Phys. Lett. 121:194–201, 1985.
3. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W. Statistical mechanics and molecular dynamics: The calculation of free energy. In: "Molecular Dynamics and Protein Structure." Hermans, J., ed. Western Springs, IL: Polycrystal Book Service. 1985:43–46.
4. McCammon, J.A., Harvey, S.C. "Dynamics of Proteins and Nucleic Acids." Cambridge: Cambridge University Press. 1987:60–73.
5. Anderson, A., Carson, M., Hermans, J. Molecular dynamics simulation study of polypeptide conformational equilibria: A progress report. Ann. N.Y. Acad. Sci. 482:51–59, 1986.
6. Richardson, J.S. The anatomy and taxonomy of protein structure. Adv. Protein Chem. 34:167–339, 1981.
7. Richardson, J.S., Richardson, D.C. Principles and patterns of protein conformation. In: "Prediction of Protein Structure and the Principles of Protein Conformation." Fasman, G., ed. New York: Plenum Press. In press.
8. Beveridge, D.L., Ravishanker, G., Mezei, M., Gedulin, B. Solvent effects on conformational stability in the Ala dipeptide: Full free energy simulations. In: "Biomolecular Stereodynamics III. Proceedings of the Fourth Conversation in the discipline Biomolecular Stereodynamics." Sarma, R.H.,

Sarma, M.H., eds. Albany, New York: Adenine Press. 1985:237–252.
9. Brant, D.A., Flory, P.J. The role of dipole interactions in determining polypeptide configurations. J. Am. Chem. Soc. 87:663–664, 1965.
10. Brant, D.A., Flory, P.J. The configuration of random polypeptide Chains. I. Experimental results. J. Am. Chem. Soc. 87:2788–2791, 1965.
11. Avignon, M., Garrigou-Lagrange, C., Bothorel, P. Conformational analysis of dipeptides in aqueous solution. II. Molecular structure of glycine and alanine dipeptides by depolarized Rayleigh scattering and laser Raman spectroscopy. Biopolymers 12:1651–1669, 1973.
12. Madison, V., Kopple, K.D. Solvent-dependent conformational distributions of some dipeptides. J. Am. Chem. Soc. 102:4855–4863, 1980.
13. Carson, M., Hermans, J. The molecular dynamics workshop laboratory. In: "Molecular Dynamics and Protein Structure." Hermans, J., ed. Western Springs, IL: Polycrystal Book Service. 1985:43–46.
14. Hermans, J., Berendsen, H.J.C., van Gunsteren, W.F., Postma, J.P.M. A consistent empirical potential for water-protein interactions. Biopolymers 23:1513–1518, 1984.