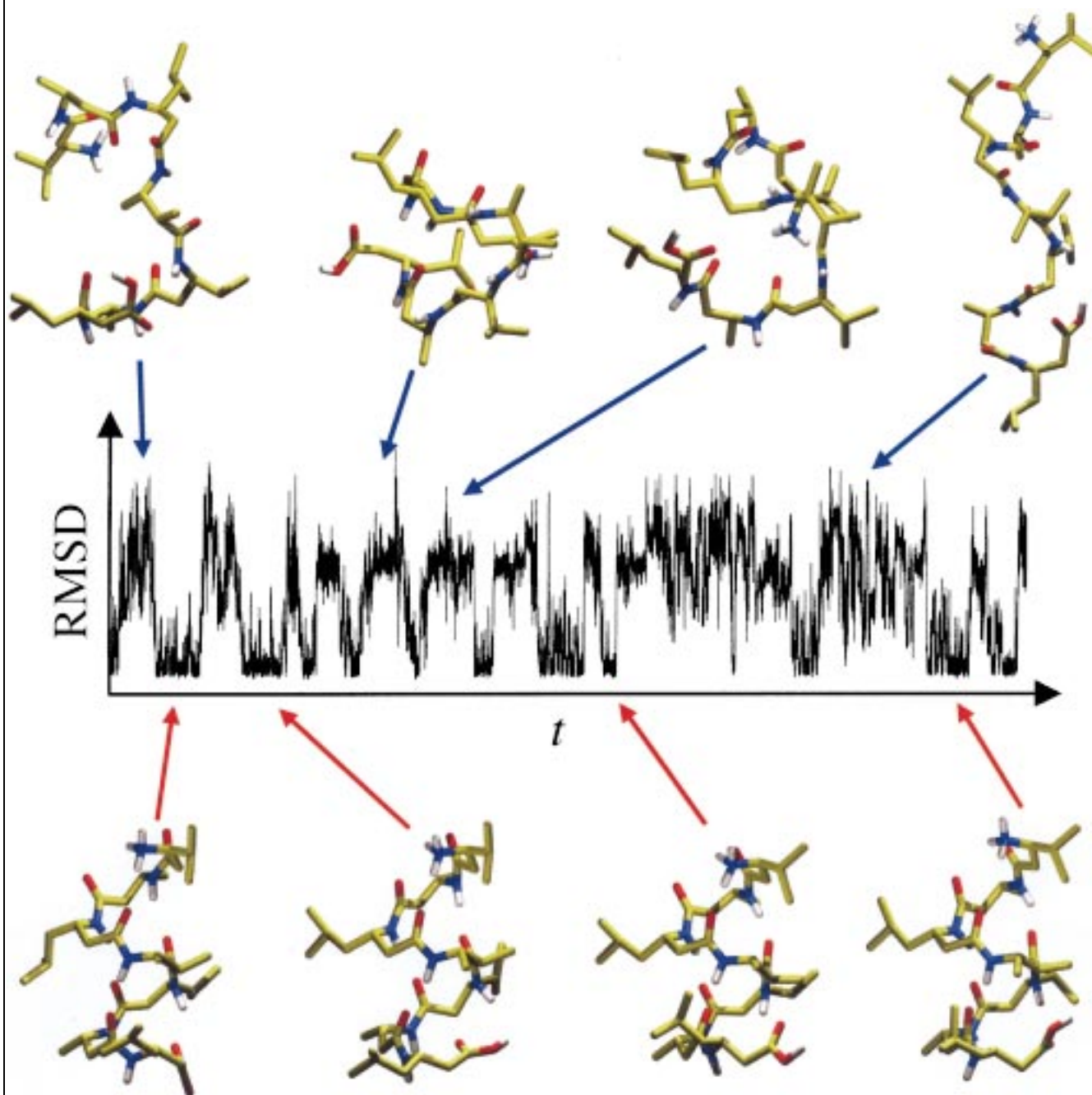


A 200 ns long simulation of the reversible folding of a β amino acid heptapeptide



The upper series of structures show “snap shots” of the various unfolded conformations at different time points of the simulation, as indicated by the blue arrows. The lower series shows the folded states (red arrows).

The helical conformations that are repeatedly occupied are also evident from the change in the root-mean-square deviation (RMSD) of the backbone atom positions during the simulation of the model structure based on the NMR experiments.

Find out more on the following pages.

The Key to Solving the Protein-Folding Problem Lies in an Accurate Description of the Denatured State**

Wilfred F. van Gunsteren,* Roland Bürgi, Christine Peter, and Xavier Daura

The prediction of the folding process of a protein (that is, its most stable conformer, the relative stability of the different accessible conformations, and the kinetics of folding) as a function of temperature and protein environment (that is, the solvent composition) is one of the fundamental long-standing challenges of molecular biology.^[1, 2] It is still unsolved due to the enormous size of the conformational space of a protein on the one hand, and the relatively small energetic differences between native and denatured conformations on the other.^[3] Looking at three possible conformers per backbone ϕ or ψ angle in an amino acid residue, a 10-residue peptide has about $3^{20} \approx 10^9$ conformers available, a 100-residue protein has about 10^{90} . The conformational space available to a polypeptide grows exponentially with its length, leaving no hope to computationally sample all conformers. Nature does not sample all conformations in the folding process either, but restricts herself to a subset, which forms a folding pathway. These pathways are very difficult, if not impossible, to elucidate experimentally at the atomic level.^[2, 4] Using the ever-increasing power of computers it has become possible to simulate the reversible folding of small peptides.^[5-7] Analysis of the thermodynamic equilibria between folded and unfolded conformers shows that the unfolded or denatured state comprises a relatively low number of conformers,^[8] which does not grow exponentially with the length of the peptide. The simulation of these folding/unfolding equilibria requires not only an accurate description of the folded state but also, most importantly, of the denatured state in order to obtain the correct relative populations of folded and unfolded conformations and their transition rates.

The most stable fold of a protein depends on its amino acid composition, solvent environment (composition, pH value), and thermodynamic state (temperature, pressure). Interactions at the atomic level determine the equilibrium between folded and unfolded conformers. Computer simulation of the statistical mechanical ensemble could be used to calculate the folded/unfolded equilibrium, molecular dynamics (MD) simulation could yield the kinetics of the folding process. However, in view of the huge number of conformations that in principle constitute the denatured state, MD simulation in atomic detail of the folding/unfolding equilibrium was considered to be out of computational reach.^[9] In folding studies, the molecular models used were of a simple nature: one interaction site per residue,^[9, 10] a neglect^[9-12] or a mean-field approximation^[13] of the solvent degrees of freedom, motion

restricted to lattice sites,^[14] etc. Often, force-field parameters were tuned to favor a (particular) fold,^[10, 11] or a particular folding pathway was postulated.^[15]

These studies gave insight into possible folding mechanisms, but could not tell which one corresponds to reality,^[16] due to the simplicity and bias of the models used. On the other hand, protein structure prediction studies^[17] concentrate on predicting the correct, most stable fold, without considering the folding/unfolding equilibrium or kinetics, and disregarding the dependence of the fold on temperature and solvent environment.

Here, we wish to approach the folding problem from a different point of view. Analyzing dynamic simulations of the reversible folding of peptides of different compositions, in various solvents and at different temperatures, into different secondary structure elements or folds, we show that the denatured state of these peptides comprises vastly less conformations than one would think based on the number of conformational degrees of freedom present. For a peptide with about 20 rotatable bonds the denatured state can be characterized by about $10^2 - 10^3$ conformers. This small size of the denatured state explains why these peptides fold on a nanosecond time scale.

Extending the argument to proteins which fold on a millisecond time scale, their denatured state will extend over say 10^9 conformers, vastly smaller than the 10^{90} quoted earlier. This makes simulation of protein folding at the atomic level amenable in the near future, provided a force field is available that characterizes the denatured state sufficiently accurately. If the denatured state is wrongly described, one can have no hope of stimulating the correct equilibrium. From this point of view, it does not help to base the force field on a statistical analysis of folded structures; an accurate description of the denatured state is also required. Due to the lack of experimental information regarding the denatured state, the only way is to base the force field on physical principles: the

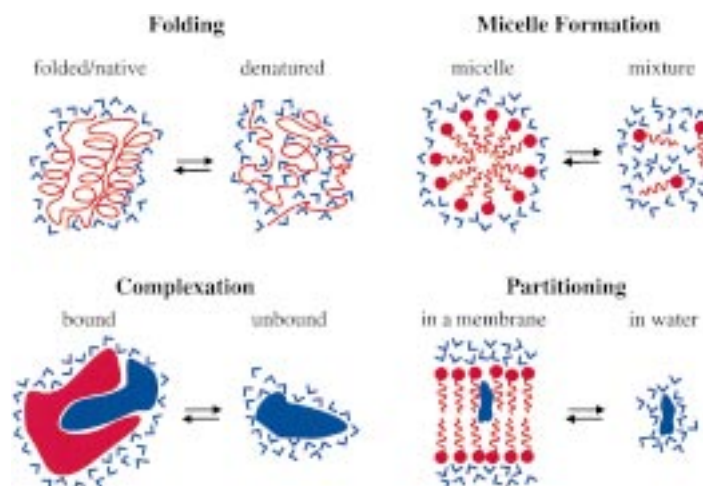


Figure 1. Thermodynamic equilibria that govern key processes in molecular biology. In order to simulate on a computer the correct distribution of, and kinetics between the molecular configurations shown, the general biomolecular force field describing the interatomic interactions^[19] has been calibrated against experimental physico-chemical partitioning data regarding small polar and nonpolar molecules in the condensed phase. Red: protein or lipid molecules. Blue: water molecules or ligand molecules (solid areas) binding to a protein or permeating through a lipid membrane.

[*] Prof. Dr. W. F. van Gunsteren, R. Bürgi, C. Peter, Dr. X. Daura
Laboratory of Physical Chemistry
Eidgenössische Technische Hochschule Zürich
ETH-Zentrum, 8092 Zürich (Switzerland)
Fax: (+41) 1-632-1039
E-mail: wfvgn@igc.phys.chem.ethz.ch

[**] Financial support from the Schweizerischer Nationalfonds (Project no. 21-50929.97) is gratefully acknowledged.

interaction between atoms and molecules in the condensed phase. Here we show that a general atomic-level biomolecular force field whose parameters have been set based on experimental hydrophilic/hydrophobic partitioning data for small molecules in the condensed phase, is able to accurately and reversibly fold a number of peptides into a variety of native folds, and this is not dependent of the starting structure used in the simulations.

Figure 1 illustrates the various thermodynamic equilibria one would like to simulate using a general biomolecular force field.^[18] One of the key factors in these equilibria is the relative strength of polar–polar, polar–nonpolar, and nonpolar–nonpolar interactions. In the GRO-MOS96 force field used,^[19] the nonbonded interaction parameters have been calibrated with solute/solvent partitioning data. The six different peptides studied are shown in Figure 2 with their most stable fold as determined by experiment. The major parameters describing the six systems and simulations are given in Table 1. A successful simulation study of peptide folding would involve the following results:

1. The folded structure should be within a given structural difference from the experimental model structure.
2. The folding should be reversible, more than one folding event should be observed.
3. The distribution of folded versus unfolded conformers in the ensemble should be correct, for example, the free energy of folding ΔG_{fold} and melting temperature T_m should be correct.
4. The rate or kinetics of folding should be correct.
5. All this should be achieved using one general force field for a variety of peptides, solvents, and folds.

Figures 2 and 3, and Table 1 show that the six peptides are

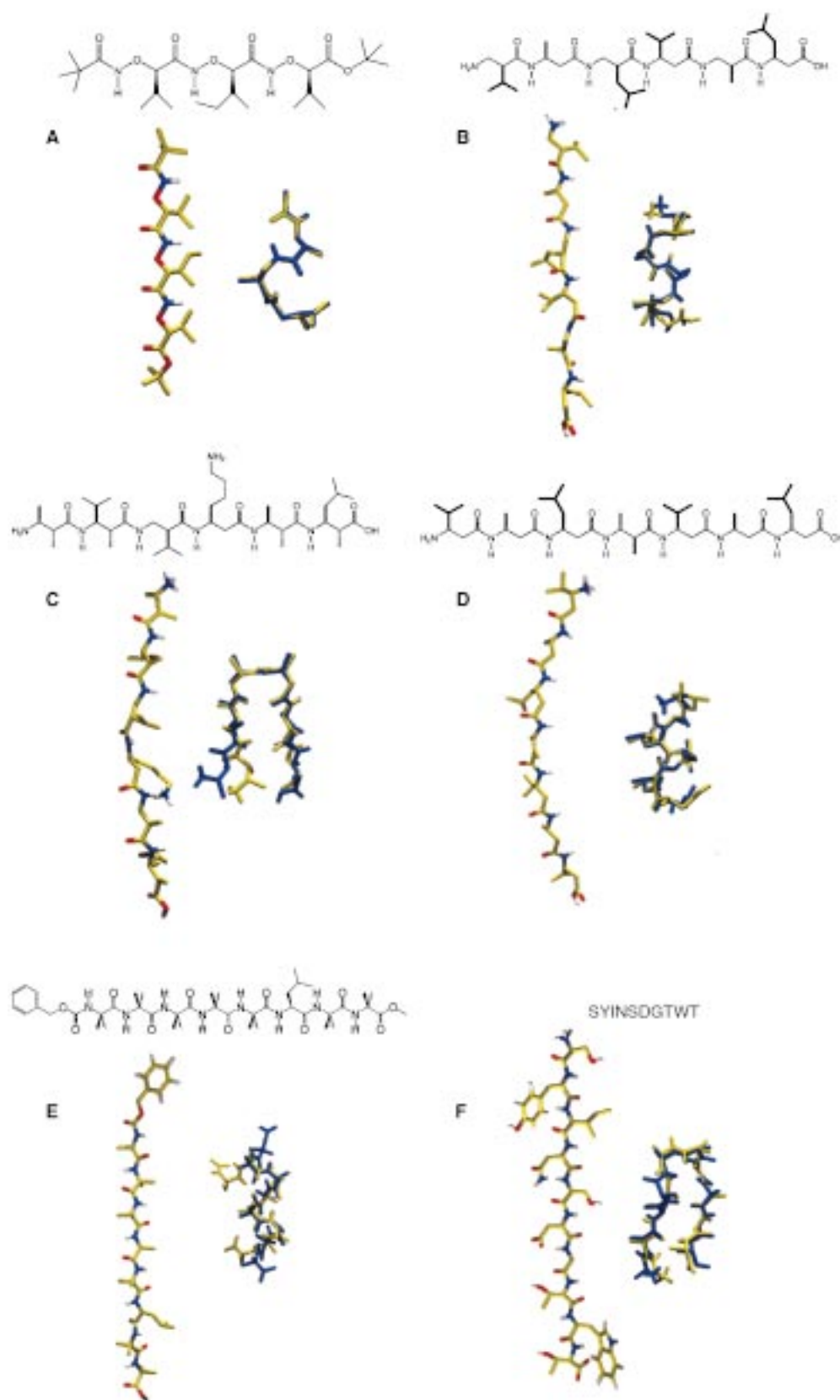


Figure 2. Folding of peptides into different folds. For each of the six different peptides (see Table 1) the following is shown: the chemical formula, an extended structure used as the initial configuration in the MD simulations, a folded structure (yellow) modeled on the available NMR experimental data, and, superimposed on the latter, a folded structure (blue) from the MD simulations of the folding/unfolding equilibrium. For the eight-residue α -amino acid peptide **E**, both the experiments and the simulation indicate the presence of roughly equally populated conformers (*M* and *P*, α and 3_{10} helices), of which one only is shown.

Table 1. Reversible folding of six peptides.^[a]

System	A	B	C	D	E	F
type of peptide	α -aminoxy		β -amino		α -amino	
stable structure (expt)	helix	helix	hairpin	helix	helix	β -hairpin
type of structure	P-1.8 ₈	P-12/10	10-member turn	M-3 ₁₄	M,P, α ,3 ₁₀	3:5
type of solvent	CHCl ₃	MeOH	MeOH	MeOH	DMSO	H ₂ O
Simulation parameters						
no. of residues	3	6	6	7	8	10
no. of peptide atoms	39	56	64	64	75	110
no. of solvent molecules	672	1435	1453	962	1119	4585
no. of atoms in the system	3399	4361	4423	2950	4551	13865
backbone torsional degrees of freedom	9	18	18	21	16	20
temperature <i>T</i> [K]	340	340	340	340	340	353
simulation length [ns]	72	50	100	200	150	32
Simulated folding process						
folded structure (sim.)	helix	helix	hairpin	helix	helix	β -hairpin
	P-1.8 ₈	P-12/10	10-member turn	M-3 ₁₄	M,P, α ,3 ₁₀	3:5
RMSD between folded structure and experimental model [nm]	0.01	0.01	0.03	0.01	0.03	0.06
RMSD between extended conformation and experimental model [nm]	0.20	0.36	0.40	0.52	0.32	0.70
RMSD criterion folded/unfolded [nm]	0.07	0.08	0.08	0.10	0.10	0.12
no. of unfolded conformers	148	129	286	360	344	57
no. of folding events	40	6	15	23	9	2
estimated free energy of folding ΔG_{fold} [kJ mol ⁻¹]	5	5	6	2	11	14
estimated melting temperature T_m [K]	330 ± 20	–	–	330 ± 10	320 ± 10	–

[a] The upper part of the table shows the type of peptide, the most stable conformation (as determined by NMR experiments; see also Figure 2) and the solvent used in both experiment and simulation. The middle part of the table shows the most important simulation parameters characterizing the different systems. The number of peptide conformational degrees of freedom is indicated by the number of backbone rotatable (i.e. nonpeptidic) torsional angles. Although each peptide was simulated at various temperatures, only results for one temperature (close to the melting temperature of the fold) per peptide are shown. More data can be found in ref. [21] for the α -aminoxy acid peptide, in ref. [6, 22] for the six-residue β -amino acid peptides, in ref. [5] for the seven-residue β -amino acid peptide, in ref. [23] for the eight-residue α -amino acid peptide, and in ref. [24] for the ten-residue α -amino acid peptide. The lower part of the table shows the results of the simulations: the most stable conformation, its backbone atom-positional RMSD value with respect to the experimental NMR model structure, the backbone atom-positional RMSD value of an extended structure from the NMR model structure, the backbone atom-positional RMSD value that is used to distinguish correctly folded conformations from unfolded conformations, the number of unfolded conformations, which form the denatured state of the peptide, the number of (un)folding events (see also Figure 3), very rough estimates of the free energy of folding, ΔG_{fold} , determined by the ratio of folded versus unfolded conformations, and, when sufficient data is available, the melting temperature of the peptide, T_m (defined by a 1:1 ratio of folded and unfolded conformations).

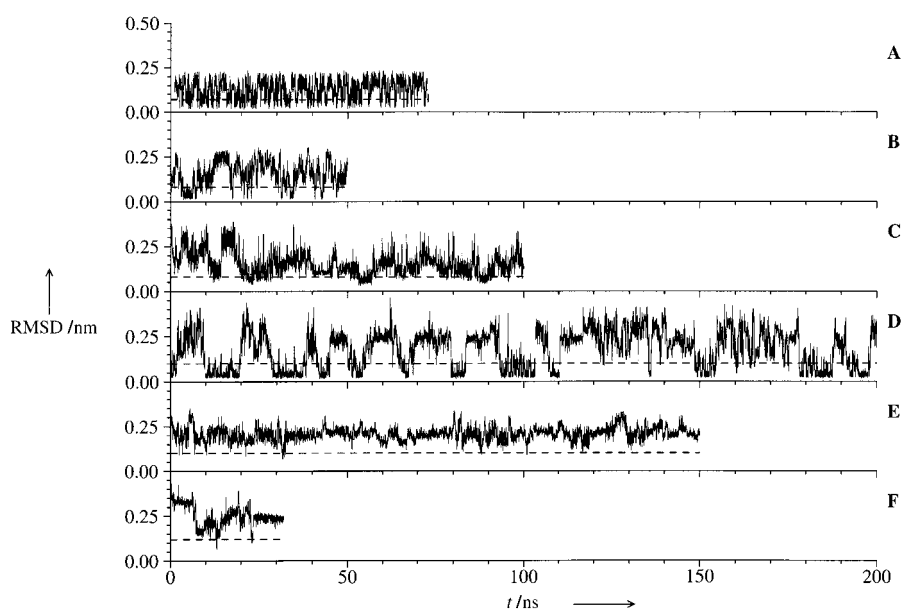


Figure 3. Dynamics of the reversible folding of peptides. For each of the six different peptides (see Table 1 and Figure 2) the backbone atom-positional RMSD of the simulated structures from the experimentally derived NMR model structure is shown as a function of time. Except for the α -aminoxy acid tripeptide A, the atoms of the first and last residues of the peptide chains were excluded when performing the translational and rotational superposition of the structures and when calculating the RMSD values. The dashed horizontal lines indicate the maximum RMSD value which serves as the criterion to identify correctly folded structures.

reversibly folding into the correct fold, with the backbone atom-positional root-mean-square deviations (RMSD) from the NMR-based experimental model structures being smaller than 0.06 nm. No reliable experimental data, either on ΔG_{fold} as function of temperature and on T_m , or on rates of folding, are available for comparison.

Two results stand out in Table 1. The force field used is able to accurately predict the most stable type of secondary structure as function of peptide composition and solvent type. Secondly, the number of different conformers making up the denatured (unfolded) state is very, very low. This leads to the following conclusions. The denatured state of peptides is much smaller than expected, only a tiny fraction of the possible conformers is populated in the temperature range between 300 and 360 K. This implies that it is possible to simulate the equilibrium between folded and

denatured states, given a force field that accurately describes the characteristics of this limited denatured state. Study of only folded states^[20] will be insufficient to solve the folding problem. The key to solving the protein folding problem lies rather in a good understanding of the denatured state.

Methods

The MD simulations were carried out using the GROMOS96 program^[18] and the force field 43A1.^[19] Aliphatic CH_n groups were treated as united atoms, both in the peptides and the solvents. Periodic boundary conditions were applied in a rectangular or truncated octahedron box. The nonbonded interactions were cutoff at a distance of 1.4 nm, and a Poisson–Boltzmann reaction field force was used to approximate electrostatic interactions beyond the cutoff in water, DMSO, and chloroform. Bond lengths in both peptides and solvents and bond angles in the solvent molecules were kept fixed. The integration time step was 2 fs and systems were kept at constant temperature and pressure (1 atm) by weak coupling to temperature and pressure baths.

The results of the simulations are independent of the various initial structures used because of the relatively long lengths of the simulations. The trajectory structures (at 0.01 ns intervals) of the peptides were clustered into conformations as follows:^[8] The number of neighbors (that is, the number of structures satisfying the similarity criterion) was determined for each trajectory structure, with the criterion of similarity between two structures being the positional RMSD value of their main chain atoms. The structure with the highest number of neighbors was then taken as representing the first (most populated) conformation or cluster of structures. After removing the structures belonging to the first cluster from the trajectory, the procedure is repeated to find the second cluster or conformation, and so on.

Received: September 20, 2000 [Z15837]

- [1] a) H. J. C. Berendsen, *Science* **1998**, *282*, 642–643; b) W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 5897–5899; c) N. D. Socci, J. N. Onuchic, P. G. Wolynes, *Proteins Struct. Funct. Genet.* **1998**, *32*, 136–158; d) C. M. Dobson, M. Karplus, *Curr. Opin. Struct. Biol.* **1999**, *9*, 92–101.
- [2] a) A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, Freeman, New York, **1999**; b) D. J. Brockwell, D. A. Smith, S. E. Radford, *Curr. Opin. Struct. Biol.* **2000**, *10*, 16–25.
- [3] Y. Duan, P. A. Kollman, *Science* **1998**, *282*, 740–744.
- [4] a) S. J. Hagen, W. A. Eaton, *J. Mol. Biol.* **2000**, *297*, 781–789; b) N. A. J. van Nuland, V. Forge, J. Balbach, C. M. Dobson, *Acc. Chem. Res.* **1998**, *31*, 773–780.
- [5] X. Daura, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *J. Mol. Biol.* **1998**, *280*, 925–932.
- [6] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *Angew. Chem.* **1999**, *111*, 249–253; *Angew. Chem. Int. Ed.* **1999**, *38*, 236–240.
- [7] M. Takano, T. Yamato, J. Higo, A. Suyama, K. Nagayama, *J. Am. Chem. Soc.* **1999**, *121*, 605–612.
- [8] X. Daura, W. F. van Gunsteren, A. E. Mark, *Proteins Struct. Funct. Genet.* **1999**, *34*, 269–280.
- [9] B. Erman, K. Dill, *J. Chem. Phys.* **2000**, *112*, 1050–1056.
- [10] Y. Zhou, M. Karplus, *Nature* **1999**, *401*, 400–403.
- [11] M.-H. Hao, H. A. Scheraga, *J. Mol. Biol.* **1998**, *277*, 973–983.
- [12] a) E. Alm, D. Baker, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11 305–11 310; b) R. L. Baldwin, G. D. Rose, *Trends Biochem. Sci.* **1999**, *24*, 26–33.
- [13] a) M. Schaefer, C. Bartels, M. Karplus, *J. Mol. Biol.* **1998**, *284*, 835–848; b) H. Wang, S.-S. Sung, *J. Am. Chem. Soc.* **2000**, *122*, 1999–2009.
- [14] a) E. I. Shakhnovich, *Fold Design* **1998**, *3*, R108–R111; b) D. K. Klimov, D. Thirumalai, *J. Chem. Phys.* **1998**, *109*, 4119–4125.
- [15] F. B. Sheinerman, C. L. Brooks III, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1562–1567.

- [16] G. M. Crippen, Y. Z. Ohkubo, *Proteins Struct. Funct. Genet.* **1998**, *32*, 425–437.
- [17] a) J. Moult, T. Hubbard, K. Fidelis, J. T. Pederson, *Proteins Struct. Funct. Genet.* **1999**, *37*, 2–6, Supplement: Third Meeting on the Critical Assessment of Techniques for Protein Structure Prediction; b) E. Lacroix, T. Kortemme, M. López de la Paz, L. Serrano, *Curr. Opin. Struct. Biol.* **1999**, *9*, 487–493.
- [18] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, W. F. van Gunsteren, *J. Phys. Chem. A* **1999**, *103*, 3596–3607.
- [19] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Vdf Hochschulverlag AG, Zürich, **1996**.
- [20] D. Baker, *Nature* **2000**, *405*, 39–42.
- [21] C. Peter, X. Daura, W. F. van Gunsteren, *J. Am. Chem. Soc.* **2000**, *122*, 7461–7466.
- [22] X. Daura, K. Gademann, H. Schäfer, B. Jaun, D. Seebach, W. F. van Gunsteren, *J. Am. Chem. Soc.*, submitted.
- [23] R. Bürgi, X. Daura, A. Mark, M. Bellanda, S. Mammi, E. Peggion, W. F. van Gunsteren, *J. Peptide Res.*, in press.
- [24] C. Santiveri, M. A. Jiménez, M. Rico, W. F. van Gunsteren, X. Daura, personal communication.

Electronic Insight into an Antithrombotic Agent by High-Resolution X-Ray Crystallography**

Ralf Flaig, Tibor Koritsánszky, Rainer Soyka, Ludger Häming, and Peter Luger*

In medicinal chemistry it is generally accepted that steric and electronic properties play a dominant role in drug–target recognition processes. Molecular activity, similarity, and recognition, which are ambiguously defined properties, are therefore most often screened in terms of their relationship to a definite and readily accessible property, the molecular structure. However, all of these properties can be deduced unambiguously from the distribution of the electronic charge,

[*] Prof. Dr. P. Luger, Dr. R. Flaig
Institut für Chemie/Kristallographie
der Freien Universität Berlin
Takustrasse 6, 14195 Berlin (Germany)
Fax: (+49) 30-838-53464
E-mail: luger@chemie.fu-berlin.de
Prof. Dr. T. Koritsánszky
Department of Chemistry, University of the Witwatersrand
Private Bag 3, WITS 2050, Johannesburg (South Africa)
Dr. R. Soyka
Research and Development, Boehringer Ingelheim Pharma KG
Birkendorfer Strasse 65, 88397 Biberach (Germany)
Dr. L. Häming
Bruker Analytical X-ray Systems
Östliche Rheinbrückenstrasse 50, 76187 Karlsruhe (Germany)

[**] This work was funded by the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF), Bonn (Germany) (grant no. 05 SM8KEA0) and by the Fonds der Chemischen Industrie, Frankfurt (Germany).