The Influence of Side Chains on the Properties of Simple Model Biopolymers: Monte Carlo Simulations

*Mark Hofstetter, Gerhard Zifferer**

Institut für Physikalische Chemie der Universität Wien, Währinger Str. 42, A-1090 Wien, Austria

Summary: A model for the simulation of proteins is introduced which is based on a new set of bond vectors and a new method for modeling the side chains of proteins. The drawbacks of united atoms models are summarized and the motivation for this new model is given. Some preliminary results are shown which shall demonstrate the suitability of the model proposed.

Introduction

In recent years a lot of simple models have been proposed for the simulation of biopolymers, especially proteins, by use of Monte Carlo techniques. Most of them aimed at learning more about the *principle* properties of protein folding (disregarding details of *specific* proteins) and, therefore, used simple lattice models (e.g. cubic[1], tetrahedral[2,3] and high coordinative lattices[4,5]). Usually, the properties of the side chains were either neglected or incorporated in an united atom approach.

This work describes the development of an innovative model based of a new set of bond vectors (located in a high coordinate lattice) for the representation of the backbone combined with a new approach to portray side chains and their geometrical properties (independent of the underlying lattice). Thus, being still a model in the sense of an equivalent chain some aspects of an atomistic presentation are included.

Motivation

Although important features of proteins may be simulated by simple united atoms approaches, models of polypeptide chains without explicit side chains have several disadvantages:

(1) a straightforward definition of handedness and bond angles between the different amino acid segments along the chain is rather difficult and to some extent arbitrary; thus, geometrical conditions have to be defined over a few segments and not only based on a single unit representing an amino acid.

 CCC 1022-1360/00/$ 17.50+.50/0

(2) a model using only the backbone without side chains is geometrically unsatisfactory as the side chains occupy most of the space in a protein, so neglecting them leaves most of the space free.

(3) polar/hydrophobic forces may lead the side chains to point toward the core of the protein or outside toward the solvent; thus an amino acid is anisotropic from a geometrical point of view, a fact which is also disregarded if no explicit side chains are considered.

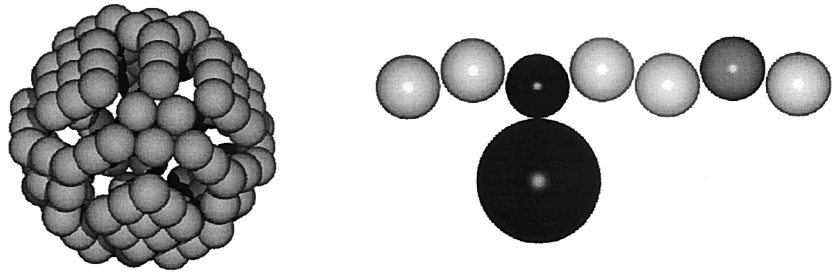

Figure 1. Left: all possible vectors of the model originate from the centre of the object and the tips of the vectors are located in the centre of the surrounding spheres. Right: side chains are attached to every third segment of the chain and are simulated as equivalence spheres (not restricted to lattice positions), e.g. glutamin and glycin (no side chain) as shown above. Visualization by use of POV-ray (http://www.povray.org).

Polypeptide model

Backbone

As a consequence of the points discussed above, we decided against the classical united atoms approach representing the whole amino acid (or at least the backbone element) as a single segment. Instead, the backbone of an amino acid is constituted of three segments, the vectors between the segments chosen from a set of vectors belonging to the following classes, depicted in Figure 1:

[4,1,1], [4,1,0], [3,3,1], [3,3,0], [3,2,2], [4,0,0]

This leads to 114 different bond vectors. Their maximum length differences are about 10% which is in good accordance with the distribution in polypeptides. Additionally, only sequences of vectors were allowed which led to a bond angle larger than 90° and consecutive vectors of the same type (leading to an angle 180°) were forbidden either.

The choice of this model has the advantage that the continuum behavior is better approximated than in low coordinate lattices while integer arithmetic is still applicable to this model.

Side chains

The side chains (being attached to every third backbone segment) are simulated by *off lattice* spheres: for each amino acid the positions of its atoms were extracted from the Protein Data Bank[6] and the distance between the C_α atom and the center of mass of all non-hydrogen atoms belonging to the side chain was calculated; in this way a sphere of equivalent volume was estimated (glycin clearly does not have a side chain) mapping the average bond length of polypeptides onto the average bond length of the model and considering the thickness of the backbone, see Figure 1. The vector **d** pointing from the linking backbone segment to the center of the sphere is located in the plane halving the angle between the two adjacent bond vectors **a** and **b** while any orientation with respect to the plane defined by theses two vectors may be given. In this way handedness, i.e. whether the side chain is "above" or "below" the plane defined by the vectors **a** and **b** may be governed. In the present investigation the vector **d** simply is located in the plane defined by **a** and **b** pointing in that direction which takes the larger angle with respect to **a** or **b**, respectively; investigation of the influence of the handedness on certain global parameters is deferred to a forthcoming publication[7].

Relaxation mechanisms

For the simulation a dynamic Monte Carlo method was used. Three different sets of moves have been employed, i.e. (i) single spike moves which are comparable to the L-flip in the cubic lattice[8], (ii) pivot-like[9] moves where a whole subchain is transformed into a new position and (iii) the replacement of two bonds in the chain which may be regarded as some sort of internal reptation[10] algorithm. Acceptance of a Monte Carlo step was controlled by use of the Metropolis Rosenbluth criterion[11], i.e. the new configuration is accepted if the energy of the new configuration is smaller than or equal to the energy of the old one; if not, it is accepted with the probability given by the Boltzmann factor of the energy-difference only.

Potential

The method (i.e. the underlying program) is flexible concerning the potential used. Presently, only the geometric effect is implemented: Backbone segments are treated as hard spheres whereas the "side chains" are attributed a (strongly repulsive) soft core within a sphere around their centre, the radius being 80% of the calculated equivalence radius. Therefore, overlaps between backbone segments are strictly forbidden while the soft cores of the side chains may intersect with each other as well as with backbone

segments, however, with an energy penalty imposed by the Metropolis-Rosenbluth criterion, the energy (in multiples of kT) of a configuration in the current version simply being given by the number of intersections (provided that the backbone segments are self avoiding).

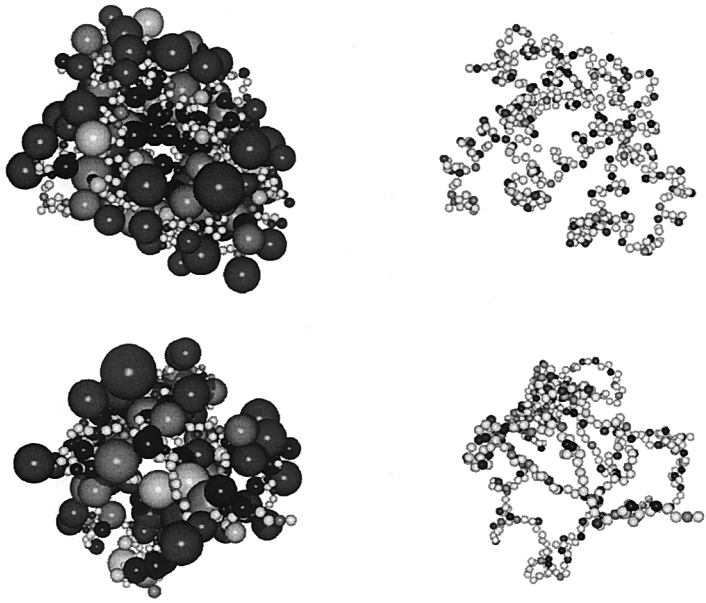

Figure 2. Model of myoglobin (the backbone of each amino acid represented by three segments) including the equivalent sphere of the side chains (left) and omitting them (right). The upper row represents crystallographic data given in the 1mbd.pdb file[6] and the lower row a single representative configuration obtained by simulation; visualization by use of POV-ray (http://www.povray.org).

Preliminary results

The suitability of the model was tested with the help of the geometrical properties of the myoglobin protein. The chain length of the model chain as well as the sequence and volume of the side chains were taken according to this protein the coordinates taken from the 1mbd.pdb file given in the Protein Data Bank[6]. In Figure 2 the myoglobin structure is shown as well as a representative snapshot of a coil which was obtained after a certain number of relaxation trials (large enough to ensure equilibrium conformations) starting from an arbitrary rod-like configuration. Obviously, the relaxation mechanisms work and considering the geometric effect only we come up with a protein-like structure. Admittedly, as no specific interactions between amino acids are

contained in the model yet, the rough resemblance with the myoglobin structure probably is a pure coincidence. Anyway, the bulky sidechains give a realistic picture of the space occupied by the amino acids and make the intramolecular distances between segments more realistic. As a matter of fact, the most important and necessary step toward refinement of the model will be the introduction of a further soft shell surrounding the repulsive core of the side chain in order to manage repusive and/or attractive interactions between the groups characteristic of the type of amino acids coming into contact.

Conclusions

The model introduced in this communication appears to by greatly promising for the study of protein folding, as (i) the backbone is quite realistic, (ii) a straightforward definition of handedness, excluded volume, bond angles and torsional angles etc. is possible at least in principle and (iii) a simple, yet realistic representation of side chains is utilized.

After introduction of the refinements mentioned above, formation of secondary – and later on to tertiary – structures should be within reach.

Acknowledgment: We are greatly indebted to the Vienna University Computer Center for making available a lot of computer resources.

---

[1] A. R. Dinner, M. Karplus, *J. Chem. Phys. B* **1999**, *103*, 7976.
[2] J. Skolnick, A. Kolinski, *Annu. Rev. Phys. Chem.* **1989**, *40*, 207.
[3] G. Zifferer, M. Hofstetter, *J. Chem. Phys.* **2001**, *115*, 6236.
[4] A. Kolinski, J. Skolnick, *J. Chem. Phys.* **1997**, *107*, 953.
[5] P. Romiszowski, A. Sikorski, *Physica A* **1999**, *273*, 190.
[6] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne: The Protein Data Bank. *Nucleic Acids Research*, **2000**, *28*, 235-242.
[7] G. Zifferer, M. Hofstetter, in preparation
[8] P. H. Verdier, W. H. Stockmayer, *J. Chem. Phys.* **1962**, *36*, 227.
[9] M. Lal, *Molecular Physics* **1969**, *17*, 57; N. Madras, A. D. Sokal, *J. Stat. Phys.* **1988**, *50*, 109; G. Zifferer, *Macromolecules* **1990**, *23*, 3166.
[10] F. T. Wall. F. Mandel, *J. Chem. Phys.* **1975**, *63*, 4592.
[11] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *J. Chem. Phys.* **1953**, *21*, 1087.