

Chapter 2

Folding

2.1 An off-lattice Monte Carlo method for protein folding

Of the numerous projects of the Knapp group, the development of an off-lattice Monte Carlo (MC) method for the long term dynamics of proteins has the oldest tradition. Ernst-Walter Knapp himself had already developed the basic ideas as he founded his group at the Freie Universität Berlin (Knapp & Irgens-Defregger, 1991; Knapp, 1992; Knapp & Irgens-Defregger, 1993). The core of the method is the protein model (section 2.1.1) together with a suitable energy function (section 2.1.2) and the so-called window move (WM, section 2.1.3).

2.1.1 The protein model

The protein model is so far only a model of the protein backbone. In the model, the amide planes are rigid and the degrees of freedom are reduced to the two torsion angles ϕ and ψ (Figure 2.1). For more details, see Hoffmann (1996), Sartori (1997).

2.1.2 Tuning the energy function

Fredo Sartori developed in our group a suitable energy function for this protein model (Sartori, 1997; Sartori *et al.*, 1998). By performing MD simulations of a protein with constraints fixing bond angles and bond lengths, it was found that the protein remained relatively immobile (van Gunsteren & Karplus, 1981). Hence, it became evident that an energy function that is tuned for fully-flexible all-atom molecular models cannot directly be applied to our protein model. With the exception of small molecules, as for instance water, small variations in particular of bond angles help to avoid atomic overlaps of atom pairs, which otherwise could hinder changes of torsion angles. The dominant contribution of atomic overlaps in rigid geometry models of the protein backbone is due to the Lennard-Jones interactions of the hydrogen and oxygen atoms in neighboring amide planes. With these interactions, the rotational barrier with respect to the dihedral angles ϕ and ψ becomes so large that even at temperatures as high as 800 K no full rotation is possible and the regime of positive ϕ angles cannot be reached (Sartori *et al.*, 1998). The mobility of the rigid protein backbone model can be regained, if the interaction between two neighboring amide planes is described by a suitable two-dimensional (ϕ, ψ) torsion energy landscape. This energy landscape can be generated from the probability distribution of the (ϕ, ψ) torsion angles obtained from a conventional MD simulation of suitable model dipeptides (Sartori *et al.*, 1998). The resulting potential of mean force is temperature dependent. For the case of alanine the suitable model dipeptide is (*N*-acetyl-alanyl)-methylamide, for the case of glycine, it is correspondingly (*N*-acetyl-glycyl)-methylamide. The resulting (ϕ, ψ) torsion potentials for alanine and glycine are visualized in Figure 2.2.

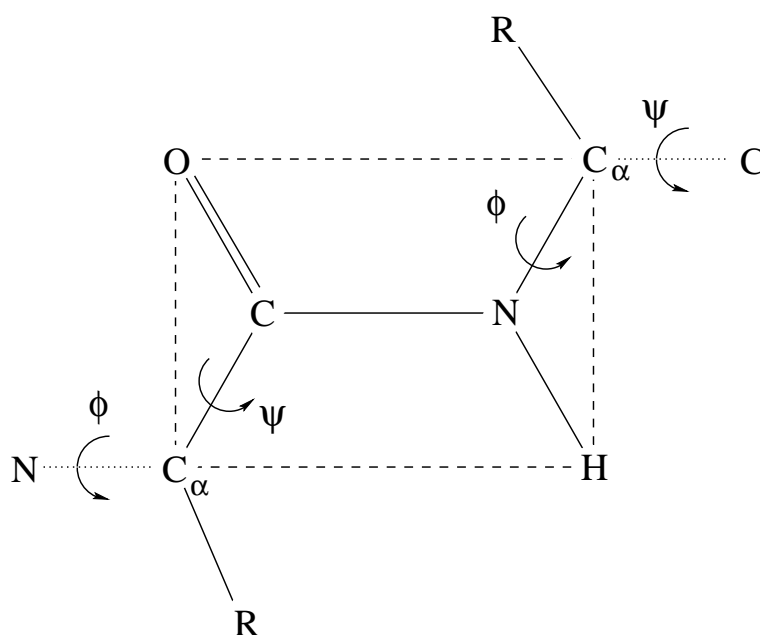


Figure 2.1: Schematic representation of a single amide plane of the protein model. The only degrees of freedom are the torsion angles ϕ and ψ . All other torsion angles, bond angles and bond lengths are fixed to idealized values. R is H for glycine and an united CH_3 atom for alanine. More complex sidechains have additional degrees of freedom and are treated later in this chapter.

The (ϕ, ψ) torsion energy landscape replaces all energy terms where only atoms of two neighboring amide planes are involved. These are all non-bonded interactions of atom pairs belonging to the same or to neighboring amide planes as well as the torsion potentials used in the energy function for a conventional MD simulations. The energy terms accounting for bond stretching and bond angle bending do not occur for molecular models with rigid bond geometry anyway. All other atom pairs are separated by at least four torsional degrees of freedom in the polypeptide backbone. Therefore, we assume these atoms can move enough independently, such that their non-bonded interactions can be taken unchanged from an energy function that is suitable for conventional MD simulations.

Adjusting the energy function in this way to compensate for the lack of flexibility in the rigid polypeptide model, the backbone mobility of polyalanine is only slightly increased. The rotational barrier for the backbone torsion angle ψ is still a factor of ten to high and in the angle ϕ a full rotation is not possible yet. But what is even worse, at a given temperature an α -helical structure of polyalanine with rigid bond geometry is much more stable as compared to the one with flexible bond geometry (Sartori *et al.*, 1998). The reason for this deviation is the absence of bond angle oscillations of hydrogen atoms described by molecular models with rigid bond geometry. These oscillations can weaken hydrogen bonds. For the same reason, hydrogen bonds formed with deuterium are generally stronger than those formed with protons. This is a true quantum effect appearing even at room temperature (Gutowsky *et al.*, 1985; McDowell & Buckingham, 1991). Since vibrational frequencies involving hydrogen atoms are very high, even at room temperature only the vibrational ground state is occupied. Due to the difference in mass the amplitude of the zero point oscillations is a factor of $\sqrt{2}$ smaller for deuterons than for protons, which explains the effect.

To overcome this problem, the hydrogen bonds of the protein backbone are weakened by introduction of a correction value Δr in the term for the electrostatic interaction energy:

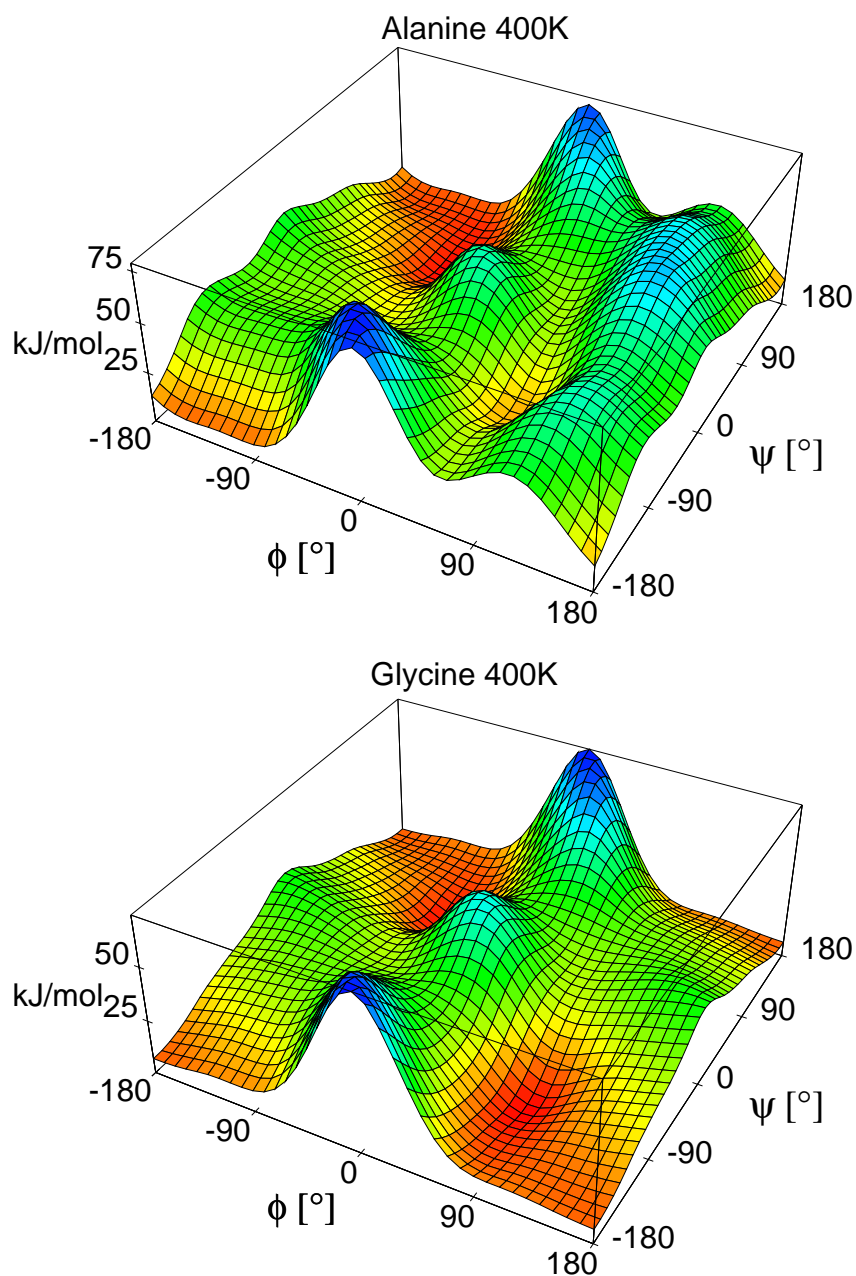


Figure 2.2: (ϕ, ψ) torsion potentials for alanine and glycine, derived from an MD simulation using the CHARMM22 parameter set at 400 K.

$$E_{ij} = \frac{q_i q_j}{|\vec{r}_{ij}| + \Delta r} \quad (2.1)$$

Typical values for Δr are in the range from 0.18 Å to 0.27 Å, weakening hydrogen bonds by 10 to 15%. The optimal value depends on the temperature. For more details of deriving this correction see Sartori (1997), Sartori *et al.* (1998), Rabenstein *et al.* (1999).

2.1.3 The window move

The obvious MC move applicable to our protein model is the random change of one randomly chosen torsion angle. This move is called simple move (SM) and generates large global conformational changes. In globular protein, SMs often result in structures with atomic clashes, which are energetically unacceptable (Hoffmann & Knapp, 1996a). An MC sampling using exclusively SMs is therefore extremely inefficient.

This problem is avoided by the WM, where cooperative rotations are applied to backbone torsion angles in a window of subsequent amide planes, such that the polypeptide conformation outside of the window remains invariant (Figure 2.3). A single WM generates a local conformational change

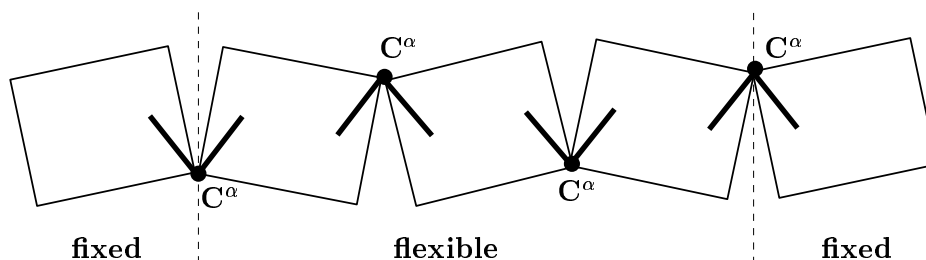


Figure 2.3: Sketch of a window of three amide planes within the protein backbone. The torsion axes corresponding to the eight degrees of freedom are indicated by solid bars (Hoffmann & Knapp, 1996b).

only. Global conformational changes can be obtained by applying WMs many times. These changes evolve the conformation gradually during a MC simulation. Thus, a trajectory of MC moves can be interpreted as a real time evolution. Since for compact molecular structures local MC moves can avoid atomic overlaps better than global MC moves they are much more efficient for the generation of large conformational changes. The elementary time step for MC dynamics is one scan of MC moves, where on the average the MC moves are applied once to each possible position at the protein backbone.

Window MC moves were first applied to generate conformational changes of molecular systems by Go and Scheraga (1970). These authors used window moves to prove that a cyclic tripeptide cannot exist. Since then window moves were used in different applications to model structures of biological macromolecules (Braun, 1987; Sklenar *et al.*, 1986; Palmer & Scheraga, 1991; Hubbard *et al.*, 1994). The works of Knapp and Irgens-Defregger (1991), Knapp (1992), Knapp and Irgens-Defregger (1993) were the first attempts to use this MC algorithm for the simulation of polypeptide dynamics. Using the WM for the simulation of polymers, it was found that the torsion angles of the generated molecular conformations are biased unless one applies a correction with suitable Jacobians (Dodd *et al.*, 1993). This behavior was also observed and fixed in our group by Daniel Hoffmann. He applied these corrections then also to the simulation of polypeptide dynamics (Hoffmann & Knapp,

1996b, 1996a, 1997). (In these references, more details about the method to compute the window moves are given.) He was able to fold a helix–turn–helix model protein (Hoffmann & Knapp, 1997).

Recently, Wu and Deem (1999) found an analytical solution for the WM equivalent to our own, numerical solution.

2.1.4 The CAMLAB program

Daniel Hoffmann left our group after finishing his PhD. He implemented the WM with its Jacobian corrections among a lot of other features in his new program CAMLAB (Hoffmann *et al.*, 1998), which is more general and efficient than the implementations done before. Hoffmann provided us with the source code of CAMLAB, so that we could modify the program for our own applications. It was one of my tasks to fit CAMLAB to our needs. The main extension at this stage was the implementation of the energy function developed by Fredo Sartori within CAMLAB. However, I did also several other modification, which are documented in appendix D. To distinguish our modified program from the original CAMLAB program, it is called CAMLAB++.

Our first application of CAMLAB++ was to fold not only a helix–turn–helix motif as before, but also a β -hairpin. This application is reported in the next section 2.2. To account also for the influence of the solvent, Björn Kleier implemented the ACS potential of Schaefer *et al.* (1998) (see chapter 1) in C. Then we included this implementation in the CAMLAB++ program. First applications of CAMLAB++ to peptides in solution are presented in section 2.3.

2.2 Folding *in vacuo*

2.2.1 To be investigated: the AGA and VGV model peptides

After folding α -helices (Sartori, 1997) and a helix–turn–helix motif (Hoffmann & Knapp, 1997), it is obvious to ask whether also the folding of a β -structure is possible. The first problem to solve is to find a suitable model protein. For doing this, it might help to have a look on the relative propensities of amino acids for a certain secondary structure (Table 2.1).

Since polyalanine indeed readily folds to an α -helix, one might assume that polyvaline will fold to something like a long, single β -strand. In the MC simulation (results not shown), polyvaline adopts for a short while a β -strand like conformation (as determined by the location of the individual residues in the Ramachandran plot, see section 2.2.2.1). However, finally it folds also to an α -helix, although it needs a longer sampling time than polyalanine to reach this structure. This is not too surprising, since in a single β -strand in vacuum there are no hydrogen bonds at all, whereas in an α -helix the optimal number of hydrogen bonds is established. This situation could be relaxed by introducing a turn in the middle of a polyvaline peptide. Then a β -hairpin, *i. e.* the smallest possible anti-parallel β -sheet, would emerge, where about half as much hydrogen bonds are possible as in an α -helix. To introduce the turn, we take the residue with the largest turn propensity, which is glycine. So our model peptide sequence for folding a β -hairpin is VVVVVVGGVVVVVV (called the VGV model peptide). For comparison, we simulate also a peptide with the sequence AAAAAAAGGAAAAAA (called the AGA model peptide), which should fold to a helix–turn–helix motif. Both peptides are capped to avoid charged end groups: The N-terminus is acylated, and the C-terminus is (*N*-methyl)-amidated.

2.2.2 Methods

2.2.2.1 Determination of secondary structure

The determination of the secondary structure from the atomic coordinates of a protein depends on the definition and is by far not trivial. Numerous methods are discussed in the literature. To mention

Table 2.1: Amino acid propensities for helix, strand and turn according to (Williams *et al.*, 1987). The propensities of Ala, Val, and Gly for helix, strand, and turn are highlighted by bold letters.

amino acid	helix	strand	turn
Glu	1.59	0.52	1.01
Ala	1.41	0.72	0.82
Leu	1.34	1.22	0.57
Met	1.30	1.14	0.52
Gln	1.27	0.98	0.84
Lys	1.23	0.69	1.07
Arg	1.21	0.84	0.90
His	1.05	0.80	0.81
Val	0.98	1.87	0.41
Ile	1.09	1.67	0.47
Tyr	0.74	1.45	0.76
Cys	0.66	1.40	0.54
Trp	1.02	1.35	0.65
Phe	1.16	1.33	0.59
Thr	0.76	1.17	0.90
Gly	0.43	0.58	1.77
Asn	0.76	0.48	1.34
Pro	0.34	0.31	1.32
Ser	0.57	0.96	1.22
Asp	0.99	0.39	1.24

only a classical and a more recent example: the DSSP program by Kabsch and Sander (1983), whose method is used by the RASMOL visualization program (Sayle & Bissell, 1992), and the STRIDE program by Frishman and Argos (1995), whose method is used by the VMD visualization program (Humphrey *et al.*, 1996).

In the present work, we use two methods. The first is based on the Ramachandran plot (Ramachandran *et al.*, 1963). We divide the Ramachandran plot into different regions (Figure 2.4) that are assigned to specific secondary structures (Robson & Garnier, 1988; Ullmann, 1995; Sartori, 1997). The secondary structure of a specific residue is determined by the (ϕ, ψ) location of its torsional angles in the Ramachandran plot. We denote this method by the abbreviation RP.

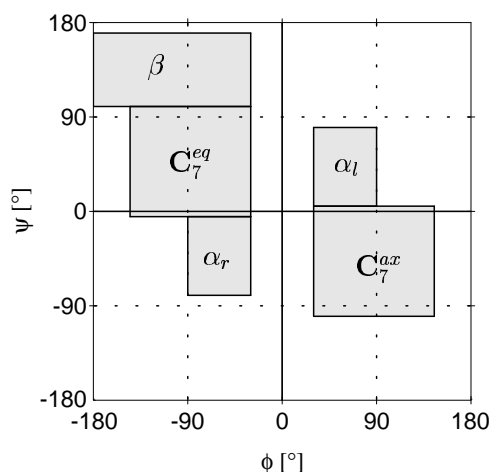


Figure 2.4: Areas of secondary structure motifs in the Ramachandran plot (Sartori, 1997).

The second method used here is the afore mentioned DSSP method by Kabsch and Sander (1983), which is based on the topology of the hydrogen bond pattern. This method is more complex than the RP method, but especially the assignment of β -sheets is more in agreement with our intuitive understanding using DSSP than RP. It is well possible for a peptide to have all (ϕ, ψ) torsion angles in the β region of the Ramachandran plot (Figure 2.4), but the overall structure does not resemble a β -sheet like structure at all (*e.g.* a single β -strand without any hydrogen bonds). In the opposite case, a β -sheet could be distorted such that its (ϕ, ψ) torsion angles are outside of the β region of the Ramachandran plot without losing the overall β -sheet like structural features. Since DSSP analyzes the hydrogen bond pattern, it would identify the latter case, but not the former, as a β -sheet structure.

2.2.2.2 Treatment of sidechains

So far, side chains are not included in our protein model. To include side chains in the spirit of the existing model, one would fix the bond length and bond angles also of the side chains and thus reduce their degrees of freedom to the torsion angles χ_n . Then, a kind of multi-dimensional effective torsion potential $(\phi, \psi, \chi_1, \dots, \chi_n)$ had to be generated. We wanted to avoid this task and therefore used only the already existing two-dimensional (ϕ, ψ) torsion potentials. For the side chains, we used the conventional one-dimensional torsion potentials from the CHARMM force field. To avoid the problems of the fixed geometry, we took into account the complete CHARMM force field (including terms for bond length and bond angles) and included additional Cartesian moves (CM) for the side chain atoms, *i.e.* we modified the side chain conformation not only by SMs, but also by small Cartesian displacements. This effectively removes the fixed geometry and leads to the same sampling behavior as a fully flexible model (Hoffmann *et al.*, 1998). This method could also be used for

the whole molecule, thereby removing the need for the two-dimensional torsion potentials at all, but this increases the sampling time to reach a properly folded structure by about an order of magnitude compared to simulations using fixed geometry and two-dimensional torsion potentials (results not shown). However, we think that the increase of folding efficiency gained by the reduced number of degrees of freedom is much more effective for the backbone than for the side chains, so that we can apply the fully flexible model to the side chains without losing too much.

In the case of the AGA model peptide, there are no sidechains at all. Alanines were treated by the (ϕ, ψ) torsion potential for alanine and glycine by the (ϕ, ψ) torsion potential for glycine. In the case of the VGV model peptide, the glycines were treated as before. The valines were also treated by the (ϕ, ψ) torsion potential for alanine, but whenever a valine residue was involved in a WM, its χ torsion angle of its side chain was modified at the same time by an SM, and its two C_γ atoms were displaced by a CM.

2.2.2.3 Mixing window moves and simple moves

Applying exclusively WMs is much more efficient for peptide folding than applying exclusively SMs (Hoffmann, 1996; Hoffmann & Knapp, 1996b, 1996a, 1997). However, we found that applying both, WMs and SMs, in a ration of 10:1, the correctly folded structure is reached much earlier, although SMs are much more expensive in computer time. This can be explained by the need for both, local and global conformational changes. If only SMs are used, it is extremely difficult for a local structure like an α -helix to evolve. However, if only WMs are used, a global concerted movement of a lot of atoms as it is necessary for creating the turn in a helix–turn–helix motif or a β -hairpin can only happen very laborious by a lot of WMs, where the global conformational change propagates slowly through the backbone of the peptide.

2.2.2.4 Parameters of the MC simulation

The elementary time step of the MC dynamics corresponding roughly to the amount of CPU time needed for a single step of time evolution with conventional MD simulation is an MC scan. In an MC scan, the moves are applied on the average one time at each possible position. The two oligopeptides considered consist of 16 residues involving 17 amide planes, where two involve the capped end groups. The MC scan used in the present application consisted of 15 window moves changing the conformation of three consecutive amide planes and 1.5 simple moves. Window moves for the first and last possible window in the polypeptide sequence can be applied without constraints. For a simple move both ϕ and ψ torsion angles at one C_α atom were changed simultaneously. To avoid atomic overlaps and to obtain a smooth and continuous time evolution, the interval for angular changes was restricted to $[-5, +5]$ degrees for simple moves, whereas for window moves no restrictions were imposed. The CM were restricted to a maximal displacement per move of 0.02 Å.

Using the program CAMLAB++, we generated six trajectories of four million scans each for the VGV model peptide, and three trajectories of one million scans each for the AGA model peptide. All trajectories were generated at 400 K using the CHARMM19 force field (Brooks *et al.*, 1983), as far as the energy terms were not included in the (ϕ, ψ) torsion potentials. The torsion potentials were generated by an MD simulation using umbrella sampling (Sartori *et al.*, 1998) and the all-atom CHARMM22 force field (MacKerell *et al.*, 1992). The Δr value for the hydrogen bond correction was 0.18 Å. The starting conformations for both model peptides was the extended C_7 -equatorial conformation with torsion angles $\phi = -90$ and $\psi = +90$. This extended start conformation was chosen because it is a low energy conformation just half way between a β -strand and an α -helix conformation.

2.2.3 Results and Discussion

2.2.3.1 Investigating the trajectories

Figures 2.5 and 2.6 show snapshots from those trajectories that lead to the lowest energy conformations. However, the time evolution of all three (AGA) or six (VGV) trajectories is essentially similar, albeit in three of the six VGV trajectories the folding into the β -hairpin conformation was incomplete, indicating that this conformation is only marginally stable under the simulation conditions used.

Due to the simple moves, a bend of the AGA model peptide can form at the two glycines already after 5000 MC scans. After a few ten thousand MC scans, the formation of secondary α -helical structure can be observed. After a few hundred thousand MC scans, the helix–turn–helix motif is completed. In order to generate a single α -helix of 18-alanine with conventional MD simulation using the corresponding energy function and simulation conditions suitable for conventional MD, several ten nanoseconds corresponding to more than 10^7 time steps are needed (Sartori *et al.*, 1998). The time required to generate a helix–turn–helix conformation with conventional MD simulation methods would be much larger.

It turned out that the formation of the VGV β -hairpin took about ten times longer than the formation of the alternative helix–turn–helix motif of the AGA polypeptide (see Figures 2.5 and 2.6). This may be simply due to the fact that for the VGV model peptide the β -hairpin conformation is only marginally stable as opposed to the AGA model peptide, whose helix–turn–helix conformation is very stable. But since during the time evolution of the β -hairpin conformation no significant content of α -helical structures could be observed, this conclusion is not mandatory. Another explanation would be that the free energy funnel for the formation of the β -hairpin conformation is relatively flat so that it takes more time for the molecular system to find its way into the appropriate folded structure. This would also be consistent with experiments where it was found that the formation of a β -hairpin conformations takes more time than the formation of α -helices (Muñoz *et al.*, 1997).

In Figure 2.6, snapshots of one trajectory evolving into the β -hairpin conformation are shown. It takes also about ten times longer than for the sequence forming the helix–turn–helix motif before a bend forms at the center where the glycines are situated. After 10^5 MC scans the signature of β -strand conformations appears. But it takes up to $4 \cdot 10^6$ MC scans to form a complete β -hairpin conformation. During this time evolution, patches of β -strand conformations appear and decompose again at different positions in the sequence. This hesitation in forming a β -sheet conformation also indicates that the free energy funnel is relatively flat slowing down the formation of a β -hairpin.

The hydrogen donor and acceptor groups of an α -helix are saturated intramolecular except for the last winding at both ends. For a β -sheet motif, only half of the molecular groups involved in hydrogen bonds are forming these with other parts of the β -sheet. In a detailed molecular model including solvent effects, the remaining molecular groups form hydrogen bonds with water molecules from the solvent or they interact with mirror charges of opposite sign from a reaction field, which mimics the electrostatic interaction with the solvent. In the present application, no solvent model is included. Under these circumstances a polypeptide conformation where the hydrogen donor and acceptor groups are saturated intramolecular will generally be preferred. This was corroborated by the fact that the energies of the α -helical conformations had generally a lower energy than the β -sheet structures. Hence, it was surprising to observe, that in spite of this disadvantage the VGV model peptide was capable to form a β -hairpin.

For a β -hairpin in vacuum, only half the total number of possible hydrogen bonds of the backbone can be formed as compared to an α -helix, where nearly all possible hydrogen bonds, *i. e.* one per residue, are formed. Assuming an energy value of 12.6 kJ/mol for the formation of an hydrogen bond, one has to conclude that in vacuum an α -helical structure is 6.3 kJ/mol per residue more favorable than a β -hairpin. An analysis of the MC dynamics of valine showed that the side chain of valine was about two times more mobile in the β -hairpin than in the α -helix conformation. This difference

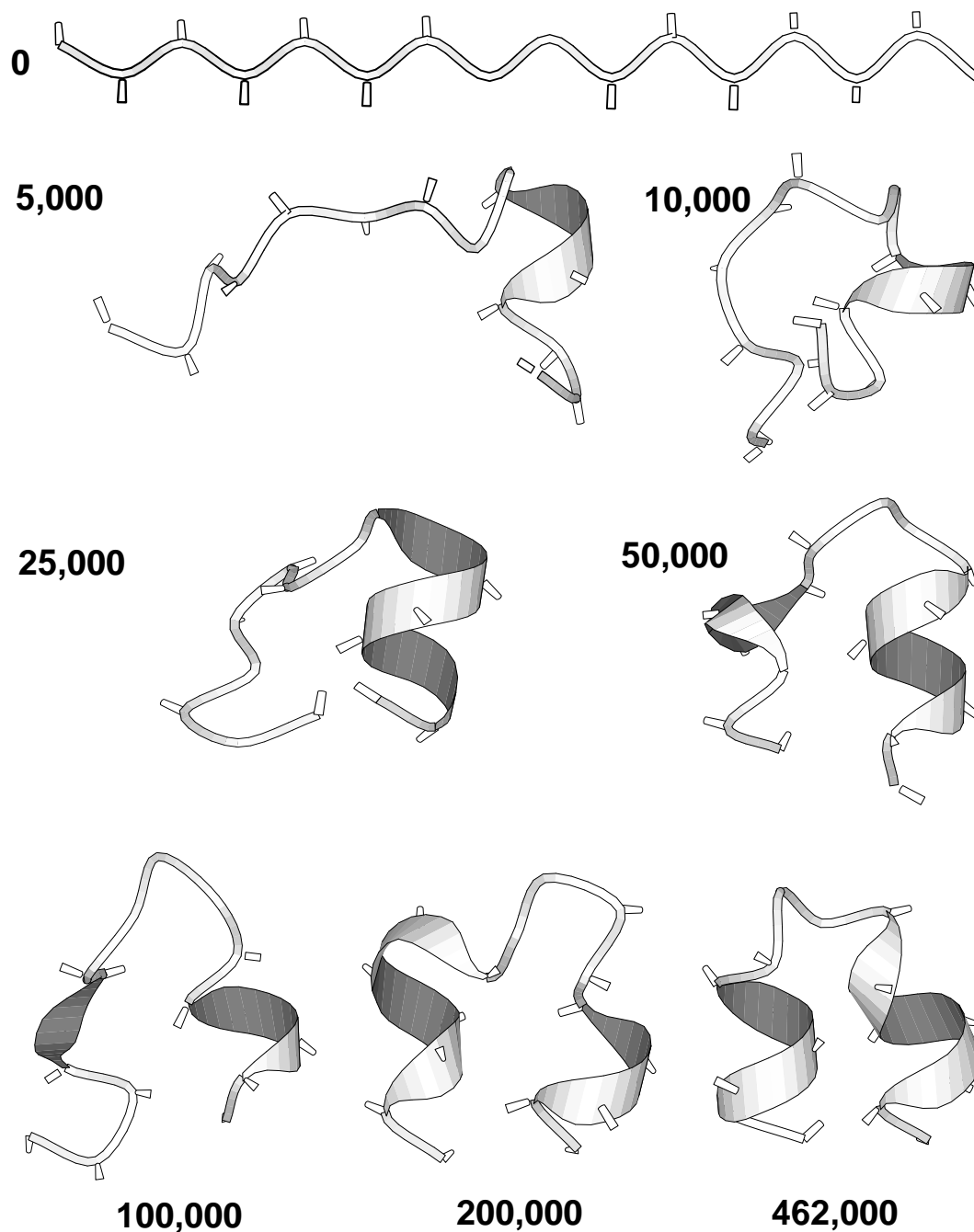


Figure 2.5: Snapshots from the MC dynamics of the AGA model peptide evolving into a helix–turn–helix motif. The initial extended conformation (0 MC scans) is C_7 -equatorial with torsion angle values of $\phi = -90$, $\psi = +90$. The arbitrary time unit is given by the number of MC scans. One MC scan corresponds roughly to the amount of CPU time needed for one time step of conventional MD simulation. For the definition of MC scans see text. The last snapshot at 462000 MC scans displays the conformation with the smallest value of the potential energy from the whole trajectory of 10^6 MC scans. The criterion for an α -helical structure was taken from the DSSP method.

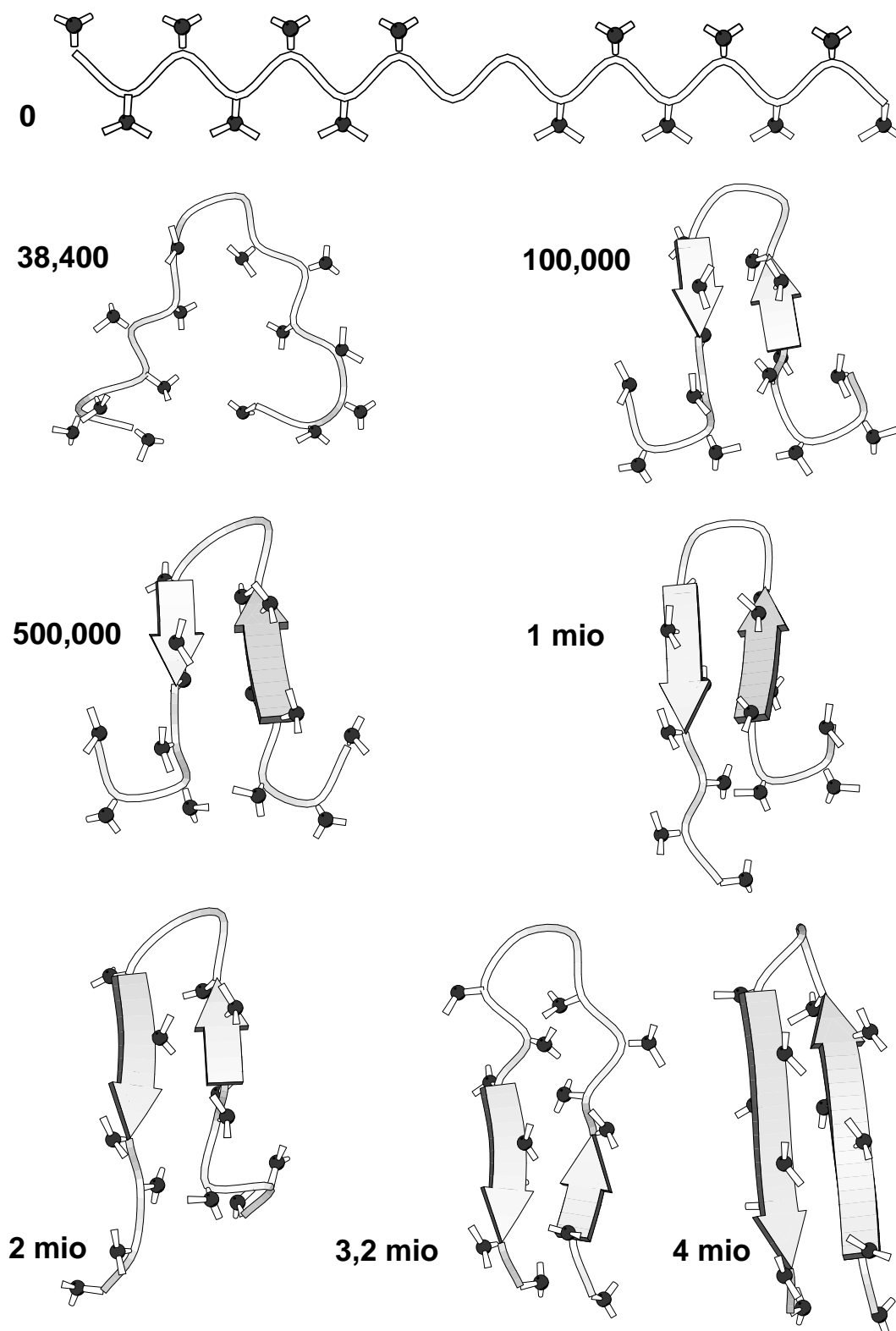


Figure 2.6: Snapshots from the MC dynamics of the VGV model peptide evolving into a β -hairpin. The initial extended conformation is the same as for the AGA model peptide shown in the previous figure. The second snapshot at 38,400 MC scans displays the conformation, where the formation of the bend at the two central glycines just appeared. Also here the criterion for assigning secondary structure was taken from the DSSP method.

in entropy corresponds at the temperature of 400K, where the simulation was performed, roughly to 2 kJ/mol of free energy per residue. That is not enough to explain the marginal preference of valine for a β -hairpin as compared to alanine which prefers the α -helical structure. A more detailed analysis showed that also the fluctuations of the (ϕ, ψ) -torsion angles of the polypeptide backbone are about a factor of two larger in the β -hairpin than in the α -helix. A rough estimate of the corresponding entropy contribution would yield another 4 kJ/mol per residue free energy favoring the β -hairpin. This would be just enough to explain the results.

2.2.3.2 Loosing of timescales

By comparing the sequence of folding events in Figures 2.5 and 2.6 with the earlier simulation of a helix–turn–helix motif folding by Hoffmann and Knapp (1997), an interesting phenomenon springs into mind: In the earlier simulation, at first local structural elements, *i. e.* the α -helical parts, develop in agreement with the diffusion–collision model of Karplus and Weaver (1994), before the global turn structure is formed. In the trajectories of the present study, the non-local structural element, *i. e.* the global turn structure, evolves first. After that the helical (or strand) parts are completed. The earlier simulation was done using exclusively WMs, whereas in the present study we added some SMs. This mixing increased the folding efficiency, but obviously changed the sequence of folding event dramatically. This is not surprising if one takes into account that a WM, by applying only small local conformational changes, is very similar to short-term events in an MD simulation. An SM causes large global structural changes that occur in reality at large timescales. Therefore, the forming of global structural elements occurs earlier if SMs are applied. This means that the pseudo-timescale preserved in MC simulations using exclusively WMs (Hoffmann, 1996) is lost by adding SMs.

2.2.3.3 Possible solvent and vacuum effects

The lack of a solvent model causes mainly two effects: A hydrophobic effect is missing, and the influence of intramolecular hydrogen bonds is too strong since the solvent molecules as alternative hydrogen bond partners are not available. The latter effect favors α -helices more than β -sheets as discussed above. However, still the folding of the β -hairpin is possible. So there must be something stabilizing the β -hairpin structure, which is present for valine but absent for alanine. Possibly, this is merely the conformational entropy as discussed above, that is gained in a greater amount by the more complex valine sidechains if α -helix is compared to β -sheet.

Isolated β -hairpins are very unstable in solutions and do not occur in nature. It was very hard to design an amino acid sequence that folds to a marginally stable, non aggregating β -hairpin in solution (Muñoz *et al.*, 1997; Kortemme *et al.*, 1998). Probably, the hydrophobic effects drives β -sheet formation in a complex manner (Maynard *et al.*, 1998). In native proteins, β -sheets are in general amphiphatic and have the hydrophobic face buried in the interior to partially constitute a hydrophobic core (Pham *et al.*, 1998). A β -sheet without a hydrophobic core is extremely rare (Koide *et al.*, 2000). However, the fact that they exist shows that there are also other driving forces for β -sheet formation (Koepf *et al.*, 1999; Street & Mayo, 1999).

A vacuum simulation does not mimic the hydrophobic effect, but nevertheless it has some similarity to the situation within the hydrophobic core of a protein. So it may be possible that simulation of β -hairpin formation including a solvent model will be much more difficult than *in vacuo*. In the next section, this situation will be studied.

2.3 Folding in solution

There are several solvent models possible for molecular simulation. An explicit simulation of the individual solvent molecules is not suitable for our MC method, since this would require a complete rearrangement of the solvent structure after each MC move and is therefore very time consuming. The simplest solvent model is to use a distant dependant dielectric constant (*e. g.* $\epsilon = r_{ij}/(1 \text{ \AA})$). This mimics screening effects of the solvent at large distances. Surprisingly, this model does not weaken hydrogen bonds, but increases their strength (by 44 % if using the CHARMM19 force field).

An appropriate model for our MC method is to describe the solvent as a dielectric medium. This could be done by solving the PBE (section 1.1). CAMLAB has already included a very fast PBE solver on the basis of a multigrid method (Hoffmann *et al.*, 1998). However, this solver is still not fast enough to permit folding simulations. Therefore, we decided to use an analytical approximation of the PBE, the so-called Analytical Continuum Solvent (ACS) by Schaefer *et al.* (1998) described in section 1.2.

The first applications of this approach are presented in this section. However, the integration of a solvent model into our MC method is still work in progress. So this cannot be more than a first glance of what will be possible in the near future.

2.3.1 To be investigated: the C-terminal helix of RNase A and the synthetic peptide BH8

For this application, we simulated first 18-alanine with termini capped as before. After that, we chose two really existing peptides:

- An analogue of the C-terminal helix of ribonuclease A termed RN24 with 13 residues (Osterhout Jr. *et al.*, 1989). Sequence: AETAAAKFLRAHA with succinylated N-terminus and amidated C-terminus.
- The polypeptide BH8 with 12 residues (Ramírez-Alvarado *et al.*, 1996), which was specifically designed to form a stable β -hairpin in solution. Sequence: RGITVNGKTYGR

Both peptides were investigated experimentally, but also by MD using the same ACS solvent model (Schaefer *et al.*, 1998). The latter reason was most important for us because this gives us the opportunity to directly compare MD and MC and to check the correct implementation of our ACS version. See the paper of Schaefer *et al.* (1998) for an overview of the experimental studies performed on the two peptides.

2.3.2 Methods

The simulation were done as before with the CAMLAB++ program, now using our new implementation of the ACS solvent model in addition to the CHARMM19 force field. We set the value of the dielectric constant of the solute to $\epsilon_p = 1$ and the dielectric constant of the solvent to $\epsilon_s = 80$. As empirical parameter for the non-polar term (section 1.2.7), we chose $\sigma = 10.5 \text{ J}/(\text{mol \AA}^2)$. The correction value for the hydrogen bonds was $\Delta r = 0.18 \text{ \AA}$. Again, we mixed backbone WMs and backbone SMs by the ratio 10:1. All side chains of residues other than alanine and glycine were treated as before by SMs for their χ torsion angles and CMs to overcome rotation barriers. All simulations started in the extended C_7 -equatorial conformation.

We simulated 18-alanine for $6 \cdot 10^6$ moves (about $3 \cdot 10^5$ scans). In the first half of these trajectories, we decreased the temperature from 1000 K to 125 K. In the second half, we increased the temperature from 125 K to 1000 K. The maximum change of a torsion angle in a backbone SM was

$\pm 5^\circ$. The maximum change of the prerotation torsion angles of a WM was $\pm 10^\circ$. The remaining torsion angle were changed in a WM by a maximum of $\pm 20^\circ$.

For the peptides RN24 and BH8, the trajectories had a length of $3 \cdot 10^6$ moves (about $2.5 \cdot 10^5$ scans). The temperature was constantly 300 K. The maximum change of a torsion angle in a backbone SM was again $\pm 5^\circ$. For sidechain SMs, the maximum change was $\pm 10^\circ$. The change of torsion angles in a WM was restricted to a maximum of $\pm 10^\circ$. The atom displacement used by a side chain CM was restricted to a maximum of 0.02 Å.

For the BH8 peptide, we performed also two additional type of trajectories. For the first type, the hydrogen bond correction Δr was set to zero for interaction of residues whose residue number differed by more than 4. For the second type, CMs were applied everywhere in the window of a WM (not only for the side chains), and the complete CHARMM 19 force field was used instead of the (ϕ, ψ) torsion potentials. No hydrogen bond correction was necessary in this case, since it corresponds to a fully flexible protein model (see above).

2.3.3 Results and Discussion

2.3.3.1 18-alanine

The helix content during a trajectory simulating 18-alanine is plotted in Figure 2.7. Helix formation occurs between 400 and 500 K. The helix melting occurs at slightly higher temperature. These values cannot be compared to experimental results since poly-alanine is insoluble in water. However, our results are in reasonable agreement with recent MD calculations using an explicit solvent model (Takano *et al.*, 1999).

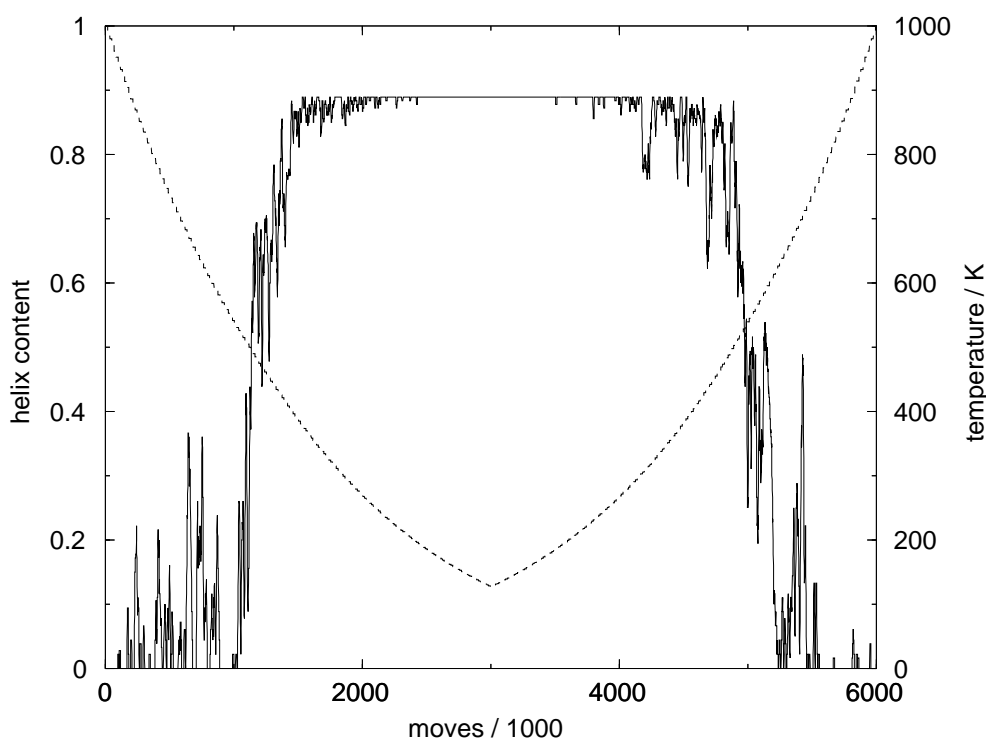


Figure 2.7: Helix formation and melting of 18-alanine. The helix content (left ordinate, determined by DSSP) is plotted by a solid line, the temperature (right ordinate) by a dashed line.

The formation of the helix is extremely fast. Our method needs only about $5 \cdot 10^5$ moves (about

$2.5 \cdot 10^4$ scans, equivalent to the computational effort of about 25 ps of MD simulations) to get from a zero helix content to nearly 90 % (which is the upper limit by definition since DSSP will never assign the α -helix attribute to terminal residues). The folding time scale for poly-alanine in MD simulation at 400 to 450 K is in the order of several nanoseconds (Takano *et al.*, 1999).

2.3.3.2 Peptide RN24

The peptide RN24 readily adopts the structure shown in Figure 2.8. The rmsd of this structure compared to a low energy structure from the MD simulation of Schaefer *et al.* (1998) is 3.5 Å for all atoms together and 1.7 Å considering only backbone atoms. Since such a small peptide as RN24 is not expected to fold to a uniquely defined structure, these values indicate a reasonable structural agreement.

Most of the helical part forms after only about $2.5 \cdot 10^5$ MC moves (about $2 \cdot 10^4$ MC scans). A small N-terminal part finally forms after $2.0 \cdot 10^6$ MC moves (about $1.6 \cdot 10^5$ MC scans). At 300 K, the folding kinetics of this peptide in experiment should be in the order of hundreds of nanoseconds (Takano *et al.*, 1999). In the MD simulation using ACS (Schaefer *et al.*, 1998), the folding time was much slower (three folding events in a 10 ns trajectory). However, Schaefer *et al.* (1998) used an adaptive umbrella sampling, which effectively increases the temperatures in parts of the trajectory, thereby increasing also folding speed. In addition, the viscosity of the solvent has an important influence on folding speed. The effective viscosity of an implicit solvent model like ACS should be quite low, which would again shorten the folding time.

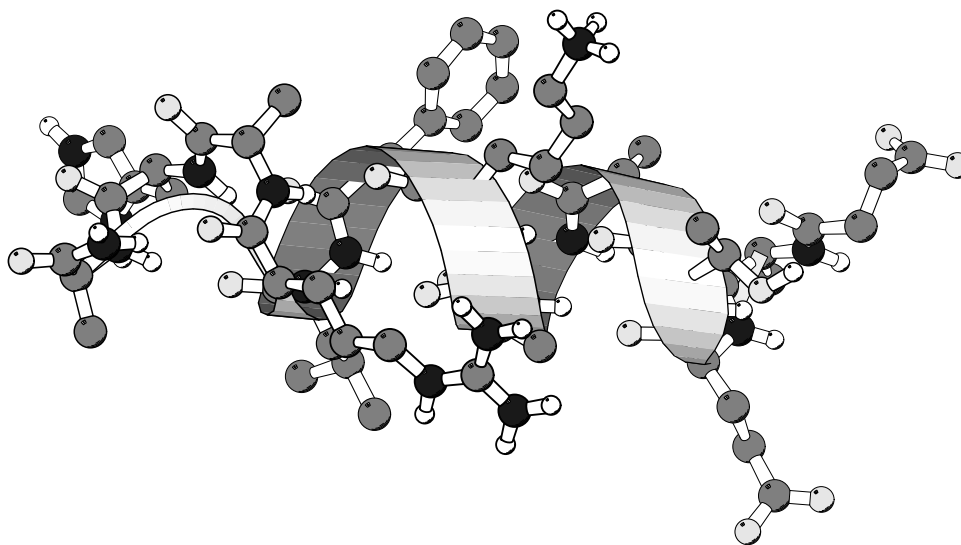


Figure 2.8: Lowest energy structure of the peptide RN24.

The speed-up factor of MD simulations using ACS compared to MD simulations using an explicit solvent model, is about 60 (Schaefer *et al.*, 1998). This speed-up is only with respect to CPU time necessary for the same simulated time interval. It does not include effects due to the adaptive umbrella sampling and the reduced viscosity of the solvent. Altogether, the speed-up of folding is three to four orders of magnitude. Our MC method yields an additional speed-up of about two orders of magnitude compared to the MD simulation using ACS. So the total speed-up of our MC method compared to conventional MD at temperatures of about 300 K is five to six orders of magnitude, which means that extremely long MD calculations using up to hundreds of CPU years on highly parallel supercomputers

(Duan & Kollman, 1998; Daura *et al.*, 1998) are reduced to one day jobs on commodity personal computers.

2.3.3.3 Peptide BH8

The folding of the peptide BH8 was less successful. Using the (ϕ, ψ) torsion potentials and the rigid protein model, the peptide folded sooner or later to a more or less α -helical structure. Variations of the temperature and the Δr correction parameter did not change this behavior fundamentally. It was of course possible to find conditions where no folding at all was possible. But folding of β -structures never occurred. Thus, the rigid peptide geometry combined with the (ϕ, ψ) torsion potentials and the Δr correction parameter are able to yield correct sampling behavior for α -helical structures but fail to simulate folding of β -structures in solution. A possible explanation for this lack of success is that the hydrogen bond correction may only be suitable for residues near to each other in the sequence (as it is the case for hydrogen bond partners in α -helices). Due to the larger flexibility of a long backbone between residues far away from each other in the sequence, the hydrogen bond correction may be unnecessary. If the hydrogen bond correction is applied everywhere, the non-local contacts as those occurring in β -sheets, may be too weak. So we calculated additional trajectories where the hydrogen bond correction Δr was set to zero for interactions of residues whose residue number differed by more than 4. However, with this method we were not able to fold a β -hairpin either. Depending on the parameters, the peptide folded to α -helical structures or secondary structure did not form at all.

Including full flexibility by applying CMs, the peptide BH8 readily folds to a β -hairpin as shown in Figure 2.9. Thus, the failure of folding a β -hairpin is clearly an artifact of the rigid protein model combined with the (ϕ, ψ) torsion potentials. This artifact must be eliminated before we can profit from the much faster folding behavior using the rigid protein model. To accomplish this, we need a deeper understanding of β -hairpin folding in solution. For this, it may help to study carefully the numerous recent MD simulations of β -hairpins and β -sheets in explicit and implicit solvent (Roccatano *et al.*, 1999; Ma & Nussinov, 1999; Wang *et al.*, 1999; Pande & Rokhsar, 1999; Dinner *et al.*, 1999; Wang & Sung, 2000; Ma & Nussinov, 2000; Bonvin & van Gunsteren, 2000). For some other aspects of possible improvement of the (ϕ, ψ) torsion potentials see chapter 4.

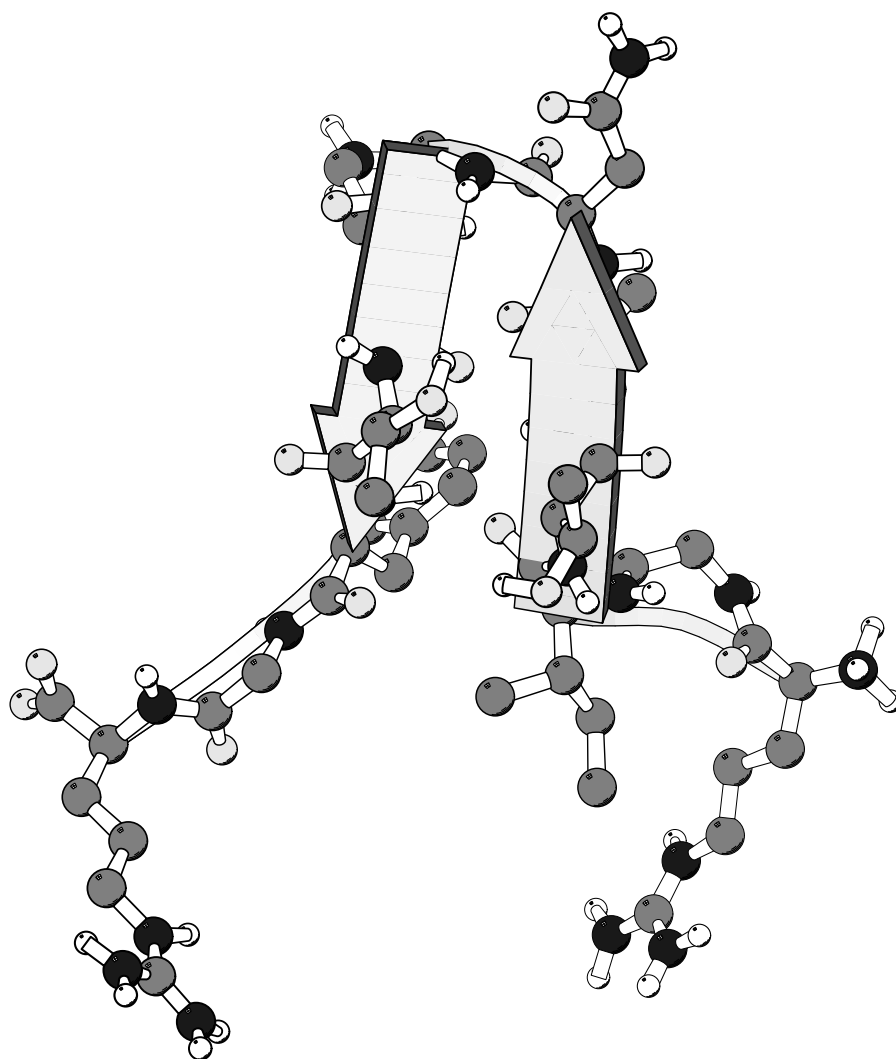


Figure 2.9: Lowest energy structure of the peptide BH8 from the trajectory using full conformational flexibility.

