Chapter 4

Outlook

4.1 Improving the Monte Carlo folding

As shown in chapter 2, the MC method for the long-term dynamics of proteins has a great potential to solve the protein-folding problem in a very intriguing way. In contrast to other methods, our method is not knowledge-based, but needs only a generic force field like CHARMM as input. From this point of view, it is really an *ab initio* method. Our method is also in atomic detail and off-lattice. It may be harder using our method to successfully fold proteins or predict structures, but the possible gain of insight is much larger. However, before this gain of insight becomes reality, several problems and necessary improvements have to be faced:

- The problems with the rigid protein model and the torsion potentials were already described in section 2.3.3.3. A possible reason for the problems are the correlations of adjacent torsion angles that are not included in the torsion potentials. It is possible to break these correlations such that all pair correlations are considered exactly. Members of our group are working on this problem (Kleier & Knapp, manuscript in preparation).

- The present treatment of side chain dynamics leads into problems if larger proteins are simulated, where some of the side chains are tightly packed in the core of the protein such that most of the possible side chain moves will result in atomic clashes and therefore be rejected by the Metropolis criterion. This could be solved by a correlated movement of multiple side chains or by a special configurational bias (Escobedo & de Pablo, 1995; Bates *et al.*, 1997) or dead-end elimination (Leach & Lemon, 1998). Also side chain placement by using a rotamer library is possible (Ullmann, 1995; Mendes *et al.*, 1999). A more extreme approach is to give up an explicit representation of the side chains at all and to treat them implicitly by contact potentials (Clementi *et al.*, 1999; Vendruscolo *et al.*, 2000). This approach is also being tackled in our group (Bastolla *et al.*, 2000). However, the latter two solutions are leaving the path of real *ab initio* folding at atomic detail.

- For the simulation of larger proteins, it is necessary to introduce a kind of cut-off criterion in CAMLAB++. This could be done by a so-called cell algorithm (Ullmann & Knapp, personal communication), which is especially suitable for our MC method.

- Multicanonical or non-Boltzmann sampling can of course also be applied to our MC dynamics. The application of parallel tempering as described in section 3.3.2.4 is straight forward. But also the numerous similar approaches like Tsallis statistics, adaptive umbrella sampling and so on are applicable (Curado & Tsallis, 1991; Berg & Neuhaus, 1992; Tsallis & Stariolo, 1996; Hansmann *et al.*, 1996; Bartels & Karplus, 1997, 1998; Bartels *et al.*, 1998; Moret *et al.*,

1998; Pak & Wang, 1999). However, it has to be kept in mind, which of the effects of these methods are advantageous for our purposes and which are not. For example, a completely equal distribution of all possible conformational states as obtained by a perfect umbrella sampling will directly lead into the Levinthal paradoxon (Zwanzig *et al.*, 1992), which is definitively not what we want,

4.2 Improving the Monte Carlo titration of conformational ensembles

There are several methods to consider conformational flexibility and ensembles of conformations for the computation of the titration behavior of proteins (You & Bashford, 1995; Beroza & Case, 1996; Buono *et al.*, 1994; Sham *et al.*, 1997; Schaefer *et al.*, 1997; Alexov & Gunner, 1997; Rabenstein *et al.*, 1998a; Alexov & Gunner, 1999; Rabenstein & Knapp, 2000a). In this section, I will describe in more detail the method of Beroza and Case (1996), which is very similar to the method developed slightly later by Alexov and Gunner (1997), and compare it to one of our own methods (described in section 3.3). Both methods have their assets and drawbacks. In the end of this section, I will suggest a new method that overcomes the problems of both approaches.

4.2.1 The generalized MC method of Paul Beroza

In the generalized MC method of Beroza and Case (1996), each titratable group can adopt more than only two possible states. In a less general manner, this principle was already applied here: The histidines were treated as titratable groups with three possible states. However, the relatively complicated treatment was only due to our software, which is only capable to treat two-state titratable groups, so that we had to split one group with three possible states into two groups with two possible states each (resulting in four possible combinations, of which one was forbidden). In a program allowing more than two states, the treatment would have been straight forward. The W matrix has to become more complex. Now there is not any longer only one entry W_{ij} for a couple of the groups i and j, but $n_i \cdot n_j$ entries, where n_i is the number of possible states of group i and n_j is the number of possible states of group j. In addition, the equivalent of the intrinsic pK_a value must be determined for all transitions of individual groups from the reference state to each other possible state. The PBE must be solved twice for each possible state of each individual group (once for the model compound and once within the protein). After that, the additional entries for the W matrix are relatively easy to calculate, since all necessary electrostatic potentials are already determined. The increased number of states taking into account all possible combinations is sampled by an MC method as before.

Conformational flexibility is included in this method by not only allowing different protonation states, but also different conformations of titratable group. Actually, the group need not necessarily to be titratable. It is perfectly possible to include *e. g.* a water molecule in different orientations as a "titratable" group that has only different conformational, but not different protonation states. By this method, a huge number of possible conformational states is sampled (exponentially growing with the number of groups as the exponent and the number of possible states per group as basis). However, all the possible conformational substates per group have to be pregenerated and put into the calculation. So there is some vulnerability to a bias due to the selection of conformers. Also the conformational changes are limited to the atoms of "titratable" groups. Background atoms are not allowed to change their position, and – perhaps an even more severe restriction – the dielectric boundary is not allowed to change its shape.

If this method is compared to our method as applied to myoglobin in section 3.3, several fundamental differences become evident (see also Table 4.1). Using our method, the enormous combinatorial increase of uncorrelated conformations of the individual titratable sites is avoided. The conformational changes in our method are correlated, so that only a small number of conformers are

considered. These conformers can be arbitrarily different. Conformational changes can occur at non-titrating residues, and even the dielectric boundary can be modified. However, since also here the conformers must determined in advance, our method is also afflicted by a conformational bias. The vulnerability is even greater due to the small number of conformations.

Table 4.1: Comparison of different methods for titration with conformational flexibility.

	Beroza	this work (section 3.3)	CAMLab++ titration
Number of conformers	exponentially growing (n^m)	small (n)	arbitrary
Predetermination of conformers	local (per residue)	global (whole structure)	none (conformers are generated during MC simulation)
Conformational changes	only in sidechains	everywhere	everywhere
Vulnerability to biases by the selection of conformers	strong	very strong	none
Change of electrostatic boundary	impossible	possible	possible
Non-electrostatic energy contributions	no (but in principle possible)	yes	yes
Field of application	sampling of a huge ensemble of conformations (with a fixed backbone conformation), which are generated by independently combining a few given conformers per sidechain	sampling a small number of (completely) given structures (with arbitrary conformational differences)	sampling unknown conformations, prediction of pH-dependent conformational changes

4.2.2 Combining Monte Carlo dynamics and Monte Carlo titration

The solution to overcome the drawback of both methods, our own and that of Beroza, is to avoid the predetermination of possible conformers by generating the conformational changes “on the fly”. This is possible by combining the MC titration with a kind of protein dynamics method. MD is not suitable for such a task, since it is impossible to change the protonation pattern discontinuously, as it is necessary for the MC titration. However, for the MC dynamics method as presented in chapter 2, this problem does not occur. Protonation moves in addition to the usual conformational moves can be perfectly combined. I call this method CAMLAB++ titration. It is compared to the other two

methods in Table 4.1.

In contrast to previous titration methods, where all necessary solutions of the PBE are done prior to the MC sampling, now we must solve the PBE for each MC move. This is to much computational effort to be feasible. However, as before on the folding chapter, we apply also here ACS, which is an approximation of the results of the PBE. In this way, the computational effort should be tolerable. If one is only interested in small relaxation movements, a fully flexible model for the protein can be used. After solving the problems with the rigid protein model, possibly even large pH induced conformational changes like protein denaturation or the dramatic conformational transition of hemagglutinine can be simulated.