This final chapter presents a summary of the work presented in this thesis in the context of the protein folding problem. The quality of the simulations in this thesis is discussed, and problem areas are designated. An outlook and proposal for further work is given.
8.1 The Quest of the Theoretician

This thesis presents an overview of long molecular dynamics simulations of peptides. The total length of the simulations presented in this thesis is approximately 55 ns\(^1\), which is probably unprecedented. These long simulations were only possible because we have constructed our own parallel computers for MD simulations with supercomputer performance [51, 52]. Alongside the hardware design and production, a complete new MD software package was written, GROMACS [13], which is now being distributed to academic as well as industrial users. The author of this thesis, being the main author of the GROMACS software, also had almost fulltime access to the computers, leading to the large number of simulations that are currently presented\(^2\).

The work in this thesis is theoretical in nature, but there is a clear link to biology and, as far as possible, to experimental data. We have argued (chapter 2 of this thesis) that it is crucial for theoreticians to make the link from theory to experimental data. The tools, i.e. the equations to extract the relevant experimental observables from the simulation data were presented. In the GROMACS package a large number of analysis programs are present which calculate such observables, like J-coupling constants or order parameters. With these programs it becomes almost trivial to analyse the simulation trajectory and produce a large amount of data ready for comparison with experimental data. We have presented such comparisons, using chemical shifts (chapter 3, chapter 5), \(\alpha\)-helicity (chapter 5) and root mean square deviations (chapter 6, chapter 7). After validating our simulations in this way, we went on, trying to extract interesting biological information from our trajectories. This way of working, which basically follows the experimental work, may aid in “filling in the details”, and this is what we did in chapter 3, chapter 4, chapter 5 and chapter 7. It is not a very adventurous way of working, but it may provide new or improved insights into the system under study. In the case of the CaM central helix for instance, we concluded that electrostatic interactions are important for the stability of the \(\alpha\)-helix. A few months after we submitted this work for publication, a paper was published that confirms this conclusion [136].

Although we think experimental verification of simulations is necessary when possible, this should not withhold theoreticians from studying systems for which there is no experimental information. After all, the prediction of properties of molecules (in solution) is the main reason to do theoretical work at all. It is very well conceivable that a careful theoretical analysis invokes experimental verification. In chapter 6 we have presented a somewhat speculative analysis of peptide simulations in the context of co-translational protein folding (i.e. during synthesis), and we have predicted that one of our peptides may be a nucleus for protein folding. We sincerely hope that a similar analysis will be performed by experimental means, since we realise that theoretical studies are not reliable enough yet) to unequivocally prove things.

Ideally, theoreticians should go both ways, following experimental work, and trying to

\(^{1}\)This is a meaningless number, but it will save Herman Berendsen some computing time

\(^{2}\)The average simulation length was 4 days, which means that there is about 7 months of computer time in this thesis
predict things. Moreover theoreticians should also work on methodological improvements. The humble opinion of the author of this thesis is, that he has done all of this. Apart from the software development efforts mentioned above, we are, to our knowledge, the first ones to calculate chemical shifts from a MD trajectory for comparison with experimental data, although chemical shifts have been used for refinement purposes. Furthermore we have introduced a new analysis method for dihedral transitions (chapter 4 of this thesis).

8.2 Success and Limitations

The simulations described in chapter 3, chapter 4 and chapter 5 show that qualitative agreement between simulations and experiment is possible. Sometimes however, almost quantitative agreement is obtained. This thesis presents a number of such cases:

1. The amide chemical shift in a tetrapeptide from BPTI and synthetic analogues was reproduced very well (chapter 3), and one of these peptides was shown to hop between two conformations in a dynamic equilibrium.

2. In a simulation of the CaM central helix we found that it bends in the center, exactly at the location where it is known to bend from experimental data (chapter 7).

It should be borne in mind, that such successes may be lucky shots. In the case of the CaM central helix for instance, the results from a second simulation with a slightly different starting structure (another crystal structure) did not break in the same location. However, the eigenvectors from an essential dynamics analysis [301] were very similar to the simulation described in chapter 7 of this thesis, which indicates that the motions in these simulations were very similar.

On the other hand, this thesis also shows where the limitations of current simulation methodology are. We will summarize the most important points, without going into details.

1. The long range electrostatics problem. In chapter 5 of this thesis as well as chapter 6 of this thesis the treatment of electrostatic interactions using cut-offs, significantly reduced the value of our analyses. It should be noted that some methods to treat the long range electrostatics are available in the literature [53, 236, 237]. One of these methods will be implemented in the GROMACS software Real Soon Now.

2. The sampling problem. Due to the intrinsic flexibility of peptides in solution, many different conformations may be present with almost equal probability. We demonstrated that the motion of a single side chain in a peptide can not be sampled completely in 2.0 ns. Since the backbone of a peptide is considerably more rigid, there is no hope of complete sampling in the near future, although methods targeted at this problem are under development [316, 317].
3. The **polarizability** problem. In many cases, strong interactions between atoms involve electronic polarizability. This effect is not accounted for in any protein force field, which means that subtle interactions like the amino-aromatic interaction we analyzed in chapter 3 of this thesis can not be described very well. A number of water models have been developed with explicit polarizability (for a review see [138]) and in our group a polarizable model for Nitrogen was developed and simulated with the *GROMACS* software [176]. Although the application of such models will make simulations more costly in CPU time, for some applications like quantum/classical simulation aimed at studying reactions in solution the inclusion of polarizability is crucial [318].

8.3 What’s Next?

In some of the chapters in this thesis, work is presented which lends itself for continuation. Here we list some of the possibilities

**Simulations of YTGP, FTGP & YTAP in TIP4P water**

In chapter 3 we presented simulations of tetrapeptides from BPTI using different water models and forcefields. At the time that these simulations were performed we were not able to use the TIP4P water model [166], because it contains a virtual site (a dummy particle, on the bisector of the water molecule). Since the use of virtual sites was implemented in the *GROMACS* software recently, it would be interesting to study whether this water model performs better than TIP3P [166] with either the G-94 or the OPLS force field.

**Refinement of pep25**

The simulations as described in chapter 5, could be continued by structure determination of pep25 from CCMV coat protein, using restrained MD. It was derived from distance-geometry calculations that there are two structure classes of pep25 in solution [222]. Since the difference between the two classes is quite large (about 0.5 nm RMS Deviation for the a-helical part) a special procedure must be used to refine the structure. We propose to perform MD simulations with the following setup: 2 pep25 molecules in a large box of approximately 10000 water molecules, corresponding to a concentration of 11 mM, which is only slightly higher than the concentration of 7mM used in NMR experiments. Furthermore, the box should contain ions in the same concentration as was used in NMR experiments (10 Pi4, 18 Acetate, 40 Na+). A proper electrostatics method should be applied, like PPPM [237] or a method based on a Poisson solver [233], and NOE restraints should be imposed using ensemble averaging [54, 319] and time averaging [115]. Ensemble averaging will allow the two pep25 molecules to find two different structure classes simultaneously, thus satisfying the NOE restraints, while time averaging will allow the peptides to interconvert between structure classes. Because interconversion may take quite long, such a simulation should again be in the nanosecond time range.
8.4 and the Protein Folding Problem...

You are almost at the end of another thesis that does not solve the protein folding problem. Or does it? The great intellectual challenge of the protein folding problem is to understand the folding pathway of a protein from an unfolded conformation to the native one. It has been proven very hard to study such pathways experimentally, although folding pathways for a few proteins have been proposed, like Lysozyme [276]. However, the experimental techniques for the study of folding pathways have evolved dramatically over the last two decades, and consequently, the recent literature on protein folding studies has evolved into a giant media-hype, with weekly publications in all leading scientific journals. A very important aspect of the protein folding problem is usually ignored in these studies. If we know a folding pathway of a certain protein, how does this help us in predicting the folding pathway of another protein?

It has long been recognized that the conformation of a protein is determined by the non-covalent interactions [1], and in our opinion it seems to be non-sensical to assume that the folding pathway is not governed by such interactions. In this respect, it is not clear what detailed experimental resolution of folding pathways contributes to the understanding of these interactions, especially since folding pathways can not be determined in atomic detail.

We think that real progress in resolving the protein folding problem can only be achieved by a better understanding of the interactions within proteins and between proteins and their environment. As discussed in chapter 1 of this thesis, a theoretical model for the interactions in proteins is available, although the model requires a set of parameters, the force field, which still is not perfect. Experimental studies involving site-directed mutagenesis combined with protein structure determination, and measuring thermodynamic properties like relative free energy of folding may aid in improving theoretical force field based models. Within 15 years, computers will be fast enough to perform a 1 millisecond MD simulation of a protein in solution, which is about the time of folding for a small protein. If the quality of force fields increases accordingly, the protein folding problem will be history in the near future.