Chapter 6

Summary & general discussion

In the standard crystallographic experiment, information about the phases of the reflections is lost. From the observed intensities alone, the electron density of the unit cell cannot be reconstructed and this problem is known as the crystallographic phase problem. At the resolution limits typically obtained in protein crystallography the phase problem is underdetermined and requires incorporation of additional information. Chapter 1 of this thesis describes existing, experimental and computational methods to determine and improve phases. The computational methods exploit prior knowledge about the content of the unit cell to supplement the limited experimental information. The quality of the available phase information typically determines how much prior knowledge can be expressed. In chapter 2 a novel refinement method is presented, called conditional optimization. The conditional formalism offers an N-particle solution for the assignment of topology to loose, unlabelled atoms. Thereby, conditional optimization allows expression of large amounts of geometrical knowledge about protein structures without the requirement of a molecular model, and thus potentially in the absence of phase information. Initial tests with a simplified structure and calculated data show that with this method, in principle, random atom distributions can be successfully optimized and *ab initio* phasing of medium-resolution data can be achieved. Conventional methods for the estimation of phase probabilities fail for models of very low phase quality and a novel multiple-model procedure to estimate σ_A-values was necessary for these optimizations. In chapter 3 a mean-force potential for conditional optimization of protein structures is presented, which expresses knowledge about common protein conformations like α-helices, β-strands and loops. Using this generally applicable parameter set, conditional optimization of three small protein structures against 2.0 Å observed diffraction data shows a large radius of convergence, validating the presented force field and illustrating the feasibility of the approach. The application of conditional optimization to automated model building is explored in chapter 4. For three test cases with data to medium resolution and good experimental phases, conditional optimization yields models of comparable quality as obtained by the commonly used programs *ARP/wARP* and *RESOLVE*. Chapter 5 describes the application of conditional optimization to *ab initio* phasing of observed diffraction data. Low-resolution reflections and reliable phase probability estimates appear to be important for convergence. For the presented test case promising results are obtained, indicating that also with observed

diffraction data successful optimization of random atom distributions may be possible.

Given the success of direct methods in small-molecule crystallography, a commonly applicable method for *ab initio* protein structure determination would be a valuable tool for structural biology. The results described in this thesis indicate that conditional optimization is a promising candidate for *ab initio* phasing of medium-resolution, protein diffraction data. On the other hand, major advances have been obtained with experimental approaches to solve protein structures. The possibility to incorporate seleno-methionine in protein molecules using bacterial expression systems has contributed significantly to the successful application of anomalous dispersion methods, and eukaryotic expression systems for this purpose are being developed. Although these methods still require biochemical modifications to the native protein molecule, techniques exploiting the anomalous signal of sulfur atoms in cysteines and methionines require only a single diffraction experiment on a native protein crystal. Currently, the weak anomalous signal of sulfur is still difficult to assess experimentally, but improvements in data collection and processing techniques may also render these experimental methods a serious alternative for *ab initio* approaches. In addition, also molecular replacement may be expected to become more generally applicable given the rapidly increasing number of different folds in the protein structure database.

All the approaches mentioned here share the ultimate goal of providing a commonly applicable solution to the phase problem in protein crystallography. Conditional optimization may benefit from advances in the other approaches and become suitable for common practice before reaching this ultimate goal. Starting from initial phases as provided by the other methods, the flexible searching behaviour with unlabelled atoms and the incorporation of prior geometric knowledge by the conditional formalism may be exploited to extend phase information in a commonly applicable way. As expressed in chapter 4 and 5, excessive computational costs form a stumble block for the applicability of conditional optimization. Parallelization of the program may be a way of dealing with this problem and such a development is underway. Common application of this method may then induce developments to further improve the applied procedures. Improvements in the rate of convergence may be expected from extension of the force field (especially concerning loops and side chains), optimization of hybrid models with explicit topology assignments for recognizable protein fragments, and development of discrete modelling steps in the optimization protocols. Estimation of reliable phase probabilities for models of low phase quality is also a point of interest. Various multiple-model procedures have been postulated in this thesis, but none of them proved generally applicable and further investigation in this direction is needed. Eventually, also the goal of *ab initio* phasing by conditional optimization would favour from these developments. From an optimistic point of view, and given the promising results described in chapter 2 and 5, this method may then become, perhaps among others, a commonly applicable tool for protein structure determination.