Computational protein modelling supported by NMR data

1 Purpose and aims

The aim of this project is to improve the speed and accuracy with which three-dimensional models of protein structures can be generated from easily and rapidly obtainable nuclear magnetic resonance (NMR) data. The computational approach that will be developed will deepen the understanding of how nature folds and assembles proteins into larger molecular complexes, and provide fast and economic access to structural information, for example on targets in drug discovery.

The protein modelling program that will be built consists of two main parts: a *search algorithm* for exploring the vast conformational search space, and a *scoring function* for evaluating the model protein structures. The search algorithm proposed here will be the first to combine (i) the "zipping and assembly" dynamic programming algorithm, (ii) a fragment-based approach for selecting and combining molecular substructures and (iii) conformational clues from NMR experiments. The scoring function proposed here will be the first to include atomic group packing propensity scores and evolutionary information.

The proposed project is a new multi-disciplinary collaboration between the Computer Science and Engineering at Chalmers, and Chemistry at the University of Gothenburg. The project will benefit from close connections to the Swedish NMR Centre, which will be able to provide new experimental data that can be used to test and refine the system that will be developed. It will also benefit from close connections to the Chalmers e-Science Centre, which will provide access to a local network of expertise in large-scale scientific computation.

Background

Proteins are important biological macromolecules that consist of chains of amino acid residues (small molecular subunits that are the building-blocks of proteins). Each higher organism has tens of thousands of different proteins, and various proteins play important functional roles in every biological process. In order to fully understand a protein's biological function, it is necessary to know its three-dimensional structure. This is because biological function is ultimately due to interactions between molecules, and a full understanding of these interactions requires detailed knowledge of the shapes and chemical properties of the interacting molecules.

Nuclear magnetic resonance (NMR) is one of the main experimental methods for determining the three-dimensional structure. NMR experiments provide a variety of restraints that can be used when constructing a model structure. The most important NMR restraints for determining a protein structure are the nuclear Overhauser effect (NOE) restraints that give distance estimates for pairs of atoms that are close together in three-dimensional space. Other restraints relate to torsion angles and also to the orientation (in a fixed coordinate system) of selected bonds, often the N-H bonds of the backbone. These latter restraints can be derived from residual dipolar couplings (RDCs), which are considered the easiest data to obtain for large proteins.

Computational methods play an important role in NMR structure determination. Having obtained a set of distance, torsion angle and orientation restraints, the challenge is to find a three-dimensional structure, or more typically an ensemble of structures, compatible with those restraints. Traditionally, this has involved the use of distance geometry, but today molecular dynamics simulation methods and simulated annealing are most commonly used.

The term NMR covers a wide range of experiments that differ in their cost, time and difficulty. While some experiments are relatively straightforward and are routinely performed when studying a new protein, there are other experiments that are very time-consuming and more expensive. The aim of this project is to develop a protein modelling program that can be used together with data from the relatively straightforward NMR experiments to obtain accurate model structures quickly, reducing or even eliminating the need for more expensive and complex multidimensional NMR experiments that require alternative isotope labelling to be done and take longer to perform.

2 Survey of the field

2.1 Protein modelling with NMR data

The value of using fragments from known protein structures when building three-dimensional model structures based on NMR data has long been recognised. Kraulis and Jones [14] derived distance matrices based on short-range NOEs (NOEs between amino acid residues that are relatively close to each other in sequence) and matched these with corresponding distance matrices derived from fragments from known protein structures. Those fragments that matched were clustered, and the fragment closest to the centre of the largest cluster was selected as the most likely conformation for part of the protein being modelled. Delaglio *et al.* [6] use the additional information available from RDCs in their molecular fragment replacement method (MFR), identifying 7-residue fragments in the Protein Data Bank [1] whose backbone torsion angles are compatible with the measured dipolar couplings.

More recently, there has been interest in using *de novo* structure prediction methods in combination with NMR restraints to obtain three-dimensional structures that are compatible with those restraints (e.g. [3, 24, 21]). Existing work in this area is based on the Rosetta *de novo* structure prediction method (see Section 2.4). However, we believe that the zipping and assembly method for exploring the conformational search space (see Section 2.2) is better suited to this task than the approach taken by Rosetta (see later) and will also offer the benefits that are outlined at the end of Section 2.2. Therefore, we propose to implement a zipping and assembly method for protein folding, and to use this with NMR data.

2.2 Zipping and assembly

The zipping and assembly method (Figure 1) was first described in the context of lattice models, where a sequence of hydrophobic and polar "amino acid residues" are folded onto a regular square grid [7, 9]. Lattice models are a gross over-simplification of the real

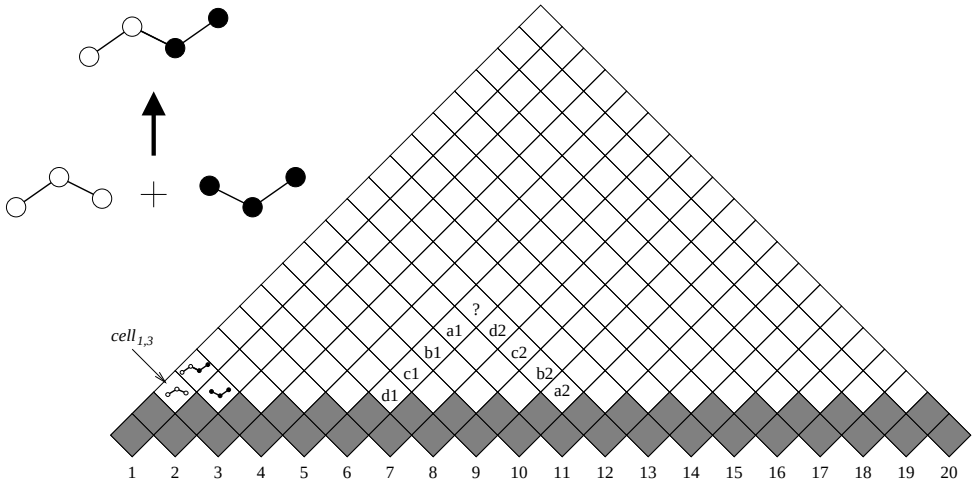

Figure 1: Zipping and assembly

protein folding problem, but they do provide a convenient framework for experimenting with search strategies and simplified scoring functions [15].

The zipping and assembly search method uses a dynamic programming algorithm (a variant of the Cocke-Kasami-Younger chart parsing algorithm that is commonly used for parsing sentences) to construct larger molecular fragments from smaller ones. The cells in Figure 1 contain sets of fragments that are candidates for modelling the conformation of part of the target protein. Suppose we splice two fragments together by overlapping the last two amino acid residues of one fragment with the first two residues of the next. Then the smallest fragments that can be considered are fragments of length three. The leftmost white cell in Figure 1, $cell_{1,3}$, will contain a set of possible conformations for modelling the fragment from residue 1 to residue 3 in the target protein. The cell to its right, $cell_{2,4}$, contains a set of possible conformations for modelling positions 2 to 4. To generate a model of the four-residue fragment from positions 1 to 4, a fragment chosen from $cell_{1,3}$ is combined with one chosen from $cell_{2,4}$, and the resulting fragment is stored in $cell_{1,4}$. A large set of candidate conformations can be constructed by choosing and combining fragments from lower cells. Consider next the seven-residue fragment from positions 6 to 12. Possible conformations for this fragment will be stored in $cell_{6,12}$, which contains a question mark in Figure 1. These can be constructed by combining a fragment chosen from the cell labelled a1 (six residues, spanning positions 6 to 11) with one from the cell labelled a2 (three residues, spanning positions 10 to 12), or combining a fragment chosen from cell b1 with one chosen from cell b2, and so on. Similarly, all cells in the diagram can be filled with fragment conformations that are the result of combining a fragment chosen from one of the cells to the lower left of that cell, with a fragment from a cell to the lower right. Finally, the cell at the apex of Figure 1 will contain a set of possible conformations for the entire protein (consisting of 20 residues in this toy example).

Claims made for the zipping and assembly method include [7]:

— its local-first-global-later search explains quick folding, and avoidance of vast stretches of conformational space ("local" here refers to local in sequence);

— it reflects the parallel nature of physical kinetics;

— it captures the relationship between contact order (whether pairs of amino acid

residues that are close together in 3-D space also tend to be close to each other along the protein chain, or tend to be distant from each other along the protein chain) and folding rate;

— it identifies slow- and fast-folding proteins, and slow- and fast-folding routes.

2.3 Rosetta

Rosetta [25, 11] is perhaps the best known *de novo* protein structure prediction method. To understand why we are proposing to use a zipping and assembly method instead of Rosetta, we need to consider the search strategy used by Rosetta, and how it differs from zipping and assembly. Rosetta proceeds by first building an entire extended protein chain. Then the conformation of the chain is repeatedly modified by replacing the conformation of a randomly chosen segment of fixed length (usually 9 residues or 3 residues) in the protein chain with the conformation of a fragment taken from another protein. Sets of potential replacement conformations are compiled in advance as described in [25]. Thus Rosetta works with a model of the entire protein at all times, and the scoring function used to evaluate the generated models will evaluate the entire protein model. In contrast, a zipping and assembly method builds a model structure incrementally, and the partial structures are all valid substructures. As stated in [7], "the [zipping and assembly] search method is efficient because it never searches more than a few degrees of freedom at a time, and eliminates high energy conformations early in the search". While Rosetta can claim some successes, it can be seen from Figure 1 of [21] that Rosetta is often unsuccessful even with proteins that are much smaller than 100 residues, even when the scoring function is supplemented with experimental data from NMR chemical shifts.

2.4 Rosetta supported by NMR data

The Rosetta *de novo* protein structure prediction method has been used to support structure determination by NMR. Bowers *et al.* [3] use a variant of the method described in [25] in which the three- and nine-residue fragments used for modelling segments of the model protein are scored according to agreement with a multiple sequence alignment, but also compatibility with backbone torsion angle restraints derived from NMR chemical shift assignments. Further, those fragments that are incompatible with short-range NOE distance restraints are discarded. That approach is extended and evaluated in chemical-shift-Rosetta (CS-Rosetta) [24]. A further extension is to use RDCs as part of the structure evaluation function during Monte Carlo conformational sampling [21]. The resulting protocol, CS-RDC-Rosetta, uses both backbone NOEs and backbone RDCs, and is found to give improved convergence over CS-Rosetta for some test cases. However, Figure 1 of [21] shows that this protocol does not always produce accurate models, and sometimes substantial parts of the model do not converge to an unambiguous conformation. Further, large proteins are difficult to model using this approach: "for proteins with over 120 residues, conformational sampling becomes limiting" [21]. This limitation is due to the conformational search strategy used by Rosetta, and is not a limitation of the NMR experiment or the RDC data.

The zipping and assembly method has not been used previously with NMR data. Advantages of the zipping and assembly approach to conformational sampling are mentioned

at the end of Section 2.2. Further, we believe that the bottom-up approach to structure generation used in zipping and assembly will allow it to scale better than Rosetta when applied to larger proteins. This should also make it better suited to modelling multi-domain proteins, since each domain can be modelled independently of the others. It is for these reasons that we are proposing here to use zipping and assembly together with NMR data.

3 Project description

In this project we aim to build a protein folding program that uses a zipping and assembly search method, and then use this together with restraints derived from NMR experiments to obtain three-dimensional structures. The proposed system is an example of the "future NMR pipeline" in a "new paradigm for NMR structure determination" proposed in [21]. In that vision, protein backbone restraints derived from relatively straightforward NMR experiments are used as input to the Rosetta protein folding program. While our aims are the same, we intend to use a different protein modelling strategy which we believe is better suited to the task than Rosetta.

By "relatively straightforward NMR experiments", we mean those that do not require special isotope labelling of the protein, or a complete resonance assignment for the protein. Carrying out resonance assignment is a necessary step for almost all NMR characterisations of proteins. The difficulty of performing this step depends on the protein size, the isotope labelling available and the expected results. Thus, assignments of backbone nuclei only can be obtained much more easily than complete assignments, requiring a lower extent of (or no) labelling and achievable by automatic procedures. The goal here is to use, for various types of proteins, only the easily obtainable NMR data, typically limited to backbone data.

Implementation of the tasks described below will be done mainly by a PhD student. The student will be supervised jointly by Graham Kemp and Martin Billeter, and thus will have access to expertise in computational molecular modelling and NMR structure determination.

3.1 Conformational search

The proposed method for constructing three-dimensional models is based on "zipping and assembly" (e.g. [20]), but will make greater use of structural information in the rapidly growing Protein Data Bank of known protein structures [1] than previous work with this search method, and will make explicit use of geometrical restraints obtained from NMR experiments, where available.

NMR data will be used both in generating or selecting conformations for the fragments to be assembled, and also within the scoring function that will be used to evaluate partial substructures and also the eventual models of the entire protein.

NMR restraints can characterise local secondary structures such as α-helices, β-strands and tight turns. In these cases, fragments with ideal geometry can be incorporated directly into the model structure, eliminating the need for conformational search in that part of the protein. For other parts of the main chain, the approach taken by Delaglio *et al.* [6] to

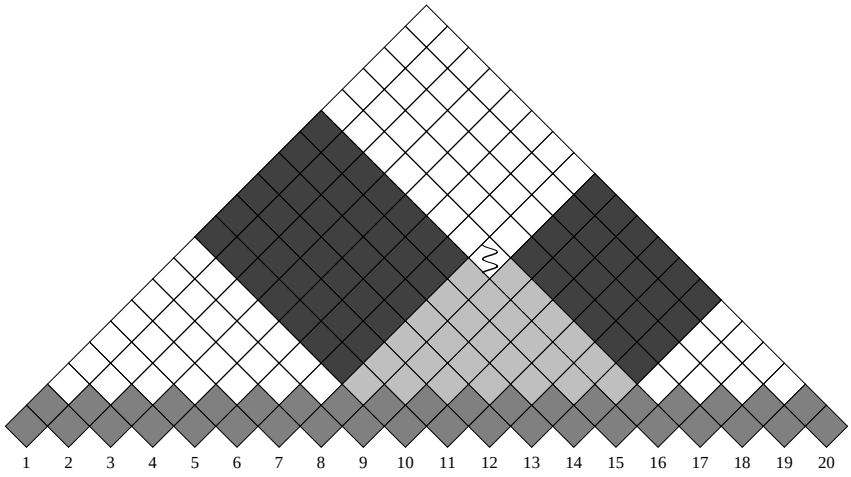

Figure 2: Zipping and assembly — pruning the search space where a region is known to adopt a particular conformation. Suppose NMR data indicate that the segment of the protein chain from positions 8 to 16 will have a particular conformation (e.g. an α-helix), then that part of the chain can be modelled directly, and the conformation can be placed into $cell_{8,16}$ of the solution matrix. This has the effect of pruning the search space: conformations do not have to be constructed for the fragments represented by the cells below $cell_{8,16}$ with light shading, since these are subsumed by the longer fragment from 8 to 16; conformations should not be constructed for the fragments represented by the cells with dark shading, since these overlap with one end of the fragment from 8 to 16 with known conformation, and the region of overlap should not be remodelled.

search the Protein Data Bank for fragment conformations that are compatible with local backbone NMR restraints. The computational framework for zipping and assembly can be adapted to allow fragments with "known" conformation to be incorporated into the model, thus taking advantage of structural knowledge, where available, to reduce computation times (Figure 2). Using fragments from known structures (either based on compatibility with RDCs [6], or local sequence-structure preferences [4, 8, 2]) would be different to previous off-lattice work with zipping and assembly [20], which uses molecular mechanics (rather than database search, as proposed here) to generate fragment conformations.

While the zipping and assembly method is most straightforwardly described when we have a contiguous sequence of residues (as shown from left-to-right in Figure 1), the method could be adapted to allow the chain to be segmented, and for the segments to be rearranged along the base of the triangle in Figure 1. This would make the code for combining fragments more complicated, but could bring overall performance benefits for proteins with many contacts between residues that are far apart in the protein's sequence by allowing distance restraints due to long range NOEs to be included earlier in the computation. Graphically, this modification would have the effect of identifying cells with many restraints that are in the higher part of the matrix in Figure 1, and moving these down to a lower part of the matrix, where these restraints can be applied earlier in the structure calculation.

The zipping and assembly method could be adapted to make use of parallel computation. For example, the task of generating and evaluating conformations for each cell in Figure 1 could be allocated to a different process, with a master process coordinating the computation. This would enable the software proposed here to benefit from multiple processors

in a Grid environment.

3.2 Scoring function

Finding a suitable scoring function is one of the challenges in any protein modelling program. This is even a weakness in Rosetta, the leading program in the protein folding area — a recent study of four modelling cases highlights "the poor discrimination of the Rosetta all-atom energy function" [5].

We shall use long range NOEs and orientation restraints as part of the scoring function used with a zipping and assembly search strategy, with the aim of rapidly constructing ensembles of structures, built from fragments with real protein geometry, that are compatible with NMR data.

We shall use new methods under development in our group for predicting interacting surfaces based on local surface patch descriptors [17, 18] as part of the scoring function used to evaluate the packing between fragments during assembly steps. In doing this, there are similarities with the *de novo* protein modelling approach described by [10], however (i) that work [10] "docks" molecular fragments but uses a different approach to scoring the interactions to the one proposed here, (ii) that work does not use zipping and assembly as a search method, and (iii) that work does not use NMR data in identifying conformations for parts of the protein chain.

A further extension is to include evolutionary information in the scoring function. If we consider a multiple alignment of the sequences of a family of related proteins, we typically find that those positions that are located in the core of a protein tend to be more strongly conserved during evolution than those positions that are on the exterior surface of the protein. This is due to there being selective forces favouring groups of interacting residues that fit well together in the protein core being conserved across the members of the family. We shall use this observation directly in the scoring function by rewarding packing between pairs and groups of residues that are at relatively conserved positions in a multiple sequence alignment.

We shall experiment with the use of different kinds of restraints from NMR experiments together with a zipping and assembly search strategy. The software developed in this project could also serve as a framework for incorporating restraints from sources other than NMR. Further, in an extreme case, the method could also be used in the absence of any NMR, and in this case would perform as a *de novo* protein structure prediction method.

3.3 Timetable

Year 1: implement conformational search method; implement scoring function that includes NMR restraints.
Year 2: analyse atomic group packing in protein cores; construct interaction propensity tables, and incorporate these into the scoring function; incorporate conservation information from molecular phylogenetic trees into the scoring function.
Year 3: test the software with NMR data; implement a parallel version of the software.
Years 4 and 5: initiate closer collaboration with other NMR groups; extend and refine the method; package the software.

4 Significance

The research proposed here is targeted towards improving our understanding of protein folding and how nature (almost) always achieves the correct fold. Thus, the work will address the basic research challenges presented by the protein folding problem.

The proposed project will enable us to obtain informative three-dimensional protein models using only those restraints that are obtained from relatively quick NMR experiments. This will reduce or even eliminate the need for more expensive multidimensional NMR experiments that require complex isotope labelling to be done and take longer to perform.

The new approach will be a competitive alternative to Rosetta. The "divide-and-conquer" nature of the dynamic programming algorithm used for zipping and assembly should allow larger proteins to be tackled; today this is still a big problem.

The software developed in this project will be of use to NMR groups in academia and in the pharmaceutical industry. The aim is that by combining a zipping and assembly modelling approach, together with relatively easily obtained data from NMR experiments, experimental groups will be able to obtain useful models without the need to invest time and money in performing additional experiments.

5 Preliminary results

A recent student on the Bioinformatics and Systems Biology programme at Chalmers undertook a masters project in this area with Graham Kemp [22]. A program that constructs "decoy" alpha-carbon structures (in the spirit of [12]) using a zipping and assembly approach has been written. The scoring function used in this program is very simple, and does not use NMR data, but this has given us experience in implementing the data structures and search algorithm that will be needed in the project proposed here. This prototype software will be available for use in the proposed project.

Surface triplet propensities, proposed by Kemp and implemented in a Masters project in his group [17], have been used successfully to predict the location of protein binding surfaces [18]. We propose to adapt that method to analyse atomic group packing in protein cores, and use the derived propensity values when scoring conformations during protein modelling.

Graham Kemp's group has generated a suite of software routines for analysing, comparing, transforming and superposing molecular structures, and the proposed project will benefit from being able to build on these. This includes a library of 3-D geometry routines, a library of routines for working with 3-D transformations. The group has experience in devising and implementing new methods for modelling proteins (e.g. [23, 26]), and also have some experience in using genetic algorithms to model carbohydrates [19]. Martin Billeter's experience includes developing parallel molecular dynamics simulation methods [13] and tools for automated analysis of NMR data [16]. This combination of relevant software tools and molecular modelling experience makes us ideally prepared for the project proposed here.

6 Part of project cost

We request the full project cost from the Swedish Research Council.

7 Budget

The main item in the budget is salary costs for a PhD student to work on the project proposed here. A contribution to cover a percentage of the applicants' time for work on this project is also requested.

8 International and national collaboration

Graham Kemp has had recent collaborations on predicting protein-ligand binding sites with Prof Malcolm Walkinshaw (University of Edinburgh) and on conformational analysis of MHC class II binding grooves with Prof Nagasuma Chandra (IISc Bangalore). He is the main supervisor for an industrial PhD student who works at Biognos AB, a computational drug design company based in Gothenburg, and another PhD student working at the Department of Radiation Physics at the University of Gothenburg. He has had recent collaborations with researchers at Cell & Molecular Biology and Zoology at the University of Gothenburg, and has been examiner for Masters student projects at AstraZeneca.

Martin Billeter's collaborations include (1) a NMR study on Zn-containing antitoxin from E. coli that is involved in the resistance of bacteria to antibiotics (Profs. A. Vlamis and A. Holmgren, Karolinska Institutet, and Prof. A. Grslund, Stockholm University), (2) a binding study by NMR of a Ru-complex to DNA (Profs. B. Norden and P. Lincoln, Chalmers Technical University), (3) a spectroscopic study of the interactions of defensins with oligosaccharides (Prof. Siebert, Kiel) and (4) an application to the EU (FP7-PEOPLE-2011-ITN), labeled TYROSINE, with the goals to develop a network between experts covering a wide range of experimental techniques and bioinformatics approaches to study biomolecular interactions (full partners: universities in Brussels, Leicester, Nijmegen, Gothenburg, Barcelona, Oxford, London plus 3 companies; associated partners: Protein Data Bank Europe and 7 companies).

References

[1] H. Berman, K. Henrick, and H. Nakamura. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.*, 10:980, 2003.

[2] W. Boomsma, K. V. Mardia, C. C. Taylor, J. Ferkinghoff-Borg, A. Krogh, and T. Hamelryck. A generative, probabilistic model of local protein structure. *Proc. Natl. Acad. Sci. USA*, 105:8932–8937, 2008.

[3] P. M. Bowers, C. E. M. Strauss, and D. Baker. De novo protein structure determination using sparse NMR data. *J. Biomol. NMR*, 18:311–318, 2000.

[4] C. Bystroff, K. T. Simons, K. F. Han, and D. Baker. Local sequence-structure correlations in proteins. *Current Opinion in Biotechnology*, 7:417–421, 1996.

[5] R. Das. Four small puzzles that Rosetta doesn't solve. *PLoS One*, in press, preprint at:http://arxiv.org/abs/1103.3042, 2011.

[6] F. Delaglio, G. Kontaxis, and A. Bax. Protein Structure Determination Using Molecular Fragment Replacement and NMR Dipolar Couplings. *J. Am. Chem. Soc.*, 122:2142–2143, 2000.

[7] K. A. Dill, A. Lucas, J. Hockenmaier, L. Huang, D. Chiang, and A. K. Joshi. Computational linguistics: A new tool for exploring biopolymer structures and statistical mechanics. *Polymer*, 48:4289–4300, 2007.
[8] T. Hamelryck, J. T. Kent, and A. Krogh. Sampling Realistic Protein Conformations Using Local Structural Bias. *PLoS Computational Biology*, 2:e131, 2006.
[9] J. Hockenmaier, A. K. Joshi, and K. A. Dill. Routes Are Trees: The Parsing Perspective on Protein Folding. *Proteins: Structure, Function and Bioinformatics*, 66:1–15, 2007.
[10] Y. Inbar, H. Benyamini, R. Nussinov, and H. Wolfson. Combinatorial docking approach for structure prediction of large proteins and multi-molecular assemblies. *Phys. Biol.*, 2:S156–S165, 2005.
[11] K. W. Kaufmann, G. H. Lemmon, S. L. DeLuca, J. H. Sheehan, and J. Meiler. Practically Useful: What the Rosetta Protein Modeling Suite Can Do for You. *Biochemistry*, 49:2987–2998, 2010.
[12] R. Kolodny and M. Levitt. Protein Decoy Assembly Using Short Fragments Under Geometric Constraints. *Biopolymers*, 68:278–285, 2003.
[13] R. Koradi, M. Billeter, and P. Güntert. Point-centered domain decomposition for parallel molecular dynamics simulation. *Computer Physics Communications*, 124:139–147, 2000.
[14] P. J. Kraulis and T. A. Jones. Determination of Three-Dimensional Protein Structures From Nuclear Magnetic Resonance Data Using Fragments of Known Structures. *Proteins: Structure, Function and Genetics*, 2:188–201, 1987.
[15] K. F. Lau and K. A. Dill. A Lattice Statistical Mechanics Model of the Conformational and Sequence Spaces of Proteins. *Macromolecules*, 22:3986–3997, 1989.
[16] D. Malmodin and M. Billeter. Multiway decomposition of NMR spectra with coupled evolution periods. *J. Am. Chem. Soc.*, 127:13486–13487, 2005.
[17] W. Mehio. A Novel Method to Predict Protein-Protein Binding Surfaces. Master's thesis, Chalmers University of Technology, 2007.
[18] W. Mehio, G. J. L. Kemp, P. Taylor, and M. D. Walkinshaw. Identification of protein binding surfaces using surface triplet propensities. *Bioinformatics*, 26:2549–2555, 2010.
[19] A. Nahmany, F. Strino, J. Rosen, G. J. L. Kemp, and P.-G. Nyholm. The use of a genetic algorithm search for molecular mechanics (MM3)-based conformational analysis of oligosaccharides. *Carbohydrate Research*, 340:1059–1064, 2005.
[20] S. B. Ozkan, G. A. Wu, J. D. Chodera, and K. A. Dill. Protein folding by zipping and assembly. *Proc. Natl. Acad. Sci. USA*, 104:11987–11992, 2007.
[21] S. Raman, O. F. Lange, P. Rossi, M. Tyka, X. Wang, J. Aramini, G. Liu, T. A. Ramelot, A. Eletsky, T. Szyperski, M. A. Kennedy, J. Prestegard, G. T. Montelione, and D. Baker. NMR Structure Determination for Larger Proteins Using Backbone-Only Data. *Science*, 327:1014–1018, 2010.
[22] F. Rashid. A System that Builds Decoy Protein Backbones by Assembling Smaller Fragments. Master's thesis, Chalmers University of Technology, 2010.
[23] D. W. Ritchie and G. J. L. Kemp. Modeling Antibody Side Chain Conformations Using Heuristic Database Search. In T. Gaasterland, P. Karp, K. Karplus, C. Ouzounis, C. Sander, and A. Valencia, editors, *Proceedings of the Fifth International Conference on Intelligent Systems for Molecular Biology*, pages 237–240, 1997.
[24] Y. Shen, O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, G. Liu, A. Eletsky, Y. Wu, K. K. Singarapu, A. Lemak, A. Ignatchenko, C. H. Arrowsmith, T. Szyperski, G. T. Montelione, D. Baker, and A. Bax. Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. USA*, 105:4685–4690, 2008.
[25] K. T. Simons, C. Kooperberg, E. Huang, and D. Baker. Assembly of Protein Tertiary Structures from Fragments with Similar Local Sequences using Simulated Annealing and Bayesian Scoring Functions. *J. Mol. Biol.*, 268:209–225, 1997.
[26] M. T. Swain and G. J. L. Kemp. A CLP approach to the protein side-chain placement problem. In T. Walsh, editor, *Principles and Practice of Constraint Programming — CP2001*, pages 479–493. Springer-Verlag, 2001.