Abstract

In traditional protein structure determination from solution Nuclear Magnetic Resonance (NMR) data, the nuclear Overhauser effect (NOE) distance restraints are primarily used in a simulated annealing/molecular dynamics (SA/MD) protocol to compute the protein structure. It is only in the final stages of structure computation, the orientational restraints such as residual dipolar couplings (RDCs) and residual chemical shift anisotropies (RCSAs) are incorporated to refine the structure. While SA/MD protocols may perform adequately in a data-rich setting, it is difficult to determine protein structures accurately using only sparse data, since SA/MD protocols provide no guarantee on the uniqueness or global optimality of the solutions, and can get trapped in local minima. Sparse data arises not only in high-throughput settings, but also for larger proteins, membrane proteins and symmetric protein complexes. Sparse-data algorithms require guarantees of completeness and correctness to ensure that solutions are not missed and local minima are evaded.

Here we present algorithms that use global orientational restraints from minimal amount of RDCs and RCSAs data and sparse set of NOEs to compute high-resolution protein backbone structure in linear time. We have also developed new algorithms to do simultaneous structure determination and packing of beta-sheets using RDCs and NOEs that involve only amide and methyl protons from ‘H-13C-ILV methyl-labeled proteins. Results from beta-sheet computation for human ubiquitin show that our algorithm can compute beta sheets with backbone RMSD ≤ 0.8 Å from X-ray reference structure (PDB id: 1UBQ). Our tests on different combinations of RDCs for human ubiquitin and 2-Dimensional Staphylococcal Protein A (SpA) demonstrate the ability of our algorithms to compute global backbone fold accurately from sparse data. These results show that our structure determination algorithms and software can be successfully applied to compute the high-resolution protein backbone from sparse NMR data, which can be used to do automated NDE assignments [2].

We are also developing an algorithm for structure determination of the backbone secondary structure elements, loops, and Saupe matrix elements using sparse RDC data and minimal a priori modeling. The data used in this method are NH and CH RDCs in two alignment media. The method is based on the exact computation of all of the local minima of the function determined over the structure parameters and Saupe matrix elements. The minima computed by the algorithm correspond to the structures minimizing the fit to the RDC data subject to some preliminary results obtained using the algorithm. The preliminary results demonstrate that very little NMR data contains enough information to perform truly de novo protein structure determination, but the Saupe matrices is a new and attractive benefit over all of the previous approaches.

Schematic of RDCAnalytic Engine

Backbone Structure From Sparse RDC Restraints and Minimal a Priori Modeling

RDC Fit Function

RDC Data Requirements

Conclusions and Future Work

Our algorithms for structure determination from sparse NMR data have provable guarantees on correctness, completeness, accuracy and time complexity, and require less data compared to other algorithms in the literature. We have been working on the following extensions:

- Extension of RDCAnalytic to incorporate the analytic solutions that has been derived for three planar RDCs and RCSAs and sparse (ambiguous) NOEs.
- Extension of beta sheet computation to compute more curvy strands.
- Algorithms to compute side-chain conformations from RDCs.

References


Funding

This work is supported by the following grant to B.R.D.: National Institutes of Health (R01 GM 65982).