SPROS: An SDP-Based Protein Structure Determination from NMR Data

Babak Alipanahi[1], Nathan Krislock[2], Ali Ghodsi[3], Henry Wolkowicz[4], Logan Donaldson[5], and Ming Li[1⋆]

[1] David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
[2] BiPop Research Group, INRIA Grenoble Rhône-Alpes, France
[3] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
[4] Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
[5] Department of Biology, York University, Toronto, Ontario, Canada M3J 1P3

Protein structure is the key to deciphering its function and biological role. Nuclear Magnetic Resonance (NMR) spectroscopy is one of the major experimental protein structure determination methods. In protein NMR, the 3D structure is determined by making use of a set of distance restraints between proton pairs and exploiting the domain knowledge about proteins. Most of the existing prominent methods are based on torsion angle/Cartesian coordinate molecular dynamics coupled with a complex simulated annealing approach. In these methods, usually an objective function representing the error between observed and given distance restraints is minimized; these objective functions are highly nonconvex and very hard to optimize.

Euclidean geometry methods based on semidefinite programming (SDP) provide a natural formulation for realizing a 3D structure from a set of given distance constraints. However, the complexity of SDP solvers is a major obstacle in their applicability to the protein NMR problem.

We propose a novel SDP-based protein structure determination from NMR data, called SPROS, which is both fast and robust. By using a technique called 'semidefinite facial reduction,' the SDP matrix size and the number of equality constraints are approximately one quarter of the original problem, which results in a fifty-fold decrease in the running time required by the SDP solver. SPROS is applied to proteins with a molecular mass less than 15 kDa and the predicted structures are accurate.

⋆ All correspondence should be addressed to mli@uwaterloo.ca.