

# Lecture 4

## Solution Structures of Native and Denatured Proteins Using RDCs

In the previous lectures, we introduced the principles of the Residual Dipolar Coupling (RDC), which can provide global orientational restraints on internuclear vectors. In this lecture, we introduce some RDC applications, that is, using RDCs to solve the solution structures of native [1, 2] and denatured [3] proteins.

### 1 Determining Exact Protein Native Structure

In the section, we first introduce a quartic equation and two simple trigonometric equations that can compute, *exactly* and *in constant time*, the backbone dihedral angles for a residue from RDCs in two media on any single backbone vector type. Furthermore, based on these exact solutions we introduce a systematic algorithm for determining protein backbone substructure consisting of both  $\alpha$ -helices and  $\beta$ -sheets [1].

#### 1.1 Theoretical Background

Based on the equations of NH RDCs measured in two media, we derived a quartic equation which can compute the vector orientation, for instance, with respect to  $x$ , we have

$$f_4u^4 + f_3u^3 + f_2u^2 + f_1u + f_0 = 0, \quad (1)$$

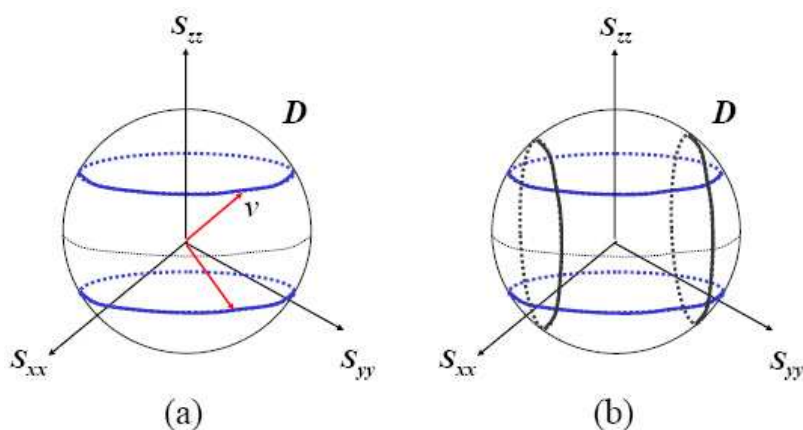
$$u = 1 - 2 \left( \frac{x}{a} \right)^2$$

where the coefficients  $f_0, f_1, f_2, f_3, f_4, a$  as well as  $y$  and  $z$  can be computed from the equations of NH RDCs (Appendix A in [1]). As shown in Figure 1, the number of real solutions of Equation 1 is at most 8, or in other words, given NH RDCs measured in two media, we can obtain up to 8 possible vector orientations, one of which is the real orientation.

In order to compute the dihedral angles for a single residue, we showed that if the directions of any two vectors  $\mathbf{v}_i$  and  $\mathbf{v}_{i+1}$  in consecutive peptide planes  $i$  and  $i+1$  are known, the intervening backbone angles  $(\phi_i, \psi_i)$  can be computed from the following two trigonometric equations:

$$\sin(\phi_i + a_1) = b_1, \quad \sin(\psi_i + a_2) = b_2, \quad (2)$$

where  $a_1, b_1, a_2$  and  $b_2$  are computed from the six angles between two consecutive residues,  $\mathbf{v}_i$  and  $\mathbf{v}_{i+1}$ . The full expressions for all these four coefficients are computed from backbone kinematics (Appendix B in [1]). Note that these two equations can be solved exactly for



**Figure 1:** An example of extracting NH unit vector using RDCs in two media: (a) unit vector determined by one RDC; (b) unit vector determined by two RDCs. Consequently, 0, 2, 4, 6 or 8 solutions can be obtained from 2 RDCs.

$\sin \phi_i$  and  $\sin \psi_i$ , and therefore 2 possible  $(\phi_i, \psi_i)$  solutions exist for each orientation of the pair  $(v_i, v_{i+1})$  of bond vectors.

Furthermore, if the first peptide plane is given, we can in principle compute all the possible  $(\phi, \psi)$  angles of a fragment based on the following two observations: (1) A peptide plane  $i$  with respect to a principle order frame (POF) for medium 1 can be determined by its  $\text{NC}_\alpha$  vector and an NH vector. (2) A unique  $\text{NC}_\alpha$  vector for the peptide plane  $i + 1$  can be computed from the  $\text{NC}_\alpha$  vector of the peptide plane  $i$  and  $(\phi_i, \psi_i)$ . Therefore, we can consecutively compute all the possible discrete  $(\phi, \psi)$  solutions for all the residues of the fragment by using a depth-first search (DFS) strategy.

## 1.2 The algorithm

The algorithm of determining a 3-dimensional backbone substructure with the  $\alpha$ -helices and  $\beta$ -sheets employs the following inputs: (1) assigned backbone NH RDCs in two media, (2) identified  $\alpha$ -helices and  $\beta$ -sheets with known H-bonds between paired strands, (3) very sparse NOE distance restraints. The algorithm is divided into three strategies:

1. **Computation of alignment tensors:** Using an ideal helix model built with the backbone, the Saupe matrices for both media can be computed from the model by a SVD method. Then, an optimal first  $\text{NC}_\alpha$  vector is computed using an  $m$  residue fragment built with the average dihedral angles from either an  $\alpha$ -helix or a  $\beta$ -strand, and standard bond lengths and angles using a grid search method.
2. **Refinement of secondary structure elements:** Give the fact that if the orientation of the first peptide plane of an  $m$  residue secondary structure element is given, the conformation can be specified uniquely by a sequence of backbone dihedral angles. Here, we refine the fragment by optimizing both the direction of individual NH vector

and also the dihedral angles of the fragment using RDCs alone while leaving the bond lengths and the six angles fixed. Furthermore, The plausible conformations are computed from a DFS-based refinement method, in which every computed  $(\phi, \psi)$  pair is filtered through the favorable Ramachandran regions for the corresponding secondary structure type. Consequently, an optimal conformation is computed from the set of all the plausible conformation vectors using the scoring function defined in [1].

3. **Backbone structure determination:** We employ a few NOE-derived distances, including the distances between an amide proton and a  $C_\alpha$  nucleus or between two  $C_\alpha$  nuclei, to compute the relative position of the helix and the single sheet of ubiquitin.

### 1.3 Results

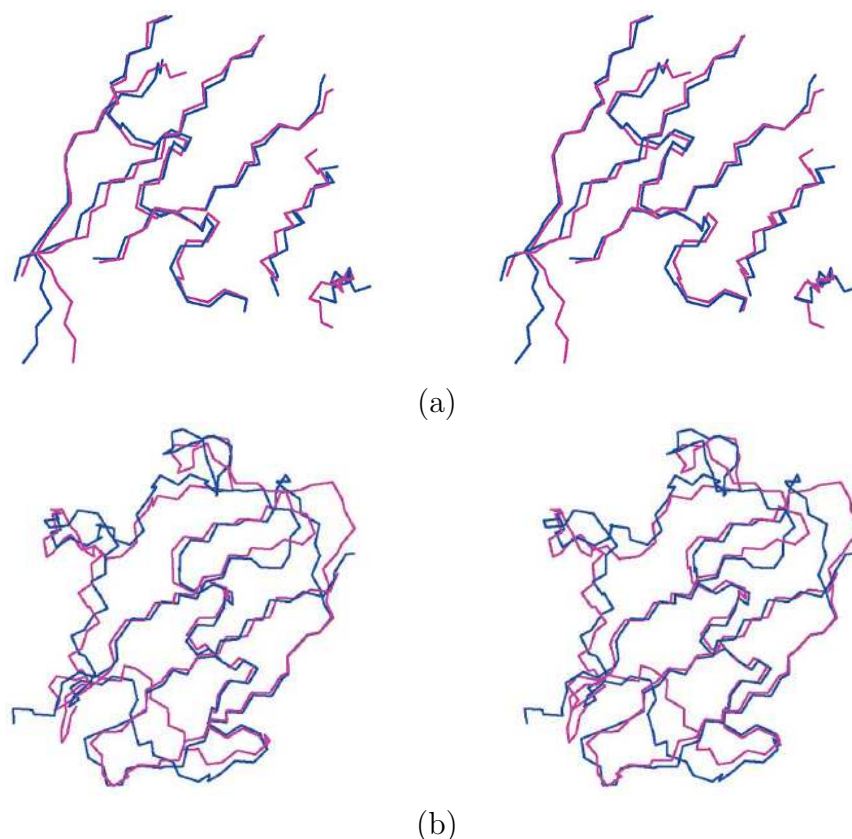
The work [2] has been successfully extended to compute a complete backbone structure, including turns and loops (connecting the secondary structure elements) using only NH and CH RDCs in a single medium (i.e., only two RDCs per residue) and two unambiguous NOEs. Figure 2a shows the structure of ubiquitin backbone without loops, while Fig 2b compares the structures of ubiquitin backbone with loops.

## 2 Determination of Denatured or Disordered Proteins

In this section, we introduce a data-driven algorithm capable of computing a set of structures for denatured or disordered proteins directly from sparse experimental restraints, including the orientational restraints from RDC, and distance restraints from paramagnetic relaxation enhancement (PRE) experiments.

### 2.1 A probabilistic interpretation of restraints in the denatured state

In contrast to the traditional algorithms for the structure determination of native proteins, [3] employs a *set* of tensors to interpret the RDCs measured in the denatured state. Each tensor in the set represents a cluster of similar denatured structures, and the set of RDCs corresponding to each tensor is sampled from the individual distributions associated with each measured RDC. That is, the experimentally-measured RDC value in the denatured state is the expectation, and the different clusters of tensors represent different conformations that are oriented differently in the aligning medium. Similarly, the PRE-derived distance is also considered a random variable, where the measured value is an average over all the possible structures in the denatured state. Hence, the structure determination problem for denatured proteins can be formulated as the computation of a set of conformation vectors, given the distributions of all the RDCs and PREs.

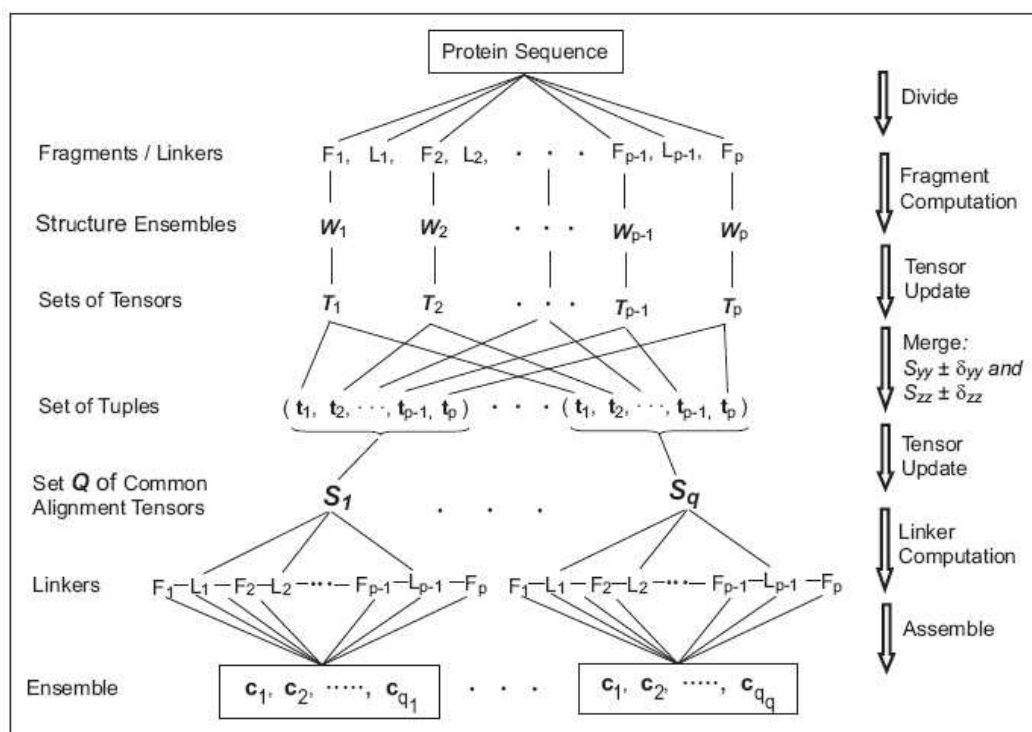


**Figure 2:** (a) Structure of ubiquitin backbone without loops. The ubiquitin backbone structure (blue) was computed using 37 NH and 39 CH RDCs, 12 hydrogen bonds, and four NOEs. (b) Structure of ubiquitin backbone with loops. The ubiquitin backbone structure (blue) was computed by extending our algorithm to handle loop regions along the protein backbone. The structure was computed using 59 NH and 58 CH RDCs, 12 H-bonds and two unambiguous NOEs.

## 2.2 The algorithm

For a denatured protein, the algorithm in [3] computes a presumably heterogeneous *ensemble* of structures that are consistent with the experimental data within a large range. The input of the algorithm includes the protein sequence, at least two RDCs per residue in a single medium and PREs. For efficiency, a divide-and-conquer strategy is used in the algorithm to compute the ensemble, which consists of six steps (Figure 3):

- (1) **Fragment division:** The entire protein sequence is divided into  $p$  fragments  $F_{0,\dots,i}$  and  $p-1$  linkages  $L_{0,\dots,i}$ , where a linker consists of the residues between two neighboring fragments.
- (2) **Fragment computation:** the algorithm computes an ensemble of structures  $W_i$  for each fragment  $F_i$ , independently.



**Figure 3: Divide-and-conquer strategy.** The terms  $c_i$  denote conformation vectors for the complete backbone structure.

- (3) **Tensor computation:** for each structure in ensemble  $W_i$ , the corresponding tensor  $t_i$  is computed by SVD and save into a set  $T_i$ . The SVD procedure is to minimize the RMSD between the experimental RDC data and the RDC back-computed from the structure using the tensor.
- (4) **Tensor merge:** all the tensors in the sets  $T_i$  are merged into  $p$ -tuples,  $(t_1, \dots, t_p)$ , where  $t_i$  is from the set  $T_i$ . All the  $p$  tensors in a  $p$ -tuple have their  $S_{yy}$  and  $S_{zz}$  (two of three diagonal elements of the diagonalized Saupe matrix) values in the certain ranges of  $[S_{yy} - \delta_{yy}, S_{yy} + \delta_{yy}]$  and  $[S_{zz} - \delta_{zz}, S_{zz} + \delta_{zz}]$ , where  $\delta_{yy}$  and  $\delta_{zz}$  are thresholds.
- (5) **Tensor update:** for each merged  $p$ -tuple, the algorithm computes their common tensor by SVD using the corresponding structures in  $W_i$  and all the experimental RDCs for the fragments  $F_i$ . Note that the SVD computation can output not only the diagonal elements  $S_{xx}, S_{yy}, S_{zz}$ , but also the orientation for each fragment in the common POF as well.
- (6) **Linker computation and assembly:** the algorithm computes each linker  $L_i$  using every common tensor and assembles the corresponding fragments and linkers into complete backbone structures.

## 2.3 Applications to real biological systems

The algorithm described in [3] has applied to compute the structure ensembles of two proteins, an acid-denatured denatured ACBP and a urea-denatured eglin C, from real experimental NMR data. For acid-denatured ACBP, an ensemble of 231 structures has been computed at pH 2.3, and all the structures have no vdW repulsion larger than 0.1Å except for a few vdW violations as large as 0.35 Å between the two nearest neighbors of a proline and the proline itself. Further analysis indicates that the acid-denatured ACBP is neither random coil nor native-like. The work [3] also draw a similar conclusion for the urea-denatured eglin C, for which an ensemble of 160 structures was computed at 8 M urea.

## References

- [1] Wang. L, Donald. BR. Exact solutions for internuclear vectors and backbone dihedral angles from NH residual dipolar couplings in two media, and their application in a systematic search algorithm for determining protein backbone structure. *Jour. Biomolecular NMR*, 29(3):223-242, 2004.
- [2] Wang. L, Mettu. RR, Donald BR. A Polynomial-Time Algorithm for De Novo Protein Backbone Structure Determination from NMR Data. *Journal of Computational Biology*, 13(7): 1276-1288, 2006.
- [3] Wang. L, Donald BR. A Data-Driven, Systematic Search Algorithm for Structure Determination of Denatured or Disordered Proteins. *The Computational Systems Bioinformatics Conference (CSB)*, Stanford CA. (August, 2006) Pages 67-78. ISBN 1-86094-700-X.