A Note on Computing hSRI

This note describes computing the structure of hSRI using the newest distribution of the software RDC-ANALYTIC.

Computing SSEs of hSRI using RDC-ANALYTIC

In hSRI there are three α-helices which are denoted here by H_1, H_2 and H_3 with residue boundaries [Glu13, Leu34], [Glu51, Tyr71] and [Glu82, Met96], respectively. For H_1, H_2 and H_3 respectively 33, 36, and 22 RDCs (N-H^N and C^α-H^α RDCs in total) are present for 44, 42, and 30 internuclear vectors (N-H^N and C^α-H^α internuclear vectors in total). RDC-ANALYTIC was used to compute the conformations and orientations of these three helices. Since RDC-ANALYTIC is a program that computes the backbone secondary structure element (SSE) conformations and orientations using only RDCs, it generally performs well when at least 90% of the N-H^N and C^α-H^α RDCs are present. However, hSRI has many RDCs (more than 20%) missing, and therefore posed a difficult but interesting case for computing the SSEs.

RDC-ANALYTIC does not randomly search the conformation space to find solutions consistent with the RDC data. Rather, it formulates the problem such that the structures computed are exact solutions of a system of quartic monomial equations derived from the RDC equation. Hence, these roots, and therefore the conformations, are discrete, finite and algebraic. All dihedral angles for each residue are solved exactly from the quartic RDC equations. A depth-first systematic search algorithm is applied to search over all possible roots (conformations) to find an optimal solution for a SSE.

When an RDC is missing, the algorithm substitutes a synthetic RDC derived from the corresponding internuclear vector in an ideal helix during the initial phase of computation, which requires fitting an ideal helix with the RDC data given the diagonalized alignment tensor components. Subsequent phase(s) involve computing a maximum likelihood solution

(a helix) over all sampled set of RDCs and then using that solution, the RDC data, and the diagonalized alignment tensor to obtain the oriented helix corresponding to the best-fit with the RDC data. Then the first peptide plane of the helix is extracted and used in the next phase with the sampled sets of RDCs to compute a maximum likelihood solution.

Our experience says that such a choice of substitution for missing RDCs usually works well when a few RDCs are missing. However, it might not always work, especially, when more than 20% of RDCs are missing as in the case of hSRI. Some heuristic techniques may be used in such situations.

For hSRI, during the computation of H_2 our algorithm found no real solution for dihedral angles within the Ramachandran region for helix for the residue Thr61. Upon replacing the C^{α}-H^{α} RDC for residue Leu60 with a synthetic RDC derived from ideal helix, the algorithm then was able to compute the conformation and orientation of H_2. We think such error accumulation (at residue Thr61) is partly due to the inaccuracy in the orientation of the first peptide plane of H_2 determined (as described above) when 6 RDCs are missing, and partly due to the decision made earlier to replace the missing RDCs (e.g., C^{α}-H^{α} RDC for Arg58 and N-H^{N} RDC for Lys59) by synthetic RDCs derived from ideal helix. Also, in a general scenario a single big outlier in RDC data can cause such a problem. Similarly, while computing H_3 (for which more than 25% of RDCs are missing) C^{α}-H^{α} RDC of Val84 and N-H^{N} RDC of Lys85 were replaced with a synthetic RDC derived from ideal helix which allowed the algorithm to compute the conformation and orientation of H_3.

The helices H_1, H_2 and H_3 computed by RDC-ANALYTIC have the respective backbone RMSDs 0.96 Å, 0.95 Å and 0.63 Å vs. the first model of the NMR reference structure (PDB ID: 2A7O).

Computing the Loops of hSRI

When computing the long loops between residues [Leu34, Glu51] and residues [Tyr71, Glu82], we first compute a set of ambiguous NOE assignments for these two loops using only the chemical shift information. These ambiguous NOE assignments are computed as follows. We consider all pairs of protons that are possibly assigned to an NOE cross peak if the resonances of corresponding atoms fall within a tolerance window around each NOESY peak. We use error windows of 0.06 ppm for heavy atoms (15N and 13C), and 0.03 ppm for protons in assigning these ambiguous NOE assignments. All ambiguous NOE assignments for each NOESY peaks are unified using the logical "OR" operation. For more details please see [1] (page 268). Below shows an example of an ambiguous NOE assignment:

```
! peak ID : 2267
assign (( resid 102 and name HA ))
      (( resid 36 and name HG1 ) or
       ( resid 36 and name HG2 ) or
       ( resid 44 and name HB ))    6.0   4.2  0.0
```

In this example, three ambiguous NOE assignments exist for the NOESY peak with ID 2267, namely, 102H$^{\alpha}$-36H$^{\gamma_1}$, 102H$^{\alpha}$-36H$^{\gamma_2}$ and 102H$^{\alpha}$-44H$^{\beta}$, since all their corresponding pairs of atom chemical shifts are within the error windows from the NOESY peak 2267.

Above procedure of NOE assignment using only the chemical shift information has been implemented as an RDC-PANDA command. More specifically, these ambiguous NOE assignments can be assigned using the following RDC-PANDA command:

```
RDC-PANDA> NOEAsgFromCS asg_noe_cs.input
```

In the input parameter file `NOEAsgFromCS asg_noe_cs.input`, which is used to by the above command, the flag `isOutORFormat` is set to be equal to 1.

All these ambiguous NOE assignments unified in "OR" format are supported by XPLOR, and used in the local minimization approach to compute the long loops between residues [Leu34, Glu51] and residues [Tyr71, Glu82].

References

[1] Jianyang Zeng, Jeffrey Boyles, Chittaranjan Tripathy, Lincong Wang, Anthony Yan, Pei Zhou, and Bruce Randall Donald. High-resolution protein structure determination starting with a global fold calculated from exact solutions to the RDC equations. *Journal of Biomolecular NMR*, 45(3):265–281, 2009.