The process of assignment validation and repositioning begins with the sequence-specific resonance assignment of intrinsically disordered proteins (IDPs). The assignment plugin greatly facilitates the effective matching of a set of connected resonances to the correct position in the sequence by making use of IDP random coil chemical shifts. The superior predictive power of ncIDP-assign translates into detection sensitivity of chemical shift deviations from the sequence-specific 'random-coil' values due to resonance misassignment.
Table 1. Comparative analysis of chemical shift back-computation for hNLG3cyt using standard chemical shift computation module available in SPARKY, and ncIDP-assign

| Nucleus (n) | Standard$^a$ | ncIDP$^a$ | Assignments |
|------------|-------------|------------|-------------|
| $^{15}$N   | 0.122       | 0.044      | 120         |
| $^{13}$C$^\alpha$ | 0.148       | 0.111      | 128         |
| $^{13}$C$^\beta$ | 0.910       | 0.324      | 128         |
| $^{13}$C$^\gamma$ | 0.410       | 0.212      | 120         |
| $^{13}$C$^\delta$ | 0.717       | 0.393      | 131         |
| $^{13}$C      | 1.314       | 0.664      | 129         |

$^a$Chemical shift RMSD computed after removal of mean systematic offsets between the computed and experimental resonance assignments for hNLG3cyt in order to minimize chemical shift referencing errors. Root mean square difference (RMSD) values are given in ppm.

Another critical parameter in the assignment process is the relative separation of the ‘best-fit’ score $S_{\text{best}}$ with respect to the second-best scoring suggestion $S_{\text{second-best}}$, expressed here as the sensitivity $S_{\text{second-best}}/S_{\text{best}}$ (Fig. 1B). Values for the sensitivity close to 1.0 indicate ambiguity. Given a sequentially assigned dipeptide, ncIDP-assign generates a list of solutions in which the ‘best-fit’ scenario scores appreciably better than the next considered option. Already significant improvements are observed in the analysis of $^1$H-$^{15}$N HSQC resonance lists, clearly demonstrating that information content of $^1$H-$^{15}$N resonance pairs in sequentially connected dipeptides can be effectively used in the assignment process of an intrinsically disordered protein.

In conclusion, we have shown here that ncIDP-assign is an effective tool to aid the sequential NMR resonance assignment of (intrinsically) disordered proteins.

ACKNOWLEDGEMENTS

Renee Otten (University of Groningen) and Thomas Goldfarb (University of California, San Francisco) are gratefully acknowledged for help with Python code development and valuable suggestions.

Funding: Netherlands Organization for Scientific Research (NWO) (VIDI grant to F.A.A.M.).

Conflict of Interest: none declared.

REFERENCES

Bormel, W. et al. (2006) Proteinless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. J. Am. Chem. Soc., 128, 3918–3919.

Boekhoven, G. and Ruben, D.J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. Chem. Phys. Lett., 80, 185–189.

Dijkstra, K. et al. (1994) The COCAHA experiment to correlate intrinsically disordered (c)[alpha] and h[alpha] resonances in proteins. J. Magn. Reson. Ser. A, 107, 102–105.

Elleman, E.D. (2009) Biophysical characterization of intrinsically disordered proteins. Curr. Opin. Struct. Biol., 19, 23–30.

Goldfarb, T.D. and Keeler, D. (2006) SPARKY 3. University of California, San Francisco.

Mantylä, S. et al. (2010) HA-detected experiments for the backbone assignment of intrinsically disordered proteins. J. Biomol. NMR, 47, 171–181.

O’Hare, B. et al. (2009) Incorporating $^1$H chemical shift determination into $^{13}$C-direct detected spectroscopy of intrinsically disordered proteins in solution. J. Magn. Reson., 200, 354–358.

Paz, A. et al. (2008) Biophysical characterization of the unstructured cytoplasmic domain of the human neuronal adhesion protein neurolgin. J. Biophys., J., 95, 1926–1944.

Tamiola, K. et al. (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. J. Am. Chem. Soc., 132, 18000–18003.

Weiterkant, M. and Mudde, J. (1995) HNCA/CB: a High-Sensitivity 3D NMR experiment to correlate Amide-Proton and nitrogen resonances with the alpha - and Beta-Carbon resonances in proteins. J. Magn. Reson. Ser. B, 101, 204–205.

Fig. 1. Comparative analysis of accuracy of standard and ncIDP SPARKY sequence repositioning extensions, using resonance assignments in $^1$H-$^{15}$N HSQC, COCAHA and HNCACB experiments, for the intrinsically disordered protein hNLG3cyt. (A) Normalized frequency of correct repositioning solutions as a function of fragment length for: standard (blue) and ncIDP-enhanced (green) repositioning, respectively. (B) Sensitivity as a function of sequence length for: standard (blue) and ncIDP (green) repositioning extensions.

This point is demonstrated by Fig. 1 where the repositioning performance of the two methods is compared against known resonance assignments for hNLG3cyt. ncIDP-assign identifies correct solutions much more readily, and already at the level of dipeptides for the considered experiments (Fig. 1A). Further expansion of the length of a query fragment to tripeptide rapidly shifts the probability of assigning the correct solution. Consequently, the information content contained within a combination of resonance frequencies in short peptides (length $\geq 2$) is unique enough to make the correct position guess in most cases.