ncIDP-assign: a SPARKY extension for the effective NMR assignment of intrinsically disordered proteins

Kamil Tamiola* and Frans A. A. Mulder*

Department of Biophysical Chemistry, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 7, Groningen 9747AG, The Netherlands

Associate Editor: Jonathan Wren

ABSTRACT

Summary: We describe here the **ncIDP-assign** extension for the popular NMR assignment program SPARKY, which aids in 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.

Availability: The ncIDP-assign extension is available at http://www.protein-nmr.org/.

Contact: k.tamiola@rug.nl; f.a.a.mulder@rug.nl; support@protein-nmr.org

1 INTRODUCTION

The structural characterization of intrinsically disordered proteins (IDPs) is a rapidly growing field in structural molecular biology. Over the past decade NMR spectroscopy has proven to be singular in its capacity to provide detailed structural characterization of these dynamic entities (Eliezer, 2009). Although diverse experimental approaches have been developed (Bermel *et al.*, 2006; Mäntylahti *et al.*, 2010; O'Hare *et al.*, 2009), the sequential resonance assignment of >10 kDa natively unfolded polypeptides is still not a trivial task.

Here, we describe an enhanced version of the SPARKY (Goddard and Kneller, 2006) sequence repositioning extension, which assist in matching, connecting and assigning consecutive residues, and is specifically designed for intrinsically disordered proteins: **ncIDP-assign** (neighbor-corrected IDP chemical shift assignment). This tool makes use of a novel random-coil chemical shift library, enabling the accurate prediction of the chemical shifts of a queried protein on a basis of tripeptides (Tamiola *et al.*, 2010). Predicted sequence-specific chemical shifts are used as a template for re-assignment and validation of existing resonance assignments. The newly designed **ncIDP-assign** greatly accelerates the process of sequential resonance assignment of large intrinsically disordered proteins by drastically reducing the level of assignment ambiguities.

2 IMPLEMENTATION

The process of assignment validation and repositioning begins with the computation of the sequence-specific random-coil chemical

shifts for the intrinsically disordered protein under study, based on its primary sequence. The random-coil chemical shift for a nucleus $n \in \{^1H^\alpha, ^1H^N, ^{13}C^\alpha, ^{13}C^\beta, ^{13}C^O, ^{15}N\}$ of amino acid residue a, within a tripeptide $x - a - y$, can be expressed as,

$$\delta^n_{pred}(x,a,y) = \Delta^n_{-1}(x) + \delta^n_{rc}(a) + \Delta^n_{+1}(y) \tag{1}$$

where $\delta^n_{rc}(a)$ is the 'random-coil' chemical shift in the reference peptide $Gly - a - Gly$, $\Delta^n_{-1}(x)$ and $\Delta^n_{+1}(y)$, are the neighbor corrections due to the preceding and the sequential residue, respectively. Subsequently, a collection of experimentally assigned resonances in a range $k \ldots k+l$ specified by the user is retrieved and compared against the predicted chemical shifts from Equation (1). All plausible locations for the fragment along the sequence of a protein are considered. Each investigated position is characterized with a chemical shift deviation score S,

$$S(k,l) = \sum_{i=k}^{k+l} \sum^{n} Z^n(i) \tag{2}$$

where $Z^n(i) = \left| \frac{\delta^n_{exp}(i) - \delta^n_{pred}(i)}{\sigma^n_{pred}(i)} \right|$, δ^n_{exp} and σ^n_{pred} are experimental chemical shifts and the standard deviations of expected chemical shieldings for $k \ldots k+l$ sequence combination, respectively. A score of 1.0 implies that at the queried sequence position, the experimental shifts are one standard deviation from the computed values, on average. Hence, the 'best-fit' solutions are characterized by the lowest S values.

3 RESULTS AND CONCLUSION

To assess the performance of **ncIDP-assign** we used the NMR chemical shift assignments for the 140-residue intrinsically disordered, cytoplasmic domain of human neuroligin-3 (hNLG3cyt) (Paz *et al.*, 2008). The robustness of the assignment procedure was established using the data obtained in ^1H-^{15}N HSQC (Bodenhausen and Ruben, 1980), COCAHA (Dijkstra *et al.*, 1994) and HNCACB (Wittekind and Mueller, 1993) experiments, providing access to backbone and $^{13}C^\beta$ chemical shifts. Table 1 displays the level of completeness of the resonance assignment for hNLG3cyt, and the accuracy of the chemical shift back computation modules in the standard and ncIDP versions of the SPARKY sequence repositioning plugins. As borne out by Table 1, **ncIDP-assign** offers an almost two-fold improvement in the estimation of chemical shifts for hNLG3cyt. 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.

---

*To whom correspondence should be addressed.

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
$^1H^\alpha$	0.122	0.044	120
$^1H^N$	0.148	0.111	128
$^{13}C^\alpha$	0.810	0.324	128
$^{13}C^\beta$	0.410	0.212	120
$^{13}C^O$	0.717	0.393	131
^{15}N	1.314	0.664	129

[a]Chemial 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.
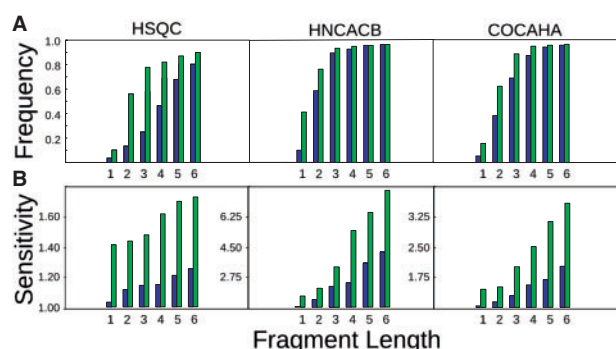
Fig. 1. Comparative analysis of accuracy of standard and ncIDP SPARKY sequence repositioning extensions, using resonance assignments in: ^{1}H-^{15}N HSQC, COCAHA and HNCACB experiments, for the intrinsically disordered protein hNLG3cyt. (**A**) Normalized frequency of correct repositioning solutions as a function of fragment length in: 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 ≥ 2) is unique enough to make the correct position guess in most cases.

Another critical parameter in the assignment process is the relative separation of the 'best-fit' score S_{best} with respect to the second-best scoring suggestion $S_{second-best}$, expressed here as the sensitivity $S_{second-best}/S_{best}$ (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.

REFERENCES

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

Bodenhausen,G. and Ruben,D.J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.*, **69**, 185–189.

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

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

Goddard,T.D. and Kneller,D. (2006) SPARKY 3. University of California, San Francisco.

Mäntylahti,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 1H chemical shift determination into 13C-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 neuroligin 3. *Biophys. J.*, **95**, 1928–1944.

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

Wittekind,M. and Mueller,L. (1993) HNCACB, 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**, 201–205.