Open Access MAGNETIC RESONANCE Discussions

Magn. Reson. Discuss., referee comment RC2
https://doi.org/10.5194/mr-2022-22-RC2, 2023

Comment on mr-2022-22

Anonymous Referee #2

Referee comment on "Facilitating the structural characterisation of non-canonical amino acids in biomolecular NMR" by Sarah Kuschert et al., Magn. Reson. Discuss., https://doi.org/10.5194/mr-2022-22-RC2, 2023

This manuscript describes a largely automatic procedure for incorporating non-canonical amino-acid residues (ncAAs) and other post-translational modifications (PTMs) into NMR-based structure calculations using the programs CYANA and CNS. The procedure builds on a previously published webserver-based facility called the Automated Topology Builder (ATB) that creates optimised geometries, topology files and force-field parameters for independent molecules starting from a set of (non-optimised) 3D co-ordinates. The new aspects described in the present manuscript are concerned with extensions that can be used with the ATB to facilitate building structures in a protein context, i.e. allowing for partial structures to be attached in place of a canonical sidechain or at one of the polypeptide chain termini. A significant part of this process is the development of software for naming atoms within ncAAs or PTMs in a standardised fashion consistent with IUPAC recommendations; this might sound like an bureaucratic nicety, but inconsistencies in atom naming can all too easily create tedious obstacles to the development or sharing of methodology or even results. These are important issues, and I believe designing a generally applicable scheme to make the setting up of such calculations more convenient and standardised is a very worthwhile goal. I therefore welcome this contribution and support its publication, subject to attention being paid to the following points:

1) While using successive letters from the Greek alphabet to indicate the (smallest) number of bonds separating a sidechain atom from the protein mainchain in an ncAA is clearly consistent with the spirit of the IUPAC recommendations, it seems to me that formally speaking this is a step beyond the IUPAC recommendations themselves that, as far as I am aware, refer only to atoms in the 20 canonical amino acids. I don't believe the two references for the IUPAC recommendations given in the manuscript (Huang et al., 1970 and Markley et al., 1998) describe this additional step to include atom naming in ncAAs; are there other references that could be cited that would formally set such a precedent?

2) The Greek alphabet contains only 24 letters, and while this is doubtless sufficient for

uniquely naming atoms in the great majority of ncAAs or PTMs following the approach outlined in this paper, it is not sufficient in all cases. For instance, glycans containing linear chains of more than 4 sugars would exceed a 24-atom chain-length limit, and the structure of glycosylphosphatidylinositol (GPI) anchors, which are a not uncommon form of N-terminal PTM, involve more than 40 bonding steps from the polypeptide backbone. Do the IUPAC recommendations have anything to say about atom naming in such cases? Assuming they do not, what do the present authors propose in such cases when the Greek alphabet runs out of letters?

3) In some of the more detailed sections of the manuscript it becomes apparent that some steps in creating files to represent complicated structures do require some manual intervention, e.g. to complete the bonding scheme for some ring structures. I can see that it may well be very much more difficult to write software that correctly handles such complications fully automatically, and I don't believe that the need for manual intervention to complete the implementation in such cases is necessarily much of a problem, but I do feel the issue should be more clearly discussed in the manuscript. The authors could comment briefly on whether they are planning to attempt the automatic handling of such remaining cases, or whether that would be impractical. I think it would also be helpful to add a short but clear statement in a rather more visible location in the paper as to what are the fundamental limitations on fully automatic operation in the present implementation.

4) It is probably inevitable that the implementation of a new approach such as this is rooted in the environment of the particular program using which it was developed, in this case CYANA. However, the transfer of the approach to a different program environment is important if the method is to be widely adopted, as presumably the authors of this contribution hope it will be. It may not be practical to go very far down this road, but it might have been nice to see the method worked through using, for instance, XPLOR-NIH, ARIA or AMBER. Are there steps for which the use of CYANA or associated programs is currently unavoidable?