Open Access MAGNETIC RESONANCE Discussions

Magn. Reson. Discuss., author comment AC2
https://doi.org/10.5194/mr-2022-22-AC2, 2023

Reply on RC2

Sarah Kuschert et al.

Author 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-AC2, 2023

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?

The reviewer is correct, IUPAC have not made a specific recommendation for such cases. We invited IUPAC to review our paper and provide comments on our approach. They responded by confirming that no convention exists and suggested additional references to IUPAC documents, which have been included.

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?

We have expanded the standard nomenclature from 24 to (600) by using 2 letter symbols from position 25 (AA, AB, … AW | BA, BB, … BW |…| WA, WB, .. WW). An explanatory sentence has been added to the manuscript. We have also highlighted the problem to IUPAC. Note our code on GitHub is intended to be dynamic and will be amended to match any official recommendation by IUPAC.

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.

A paragraph has been added to the discussion regarding both the existing limitations and future directions.

"The manual modifications that were noted in specific cases are largely due to limitations and compatibility issues of the tools used in the pipeline. The treatment of ring closures by CYLIB has been noted previously. The need for a torsion angle between connected aromatic rings is a consequence of existing rules within CYLIB. The limitations within the import function in CcpNmr is subject of current development by the authors of that software. The manual steps required for compatibility with CNS are largely due to the use of atom types to define both Lennard-Jones parameters as well as the bonded terms. Name clashes in the atom type definitions that arise from combining multiple ATB generated building blocks within a single system must be addressed to ensure the intended parameters are used. Further refinement of the ATB outputs to improve compatibility with different packages (e.g., NIH-Xplore, AMBER, ARIA) will be the subject of future work."

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?

In section 2.6 we outline how we use CNS to perform water refinement. The topology and parameter files required by CNS are in principle compatible with XPLOR-NIH, ARIA and AMBER (i.e. these can be used for cartesian based structure calculations without the need to use CYANA). The additional paragraph in the discussion now also emphasise this.