Defining distance restraints in HADDOCK

To the Editor — In a recent issue of *Nature Protocols*, Orbán-Németh et al.[1] present a protocol to predict structural models of proteins and their complexes from mass spectrometry (MS) cross-linking data. We read the protocol with interest, as it uses third-party software including the HADDOCK web portal[2] (http://haddock.science.uu.nl/) that we developed and maintain. While we endorse and encourage the inclusion of our software in other protocols and pipelines, it is important that its usage be accurately and correctly described to avoid problems and incorrect results that we, as primary developers, will have to troubleshoot.

Distance restraints are implemented in HADDOCK to force groups of atoms to be at a specific distance from each other. As stated in our *Nature Protocols* paper describing the web server[3], a distance restraint is defined using Crystallography and NMR system (CNS)[4] syntax, by two atom selections followed by three numbers —the target distance (d^0), a lower margin (d^-) and an upper margin (d^+). These three numbers are used to define a distance range, by subtracting and adding, respectively, the lower and upper margins to the target distance. Within this range, the potential energy of the restraint is zero. Further, this flexible syntax allows for the same distance range to be expressed differently, with practically no implications for the quality of the final models.

In step 21 of their protocol, Orbán-Németh et al.[1] erroneously describe how to define distance restraints, which can have severe consequences for the resulting models. Specifically, the lower and upper (d^-, d^+) distance margins are swapped in their definition. In their examples, reproduced below, the authors intend to give distance restraints with ranges of 0–35 Å and 0–23 Å, respectively. Instead, their syntax results in distance ranges of 0–18 Å and 0–12 Å, which are substantially shorter than the maximum cross-linker distance.

```
assign (resid 152 and segid B)
(resid134 and segid A) 18 35 0
assign (resid 152 and segid B)
(resid137 and segid A) 18 35 0
assign (resid 235 and segid B)
(resid147 and segid A) 12 23 0
```

This change in the distance range impacts the energy landscape of the system and can ultimately lead to different, possibly incorrect, models. The correct syntax should be

```
assign (resid 152 and segid B)
(resid 134 and segid A) 35 35 0
assign (resid 152 and segid B)
(resid 137 and segid A) 35 35 0
assign (resid 235 and segid B)
(resid 147 and segid A) 23 23 0
```

or, the following, in an example (out of the many possible combinations) of using both upper and lower margins to achieve the same distance range.

```
assign (resid 152 and segid B)
(resid 134 and segid A) 18 18 17
assign (resid 152 and segid B)
(resid 137 and segid A) 18 18 17
assign (resid 235 and segid B)
(resid 147 and segid A) 12 12 11
```

Finally, in previous publications, we have used similar protocols to model protein–protein complexes with MS cross-linking data (for example, interactome-wide docking[5], GPCR–GRK docking[6] and protein–RNA docking[7]). In these publications, we define distance restraints between specific atoms on the cross-linked residues (for example, C_α–C_α or C_β–C_β) and assert only a maximum distance based on the length of the extended linker, to ensure an easier interpretation of the distance restraint value. We have also provided an online tutorial about the use of cross-linking data in HADDOCK (http://www.bonvinlab.org/education/HADDOCK-Xlinks/). ❐

Alexandre M. J. J. Bonvin[1]*, Ezgi Karaca[2], Panagiotis L. Kastritis[3] and João P. G. L. M. Rodrigues[4]

[1]*Computational Structural Biology, Utrecht University, Utrecht, The Netherlands.* [2]*Bioinformatics Unit, Izmir Biomedicine and Genome Center, Izmir, Turkey.* [3]*Structural and Computational Biology Unit, EMBL-Heidelberg, Heidelberg, Germany.* [4]*Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, USA.*
*e-mail: a.m.j.j.bonvin@uu.nl

References
1. Orbán-Németh, Z. et al. *Nat. Protoc.* **13**, 478–494 (2018).
2. van Zundert, G. C. P. et al. *J. Mol. Biol.* **428**, 720–725 (2016).
3. de Vries, S. J., van Dijk, M. & Bonvin, A. M. J. J. *Nat. Protoc.* **5**, 883–897 (2010).
4. Brunger, A. T. *Nat. Protoc.* **2**, 2728–2733 (2007).
5. Kastritis, P. L. et al. *Mol. Syst. Biol.* **13**, 936 (2017).
6. Komolov, K. E. et al. *Cell* **169**, 407–421 (2017).
7. Karaca, E., Rodrigues, J. P. G. L. M., Graziadei, A., Bonvin, A. M. J. J. & Carlomagno, T. *Nat. Methods* **14**, 897–902 (2017).

Competing interests
The authors declare no competing interests.

Reply to 'Defining distance restraints in HADDOCK'

Orbán-Németh et al. reply — In our recently published protocol in *Nature Protocols*, we provide instructions on how to use cross-linking mass spectrometry (XL-MS) data to predict structural models of proteins and their complexes[1]. In our protocol, we describe the use of several software programs, including HADDOCK[2], which is used