Structure Modeling

DOI: 10.1002/anie.201201783

# Efficient Modeling of Symmetric Protein Aggregates from NMR Data**

*Benjamin Bardiaux, Barth-Jan van Rossum, Michael Nilges,* and Hartmut Oschkinat**

An estimated 60% of genes in every genome code for homo-oligomeric proteins.[1] The symmetry ambiguity, namely the necessity to distinguish intra- and intermolecular correlations, is a major challenge in NMR spectroscopic structure determination of such oligomeric symmetric proteins. Next to the size limitation, this is the main reason that homo-oligomeric structures are severely under-represented among the structures solved by NMR spectroscopy (Supporting Information, Figure S1). So far, algorithmic support is based on ambiguous distance restraints (the symmetry-ADR method[2]), but is limited to simple types of symmetry, and calculations do not always converge. Other methods based on branch-and-bound algorithms have been proposed, but they rely on the knowledge of the protomer structure, which is not known in structure determination projects.[3]

Herein, we present a general method for structure calculation of symmetric aggregates that is valid for both non-crystallographic and crystallographic symmetries (including the lattice translational periodicity). In the context of NMR applications, protein complexes with point symmetries (cyclic or dihedral) as well as symmetries found in fibrils and pili are a major focus. The method has implications well beyond NMR structure determination for modeling of homomeric complex structures from any distance information,[4] as obtained for example from chemical cross-links,[5] paramagnetic relaxation enhancement (PRE),[6] double electron–electron resonance (DEER),[7] or Förster resonance energy transfer (FRET).[8]

Our method relies on a global nonlinear optimization scheme using experimentally derived distance restraints, starting from a random conformation placed in random position and orientation in space. In contrast to previous methods[2] where all subunits are calculated at the same time, only one (primary) protomer is explicitly modeled, and the others are generated "on the fly" using appropriate symmetry operators as images of the primary protomer (Figure 1). This procedure (later referred to as strict-symmetry) requires the number of subunits in an oligomer to be known, for example, from analytical ultracentrifugation. Symmetry relationships between the primary protomer and the images (for example axes, rotation angles, and translations) are imposed and fixed at the start of the calculation. The primary molecule rotates and translates freely with respect to the axes. The required symmetry group and the symmetry operators can be inferred from the number of subunits for point symmetries (Supporting Information, Table S4). If it is not possible to distinguish between different symmetries with the same number of protomers (for example between $C_4$ or $D_2$ for a tetramer), all possibilities can be tested in calculations and the energies and convergence rates can be compared. Among the examples presented herein, the number of images ranges from 1 (for dimers) to 5 (for hexamers) to 107 (for the $P2_12_12_1$ crystallographic symmetry). For every distance restraint, the degree of ambiguity can be specified (intramolecular, intermolecular, or ambiguous) if it is known from experiments (such as X-filtered NOE experiments[9]). The symmetry ambiguity outlined herein can be applied in conjunction with the standard chemical-shift ambiguity used in automated NOE assignment of monomeric proteins. The calculation protocols are described in detail in the Supporting Information.

Published experimental restraints for eight proteins exhibiting different point symmetries served for test calculations (Figure 2; Supporting Information, Table S1). The precision (RMS deviation around the average structure) of the obtained ensembles of oligomers is similar to the standard

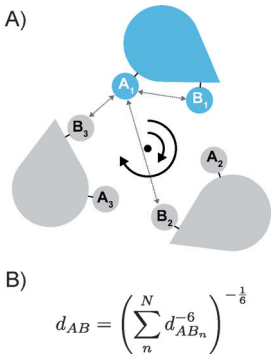


***Figure 1.*** Illustration of the strict-symmetry method for a trimer with $C_3$ point symmetry. A) The two symmetry relationships to produce the images (gray) of the primary protomer (blue) are rotations around the symmetry axis (indicated by a black dot) of 120° and 240° (circular arrows). B) The equation to calculate an effective distance over the images. *n* is the protomer number, with a total of *N* protomers in the oligomer.

[*] Dr. B. Bardiaux, Dr. B.-J. van Rossum, Prof. H. Oschkinat
NMR Supported Structural Biology
Leibniz-Institut für Molekulare Pharmakologie (FMP)
Robert-Rössle-Strasse 10, 13125 Berlin (Germany)
E-mail: oschkinat@fmp-berlin.de
Prof. M. Nilges
Dept. of Structural Biology and Chemistry,
Institut Pasteur; CNRS URA 2185
25–28 rue du docteur Roux, 75015 Paris (France)
E-mail: michael.nilges@pasteur.fr

[**] This work was supported by EU grants FP6 Extend-NMR no. 18988 (to M.N.) and FP7 Bio-NMR no. 261863 (to H.O.).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201201783.


WILEY ONLINE LIBRARY
© 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

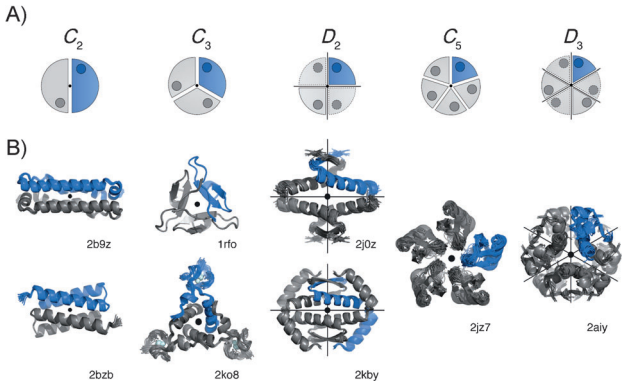


***Figure 2.*** Calculation of oligomers with different cyclic or dihedral symmetries. In each case, the primary protomer is indicated in blue. $C_2$, $C_3$, $C_5$: cyclic point symmetries (one symmetry axis); $D_2$, $D_3$: dihedral symmetries, with three symmetry axes for $D_2$ (tetramer), and four symmetry axes for $D_3$ (hexamer). Labels are Protein Data Bank (PDB) identifiers.

symmetry-ADR method (Supporting Information, Table S2), except for the insulin R6 hexamer (PDB: 2aiy). In this case, only the strict-symmetry protocol achieves convergence starting from random atomic coordinates. Futhermore, for dimers, tetramers, and hexamers, our method yields oligomeric structures closer to the known X-ray structure when compared to the originally reported ensembles (Supporting Information, Table S3), except for the p53 tetramerization domain. While the overall structural quality of the oligomeric ensembles is comparable, structures determined with the method presented herein consistently exhibit a greatly reduced number of steric clashes, resulting from a better usage of the experimental restraints (Supporting Information, Table S7).

In the symmetry-ADR method, internal symmetry is imposed by a large number of artificial distance restraints, and experimental restraints need to be introduced for all protomers. The main advantage of the strict-symmetry method is the reduced computational cost (Supporting Information, Figure S3) and the convenient way of defining the constraints imposing the symmetry. Such advantages are imperative when calculating structures of assemblies with complex symmetries. For instance, when symmetry involves translation as well as rotation, it is cumbersome to treat the system with the symmetry-ADR method owing to its "infinite" nature. An example is the helical symmetry found in fibrils[10] and pili.[4b] The simpler case of fibrils can be treated with an ad hoc adaptation of the symmetry-ADR method,[11] but it becomes intractable for pili where eight or more neighbors are in direct contact with the primary protomer.[4b]

To demonstrate the benefit of our method for such complex systems, we recalculated the fibrillar structure of the prion domain of HET-s, HET-s(218-289), from the published data.[10] The helical symmetry was imposed by a translation of $2 \times 4.7$ Å and a fixed rotational twist between the protomers. These parameters can be readily obtained from diffraction and cryo-electron microscopy.[12] Here, they were derived from the solid-state NMR structure for sake of comparison. Starting from a random configuration for the protomer, calculations converged to the published structure, even with ambiguous restraints (Figure 3; Supporting Information, Table S5).

Another example that cannot be easily modeled is the crystal packing in microcrystals as used in solid-state NMR (ssNMR) structure determination.[13] To explore the ability of

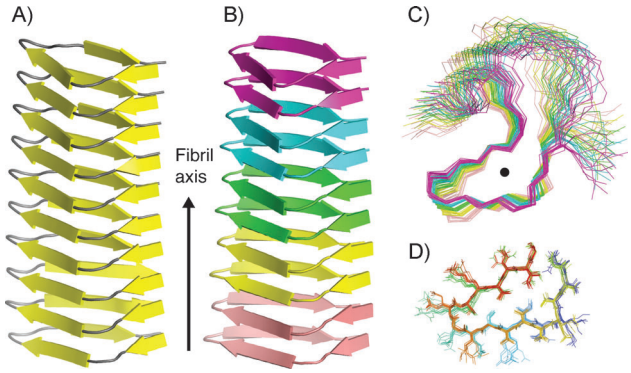


***Figure 3.*** HET-s(218-289) structures. A) Published structure (PDB: 2RNM[10]). B) Recalculated structure with strict symmetry with the fibril axis indicated (core region). C) Top view of the 10 conformers. D) Top view of the central protomer (core region) of the 10 conformers. The two stacked β-strand layers (residue 226 to 242 and residue 262 to 278) are colored according to the residue numbering.

our approach to cope with the ambiguity and noise present in ssNMR spectra recorded on crystalline samples, we applied it to the spectrin SH3 domain data (Figure 4). First, we evaluated the contribution of the intermolecular interactions to the definition of the final structure (Supporting Information). We used a two-phase protocol where an approximate

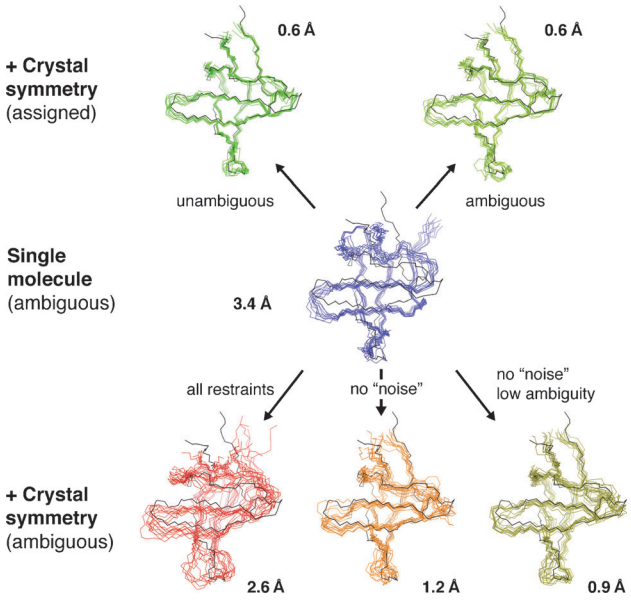


***Figure 4.*** Refinement of SH3 domain structure with ambiguous restraints as a monomeric molecule (center) or with crystalline symmetry (top: unambiguous restraints; bottom: ambiguous restraints). In each case, the accuracy from the X-ray structure (black) is given.

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim 

structure was first calculated from a raw peaklist without the crystal context, and with the ARIA program[14] (incorporating SOLARIA[15] capability of assigning ssNMR proton-driven spin diffusion spectra). Here, intermolecular correlations were treated as noise, as no neighboring molecules were considered. Then the initial data were re-interpreted within the crystal context to refine the structure. The final RMS difference to the X-ray crystal structure was as small as 0.6 Å, which is the best result achieved by us to date (Supporting Information, Table S6).[13a,15,16] This indicates that not only the crystal context enables the correct interpretation of intermolecular correlations that may not always be identified as noise by ARIA, but also that inclusion of these intermolecular interactions leads to a more accurate structure. It is also encouraging that even with chemical shift ambiguity left in the data, our procedure can help to directly refine a structure to an accuracy below 1 Å compared to the X-ray crystal structure.

The procedure presented herein allows for efficient structure calculation of symmetric aggregates on the basis of experimental distance restraints. It is valid for all types of symmetry and long lists of duplicate restraints are not needed. Experimental information to distinguish between intra- and intermolecular NOEs or spatial correlations, such as results from asymmetric labeling and X-filtered experiments can be readily incorporated. Also, other experimental information that can help to define the symmetric arrangement, for example residual dipolar coupling (RDC) measurements, can easily be included in the calculation. In contrast to the original method[17] where both the coordinate system and the molecule can rotate, the axes of the anisotropy tensor are then fixed to the external coordinate system and only the molecule rotates. This is possible because 1) one axis of the tensor always coincides with a symmetry axis, and 2) the orientation of the coordinate system around the first axis becomes arbitrary for threefold or higher symmetry (axially symmetric tensor).[18] This integrated approach is more straightforward and reliable than first calculating a monomer structure and then using RDCs to dock it into the oligomer structure.[19]

Our method requires the number of subunits to be known and the exact specification of all symmetry operations. In general, such information can be determined experimentally. For point symmetries, the number of protomers in the oligomer can be readily obtained from ultracentrifugation or native gel experiments. Parameters for helical symmetries can be measured by electron microscopy. Should such measurements not be available, these parameters can be optimized by a grid search (illustrated in the Supporting Information, Figure S5, for the case of the rotation angle between individual units). Alternatively, these symmetry parameters could be left "floating", similar to the alignment tensor parameters in the Sculptor approach to fit RDC data.[20]

For the study on SH3, the crystal symmetry found in the single crystal used for the X-ray was (correctly) assumed to be identical to the microcrystals used for the ssNMR experiments.[13a] It is by no means certain that this is always the case. However, the number of possible X-ray symmetries can be quickly reduced by mass measurements. Then, different possibilities can be compared. In the present study, we included the microcrystal example to illustrate the generality of the approach, and the usefulness of intermolecular interactions to obtain highly accurate structures. If information on the crystal symmetry is available, this is currently the only method to include them in the calculation. Another advantage of our method is that the computational cost does not rise significantly with the number of subunits. It can therefore readily be applied to very large systems, such as pili[4b] or the nuclear pore.[4a]

## Experimental Section

We extended the NOE module of the program CNS[21] for calculation of distances, energies, and forces with respect to symmetry-related images. Non-bonded interactions between the primary molecule and the images could be treated with the existing setup of non-crystallographic and crystallographic symmetries in CNS. All calculations were performed with CNS. The modified version of CNS and the calculation script will be distributed as part of the ARIA package (http://aria.pasteur.fr).

Received: March 5, 2012
Published online: May 31, 2012

**Keywords:** ambiguous distance restraints · NMR spectroscopy · protein structures · structure elucidation · symmetric oligomers

[1] E. D. Levy, J. B. Pereira-Leal, C. Chothia, S. A. Teichmann, *PLoS Comput. Biol.* **2006**, *2*, e155.
[2] a) B. Bardiaux, A. Bernard, W. Rieping, M. Habeck, T. E. Malliavin, M. Nilges, *Proteins: Struct. Funct. Bioinf.* **2008**, *75*, 569–585; b) M. Nilges, *Proteins Struct. Funct. Genet.* **1993**, *17*, 297–309.
[3] a) H. Chandola, A. K. Yan, S. Potluri, B. R. Donald, C. Bailey-Kellogg, *J. Comput. Biol.* **2011**, *18*, 1757–1775; b) S. Potluri, A. K. Yan, B. R. Donald, C. Bailey-Kellogg, *Protein Sci.* **2007**, *16*, 69–81; c) C. S. Wang, T. Lozano-Perez, B. Tidor, *Proteins Struct. Funct. Genet.* **1998**, *32*, 26–42.
[4] a) F. Alber, S. Dokudovskaya, L. M. Veenhoff, W. Zhang, J. Kipper, D. Devos, A. Suprapto, O. Karni-Schmidt, R. Williams, B. T. Chait, A. Sali, M. P. Rout, *Nature* **2007**, *450*, 695–701; b) M. Campos, O. Francetic, M. Nilges, *J. Struct. Biol.* **2011**, *173*, 436–444.
[5] A. Maiolica, D. Cittaro, D. Borsotti, L. Sennels, C. Ciferri, C. Tarricone, A. Musacchio, J. Rappsilber, *Mol. Cell. Proteomics* **2007**, *6*, 2200–2211.
[6] J. L. Battiste, G. Wagner, *Biochemistry* **2000**, *39*, 5355–5365.
[7] G. Jeschke, Y. Polyhach, *Phys. Chem. Chem. Phys.* **2007**, *9*, 1895–1910.
[8] A. T. Brunger, P. Strop, M. Vrljic, S. Chu, K. R. Weninger, *J. Struct. Biol.* **2011**, *173*, 497–505.
[9] G. Otting, K. Wuthrich, *Q. Rev. Biophys.* **1990**, *23*, 39–96.
[10] C. Wasmer, A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek, B. H. Meier, *Science* **2008**, *319*, 1523–1526.
[11] H. Van Melckebeke, C. Wasmer, A. Lange, E. Ab, A. Loquet, A. Bockmann, B. H. Meier, *J. Am. Chem. Soc.* **2010**, *132*, 13765–13775.
[12] a) N. Mizuno, U. Baxa, A. C. Steven, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3252–3257; b) A. Sen, U. Baxa, M. N. Simon, J. S. Wall, R. Sabate, S. J. Saupe, A. C. Steven, *J. Biol. Chem.* **2007**, *282*, 5545–5550.
[13] a) F. Castellani, B. van Rossum, A. Diehl, M. Schubert, K. Rehbein, H. Oschkinat, *Nature* **2002**, *420*, 98–102; b) A. J.

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim 

Nieuwkoop, C. M. Rienstra, *J. Am. Chem. Soc.* **2010**, *132*, 7570 – 7571.
[14] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T. E. Malliavin, M. Nilges, *Bioinformatics* **2007**, *23*, 381 – 382.
[15] M. Fossi, F. Castellani, M. Nilges, H. Oschkinat, B. J. van Rossum, *Angew. Chem.* **2005**, *117*, 6307 – 6310; *Angew. Chem. Int. Ed.* **2005**, *44*, 6151 – 6154.
[16] F. Castellani, B. J. van Rossum, A. Diehl, K. Rehbein, H. Oschkinat, *Biochemistry* **2003**, *42*, 11476 – 11483.
[17] N. Tjandra, *Structure* **1999**, *7*, R205 – 211.
[18] M. Zweckstetter, *Nat. Protoc.* **2008**, *3*, 679 – 690.
[19] K. Oxenoid, J. J. Chou, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10870 – 10875.
[20] J. C. Hus, D. Marion, M. Blackledge, *J. Mol. Biol.* **2000**, *298*, 927 – 936.
[21] A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, G. L. Warren, *Acta Crystallogr. Sect. D* **1998**, *54*, 905 – 921.

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim 