

# SCIENTIFIC REPORTS

OPEN

## An Exhaustive Search Algorithm to Aid NMR-Based Structure Determination of Rotationally Symmetric Transmembrane Oligomers

Jing Yang<sup>1,2</sup>, Alessandro Piai<sup>2</sup>, Hong-Bin Shen<sup>1</sup> & James J. Chou<sup>2,3</sup>

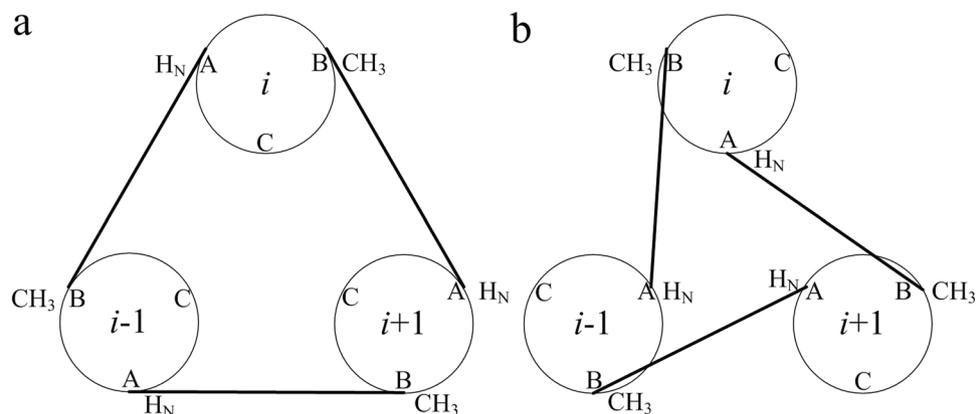
Nuclear magnetic resonance (NMR) has been an important source of structural restraints for solving structures of oligomeric transmembrane domains (TMDs) of cell surface receptors and viral membrane proteins. In NMR studies, oligomers are assembled using inter-protomer distance restraints. But, for oligomers that are higher than dimer, these distance restraints all have two-fold directional ambiguity, and resolving such ambiguity often requires time-consuming trial-and-error calculations using restrained molecular dynamics (MD) with simulated annealing (SA). We report an Exhaustive Search algorithm for Symmetric Oligomer (ExSSO), which can perform near-complete search of the symmetric conformational space in a very short time. In this approach, the predetermined protomer model is subject to full angular and spatial search within the symmetry space. This approach, which can be applied to any rotationally symmetric oligomers, was validated using the structures of the Fas death receptor, the HIV-1 gp41 fusion protein, the influenza proton channel, and the MCU pore. The algorithm is able to generate approximate oligomer solutions quickly as initial inputs for further refinement using the MD/SA method.

Constructing molecular models by satisfying experimentally derived spatial and angular restraints is a general framework for the generation of three-dimensional protein structures by NMR<sup>1</sup>. The most common method for NMR structure calculation is using restrained molecular dynamics (MD) with simulated annealing (SA)<sup>2-6</sup>. In the MD/SA method, structural restraints are implemented as pseudo potentials that drive the dynamics, but such implementation is difficult for ambiguous restraints as they generate potentials with multiple minima. Distance geometry (DG) is another structure calculation method that was very popular in early NMR applications to structural biology<sup>7,8</sup>. This algorithm, however, is sensitive to small uncertainties in the distance matrix. Furthermore, the Bayesian inference has been proposed for NMR structure determination<sup>9</sup>. This method, which derives a probability distribution for the unknown structure, is more computationally challenging. In general, all these methods are not very effective in handling ambiguous restraints.

In NMR-based structure determination of transmembrane (TM) oligomers, the key structural restraints are inter-protomer distance restraints derived from nuclear Overhauser enhancement (NOE). These NOEs are typically between the backbone amide proton of one protomer and aliphatic protons of the neighboring protomers<sup>10-13</sup>. For oligomers with  $n$ -fold rotational symmetry, each NOE restraint between a pair of protomers is duplicated  $n$  times and assigned respectively to all equivalent pairs of protomers to satisfy the condition of symmetry. In symmetric dimers ( $n = 2$ ), the inter-protomer NOE restraint can be assigned unambiguously between atom  $j$  of protomer 1 and atom  $k$  of protomer 2, and between atom  $j$  of protomer 2 and atom  $k$  of protomer 1.

<sup>1</sup>Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, 200240, China. <sup>2</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, 02115, USA.

<sup>3</sup>National Center for Protein Science Shanghai, Shanghai Institute of Biochemistry and Cell Biology, Shanghai Science Research Center, Chinese Academy of Sciences, Shanghai, 200031, China. Correspondence and requests for materials should be addressed to H.-B.S. (email: [hbsen@sjtu.edu.cn](mailto:hbsen@sjtu.edu.cn)) or J.J.C. (email: [james\\_chou@hms.harvard.edu](mailto:james_chou@hms.harvard.edu))



**Figure 1.** Illustration of the two-fold directional ambiguity of an inter-protomer NOE restraint between the backbone  $H_N$  and side-chain  $CH_3$ . **(a)** Restraint between  $H_N$  of residue A in protomer  $i$  and  $CH_3$  of residue B in protomer  $i - 1$  in counterclockwise arrangement. **(b)** Restraint between  $H_N$  of residue A in protomer  $i$  and  $CH_3$  of residue B in protomer  $i + 1$  in clockwise arrangement.

For  $n \geq 3$ , however, each of the NOE-derived restraints has two-fold directional ambiguity. Taking a symmetric trimer as an example, supposing an inter-protomer NOE cross peak between the amide proton of residue A ( $H_N(A)$ ) and the methyl proton of residue B ( $CH_3(B)$ ) has been identified, it can represent a restraint between  $H_N(A)$  of protomer  $i$  and  $CH_3(B)$  of protomer  $i-1$  (Fig. 1a), or between  $H_N(A)$  of protomer  $i-1$  and  $CH_3(B)$  of protomer  $i$  (Fig. 1b), because the NMR resonances of the protomers are identical. The MD/SA method is suitable for restraints that can be implemented as pseudo potentials, but such potentials cannot be implemented precisely in the case of ambiguous restraints, posing serious problems for energy minimization calculations.

Obviously, algorithms that do not depend on restraint-derived energy landscape are more suitable for resolving ambiguous restraints, and several exhaustive search methods have been proposed previously, including the AmbiPack<sup>14</sup> and SYMBRANE<sup>15</sup>. These methods are based on the branch-and-bound algorithm to exhaustively search molecular symmetry axis – which is then translated to the oligomer structure – by recursively dividing a cell representing the symmetry space into smaller sub-cells until finding a cell in which the symmetry axes satisfy all the restraints. The search time of this algorithm, however, increases exponentially with fewer restraints because more cells remain, which will be further divided in each of the following iterations of the search. Moreover, it is difficult to implement other restraints such as orientation restraints or the unconventional restraints such as solvent or membrane accessibility<sup>16</sup>.

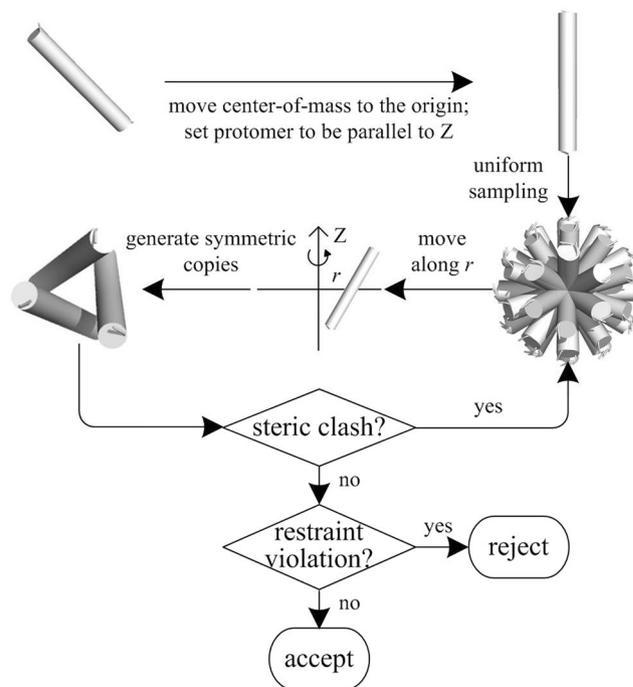
Inspired by these studies, we sought to develop an exhaustive search algorithm for symmetric oligomers with complexity and searching time unaffected by the amount or form of restraints. The purpose of the program is to allow convenient and fast evaluation of whether the experimental restraints are sufficient to achieve a unique mode of oligomerization. The representative conformations from this program can then be further refined in the standard restrained MD and SA programs.

## Results

In the proposed method, named ExSSO and schematically illustrated in Fig. 2, each protomer is treated as a rigid body whose orientation and position relative to the symmetry axis are evaluated. The protomer structure and oligomeric state must be predetermined. In the case of small transmembrane domains (TMDs), e.g., a TM helix, the protomer backbone structure can be initially constructed with the backbone dihedral angles derived from chemical shifts (using, e.g., the TALOS+ program<sup>17</sup>). The algorithm assigns the Z-axis as the axis of symmetry and samples the orientation of the protomer by performing an Euler rotation around its center-of-mass with Euler angles  $\alpha$ ,  $\beta$ , and  $\gamma$ . To ensure near-complete and uniform conformational sampling, we use the following search grid:  $\alpha = 0 - 2\pi$ ,  $\Delta\alpha = 5^\circ$ ;  $\beta = 0 - \pi/2$ ,  $\Delta\beta = 5^\circ$ ;  $\gamma = 0 - 2\pi$ ,  $\Delta\gamma = 5^\circ/\sin(\beta)$ . Subsequently, the oriented protomer is placed at distance  $r$  between the Z-axis and its center-of-mass. By default, the distance  $r$  is set to the range 3–15 Å and the step size  $\Delta r = 0.5$  Å is used, because these values were found optimal for the sizes of most TM oligomers investigated by NMR. Moreover, the user is given the option to adjust these settings if needed. For each configuration of the protomer, the oligomer structure is then constructed by generating symmetric copies of the protomer around the Z-axis using the rotational symmetry operator. Structures with steric clashes, as indicated by inter-protomer distances between C $\beta$  atoms (see Supplementary Fig. 1), are not considered. Finally, each of the oligomer structures is evaluated against the inter-protomer restraints using the following scoring system, which quantifies the agreement between each structural model and inter-protomer restraints as follows:

$$\Delta = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}, \quad (1)$$

where  $N$  is the number of restraints and  $\delta_i$  is the deviation in the model from the  $i$ th restraint.  $\delta_i$  is defined as:



**Figure 2.** Schematic diagram of the ExSSO algorithm for uniformly searching the conformations of a symmetric oligomer that satisfy experimental restraints. A steric clash is reported when the distance between C $\beta$  atoms from different protomers is less than 3.8 Å (Supplementary Fig. 1).

$$\delta_i = \begin{cases} d - D_i - \sigma_i, & d > D_i + \sigma_i \\ d - D_i + \sigma_i, & d < D_i - \sigma_i, \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

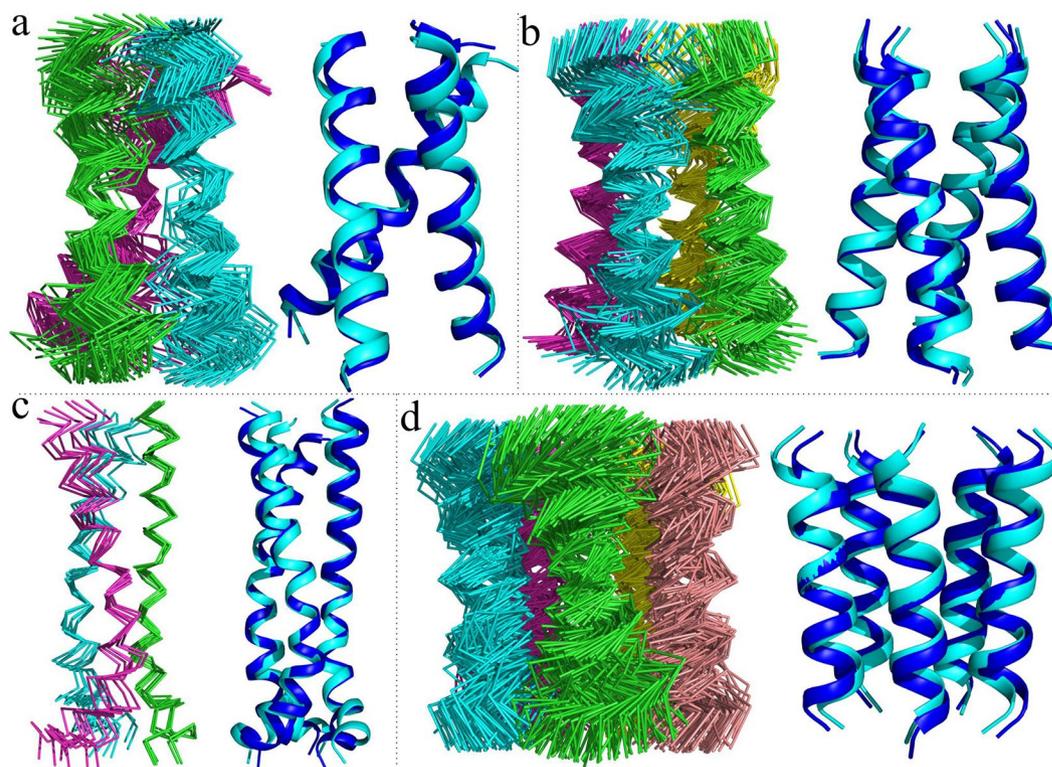
where  $D_i$  and  $\sigma_i$  are the value and uncertainty of the  $i$ th restraint, respectively, and  $d$  is the corresponding distance calculated from the structural model. As described above, an inter-protomer NOE restraint has two-fold directional ambiguity (Fig. 1). Hence, only the one that is better satisfied by the model is used to represent that NOE restraint.

The NOE-derived inter-protomer restraints typically involve protein side-chain methyl and aromatic groups, which are usually mobile due to side-chain flexibility. To enable the rigid-body conformational search of the protomers without sampling side-chain flexibility, the ExSSO converts each of the NOE restraints to pseudo restraint between the protein backbone heavy atoms including C $\alpha$  and C $\beta$ . The protons in the NOE restraints are grouped into two groups: (1) those that are close to the backbone (H $_N$ , H $\alpha$ , and H $\beta$ ), and (2) those that are farther away from the backbone (protons at  $\gamma$ ,  $\delta$ , and  $\epsilon$  positions and aromatic protons). The NOE restraints are then classified into three types: (I) between group 1 protons, (II) between group 1 and 2 protons, and (III) between group 2 protons. The type I, II, and III restraints are represented by pseudo restraints between C $\alpha$  and C $\alpha$ , between C $\alpha$  and C $\beta$ , and between C $\beta$  and C $\beta$ , respectively. To assign the proper distance range for the three types of pseudo restraints above, we performed a statistical analysis using 26 membrane protein structures solved by NMR (Supplementary Table 1). For each type (I, II, and III) of observed long-range NOE restraint (inter-protomer or inter-helical), the distance between the two corresponding backbone atoms (C $\alpha$ , C $\beta$ ) was extracted from the structure. Then, by fitting the distribution for each of the three types of distances to Gaussian function (Supplementary Fig. 2), we derived the mean distance and standard deviation for the three restraint types:  $6.3 \pm 1.7$  Å (type I),  $6.8 \pm 1.5$  Å (type II), and  $7.3 \pm 1.5$  Å (type III). For each restraint type, the mean distance was assigned to  $D_i$  in Eq. 2 and the standard deviation to the associated uncertainty  $\sigma_i$ .

During the search, the ExSSO keeps a *conformation queue* of representative models with  $\Delta \leq \bar{\sigma}$ , where  $\bar{\sigma}$  is the average restraint uncertainty ( $\sim 1.5$  Å). The use of  $\bar{\sigma}$  in collecting the models is based on the argument that the discrepancy between a model and restraints is acceptable if it is within the restraint uncertainty. Initially, the queue is empty. Then, models with  $\Delta \leq \bar{\sigma}$  are added to the queue and ranked according to  $\Delta$  in ascending order. The first model in the queue (or the model with the smallest  $\Delta$ ) is kept by default. Then, starting from the second model in the queue, each model is compared with the other existing models in the queue and is removed from the queue if it is similar to another model with  $\text{RMSD} \leq 0.5$  Å. Lastly, representative models are identified using a clustering algorithm, i.e., only cluster centers in the *conformation queue* are collected as the *conformational ensemble*. In the clustering algorithm<sup>18</sup>, the first cluster center is identified as the model with the largest number of similar models (with  $\text{RMSD} \leq 1$  Å). Once the cluster is found, the models within the cluster are removed. The procedure is then repeated iteratively to find remaining cluster centers until all models in the *conformation queue* are processed.

TMD	# of residues	# of restraints <sup>b</sup>	Avg. pairwise RMSD of the ensemble (Å)	RMSD from known structure (Å) <sup>c</sup>
Fas	22	13	4.5	1.5
gp41	34	18	2.8	2.1
M2	23	11	3.8	0.8
MCU	17	9	4.2	1.3

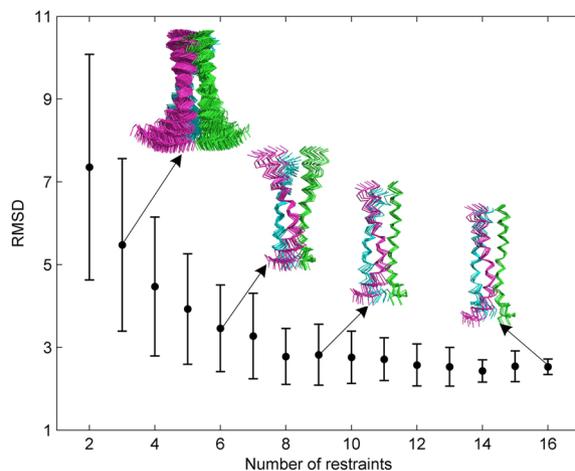
**Table 1.** Summary of ExSSO calculation results<sup>a</sup>. <sup>a</sup>More details in Supplementary Table 2. <sup>b</sup>NOE-derived inter-protomer restraints. <sup>c</sup>Backbone C $\alpha$  difference between the best model (with the smallest  $\Delta$  from ExSSO) and the deposited NMR structure.



**Figure 3.** ExSSO-derived *conformational ensemble* and comparison with the known structures. In each sub-figure, the left panel shows the final *conformational ensemble*, and the right panel shows the overlay of the best conformation (cyan) to the known structure (blue). **(a)** The trimeric TMD of the human Fas (PDB ID: 2na7). **(b)** The tetrameric TMD of influenza M2 (PDB ID: 2r1f). **(c)** The trimeric TMD of the HIV-1 gp41 (PDB ID: 5jyn). **(d)** The pentameric complex formed by the second TM helix of *C.elegans* MCU (PDB ID: 5id3).

The above algorithm was tested for several oligomeric TMD structures for which inter-protomer NMR restraints are available: the trimeric TMD of the Fas receptor<sup>19</sup>, the trimeric TMD of the HIV-1 gp41 fusion protein<sup>12</sup>, the tetrameric TMD of influenza M2 channel<sup>20</sup>, and the pentameric TMD of MCU channel pore<sup>21</sup>. The ExSSO parameters used for these applications were:  $\Delta\alpha = 5^\circ$ ,  $\Delta\beta = 5^\circ$ ,  $\Delta\gamma = 5^\circ/\sin(\beta)$ , and  $\Delta r = 0.5 \text{ \AA}$ . The calculations were performed in Mac OS X with a 2.5 GHz Intel Core i5 processor and the results are listed in Table 1 and shown in Fig. 3. The algorithm demonstrated high efficiency as the search time was typically within 20 seconds for each of the four cases (Supplementary Table 2).

The *conformational ensembles* (i.e. all the cluster centers found in the *conformation queue*) generated by the ExSSO for each of the four TMDs are displayed in Fig. 3. As can be seen, the spread of the ensemble varied substantially among the four TMDs despite the fact that the average number of inter-protomer restraints per residue is similar for all four cases. Among them, the Fas TMD showed the worst convergence (RMSD = 4.5 Å) (Table 1); it has 62 cluster centers (Fig. 3a), identified from 320 representative models in the *conformation queue*. Notably, although the Fas TMD has more restraints than the M2 TMD, the latter showed better convergence (RMSD = 3.8 Å) (Table 1; Fig. 3b). A careful examination of the restraint list revealed that the inter-protomer restraints of the M2 TMD are better distributed along the TM helix than the Fas TMD. The Fas TMD trimerizes around the central proline, where most of the inter-protomer restraints were found. There are, however, no restraints near the two ends of the TMD, which is consistent with the greater structural divergence observed



**Figure 4.** Plot of (*average ensemble RMSD*) vs. (*number of restraints*), showing the structural convergence of the HIV-1 gp41 TMD as a function of the number of inter-protomer restraints used in the ExSSO calculation. The ensemble RMSD is averaged over 100 repetitions and the error bar is the standard deviation ( $\pm\sigma$ ).

moving away from the core region of the TMD (Fig. 3a). The HIV-1 gp41 TMD had the best convergence with backbone RMSD of 2.8 Å (Table 1), calculated from 6 cluster centers (Fig. 3c). Finally, the MCU pore TMD also showed mediocre convergence with RMSD of 4.2 Å, calculated from 86 cluster centers (Fig. 3d). Despite the variation in the ensemble spread, the best models (those with the smallest  $\Delta$ ) from ExSSO in the four cases agree remarkably well with the respective known structures, with backbone RMSD from the known models in the range 0.8–2.1 Å (Table 1; Fig. 3).

We next investigated the influence of reducing the number of restraints on the structural convergence using the TMD of HIV-1 gp41 as a model system. The ExSSO algorithm was tested for different number of inter-protomer restraints, from 16 to 2. For each case, restraints were randomly taken out for an ExSSO calculation and the process was repeated 100 times. The plot of the *average ensemble RMSD* versus the *number of restraints* showed rapid improvement in convergence from 1–5 restraints, but reached steady state at ~9 restraints (Fig. 4). The ensembles generated with 16, 9, 6, and 3 restraints show that the RMSD increased from 2.5 to 5.5 Å with less restraints (Fig. 4). Consistent results were also obtained from performing an identical analysis for the pore-forming TM helix of MCU (Supplementary Fig. 3). In both cases, the plots conform to the fundamental principle of restraint-driven structure determination, and thus further validate the ExSSO algorithm.

## Discussion

In this study, we have developed a fast and efficient algorithm, ExSSO, for uniformly and exhaustively searching for structures of symmetric TM oligomers that satisfy ambiguous and non-ambiguous restraints. We have shown, for several TMDs with known structure and available NMR restraints, that ExSSO can generate not only the highest-score models that agree well with the known structures but also all possible representative models that satisfy the experimental data to within uncertainties.

Historically, exhaustive search has not been commonly used in NMR-based structure determination because it is computationally unrealistic to search the entire conformational space starting from linear polypeptide chains. The search space, however, can be greatly reduced using predetermined secondary structures of the protein. Moreover, in the current application to homo-oligomeric TMD structures, the symmetry constraint further reduces the conformational space, allowing for structure calculation in as little as a few seconds (Supplementary Table 2) on a single 2.5 GHz Intel Core i5 CPU. This type of fast structure turnaround makes the program a useful tool for iterative assignment and resolution of the two-fold ambiguous inter-protomer NOEs during the process of structure determination.

The size of search grid is obviously a central parameter of the algorithm as it has an important consequence on the compromise between the speed and completeness of the search. We found that 5° is an optimal grid size, which produced good results for many cases in a short time. A smaller grid size would take much more time and result in more redundant structures; a larger grid size could miss good models.

The uncertainty in the restraints is another important parameter. In this study, we used a very generous uncertainty ( $\pm 1.5$  Å) to account for side-chain flexibility as well as potentially wrongly assigned restraints (though their number must be much lower than that of the correct restraints). In the case in which a sufficient number of restraints still cannot generate a convergent ensemble, a valid option is to slightly reduce the uncertainties of all the restraints or only of those that are absolutely correct.

To address the tolerance of the ExSSO to inaccurate starting protomer model, we tested the ExSSO calculation using a protomer model of the HIV-1 gp41 TMD that was generated using only the TALOS-derived dihedral restraints. As shown in Supplementary Fig. 4a, this protomer structure is significantly different from the one determined using both TALOS-derived dihedral restraints and local NOE restraints. Despite the substantial deviation in the protomer structure, ExSSO correctly determined the mode of trimer assembly with the inter-protomer NOEs restraints (Supplementary Fig. 4b). This preliminary trimer structure could then be refined

to the accurate structure in XPLOR using all NMR restraints (Supplementary Fig. 4c). Therefore, ExSSO proved itself to be an effective tool for generating proximal but correct oligomer models for final refinement using the conventional MD/SA methods.

Finally, the exhaustive conformational search in structure calculation has the obvious advantage of evaluating all ambiguous restraints in a completely unbiased manner, as the search result does not depend on the starting models. The search is also not affected by the complexity introduced by ambiguity in restraints because it systematically evaluates all restraints for all possible conformations. While the current study demonstrates such advantage for ambiguous distance restraints, the exhaustive search approach is in principle generally applicable to all types of structural constraints, including those that are difficult to implement in the form of pseudo potential required by the MD/SA calculation. For example, confinement of TM helices in a lipid bilayer or exclusion of extramembrane domains from the lipid bilayer can be implemented with simple conditional statements to be included in the scoring function (see Supplementary Information). The reported ExSSO program thus represents a versatile framework with which experimental data other than inter-protomer distance restraints can be explored for determining oligomeric TMD structures.

## Method

The calculation modules of the ExSSO program were written in the C++ language. Python scripts were used to operate these modules. All ExSSO calculations were performed in Mac OS X with a 2.5 GHz Intel Core i5 processor. The experimental inter-protomer NOE restraints for the TMDs of the Fas death receptor, the HIV-1 gp41 fusion protein, the influenza M2 channel, and the MCU pore were taken from PDB depositions with PDB IDs 2na7, 5jyn, 2rlf, and 5id3, respectively.

The restrained MD/SA calculation for refining the initial model of the HIV-1 gp41 TMD derived from ExSSO (with inaccurate protomer model) in Supplementary Fig. 4 was performed using the program XPLOR-NIH (version 2.41.1)<sup>5</sup>. In this calculation, the best trimer model from ExSSO (model with the smallest  $\Delta$  in Eq. 1) was used as the starting model. The model was refined against the complete set of NMR restraints deposited with PDB ID 5jyn using a SA protocol in which the temperature of the bath was cooled from 1000 to 200 K with steps of 40 K. The NOE restraints were enforced by flat-well harmonic potentials, with the force constant ramped from 2 to 50 kcal/mol  $\text{\AA}^{-2}$  during annealing. Backbone dihedral angle restraints, all with a flat-well ( $\pm$  the corresponding uncertainties) harmonic potential with force constant ramped from 10 to 30 kcal/mol  $\text{rad}^{-2}$ . A total of 50 structures were calculated and 10 lowest energy structures were selected as the final structural ensemble (shown in Supplementary Fig. 4c).

## References

1. Wuthrich, K. NMR studies of structure and function of biological macromolecules (Nobel Lecture). *J Biomol NMR* **27**, 13–39 (2003).
2. Brunger, A. T., Adams, P. D. & Rice, L. M. New applications of simulated annealing in X-ray crystallography and solution NMR. *Structure* **5**, 325–336 (1997).
3. Nilges, M., Clore, G. M. & Gronenborn, A. M. Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. *Circumventing problems associated with folding. FEBS Lett* **239**, 129–136 (1988).
4. Nilges, M., Gronenborn, A. M., Brunger, A. T. & Clore, G. M. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng* **2**, 27–38 (1988).
5. Schwieters, C. D., Kuszewski, J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 66–74 (2002).
6. Nilges, M. Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J Mol Biol* **245**, 645–660 (1995).
7. Havel, T. F., Crippen, G. M., Kuntz, I. D. & Blaney, J. M. The combinatorial distance geometry method for the calculation of molecular conformation. II. *Sample problems and computational statistics. J Theor Biol* **104**, 383–400 (1983).
8. Havel, T. F., Kuntz, I. D. & Crippen, G. M. The combinatorial distance geometry method for the calculation of molecular conformation. I. A new approach to an old problem. *J Theor Biol* **104**, 359–381 (1983).
9. Rieping, W., Habeck, M. & Nilges, M. Inferential structure determination. *Science* **309**, 303–306 (2005).
10. MacKenzie, K. R., Prestegard, J. H. & Engelman, D. M. A transmembrane helix dimer: structure and implications. *Science* **276**, 131–133 (1997).
11. Oxenoid, K. & Chou, J. J. The structure of phospholamban pentamer reveals a channel-like architecture in membranes. *Proc Natl Acad Sci USA* **102**, 10870–10875 (2005).
12. Dev, J. *et al.* Structural basis for membrane anchoring of HIV-1 envelope spike. *Science* **353**, 172–175 (2016).
13. Ma, D. *et al.* NMR studies of a channel protein without membranes: structure and dynamics of water-solubilized KcsA. *Proc Natl Acad Sci USA* **105**, 16537–16542 (2008).
14. Wang, C. S., Lozano-Perez, T. & Tidor, B. AmbiPack: a systematic algorithm for packing of macromolecular structures with ambiguous distance constraints. *Proteins* **32**, 26–42 (1998).
15. Potluri, S., Yan, A. K., Chou, J. J., Donald, B. R. & Bailey-Kellogg, C. Structure determination of symmetric homo-oligomers by a complete search of symmetry configuration space, using NMR restraints and van der Waals packing. *Proteins* **65**, 203–219 (2006).
16. Wang, Y., Schwieters, C. D. & Tjandra, N. Parameterization of solvent-protein interaction and its use on NMR protein structure determination. *J Magn Reson* **221**, 76–84 (2012).
17. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213–223 (2009).
18. Barth, P., Wallner, B. & Baker, D. Prediction of membrane protein structures with complex topologies using limited constraints. *Proc Natl Acad Sci USA* **106**, 1409–1414 (2009).
19. Fu, Q. *et al.* Structural Basis and Functional Role of Intramembrane Trimerization of the Fas/CD95 Death Receptor. *Molecular cell* **61**, 602–613 (2016).
20. Schnell, J. R. & Chou, J. J. Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* **451**, 591–595 (2008).
21. Oxenoid, K. *et al.* Architecture of the mitochondrial calcium uniporter. *Nature* **533**, 269–273 (2016).

## Acknowledgements

We thank Liqiang Pan and Qingshan Fu for insightful discussions. This work was supported by the CAS grant XDB08030301 and the US National Institutes of Health grant GM116898 to J.J.C., the National Natural Science Foundation of China (No. 61671288, 91530321, 61725302) and Science and Technology Commission of Shanghai Municipality (No. 16JC1404300, 17JC1403500) to H.S.

## Author Contributions

J.Y., H.S., and J.J.C. conceived of the study; J.Y., H.S., and J.J.C. designed the algorithm; J.Y. wrote the ExSSO program; A.P. tested the ExSSO program; J.Y., A.P., and J.J.C. wrote the paper and all authors contributed to editing of the paper.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-17639-w>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017