DR-SIP: Protocols for Higher Order Structure Modeling with Distance Restraints- and Cyclic Symmetry-Imposed Packing

Justin Chan[1,2], Jinhao Zou[1], Chi-Hong Chang Chien[1], Rong-Long Pan[1], Lee-Wei Yang[1,2,3,*]

[1]Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, Taiwan
[2]Bioinformatics Program, Taiwan International Graduate Program, Institute of Information Sciences, Academia Sinica, Taipei, Taiwan
[3]Physics Division, National Center for Theoretical Sciences, National Tsing Hua University, Hsinchu, Taiwan

*Correspondence: lwyang@life.nthu.edu.tw

Abstract

Motivation

Quaternary structure determination for proteins is difficult especially so for transmembrane proteins. Even if the monomeric constituents of complexes have been experimentally resolved, computational prediction of quaternary structures is still a challenging task particularly for complexes higher than dimers. It is essential to have a reliable computational protocol to predict quaternary structures of both transmembrane and soluble proteins with options to take advantage of experimentally determined distance restraints and general cyclic symmetry (C_n symmetry) found in majority of homo-oligomeric transmembrane proteins.

Results

We survey the structures of all α-helical homo-oligomeric transmembrane proteins (HoTPs) that are structurally solved by X-ray crystallography and find that 90% of them are C_n symmetric. Given the prevalence of C_n symmetric HoTPs and the benefits of incorporating geometry restraints in aiding structure determination of oligomers, we introduce two new filters, the distance-restraints (DR) filter and the Symmetry-Imposed Packing (SIP) filter. These filters are incorporated into two new docking protocols, one for HoTPs and another for soluble proteins. Our software package, DR-SIP, integrates experimentally characterized distance restraints to confine the spatial arrangement of HoTPs and select docking poses (homo-dimers) where the monomers are related by C_n symmetry without prior knowledge of the number ("*n*") of monomers in the complex. Our results show that DR-SIP can identify near-native poses among 54,000 possible decoys for each of the four systems, as the first ranked pose in three cases (two HoTPs and one soluble protein) and the second ranked pose in one HoTP.

Availability and Implementation

https://github.com/capslockwizard/drsip

Supplementary information

Supplementary methods and results are available.

Introduction

The quaternary structure of proteins provides atomistic details which can be used to study the mechanisms underlying the function of these proteins. A large subset of these protein complexes are transmembrane proteins which constitute 20-30% of the proteome (Fagerberg *et al.*, 2010) and are the second most common drug targets (Rask-Andersen *et al.*, 2014), after enzymes. A quick search on RCSB PDB (Rose *et al.*, 2017) shows that about 66% of α-helical transmembrane proteins are homo-oligomeric transmembrane proteins (HoTPs), while 92% of HoTPs are cyclic (C_n) symmetric (**Fig. 1**).

Experimentally solving the ternary and quaternary structures of transmembrane proteins is challenging. The general difficulty is in purifying proteins in sufficient quantity and purity. For X-ray crystallography, the major challenge after obtaining the protein is in screening for the optimal conditions to reconstitute and crystalize the protein (Moraes *et al.*, 2014). On the other hand, solution NMR is limited to protein complexes with a high tumbling rate (Liang and Tamm, 2016) which limits the size of the complex while solid-state NMR is limited by spectral crowding, which is further aggravated by large complexes (Liang and Tamm, 2016).

Computational methods such as homology docking has been applied to predict quaternary structures which are evolutionarily conserved (Levy *et al.*, 2008). These methods (Szilagyi and Zhang, 2014; Bertoni *et al.*, 2017) identify structurally solved homologous protein complexes, not just constituent proteins, as the templates for the homology modelling. However, to predict higher order structures for proteins that lack any structural template of homologs, physiochemically driven molecular docking could be the only option to go with.

Molecular docking is a method that usually generates tens of thousands of orientations (poses) of two proteins bound to one another which are then rank-ordered by a scoring function that evaluates the fitness (or "energy") of these poses (Huang, 2014). Identifying near-native poses among the >10,000 generated poses is difficult because of the efficient sampling in molecular docking comes at the expense of accurately evaluating the solvent and entropy's energy (Brooijmans and Kuntz, 2003). Common docking packages include ZDOCK (Pierce *et al.*, 2011), ClusPro (Kozakov *et al.*, 2017), GRAMM-X (Tovchigrechko and Vakser, 2006), RosettaDock (Chaudhury and Gray, 2008) and HADDOCK (van Zundert *et al.*, 2016), where the latter two allow for experimentally determined restraints or prior knowledge of pairwise distances to constraint and filter the docking results.

Predicting interactions between membrane proteins, DOCK/PIERR (Viswanath *et al.*, 2015), ROSETTA:MPdock (Alford *et al.*, 2015) and Memdock (Hurwitz *et al.*, 2016) generate poses whose transmembrane helices in the constituent monomers are embedded in the membrane bilayer, which enriches the near-native poses among the decoys. There are also docking packages that specifically perform C_n symmetric docking such as SymmDock (Schneidman-Duhovny *et al.*, 2005), ClusPro (Kozakov *et al.*, 2017), M-ZDOCK (Pierce *et al.*, 2005), ROSETTA:MPsymdock (Alford *et al.*, 2015) and HADDOCK (van Zundert *et al.*, 2016). Although these methods are successful at predicting the native complexes, they require prior knowledge of the number of constituent monomers in the complex, or the "order of symmetry".

Integrative approaches, combining multiple sources of experimental data with computational methods, were used to confine the search space during generation of poses or to filter poses that

are incompatible with the data thereby enriching the near-native poses among the docking decoys (Alber *et al.*, 2008). Biochemically and biophysically determined distance restraints include but are not limited to the data obtained from mutagenesis, cross-linking, hydrogen/deuterium exchange, electron microscopy (Mitra and Frank, 2006), small-angle X-ray scattering (Yamagata and Tainer, 2007), NMR chemical shift perturbations (Gupta *et al.*, 2013; Chang *et al.*, 2016; Khan *et al.*, 2018) and single molecule resonance energy transfer (smFRET) (Dimura *et al.*, 2016; Choi *et al.*, 2010; Muschielok *et al.*, 2008), which have been used to predict quaternary structures (Alber *et al.*, 2008).

Among these biophysical methods, smFRET measures the distance between two fluorophore dyes acting as labels attached to residues in proteins or nucleotides to study protein folding, protein-protein interaction and protein-DNA/RNA interaction (Sasmal *et al.*, 2016). However, these distances suffer from uncertainties due to the effect of the local environment on the dye pairs (Dimura *et al.*, 2016; Kalinin *et al.*, 2012; Muschielok *et al.*, 2008).

In this study, we introduce the Distance Restraints and Cyclic Symmetry-Imposed Packing (DR-SIP) software package containing two new docking protocols and two new filters to facilitate quaternary structure determination for soluble and transmembrane proteins. The Symmetry-Imposed Packing ("SIP") filter measures the deviation of a docking pose (homo-dimer) from the ideal C_n symmetry without prior knowledge of the order of symmetry (the *n* in C_n). The pose is removed if it is larger than the statistically-suggested cutoff (see **Results**). Given the uncertainties of the absolute distances obtained from smFRET, our Distance Restraints (DR) filter instead uses the agreement in the relative ranking of these distances by correlating the smFRET-reported intramolecular distances and their counterparts in the docking poses. The cutoffs of these filters are carefully derived from the statistical analyses of all experimentally solved α-helical HoTPs (see **Results**). DR-SIP is validated with four systems, Syt1-SNARE complex (soluble complex) and three HoTPs - MscL, *Vr*H^+-PPase and *Ct*H^+-PPase. A near-native pose can be found within the top two out of 54,000 poses generated by ZDOCK for all four systems, while it is ranked first in three out of four systems. In addition to that, applying DR-SIP to the docking results of GRAMM-X shows that it can enrich the near-native poses such that they make up one out of every two remaining poses.

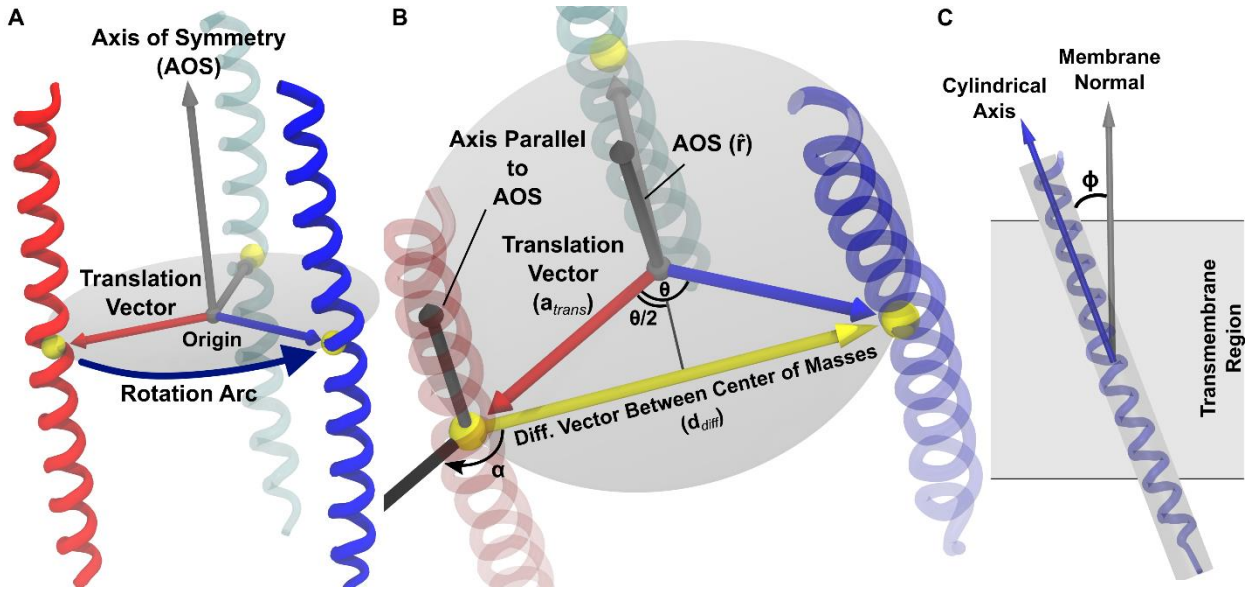

Fig. 1. C_n symmetry and tilt angle of HoTPs. **(A)** Three helices with C_3 symmetry about the axis of rotation/symmetry (AOS). **(B)** The translation vector ($\mathbf{a}_{trans}$) can be expressed as a function of angle of rotation (θ) and difference vector between the centers of mass of the two monomers ($\mathbf{d}_{diff}$) such that $\mathbf{a}_{trans}$ is obtained by rotating $\mathbf{d}_{diff}$ by $\boldsymbol{\alpha} = 90^{o} + \theta/2$ degree about the AOS ($\hat{\mathbf{r}}$) and then rescaled to $\|\mathbf{d}_{diff}\|/(2\sin(\theta/2))$ for its magnitude. **(C)** Tilt angle (ϕ) is the acute angle between the cylindrical axis and membrane normal (assumed to be parallel to AOS).

Methods

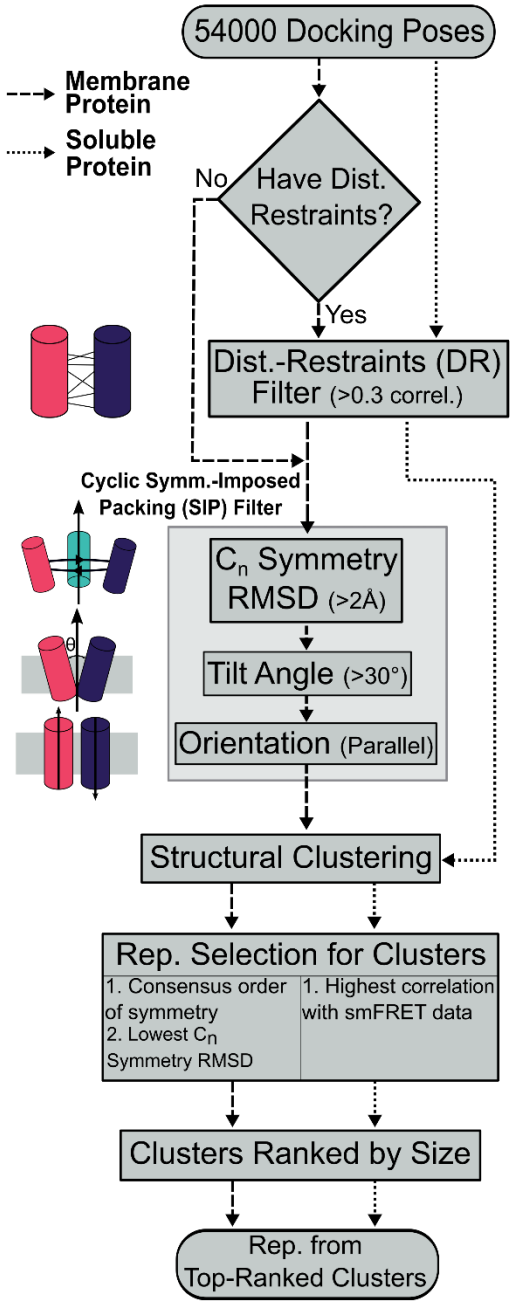

Fig. 2. Flowchart of DR-SIP's membrane and soluble protein docking protocols. Each cluster is represented by one member of the cluster. The final results are the representatives ranked based on the size of the clusters.

Docking Protocols

DR-SIP contains two docking protocols (**Fig. 2**). The first protocol is for predicting the quaternary structures of HoTPs (dashed line in **Fig. 2**) and the other for soluble proteins (dotted line in **Fig. 2**). The docking protocols make use of the DR and SIP filters. The SIP filter (see below and **Supplementary Methods**) consists of the C_n symmetry root-mean squared deviation (RMSD), tilt angle, and orientation criteria.

For all four systems, 54,000 docking poses were generated using ZDOCK 3.0.2 (Pierce *et al.*, 2011) with 6° rotational sampling (dense sampling).

C_n symmetry root-mean-squared deviation (C_n RMSD) measures the current pose's deviation from its closest ideal C_n symmetric pose (see **Methods**). The axis and order of symmetry (*n*) are estimated from each docking pose containing two monomers and used to generate the ideal C_n symmetric complex. Poses with >2Å C_n RMSD were removed (see **Figs. 3** and **4**).

While, tilt angle is the acute angle between the membrane's normal and the direction the monomer spans the membrane (**Fig. 1C**). When the tilt angle is large, the protein might not be able to span the full thickness of the membrane. Assuming the membrane normal is the axis of symmetry (AOS) of HoTPs, poses having tilt angles >30° are removed (see **Fig. 3**).

Orientation of HoTPs is defined as the relative direction that the monomers span the membrane. Poses where both monomers' N-termini face the same side of the membrane are defined as "parallel". A pose is parallel when the dot product of the vector pointing from the center of mass to the first residue of each monomer is positive. Anti-parallel poses are filtered out.

The distance-restraints (DR) filter removes non-native poses that are incompatible with the experimental data (smFRET measured distance data) using relative distances rather than absolute distances. This is measured by the Pearson and Spearman correlations

between the smFRET-measured distances and the distances of the corresponding labeled residues (between Cα atoms) in the docking pose. Poses are removed if they have ≤0.3 correlation.

After applying the filters, the remaining poses are clustered using agglomerative clustering with the average linkage criteria (Manning *et al.*, 2008), see **Supplementary Methods** for details. RMSD is used to measure the difference/distance between all pairs of poses. A cutoff of 12Å is employed and clusters with fewer than three poses are removed. The clusters for proteins are ranked based on their sizes in descending order.

For HoTPs, each cluster is assigned a consensus order of symmetry (*n*), such that most members in the cluster have this '*n*'. The representative pose of each cluster must have the consensus order of symmetry and is the closest to the ideal C_n symmetry as measured by the smallest C_n symmetry RMSD.

For soluble proteins, the representative pose of each cluster has the highest Spearman correlation with the distance restraints data (smFRET data). If there are ≥2 poses with the same Spearman correlation, the one with the higher Pearson correlation is chosen as the representative.

The final results returned to the user are the representative poses from the top clusters. Details of the filters and clustering method are shown in the **Supplementary Methods**.

Extracting C_n Symmetry Parameters from an Ideal C_n Symmetric Pose

As explained before, also true for this and the next sections, each ideal C_n symmetric HoTP pose contains two identical and neighboring monomers (the blue and red monomers of the C_3 complex in **Fig. 1A**) that is a part of ($C_{n>2}$) or is the full (C_2) protein complex. The number of monomers in a complex can be determined by dividing 2π with the angle of rotation between these two monomers (see below).

The two monomers, A and A', are related by

$$\mathbf{A}' = \mathbf{R}(\mathbf{A}) = \mathbf{R}(\mathbf{A}_0 + \mathbf{A}_{trans}) = \mathbf{R}\mathbf{A}_0 + \mathbf{R}\mathbf{A}_{trans} = \mathbf{A}'_0 + \mathbf{R}\mathbf{A}_{trans} \quad (1)$$

where **A** and **A'** are 3x*M* coordinate matrices of *M* atoms for A and A', respectively. **R** is a rotation matrix that rotates about the AOS by 2π/*n* rad (*n* is the order of symmetry). In **Equation (1)**, the COM of the n-mer complex is assumed to be at the origin (e.g. the center of the trimer in **Fig 1A**). When the COM of A and A' are placed at the origin, their coordinate matrices are $\mathbf{A}_0$ and $\mathbf{A}'_0$, respectively. Let $\mathbf{A}_{trans} = \mathbf{a}_{trans}\,[1, 1, \ldots 1]_{1xM}$ where $\mathbf{a}_{trans}$ is the 3x1 vector pointing from the origin to the COM of A.

The docking pose provides **A** and **A**' which are used to compute the AOS ($\hat{\mathbf{r}}$), angle of rotation (θ_r), and the translation vector ($\mathbf{a}_{trans}$). These three parameters define the C_n symmetric relationship between A and A' which can be used to reconstruct the n-mer complex.

$\hat{\mathbf{r}}$ and θ_r can be obtained from **R**. As for **R**, from **Equation (1)**, $\mathbf{A}'_0 = \mathbf{R}\mathbf{A}_0$, the Kabsch algorithm is used to find an **R** that minimizes the RMSD between $\mathbf{A}'_0$ and $\mathbf{R}\mathbf{A}_0$ (Kabsch, 1976; Cock *et al.*, 2009).

Given **R**, $\hat{\mathbf{r}}$ is the eigenvector associated with the unity eigenvalue such that $\mathbf{R}\hat{\mathbf{r}} = 1\hat{\mathbf{r}}$, because the AOS is invariant when operated on by **R**.

As for θ_r, it can be found by extracting cos θ_r and sin θ_r from **R** which can be formulated as (Jia, 2017):

$$\mathbf{R} = \begin{bmatrix} \cos\theta_r + r_x^2(1-\cos\theta_r) & r_x r_y(1-\cos\theta_r) - r_z\sin\theta_r & r_x r_z(1-\cos\theta_r) + r_y\sin\theta_r \\ r_x r_y(1-\cos\theta_r) + r_z\sin\theta_r & \cos\theta_r + r_y^2(1-\cos\theta_r) & r_y r_z(1-\cos\theta_r) - r_x\sin\theta_r \\ r_x r_z(1-\cos\theta_r) - r_y\sin\theta_r & r_y r_z(1-\cos\theta_r) + r_x\sin\theta_r & \cos\theta_r + r_z^2(1-\cos\theta_r) \end{bmatrix} \quad (2)$$

where θ_r is the rotation angle and $\hat{\mathbf{r}} = (r_x, r_y, r_z)$ is the AOS (unit vector). The trace of **R** is 1 + 2cos θ_r. Thus cos θ_r can be obtained.

While, sin θ_r is obtained as follows

$$\mathbf{R} - \mathbf{R}^{\boldsymbol{T}} = \begin{bmatrix} 0 & -2r_z\sin\theta_r & 2r_y\sin\theta_r \\ 2r_z\sin\theta_r & 0 & -2r_x\sin\theta_r \\ -2r_y\sin\theta_r & 2r_x\sin\theta_r & 0 \end{bmatrix} \quad (3)$$

where $2r_x\sin\theta_r + 2r_y\sin\theta_r + 2r_z\sin\theta_r = 2\sin\theta_r\,(r_x + r_y + r_z)$. With $\hat{\mathbf{r}}$, we solve for sin θ_r.

Finally, θ_r = sign(sin θ_r) arccos θ_r. Alternatively, θ_r can be computed using the atan2 function which is implemented in most mathematical software libraries (van der Walt *et al.*, 2011).

$\mathbf{a}_{trans}$ is obtained by rotating $\mathbf{d}_{diff}$ about $\hat{\mathbf{r}}$ by α radians (**Fig. 1B**), which is then rescaled to $\|\mathbf{a}_{trans}\|$ based on its trigonometric relationship with $\mathbf{d}_{diff}$

$$\|\mathbf{a}_{trans}\| = \left|\frac{\|\mathbf{d}_{diff}\|}{2\sin\left(\frac{\theta_r}{2}\right)}\right| \quad (4)$$

$$\mathbf{a}_{trans} = \mathbf{R}_t\frac{\mathbf{d}_{diff}}{\|\mathbf{d}_{diff}\|}\|\mathbf{a}_{trans}\| \quad (5)$$

where $\mathbf{d}_{diff}$ is the vector from the COM of **A** to **A'** and $\mathbf{R}_t$ is a rotation matrix, rotating α = [-sign(θ_r) (|θ_r|/2 + π/2)] rad about $\hat{\mathbf{r}}$ (**Fig. 1B**).

Constructing the Ideal C_n Symmetric Pose That Is Closest to a Non-Ideal Docking Pose or X-Ray Structure

No poses obtained from docking or experimentally resolved structures are in perfect C_n symmetry (non-ideal). Our goal here is to obtain an ideal C_n symmetric pose that is closest to the non-ideal pose. The deviation of the non-ideal pose from the ideal pose is used to determine if the non-ideal pose is C_n symmetric (see C_n Symmetry RMSD in **Supplementary Methods**).

To construct the ideal C_n symmetric pose (A'*) from a non-ideal pose, the angle of rotation and translation vector are computed as follows

$$\mathbf{d}_{diff}^{*} = \mathbf{d}_{diff} - \left(\mathbf{d}_{diff}^{T}\hat{\mathbf{r}}\right)\hat{\mathbf{r}} \quad (6)$$

where n = [$|2\pi/\theta_r|$ rounded to the nearest integer] (such that $\theta_n = 2\pi / n$) and $\mathbf{d}^*_{diff}$ is the vector component of $\mathbf{d}_{diff}$ that is orthogonal to $\hat{\mathbf{r}}$. Note that, $\left|\mathbf{d}^T_{diff}\hat{\mathbf{r}}\right|$ is the shift along the AOS (see **Supplementary Methods**).

Then, $\mathbf{a}_{trans}$ is computed with $\mathbf{d}^*_{diff}$ and θ_n (see **Equation (4)** and **(5)**). To obtain the A'* pose, $\mathbf{A}_0$, $\mathbf{a}_{trans}$ and $\mathbf{R}^*$ are substituted into **Equation (1)**, where the rotation matrix $\mathbf{R}^*$ rotates about $\hat{\mathbf{r}}$ by θ_n.

Data Sets

Non-redundant X-ray structures of α-helical HoTPs

The PDB IDs of α-helical transmembrane proteins were taken from the mpstruct database (http://blanco.biomol.uci.edu/mpstruc/) and filtered with RCSB PDB's advanced filter (Rose *et al.*, 2017) to remove redundant proteins and select for high-resolution X-ray structures (see **Supplementary Methods**). The structures were manually verified to be transmembrane and homo-oligomeric proteins resulting in the final data set containing 129 PDB IDs.

smFRET and structural data used for validation

The Syt1-SNARE smFRET data are obtained from Choi *et al.* (Choi *et al.*, 2010) and filtered such that each residue pair's transfer efficiency histogram has a single peak (**Supplementary Fig. S1, Table S1 and S2**). Syt1-SNARE's reference structure is PDB ID: 5CCG (Zhou *et al.*, 2015) but the C2A and C2B domains of Syt1 are from PDB ID: 2R83 (Huang *et al.*, 2013) which were used to fix the missing residue GLU130 (see **Supplementary Methods**). For MscL, the smFRET data (**Supplementary Fig. S2 and Table S3**) for the closed state *E. coli* MscL are from Wang *et al.* (Wang *et al.*, 2014) while the reference structure is PDB ID: 2OAR (Steinbacher *et al.*, 2007). The smFRET data (**Supplementary Table S4**) for the *Ct*H^+-PPase system are from Huang *et al.* (Huang *et al.*, 2013) and the reference structure is predicted by homology modeling using SwissModel (Biasini *et al.*, 2014) with the template structure from the *Vr*H^+-PPase system sharing a sequence identity of 45.1% (PDB ID: 4A01 (Lin *et al.*, 2012); the smFRET data for *Vr*H^+-PPase are mapped from *Ct*H^+-PPase to *Vr*H^+-PPase based on the pairwise sequence alignment (**Supplementary Fig. S3)**. Further details can be found in the **Supplementary Methods**.

Implementation of DR-SIP

The DR-SIP package is implemented in Python 2.7 (McKinney, 2010; Gowers *et al.*, 2016; Lam *et al.*, 2015; Cock *et al.*, 2009; Millman and Aivazis, 2011) and contains four modules: drsip, drsip-common, zdock-parser and docking-eval. The drsip module implements the docking protocols, filters and a command-line-interface (CLI) for users to run the DR-SIP docking protocols. When the docking results are derived from ZDOCK, the zdock-parser module is available to parse the ZDOCK output file, generate the docking poses and write the coordinates out to PDB files. On the other hand, the docking-eval module implements the CAPRI criteria used to evaluate the quality of docking poses. Lastly, drsip-common contains functions that are commonly used by the other modules.

Regular users can perform standard DR-SIP docking protocols with the CLI while more advanced users can import the specific modules and functions to perform their own customized docking protocol.

The source code and documentation on how to use the DR-SIP package are provided at https://github.com/capslockwizard/drsip. The packages are also distributed through the PyPi and Anaconda Cloud repositories.

Results

Statistical Analysis of All Experimentally Solved Structures of α-Helical HoTPs

To observe how common C_n symmetric structures are and how well the properties of C_n symmetry holds for α-helical HoTPs, 152 non-redundant structures containing HoTPs were taken from the mpstruct database (http://blanco.biomol.uci.edu/mpstruc/), see **Supplementary Tables**. Out of the 152 structures, 23 of them are monomers or do not contain a transmembrane domain. The remaining 129 structures (64% are observed within asymmetric units) contain 118 structures (includes two dihedral symmetric structures, see **Supplementary Methods**) with parallel orientation (N-termini of constituent monomers face the same side of the membrane) and 11 anti-parallel structures (see **Supplementary Results**). The 118 parallel structures which constitute 91.5% of the 129 HoTPs were used for subsequent statistical analyses. The anti-parallel structures were not included because most have not been verified as their functional forms through supporting experiments (see **Supplementary Results**).

The distribution of the sizes of the 118 parallel HoTPs (**Fig. 3A**) show that most are dimers (47 HoTPs), trimers (28 HoTPs) and tetramers (26 HoTPs), for a total of 101 HoTPs (85.6%) out of the 118 HoTPs.

We further identified 115 of the 118 parallel HoTPs to be C_n symmetric (**Fig. 3B-E**, C_n symmetry RMSD <2Å) concluding that most of the HoTPs (89.1% or 115/129) are C_n symmetric and parallel in orientation.

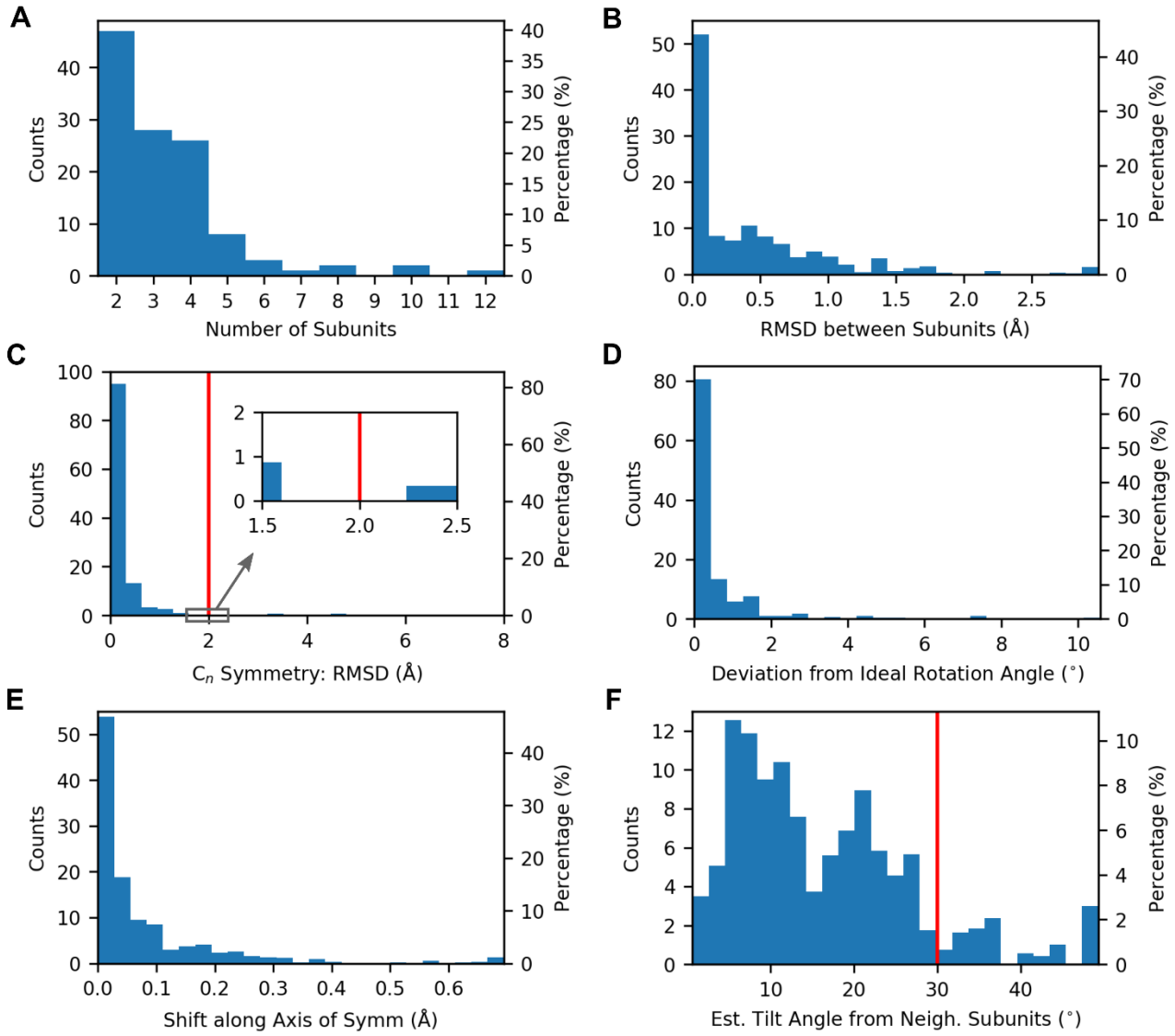

Fig. 3. Statistical distributions of the 118 HoTPs with parallel orientation. The left axis shows the counts of each bin while their percentages (out of 115 or 118 structures) are shown on the right axis. (**A**) Distribution of oligomeric states of the HoTPs reveals 85.6% (101/118) of the structures are dimers, trimers and tetramers. (**B**) Conformational differences between neighboring monomers measured by root-mean squared deviation (RMSD) of Cα atoms, shows that the conformational differences are small, all are <3Å while 87.9% <1Å. (**C**) The C_n symmetry RMSD measures the translational and rotational differences for a pair of neighboring monomers caused by deviation from ideal C_n symmetry (see **Methods**). The distribution includes the HoTPs within 8Å (except for PDB ID: 5HK1). The cutoff of <2Å (see red line and **inset**) is used to identify C_n symmetric HoTPs. 115 (97.5%) out of the 118 structures are within this cutoff. The remaining three structures (**Supplementary Results**), PDB IDs: 2GIF, 2QTS and 5HK1, with the lowest RMSD between neighboring monomers of 2.5Å, 4.6Å, 21.1Å, respectively, were classified as non-C_n symmetric. (**D** and **E**) The deviation from ideal C_n symmetry for the remaining 115 structures can be decomposed into *rotational differences* and *translational differences* which are measured by the (**D**) deviation from ideal rotation angle and

(**E**) shift along (parallel to) the AOS, respectively. The shift here refers to the amount of the translational difference between the two monomer's COM (**d**$_{diff}$, **Fig. 1B**) that are parallel to the AOS. (**F**) The tilt angle distribution shows the acute angle between the direction that the monomers span the membrane and the membrane normal (AOS). 89.9% of the 115 C_n symmetric structures have tilt angles <30° (red line) which is the cutoff employed in the SIP filter. (**B**-**F**) are calculated using all neighboring pairs of monomers in each HoTP. Each pair is weighted to contribute 1/*n* to the counts except for homo-dimers where the weight is one.

Decomposition of the Deviation from Ideal C_n Symmetry

There are two potential contributors to the deviation from ideal C_n symmetry. The first is the conformational difference between monomers in the complex. The other is the translational and rotational differences between a docking pose and its closest ideal C_n symmetric pose.

Our results show that conformational differences between monomers is not a factor preventing the formation of C_n symmetry. All the monomers in the 118 parallel HoTPs have small structural differences (**Fig. 3B**) – most (~90%) are less than 1Å, while all are less than 3Å.

On the other hand, the C_n symmetry RMSD (**Fig. 3C**) measures the translational and rotational differences between the ideal C_n symmetric pose and the pose from the X-ray structure for all unique pairs of neighboring monomers. When the C_n symmetry RMSD is <2.0Å (**Fig. 3C**, red line and inset), the pose is classified as C_n symmetric. This is the same cutoff used in the SIP filter. Among the 118 parallel HoTPs, 115 (97.5%) were classified as C_n symmetric.

The cutoff of <2.0Å was chosen because it is the midpoint between 1.6Å and 2.4Å where there are no structures (**Fig. 3C**). There are only three HoTPs (**Supplementary Fig. S4 and Table S5**) that have C_n symmetry RMSD >2.4Å and upon further examination were found to be non-C_n symmetric (see **Supplementary Results** for details on the three HoTPs). **Figure 4** compares the X-ray structures of the two HoTPs closest to the 2.0Å cutoff, one above it and the other below it, with their respective ideal C_n symmetric complex. The ExbB/ExbD complex (**Fig. 4A**) which is below the cutoff does not deviate much (see next section) from the ideal C_n symmetric complex while the multidrug efflux pump subunit AcrB (**Fig. 4B**) shows a visibly larger deviation (see next section).

Decomposition of Deviations from Ideal C_n Symmetry into Translation and Rotational Differences

Deviations from ideal C_n symmetry are decomposed into translational (shift along AOS, sAOS) and rotational (deviation from ideal rotation angle, devRot) differences. The statistics show that all 115 C_n symmetric HoTPs have <11° deviation from the ideal rotation angle (94% is <2°, **Fig. 1D**) and <0.7Å shift along AOS (95% is <0.3Å**, Fig. 1E**).

All the three non-C_n symmetric structures (**Supplementary Table S5**) have ≥0.7Å sAOS while two out of the three (acid-sensing ion channel and the trimeric AcrB [**Fig. 4B**]) have ≥11° devRot with the remaining one >6.0°. For comparison, the ExbB/ExbD complex (**Fig. 4A**) has 0.5Å sAOS and 4.0° devRot which are smaller than the three non-C_n symmetric structures.

Tilt Angles Distribution

The distribution of tilt angles (**Fig. 1F**) show that all the 115 structures have tilt angles <50°, while 90% of them is <30°; hence 30° (red line in **Fig. 1F**) is used as the tilt angle cutoff for the SIP filter.

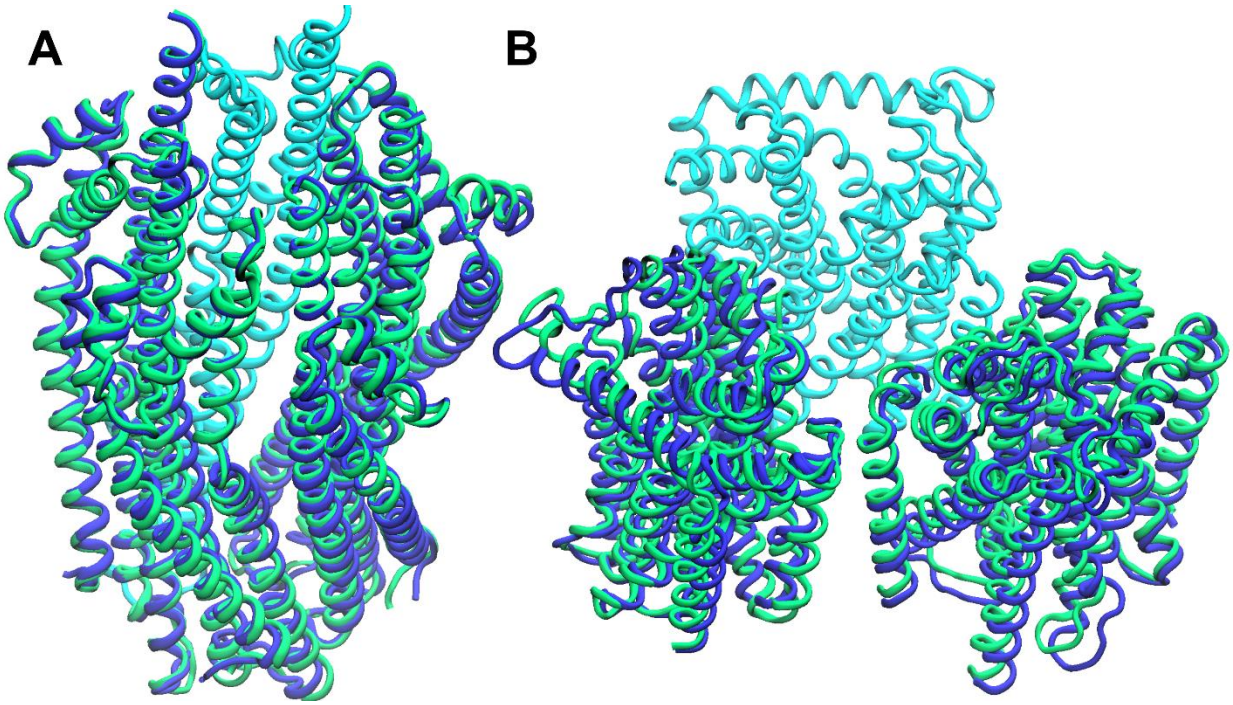

Fig. 4. Comparison of the X-ray structure (blue) and the ideal C_n symmetric complex (green) for the two HoTPs closest to the C_n symmetry RMSD cutoff of 2Å. The complexes are superimposed at chain A (cyan). (**A**) The protein that is closest to and below the cutoff is the ExbB/ExbD complex (PDB ID: 5SV0, C_n symmetry RMSD: 1.6 Å). (**B**) The protein that is closest to, but above the cutoff is the multidrug efflux pump subunit AcrB (PDB ID: 2GIF, RMSD: 2.5Å). There are more visible differences when the RMSD is above the chosen cutoff of 2Å.

Evaluation of DR-SIP: Near-Native Poses can be Found Within the Top 2 Ranking Results

The performance of DR-SIP is evaluated by applying the docking protocols to one soluble protein system (Syt1-SNARE) and three HoTP systems (MscL, *Vr*H^+-PPase and *Ct*H^+-PPase) (see **Fig 2**). DR-SIP's ranking of the near-native poses was compared to their corresponding ZDOCK ranking (**Table 1**).

In most cases DR-SIP performed substantially better than the original ZDOCK ranking. For all systems (**Table 1** and **Fig. 5**), a near-native pose can be found in the top two results. A near-native pose of Syt1-SNARE (**Fig. 5A**, best ranking near-native pose by original ZDOCK ranking: 62nd) was ranked 1st by DR-SIP, 2nd for MscL (**Fig. 5B**, ZDOCK: $2{,}231^{st}$) and 1st for *Ct*H^+-PPase and *Vr*H^+-PPase (**Fig. 5C**, ZDOCK: 1st for both systems).

The docking poses were evaluated by comparing them to their respective X-ray structures by the CAPRI criteria (Méndez *et al.*, 2003) which classifies them into one of four ranks: High, Medium, Acceptable and Incorrect, respectively (see **Supplementary Methods** and **Table S6**). Docking poses that are not "Incorrect" (<10% native contact or [>4.0 iRMSD and >10.0 lRMSD]) are considered a near-native pose. The ranks of the top near-native docking pose for all systems are shown in **Table 1**.

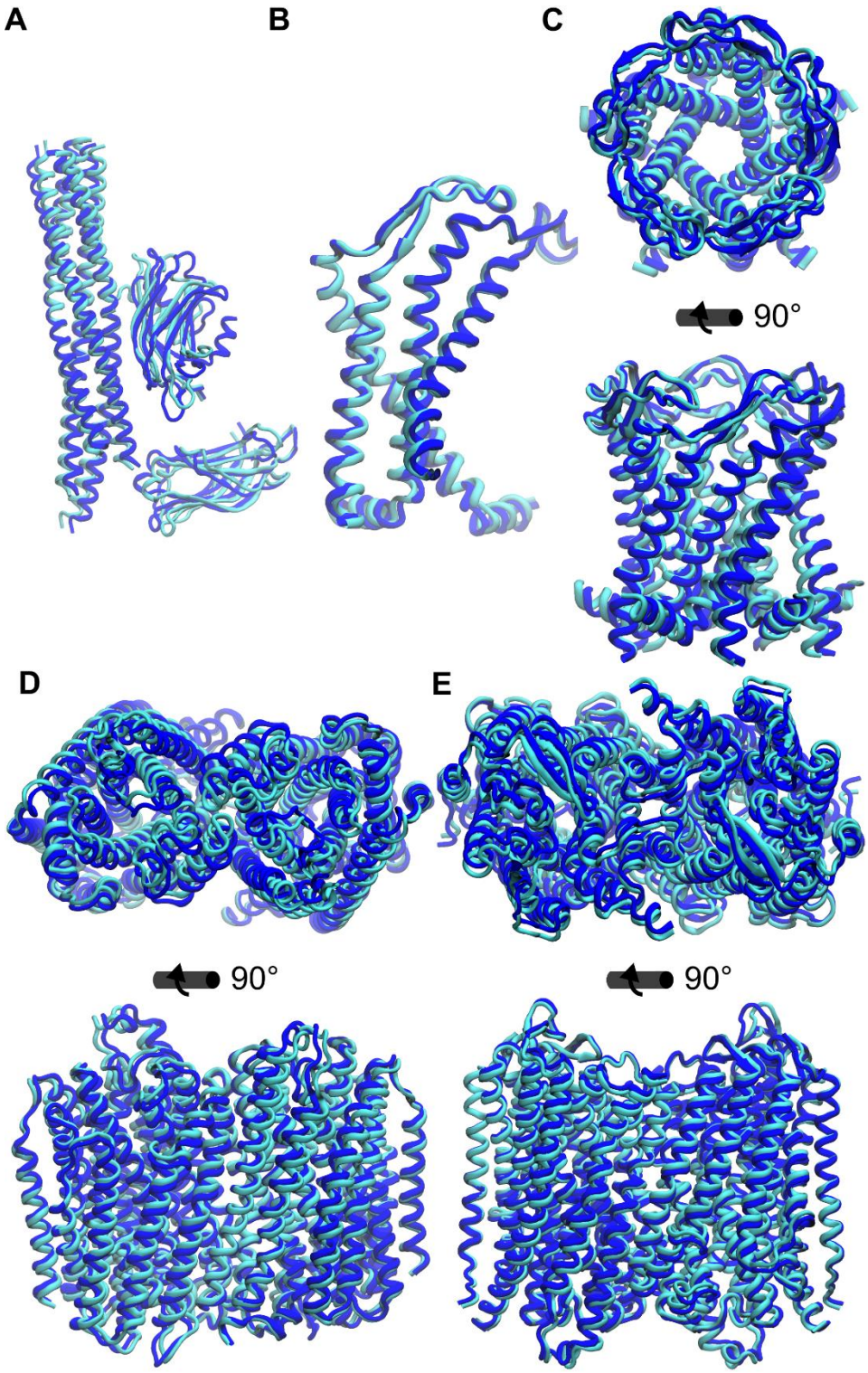

Fig. 5. The top near-native docking pose (cyan) predicted by DR-SIP superimposed onto the reference structure (blue) for all four systems. (**A**) Syt1-SNARE complex: the 1st ranked DR-SIP pose with the reference structure PDB ID: 5CCG. **(B-C)** MscL: 2nd ranked pose with the reference structure from PDB ID: 2OAR. (**B**) The docking pose (dimer) superimposed onto chain A and B of the reference structure. (**C**) The predicted pentamer complex with the reference

structure. (**D**) *Ct*H^+-PPase: 1st ranked pose with the homology modeled structure based on the template PDB ID: 4A01. (**E**) *Vr*H^+-PPase: 1st ranked pose with the reference structure PDB ID: 4A01.

Table 1. Characteristics of the top-ranked near-native predicted docking poses for the Syt1-SNARE, MscL, *Ct*H$^+$-PPase and *Vr*H$^+$-PPase systems.

		Syt1-SNARE	MscL	*Ct*H$^+$-PPase	*Vr*H$^+$-PPase
DR-SIP Rank[a]		1	2	1	1
Symmetry-Imposed Packing (SIP)[b]	**C_n Symm. RMSD (Å)**	-	0.53	0.8	1.36
	Tilt Angle (°)	-	28.01	5.95	14.21
	Orientation	-	Parallel	Parallel	Parallel
Distance Restraints (DR)[b]	**Pearson Corr.**	0.95	0.61	0.5	0.48
	Spearman Corr.	0.9	0.66	0.6	0.6
		Comparison between the top predictions and the X-ray structure (Native Pose)			
CAPRI Rank[c]		Acceptable	Medium	Medium	Medium
% Native Contacts		77.78	94.59	60.82	64.05
iRMSD (Å)		3.22	1.01	1.96	1.59
lRMSD (Å)		8.46	1.77	3.78	3.94

[a] DR-SIP rank is the rank of the cluster of poses ordered by size in descending order. The data shown are for the representative pose of the respective clusters (see **Methods** and **Fig. 2**).

[b] A large portion of HoTP poses were filtered out by the SIP filter (C_n symmetry >2Å, tilt angle >30°, orientation is anti-parallel) and DR filter (≤0.3 correlation), see **Methods**, **Fig. 2**. and **Table 2**, while Syt1-SNARE was filtered using only the DR filter.

[c] The CAPRI rank is determined based on the % Native contacts, iRMSD and lRMSD (details in the **Supplementary Methods** and **Table S6**). The reference native poses are from PDB ID: 5CCG, 2OAR, and 4A01 for Syt1-SNARE, MscL and *Vr*H$^+$-PPase, respectively. For *Ct*H$^+$-PPase, the complex obtained from homology modeling with PDB ID: 4A01 as the template is the reference native pose.

Table 2. Contribution of the DR-SIP Filters to Enriching the Near-Native Poses Generated by ZDOCK.

	% Native[c]	Enrichment[d]	Near-Native Poses			Incorrect	Total
			High[e]	Medium[e]	Acceptable[e]		
Syt1-SNARE							
Unfiltered	0.25	1.00	3	23	108	53866	54000
DR Filter	2.70	10.80	3	23	108	4824	4958
Filtered[a] & Clustered	1.65	6.60	0	1	6	418	425
Top 2 Clusters	50.00	**200.00**	0	0	1	1	2
MscL							
Unfiltered	0.18	1.00	5	30	63	53902	54000
All Filters[b]	26.23	145.72	4	9	3	45	61
Ex. DR Filter	6.23	34.61	4	9	3	241	257
Ex. C_n Symm. RMSD	1.86	10.33	4	11	9	1263	1287
Ex. Tilt Angle	9.31	51.72	5	21	13	380	419
Ex. Orientation	17.78	98.78	4	9	3	74	90
Filtered[b] & Clustered	20.00	111.11	0	1	0	4	5
Top 2 Clusters	50.00	**277.78**	0	1	0	1	2
***Ct*H$^+$-PPase**							
Unfiltered	0.14	1.00	0	29	45	53926	54000
All Filters[b]	32.39	231.36	0	12	11	48	71
Ex. DR Filter	12.78	91.29	0	12	11	157	180
Ex. C_n Symm. RMSD	3.16	22.57	0	20	38	1777	1835
Ex. Tilt Angle	14.65	104.64	0	12	11	134	157
Ex. Orientation	32.08	229.14	0	19	15	72	106
Filtered[b] & Clustered	25.00	178.57	0	1	0	3	4
Top 2 Clusters	50.00	**357.14**	0	1	0	1	2
***Vr*H$^+$-PPase**							
Unfiltered	0.17	1.00	4	29	58	53909	54000
All Filters[b]	24.22	143.72	3	13	15	97	128
Ex. DR Filter	12.76	75.70	3	13	15	212	243
Ex. C_n Symm. RMSD	3.22	19.12	4	29	58	2734	2825
Ex. Tilt Angle	11.65	69.16	3	13	15	235	266
Ex. Orientation	20.39	121.02	3	13	15	121	152
Filtered[b] & Clustered	11.11	65.93	0	1	0	8	9
Top 2 Clusters	50.00	**296.70**	0	1	0	1	2

The "Ex." here means that all the other filters/criteria were applied except the listed filter/criterion. The CAPRI criteria are used to rank the docking poses into one of the following: High, Medium, Acceptable or Incorrect (details in **Supplementary Methods** and **Table S6**).

[a] Filtered with only the DR filter.

[b] Filtered by both the DR and SIP.

[c] % Native = (High + Medium + Acceptable) Poses / Total Poses * 100.

[d] Enrichment = (% Native) / (% Native when unfiltered). For example, MscL's final enrichment of 277.8 is obtained by dividing the 50% near-native poses from the top 2 clusters with the ~0.18% near-native poses in the unfiltered data.

Contribution of the Filters and Clustering to the Enrichment of Near-Native Poses

Among the 54,000 docking poses generated by ZDOCK for each of the four systems, <0.3% (**Table 2**) of the poses are near-native poses. The docking protocols enriched the near-native poses by ~200-360 fold such that the near-native poses can be found within the top two ranking results (**Table 1 and 2**).

The contribution of the filters is measured by excluding one of the four filters/criteria and comparing the decrease in the enrichment of near-native poses relative to the enrichment when using all filters/criteria (see **Table 2**). The larger the drop in the enrichment, the larger the contribution and vice versa.

The C_n symmetry criteria has the largest effect on the enrichment resulting in a drop of ~90% when it is excluded. This is followed by the tilt angle and relative distance filters which are of equal importance where they reduce the enrichment by ~50-70%. Lastly, the orientation filter shows the smallest reduction in the enrichment of ~1-30%.

After applying all the filters on the HoTP systems, the near-native poses make up ~20%-30% of the remaining poses. This is a significant enrichment (up to 230 fold) of the near-native poses from the <0.3% observed in the 54,000 unfiltered poses. The near-native poses are further enriched by the clustering step such that the near-native pose can be found within the top two results.

For the soluble case, the Syt1-SNARE complex, the DR filter enriches the near-native poses by ~11 fold. Clustering the remaining poses and ranking of the clusters results in a near-native pose residing in the top two results, further enriching the near-native poses by another ~19 fold.

Performance of DR-SIP with GRAMM-X

Besides ZDOCK, we would like to examine how DR-SIP performs for poses generated by other software. DR-SIP was therefore applied to the docking poses generated from GRAMM-X for all four systems.

From the maximum 300 poses that could be obtained for each system, GRAMM-X results contain the near-native pose(s) for MscL, *Vr*H^+-PPase and *Ct*H^+-PPase but not any near-native pose for the Syt1-SNARE complex. Applying DR-SIP (excluding Syt1-SNARE) enriches the near-native poses to >50% from the otherwise unfiltered chance of 0.3-2.7% (**Table 3**), which could qualify DR-SIP as a general filter for docking results obtained from different docking packages.

Table 3. GRAMM-X Docking Results Filtered by the DR-SIP Protocols.

System	% Native Poses [Before Filter]	% Native Poses [After Filter]	Enrichment[a]
MscL	2.7 (8/300)	66.7 (2/3)	24.7
***Ct*-PPase**	0.3 (1/300)	100.0 (1/1)	333.3
***Vr*-PPase**	0.3 (1/300)	50.0 (1/2)	166.7
Syt1-SNARE	0.0 (0/300)	0.0 (0/300)	-

GRAMM-X returns a maximum of 300 top docking poses. The parentheses show the number of (near-native poses/total poses). Near-native poses were identified by the CAPRI criteria (see **Supplementary Methods** and **Table S6**).

[a] Enrichment = % Native Poses [After Filter] / % Native Poses [Before Filter]

Discussions

The distribution of the "*n*" of C_n symmetric HoTPs (**Fig. S5A** and **Table S7**) is not uniformly distributed. Instead, the dimers and trimers make up most, 63.5%, of the structures. We wonder whether such a distribution is due to a natural outcome of physiochemical/geometric complementarity of constituent monomers (physical reason) or other selection criteria to meet functional needs of living (biological reason).

We assume the observed distribution is a realization of all the possible C_n symmetry a single constituent monomer could form. Resembling the idea of ergodicity where ensemble average/distribution is equal to time average/distribution for a single molecule that is sampled for long time, here we assume the probability distribution of how often a certain C_n symmetry is chosen by a particular HoTP as the global minimum is close to how frequent that C_n symmetry is visited by any (or a random) HoTP's constituent monomer. In other words, we consider a constituent monomer that keeps sampling different C_n symmetries for a long time and these C_n symmetries are preferred in the same way as the observed distribution of the "*n*" of C_n symmetric HoTPs. Instead of examining the possible C_n symmetries sampled by a given HoTP's monomer using MD simulations, which is technically unfeasible at the moment, we use ZDOCK as an off-lattice model (Huang, 2014) that rewards physiochemical/geometric complementarity of two interacting monomers to explore a distribution of "*n*" using three HoTP systems (MscL, *Vr*H^+-PPase and *Ct*H^+-PPase). Non-C_n symmetric poses are removed with the C_n symmetry criterion.

The results show that the size distributions of the C_n symmetric poses from docking are over-represented by dimers and trimers, making up ~83% of total (**Fig. S5A, S5C** and **Table S7**), which is much higher than 63.5% in the HoTP dataset. Not unique to HoTPs, C_n symmetric soluble proteins (Pentameric Capsid Protein (PDB ID: 4dmi) and GH3 Beta-Glucosidase (PDB ID: 5xxl) that are selected for having a similar size as MscL and *Vr*H^+-PPase/*Ct*H^+-PPase, respectively; see Supporting Methods for details) also show an over-representation of dimers and trimers of ~75%.

As we know, transmembrane proteins need to span the membrane bilayer which imposes geometrical restraints on the protein. One such restraint is the tilt angle. When the transmembrane regions of the monomers in the protein complex are longer than the thickness of the membrane, the monomers undergo a tilt (Park and Opella, 2005; Kim and Im, 2010; de Jesus and Allen, 2013). Our statistics (**Fig. 3F**) shows that smaller tilt angles ($<30^o$) are preferred, which could indicate that most of the transmembrane regions are just long enough to span the membrane but not too long. Also, overly large tilting angle may reduce contact interface of monomers in HoTPs. However, there is no straightforward geometrical or physical explanation of why tilt angles are populated in 10° and 20° (**Fig. 3F**), and therefore we have to surmise that the "selection" is made for functional reason.

After applying the tilt angle criterion ($<30^o$) to the C_n symmetric docking poses, we start to see a distributions similar (dimers and trimers make up ~68%) to that of the HoTP dataset (**Fig. S5B, S5D** and **Table S7**). As a result, the tilting angle of monomers determines the observed distribution. The spatial restriction of HoTPs to situate inside the 2D membrane imposes a restraint on the possible tilting angle. However, why $<30^o$ instead of employing a slightly smaller or larger angle could be due to genetic or functional reasons in the evolution.

Taken together, DR-SIP incorporates the physicochemical restraints by the membrane on HoTPs such as the preference for C_n symmetry and tilt angles while leveraging experimentally determined distance restraints providing a suite of powerful docking protocols to cull the near-native quaternary structures of hetero-/homo-oligomers from tens of thousands possible docking decoys with high confidence.

Acknowledgments

Vast computational resources for this project are provided by National Center for High-performance Computing (NCHC) of National Applied Research Laboratories (NARLabs), Taiwan. J.C. acknowledges financial support from Taiwan International Graduate Program (TIGP), Academia Sinica, Taipei, Taiwan. This work is supported by the Ministry of Science and Technology, Taiwan (103-2627-M-007-001 and 104-2113-M-007-019 to L.-W. Y.) and Academia Sinica, Taipei, Taiwan (AS 103-TP-B11 to Dr. Tzyy-Jen Chiou). We would like to thank Dr. Tzyy-Jen Chiou for her support and advice.

References

Alber,F. *et al.* (2008) Integrating diverse data for structure determination of macromolecular assemblies. *Annu. Rev. Biochem.*, **77**, 443–477.

Alford,R.F. *et al.* (2015) An integrated framework advancing membrane protein modeling and design. *PLoS Comput. Biol.*, **11**, e1004398.

Bertoni,M. *et al.* (2017) Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci. Rep.*, **7**, 1–15.

Biasini,M. *et al.* (2014) SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, **42**, W252-8.

Brooijmans,N. and Kuntz,I.D. (2003) Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 335–73.

Chang,C.-C. *et al.* (2016) Blocking the interaction between S100A9 and RAGE V domain using CHAPS molecule: A novel route to drug development against cell proliferation. *Biochim. Biophys. Acta - Proteins Proteomics*, **1864**, 1558–1569.

Chaudhury,S. and Gray,J.J. (2008) Conformer selection and induced fit in flexible backbone protein-protein docking using computational and NMR ensembles. *J. Mol. Biol.*, **381**, 1068–87.

Choi,U.B. *et al.* (2010) Single-molecule FRET-derived model of the synaptotagmin 1-SNARE fusion complex. *Nat. Struct. Mol. Biol.*, **17**, 318–24.

Cock,P.J.A. *et al.* (2009) Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–3.

Dimura,M. *et al.* (2016) Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. *Curr. Opin. Struct. Biol.*, **40**, 163–185.

Fagerberg,L. *et al.* (2010) Prediction of the human membrane proteome. *Proteomics*, **10**, 1141–1149.

Gowers,R. *et al.* (2016) MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations. In, Benthall,S. and Rostrup,S. (eds), *Proceedings of the 15th Python in Science Conference*. Austin, TX, pp. 98–105.

Gupta,A.A. *et al.* (2013) Structural insights into the interaction of human S100B and basic fibroblast growth factor (FGF2): Effects on FGFR1 receptor signaling. *Biochim. Biophys. Acta - Proteins Proteomics*, **1834**, 2606–2619.

Huang,S.Y. (2014) Search strategies and evaluation in protein-protein docking: Principles, advances and challenges. *Drug Discov. Today*, **19**, 1081–1096.

Huang,Y.-T. *et al.* (2013) Squeezing at entrance of proton transport pathway in proton-translocating pyrophosphatase upon substrate binding. *J. Biol. Chem.*, **288**, 19312–20.

Hurwitz,N. *et al.* (2016) Memdock: An α-helical membrane protein docking algorithm. *Bioinformatics*, **32**, 2444–50.

de Jesus,A.J. and Allen,T.W. (2013) The determinants of hydrophobic mismatch response for transmembrane helices. *Biochim. Biophys. Acta*, **1828**, 851–863.

Jia,Y.-B. (2017) Rotation in the space. *Com S 477/577 Notes*, 1–11.

Kabsch,W. (1976) A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. Sect. A*, **32**, 922–923.

Kalinin,S. *et al.* (2012) A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods*, **9**, 1218–1227.

Khan,M.I. *et al.* (2018) S100B as an antagonist to block the interaction between S100A1 and the RAGE V domain. *PLoS One*, **13**, e0190545.

Kim,T. and Im,W. (2010) Revisiting hydrophobic mismatch with free energy simulation studies of transmembrane helix tilt and rotation. *Biophys. J.*, **99**, 175–183.

Kozakov,D. *et al.* (2017) The ClusPro web server for protein–protein docking. *Nat. Protoc.*, **12**, 255–278.

Lam,S.K. *et al.* (2015) Numba. In, *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC - LLVM '15*. ACM Press, New York, New York, USA, pp. 1–6.

Levy,E.D. *et al.* (2008) Assembly reflects evolution of protein complexes. *Nature*, **453**, 1262–1265.

Liang,B. and Tamm,L.K. (2016) NMR as a tool to investigate the structure, dynamics and function of membrane proteins. *Nat. Struct. Mol. Biol.*, **23**, 468–474.

Lin,S.-M. *et al.* (2012) Crystal structure of a membrane-embedded H+-translocating pyrophosphatase. *Nature*, **484**, 399–403.

Manning,C.D. *et al.* (2008) Introduction to information retrieval 1st ed. Cowles,L. (ed) Cambridge University Press, New York, NY, USA.

McKinney,W. (2010) Data structures for statistical computing in Python. In, *Proceedings of the 9th Python in Science Conference*., pp. 51–56.

Méndez,R. *et al.* (2003) Assessment of blind predictions of protein-protein interactions: Current status of docking methods. *Proteins Struct. Funct. Genet.*, **52**, 51–67.

Millman,K.J. and Aivazis,M. (2011) Python for scientists and engineers. *Comput. Sci. Eng.*, **13**, 9–12.

Mitra,K. and Frank,J. (2006) Ribosome dynamics: Insights from atomic structure modeling into cryo-electron microscopy maps. *Annu. Rev. Biophys. Biomol. Struct.*, **35**, 299–317.

Moraes,I. *et al.* (2014) Membrane protein structure determination — The next generation.

Biochim. Biophys. Acta - Biomembr., **1838**, 78–87.

Muschielok,A. *et al.* (2008) A nano-positioning system for macromolecular structural analysis. *Nat. Methods*, **5**, 965–971.

Park,S.H. and Opella,S.J. (2005) Tilt angle of a trans-membrane helix is determined by hydrophobic mismatch. *J. Mol. Biol.*, **350**, 310–318.

Pierce,B. *et al.* (2005) M-ZDOCK: A grid-based approach for Cn symmetric multimer docking. *Bioinformatics*, **21**, 1472–1478.

Pierce,B.G. *et al.* (2011) Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS One*, **6**, e24657.

Rask-Andersen,M. *et al.* (2014) The druggable genome: Evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. *Annu. Rev. Pharmacol. Toxicol.*, **54**, 9–26.

Rose,P.W. *et al.* (2017) The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.*, **45**, D271–D281.

Sasmal,D.K. *et al.* (2016) Single-molecule fluorescence resonance energy transfer in molecular biology. *Nanoscale*, **8**, 19928–19944.

Schneidman-Duhovny,D. *et al.* (2005) PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res.*, **33**, W363–W367.

Steinbacher,S. *et al.* (2007) Structures of the prokaryotic mechanosensitive channels MscL and MscS. In, *Mechanosensitive Ion Channels, Part A*., pp. 1–24.

Szilagyi,A. and Zhang,Y. (2014) Template-based structure modeling of protein-protein interactions. *Curr. Opin. Struct. Biol.*, **24**, 10–23.

Tovchigrechko,A. and Vakser,I.A. (2006) GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res.*, **34**, W310–W314.

Viswanath,S. *et al.* (2015) Extension of a protein docking algorithm to membranes and applications to amyloid precursor protein dimerization. *Proteins*, **83**, 2170–2185.

van der Walt,S. *et al.* (2011) The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.*, **13**, 22–30.

Wang,Y. *et al.* (2014) Single molecule FRET reveals pore size and opening mechanism of a mechano-sensitive ion channel. *Elife*, **2014**, 1–21.

Yamagata,A. and Tainer,J.A. (2007) Hexameric structures of the archaeal secretion ATPase GspE and implications for a universal secretion mechanism. *EMBO J.*, **26**, 878–90.

Zhou,Q. *et al.* (2015) Architecture of the synaptotagmin-SNARE machinery for neuronal exocytosis. *Nature*, **525**, 62–67.

van Zundert,G.C.P. *et al.* (2016) The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.*, **428**, 720–725.