Modeling Symmetrical CASP/CAPRI Complexes using SAM

Dave Ritchie
Inria Nancy

Symmetry in Protein Quaternary Structures
Many protein complexes have quaternary symmetry
- How? – Two or more asymmetric monomers related by symmetry operators

Cn: cyclic n = 1, 2, 3, ...
Dn: dihedral n = 2, 3, 4, ...
T: tetrahedral n = 12
O: octahedral n = 24
I: icosahedral n = 60
H: helical n = ...

2XQT: C15 1A6D: D4 1J2Y: T 1EAB: O 1A34: I 1CGM: H

D = 4  D = 6  D = 6  D = 6  D = 6  D = 6


Symmetry reduces the number of degrees of freedom (D)
- So, symmetrical complexes should be easy for CAPRI dockers?

3D-Complex Shows Many Symmetrical Structures in PDB
Levy et al. (2008), Nature, 453, 1262–1265
Existing Symmetry Docking Approaches

Post-Filter 3D FFT Grid
- Molfit ($D_2$): Berchanski et al. (2003), Proteins, 53, 817–829
- Cluspro ($C_n, D_2, D_3$): Comeau et al. (2005), JSB, 150, 233-244

Symmetry-Constrained 3D FFT / Geometric Hashing
- M-ZDOCK ($C_n$): Pierce et al. (2005), Bioinformatics, 21, 1472–1478
- SymmDock ($C_n$): Schneidman et al. (2005), Proteins, 60, 224–231

MD-based Energy-Minimisation
- Rosetta3 (any): André et al. (2007), PNAS, 104, 17676–17661
- Haddock ($C_n, D_n$): Karaca et al. (2010), Mol Cell Prot, 9, 1784–1794

Spherical Polar Fourier (SPF) Representations

Represent protein shape as a 3D shape-density function...

$$\tau(r) = \sum_{n,m} a_{nlm} R_{nl}(r) y_{lm}(\theta, \phi)$$

...using spherical harmonic, $y_{lm}(\theta, \phi)$, and radial, $R_{nl}(r)$, basis functions

Image Order Coefficients
A Gaussians -
B $N = 16$ 1,496
C $N = 25$ 5,525
D $N = 30$ 9,455

Protein Docking Using SPF Density Functions

Favourable:
$$\int (\sigma_A(\mathbf{r}_A) \tau_B(\mathbf{r}_B) + \tau_A(\mathbf{r}_A) \sigma_B(\mathbf{r}_B)) dV$$

Unfavourable:
$$\int \tau_A(\mathbf{r}_A) \tau_B(\mathbf{r}_B) dV$$

Score:
$$S_{AB} = \left(\int (\sigma_A(\mathbf{r}_A) + \tau_A(\mathbf{r}_A)) \sigma_B(\mathbf{r}_B) dV\right)^2 - P$$

Orthogonality:
$$S_{AB} = \sum_{n,m} (a_{nlm}^* b_{nlm}^* + b_{nlm} a_{nlm})$$

Search:
6D space = 1 distance + 5 Euler rotations: $(R, \beta_A, \gamma_A, \alpha_B, \beta_B, \gamma_B)$

Coordinate Operators and Docking Equations

Describe search space using operators
- Rotation: $R(\alpha, \beta, \gamma)$
- Translation: $T_z(R)$

Describe interaction as an “equation”
- $\hat{R}(0, \beta_A, \gamma_A) A(z) \leftrightarrow \hat{T}_z(R) \hat{R}(\alpha_B, \beta_B, \gamma_B) B(z)$

Can re-write this in many ways...
- $\hat{T}_x(R)^{-1} \hat{R}(0, \beta_A, \gamma_A) A(z) \leftrightarrow \hat{R}(\alpha_B, \beta_B, \gamma_B) B(z)$

In SPF basis, score as an overlap integral
- $S_{AB} = \int (\hat{T}_x(R)^{-1} \hat{R}(0, \beta_A, \gamma_A) A(z))^* [\hat{R}(\alpha_B, \beta_B, \gamma_B) B(z)]$
The Docking Equation for $C_n$

SPF translations are easiest in the $z$ direction
- So let $y$ be principal symmetry axis, $\omega$ be symmetry angle

$$\hat{R}(\omega_{y+1}) \hat{T}_z(D) \hat{R}(\alpha, \beta, \gamma) B(\omega) \rightarrow \hat{R}(\omega) \hat{T}_z(D) \hat{R}(\alpha, \beta, \gamma) A(\omega)$$
- $0 \leq \alpha < \pi, 0 \leq \beta < \pi, 0 \leq \gamma < 2\pi$ (variable Euler angles)
- $0 \leq D \leq 50$ Ångstrom (variable inter-molecular distance)
- $\omega_j = 2\pi j/n$ (fixed by symmetry; just need $j = 0$ and $j = 1$)

One-Dimensional FFT Docking with $C_n$ Symmetry

Re-order the docking equation to collect terms in $\alpha$
- $\hat{T}_z(D)^{-1} \hat{R}(0, \beta', \gamma') A(\omega) \rightarrow R_z(\alpha')^{-1} \hat{R}(\omega) \hat{R}_z(\alpha') \hat{R}(0, \beta', \gamma') B(\omega)$

Apply operators to SPF expansion coefficients
- $B_{nmp} = \sum_{m'} D_{m'm}^{(0, \beta', \gamma')} B_{nmp'}$
- $A_{nmp} = \sum_{j,k} T_{n',k} (-D) B_{nk'm'}$

Expand the $\alpha$ and $\omega$ rotations in a similar way
- $S(\alpha', \omega, D, \beta', \gamma') = \sum_{nlp} A_{nmp} e^{-iD(n-m)\alpha'} B_{nmp}^{*}$

Scale onto $2\pi$ ($\alpha'' = 2\alpha'$) and collect coeffs for 1D FFT in $\alpha''$
- $S(\alpha', \omega, D, \beta', \gamma') = \sum_{l} C_l e^{-i2\pi \alpha''}$

Technical Note – Working Near the Origin

- With Polar Fourier correlations, it is best to work near the origin
- Computational frame (left) for FFT
- Symmetry frame (right) for results

$\hat{R}(\alpha', \beta', \gamma') A(\omega) \rightarrow \hat{T}_z(D) \hat{R}(\omega) \hat{R}(\alpha', \beta', \gamma') B(\omega)$
- SAM does the calculation in the computational frame (left)...
- ... and transforms solutions back to the symmetry frame (right)

$C_3$ Example (Rigid-Body Reconstruction)

1F7O: Feline immuno-deficiency virus DUTP pyrophosphatase
- 107 AA, 2.2Å resolution
- SPF expansions to $N=30$
- 0.8 Å steps in $D$
- 2.8° steps in $\alpha$ (by FFT)
- 7.5° steps in ($\beta$, $\gamma$)

Performance Comparison (dual 6-core Intel X5650 2.67 GHz)
- SAM: 1st solution, 2.82 Å RMSD, 48 seconds
- M-ZDOCK: 1st solution, 2.33 Å RMSD, 4641 seconds
- SymmDock: 1st solution, 2.32 Å RMSD, 14 seconds
Building $D_n$ Complexes From Two $C_n$ Solutions

- For $D_n$, define 2 more operators: $\hat{T}_z(E)$ and $\hat{R}_z(\eta)$

- Also need allow a possible flip of one $C_n$ plane relative to the other...
- Then dock $k$ pairs of trial $C_3$ pseudo-molecules ($i = 0$ or $1$ for flips):
  \[ P_{ki}(t) = A_{ki}(t) + B_{ki}(t) + C_{ki}(t) \]
  (just do by brute-force search in $E$ and $\eta$, no FFT here)

Making Higher Symmetries From $C_3$ Trimers

- $T$, $O$, and $I$ all have multiple $C_3$ axes
- But like $D_n$, we still have only $4+2$ degrees of freedom
- So make positioning operators from geometry of platonic solids...

- Put a $C_3$ pseudo-molecule at two vertices and score as before...

Results – Selected 3D-Complex Examples

- All except 2 solutions are rank-1, RMSD < 3 Å
- D5/1l6w: rank-5, 1.3/3.6 Å RMSD; D8/1q3r: rank-25, 3.6/10.8 Å RMSD

Results – All 3D-Complex Structures

- Tested on (nearly) ALL point-group symmetry structures in 3D-Complex

- Main limitation is size of monomer (approx 500 residue limit)
- About 89% of 3D-Complex structures have monomers ≤500 residues...
Using KBDOCK to Find Symmetric DDI Templates

- PfamScan to find Pfam family
- Select “homo” DDIs...
- Then, eye-ball for symmetry

Ghoorah et al. (2014), Nucleic Acids Research, 42, D389–D395

Using Kpax to Select a “Representative” CASP Model

- Download the “Stage-2” CASP models...
- Find centre model using Kpax all-vs-all structural alignments (e.g., T85)

kpx -all T0813/*
Zhang_Server_TS4

Ritchie et al. (2012), Bioinformatics, 28, 3274–3281

The Round 30 Targets That We Attempted

<table>
<thead>
<tr>
<th>Target</th>
<th>Category</th>
<th>CASP Stage-2</th>
<th>KBDOCK</th>
<th>KBDOCK/SAM</th>
<th>CAPRI Top-10</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>T79</td>
<td>DIFFICULT</td>
<td>FASA03_T51</td>
<td>JRCG</td>
<td>1 / 0</td>
<td>M01 / KBDOCK</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>T80</td>
<td>EASY</td>
<td>FALCON_MANUAL_T52</td>
<td>2OGA 3NYU</td>
<td>1–2 / 3–10</td>
<td>M01 / KBDOCK</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>T81</td>
<td>HETERO</td>
<td>STRINGS_T51</td>
<td>G2IN</td>
<td>1–2 / 3–10</td>
<td>M01 / KBDOCK</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>T82</td>
<td>EASY</td>
<td>FALCON_ENV_FOLD_T54</td>
<td>3ED8</td>
<td>1 / 2–10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T83</td>
<td>CANCELED</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T84</td>
<td>EASY</td>
<td>PHYX_T51</td>
<td>4B5B</td>
<td>1 / 2–10</td>
<td>M01 / KBDOCK</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>T85</td>
<td>EASY</td>
<td>ZHANG_SERVER_T54</td>
<td>2FKK 3GCP</td>
<td>1–2 / 3–10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T86</td>
<td>DIFFICULT</td>
<td>FALCON_ENV_FOLD_T55</td>
<td>3IHU 3QY 3ER7</td>
<td>1–3 / 4–10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T87</td>
<td>EASY</td>
<td>FALCON_ENV_FOLD_T51</td>
<td>2ROR</td>
<td>0 / 1–10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T88</td>
<td>DIFFICULT</td>
<td>PHYX_T51</td>
<td>2X1U</td>
<td>1 / 2–10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T89</td>
<td>HETERO</td>
<td>--</td>
<td>0 / 1–10 (*)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T90</td>
<td>EASY</td>
<td>--</td>
<td>0 / 1–10 (*)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* used Hex instead of SAM

- We were the only predictor group to get acceptable models for T86
- Overall, we modeled 4 / 9 symmetric homo-dimers (2 KBDOCK, 2 SAM)
- But our models for T84 and T87 were rejected due to steric clashes...
Our Best Solution With KBDOCK – T80

- T80–M01: $F_{nat} = 0.493$, LRMSD = 3.77 Å, IRMSD = 1.97 Å

Our Best Solution With SAM – T86

- T86–M09: $F_{nat} = 0.667$, LRMSD = 8.97 Å, IRMSD = 2.55 Å

Question: Is “LRMSD” really appropriate for symmetry targets?

No doubt, RMSD(Receptor+Ligand) would give a smaller number...

Our Rejected Solutions – T84 and T87

- T84–M01 (KBDOCK), $F_{nat}=0.718$, LRMSD=3.22, IRMSD=1.28
- T87–M02 (SAM), $F_{nat}=0.473$, LRMSD=2.45, IRMSD=2.20

Dear Organisers, Please Stop Rejecting Our Best Solutions!

Conclusions

SAM
- New docking code for arbitrary point group symmetry
  (every model is perfectly symmetrical)

Docking Equations
- Very useful for describing the docking problem
  (not only for doing FFTs!)

CAPRI Assessment
- Is LRMSD the right measure for symmetry targets?
- How about a new category for homology / soft docking?
Thank You!

SAM: Ritchie & Grudinin (2016), J Appl Cryst, 49, 158–167

http://sam.loria.fr/
http://kbdock.loria.fr/
http://kpax.loria.fr/