Modeling Symmetrical Protein Complexes using SAM Soft Docking

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

![Symmetry in Protein Quaternary Structures](http://www.rcsb.org/pdb/staticHelp.do?p=help/viewers/jmol_symmetry_view.html)

- Symmetry reduces the number of degrees of freedom (D)

![Images of protein structures with different symmetries](http://www.rcsb.org/pdb/staticHelp.do?p=help/viewers/jmol_symmetry_view.html)
3D-Complex Shows Many Symmetrical Structures in PDB

Levy et al. (2008), Nature, 453, 1262–1265
Spherical Polar Fourier Representations

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

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

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

<table>
<thead>
<tr>
<th>Image</th>
<th>Order</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Gaussians</td>
<td>-</td>
</tr>
<tr>
<td>B</td>
<td>N = 16</td>
<td>1,496</td>
</tr>
<tr>
<td>C</td>
<td>N = 25</td>
<td>5,525</td>
</tr>
<tr>
<td>D</td>
<td>N = 30</td>
<td>9,455</td>
</tr>
</tbody>
</table>
Protein Docking Using Shape-Density Functions

\begin{align*}
\text{Favourable:} & \quad \int (\sigma_A(r_A)\tau_B(r_B) + \tau_A(r_A)\sigma_B(r_B))dV \\
\text{Unfavourable:} & \quad \int \tau_A(r_A)\tau_B(r_B)dV \\
\text{Score:} & \quad S_{AB} = \int (\sigma_A\tau_B + \tau_A\sigma_B - Q\tau_A\tau_B)dV, \quad \text{Penalty Factor:} \quad Q = 11 \\
\text{Orthogonality:} & \quad S_{AB} = \sum_{nlm} \left(a_{nlm}^\sigma b_{nlm}^\tau + a_{nlm}^\tau \left(b_{nlm}^\sigma - Qb_{nlm}^\tau\right)\right) \\
\text{Blind Search:} & \quad 6D \text{ space } = 1 \text{ distance } + 5 \text{ Euler rotations: } (R, \beta_A, \gamma_A, \alpha_B, \beta_B, \gamma_B)
\end{align*}
Modeling Symmetric $C_n$ Complexes

Let $y$ be principal symmetry axis, $\omega$ be symmetry angle

Describe the system using a “Docking Equation”

\[
\hat{R}_y(\omega_{j+1}) \hat{T}_z(D) \hat{R}(\alpha, \beta, \gamma) B(r) \leftrightarrow \hat{R}_y(\omega_j) \hat{T}_z(D) \hat{R}(\alpha, \beta, \gamma) A(r)
\]

- Search over 3 rotations ($\hat{R}(\alpha, \beta, \gamma)$) and 1 translation ($\hat{T}_z(D)$)
- Accelerate search using polar Fourier FFT (details not shown)
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):

$P_{ki}(r) = A_{ki}(r) + B_{ki}(r) + C_{ki}(r)$

(just do by brute-force search in $E$ and $\eta$, no FFT here)
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
Docking Model Structures in CASP/CAPRI Round 30

- CASP – community-wide protein modeling experiment (start from sequence)
- CAPRI – similar experiment for protein docking (start from monomers)
- Round 30 challenge: make $C_2$ and $D_2$ symmetric complexes from CASP models

Example: 150 models for T85
Centre: Zhang_Server_TS4
## The CASP 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</th>
<th>Top-10</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>T79</td>
<td>DIFFICULT</td>
<td>FFAS03_Ts1</td>
<td>3RCo</td>
<td>1 / 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T80</td>
<td>EASY</td>
<td>FALCON_MANUAL_X_Ts2</td>
<td>2OGa 3NYu</td>
<td>1–2 / 3–10</td>
<td>M01 / KBDOCK</td>
<td>M07 / SAM</td>
<td>MEDIUM ACCEPTABLE</td>
</tr>
<tr>
<td>T81</td>
<td>HETERO</td>
<td>STRINGS_Ts1</td>
<td>2G0N</td>
<td>1–2 / 3–10</td>
<td>(*)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T82</td>
<td>EASY</td>
<td>FALCON_EnvFold_Ts4</td>
<td>3E08</td>
<td>1 / 2–10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T83</td>
<td>CANCELED</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T84</td>
<td>EASY</td>
<td>PhyreX_Ts1</td>
<td>4BI5</td>
<td>1 / 2–10</td>
<td>M01 / KBDOCK</td>
<td>M02 / SAM</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>T85</td>
<td>EASY</td>
<td>Zhang-Server_Ts4</td>
<td>2F1K 3GGP</td>
<td>1–2 / 3–10</td>
<td>M07 / SAM</td>
<td>M09 / SAM</td>
<td>ACCEPTABLE</td>
</tr>
<tr>
<td>T86</td>
<td>DIFFICULT</td>
<td>FALCON_EnvFold_Ts5</td>
<td>43HU 3I0Y 3ER7</td>
<td>1–3 / 4–10</td>
<td>M07 / SAM</td>
<td>M09 / SAM</td>
<td>ACCEPTABLE</td>
</tr>
<tr>
<td>T87</td>
<td>EASY</td>
<td>FALCON_EnvFold_Ts1</td>
<td>200R</td>
<td>0 / 1–10</td>
<td>M01 / SAM</td>
<td>M02 / SAM</td>
<td>ACCEPTABLE</td>
</tr>
<tr>
<td>T88</td>
<td>DIFFICULT</td>
<td>PhyreX_Ts1</td>
<td>2X1U</td>
<td>1 / 2–10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T89</td>
<td>HETERO</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T90</td>
<td>EASY</td>
<td>–</td>
<td>–</td>
<td>–</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...

El Houasli et al. (2016), Proteins, 84, in press
Conclusions

SAM

- New docking code for arbitrary point group symmetry
- Every solution is perfectly symmetrical
- But some solutions can contain steric clashes
- So should refine / energy-minimise using molecular mechanics

Current Work

- We are developing an automated docking pipeline on our server
  - SAM: symmetric complexes
  - Hex: non-symmetric complexes
  - NAMD: flexible energy minimisation and refinement
Thank You!

http://sam.loria.fr/
http://hex.loria.fr/
http://kbdock.loria.fr/
Extra Slides
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
Our Best Solution With KBDOCK – T80

- T80–M01: $F_{\text{nat}} = 0.493$, LRMSD = 3.77 Å, IRMSD = 1.97 Å
Our Best Solution With SAM – T86

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

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

No doubt, $\text{RMSD(Receptor+Ligand)}$ 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!