Protein Docking – Predicting Protein-Protein Interactions at The Molecular Level

Dave Ritchie
Department of Computing Science
University of Aberdeen

Recent Growth of Protein-Protein Interaction (PPI) Literature

Citations of key yeast functional genomics papers (per year):
- Red: Ito et al., Uetz et al. (Y2H)
- Blue: Ho et al., Gavin et al. (TAP-MS)
- Black: All protein-protein interaction papers

High-Throughput Determination of PPIs

Experimental
- Yeast-2-hybrid (Y2H)
- Tandem-Affinity-Purification Mass Spectrometry (TAP-MS)

Bionformatics
- Gene Fusion Analysis (Rosetta Stone)
- Phylogenetic Profiling & Ortholog Transfer
- Evolutionary Trace & Correlated Mutations
- Threading
- Literature Data Mining
- All approaches have high false-positive rates

Example TAP-MS PPI Network - Yeast DNA Damage Response

- Blue: known previously; red: new; arrows: bait → target

Figure from: Ho et al. Nature (2002) 415 180–183
How Complete Are PPI Networks?

• Every protein has an average of \( \sim 9 \) interaction partners

Yeast

\( \sim 6,000 \) proteins

\( \sim 38-75,000 \) interactions

\( \sim 50\% \) complete

Human

\( \sim 30,000 \) proteins

\( \sim 154-370,000 \) interactions

\( \sim 10\% \) complete


GT Hart, AK Ramani, EM Marcotte, Genome Biology (2006) 7 120

Docking - Predicting PPIs at the 3D Molecular Level

Ab Initio

• Soft Docking – FFT, Polar Fourier Correlations (\( \sim \) hours)
• MC/MD – Flexible side chains + backbones (\( \sim \) days)

Re-Scoring

• Knowledge-based potentials

Data-Driven

• Biochemical: mutagenesis hot-spot residues
• Biophysical: NMR CSP/RDC, H/D exchange, \(^{13}\)C labeling, ...
• ET + Correlated mutations
• Structural Databases (docking by homology)

Example Predicted PPI Network - Yeast Prion Sup35

• Yellow: synthesis & folding
• Red: tRNA/mRNA synthesis/splicing
• Blue: protein targeting
• Clear: other
• Dash: experimentally determined
• Thin: from phylogenetic profiles
• Thick: from multiple sources

Figure from: DEisenberg, EM Marcotte, I Xenarios, TO Yeates, Nature (2000) 405 823–826

The Basic Goal of Protein-Protein Docking

Find minimum potential energy of the system as rapidly as possible:

\[ E = \int \phi(r) \rho(r) dV \]

For two proteins

\[ \phi(r) = \phi_A(r) + \phi_B(r) \]
\[ \rho(r) = \rho_A(r) + \rho_B(r) \]

and so

\[ E = \int (\phi_A(r) \rho_B(r) + \phi_B(r) \rho_A(r)) dV \]

• With brute-force search, typically need \( \sim 10^9 \) such integrals
• Current algorithms often sum several such potential/density terms...
• ... and often use 3D Cartesian FFTs to accelerate the calculation
Spherical Polar Fourier (SPF) Expansions

Real functions:
\[ \phi_A(r) = \sum_{n,l,m} a_{n,l,m} R_{n,l}(r) y_{l,m}(\theta, \phi) \]

Complex functions:
\[ \phi_A(r) = \sum_{n,l,m} A_{n,l,m} R_{n,l}(r) Y_{l,m}(\theta, \phi) \]

Unitary transforms:
\[ y_{l,m}(\theta, \phi) = \sum_l U(l)_{m,t} Y_{l,t}(\theta, \phi); \quad A_{n,l,m} = \sum_l U(l)_{im} a_{n,l,t} \]

Gaussian-type (GTO):
\[ R_{n,l}(r) = N_{n,l} e^{-\rho^2/2} L_{n-l-1}^{(l+1/2)}(\rho); \quad \rho = r^q/q, \quad q = 20. \]

Exponential-type (ETO):
\[ R_{n,l}(r) = N_{n,l} e^{-\rho^2/2} L_{n-l-1}^{(2l+2)}(\rho); \quad \rho = 2\Lambda r, \quad \Lambda = 1/2. \]

3D Shape Density Reconstruction – CAPRI T21: Orc1/Sir1

Docking Using 3D Polar Fourier Density Functions - “Hex”

Densities:
\[ \sigma(r) = \sum_{n,l,m} a_{n,l,m} R_{n,l}(r) y_{l,m}(\theta, \phi) \quad \tau(r) = \sum_{n,l,m} a_{n,l,m}^* R_{n,l}(r) y_{l,m}(\theta, \phi) \]

Favourable:
\[ \int (\sigma_A(r_A)\tau_B(r_B) + \tau_A(r_A)\sigma_B(r_B))dV \]

Unfavourable:
\[ \int \tau_A(r_A)\tau_B(r_B)dV \]

Score:
\[ S_{AB} = \int (\sigma_A\tau_B + \tau_A\sigma_B - Q\tau_A\tau_B)dV \quad \text{Penalty Factor: } Q = 11 \]
Correlations - Overlap as a Function of Coordinate Operations

Rotation:

\[  \hat{R}(\alpha, \beta, \gamma) \sigma_A(r) = \sum_{n,l,m} a'_{nlm} R_{nl}(r) y_{lm}(\theta, \phi) \]

Rotated Coefficients:

\[ a''_{nlm} = \sum_{m'=-l}^{l} R_{nm'}^{(l)}(\alpha, \beta, \gamma) a'_{nlm} \]

Translation:

\[ \hat{T}(R) \sigma_A(r) = \sum_{n,l,m} a''_{nlm} R_{nl}(r) y_{lm}(\theta, \phi) \]

Translated Coefficients:

\[ a''_{nlm} = \sum_{n'lm'} T_{n'lm'}^{(l)}(R) a''_{nlm} \]

Hence:

\[ \int \sigma_A'(r) r''_{B}(r) dV = \sum_{n,l,m} a''_{nlm} b''_{nlm} \text{ etc.} \]

Search Space:

\(~ 10^9\) orientations (~ 10^6 orientations/sec)


Multi-Sample Docking for Very Large Molecules - Antibody-VP2

Show Docking Movie!

CAPRI – Critical Assessment of Predicted Interactions

- Started in 2001/2 following CASP with 19 groups & 7 targets...
- At least one protein presented in its unbound form
- Any predictive approach allowed: homology/literature, etc.

<table>
<thead>
<tr>
<th>Target</th>
<th>Receptor</th>
<th>Ligand</th>
<th>Type</th>
<th>Complex</th>
<th>Lab</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HPPr</td>
<td>HPPr</td>
<td>U/U</td>
<td>Fieulaine et al.</td>
<td>Janin</td>
</tr>
<tr>
<td>2</td>
<td>Rotavirus VP6</td>
<td>MCV</td>
<td>U/B</td>
<td>Vaney et al.</td>
<td>Rey</td>
</tr>
<tr>
<td>3</td>
<td>Hemagglutinin</td>
<td>HC63</td>
<td>U/B</td>
<td>Barbey-Martin al.</td>
<td>Knossow</td>
</tr>
<tr>
<td>4</td>
<td>(\alpha)-Amylase</td>
<td>AMD10</td>
<td>U/B</td>
<td>Desmyter et al.</td>
<td>Cambillau</td>
</tr>
<tr>
<td>5</td>
<td>(\alpha)-Amylase</td>
<td>AMB7</td>
<td>U/B</td>
<td>Desmyter et al.</td>
<td>Cambillau</td>
</tr>
<tr>
<td>6</td>
<td>(\alpha)-Amylase</td>
<td>AMD9</td>
<td>U/B</td>
<td>Desmyter et al.</td>
<td>Cambillau</td>
</tr>
<tr>
<td>7</td>
<td>SpeA</td>
<td>TCR 14.3.D</td>
<td>U/U</td>
<td>Sundberg et al.</td>
<td>Mariuzza</td>
</tr>
</tbody>
</table>

- Now > 40 groups; Currently on Targets 28 ...
- 3 Sections - Predictors, Servers, Scorers

**CAPRI Target 1 - Lactobacillus HPr / HprK**

**Docked Orientation (Hex) for Target 3 - Hemagglutinin/HC63**

- **CAPRI “medium accuracy”** (1 Å ≤ Ligand RMSD ≤ 5 Å)

**CAPRI Results: Targets 1–7**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Software</th>
<th>Algorithm</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abagyan</td>
<td>ICM</td>
<td>FF</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>**</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Camacho</td>
<td>CHARMM</td>
<td>FF</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eisenstein</td>
<td>MolFit</td>
<td>FFT</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sternberg</td>
<td>FTDOCK</td>
<td>FFT</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ten Eyck</td>
<td>DOT</td>
<td>FFT</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gray</td>
<td>MC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ritchie</td>
<td>Hex</td>
<td>SPF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weng</td>
<td>ZDOCK</td>
<td>FFT</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wolfson</td>
<td>BUDDA/PPD</td>
<td>GH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bates</td>
<td>Guided Docking</td>
<td>FF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palma</td>
<td>BIGGER</td>
<td>GF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gardiner</td>
<td>GAPDOCK</td>
<td>GA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Olson</td>
<td>SurfDock</td>
<td>SH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valencia</td>
<td>ANN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vakser</td>
<td>GRAMM</td>
<td>FFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* low, ** medium, *** high accuracy prediction; – no prediction


**Docked Orientation (Hex) for Target 6 - Amylase/AMD9**

- **CAPRI “high accuracy”** (Ligand RMSD ≤ 1 Å)
Subsequent CAPRI Targets (Rounds 3 – 5)

<table>
<thead>
<tr>
<th>Target</th>
<th>Description</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>T8</td>
<td>Nidogen-γ3-Laminin</td>
<td>U/U</td>
</tr>
<tr>
<td>T9</td>
<td>LiCT homodimer</td>
<td>build from monomer – 12Å RMS deviation</td>
</tr>
<tr>
<td>T10</td>
<td>TBEV trimer</td>
<td>build from monomer – 11Å RMS deviation</td>
</tr>
<tr>
<td>T11</td>
<td>Cohesin - dockerin</td>
<td>U/U; model-build dockerin</td>
</tr>
<tr>
<td>T12</td>
<td>Cohesin - dockerin</td>
<td>U/B</td>
</tr>
<tr>
<td>T13</td>
<td>SAG1 - antibody Fab</td>
<td>SAG1 conformational change: 10Å RMS</td>
</tr>
<tr>
<td>T14</td>
<td>MYPT1 - PP1 δ</td>
<td>U/U; model-build PP1 α → PP1 δ</td>
</tr>
<tr>
<td>T18</td>
<td>TAXI - xylanase</td>
<td>U/B</td>
</tr>
<tr>
<td>T19</td>
<td>Ovine prion - antibody Fab</td>
<td>model-build prion</td>
</tr>
</tbody>
</table>

- T15-T17 cancelled due to structures released prematurely - Google!
- T11, T14, T19 involved homology model-building step...

Docked Orientation (Hex) for Target 12 - Cohesin/Dockerin

- Here, we assumed “molecular mimicry”
- First superposed dockerin onto cohesin dimer, then docked...

- CAPRI “high accuracy” (Interface RMSD ≤ 1Å)

Recent Progress and Future Directions in Protein Docking

Recent Progress

- Increasing use of biophysical information - data-driven docking
- Better predictions of “hot-spot” interaction residues
- Increasing use of statistical potentials from known interfaces
- Increasing availability of interaction databases (e.g. STRING, SCOPPI, PIBASE, ...)
- On-line docking servers (unsupervised docking) have success in CAPRI
- Recent successes in multimeric assembly

Future Directions

- Closer integration of diverse bioinformatics and biophysical information
- Special hardware - GPUs and FPGAs look increasingly promising
- 5D FFT docking coming soon!


Using Low Resolution Docking to Cross-Validate Predicted PPIs?

Low resolution docking of Tripsin + BPTI

- Crystal structure + low res FFT
  - Gold: BPTI location in crystal
  - Red: centroid of calculated BPTI solutions

- Model-built structure (green) + low res FFT
  - RMS = 6 Å

Figure from: Tovchigrechko et al., Prot. Sci. (2002) 11 1888–1896

Acknowledgments

BBSRC 1996–2000

- Hex
  - http://www.csd.abdn.ac.uk/hex/

- HexServer
  - http://www.csd.abdn.ac.uk/hex_server/

- Preprints
  - http://www.csd.abdn.ac.uk/~dritchie/

- For a Recent Review
  - email: dritchie@csd.abdn.ac.uk

- PSFB Special Issue: Third CAPRI Evaluation Meeting December 2007
  - Google: Proteins Wiley

Hex
  - http://www.csd.abdn.ac.uk/hex/

HexServer
  - http://www.csd.abdn.ac.uk/hex_server/

Preprints
  - http://www.csd.abdn.ac.uk/~dritchie/

For a Recent Review
  - email: dritchie@csd.abdn.ac.uk

PSFB Special Issue: Third CAPRI Evaluation Meeting December 2007
  - Google: Proteins Wiley