Molecular Shape Recognition Using 3D Polar Fourier Correlations

Dave Ritchie
Department of Computing Science
King’s College, University of Aberdeen
Molecular Shape Recognition Using 3D Polar Fourier Correlations

Contents

- Spherical Polar Fourier Expansions
- Spherical Polar Correlations
- Protein Docking
- CAPRI: International Blind Docking Trial
- Small Molecule 3D Shape Matching Examples
- Future Prospects & Conclusions
Real Spherical Harmonics: \( y_{lm}(\theta, \phi) \)

**Orthogonality:**
\[
\int y_{lm}(\theta, \phi) y_{l'm'}(\theta, \phi) d\Omega = \delta_{ll'} \delta_{mm'}
\]

**Rotation:**
\[
y_{lm}(\theta', \phi') = \sum_{m'=-l}^{l} R_{m'm}(\alpha, \beta, \gamma) y_{lm'}(\theta, \phi)
\]
Spherical Harmonic Surfaces

Example: 2D Radial Expansions (256 Basis Functions)

$$r(\theta, \phi) = \sum_{l=0}^{15} \sum_{m=-l}^{l} a_{lm} y_{lm}(\theta, \phi)$$

- Good for matching similar shapes, not so good for docking (complementary shapes) ...
Radial Basis Functions: $R_{nl}(r)$

**HO (shape):**

$$R_{nl}(r) = N_{nl}^{(q)} e^{-\rho/2} \rho^{l/2} L_{n-l-1}^{(l+1/2)}(\rho);$$

$$\rho = r^2/q, \quad q = 20.$$

**Coulomb (electro):**

$$R_{nl}(r) = N_{nl}^{(\Lambda)} e^{-\rho/2} \rho^l L_{n-l-1}^{(2l+2)}(\rho);$$

$$\rho = 2\Lambda r, \quad \Lambda = 1/2.$$

**Orthogonality:**

$$\int_0^\infty R_{nl}(r) R_{n'l}(r) r^2 dr = \delta_{nn'}$$
Spherical Polar Shape Reconstruction

(Looking at MCV Antibody CDR Loops)

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

<table>
<thead>
<tr>
<th>Image</th>
<th>Expansion</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 Density Representations

Surface Density: $\sigma(r) = \begin{cases} 1; & r \in \text{surface skin volume} \\ 0; & \text{otherwise} \end{cases}$

Interior Density: $\tau(r) = \begin{cases} 1; & r \in \text{VDW volume} \\ 0; & \text{otherwise} \end{cases}$

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: $S_{AB} = \int (\sigma_A\tau_B + \tau_A\sigma_B - Q\tau_A\tau_B)dV$; Penalty Factor: $Q = 11$
Correlations – Overlap as a Function of Coordinate Operations

Basic Overlap:
\[ \int \sigma_A(\mathbf{r}) \tau_B(\mathbf{r}) \, dV = \sum_{nlm}^N a_{nlm}^\sigma b_{nlm}^\tau \]

Rotation:
\[ \hat{R}(\alpha, \beta, \gamma) \sigma_A(\mathbf{r}) = \sum_{nlm}^N a_{nlm}^{\sigma'} R_{nl}(r) y_{lm}(\theta, \phi) \]

Translation:
\[ \hat{T}(R) \sigma_A(\mathbf{r}) = \sum_{nlm}^N a_{nlm}^{\sigma''} R_{nl}(r) y_{lm}(\theta, \phi) \]

Rotated Coefficients:
\[ a_{nlm}^{\sigma'} = \sum_{m'=-l}^l R_{mml'}^{(l)}(\alpha, \beta, \gamma) a_{nlm'}^{\sigma} \]

Translated Coefficients:
\[ a_{nlm}^{\sigma''} = \sum_{n'l'}^N T_{nl,n'l'}^{(m'l)}(R) a_{n'l'm}^{\sigma} \]

Hence:
\[ \int \sigma_A'(\mathbf{r}) \tau''_B(\mathbf{r}) \, dV = \sum_{nlm}^N a_{nlm}^{\sigma'} b_{nlm}^{\tau''} \text{ etc.} \]
6D Docking Search as a Nested Sequence of Transformations

Get 4 rotations from icosahedral tessellations...

Rotate A \((\times 492 @ 9.9^\circ)\): 
\[ A'(\mathbf{r}) = \hat{R}(0, \beta_1, \gamma_1)A(\mathbf{r}) \]

Translate A \((\times 50 @ 0.75\text{\AA})\): 
\[ A''(\mathbf{r}) = \hat{T}(-R)A'(\mathbf{r}) \]

Rotate B \((\times 642 @ 8.5^\circ)\): 
\[ B'(\mathbf{r}) = \hat{R}(0, \beta_2, \gamma_2)B(\mathbf{r}) \]

Twist B \((\times 64 @ 5.6^\circ)\): 
\[ B''(\mathbf{r}) = \hat{R}(\alpha_2, 0, 0)B'(\mathbf{r}) \]

Search Space: 
\[ 492 \times 50 \times 642 \times 64 \simeq 10^9 \sim 10^6/s \text{ on a 1GHz PIII Xeon} \]
Summary of Hex Docking Strategy

- Contour VDW & SAS surface onto \((0.6\text{Å})^3\) grid
- Calculate shape & electrostatic coefficients – once
- Scan search space with \(N=16\) correlations (shape only)
- (Optionally constrain search to one or both epitopes)
- Re-score top 10,000 with \(N=30\) (+electrostatics)
- Refine top 500 by soft OPLS minimisation & cluster solutions
- Output best member of each cluster...
CAPRI – Critical Assessment of Predicted Interactions

- Following from CASP: Janin, Wodak, et al.
- During 2001/2: seven target complexes made available...

<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>HPr Kinase</td>
<td>HPr</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 et 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>

- At least one docking partner presented in its unbound form
- Participants permitted 5 attempts for each target
- Any predictive approach allowed: homology/literature, etc.
- 19 international groups attempted predictions...
# CAPRI Results: Targets 1–7
*(Mendez et al. (2003) PSFG 52 51–67)*

<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>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>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>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>Gray</td>
<td>MC</td>
<td></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>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>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>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>Vakser</td>
<td>GRAMM</td>
<td>FFT</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>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>
</tbody>
</table>

+ medium or high accuracy prediction; – low accuracy prediction
Best Hex Orientation for Target 3
Best Hex Orientation for Target 6
3D Molecular Field Similarity Scores

- Define “distance” between two 3D functions:

\[ \int |\rho_A - \rho_B|^2 dV = \int \rho_A^2 dV + \int \rho_B^2 dV - 2 \int \rho_A \rho_B dV \]

- A related similarity index (Carbo) is given by:

\[ C_{AB} = \frac{\int \rho_A \rho_B dV}{\left[ \int \rho_A^2 dV \int \rho_B^2 dV \right]^{1/2}} \]

- So, we wish to evaluate 6D correlations of the form:

\[ S_{AB} = \int [\hat{T}_z(-R)\hat{R}(0, \beta_A, \gamma_A)\rho_A(\vec{r})][\hat{R}(\alpha_B, \beta_B, \gamma_B)\rho_B(\vec{r})]dV \]

- With \( N = 8 \) (204 coeffs), get \( 5 \times 10^6/\text{sec} \) (1.8GHz PIII Xeon)
3D Shape Matching Example: Benzodiazepines

A: CGS-8216
B: Methyl-β-carboline-3-carboxylate
C: Ro15-1788

A–B: 0.95
B–C: 0.87
A–C: 0.85
3D Partial Shape Matching Example: Leu-Enkephalin/Morphine

- Top (N-TYR): 0.57; Bottom (C-LEU): 0.60
- $\sim 10^8$ trial orientations: 25 seconds on 1.8GHz PIII Xeon
Enkephalin/Morphine Shape Functions Used

- N=8: 204 coefficients
- N=16: 1,496 coefficients
- Scan search space at N=8; Re-score top 2,000 at N=16...
Practical Issues & Future Improvements

- Improve search method (aiming for $<< 1$s per comparison)...
- Scale radial functions for small ligands...
- Correlate other surface properties...
- Pre-calculate & store shape/property vectors...
- Handle multiple conformations...
- Dock database hits into binding sites...
- Better data management & user interface etc...
Conclusions

- **Protein Docking ("Hex"):**
  - Novel, very fast, & (fairly) accurate docking algorithm
  - Hex performed well in CAPRI: 2 out of 7 ain’t bad!

- **Small-Molecule Applications:**
  - Fast 3D shape/field matching...
  - 3D database search for drug-like ligands now feasible...
  - Extensible to pharmacophore searching & ligand docking...

Software: http://www.biochem.abdn.ac.uk/hex/
Preprints: http://www.csd.abdn.ac.uk/~dritchie/