Swarming Binding Funnels: Ranking Native over False Docking Poses

Motivation I: Docking Mechanisms

Crowding

TP docking funnel
Mutation space more robust?

FP docking Funnel
Mutational space less robust?
Motivation II: Model Refinement (MD)

Erik Pfeiffenberger’s poster

Crystal Structure
Starting Model
Refined Model

T39: Cluster 44 --> Cluster 9
1. Find the correct funnel?
2. **Confidence** in finding TP funnel in the top 10?
3. Get to the bottom of the correct funnel?
4. Do the proteins dock at all?
### Score_set: Decoy Statistics

(Lensink & Wodak Proteins(2014))

<table>
<thead>
<tr>
<th>Target</th>
<th>Total</th>
<th>High</th>
<th>Medium</th>
<th>Acceptable</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2 9</td>
<td>2083</td>
<td>2</td>
<td>78</td>
<td>87</td>
<td>1916</td>
</tr>
<tr>
<td>T3 0</td>
<td>1343</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1341</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>599</td>
<td>0</td>
<td>3</td>
<td>12</td>
<td>584</td>
</tr>
<tr>
<td></td>
<td>499</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>496</td>
</tr>
<tr>
<td></td>
<td>309</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>309</td>
</tr>
<tr>
<td></td>
<td>1500</td>
<td>11</td>
<td>46</td>
<td>42</td>
<td>1401</td>
</tr>
<tr>
<td></td>
<td>899</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>899</td>
</tr>
<tr>
<td></td>
<td>1400</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1396</td>
</tr>
<tr>
<td></td>
<td>2180</td>
<td>193</td>
<td>206</td>
<td>189</td>
<td>1592</td>
</tr>
<tr>
<td></td>
<td>2083</td>
<td>2</td>
<td>78</td>
<td>87</td>
<td>1916</td>
</tr>
</tbody>
</table>
Docking Poses Scored with DCOMPLEX (Liu et al. Proteins (2004))

T47 (easy)

T40 (easy)

T32 (hard)

T39 (hard)
Methodology

Score_Set dataset (Lensink & Wodak; Proteins 2014)

Removing models with clashes

Feature calculation and normalization

Cluster (10 Angstrom)

Aggregation of feature values for each cluster

Construction of pairwise cluster comparision matrix for classification learning
Number of Models per Cluster

Cluster

Number of Models

250 PSO Particles
10 Angstrom Clusters and PSO Ensembles

T39 Cluster 1: 95

T39 Cluster 1: PSO 250 Particles

T39 Cluster 44: 5

T39 Cluster 1: PSO 250 Particles
SwarmDock

- **Swarm / Flock / Colony**
  - Group of Individuals

- **Individual**
  - Follows Simple Rules

- **Collective Behaviour**
  - Solve Complex Problems
Particle Swarm Optimisation

\[ \chi_i = \begin{bmatrix} x \\ y \\ z \\ \theta_x \\ \theta_y \\ \theta_z \\ M_1 \\ M_2 \\ \vdots \\ M_m \end{bmatrix} \]

\[ V_i = \begin{bmatrix} \delta x \\ \delta y \\ \delta z \\ \delta \theta_x \\ \delta \theta_y \\ \delta \theta_z \\ \delta M_1 \\ \delta M_2 \\ \vdots \\ \delta M_m \end{bmatrix} \]

A Particle Position

A Particle Velocity

Repeat Until Convergence

Initialise Swarm

Calculate scores

Update Velocities

Update Positions

Return Best Structure
PSO Simulation for T32 Cluster 6
Compute Molecular Descriptors for each member of each cluster

Molecular descriptors of all models for a cluster are represented as distributions

Extremely Randomized Trees

110 molecular descriptors

Clustering

Cluster A

Cluster B

Comparison

Min, Q1, Mean, Median, Q3, Max

Cluster A or B Better?

1100 Features to compare two clusters
Cluster Comparison Score

T29

Number of times better

Min LRMSD (nm)

Predicted

Actual
Cluster Comparison Profiles for 11 Targets
1100 Features

Ranked of Features

Tobi Potential: Residue Level (#1)
(Tobi & Bahar Proteins (2006))

ZRANK (#96)
Pierce & Weng: Proteins (2007)

DCOMPLEX (#241)
Liu et al Proteins (2004))

Top 10

- C1_MIN_N_CP_TB
- C2_Q1N_CP_D1
- C2_MIN_N_CP_TB
- C1_Q1_N_CP_TB
- C2_MIN_N_CP_D1
- C2_AVG_N_CP_D1
- C1_MIN_N_CP_D1
- C1_AVG_N_ODA
- C2_Q1_N_CP_TB
- C1_MAX_N_CP_PIE
Pearson Product-Moment Correlation Coefficient
(Many features calculated using the CCharPPI server; Moal et al Bioinformatics (2015))

Molecular Descriptor

Positively correlated
Not correlated
Negatively correlated
Transforming the Feature Space (PCA)

- Accuracy
- F1
- Precision
- Recall

Minimised False negative rate

Percentage Dimensions
Final Ranking of Clusters with at Least One Acceptable Docking Pose

<table>
<thead>
<tr>
<th>Target</th>
<th>Pred (ML)</th>
<th>ZRANK</th>
<th>DCOMPLEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>T29</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>T30</td>
<td>8</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>T32</td>
<td>10</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>T35</td>
<td>9</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>T39</td>
<td>9</td>
<td>3</td>
<td>29</td>
</tr>
<tr>
<td>T40</td>
<td>8</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>T41</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>T46</td>
<td>5</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>T47</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>T53</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>T54</td>
<td>5</td>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>

**Graph:** Percentage of Targets

- **1st**
- **Top 5**
- **Top 10**