Exploring Multi-conformational Modeling and Flexibility of Molecular Recognition Features In Improving Drug Docking

Andrew Li
Mentor: Dr. Gil Alterovitz

May 17, 2015
Intrinsically Disordered Proteins

- Proteins are biological structures, made of chains of amino acids.
Intrinsically Disordered Proteins

- Proteins are biological structures, made of chains of amino acids.
- (Ordered) proteins generally have four levels of structure.
Intrinsically Disordered Proteins

- Proteins are biological structures, made of chains of amino acids.
- (Ordered) proteins generally have four levels of structure.
- A *intrinsically disordered protein* (IDP) is a protein containing regions of disorder.
**Intrinsically Disordered Proteins**

- Proteins are biological structures, made of chains of amino acids.
- (Ordered) proteins generally have four levels of structure.
- A *intrinsically disordered protein* (IDP) is a protein containing regions of disorder.
- They lack a fixed tertiary, or 3-D structure.
**INTRINSICALLY DISORDERED PROTEINS**

- Proteins are biological structures, made of chains of amino acids.
- (Ordered) proteins generally have four levels of structure.
- A *intrinsically disordered protein* (IDP) is a protein containing regions of disorder.
- They lack a fixed tertiary, or 3-D structure.
- IDPs are potential drug targets and are now closely studied.
The protein disorder continuum

Disordered  Structured
Molecular Recognition Features (MoRFs)

- MoRFs are small, interaction-prone segments of disorder within larger proteins.
**Molecular Recognition Features (MoRFs)**

- MoRFs are small, interaction-prone segments of disorder within larger proteins.

- Their presence indicates the ability for recognition and binding.
**MOLECULAR RECOGNITION FEATURES (MoRFs)**

- MoRFs are small, interaction-prone segments of disorder within larger proteins.
- Their presence indicates the ability for recognition and binding.
- They are usually defined to be between 10-70 residues long.
**Motivating Questions**

- How can we utilize the flexible nature of IDPs in improving docking ability?
Motivating Questions

- How can we utilize the flexible nature of IDPs in improving docking ability?
- What are different paradigms within which we can analyze binding affinities of flexible regions?
Motivating Questions

▶ How can we utilize the flexible nature of IDPs in improving docking ability?

▶ What are different paradigms within which we can analyze binding affinities of flexible regions?

▶ How can these results be applied to finding new drugs for diseases such as cancer?
**DATA COLLECTION AND PROCESSING**

- A pipeline was written to fully automate the process of drug-protein matching.
**Data Collection and Processing**

- A pipeline was written to fully automate the process of drug-protein matching.
- From the Protein Data Bank, proteins were gathered related to major pathogens.
**Data Collection and Processing**

- A pipeline was written to fully automate the process of drug-protein matching.
- From the Protein Data Bank, proteins were gathered related to major pathogens.
- The MoRF segments were isolated from the PDB files, and ran through the pipeline to find drugs that might bind with the MoRFs.
**SIMULATION OF FLEXIBILITY**

- TraDES was used to generate 200 conformations of each protein analyzed.
SIMULATION OF FLEXIBILITY

- TraDES was used to generate 200 conformations of each protein analyzed.
- These proteins were re-screened through the pipeline.
SIMULATION OF FLEXIBILITY

- TraDES was used to generate 200 conformations of each protein analyzed.
- These proteins were re-screened through the pipeline.
- (two pictures of different conformations side by side)
Drug Results

Based on the process used, six drugs have been found to address *Pseudomas Aeruginosa* (which affects airways and can cause blood infections)

<table>
<thead>
<tr>
<th>PubchemID</th>
<th>Prob-SSB</th>
</tr>
</thead>
<tbody>
<tr>
<td>46507215</td>
<td>0.926</td>
</tr>
<tr>
<td>46506020</td>
<td>0.929</td>
</tr>
<tr>
<td>46508185</td>
<td>0.926</td>
</tr>
<tr>
<td>45406770</td>
<td>0.926</td>
</tr>
<tr>
<td>45406528</td>
<td>0.926</td>
</tr>
<tr>
<td>46507414</td>
<td>0.928</td>
</tr>
</tbody>
</table>
COMPARISON OF METHODS

▶ Using a matched pairs test between all MoRFs analyzed and the top score from their conformations, I obtain a p-value of 0.02.
COMPARISON OF METHODS

- Using a matched pairs test between all MoRFs analyzed and the top score from their conformations, I obtain a p-value of 0.02.

- At the $\alpha = 0.05$ level, this is significant, and shows an improvement in docking score.
ALTERNATIVE METHOD OF FLEXIBLE DOCKING

- A program was written in order to dock multiple pieces of the MoRF with the drug individually.
ALTERNATIVE METHOD OF FLEXIBLE DOCKING

- A program was written in order to dock multiple pieces of the MoRF with the drug individually.
- This is possible because of the difference in size between the MoRF (or protein) and the drug.
**Analysis of Method Runtime**

- All bonds which can rotate are kept rotatable, and if the sections are divided correctly only one will bind to the drug.
- Further work must be done in automating this process.
CONCLUSION

The property of flexibility for MoRFs was utilized to improve docking score by generating a large number of conformations, and binding them with the appropriate drugs. Additionally, a new method of docking with flexible proteins was developed to reduce docking runtime significantly.
ACKNOWLEDGEMENTS

I would like to thank the following people for their essential role in allowing this project to succeed:

- Dr. Gil Alterovitz
- Anvita Gupta
- MIT PRIMES
- Parents!