Comparative Modeling and Docking

Andrew Kohlway
Dr. Jeffry D. Madura
Outline

- Background Information – GPCRs/Melatonin Receptors
- Comparative Modeling Techniques
- Docking
- Data
- Conclusion
G-Protein Coupled Receptors (GPCRs)

- Seven transmembrane helical protein
- The binding of a ligand will alter the conformation of the receptor and the heterotrimeric G-protein coupled to it.
- A GTP will replace a GDP attached to the alpha subunit of the G-protein, resulting in a change in a nearby effector molecule, which in turn sends out second messengers such as cAMP throughout the cell.

http://www.emory.edu/CHEMISTRY/justice/seminar/receptor1.htm
Problem and Solution

- The only complete structure to be discovered so far has been bovine rhodopsin (SCIENCE 2002).
- The structures of other GPCRs are very important because most drugs will bind to them.
- So far a technique called “comparative modeling” has been used to produce acceptable structures for proteins that have been sequenced.
Melatonin Receptors

- Attempts have been made to model the melatonin receptors to help further studies into drug binding interactions.
- These receptors have been linked to depression and also to the sleep-cycle.
- There are two specific motifs of interest that may be involved with drug binding and ultimately the conformational changes needed for G-protein interactions.
- These motifs have been conserved in most rhodopsin-related GPCRs, however the melatonin receptors contain key mutations in these motifs that need to be analyzed.
Important Motifs

A structural motif or domain is a 3-dimensional fold which may be self-stabilizing or have an important biological function in the protein.

“Aromatic Cluster” Motif (FXXXWXXXF)
- Assumed to interact with the ligand through aromatic interactions.
- Ligand interaction minimizes a kink in TM 6 and disrupts the “arginine cage”, supposedly activating the receptor.

“NPXXY” Motif
- Provides stabilization of the inactive GPCR.
- Interaction between TM2 and TM7 occurs and also between residues in TM7 itself.

Figure 2. TM 2 and TM 7 interactions confirmed for rhodopsin (left) and assumed for MT1R (right).
Comparative Modeling

- Template Selection
- Alignment
- Model Building
- Evaluation of the Model
Template Selection

- Find as many protein structures as possible similar to your target sequence.
- Three different methods for identifying related structures.
  - Pairwise sequence comparison (BLAST and FASTA) between a target sequence and the database sequence – detects half of the relationships near 20-30% sequence identity.
  - Multiple sequence comparisons (PSI-BLAST and SAM) – detects three times more relationships than pairwise sequence comparisons. These comparisons compare a target sequence to a set of related sequences.
  - Threading methods (THREADER and 3D-PSSM) in which a sequence is compared using a “structure-dependent” scoring function. Best method to use when there are very few related sequences.

Modeller Script
Alignment

- Four different types of scoring methods:
  - Identity
  - Genetic Code
  - Chemical Similarity
  - Observed Substitutions

- PAM and BLOSUM scoring matrices

- Each pair of amino acids in the alignment receives a score based on observed frequencies from other alignments and also based on the frequency of occurrence of the amino acids.
  - Identities and frequently observed substitutions = positive score
  - Chance substitutions = negative score
Model Building

- First approach:
  - Homologous Modeling

- Second approach:
  - Choose a fold from a library
  - Score the compatibility between the sequence and fold
  - Align the sequence along the fold
  - Choose the best scoring fold

- Third approach:
  - Ab initio structure prediction
  - Generate a library of folds using Monte Carlo methods
  - Score folds
Model Building

- Modeller uses a method of satisfying spatial restraints (such as Cα-Cα distances, hydrogen bonds, dihedral angles) in order to create a model.

- MOE uses databases of backbone fragments and sidechains obtained from the Protein Data Bank to create many intermediate models which are then scored and averaged (or in some cases the best scoring model can be chosen).
Evaluation of the Model

- Programs like PROCHECK and 3D-Profiler are used to analyze final structures created through homology modeling.
- Criteria analyzed for correctness includes:
  - Dihedral Angles
  - Peptide Bonds
  - Side Chain Conformations
  - Hydrogen Bonding
  - Hydrophobic/Hydrophilic Residues
  - Atom Contacts
Ligand-Protein Docking

- Purpose – to find the optimal binding site between a ligand and in this case a protein.

- Matching Methods (Dock)
  - Active site is modeled using spheres.
  - Match ligands to the active site based on shape.

- Grid-Based Methods (Autodock)
  - A more flexible ligand is allowed to move around a specified docking box of a protein in order to find its optimal binding site.
    - Simulated Annealing
    - Genetic Algorithm
Lemarckian Genetic Algorithm (BioCache)

- The genetic algorithm is based on the biological principles of evolution and natural selection.
- Data is stored in chromosomes and segregated into genes that represent the orientation or phenotype of the protein/ligand.
- Using multiple operators, including crossover (recombination), mutation, and maturation operators, “parent” chromosomes will generate offspring that will gradually lead to an optimized “solution.”
- The Lamarckian aspect of this algorithm pertains to the local searches performed on the offspring to minimize the phenotype (structure), and then in essence turn the phenotype into the genotype of the new offspring.
Monte Carlo Docking (MOE)

- The algorithm in MOE creates small changes in the conformation, orientation, or displacement of the ligand and then scores the new structure based on empirical force field equations.
- If the move lowers the energy of the system, the new conformation is accepted, otherwise the Boltzmann factor ($p = e^{-\Delta E/kT}$) is computed.
- The new conformation is then accepted only if the probability is greater than a random number between zero and one.
- The environment of the ligand/receptor begins at a very high temperature allowing for large global searches of conformational space. As time progresses the temperature is lowered so that local searches can be made to optimize the final structure.
Methods

- Used NCBI Blast to find proteins similar to the Melatonin Receptors.

**Related Structures**

<table>
<thead>
<tr>
<th>Sequences producing significant alignments:</th>
<th>Score</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>10121075</td>
<td>pdb</td>
</tr>
<tr>
<td>gi</td>
<td>16975387</td>
<td>pdb</td>
</tr>
<tr>
<td>gi</td>
<td>14718455</td>
<td>pdb</td>
</tr>
<tr>
<td>gi</td>
<td>2914604</td>
<td>pdb</td>
</tr>
<tr>
<td>gi</td>
<td>21465824</td>
<td>pdb</td>
</tr>
</tbody>
</table>

- Picked Bovine Rhodopsin PDB 1F88.
Methods

- Used MOE (BLOSUM 62) and Modeller to generate two separate alignments.

**MOE Alignment**

**Modeller Alignment**
Methods

- Generated Comparative Models in MOE.
  - Used Modeller and MOE Alignments.
  - Used the Best Cartesian Average and Best Intermediate Methods using the MOE Alignments.

- Generated Comparative Models in Modeller.
  - Used Modeller and MOE Alignments.
RMS deviations

**MT1 RMSD Values**
- Modeller/ModellerAlign – Modeller/MOEAlign : 5.18247
- Modeller/ModellerAlign – MOE/ModellerAlign : 3.41042
- Modeller/ModellerAlign – MOE/MOEAlign : 5.3215
- Modeller/MOEAlign – MOE/ModellerAlign : 5.40229
- Modeller/MOEAlign – MOE/MOEAlign : 3.82068
- MOE/ModellerAlign – MOE/MOEAlign : 5.22865

**MT2 RMSD Values**
- Modeller/ModellerAlign – Modeller/MOEAlign : 5.108888
- Modeller/ModellerAlign – MOE/ModellerAlign : 5.05583
- Modeller/ModellerAlign – MOE/MOEAlign : 5.28168
- Modeller/MOEAlign – MOE/ModellerAlign : 3.91024
- Modeller/MOEAlign – MOE/MOEAlign : 3.70095
- MOE/ModellerAlign – MOE/MOEAlign : 2.63958

**MT1 CartesianAverage – Best Intermediate** : 2.58633
**MT2 CartesianAverage – Best Intermediate** : 2.49331
Ramachandran Plot

mtr1proteinBEST1F88MODELLERALIGNMENT

Plot statistics

- Residues in main chain allowed regions: 470 (97.8%)
- Residues in generously allowed regions (4.5-12.5): 4 (1.7%)
- Residues in disallowed regions: 0
- Number of core glycine and unpaired residues: 23
- Number of residues with classical beta (C-beta) conformation: 1
- Number of residues with unusual phi/psi angles: 15
- Number of polar residues: 110
- Total number of residues: 483

Based on analysis of 11 Ramachandran plots of proteins from 95 archaea and eukaryotic species, the above model is considered to have a good overall quality.
<table>
<thead>
<tr>
<th></th>
<th>G-factors</th>
<th></th>
<th>Covalent:</th>
<th></th>
<th>Overall:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MT1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MODELLER/MODELLERALIGNMENT</td>
<td>G-factors</td>
<td></td>
<td>Dihedrals:</td>
<td>-0.04</td>
<td>Covalent:</td>
</tr>
<tr>
<td>MODELLER/MOEALIGNMENT</td>
<td>G-factors</td>
<td></td>
<td>Dihedrals:</td>
<td>-0.09</td>
<td>Covalent:</td>
</tr>
<tr>
<td>MT1MOE/MODELLERALIGNMENT</td>
<td>G-factors</td>
<td></td>
<td>Dihedrals:</td>
<td>-0.55</td>
<td>Covalent:</td>
</tr>
<tr>
<td>MT1MOE/MOEALIGNMENT</td>
<td>G-factors</td>
<td></td>
<td>Dihedrals:</td>
<td>-0.74</td>
<td>Covalent:</td>
</tr>
<tr>
<td><strong>MT2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MODELLER/MODELLERALIGNMENT</td>
<td>G-factors</td>
<td></td>
<td>Dihedrals:</td>
<td>-0.02</td>
<td>Covalent:</td>
</tr>
<tr>
<td>MODELLER/MOEALIGNMENT</td>
<td>G-factors</td>
<td></td>
<td>Dihedrals:</td>
<td>-0.14</td>
<td>Covalent:</td>
</tr>
<tr>
<td>MOE/MODELLERALIGNMENT</td>
<td>G-factors</td>
<td></td>
<td>Dihedrals:</td>
<td>-0.86</td>
<td>Covalent:</td>
</tr>
<tr>
<td>MOE/MOEALIGNMENT</td>
<td>G-factors</td>
<td></td>
<td>Dihedrals:</td>
<td>-0.79</td>
<td>Covalent:</td>
</tr>
</tbody>
</table>
Data / ‘NPXXY’ Motif

MT1
Y7.53 - F7.60 :
Modeller/MoeAlign: 2.94 Angstroms
Modeller/ModellerAlign: 2.94 Angstroms
MOE/ModellerAlign: 3.69 Angstroms
MOE/MoeAlign: 3.72 Angstroms

MT2
Y7.53 - F7.60 :
Modeller/MoeAlign: 2.95 Angstroms
Modeller/ModellerAlign: 2.94 Angstroms
MOE/ModellerAlign: 3.65 Angstroms
MOE/MoeAlign: 3.72 Angstroms
Data / ‘NPXXY’ Motif

MT1
N7.49 - D2.50 :
  Modeller/MoeAlign: 6.16 Angstroms
  Modeller/ModellerAlign: 6.24 Angstroms
  MOE/ModellerAlign: 4.82 Angstroms
  MOE/MoeAlign: 4.83 Angstroms

MT2
N7.49 - D2.50 :
  Modeller/MoeAlign: 6.89 Angstroms
  Modeller/ModellerAlign: 5.79 Angstroms
  MOE/ModellerAlign: 4.64 Angstroms
  MOE/MoeAlign: 6.20 Angstroms
Data / ‘NPXXY’ Motif

- Mutated 7.50A to 7.50P to test the influence of a proline kink in TM7.
  - MT1 Moe/MoeAlign – 4.52 (4.83) Angstroms
  - MT2 Moe/MoeAlign – 4.82 (6.20) Angstroms
  - MT1 Modeller/ModellerAlign – 5.82 (6.24) Angstroms
  - MT2 Modeller/ModellerAlign – 4.82 (5.79) Angstroms
Docking

<table>
<thead>
<tr>
<th>Color</th>
<th>Energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-9.31</td>
</tr>
<tr>
<td></td>
<td>-5.66</td>
</tr>
<tr>
<td></td>
<td>-5.47</td>
</tr>
<tr>
<td></td>
<td>-5.26</td>
</tr>
<tr>
<td></td>
<td>-5.03</td>
</tr>
</tbody>
</table>
Docking

<table>
<thead>
<tr>
<th>Color</th>
<th>Energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3.94</td>
</tr>
<tr>
<td></td>
<td>-3.92</td>
</tr>
<tr>
<td></td>
<td>-3.84</td>
</tr>
<tr>
<td></td>
<td>-2.73</td>
</tr>
<tr>
<td></td>
<td>3.01</td>
</tr>
</tbody>
</table>
The stereochemical aspects of the models created by Modeller were better than MOE when evaluated by PROCHECK.

The NAXXY motif of TM7 seem to serve the same function as the ‘NPXXY’ motif in other rhodopsin GPCRs.

The ‘Aromatic Cluster’ motif of the melatonin receptors seems to be similarly active in binding ligands.
References

- Current Protocols in Bioinformatics
- http://cnx.rice.edu/content/m11062/latest/
- Modeller Manual
- MOE Contents
- http://www.ebi.ac.uk/interpro/
- http://www.biochem.vt.edu/modeling/homology.html
- http://www.cis.upenn.edu/~krice/receptor.html
- Structural Motifs underlying G-Protein-Coupled Receptor Function
- BioCache “Getting Started” manual
- Devillers J. Genetic Algorithms in Molecular Modeling.
- http://www.biochem.vt.edu/modeling/homology.html
- http://biotech.ebi.ac.uk:8400/
- Structural Motifs Research Grant Proposal
- http://structure.icm.edu.pl/procheck/
Acknowledgements

- Bioengineering and Bioinformatics Summer Institute @PITT
- Department of Chemistry and Biochemistry, Center for Computational Sciences, Duquesne University
- The University of Pittsburgh
- Pittsburgh Supercomputing Center
- Carnegie Mellon University
- The National Institutes of Health
- The National Science Foundation
- Dr. Ivet Bahar, BBSI Director, CCBB@PITT
- Dr. Rajan Munshi, BBSI coordinator, CCBB@PITT
- BBSI Instructors and Students
- Carmen Rios, Symposium Coordinator, Duquesne University
- Sandy Russell, Duquesne University
- The Theory Group @ Duquesne University
- Dr. Jeffry D. Madura
- Dr. Surratt
- Dr. Pranav Dalal
- Dr. Susan Gregurick
- Laura Thomas
Alignment

- **PAM Matrices** – (Percentage of Acceptable point Mutations)
  - Lower PAM matrices find short alignments of similar sequences.
  - Longer PAM matrices are used for longer more divergent sequences.

- **BLOSUM Matrices** –
  - Unlike PAM matrices these matrices were created using more distant sequences.
  - Matrices differ depending on the percent identity of the sequences they are derived from.
G-Protein Coupled Receptors (GPCRs)

- Integral membrane proteins made up of seven alpha helices and most notably involved in signal transduction.
- The binding of a ligand will alter the conformation of the receptor enough to cause a change in the heterotrimeric G-protein coupled to it.
- A GTP will replace a GDP attached to the alpha subunit of this G-protein, resulting in a change in a nearby effector molecule, which in turn sends out second messengers such as cAMP throughout the cell.
- This signal pathway is amplified through each step as many G-proteins can activate many effectors which in turn produce a large number of second messengers.

http://www.emory.edu/CHEMISTRY/justice/seminar/receptor1.htm