Knowledge Discovery of Protein-Ligand Interaction Network

Lei Xie, Li Xie, Cuong Dao, Augustus Lu, Philip E. Bourne
San Diego Supercomputer Center
University of California, San Diego
Protein-Ligand Interaction Network Analysis

- Identification of protein-drug interactions on a genome scale.
- Association of genes, biological networks, and pathways with their biological context (phenotype, disease, drug efficacy, side effects etc.)
- Polypharmacology design, drug efficacy, and side effect
Study of Interactome-Phenome Correlation by Integrating Semantic Techniques with Molecular Modeling

- Binding Site Similarity
- Small Molecule Similarity
- Protein-Ligand Docking & MD Simulation
- Network Analysis & Systems Biology

Text Mining & Ontology
Issues in Mining Protein-Ligand Interaction

• Chemical and biological name entity recognition and object mapping

• Protein-ligand relation detection

• Benchmark to evaluate the performance
A Biomedical Search Engine
(http://www.novoseek.com)

- Index medline abstract, pubmed central full text and NIH grant

- Entity disambiguation, recognition and mapping
  - Chemical: alternative names, links to pubchem, drugbank, ZINC, chemidplus, CAS etc., ontology (MESH, CHEBI)
  - Protein: alternative names, links to uniprot, refseq, pir, PDB, Pfam, interpro, reactome, kegg etc., ontology (MESH, GO)

- Programming API
A Protein-Ligand Interaction Corpus

~2,000 literatures cited in Drugbank
500 non-redundant abstracts that describe the protein-ligand interaction

@note annotation tool:  http://sysbio.di.uminho.pt/anote
## Performance of Name Entity Recognition

<table>
<thead>
<tr>
<th></th>
<th><strong>Chemical</strong></th>
<th></th>
<th><strong>Protein</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>precision</td>
<td>recall</td>
<td>precision</td>
<td>recall</td>
</tr>
<tr>
<td>novoseek</td>
<td>84.6%</td>
<td>43.7%</td>
<td>66.8%</td>
<td>17.5%</td>
</tr>
<tr>
<td>HMM (genetag)</td>
<td>-</td>
<td>-</td>
<td>63.1%</td>
<td>17.8%</td>
</tr>
<tr>
<td>Oscar3</td>
<td>36.8%</td>
<td>66.8%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dictionary*</td>
<td></td>
<td></td>
<td>67.1%</td>
<td>17.2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>23.2%</td>
<td>31.3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13.4%</td>
<td>52.3%</td>
</tr>
<tr>
<td>Onto-BLAST*</td>
<td></td>
<td></td>
<td>78.3%</td>
<td>24.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>63.8%</td>
<td>31.7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>23.6%</td>
<td>53.5%</td>
</tr>
</tbody>
</table>

* PRotein Ontology (PRO): [http://www.obofoundry.org/]
Onto-Blast: A New Algorithm for Name Entity Recognition

Ontology

>Kinase
AGCTACCTCAAACG
AACT

Free Text

>Kinase inhibitor …
AGCTACCTCAAACG
AACTCATTCT……

BLAST

Post Processing
Protein-Ligand Relation Detection Through Cross-Document Association

Ligand

Protein

Pathway, Phenotype, etc.

?
Protein-Ligand Interaction Modeling Ontology (PLIMO)

- Modeling protein-ligand interaction on multi-scales from atomic level to biological network
- Correlation of protein-ligand interaction to cellular functions
- Maximum reuse of existing ontology (BFO, PRO, CHEBI, phenotype, disease etc.)

Basic Formal Ontology
Examples of Entities and Relations in PLIMO

- Systems Biology Simulation
  - biological network
  - biological module
  - protein-ligand complex
  - chemical
  - protein biological unit
  - protein
  - protein fragment
  - protein domain
  - protein functional site
  - ligand binding site

- Molecular Modeling of Protein-Ligand Interaction
  - binding activity
  - interaction type
  - physical contact
  - hydrophobic interaction
  - hydrogen bonding interaction

- Bioinformatics
  - participate_in

- Text Mining
Molecular Modeling & Systems Biology Simulation

PLIM Ontology

Cross document relation association

Onto-BLAST

novo|seek
Case Study: Side Effect Profile of Cholesteryl Ester Transfer Protein Inhibitors

- CETP inhibitors are developed to lower cholesterol.
- Torcetrapib causes deadly side-effect of hypertension. It was withdrawn from Phase III clinical trial.
- Unknown off-targets may be involved in the control of aldosterone level in the kidney.
- No extensive hypertension has been observed for two other CETP inhibitors anacetrapib and JTT-705.

- Off-targets of CETP inhibitors ???
Association Search

Torcetrapib
Anacetrapib
JTT-705

ACE
renin
AT1

Mineralocorticoid receptor
Vitamin D receptor
PPAR-alpha

Renin-angiotensin system

NR
Structural Proteome-Wide Ligand Binding Site Similarity to CETP

*SMAP* ([http://funsite.sdsc.edu](http://funsite.sdsc.edu))

Off-target Binding Profiles Of CETP Inhibitors

Predicated binding affinity - red: strong, purple: weak, blue: not binding

<table>
<thead>
<tr>
<th>Protein</th>
<th>Normalized Docking Score</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Anacetrapib</td>
<td>Torcetrapib</td>
<td>JTT-705</td>
<td></td>
</tr>
<tr>
<td>CETP</td>
<td>-4.6705</td>
<td>-5.6024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retinoid X receptor (agonist)</td>
<td>-4.1922</td>
<td>-5.5803</td>
<td>-0.9344</td>
<td></td>
</tr>
<tr>
<td>PPARδ (agonist)</td>
<td>-3.8384</td>
<td>-3.8703</td>
<td>-1.5662</td>
<td></td>
</tr>
<tr>
<td>PPARα (agonist)</td>
<td>6.6785</td>
<td>-4.0828</td>
<td>-3.0660</td>
<td></td>
</tr>
<tr>
<td>PPARγ (agonist)</td>
<td>6.0096</td>
<td>-3.9838</td>
<td>-2.0316</td>
<td></td>
</tr>
<tr>
<td>LXRα (agonist)</td>
<td>6.3052</td>
<td>5.7793</td>
<td>-0.6900</td>
<td></td>
</tr>
<tr>
<td>LXRβ (agonist)</td>
<td>5.5450</td>
<td>5.0882</td>
<td>-1.7543</td>
<td></td>
</tr>
<tr>
<td>Vitamin D receptor (agonist)</td>
<td>6.1759</td>
<td>5.7622</td>
<td>-1.1761</td>
<td></td>
</tr>
<tr>
<td>Glucocorticoid receptor (agonist)</td>
<td>6.1432</td>
<td>5.5504</td>
<td>-2.0131</td>
<td></td>
</tr>
</tbody>
</table>
Combinatorial Control May Play a Role in Clinical Indications of CETP Inhibitors

Drugs: Torcetrapib, Anacetrapib, JTT-705
Targets: RXR, PPARα, PPARδ, PPARγ, LXRα, LXRβ, VDR, GCR, FABP
Pathways: RAAS, NF-κB pathway, Repression of inflammatory genes, Activation of M2 macrophage
Clinical indications: Hypertension, Inflammation, Cancer
Summary and Future Works

- Integration of text mining, ontology and molecular modeling is a valuable tool to generate testable hypothesis that associates interactome with phenome.
- Biological and chemical name entity recognition and relationship detection are still challenges. Incorporation of linguistics features into Onto-BLAST may improve its performance.
Acknowledgements

- Prof. Phil Bourne (School of Pharmacy, UCSD)
- Dr. Li Xie (School of Pharmacy, UCSD)
- Mr. Augustus Lu (Bioinformatics, UCSD)
- Mr. Cuong Dao (Bioinformatics, UCSD)

GM078596