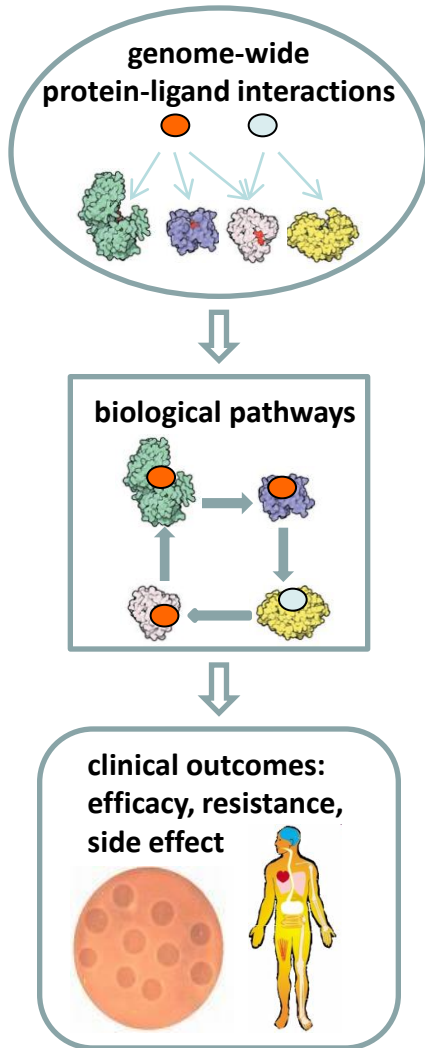


Knowledge Discovery of Protein- Ligand Interaction Network

*Lei Xie, Li Xie, Cuong Dao, Augustus
Lu, Philip E. Bourne*

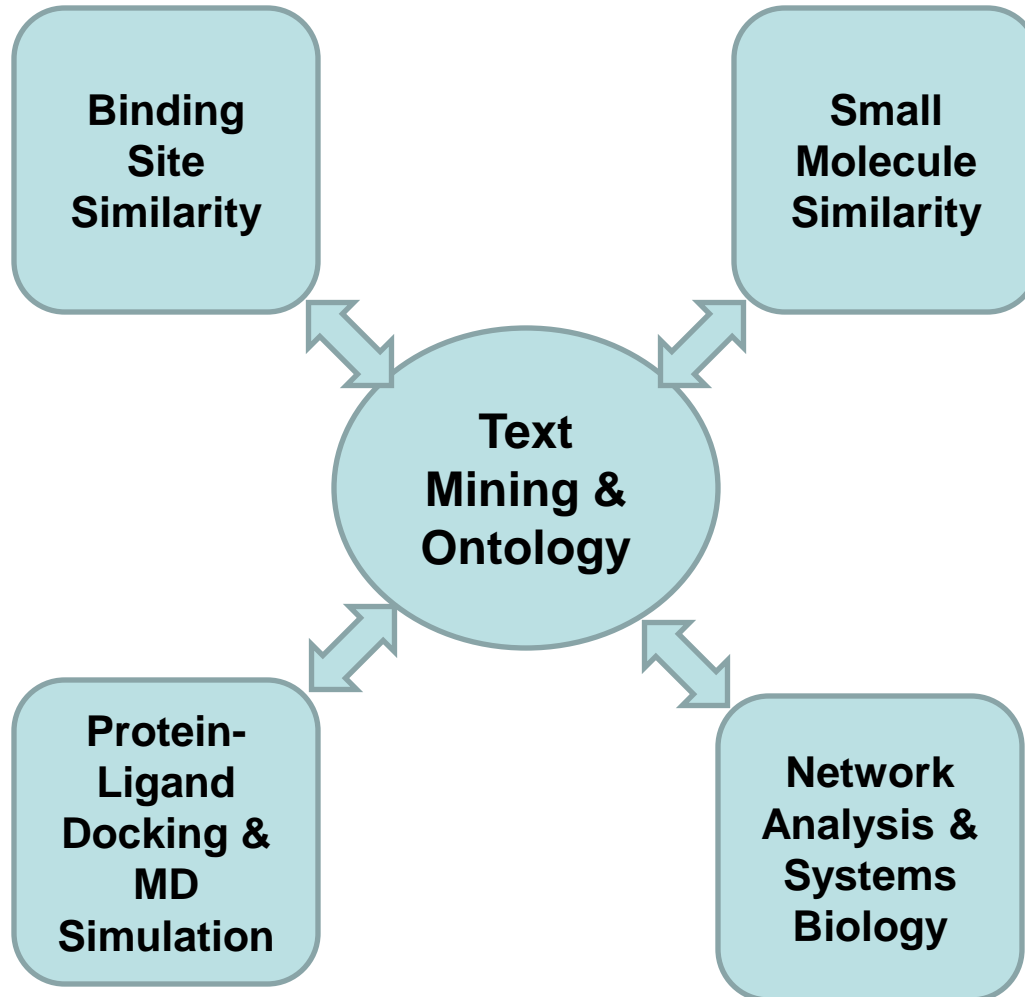
San Diego Supercomputer Center
University of California, San Diego

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



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

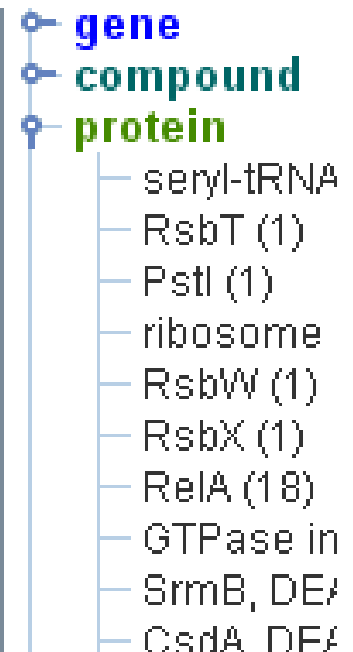
A Protein-Ligand Interaction Corpus

~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>

. **RelA** generates the alarmone (p)ppGpp under amino acid starvation, whereas **SpoT** under a variety of cellular stress conditions. It is widely accepted that **RelA** has been controversial. **SpoT** physically interacts with the ribosome-associated **CgtA**, **SpoT** is also associated with a pre-50S particle. Analysis of **spoT** and **cgtA** genes of **SpoT** and **CgtA** are mutually independent. The steady-state level of (p)ppGpp during amino acid starvation is not affected, providing strong evidence that **CgtA** is a stringent response. We show that **CgtA** is not associated with pre-50S particles. (p)ppGpp accumulates, **CgtA** is not bound either to the pre-50S particle or to **CgtA** promotes **SpoT** (p)ppGpp degradation activity on the ribosome and that the **CgtA** function under stress conditions. Intriguingly, we found that in the absence of **spoT**

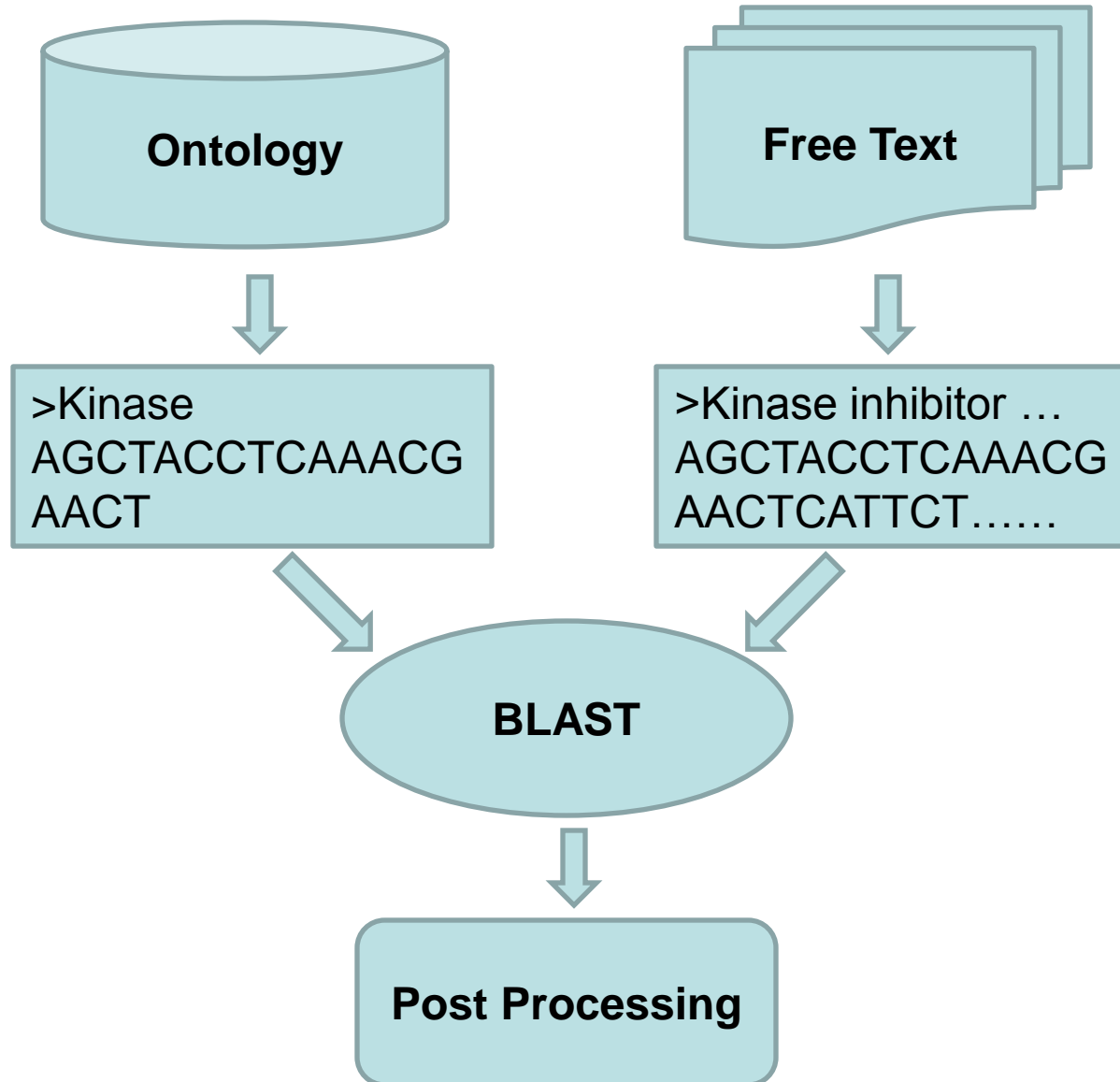


Performance of Name Entity Recognition

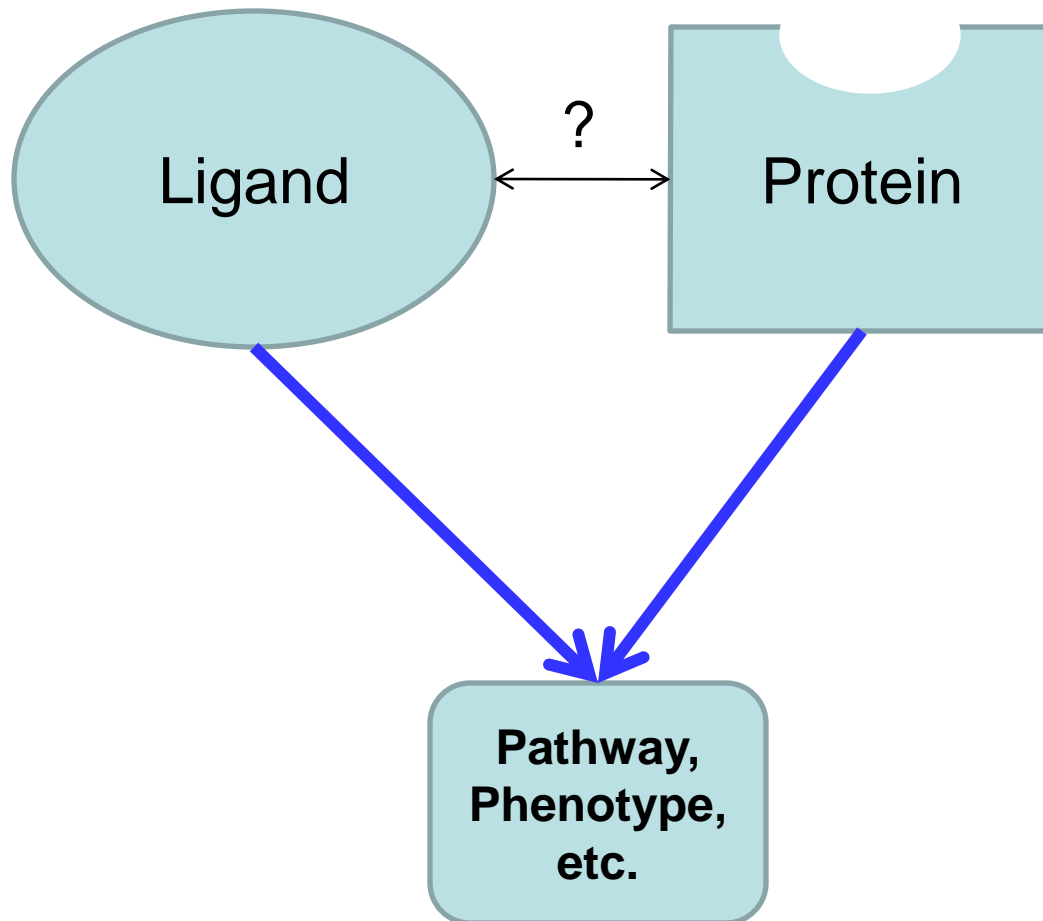
	Chemical		Protein	
	precision	recall	precision	recall
novoseek	84.6%	43.7%	66.8%	17.5%
HMM (genetag)	-	-	63.1%	17.8%
Oscar3	36.8%	66.8%	-	-
Dictionary*			67.1% 23.2% 13.4%	17.2% 31.3% 52.3%
Onto-BLAST*			78.3% 63.8% 23.6%	24.8% 31.7% 53.5%

* PRotein Ontology (PRO) : <http://www.obofoundry.org/>

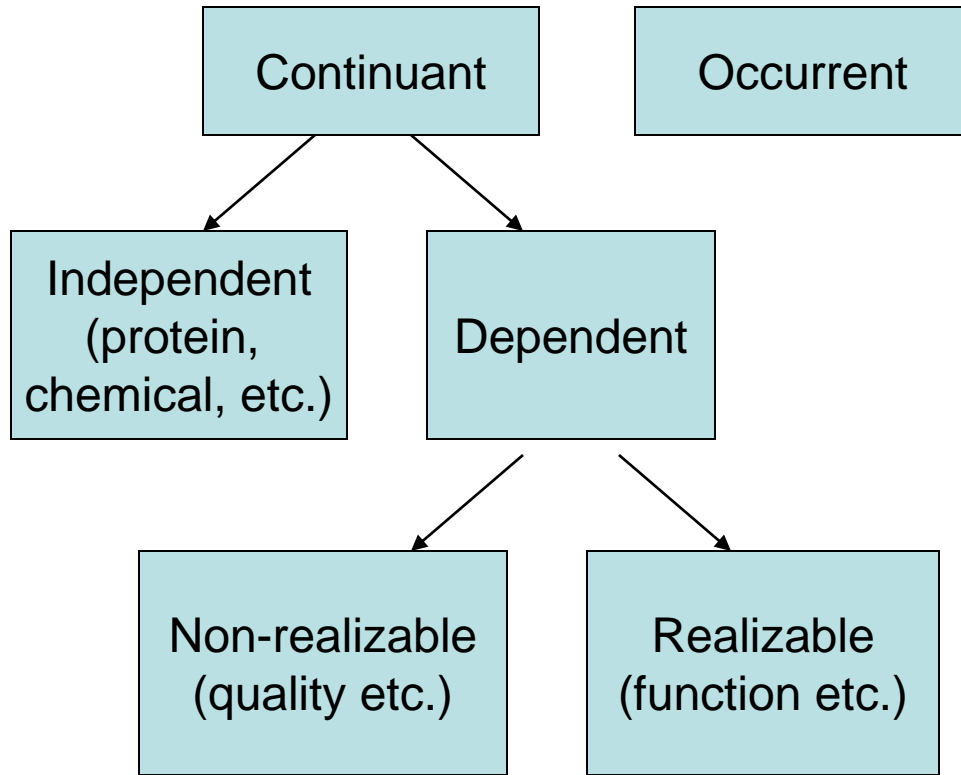
Onto-Blast: A New Algorithm for Name Entity Recognition



Protein-Ligand Relation Detection Through Cross-Document Association



Protein-Ligand Interaction Modeling Ontology (PLIMO)

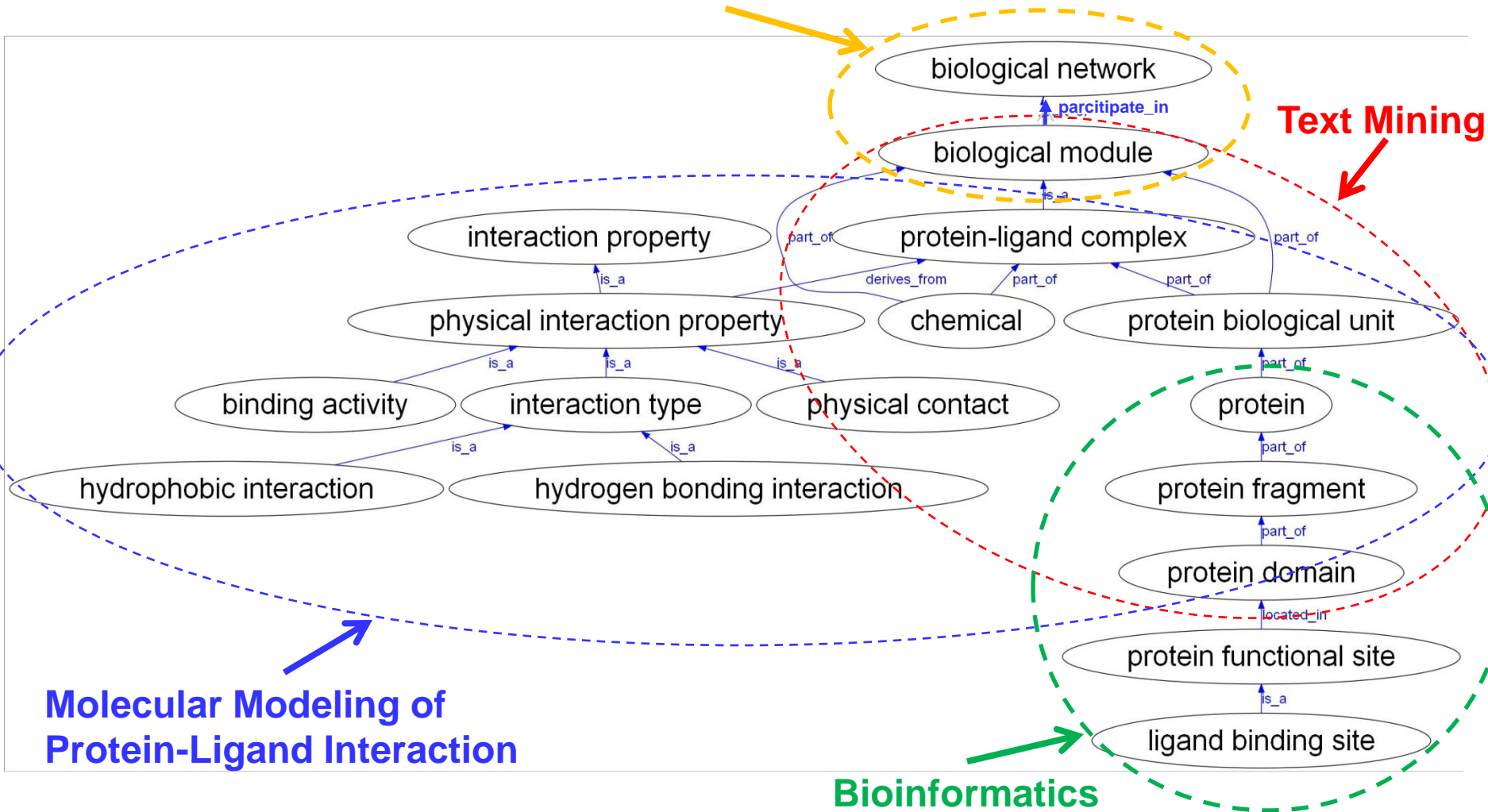


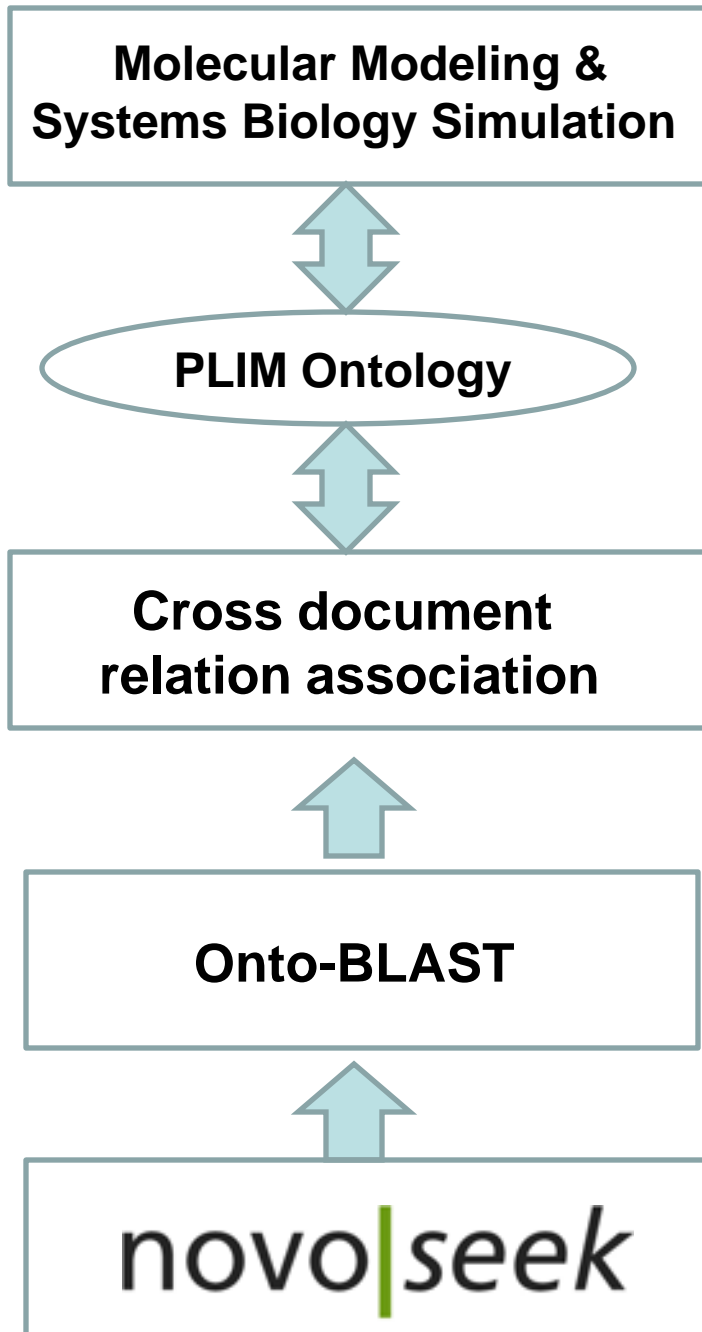
Basic Formal Ontology

- 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.)

Examples of Entities and Relations in PLIMO

Systems Biology Simulation

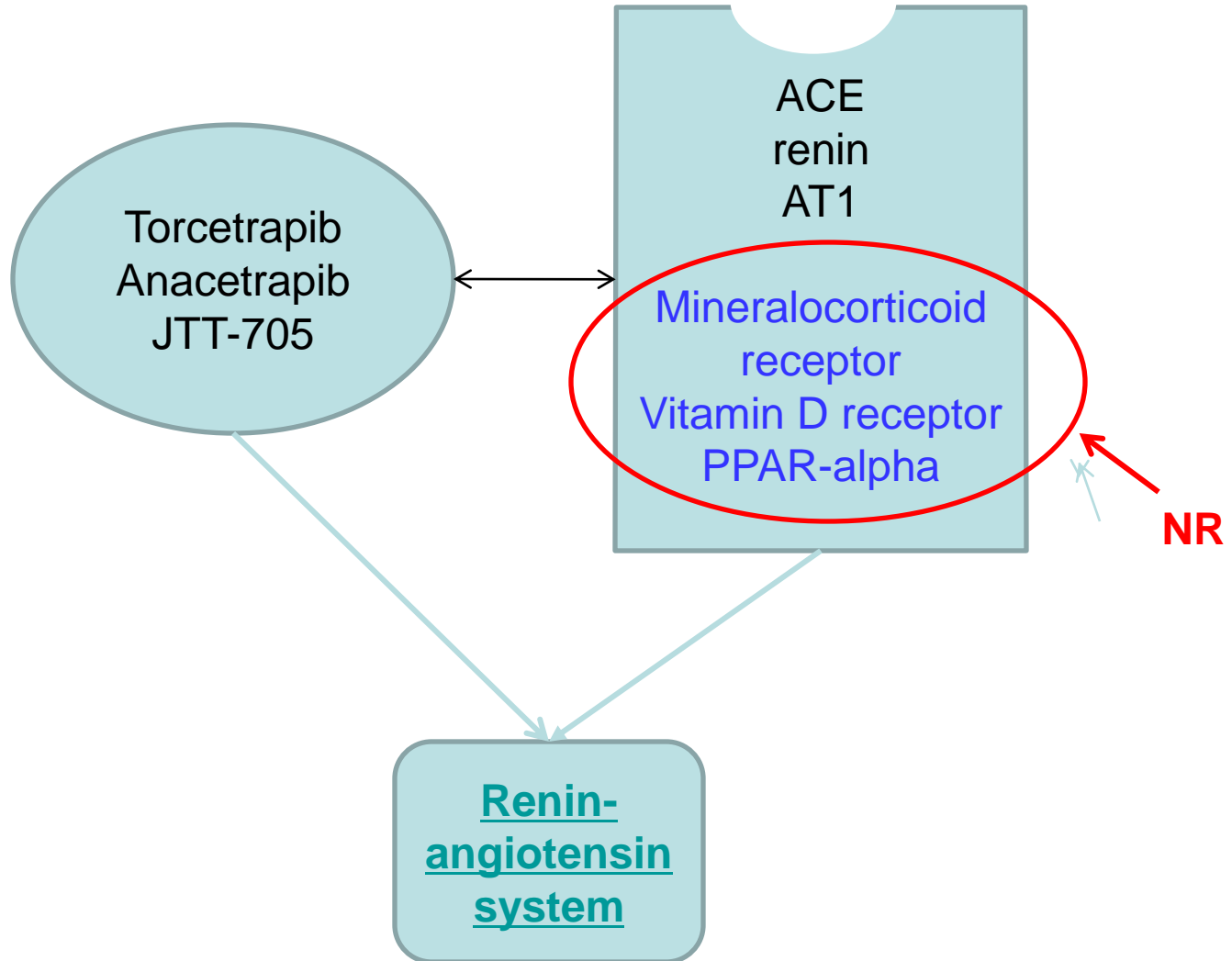




Case Study: Side Effect Profile of Cholesteryl Ester Transfer Protein Inhibitors

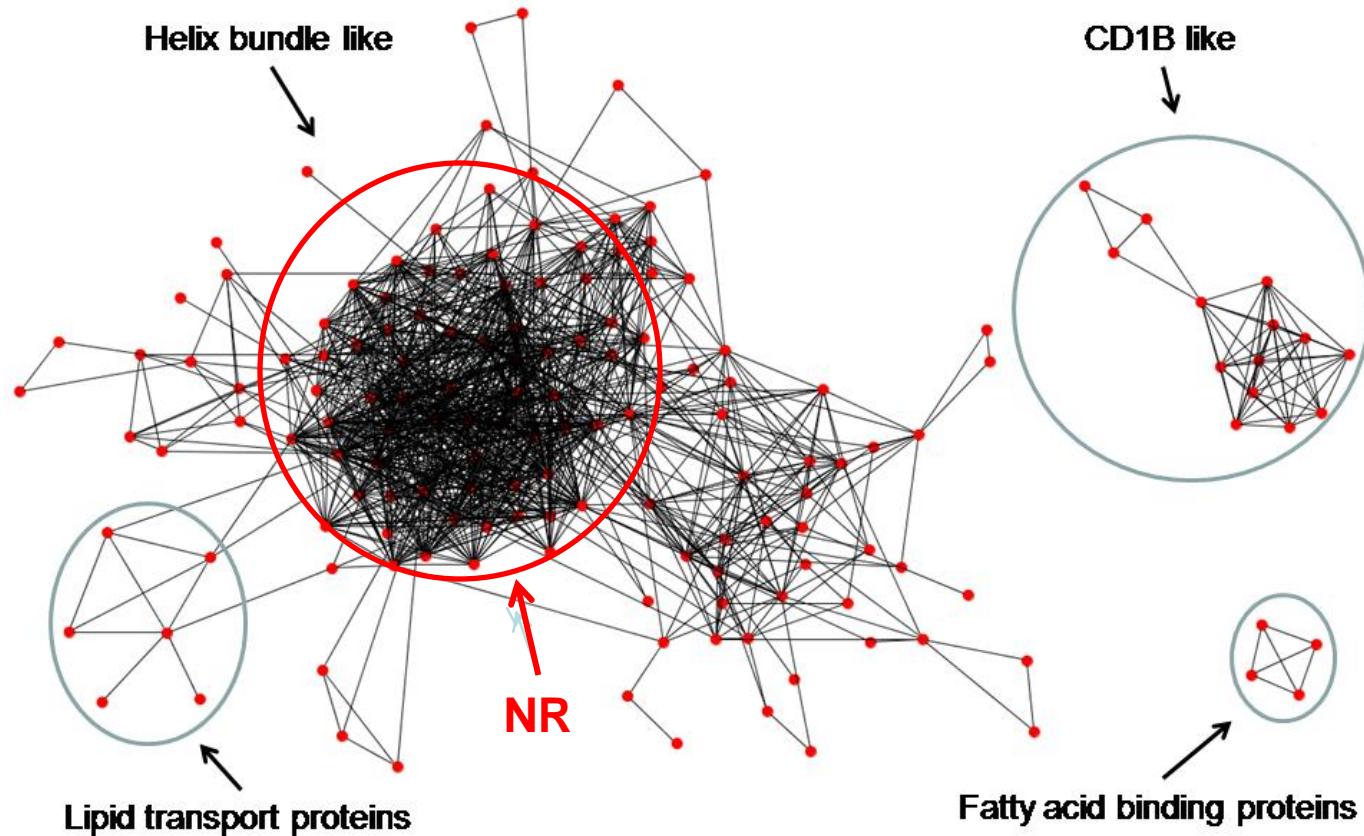
- CETP inhibitors are developed to lower cholesterol.
- Torcetrapib causes deadly side-effect of hypertension. It was withdrawn from Phase III clinical trial
- Unknown off-targets may be involved in the control of aldosterone level in the kidney
- No extensive hypertension has been observed for two other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

Association Search



Structural Proteome-Wide Ligand Binding Site Similarity to CETP

SMAP (<http://funsite.sdsc.edu>)



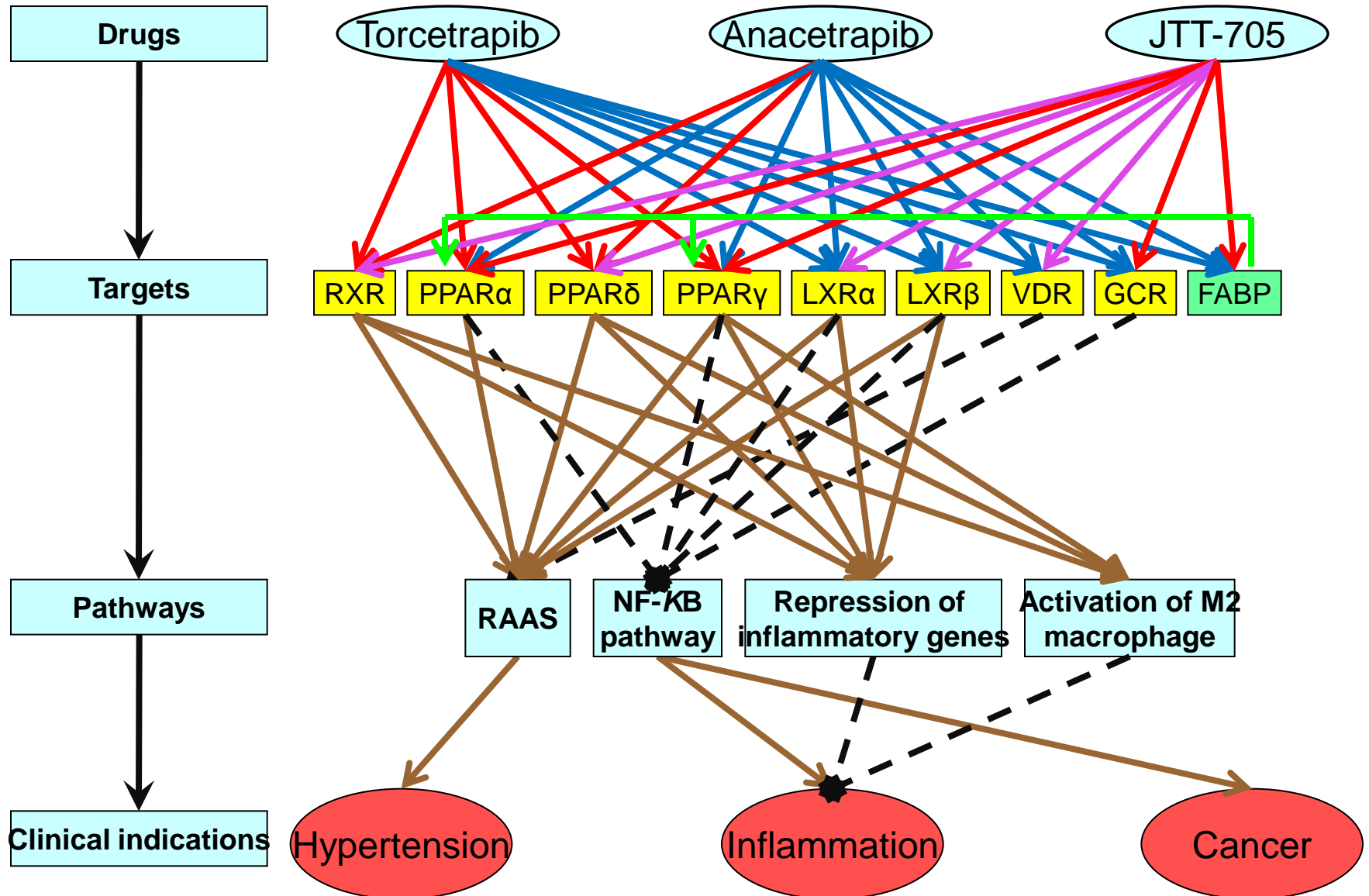
Xie L, Li J, Xie L, Bourne PE (2009) *PLoS Comput Biol* 5(5): e1000387

Off-target Binding Profiles Of CETP Inhibitors

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

Protein	Normalized Docking Score		
	<i>Anacetrapib</i>	<i>Torcetrapib</i>	<i>JTT-705</i>
CETP	-4.6705	-5.6024	-
Retinoid X receptor (agonist)	-4.1922	-5.5803	-0.9344
PPAR δ (agonist)	-3.8384	-3.8703	-1.5662
PPAR α (agonist)	6.6785	-4.0828	-3.0660
PPAR γ (agonist)	6.0096	-3.9838	-2.0316
LXR α (agonist)	6.3052	5.7793	-0.6900
LXR β (agonist)	5.5450	5.0882	-1.7543
Vitamin D receptor (agonist)	6.1759	5.7622	-1.1761
Glucocorticoid receptor (agonist)	6.1432	5.5504	-2.0131

Combinatorial Control May Play a Role in Clinical Indications of CETP Inhibitors



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