Chemical aspects of the cell

Integration of biological systems and the use of chemical knowledge

Part 1 - Systems biology

Network topology

Nodes: these are the objects under analaysis (for example proteins in a proteome study).

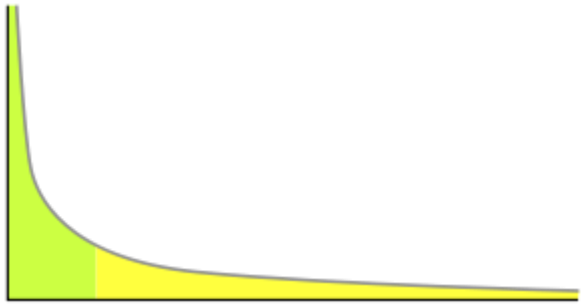
Connections: these nodes (objects) connect to each other in different ways. It could, for example, be defined as intermolecular interactions between proteins. Therefore, proteins are nodes and when two or more proteins interact it is represented by the connections.

Centrality: based on the assumption that a network is not random, it is possible to identify central nodes that are the most important ones in terms of connetions (i.e. the most interconnected nodes). This means that these nodes are central to the system.

Hubs: highly connected nodes are known as hubs, which are responsible to intermediate many events in a system.

x

Network topology



Pareto distribution: 20/80!

Exponential topology:

Scale-free: most uncoordinated network, with randomnes increment achieved as the proteins connects to each other without any order and having the same importance to the network.

Power law: it is a nonlinear functional relationship between two quantities.

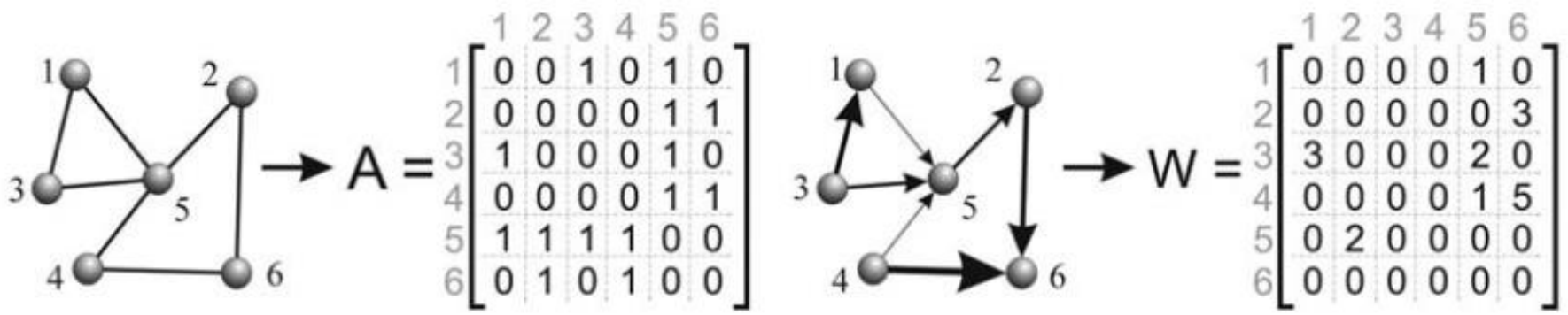
Example: $y = ax^2$

What type of intermolecular interaction works that way?

x

Visual intepretation of networks

Examples of (a) an undirected network (graph) and its mapping on an adjacency matrix A; and (b) a directed weighted network (weighted digraph) and its respective mapping on a weight matrix W.



Costa, L.F.; Rodrigues, F.A.; Cristino, A.S. *Gen. Mol. Biol.* **2008**, *31*, 591-601

Visual intepretation of networks

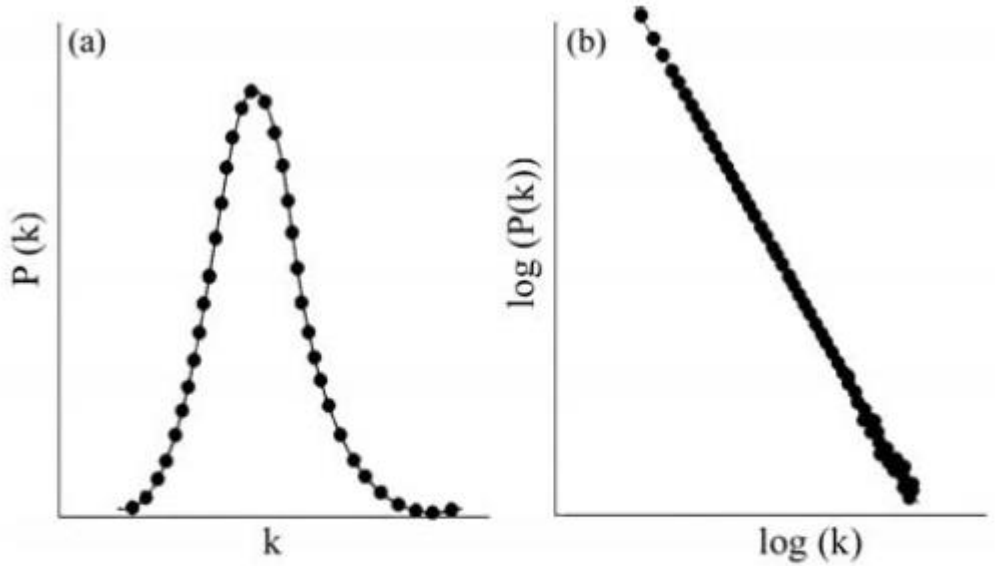
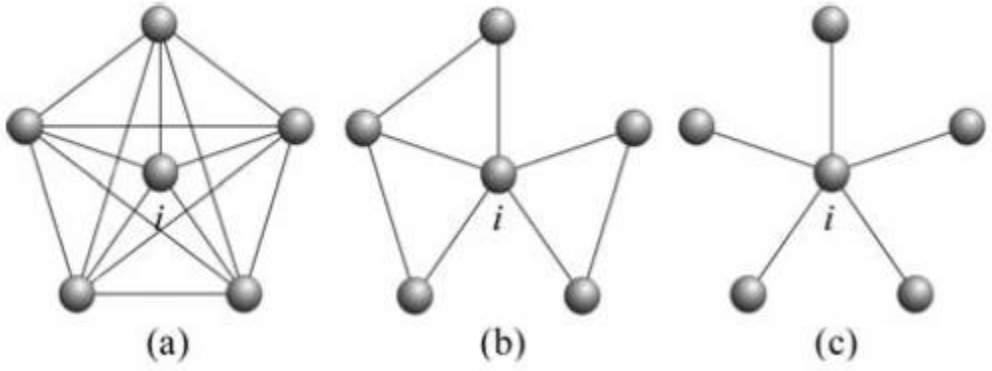


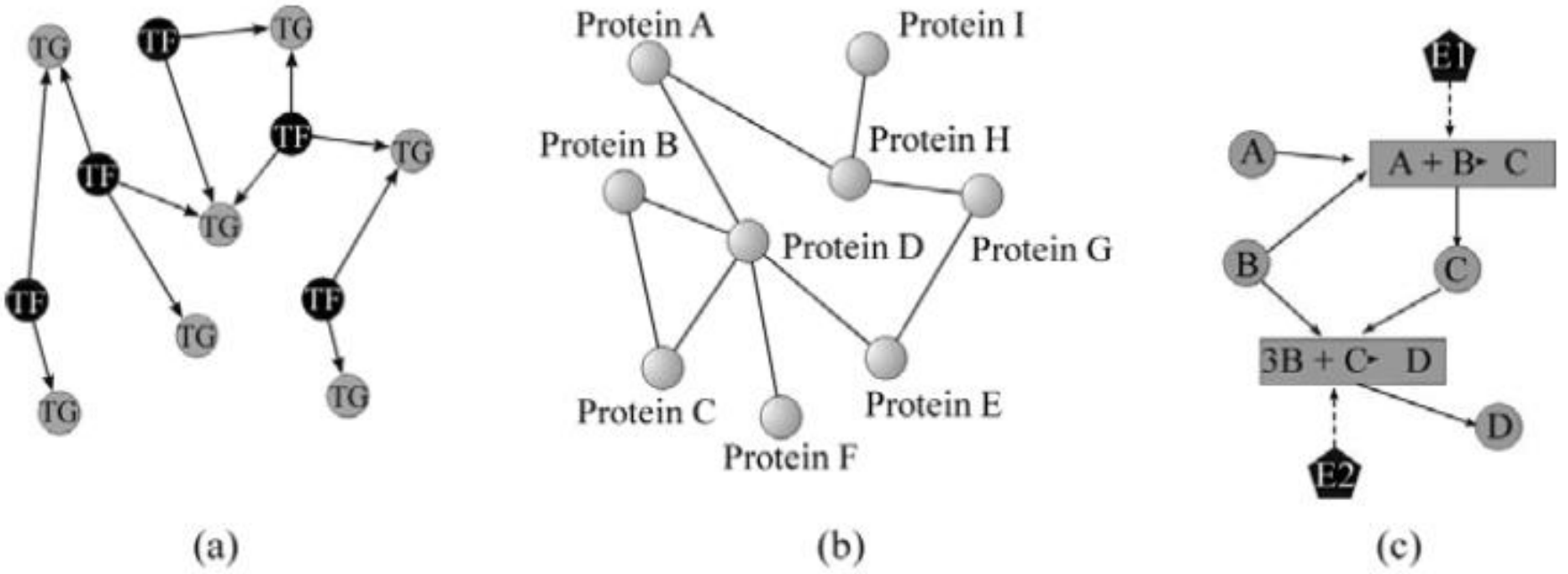
Figure 2 - Degree distributions for (a) random and (b) scale-free networks. While random networks present a peak distribution, scale-free networks present a straight line in the log-log plot.



Cluster coefficients from high (a), down to low (b) and none (c)

Costa, L.F.; Rodrigues, F.A.; Cristino, A.S. *Gen. Mol. Biol.* **2008**, *31*, 591-601

Visual intepretation of networks



The three main types of biological networks: (a) a transcriptional regulatory network has two components: transcription factor (TF) and target genes (TG), where TF regulates the transcription of TGs; (b) protein-protein interaction networks: two proteins are connected if there is a docking between them; (c) a metabolic network is constructed considering the reactants, chemical reactions and enzymes.

Costa, L.F.; Rodrigues, F.A.; Cristino, A.S. *Gen. Mol. Biol.* **2008**, *31*, 591-601

Databases

Accumulation of data for different macromolecules and systems in a set of databases:

Genes
Entrez gene: http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene
Proteins
SwissProt: http://expasy.org/sprot/
Structures of biological macromolecules
PDB: http://www.rcsb.org/pdb/home/home.do
Structural genomics consortium: http://www.sgc.utoronto.ca/
Pathways
KEGG: http://www.genome.jp/kegg/
MetaCyc: http://metacyc.org/
BioCarta: http://www.biocarta.com/genes/index.asp
Reactome: http://www.reactome.org/

Oprea, T.I.; May, E.E.; Leitão, A.; Tropsha, A. Computational Systems Chemical Biology. In: Chemoinformatics and Computational Chemical Biology, Jürgen Bajorath (ed.) vol. 672, chapter 18, Springer, 2011

Databases

Receptors
GPCRdb: http://www.gpcr.org/7tm/
NHRs: http://www.nursa.org/
Ion channels: http://www.iuphar-db.org/iuphar-ic/index.html
Biochemical pathway reaction kinetics
SABIORK: http://sabio.villa-bosch.de/SABIORK/
BRENDA: http://www.brenda.uni-koeln.de/
Annotated biological models
http://www.ebi.ac.uk/biomodels/
Other MLI initiatives
NIH Roadmap: http://nihroadmap.nih.gov/

[a]Non-exhaustive list

Oprea, T.I.; May, E.E.; Leitão, A.; Tropsha, A. Computational Systems Chemical Biology. In: Chemoinformatics and Computational Chemical Biology, Jürgen Bajorath (ed.) vol. 672, chapter 18, Springer, 2011

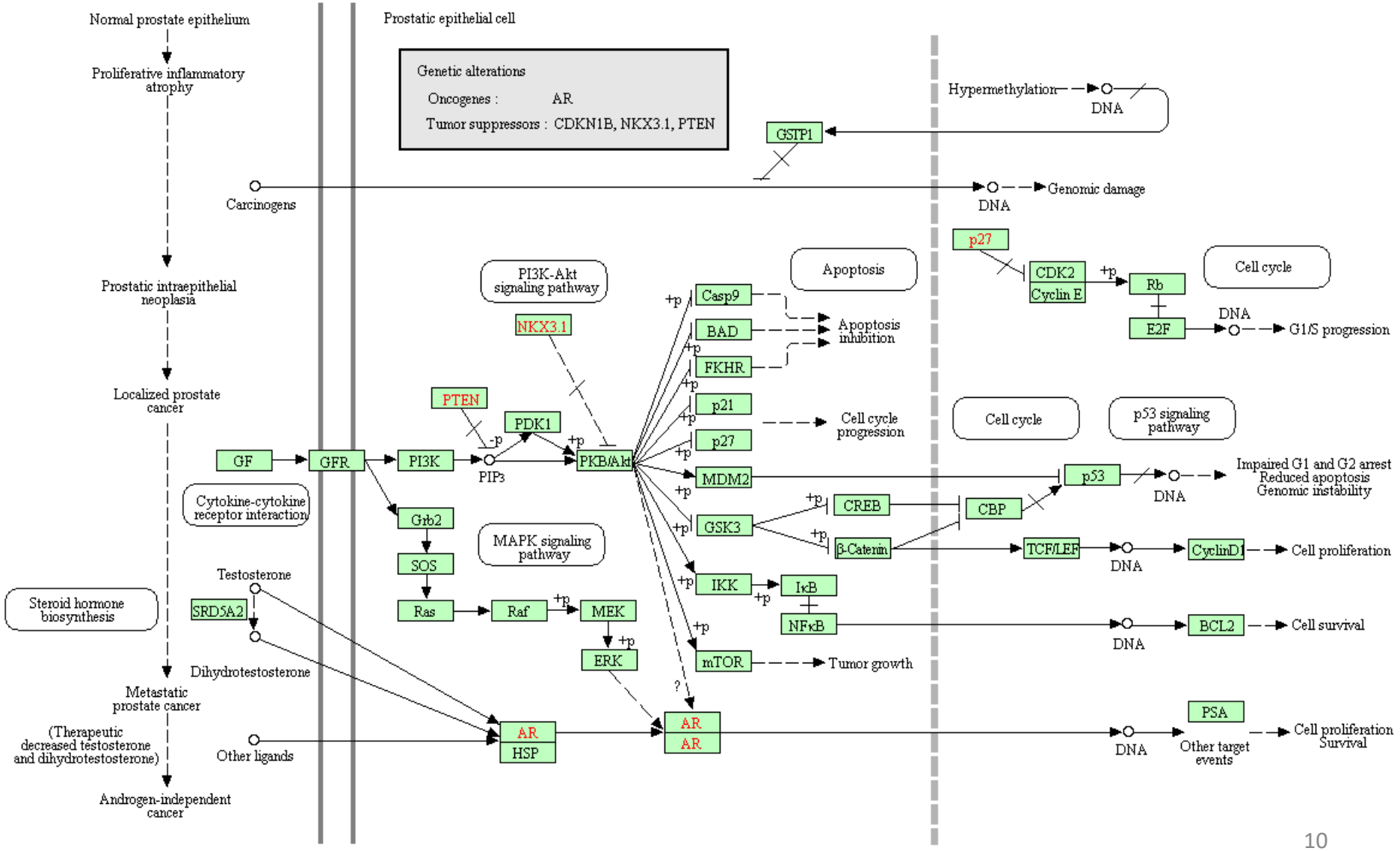
Databases

Relational databases also contain bioactivity assays:

Small molecules
PubChem: http://pubchem.ncbi.nlm.nih.gov/
NCI: http://dtp.nci.nih.gov/docs/dtp_search.html
WOMBAT: http://sunsetmolecular.com/
BINDING DB: http://www.bindingdb.org/bind/index.jsp
Metabolites: http://www.hmdb.ca/
Drugs and clinical candidates
NLM's Dailymed: http://dailymed.nlm.nih.gov/
DrugBank: http://drugbank.ca/
FDA: http://www.accessdata.fda.gov/scripts/cder/drugsatfda/
WHO essential drugs: http://www.who.int/medicines/publications/essentialmedicines/en/
Toxicology data
NIEHS: http://ntp.niehs.nih.gov/ntpweb/
EPA DSS-Tox: http://www.epa.gov/ncct/dsstox/index.html

Oprea, T.I.; May, E.E.; Leitão, A.; Tropsha, A. Computational Systems Chemical Biology. In: Chemoinformatics and Computational Chemical Biology, Jürgen Bajorath (ed.) vol. 672, chapter 18, Springer, 2011

Example: KEGG database for prostate cancer



KEGG database

Visual intepretation of networks

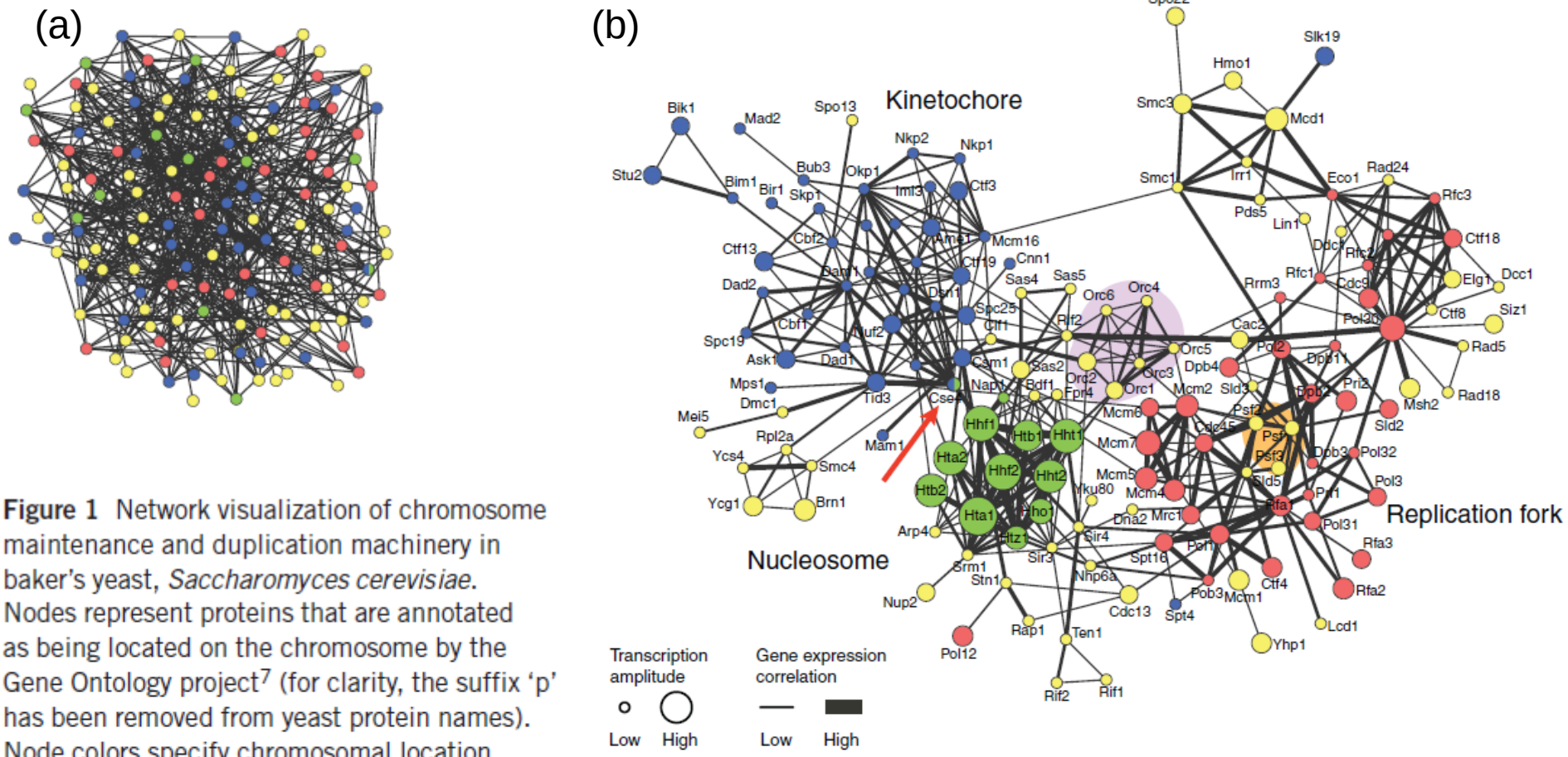
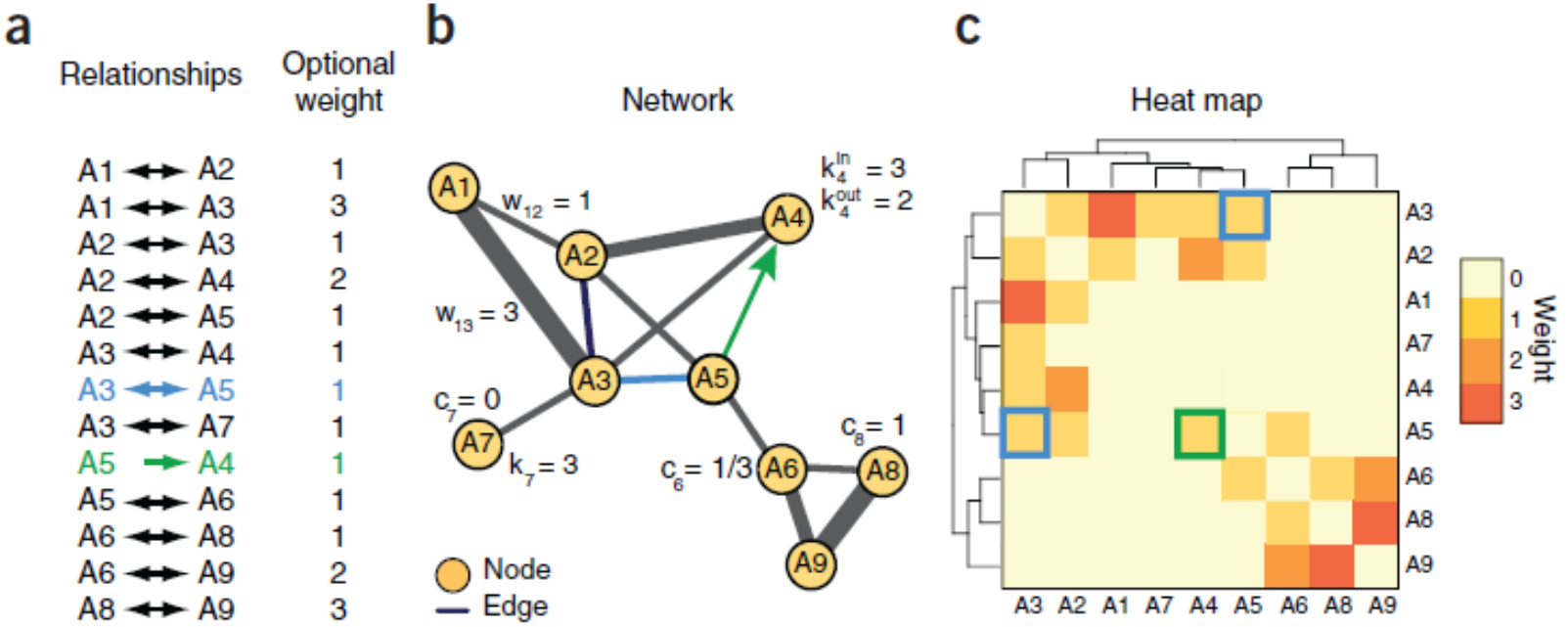


Figure 1 Network visualization of chromosome maintenance and duplication machinery in baker's yeast, *Saccharomyces cerevisiae*. Nodes represent proteins that are annotated as being located on the chromosome by the Gene Ontology project[7] (for clarity, the suffix 'p' has been removed from yeast protein names). Node colors specify chromosomal location subcategories: red, replication fork; green, nucleosome; blue, kinetochore; yellow, other chromosome components. Edges represent protein-protein interactions that were manually extracted from publications by BioGRID database

(**a**) Without specific layout, the network looks like a 'jumbled mess' and cannot be interpreted. (**b**) The same network after applying the force-directed layout and adding gene expression data of cells monitored during one round of the cell cycle are visually annotated on the network (data are from ref. 8). Edges are drawn thicker when the Pearson correlation between transcript profiles is higher.

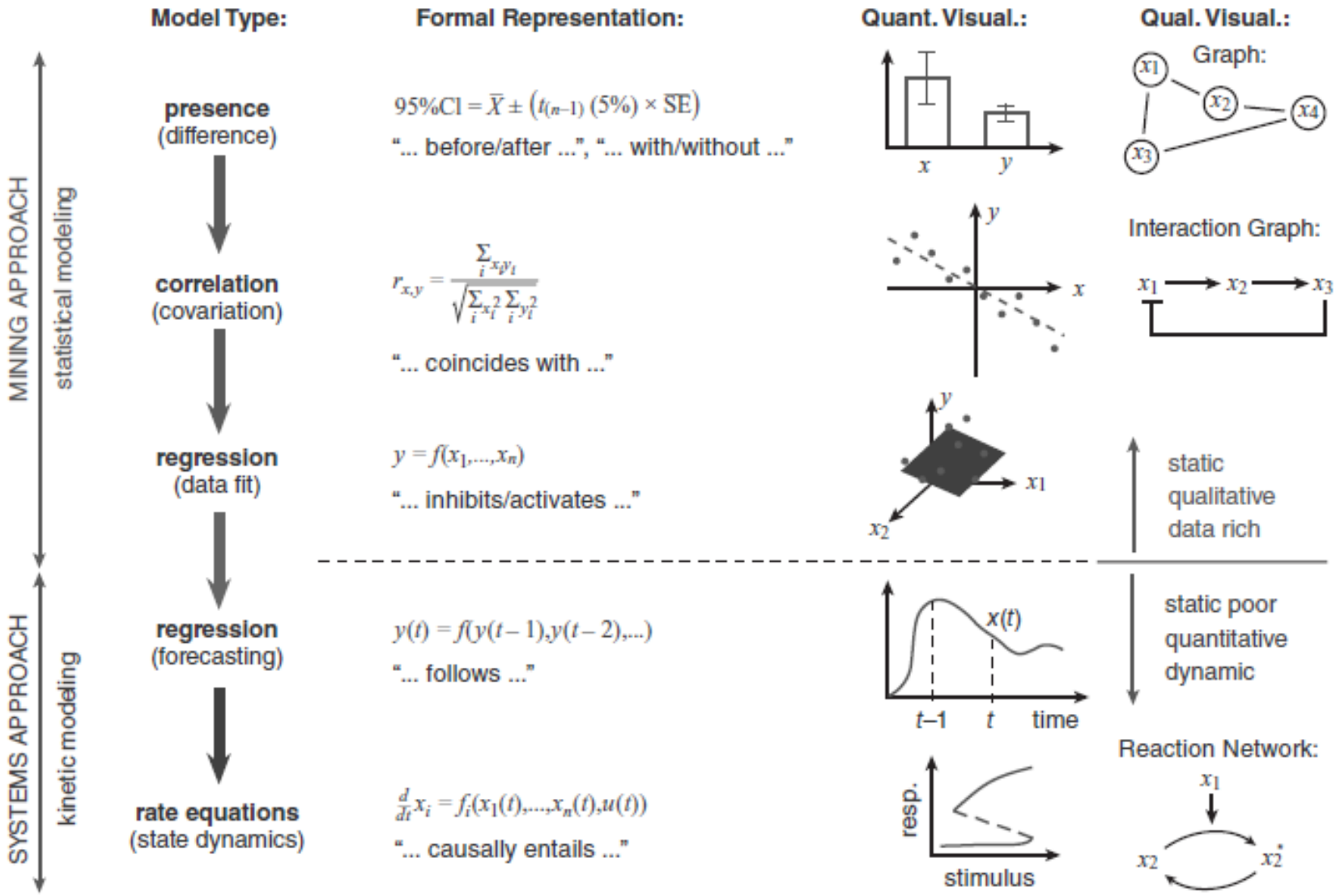
Merico, D.; Gfeller, D.; Bader, G.D. *Nat. Biotech.* **2009**, *27*, 921

Visual intepretation of networks



Merico, D.; Gfeller, D.; Bader, G.D. *Nat. Biotech.* **2009**, *27*, 921

Models for the systems



Young, D.L.; Michelson, S. "Systems Biology in Drug Discovery and Development" John Wiley & Sons, Inc., Hoboken, New Jersey, 2012

Integration in drug discovery

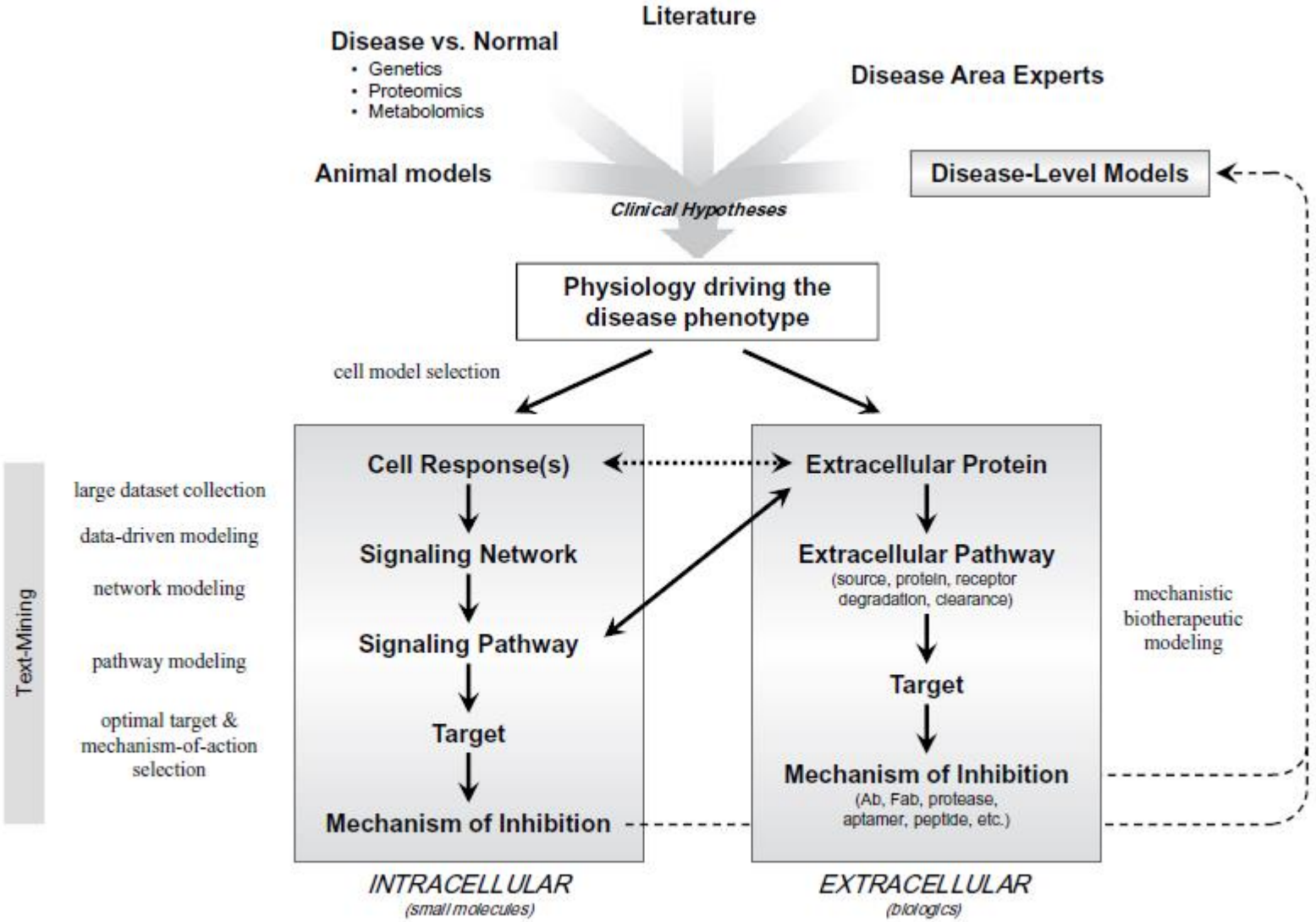


Figure 5.1 Conceptual workflow for integrating systems biology approaches across cell and disease scales.

Young, D.L.; Michelson, S. "Systems Biology in Drug Discovery and Development" John Wiley & Sons, Inc., Hoboken, New Jersey, 2012

Part 2
Systems chemical biology (SCB)

Systems chemical biology (SCB)

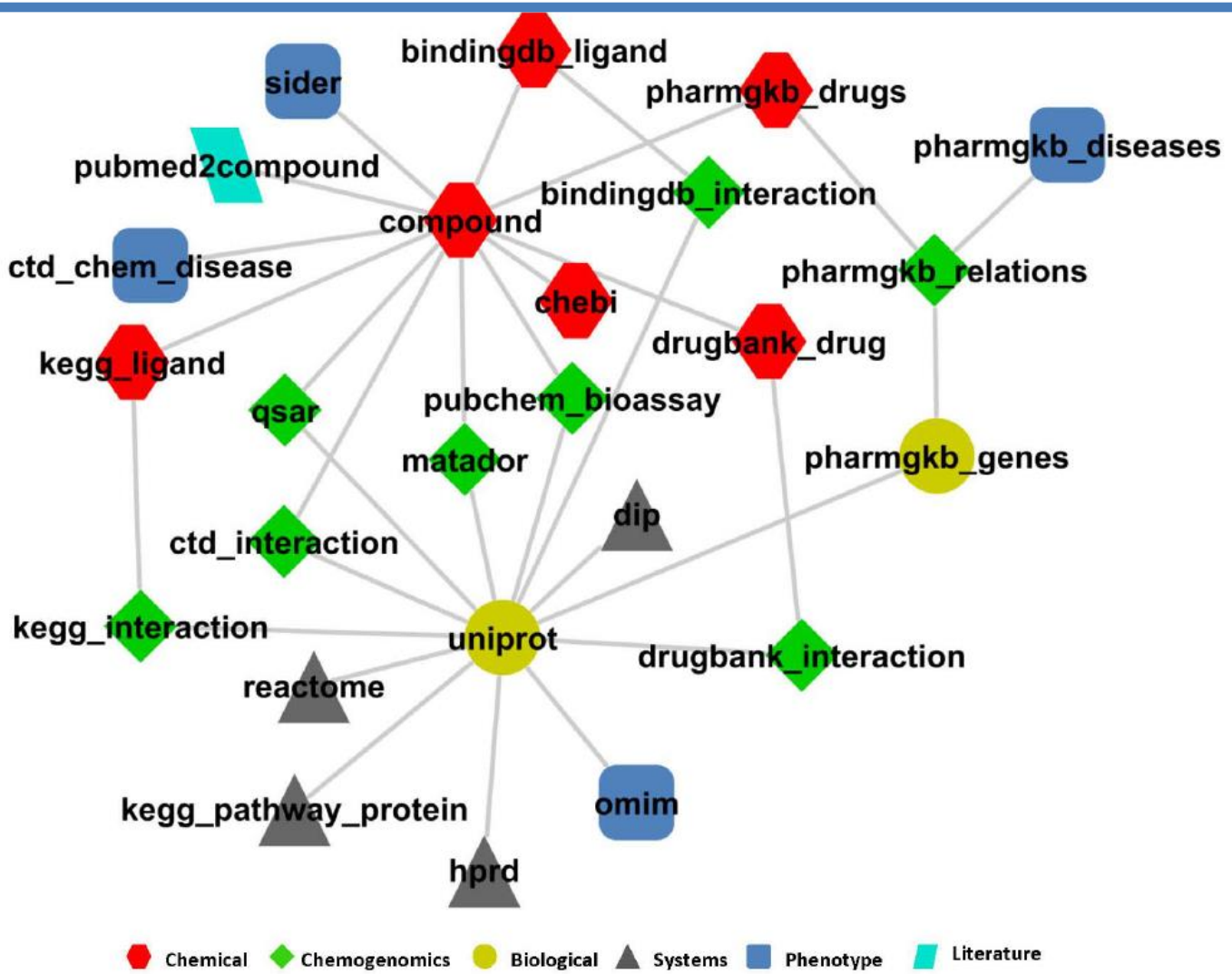


Figure 1 Chem2Bio2RDF datasets. Nodes represent data sources. Two nodes are linked if the data of one source is directed to the data of another source. The node is shaped and colored by its type, which is organized into six categories. Some databases map to multiple sources.

Chen, B.; *et al. BMC Bioinformatics* **2010**, *11*, 255

Systems chemical biology (SCB)

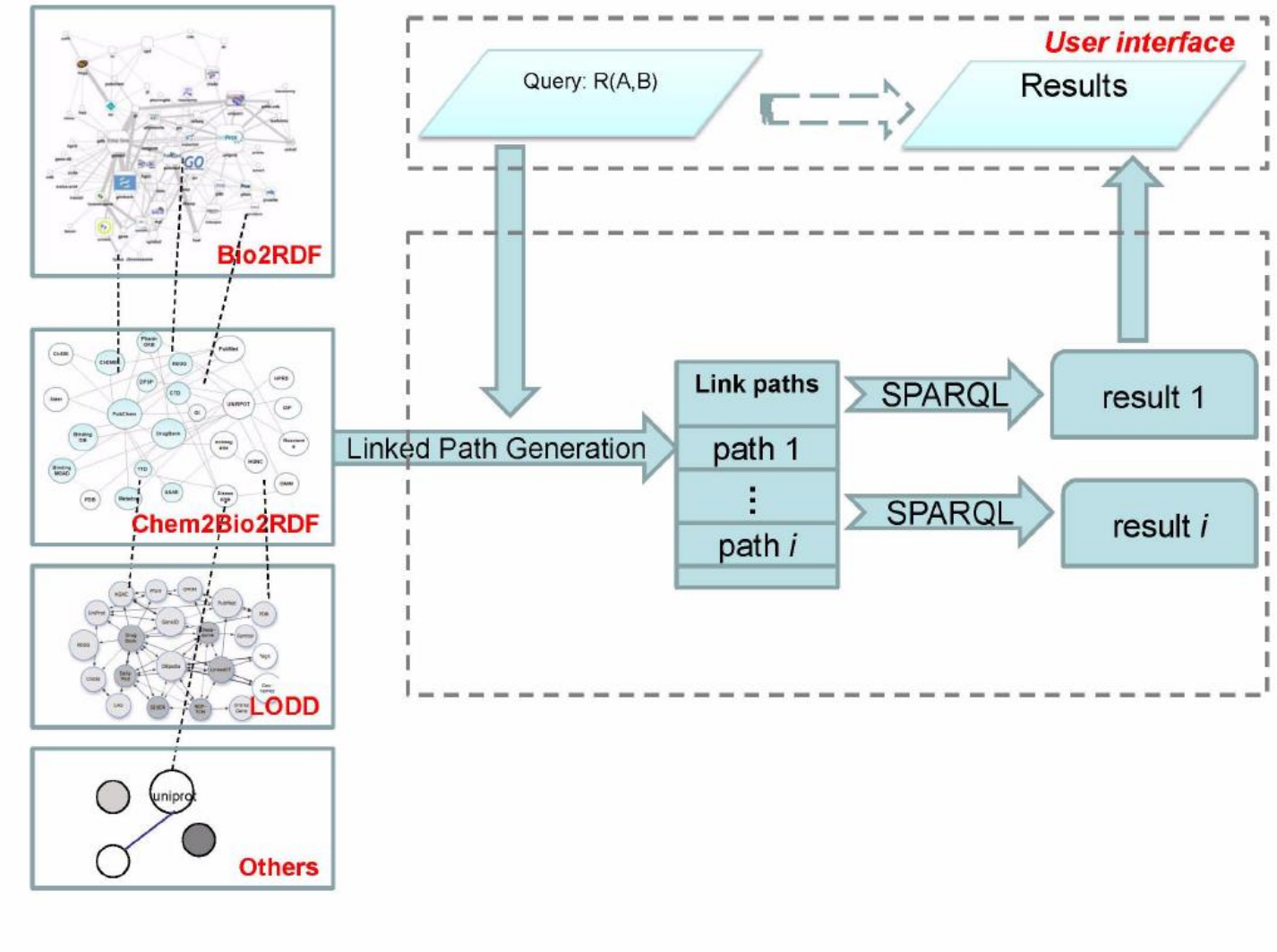


Figure 2 Chem2Bio2RDF querying architecture. Chem2Bio2RDF is linked to Bio2RDF, LODD and other RDF resources. LPG refers to prototype methods used for automatically generating links between two given objects and automated generation of SPARQL queries.

Chen, B.; *et al. BMC Bioinformatics* **2010**, *11*, 255

Systems chemical biology (SCB)

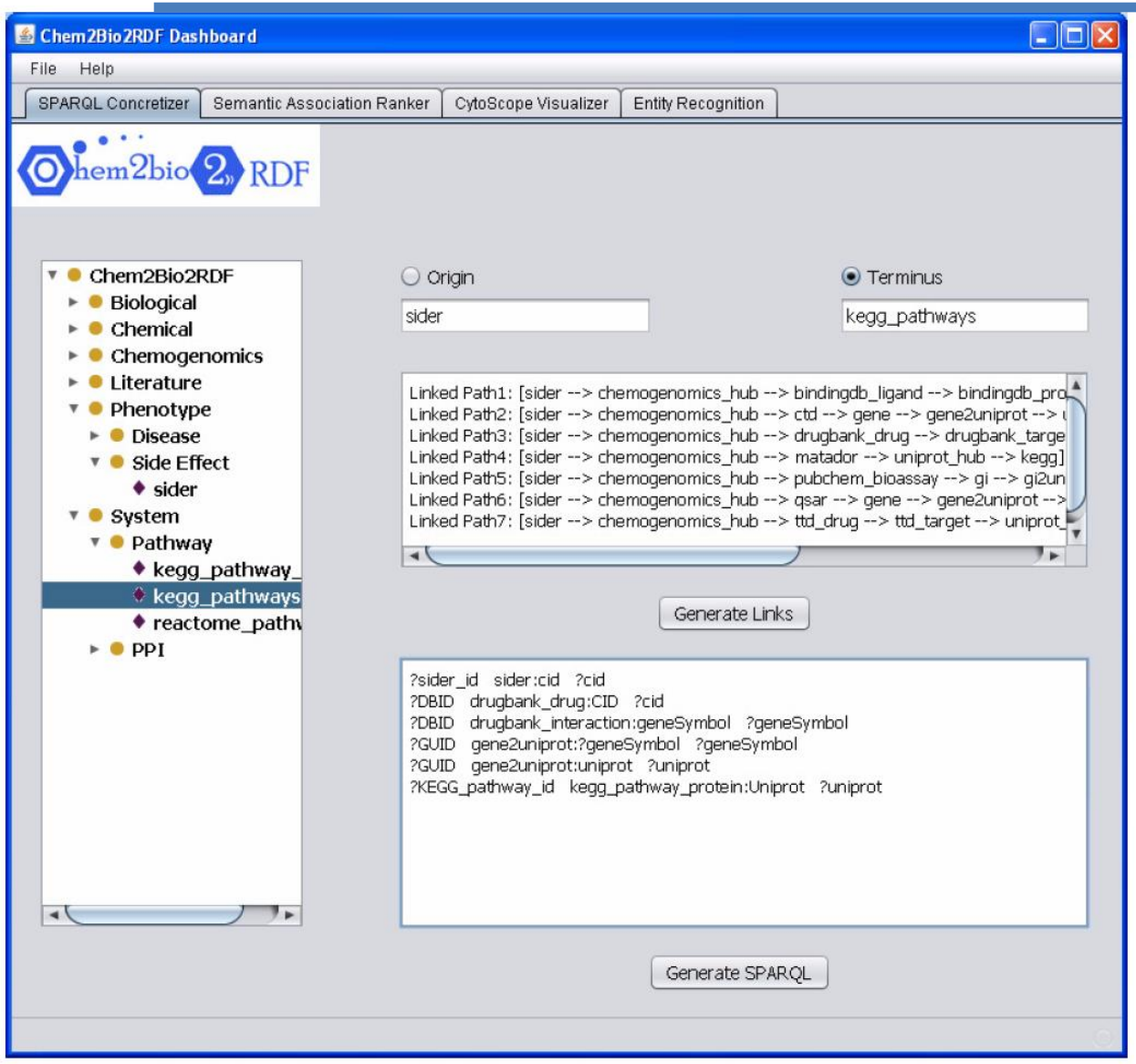


Figure 3 Prototype linked path generation. A prototype of a tool that allows users to select origin and terminal data sources. The tool will generate all the possible paths between the two data sources, will allow the user to select individual paths, and will then convert these into SPARQL queries.

Chen, B.; *et al. BMC Bioinformatics* **2010**, *11*, 255

Systems chemical biology (SCB)

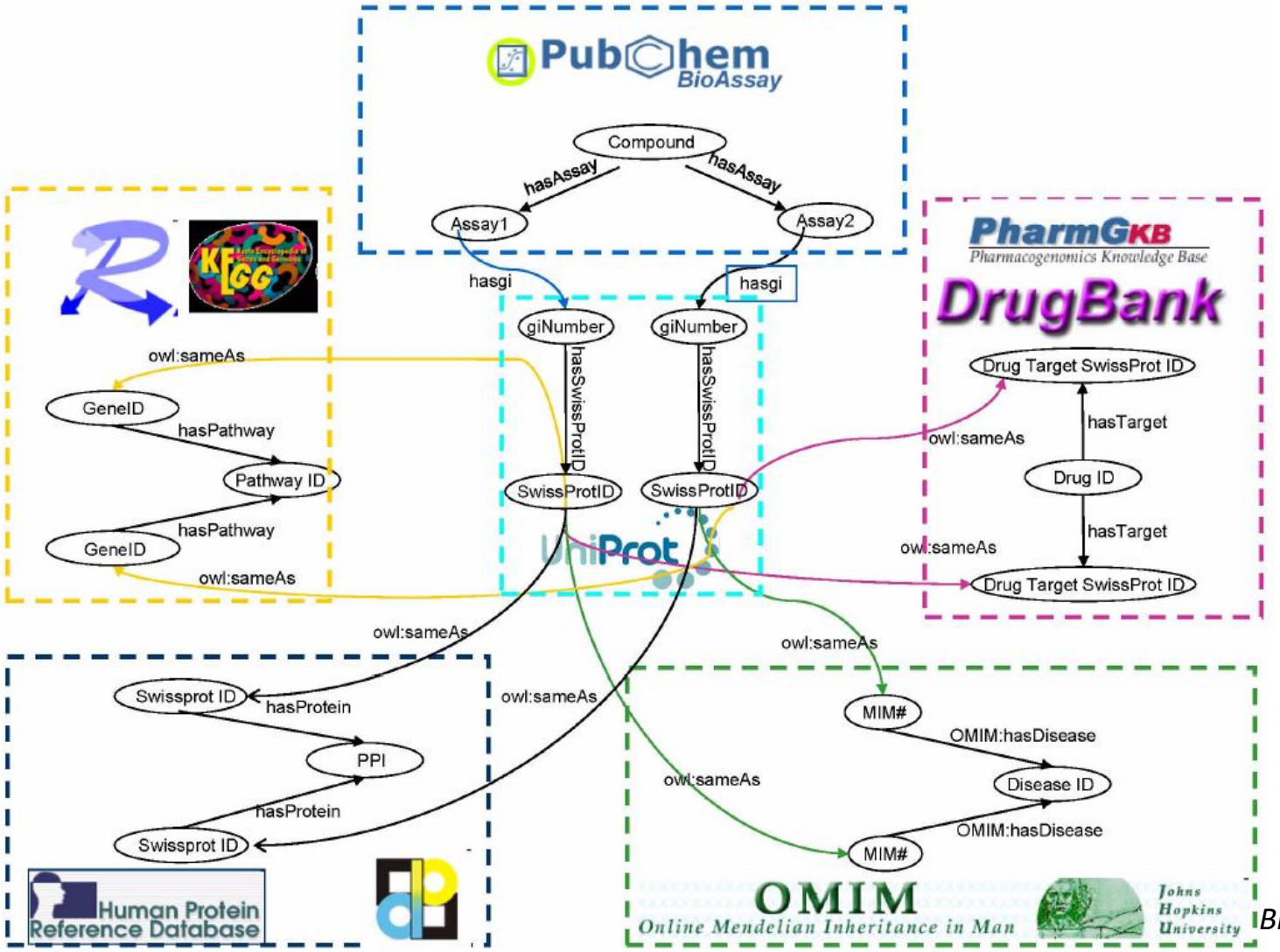


Figure 4 Class links for polypharmacology. Includes the classes: Bioassay, Drug Target, Pathway, Protein-Protein Interaction, and Disease. Some classes include more than one data source. Two nodes in different classes are linked through two paths. For instance, drug *X* is linked to compound *Y* if targets *A* and *B* of drug *X* are linked to assays *A* and *B* of compound *Y* via UNIPROT ID.

Chen, B.; *et al. BMC Bioinformatics* **2010**, *11*, 255

Systems chemical biology (SCB)

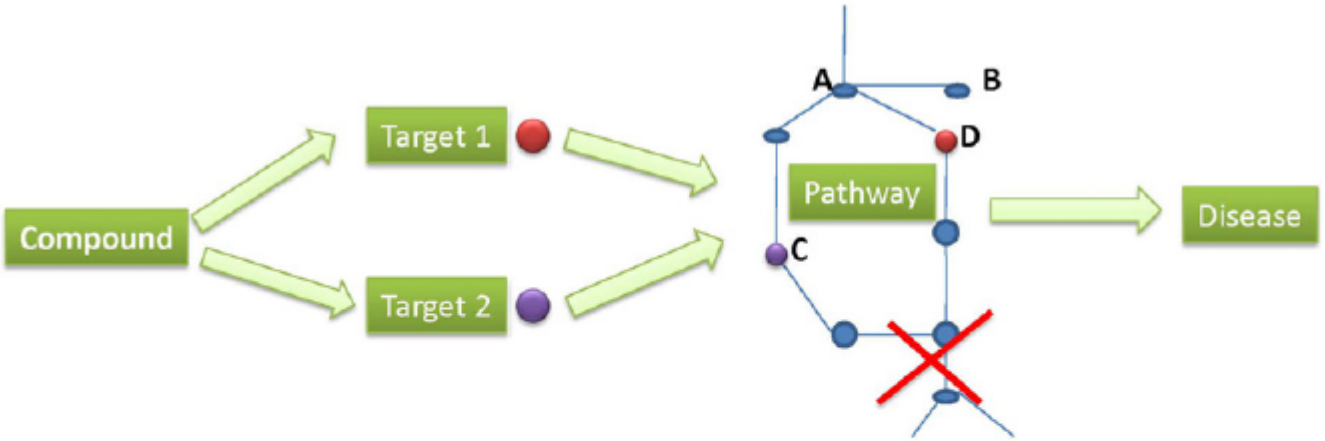


Figure 6 Illustration of polypharmacology in pathways. The compound is active against two proteins that are located in the two branches of the pathway that is associated with one disease. Targeting either node C or node D is not able to block the whole pathway.

Chen, B.; *et al. BMC Bioinformatics* **2010**, *11*, 255

Systems chemical biology (SCB)

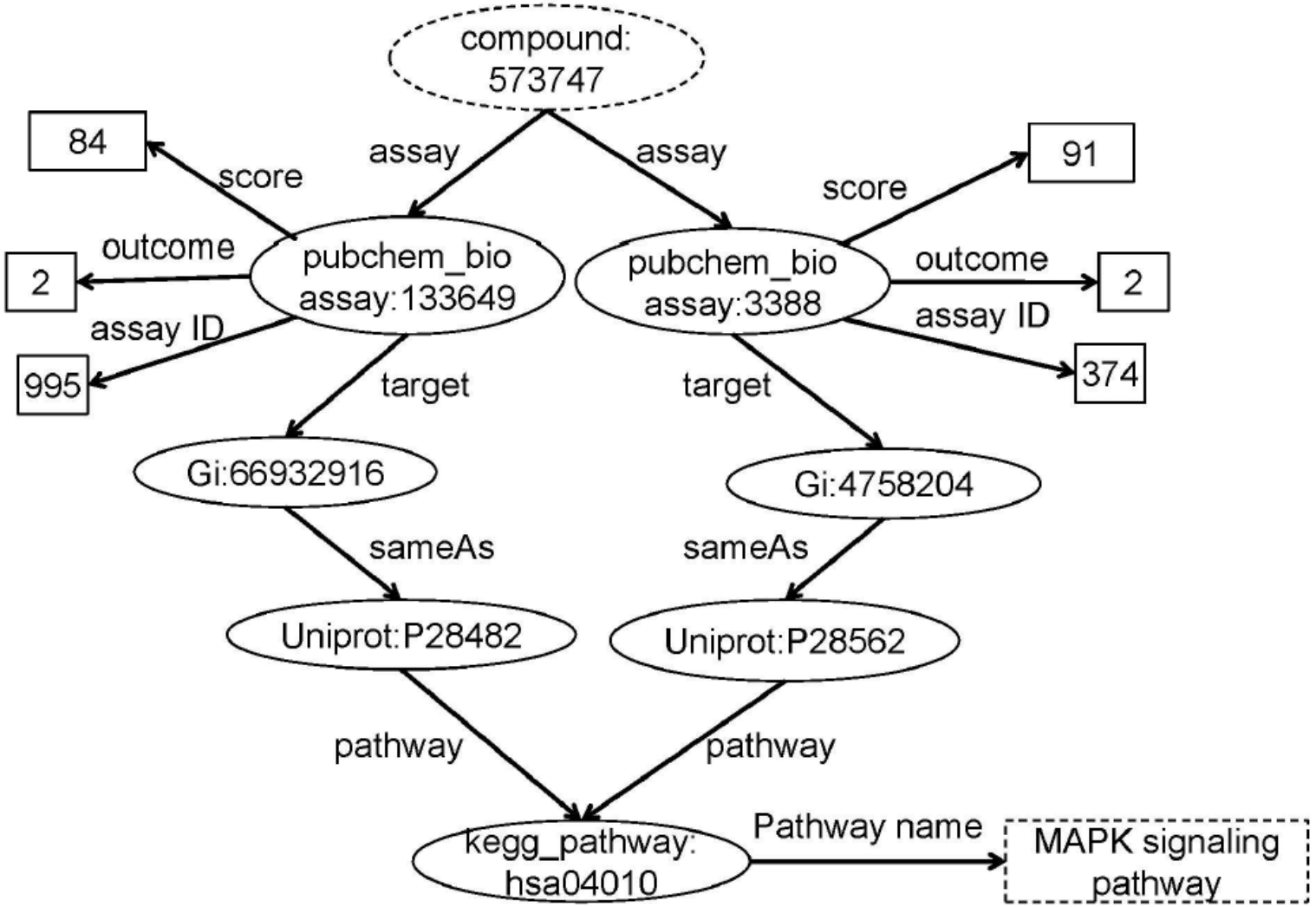


Figure 7 Graphical representation of the SPARQL query for Case Study 2. PubChem compounds (e.g. CID 573747) are identified that are active in bioassays that are associated with protein targets, which are associated with genes (via UNIPROT) which are identified as being part of the MAPK signalling pathway (via KEGG). We thus identify compounds which have multiple paths, and thus which interact with multiple targets in this protein.

Chen, B.; *et al. BMC Bioinformatics* **2010**, *11*, 255

Systems chemical biology (SCB)

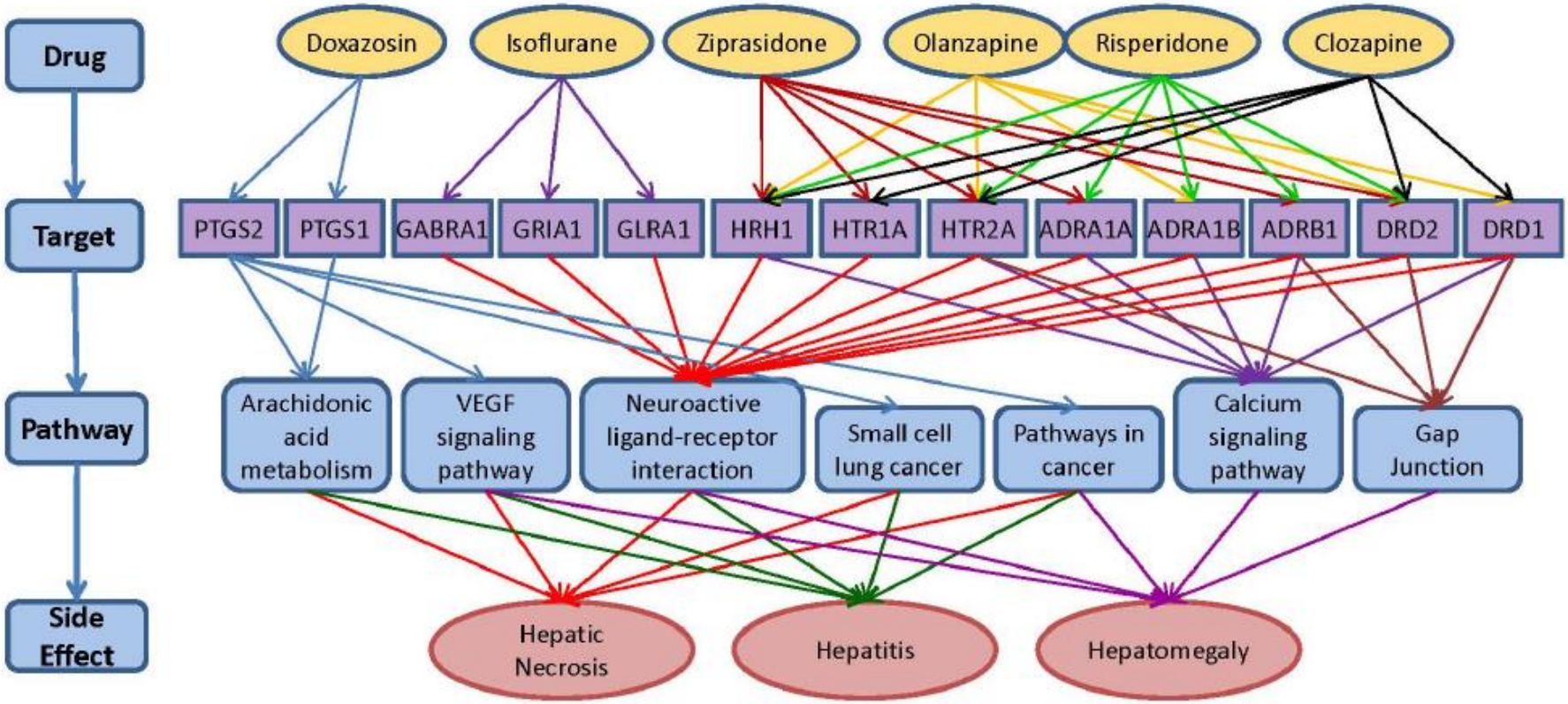
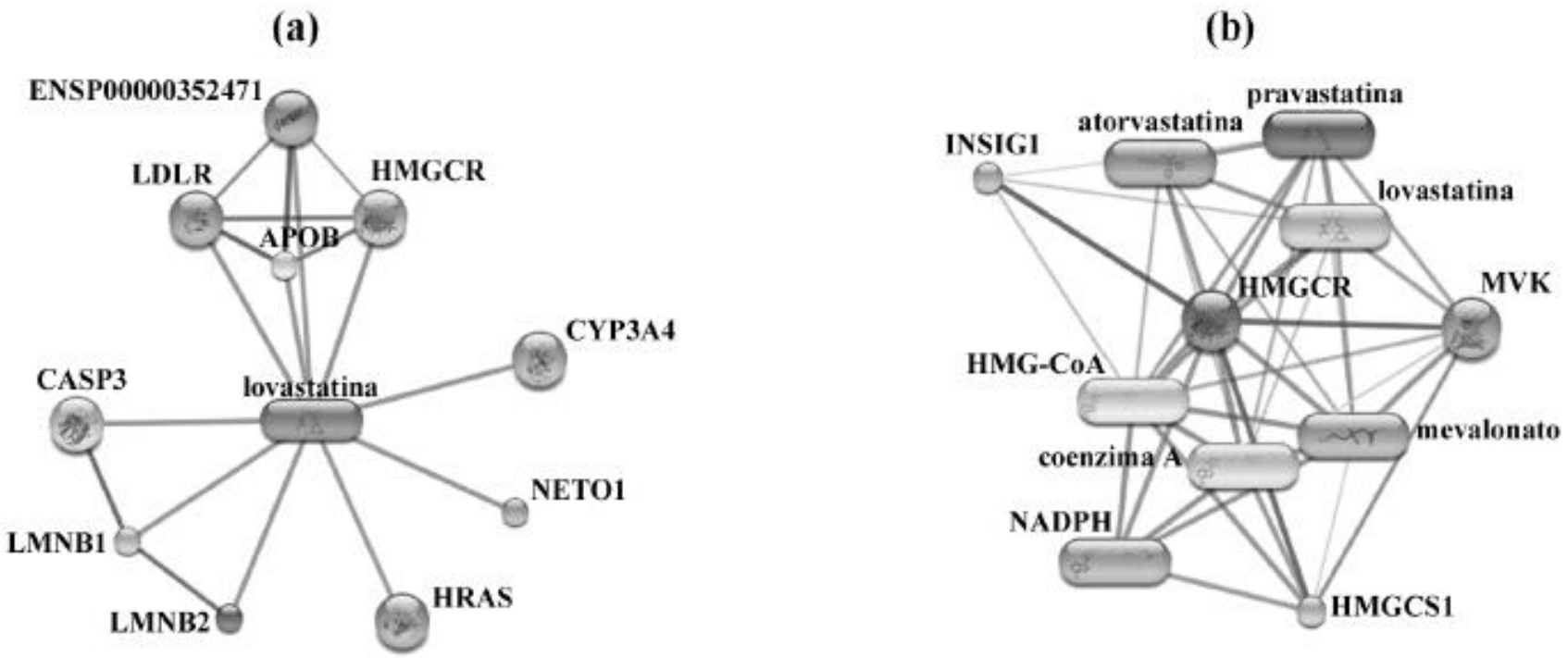


Figure 8 Associating pathways with hepatotoxic effects. The drugs that are associated with hepatotoxicity-related side effects are associated with their targets using DrugBank. The targets are associated with pathways using KEGG to establish association chains between pathways and side-effects.

Chen, B.; *et al. BMC Bioinformatics* **2010**, *11*, 255

Relational database

Example: Stitch database for lovastatin and atorvastatin



Leitão, A.; Oprea, T. I. . Integração da Química Medicinal na Biologia Química de Sistemas. In: Montanari, C.A. (Org.) Química medicinal: Métodos e fundamentos em planejamento de fármacos. 1ed.São Paulo: EDUSP, 2011, 1073-1093.

Dynamic SCB

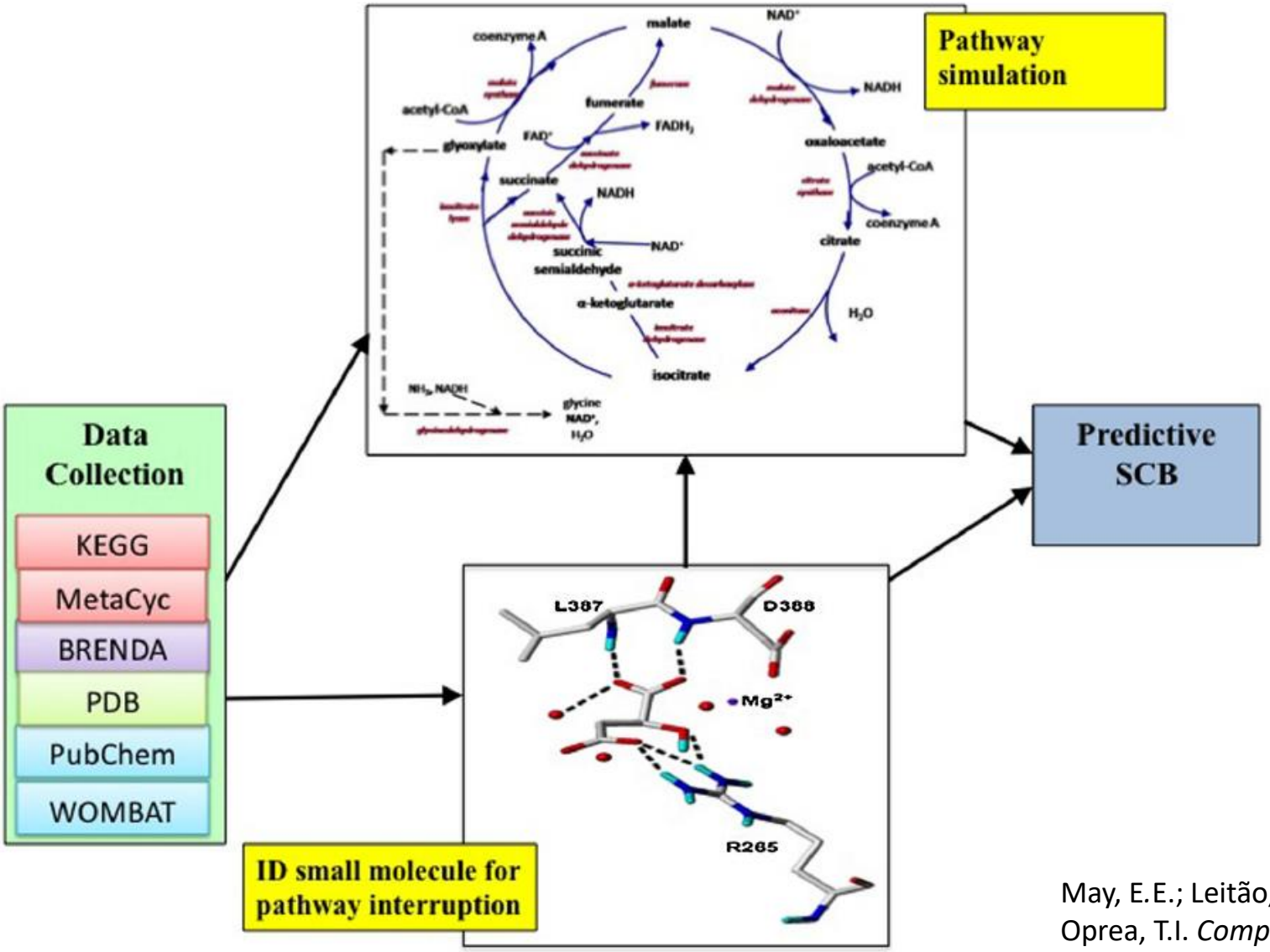
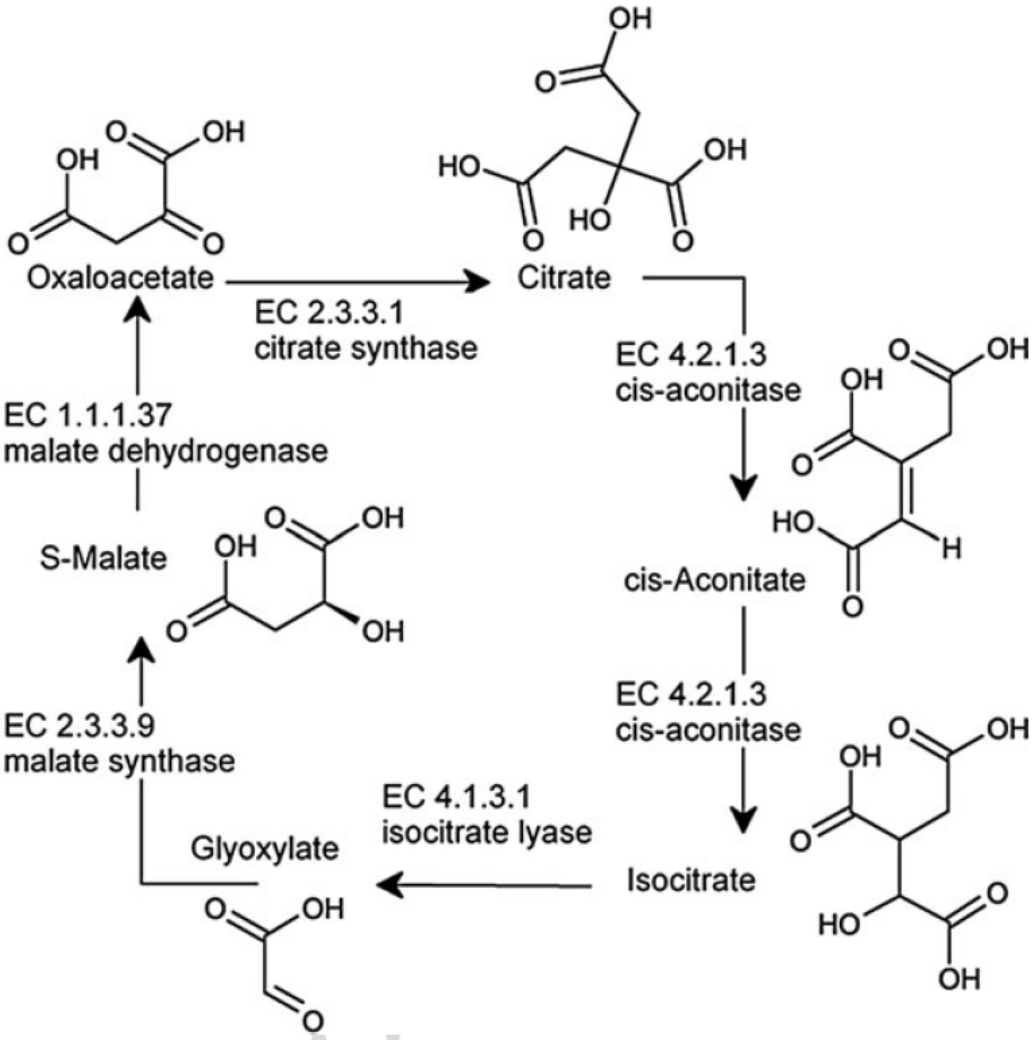
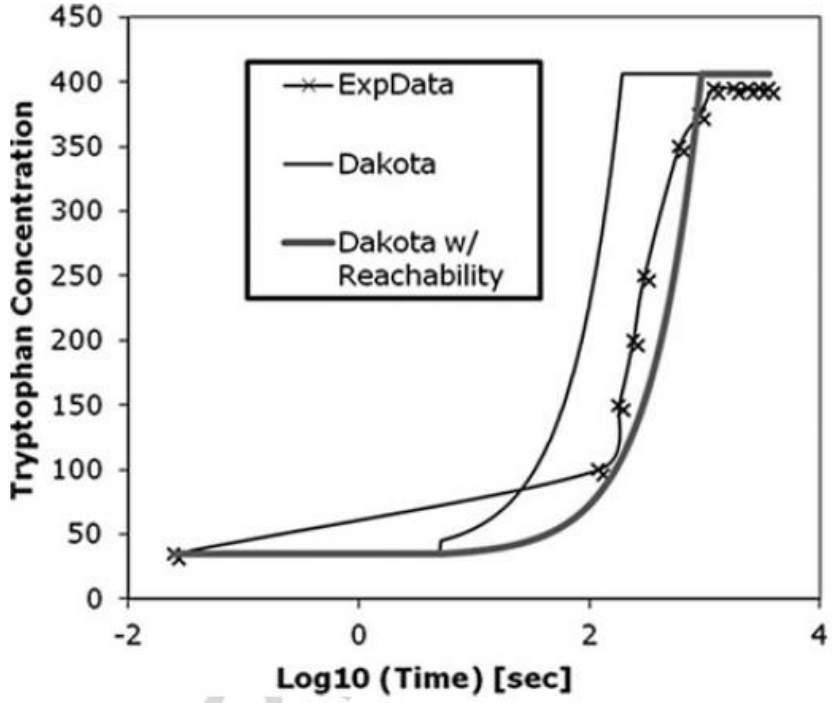


Fig. 1. Computational systems biology workflow.

May, E.E.; Leitão, A.; Tropsha, A.; Oprea, T.I. *Comput.. Biol.Chem.* **2013**, *47*, 167

Dynamic SCB

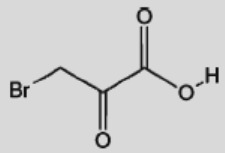
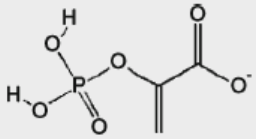
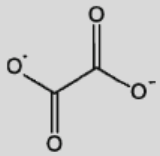
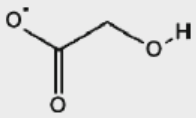




Oprea, T.I.; May, E.E.; Leitão, A.; Tropsha, A. Computational Systems Chemical Biology. In: Chemoinformatics and Computational Chemical Biology, Jürgen Bajorath (ed.) vol. 672, chapter 18, Springer, 2011

Dynamic SCB

Malate synthase ligands

Compound	Structure	K_i (μM)
Bromopyruvate (inhibitor)		60
Phosphoenol-pyruvate (weak inhibitor)		200
Oxalate (weak inhibitor)		400
Glycolate (very weak inhibitor)		900

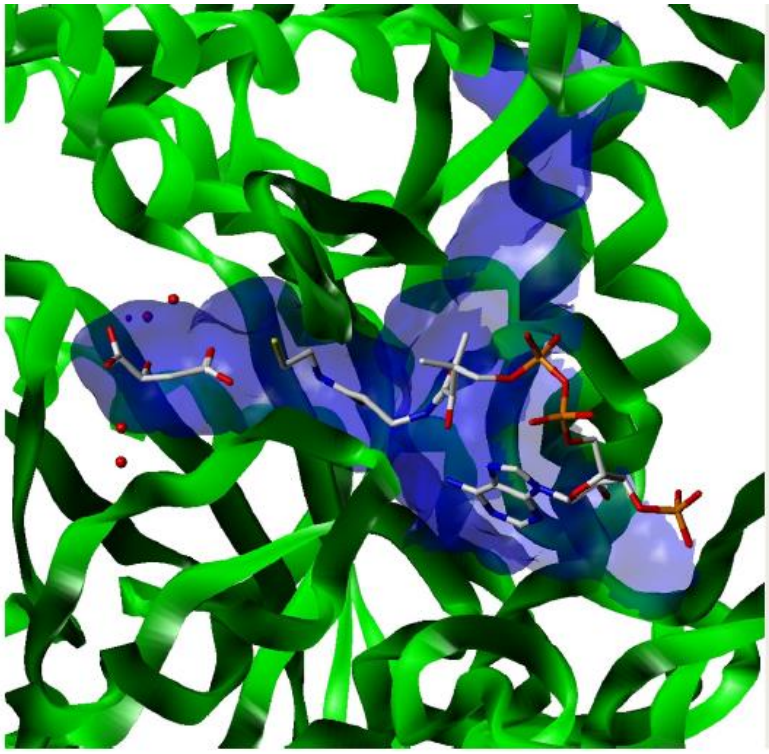
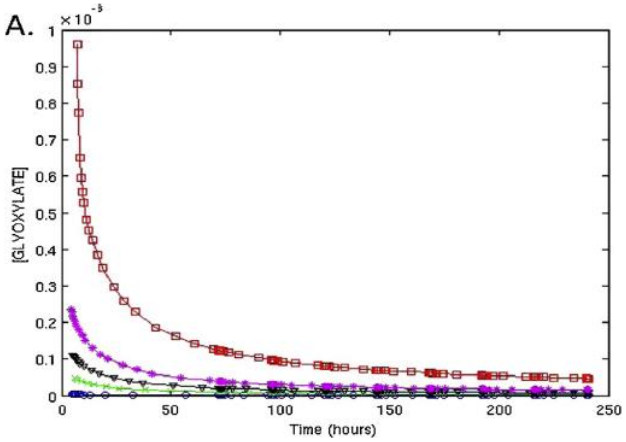
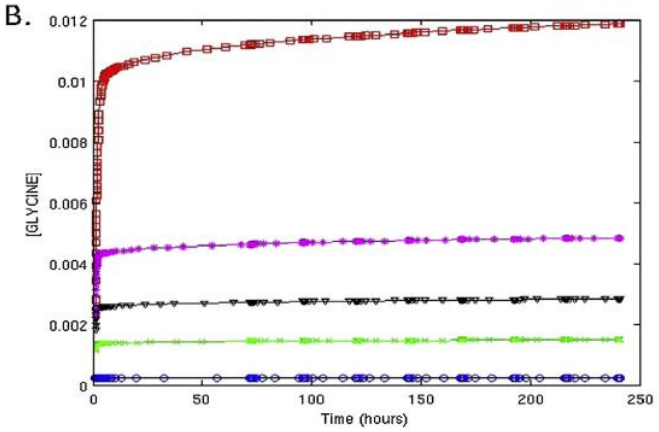


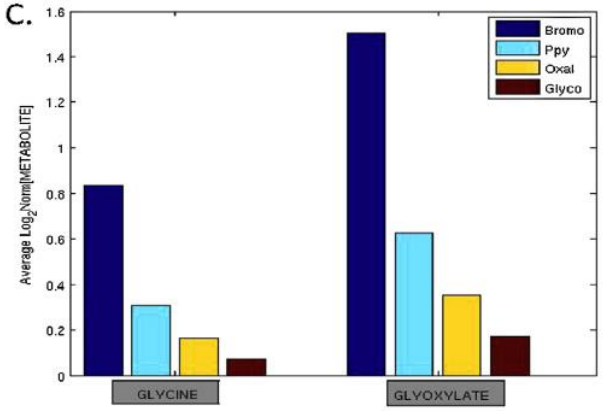
Fig. 4. Translucent view of the binding pockets surface of malate synthase showing malate (left), coenzyme A (right), water molecules and magnesium (spheres). Part of the CoA side chain is pointing outward.

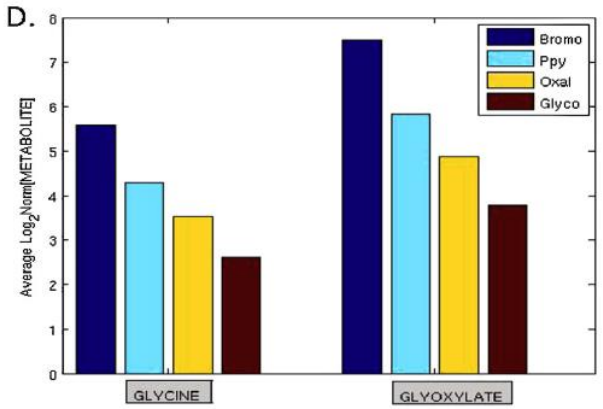
May, E.E.; Leitão, A.; Tropsha, A.; Oprea, T.I. *Comput.. Biol.Chem.* **2013**, *47*, 167

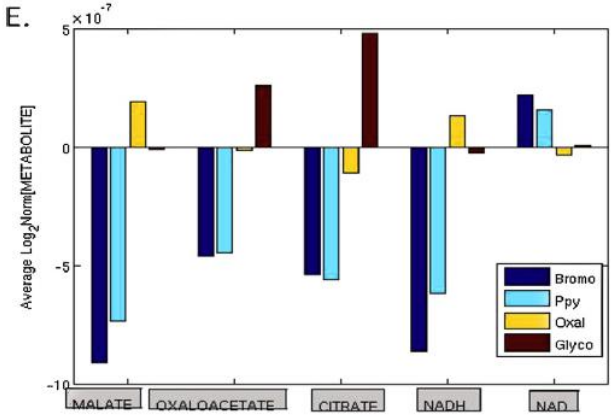
Real time analysis

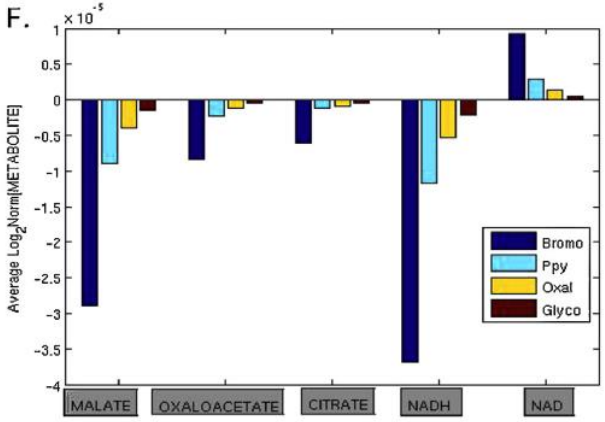








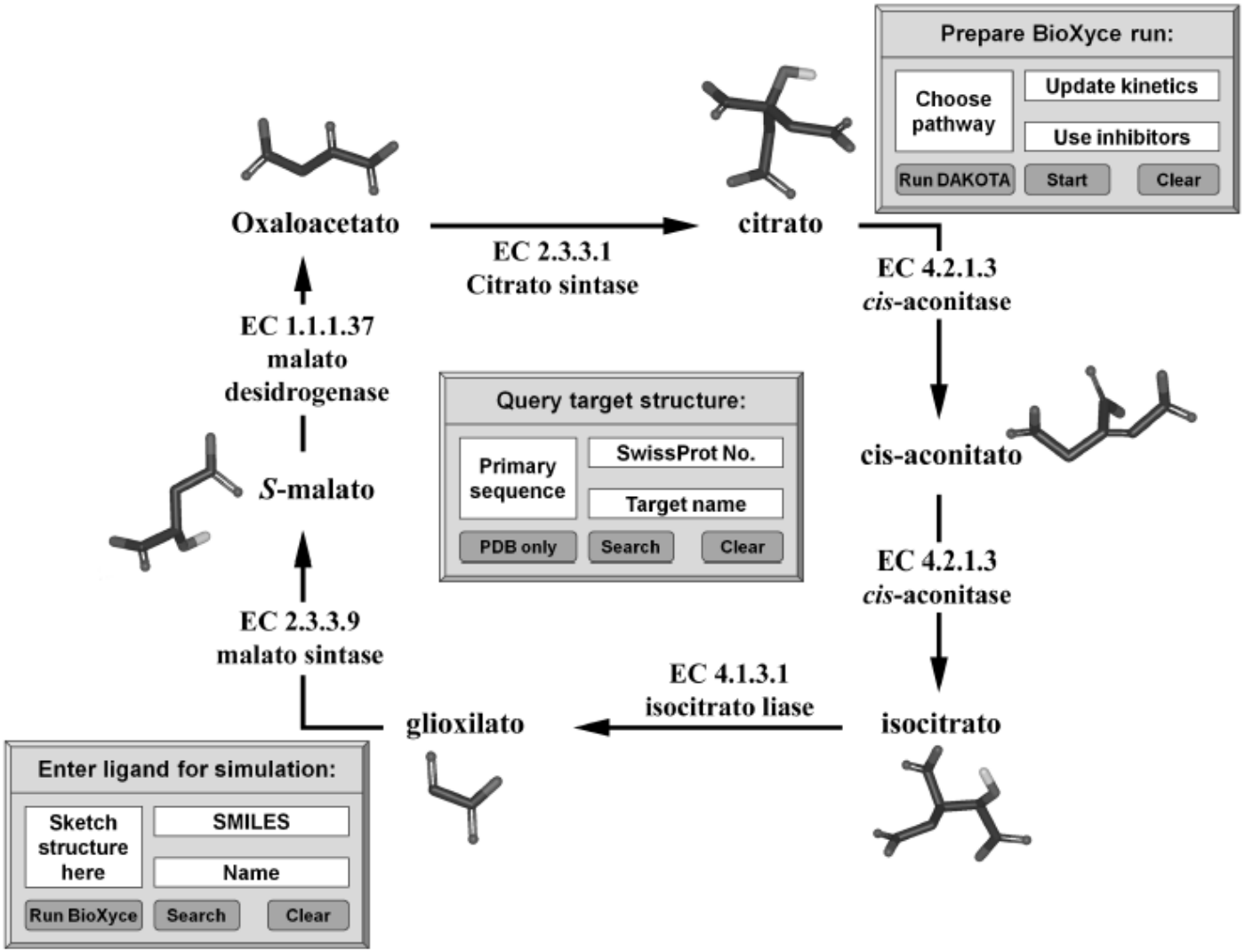




Greatest impact of MS inhibition on glyoxylate and glycine concentration for inhibitor levels at 100× the initial concentration of glyoxylate

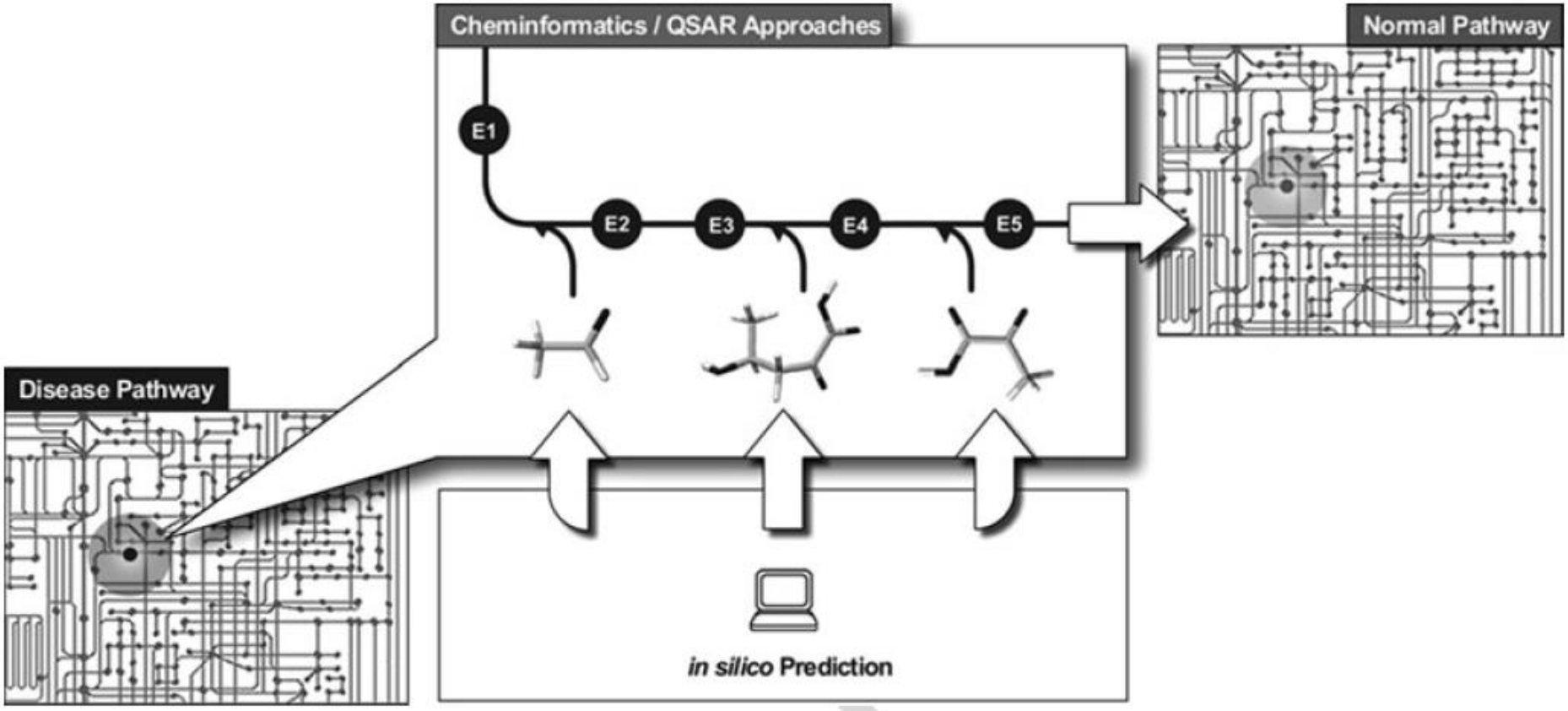
May, E.E.; Leitão, A.; Tropsha, A.; Oprea, T.I. *Comput.. Biol.Chem.* **2013**, *47*, 167

Real time analysis



Leitão, A.; Oprea, T. I. . Integração da Química Medicinal na Biologia Química de Sistemas. In: Montanari, C.A. (Org.) Química medicinal: Métodos e fundamentos em planejamento de fármacos. 1ed.São Paulo: EDUSP, 2011, 1073-1093.

Real time analysis



Oprea, T.I.; May, E.E.; Leitão, A.; Tropsha, A. Computational Systems Chemical Biology. In: Chemoinformatics and Computational Chemical Biology, Jürgen Bajorath (ed.) vol. 672, chapter 18, Springer, 2011