

## The Quantitative Human Interactome:

### *Specification of quantitative parameters for mechanistic models: Relevance for Systems Pharmacology*

#### Executive Summary

The overall goal of the Quantitative Human Interactome Project is to experimentally obtain kinetic constants for cellular interactions between all of the proteins encoded by the human genome and construct a database of these parameters. Such a database of quantitative parameters will enable the development of differential equation-based kinetic models of cellular processes. Such models are necessary for the understanding of drug action and connections between molecular events and PK, PD data. These models will also serve as a major driver for the discovery of new drugs for many complex diseases.

This is the report from the second one-day workshop held by the Systems Biology Center New York to initiate discussion on the feasibility of a Quantitative Human Interactome Project. The focus of the second workshop was on obtaining the view of expert modelers as well as graduate students and postdoctoral fellows who are currently working on biochemical kinetic modeling projects. The workshop focused on discussing what a database (the Quantitative Interaction Database, QID) for the Quantitative Human Interactome should contain and how might be organized.

It was recommended that the QID consists of three sub-databases: Concentrations, Reactions and Constrains. The **concentration** sub-database would contain the values of experimentally measured concentrations of cellular components under standard conditions. Components would be grouped together on the basis of functional modularity. Since these concentrations are likely to vary from cell type to cell type it was recommended that concentrations be measured in 5-10 cell types. In a pilot project concentrations should be measured in at least 3 cell types. It was also recommended when ever feasible concentrations of components that have been previously measured or estimated from biochemical experiments be recorded in this sub-database. These values should be clearly labeled as non-standard measurements and explicitly associated with primary publications. The **reactions** sub-database would contain both the equilibrium constants and forward and reverse rate constants for reactions between the components listed in concentration database. Enzymatic reactions should contain  $K_m$  and  $K_{cat}$  values. All of these values would be gathered under standardized experimental conditions. The reactions sub-database should explicitly deal with complexes. The reactions sub-database could also contain kinetic parameters that been previously measured or estimated. These values, like the concentrations should be clearly labeled as non-standard and associated with primary publications. The **constrains** sub-database would contain time courses and dose response curve data from biochemical and cell physiology experiments that could be used to constrain models. These data would be specified with respect to the cell type and can serve as the basis for integrating the vast regulatory biochemistry and cell physiology data in the current experimental literature with the standard measurements made in the Quantitative Human Interactome Project. Data in the QID would be SMBL compatible and exportable to standard simulation software such as MatLab and Virtual Cell.

## **SBCNY Workshop Wednesday November 19, 2008**

**held at Mount Sinai School of Medicine**

### Participants:

#### *Invited Participants*

|                  |   |
|------------------|---|
| Walter Kolch     | Beatson Institute for Cancer Research, Glasgow, UK          |
| Boris Kholodenko | Thomas Jefferson University, Philadelphia, PA, USA          |
| Dean Bottino     | Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA |

#### *SBCNY- MSSM Participants*

Ravi Iyengar  
Fernand Hayot  
Susana Neves  
Eric Sobie  
Avi Maayan  
Simon Hardy  
Azi Lipshtat  
Anamika Sarkar  
Padmini Rangamani  
Seth Berger

This workshop was held as part of the ongoing activities of the Systems Biology Center New York and coincided with the visit of Professor Walter Kolch from the Beatson Institute, University of Glasgow as Senior Visiting Scholar to the Center. This workshop continued the discussions initiated in the SBCNY Spring 2008 workshop where we focused on the feasibility of experimentally measuring levels of cellular proteins and obtaining kinetic constants for cellular interactions between all of the proteins encoded by the human genome. This workshop focused on theoretical aspects and assessed the currently available kinetic data for the development of quantitative models of signaling pathways. Approaches to organize the data in a format useful for kinetic modeling in standard modeling platforms such as MatLab, Mathematica and Virtual Cell were discussed. Currently available databases of metabolic pathways were analyzed. The workshop focused on identifying the types of kinetic data that will be most useful for the development of models with the ultimate goal of drug targets discovery and drug action analysis. The cost-benefit considerations of obtaining equilibrium constants vs. rate parameters were discussed. Specification of initial concentrations of reactants in different cell types was also discussed. Development of simple and useful databases for storage of quantitative information was considered.

### **Long-Term Goal**

The long term goal of this project is to develop a cohesive informatics approach so that the quantitative parameters that are experimentally obtained under standard conditions can be used in the most effective manner for the development of mechanistic kinetic models. The development of such models, at the level of cellular components and their interactions, should have the capability to be readily integrated with cellular and tissue/organ level physiology and

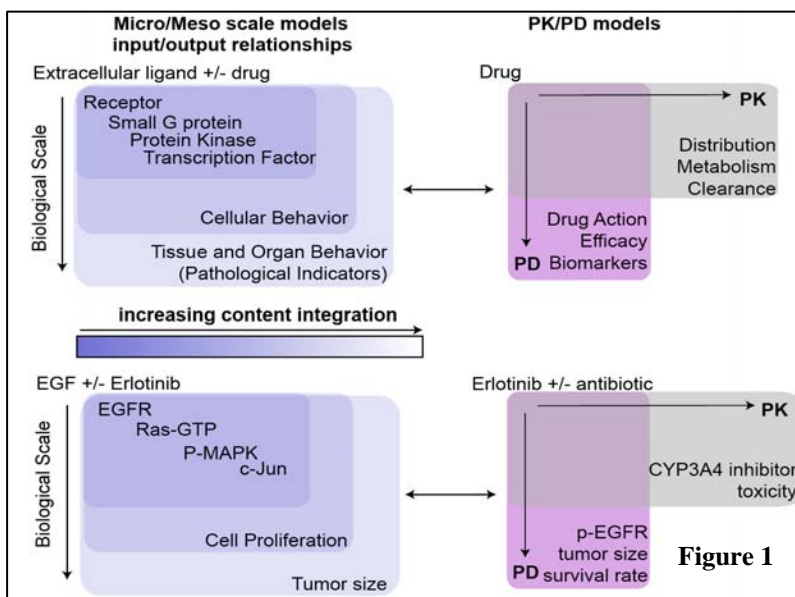
disease models.

### **Current Status and Statement of Problem**

Currently, (Dec 08) there is a need to develop an easily accessible database that lists quantitative parameters for mammalian signaling pathways components and their interactions. Although there have been some efforts in the past, development and population of a database has not been a sustained effort and nor has there been dedicated funding for the development of such a quantitative database. The lack of such databases have seriously hampered the development of cell signaling computational models that can be integrated with higher order models for studying pathophysiology and aid in drug discovery.

### **Relevance to Pharmacology and Drug Discovery**

Quantitative models of signaling pathways and networks are important for drug target discovery and understanding drug action and adverse effects. Almost all drugs act at one or more cellular component to change their activity. It is this change in activity when propagated across scales of organization results in both therapeutic and adverse effects of the drugs. Since most drugs are targeted to cell signaling molecules such as receptors, quantitative analyses of drug effects on signaling pathways are required for understanding drug action. Kinetic models of cell signaling networks, if appropriately constructed, can be seamlessly integrated into dynamic models of physiological functions at both the cellular and tissue/organismal levels. Such integration will allow for the development and mechanistic analysis of pharmacokinetic and multiscale pharmacodynamic models. This relationship is shown schematically in Figure 1, with general example in the upper panels and a specific example of the EGF receptor and a cancer drug that is being currently used. The general usefulness of the database of quantitative parameters will depend on its ability to sustain the development and analyses of multiscale models for drug interactions such as that shown in Figure 1.



**Figure 1**

### **Rationale**

The experimental determination of quantitative parameters on a genome-wide scale would not have value if this information was not available for use in building models. These models could cover a wide variety of cell types and tissues for different physiological and pathophysiological functions as well as drug action. For this the quantitative data at the level of components need to be suitably annotated and framed. Consequently the database containing the quantitative parameters

should also have information specifying how the parameters were obtained but also information of how the components function as part of cellular and tissue/organ level systems.

It was recognized that even in a genome-wide analysis, there will not be a single set of values that define a component (gene product) or interactions. For example, a gene product will have different levels in different cell types. Consequently, initial concentrations of proteins will have to be defined in a context specific manner. Similarly post-translational modifications of proteins are also likely to occur in a cell-type specific manner and in specific locations within the cell. These variations will have to be accounted for. Consequently, it was recommended that determination of quantitative kinetic parameters be done in multiple cell types. The approach being utilized by the Human Protein Atlas ([www.proteinatlas.org](http://www.proteinatlas.org)) documenting qualitative data in numerous tissues and cell types should be studied and considered for usage in this project as well.

Given the central role of cell signaling pathways in drug discovery, it was agreed upon that the project should focus on receptors, G proteins (small and heterotrimeric), scaffolds, protein kinases, phosphatases, channels, transcriptional regulators and key metabolic enzymes that are targets of kinases and phosphatases. Even this group of proteins is a sufficiently large and may well run into the thousands. For the pilot project further culling may be necessary.

## **Approach**

### **Measuring cellular concentrations of components**

Several participants in the workshop emphasized the importance of obtaining initial concentrations of proteins and other reactants, as essential for the development of quantitative kinetic models. It was emphasized that however accurate, measurement in a single cell type would not be broadly useful. It was recommended that measurements be made for a thousand proteins in 30 human cell lines, used widely in academia and in pharmaceutical industry for preclinical testing of drug candidates. If this was not immediately possible, then measurements should be made for 100-200 proteins in at least 5-10 human cells from different tissues. It was recommended that that this phase of the project be coordinated with the Human Protein Atlas Project since that project has a large number of antibodies that could potentially be used for quantitative immunoblotting. Criteria for selecting proteins for the pilot project could use pathways and networks involved in specific diseases such as pancreatic cancer or Type 2 diabetes for which there is an urgent need for new drugs. It was recommended that the initial focus be on developing rigorous ODE models and consequently it would be best to determine whole cell concentrations without focusing on specific subcellular localization.

In addition to protein components it will be necessary to obtain cellular concentrations of key lipids, nucleotides and ions. For this a metabolomic approach will be required. Interactions with the LipidMAPS consortium and other metabolomics groups could be useful for this purpose and could be pursued at the appropriate time.

## Measuring Kinetic Constants

The workshop participants agreed that both kinetic and equilibrium measurements be obtained. At this time the most feasible approach for these measurements involves the use of purified proteins. It was agreed that these numbers needed to be obtained. The kinetic parameters obtained in vitro should over a period of time be verified in the intact cells. From a modeling standpoint it was felt that the kinetics interaction parameters ( $k_{\text{on}}$  and  $k_{\text{off}}$ ) of the more scarce molecules are more important than those of larger abundance. For most interactions, obtaining  $K_d$  will be sufficient to build first generation models. When obtaining experimental measurements, care must be taken to explicitly specify how parameters are collected for multi-component complexes (e.g. parameters for A binding to B and AB binding to C). Another area where explicit specification is essential is how different species of the same protein (phosphorylated vs. non-phosphorylated) interact with their respective partners.

When measuring enzyme activities, simple  $K_m$  and  $V_{\text{max}}$  measurements are insufficient for the development of kinetic models.  $V_{\text{max}}$  measurements in vitro often lack specification of enzyme concentration thus making the determination of  $K_{\text{cat}}$  not feasible. For modeling purposes it is important that  $K_{\text{cat}}$  be determined for all of the enzymes in the model in a standard fashion.

## Relationships between biochemical reactions and physiological events

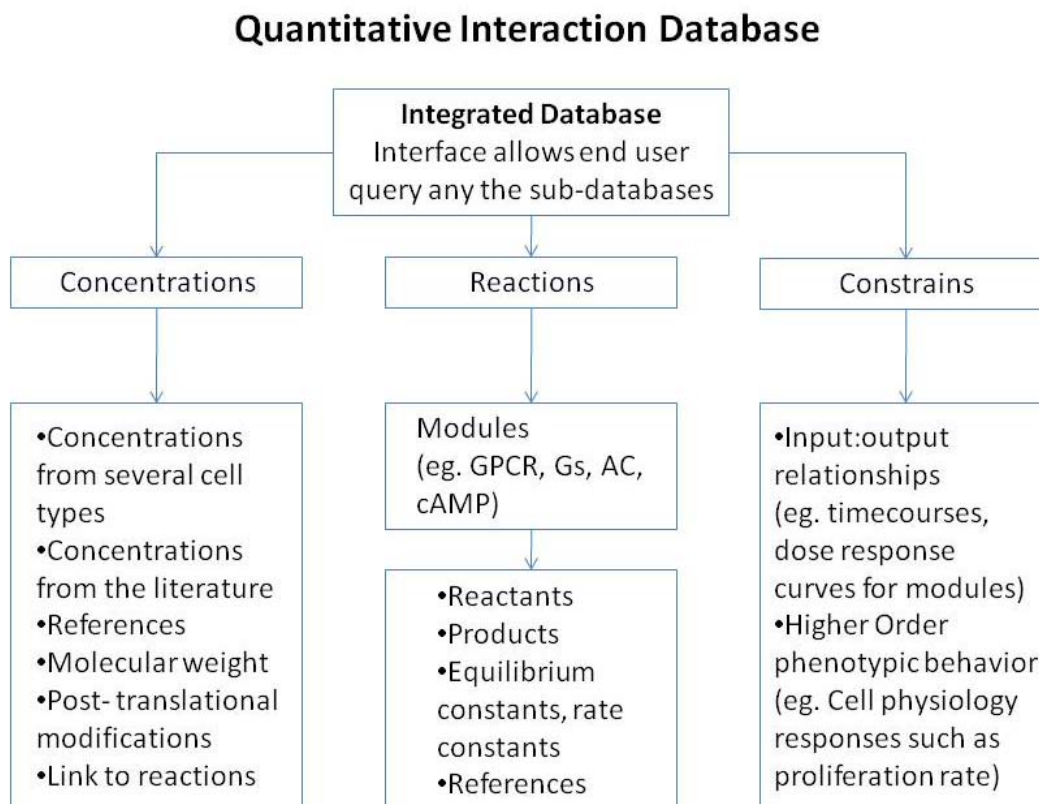
The usefulness of a project to obtain biochemical kinetic parameters and building a database to develop quantitative models of signaling pathways to aid in drug discovery will rest in the ability of such models to be integrated with measurement of physiological functions. For such integration, it will be essential to develop connections between events that occur in the seconds and minute time scale to those that occur in the hours and days time scale. For example, for a study of proliferation, the modeling of the MAP-kinase pathway would need to be connected to time courses of activation of the transcription factor Elk, to time courses of production and degradation of cyclins to time course of entry and exit of the various phases of the cell cycle. Such data is often available in the primary cell biology and physiology literature and should be incorporated in a machine readable format into the database.

## Database Organization: A Quantitative Interaction Database

All participants agreed that a database must be user-friendly, accessible to non-experts users, such as experimental biologists who may occasionally develop kinetic models to drive their experimental design. This situation would be similar to experimental biochemists and molecular biologists who look at 3D crystal structures to decide where to target mutations in structure function studies.

The database should contain different levels of information necessary to completely understand a published kinetic model. Also, extraction of relevant information from the database must be straight-forward for incorporation into future models. The database design must also be expandable as more computational models are made available in the public-domain. The NIST chemistry webbook (<http://webbook.nist.gov>) is an excellent example of a chemical reaction information database.

There are different levels of information that need to be incorporated into a database – concentrations, reactions, validation constraints. The overall organization of the database can be envisioned to be as shown in Figure 2.



**Figure 2: Schematic of the Quantitative Interaction Database (QID)**

**The concentration sub-database** will contain the concentration of the components obtained experimentally in a standardized project and also concentration from purification experiments in the literature along with the appropriate reference.

- Each component must be annotated with Gene ID and other standard identifiers. For data from the literature, the particular experimental procedure specific to this component should be identified along with the cell/tissue source. These concentrations can be used as initial concentrations in model development.
- The concentration level for each component in multiple cell types should be given.
- Data from both the standardized project and numbers estimated or guesstimated from the experimental and modeling literature should be provided and clearly labeled.

**The reactions sub-database** will contain the biochemical interactions along with the kinetic parameters associated with the reaction.

- The reactions could be arranged in a modular fashion following well described pathways and regulatory loops within the pathways. For example, a module would include all the reactions from ligand binding to GPCRs, Gs activation to cAMP production.
- Reactions should take into account the progressive formation and dissociation of macromolecular complexes. Complexes should have unique identifiers and be associated with individual components (and their corresponding gene IDs). Reactions within complexes should be specified at the level individual components (if AB complex interacts with C, is it thru binding to A or B or both?).
- Both data from experiments conducted in the standardized project and rates estimated from experiments in the primary literature would be included in the sub-database in a clearly identified manner.
- 

**The constrains sub-database** will contain multi-scale input-output constraints that relate model output to phenotypic behavior. The phenotypic behavior will be broadly classified under types of cellular behaviors such as action potential generation, proliferation, motility and apoptosis. Often such constraint data take the form of time courses or dose response curves. One example that was discussed was the development of qualitative models that relate model output for e.g. time course of the concentration profile of cyclin D to proliferation. For such model, the time course of MAPK-kinase and Elk activation could serve as constrains for the initial signaling model. The qualitative models need to be rooted in biological observations while providing insight into global cellular behavior. Also, the input-output relationships listed in the validation database will provide model developments constraints for new models. All workshop participants agreed that this will be a valuable resource for the generation of new models.

The major concern regarding such a database is user-friendliness and accessibility. All members agreed that this database should be better than previously developed databases. The database must also be compatible with current model exchange standards such as SBML. It was decided that the best way to approach this was to build a test database and survey potential users in varied academic and drug industry setting to obtain input on what would be of most use to them.

## **Pilot Project**

The pilot project should be a database that

- shows the layout of how a Quantitative Interaction Database would display the data from the standardized project and data from the experimental and modeling literature.
- demonstrates how the truth-table sub-database can be used to develop the model constraints.
- shows how the data can be exported for model development and numerical simulations in MatLab and *Virtual Cell*.

## Outreach

In developing the experimental and bioinformatics components of the project, it is essential there be continuing outreach so that the database is useful in a contemporaneous manner as well as have the capability to evolve as new types of data become available. Experts who could be useful in providing further advice include Marc Vidal, Douglas Lauffenberger, Lewis Cantley, for basic biochemical models, Bill Jusko and David D'Argenio for interface with PK, PD models and Eric Schadt , David DeGraaf , John Reynders, Adriano Henney from the pharmaceutical industry.

## Currently Available Quantitative Databases

### *Databases of kinetic parameters:*

|        |   |
|--------|---|
| KDBI   | <a href="http://bioinf.xmu.edu.cn/databases/kdbi/kdbi.php">http://bioinf.xmu.edu.cn/databases/kdbi/kdbi.php</a> |
| BIND   | <a href="http://bond.unleashedinformatics.com/Action">http://bond.unleashedinformatics.com/Action</a>           |
| BRENDA | <a href="http://www.brenda-enzymes.info/">http://www.brenda-enzymes.info/</a>                                   |
| Kmed   | <a href="http://sysbio.molgen.mpg.de/KMedDB">http://sysbio.molgen.mpg.de/KMedDB</a>                             |

### *Databases of kinetic models:*

|           |   |
|-----------|---|
| BioModels | <a href="http://www.ebi.ac.uk/biomodels-main/static-pages.do?page=home">http://www.ebi.ac.uk/biomodels-main/static-pages.do?page=home</a> |
| DOQCS     | <a href="http://doqcs.ncbs.res.in/">http://doqcs.ncbs.res.in/</a>   |
| JWS       | <a href="http://jjj.biochem.sun.ac.za/">http://jjj.biochem.sun.ac.za/</a>   |
| SABIO-RK  | <a href="http://sabio.villa-bosch.de/">http://sabio.villa-bosch.de/</a>   |