# Pharmacophore-Based Techniques
# For the Construction of Biochemical Reaction Networks

Michael Binns and Constantinos Theodoropoulos

School of Chemical Engineering and Analytical Science,
University of Manchester, Sackville St, Manchester, M60 1QD, United Kingdom

In silico models for biochemical systems are usually built by extracting in vitro kinetic expressions for each enzyme-catalysed reaction from databases. This method, however, assumes that the enzyme kinetics measured in isolation will be the same as those in complex systems (in vivo). Also, these models require that all the kinetic expressions and all the reactions are known. If this knowledge is missing and there are reactions occurring which are not considered or which have no kinetic data then the models built will be inaccurate.

Known reactions can be found from databases for organisms such as E. coli from which pathways are built. However these databases do not contain all possible reactions and the unknown reactions can form new pathways or fill gaps in existing pathways. The objective of this work is to construct an automated sequence of methodologies for generating all the feasible enzymatic reactions which can occur in a biological system. New and known reactions are generated simultaneously to give more complete reaction networks. Kinetics are then computed using parameter estimation techniques in conjunction with model reduction technologies.

The procedure for generating biochemical reaction networks involves extracting binding information from known enzyme reactions. This is achieved by using a graph theory-based method to compare the known binding species and hence to find the corresponding pharmacophore (structure and functional groups necessary for the binding). New binding species are then be obtained by using graph theory to search for alternative species which contain the pharmacophore. The structure comparisons necessary for this process involve a 2-D method [1] which only considers atoms and bonds. Quantum mechanical computations [2] are then employed to compute 3-D molecular and electronic structures. A 3-D screening method is employed next that uses flexible structures and therefore can accommodate a large number of conformations [3] modified appropriately to eliminate infeasible molecules and to use the 3-D coordinates of the atoms involved. This 3-D screening is important because a 2-D comparison method will identify many molecular structures which can not fit into the physical structure of the relevant binding site.

New reactions are then constructed using a biochemical reaction generation procedure based on combinatorial chemistry to find all possible reactions which fit the stoichiometries given by the locations of the binding sites. These new reactions are tested for feasibility ensuring the numbers and types of atoms are conserved. In addition only linearly independent sets of reactions are considered.

The produced sets of biochemical reactions are then analysed through pathway construction which involves generating combinations of reactions subject to constraints. These pathways show the capabilities of the network, i.e. whether it can perform certain conversions, which reactions are important and how it can be modified through flux analysis to enhance its performance (increasing yields, removing feedback inhibition etc.). Pathway construction is performed by selecting raw materials and desired products, computing all possible pathways between them, excluding redundant and cyclic pathways [4]. The relative importance of different metabolites is determined by comparing the number of pathways each metabolite is involved in. Furthermore, the role of metabolites in the pathways is investigated, i.e. if they are consumed, produced or used as

intermediates. Pathway construction is a non-polynomial problem, thus the computational effort required finding all pathways increases exponentially with the number of reactions. However an efficient way of generating pathways involving the use of appropriately adapted P-graphs [4] makes this possible for reasonably large reaction networks. Finally kinetic expressions are obtained using parameter estimation so that dynamic sensitivity analysis of the system can be performed. Parameter estimation starts from a set of possible kinetic expressions which are derived from the stoichiometry and possible inhibitors for each reaction. The important kinetic parameters are then identified through model reduction (e.g.[5]) exploiting the time-scale separation the system exhibits, and computing the "slow" dynamics of the system. A subsequent fitting procedure is used to match the system dynamics with experimental observations. We believe that with sufficient computation power this methodology could be extended to build models for entire cells.

## REFERENCES


1 Hattori, M., Okuno, Y., Goto, S., Kanehisa, M., "Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways", Journal of the American Chemical Society, (2003).

2 Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel et al. Gaussian, Inc., Wallingford CT, 2004

3 Raymond J.W and Willet P, "Similarity searching in Databases of flexible 3- structures using smoothed bounded distance matrices", J. Chem. Inf. Comput. Sci. 43: 908-916 (2003).

4 Fan, L.T., Bertok, B., Friedler, F., "A graph-theoretic method to identify candidate mechanisms for deriving the rate law of a catalytic reaction", Computers and chemistry, 26: 265-292, (2002).

5 Maas U and Pope, "Simplifying chemical kinetics: intrinsic low-dimensional manifolds in composition space", Combust. & Flame 88: 239-264 (1992).