MANCHESTER 1824

The University of Manchester

# 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, Manchester, M60 1QD, UK

MANCHESTER 1824

The University of Manchester

# Objectives

**Using Automated methods to construct reaction networks, searching for new reactions.**

**Generating new pathways through large reaction networks**

**Obtaining kinetic expressions and fitting parameters using reduced models**

# Applications

- Improving Drug Design Options
- Control Flux Analysis
- Improving Yields in Bioreactors

- Designing New Processes

- Determine the effect of New drugs
- Explore Genetic Modification and Diseases
- Adding New Capabilities to a Process

MANCHESTER 1824

The University of Manchester

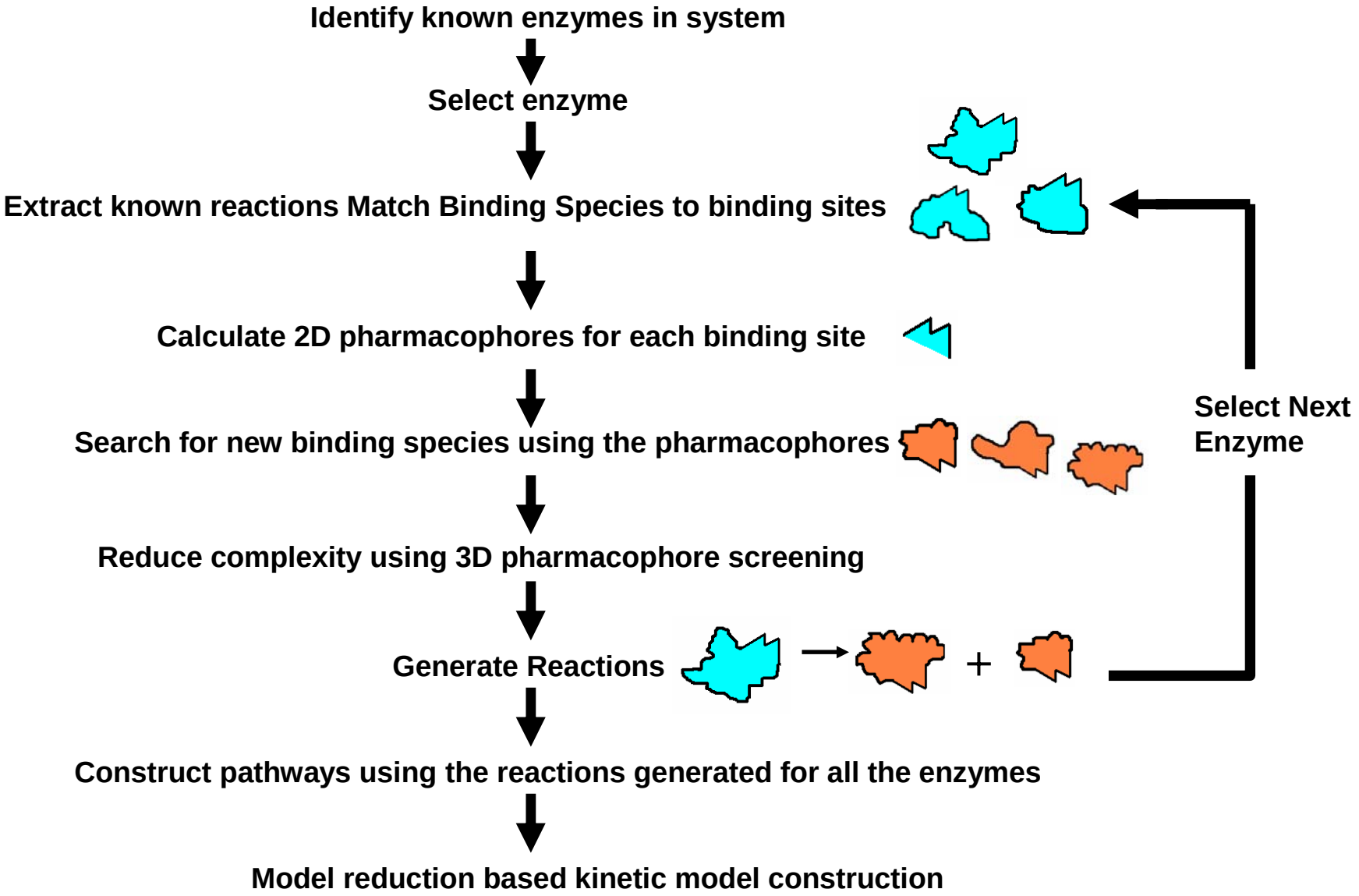
# Metabolic Network Development: Main Ideas


- **Initially assume that all enzymes are known**
- **Treat enzymes as black boxes**
- **Compare ligands which bind to the same sites**
- **Use this information to predict**
    - **new binding species**
    - **new reactions**

MANCHESTER 1824

The University of Manchester

# Metabolic Network Development: A knowledge-based method for generating Biochemical reaction networks.



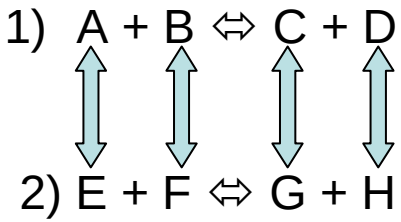
# Matching Binding Species to Binding Sites

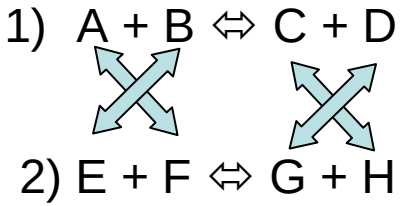
An example enzyme catalyses the following reactions

1) A + B ⇔ C + D

2) E + F ⇔ G + H

A comparison (Using Simcomp[1]) of the species in reaction 1 with those in reaction 2 should reveal which species bind to which sites.

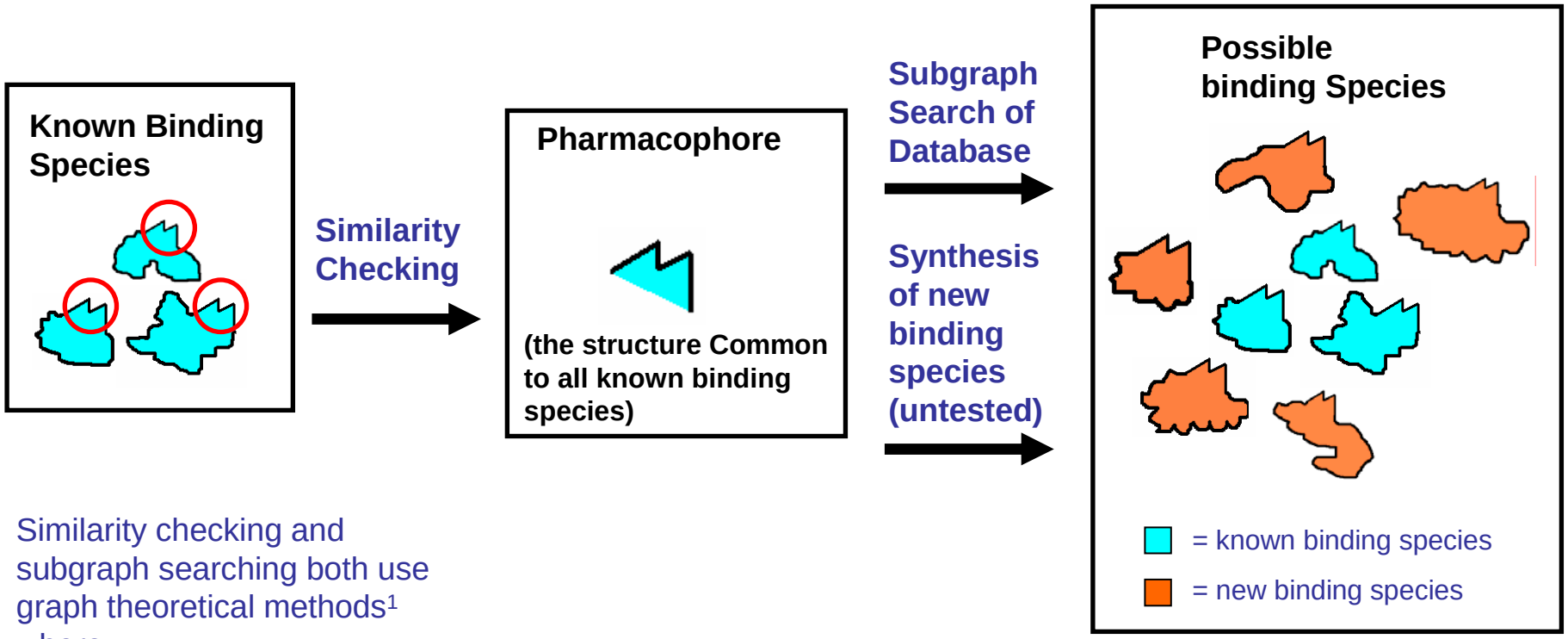



Assuming that species which bind to the same site are similar

1. Hattori et. al., Journal of the American chemical society, 125: 11853-11865 (2003).

# Computation of 2D Pharmacophores




Similarity checking and subgraph searching both use graph theoretical methods[1] where

functional groups = nodes
bonds = edges.


1. Hattori et. al., Journal of the American chemical society, 125: 11853-11865 (2003).

# Graph Theory

We have used the graph theory comparison methods in the program simcomp[1] to compute similarity.

Pharmacophores are computed by finding the maximal common subgraphs between binding species.

Searching for binding species is possible using a subgraph isomorphism algorithm which searches for 1 graph inside another

1. Hattori et. al., Journal of the American chemical society, 125: 11853-11865 (2003).

# Example using the enzyme 1.1.1.27

This enzyme catalyses 3 known reactions

R00703: C00186 + C00003 ⇔ C00022 + C00004 + C00080

R01000: C05984 + C00003 ⇔ C00109 + C00004 + C00080

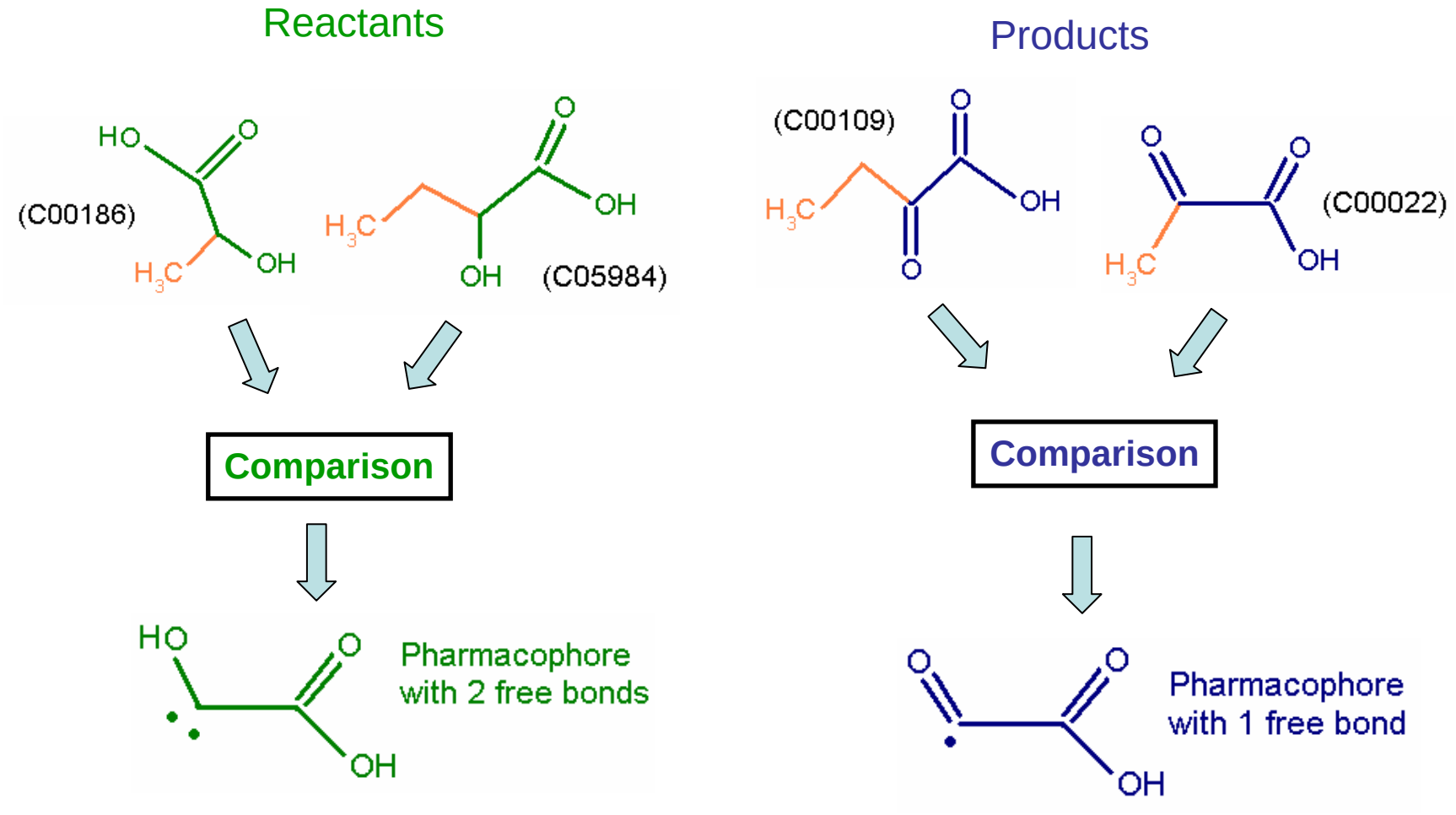
R03104: C05823 + C00003 ⇔ C00958 + C00004

In this case only two pairs of species need to be compared.

**All Calculations were performed using raw data from the KEGG: Ligand databases[2] which contain biochemical reactions and the 2D structure of the compounds involved**

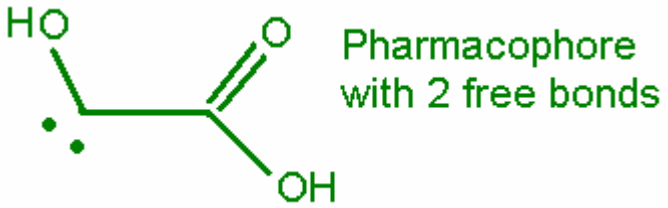
2. Goto et. al., Nucleic Acids Research, 30: 402-404, (2002).

# Comparing the reactants and products

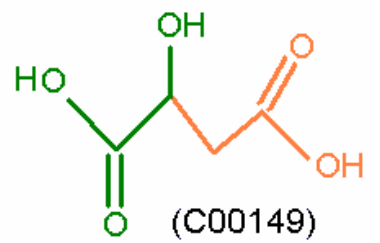


# Searching for binding species containing the 2D pharmacophores

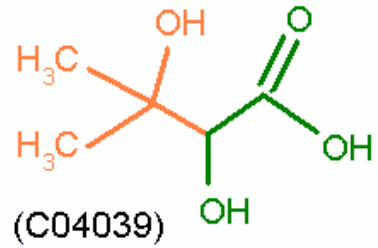
## Reactants



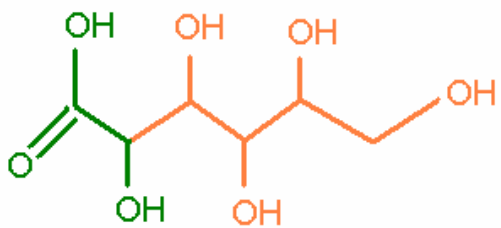

**118 possible binding species**




(C00149)

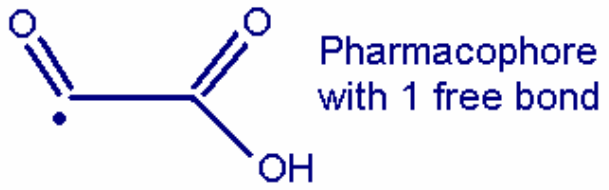



(C04039)

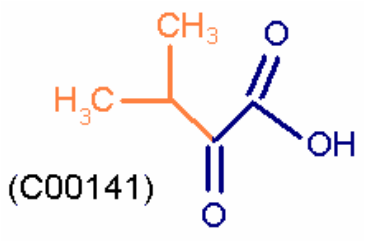



(C00817)

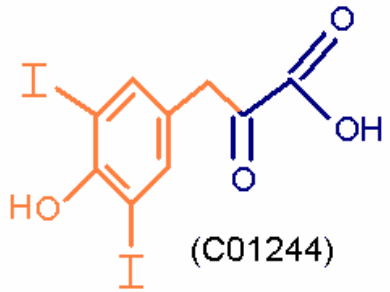
## Products



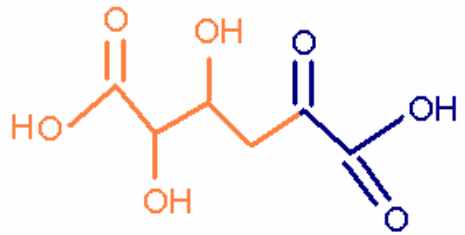

**139 possible binding species**




(C00141)




(C01244)




(C00679)

# Computing 3D Structures

- Using an optimization algorithm together with an expert system
- The approximate bond lengths are found from the covalent radii of the atoms
  - With the 1st atom being placed at coordinates (0,0,0)
- All subsequent atom positions are found through the bond angles of each bond
  - Initially these bond angles are randomly generated
- The bond angles are then optimized using the following objective function
  - The 1st part of this function ensures that the correct bondlengths are maintained
  - The 2nd part maximises the distance between non-bonded pairs to avoid overlap

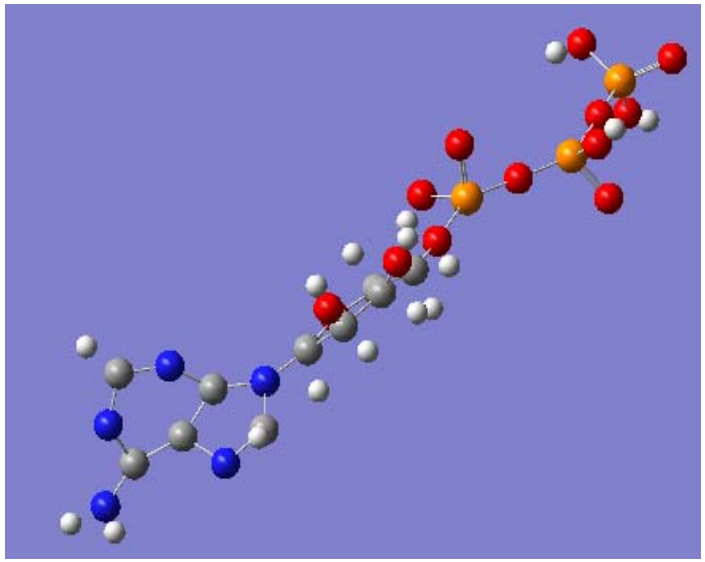
$$\min \sum_{bonded} (d_{i,j} - d_{i,j}^{ex})^2 + \sum_{non-bonded} e^{2 \times d_{i,j}^{ex} / d_{i,j}}$$

$d_{i,j}^{ex}$ Expected distances between atoms if they were bonded

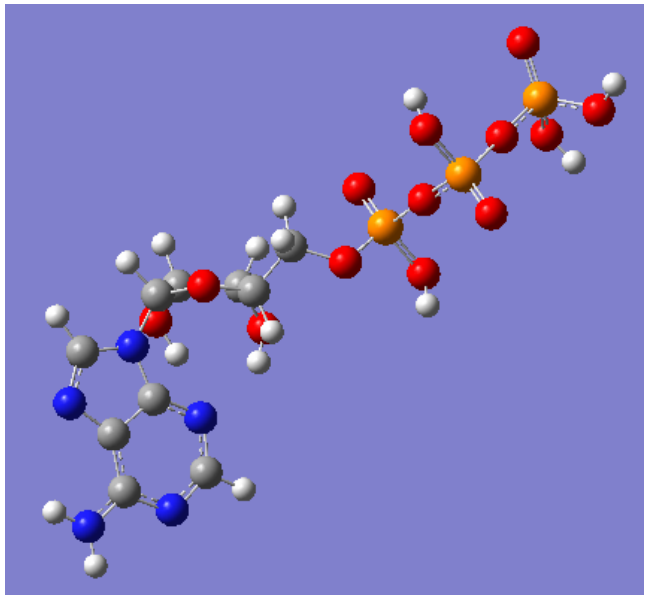
$d_{i,j}$ Actual distances between the atoms

# Structure Optimization and Charge

Approximate 3D structures can be optimized and charges can be calculated using the quantum chemical software Gaussian[5]. Here we have used the semi-empirical method AM1 with basis set STO-3G.



Approximate 3D Structure

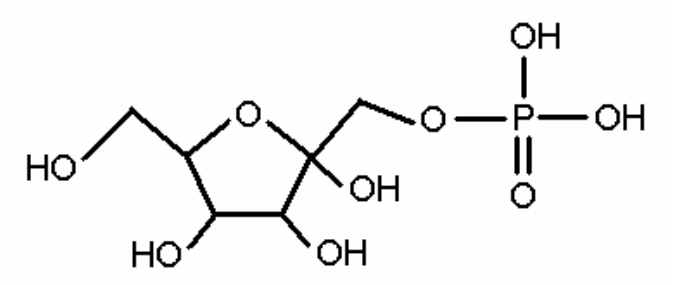


Exact 3D structure

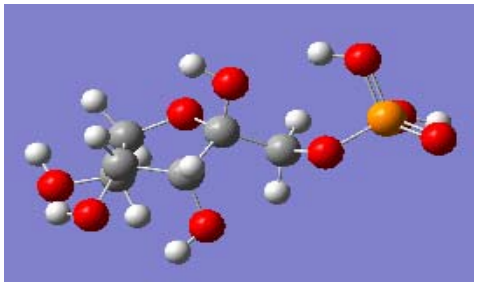
3. Frisch, M., et. Al., Gaussian-03, Inc., Wallingford CT, (2004).

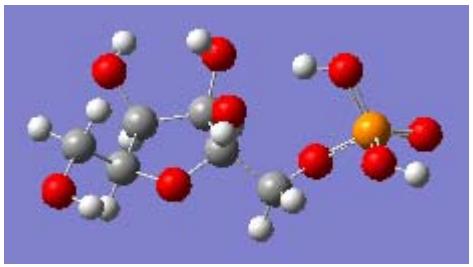
# 3D Pharmacophore Generation

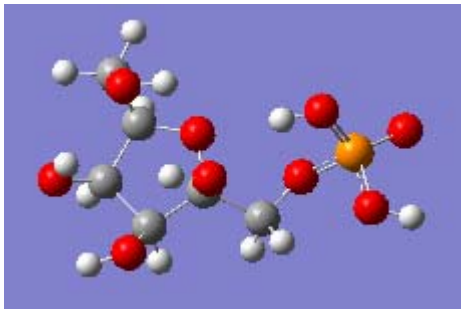
One 2D Structure




Many 3D conformations





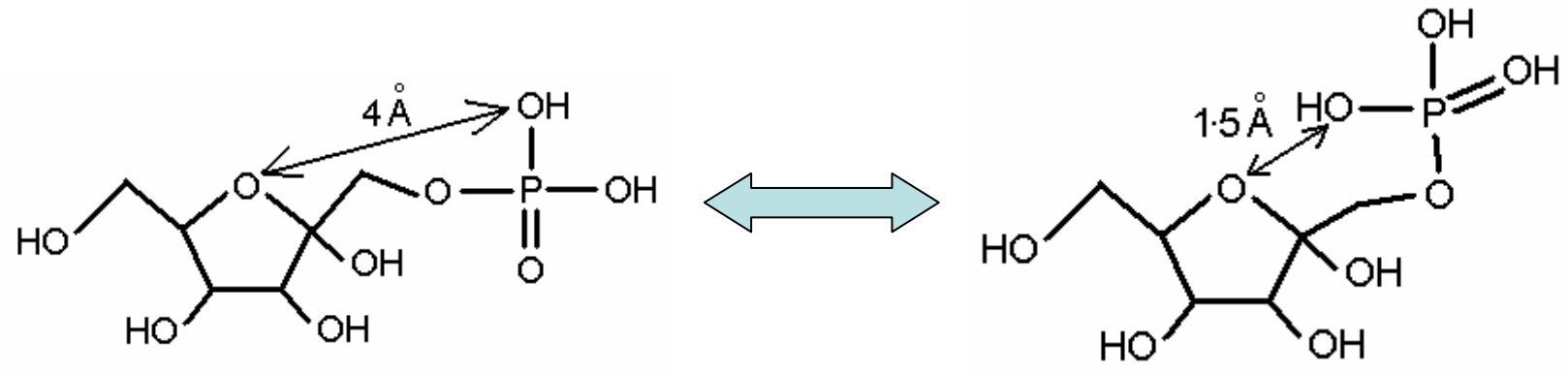


3D structures obtained using Gaussian[3]

3. Frisch, M., et. Al., Gaussian-03, Inc., Wallingford CT, (2004).

# Flexible 3D Bounds: A Model Reduction Step

Molecules can bend and twist



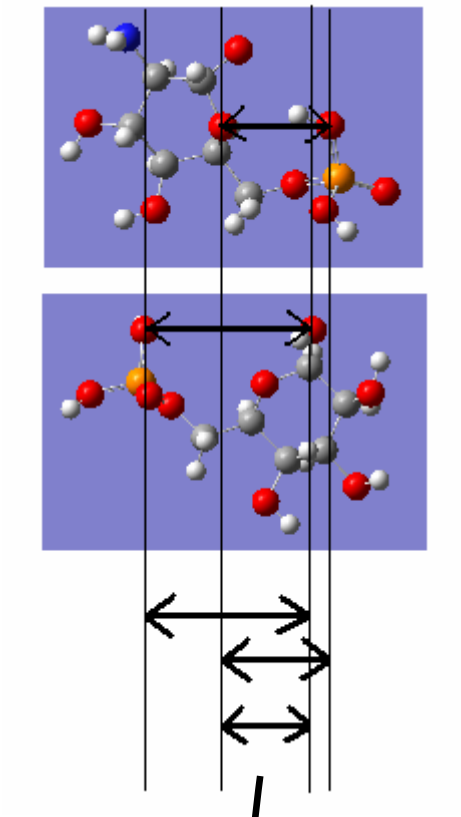

Giving a range of distances between each atom-atom pair

e.g. Oxygen-Oxygen distance = 1.5 – 4 angstroms

A method called Bound Smoothing[4] can be used to obtain these distance ranges

4. Raymond, J., Willett, P., Journal of Chemical Information and Computer Sciences, 44:908-916, (2003).

# 3D Comparisons



**Matching flexible structures have overlapping atom-atom distances**

Comparing of 2 molecules

Gives a smaller bound in the pharmacophore

New Binding species must be able to squeeze in this gap

# 3D Molecule Checks: Computational Requirements

A 3D graph-theoretical method is used based on atom-atom distance specification

**A typical molecule**

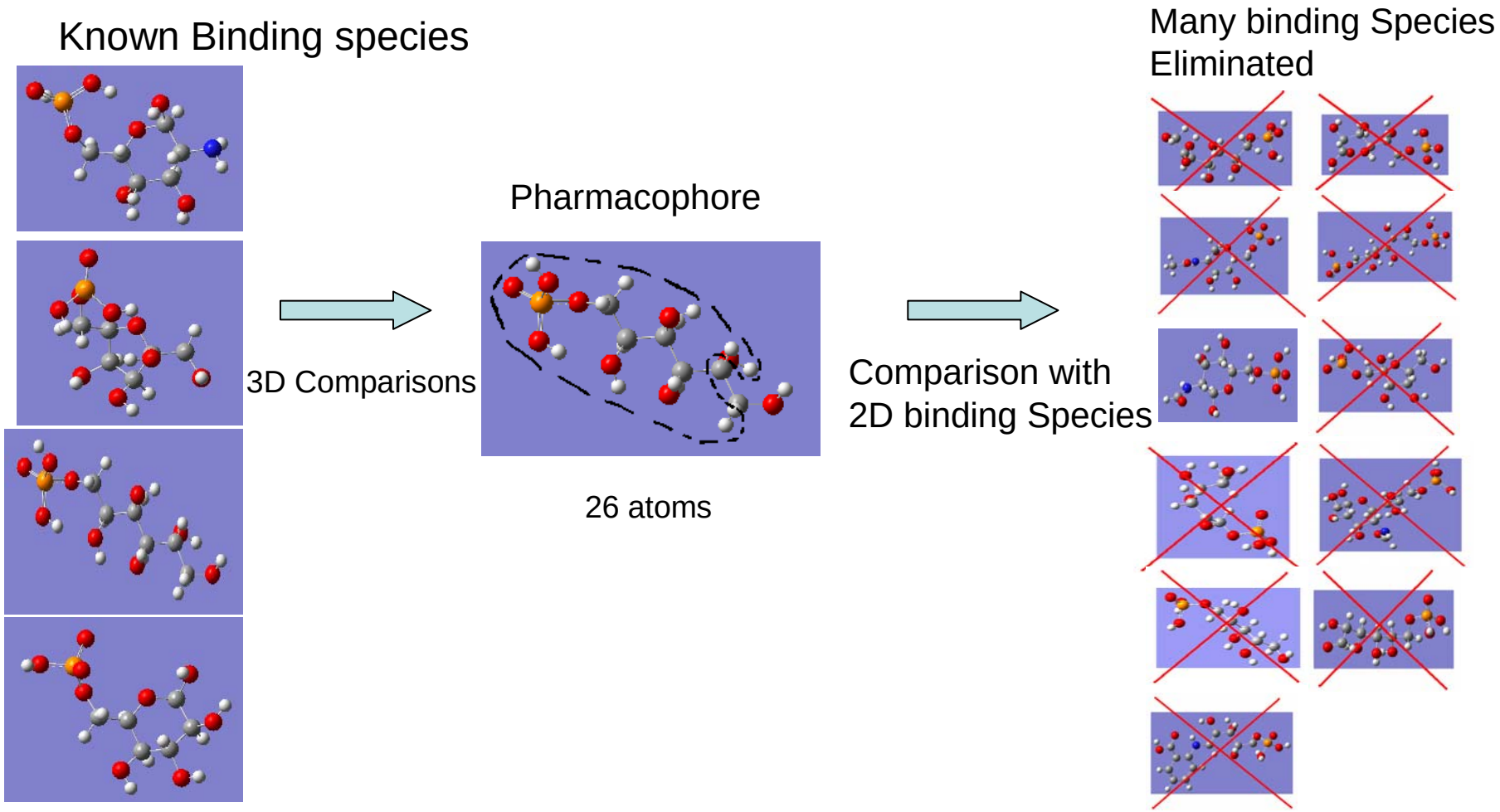
- **30 atoms**
- **30 bonds**
- **435 atom-atom distances**

Where a 2D method would use 30 bonds, the 3D method uses 435 atom-atom distances

Hence the 3D method is much more expensive

# Example: 3D Pharmacophore

A 3D graph theoretical method[4] is used to eliminate 2D binding species




4. Raymond, J., Willett, P., Journal of Chemical Information and Computer Sciences, 44:908-916, (2003).

# Reaction Generation

For each enzyme generate combinations of binding species

Binding Sites

$$A + B \Leftrightarrow C + D$$

Computed Binding species

| $A_1$ | $B_1$ | $C_1$ | $D_1$ |
|---|---|---|---|
| $A_2$ | $B_2$ | $C_2$ | $D_2$ |
| | $B_3$ | $C_3$ | |
| | | $C_4$ | |

Combinations of binding species which can form reactions

- Contain no more than 1 species from each binding site on each side of the reaction

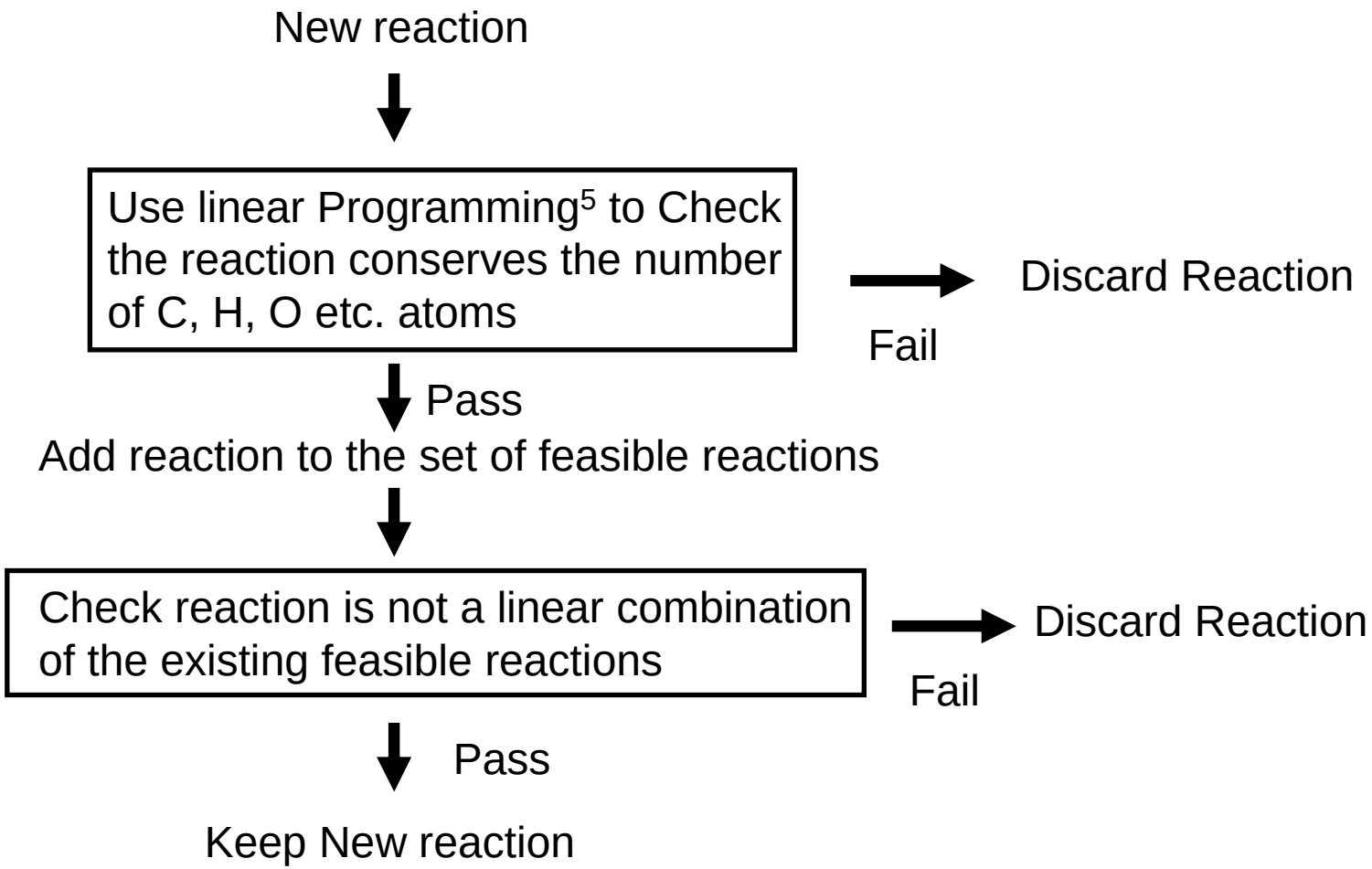
For example $A_1 + A_2 \rightarrow$ products would not be allowed

- Do not involve stoichiometries greater than that of the known reactions

For example $A_1 + B_1 + C_1 \rightarrow A_2 + D_2$ would not be allowed in this case

# Reaction Feasibility

New reactions are tested for feasibility in the following ways

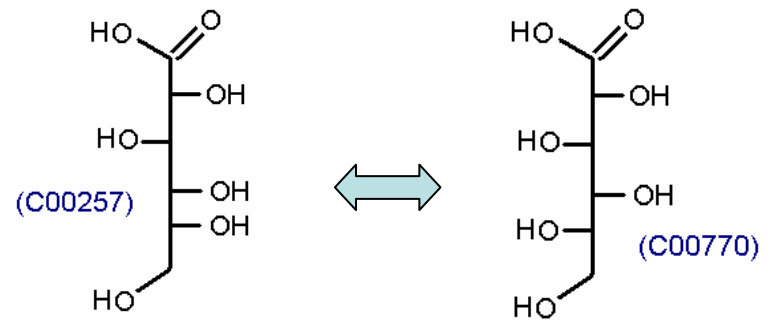



5. Zhang, W., PhD Thesis, UMIST, (2004).

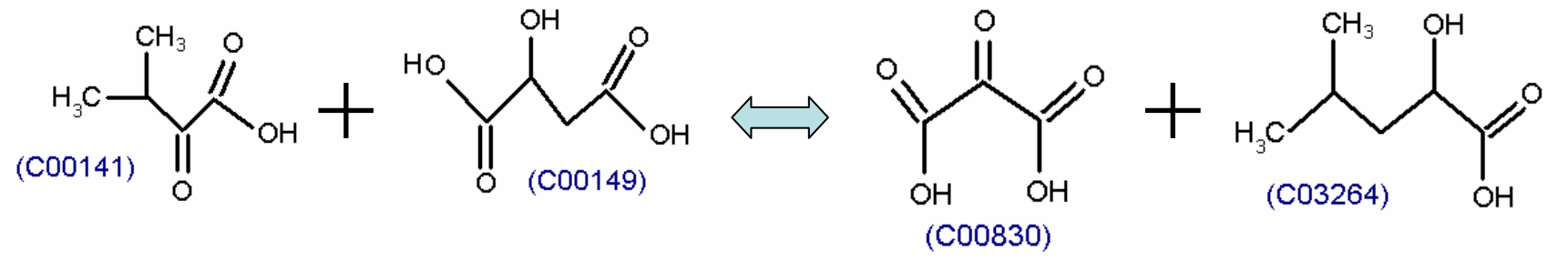
# Example: Reaction Generation for Enzyme 1.1.1.27

**Reaction generation yields 241 linearly independent reactions**

**Including Chirality changing Reactions**



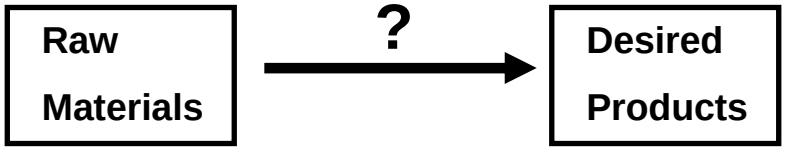

**And many other new reaction involving the new binding species**



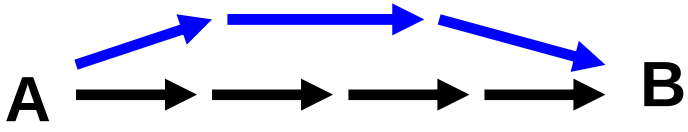
# Pathway Construction

**Allows us to answer the following questions:**

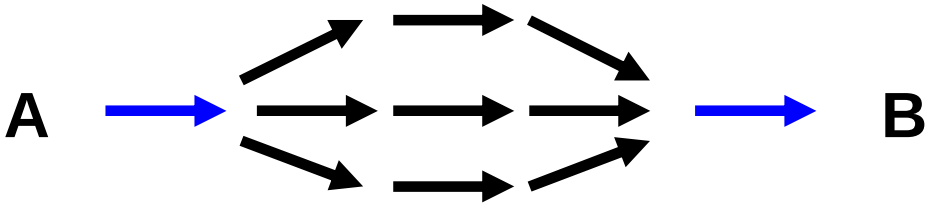
**Can a Reaction network convert a set of raw materials into desired products?**



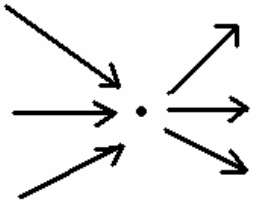

**Are there any alternative pathways converting A into B?**




**Which reactions are important for converting A into B?**




**Pathway construction is a complex problem for large numbers of reactions**

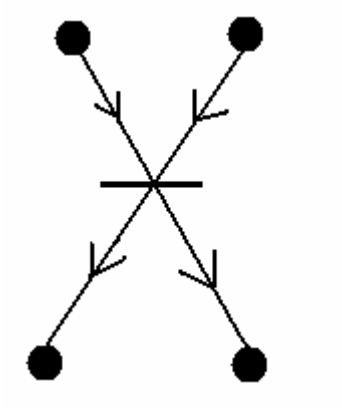


**Pathway construction using P-graphs based method[6]**

6. Seo, H., et. al., Biotechnology Letters, 23: 1551-1557, (2001).

# P-Graph Based Pathway Generation

Pathway generation is possible using an efficient implementation of p-graphs[7]



Dots = species/components

Bars = reactions/processes

Generates pathways according to a set of rules or axioms

e.g. Specified Products must be produced

Using combinatorial algorithms to generate pathways

Based on branch-and-bound methods

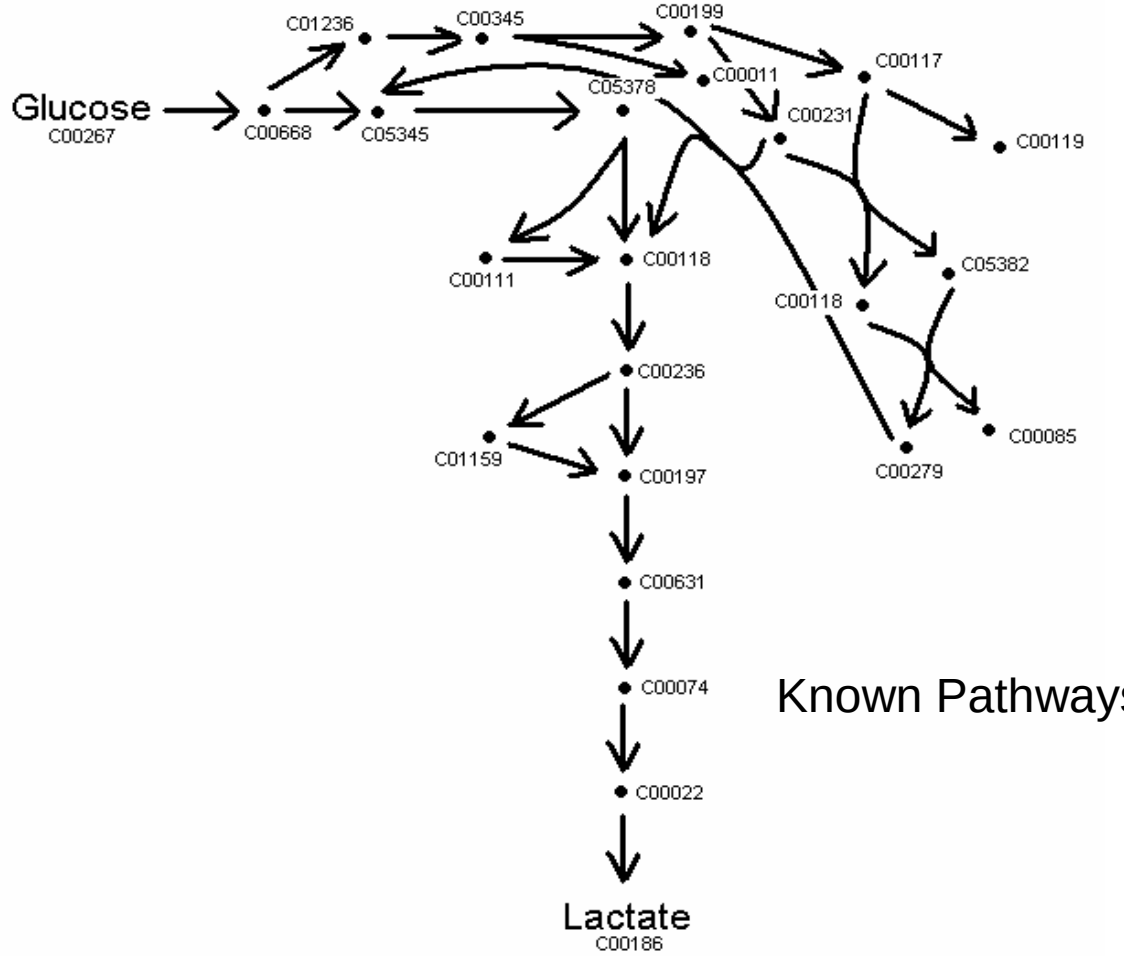
subject to a set of rules and constraints

- Currently limited to fixed overall reactions
- Eliminates cyclic and dependent pathways
- Low memory requirements

6. Seo, H., et. al., Biotechnology Letters, 23: 1551-1557, (2001).

# Human Red Blood Cell: Glycolysis

**We have applied our procedure on 24 enzymes involved in the glycolysis and pentose phosphate pathways**




Known Pathways

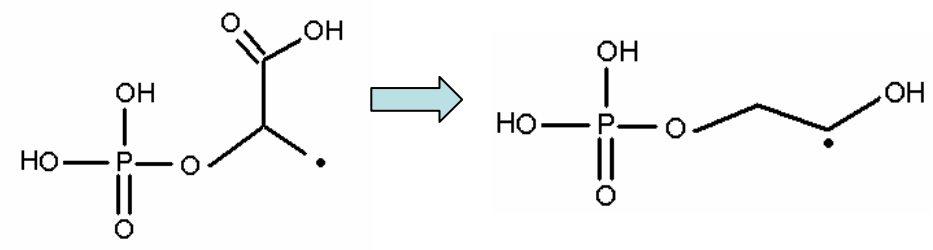
| | |
|---|---|
| C00267 | Glucose |
| C00668 | Glucose 6-Phosphate |
| C05345 | Fructose 6-phosphate |
| C05378 | Fructose 1,6 bisphsphate |
| C00118 | Glyceraldehyde 3-phosphate |
| C00111 | Dihydroxyacetone phosphate |
| C00236 | 1,3-Bisphospho-D-glycerate |
| C00197 | 3-Phospho-D-glycerate |
| C00631 | 2-Phospho-D-glycerate |
| C00074 | Phosphoenolpyruvate |
| C00022 | Pyruvate |
| C00186 | Lactate |
| C00085 | Fructose 6-phosphate (different chirality to C05345) |
| C00345 | 6-Phospho-D-gluconate |
| C00199 | Ribulose 5-phosphate |
| C00011 | Carbon dioxide |
| C00231 | Xylulose 5-phosphate |
| C00117 | Ribose 5-phosphate |
| C05382 | Sedoheptulose 7-phosphate |
| C00279 | Erythrose 4-phosphate |

# Calculating Pharmacophores

**Our procedure calculates pharmacophores for the 24 enzymes (107 overall)**

| Enzyme | Stoichiometry |
|---|---|
| 1.1.1.27 | 2 → 3 and 2 → 2 |
| 1.1.1.44 | 2 → 4 |
| 1.1.1.49 | 2 → 3 |
| 1.2.1.12 | 3 → 3 and 3 → 3 |
| 2.2.1.1 | 2 → 2 |
| 2.2.1.2 | 2 → 2 |
| 2.7.1.1 | 2 → 2 |
| 2.7.1.11 | 2 → 2 |
| 2.7.1.2 | 2 → 2 |
| 2.7.1.40 | 2 → 2 |
| 2.7.2.3 | 2 → 2 |
| 2.7.6.1 | 2 → 2 |
| 3.1.1.31 | 2 → 1 |
| 3.1.3.11 | 2 → 2 |
| 3.1.3.13 | 2 → 2 |
| 3.6.1.5 | 3 → 3 and 2 → 2 |
| 4.1.2.13 | 1 → 2 |
| 4.2.1.11 | 1 → 2 |
| 5.1.3.1 | 1 → 1 |
| 5.3.1.1 | 1 → 1 |
| 5.3.1.6 | 1 → 1 |
| 5.3.1.9 | 1 → 1 |
| 5.4.2.1 | 1 → 1 |
| 5.4.2.4 | 2 → 2 and 1 → 1 |

**Enzyme 5.3.1.9**



MANCHESTER 1824

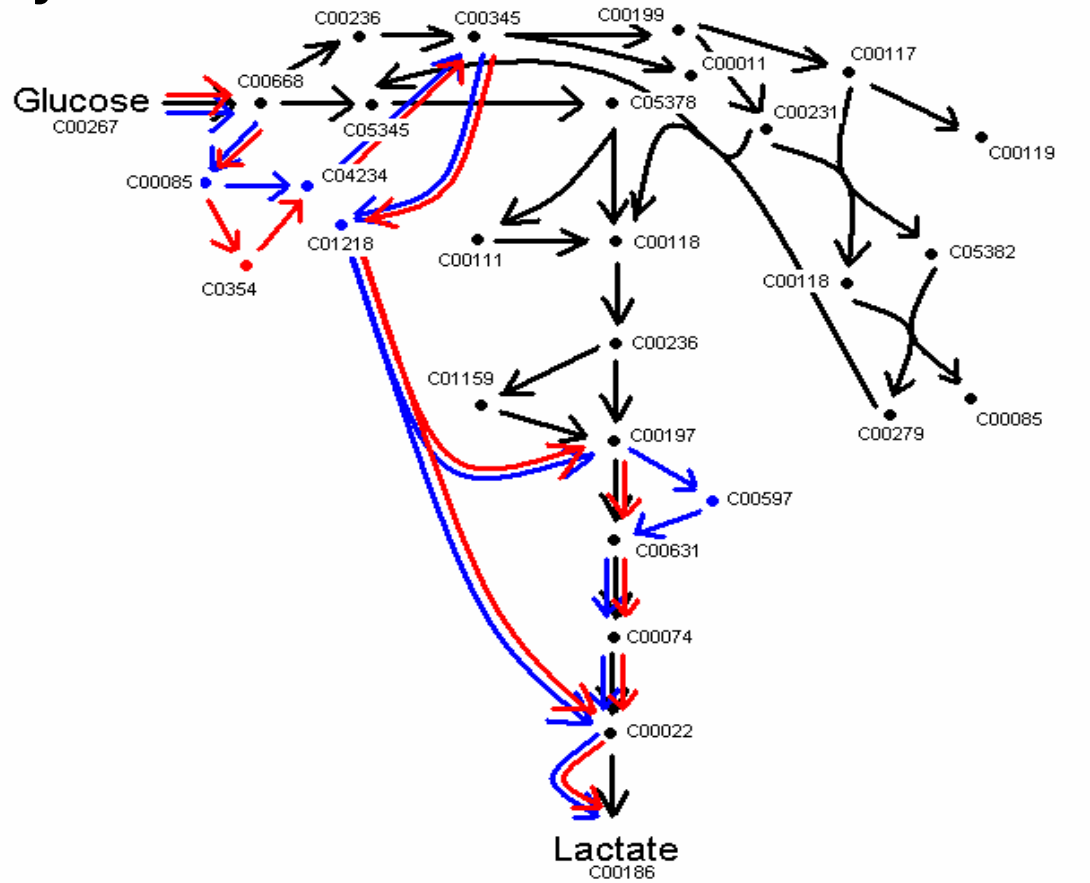
The University of Manchester

# Binding Species and Reactions

| Enzyme | Binding Species | | | Reactions | |
|---|---|---|---|---|---|
| | Known | Calculated | | Known | Calculated |
| 1.1.1.27 | 9 | 254 | | 3 | 241 |
| 1.1.1.44 | 6 | 9 | | 1 | 5 |
| 1.1.1.49 | 6 | 13 | | 2 | 10 |
| 1.2.1.12 | 12 | 17 | | 3 | 2 |
| 2.2.1.1 | 9 | 237 | | 6 | 222 |
| 2.2.1.2 | 4 | 10 | | 1 | 7 |
| 2.7.1.1 | 24 | 825 | | 22 | 789 |
| 2.7.1.11 | 16 | 56 | | 13 | 41 |
| 2.7.1.2 | 8 | 35 | | 3 | 31 |
| 2.7.1.40 | 18 | 46 | | 8 | 35 |
| 2.7.2.3 | 4 | 5 | | 1 | 2 |
| 2.7.6.1 | 4 | 5 | | 1 | 2 |
| 3.1.1.31 | 3 | 3 | | 1 | 1 |
| 3.1.3.11 | 8 | 49 | | 4 | 33 |
| 3.1.3.13 | 4 | 10 | | 1 | 3 |
| 3.6.1.5 | 20 | 312 | | 13 | 297 |
| 4.1.2.13 | 10 | 434 | | 5 | 180 |
| 4.2.1.11 | 7 | 2422 | | 3 | 2039 |
| 5.1.3.1 | 2 | 4 | | 1 | 3 |
| 5.3.1.1 | 2 | 3 | | 1 | 2 |
| 5.3.1.6 | 2 | 6 | | 1 | 5 |
| 5.3.1.9 | 6 | 26 | | 4 | 16 |
| 5.4.2.1 | 4 | 76 | | 2 | 28 |
| 5.4.2.4 | 7 | 82 | | 3 | 31 |

# Glycolysis: Pathway Construction

**Generating new reactions and computing pathways leads to some new routes through the system**




| | |
|---|---|
| C00267 | Glucose |
| C00668 | Glucose 6-Phosphate |
| C05345 | Fructose 6-phosphate |
| C05378 | Fructose 1,6 bisphsphate |
| C00118 | Glyceraldehyde 3-phosphate |
| C00111 | Dihydroxyacetone phosphate |
| C00236 | 1,3-Bisphospho-D-glycerate |
| C00197 | 3-Phospho-D-glycerate |
| C00631 | 2-Phospho-D-glycerate |
| C00074 | Phosphoenolpyruvate |
| C00022 | Pyruvate |
| C00186 | Lactate |
| C00085 | Fructose 6-phosphate (different chirality to C05345) |
| C00354 | Fructose 1,6-bisphosphate (different chirality to C05378) |
| C00345 | 6-Phospho-D-gluconate |
| C00199 | Ribulose 5-phosphate |
| C00011 | Carbon dioxide |
| C00231 | Xylulose 5-phosphate |
| C00117 | Ribose 5-phosphate |
| C00119 | 5-Phospho-alpha-D-ribose 1-diphosphate (PRPP) |
| C05382 | Sedoheptulose 7-phosphate |
| C00279 | Erythrose 4-phosphate |
| C00597 | 3-Phosphoglycerate (different chirality than C00197) |
| | |
| C01218 | 6-Phospho-2-dehydro-D-gluconate |
| C04234 | 2-Carboxy-D-arabinitol 1-phosphate |

# Metabolic Network Development: A knowledge-based method for generating Biochemical reaction networks.

**Identify known enzymes in system**

↓

**Select enzyme**

↓

**For known reactions Match Binding Species to binding sites**

↓

**Calculate 2D pharmacophores for each binding site**

↓

**Search for new binding species using the pharmacophores**

↓

**Reduce complexity using 3D pharmacophore screening**

↓

**Generate Reactions**

↓

**Construct pathways using the reactions generated for all the enzymes**

↓

**Model reduction based kinetic model construction**

**Select Next Enzyme**

# Low Dimensional Manifolds

$$\frac{dc}{dt} = F(c)$$

System of ODEs

$$Q^T J Q = \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix}$$

Decoupling fast and slow timescales

$$S_{11} Z - Z S_{22} = -S_{12}$$

$$T = Q\left( I + \begin{pmatrix} 0 & Z \\ 0 & 0 \end{pmatrix} \right) \qquad T^{-1} J T = \begin{pmatrix} S_{11} & 0 \\ 0 & S_{22} \end{pmatrix}$$

Projection onto fast/slow subspace

$$\begin{pmatrix} g_{slow} \\ g_{fast} \end{pmatrix} = T^{-1} F(T\,\cdot) \qquad \begin{pmatrix} x_{slow} \\ x_{fast} \end{pmatrix} = T^{-1} c$$

QSSA for fast species

$$\frac{dx_{slow}}{dt} = g_{slow}\left(x_{slow}, x_{fast}\right)$$

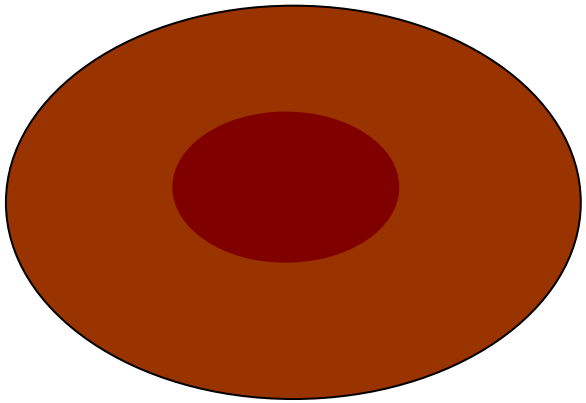
$$\frac{dx_{fast}}{dt} = 0$$


7. Maas, U., Pope, S. B., Combustion and Flame, 88:239-264, (1992).
8. Shaik, O. S., et. al., The Journal of Chemical Physics, 123, 234103, (2005)

# Human Red Blood Cell

**A biochemical system representing the function of a red blood cell[9]**

- 49 Species
  - 5 extracelluar species
  - 44 intracellular species
- 41 Reactions



A simple system

- No Nucleus
- Involves mostly Glycolysis

9. Ni, T., Savageau, M., Journal of Theoretical Biology, 179: 329-368, (1996)

# Example: Human Red Blood Cell

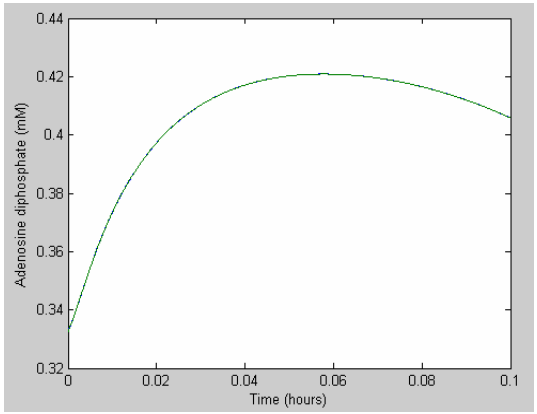
**Applying the LDM procedure**

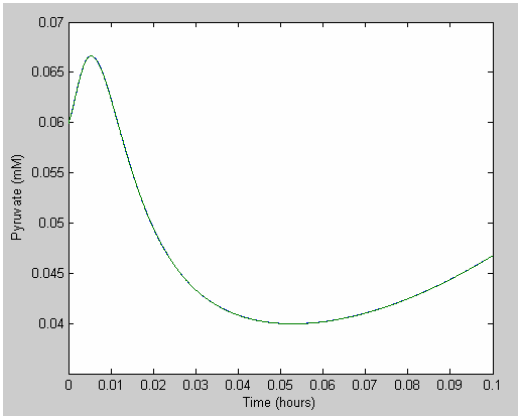
| | $\varepsilon\left\|g_{slow}(c_0)-g(c_{DAE})\right\|$ |
|---|---|
| 44 ODEs and 5 algebraic equations | 0.00074 |
| 40 ODEs and 9 algebraic equations | 0.00528 |
| 37 ODEs and 12 algebraic equations | 0.0441 |

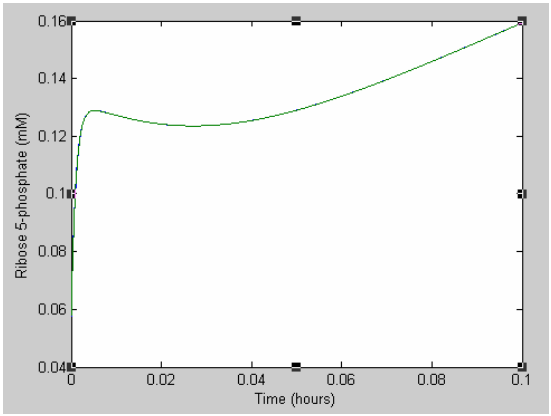
For this system a large number of ODEs are required

# Comparison with Full Model

37 ODEs and 12 algebraic equations




Comparison with the full kinetic model showed an almost exact match for 0.1 hours

# Species contributing to the Slow Dynamics

By analysing the projection matrix the species contributing to the fast or slow dynamics can be assessed[8]

$$x_{slow} = T_r^{-1} c$$

$$p_i^{slow} = \sum_{j=1}^{r} t_{j,i} t_j \qquad p_i^{fast} = \sum_{j=r+1}^{n} t_{j,i} t_j$$

$$d_i^{slow} = \left| e_i - p_i^{slow} \right| \qquad d_i^{fast} = \left| e_i - p_i^{fast} \right|$$

$$r_i^{slow} = \frac{d_i^{fast} \arccos d_i^{slow}}{d_i^{fast} \arccos d_i^{slow} + d_i^{slow} \arccos d_i^{fast}}$$

$$r_i^{fast} = \frac{d_i^{slow} \arccos d_i^{fast}}{d_i^{slow} \arccos d_i^{fast} + d_i^{fast} \arccos d_i^{slow}}$$

$t_j$ = column vectors of T

$t_{j,i}$ = ith component of $t_j$

$e_i$ = unit vector (species i)

$d_i$ = distance of species i from slow/fast subspace

$p_i$ = projection of species i on the slow/fast subspace

8. Shaik, O. S., et. al., The Journal of Chemical Physics, 123, 234103, (2005)

# Species contributing to the Slow Dynamics

Analysis of $r_i^{fast}$ and $r_i^{slow}$

For the system with 37 slow modes and 12 fast modes shows

7 species are most associated with the fast dynamics

GL6P, GA3P, 1,3DPG, 2PG, PYR, INO and R1P

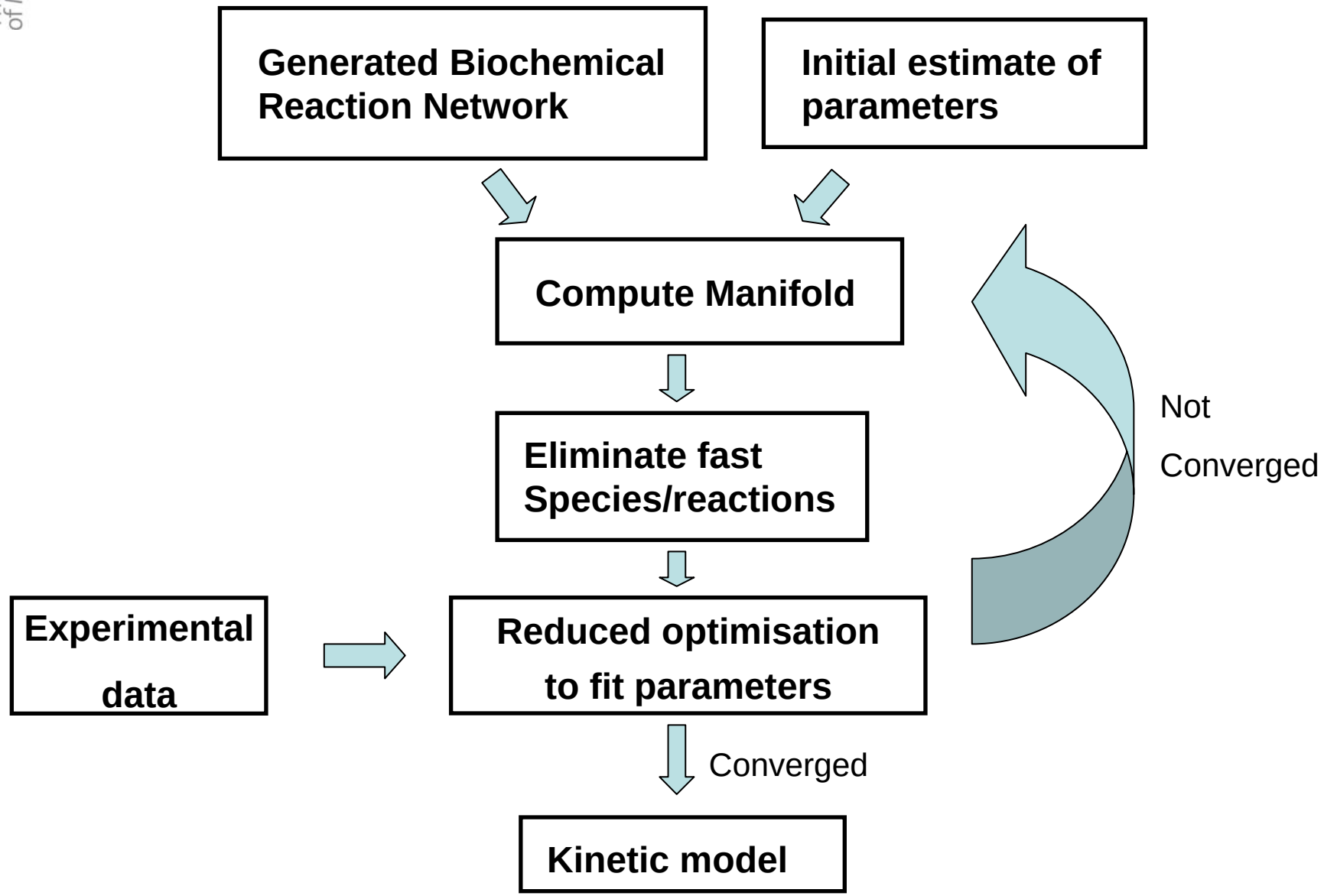
31 species are most associated with the slow dynamics

e.g. ATP, Glc, 2,3DPG, FDP, NADP, GSSG, LAC, etc.

And the remaining 11 species are in-between

5. Shaik, O. S., et. al., The Journal of Chemical Physics, 123, 234103, (2005)

# Parameter Estimation



MANCHESTER 1824

The University of Manchester

# Conclusions

We have developed a complete procedure to construct a wide range of metabolic networks with minimal knowledge of the system

The Procedure:

- Finds New Binding Species
- Calculates New reactions
- Generates New pathways

We have eliminated redundant data at each step to reduce the complexity

Finding kinetic expressions and fitting parameters for large reaction networks

Should be possible if a sufficiently reduced model can be found

- Using a Low Dimensional Manifold method
- Identifying and eliminating species which do not contribute to the slow dynamics

MANCHESTER 1824

The University of Manchester

# Acknowledgements

The financial support of the EU programme AITEKIN C00P-CT-2003-506667 is gratefully acknowledged