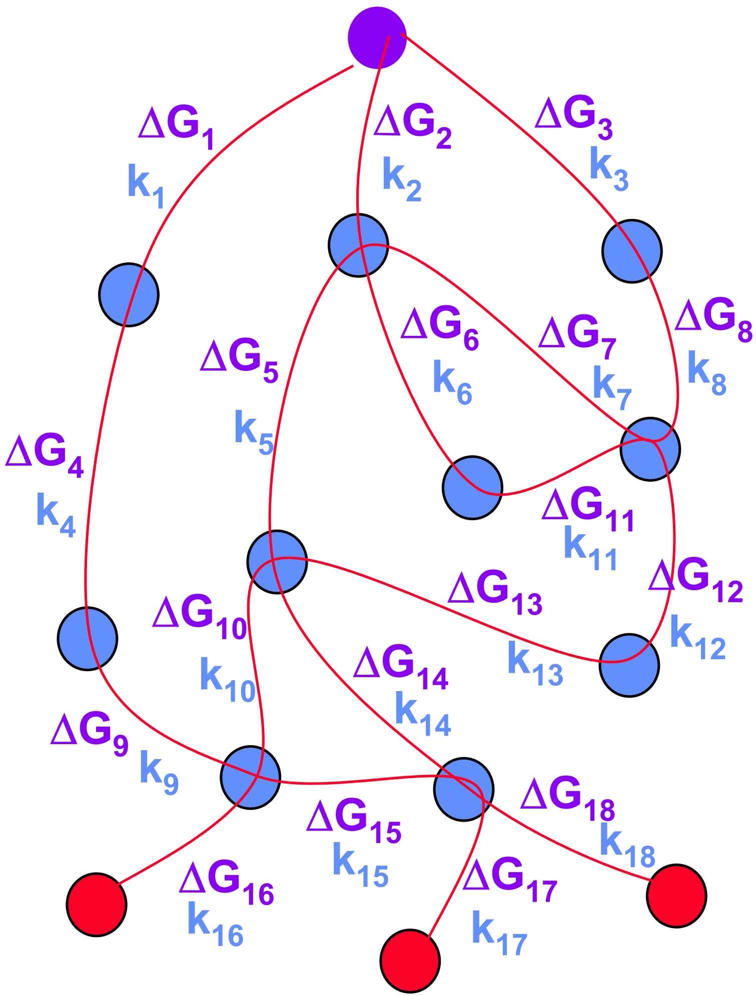
Methods for Establishing Reaction Networks

Linda J. Broadbelt, Maria C. Curet-Arana and Shumaila Khan

Department of Chemical Engineering
Northwestern University
Evanston, IL 60208

Components of the Reaction Network

- Reactants, intermediates and products
- Reactions
- Thermodynamic parameters
- Kinetic parameters



Challenges for Reaction Network Development

- Reactive intermediates have not been detected
- Pathways have not been elucidated experimentally
- Thermodynamic and kinetic parameters are unknown
- Reaction networks are large
- Construction is tedious and prone to user's bias and errors

Approaches for Elucidating Reaction Networks

- Experimental
  - Analysis of kinetic data
  - Mechanism analysis
- Theoretical
  - Automated network generation
  - Network reduction
  - Quantum chemical calculations
  - Emerging techniques

Approaches for Elucidating Reaction Networks

- Experimental
  - Analysis of kinetic data
  - Mechanism analysis
- Theoretical
  - Automated network generation
  - Network reduction
  - Quantum chemical calculations
  - Emerging techniques

Automated Network Generation

- Complex chemistry can be represented by a small number of reaction types
- Complexity arises because of application of reaction types to many different substrates
- Chemistry can be represented mathematically as local changes of bonds and electrons
- Mathematical operators can be implemented in the computer to generate reaction networks automatically

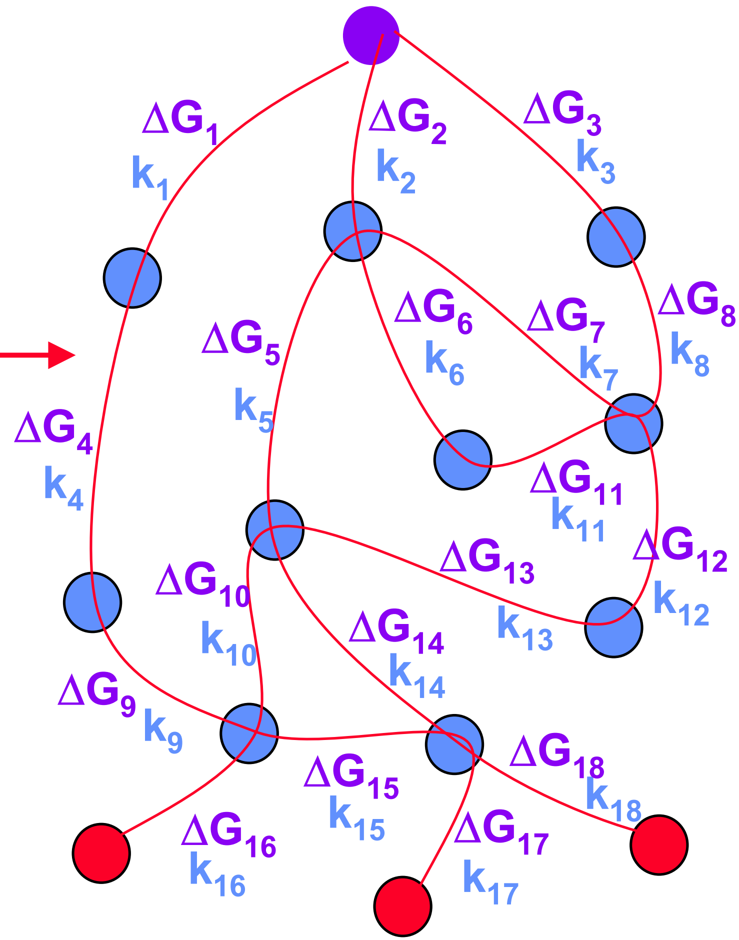
Elements of Computer Generated Reaction Networks

Reactants

Reaction Types

Reaction Rules

- Graph Theory
- Reaction Matrix Operations
- Connectivity Scan
- Uniqueness Determination
- Property Calculation
- Termination Criteria



Bond-Electron Representation Allows Implementation of Chemical Reaction

C	0	1	1	1	1
H	1	0	0	0	0
H	1	0	0	0	0
H	1	0	0	0	0
H	1	0	0	0	0

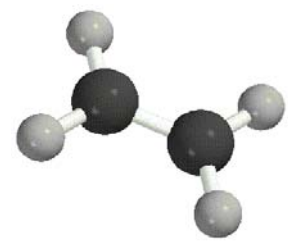
methane

C	1	1	1	1
H	1	0	0	0
H	1	0	0	0
H	1	0	0	0

methyl radical

C	0	2	1	0	0	1
C	2	0	0	1	1	0
H	1	0	0	0	0	0
H	0	1	0	0	0	0
H	0	1	0	0	0	0
H	1	0	0	0	0	0

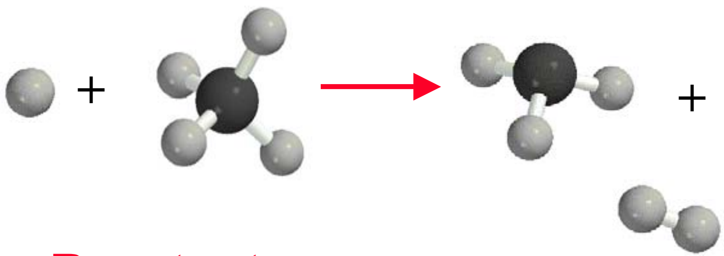
ethylene



- ij entries denote the bond order between atoms i and j
- ii entries designate the number of nonbonded electrons associated with atom i

Chemical Reaction as a Matrix Addition Operation

$H\bullet + CH_4 \longrightarrow \bullet CH_3 + H_2$



Reaction Operation

$$\begin{array}{c} H \\ C \\ H\bullet \end{array}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \longrightarrow \begin{array}{c} H \\ C\bullet \\ H \end{array}\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Reactant Matrices

$$\begin{array}{c} C \\ H \\ H \\ H \\ H \end{array}\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$H\bullet \begin{bmatrix} 1 \end{bmatrix}$$

Reactant Matrix

$$\begin{array}{c} C \\ H \\ H \\ H \\ H \\ H\bullet \end{array}\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Reordered Reactant Matrix

$$\begin{array}{c} H \\ C \\ H\bullet \\ H \\ H \\ H \end{array}\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Product Matrix

$$\begin{array}{c} H \\ C\bullet \\ H \\ H \\ H \\ H \end{array}\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Formulation of Reaction Matrices Using Enzyme Classification System

- Enzyme commission (EC) code number provides systematic names for enzymes

- EC i.j.k.l → unique enzyme
  - i → the main class
  - j → the specific functional groups
  - k → cofactors
  - l → specific to the substrates

Generalized Enzyme Function

EC 1.1.1.1 : *Alcohol dehydrogenase*

1. : *Oxidoreductase*

1.1. : *Acting on the CH-OH group of donors*

1.1.1. : *Using* NAD^+ *or* $NADP^+$ *as acceptor*

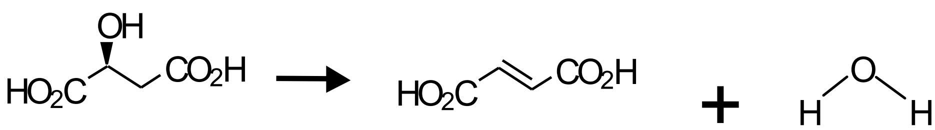
1.1.1.1 : *Alcohol dehydrogenase*

1.1.1.3 : *Homoserine dehydrogenase*

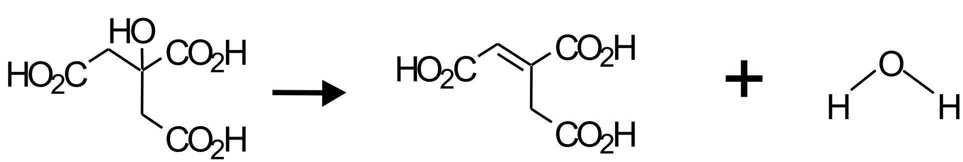
1.1.1.6 : *Glycerol dehydrogenase*

Example of a Generalized Enzyme Reaction

- EC 4.2.1.2 (fumarate hydratase)



- EC 4.2.1.3 (aconitate hydratase)



Generalized enzyme reaction
(EC 4.2.1)

H-C-C-O-H

↓

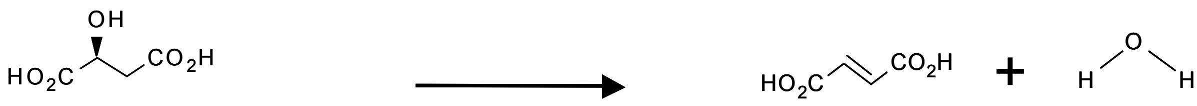
C=C + H_2O

Generalized Enzyme Function Examined at the i.j.k Level

- More than 5,000 specific enzyme functions (i.j.k.l)
- Fewer than 250 generalized enzyme functions (i.j.k)
- Novel enzyme functions should be expected through genomic sequencing, proteomics and protein engineering

Matrix Representation of Generalized Enzyme Function (i.j.k)

Generalized enzyme reaction EC 4.2.1



H-C-C-O-H ⟶ C=C + HOH

Reactant

	H	C	C	O
H	0	1	0	0
C	1	0	1	0
C	0	1	0	1
O	0	0	1	4

+

Reaction operator

	H	C	C	O
H	0	1	0	1
C	1	0	1	0
C	0	1	0	1
O	1	0	1	0

=

Products

	H	C	C	O
H	0	0	0	1
C	0	0	2	0
C	0	2	0	0
O	1	0	0	4

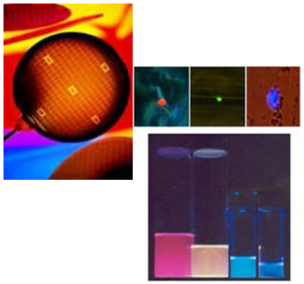
Implications for Novel Pathway Development

Given a novel reaction (reactant/product), can we identify enzymes (catalysts) that could be engineered (evolved) to carry this novel biotransformation ?

If A gives B under 2.4.1 action,

then target enzymes within the 2.4.1 class

Complex Chemistry Summarized in Terms of Reaction Matrices



Silicon nanoparticle production

4 reversible reaction families


Tropospheric ozone formation

15 thermal reaction families
4 photolysis reaction families
9 small molecule reactions


Biochemical transformations

205 unique enzyme actions in KEGG database at i.j.k level

Application of Reaction Matrix Approach

Step 1

Enumerate all enzymes in the EC system

Step 2

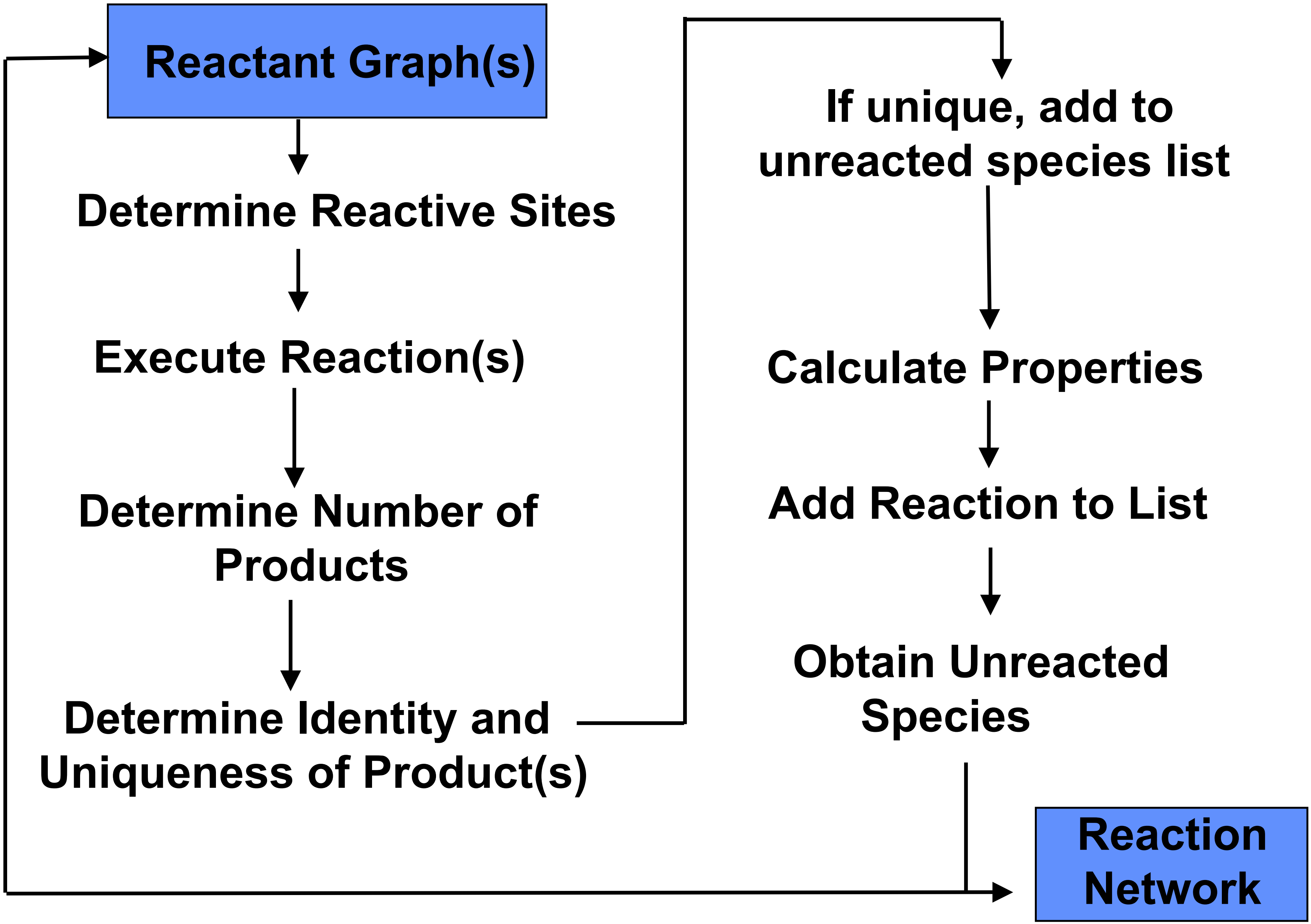
Choose a specific pathway to explore its synthetic ability

Example

Tryptophan biosynthetic pathway

- Exists in higher plants and microorganisms
- Pathway does not exist in mammals
- Tryptophan and its derivatives have considerable market value

Logic of the Algorithm



Tryptophan Biosynthesis Pathway

Input Molecules

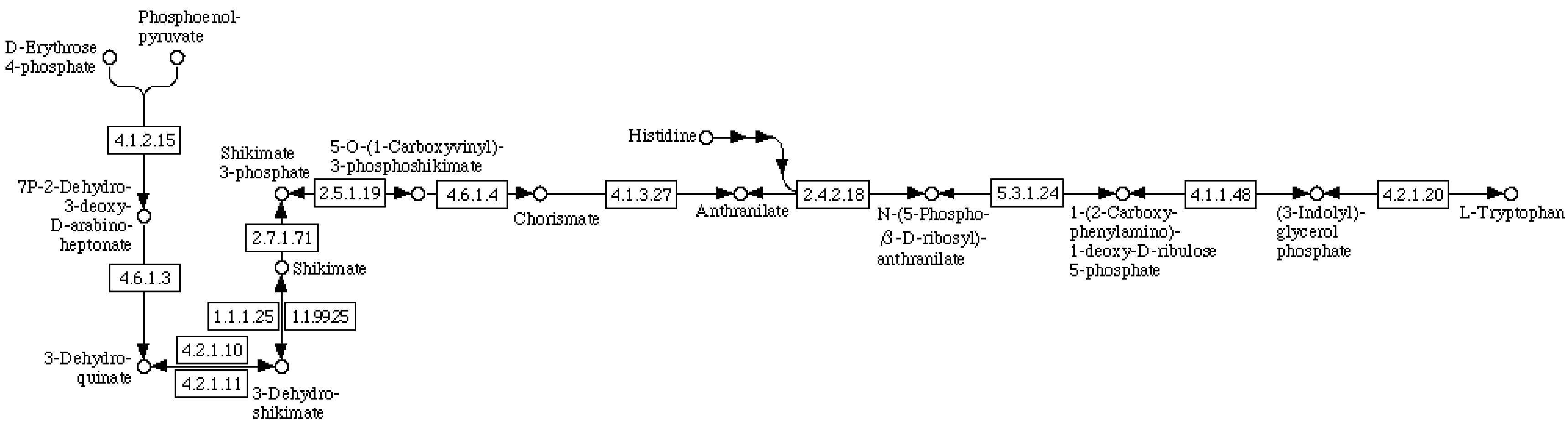
phosphoenolpyruvate (PEP), erythrose-4-phosphate (E4P), glutamine, serine, ribose-5-phosphate (R5P)

Cofactors

ATP, NADPH

Specific Enzyme Actions

12



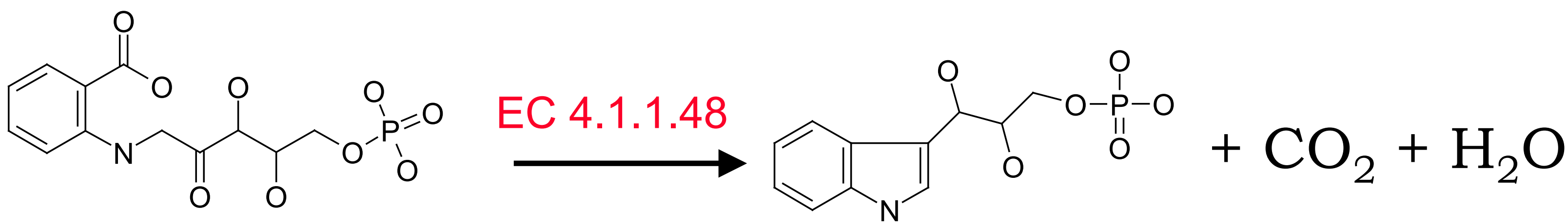
Addition of Chemical Transformations

Generalized reaction for EC 4.1.1 Carboxylyases

C (or O)=C-C-COOH
HN -C-COOH
Or O=C-COOH

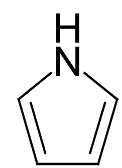
$\xrightarrow{\text{EC 4.1.1}}$ R-CH + CO_2

Specific member of EC 4.1.1



Requires chemical transformation

O=C-CH-NH-C=CH →



+ H_2O

New Pathway from Chorismate to Tryptophan

- Four input molecules: Chorismate, glutamine, R5P, Serine
- The following numbers are unique molecule IDs

Original pathway

1+2 →(EC413) 11 →(EC242) 32 →(CT2) 120 →(EC531) 472 →(EC411) 537 → 2031 → 7181 → 17905

\+ 3 → 26 (into 7181 → 17905)

Tryptophan (17905)

New pathway

1 →(EC412) 5 + 16 →(EC413) 37 →(CT2) 138 →(EC531) 537 → 2031 → 7181 → 17905

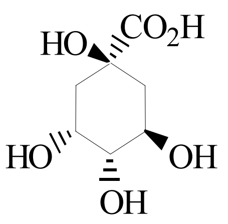
3+2 ↑EC242 (into 37 → 138)

\+ 3 → 26 (into 7181 → 17905)

Tryptophan (17905)

Exploring Novel Pathways and Molecules

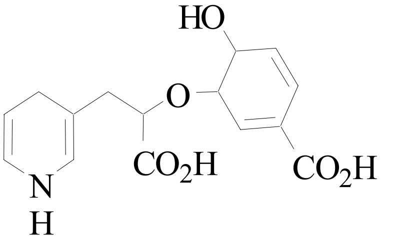
New routes to bioavailable species



1,3,4,5-Tetrahydroxy
Cyclohexanecarboxylic acid

Present in KEGG
(Kyoto Encyclopedia of Genes and Genomes)

New molecules

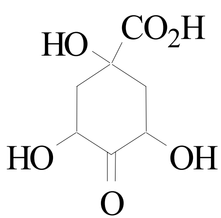


3-[1-Carboxy-2-(1,4-dihydro-pyridin
-3-yl)-ethoxy]-4-hydroxy-cyclohexa-
-1,5-dienecarboxylic acid

NOT present in KEGG
NOT present in CAS REGISTRY

Migration to Biocatalytic Processes

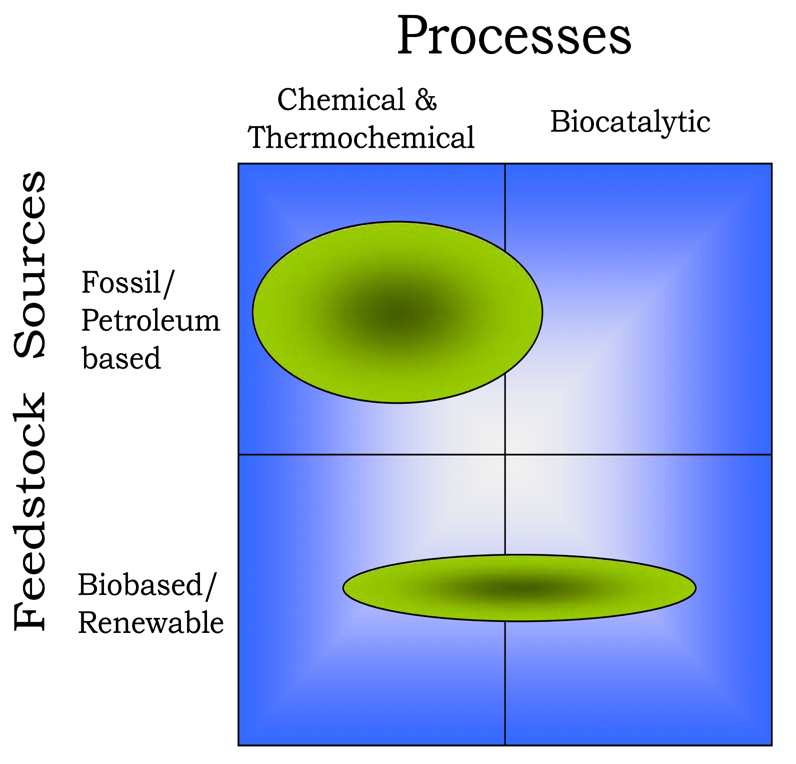
New biochemical routes to existing chemicals



1,3,5-Trihydroxy-4-oxo-cyclohexane carboxylic acid

NOT present in KEGG

Present in CAS REGISTRY



When Does Reaction Network Generation Halt?

- For some chemistries, unique species may continually be formed

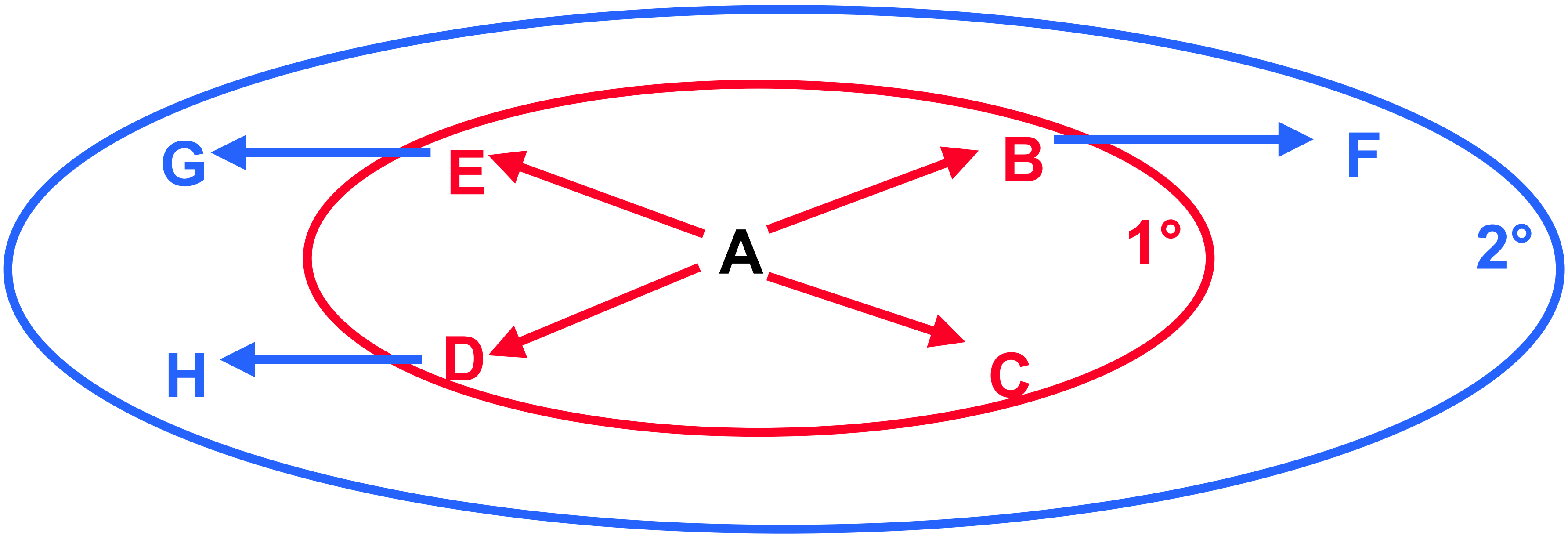
$\bullet CH_3 + C_2H_4 \rightarrow \bullet CC_2H_7$
$\bullet CC_2H_7 + C_2H_4 \rightarrow \bullet CC_4H_{11}$
$\bullet CC_4H_{11} + C_2H_4 \rightarrow \bullet CC_6H_{15}$
•
•

- Without external termination criteria, the network will grow to infinite size

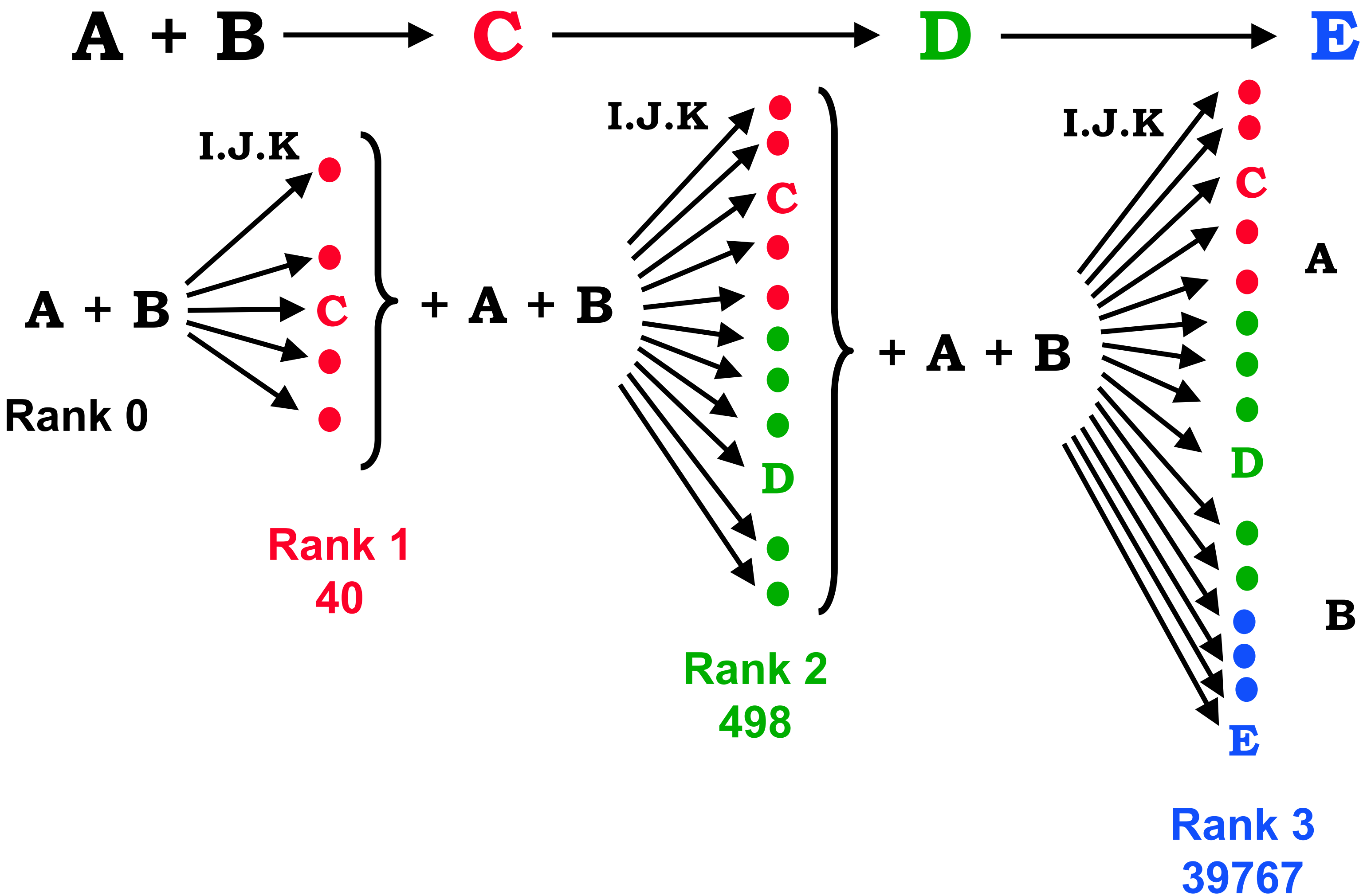
- Rank-based criterion prevents infinite network generation

Rank-Based Termination of Reaction Network Generation

- Generation of the reaction network is terminated when all species of a specified product rank have been allowed to react



Reaction Network Growth is Controlled but Rapid



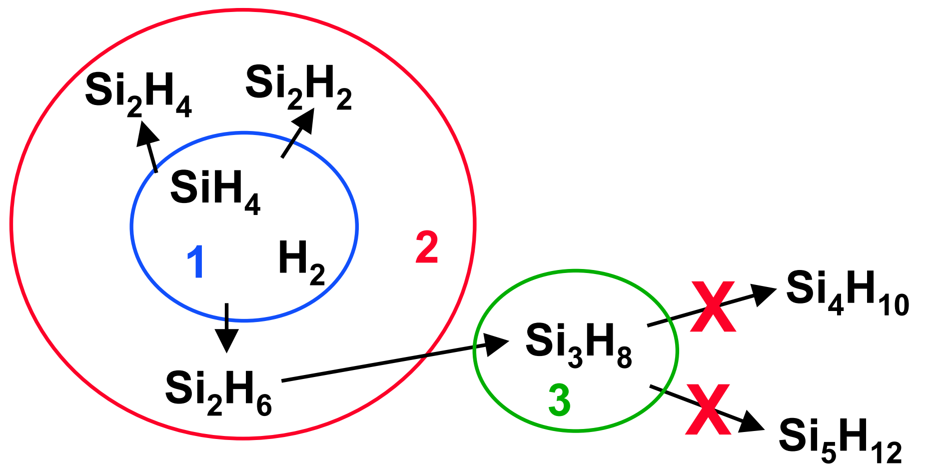
What Other Strategies Can We Use to Generate Networks Intelligently?

Rank

Limits the rank of the reactants

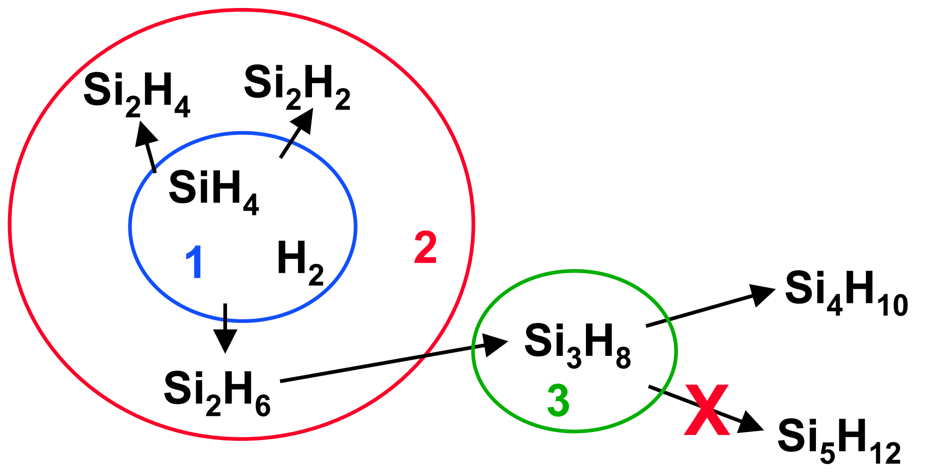
Heavy atom count

Places a bound on the number of heavy atoms in the **reactant**



Heavy atom shell

Place a bound on the number of heavy atoms in the **product**



Growth of Reaction Network is Explosive

Rank-based termination

Includes many insignificant species of lower rank while excluding important species of higher rank

Heavy atom count (HAC)

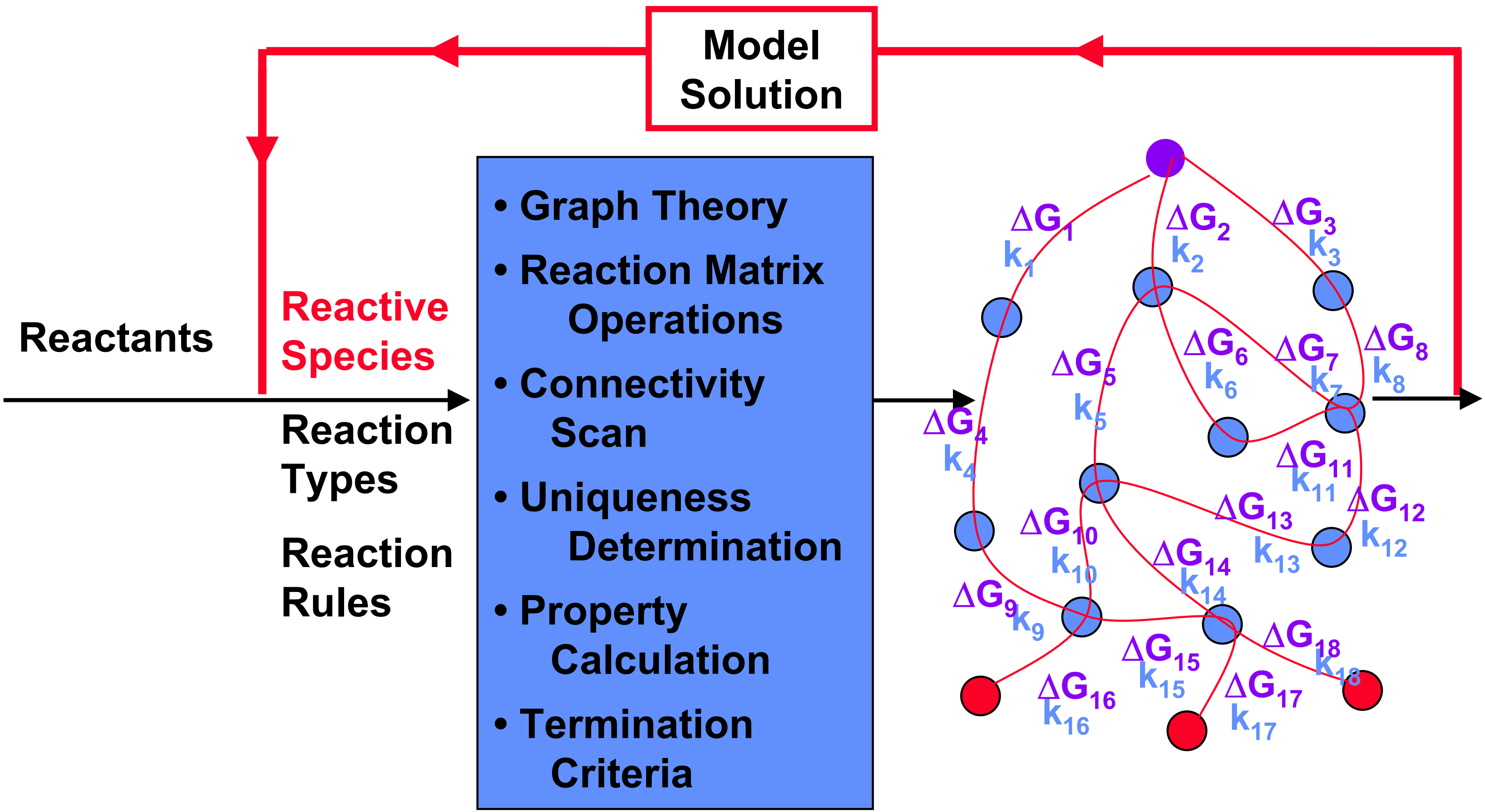
Offers no method to prevent the formation of a large number of chemically insignificant species

Heavy atom shell (HAS)

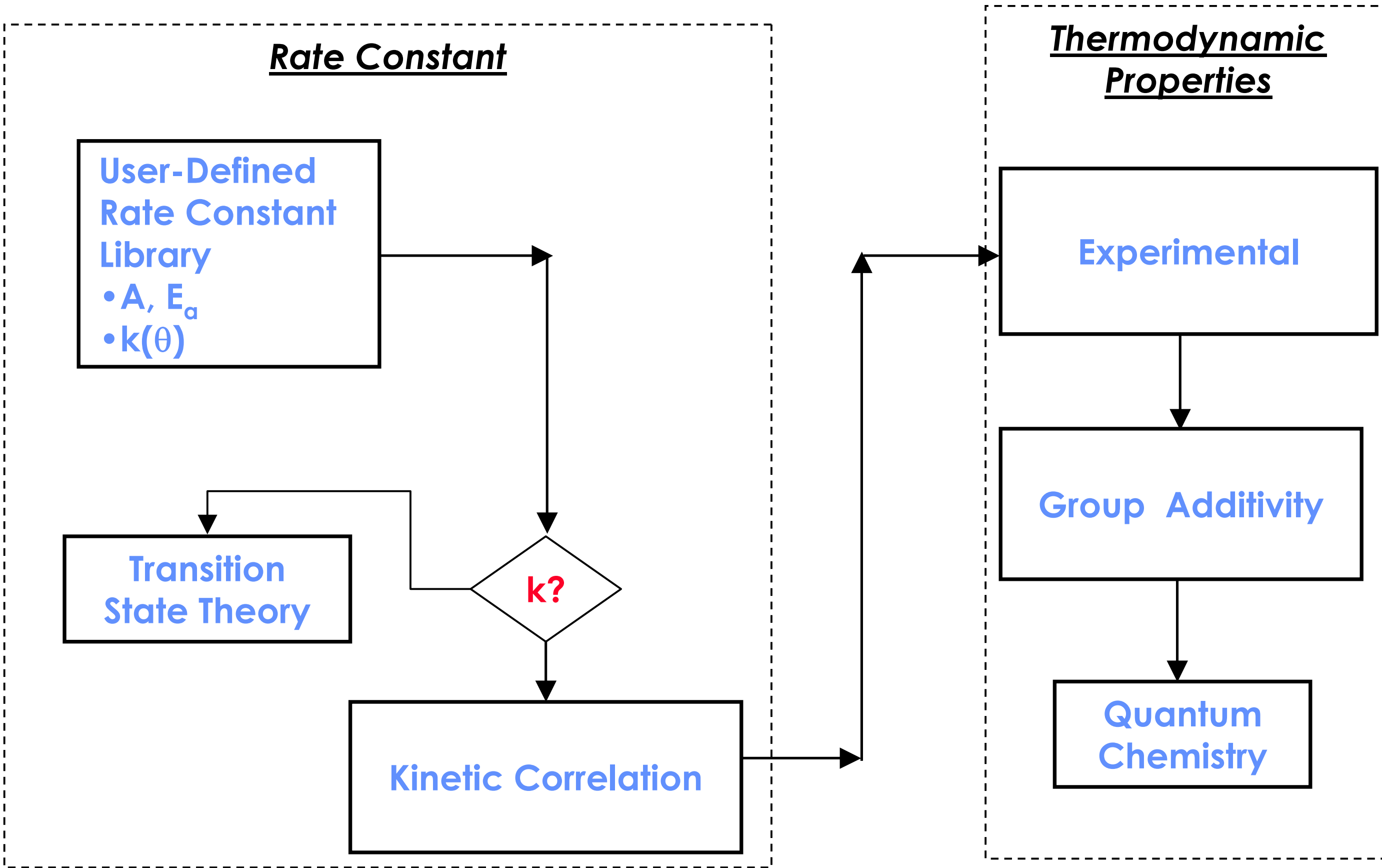
Generates comprehensive set of molecules up to a certain size but includes chemically insignificant ones

	Bound allowed	Number of species	Number of reactions	Maximum Si size
Rank allowed	0	3	2	1
	1	5	8	2
	2	12	26	4
	3	88	274	8
	4	16,279	48,168	16
Si # bound (HAC)	1	5	8	2
	2	15	36	4
	3	82	218	6
	4	701	1,794	8
	5	11,434	26,976	10
Si # bound (HAS)	8	12,527	87,938	8

Iterative Rate-Based Network Construction



Properties of Reactions and Molecules are Estimated "On-the-fly"



Kinetic Correlations Facilitate Rate Constant Estimation

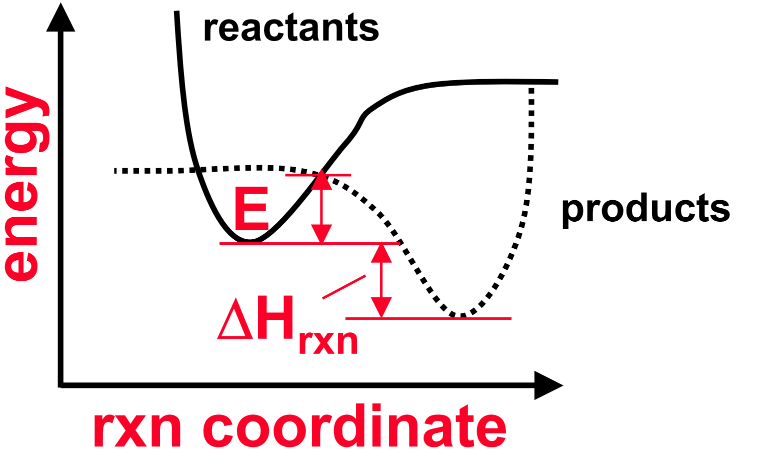
- **Kinetic correlations relate rate constants to reactivity indices**

$$\log k_i = a_i + b_i \, RI_i$$

- **Reactivity indices are easier to calculate than rate constants**
- **Thermodynamic properties are commonly used**

Evans-Polanyi relationship

$$E = E_o + \alpha \Delta H_{rxn}$$



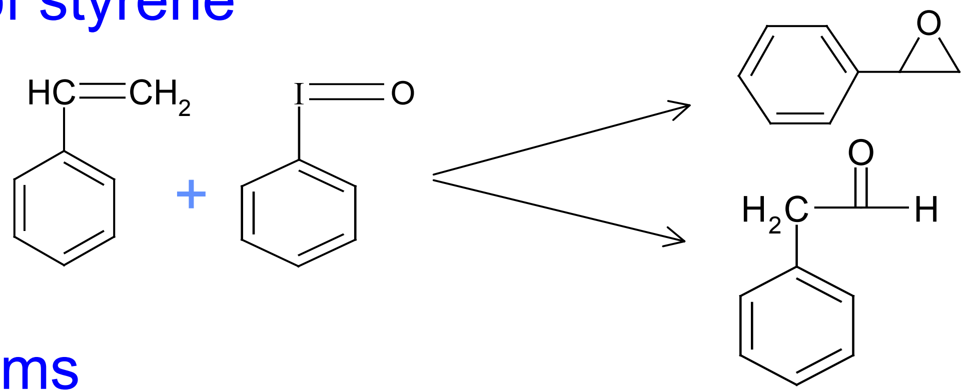
Quantum Chemical Calculations

$$H\Psi = E\Psi$$

- Information about geometries, energetics and transition states not available experimentally
- Implications for reaction network elucidation
  - Thermodynamic properties for use in kinetic correlations
  - Rate constant values using transition state theory
  - Likelihood of different reaction channels based on kinetics and thermodynamics

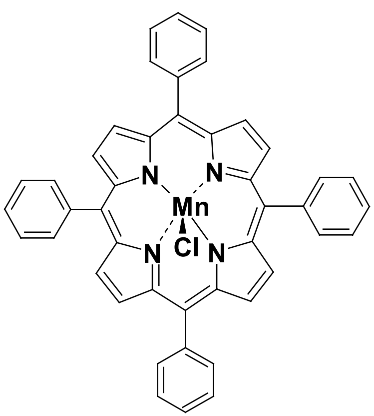
Homogeneous Olefin Epoxidation Provides Opportunities for Reaction Network Elucidation

Epoxidation of styrene

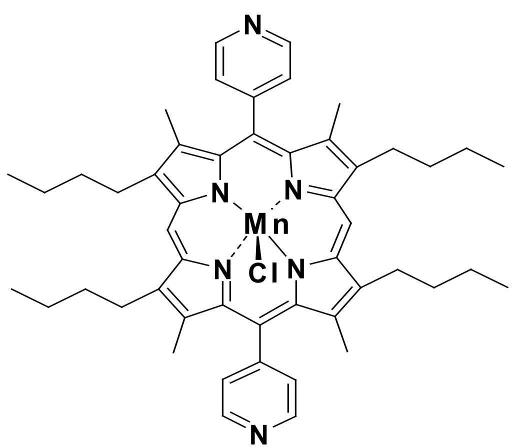


Catalyst systems

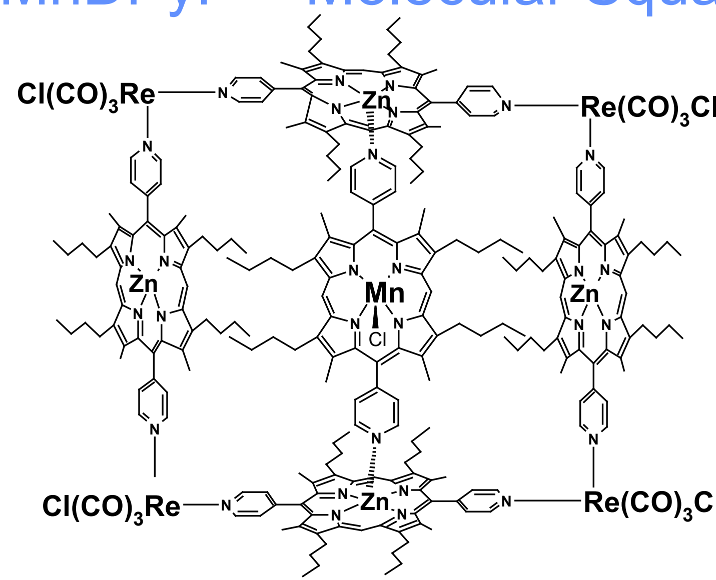
MnTPP-Cl



MnDPyP-Cl



MnDPyP + Molecular Square



Proposed Mechanism for Mn–Porphyrin Systems

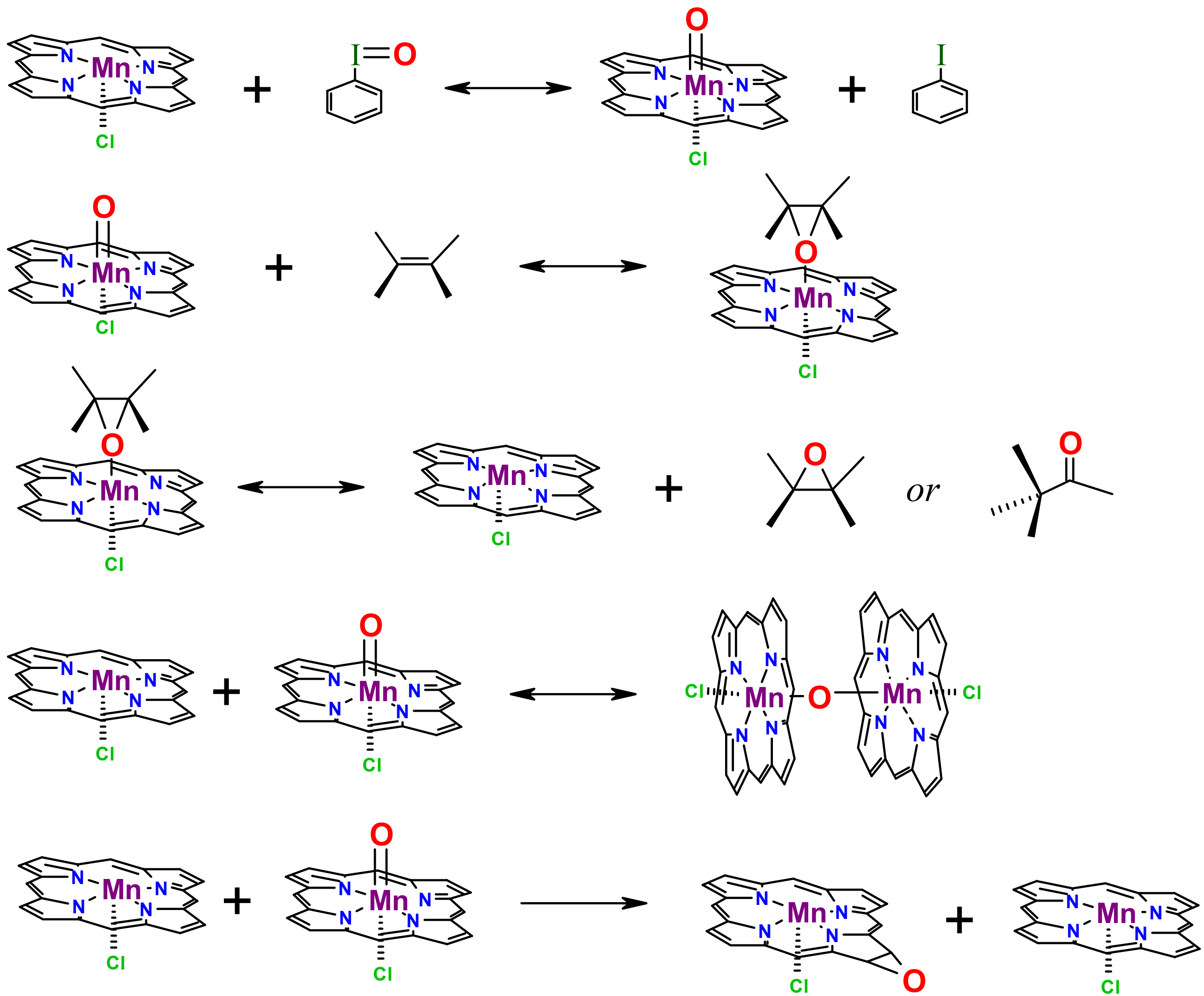
1)

2)

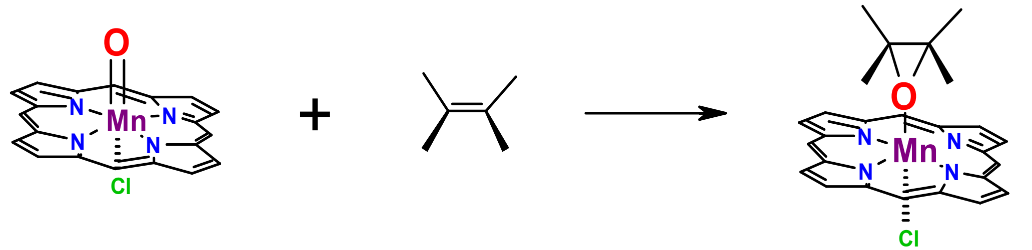
3)

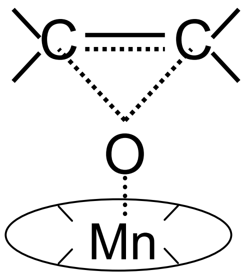
4)

5)

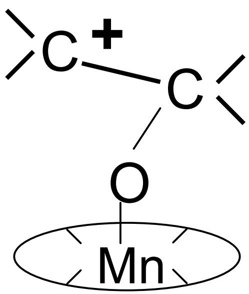


Structure of Intermediates Unresolved

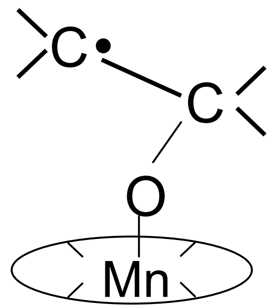




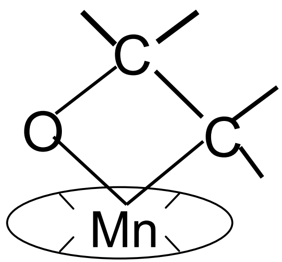
concerted [1]



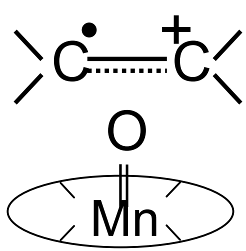
carbocation [2]



carbon radical [2]



metallaoxetane [3]

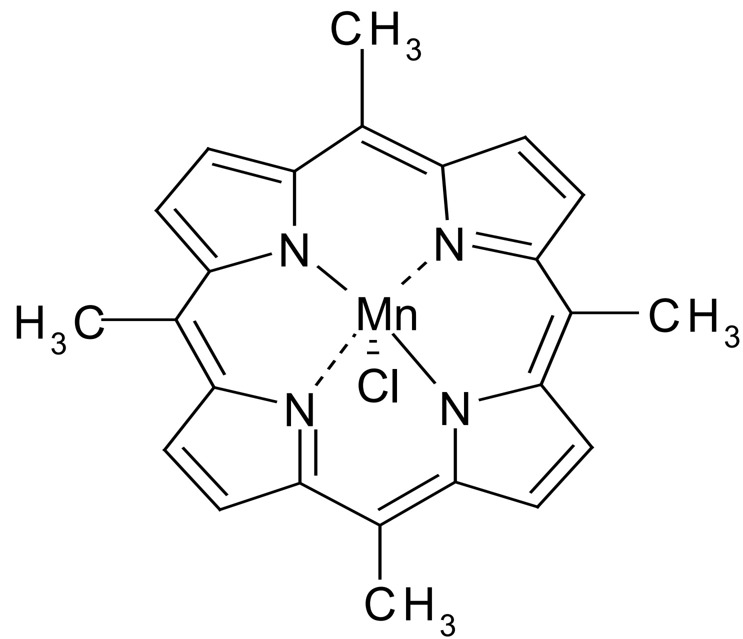


π-radical cation [4]

[1] Arasasingham et al., *J. Am. Chem. Soc.*, **1993**, *115*, 7985.
[2] Nolte et al., *J. Am. Chem. Soc.*, **1986**, *108*, 2751, and Nolte et al., J. Molec. Catal., **1985**, *31*, 271.
[3] J. Collman et al., *J. Am. Chem. Soc.*, **1985**, *107*, 2000, and Collman, et al., *J. Am. Chem. Soc.*, **1990**, *112*, 1980.
[4] He et al., *J. Am. Chem. Soc.*, **1991**, *113*, 9828.

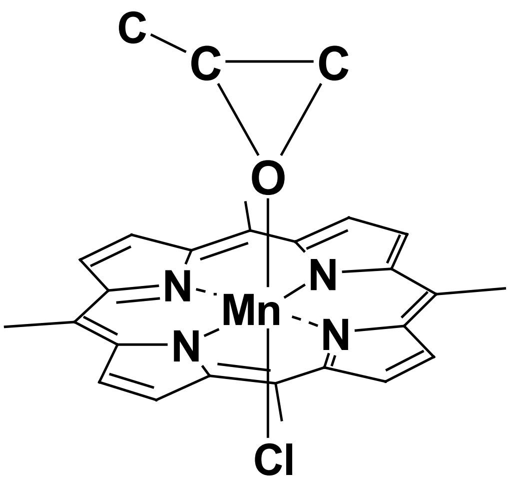
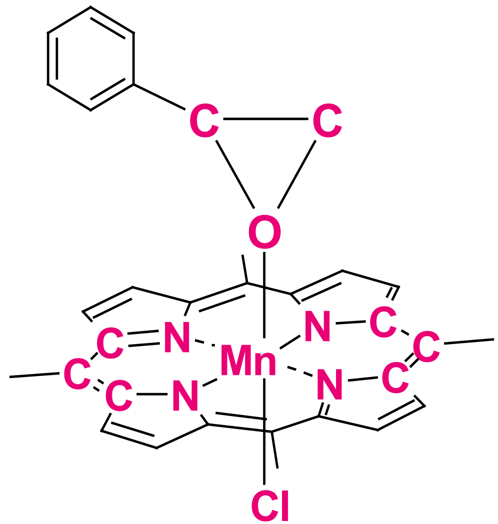
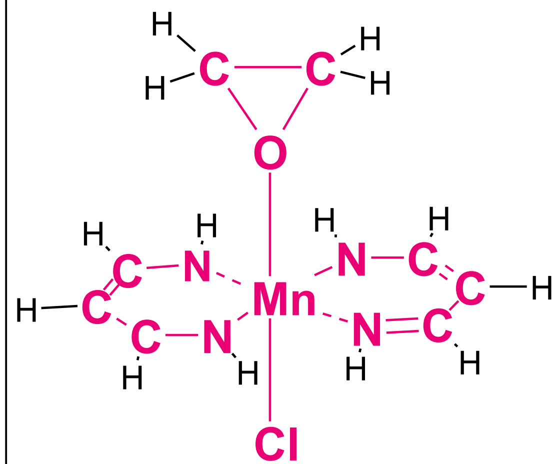
Quantum Chemical Calculations Using DFT

- Model porphyrin examined using density functional theory
  - PW91 exchange-correlation functional
  - Basis set: LANL2DZ
  - Full geometry optimization with no symmetry constraints



Mn-porphyrin system used for calculations

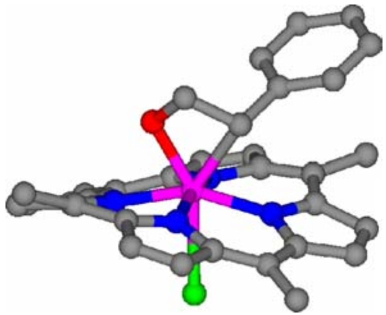
Size of Catalyst Demands Hybrid Quantum/Classical Approach: ONIOM

Model for DFT	Model for ONIOM	
		

high level =DFT with PW91 functional and LANL2DZ basis set
low level = UFF force field

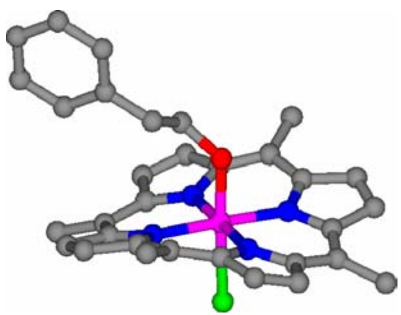
Oxidized Porphyrin Intermediates

ONIOM



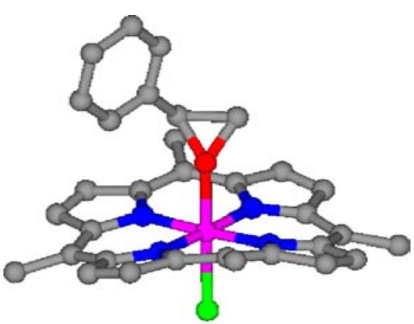
metallaoxetane

ΔE = 154.5 kJ/mol



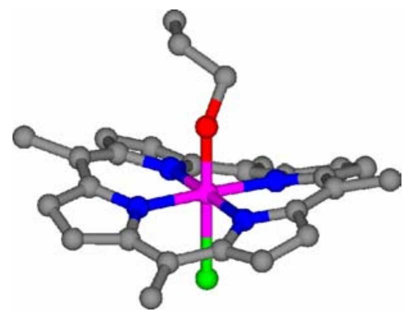
radical

ΔE = 12.6 kJ/mol



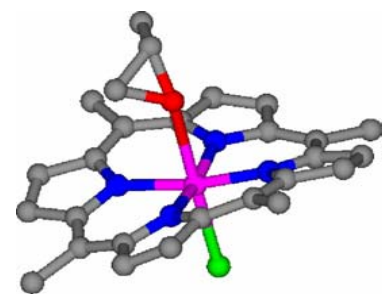
concerted

DFT



radical

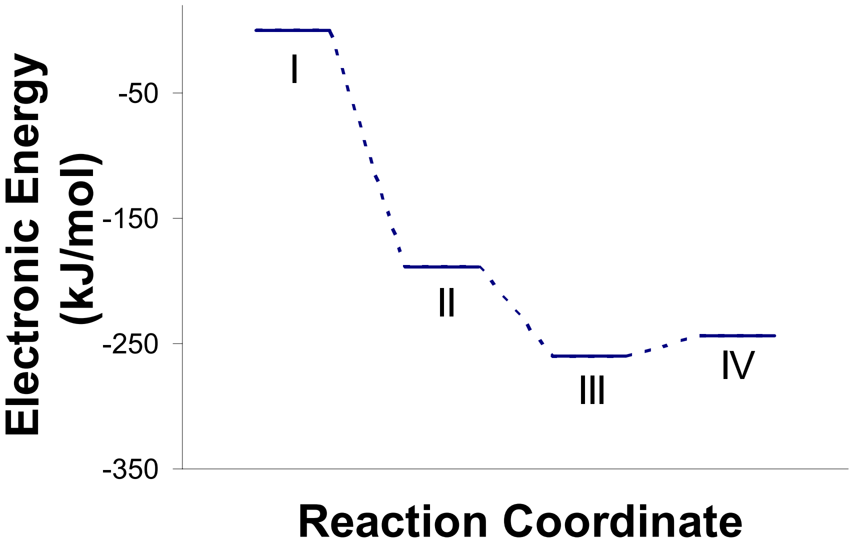
ΔE = 10.8 kJ/mol



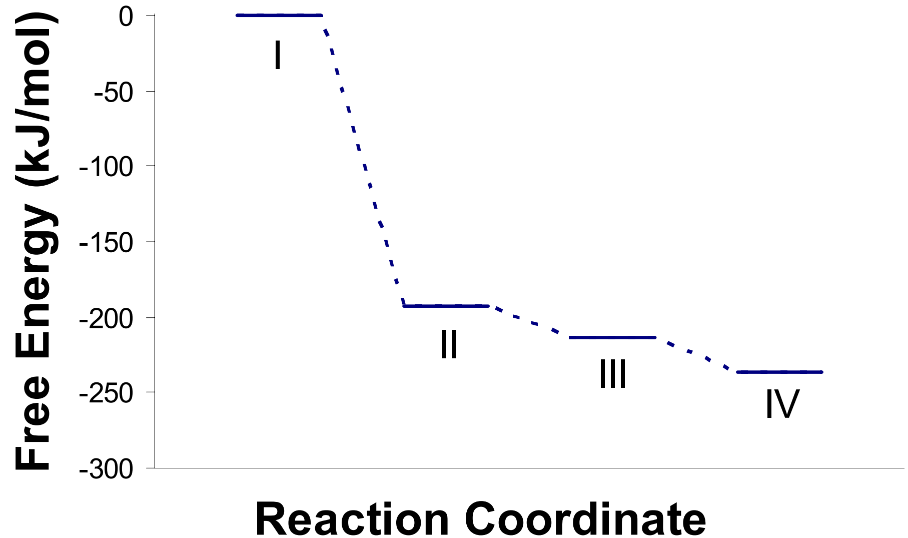
concerted

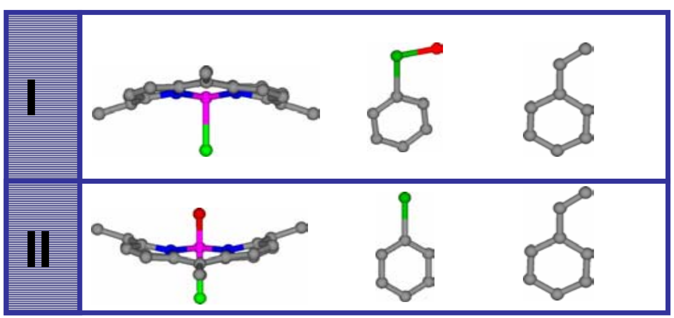
Energies for Reaction Pathway

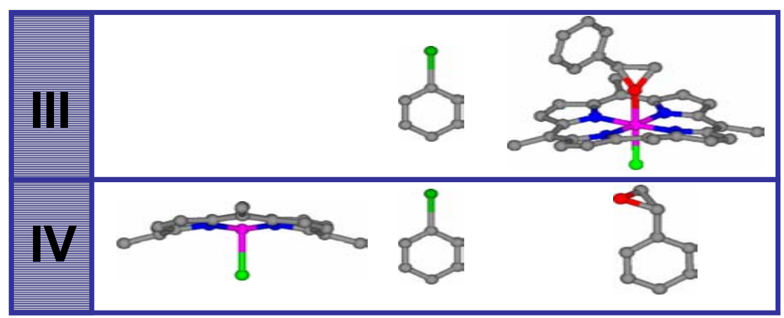
Electronic Energy



Free Energy



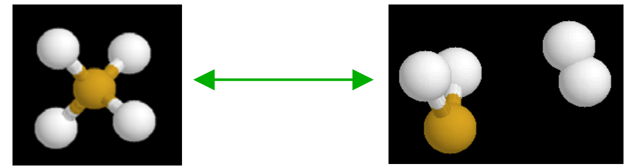




Reaction Families for Silicon Nanoparticle Production

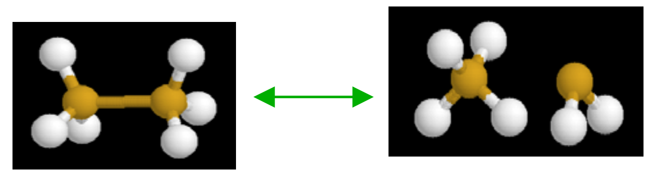
Hydrogen elimination / Hydrogen addition

$(\text{silane})_n \leftrightarrow (\text{silylene})_n + H_2$



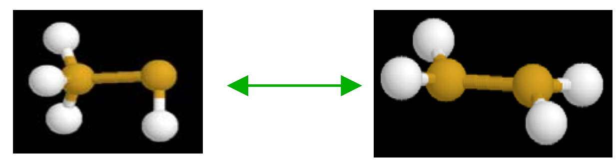
Silylene elimination / Silylene addition

$(\text{silane/silylene/silene})_n \leftrightarrow$
$(\text{silane/silylene/silene})_m + (\text{silylene})_{n\text{-}m}$



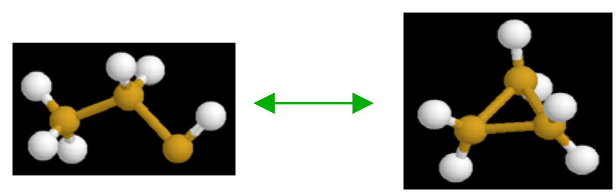
Silylene to silene isomerization

$(\text{silylene})_n \leftrightarrow (\text{silene})_n$



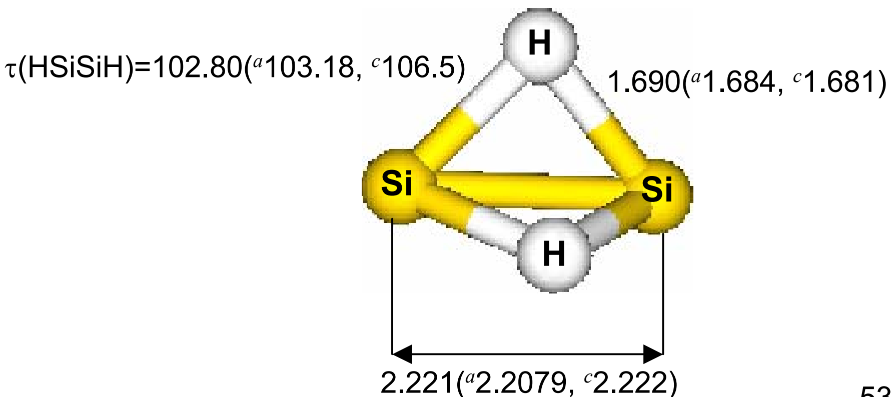
Ring formation / Ring opening

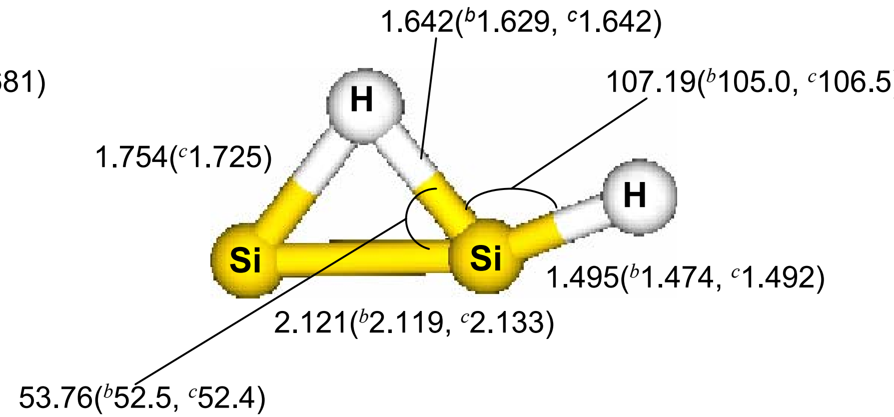
$(\text{acyclic silylene})_n \leftrightarrow (\text{cyclic silane})_n$

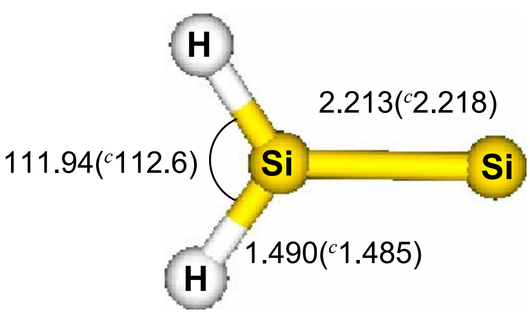


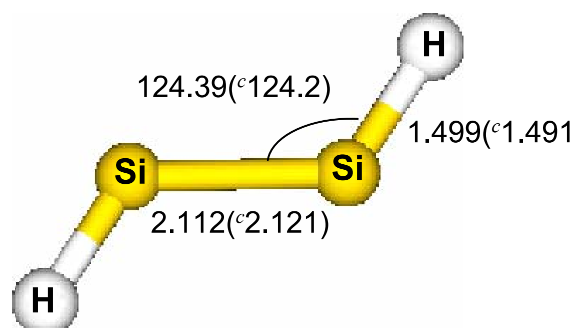
Si_2H_2 is a Critical Intermediate for Growth

Four likely structures of Si_2H_2







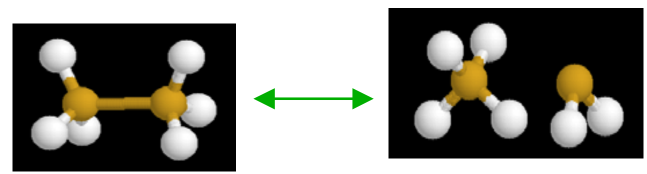


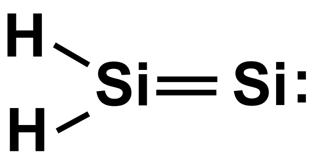
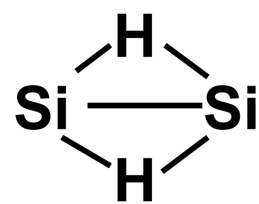
G3//B3LYP geometries

Can Addition of Si_2H_2 into Si-H Bonds Be Described Using the Silylene Addition Reaction Family?

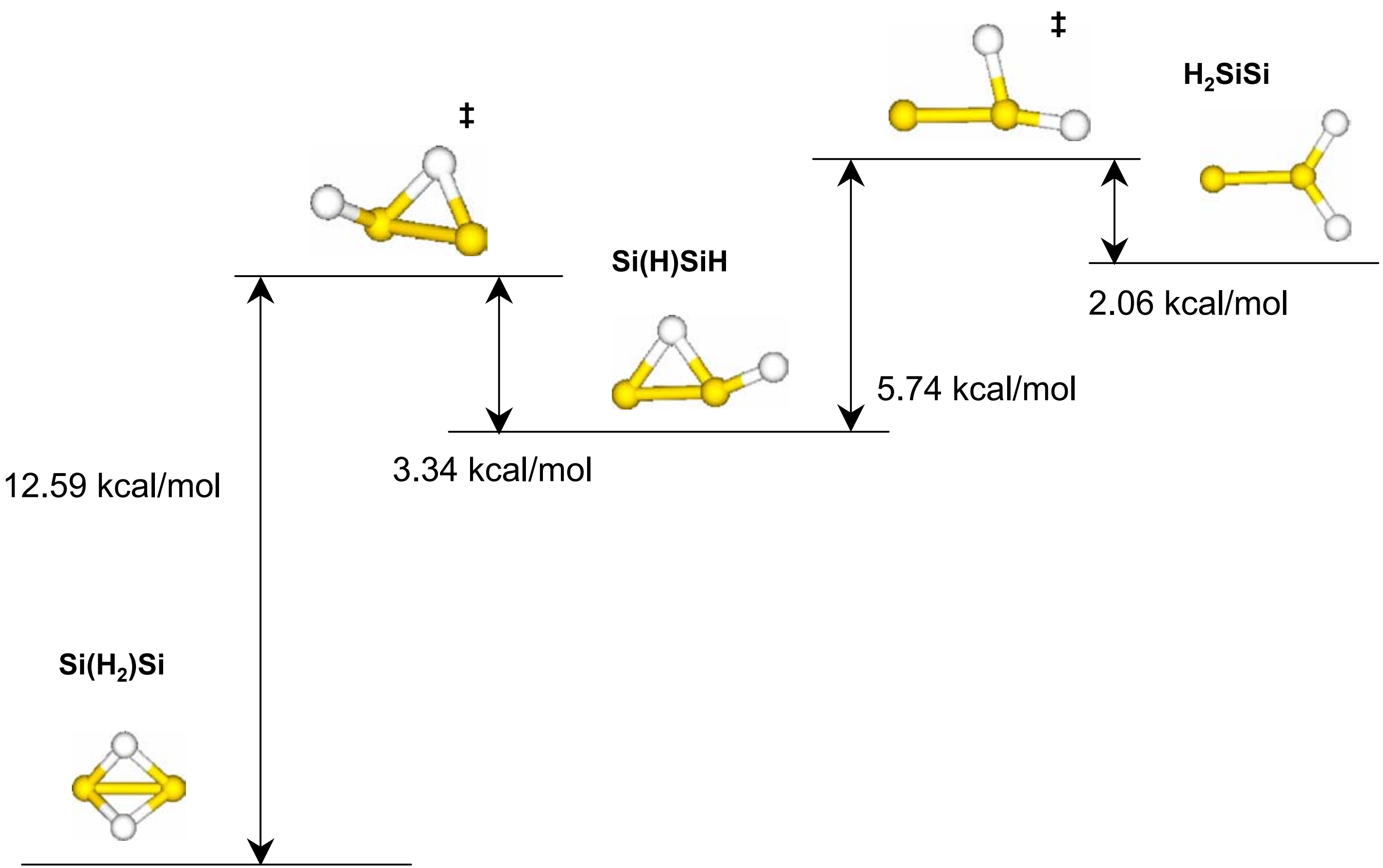
Silylene elimination / Silylene addition

$(\text{silane/silylene/silene})_n \leftrightarrow$
$(\text{silane/silylene/silene})_m + (\text{silylene})_{n\text{-}m}$

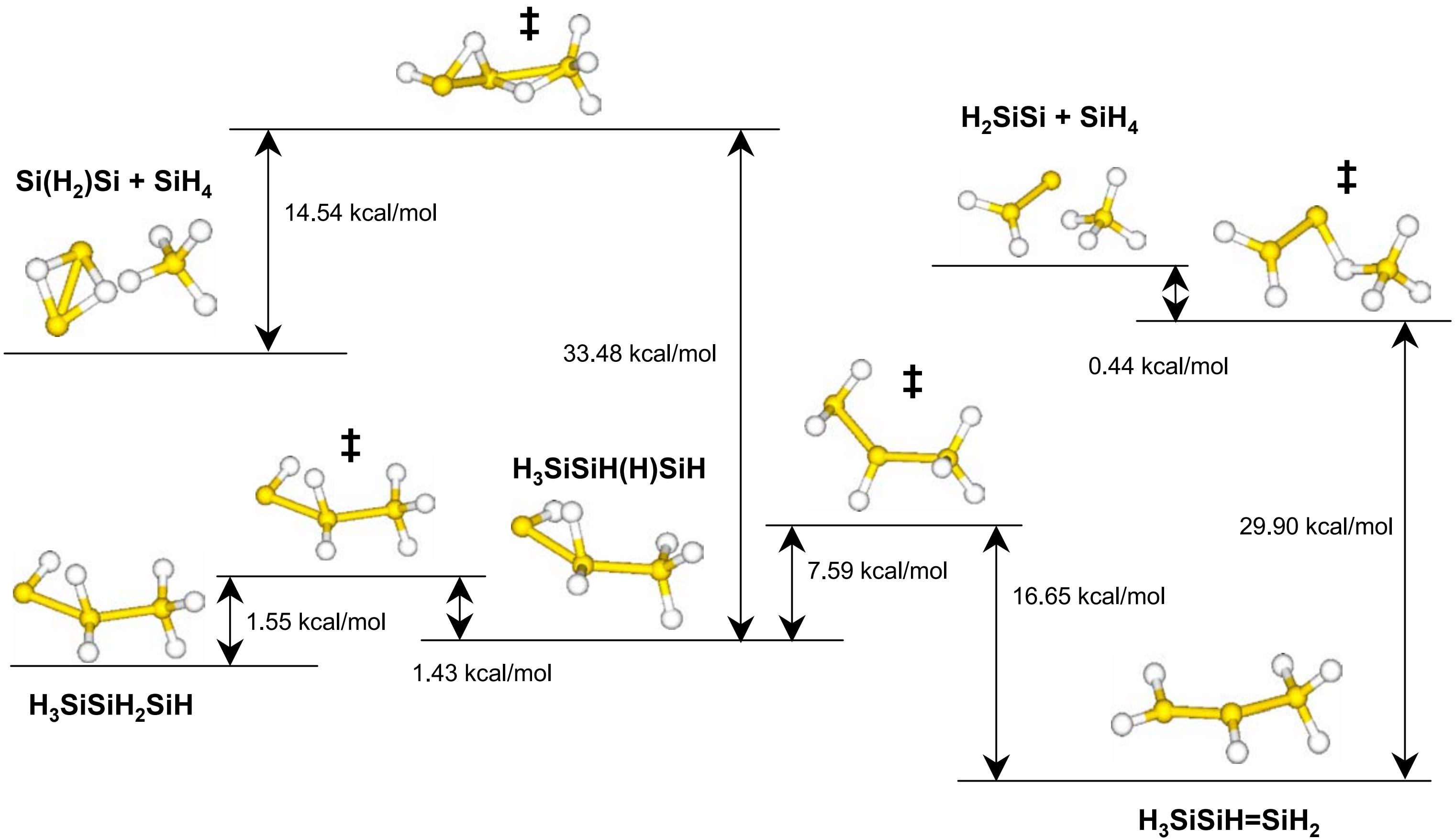


Is  **with the properties of**  **a good representation of the detailed chemistry of Si_2H_2 isomers?**

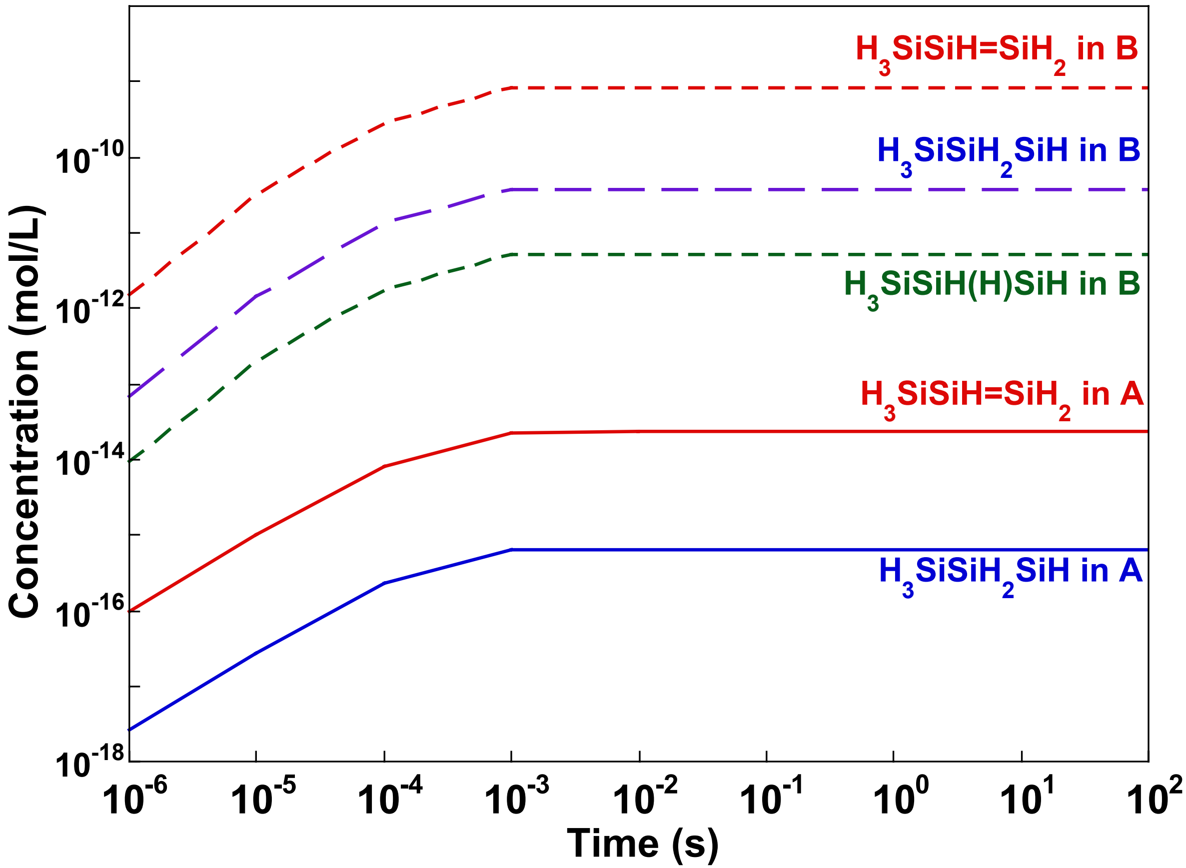
Potential Energy Surface for Si_2H_2 Isomerization from G3//B3LYP



Potential Energy Surface for Reactions between Si_2H_2 and SiH_4 from G3//B3LYP



Comparison of Product Yields for Simplified and Detailed Mechanisms



Mechanism B: Detailed Mechanism

Mechanism A: Simplified Mechanism

Summary

- **Automated network generation** can be used to build complex reaction networks for a wide range of chemistries
- Reaction networks require specification of species, reactions, thermodynamic properties, and kinetic parameters
- **Quantum chemical calculations** are increasingly valuable in reaction network elucidation

Acknowledgments

Funding

National Science Foundation
Environmental Protection Agency
Department of Energy

Collaborators

Hsi-Wu Wong	Mark Swihart	Xuegeng Li
Shumaila Khan	Qizhi Zhang	
Vassily Hatzimanikatis	Chunhui Li	Chris Henry
Joanna Gonzalez	Matt Jankowski	Justin Ionita