Reaction Networks, A Simplification Attempt

Dmitrij Rappoport
Department of Chemistry and Chemical Biology, Harvard University

Development from hieroglyphics to Proto-Sinaitic, Phoenician, Aramaic, and Greek letters

What?

I. Bilibin, *Ivan Tsarevitch Catching the Firebird's Feather* (1899)

Formose Reaction

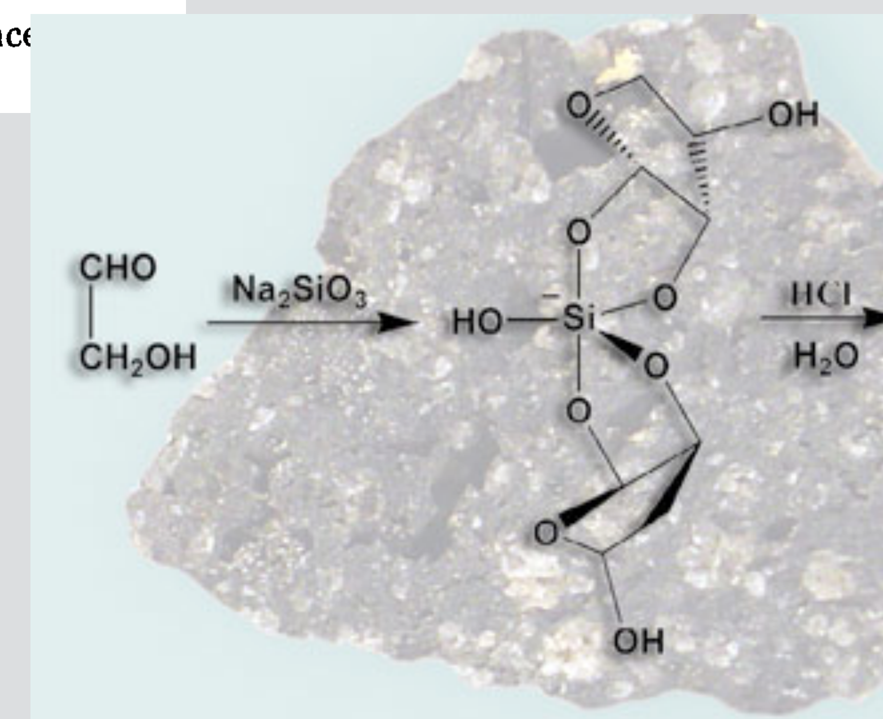
CHIMIE ORGANIQUE. — *Formation synthétique d'une substance sucrée ;* *par* **M. A. Boutlerow.**

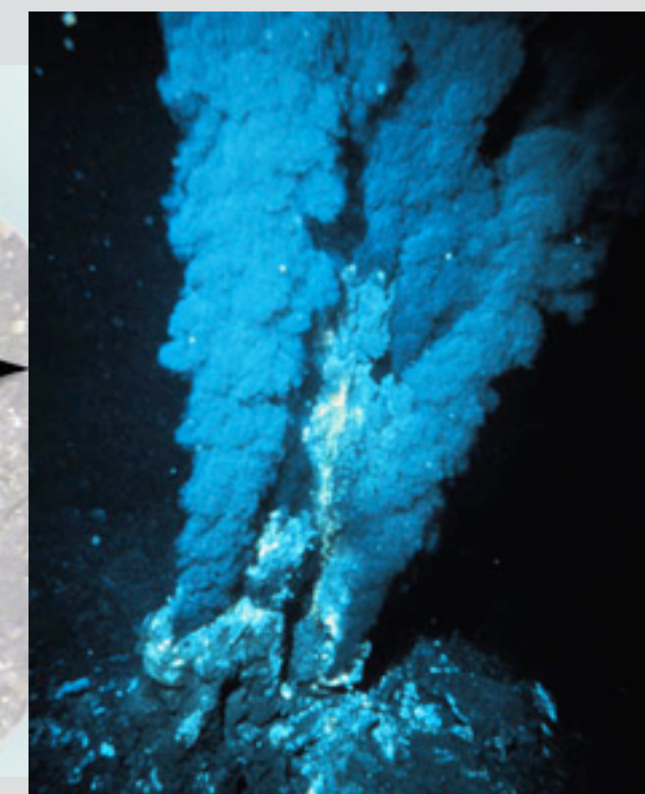
(Commissaires, MM. Balard, Peligot.)

« Le dioxyméthylène ~~$C^2H^4O^2$~~ que j'ai obtenu en traitant l'iodure de méthylène par l'oxalate d'argent se dissout facilement, surtout si l'on chauffe dans des solutions de potasse ou de soude étendues, dans l'eau de baryte ou de chaux, et ne tarde pas à subir une transformation totale sous l'influence de ces réactifs. En faisant bouillir le dioxyméthylène avec de l'eau de chaux, on voit la solution incolore se colorer bientôt en jaune et prendre, finalement, une teinte jaune-brunâtre. En même temps, l'odeur caractéristique du dioxyméthylène disparaît complétement pour faire place du sucre brûlé. Aucun gaz ne se dégage dans cette réaction.

$$n\ \mathrm{CH_2O} \xrightarrow{\mathrm{OH^-}} \mathrm{C}_n\mathrm{H}_{2n}\mathrm{O}_n$$

Alexander Butlerov
(1828–1886)



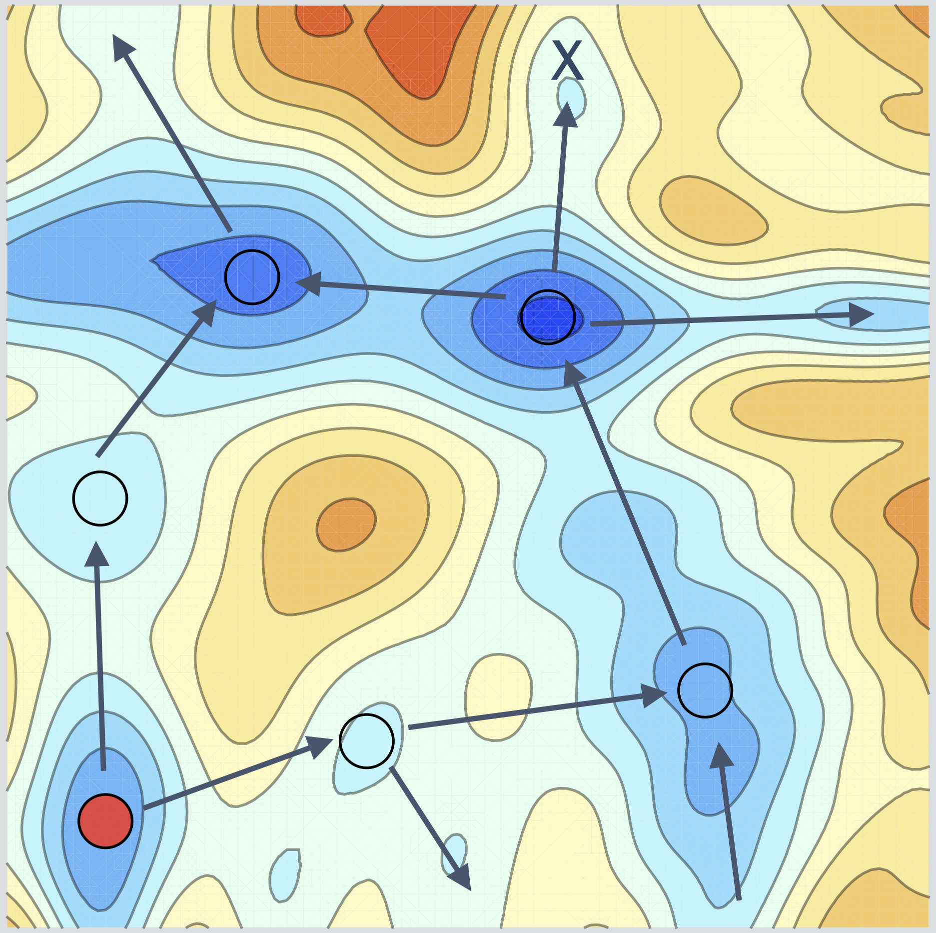


A. Boutlerow, *C. R. Acad. Sci. Paris* (1861) 53, 145.

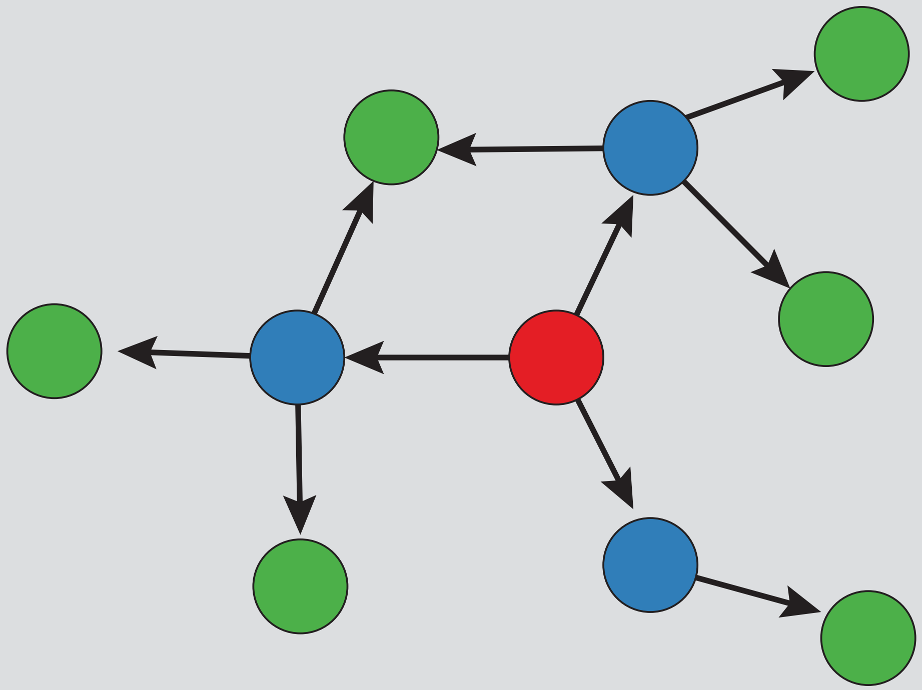
Borate/Silicate Minerals

Deep Sea Vents

Mapping Potential Energy Surfaces (PES) onto Networks

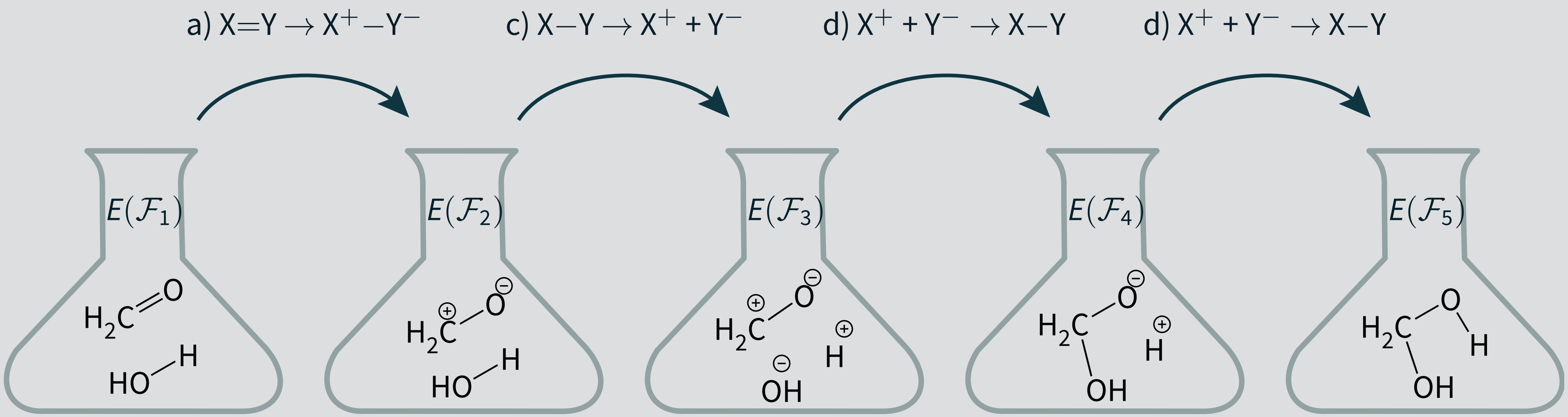


Continuous Energy Function of 3*N*–6 Coordinates



Discrete, Enumerable Network Graph

Nodes and Edges in the Transition Network Representation



- Flask $\mathcal{F}$ is a Molecular Collection Subject to Fixed Stoichiometry
- Transformation $\mathcal{F}_i \rightarrow \mathcal{F}_j$ Must Conserve Stoichiometry and Describe Making/Breaking of Specific Bond Type

Heuristically-Aided Quantum Chemistry (HAQC) Approach

Empirical rules for trial moves on PES

Quantum chemistry to accept/reject moves

Bond dissociation	Bond association
$C{-}H \rightarrow C^{-} \cdots H^{+}$	$C^{-} \cdots H^{+} \rightarrow C{-}H$
$C{-}C \rightarrow C^{+} \cdots C^{-}$	$C^{+} \cdots C^{-} \rightarrow C{-}C$
$C{-}O \rightarrow C^{+} \cdots O^{-}$	$C^{+} \cdots O^{-} \rightarrow C{-}O$
$C{-}X \rightarrow C^{+} \cdots X^{-}$	$C^{+} \cdots X^{-} \rightarrow C{-}X$
$C{-}X \rightarrow C^{-} \cdots X^{+}$	$C^{-} \cdots X^{+} \rightarrow C{-}X$
$O{-}H \rightarrow O^{-} \cdots H^{+}$	$O^{-} \cdots H^{+} \rightarrow O{-}H$
$X{-}H \rightarrow X^{-} \cdots H^{+}$	$X^{-} \cdots H^{+} \rightarrow X{-}H$
$X{-}X \rightarrow X^{-} \cdots X^{+}$	$X^{-} \cdots X^{+} \rightarrow X{-}X$
Polarization	**Depolarization**
$C{=}O \rightarrow C^{+}{-}O^{-}$	$C^{+}{-}O^{-} \rightarrow C{=}O$
$C{=}C \rightarrow C^{+}{-}C^{-}$	$C^{+}{-}C^{-} \rightarrow C{=}C$
$C{\equiv}C \rightarrow C^{+}{=}C^{-}$	$C^{+}{=}C^{-} \rightarrow C{\equiv}C$

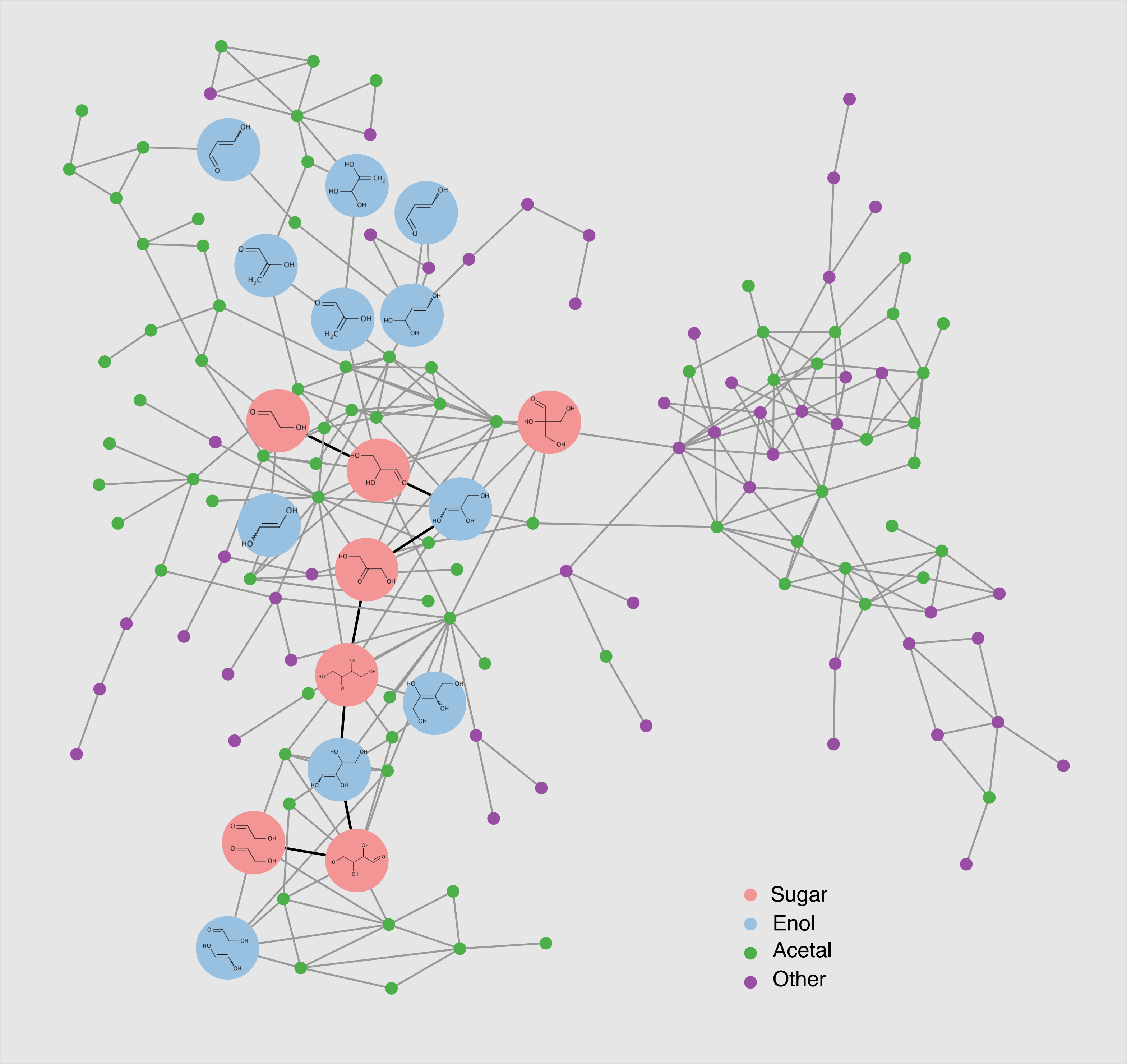
Structure Builder: UFF

Energy Calculation: PM7/COSMO (ε = 78.4)

Intermolecular Forces: None

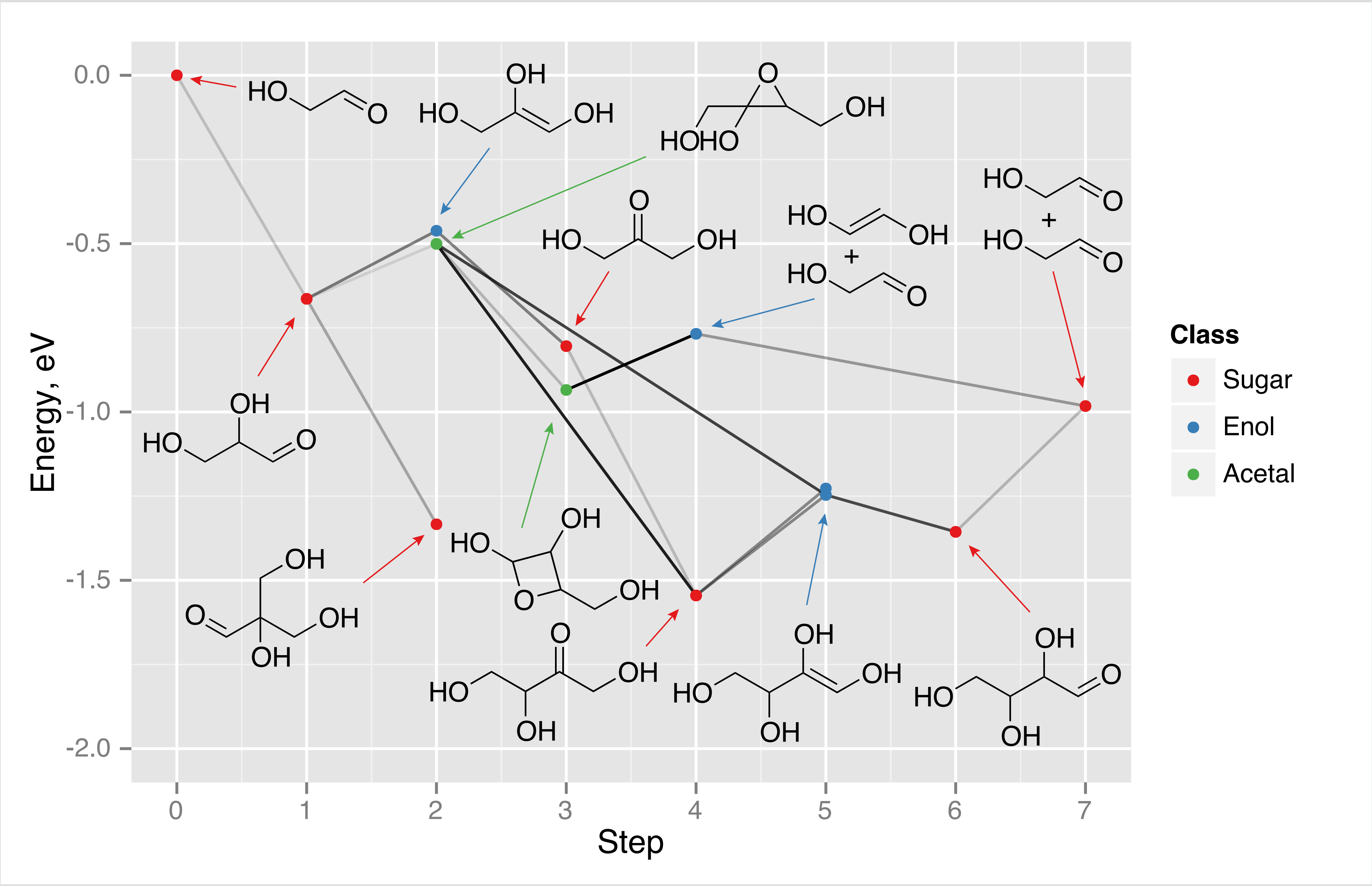
J. Tenniel, *Chess Board* (1865)

Formose Reaction Network ($C_4H_8O_4$ / Tetrose)



9 generations
149 nodes
445 edges
146 molecules

Sugar Formation in Formose Reaction



Aldol condensation/aldose–ketose isomerizations are major pathways

Pros & Cons

- Unambiguous Mapping from PES to Network
- Well-Defined Hierarchy of Networks
- Know How to Analyze and Search in Networks
- Any Symbolic Chemistry Representable

- Needs Heuristics for Kinetic Feasibility
- How to Represent Concerted Reaction Mechanisms?
- Lumps Together Molecular Conformations
- Exponential Growth with # Atoms, # Rules

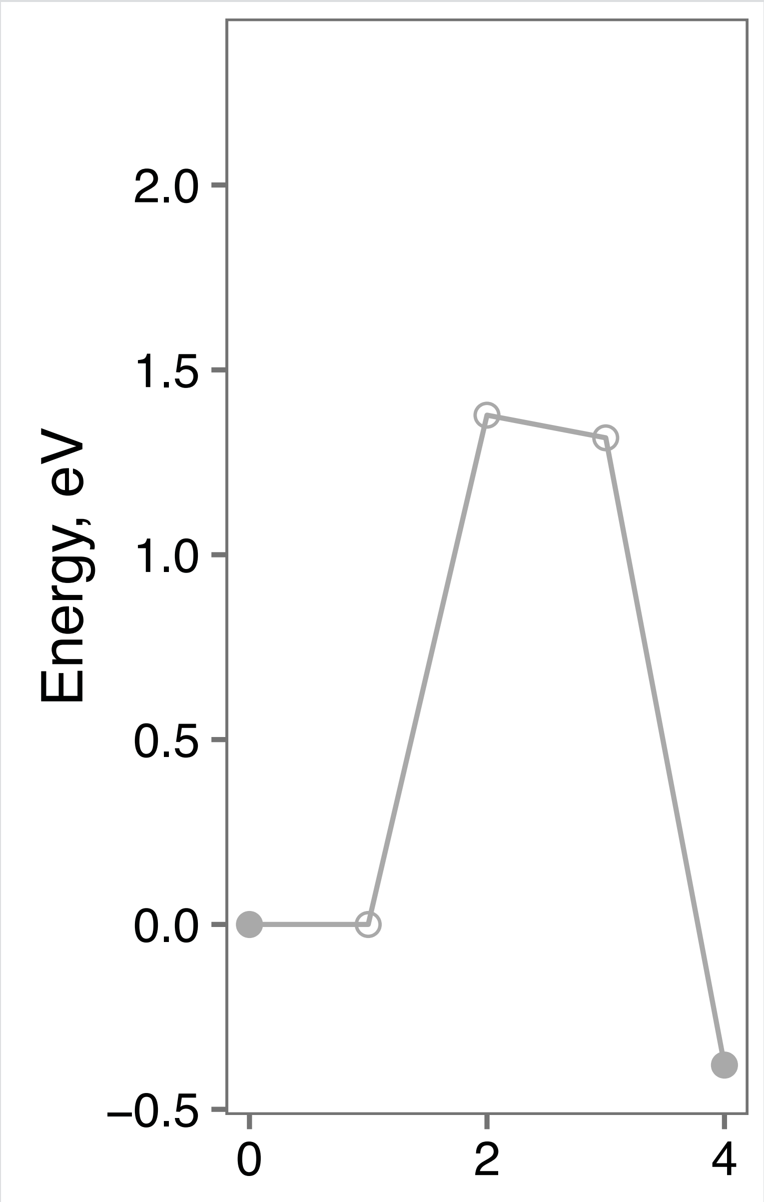
A. Rackham, *The Town Musicians of Bremen* (1909)

How?

V. Ermolaeva, *Monkey and Spectacles* (1929)

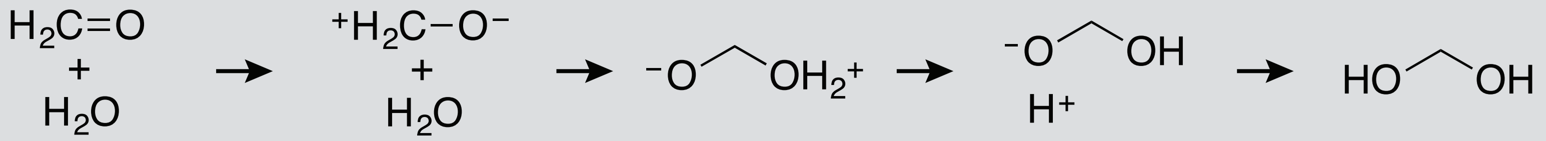
L. L. Brooke, *Humpty Dumpty* (1914)

Heuristic Kinetic Feasibility Criteria



- Non-Negativity $W_{K \to L} \geq 0$
- Additivity $W_{K \to L \to M} = W_{K \to L} + W_{L \to M}$
- Hammond's Postulate
- Length Penalty
- Detailed Balance

$$W_{K \to L} - W_{L \to K} = \Delta E_{K \to L} = E_L - E_K$$



Heuristic Kinetic Feasibility Criteria

- Climb Criterion

$$W_{c,\,K\to L} = \sum_{i=0}^{N} \max(E_{K_{i+1}} - E_{K_i}, 0)$$

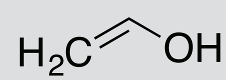
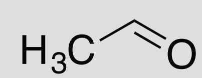
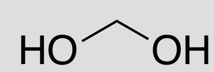
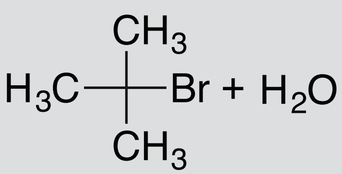
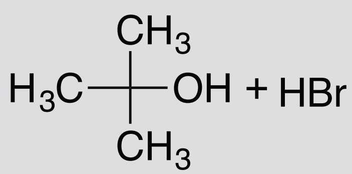
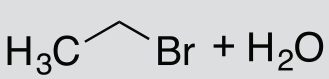
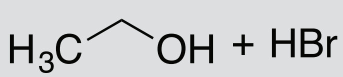
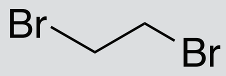
- Arc Criterion

$$W_{a,\,K\to L} = \sum_{i=0}^{N} \left[(E_{K_{i+1}} - E_{K_i})^2 + \alpha^2\right]^{1/2}$$

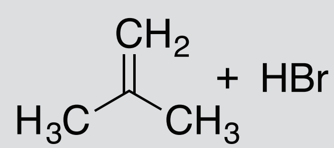
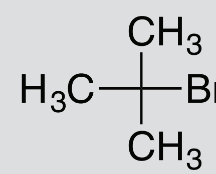
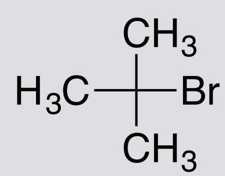
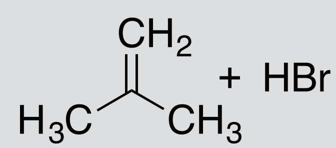
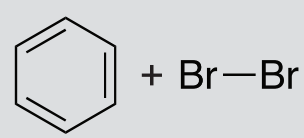
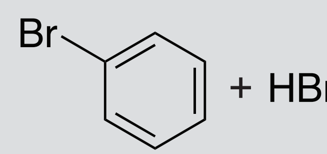
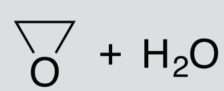
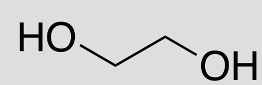
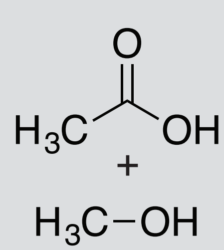
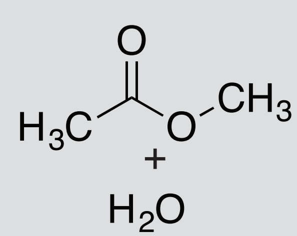
- Karc Criterion

$$W_{k,\,K\to L} = \sum_{i=0}^{N} \left[(E_{K_{i+1}} - E_{K_i})^2 + \alpha^2 \left(\frac{1 - \exp(-\kappa(E_{K_{i+1}} - E_{K_i})^2)}{1 + \exp(-\kappa(E_{K_{i+1}} - E_{K_i})^2)}\right)\right]^{1/2}$$

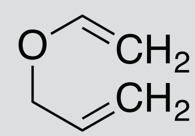
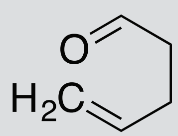
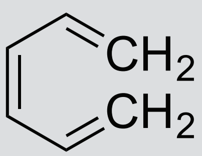
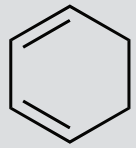
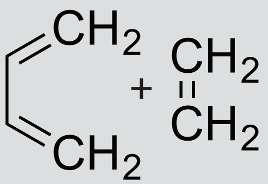
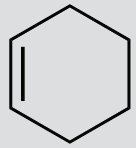
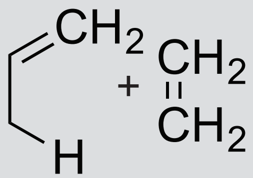
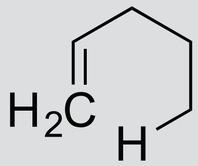
Simple Polar Reactions

	Reaction		Pathways		
	Reactants	Products	n_{f}	n_{p}	$N^{\max}$
Tautomerizaton (TA)			1	2	8
Diol formation (DI)	$H_2C{=}O + H_2O$		7	32	8
S_N1 substitution (S1)			5	94	6
S_N2 substitution (S2)			5	102	6
Br_2 addition (BA)	$H_2C{=}CH_2 + Br{-}Br$		2	14	6

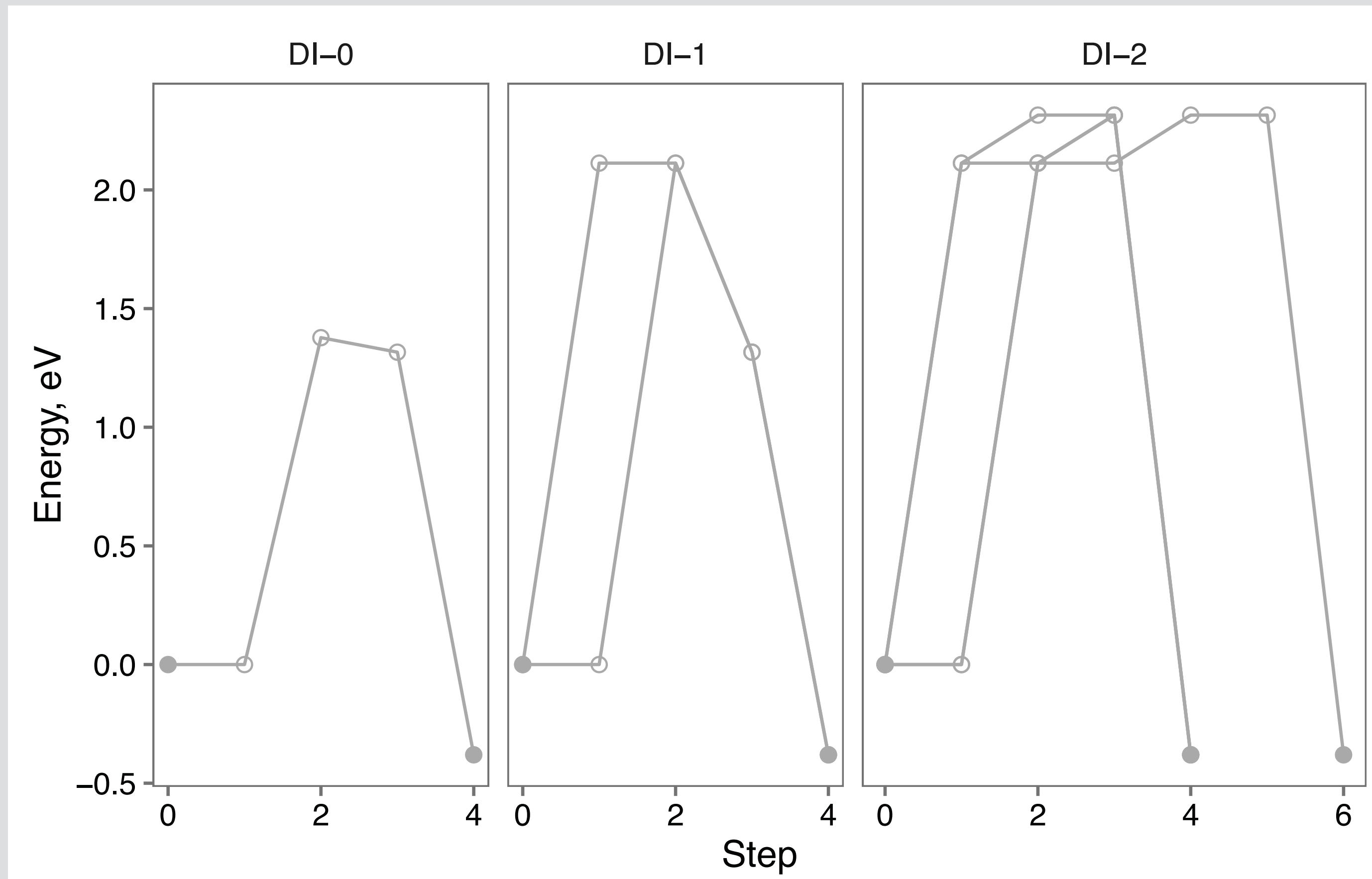
Simple Polar Reactions (Cont'd)

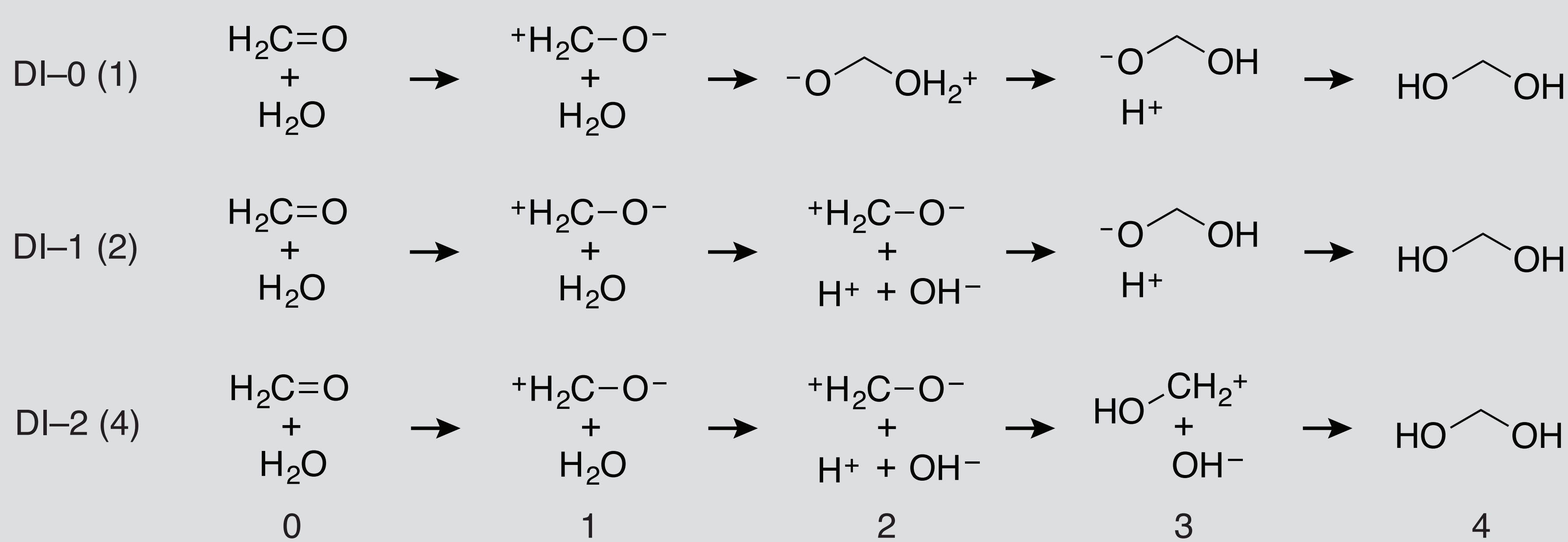
Reaction	Reactants	Products	n_f	n_p	N^max
HBr addition (HA)			3	28	6
E1 elimination (E1)			3	28	6
S_EAr substitution (SA)			4	384	6
Epoxide hydrolysis (EP)			2	61	6
Esterification (ES)			9	523	6

Pericyclic Reactions

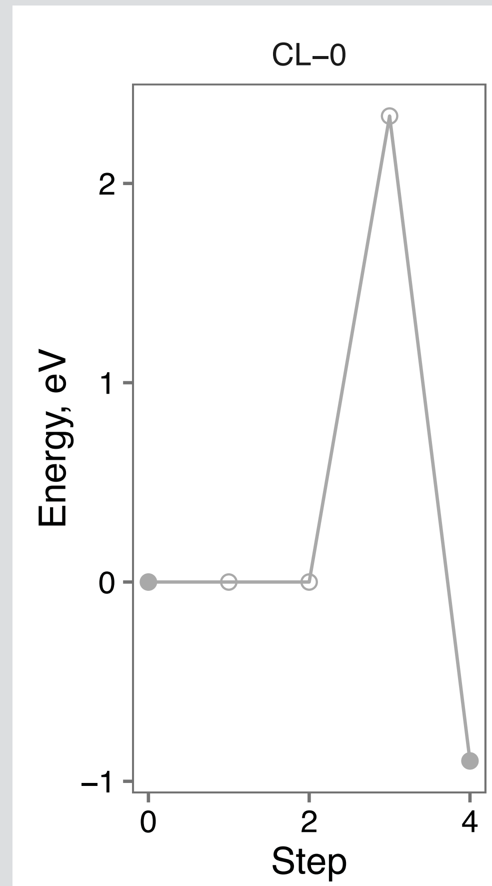
Reaction	Reactants	Products	n_{f}	n_{p}	N^{max}
Claisen rearrangement (CL)			1	4	6
6π electrocyclization (6C)			52	9685	8
Diels–Alder reaction (DA)			248	4958	8
Ene reaction (EN)			146	9193	8

Diol Formation

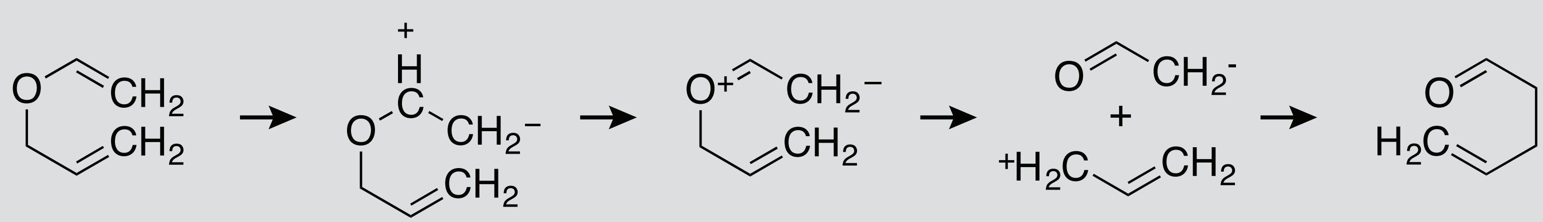




Claisen Rearrangement

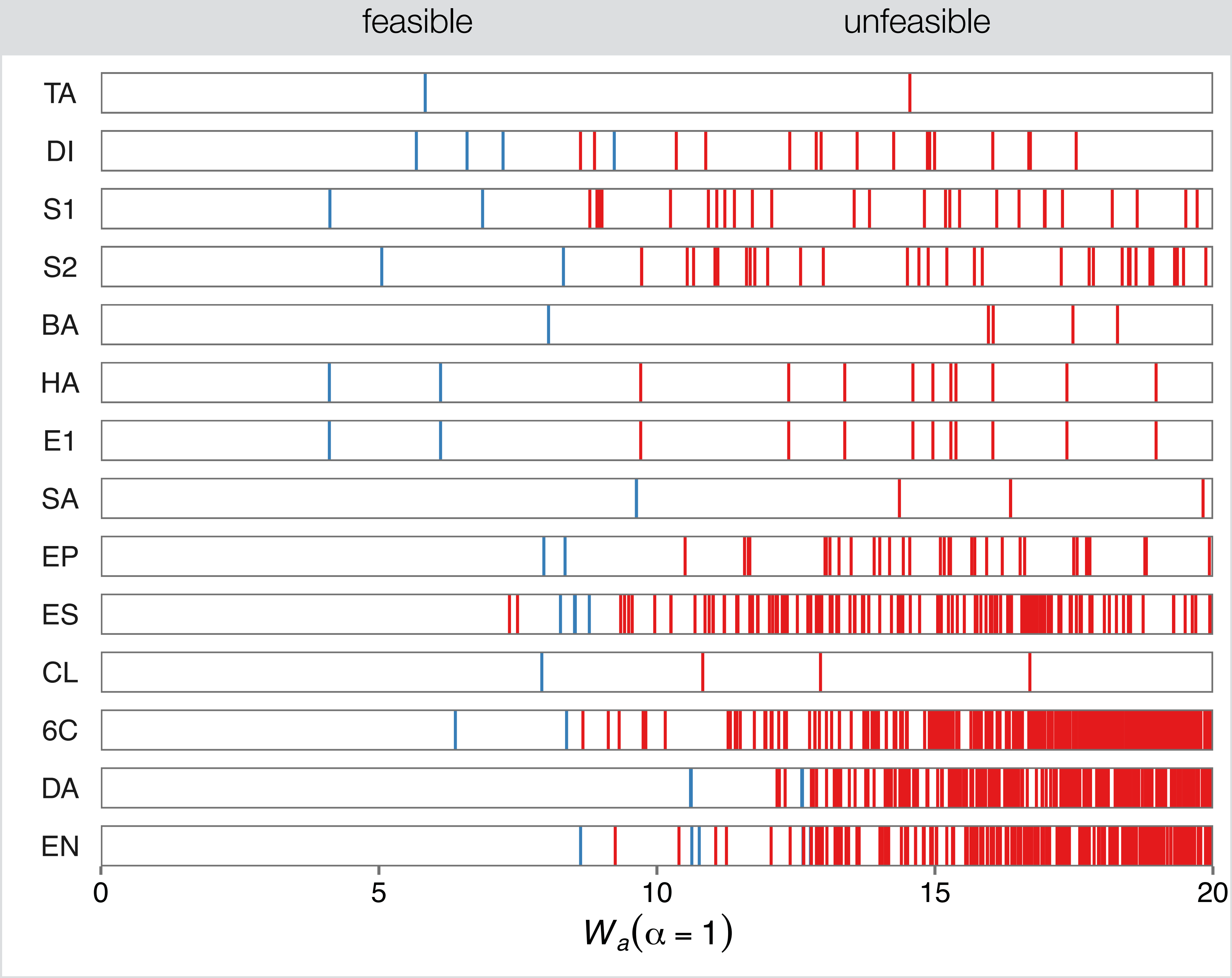


CL–0 (1)

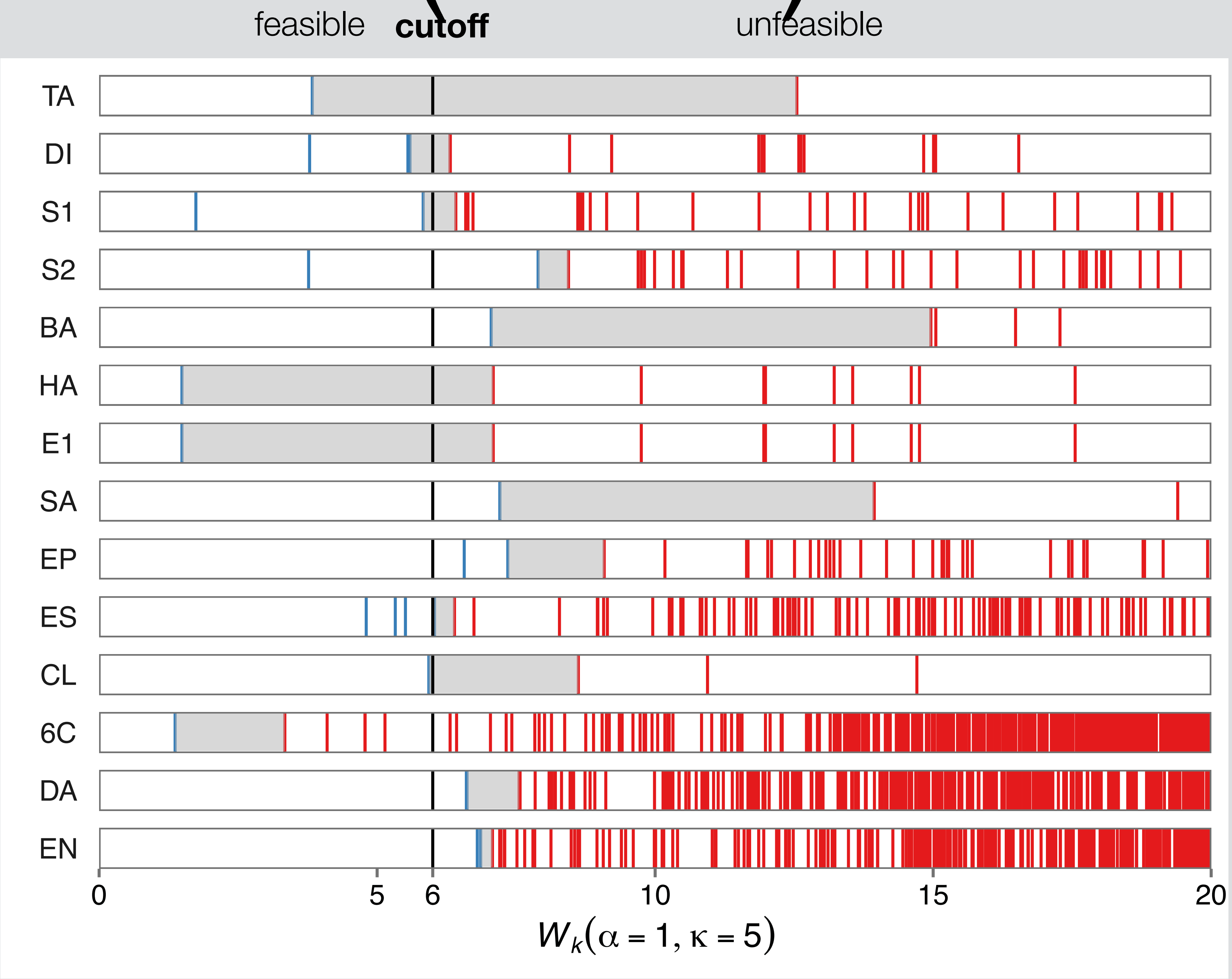


0 1 2 3 4

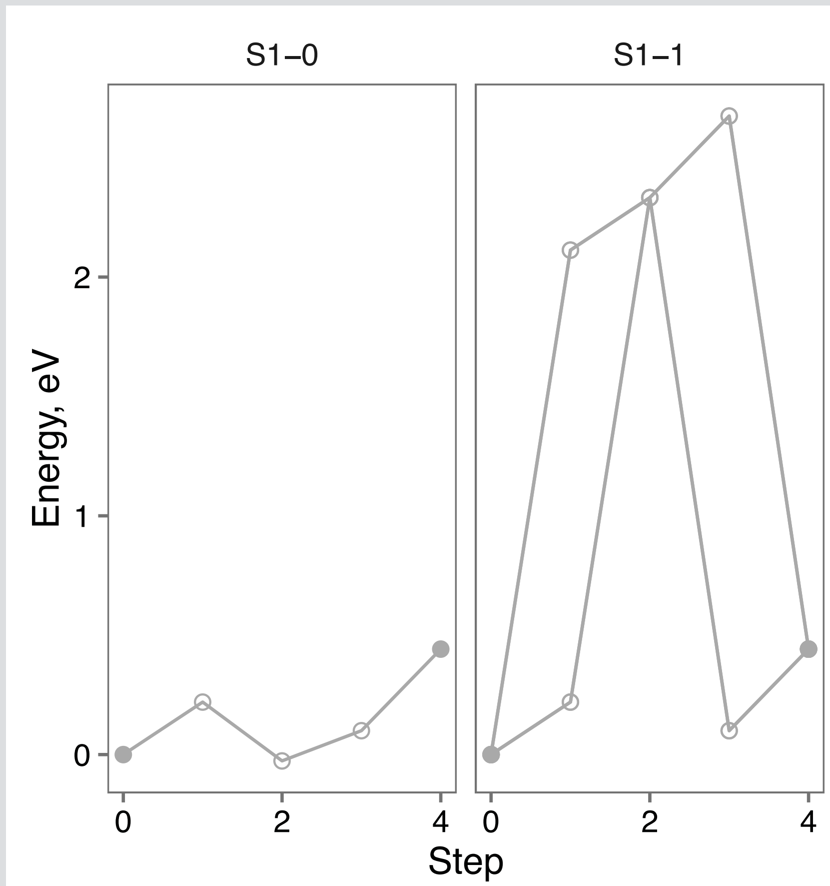
Heuristics for the Masses

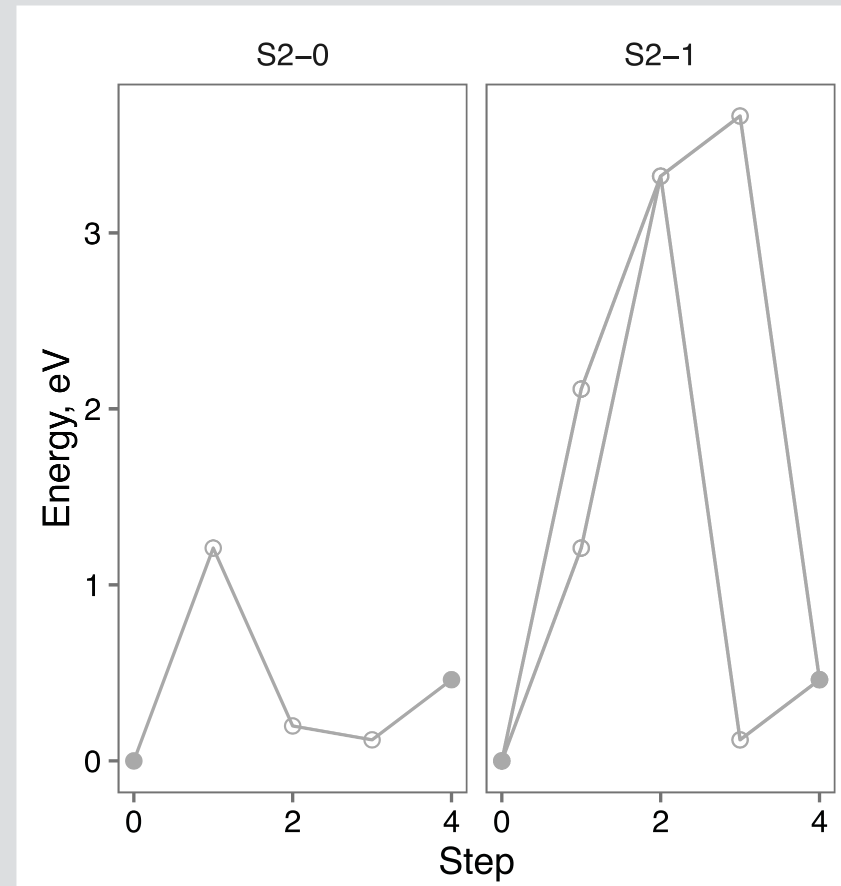


Heuristics for the Masses (Cont'd)

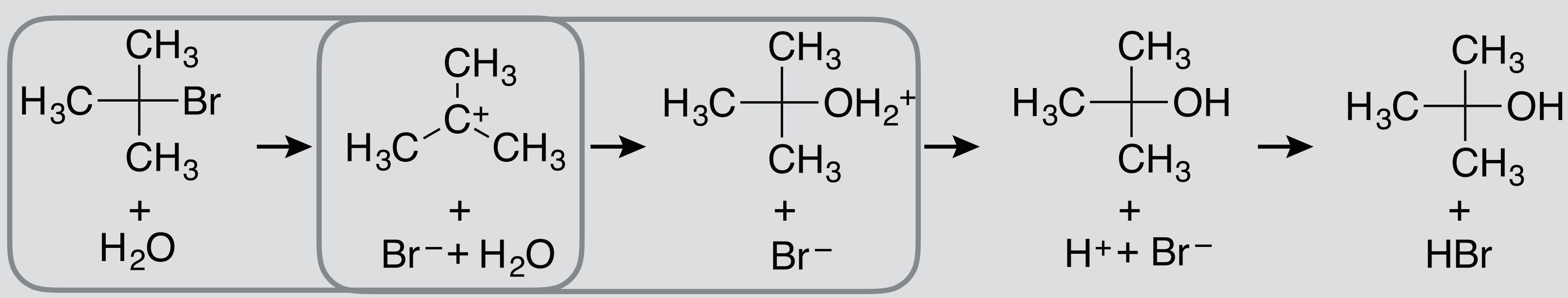


Concerted Reactions

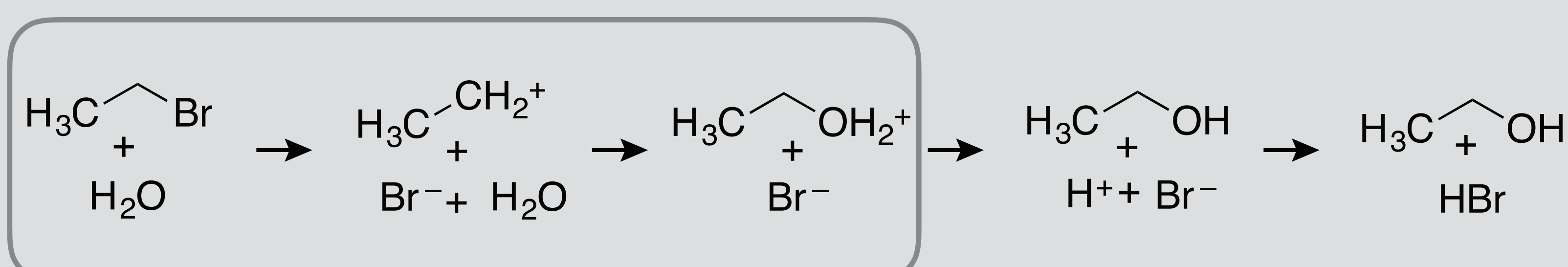




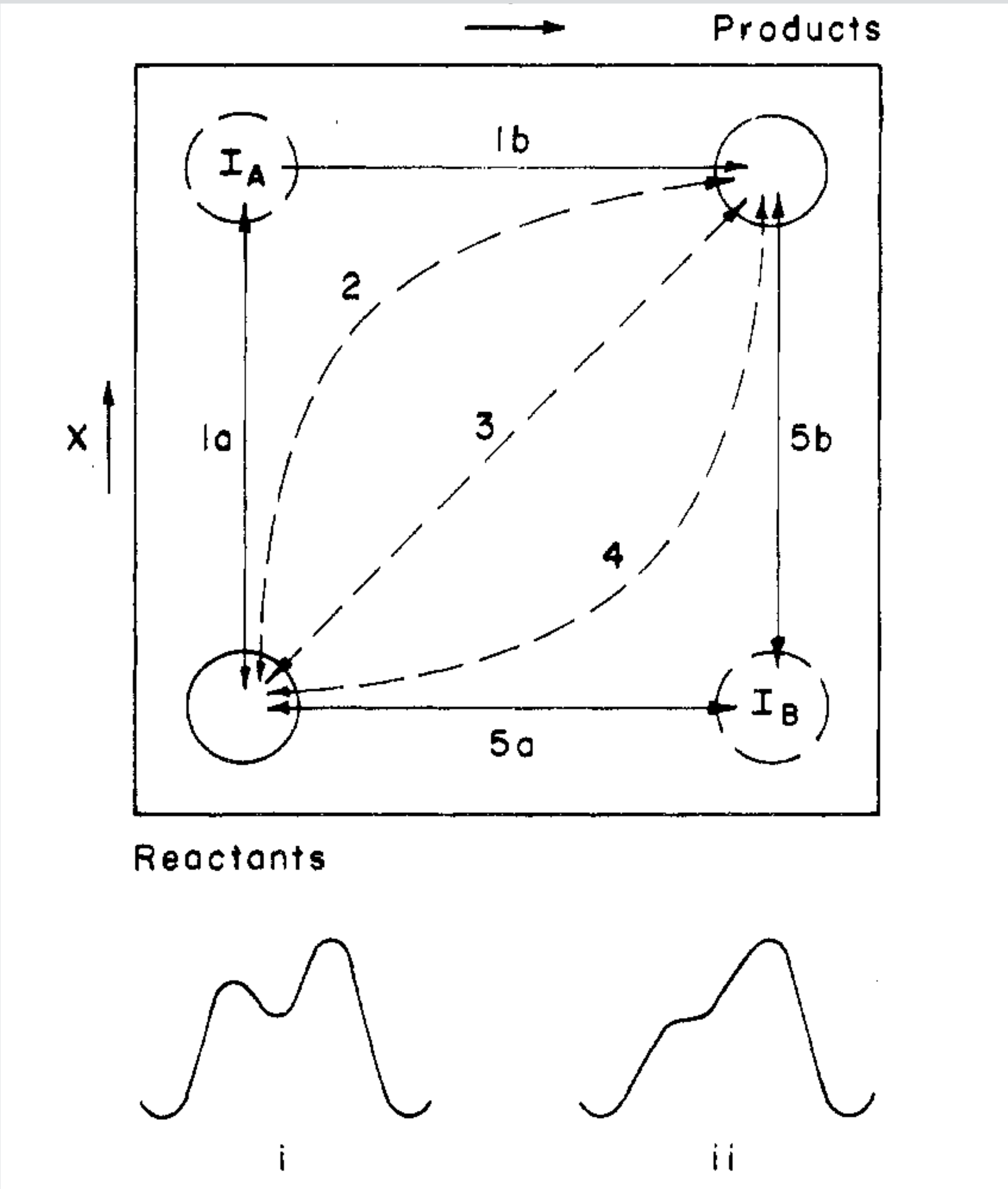
S1–0 (1)



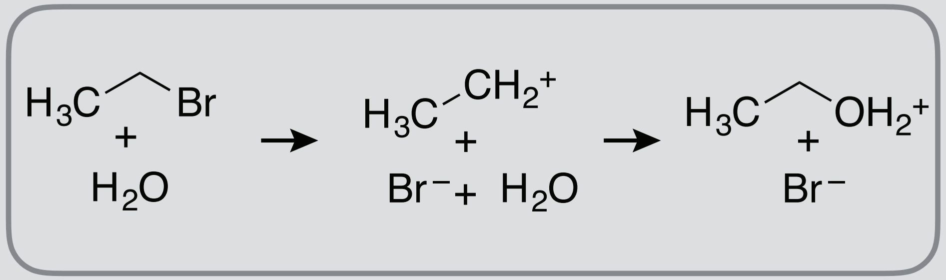
S2–0 (1)



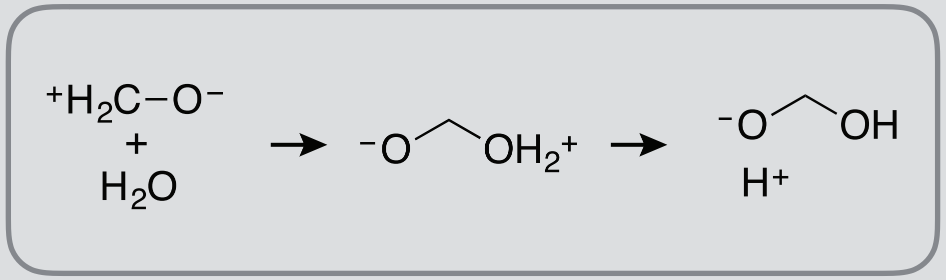
Concerted Reactions (Cont'd)



- Dissociation / Association Mechanism

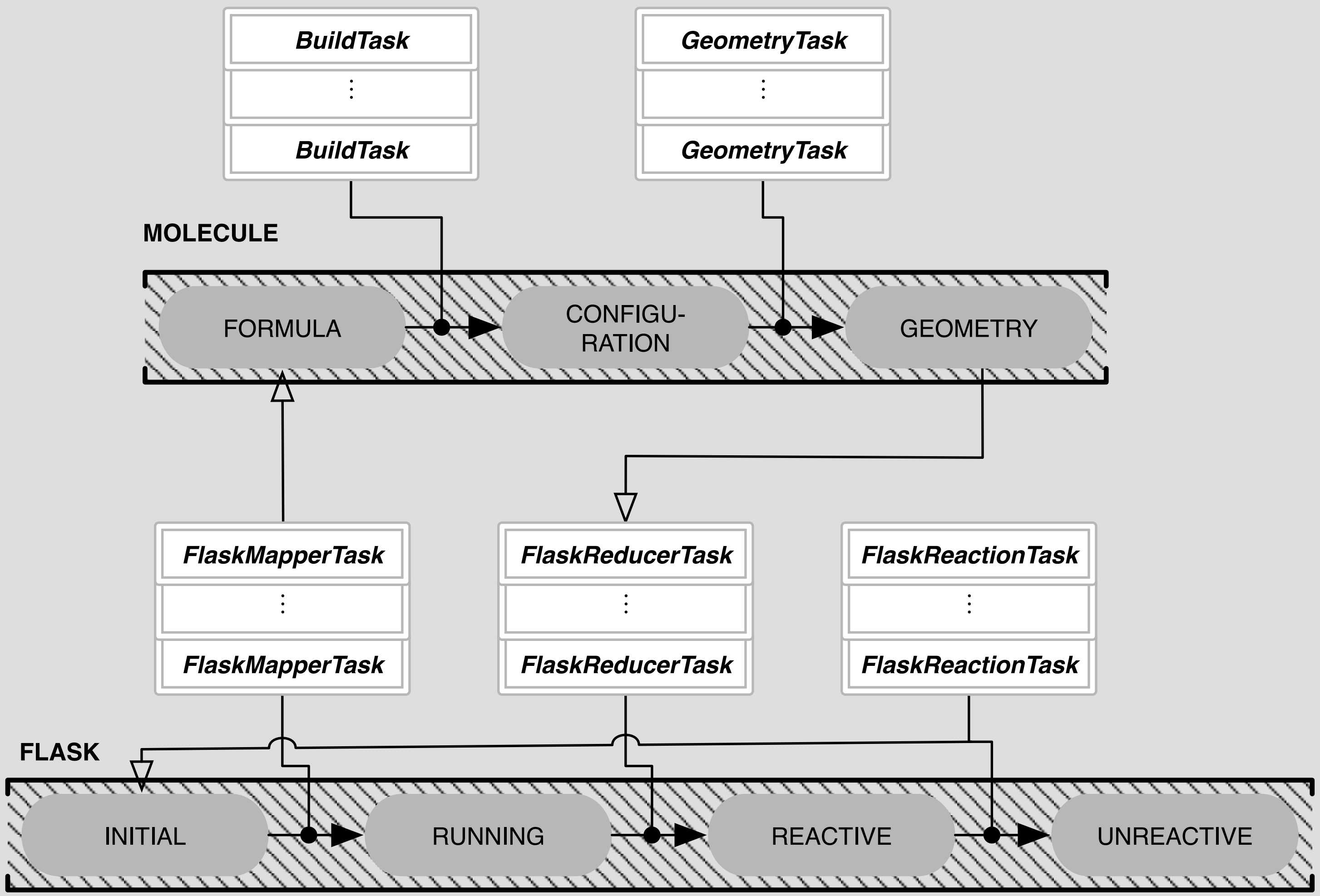


- Association / Dissociation Mechanism



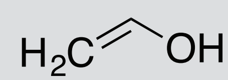
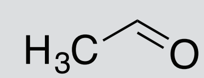
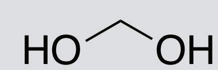
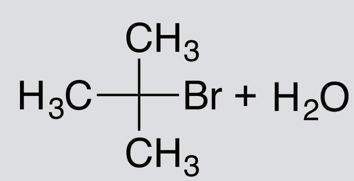
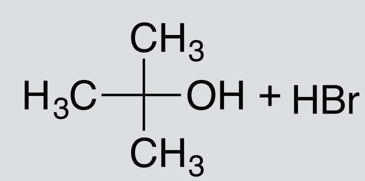
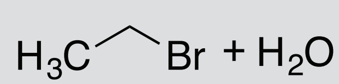
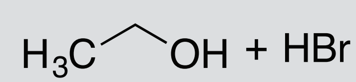
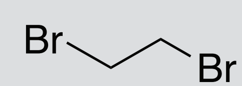
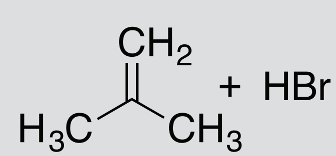
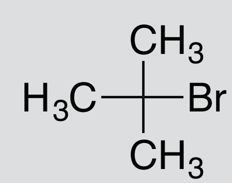
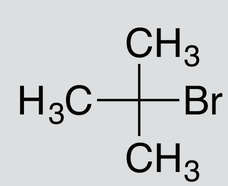
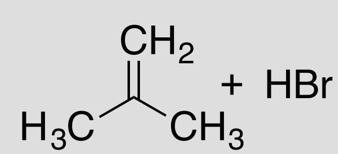
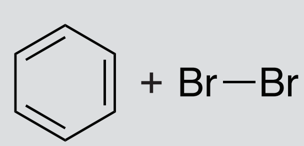
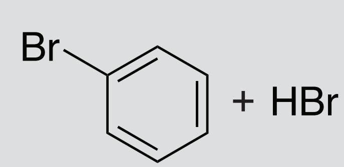
W. P. Jencks, *Acc. Chem. Res.* (1980) *13*, 161.

colibri

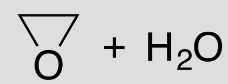
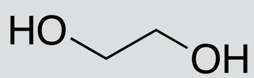
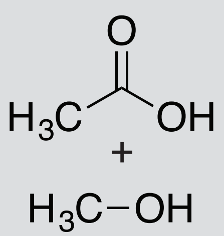
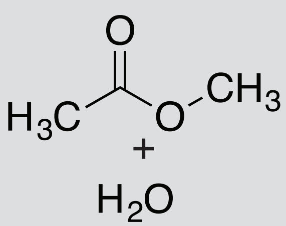
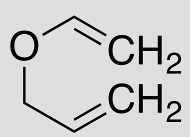
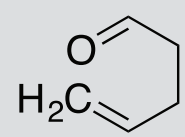
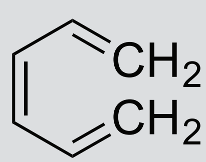
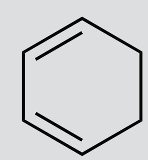
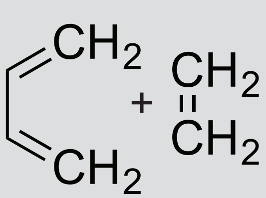
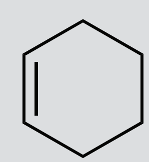
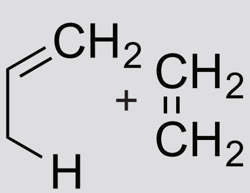
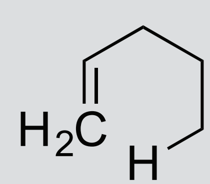


- Written in Python
- Database-centric; embarrassingly parallel
- MongoDB for data interchange; Neo4J for graph store & exploration
- To be released soon under BSD License

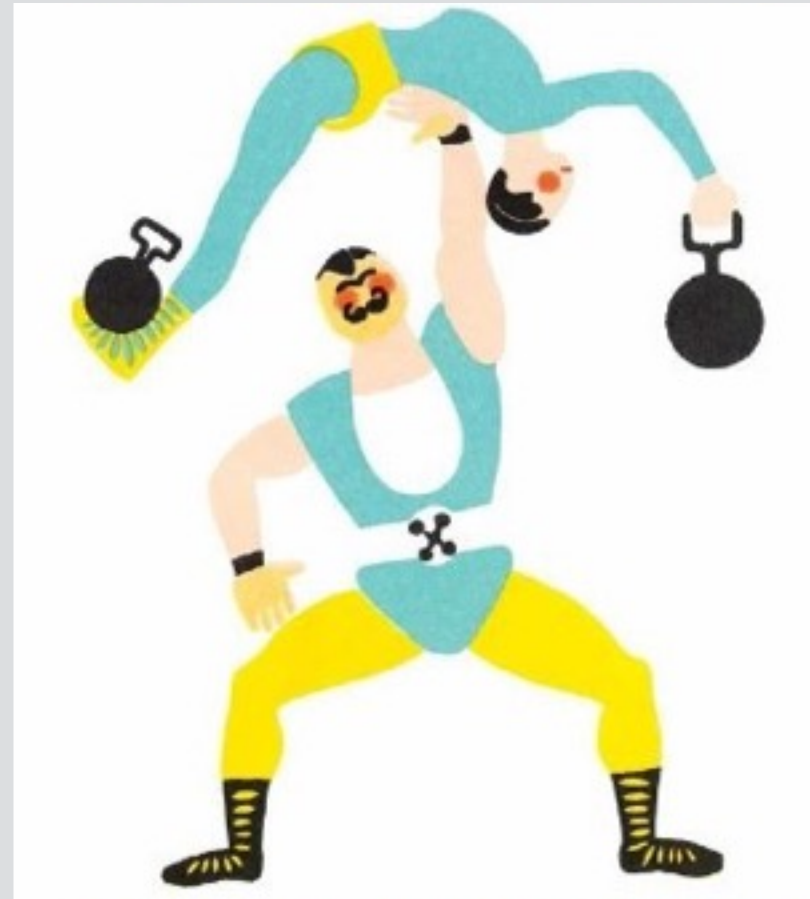
Network Sizes

	Reaction		Network	
	Reactants	Products	Nodes	Edges
Tautomerizaton (TA)			49	144
Diol formation (DI)	$H_2C{=}O + H_2O$		20	60
S_N1 substitution (S1)			7100	34563
S_N2 substitution (S2)			145	610
Br_2 addition (BA)	$H_2C{=}CH_2$ + Br—Br		53	168
HBr addition (HA)			738	3780
$E1$ elimination (E1)			738	3914
S_EAr substitution (SA)			114652	485814

Network Sizes (Cont'd)

	Reactants	Products	Nodes	Edges
Epoxide hydrolysis (EP)			249	1076
Esterification (ES)			6563	63536
Claisen rearrangement (CL)			15397	127408
6π electrocyclization (6C)			14779	87535
Diels–Alder reaction (DA)			11176	66895
Ene reaction (EN)			1034	5630

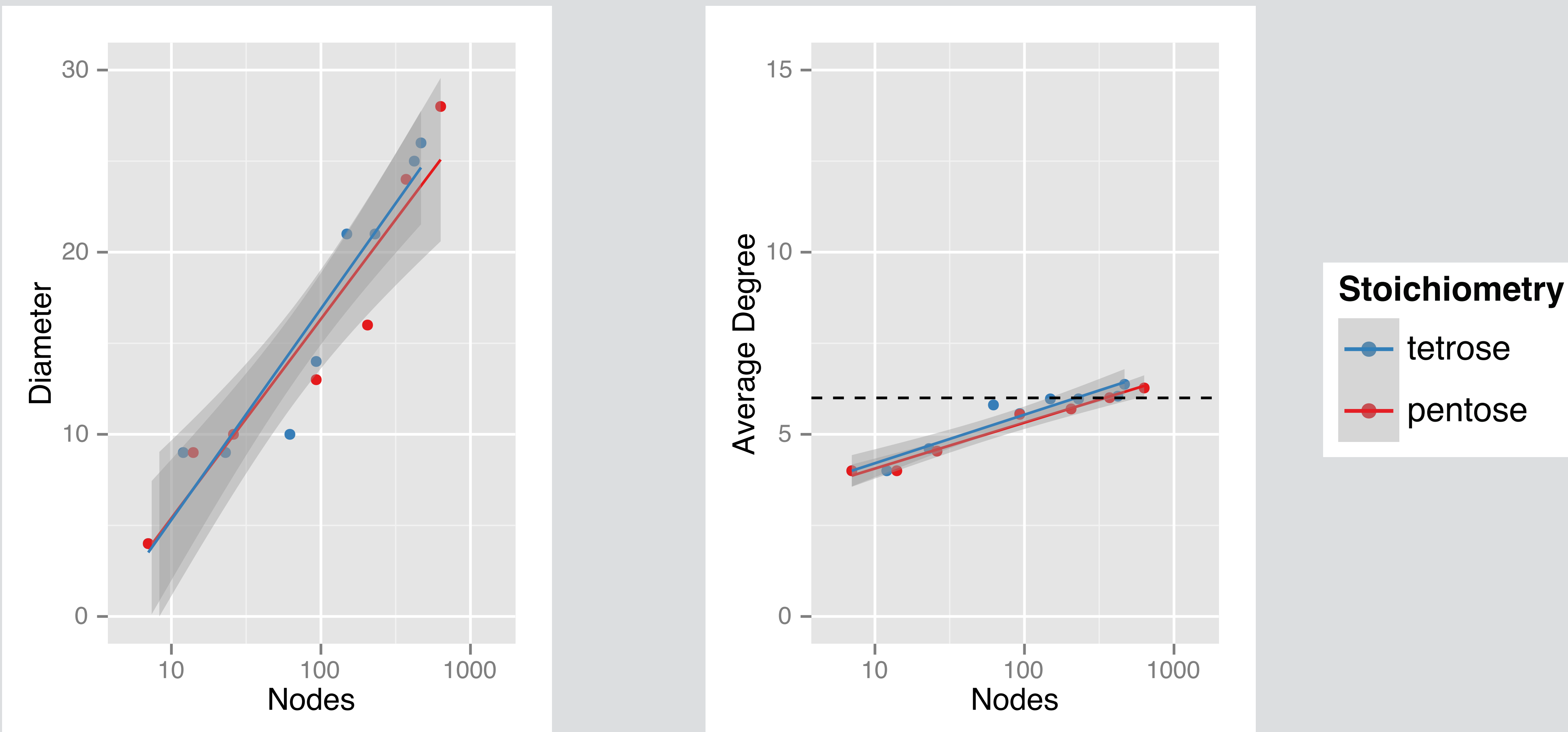
One More Thing



V. Lebedev, *Circus* (1925)

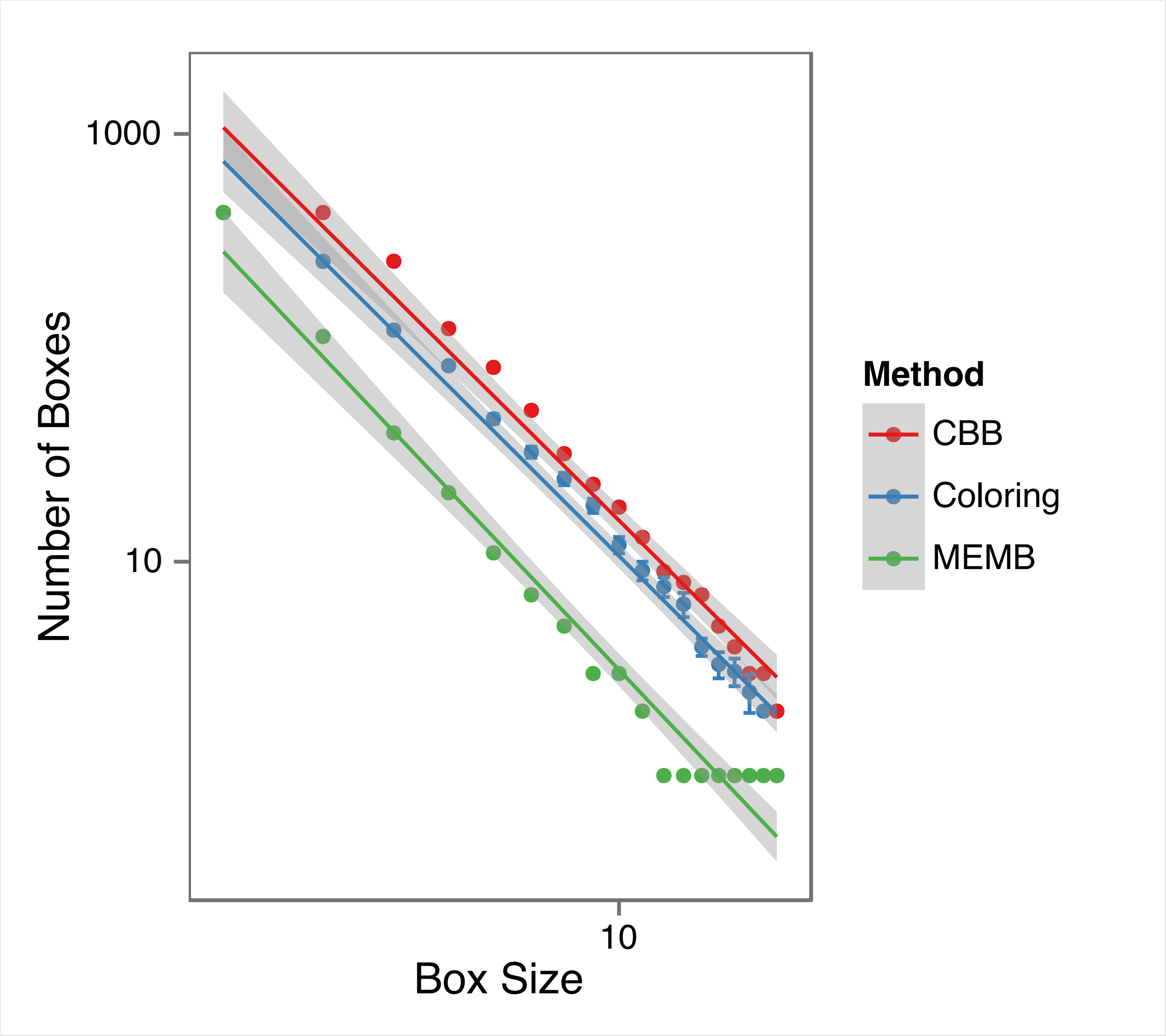
V. Lebedev, *Baggage* (1926)

Network Scaling Properties



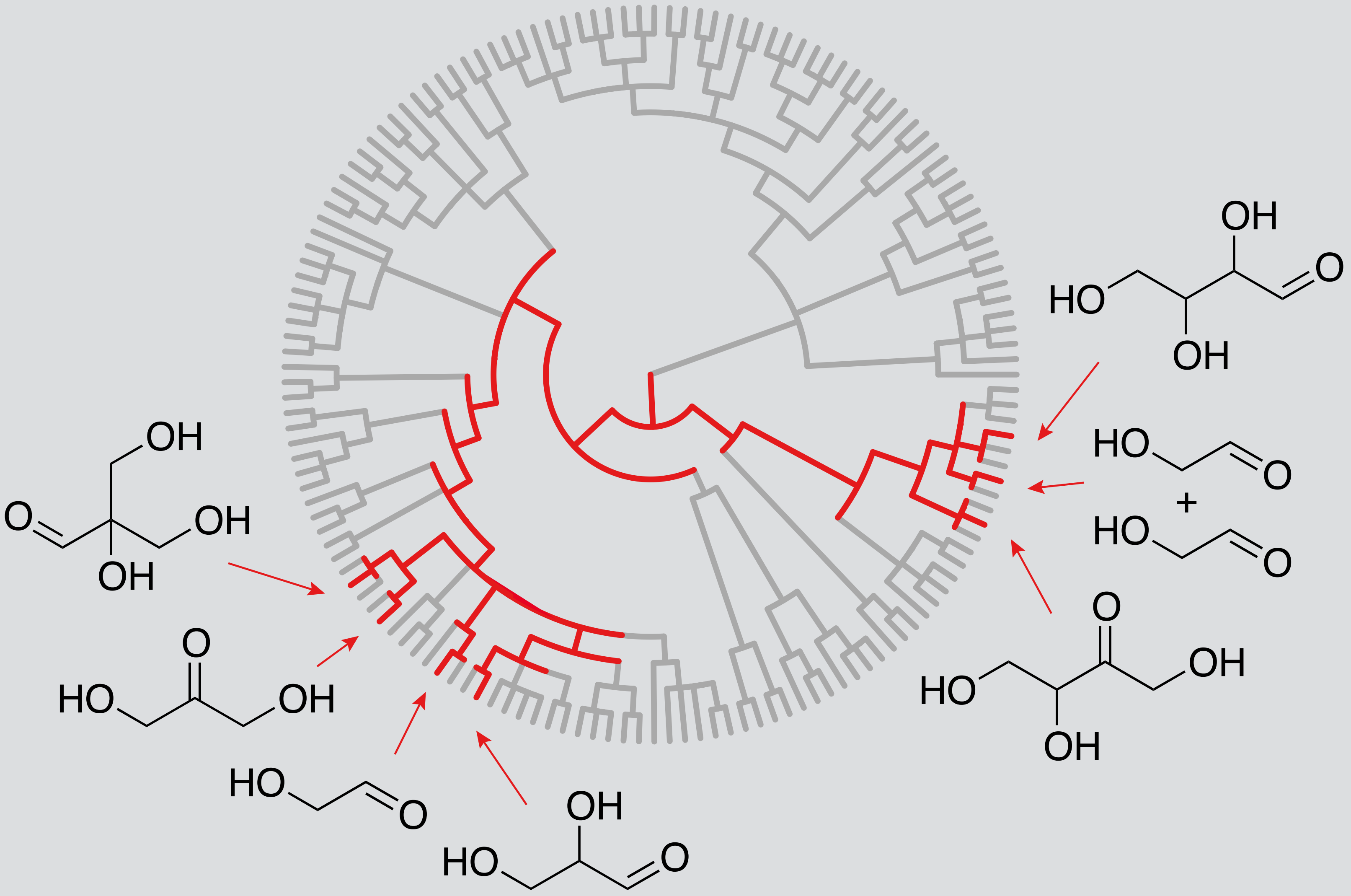
Network diameter grows logarithmically with number of nodes (small–world networks); favorable for local search

Fractal Dimension of Networks



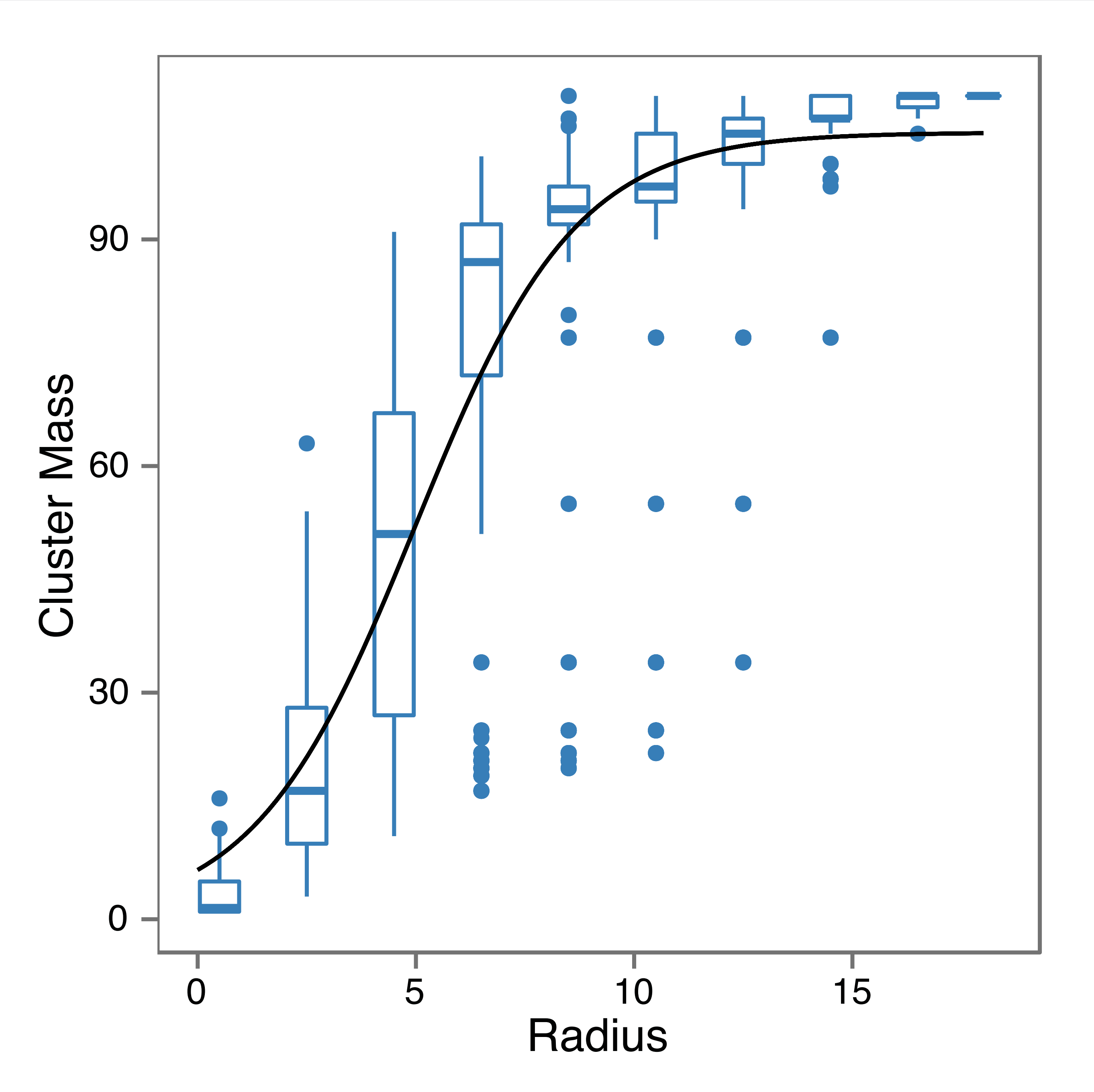
CH_2O + HCOOH
432 nodes
1612 edges
$d \approx 2.1$

Hierarchical Clustering in Formose Networks



Hierarchical network structure reflects reaction mechanism and helps identify mechanistic bottlenecks

Reaction Network Growth ($C_3H_6O_3$ / Triose)



32 Generations
6620 Nodes
31359 Edges
982 Molecules

Looking Ahead

- Larger Networks (> 10^9 Nodes)
- Better Quantum Chemistry
- Incorporation of Transition State Searches and Experimental Data
- Stochastic Search, Dynamics & Statistical Mechanics of Networks
- More Reactions (Radicals, Photochemistry)
- Learn on and from Networks

A. Mucha, *Nestlé’s Food for Infants* (1897)

Reaction Networks for Fun and Non-Profit

- Define Cost Function $C(\mathcal{P}) = C(\{E(\mathcal{F}_i)\})$ for Pathway

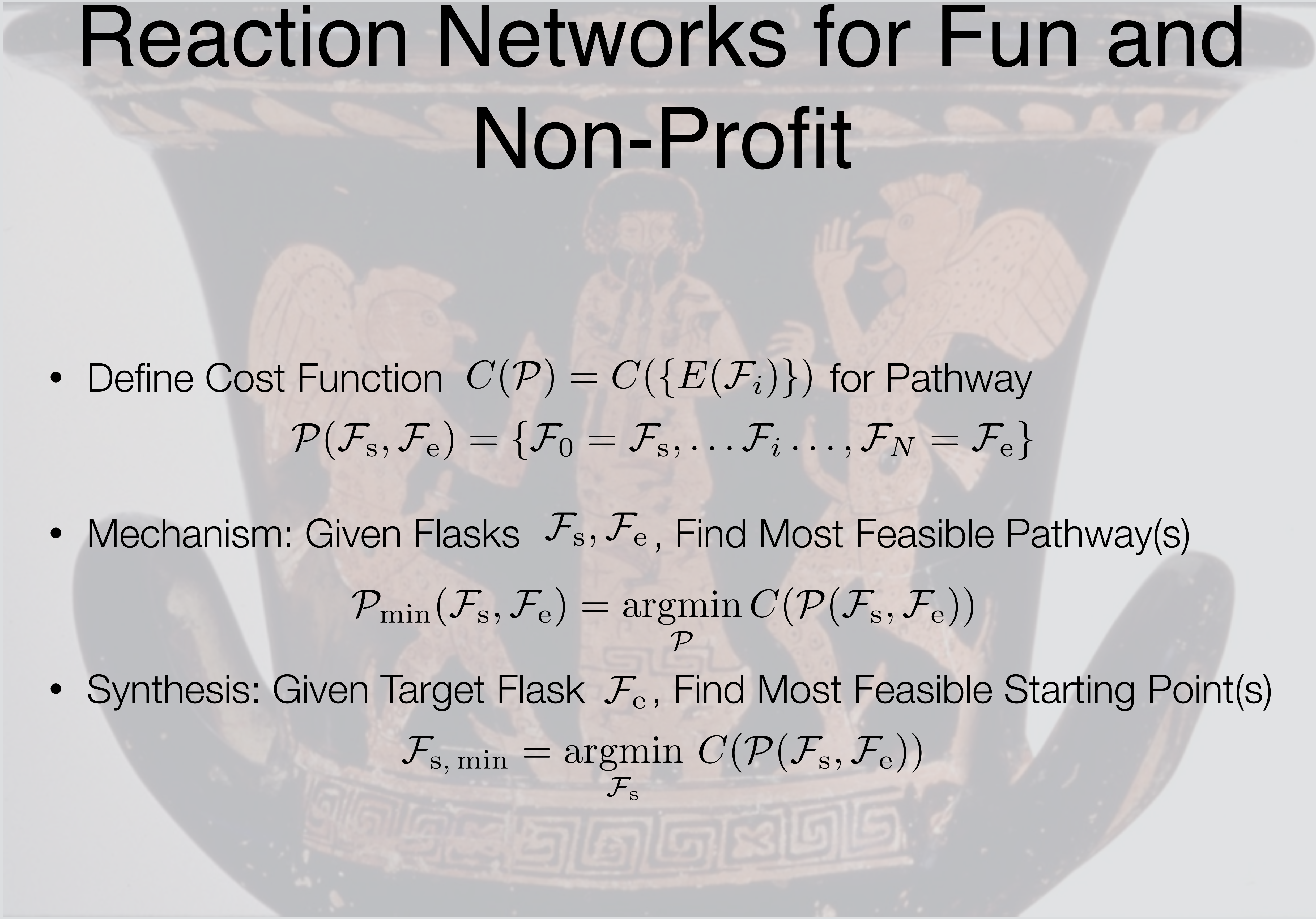
$$\mathcal{P}(\mathcal{F}_{\mathrm{s}}, \mathcal{F}_{\mathrm{e}}) = \{\mathcal{F}_0 = \mathcal{F}_{\mathrm{s}}, \ldots \mathcal{F}_i \ldots, \mathcal{F}_N = \mathcal{F}_{\mathrm{e}}\}$$

- Mechanism: Given Flasks $\mathcal{F}_{\mathrm{s}}, \mathcal{F}_{\mathrm{e}}$, Find Most Feasible Pathway(s)

$$\mathcal{P}_{\min}(\mathcal{F}_{\mathrm{s}}, \mathcal{F}_{\mathrm{e}}) = \underset{\mathcal{P}}{\operatorname{argmin}}\, C(\mathcal{P}(\mathcal{F}_{\mathrm{s}}, \mathcal{F}_{\mathrm{e}}))$$

- Synthesis: Given Target Flask $\mathcal{F}_{\mathrm{e}}$, Find Most Feasible Starting Point(s)

$$\mathcal{F}_{\mathrm{s,\,min}} = \underset{\mathcal{F}_{\mathrm{s}}}{\operatorname{argmin}}\; C(\mathcal{P}(\mathcal{F}_{\mathrm{s}}, \mathcal{F}_{\mathrm{e}}))$$



Attic Red Figure, *Aristophanes' Birds* (ca. 415–400 BCE)

Weaving Molecules into Networks

- Define Partial Ordering $N \leq N'$ Between Networks $N = \{C_x H_y N_z O_w \ldots\}$ and $N' = \{C_{x'} H_{y'} N_{z'} O_{w'} \ldots\}$ if $x \leq x'$, $y \leq y'$, $z \leq z'$, $w \leq w'$.

- Catalysis: Given Network N and $C_0 \geq 0$, Find Smallest $N' \geq N$ with $C(\mathcal{P}') \leq C_0$

- Analysis: Determine Network Parameters of $N = \{C_x H_y N_z O_w \ldots\}$ (Size, Degree Distribution, Clustering Coefficient, …) as Functions of x, y, z, w,…

Flemish Tapestry, *The Three Fates* (prob. 1510–1520)