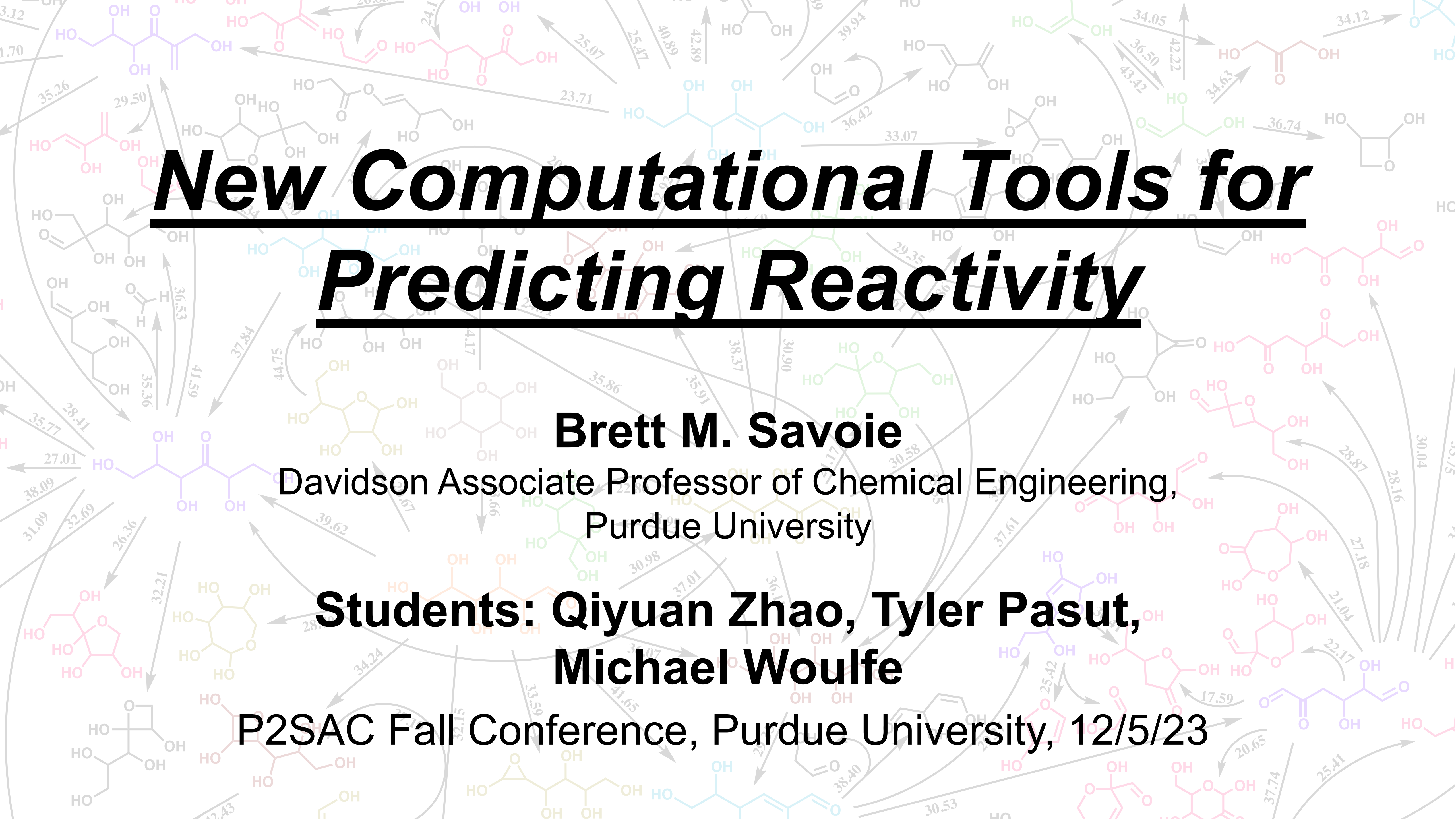


New Computational Tools for Predicting Reactivity

Brett M. Savoie

Davidson Associate Professor of Chemical Engineering, Purdue University

Students: Qiyuan Zhao, Tyler Pasut, Michael Woulfe

P2SAC Fall Conference, Purdue University, 12/5/23

"Known Unknowns" and "Unknown Unknowns"

A → B

- To safely plan a known reaction, we need access to solid thermodynamic data (e.g., ΔH_f, S^o, C_v) to understand and classify risks.

- This is a "**known unknown**" in that we know the reaction, A → B, but we need values for a few unknown variables.

A → ? → B ; A → B + ? ; A → ?

- **A → ? → B**, means that we know the net reaction, but there may be a consequential (e.g., potentially reactive) intermediate. Even if we have accurate thermodynamic data on A/B, neglecting the intermediate could be disastrous.

- The **A → B + ?** (unknown side-reaction) and **A → ?** (unknown main product), problems have similar "**unknown unknown**" characteristics.

“Known Unknowns” and “Unknown Unknowns”

A → B

- To safely plan a known reaction, we need access to solid thermodynamic data (e.g., ΔH_f, S°, C_v) to understand and classify risks.
- This is a “known unknown” in that we know the reaction, A → B, but we need values for a few unknown variables.

TAFFI Component Increment Theory (TCIT)

A → ? → B ; A → B + ? ; A → ?

- **A → ? → B**, means that we know the net reaction, but there may be a consequential (e.g., potentially reactive) intermediate. Even if we have accurate thermodynamic data on A/B, neglecting the intermediate could be disastrous.
- The **A → B + ?** (unknown side-reaction) and **A → ?** (unknown main product), problems have similar “unknown unknown” characteristics.

Yet Another Reaction Program (YARP)

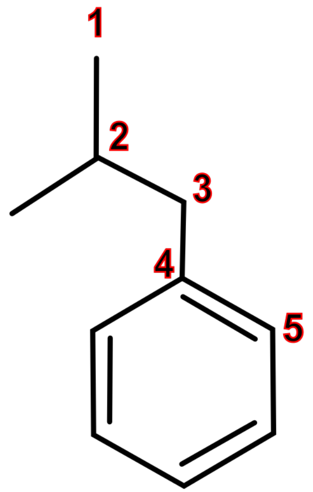
Challenges of Contemporary Group Theories

Benson Group Theory:

- The idea is to decompose molecular properties (ΔH_f, S°, C_v) as the sum of "group" contributions.

- Group contributions are calculated based on trusted experimental or computational data, and transferability is assumed.

From Anslyn and Dougherty's Textbook



1) C -(C)$(H)_3$	2(-10.20)
2) C -$(C)_3$(H)	-1.90
3) C -(C_B)(C)$(H)_2$	-4.86
4) C_B -(C)	5.51
5) C_B -(H)	5(3.30)
	-5.15 kcal/mole (-21.6 kJ/mole)

Experimental ΔH_f: -5.15 +/- 0.34 kcal/mol

Problems we want to address:

- **Specificity:** the definition of a "group" has never been formalized and inconsistent granularity is applied.

- **Provenance:** inconsistent thermodynamic data is available/used to determine group contributions.

- **Extensibility:** because of the provenance and specificity problems, it isn't possible to develop new groups in a consistent way.

Challenges of Contemporary Group Theories

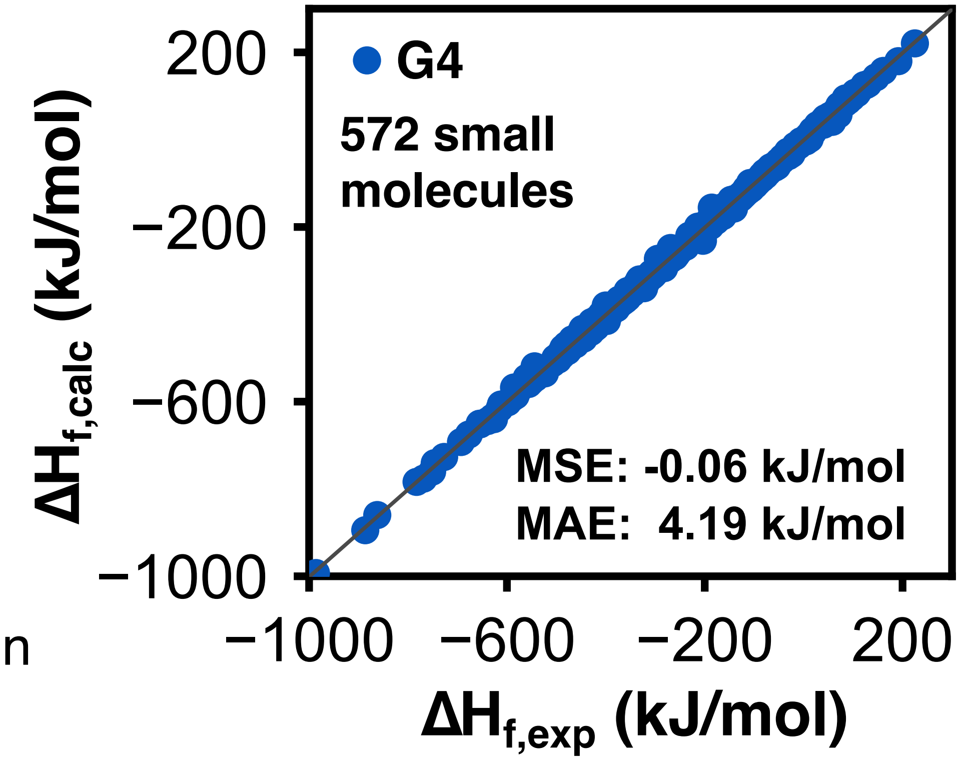
Benson Group Theory:

- The idea is to decompose molecular properties (ΔH_f, S°, C_v) as the sum of “group” contributions.
- Group contributions are calculated based on trusted experimental or computational data, and transferability is assumed.

Problems we want to address:

- **Specificity:** the definition of a “group” has never been formalized and inconsistent granularity is applied.
- **Provenance:** inconsistent thermodynamic data is available/used to determine group contributions.
- **Extensibility:** because of the provenance and specificity problems, it isn’t possible to develop new groups in a consistent way.

ΔH_f from modern quantum chemistry



Zhao, Q.; Savoie, B. M.; Enthalpy of Formation Prediction via a fully Self-Consistent Component Increment Theory. *J. Chem. Info. Model.* **2020**, 60, 2199-2207

Challenges of Contemporary Group Theories

Benson Group Theory:

- The idea is to decompose molecular properties (ΔH_f, S°, C_v) as the sum of “group” contributions.

- Grou
on trus
data, a

Can we circumvent the provenance and extensibility challenges using the <u>throughput</u> and <u>accuracy</u> of modern quantum chemistry?

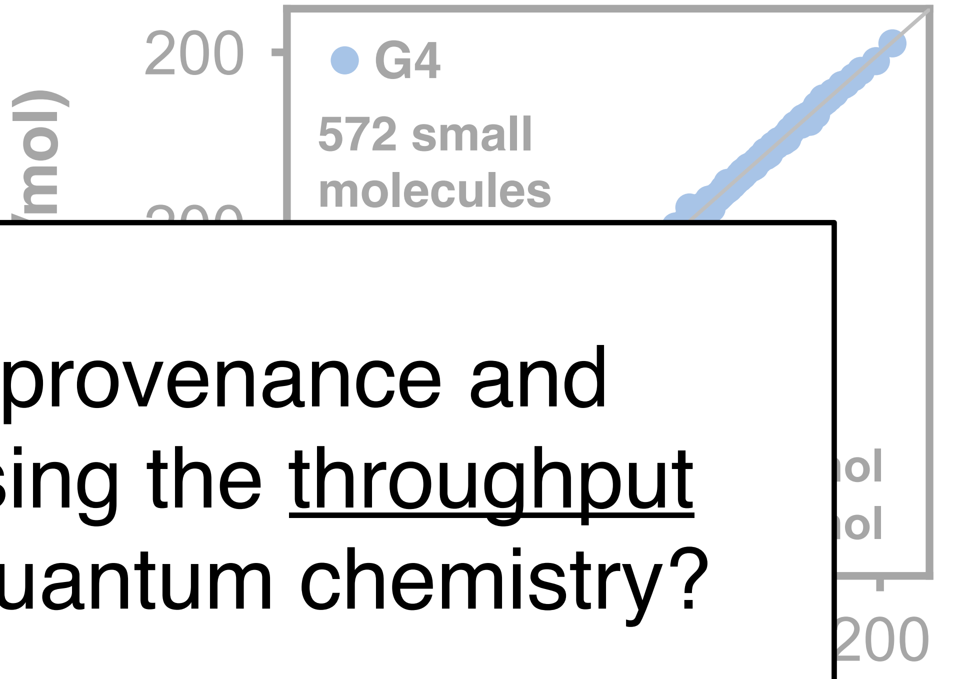
Prob

- **Spec**
formal

- **Provenance:** inconsistent thermodynamic data is available/used to determine group contributions.

- **Extensibility:** because of the provenance and specificity problems, it isn't possible to develop new groups in a consistent way.

ΔH_f from modern quantum chemistry



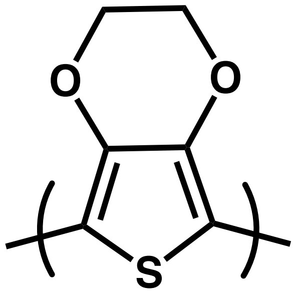
Zhao, Q.; Savoie, B. M.; Enthalpy of Formation Prediction via a fully Self-Consistent Component Increment Theory. *J. Chem. Info. Model.* **2020**, 60, 2199-2207

TAFFI Component Increment Theory (TCIT)

The fundamental idea

- Systematize component-definitions and model compound selection with rigorous graph-based typing.

TCIT is a component theory (2-bond specific)



Topology Automated Force Field Interactions

graph/structure equivalence

$$
\begin{array}{c}
S \\ C \\ C \\ C \\ C \\ O \\ C \\ C \\ H \\ H \\ O \\ H \\ H
\end{array}
\left[\begin{array}{ccccccccccccc}
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0
\end{array}\right]
$$

Adjacency matrix for PEDOT monomer

Zhao, Q.; Savoie, B. M.; “Enthalpy of Formation Prediction via a fully Self-Consistent Component Increment Theory”. *J. Chem. Info. Model.* **2020**, 60, 2199-2207

Zhao, Q.; Iovanac, N.; Savoie, B. M.; “Transferable Ring Corrections for Predicting Enthalpy of Formation of Cyclic Compounds” *J. Chem. Info. Model.* **2021**, 61, 5013-5027

Seo, B.; Lin, Z.-Y.; Zhao, Q.; Webb, M. A.; Savoie, B. M. Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force Fields. *J. Chem. Inf. Model.* **2021**, 61 (10), 5013–5027. https://doi.org/10.1021/acs.jcim.1c00491.

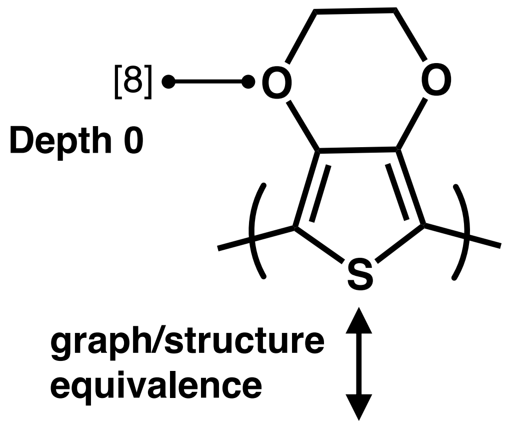
P2SAC Publications

TAFFI Component Increment Theory (TCIT)

The fundamental idea

- Systematize component-definitions and model compound selection with rigorous graph-based typing.

TCIT is a component theory (2-bond specific)



[8]

Depth 0

graph/structure equivalence

Topology Automated Force Field Interactions

TAFFI

$$\begin{array}{c|ccccccccccccc} \mathrm{S} & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathrm{C} & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathrm{C} & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathrm{C} & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \mathrm{C} & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathrm{O} & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathrm{C} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ \mathrm{C} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ \mathrm{H} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathrm{H} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathrm{O} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \mathrm{H} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \mathrm{H} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{array}$$

Adjacency matrix for PEDOT monomer

Zhao, Q.; Savoie, B. M.; “Enthalpy of Formation Prediction via a fully Self-Consistent Component Increment Theory”. *J. Chem. Info. Model.* **2020**, 60, 2199-2207

Zhao, Q.; Iovanac, N.; Savoie, B. M.; “Transferable Ring Corrections for Predicting Enthalpy of Formation of Cyclic Compounds” *J. Chem. Info. Model.* **2021**, 61, 5013-5027

Seo, B.; Lin, Z.-Y.; Zhao, Q.; Webb, M. A.; Savoie, B. M. Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force Fields. *J. Chem. Inf. Model.* **2021**, 61 (10), 5013–5027. https://doi.org/10.1021/acs.jcim.1c00491.

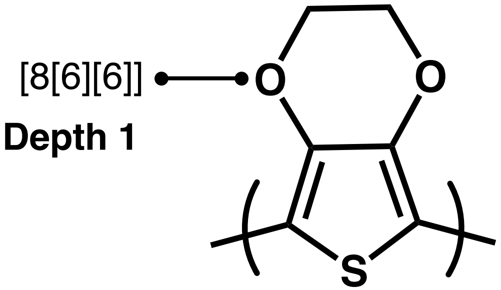
P2SAC Publications

TAFFI Component Increment Theory (TCIT)

The fundamental idea

- Systematize component-definitions and model compound selection with rigorous graph-based typing.

TCIT is a <ins>component</ins> theory (2-bond specific)



Topology **A**utomated **F**orce **F**ield **I**nteractions

graph/structure equivalence

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

Adjacency matrix for PEDOT monomer

Zhao, Q.; Savoie, B. M.; “Enthalpy of Formation Prediction via a fully Self-Consistent Component Increment Theory”. *J. Chem. Info. Model.* **2020**, 60, 2199-2207

Zhao, Q.; Iovanac, N.; Savoie, B. M.; “Transferable Ring Corrections for Predicting Enthalpy of Formation of Cyclic Compounds” *J. Chem. Info. Model.* **2021**, 61, 5013-5027

Seo, B.; Lin, Z.-Y.; Zhao, Q.; Webb, M. A.; Savoie, B. M. Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force Fields. *J. Chem. Inf. Model.* **2021**, 61 (10), 5013–5027. https://doi.org/10.1021/acs.jcim.1c00491.

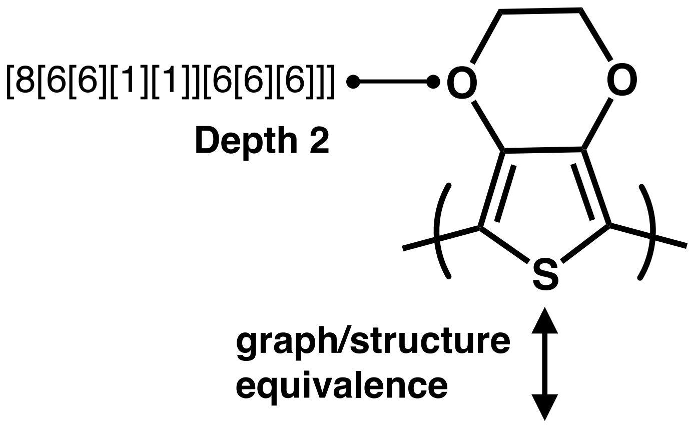
P2SAC Publications

TAFFI Component Increment Theory (TCIT)

The fundamental idea

- Systematize component-definitions and model compound selection with rigorous graph-based typing.

TCIT is a <u>component</u> theory (2-bond specific)



Topology **A**utomated **F**orce **F**ield **I**nteractions

$$\begin{array}{c|ccccccccccccc}
S & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
C & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
O & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
C & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\
H & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
H & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
O & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
H & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
H & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0
\end{array}$$

Adjacency matrix for PEDOT monomer

Zhao, Q.; Savoie, B. M.; "Enthalpy of Formation Prediction via a fully Self-Consistent Component Increment Theory". *J. Chem. Info. Model.* **2020**, 60, 2199-2207

Zhao, Q.; Iovanac, N.; Savoie, B. M.; "Transferable Ring Corrections for Predicting Enthalpy of Formation of Cyclic Compounds" *J. Chem. Info. Model.* **2021**, 61, 5013-5027

Seo, B.; Lin, Z.-Y.; Zhao, Q.; Webb, M. A.; Savoie, B. M. Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force Fields. *J. Chem. Inf. Model.* **2021**, 61 (10), 5013–5027. https://doi.org/10.1021/acs.jcim.1c00491.

P2SAC Publications

TAFFI Component Increment Theory (TCIT)

The fundamental idea

- Systematize component-definitions and model compound selection with rigorous graph-based typing.

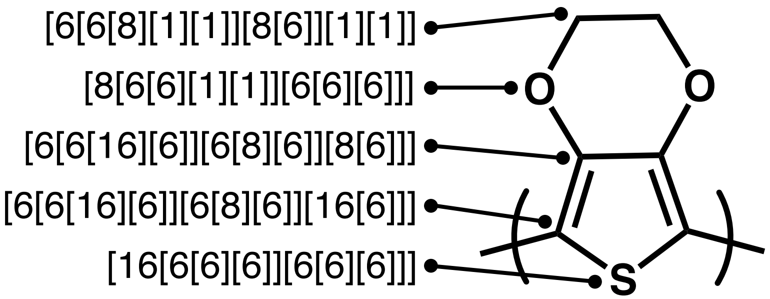
Zhao, Q.; Savoie, B. M.; “Enthalpy of Formation Prediction via a fully Self-Consistent Component Increment Theory”. *J. Chem. Info. Model.* **2020**, 60, 2199-2207

Zhao, Q.; Iovanac, N.; Savoie, B. M.; “Transferable Ring Corrections for Predicting Enthalpy of Formation of Cyclic Compounds” *J. Chem. Info. Model.* **2021**, 61, 5013-5027

Seo, B.; Lin, Z.-Y.; Zhao, Q.; Webb, M. A.; Savoie, B. M. Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force Fields. *J. Chem. Inf. Model.* **2021**, 61 (10), 5013–5027. https://doi.org/10.1021/acs.jcim.1c00491.

P2SAC Publications

TCIT is a component theory (2-bond specific)



Topology **A**utomated **F**orce **F**ield **I**nteractions

graph/structure equivalence

$$
\begin{array}{c|ccccccccccccc}
S & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
C & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
O & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
C & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\
H & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
H & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
O & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
H & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
H & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0
\end{array}
$$

Adjacency matrix for PEDOT monomer

TAFFI Component Increment Theory (TCIT)

The fundamental idea

- Systematize component-definitions and model compound selection with rigorous graph-based typing.
- Two-bond specificity should improve both the accuracy and transferability of the resulting components.
- Parameterizing a component model **would not be feasible with only experimental data**.

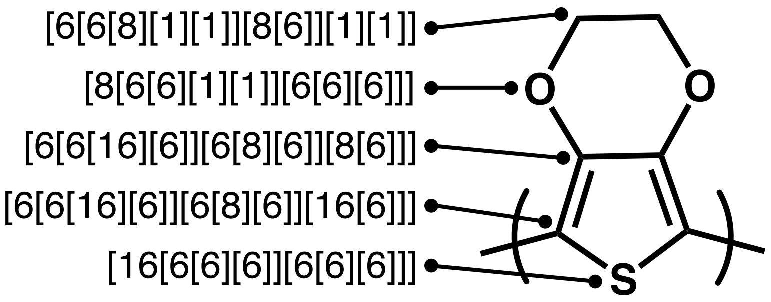
Zhao, Q.; Savoie, B. M.; "Enthalpy of Formation Prediction via a fully Self-Consistent Component Increment Theory". *J. Chem. Info. Model.* **2020**, 60, 2199-2207

Zhao, Q.; Iovanac, N.; Savoie, B. M.; "Transferable Ring Corrections for Predicting Enthalpy of Formation of Cyclic Compounds" *J. Chem. Info. Model.* **2021**, 61, 5013-5027

Seo, B.; Lin, Z.-Y.; Zhao, Q.; Webb, M. A.; Savoie, B. M. Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force Fields. *J. Chem. Inf. Model.* **2021**, 61 (10), 5013–5027. https://doi.org/10.1021/acs.jcim.1c00491.

P2SAC Publications

TCIT is a <u>component</u> theory (2-bond specific)



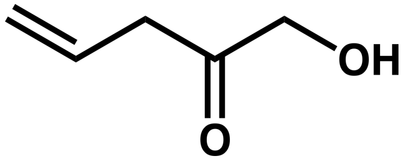
Topology **A**utomated **F**orce **F**ield **I**nteractions

graph/structure equivalence

$$\begin{array}{c|ccccccccccccc}
S & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
C & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
O & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
C & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\
H & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
H & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
O & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
H & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
H & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0
\end{array}$$

Adjacency matrix for PEDOT monomer

Graphical Decomposition of Model Compounds



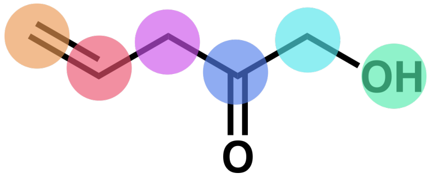
1-hydroxy-pent-2-ene-2-one

How will we select molecules for parameterizing TCIT components?

Zhao, Q.; Savoie, B. M. *J. Chem. Info. Model.* 2020, 60, 2199-2207.

Graphical Decomposition of Model Compounds

identify components*



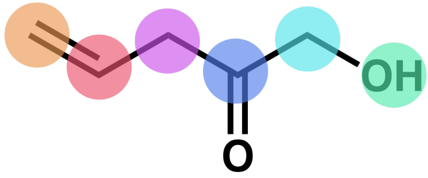
1-hydroxy-pent-2-ene-2-one

How will we select molecules for parameterizing TCIT components?

Zhao, Q.; Savoie, B. M. *J. Chem. Info. Model.* 2020, 60, 2199-2207.

Graphical Decomposition of Model Compounds

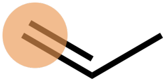
identify components*



1-hydroxy-pent-2-ene-2-one

How will we select molecules for parameterizing TCIT components?

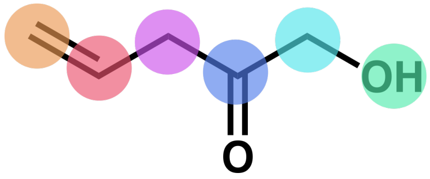
Recursively generate smallest acyclic model compounds



Zhao, Q.; Savoie, B. M. *J. Chem. Info. Model.* 2020, 60, 2199-2207.

Graphical Decomposition of Model Compounds

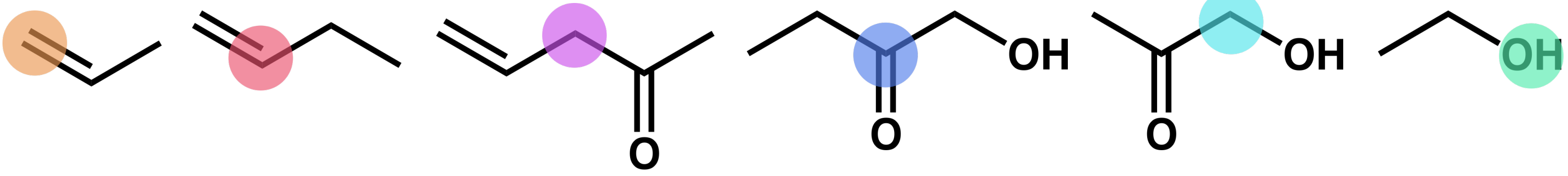
identify components*



1-hydroxy-pent-2-ene-2-one

How will we select molecules for parameterizing TCIT components?

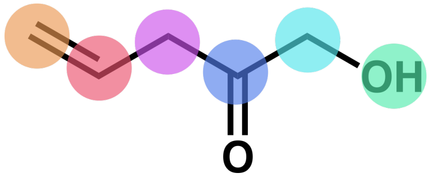
Recursively generate smallest acyclic model compounds



Zhao, Q.; Savoie, B. M. *J. Chem. Info. Model.* 2020, 60, 2199-2207.

Graphical Decomposition of Model Compounds

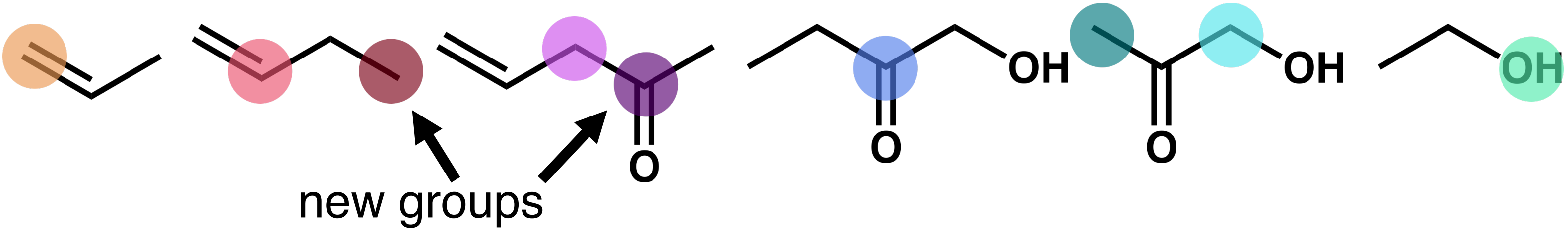
identify components*



1-hydroxy-pent-2-ene-2-one

How will we select molecules for parameterizing TCIT components?

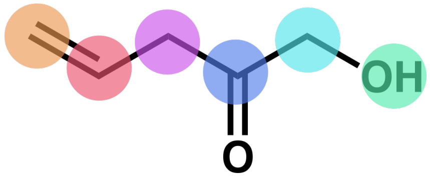
Recursively generate smallest acyclic model compounds



Zhao, Q.; Savoie, B. M. *J. Chem. Info. Model.* 2020, 60, 2199-2207.

Graphical Decomposition of Model Compounds

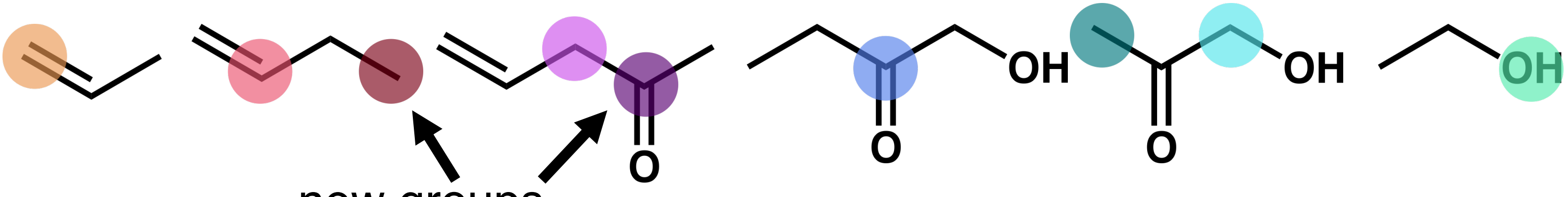
identify components*



1-hydroxy-pent-2-ene-2-one

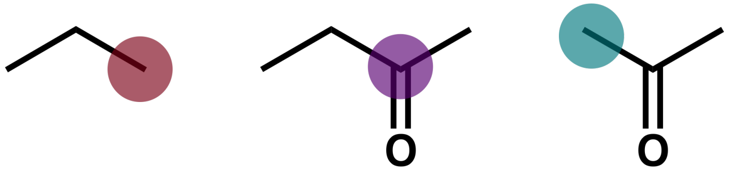
How will we select molecules for parameterizing TCIT components?

Recursively generate smallest acyclic model compounds



new groups

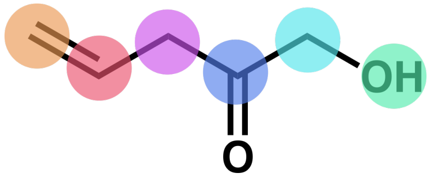
Resolve dependencies



Zhao, Q.; Savoie, B. M. *J. Chem. Info. Model.* 2020, 60, 2199-2207.

Graphical Decomposition of Model Compounds

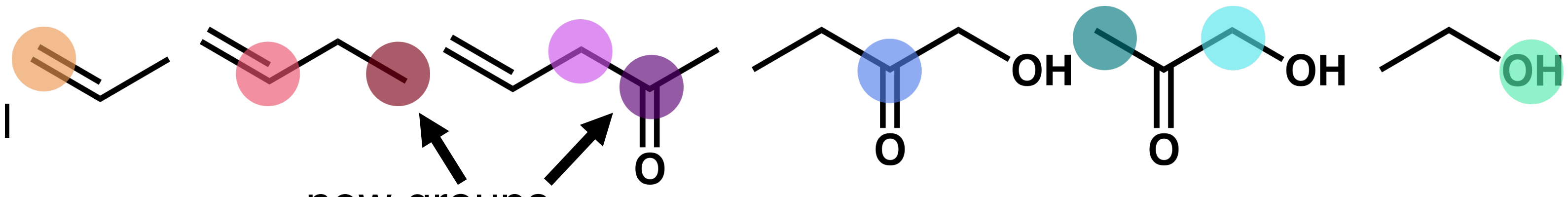
identify components*



1-hydroxy-pent-2-ene-2-one

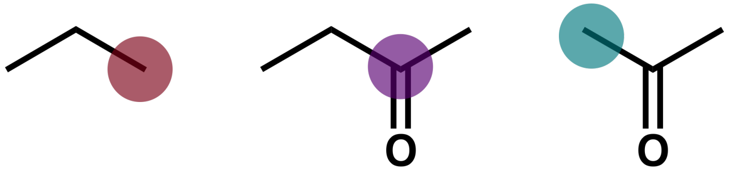
How will we select molecules for parameterizing TCIT components?

Recursively generate smallest acyclic model compounds

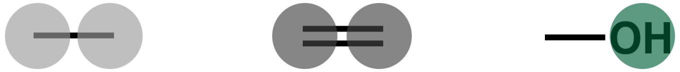


new groups

Resolve dependencies

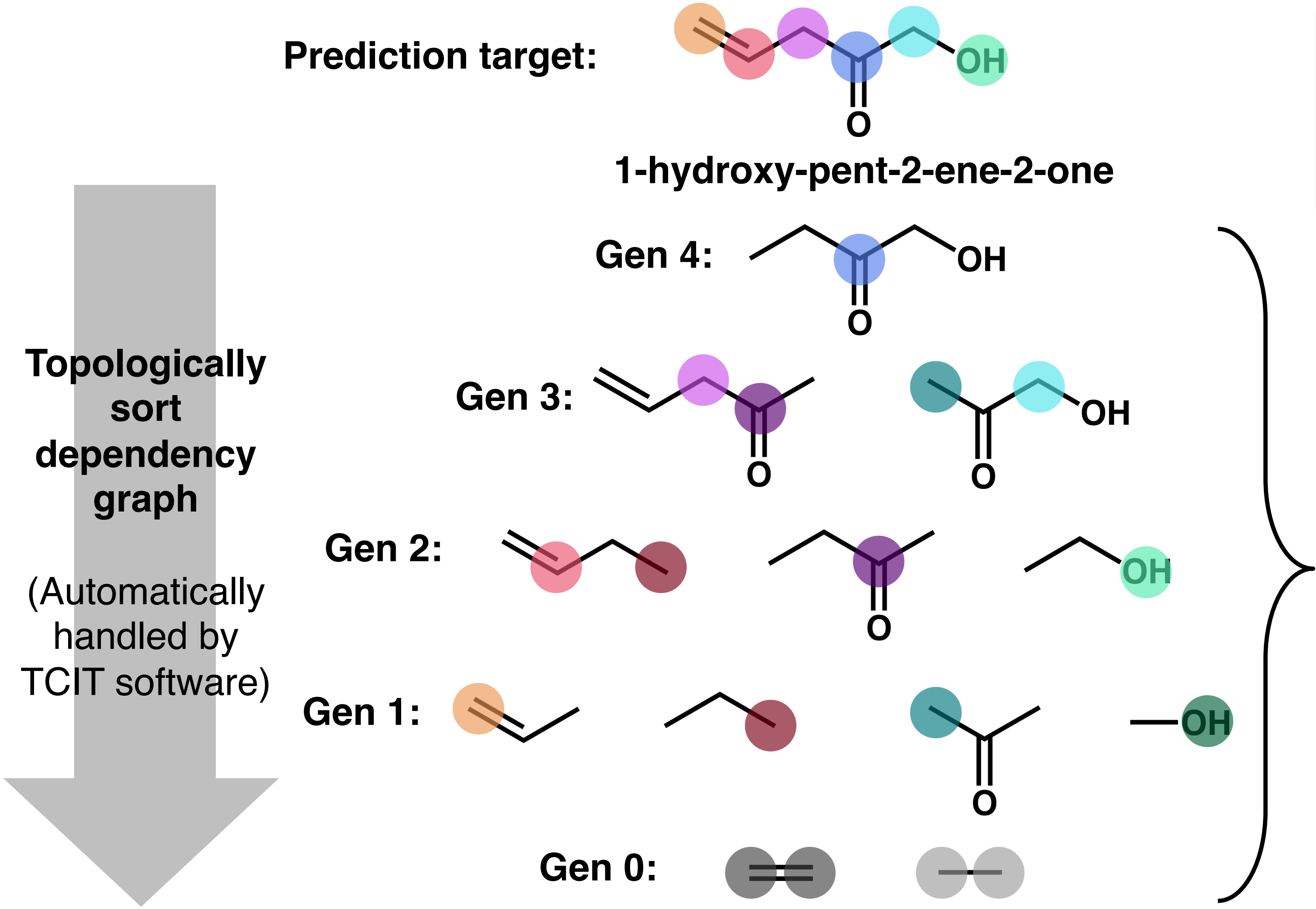


Resolve rank deficiency with elementary constraints



Zhao, Q.; Savoie, B. M. *J. Chem. Info. Model.* 2020, 60, 2199-2207.

Graphical Decomposition of Model Compounds



1-hydroxy-pent-2-ene-2-one

Topologically sort dependency graph

(Automatically handled by TCIT software)

$\Delta H_{f,G4}$ = -259.9 kJ/mol
$\Delta H_{f,TCIT}$ = -259.3 kJ/mol

no experimental data

Model compounds are small enough to perform the highest quality quantum chemistry calculations (G4 throughout)

Zhao, Q.; Savoie, B. M. *J. Chem. Info. Model.* 2020, 60, 2199-2207.

Graphical Decomposition of Model Compounds

Have we solved the specificity problem?

All components are unique out to a graph depth of two, no exceptions.

Have we solved the provenance problem?

All ΔH_f data is calculated at the G4 composite level, no exceptions.

Have we solved the extensibility problem?

Model compounds exist for all conceivable components, no exceptions.

Zhao, Q.; Savoie, B. M. *J. Chem. Info. Model.* 2020, 60, 2199-2207.

Benchmarking $\Delta H_{f,gas}$ Predictions Against the PNK Dataset

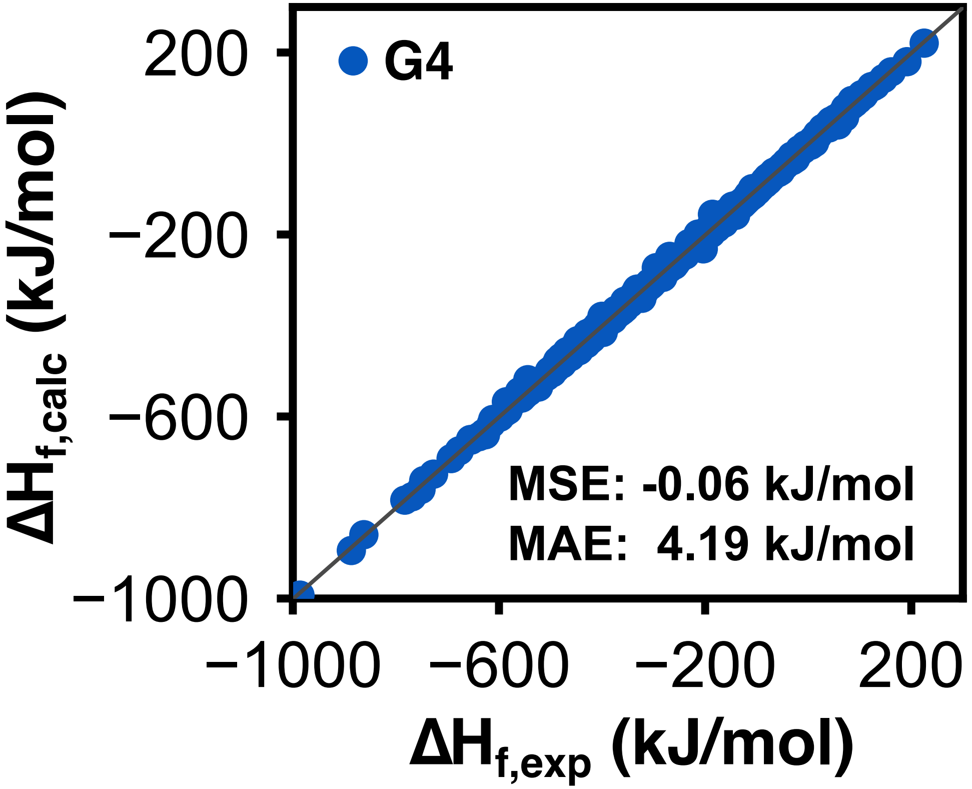
- Initial benchmarking set consists of ~1100 **linear** C,H, and O containing compounds from PNK[1]

[1] J. B. Pedley, R. D. Naylor, S. P. Kirby “Thermochemical Data of Organic Compounds” 2nd ed. 1986

- PNK is a core dataset for fitting Benson groups

- ~600 PNK compounds are small enough for G4 calculations and comparison with experiment.

572 small compounds from PNK



Zhao, Q.; Savoie, B. M.; Enthalpy of Formation Prediction via a Fully Self-Consistent Component Increment Theory. *J. Chem. Info. Model.* **2020**, 60, 2199-2207

Benchmarking $\Delta H_{f,gas}$ Predictions Against the PNK Dataset

- Initial benchmarking set consists of ~1100 **linear** C,H, and O containing compounds from PNK[1]

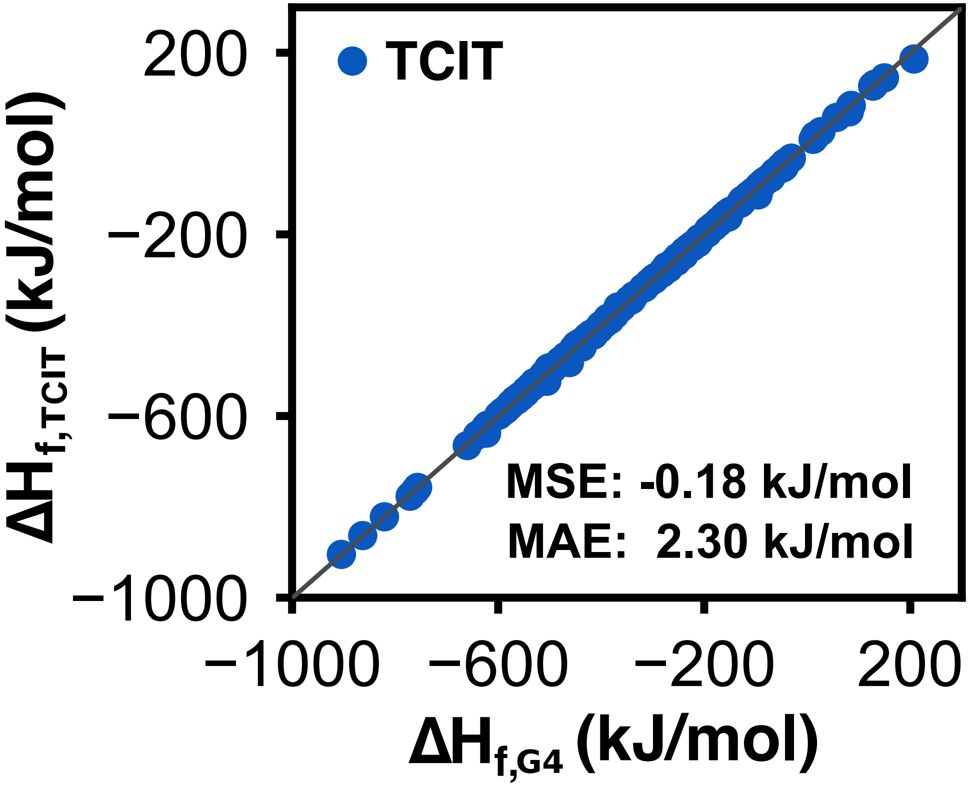
[1] J. B. Pedley, R. D. Naylor, S. P. Kirby “Thermochemical Data of Organic Compounds” 2nd ed. 1986

- PNK is a core dataset for fitting Benson groups

- ~600 PNK compounds are small enough for G4 calculations and comparison with experiment.

- ~150 PNK compounds are large enough for direct G4 calculation and comparison with TCIT.

150 medium compounds from PNK



Zhao, Q.; Savoie, B. M.; Enthalpy of Formation Prediction via a Fully Self-Consistent Component Increment Theory. *J. Chem. Info. Model.* **2020**, 60, 2199-2207

Benchmarking $\Delta H_{f,gas}$ Predictions Against the PNK Dataset

- Initial benchmarking set consists of ~1100 **linear** C,H, and O containing compounds from PNK[1]

(1) J. B. Pedley, R. D. Naylor, S. P. Kirby "Thermochemical Data of Organic Compounds" 2nd ed. 1986

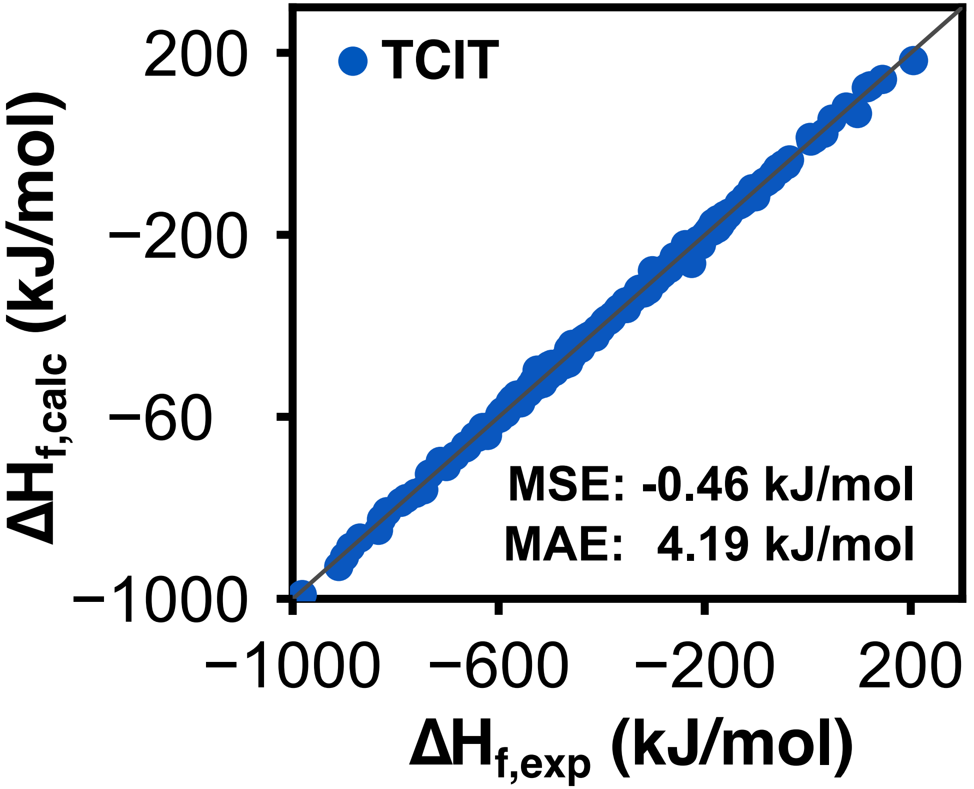
- PNK is a core dataset for fitting Benson groups

- ~600 PNK compounds are small enough for G4 calculations and comparison with experiment.

- ~150 PNK compounds are large enough for direct G4 calculation and comparison with TCIT.

- ~500 PNK compounds are large enough to evaluate the predictive accuracy of the increment theories.

~500 large compounds from PNK



Zhao, Q.; Savoie, B. M.; Enthalpy of Formation Prediction via a Fully Self-Consistent Component Increment Theory. *J. Chem. Info. Model.* **2020**, 60, 2199-2207

Benchmarking $\Delta H_{f,gas}$ Predictions Against the PNK Dataset

- Initial benchmarking set consists of ~1100 **linear** C,H, and O containing compounds from PNK[1]

[1] J. B. Pedley, R. D. Naylor, S. P. Kirby "Thermochemical Data of Organic Compounds" 2nd ed. 1986

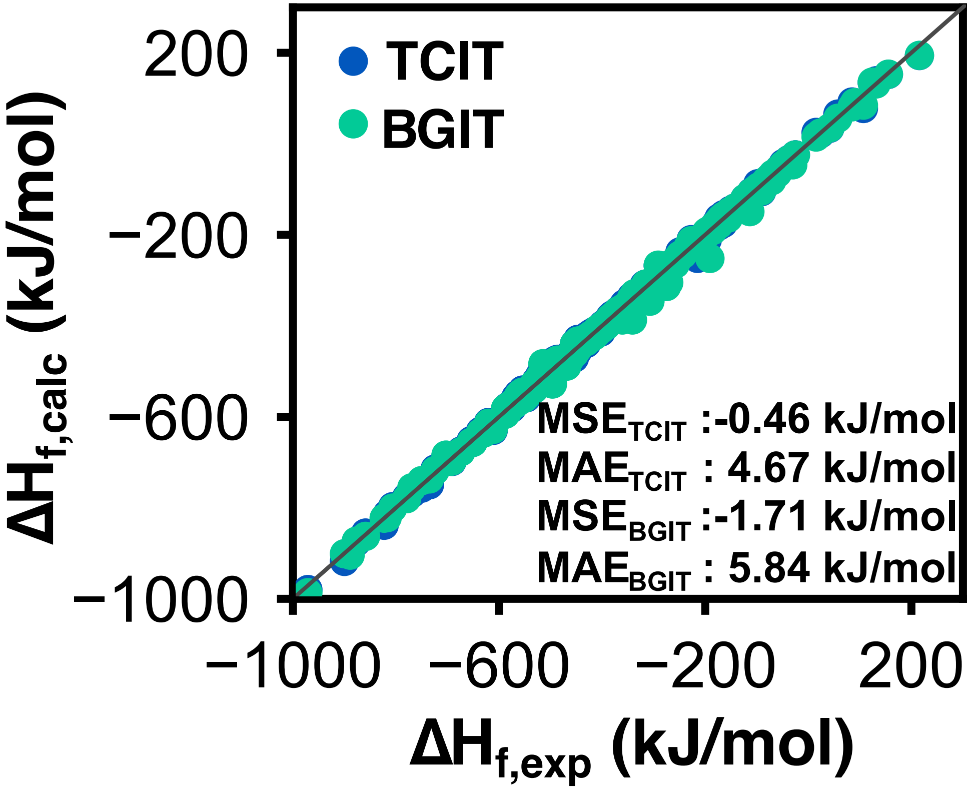
- PNK is a core dataset for fitting Benson groups

- ~600 PNK compounds are small enough for G4 calculations and comparison with experiment.

- ~150 PNK compounds are large enough for direct G4 calculation and comparison with TCIT.

- ~500 PNK compounds are large enough to evaluate the predictive accuracy of the increment theories.

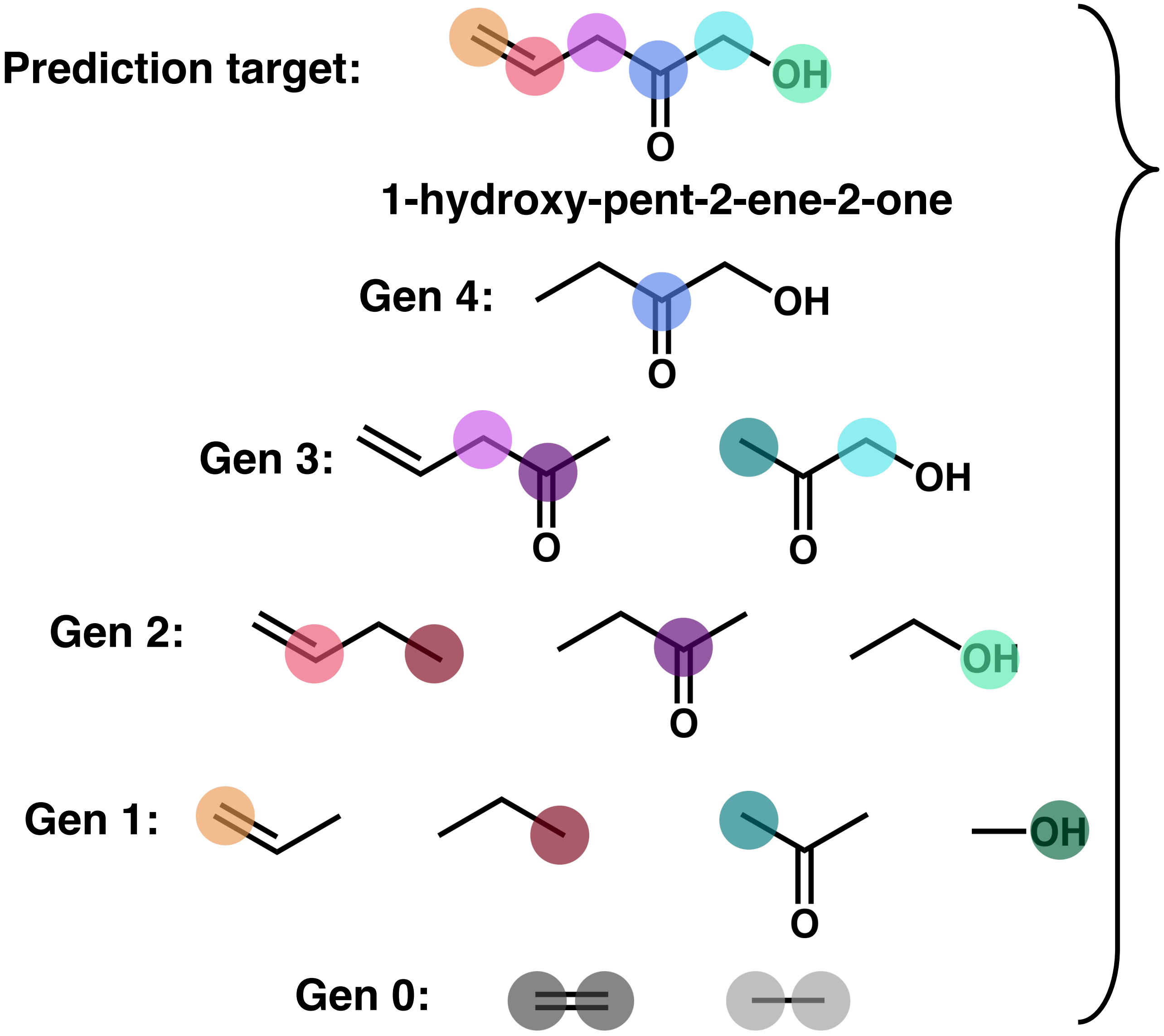
~500 large compounds from PNK



Zhao, Q.; Savoie, B. M.; Enthalpy of Formation Prediction via a Fully Self-Consistent Component Increment Theory. *J. Chem. Info. Model.* **2020**, 60, 2199-2207

TCIT shows comparable performance to BGIT/CHETAH but is derived exclusively from extensible G4 data.

How Many Components are Possible?



Prediction target:

1-hydroxy-pent-2-ene-2-one

Gen 4:

Gen 3:

Gen 2:

Gen 1:

Gen 0:

We database all model compounds and components for reuse.

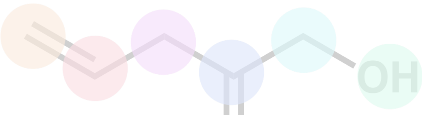
Over the past three years, we have parameterized new components in response to distinct project needs **(many from P2SAC Pharma Members)**

Current Database:

- ~35k distinct components for ΔH_f relevant to organic chemistry
- ~35k distinct G4 calculations on organic molecules.
- ~450 distinct ring corrections

How Many Components are Possible?

Prediction target:



We database all model compounds

How many components are required to predict the ΔH$_f$ of **all** (physically relevant) organic molecules?

response to distinct project needs (many from P2SAC Pharma Members)

Gen 3:

Current Database:

How many P2SAC funding periods would it take to make a “complete” or “gapless” component theory?

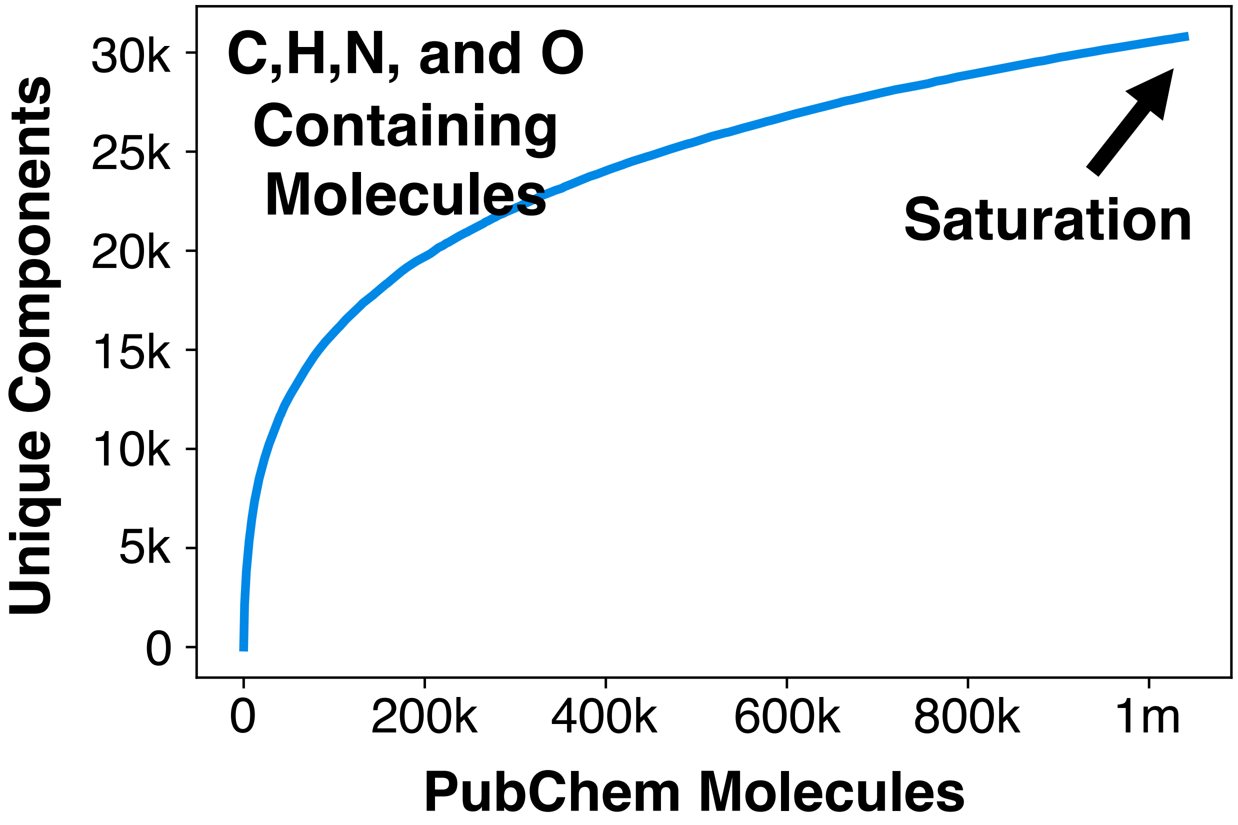
• ~450 distinct ring corrections

Gen 0:

Treating PubChem as a Model of Organic Chemical Space

PubChem is a repository of chemical properties that contains many millions of organic species ranging from small molecules to oligonucleotides.

We recently started mining PubChem's H,C,N, and O containing molecules for distinct components and the model compounds necessary to predict ΔH_f

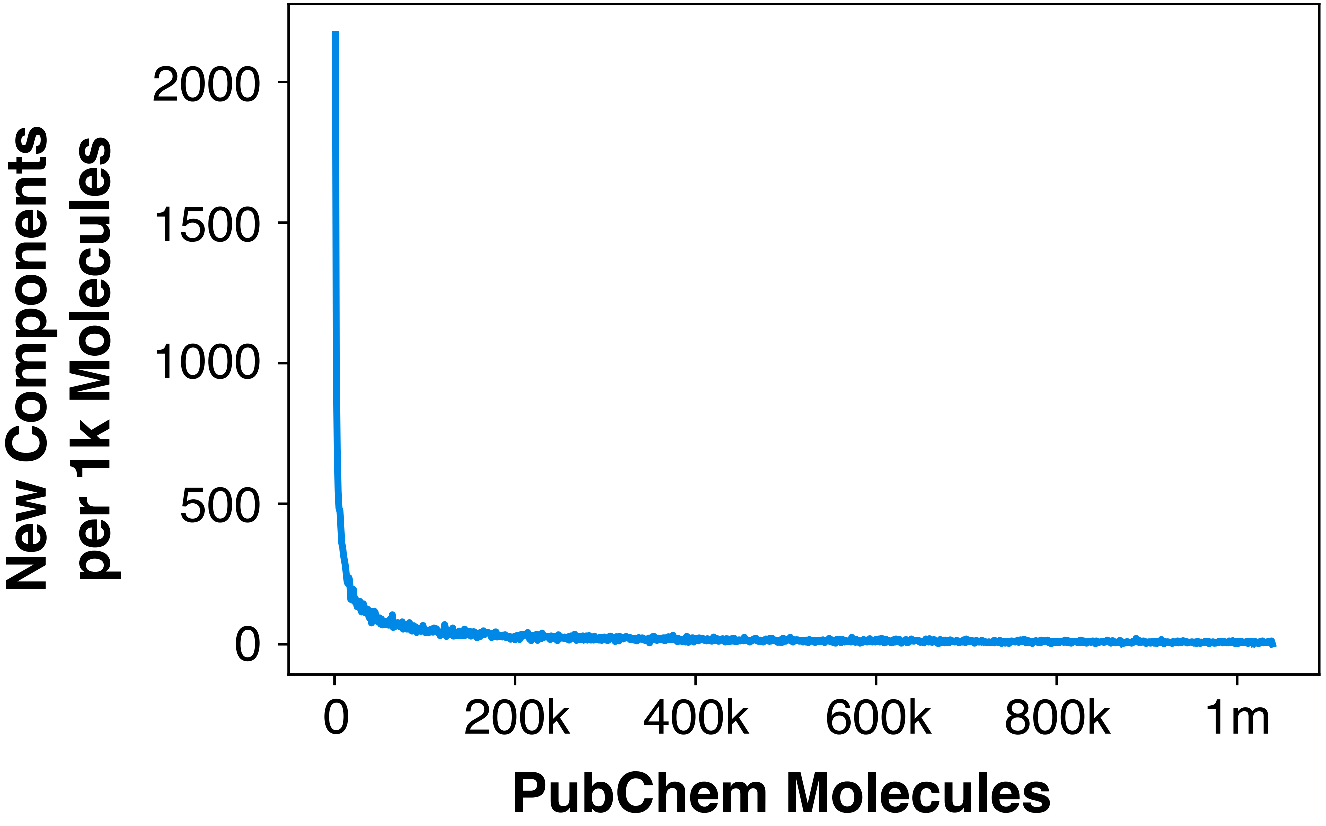


Treating PubChem as a Model of Organic Chemical Space

PubChem is a repository of chemical properties that contains many millions of organic species ranging from small molecules to oligonucleotides.

We recently started mining PubChem's H,C,N, and O containing molecules for distinct components and the model compounds necessary to predict ΔH_f

The derivative plot shows that TCIT initially generates ~2 new components per molecule, but by the end of the sampling ~100 molecules need to be sampled to find a new component.

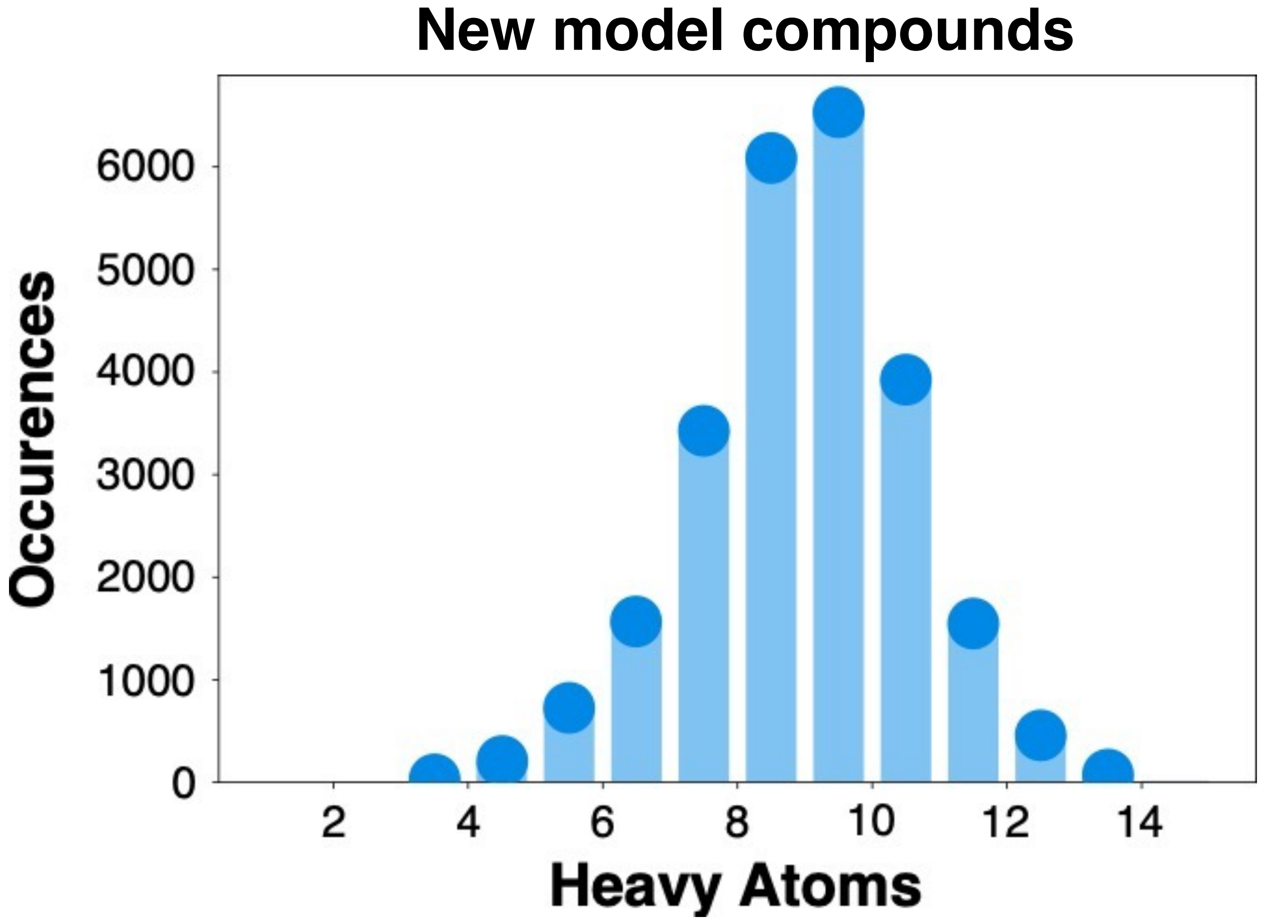


Treating PubChem as a Model of Organic Chemical Space

PubChem is a repository of chemical properties that contains many millions of organic species ranging from small molecules to oligonucleotides.

We recently started mining PubChem's H,C,N, and O containing molecules for distinct components and the model compounds necessary to predict ΔH_f

The derivative plot shows that TCIT initially generates ~2 new components per molecule, but by the end of the sampling ~100 molecules need to be sampled to find a new component.



Treating PubChem as a Model of Organic Chemical Space

PubChem is a repository of chemical properties that contains many millions of organic

TCIT now contains all CAVs necessary to predict ΔH_f of all N, H, O, and C-containing molecules in pubchem. **This is the largest repository of G4 calculations on large molecules in the world.**

W

H,

C

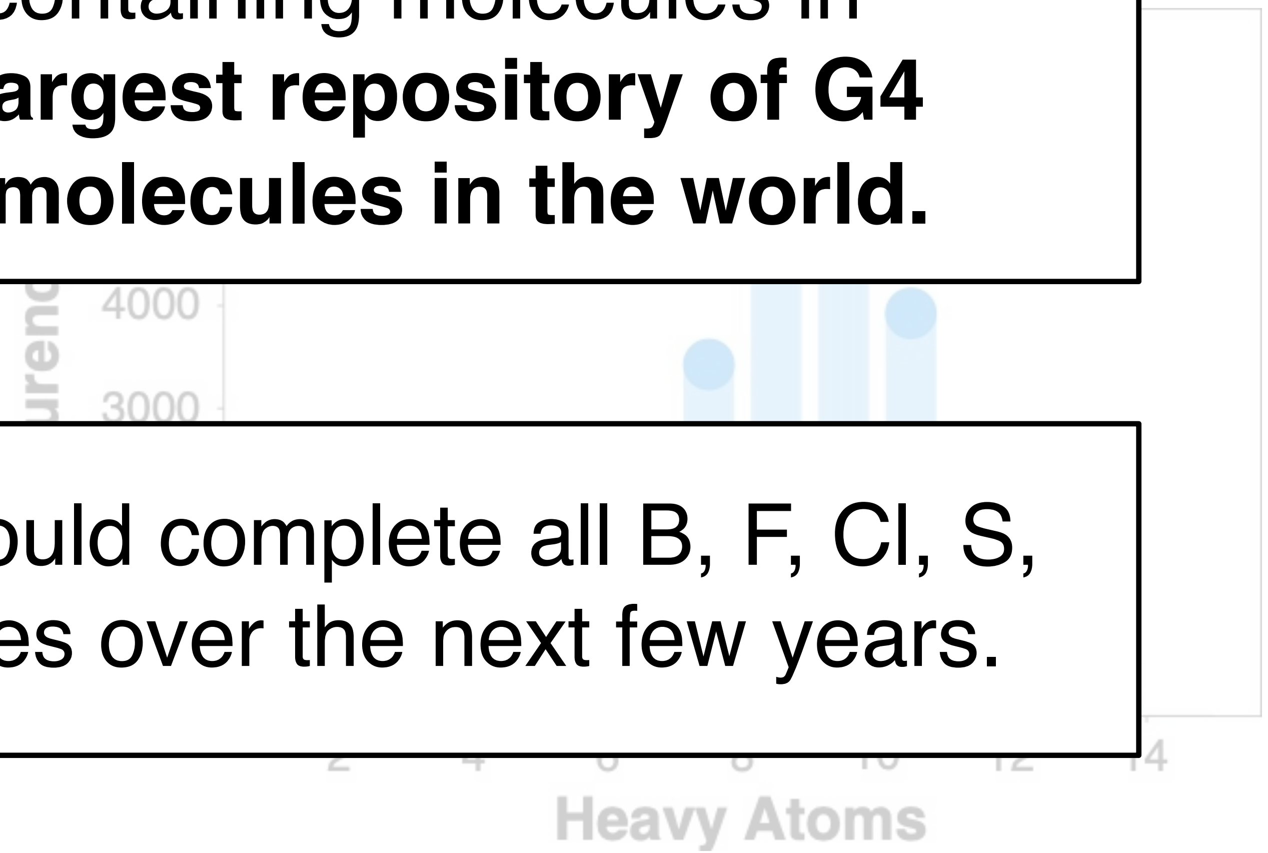
It is foreseeable that we could complete all B, F, Cl, S, and P containing structures over the next few years.

T

ini

sa

sampled to find a new component.

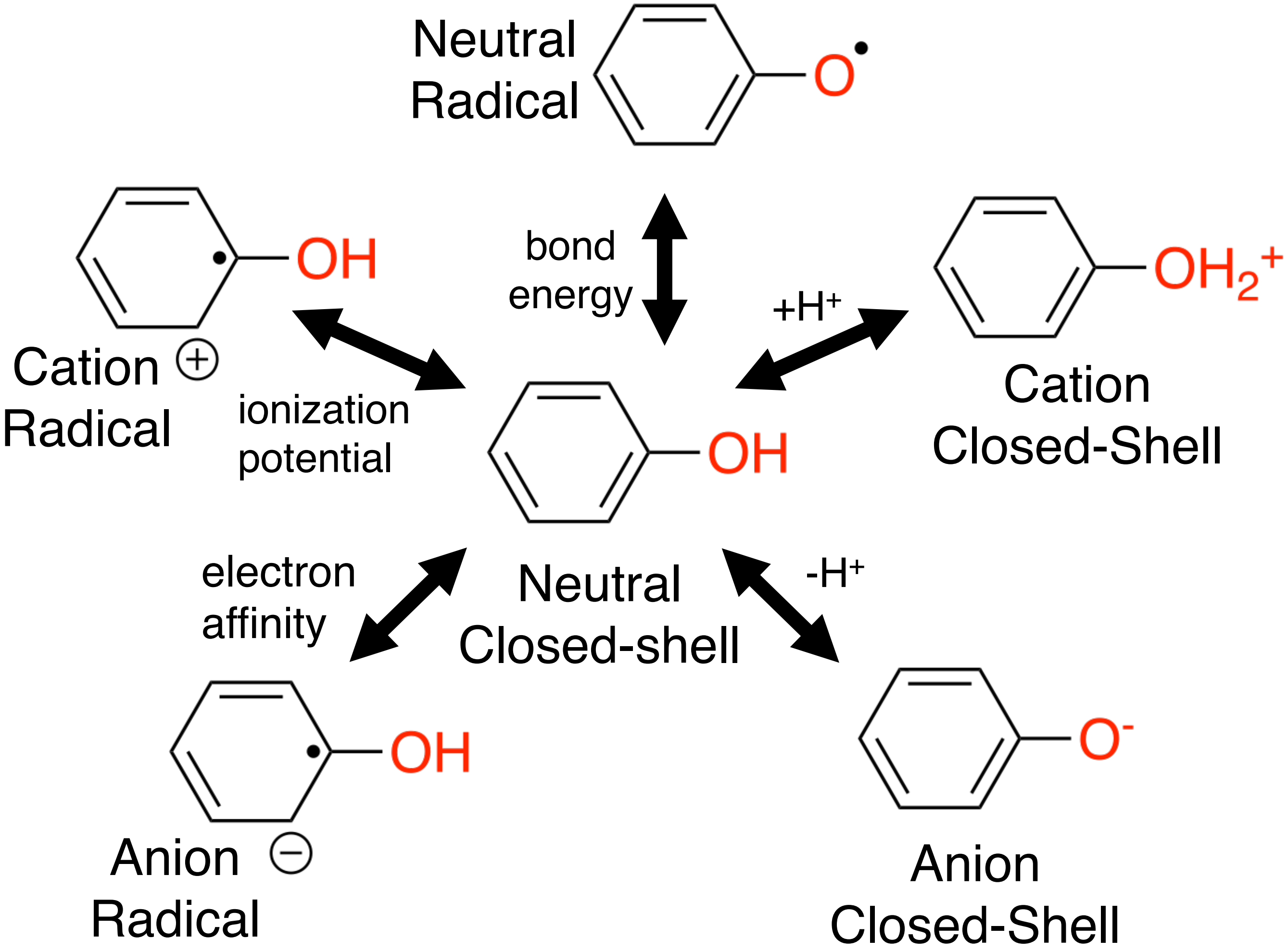


Extending TCIT to Radicals and Ions

A recurring question is when will TCIT support predictions on **radicals** and **ions**?

TCIT already covers neutral close-shell species, so these extensions require us only to predict the difference between the target and the **nearest closed-shell neutral**.

This amounts to developing models to predict IP/EA/+H⁺/-H⁺

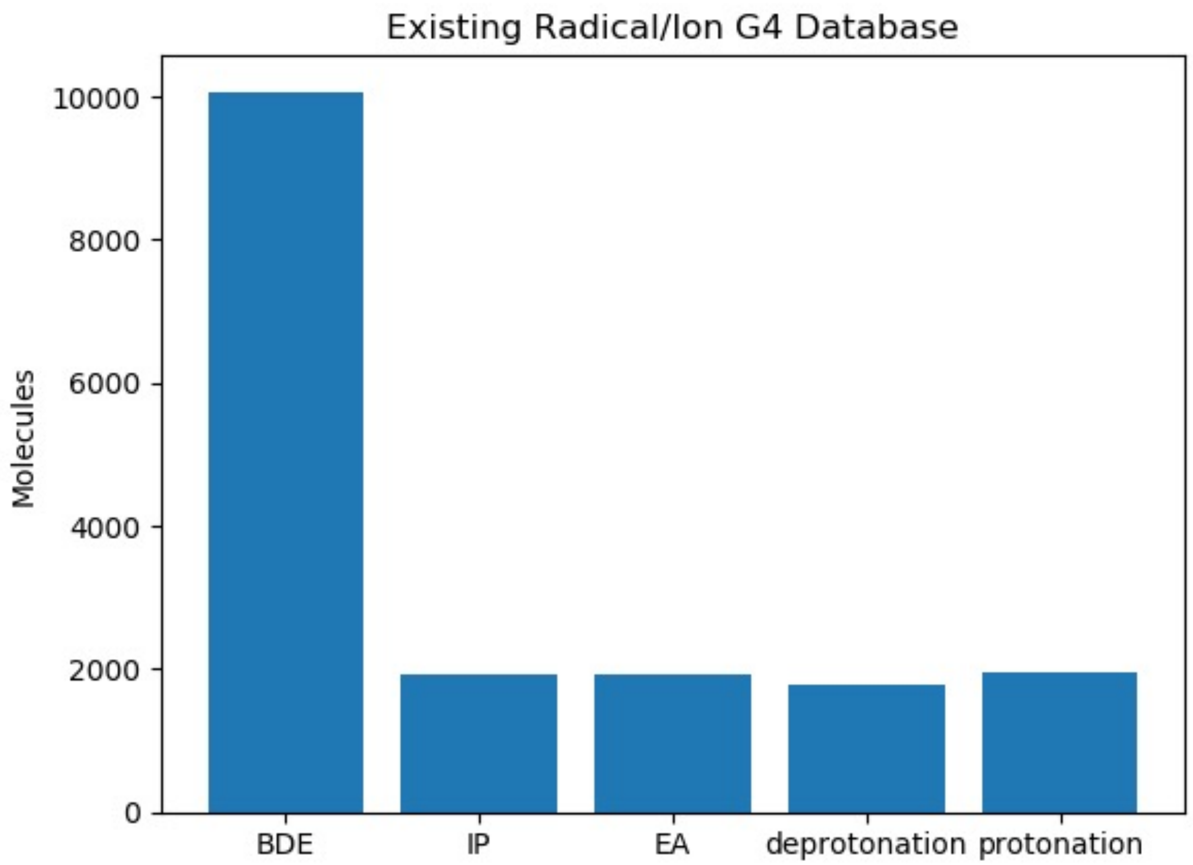


Extending TCIT to Radicals and Ions

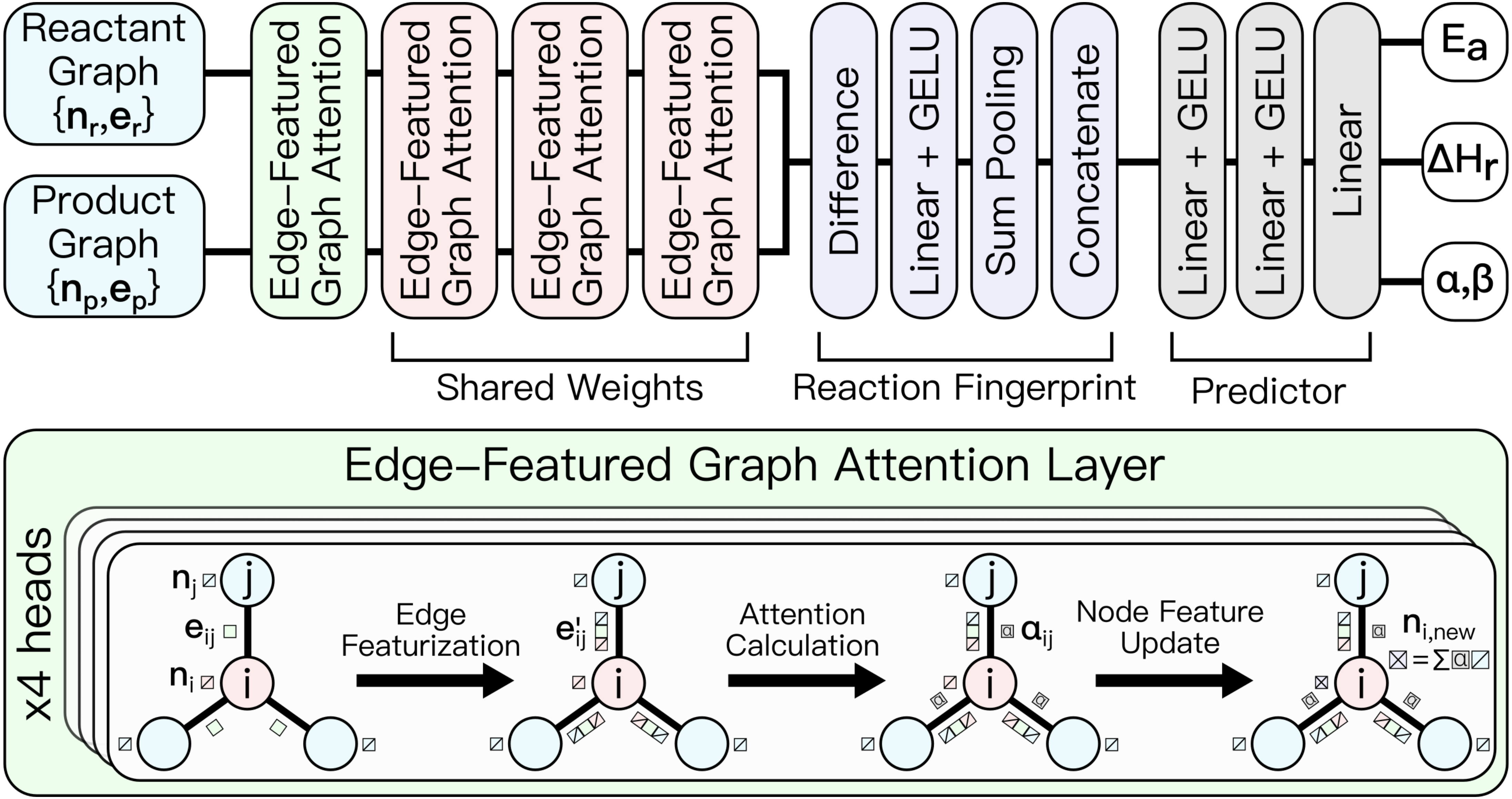
A recurring question is when will TCIT support predictions on **radicals** and **ions**?

TCIT already covers neutral close-shell species, so these extensions require us only to predict the difference between the target and the **nearest closed-shell neutral**.

This amounts to developing models to predict IP/EA/+H$^+$/-H$^+$



Graph → IP/EA/$+H^+$/$-H^+$ Models



The EGAT architecture has already been developed by our group for activation energy prediction

Vadaddi, S. M.; Zhao, Q.; Savoie, B. M. *ChemRxiv.* 2023, https://doi.org/10.26434/chemrxiv-2023-tdl2v

“Known Unknowns” and “Unknown Unknowns”

A → B

- To safely plan a known reaction, we need access to solid thermodynamic data (e.g., ΔH_f, S^o, C_v) to understand and classify risks.

- This is a “**known unknown**” in that we know the reaction, A → B, but we need values for a few unknown variables.

A → ? → B ; A → B + ? ; A → ?

- **A → ? → B**, means that we know the net reaction, but there may be a consequential (e.g., potentially reactive) intermediate. Even if we have accurate thermodynamic data on A/B, neglecting the intermediate could be disastrous.

- The **A → B + ?** (unknown side-reaction) and **A → ?** (unknown main product), problems have similar “**unknown unknown**” characteristics.

The Reaction Prediction Problem

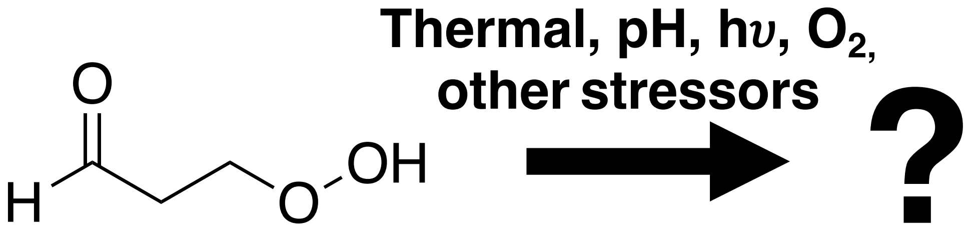
A → B : When we know the reactants and products, mature quantum chemistry tools exist to characterize transition states and establish pathways

A → ? : For degradation reactions, plausible reactions are often unknown.

The Reaction Prediction Problem

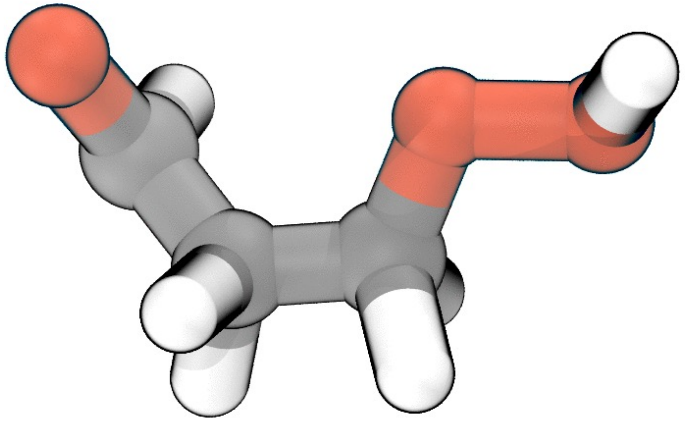
A → B : When we know the reactants and products, mature quantum chemistry tools exist to characterize transition states and establish pathways

A → ? : For degradation reactions, plausible reactions are often unknown.



Thermal, pH, hν, O_2, other stressors

3-hydroperoxypropanal

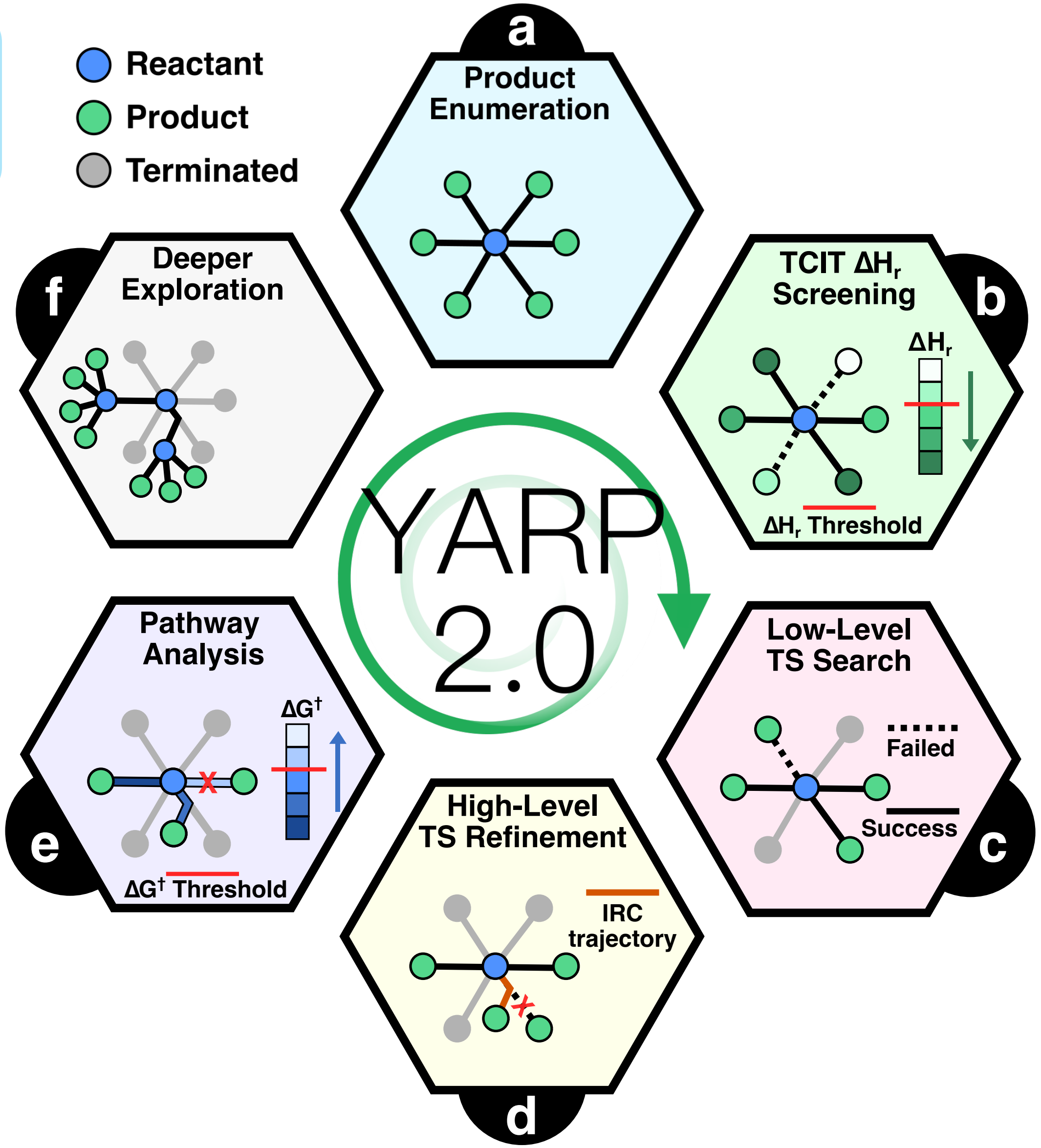


Yet Another Reaction Program (YARP)

Idea: Turn the **A→?** problem into tractable (and parallelizable) **A→B** problems.

Observations:

- Product enumeration is easier than transition state enumeration.
- Transition state algorithms for A→B problems are mature. Let the TS algorithm identify physical reactions.
- Recent developments in semi-empirical models and ML create opportunities.
- Solving the **A→?** problem is the prerequisite for reaction network prediction.

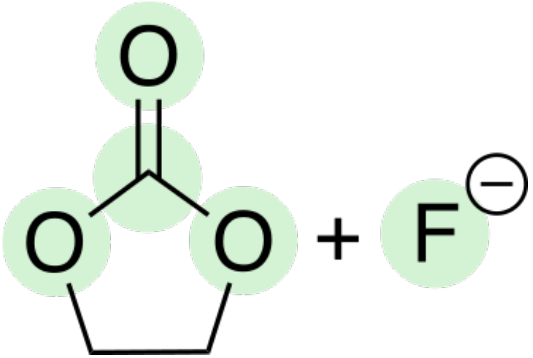


Zhao, Q.; Savoie, B. M. *Nature Comput. Sci.* 2021 ; Zhao, Q.; Savoie, B. M. *Angew. Chemie* 2022.

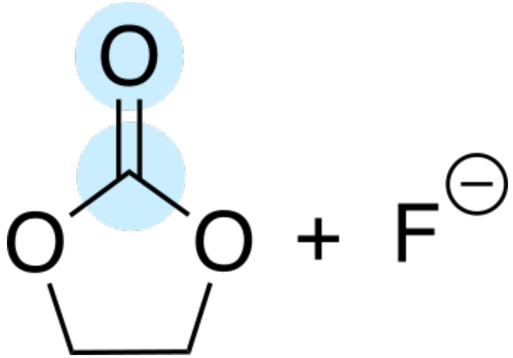
YARP: Elementary Reaction Step(s)

Polar and pericyclic organic reactions are decomposed into elementary electron donor and acceptor reactions with concomitant σ-bond breaks

bnfn
will refer to σ-bond changes, π-bonds are allowed to arbitrarily rearrange.



Lone-Pair Donors

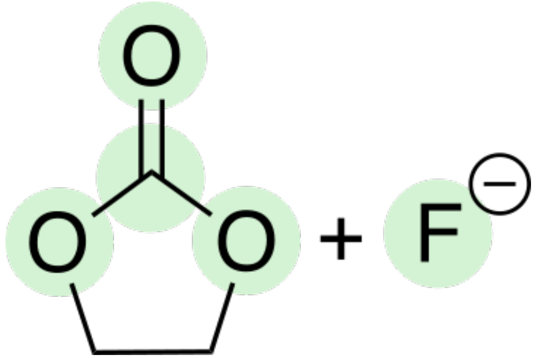


Lone-Pair Acceptors

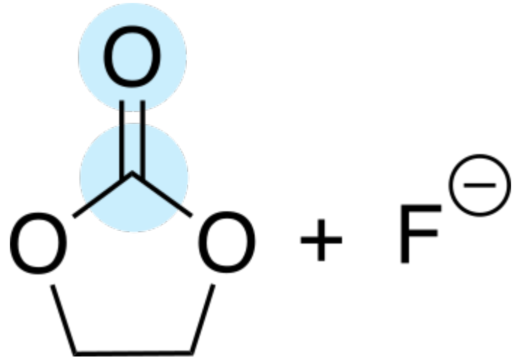
YARP: Elementary Reaction Step(s)

Polar and pericyclic organic reactions are decomposed into elementary electron donor and acceptor reactions with concomitant σ-bond breaks

bnfn
will refer to σ-bond changes, π-bonds are allowed to arbitrarily rearrange.

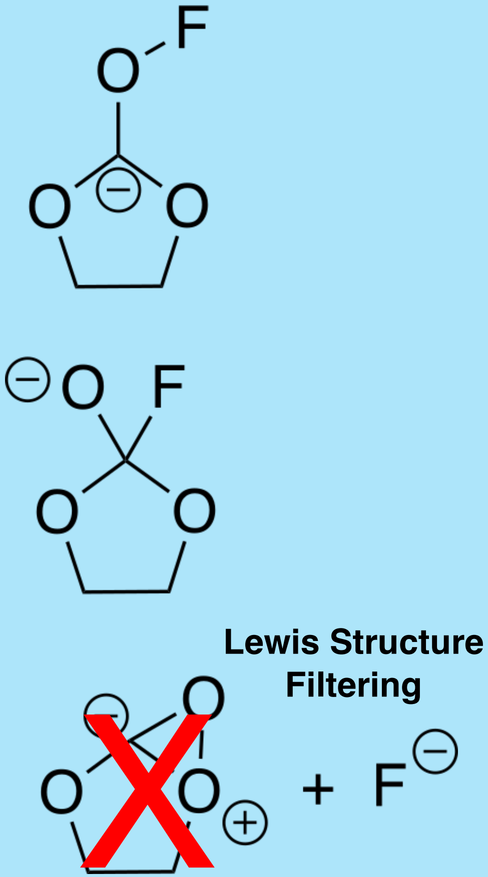


Lone-Pair Donors



Lone-Pair Acceptors

Form 1 Products



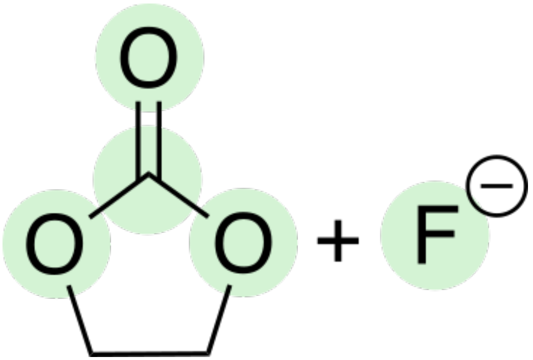
Lewis Structure Filtering

Hsu, H. H., Zhao, Q., Savoie, B. M. *In Preparation*

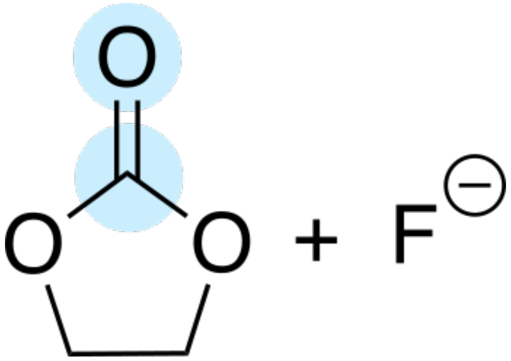
YARP: Elementary Reaction Step(s)

Polar and pericyclic organic reactions are decomposed into elementary electron donor and acceptor reactions with concomitant σ-bond breaks

bnfn will refer to σ-bond changes, π-bonds are allowed to arbitrarily rearrange.

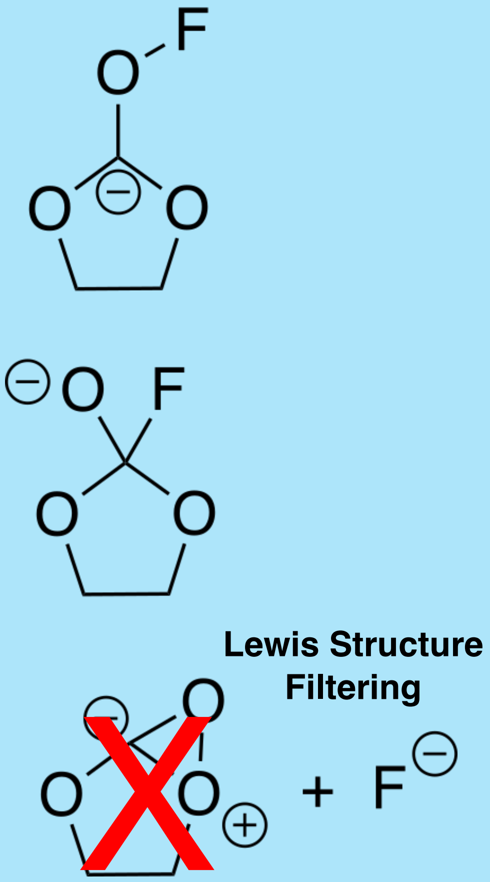


Lone-Pair Donors



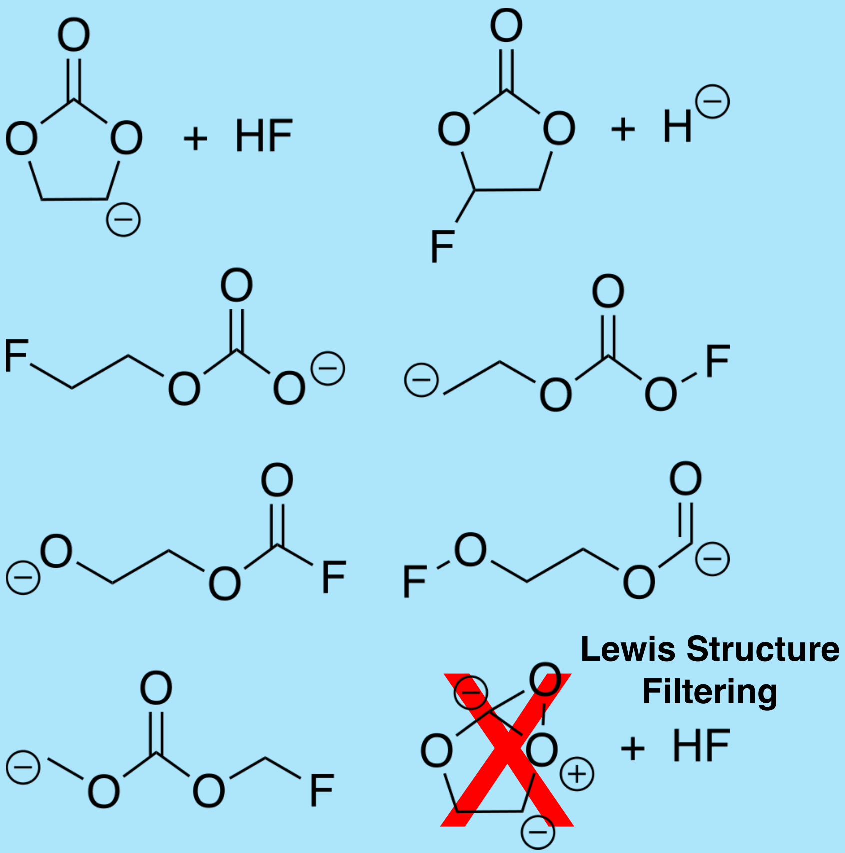
Lone-Pair Acceptors

Form 1 Products



Lewis Structure Filtering

Break 1 Form 1 Products



Lewis Structure Filtering

+ 28 others

Hsu, H. H., Zhao, Q., Savoie, B. M. *In Preparation*

YARP: Elementary Reaction Step(s)

Polar and pericyclic organic reactions are decomposed into elementary electron donor and acceptor reactions with concomitant σ-bond breaks

Form 1 Products

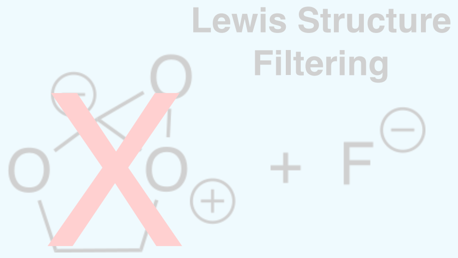
Break 1 Form 1 Products

All bnfn products are b(n-1)f(n-1) decomposable

This means that using only “break 1 bond form 1 bond” (b1f1) for radicals and ions won’t miss any products, but it will potentially miss important transition states (i.e., by predicting a sequential mechanism when a concerted mechanism is favored)

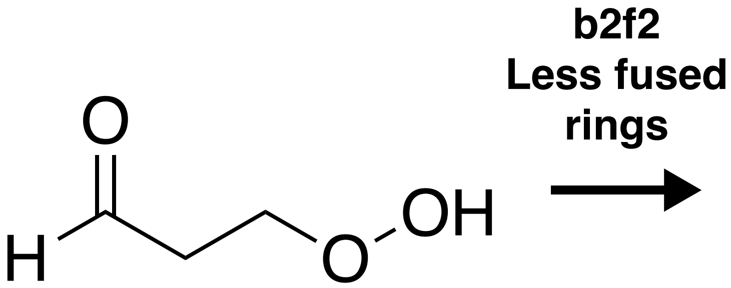
wil
c
ch
π-b
a
to arbitrarily rearrange.

Lone-Pair Acceptors



+ 28 others

Testing YARP on a Unimolecular Decomposition Problem

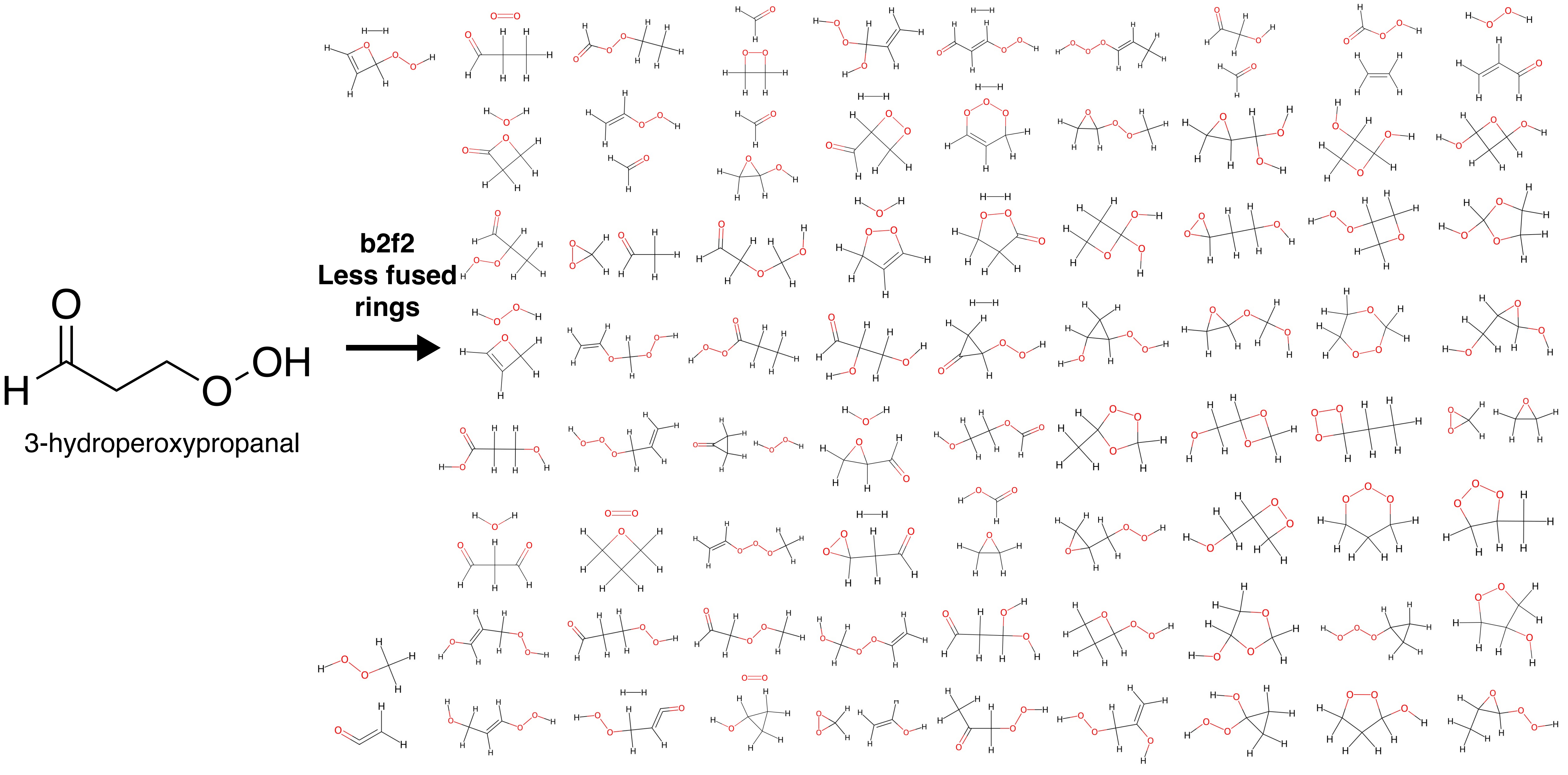


b2f2
Less fused
rings

3-hydroperoxypropanal

Zhao, Q.; Savoie, B. M. *Nature Computational Science* 2021, 479-490.

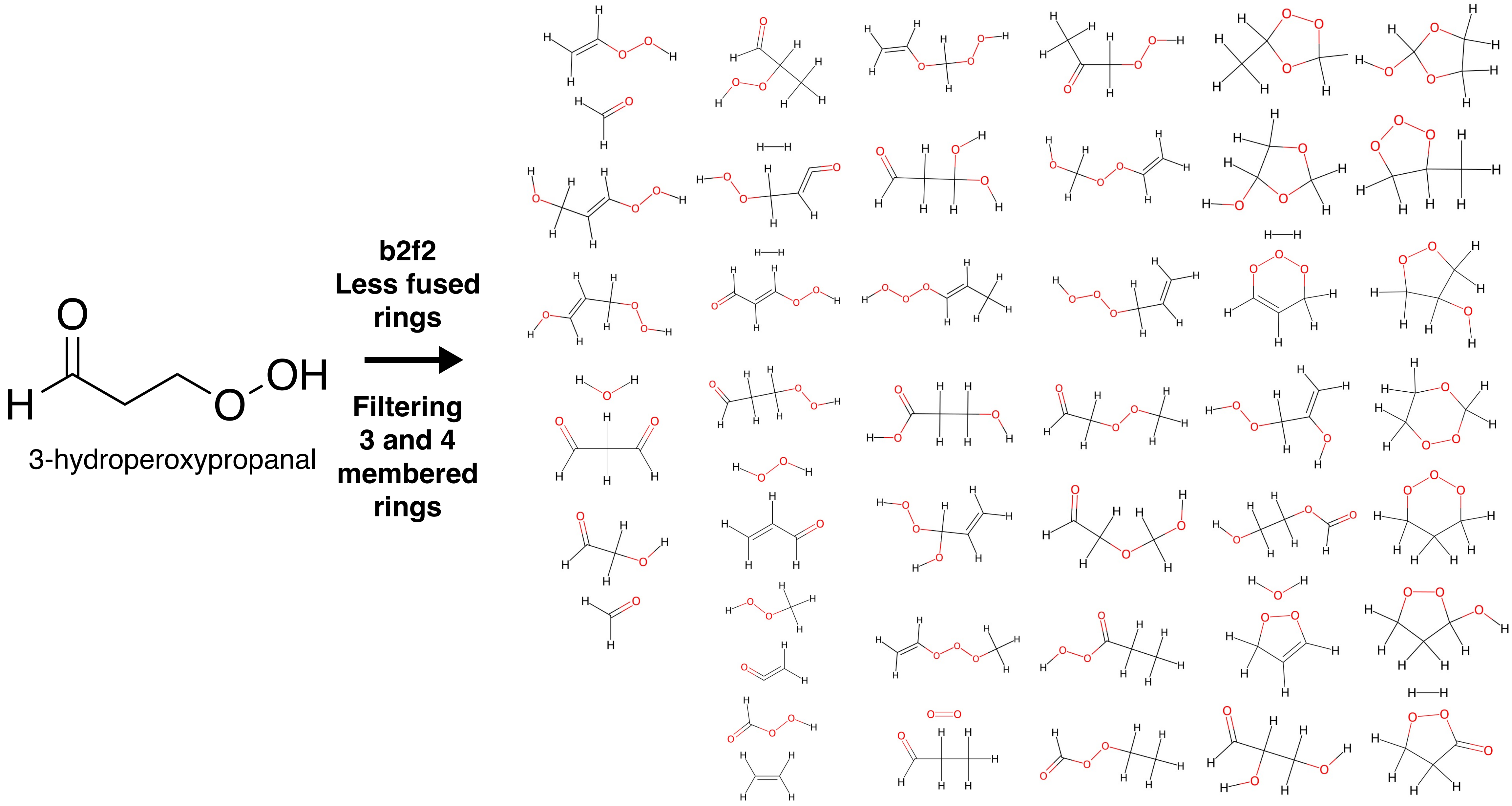
Testing YARP on a Unimolecular Decomposition Problem



3-hydroperoxypropanal

Zhao, Q.; Savoie, B. M. *Nature Computational Science* 2021, 479-490.

Testing YARP on a Unimolecular Decomposition Problem



3-hydroperoxypropanal

b2f2
Less fused rings

Filtering 3 and 4 membered rings

Zhao, Q.; Savoie, B. M. *Nature Computational Science* 2021, 479-490.

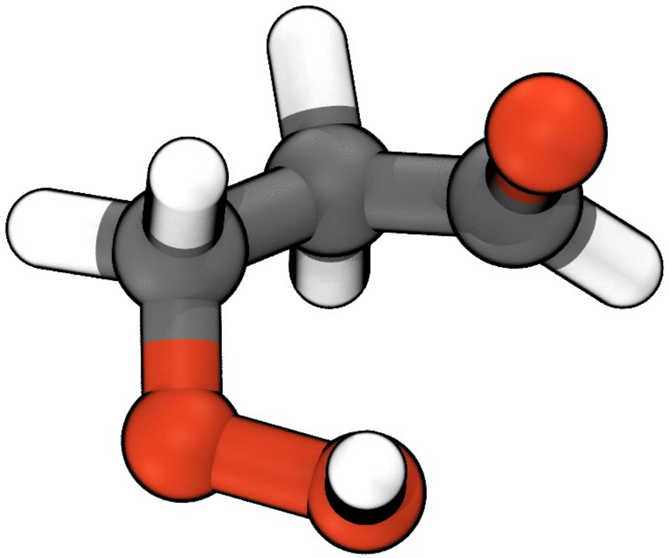
What Happens First?

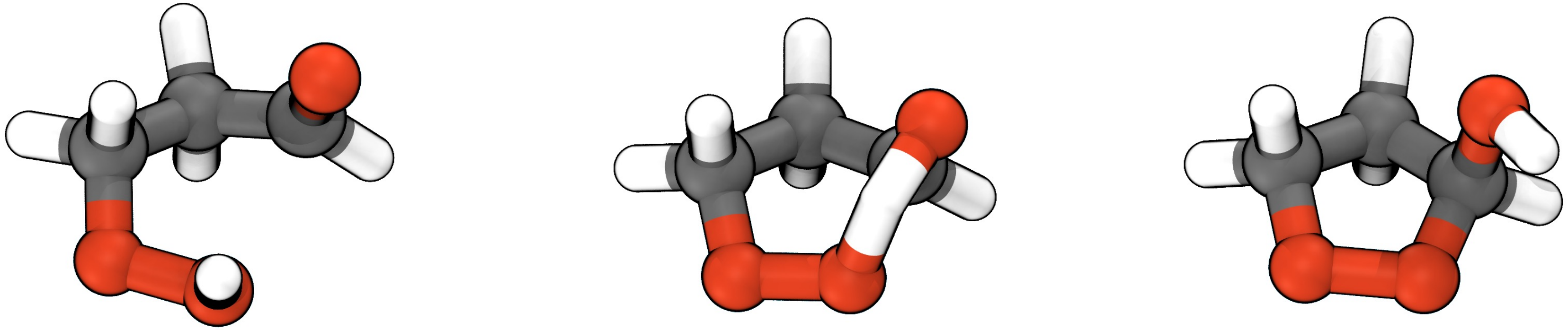
Jensen, R. K.; Korcek, S.; Mahoney, L. R.; Zinbo, M. *JACS* **1979**, 101, 7574

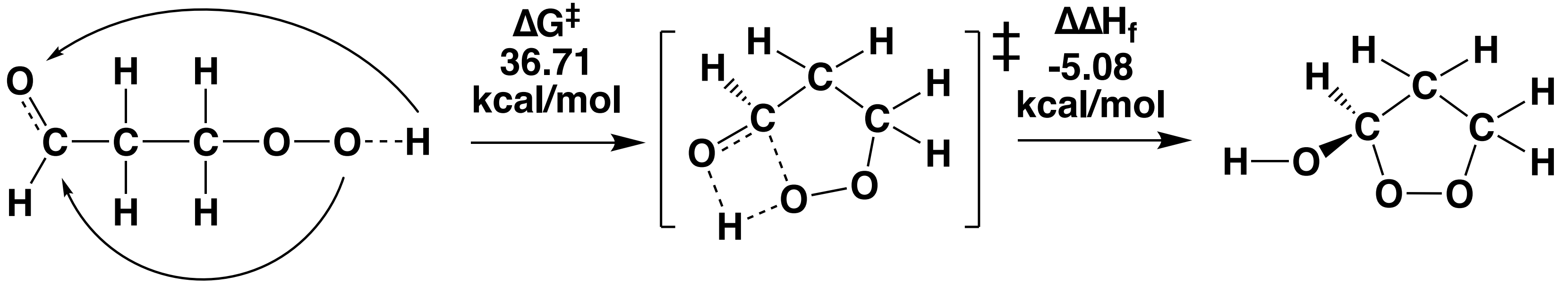
The Korcek Mechanism

According to YARP, this is the lowest barrier unimolecular reaction.

MEP (B3LYP/6-31+G*)







Fully resolved (along with subsequent ROOH and R=O formation) 30 years later by Green and Truhlar: Jalan, A.; Alecu, I. M.; Meana-Pañeda, R.; Aguilera-Iparraguirre, J.; Yang, K. R.; Merchant, S. S.; Truhlar, D. G.; Green, W. H. *JACS* **2013**, *135* (30), 11100–11114.

Reaction Network Case Study: β-D-Glucose Pyrolysis

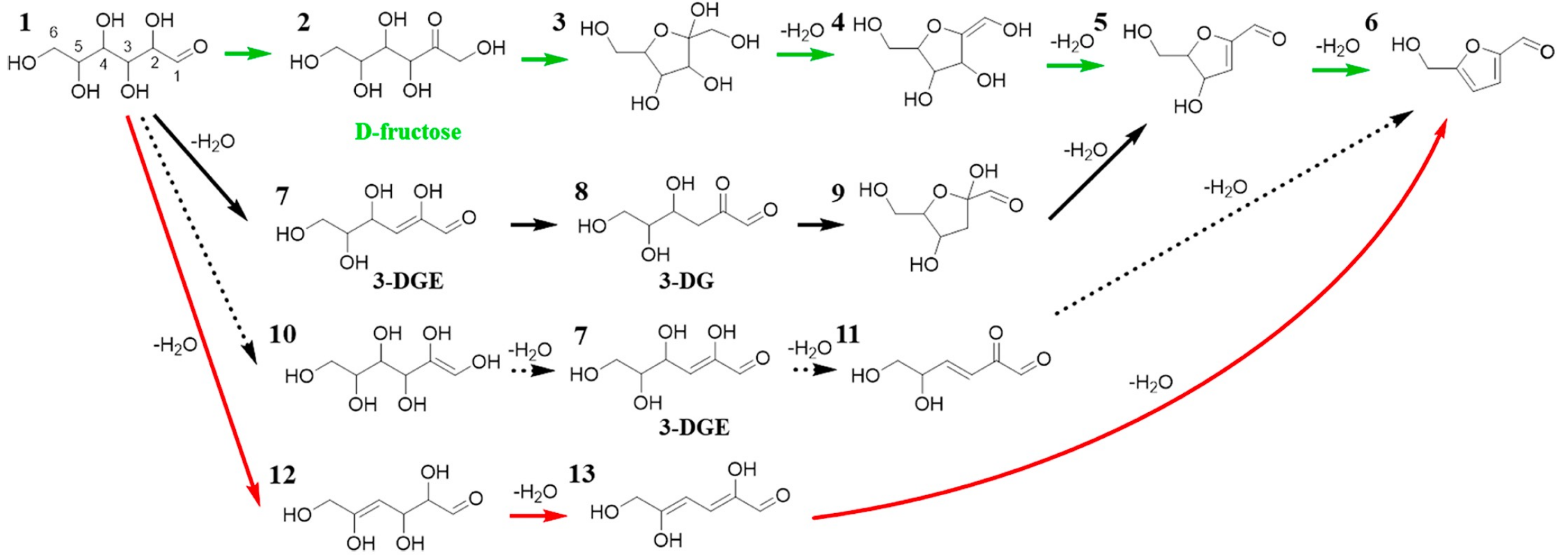
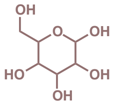


Figure 1. Proposed pathways in literature from glucose to HMF, namely the fructose path (green), 3-DG paths (black and black dotted), and direct path (red). The molecules are indicated by numbers and some key molecules are named as follows: **1**. D-glucose; **2**. D-fructose; **3**. D-fructofuranose; **6**. 5-hydroxymethylfurfural (5-HMF); **7**. 3-deoxyglucos-2-ene (3-DGE); **8**. 3-deoxyglucosone (3-DG); and **10**. hex-1-ene-1,2,3,4,5,6-hexaol (enol form of glucose).

Kang, P.L., Shang, C. and Liu, Zhi-Pan, **2019**. *JACS,* 141, 20525-20536

β-D-Glucose Pyrolysis Network Exploration

Zhao, Q.; Savoie, B. M. Proc. Nat. Acad. Sci. 120, 2023



Depth 1: ▨

To perform a deep network exploration, we've implemented a modified version of Dijkstra's algorithm

At each iteration:

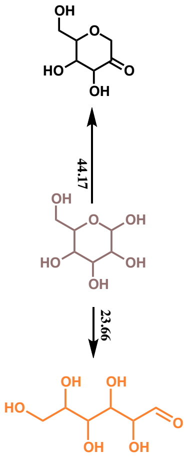
(1) all b2f2 reactions are explored for active nodes.

(2) Active nodes are determined by the minimum barrier to a given product (with a window)

(3) Water catalyzed reactions are considered for all H-transfers

β-D-Glucose Pyrolysis Network Exploration

Zhao, Q.; Savoie, B. M. Proc. Nat. Acad. Sci. 120, 2023



Depth 1: 🟫

Depth 2: 🟧

To perform a deep network exploration, we've implemented a modified version of Dijkstra's algorithm

At each iteration:

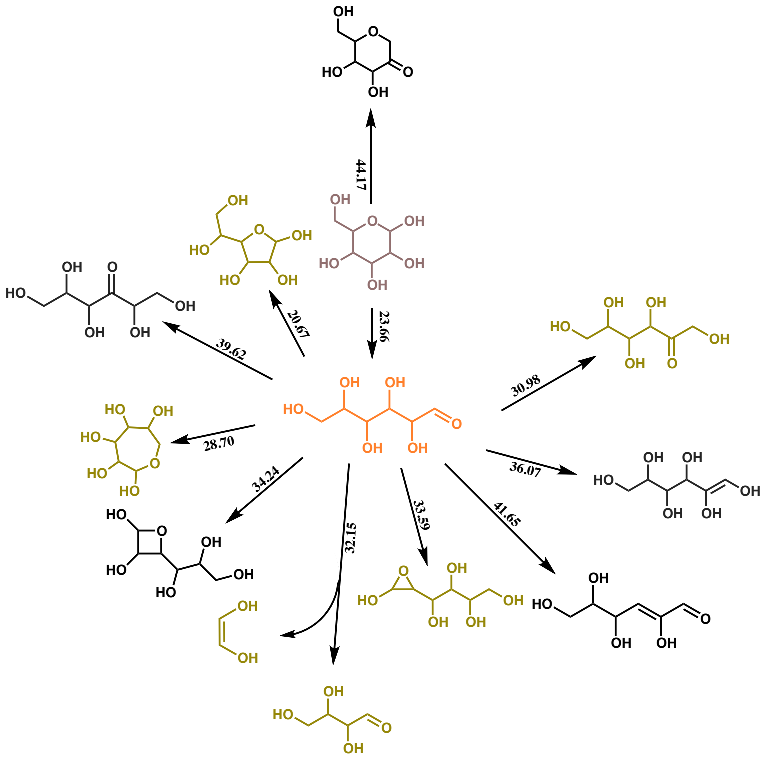
(1) all b2f2 reactions are explored for active nodes.

(2) Active nodes are determined by the minimum barrier to a given product (with a window)

(3) Water catalyzed reactions are considered for all H-transfers

β-D-Glucose Pyrolysis Network Exploration

Zhao, Q.; Savoie, B. M. Proc. Nat. Acad. Sci. 120, 2023



Depth 1: 🟫

Depth 2: 🟧

Depth 3: 🟨

To perform a deep network exploration, we've implemented a modified version of Dijkstra's algorithm

At each iteration:

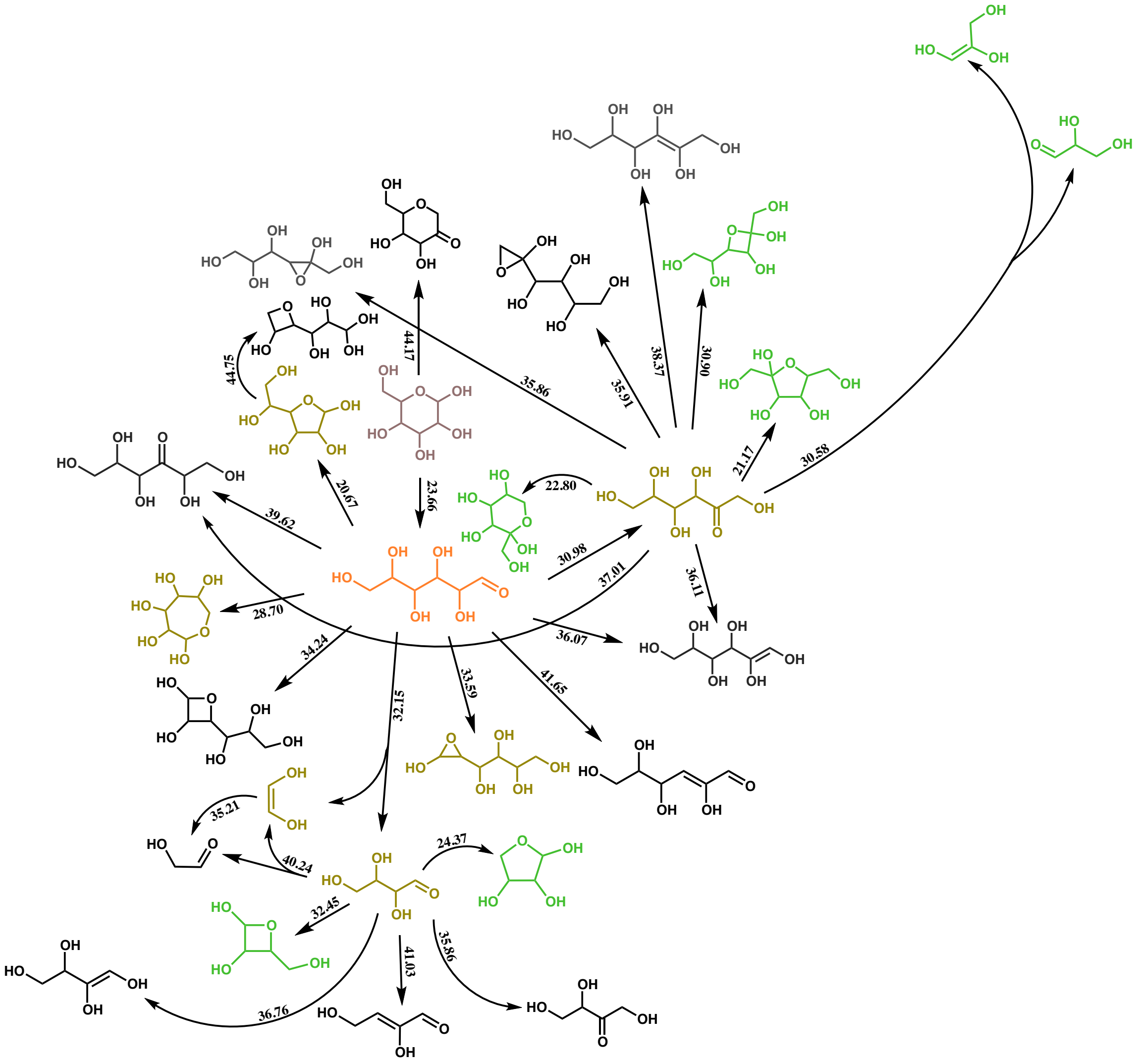
(1) all b2f2 reactions are explored for active nodes.

(2) Active nodes are determined by the minimum barrier to a given product (with a window)

(3) Water catalyzed reactions are considered for all H-transfers

β-D-Glucose Pyrolysis Network Exploration

Zhao, Q.; Savoie, B. M. Proc. Nat. Acad. Sci. 120, 2023



Depth 1: 🟫

Depth 2: 🟧

Depth 3: 🟨

Depth 4: 🟩

To perform a deep network exploration, we've implemented a modified version of Dijkstra's algorithm

At each iteration:

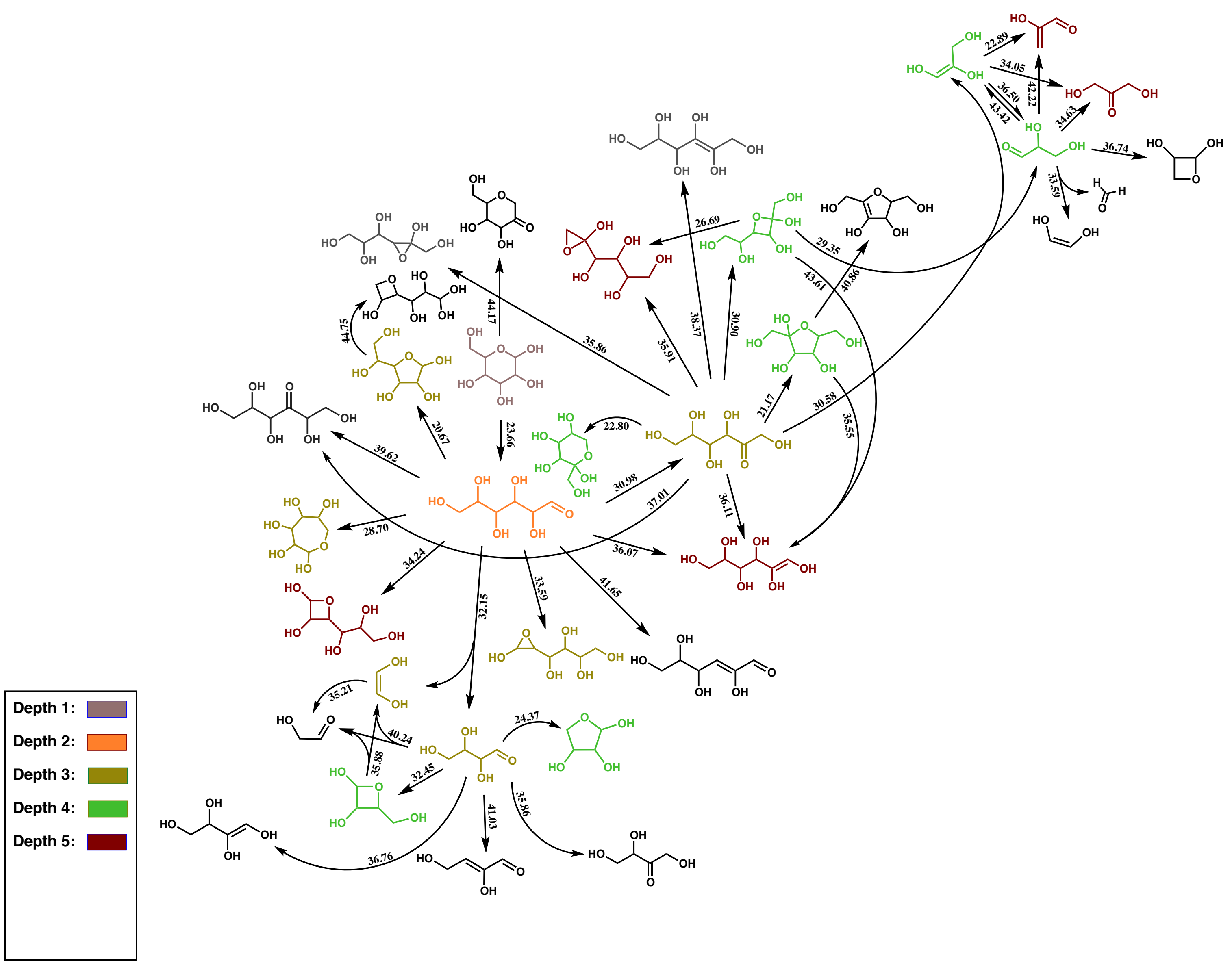
(1) all b2f2 reactions are explored for active nodes.

(2) Active nodes are determined by the minimum barrier to a given product (with a window)

(3) Water catalyzed reactions are considered for all H-transfers

β-D-Glucose Pyrolysis Network Exploration

Zhao, Q.; Savoie, B. M. Proc. Nat. Acad. Sci. 120, 2023



To perform a deep network exploration, we've implemented a modified version of Dijkstra's algorithm

At each iteration:

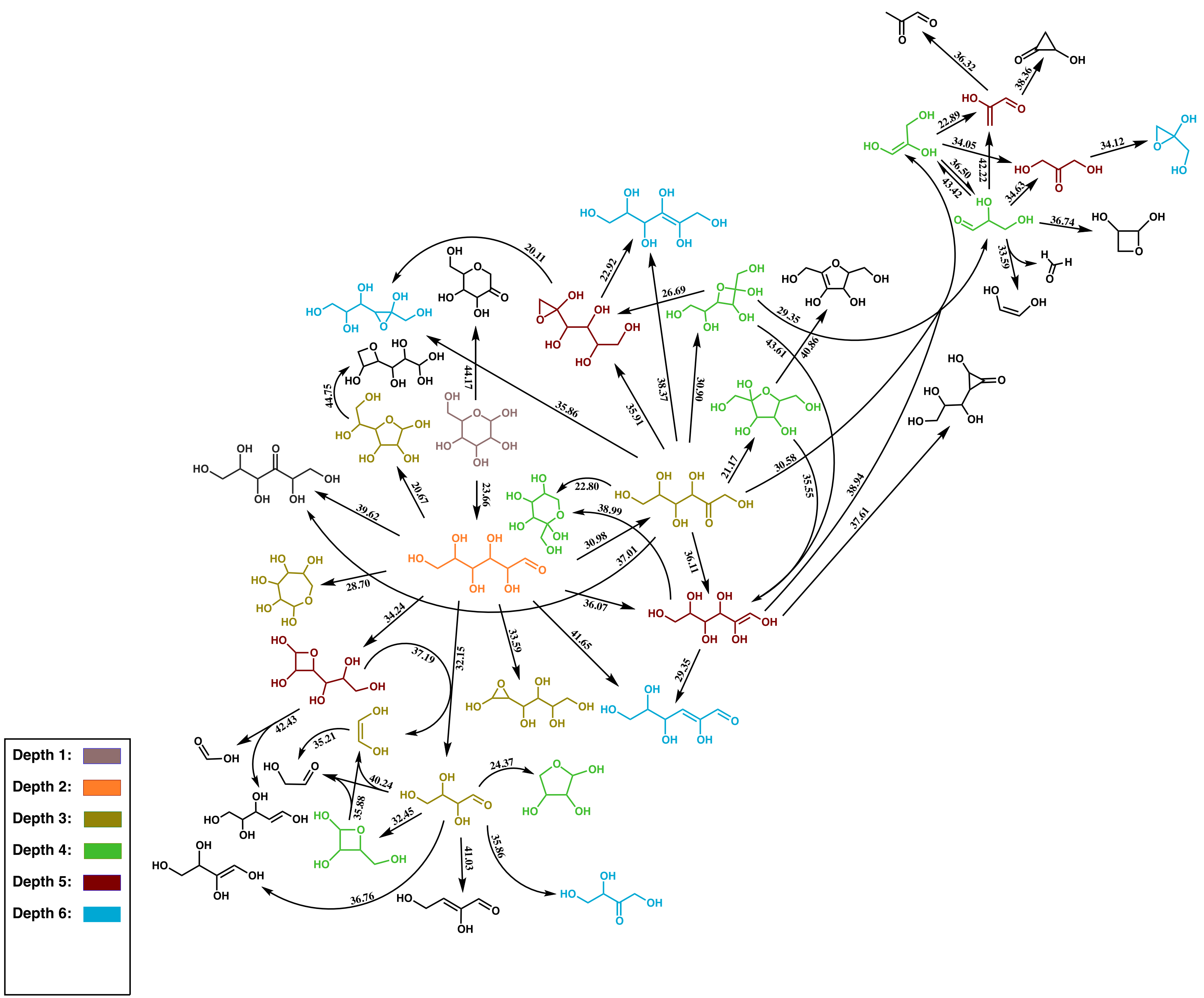
(1) all b2f2 reactions are explored for active nodes.

(2) Active nodes are determined by the minimum barrier to a given product (with a window)

(3) Water catalyzed reactions are considered for all H-transfers

β-D-Glucose Pyrolysis Network Exploration

Zhao, Q.; Savoie, B. M. Proc. Nat. Acad. Sci. 120, 2023



To perform a deep network exploration, we've implemented a modified version of Dijkstra's algorithm

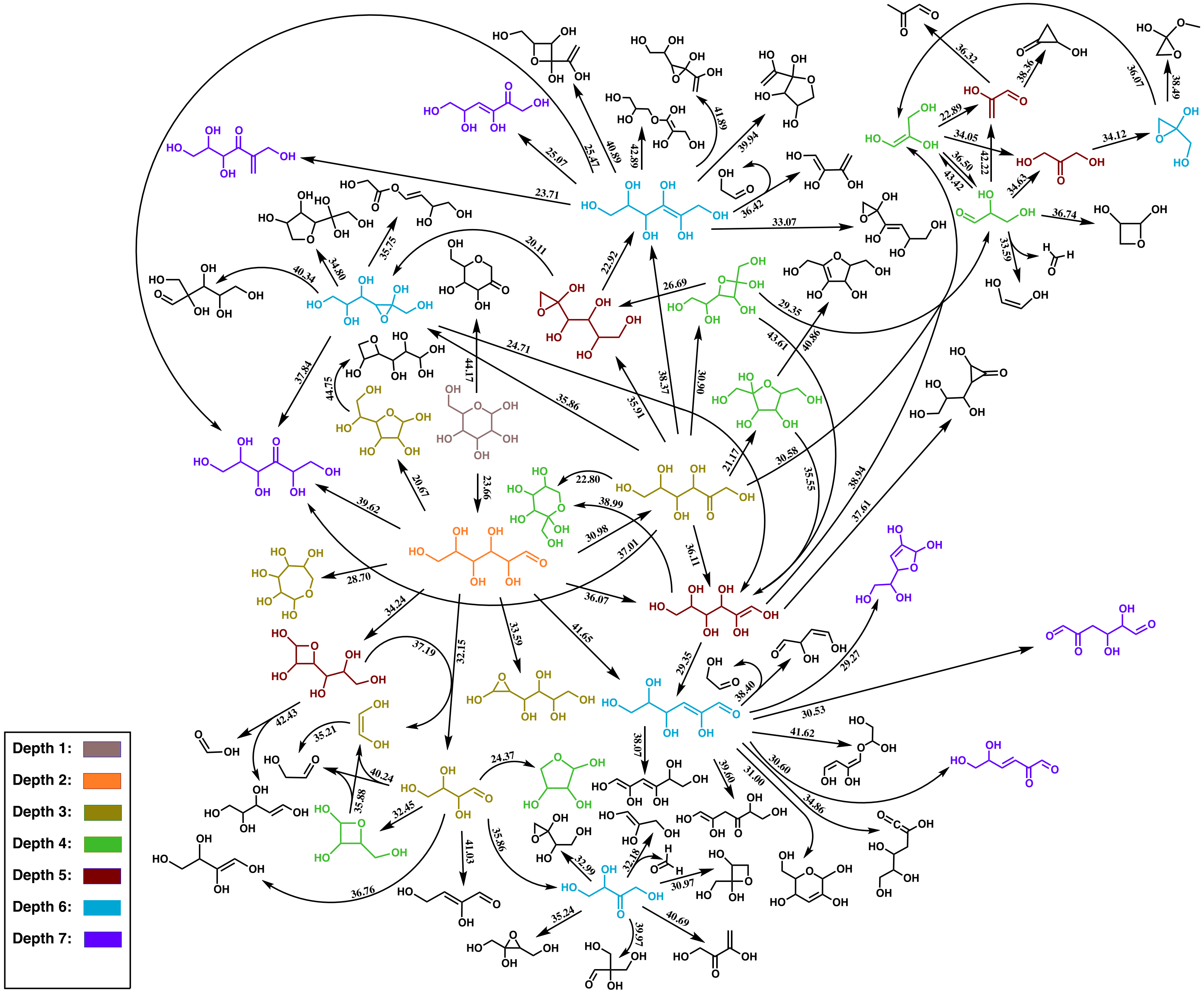
At each iteration:

(1) all b2f2 reactions are explored for active nodes.

(2) Active nodes are determined by the minimum barrier to a given product (with a window)

(3) Water catalyzed reactions are considered for all H-transfers

β-D-Glucose Pyrolysis Network Exploration



Zhao, Q.; Savoie, B. M. Proc. Nat. Acad. Sci. 120, 2023

To perform a deep network exploration, we've implemented a modified version of Dijkstra's algorithm

At each iteration:

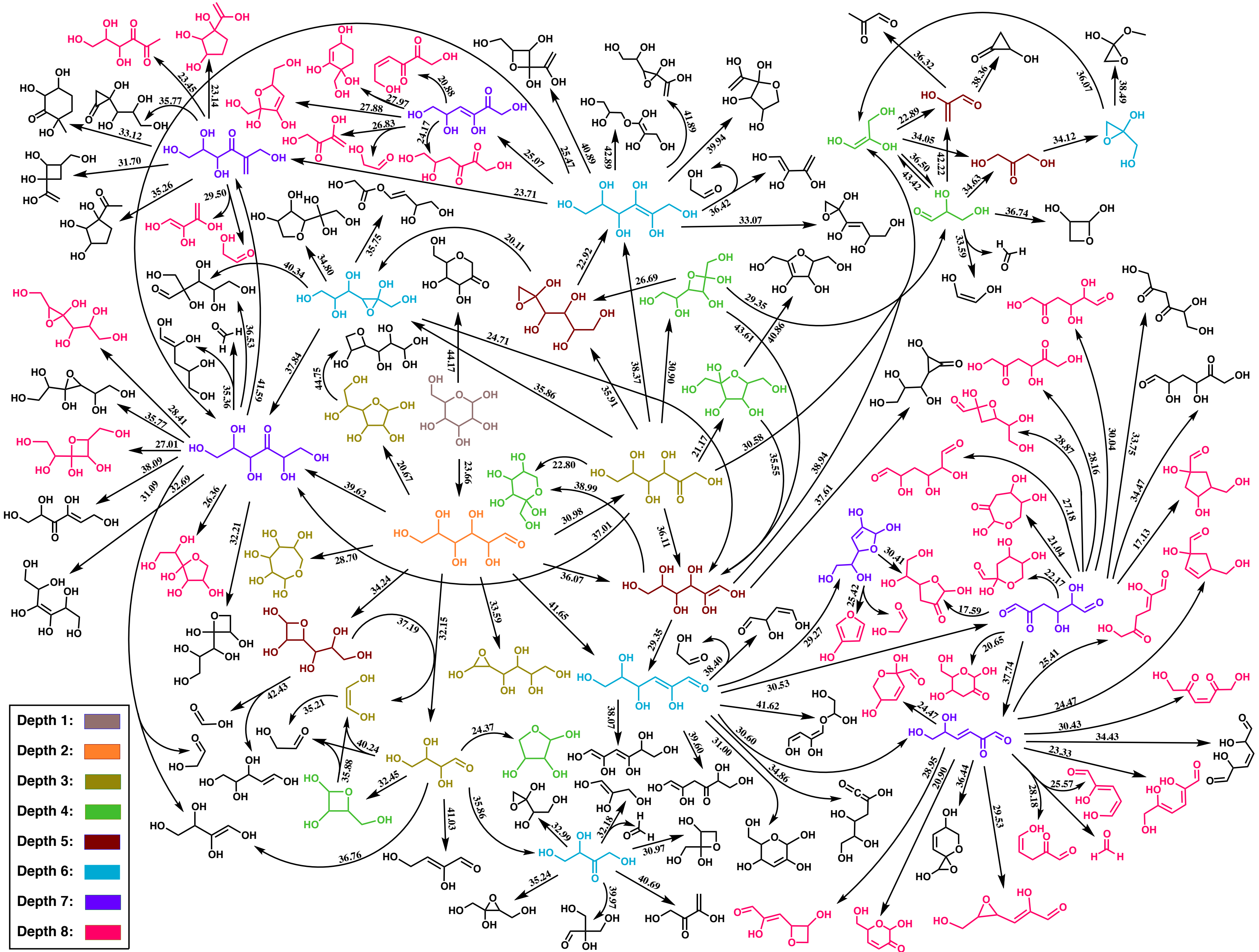
(1) all b2f2 reactions are explored for active nodes.

(2) Active nodes are determined by the minimum barrier to a given product (with a window)

(3) Water catalyzed reactions are considered for all H-transfers

β-D-Glucose Pyrolysis Network Exploration

Zhao, Q.; Savoie, B. M. Proc. Nat. Acad. Sci. 120, 2023



To perform a deep network exploration, we've implemented a modified version of Dijkstra's algorithm

At each iteration:

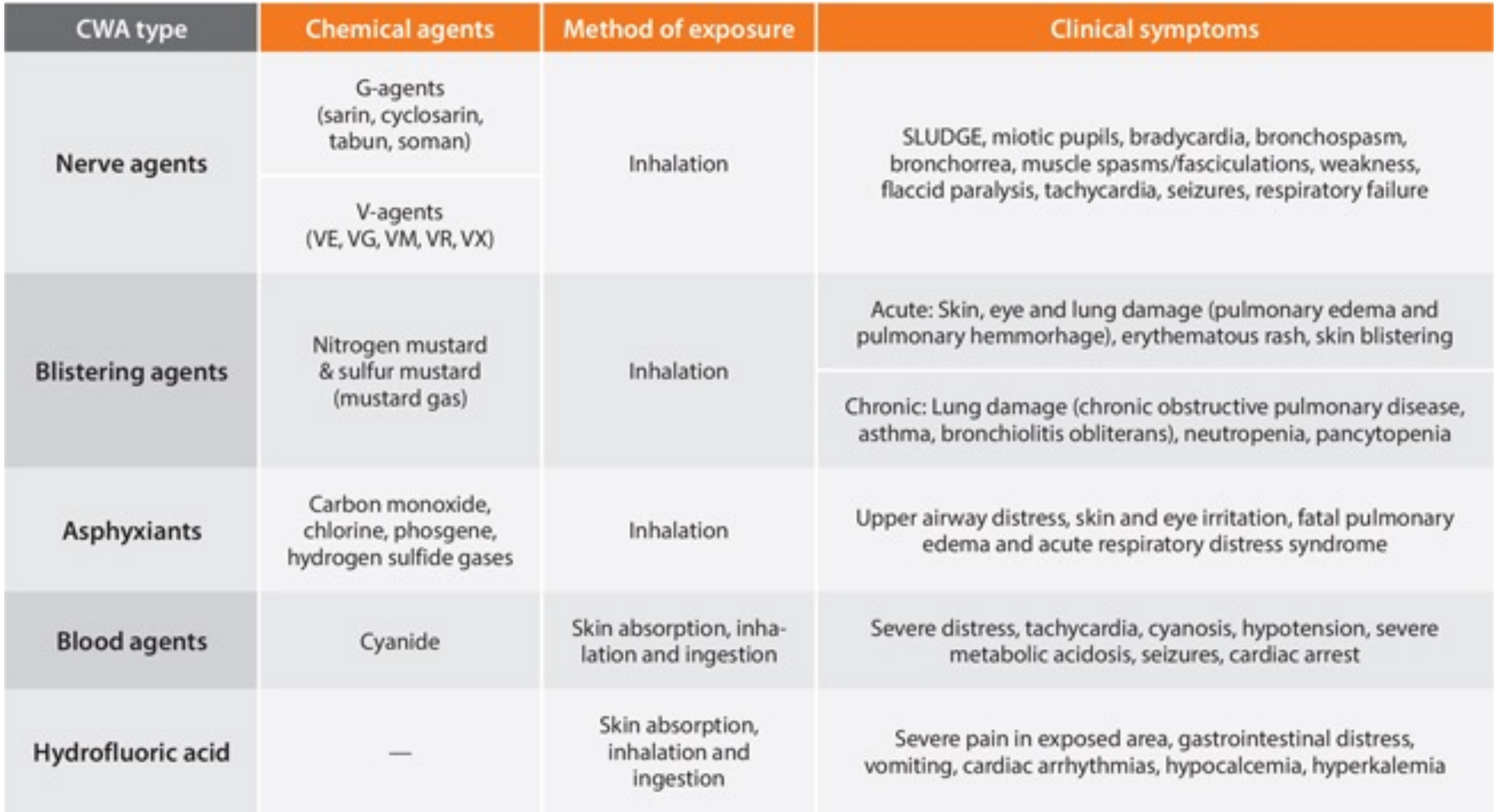
(1) all b2f2 reactions are explored for active nodes.

(2) Active nodes are determined by the minimum barrier to a given product (with a window)

(3) Water catalyzed reactions are considered for all H-transfers

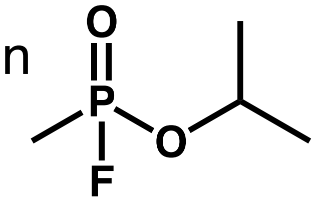
Studying Chemical Warfare Agents (CWAs) with YARP

Degradation products are often the only evidence of CWA use or existence. Establishing mechanistic pathways provides evidentiary value to investigators.

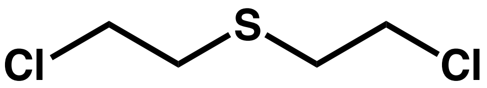


CWA type	Chemical agents	Method of exposure	Clinical symptoms
Nerve agents	G-agents (sarin, cyclosarin, tabun, soman)	Inhalation	SLUDGE, miotic pupils, bradycardia, bronchospasm, bronchorrea, muscle spasms/fasciculations, weakness, flaccid paralysis, tachycardia, seizures, respiratory failure
	V-agents (VE, VG, VM, VR, VX)		
Blistering agents	Nitrogen mustard & sulfur mustard (mustard gas)	Inhalation	Acute: Skin, eye and lung damage (pulmonary edema and pulmonary hemmorhage), erythematous rash, skin blistering
			Chronic: Lung damage (chronic obstructive pulmonary disease, asthma, bronchiolitis obliterans), neutropenia, pancytopenia
Asphyxiants	Carbon monoxide, chlorine, phosgene, hydrogen sulfide gases	Inhalation	Upper airway distress, skin and eye irritation, fatal pulmonary edema and acute respiratory distress syndrome
Blood agents	Cyanide	Skin absorption, inhalation and ingestion	Severe distress, tachycardia, cyanosis, hypotension, severe metabolic acidosis, seizures, cardiac arrest
Hydrofluoric acid	—	Skin absorption, inhalation and ingestion	Severe pain in exposed area, gastrointestinal distress, vomiting, cardiac arrhythmias, hypocalcemia, hyperkalemia

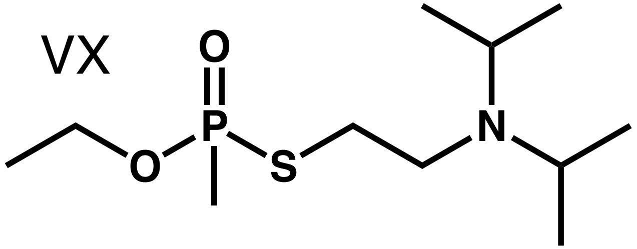
Sarin



Mustard Gas (HB)

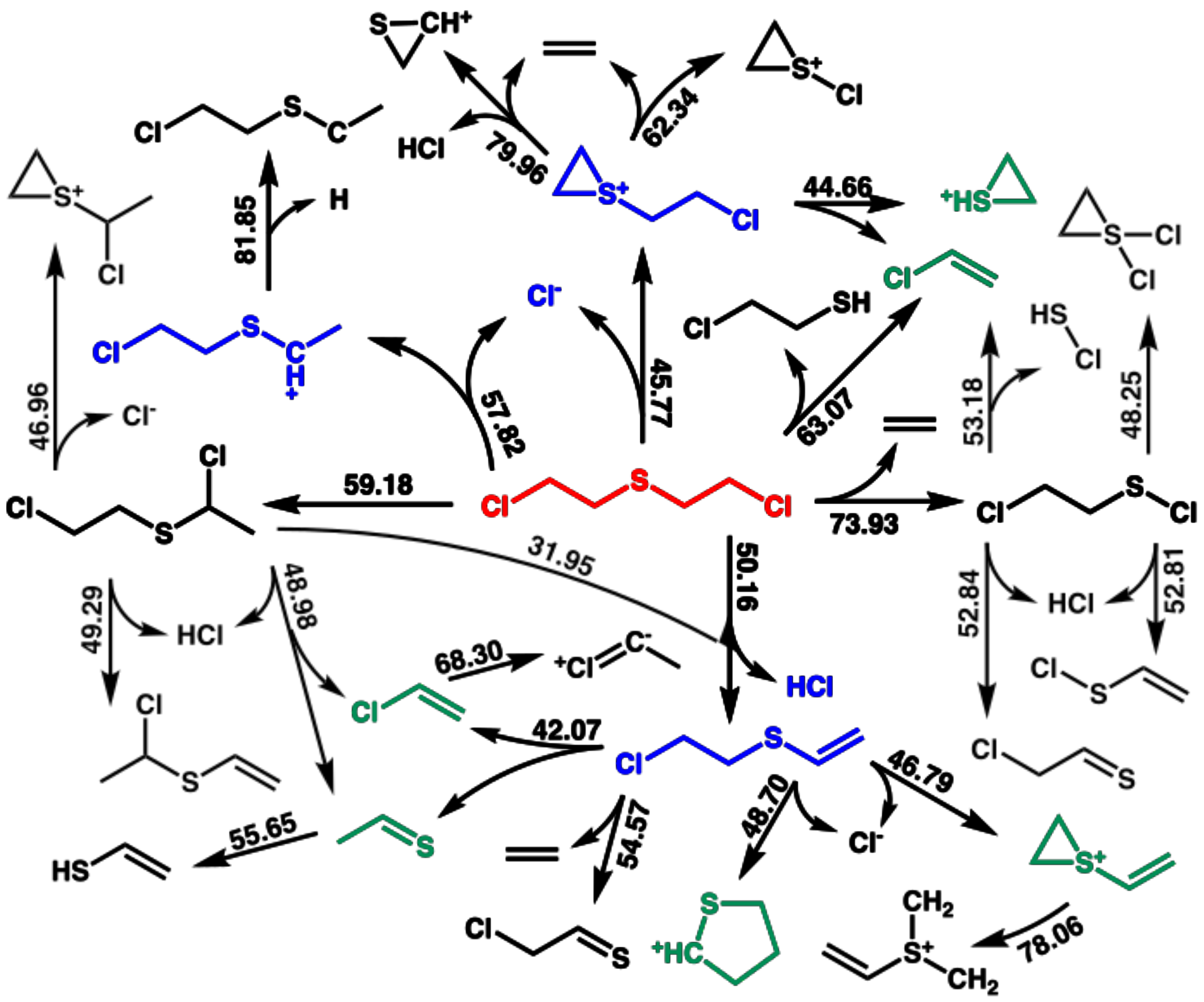


VX

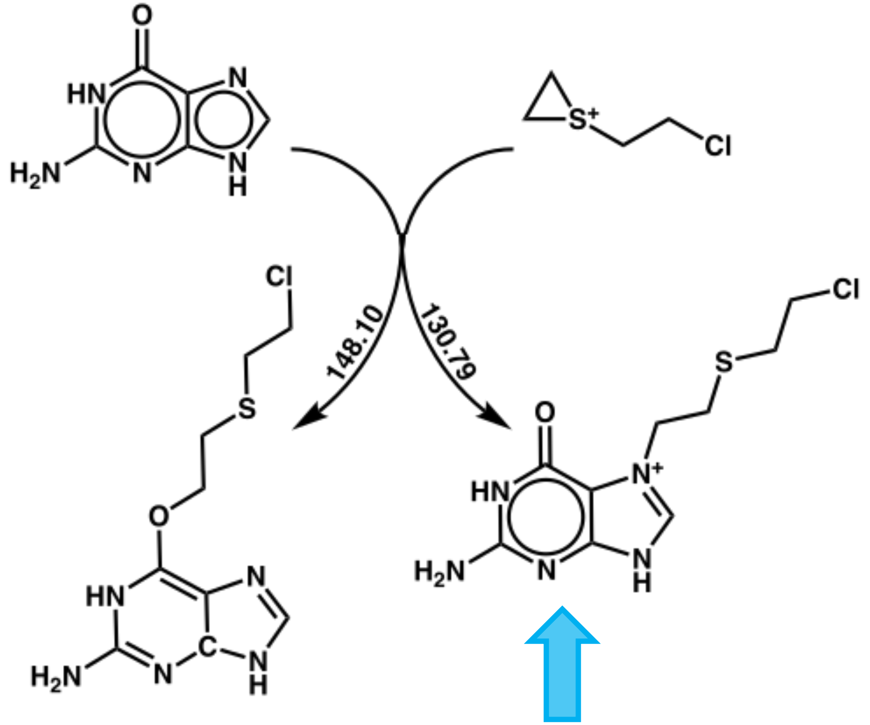


https://www.jems.com/major-incidents/a-review-of-chemical-warfare-agents-and-treatment-options/

YARP Prediction for Sulfur Mustard (HD) Reactivity



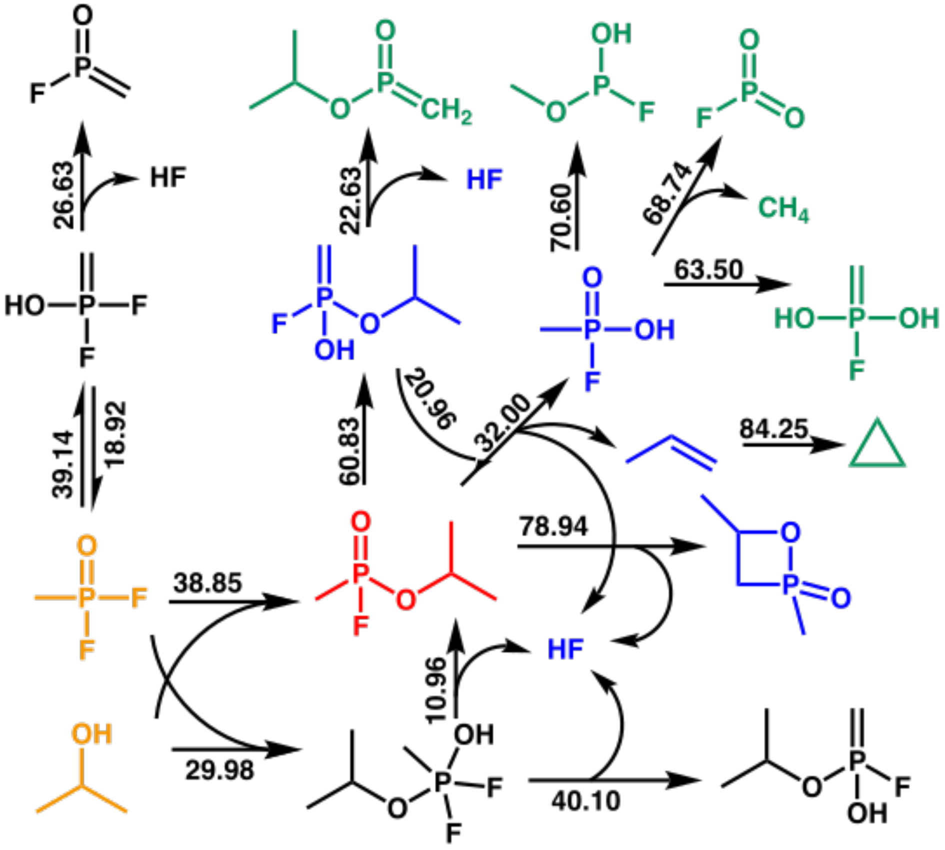
Mechanism of Action



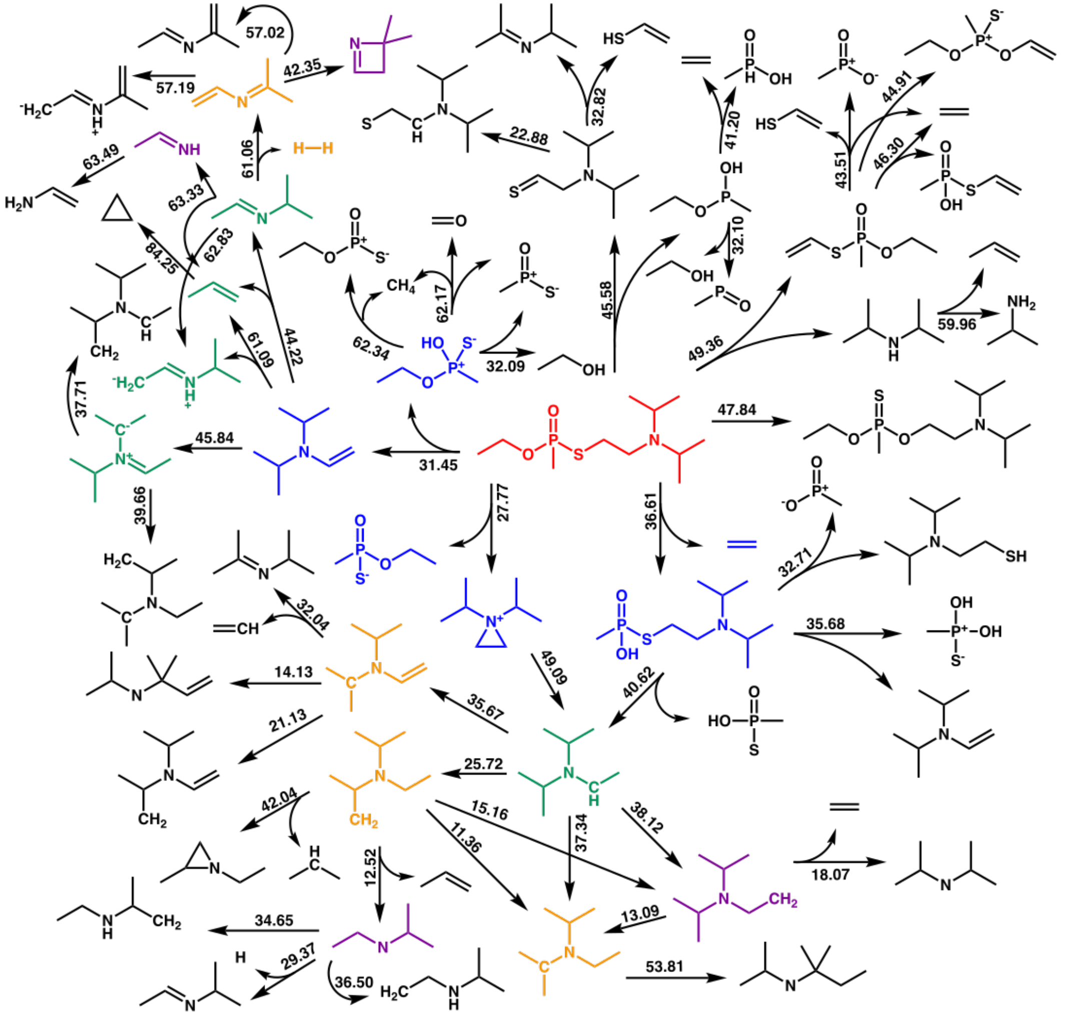
Lowest barrier bimolecular reaction

Predicted Reactivity for Organophosphorus Nerve Agents

Sarin (GB)



VX



Outlook and Acknowledgements

Students: Qiyuan Zhao, Tyler Pasut, Michael Woulfe

State-of-the-art:

- The accurate calculation of thermodynamic properties has become routine in many scenarios. Major opportunities lie in automation, systemization, and low-cost models.

- Practical solutions to the A→?→B, A→B+?, and A→? problems are now available. We envision black-box tools for non-experts in the near future that will assist in hypothesis generation and potentially reactivity screening.

- P2SAC and ONR for funding.
- Ray Mentzer (Purdue)
- Spencer Goldrich(PMP)

P2SAC
Purdue Process Safety & Assurance Center