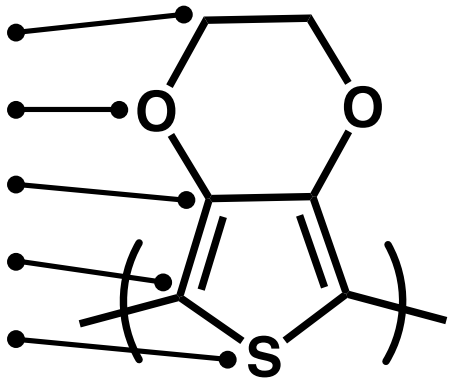
Quantum Chemical Prediction of Molecular Thermodynamics to Assess Reaction Safety and Scale-up

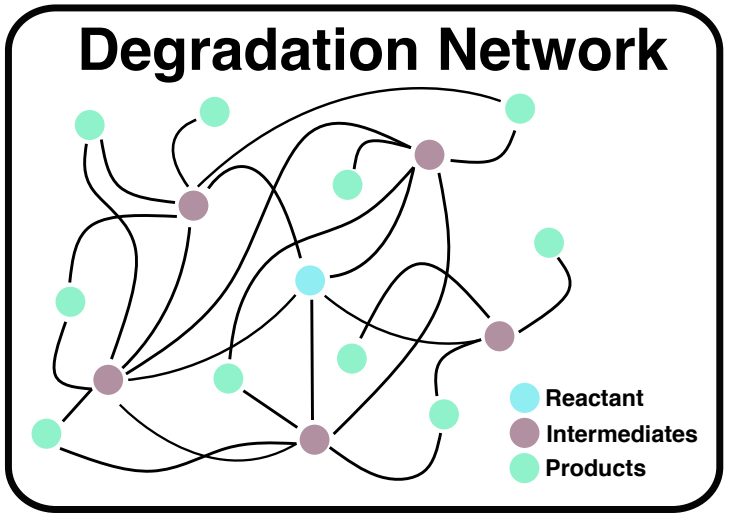
A new component theory:

[6[6[8][1][1]][8[6]][1][1]]
[8[6[6][1][1]][6[6][6]]]
[6[6[16][6]][6[8][6]][8[6]]]
[6[6[16][6]][6[8][6]][16[6]]]
[16[6[6][6]][6[6][6]]]



ΔH_f , S^o, C_v

Predicting Decomposition Pathways:



Qiyuan Zhao (4th year PhD student)

Brett M. Savoie
Charles Davidson Assistant Professor of Chemical Engineering
12/7/21, P2SAC Fall Meeting

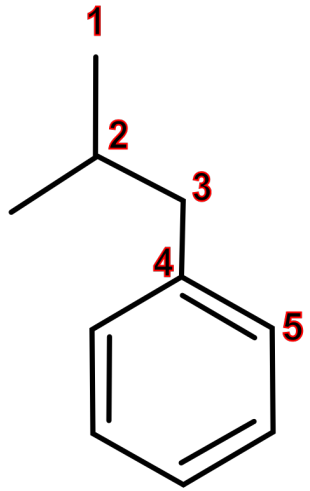
Challenges of Contemporary Group Theories

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^o, C_v) as the sum of "group" contributions.

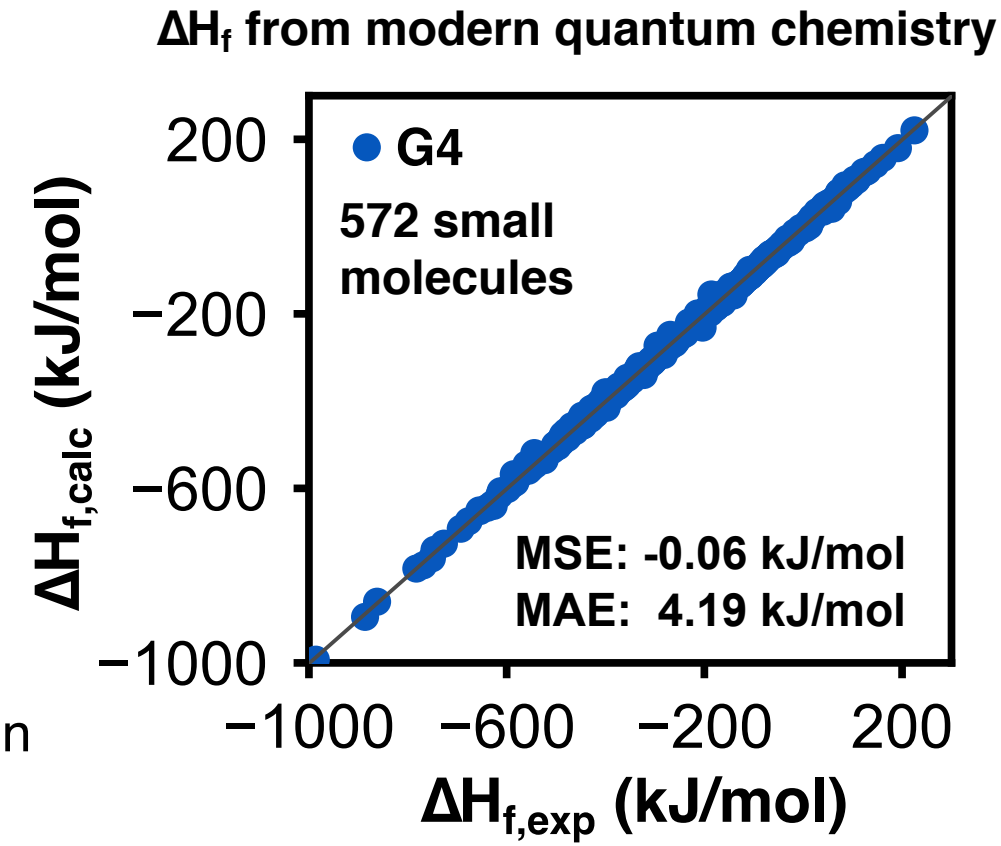
• Group contributions are calculated based on trusted experimental or computational data, and transferability is assumed.

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.



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

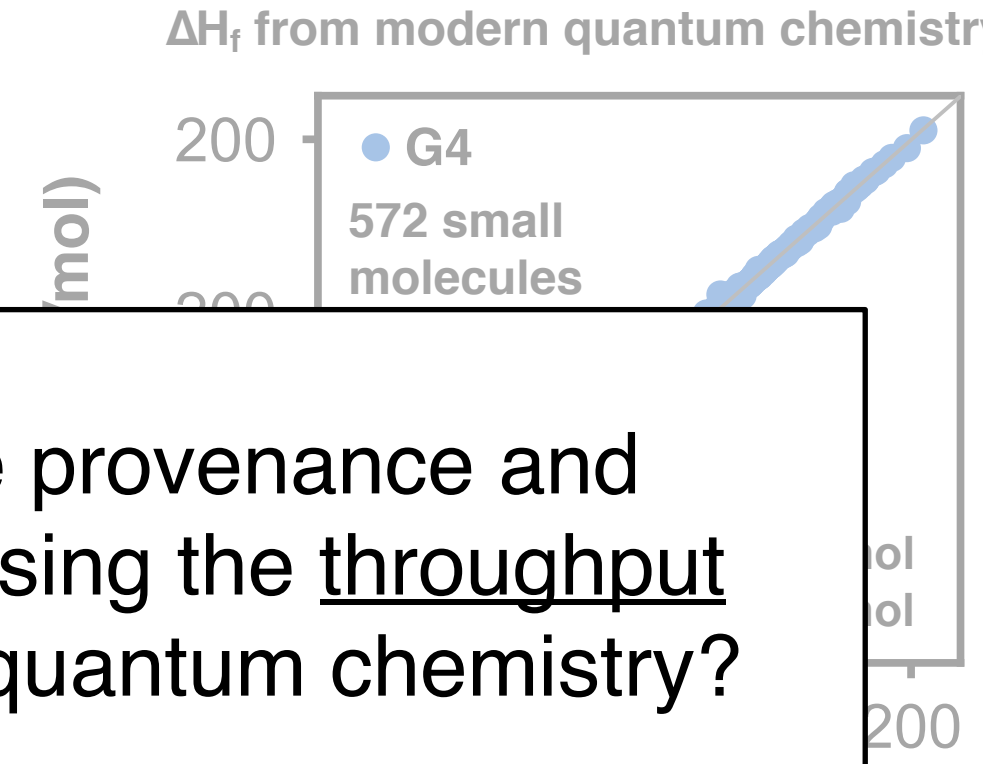
Benson Group Theory:

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

• Grou
on trus
data, a

Prob

• **Spec**
formal



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

• **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.

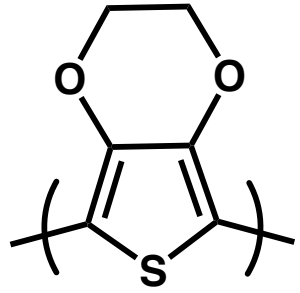
Zhao, Q.; Savoie, B. M.; Enthalpy of Formation Prediction via a fully Self-Consistent Component Increment Theory. *J. Chem. Info. Model.* **2020**, 60, 2199-2207

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

graph/structure equivalence

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

Adjacency matrix for PEDOT monomer

P2SAC Publications

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

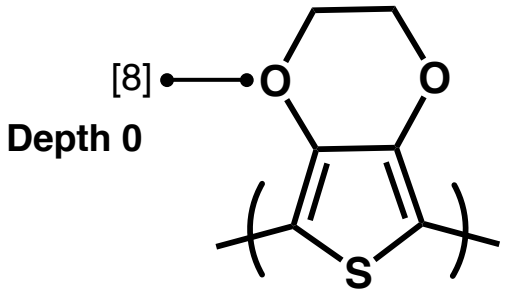
Seo, B.; Lin, Z.-Y.; Zhao, Q.; Webb, M. A.; Savoie, B. M. Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force Fields. *J. Chem. Inf. Model.* **2021**, 61 (10), 5013–5027. https://doi.org/10.1021/acs.jcim.1c00491.

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

graph/structure equivalence

	S	C	C	C	C	O	C	C	H	H	O	H	H
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

Adjacency matrix for PEDOT monomer

P2SAC Publications

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

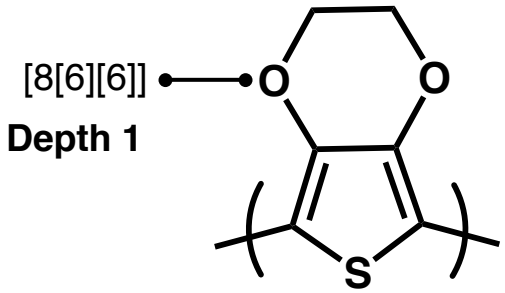
Seo, B.; Lin, Z.-Y.; Zhao, Q.; Webb, M. A.; Savoie, B. M. Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force Fields. *J. Chem. Inf. Model.* **2021**, 61 (10), 5013–5027. https://doi.org/10.1021/acs.jcim.1c00491.

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

graph/structure equivalence

S	0	1	0	0	1	0	0	0	0	0	0	0
C	1	0	1	0	0	0	0	0	0	0	0	0
C	0	1	0	1	0	1	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

Adjacency matrix for PEDOT monomer

P2SAC Publications

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.

TAFFI Component Increment Theory (TCIT)

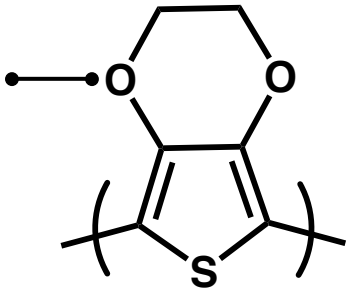
The fundamental idea

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

TCIT is a component theory (2-bond specific)

[8[6[6][1][1]][6[6][6]]]

Depth 2



graph/structure equivalence

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

S	0	1	0	0	1	0	0	0	0	0	0	0
C	1	0	1	0	0	0	0	0	0	0	0	0
C	0	1	0	1	0	1	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
O	0	0	1	0	0	0	1	0	0	0	0	0
C	0	0	0	0	0	1	0	1	1	1	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

Adjacency matrix for PEDOT monomer

P2SAC Publications

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

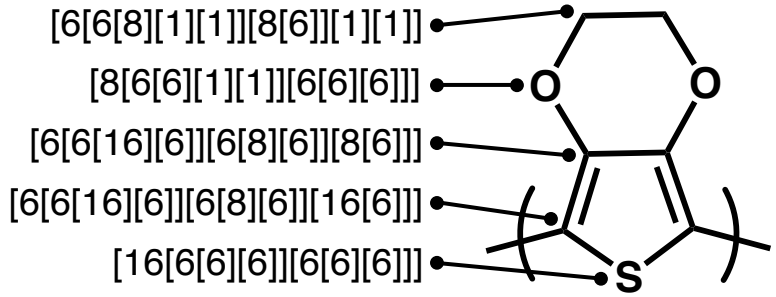
Seo, B.; Lin, Z.-Y.; Zhao, Q.; Webb, M. A.; Savoie, B. M. Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force Fields. *J. Chem. Inf. Model.* **2021**, 61 (10), 5013–5027. https://doi.org/10.1021/acs.jcim.1c00491.

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)



graph/structure equivalence

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

TAFFI

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

Adjacency matrix for PEDOT monomer

P2SAC Publications

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

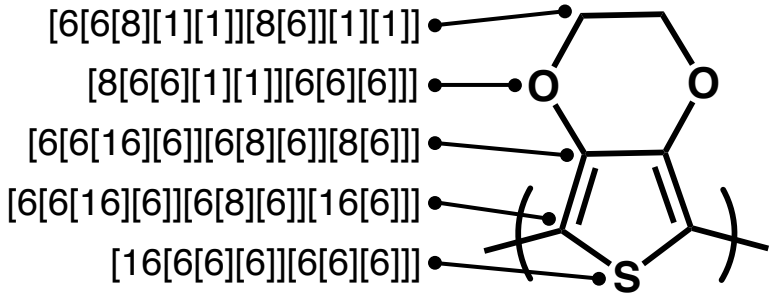
Seo, B.; Lin, Z.-Y.; Zhao, Q.; Webb, M. A.; Savoie, B. M. Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force Fields. *J. Chem. Inf. Model.* **2021**, 61 (10), 5013–5027. https://doi.org/10.1021/acs.jcim.1c00491.

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**.

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



graph/structure equivalence

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

TAFFI

	S	C	C	C	C	O	C	C	H	H	O	H	H
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

Adjacency matrix for PEDOT monomer

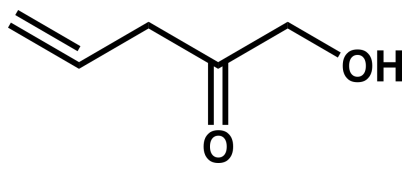
P2SAC Publications

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.

Graphical Decomposition of Model Compounds



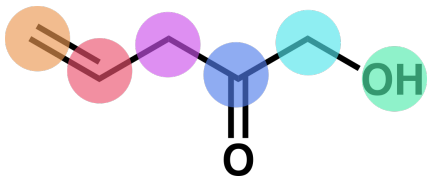
1-hydroxy-pent-2-ene-2-one

How will we select molecules for parameterizing TCIT components?

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

Graphical Decomposition of Model Compounds

identify components*



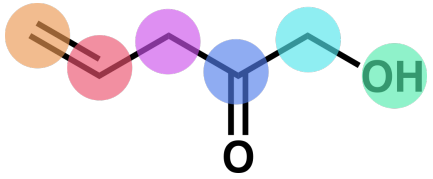
1-hydroxy-pent-2-ene-2-one

How will we select molecules for parameterizing TCIT components?

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

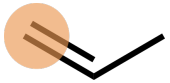
Graphical Decomposition of Model Compounds

identify components*



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

Recursively generate model compounds

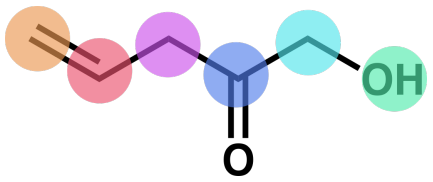


How will we select molecules for parameterizing TCIT components?

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

Graphical Decomposition of Model Compounds

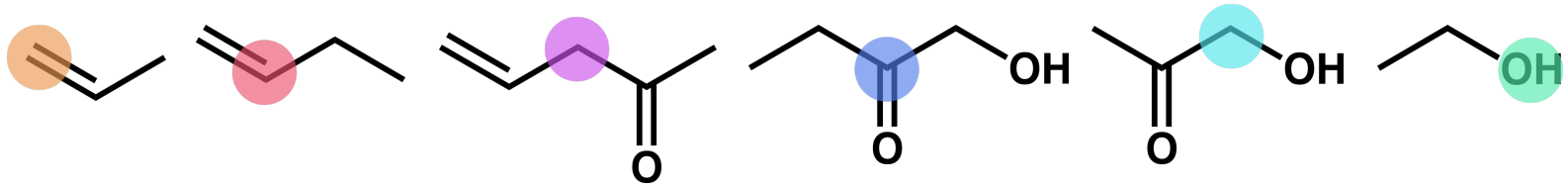
identify components*



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

How will we select molecules for parameterizing TCIT components?

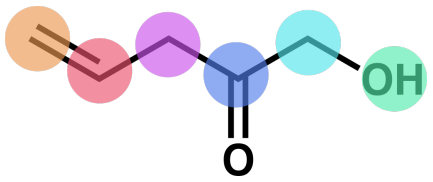
Recursively generate model compounds



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

Graphical Decomposition of Model Compounds

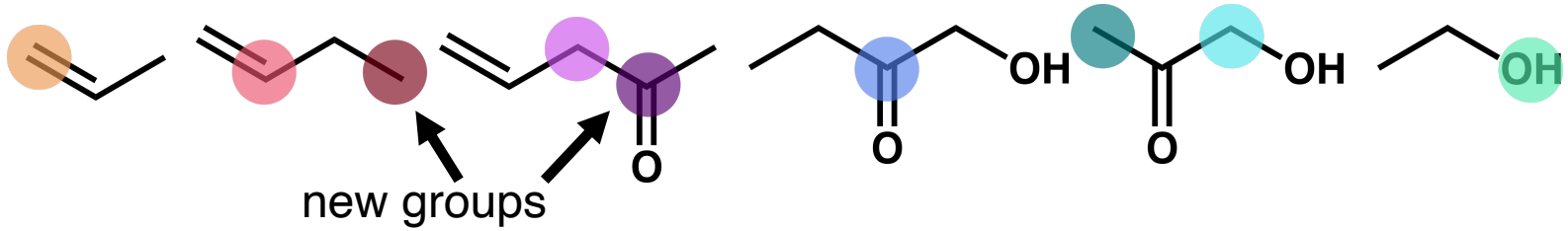
identify components*



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

How will we select molecules for parameterizing TCIT components?

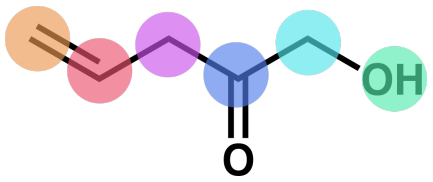
Recursively generate model compounds



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

Graphical Decomposition of Model Compounds

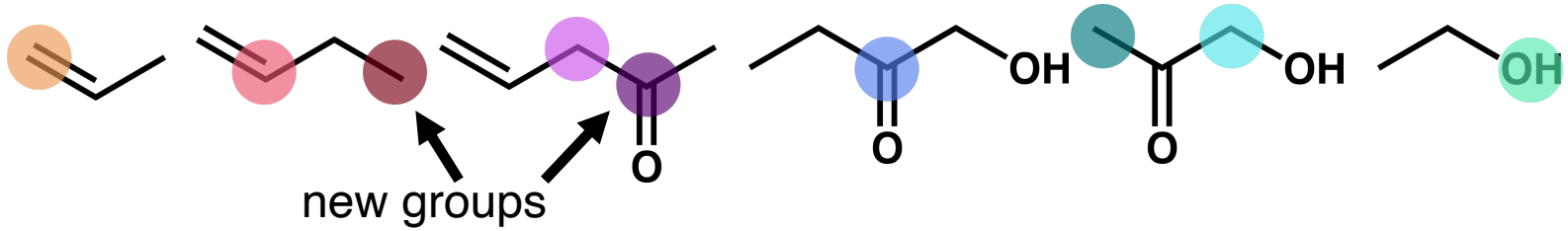
identify components*



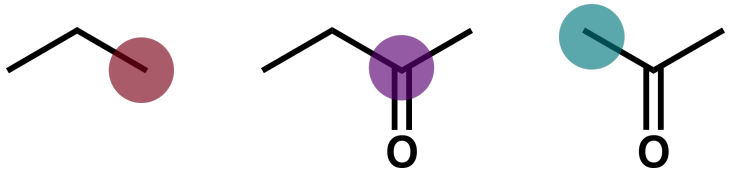
1-hydroxy-pent-2-ene-2-one

How will we select molecules for parameterizing TCIT components?

Recursively generate model compounds



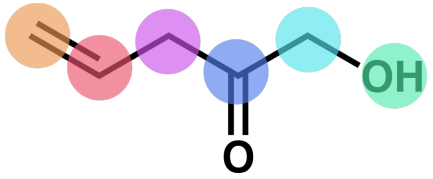
Resolve dependencies



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

Graphical Decomposition of Model Compounds

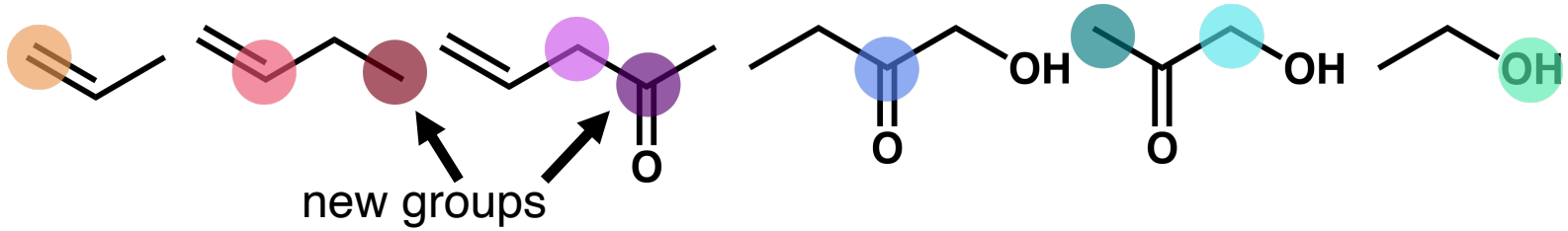
identify components*



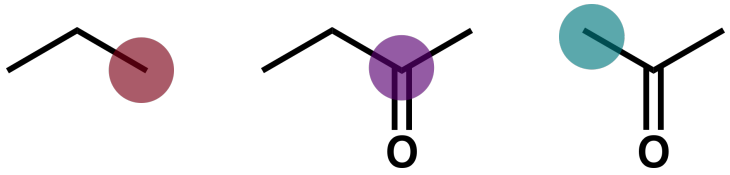
1-hydroxy-pent-2-ene-2-one

How will we select molecules for parameterizing TCIT components?

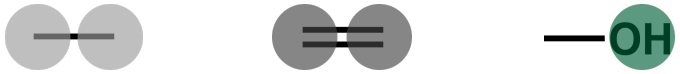
Recursively generate model compounds



Resolve dependencies



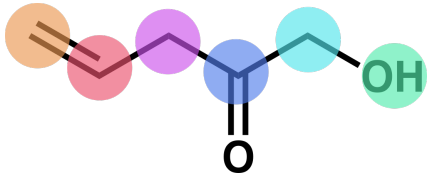
Resolve rank deficiency with elementary constraints



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

Graphical Decomposition of Model Compounds

Prediction target:



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

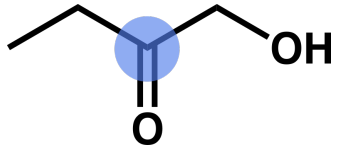
$\Delta H_{f,G4}$ = -259.9 kJ/mol
$\Delta H_{f,TCIT}$ = -259.3 kJ/mol

no experimental data

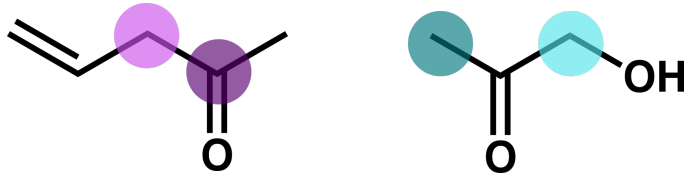
Topologically sort dependency graph

(Automatically handled by TCIT software)

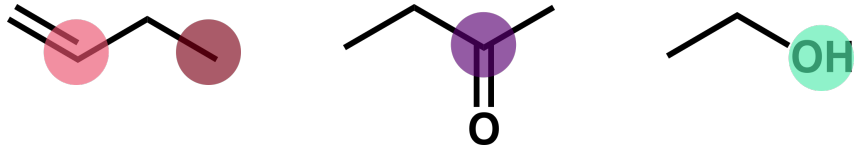
Gen 4:



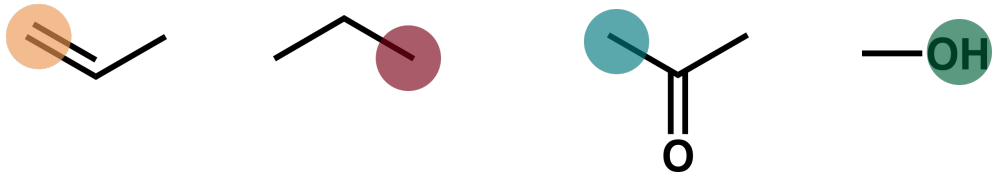
Gen 3:



Gen 2:



Gen 1:



Gen 0:

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

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.

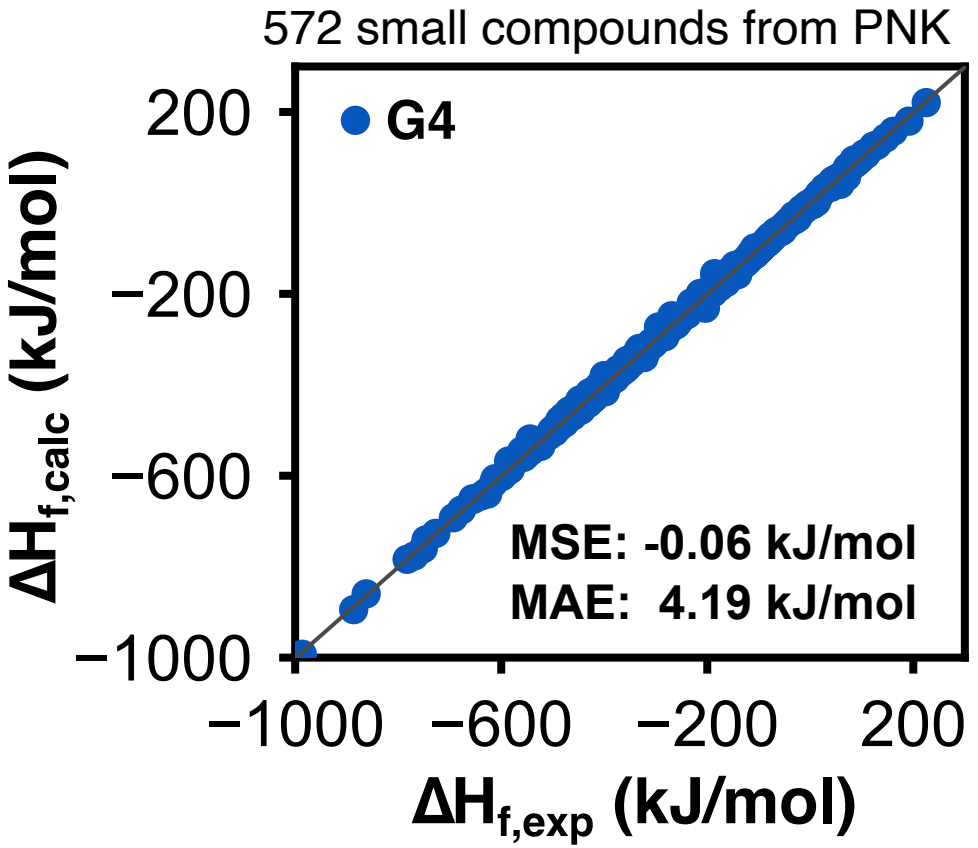
Benchmarking $\Delta H_{f,gas}$ Predictions Against the PNK Dataset

- 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.



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

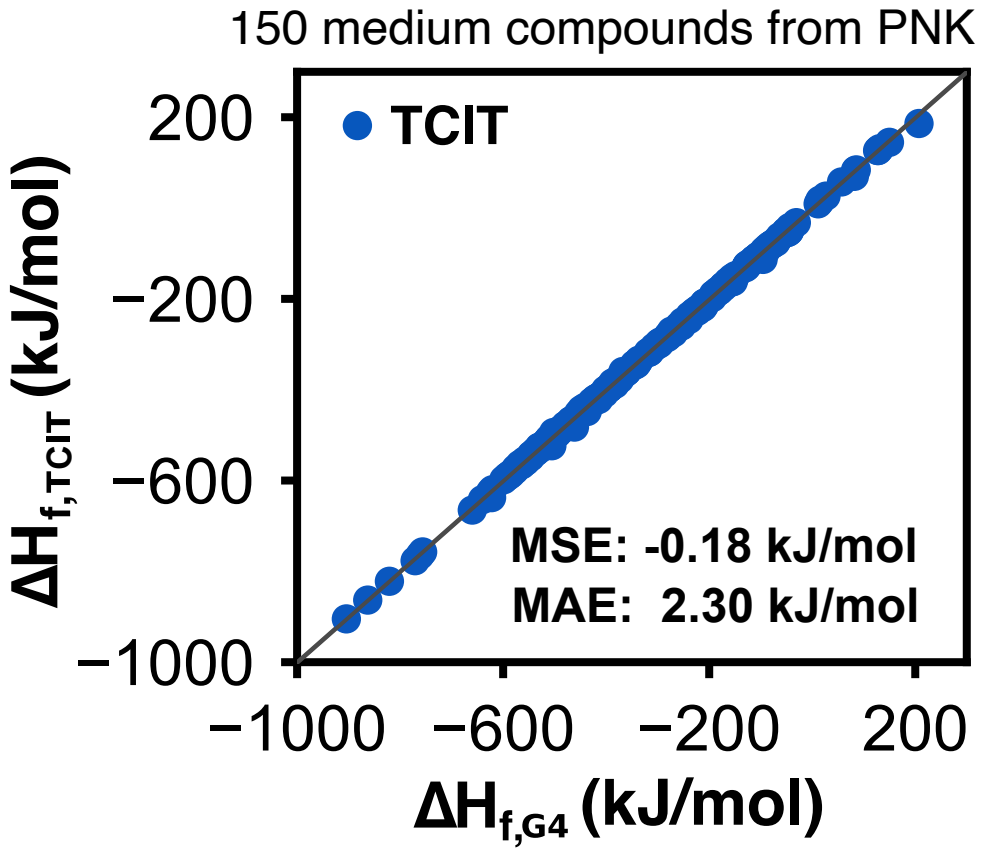
- Initial benchmarking set consists of ~1100 **linear** C,H, and O containing compounds from PNK[1]

(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.



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]

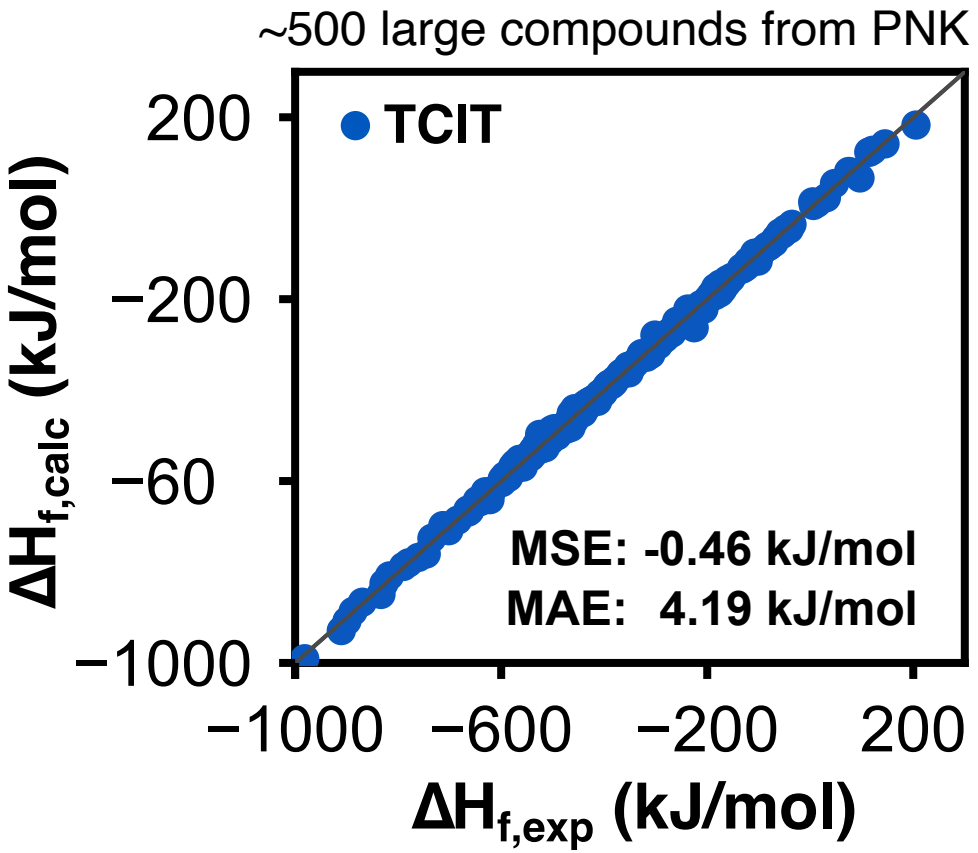
(1) J. B. Pedley, R. D. Naylor, S. P. Kirby "Thermochemical Data of Organic Compounds" 2nd ed. 1986

- 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.



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]

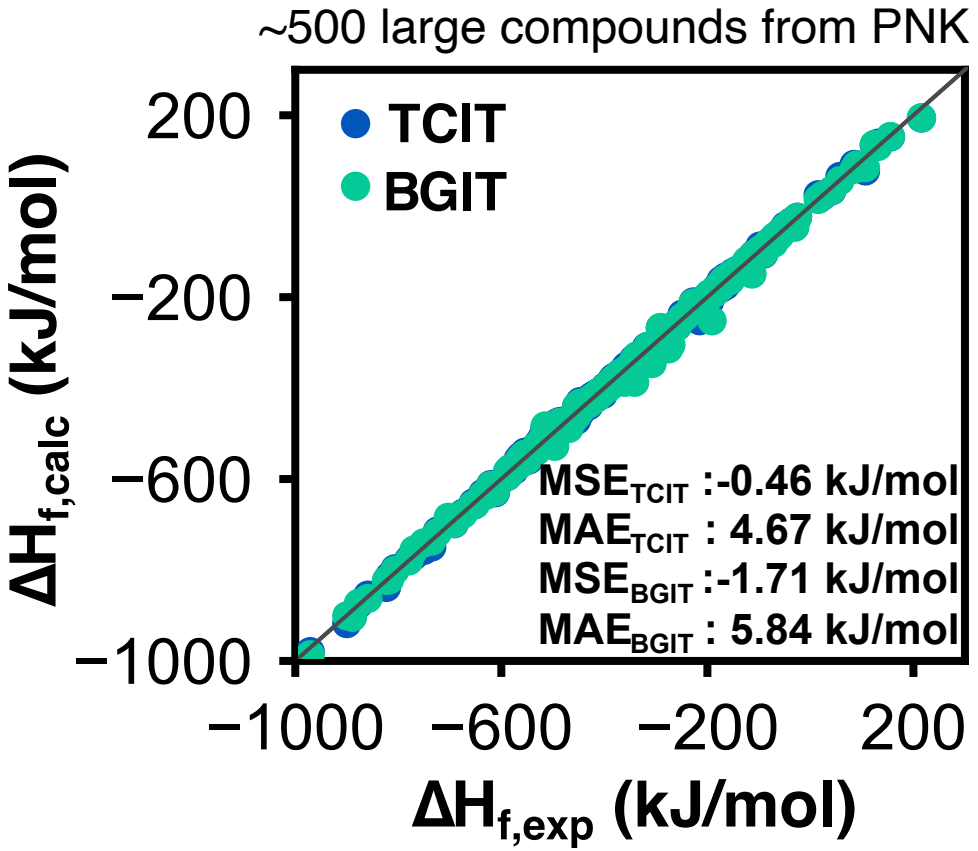
(1) J. B. Pedley, R. D. Naylor, S. P. Kirby "Thermochemical Data of Organic Compounds" 2nd ed. 1986

- 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.



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

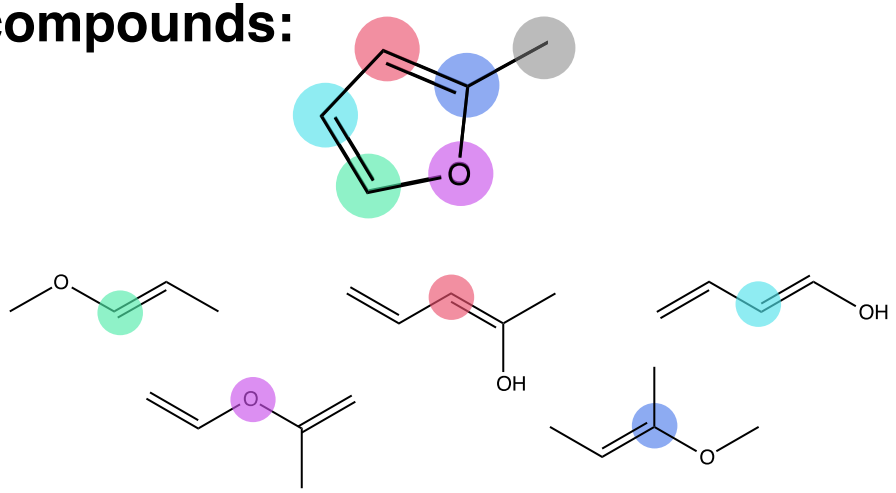
TCIT shows comparable performance to BGIT/CHETAH but is derived exclusively from extensible G4 data.

Extension to Ring-Containing Molecules

- **Ring-containing molecules have additional strain and/or conjugation corrections that exacerbate the extensibility issues of Benson Theory.**

- **In TCIT we are addressing this through chemically specific ring corrections that account for differences in substitution pattern and topology:**

1. Decompose ring into acyclic model compounds:

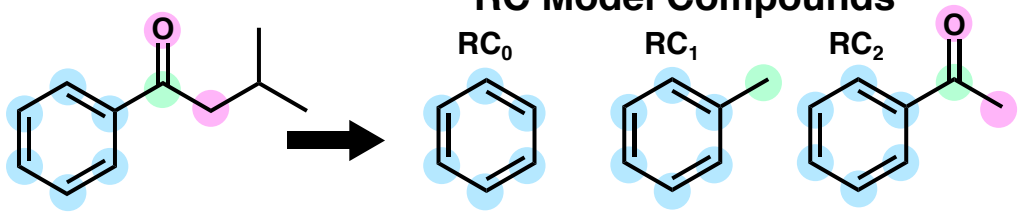


2. Add ring correction (RC) to final prediction:

$$\mathrm{RC} = H_f(ring) - H_f(\bigcirc)\text{-}\ H_f(\bigcirc)\text{-}\ H_f(\bigcirc)\text{-}\ H_f(\bigcirc)\text{-}\ H_f(\bigcirc)\text{-}H_f(\bigcirc)$$

Technical Developments

RC Model Compounds



: Depth 0 : Depth 1 : Depth 2

Method 1: Use RC_1 based model parameterized to G4 data.

Method 2: Use graph-NN to predict RC_0-RC_2

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

Benchmarking Ring-Correction Performance

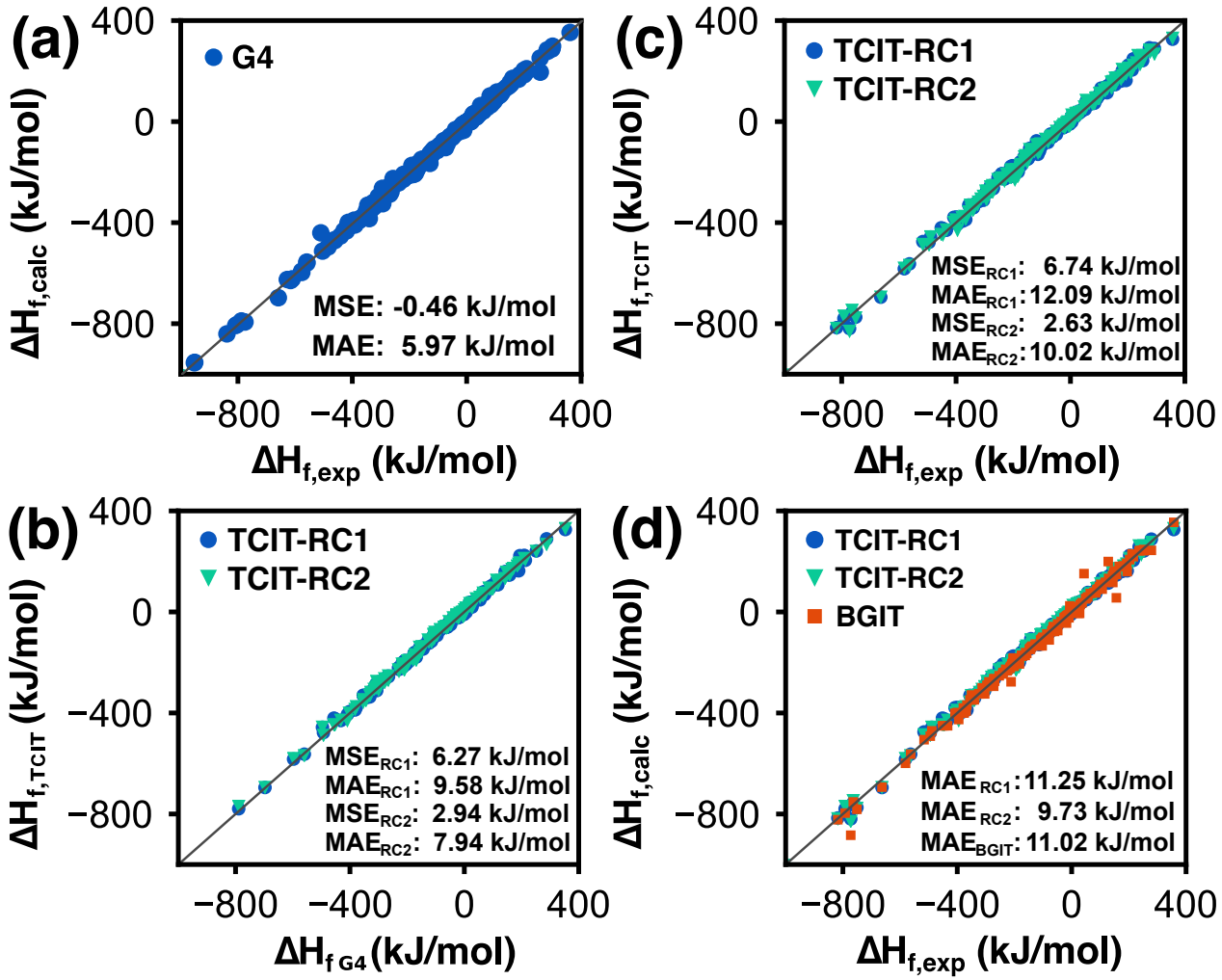
(a) G4 errors are marginally larger for ring-containing compounds but still very accurate

(b) The neural-network based ring-correction exhibits excellent reproduction of the G4 predictions (MSE: ~3kJ/mol; MAE: ~8 kJ/mol).

(c) TCIT is completely transferable to new testing compounds that are experimentally characterized. Errors are consistent with G4 comparison

(d) The TCIT-R2 model outperforms BGIT on the large molecule benchmark while being extensible. Significantly, these compounds are within BGIT's training data.

~120 ring-containing compounds from PNK (excluding training)



BGIT cannot make predictions for ~2% of PNK compounds

TCIT Extension to Other Properties and Phases

Condensed Phases: The condensed-phase and gas-phase standard enthalpies of formation differ by the heats of sublimation and vaporization[1]:

$$\Delta_{\mathrm{f}} H^{\circ}_{(\mathrm{s})} = \Delta_{\mathrm{f}} H^{\circ}_{(\mathrm{g})} - \Delta_{\mathrm{sub}} H^{\circ}$$
$$\Delta_{\mathrm{f}} H^{\circ}_{(\ell)} = \Delta_{\mathrm{f}} H^{\circ}_{(\mathrm{g})} - \Delta_{\mathrm{vap}} H^{\circ}$$

We have implemented group contribution models for heat of vaporization[2] and sublimation[3], respectively. The group assignments and group values associated with these models have been automated within the context of TCIT.

Standard Molar Entropy (S^{o}) and heat capacity (C_v): The molar entropies and constant volume heat capacities are accessible from quantum chemistry using the harmonic oscillator approximation for the molecular partition function and corrections based on the number of rotatable bonds (N_{rot}) and molecular symmetry:

$$S^{\circ} = \langle S^{\circ}_{\mathrm{harm}} \rangle + R N_{\mathrm{rot}} + R \log \sigma \qquad C_{\mathrm{v}} = \langle C_{\mathrm{v,harm}} \rangle + \alpha N_{\mathrm{rot}} + \beta$$

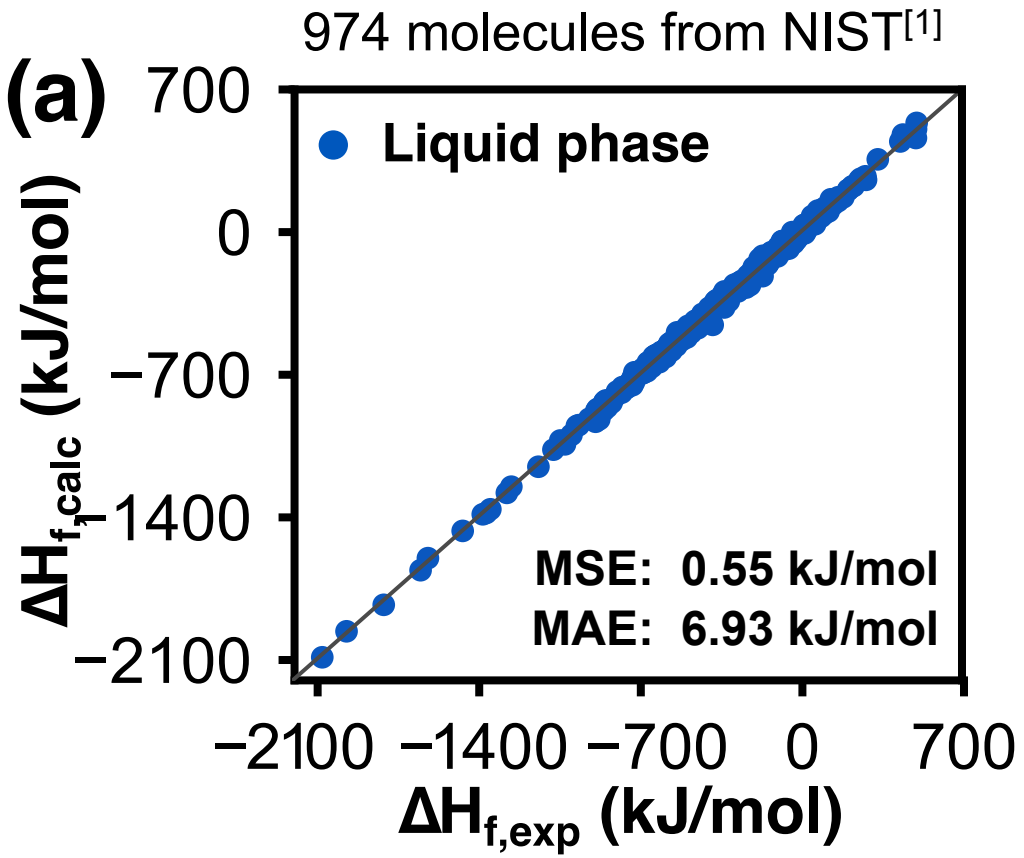
[1] Murray, J.S., Brinck, T. and Politzer, P., **1996**. *Chemical physics*, 204, 289-299.

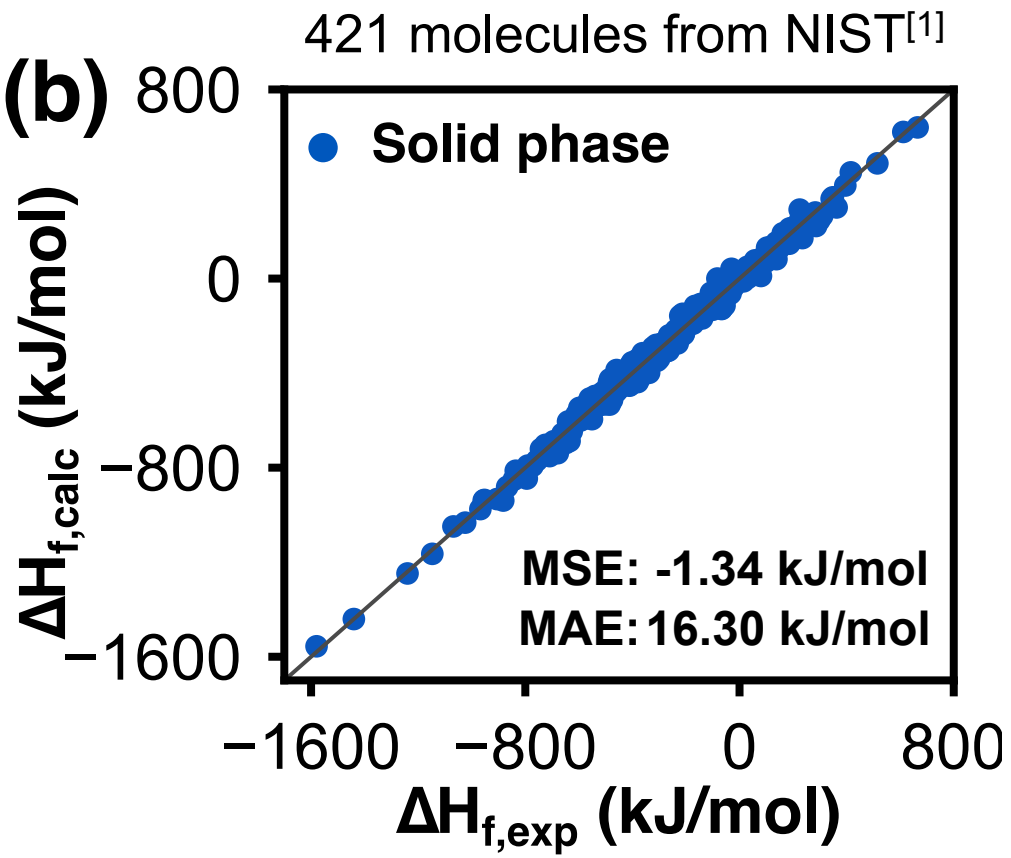
[2] Pankow, J.F. and Asher, W.E., **2008**. *Atmospheric Chemistry and Physics.*

[3] Bagheri, M.; Gandomi, A. H.; Golbraikh, A. **2012**,*Thermochim. Acta*, 543, 96–106

<.> Indicates conformational averaging
R: ideal gas constant
σ: symmetry number
α, β: regressed constants

Benchmarking Condensed Phase ΔH_f Predictions





- Testing set includes both linear and cyclic compounds with number of heavy atoms varying from 1 to 30.

- Low MSE indicates no systematic bias, larger absolute errors result from the quality of the ΔH_{vap} and ΔH_{sub} models.

[1] Linstrom, P.J. and Mallard, W.G., 2001. *Journal of Chemical & Engineering Data*, 46(5), pp.1059-1063.

Benchmarking TCIT S° and C_v Predictions

(a) G4/TCIT S° comparison for 314 medium sized molecules.

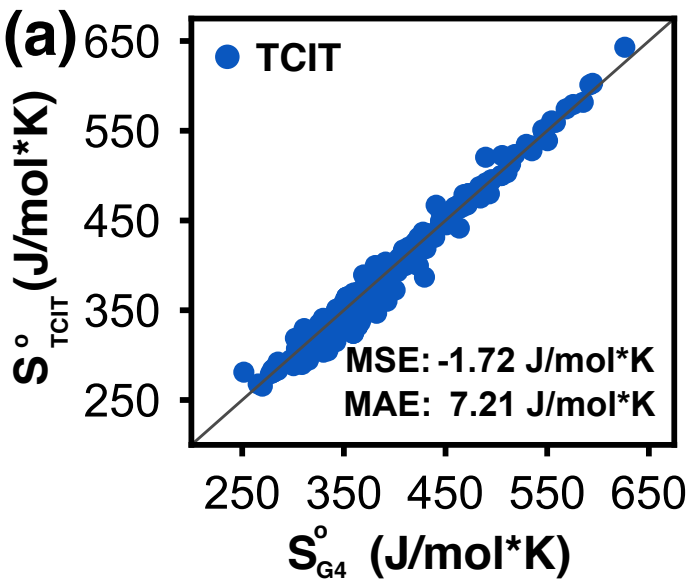
(b) G4/TCIT S° comparison for 314 medium sized molecules.

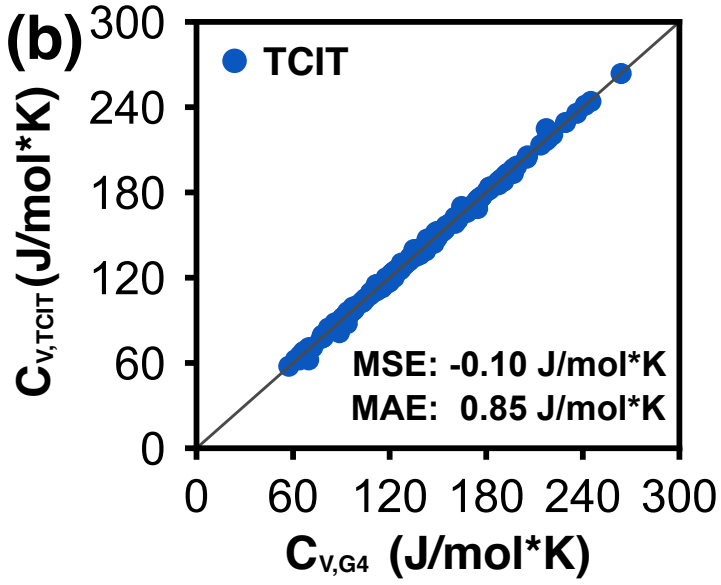
(c) TCIT S° comparison for 439 large molecules from NIST[1]

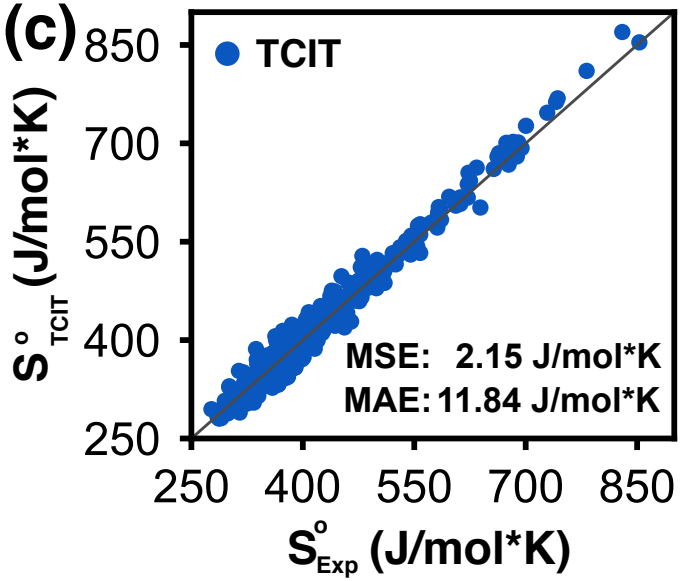
(d) TCIT heat capacity comparison for 904 large molecules from NIST[1]

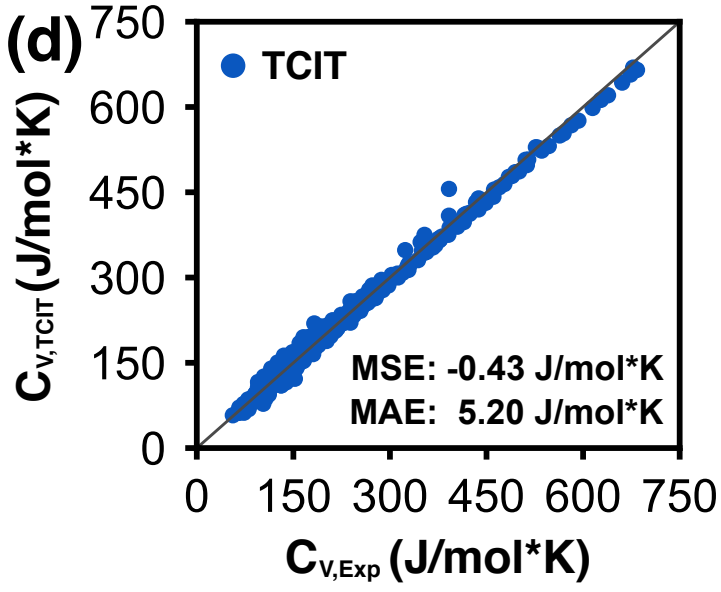
• The TCIT errors are consistent with error propagation of G4:exp and TCIT:G4 errors.

TCIT now supports S° and C_v predictions with accuracies comparable to G4 model chemistry.









[1] Linstrom, P.J. and Mallard, W.G., 2001. *Journal of Chemical & Engineering Data*, 46(5), pp.1059-1063.

Benchmarking TCIT S° and C_v Predictions

(a) G4/TCIT S° comparison for 314 medium sized molecules.

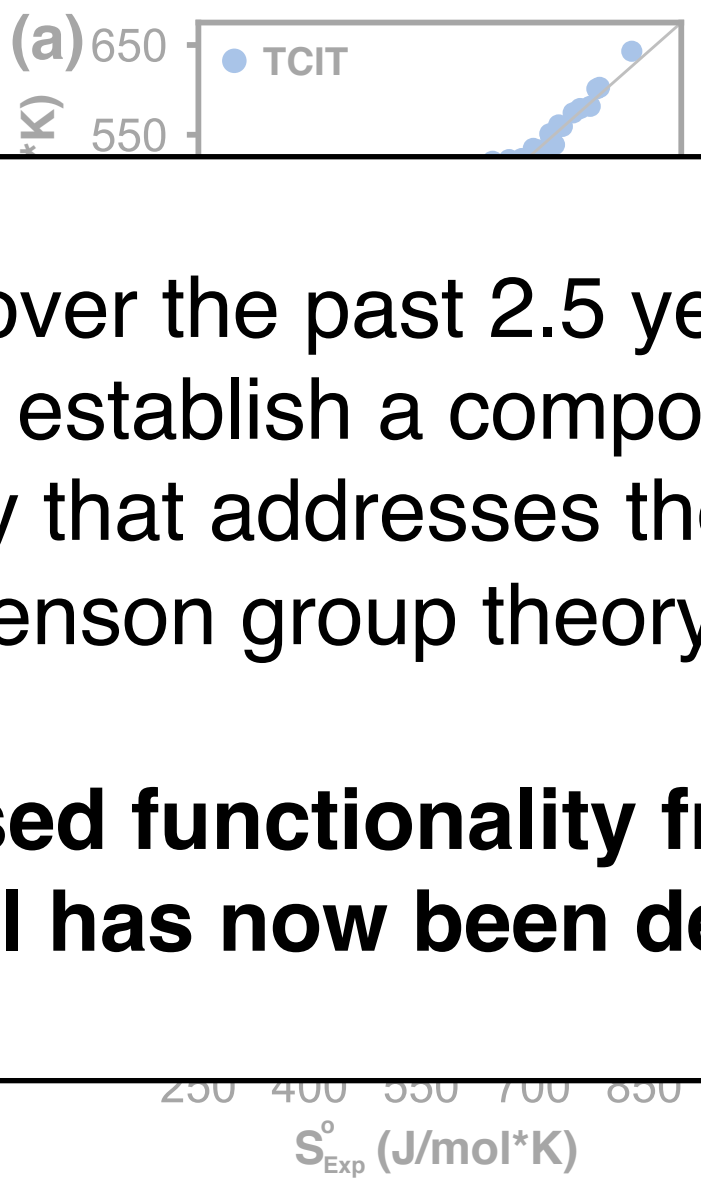
(b) G4/
medium

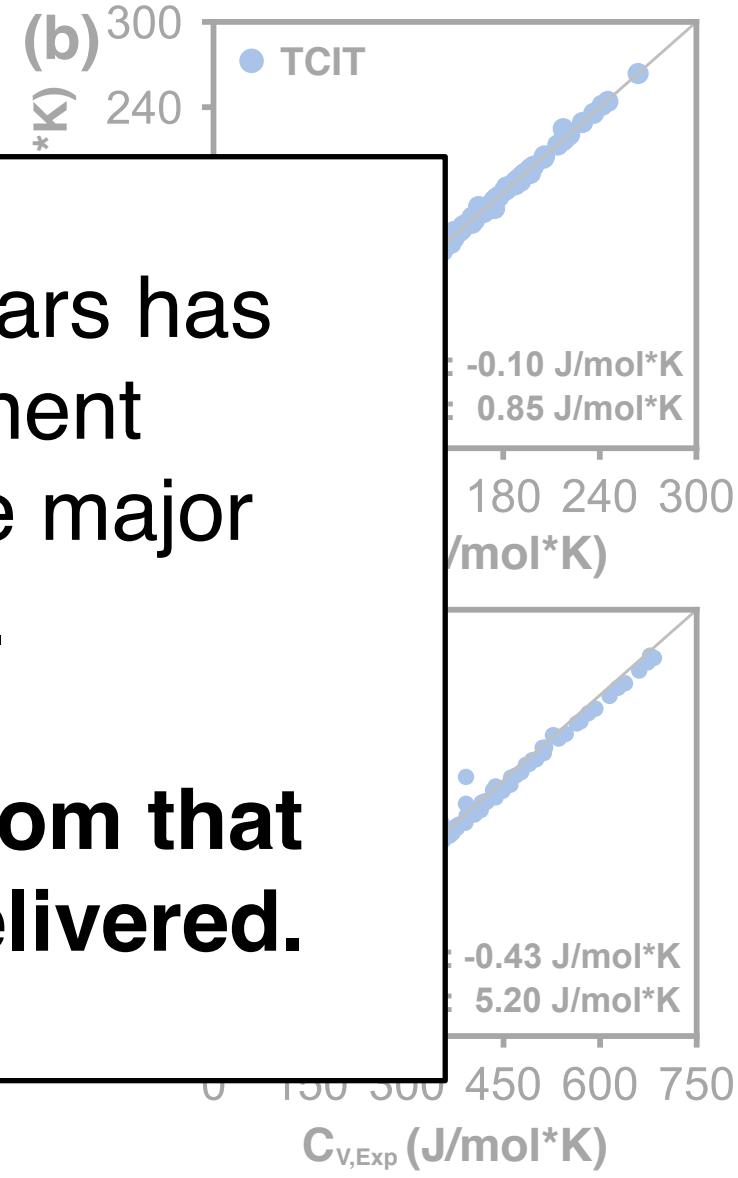
(c) TCI
molecul

(d) TCI
904 larg

• The T
error p
TCIT:G4

TCI
predictions with accuracies
comparable to G4 model chemistry.





P2SAC funding over the past 2.5 years has allowed us to establish a component increment theory that addresses the major gaps in Benson group theory.

All of the promised functionality from that original proposal has now been delivered.

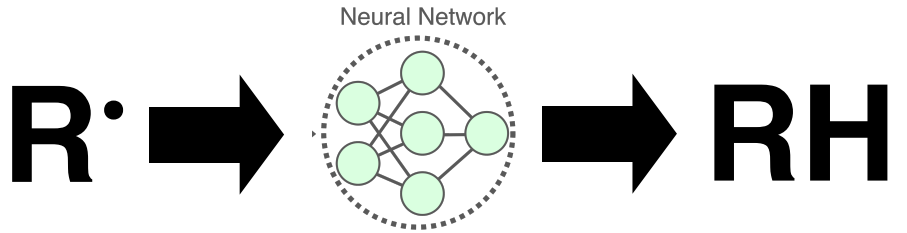
[1] Linstrom, P.J. and Mallard, W.G., 2001. *Journal of Chemical & Engineering Data*, 46(5), pp.1059-1063.

Extending TCIT to Radicals and Ions

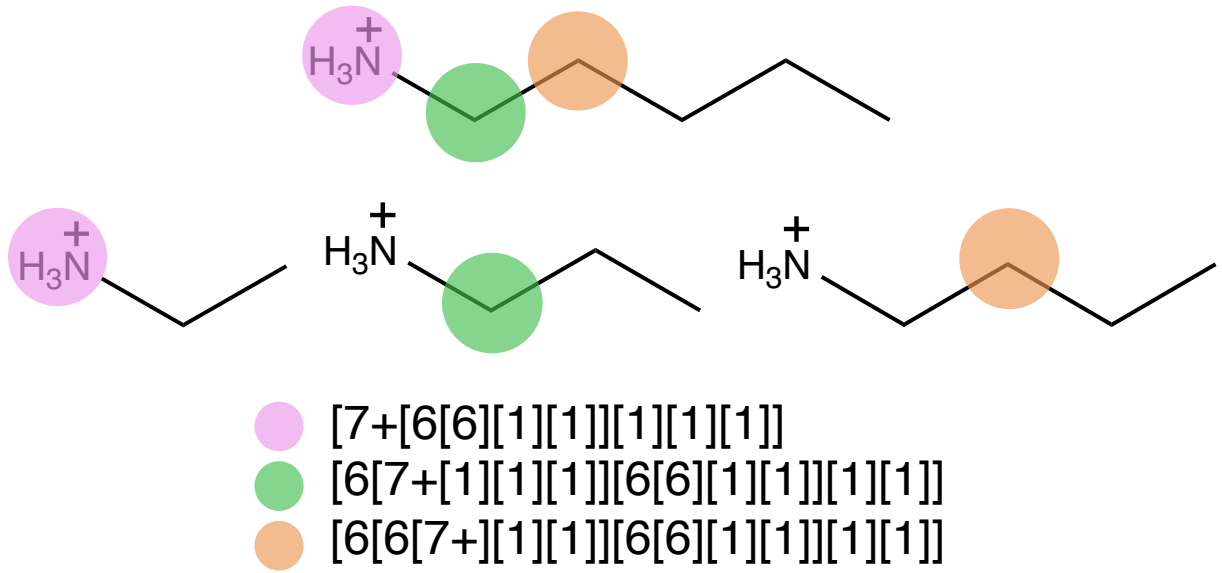
Ray Mentzer, Katherine Young, and the Pharma P2SAC Subgroup have spent the semester trying to break TCIT against a range of new functional groups and reactions.

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

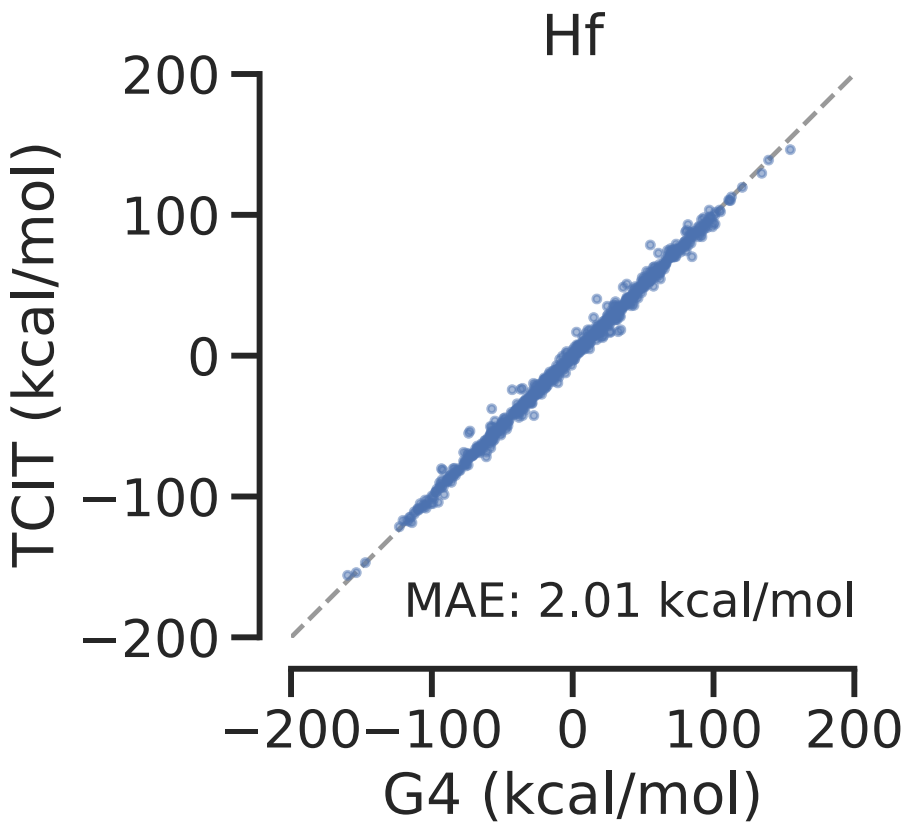
For radicals we've tested adding a hydrogen-bond increment for a closed-shell analog that TCIT can already predict.

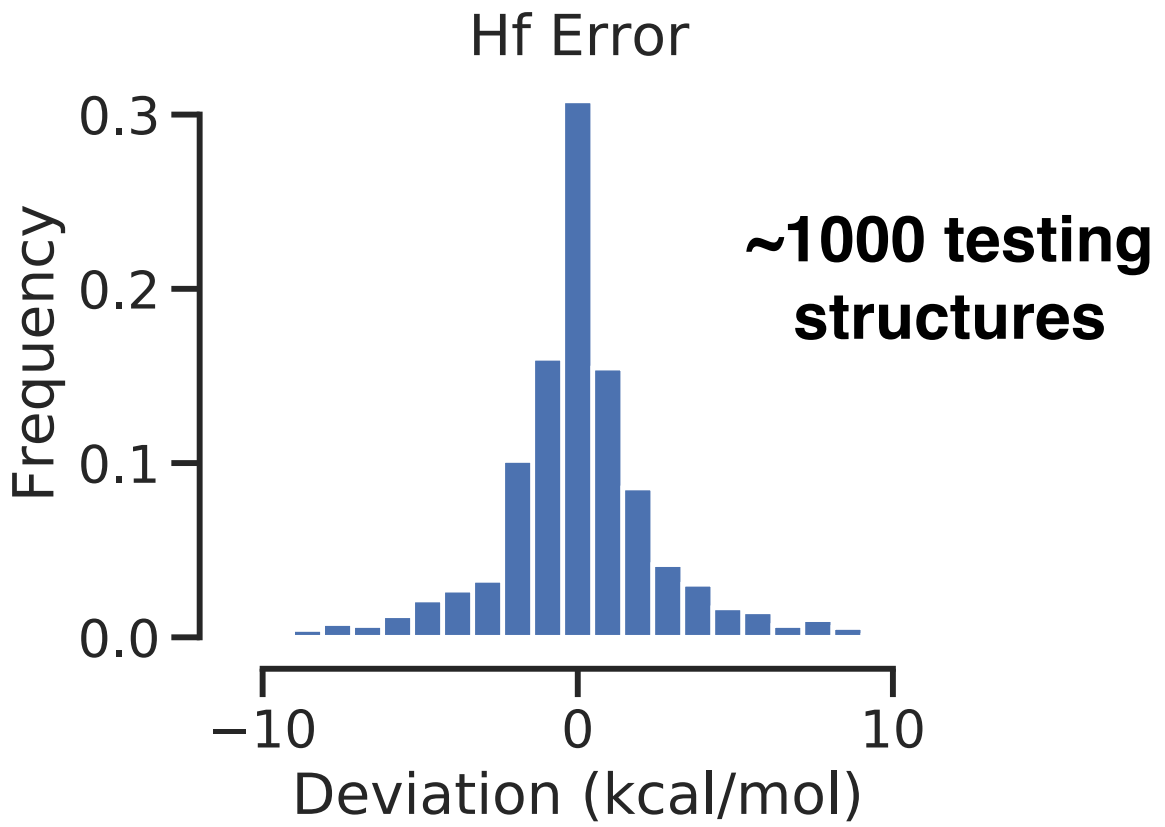


For ions we've implemented component typing and model compound generation that respects formal charges. Thus, predictions will proceed as for neutral species.



Preliminary TCIT Radical Benchmarks (ΔH_f)

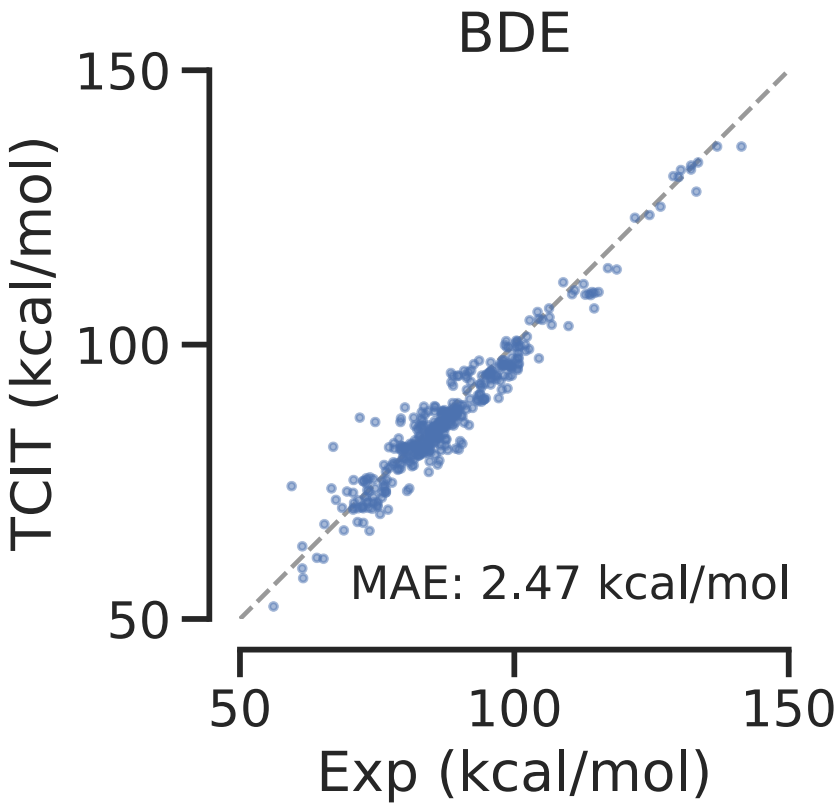


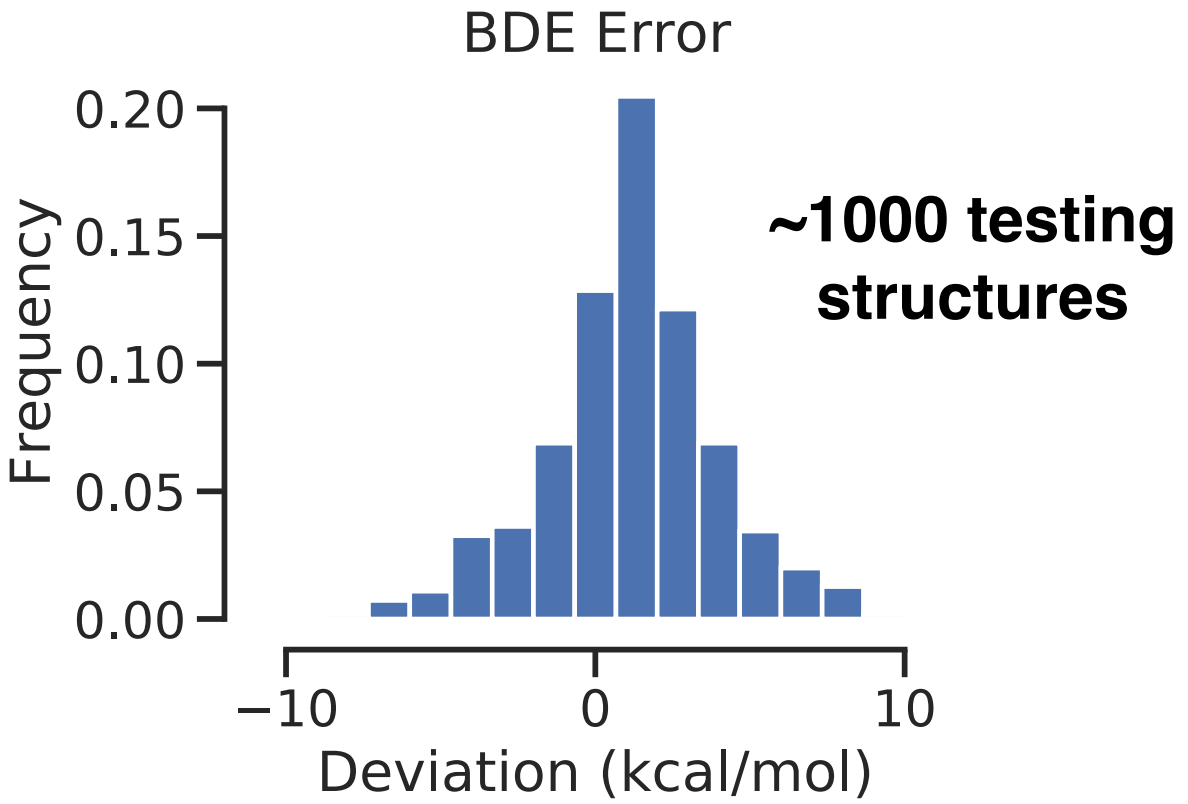


~1000 testing structures

- In this case, there simply isn't much experimental data available, so most of our validations are occurring at the G4 level (some experimental comparisons are presented later).
- ΔH_f errors are consistent with previous benchmark for closed-shell species.

Preliminary TCIT Radical Benchmarks (BDE)

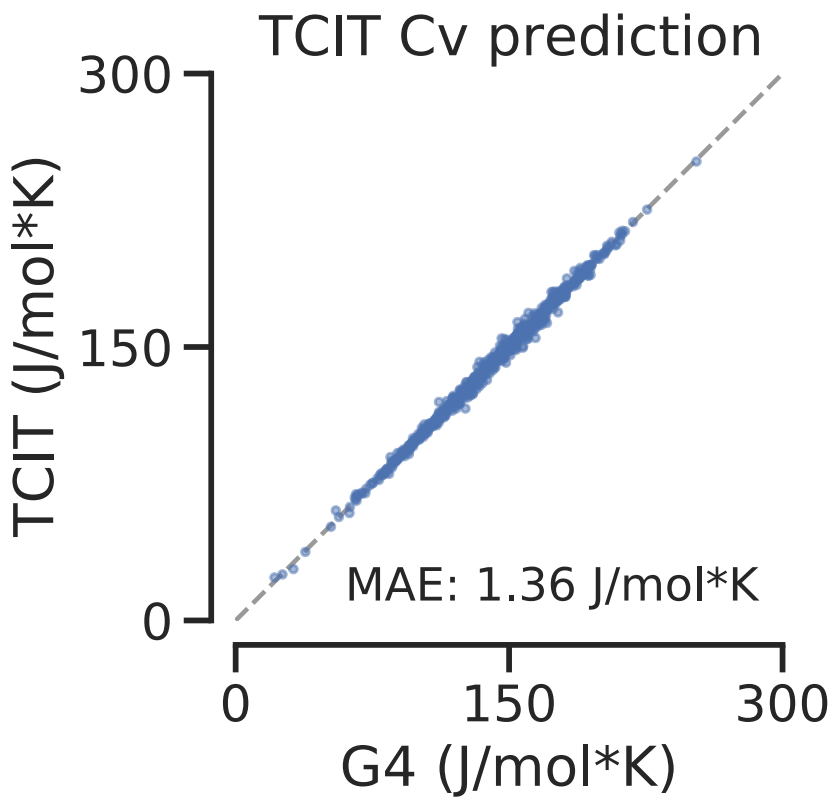


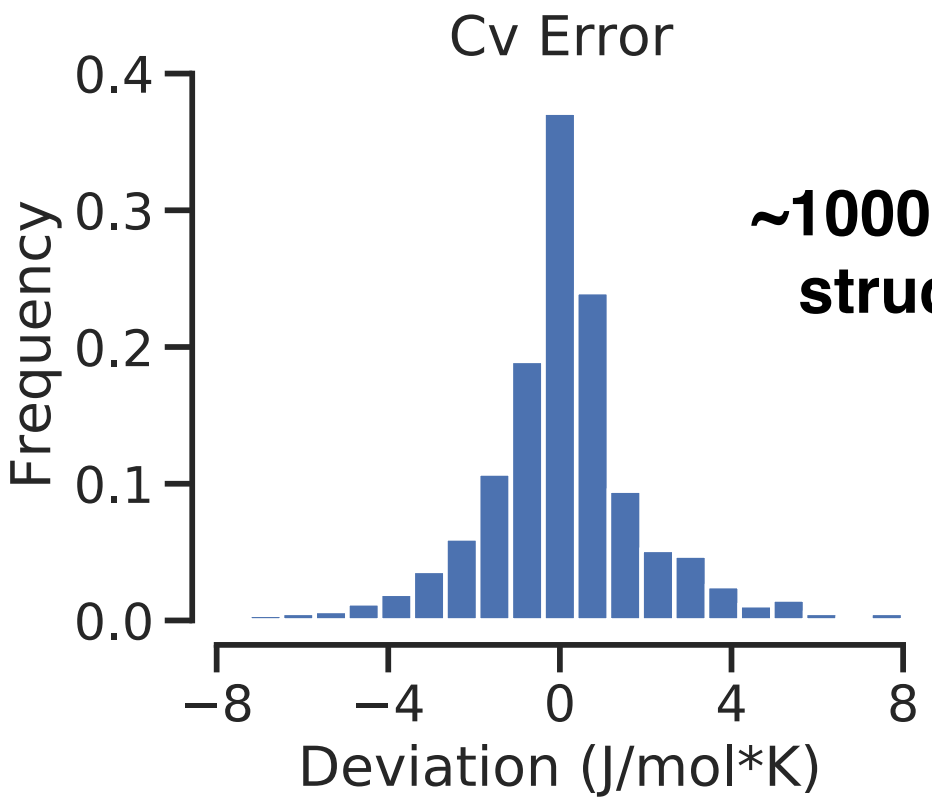


~1000 testing structures

- The bond-dissociation energy (BDE) is another test that is relevant to radicals.
- Errors are consistent with error propagation (**Note:** BDE is calculated as the enthalpy difference between the neutral and two radicals generated by the scission)

Preliminary TCIT Radical Benchmarks (C_v)

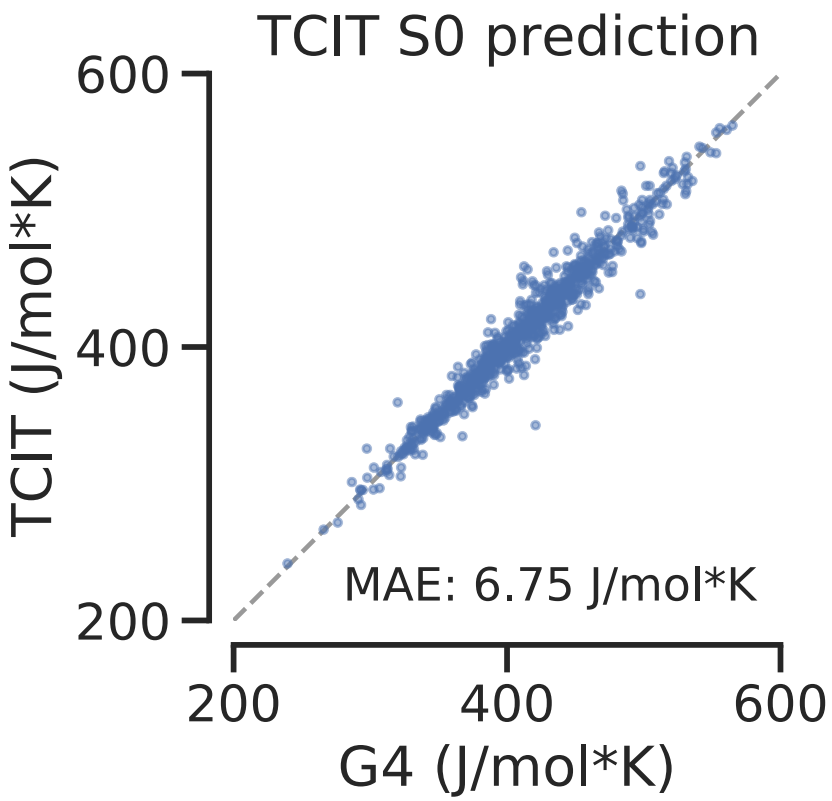


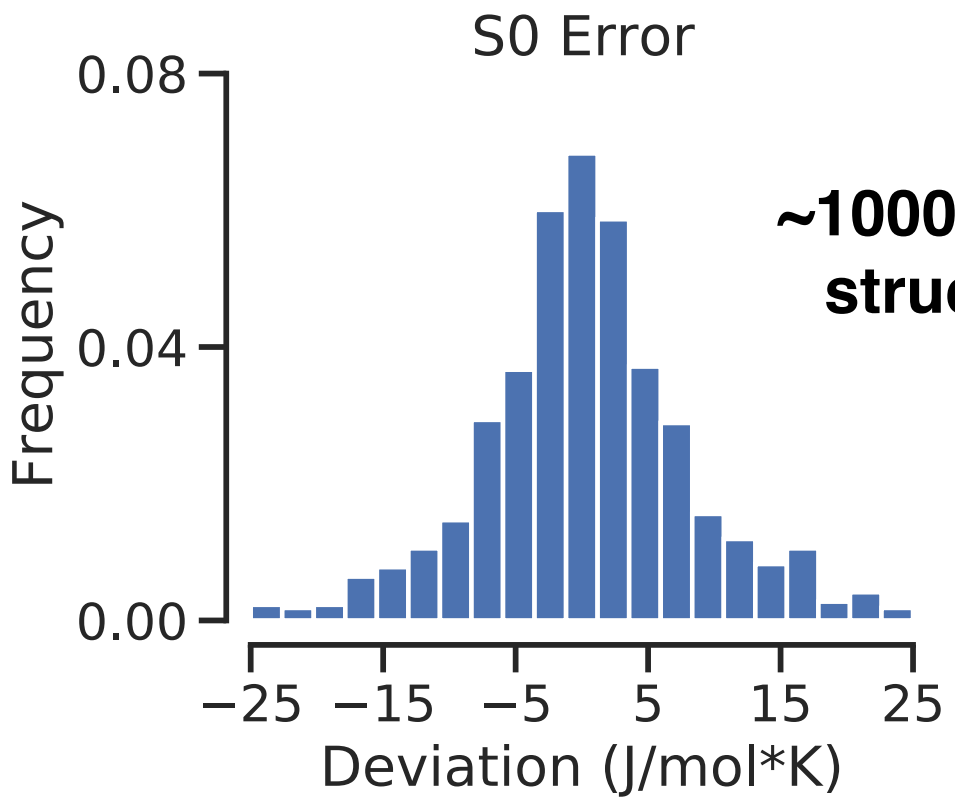


~1000 testing structures

- Similar accuracy compared with closed-shell species.

Preliminary TCIT Radical Benchmarks (S^o)





~1000 testing structures

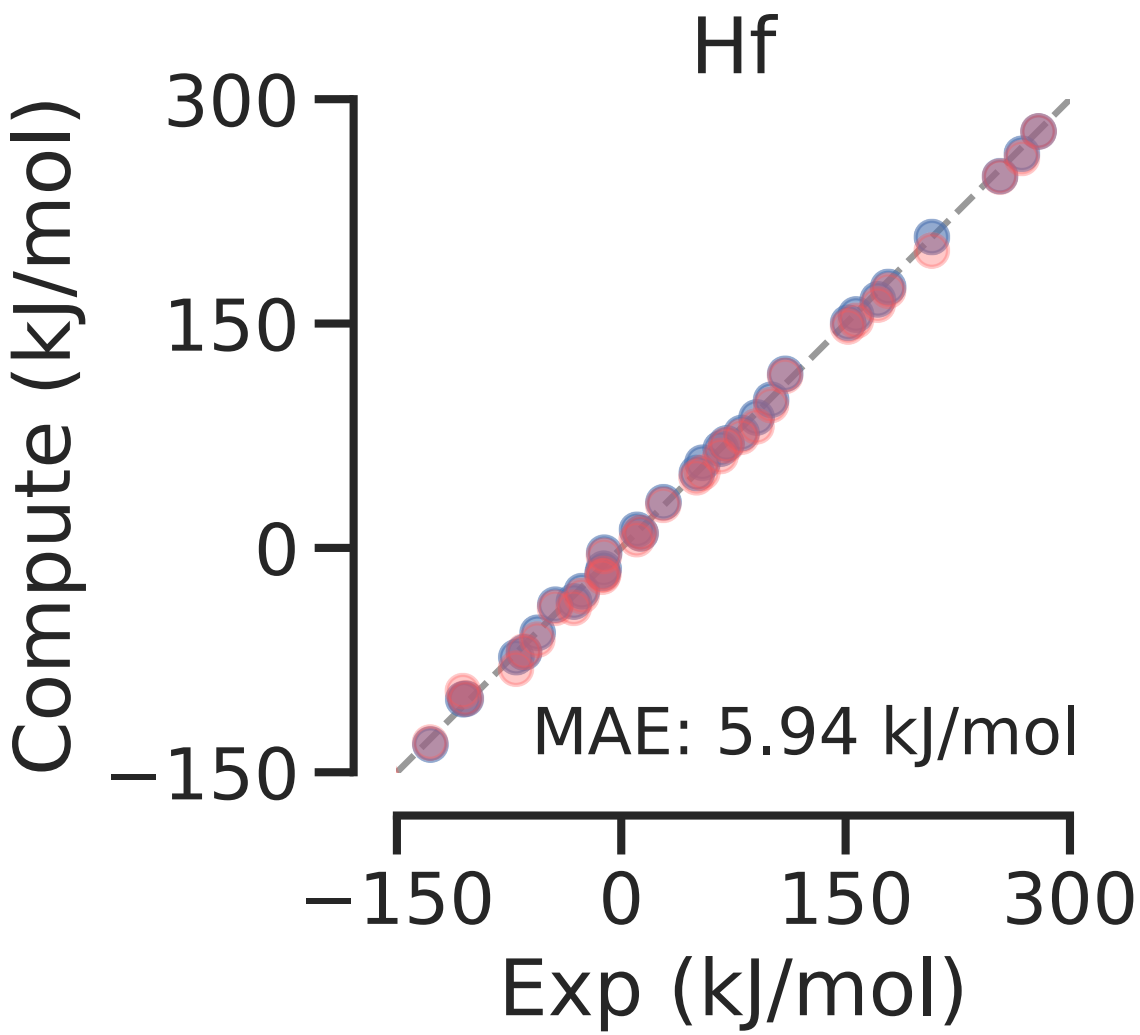
- Similar accuracy compared with closed-shell species.

Preliminary TCIT Radical Benchmarks (ΔH_f exp)

~40 experimental values have been cobbled together from NIST and ATCT testing structures

- Preliminary validation is looking promising, with similar overall performance for TCIT on radical and closed-shell species.

- This is a big victory for TCIT, since the limited experimental data for radical species makes CHETAH predictions impossible in many cases.



The Reaction Prediction Problem

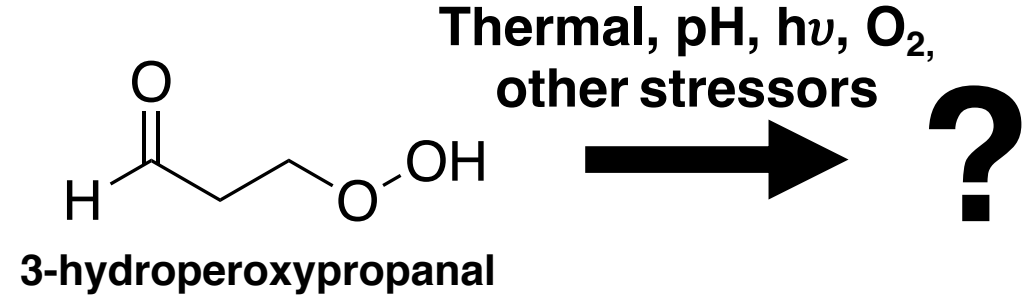
A → B : When we know the reactants and products, mature quantum chemistry tools exist to characterize transition states and establish pathways

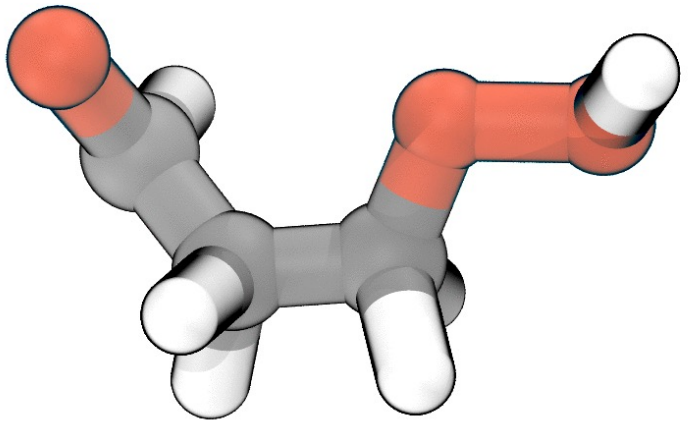
A → ? : For degradation reactions, plausible reactions are often unknown.

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.



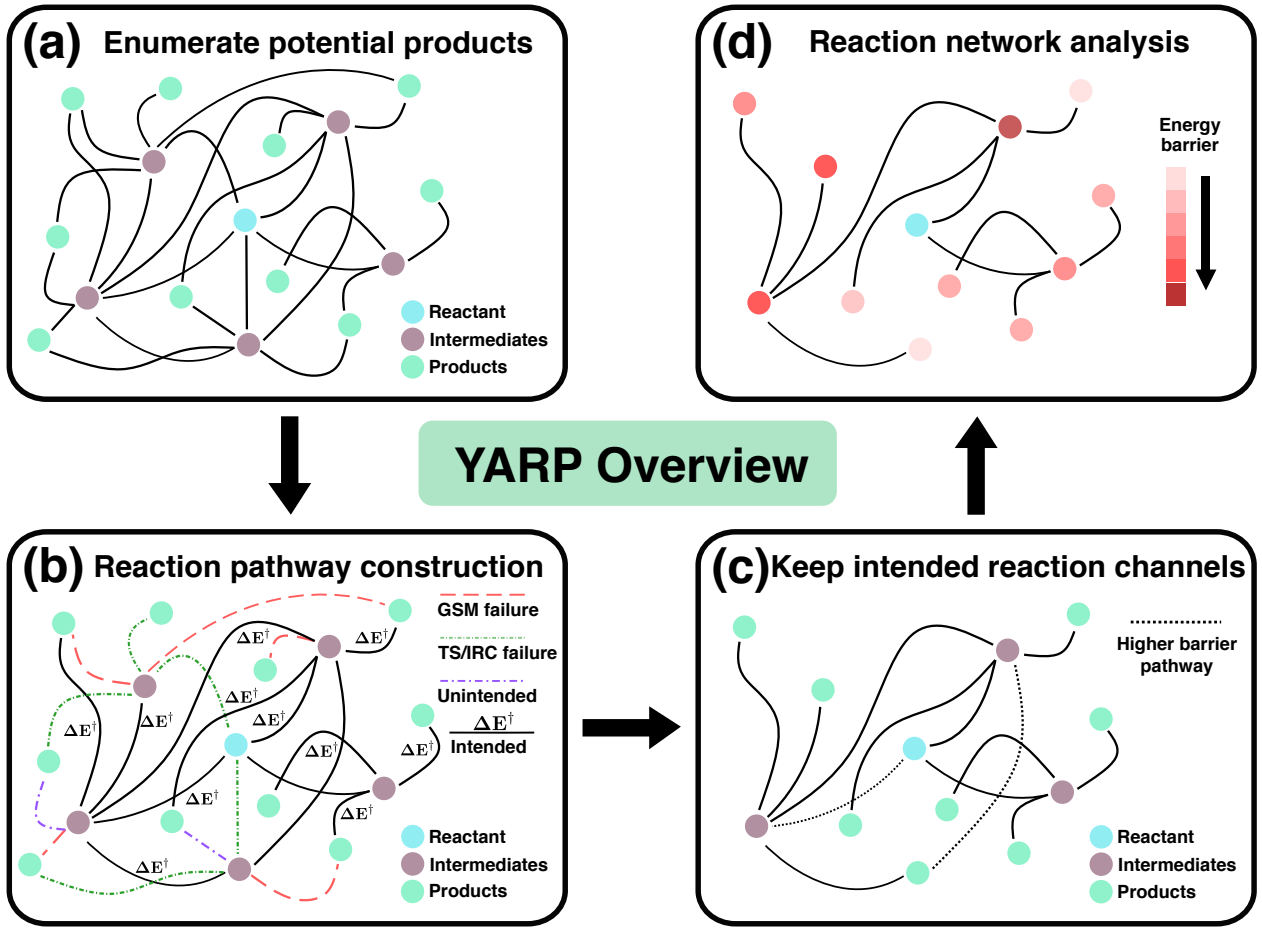


Yet Another Reaction Program (YARP)

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

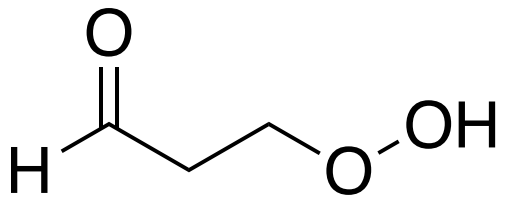
Observations:

- **Product enumeration is easier than reaction enumeration.**
- **Transition state algorithms for A→B problems are mature. Let the TS algorithm identify physical reactions.**
- **Recent developments in semi-empirical quantum chemistry (and ML) can be leveraged here.**



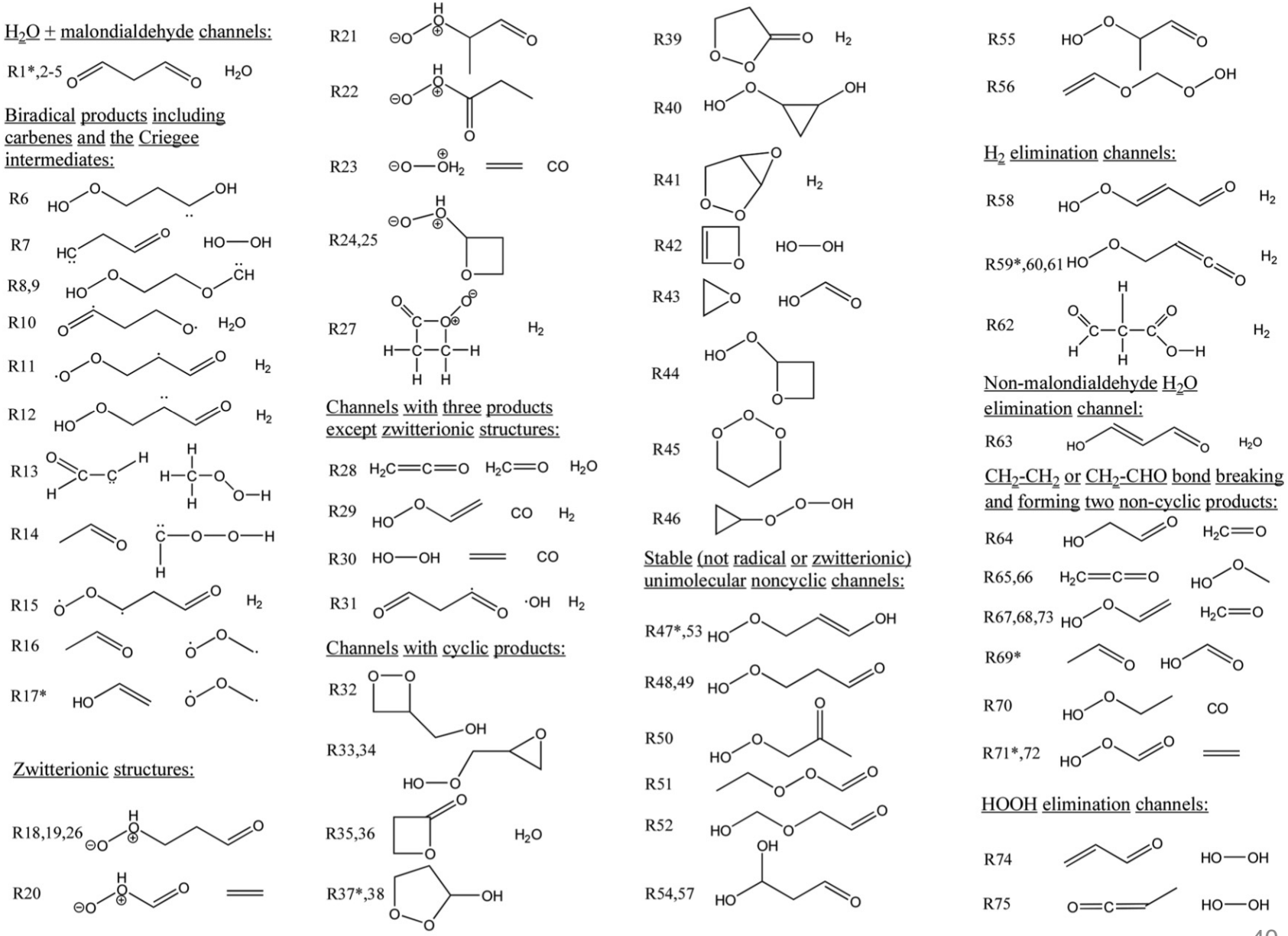
Qiyuan, Z.; Savoie, B. M. "Simultaneously Improving Reaction Coverage and Computational Cost in Automated Reaction Prediction Tasks." Nature Computational Science **2021**, 1, 479-490. ***(P2SAC Publication!)***

Testing YARP on a Unimolecular Decomposition Problem

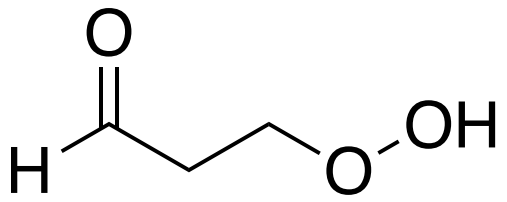


The 3-hydroperoxypropanal reaction network out to b4f4 was recently published as a benchmark for 5 reaction discovery methods.

Grambow, C. A, Suleimanov, Y. V. et al. *J. Am. Chem. Soc.* **2018**, 140 (3), 1035–1048.



Testing YARP on a Unimolecular Decomposition Problem



The 3-hydroperoxypropanal reaction network out to b4f4 was recently published as a benchmark for 5 reaction discovery methods.

Grambow, C. A, Suleimanov, Y. V. et al. *J. Am. Chem. Soc.* **2018**, 140 (3), 1035–1048.

55 distinct products
(neglecting charged species)

562 reactions attempted

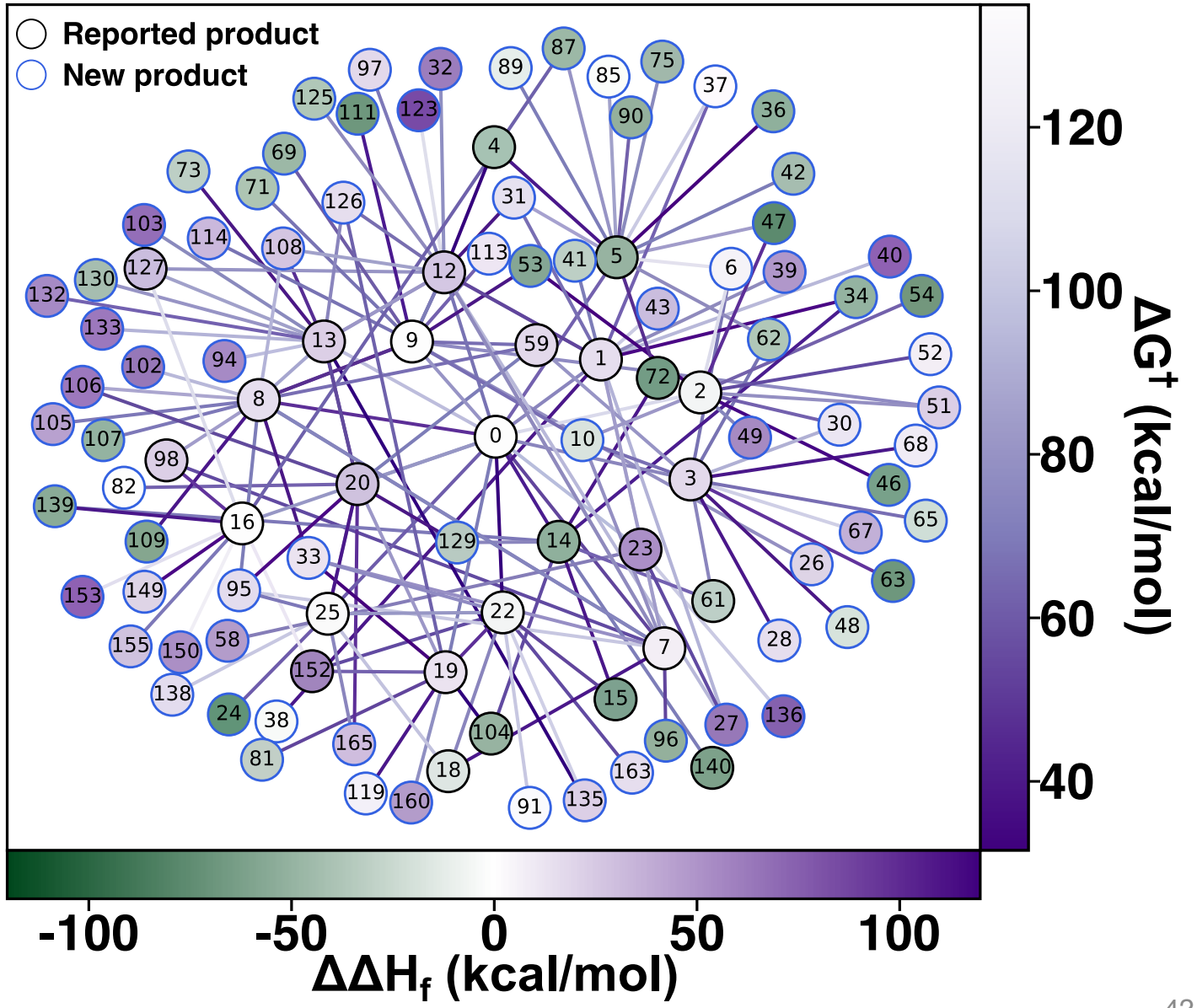
75 intended reactions discovered

None of the methods found all of the products of the others

3-Hydroperoxypropanal - Reaction Network

We used YARP to recursively elucidate the 3-hydroperoxypropanal unimolecular thermal degradation network for comparison with Grambow et al.

YARP finds **all known products** of this thermal decomposition network, as well as new products (77), and new reactions (157).

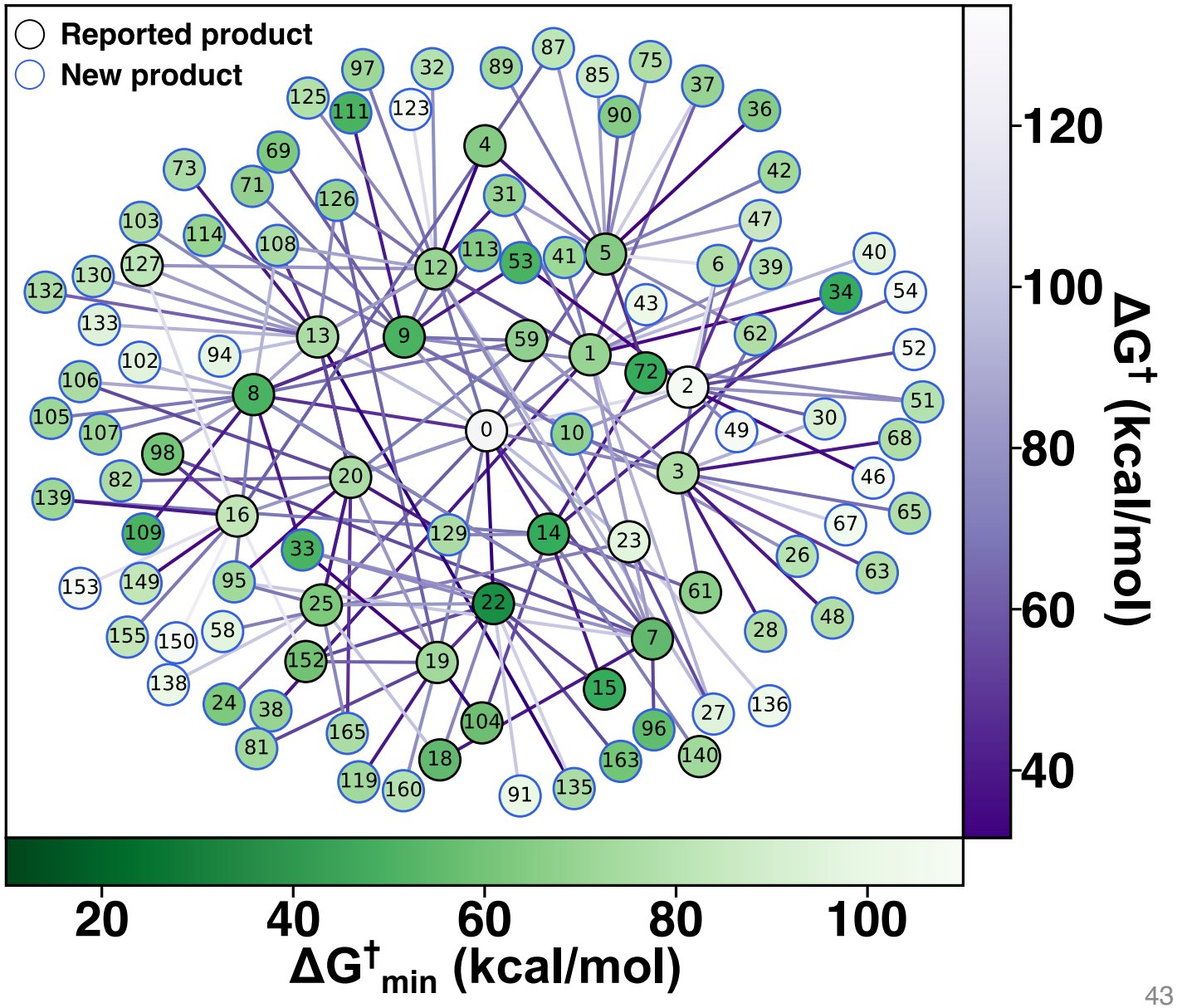


Qiyuan, Z.; Savoie, B. M. "Simultaneously Improving Reaction Coverage and Computational Cost in Automated Reaction Prediction Tasks." Nature Computational Science **2021**, 1, 479-490.

3-Hydroperoxypropanal - Reaction Network

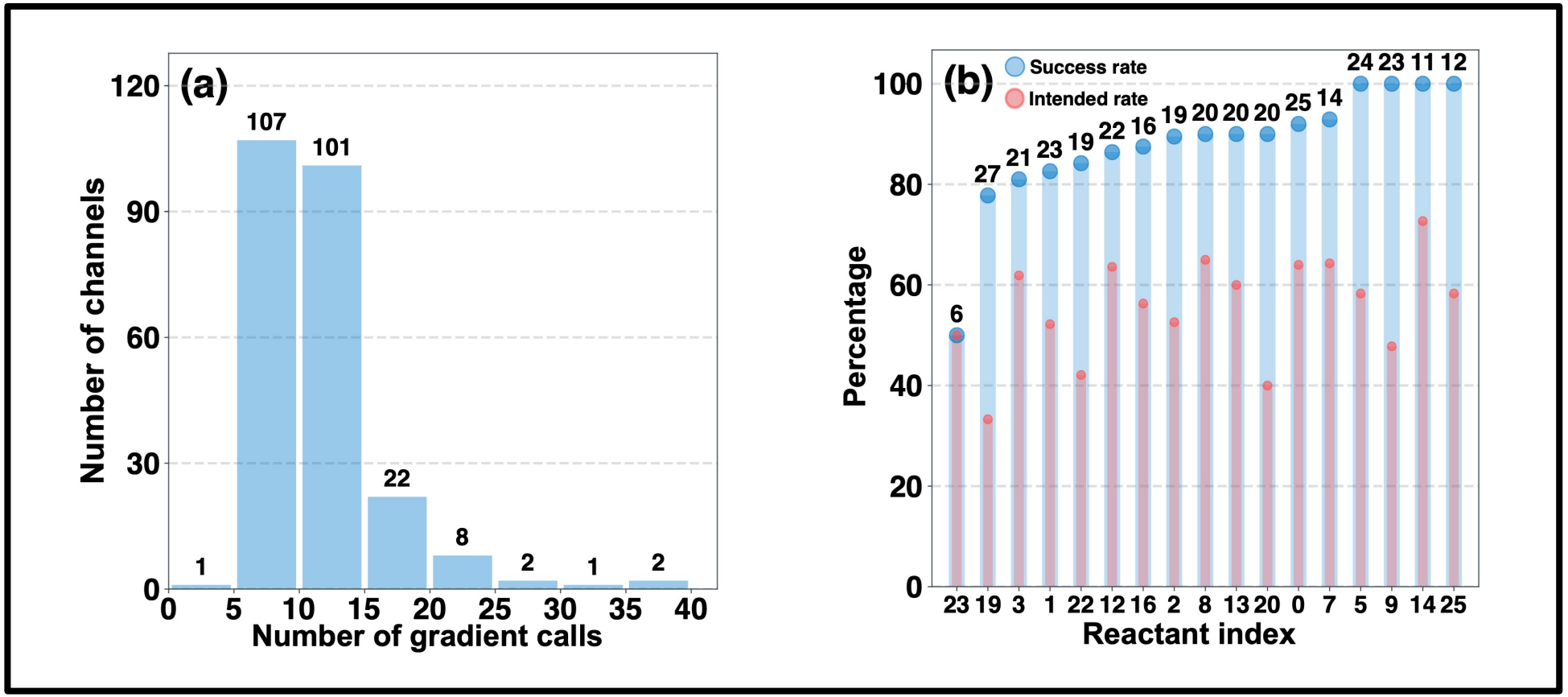
We used YARP to recursively elucidate the 3-hydroperoxypropanal unimolecular thermal degradation network for comparison with Grambow et al.

YARP finds **all known products** of this thermal decomposition network, as well as new products (77), and new reactions (157).



Qiyuan, Z.; Savoie, B. M. "Simultaneously Improving Reaction Coverage and Computational Cost in Automated Reaction Prediction Tasks." Nature Computational Science **2021**, 1, 479-490.

Predicting More (Reactions) with Less (Cost)

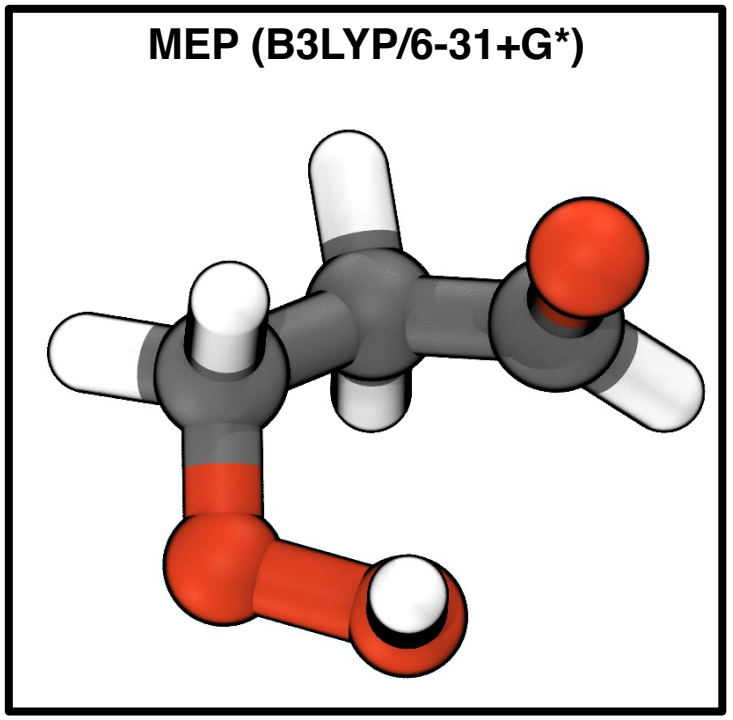


- Constructing the whole network required **8364** DFT gradient calls for YARP compared with **756,227** for the earlier benchmark (**100-fold reduction**)
  - Average success and intended rates for YARP are **81.4%** and **41.1%**, respectively, compared with **38%** and **4%**, in the earlier benchmark.

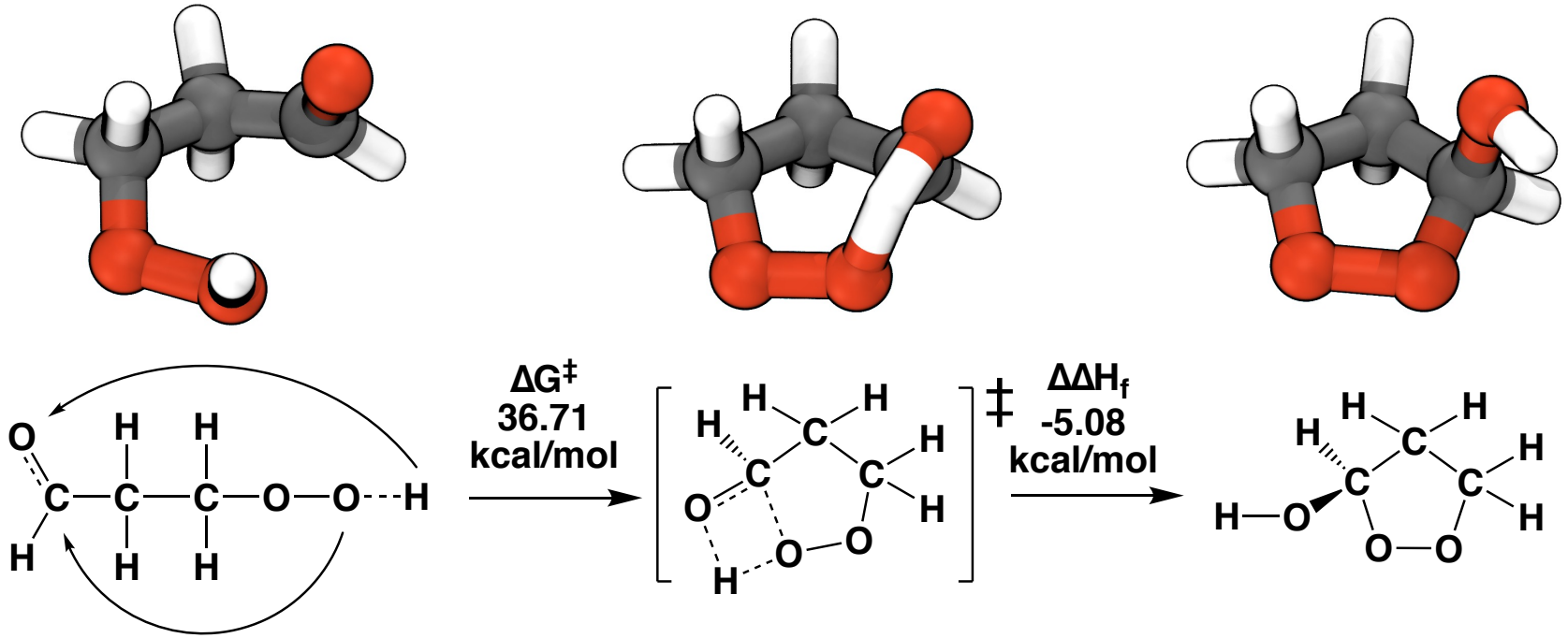
What Happens First?

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 degradation product.



Validated 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.

Outlook

The throughput enabled by YARP creates many new opportunities:

(i) Broader reaction discovery → Lowe Dataset as a discovery testbed

(ii) Generating positive and negative exemplary reaction datasets

(iii) Exploring deeper networks (e.g., materials degradation, catalysis)

In the process safety space, it seems like predicting kinetics in addition to thermochemistry could be useful.

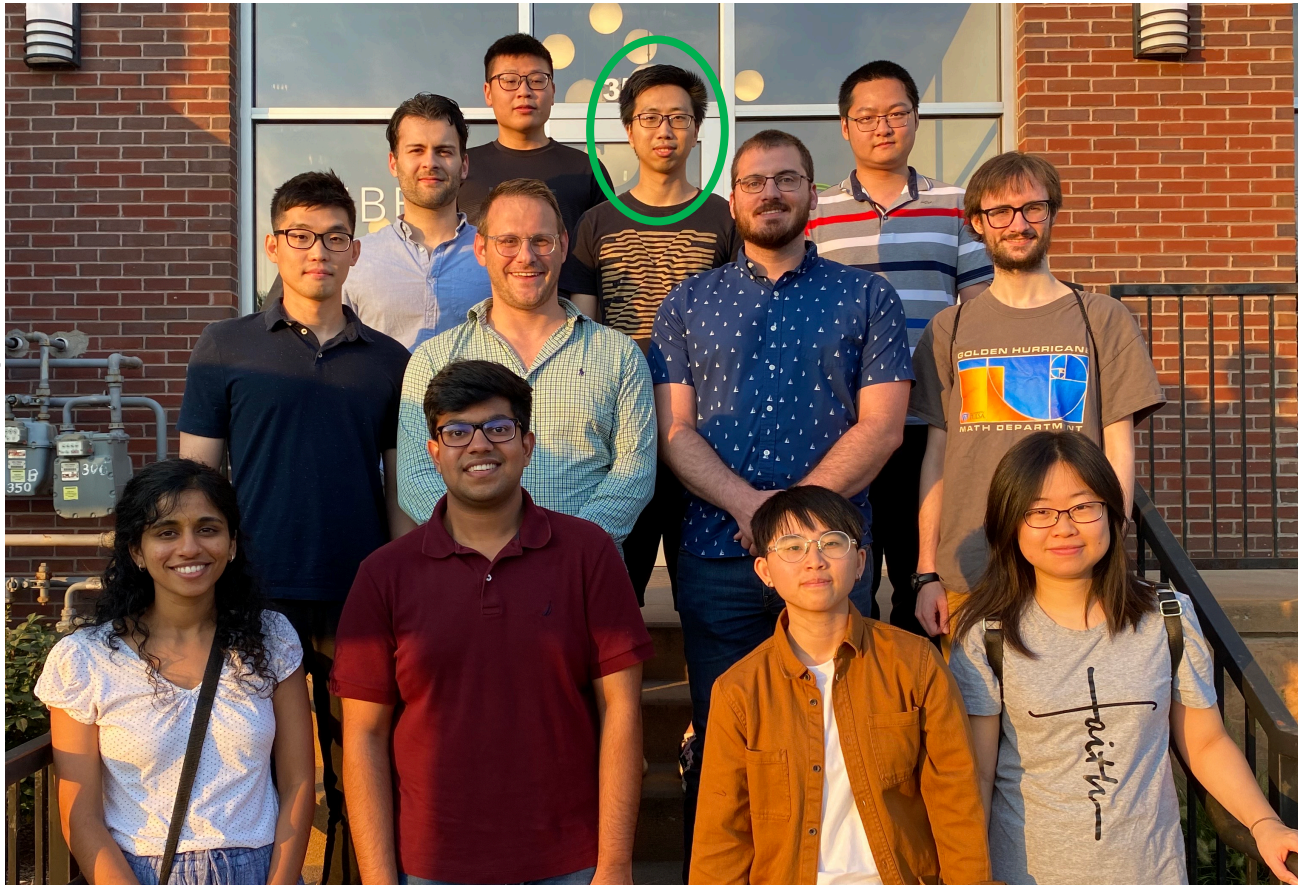
Qiyuan, Z.; Savoie, B. M. “Simultaneously Improving Reaction Coverage and Computational Cost in Automated Reaction Prediction Tasks.” Nature Computational Science **2021**, 1, 479-490.

Outlook and Acknowledgements

Qiyuan Zhao performed all of the work.

Project Accomplishments:

- Implemented a fully-consistent 2-bond (i.e., component) increment theory based on G4 data
- Automated model compound generation and fitting algorithms.
- Built a database infrastructure for reusing calculations and parameter fitting.
- Developed a ring-correction for TCIT to improve performance on conjugated and non-benzene structures.
- Extended TCIT to condensed phases and new thermodynamic properties and radicals.



- P2SAC for funding.
- Ray Mentzer (Purdue)
- Katherine Young (Purdue UG)

P2SAC
Purdue Process Safety & Assurance Center