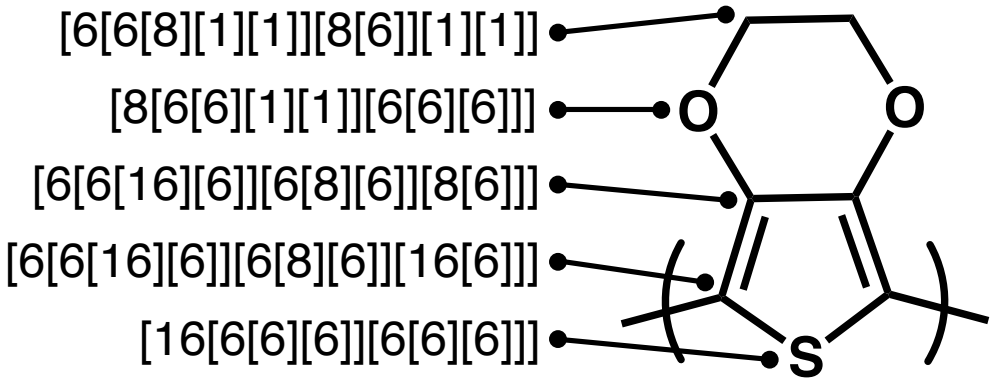
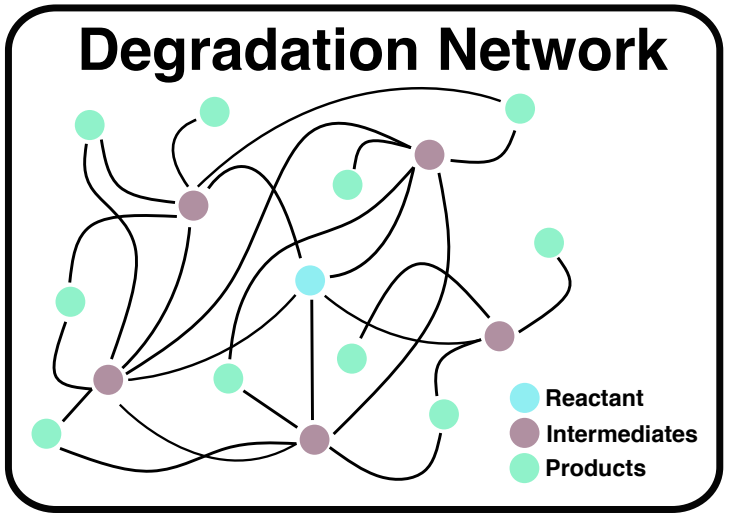
High-Throughput Predictions of Molecular Thermodynamics and Reactivity

A new component theory:



ΔH_f , S^o, C_v

Predicting Decomposition Pathways:



Qiyuan Zhao (3rd year PhD student)

Brett M. Savoie

Charles Davidson Assistant Professor of Chemical Engineering

5/17/21, P2SAC Spring Meeting (Day 3)

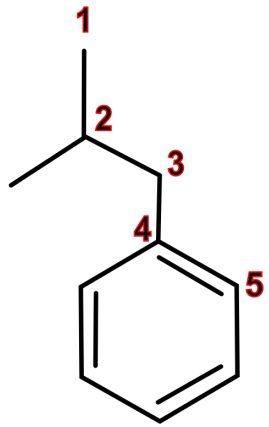
Challenges of Contemporary Group Theories

Benson Groups:

• 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



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: -5.15 +/- 0.34 kcal/mol

Problems we want to address:

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

• **Specificity:** the definition of a "group" has never been formalized and inconsistent granularity is applied. Limited information on group interactions.

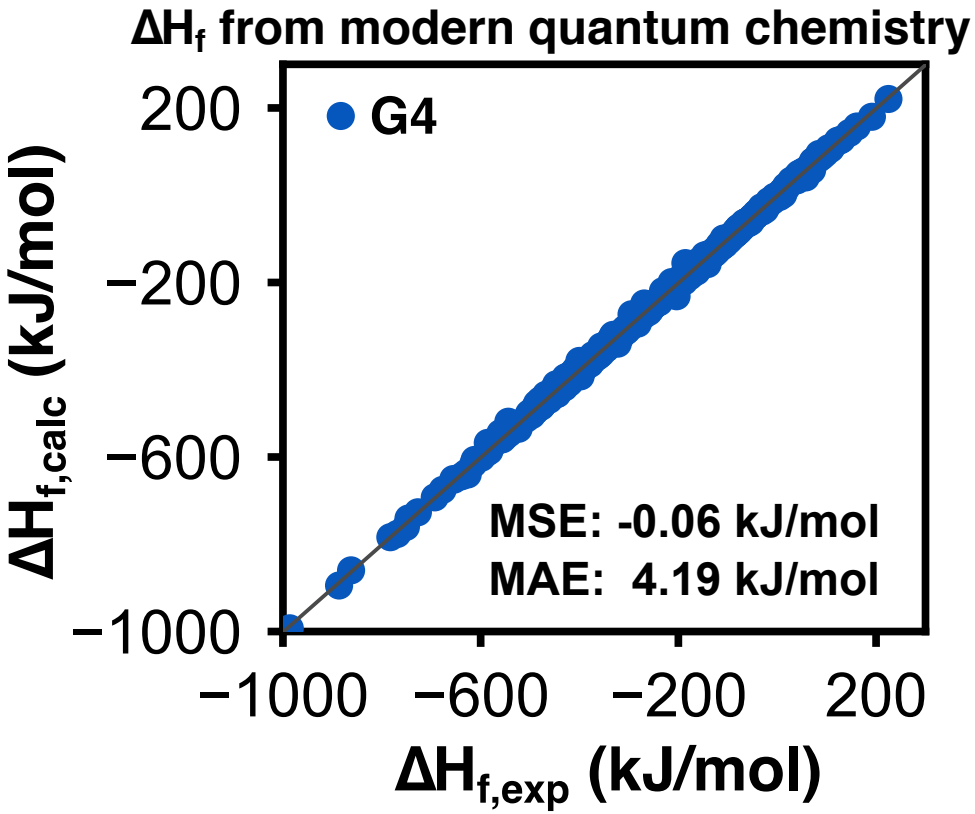
• **Extensibility:** because of the provenance and specificity problems, it isn't possible to develop new groups in a consistent way.

Challenges of Contemporary Group Theories

Benson Groups:

• 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.



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

Problems we want to address:

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

• **Specificity:** the definition of a "group" has never been formalized and inconsistent granularity is applied. Limited information on group interactions.

• **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

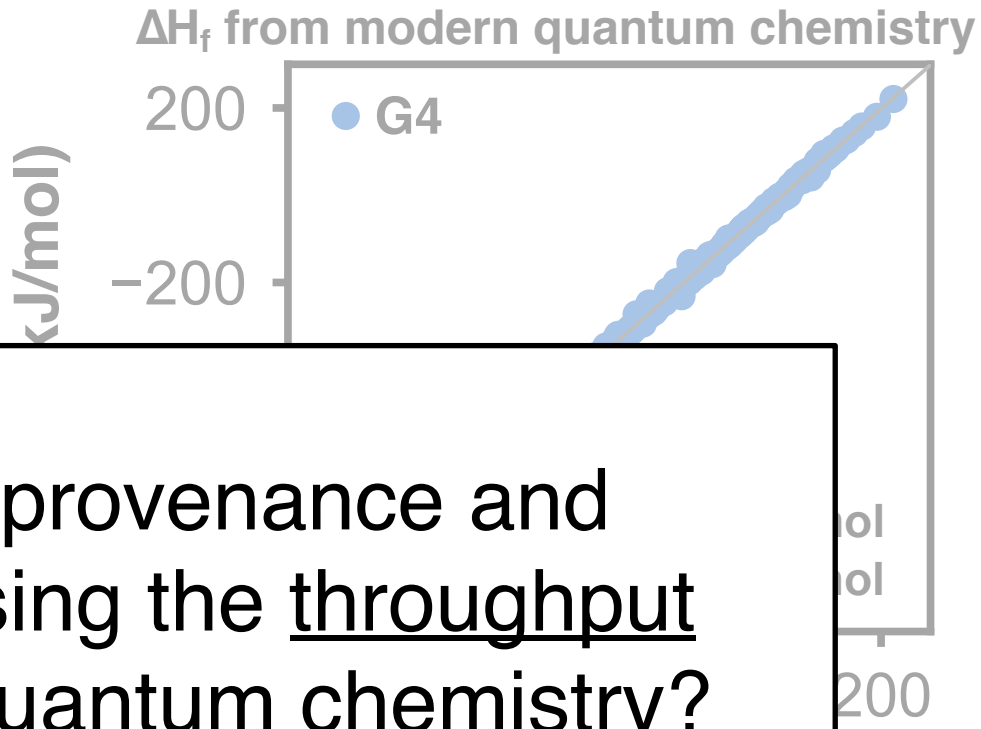
Benson Groups:

• The idea is to decompose molecular properties (ΔH_f, ΔS_f, ΔC_p) as the sum of "group" contributions.

• Grou
on trus
data, a

Prob

• **Prov**
is avai



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

Self-Consistent
2199-2207

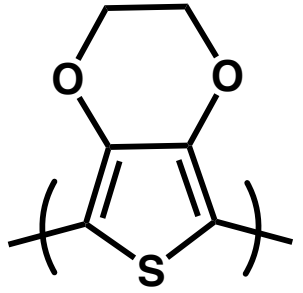
• **Specificity:** the definition of a "group" has never been formalized and inconsistent granularity is applied. Limited information on group interactions.

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

TAFFI Component Increment Theory (TCIT)

The fundamental idea:

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



Topology **A**utomated
Force **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

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.* In Press.

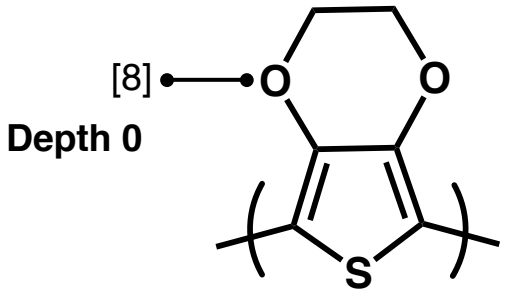
Seo, B.; Lin, Z-Y.; Zhao, Q.; Webb, M. A.; Savoie, B. M. "Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force-Fields." **2021** ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.14527299.v1

TAFFI Component Increment Theory (TCIT)

The fundamental idea:

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

TAFFI syntax for machine-readable atom types



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
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

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.* In Press.

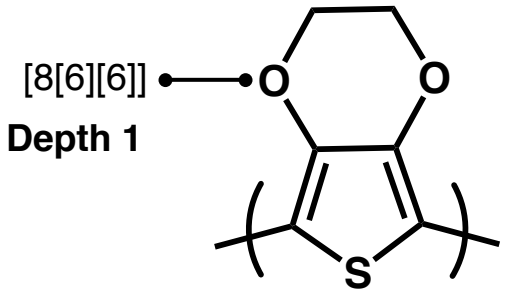
Seo, B.; Lin, Z-Y.; Zhao, Q.; Webb, M. A.; Savoie, B. M. "Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force-Fields." **2021** ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.14527299.v1

TAFFI Component Increment Theory (TCIT)

The fundamental idea:

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

TAFFI syntax for machine-readable atom types



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

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.* In Press.

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." **2021** ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.14527299.v1

TAFFI Component Increment Theory (TCIT)

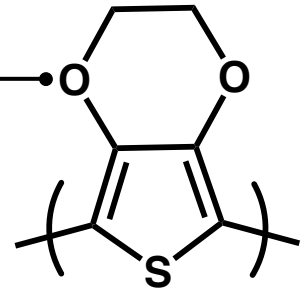
The fundamental idea:

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

TAFFI syntax for machine-readable atom types

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

Depth 2



Topology Automated Force Field Interactions

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	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

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.* In Press.

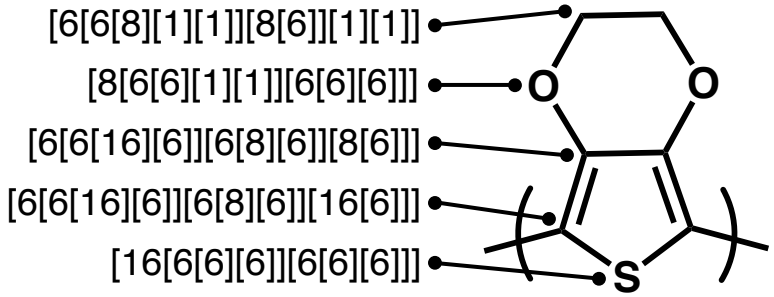
Seo, B.; Lin, Z-Y.; Zhao, Q.; Webb, M. A.; Savoie, B. M. "Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force-Fields." **2021** ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.14527299.v1

TAFFI Component Increment Theory (TCIT)

The fundamental idea:

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

TAFFI syntax for machine-readable atom types



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
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

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.* In Press.

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." **2021** ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.14527299.v1

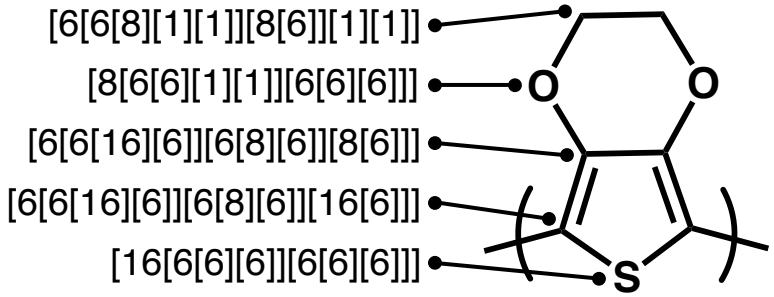
TAFFI Component Increment Theory (TCIT)

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. But parameterizing a component model **would not be feasible with only experimental data**.

TAFFI syntax for machine-readable atom types



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

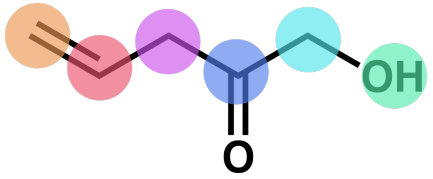
Zhao, Q.; Savoie, B. M.; "Enthalpy of Formation Prediction via a fully Self-Consistent Component Increment Theory". *J. Chem. Info. Model.* **2020**, 60, 2199-2207

Zhao, Q.; Iovanac, N.; Savoie, B. M.; "Transferable Ring Corrections for Predicting Enthalpy of Formation of Cyclic Compounds" *J. Chem. Info. Model.* In Press.

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." **2021** ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.14527299.v1

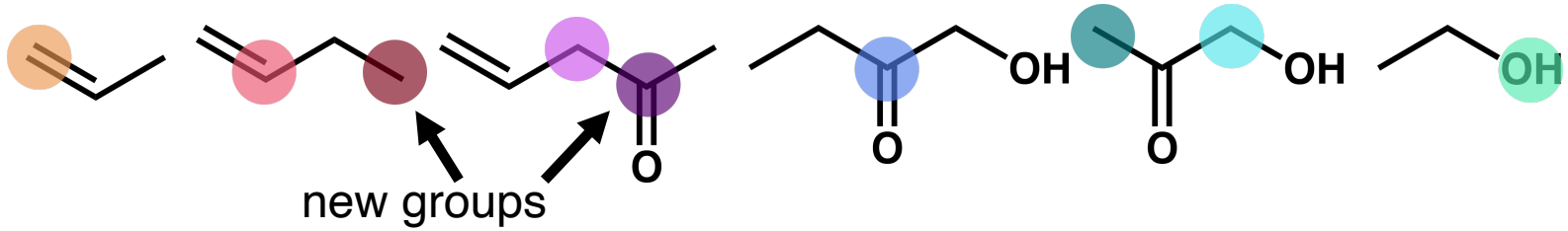
Graphical Decomposition of Model Compounds

identify components*

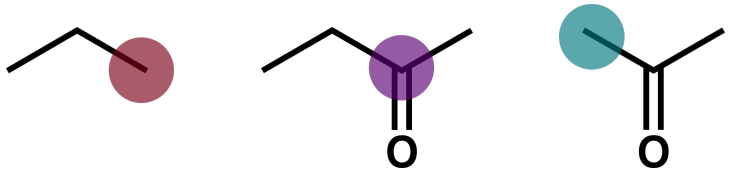


1-hydroxy-pent-2-ene-2-one

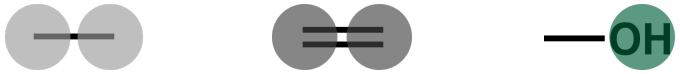
Recursively generate model compounds



Resolve dependencies



Resolve rank deficiency with elementary constraints

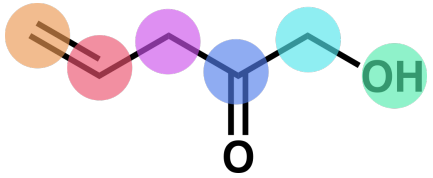


Zhao, Q.; Savoie, B. M.; Enthalpy of Formation Prediction via a fully Self-Consistent Component Increment Theory. *J. Chem. Info. Model.* **2020**, 60, 2199-2207

Graphical Decomposition of Model Compounds

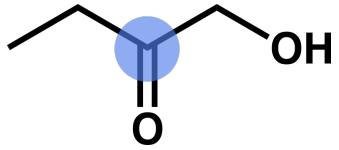
Topologically sort dependency graph

Prediction target:

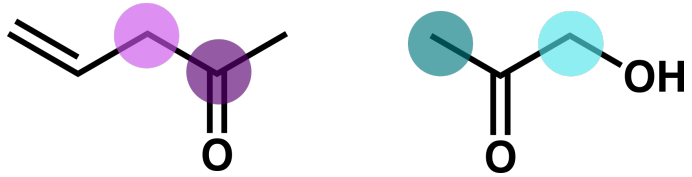


1-hydroxy-pent-2-ene-2-one

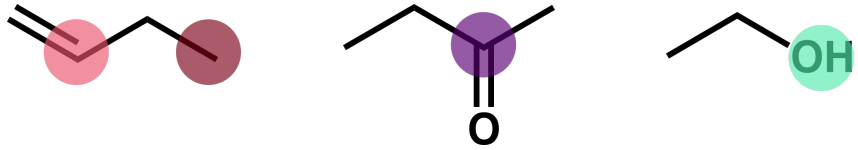
Gen 4:



Gen 3:



Gen 2:



Gen 1:

Gen 0:

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

no experimental data

Our implementation automatically plans calculations accounting for dependencies.

All calculations and components are databased for future use

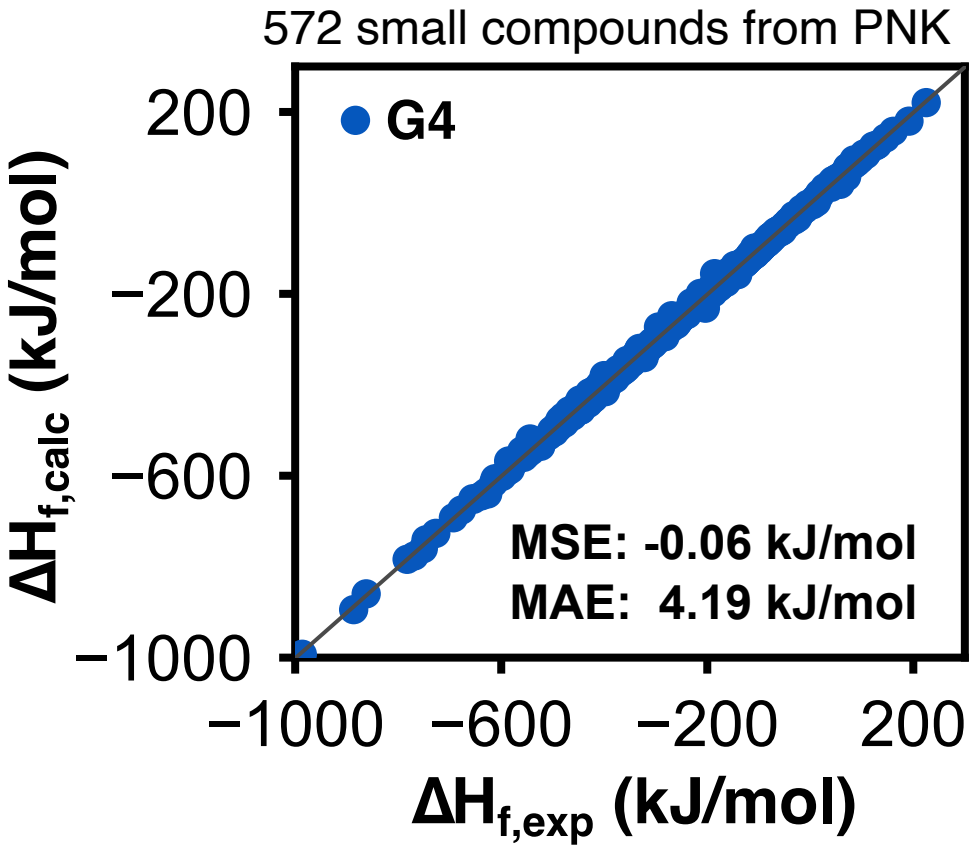
Benchmarking $\Delta H_{f,gas}$ Predictions Against the PNK Dataset

- 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

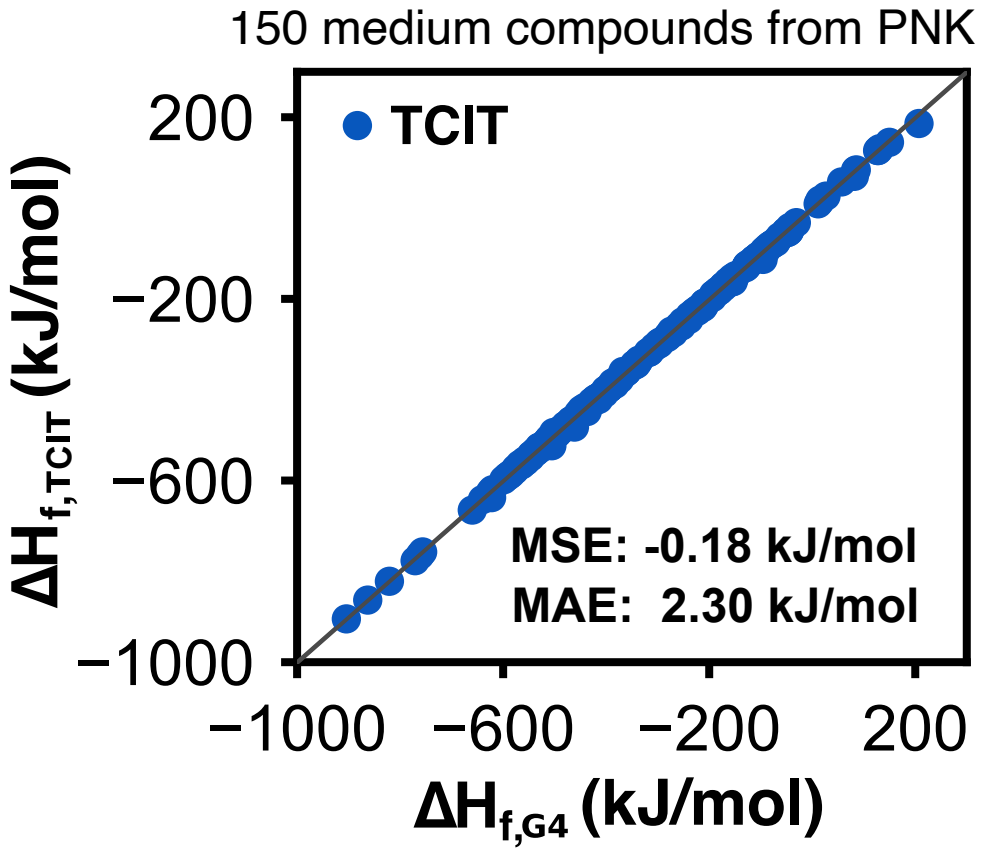
- Initial benchmarking set consists of ~1100 **linear** C,H, and O containing compounds from PNK[1]

(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]

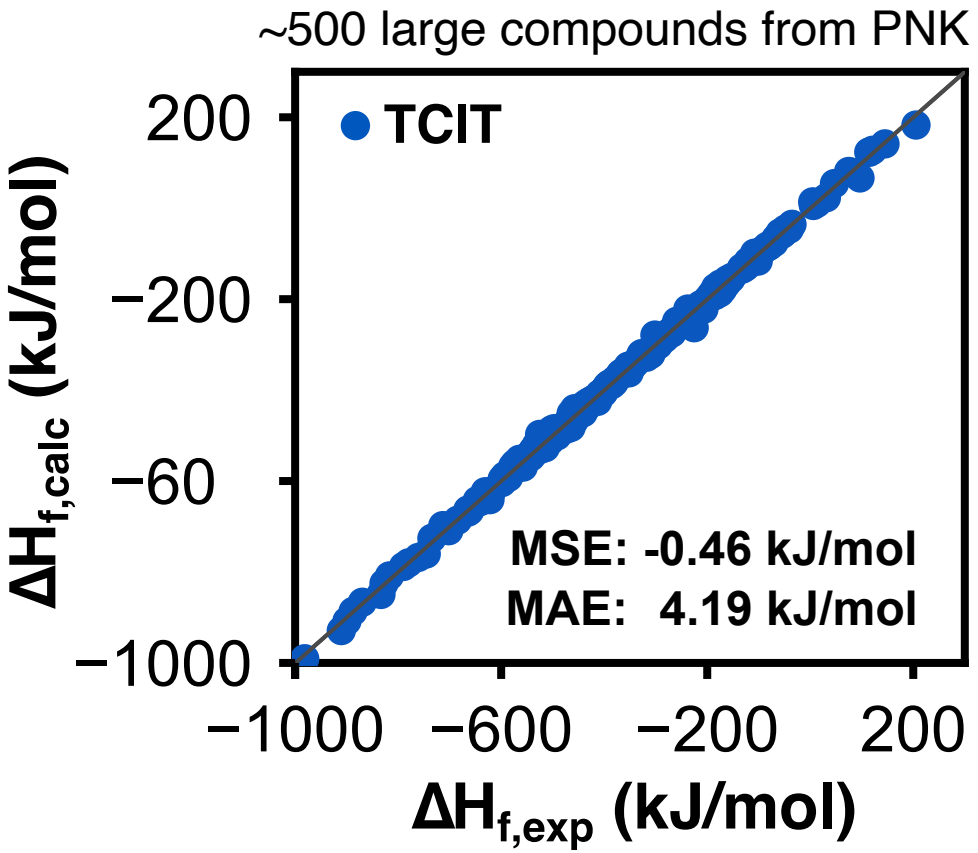
(1) J. B. Pedley, R. D. Naylor, S. P. Kirby "Thermochemical Data of Organic Compounds" 2nd ed. 1986

- 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]

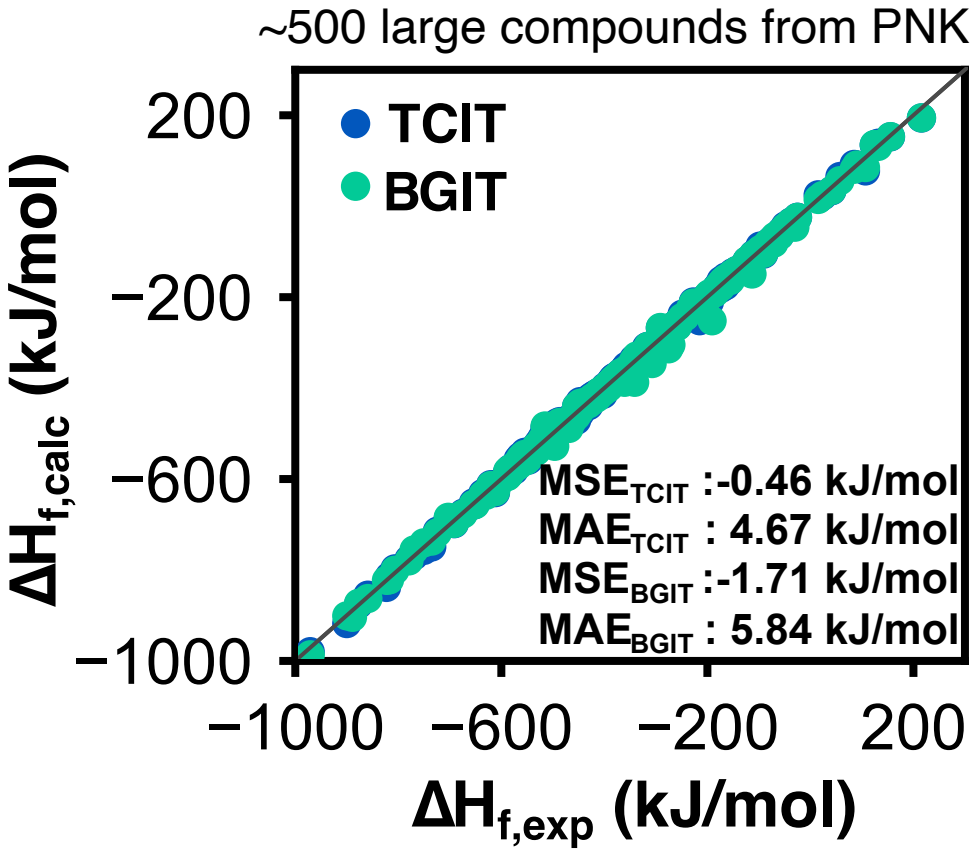
(1) J. B. Pedley, R. D. Naylor, S. P. Kirby "Thermochemical Data of Organic Compounds" 2nd ed. 1986

- 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

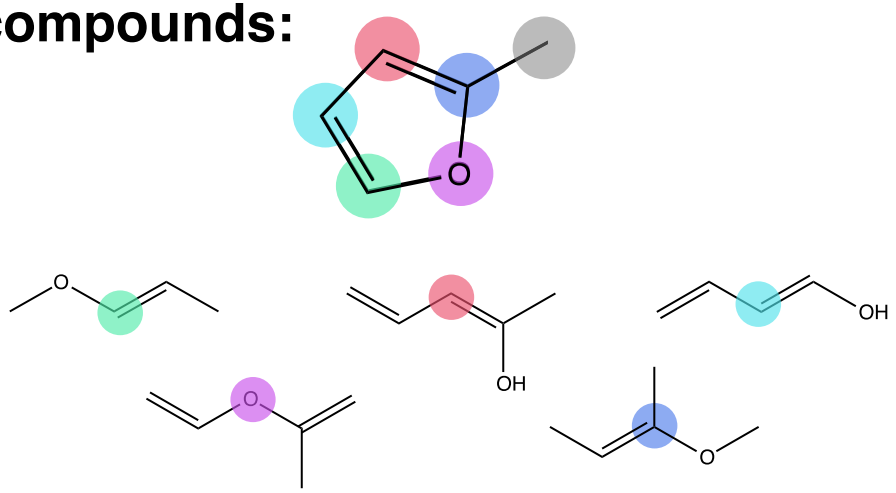
TCIT shows comparable performance to BGIT/CHETAH but is derived exclusively from extensible G4 data.

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:

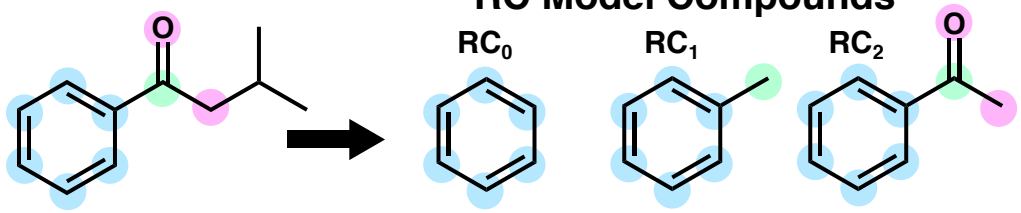


2. Add ring correction (RC) to final prediction:

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

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. In Press.

Benchmarking Ring-Correction Performance

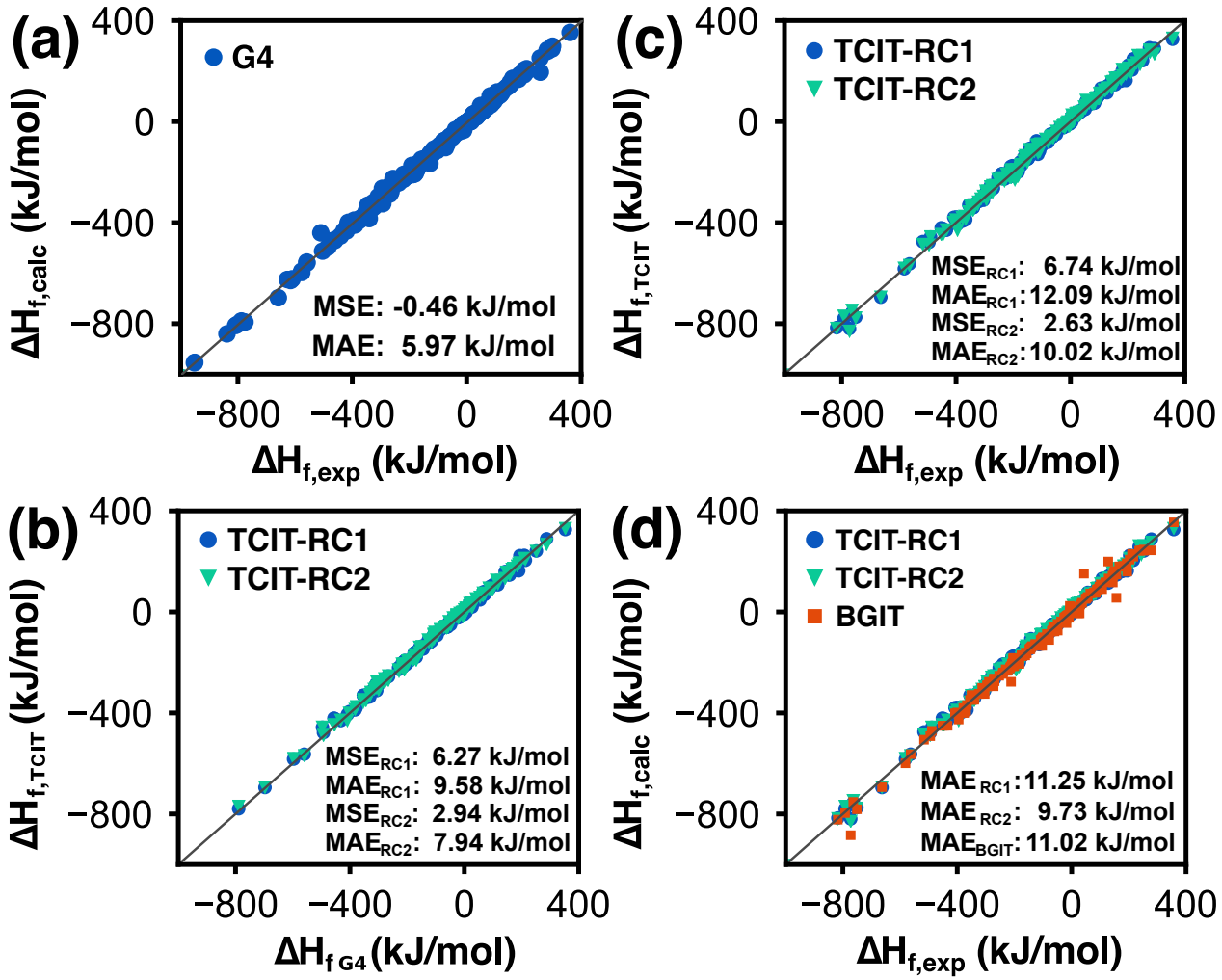
(a) G4 errors are marginally larger for ring-containing compounds but still very accurate

(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

Benchmarking Ring-Correction Performance

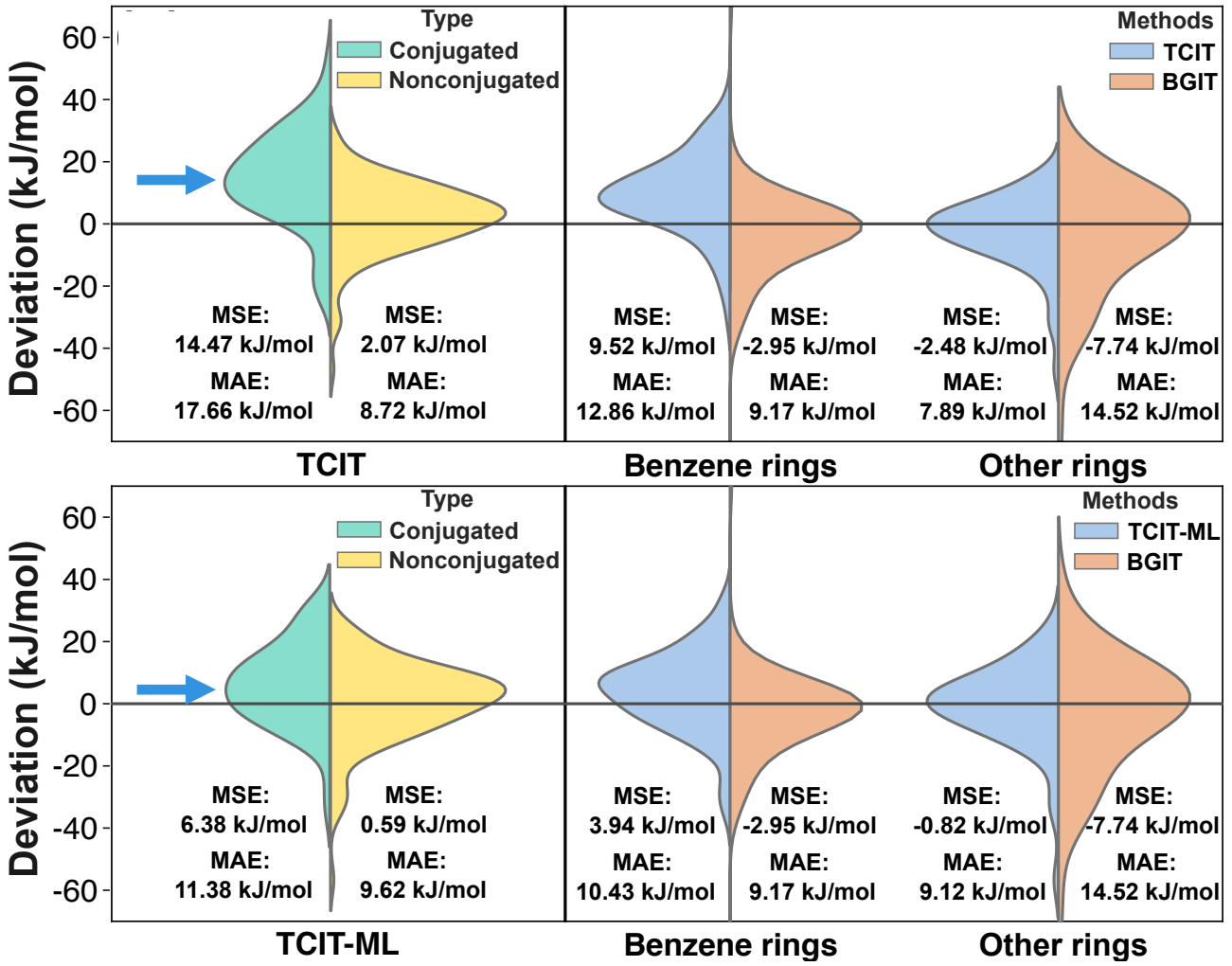
Breaking down distinct subsets:

- Conjugated systems are challenging to accurately predict with a local ring correction.

- BGIT has excellent performance on benzene rings due to the prevalence of experimental data, but poor performance on novel rings.

- The ML ring-correction shows the strongest overall performance. This strategy could also be used to generically correct for long-range conjugation effects.

~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:

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

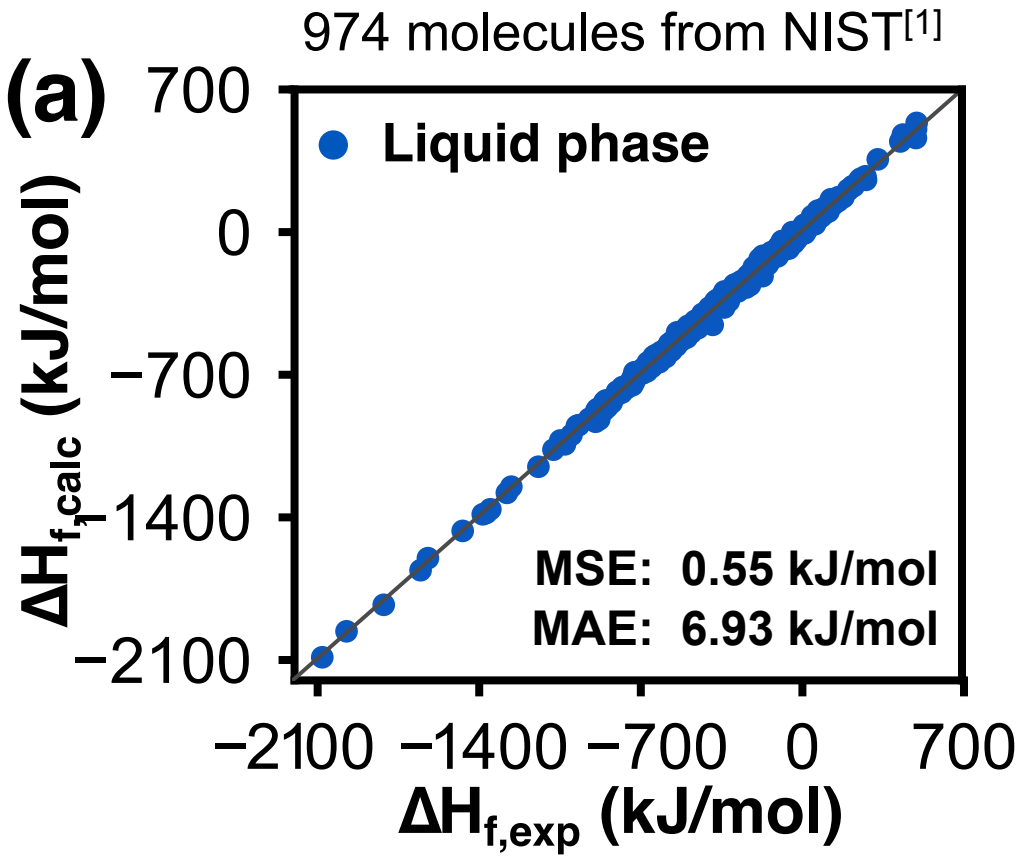
[1] Murray, J.S., Brinck, T. and Politzer, P., **1996**. *Chemical physics*, 204, 289-299.

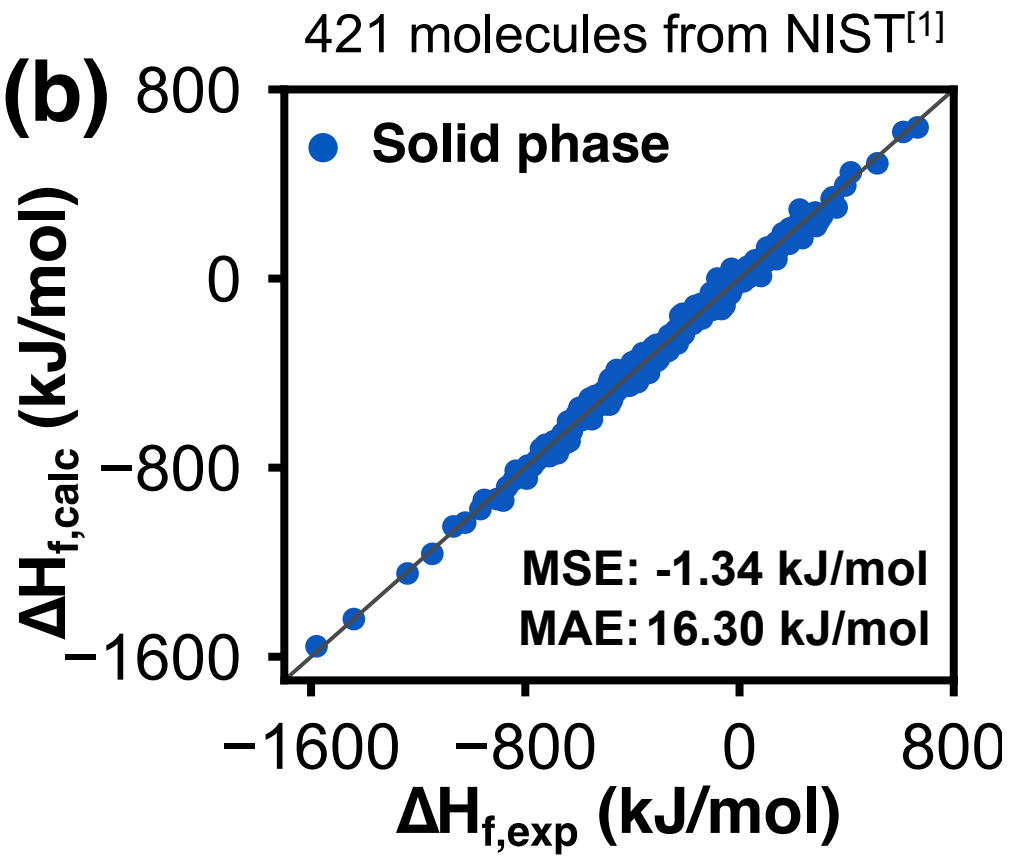
[2] Pankow, J.F. and Asher, W.E., **2008**. *Atmospheric Chemistry and Physics.*

[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.

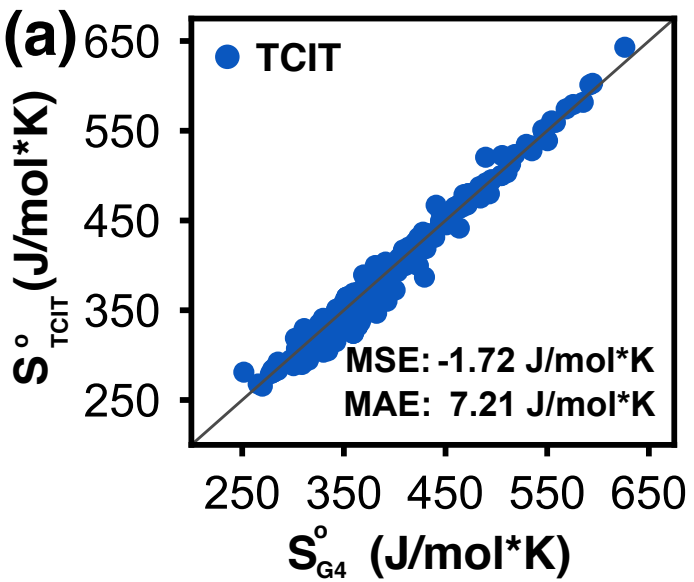
(b) G4/TCIT S° comparison for 314 medium sized molecules.

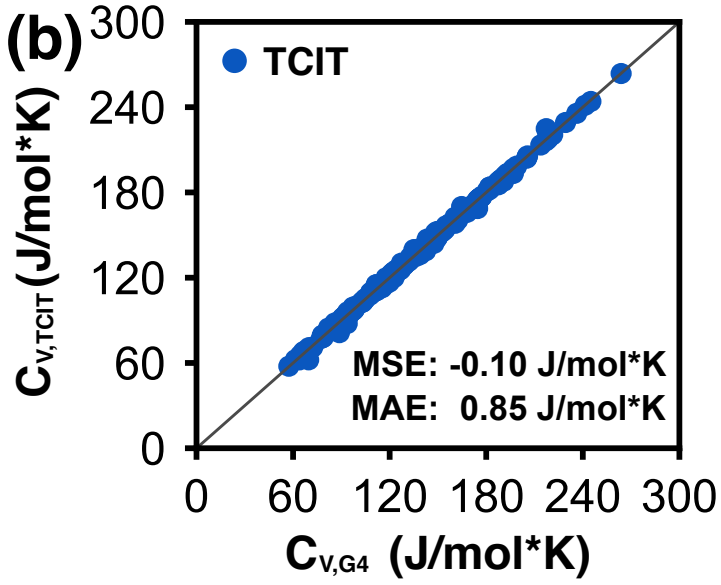
(c) TCIT S° comparison for 439 large molecules from NIST[1]

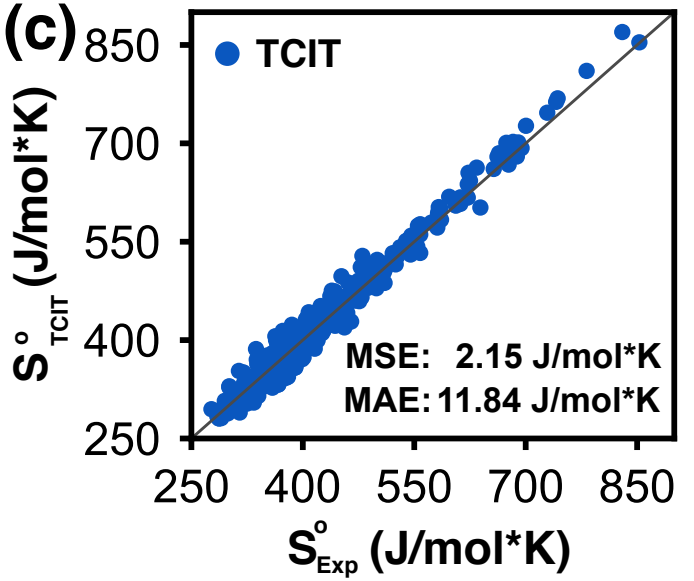
(d) TCIT heat capacity comparison for 904 large molecules from NIST[1]

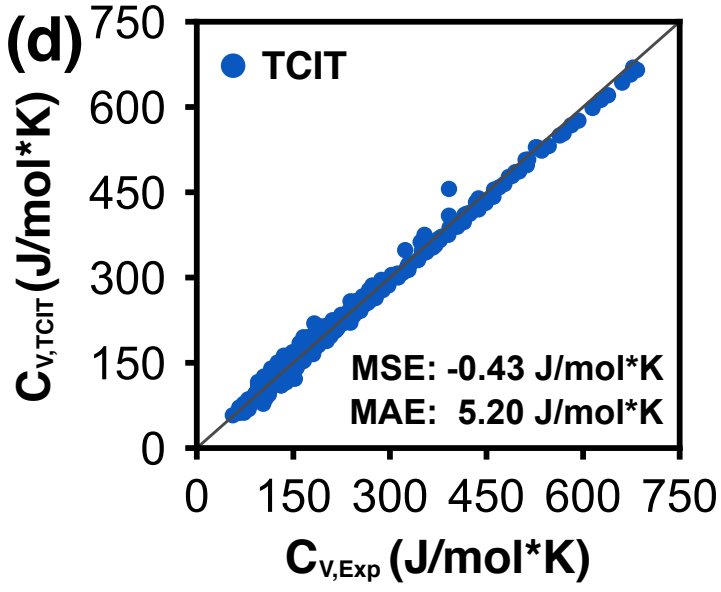
• The TCIT errors are consistent with error propagation of G4:exp and TCIT:G4 errors.

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.

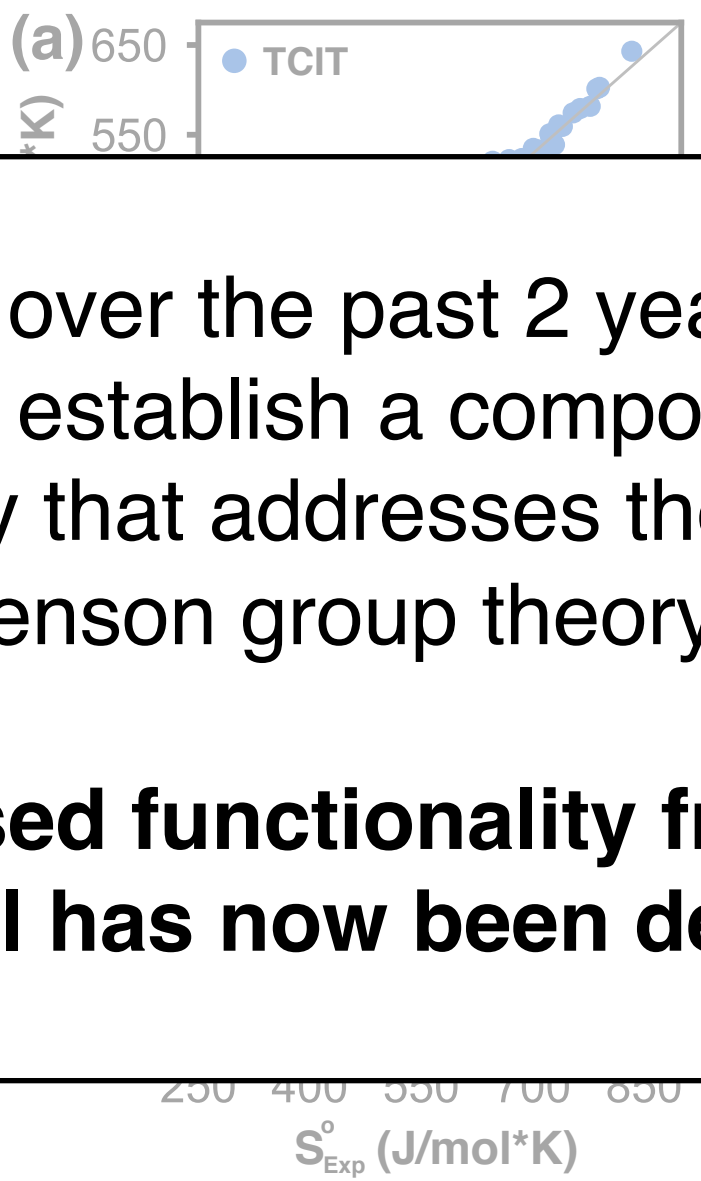
(b) G4/
medium

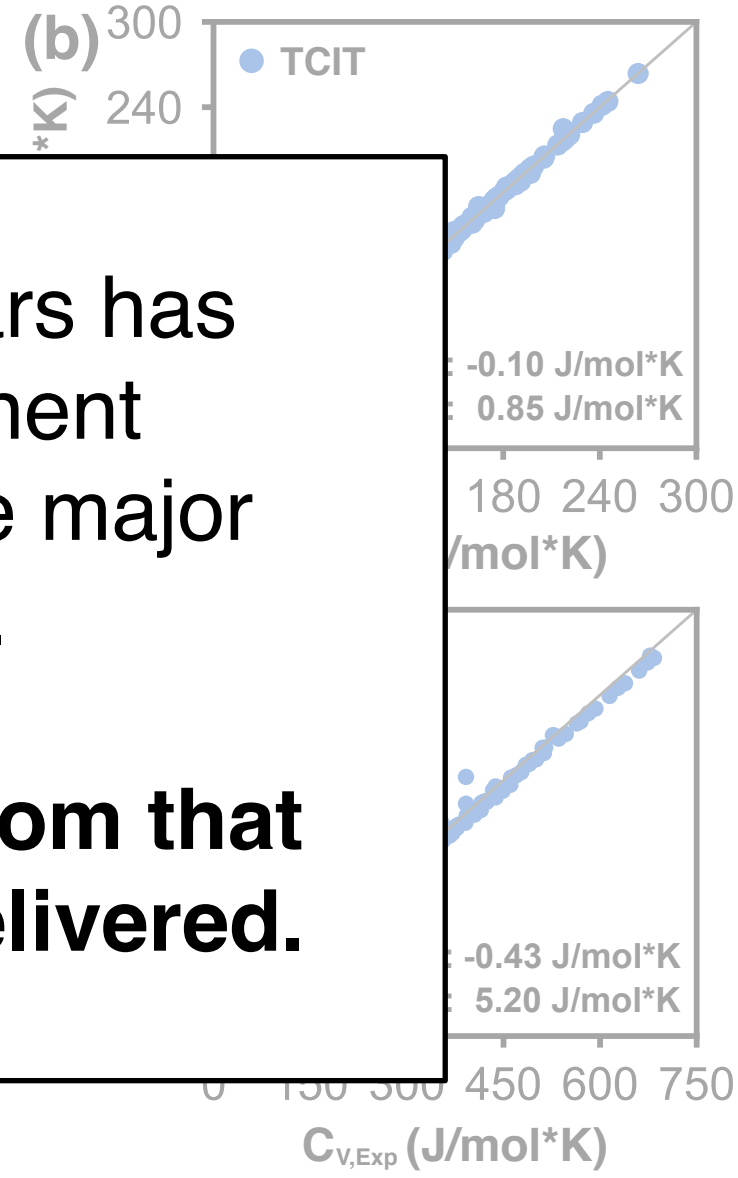
(c) TCI
molecul

(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 years has allowed us to establish a component increment theory that addresses the major gaps in Benson group theory.

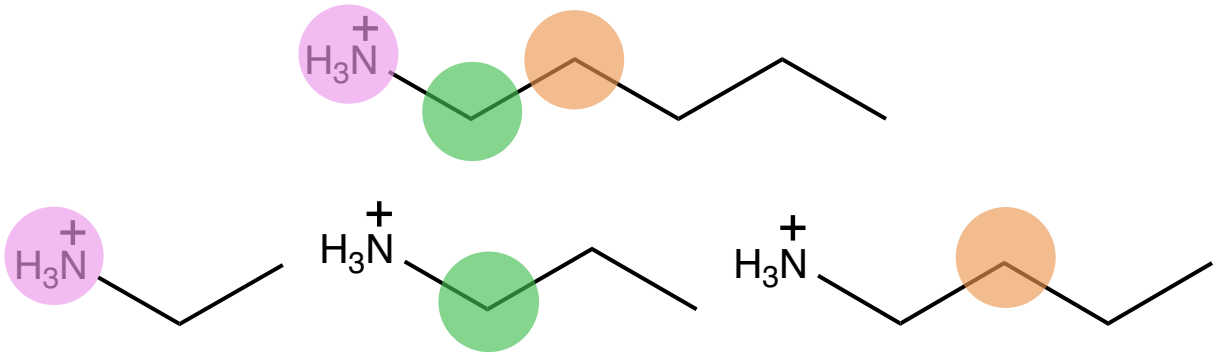
All of the promised functionality from that original proposal has now been delivered.

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

What's Next: Ions, Radicals, and other Properties

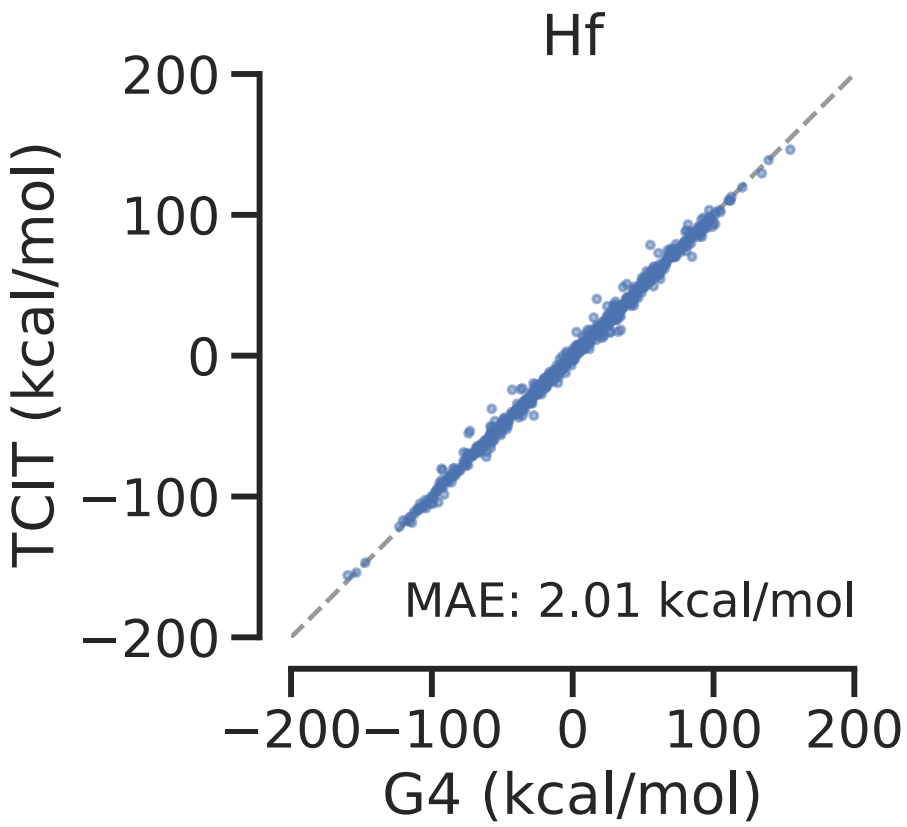
Ionic and radical species present acute challenges for experimental data collection and have limited representability within BGIT. **We propose to continue leveraging the intrinsic advantages of TCIT to address these gaps.**

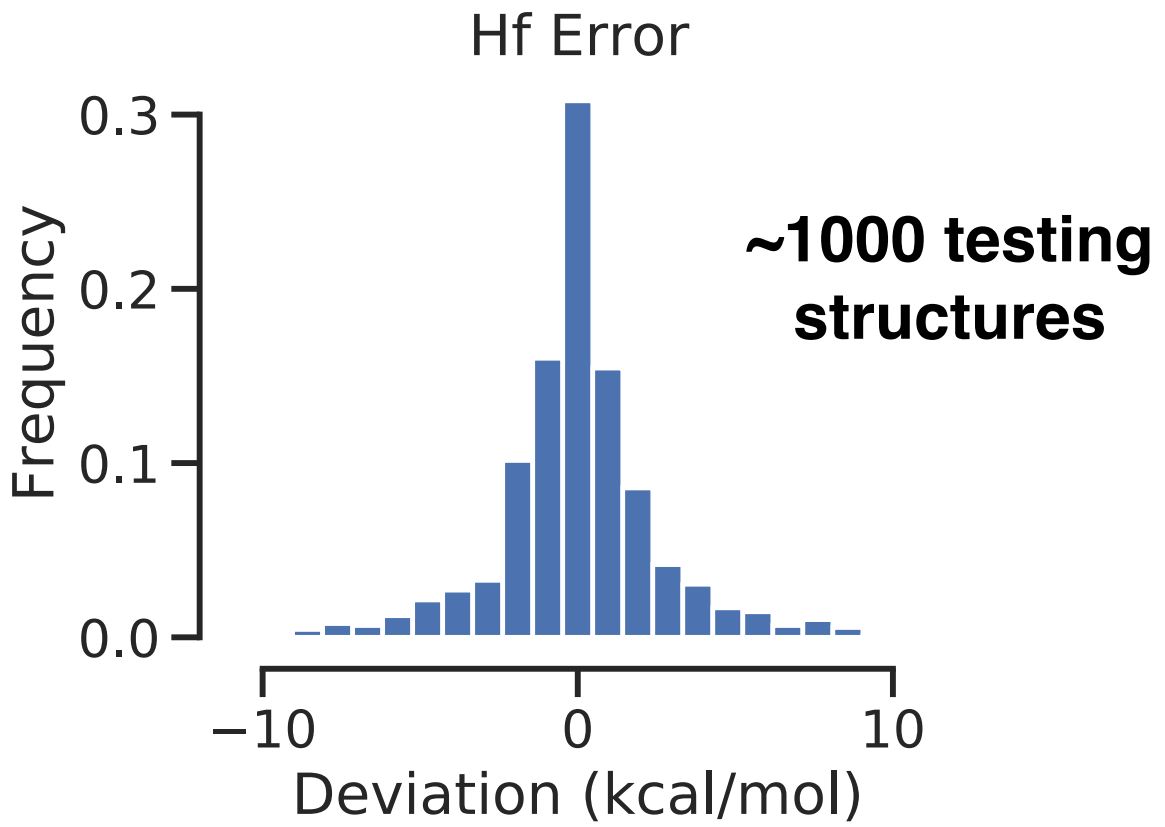
We are extending TCIT to cover ionic and radical species by propagating formal charge information through model compound generation and component definitions:



Because of the way that we have developed TCIT, the extension is relatively straightforward, it just requires the additional prediction of hydrogen-bond increments for conversion from the radical to the closed-shell analogue.

Preliminary TCIT Radical Benchmarks (ΔH_f)

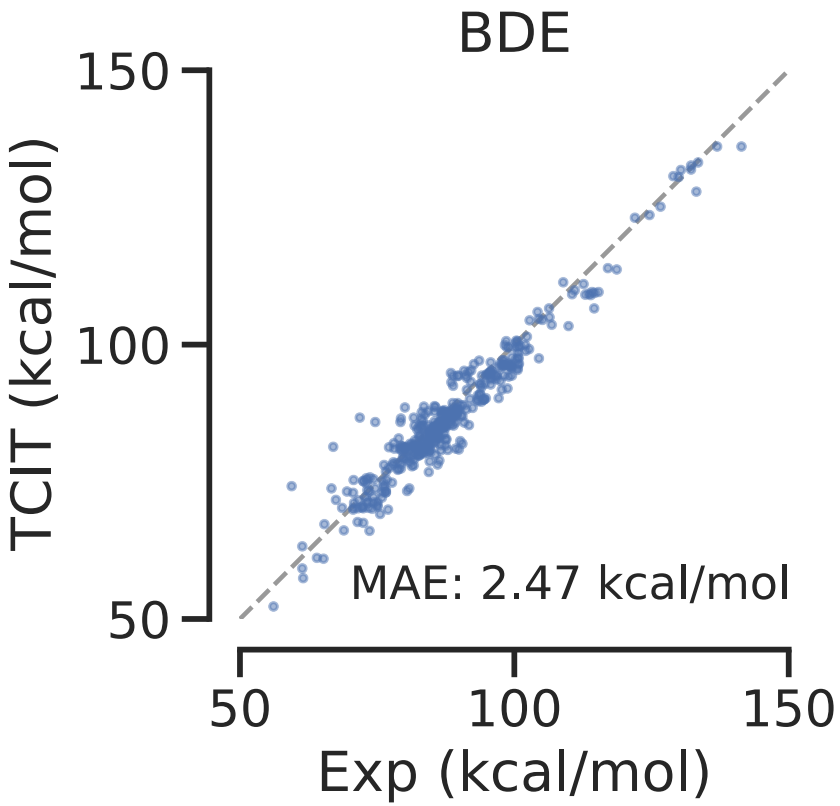


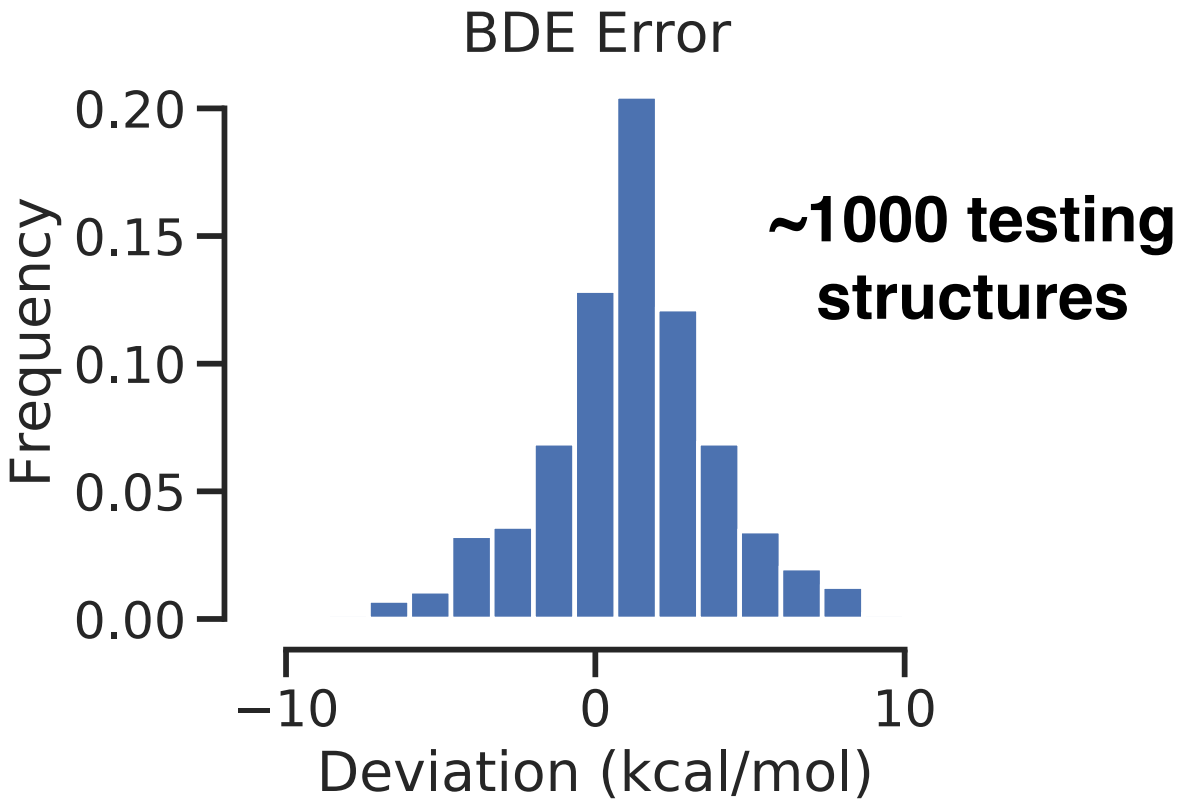


~1000 testing structures

- 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)

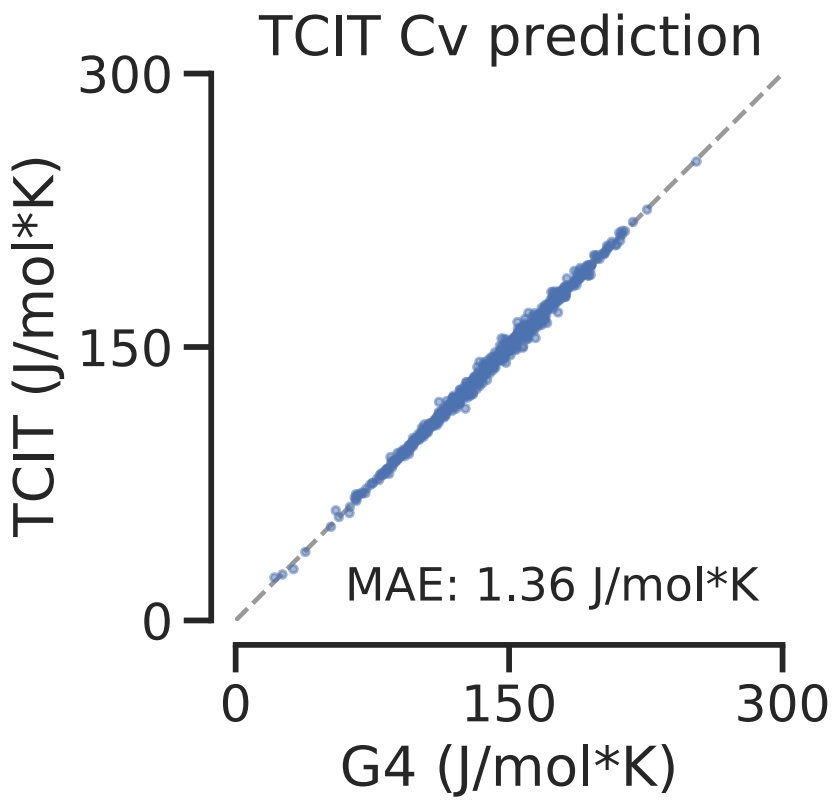


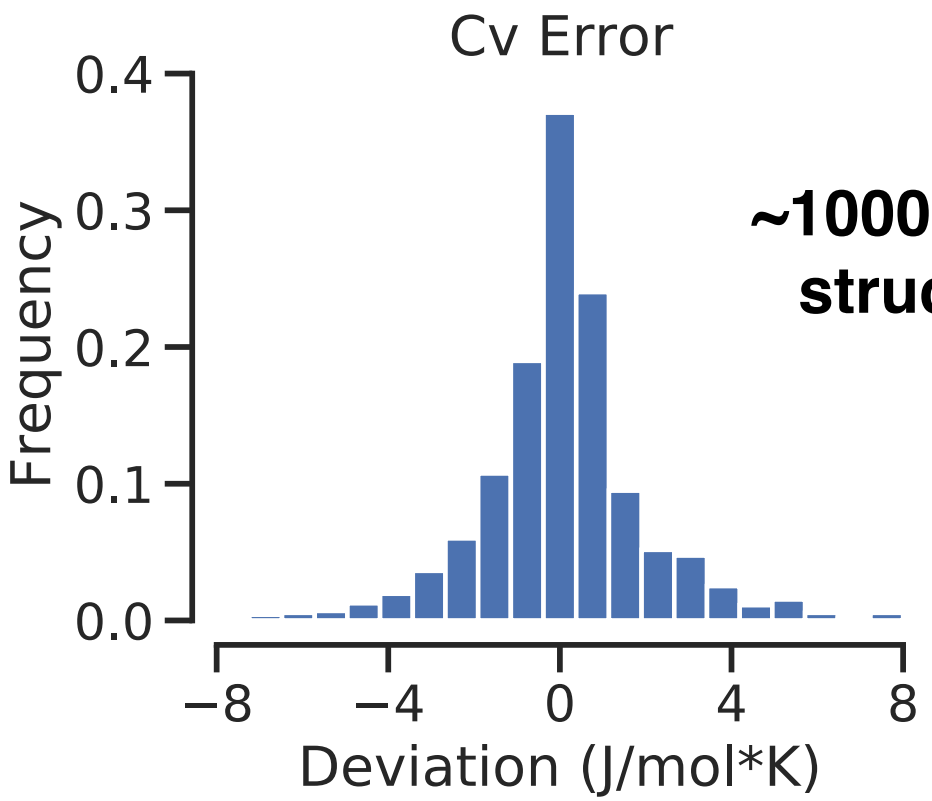


~1000 testing structures

- 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)

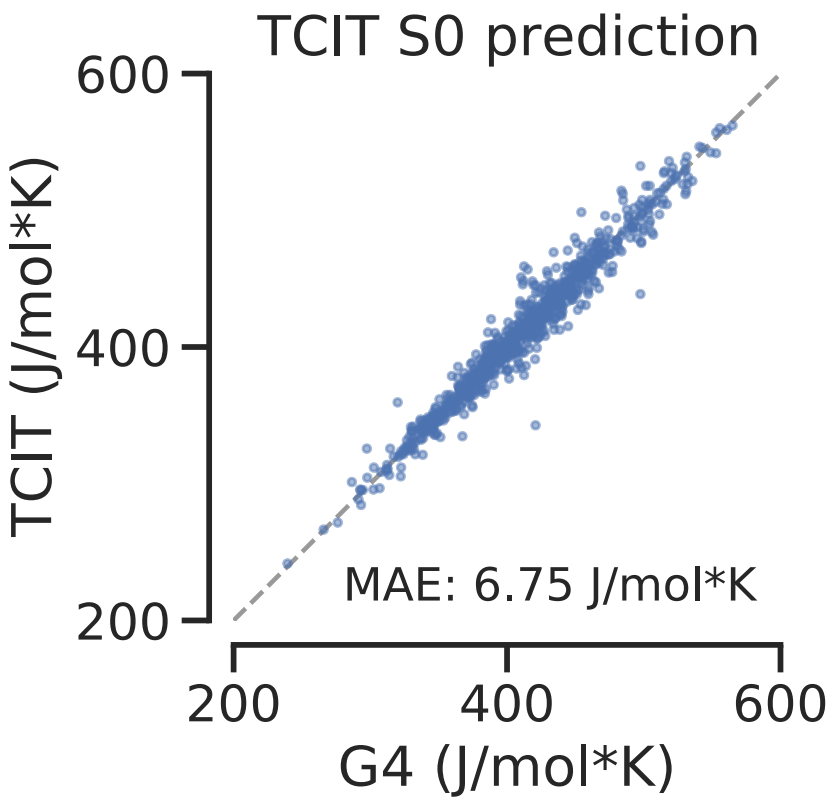


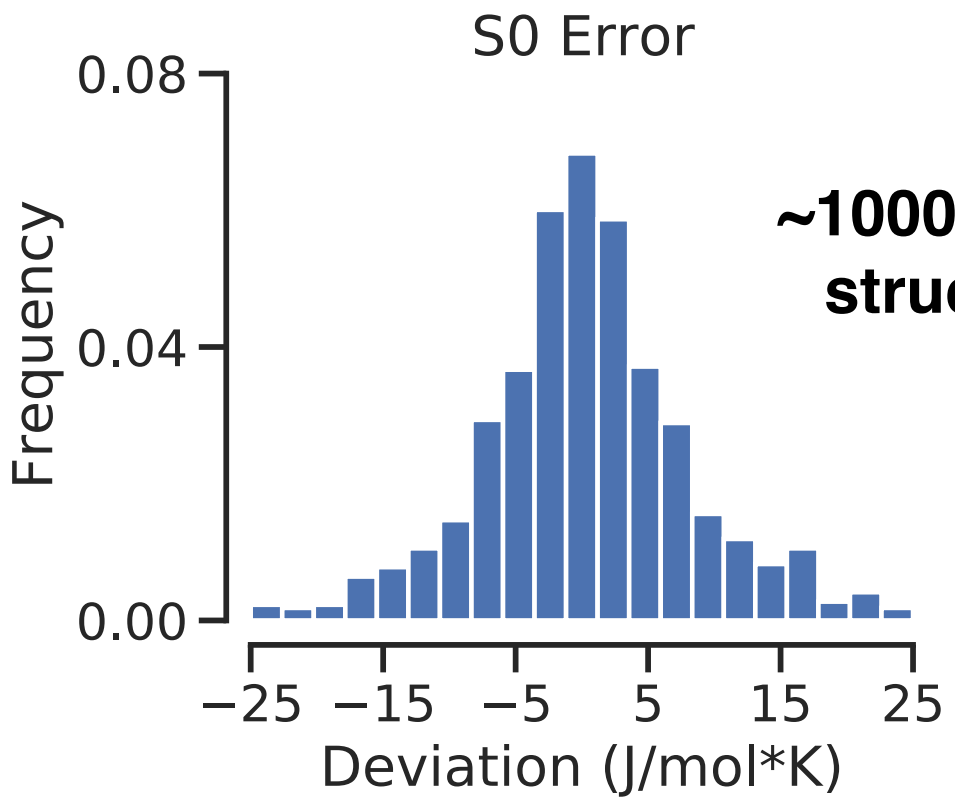


~1000 testing structures

- Similar accuracy compared with closed-shell species.

Preliminary TCIT Radical Benchmarks (S^o)





~1000 testing structures

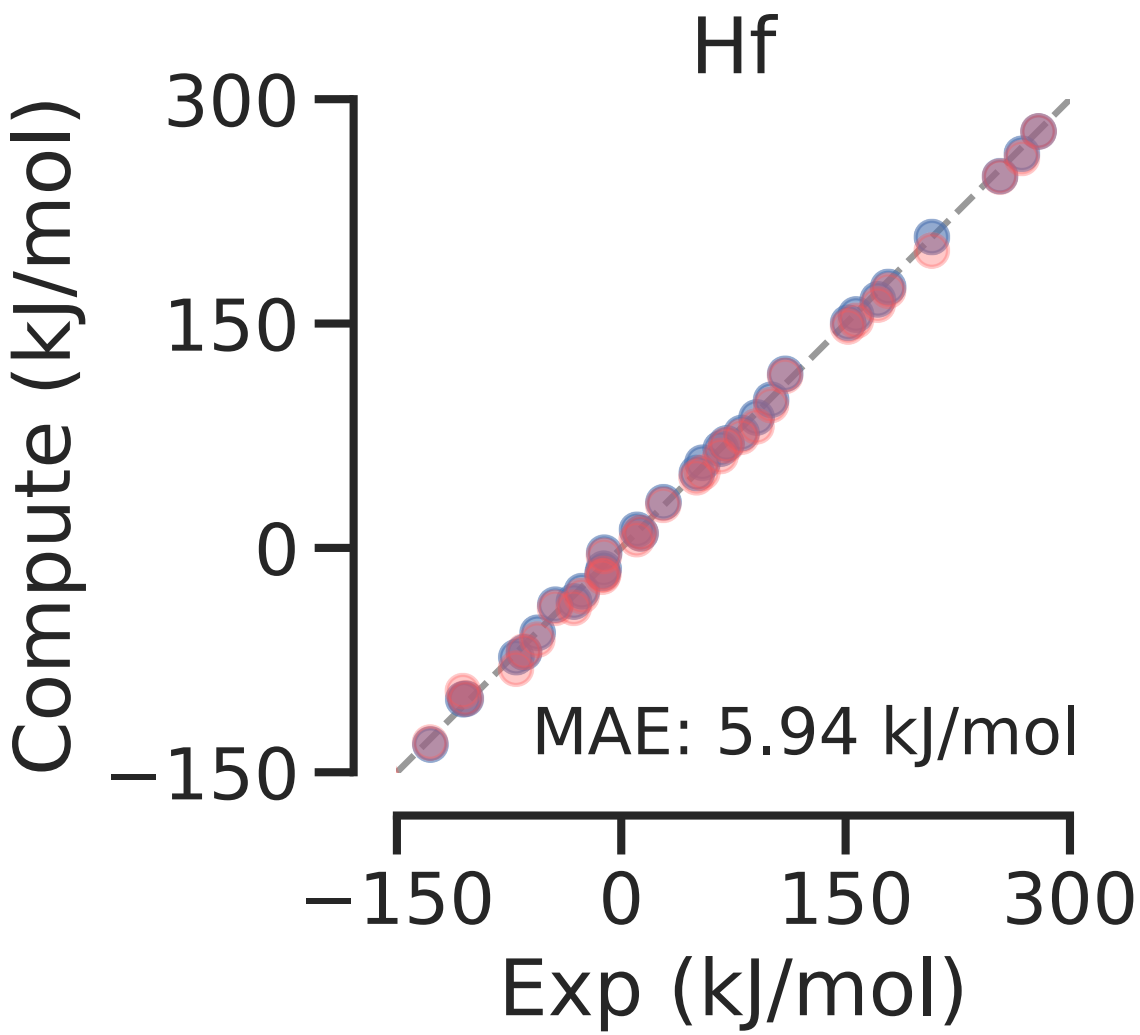
- Similar accuracy compared with closed-shell species.

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.

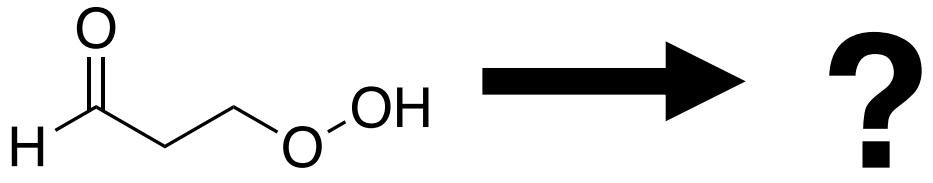
- This is a big victory for TCIT, since the limited experimental data for radical species makes CHETAH predictions impossible in many cases.



Trying to Solve the Reaction Prediction Problem

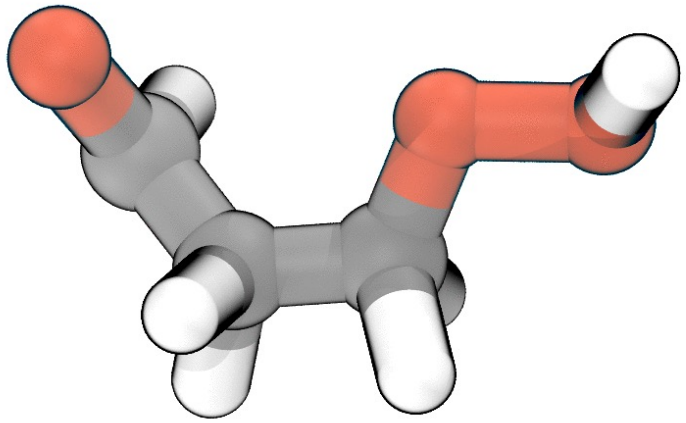
A → B : When we know the reactants and products, mature quantum chemistry tools exist to characterize transition states and establish pathways

A → ? : For many interesting (safety) applications, plausible reactions are unknown.



3-hydroperoxypropanal

→ **?**

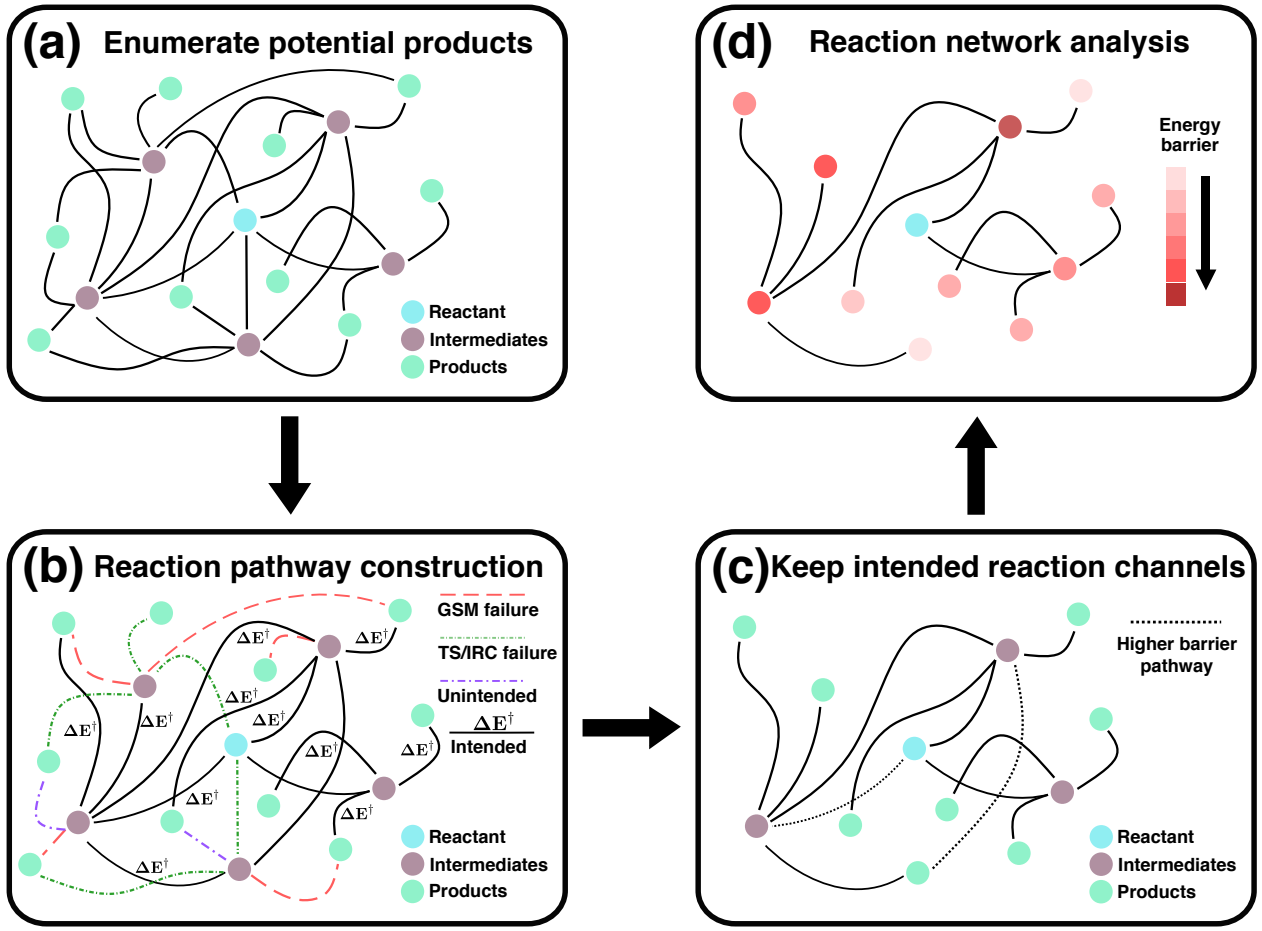


Yet Another Reaction Program (YARP)

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

Observations:

- **Product enumeration is easier than 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 can be leveraged here.**

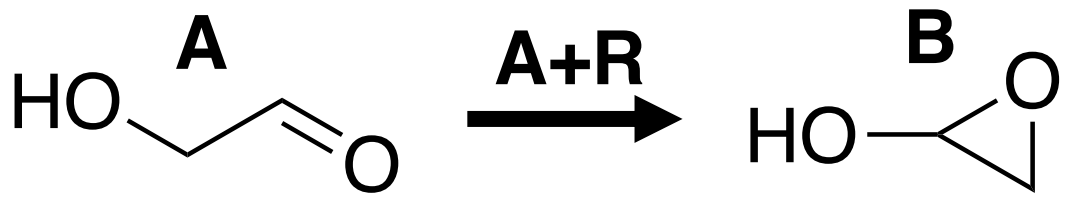


Qiyuan, Z.; Savoie, B. M. "More and Faster: Simultaneously Improving Reaction Coverage and Computational Cost in Automated Reaction Prediction Tasks." **(2020)** *ChemRxiv*. https://doi.org/10.26434/chemrxiv.13076087.v1

Converting Reactions into a Machine-Readable Grammar

Bond-Electron Matrix Formalism: matrix representation of molecules with bond order indicated in off-diagonal elements and lone electrons along the diagonal.

Ugi, I. et al. "New Applications of Computers in Chemistry." *Angew. Chem.* **1979**, 18 (2), 111–123.



$$\mathbf{R = B - A}$$

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

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

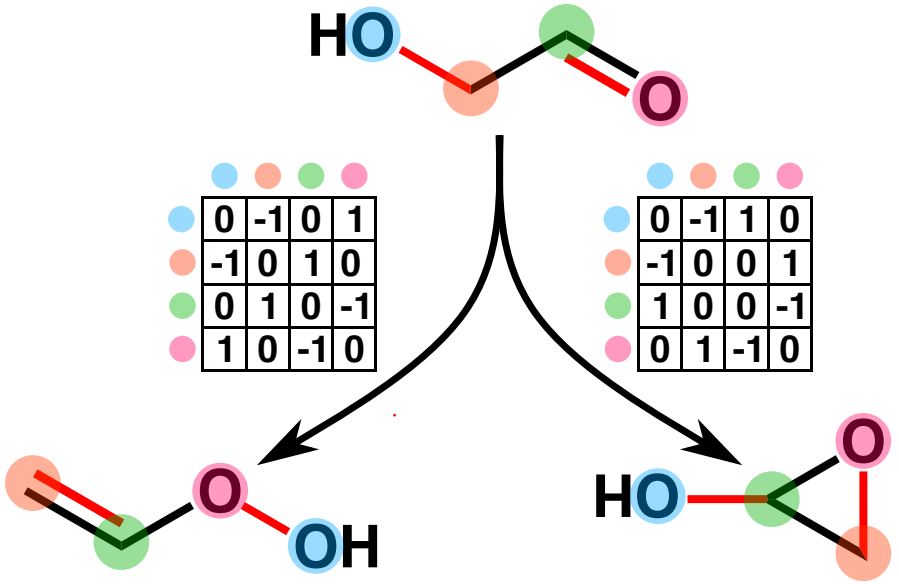
O	0	−1	1	0	0	0	0	0
C	−1	0	0	0	0	1	0	0
C	1	0	0	0	0	−1	0	0
H	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0
O	0	1	−1	0	0	0	0	0
H	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0

This is essentially a way of converting arrow pushing into something we can automate and interpret by a program

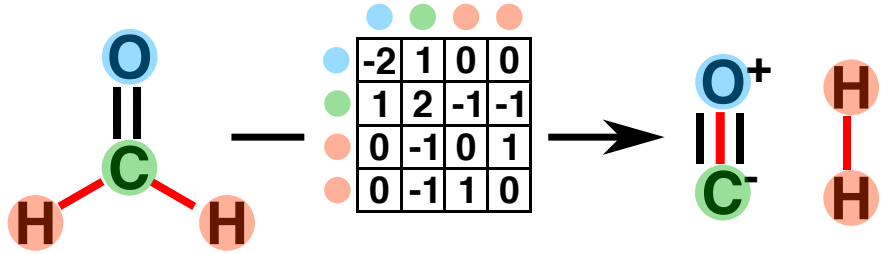
YARP: Elementary Reaction Step(s)

Elementary Reaction Step: a reaction with a single saddle point between products and reactants (we can't always know this in advance).

For closed-shell neutral organic molecules, "break 2 bonds form 2 bonds" is the simplest ERS that yields non-trivial closed-shell neutral products.



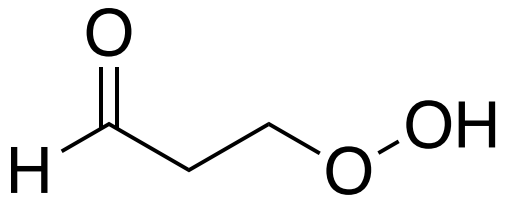
Example with concomitant e-transfer



Note: b3f3 and b4f4 products are decomposable into b2f2 ERS(s).

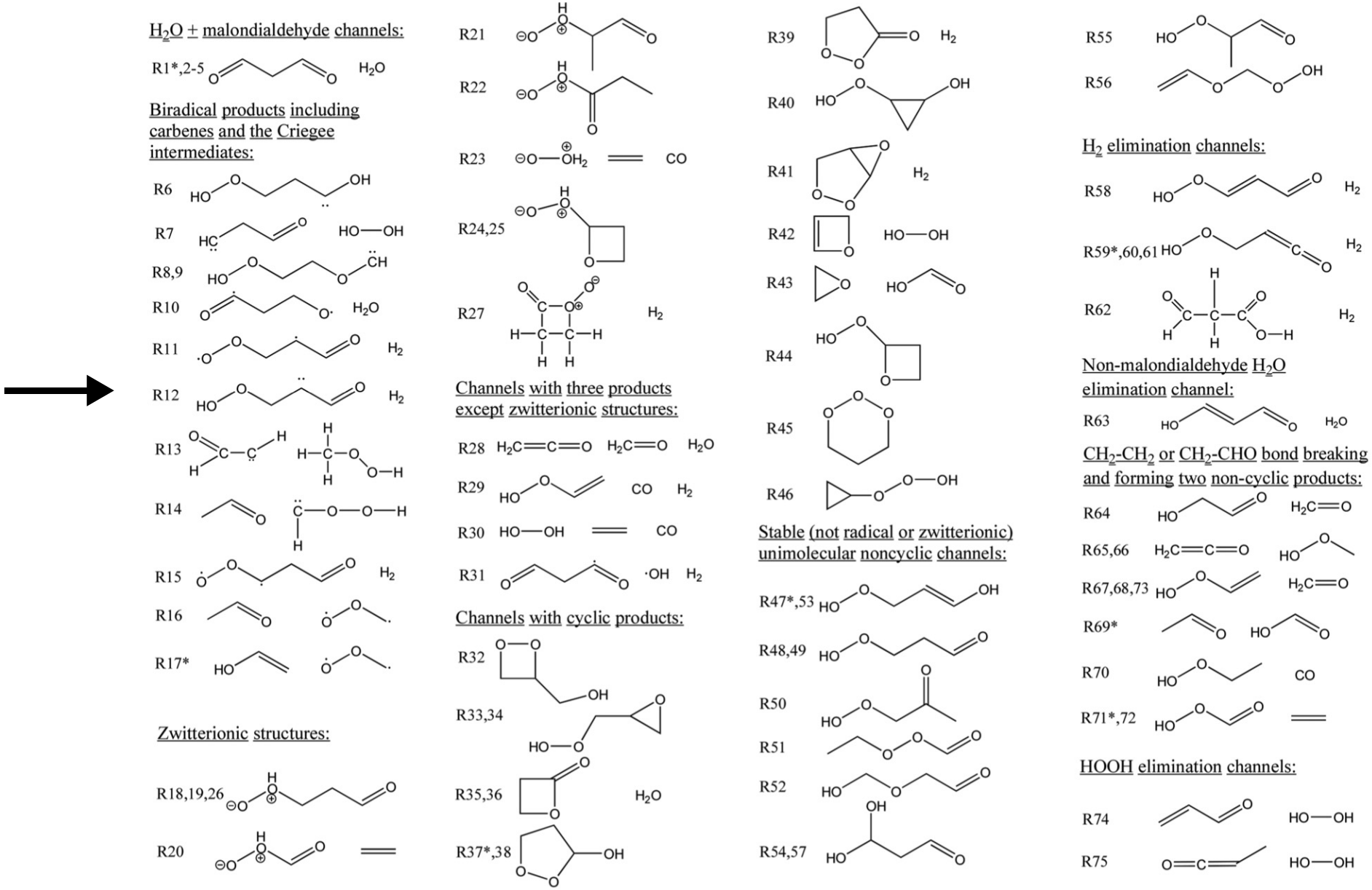
A more general suite of ERS(s) are certainly possible. These will certainly be required for radicals, organics with expanded octets, and metals.

Testing YARP on a Network Problem

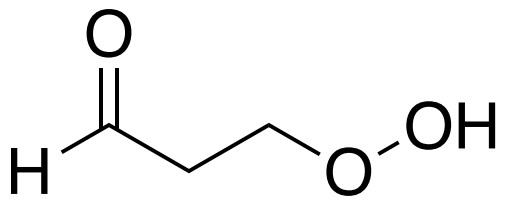


The 3-hydroperoxypropanal reaction network out to b4f4 was recently published as a benchmark for 5 reaction discovery methods.

Grambow, C. A, Suleimanov, Y. V. *JACS* **2018**, 140 (3), 1035–1048.



Testing YARP on a Network 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. *JACS* **2018**, 140 (3), 1035–1048.

55 distinct products
(neglecting charged species)

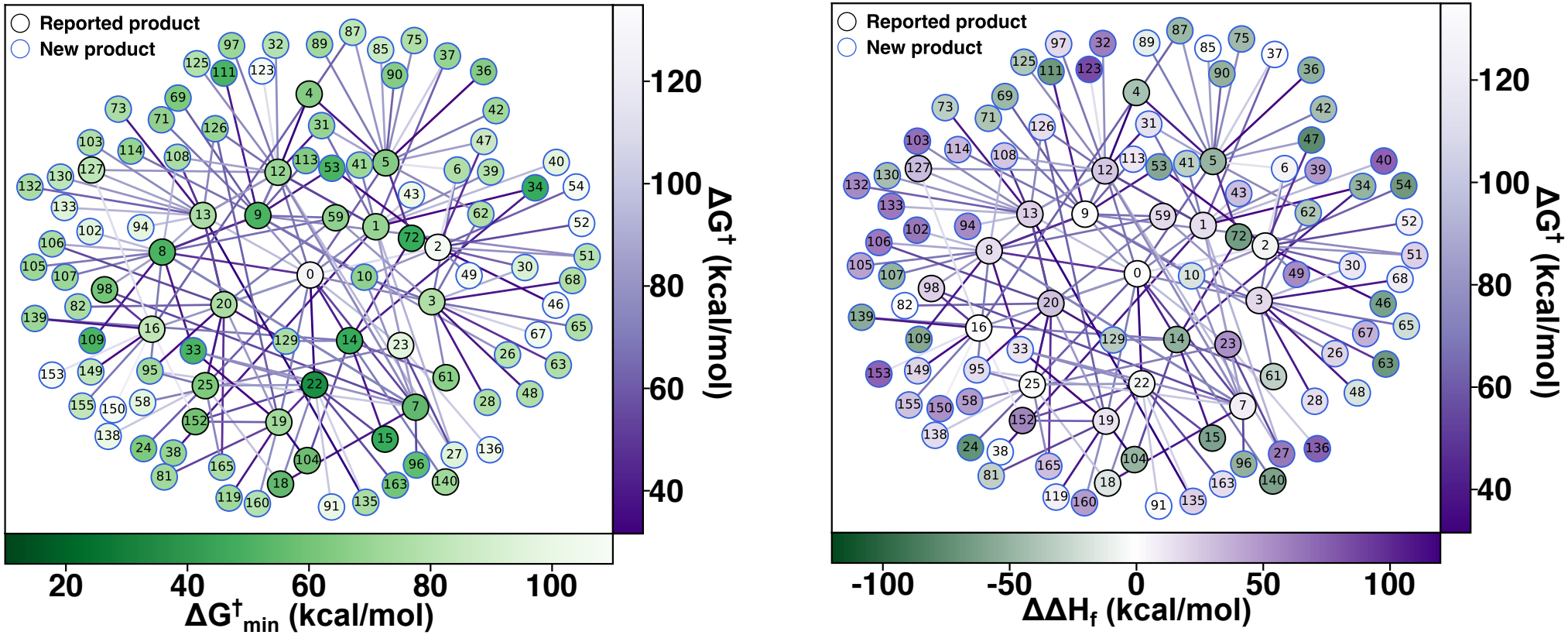
562 reactions attempted

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 all b2f2 products out to depth two, in order to achieve parity with Grambow et al.

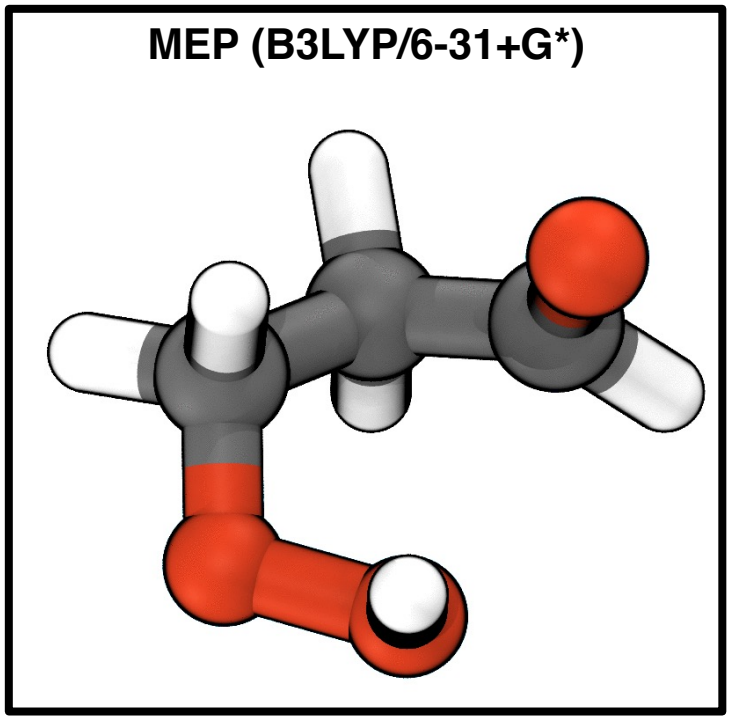


YARP finds all known products of this thermal decomposition network, as well as new products (77), and new reactions (157).

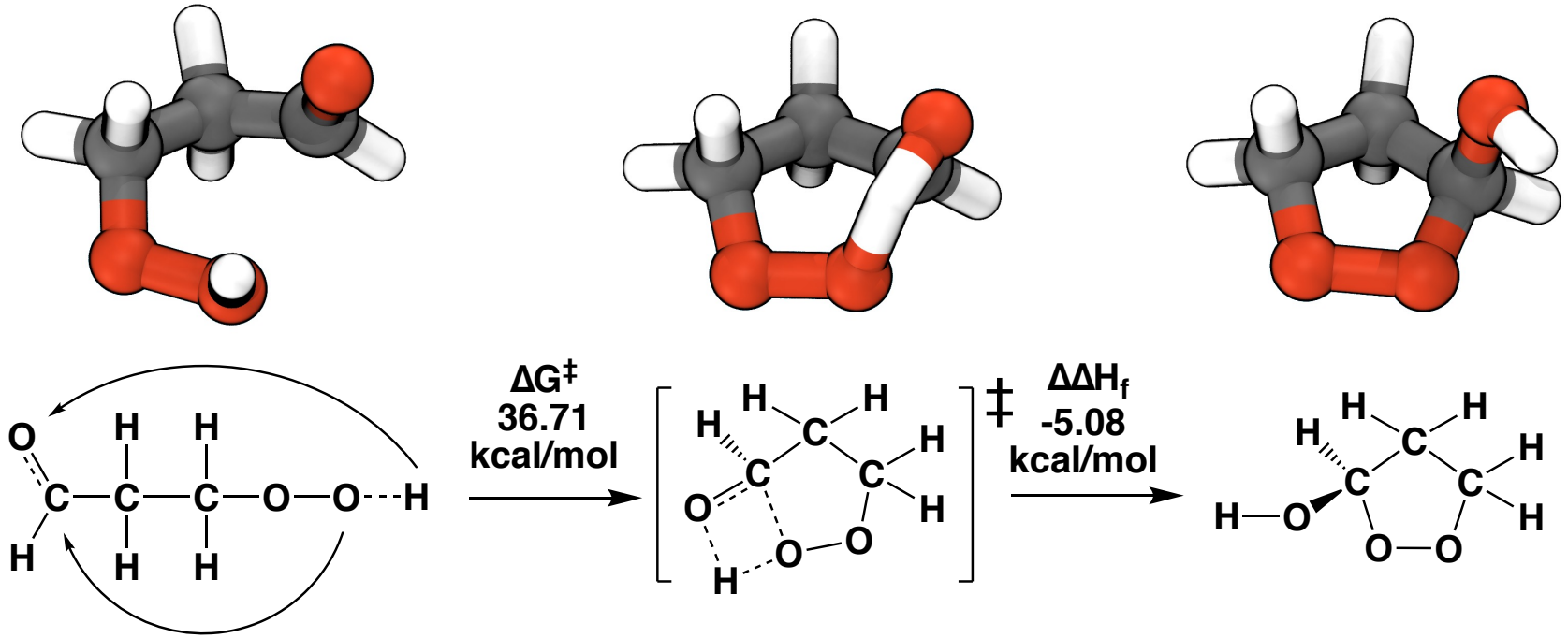
What Happens First?

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)

YARP needs to be extended:

(i) Broader suite of ERS(s) for organometallics and expanded octets

(ii) Validation of our treatment of ionic, radical, and resonance species

(iii) Kinetic modeling for prioritizing branches of deep networks

Qiyuan, Z.; Savoie, B. M. “More and Faster: Simultaneously Improving Reaction Coverage and Computational Cost in Automated Reaction Prediction Tasks.” **(2020)** *ChemRxiv*. https://doi.org/10.26434/chemrxiv.13076087.v1

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.

- P2SAC for funding.
- Ray Mentzer (Purdue)
- Madison Sprecher (Purdue UG)
- Caitlin Justice (Purdue UG)

P2SAC
Purdue Process Safety & Assurance Center