Extensive High-Accuracy Thermochemistry and Group Additivity Values for Halocarbon Combustion Modeling

David S. Farina Jr., Sai Krishna Sirumalla, Emily J. Mazeau, and

Richard H. West*

Department of Chemical Engineering, Northeastern University, Boston, MA 02115, USA

E-mail: r.west@northeastern.edu

Abstract

Standard enthalpies, entropies, and heat capacities are calculated for 16,813 halocarbons using an automated high-fidelity thermochemistry workflow. This workflow generates conformers at density functional tight binding (DFTB) level, optimizes geometries, calculates harmonic frequencies, and performs 1D hindered rotor scans at DFT level, and computes electronic energies at G4 level. The computed enthalpies of formation for 400 molecules show good agreement with literature references, but the majority of the calculated species have no reference in the literature. Thus, this work presents G4-computed thermochemistry for thousands of novel halocarbons. This new data set is used to train an extensive ensemble of group additivity values and hydrogen bond increment groups within the Reaction Mechanism Generator (RMG) framework. On average, the new group values estimate standard enthalpies for halogenated hydrocarbons within 3 kcal/mol of their G4 values. A significant contribution towards automated mechanism generation of halocarbon combustion, this research provides thermochemical data for thousands of novel halogenated species and presents a self-consistent set of halogen group additivity values.

This preprint was published as: "Extensive High-Accuracy Thermochemistry and Group Additivity Values for Halocarbon Combustion Modeling" David S. Farina Jr., Sai Krishna Sirumalla, Emily J. Mazeau, and Richard H. West *Industrial & Engineering Chemistry Research* 2021 60 (43), 15492–15501. DOI: 10.1021/acs.iecr.1c03076

Introduction

Halogenated Hydrocarbons (HHCs)

Halogenated hydrocarbons (HHCs) are commonly used as flame suppressants and refrigerant working fluids. The first generation of these compounds, chlorofluorocarbons (CFCs) and hydrochlorofluorocarbons (HCFCs), depleted the ozone layer and were banned worldwide under the Montreal Protocol in the 1980s.[1] The second generation, hydrofluorocarbons (HFCs), are ozone-friendly but are currently being phased out due to their high global warming potentials (GWPs).[2] Despite these controls on high-GWP HFC production, a recent study discovered that emissions of HFC-23 (CHF_3), a potent greenhouse gas, reached an historic high in 2018.[3]

To address these environmental concerns, several low-GWP HHC refrigerants and suppressants have been proposed. However, the chemical properties that make these HFCs more environmentally friendly also increase their flammability.[4] Therefore, the combustion properties of these proposed HHCs are of the utmost concern. Since experimental studies of these properties are complex and costly, predictive kinetic modeling of HHC combustion is crucial in screening proposed compounds in order to facilitate their innovation and implementation.

Understanding the complex chemistry of new compounds and predicting their combustion behavior under different conditions requires the compilation and simulation of detailed kinetic mechanisms (or microkinetic models) which often contain thousands of elementary reactions among hundreds of intermediate species. Building these models by hand is extremely challenging and error-prone due to the vast number of possible species and reactions to consider, sparse thermokinetic data available in the scientific literature, and biases of the human

choosing which pathways to select. Thus, a tool that generates these models automatically by enumerating and evaluating the many potential pathways by which HHCs combust will be instrumental in screening the flammabilty of greener refrigerants and suppressants.

Reaction Mechanism Generator

Reaction Mechanism Generator (RMG) is an open-source software package that automatically builds detailed kinetic models by proposing elementary reactions and estimating chemical properties (physical, thermochemical, kinetic, solvation, etc.) using a database of reaction templates, thermokinetic data, and estimation methods.[5,6] These chemical properties are first sought in a database of known parameters, but are more commonly estimated using hierarchical decision trees. Thermochemical parameters ($\Delta_f H^\circ_{298K}$, S°_{298K}, C_p^T) are usually estimated using Benson's group additivity method[7] for closed-shell species and the Hydrogen Bond Increment (HBI) scheme[8] for radicals. RMG's group additivity values are derived from high fidelity experimental data supplemented with high level quantum chemistry data.

The success of RMG's rate-based algorithm in generating reliable kinetic models that capture all the essential chemistry in complex reacting systems depends heavily on the accuracy of thermokinetic parameters. As RMG was originally developed to study the kinetics of hydrocarbon combustion, its databases contain extensive, although not exhaustive, thermokinetic data for CHO chemistry (molecules solely comprised of carbon, hydrogen, and oxygen). Since many short-lived intermediate species and elementary reactions are impossible to isolate and investigate experimentally, quantum chemistry methods are needed to calculate thermokinetic parameters. Recent progress on expanding RMG to model hydrocarbons substituted with nitrogen,[9] sulfur,[10] and silicon[11] has shown that quantum chemistry calculations are a viable approach to expand RMG's databases and estimation methods to new chemical systems. Thus, adding high-accuracy halogen thermochemical data to RMG's thermodynamic libraries and group additivity trees is essential to extending RMG to model halocarbon combustion.

Thermochemistry for HHCs

As kinetic modeling of HHCs has been been studied for many decades, thermochemical data for these molecules and their derivatives have accumulated through various theoretical investigations (Table 1). The Active Thermochemical Tables (ATcT)[12,13] and the Third Millennium Ideal Gas and Condensed Phase Thermochemical Database for Combustion[14,15] are two of the most expensive databases containing experimental and theoretical thermochemistry for HHCs. With advances in *ab-initio* quantum chemistry methods and codes, recent studies use high-level *ab initio* composite schemes and/or isodesmic reactions to calculate thermochemical properties such as $\Delta_f H^{\circ}_{298K}$ within or close to chemical accuracy (1 kcal/mol). Included in many of these studies are group additivity values (GAVs), in particular the works of Bozzelli and coworkers.[16–22] Since conventional group additivity does not work well for chlorocarbons, Chen and Bozzelli introduced a modified scheme which uses non-next-nearest neighbor interaction terms (NNNI) to account for the destabilizing interaction between electronegative chlorine atoms on adjacent carbons.[23] This modified NNNI-GAV scheme was later extended to fluorine-fluorine interactions in HFCs[16] and chlorine-oxygen interactions in chlorinated hydroperoxides.[17]

Although there are highly accurate thermochemical data in the literature, the data are very sparse and only cover a very small fraction of chemical space containing C, H, O, F, Cl, and Br chemistry. Since RMG explores a significant portion of this chemical space as it builds a mechanism, more HHC thermochemical data are needed to train RMG's group additivity schemes to improve its accuracy and reliability in estimating thermochemical properties. Thus, the focus of this work is not to more accurately calculate thermochemistry of previously studied molecules; instead, the objectives are: (1) to compute thermochemistry for a large and diverse set of halocarbons, and (2) derive an extensive ensemble of halogen GAVs with NNNIs within RMG's framework.

Table 1: Decades of quantum chemistry calculations for halocarbons. (GAVs- group additivity values, CC-high level coupled-cluster based model chemistry, NNNI-non-next-nearest neighbor interaction)

Year	Species	Methods	GAVs	Ref
1996	∼ 100 C1 and C2 HFCs	BAC-MP4		24
1998	fluorinated and chlorinated C1 and C2 species	G2, G2(MP2), CBS-Q, CBS-4		25
1999	2-fluoro-2-methylpropane	G2(MP2) (isodemic reactions)	C/C3/F (with NNNI)	16
2000	∼ 120 C1 and C2 HFCs	G3		26
2003	chloroethyl hydroperoxides	B3LYP/6-311+G(3df,-2p), QCISD(T)/6-31G(d,p), CBSQ//B3LYP/6-31G(d,p)	INT/OO/Cl, INT/OO/Cl2, INT/OO/CL3 NNNI terms	17
2009	fluorine and chlorine oxides	CC		27
2010	halomethane derivatives	CC		28
2014	$C_2H_xF_{6-x}$ (x=0–5)	CC		29
2015	14 C2–C4 HFCs	B3LYP, CBS-QB3, CBS-APNO, M06, M06-2X, ωB97X, G4, G4(MP2)-6X, W1U (isodesmic reactions)	C/C/F/H2, C/C2/F/H, C/C/F2/H, C/C2/F2, C/C/F3 (with NNNI)	18
2016	FO_xH and FO_xF (x=1–3)	ccCA, G3, G3B3, M06, M06-2X		30
2016	$CH_{3-x}F_xOH$ and $CH_{3-x}F_xOOH$ (x=1–3)	CBS-QB3, CBS-APNO, G4 (isodesmic reactions)	C/F/H2/O, C/F2/H/O, C/F3/O and HBI groups (with NNNI)	19
2017	C1–C2 HFC derivatives and radicals	CC		31
2018	65 C1 and C2 brominated halocarbons	G3, G4, ccCA, CCSD(T)/CBS		32
2018	fluorinated and chlorinated methane and ethane derivatives	diet-HEAT, diet-HEAT-F12		33
2019	organofluorine compounds	CCSD(T)/aug-cc-pvqz//B3LYP/def2-TZVP-D3(BJ)		34
2019	C2–C3 fluorinated aldehydes	CBS-APNO, CBS-4M, CBS-QB3, M06-2X, ωB97X, B3LYP, G2, G3, G4, W1U (isodesmic reactions)	CO/C/F, C/CO/F3, C/CO/F/H2, C/C/CO/F/H, C/C/CO/F2, C/C/CO/F/H (with NNNI)	20
2019	1,1,2-trifluoroethene and oxidized derivatives	B2PLYP, M06-2X, G3MP2B3, CBS-QB3, CBS-APNO (isodesmic reactions)	C/C/F2/OH, C/C/F/H/OH (with NNNI)	21
2020	dimethyl and ethyl methyl fluorinated ethers	MN15/cc-pVTZ, CBS-QB3, and CBS-APNO	C/C/F/H2, C/C/F/H/O, C/F/H2/O and HBI groups	22
2021	16,813 CHO-F,Cl,Br molecules (see Table 2)	G4//M062X-D3/jun-cc-pvtz	2,041 GAVs (see Table 4)	This work

Methods

The *enum-halocarb4* dataset

Due to a scarcity of thermochemical data for HHCs in the literature, we generated a new dataset called *enum-halocarb4*. In order to obtain high coverage and diversity of CHO-(F,Cl,Br) chemical space, this dataset was created by "halogenating" the complete chemical subspace for CHO containing up to 4 carbon and/or oxygen atoms, denoted $[CHO]_{1-4}$, as generated by Margraf et al.[35] Margraf defines the complete chemical subspace as a set of molecules such that breaking any bond in any molecule makes fragments that are also included in the set. $[CHO]_{1-4}$ was created by starting with CH_4, H_2O, and H_2, the smallest saturated closed-shell molecules containing C, O, and H, respectively. Then, the set was iteratively enlarged by replacing hydrogens with methyl ($-CH_3$) and hydroxyl ($-OH$) groups. Subsequently, rings were added, bond orders were incremented, and radicals were introduced via hydrogen abstractions. Along the way, duplicate strucutres were removed using canonical SMILES from RDKit.[36]

In this work, *enum-halocarb4* was created by systematically substituting halogen atoms (F,Cl,Br) for hydrogens in $[CHO]_{1-4}$ using RDKit. Molecules that contain more than one halogen element (CF_3Br for example) are referred to as "mixed-halogens" and molecules with only a single halogen element (CF_4 for example) are called "mono-halogens". To reduce the size of the dataset, *enum-halocarb4* was pruned by removing:

- radical species with more than one unpaired electron
- mixed-halogens with more than 8 heavy atoms
- cyclic mixed-halogens

The chemical composition of the 16,813 molecules in the *enum-halocarb4* is shown in Table 2. Since the dataset lacked singlet carbenes, which are more prominent in halocarbon combustion than in hydrocarbon combustion, 47 singlet carbenes (18 CHOF, 16 CHOCl, and 13 CHOBr) were generated manually and added to *enum-halocarb4* (see supplementary information).

Table 2: Composition of 16,813 molecules in *enum-halocarb4*

Composition	Noncyclic Closed	Noncyclic Radical	Cyclic Closed	Cyclic Radical	Total
CHOF	1,048	1,591	278	441	3,358
CHOCl	1,040	1,552	264	393	3,249
CHOBr	721	721	254	382	2,078
CHOFCl	852	923	0	0	1,775
CHOFBr	1,540	1,858	0	0	3,398
CHOClBr	932	1,108	0	0	2,040
CHOFClBr	447	468	0	0	915
Total	6,580	8,221	796	1,216	16,813

Thermochemistry Workflow

The automated thermochemistry workflow used to calculate high-level thermochemical parameters ($\Delta_f H^\circ_{298K}$, S°_{298K}, Cp°_{1000K}) of the *enum-halocarb4* dataset is shown in Figure 1. First, a SMILES representation of the molecule is used to generate a molecular graph of the species using RMG. Then, the molecule is embedded with RDKit. to create a 3D geometry. After embedding, conformers are investigated using the systematic conformer generation algorithm implemented in AutoTST.[37,38] This algorithm explores conformers by rotating dihedrals in 120° increments, varying cis/trans isomerism of double bonds, and alternating R/S sterochemistry for chiral centers. Conformers are optimized in ASE[39] using the dftb+ calculator[40] with the halorg-0-1 parameter set.[41] A unique set of conformers is gathered by discarding geometrically similar molecules with a root-mean-square deviation below 0.1 Å of a conformer in the unique set. Subsequently, to obtain a unique set of low-energy conformers, high energy conformers are discarded with electronic energies above 5 kcal/mol of the lowest energy conformer in the set. A maximum conformer cutoff of 10 was used in this work to limit the number of DFT calculations and save computational time.

These conformers are re-optimized with DFT in Gaussian 16[42] using M06-2X-D3/jun-cc-pvtz (the M06-2X functional with Grimme's D3 empirical dispersion[43] and the jun-cc-pvtz basis set[44]). Following all geometry optimizations, RMG's graph isomorphism algorithm is used to ensure that the optimized molecules match the corresponding input SMILES. If

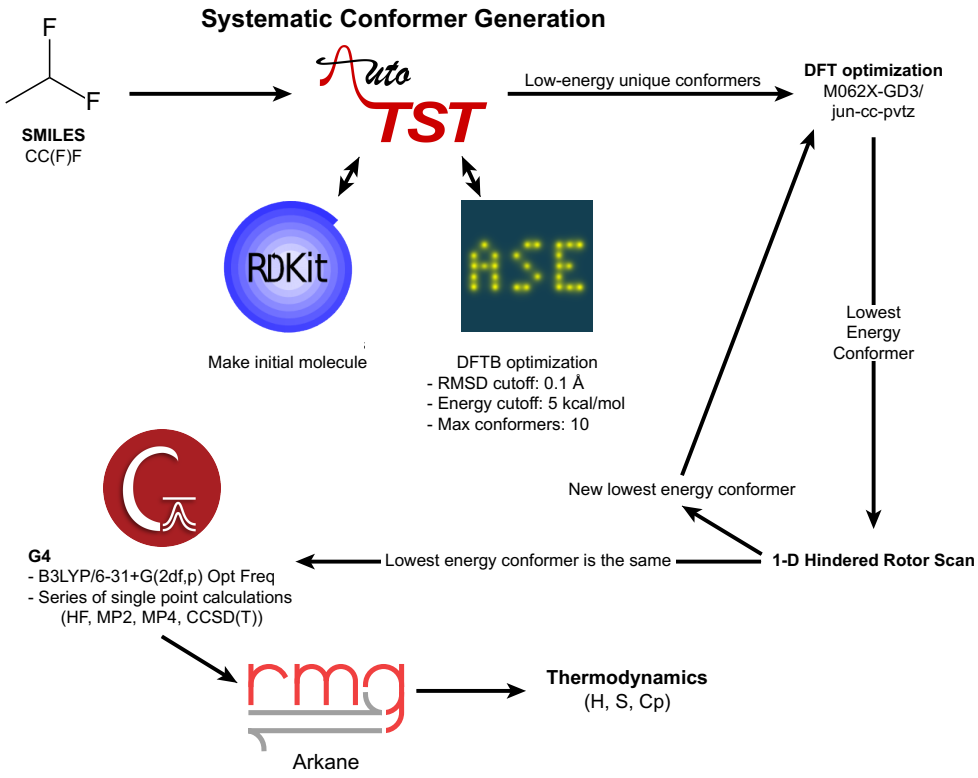


Figure 1: *ab-initio* Thermochemistry Workflow

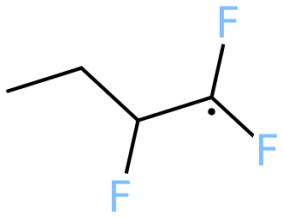
the optimized conformer is not isomorphic with the input molecular graph, the conformer is discarded from the set. If all of the optimized conformers were not isomorphic, that molecule was deemed unstable and removed from *enum-halocarb4*.

After identifying the lowest energy conformer with M06-2X-D3/jun-cc-pvtz, 1D hindered rotors calculations are performed in Gaussian 16 for molecules with internal rotors. For CHOF-containing molecules, the lowest energy conformer was optimized using B3LYP/6-31+G(2df,p) and rotor scans were performed in 10° increments with the same method. For molecules containing chlorine and bromine, the rotor scans were performed in 15° increments using M06-2X-D3/jun-cc-pvtz. If a lower energy conformer is found during the rotor scans, the conformer is reoptimized starting with the lower energy coordinates with DFT and the rotor scans are redone. If the rotor scan is discontinuous, the rotor scan is rerun with all of the torsions frozen in the molecule except for the scan torsion. If the rotor scan is still discontinuous, the RRHO approximation is used.

To obtain more accurate electronic energies, the lowest energy conformer is re-calculated with the Gaussian 4 (G4) compound method[45] in Gaussian 16. The G4 method was selected because it is an economical composite scheme that is accurate for halogenated hydrocarbons.[46–49] Lastly, to obtain thermochemistry as a seven coefficient NASA polynomial,[50] the G4 energies and harmonic frequencies and the DFT rotor scans are passed to RMG's statistical mechanics calculator Arkane,[51] a tool which uses quantum chemistry calculation results to compute thermochemical or kinetic parameters. For species with internal rotors, Arkane's best fit algorithm is used to determine whether a cosine function or a Fourier series, with 5, 7, or 9 terms, better fits the energy profile of the rotor scan.

Group Additivity Values

Table 3: Group additivity enthalpy estimate for $CH_3CH_2CHFCF_2$.



Group	$\Delta_f H^\circ_{298K}$ (kcal/mol)
Benson	
Cs-CsHHH	−10.2
Cs-CsCsHH	−4.9
Cs-CsCsFH	−55.2
Cs-CsFFH	−107.4
Long-Distance	
Cs(F)2-Cs(F)	+4.7
HBI	
CsjF2(Cs-CsFH)	+100.2
H atom	−52.1
Total	−124.9
G4	−126.2
Error	+1.3

The thermochemical data in the *enum-halocarb4* dataset was used as training data to fit 2,041 new halogen thermo group addivitity values (GAVs) in RMG. Five types of thermo

Table 4: Halogen GAVs derived in this work

Type	Groups
Benson	226
HBI	1,421
Ring Correction	247
Long-distance noncyclic	91
Long-distance cyclic	56
Total	2,041

additivity groups were derived: conventional nearest-neighbor Benson groups, hydrogen bond increment groups (HBI), cyclic and noncyclic non-next-nearest neighbor interaction groups (long-distance), and ring correction groups. Table 4 shows the number of groups derived for each group type, and Table 3 demonstrates how these groups combine to estimate the enthalpy of formation for a hydrofluorocarbon radical $CH_3CH_2CHFCF_2$ calculated in *enum-halocarb4*. *Cs* represents an sp^3 hybridized carbon with four single bonds, and *j* represents a radical electron.

The new Benson groups were systematically created by generating every possible nearest-neighbor group with either a carbon or oxygen center atom, bonded to at least one halogen atom. To maintain RMG's performance for estimating CHO thermochemistry, existing groups were held fixed while new halogen groups were fitted using scikit-learn's[52] linear least square ridge regression method with a regularization strength (alpha) of 1×10^{-6}. In total, 226 Benson groups and 91 long-distance interaction groups were derived using thermochemical data ($\Delta_f H^\circ_{298K}$, S°_{298K}, C_p^T) for 6,578 noncyclic closed-shell species in *enum-halocarb4*. To derive HBIs, reactions with closed-shell conjugates were created by saturating radicals with a hydrogen atom as shown in equation 1. Then, the HBIs were fitted by subtracting the thermochemical properties of the radical and hydrogen atom from the thermo of the saturated conjugate species as shown in equation 2. 3,825 radicals and their saturated closed-shell conjugates from *enum-halocarb4* were used to fit 76 nearest-neighbor HBIs. To improve the accuracy of the HBI scheme, 1025 second nearest-neighbor HBIs and 320 three and four member ring HBIs were derived using 6,064 noncyclic radicals and 717 cyclic radicals,

respectively.

For the radical $R\cdot$ the reaction

$$R\cdot + H\cdot \rightarrow RH \tag{1}$$

is used to define the HBI corrections as

$$X[HBI(R\cdot)] = X[RH] - X[R\cdot] - X[H\cdot] \tag{2}$$

where X is a thermochemical function $\Delta_f H^\circ_{298K}$, S°_{298K}, or C_p^T.

Cyclic groups were also included to account for ring strain in halogen-substituted rings. 247 cyclic corrections and 56 long-distance cyclic interactions were derived from 796 closed-shell three and four member ring species in *enum-halocarb4*. No corrections were included for ring sizes greater than four atoms since the maximum ring size in *enum-halocarb4* is four.

Test Set Generation

To assess the accuracy of RMG's new halogen thermo groups in estimating thermochemistry of intermediates created during automated generation of HHC combustion models, an RMG model was constructed for 2-Bromo-3,3,3-trifluoropropene (2-BTP) and CF_3Br in methane flames. Before generating a model, a literature mechanism for 2-BTP from NIST[53] was imported into RMG. In order to teach RMG how these two flame suppressants behave in hydrocarbon flames, 727 of the 1,610 reactions in the literature mechanism were added as training reactions to RMG's reaction families. Adding these reactions into training helps RMG generalize and improves rate estimates of reactions with similar functional groups. Then, an RMG model was built using the *Foundational Fuel Chemistry Model Version 1.0*[54] in RMG-database as a seed mechanism. *enum-halocarb4*, the thermochemical dataset gen-

erated in this work, was used as an RMG thermo library during 2-BTP model generation, should RMG need thermochemistry for an intermediate in that dataset; other HHCs were estimated using the new GAV scheme. Our description is brief because the goal here is not to produce a better kinetic model, but to generate a test-set of molecules from a realistic scenario of automated model generation with RMG. 104 HHCs from the resulting RMG mechanism that were not in the *enum-halocarb4* GAV training set were recalculated at G4 level using the automated thermochemistry workflow previously discussed, and the calculated thermochemical properties were compared to GAV estimates.

Results and discussion

Thermochemistry Workflow Benchmark

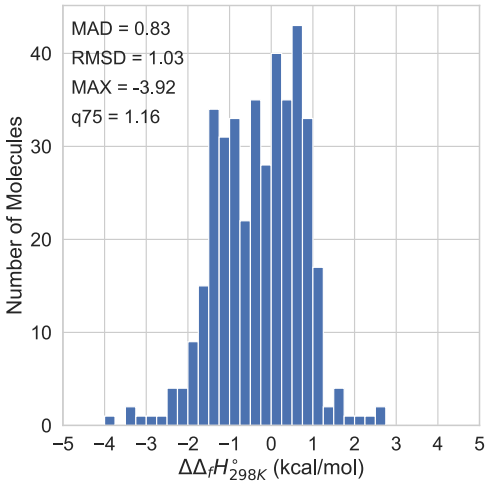


Figure 2: Enthalpies for halocarbons calculated with G4 show good agreement with reference values. (MAD = mean absolute deviation, RMSD = root-mean-square deviation, MAX = maximum deviation, q75 = 0.75 quantile of absolute error)

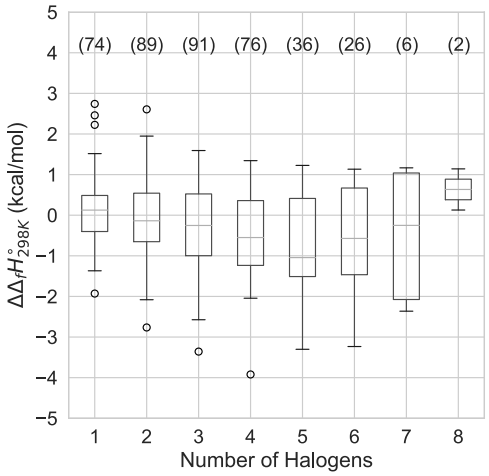


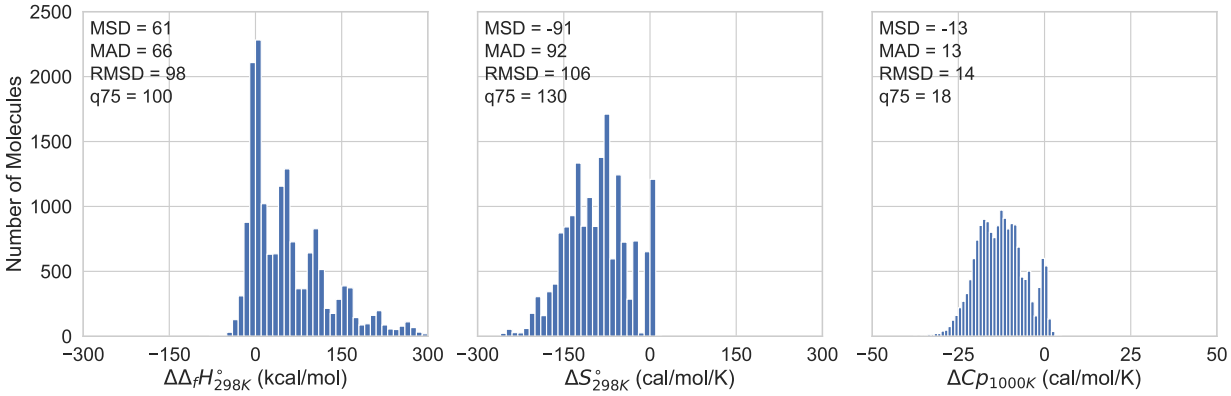
Figure 3: Error distribution by number of halogens in the molecule. G4 tends to slightly underestimate enthalpies of formation for molecules with more halogens. (Number in parentheses indicates the number of benchmark molecules with number of halogen atoms on x axis)

To evaluate the accuracy of our automated workflow in calculating $\Delta_f H^\circ_{298K}$ using the

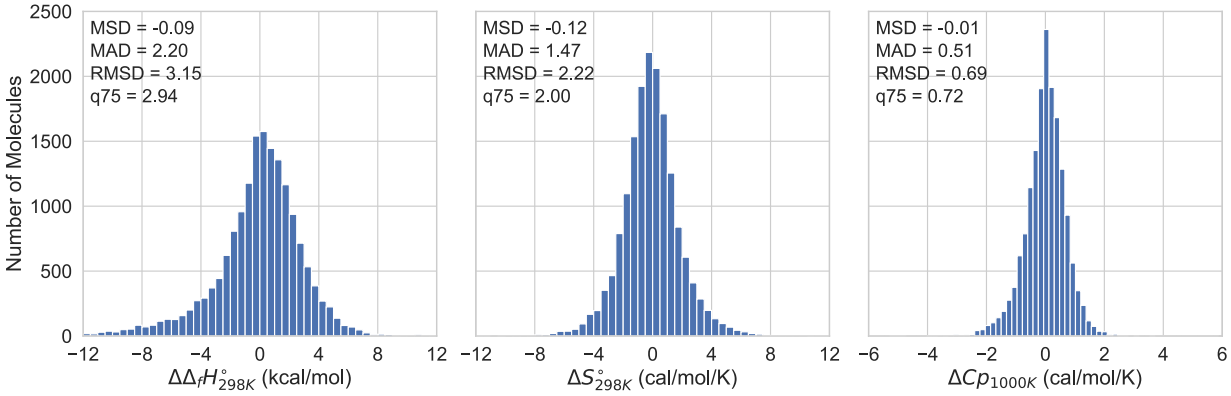
G4 method, 400 *enum-halocarb4* species were benchmarked against literature data. Figure 2 shows the distribution of the error for the benchmark set compared to reference $\Delta_f H^\circ_{298K}$ from the Active Thermochemical Tables (ATcT)[13] and various literature sources.[17–22,26–28,31–34,55] With a mean absolute error (0.83 kcal/mol) within chemical accuracy ($\leq$ 1 kcal/mol), G4 is a suitable, relatively low cost composite quantum chemistry method for high-fidelity and high-throughput calculations of HHCs. However, for heavily halogenated systems, G4 and other composite methods do not compute enthalpies within chemical accuracy.[56] Calculated G4 enthalpies for C_2Cl_5 and C_2Cl_6 in *enum-halocarb4* are more than 3 kcal/mol lower than ATcT values. Illustating this trend, Figure 3 shows that G4 tends to slightly underpredict enthalpies for molecules containing more halogen atoms in the benchmark set. Thus, it is likely that more heavily halogenated molecules in *enum-halocarb4* have higher errors than their more sparsely halogenated counterparts. To more accurately compute thermochemistry for heavily halogenated molecules, error-cancelling reactions, such as isodesmic reactions, could be used in place of the atomization approach used in this work. However, incorporating automated error-canceling reaction generation within our workflow was beyond the scope of the present work. The benchmark set is available in the supplementary material.

Group Additivity Values Performance

The new GAVs derived in this work dramatically improve RMG's estimates of $\Delta_f H^\circ_{298K}$, S°_{298K}, Cp°_{1000K} for 16,456 species in *enum-halocarb4*, as shown in Figure 4. Overall, the mean absolute deviation (MAD) of RMG's GAV estimates for $\Delta_f H^\circ_{298K}$ is reduced from 66 kcal/mol to 2.21 kcal/mol with the new groups, which is now on-par with RMG's CHO GAVs which typically estimate enthalpy within 2–3 kcal/mol. The new groups also significantly improve accuracy at estimating S°_{298K} (MAD reduced from 92 to 1.48 cal/mol/K) and Cp°_{1000K} (MAD reduced from 13 to 0.51 cal/mol/K). Based on the halogen GAV performance on the large and diverse *enum-halocarb4* dataset, RMG's new halogen GAVs appear to be reliable at accurately predicting thermochemical properties for a wide range of halocarbons.



(a) RMG GAV estimates for *enum-halocarb4* before new groups



(b) RMG GAV estimates for *enum-halocarb4* with new groups

Figure 4: New GAVs significantly improve RMG's estimates of $\Delta_f H^\circ_{298K}$, S°_{298K}, Cp°_{1000K} for HHCs in *enum-halocarb4* computed using G4 level of theory. (MSD = mean signed deviation, MAD = mean absolute deviation, RMSD = root-mean-square deviation, q75 = 0.75 quantile of absolute error)

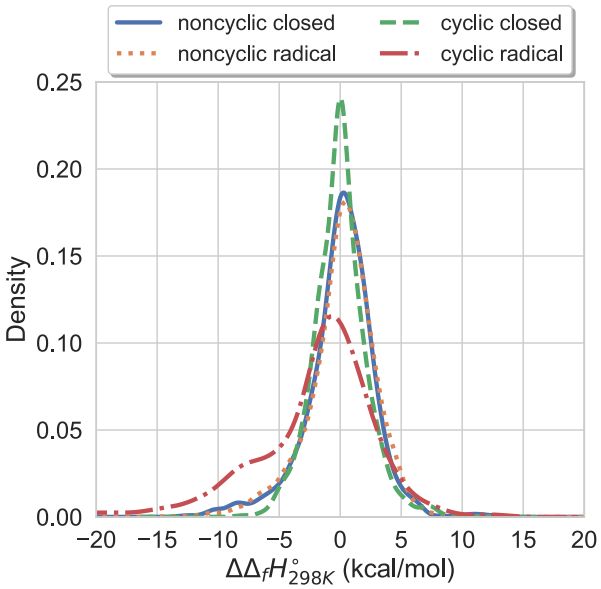


Figure 5: Probability density functions of GAV error for 4 classes of molecules in *enum-halocarb4*: noncyclic closed-shell, noncyclic radicals, cyclic closed-shell, cyclic radical molecules.

To further examine their accuracy and fidelity, the GAV performance with the new groups was evaluated for four different classes of molecules within the *enum-halocarb4* dataset: noncyclic closed-shell, noncyclic radical, cyclic closed-shell, and cyclic radical. Figure 5 shows the GAV performance for each molecule class. For noncyclic HHCs, which are a majority of the dataset ($\sim$ 88%), the GAVs are accurate at estimating thermochemical parameters for both closed-shell molecules and radicals. For non-cyclic closed-shell molecules, the MAD for $\Delta_f H^\circ_{298K}$ is 2.07 kcal/mol which is slightly less than the 2.13 kcal/mol MAD for noncyclic radicals. However, for cyclic species, the GAVs are significantly more accurate for cyclic closed-shell species (MAD = 1.56 kcal/mol) than for cyclic radicals (MAD = 3.75 kcal/mol). This disparity is most likely a result of only including closed-shell cyclic species in the derivation of the cyclic corrections and long-distance cyclic interaction groups. Therefore, since the radical cyclics are a test set for these groups whereas the closed-shell cyclics were included in training, it follows that the GAVs perform worse for the cyclic radicals.

Long-distance groups are included in the new GAV scheme to account for halogen/halogen

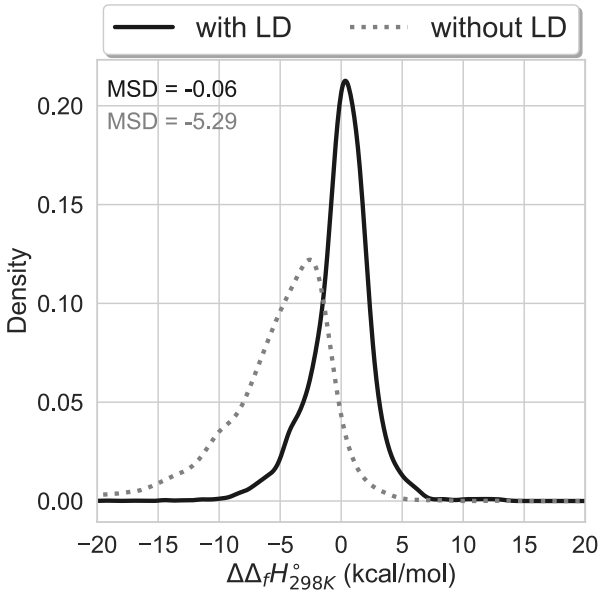


Figure 6: GAV enthalpy of formation estimates with long-distance groups (solid) and without long-distance interaction groups (dashed). Without long-distance groups, GAVs systematically underpredict enthalpies for HHCs.

and halogen/oxygen interactions on adjacent carbons. This modified group additivity approach was shown to significantly reduce errors in estimates of fluorinated and chlorinated hydrocarbons.[16,23] To investigate their impact in this work, the long-distance interactions were removed from the ensemble of GAVs, and estimated $\Delta_f H^\circ_{298K}$ without these groups were compared to estimates with the groups included for molecules in *enum-halocarb4* that have halogens on adjacent carbons. Figure 6 shows that, without the contributions from long-distance interactions, the GAVs systematically underestimate the $\Delta_f H^\circ_{298K}$ of these molecules by over 5 kcal/mol on average. In other words, GAVs overpredict stability of HHCs if lacking a long-distance term to capture the destabilizing interaction between halogens or oxygens on adjacent carbons. This indicates that these long-distance interactions are essential to accurately predict thermochemical properties of HHCs using a group additive scheme.

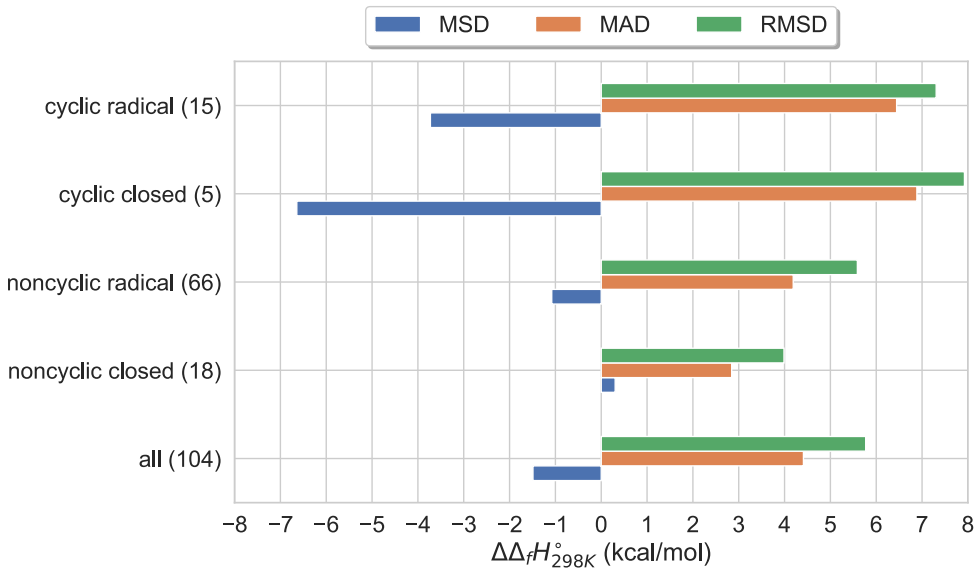


Figure 7: GAV $\Delta_f H^\circ_{298K}$ estimates for 104 fluorinated and/or brominated hydrocarbons in the 2-BTP RMG test set computed using G4 level of theory. The number of molecules in each category is indicated in parentheses. (MSD = mean signed deviation, MAD = mean absolute deviation, RMSD = root-mean-square deviation

RMG Test Set

As automated mechanism generation is the intended application, the new halogen GAVs were implemented in RMG, and RMG was used to construct a test set. Since RMG explores a wide variety of molecules as it builds a mechanism, RMG molecules provide a challenging test for GAVs and other graph-based molecular property estimators. The performance of the GAVs for the 2-BTP RMG test set is shown in Figure 7. Expectedly, the GAVs showed poorer performance for estimating $\Delta_f H^\circ_{298K}$ on the RMG test set with an overall mean absolute deviation of 4.69 kcal/mol compared to 2.21 kcal/mol on the *enum-halocarb4* training set. The GAVs performed well for noncyclic test set molecules with mean absolute deviations of 2.85 kcal/mol for closed-shell species and 4.19 kcal/mol for radicals. However, the GAVs performed relatively poorly for cyclic compounds, with mean absolute deviations of 6.89 kcal/mol for closed-shell molecules and 6.44 kcal/mol for cyclic radicals. The poorer performance for the cyclic species can mainly be attributed to the following three factors. First, due to a lack of training data, the GAVs do not include halogen-specific ring corrections for

ring sizes greater than 4 atoms. The mean absolute deviation for cyclic species with ring sizes greater than 4 was 7.42 kcal/mol compared to 6.09 kcal/mol for three and four member rings. Second, the long-distance corrections in RMG are not applied for neighboring atoms if one atom is in a ring and the other atom is outside the ring. A missing long-distance interaction between halogens on atoms in and out of a ring would lead to systematic underprediciton of $\Delta_f H^\circ_{298K}$ for these types of molecules, which may be reflected in the negative mean signed deviation for cyclic compounds in the test set. Third, because RMG was used to build the test set and RMG prefers to incorporate low energy molecules in its models (endothermic reactions are typically slower), there is a bias for underestimated (i.e. lower energy) molecules to be selected for the test set. In other words, it is more likely for molecules that the GAVs underestimated (negative MSD) to end up in the test set than overestimated or correctly estimated molecules. Although the current GAV estimates do not perform as well for cyclic halocarbons as noncyclic ones, the estimates are a vast improvement on RMG's estimates before this work and will help in the automated construction of microkinetic models.

Conclusions

This research provides thermochemical data for thousands of novel halogenated species and presents a comprehensive, self-consistent set of halogen group additivity values within the Reaction Mechanism Generator framework. The new GAVs accurately estimate enthalpies of formation for noncyclic closed-shell and radical species in *enum-halocarb4* and an RMG test set, but show poorer performance for rings for which more thermochemical training data are needed. Overall, the new halogen GAVs substantially improve RMG's thermo estimates for halocarbons, reducing the mean-absolute deviation of $\Delta_f H^\circ_{298K}$ for the *enum-halocarb4* dataset from 66 to 2.21 kcal/mol. Importantly, these new groups will enable rapid and accurate on-the-fly estimation of halocarbon thermochemistry during automated model generation, thereby improving the fidelity and reliability of RMG's halocarbon combustion models

for next-generation eco-friendly refrigerants and flame suppressants. These models can be used to determine important molecules for more rigorous thermochemical study. This work also contributes a new data set, *enum-halocarb4*, which provides essential thermochemical data in a sparsely populated region of chemistry for training other machine learning estimation methods We recommend a few areas of chemical space not thoroughly studied in this work for future thermochemistry measurements: triplet halocarbons, aromatic and cyclic halocarbons with rings larger than 4 atoms, and hypervalent chlorinated and brominated molecules.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. 1751720. The authors acknowledge Information Technology Services, Research Computing at Northeastern University for providing high performance computing and storage, related software, visualization, and consulting resources. The authors thank Professor Bozzelli for sharing work on fluorine thermochemistry and GAVs, and also thank Dr. Linteris and Dr. Burgess for providing their HFC and 2-BTP/CF_3Br kinetic models. Sai Krishna Sirumalla acknowledges funding from Entos Inc through LEADERs program at Northeastern University.

Supporting Information Available

The *enum-halocarb4* dataset and group additivity values are publicly available in RMG's open-source database (`https://github.com/ReactionMechanismGenerator/RMG-database`) and can also be accessed through the *molecule search* function on the RMG website (`https://rmg.mit.edu/molecule_search`).

References

(1) Montreal Protocol on Substances that Deplete the Ozone Layer Final Act 1987. *J. Environ. Law* **1989**, *1*, 128–136, DOI: 10.1093/jel/1.1.128.

(2) Harby, K. Hydrocarbons and their mixtures as alternatives to environmental unfriendly halogenated refrigerants: An updated overview. *Renew. Sust. Energ. Rev.* **2017**, *73*, 1247–1264, DOI: 10.1016/j.rser.2017.02.039.

(3) Stanley, K. M.; Say, D.; Mühle, J.; Harth, C. M.; Krummel, P. B.; Young, D.; O'Doherty, S. J.; Salameh, P. K.; Simmonds, P. G.; Weiss, R. F.; Prinn, R. G.; Fraser, P. J.; Rigby, M. Increase in global emissions of HFC-23 despite near-total expected reductions. *Nature Commun.* **2020**, *11*, 397, DOI: 10.1038/s41467-019-13899-4.

(4) Linteris, G.; Babushok, V. Numerically-Predicted Burning Velocities of C1 and C2 Hydrofluorocarbon Refrigerant Flames with Air. 17th International Refrigeration and Air Conditioning Conference at Purdue, West Lafayette, IN, 2018; https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=925585.

(5) Gao, C. W.; Allen, J. W.; Green, W. H.; West, R. H. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Comput. Phys. Commun.* **2016**, *203*, 212–225, DOI: 10.1016/j.cpc.2016.02.013.

(6) Liu, M.; Grinberg Dana, A.; Johnson, M. S.; Goldman, M. J.; Jocher, A.; Payne, A. M.; Grambow, C. A.; Han, K.; Yee, N. W.; Mazeau, E. J.; Blondal, K.; West, R. H.; Goldsmith, C. F.; Green, W. H. Reaction Mechanism Generator v3.0: Advances in Automatic Mechanism Generation. *J. Chem. Inf. Model.* **2021**, *61*, 2686–2696, DOI: 10.1021/acs.jcim.0c01480.

(7) Benson, S. W.; Cruickshank, F. R.; Golden, D. M.; Haugen, G. R.; O'Neal, H. E.;

Rodgers, A. S.; Shaw, R.; Walsh, R. Additivity rules for the estimation of thermochemical properties. *Chem. Rev.* **1969**, *69*, 279–324, DOI: 10.1021/cr60259a002.

(8) Lay, T. H.; Bozzelli, J. W.; Dean, A. M.; Ritter, E. R. Hydrogen Atom Bond Increments for Calculation of Thermodynamic Properties of Hydrocarbon Radical Species. *J. Phys. Chem.* **1995**, *99*, 14514–14527, DOI: 10.1021/j100039a045.

(9) Dana, A. G.; Buesser, B.; Merchant, S. S.; Green, W. H. Automated Reaction Mechanism Generation Including Nitrogen as a Heteroatom. *Int. J. Chem. Kinet.* **2018**, *50*, 243–258, DOI: 10.1002/kin.21154.

(10) Class, C. A.; Vasiliou, A. K.; Kida, Y.; Timko, M. T.; Green, W. H. Detailed kinetic model for hexyl sulfide pyrolysis and its desulfurization by supercritical water. *Phys. Chem. Chem. Phys.* **2019**, *21*, 10311–10324, DOI: 10.1039/C9CP00234K.

(11) Slakman, B. L.; Simka, H.; Reddy, H.; West, R. H. Extending Reaction Mechanism Generator to Silicon Hydride Chemistry. *Ind & Eng Chem Res* **2016**, *55*, 12507–12515, DOI: 10.1021/acs.iecr.6b02402.

(12) Ruscic, B.; Pinzon, R. E.; Laszewski, G. V.; Kodeboyina, D.; Burcat, A.; Leahy, D.; Montoy, D.; Wagner, A. F. Active Thermochemical Tables: Thermochemistry for the 21st century. *Journal of Physics: Conference Series* **2005**, *16*, 561–570, DOI: 10.1088/1742-6596/16/1/078.

(13) Ruscic, B.; Bross, D. H. Active Thermochemical Tables (ATcT) values based on ver. 1.122p of the Thermochemical Network (2020); available at ATcT.anl.gov. https://atct.anl.gov/ThermochemicalData/version1.122p/, Accessed: 2021–02-20.

(14) Burcat, A.; Ruscic, B. *Third Millenium Ideal Gas and Condensed Phase Thermochemical Database for Combustion with Updates from Active Thermochemical Tables*; 2005.

(15) Goos, E.; Burcat, A.; Ruscic, B. Extended Third Millennium Ideal Gas and Condensed Phase Thermochemical Database for Combustion. http://burcat.technion.ac.il, 2017; http://burcat.technion.ac.il.

(16) Yamada, T.; Bozzelli, J. W. Thermodynamic Properties Hf 298°, S°298, and Cp(T) for 2-Fluoro-2-MethyIpropane, Hf 298° of Fluorinated Ethanes, and Group Additivity for Fluoroalkanes. *Phys. Chem. A* **1999**, *103*, 7373–7379, DOI: 10.1021/jp9919516.

(17) Sun, H. Y.; Bozzelli, J. W. Structures, rotational barriers, and thermochemical properties of beta-chlorinated ethyl hydroperoxides. *Phys. Chem. A* **2003**, *107*, 1018–1024, DOI: 10.1021/Jp022298g.

(18) Wang, H.; Álvaro Castillo,; Bozzelli, J. W. Thermochemical Properties Enthalpy, Entropy, and Heat Capacity of C1-C4 Fluorinated Hydrocarbons: Fluorocarbon Group Additivity. *Phys. Chem. A* **2015**, *119*, 8202–8215, DOI: 10.1021/acs.jpca.5b03912.

(19) Wang, H.; Bozzelli, J. W. Thermochemical Properties and Bond Dissociation Energies for Fluorinated Methanol, CH3-xFxOH, and Fluorinated Methyl Hydroperoxides, CH3-xFxOOH: Group Additivity. *Phys. Chem. A* **2016**, *120*, 6998–7010, DOI: 10.1021/acs.jpca.6b05293.

(20) Purnell, D. L.; Bozzelli, J. W. Thermochemical Properties: Enthalpy, Entropy, and Heat Capacity of C2-C3 Fluorinated Aldehydes. Radicals and Fluorocarbon Group Additivity. *Phys. Chem. A* **2019**, *123*, 650–665, DOI: 10.1021/acs.jpca.8b09065.

(21) Snitsiriwat, S.; Yommee, S.; Bozzelli, J. W. Thermochemistry of Intermediates and Products in the Oxidation Reaction of 1,1,2-Trifluoroethene via OH Radical. *Phys. Chem. A* **2019**, *123*, 8017–8027, DOI: 10.1021/acs.jpca.9b06647.

(22) Hudzik, J. M.; Stoler, L. R.; Bozzelli, J. W.; Liebman, J. F. Thermochemistry of Fluorinated Dimethyl and Ethyl Methyl Ethers and Corresponding Radical Species. *Chem. Eng. Data* **2020**, *65*, 1594–1616, DOI: 10.1021/acs.jced.9b01034.

(23) Chen, C. J.; Wong, D.; Bozzelli, J. W. Standard chemical thermodynamic properties of multichloro alkanes and alkenes: A modified group additivity scheme. *Phys. Chem. A* **1998**, *102*, 4551–4558, DOI: 10.1021/jp980447i.

(24) Zachariah, M. R.; Westmoreland, P. R.; Burgess, D. R.; Tsang, W.; Melius, C. F. BAC-MP4 Predictions of Thermochemical Data for C_1 and C_2 Stable and Radical Hydrofluorocarbons and Oxidized Hydrofluorocarbons. *J. Phys. Chem.* **1996**, *100*, 8737–8747, DOI: 10.1021/jp952467f.

(25) Berry, R. J.; Schwartz, M.; Marshall, P. *Computational Thermochemistry*; Chapter 18, pp 341–358, DOI: 10.1021/bk-1998-0677.ch018.

(26) Haworth, N. L.; Smith, M. H.; Bacskay, G. B.; Mackie, J. C. Heats of formation of hydrofluorocarbons obtained by Gaussian-3 and related quantum chemical computations. *Phys. Chem. A* **2000**, *104*, 7600–7611, DOI: 10.1021/jp000475c.

(27) Karton, A.; Parthiban, S.; Martin, J. M. Post-CCSD(T) ab initio Thermochemistry of Halogen Oxides and Related Hydrides XOX, XOOX, HOX, XO, and HXO, (X = F, Xl), and Evaluation of DFT Methods for These Systems. *Phys. Chem. A* **2009**, *113*, 4802–4816, DOI: 10.1021/jp8087435.

(28) Csontos, J. J.; Rolik, Z. Z.; Das, S.; Kállay, M.; Kállay, M. High-Accuracy Thermochemistry of Atmospherically Important Fluorinated and Chlorinated Methane Derivatives. *Phys. Chem. A* **2010**, *114*, 13093–13103, DOI: 10.1021/jp105268m.

(29) Nagy, B.; Csontos, B.; Csontos, J.; Szakács, P.; Kállay, M. High-accuracy theoretical thermochemistry of fluoroethanes. *Phys. Chem. A* **2014**, *118*, 4824–4836, DOI: 10.1021/jp503492a.

(30) Alsunaidi, Z. H.; Wilson, A. K. DFT and ab initio composite methods: Investigation of oxygen fluoride species. *Comput. Theor. Chem.* **2016**, *1095*, 71–82, DOI: 10.1016/j.comptc.2016.08.024.

(31) Ádám Ganyecz,; Kállay, M.; Csontos, J. Accurate Theoretical Thermochemistry for Fluoroethyl Radicals. *Phys. Chem. A* **2017**, *121*, 1153–1162, DOI: 10.1021/acs.jpca.6b12404.

(32) Jorgensen, K. R.; Cadena, M. Theoretical study of bromine halocarbons: Accurate enthalpies of formation. *Comput. Theor. Chem.* **2018**, *1141*, 66–73, DOI: 10.1016/j.comptc.2018.08.016.

(33) Ganyecz, Á.; Kállay, M.; Csontos, J. High Accuracy Quantum Chemical and Thermochemical Network Data for the Heats of Formation of Fluorinated and Chlorinated Methanes and Ethanes. *Phys. Chem. A* **2018**, *122*, 5993–6006, DOI: 10.1021/acs.jpca.8b00614.

(34) Paulechka, E.; Kazakov, A. Critical Evaluation of the Enthalpies of Formation for Fluorinated Compounds Using Experimental Data and High-Level Ab Initio Calculations. *Chem. Eng. Data* **2019**, *64*, 4863–4874, DOI: 10.1021/acs.jced.9b00386.

(35) Margraf, J. T.; Reuter, K. Systematic Enumeration of Elementary Reaction Steps in Surface Catalysis. *ACS Omega* **2019**, *4*, 3370–3379, DOI: 10.1021/acsomega.8b03200.

(36) Landrum, G. RDKit: Open-source cheminformatics. http://www.rdkit.org, Accessed: 2021–01-29.

(37) Bhoorasingh, P. L.; Slakman, B. L.; Seyedzadeh Khanshan, F.; Cain, J. Y.; West, R. H. Automated Transition State Theory Calculations for High-Throughput Kinetics. *J. Phys. Chem. A* **2017**, *121*, 6896–6904, DOI: 10.1021/acs.jpca.7b07361, PMID: 28820268.

(38) Harms, N.; Underkoffler, C.; West, R. Advances in automated transition state theory calculations: improvements on the AutoTST framework. *ChemRxiv Preprint.* **2020-12-07**, DOI: 10.26434/chemrxiv.13277870.

(39) Larsen, A. H., *et al.* The atomic simulation environment—a Python library for working with atoms. *J. Condens. Matter Phys.* **2017**, *29*, 273002, DOI: 10.1088/1361-648X/aa680e.

(40) Aradi, B.; Hourahine, B. dftbplus/dftbplus [DFTB+]: Release 17.1. 2017; https://doi.org/10.5281/zenodo.809965, Accessed: 2020–05-29.

(41) Kubař, T.; Bodrog, Z.; Gaus, M.; Köhler, C.; Aradi, B.; Frauenheim, T.; Elstner, M. Parametrization of the SCC-DFTB method for halogens. *J. Chem. Theory Comput* **2013**, *9*, 2939–2949, DOI: 10.1021/ct4001922.

(42) Frisch, M. J., *et al.* Gaussian 16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.

(43) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *Chem Phys* **2010**, *132*, DOI: 10.1063/1.3382344.

(44) Papajak, E.; Zheng, J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions. *J. Chem. Theory Comput.* **2011**, *7*, 3027–3034, DOI: 10.1021/ct200106a.

(45) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 theory. *Chem Phys* **2007**, *126*, 084108, DOI: 10.1063/1.2436888.

(46) Somers, K. P.; Simmie, J. M. Benchmarking Compound Methods (CBS-QB3, CBS-APNO, G3, G4, W1BD) against the Active Thermochemical Tables: Formation Enthalpies of Radicals. *Phys. Chem. A* **2015**, *119*, 8922–8933, DOI: 10.1021/acs.jpca.5b05448, PMID: 26171842.

(47) Simmie, J. M.; Somers, K. P. Benchmarking Compound Methods (CBS-QB3, CBS-APNO, G3, G4, W1BD) against the Active Thermochemical Tables: A Litmus Test

for Cost-Effective Molecular Formation Enthalpies. *Phys. Chem. A* **2015**, *119*, 7235–7246, DOI: 10.1021/jp511403a.

(48) Dávalos, J. Z.; Notario, R.; Cuevas, C. A.; Oliva, J. M.; Saiz-Lopez, A. Thermochemistry of halogen-containing organic compounds with influence on atmospheric chemistry. *Comput. Theor. Chem.* **2017**, *1099*, 36–44, DOI: 10.1016/j.comptc.2016.11.009.

(49) Das, S. K.; Chakraborty, S.; Ramakrishnan, R. Critical benchmarking of popular composite thermochemistry models and density functional approximations on a probabilistically pruned benchmark dataset of formation enthalpies. *J. Chem. Phys.* **2021**, *154*, DOI: 10.1063/5.0032713.

(50) Gordon, S.; McBride, B. J. Computer Program for Calculation of Complex Chemical Equilibrium Composition, Rocket Performance, Incident and Reflected Shocks and Chapman-Jouguet Detonations. 1976; http://hdl.handle.net/2060/19780009781, Technical Report NASA-SP-273.

(51) Arkane: Automated Reaction Kinetics and Network Exploration. https://reactionmechanismgenerator.github.io/RMG-Py/users/arkane/, Accessed: 2019–01-29.

(52) Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(53) Burgess, D. R.; Babushok, V. I.; Linteris, G. T.; Manion, J. A. A Chemical Kinetic Mechanism for 2-Bromo-3,3,3-trifluoropropene (2-BTP) Flame Inhibition. *Int. J. Chem. Kinet.* **2015**, *47*, 533–563, DOI: 10.1002/kin.20923.

(54) Smith, G. P.; Tao, Y.; Wang, H. Foundational Fuel Chemistry Model Version 1.0. 2016; http://nanoenergy.stanford.edu/ffcm1, Accessed: 2021–03-20.

(55) Manion, J. A. Evaluated Enthalpies of Formation of the Stable Closed Shell C1 and C2 Chlorinated Hydrocarbons. *J. Phys. Chem. Ref. Data* **2002**, *31*, 123–172, DOI: 10.1063/1.1420703.

(56) Karton, A.; Sylvetsky, N.; Martin, J. M. W4-17: A diverse and high-confidence dataset of atomization energies for benchmarking high-level electronic structure methods. *Comp. Chem.* **2017**, *38*, 2063–2075, DOI: 10.1002/jcc.24854.