

Building Towards the Future in Chemical and Materials Simulation with Accessible and Intelligently Designed Web Applications

Eric J. Bylaska¹0000-0001-6405-599X, Duo Song¹, Eugene S. Ilton¹, Shaun O’Leary², Tiffany L. Torralba-Sánchez³, and Paul G. Tratnyek³

¹ Physical Sciences Division, Pacific Northwest National Laboratory
Richland, WA 99354, USA

² Research Computing Division, Pacific Northwest National Laboratory
Richland, WA 99354, USA

³ OHSU-PSU School of Public Health (Mailcode HRC-3), Oregon Health Science University (OHSU),
3181 SW Sam Jackson Park Road
Portland, Oregon 97239-3098, USA
<https://arrows.emsl.pnnl.gov/api>
<https://github.com/ebylaska/TinyArrows>

Abstract. Over the last few decades, significant progress has been made in the development and use of electronic structure and other molecular simulation methods. As these methods become more mature and are able to simulate larger and more complex chemical simulations, the need for improvement in scientific visualization, molecular builders, simplified input to simulation methods, and the development of new approaches and languages to describe simulations, along with workflows to carry them out, becomes more apparent. In this chapter, we describe our recent efforts in developing a prototype open-source computational tool called Arrows that combines NWChem, SQL and NoSQL databases, email, web APIs, and web applications in a way that makes molecular and materials modeling accessible to all scientists and engineers. At the same time, because of its simplified input, it provides a framework for expert users to carry out large numbers of calculations and run complex workflows.

Keywords: Computational chemistry workflows, molecular builders, JSIMol, NWChem, Mopac, DFT, CCSD(T), Plane-Wave DFT, PAW, PSPW, Chemistry, Ab initio Molecular Dynamics, AIMD, AIMD/MM, QM/MM, AIMD-EXAFS, hematite, Zn(II), carbon tetrachloride, trichloropropane, 2,4-dinitroanisole, DNAN, hydrolysis, hydrogenolysis, reaction pathways, kinetics, reaction prediction

1 Introduction

Computational chemistry has played a large and growing role in chemistry, material science, and biology, since the appearance of the first electronic computing systems and subsequent programmable computers in the early to mid-part of the last century. Some of the first calculations, which were used to generate atomic form factors for X-ray and electron scattering [1], had an immediate scientific impact that continues to this day. From these beginnings, the growth and impact of the field has continued, so much so, that today it has become a necessary tool across multiple domains of science and engineering. This is an extraordinary impact for a field still undergoing rapid development. In fact, recent economic analysis concluded that molecular modelling contributes $\sim 1\%$ of the GDP of a large technologically advanced country [2].

The success and versatility of computational chemistry is partly due to increases in processing power coupled with advances in algorithms and software. With these advances, computational chemistry is now able to probe various physical observables through the use of predictive atomic level simulations. Combined with the freedom to utilize diverse external constraints and fields, these methods are able to explore larger spatial and time scales of chemical processes under a variety of conditions. Given this versatility, along with the new insight that comes from having an atomic picture of processes, not easily obtained from experiment, it is already very common to augment experimental studies with simulation. For example, simulation is actively used in the interpretation of molecular spectroscopies. It is also actively used to systematically and efficiently extend experimentally measured thermodynamic and kinetic properties of chemical systems. However, augmenting and replacing experiment with simulation is nontrivial, since simulating new materials is complicated by the sensitivity of the processes at the macroscopic scale to the atomic scale; the unusual and unexpected bonding behaviors of materials; cases where extreme temperature and pressure environments could be encountered; and the requirements that simulations be as parameter free as possible and extremely reliable.

The tools of electronic structure theory and statistical mechanics combined with advanced parallel packages such as NWChem [3–6] have proved to be very effective and productive computational chemistry methods. Indeed, computer programs that implement these types of methods consume a large fraction of supercomputer cycles. Even with these hugely successful developments, reliable calculations of this type require considerable theoretical expertise, computational effort, and often costly software licenses that are out of reach for many scientists, engineers, and students.

One of the biggest barriers to using computational chemistry is the need for efficient and intuitive input to the software. This is particularly true for open-source programs such as NWChem that are trying to stay on the cutting edge with the latest theoretical and computational developments, and as a result the complexity of these programs continues to grow unabated. For example, NWChem implements a robust and diverse set of molecular theories, including Gaussian basis set DFT, high-level chemistry methods (e.g. MP2 [7], CCSD(T) [8, 9]), molecular dynamics (MD) [10], hybrid quantum mechanics/molecular mechanics (QM/MM), plane-wave density functional theory (DFT) [11], ab initio molecular dynamics (AIMD), and hybrid ab initio molecular dynamics/molecular dynamics (AIMD/MM) [12, 13] that can estimate the thermodynamics, kinetics, and spectra for molecules, nanoparticles, bio-materials, and minerals. It arguably has the most capabilities of any molecular modeling code today. As a result, new users of NWChem and other computational chemistry programs may feel like they need to take classes on how to use them, or at the very least obtain a certification akin to what used to be thought of as compulsory to use VisiCalc and Lotus spreadsheets in the 1980’s.

In addition to the input challenge, other problems with NWChem and other computational chemistry codes are that:

- Molecular modeling software is extremely complex, contains millions of lines of code, takes a long time to set up and to learn how to use.
- Even the most basic input for molecular modeling software requires the use of other software to generate it.
- It is not uncommon for simulations to take days, weeks and even months to complete.
- Multiple simulations are usually needed to calculate molecular properties and reaction energies.
- Because of this complexity, people unnaturally identify with codes and molecular theories, and they are hesitant to learn new codes and new molecular simulation techniques.

Many groups have been trying to fill this need by developing molecular graphics programs, graphical user interfaces (GUIs), and workflows [15–43] and standardized formats to make molec-

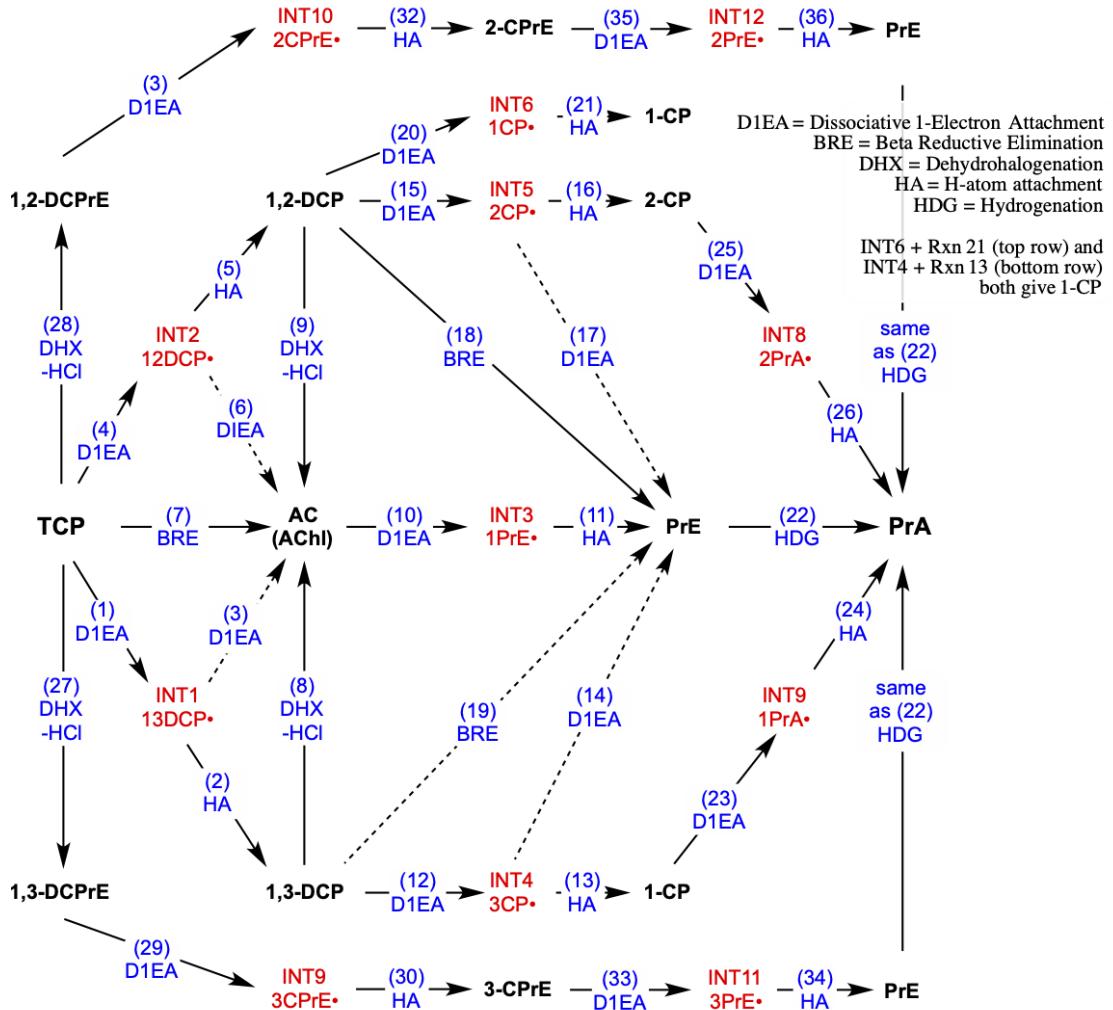


Fig. 1: Example of a reaction network in environmental chemistry from work of authors. Likely reactions for trichloropropane (TCP) under reducing conditions. Reaction network image in figure from Torralba et al. [14].

ular and materials data machine-readable and writable [44–46]. While extremely useful, these interface programs, which are hard to design and implement, sometimes end up being too narrowly focused and at the same time just as complicated to use as the underlying molecular modeling software. Unfortunately, even with the plethora of available GUIs it is not uncommon for a user of quantum molecular modeling software (especially open source software) to become overwhelmed by the proemial state of the software, and having to resort to using multiple GUIs just to run a basic calculation. For example, to perform a simple DFT calculation for a molecule with an open source code like NWChem [3–6], one may have to use separate steps to build the molecule, fetch a basis set or pseudopotentials from a website, pre-optimize the molecule with a molecular force field or semi-empirical theory, submit and manage the DFT calculation, and lastly visualize the results.

To make matters worse, when performing chemical studies, one rarely just needs the results of a single molecular calculation; in fact, it is very common to carry out many hundreds to thousands of molecular calculations. As shown in Fig. 1, this can be true even for chemical studies involving relatively simple molecules, like the degradation of chlorinated hydrocarbons in natural ground waters.

In response to these challenges, we have been investigating and prototyping intelligently-designed web interfaces to carry out complex and difficult chemical and materials simulations, while at the same time developing simplified input to make these simulations accessible to more scientists, engineers, and students. In this chapter, an overview of the open-source computational chemistry and materials tool called Arrows is presented. Arrows is a web tool that combines NWChem, SQL [47, 48] and NoSQL [49] databases, email, back-end web servers, external web services, and web applications to perform molecular and materials modeling from a web browser. There are two implementations of Arrows called EMSL Arrows and TinyArrows. EMSL Arrows [50] is hosted by the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) at the Pacific Northwest National Laboratory (PNNL) [51, 52], and TinyArrows is a stand-alone implementation of Arrows available on GitHub.

2 Evolution of Chemical and Materials Computation

Scientific computing is undergoing its biggest transition in over 60 years. The old way in which scientific computations were performed was by modelers and technicians who had specialized access to large and mid-scale computers at scientific user facilities. The new way will be for scientific calculations to be carried out by individual non-specialized scientists (and AI engines) using smart web 4.0 [53] (web 5.0) interfaces, where the computing is carried out across a federated network of computers instead of at scientific user facilities. How we think about scientific computing is changing.

- In the old way:
 - Computation was expensive.
 - Most computing cycles were obtained from scientific user facilities that were expensive to build and operate.
 - In this hegemonic system based on scarcity, calculations were primarily carried out by experts, running difficult and large numbers of calculations.
 - Routine simulations supporting single investigator scientific studies were given low priorities.
- In the new way:
 - Bezos's Law [54–58] - The price of a unit of computing power decreases by approximately one-half every three years.
 - Cloud computing and client-side computing will be where we get many of our cycles.
 - Most calculations will be automated and easy to carry out with the help of smart AI assistants.
 - As computational resources become more available and upend the system based on scarcity, modelers will be more than just be gatekeepers to computing cycles and providers of expert codes. They will be tasked with exploring scientific hypotheses through simulation.
 - Modern web technologies will play a critical role in carrying out future scientific simulations.

These are big changes, but even bigger changes are to be expected with the rapid development and deployment of new web based technologies combined with database, artificial intelligence (AI), machine learning (ML), and other types of expert system software, making it possible to control devices, organize and plan work, and monitor events remotely.

A common term that encompasses most modern web technologies is "HTML5". It refers to the fifth and latest version of HTML (Hyper Text Markup Language) for creating web pages that are more interoperable than in the past [59]. Besides improving the markup available for web documents, it also introduces into the standard, various application programming interfaces (APIs), designed to work with low-powered devices, to help programmers build more robust and efficient web (and cross-platform mobile) applications. New features include advanced vector graphics (notably a full implementation of OpenGL through WebGL [60, 61]), audio (<audio> HTML element), video (<video> HTML element), and voice recognition (Web Speech [62]). There is even available software that implements a flexible database (indexedDB [63, 64]). While HTML5 is still being developed and at the moment there is no browser that it fully supports, however, this situation has significantly improved in recent years, and many of its new features are now supported on most major web browsers.

Now that HTML5 is widely available, the time is right for computational chemistry to evolve from being a tool used only by specialists to a tool used by all chemists, material scientists and other scientific and engineering researchers. A strength of molecular modeling, in particular electronic structure methods, that make it viable to new ways of computing is the development of many human-readable and machine-readable input and output formats (e.g., geometry formats: SMILES [65, 66], IUPAC names [67], common names, InChI [68–70], InChIKey [68, 71], PDB [72], XYZ, mol, ...; input and output formats: POSCAR [73], CIF [74], gau, gam, nw, ...; number formats: CAS, KEGG numbers [75], CID [76], CSID [77, 78], ChEMBL [79], EC Number [80], DrugBank [81], ICSD [82], ...), and various standard acronyms to describe various calculation details such as theories (e.g., CCSD(T) [8, 9], MP2 [7], DFT [83, 84], HF [85], PSPW [11], MCSCF [86], and CI [87, 88]), basis sets and exchange correlation functionals (e.g. 6-31G* [89–92], 6-311++G(2d,dp) [89, 93–96] B3LYP [97, 98], PBE0 [99], M06-2x [100], LDA [101], PBE [102], and BLYP [98, 103]), and simulation methods (e.g., AIMD, NEB [104], DOS [105], PDOS, BSSE [106], DIIS [107], and COSMO [108]). This strength is also one of its weaknesses, as the downside to all these formats is it gives the appearance that input for molecular modeling is complicated and requires extensive expertise to use.

Input does not need to be complicated, and extensive software expertise should not be required for most molecular modeling calculations. Every scientist and engineer has been trained to write balanced chemical reactions in their high-school chemistry class. Indeed, it is possible to use optical chemical structure recognition (OCSR) to recognize handwritten chemical reactions, although these tools probably won't be completed and widely available for a few years [109–112]. Further, today there are a variety of string formats that can be used to enter chemical compounds. Popular string formats such as IUPAC names, SMILES [65, 66], and InChI strings [69, 70] are all capable of describing chemical structures. For defining chemical structures, IUPAC names and SMILES have an advantage over other string formats in that they are human-readable. However, InChI strings, which are not designed to be human-readable, have an advantage over other formats in that every structure has a unique InChI string, which is important in database applications. This property, known as canonicalization, has a long history of being used to register drawings of chemical structures and indexing chemical databases [66, 113–115]. The earliest solution to this problem was developed by Harry Morgan of CAS (Chemical Abstracts Service), who developed an algorithm that generates a unique and unambiguous record of a molecule's structure [113]. This algorithm, today known as the Morgan algorithm, became the basis of the CAS Chemical Registry System (CAS Registry) [116]. Not surprisingly, as we build and design

intelligent web interfaces for chemistry, canonicalization plays an important role in the design and implementation of these services, especially for their back-end or server-side implementations.

3 Overview of the Arrows Web Applications and Web Services

Arrows is designed to be used by experts and non-experts alike. Experts can carry out and keep track of large numbers of complex calculations with diverse levels of theories present in their workflows. Additionally, due to a streamlined and easy-to-use input, non-experts will be able to carry out a wide variety of molecular modeling calculations previously not accessible to them. The Arrows software uses NWChem and chemical computational databases to make materials and chemical modeling accessible via a broad spectrum of digital communications including web applications (web apps), posts to web APIs, and traditional email. There are two versions of Arrows. The first version is called EMSL Arrows. It can be accessed at the <https://arrows.emsl.pnnl.gov/api> URL. The second version, called TinyArrows, is a stand-alone version, which can be installed on a laptop or workstation that is behind a network firewall. The source code for TinyArrows can be obtained from the <https://github.com/ebylaska/TinyArrows> URL.

Arrows is very simple to use: One just enters a chemical reaction as a string into a web application (or emails it to arrows@emsl.pnnl.gov) and then, for example, thermodynamic properties, reaction pathways (kinetics), and spectra are returned. For those interested in the infrastructure, Arrows first parses the input, and then searches an SQL [47, 48] (Structured Query Language) database for the compounds in the reactions. To avoid the duplication of database entries, the compounds are searched on a sequence of indexes including canonicalized string formats to describe the molecular structure, charge, multiplicity, theory, exchange-correlation, basis sets, and solvation type. If a compound isn't there, an NWChem calculation is set up and placed into a NoSQL [49] (Not Only Structured Query Language) container, and then the container is submitted to an internal queue awaiting calculation. The internal Arrows queue is used to manage NWChem and other jobs on machines that have HTTP [117] access to a running Arrows web service. Once the calculation is finished (and uploaded back to the Arrows queue), the results (downloaded from the Arrows queue) are parsed and then entered into the SQL database. This whole process is completely automated.

The following subsections 3.1 and 3.2, are organized to highlight basic capabilities for novice users, and the more advanced capabilities that were designed to be used by experts in molecular and materials simulation. Detailed instruction on how to use Arrows can be found in the online manuals (https://nwchemgit.github.io/EMSL_Arrows.html# or <https://ebylaska.github.io/TinyArrows/>).

3.1 Basic Arrows Calculations for Molecules and Chemical Reactions

A simple string is given as input to calculate a molecule or a chemical reaction with Arrows. The input for a molecule is called an "extended smiles" or *esmiles* for short, and it has the following form, `Molecule_Input keyword1{option1} keyword2{option2}... keywordN{optionN}`. `Molecule_Input` can be specified using a variety of formats including a SMILES [65, 66] string, common names, IUPAC [67], KEGG numbers [75], CAS Registry numbers [118], PubChem IDs [76], ChemSpider IDs [77, 78], ChEMBL [79], InChI strings [69, 70], and XYZ data. The `keyword{option}` tags are used to enter different theories (e.g., pspw theory, ccsd(t), or pbe0 exchange correlation functional, ...) and calculation types (e.g., energy, optimization, vibrational spectra, ...) for a molecule. Examples of *esmiles* input for molecules that can be used with Arrows are as follows:

```

CN1C=NC2=C1C(=O)N(C(=O)N2C)C
CN1C=NC2=C1C(=O)N(C(=O)N2C)C theory{pspw}
CN1C=NC2=C1C(=O)N(C(=O)N2C)C xc{pbe}
CN1C=NC2=C1C(=O)N(C(=O)N2C)C xc{m06-2x}
CN1C=NC2=C1C(=O)N(C(=O)N2C)C theory{mp2}
PubChem=2519
ChemSpider=2424
cas=58-08-2 theory{blyp}
kegg=D00528
InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3
Caffeine theory{ccsd(t)}
CHEMBL113 theory{pspw} xc{lda}
1,3,7-trimethylpurine-2,6-dione
Cinnamaldehyde
Cubane
[C@H] 12 [C@H] 3 [C@H] 4 [C@H] 2 [C@H] 2 [C@H] 1 [C@H] 3 [C@H] 42 theory{pspw4}
[C] 1=C=C=C=C=C=C=C=[C] 1

```

The following *esmiles* input shows the form of how to enter XYZ data.

xyzdata{C	3.26140	1.16120	-0.00560	
N	2.28620	0.06800	-0.00280	C
N	1.44461	-1.93420	0.00230	C
C	0.91140	0.19390	0.00350	C
O	0.43790	2.42800	0.00450	N
C	-2.24660	2.17610	-0.00080	C
O	-2.99380	-0.38390	-0.00470	N
C	-1.53370	-2.62940	0.00020	H
H	2.83960	2.02580	0.50690	H
H	3.55200	-1.67960	-0.00020	H
H	-3.16470	1.88780	0.51110	H
H	-2.62210	-2.57040	-0.00560	H
H	-1.20340	-3.16190	0.89210}	

To run a basic Arrows calculation, an *esmiles* string (or *esmiles reaction* string) is entered as input into the Arrows web app as shown in Fig. 2, and then the "Run Arrows" input button is selected. The results returned by Arrows are a combination of text and graphical output as shown in Fig. 3.

Tiny Arrows Expert Periodic and Molecular Editor 3D Molecular And Reaction Editor Molecular Combo Editor Microsoft Quantum Editor EMSL Aerosol Workshop Editor
Manual

Tiny Arrows 2D Molecular And Reaction Editor

Tiny Arrows is a revolutionary approach to materials and chemical simulations that uses NWChem and chemical computational databases to make materials and chemical modeling accessible via a broad spectrum of digital communications.

389701 Arrows calculations carried out to date
59513 Molecular calculations in database to date

Enter an esmiles, esmiles reaction, or other Arrows input, then push the "Run Arrows" button.

theory(dft) theory(pspw) theory(pspw4) theory(mp2) theory(ccsd(t)) theory(msm3) basis(b-31G*) basis(g-311++Q(2s2p))
 xc[lda] xc[pbe] xc[pbe0] xc[blyp] xc[m06-2x] solvation_type(None) solvation_type(COSMO) solvation_type(COSMO-SMD) (water default)
 [] 1 2 3 4 5 6 7 8 9 0 change multi

Arrows input to JSME

NEW X R M Z C S F Cl Br I P X
 = ~ Δ □ ○ ◊ ◇ FG Y A

O=C(O)c1ccccc1C(=O)O

JSME Editor Citation: B. Bienfait and P. Ertl, JSME: a free molecule editor in JavaScript, *J. Cheminformatics* 5:24 (2013)

The following are examples of input that can be used with Tiny Arrows. Clicking on a button places the command into the above entry box.

Molecular calculations: [CN1C=NC2=C1C\(=O\)N\(C\(=O\)N2C\)C](#) [Pubchem=2519](#) [ChemSpider=2424](#) [cas=58-08-2](#) [kegg=D000528](#) [InChI=1S/C8H10N4O2/c1-10-4-9-6-5\(10\)7\(13\)12\(3\)8\(14\)11\(6\)2/h4H,1-3H3](#) [Caffeine](#) [CHEMBL113](#) [CN1C=NC2=C1C\(=O\)N\(C\(=O\)N2C\)C theory\(pspw\)](#) [CN1C=NC2=C1C\(=O\)N\(C\(=O\)N2C\)C xc\[pbe\]](#) [CN1C=NC2=C1C\(=O\)N\(C\(=O\)N2C\)C xc\[m06-2x\]](#) [CN1C=NC2=C1C\(=O\)N\(C\(=O\)N2C\)C theory\(mp2\)](#)

Reaction calculations: [CC + hydrogen gas -> 2 methane](#) [C+C + ClCl -> CICCI](#) [TNT + hydroxide -> TNT-2-OH + nitrite](#) [TNT + hydroxide -> TNT-2-OH + nitrate - xc\(m06-2x\)](#) [TNT + hydroxide -> TNT-2-OH + nitrite - theory\(pspw\)](#) [TNT + water -> TNT-2-OH + nitrous acid - theory\(pspw\)](#)

NMR calculations: [nmr for O=N\(=O\)c1cc\(N\(=O\)=O\)c\(c1\)N\(=O\)=O/C basis\(b-31G*\)](#) [nmr for O=N\(=O\)c1cc\(N\(=O\)=O\)c\(c1\)N\(=O\)=O/C basis\(b-31G*\) solvation_type\(None\)](#) [nmr for aspirin](#)

Predict reactions: [TNT + hydroxide ->](#) [CcC + ClCl ->](#) [TNT + hydroxide -> ~theory\(pspw4\)](#) [predict for TNT + hydroxide](#) [predict for tetral + hydroxide](#)

SMARTS searching: [SMARTS cN\(=O\)=O xc\[m06-2x\]](#) [SMARTS c1ccccc1 theory\(pspw4\)](#) [SMARTS \[C#H\]\(=O\)\[#6\]](#) [SMARTS \[CX3\]\(=O\)C theory\(mp2\)](#)

Generate NWChem input, Fetch NWChem output, Fetch XYZ and Fetch JSON: [input deck for aspirin](#) [input deck for aspirin theory\(pspw\)](#) [input deck for aspirin theory\(mp2\)](#) [input deck for PubChem=2519 xc\[pbe\]](#) [output deck for aspirin](#) [mwoutput for aspirin theory\(pspw\)](#) [xyz for aspirin theory\(pspw\)](#) [json for aspirin theory\(pspw\)](#)

Other Arrows commands: [list all reactions](#) [list all esmiles 100](#) [queue](#) [download fetch_nwchem_input](#) [download upload_nwchem_outfiles](#) [download arrows_esmiles2xyz](#) [submitsmiles for ethene theory\(pspw\)](#)

Fig. 2: Snapshot of the basic Arrows web app (<http://localhost:5000/api/>). Below the input bar in the web app, there are several options available that can be used to change the theory, exchange-correlation, basis set, solvation method, charge, and multiplicity of the *esmiles* and *esmiles reaction* input. At the bottom of the default Arrows web page, there are examples of commands that can be placed into the entry box. A JSME molecular editor [42] is integrated into the web-page that can be used to generate SMILES input. There are also web apps that are integrated with a JSMol molecular builder [34,36] (e.g., <http://localhost:5000/api/3dbuilder>).

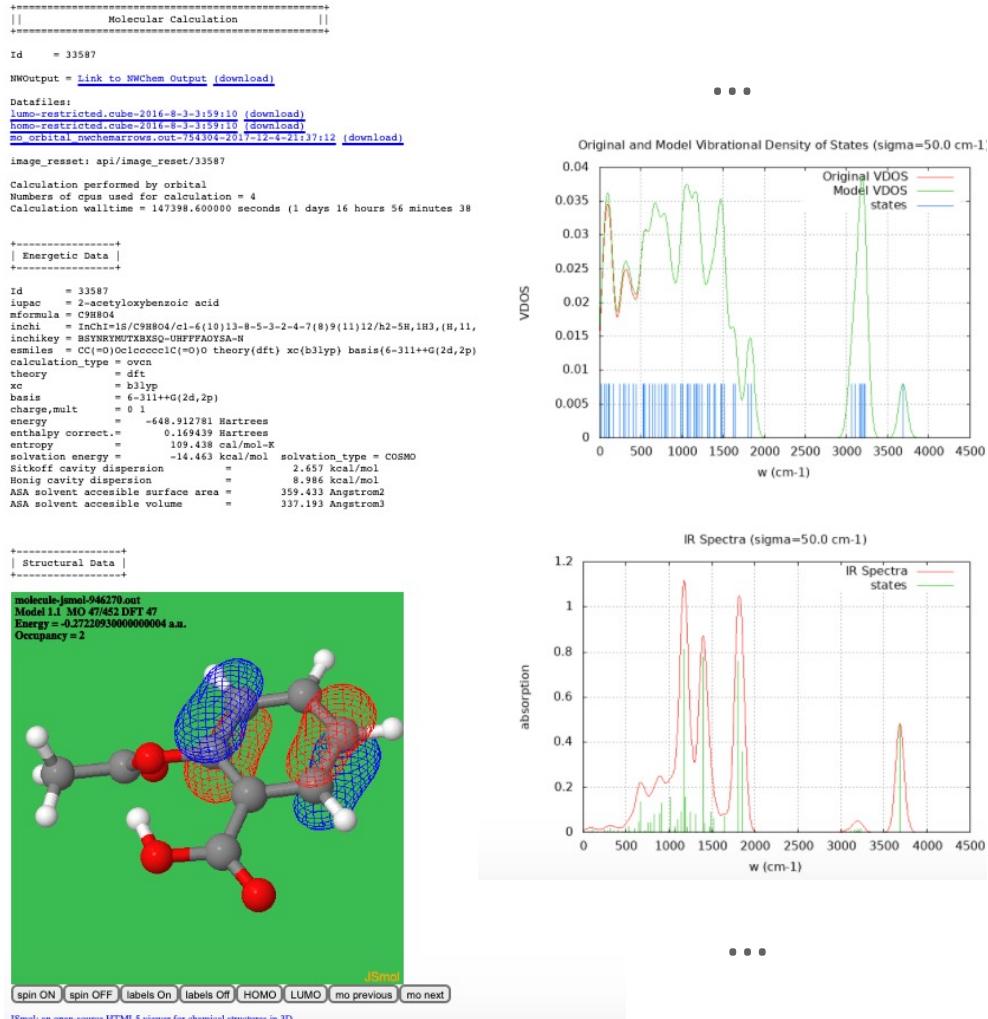


Fig. 3: Illustration of the output from an Arrows molecular calculation. The complete molecular output contains sections for "Energetic Data", "Structural Data", "Eigenvalue Data", "Vibrational Density of States Analysis", and "Reactions Contained in the Database" (which contain the molecule).

The input for a chemical reaction is called an *esmiles reaction* in Arrows. It is composed of a list of *esmiles* separated by the "+" and "-->" (dash-dash-greater than) tokens. To change the calculation options for an *esmiles reaction* a map function has been added to the reaction input, where the format for the mapping function is to append the reaction with the tilde, "~", symbol followed by keyword{option} tags. Examples of the basic input for chemical reactions are as follows:

```
C(Cl)(Cl)(Cl)O + C --> C(Cl)(Cl)Cl + CO
C(Cl)(Cl)(Cl)O + C --> C(Cl)(Cl)Cl + CO ~ theory{pspw}
C(Cl)(Cl)(Cl)S + C --> C(Cl)(Cl)Cl + CS
C(Cl)(Cl)(Cl)S + C --> C(Cl)(Cl)Cl + CS ~ theory{pm3}
TNT + 3 benzene --> toluene + 3 nitrobenzene ~ xc{pbe}
cid=8376 + hydroxide --> O=N(=O)c1cc(O)c(c1N(=O)=O)C + nitrite ~ theory{mp2}
carbon tetrachloride + 2 SHE + [H+] --> chloroform + chloride
```

To run a chemical reaction with Arrows, a chemical reaction (i.e., *esmiles reaction*) is entered as input into the Arrows web app as shown in Fig. 4, and then the "Run Arrows" input button is selected. The results returned by Arrows are a combination of text and graphical output as shown in Fig. 5.

3.2 Other Kinds of Arrows Calculations for More Advanced Computational Scientists

This subsection highlights a variety of other calculations that can be performed with Arrows. Many of these calculations are designed to be used by expert users of Arrows who have experience in using web services, and who also have expertise in carrying out advanced electronic structure calculations.

NMR spectra can be calculated by entering into Arrows an *esmiles* that is preceded by the words “nmr for”, e.g.

```
nmr for c1ccccc1 basis{6-31G*} solvation_type{None}
```

XYZ, JSON, NWChem input, and NWChem output can be fetched by entering into Arrows an *esmiles* that is preceded by “xyz for”, “json for”, “input deck for”, or “output deck for”, e.g.

```
xyz for c1ccccc1 basis{6-31G*}
json for c1ccccc1 basis{6-31G*}
input deck for c1ccccc1 basis{6-31G*}
output deck for c1ccccc1 basis{6-31G*}
```

SMARTS searching can be performed by entering into Arrows an *esmiles* that is preceded by the words “smarts”, e.g.

```
smarts cN(=O)=O xc{m06-2x}
smarts c1ccccc1 theory{pspw4}
smarts [CX3H1](=O)[#6]
smarts [CX3](=[OX1])C theory{mp2}
```

Input for quantum computing simulations with the Microsoft Quantum Development Kit - Arrows can be used to generate a Broombridge YAML file containing the 1-electron and 2-electron integrals over a user defined select-CI orbital subspace. This YAML file can be used by the Microsoft Quantum Development Kit chemistry library for quantum chemistry simulations. The easiest way to generate a Broombridge YAML file is to use Microsoft Quantum

Enter an esmiles, esmiles reaction, or other Arrows input, then push the "Run Arrows" button.

Append SMILES from JSMET Molecular Editor
Reaction on
 no html5
 Clear

theory{pspw}
 theory{pspw4}
 theory{mp2}
 theory{ccsd(t)}
 theory{pm3}
 basis{6-31G*}

xc{da}
 xc{pbe}
 xc{pbe0}
 xc{b3lyp}
 xc{m06-2X}
 solvation_type{COSMO}
 solvation_type{COSMO-SMD}

water (default)
>

->
+ ~ 1 2 3 4 5 6 7 8 9 0 □ charge mult

Run Arrows
Email Submit

Fig. 4: Example of the basic Arrows input for a chemical reaction (i.e., *esmiles reaction*). Below the Arrows input, there are several options available that can be used to change the theory, exchange-correlation, basis set, solvation, charge and multiplicity of the esmiles and *esmiles reaction* input.

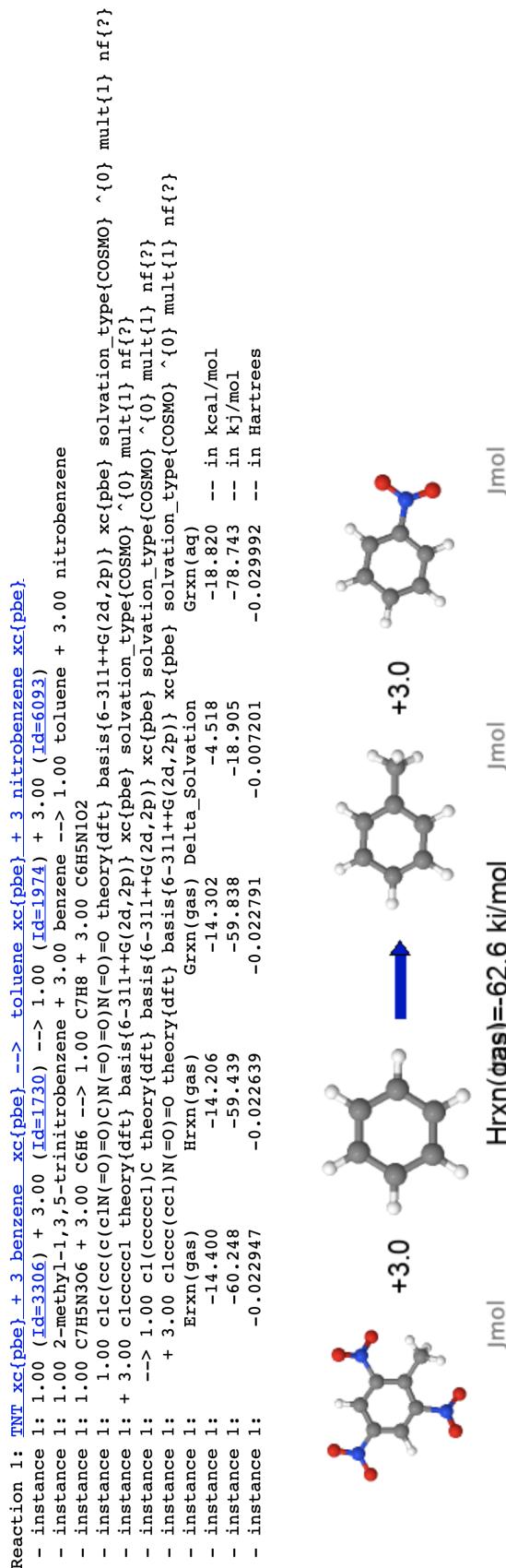


Fig. 5: Example of chemical reaction output generated by Arrows using an *esmiles reaction* input. The reaction energy output is given in kcal/mol, kj/mol and Hartree units, and contains the gas-phase energies, enthalpies and free energies of reaction, as well as solvation and aqueous reaction free energies of reaction.

Editor web app, which contains a few addition options to help set up the input. The following steps can be used to generate the input and run Arrows.

1. Press the “Append SMILES from JSME Editor” button to populate the Arrows text box with a SMILES string defined from the structure drawn in the JSME editor [42] to the Arrows input.
2. Press the "filling" button and enter the number of orbitals and electrons. This input defines the orbital subspace. The default filling for qsharp_chem theory is qsharp_chem_filling{2 1 1}.
3. Press the "nroots" button and enter the number excited states. The default is to have 0 excited states.
4. Press the basis, charge, and mult buttons to change the basis set, molecular charge and molecular spin multiplicity.
5. Press the "theoryqsharp_chem" button to add the Qsharp theory to the Arrows input.
6. Press "Run Arrows"
7. If the data needed to complete the Arrow request is not available, the web-page will contain after the "Link back to Microsoft Quantum Editor" link the text "No molecule data for esmiles". When this happens Arrows will automatically submit a job to one of its calculation queues,
8. If the data to complete the request is available, a link to the YAML file will be available in the "Molecular Calculation" section. The section will look as below.

```
...
+=====+
||          Molecular Calculation          ||
+=====+
Id      = 48414

NWOutput = Link to NWChem Output (download)

Datafiles:
qsharp\_chem.yaml-2018-10-18-13:44:50 (download)

Calculation performed by Eric Bylaska - we29676.emsl.pnl.gov
...
```

In the Datafiles: subsection click on the (download) link next to the YAML file (e.g. qsharp_chem.yaml-2018-10-18-13:44:50 in the above example).

Accessing data using web APIs - The following restful web APIs that can be used to fetch data from Arrows. The URL formats available are:

Restful Web API URL Format:

- [-](https://arrows.emsl.pnnl.gov/api/molecule/)

where the *esmiles* and *esmiles reaction* input are appended by the parse output option, `parse_output{option}`. The parse output option can be any of the following: `e(gas)`, `h(gas)`, `g(gas)`, `g(aq)`, `erxn(gas)`,

`hrxn(gas), grxn(gas), grxn(aq), delta_solvation, homo, lumo, alpha_homo, alpha_lumo, beta_homo, or beta_lumo.`

Here are some examples of calling restful web APIs in Arrows:

Examples of restful Web API URL input:

- ["](https://arrows.emsl.pnnl.gov/api/molecule/)

To use the above APIs, just copy and paste the following lines into web browser URL input, or use with an HTML GET request method in a program written in a programming language such as Python, JavaScript, TypeScript, Java, C#, C++, Go, Julia, etc..

Unfortunately, the “#” symbol, which is used to designate a triple bond in a SMILES string, causes problems with URL input. There are a variety always around this problem. The Arrows APIs have been programmed to use “!” symbol as a replacement for “#” symbol. There are also several alternatives to using the “#” symbol, that can be used in Arrows web API, as listed below.

Alternative restful web API URL input for molecules, such as the ethyne molecule, that use the # symbol:

- ["](https://arrows.emsl.pnnl.gov/api/molecule/)

Using these web APIs, one can fetch data from Arrows directly into Google sheets (UrlFetchApp script) and Microsoft Excel (Windows version only) spreadsheets using scripting (i.e. the UrlFetchApp in Google sheets, and the fetch API in Excel).

SpreadSheet Input - There is also a spreadsheet capability, Jspreadsheet [119], that has been integrated in Arrows as shown in Fig. 6. To use it, navigate to the "Expert Periodic and Molecular Editor" web app, and then select the "SpreadSheet" tab. The following steps can be used to make a table in the spreadsheet that can filled in with calculations from Arrows:

Arrows Periodic 3D Builder										NWChem Input Editor										Arrows Entry - 2D Builder (not periodic)										Notes		SpreadSheet		Output		Queue		Reaction Predictor - Beta version		Instructions		
Arrows					Periodic					Molecular					Reaction					Editor					Molecular		Combo		Editor		Quantum		Editor		Aerosol		Workshop		Editor		Manual	
32	1-a-1-b	CO --> [C](O) + [H+] + [Cl-]	A	B	C	D	E	F	G	H	I	J																														
33					0.450243	0.464494	0.490144	0.447858	0.448266	0.464408	0.461045																															
34	1-a-2-a	C(C)O ^{-1} + [H+] --> C(C)C(Cl)O		-0.503215	-0.486647	-0.586277	-0.477251	-0.503877	-0.582595	-0.526625																																
35	1-a-2-b	C(C)O ^{-1} --> C(=O)C1 + [Cl-]		0.015319	0.039154	-0.055734	0.050504	0.025692	-0.055456	0.007263																																
36																																										
37	1-a-2-b-1	C(=O)Cl + [OH-] --> C(=O)O + [Cl-]		-0.104751	-0.099233	-0.107839	-0.112383	-0.101220	-0.090786	-0.102934																																
38																																										
39																																										
40																																										
41	rxn	Aqueous Reaction Free Energy (Hartrees)					xc(b3lyp)	xc(pbe)	xc(pbe0)	xc(m06-2x)	theory(ccsd(t)}	theory(ppw4)	theory(ppw4)	theory(pbe0)	theory(ccsd(t)}	theory(ppw4)	theory(ccsd(t)}	theory(ppw4)	theory(ccsd(t)}	theory(ppw4)	theory(ccsd(t)}	theory(ppw4)	theory(ccsd(t)}	theory(ppw4)	theory(ccsd(t)}	theory(ppw4)	theory(ccsd(t)}	theory(ppw4)	theory(ccsd(t)}	theory(ppw4)	theory(ccsd(t)}	theory(ppw4)	theory(ccsd(t)}									
42	1	C(C)(C)(C)Cl + [OH-] --> C(C)(C)(C)Cl + [Cl-]		-0.085821	-0.081013	-0.087766	-0.090578	-0.080716	-0.080877																																	
43	2	C(C)(C)(C)Cl + SHE --> C(C)Cl + [Cl-]		-0.025673	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)											
44	3	C(C)(C)(C)Cl + SHE --> [C-](C)(C)Cl(C)Cl		grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)																	
45																																										
46	1-a	C(C)(C)(C)O + SHE --> C(C)(C)O + [Cl-]		grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)																
47	1-b	C(C)(C)(C)O + SHE --> C(C)(C)O ^{-1}		grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)	grxn(eq)																	
48																																										
49	2-a	C(C)Cl + [OH-] --> C(C)(C)(C)O ^{-1}		-0.079006	-0.072442	-0.067949	-0.072484	-0.072690	-0.056166	0.005301																																
50	2-b	C(C)Cl + SHE --> C(C)Cl ^{-1}		-0.010979	-0.007624	-0.002350	-0.011828	-0.002835	0.002467	0.007151																																
51	2-c	C(C)Cl + SHE + [H+] --> C(C)(C)Cl		-0.069867	-0.062950	-0.056913	-0.070409	-0.069309	-0.06050	-0.063803																																
52	2-d	C(C)Cl + [H+] --> C(C)(C)Cl ^{-1}		0.076158	0.070866	0.0712154	0.090655	0.086744	0.086924	0.086924																																
		table db number: 2 table type: [Gas-Phase Reaction Enthalpy(Hartrees) column theories: [theory(ppw4) : xc(pbe) : xc(b3lyp) : xc(m06-2x)]																																								
					</td																																					

Fig. 6: Illustration of using the Arrows spreadsheet capability to calculate reaction energies for the reactions contained in the CCl_4 reaction network illustrated in Fig. 10 (vida infra).

Using Spreadsheet to Create an Arrows Calculation Table:

1. Just below the spreadsheet, select a table type from the list of table types (e.g., "Gas-Phase Reaction Enthalpy (Hartrees)").
2. To right the right of the table type input, enter *esmiles* options, separated by ":" , in the column theory input.
3. In the next input to the right enter the number of table rows, which is the number of *esmiles* or *esmiles reaction* that will be calculated.
4. Once these inputs are set, select a cell in the spreadsheet where to place a table template. To place the template, select the table icon (7th icon from the right) in the options at the top of the spreadsheet.
5. To complete the table input, enter *esmiles* or *esmiles reaction* as the row labels, replacing the row labels temporarily filled in with the *esmiles* or *esmiles reaction* string. Note, the smiley face icon can be used to place a smiles string for the structure currently residing in the 2D builder tab.
6. To fill in the body of the table with values calculated from Arrows, just select the "Submit Arrows" button below the spreadsheet.

The Arrows Queues are used to manage NWChem and other jobs on machines that have HTTP access to a running Arrows web service. There are two queues in Arrows. The standard queue in Arrows, called "queue", is used to keep track of NWChem calculations that were generated when a request was made for a compound that was not available in the Arrows databases. The jobs on the queue can be obtained by entering "queue" in the Arrows input. In addition, to the standard queue in Arrows, there is also a second queue, called "queue_nwchem", that can be used to manage user specified NWChem (and other program) input and output decks. A list of jobs on the queue_nwchem can be obtained by entering "queue_nwchem" in the Arrows input. The two Arrows queues can also be accessed using the following URLs.

URL commands to list jobs on the queue and nwchem_queue:

- <https://localhost:5000/api/queue> [GET]
- https://localhost:5000/api/queue_html [GET]
- https://localhost:5000/api/queue_nwchem [GET]
- https://localhost:5000/api/queue_nwchem_html [GET]

URL commands to fetch and delete from from the queue and nwchem_queue:

- https://localhost:5000/api/queue_fetch/<id> [GET]
- https://localhost:5000/api/queue_nwchem_fetch/<id> [GET]
- https://localhost:5000/api/queue_delete/<id> [GET]
- https://localhost:5000/api/queue_nwchem_delete/<id> [GET]

URL commands to upload jobs to the queue and nwchem_queue

- <https://localhost:5000/api/upload/> [POST:files]
- https://localhost:5000/api/submit_output [GET]
- https://localhost:5000/api/submit_nwchem_output [GET]

The Advanced Builders Implemented in Arrows can also be used to carry out computational chemistry and other electronic structure calculations using its more traditional web-based

periodic and molecular builders. These builders are written in JavaScript, and they contain the following capabilities:

Periodic and Molecular Builder Web Applications Implemented in Arrows:

1. Works on any web browser on workstations, laptops, tablets, smart phones.
2. Extensive options to build and modify periodic and cluster systems.
3. Create, edit, load and save NWChem input decks and submit jobs.
4. Generates and cleaves surfaces in seconds.
5. Options for setting up initial reaction pathways, free energy simulations, electron transfer calculations.
6. Space-group symmetry.
7. Text and spreadsheet editors.
8. MM optimization.
9. Solvation for periodic systems via web API.
10. Workflows to carry out various reaction pathway/TST methods including
 - NEB method
 - WHAM reaction free energy with Hausdorff moments fitting of histograms
 - String method,
 - Reaction coordinate penalty functions (e.g. $E_{\text{penalty}} = \frac{K}{2}(\gamma(R_I) - \gamma_0)^2$ where $\gamma(R_I)$ is the order parameter or collective variable of the reaction and γ_0 is a constant),
 - Potential of mean force,
 - WHAM reaction free energy with Hausdorff moments fitting of histograms
 - Metadynamics,
 - TAMD,
 - Single sweep AIMD free energy methods,
11. AIMD-EXAFS workflow.
12. User defined workflows through the use of a scripting language implemented directly in the advanced builders.
13. Reaction prediction.
14. View NWChem output, structures, trajectories, Gaussian cube files, free-energy surfaces, etc.

Predicting Reaction Pathways - A variety of chemical reaction pathway prediction toolboxes have become available in recent years [120–127]. The development of rules-based systems like LHASA [120], EPA-CTS [128] (<https://www.epa.gov/cts/>) and EAWAG-BBD [129] (<http://eawag-bbd.ethz.ch/predict/>), KinBot [130, 131], RMG [132], and IBM RXN for Chemistry(<https://rxn.res.ibm.com/>) ultimately can provide a universal solution to path pathway prediction. However, these methods require lots of input and detailed knowledge of the chemistries being modeled. As a result, these developments have been spanning over decades, even though the underlying computing hardware needed to run these methods is not changing. Another approach for pathway prediction is to use brute force combinatorics, which generate possible reaction mechanisms with electronic structure calculations to evaluate the reaction pathways. These methods break N bonds and make M bonds, where M and N are defined to be relatively small to limit the number of results. Because of the large number of products produced even with $M,N < 4$, these approaches need to be augmented with quick energetic filters such as classical molecular dynamics potentials, machine-learning potentials, and Benson group methods [132].

We have been developing an alternative learning approach for predicting reactions in Arrows using predicate calculus learning [133] based on defining canonical reaction hashes. Reaction

Algorithm to generate reaction hashes

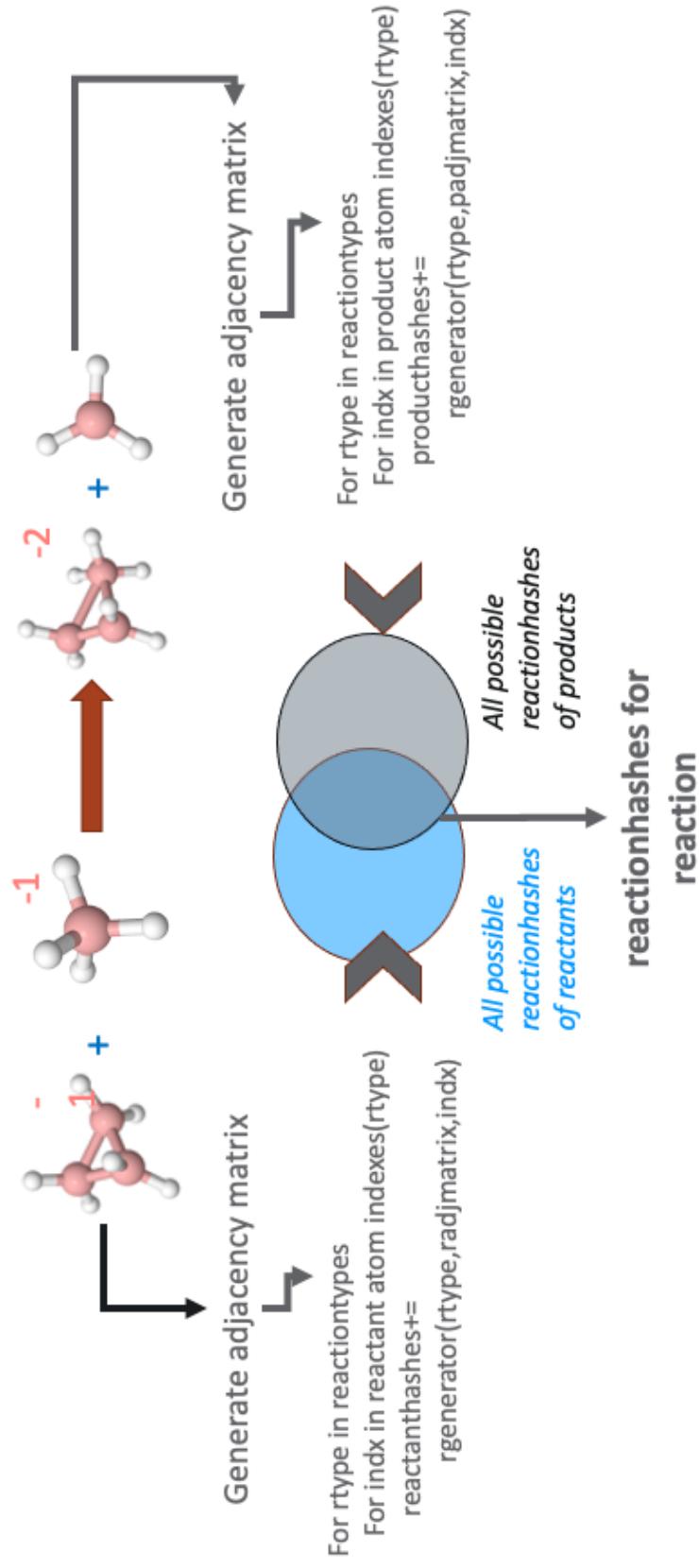


Fig. 7: Illustration of algorithm to generate a reaction hash from a balanced chemical reaction.

hashes classify reactions by defining the connectivity of the atoms relative to a reaction coordinate. The hashes generated in this way can be quite big for larger systems, however, they can easily be made more manageable by restricting how far the connectivity is spanned. Given the adjacency matrix of a chemical system (where the XYZ coordinates of reactants or products of a reaction can be used to define an adjacency matrix), a reaction hash is created by specifying the type of reaction (e.g., A + B \rightarrow AB, AB + C \rightarrow AC + B, AB + CD \rightarrow AC + BD, EA + BCD \rightarrow AB + CDE, ...) and the indexes of atoms involved in the reaction coordinate. In other words, a hash function is defined in Arrows that generates a hash string

```
hash_string = ReactionHash(XYZ string, "2 1 8", "AB + C -> AC + B"),
```

that contains information about the connectivity of the chemical system, where the three input strings describe the XYZ coordinates for the reactants, atom indexes (e.g. "2 1 8"), and the reaction type (e.g. "AB + C \rightarrow AC + B").

Using these hash functions, canonical reaction hashes for a balanced chemical reaction can be generated using the algorithm illustrated in Fig. 7. These canonical hashes can be used to find products given the reactants, and to define initial reaction pathways for Nudged Elastic Band (NEB) [104,134], and string calculations [135]. By making a database of reaction hashes from a set of balanced chemical reactions, it is possible to generate future chemical reactions by filtering the hash to reduce how far it spans the chemical graph (Fig. 8). Both the initial reaction pathway and reaction learning capabilities are available in Arrows.

- Initial reaction pathways can be generated using the web-based periodic and molecular builders web application by selecting the "reaction input" in the "Arrows Periodic 3D Builder" tab.
- Reaction predictions can be calculated by using the "Reaction Predictor - (Beta Version)" tab in the web-based periodic and molecular builders" web application, or by entering into the Arrows input (or using the reaction web api) only the reactants of an *esmiles reaction* followed by the reaction arrow " \rightarrow ", e.g.,

```
C=C + ClCl -->
TNT + hydroxide -->
TNT + hydroxide --> ~ theory{pspw4}
DNAN + [SH-] -->
```

4 Applications

Arrows or early versions of capabilities in the Arrows have been used to carry out a variety of scientific studies [5, 11, 14, 136–148]. Most of these studies fall into two categories. The first use cases are for performing large numbers of molecular and reaction energy calculations. The second use cases employ the workflow capabilities in Arrows to carry out complex simulations, e.g. ab initio molecular dynamics informed extended X-ray absorption fine structure (AIMD-EXAFS [149]), and hybrid AIMD and molecular dynamics (AIMD/MM [12, 13]) free energy simulations.

The following subsections highlight using Arrows to carry out molecular computations in two areas of the authors' research: environmental chemistry and materials computations for geochemistry (and catalytic chemistry).

4.1 AIMD Simulations and Workflows in Geochemistry and Catalysis

The interpretation of geochemical and related catalytic surface processes are complicated by chemical complexity and structural heterogeneity of natural materials and their interfaces, as well

Algorithm to generate reaction coordinates

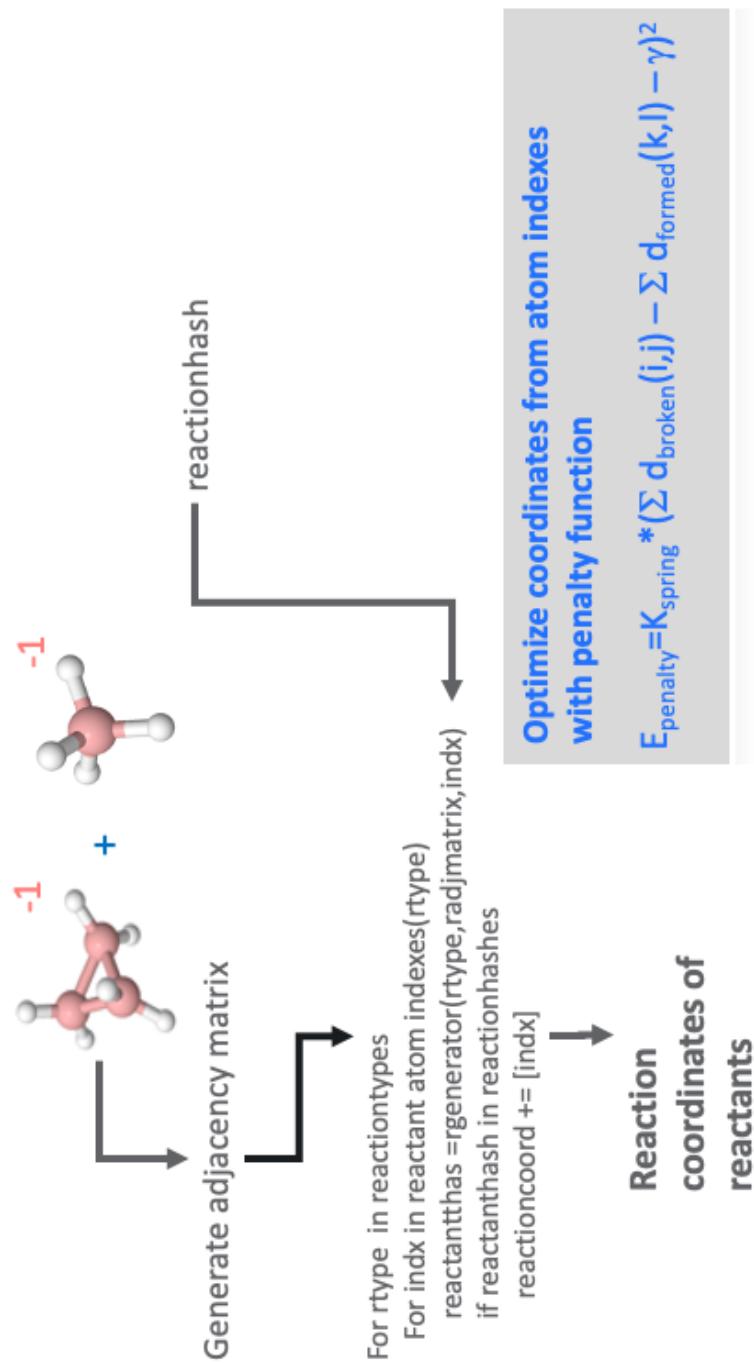


Fig. 8: Illustration of algorithm to generate reaction coordinates given a reaction hash and reactants.

as by the wide range of temperature and pressure conditions under which they occur. Unraveling the nature of these processes in terms of key molecular-level reactions under realistic conditions is a grand challenge in geochemistry and catalysis. AIMD simulations are well suited to make inroads in this area of chemistry [13, 137, 142, 143, 149–163].

In recent years, Bylaska and collaborators have demonstrated that AIMD simulations can provide accurate hydration structures for a series of divalent and trivalent dissolved cations (Al^{3+} , Ca^{2+} , Cr^{3+} , Mn^{2+} , Fe^{3+} , Co^{2+} , Ni^{2+} , and Zn^{2+}) that yields near-quantitative agreement with their EXAFS spectra [149, 157, 159, 160]. The AIMD-EXAFS [149, 158] method has been applied to several systems including the U^{4+} , U^{5+} , U^{6+} , and Cm^{3+} aqueous species [152, 154, 164]. Bylaska and Ilton recently used the AIMD-EXAFS approach to study U(IV), U(V), and U(VI) substitution into goethite, hematite, and lepidocrocite and the associated defects [143, 161, 163, 165].

The AIMD-EXAFS approach was recently extended to characterize the composition and structure of Zn(II) and Cu(II) point defects in hematite [137, 147] (see Fig. 9). Point defects play an important role in the physics and chemistry of materials including ionic conductivity, diffusion, general reactivity, creep, optical behavior, radiation damage, and many other phenomena. They also play a critical role in regulating the fate and transport of minor yet critical metals, including micronutrients and contaminants, in the environment. For example, at normal background concentrations, Zn(II) is an essential nutrient. However, due to anthropogenic activities, it can be released in large enough concentrations that it can exhibit high levels of toxicity. A salient point is that the potential incorporation of Zn in secondary refractory minerals such as Fe(III) (oxyhydr)oxides can decrease bioavailable Zn below levels indicated by measured total concentrations.

It was previously predicted by Catalano et al. that Zn was immobilized as a tetrahedral coordinated defect in hematite. The predicted structure of this defect was based on EXAFS using a traditional shell by shell fitting procedure. The authors noted that the tetrahedral coordination of Zn in hematite was problematic, as the distances and CN to second-nearest neighbor Fe^{3+} atoms were consistent with Zn substituting for Fe^{3+} in regular octahedral sites. Furthermore, the Debye-Waller factor for first shell Zn-O in hematite was unusually high. This suggested possible unresolved configurational disorder.

In order to resolve these inconsistencies, the fitting of the experimental EXAFS spectra was revisited using an AIMD-EXAFS approach. In contrast to the earlier work, these follow-on studies, which produced fittings with reasonable Debye-Waller factors, indicated the presence of three configurations that were dominated by octahedral coordination.

While AIMD-EXAFS simulations have proven to be useful, and they do not have the limitations of shell-by-shell fitting that (often) leads to simplifying assumptions, they are difficult and time-consuming to carry out. They require setting up and running multiple different AIMD simulations of large unit cells containing defects with specified spin and charge configurations, running thousands of FEFF calculations per trajectory, Fourier transforming and averaging the spectra from each trajectory, and finally comparing the experimental spectra with linear combinations of the calculated fingerprint spectra to extract likely structures. Moreover, depending on the size of the unit cell, each AIMD simulation may need to be run in stages for the AIMD trajectory to be long enough (e.g. each AIMD simulation is restarted 25 times). The following workflow details the steps used to carry out the AIMD-EXAFS simulations of the Zn(II) and Cu(II) in hematite.

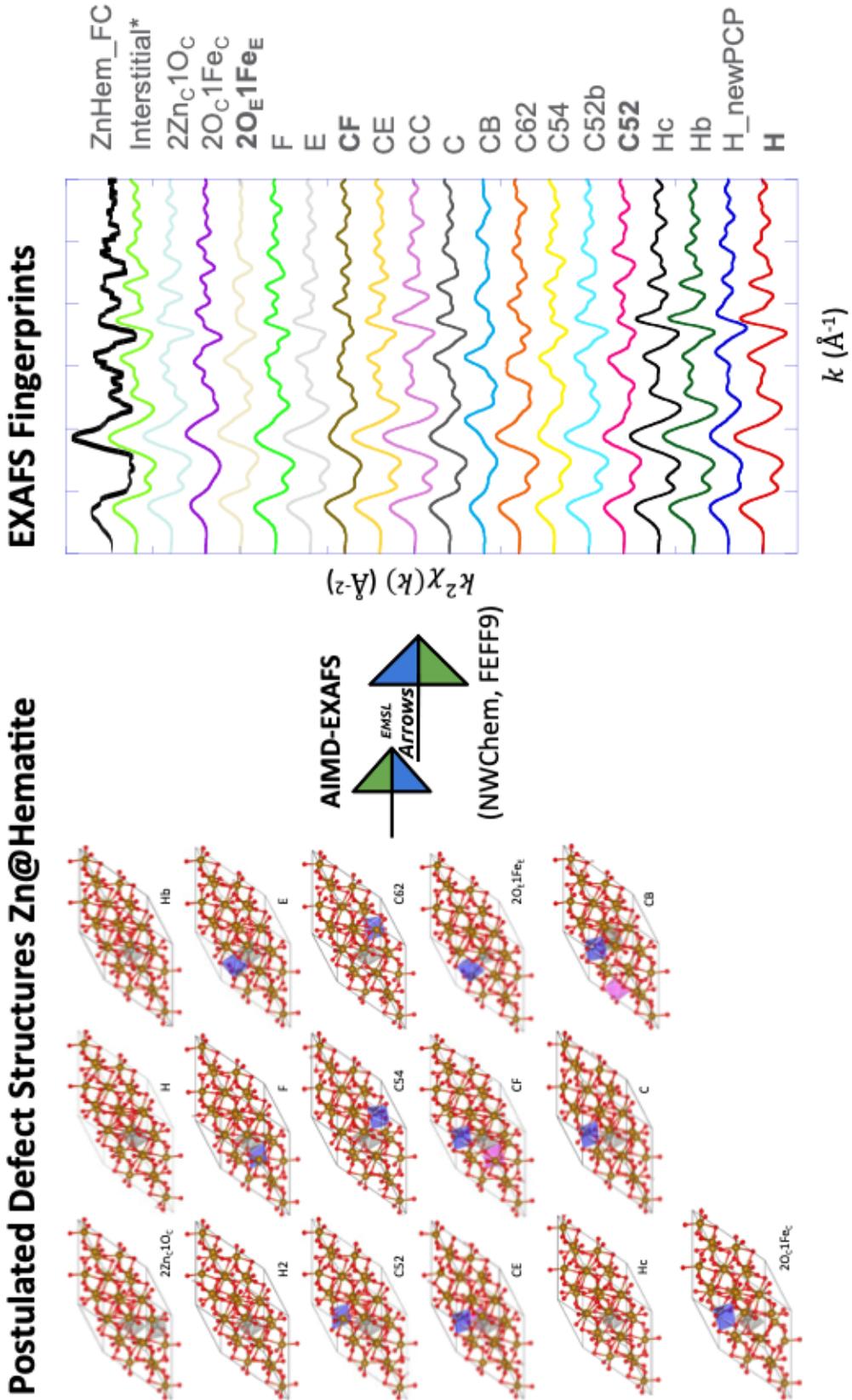


Fig. 9: The simulated EXAFS spectra for 16 possible defect structures of Zn in hematite, excluding a structure for the interstitial site. Note H_newPCP and H are same structure, but spectra was calculated using different pseudopotentials.

AIMD-EXAFS Workflow Using Arrows:

1. Create the unit cell with a defect, set up the AIMD simulation, and submit to the Arrows queue.
2. If using TinyArrows, SSH with port forwarding to a computer center from a machine running the TinyArrows web server.
3. At the external computer center:
 - Fetch the AIMD job from the Arrows queue and submit the job to the computer center's queue.
 - When the simulation is finished, upload simulation results to the Arrows queue. (When uploaded, the Arrows queue processes the results and automatically submits a restart job if more simulation time is needed).
 - Repeat until there are no more jobs on Arrows queue
4. Fetch the trajectory data from the Arrows queue to a local workstation.
5. Parse the trajectory and submit 1000s of EXAFS (FEFF) jobs to the Arrows queue.
6. These jobs are then fetched, run with EXAFS calculations, and results uploaded using available workstations.
7. EXAFS results on the Arrows queue are then collated and averaged into fingerprint spectra for use in analysis of experimental spectra.

This workflow, which heavily utilizes the Arrows queue to move data between different computer systems, was implemented with Python scripts running as cron jobs on various machines. These scripts just run through a simple cycle of repeated steps: (i) List the Arrows queue (`arrows_queue_nwchem_get_url`), (ii) Download an input deck from arrows queue (`arrows_queue_fetch_get_url`), (iii) Run the input deck, (iv) Upload the output deck to the Arrows queue (`arrows_post_url`), (v) Submit the output (`arrows_submit_output_get_url`), and go to (i). Access to the Arrows queue in above workflow and its associated scripts used the HTTPS GET and POST requests described in Arrows Queue subsection in section 3.2. Once the workflow is started, all the succeeding steps were done automatically.

4.2 New insight into Environmental Reactions Using Chemical Reaction Networks

As illustrated previously in Fig. 1, the degradation of many pollutants in natural ground-waters will involve large reaction networks consisting of numerous reaction pathways. Obtaining all the parameters needed from experimental measurements, while hypothetically possible, would end up being time-consuming and not practical for most systems. Even with computation, this can become an arduous task. In this subsection, we demonstrate how Arrows can be used to map out possible reaction pathways for the degradation of a prototypical pollutant, CCl_4 .

Over the years several groups have used the standard tools of quantum chemistry, to estimate the thermodynamics and kinetics of the possible degradation reactions for CCl_4 with a variety of theories and methods, including high-level electronic structure and DFT [83, 84] methods, QM/MM and continuum solvent models, and advanced free energy methods [166–179]. Despite the extensive literature using electronic structure methods to model CCl_4 , these studies were primarily focused on individual reactions or particular classes of reactions, rather than considering the reaction network as a whole. An advantage of Arrows is that it becomes feasible to study reaction networks in their entirety.

For example, one of the motivations of this work has been the "chlorform problem" in which it has been suggested there are competing reaction pathways for the degradation of CCl_4 that can produce unfavorable and favorable metabolites [180–189]. Unfortunately, many groundwater

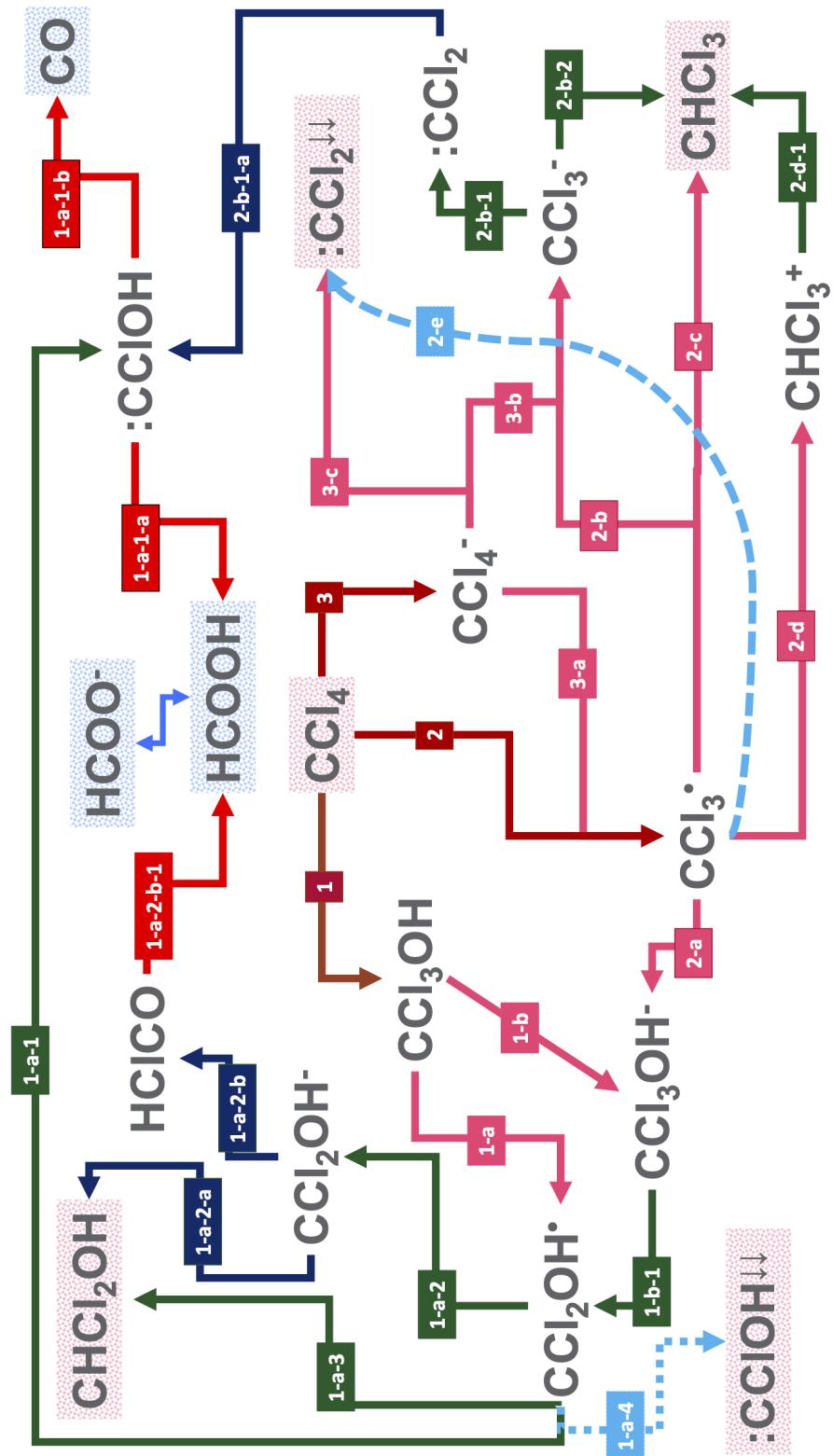


Fig. 10: Hypothesized reaction network for the degradation of CCl_4 in the environment under reducing conditions.

remediation strategies follow an unfavorable pathway in which they produce chloroform (CHCl_3) as the major product and methylene chloride (CH_2Cl_2) as a minor product. Both of these products are nearly as persistent and problematic as the parent compound, but there are competing reaction pathways that could produce more desirable products, carbon monoxide (CO) and/or formic acid (HCOOH). Results scattered throughout the chemical and environmental engineering literature have suggested the branching between these reaction pathways is highly variable, but the controlling factors have not been identified.

In Fig. 10, a hypothesized reaction network, containing 27 elementary reduction and hydrolysis steps, is shown for the first few steps of CCl_4 degradation. This network was designed to capture common elements in the degradation of CCl_4 , regardless of whether it occurs as an intracellular metabolic process, an enzyme-catalyzed process *in vitro*, an abiotic reaction catalyzed on mineral surfaces, or even an electrode reaction. If we understood the fundamental chemistry that controls the branching among these, and related, product-formation pathways, we could improve the applicability of a host of remediation technologies (both chemical and biological) to the large plumes of CCl_4 that contaminate DOE sites across the country.

Thermodynamic calculations for the hypothesized reactions in solution are given in Table 1. Seven different electronic structure methods were used to carry out the calculations, including four DFT [83, 84] calculations using B3LYP [97, 98], PBE [102], PBE0 [99], and M06-2x [100] exchange-correlation functionals with the 6-311++G(2d,2p) [89, 93–96] Gaussian basis set, two plane-wave DFT calculations with PBE and PBE0 exchange-correlation functionals, and frozen core CCSD(T)/6-311++G(2d,2p) calculations. The CCSD(T) [8, 9] calculations used geometries optimized at the B3LYP/6-311++G(2d,2p) level. The COSMO [108] method was used to model the effect of solution in the Gaussian DFT calculations. The solvation energies for the PW DFT and CCSD(T) methods were taken from the PBE, PBE0 and B3LYP COSMO Gaussian DFT calculations respectively. These calculations were performed using the spreadsheet capability in Arrows (see Fig. 6 in subsection 3.1).

Table 1: Aqueous reaction free energies (kcal/mol) for the CCl_4 reaction network given in Fig. 10. The reactions are labeled using both the network reaction labels and *esmiles reaction* input (defined in section 3).

Label	<i>esmiles reaction</i>	B3LYP	PBE	PBE0	M06-2x	CCSD(T)	PW	PBE	PW	PBE0
1	$\text{C(Cl)(Cl)(Cl)Cl} + [\text{OH}^-] \rightarrow \text{C(Cl)(Cl)(Cl)O} + [\text{Cl}^-]$	-53.9	-50.8	-55.1	-56.8	-50.6	-44.5	-50.8		
2	$\text{C(Cl)(Cl)(Cl)Cl} + \text{SHE} \rightarrow [\text{C}](\text{Cl})(\text{Cl})\text{Cl} + [\text{Cl}^-]$	-16.1	-7.2	-9.4	-7.3	-3.0	71.0	-10.1		
3	$\text{C(Cl)(Cl)(Cl)Cl} + \text{SHE} \rightarrow \text{C(Cl)(Cl)(Cl)Cl}^{\{-1\}}$	1.6	8.4	7.3	7.6	14.7	13.9	8.6		
1-a	$\text{C(Cl)(Cl)O} + \text{SHE} \rightarrow [\text{C}](\text{Cl})(\text{Cl})\text{O} + [\text{Cl}^-]$	-10.8	-2.0	-3.9	-2.5	1.8	-2.6	-4.3		
1-b	$\text{C(Cl)(Cl)O} + \text{SHE} \rightarrow [\text{C}](\text{Cl})(\text{Cl})\text{O}^{\{-1\}}$	-11.8	-2.0	3.0	4.1	2.0	0.9	44.0		
2-a	$[\text{C}](\text{Cl})(\text{Cl})\text{Cl} + [\text{OH}^-] \rightarrow \text{C(Cl)(Cl)(Cl)O}^{\{-1\}}$	-49.6	-45.6	-42.6	-45.5	-45.6	-35.2	3.3		
2-b	$[\text{C}](\text{Cl})(\text{Cl})\text{Cl} + \text{SHE} \rightarrow [\text{C}](\text{Cl})(\text{Cl})\text{Cl}^{\{-1\}}$	-6.9	-4.8	-1.5	-7.4	-1.8	1.5	4.5		
2-c	$[\text{C}](\text{Cl})(\text{Cl})\text{Cl} + \text{SHE} + [\text{H}^+] \rightarrow \text{C(Cl)(Cl)Cl}$	-43.8	-39.5	-41.4	-44.2	-43.5	-37.7	-40.0		
2-d	$[\text{C}](\text{Cl})(\text{Cl})\text{Cl} + [\text{H}^+] \rightarrow \text{C(Cl)(Cl)Cl}^{\{+1\}}$	47.9	44.5	51.6	56.9	54.1	43.2	50.8		
2-3	$[\text{C}](\text{Cl})(\text{Cl})\text{Cl} + \text{SHE} \rightarrow [\text{C}](\text{Cl})\text{Cl} \text{ mult3} + [\text{Cl}^-]$	5.6	14.4	11.8	10.2	14.5	14.1	12.1		
3-a	$\text{C(Cl)(Cl)Cl}^{\{-1\}} \rightarrow [\text{C}](\text{Cl})(\text{Cl})\text{Cl} + [\text{Cl}^-]$	-17.7	-15.6	-16.7	-14.9	-17.7	-22.3	-18.7		
3-b	$\text{C(Cl)(Cl)Cl}^{\{-1\}} + \text{SHE} \rightarrow [\text{C}](\text{Cl})(\text{Cl})\text{Cl}^{\{-1\}} + [\text{Cl}^-]$	-24.6	-20.4	-18.2	-22.3	-19.5	-20.7	-14.2		
3-c	$\text{C(Cl)(Cl)Cl}^{\{-1\}} + \text{SHE} \rightarrow [\text{C}](\text{Cl})\text{Cl} \text{ mult3} + 2 [\text{Cl}^-]$	-12.1	-1.2	-4.9	-4.7	-3.2	-8.2	-6.7		
1-a-1	$[\text{C}](\text{Cl})(\text{Cl})\text{O} + \text{SHE} \rightarrow [\text{C}](\text{Cl})\text{O} + [\text{Cl}^-]$	-28.6	-19.9	-18.9	-25.2	-21.8	-15.5	-14.4		
1-a-2	$[\text{C}](\text{Cl})(\text{Cl})\text{O} + \text{SHE} \rightarrow [\text{C}](\text{Cl})(\text{Cl})\text{O}^{\{-1\}}$	-60.9	-58.1	1.3	-67.0	-59.7	2.0	-35.2		
1-a-3	$[\text{C}](\text{Cl})(\text{Cl})\text{O} + \text{SHE} + [\text{H}^+] \rightarrow \text{C(Cl)(Cl)O}$	-46.0	-42.0	-43.7	-46.1	-45.2	-40.6	-42.7		
1-a-4	$[\text{C}](\text{Cl})(\text{Cl})\text{O} + \text{SHE} \rightarrow [\text{C}](\text{Cl})\text{O} \text{ mult3} + [\text{Cl}^-]$	10.0	18.6	16.1	14.5	18.6	18.5	22.3		
1-b-1	$[\text{C}](\text{Cl})(\text{Cl})\text{O}^{\{-1\}} \rightarrow [\text{C}](\text{Cl})(\text{Cl})\text{O} + [\text{Cl}^-]$	1.0	0.0	-7.0	-6.6	-0.2	-3.4	-48.3		
2-b-1	$[\text{C}](\text{Cl})(\text{Cl})\text{Cl}^{\{-1\}} \rightarrow [\text{C}](\text{Cl})\text{Cl} + [\text{Cl}^-]$	-7.4	-2.0	-3.0	-2.4	-5.1	-3.4	-4.0		
2-b-2	$[\text{C}](\text{Cl})(\text{Cl})\text{Cl}^{\{-1\}} + [\text{H}^+] \rightarrow \text{C(Cl)(Cl)Cl}$	-37.0	-34.7	-39.9	-36.8	-41.7	-39.2	-44.5		
2-d-1	$\text{C(Cl)(Cl)Cl}^{\{+1\}} + \text{SHE} \rightarrow \text{C(Cl)(Cl)Cl}$	-91.8	-84.0	-92.9	-101.1	-97.6	-80.9	-90.8		
1-a-1-a	$[\text{C}](\text{Cl})\text{O} + [\text{OH}^-] \rightarrow \text{C}(=\text{O})\text{O} + [\text{Cl}^-]$	-96.0	-92.8	-98.5	-99.0	-92.7	-89.1	-96.9		
1-a-1-b	$[\text{C}](\text{Cl})\text{O} \rightarrow [\text{C}][\text{O}] + [\text{H}^+] + [\text{Cl}^-]$	-61.9	-52.9	-56.4	-64.2	-63.2	-53.9	-55.8		
1-a-2-a	$[\text{C}](\text{Cl})(\text{Cl})\text{O}^{\{-1\}} + [\text{H}^+] \rightarrow \text{C(Cl)(Cl)O}$	14.8	16.1	-44.9	20.9	14.4	-44.9	-7.5		
1-a-2-b	$[\text{C}](\text{Cl})(\text{Cl})\text{O}^{\{-1\}} \rightarrow \text{C}(=\text{O})\text{Cl} + [\text{Cl}^-]$	-10.6	-4.9	-63.5	0.3	-4.1	-64.4	-24.0		
1-b-1-a	$[\text{C}](\text{Cl})\text{Cl} + [\text{OH}^-] \rightarrow [\text{C}](\text{Cl})\text{O} + [\text{Cl}^-]$	-62.8	-58.8	-64.0	-67.5	-60.7	-52.4	-59.8		
1-a-2-b-1	$\text{C}(=\text{O})\text{Cl} + [\text{OH}^-] \rightarrow \text{C}(=\text{O})\text{O} + [\text{Cl}^-]$	-53.1	-49.8	-55.1	-57.5	-50.8	-44.5	-52.0		

The reaction energies fall into three classes: strongly exothermic ($G_{rxn} \ll -25$ kcal/mol), moderately exothermic (-25 kcal/mol $\lesssim G_{rxn} \lesssim -10$ kcal/mol), and mildly endothermic (0 kcal/mol $\lesssim G_{rxn} \lesssim 18$ kcal/mol). For many of the reactions, the different levels of theory have results that are in reasonable agreement with each other. However, there are several reactions highlighted in "red" in Table 1 (i.e. the 1-b, 2-a, 1-a-2, 1-b-1, 1-a-2-a, and 1-a-2-b reactions) in which the different methods produce energies that are up to 60 kcal/mol different. What these reactions have in common is that they contain either a CCl_3OH^- or CCl_2OH^- negatively charged alcohol metabolite. On closer inspection of these two species, it was observed that it is energetically favorable to fragment them into a carbon monoxide and chloride species. The discrepancies between different methods was due to the fact that in some cases, the geometry optimization trapped into a highly energetic local minimum that stayed together, and in other cases the optimization immediately fell into a lower energy structure that fragmented. These energetic results suggest that there are missing reactions in the reaction network given in Fig. 10 that are significant and may help explain the origin of competing pathways in which carbon monoxide is produced instead of chloroform.

AIMD and AIMD/MM simulations were used to check the stability of these high-energy local minima. Simulations of CCl_3OH^- in the gas-phase and solution using AIMD and AIMD/MM [12, 13] simulations are shown in Fig. 11. The "Expert Periodic and Molecular Builder" in Arrows (<https://arrows.emsl.pnnl.gov/api/periodic> or <http://localhost:5000/api/periodic>) was used to set up and run the simulations. Results from both these simulations clearly showed that CCl_3OH^- immediately decomposes into carbon monoxide and chloride fragments. AIMD and AIMD/MM simulations for the radical ion, CCl_2OH^- , also showed that it immediately decomposed as well (not shown). Both these sets of results are in agreement with the observation that it is energetically favorable to fragment the negatively charged trichloromethanol and dichloromethanol species into a carbon monoxide and chloride species, further confirming that electron-transfer to these chloro-alcohol species may be what controls the presence of alternative pathways in the "chloroform problem".

4.3 Reaction Pathways

+As illustrated in the previous section, it is quite common to hypothesize a chemical reaction network that contains multiple branching pathways that are all thermodynamically favorable. As a result of this, relying only on reaction energies can be limiting, and approaches to modeling reaction kinetics, e.g. transition-states and reaction pathways are needed. Unfortunately, these types of calculations involve difficult optimizations that can easily fail. Even in best case scenarios with an expert user running the calculation, this type of calculation will still end up being 10-20 times more expensive than a reaction energy calculation. Modeling reactions in solution is even worse. In addition, transition-states usually contain non-bonding electronic states that are not well described by lower levels of electronic structure theories, and moreover many reactions, even intrinsic one-step reactions, end up having multiple pathways containing multiple barriers. In short, these calculations are time-consuming and difficult, and not surprisingly, automating calculations for transition-states and reaction pathways is an active area of research. Even though transition-state and reaction pathway calculation are not completely automated in Arrows, there are several workflows implemented in Arrows that can be used to perform these types of simulations (see the online Arrows manuals https://nwchemgit.github.io/EMSL_Arrows.html# or <https://ebylaska.github.io/TinyArrows/>).

One of the more important reactions, in which reaction pathway calculations are needed, that affects the environmental fate of most organic compounds, is alkaline hydrolysis. Owing

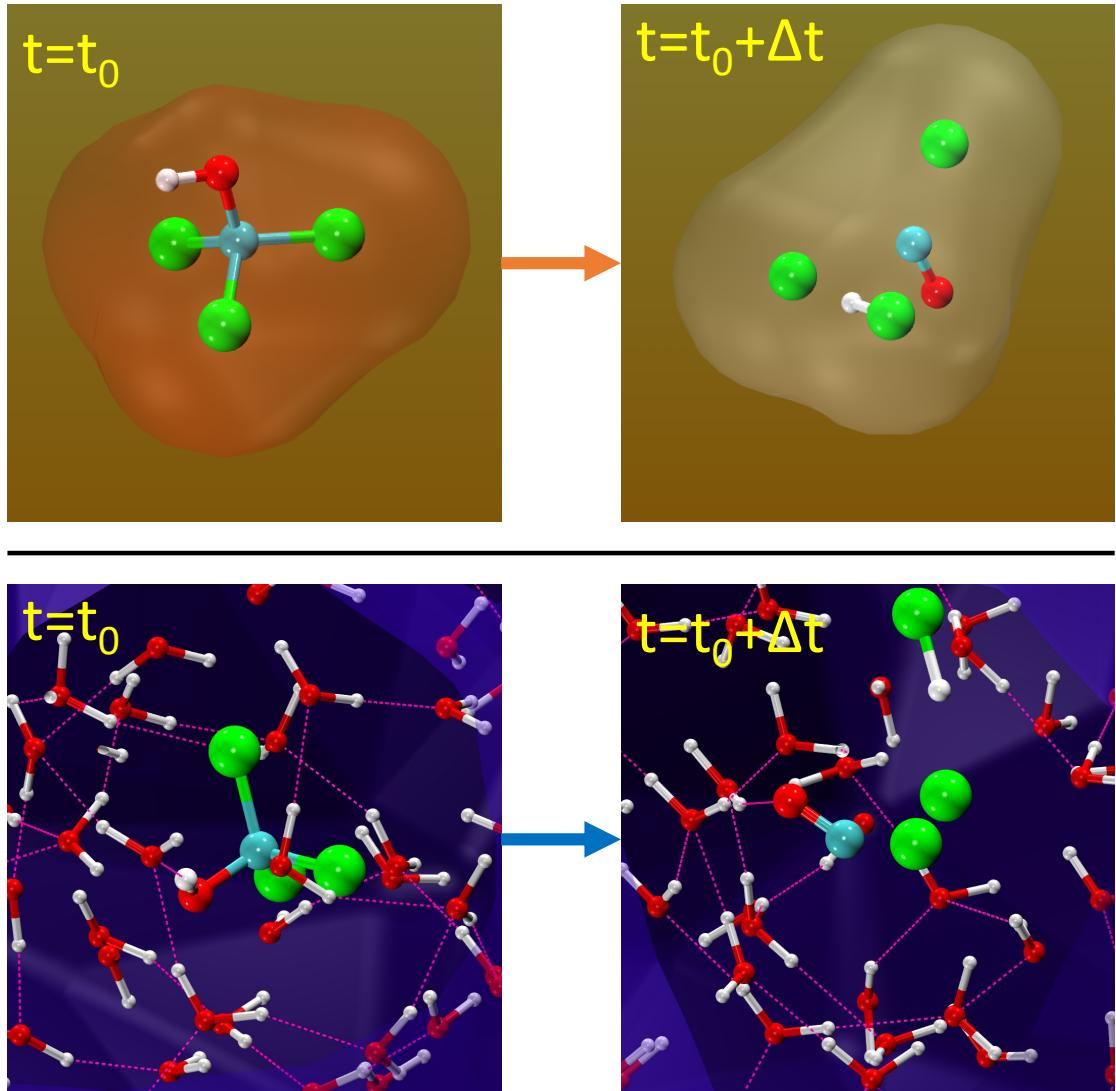


Fig. 11: Gas-phase AIMD (top panels) and solution phase AIMD/MM (bottom panels) simulations of the decomposition of $\text{CCl}_3\text{OH}^{-1}$.

to its ever-presence, it is often used as a baseline against which the significance of all other environmental processes are considered. Typically, for many molecules, hydrolysis can proceed via multiple different reactions (e.g., nucleophilic aromatic substitution, hydrogen abstraction, and Meisenheimer addition reactions are important reactions for nitroaromatic hydrolysis) that are all thermodynamically favorable, and thus it is a type of reaction for which transition-states and reaction pathways are needed.

To illustrate using Arrows to calculate a reaction pathway for the alkaline hydrolysis of 2,4-dinitroanisole (DNAN) in solution by nucleophilic aromatic substitution are shown in Fig. 12. These results show that the $\text{2,4-dinitroanisole} + \text{OH}^- \rightarrow \text{2-methoxy-5-nitrophenol} + \text{NO}_2^-$ nucleophilic aromatic reaction, while strongly exothermic, has a high barrier at both the PBE [102] and

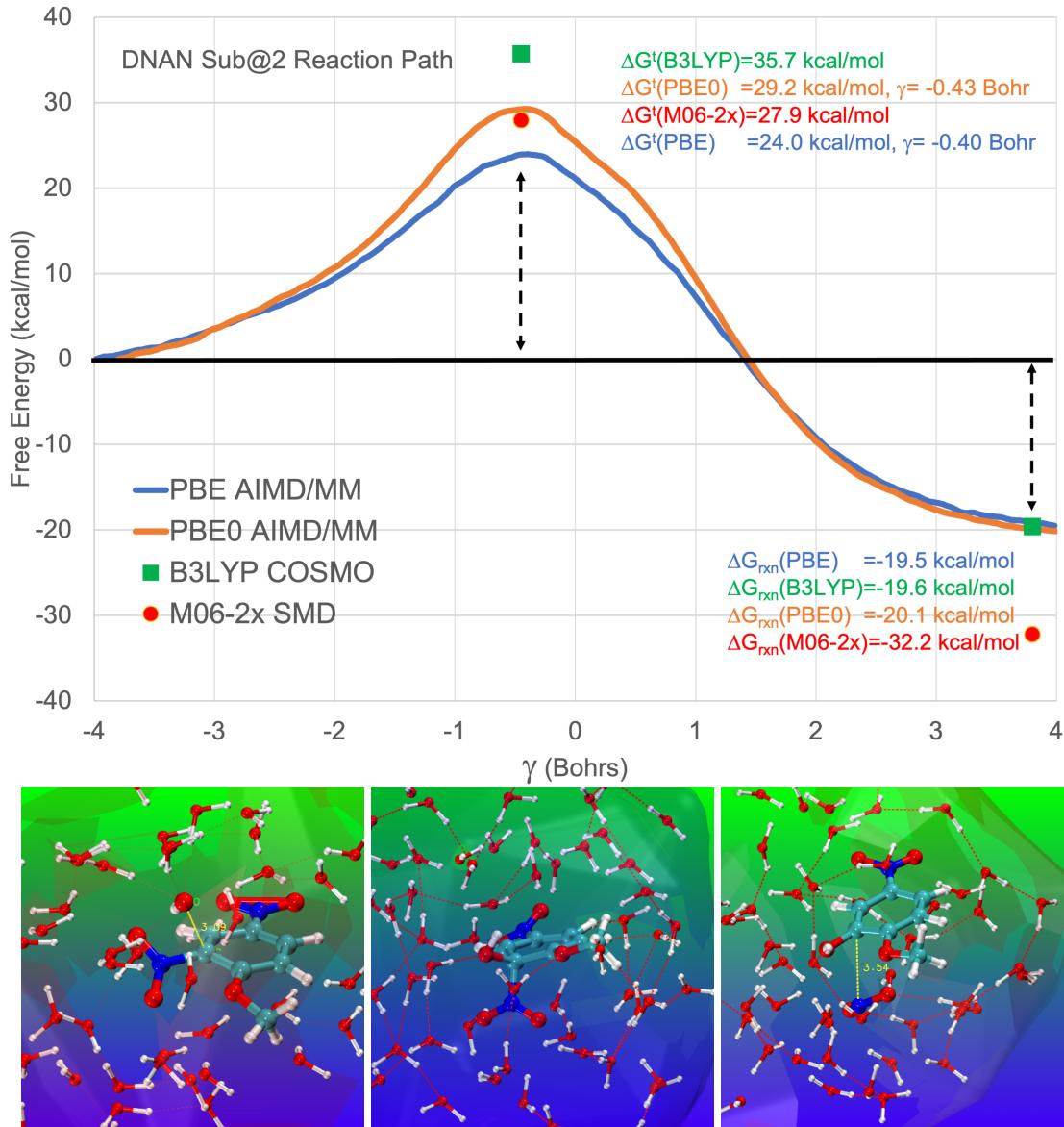


Fig. 12: (Top) Free energy reaction pathway for the $2,4\text{-dinitroanisole} + \text{OH}^- \rightarrow 2\text{-methoxy-5-nitrophenol} + \text{NO}_2^-$ nucleophilic aromatic reaction of 2,4-dinitroanisole (DNAN). Also shown in the figure are results of the reaction energy and barrier from the prior calculations of Salter-Blanc et al. [190] and Hill et al. [191], which used the COSMO [108] and SMD [192] solvent models respectively to describe solvation. (Bottom) Snapshots from different windows of the WHAM simulations for the reaction.

PBE0 [99] levels (Note the barrier height from the PBE0 simulation is expected to be more accurate than the PBE simulation, which is usually underestimated) suggesting that it will proceed very slowly and likely be out competed by other hydrolysis reactions. These reaction pathway simulations were performed with the weighted histogram analysis method (WHAM) [193] using the AIMD/MM [12,13] method with PBE96 [102] and PBE0 [99] exchange correlation function-

als in a bath of 56 water molecules (i.e., 1 M concentration). A short write-up explaining the WHAM method and how to use it with NWChem is available for download [194]. The “bonds” collective variable, which is defined as the sum of the bond distances (in Bohr) of the bonds destroyed subtracted from the sum of bond distances (in Bohr) of the bonds created during the reaction, was used for the reaction coordinate in these calculations.

WHAM simulations are another example of a molecular simulation that can benefit from the workflow capabilities in Arrows. Examples of how to run WHAM and other types of free energy simulations, including potential of mean force (PMF), metadynamics, and temperature accelerated molecular dynamics (TAMD), can be found in the online Arrows manuals (https://nwchemgit.github.io/EMSL_Arrows.html# or <https://ebylaska.github.io/TinyArrows/>).

4

5 Conclusions

This chapter provides a brief overview and demonstrations of the various functionalities of the Arrows web-based tools. These tools combine NWChem, SQL [47, 48] and NoSQL [49] databases, email, back-end web servers, external web services, and web applications, and it was designed to carry out complex and difficult molecular and materials simulations. It was also designed to provide a starting infrastructure for researchers in molecular simulation to investigate and prototype intelligently designed web interfaces. There are two implementations of Arrows called EMSL Arrows and TinyArrows, which are freely available. EMSL Arrows is hosted by the William R. Wiley Environmental Molecular Sciences Laboratory at PNNL, and TinyArrows is a stand-alone implementation of Arrows available on GitHub.

Currently, Arrows can be used to calculate reaction thermodynamics and reaction pathways for molecular systems, and broad range of advanced computational chemistry simulations (e.g., IR, NMR, EXAFS spectra, and free energy reaction pathways). Arrows has potential benefits to both users and developers of molecular simulation software. For users, it makes molecular simulation more accessible, and it gives them the ability to design and carry out complex scientific workflows (e.g., AIMD-EXAFS [149,158], chemical reaction networks, AIMD and AIMD/MM [12, 13] simulations, AIMD free energy simulations). Given that it is easy to run large numbers of calculations and carry out complex workflows, another potential benefit of Arrows is to help developers both improve and find bugs in their computational chemistry codes.

In the future, there are many ways that Arrows and other web-based tools could be improved. As new molecular (exascale) simulation methods are able to handle thousands to millions of atoms, even at the level of DFT, new ways to enter large scale chemical and materials systems will be needed. As discussed in section 3.2, reaction path generators can be made more automated, and even generalized to be reaction predictors, using algorithms and techniques from AI and ML. Finally, several improvements could be made in modeling chemical reaction networks. In the near future, it should become not only possible to routinely map out chemical reaction networks, but extend them with ML surrogate models that are differentiable, based on AIMD and other first principles rare event modeling. Furthermore, if AIMD rare-event modeling is coupled with multi-physics and mesoscopic models with continuum descriptions, such as classical density-functional theory, it could be possible to enable the optimization of reaction environmental conditions to obtain desired critical pathways through a chemical reaction network and vice versa (e.g., optimize catalytic processes, infer geochemical environments, develop new approaches for environmental degradation, and design new chemical processes for polymer upcycling that are able to destroy the macromolecular structures in plastics).

Finally, the goals of computational chemistry automation, as implemented by Arrows, are quickly evolving. In our experience, Arrows, which was initially designed to make complex calcu-

lations easier to carry out, reduce input errors, detect computational errors, aid reproducibility, provide a log of calculation workflows, and visualize results, is now allowing scientists to carry out new types of combined scientific studies, where computation and experiment are used either in tandem or in sequence to test a scientific hypothesis.

Acknowledgements

This work was supported by the NWChem project in the DOE BER William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), the DoD Strategic Environmental Research and Development Program (SERDP) under ER1735 and ER-2725, the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research ECP program (NWChemEx project), and was also supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division (Sandia-Livermore CCS, PNNL QIS, and PNNL Geosciences projects) DE-AC05-76RL01830. EMSL operations are supported by the DOE's Office of Biological and Environmental Research. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a User Facility supported by the Office of Science of the U.S. DOE under Contract No. DE-AC02-05CH11231, to carry out the AIMD-EXAFS studies highlighted in this work. This report has not been subject to review by SERDP and therefore does not necessarily reflect their views and no official endorsement should be inferred.

Author Contributions

E.J.B. was the primary author of the manuscript, and he developed the EMSL Arrows and Tiny Arrows software. D.S. and T.L.S helped test and debug the Arrows software. All authors have read and contributed to the editing of the manuscript, and all authors have provided feedback to the design and implementation of Arrows. E.J.B. and D.S. carried out the AIMD-EXAFS simulations and E.S.I analyzed the AIMD-EXAFS results given in section 4.1. T.L.S. and P.G.T. generated Fig. 1. S.O. with E.J.B. led the deployment of EMSL Arrows at Pacific Northwest National Laboratory. S.O. was also responsible for installing and testing the HTTP security protocols used for accessing the EMSL Arrows Web APIs.

Conflict of Interest

The authors declare that they have no competing interests.

References

1. Don Tiffany Cromer and James Thomas Waber. Scattering factors computed from relativistic Dirac-Slater wave functions. *Acta Crystallographica*, 18(1):104–109, 1965.
2. Gerhard Goldbeck. The economic impact of molecular modelling of chemicals and materials. Goldbeck Consulting, <https://materialsmodelling.com/wp-content/uploads/2014/01/the-economic-impact-of-modelling.pdf>, 2012. Accessed: 2021-07-31.
3. Ricky A Kendall, Edoardo Aprà, David E Bernholdt, Eric J Bylaska, Michel Dupuis, George I Fann, Robert J Harrison, Jialin Ju, Jeffrey A Nichols, Jarek Nieplocha, et al. High performance computational chemistry: An overview of NWChem a distributed parallel application. *Computer Physics Communications*, 128(1-2):260–283, 2000.
4. Marat Valiev, Eric J Bylaska, Niranjan Govind, Karol Kowalski, Tjerk P Straatsma, Hubertus JJ Van Dam, Dunyou Wang, Jarek Nieplocha, Edoardo Aprà, Theresa L Windus, et al. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Computer Physics Communications*, 181(9):1477–1489, 2010.
5. Edoardo Aprà, Eric J Bylaska, Wibe A De Jong, Niranjan Govind, Karol Kowalski, Tjerk P Straatsma, Marat Valiev, Hubertus JJ van Dam, Yuri Alexeev, James Anchell, et al. NWChem: Past, present, and future. *Journal of Chemical Physics*, 152(18):184102, 2020.
6. Karol Kowalski, Raymond Bair, Nicholas P Bauman, Jeffery S Boschen, Eric J Bylaska, Jeff Daily, Wibe A de Jong, Thom Dunning Jr, Niranjan Govind, Robert J Harrison, et al. From NWChem to NWChemEx: Evolving with the Computational Chemistry Landscape. *Chemical Reviews*, 121(8):4962–4998, 2021.
7. Chr. Møller and Milton S. Plesset. Note on an approximation treatment for many-electron systems. *Physical Review*, 46(7):618, 1934.
8. Rodney J Bartlett and John F Stanton. Applications of Post-Hartree–Fock Methods: A Tutorial. *Reviews in Computational Chemistry*, pages 65–169, 1994.
9. Rodney J Bartlett and Monika Musiał. Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics*, 79(1):291, 2007.
10. Michael P Allen and Dominic J Tildesley. *Computer Simulation of Liquids*. Oxford university press, 2017.
11. Eric J Bylaska. Plane-Wave DFT methods for Chemistry. In *Annual Reports in Computational Chemistry*, volume 13, pages 185–228. Elsevier, 2017.
12. Alessandro Laio, Joost VandeVondele, and Ursula Rothlisberger. A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations. *Journal of Chemical Physics*, 116(16):6941–6947, 2002.
13. Emilie Cauët, Stuart Bogatko, John H Weare, John L Fulton, Gregory K Schenter, and Eric J Bylaska. Structure and dynamics of the hydration shells of the Zn²⁺ ion from ab initio molecular dynamics and combined ab initio and classical molecular dynamics simulations. *Journal of Chemical Physics*, 132(19):194502, 2010.
14. Tifany L Torralba-Sanchez, Eric J Bylaska, Alexandra J Salter-Blanc, Douglas E Meisenheimer, Molly A Lyon, and Paul G Tratnyek. Reduction of 1,2,3-trichloropropane (TCP): Pathways and mechanisms from computational chemistry calculations. *Environmental Science: Processes & Impacts*, 22(3):606–616, 2020.
15. CK Johnson. Ortep. Report ORNL-3794. Oak Ridge National Laboratory, 1965.
16. C Wasikowski and S Klemm. XMol Version 1.3. 1, 1990, 1991, 1992, 1993 Research Equipment, Inc. d/b/a Minnesota Supercomputer Center, Inc. <https://www.msi.umn.edu/>, 1990, 1991, 1992, 1993.
17. J Labanowski and E Wimmer. UniChem software package, Cray Research Inc., 1992.
18. MT Michalewicz, VM Norling, IE McCarthy, and M Brunger. Combining UniChem and electron momentum spectroscopy to study valence electronic structure of molecules. Technical report, Flinders Univ. of South Australia, 1995.
19. Kim K Baldridge and Jerry P Greenberg. QMView: A computational chemistry three-dimensional visualization tool at the interface between molecules and mankind. *Journal of Molecular Graphics*, 13(1):63–66, 1995.

20. Thomas James Smith. MolView: A program for analyzing and displaying atomic structures on the Macintosh personal computer. *Journal of Molecular Graphics*, 13(2):122–125, 1995.
21. William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996.
22. Brett M Bode and Mark S Gordon. MacMolPlt: A graphical user interface for GAMESS. *Journal of Molecular Graphics and Modelling*, 16(3):133–138, 1998.
23. Anton Kokalj. XCrySDen—A new program for displaying crystalline structures and electron densities. *Journal of Molecular Graphics and Modelling*, 17(3-4):176–179, 1999.
24. A Frisch, AB Nielson, AJ Holder, et al. Gaussview User Manual. *Gaussian Inc., Pittsburgh, PA*, 556, 2000.
25. Gijs Schaftenaar and Jan H Noordik. Molden: A pre-and post-processing program for molecular and electronic structures. *Journal of Computer-Aided Molecular Design*, 14(2):123–134, 2000.
26. Stefan Portmann and Hans Peter Lüthi. MOLEKEL: An interactive molecular graphics tool. *CHIMIA International Journal for Chemistry*, 54(12):766–770, 2000.
27. Jason D Gans and David Shalloway. Qmol: a program for molecular visualization on Windows-based PCs. *Journal of Molecular Graphics and Modelling*, 19(6):557–559, 2001.
28. Sune R Bahn and Karsten W Jacobsen. An object-oriented scripting interface to a legacy electronic structure code. *Computing in Science & Engineering*, 4(3):56–66, 2002.
29. Karen Schuchardt, Brett Didier, and Gary Black. ECCE — A problem-solving environment’s evolution toward Grid services and a Web architecture. *Concurrency and Computation: Practice and Experience*, 14(13-15):1221–1239, 2002.
30. Anton Kokalj. Computer graphics and graphical user interfaces as tools in simulations of matter at the atomic scale. *Computational Materials Science*, 28(2):155–168, 2003.
31. Robert J Gdanitz, Gary D Black, Carina S Lansing, Bruce J Palmer, and Karen L Schuchardt. Registering the amica electronic structure code in the extensible computational chemistry environment. *Journal of Computational Chemistry*, 26(3):214–225, 2005.
32. T. Verstraelen, V. Van Speybroeck, and M. Waroquier. ZEOBUILDER: A GUI Toolkit for the Construction of Complex Molecular Structures on the Nanoscale with Building Blocks. *Journal of Chemical Information and Modeling*, 48(7):1530–1541, 2008. PMID: 18543904.
33. J Thomas and P Sherwood. CCP1-GUI, A general-purpose visualization code for electronic structure codes. *version 0.8, Obtained from http://sourceforge.net/projects/ccp1gui/[Last accessed: 12 September 2014] Search PubMed*, 2010.
34. Robert M Hanson. Jmol—A paradigm shift in crystallographic visualization. *Journal of Applied Crystallography*, 43(5):1250–1260, 2010.
35. Koichi Momma and Fujio Izumi. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *Journal of Applied Crystallography*, 44(6):1272–1276, 2011.
36. Robert M Hanson, Jaime Prilusky, Zhou Renjian, Takanori Nakane, and Joel L Sussman. Jsmol and the next-generation web-based representation of 3d molecular structure as applied to proteopedia. *Israel Journal of Chemistry*, 53(3-4):207–216, 2013.
37. Marcus D Hanwell, Chris Harris, Alessandro Genova, Mojtaba Haghatali, Muammar El Khatib, Patrick Avery, Johannes Hachmann, and Wibe Albert de Jong. Open Chemistry, JupyterLab, REST, and quantum chemistry. *International Journal of Quantum Chemistry*, 121(1):e26472, 2021.
38. Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, et al. The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, 2017.
39. Hai Nguyen, David A Case, and Alexander S Rose. NGLview—interactive molecular graphics for Jupyter notebooks. *Bioinformatics*, 34(7):1241–1242, 2018.
40. Søren Smidstrup, Troels Markussen, Pieter Vancraeyveld, Jess Wellendorff, Julian Schneider, Tue Gunst, Brecht Verstichel, Daniele Stradi, Petr A Khomyakov, Ulrik G Vej-Hansen, et al. QuantumATK: An integrated platform of electronic and atomic-scale modelling tools. *Journal of Physics: Condensed Matter*, 32(1):015901, 2019.
41. Andrew Gilbert. Introduction to IQmol. <http://www.iqmol.org/downloads/IQmolUserGuide.pdf>, 2019. Accessed: 2021-07-31.

42. Bruno Bienfait and Peter Ertl. JSME: a free molecule editor in JavaScript. *Journal of Cheminformatics*, 5(1):1–6, 2013.
43. Mingyang Chen, Amanda C Stott, Shenggang Li, and David A Dixon. Construction of a robust, large-scale, collaborative database for raw data in computational chemistry: The collaborative chemistry database tool (ccdbt). *Journal of Molecular Graphics and Modelling*, 34:67–75, 2012.
44. Peter Murray-Rust and Henry S Rzepa. CML: Evolution and design. *Journal of Cheminformatics*, 3(1):1–15, 2011.
45. Peter Murray-Rust, Joe A Townsend, Sam E Adams, Weerapong Phadungsukanan, and Jens Thomas. The semantics of Chemical Markup Language (CML): dictionaries and conventions. *Journal of Cheminformatics*, 3(1):1–12, 2011.
46. Wibe A De Jong, Andrew M Walker, and Marcus D Hanwell. From data to analysis: linking NWChem and Avogadro with the syntax and semantics of Chemical Markup Language. *Journal of Cheminformatics*, 5(1):1–12, 2013.
47. Donald D Chamberlin and Raymond F Boyce. Sequel: A structured english query language. In *Proceedings of the 1974 ACM SIGFIDET (now SIGMOD) workshop on Data description, access and control*, pages 249–264, <https://dl.acm.org/doi/proceedings/10.1145/800296>, Ann Arbor, MI, 1974.
48. Carolyn J Hursch and Jack L Hursch. *SQL: The Structured Query Language*. Tab Books, New York, 1988.
49. Michael Stonebraker. SQL databases v. NoSQL databases. *Communications of the ACM*, 53(4):10–11, 2010.
50. Eric J Bylaska. EMSL Arrows. <https://arrows.emsl.pnnl.gov/api/>, 2021. Accessed: 2021-08-31.
51. Nancy S Foster. The Environmental Molecular Sciences Laboratory, a USDOE national scientific user facility. In *60th Northwest Regional ACS Meeting*, 2005.
52. Pacific Northwest National Laboratory. Environmental Molecular Sciences Laboratory (EMSL), A DOE Office of Science User Facility. <https://www.emsl.pnnl.gov/>, 2021. Accessed: 2021-08-31.
53. Karan Patel. Incremental journey for World Wide Web: Introduced with Web 1.0 to recent Web 5.0—A survey paper. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(10), 2013.
54. Greg O'Connor. Moore's law gives way to Bezos's law. GigaOm, <https://gigaom.com/2014/04/19/moores-law-gives-way-to-bezos-law/>. Accessed: 2021-07-31.
55. 2600Hz Blog. Bezos law. <http://blog.2600hz.com/bezos-law>. Accessed: 2021-07-31.
56. Qazi Zia Ullah, Shahzad Hassan, and Gul Muhammad Khan. Adaptive resource utilization prediction system for infrastructure as a service cloud. *Computational intelligence and neuroscience*, 2017, 2017.
57. Joe Weinman. The nuances of cloud economics. *IEEE Cloud Computing*, 1(4):88–92, 2014.
58. Jimmy Lin. Is big data a transient problem? *IEEE Internet Computing*, 19(5):86–90, 2015.
59. W3C. Html living standard. <https://html.spec.whatwg.org/>, 2021. Accessed: 2021-08-31.
60. Brian Danchilla. *Beginning WebGL for HTML5*. Apress, 2012.
61. Nicholas Rego and David Koes. 3Dmol.js: Molecular visualization with WebGL. *Bioinformatics*, 31(8):1322–1324, 2015.
62. Julius Adorf. Web speech API, KTH Royal Institute of Technology. <https://www.juliusadorf.com/pub/web-speech-api.pdf>, 2013. Accessed: 2021-07-31.
63. Mark J Collins. Indexed DB. In *Pro HTML5 With Visual Studio 2012*, pages 255–279. Springer, New York, 2012.
64. Stefan Kimak and Jeremy Ellman. The role of HTML5 IndexedDB, the past, present and future. In *2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)*, pages 379–383. IEEE, 2015.
65. David Weininger. SMILES, A chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1):31–36, 1988.
66. David Weininger, Arthur Weininger, and Joseph L Weininger. SMILES. 2. Algorithm for generation of unique SMILES notation. *Journal of Chemical Information and Computer Sciences*, 29(2):97–101, 1989.

67. Henri A Favre and Warren H Powell. *Nomenclature of organic chemistry: IUPAC recommendations and preferred names 2013*. Royal Society of Chemistry, London, 2013.
68. Stephen R Heller and Alan D McNaught. The IUPAC international chemical identifier (InChI). *Chemistry International*, 31(1):7, 2009.
69. Stephen Heller, Alan McNaught, Stephen Stein, Dmitrii Tchekhovskoi, and Igor Pletnev. InChI—the worldwide chemical structure identifier standard. *Journal of Cheminformatics*, 5(1):1–9, 2013.
70. Stephen R Heller, Alan McNaught, Igor Pletnev, Stephen Stein, and Dmitrii Tchekhovskoi. InChI, the IUPAC international chemical identifier. *Journal of Cheminformatics*, 7(1):1–34, 2015.
71. Christopher Southan. InChI in the wild: an assessment of InChIKey searching in Google. *Journal of Cheminformatics*, 5(1):1–10, 2013.
72. John D Westbrook and Paula MD Fitzgerald. The PDB format, mmCIF formats, and other data formats. *Structural bioinformatics*, 44:159–179, 2003.
73. Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L Chevrier, Kristin A Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.
74. Sydney R Hall, Frank H Allen, and I David Brown. The crystallographic information file (cif): a new standard archive file for crystallography. *Acta Crystallographica Section A: Foundations of Crystallography*, 47(6):655–685, 1991.
75. Minoru Kanehisa and Susumu Goto. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.
76. Sunghwan Kim, Paul A Thiessen, Evan E Bolton, Jie Chen, Gang Fu, Asta Gindulyte, Lianyi Han, Jane He, Siqian He, Benjamin A Shoemaker, et al. Pubchem substance and compound databases. *Nucleic acids research*, 44(D1):D1202–D1213, 2016.
77. Antony Williams and Valery Tkachenko. The Royal Society of Chemistry and the delivery of chemistry data repositories for the community. *Journal of Computer-Aided Molecular Design*, 28(10):1023–1030, 2014.
78. Meredith Ayers. ChemSpider: The free chemical database. *Reference Reviews*, 26(7), <https://www.emerald.com/insight/content/doi/10.1108/09504121211271059/full/html>, 2012.
79. Anna Gaulton, Anne Hersey, Michał Nowotka, A Patricia Bento, Jon Chambers, David Mendez, Prudence Mutowo, Francis Atkinson, Louisa J Bellis, Elena Cibrián-Uhalte, et al. The ChEMBL database in 2017. *Nucleic Acids Research*, 45(D1):D945–D954, 2017.
80. Amos Bairoch. The ENZYME database in 2000. *Nucleic acids research*, 28(1):304–305, 2000.
81. David S Wishart, Craig Knox, An Chi Guo, Savita Srivastava, Murtaza Hassanali, Paul Stothard, Zhan Chang, and Jennifer Woolsey. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 34(suppl_1):D668–D672, 2006.
82. Mariette Hellenbrandt. The inorganic crystal structure database (ICSD)—present and future. *Crystallography Reviews*, 10(1):17–22, 2004.
83. Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B):B864, 1964.
84. Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133, 1965.
85. CCJ Roothaan. Self-consistent field theory for open shells of electronic systems. *Reviews of modern physics*, 32(2):179, 1960.
86. Michael W Schmidt and Mark S Gordon. The construction and interpretation of MCSCF wavefunctions. *Annual review of physical chemistry*, 49(1):233–266, 1998.
87. Attila Szabo and Neil S Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012.
88. I Ge Ross. Calculations of the energy levels of acetylene by the method of antisymmetric molecular orbitals, including σ - π interaction. *Transactions of the Faraday Society*, 48:973–991, 1952.
89. Michelle M. Francl, William J. Pietro, Warren J. Hehre, J. Stephen Binkley, Mark S. Gordon, Douglas J. DeFrees, and John A. Pople. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *Journal of Chemical Physics*, 77:3654–3665, 1982.

90. Mark S. Gordon, J. Stephen Binkley, John A. Pople, William J. Pietro, and Warren J. Hehre. Self-consistent molecular-orbital methods. 22. small split-valence basis sets for second-row elements. *Journal of the American Chemical Society*, 104:2797–2803, 1982.
91. P. C. Hariharan and J. A. Pople. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretical Chemistry Accounts*, 28:213–222, 1973.
92. W. J. Hehre, R. Ditchfield, and J. A. Pople. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *Journal of Chemical Physics*, 56:2257–2261, 1972.
93. Timothy Clark, Jayaraman Chandrasekhar, Günther W. Spitznagel, and Paul Von Ragué Schleyer. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li-F. *Journal of Computational Chemistry*, 4:294–301, 1983.
94. R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople. Self-consistent molecular orbital methods. xx. a basis set for correlated wave functions. *Journal of Chemical Physics*, 72:650–654, 1980.
95. A. D. McLean and G. S. Chandler. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11–18. *Journal of Chemical Physics*, 72:5639–5648, 1980.
96. Günther W. Spitznagel, Timothy Clark, Paul von Ragué Schleyer, and Warren J. Hehre. An evaluation of the performance of diffuse function-augmented basis sets for second row elements, Na-Cl. *Journal of Computational Chemistry*, 8:1109–1116, 1987.
97. Axel D Beck. Density-functional thermochemistry. III. The role of exact exchange. *Journal of Chemical Physics*, 98(7):5648–6, 1993.
98. Chengteh Lee, Weitao Yang, and Robert G Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2):785, 1988.
99. Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *Journal of Chemical Physics*, 110(13):6158–6170, 1999.
100. Yan Zhao and Donald G Truhlar. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts*, 120(1):215–241, 2008.
101. Seymour H Vosko, Leslie Wilk, and Marwan Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of physics*, 58(8):1200–1211, 1980.
102. John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865, 1996.
103. Axel D Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A*, 38(6):3098, 1988.
104. Hannes Jónsson, Greg Mills, and Karsten W Jacobsen. Nudged elastic band method for finding minimum energy paths of transitions. 1998.
105. Neil W Ashcroft, N David Mermin, et al. Solid state physics, 1976.
106. Frans B Van Duijneveldt, Jeanne GCM van Duijneveldt-van de Rijdt, and Joop H van Lenthe. State of the art in counterpoise theory. *Chemical Reviews*, 94(7):1873–1885, 1994.
107. Péter Pulay. Convergence acceleration of iterative sequences. the case of scf iteration. *Chemical Physics Letters*, 73(2):393–398, 1980.
108. Andreas Klamt and GJGJ Schüürmann. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2*, (5):799–805, 1993.
109. Joe R McDaniel and Jason R Balmuth. Kekule: OCR-optical chemical (structure) recognition. *Journal of Chemical Information and Computer Sciences*, 32(4):373–378, 1992.
110. Richard Casey, Stephen Boyer, Paul Healey, Alex Miller, Bernadette Oudot, and Karl Zilles. Optical recognition of chemical graphics. In *Proceedings of 2nd International Conference on Document Analysis and Recognition (ICDAR'93)*, pages 627–631, <https://ieeexplore.ieee.org/document/395796>, Tsukuba, Japan, 1993. IEEE.
111. Kohulan Rajan, Henning Otto Brinkhaus, Achim Zielesny, and Christoph Steinbeck. A review of optical chemical structure recognition tools. *Journal of Cheminformatics*, 12(1):1–13, 2020.

112. Martijn Oldenhof, Adam Arany, Yves Moreau, and Jaak Simm. ChemGrapher: Optical Graph Recognition of Chemical Compounds by Deep Learning. *Journal of Chemical Information and Modeling*, 60(10):4506–4517, 2020.
113. Harry L Morgan. The generation of a unique machine description for chemical structures—a technique developed at chemical abstracts service. *Journal of Chemical Documentation*, 5(2):107–113, 1965.
114. Julian R Ullmann. An algorithm for subgraph isomorphism. *Journal of the ACM (JACM)*, 23(1):31–42, 1976.
115. Nadine Schneider, Roger A Sayle, and Gregory A Landrum. Get Your Atoms in Order: An Open-Source Implementation of a Novel and Robust Molecular Canonicalization Algorithm. *Journal of Chemical Information and Modeling*, 55(10):2111–2120, 2015.
116. Judah B Ginsberg. Overview of the National Historic Chemical Landmarks Program. In *Abstracts of Papers of the American Chemical Society*, volume 233, pages 657–657. American Chemical Society 1155 16TH ST, NW, Washington, DC 20036 USA, 2007.
117. Roy Fielding, Jim Gettys, Jeffrey Mogul, Henrik Frystyk, Larry Masinter, Paul Leach, and Tim Berners-Lee. Hypertext transfer protocol—HTTP/1.1, 1999.
118. David W Weisgerber. Chemical abstracts service chemical registry system: history, scope, and impacts. *Journal of the American Society for Information Science*, 48(4):349–360, 1997.
119. Paul Hodel. Jspreadsheet v4: The JavaScript spreadsheetMS Windows NT. <https://bossanova.uk/jsspreadsheet/v4/>, 2021. Accessed: 2021-08-31.
120. DAVID A. PENSAK and E. J. COREY. *LHASA—Logic and Heuristics Applied to Synthetic Analysis*, chapter 1, pages 1–32.
121. Peter D Karp, Suzanne Paley, and Pedro Romero. The pathway tools software. *Bioinformatics*, 18(suppl_1):S225–S232, 2002.
122. Ingrid M Socorro, Keith Taylor, and Jonathan M Goodman. ROBIA: A reaction prediction program. *Organic Letters*, 7(16):3541–3544, 2005.
123. Manyi Yang, Jingxiang Zou, Guoqiang Wang, and Shuhua Li. Automatic reaction pathway search via combined molecular dynamics and coordinate driving method. *Journal of Physical Chemistry A*, 121(6):1351–1361, 2017.
124. Sina Stocker, Gábor Csányi, Karsten Reuter, and Johannes T Margraf. Machine learning in chemical reaction space. *Nature Communications*, 11(1):1–11, 2020.
125. Connie W Gao, Joshua W Allen, William H Green, and Richard H West. Reaction mechanism generator: Automatic construction of chemical kinetic mechanisms. *Computer Physics Communications*, 203:212–225, 2016.
126. Robert S. Weber and Michael A. Reynolds. Special issue of energy & fuels honoring michael t. klein. *Energy & Fuels*, 34(12):15079–15081, 2020.
127. Dmitrij Rappoport and Alán Aspuru-Guzik. Predicting feasible organic reaction pathways using heuristically aided quantum chemistry. *Journal of chemical theory and computation*, 15(7):4099–4112, 2019.
128. Kurt Wolfe, Nicholas Pope, Rajbir Parmar, Mike Galvin, Caroline Stevens, Eric Weber, Jonathan Flaishans, and Tom Purucker. Chemical Transformation System: Cloud Based Cheminformatic Services to Support Integrated Environmental Modeling. 2016.
129. Jörg Wicker, Kathrin Fenner, Lynda Ellis, Larry Wackett, and Stefan Kramer. Predicting biodegradation products and pathways: a hybrid knowledge-and machine learning-based approach. *Bioinformatics*, 26(6):814–821, 2010.
130. Judit Zádor and Habib N Najm. KinBot 1.0: A code for automatic PES exploration. Technical report, Sandia National Lab.(SNL-CA), Livermore, CA (United States), 2013.
131. Ruben Van de Vijver and Judit Zádor. KinBot: Automated stationary point search on potential energy surfaces. *Computer Physics Communications*, 248:106947, 2020.
132. Yury V Suleimanov and William H Green. Automated discovery of elementary chemical reaction steps using freezing string and berny optimization methods. *Journal of Chemical Theory and Computation*, 11(9):4248–4259, 2015.
133. Eugene Charniak and McDermott. *Introduction to Artificial Intelligence*. Addison-Wesley, Reading, Massachusetts, 1985.

134. Graeme Henkelman, Blas P Uberuaga, and Hannes Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics*, 113(22):9901–9904, 2000.
135. E Weinan, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. *Physical Review B*, 66(5):052301, 2002.
136. Alexandra J Salter-Blanc, Eric J Bylaska, Hayley J Johnston, and Paul G Tratnyek. Predicting reduction rates of energetic nitroaromatic compounds using calculated one-electron reduction potentials. *Environmental Science & Technology*, 49(6):3778–3786, 2015.
137. Eric J Bylaska, Jeffrey G Catalano, Sebastian T Mergelsberg, Sarah A Saslow, Odetta Qafoku, Micah P Prange, and Eugene S Ilton. Association of defects and zinc in hematite. *Environmental Science & Technology*, 53(23):13687–13694, 2019.
138. Guang Hao Low, Nicholas P Bauman, Christopher E Granade, Bo Peng, Nathan Wiebe, Eric J Bylaska, Dave Wecker, Sriram Krishnamoorthy, Martin Roetteler, Karol Kowalski, et al. Q# and NWChem: Tools for scalable quantum chemistry on quantum computers. *arXiv preprint arXiv:1904.01131*, 2019.
139. Justin A North, Adrienne B Narrowe, Weili Xiong, Kathryn M Byerly, Guanqi Zhao, Sarah J Young, Srividya Murali, John A Wildenthal, William R Cannon, Kelly C Wrighton, et al. A nitrogenase-like enzyme system catalyzes methionine, ethylene, and methane biogenesis. *Science*, 369(6507):1094–1098, 2020.
140. Emma Leverich Trainer, Matthew Ginder-Vogel, and Christina K Remucal. Organic structure and solid characteristics determine reactivity of phenolic compounds with synthetic and reclaimed manganese oxides. *Environmental Science: Water Research & Technology*, 6(3):540–553, 2020.
141. Ashley S McNeill, Brooke H Dallas, John M Eiler, Eric J Bylaska, and David A Dixon. Reaction energetics and ^{13}C fractionation of alanine transamination in the aqueous and gas phases. *Journal of Physical Chemistry A*, 124(10):2077–2089, 2020.
142. Eric J Bylaska, Kevin Waters, Eric D Hermes, Judit Zádor, and Kevin M Rosso. A Filon-like integration strategy for calculating exact exchange in periodic boundary conditions: a plane-wave DFT implementation. *Materials Theory*, 4(1):1–31, 2020.
143. Eugene S Ilton, Libor Kovarik, Elias Nakouzi, Sebastian T Mergelsberg, Martin E McBriarty, and Eric J Bylaska. Using Atom Dynamics to Map the Defect Structure Around an Impurity in Nano-Hematite. *Journal of Physical Chemistry Letters*, 11(24):10396–10400, 2020.
144. Khadouja Harouaka, Caleb Allen, ML diVacri, Eric Jon Bylaska, Gregory Eiden, Eric Hoppe, and Isaac Arnequist. Gas phase ion-molecule interactions in a collision reaction cell with QQQ-ICP-MS: Investigations with N_2O as the reaction gas. In *AGU Fall Meeting Abstracts*, volume 2020, pages V033–0009, 2020.
145. Eric J Bylaska, Duo Song, and Kevin M Rosso. Electron transfer calculations between edge sharing octahedra in hematite, goethite, and annite. *Geochimica et Cosmochimica Acta*, 291:79–91, 2020.
146. Eric J Bylaska, Duo Song, Nicholas P Bauman, Karol Kowalski, Daniel Claudino, and Travis S Humble. Quantum Solvers for Plane-Wave Hamiltonians: Abridging Virtual Spaces Through the Optimization of Pairwise Correlations. *Frontiers in Chemistry*, 9:26, 2021.
147. Sebastian T Mergelsberg, Micah Prange, Duo Song, Eric J Bylaska, Sarah A Saslow, Jeffrey G Catalano, and Eugene S Ilton. Resolving Configurational Disorder for Impurities in a Low-Entropy Phase. *Journal of Physical Chemistry Letters*, 12:5689–5694, 2021.
148. Yidan Gao, Shifa Zhong, Tiffany L Torralba-Sánchez, Paul G Tratnyek, Eric J Weber, Yiling Chen, and Huichun Zhang. Quantitative structure activity relationships (QSARs) and machine learning models for abiotic reduction of organic compounds by an aqueous Fe(II) complex. *Water Research*, 192:116843, 2021.
149. John L Fulton, Eric J Bylaska, Stuart Bogatko, Mahalingam Balasubramanian, Emilie Cauët, Gregory K Schenter, and John H Weare. Near-quantitative agreement of model-free DFT-MD predictions with XAFS observations of the hydration structure of highly charged transition-metal ions. *Journal of Physical Chemistry Letters*, 3(18):2588–2593, 2012.
150. Thomas W Swaddle, Jörgen Rosenqvist, Ping Yu, Eric Bylaska, Brian L Phillips, and William H Casey. Kinetic evidence for five-coordination in $\text{AlOH}_{(aq)}^{2+}$ ion. *Science*, 308(5727):1450–1453, 2005.

151. James R Rustad and Eric J Bylaska. Ab initio calculation of isotopic fractionation in $\text{B}(\text{OH})_{3(aq)}$ and $\text{B}(\text{OH})_{4(aq)}^-$. *Journal of the American Chemical Society*, 129(8):2222–2223, 2007.
152. Raymond Atta-Fynn, Donald F Johnson, Eric J Bylaska, Eugene S Ilton, Gregory K Schenter, and Wibe A De Jong. Structure and hydrolysis of the U(IV), U(V), and U(VI) aqua ions from ab initio molecular simulations. *Inorganic Chemistry*, 51(5):3016–3024, 2012.
153. Samuel O Odoh, Eric J Bylaska, and Wibe A de Jong. Coordination and Hydrolysis of Plutonium Ions in Aqueous Solution Using Car–Parrinello Molecular Dynamics Free Energy Simulations. *Journal of Physical Chemistry A*, 117:12256–12267, 2013.
154. Raymond Atta-Fynn, Eric J Bylaska, and Wibe A de Jong. Importance of Counteranions on the Hydration Structure of the Curium Ion. *Journal of Physical Chemistry Letters*, 4(13):2166–2170, 2013.
155. James R Rustad, William H Casey, Qing-Zhu Yin, Eric J Bylaska, Andrew R Felmy, Stuart A Bogatko, Virgil E Jackson, and David A Dixon. Isotopic fractionation of $\text{Mg}_{(aq)}^{2+}$, $\text{Ca}_{(aq)}^{2+}$, and $\text{Fe}_{(aq)}^{2+}$ with carbonate minerals. *Geochimica et Cosmochimica Acta*, 74(22):6301–6323, 2010.
156. James R Rustad, Eric J Bylaska, Virgil E Jackson, and David A Dixon. Calculation of boron-isotope fractionation between $\text{B}(\text{OH})_{3(aq)}$ and $\text{B}(\text{OH})_{4(aq)}^-$. *Geochimica et Cosmochimica Acta*, 74(10):2843–2850, 2010.
157. Eric J Bylaska, Marat Valiev, James R Rustad, and John H Weare. Structure and dynamics of the hydration shells of the Al^{3+} ion. *Journal of Chemical Physics*, 126(10):104505, 2007.
158. Patrick Nichols, Eric J Bylaska, Gregory K Schenter, and Wibe de Jong. Equatorial and apical solvent shells of the UO_2^{2+} ion. *Journal of Chemical Physics*, 128(12):124507, 2008.
159. Stuart A Bogatko, Eric J Bylaska, and John H Weare. First principles simulation of the bonding, vibrational, and electronic properties of the hydration shells of the high-spin Fe^{3+} ion in aqueous solutions. *Journal of Physical Chemistry A*, 114(5):2189–2200, 2010.
160. Stuart Bogatko, Emilie Cauët, Eric Bylaska, Gregory Schenter, John Fulton, and John Weare. The aqueous Ca^{2+} system, in comparison with Zn^{2+} , Fe^{3+} , and Al^{3+} : an ab initio molecular dynamics study. *Chemistry—A European Journal*, 19(9):3047–3060, 2013.
161. Martin E McBriarty, Jennifer A Soltis, Sebastien Kerisit, Odeta Qafoku, Mark E Bowden, Eric J Bylaska, James J De Yoreo, and Eugene S Ilton. Trace uranium partitioning in a multiphase nano- FeOOH system. *Environmental Science & Technology*, 51(9):4970–4977, 2017.
162. Martin E McBriarty, Guido Falk von Rudorff, Joanne E Stubbs, Peter J Eng, Jochen Blumberger, and Kevin M Rosso. Dynamic stabilization of metal oxide–water interfaces. *Journal of the American Chemical Society*, 139(7):2581–2584, 2017.
163. Sebastien Kerisit, Eric J Bylaska, Michael S Massey, Martin E McBriarty, and Eugene S Ilton. Ab initio molecular dynamics of uranium incorporated in goethite ($\alpha\text{-FeOOH}$): Interpretation of X-ray absorption spectroscopy of trace polyvalent metals. *Inorganic Chemistry*, 55(22):11736–11746, 2016.
164. Raymond Atta-Fynn, Eric J Bylaska, Gregory K Schenter, and Wibe A De Jong. Hydration shell structure and dynamics of Curium(III) in aqueous solution: First principles and empirical studies. *Journal of Physical Chemistry A*, 115(18):4665–4677, 2011.
165. Martin E McBriarty, Sebastien Kerisit, Eric J Bylaska, Samuel Shaw, Katherine Morris, and Eugene S Ilton. Iron vacancies accommodate uranyl incorporation into hematite. *Environmental Science & Technology*, 52(11):6282–6290, 2018.
166. SS Kumaran, M-C Su, KP Lim, JV Michael, AF Wagner, Lawerence B Harding, and David A Dixon. Ab initio calculations and three different applications of unimolecular rate theory for the dissociations of CCl_4 , CFCl_3 , CF_2Cl_2 , and CF_3Cl . *The Journal of Physical Chemistry*, 100(18):7541–7549, 1996.
167. Eric J Bylaska, David A Dixon, and Andrew R Felmy. The free energies of reactions of chlorinated methanes with aqueous monovalent anions: Application of ab initio electronic structure theory. *Journal of Physical Chemistry A*, 104(3):610–617, 2000.
168. Yurii A Borisov, Edgar E Arcia, Steven L Mielke, Bruce C Garrett, and Thom H Dunning. A systematic study of the reactions of oh-with chlorinated methanes. 1. benchmark studies of the gas-phase reactions. *Journal of Physical Chemistry A*, 105(32):7724–7736, 2001.

169. Eric J Bylaska, David A Dixon, Andrew R Felmy, and Paul G Tratnyek. One-electron reduction of substituted chlorinated methanes as determined from ab initio electronic structure theory. *Journal of Physical Chemistry A*, 106(47):11581–11593, 2002.
170. Eric J Bylaska. Estimating the thermodynamics and kinetics of chlorinated hydrocarbon degradation. *Theoretical Chemistry Accounts*, 116(1):281–296, 2006.
171. Marat Valiev, Eric J Bylaska, Michel Dupuis, and Paul G Tratnyek. Combined quantum mechanical and molecular mechanics studies of the electron-transfer reactions involving carbon tetrachloride in solution. *Journal of Physical Chemistry A*, 112(12):2713–2720, 2008.
172. David Feller and David A Dixon. Extended benchmark studies of coupled cluster theory through triple excitations. *The Journal of Chemical Physics*, 115(8):3484–3496, 2001.
173. David Feller, Kirk A Peterson, Wibe A De Jong, and David A Dixon. Performance of coupled cluster theory in thermochemical calculations of small halogenated compounds. *The Journal of Chemical Physics*, 118(8):3510–3522, 2003.
174. Jie Chen, Hongyun Yin, Dunyou Wang, and Marat Valiev. Water assisted reaction mechanism of oh- with ccl4 in aqueous solution–hybrid quantum mechanical and molecular mechanics investigation. *Chemical Physics Letters*, 559:30–34, 2013.
175. William H Robertson, Gary H Weddle, Jude A Kelley, and Mark A Johnson. Solvation of the cl- h2o complex in ccl4 clusters: The effect of solvent-mediated charge redistribution on the ionic h-bond. *The Journal of Physical Chemistry A*, 106(7):1205–1209, 2002.
176. Tingting Wang, Hongyun Yin, Dunyou Wang, and Marat Valiev. Hybrid Quantum Mechanical and Molecular Mechanics Study of the S_N2 Reaction of CCl₄ + OH⁻ in Aqueous Solution: The Potential of Mean Force, Reaction Energetics, and Rate Constants. *The Journal of Physical Chemistry A*, 116(9):2371–2376, 2012.
177. Marc TM Koper and Gregory A Voth. A theory for adiabatic bond breaking electron transfer reactions at metal electrodes. *Chemical Physics Letters*, 282(1):100–106, 1998.
178. Jean Michel Saveant. A simple model for the kinetics of dissociative electron transfer in polar solvents. Application to the homogeneous and heterogeneous reduction of alkyl halides. *Journal of the American Chemical Society*, 109(22):6788–6795, 1987.
179. Donald M Camaioni, Bojana Ginovska, and Michel Dupuis. Modeling the Reaction of Fe Atoms with CCl₄. *The Journal of Physical Chemistry C*, 113(5):1830–1836, 2009.
180. PG Tratnyek, TL Johnson, SD Warner, HS Clarke, and JA Baker. In situ treatment of organics by sequential reduction and oxidation. In *Proceedings of the First International Conference on Remediation of Chlorinated and Recalcitrant Compounds*, volume 1, page 5, 1998.
181. National Research Council (US). Panel on Low Molecular Weight Halogenated Hydrocarbons, National Research Council Staff, et al. *Chloroform, Carbon Tetrachloride, and Other Halomethanes: An Environmental Assessment: A Report*. National Academy Press, Washington, D.C., 1978.
182. James F Pankow and John A Cherry. *Dense chlorinated solvents and other DNAPLs in groundwater: History, behavior, and remediation*. Waterloo Press, Waterloo, IN, 1996.
183. Michael J Truex, Christopher J Murray, Charles R Cole, Richard J Cameron, Michael D Johnson, Rodney S Skeen, and Christian D Johnson. Assessment of carbon tetrachloride groundwater transport in support of the hanford carbon tetrachloride innovative technology demonstration program. Technical report, Pacific Northwest National Lab., Richland, WA (US), 2001.
184. James E Amonette, Darla J Workman, David W Kennedy, Jonathan S Fruchter, and Yuri A Gorby. Dechlorination of carbon tetrachloride by Fe(II) associated with goethite. *Environmental Science & Technology*, 34(21):4606–4613, 2000.
185. Michelle R Kriegman-King and M Reinhard. Transformation of carbon tetrachloride by pyrite in aqueous solution. *Environmental Science & Technology*, 28(4):692–700, 1994.
186. Michelle R Kriegman-King and Martin Reinhard. Transformation of carbon tetrachloride in the presence of sulfide, biotite, and vermiculite. *Environmental Science & Technology*, 26(11):2198–2206, 1992.
187. Klaus Pecher, Stefan B Haderlein, and René P Schwarzenbach. Reduction of polyhalogenated methanes by surface-bound Fe(II) in aqueous suspensions of iron oxides. *Environmental Science & Technology*, 36(8):1734–1741, 2002.

188. Nada Assaf-Anid, Kim F Hayes, and Timothy M Vogel. Reduction dechlorination of carbon tetrachloride by cobalamin(II) in the presence of dithiothreitol: mechanistic study, effect of redox potential and pH. *Environmental Science & Technology*, 28(2):246–252, 1994.
189. Ute E Krone, Rudolf K Thauer, Harry PC Hogenkamp, and Klaus Steinbach. Reductive formation of carbon monoxide from carbon tetrachloride and FREONS 11, 12, and 13 catalyzed by corrinoids. *Biochemistry*, 30(10):2713–2719, 1991.
190. Alexandra J Salter-Blanc, Eric J Bylaska, Julia J Ritchie, and Paul G Tratnyek. Mechanisms and kinetics of alkaline hydrolysis of the energetic nitroaromatic compounds 2,4,6-trinitrotoluene (TNT) and 2,4-dinitroanisole (DNAN). *Environmental Science & Technology*, 47(13):6790–6798, 2013.
191. Frances C Hill, Liudmyla K Sviatenko, Leonid Gorb, Sergiy I Okovytyy, Gail S Blaustein, and Jerzy Leszczynski. DFT M06-2X investigation of alkaline hydrolysis of nitroaromatic compounds. *Chemosphere*, 88(5):635–643, 2012.
192. Aleksandr V Marenich, Christopher J Cramer, and Donald G Truhlar. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *Journal of Physical Chemistry B*, 113(18):6378–6396, 2009.
193. Marc Souaille and Benoit Roux. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer Physics Communications*, 135(1):40–57, 2001.
194. Raymond Atta-Fynn, Eric J Bylaska, and Wibe A de Jong. Finite temperature free energy calculations in NWChem: Metadynamics and umbrella sampling-WHAM. <https://www.nwchem-sw.org/images/Nwchem-new-pmf.pdf>, 2015. Accessed: 2021-07-31.