

Predictive stochastic analysis of massive filter-based electrochemical reaction networks

Daniel Barter,^{a,†} Evan Walter Clark Spotte-Smith,^{b,c,†} Nikita S. Redkar,^{c,d} Aniruddh Khanwale,^{b,c,e} Shyam Dwaraknath,^f Kristin A. Persson^{b,g} and Samuel M. Blau^{a*}

Received Date

Accepted Date

DOI: 00.0000/xxxxxxxxxx

Chemical reaction networks (CRNs) are powerful tools for obtaining insight into complex reactive processes. However, they are difficult to employ when reaction mechanisms and products are not thoroughly understood. Here we report new methods of CRN generation and analysis that seek to overcome these limitations. We construct CRNs by enumerating and then filtering all stoichiometrically valid reactions, avoiding the need to know reaction templates *a priori*. By applying efficient stochastic algorithms, we can interrogate CRNs to predict network products and reveal reaction pathways to species of interest. We apply this methodology to study solid-electrolyte interphase (SEI) formation in Li-ion batteries, automatically recovering products from the literature and predicting previously unknown species. We validate these results by combining CRN-predicted pathways with first-principles mechanistic analysis, discovering novel mechanisms which could realistically occur during SEI formation. This methodology enables the *de novo* exploration of vast chemical spaces, with the potential for diverse applications throughout electrochemistry.

Introduction

Electrochemistry has the power to unlock extremely useful reactions that are otherwise inaccessible. To design next-generation electrochemical technologies – ranging from batteries and fuel cells to electrosynthesis of value-added products – it is imperative to understand and eventually control reactions on the elementary level. Traditionally, studies of (electro)chemical reactivity have been conducted by hand using either trial-and-error experiments or low-throughput molecular simulations. Recent years have seen the development of a range of computational methods to automatically explore chemical reaction networks (CRNs),¹ which are defined by a set of reactions **R** between species **S**. CRNs can facilitate the rapid discovery of key species and reactions in complex

systems with minimal manual intervention. However, standard CRN construction approaches have thus far not been capable of describing electrochemical systems.

Commonly, CRNs are generated² by applying quantum chemical methods to explore a potential energy surface (PES).³ PES exploration techniques – including *ab initio* molecular dynamics,⁴ artificial force-induced reactions,⁵ and stochastic surface walking,⁶ among others – are useful for identifying intermediates, reactive products, and elementary reaction steps (including energy barriers) but typically suffer from prohibitively high cost. They are therefore limited to simple systems involving only small molecules or exploring reactivity over very short (~ 10 ps) time scales.

Recently, Zhao and Savoie devised Yet Another Reaction Program (YARP) to accelerate the generation of CRNs by PES exploration for systems of gas-phase organic molecules.⁷ By combining the semi-empirical quantum chemistry code GFN2-xTB⁸ with density functional theory (DFT), YARP is able to achieve an admirable 100-fold reduction in cost compared to using DFT alone. In spite of YARP's considerable promise for conventional organic chemistry, its applicability in electrochemistry is presently limited. Since YARP relies on GFN2-xTB to pre-optimize reaction pathways, it can only be expected to perform well in domains where GFN2-xTB produces reasonable geometries. However, we have previously shown⁹ that GFN2-xTB is not an effective tool for accelerating geometry optimization of many classes of molecules

^a Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States.

^b Department of Materials Science and Engineering, University of California, Berkeley, CA 94720, United States.

^c Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States.

^d Department of Chemical and Biomolecular Engineering, University of California, Berkeley, Berkeley, CA 94720, United States.

^e Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, CA 94720, United States.

^f Luxembourg Institute of Science and Technology, Luxembourg.

^g Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States.

† These authors contributed equally.

* smblau@lbl.gov

that are important in electrochemistry. Specifically, we found that for complex molecules including radicals, charged species, and metal-coordinated species in solution, GFN2-xTB produces geometries that are inconsistent with DFT nearly half of the time. Given the unreliability of GFN2-xTB and the intractable cost of pure DFT PES exploration, new methods must be developed to allow for the study of reactive electrochemistry using CRNs.

When PES exploration is not used, CRNs are most commonly constructed based on human chemical intuition. By applying reaction templates to include only commonly observed mechanisms^{10–14} or pruning by the “chemical distance” between species (the number of bonds that must change for a reaction to occur, or the number of reactions required to transform reactants to products) to focus only on starting species and known products of interest,¹⁵ it is possible to create networks capable of elucidating reaction pathways in thermochemical systems. However, chemical intuition is limited and unreliable when describing new reactive spaces.¹⁶ In electrochemistry, studies of reaction mechanisms^{17,18} and characterization of reaction products¹⁹ are very challenging. In addition, the linear scaling relations (*i.e.* Bell-Evans-Polanyi^{20,21}) that are widely used to predict the rates of families of similar reactions in thermochemistry have not been well established in electrochemistry. As a result, CRN methods that rely on templates or the chemical distance to known products cannot yet be used to study electrochemical reactivity.

Aiming to bypass both the cost of PES exploration and the intuition required for template-based CRN generation, we recently developed the first method to construct and analyze electrochemical CRNs, which we used to study the formation of the solid electrolyte interphase (SEI) in lithium-ion batteries. After generating graph-based CRNs containing thousands of species and millions of reactions, we used shortest-path algorithms to identify optimal pathways to two key SEI products, lithium ethylene dicarbonate (LEDC)²² and lithium ethylene monocarbonate (LEMC).²³ With this approach, we recovered known and proposed reaction mechanisms and predicted pathways that had not been previously reported.

In our prior work, the networks that we studied were limited by the computational cost of both network construction and network analysis (particularly due to the poor scaling of shortest-path graph algorithms). These costs constrained the number of species as well as the number and types of reactions contained in the networks. Even more critically, our prior graph-based analysis approach was not predictive; in order to apply shortest-path algorithms, products of interest had to be known *a priori*. Here we confront the more challenging problem of exploring a reactive space without significant *a priori* knowledge of end products. Specifically, we seek to search for many feasible pathways under various starting conditions to a range of products, byproducts, and intermediates, including products that might not be known to be important at the time of network construction.

We present a new approach to automatically construct and explore CRNs in electrochemistry that is capable of extracting unique, predictive insights and generating hypotheses to guide further in-depth analysis. We first describe our method of High-Performance Reaction Generation (HiPRGen). HiPRGen enu-

merates all stoichiometrically valid reactions and uses extensible filters to eliminate reactions based on a variety of physical or practical criteria while aiming to retain a diverse and chemically reasonable set **R**. To overcome the scaling limitations of graph-based pathfinding, we explore CRNs with a Monte Carlo-based approach, stochastically sampling the reactive space based on only reaction thermodynamics. We can then extract paths to any molecule formed in the trajectories and heuristically identify the products of the network as a function of initial conditions. The combination of HiPRGen with stochastic network analysis allows for the investigation of electrochemical reactivity without prior knowledge of reaction mechanisms or end products for the first time. We demonstrate and validate this approach with an application to SEI formation. We first identify 36 products of a HiPRGen-constructed network of roughly 5,200 molecules and 86,000,000 reactions by analyzing thermodynamically bounded Monte Carlo trajectories. The identified network products include many species reported in the SEI literature as well as a range of unreported species. To demonstrate the plausibility of these network products and their associated formation pathways, we use first-principles calculations to refine the shortest thermodynamic paths to two previously unreported products, discovering chemically plausible elementary mechanisms. Further bespoke calculations indicate that several previously-unreported network products (or related intermediates) could reasonably emerge during SEI formation and could contribute to the production of other, experimentally-identified SEI products. The methods described here serve as a starting point for predictive studies of reactivity in electrochemistry and other domains where existing knowledge is limited.

Template-Free Reaction Network Generation

Inspired by the previous work of Kim¹⁵ and Xie²³ where the chemical distance between species was used to selectively include reactions in a CRN without employing templates, we have devised HiPRGen to construct CRNs by applying filters to initial collections of species and reactions.

HiPRGen begins with some large dataset of species, the properties of which are known from *e.g.* quantum chemical calculations. We note that this work does not consider species dataset construction and assumes that an appropriate species dataset is already available (see Future Work for further details). We then apply a series of filters, where each filter can remove species that are chemically unreasonable or otherwise undesirable under the conditions studied (Figure 1-1). A list of species filters that we have designed and employed is described in Methods and is discussed in more detail in the Supplementary Information.

The filtered set of species S_{filtered} is then used to populate buckets that are each defined by a unique composition (Figure 1-2). Buckets are populated by members containing either one or two species where the total composition of each member matches the composition of the bucket. This means that any pair of members in a given bucket define the reactants and products of a stoichiometrically balanced chemical reaction containing one or two reactants and one or two products; we do not allow ternary reactions. For each bucket, all combinations of two unique members yield

HiPRGen

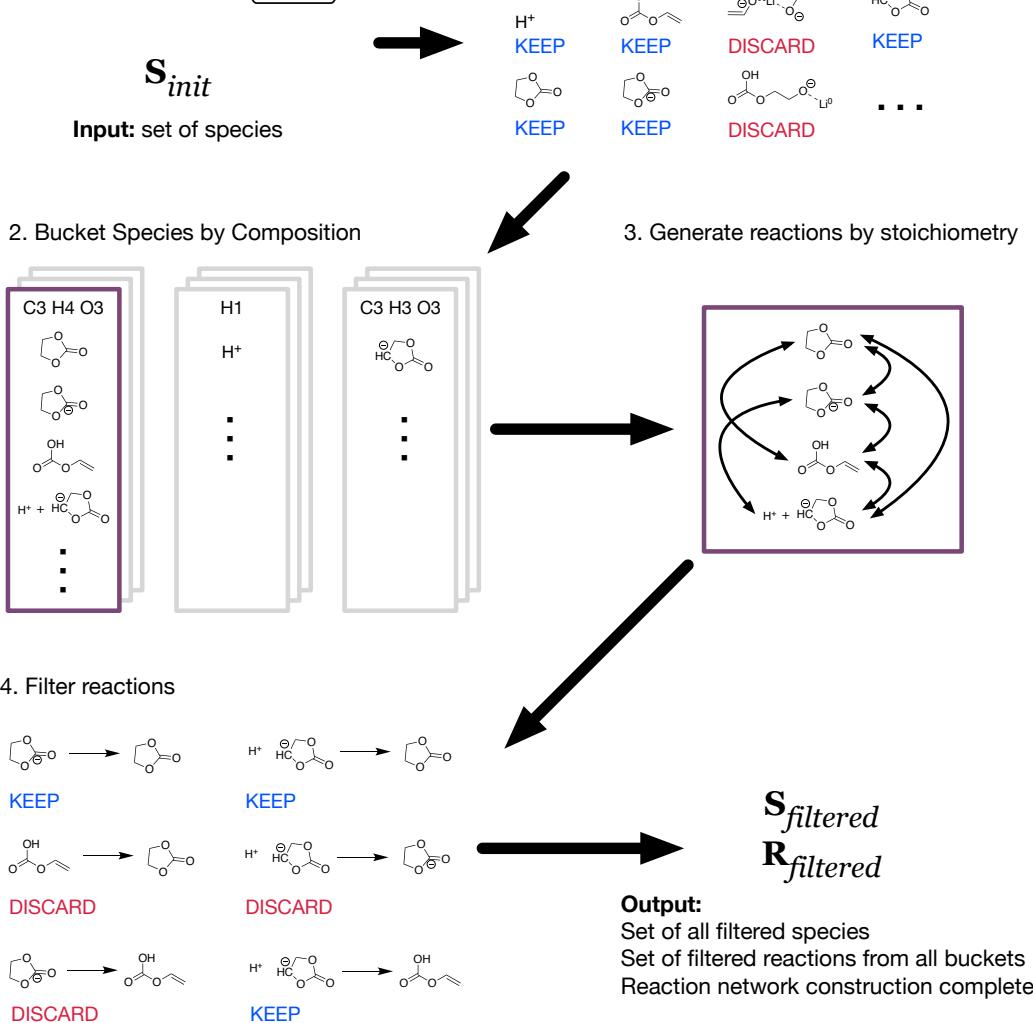


Fig. 1 A schematic overview of the High-Performance Reaction Generation (HiPRGen) method. A set of species S_{init} is provided as input. **1.** The species contained in S_{init} are filtered via user-defined criteria. Here, species including non-ionic lithium (Li^0) and species where multiple fragments are connected only by coordination bonds are removed. **2.** Species are grouped and bucketed based on composition. Each bucket is populated by entries that contain either a single molecule or a pair of molecules that together have the composition of the bucket. **3.** Within each bucket, all stoichiometrically valid reactions are generated. This corresponds to generating all combinations of two members of the bucket. **4.** The generated, stoichiometrically valid reactions are then passed through user-defined reaction filters. Here, dissociative redox reactions (where changes in bonding occur simultaneously with reduction or oxidation) and reactions involving more than two bonds changing are removed. After aggregating the reactions generated from each bucket, the end result of the HiPRGen procedure is a set of filtered species $S_{filtered}$ and a set of filtered reactions $R_{filtered}$ constituting a reaction network.

unique reactions (Figure 1-3). Note that, because we allow for electrochemical reactions, charge is not necessarily balanced in these reactions. For a system of several thousand species, there can easily be hundreds of billions or even trillions of stoichiometrically valid reactions. Reaction filters are therefore employed to remove reactions that, despite being stoichiometrically valid, are chemically implausible or otherwise undesirable (Figure 1-4). All reaction filters employed in this work are described in Methods and are discussed in more detail in the Supplementary Information. Finally, the reactions from each bucket that pass all filters are aggregated. The result of HiPRGen is a set of filtered species $S_{filtered}$ and filtered reactions $R_{filtered}$, which constitute a CRN.

HiPRGen can enumerate and filter all possible reactions between up to approximately 10,000 species, overcoming the scaling limitations of our previous approach.²³ Further, to the best of our knowledge, HiPRGen is the first method that combines an exhaustive enumeration of stoichiometrically valid reactions with a range of chemically-motivated filters that leverage pre-computed molecular properties. In addition to improving the thoroughness and efficiency of reaction generation compared to previous methods (see Supplementary Information), HiPRGen has the benefit that the filtering infrastructure was designed to be easily modified and extended by future users, making it facile to apply HiPRGen to new chemical domains.

We note that HiPRGen is an inherently inefficient method compared to *e.g.* template-based CRN generation. Many of the reactions generated by HiPRGen may not occur in a single step, may not be kinetically accessible (due to excessively high energy barriers), or may not ever occur in the chemical system of interest because they require a reactant that will never form. While we are continuing to improve HiPRGen's filters in order to better avoid non-elementary or inaccessible reactions, in the absence of a general method to robustly identify plausible species and reactions in electrochemistry, this inefficiency cannot presently be avoided.

From the CRN generated by HiPRGen, it becomes possible to search for diverse products and reaction pathways to those products. However, even after filtering the set of stoichiometrically valid reactions, the number of remaining reactions can be so vast that a highly scalable method of network analysis is required.

Thermodynamically Bounded Network Analysis Via Stochastic Methods

While it might be desirable to use shortest-path algorithms to identify reaction pathways in graph-based CRNs, as we did previously,^{22,23} such algorithms become computationally intractable as network size increases. We therefore turn to the kinetic Monte Carlo (kMC) algorithm of Gillespie,²⁴ which, with appropriate modifications,²⁵ can scale sublinearly with number of reactions. In a kMC simulation, a system evolves from some user-defined initial state in a manner that is non-deterministic but consistent with the rate coefficients provided to the model.

When templates are viable and accurately describe the reactivity in a system, they can be used to approximate reaction kinetics with minimal cost.^{10,12} However, in a template-free network of

potentially millions of reactions, it is completely impossible to include accurate rate coefficients for all reactions. For the purposes of stochastic network exploration and analysis, we therefore assign rate coefficients based on the reaction free energy ΔG rather than the reaction energy barrier ΔG^\ddagger , which is unknown at this point. Exergonic reactions ($\Delta G \leq 0$) are assumed to be barrierless and all have the same rate coefficient given by the Eyring prefactor ($\frac{k_B T}{h}$), while endergonic reactions ($\Delta G > 0$) have rate coefficients given by the Eyring equation ($\frac{k_B T}{h} \exp(-\Delta G^\ddagger/k_B T)$) with the reaction free energy serving as an effective barrier ($\Delta G^\ddagger = \Delta G$).²⁶ We note that, given sufficient sampling, all energetically favorable reactions that would occur given accurate kinetics will be observed in kMC simulations with rates based on reaction thermodynamics. In this way, we can say that the results of these kMC simulations are bounded by reaction thermodynamics. Further, the analysis of CRNs with only thermodynamic and no kinetic information is not without precedent; for instance, Stocker et al.²⁷ have previously used reaction thermodynamics and arbitrary energy barriers to explore a combustion reaction network.

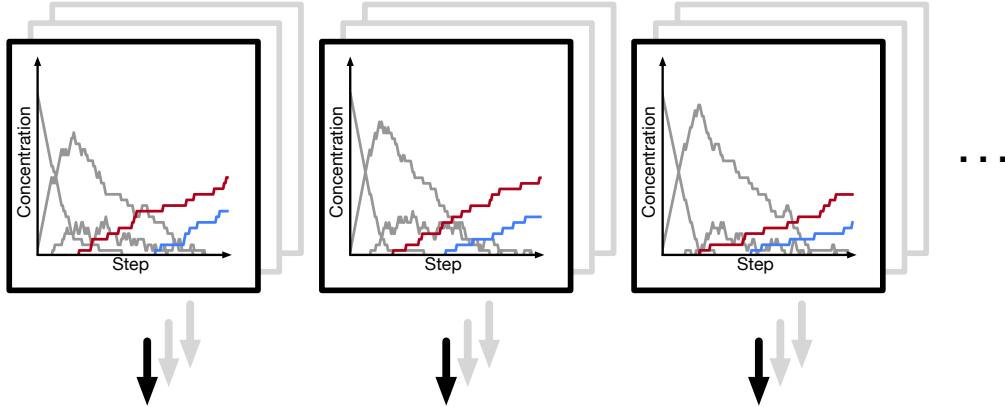
To analyze a CRN with thermodynamically bounded kMC, we perform a large number of Monte Carlo simulations in parallel (Figure 2a). The result of each simulation is a series of reactions defining a trajectory of the system state. If a molecule of interest is known, we can use these trajectories to identify potential formation pathways to that molecule. We trace each trajectory; if the molecule of interest is formed at any point, we then identify the shortest sequence of reactions leading to its first formation (Figure 2b). Performing this method of stochastic pathfinding over many trajectories, we identify a range of possible pathways to the molecule of interest. We then rank the identified paths in order to identify the “best” paths among those observed, as defined by some cost function (see Methods). The thermodynamic pathways obtained from network analysis can then be subjected to further analysis to identify complete mechanisms, including transition-states (TS) and energy barriers.

However, pathfinding is useful only if one already knows what molecule to search for. Thermodynamically bounded kMC, unlike graph-based pathfinding, enables the exploration of a reactive space without a specific target. This is because, while kMC trajectories can be used to search for a specific species, they are neither produced with any species in mind, nor are they biased towards any species. As a result, a unique capability of stochastic network analysis is the ability to identify products of a CRN with minimal prior knowledge. To do this, we apply a set of heuristic criteria to the collection of trajectories (Figure 2c). In line with the common-sense notion of a reaction product, we define a network product as any species that i) is on average formed significantly more than it is consumed; ii) accumulates significantly in an average trajectory; and iii) can be reached by low-cost reaction pathways (see Methods). We note that the specific products that are identified depend on threshold values for these heuristics, which are arbitrarily selected.

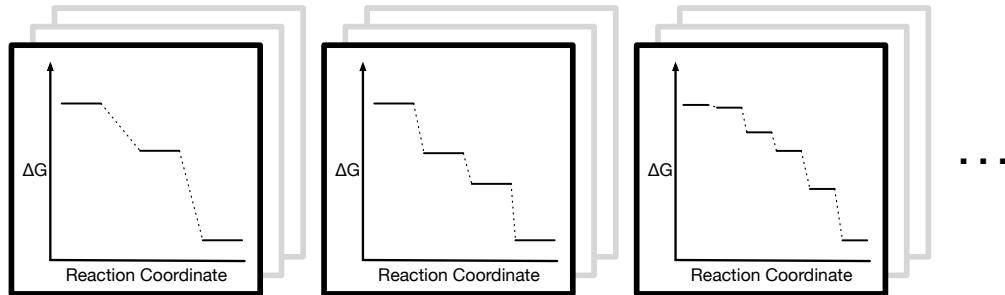
Using this heuristic method, we are able to analyze the structure of the CRN itself. The average trajectory (Figure 2c) satisfies a rate equation of the system.^{28,29} We observe (see Supplementary Information) that due to extensive sampling of the reactive

Inputs: $\mathbf{S}_{filtered}$ $\mathbf{R}_{filtered}$ $[x_i, x_j, \dots]_0$

- a) Perform many thermodynamically bounded Monte Carlo trajectories



- b) Extract shortest reaction pathways from each trajectory to each species of interest



- c) Heuristically identify network products

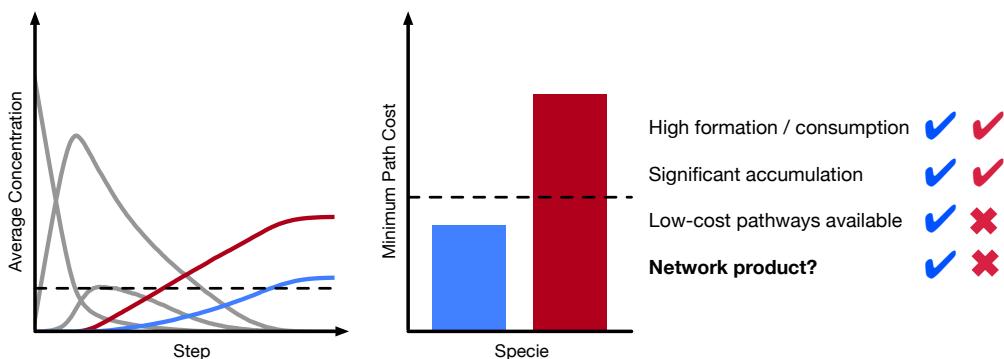


Fig. 2 Methods for analyzing CRNs from stochastic simulations. **a)** A large number of kinetic Monte Carlo trajectories with rates based only on reaction thermodynamics are calculated, beginning with the same network (defined by $\mathbf{S}_{filtered}$ and $\mathbf{R}_{filtered}$) and the same initial state ($[x_i, x_j, \dots]_0$, where x_q is the quantity of species q). **b)** In each trajectory, the shortest reaction pathway to some species of interest can be identified. Note that because these trajectories are stochastic, different trajectories will often yield different shortest pathways to the same product. **c)** To identify products of the network, a set of heuristics are applied. In order to be considered a product of the CRN, a species must be formed substantially more than it is consumed and must accumulate to a significant degree on average (that is, its average final concentration must be higher than some threshold). In addition, a product species must be reachable by some low-cost path. In the example provided, both the red and the blue species are formed significantly more than they are consumed, and both accumulate, but only the blue species can be reached by a low-cost pathway. Therefore, by this heuristic, the blue species is a network product, while the red species is not.

space, the average stochastic trajectory is smooth, indicating convergence to the exact expected system dynamics with rates based only on thermodynamics. This means that identifying network products based on the average stochastic trajectory is mathematically rigorous and well-defined. The products of the network are not necessarily the metastable or stable products that would be observed experimentally, nor are they necessarily exhaustive. Nonetheless, the network products provide useful hypotheses regarding what might form in an actual reactive system. We can then interrogate these hypotheses and validate them by either theoretical or experimental means. We note that in addition to the choice of heuristic thresholds, the choice of initial state can affect the network products identified via this method.

Results and Discussion

Automatic Identification of Battery SEI Network Products

Using HiPRGen, we constructed a reaction network that seeks to describe SEI formation in lithium-ion batteries. We begin with an initial set of 8,904 species taken from the Lithium-Ion Battery Electrolyte (LIBE) dataset,³⁰ which was strategically designed to facilitate studies of electrolyte reactivity and interphase formation (see Methods). Network construction resulted in a CRN containing 5,193 filtered species and 86,001,275 filtered reactions. With this network, we conducted 100,000 thermodynamically bounded kMC trajectories under four conditions, with combinations of two different applied potentials (+0.0V vs. Li/Li⁺ and +0.5V vs. Li/Li⁺) and two different initial states (one consisting only of Li⁺ and ethylene carbonate (EC) and the other consisting of Li⁺, EC, and CO₂). Average trajectories for each condition are shown in the Supplementary Information. We emphasize that our goal is not to compute and observe the dynamics of SEI formation, but rather to identify key species and reaction pathways. We further note that we do not consider the effect of the electrode surface in our simulations. However, since the SEI can grow to a thickness of ~10nm, the effect of the electrode on the SEI chemistry should be small after the first reactions occur. Moreover, the products of the SEI are in general insensitive to the identity of the anode.³¹

The utility of our approach is demonstrated through analysis of the 36 network products collected from the set of four conditions previously described (Figure 3). We first note that our automated procedure recovers 16 species that include a majority of the experimentally observed products of SEI formation (Fig. 3 solid dark green). These include gases (H₂, C₂H₄, CO),³² inorganic species (lithium carbonate (Li₂CO₃) and lithium oxalate (Li₂C₂O₄)),^{33,34} and alkyl carbonates (including species closely related to LEDC^{33–35} and LEMC,^{19,34} as well as lithium methyl carbonate or LMC, lithium butylene dicarbonate or LBDC,³⁴ and lithium vinyl carbonate).³⁶ We emphasize that these species are recovered even though network exploration is based only on reaction thermodynamics, with reaction kinetics entirely ignored.

In addition to these well-known species, there are also a number of novel products that have not previously been proposed to participate in SEI formation. Among these are six additional alkyl carbonates (Fig. 3 dotted light green) which are each very sim-

ilar to known products in molecular size, composition, bonding, and contained functional groups. Due to the extreme difficulty of experimentally characterizing the SEI and the resulting limited ability to resolve small signal to noise,³⁷ the likely spectroscopic similarity³⁸ of these species to the known products means that they may be present in the SEI in small quantities but that they could not easily be positively identified.

Other network products include species with ester, carboxylate, and oxide functional groups, such as lithium 2-(formyloxy)ethan-1-olate, which we abbreviate as LFEO, as well as a number of cyclic species, such as 4,4',5,5'-tetrahydro-2,2'-bi(1,3-dioxolylidene), which we abbreviate as bi-dioxolylidene. LFEO and bi-dioxolylidene (Fig. 3 dashed purple) were particularly unexpected given how distinct they are from other predicted SEI products and, in particular, the experimentally identified products. Evaluating whether or not these products will actually form in the SEI necessitates considering energy barriers, kinetics, and reactive competition. Using the thermodynamically shortest paths to guide automated transition-state calculations, we identified elementary formation mechanisms to both LFEO and bi-dioxolylidene to evaluate their potential to participate in SEI formation (see CRN-Derived Elementary Mechanisms to Form Unexpected Network Products).

On the other hand, there are some network products which do not reflect the corresponding chemical system in a real battery. Specifically, both vinylene carbonate (VC) and propylene carbonate (PC) are known to rapidly decompose when included in battery electrolytes.^{39,40} This contradiction indicates that there are missing reactions or missing species necessary to facilitate the decomposition of VC and PC. The identification of this gap through the use of stochastic network analysis and network product prediction provides a tractable path forward to expand the CRN via selective addition of missing molecules that enable redox, decomposition, or recombination of network products with other abundant intermediate or product species.

CRN-Derived Elementary Mechanisms to Form Unexpected Network Products

The reaction pathways produced by our stochastic approach are based solely on reaction thermodynamics, with no knowledge of reaction kinetics. In actuality, the dominant reaction pathways are heavily dependent on reaction energy barriers ΔG^\ddagger and rate coefficients. In order for thermodynamically bounded kMC to provide useful chemical insights, it is critical that the thermodynamic reaction pathways can be reasonably translated into elementary reaction mechanisms including TS and energy barriers.

Using our stochastic approach, we can identify the N shortest thermodynamic reaction pathways to the network products, ranked by a cost function that we have employed previously²² (see Methods). We selected two unexpected network products – LFEO and bi-dioxolylidene – and subjected their shortest pathways in order of cost to an automated procedure to identify the TS for each step along each pathway, allowing for the construction of complete reaction mechanisms. Figure 4 highlights two formation paths obtained using this procedure, emphasizing the

Collected network products

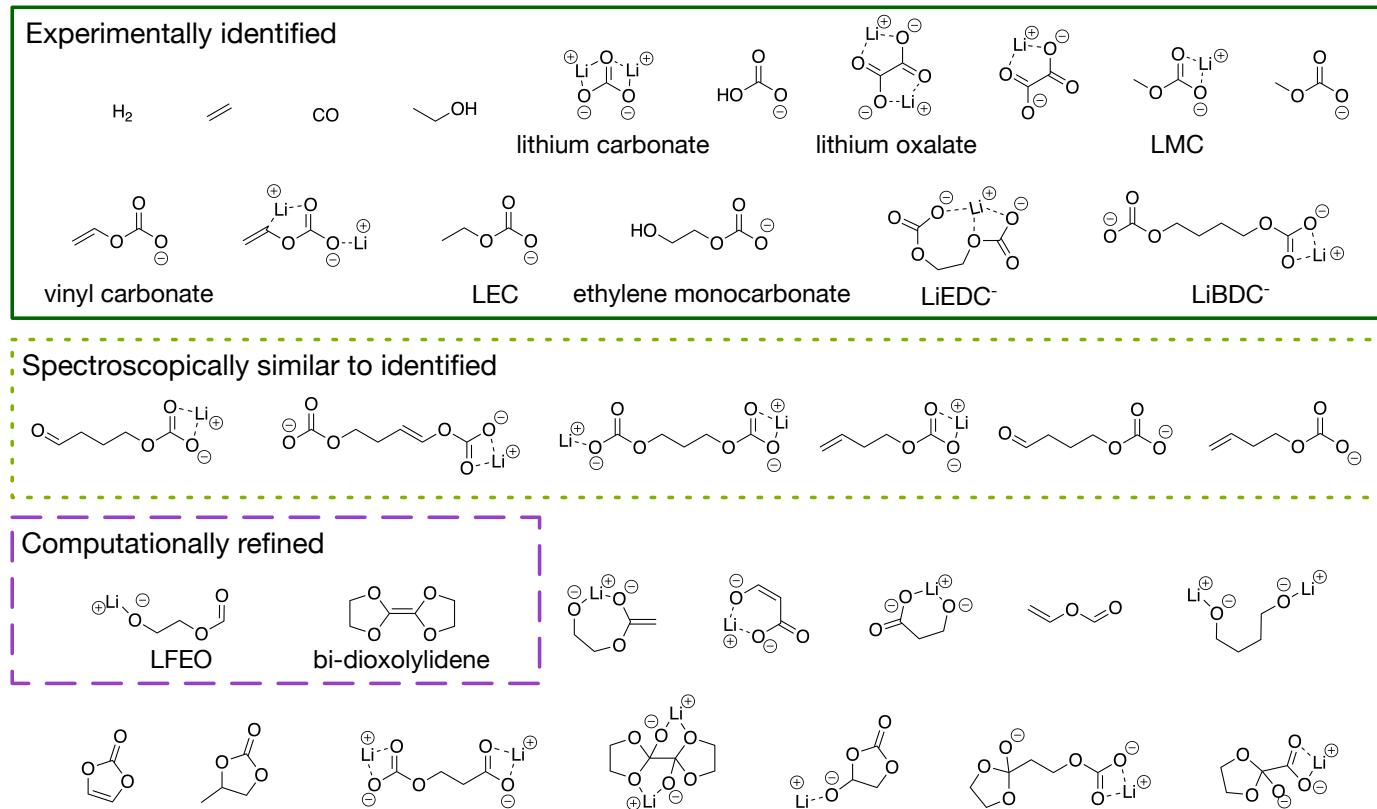


Fig. 3 The 36 total collected network products from four different initial conditions (+0.0V vs. Li/Li⁺ with Li⁺ and EC as starting species; +0.0V vs. Li/Li⁺ with Li⁺, EC, and CO₂ as starting species; +0.5V vs. Li/Li⁺ with Li⁺ and EC as starting species; and +0.5V vs. Li/Li⁺ with Li⁺, EC, and CO₂ as starting species). The 16 network products outlined in green have previously been experimentally identified in the SEI; these include the major gaseous products, molecular inorganic components, and organic components (including lithium methyl carbonate or LMC, vinyl carbonate, lithium ethyl carbonate or LEC, ethylene monocarbonate, lithium ethylene dicarbonate or LiEDC⁻, and lithium butylene dicarbonate or LiBDC⁻). Six of the network products, outlined in dotted light green, are species which have very similar spectroscopic signatures to the dominant organic components, and thus may be present in the SEI in small quantities without being detected. Two of the network products outlined in dashed purple, lithium 2-(formyloxy)ethan-1-olate or LFEO and 4,4',5,5'-tetrahydro-2,2'-bi(1,3-dioxolylidene) or bi-dioxolylidene, have not been previously reported and were subjected to further mechanistic analysis. Finally, the remaining 12 network products (which have also not been previously reported as SEI products) may be kinetically inaccessible, may indicate that our CRN is missing species or reactions, or may be true SEI products, motivating future calculations.

utility of network-generated thermodynamic reaction pathways to construct elementary mechanisms.

The network pathway shown in Figure 4a has the 12th lowest cost with only Li⁺ and EC as starting species (no CO₂) at +0.0V vs. Li/Li⁺. In this pathway, Li⁺ coordinates with EC, and the Li⁺EC reduces twice. The doubly reduced Li⁺EC²⁻ then ring-opens at the shoulder bond, after which this shoulder-ring-opened species can abstract a proton from an additional Li⁺EC, forming LFEO with a Li⁺EC-H⁻ as a byproduct. The identified elementary mechanism follows this path exactly, with two TS – one for the ring-opening of Li⁺EC²⁻ with a barrier $\Delta G^\ddagger = 0.10$ eV and one for proton abstraction from EC to form LFEO with $\Delta G^\ddagger = 0.11$ eV.

A path to form bi-dioxolylidene is shown in Figure 4b. This network pathway has the 3rd lowest cost for simulations with CO₂ was present as a starting species at 0.0V vs. Li/Li⁺. In the pathway, CO₂ reduces twice and coordinates with Li⁺, forming Li⁺CO₂²⁻. This Li⁺CO₂²⁻ species reacts with EC to form

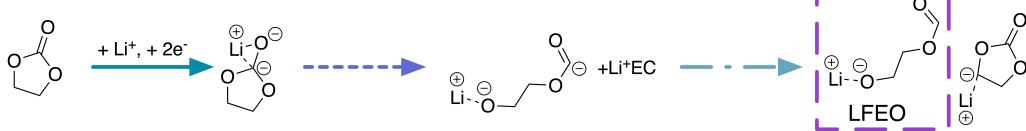
Li⁺CO₃²⁻ and the 1,3-dioxolylidene carbene, which we abbreviate as dioxolylidene. Two of these carbenes can then combine to form the dimer bi-dioxolylidene. The identified elementary mechanism follows the same general steps as the network pathway – coordinate and reduce, form dioxolylidene, and then dimerize two carbene – but differs in two main ways. First, we found that the carbene formation actually occurs via an addition-elimination mechanism with two elementary steps. The addition, which results in an EC-CO₂ adduct, has a barrier $\Delta G^\ddagger = 0.20$ eV, and the elimination to produce Li⁺CO₃²⁻ and dioxolylidene has a barrier $\Delta G^\ddagger = 0.01$ eV. Second, it is more favorable to reduce EC than CO₂, which changes the initial steps of the mechanism.

The identified elementary mechanisms to LFEO and bi-dioxolylidene involve steps that are predicted to be competitive with other known SEI formation processes. Both mechanisms rely on Li⁺EC²⁻, which can form easily at low potentials.^{31,41} After breaking the shoulder bond, the ring-opened Li⁺EC²⁻ is known to decompose unimolecularly to Li⁺OCH₂CH₂O²⁻ and CO with

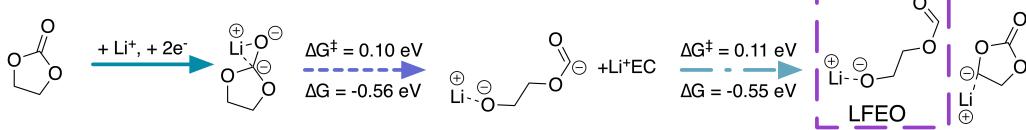
Mechanistic refinement of network paths to novel products

a) lithium 2-(formyloxy)ethan-1-olate (LFEO)

Network pathway:

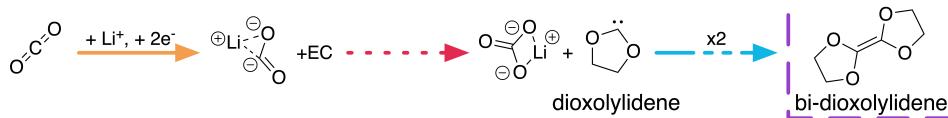


Elementary mechanism:



b) 4,4',5,5'-tetrahydro-2,2'-bi(1,3-dioxolylidene) (bi-dioxolylidene)

Network pathway:



Elementary mechanism:

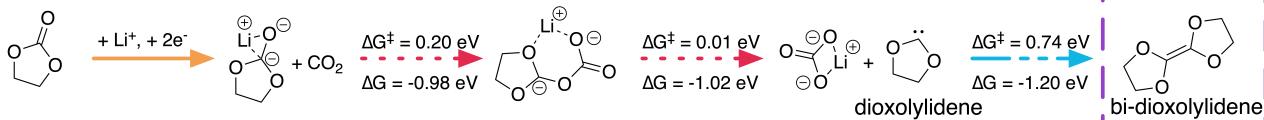


Fig. 4 Comparison of network-identified reaction pathways and elementary mechanisms obtained from kinetic refinement to form (a) lithium 2-(formyloxy)ethan-1-olate (LFEO) and (b) 4,4',5,5'-tetrahydro-2,2'-bi(1,3-dioxolylidene) (bi-dioxolylidene). For elementary steps involving a transition-state, energy barriers (ΔG^\ddagger) and reaction free energies (ΔG) are provided. Corresponding reaction steps between the network-identified pathways and elementary mechanisms are indicated via line color and style. Coordination and reduction steps are combined for brevity; in reality, these occur as separate steps in both network-identified pathways and elementary mechanisms.

a predicted energy barrier between 0.09 eV^{31} and 0.22 eV^{41} depending on the level of theory used. Considering that the necessary precursor to LFEO formation, Li^+EC , should be present in abundance during early SEI formation, this implies that LFEO could actually be a significant product during early SEI formation at low potentials.

The formation of bi-dioxolylidene is predicted to be less kinetically favorable than that of LFEO. The dimerization reaction has an energy barrier that is considerably higher than many major SEI formation pathways,^{31,41–44} implying that bi-dioxolylidene should not be a significant product. The formation of the carbene monomer, on the other hand, is plausible. Using the Eyring equation,²⁶ the addition of CO_2 to $\text{Li}^+\text{EC}^{2-}$ with a 0.20 eV barrier has a predicted rate coefficient roughly 70 times lower than that of the shoulder ring-opening of $\text{Li}^+\text{EC}^{2-}$ with a 0.09 eV barrier.

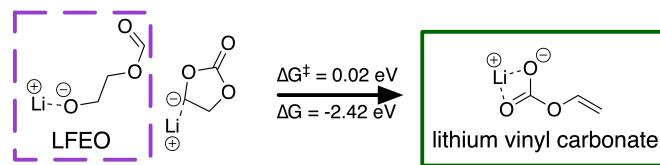
However, dioxolylidene formation could be significant if CO_2 is abundant, a plausible scenario considering that CO_2 is highly soluble in carbonate solvents⁴⁵ and can form at either the anode⁴⁶ or the cathode⁴⁷ in Li-ion batteries. On this basis, we predict that while LFEO, $\text{Li}^+\text{OCH}_2\text{CH}_2\text{O}^{2-}$, and CO will be favored, dioxolylidene should at least form as a short-lived minority intermediate (considering the general reactivity of carbenes).⁴⁸

CRN Pathways and Products Guide Investigation of Expanded Mechanisms

Leveraging CRN analysis, we have predicted network products and reaction paths that could be important to a complex electrochemical system but which have not been seriously studied in the literature before. We can now expand on these paths, using the calculated elementary formation mechanisms of LFEO and bi-

Mechanistic extension of network product pathways

a) Byproducts of LFEO formation



b) Carbene catalysis

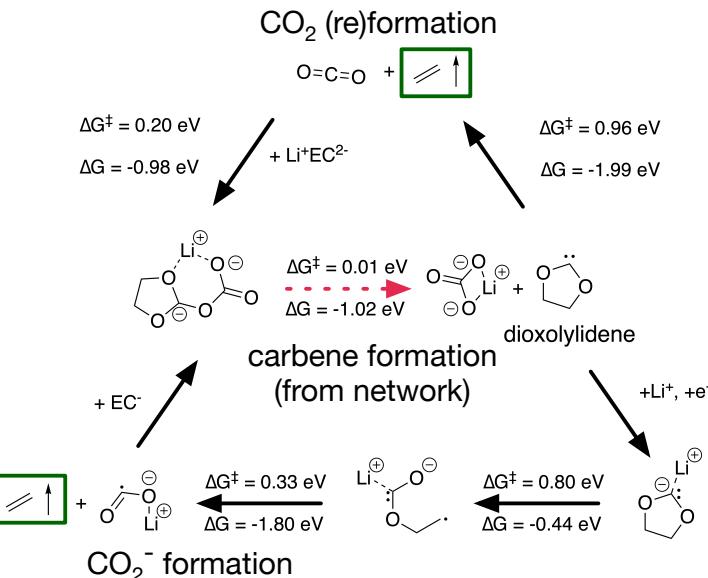


Fig. 5 Extended mechanisms based off of network-identified pathways: a) formation of lithium vinyl carbonate, a known SEI product, via the ring-opening of the deprotonated $\text{Li}^+\text{EC}-\text{H}^-$, a byproduct of LFEO formation; b) two possible catalytic cycles yielding ethylene gas and reforming dioxolylidene via the production of either CO_2 or Li^+CO_2^- . For elementary steps involving a transition-state, energy barriers (ΔG^\ddagger) and reaction free energies (ΔG) are provided. Green boxes indicate species that have been experimentally identified as products or byproducts of SEI formation.

dioxolylidene as starting points for studying how species along these paths may further react. In doing so, we demonstrate the utility of CRN-generated pathways as a tool for hypothesis generation to guide follow-up investigation (see Figure 5).

In the mechanism for LFEO formation identified in Figure 4a, a byproduct is the deprotonated EC species $\text{Li}^+\text{EC}-\text{H}^-$. We suspected that this byproduct would be highly reactive and would likely decompose. Indeed, we found (Figure 5a) that $\text{Li}^+\text{EC}-\text{H}^-$ can open at the waist bond with an extremely low barrier ($\Delta G^\ddagger = 0.02 \text{ eV}$), forming lithium vinyl carbonate. We note that lithium vinyl carbonate has previously been identified as an SEI product,³⁶ though its formation mechanism has not been thoroughly studied. Therefore, not only is the formation of LFEO plausible on the basis of the low reaction barriers identified, but LFEO formation can potentially help to explain the formation of another SEI product.

As another example, we considered the reactivity of the dioxolylidene carbene (Figure 5b). In addition to the dimerization shown in Figure 4b, we found that dioxolylidene could react in two other ways, either decomposing in a single step to form CO_2 and C_2H_4 or decomposing to Li^+CO_2^- and C_2H_4 via a two-step

process after coordination with Li^+ and reduction. All reactions identified – dimerization and both decomposition mechanisms – involve relatively high energy barriers. Therefore, our understanding of the role of dioxolylidene in SEI formation remains incomplete, and further work must be done to elucidate its decomposition routes. However, we note that if the barriers to dioxolylidene decomposition were lowered by e.g. a solvent effect⁴⁹ or a reactive surface,⁵⁰ the possibility exists for a catalytic loop in which dioxolylidene is repeatedly reformed via the reaction of CO_2 with $\text{Li}^+\text{EC}^{2-}$ or Li^+CO_2^- with EC^- .

Future Work

The combination of CRN generation via HiPRGen and stochastic CRN analysis can lead to predictive and unique insights, as we have shown through our case study of SEI formation. Although the SEI has been intensely studied for roughly 30 years, our approach is able to predict new product and intermediate species that are likely to form in a real battery based on reaction kinetics and could perhaps even compete with known reaction mechanisms. However, there remain both methodological and chemical challenges that limit the widespread applicability of our method-

ology.

When PES exploration can be used to generate CRNs, new intermediate and product species can be identified automatically via reactions starting from known species. Likewise, when templates can be used, new species are generated by repeatedly applying reaction motifs to provided or previously generated reactants. However, as we have explained, neither PES exploration nor templates can presently be applied to electrochemistry. Instead, HiPRGen currently requires an input set of species containing all molecules to be included in the CRN and their calculated properties. For applications to Li-ion SEI formation, where the LIBE dataset of molecular thermochemistry already exists,³⁰ this is not a significant limitation. However, in general, the reliance on a pre-computed dataset of species is problematic, as it potentially creates a high computational barrier to entry for the user and limits predictive discovery of novel species and pathways they participate in. We are actively working to overcome this requirement, developing a framework to automatically generate CRNs from only a set of known starting molecules by leveraging calculated network products to strategically guide iterative network expansion and machine learning (ML) to reduce the number of required high-throughput DFT calculations.

An additional improvement could be made by incorporating reaction kinetics into CRN generation and stochastic analysis, as is common in the study of CRNs. We have shown that reaction thermodynamics alone can be sufficient to predict reasonable reaction pathways as well as network products, yet a network based on only thermodynamics certainly contains many reactions that are kinetically limited and therefore not important in practice. Although there are not currently any methods capable of predicting energy barriers or rate coefficients for solution-phase electrochemical reactions without relying on expensive electronic structure methods (which cannot be applied to millions of reactions), we believe that ML could eventually provide sufficiently accurate estimates to realistically constrain network construction at minimal cost given sufficient training data.⁵¹

Conclusions

In this article, we have described our approach to explore electrochemical reactivity using CRNs with minimal reliance on prior mechanistic knowledge. The HiPRGen method allows for the construction of CRNs in domains where reaction templates cannot be applied and where reaction mechanisms are poorly understood. The resulting networks can then be analyzed using stochastic simulations based only on reaction thermodynamics, allowing for the identification of network products and reaction pathways which can be further refined to discover elementary mechanisms. We applied this methodology to study SEI formation in Li-ion batteries, generating a network consisting of over 5,000 species and over 86,000,000 reactions. Network product prediction yielded many known SEI components as well as several species that had not previously been proposed. We identified elementary mechanisms for the formation of multiple novel network products (LFEO and bi-dioxolylidene) based on pathways obtained from CRN analysis and refined using automated DFT. CRN-derived pathways further guided the expansion of automati-

cally obtained mechanisms, yielding additional insight. The novel mechanisms that we discovered validate our methods of product prediction and pathway identification and moreover indicate that LFEO could be a component of the early SEI. Our findings indicate that thermodynamic reaction pathways obtained via analysis of appropriately filtered CRNs can be used to efficiently search for transition-states of chemically meaningful reaction mechanisms, facilitating the construction of microkinetic models (an example of which we report in a separate publication).³¹ With the approach presented here, it will be possible to provide insights into a range of domains where fundamental understanding of reactivity is limited.

Computational Methods

Species Selection

Initial species were taken from the Lithium Ion Battery Electrolyte (LIBE) dataset.³⁰ The LIBE dataset contains the properties of 17,190 species of various charges (-1, 0, 1) and spin multiplicities (1, 2, 3) calculated using density functional theory (DFT) at the ω B97X-V/def2-TZVPPD/SMD level of theory.⁵²⁻⁵⁴ LIBE was designed for study of reactivity in battery electrolytes and was constructed so as to include diverse and novel reactive species from automated molecular fragmentation and recombination, making it well suited for the construction of CRNs and especially well suited for interrogating SEI formation.

For this work, network construction began with a subset of LIBE containing all species comprised of only carbon, hydrogen, oxygen, and/or lithium. This subset, which we call LIBE-CHOLi, contains 8,904 species.

Modifying Species Thermodynamics

All calculations on the species present in LIBE were conducted in an implicit solvation environment. While implicit solvation is generally accurate enough for the calculation of properties such as the solvation energy⁵⁵ and redox potentials of organic molecules,⁵⁶ we have found that even highly accurate implicit solvation methods severely underestimate the stabilization of small ions, especially metal ions, by solvent. This means that species in LIBE containing Li^+ ions with many coordination bonds are in many cases vastly more stable according to DFT than those with fewer coordination bonds, even if the corresponding species without lithium present are significantly less stable (an example is provided in the Supplementary Information). This insufficient stabilization led to inaccurate thermodynamics for reactions where the overall charge of the system was constant but the number of coordinate bonds changed (non-redox reactions).

To correct for insufficient metal ion stabilization, we optimized Li^+EC_n clusters at the ω B97X-D/def2-SVPD/PCM// ω B97X-V/def2-TZVPPD/SMD level of theory, with $n \in 0, 1, 2, 3, 4$ to estimate the stabilizing effect of each solvent molecule on Li^+ . The lower level of theory (ω B97X-D/def2-SVPD/PCM, $\epsilon = 18.5$) was used for optimization due to the considerable computational cost of optimizing large clusters. We found (see Supplementary Information) that each EC stabilized Li^+ by ~ 0.7 eV.

During reaction network construction, we consider two free

energies for each species: one uncorrected, and one solvent-corrected. The uncorrected free energy is taken directly from LIBE. For the solvent-corrected free energy, we count the total number of coordinate bonds to all Li^+ ions (see the Supplementary Information for a description of the method for deducing metal coordination) and compare this to the maximum expected number of coordinate bonds (assuming that each Li^+ would prefer to be coordinated by four neighbors).⁵⁷ If any Li^+ are under-coordinated, then the free energy is lowered by 0.68 eV for each “missing” coordinate bond. When calculating redox free energies, the uncorrected free energy is used; otherwise, the corrected free energy is used (see the Supplementary Information).

Species Filtering

In the HiPRGen package (see Code Availability), we implement a number of filters that remove undesirable species. These filters take as input an object containing information about a molecule, including its species, coordinates, charge, spin multiplicity, partial charges, connectivity, and thermodynamics. Each filter, based on this information, can discard the molecule or pass it onto the next filter. For terminal filters, if the molecule passes, then it is included in the final filtered set $\mathbf{S}_{\text{filtered}}$. For this work, the following molecules were filtered out:

- Molecules composed of two or more disconnected fragments
- Metal-centric complexes, where two or more non-metal fragments are connected only by coordinate bonds to Li^+
- Molecules containing neutral or negative metal ions, where the charges are calculated by applying the Natural Bonding Orbital (NBO) program version 5.0^{58,59} to a single-point energy calculation at the $\omega\text{B97X-V}/\text{def2-TZVPPD}/\text{SMD}$ level of theory in Q-Chem.
- Molecules with charge less than -1 or greater than 1

In addition to these filters, which define types of molecules to be excluded from the final network, we further reduce the molecules in the network by removing redundant species. In LIBE, all molecules are unique based on the combination of their charge, spin multiplicity, and molecular connectivity. This means that there could be several molecules that differ only by spin multiplicity, or that differ only by the coordination environment of Li^+ ions (what we call “coordimers”). When this occurs (when there are multiple molecules with the same covalent connectivity and charge but potentially with different coordination environments or spin multiplicities), we include only that species with the lowest solvation-corrected free energy in the final filtered set of species $\mathbf{S}_{\text{filtered}}$.

These filters are explained further in the Supplementary Information. We emphasize that these filters are particular to the chemistry being studied in this work, but that HiPRGen has been engineered to enable straightforward addition, removal, or modification of filters in order to be easily applied across diverse chemical applications.

Reaction Filtering

Reaction filters take as input a reaction, defined by a collection of reactants and a collection of products, and either discard the reaction or pass it onto the next filter until a terminal filter is reached. The following types of reactions were filtered out in this work:

- Endergonic reactions with $\Delta G > 0\text{eV}$. The reaction free energies for non-redox reactions were taken as $\Delta G = \sum_{i=1}^m G_{\text{product},i}^C - \sum_{j=1}^n G_{\text{reactant},j}^C$, where G^C is the solvation-corrected free energy, m is the length of the product collection, and n is the length of the reactant collection. For redox reactions, $\Delta G = G_{\text{product}}^0 - G_{\text{reactant}}^0 + \Delta q(G_e)$, where G^0 is the uncorrected free energy, Δq is the change in charge of the reaction, and G_e is the free energy of the electron (for +0.0 V vs. Li/Li^+ , $G_e = -1.4\text{eV}$; for +0.5 V vs. Li/Li^+ , $G_e = -1.9\text{eV}$)
- Reactions involving a charge change $|\Delta q| > 1$
- Redox reactions involving more than one reactant or more than one product
- Unimolecular dissociative redox reactions in which $|\Delta q| > 0$ and covalent bonds form or break
- Reactions involving more than two reactants or products (this is actually enforced by the species bucketing procedure of HiPRGen and is not a separate filter)
- Reactions involving spectators (species that do not change as a result of the reaction) e.g. A + B to A + C.
- Rections involving more than two bond changes
- Reactions in which two bonds form simultaneously or two bonds break simultaneously
- Reactions in which covalent bonds change and metal ions coordinate/decoordinate (note that reactions in which metal ions remain coordinated but change their coordinate bonds are allowed)

Motivations for each of these filters, along with examples, are provided in the Supplementary Information. Like species filters, the reaction filters can be easily modified and extended by end users to suit a broad range of chemical applications. Of course, removing filters will yield a larger final collection of reactions. We note that for the size of the species collection presented in this work, some filters are necessary to obtain a tractable number of reactions in the final collection. We further note that for thousands of species, it is necessary to filter reactions in parallel and for each filter to be computationally efficient in order to allow filtering to complete in a reasonable amount of time (hours to days).

Monte Carlo Methods

We developed a high-performance implementation of Gillespie’s direct method²⁴, with dependency graph and logarithmically scaling sampler optimizations²⁵, which we call Reaction

Network Monte Carlo (RNMC). RNMC is heavily based on the Stochastic Parallel Particle Kinetic Simulator (SPPARKS) package^{60,61} but with modifications to allow simulating networks with hundreds of millions of reactions and thousands of species. RNMC shares the reaction network and dependency graph between all running simulators and uses a lockless data structure for the dependency graph that allows it to be computed dynamically by all of the simulators in parallel.

Using RNMC, we performed 100,000 simulations under each of the four chosen conditions (+0.0V without CO₂, +0.0V with CO₂, +0.5V without CO₂, and +0.5V with CO₂). For simulations without CO₂, the initial state consisted of 30 Li⁺ and 30 EC; for those with CO₂, the initial state also included 30 CO₂. Because all reactions were exergonic and no energy barriers were considered, all rate coefficients were constant and equal. Each simulation was conducted to “completion” – that is, until there were no further reactions available for further simulation. Due to the relatively small number of initial species, most simulations took between roughly 200 and 500 steps. We note that simulating to completion – especially with so few simulation steps – is only possible because the system contains only exergonic reactions and therefore contains no loops.

Identification of Thermodynamic Reaction Pathways

To identify a single reaction pathway to a species of interest, we trace through an individual Monte Carlo trajectory. If the species of interest is formed in that trajectory, then we trace back the series of reactions leading to the first formation of that species (see Supplementary Information for an illustration of this process). For instance, if we are searching for pathways to species *X*, we might find that it is first formed by the reaction $V + W \rightarrow X$. We then look for the first reaction(s) forming *V* and *W*, and then for the first reaction(s) forming the reactants of those formation reactions, until all reactions can occur from only starting species of the simulation. The series of reactions obtained in this way define a reaction pathway to *X*.

In general, we are not interested in a single reaction pathway but rather the myriad pathways to the species of interest. Therefore, for each species of interest, we repeat the pathway identification procedure above for each trajectory, collecting all unique pathways. We then rank these pathways by some cost function. Here, the cost Φ of a given reaction is defined as $\Phi = \exp(\Delta G/k_B T) + 1$, where ΔG is the reaction free energy (uncorrected for a redox reaction, and solvation-corrected otherwise).²² The total cost of a reaction pathway is the sum of the costs of the individual reactions. We note that, because all reactions included in our network are exergonic, the constant term tends to dominate, though this cost function retains a preference for highly exergonic reactions over those that are only slightly exergonic.

Identification of Network Products

After all simulations have completed, the resulting trajectories are analyzed to determine product species. Products are defined by three criteria: the ratio of formation and consumption, relative

accumulation, and availability of low-cost pathways.

To determine the ratio of formation and consumption, each trajectory was interrogated to find all reactions involving each specie. If a given species is a reactant of an identified reaction, then that means it was consumed; if it is a product of the reaction, then that means it was formed. If the ratio of the total number of instances of formation across all trajectories to the total number of instances of consumption across all trajectories is greater than some threshold (here chosen as 1.5, meaning that three of the species are produced for every two consumed), then it could be a network product.

For relative accumulation, we take the average of all trajectories. The expected value of a species is the average of the final state – how many of the molecule will persist once the average simulation has completed. If this expected value is greater than some threshold (here 0.1, meaning that one of this species is produced and is present in the final state for every ten simulations), then that species could be a product.

Finally, for those species with formation/consumption ratios and expected values that pass the chosen thresholds, we perform thermodynamic pathfinding analysis. If the pathway with the lowest cost has a cost less than some threshold (here 10.0), then we consider the species to be a product of the network.

The species reported in Figure 3 are network products in at least one – but not necessarily all – of the four conditions considered (see Supplementary Information for details). We note that we add one additional constraint to the products reported here: spin multiplicity. While open-shell species can be products of the network, they are highly unlikely to be stable or meta-stable (long-lived radicals are generally rare). In the hopes of extracting useful chemical insights from network products, we therefore only consider network products that are singlets.

Kinetic Refinement of Reaction Mechanisms

The thermodynamic reaction pathways obtained via stochastic analysis were interrogated to determine the actual elementary steps. For the network products considered here (LFEO and bi-dioxolylidene), several low-cost thermodynamic reaction pathways were selected. For each elementary step along these pathways – excluding coordination reactions and redox reactions – we attempted to locate the TS using the AutoTS workflow,⁶² an end-to-end workflow to identify TS and reaction pathways that is built on top of the Jaguar electronic structure code (version 11.2).⁶³ All AutoTS calculations were conducted at a ω B97X-D/def2-SVPD(-f)/PCM level of theory,^{53,64,65} with water as the solvent. In some cases, for reactions involving two bonds changing, AutoTS identified two TS (for instance, one to form a bond and one to break a bond); these were optimized separately.

In cases where AutoTS was unable to find a TS for a given reaction, we searched using the single-ended growing string method (SE-GSM), as implemented in the pyGSM code.⁶⁶ SE-GSM calculations were conducted with a Q-Chem backend (version 5.3.2) at the ω B97X-D/def2-SVPD/PCM level of theory.⁶⁷ To be as consistent as possible, TS found using SE-GSM in Q-Chem were re-optimized in Jaguar at the ω B97X-D/def2-SVPD(-f)/PCM level of

theory.

For each TS, we confirmed that the optimized structure possessed one imaginary frequency and confirmed that it connected the expected endpoints. For cases where the endpoints consist of two molecules that are not covalently bound (typically bound only by coordination to Li⁺), we allow small imaginary frequencies (less than 75i cm⁻¹). These small imaginary modes can prove extremely difficult to remove using conventional geometry optimization methods, especially when they involve the motion of Li⁺ ions, and typically do not significantly affect the free energy. We note that in some cases, the barriers that we report are based on the difference between the TS and the reactants or products at infinite separation, rather than the entrance or exit complex. The electronic energies of all optimized structures (TS and endpoints) were corrected using a single-point calculation at a higher level of theory (ω B97X-V/def2-TZVPPD/SMD) in Q-Chem. The SMD parameters used were the same used for the construction of the LIBE dataset.³⁰ We note that we used Q-Chem for these calculations, rather than Jaguar, because the SMD implicit solvent model is not implemented in Jaguar at the time of this writing.

All AutoTS and pyGSM calculations were automated using workflows that we have implemented in the MPcat code (see Code Availability). These workflows are designed for high-throughput transition-state searches and reaction pathway analysis. Note that we use a fork of the original pyGSM code for SE-GSM (see Code Availability).

Data Availability

Molecular data used for network construction are the CHOLI subset of the Lithium Ion Battery Electrolyte (LIBE) dataset. LIBE is provided in Javascript Object Notation (JSON) format at <https://doi.org/10.6084/m9.figshare.14226464.v2>. All data used to construct mechanisms (molecular structures, thermodynamics, vibrational frequencies, and frequency modes) are also provided in JSON format in the supplementary file “reaction_pathways.json”.

Code Availability

All codes discussed here (HiPRGen, RNMC, MPcat, and pyGSM) are released open source on Github. A Python implementation of the HiPRGen method can be found at <https://github.com/BlauGroup/HiPRGen>. Please refer to the v0.1 release. RNMC, a performant kinetic Monte Carlo code in C++ and based on SPPARKS, can be found at <https://github.com/BlauGroup/RNMC>. Please refer to the v0.1 release. AutoTS and SE-GSM calculations were performed using the automated workflows defined in MPcat, which can be found at <https://github.com/espottesmith/MPcat>. Please refer to the v0.0.1 release. SE-GSM calculations specifically used a fork of the original pyGSM code, which can be found at <https://github.com/espottesmith/pyGSM/tree/c8cd99fcac451b1584f3f75e676f9d325e7ad6d4>.

Acknowledgements

This work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-

05CH11231, the Joint Center for Energy Storage Research, an Energy Innovation Hub funded by the US Department of Energy, Office of Science, Basic Energy Sciences, and the Silicon Consortium Project directed by Brian Cunningham under the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Vehicle Technologies of the U.S. Department of Energy, Contract No. DE-AC02-05CH11231.. Computational resources were provided by the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility under Contract No. DE-AC02-05CH11231, and by the Department of Energy’s Office of Energy Efficiency and Renewable Energy located at the National Renewable Energy Laboratory. This research also used the Lawrencium computational cluster resource provided by the IT Division at the Lawrence Berkeley National Laboratory (Supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231). Schrödinger, Inc. provided access to Jaguar and AutoTS software, as well as technical advice regarding their use.

Author Contributions

D.B. and E.W.C.S.-S. both have the right to list their names first when presenting this research or listing it in their CVs. D.B., E.W.C.S.-S., and S.M.B. conceived of the study and approach; D.B. implemented HiPRGen, RNMC, and stochastic analysis methods with feedback from S.M.B.; D.B., E.W.C.S.-S., S.D., and S.M.B. designed species and reaction filters; D.B., S.D., and S.M.B. designed the methods for network product identification; E.W.C.S.-S., N.S.R., and A.K. performed quantum chemical calculations; E.W.C.S.-S., A.K., and S.M.B. analyzed data; K.A.P. and S.M.B. secured funding; D.B., E.W.C.S.-S., and S.M.B. wrote the original manuscript; all authors edited the manuscript.

Notes and references

- 1 J. P. Unsleber and M. Reiher, *Annual Review of Physical Chemistry*, 2020, **71**, 121–142.
- 2 S. Maeda and K. Morokuma, *J. Chem. Theory Comput.*, 2012, **8**, 380–385.
- 3 A. L. Dewyer, A. J. Argüelles and P. M. Zimmerman, *WIREs Computational Molecular Science*, 2018, **8**, e1354.
- 4 J. S. Tse, *Annual Review of Physical Chemistry*, 2002, **53**, 249–290.
- 5 S. Maeda, Y. Harabuchi, M. Takagi, T. Taketsugu and K. Morokuma, *The Chemical Record*, 2016, **16**, 2232–2248.
- 6 C. Shang and Z.-P. Liu, *J. Chem. Theory Comput.*, 2013, **9**, 1838–1845.
- 7 Q. Zhao and B. M. Savoie, *Nat Comput Sci.*, 2021, **1**, 479–490.
- 8 C. Bannwarth, S. Ehler and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.
- 9 S. M. Blau, E. W. C. Spotte-Smith, B. Wood, S. Dwaraknath and K. A. Persson, *ChemRxiv*, 2020.
- 10 C. W. Gao, J. W. Allen, W. H. Green and R. H. West, *Computer Physics Communications*, 2016, **203**, 212–225.
- 11 C. F. Goldsmith and R. H. West, *J. Phys. Chem. C*, 2017, **121**, 9970–9981.
- 12 M. Liu, A. Grinberg Dana, M. S. Johnson, M. J. Goldman, A. Jocher, A. M. Payne, C. A. Grambow, K. Han, N. W. Yee, E. J. Mazeau, K. Blondal, R. H. West, C. F. Goldsmith and W. H. Green, *J. Chem. Inf. Model.*, 2021, **61**, 2686–2696.
- 13 D. Rappoport and A. Aspuru-Guzik, *J. Chem. Theory Comput.*, 2019, **15**, 4099–4112.
- 14 A. Wolos, R. Roszak, A. Źródło Dobrowolska, W. Beker, B. Mikulak-Klucznik, G. Spólnik, M. Dygas, S. Szymkuć and B. A. Grzybowski, *Science*, 2020, **369**, 1–12.
- 15 Y. Kim, J. W. Kim, Z. Kim and W. Y. Kim, *Chem. Sci.*, 2018, **9**, 825–835.
- 16 X. Jia, A. Lynch, Y. Huang, M. Danielson, I. Lang’at, A. Milder, A. E. Ruby, H. Wang, S. A. Friedler, A. J. Norquist and J. Schrier, *Nature*, 2019, **573**, 251–255.
- 17 F. Calle-Vallejo and M. T. M. Koper, *Electrochimica Acta*, 2012, **84**, 3–11.
- 18 Y. Y. Birdja, E. Pérez-Gallent, M. C. Figueiredo, A. J. Göttle, F. Calle-Vallejo and M. T. M. Koper, *Nat Energy*, 2019, **4**, 732–745.
- 19 L. Wang, A. Menakath, F. Han, Y. Wang, P. Y. Zavalij, K. J. Gaskell, O. Borodin, D. Iuga, S. P. Brown, C. Wang, K. Xu and B. W. Eichhorn, *Nature Chemistry*, 2019, **11**, 789–796.
- 20 R. P. Bell and C. N. Hinshelwood, *Proc. R. Soc. London A - Math Phys. Sci.*, 1936, **154**, 414–429.
- 21 M. Evans and M. Polanyi, *Trans. Faraday Soc.*, 1936, **32**, 1333–1360.

- 22 S. M. Blau, H. D. Patel, E. W. Clark Spotte-Smith, X. Xie, S. Dwaraknath and K. A. Persson, *Chemical Science*, 2021, **12**, 4931–4939.
- 23 X. Xie, E. W. Clark Spotte-Smith, M. Wen, H. D. Patel, S. M. Blau and K. A. Persson, *J. Am. Chem. Soc.*, 2021, **143**, 13245–13258.
- 24 D. T. Gillespie, *Journal of Computational Physics*, 1976, **22**, 403–434.
- 25 M. A. Gibson and J. Bruck, *J. Phys. Chem. A*, 2000, **104**, 1876–1889.
- 26 H. Eyring, *The Journal of Chemical Physics*, 1935, **3**, 107–115.
- 27 S. Stocker, G. Csányi, K. Reuter and J. T. Margraf, *Nature communications*, 2020, **11**, 1–11.
- 28 J. C. Baez, *Advances in Mathematical Physics*, 2018, **2018**, e7676309.
- 29 J. C. Baez and J. Biamonte, *arXiv:1209.3632 [math-ph, physics:quant-ph]*, 2019.
- 30 E. W. C. Spotte-Smith, S. M. Blau, X. Xie, H. D. Patel, M. Wen, B. Wood, S. Dwaraknath and K. A. Persson, *Sci Data*, 2021, **8**, 203.
- 31 E. W. C. Spotte-Smith, R. L. Kam, D. Barter, X. Xie, T. Hou, S. Dwaraknath, S. M. Blau and K. A. Persson, *ACS Energy Lett.*, 2022, **7**, 1446–1453.
- 32 B. Rowden and N. García-Araez, *Energy Reports*, 2020, **6**, 10–18.
- 33 P. Verma, P. Maire and P. Novák, *Electrochimica Acta*, 2010, **55**, 6332–6341.
- 34 B. L. D. Rinkel, D. S. Hall, I. Temprano and C. P. Grey, *J. Am. Chem. Soc.*, 2020, 22.
- 35 S. J. An, J. Li, C. Daniel, D. Mohanty, S. Nagpure and D. L. Wood, *Carbon*, 2016, **105**, 52–76.
- 36 R. Mogi, M. Inaba, Y. Iriyama, T. Abe and Z. Ogumi, *Journal of Power Sources*, 2003, **119–121**, 597–603.
- 37 J. Nanda, G. Yang, T. Hou, D. N. Voylov, X. Li, R. E. Ruther, M. Naguib, K. Persson, G. M. Veith and A. P. Sokolov, *Joule*, 2019, **3**, 2001–2019.
- 38 S. Tsubouchi, Y. Domi, T. Doi, M. Ochida, H. Nakagawa, T. Yamanaka, T. Abe and Z. Ogumi, *Journal of The Electrochemical Society*, 2012, **159**, A1786–A1790.
- 39 M. Nie, D. P. Abraham, D. M. Seo, Y. Chen, A. Bose and B. L. Lucht, *J. Phys. Chem. C*, 2013, **117**, 25381–25389.
- 40 M. Nie, J. Demeaux, B. T. Young, D. R. Heskett, Y. Chen, A. Bose, J. C. Woicik and B. L. Lucht, *J. Electrochem. Soc.*, 2015, **162**, A7008.
- 41 K. Leung, *Chemical Physics Letters*, 2013, **568**, 1–8.
- 42 Y. Wang, S. Nakamura, M. Ue and P. B. Balbuena, *Journal of the American Chemical Society*, 2001, **123**, 11708–11718.
- 43 L. D. Gibson and J. Pfaendtner, *Phys. Chem. Chem. Phys.*, 2020, **22**, 21494–21503.
- 44 D. Kuai and P. B. Balbuena, *ACS Appl. Mater. Interfaces*, 2022, **14**, 2817–2824.
- 45 M. Anouti, Y. R. Dougassa, C. Tessier, L. El Ouattani and J. Jacquemin, *J. Chem. Thermodyn.*, 2012, **50**, 71–79.
- 46 M. Onuki, S. Kinoshita, Y. Sakata, M. Yanagidate, Y. Otake, M. Ue and M. Deguchi, *Journal of The Electrochemical Society*, 2008, **155**, A794.
- 47 B. L. D. Rinkel, D. S. Hall, I. Temprano and C. P. Grey, *J. Am. Chem. Soc.*, 2020, **142**, 15058–15074.
- 48 R. A. Moss and M. P. Doyle, *Contemporary carbene chemistry*, John Wiley & Sons, 2013.
- 49 M. J. Boyer and G. S. Hwang, *J. Phys. Chem. C*, 2019, **123**, 17695–17702.
- 50 J. Young, P. M. Kulick, T. R. Juran and M. Smeu, *ACS Appl. Energy Mater.*, 2019, **2**, 1676–1684.
- 51 C. A. Grambow, L. Pattaiaik and W. H. Green, *J. Phys. Chem. Lett.*, 2020, **11**, 2992–2997.
- 52 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.
- 53 D. Rappoport and F. Furche, *J. Chem. Phys.*, 2010, **133**, 134105.
- 54 N. Mardirossian and M. Head-Gordon, *Physical Chemistry Chemical Physics*, 2014, **16**, 9904–9924.
- 55 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 4538–4543.
- 56 J. J. Guerard and J. S. Arey, *J. Chem. Theory Comput.*, 2013, **9**, 5046–5058.
- 57 M. I. Chaudhari, J. R. Nair, L. R. Pratt, F. A. Soto, P. B. Balbuena and S. B. Rempe, *J. Chem. Theory Comput.*, 2016, **12**, 5709–5718.
- 58 E. Glendening, J. Badenhoop, A. Reed, J. Carpenter, J. Bohmann, C. Morales and F. Weinhold, *Theoretical Chemistry Institute, University of Wisconsin: Madison, WI, USA*, 2001.
- 59 F. Weinhold, C. Landis and E. Glendening, *International Reviews in Physical Chemistry*, 2016, **35**, 399–440.
- 60 S. Plimpton, A. Thompson and A. Slepoy, *Stochastic Parallel PARticle Kinetic Simulator*, Sandia National Lab. (SNL-NM), Albuquerque, NM (United States) Technical Report SPPARKS, 2008.
- 61 S. Plimpton, C. Battaile, M. Ch, L. Holm, A. Thompson, V. Tikare, G. Wagner, X. Zhou, C. G. Cardona and A. Slepoy, *Crossing the Mesoscale No-Man's Land via Parallel Kinetic Monte Carlo*, 2009.
- 62 L. D. Jacobson, A. D. Bochevarov, M. A. Watson, T. F. Hughes, D. Rinaldo, S. Ehrlich, T. B. Steinbrecher, S. Vaitheeswaran, D. M. Philipp, M. D. Halls and R. A. Friesner, *J. Chem. Theory Comput.*, 2017, **13**, 5780–5797.
- 63 A. D. Bochevarov, E. Harder, T. F. Hughes, J. R. Greenwood, D. A. Braden, D. M. Philipp, D. Rinaldo, M. D. Halls, J. Zhang and R. A. Friesner, *International Journal of Quantum Chemistry*, 2013, **113**, 2110–2142.
- 64 J.-D. Chai and M. Head-Gordon, *Physical Chemistry Chemical Physics*, 2008, **10**, 6615–6620.
- 65 B. Mennucci, *WIREs Computational Molecular Science*, 2012, **2**, 386–404.
- 66 C. Aldaz, J. A. Kammeraad and P. M. Zimmerman, *Physical Chemistry Chemical Physics*, 2018, **20**, 27394–27405.
- 67 E. Epifanovsky, A. T. B. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, Z. Gan, D. Hait, P. R. Horn, L. D. Jacobson, I. Kaliman, J. Kussmann, A. W. Lange, K. U. Lao, D. S. Levine, J. Liu, S. C. McKenzie, A. F. Morrison, K. D. Nanda, F. Plasser, D. R. Rehn, M. L. Vidal, Z.-Q. You, Y. Zhu, B. Alam, B. J. Albrecht, A. Aldosary, E. Alguire, J. H. Andersen, V. Athavale, D. Barton, K. Begam, A. Behn, N. Bellonzi, Y. A. Bernard, E. J. Berquist, H. G. A. Burton, A. Carreras, K. Carter-Fenk, R. Chakraborty, A. D. Chien, K. D. Closser, V. Cofer-Shabica, S. Dasgupta, M. de Wergifosse, J. Deng, M. Diedenhofen, H. Do, S. Ehler, P.-T. Fang, S. Fatehi, Q. Feng, T. Friedhoff, J. Gayvert, Q. Ge, G. Gidofalvi, M. Goldey, J. Gomes, C. E. González-Espinoza, S. Gulania, A. O. Gunina, M. W. D. Hanson-Heine, P. H. P. Harbach, A. Hauser, M. F. Herbst, M. Hernández Vera, M. Hodecker, Z. C. Holden, S. Houck, X. Huang, K. Hui, B. C. Huynh, M. Ivanov, A. Jas, H. Ji, H. Jiang, B. Kaduk, S. Kähler, K. Khistyayev, J. Kim, G. Kis, P. Klunzinger, Z. Koczor-Benda, J. H. Koh, D. Kosenkov, L. Koulias, T. Kowalczyk, C. M. Krauter, K. Kue, A. Kunita, T. Kus, I. Ladjanszki, A. Landau, K. V. Lawler, D. Lefrancois, S. Lehtola, R. Li, Y.-P. Li, J. Liang, M. Liebenthal, H.-H. Lin, Y.-S. Lin, F. Liu, K.-Y. Liu, M. Loipersberger, A. Luenser, A. Manjanath, P. Manohar, E. Mansoor, S. F. Manzer, S.-P. Mao, A. V. Marenich, T. Markovich, S. Mason, S. A. Maurer, P. F. McLaughlin, M. F. S. J. Menger, J.-M. Mewes, S. A. Mewes, P. Morgante, J. W. Mullinax, K. J. Oosterbaan, G. Paran, A. C. Paul, S. K. Paul, F. Pavošević, Z. Pei, S. Prager, E. I. Proynov, A. Rak, E. Ramos-Cordoba, B. Rana, A. E. Rask, A. Rettig, R. M. Richard, F. Rob, E. Rossomme, T. Scheele, M. Scheurer, M. Schneider, N. Sergueev, S. M. Sharada, W. Skomorowski, D. W. Small, C. J. Stein, Y.-C. Su, E. J. Sundstrom, Z. Tao, J. Thirman, G. J. Tornai, T. Tsuchimochi, N. M. Tubman, S. P. Veccham, O. Vydrov, J. Wenzel, J. Witte, A. Yamada, K. Yao, S. Yeganeh, S. R. Yost, A. Zech, I. Y. Zhang, X. Zhang, Y. Zhang, D. Zuev, A. Aspuru-Guzik, A. T. Bell, N. A. Besley, K. B. Bravaya, B. R. Brooks, D. Casanova, J.-D. Chai, S. Coriani, C. J. Cramer, G. Cseray, A. E. DePrince, R. A. DiStasio, A. Dreuw, B. D. Dunietz, T. R. Furlani, W. A. Goddard, S. Hammes-Schiffer, T. Head-Gordon, W. J. Hehe, C.-P. Hsu, T.-C. Jagau, Y. Jung, A. Klamt, J. Kong, D. S. Lambrecht, W. Liang, N. J. Mayhall, C. W. McCurdy, J. B. Neaton, C. Ochsenfeld, J. A. Parkhill, R. Peverati, V. A. Rassolov, Y. Shao, L. V. Slipchenko, T. Stauch, R. P. Steele, J. E. Subotnik, A. J. W. Thom, A. Tkatchenko, D. G. Truhlar, T. Van Voorhis, T. A. Wesolowski, K. B. Whaley, H. L. Woodcock, P. M. Zimmerman, S. Faraji, P. M. W. Gill, M. Head-Gordon, J. M. Herbert and A. I. Krylov, *J. Chem. Phys.*, 2021, **155**, 084801.