Electronic Supplementary Material (ESI) for Digital Discovery.

Supplementary Information: 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 *]

[a] *Energy Storage and Distributed Resources, Lawrence Berkeley National Laboratory*
[b] *Department of Materials Science and Engineering, University of California Berkeley*
[c] *Materials Science Division, Lawrence Berkeley National Laborory*
[d] *Department of Chemical and Biomolecular Engineering, University of California, Berkeley*
[e] *Department of Electrical Engineering and Computer Science, University of California, Berkeley*
[f] *Luxembourg Institute of Science and Technology, Luxembourg*
[g] *Molecular Foundry, Lawrence Berkeley National Laboratory*
[†] *These authors contributed equally to this work*
[*] *Correspondence: smblau@lbl.gov*

Pathway Identification from Monte Carlo Trajectories

A reaction network consists of a set of species and a set of reactions linking them. Given a chemical system, we are interested in exploring the reaction pathways which produce particular species of interest. Stochastic trajectories are a useful tool for approaching the problem because they can be simulated efficiently, even when the network has hundreds of millions of reactions.[1] Unfortunately, since we collapse all spatial aspects of the system, identical molecules become indistinguishable (in reality, identical molecules can be distinguished by their locations in space). This creates the following problem: Suppose we are interested in the production of species G from A and have the following simulation trajectory:

$$\begin{aligned} A &\to Z + F \\ F &\to X \\ A &\to B + F \\ A &\to C + H \\ A &\to D + H \\ C &\to E \\ D &\to E \\ E + F &\to G \end{aligned}$$

It is impossible to decide between the two pathways

$$\begin{array}{cccc} A \to B + F & A \to C + H & C \to E & E + F \to G \\ A \to B + F & A \to D + H & D \to E & E + F \to G \end{array}$$

If our model had a spacial aspect, we would be able to trace the specific E used back to either a C or a D. Fundamentally, this ambiguity is caused by sequence

$$A \to C + H \qquad C \to E \qquad E \to D \qquad D + H \to A$$

which is called a deficiency loop.[2] To avoid this problem, we extract a pathway which produces the target in the following way: Take the first reaction x which produces the target molecule. Then recursively, take the first reactions which produced the reactants of x. If a reactant is a starting molecule, then stop. Applying the procedure to the above sequence gives the pathway

$$A \to C + H \qquad C \to E \qquad A \to Z + F \qquad E + F \to G$$

Intuitively, this procedure is producing pathways which don't take into account competition, since there is no guarantee that the first molecule which is produced is not immediately consumed by some competing reaction. Since, in this work, we identify pathways using Monte Carlo simulation with thermodynamically bounded rate coefficients, the problem is not as bad as it seems. Since all reactions have the same rate coefficient, the only thing which can cause one reaction to out-compete another is relative abundance of reactant species. Reaction pathways do not get shut down by some competing reaction with a much higher rate coefficient, so if we have two reactions $A + B \rightarrow C$ and $A + D \rightarrow E$, if the number of Bs and Ds have equalized, then both reactions are equally likely to fire.

Since we collapse trajectories to pathways in this way and then rank them using the cost function $\Phi_{total} = \sum_x \Phi_x = \sum_x 1 + \exp(\Delta G_x / k_B T)$, where the sum is over all reactions and ΔG_x is the free energy of reaction x, frequently occurring pathways do not necessarily have low costs, and visa versa. This means we are not guaranteed to find the lowest cost pathway via exhaustively sampling for large networks. In practice, we do not believe that this is an issue, as our approach nonetheless produces many low-cost, chemically reasonable paths.

To verify that our stochastic approach, while not theoretically guaranteed to identify the shortest pathway, nonetheless can, we applied it to a previously constructed network of roughly 4.5 million reactions.[3] We note that this network was not constructed using the High-Performance Reaction Generation (HiPRGen) method, but rather as described by Blau et al. For this network, our stochastic method finds exactly the same top pathways to lithium ethylene dicarbonate (LEDC) as our previously published method which used graph-based pathfinding algorithms that are guaranteed to find the shortest paths.

Comparing HiPRGen to Previous Methods of Reaction Enumeration

The first step towards using a CRN to analyze reactivity is to construct a network by selecting species to include and reactions between those species. In the main text, we present the HiPRGen method, which relies on a series of user-extensible filters to selectively remove species and reactions. Here, we place this method in context by comparing it to our previous work,[3,4] pointing out the benefits of HiPRGen.

Our first attempt to interrogate electrochemical reactivity using CRNs - in the work by Blau et al.[3] - began similarly to HiPRGen. A large set of species was provided as input, and these species were then filtered. We note that the filters used by Blau et al. were more rudimentary than those in HiPRGen, only removing species with identical charge and connectivity. Blau et al., furthermore, allowed for multiple “coordimers” (species with identical composition, charge, and covalent connectivity, but with different placements of lithium ions), leading to many effectively degenerate reactions with covalently equivalent reactants and products being added to the network.

For reaction addition, Blau et al. first included all reactions involving no change in bonding but a change in charge of ±1 (one-electron redox) or a change in one bond but no change in charge (one-bond formation or cleavage). Then, based on the one-bond formation or cleavage reactions, more complex mechanisms involving two bonds changing were added. Specifically, exergonic two-bond reactions were added if and only if they could combine two one-bond reactions while allowing a high-energy intermediate to be avoided. That is, if there were two one-bond reactions, $A \rightarrow B$ with $\Delta G_{A \rightarrow B} > 0$ and $B \rightarrow C$ with $\Delta G_{B \rightarrow C} < 0$, then B is a high-energy intermediate. In this situation, the reaction $A \rightarrow C$ would be included only if the overall free energy $\Delta G_{A \rightarrow C} < 0$. Alternatively, if $\Delta G_{A \rightarrow B} < 0$, then $A \rightarrow C$ will not be added to the network, because both $A \rightarrow B$ and $B \rightarrow C$ are exergonic and there is therefore no high-energy intermediate. This method relies on the presence of relatively unstable species for the identification of any reaction more complex than the cleavage or formation of a single bond. Because such species might be especially difficult (or even impossible) to optimize using DFT, it is likely that many concerted mechanisms are not identified and added due to missing intermediates.

In the work of Xie et al.,[4] high-energy intermediates were not used to identify concerted mechanisms. Rather, mixed-integer linear programming (MILP) was employed to identify the minimum number of bonds that had to change for a stoichiometrically balanced reaction to occur. Given some set of species, all reactions with a "chemical distance" between reactants and products less than some threshold were included in the network (in Xie et al., networks with a maximum chemical distance of 2 and a maximum chemical distance of 5 were constructed). As the maximum chemical distance is increased, the number of possible reactions increases dramatically. A network with only 570 species had nearly 9,000,000 reactions when the maximum chemical distance was set to 5. For this method to be viable in practice, the number of species included must be kept small, which makes this method less suitable for expansive exploration of reactive spaces. Moreover, since no filtering was applied after the initial MILP-based reaction enumeration, many of the reactions included in the network were not actually chemically reasonable and could not proceed in a single step, meaning that significant manual work had to be undertaken to turn thermodynamic, network-predicted pathways into full mechanisms consisting only of elementary steps.

HiPRGen, like both the work of Blau et al. and Xie et al., begins with a set of species. This set is then filtered such as to not severely limit the possible species that can form (as in Xie et al.) but to remove chemically unreasonable or redundant molecules (improving upon the method of Blau et al.). As in Xie et al., all stoichiometrically valid reactions are first generated, but rather than relying solely on chemical distance, a suite of filters are used to limit the reactions to a set that is both computationally tractable and chemically reasonable. As a result, an automated transition-state workflow is able to kinetically refine many pathways with limited human intervention, turning thermodynamic network-predicted pathways into full mechanisms with a fraction of the manual work that was necessary for a similar conversion by Xie et al. The ability to refine pathways into full kinetic mechanisms with minimal human intervention will be essential for rapidly exploring such massive reactive spaces towards the construction of microkinetic models in the future. We note that, using our filters (as described below), we effectively limit our network to a maximum chemical distance of 2.

It is worth noting that HiPRGen is drastically more computationally efficient than the method of Xie et al. The network in the work of Xie et al. containing roughly 9,000,000 unfiltered reactions between 570 species took approximately 2 weeks to generate using 20 Intel Xeon Gold 6230 compute nodes, while the network reported here containing over 86,000,000 filtered reactions between nearly 5,200 species was generated in roughly 12 hours on 10 Intel Xeon Phi 7250 "Knight's Landing" (KNL) nodes. Because, in our benchmarking, the Xeon Gold 6230 nodes are at least two times as fast as the KNL nodes, this means that HiPRGen is at least 1,200 times more efficient in terms of either cost per reaction or cost per species.

Complexity Reduction Through Reaction Network Construction and Analysis

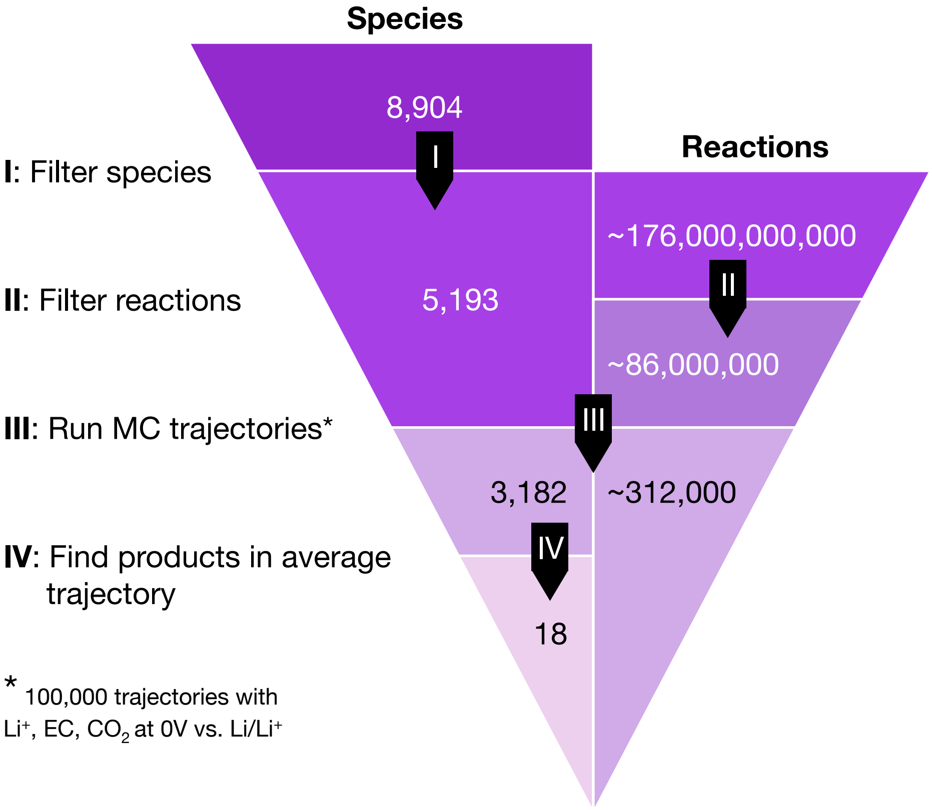


Figure 1: **CRN construction and stochastic analysis for SEI formation yields a systematic reduction of complexity. I.** 8,904 initial species are reduced to 5,193 species via HiPRGen species filters. **II.** Of the approximately 176 billion possible stoichiometric reactions between the 5,193 species, just over 86 million pass HiPRGen reaction filters. **III.** 100,000 Monte Carlo trajectories are run to completion with Li^+, EC, and CO_2 as starting species at +0.0V vs. Li/Li^+, and in total only 3182 species and approximately 312,000 unique reactions are observed. **IV.** Heuristic analysis of the average trajectory yields only 18 network product species.

Our approach is unique in its capacity to describe the complexity of a reactive space including thousands of charged and open shell species. Figure 1 graphically depicts the systematic reduction in complexity that is inherent to filter-based network construction followed by stochastic analysis. While HiPRGen seeks to preserve the vast range of chemically reasonable reactions that could occur in a system containing charged and radical species, only 312,034 of the over 86 million unique reactions in the network are actually observed in a set of 100,000 kMC trajectories run to completion from one initial condition. Further, of the 3,182 unique species observed to form in the same set of trajectories, just 18 are identified as network products. Only an automated computational infrastructure could possibly navigate such a vast space of interconnecting species and reactions to guide further investigation, and no previously reported method can be applied on this scale.

Solvent Corrections to Electronic Energy

Selection of Correction Factor

As noted in the main text, to correct for the inaccurate Li^+ stabilization of SMD, we sought to approximate the stabilizing effect of a solvation shell comprised of ethylene carbonate (EC). Specifically, we assume that all species in our CRN are sufficiently stable such that all Li^+ ions have full solvation shells consisting of four coordinate bonds (either from solvent or other coordinating species). To correct for insufficient Li^+ stabilization and approximate the effect of an explicit solvent shell, we calculated the electronic energy of $\mathrm{Li}^+\mathrm{EC}_n$ in implicit solvent (optimized using PCM and then single-point corrected with SMD to reduce cost). The relative stabilization was calculated as

$$\Delta E = E_{Li^+EC_n} - E_{Li^+} - nE_{EC_n} \tag{1}$$

Performing this calculation with $n = 0, 1, 2, 3, 4$, we obtain a roughly linear trend (Figure 2) with slope -0.746 eV/(molecule EC) ($R^2 = 0.994$). When free energy is accounted for, this solvent correction to the electronic energy (-0.746 eV per missing coordinate bond) actually results in the coordination of Li^+ with EC being slightly endergonic. We therefore alter this value to -0.68 eV, at which point the coordination of Li^+ with EC is essentially isergonic (-0.01 eV).

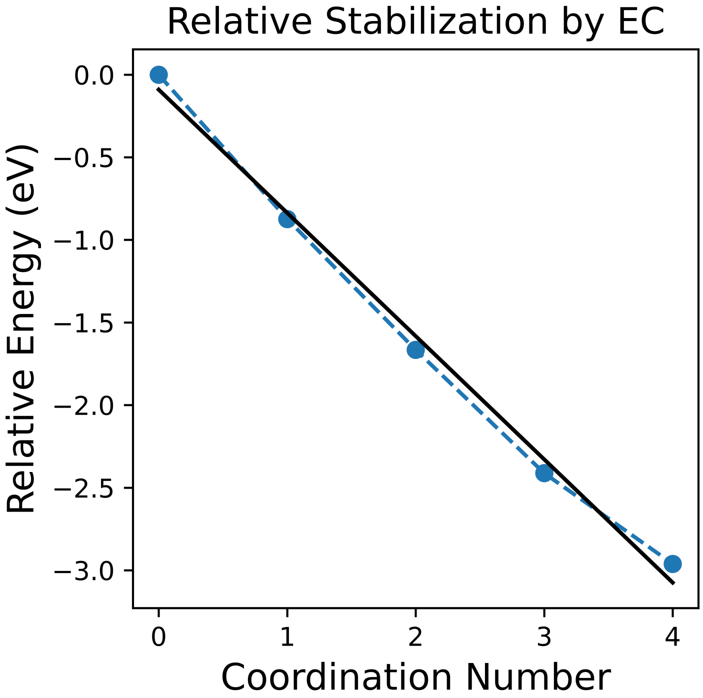


Figure 2: **Relative stabilization of Li^+ by increasing numbers of EC.**

Selection of Coordimers

We found that the SMD model,[5] which was used to construct the LIBE dataset, insufficiently stabilized Li^+ ions. This is somewhat unsurprising, given that small ions like Li^+ were not included in the training and parameterization of SMD. As a result of this insufficient stabilization, species where Li^+ had more coordinate bonds were more stable than species where Li^+ had fewer coordinate bonds to an unphysical degree. This affects the selection of stable coordimers. As an example, we show LEDC (Figure 3). In previous molecular dynamics studies,[6–8] it has generally been found that LEDC prefers a roughly linear conformation both when ordered in the solid state and when present in an amorphous SEI or liquid electrolyte. However, when using SMD with DFT, a "puckered" conformation is preferred by 0.68 eV because each Li^+ has 3 coordinate bonds (compared to only two in the linear LEDC conformation). As further evidence that this result is based solely on the insufficient stabilization of the Li^+, upon removing the Li^+ from the puckered conformation, the structure optimizes to a roughly linear conformation. When the solvation correction is applied, the linear conformer is identified as the more stable by 0.68 eV.

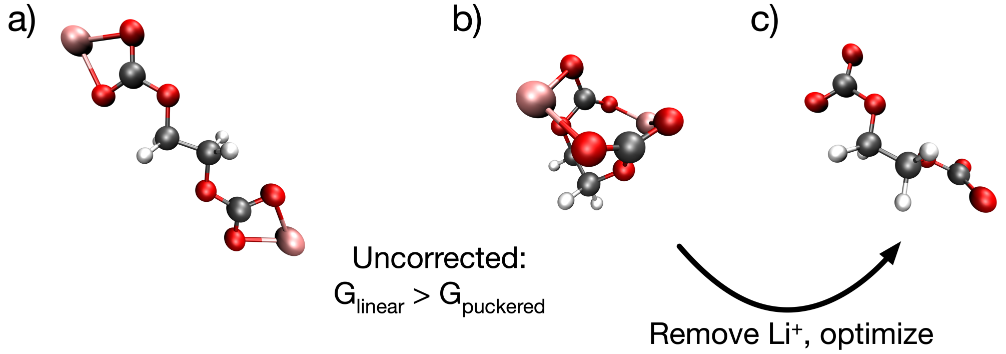


Figure 3: **Conformations of LEDC.** The linear conformer (libe-115795, **a**) is less stable than the puckered conformer (libe-652486, **b**) by 0.68 eV, although the linear conformer is more commonly observed in all-atom molecular dynamics simulations. When the Li^+ are removed from the puckered conformer and the structure is re-optimized (charge -2), a roughly linear conformer is again obtained (**c**), indicating that the puckered conformer is only preferred because of the insufficient stabilization of Li^+ in implicit solvent.

Reactions Involving Changes in Coordination Number

In addition to facilitating the selection of reasonable coordimers, the solvation correction also impacts reactions involving bond changes. For reactions where the number of coordination bonds does not change (Figure 4a), the same correction is applied to the reactants and the products, so there is no net effect on the reaction thermodynamics. However, if there is a change in the number of coordinate bonds between reactants and products (Figure 4b), then the solvation correction alters the reaction free energy ($\Delta\Delta G = 0.68(\Delta n_c)$), where Δn_c is the change in the number of coordinate bonds in the reaction ($\Delta n_c = n_{c,products} - n_{c,reactants}$). In some cases, this change can make reactions that are endergonic without a solvation correction exergonic, and vice versa.

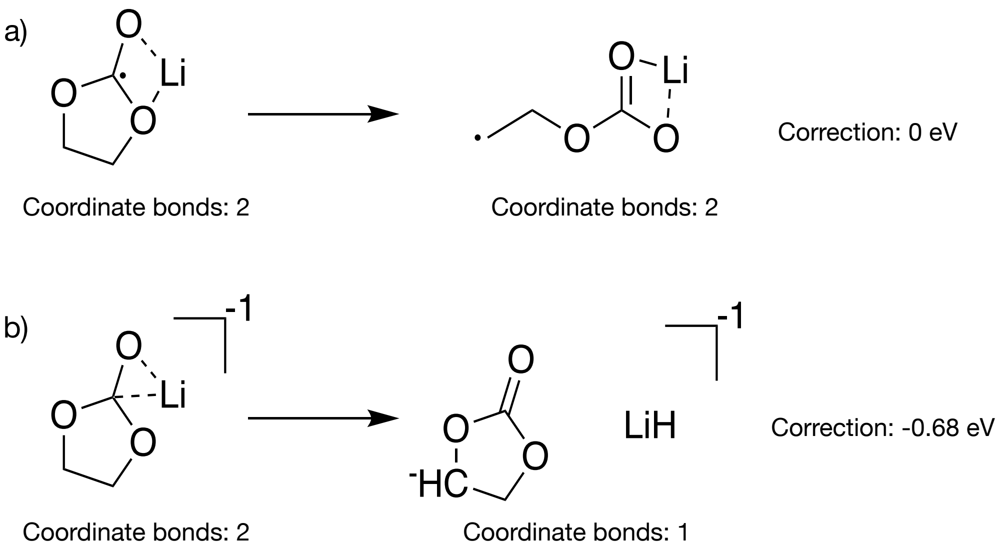


Figure 4: **Solvation corrections applied to reactions with bonds changing. a)** When the number of coordinate bonds in the reactants and the products are the same, then the correction to the reaction free energy is 0 eV. **b)** However, when there is a change in the number of coordinate bonds (here, there are two coordinate bonds in the reactants and only one in the products), then the reaction free energy changes based on the change in coordinate bonds.

When Not to Apply a Solvation Correction

Redox Reactions

Initially, we applied a solvation correction to all reactions, including reduction and oxidation reactions. However, based on previous studies that showed that SMD can be used for the accurate prediction of redox potentials,[9] we evaluated this choice using a small benchmark, comparing the uncorrected redox potentials calculated using the data in LIBE to experiment.

The experimental data relevant to molecules in lithium-ion battery electrolytes is relatively scarce. While some redox potentials for common electrolyte species have been reported in the literature, even comparing to this data must be done with caution because there can be considerable uncertainty and disagreement between measurements. For instance, the reduction potential of ethylene carbonate (EC) has been measured to be as low as +0.35V[10] and as high as +1.36V[11] vs. Li/Li$^+$ (though the most common value seen in the literature is around +0.75V). Moreover, side reactions like the trans-esterification of linear carbonates[12] make the interpretation of experimental data challenging. We calculate reduction potentials as

$$E(Li/Li^{+}) = -(G_{reduced} - G_{non-reduced}) - 1.4\ V \tag{2}$$

and oxidation potentials as

$$E(Li/Li^{+}) = (G_{oxidized} - G_{non-oxidized}) - 1.4\ V \tag{3}$$

where all free energies are given in units of eV. The scaling factor of $\sim -1.4\ V$, derived from the difference between the energy of an electron in vacuum ($-4.44\ eV$) and the absolute potential of a Li electrode ($-3.05\ eV$),[13] converts the calculated reduction potential to

Molecule	**ID (non-reduced)**	**ID (reduced)**	E_{red} **(V) (calc)**	E_{red} **(V) (exp)**
LiEC	libe-115918	libe-115782	0.65	0.75[14]
LiVC	libe-115855	libe-120798	0.87	1.0[14], 1.1[15]
LiFEC	libe-135858	libe-140687	0.97	0.9, 1.15[14]
LiDEC	libe-173831	libe-173906	0.60	0.71 - 0.90[16]
LiEMC	libe-173686	libe-173778	0.62	≈0.45[14]

Table 1: : Comparison of computed and experimental reduction potentials for lithium-ion battery electrolyte solvents (ethylene carbonate (EC), vinylene carbonate (VC), fluoroethylene carbonate (FEC), diethyl carbonate (DEC), and ethyl methyl carbonate (EMC)) coordinated with Li^+.

Molecule	**ID (non-oxidized)**	**ID (oxidized)**	E_{ox} **(V) (calc)**	E_{ox} **(V) (exp)**
EC	libe-115834	libe-115879	6.98	6.2[17]
DMC	libe-202458	libe-202592	6.84	6.7[17]
DEC	libe-175140	libe-175285	6.69	6.7[17]
EMC	libe-185427	libe-186936	6.74	6.7[17,18]

Table 2: : Comparison of computed and experimental oxidation potentials for lithium-ion battery electrolyte solvents (ethylene carbonate (EC), dimethyl carbonate (DMC), diethyl carbonate (DEC), and ethyl methyl carbonate (EMC)).

be relative to a Li/Li^+ reference electrode, which is standard in the lithium-ion battery community.

As can be seen, our calculations tend to somewhat underestimate reduction potentials (except for EMC, which we note has a somewhat more approximate experimental reduction potential). However, we are generally quantitatively close to the experimental value (within roughly 0.2V), and we correctly capture trends - for instance, that the electrolyte additives FEC and VC should reduce before EC, and that cyclic carbonates are in general easier to reduce than linear carbonates. With the exception of EC, for which we significantly overestimate the oxidation potential, our calculated E_{ox} are in excellent agreement with experiment. Because of this generally good agreement, we chose not to include a solvation correction for redox reactions.

Full Mechanisms

Perhaps surprisingly, we only apply a solvation correction to reactions where energy barriers are unknown. The reason is based on the initial assumption of the solvation correction - that at all points, Li^+ has a full solvation shell. This is effectively an equilibrium or quasi-equilibrium argument. At any particular moment in time, fluctuations in the electrolyte may cause Li^+ to have either less than or more than the preferred number of coordinate bonds. These fluctuations should be relatively rare, however, and on average we expect the coordination number of Li^+ to be four.

This argument breaks down at the transition-state (TS). TS are inherently points of instability that are typically occupied for vanishingly small times. Moreover, at the TS, bonds are commonly in a state of rearrangement - forming and breaking. While it is possible that all Li^+ might be fully coordinated at a TS, there is no reason why this should be so in general. Rather than try to apply an equilibrium argument to a non-equilibrium point of the potential energy surface, we choose not to apply a solvation correction at TS. To be consistent with this choice, when reporting full mechanisms, all free energies are uncorrected. We note that devising reasonable corrections for the energy or free energy of full reaction mechanisms in solution, including TS, is a topic for future research.

Identification of Bonding In Molecules

Much of the HiPRGen method of CRN generation relies on bonding - to identify coordimers, to filter species, and to filter reactions (more details below). It is therefore worthwhile to describe how bonds are identified in HiPRGen.

In this work, HiPRGen took as input entries from the LIBE dataset. The bonds in the molecules contained in LIBE are identified using a combination of the bond detection algorithms in OpenBabel[19,20] and Critic2,[21,22] as well as a heuristic method.[23] Two possible modifications can be made to the LIBE-defined bonding. First, in rare cases, molecules in LIBE contain hydrogen atoms with two covalent bonds. Where this occurs, we discard the longer of the two bonds. Second, we re-assess coordinate bonds. In addition to any coordinate bonds present in the molecular graph representation in LIBE, we include any coordinate bonds between Li^+ and an atom X if

1. X has a partial charge less than 0, as calculated by either the Mulliken or Restrained Electrostatic Potential (RESP) methods

2. The distance between Li^+ and X is less than or equal to 2.4 Å (note that the distance cutoff in LIBE is slightly larger, at 2.5 Å).

Explaining Species Filters

Species filters are an important component of the HiPRGen method, removing unphysical or redundant molecules. We note that, while effectively a species filter, coordimer selection can only be carried out if a set of otherwise acceptable coordimers are known. Therefore, it is carried out after application of the filters described below:

1. **Metal non-cation filter**: Remove species with a lithium atom with an NBO partial charge less than 0.1. We assume that if a Li^+ ion is reduced, it should rapidly desolvate and either plate or intercalate. Therefore, we do not want to include any species containing Li^0 or Li^- in the CRN.

2. **Molecule not connected**: Remove species that do not have a connected full molecule graph as defined by both covalent and coordinate bonding. This indicates that the structure contains separate molecules and should not be treated as a single species in the CRN. Each network species being an individual molecule is essential for bounding complexity. If a species is allowed to be composed of multiple molecules, the number of species grows substantially without any benefit, e.g. [A, B, C] becomes [A, B, C, A+B, A+C, B+C, A+A, B+B, C+C, etc]. Meanwhile, the number of reactions

similarly grows without encoding any new information, e.g. "A and B react to form C" ends up being represented additionally as "A+B reacts to form C" and "A and B+C reacts to form C+C" and "A+B and C reacts to form C+C" and "A+A and B react to form A+C", etc.

3. **Metal-centric complex**: Remove species which only have a connected full molecule graph due to the presence of a metal ion, e.g. A — Li^+ — B. Since metal coordinate bonds are electrostatic interactions that are typically much weaker than covalent bonds, such a species is effectively disconnected A and B molecules which can in principle participate in all of the same reactions that A or B or Li^+A or Li^+B could participate in individually. Therefore, for the same reasons stated above, such species must be removed. Chemically, this is equivalent to assuming that non-solvent molecules exist in sufficiently low concentration that each Li^+ is always coordinated to at most one non-solvent molecule. In the future, we will explore situations where this approximation breaks down, but it is presently necessary.

We note that species filters can be easily modified or extended by users to suit their chemical application.

Explaining Reaction Filters

The heart of the HiPRGen method is the massively parallelized application of reaction filters. The computational cost and memory footprint of each filter described below has been carefully engineered to eliminate bottlenecks and allow HiPRGen to be tractably applied to systems on the scale of our Li-ion SEI application. While chemical examples and TS from our specific system motivate and justify some of the filters developed, we believe the filters are generally applicable as a starting point to model a diverse range of complex reactivity. Further, HiPRGen has been intentionally engineered to enable the straightfoward addition of new filters as well as the modification or removal of any of the filters employed in this work in order to be easily customized to suit any system of interest.

1. **Redox reactions**, as defined by a change in total charge between reactants and products:
   (a) **Too many reactants or products**: Remove redox reactions with more than one reactant or more than one product. This partially enforces the assumption that redox reactions occur separately from covalent bond breakage/formation. While there are known dissociative redox reactions, in which a redox process occurs simultaneous with a covalent bond breakage,[24] they are uncommon, and thus we exclude them in general.
   (b) **Reactant and product not covalent isomorphic**: Remove redox reactions with one reactant and one product where the covalent bonds of the reactant and product are not equivalent. In combination with the previous filter, this completes enforcement of the assumption that redox reactions occur separately from covalent bond breakage/formation. Note that Li^+ coordinate bonds are allowed to change

during a redox reaction as long as there is at least one coordinate bond in both the reactant and the product.

   (c) **Change in charge too large**: Remove redox reactions involving the simultaneous addition or removal of two electrons. As above, there are known simultaneous two-electron redox processes, but they are very rare, and thus we exclude them in general.

2. **Both redox and non-redox reactions**:

   (a) **ΔG above threshold**: Remove all reactions with $\Delta G > 0$ eV. If we include endergonic reactions over even a very small range - e.g. corresponding to the expected magnitude of DFT uncertainty or what would be thermally accessible - then we introduce loops into our CRN, as the corresponding exergonic reverse reactions will also be present. Such loops are detrimental to pathfinding, as trajectories become filled with unimportant back-and-forth processes. This in turn obscures network product identification which depends on meaningful counts of species formation and consumption. Such values are only rigorously well-defined when trajectories are all guaranteed to run to completion, which is only guaranteed when the network contains no loops. Therefore, at present, we enforce the approximation that the network is composed of only exergonic reactions.

3. **Non-redox reactions**:

   (a) **Star count difference above threshold**:

   A molecule can be characterized by its "stars", where a star is defined by the identity of a central atom and the identity of the neighboring atom(s) it is bonding with. Each atom in a molecule will be the center of a star, and thus a molecule will have the same number of stars as it has atoms. For example, the molecule CH_3OH (methanol) will have six stars:

   - Center: C, bonded atoms: H, H, H, O
   - Center: O, bonded atoms: H, C
   - Center: H, bonded atoms: O
   - (Center: H, bonded atoms: C) x3

   shown graphically in Figure 5a. If we break the C-O bond in methanol, the product stars would be:

   - Center: C, bonded atoms: H, H, H
   - Center: O, bonded atoms: H
   - Center: H, bonded atoms: O
   - (Center: H, bonded atoms: C) x3

   shown graphically in Figure 5b. As could be expected, both the C-centric and the O-centric stars are different than they were previously. Therefore, as shown in Figure 6, the total star count difference of this reaction is four; there are two stars in the reactant that are not found in the products, and there are two stars in the products not found in the reactant.

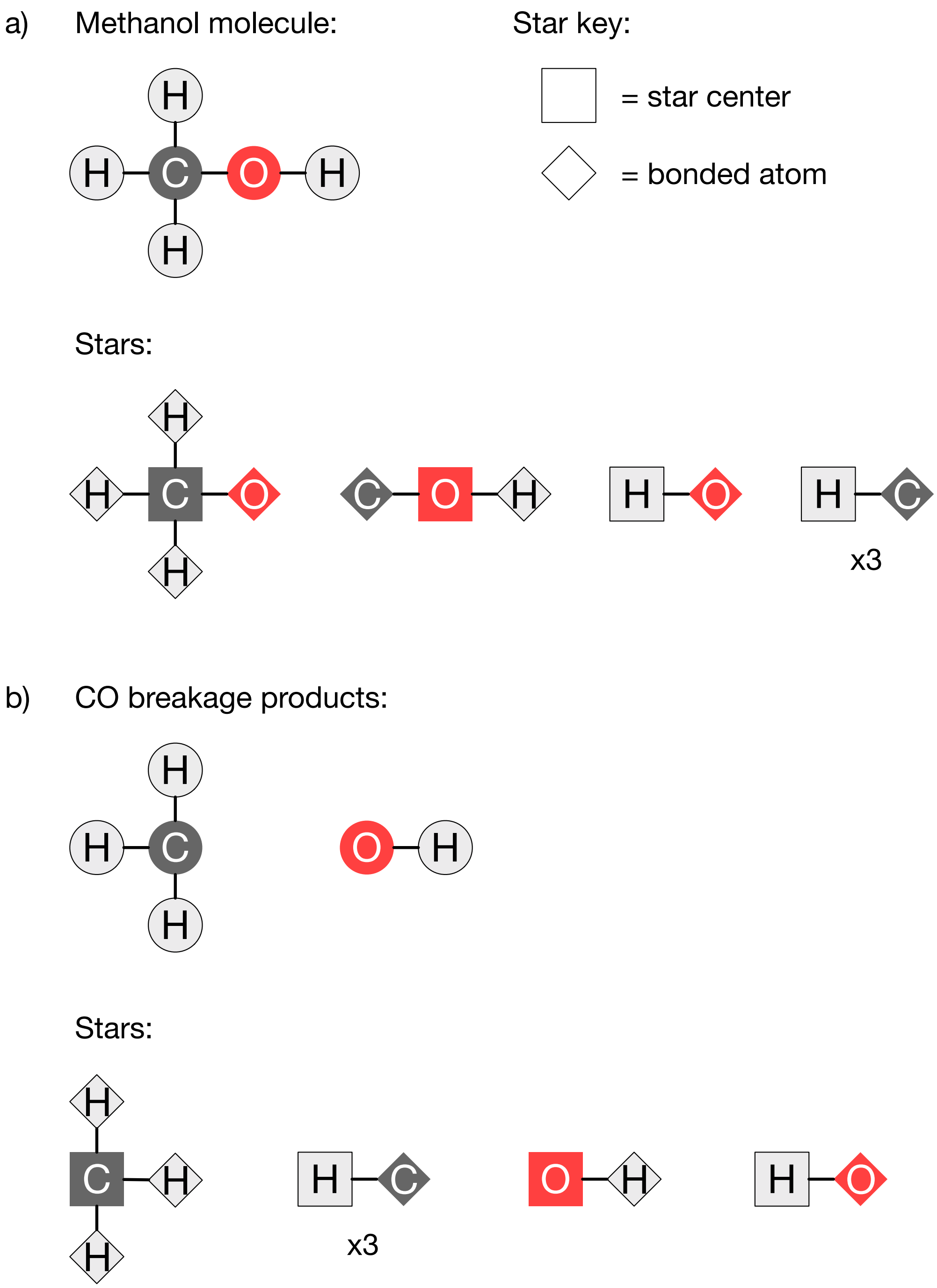


Figure 5: **Star examples.** **a)** The stars of methanol. **b)** The stars of two product molecules following the breakage of the CO bond in methanol.

We pre-compute the stars of each molecule during species filtering and are able to rapidly compute the number of stars that are different between reactant(s) and product(s) in a given reaction during reaction filtering. Our goal is to remove reactions involving the simultaneous breakage/formation of more than two bonds in total or which are break-1-form-1 but do not have a reaction center involved in

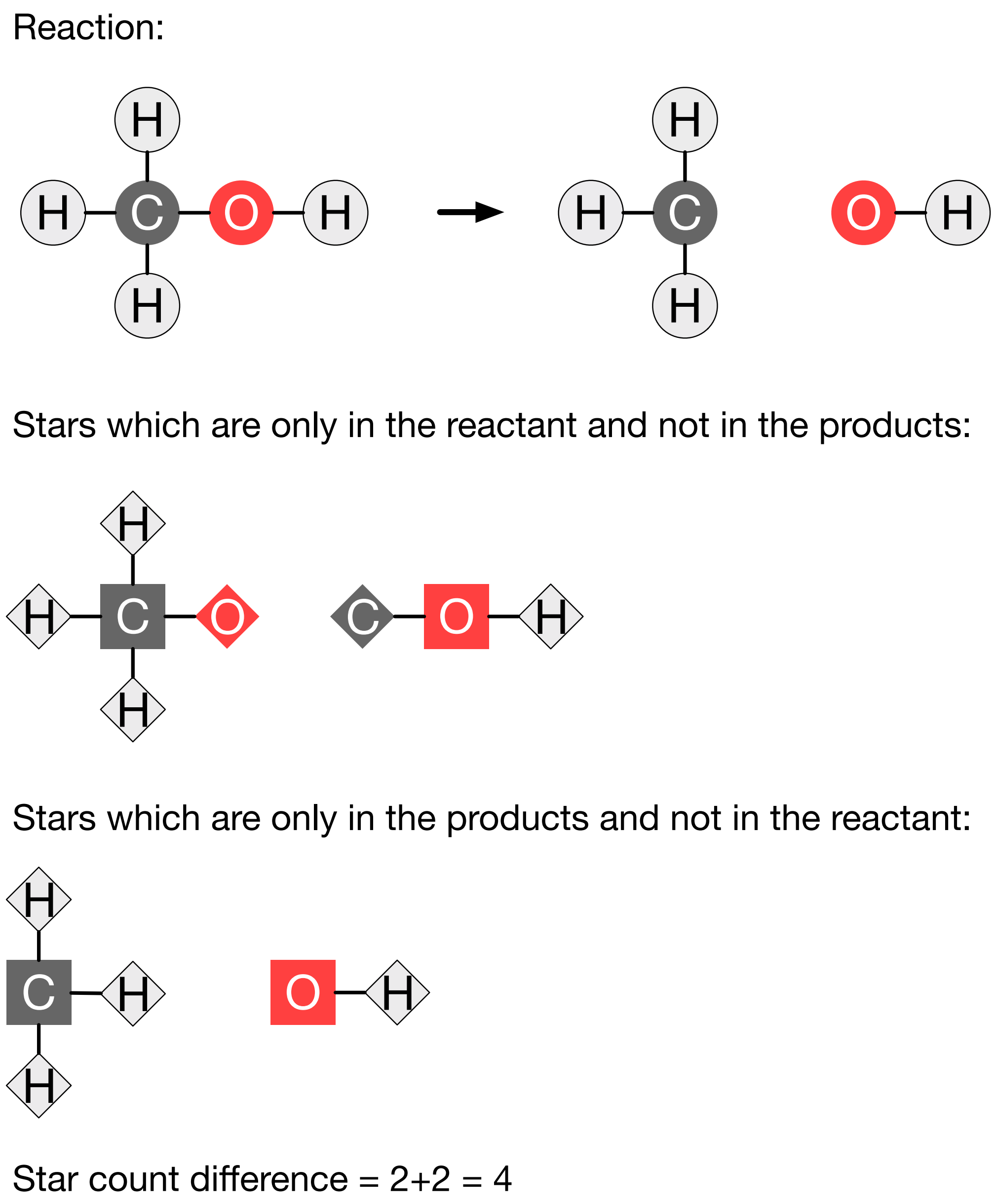


Figure 6: **The star count difference of the reaction breaking the CO bond in methanol.**

both the breakage and formation, as such reactions are unlikely to occur in a single concerted step. This is based on the concept that reactions involving fewer bonds changing should be more likely to occur in a single step and more energetically favorable, in general, than those involving many bonds changing.[25] We note that there are well known examples of mechanisms which are outside of our stated scope, including E2 reactions (most often break-2-form-1) and hydrolysis reactions (up to break-2-form-2). However, these reactions are notable precisely because of their uniqueness, reinforcing the idea that break-2-form-1 and break-2-form-2 mechanisms should not be allowed in general as the vast majority will not be single-step reactions. We continue to consider possibilities for how to include such unique mechanisms in our approach in the future.

We remove all reactions where the stars contained in the reactants and products differ by more than a user-defined threshold. Throughout this work, we use a threshold of six. In order to explain our cutoff of six, we will use a simplified model system of elements [A, B, C, D] which can each form two bonds. Some example reactions and star count differences:

- $AB \rightarrow A + B$, star count difference of 4 which is accepted. This is a break-1 reaction like the example above.
- $AB + C \rightarrow A + BC$, star count difference of 6 which is accepted. This is a break-1-form-1 reaction where B is the reaction center as it is participating in both the bond being broken and the bond being formed.
- $ABC + D \rightarrow AB + CD$, star count difference of 6 which is accepted. This is also a break-1-form-1 reaction, and C is the reaction center.
- $ABC + D \rightarrow A + BCD$, star count difference of 8 which is not accepted. This is a break-1-form-1 reaction without a reaction center. A and B are participating in the bond breakage, but C and D are participating in the bond formation.
- $ABC + D \rightarrow CABD$, star count difference of 8 which is not accepted. This is a break-1-form-2 reaction, where the BC bond is breaking, the CA bond is forming, and the BD bond is forming.

For molecules composed entirely of uniquely distinguishable atoms, a star count difference cutoff of six is 100% effective at enforcing the removal of only reactions without a reaction center or those involving the breakage/formation of more than two covalent bonds. However, once a molecule contains multiple indistinguishable atoms, this cutoff becomes only partially effective. In the examples below, I bold atoms whose stars cannot be matched u to an equivalent star on the other side of the reaction. Therefore, the star count difference is the total number of bolded atoms:

- $ABC\mathbf{AB} + \mathbf{A} \rightarrow A\mathbf{BA}BC\mathbf{A}$, star count difference of 6 which is accepted. This is a break-1-form-2 reaction, where the CA bond is breaking, the BA bond is forming, and the CA bond is forming.
- $AB\mathbf{AAC}+AB \rightarrow ABA+\mathbf{ACA}B$, star count difference of 6 which is accepted. This is a break-1-form-1 reaction without a reaction center, where the AA bond is breaking and the CA bond is forming.

Therefore, while the star filter is useful in that it is extremely performant and removes all reactions which are unambiguously violating our stated goal, a later filter that is more computationally intensive - the fragment filter - will be necessary to remove most of the remaining ambiguous but undesirable reactions.

(b) **Covalently decomposable**: Remove reactions of the form $A + B \rightarrow A + C$ as the reaction is covalently decomposable and the presence of A has no impact on the ΔG. We note that in this type of scenario A could be acting catalytically, which could non-trivially impact the reaction kinetics, but that does not need to be considered until we are attempting to kinetically refine a given reaction or pathway.

(c) **Concerted metal coordination**: Remove reactions in which a bare metal ion is coordinated or uncoordinated simultaneous with covalent bond breakage and/or formation since these processes occur on different timescales and are very unlikely to happen simultaneously.

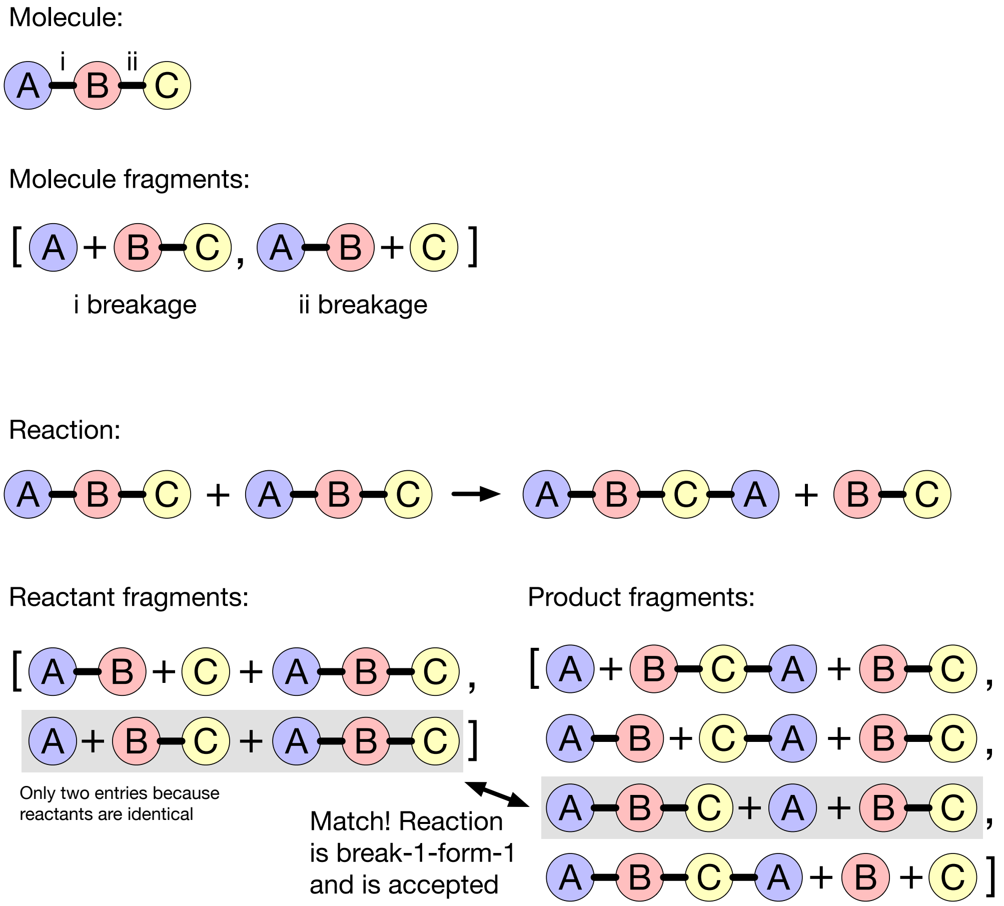


Figure 7: **Generating fragments from the breakage of one bond for a molecule and then comparing reactant and product fragments for a reaction.**

(d) **Fragment filter**: Remove reactions without a viable fragment matching. As stated previously, our goal at present is to remove reactions in which more than two covalent bonds are breaking/forming simultaneously or which do not have a reaction center involved in both the breakage and formation. We achieve this by precomputing and saving the fragments corresponding to the breaking of each individual bond in a molecule and then systematically comparing those fragments between reactant(s) and product(s) during reaction filtering.

Figure 7 shows a graphical example of this process. The ABC molecule has two sets of fragments: A+BC, from breaking the A-B bond (labeled i), and AB+C, from breaking the B-C bond (labeled ii). In the context of a reaction R1 + R2 → P1 + P2, we obtain reactant fragment entries by taking the precomputed fragments from R1 and adding them to the full molecule R2 and vice versa. Thus each reactant fragment set defines the fragments that would be obtained from breaking one bond. We do the same for the products. In this case, there are only two reactant fragment sets because the two reactants are identical, but there are four product fragment entries corresponding to the four distinct bonds that can be broken. We then compare each reactant fragment set to each product fragment set. If there is a match, that means that one bond can be broken in the reactants and one bond can be broken in the products and the identical fragments are obtained. In other words, the reaction is a break-1-form-1 reaction. A similar but simplified procedure is used to identify break-1 and form-1 reactions as well.

As currently implemented, our procedure enforces that only break-1, form-1, or break-1-form-1 non-redox reactions are allowed. In the near future it will also be possible to include break-2 reactions or form-2 reactions, such as Diels-Alder reactions.

The fragment filter is both the most important and by far the most computationally intensive filter. Even with all of the filters beforehand, it could still be performed hundreds of millions of times or more, so it must be extremely fast. Comparing fragment sets involves many operations checking whether two graphs are isomorphic. Standard graph isomorphism operations are fairly slow and must be accelerated for our purposes.

We achieve this using a graph hashing function. Graph hashing functions take graphs and return strings (with a fixed length) such that isomorphic graphs are assigned identical strings. Like all hash functions, graph hash functions have collisions, but for a well defined hashing functions, such collisions will be extremely rare. In HiPRGen, we use the Weisfeiler-Lehman graph hash[26], which is implemented in networkx[27]. During species filtering, we perform fragmentation and graph hashing pre-computation and store the hashes so that they can be referred back to during reaction filtering, as described above.

We restate the procedure for fragment filtering here in terms of hashes in order to be precise: for each molecule, for each bond, we remove the bond and hash the resulting connected components. If the bond was part of a ring, there may be only one component. We call the list of hashes associated with the removal of one bond a fragment hash list. We also store the hash of the whole molecule which we call the total hash. Then, in the fragment filter, a reaction is break-1-form-1 if there is a way of choosing one bond in the reactants and one bond in the products such that the fragment hash list + total hash is the same on the left and the right (up to permutation). Searching for such a hash matching is quadratic in the expected number of bonds. While over 99% of the reactions being filtered are $A + B \to C + D$, for reactions with only one reactant or only one product, a similar procedure is used to also identify and accept break-1 and form-1 reactions.

Returning to the final two model examples discussed in the context of the star count filter, the $ABCAB + A \to ABABCA$ reaction will be successfully removed by the fragment filter because there is no way to either just break one bond on the right or break one bond on the left and one bond on the right and obtain matching sets of fragments. In contrast, the $ABAAC + AB \rightarrow ABAACAB$ reaction will not be filtered out as breaking the AA bond on the left and the CA bond on the right does yield a matching set of fragments, consistent with the fact that the reaction is break-1-form-1 despite not having a reaction center. Thus, while this filter does effectively remove reactions involving more than two bonds simultaneously breaking and/or forming, it does not enforce that a reaction center must be present. Molecular symmetry has thus far prevented us from tractably filtering out all reactions without a reaction center which make it past the star count filter, but none have yet been observed during pathfinding analysis.

(e) $A \rightarrow B$ **non-hydrogen transfer**: Remove break-1-form-1 $A \rightarrow B$ reactions involving the movement of a non-hydrogen fragment within a molecule. While single step intramolecular hydrogen transfer is not uncommon, the concerted movement of e.g. a CO_2 fragment from one end of a molecule to another seemed extremely unlikely in general, for instance in the reaction shown below in Figure 8 which frequently showed up in paths to LEDC.

Given that all evidence from TS searching indicated that this reaction and similar reactions do not occur in a single step, we constructed this filter to remove such mechanisms in general.

(f) $A \rightarrow B + C$ **ring closing**: Remove break-1-form-1 $A \rightarrow B + C$ reactions in which the number of reactants and products necessitates that the bond being formed is a ring-closing. Ring closing reactions in general are known to be highly entropically unfavorable. For a ring closing to occur simultaneous with another bong breaking should thus be extremely unlikely, such as in the reaction shown below in Figure 9 in which LEDC is able to decompose into lithium carbonate while reforming the EC ring.

As with the previous filter, evidence from TS searching indicated that this reaction and similar reactions cannot occur in a single step, motivating the construction of this filter to remove such mechanisms in general.

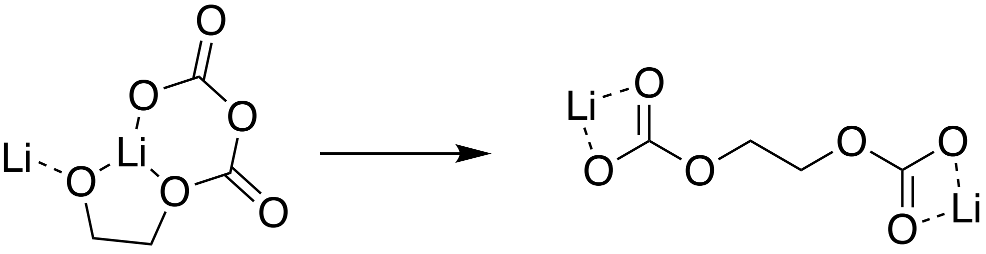


Figure 8: **Unphysical single step formation of LEDC that motivated the non-hydrogen transfer filter.**

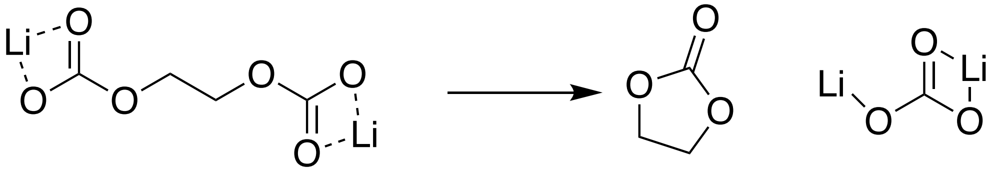


Figure 9: **Unphysical single step decomposition of LEDC that motivated the ring closing filter.**

Network Products

Average Monte Carlo Trajectories

If we plot a single trajectory of a reaction network simulation, it is a step function. On the other hand, the expected values of the species counts satisfy a rate equation, so the average trajectory should be smooth as demonstrated in Figures 10, 11, 12, 13. Moreover, for the purpose of identifying proposed products, we need to sample until the average trajectory smooths out and no longer looks like a step function, which is indicative of sampling convergence.

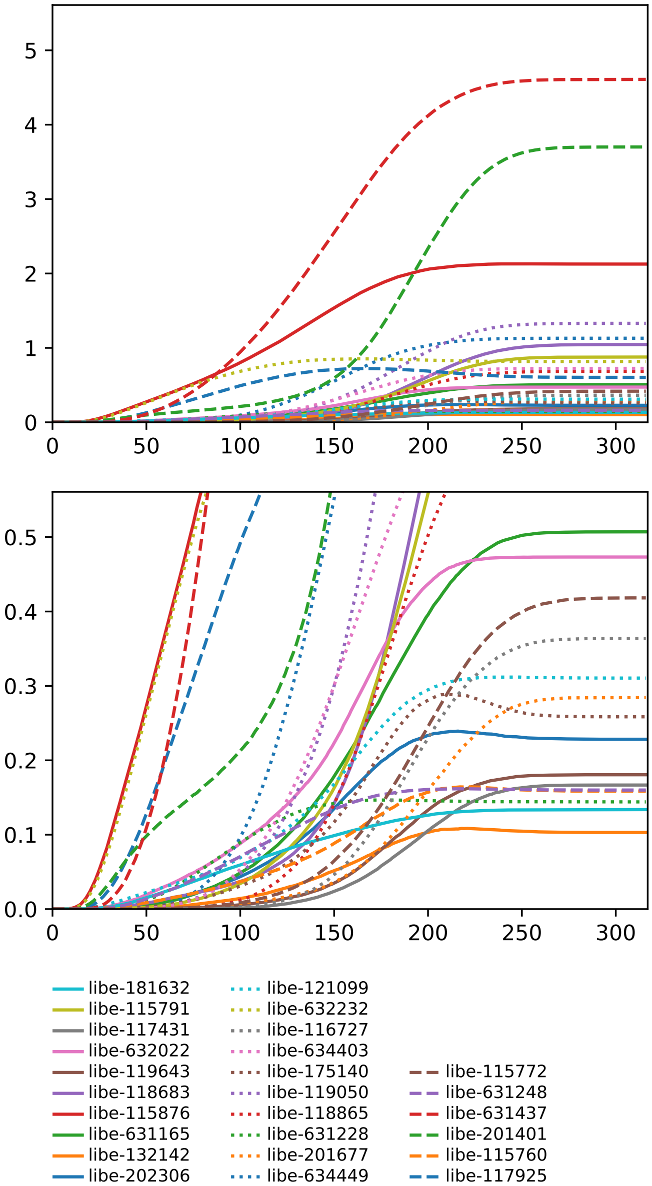


Figure 10: Average trajectory from 100,000 stochastic trajectories of our network with electron free energy −1.4 and initial condition 30 Li^+ ions and 30 EC. Only network products are shown.

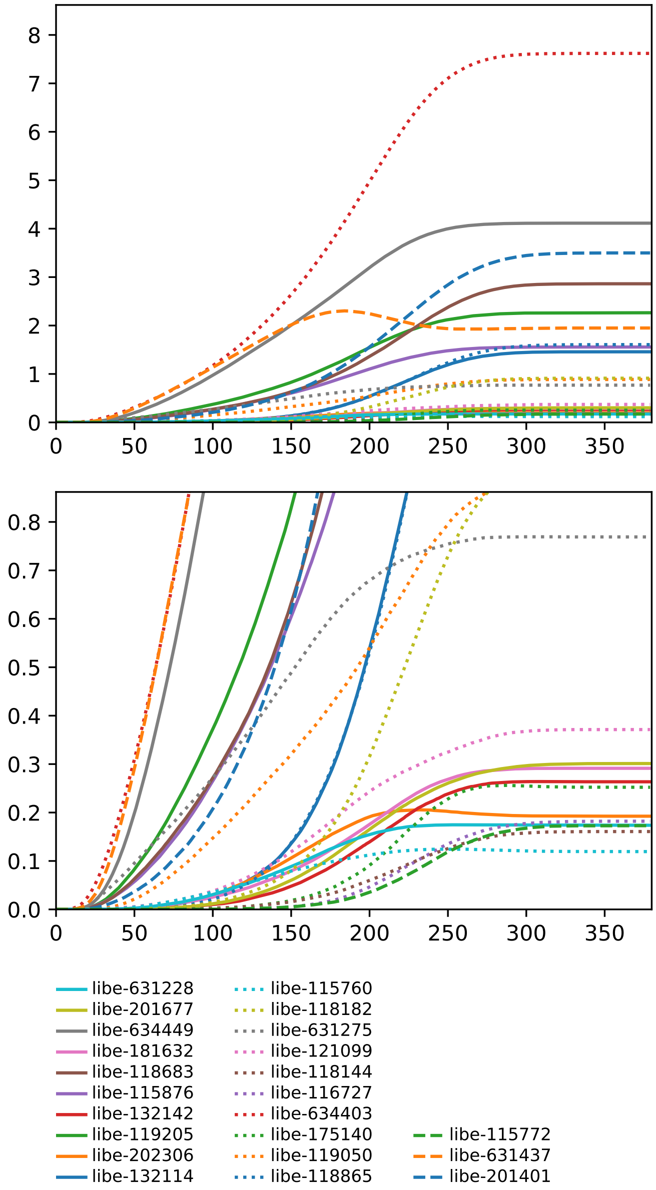


Figure 11: Average trajectory from 100,000 stochastic trajectories of our network with electron free energy -1.4 and initial condition 30 Li^+ ions, 30 EC and 30 CO_2. Only network products are shown.

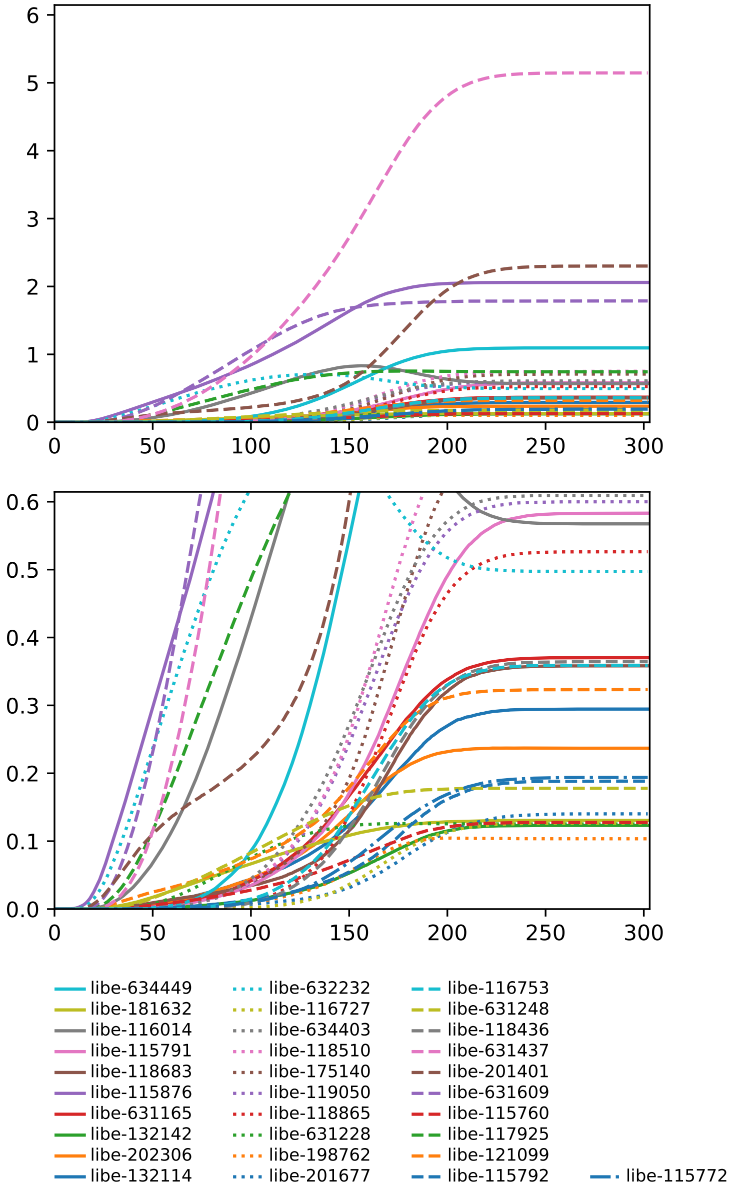


Figure 12: Average trajectory from 100,000 stochastic trajectories of our network with electron free energy -1.9 and initial condition 30 Li^+ ions and 30 EC. Only network products are shown.

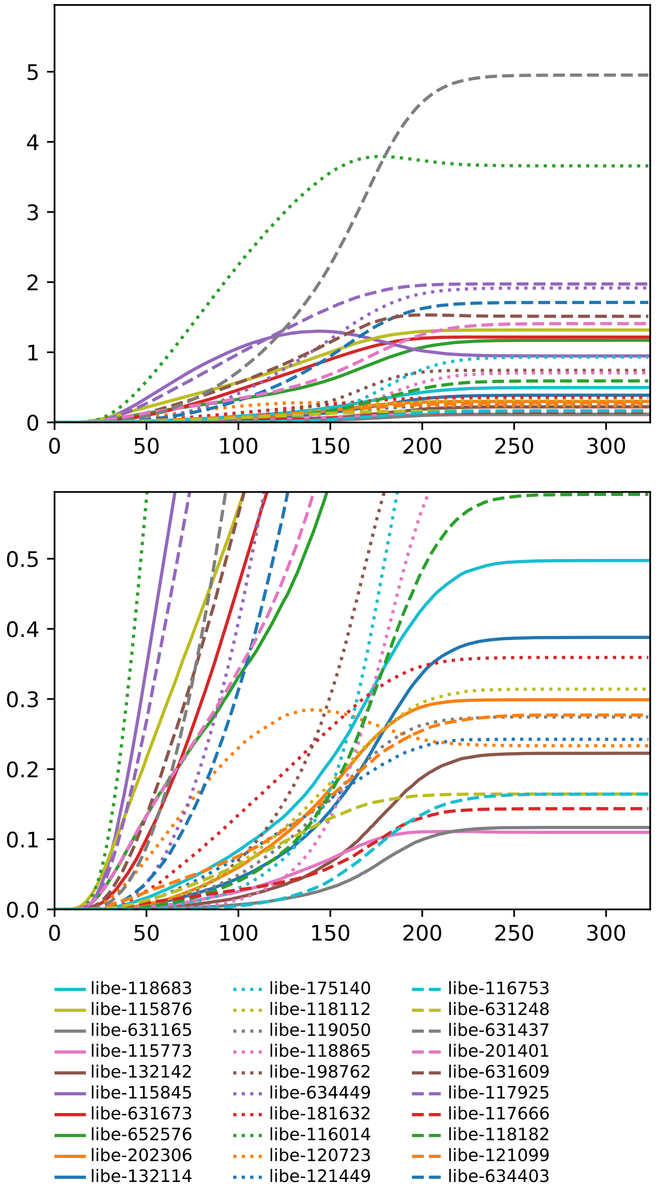
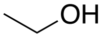
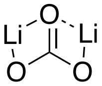
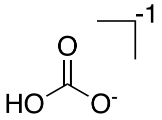
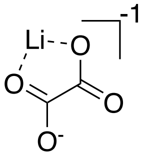
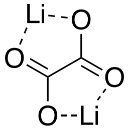
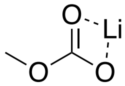
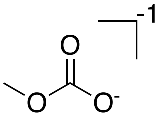
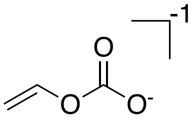
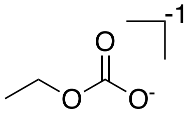
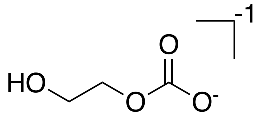
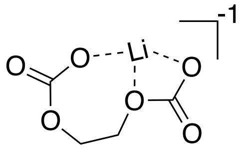
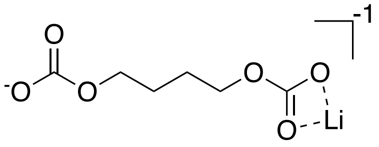
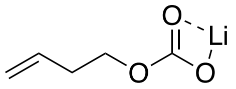
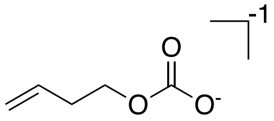
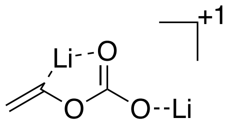
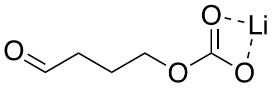
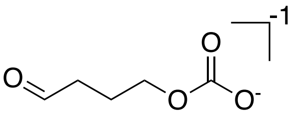
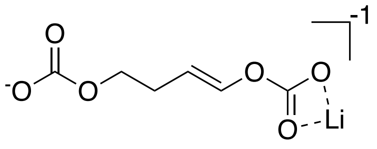
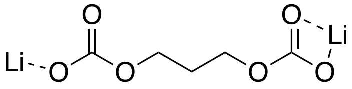
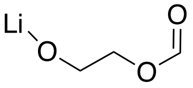
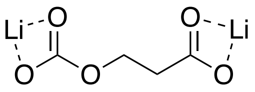


Figure 13: Average trajectory from 100,000 stochastic trajectories of our network with electron free energy −1.9 and initial condition 30 Li^+ ions, 30 EC and 30 CO_2. Only network products are shown.

Dependence of Network Products on Conditions

The network products shown in Figure 3 of the main text emerge from four different sets of thermodynamically bounded kMC trajectories with different initial conditions. While some species in Figure 3 of the main text are products under all four initial conditions (such as Li_2CO_3), most are not. Here, we show how network products vary with initial conditions. We also show the values for the heuristic criteria used to decide if a species is a network product (expected value, formation/consumption ratio, and minimum path cost).

Molecule	**LIBE ID**	**Product (+0.0V, w/o CO_2)**	**Product (+0.0V, w/ CO_2)**	**Product (+0.5V, w/o CO_2)**	**Product (+0.5V, w/ CO_2)**
H_2	libe-115760	YES	YES	YES	NO
	libe-115876	YES	YES	YES	YES
CO	libe-116014	NO	NO	YES	YES
	libe-115773	NO	NO	NO	YES
	libe-631437	YES	YES	YES	YES
	libe-118683	YES	YES	YES	YES
	libe-634403	NO	YES	NO	YES
	libe-634449	NO	YES	NO	YES
	libe-202306	YES	YES	YES	YES
	libe-201677	YES	YES	YES	NO
	libe-119205	NO	YES	NO	NO
	libe-201401	YES	YES	YES	YES

	libe-117666	NO	NO	NO	YES
	libe-652576	NO	NO	NO	YES
	libe-121099	YES	YES	YES	YES
	libe-118112	NO	NO	NO	YES
	libe-118182	NO	YES	NO	NO
	libe-120723	NO	NO	NO	YES
	libe-121449	NO	NO	NO	YES
	libe-132114	NO	YES	YES	YES
	libe-132142	NO	NO	YES	NO
	libe-631248	YES	NO	YES	YES
	libe-631609	NO	NO	YES	YES
	libe-631673	NO	NO	NO	YES

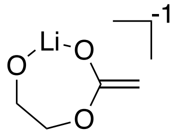
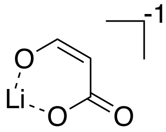
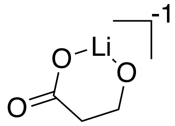
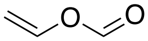
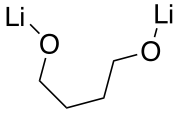
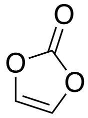
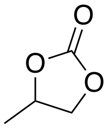
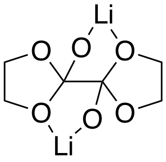
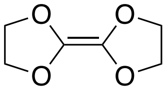
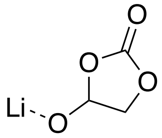
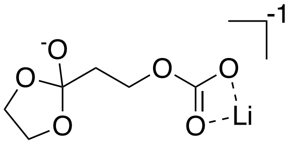
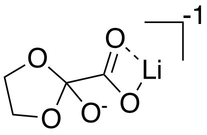
	libe-631165	YES	NO	YES	YES
	libe-118865	NO	YES	NO	NO
	libe-119050	YES	YES	YES	YES
	libe-198762	NO	NO	NO	YES
	libe-115791	YES	NO	YES	NO
	libe-115845	NO	NO	NO	YES
	libe-181632	YES	YES	YES	YES
	libe-632232	YES	NO	YES	NO
	libe-631228	YES	YES	NO	NO
	libe-117925	YES	NO	YES	YES
	libe-631275	NO	YES	NO	NO
	libe-632022	YES	NO	NO	NO

Table 3: **Products of the CRN** under four initial conditions (+0.0V vs. Li/Li$^+$ with Li$^+$ and EC as starting species; +0.0V vs. Li/Li$^+$ with Li$^+$, EC, and CO$_2$ 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 CO$_2$ as starting species). Molecule IDs for the LIBE dataset[23] are provided.

H_2 **libe-115760**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	2.0809299788641167	0.15854	7.169848111578256	Yes
0.0V, w/ CO_2	2.5723883546304833	0.11936	7.169848111578256	Yes
+0.5V, w/o CO_2	4.145571181840612	0.12749	9.173815815086435	Yes
+0.5V, w/ CO_2	7.5390625	0.06696	9.173815815086435	No

Table 4: **Heuristic parameters for libe-115760 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

 **libe-115876**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.795232040686586	2.12653	4.5849654419948065	Yes
0.0V, w/ CO_2	2.1908419723805515	1.55649	4.5849654419948065	Yes
+0.5V, w/o CO_2	1.8479559366548248	2.05988	4.587995692565772	Yes
+0.5V, w/ CO_2	1.9779571607695332	1.31812	4.587995692565772	Yes

Table 5: **Heuristic parameters for libe-115876 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

CO **libe-116014**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.2742171875238546	0.35923	5.609549851288864	No
0.0V, w/ CO_2	1.031377139455868	0.23669	4.071475466582235	No
+0.5V, w/o CO_2	1.557506731392858	0.56733	5.613513125184876	Yes
+0.5V, w/ CO_2	2.655159462785983	3.65777	5.58797868113481	Yes

Table 6: **Heuristic parameters for libe-116014 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

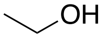
 **libe-115773**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.0400195626353665	0.02864	9.170500067647772	No
0.0V, w/ CO_2	1.0398831449256953	0.00942	8.684811785955393	No
+0.5V, w/o CO_2	1.1943197467494142	0.10865	9.173956350524378	No
+0.5V, w/ CO_2	1.8518949081608929	0.10992	9.173956350524378	Yes

Table 7: **Heuristic parameters for libe-115773 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

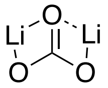
 **libe-631437**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	4.47576813430472	4.60866	6.584910828905317	Yes
0.0V, w/ CO_2	1.6128225928161788	1.94907	5.071475466582235	Yes
+0.5V, w/o CO_2	5.229718287860977	5.14541	6.587941083019205	Yes
+0.5V, w/ CO_2	5.094708902835314	4.9521	6.587941083019205	Yes

Table 8: **Heuristic parameters for libe-631437 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

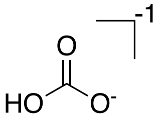
 **libe-118683**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	2.1531263104460483	1.04494	6.634026862214116	Yes
0.0V, w/ CO_2	2.0980338748968985	2.86219	5.58491088771418	Yes
+0.5V, w/o CO_2	1.6125055568854085	0.35823	6.637057116328004	Yes
+0.5V, w/ CO_2	1.7691772455274999	0.49745	6.637057116328004	Yes

Table 9: **Heuristic parameters for libe-118683 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

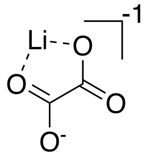
 **libe-634403**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	7.0852368155861605	0.72463	10.170755269329723	No
0.0V, w/ CO_2	9.617551916029498	7.61895	4.071475466582235	Yes
+0.5V, w/o CO_2	6.57054036755966	0.60925	10.176815770471654	No
+0.5V, w/ CO_2	5.256276280759374	1.71064	6.588013231328022	Yes

Table 10: **Heuristic parameters for libe-634403 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

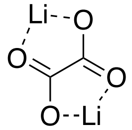
 **libe-634449**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	30.599161645271156	1.1298	10.169823371709874	No
0.0V, w/ CO_2	15.950174443959877	4.11369	5.071475466582239	Yes
+0.5V, w/o CO_2	41.29102281089036	1.09511	10.374445183151913	No
+0.5V, w/ CO_2	39.37077077077077	1.91662	7.588013231328026	Yes

Table 11: **Heuristic parameters for libe-634449 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

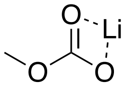
 **libe-202306**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.8668641300220248	0.22828	6.819415641753605	Yes
0.0V, w/ CO_2	2.1151416816364375	0.19244	6.819415641753605	Yes
+0.5V, w/o CO_2	2.8269853508095606	0.23696	6.822445892324571	Yes
+0.5V, w/ CO_2	5.455507527202266	0.29892	6.822445892324571	Yes

Table 12: **Heuristic parameters for libe-202306 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

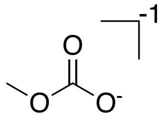
 **libe-201677**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	3.0916513424052963	0.28436	7.769132827847294	Yes
0.0V, w/ CO_2	3.220125322521194	0.30116	7.769132827847294	Yes
+0.5V, w/o CO_2	3.9557428872497367	0.14025	7.82244820817826	Yes
+0.5V, w/ CO_2	3.346879150066401	0.08836	7.772163081961182	No

Table 13: **Heuristic parameters for libe-201677 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

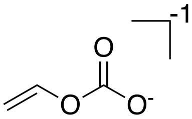
 **libe-119205**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.3962930954659094	0.51358	5.584911160548076	No
0.0V, w/ CO_2	1.9572016843671785	2.26405	5.07147549707084	Yes
+0.5V, w/o CO_2	1.4076885600661675	0.35983	5.587941414661964	No
+0.5V, w/ CO_2	1.3823773842422113	0.56593	5.587941083019205	No

Table 14: **Heuristic parameters for libe-119205 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

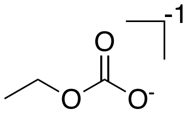
 **libe-201401**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	3.6691983670172097	3.70063	5.584910857879308	Yes
0.0V, w/ CO_2	4.117233085385951	3.49925	5.584910857879308	Yes
+0.5V, w/o CO_2	3.4562465308910806	2.30111	5.587941111993196	Yes
+0.5V, w/ CO_2	2.716957210776545	1.40842	5.587941111993196	Yes

Table 15: **Heuristic parameters for libe-201401 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

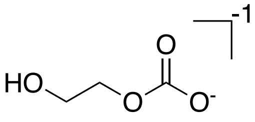
 **libe-117666**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.0226210196944598	0.01229	8.184610595183646	No
0.0V, w/ CO_2	1.1091227944597546	0.06098	8.173139943265971	No
+0.5V, w/o CO_2	1.0339305689110427	0.01729	8.190671096325577	No
+0.5V, w/ CO_2	1.7399793920659454	0.14363	8.179200444407902	Yes

Table 16: **Heuristic parameters for libe-117666 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

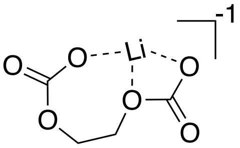
 **libe-652576**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.046725193485041	0.05663	6.584984771540545	No
0.0V, w/ CO_2	1.3665448282541595	1.74089	5.071475466665193	No
+0.5V, w/o CO_2	1.0363261249070757	0.04349	6.588015025654433	No
+0.5V, w/ CO_2	1.6211635048342359	1.17121	6.000223350559122	Yes

Table 17: **Heuristic parameters for libe-652576 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

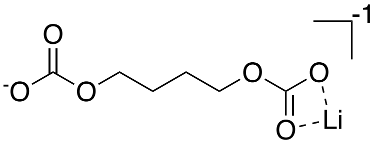
 **libe-121099**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.9487594720117332	0.31051	5.584910828905317	Yes
0.0V, w/ CO_2	1.7338102484042448	0.37133	5.584910828905317	Yes
+0.5V, w/o CO_2	2.1176348547717843	0.32322	5.587941083019205	Yes
+0.5V, w/ CO_2	1.6751693222238464	0.27713	5.587941083019205	Yes

Table 18: **Heuristic parameters for libe-121099 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

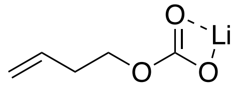
 **libe-118112**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.0143047431516765	0.01482	7.634026862214116	No
0.0V, w/ CO_2	1.1522555424103698	0.0592	7.634026862214116	No
+0.5V, w/o CO_2	1.1735181528790748	0.21531	7.637057116328004	No
+0.5V, w/ CO_2	1.5731518148803767	0.31407	7.637057116328004	Yes

Table 19: **Heuristic parameters for libe-118112 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

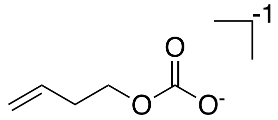
 **libe-118182**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.0980854172472911	0.11614	10.169821657810868	No
0.0V, w/ CO_2	3.050596776731327	0.91229	9.585844477705095	Yes
+0.5V, w/o CO_2	1.3746357017374615	0.36893	10.175882159260238	No
+0.5V, w/ CO_2	3.0693356643356644	0.59183	10.175882159260004	No

Table 20: **Heuristic parameters for libe-118182 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

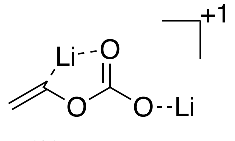
 **libe-120723**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.0	0.0	6.584911417099472	No
0.0V, w/ CO_2	1.0	0.0	6.584910862700847	No
+0.5V, w/o CO_2	1.2234203991538675	0.19434	7.172851911915762	No
+0.5V, w/ CO_2	1.5425628386076686	0.23334	7.172851911915762	Yes

Table 21: **Heuristic parameters for libe-120723 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

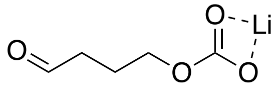
 **libe-121449**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.0037097542713225	0.00218	7.585844440440368	No
0.0V, w/ CO_2	1.0927785621054313	0.03483	7.584910862700735	No
+0.5V, w/o CO_2	1.066271305934558	0.04794	7.588874691011334	No
+0.5V, w/ CO_2	1.5977767918761707	0.24253	7.588874691011334	Yes

Table 22: **Heuristic parameters for libe-121449 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

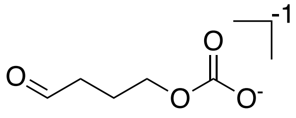
 **libe-132114**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.1801598015250465	0.14814	7.585844440424406	No
0.0V, w/ CO_2	3.7444040738718725	1.4578	7.584910862700734	Yes
+0.5V, w/o CO_2	1.5058207417582417	0.29459	7.588874690995372	Yes
+0.5V, w/ CO_2	3.416812063808574	0.38785	7.588874690995372	Yes

Table 23: **Heuristic parameters for libe-132114 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

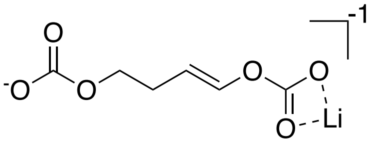
 **libe-132142**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.5646142880640386	0.10298	11.169822246007575	No
0.0V, w/ CO_2	1.9197849762636134	0.2635	10.169821928614454	No
+0.5V, w/o CO_2	2.1702612826603325	0.12317	9.607573190745873	Yes
+0.5V, w/ CO_2	2.5124320652173915	0.22263	11.181362366315046	No

Table 24: **Heuristic parameters for libe-132142 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

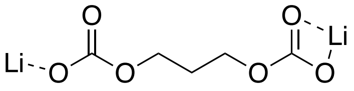
 **libe-631248**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	5.9218701937865275	0.16001	9.169821657810662	Yes
0.0V, w/ CO_2	6.175968109339408	0.09089	8.582103937325135	No
+0.5V, w/o CO_2	11.978408389882789	0.17796	9.175882158952593	Yes
+0.5V, w/ CO_2	11.883597883597883	0.16456	8.588164438467066	Yes

Table 25: **Heuristic parameters for libe-631248 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

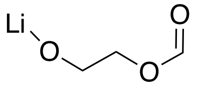
 **libe-631609**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.202614180618625	0.73398	5.584911417100697	No
0.0V, w/ CO_2	1.0245670349790543	0.0424	5.584911417100697	No
+0.5V, w/o CO_2	1.7022313202076123	1.78727	6.587941431513009	Yes
+0.5V, w/ CO_2	2.2154438231135107	1.51279	6.587941431513009	Yes

Table 26: **Heuristic parameters for libe-631609 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

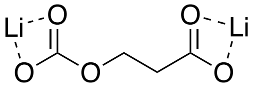
 **libe-631673**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.2236191124859286	0.47278	9.169821658029703	No
0.0V, w/ CO_2	1.3143242429220743	0.45919	5.071475466582285	No
+0.5V, w/o CO_2	1.3463893456653628	0.6172	9.175882162714556	No
+0.5V, w/ CO_2	2.119064930283926	1.21592	6.587941083019269	Yes

Table 27: **Heuristic parameters for libe-631673 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

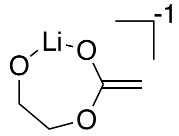
 **libe-631165**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	2.2732705586942874	0.50708	9.169822246004685	Yes
0.0V, w/ CO_2	1.175621107384847	0.02566	9.169822246004685	No
+0.5V, w/o CO_2	2.9998379517096097	0.37023	9.17681577047882	Yes
+0.5V, w/ CO_2	3.3279537756525204	0.11684	9.17681577047882	Yes

Table 28: **Heuristic parameters for libe-631165 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

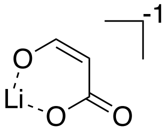
 **libe-118865**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	12.824991370383154	0.68514	11.17600096390089	No
0.0V, w/ CO_2	4.6917737375594735	1.60618	8.58498396760627	Yes
+0.5V, w/o CO_2	12.922745808790213	0.52627	12.397133972210726	No
+0.5V, w/ CO_2	4.580597771023303	0.70681	11.182060540203874	No

Table 29: **Heuristic parameters for libe-118865 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

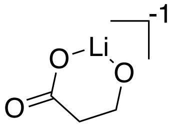
 **libe-119050**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	3.296284145457057	1.32987	8.170755269329968	Yes
0.0V, w/ CO_2	1.6715770967839714	0.88624	6.5849132421220276	Yes
+0.5V, w/o CO_2	2.573782103412996	0.59991	8.1768157704719	Yes
+0.5V, w/ CO_2	1.8055898068228524	0.2744	6.588874690995372	Yes

Table 30: **Heuristic parameters for libe-119050 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

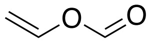
 **libe-198762**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.149322096005863	0.0163	11.939360254747937	No
0.0V, w/ CO_2	1.2649337568761136	0.06839	7.584963941145458	No
+0.5V, w/o CO_2	1.8487332950725588	0.10352	12.176868886254718	No
+0.5V, w/ CO_2	3.9591991385155345	0.74196	7.587994195259346	Yes

Table 31: **Heuristic parameters for libe-198762 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

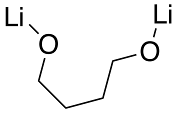
 **libe-115791**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	2.3610417022333277	0.87696	9.171688880848812	Yes
0.0V, w/ CO_2	1.0433257055682685	0.00284	9.171688880848812	No
+0.5V, w/o CO_2	2.0824577687024317	0.58312	9.177749381990743	Yes
+0.5V, w/ CO_2	1.270214414879876	0.01046	9.177749381990743	No

Table 32: **Heuristic parameters for libe-115791 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

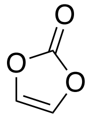
 **libe-115845**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.0510584764879403	0.02354	7.169822246359074	No
0.0V, w/ CO_2	1.087357189049364	0.19452	5.584910962612366	No
+0.5V, w/o CO_2	1.2013348401816748	0.09442	8.174229986186393	No
+0.5V, w/ CO_2	1.6457096912667535	0.94765	6.587941186891381	Yes

Table 33: **Heuristic parameters for libe-115845 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

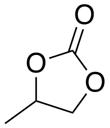
 **libe-181632**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	28.234215885947048	0.13372	5.819414839753749	Yes
0.0V, w/ CO_2	57.245173745173744	0.29135	5.819414839753749	Yes
+0.5V, w/o CO_2	43.41883116883117	0.13065	5.822445090324715	Yes
+0.5V, w/ CO_2	100.24309392265194	0.35926	5.822445090324715	Yes

Table 34: **Heuristic parameters for libe-181632 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

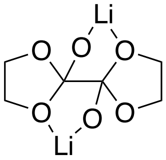
 **libe-632232**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	4.237353931471579	0.81727	6.169821657810633	Yes
0.0V, w/ CO_2	1.2596496633070728	0.11799	6.169821657810633	No
+0.5V, w/o CO_2	1.9693986940843973	0.49735	6.1758821589525645	Yes
+0.5V, w/ CO_2	1.0	0.0	6.1758821589525645	No

Table 35: **Heuristic parameters for libe-632232 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

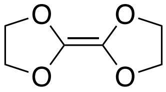
 **libe-631228**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	7.386081560283688	0.14407	9.169843331665913	Yes
0.0V, w/ CO_2	6.699803149606299	0.17373	8.657060757907614	Yes
+0.5V, w/o CO_2	9.473292765382014	0.12532	10.175902664048358	No
+0.5V, w/ CO_2	7.908629441624366	0.02722	10.175902664048358	No

Table 36: **Heuristic parameters for libe-631228 under various initial network conditions.** “Ratio” refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). “EV” refers to the expected value of the species (the final quantity present in the average trajectory). “Min. cost” is the cost of the lowest-cost pathway to form the specie.

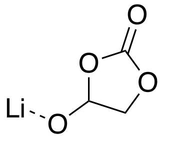
 **libe-117925**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	2.0027242977458095	0.60363	6.62011589587385	Yes
0.0V, w/ CO_2	1.4289213829330505	0.67936	5.5849108877143525	No
+0.5V, w/o CO_2	2.734500466853408	0.74306	6.623146146444816	Yes
+0.5V, w/ CO_2	3.149370093315476	1.97396	6.587941111993368	Yes

Table 37: **Heuristic parameters for libe-117925 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

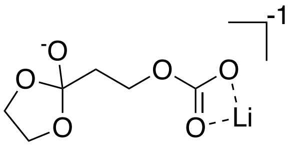
 **libe-631275**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.2583843340417187	0.28329	4.5849114170993595	No
0.0V, w/ CO_2	1.7037464082431963	0.76904	4.5849114170993595	Yes
+0.5V, w/o CO_2	1.2616093983303127	0.27858	4.6075702730672266	No
+0.5V, w/ CO_2	1.4028539871879833	0.45593	4.6075702730672266	No

Table 38: **Heuristic parameters for libe-631275 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

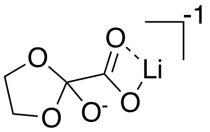
 **libe-632022**	**Ratio**	**EV**	**Min. Cost**	**Product?**
0.0V, w/o CO_2	1.6269558420222843	0.47322	6.584912059145902	Yes
0.0V, w/ CO_2	1.0457656764312964	0.09969	4.0714754665822355	No
+0.5V, w/o CO_2	1.3911786183730102	0.26784	6.588013873374227	No
+0.5V, w/ CO_2	1.0	0.0	6.588013873374227	No

Table 39: **Heuristic parameters for libe-632022 under various initial network conditions.** "Ratio" refers to the ratio of formation (how many times is a species formed across all trajectories) to consumption (how many times is a species consumed across all trajectories). "EV" refers to the expected value of the species (the final quantity present in the average trajectory). "Min. cost" is the cost of the lowest-cost pathway to form the specie.

Rendering Large Reaction Networks

In order to comprehend the scale of the computations performed in this paper, it is useful to try and visualize the complexity. Reaction networks are not efficiently modeled as graphs, but for the purposes of visualization, it is useful to map them onto graphs. We do so as follows: The nodes are the species and a reaction $A + B \to C + D$ gives rise to four edges, $(A, C), (A, D), (B, C), (B, D)$. Our renders are for the network with electron free energy -1.4 eV and starting species Li+, EC and CO_2. Recall that the full network has 5,193 species and 86,001,275 reactions. The starting species are shown as larger dots on the left. In Figure 15 we only render the 312,034 reactions which fired during 100,000 thermodynamically bounded Monte Carlo trajectories. In Figure 18, we render the 290 reactions in the top 10 pathways to the 18 network products identified from this specific initial condition. In Figure 19, we highlight the 3rd shortest pathway to bi-dioxolylidine.

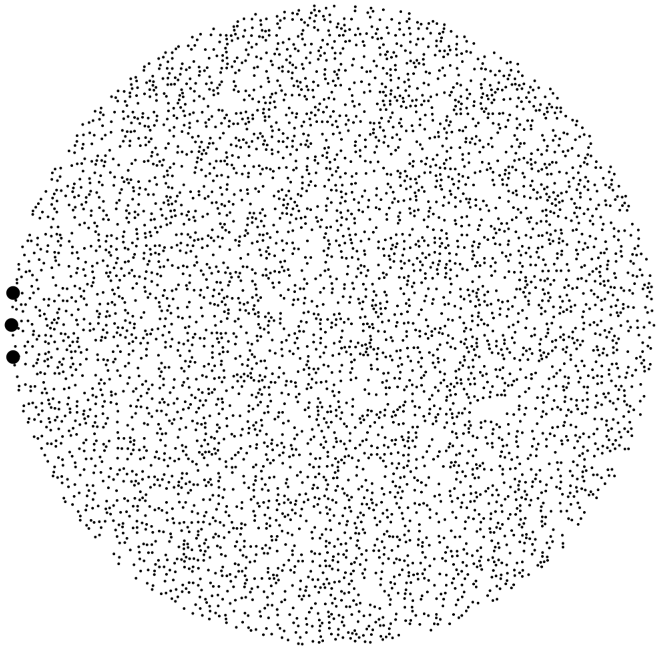


Figure 14: Plot of all 5193 species nodes, with initial species Li+, EC and CO_2 shown as larger dots on the left.

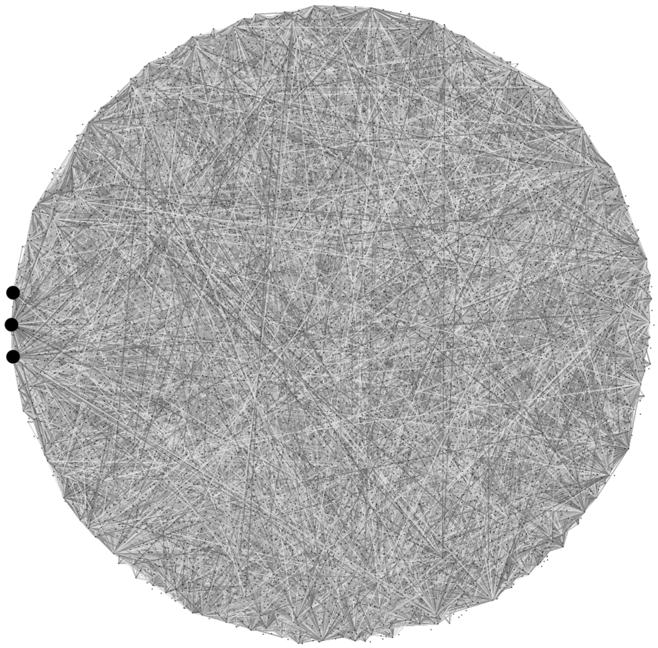


Figure 15: Plot of all 312,034 unique reactions which fired in 100,000 thermodynamically bounded Monte Carlo trajectories with electron free energy −1.4 eV and starting species Li+, EC and CO_2.

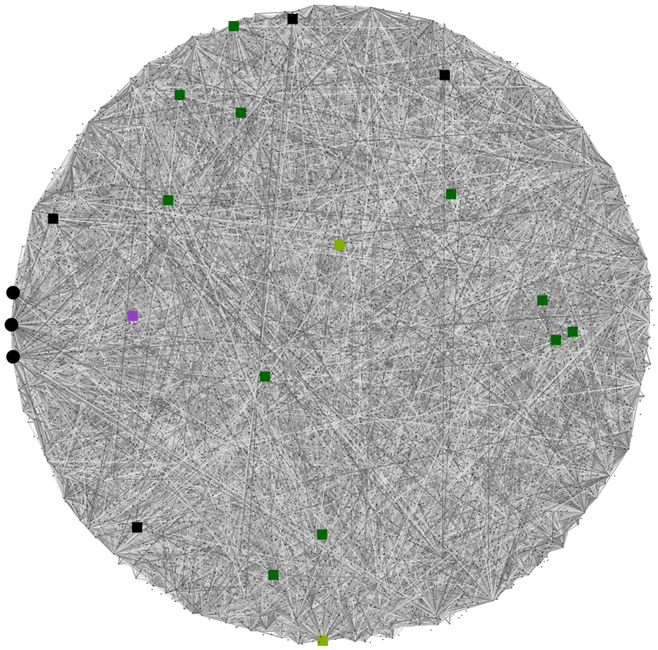


Figure 16: Plot of all unique reactions which fired under previously noted initial condition as well as the 18 identified network products indicated.

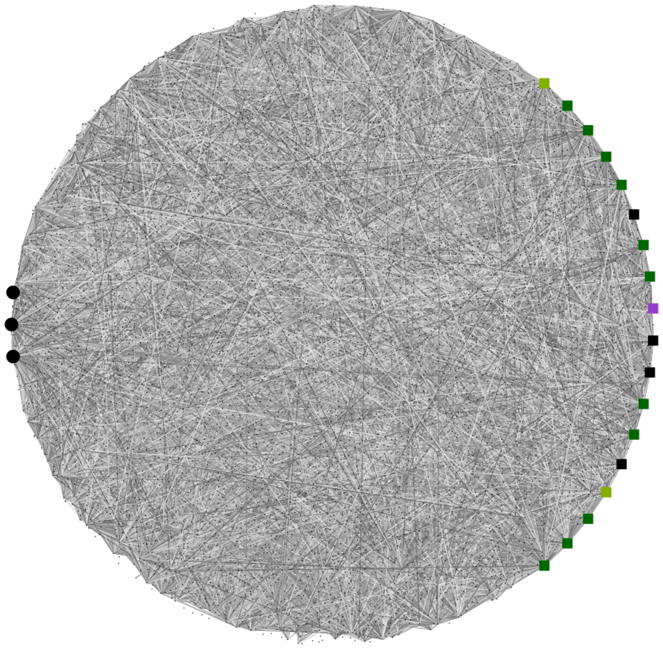


Figure 17: Moving network products to the right for graphical clean-up.

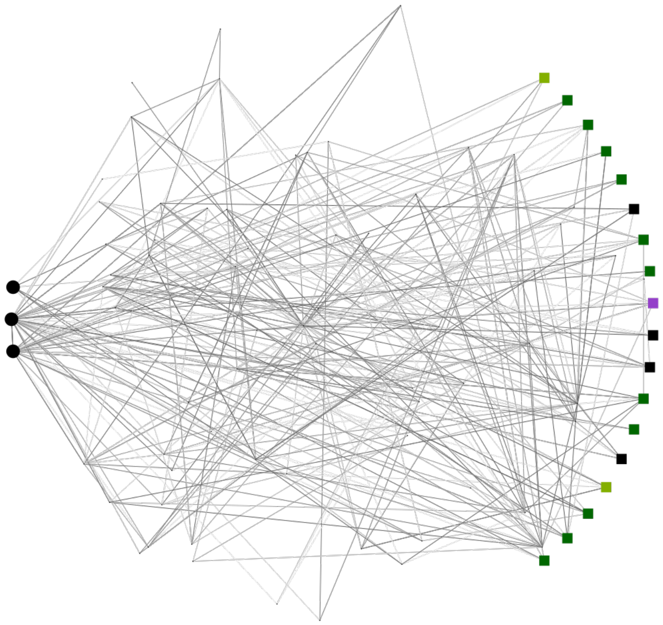


Figure 18: Plot of reactions in top 10 pathways to 18 network products under previously noted initial condition.

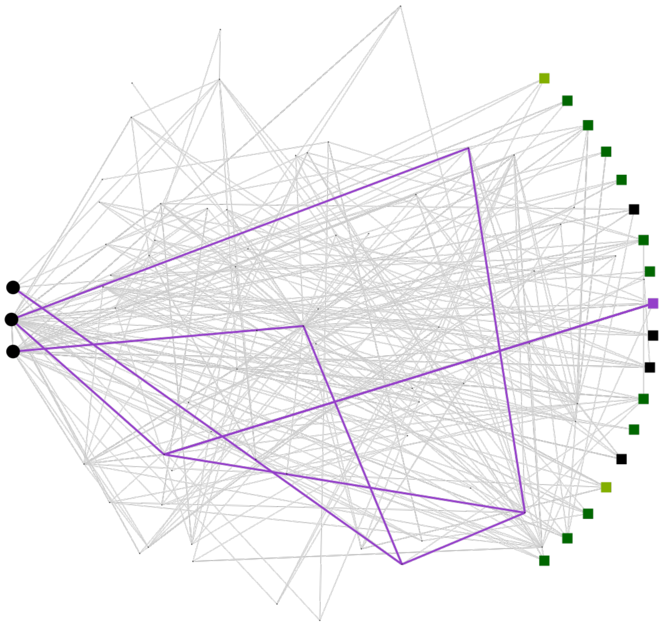


Figure 19: Third shortest pathway to bi-dioxolylidine.

References

[1] 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.

[2] J. C. Baez, *Advances in Mathematical Physics*, 2018, **2018**, e7676309.

[3] 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.

[4] 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.

[5] A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.

[6] D. Bedrov, G. D. Smith and A. C. T. van Duin, *J. Phys. Chem. A*, 2012, **116**, 2978–2985.

[7] O. Borodin, G. V. Zhuang, P. N. Ross and K. Xu, *J. Phys. Chem. C*, 2013, **117**, 7433–7444.

[8] L. Alzate-Vargas, S. M. Blau, E. W. C. Spotte-Smith, S. Allu, K. A. Persson and J.-L. Fattebert, *J. Phys. Chem. C*, 2021, **125**, 18588–18596.

[9] J. J. Guerard and J. S. Arey, *J. Chem. Theory Comput.*, 2013, **9**, 5046–5058.

[10] Y. Horowitz, H.-L. Han, P. N. Ross and G. A. Somorjai, *J. Am. Chem. Soc.*, 2016, **138**, 726–729.

[11] X. Zhang, R. Kostecki, T. J. Richardson, J. K. Pugh and P. N. Ross, *J. Electrochem. Soc.*, 2001, **148**, A1341.

[12] L. Ma, S. L. Glazier, R. Petibon, J. Xia, J. M. Peters, Q. Liu, J. Allen, R. N. C. Doig and J. R. Dahn, *J. Electrochem. Soc.*, 2016, **164**, A5008.

[13] S. Trasatti, *Pure and Applied Chemistry*, 1986, **58**, 955–966.

[14] J. Xia, R. Petibon, D. Xiong, L. Ma and J. R. Dahn, *Journal of Power Sources*, 2016, **328**, 124–135.

[15] H. Ota, Y. Sakata, A. Inoue and S. Yamaguchi, *J. Electrochem. Soc.*, 2004, **151**, A1659.

[16] A. M. Haregewoin, E. G. Leggesse, J.-C. Jiang, F.-M. Wang, B.-J. Hwang and S. D. Lin, *Electrochimica Acta*, 2014, **136**, 274–285.

[17] M. Ue, M. Takeda, M. Takehara and S. Mori, *J. Electrochem. Soc.*, 1997, **144**, 2684–2688.

[18] K. Xu and C. A. Angell, *J. Electrochem. Soc.*, 1998, **145**, L70.

[19] N. M. O'Boyle, C. Morley and G. R. Hutchison, *Chemistry Central Journal*, 2008, **2**, 5.

[20] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *Journal of Cheminformatics*, 2011, **3**, 33.

[21] A. Otero-de-la Roza, M. A. Blanco, A. M. Pendás and V. Luaña, *Computer Physics Communications*, 2009, **180**, 157–166.

[22] A. Otero-de-la Roza, E. R. Johnson and V. Luaña, *Computer Physics Communications*, 2014, **185**, 1007–1018.

[23] 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.

[24] J. M. Saveant, *Journal of the American Chemical Society*, 1987, **109**, 6788–6795.

[25] C. A. Grambow, L. Pattanaik and W. H. Green, *Sci Data*, 2020, **7**, 137.

[26] N. Shervashidze, P. Schweitzer, E. J. van Leeuwen, K. Mehlhorn and K. M. Borgwardt, *J. Mach. Learn. Res.*, 2011, **12**, 2539–2561.

[27] A. A. Hagberg, D. A. Schult and P. J. Swart, Proceedings of the 7th Python in Science Conference, Pasadena, CA USA, 2008, pp. 11 – 15.