JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

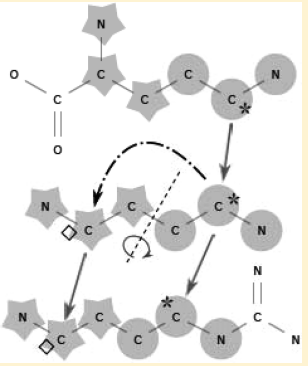
pubs.acs.org/jcim

Quantifying and Assessing the Effect of Chemical Symmetry in Metabolic Pathways

Wanding Zhou*,† and Luay Nakhleh*,‡

†Department of Bioengineering and ‡Department of Computer Science, Rice University, Houston, Texas 77005, United States

ABSTRACT: Atom tracing provides valuable information in many analyses of metabolic networks including pathway inference and flux estimation. Symmetries—mapping operations that produce atom equivalencies—introduce alternative tracings when multiple atom mappings are aggregated. Although several attempts have been made to consider symmetry while curating atom mappings, a definition of the symmetry amenable to automated computation and a systematic quantification of the extent of symmetries in both compounds and reactions is still lacking. Moreover, the impact of symmetries on the calculation of the atom economy of pathways and the simulation of isotopomer distribution is yet to be assessed. In this study, we formulate the symmetries of both compounds and reactions as automorphic mappings of the corresponding graph representations. We investigate the extent of both compound and reaction symmetries in several metabolic systems. We find, through random walking in the metabolic network of *E.coli*, that alternative tracings originated from symmetries could give rise to considerable amount of differential conservation of atoms and distinct transition patterns of the isotopomer distribution.



INTRODUCTION

The study of metabolism at the atomic level has made significant progress in the last few decades.[1−4] With detailed information of atom transition for each metabolic reaction, the fate of atoms can be traced in a pathway composed of concatenated reactions.[5,6] Besides the study of the circulation of the mass,[5,7] tracing atoms in a metabolic pathway or network has found two major applications, namely pathway inference and the simulation of isotopomer distribution for the purpose of estimating reaction fluxes. Known and novel metabolic pathways can be inferred from a metabolic network under the principle of atom economy,[8] which seeks, among all the pathways that connect a source compound to a target compound, the one that maximizes the number of atoms traced.[9−11] Simulating isotopomer distribution is a task of predicting for a compound the relative abundance of its isotope-labeled forms, called isotopomers, from (1) the isotopomer distribution of a given source compound; (2) the reaction fluxes of the metabolic network; and (3) the atom mapping relationship as yielded by each reaction mechanism.[5] Experimentally, information on the steady-state isotopomer distribution, such as the mass distribution from mass spectrometry[12] or the distribution of atoms in certain microenvironments as can be detected by nuclear magnetic resonance,[13,14] can be used to derive the steady-state flux distribution. Isotopomer distribution is represented in silico as an isotopomer distribution vector (IDV) which contains the relative abundance of each isotopomer whose location in the IDV can be coded into a binary vector with each bit projected to an atom in the compound with its value indicating whether the atom is isotopically labeled.[15] The transition from the IDV of one compound to the IDV of another compound can be summarized in a matrix called the isotopomer mapping matrix (IMM) whose rows are IDVs of the source compound and columns are IDVs of the target compound.[15] Computing an IMM reduces to tracing atoms from the source compound to the target compound.

Computational tracing of atoms in a reaction requires knowledge of the full atom mapping of each reaction on the pathway. Compounds in a reaction atom mapping are usually represented with atoms labeled to distinguish different atom instances. However, there are cases where atoms or compound instances are chemically indistinguishable. For example, in Figure 1A when the σ-bond linking the central atom and the star rotates, atoms 1, 2, and 3 are replaced by atoms 2, 3, and 1, respectively, without changing the nature of the compound. Linking, for tracing purposes, multiple reactions via compounds that contain indistinguishable atoms might bring about alternative tracings (see the Methods section). This equivalence in atoms or compound instances is generally referred to as symmetry.

Three-dimensional (3D) molecular symmetry has been well studied for more than half a century using the theory of point groups.[16] The main purpose of studying molecular symmetry is to elucidate chemical properties of a molecule such as selection rules in vibration spectroscopy.[17] However, there are several challenges facing a large-scale analysis that applies this theory. First, the identification of the 3D symmetry operations is hard to automate. The complete and accurate curation of symmetry groups and their resolution into subgroups requires ad hoc knowledge from the molecular point group theory. In fact, most symmetry studies have been conducted on small compounds.[18] Second, most reaction atom mapping data available, including manual curations such as the RPAIR data[19] and automated methods such as ones based on the maximum common subgraph (MCS) heuristics[20,21] and the ones based on the minimum

Received: June 7, 2012
Published: September 17, 2012

   dx.doi.org/10.1021/ci300259u | *J. Chem. Inf. Model.* 2012, 52, 2684−2696

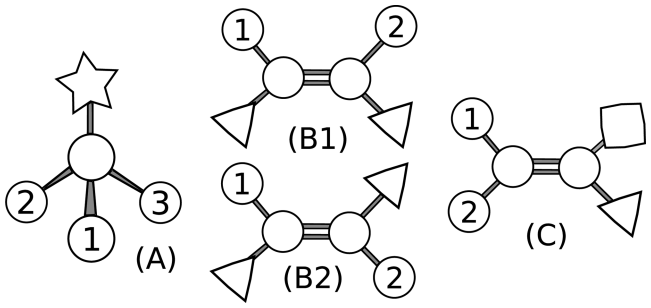


Figure 1. Comparison between 3D symmetry operations and automorphisms of the graph representation. (A) Under graph representation, every permutation of the three labeled atoms is a valid symmetry mapping. $\{1 \rightarrow 2;2 \rightarrow 3;3 \rightarrow 1\}$ is a rotational symmetry operation and $\{1 \rightarrow 1;2 \rightarrow 3;3 \rightarrow 2\}$ is a mirror symmetry opertation. The mirror symmetry operation is not physically feasible. (B) Swapping 1 and 2 is a mirror symmetry in B1 and rotational symmetry in B2. However, under graph representation, B1 and B2 are indistinguishable. (C) Atoms 1 and 2 cannot be swapped. However, under graph representation, swapping 1 and 2 is a false positive symmetry.

graph edits,[22,23] fail to consider 3D molecular symmetries. It is impossible to generate correct alternative atom mappings if the reaction atom mappings are not produced aware of the 3D configuration.

In this work, we take a graph-theoretic approach that represents both compounds and reactions as *chemical graphs*, which are attributed relational graphs,[24] and formulates the problem of the symmetry as an automorphism problem on the chemical graphs to facilitate automated computing. Although this approach fails to capture all the symmetry operations of compounds in 3D (see Figure 1 case C), it captures most symmetry mappings as well as the group structure (see Figure 1 cases A and B). Its calculation can be automated and the resulting symmetries integrate well with reaction atom mapping data that do not consider 3D features such as prochirality[25] and cis−trans isomerization. We developed automated methods to identify symmetries of both compounds and reactions. Symmetry-breaking atom mappings are defined and demonstrated to impact atom tracing. We find that 257 out of 1251 reactions in *E.coli* and 176 atom mappings from the KEGG RPAIR database[19] are symmetry-breaking with respect to a selection of symmetry subgroups. We devise a decomposition scheme for efficient storage of symmetry in the database. Alternative atom transitions in KEGG are explicitly listed when symmetric compounds are present. Finally, we evaluate, using metabolic pathways in the model organism *E.coli*, the impact of alternative atom tracings on pathway inference, in terms of atom economy, and isotope labeling analysis, in terms of differential IMM calculation. We regenerate atom mapping data between reactants by first augmenting the RPAIR data into full reaction atom mappings and then composing the reaction atom mappings with symmmetries on both sides of a reaction to get all the alternative atom mappings in cases where symmetries are broken by the reaction. Compared to a previous approach that considers symmetry and curates atom mapping in one organism and one specific model,[1] our work can be readily extended to many organisms.

METHODS

Formulation of the Symmetry of Chemical Graphs. In chemical informatics, a chemical compound can be represented as an attributed relational graph G,[24] whose nodes $V(G)$ correspond to atoms and edges $E(G)$ correspond to chemical bonds. Each node $v \in V(G)$ refers to an atom, and each edge $(u,v) \in E(G)$ refers to a chemical bond linking the atoms to which nodes u and v refer. Nodes in $V(G)$ inherit attributes such as elements and isotopic status from the atoms they refer to. Likewise, bonds in $E(G)$ inherit attributes such as bond types (single, double, aromatic, etc.). Multiple compounds, such as all compounds in a reaction, can be represented as the individual connected components in one graph. Together with graphs for a single compound, they are collectively referred to as *chemical graphs* in this study. For example, all the compounds that participate in a reaction can be represented in a chemical graph. So can all compounds produced from a reaction. In this study, we call the two graphs *reactant graph* and *product graph*, respectively. Atoms in a chemical graph are labeled to distinguish their identities in the atom mapping.

An *atom mapping* from a chemical graph G to a chemical graph H is a bijection $g:V(G) \rightarrow V(H)$ such that if $g(u) = v$, then u and v refer to the same phyiscal atom identity. When defined for a reaction atom mapping, $g:V(G_r) \rightarrow V(G_p)$, where G_r and G_p are the reactant graph and product graph of the reaction. A node $u \in V(G_r)$ is mapped to $v \in V(G_p)$ if u,v refer to the same atom before and after the reaction process, assuming the chemical reaction does not incur atom-level modifications such as nuclear fission/fusion, which is true for most biochemical reactions. We denote the domain and image of a function by $\mathrm{Dom}(\cdot)$ and $\mathrm{Img}(\cdot)$.

A symmetry of a chemical graph is defined as an automorphism f of G (that is, a bijection $f:V(G) \rightarrow V(G)$), that preserves the chemical properties which may include, but are not limited to the following: (1) The atom's element. For all $p \in V(G)$, p and $f(p)$ have the same element. (2) The atom's incident bonds. For every pair of atoms p and q, there is a bond between $f(p)$ and $f(q)$ if and only if there is a bond between p and q with the same bond type. The bond types we consider here includes single, double, triple, and aromatic. A nontrivial symmetry is a symmetry mapping that is not the identity mapping (where each atom maps to itself). A compound with at least one nontrivial symmetry is said to be symmetric. For simplicity, we do not consider 3D chemical features such as prochirality[25] in this study. We target this as future directions and discuss the issue in the Discussion and Conclusion section.

Chemical graphs are graphs of bounded valence, which means that the degrees of the nodes are bounded by a small constant (the number of electrons available for forming covalent bonds). The more general isomorphism problem, the problem of finding a bijection with aforementioned properties but between two arbitrary graphs and of which automorphism is a special case, can be tested in polynomial time on graphs of bounded valence.[26] Due to the constraints imposed by the nodes' and edges' attributes, the problem is in practice not hard to solve. We devise the subroutine $\mathrm{GraphIso}(G_1,G_2)$ for this problem, which is a straightforward adaptation from the VF2 algorithm[27] by further testing the nodes' element and edges' bond types in the feasibility function. In other words, GRAPHISO takes two chemical graphs G_1 and G_2 and returns all the isomorphisms preserving the chemical properties between the two. A compound's symmetries are found by invoking $\mathrm{GRAPHISO}(G,G)$ where G is the compound's graph representation. The method takes seconds to return all the isomorphisms for all the compounds in KEGG,[28] each of which may contain up to hundreds of atoms.

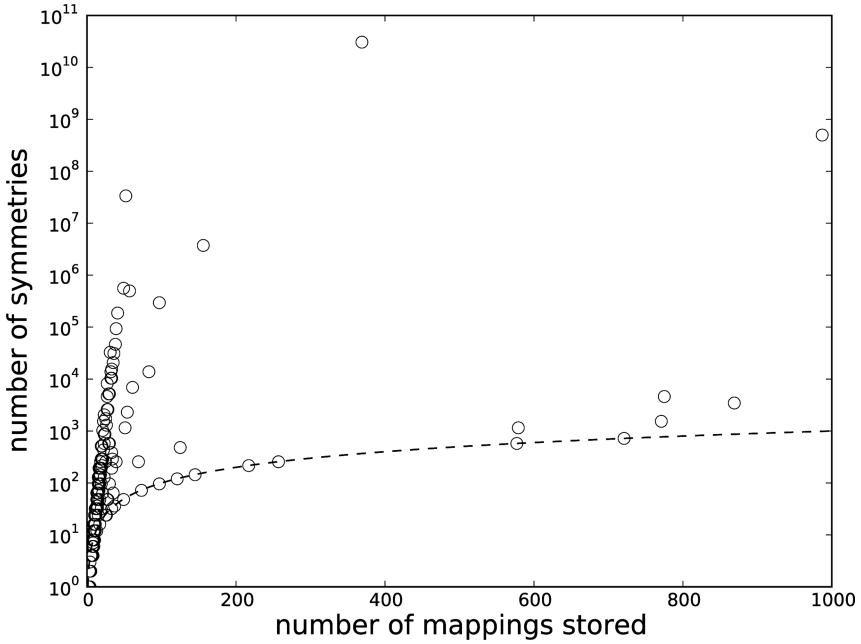
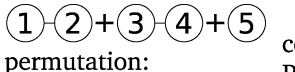


Figure 2. Reduction of the number of explicit mappings stored after decomposition. All the compounds in KEGG ligand database are plotted. The x-axis is the number of symmetry mappings explicitly stored in the database. The y-axis is the number of actual symmetries (on the log scale). The dashed line is the $y = x$ line.

Symmetry Decomposition and Reaction Symmetries. All the symmetry mappings of a compound form a symmetry group [A group is a set together with a defined group operation satisfying the group properties including the existence of identity, closure, associativity, etc.[29]] (in the algebraic sense). The group operation is function composition, and the identity is the mapping that projects every atom to itself. Enumerating and storing all the symmetry mappings explicitly can be infeasible for compounds with multiple symmetry subgroups. The number of mappings increases exponentially with the number of subgroups (as can be seen from Lagrange's theorem in group theory which states that the order of the subgroup divides the order of the original group). In this study, we sought a nonexhaustive decomposition of symmetry groups. Certain chemical substructures of putative symmetry subgroups are detected from the chemical graph. They include the following: (1) subgroups of the form XYn where n Y atoms are covalently linked only to X, forming a permutation group; e.g., PO4, PO3, PO2, CO2, SO4, SO3, SO2, NO3, CC2, CC3, CN2, RuN4, and RuN5. Although Y atoms in some of these substructures are not all equivalent in their microenvironment (e.g., the four oxygens in PO4 are not all equivalent and cannot be freely exchanged as they appear in the graph representation), they are treated equivalently in the RPAIR data that we use to expand the reaction atom mapping, with the rationale that these bonds exchange electrons at a rate much faster than what can be captured by a GC/MS analysis (quick equilibration of labeling caused by resonance stabilization[30]). (2) Single-atom ions forming a permutation group, e.g., Na^+, Cl^-. (3) Hydration water. (4) Mirror image in benzene groups and cyclohexane groups (analogous to the σ_h operation in the 3D symmetry theory). (5) Remaining symmetries. The remaining subgroup may be further reduced or may be irreducible. The symmetry mappings from the remaining symmetry subgroup are explicitly computed. As Figure 2 shows, most data points (circles) are above the $y = x$ line which means that the above decomposition greatly reduces the number of explicit mappings one needs to store (x-axis of Figure 2) compared to the actual number of symmetries (y-axis of Figure 2). To detect benzene and cyclohexane substructure, we employ the ring perception algorithm by Balducci and Pearlman.[31]

The symmetry of each reactant graph and product graph of a reaction is obtained by composing the inherent symmetries of their constituent compounds and the permutation of those compounds (see Figure 3 for an illustration). For inherent symmetries, only symmetries composed from subgroups CC2, CC3, benezene and cyclohexane mirror, and remaining symmetries are considered since the other symmetry subgroups do not involve carbon, which is of utmost interest in tasks such as the prediction of the carbon fate and carbon-13 simulation.[5] Such selective assembly of the compound symmetry greatly



permutation:
P_0 {1:1;2:2;3:3;4:4;5:5}
P_1 {1:3;2:4;3:1;4:2;5:5}
Inherent:
I_0 {1:1;2:2;3:3;4:4;5:5}
I_1 {1:1;2:2;3:4;4:3;5:5}
I_2 {1:2;2:1;3:3;4:4;5:5}
I_3 {1:2;2:1;3:4;4:3;5:5}

composition:
$P_0 \circ I_0$ {1:1;2:2;3:3;4:4;5:5}
$P_0 \circ I_1$ {1:1;2:2;3:4;4:3;5:5}
$P_0 \circ I_2$ {1:2;2:1;3:3;4:4;5:5}
$P_0 \circ I_3$ {1:2;2:1;3:4;4:3;5:5}
$P_1 \circ I_0$ {1:3;2:4;3:1;4:2;5:5}
$P_1 \circ I_1$ {1:4;2:3;3:1;4:2;5:5}
$P_1 \circ I_2$ {1:3;2:4;3:2;4:1;5:5}
$P_1 \circ I_3$ {1:4;2:3;3:2;4:1;5:5}

Figure 3. Illustration of calculating reaction symmetries on a graph with three compounds. Two compounds are of the same type and composed of two bonded atoms of the same element. Each circle represents an atom. The *permutation* mapping (labeled as P_i, i being the index) and the *inherent* symmetry mapping (labeled as I_i, i being the index) of each component compounds are listed in curly braces. The enumeration of the final reaction symmetry mappings (the right column) are realized by iteratively composing each permutation mapping with each inherent mapping. Function composition of two functions P_i and I_j is defined as a new function $P_i \circ I_j$: $x \rightarrow I_j(P_i(x))$ for each $x \in \mathrm{Dom}(P_i)$ given $\mathrm{Img}(P_i) = \mathrm{Dom}(I_j)$.

reduces the amount of space needed for the explicit storage of the symmetries of reactant graphs and product graphs of all the metabolic reactions in *E.coli*, facilitating the detection of symmetry breaking reactions (shown below). The only exception is reaction R06447 which contains 1 290 240 symmetry mappings. The number of symmetries of a reactant graph or product graph can be calculated as

$$\prod_i s_i! \times c_i^{s_i}$$

where s_i is the stochiometric coefficient of the ith compound in the chemical graph and c_i is the number of inherent symmetries of the ith compound.

Detecting Symmetry-Breaking Atom Mappings. Consider an atom mapping g from compound x to compound y. Compound x has a symmetry mapping f_x and compound y has a symmetry mapping f_y. In theory, one can generate all the alternative mappings from g through function composition: $f_x \bigcirc g \bigcirc f_y$. However, not all alternative mappings constructed this way can introduce alternative routes in pathway inference. The iterative composition could exponentially increase the number of alternative tracings as more and more mappings are traced (see Figure 4). Only symmetries that are broken by the mapping of each reaction need to be considered. Here we give the formal definition of symmetry breaking in chemical graphs.

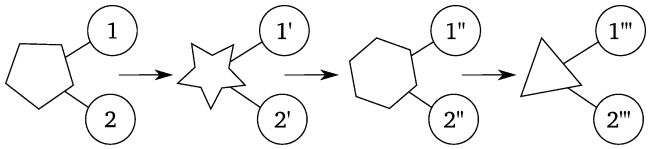


Figure 4. Illustration of Nonsymmetry-Breaking Reactions. The circles represent atoms of the same element. Tracing from the left-most compound to the right-most compound through three atom mappings gives rise to 2^3 alternative tracings (given by $g_1 \bigcirc g_2 \bigcirc g_3$ for $(g_1, g_2, g_3) \in \{\{1{:}2',2{:}1'\},\{1{:}2',2{:}1'\}\} \times \{\{1'{:}2'',2'{:}1''\},\{1'{:}1'',2'{:}2''\}\}$, $\{\{1''{:}2''',2''{:}1'''\} \times \{1''{:}1''',2''{:}2'''\}\}$).

Definition 1 *Consider two chemical graphs G,H and an atom mapping g from V(G) to V(H). A symmetry f on G is broken by g if either (1) Dom(g) ≠ f(Dom(g)); or (2) there is no symmetry of H whose restriction to Img(g) is* $g \bigcirc f \bigcirc g^{-1}$.

If the atom mapping is a complete reaction atom mapping, then $\mathrm{Dom}(g) = V(G) = f(V(G)) = f(\mathrm{Dom}(g))$. The atom mapping breaks the symmetry only by the second condition of Defintion 1. For example, myo-inositol (see Figure 5) has 11 nontrivial symmetries, 5 by rotating the carbon ring, and 6 by reflecting with respect to 6 mirror symmetry axes. Only one mirror symmetry (marked with * in Figure 5) is preserved through its transformation into myo-inositol-3-phosphate. The algorithm used to find symmetry-breaking atom mapping strictly follows Definition 1.

For each atom mapping g in the RPAIR database,[19] we evaluated the symmetry of both of its constituent compounds. Consider a compound x and its symmetry f_x, y is the other compound reached through the atom mapping from x. If $\mathrm{Dom}(g) = f_x(\mathrm{Dom}(g))$ and if there exists a symmetry f_y of compound y such that for each atom $a \in \mathrm{Dom}(g)$, $g(f_x(a)) = f_y(g(a))$ ($g \bigcirc f_x \bigcirc g^{-1}$ can be embedded into f_y), the atom mapping g is said to preserve the symmetry f_x. Otherwise, g breaks f_x.

Breaking the compound/reaction symmetry in a reaction has an implication when atoms are traced across multiple reactions. In Figure 6, carbon labeled * in ornithine can reach carbon * in agmatine via carbon * in putrescine according to the annotated atom mapping in RPAIR. Since putrescine is symmetric (as shown in the mirror symmetry axis) and the symmetry is broken when it gets transformed into agmatine, carbon labeled * in ornithine can as well be mapped into carbon ◇ in putrescine and

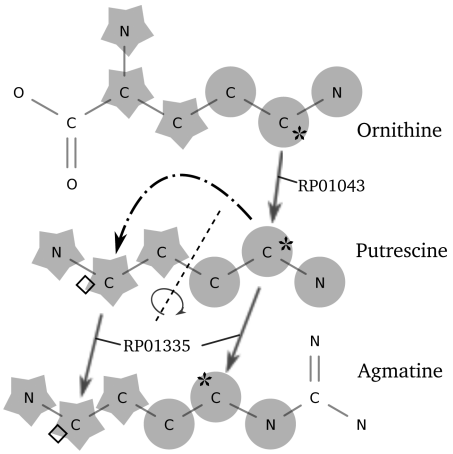


Figure 6. Symmetry breaking which introduces alternative atom fates. The symmetry in putrescine brings its atoms shaded in circles to atoms shaded in stars. Atoms in ornithine and agmatine are shaded in the same way as they are in putrescine. The dashed line corresponds to the symmetry mapping from C4 to C3 in putrescine. Note that these are not complete reaction atom mappings (and therefore not mass-balanced), and irrelevant reactants are omitted from the figure for simplicity.

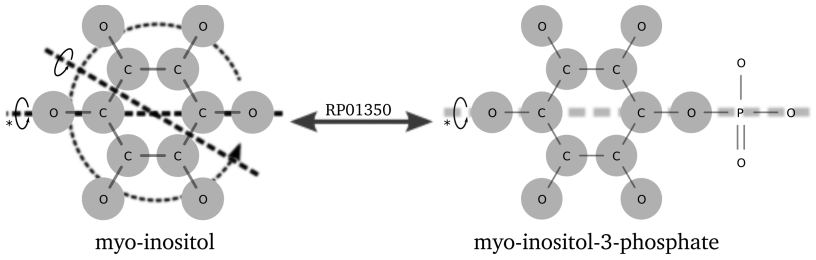


Figure 5. Symmetry breaking between compounds. Each compound is shown with its atoms' elements and labeling. Symmetry axes and operations corresponding to these symmetries are marked in dashed lines. Only two mirror symmetry axes of myo-inositol are shown. Atoms mapped in the two compounds are shaded in circles. RP01350 is the ID corresponding to the atom mapping given in RPAIR database. Atoms mapped are shaded. Note that this is not a complete reaction atom mapping (and therefore not mass-balanced), and irrelevant reactants are omitted from the figure for simplicity.

dx.doi.org/10.1021/ci300259u | J. Chem. Inf. Model. 2012, 52, 2684−2696

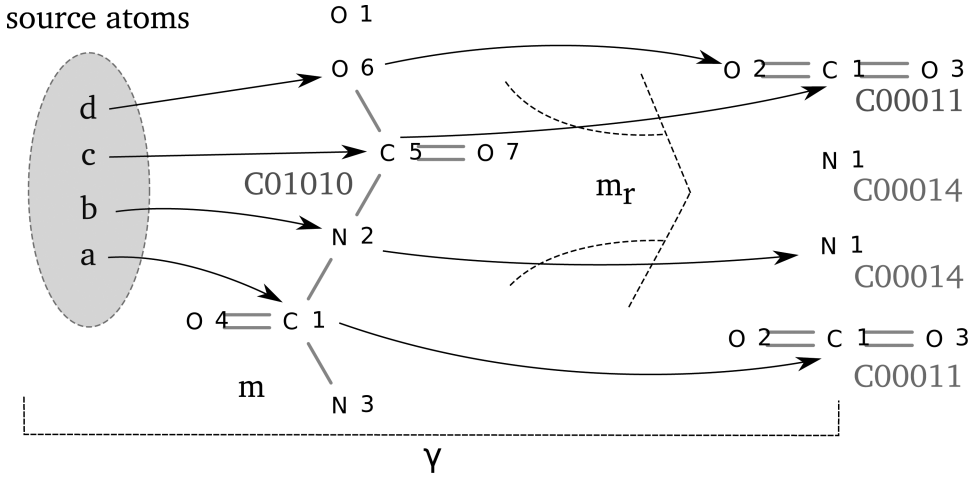


Figure 7. Schematic illustration of a tracking result. Hydrogens and their associated bonds are not shown.

carbon ◇ in agmantine. Therefore carbon * in ornithine has alternative fates if one considers symmetry while tracing. It is easy to see that the symmetry of a reaction graph is broken only if either the permutation symmetry of its components or any of their inherent symmetry is broken.

Evaluating the Impact of Alternative Tracings Using Tabu Search. We devised a simple tabu search algorithm (Algorithm 1 below) to investigate the impact of alternative tracings by conducting a random walk on the metabolic network from a given source metabolite. Three consequences of alternative tracing are investigated along the walk.

1. Differential in isotopomer mapping matrix (IMM)
   This is the case when source IDVs are traced to different target IDVs. In other words, under different atom mapping matrices, the isotopomer mapping matrices are different. For example, consider atoms 1, 2, 3 in the source metabolite and atoms 7, 8, 9 in the target metabolite; mapping $\{1 \Rightarrow 7, 2 \Rightarrow 8, 3 \Rightarrow 9\}$ and mapping $\{1 \Rightarrow 8, 2 \Rightarrow 7, 3 \Rightarrow 9\}$ are two alternative tracings that will give rise to differential IDVs. For instance, a source IDV with only atom 1 labeled will be mapped to a target IDV with either atom 7 labeled or atom 8 labeled depending on the tracing.
2. Differential in conservation
   This is the case when source atoms are traced to different sets of target atoms. For example, mapping $\{1 \Rightarrow 7, 2 \Rightarrow 8, 3 \Rightarrow 10\}$ and mapping $\{1 \Rightarrow 8, 2 \Rightarrow 7, 3 \Rightarrow 9\}$ are two alternative tracings that will give rise to differential conservation, namely set {7, 8, 10} and set {7, 8, 9}. However, the previous example raised for differential IDV is not differential in conservation (both results in set {7, 8, 9}). It is easy to see that every pair of alternative tracings differential in conservation is differential in IMM.
3. Differential in conservation size
   This is the case when source atoms are traced to different numbers of target atoms. For example, mapping $\{1 \Rightarrow 7, 2 \Rightarrow 8\}$ and mapping $\{1 \Rightarrow 8, 2 \Rightarrow 7, 3 \Rightarrow 9\}$ are differential in conservation size with the former conserving two atoms with the source and the later conserving three atoms with the source. Differential in conservation size has an implication in calculating the atom economy and thus affects pathway inferences based on such criterion. Likewise, every pair of alternative tracings differential in conservation size are differential in conservation.

In this algorithm, a priority queue of so-called *pathway states* is kept. Each pathway state belongs to a metabolite and specifies the set of its atoms conserved along the pathway to the source. Pathway states are sorted by the size of conservation. Initially, the only pathway state in the queue is the chosen source metabolite and the set of all its atoms. In each round, a pathway state is popped from the queue and tracked through all the reactions in which the metabolite is involved. New pathway state spawned from the tracking is pushed into the queue only when its conservation cannot be included in any of the existing pathway states of the metabolite to whom the tracking leads.

We first describe an auxiliary subroutine ReactionTrack, which takes a reaction r, its atom mapping m_r, a set s of atoms, and a tracking direction d as inputs and returns a tracking result. The tracking result γ is a list containing ordered pairs (m, ψ), where m is a metabolite that a subset of s is tracked into and ψ is a mapping that projects the atom(s) tracked to atom(s) in the source metabolite. For example, consider reaction R00005 (C01010 + C00001 ⇔ 2C00011 + 2C00014; see Figure 7); a pathway state of metabolite C01010 with source projection $\{1 \Rightarrow a, 2 \Rightarrow b, 5 \Rightarrow c, 6 \Rightarrow d\}$ (suppose a, b, c, and d are source atoms) is tracked through the reaction and the tracking direction d is from left to right. According to one of the reaction mechanisms, atom 1 maps to atom 1 in C00011, atom 2 maps to atom 1 in C00014, atoms 5 and 6 map to atoms 1 and 2 in the second C00011. The resulting list is $\gamma = [(\text{C00011}, \{1 \Rightarrow a\}), (\text{C00014}, \{1 \Rightarrow b\}), (\text{C00011}, \{1 \Rightarrow c, 2 \Rightarrow d\})]$. Note that in γ one metabolite can appear multiple times with different projections to the source if its stoichiometry in the reaction is higher than 1 (as is the case for C00011 in the example). Three other auxiliary subroutines CheckDiffIMM (Algorithm 2), CheckDiffConserv (Algorithm 3), and CheckDiffConservSize (Algorithm 4) test if a set Γ of tracking results is differential in IMM, conservation, or conservation size, respectively.

Computing Whole Reaction Atom Mappings by Minimizing the Graph Edit Distance. The whole reaction atom mappings are found in two steps. For each reaction, we aimed at finding only one valid reaction atom mapping. Subsequent symmetry analysis will give all the possible reaction atom mappings. This is done by first obtaining a partial mapping for a reaction via composing all the associated RPAIRs (subroutine GreedySeed, elaborated below) and then expanding

Algorithm 1: RandomTabuSearch

input : maximum path length l, source metabolite m_s
output: boolean variables diff_I, diff_c, diff_s, indicating whether the tracking from m_s has given rise to alternative tracings differential in IMM, conservation or the conservation size respectively.

diff_I = False;
diff_c = False;
diff_s = False;
$m_c = m_s$;
c = the set of all atoms of m_s;
while path length $< l$ **do**
  r = a randomly chosen reaction that m_c is involved;
  m_r^0 = a reaction atom mapping of r;
  d = a randomly chosen tracking direction if the metabolite appears on both sides of r, or from the side where the metabolite appears if otherwise.;
  tracking_results: $\Gamma = \emptyset$;
  if r is symmetry-breaking from d **then**
    spawn the set of reaction atom mappings $M_r = \{m_r\}^i$ by composing m_r^0 with the reaction symmetries that are broken from d.;
    foreach $m_r^i \in M_r$ **do**
      $\Gamma = \Gamma \cup \{\text{REACTIONTRACK}\ (r, m_r^i, c, d)\}$;
    end
    if not diff_I **then**
      diff_I = CHECKDIFFIMM(m_s, M_r);
    end
    if not diff_c **then**
      diff_c = CHECKDIFFCONSERV(Γ);
    end
    diff_s = CHECKDIFFCONSERVSIZE(Γ);
    if diff_s **then**
      Break;
    end
  else
    Γ = {REACTIONTRACK (r, m_r^0, c, d)} ;
  end
  m_c, c = randomly choose a metabolite and its conservation from Γ;
end
return diff_I, diff_c, diff_s;

Algorithm 2: CheckDiffIMM

input : source metabolite m_s, reaction atom mappings M_r
output: a boolean variable indicating whether the tracking from m_s has given rise to alternative tracings differential in IMM.

X = a random sample of IDVs of m_s;
foreach $x \in X$ **do**
  foreach $m_r^i, m_r^j \in M_r \times M_r \wedge m_r^i \neq m_r^j$ **do**
    if $m_r^i(x) \neq m_r^j(x)$ **then**
      return True;
    end
  end
end
return False;

the partial mapping to a complete reaction atom mapping by minimizing the graph edit distance based on a cost assignment.[23] In this way, we both take advantage of the manual curation of the reactant pair mapping from the RPAIR database[19] and the power of the automated reaction atom mapping analysis by the criterion of minimum graph edit distance. The method only

Algorithm 3: CheckDiffConserv

input : a list of tracking results Γ
output: a boolean variable indicating whether the tracking from m_s yields alternative tracings differential in conservation.

$S = \emptyset$;
diff$_c$ = False;
foreach $\gamma \in \Gamma$ **do**
  $s = \emptyset$;
  foreach $(m, \psi) \in \gamma$ **do**
    $s = s \cup \{(m, Dom(\psi))\}$
  end
  if $s \notin S \wedge S \neq \emptyset$ **then**
    diff$_c$ = True;
    Break;
  else
    $S = S \cup \{s\}$;
  end
end
return diff$_c$;

Algorithm 4: CheckDiffConservSize

input : a list of tracking results Γ
output: boolean variable diff$_s$ indicating whether the tracking from m_s yields alternative tracings differential in conservation size.

$S = \emptyset$;
diff$_s$ = False;
foreach $\gamma \in \Gamma$ **do**
  $s = \emptyset$;
  foreach $(m, \psi) \in \gamma$ **do**
    $s = s \cup \{(m, |Dom(\psi)|)\}$
  end
  if $s \notin S \wedge S \neq \emptyset$ **then**
    diff$_s$ = True;
    Break;
  else
    $S = S \cup \{s\}$;
  end
end
return diff$_s$;

fails (in minutes) to yield a atom mapping for reaction R06447 because of its high graph edit distance. The resulting reaction atom mappings are manually inspected for correctness.

In GreedySeed, we first find the set Θ of all the possible atom mappings that are derived from any of the RPAIRs linked to the reaction in either direction. Each atom mapping $\theta \in \Theta$ maps atoms in a substrate to atoms in a product. We make use of a subroutine SetMapping that takes as input a partial mapping m, augments it by the mappings in Θ, and returns the new partial mapping that is expanded. We verify that the domain of θ has no overlap with the domain of m before the augmentation.

Chemical Graph Standardization. Because of inconsistency in the representations of compounds in the KEGG RPAIR database (multiple labeling for the same compound in different RPAIRs), compound graphs are first standardized to the representation used in KEGG LIGAND database.[28] We identified 394 inconsistencies in compound representations from the KEGG RPAIR database. These representations are standardized by running GRAPHISO(G_r, G_L) (where G_r is the representation used in RPAIR and G_L is the (standard) representation used in LIGAND) and finding a graph isomorphism (an arbitrary one if there are many) preserving chemical facts. Four compounds contain irreconcilable representations which are not amenable to standardization because of discrepancy in the compound structure. Three such compounds (C02419, C03624, and C06185) are manually standardized.

RESULTS

Compound Symmetry in Metabolic Networks. We study the symmetry of the entire reaction by first considering the inherent symmetry of each reactant. A total of 14 066 compounds in the KEGG COMPOUND database[28] with a defined representation were analyzed for symmetry. We identified 9131 compounds in the KEGG LIGAND database that possess nontrivial symmetry mappings assembled from all subgroups and 5912 compounds with nontrivial symmetry mappings assembled from the chosen subgroups (see Methods). They include the most studied examples of rotationally symmetric compounds such as

Algorithm 5: GreedySeed

```
input  : Θ the set of all mapping derived from RPAIRs linked to a given reaction.
output: a partial mapping of the reaction
Q = an empty priority queue of partial mappings prioritized by the number of atoms
mapped;
foreach θ ∈ Θ do
    m = an empty partial mapping;
    m = SETMAPPING(m, θ);
    push m into Q;
end
expandable = True;
while expandable do
    m = first element in Q;
    expandable=False;
    foreach θ ∈ Θ do
        if θ does not contradict m then
            m′ = SETMAPPING(m, θ);
            push m′ into Q;
        end
    end
end
return first element in Q;
```

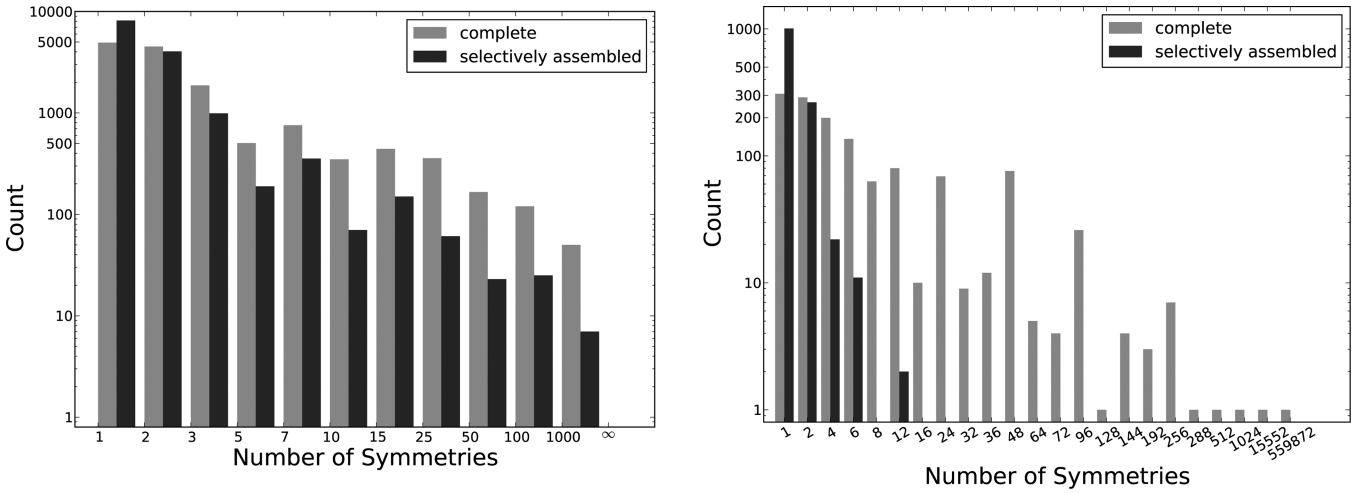


Figure 8. Distribution of symmetric compounds. (left) All compounds in KEGG LIGAND. (right) Compounds in assembled *E.coli* metabolism. Bar heights are in log scale. The number of symmetries includes the identity mappings, meaning that compounds with one symmetry are essentially asymmetric. Selectively assembled symmetries exclude noncarbon subgroups (using only CC2, CC3, benzene mirror, and the remaining subgroup; see the Methods section.).

fumarate and succinate.[30] In Figure 8, we plot the distribution of the number of symmetry mappings in the entire KEGG COMPOUND database and in assembled networks of the microbial organism *E.coli*. Table 1 tabulates some exceptionally highly symmetric compounds (with more than 1000 symmetry mappings found in KEGG). We see that most compounds have a small number of symmetry mappings. Most highly symmetric compounds are drugs or nonbiomolecules (see Table 1) and do not appear in the *E.coli* metabolism (see Figure 8).

Carbon and oxygen are the two mostly exchanged elements in a symmetry mapping (shown in Figure 9). Next are nitrogen and chloride. Seletive assembly reduces symmetries involving oxygen, nitrogen and sodium most significantly.

Reaction Symmetry. Besides inherent compound symmetry, the symmetry of atoms in a reactant/product graph also results from multiple occurrences of the same reactant or product. This is referred to as non-1−0 stoichiometry, which RPAIR data fails to handle properly.[11] Using pruned symmetries subgroups of compounds, we explicitly enumerated all the reaction symmetries originated from inherent compound symmetry as well as from permutating compounds with non-1−0 stoichiometry. Twenty reactions from KEGG database are intractable due to high stoichiometry which leads to the difficulty in enumerating all the permutations. They are R00918, R05185, R05464, R06448, R06453, R06458, R06459, R06480, R06481, R06482, R06483, R06635, R06636, R06637, R06641, R06643, R06644, R06645, R07251, and R08649. And these reactions are lumped representations of multiple elementary reaction steps and do not appear in the assembled *E.coli* metabolic network. Out of 8163 KEGG reactions that are amenable to explicit symmetry enumeration,

Table 1. Highly Symmetric Compounds in the KEGG LIGAND Database[a]

ID	common name	classification	no. symmetries	no. selectively assembled symmetries
C00374	heparin	drug	31104	1
C00925	heparan sulfate	drug	31104	1
C01204	myo-inositol hexakisphosphate	phytic acid	559872	12
C06042	lipoteichoic acid	polymer	33554432	1
C06043	D-Alanyl-lipoteichoic acid	polymer	33554432	1
C07373	probucol	drug	20736	20736
C07974	suramin	drug	93312	2
C11174	1-diphosinositol pentakisphosphate	metabolite	186624	2
C11526	5-diphosphoinositol pentakisphosphate	metabolite	186624	2
C12933	rolitetracycline nitrate	drug (antibiotic)	13824	64
C13523	secretin	polymer	32768	512
C13553	collistin sodium methanesulfonate	drug (antibiotic)	3732480	4
C13704	TX-TM-calixarene	polymer	10368	8
C13723	dextran sulfate	polymer	46656	1
C13768	collistin sulfate	drug (antibiotic)	30576476160	3840
C13932	Ruthenium Red	dye	497664000	2
C14287	4,4′-methylenebis(2,6-ditert-butylphenol)	curing agent	10368	10368
C15435	fenbutatin oxide	pesticide	294912	294912
C15990	spheroidene	metabolite	15552	2
C15991	myo-inositol pentakisphosphate	metabolite	15552	2
C16001	Reactive Black 5	dye	497664	4
C17142	[heparan sulfate]-*N*-sulfoglucosamine	drug	31104	1

[a]Selectively assembled symmetries exclude non-carbon subgroups (using only CC2, CC3, benzene mirror, and the remaining subgroup).

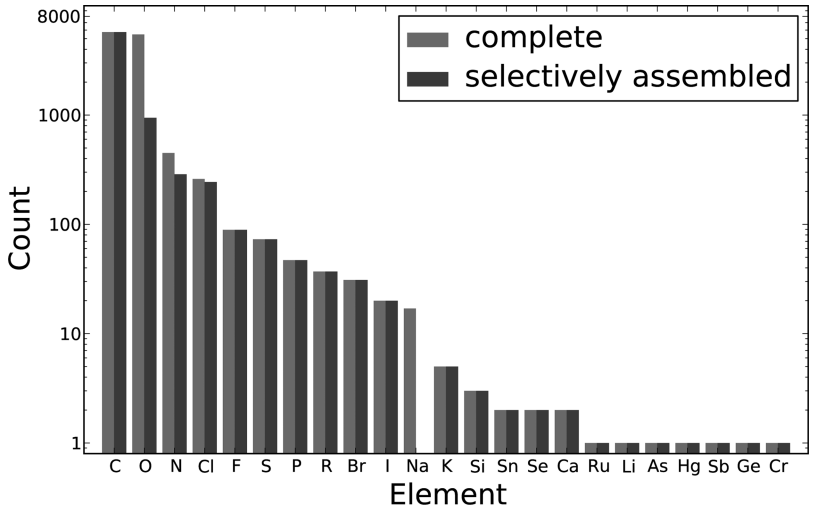


Figure 9. Element composition of symmetry mappings. The height of each bar is the number of symmetry mappings that alter at least one atom of the element. Selectively assembled symmetries have reduced composition in oxygen, nitrogen, and sodium. Selectively assembled symmetries exclude noncarbon subgroups (using only CC2, CC3, benzene mirror, and the remaining subgroup; see the Methods section).

7194 reaction sides contain nontrivial symmetries (every reaction has two sides). When restricted onto the *E.coli* metabolic network, we found that out of 5647 cases where a compound appears in a reaction, only 149 (~2.6%) appear more than once. In 121 of these 149 cases, compounds participate exactly twice. In only 27 cases, compounds participate more than twice. Out of these non-1−0 stoichiometry compounds, even fewer contain inherent symmetry (see the left panel of Figure 10), indicating rare cases where the total number of symmetries is attributed to both non-1−0 stoichiometry and inherent compound structures. Out of 1398 metabolic reactions in *E.coli*, 539 of the reactions have nontrivial symmetries on either side. Among them, 77 reactions have nontrivial symmetry from non-1−0 stoichiometry. Also, 494 reactions inherit their symmetries from inherent symmetries of the component reactants or products. Here, 32 reactions have symmetries coming from both sources. The distribution of the number of symmetry mappings in all reactant and product graphs of all reactions in *E.coli* has a power-law shape, as shown in the right panel of Figure 10.

Symmetry-Breaking Reactions. In studying symmetry-breaking reactions, we studied two classes of atom mappings, atom mappings coming from entire reactions and atom mappings restricted to the atom transition between two compounds (as is the case of the RPAIR data).

We first look at the number of symmetries on different sides of the reaction. In Figure 11, we observe that a substantial number of reactions have discrepant numbers of symmetries on the two sides. This signifies the breaking and emergence of symmetries. On the

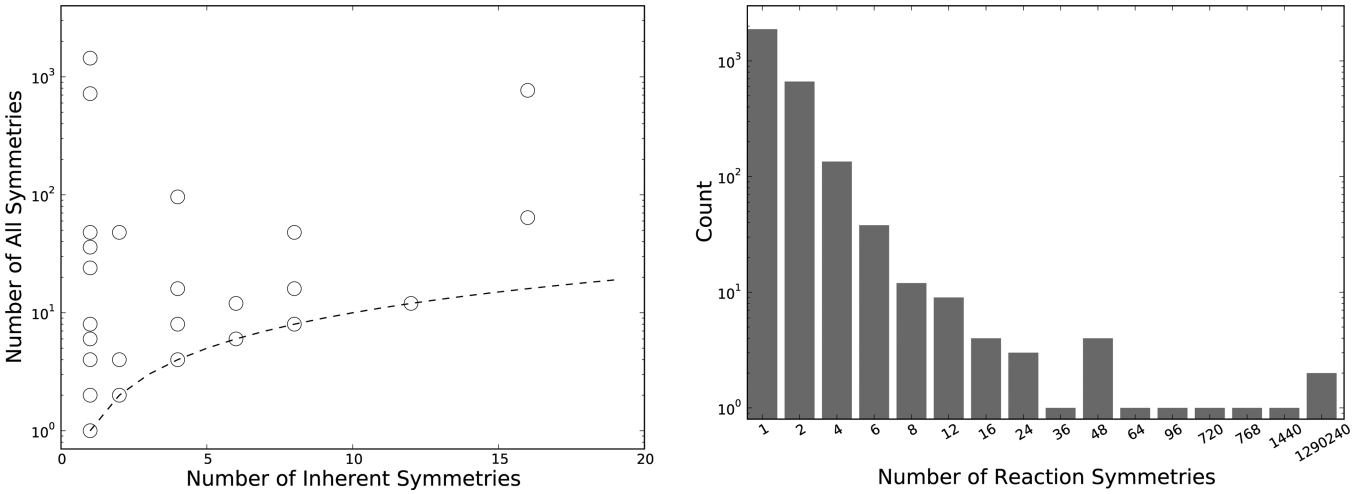


Figure 10. Distribution of reaction symmetries. (left) Increase in the number of symmetries from non-1−0-stoichiometry. The dashed line corresponds to $x = y$, that is, symmetries come solely from inherent structures of the compounds. (right) Distribution of the number of symmetric forms of both reactant graphs and product graphs of all reactions in the metabolic network of *E.coli*.

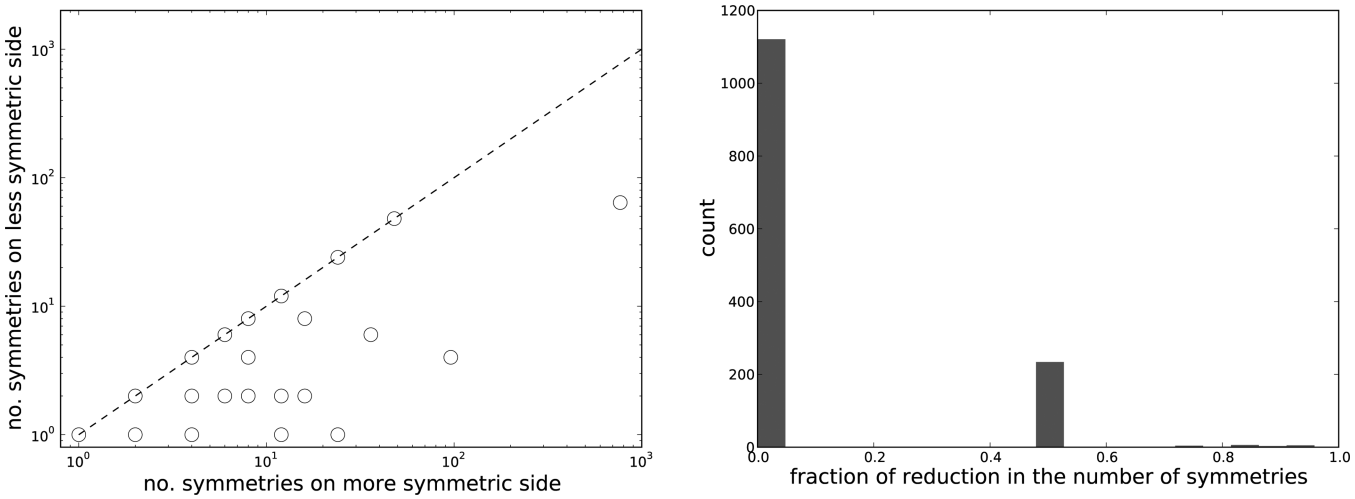


Figure 11. Distribution of symmetry-breaking reactions in the *E.coli* metabolic network. (left) Only reactions with all reactants structually defined plotted. One dot on the plot can contain more than 1 reaction. (right) Distribution of the reduction in the number of symmetry mappings of reactions in the metabolic network of *E.coli*. The relative reduction is calculated by normalizing the difference between the number of symmetric forms of reactant graph and of product graph by the larger of the two. Symmetry mappings exclude noncarbon subgroups (using only CC2, CC3, benzene mirror, and the remaining subgroup; see the Methods section). Note that having equal numbers of symmetries on the two sides of a reaction does not exclude the possibility of symmetry-breaking.

basis of the source of the symmetry the atom mappings break, we classify all the symmetry breaking reactions into two types. Type 1 involves the breaking of symmetries from non-0−1 stoichiometry. Type 2 involves the breaking of reaction symmetries that are inherent to the participating compounds. The inherent symmetries are assembled from compound symmetries excluding noncarbon subgroups (using only CC2, CC3, benzene mirror, and the remaining subgroup). There are 1398 metabolic reactions in *E.coli*. Out of 1257 reactions with defined atom mappings, we found 56 reactions having type 1 symmetry breaking of either the reactant or the product graph (12 of which have type 1 symmetry breaking on both sides). For type 2 symmetry breaking, 217 reactions are identified and 15 of them are broken from both sides of the reaction. Here, 257 reactions have symmetry breaking of either types. Sixteen of them have symmetry breaking of both types.

Regeneration of RPAIR Atom Mapping. Several studies have depended on the KEGG's RPAIR database for inferring metabolic pathways.[9−11] But due to the lack of explicit incorporation of symmetry-induced alternative mappings, RPAIR data is insufficient for the task of pathway inference.[1]

In this study, atom mappings in RPAIR database are first checked to see whether they break the computed symmetries of the constituent compounds. Out of 2369 reactant pairs that are involved in reactions from the *E.coli* metabolome, we identified 176 reactant pairs that break the symmetry of at least one of the two constituent compounds. The identification is based on the definition of symmetry breaking of compound graphs (see Methods). Also, 257 reactant pairs are added to the data of KEGG RPAIR database by composing the compound(s) symmetry that each rectant pair breaks with the atom mapping from the reactant pair itself.

Note that in some cases where the stoichiometry is higher than 1, the RPAIR database already provides multiple atom mappings for the reaction if these atom mappings are different (e.g., R00006 has RP00440 and RP12733 for pyruvate). But due to the incompleteness of RPAIR database in covering the

dx.doi.org/10.1021/ci300259u | J. Chem. Inf. Model. 2012, 52, 2684−2696

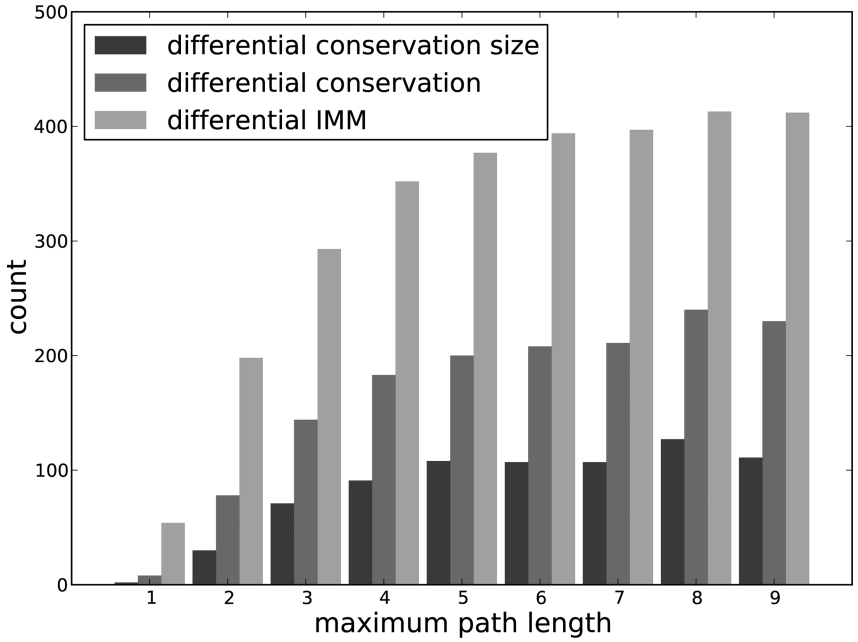


Figure 12. Distribution of the three consequences of alternative tracings in tabued random walks of the metabolite network of *E.coli*.

reaction atom mappings, we regenerated the atom mapping between every pair of reactants in all the KEGG reactions except the 20 reactions with high stoichiometries and infeasible to explicit enumeration of all their symmetry mappings (see the Reaction Symmetry section). We built 15 196 pairwise compounds atom mapping data from 6809 reactions. The pairwise reactant atom mapping data regenerated not only covers all the atoms in each reactions but also has taken into consideration of the reaction symmetry.

Impact of Alternative Tracing to Atom Economy and Isotopomer Distribution Vector. Taking each of the 1261 metabolites from E.coli network as the source, we repeated 200 times the tabu search (see Methods) and collected, upon encountering of a symmetry-breaking reaction, the information on whether it gives rise to any of the three consequences: differential in IMM, differential in conservation, and differential in size (see Methods). We plotted the number of sources from which any of the 200 searches within a prespecified maximum length (the search scope) would result in alternative tracings differential in IMM, conservation, and conservation size, respectively (see Figure 12). We observe that as the search scope increases, so does the chance of encountering alternative tracings with any of the three consequences. The increase in the probability of seeing any alternative tracing plateaus after the maximum path length exceeds 5. For more than 350 metabolites out of 1261 (~27.8%), at least one search out of 200 random paths of length higher than 5 yielded alternative tracings differential in IMM. More than 200 (~15.9%) of them are differential in conservation, and more than 100 (~7.9%) exhibit difference in the size of the atom set conserved. In fact, any difference in atom tracing would be reflected in the atom mapping matrix (AMM), hence changing the isotopomer mapping matrix (IMM). Indeed, in every case where an alternative tracing arises, there is a difference in IMM.

■ DISCUSSION AND CONCLUSION

Although previous studies on pathway inference largely ignored the impact of symmetry in compound and reaction,[32] several automated methods exist that are able to generate alternative mapping directly from scratch.[1,22,23,30] BioNetGen provides a method that makes proper corrections for symmetries in counting molecule observables.[33] Antoniewicz et al.[30] provided a method for computing elementary metabolite units (EMU) considering equivalent atoms without giving a solution to how these equivalent atoms can be identified. Algorithms designed by Heinonen et al.[23] as well as Crabtree and Mehta[22] can compute alternative reaction atom mappings in the form of equally optimal solutions. However, there is no apparent extension to predict, using their algorithm, the number of equally optimal solutions, which can be prohibitively large when the number of symmetry mappings are high. Our work can supplement the automatic generation of reaction atom mapping in estimating the number of equally optimal solutions.

Ravikirthi et al.[1] curated the atom mappings for 2077 reactions present in a genomewise construction of the *E.coli* network, combining an automated method based on heuristics such as maximum common subgraph[20] with manual curation of not only symmetry but also chirality and prochirality.[25] Particularly, these authors manually treated "equivalent oxygen atoms" and "rotational symmetric molecules". The authors found 653 reactions that contained compounds with at least one kind of symmetry. For each atom mapping they curated, alternative mappings ("mapping degeneracies") are investigated. However, the maximum common subgraph heuristic is reported to be less accurate in returning the true reaction atom mapping[22,23] compared to ones that minimize the number of bond break and formation. Moreover, the curation in Ravikirthi et al.[1] is limited to one model organism, *E.coli*, and one of its reconstructed model, iAF1260.[34] On the contrary, the curation from RPAIR is organism-independent but does not handle non-1−0 stoichiometry properly[11] and is not complete in terms of covering all the atoms in the reaction. We took advantage of this but employed a graph-theoretic, minimum graph edit based method for expanding the atom mappings from RPAIR to complete the atom mapping for each entire reaction.

In all the previously mentioned studies, symmetry was only implicitly accounted for. No symmetry mapping was computed. In this study, we explicitly computed compound and reaction symmetries by using an adapted method for finding graph automorphisms.

From the random walking experiment, we observe that symmetry-initiated alternative tracings is nonnegligible (with around 27.8% chance of emergence in the *E.coli* metabolism) and could result in miscalculation of atom economy (around 7.9% in the *E.coli* metabolism) and isotopomer mapping matrices (everytime an alternative tracing is seen). This observation highlights the significance of having symmetry-aware atom mapping data when one calculates atom economy in pathway inference and computes the distribution of isotopomers.

There are some obstacles in the accurate in silico computation of whole reaction atom mappings. The first involves reactions that are formed from multiple elementary reactions aggregated together. On those reactions, the automated computation of atom mapping becomes infeasible when the graph edit distance is too large. Moreover, the high stoichiometry gives high symmetry from permutating the reactants of the same kind.

The second involves the curation of 3D symmetry operations. In this work, we have ignored prochirality for simplicity. When we consider 3D configurations, some prochiral substituents can be differentiated, leaving some symmetries detected using graph-theoretic methods invalid. In other words, symmetries that arises soley from exchanging two prochiral substituents are invalid under a 3D point of view. We employed the following ad hoc method for a preliminary detection of prochiral carbon centers. For each carbon of a compound, we studied all of its covalently linked substituents. We compared every pair of the substituents by whether they are isomorphic. When testing graph isomorphism, in each chemical graph of the substituent, we disconnect the center carbon atom from all of its neighbors except from the substituent under study. The center carbon's element attribute is uniquely relabeled to differentiate from other carbons. The number of distinct (in the graph isomorphism sense) substituents are then coded in a sorted list. For example, a carbon has the code (2,1) if it has three substituents, two of them are isomorphic, and none of the two is isomorphic to the third substituent. We detect among all the KEGG compounds those that have carbons with code (2,1) or (2,1,1). These carbons are putative prochiral centers, and the exchange of the two isomorphic substituents are invalid symmetries in the 3D point of view. We find 2904 out of 14 066 compounds from KEGG with such prochiral carbon center(s). These prochiral carbon centers are special cases of the more general 3D restriction which is beyond the scope of this paper. We identify the full resolution of 3D symmetry as a future direction.

Third, there are further uncertainties from not knowing detailed reaction mechanisms in this automated approach. These mechanisms could alter atom mappings and symmetry operations in ways such as triggering uncommon atom transitions which could be of very high graph edit distance, or having special contraints from timing and synchronization that prohibits certain symmetry operations. How much these uncertainties affect our reported results remains a question whose answering requires more detailed information on enzymes and reaction thermodynamics.

To summarize, in this study, we formulated the problem of compound and reaction symmetry as a graph automorphism problem. We explicitly computed symmetry mappings of reactions either from non-1−0 stoichiometry or inherent symmetries of the reactants/products. We motivated the concept of symmetry-breaking reactions and studied its extent. Random walk on the metabolic network revealed significant impact of alternative tracings to pathway inference and isotopomer distribution simulation. Technically, we augmented the KEGG RPAIR data by first expanding atom mappings from RPAIR and then composing symmetries that are broken to complete the whole reaction atom mapping.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Wanding.Zhou@rice.edu (W.Z.); nakhleh@rice.edu (L.N.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported in part by NSF grant CCF-0622037 and an Alfred P. Sloan Research Fellowship. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NSF or the Alfred P. Sloan Foundation. We also thank editor Wendy A. Warr and two anonymous reviewers for valuable comments that helped improve the manuscript.

REFERENCES

(1) Ravikirthi, P.; Suthers, P. F.; Maranas, C. D. Construction of an E. coli genome-scale atom mapping model for MFA calculations. *Biotechnol. Bioeng.* **2011**, *108*, 1372−82.

(2) Szyperski, T. 13C-NMR, MS and metabolic flux balancing in biotechnology research. *Q. Rev. Biophys.* **1998**, *31*, 41−106.

(3) Blum, T.; Kohlbacher, O. Using atom mapping rules for an improved detection of relevant routes in weighted metabolic networks. *J. Comput. Biol.* **2008**, *15*, 565−576.

(4) Hogiri, T.; Furusawa, C.; Shinfuku, Y.; Ono, N.; Shimizu, H. Analysis of metabolic network based on conservation of molecular structure. *Biosystems* **2009**, *95*, 175−178.

(5) Mu, F.; Williams, R. F.; Unkefer, C. J.; Unkefer, P. J.; Faeder, J. R.; Hlavacek, W. S. Carbon-fate maps for metabolic reactions. *Bioinformatics* **2007**, *23*, 3193−9.

(6) Tohsato, Y.; Nishimura, Y. Reaction similarities focusing substructure changes of chemical compounds and metabolic pathway alignments. *Inf. Media Tech.* **2009**, *4*, 390−399.

(7) Rivas-Ubach, A.; Sardans, J.; Pérez-Trujillo, M.; Estiarte, M.; Peñuelas, J. Strong relationship between elemental stoichiometry and metabolome in plants. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 4181−6.

(8) Trost, B. M. Atom Economy − A Challenge for Organic Synthesis: Homogeneous Catalysis Leads the Way. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 259−281.

(9) Heath, A. P.; Bennett, G. N.; Kavraki, L. E. Finding metabolic pathways using atom tracking. *Bioinformatics* **2010**, *26*, 1548−1555.

(10) Faust, K.; Croes, D.; van Helden, J. Metabolic pathfinding using RPAIR annotation. *J. Mol. Biol.* **2009**, *388*, 390−414.

(11) Pitkänen, E.; Jouhten, P.; Rousu, J. Inferring branching pathways in genome-scale metabolic networks. *BMC Syst. Biol.* **2009**, *3*, 103.

(12) Schellenberger, J.; Zielinski, D. C.; Choi, W.; Madireddi, S.; Portnoy, V.; Scott, D. A.; Reed, J. L.; Osterman, A. L.; Palsson, B. O. Predicting outcomes of steady-state 13C isotope tracing experiments with Monte Carlo sampling. *BMC Syst. Biol.* **2012**, *6*, 9.

(13) Wiechert, W.; Siefke, C.; de Graaf, A. A.; Marx, A. Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis. *Biotechnol. Bioeng.* **1997**, *55*, 118−35.

(14) Szyperski, T. Biosynthetically directed fractional 13C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism. *Eur. J. Biochem.* **1995**, *232*, 433−48.

(15) Schmidt, K.; Carlsen, M.; Nielsen, J.; Villadsen, J. Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnol. Bioeng.* **1997**, *55*, 831−40.
(16) Rosenthal, J.; Murphy, G. Group theory and the vibrations of polyatomic molecules. *Rev. Mod. Phys.* **1936**, *8*, 317−346.
(17) Harrisand, D.; Bertolucci, M. *Symmetry and spectroscopy: an introduction to vibrational and electronic spectroscopy*; Dover Publications: Mineola, New York, 1989.
(18) Atkins, P.; de Paula, J. *Physical Chemistry*, 9th ed.; W. H. Freeman: New York, New York, 2009.
(19) Kotera, M.; Hattori, M.; Oh, M.; Yamamoto, R. T.; Komeno, T.; Yabuzaki, J.; Tonomura, K.; Goto, S.; Kanehisa, M. RPAIR: a reactant-pair database representing chemical changes in enzymatic reactions. *Genome Inf.* **2004**, P062.
(20) Hattori, M.; Okuno, Y.; Goto, S.; Kanehisa, M. Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.* **2003**, *125*, 11853−11865.
(21) Hattori, M.; Tanaka, N.; Kanehisa, M.; Goto, S. SIMCOMP/SUBCOMP: chemical structure search servers for network analyses. *Nucleic Acids Res.* **2010**, *38*, W652−6.
(22) Crabtree, J. D.; Mehta, D. P. Automated reaction mapping. *J. Exp. Algorithmics* **2009**, *13*, 15−29.
(23) Heinonen, M.; Lappalainen, S.; Mielikäinen, T.; Rousu, J. Computing atom mappings for biochemical reactions without subgraph isomorphism. *J. Comput. Biol.* **2011**, *18*, 43−58.
(24) Shapiro, L. G.; Haralick, R. M. Structural descriptions and inexact matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **1981**, *PAMI-3*, 504−519.
(25) Moss, G. P. Basic terminology of stereochemistry (IUPAC Recommendations 1996). *Pure Appl. Chem.* **1996**, *68*, 2193−2222.
(26) Luks, E. M. Isomorphism of graphs of bounded valence can be tested in polynomial time. In *Proceedings of 21st Annual Symposium on Foundations of Computer Science*, Syracus, New York, Oct 13−15, 1980; pp 42−49.
(27) Cordella, L. P.; Foggia, P.; Sansone, C.; Vento, M. An improved algorithm for matching large graphs. In *Proceedings of 3rd IAPR-TC15 Workshop on Graph-based Representations in Pattern Recognition*, Cuen, May 23−25, 2001; pp 149−159.
(28) Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27−30.
(29) Artin, E. *Galois Theory: Lectures Delivered at the University of Notre Dame by Emil Artin (Notre Dame Mathematical Lectures, Number 2)*; Dover Publications: Mineola, New York, 1997.
(30) Antoniewicz, M. R.; Kelleher, J. K.; Stephanopoulos, G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab. Eng.* **2007**, *9*, 68−86.
(31) Balducci, R.; Pearlman, R. S. Efficient exact solution of the ring perception problem. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 822−831.
(32) Pitkänen, E. Ph.D. Computational Methods for Reconstruction and Analysis of Genome-Scale Metabolic Networks. Thesis, University of Helsinki, 2010.
(33) Blinov, M. L.; Faeder, J. R.; Goldstein, B.; Hlavacek, W. S. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics* **2004**, *20*, 3289−3291.
(34) Feist, A.; Henry, C.; Reed, J. L.; Krummenacker, M.; Joyce, A.; Karp, P.; Broadbelt, L.; Hatzimanikatis, V.; Palsson, B. O. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **2007**, *3*, 121.