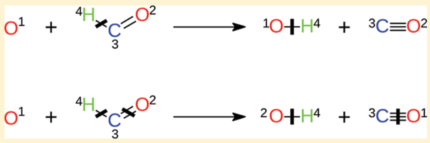
Journal of
CHEMICAL INFORMATION
AND MODELING

Stereochemically Consistent Reaction Mapping and Identification of Multiple Reaction Mechanisms through Integer Linear Optimization

Eric L. First, Chrysanthos E. Gounaris, and Christodoulos A. Floudas*

Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, United States

ABSTRACT: Reaction mappings are of fundamental importance to researchers studying the mechanisms of chemical reactions and analyzing biochemical pathways. We have developed an automated method based on integer linear optimization, ILP, to identify optimal reaction mappings that minimize the number of bond changes. An alternate objective function is also proposed that minimizes the number of bond order changes. In contrast to previous approaches, our method produces mappings that respect stereochemistry. We also show how to locate multiple reaction mappings efficiently and determine which of those mappings correspond to distinct reaction mechanisms by automatically detecting molecular symmetries. We demonstrate our techniques through a number of computational studies on the GRI-Mech, KEGG LIGAND, and BioPath databases. The computational studies indicate that 99% of the 8078 reactions tested can be addressed within 1 CPU hour. The proposed framework has been incorporated into the Web tool DREAM (http://selene.princeton.edu/dream/), which is freely available to the scientific community.



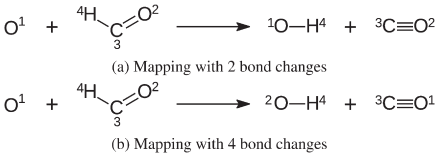
INTRODUCTION

Reaction mappings, that is, one-to-one correspondences between reactant and product atoms, serve as a basis for the study of chemical reactions, as each mapping implies a possible reaction mechanism. Particularly with biochemical pathways, which can involve networks of complex reactions, an automated approach for stereochemically consistent reaction mappings is of great benefit to researchers. Knowledge of reaction mechanisms is also important for the calculation of chemical kinetics and the generation of transition state structures.[1−3]

In general, reactions can have multiple mappings, and those mappings that most likely correspond to an exhibited mechanism are of interest. For example, consider the elementary reaction O + HCO → OH + CO, illustrated in Figure 1. The reaction has two possible mechanisms that correspond to two different mappings: the first involves breaking the C−H bond and forming an O−H bond, a total of two bonds breaking and forming; the second involves breaking both of the bonds in reactant HCO and forming bonds between the C and isolated O and between the O and H that were connected to the C in the reactants, a total of four bond changes. An objective of our approach is to identify a mapping with the fewest number of bond changes, the first mapping in this example.

Previously, reaction mapping algorithms were based on finding the maximum common subgraph[4−7] or maximum common edge subgraph[8] between reactants and products. Such methods, in essence, minimize the number of reaction centers, but this may not necessarily correspond to the optimal reaction mechanism. Another approach based on the weighted maximum common edge subgraph problem minimizes an approximate reaction energy referred to as the "imaginary transition state energy".[9,10] Other methods have been developed for special classes of reactions, such as the algorithms based on graph partitioning and isomorphism presented by Akutsu.[11]

An important advance in automated reaction mapping was made by Crabtree and Mehta[12,13] when they developed a method based on canonical graph naming. Unlike maximum common subgraph methods, their approach minimizes the number of bonds that must be broken on each side of the reaction in order to reduce it to an identity reaction. Much effort has been made to improve their algorithms,[14] which involve a clever enumeration of possible mappings.



(a) Mapping with 2 bond changes

(b) Mapping with 4 bond changes

Figure 1. Two possible mappings for the elementary reaction O + HCO → OH + CO.

An optimization-based approach to solving the reaction mapping problem has the advantage that efficient algorithms already exist for locating an optimal solution without an exhaustive enumeration of possibilities. The approach allows for additional features not present in other methods, such as identification of multiple mechanisms, alternative objective functions, and simultaneous accommodation of stereochemical consistency. Optimization models also have a degree of modularity, with the possibility for future functionality to be incorporated through additional blocks of constraints.

For the first time, we present a reaction mapping method that correctly considers stereochemistry, which has been previously neglected by other approaches. Ignoring stereochemical information can lead to suboptimal solutions where a change in chirality is required but not accounted for. Stereochemical consistency is particularly important for the generation of transition

Received: July 27, 2011
Published: November 19, 2011

state structures, where the spatial arrangement plays a key role in determining the result. In contrast to most other approaches, we also include hydrogen atoms in our model, which is important because they can play an important role in the reaction mechanism and affect the optimal solution.

We also present for the first time in the open literature an automated method to detect the equivalence over symmetry of two or more reaction mappings. This enables us to distinguish mappings that correspond to distinct reaction mechanisms. Therefore, by locating multiple mappings for a chemical reaction, we are able to identify each of its possible mechanisms.

Finally, we have incorporated the proposed framework into the Web tool DREAM (http://selene.princeton.edu/dream/), which is freely available to the scientific community.

MATHEMATICAL MODEL FORMULATION

Most generally, mixed-integer linear optimization (MILP) problems have the form:[15]

$$\min c^T x + d^T y$$

$$\text{s.t. } Ax + By \le b$$

$$x \in X \subset \mathbb{R}^n$$

$$y \in \{0, 1\}^q$$

where x is a vector of n continuous variables; y is a vector of q binary variables; b, c, and d are vectors of parameters; and A and B are parameter matrices. Standard solution techniques exist for solving MILP models (e.g., branch and bound, branch and cut) and are efficiently implemented in commercial solvers. Briefly, the branch and bound technique employs a binary tree to represent possible combinations of the y variables. At the root node of the binary tree, all y variables are relaxed to continuous variables on [0,1], and the now continuous linear model (LP) is solved using standard linear optimization methods (e.g., simplex method). For each branch in the tree, a y variable is selected and fixed to 0 or 1 in each of the two child nodes, and the resulting LP problems are solved. Fathoming criteria are used to locate a guaranteed optimal solution without exploring the entire binary tree.

We aim to express the reaction mapping problem as an MILP model where we solve for a mapping that minimizes the number of bonds broken and formed. Such a model is developed in the remainder of this section.

Sets and Parameters.

We define sets of atoms and bonds for the reaction:

$\mathrm{A} = \{1,2,...,n\}$: atom indices involved in the reaction

$\mathrm{B_R} = \{(i,j)\text{: } i, j \in \mathrm{A}, i < j, i \text{ and } j \text{ bonded in reactants}\}$: reactant bonds

$\mathrm{B_P} = \{(k,l)\text{: } k, l \in A, k < l, k \text{ and } l \text{ bonded in products}\}$: product bonds

We define sets related to the stereochemistry of tetrahedral atoms:

$\mathrm{C_R} \subset \mathrm{A}$: indices of tetrahedral atoms in reactants

$\mathrm{C_P} \subset \mathrm{A}$: indices of tetrahedral atoms in products

$\mathrm{CA_R^{\mathit{i}}}\ \forall i \in \mathrm{C_R}$: indices of atoms bonded to tetrahedral atom i in reactants (ordered set)

$\mathrm{CA_P^{\mathit{k}}}\ \forall k \in \mathrm{C_P}$: indices of atoms bonded to tetrahedral atom k in products (ordered set)

$\mathrm{CQ} = \{\{1,4,3,2\}, \{1,3,2,4\}, \{1,2,4,3\}, \{2,1,3,4\}, \{2,3,4,1\}, \{2,4,1,3\}, \{3,1,4,2\}, \{3,4,2,1\}, \{3,2,1,4\}, \{4,1,2,3\}, \{4,2,3,1\}, \{4,3,1,2\}\}$: mappings that change tetrahedral stereochemistry (ordered set)

We define sets related to the stereochemistry of double bonds:

$\mathrm{D_R} \subset \mathrm{B_R}$: stereochemical double bonds in reactants

$\mathrm{D_P} \subset \mathrm{B_P}$: stereochemical double bonds in products

$\mathrm{DA_R^{(\mathit{i,j})}}\ \forall (i,j) \in \mathrm{D_R}$: indices of atoms connected to double bond (i,j) in reactants (ordered set)

$\mathrm{DA_P^{(\mathit{k,l})}}\ \forall (k,l) \in \mathrm{D_P}$: indices of atoms connected to double bond (k,l) in products (ordered set)

$\mathrm{DQ} = \{\{1,2,4,3\}, \{2,1,3,4\}, \{3,4,2,1\}, \{4,3,1,2\}\}$: mappings that change double bond stereochemistry (ordered set)

We define parameters and sets for the atom types:

$T_R^i\ \forall i \in \mathrm{A}$: type of atom i in reactants

$T_P^k\ \forall k \in \mathrm{A}$: type of atom k in products

$\mathrm{N_R^{\mathit{i}}} = \{T_R^j\ \forall j \in \mathrm{A}\text{:}(j,j) \in \mathrm{B_R} \text{ or } (j,i) \in \mathrm{B_R}\}\ \forall i \in \mathrm{A}$: types of atoms that neighbor atom i in reactants

$\mathrm{N_P^{\mathit{k}}} = \{T_P^l\ \forall l \in \mathrm{A}\text{:}(k,l) \in \mathrm{B_P} \text{ or } (l,k) \in \mathrm{B_P}\}\ \forall k \in \mathrm{A}$: types of atoms that neighbor atom k in products

We use square bracket notation to index within an ordered set. For example, $\mathrm{CA_R^3}[1]$ is the index of the first atom bonded to tetrahedral atom 3 in the reactants. $\mathrm{CA_P^2}[4]$ is the index of the fourth atom bonded to tetrahedral atom 2 in the products. CQ[1][4] = 2 is the fourth element of the first permutation ({1,4,3,2}) that changes tetrahedral stereochemistry. $\mathrm{DA_R^{(1,2)}}[4]$ is the index of the fourth atom connected to double bond (1,2) in the reactants. $\mathrm{DA_P^{(3,4)}}[2]$ is the index of the second atom connected to double bond (3,4) in the products. DQ[2][1] = 2 is the first element of the second permutation ({2,1,3,4}) that changes double bond stereochemistry.

Variables. We introduce the following binary variables:

$$y_{ik} = \left\{\begin{array}{ll} 1 & \text{if atom } i \text{ in reactants is} \\ & \quad \text{mapped to atom } k \text{ in products} \\ 0 & \text{otherwise} \end{array}\right\} \quad \forall i, k \in A$$

$$\alpha_{ijkl} = \left\{\begin{array}{ll} 1 & \text{if bond } (i,j) \text{ is mapped} \\ & \quad \text{to bond } (k,l) \\ 0 & \text{otherwise} \end{array}\right\} \quad \begin{array}{l} \forall (i,j) \in \mathrm{B_R}; \\ \forall (k,l) \in \mathrm{B_P} \end{array}$$

$$\beta_{ik} = \left\{\begin{array}{ll} 1 & \text{if tetrahedral atom } i \\ & \quad \text{and its neighbors } \mathrm{CA_R^{\mathit{i}}} \\ & \quad \text{are mapped to} \\ & \quad \text{tetrahedral atom } k \\ & \quad \text{and its neighbors } \mathrm{CA_P^{\mathit{k}}} \\ & \quad \text{with a change in} \\ & \quad \text{stereochemistry} \\ 0 & \text{otherwise} \end{array}\right\} \quad \begin{array}{l} \forall i \in \mathrm{C_R}; \\ \forall k \in \mathrm{C_P} \end{array}$$

$$\gamma_{ijkl} = \left\{\begin{array}{ll} 1 & \text{if double bond } (i,j) \\ & \quad \text{and its neighbors } \mathrm{DA_R^{(\mathit{i,j})}} \\ & \quad \text{are mapped to} \\ & \quad \text{double bond } (k,l) \\ & \quad \text{and its neighbors } \mathrm{DA_P^{(\mathit{k,l})}} \\ & \quad \text{with a change in} \\ & \quad \text{stereochemistry} \\ 0 & \text{otherwise} \end{array}\right\} \quad \begin{array}{l} \forall (i,j) \in \mathrm{D_R}; \\ \forall (k,l) \in \mathrm{D_P} \end{array}$$

Integer Linear Optimization Model. The following MILP model is presented to identify a reaction mapping that minimizes the number of bonds broken and formed:

$$\zeta = \min \sum_{(i,j)\,\in\,\mathrm{B_R}} \Big(1 - \sum_{(k,l)\,\in\,\mathrm{B_P}} \alpha_{ijkl}\Big) + \sum_{(k,l)\,\in\,\mathrm{B_P}} \Big(1 - \sum_{(i,j)\,\in\,\mathrm{B_R}} \alpha_{ijkl}\Big) + 2 \sum_{i\,\in\,\mathrm{C_R}} \sum_{k\,\in\,\mathrm{C_P}} \beta_{ik} + 2 \sum_{(i,j)\,\in\,\mathrm{D_R}} \sum_{(k,l)\,\in\,\mathrm{D_P}} \gamma_{ijkl} \tag{1}$$

$$\text{s.t.} \sum_{k\,\in\,\mathrm{A}} y_{ik} = 1 \quad \forall i \in \mathrm{A} \tag{2}$$

$$\sum_{i\,\in\,\mathrm{A}} y_{ik} = 1 \quad \forall k \in \mathrm{A} \tag{3}$$

$$y_{ik} = 0 \quad \forall i,k \in \mathrm{A} : T_{\mathrm{R}}^{i} \neq T_{\mathrm{P}}^{k} \tag{4}$$

$$\alpha_{ijkl} \leq y_{ik} + y_{il} \quad \forall \in \mathrm{B_R} \quad \forall (k,l) \in \mathrm{B_P} \tag{5}$$

$$\alpha_{ijkl} \leq y_{jk} + y_{jl} \quad \forall (i,j) \in \mathrm{B_R} \quad \forall (k,l) \in \mathrm{B_P} \tag{6}$$

$$\beta_{ik} \geq y_{ik} + \sum_{m=1}^{4} y_{\mathrm{CA}_{\mathrm{R}}^{i}[m],\,\mathrm{CA}_{\mathrm{P}}^{k}[\mathrm{CQ}][q][m]]} - 4 \quad \forall q \in \{1, ..., 12\}\ \forall i \in \mathrm{C_R}\ \forall k \in \mathrm{C_P} \tag{7}$$

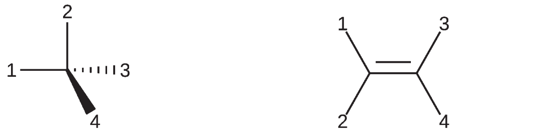
$$\gamma_{ijkl} \geq y_{ik} + y_{jl} + \sum_{m=1}^{4} y_{\mathrm{DA}_{\mathrm{R}}^{(i,j)}[m],\,\mathrm{DA}_{\mathrm{P}}^{(k,l)}[\mathrm{DQ}][q][m]]} - 5 \quad \forall q \in \{1, 2\}\ \forall (i,j) \in \mathrm{D_R}\ \forall (k,l) \in \mathrm{D_P} \tag{8}$$

$$\gamma_{ijkl} \geq y_{il} + y_{jl} + \sum_{m=1}^{4} y_{\mathrm{DA}_{\mathrm{R}}^{(i,j)}[m],\,\mathrm{DA}_{\mathrm{P}}^{(k,l)}[\mathrm{DQ}][q][m]]} - 5 \quad \forall q \in \{3, 4\}\ \forall (i,j) \in \mathrm{D_R}\ \forall (k,l) \in \mathrm{D_P} \tag{9}$$

The objective function 1 consists of four summation terms: the first is over reactant bonds with each term equal to one if the bond breaks (i.e., does not exist in the products). The second is over product bonds with each term equal to one if the bond forms (i.e., does not exist in the reactants). The third is over tetrahedral atoms with each term equal to one if the stereochemistry changes. The fourth is over stereochemical double bonds with each term equal to one if the stereochemistry changes. The value of the objective function, ζ, can be interpreted as the total number of bonds that break and form in the chemical reaction mechanism implied by the mapping.

Constraint 2 requires that each atom in the reactants maps to exactly one atom in the products. Constraint 3 requires that each atom in the products maps to exactly one atom in the reactants. Together, these impose a one-to-one mapping between the atoms in the reactants and products. [We require that the reaction is balanced; i.e., the quantity of each type of atom is the same on both sides of the reaction. The problem is ill-posed for unbalanced reactions.] Constraint 4 allows only atoms of the same type to map to one another.

Constraints 5 and 6 define each α_{ijkl} variable, permitting it to take the value of one only if the reactant bond (i,j) maps to the product bond (k,l). For example, if i maps to k and j maps to l, then $y_{ik} = 1$ and $y_{jl} = 1$. In this case, constraints 5 and 6 both reduce to $\alpha_{ijkl} \leq 1$, and since α_{ijkl} appears in the objective function with a negative coefficient, it will assume the value $\alpha_{ijkl} = 1$ whenever it is permitted.



(a) Tetrahedral atom (b) Stereochemical double bond

Figure 2. Forms of stereochemistry treated by the model.

Constraints 7, 8, and 9 detect changes in stereochemistry and are described in detail in the next section. The complexity of the model for n atoms (assume same type) and m bonds is, in the worst case, $O(n^2 + m^2)$ binary variables and $O(n^2 + m^2)$ linear constraints.

Each feasible solution of the model corresponds to a one-to-one mapping between reactant and product atoms. The mapping is determined by the y_{ik} variables—each taking the value one represents a mapping between reactant atom i and product atom k.

The model is of appropriate form to be readily solved by commercial MILP solvers. For balanced reactions, the model is always feasible. Thus, solving it with standard MILP solution techniques (e.g., branch and bound) will always locate a guaranteed optimal solution.

Stereochemistry. Reaction mappings can be used to gain insights into the three-dimensional transition state structures of chemical reactions. Therefore, an accurate treatment of stereochemistry is necessary for consistency in these structures. The two most common forms of stereochemistry, tetrahedral atoms and stereochemical double bonds, are implemented in the model through simple constraints. Because of the modularity of the optimization-based approach, other types of stereochemistry, such as square planar, conjugated allenes, folding, and steric effects, could be included through similar constraints as needed.

The tetrahedral atom is depicted in Figure 2a as a central atom (not necessarily carbon) connected to four neighbors in a tetrahedral geometry. Even when it is not a chiral center, as is the case in methane where the four adjacent atoms are identical, stereochemistry still plays a role once the atoms are labeled. Of the 24 possible permutations of the four adjacent atoms, half of them will change the stereochemistry. These 12 restricted permutations are captured by constraint 7. Such a stereochemical violation results in an objective function value increase of two, which represents the notion that a bond must break and reattach on the opposite side of the central atom to carry out the implied mechanism.

Constraint 7 captures the restricted permutations as follows. Suppose that tetrahedral atom i in the reactants maps to tetrahedral atom k in the products, and the four neighboring atoms in the reactants, r_1, r_2, r_3, and r_4, map to the four neighboring atoms in the products, p_1, p_2, p_3, and p_4, as arranged in Figure 2a. Here, $\mathrm{CA}_{\mathrm{R}}^{i} = \{r_1,r_2,r_3,r_4\}$ and $\mathrm{CA}_{\mathrm{P}}^{k} = \{p_1,p_2,p_3,p_4\}$. A mapping for which the stereochemistry changes is r_1 maps to p_1, r_2 maps to p_4, r_3 maps to p_3, and r_4 maps to p_4. This permutation is captured by constraint 7 for $q = 1$, which reads:

$$\beta_{ik} \geq y_{ik} + y_{r_1p_1} + y_{r_2p_4} + y_{r_3p_3} + y_{r_4p_2} - 4$$

If this mapping were selected, all five of the binary variables on the right-hand side of this constraint would be equal to 1, requiring that $\beta_{ik} = 1$, which increases the objective function value.

Stereochemical double (or triple) bonds are characterized by four planar atoms adjacent to the multiple bond (see Figure 2b), which again is not necessarily formed by carbons. Assuming that the adjacency remains constant, there are four possible rearrangements of the connected atoms. Half of these change the stereochemistry, known as a cis/trans isomerization. The restricted permutations are captured by constraints 8 and 9 and similarly result in an objective function increase if realized.

Alternate Objective Function. The objective function 1 accounts for the formation and/or breaking of bonds during a reaction but does not distinguish between bond orders. Though this is typically sufficient for most reactions of interest, one may seek to obtain a mapping that minimizes the total number of bond order changes in the implied chemical reaction mechanism. This can be achieved by using the alternate objective function 10:

$$\begin{aligned}\zeta' = \min & \sum_{(i,j)\,\in\,\mathrm{B_R}} O_{\mathrm{R}}^{(i,j)}\Big(1 - \sum_{(k,l)\,\in\,\mathrm{B_P}} \alpha_{ijkl}\Big) \\ & + \sum_{(k,l)\,\in\,\mathrm{B_P}} O_{\mathrm{P}}^{(k,l)}\Big(1 - \sum_{(i,j)\,\in\,\mathrm{B_R}} \alpha_{ijkl}\Big) \\ & + \sum_{(i,j)\,\in\,\mathrm{B_R}} \sum_{(k,l)\,\in\,\mathrm{B_P}} |O_{\mathrm{R}}^{(i,j)} - O_{\mathrm{P}}^{(k,l)}|\alpha_{ijkl} \\ & + 2\sum_{i\,\in\,\mathrm{C_R}} \sum_{k\,\in\,\mathrm{C_P}} \beta_{ik} + 2\sum_{(i,j)\,\in\,\mathrm{D_R}} \sum_{(k,l)\,\in\,\mathrm{D_P}} \gamma_{ijkl}\end{aligned} \tag{10}$$

where $O_{\mathrm{R}}^{(i,j)}$ is the order of bond (i,j) in the reactants and $O_{\mathrm{P}}^{(k,l)}$ is the order of bond (k,l) in the products (i.e., 1 for a single bond, 2 for a double bond, 3 for a triple bond). Note that no changes are proposed for the stereochemistry terms of the objective function, since those address only single bonds.

EFFICIENT SOLUTION PROCEDURE

In this section, we present a number of techniques that can be applied to expedite the branch and bound process and obtain an optimal solution faster. While these are not required to solve the model 1−9, we find that they can dramatically reduce the solution time.

Tightening Constraints. In the context of mathematical optimization, *tightening* refers to the addition of constraints to a model that do not eliminate any integer feasible solutions, yet they eliminate fractional solutions that would otherwise be feasible in the continuous relaxation. Thus, they restrict the amount of relaxation at each node in the branch and bound tree. This leads to consistently better LP solutions at each node, allowing a faster rate of node fathoming and identification of the optimal solution after exploring fewer nodes.

Constraints 5 and 6 determine whether reactant bond (i,j) maps to product bond (k,l) by checking that reactant atom i maps to product atom k or l and reactant atom j maps to product atom k or l. These are sufficient because there is a one-to-one mapping between reactant and product atoms. Equally valid constraints 11 and 12 check that reactant atom i or j maps to product atom k and reactant atom i or j maps to product atom l.

$$\alpha_{ijkl} \leq y_{ik} + y_{jk} \quad \forall (i,j) \in \mathrm{B_R} \quad \forall (k,l) \in \mathrm{B_P} \tag{11}$$

$$\alpha_{ijkl} \leq y_{il} + y_{jl} \quad \forall (i,j) \in \mathrm{B_R} \quad \forall (k,l) \in \mathrm{B_P} \tag{12}$$

Keeping the original constraints in place, the inclusion of these additional constraints tightens the model.

Another form of tightening constraint is derived from the observation that in each summation of α_{ijkl} variables in the objective function, at most one can take the value of one. That is, if reactant bond (i,j) maps to product bond (k,l), then reactant bond (i,j) can map to no other product bond (i.e., $\Sigma_{(k,l)\in \mathrm{B_P}}\alpha_{ijkl} \leq 1\ \forall (i,j) \in \mathrm{B_R}$) and no other reactant bond can map to product bond (k,l) (i.e., $\Sigma_{(i,j)\in \mathrm{B_R}}\alpha_{ijkl} \leq 1\ \forall (k,l) \in \mathrm{B_P}$). Furthermore, conclusions can be drawn even before a bond is fully mapped. For example, if reactant atom i maps to product atom k and k is not connected to any atom in the products of type T_{R}^{j}, then reactant bond (i,j) cannot map to product bond (k,l) for any l. Considering other possibilities for this last remark and combining with the first observation leads to four additional forms of tightening constraints:

$$\sum_{(k,l)\,\in\,\mathrm{B_P}} \alpha_{ijkl} + \sum_{k\,\in\,A:T_{\mathrm{R}}^{j}\,\notin\,\mathrm{N}_{\mathrm{P}}^{k}} y_{ik} \leq 1 \quad \forall (i,j) \in \mathrm{B_R} \tag{13}$$

$$\sum_{(k,l)\,\in\,\mathrm{B_P}} \alpha_{ijkl} + \sum_{k\,\in\,A:T_{\mathrm{R}}^{i}\,\notin\,\mathrm{N}_{\mathrm{P}}^{k}} y_{jk} \leq 1 \quad \forall (i,j) \in \mathrm{B_R} \tag{14}$$

$$\sum_{(i,j)\,\in\,\mathrm{B_R}} \alpha_{ijkl} + \sum_{i\,\in\,A:T_{\mathrm{P}}^{l}\,\notin\,\mathrm{N}_{\mathrm{R}}^{i}} y_{ik} \leq 1 \quad \forall (k,l) \in \mathrm{B_P} \tag{15}$$

$$\sum_{(i,j)\,\in\,\mathrm{B_R}} \alpha_{ijkl} + \sum_{i\,\in\,A:T_{\mathrm{P}}^{k}\,\notin\,\mathrm{N}_{\mathrm{R}}^{i}} y_{il} \leq 1 \quad \forall (k,l) \in \mathrm{B_P} \tag{16}$$

It is important to note that the addition of tightening constraints increases the size of the model, so each LP may take longer to solve. We have observed that, for this model, they are consistently effective in speeding up the overall solution process, so we have opted to always include them.

Symmetry Breaking. The model may have multiple optimal solutions corresponding to the same reaction mechanism; that is, the model may exhibit some level of isomorphism. For example, in the water molecule, H_2O, the two hydrogen atoms are equivalent. Each feasible solution has a corresponding solution where the labels on the hydrogen atoms are swapped. Since the labeling is arbitrary, reshuffling the labels does not yield a distinct reaction mechanism.

Each of these equivalent mappings is the result of symmetries in the reactants or products. In principle, the identification of all symmetries can be difficult for some instances. In a subsequent section, we present a formal method to identify all symmetries; however, this can be computationally expensive in practice and may not be justified if the aim is to expedite the solution process.

There exist special cases of symmetries that we can quickly identify and readily address via the addition of simple constraints. These include isolated components, diatomics, leaf components, and simple five- and six-member rings. Here, a component takes the form AX_y, which is either a single atom ($y = 0$) or an atom connected to one or more leaf atoms of the same type. This includes popular groups such as OH, NH_2, and CH_3. Multiple instances of simple molecules such as H_2O, NH_3, CH_4, and CO_2 are covered by the isolated components case.

The basic idea for breaking these symmetries is to specify an ordering for each pair of equivalent atoms; that is, one atom in the pair is to map to an atom with a higher index than the other. To ensure consistency when symmetries exist on both sides of the reaction, we choose the atom with a higher index in the equivalent pair to map to an atom with a higher index than that which the lower indexed atom maps to in the equivalent pair. For example, the hydrogen atoms in a water molecule form an equivalent pair. With the hydrogen atoms arbitrarily labeled with indices 1 and 2, we should require that hydrogen atom 2 maps to an atom with a higher index than that which hydrogen atom 1

Table 1. Special Cases of Noncyclic Molecular Symmetries

type	representation	equivalent pairs
isolated components	$(A{:}1)X_y$, $(A{:}2)X_y$, ..., $(A{:}m)X_y$	$(i-1,i)\ \forall\ i = 2,3,...,m$
diatomics	$(A{:}1)(A{:}2)$	(1,2)
leaf components, no stereo.	$R-C-(AX_y)_m$	$(i-1,i)\ \forall\ i = 2,3,...,m$
leaf components, stereochemistry	$R-C-(AX_y)_m$	$(1,3),(i-1,i)\ \forall\ i = 3,4,...,m$

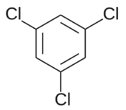


Figure 3. 1,3,5-trichlorobenzene is a six-member ring matching symmetry pattern ABABAB.

maps to. This idea extends to symmetries beyond two atoms and is summarized in Table 1 for the noncyclic special cases that we consider. We use the atom labeling notation ($A{:}m$) to indicate an atom of type A with label m.

Symmetric rings are identified by locating cycles in the molecular graph of length 5 and 6 composed of atoms of the same type. On the basis of the attachments to each atom of the cycle (which must be zero or more identical components), the pattern of the ring is determined. For example, 1,3,5-trichlorobenzene (depicted in Figure 3) is a six-member ring with the pattern ABABAB. The ring pattern is matched to one of the symmetry patterns in Table 2 (for five-member rings) or Table 3 (for six-member rings), allowing cyclic permutations and reversals. The atoms in the cycle are relabeled to match the ordering in the standard rings (Figure 4) for consistency.

Note that for leaf components, when atom C is involved in stereochemistry (tetrahedral atom or part of a stereochemical double bond), the first equivalent pair (1,2) is relaxed to (1,3) to ensure that stereochemistry is not violated. Also, for components AX_y for $y \geq 2$, the symmetry of the X atoms is additionally treated as a leaf component case.

We have developed three alternative forms for symmetry breaking constraints that enforce a consistent ordering between a pair of atoms. For reactant symmetries, any one of the constraints 17, 18, or 19 applies and should be included for each equivalent pair (i,j). Similar constraints are used to break symmetries on the product side.

$$y_{il} + y_{jk} \leq 1 \quad \forall k, l \in A : k < l \tag{17}$$

$$y_{jk} \leq \sum_{l \in A : l \leq k} y_{il} \quad \forall k \in A \tag{18}$$

$$\sum_{l \in A} l y_{il} \leq \sum_{k \in A} k y_{jk} \tag{19}$$

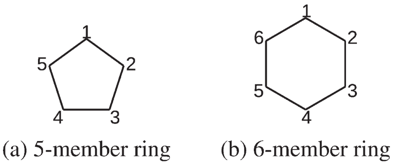
Form 17 involves a large number of simple constraints. Form 18 involves a small number of constraints with increasing density. Form 19 involves a single dense constraint. For the computational results presented in this paper, we choose to use form 19 because we have observed that it is often most effective in expediting the solution process for our reactions of interest.

Table 2. Special Cases of Five-Member Ring Symmetries

pattern	equivalent pairs
AAAAA	(1,2),(1,3),(1,4),(1,5),(2,5)
ABBBB	(2,5)
ABCCB	(2,5)

Table 3. Special Cases of Five-Member Ring Symmetries

pattern	equivalent pairs
AAAAAA	(1,2),(1,3),(1,4),(1,5),(1,6),(2,6)
ABABAB	(1,3),(1,5),(2,6)
ABCCBA	(2,5)
ABBBBA	(2,5)
ABAABA	(1,3),(2,5)
ABCABC	(1,4)
ABCDCB	(2,6)
ABCACB	(2,6)
ABBBBB	(2,6)
ABBCBB	(2,6)



(a) 5-member ring (b) 6-member ring

Figure 4. Numbering for symmetric rings.

Branching Priorities. The order in which the binary variables are selected for branching in the branch and bound process is important because, although it does not affect the theoretical guarantee of obtaining an optimal solution, it can greatly influence the extent of the search required. Observe that the y_{ik} variables are the only true decision variables in our model, as they encode the actual mapping. The other variables (α_{ijkl}, β_{ik}, and γ_{ijkl}) are auxiliary and are used to count the number of bonds that break and form. In fact, it can be shown that, given an assignment of binary y_{ik} variables, the values of the other variables are uniquely defined. The implication is that branching a y_{ik} variable resolves a greater amount of uncertainty in the solution than branching another variable. Thus, the y_{ik} variables take branching priority over the other variables. For similar reasons, we have observed that, within the y_{ik} variables, it is beneficial to prioritize branching based on valence (i.e., the number of neighbors). Let the valence of reactant atom i be V_R^i and the valence of product atom k by V_P^k. For two variables y_{ik} and y_{jl} for which $V_R^i + V_P^k > V_R^j + V_P^l$, variable y_{ik} should take branching priority over variable y_{jl}.

As a consequence, the variables α_{ijkl}, β_{ik}, and γ_{ijkl} would become branching candidates only after all y_{ik} variables have been determined. But, on the basis of the previous discussion, by that point the other variables would have also been determined and the solution would be integral, so no additional branching will occur. Therefore, the variables α_{ijkl}, β_{ik}, and γ_{ijkl} do not need to be declared as binary variables and can simply be declared as continuous variables on [0,1].

Implicit Hydrogen Treatment. In contrast to other approaches, our method explicitly takes hydrogen atoms into account. As a result, the size of the MILP can become large for reactions involving many atoms. On the basis of the observation that hydrogen atoms appear only as leaf nodes in a molecular graph, we can treat hydrogen atoms in our model implicitly. For each reactant atom i and each product atom k, let H_R^i and H_P^k be the number of attached hydrogen atoms, respectively. In calculating an optimal reaction mapping, we assume that bonds to hydrogen atoms are preserved whenever possible. Therefore, when reactant atom i maps to product atom k, the difference in the number of attached hydrogen atoms, $|H_R^i - H_P^k|$, is the number of bonds that must be broken or formed to account for the hydrogen atoms. We incorporate this expression into a modified objective function 20:

$$\begin{aligned}\zeta = \min &\sum_{(i,j)\,\in\,B_R}\Big(1 - \sum_{(k,l)\,\in\,B_P}\alpha_{ijkl}\Big)\\ &+ \sum_{(k,l)\,\in\,B_P}\Big(1 - \sum_{(i,j)\,\in\,B_R}\alpha_{ijkl}\Big)\\ &+ \sum_{i\,\in\,A}\sum_{k\,\in\,A}|H_R^i - H_P^k|y_{ik} + 2\sum_{i\,\in\,C_R}\sum_{k\,\in\,C_P}\beta_{ik}\\ &+ 2\sum_{(i,j)\,\in\,D_R}\sum_{(k,l)\,\in\,D_P}\gamma_{ijkl} + |(H_2)_R - (H_2)_P| \qquad (20)\end{aligned}$$

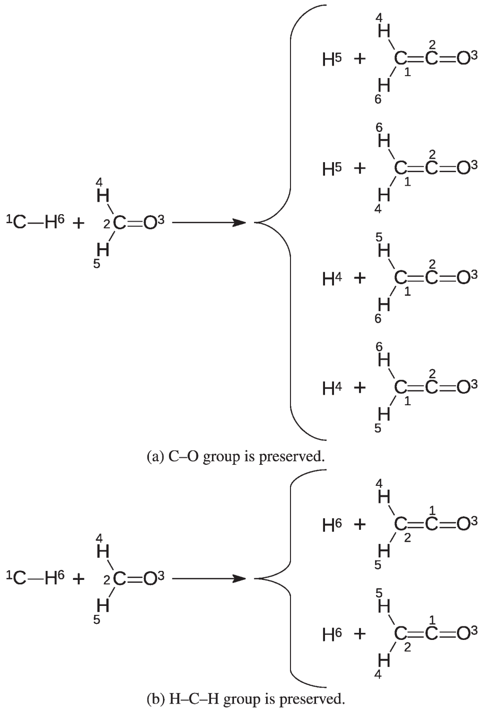
where we have removed the hydrogen atoms from the set of atoms, A, and updated the sets of bonds, B_R and B_P, accordingly. The constant term $|(H_2)_R - (H_2)_P|$ refers to the difference in number of diatomic hydrogens in the reactants, $(H_2)_R$, and products, $(H_2)_P$, which is an additional number of bonds that must break or form.

Minor changes must also be made to the stereochemistry constraints. For tetrahedral atoms with two or more attached hydrogens, constraint 7 is removed, since the mapping of the hydrogen atoms will determine the stereochemistry. Similarly, for stereochemical double bonds with two hydrogens attached to the same atom, constraints 8 and 9 are omitted. In the remaining stereochemistry constraints, binary variables that refer to hydrogen atoms are replaced with the value 1, which assumes that the hydrogens are mapped appropriately.

With the implicit treatment of hydrogen atoms, the computational complexity of the model is decreased without compromising solution accuracy. Though the value of the objective function reflects the total number of bonds that must break and form, the model no longer provides a mapping for the hydrogen atoms. Hydrogen atoms are optimally mapped using the following simple postprocessing procedure: (1) For each mapped atom, the attached hydrogens in common are mapped to each other. (2) Isolated hydrogen atoms on both sides of the reaction are mapped to each other. (3) Diatomic hydrogen on both sides of the reaction are mapped to each other. (4) The remaining hydrogen atoms have bonds that broke or formed and are mapped arbitrarily. (5) Stereochemistry constraints are checked for violations, and hydrogen atom mappings are switched where possible to remove the violations.

MULTIPLE REACTION MECHANISMS

Our optimization model will generally have multiple feasible solutions, which include the optimal one. Each solution is a mapping corresponding to a potential reaction mechanism. In this section, we describe a procedure for locating multiple mechanisms for a chemical reaction.



(a) C–O group is preserved.

(b) H–C–H group is preserved.

Figure 5. All optimal mappings for the elementary reaction CH + CH_2O → H + CH_2CO.

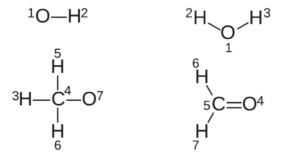
Locating Multiple Solutions.

With an approach based on integer linear optimization, one can locate more than one solution to the model. In fact, one can locate all integer-feasible solutions when practical (e.g., no memory limitations). One technique to enumerate multiple solutions is through the use of *integer cuts*, which is the iterative application of a constraint that eliminates only a single solution. This is accomplished by enforcing the Hamming distance for this undesirable solution to be greater than or equal to one (i.e., if $\bar{y}_{ik}$ is a known solution to the model, it can be eliminated by appending constraint 21 and solving the model again to obtain an additional solution). This can be repeated until the model becomes integer infeasible and no further solution can be found. The drawback of this technique is that the computational burden scales linearly with the number of additional solutions sought.

$$\sum_{i,k\,\in\,A:\bar{y}_{ik}=0} y_{ik} + \sum_{i,k\,\in\,A:\bar{y}_{ik}=1}(1 - y_{ik}) \geq 1 \qquad (21)$$

Fortunately, an alternative, more efficient approach exists. Rather than having to restart the solution process each time an integer cut is added, the existing search tree can continue to be explored by relaxing the fathoming criteria through a technique known as *solution pool*. For the sake of brevity, we will omit implementation details, but the reader should be aware that modern MILP solvers readily provide access to such a solution pool facility.

An example with multiple solutions is the elementary reaction CH + CH_2O → H + CH_2CO, illustrated in Figure 5. There exist a total of six mappings with the optimal objective function value of $\zeta = 4$, corresponding to two distinct mechanisms. In particular, the four mappings in Figure 5a correspond to a mechanism in which the C−O group is preserved. These mappings are equivalent and are the result of swapping the hydrogen atoms



(a) Reactant labeling. (b) Product labeling.

Figure 6. Arbitrary labeling of the reactant and product atoms for the elementary reaction $OH + CH_3O \rightarrow H_2O + CH_2O$.

Table 4. Equivalent Permutations of Reactants $OH + CH_3O$ and Products $H_2O + CH_2O^a$

reactants			products			
A	B	C	A	B	C	D
1	1	1	1	1	1	1
2	2	2	2	3	3	2
3	5	6	3	2	2	3
4	4	4	4	4	4	4
5	6	3	5	5	5	5
6	3	5	6	6	7	7
7	7	7	7	7	6	6

[a] Note: permutation A in each case corresponds to the initial atomic labeling.

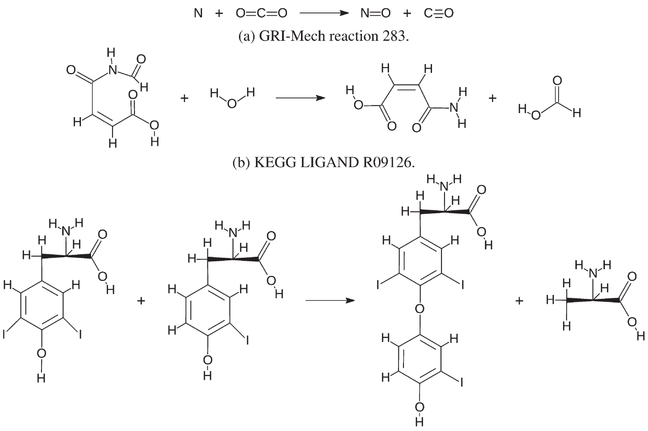
in reactant CH_2O and/or swapping the hydrogen atoms in product CH_2CO. In addition, the two mappings in Figure 5b are equivalent to each other and correspond to a different mechanism where the H—C—H group is preserved.

In the absence of symmetry, each solution would be a distinct reaction mechanism. In the next section, we discuss how to eliminate all symmetries in the reaction. Thus, for the first time in the open literature, we present a method for the automated identification of all distinct mechanisms for a chemical reaction.

Eliminating Equivalent Mappings. In order to locate distinct reaction mechanisms, we need to identify all symmetries in the reactants and products. A method to describe all molecular symmetries based on graph naming has been previously developed by Chen et al.[16] Here, we present an alternative approach that leverages our optimization model and provides symmetries in a form that can readily be used to filter out equivalent mappings.

Symmetries in the reactants and products exist when atom labels can be rearranged without changing the adjacency of any labels. For example, if the atoms in water were labeled as (H:1)—(O:3)—(H:2), the labels of hydrogen atoms 1 and 2 could be swapped and atoms 1 and 3 would still be adjacent, as would atoms 2 and 3. Therefore, the hydrogen atoms are symmetric and the rearrangement of the atom labels (1, 2, 3) to (2, 1, 3) is an equivalent permutation.

To calculate all equivalent permutations of the reactant atoms, an identity reaction is constructed with the reactants on both sides, that is, the "dummy" reaction where reactants go to reactants. The identity reaction is mapped without including any symmetry breaking constraints using a full solution pool to find all mappings with an objective value $\zeta = 0$, which corresponds to no change in atomic adjacency. Similarly, all equivalent permutations of the product atoms are identified using an identity reaction with the products on both sides.



(a) GRI-Mech reaction 283.

(b) KEGG LIGAND R09126.

(c) BioPath RXN00054.

Figure 7. Sample reactions from each database.

Table 5. Reaction Databases and Mapping Results

database	number of reactions	median time (sec)	median number of atoms	median objective value
GRI-Mech	325	$<10^{-2}$	5	2
KEGG LIGAND	6208	0.43	82	5
BioPath	1545	0.44	72	6

Table 6. Distribution of Computational Times

	mapped within specified time			
database	1 s	1 min	1 h	1 day
GRI-Mech	100%	100%	100%	100%
KEGG LIGAND	66.4%	95.0%	99.0%	99.7%
BioPath	57.6%	86.0%	97.8%	99.2%

Once we have identified all equivalent permutations of the reactant and product atoms, we can then locate distinct mechanisms for the chemical reaction. Multiple mappings are generated using the solution pool approach described in the previous section (symmetry breaking constraints may be kept in place). We determine whether two mappings are equivalent by considering all possible pairs of the I_R equivalent reactant permutations with the I_P equivalent product permutations. A mapping is transformed using these permutations by relabeling the reactant and product atoms. Equivalent mappings that correspond to the same reaction mechanism are removed by transforming each mapping over all $I_R \times I_P$ permutation pairs and checking whether the transformation results in another mapping. Potential mechanisms could also correspond to suboptimal solutions, and when these are included in the solution pool, mappings with different objective values need not be checked for equivalence.

This procedure is demonstrated with the elementary reaction $OH + CH_3O \rightarrow H_2O + CH_2O$ where the reactant and product atoms are labeled arbitrarily as indicated in Figure 6. The identity reactions locate $I_R = 3$ equivalent permutations of the reactant atoms and $I_P = 4$ equivalent permutations of the product atoms, enumerated in Table 4. A solution pool of the reaction identifies two optimal mappings: (1,2,3,4,5,6,7) ↔ (1,2,6,5,3,7,4) and (1,2,3,4,5,6,7) ↔ (1,2,3,5,6,7,4). The second mapping can be

Table 7. Distinct Mechanisms for Reactions in the GRI-Mech Database

	number of distinct mechanisms per reaction		number of reactions with specified number of distinct mechanisms				
deviation from optimal objective value	mean	std. dev.	0	1	2	3	≥4
0	1.08	0.31	0	304	17	4	0
+2	1.16	2.08	159	97	29	13	27
+4	2.67	8.40	181	49	38	13	44

transformed into the first using reactant transformation B and product transformation D, as listed in the table. Thus, the mappings are equivalent and correspond to a single reaction mechanism.

COMPUTATIONAL RESULTS

We applied our approach to three databases of chemical reactions. GRI-Mech 3.0[17] is a database of 325 elementary reactions involving small molecules relevant to gas combustion. The KEGG LIGAND release 29.0 of 2004[18] contains 8447 reactions found in biological pathways, 6208 of which were available for study after those that were unbalanced or without species data were removed. The BioPath database version 1.0 of 2006[19] is comprised of 1545 biochemical reactions and includes manually annotated mappings. Sample reactions from each database are presented in Figure 7.

The model was solved for each reaction from the three databases using ILOG CPLEX Optimizer 12.1[20] running as a single thread on an Intel Nehalem 2.66 GHz processor with 3 GB of available memory. We included the hydrogen atoms that were omitted in the KEGG LIGAND database and used the annotated stereochemistry provided for each species. The stereochemistry could also have been inferred directly from three-dimensional structure input. Note that 85% of the reactions studied included some form of stereochemistry. The computational results are summarized in Table 5. Most reactions were mapped within seconds, though some reactions took longer (see distribution in Table 6). Out of a total of 8078 reactions, only 32 reached the wallclock limit of 24 h, and thus the optimality of these mappings was not guaranteed. Solutions to these problems could be obtained with a parallel implmenetation of the MILP algorithm.

For the GRI-Mech 3.0 database, a computational study was performed with full solution pool. The search was limited to solutions with up to four bond changes more than optimal. The total computational time, using the configuration described previously, was 17 s. Most reactions had a single mapping at the optimal objective value, though some reactions had up to 8. We applied our automated procedure for detecting and eliminating equivalent mappings to identify distinct mechanisms for each reaction. An example of a reaction from this database with multiple distinct mechanisms at the optimal objective value is depicted in Figure 5. The results are grouped by deviation from the optimal objective value and presented in Table 7.

We also studied the effect of the alternate objective function 10, which minimizes the total number of bond order changes, by studying a subset 71 of reactions from the biochemical databases that included all three of single, double, and triple bonds. For all but four of these reactions, the alternate objective function resulted in the same solution as the original objective function.

CONCLUSIONS

For the first time, we have presented an automated approach to the reaction mapping problem based on integer linear optimization. It has the advantages of providing stereochemical consistency and locating multiple reaction mappings. Using an automated method for the detection of symmetries and elimination of equivalent mappings, we have presented for the first time in the open literature a method capable of identifying all distinct reaction mechanisms. Our methods utilize efficient branch and bound algorithms for integer linear optimization that locate guaranteed optimal solutions. We have demonstrated our techniques by studying several databases of reactions, including those important to gas combustion and biochemical pathways.

While minimizing the number of bonds that break and form is a reasonable choice for the objective function, particularly for elementary reactions, and is consistent with previous work, other objective functions could be substituted into the model. We have formulated and demonstrated an alternate objective function that penalizes changes in bond orders. Other objective functions could weigh each broken bond by its dissociation energy or consider electron rearrangement. In a postprocessing method, energetic calculations could also be used to select the best mappings from a solution pool. Because of its modularity, our integer linear optimization model is also receptive to more complex enhancements, such as the incorporation of additional types of chemistry, including radicals, charges, and energy states.

A Web tool called DREAM (Determination of REAction Mechanisms) has been developed that implements the reaction mapping approach described in this paper. DREAM enables a user to submit chemical reactions of interest in a variety of formats to be processed by our method. It provides a choice of objective functions and the option of identifying either a single mapping or multiple mechanisms. DREAM is freely available to the scientific community at http://selene.princeton.edu/dream/

AUTHOR INFORMATION

Corresponding Author

*E-mail: floudas@titan.princeton.edu.

ACKNOWLEDGMENT

The authors gratefully acknowledge support from the National Science Foundation (NSF EFRI-0937706). E.L.F. is also thankful for his National Defense Science and Engineering Graduate (NDSEG) fellowship. The authors would also like to thank Oliver Sacher from Molecular Networks GmbH for providing the BioPath database in machine-readable format.

REFERENCES

(1) Westerberg, K. M.; Floudas, C. A. Locating all transition states and studying the reaction pathways of potential energy surfaces. *J. Chem. Phys.* **1999**, *110*, 9259–9295.

(2) Westerberg, K. M.; Floudas, C. A. Dynamics of Peptide Folding: Transition States and Reaction Pathways of Solvated and Unsolvated Tetra-Alanine. *J. Global Optim.* **1999**, *15*, 261–297.

(3) First, E. L.; Gounaris, C. E.; Wei, J.; Floudas, C. A. Computational characterization of zeolite porous networks: an automated approach. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17339–17358.

(4) Lynch, M. F.; Willett, P. The Automatic Detection of Chemical Reaction Sites. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 154–159.

(5) McGregor, J. J.; Willett, P. Use of a maximum common subgraph algorithm in the automatic identification of ostensible bond changes occurring in chemical reactions. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 137–140.

(6) McGregor, J. J. Backtrack search algorithms and the maximal common subgraph problem. *Softw. Pract. Exper.* **1982**, *12*, 23–34.

(7) Raymond, J. W.; Willett, P. Maximum common subgraph isomorphism algorithms for the matching of chemical structures. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 521–533.

(8) Raymond, J. W.; Gardiner, E. J.; Willett, P. RASCAL: Calculation of Graph Similarity using Maximum Common Edge Subgraphs. *Comput. J.* **2002**, *45*, 631–644.

(9) Körner, R.; Apostolakis, J. Automatic Determination of Reaction Mappings and Reaction Center Information. 1. The Imaginary Transition State Energy Approach. *J. Chem. Inf. Model.* **2008**, *48*, 1181–1189.

(10) Apostolakis, J.; Sacher, O.; Körner, R.; Gasteiger, J. Automatic Determination of Reaction Mappings and Reaction Center Information. 2. Validation on a Biochemical Reaction Database. *J. Chem. Inf. Model.* **2008**, *48*, 1190–1198.

(11) Akutsu, T. Efficient Extraction of Mapping Rules of Atoms from Enzymatic Reaction Data. *J. Comput. Biol.* **2004**, *11*, 449–462.

(12) Crabtree, J. D.; Mehta, D. P. Automated reaction mapping. *ACM J. Exp. Algor.* **2009**, *13*, Article 1.15.

(13) Crabtree, J. D.; Mehta, D. P.; Kouri, T. M. An Open-Source Java Platform for Automated Reaction Mapping. *J. Chem. Inf. Model.* **2010**, *50*, 1751–1756.

(14) Kouri, T.; Mehta, D. In *Experimental Algorithms*; Pardalos, P. M., Rebennack, S., Eds.; Springer: Berlin, Germany, 2011; Vol. 6630, pp 157—168.

(15) Floudas, C. A. *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*; Oxford University Press: Oxford, U.K., 1995.

(16) Chen, W.; Huang, J.; Gilson, M. K. Identification of Symmetries in Molecules and Complexes. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1301–1313.

(17) Smith, G. P.; Golden, D. M.; Frenklach, M.; Moriarty, N. W.; Eiteneer, B.; Goldenberg, M.; Bowman, C. T.; Hanson, R. K.; Song, S.; Gardiner, Jr., W. C.; Lissianski, V. V.; Qin, Z. GRI-Mech 3.0. http://www.me.berkeley.edu/gri_mech/ (accessed Apr 18, 2011).

(18) Goto, S.; Okuno, Y.; Hattori, M.; Nishioka, T.; Kanehisa, M. LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.* **2002**, *30*, 402–404.

(19) Reitz, M.; Sacher, O.; Tarkhov, A.; Trümbach, D.; Gasteiger, J. Enabling the exploration of biochemical pathways. *Org. Biomol. Chem.* **2004**, *2*, 3226–3237.

(20) *ILOG CPLEX Optimizer 12.1.0*; IBM Corp.: Armonk, NY, 2009.