






Article

Partial Imaginary Transition State (ITS) Graphs: A Formal Framework for Research and Analysis of Atom-to-Atom Maps of Unbalanced Chemical Reactions and Their Completions

Marcos E. González Laffitte ^{1,2,*}, Klaus Weinbauer ^{1,3}, Tieu-Long Phan ^{1,4}, Nora Beier ¹, Nico Domschke ¹, Christoph Flamm ⁵, Thomas Gatter ^{1,2}, Daniel Merkle ^{4,6} and Peter F. Stadler ^{1,2,5,7,8,9,10,*}

- ¹ Bioinformatics Group, Department of Computer Science & Interdisciplinary Center for Bioinformatics, Leipzig University, Härtelstraße 16-18, D-04107 Leipzig, Germany; klaus@bioinf.uni-leipzig.de (K.W.); tieu@bioinf.uni-leipzig.de (T.-L.P.); nora@bioinf.uni-leipzig.de (N.B.); dnico@bioinf.uni-leipzig.de (N.D.); thomas@bioinf.uni-leipzig.de (T.G.)
- ² Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Leipzig University, D-04103 Leipzig, Germany
- ³ Machine Learning Research Unit, TU Wien Informatics, Erzherzog-Johann-Platz 1 (FB02), A-1040 Wien, Austria
- ⁴ Department of Mathematics and Computer Science, University of Southern Denmark, DK-5230 Odense, Denmark; daniel@imada.sdu.dk
- ⁵ Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Vienna, Austria; xtof@tbi.univie.ac.at
- ⁶ Algorithmic Cheminformatics Group, Faculty of Technology, Bielefeld University, 33615 Bielefeld, Germany
- ⁷ Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany
- ⁸ Center for Non-Coding RNA in Technology and Health, University of Copenhagen, DK-1870 Fredriksberg, Denmark
- ⁹ Facultad de Ciencias, Universidad Nacional de Colombia, Sede Bogotá, Bogotá CO-111321, Colombia
- ¹⁰ Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA
- * Correspondence: marcos@bioinf.uni-leipzig.de (M.E.G.L.); studla@bioinf.uni-leipzig.de (P.F.S.); Tel.: +49-431-97-16690 (P.F.S.)



Citation: González Laffitte, M.E.; Weinbauer, K.; Phan, T.-L.; Beier, N.; Domschke, N.; Flamm, C.; Gatter, T.; Merkle, D.; Stadler, P.F. Partial Imaginary Transition State (ITS) Graphs: A Formal Framework for Research and Analysis of Atom-to-Atom Maps of Unbalanced Chemical Reactions and Their Completions. *Symmetry* **2024**, *16*, 1217. <https://doi.org/10.3390/sym16091217>

Academic Editor: Zhibin Du

Received: 30 July 2024

Revised: 28 August 2024

Accepted: 5 September 2024

Published: 16 September 2024

Abstract: Atom-to-atom maps (AAMs) are bijections that establish the correspondence of reactant and product atoms across chemical reactions. They capture crucial features of the reaction mechanism and thus play a central role in modeling chemistry at the level of graph transformations. AAMs are equivalent to so-called “imaginary transition state” (ITS) graphs, making it possible to reduce tasks such as the computational comparison of AAMs to testing graph isomorphisms. In many application scenarios, nonetheless, only partial information is available, i.e., only partial maps or, equivalently, only subgraphs of the ITS graphs, are known. Here, we investigate whether and how, and to what extent, such partial chemical data can be completed and compared. The focus of this contribution is entirely on the development of a solid mathematical foundation for the analysis of partial AAMs and their associated partial ITS graphs.

Keywords: chemical graphs; cheminformatics; chemical reaction mechanisms; atom-to-atom maps; imaginary transition state; condensed graph of the reaction; (sub)graph isomorphism; (chemical) graph alignment; rebalancing of reactions; tautomerization reactions; hydrogen atom mapping



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Chemical reactions, by definition, comprise the rearrangement of bonds among the atoms of a set of reactant molecules. That is, chemical reactions preserve the atoms involved. This fundamental property is formalized by an atom-to-atom map (AAM). Mathematically, an AAM is simply a bijection between corresponding atoms in the reactants and products that preserves atom types. From a chemical point of view, the AAM encodes the mechanism of the reaction by implying the bonds that are formed and broken, thereby providing a condensed description of the transition from reactants to products.

Despite the fundamental conceptual importance of AAMs in chemical reactions, and its many practical applications, AAMs constitute data that are not commonly available. Large-scale reaction databases such as Reaxys[®] [1] typically store each reaction as a pair comprising the (multi)sets of reactant and product molecules, respectively, recording with a stoichiometric coefficient whose molecule appears more than once on either side of the reaction. The computational reconstruction of AAMs, on the other hand, turns out to be a highly non-trivial problem [2,3]. Following the development of combinatorial approaches maximizing common subgraphs or minimizing chemical distance [4,5], several machine learning tools have recently become available for this purpose; see refs. [6–10]. While most tools report bijections between non-hydrogen atoms for balanced reactions, usually only partial maps are reported for unbalanced, i.e., incomplete reaction data. Moreover, some tools deliberately focus on the reaction center only [11]. This raises the question of if and when such partial AAM information can be extended to a complete description of a reaction.

The task of benchmarking tools that compute AAMs and/or perceive reaction centers turns out to be more challenging than expected because it requires a precise comparison of AAMs or parts thereof. The difficulty arises from the fact that different tools report (partial) AAMs using essentially arbitrary numberings of reactant and product atoms. Hence, the problem is not the comparison of two AAMs on the same pair of reactant and product graphs but on isomorphic pairs. For balanced reactions, some of us showed in [12] that the problem reduces to determining a solution for an isomorphism search between certain graphs, in particular, the Imaginary Transition State (ITS) graph of a reaction. This graph is obtained by “superimposing” the reactant and product molecules over corresponding pairs of atoms obtained with respect to an AAM and assigning special labels to the bonds that change during the reaction. This graph representation of a reaction was introduced by Shinsaku Fujita [13] and Wilcox and Levinson [14] with the aim of facilitating storage, retrieval, and processing of chemical information. Essentially the same construction was later proposed as Condensed Graph of the Reaction (CGR) for machine learning applications [15]. ITS graphs and AAMs are equivalent in that one can be constructed uniquely from the other. As a consequence, two AAMs are equivalent if and only if their ITS graphs are isomorphic [12].

In this contribution, we are interested in situations where only partial information on AAMs is available, naturally leading to the concept of partial ITS graphs. This is particularly the case for unbalanced reaction data. Nonetheless, it should be noted first that even in balanced reactions, however, it becomes more difficult to formalize the equivalence of partial information. In addition, partial AAMs that are not defined on isomorphic subgraphs may still provide consistent information. The situation is aggravated if the reaction data are not balanced. A large fraction of the millions of chemical reactions stored in large-scale databases such as Reaxys[®] [1] or (USPTO) (United States Patent and Trademark Office) [16] omits co-reactants or co-products. While these are usually “obvious” to a trained chemist, these omissions are a serious problem for computational methods, limiting the use of the data, particularly for machine learning applications. Several methods to alleviate the problem have been devised to rebalance reactions [17–19], highlighting the difficulty of this problem.

Partial ITS graphs that completely cover the reaction center, moreover, encode the “reaction pattern”, i.e., a combination of (a) the prerequisites and adjacent context of the reaction in terms of functional groups and their structural arrangement, and (b) the bonds that change in the transition from reactants to products. Thus, there is a 1–1 correspondence; furthermore, between such “good” reaction patterns and graph rewriting rules that implement reaction mechanisms in the so-called Double Pushout (DPO) graph grammars, a powerful tool to model large reaction networks at the atomic level, and which is part of our ongoing research [20,21]. In this context, a key open problem is the inference of comprehensive sets of rules, i.e., the determination of suitable partial ITS graphs from sets of ITS graphs for reactions that share the same mechanism.

The framework presented here is based on the assumption that there is a 1–1 correspondence between molecules and their graph-theoretic representation, i.e., their structural formula. This is an approximation, albeit an extremely useful one. Mesomerism and tautomerism, however, introduce further ambiguities. While mesomeric formulas [22,23] describe the same molecule by equivalent but distinct labeled graphs that represent delocalized electron systems, e.g., by different sequences of single and double bonds, the situation is more complicated for tautomers. Tautomers are distinct molecules that are related by well-defined intra-molecular chemical reactions that are fast enough to establish an equilibrium between the alternatives [24,25]. This begs the question of whether such equivalence classes of molecules or reactions can be described and analyzed with a similar formalism.

Reaction rebalancing, the construction of AAMs, reaction center perception, the extraction of reaction rules or patterns from reaction data, as well as the classification of reactions, are problems that share partial bijections and concepts of equivalence, consistency, isomorphism, and extensions, and also convey algorithmic problems concerning the comparison of associated graph representations. The purpose of this contribution is to provide a solid mathematical foundation by investigating the properties of partial AAMs and their associated partial ITS graphs and to clarify the relationship with conceptually similar structures, in particular the pairwise graph alignments explored in [26].

This contribution is organized as follows: we first introduce AAMs and ITS graphs as mathematical objects and fix the notation that will be used throughout. In Section 3, we present the partial ITS graphs by considering subgraphs of ITS graphs and the corresponding incomplete AAMs. The focus of this section is on the conditions required for completing such partial structures to full AAMs and complete ITS graphs, as well as on their connection to graph alignments. In Section 4, we introduce the notions of equivalence and consistency of partial AAMs and study the connection to isomorphisms of the corresponding partial ITS graphs. We proceed in Section 5 to generalize these concepts to tautomers, showing that more general maps take the place of isomorphisms. We complete our contribution with a short overview of the issues caused by hydrogen atoms, which are often omitted in chemical structural formulae. The concluding remarks focus on important open questions.

2. Preliminaries

2.1. Notation

For standard notation and terminology in Graph Theory, we reference [27]. Molecules are represented as vertex and edge labeled (simple) graphs, i.e., without loops or multiple edges. For a graph G , we write $V(G)$ and $E(G)$ for its vertex and edge sets, and $a_G : V(G) \rightarrow L_v$ and $b_G : E(G) \rightarrow L_e$ for the labeling function on the vertices and edges, respectively. We choose L_v and L_e to be non-empty and disjoint. Moreover, we reserve the symbol $\emptyset \notin L_v \cup L_e$. Given vertices x and y of G , if there is an edge $e \in E(G)$ joining them, we say that x and y are *adjacent* to each other and *incident* with e . As is customary, we write $e = xy$ throughout. Two labeled graphs G and H are said to be *isomorphic* if there exists a bijection $\varphi : V(G) \rightarrow V(H)$, preserving adjacency, and vertex and edge labels, i.e., preserving edges and non-edges, such that $a_H(\varphi(x)) = a_G(x)$ for all $x \in V(G)$, and $b_H(\varphi(x)\varphi(y)) = b_G(xy)$ for every $xy \in E(G)$. We say that φ is an *isomorphism* from G to H and write $G \cong H$. On the other hand, φ is said to be an *automorphism* of G if it is an isomorphism from G to G itself. In general, the composition or concatenation of maps $\alpha : X \rightarrow Y$ and $\beta : Y \rightarrow Z$ will be written $\beta \circ \alpha(x) := \beta(\alpha(x))$.

A chemical reaction will be considered as a pair of graphs $G \longrightarrow H$. Connecting to the theory of chemical reaction networks [28,29], the vertex- and edge-labeled graphs G and H represent the so-called “complexes” of the reaction. In this setting, the connected components of G and H are the reactant and product molecules, respectively.

2.2. Atom-to-Atom Map, ITS Graph, and Reaction Center

A chemical reaction $G \longrightarrow H$ is *balanced* if for every reactant atom $x \in V(G)$, there is a (unique) corresponding product atom $\alpha(x) \in V(H)$, and vice versa.

The map $\alpha : V(G) \rightarrow V(H)$ is usually called the atom-to-atom map (AAM) of the reaction. More formally, we have the following.

Definition 1. Let G and H be two labeled graphs. A map $\alpha : V(G) \rightarrow V(H)$ is an AAM if it is a bijection and preserves vertex labels, i.e., if $a_H(\alpha(x)) = a_G(x)$ for all $x \in V(G)$.

The concept of an imaginary transition state (ITS) representing the superposition of reactants and products as well as the changes in bonds was introduced by Shinsaku Fujita [13] and Wilcox and Levinson [14]:

Definition 2. Let $\alpha : V(G) \rightarrow V(H)$ be an AAM for two labeled graphs G and H . Then, the imaginary transition state (ITS) $Y := Y(G, H, \alpha)$ is the graph defined by all the following:

- (i) There is a bijection $\eta : V(Y) \rightarrow V(G)$,
- (ii) For $x, y \in V(Y)$ we have $xy \in E(Y)$ iff $\eta(x)\eta(y) \in E(G)$ or $\alpha(\eta(x))\alpha(\eta(y)) \in E(H)$,
- (iii) $x \in V(Y)$ is labeled by the ordered pair $a_Y(x) = (a_G(\eta(x)), a_H(\alpha(\eta(x))))$ and $xy \in E(Y)$ is labeled by the pair

$$b_Y(xy) = \begin{cases} (b_G(\eta(x)\eta(y)), b_H(\alpha(\eta(x))\alpha(\eta(y)))) & \text{if } \eta(x)\eta(y) \in E(G) \text{ and } \alpha(\eta(x))\alpha(\eta(y)) \in E(H) \\ (b_G(\eta(x)\eta(y)), \emptyset) & \text{if } \eta(x)\eta(y) \in E(G) \text{ and } \alpha(\eta(x))\alpha(\eta(y)) \notin E(H) \\ (\emptyset, b_H(\alpha(\eta(x))\alpha(\eta(y)))) & \text{if } \eta(x)\eta(y) \notin E(G) \text{ and } \alpha(\eta(x))\alpha(\eta(y)) \in E(H) \end{cases}$$

Note that by definition $\eta' = \alpha \circ \eta$ is a bijection $\eta' : V(Y) \rightarrow V(H)$. It should be noted as well that η is not unique in general but only determined up to automorphisms of G and H . Since α is an AAM for G and H , we note $a_H(\eta'(x)) = a_H(\alpha(\eta(x))) = a_G(\eta(x))$. Given an arbitrary ITS graph Y , we shall write $a_Y(x) = (a_Y^1(x), a_Y^2(x))$ for the vertex labels, and $b_Y(xy) = (b_Y^1(xy), b_Y^2(xy))$ for the edge labels of Y . The coordinates $a_Y^i(x)$ and $b_Y^i(xy)$, with $i = 1, 2$, are variables that take specific values determined by both the presence and labels of the vertices and edges of the graphs in the underlying reaction $G \rightarrow H$ from which Y is derived. For a concrete example of the usage of this notation, see Figure 1. In the literature, the vertex sets $V(Y)$, $V(G)$ and/or $V(H)$ are often identified to simplify the notation. We will, however, need the notational distinction between these three sets.

The ITS graph is closely related to earlier matrix representations of chemical reactions [30]. In the context of machine learning applications, essentially the same construction has appeared more recently as Condensed Graph of the Reaction (CGR) [15]. The notion of a reaction center arose for example, in [31–33].

Definition 3. Let $\alpha : V(G) \rightarrow V(H)$ be an AAM for two labeled graphs G and H and let $Y := Y(G, H, \alpha)$ be the corresponding ITS. The reaction center $\Gamma := \Gamma(G, H, \alpha)$ is the subgraph of Y induced by the reaction edges, i.e., edges $xy \in E(Y)$ with $b_Y^1(xy) \neq b_Y^2(xy)$ and comprising only the reaction vertices incident with these edges.

We say that $xy \in E(G)$ is a reaction edge in G if $\eta^{-1}(x)\eta^{-1}(y)$ is a reaction edge in $Y(G, H, \alpha)$, and similarly for edges in H with respect to $\eta^{-1} \circ \alpha^{-1}$. Denote by \hat{G}_α , \hat{H}_α , and \hat{Y} the subgraphs of G , H , and Y obtained by removing all the reaction edges in these graphs, i.e., all those edges that correspond to reaction edges in Y as defined above.

Lemma 1. If $\alpha : V(G) \rightarrow V(H)$ is an AAM for two labeled graphs G and H , then α is an isomorphism from \hat{G}_α to \hat{H}_α .

Proof. First note that $V(G) = V(\hat{G}_\alpha)$ and $V(H) = V(\hat{H}_\alpha)$. Let Y and η be as in Definition 2. By definition $xy \in E(\hat{G}_\alpha)$ is not a reaction edge, i.e., $b_Y^1(\eta^{-1}(x)\eta^{-1}(y)) = b_Y^2(\eta^{-1}(x)\eta^{-1}(y))$, which can only happen if $b_Y^2(\eta^{-1}(x)\eta^{-1}(y)) \neq \emptyset$ by the selection of the label sets, implying $\alpha(x)\alpha(y) \in E(H)$ with $b_G(xy) = b_H(\alpha(x)\alpha(y))$, and thus $\alpha(x)\alpha(y) \in E(\hat{H}_\alpha)$. Similarly, any edge $uv \in E(\hat{H}_\alpha)$ is not a reaction edge and thus $\alpha^{-1}(u)\alpha^{-1}(v) \in E(\hat{G}_\alpha)$ and $b_G(\alpha^{-1}(u)\alpha^{-1}(v)) = b_H(uv)$.

Hence, α is a bijection, preserves the adjacency of vertices in \hat{G}_α , \hat{H}_α , respectively, and preserves vertex and edge labels. Therefore, α is an isomorphism. \square

In addition to the doubling of the labels, moreover, \hat{G}_α and \hat{H}_α are also isomorphic to \hat{Y} , that is, isomorphic to \hat{Y} up to the first and second coordinates of the labels of each vertex $x \in V(\hat{Y})$ and every edge $uv \in E(\hat{Y})$, respectively, with η and $\eta' = \alpha \circ \eta$ being the corresponding isomorphisms; see Figure 1.

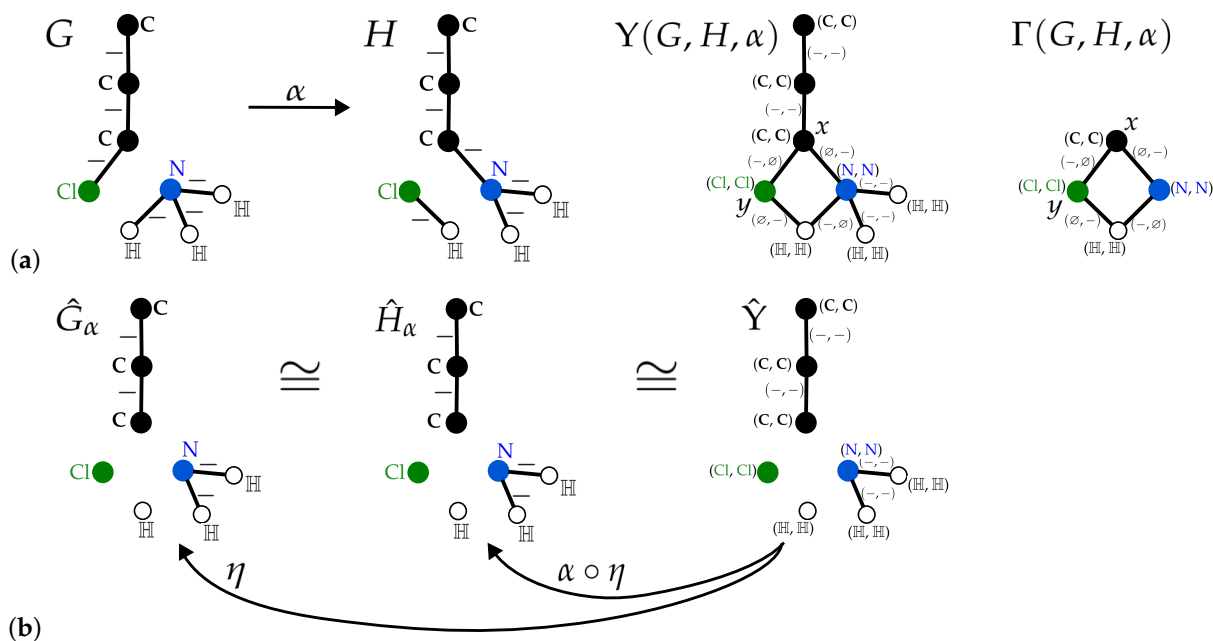


Figure 1. Reaction illustrating the formalism introduced in Section 2. Atom types are determined by the color of vertices and their labels (N ●, H ○, C ●, and Cl ●), while edges represent single bonds depicted by label “-”. Row (a) shows an AAM α assigned to $G \longrightarrow H$, together with the ITS graph $Y := Y(G, H, \alpha)$ and the reaction center $\Gamma(G, H, \alpha)$. Here, α is implied by the position of the vertices in the drawing. Row (b) shows the graphs \hat{G}_α , \hat{H}_α and \hat{Y} obtained by removing the reaction edges from their counterparts, i.e., edges belonging to, or leading to edges in, $\Gamma(G, H, \alpha)$. Since the remaining edges conserve their labels under α , the graphs \hat{G}_α and \hat{H}_α are *isomorphic* to \hat{Y} , with the exception that the latter is, formally, labeled by ordered pairs of the repeated values for \hat{G}_α and \hat{H}_α . The maps η and $\alpha \circ \eta$ provide the isomorphisms between \hat{G}_α and \hat{Y} , and \hat{H}_α and \hat{Y} , respectively. Moreover, as an example of the usage of the variables a_Y^i and b_Y^j , consider the vertex x and the edge xy of Y . In this case, it holds $(a_Y^1(x), a_Y^2(x)) = (C, C)$, meaning that both variables hold the value $a_Y^1(x) = C$. Similarly, we have $(b_Y^1(xy), b_Y^2(xy)) = (-, \emptyset)$, meaning that $b_Y^1(xy) = -$ and $b_Y^2(xy) = \emptyset$. Introducing this notation is necessary to speak formally of arbitrary ITS graphs.

2.3. Connectedness of the ITS Graphs, Reaction Center and AAMs

We say that an AAM $\alpha : V(G) \rightarrow V(H)$ is *connected* if its corresponding ITS graph is connected. From Definition 2 follows that there exist subgraphs G' and H' of Y whose underlying graphs are isomorphic to the underlying graphs of G and H . Hence,

Observation 1. Let α be an AAM for $G \longrightarrow H$. If G or H are connected, then $Y(G, H, \alpha)$, and thus α , are connected.

From a chemical point of view, moreover, connectedness implies that every molecule takes part in the reaction in the following sense.

Lemma 2. Let $\alpha : V(G) \rightarrow V(H)$ be a connected AAM for the reaction $G \longrightarrow H$. Then, every connected component of G and H contains at least one reaction vertex.

Proof. Suppose this is not the case. Then, there is, without loss of generality, a connected component $G_1 \subsetneq G$ without a reaction vertex, i.e., G_1 and $H[\alpha(V(G_1))]$ are isomorphic, and $H[\alpha(V(G_1))]$ contains no reaction vertex. Thus, for every $x \in \alpha(V(G_1))$ and $xy \in E(H)$, $\alpha^{-1}(x)\alpha^{-1}(y) \in E(G_1)$ holds. Therefore, $H[\alpha(V(G_1))] =: H_1$ is a connected component in H , and the restriction $\alpha_1 : V(G_1) \rightarrow V(H_1)$ is a label-preserving isomorphism. It follows immediately that $Y(G_1, H_1, \alpha_1)$ is a connected component of the ITS $Y(G, H, \alpha)$. Since $G_1 \neq G$, $Y(G_1, H_1, \alpha_1) \neq Y(G, H, \alpha)$, contradicting the connectedness of α . \square

Nonetheless, neither the connectedness of G or H , nor of Y , imply the connectedness of the reaction center; see Figure 2. In particular, Figure 2a exhibits an AAM defining a connected ITS but a disconnected reaction center. Moreover, Figure 2b shows that the connectedness of the reaction center depends on the selection of the AAM between the same pair of graphs. We note, finally, that the converse of Lemma 2 is not true. To see this, consider two reactions $G \rightarrow H$ and $P \rightarrow Q$, each with connected ITS graphs that do not share common molecules (i.e., isomorphic connected components). By Lemma 2, every connected component of G , P , H and Q contains a reaction vertex, and thus, this is also true for the disjoint unions $G \cup P$ and $H \cup Q$. From a mathematical point of view, $(G \cup P) \rightarrow (H \cup Q)$ is also a valid chemical reaction. Its ITS graph is, by construction, the disjoint union of the ITS graphs for $G \rightarrow H$ and $P \rightarrow Q$, and thus not connected.

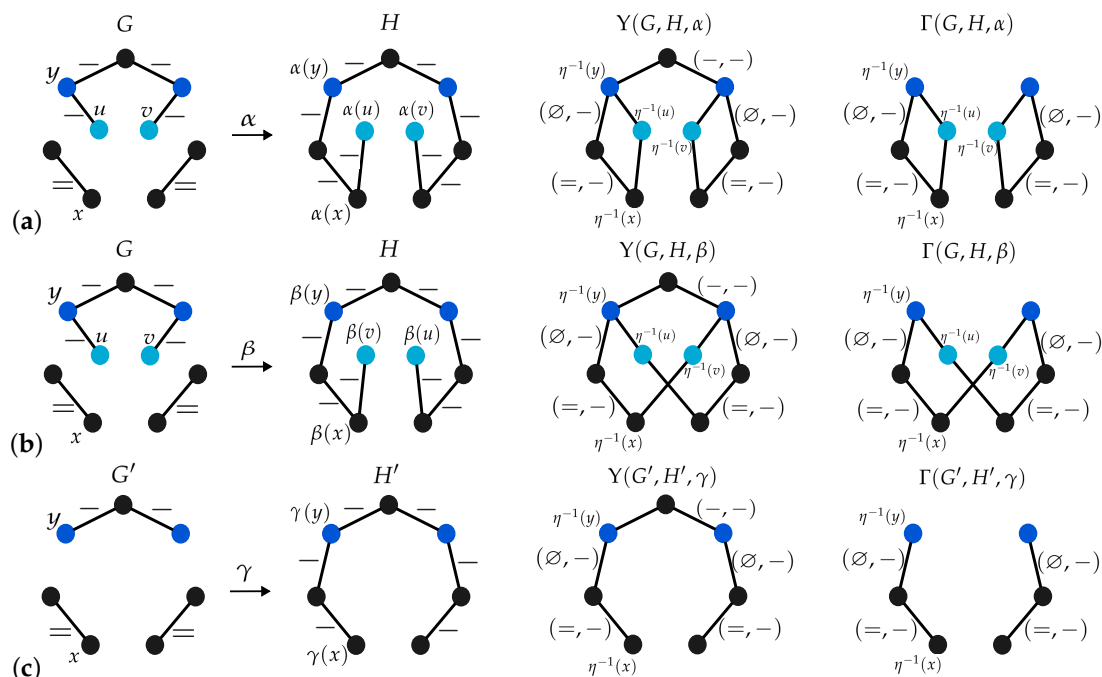


Figure 2. Reactions relating the selection of AAMs with the (dis-)connectedness of the ITS graphs and the reaction centers. Vertex colors depict atom types: N ●, H ●, and C ●. Only hydrogens that take part in the reaction are shown. Similarly, edge labels “—” and “=” denote single and double bonds, respectively. To simplify the drawing, not all edge labels are shown. The remaining edges have labels according to Definitions 2 and 3. We use the convention that the AAMs are implied by the relative positions of the vertices unless explicitly stated by their labels. (a) shows the reaction $G \rightarrow H$ subject to an AAM α that defines a connected ITS graph $Y(G, H, \alpha)$. The reaction center $\Gamma(G, H, \alpha)$, however, is disconnected. In (b), the same reaction $G \rightarrow H$ is explained by a different AAM β , differing from α only in the map of the hydrogens u and v of G , while preserving the image of every other atom. This small change produces a connected reaction center $\Gamma(G, H, \beta)$. Finally, (c) shows the reaction $G' \rightarrow H'$ between the graphs G' and H' obtained by removing the hydrogen atoms from G and H , and the atom map γ corresponding to the restriction of both α and β to non-hydrogen atoms. Ignoring the hydrogen atoms produces a disconnected reaction center $\Gamma(G', H', \gamma)$, from which neither $\Gamma(G, H, \alpha)$ nor $\Gamma(G, H, \beta)$ can be recovered unambiguously.

From a chemical perspective, every reaction can be explained as a sequence of elementary reactions, each of which proceeds via a single transition state and can be described by cyclic electron pushing. The latter implies that an elementary reaction necessarily has a connected ITS that, in fact, can be described by a closed circular walk of broken and formed bonds [34]. In reaction databases, however, reactions are rarely classified with respect to their detailed reaction mechanism. Instead, “single-step reaction” may designate a single step in a synthesis plan, which may involve a more complex sequence of steps that, however, proceed without the need or even the practical possibility of isolating intermediate products. In a sequence of elementary reactions, it is entirely possible that the reaction centers of the different steps are vertex disjoint. It is also possible that bonds that are formed in an earlier step are broken again in a later step. This may also lead to an AAM with a disconnected reaction center.

Even elementary reactions, however, may appear to have a disconnected reaction center. This is typically the case if the conventional structural formulas are used to represent molecules. In general, these do not represent hydrogen atoms and thus, the ITS graphs defined by an AAM on the non-hydrogen atom, by construction, also does not include hydrogens. For example, see Figure 2c. The making and breaking of bonds to hydrogen atoms is an important part of reaction mechanisms. We shall return to this and other complications that are often associated in particular with hydrogen atoms in Section 6.

2.4. Good and Bad Reaction Patterns

Chemical reactions $G \longrightarrow H$ can be grouped into equivalence classes that are defined by their “mechanism”, i.e., the bonds that are broken and formed and the so-called functional groups that are directly or indirectly involved. From a mathematical point of view, this amounts to a common subgraph L that appears in the different reactants G that can undergo the reaction of interest, and a common subgraph R in the product H that is the result of breaking and making bonds in L . In the simplest case, L and R cover only the reaction centers; see [31–33,35]. In general, however, atoms adjacent to the reaction center may be critical for enabling the transformation of the reaction center. Hence, it is of interest to consider also larger subgraphs L and R . It is worth noting that several subgraphs of the ITS were considered already in [36].

Definition 4. Let $G \longrightarrow H$ be a reaction with AAM $\alpha : V(G) \rightarrow V(H)$, ITS $Y := Y(G, H, \alpha)$, and reaction center $\Gamma := \Gamma(G, H, \alpha)$. A good reaction pattern for $G \longrightarrow H$ is a pair of subgraphs $L \subseteq G$ and $R \subseteq H$, together with the restriction $\alpha' : V(L) \rightarrow V(R)$ of α such that $\Gamma \subseteq Y(L, R, \alpha') \subseteq Y$.

Note that by construction, R is induced from H by the image $\alpha(V(L))$ of $V(L)$ under α ; that is, $R = H[\alpha'(V(L))]$. If (L, R, α') is a good reaction pattern, then $\Gamma(L, R, \alpha') = \Gamma(G, H, \alpha)$. A reaction pattern (L, R, α') is uniquely defined by a subgraph $\Psi \subseteq Y(G, H, \alpha)$ by virtue of the restriction of η and η' to $V(\Psi)$ and defining $\eta(u)\eta(v) \in E(L)$ if and only if $uv \in E(\Psi)$ and $b_Y^1(uv) \neq \emptyset$; similarly, $\eta'(u)\eta'(v) \in E(R)$ if and only if $uv \in E(\Psi)$ and $b_Y^2(uv) \neq \emptyset$. This implies the following.

Observation 2. A subgraph Ψ of $Y(G, H, \alpha)$ defines a good reaction pattern (L, R, α') satisfying $Y(L, R, \alpha') = \Psi$ if and only if $\Gamma(G, H, \alpha) \subseteq \Psi \subseteq Y(G, H, \alpha)$.

We also note that $\hat{L}_{\alpha'}$ and $\hat{R}_{\alpha'}$ are isomorphic, and α' serves as isomorphism. These graphs are also isomorphic to $\hat{\Psi}$ up to the doubling of the labels.

3. Incomplete Atom-to-Atom Maps and partial ITS Graphs

3.1. Good and Bad Partial AAMs

Some atom mapping tools such as Indigo [7] do not compute a complete AAM but instead only produce a bijection between subsets of vertices of the graphs in the reaction.

Formally, they produce a bijection $\beta : U \rightarrow W$, preserving vertex labels between subsets $U \subseteq V(G)$ and $W \subseteq V(H)$ for the reaction $G \rightarrow H$. More recent tools such as RXNMapper [8], or LocalMapper [10] produce global AAMs by default but switch to a “local mode” if the input reactions are unbalanced. GraphormerMapper [9] also can be used to rebalance an unbalanced reaction while computing a full AAM. KEGG RCLASSES [37] in general also provide only partial AAMs; see [38].

By definition, β defines an AAM from the induced subgraph $G[U]$ to the induced subgraph $H[W]$; hence, both the *partial* ITS graph $Y(G[U], H[W], \beta)$ and the subgraph $\Gamma(G[U], H[W], \beta)$ are well defined. The subgraph $\Gamma(G[U], H[W], \beta)$ describes a possibly partial reaction center; that is, it may or may not contain already all the reaction edges implied by $G \rightarrow H$. Let us write \hat{G}_β and \hat{H}_β , respectively, for subgraphs obtained from G and H by removing the reaction edges determined by β in $G[U]$ and $H[W]$. We are now interested in using β as an “anchor”, which we intend to extend into an AAM that is as complete as possible.

Definition 5. Let $G \rightarrow H$ be a reaction and $\beta : U \rightarrow W$ a partial AAM, for $U \subseteq V(G)$ and $W \subseteq V(H)$. A proper extension of β is a bijective map $\gamma : U' \rightarrow W'$, with $U \subseteq U' \subseteq V(G)$ and $W \subseteq W' \subseteq V(H)$, that (i) preserves vertex labels, i.e., $a_G(x) = a_H(\gamma(x))$ for all $x \in U'$, and (ii) coincides with β on U , i.e., $\gamma(x) = \beta(x)$ for all $x \in U$.

Clearly, a proper extension γ of β is again a (partial) AAM for $G \rightarrow H$ and thus defines a partial ITS graph $Y(G[U'], H[W'], \gamma)$, as well as reaction edges and subgraphs \hat{G}_γ and \hat{H}_γ . Since β is a bijection that conserves vertex labels, and a reaction is balanced if there exists a bijection from $V(G) \rightarrow V(H)$ preserving vertex labels, such a bijection must also exist on $V(G) \setminus U \rightarrow V(H) \setminus W$. Thus, we have the following.

Observation 3. If $G \rightarrow H$ is balanced and $\beta : U \rightarrow W$ with $U \subseteq V(G)$ and $W \subseteq V(H)$ is a partial AAM, then there exists a proper extension $\gamma : V(G) \rightarrow V(H)$ of β , i.e., a complete AAM.

It is important to note that γ is not necessarily unique, even for fixed graphs G and H . It is possible, moreover, for the partial map β to explain reactions with distinct (i.e., non-isomorphic) reactants-graph G and/or distinct products-graph H , see Figure 3. This is relevant, for example, for the KEGG RCLASSES [37]. Each RCLASS specifies a partial AAM between a connected component of G and a connected component of H ; see ref. [38]. Such a “seed” is not sufficient, in general, to completely determine an AAM.

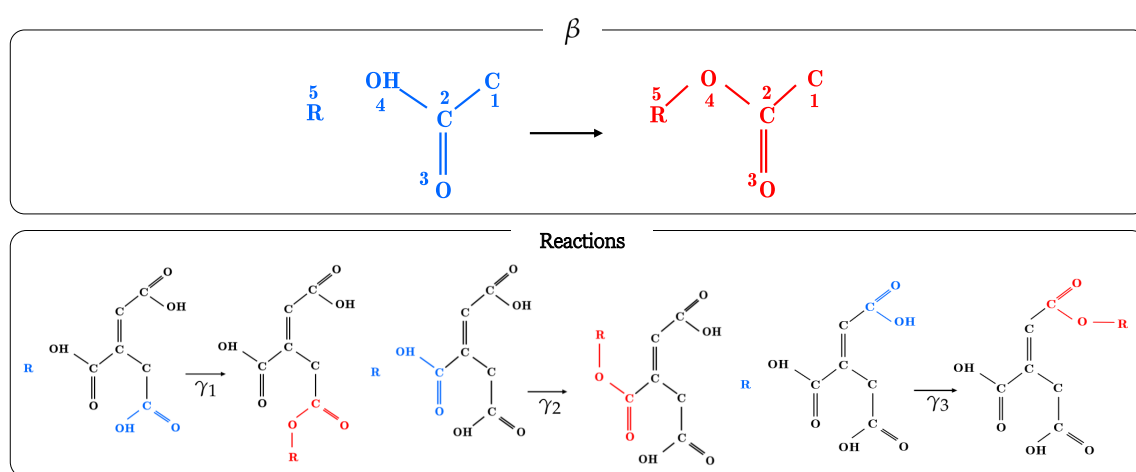


Figure 3. The KEGG RCLASS entry RC00460 corresponds to the partial AAM β shown in the upper panel. The lower panel exhibits 3 distinct reactions with maps γ_1 , γ_2 and γ_3 extending β , with blue and red patterns portraying the embeddings of the subgraphs implied by β . Partial AAMs therefore are not sufficient to guarantee a unique global AAM.

Definition 6. Let $G \longrightarrow H$ be a (not necessarily balanced) reaction. We say that a partial AAM $\beta : U \rightarrow W$ with $U \subseteq V(G)$ and $W \subseteq V(H)$, is good if U and W contain all reaction vertices and thus $Y(G[U], H[W], \beta)$ contains the reaction center as a subgraph. Otherwise, β is called bad.

From the chemical point of view, the following observation is of use.

Observation 4. A partial AAM β for a (not necessarily balanced) reaction $G \longrightarrow H$, is good if and only if \hat{G}_β and \hat{H}_β contain no more reaction edges of $G \longrightarrow H$.

In many applications, however, the true AAM will be unknown. From a mathematical perspective, therefore, Definition 6 does not refer to a ground truth but only requires that the reaction can be explained in such a way that β captures all bonds that change. In the case of balanced reactions, we obtain a simple characterization of good partial AAMs together with a simple algorithmic approach to construct a full AAM.

Proposition 1. Let $G \longrightarrow H$ be a balanced reaction and let $\beta : U \rightarrow W$ with $U \subseteq V(G)$ and $W \subseteq V(H)$ be a partial AAM. Then, β is a good partial AAM if and only if there is a proper extension γ of β such that γ is an isomorphism for \hat{G}_β and \hat{H}_β .

Proof. Suppose first that γ is an extension of β and an isomorphism from \hat{G}_β and \hat{H}_β , i.e., it is, in particular, a bijection that preserves adjacency, and vertex and edge labels, and thus is a complete AAM. Then, \hat{G}_γ and \hat{H}_γ contain no reaction edges. Moreover, the set of reaction edges implied by β , called $E(\Gamma(G, H, \beta))$, is contained in the set of reaction edges $E(\Gamma(G, H, \gamma))$ implied by γ . Since γ is also an isomorphism for \hat{G}_β and \hat{H}_β , then γ does not imply more reaction edges than β , that is, $E(\Gamma(G, H, \beta)) = E(\Gamma(G, H, \gamma))$. Thus, $\hat{G}_\beta = \hat{G}_\gamma$ and $\hat{H}_\beta = \hat{H}_\gamma$, and therefore β is a good partial AAM by Observation 4.

Suppose, on the other hand, that such an extension γ does not exist, i.e., for every bijection $\gamma : V(G) \rightarrow V(H)$ that preserves vertex labels and satisfies $E(\Gamma(G, H, \beta)) \subseteq E(\Gamma(G, H, \gamma))$, there is an edge $e = uv \in E(\Gamma(G, H, \gamma)) \setminus E(\Gamma(G, H, \beta))$. By definition, therefore, $E(Y(G[U], H[W], \beta))$ does not contain the complete reaction center and hence, $\beta : U \rightarrow W$ is a bad partial AAM. \square

Note that for a good partial AAM β , we have $\Gamma(G, H, \gamma) = \Gamma(G, H, \beta) \subseteq Y(G, H, \beta) \subseteq Y(G, H, \gamma)$. Moreover, it is feasible in practice to compute a proper extension γ using any algorithm for testing graph isomorphism or subgraph isomorphism provided it is amenable to specifying a partial isomorphism, i.e., β , as a constraint. Proposition 1 guarantees that finding γ can be performed by extending β to an isomorphism of \hat{G}_β and \hat{H}_β . A suitable family of algorithms for this purpose is the VF2 and relatives [39,40]. Supplied with the vertex matches defined by β as starting point, VF2 extends the set of vertices matched between $V(G)$ and $V(H)$ stepwisely such that each added match has the same adjacencies to all previously matched vertices in both G and H . Upon termination, it therefore either returns a bijection of $V(G)$ and $V(H)$ guaranteed to be the desired extension γ , or unmatched vertices remain, in which case no extension of β exists. Although the worst-case running time of VF2 is exponential, this algorithm is applicable in practice for chemical graphs. We note that it is not sufficient to extract \hat{G}_β and \hat{H}_β and check for their isomorphism since we are only concerned with isomorphisms γ that extend the given partial AAM β . It is possible, however, to tag pairs of vertices matched by β with the same unique label in both \hat{G}_β and \hat{H}_β , and test for the isomorphism of these modified graphs.

Before we proceed to the general case, let us consider a balanced reaction with partial AAM that does not cover all bonds that are broken or formed. This case is of practical importance since most AAM prediction tools ignore hydrogens. Consider the nucleophilic substitution reaction depicted in Figure 4. We can directly determine the ITS and the reaction center of reactions with a complete AAM, illustrated in Figure 4A. In this process, the C–Cl bond breaks, and a C–N bond forms.

From the mathematical point of view, the corresponding AAM β is trivially good *on the graphs without hydrogens* since it is a complete AAM for these. However, if the hydrogens are attached, β becomes a *bad* partial AAM with respect to such new graphs since now the reaction center involves a hydrogen atom that dissociates from the nitrogen and attaches to the chlorine. Indeed, β cannot be extended to an isomorphism between \hat{G}_β and \hat{H}_β . Instead, the correct reaction center contains a new edge representing the breaking bond between nitrogen and hydrogen, as well as the newly formed bond between hydrogen and chloride (Figure 4B). With respect to a detailed understanding of the reaction mechanism, however, we have to distinguish between hydrogens that are exchanged with the medium, and hydrogens that can be traced as individual entities through a reaction. This issue is not strictly limited to H atoms, however. We shall return to this point in Section 6 below.

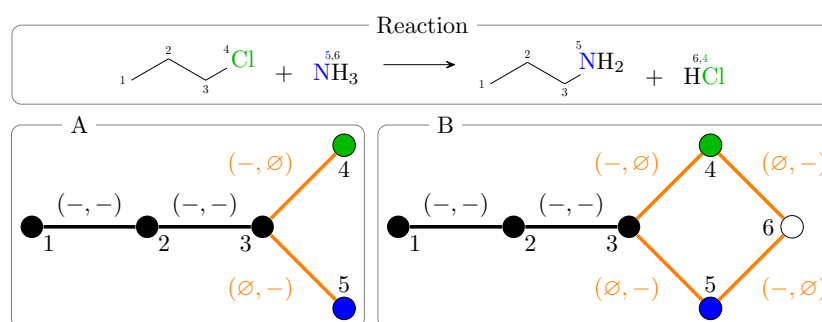


Figure 4. The process of hydrogen insertion in a nucleophilic substitution reaction. Vertex colors correspond to the atom types stated in the upper panel. (A) ITS graph with orange edges representing the reaction center. (B) The ITS graph following the insertion of hydrogen. The AAM β (when defined only on atoms 1–5) is a good AAM in representation (A) but a bad partial AAM in (B).

If $G \rightarrow H$ is balanced, then by Observation 3, an extension γ of any bad partial AAM β to a complete AAM exists. By construction, γ is then a good AAM. Of course, γ is not unique.

Among all such complete extensions, it is most natural to consider those that maximize $|E(\hat{G}_\gamma)| = |E(\hat{H}_\gamma)|$, i.e., those that maximize the number of edges in \hat{G}_γ and \hat{H}_γ , which is equivalent to attributing as few reaction edges as possible to G and H . This cost function [30,41] is closely related with the Principle of Minimal Chemical Distance, which stipulates that most chemical reactions proceed via the redistribution of a minimum number of valence electrons, which in most cases is the same as minimizing the total number of broken and form bonds. Since the total number of valence electrons is constant in balanced reactions, this measure is equivalent to maximizing the number $|E(\hat{Y})|$ of unchanged bonds in the ITS. We remark that for reactions that also involve the shifting of individual electrons it may be of interest to use slightly different quantifications of the “chemical distance”. This issue will be further investigated elsewhere.

3.2. An ILP for Completing Bad Partial AAMs on Balanced Reactions

Given a partial AAM β , the task of finding an extension γ that maximizes $|E(\hat{Y})|$ can be phrased as an Integer Linear Program (ILP). To this end, denote by \mathbf{A}_G and \mathbf{A}_H the adjacency matrices of G and H . The AAM is a bijection and thus can be expressed as a permutation matrix \mathbf{X} of size $|V(G)| \times |V(G)|$ with entries $X_{ij} = 1$ if $\alpha(x_i) = y_j$ where $x_i \in V(G)$ and $y_j \in V(H)$, and $X_{ij} = 0$ otherwise. To ensure that \mathbf{X} is a permutation matrix, it suffices to ensure

$$\sum_{i=1}^m X_{ij} = 1 \text{ for all } j \quad \text{and} \quad \sum_{j=1}^m X_{ij} = 1 \text{ for all } i.$$

Two graphs are isomorphic if a permutation matrix, such as that defined above, exists with $\mathbf{A}_H = \mathbf{X}\mathbf{A}_G\mathbf{X}^T$ [42], motivating our optimization function below.

In addition, we have the constraint $X_{ij} = 0$ if $a_G(x_i) \neq a_H(y_j)$ to enforce the preservation of atom labels, and we have $X_{ij} = 1$ if $\beta(x_i) = y_j$ as a means of constraining γ to coincide with β of all vertices mapped by β . The difference of the bonds is expressed by $\mathbf{D} := \mathbf{A}_H - \mathbf{X}\mathbf{A}_G\mathbf{X}^T$. As $\mathbf{X}^T = \mathbf{X}^{-1}$ for permutation matrices, $\mathbf{D} := \mathbf{A}_H\mathbf{X} - \mathbf{X}\mathbf{A}_G$ yields an equivalent definition suitable for linear programming. We have $\mathbf{D}_{ij} = 0$ if and only if there is no change in bond between G and H under the mapping γ encoded by \mathbf{X} . \mathbf{D}_{ij} takes bounded positive and negative integer values when a bond is formed or broken, respectively. Thus, the goal is to maximize the number of zero entries in \mathbf{D} . This can be achieved, e.g., by introducing binary indicator variables with $g_{ij} = 1$ if $\mathbf{D}_{ij} > 0$ and $g_{ij} = 0$ otherwise, and similarly $s_{ij} = 1$ if $0 > \mathbf{D}_{ij}$ and else $s_{ij} = 0$, employing the usual trick:

$$\begin{aligned} \mathbf{D}_{ij} &\leq Kg_{ij} & \text{and} & & -\mathbf{D}_{ij} &\leq Ks_{ij} \\ \mathbf{D}_{ij} &\geq -K + g_{ij}(K + 1) & & & -\mathbf{D}_{ij} &\geq -K + s_{ij}(K + 1) \end{aligned} \quad ,$$

where K is the maximal bond order in \mathbf{A}_G and \mathbf{A}_H , respectively. Then, $f = \sum_{ij}(g_{ij} + s_{ij})$ counts the number of non-zero entries in \mathbf{D} and needs to be minimized.

An implementation of the proposed ILP completion algorithm is available as the Python package `aamutils` in <https://github.com/klausweinbauer/AAMUtils>. It was evaluated on a test set of 1000 balanced reactions with correct complete AAMs. The test set is derived from the validation dataset used in [19], which is a composition of corrected reactions from USPTO 50k, Jaworski and Golden datasets. AAMs are generated by three atom-to-atom mapping tools (RXNMapper [8], Graphormer [9], LocalMapper [10]) and are considered correct if all three AAMs are equivalent. To test the completion rate, we construct (bad) partial AAMs (1) by removing the mapping from nodes outside the reaction center, (2) by removing the mapping from nodes in the reaction center and (3) by removing any mappings. These partial AAMs are then extended by `aamutils` and compared to the initial AAM. For this benchmark, the CBC solver [43] is used in version 2.10.3. The success rate is perfect (100%) for (1) for extending the AAM when only having the reaction center mapped, and (2) for up to 35% of missing nodes in the reaction center. This rate drops to 87.4% when all mappings are removed, which corresponds to performing atom mapping based on the Minimal Chemical Distance. This is expected because minimizing the number of bond changes is not a suitable mapping method for some reaction types. Using the Diels–Alder reaction as an example, this method yields a disconnected reaction center with twice a change of one double bond into two single bonds (4 reaction edges) instead of the expected [4 + 2] cycloaddition reaction center (6 reaction edges). However, it performs well when expanding a reaction center that is already partially mapped. Figure 5 depicts the success rate of the benchmark results for missing atom mappings between 0% and 100%.

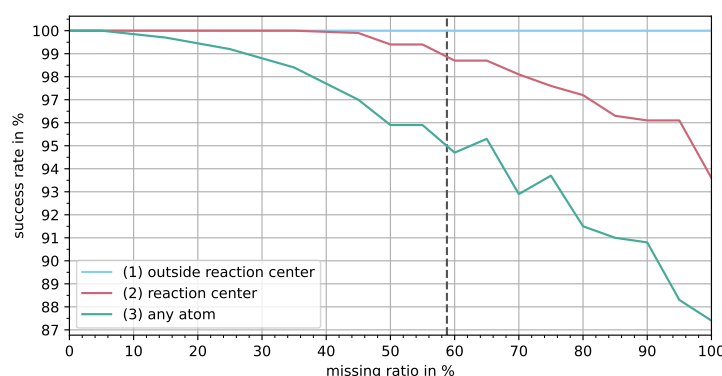


Figure 5. The completion rate (success rate) for reactions with a given fraction of mapped atoms. (1) For missing mappings outside the reaction center, (2) for incompletely mapped reaction centers, and (3) for partially mapped reactions in general. The dashed line marks the interpolated point where the success rate first drops below 95%. Based on the figure, we estimate that around 40% of the atoms in a reaction should be mapped to obtain completions with a small error rate of at most 5%.

3.3. Unbalanced Reactions and Graph Alignments

In the general case, the reaction is unbalanced and there is no bijection $\alpha : V(G) \rightarrow V(H)$ that preserves vertex labels. In other words, there will be unmapped vertices in G , H , or both graphs. Still, any proper extension $\gamma : U \rightarrow W$ defines an isomorphism between $\hat{G}_\gamma[U]$ and $\hat{H}_\gamma[W]$, while the rest of the vertices, i.e., $V(G) \setminus U$ and $V(H) \setminus W$, remain unmatched. To handle this situation, we introduce the following.

Definition 7. A graph \mathbb{A} with vertices and edges labeled by pairs $a_{\mathbb{A}}(x) = (a_{\mathbb{A}}^1(x), a_{\mathbb{A}}^2(x))$ with $a_{\mathbb{A}}^i(x) \in L_v \cup \{\emptyset\}$, and $b_{\mathbb{A}}(xy) = (b_{\mathbb{A}}^1(xy), b_{\mathbb{A}}^2(xy))$ with $b_{\mathbb{A}}^i(xy) \in L_e \cup \{\emptyset\}$, respectively, is said to be an alignment of two labeled graphs \mathcal{G} and \mathcal{H} if \mathcal{G} and \mathcal{H} are isomorphic, respectively, to the subgraphs G and H of \mathbb{A} induced, each, by $V_1 = \{x \in V(\mathbb{A}) \mid a_{\mathbb{A}}^1(x) \neq \emptyset\}$ and $V_2 = \{x \in V(\mathbb{A}) \mid a_{\mathbb{A}}^2(x) \neq \emptyset\}$ with vertex labels $a_G(x) = a_{\mathbb{A}}^1(x)$ for $x \in V_1$ and $a_H(x) = a_{\mathbb{A}}^2(x)$ for $x \in V_2$, and edge labels $b_G(xy) = b_{\mathbb{A}}^1(xy)$ for $x, y \in V_1$ and $xy \in E(\mathbb{A})$, and $b_H(xy) = b_{\mathbb{A}}^2(xy)$ for $x, y \in V_2$ and $xy \in E(\mathbb{A})$.

We note that, in practice, the analysis of unbalanced reactions becomes more difficult when compounds are missing on both sides, see ref. [19]. From a formal point of view, which is the emphasis of this contribution, it is more natural to consider the general case.

Each partial AAM $\gamma : U \rightarrow W$, thus, defines an alignment \mathbb{A}_γ of \hat{G}_γ and \hat{H}_γ , whose vertex set consists of *alignment columns* of two types: *match columns* defined to be the ordered pairs $(x, \gamma(x))$ for every $x \in U$, while all unmatched vertices $x \in V(G) \setminus U$ and $y \in V(H) \setminus W$ correspond, respectively, to *in/del columns* (x, \emptyset) and (\emptyset, y) . Given $v \in V(\mathbb{A}_\gamma)$, we write v^1 and v^2 to denote, respectively, the first and second entries of v . This alignment \mathbb{A}_γ has, furthermore, an edge uv whenever either $u^1 \neq \emptyset \neq v^1$ and $u^1 v^1 \in E(\hat{G}_\gamma)$, or $u^2 \neq \emptyset \neq v^2$ and $u^2 v^2 \in E(\hat{H}_\gamma)$, i.e., the respective non-gap entries (i.e., not \emptyset) of u and v are adjacent in \hat{G}_γ or \hat{H}_γ . Labels are inherited, entrywise, from the vertex and edge labels in \hat{G}_γ and \hat{H}_γ in each case.

Thus, given a partial AAM γ for an unbalanced reaction, the corresponding alignment \mathbb{A}_γ can be used to infer a completion, i.e., a rebalancing of the reaction. This idea has been proposed also in the CGRtools manual [44], which, however, cautions that the “Decomposition of CGRs for unbalanced reactions back to reaction may lead to strange (and erroneous) structures”. We proceed by showing that ITS graphs obtained from alignments yield at least a mathematically sound avenue for reaction rebalancing. Whether this yields a chemically correct solution of course depends on the correctness of the alignment.

The basic idea is to assume that vertices of G and H that appear in columns of the form (x, \emptyset) or (\emptyset, x) in the alignment \mathbb{A}_γ , refer to atoms x that should be present in both the reactants and the products but that, in the reaction data, were omitted in either the reactant graph or in the product graph. The alignment then specifies the partial AAM and the balancing of the reaction. More formally, we have the following.

Proposition 2. Let γ be a partial AAM for an unbalanced reaction $G \rightarrow H$. Then, the corresponding alignment \mathbb{A}_γ of \hat{G}_γ and \hat{H}_γ uniquely defines (up to isomorphism) a balanced reaction $G^* \rightarrow H^*$ and a corresponding complete AAM $\gamma^* : V(G^*) \rightarrow V(H^*)$.

Proof. We observe that the alignment graph \mathbb{A}_γ can be converted into an ITS graph $Y^*(G, H, \mathbb{A}_\gamma)$ by replacing gap labels \emptyset of in/del columns with the label of the corresponding atom that is present in the reactant or product graph, respectively. Similarly, edge labels are inherited. Since the alignment \mathbb{A}_γ again refers to the graphs \hat{G}_γ and \hat{H}_γ , it does not include the reaction edges. The construction of $Y^* := Y^*(G, H, \mathbb{A}_\gamma)$ thus requires the re-insertion of the reaction edges. To this end, we proceed as follows:

(A1) In in/del columns $x \in V(\mathbb{A}_\gamma)$, a gap label \emptyset is replaced by the non-gap atom label $a_G(x^1)$ or $a_H(x^2)$, respectively, that appears in the same column.

- (A2) If $xy \in E(\mathbb{A}_\gamma)$, then $b_{\mathbb{A}_\gamma}(xy) = (b_G(x^1y^1), \emptyset)$ or $b_{\mathbb{A}_\gamma}(xy) = (\emptyset, b_H(x^2y^2))$ are replaced, respectively, by $b_{Y^*}(xy) = (b_G(x^1y^1), b_G(x^1y^1))$ and $b_{Y^*}(xy) = (b_H(x^2y^2), b_H(x^2y^2))$, only if x and y are in/del columns of the same type, i.e., they have the same gap entry; otherwise, the label is left unchanged. This step handles all non-reaction edges.
- (A3) If $x, y \in V(\mathbb{A}_\gamma)$ and $xy \notin E(\mathbb{A}_\gamma)$ then let $b^1(xy) = \emptyset$ if $x^1y^1 \notin E(G)$ or $b^1(xy) = b_G(x^1y^1)$ if $x^1y^1 \in E(G)$, and also $b^2(xy) = \emptyset$ if $x^2y^2 \notin E(H)$ or $b^2(xy) = b_H(x^2y^2)$ if $x^2y^2 \in E(H)$. Set $b(xy) = (b^1(xy), b^2(xy))$. If $b(xy) = (\emptyset, \emptyset)$, or if either x or y are in/del columns, then there is no edge between them; otherwise, the reaction edge xy with label $b(xy)$ is added to $E(Y^*(G, H, \mathbb{A}_\gamma))$. This step pertains to reaction edges.

By construction, the graph $Y^*(G, H, \mathbb{A}_\gamma)$ built from the alignment \mathbb{A}_γ has the form of an ITS graph $Y^*(G, H, \mathbb{A}_\gamma) = Y(G^*, H^*, \gamma^*)$. In other words, the vertex sets of the graphs G^* and H^* correspond to the columns of \mathbb{A}_γ , and thus contain exactly the atoms present in at least one of the reactant graph G and the product graph H . The map $\gamma^* : V(G^*) \rightarrow V(H^*)$ is the natural one-to-one correspondence between $V(G^*)$ and $V(H^*)$ implied by the alignment columns. Since the labels in \mathbb{A}_γ are pairs that either contain a gap symbol or are of the form (a, a) with identical labels for the aligned atoms, the map $\gamma^* : V(G^*) \rightarrow V(H^*)$ preserves the vertex labels of G^* and H^* , and thus the same vertex label appears in $Y(G^*, H^*, \gamma^*)$. The edges and edge labels, on the other hand, are given by the first component for G^* , and the second component for H^* , of the edge labels of $Y^*(G, H, \mathbb{A}_\gamma)$ distinct from \emptyset as specified in (A2) and (A3) above, together with the reaction edges and their labels that were already defined by the partial AAM γ between G and H . Taken together, \mathbb{A}_γ therefore defines a unique ITS graph $Y(G^*, H^*, \gamma^*)$ that describes a balanced reaction $G^* \longrightarrow H^*$ with AAM γ^* . Moreover, γ^* by construction extends γ . \square

A concrete example is shown in Figure 6. By slight abuse of the notation, moreover, it follows in general that $G^*[V(G)] \cong G$ and $H^*[V(H)] \cong H$, i.e., adjacency and both vertex and edge labels are preserved between $G^*[V(G)]$ and G as well as between $H^*[V(H)]$ and H . In other words, G^* and H^* contain G and H as induced subgraphs with preserved labels.

Similarly, we consider now two vertices $x \in V(G)$ and $y \in V(G^*) \setminus V(G)$. The corresponding columns in \mathbb{A} are of the form (x^1, x^2) or (x^1, \emptyset) and (\emptyset, y^2) , and thus $xy \notin E(G^*)$, i.e., there is no edge connecting $V(G^*) \setminus V(G)$ and $V(G)$. An analogous argument holds for $V(H^*) \setminus V(H)$ and $V(H)$. Thus, G^* and H^* are extensions of G and H by one or more connected components. Therefore, the rebalancing of an unbalanced reaction $G \longrightarrow H$ to a balanced reaction $G^* \longrightarrow H^*$ thus only attaches additional connected components to G and H . Of course, these connected components correspond to the molecules that are missing in the unbalanced input reaction.

Our discussion so far suggests to further investigate algorithmic approaches for the inference of alignment-based AAMs that address the following scenario, which is of obvious practical relevance: Suppose we have obtained, for an unbalanced reaction $G \longrightarrow H$, the partial AAM $\beta : U \rightarrow W$ with $U \subseteq V(G)$ and $W \subseteq V(H)$. The task is then to compute an extension $\gamma : U' \rightarrow W'$ of β that (a) covers as many vertices of $V(G)$ and $V(H)$ as possible, and (b) at the same time implies as few (additional) reaction edges as possible. Once γ has been computed, both the rebalanced reaction $G^* \longrightarrow H^*$ and the complete AAM γ^* are then uniquely defined by Proposition 2.

Proposition 2 sets, therefore, the stage for considering the rebalancing and inference of AAMs as a joint optimization problem, possibly given an input partial AAM as a constraint. Our discussion above shows that this can also be viewed as a constrained graph alignment problem. The problem simplifies considerably for unbalanced reactions, where compounds are missing either only on the reactant side or the product side since in this case, \hat{G} is an induced subgraph of \hat{H} , or vice versa, i.e., the alignment consists only of matches and insertions, or matches and deletions. We will pursue this topic in separate work.

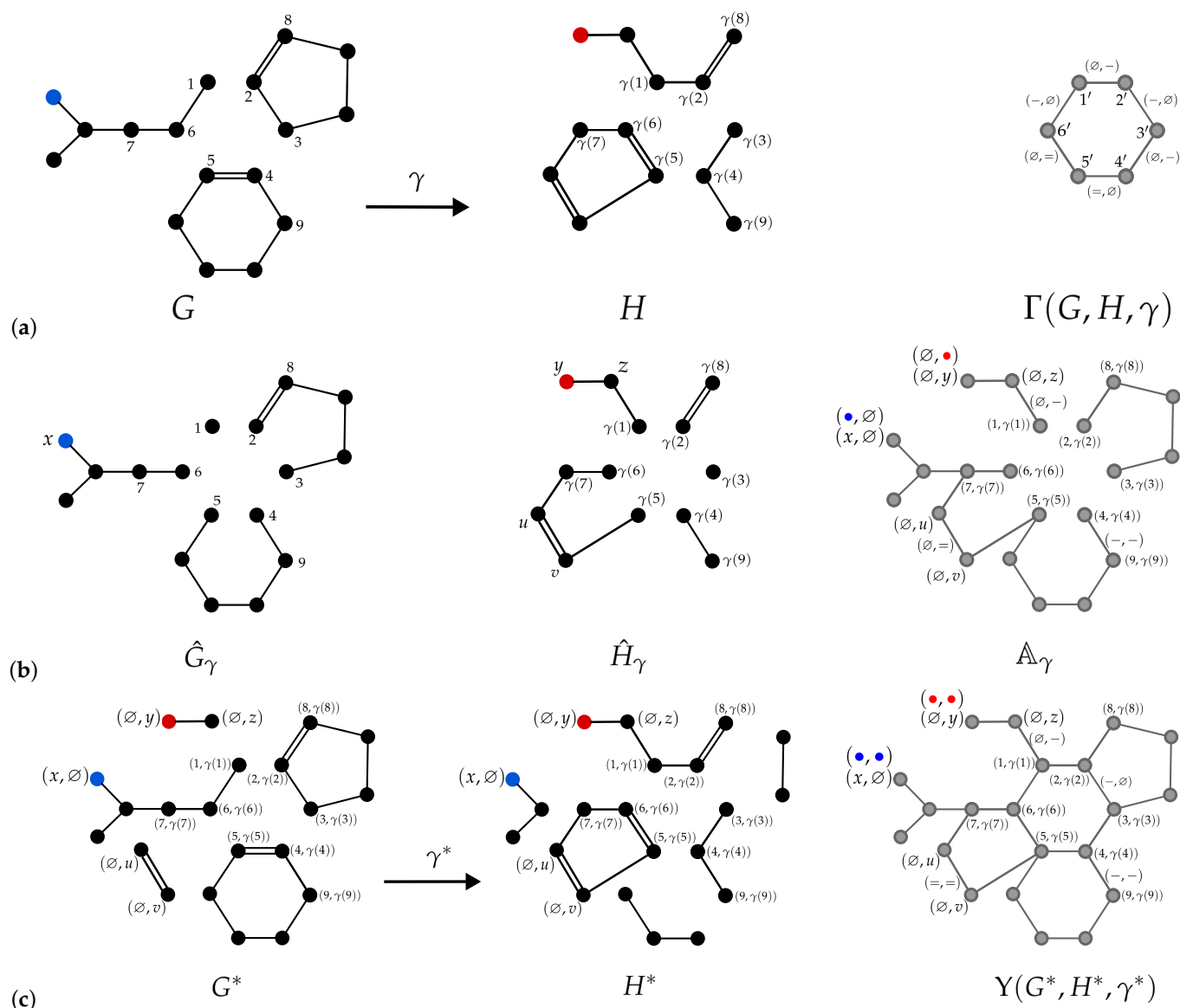


Figure 6. Construction of an alignment \mathbb{A}_γ and the ITS graph $Y(G^*, H^*, \gamma^*)$ produced by a partial AAM γ for $G \rightarrow H$. For clarity of the figure, double bonds are drawn explicitly in the first two graphs of each row and only specified as labels in the last one. (a) shows the reaction center $\Gamma(G, H, \gamma)$ comprising the reaction edges implied by γ , which are removed from G and H to obtain \hat{G}_γ and \hat{H}_γ . Then, in (b), the alignment \mathbb{A}_γ is produced by “gluing” the graphs \hat{G}_γ and \hat{H}_γ through γ . Formally, this operation is performed with respect to the alignment columns and assigns labels to them depending on their presence and labels in the input graphs, e.g., column (x, \emptyset) obtains the label (\bullet, \emptyset) , and similarly for the edges. (c) exhibits the ITS graph $Y(G^*, H^*, \gamma^*)$ obtained by applying (A1)–(A3) to \mathbb{A}_γ . In particular, vertex labels are duplicated and so are edge labels only if their end points are in/dels of the same type; otherwise, they remain unchanged, e.g., $(\emptyset, z)(1, \gamma(1)) \in E(\mathbb{A}_\gamma)$ conserves its label. This ensures that an edge is created in G^* and H^* only if both vertices are present in the same input graph.

4. Comparison of Partial AAM

4.1. Equivalence of Partial AAMs

Consider the dehydration reaction in Figure 7. From a chemical point of view, the two hydrogen atoms H7 and H8 are equivalent. This is not necessarily inherently obvious for the computational methods for inferring AAMs; an enumeration of atom maps thus may yield two alternative reactions.

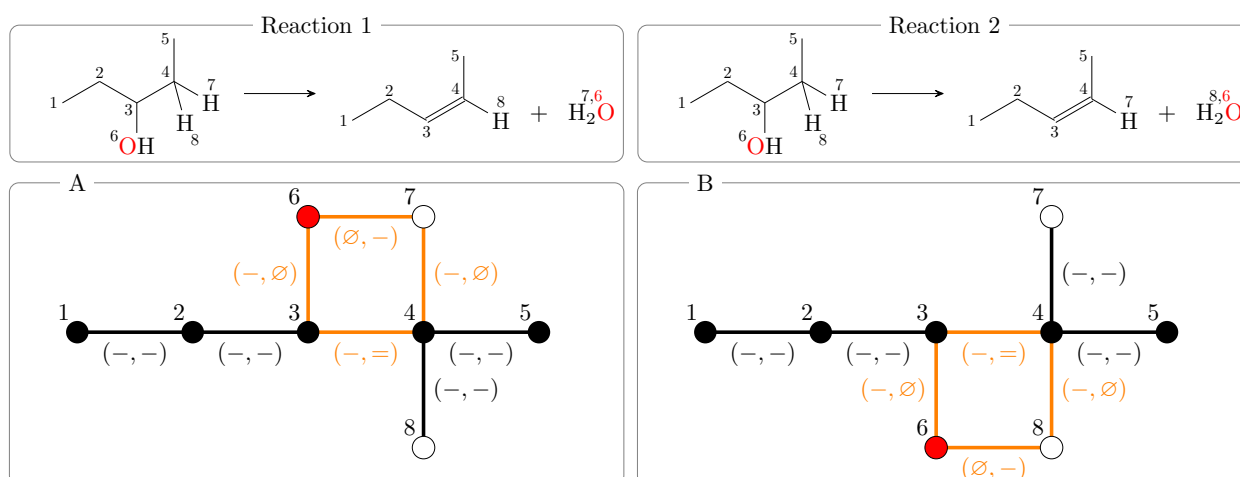


Figure 7. Reactions 1 and 2 illustrate two different inferred AAMs for hydrogen atom insertion (white vertices in the lower panels). Panel (A) shows the ITS of Reaction 1 and Panel (B) the ITS of Reaction 2. The two ITS graphs in (A,B) are isomorphic meaning that the two AAMs are equivalent.

In a more general setting, different AAM tools may return AAMs using their own numbering of atoms. It is of interest, therefore, to ask under which conditions two complete AAMs $\alpha : V(G) \rightarrow V(H)$ and $\alpha' : V(G') \rightarrow V(H')$ for isomorphic representations of the same reaction, i.e., $G \cong G'$ and $H \cong H'$, are equivalent. More formally, we ask whether there is a pair of isomorphisms $\varphi : V(G) \rightarrow V(G')$ and $\psi : V(H) \rightarrow V(H')$ such that $\psi(\alpha(x)) = \alpha'(\varphi(x))$ for all $x \in V(G)$. In [12], we showed that this condition is equivalent to the isomorphism of the complete ITS graphs $Y(G, H, \alpha)$ and $Y(G', H', \alpha')$. The two alternative ITS graphs for the dehydration reaction in Figure 7 indeed are isomorphic, and hence the corresponding AAMs are equivalent.

In this section, we extend the notion of equivalence to partial AAMs. The basic idea is that partial AAMs can only be equivalent if they cover “the same” parts of the reactant and product molecules in “the same” way. That is, the mapped parts of the molecules must match just like the full AAMs in [12] under isomorphisms for reactants and products, respectively.

Definition 8. The partial AAMs $\beta : U \rightarrow W$ with $U \subseteq V(G)$ and $W \subseteq V(H)$ and $\beta' : U' \rightarrow W'$ with $U' \subseteq V(G')$ and $W' \subseteq V(H')$ are equivalent, $\beta \equiv \beta'$, if there are extensions $\beta_* : V(G) \rightarrow V(H)$ and $\beta'_* : V(G') \rightarrow V(H')$, respectively, of β and β' , such that there exist isomorphisms $\varphi : V(G) \rightarrow V(G')$ for $G \cong G'$ and $\psi : V(H) \rightarrow V(H')$ for $H \cong H'$, for which $\beta'_*(\varphi(x)) = \psi(\beta_*(x))$ holds for all $x \in V(G)$.

Note that in particular, we have $U' = \varphi(U)$ and $W' = \psi(W)$.

As an immediate consequence of the definition and the results in [12], we obtain the following.

Observation 5. Let for $G \cong G'$, $H \cong H'$, $\beta : U \rightarrow W$ with $U \subseteq V(G)$ and $W \subseteq V(H)$, and $\beta' : U' \rightarrow W'$ with $U' \subseteq V(G')$ and $W' \subseteq V(H')$. Then, $\beta \equiv \beta'$ implies $Y(G[U], H[W], \beta) \cong Y(G'[U'], H'[W'], \beta')$.

In general, the isomorphism of the two partial ITS graphs is of course not sufficient to conclude that ITS graphs are isomorphic. This statement remains true even if, in addition, the reactant graphs (or the product graphs) are isomorphic; see Figure 8 for an example.

Here, the Diels–Alder reaction yields two distinct products, ortho (A) and meta (B). These products, despite their differences, still result in identical partial ITS graphs, in this case, the reaction centers. In fact, the isomorphism of partial ITS graphs and, in particular, reaction centers, is a very natural starting point to the defined classes of reactions [35].

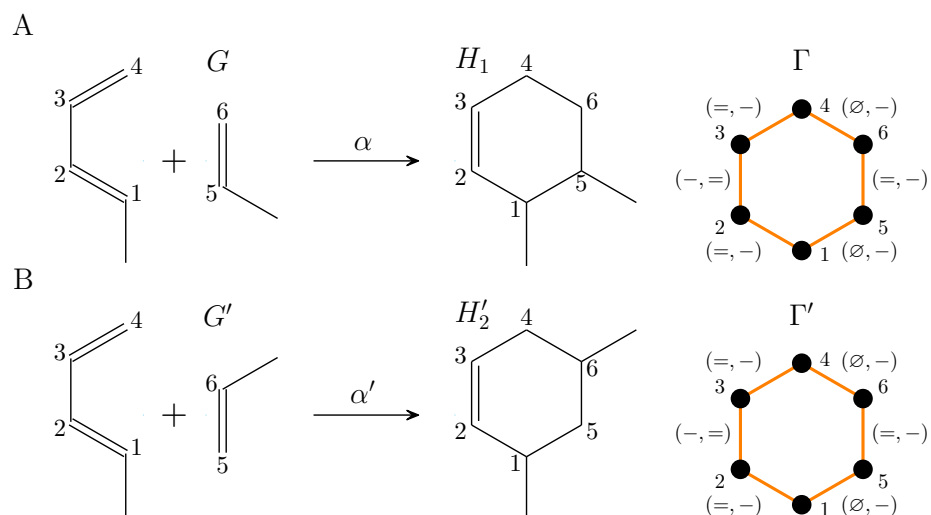


Figure 8. Diels–Alder reactions (A) ortho and (B) meta. The isomorphism of reactant graphs $G \cong G'$ and of partial ITS graphs, here of the reaction centers $\Gamma(G, H_1, \alpha) \cong \Gamma(G', H_2', \alpha')$ on the right and depicted with orange edges, does not imply isomorphisms of product graphs H_1 and H_2' .

The situation becomes less obvious if we assume $G \cong G'$ and $H \cong H'$. The esterification in Figure 9 shows an example with isomorphic molecular graphs and isomorphic partial ITS graphs. Nevertheless, there are two non-isomorphic extensions. The example suggests that this is a consequence of the reaction edges that are not contained in the given partial ITS graphs.

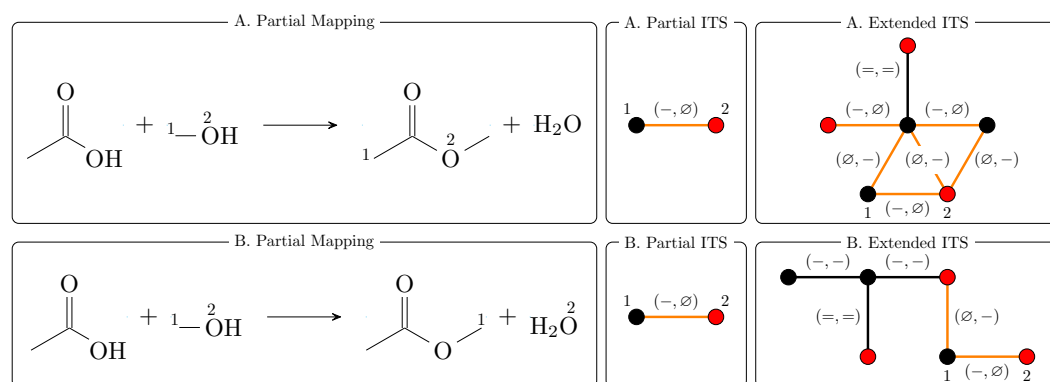


Figure 9. Isomorphism of reactant graphs $G \cong G'$ in row (A) and of product graphs $H \cong H'$ in row (B). These are two non-equivalent AAMs for the esterification of acetic acid and methanol, having isomorphic partial ITSs that capture only the breaking of C–O bonds. The example shows that conditions $G \cong G'$, $H \cong H'$, and $Y(G[U], H[W], \beta) \cong Y(G'[U'], H'[W'], \beta')$ are not sufficient to ensure that all extensions of β are equivalent. The key feature is that the non-equivalent ITS graphs contain reaction edges that are not present in the partial ITS graphs.

Examples of non-equivalent AAMs with isomorphic partial ITS graphs as shown in Figure 9 seem to involve partial ITS graphs that are bad reaction patterns. This suggests that the isomorphism of the reactant and product graphs together with the isomorphism of a good reaction pattern is sufficient. This is not the case, however, as the example in Figure 10 shows: two isomorphic good reaction patterns $Y(G[U], H[W], \beta)$ and $Y(G'[U'], H'[W'], \beta')$, for $G \longrightarrow H$ and $G' \longrightarrow H'$ with $G \cong G'$ and $H \cong H'$. Nevertheless, the full ITS graphs $Y(G, H, \beta)$ and $Y(G', H', \beta')$ are non-isomorphic. This is possible, in particular, whenever there are automorphisms of the reaction centers of G and H preserving the structure of the partial ITS graphs but not of the extensions of such good AAMs.

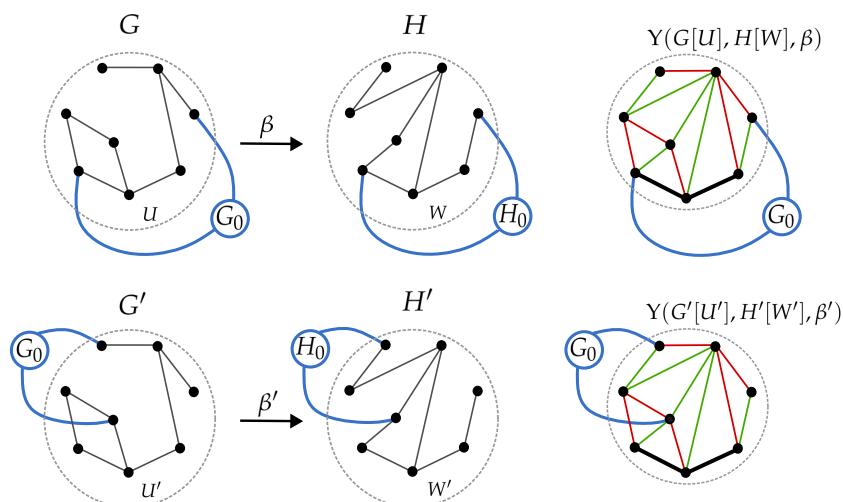


Figure 10. The vertices where G_0 is being attached to U and U' are (pairwise) in the same orbits in G and G' , and the same for H_0 and thus $G \cong G'$ and $H \cong H'$. The partial ITSs produced by β and β' are also isomorphic, with red, green, and bold edges depicting the edges present, correspondingly, only in $G[U]$ and $G'[U']$, only in $H[W]$ and $H'[W']$, and in both graphs of each reaction. Let $G_0 \cong H_0$. Then, the edges in G_0 , H_0 and in their attachments (blue edges) can be preserved by the extended maps in each case, so they contain no reaction edges and thus the maps are good. Note, nonetheless, the non-isomorphism of the full-ITS graphs due to the bold and blue edges. We emphasize, moreover, that in this counterexample, the maps β and β' are degree-preserving, which is a property of interest for our research and has been studied in more depth in related work [34].

In principle, it is possible to test whether $\beta \equiv \beta'$ by iterating over all isomorphisms φ of G and G' and ψ of H and H' and to test the condition in the definition directly.

For computational efficiency, it is of interest, however, to make use of the necessary condition in Observation 5. To this end, we note first that there is a 1–1 correspondence between U and W with the vertices $V(Y(G, H, \beta))$. To see this, let $\eta_G : U \rightarrow V(Y)$, $\eta_H : W \rightarrow V(Y)$, $\eta_{G'} : U' \rightarrow V(Y')$, $\eta_{H'} : W' \rightarrow V(Y')$ be known embeddings mapping the parts of G and H into $Y = Y(G, H, \beta)$ and G' and H' into $Y' = Y(G', H', \beta')$, respectively, as depicted by Figure 11. Note that in practical applications, these maps are part of the output of a tool that computes partial AAMs. Thus, we can assume that they are known and fixed. Now consider an isomorphism $\vartheta : V(Y) \rightarrow V(Y')$ of the partial ITS graphs. The construction of the ITS graph immediately implies that the two maps

$$\tilde{\varphi}_\vartheta := \eta_{G'}^{-1} \circ \vartheta \circ \eta_G \quad \tilde{\psi}_\vartheta := \eta_{H'}^{-1} \circ \vartheta \circ \eta_H \quad (1)$$

are the isomorphisms $\tilde{\varphi}_\vartheta : U \rightarrow U'$ from $G[U]$ and $G'[U']$, and $\tilde{\psi}_\vartheta : W \rightarrow W'$ for $H[W]$ and $H'[W']$ induced by the isomorphism ϑ of the ITS graphs. This discussion constitutes a proof of the following.

Proposition 3. *Let for $G \cong G'$, $H \cong H'$, $\beta : U \rightarrow W$ with $U \subseteq V(G)$ and $W \subseteq V(H)$, and $\beta' : U' \rightarrow W'$ with $U' \subseteq V(G')$ and $W' \subseteq V(H')$. Then, $\beta \equiv \beta'$ if and only if there is an isomorphism $\vartheta : V(Y) \rightarrow V(Y')$ from $Y = Y(G[U], H[W], \beta)$ and $Y' = Y(G'[U'], H'[W'], \beta')$ such that $\tilde{\varphi}_\vartheta$ can be extended to an isomorphism $\varphi_\vartheta : V(G) \rightarrow V(G')$ and $\tilde{\psi}_\vartheta$ can be extended to an isomorphism $\psi_\vartheta : V(H) \rightarrow V(H')$.*

This result provides us with a more efficient algorithm to check the equivalence of partial AAMs: it is sufficient to list all isomorphisms ϑ of the partial ITS graphs $Y(G[U], H[W], \beta)$ and $Y(G'[U'], H'[W'], \beta')$ and to use, for example, the VF2 algorithm to inspect whether the implied partial isomorphisms $\tilde{\varphi}_\vartheta$ and $\tilde{\psi}_\vartheta$ can be extended to full automorphisms of the reactant and product graphs, respectively. As soon as one such isomorphism ϑ is encountered, the answer is affirmative.

Analogous to the discussion following Proposition 1, one could also use an isomorphism test on suitable relabeled graphs that encode the partial isomorphisms $\tilde{\varphi}_\vartheta$ and $\tilde{\psi}_\vartheta$. Figure 11 summarizes the most important (sub)graphs we have encountered so far and the maps between them.

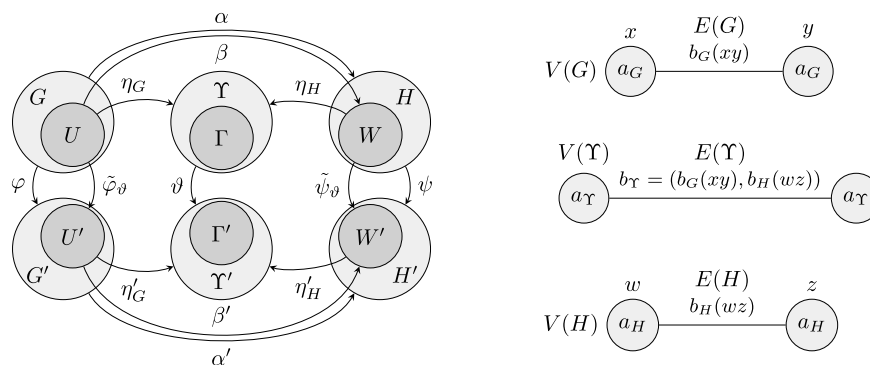


Figure 11. Visual summary of the most important graphs (e.g., reactant graph G , product graph H , and ITS graph $Y(G, H, \alpha)$) and the maps that connect them. The figure is intended as a visual guideline to help the reader to keep track of the most important notations.

This notion of equivalence of partial ITS graphs does not require that the reaction $G \rightarrow H$ is balanced. In fact, the possibility to extend that partial ITS to isomorphisms $G \cong G'$ and $H \cong H'$ immediately implies that one can also obtain isomorphic alignment graphs. It should be noted, however, that score-optimal alignments, and thus the ITS graphs obtained from them as described in Section 3.3, are not necessarily unique (up to isomorphism). In general, therefore, it will not suffice to only check the isomorphism of a score-optimal alignment-based ITS graph for $G \rightarrow H$ and $G' \rightarrow H'$, unless all co-optimal solutions are considered.

4.2. Consistency of Partial AAMs

There is no guarantee, however, that two distinct AAM computations will produce maps on “the same” atoms. The AAMs then are clearly non-equivalent, but they may still be consistent in the sense that their “union” is a partial AAM. More formally, we consider $U, U' \subseteq V(G)$ and $W, W' \subseteq V(H)$ and two bijective maps $\beta : U \rightarrow W$ and $\beta' : U' \rightarrow W'$. Then, their union $\beta \cup \beta' : U \cup U' \rightarrow W \cup W'$ given by $\beta \cup \beta'(x) = \beta(x)$ if $x \in U$ and $\beta \cup \beta'(x) = \beta'(x)$ if $x \in U'$ is well defined if and only if $\beta(x) = \beta'(x)$ for all $x \in U \cap U'$. Moreover, $\beta \cup \beta'$ is again a bijection if and only if $\beta(U \setminus U') = W \setminus W'$ and $\beta'(U' \setminus U) = W' \setminus W$.

The problem with this idea is again that the notions of “the same atoms” and of “union” in the context of AAMs is again subject to isomorphisms.

Definition 9. The partial AAMs $\beta : U \rightarrow W$ with $U \subseteq V(G)$ and $W \subseteq V(H)$ and $\beta' : U' \rightarrow W'$ with $U' \subseteq V(G')$ and $W' \subseteq V(H')$ are consistent if there are isomorphisms $\varphi : V(G) \rightarrow V(G')$ for $G \cong G'$ and $\psi : V(H) \rightarrow V(H')$ for $H \cong H'$ such that the union $\gamma := \beta \cup (\psi^{-1} \circ \beta' \circ \varphi) : U \cup \varphi^{-1}(U') \rightarrow W \cup \psi^{-1}(W')$ is well defined and bijective.

In particular, therefore $\beta = \psi^{-1} \circ \beta' \circ \varphi$ on $\tilde{U} := U \cap \varphi^{-1}(U')$. Equivalently, $\psi(\beta(x)) = \beta'(\varphi(x))$ for all $x \in \tilde{U}$. Let us write $\tilde{W} := W \cap \psi^{-1}(W')$ and $\tilde{\beta} : \tilde{U} \rightarrow \tilde{W}$ for the restriction of β to the vertices also mapped under the alternative atom map $\psi^{-1} \circ \beta' \circ \varphi$. Then, there is a partial ITS graph $Y(G[\tilde{U}], H[\tilde{W}], \tilde{\beta})$ that is isomorphic to a common induced subgraph of $Y(G[U], H[W], \beta)$ and $Y(G'[U'], H'[W'], \beta')$.

Moreover, if $\beta : U \rightarrow W$ and $\beta' : U' \rightarrow W'$ are consistent with associated isomorphisms $\varphi : V(G) \rightarrow V(G')$ and $\psi : V(H) \rightarrow V(H')$, then the union $\gamma = \beta \cup (\psi^{-1} \circ \beta' \circ \varphi)$ is a partial AAM and thus there is a common partial ITS graph $Y(G[U \cup \varphi^{-1}(U')], H[W \cup \psi^{-1}(W')], \gamma)$ that contains both $Y(G[U], H[W], \beta)$ and $Y(G'[U'], H'[W'], \beta')$ as subgraphs.

Furthermore, $\gamma' = \psi \circ \gamma \circ \varphi^{-1}$ is the corresponding union of AAMs for $G' \longrightarrow H'$, and by construction, we have $\gamma \equiv \gamma'$.

The problem of consistency between partial ITS that may overlap at least partially arises, for example, when reconstructing AAMs from KEGG RCLASSES [38]. An example of partial AAMs obtained from KEGG RCLASSES pertaining to the same reaction that turn out to be inconsistent is given in Appendix A.1. This real-life example emphasizes the utility of a mathematically sound formalization of partial AAM data. The question also naturally arises when partial AAMs obtained from different tools are computed for unbalanced reactions. In this setting, it is a natural task to reconcile these partial AAMs before computing extensions and alignment-based ITS graph as outlined in Section 3.2.

5. Atom-to-Atom Maps and Tautomerism

5.1. The Problem

Up to this point, we have studied molecules and molecular graphs. The phenomenon of tautomerism, however, rapidly interconverts certain structural isomers. Tautomers are distinct chemical species whose interconversion depends on external conditions such as solvent, temperature, or other physicochemical parameters. Tautomerization is therefore a *bona fide* chemical reaction that, in the simplest case, matches the scheme



where H is a mobile group, typically a hydrogen, and Q is a neutral atom [24,25].

A difficulty arising from tautomerism is that in some cases atoms are “chemically equivalent” despite the fact that they are not in a common orbit of any automorphism of the molecular graph. This is of relevance when comparing AAMs and their ITS graphs because non-isomorphic maps may still both correctly describe a given reaction. A well-known example is the two oxygen atoms in a carboxylic acid R-COOH. In the esterification reaction shown in Figure 12, the oxygen atom O³ and O⁴ on the left-hand side of the reactions are equivalent from a chemical point of view; thus, both atom maps can be viewed as chemically correct.

More generally, whenever M and Z in Equation(2) are atoms of the same type, both carrying no isomorphic additional substituents, M and Z become chemically equivalent due to the migration of H and the concurrent changes in bonding.

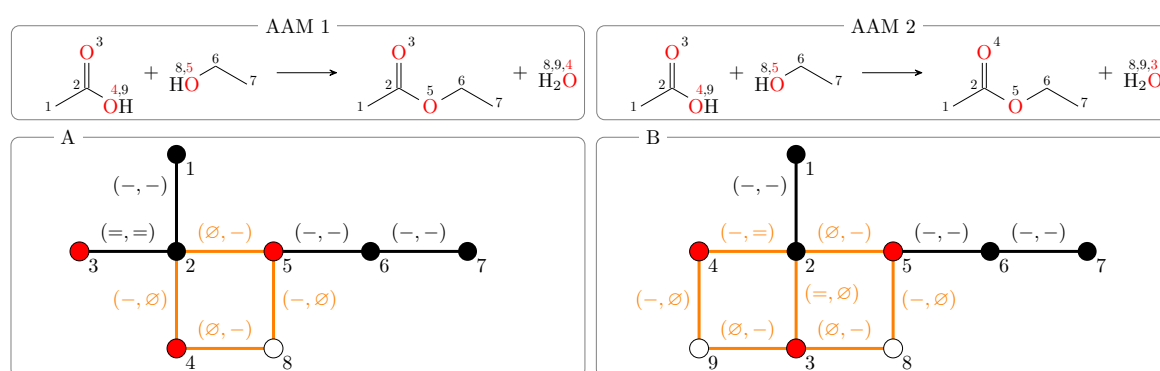


Figure 12. Identical AAMs in chemistry with O³ and O⁴ being symmetric due to tautomerism. However, there is no isomorphism present between their ITS graphs.

5.2. Tautomerism Equivalence

In order to handle such situations mathematically, it seems natural to combine the reaction of interest with possible tautomerizations on both the reactant and the product sides. Two reactions $G \longrightarrow H$ and $G' \longrightarrow H'$ with AAMs $\alpha : V(G) \rightarrow V(H)$ and $\alpha' : V(G') \rightarrow V(H')$ then could be considered “chemically the same” if G and G' as well as H and H' differ by tautomerizations of one or more of their constituent molecules.

From a mathematical point of view, this means that there are tautomerization reactions $G \rightarrow G'$ and $H \rightarrow H'$ with corresponding AAMs $\tau : V(G) \rightarrow V(G')$ and $\pi : V(H) \rightarrow V(H')$. Since the atoms in G and G' as well as H and H' will in general be numbered independently, we think of τ and π as maps that already include, i.e., are already composed with, the respective graph isomorphisms. The set of admissible maps τ and π is, in this new formulation, not only graph isomorphisms. Instead, we now admit ITS graphs $Y(G, G', \tau)$ and $Y(H, H', \pi)$ that describe tautomerizations.

Definition 10. Two reactions $G \longrightarrow H$ with AAM $\alpha : V(G) \rightarrow V(H)$ and $G' \longrightarrow H'$ with AAM $\alpha' : V(G') \rightarrow V(H')$ are tautomerism equivalent if there are bijections $\tau : V(G) \rightarrow V(G')$ and $\pi : V(H) \rightarrow V(H')$ such that both $Y(G, G', \tau)$ and $Y(H, H', \pi)$ are ITS graphs of valid tautomerization reactions and $\pi \circ \alpha = \alpha' \circ \tau$.

Of course, we have to include isomorphisms as “trivial tautomerizations” which correspond to ITS graphs without reaction edges. As an immediate consequence of this definition, a pair of equivalent AAMs is also tautomerism equivalent, but the converse is not true. While this may be a non-trivial problem in its own right, we assume for the remainder of this section that there is a computational procedure that decides whether or not a given ITS graph $Y(G, H, \tau)$ such that τ describes a tautomerization reaction.

Observation 6. The reactions $G \longrightarrow H$ and $G' \longrightarrow H'$ are tautomerism equivalent (up to automorphisms of G and H') if and only if there are tautomerization AAMs τ for $G \longrightarrow G'$ and π for $H \longrightarrow H'$ such that $Y(G, H', \pi \circ \alpha)$ and $Y(G, H', \alpha' \circ \tau)$ are isomorphic.

Proof. Each of τ , π , α , and α' is an AAM, i.e., a bijection that preserves vertex labels; thus, this remains true for their compositions. That is, both $\pi \circ \alpha$ and $\alpha' \circ \tau$ are AAMs for the composed reaction $G \longrightarrow H'$. But by Cor.1 in [12], the equivalence of the AAMs is equivalent to the isomorphism of the corresponding ITS graphs, from where the statement follows. \square

The check for tautomerization equivalence thus again reduces to checking the graph isomorphism of the ITS graph. However, now we have to check all concatenations of the given AAMs with tautomerization reactions on the reactant and product sides.

An alternative approach is to compare the ITS graphs $Y := Y(G, H, \alpha)$ and $Y' := Y(G', H', \alpha')$ directly. While they are not isomorphic in general, there is still a bijection $\zeta : V(Y) \rightarrow V(Y')$ that preserves the vertex labels. Recall that both vertex labels and edge labels in the ITS graphs $Y(G, H, \alpha)$ and $Y(G', H', \alpha')$ are pairs, whose individual components are labels of G and H , and G' and H' , respectively. We can thus use ζ to construct a graph Ξ whose vertices have labels $a_{\Xi}(x) = (a_G(x), a_H(x), a_{G'}(\zeta(x)), a_{H'}(\zeta(x)))$, all of which are identical by definition. There are edges $xy \in E(\Xi)$ if $xy \in E(Y(G, H, \alpha))$ or $\zeta(x)\zeta(y) \in E(Y(G', H', \alpha'))$ with labels $b_{\Xi}(xy) = (b_G(xy), b_H(xy), b_{G'}(\zeta(x)\zeta(y)), b_{H'}(\zeta(x)\zeta(y)))$, with the convention that the label is set to $(b_G(xy), b_H(xy)) = (\emptyset, \emptyset)$ if $xy \notin E(Y(G, H, \alpha))$ and $(b_{G'}(\zeta(x)\zeta(y)), b_{H'}(\zeta(x)\zeta(y))) = (\emptyset, \emptyset)$ if $\zeta(x)\zeta(y) \notin E(Y(G', H', \alpha'))$. Thus, for an arbitrary graph Ξ of this type with vertex labels $a_{\Xi}(x) = (a_{\Xi}^1(x), a_{\Xi}^2(x), a_{\Xi}^3(x), a_{\Xi}^4(x))$ and edge labels $b_{\Xi}(xy) = (b_{\Xi}^1(xy), b_{\Xi}^2(xy), b_{\Xi}^3(xy), b_{\Xi}^4(xy))$, we find:

Observation 7. $Y(G, G', \tau)$ is obtained from Ξ by retaining only the vertex labels $(a_{\Xi}^1(x), a_{\Xi}^3(x))$ and edge labels $(b_{\Xi}^1(xy), b_{\Xi}^3(xy))$ and removing all edges with empty labels, i.e., $(b_{\Xi}^1(xy), b_{\Xi}^3(xy)) = (\emptyset, \emptyset)$. $Y(H, H', \pi)$ is obtained analogously from the second and fourth components of the labels.

Thus, it is possible, at least in principle, to test whether two ITS graphs are tautomerism equivalent by constructing, for a given vertex label-preserving bijection ζ , the auxiliary graph Ξ and checking whether the subgraphs $Y(G, G', \tau)$ and $Y(H, H', \pi)$, which are obtained as “projections” from Ξ , describe tautomerization reactions. In practice, it will not be necessary to consider all bijections ζ .

Instead, we expect that it will usually suffice to compute the graph Ξ , which (almost) minimizes the total number of unmatched edges and matching edges with different labels. This corresponds to computing the minimal chemical distance of the two ITS graphs $Y(G, G', \tau)$ and $Y(H, H', \pi)$.

5.3. Tautomerism ITS Graphs

In order to characterize the ITS graphs of a valid tautomerization reaction with AAM τ , we first note that tautomerization reactions are by definition intramolecular. Thus, τ is a disjoint union of bijections between the connected components of G, G' , and thus, $Y(G, G', \tau)$ comprises connected components that are in 1–1 correspondence with the connected components of both G and G' . It therefore suffices to consider tautomerism ITS graphs for single molecules. That is, $Y(G, G', \tau)$ is a tautomerism ITS graph for the reactants if and only if every connected component of $Y(G, G', \tau)$ is a tautomerism ITS graph. In other words, it is sufficient to specify what constitutes connected tautomerism ITS graphs.

Though Equation (2) defines elementary tautomeric reactions, it is no trivial task to determine whether a graph $Y(G, G', \tau)$ connects two tautomers. The reason is that these may be connected by a sequence of reactions of the form of Equation (2) rather than by a single transformation of this type. This is complicated by the fact that tautomerization reactions in general cannot be applied in arbitrary order, and the composition of tautomerization reactions does not necessarily conform to the pattern in Equation (2). For an elaborate example, see Figure 13.

These issues are studied in considerable detail in the theory of rule composition for graph rewriting systems [20,45]. While it may be possible to construct a formal system to recognize arbitrary tautomers on this basis, such an endeavor goes beyond the scope of the present contribution. As a pragmatic alternative, tautomer databases [46] will be helpful.

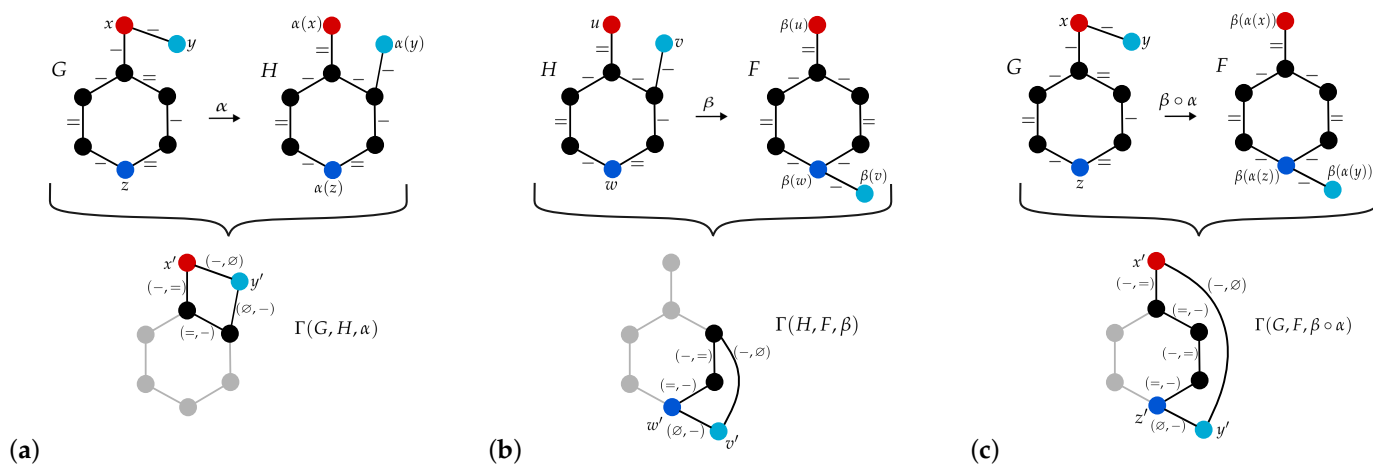


Figure 13. Three AAMs describing *Keto–Enol* tautomerizations between (G) *pyridine-4-ol*, (H) *pyridin-4(5H)-one* and (F) *4-pyridone*. Atom types are shown as vertex colors (N ●, H ●, C ●, and O ●). Single and double bonds are labeled as – and =, respectively. The AAMs α in (a) and β in (b) produce reaction centers comprising similar tautomerization mechanisms, i.e., reaction centers isomorphic up to edge labels and hydrogen labels. The atom map in (c) from G to F given by their composition $\beta \circ \alpha$, however, does not conform to Equation (2). It is important to emphasize that the reaction centers shown here do not depict the physical migration of hydrogen atoms, but rather they can be thought of as computational diagrams conveying the graph transformations encoding the reactions, and thus this example exhibits some limitations of the composition of AAMs as mathematical functions.

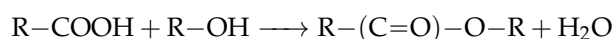
We note, finally, that mesomeric (resonance) structures could be handled in the same manner. This pertains in particular to aromatic systems. Alternatively, however, in this case, *aromatic bonds* may also be used as a separate bond types.

6. Hydrogens—A Necessary Nuisance

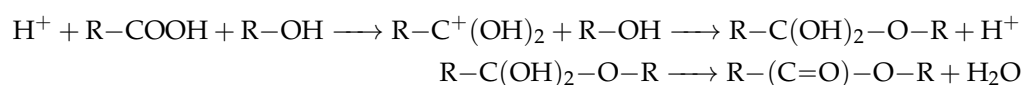
The representation of hydrogen atoms causes a variety of problems in formal descriptions of chemistry. There are good reasons to omit them partially or entirely in structural formulae. There are, however, as many good reasons to include them in other contexts. The omission of hydrogens without loss of information is possible because of the valency rules for carbon C and nitrogen N. The number of adjacent hydrogen is simply given as the difference between the expected valency of 4 or 3 and the number of incident bonds. This requires a complete molecule, possibly with explicitly given wildcard residues (e.g., -R). In reaction rules, on the other hand, the presence of a hydrogen taking part in a reaction, is essential for the reaction pattern since the reaction cannot take place in its absence.

Apart from a more compact structural formula, the omission of hydrogens simplifies the computation of automorphisms and isomorphisms. If hydrogens are represented explicitly, a -CH₃ group adds an extra six automorphisms, just mapping the three hydrogens onto each other. The omission of hydrogens is therefore also common practice for AAM tools, even though this leads to bad reaction patterns whenever hydrogens are part of the reaction center.

A more subtle issue at least closely related to the common omission of hydrogen is the question of which hydrogens are actually *preserved* through a formally balanced reaction. As an example, consider the balanced esterification reaction



At face value, the hydrogen atoms in H₂O on the product side come from the acid and alcohol reactants, respectively. A closer look at the reaction mechanism, however, shows that this is indeed a multi-step reaction that can be written as



Here, the H⁺ of the alcohol is exchanged for the H⁺ reactant that is taken from the solvent, while two hydrogens in the H₂O in the second line are the solvent-derived H⁺ and the hydrogen from the carboxylic acid. Assuming an aqueous solvent, all protons are thus exchanged with the solvent here. In reaction mechanisms such as unimolecular nucleophilic substitution (S_N1) and unimolecular elimination (E₁), the involvement of a solvent-borne proton is crucial, and it is incorrect to assume that the hydrogen exchange occurs solely within the internal reaction system.

On the other hand, it is also not correct to assume that hydrogens are always exchanged with the solvent system. In the last step of the biosynthesis of lanosterol in Figure 14, for instance, the hydrogen atoms are relocated by 1,2-hydrogen and 1,2-methyl shifts that keep them within the system.

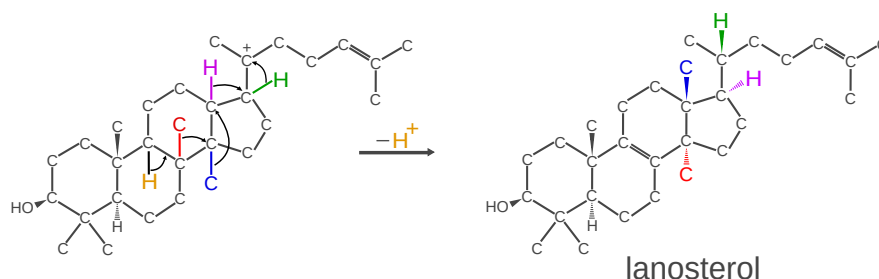


Figure 14. The last step of the biosynthesis of lanosterol showcases the problems arising from partial maps and hydrogen maps. Colors in the figure represent the correspondence of the hydrogen atoms and functional groups preserved by the reaction. Here, the hydrogen atoms are being *rewired* but conserved during multiple 1,2-hydrogen and 1,2-methyl shifts [47]. Only the hydrogen colored in yellow is being lost to the media as a proton and thus has no correspondence on the product side.

This suggests to address the issue in a more formal way that accommodates the correct mappings both with and without exchange with the medium. From a formal point of view, the best solution would be a decomposition of reactions into sequences of reactions that explicitly represent necessary hydrogens “borrowed” from the environment. Unfortunately, few data have this level of resolution.

The next-best approach is to attach an additional attribute label to hydrogens that can be exchanged with the medium in the course of a reaction. Together with this extra label, AAMs on the available reaction data could carry the pertinent information. Still, it is not clear how to compute such labels efficiently. Learning approaches are appealing for this purpose but will depend on well-curated datasets.

7. Concluding Remarks

We have surveyed the connections between incomplete atom-to-atom mapping data and their natural graph-theoretical description. In particular, in the context of unbalanced reactions, there is a natural but non-trivial connection to graph alignments.

The focus of this work is entirely on the mathematical framework and the formal properties of partial AAMs and their equivalent ITS graphs. Nevertheless, we show that in the case of balanced chemical reactions, the partial AAM can be extended to a global AAM with very high accuracy by solving a purely combinatorial optimization problem (framed here as an ILP that minimized chemical distance). To this end, only a small “anchor” is needed, e.g., a partial reaction center.

This task appears to become much more difficult on unbalanced reaction data. While we provide here a solid formal basis for this problem, practical implementations and an empirical analysis is the subject of ongoing work that will be published elsewhere. In particular, we have not attempted to devise efficient algorithms for the various restricted graph isomorphism problems appearing throughout the text. In this context, it may be of interest to invoke topological graph indices. While global graph invariants [48] may help to rule out the existence of isomorphisms, local versions, so-called vertex invariants, may be used to reduce the possible mappings and hence result in more efficient subgraph isomorphism checkers; see refs. [49,50].

Beyond the question of how AAMs can be obtained and completed, their comparison is a problem of practical significance, particularly in the context of benchmarking the increasing portfolio of atom-to-atom mapping tools. Again, our focus lies on the treatment of partial data, and the connection to testing isomorphisms of the corresponding subgraphs.

Not surprisingly, the comparison of subgraphs alone provides a necessary but not sufficient condition, which has to be complemented by additional requirements on the embeddings of the subgraphs. Interestingly, this framework extends to weaker forms of the equivalence of AAMs that naturally encompass tautomerism. As an open problem, however, the question remains of how to formally determine when two molecular graphs are tautomers of each other. Mesomeric formulas are captured by the same formalism.

A notoriously difficult problem is the handling of hydrogen atoms and their persistence through chemical reactions. Here, we have only touched upon this topic to the extent that it has a direct bearing on AAMs and their completion. Our treatment of the topic is far from complete or comprehensive, although we highlight some of the major issues and sketch some possible approaches that could be explored in future work.

The detailed presentation of the partial AAMs and ITS graphs results in a concise specification of several optimization problems associated with the completion of partial reaction data. As an example, we considered the completion of bad partial AAMs in the setting of balanced reactions (Section 3.2).

Related problems, touched upon in our presentation, most likely can be tackled by similar ILP formulations, providing a solid starting point for dedicated computational investigations. Algorithmic aspects, however, are beyond the scope of this contribution.

It should not be forgotten, furthermore, that even in the presence of partial chemically correct information, there is no guarantee that purely combinatorial objectives, such as the minimum edit distance or maximum common subgraphs, will lead to chemically correct solutions. We expect, however, that this will be the case in most instances.

The main practical use of the framework outlined here, thus, is to expose how partial information, that we expect to become increasingly available with the rise in learning-based approaches in computational chemistry, can be incorporated into optimization-centered approaches.

Author Contributions: M.E.G.L. and P.F.S. designed the study. K.W. and T.G. performed the computational experiments. M.E.G.L., K.W., T.-L.P., N.B., N.D., C.F., T.G., D.M. and P.F.S. contributed to the derivation of the theoretical results, the discussion, and the writing of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: M.E.G.L. acknowledges financial support by the Federal Ministry of Education and Research of Germany and by the Sächsische Staatsministerium für Wissenschaft Kultur und Tourismus in the program Center of Excellence for AI-research “Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig”, project identification number: ScaDS.AI. The research of K.W. and T.-L.P. received funding from the European Unions Horizon Europe Doctoral Network program under the Marie-Sklodowska-Curie grant agreement No 101072930 (TACsy—Training Alliance for Computational systems chemistry). The research of N.D. is supported by the German Research Foundation (DFG) as part of the Priority Program SPP 2363 (grant no. STA 850/58-1 497135079 to P.F.S.). The research of N.B. and T.G. is supported by the Novo Nordisk Foundation (grant NNF21OC0066551 “MATOMIC” to C.F., D.M., and P.F.S.). P.F.S. acknowledges support by the German Federal Ministry of Education and Research BMBF through DAAD project 57616814 (SECAI, School of Embedded Composite AI). Publication costs were covered in part by ScaDS.AI and in part by the Open Access Publishing Fund of Leipzig University supported by the German Research Foundation within the program Open Access Publication Funding.

Data Availability Statement: The software package as well as the scripts and data to re-run the validation experiments are available at <https://github.com/klausweinbauer/AAMUtils/tree/paper> (accessed on 27 August 2024).

Acknowledgments: We thank Maria Waldl for the valuable comments on the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AAM	atom-to-atom map
CGR	condensed graph of the reaction
DPO	double pushout (graph grammars)
ILP	integer linear program
ITS	imaginary transition state

Appendix A

Appendix A.1. Inconsistent AAMs from KEGG RCLASSES

The KEGG RCLASSES shown in Figure A1 determine two partial AAMs for the KEGG reaction R01144. The map AAM1 in this figure comprises atoms maintained between oxaloacetate 4-methyl ester and oxaloacetate, while AAM2 connects oxaloacetate 4-methyl ester and methanol.

The two partial AAMs are inconsistent because the O with label 2 in the reactants is being mapped to two non-equivalent atoms on the product side, i.e., there is no isomorphism between the product graphs mapping the corresponding images of this atom to each other.

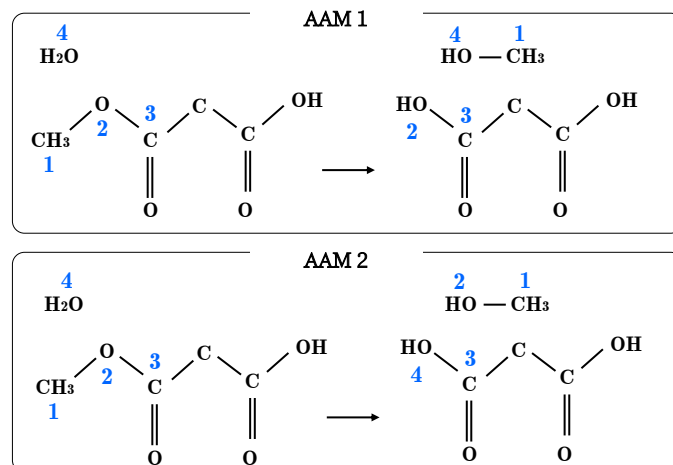


Figure A1. The KEGG RCLASSes RC00460 (upper panel) and RC00461 (lower panel) available for the KEGG reaction R01144.

References

- Elsevier “Reaxys”. Available online: <https://www.elsevier.com/products/reaxys> (accessed on 20 July 2024).
- Crabtree, J.D.; Mehta, D.P. Automated reaction mapping (ACM). *J. Exp. Algorithmics* **2009**, *13*, 15. [[CrossRef](#)]
- Chen, W.L.; Chen, D.Z.; Taylor, K.T. Automatic reaction mapping and reaction center detection. *WIREs Comput. Mol. Sci.* **2013**, *3*, 560–593. [[CrossRef](#)]
- Fooshee, D.; Andronico, A.; Baldi, P. ReactionMap: An efficient atom-mapping algorithm for chemical reactions. *J. Chem. Inf. Model.* **2013**, *53*, 2812–2819. [[CrossRef](#)] [[PubMed](#)]
- Mann, M.; Nahar, F.; Schnorr, N.; Backofen, R.; Stadler, P.F.; Flamm, C. Atom mapping with constraint programming. *Alg. Mol. Biol.* **2014**, *9*, 23. [[CrossRef](#)] [[PubMed](#)]
- Rahman, S.A.; Torrance, G.; Baldacci, L.; Martínez Cuesta, S.; Fenninger, F.; Gopal, N.; Choudhary, S.; May, J.W.; Holliday, G.L.; Steinbeck, C.; et al. Reaction decoder tool (RDT): Extracting features from chemical reactions. *Bioinformatics* **2016**, *32*, 2065–2066. [[CrossRef](#)]
- Savelev, A.; Puzanov, I.; Samoilov, V.; Karnaukhov, V. Indigo Toolkit. 2019. Available online: <https://lifescience.opensource.epam.com/indigo/index.html> (accessed on 27 August 2024).
- Schwaller, P.; Hoover, B.; Reymond, J.L.; Strobelt, H.; Laino, T. Extraction of organic chemistry grammar from unsupervised learning of chemical reactions. *Sci. Adv.* **2021**, *7*, eabe4166. [[CrossRef](#)]
- Nugmanov, R.; Dyubankova, N.; Gedich, A.; Wegner, J.K. Bidirectional graphormer for reactivity understanding: Neural network trained to reaction atom-to-atom mapping task. *J. Chem. Inf. Model.* **2022**, *62*, 3307–3315. [[CrossRef](#)]
- Chen, S.; An, S.; Babazade, R.; Jung, Y. Precise atom-to-atom mapping for organic reactions via human-in-the-loop machine learning. *Nat. Commun.* **2024**, *15*, 2250. [[CrossRef](#)]
- Lan, Z.; Zeng, Z.; Hong, B.; Liu, Z.; Ma, F. RCsearcher: Reaction center identification in retrosynthesis via deep Q-learning. *Pattern Recogn.* **2024**, *150*, 110318. [[CrossRef](#)]
- González Laffitte, M.E.; Beier, N.; Domschke, N.; Stadler, P.F. Comparison of atom maps. *MATCH Comm. Math. Comp. Chem.* **2023**, *90*, 75–102. [[CrossRef](#)]
- Fujita, S. Description of organic reactions based on imaginary transition structures. 1. Introduction of new concepts. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 205–212. [[CrossRef](#)]
- Wilcox, C.S.; Levinson, R.A. A self-organized knowledge base for recall, design, and discovery in organic chemistry. In *Artificial Intelligence Applications in Chemistry*; Pierce, T.H., Hohne, B.A., Eds.; Am. Chem. Soc. Symposium Series, Chapter 18; American Chemical Society: Washington, DC, USA, 1986; Volume 306, pp. 209–230. [[CrossRef](#)]
- Hoonakker, F.; Lachiche, N.; Varnek, A.; Wagner, A. A representation to apply usual data mining techniques to chemical reactions—Illustration on the rate constant of S_N2 reactions in water. *Int. J. Artif. Intell. Tools* **2011**, *20*, 253–270. [[CrossRef](#)]
- Lowe, D.M. *Extraction of Chemical Structures and Reactions from the Literature*; Technical Report, Apollo—University of Cambridge Repository: Cambridge, UK, 2012.
- Nugmanov, R.I.; Mukhametgaleev, R.N.; Akhmetshin, T.; Gimadiev, T.R.; Afonina, V.A.; Madzhidov, T.I.; Varnek, A. CGRtools: python library for molecule, reaction, and condensed graph of reaction processing. *J. Chem. Inf. Mod.* **2019**, *59*, 2516–2521. [[CrossRef](#)] [[PubMed](#)]
- Zhang, C.; Arun, A.; Lapkin, A.A. Completing and balancing database excerpted chemical reactions with a hybrid mechanistic-machine learning approach. *ACS Omega* **2024**, *9*, 18385–18399. [[CrossRef](#)]
- Phan, T.L.; Weinbauer, K.; Gärtner, T.; Merkle, D.; Andersen, J.L.; Fagerberg, R.; Stadler, P.F. Reaction rebalancing: A novel approach to curating reaction databases. *J. Cheminf.* **2024**, *16*, 82. [[CrossRef](#)]

20. Andersen, J.L.; Flamm, C.; Merkle, D.; Stadler, P.F. Inferring chemical reaction patterns using graph grammar rule composition. *J. Syst. Chem.* **2013**, *4*, 4. [CrossRef]
21. Andersen, J.L.; Flamm, C.; Merkle, D.; Stadler, P.F. A software package for chemically inspired graph transformation. In Proceedings of the Graph Transformation, ICGT 2016, Vienna, Austria, 5–6 July 2016; Echahed, R., Minas, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2016; Volume 9761, pp. 73–88. [CrossRef]
22. IUPAC Chemical Nomenclature and Structure Representation Division; Brecher, J. Graphical representation standards for chemical structure diagrams (IUPAC Recommendations 2008). *Pure Appl. Chem.* **2008**, *80*, 277–410. [CrossRef]
23. Mann, M.; Thiel, B. Kekulé structure enumeration yields unique SMILES. In Proceedings of the WCB13: Workshop on Constraint Based Methods for Bioinformatics, Uppsala, Sweden, 16–20 September 2013; pp. 57–65.
24. Connolly Martin, Y. Let's not forget tautomers. *J. Comput. Aided. Mol. Des.* **2009**, *23*, 693–704. [CrossRef]
25. Sayle, R.A. So you think you understand tautomerism? *J. Comput. Aided. Mol. Des.* **2010**, *24*, 485–496. [CrossRef]
26. González Laffitte, M.E.; Stadler, P.F. Progressive multiple alignment of graphs. *Algorithms* **2024**, *17*, 116. [CrossRef]
27. Harary, F. *Graph Theory*; CRC Press: Boca Raton, FL, USA, 2018.
28. Horn, F.J.M. Necessary and sufficient conditions for complex balancing in chemical kinetics. *Arch. Ration. Mech. Anal.* **1972**, *49*, 172–186. [CrossRef]
29. Feinberg, M. Complex balancing in general kinetic systems. *Arch. Ration. Mech. Anal.* **1972**, *49*, 187–194. [CrossRef]
30. Dugundji, J.; Ugi, I. An algebraic model of constitutional chemistry as a basis for chemical computer programs. *Topics Curr. Chem.* **1973**, *39*, 19–64. [CrossRef]
31. Lynch, M.F.; Willett, P. The automatic detection of chemical reaction sites. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 154–159. [CrossRef]
32. Funatsu, K.; Endo, T.; Kotera, N.; Sasaki, S.I. Automatic recognition of reaction site in organic chemical reactions. *Tetrahedron Comput. Methodol* **1988**, *1*, 53–69. [CrossRef]
33. Fujita, S. Imaginary transition structures. A novel approach to computer oriented representation of organic reactions. *J. Synth. Org. Chem. Japan* **1989**, *47*, 396–412. [CrossRef]
34. Flamm, C.; Müller, S.; Stadler, P.F. Every atom-atom map for neutral molecules can be explained by electron pair pushing diagrams. *Discr. Math. Chem.* **2024**, in press. [CrossRef]
35. Hendrickson, J.B. Descriptions of reactions: Their logic and applications. *Red. Trav. Chim. Pays-Bas* **1992**, *111*, 323–334. [CrossRef]
36. Fujita, S. A novel approach to the enumeration of reaction types by counting reaction-center graphs which appear as the substructures of imaginary transition structures. *Bull. Chem. Soc. Japan* **1988**, *61*, 4189–4206. [CrossRef]
37. 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. [CrossRef]
38. Beier, N.; Gatter, T.; Stadler, P.F. Converting KEGG RCLASS data into DPO graph rewriting rules. 2024, in preparation.
39. Cordella, L.P.; Foggia, P.; Sansone, C.; Vento, M. A (sub)graph isomorphism algorithm for matching large graphs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1367–1372. [CrossRef] [PubMed]
40. Alpár Jüttner, A.; Madarasi, P. VF2++—An improved subgraph isomorphism algorithm. *Discr. Appl. Math.* **2018**, *242*, 69–81. [CrossRef]
41. Jochum, C.; Gasteiger, J.; Ugi, I. The principle of minimum chemical distance (PMCD). *Ang. Chem. Intl. Ed.* **1980**, *19*, 495–505. [CrossRef]
42. Takapoui, R.; Boyd, S. Linear programming heuristics for the graph isomorphism problem. Technical Report 1611.00711. *arXiv* **2016**, arXiv:1611.00711. [CrossRef]
43. Linderoth, J.T.; Ralphs, T.K. Noncommercial software for mixed-integer linear programming. In *Integer Programming: Theory and Practice*; Karlof, J., Ed.; CRC Press: Boca Raton, FL, USA, 2005; pp. 253–303.
44. CGRTools, Read the Docs: Data Types and Operations with Them. Available online: https://cgrtools.readthedocs.io/tutorial/1_data_types_and_operations.html#Bonds-has-order-and-p_order-attribute (accessed on 3 June 2024).
45. Behr, N.; Krivine, J. Compositionality of rewriting rules with conditions. *Compositionality* **2021**, *3*, 2. [CrossRef]
46. Dhaked, D.K.; Guasch, L.; Nicklaus, M.C. Tautomer database: A comprehensive resource for tautomerism analyses. *J. Chem. Inf. Model.* **2020**, *60*, 1090–1100. [CrossRef]
47. Huff, M.; Telford, D. Lord of the rings—The mechanism for oxidosqualene: Lanosterol cyclase becomes crystal clear. *Trends Pharmacol. Sci.* **2005**, *26*, 335–340. [CrossRef] [PubMed]
48. Dehmer, M.; Chen, Z.; Emmert-Streib, F.; Shi, Y.; Tripathi, S.; Musa, A.; Mowshowitz, A. Properties of graph distance measures by means of discrete inequalities. *Appl. Math. Model.* **2018**, *59*, 739–749. [CrossRef]
49. Bonnici, V.; Giugno, R.; Pulvirenti, A.; Shasha, D.; Ferro, A. A subgraph isomorphism algorithm and its application to biochemical data. *BMC Bioinform.* **2013**, *14*, S13. [CrossRef]
50. Aparo, A.; Bonnici, V.; Micale, G.; Ferro, A.; Shasha, D.; Pulvirenti, A.; Giugno, R. Fast subgraph matching strategies based on pattern-only heuristics. *Interdiscip. Sci.* **2019**, *11*, 21–32. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.