Quantum chemical calculations to trace back reaction paths for the prediction of reactants

Yosuke Sumiya[1], Yu Harabuchi[1,2,3], Yuuya Nagata[2,3], Satoshi Maeda[1,2,3,4]*

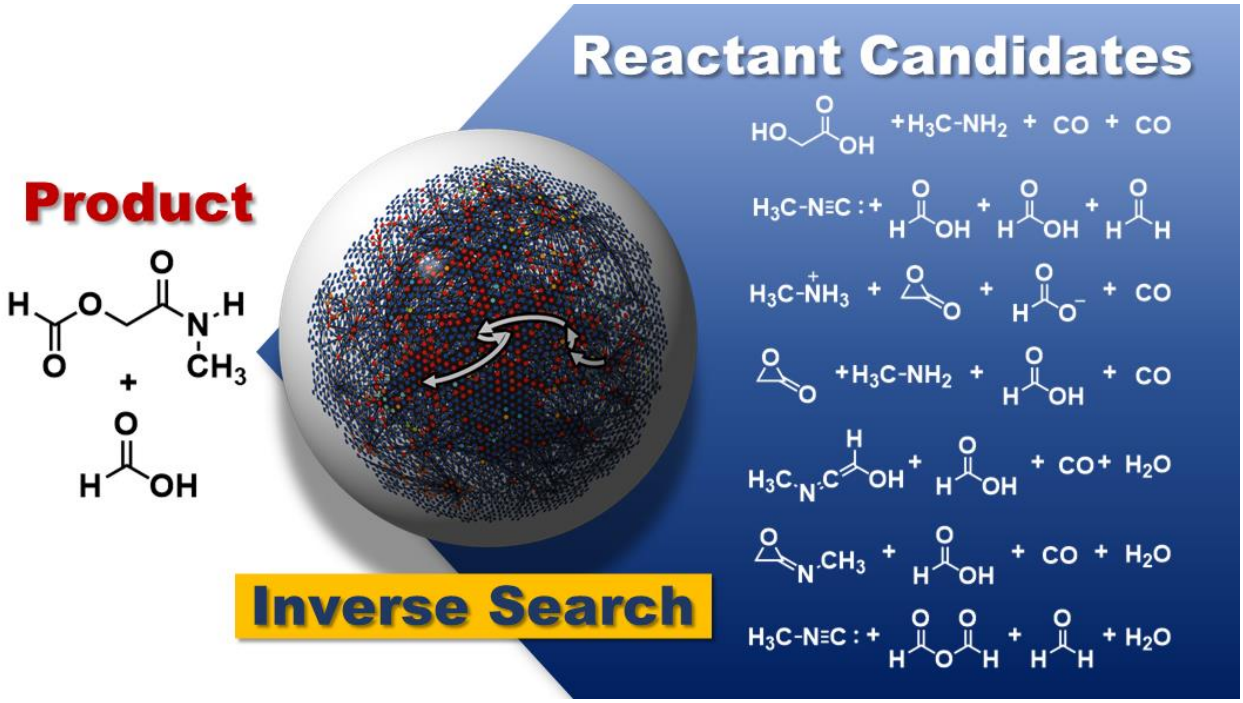
[1] Department of Chemistry, Graduate School of Science, Hokkaido University; Sapporo, Hokkaido 060-0810, Japan.

[2] Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University; Sapporo, Hokkaido 001-0021, Japan.

[3] ERATO Maeda Artificial Intelligence for Chemical Reaction Design and Discovery Project, Hokkaido University; Sapporo, Hokkaido 060-0810, Japan

[4] Research and Services Division of Materials Data and Integrated System (MaDIS), National Institute for Materials Science (NIMS); Tsukuba, Ibaraki 305-0044, Japan

*Corresponding author. Email: smaeda@eis.hokudai.ac.jp (S.M.)



Abstract:

The long-due development of a computational method for the *ab initio* prediction of chemical reactants that provide a target compound has been hampered by the combinatorial explosion that occurs when reactions consist of multiple elementary reaction processes. To address this challenge, we have developed a quantum chemical calculation method that can enumerate the reactant candidates from a given target compound by combining an exhaustive automatic reaction path search method with a kinetics method for narrowing down the possibilities. Two conventional name reactions were then assessed by tracing back reaction paths using this new method to determine whether the known reactants could be identified. Our method is expected to be a powerful tool for the prediction of reactants and the discovery of new reactions.

Since Schrödinger first reported his now fundamental equation describing wave mechanics nearly a century ago, the development of quantum chemical calculation methods has further revealed the behaviour of atoms during chemical reactions.[1–4] Quantum chemical calculations continue to be used to elucidate the behaviours of atoms. This provides a more detailed understanding of the mechanisms involved in known chemical reactions and establishes mechanistic insights that often lead to the rational design of new chemical reactions.[5–9] Such calculations have typically been performed by solving the Schrödinger equation for the motion of electrons at a series of nucleus configurations along a reaction path.[10,11] To this end, one generally needs to first project the motion of atoms along the path beforehand. The projected motion is then assessed to determine whether it is energetically feasible by solving the Schrödinger equation for the motion of electrons along this path. By performing such calculations for various paths and comparing their kinetic and thermodynamic preferences, one can identify the actual motion of atoms during a chemical reaction. The full automation of this procedure to systematically predict chemical reactions has been intensively investigated for many years.[12-19] In several cases, such methods have been successful in predicting the reaction path from a given initial state (e.g., initial mixture of reactants, additives, and catalysts) to a product, without relying on any knowledge concerning the path or the product.

The next challenge to overcome involves predicting the initial state of a chemical reaction by tracing back reaction paths using quantum chemical calculations, which is equivalent to solving an inverse problem of a chemical reaction. Once a systematic method is established, one could predict a method for synthesizing a target compound

without relying on any previous synthetic knowledge or database. We introduced such a design concept in 2013 as quantum chemistry-aided retrosynthetic analysis (QCaRA),[14] which has been utilized for predicting a very simple reaction that involves only one elementary step.[20] However, a combinatorial explosion of the number of possible paths has been a major obstacle for generalizing the concept to multistep cases and applying it to various chemical reactions. The number of paths for possible reactant candidates increases dramatically depending on the number of elementary steps from the product state. Moreover, the reactant candidates are usually less stable than the product state; therefore, a comprehensive search must be performed to find a feasible path leading to highly unstable compounds.

This study provides a solution to the aforementioned problem by combining a kinetic analysis method with an automated reaction path search method. The artificial force induced reaction (AFIR) method,[21] which is employed herein as an automated reaction path search method, is capable of exhaustive searches that include paths to highly unstable compounds.[22] The AFIR method eliminates energy barriers by applying a virtual force between fragments X and Y in a system to induce chemical transformation (Fig. 1a), where single atoms or small groups of atoms can be chosen as X and Y. Depending on the choice of fragments X and Y, different chemical transformations can be produced; two possibilities are illustrated in Fig. 1 for forces applied between X and Y, viz. cyanide anion dissociation upon application of a negative force (Fig. 1b) and water addition by applying a positive force (Fig. 1c). By applying either a negative or a positive force to a variety of fragment pairs, various stable structures are constructed.

After a systematic exploration in this manner, a network of stable structures and their transformation pathways is produced to establish a reaction path network.

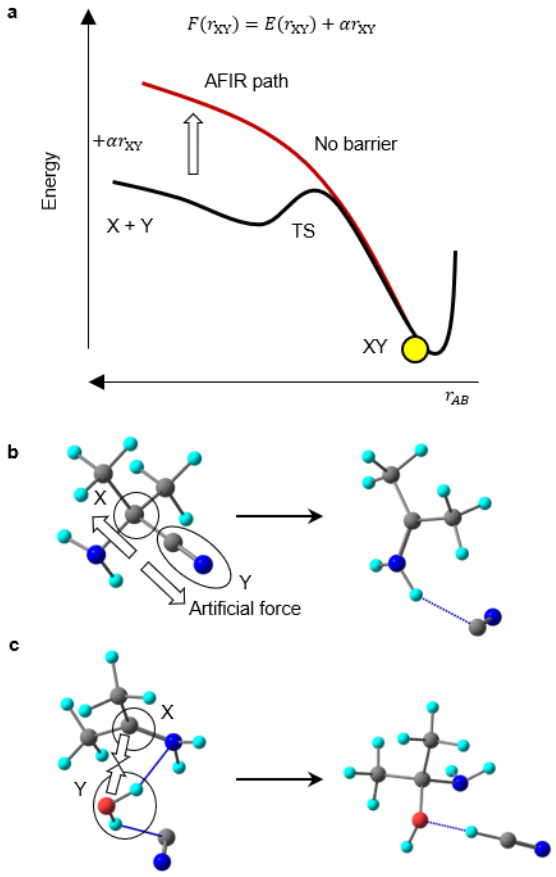


Fig. 1 | Schematic illustrations of how the AFIR method induces chemical transformations. a, Bond formation reaction between X and Y induced by the AFIR method. A function $F(r_{XY})$ consisting of the potential energy surface $E(r_{XY})$ and the force term αr_{XY} is minimized, where r_{XY} is the X-Y distance and α is a parameter regarding the force strength and direction. **b,** Bond dissociation reaction induced by the AFIR method. By setting α as negative and X and Y as the central C and CN, respectively, X and Y (circled) were pulled apart with a negative force. **c**, Bond formation reaction induced by the AFIR method. By setting α as positive and X and Y as the iminium C and H_2O, respectively, X and Y (circled) were pushed together with a positive force. Atoms shown in **b** and **c** are H (light blue), C (grey), N (dark blue), and O (red).

Without imposing any restrictions, the AFIR method tries to find all possible stable structures. Since the number of stable structures increases exponentially with increasing number of atoms in the system,[12,19] the computational cost to perform a search targeting all possible stable structures is high, even for systems composed of less than ~20 atoms. Hence, exploration in this study is guided by a kinetic analysis method called rate constant matrix contraction (RCMC).[23] The RCMC method performs a coarse-graining of a reaction path network by integrating stable structures that reach a thermal equilibrium within a given timescale (t_{MAX}). Fig. 2a shows an energy profile that consists of the minima of seven stable structures, with the minimum indicated as "Start" corresponding to the energy of the product state; thus, exploration by the AFIR method in this study starts from there. The RCMC method is then used to integrate the shallowest minimum into its adjacent minima with a certain ratio to yield a reaction profile shown in Fig. 2b. This procedure called "contraction" is applied to all shallow minima with lives shorter than t_{MAX}. Fig. 2c displays the final coarse-grained profile on which all the shallow minima have been contracted to the three deep minima. Each of the three deep minima called a "superstate" (indicated as "SS") is expressed as the sum of the main minimum to which the contraction has not been applied and all the minima contracted to it. Since the RCMC method determines the contribution ratio of each minimum to each superstate so as to reproduce the thermal equilibrium within t_{MAX},[23] the ratio directly corresponds to the reaction yield obtained after the time t_{MAX}. In other words, by using the RCMC method, the yield of the target product obtained by every reaction starting from a minimum can be computed as its contribution ratio to the superstate in which the product state is the main component.

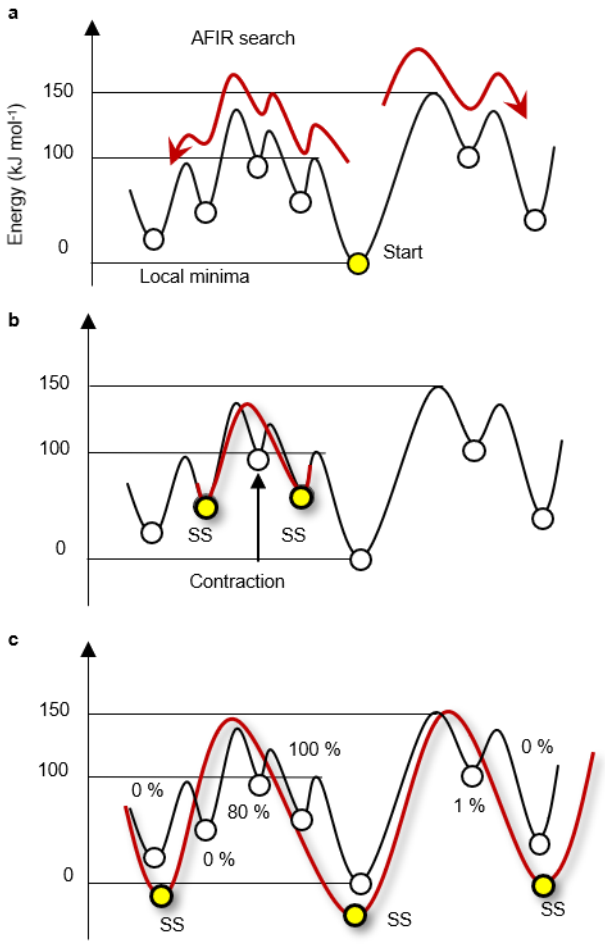


Fig. 2 | Schematic illustrations of how the RCMC method evaluates the kinetic importance of each local minimum. a, Energy profile consisting of seven local minima obtained by the AFIR method, starting from the minimum corresponding to the product state that is indicated as “Start”. **b,** Energy profile after applying a contraction to the shallowest local minimum and integrating it to the adjacent minima with certain ratios. **c,** Energy profile obtained after integrating all minima with shorter lives than t_{MAX} (i.e., one day for the applications described below) to the three deep minima by the RCMC method. The original profile consisting of the seven minima was simplified assuming equilibration within t_{MAX} in the three superstates (SSs). The contribution ratio of the original minimum to the SS that includes the “Start” point corresponds to the probability that a reaction starting from a minimum will then fall to the SS after equilibration of t_{MAX}, thus providing the yield of the state at the bottom of the SS, i.e., the product state. The AFIR search is then performed again starting from local minima with yields greater than the threshold value.

By combining the AFIR method with the RCMC method, paths to attain the product state can be selectively explored. First, the AFIR method searches for reaction paths starting from the product state to find minima adjacent to the minimum of the product state. Next, among these minima, the RCMC method identifies minima contributing to the superstate of the product state greater than a predefined threshold. Then the AFIR method is once again applied to continue the reaction path exploration from the minima that satisfy this threshold. Through alternately searching by AFIR and screening for minima by RCMC, and repeating this procedure until a termination condition is met, the paths leading to the target product can be selectively searched.

This study applied the above-mentioned combined AFIR/RCMC method to two known reactions,[24-26] i.e., the Strecker reaction (**R1**) and the Passerini reaction (**R2**) shown in Fig. 3, using each product as an input structure. Although the second HCOOH in **R2** is not stoichiometrically required, it was added due to its positive impact on the reaction as a catalyst.[27,28] The reaction temperatures were set at 200, 300, and 400 K, and the timescale for t_{MAX} was set to one day. When a stable structure gave the target product with a reaction yield of ≥0.1% at one or more of these temperatures, the path search continued beyond the stable structure. The solvent effects of water in **R1** and tetrahydrofuran in **R2** were taken into account by using an implicit solvation model. Further details of the search conditions are discussed in the Supplementary Information. The obtained stable structures were classified according to chemical species using the SMILES notation.[29] It should be emphasized that information related to the reactants of **R1** and **R2** were not included in these procedures. In other words, the focus of this study is to determine whether the AFIR/RCMC method is capable of finding the known

reactants. In the following, it is demonstrated that the AFIR/RCMC method has been successful in identifying reactants for both **R1** and **R2**, by tracing back multistep paths inverse way starting from the corresponding products.

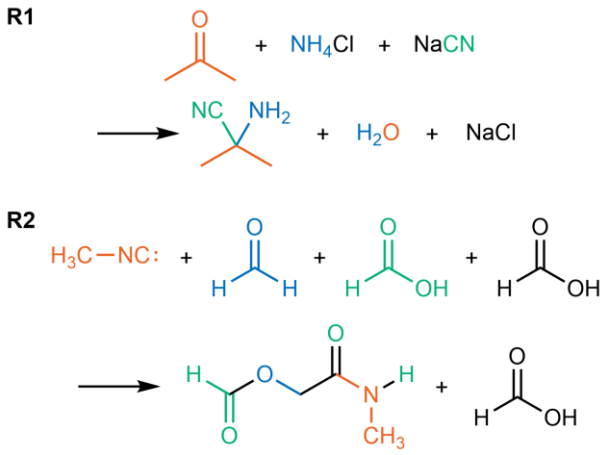


Fig. 3 | The target reactions explored in this study. The Strecker reaction (**R1**). The Passerini reaction (**R2**).

Results

Starting from the products of **R1** and **R2**, the searches found 9,208 and 12,219 stable structures, respectively, which were then classified into 1,679 and 2,407 chemical species, respectively, using their SMILES representations. Figs. 4a and 4b show the reaction path networks for **R1** and **R2** consisting of 1,679 and 2,407 nodes, respectively. Each node represents a different chemical species, and the energy of the most stable structure among those classified is indicated according to the colour scheme in the legend. These nodes are linked by 2,934 and 4,018 edges that represent reaction paths. As shown in these networks, the search yielded structures in a wide

energy range of 800 kJ mol^{-1}. The nodes in Figs. 4c and 4d indicate the contribution ratio to the superstate of the product for each chemical species. In other words, reactions initiated from red nodes are predicted to provide the target product in 100% yield, while those from dark blue nodes do not generate the target product at all. As seen from these networks, a variety of species can act as reactants to produce the target products. In Figs. 4e and 4f, the node colours identify the minimum number of elementary steps required to reach the target product; since these appear to be predominantly green, it is clear that three or four steps occupies the largest area and is most common. Although the searches found paths that consist of up to eight elementary steps, such multistep paths did not produce reaction yields that can be understood by comparing Figure 4c with 4e and Figure 4d with 4f.

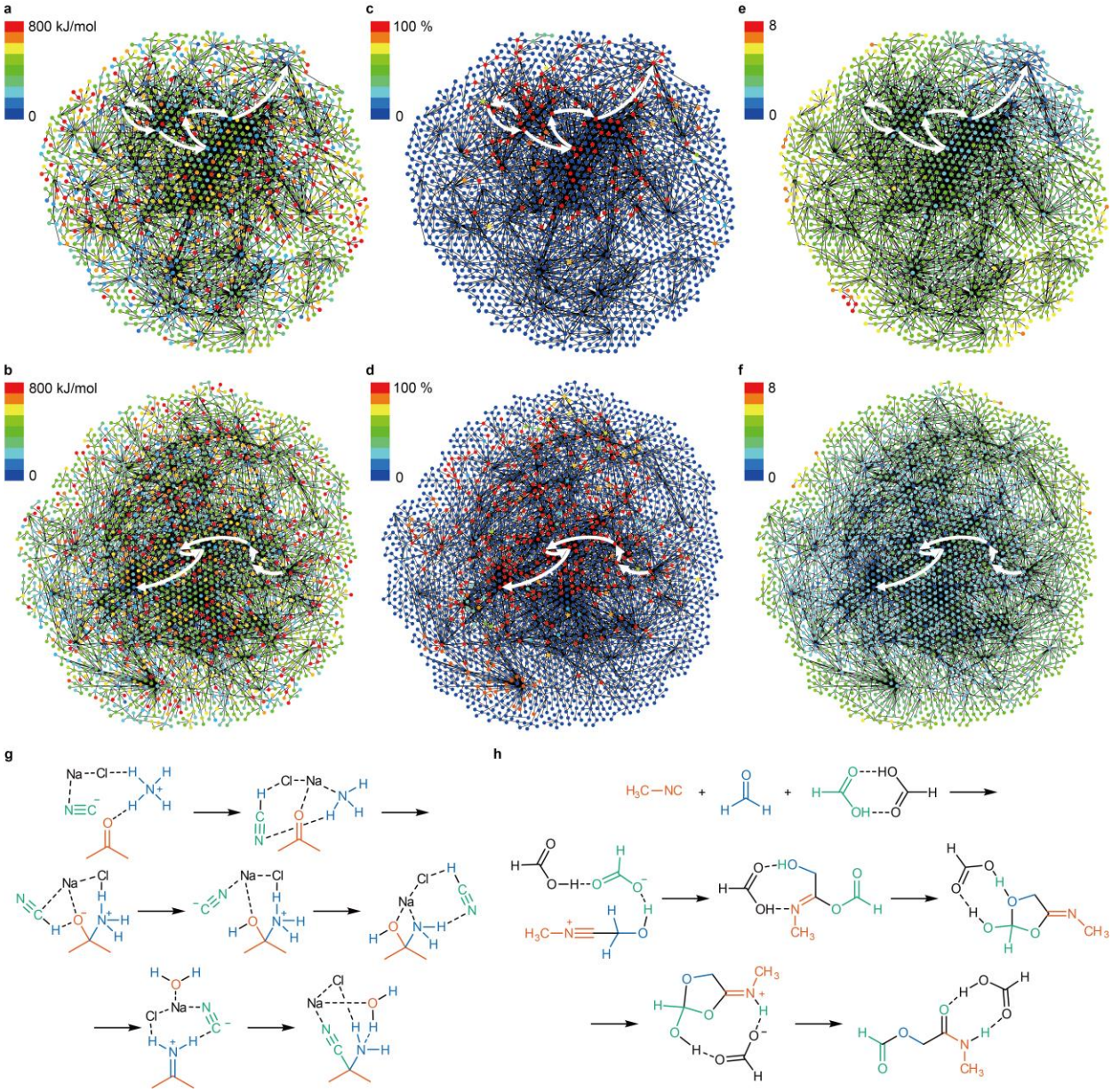


Fig. 4 | Reaction path networks for the Strecker reaction (R1) and the Passerini reaction (R2). Different nodes represent different chemical species. **a,** The network for **R1** showing the energy associated with each chemical species represented as a node. **b,** The network for **R2** showing the energy associated with each chemical species represented as a node. **c,** The network for **R1** with each node showing the contribution ratio to the product state (reaction yield of the product state). **d,** The network for **R2** with each node showing the contribution ratio to the product state (reaction yield of the product state). **e,** The network for **R1** with each node indicating the minimum number of elementary steps from the target product. **f,** The network for **R2** with each node indicating the minimum number of elementary steps from the target product. **g,** The

reaction path of **R1** as identified by its reaction path network highlighted by white arrows in **a**, **c**, and **e**. **h,** The reaction path of **R2** as identified by its reaction path network highlighted by white arrows in **b**, **d**, and **f**.

These reaction path networks include nodes representing the known reactants for **R1** and **R2**. In other words, the searches tracing back quantum chemical reaction paths were successful for both cases. In these networks, paths generating the products through well-known reaction mechanisms were identified (Figs. 4g and 4h), with each two-dimensional structure representing that of each node; the corresponding paths are highlighted with white arrows in the networks displayed in Fig. 4. As seen in Figs. 4c and 4d, these paths follow red nodes, thus indicating that these paths are kinetically feasible. In Figs. 4e and 4f, a majority of the nodes along the paths indicate that the target product is accessible in three steps from these nodes. However, the paths shown in Figs. 4g and 4h consist of six and five steps, respectively, thus demonstrating that fewer elementary steps do not correspond to kinetically feasible paths. For the path of **R1** in Fig. 4g, the reaction steps occur as follows: 1) a proton transfers from an ammonium cation to a cyanide anion to generate ammonia, 2) nucleophilic addition of ammonia to a ketone takes place, 3) proton transfer from hydrogen cyanide to a carbonyl oxygen atom occurs, 4) proton transfer from an ammonium cation to a cyanide anion generates a hemiaminal intermediate, 5) a proton transfers from hydrogen cyanide and subsequent dissociation of water generates an iminium cation, and finally 6) nucleophilic addition of a cyanide anion generates the target product, aminonitrile. For the path of **R2** in Fig. 4h, the reaction steps are as follows: 1) the nucleophilic addition of isocyanide to an aldehyde and the subsequent proton transfer from a

carboxylic acid to carbonyl oxygen atom generates a polar intermediate, 2) coupling between the charged moieties generates an imine intermediate, 3) the transfer of a proton promotes ring closure and generates a five-membered ring intermediate, 4) the transfer of a proton from a carboxylic acid to the imine nitrogen atom generates a polar intermediate, and finally 5) a proton transfer promotes ring opening and generates the target product, i.e., α-acyloxyamide. These paths are consistent with those suggested in the literature.[26-28] Therefore, the present calculations successfully predicted these synthetic paths based on a reverse path search from the corresponding products.

Discussion

Here, we further assess and discuss the species that were forecasted from the reaction path network of **R2**. Fig. 5a shows the 21 most stable species with a contribution ratio >50% at 300 K, and Fig. 5b lists the seven most stable species consisting of four molecules and having a contribution ratio >50% at 300 K. Below each species are listed the contribution ratios calculated at 250, 300, and 350 K with t_{MAX} = 1 day, i.e., their yields as reactants. These include some familiar reactions. First, S23 is a reactant of the Passerini reaction, while S15 and S20 are intermediates of the Passerini reaction, respectively, which are included in Fig. 4h as well. Therefore, it is obvious that they give the products of the Passerini reaction in high yields. Likewise, the conversion of S4 to α-acyloxyamide is a simple enol-keto tautomerism, and the conversion of S6 to α-acyloxyamide is a simple condensation of an acid anhydride and an alcohol. The anhydrous formic acid and H_2O in S28 are readily converted into two molecules of formic acid, which is equivalent to reactant S23 in the Passerini reaction. The correct

predictions of such indisputable chemical transformations confirm that the method is able to systematically predict possible chemical transformations.

On the other hand, there are many cases in which the corresponding experiments are difficult in terms of stability and solubility of the reactants. For example, it is difficult to prepare single-molecule species like S3, S5, S8, S9, S10, S16, S18, and S21 as reactants because they spontaneously change into α-acyloxyamide. Molecules in species S1, S2, S11, S14, and S19, which contain formic acid or $HCOO^-$, may also spontaneously convert to α-acyloxyamide because they have the same chemical composition as α-acyloxyamide or its protonated form. Species such as S12, S13, and S22 contain acid-base pairs; therefore, they easily crystallize and precipitate. Furthermore, S24, S25, S26, and S27 contain molecules that are difficult to isolate, such as acetolactone in S24. Among these, S7 and S17 seem to be experimentally feasible as new reactions, but it was found that formyl 2-hydroxyacetate and formyl 2-formyloxyacetate contained in them are not commercially available or synthesized. Many of the chemical reactions predicted by this method are difficult to demonstrate experimentally; however, it is an important finding that the experimentally established Strecker and Passerini reactions were included among them. It is also important to note that a new reaction was actually found in the previous application limited to one step,[20] producing a $R\text{-}CF_2\text{-}COO^-$ skeleton in one-pot from CO_2 enabled syntheses of a series of compounds that were difficult to synthesize.[30] In the future, we would like to systematically apply this method to various compounds to discover unknown chemical reactions.

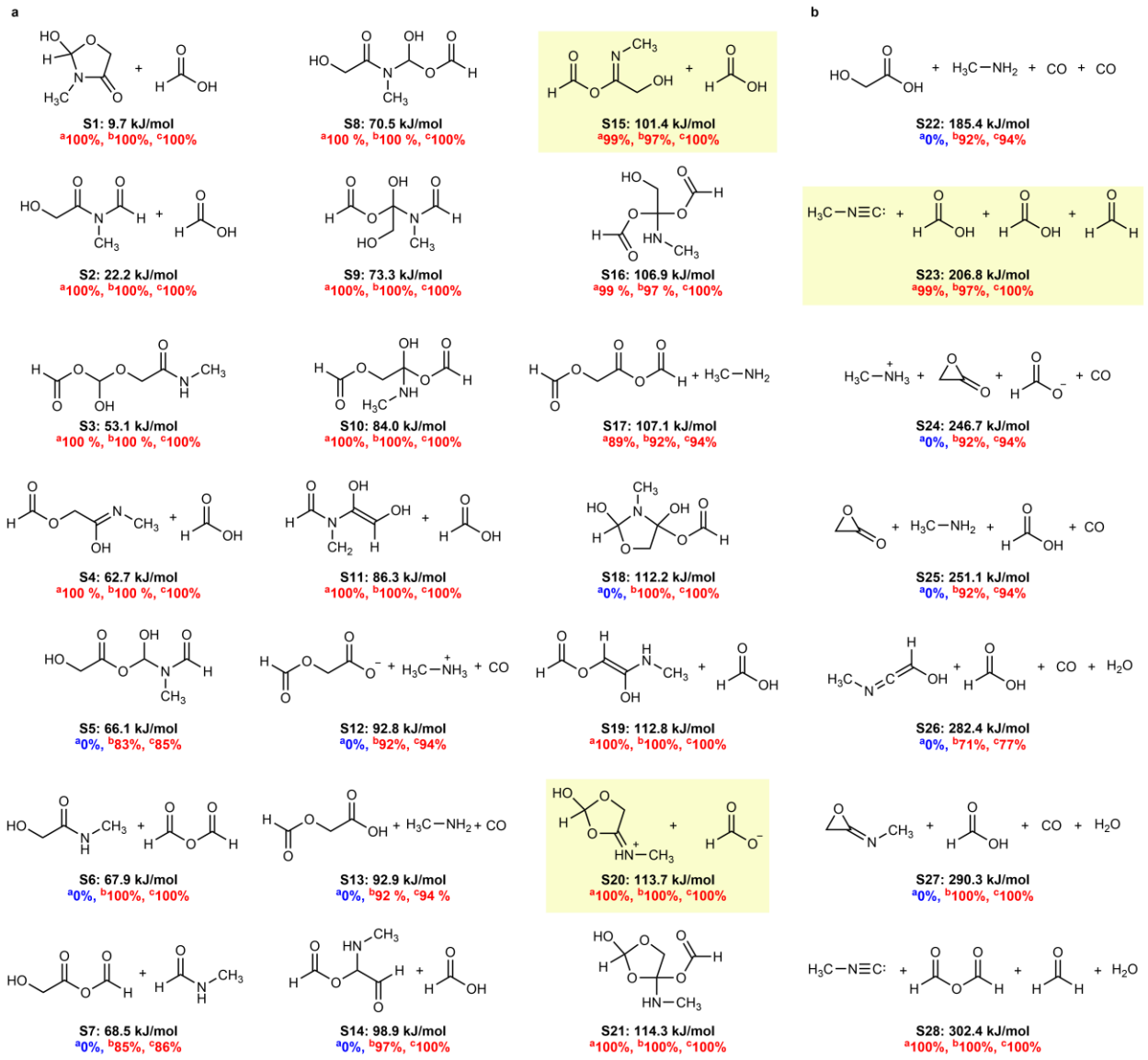


Fig. 5 | Chemical species with a contribution ratio to the product state (reaction yield of the product state) >50% at 300 K with t_{MAX} = 1 day for R2. a, The 21 most stable species. **b,** The seven most stable species consisting of four molecules. Gibbs energy at 300 K relative to the product state, and the contribution ratio to the product state at [a]250 K, [b]300 K, and [c]350 K are shown below each species.

Finally, we would like to discuss the advantages and limitations of the backward search for synthetic paths consisting of multiple elementary reaction processes achieved by this study. First of all, this method allows us to predict chemical reactions

starting from products and going backward to reactant candidates for the first time by tracing back quantum chemical reaction paths. The most important benefit of this method is that it does not require any previous knowledge or data, which makes it a powerful and innovative computational method that breaks away from the conventional norms of chemistry. Unfortunately, the search based on quantum chemical calculations involves relatively high computational costs. For the two reactions calculated in this study, it took approximately four days using 1,344 cores (Xeon Platinum 2.3 GHz). Since the computational cost of quantum chemical explorations increases exponentially with the number of atoms in the system, the current targets are limited to only simple chemical reactions such as those described here. Another limitation is that only the prediction of reaction paths that can be described by the set of atoms in the microsystem is currently possible. In other words, if a reaction path involves a catalyst, then the catalyst molecule must be explicitly included in the calculation. Furthermore, macroscopic phenomena such as the precipitation of acid-amine salts are ignored. These limitations must be overcome or mitigated in the future in order to expand the possibilities of this method. However, even with the current limitations, it is anticipated that this method will be used to discover new reactions. In the future, we plan to work on applying this method to the discovery of new reactions, while working toward mitigation of the application limits in parallel. With these efforts, we continue to develop QCaRA into an essential chemical reaction design concept complementary to computer-assisted approaches that rely on experimental data.[31-38]

References:

1. Houk, K. N. & Cheong, P. H.-Y. Computational prediction of small-molecule catalysts. *Nature* **455**, 309–313 (2008).
2. Thiel, W. Computational catalysis – Past, present, and future. *Angew. Chem. Int. Ed.* **53**, 8605–8613 (2014).
3. Sameera, W. M. C., Maeda, S. & Morokuma, K. Computational catalysis using the artificial force induced reaction method. *Acc. Chem. Res.* **49**, 763–773 (2016).
4. Houk, K. N. & Liu, F. Holy grails for computational organic chemistry and biochemistry. *Acc. Chem. Res.* **50**, 539–543 (2017).
5. Ahn, S., Hong, M., Sundararajan, M., Ess, D. H. & Baik, M.-H. Design and optimization of catalysts based on mechanistic insights derived from quantum chemical reaction modeling. *Chem. Rev.* **119**, 6509–6560 (2019).
6. Rosales, A. R. et al. Rapid virtual screening of enantioselective catalysts using CatVS. *Nat. Catal.* **2**, 41–45 (2019).
7. Falivene, L. et al. Towards the online computer-aided design of catalytic pockets. *Nat. Chem.* **11**, 872–879 (2019).
8. Foscato, M. & Jensen, V. R. Automated in silico design of homogeneous catalysts. *ACS Catal.* **10**, 2354–2377 (2020).
9. Athavale, S. V., Simon, A., Houk, K. N. & Denmark, S. E. Demystifying the asymmetry-amplifying, autocatalytic behaviour of the Soai reaction through structural, mechanistic and computational studies. *Nat. Chem.* **12**, 412–423 (2020).

10. Schlegel, H. B. Geometry optimization. *WIREs Comput. Mol. Sci.* **1**, 790–809 (2011).

11. Maeda, S., Harabuchi, Y., Ono, Y., Taketsugu, T. & Morokuma, K. Intrinsic reaction coordinate: Calculation, bifurcation, and automated search. *Int. J. Quantum Chem.* **115**, 258–269 (2015).

12. Wales, D. J. *Energy Landscapes: With Applications to Clusters, Biomolecules and Glasses* (Cambridge University, Cambridge, England, 2003).

13. Schlegel, H. B. Exploring potential energy surfaces for chemical reactions: An overview of some practical methods. *J. Comput. Chem.* **24**, 1514–1527 (2003).

14. Maeda, S., Ohno, K. & Morokuma, K. Systematic exploration of the mechanism of chemical reactions: The global reaction route mapping (GRRM) strategy using the ADDF and AFIR methods. *Phys. Chem. Chem. Phys.* **15**, 3683–3701 (2013).

15. Wang, L.-P. et al. Discovering chemistry with an *ab initio* nanoreactor. *Nat. Chem.* **6**, 1044–1048 (2014).

16. Dewyer, A. L., Argüelles, A. J. & Zimmerman, P. M. Methods for exploring reaction space in molecular systems. *WIREs Comput. Mol. Sci.* **8**, e1354 (2018).

17. Grambow, C. A. et al. Unimolecular reaction pathways of a γ-ketohydroperoxide from combined application of automated reaction discovery methods. *J. Am. Chem. Soc.* **140**, 1035–1048 (2018).

18. Simm, G. N., Vaucher, A. C. & Reiher, M. Exploration of reaction pathways and chemical transformation networks. *J. Phys. Chem. A* **123**, 385–399 (2019).

19. Sumiya, Y. & Maeda, S. “Paths of chemical reactions and their networks: From geometry optimization to automated search and systematic analysis” in *Chemical Modelling*, M. Springborg, J.-O. Joswig, Eds. (Royal Society of Chemistry, London, 2019), vol. 15 pp. 28–69.

20. Mita, T., Harabuchi, Y. & Maeda, S. Discovery of a synthesis method for a difluoroglycine derivative based on a path generated by quantum chemical calculations. *Chem. Sci.* **11**, 7569–7577 (2020).

21. Maeda, S. & Morokuma, K. A systematic method for locating transition structures of A + B → X type reactions. *J. Chem. Phys.* **132**, 241102 (2010).

22. Maeda, S. & Harabuchi, Y. Exploring paths of chemical transformations in molecular and periodic systems: An approach utilizing force. *WIREs Comput. Mol. Sci.* e1538 (2021).

23. Sumiya, Y. & Maeda, S. Rate constant matrix contraction method for systematic analysis of reaction path networks. *Chem. Lett.* **49**, 553–564 (2020).

24. Strecker, A. Ueber die künstliche Bildung der Milchsäure und einen neuen, dem Glycocoll homologen Körper. *Liebigs Ann. Chem.* **75**, 27–45 (1850).

25. Passerini, M. Sopra gli isonitrili (I). Composto del *p*-isonitril-azobenzolo con acetone ed acido acetico. *Gazz. Chim. Ital.* **51**, 126–129 (1921).

26. Solomons, T. W. G. *Organic Chemistry*. (Wiley, New York, ed. 6, 1996).

27. Maeda, S., Komagawa, S., Uchiyama, M. & Morokuma, K. Finding reaction pathways for multicomponent reactions: The Passerini reaction is a four-component reaction. *Angew. Chem. Int. Ed.* **50**, 644–649 (2011).

28. Ramozzi, R. & Morokuma, K. Revisiting the Passerini reaction mechanism: Existence of the nitrilium, organocatalysis of its formation, and solvent effect. *J. Org. Chem*. **80**, 5652–5657 (2015).

29. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).

30. Hayashi, H. et al. Synthesis of difluoroglycine derivatives from amines, difluorocarbene, and CO_2: Computational design, scope, and application. *Chem. Eur. J.* (early view).

31. Milo, A., Neel, A. J., Toste, F. D. & Sigman, M. S. A data-intensive approach to mechanistic elucidation applied to chiral anion catalysis. *Science* **347**, 737–743 (2015).

32. Szymkuć, S. et al. *Angew. Chem. Int. Ed.* **55**, 5904–5937 (2016).

33. Segler, M. H. S., Preuss, M. & Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).

34. Coley, C. W., Green, W. H. & Jensen, K. F. Machine learning in computer-aided synthesis planning. *Acc. Chem. Res.* **51**, 1281–1289 (2018).

35. Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **361**, 360–365 (2018).

36. Zahrt, A. F. et al. Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. *Science* **363**, eaau5631 (2019).

37. Shields, B. J. et al. Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **590**, 89–96 (2021).

38. Shen, Y. et al. Automation and computer-assisted planning for chemical synthesis. *Nat. Rev. Methods Primers* **1**, 23, (2021).

Acknowledgments:

We thank Ms. Takako Homma for her help in draft editing. This work was supported by the JST via ERATO grant JPMJER1903. Support was also provided by the Institute for Chemical Reaction Design and Discovery (ICReDD), which was established by the World Premier International Research Initiative (WPI), MEXT, Japan.

Author contributions:

SM designed the computational framework and conducted the calculations. YS and SM developed the method. YH and YN analysed the reaction path networks. All authors prepared, reviewed, and approved the manuscript.

Competing interests:

The authors declare that they have no competing interests.

Additional information:

Supplementary information is available for this paper at https://XXXXX.