EvolvedComplexity as a Total Synthesis Assessment Metric: Strychnine as a Case Study of Scoring Functions

Abbigayle E. Cuomo[1]‡, John-Paul Webster[1]‡, H. Ray Kelly[2], Sumon Sarkar[1], Yu Shee[1], Sanil Sreekumar[2], Haote Li[1], Frederic Buono[2], Victor S. Batista[1], Timothy R. Newhouse[1]*

Address:

[1]Department of Chemistry, Yale University, New Haven, CT 06511, United States

[2]Chemical Development, Boehringer Ingelheim Pharmaceuticals Inc, Inc, 900 Ridgebury Road, Ridgefield, Connecticut 06877, United States

Abstract

The selection of synthetic routes to a small molecule of interest is enabled by the use of various tools to assess the chemical complexity of a given intermediate. While prior approaches assess the intrinsic molecular complexity or the facility with which an intermediate can be synthesized, in this study we introduce an alternative approach that tracks the progress towards the target structure in a given synthesis. A simple metric, EvolvedComplexity, was developed that compares the chemical similarity of a pair of molecules on the basis of the Tanimoto distance between chemical fingerprints. This complementary approach to assessing progress in synthesis may prove to be a useful tool for planning synthetic routes and for developing novel chemistries.

Introduction

Retrosynthetic planning is a cornerstone in the synthesis of complex natural products, where the challenge lies in deconstructing a target molecule into simpler, commercially available building blocks through a series of transforms.[1-3] This process, traditionally guided by the expertise of synthetic chemists, has increasingly been augmented by computer-aided synthetic planning (CASP) tools, an idea originally pioneered by EJ Corey in the 1960s that has seen a resurgence in recent decades owing to advances in computation.[1, 2, 4-7] Early CASP systems relied heavily on rule-based approaches, using expert knowledge to guide retrosynthesis.[3, 8-12] Recently CASP approaches have evolved from these rule-based methods to data-driven, predictive algorithms.[5, 13, 14] This transition was enabled by the automated extraction of reaction rules from vast chemical datasets, transforming how synthetic routes are identified.[15] These tools offer to improve the feasibility of a selected chemical route in the synthesis of a complex small molecule.

Automated retrosynthesis tools benefit from an assessment of molecular complexity in order to provide directionality to a retrosynthetic operation.[16] While human-guided retrosynthesis considers complexity reduction as a combination of the application of the logic of chemical synthesis and intuition, an automated approach requires an automated assessment. Each retrosynthetic step creates several more retrosynthetic possibilities leading to a tree whose search space increases exponentially with each additional step.[17] Selecting an individual pathway can be guided by molecular complexity where further consideration of given branches can be limited on the basis of molecular complexity (i.e. routes that lead to increasingly complex intermediates do not need to be considered further). Several methods for the evaluation of molecular complexity, hereafter referred to as Scoring Functions (SFs), have been developed to quantify complexity,

providing chemists with a metric to gauge the structural and synthetic challenges of a molecule.[17-19]

SFs can generally be divided into one of two categories: complexity-based methods and reaction-based methods.[17] Complexity-based methods employ rule-based systems to estimate the complexity of target structures. The first widely applicable index of molecular complexity was introduced by Bertz, who utilized principles of graph theory combined with information theory to analyze and quantify molecular topology (**Figure 1A**).[20] Graph theory treats molecules as mathematical graphs, where atoms are represented as vertices and bonds as edges.[21] This abstraction enables rigorous analysis of molecular topology, including connectivity, cyclic structures, and branching patterns, using combinatorial and algebraic methods. Information theory, on the other hand, quantifies the amount of information contained in the system, which involves assessing how diverse or ordered the arrangement of atoms and bonds is within a molecule.[22, 23] This approach evaluates how much information is required to describe the structure of a molecule, with more complex molecules containing higher information content due to their intricate patterns of bonds and stereochemistry.[23]

Later, Whitlock developed a metric for molecular complexity that was designed to emulate a chemist's chemical intuition.[24] It comprises a size metric (*S*) and complexity metric (*H*), and is calculated by evaluating the number of atoms, bonds, and specific structural features (e.g. stereocenters or rings) in the molecule. Building on Whitlock's foundation, Barone and Chanon sought to expand and refine this approach. Choosing to neglect the chiral term, as Bertz does, they also incorporate terms to account for substituents and ring size, two features that had not been accounted for previously in the Whitlock index.[25] The Synthetic Method Complexity Metric (SMCM) was developed shortly thereafter to overcome key limitations

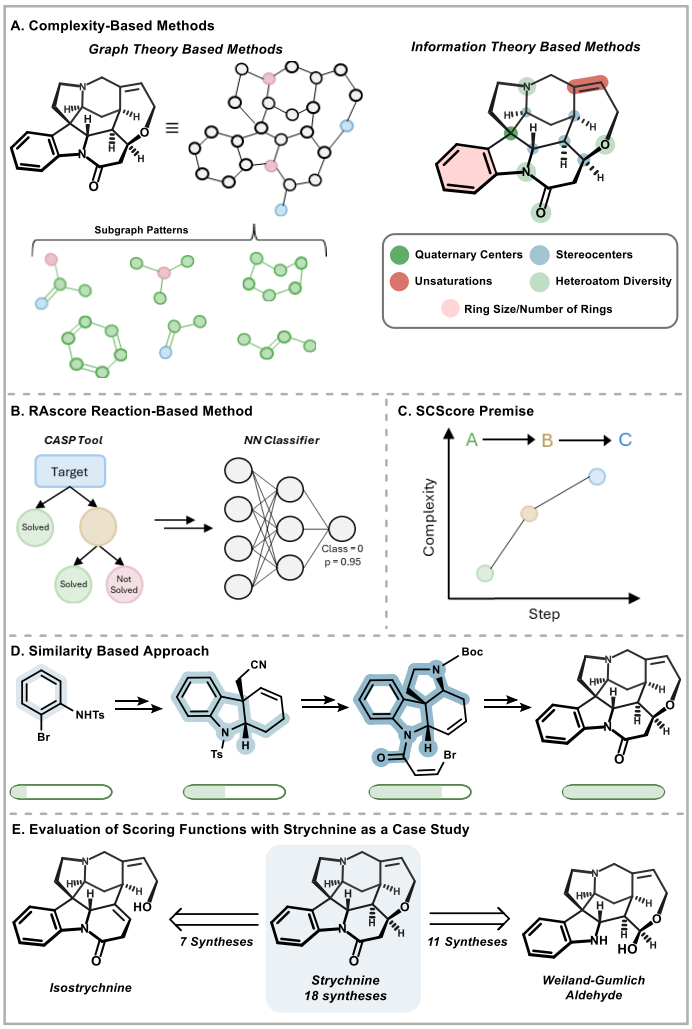

Figure 1. (A) Complexity-based methods for creating scoring functions include deriving functions from graph-based and/or information-based methods. (B) The reaction-based workflow used to develop RAscore. (C) The underlying reaction-based premise used to develop SCScore. (D) Similarity metrics for monitoring reaction progress. (E) A general representation of the 18 total syntheses of Strychnine proceeding via either the Isostrychnine or Weiland-Gumlich Aldehyde intermediate.

of earlier SFs, specifically addressing chirality, fused ring systems, and the presence of functional groups.[26] SMCM incorporates a unique set of multipliers based on the electronegativity of each atom, ring size and type, bond type, and substructures. This nuanced approach successfully minimizes the common bias that links higher molecular weight to greater complexity - a correlation that does not always hold true.

More recently, Böttcher drew from Bertz's systematic approach and Whitlock's intuitive method to create an alternative complexity index. This index measures molecular complexity by examining the information content of each atom's local environment.[23] By employing an entirely additive model, it mitigates the biases found in graph-theoretical methods (**Figure 1A**) and incorporates crucial aspects like symmetry and stereochemistry. This metric has been applied recently to demonstrate a marked increase in molecular complexity in challenging transformations, underscoring the utility of their newly developed methods for achieving these complex reactions more efficiently.[27, 28] In an effort to correlate complexity to biologically relevant properties, Waldmann and co-workers defined Spacial Score (SPS), an index engineered to closely mimic the fraction of sp^3-hybridized carbons (F_{sp3}) and the fraction of stereogenic carbons ($F_{Cstereo}$).[29] The intentional inclusion of certain molecular properties creates a SF that scales with the relative complexity of the molecule's skeletal structure. Ertl and Schuffenhauer introduced a novel approach to molecular complexity with the Synthetic Accessibility Score (SAscore). They assign fragment scores to common substructures from the PubChem database, with easier-to-synthesize fragments receiving lower scores and rarer, more complex fragments getting higher ones. The total fragmentScore is the sum of these individual scores, further adjusted by a complexity penalty for challenging structural features. SAscore has proven to be a valuable tool, having been successfully applied in the planning, design, and synthesis of numerous inhibitors targeting a variety of drug candidates.[30, 31]

An alternative approach to evaluating complexity involves reaction-based SFs. Instead of defining molecular complexity solely based on atomic properties, this approach ranks how readily a molecule can be synthesized given a certain set of chemical transforms (**Figure 1C**). Coley et al. developed a neural network approach to develop the learned metric Synthetic Complexity Score (SCScore).[32] Another approach to a reaction-based SF is the Retrosynthetic Accessibility Score (RAscore) developed by Thakkar, Reymond and co-workers.[33] This unique approach utilizes a machine learning (ML) classifier trained on the outputs of the CASP tool, AiZynthFinder. Synthetic feasibility is assessed on the basis of the retrosynthetic routes suggested by AiZynthFinder (**Figure 1B**). After training, RAscore offered synthetic accessibility predictions around 4,500 times faster when running on a GPU compared to AiZynthFinder. This increased speed highlights the benefit of using scoring functions to streamline the identification of synthetically tractable routes.

The wide array of approaches and indices for measuring molecular complexity highlights that no single method has proven entirely satisfactory, as each comes with its own limitations. Common criticisms of synthetic accessibility scores include their oversimplification of molecular complexity, failure to account for chirality and stereochemistry, and bias toward larger molecules. Additionally, many approaches are disconnected from the extrapolation of known synthetic methods to new systems and to entirely novel chemistries, limiting their predictive power and alignment with real-world synthetic challenges. Scoring functions have been continually assessed

in many different contexts.[17, 18] In this study, we evaluated a wide range of scoring functions within the context of syntheses of the natural product strychnine, which has long served as a benchmark and training ground within the field of total synthesis.

Strychnine, an alkaloid derived from the seeds of the *Strychnos nux-vomica* tree, is renowned as a structurally complex and challenging natural product to synthesize.[34, 35] Its intricate molecular architecture includes a densely packed polycyclic framework with six fused rings, seven contiguous stereocenters, and a bridged bicyclic amine. The first total synthesis of strychnine was achieved by Woodward in 1954.[36] The synthesis involved 29 steps, many of which were pioneering at the time, including the use of strategic cyclization reactions and selective functional group manipulations.[37] Since then, strychnine has become a ruler to measure synthetic innovation, with multiple total syntheses having been completed over the decades.[38-40] The synthesis of strychnine remains a hallmark of advanced synthetic chemistry, symbolizing both the intellectual challenge and the art of constructing highly complex natural products. With 18 published synthetic routes to strychnine, the extensive diversity in chemical strategies to complete its synthesis make it an ideal benchmark for evaluating the performance of various SFs.[36, 41-63]

Results and Discussion

The ten scoring functions discussed earlier—SAscore, RAscore, SCScore, SPS, nSPS, Böttcher (C_m), Barone and Chanon (BC), Whitlock (S), Bertz (C_T), and SMCM—were applied to evaluate each intermediate across the 18 total syntheses of strychnine (detailed synthetic routes can be found in the SI). For the three convergent syntheses by Rawal, Fukuyama, and Vanderwal, empirical evaluations were conducted after the point of convergence to ensure consistency throughout the analysis. In order to make comparisons across scoring functions, which have a wide range of numerical outputs on different scales, all scores were normalized between 0 and 1. Here, a score of 0 indicates the least complex molecule according to a particular SF, while a score of 1 signifies the most complex molecule.

Several factors—such as atom, redox, and step economy, alongside overall yield and the cost of commercially available starting materials—play a crucial role in determining the efficiency of a synthetic route. These considerations are integral to retrosynthetic analysis, helping chemists design pathways that balance efficiency with practicality. Nevertheless, one fundamental guideline is that molecular complexity should, on average, increase progressively throughout a multistep synthesis.[20, 32] In retrosynthetic analysis, the goal is to simplify the target molecule into readily available starting materials through sequential transformations that decrease complexity at each step. This approach helps prioritize retrosynthetic disconnections during automated synthesis planning, ensuring each step effectively breaks down the structure while moving toward accessible precursors.

Molecular complexity plots (**see SI**) were generated for each intermediate in all of the syntheses. These visualizations provide a clear, two-dimensional representation of how each reaction step builds toward the final target molecule. Beyond illustrating the incremental contributions of each step, these plots also offer a useful baseline for comparing the different SFs, allowing for a more intuitive assessment of how well each function captures the progression of complexity throughout the synthesis. To evaluate the premise that molecular complexity should increase progressively over the course of a multistep synthesis, the slope and R^2 values for each scoring function were determined. Given the varying number of steps across different syntheses, an ideal slope was calculated for each, based on a theoretical linear progression of complexity from 1 to the total number of steps (n). The absolute deviation between this ideal slope and the actual slope derived from each SF was then measured, providing a quantitative comparison of how closely each synthesis adhered to a linear increase in complexity (**Figure 2A**). It should be noted that this simplification was made with the understanding that some syntheses may have one or a few steps that rapidly increase complexity or that overly complex intermediates can still be strategic.

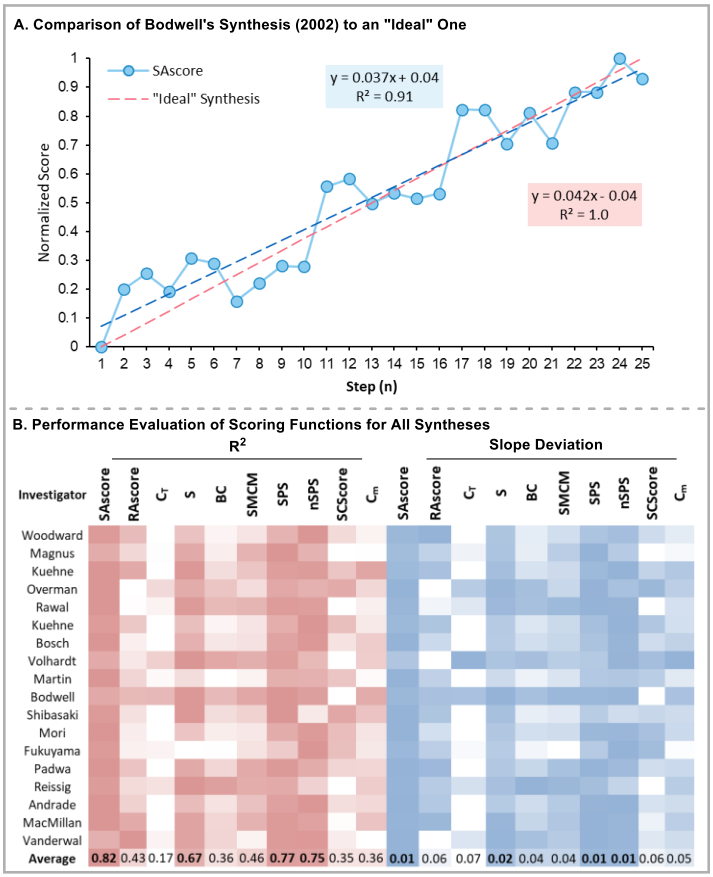

Figure 2. (A) General representation of the assumption that an ideal synthesis should follow a linear trend. The red line represents an ideal synthesis. The blue points and line represent the SAscore progression for the 2002 Bodwell synthesis (selected arbitrarily for demonstration). (B) Heat maps representing R^2 values (red) and the deviation from the ideal slope (blue). Higher R^2 and lower deviations are indicative of better scoring function performance (darker red/blue).

Heat maps were generated to visualize the R^2 and slope metrics for each SF (**Figure 2B**). For every synthesis each SF is associated with an R^2 value, where an R^2 of 1 represents a synthesis that follows a perfectly linear progression in complexity. Among the SFs, SAscore demonstrated the highest average R^2 of 0.82. SPS and nSPS followed with average R^2 values of 0.77 and 0.75, respectively. Whitlock's index also showed notable linearity, with an average R^2 of 0.67. A parallel trend emerged when examining deviations from the ideal slope. SAscore again exhibited the most consistent linearity, with an average deviation of just 0.006, while SPS, nSPS, and Whitlock showed slightly larger deviations, but still indicative of overall linear behavior.

While this type of analysis provides valuable insight, it is rarely sufficient for comparing SF due to the inherent complexity of chemical syntheses. In practice, synthetic routes often require the use of protecting groups, functional group interconversions, and non-strategic redox manipulations, which lead to deviations from the linear progression of molecular complexity.[64] These additional steps, though essential for achieving the target molecule, introduce variations that must be accounted for when evaluating the overall performance of a SF. Various strategies are typically employed when designing a synthesis, including introducing complexity either at the outset or toward the end of the sequence, or even generating excess complexity beyond that of the target molecule.[18] Our initial analysis focused on how each SF responded to the early introduction of complexity within the synthetic pathway, assessing how well they captured the increase in molecular complexity during the initial stages (**Figure 3**).

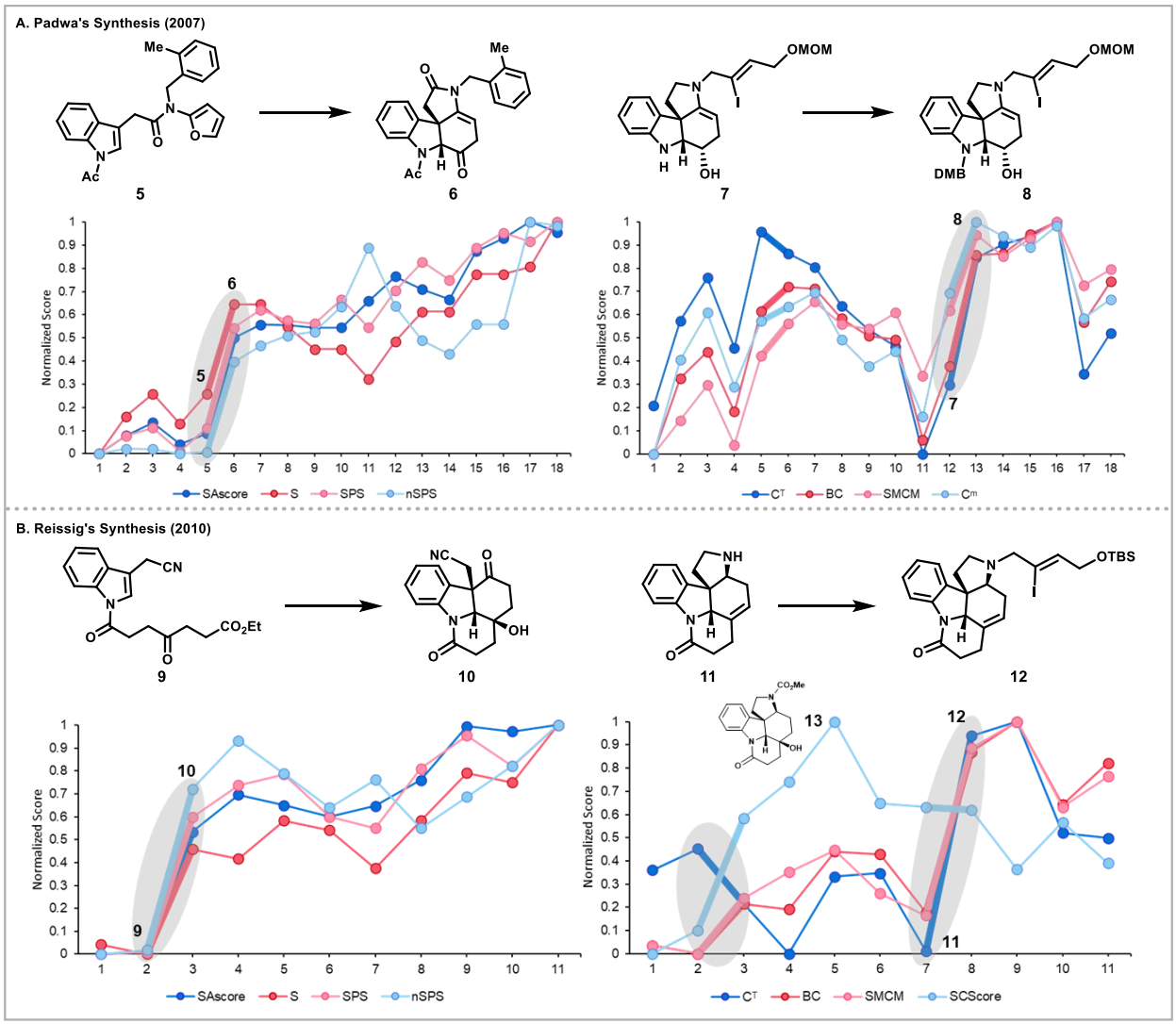

Figure 3. Examples of early-stage complexity introduction. (A) Example of SAscore, Whitlock's Index (S), SPS, and nSPS performance (left) compared to Bertz's Index (C_T), Barone and Chanon (BC), SMCM, and Böttcher's Index (C_m). (B) A similar comparison of scoring functions for Reissig's 2010 synthesis.

In 2007, Padwa reported a concise, enantioselective synthesis of strychnine, highlighting the critical role of an intramolecular [4 + 2] cycloaddition/rearrangement cascade of an indolyl-substituted amidofuran **5** to afford the aza-tetracycle **6** (**Figure 3A**).[57] Since this key reaction occurred early in the sequence, we examined how the SFs responded to the rapid increase in molecular complexity introduced at this stage. SAscore, SPS, nSPS, and the Whitlock index all capture this increase in molecular complexity from **5** to **6**. When complexity is introduced early in a synthesis, it often creates reactive functionalities that require careful mitigation throughout the remaining steps. SAscore, SPS, nSPS, and the Whitlock index effectively capture the fluctuations in complexity after the initial sharp increase, while still reflecting an overall linear upward trend. For the same synthetic step, the BC, SMCM, and Böttcher indices all show only a modest increase in complexity, while the Bertz index instead indicates a decrease in complexity. Interestingly, these 4 SFs showcase an increase in complexity from **7** to **8**. Upon inspection, this step is simply a DMB protection of the indole nitrogen. Of these 4 SFs, SMCM has the highest R^2 value of 0.73 (**Figure 2B**), yet it falls short in reflecting the change in complexity that would align with a chemist's intuitive ranking of the intermediates.

Reissig's 2010 synthesis of strychnine is another great example of the introduction of complexity at an early stage.[58] The authors denote a SmI_2-induced cascade reaction as the key step (**Figure 3B**) to afford the desired tetracycle **10**. Similar to the Padwa synthesis, SAscore, SPS, nSPS, and the Whitlock index all capture this increase in complexity while maintaining an overall linear progression in complexity for the remainder of the synthesis. While SCScore captures the sharp rise in complexity during this key step, it suggests that intermediate **13** is far more complex than the final target. The Bertz index, on the other hand, shows a decrease in complexity from intermediate **9** to **10**, while SMCM and BC reveal only modest increases. Notably, these three indices agree that the alkylation of intermediate **11** introduces the greatest complexity. This reaction, inspired by Rawal's 1994 synthesis and subsequently used in ten later syntheses[43], is consistently overemphasized by these SFs, exaggerating its complexity change compared to other innovative disconnections.

When developing Whitlock's index, it was proposed that longer syntheses should exhibit more linearity and experience smaller average changes in molecular complexity ($\Delta Score_m$) compared to shorter syntheses.[24] To evaluate this hypothesis, we analyzed two syntheses devoid of early complexity spikes, focusing on how the scoring functions responded to a gradual buildup of molecular complexity. The 2002 synthesis by Bosch[49] and the 1992 synthesis by Magnus[41] provide excellent examples (**Figure 4**). In both cases, SAscore, SPS, nSPS, and Whitlock's index showed a consistent linear increase in complexity. Conversely, RAscore, Bertz (C_T), SCScore, Böttcher (C_m), and Barone-Chanon (BC) exhibited more significant deviations across the synthetic route. Assuming a perfectly linear increase in complexity throughout a synthesis, the $\Delta Score_m$ should scale proportionally as 1/n, where n represents the number of steps. The 1/n curve was plotted by calculating the $\Delta Score_m$ for each synthesis of Strychnine. We visualized the behavior of each of the SFs and calculated the RMSE relative to the 1/n curve (**Figure 4C**). SAscore achieved the lowest RMSE (0.05), with SPS, nSPS, and Whitlock's index closely following, outperforming the remaining six scoring functions in this analysis. While this analysis may not be definitive and

could have some limitations, we showed that some SFs do not appear to be applicable to this context, and the most applicable SFs appear to have some disagreement.

In an effort to develop a scoring function that performed well across these varied methods for SF assessment, we considered an alternative approach that does not evaluate the inherent molecular complexity or the synthetic complexity. Instead, we were interested in developing a SF that assess how much progress towards the target structure is achieved in any given chemical step (**Figure 1D**). To do so, we applied the Tanimoto similarity metric to measure the amount of complexity that evolves over the course of a synthesis, defined here as EvolvedComplexity (EC).

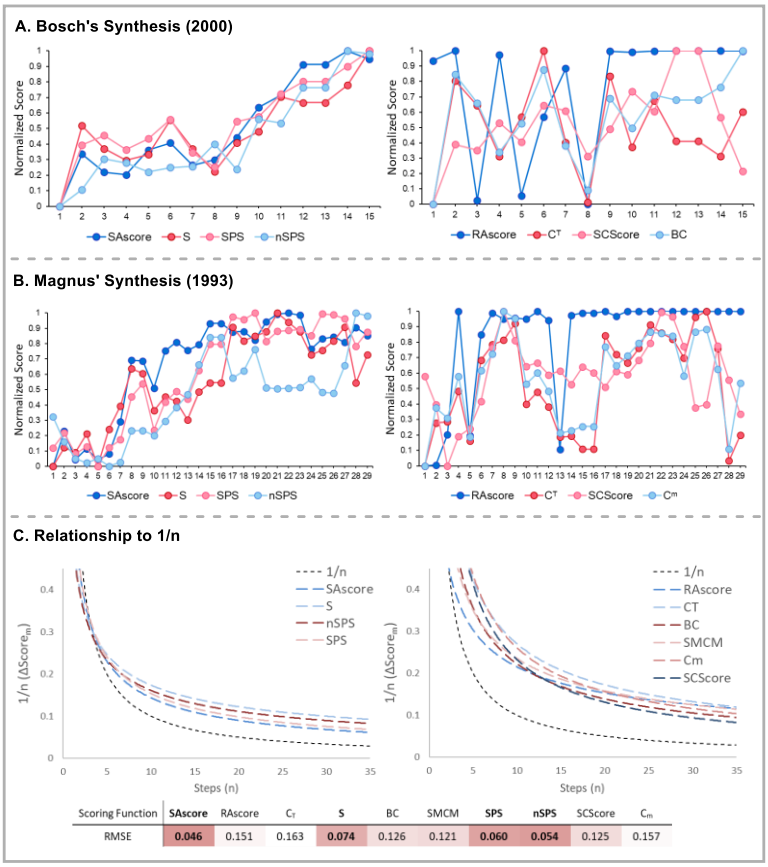

Figure 4. (A) Example of SAscore, Whitlock's Index (S), SPS, and nSPS performance (left) compared to Bertz's Index (C_T), Barone and Chanon (BC), RAscore, and SCScore (right) for a linear progression of a longer synthesis by Bosch. (B) A similar comparison for the 1993 synthesis by Magnus. (C) Comparison of SAscore, Whitlock's Index (S), SPS, and nSPS performance (left) to the remaining scoring functions' (right) relation to the ideal 1/n curve. The table displays the RMSE of each scoring function to the 1/n curve.

The Tanimoto similarity is a key metric in cheminformatics for evaluating the structural similarity of chemical compounds.[65] Derived from the Jaccard index, which measures the similarity and diversity of sets, the Tanimoto coefficient specifically assesses molecular fingerprints—binary vectors representing molecular structures (**Figure 5A**).[66-68] It calculates the ratio of shared features to total features between two vectors, yielding a similarity score from 0 (completely dissimilar) to 1 (identical). This score provides a valuable tool for chemists to compare compounds and explore their structural relationships in various contexts, including virtual screening and structure-activity relationship studies.[67, 68] Cernak and co-workers introduced a graph edit distance as a method to identify key

steps from automated outputs.[69] Relatedly, Genheden and Shields reported a method to compute a similarity score between any two routes towards the same target, enabling the filtering of redundant pathways generated by retrosynthetic algorithms.[70] We aimed to investigate the potential of employing Tanimoto similarity as a metric for tracking the progression of synthetic pathways over time. By quantifying the structural similarities between intermediates and the target molecule, we sought to gain insights into how these structural relationships mirrored the overall evolution of complexity throughout the synthesis. By treating the four selected scoring functions as benchmarks for evaluating molecular complexity in complex natural products, we used them as a reference to assess how well Tanimoto similarity scores perform as an alternative type of complexity metric. We began our analysis by calculating the Tanimoto similarity between each intermediate and the final target, strychnine. We evaluated several fingerprinting methods which include MACCS keys, MHFP6, RDKit, Avalon, Morgan, and Atom Pair fingerprints. To ensure consistency, similarities for each synthesis were normalized on a scale from 0 to 1. As an initial evaluation, we calculated the RMSE of these normalized Tanimoto similarities against the traditional SFs to determine if their behavior aligned with our previous assessments. **Figure 5B** compares the performance of Tanimoto fingerprinting methods against all SFs. MHFP6, Atom Pair, and Morgan fingerprints had the highest average RMSEs, both across all SFs and the top 4 performing SFs, suggesting a less optimal representation of molecular complexity in these cases. In contrast, Avalon and RDKit fingerprints consistently exhibited the lowest RMSEs in both categories, indicating that they are particularly well-suited for capturing the progression of molecular complexity during a synthesis. This strong alignment with the best-performing SFs suggests that these fingerprints possess advantageous properties for tracking structural changes throughout a synthetic sequence.

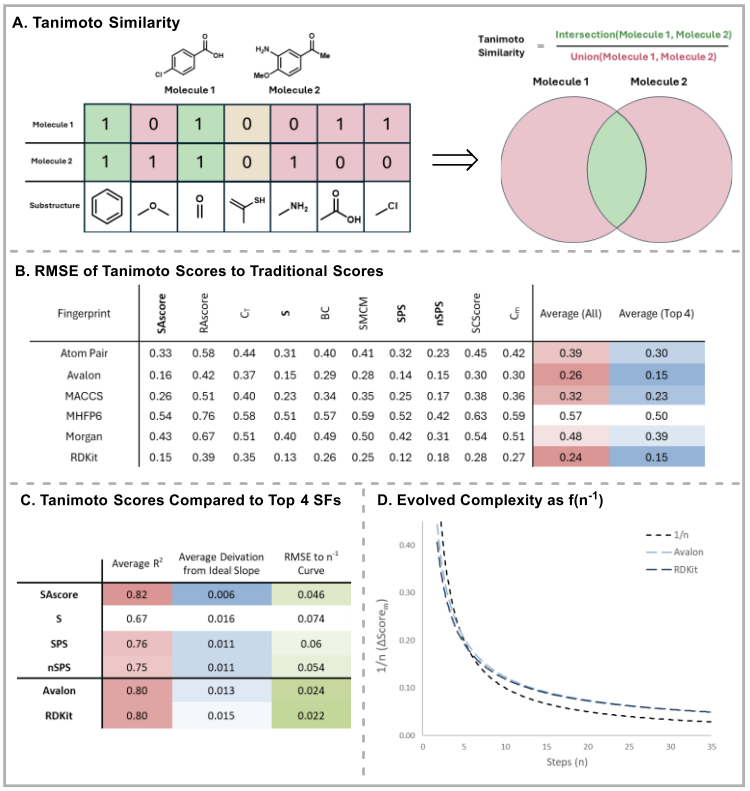

Figure 5. (A) A broad overview of calculating Tanimoto similarity scores from molecular fingerprints. The Tanimoto similarity is calculated by identifying common substructures found in each of the molecules. (B) The average RMSE across all 18 Strychnine syntheses between the various Tanimoto similarity scores based on different fingerprints and the SFs. The right shows the average RMSE for each fingerprinting method. Heat maps (red for all SFs, blue for top 4 SFs) show which fingerprints exhibited the lowest average RMSE (darker colors). (C) Average R^2, deviation from ideal slope, and RMSE to 1/n for the top 4 SFs and two best fingerprinting methods. (D) Representative 1/n curve for the top two fingerprinting methods.

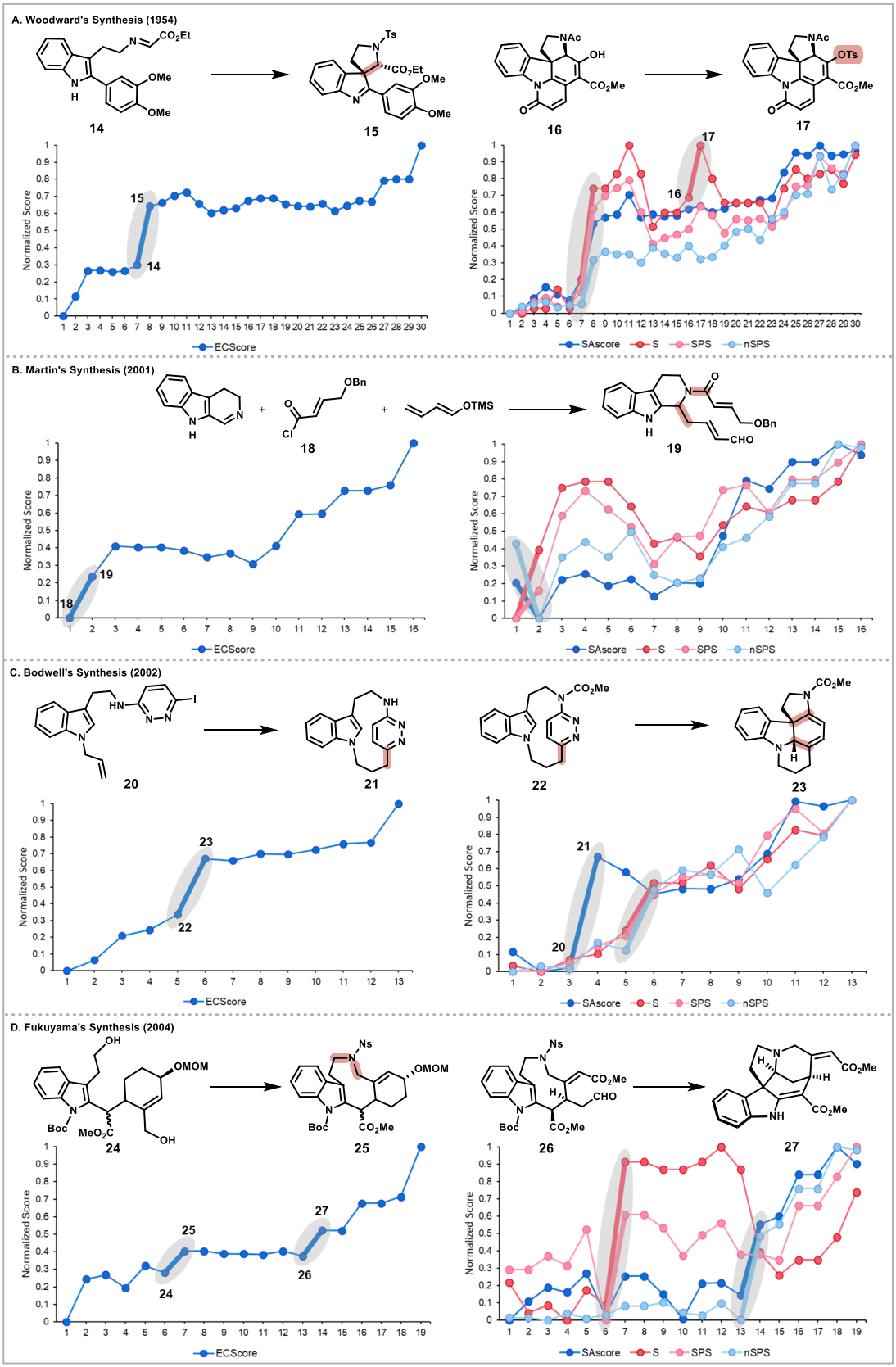

Figure 6. (A) A comparison of the RDKit fingerprinting method for Tanimoto similarity scores compared to SAscore, Whitlock's index (S), SPS, and nSPS. Highlighted here is the early-stage construction reaction and protecting group manipulation. (B) A similar analysis for the early-stage construction reaction in Martin's 2001 synthesis. (C) Analysis of SFs response to macrocycle formation in Bodwell's 2002 synthesis. (D) A similar analysis as to macrocycle formation and transannular cyclization for Fukuyama's 2004 synthesis.

Next, we assessed the average R^2, deviation from the ideal slope, and RMSE relative to the ideal 1/n curve. Once again, Avalon and RDKit fingerprints showed the best overall performance. Interestingly, while SAscore had a slightly higher R^2 value (0.82 compared to 0.80 for both fingerprinting methods), it exhibited a lower average deviation from the ideal slope—by nearly half—across all syntheses (**Figure 5C**). However, the fingerprinting methods achieved a lower RMSE to the 1/n curve (**Figure 5D**). Given these nuanced differences, we explored individual chemistries from various syntheses to further analyze their behaviors. Since RDKit and Avalon fingerprints demonstrated comparable performance, we present only the RDKit results for clarity.

Beginning with Woodward's 1954 synthesis (**Figure 6A**), all SFs capture the important construction step from **14** to **15**. Interestingly, Whitlock's index indicates a sharp increase in complexity compared to all other SFs from **16** to **17**, which is simply a protecting group installation. While the SPS score also increases, albeit only slightly, all other SFs agree that little to no change in complexity occurs at this stage of the synthesis. For the first step of Martin's 2001 synthesis, an important construction step via a vinylogous Mannich reaction of **18** is employed to afford **19** (**Figure 6B**). The Tanimoto scores, SPS, and Whitlock's index all reflect an expected increase in molecular complexity, however SAscore and nSPS mark a decrease in complexity for this bond forming step.

In Bodwell's 2002 synthesis, a marked anomaly in SAscore behavior can be observed. The sequential hydroboration/intramolecular B-alkyl Suzuki-Miyaura cross-coupling reaction from **20** to cyclophane **21** triggered a significant spike in complexity according to SAscore, while other SFs showed little to no increase (**Figure 6C**). This pronounced shift can be attributed to SAscore's built-in complexityPenalty, which imposes a substantial penalty on the presence of macrocycles (rings of size ≥ 8). A similar pattern in the treatment of macrocycle formation is evident in Fukuyama's 2004 synthesis (**Figure 6D**). When diol **24** underwent the Mitsunobu reaction yielding the nine-membered Ns-amide **25**, both SAscore and the Tanimoto scores reflected the anticipated rise in molecular complexity. However, SPS and Whitlock's index indicated a dramatic surge in complexity at this step. During the key transannular cyclization to form the pentacyclic core of strychnine, **27**, SAscore, nSPS, and the Tanimoto scores successfully captured the increase in complexity, whereas Whitlock's index, surprisingly, reported a sharp drop for this pivotal C–C bond formation step.

While by no means exhaustive, the new approach to scoring functions focusing on similarity disclosed herein (EC), generally agrees with changes in molecular and synthetic complexity throughout the course of complex natural product syntheses. By capturing these structural transformations, EC provides an insightful reflection of the dynamic nature of these processes, aligning well with the nuanced progression of molecular changes.

Conclusion

In this study, we evaluated the utility of several molecular complexity SFs alongside a novel similarity-based metric, EvolvedComplexity, derived from Tanimoto similarity scores. Our analysis revealed that the four selected SFs - SAscore, SPS, nSPS, and Whitlock's index - excelled in capturing the intricate changes in molecular complexity during the synthesis of complex natural products, such as strychnine. However, no single metric has emerged as a universal standard for

evaluating the overall progress in molecular and synthetic complexity across a trajectory of a multistep synthesis.

By incorporating Tanimoto similarity into this framework, we aimed to explore how well a similarity-based approach could complement traditional SFs. Our findings suggest that EvolvedComplexity captures key structural changes in a manner consistent with established SFs, particularly in reactions involving significant structural reorganization, such as transannular cyclizations. While not a substitute for complexity scoring, that can easily compare different classes of small molecules, EC offers a promising complimentary tool for evaluating molecular evolution across synthetic pathways that may prove to be a useful metric for synthetic planning and cheminformatics studies.

References

(1) Corey, E. J.; Wipke, W. T. Computer-Assisted Design of Complex Organic Syntheses. *Science* **1969**, *166* (3902), 178-192.

(2) Elkin, M.; Newhouse, T. R. Computational chemistry strategies in natural product synthesis. *Chem. Soc. Rev.* **2018**, *47* (21), 7830-7844.

(3) Williams, C. M.; Dallaston, M. A. The Future of Retrosynthesis and Synthetic Planning: Algorithmic, Humanistic or the Interplay? *Aust. J. Chem.* **2021**, *74* (5), 291-326.

(4) Corey, E. J. The Logic of Chemical Synthesis - Multistep Synthesis of Complex Carbogenic Molecules. *Angew. Chem. Int. Edit.***1991**, *30* (5), 455-465.

(5) Coley, C. W.; Green, W. H.; Jensen, K. F. Machine Learning in Computer-Aided Synthesis Planning. *Acc. Chem. Res.* **2018**, *51* (5), 1281-1289.

(6) Shen, Y.; Borowski, J. E.; Hardy, M. A.; Sarpong, R.; Doyle, A. G.; Cernak, T. Automation and computer-assistedplanning for chemical synthesis. *Nat. Rev. Methods Primers* **2021**, *1* (23), 23.

(7) Molga, K.; Szymkuc, S.; Grzybowski, B. A. Chemist Ex Machina: Advanced Synthesis Planning by Computers. *Acc. Chem. Res.* **2021**, *54* (5), 1094-1106.

(8) Szymkuc, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A. Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chem. Int. Edit.* **2016**, *55* (20), 5904-5937.

(9) Judson, P. *Knowledge-based Expert Systems in Chemistry: Not Counting on Computers*; RSCPublishing, **2009**.

(10) Wipke, W. T.; Ouchi, G. I.; Krishnan, S. Simulation and Evaluation of Chemical Synthesis - Secs - Application of Artificial Intelligence Techniques. *Artif. Intell.* **1978**, *11* (1-2), 173-193.

(11) Gasteiger, J.; Hutchings, M. G.; Christoph, B.; Gann, L.; Hiller, C.; Low, P.; Marsili, M.; Saller, H.; Yuki, K. A New Treatment of Chemical-Reactivity - Development of Eros, an Expert System for Reaction Prediction and Synthesis Design. *Top. Curr. Chem.* **1987**, *137*, 19-73.

(12) Hendrickson, J. B. Systematic Synthesis Design .6. Yield Analysis and Convergency. *J. Am. Chem. Soc.* **1977**, *99* (16), 5439-5450.

(13) Todd, M. H. Computer-aided organic synthesis. *Chem. Soc. Rev.* **2005**, *34* (3), 247-266.

(14) Mitchell, J. B. O. Machine learning methods in chemoinformatics. *Wires. Comput. Mol. Sci.* **2014**, *4* (5), 468-481.

(15) Coley, C. W.; Rogers, L.; Green, W. H.; Jensen, K. F. Computer-Assisted Retrosynthesis Based on Molecular Similarity. *ACS Cent. Sci.* **2017**, *3* (12), 1237-1245.

(16) Gao, W. H.; Coley, C. W. The Synthesizability of Molecules Proposed by Generative Models. *J. Chem. Inf. Model.* **2020**, *60* (12), 5714-5723.
(17) Skoraczynski, G.; Kitlas, M.; Miasojedow, B.; Gambin, A. Critical assessment of synthetic accessibility scores in computer-assisted synthesis planning. *J. Cheminform.*. **2023**, *15* (1).
(18) Wright, B. A.; Sarpong, R. Molecular complexity as a driving force for the advancement of organic synthesis. *Nat. Rev. Chem.* **2024**, *8* (10), 776-792.
(19) Li, J.; Eastgate, M. D. Current complexity: a tool for assessing the complexity of organic molecules. *Org. Biomol. Chem.* **2015**, *13* (26), 7164-7176.
(20) Bertz, S. H. The 1st General Index of Molecular Complexity. *J. Am. Chem. Soc.* **1981**, *103* (12), 3599-3601.
(21) García-Domenech, R.; Gálvez, J.; de Julián-Ortiz, J. V.; Pogliani, L. Some new trends in chemical graph theory. *Chem. Rev.* **2008**, *108* (3), 1127-1169.
(22) Nalewajski, R. F.; Parr, R. G. Information theory, atoms in molecules, and molecular similarity. *P. Natl. Acad. Sci. USA* **2000**, *97* (16), 8879-8882.
(23) Böttcher, T. An Additive Definition of Molecular Complexity. *J. Chem. Inf. Model.* **2016**, *56* (3), 462-470.
(24) Whitlock, H. W. On the structure of total synthesis of complex natural products. *J. Org. Chem.* **1998**, *63* (22), 7982-7989.
(25) Barone, R.; Chanon, M. A new and simple approach to chemical complexity. Application to the synthesis of natural products. *J. Chem. Inf. Comp. Sci.* **2001**, *41* (2), 269-272.
(26) Allu, T. K.; Oprea, T. I. Rapid evaluation of synthetic and molecular complexity for in silico chemistry. *J. Chem. Inf. Model.* **2005**, *45* (5), 1237-1243.
(27) Dang, H. T.; Nguyen, V. D.; Haug, G. C.; Arman, H. D.; Larionov, O. V. Decarboxylative Triazolation Enables Direct Construction of Triazoles from Carboxylic Acids. *JACS Au* **2023**, *3* (3), 813-822.
(28) Grant, P. S.; Meyrelles, R.; Gajsek, O.; Niederacher, G.; Maryasin, B.; Maulide, N. Biomimetic Cationic Cyclopropanation Enables an Efficient Chemoenzymatic Synthesis of 6,8-Cycloeudesmanes. *J. Am. Chem. Soc.* **2023**, *145* (10), 5855-5863.
(29) Krzyzanowski, A.; Pahl, A.; Grigalunas, M.; Waldmann, H. Spacial Score-A Comprehensive Topological Indicator for Small-Molecule Complexity. *J. Med. Chem.* **2023**, *66* (18), 12739-12750.
(30) Mengist, H. M.; Zunera, K.; Fentahun, A. In silico screening of potential SARS-COV-2 MAIN protease inhibitors from thymus Schimperi. *Clin. Chim. Acta.* **2024**, *558*, 28-28.
(31) Ouassaf, M.; Bourougaa, L.; Al-Mijalli, S. H.; Abdallah, E. M.; Bhat, A. R.; Kawsar, S. M. A. Marine-Derived Compounds as Potential Inhibitors of Hsp90 for Anticancer and Antimicrobial Drug Development: A Comprehensive In Silico Study. *Molecules* **2023**, *28* (24). 8074.
(32) Coley, C. W.; Rogers, L.; Green, W. H.; Jensen, K. F. SCScore: Synthetic Complexity Learned from a Reaction Corpus. *J. Chem. Inf. Model.* **2018**, *58* (2), 252-261.
(33) Thakkar, A.; Chadimová, V.; Bjerrum, E. J.; Engkvist, O.; Reymond, J. L. Retrosynthetic accessibility score (RAscore) - rapid machine learned synthesizability classification from AI driven retrosynthetic planning. *Chem. Sci.* **2021**, *12* (9), 3339-3349.
(34) Beifuss, U. New Total Syntheses of Strychnine. *Angew. Chem. Int. Edit.***1994**, *33* (11), 1144-1149.

(35) Cannon, J. S.; Overman, L. E. Is There No End to the Total Syntheses of Strychnine? Lessons Learned in Strategy and Tactics in Total Synthesis. *Angew. Chem. Int. Edit.* **2012**, *51* (18), 4288-4311.
(36) Woodward, R. B.; Cava, M. P.; Ollis, W. D.; Hunger, A.; Daeniker, H. U.; Schenker, K. The Total Synthesis of Strychnine. *J. Am. Chem. Soc.* **1954**, *76* (18), 4749-4751.
(37) Bonjoch, J.; Solé, D. Synthesis of strychnine. *Chem. Rev.* **2000**, *100* (9), 3455-3482.
(38) He, W. G.; Wang, P. Y.; Chen, J. H.; Xie, W. Q. Recent progress in the total synthesis of alkaloids. *Org. Biomol. Chem.* **2020**, *18* (6), 1046-1056.
(39) Seeman, J. I.; Tantillo, D. J. From Decades to Minutes: Steps Toward the Structure of Strychnine 1910-1948 and the Application of Today's Technology. *Angew. Chem. Int. Edit.* **2020**, *59* (27), 10702-10721.
(40) Shibasaki, M.; Ohshima, T. Recent Studies on the Synthesis of Strychnine. *Alkaloids-Chem. Biol.* **2007**, *64*, 103-138.
(41) Magnus, P.; Giles, M.; Bonnert, R.; Kim, C. S.; Mcquire, L.; Merritt, A.; Vicker, N. Synthesis of Strychnine Via the Wieland-Gumlich Aldehyde. *J. Am. Chem. Soc.* **1992**, *114* (11), 4403-4405.
(42) Kuehne, M. E.; Xu, F. Total Synthesis of Strychnan and Aspidospermatan Alkaloids .3. The Total Synthesis of (+/-)-Strychnine. *J. Org. Chem.* **1993**, *58* (26), 7490-7497.
(43) Rawal, V. H.; Iwasa, S. A Short, Stereocontrolled Synthesis of Strychnine. *J. Org. Chem.* **1994**, *59* (10), 2685-2686.
(44) Knight, S. D.; Overman, L. E.; Pairaudeau, G. Synthesis Applications of Cationic Aza-Cope Rearrangements .28. Asymmetric Total Syntheses of (-)-Strychnine and (+)-Strychnine and the Wieland-Gumlich Aldehyde. *J. Am. Chem. Soc.* **1995**, *117* (21), 5776-5788.
(45) Kuehne, M. E.; Xu, F. Syntheses of strychnan- and aspidospermatan-type alkaloids. 10. An enantioselective synthesis of (-)-strychnine through the Wieland-Gumlich aldehyde. *J. Org. Chem.* **1998**, *63* (25), 9427-9433.
(46) Solé, D.; Bonjoch, J.; García-Rubio, S.; Peidró, E.; Bosch, J. Total synthesis of (-)-strychnine via the Wieland-Gumlich aldehyde. *Angew. Chem. Int. Edit.* **1999**, *38* (3), 395-397.
(47) Vollhardt, K. P. C.; Eichberg, M. J.; Dorta, R. L.; Lamottke, K. Total synthesis of strychnine via a cobalt-mediated [2+2+2] cycloaddition. *Abstr. Pap. Am. Chem. S.* **1999**, *218*, U36-U36.
(48) Eichberg, M. J.; Dorta, R. L.; Lamottke, K.; Vollhardt, K. P. C. The formal total synthesis of (±)-strychnine via a cobalt-mediated [2+2+2]cycloaddition. *Org. Lett.* **2000**, *2* (16), 2479-2481.
(49) Solé, D.; Bonjoch, J.; García-Rubio, S.; Peidró, E.; Bosch, J. Enantioselective total synthesis of Wieland-Gumlich aldehyde and (-)-strychnine. *Chem. Eur. J.* **2000**, *6* (4), 655-665.
(50) Ito, M.; Clark, C. W.; Mortimore, M.; Goh, J. B.; Martin, S. F. Biogenetically inspired approach to the alkaloids.: Concise syntheses of (±)-akuammicine and (±)-strychnine. *J. Am. Chem. Soc.* **2001**, *123* (33), 8003-8010.
(51) Bodwell, G. J.; Li, J. A concise formal total synthesis of (±)-strychnine by using a transannular inverse-electron-demand Diels-Alder reaction of a [3](1,3)indolo[3](3,6)pyridazinophane. *Angew. Chem. Int. Edit.* **2002**, *41* (17), 3261.
(52) Nakanishi, M.; Mori, M. Total synthesis of (-)-strychnine. *Angew. Chem. Int. Edit.* **2002**, *41* (11), 1934.
(53) Mori, M.; Nakanishi, M.; Kajishima, D.; Sato, Y. A novel and general synthetic pathway to Strychnos indole alkaloids: Total syntheses of (-)-tubifoline, (-)-dehydrotubifoline, and (-)-

strychnine using palladium-catalyzed asymmetric allylic substitution. *J. Am. Chem. Soc.* **2003**, *125* (32), 9801-9807.
(54) Ohshima, T.; Xu, Y. J.; Takita, R.; Shimizu, S.; Zhong, D. F.; Shibasaki, M. Enantioselective total synthesis of (-)-strychnine using the catalytic asymmetric Michael reaction and tandem cyclization (vol 124, pg 14546, 2002). *J. Am. Chem. Soc.* **2003**, *125* (7), 2014-2014.
(55) Kaburagi, Y.; Tokuyama, H.; Fukuyama, T. Total synthesis of (-)-strychnine. *J. Am. Chem. Soc.* **2004**, *126* (33), 10246-10247.
(56) Ohshima, T.; Xu, Y. J.; Takita, R.; Shibasaki, M. Enantioselective total synthesis of (-)-strychnine: development of a highly practical catalytic asymmetric carbon-carbon bond formation and domino cyclization. *Tetrahedron* **2004**, *60* (43), 9569-9588.
(57) Zhang, H. J.; Boonsombat, J.; Padwa, A. Total synthesis of (±)-strychnine via a [4+2]-cycloaddition/rearrangement cascade. *Org. Lett.* **2007**, *9* (2), 279-282.
(58) Beemelmanns, C.; Reissig, H. U. A Short Formal Total Synthesis of Strychnine with a Samarium Diiodide Induced Cascade Reaction as the Key Step. *Angew. Chem. Int. Edit.***2 010**, *49* (43), 8021-8025.
(59) Mori, M. Total Synthesis of Strychnine. *Heterocycles* **2010**, *81* (2), 259-292.
(60) Sirasani, G.; Paul, T.; Dougherty, W.; Kassel, S.; Andrade, R. B. Concise Total Syntheses of (±)-Strychnine and (±)-Akuammicine. *J. Org. Chem.* **2010**, *75* (10), 3529-3532.
(61) Jones, S. B.; Simmons, B.; Mastracchio, A.; MacMillan, D. W. C. Collective synthesis of natural products by means of organocascade catalysis. *Nature* **2011**, *475* (7355), 183-188.
(62) Martin, D. B. C.; Vanderwal, C. D. A synthesis of strychnine by a longest linear sequence of six steps. *Chem. Sci.* **2011**, *2* (4), 649-651.
(63) Sirasani, G.; Andrade, R. B. Total Synthesis of Alkaloids Akuammicine, Strychnine, and Leuconicines A and B. *Strat. Tactics Org. Sy.* **2013**, *9*, 1-44.
(64) Gaich, T.; Baran, P. S. Aiming for the Ideal Synthesis. *J. Org. Chem.* **2010**, *75* (14), 4657-4673.
(65) Bajusz, D.; Racz, A.; Heberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.*. **2015**, *7* (20).
(66) Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug. Discov. Today* **2006**, *11* (23-24), 1046-1053.
(67) Todeschini, R.; Consonni, V.; Xiang, H.; Holliday, J.; Buscema, M.; Willett, P. Similarity Coefficients for Binary Chemoinformatics Data: Overview and Extended Comparison Using Simulated and Real Data Sets. *J. Chem. Inf. Model.* **2012**, *52* (11), 2884-2901.
(68) Rácz, A.; Bajusz, D.; Héberger, K. Life beyond the Tanimoto coefficient: similarity measures for interaction fingerprints. *J. Cheminform.*. **2018**, *10*.
(69) Lin, Y. F.; Zhang, R.; Wang, D.; Cernak, T. Computer-aided key step generation in alkaloid total synthesis. *Science* **2023**, *379* (6631), 453-456.
(70) Genheden, S.; Shields, J. D. A simple similarity metric for comparing synthetic routes. *Digit. Discov.* **2024**. *1*(4), 46-53.