Hybridizing organic chemistry and synthetic biology reaction networks for optimizing synthesis routes

Chonghuan Zhang[1] *and Alexei A. Lapkin*[1, 2 *]

[1] Department of Chemical Engineering and Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge CB3 0AS, United Kingdom

[2] Cambridge Centre for Advanced Research and Education in Singapore, CARES Ltd, 1 CREATE Way, CREATE Tower #05-05, 138602 Singapore

∗ Corresponding author E-mail address: aal35@cam.ac.uk

Abstract

Computer assisted synthesis planning (CASP) accelerates the development of organic synthesis routes of pharmaceuticals and industrial chemicals. CASP tools are generally developed on the rules or data of synthetic chemistry which include some enzymatic reactions. However, synthetic biology offers a new degree of freedom through the potential to engineer new synthetic steps. In this work, we present a method to hybridise conventional organic synthesis and synthetic biology to guide synthesis planning. A section of organic reactions from Reaxys® database was combined with metabolic reactions from KEGG database as reactions pools. The combined database was used to assemble synthetic pathways from multiple building blocks to a target molecule. The routes assembly was performed using reinforcement learning, which was adapted to *learn* the values of molecular structures in synthesis planning, and to develop a policy model to suggest near-optimal multi-step synthesis route choices from the available reactions pool. To quantify the added value of synthetic biology in the hybrid routes, three policy model 'decision makers' were developed from the organic, biological and hybrid reactions pools respectively. The near-optimal synthetic routes planned from the three reactions pools were evaluated and compared to discuss the benefits of the hybrid synthetic chemistry plus synthetic biology decision space in reaction routes optimisation.

Keywords*:* cheminformatics; retrosynthesis; metabolic engineering; synthetic biology, reinforcement learning

Introduction

Retrosynthetic analysis,[1] first developed decades ago, plans the synthetic routes of pharmaceuticals and industrial chemicals by transforming a target molecule into simpler precursors until available building block molecules are reached. Computer assisted synthesis

planning (CASP) has gained great interests since it accelerates the development of retrosynthetic routes.[2, 3] The emergence of chemical reaction databases and the ability to mine data from historical literature and patents made it possible for the development of numerous data-driven CASP methods.[4] Recent successes in this field include: (*i*) generation of reaction networks based on graph theory,[5] using vertices to represent molecules and directed edges to represent reactions from reactants to products, to understand connectivity among molecules,[6-11] (*ii*) manual curation[12-14] and algorithmic extraction[15, 16] of reaction rules and templates from historical reactions to predict functional group transformations, (*iii*) template-free deep learning methods to learn from historical reactions and plan new transformations,[17-19] etc.

A seemingly parallel development is that of synthetic biology, in which cells are engineered to produce target molecules. Synthetic biology is increasingly drawing attention of organic chemists for two reasons: (*i*) it can improve routes redox efficiency by finding metabolic shortcuts for the key synthetic steps, and (*ii*) most enzymatic reactions are highly selective.[20] In addition, biochemical reactions are performed under mild operating conditions and usually with benign solvents, which may lower operational costs and reduce life cycle impact of syntheses.[21] Similar to the reaction network of organic synthesis, a map to visualise synthetic biology space has been summarised to guide the biosynthetic planning.[22] A number of pharmaceutical ingredients and bulk chemicals have been produced economically through metabolic engineering approaches.[23, 24]

With the knowledge of molecular transformations from organic chemistry and synthetic biology, retrosynthesis relies on multi-step decision making to select optimal reaction routes among all feasible molecular transformations based on criteria such as exergetic efficiency, E-factor, and etc.[4, 6, 25] In linear reaction routes, which include only one-to-one (reactant(s)-

product only) ‘wiring’ (using the networks’ jargon) of the reactions, the decision making could be done through exhaustive search of all possible reaction routes, and ranking of the routes based on a set of pre-defined criteria. However, in a topological-tree-styled reaction routes, which include multiple-to-multiple wiring reactions (including co-reactants and bi-products), the number of routes options increase exponentially with the increase in the number of synthetic branches and depth; exhaustive search becomes computationally expensive.[26] In order to improve the efficiency of route design the machine learning method of reinforcement learning has been proposed for the application in synthesis planning.[18, 26-28]

Reinforcement learning (RL) mimics how an intelligent ‘decision maker’ takes multi-step actions within a specific problem environment to maximise/minimise the cumulative rewards/penalties of the actions.[29] In synthetic planning, the selection of each reaction step within a path is a decision making step. With the given rules and criteria costs, a ‘decision maker’ starts synthetic planning by trail-and-error, and algorithmically learns from the simulated experience to perform better in next iteration (an ‘episode’ within the RL jargon). For example, Schreck et al.[26] trained a policy model to understand the potential costs of candidate reactions computed from reaction templates at a certain synthetic depth, and to select retrosynthetic pathways based on costs. The method was compared with a decomposition heuristic method[2] to prove its ability for synthetic planning. Similarly, in metabolic engineering, Koch et al.[27] presented a code named RetroPath RL, which uses Monte Carlo Tree Search (MCTS) reinforcement learning to rank metabolic reaction rules to enable the development of biosynthetic routes.

Despite the interest in chemoinformatics to combine multiple datasets to have comprehensive understanding of the chemical space,[3, 30] to the best of our knowledge, none has worked on

combining organic chemistry and synthetic biology databases and understand the benefits from both in retrosynthesis planning. In this work, we mined a section of historical reactions from Reaxys® database,[31] and all metabolic reactions from Kyoto Encyclopedia of Genes and Genomes (KEGG),[32] which is an open-source manually curated bioinformatic library. We compared the influence of the presence of organic synthesis and synthetic biology past reaction data in a dataset used for identification of retrosynthesis pathways of a curated set of drug molecules, which were believed to be difficult to synthesise. To evaluate the identified routes we used atom economy, the number of reaction steps and price of molecular building blocks as key quantifiable performance criteria. Reinforcement learning method from Schreck et al.[26] was adapted to build policy models to guide the search for retrosynthesis pathways. Different from other CASP tools, the synthetic pathways from this method were not assembled from predicted reactions (i.e., using reaction template[15] or the algorithm generated reactions[17]), but using historical published data. This reduced the operational uncertainty over the identified paths to enable us to focus on the key research question of this work - how much added value could synthetic biology bring to synthetic organic chemistry in multi-step syntheses?

Methods

Building blocks

The molecular building blocks of a target molecule are smallest precursors required to build up the structure of the target molecule. To make a collection of building blocks and their prices for synthetic planning, here we defined building blocks to be commercially or naturally available small molecules. The collection of commercially available building blocks came from 'buyable' molecules crawled from ChemSpace,[33] which is an online catalogue of small molecules. For most buyable molecules listed in ChemSpace, more than one price is given from multiple suppliers. We only chose the lowest price available for these molecules.

The naturally available building blocks are freely available cofactor metabolites from cell organism in metabolic reactions, such as ATP and NADPH, and a list of such molecules curated by Blaβ et al.[34] was used for the naturally available building blocks. The price for these naturally available molecules is zero. In the enzyme-based industrial processes, although the natually occuring molecules are free to acquire, some of these molecules, specifically co-factors, are difficult to recover and recycle, which makes them economically unviable.[35] The common industrial solutions include stoichiometric design that balances each cofactor occurring in the total pathway,[36] or integrating multiple pathways to link the generation/degradation of cofactors.[35] For simplicity, in this work we only consider the acquisition price of the building blocks to demonstrate the overall approach; the cost of separation and other process 'costs' of the syntheses are deliberately left out from the current study.

Figure **1** shows the price distribution of all building block molecules. Price of approximately $1/10^{th}$ of all molecules from ChemSpace range from 10^3 to 10^6 USD/g, which is unreasonably expensive. Therefore, these molecules were removed from the set of building blocks to lower the costs of the potential synthetic routes. In total, we selected 24,282 commercially and 451 naturally available building blocks respectively. A full list of building block molecules can be found in the Supporting Information (SI).

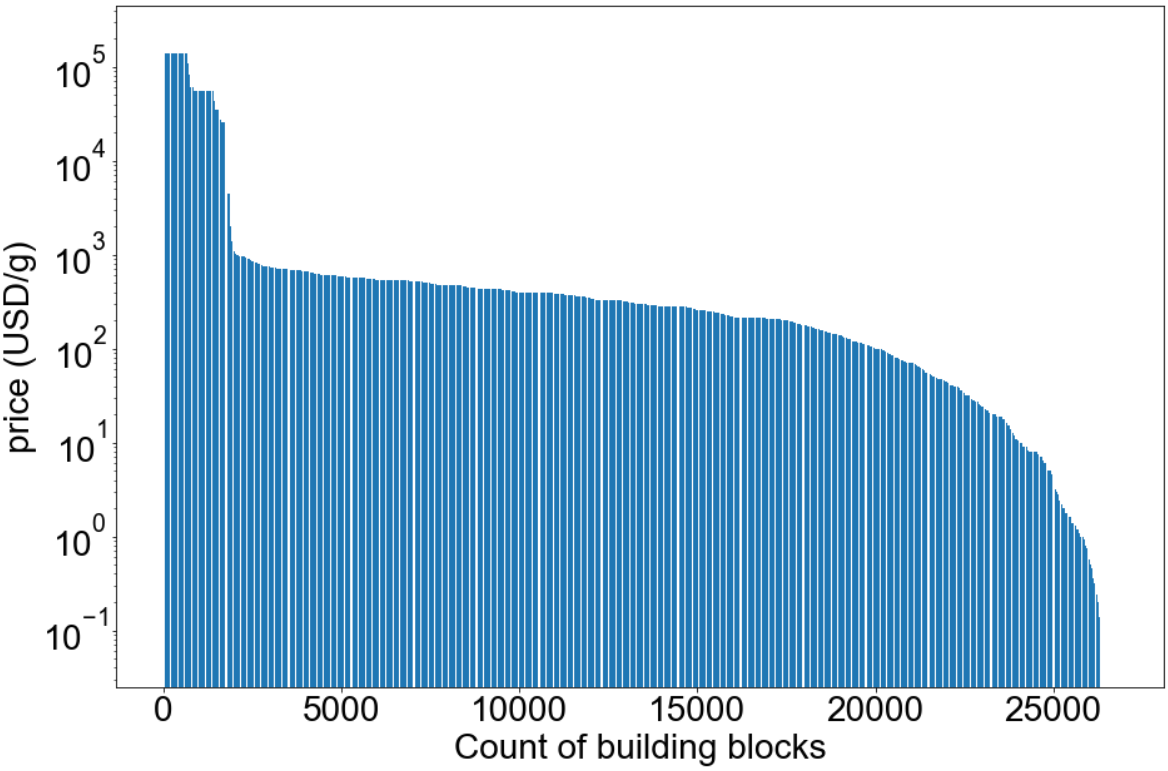


Figure 1. The price distribution of building block molecules, vertical axis shown in log-scale to distinguish the sharp corner at 10^3USD/g.

Mining and merging of Reaxys and KEGG datasets

A section of reactions was mined from Reaxys®. This is the same section as the molecules and reactions to generate the network of organic chemistry (NOC) from our previous studies.[6-10] All reactions were mined from KEGG reaction database. Since in most cases enzymes catalyse metabolic reactions from both directions, all reactions were assumed to be reversible. Therefore, for all metabolic reactions, both directions were recorded in the local dataset.

From the two datasets, all molecules and reactions were recorded with their own identification numbers. Due to the use of different identifiers, to merge both datasets we used RDKit package[37] to pairwise compare molecular canonical SMILES strings for all molecules, and reaction SMARTS for all reactions in both datasets. All KEGG molecules and reactions found

in Reaxys were renamed with Reaxys identifier in the local datasets. An example of these molecules and reactions can be found in SI. The statistics of both datasets are shown in Figure 2a.

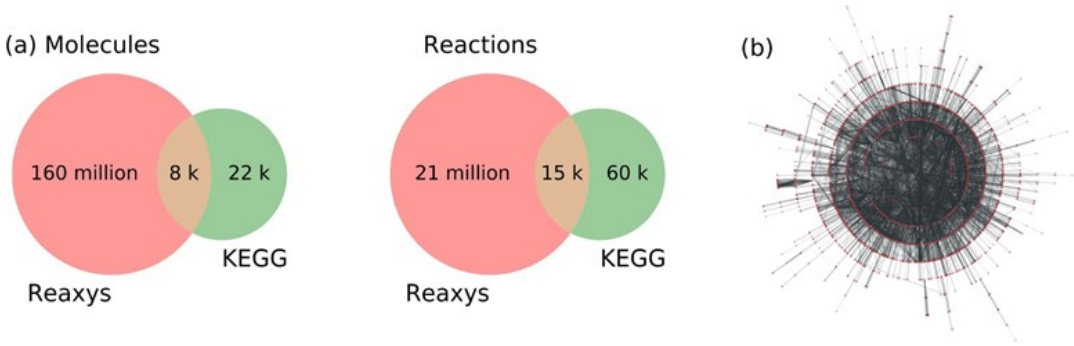


Figure 2. Visualisation of the datasets used in this stidy. (a) A Venn diagram of molecules and reactions in Reaxys and KEGG datasets. The number of KEGG reactions is the number of reactions from both directions. (b) Visualisation of the hybrid reaction network in ARF layout, (red) nodes as molecules and (black) edges as reactions.

Data from KEGG is significantly more sparse compared with the dataset of reactions mined from Reaxys. Among KEGG data, a proportion of molecules and reactions overlap with the Reaxys data, since Reaxys includes mined reaction data regardless whether they are from organic synthetic or bio-synthetic sources. To compare the optimal reaction routes computed from candidate reactions from different sources, three local reactions pools were created: reactions from KEGG were labelled as a biological reactions pool (green + brown in Figure 2a), those from Reaxys and excluding the intersection between Reaxys and KEGG, were labeled as chemical reactions pool (pink), whilst the union of the two sets became the hybrid reactions pool (pink, green + brown). The visualisation of the hybrid reaction network in node and edge representation is shown in Figure 2b.

Since KEGG reactions were manually curated, by no means could the reactions cover the entire synthetic biological space. In reality, the intersection between Reaxys and KEGG may be larger than the brown area. Thus, the chemical reactions pool (pink) possibly still includes metabolic reactions. However, by far boundary between the chemical and biological datasets has been set from the best of knowledge.

We noticed that in Reaxys multiple molecules may sometimes share one identical canonical SMILES string, which gives database noise or high order structure differences (isotopes, stereoisomers, etc.); in the intersection of the two datasets one KEGG molecule may have multiple counterparts in Reaxys. The statistics in the intersection of Venn diagram were counted based on the data from the KEGG-extracted dataset to avoid this issue.

Reaction assessment scores

In reinforcement learning multi-step decision making, each decision making step is associated with a reward/penalty, whilst the whole multi-step decision making process is associated with an expected return/cost, which is the acummulation of rewards/penalties of all decision-making steps. For each of the decision making steps, having multiple options, a well-trained policy model predicts expected returns/costs of the options by foreseeing the cumulative return/cost of the following steps from the current options, and then selects the option with maximised expected return, or the minimised expected cost. In the synthetic problem, each decision making step is to select a reaction option from the reactions pool for the current molecule, and the whole multi-step decision making process is to assemble the multi-step reaction pathway. Here to assess the reaction options, we chose the penalty-expected cost evaluation system and designed reaction assessment scores to represent the penalty to select the candidate reactions. The scores would later be also used to quantify performance of synthesis planning from the three different reactions pools.

Various criteria could be applied to evaluate candidate reactions, subject to the optimisation objectives and data availability.[6] In this work, our objective was to find efficient reaction routes whilst maintaining environmental efficiency. Although many reactions in Reaxys provide attributes such as yield, selectivity and reaction conditions, such information cannot be found from the KEGG dataset or other commonly used biological databases. Therefore, only global criteria determined from the data available in both Reaxys and KEGG databases were used to design the assessment scores. After trials to avoid failure of computation and biases, the global criteria were designed to include atom economy of reaction steps and price of building blocks to consider both route efficiency and operational costs. Of course, changes in global criteria would significantly alter the optimisation results. Here we chose a minimum set of criteria to demonstrate the overall approach.

The design of penalty based on the global criteria was adapted from Schreck et al.[26] For a reaction pathway, penalties were added to the reactions and the building block molecules. For any reaction or building block molecule in the pathway, the penalty was designed to be lower than 1. The penalty of a reaction is shown in Eq. 1.

$$penalty_{reaction} = 1 - AE_i \qquad \text{Eq. 1}$$

The penalty is lower when the atom economy of the desired product in the reaction is greater, where the atom economy (AE_i) counts for the ratio of desired products i over all products in a reaction step. Due to unavailability of reaction stoichiometry in Reaxys, atom economy was determined on the basis of molecular weights (Eq. 2).

$$AE_i = \frac{mol.weight_{product,i}}{\sum mol.weight_{product}} \qquad \text{Eq. 2}$$

The price of building blocks ranges from zero to 10^3 USD/g. To design the penalty score lower than 1, the penalty of a building block is the price of the building block divided by 1000, and cheaper building blocks are favoured by the 'decision maker'.

Apart from building blocks, the reaction pathway may also terminate at a *dead-end molecule*, which means no other reaction link with the molecule, or a *maximum-depth molecule*, which means the end point molecule reaches maximum allowed route depth from the target molecule, which was set to be 10 synthesis steps. The 'decision maker' fails to find a proper pathway in these cases, and therefore, adapted from Schreck et al.,[26] the penalty for a dead-end molecule is 100, and the penalty for a maximum-depth molecule is 10.

The expected cost of a molecule in the reaction pathway is the cumulative penalties of all reactions and end-point molecules from the sub-pathway from the molecule as target molecule to its sub-branches, (Eq. 3). The expected cost of a molecule is also equal to the penalty of the reaction linked with the molecule as a product, plus expected costs of all reactants in the reaction. For example, in Figure 4, the expected cost of m_2 is the sum of penalty of a building block m_6, and a max-length molecule m_{10}, plus the penalty of reactions r_1 and r_3.

$$Exp.Cost_{mol,resi.depth} = \sum_r penalty_{rxn} + \sum_m penalty_{end-point\ mol} \qquad \text{Eq. 3}$$

Workflow of reaction routes optimisation

A reinforcement learning approach adapted from Schreck et al.[26] was applied to create policy 'decision makers' towards identification of the optimal reaction routes. Optimisation workflow shown in Figure 3 was conducted in 20 iterations to optimise the decision-making process and provide promising policy models. The policy model was later used to suggest near-optimal reaction routes for target molecules from the reactions pool. To compare the impacts of

chemical, biological and hybrid reactions pools, the same workflow was run three times to train and produce three policy models from the three reactions pools respectively.

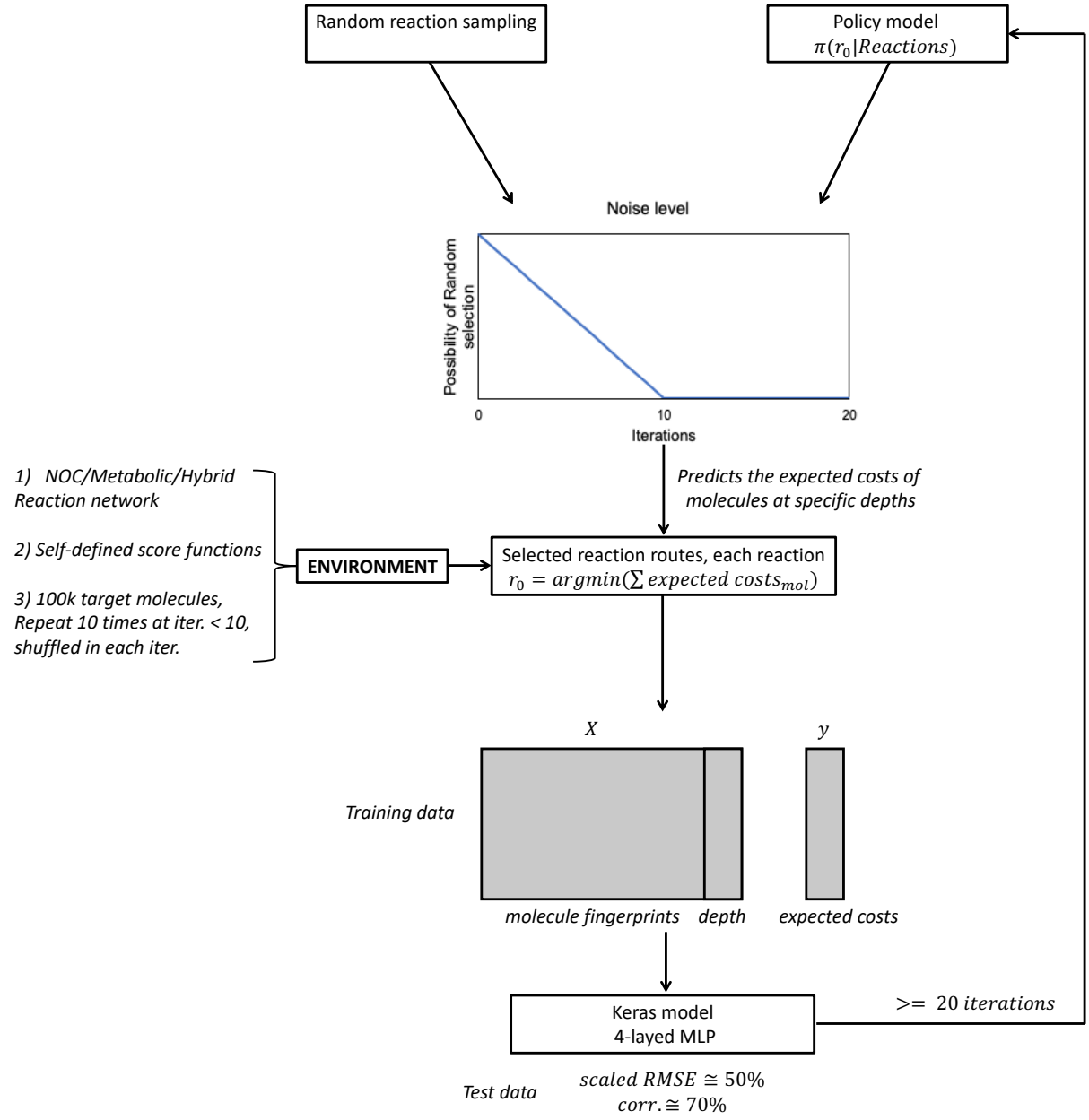


Figure 3. Workflow to collect reaction data and train reinforcement learning policy model for reaction routes optimisation.

Reinforcement learning decision making always starts from defining the decision space, and defining how the 'decision maker' interacts with the decision space (defining 'environment' within RL jargon). The workflow starts with defining a synthesis planning environment, which

includes: (*i*) a reactions pool compromising all molecules and reactions for the 'decision maker' to choose from, (*ii*) evaluation score functions to assess reactions and synthetic routes, and (*iii*) a set of molecules as target molecules to initialise retrosynthesis planning.

For the target molecule set, molecules were filtered to be in the SMILES string length of 20 to 100. This was to maintain the target molecules from different datasets with fair synthetic difficulty. The aim was to include only 100,000 molecules to maintain reasonable computational costs. This was the case for the chemical and hybrid reactions pools. 100,000 molecules (excluding the molecular building blocks) were randomly selected from the molecule set as targets. Also, in each iteration of the optimisation, the target molecule set was reshuffled to increase randomness. However, since the biological dataset records only approximately 30,000 molecules, all molecules with SMILES string length of 20 to 100 (building blocks exclusive), i.e. 12,281 molecules were included as the set of target molecules to compute synthesis planning.

For each target molecule, to compute its retrosynthesis route, all reactions in the reactions pool using the target molecule as one of the reaction products were marked as possible reaction options. If no reaction was found from the dataset, the molecule was marked as a dead-end molecule, as no synthesis step could be further added to the molecule. A dead-end molecule in the pathway is highly disfavoured by the 'decision maker'.

A 'decision maker' selected one of the reaction options as the next synthesis step for the target molecule. For each reactant in selected reaction, as shown in Figure 4, if the reactant was a building block or a dead-end molecule, no further synthesis step is required. If not, the reactant became the next step target molecule. The same procedure was repeated to add the next reaction

to the retrosynthesis route until all end-point molecules at all branches (resulted from multiple reactants reactions in the route) were building blocks, dead-end molecules, or maximum-depth molecules, where the maximum allowed depth was set to be 10 synthesis steps from the target molecule, which is also highly disfavoured.

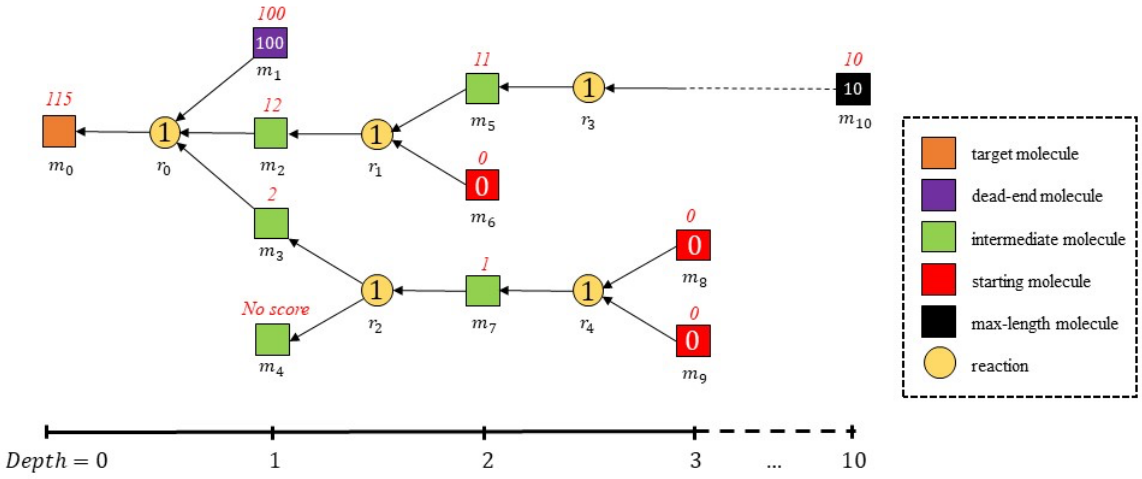


Figure 4. Schematics of retrosynthesis planning of a target molecule m_0, with labelled penalties and expected costs in the planned route. In the schematics, molecules are marked as m_i and reactions as r_i. The axis below represents the depth from the target molecule m_0, where the maximum allowed depth is 10. All r_i are selected by a 'decision maker' (a random selector or policy model) from the available reactions in the reaction network. The penalties of reactions and end-point molecules are shown inside the nodes, and the expected costs of molecules are shown in red on top of the nodes. In the schematics, all reactions have penalties of 1 and all building blocks have no penalty only for simplification. However, in most cases, reactions and building blocks always have penalties ranging from 0 to 1.

The 'decision maker' came from either random sampling of the candidate reactions, or a trained policy model (discussed below), and the possibility of random sampling follows the noise level distribution in Figure 3 and Eq. 4. The possibility of the trained 'decision maker' is one minus that of random sampling.

$$P_{random\ sampling} = \begin{cases} 1 - No._{iter.}, & 0 \leq No._{iter.} < 10 \\ 0, & 10 \leq No._{iter.} \leq 19 \end{cases} \quad \text{Eq. 4}$$

Random sampling brought noises to the decision-making problem, which randomly selected over all reaction options, and made it possible to explore over the decision space. Using the trained policy model, the expected costs of all reactants based on their molecular fingerprints and residual depths (discussed below) were predicted for each reaction option. The policy model 'decision maker' selects the candidate reaction r_o, which has the minimum sum of predicted expected costs of all reactant molecules (Eq. 5). Essentially, this means the policy model would understand the expected costs of molecular structures after exploring the reaction space, and always point to the molecule structures that are easier to synthesise as the following steps.

$$r_o = argmin(\sum_m \widetilde{Exp.Cost_{mol,resi.depth}}) \quad \text{Eq. 5}$$

At iteration 0, a policy model was not generated since no training data had been collected yet. Thus, the 'decision maker' started with random sampling to learn from trial-and-error. From iteration 1, the possibility to use the updated policy model 'decision maker' gradually increased, and from iteration 10, the reaction pathway was optimised only from the policy model.

Suggested by Schreck et al.,[26] not only does the expected cost of a molecule depend on the molecule itself, but also the depth of the molecule in the pathway. If the molecule requires a long synthesis pathway, whether the pathway of a molecule reaches maximum-depth molecules or building-block molecules relies on its residual depth, i.e. the maximum allowed depth (10 steps) subtracted by the current depth from the target molecule. To learn from the simulated experience, the residual depth and the corresponding expected costs of all molecules in the pathway were collected following the designed penalty rules. This did not include the side-

product molecules in the reactions, since the side products have no expected costs. However, the effects of side products were included in the reaction penalty, which counts for the atom economy of the reaction.

Same procedure was repeated for all target molecules to collect residual depths and expected costs of all simulated molecules. For the former 10 iterations, the simulation of each target molecule was repeated 10 times to add randomness to the built pathways. However, for the latter 10 iterations, since all pathways were built by the trained policy models, the repeated simulation results were identical, and thus, only one simulation was required for each target molecule. In each iteration, approximately a million expected cost values of molecules at their corresponding residual depths were collected. The multiple expected costs of same molecule at same residual depth were averaged to count into the training data.

As shown in Figure 3, at each iteration, the trained MLP (discussed below) was eventually updated as the policy model 'decision maker' for the next iteration. The optimisation was terminated after 20 iterations, and the policy model at the last iteration became the final 'decision maker' to predict expected costs of molecules, and select reactions based on Eq. 5 to build retrosynthesis pathways.

Machine learning to learn from molecules

Following Schreck et al.,[26] machine learning models mimic the mathematical relationship between the molecules and their residual depths as inputs and the corresponding expected costs as output. To digitise molecules into mathematical models, extended-connectivity fingerprint (ECFP)[38] was applied, which is a topological fingerprint to convert the circular structure of neighborhood of each non-hydrogen atom into bytes. In this work, the radius of the fingerprint

was four (ECFP4), which detects the multiple layers of the neighborhoods from the molecule centre, and all molecules were converted into 2048 fixed-length bit string. Overall, the input has 2049 features, from which 2048 are binary variables from ECFP, and one from the residual depth.

Multi-layered perception (MLP) neural network was used as the machine learning model to learn from the data, and this was conducted by using the deep learning API Keras.[39] Although over one million datapoints were obtained from each iteration, the structure of the MLP was simple to avoid data overfitting, especially from the 2048 binary variables. The MLP includes an input layer of 2049 nodes, followed by a batch normalization layer to standardize the inputs. Three hidden layers of 30, 15 and 5 nodes respectively using the 'exponential linear unit' activation function were added, and right after each hidden layer, three dropout layers, with dropout rate of 0.3, 0.2 and 0.1 were added to randomly reduce the size of hidden nodes to avoid overfitting. This was eventually followed by an output layer of one node, also with 'elu' activation function, which approximates the molecular expected cost. For specification, MLP used learning rate of 0.002 to slowly learn from data, 'mean square error' as the loss function, and 'adam' as the optimiser. At each iteration, the collected data was split into training data and test data at the ratio of 4:1, and digitalised into 2049 inputs and one output to fit the specified MLP model. With a slow learning rate, we set 50 epoch for the MLP to learn from the training data.

Results and Discussion

Reaction routes optimisation

The logic of the policy model 'decision maker' is as follows: it determines costs of molecules based on their functional structures and previous synthetic performance and minimises the costs

of synthetic planning by selecting the overall low-cost molecules. The expected costs of the molecules were learned through machine learning. Generally at each iteration, after 50 epochs of learning, the test data outputs usually show approximately 45% scaled root mean square error (RMSE) and 65% Pearson correlation coefficient (correlation) from the test data approximations. The RMSE and correlation equations and results for all three environments and all 20 iterations are shown in SI. The RMSEs are high since we tried to learn from 2049 features out of millions of molecules, and by no means could an MLP with three hidden layers fit all the costs of molecular structures using such a simple model structure. Also, we did not expect the MLP to grasp all details from the observations, since a large portion were from trial-and-error noise, which would eventually cause overfitting. However, approximately 70% correlation means that the model learned the overall relationship among molecular structure, retrosynthetic depth and the expected costs, which was promising for overall predictions.

With well-trained policy models, the optimisation results improved over the iterations. The statistics of the expected costs of molecules from the biological reactions pool over the 20 iterations is shown in Figure 5, and the chemical and hybrid reaction pathways show similar optimisation trends (Figure 6). At iteration 0, the median of expected costs for all target molecules reaches approximately 100, which means in most cases, the random sampling 'decision maker' picks dead-end molecules to build reaction routes for the target molecules. For a great portion of the outliers, the 'decision maker' selects multiple dead-end molecules, which approaches the expected costs of multiple hundreds. By learning from trial-and-error results, the policy model reduces the expected costs of most target molecules, with median expected costs being stabilised below 10 in the last five iterations and finalised at 5.2 at the last iteration. Along the 20 iterations, although the portion of outliers also reduces, there are still outliers that reach costs over 200 in the last five iterations. These are large protein molecules

which usually have molar weights over 500 and are believed to be hard to synthesize, which include C16-KDO2-lipid A, UDP-4-amino-4-deoxy-L-arabinose, etc. The situation of target molecule ferricytochrome c has not been improved over the 20 iterations, which stabilises at the expected costs of 704 in the biological reaction pathway in Figure 5.

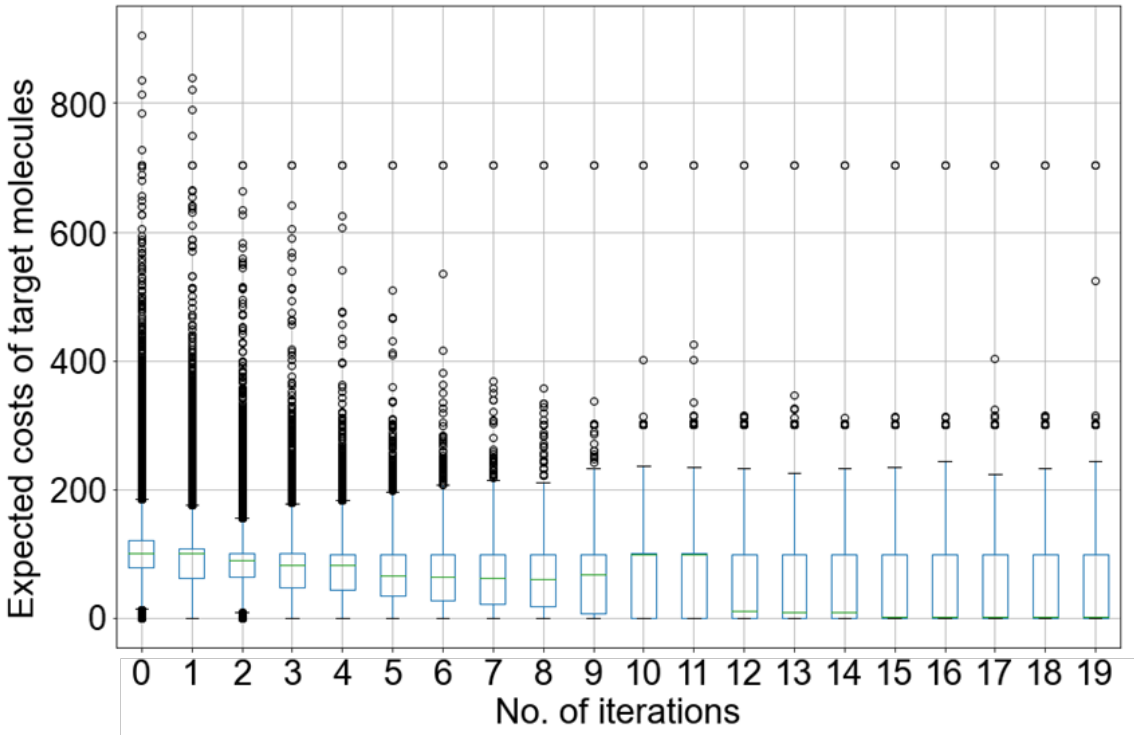


Figure 5. The statistic of expected costs of all target molecules in the biological pathway optimised along with the number of iterations, shown in boxplot.

Comparison of the optimisation results from the three datasets

The expected costs of target molecules rely on the molecule synthetic difficulty and the quality of the decision maker to reach a near-optimal synthetic route. Since we used large target molecules pools and fix the molecules synthetic difficulty by filtering the molecules SMILES string length from 20 to 100, the synthetic difficulty was fair for the chemical, biological and hybrid reactions pools. Hence, we use the median expected costs of target molecules to judge the optimisation results from the three reactions pools. As the policy model 'decision makers' being trained and optimised, whilst the median expected cost of the target molecules from the biological reactions pool has a significant jump at iteration 12, those from the chemical and the

hybrid reactions pools both reduce gradually over the 20 iterations (Figure 6). The three curves all tend to be stable in the last five iterations, which means they all approach optimisation limits. At the last iteration, the medians are 4.3, 5.2, 4.15 respectively for the three reactions pools. This can be interpreted such that in most cases, the molecule synthetic difficulty reduces in the hybrid reactions pool compared with the organic synthesis or synthetic biology ones alone. It also suggests that although the addition of the biological dataset only adds 0.36% data into the chemical dataset (Figure 2a in terms of the number of reactions), in overall, it adds value by 3.4% to the organic synthesis to reach better synthetic results (by comparing the expected costs of molecule medians of 4.3 and 4.15 in organic and hybrid synthesis respectively): it is able to improve the redox efficiency and find more opportunities for synthetic shortcuts among molecules via hybridising the reactions pools.

However, we acknowledge that this interpretation is specific to the used assessment criteria and penalty scores. Other advantages of biological reactions such as greenness and close-to-ambient reaction conditions have not been covered by the current methodology. We also did not implement any quantification of the drawbacks of biological reactions. For example, it is common for biological reactions to be highly dependent on the rest of cellular metabolic network, which increases operational costs of reactions. We also did not consider product separation for any of the reactions in the current implementation.

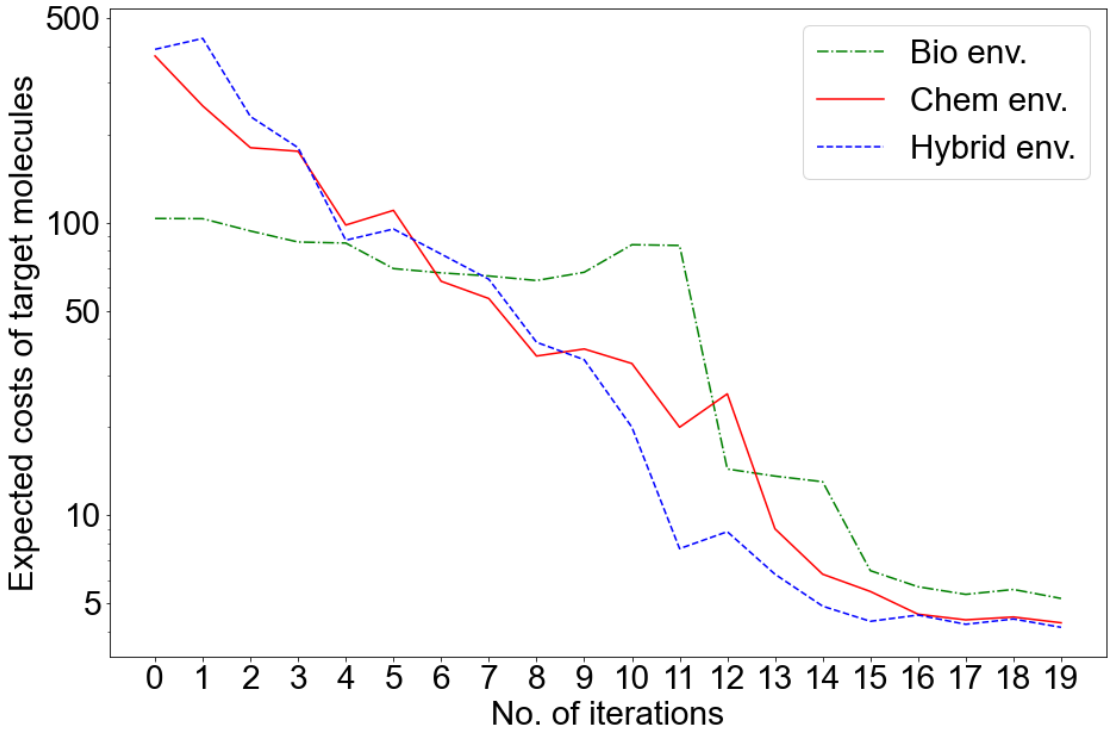


Figure 6. The optimisation results of median expected costs of target molecules from the chemical, biological and hybrid environments, shown in log-scale vertical axis to distinguish the tails at last five iterations.

Optimisation results for drug molecules

KEGG drugs database[40] gives a list of drug molecules as active pharmaceutical ingredients. To test the performance of the final policy model 'decision makers', molecules were crawled from the database to compute reaction pathways to them. Since the reaction pathways were compared by reactions pools from organic chemical, biological and hybrid databases, only 3,821 drug molecules co-existing in the chemical and biological datasets were used as the target drug molecules to compare the optimal reaction pathways from the three datasets. Here we also filtered these molecules into a set of 3,746, to contain only molecules with SMILES string length greater than 10 to increase the synthetic planning complexity, see SI.

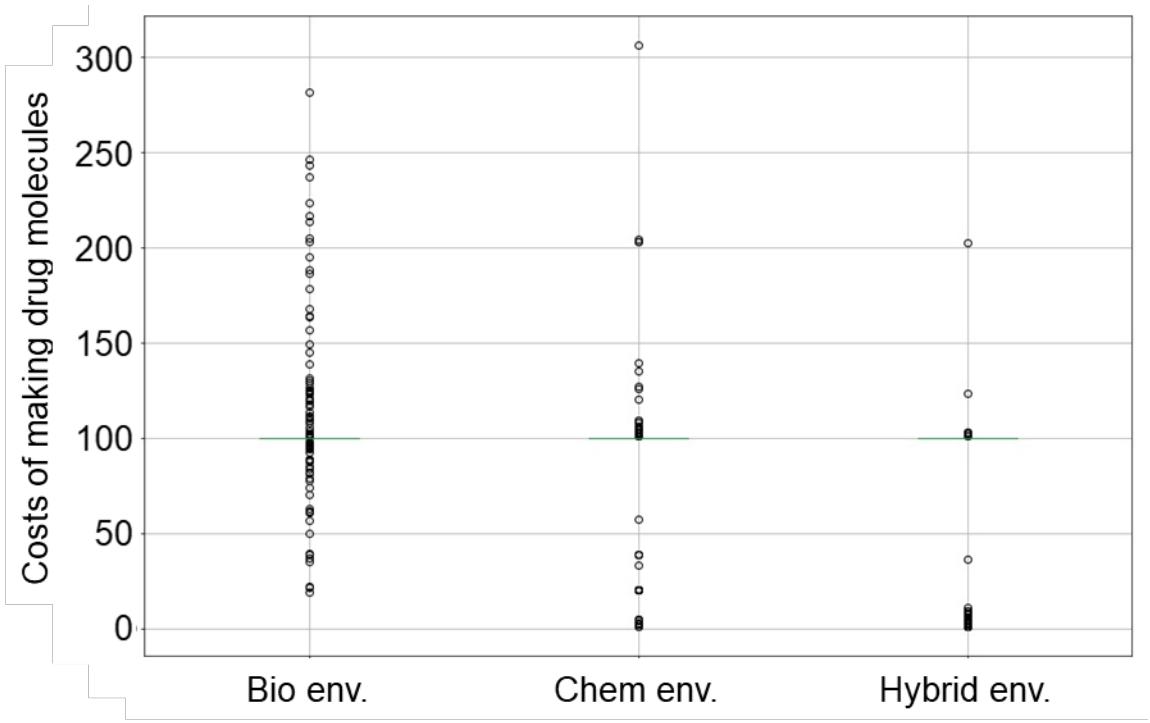


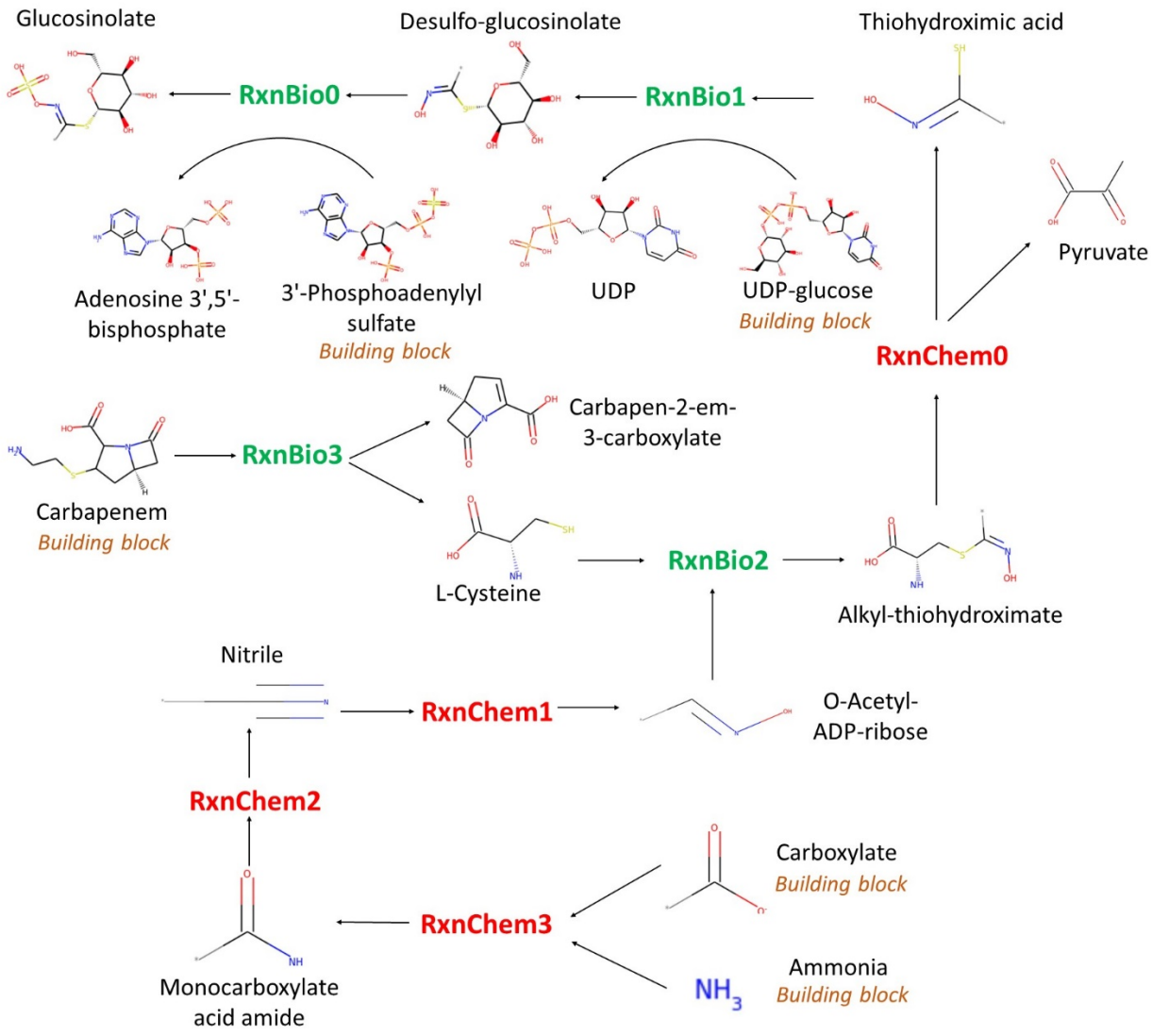
Figure 7. Statistics of costs to make drug molecules determined from the final policy model 'decision maker' using the biological, chemical and hybrid reactions pools.

The final policy model 'decision makers' from the three reactions pools were used to identify the synthetic pathways for each of the drug molecules. The results from the three environments are shown in Figure 7. Different from other random selected target molecules with shorter SMILES string lengths, the drug molecules are more difficult to find synthetic routes. Whilst the target molecules are usually being synthesised within five steps, the cost to make drug molecules reach median of 100 for the three datasets, which means the routes always point to a dead-end molecule. This indicates that due to the molecular complexity, the majority of drug molecules cannot be synthesised with the demonstrated method and datasets. For further work, we could use partial reactions to predict functional transformations. In this way, more possible solutions could be given to the synthetic routes. However, it is also seen that the hybrid environment exhibits a heavier tail towards lower costs to make the molecules. This means the

method opens possibilities to synthesise a significant proportion of drug molecules making use of the full set of chemistry combining organic synthesis and synthetic biology.

An example of these successfully synthesised molecules is glucosinolate, an active pharmaceutical ingredient of multiple Chinese medicines, which are antibacterial, antioxidant, anticarcinogenic, etc.[41] We illustrate the following 7-step synthetic route of glucosinolate in Scheme 1, suggested by the policy model. The cost of making glucosinolate in this route is 3.66, with five building block used. The depth of the longest branch is 7 steps. The route uses four organic chemical reactions and four synthetic biological reactions. Excluding free metabolites and cofactors such as NADPH, NADP+, oxygen, etc., the route produces in total three side products - UDP, pyruvate and carbapen-2-em-3-carboxylate; adenosine-3', 5'-biphosphate from the first biological reaction (RxnBio0) is a coenzyme circulating over cell organisms, and thus, it is not strictly a side product. To compare, there is no chemical route to synthesise this molecule, and the cost of the purely biological route is 7.15.

In this proposed route all reactions are existing historical literature examples. However, since the method did not consider yield, selectivity, greenness and other reaction/process parameters due to lack of data, the route needs to be further investigated using more conventional approaches.



Scheme 1. The proposed synthetic route of glucosinolate by the final hybrid policy model 'decision maker', with glucosinolate shown as target, and organic chemical, synthetic biological reactions labelled as 'RxnChem#' and 'RxnBio#' respectively. Free metabolites and cofactors including NADPH, NADP+, oxygen and water are excluded from the scheme.

Conclusions

We presented an efficient method to suggest near-optimal biochemical pathway *via* data mining from organic chemistry and synthetic biology datasets, and reinforcement learning decision making. With this method, we also proved that, in overall, hybridising chemical and biological reactions to plan synthetic pathways was better than conventional organic synthesis by 3.4%

with respect of the synthesis of all target molecules in the molecular space, due to advantages of synthetic biology to improve redox efficiency and enable synthetic shortcuts for the reaction routes. This conclusion was justified by using atom economy, numbers of reactions steps, and price of building blocks as key criteria to quantify retrosynthesis performance, and applying a comprehensive list of target molecules for the comparison of the organic chemical, synthetic biological and hybrid reactions pools. We could especially benefit from the well-trained policy model to plan the synthetic routes of a set of drug molecules. The case study of certain drug molecules indicates synthesis would be significantly eased with the help of synthetic biology reactions. This methodology could be further extended to mine more comprehensive reaction data to further understand the true costs of making biological reactions, which would make it possible to plan reaction routes with more confidence.

Acknowledgements

CZ is grateful to Cambridge Trust CSC Scholarship for funding his PhD study. We gratefully acknowledge collaboration with RELX Intellectual Properties SA and their technical support, which enabled us to mine Reaxys. Copyright © 2020 Elsevier Limited except certain content provided by third parties. Reaxys is a trademark of Elsevier Limited. Reaxys data were made accessible to our research project via the Elsevier R&D Collaboration Network.

Data Availability

Reaxys molecule and reaction data are accessible to users via Elsevier. KEGG reaction and molecule data are available via KEGG APIs. All other data are shared via Supporting Information.

References

1. Corey, E. J., Robert Robinson Lecture. Retrosynthetic thinking—essentials and examples. *Chem. Soc. Rev.* **1988,** *17* (0), 111-133.

2. Szymkuć, 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.* **2016,** *55* (20), 5904-5937.
3. Thakkar, A.; Kogej, T.; Reymond, J.-L.; Engkvist, O.; Bjerrum, E. J., Datasets and their influence on the development of computer assisted synthesis planning tools in the pharmaceutical domain. *Chem. Sci.* **2020,** *11* (1), 154-168.
4. Weber, J. M.; Guo, Z.; Zhang, C.; Schweidtmann, A. M.; Lapkin, A. A., Chemical data intelligence for sustainable chemistry. *Chem. Soc. Rev.* **2021,** *50* (21), 12013-12036.
5. Wilson, R. J., *Introduction to graph theory*. John Wiley & Sons, Inc.: 1986.
6. Jacob, P. M.; Yamin, P.; Perez-Storey, C.; Hopgood, M.; Lapkin, A. A., Towards automation of chemical process route selection based on data mining. *Green Chem.* **2017,** *19* (1), 140-152.
7. Lapkin, A. A.; Heer, P. K.; Jacob, P. M.; Hutchby, M.; Cunningham, W.; Bull, S. D.; Davidson, M. G., Automation of route identification and optimisation based on data-mining and chemical intuition. *Faraday Discuss.* **2017,** *202* (0), 483-496.
8. Jacob, P.-M.; Lapkin, A., Statistics of the network of organic chemistry. *React. Chem. Eng.* **2018,** *3* (1), 102-118.
9. Weber, J. M.; Ló, P.; Lapkin, A. A., Identification of strategic molecules for future circular supply chains using large reaction networks. *React. Chem. Eng.* **2019,** *4* (11), 1969-1981.
10. Weber, J. M.; Schweidtmann, A. M.; Nolasco, E.; Lapkin, A. A., Modelling Circular Structures in Reaction Networks: Petri Nets and Reaction Network Flux Analysis. In *Comput. Aided Chem. Eng.*, Pierucci, S.; Manenti, F.; Bozzano, G. L.; Manca, D., Eds. Elsevier: 2020; Vol. 48, pp 1843-1848.
11. Fialkowski, M.; Bishop, K. J. M.; Chubukov, V. A.; Campbell, C. J.; Grzybowski, B. A., Architecture and Evolution of Organic Chemistry. *Angew. Chem.* **2005,** *44* (44), 7263-7269.
12. Gothard, C. M.; Soh, S.; Gothard, N. A.; Kowalczyk, B.; Wei, Y.; Baytekin, B.; Grzybowski, B. A., Rewiring Chemistry: Algorithmic Discovery and Experimental Validation of One-Pot Reactions in the Network of Organic Chemistry. *Angew. Chem.* **2012,** *51* (32), 7922-7927.
13. Grzybowski, B. A.; Szymkuć, S.; Gajewska, E. P.; Molga, K.; Dittwald, P.; Wołos, A.; Klucznik, T., Chematica: A Story of Computer Code That Started to Think like a Chemist. *Chem* **2018,** *4* (3), 390-398.
14. Mikulak-Klucznik, B.; Gołębiowska, P.; Bayly, A. A.; Popik, O.; Klucznik, T.; Szymkuć, S.; Gajewska, E. P.; Dittwald, P.; Staszewska-Krajewska, O.; Beker, W.; Badowski, T.; Scheidt, K. A.; Molga, K.; Mlynarski, J.; Mrksich, M.; Grzybowski, B. A., Computational planning of the synthesis of complex natural products. *Nature* **2020,** *588* (7836), 83-88.
15. Coley, C. W.; Barzilay, R.; Jaakkola, T. S.; Green, W. H.; Jensen, K. F., Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **2017,** *3* (5), 434-443.
16. Coley, C. W.; Green, W. H.; Jensen, K. F., RDChiral: An RDKit Wrapper for Handling Stereochemistry in Retrosynthetic Template Extraction and Application. *J. Chem. Inf. Model.* **2019,** *59* (6), 2529-2537.
17. Schwaller, P.; Laino, T.; Gaudin, T.; Bolgar, P.; Hunter, C. A.; Bekas, C.; Lee, A. A., Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Cent. Sci.* **2019,** *5* (9), 1572-1583.
18. Segler, M. H. S.; Preuss, M.; Waller, M. P., Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **2018,** *555* (7698), 604-610.

19. Schwaller, P.; Petraglia, R.; Zullo, V.; Nair, V. H.; Haeuselmann, R. A.; Pisoni, R.; Bekas, C.; Iuliano, A.; Laino, T., Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chem. Sci.* **2020,** *11* (12), 3316-3325.
20. Woodley, J. M.; Turner, N. J., New Frontiers in Biocatalysis. In *Handbook of Green Chem.*, pp 73-86.
21. Ko, Y.-S.; Kim, J. W.; Lee, J. A.; Han, T.; Kim, G. B.; Park, J. E.; Lee, S. Y., Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. *Chem. Soc. Rev.* **2020,** *49* (14), 4615-4636.
22. Lee, S. Y.; Kim, H. U.; Chae, T. U.; Cho, J. S.; Kim, J. W.; Shin, J. H.; Kim, D. I.; Ko, Y.-S.; Jang, W. D.; Jang, Y.-S., A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal.* **2019,** *2* (1), 18-33.
23. Balderas-Hernández, V. E.; Sabido-Ramos, A.; Silva, P.; Cabrera-Valladares, N.; Hernández-Chávez, G.; Báez-Viveros, J. L.; Martínez, A.; Bolívar, F.; Gosset, G., Metabolic engineering for improving anthranilate synthesis from glucose in Escherichia coli. *Microb. Cell Fact.* **2009,** *8* (1), 19.
24. Thomas, K. C.; Ingledew, W. M., Production of 21% (v/v) ethanol by fermentation of very high gravity (VHG) wheat mashes. *J. Ind. Microbiol.* **1992,** *10* (1), 61-68.
25. Voll, A.; Marquardt, W., Reaction network flux analysis: Optimization-based evaluation of reaction pathways for biorenewables processing. *AIChE J.* **2012,** *58* (6), 1788-1801.
26. Schreck, J. S.; Coley, C. W.; Bishop, K. J. M., Learning Retrosynthetic Planning through Simulated Experience. *ACS Cent. Sci.* **2019,** *5* (6), 970-981.
27. Koch, M.; Duigou, T.; Faulon, J.-L., Reinforcement Learning for Bioretrosynthesis. *ACS Synth. Biol.* **2020,** *9* (1), 157-168.
28. Khan, A.; Lapkin, A., Searching for optimal process routes: A reinforcement learning approach. *Comput. Chem. Eng.* **2020,** *141*, 107027.
29. Hu, J.; Niu, H.; Carrasco, J.; Lennox, B.; Arvin, F., Voronoi-Based Multi-Robot Autonomous Exploration in Unknown Environments via Deep Reinforcement Learning. *IEEE Trans. Veh. Technol.* **2020,** *69* (12), 14413-14423.
30. Coley, C. W., Defining and Exploring Chemical Spaces. *Trends Chem.* **2021,** *3* (2), 133-145.
31. Elsevier Reaxys. https://www.reaxys.com/ (accessed 6 Feb).
32. Kanehisa, M.; Goto, S., KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000,** *28* (1), 27-30.
33. ChemSpace https://chem-space.com (accessed 21 Feb).
34. Blaß, L. K.; Weyler, C.; Heinzle, E., Network design and analysis for multi-enzyme biocatalysis. *BMC Bioinform.* **2017,** *18* (1), 366.
35. Bowie, J. U.; Sherkhanov, S.; Korman, T. P.; Valliere, M. A.; Opgenorth, P. H.; Liu, H., Synthetic Biochemistry: The Bio-inspired Cell-Free Approach to Commodity Chemical Production. *Trends Biotechnol.* **2020,** *38* (7), 766-778.
36. Shi, T.; Han, P.; You, C.; Zhang, Y.-H. P. J., An in vitro synthetic biology platform for emerging industrial biomanufacturing: Bottom-up pathway design. *Synth. Syst. Biotechnol.* **2018,** *3* (3), 186-195.
37. RDKit: Open-source cheminformatics. http://www.rdkit.org (accessed 6 Feb).
38. Rogers, D.; Hahn, M., Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010,** *50* (5), 742-754.
39. Chollet, F. Keras. https://github.com/fchollet/keras (accessed 17 July).
40. KEGG DRUG Database. https://www.genome.jp/kegg/drug/ (accessed 10 Dec).
41. Vig, A. P.; Rampal, G.; Thind, T. S.; Arora, S., Bio-protective effects of glucosinolates – A review. *LWT-Food Sci. Technol.* **2009,** *42* (10), 1561-1572.