

SOFTWARE

novoStoic2.0: An integrated framework for pathway synthesis, thermodynamic evaluation, and enzyme selection

Vikas Upadhyay, Mohit Anand, Costas D. Maranas ^{*}

Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania, United States of America

^{*} costas@psu.edu



OPEN ACCESS

Citation: Upadhyay V, Anand M, Maranas CD (2025) novoStoic2.0: An integrated framework for pathway synthesis, thermodynamic evaluation, and enzyme selection. PLoS Comput Biol 21(8): e1012516. <https://doi.org/10.1371/journal.pcbi.1012516>

Editor: Anu Raghunathan, CSIR-National Chemical Laboratory: CSIR National Chemical Laboratory, INDIA

Received: September 24, 2024

Accepted: July 3, 2025

Published: August 6, 2025

Copyright: © 2025 Upadhyay et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The data used in building the platform is available at ScholarSphere (<https://doi.org/10.26207/fox2-se27>) and the source code for running the Streamlit version locally for the web interface is available on GitHub (<https://github.com/>)

Abstract

Computational pathway design and retro-biosynthetic approaches can facilitate the development of innovative biochemical production routes, biodegradation strategies, and the funneling of multiple precursors into a single bioproduct. However, effective pathway design necessitates a comprehensive understanding of biochemistries, enzyme activities, and thermodynamic feasibility. Herein, we introduce novoStoic2.0, an integrated platform that combines tools for estimating overall stoichiometry, designing de novo synthesis pathways, assessing thermodynamic feasibility, and selecting enzymes. novoStoic2.0 offers a unified web-based interface as a part of the AlphaSynthesis platform (<http://novostoitic.platform.moleculemaker.org/>) tailored for the synthesis of thermodynamically viable pathways as well as the selection of enzymes for re-engineering required for novel reaction steps. We exemplify the utility of the platform to identify novel pathways for hydroxytyrosol synthesis, which are shorter than the known pathways and require reduced cofactor usage. In summary, novoStoic2.0 aims to streamline the process of pathway design contributing to the development of sustainable biotechnological solutions.

Author summary

Designing biosynthetic pathways for novel chemical targets often involves a series of non-trivial tasks: estimating stoichiometries, assessing thermodynamic feasibility, and selecting appropriate enzymes, typically addressed using separate tools, which can lead to inconsistencies and hinder the transition from computational design to experimental implementation. To address this, we developed novoStoic2.0, an integrated platform that unifies these tasks into a single workflow. It supports the construction of biosynthetic routes that are thermodynamically viable, meaning they are energetically favorable and chemically feasible, and helps identify which steps may require enzyme re-engineering. This streamlines

maranagroup/novoStoic2.0). The website for the final version of the software is available at <http://novostoic.platform.moleculemaker.org/>

Funding: This work is supported by the U.S. National Science Foundation funded Molecule Maker Lab Institute (MMLI), award number 2019897 supported by National AI Research Institutes Program of the Directorate for Computer and Information Science and Engineering (CISE), in collaboration with the Division of Chemistry (CHE) and the Division of Chemical, Bioengineering, and Environmental Transport Systems (CBET) awarded to CDM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

the transition from computational design to experimental implementation. We demonstrated the platform's capabilities by designing biosynthetic pathways for hydroxytyrosol, a compound with both industrial and biomedical relevance as an antioxidant. The resulting routes are shorter than known alternatives and reduce cofactor requirements, offering more efficient options for microbial production. By combining pathway construction, thermodynamic analysis, and enzyme selection in a coherent system, novoStoic2.0 simplifies and strengthens early-stage biosynthesis planning. We see this as a step toward faster, more reliable development of sustainable production routes in synthetic biology and metabolic engineering.

Introduction

Advances in synthetic biology offer considerable potential for engineering biochemical pathways in producing a diverse array of molecules, ranging from biofuels and pharmaceuticals to value-added chemicals and environmentally friendly biodegradation strategies [1–6]. Traditional approaches to biosynthesis often rely on assembling cataloged enzymatic activities, limiting the exploration of novel production routes [7]. Recent studies have demonstrated the utility of leveraging enzyme promiscuity, wherein enzymes exhibit activity on substrates beyond their native targets [8], to enable the assembly of novel pathways through alteration of enzymatic substrate or cofactor specificity. However, the search for novel conversions through enzyme modification presents both a significant enzyme engineering challenge but also an opportunity [9–11]. In an example of a successful application the promiscuous hydroxylase enzyme (4-hydroxyphenylacetate 3-monooxygenase) was used to alter substrate specificity from its native substrate 4-hydroxyphenylacetate to tyrosol and tyramine. [12] These alternative pathways not only minimized the cell metabolic burden by lowering protein synthesis costs but also improved the efficiency of hydroxytyrosol production by rearranging the metabolic flux.

Several pathway design tools are available to generate sequential reaction steps to convert a source chemical into a target molecule. Notable examples include novoStoic [7], RetroPath 2.0 [13,14], and BNICE [15], which facilitate the exploration of biochemical pathways. User-friendly web interfaces are provided by tools such as RetroBioCat and novoPathFinder [16,17]. Moreover, recent advancements in machine learning ushered transformer-based models to move beyond the simple molecular-input line-entry system (SMILES) to a Seq2Seq model [18] akin to a large-language model (LLM) [19]. In addition, sampling techniques such as Monte Carlo tree search (MCTS) [18,20] or deep learning-guided AND-OR tree search [20–22] can be used to explore the identification of routes connecting the target molecules with inexpensive precursors.

Although pathway synthesis tools can invoke novel steps to complete biosynthetic pathways, the discovery or redesign of an enzyme to carry out the hypothesized transformation remains a challenge [23–25]. Recently, we developed

EnzRank [23], which relies on utilizing convolutional neural networks (CNNs) to understand underlying residue patterns and combines with the substrate molecule signature to provide a probability score for the compatibility of enzyme-substrate pairs. However, implementing the novel steps requires *de novo* enzyme design or protein re-engineering to alter substrate specificity using directed evolution. One such example is the design of a luciferase enzyme that selectively catalyzes the oxidative chemiluminescence of novel luciferin substrates, diphenylterazine and 2-deoxycoelenterazine [26].

Thermodynamic feasibility assessment of the entire pathway and also of individual steps is an important check, as most databases used to train ML retrosynthesis tools, treat reactions as reversible, resulting in erroneously adding steps in a thermodynamically unfavorable direction. Tools such as eQuilibrator [27] and dGPredictor [28] can estimate the standard Gibbs energy change of reactions. While eQuilibrator uses an expert-defined functional group, dGPredictor uses automated chemical moieties that classify every atom in a molecule based on their surrounding atoms and bonds to feature molecules. Using structure-agnostic chemical moieties in dGPredictor allows for the estimation of the standard Gibbs energy of reactions containing novel metabolites absent from databases or molecular structures that cannot be decomposed using expert-defined functional groups.

Herein, we introduce the novoStoic2.0 framework, which integrates the aforementioned tasks (Fig 1) within a single interface. First, the optStoic [29] tool can be used to estimate the optimal overall stoichiometry of the desired conversion by maximizing the yield of the target molecule from the given starting compound. Next, novoStoic [7] attempts to identify the link between the input and output molecules of the overall conversion using both the database and novel reactions. To assess the thermodynamic feasibility of these individual steps, including novel ones, dGPredictor [28] is accessed to estimate the standard Gibbs energy changes. Finally, EnzRank [9] can be used to select enzyme candidates for (any) novel conversions identified within the pathway design.

novoStoic2.0 is a one-stop web-based interface for designing biosynthesis pathways that are not only thermodynamically feasible and carbon/energy balanced but also provide suggestions on enzyme selection for re-engineering associated with novel reaction steps.

Design and implementation

This section describes the design, implementation, and use of the novoStoic2.0 user interface. We used a Streamlit-based Python framework to build the user interface for all the tools that are integrated within novoStoic2.0. We used the MetaNetX database to extract a total of 74,612 reactions and 1,292,153 molecules. Upon processing the database by removing unbalanced reactions, transport reactions, and reactions containing generic molecules, we have 23,585 reactions. Similarly, for molecules, we removed the generic molecules, molecules with multiple MetaNetX IDs, and used molecules that were present in 23,585 reactions, which ended up using 17,154 molecules from the database. Hence, in novoStoic2.0, we use 23,585 reactions and 17,154 molecules for pathway design. Using that, we generated the molecular signature for the molecules and reaction rules for the reactions that are used in novoStoic to design *de novo* pathways for biosynthesis. Finally, we generated 9,686 unique reaction rules from 23,585 reactions. However, both dGPredictor and EnzRank use the KEGG database for standard Gibbs energy estimations and enzyme selection, respectively. We created a mapping of MetaNetX IDs with KEGG IDs, in order to use the novoStoic outputs in both tools. If a MetaNetX ID is not present in KEGG, we use InChI and SMILES string and consider them as novel molecules in dGPredictor and EnzRank. Using dGPredictor to estimate the standard Gibbs energy change for novoStoic identified reactions, we generate thermodynamically feasible pathways. Allowing the integration with EnzRank, we probe all the enzyme sequences for a given reaction rule from the KEGG and Rhea databases using the KEGG REST API [30] and Rhea API [31] along with the novel substrate SMILES string to estimate the probability score for enzyme-substrate activity and rank-order the known enzymes for the novel reaction steps.

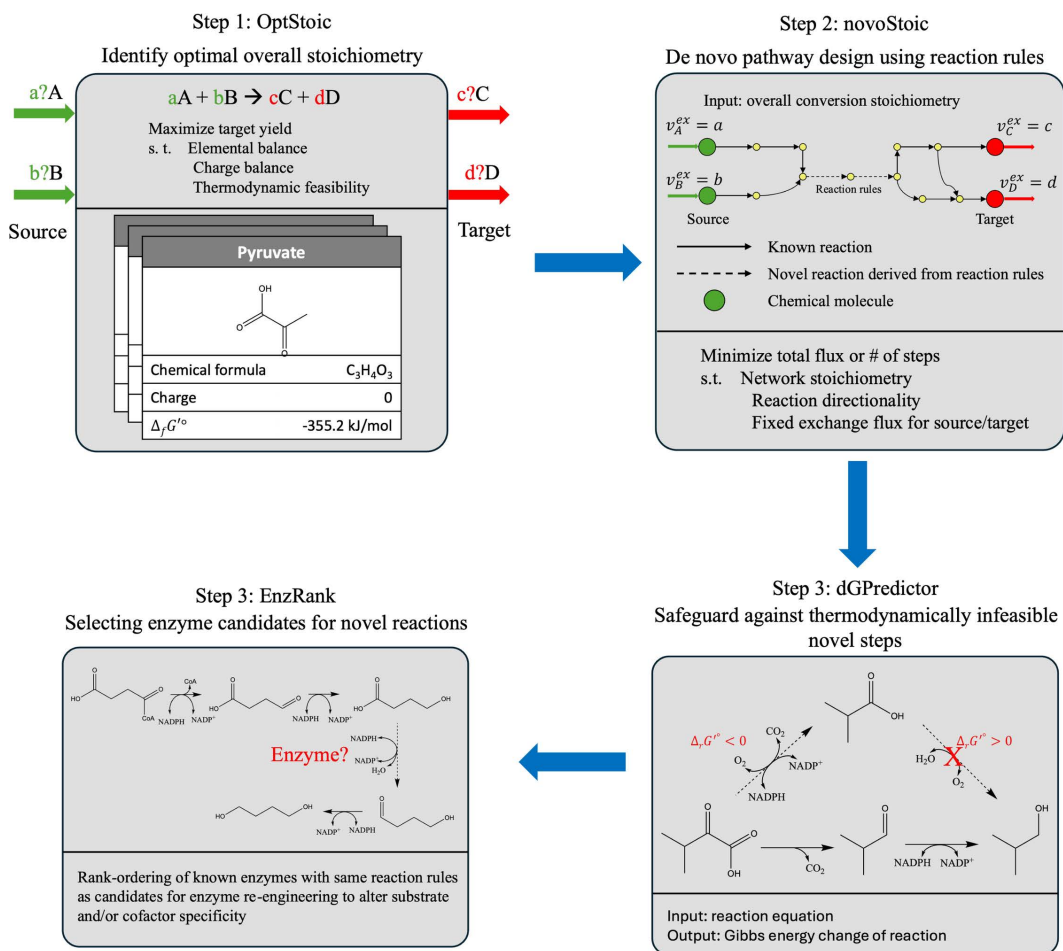


Fig 1. novoStoic2.0: overall workflow for the integrated pathway design, including tools for optimal overall stoichiometry of conversion, optStoic. De novo pathway design using optimal stoichiometry, novoStoic. Thermodynamic assessment of reaction steps using dGPredictor and rank-ordering of known enzymes for novel substrate activity to identify starting enzymes for novel steps using EnzRank.

<https://doi.org/10.1371/journal.pcbi.1012516.g001>

Results

novoStoic2.0 user interface

[Fig 2](#) illustrates the homepage of the web-based platform built using the Streamlit Python package.

It provides brief information for each tool along with relevant references. optStoic [29] uses MetaNetX [30,32] as well as KEGG database compound IDs to input both the source and target molecules. Users may also specify any other co-substrate and co-product to be considered in conjunction with the source and target molecules. Subsequently, optStoic (interface shown in [Fig 3](#)) solves an LP optimization problem, generally aiming to maximize theoretical yield while ensuring mass, energy, charge, and atom balance [29]. The resultant output of this step is the overall target stoichiometry of the conversion.

This overall stoichiometry becomes a key input for novoStoic, depicted in [Fig 4](#). novoStoic requires additional inputs, including the maximum number of (novel) steps, the maximum number of pathway designs, and the primary source and target molecules.

The output of novoStoic visualizes the identified pathways, as well as individual reactions with their standard Gibbs energy estimated using dGPredictor [28]. For pathways involving novel steps, the output provides the option of using

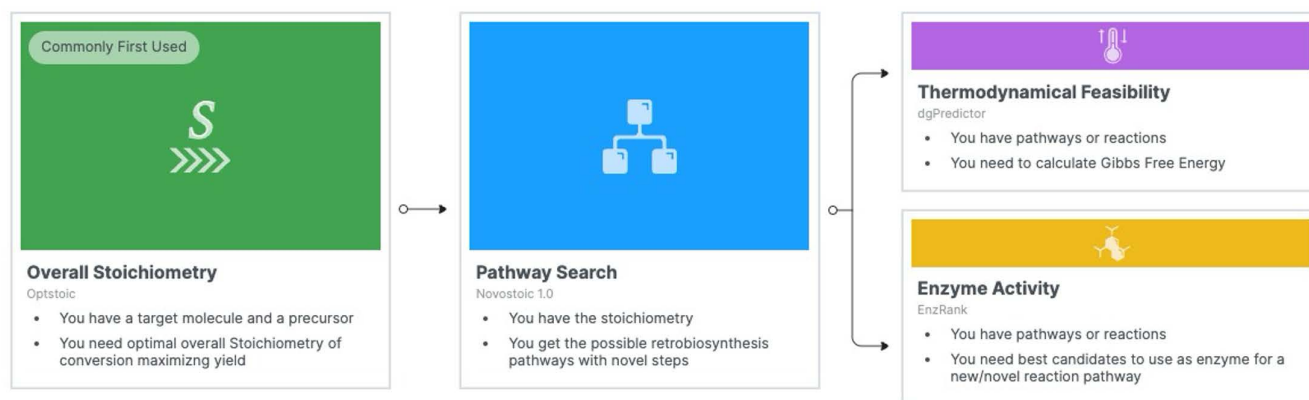
novoStoic2.0

Choose task

Which tool should I use?

I have → I need

Common Workflow of Novostoic



Integrated workflow for pathway design

Fig 2. Annotated workflow of the home page of the novoStoic2.0 user interface. This page provides an option for users to select specific tasks, standalone as well as in the given workflow, to include all the tools for designing pathways.

<https://doi.org/10.1371/journal.pcbi.1012516.g002>

optStoic: Overall Stoichiometry

Input

Primary precursor MNXM76

Target Molecule MNXM56953

Results

Overall Stoichiometry	Yield T1	ΔG° T1	Action
• + 1.0 NADH + 2.0 H ⁺ + 5.0 H ₂ O ₂ [X] 1.0 NAD ⁺ + 4.0 O ₂ + 1.0 CO ₂ + 5.0 H ₂ + 1.0 NH ₄ ⁺ + •	1.0 mol	-9.0 kJ/mol	Pathway Design →
• + 2.0 Water + 1.0 NAD ⁺ + 5.0 O ₂ + 5.0 H ₂ [X] 1.0 NADH + 1.0 CO ₂ + 1.0 NH ₄ ⁺ + 5.0 H ₂ O → •	1.0 mol	-236.0 kJ/mol	Pathway Design →
• + 4.0 NADH + 1.0 O ₂ + 5.0 H ⁺ [X] 4.0 NAD ⁺ + 1.0 CO ₂ + 3.0 H ₂ + 1.0 NH ₄ ⁺ + •	1.0 mol	-551.0 kJ/mol	Pathway Design →

Fig 3. User interface of optStoic, requires the molecule information (MetaNetX/KEGG ID or SMILES string) of the starting and target molecules and outputs the optimal overall conversion stoichiometry, i.e., used in novoStoic to find synthesis pathways.

<https://doi.org/10.1371/journal.pcbi.1012516.g003>

novoStoic: Pathway Search

Input reactant stoichiometry

Input product stoichiometry

Input parameters

Overall Stoichiometry

Use an Example Clear All

A Primary Precursor

1 MNXM76 Draw / Upload *L-tyrosine*

+

Co-Reactants

2 MNXM738702 *NADPH*

2 MNXM735438 *O₂*

+

Co-Products

2 MNXM5 *NADP(+)*

1 MNXM729302 *NH₄(+)*

1 MNXM13 *CO₂*

1 MNXM732620 *H₂O₂*

+

A Target Molecule

0 MNXM56953 Draw / Upload *Hydroxytyrosol*

1 Primary Precursor 2 Co-Reactants 4 Co-Products 1 Target Molecule Overall Stoichiometry

Pathway Search setting

Tip: More you choose, the longer it takes.

Number of reaction steps per pathway design 4

Number of pathway generated 3

Gibbs Energy Predictor

Tip: You can always adjust your selections later on the results page in the filtering section.

☐ Check the box to find pathways with thermodynamically feasible reactions

Enzyme selection for novel reactions

Rank-ordering known enzyme candidates for the novel reactions to select top enzyme candidates

Number of candidates you want to see for each novel steps 3

Results

Selected Stoichiometry

2.0 NADH + 2.0 O₂ + 3.0 H(+) -> 2.0 NAD(+) + 1.0 CO₂ + 1.0 NH₄(+) + 1.0 H₂O₂

Filter 0

PATHWAY 1

STEP 1 R1: L-tyrosine +326.04 ± 2.3 kJ/mol

STEP 2 R2: Intermediate +118.53 ± 1.1 kJ/mol

STEP 3 R3: Intermediate -32.82 ± 1 kJ/mol

Common Name: L-tyrosine

Common Name: hydroxytyrosol

Fig 4. User interface of novoStoic, requires overall balanced stoichiometry of conversion from a source to a target molecule, and input parameters such as maximum number of steps for pathway search, number of pathways to search, and the number of top known enzyme candidates ranked by EnzRank for novel steps identified in the pathways.

<https://doi.org/10.1371/journal.pcbi.1012516.g004>

EnzRank to generate the top known enzyme candidates for potential activity with a novel substrate (i.e., the default is the top five enzyme candidates for each novel step).

dGPredictor could also be used as a standalone standard Gibbs energy estimator as illustrated in Fig 5 using either the KEGG [30] reaction string or the InChI string for novel molecules that are not present in the KEGG [30] database. This feature holds broad applicability, including conducting global thermodynamic feasibility analysis of genome-scale metabolic models [15].

Finally, as shown in Fig 6, we also provide the user interface of EnzRank as a standalone tool within novoStoic2.0. Here, the users can input the enzyme's amino acid sequence and the KEGG ID or SMILES string (if there is a novel structure) of the substrate to generate the probability score for the enzyme-substrate activity. This score can rank-order known enzymes for novel substrate activity. While the tool automatically uses enzyme sequences from KEGG and RHEA

dGPredictor: Estimating $\Delta_r G'^o$ of the biochemical reaction

Input Reactions

Enter known or/and novel reactions to calculate Gibbs free energy and check thermodynamic feasibility.

The screenshot displays the dGPredictor web interface. At the top, there are tabs for 'Reactions' and 'Kegg ID'. Below this, the 'Reaction 1' section contains a toggle switch for 'Have new molecule(s) not listed in KEGG database and MetaNetX database', which is currently turned on. Below the toggle is a text input field for 'Enter the new molecule's SMILES or InChI' containing the string 'InChI=1S/C14H12O/c15-14-8-4-7-13(11-14)10-9-12-5-2-1-3-6-12/h1-11,15H/b10-9+'. To the right of this field is a button labeled 'N00001'. Below this is another text input field for 'Enter a reaction with database IDs of known compounds.' containing the string 'C01745 + C00004 ⇌ N00001 + C00003 + C00001'. At the bottom of the 'Reaction 1' section is a button labeled '+ Add more reactions'. Below the 'Reaction 1' section is the 'Settings' section, which contains two sliders: 'pH' set to 7 and 'Ionic Strength [M]' set to 0.1.

Fig 5. User interface of dGPredictor, requires the KEGG reaction string as input. An InChI string is required if a molecule is not present in the database. The user can modify settings such as pH and ionic strength to get $\Delta_r G'^o$ at different physiological conditions.

<https://doi.org/10.1371/journal.pcbi.1012516.g005>

databases [30,31], users also have the option to manually input custom enzyme sequences if they are working with enzymes not included in these databases, offering greater flexibility in enzyme selection. EnzRank is also integrated during pathway search to extract and rank-order appropriate enzyme candidates for any novel steps found.

Example application: Hydroxytyrosol synthesis

Hydroxytyrosol (HXT) is a powerful antioxidant that offers protection from free radicals. The initial effort of biosynthesis uses tyrosine as a starting substrate, using tyrosol as the intermediate, and was reported to achieve a yield close to 50% [12]. A recent effort by Chen et. al [12] uses a promiscuous hydroxylase enzyme to identify two novel pathways for HXT biosynthesis. Motivated by these results herein, we applied novoStoic2.0 to explore additional pathway designs. We show two of the identified new pathways designed using L-tyrosine as a starting precursor and HXT as the target molecule.

The first step involves the identification of the optimal stoichiometry (i.e., max carbon yield) for the overall conversion using optStoic. We explored multiple alternatives in one run by allowing as reactants/products an inclusive set of small molecules such as, CO_2 , H_2O , NH_4^+ , O_2 , H_2O_2 , and cofactors such as NAD(P)H , NAD(P)^+ to balance the overall conversion.

optStoic identified the overall stoichiometry associated with the pathway identified by Chen et. al that uses a promiscuous hydroxylase enzyme given by:

 Draw / Upload

Enzyme Selection

(S)-malate
O=C(O)C[C@H](O)C(=O)O
C00149
MNXM1107192

A0A4P8WFA8:MTKRVLVTGGAGLGSLCERLSEGVICLDNFGSGRRKNKEFEDHPSPKVNDRVRISLESPVDRIYHLASRAPDFTQFPVNIALANTQTRRLLDGARACARMVFASTSE
VYGGPKVHPQPEYTGNGVNIQRGGEQSKRFGELTVAYRKQVYDARTVRIFTYGPMPRPDGGRVPTFTGALRGDDLTLDGDEGQTRSPCYVDLLIEGLSLMRVDNIVYNGIKENERIT
IKFI AYVEYI GITDFTSDIYVEFI PDDPGGRRDDPDIKTRAKTEI DMWPKISL REGI FDTITYTNG

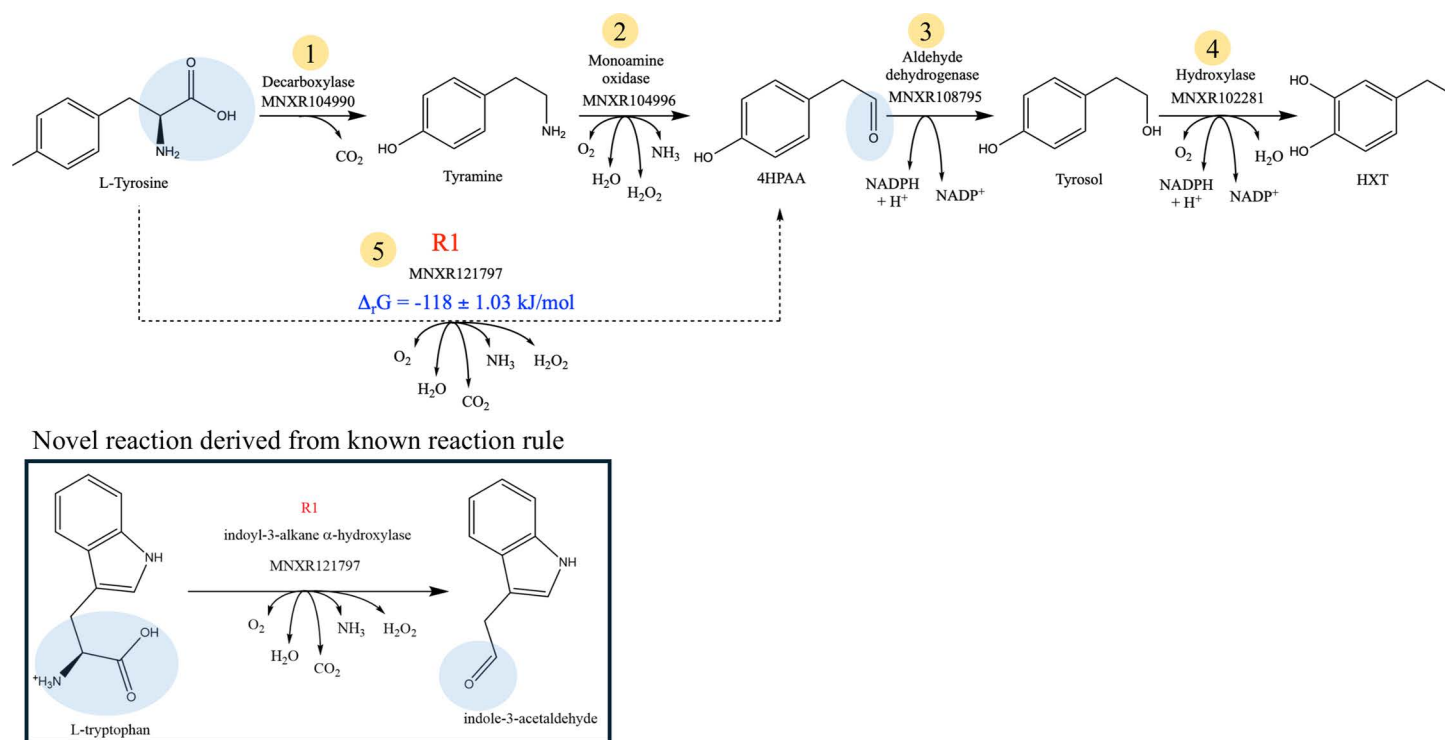


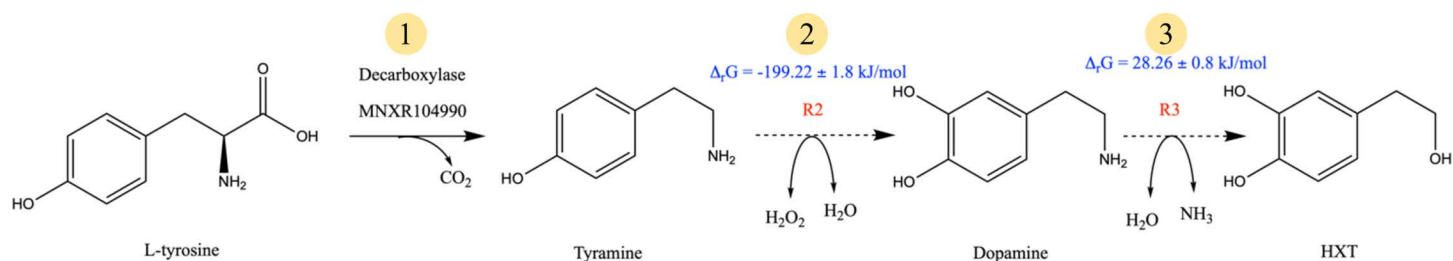
Fig 7. Pathway identified using novoStoic with overall stoichiometry 1, here is the pathway with steps 1-4 showing the known pathway from the literature. Whereas the pathway with steps 5-3-4 shows a novel pathway that bypasses steps 1 and 2 in a single step. Here, step 1 is a decarboxylation reaction, step 2 is tyramine: oxygen oxidoreductase, step 3 is aldehyde dehydrogenase, step 4 is a hydroxylase reaction, and novel step 5 comes from a reaction MNXR121797 (exact chemical transformation highlighted in blue), which is a hydroxylase enzyme.

<https://doi.org/10.1371/journal.pcbi.1012516.g007>

However, an alternative novel pathway was identified that bypasses the first two steps of converting tyrosine to 4HPAA by integrating them into a single step using the reaction rule R1 (i.e., EC 1.13.99.-) derived from reaction indoyl-3-alkane- α -hydroxylase. The pathway in Fig 7 shows the novel bypassing step 5, which follows the reaction rule of reaction MNXR121797. Using dGPredictor, we confirmed that the proposed novel step involves a negative standard Gibbs energy. Reaction MNXR121797 (EC 1.13.99.-) has a single enzyme sequence assigned to it, which can serve as the starting enzyme that would need to be re-engineered for usher activity on the new substrate L-tyrosine. This makes the use of EnzRank redundant. Even though the reaction for step 4 is not cataloged in MetaNetX, it follows the reaction rule of reaction MNXR102281. However, recent experimental evidence [12] points out that enzyme 4-hydroxyphenylacetate 3-monooxygenase has a secondary activity on tyrosol, obviating the need for enzyme discovery/engineering.

Using optimal stoichiometry 2, we identified novel pathways using three steps that do not use any cofactors and only small molecules, co-reactants and co-products (see S1 File). Fig 8 shows one of the pathways identified by novoStoic for the synthesis of HXT.

Fig 8 illustrates the three-step cofactor-free pathway. Step 1 is the same as in the known pathway shown in (Fig 7) [12]. The novel reaction associated with step 2 follows the reaction rule associated with reaction 3-methyl-L-tyrosine hydrogen peroxide oxidoreductase (EC 1.11.2.5) that performs a hydroxyl addition to convert tyramine to dopamine. Notably, there is a known reaction in the MetaNetX database that performs this conversion, but it requires NAD(P)H as a cofactor. Using dGPredictor, we assessed the thermodynamic feasibility of the desired direction and found a high negative standard Gibbs



Novel reactions derived from known reaction rules

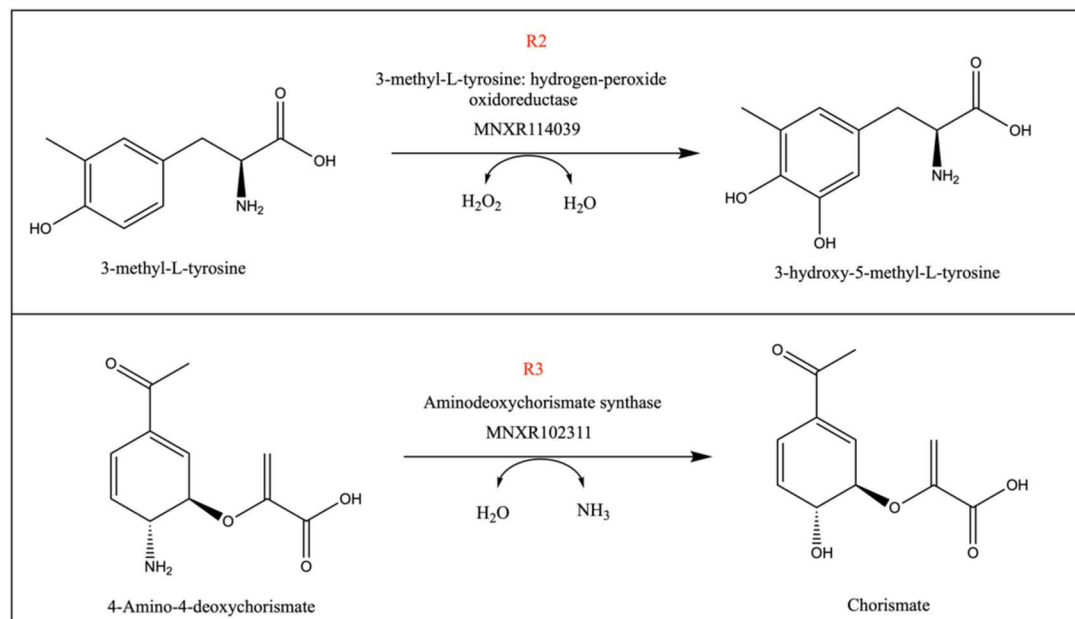


Fig 8. Pathway identified using novoStoic with overall stoichiometry 2, here is the pathway with steps 1-3 showing the shorter novel pathway. Here, step 1 is a known decarboxylation reaction, step 2 is derived from the reaction rule R2, and step 3 is derived from reaction rule R3.

<https://doi.org/10.1371/journal.pcbi.1012516.g008>

energy. Step 3 is a novel reaction that shares the same reaction rule as reaction MNXR102311 (aminodeoxychorismate synthase), which converts 4-amino-4-deoxychorismate to Chorismate. dGPredictor assesses the standard Gibbs energy change to be slightly positive, which can be made feasible by tilting the reactant concentration to move the reaction in the forward direction. Nevertheless, this example showcases that novoStoic can identify a shorter pathway with a lesser number of cofactors compared to the known pathway. The novel step R2 (Fig 8) was already suggested in the recent article by Chen et. al [12] that includes the engineering promiscuous enzyme 4-hydroxyphenylacetate 3-monooxygenase to show activity with tyramine. However, novoStoic qualified it as a novel step as the reaction was not present in the database yet. For the novel step R3 (in Fig 8), we found a total of 3,744 unique enzyme sequences from the reactions with the same rule as R3 (see S2 File, which provides the list of all the enzymes for step R3). Next, EnzRank was used to rank-order known enzymes for the same reaction rule from which the novel step R3 is derived using the probability score to identify the potential of an enzyme to act on a substrate (dopamine). Here, Table 1 shows the top 3 candidates found using EnzRank for reaction step R3, and S3 File contains the EnzRank scores for all the enzyme sequences for the same rule as R3.

Using HXT as an example/case study, we highlighted how novoStoic2.0 can be used to find multiple pathway alternatives for HXT synthesis mapping to different overall stoichiometries. These alternatives mapping to different overall

Table 1. Top 3 genes from different organisms selected by EnzRank for novel reaction step R3 with reaction rule derived from aminodeoxychorismate synthase enzyme.

Organisms	EnzRank score
<i>Streptomyces rochei</i>	0.87
<i>Streptomyces griseofuscus</i>	0.86
<i>Cellulomonas xiejunii</i>	0.80

<https://doi.org/10.1371/journal.pcbi.1012516.t001>

stoichiometries were subsequently subjected to thermodynamic feasibility analysis, and enzymes were prioritized for all novel steps. Additional examples are provided in the [S4 File](#).

In collaboration with the Molecule Maker Lab Institute (MMLI) at UIUC, novoStoic2.0 is being developed as an interactive website for the synthesis planning tool, which is a part of the AlphaSynthesis platform and is publicly available at <http://novostoitc.platform.moleculemaker.org/>.

Availability and future directions

Pathway design entails multiple tasks ranging from decisions on starting points and possible co-reactants/products, sequence of metabolic reaction steps, including novel reaction by-passes, selection of enzymes for uncharacterized reaction steps, checking on thermodynamic feasibility of chosen reaction direction, and many more. novoStoic2.0 integrates many of these steps within a single resource, thus streamlining the task of pathway discovery and evaluation. In contrast to existing pathway design tools, which focus solely on pathway generation and require separate analyses for thermodynamics, enzyme selection for individual steps, and overall reaction balance including cofactors, novoStoic2.0 integrates all these aspects into a unified framework. [Table 2](#) presents a comparative analysis of the features offered by existing biosynthetic pathway design tools and novoStoic2.0. By enabling the rapid computational exploration of design alternatives, it expands the space of alternatives explored and thus the chances of success. The development of an easy-to-use web-based interface consolidates many tasks into a single platform. The web platform is currently limited to two runs due to computational constraints, and jobs will be queued if multiple pathway design tasks are submitted. Note that optimal stoichiometry does not guarantee a viable pathway design, as optStoic solves a linear programming (LP) problem that maximizes theoretical yield while enforcing elemental and charge balance. This can sometimes result in unrealistic stoichiometries that do not generate feasible pathways. Users may also need to allow ATP hydrolysis to drive reactions with a positive Gibbs energy change, making the overall stoichiometry thermodynamically feasible. This adjustment can potentially lead to alternative biosynthetic pathways. The web platform supports pathway searches, generating up to 10 distinct pathways, each with up to 10 reactions, for any overall stoichiometry provided by optStoic or entered manually. For novel reactions identified during the search, users can view the top 10 enzyme candidates, provided the underlying reaction rule yields at least 10 candidates. Furthermore, ΔG (Gibbs free energy change) predictions are available for all novel reactions. As standalone tools, dGPredictor [38] evaluates up to 5 reactions simultaneously, while EnzRank [23] assesses a single substrate-enzyme pair at a time. Both typically complete their analyses in under 2 minutes and can be executed repeatedly. In contrast, pathway searches generally require several hours to complete.

novoStoic2.0 is by no means inclusive of all tasks needed to instantiate a pathway design. The pathway will ultimately have to be ported in a production strain. The strain will have to be engineered so that no carbon flux leaks away from the desired pathway. The expression levels of genes and translation rates of proteins along the pathway will have to be finely tuned to limit metabolic burden, and inhibitory controls will have to be ameliorated. There is already a rich literature of tools aimed at addressing these challenges (e.g., optKnock [39], optForce [40], RBSCalculator [41], etc.). Moving beyond, pathway carbon and energy efficiency, additional design considerations are

Table 2. Comparison of existing biosynthesis planning tools.

Tool	Year	Website	Input	Thermodynamic feasibility check	Enzyme candidates	Balanced	Database
novoStoic2.0	2025	https://novostoic.platform.moleculemaker.org/home	Start & Target	Yes (dGPredictor)	Yes (EnzRank)	Yes	MetaNetX, KEGG, RHEA
RetroBioCat [17]	2021	https://retrobiocat.com	Target	No	Yes (via literature)	No	Expert defined rules
RetroPath2.0 [13]	2018	https://www.myexperiment.org/workflows/4987.html	Target	Yes (group contribution)	Yes (Selenzyme [33])	No	MetaNetX, BRENDA
BioNavi [34]	2022	http://biopathnavi.qmclab.com/	Target	No	Yes (Selenzyme [33])	No	MetaCyc, KEGG, MetaNetX
PathPred [35]	2010	http://www.genome.jp/tools/pathpred/	Start & Target	No	No	No	KEGG
XTMS [36]	2014	http://xtms.issb.genopole.fr/	Target	Yes (group contribution)	Yes	Yes	MetaCyc, KEGG
enviPath [37]	2016	https://envipath.org/	Target	No	No	No	Expert curated reactions

<https://doi.org/10.1371/journal.pcbi.1012516.t002>

equally important. These may include predicting the toxicity of intermediates, safeguarding against protein misfolding/aggregation, the presence of high-affinity product exporters, etc. For all these tasks, several computational tools are available to provide estimates [42–45]. Finally, perhaps the most challenging step is the identification or redesign of enzymes with the desired substrate specificity and activity to carry out novel conversion steps. Whereas in some cases promiscuous enzymes can be found or adapted through directed evolution [12,46], rapid advances in enzyme design using ML tools promise to automate this step [47–49]. We envision that many of these aforementioned tools will be integrated into future versions of novoStoic.

We anticipate that ML techniques will likely revolutionize pathway design in the same way that they have changed the landscape in protein folding and enzyme design. For instance, leveraging Large Language Models (LLMs) on up-to-date literature information could immediately inform pathway designs. For example, for the identified pathway shown in Fig 7, novoStoic classified step 3 converting tyrosol to HXT as a novel step, but a recent article by Chen et. al [12] suggested that the desired novel activity has already been assigned an enzyme that was not present yet in the database. Therefore, LLMs offer the promise of automating data mining from the literature and moving beyond simple SMILES encodings for metabolites and reactions. novoStoic2.0 already leverages some of these ML developments and offers a versatile platform to integrate additional ones in the future.

Supporting information

S1 File. Case study: L-tyrosine to hydroxytyrosol pathway design using novoStoic2.0.
(DOCX)

S2 File. Enzyme candidates for novel reaction identified in hydroxytyrosol synthesis.
(CSV)

S3 File. EnzRank scores for novel reaction identified in hydroxytyrosol synthesis.
(CSV)

S4 File. More examples of pathway design using novoStoic2.0.
(DOCX)

Acknowledgments

I acknowledge the NCSA team at the University of Illinois Urbana-Champaign for their invaluable assistance in developing the novoStoic2.0 website, an integral component of the AlphaSynthesis platform for the Molecule Maker Lab Institute (MMLI). I extend special thanks to Matt Berry, Lijiang Fu, Kate Arneson, Bingji Guo, and Sara Lambert for their significant contributions.

Author contributions

Conceptualization: Vikas Upadhyay, Costas D. Maranas.

Data curation: Vikas Upadhyay, Mohit Anand.

Formal analysis: Vikas Upadhyay, Mohit Anand.

Funding acquisition: Costas D. Maranas.

Investigation: Vikas Upadhyay, Mohit Anand.

Methodology: Vikas Upadhyay, Costas D. Maranas.

Project administration: Costas D. Maranas.

Resources: Costas D. Maranas.

Software: Vikas Upadhyay, Mohit Anand, Costas D. Maranas.

Supervision: Costas D. Maranas.

Validation: Vikas Upadhyay.

Visualization: Vikas Upadhyay, Mohit Anand.

Writing – original draft: Vikas Upadhyay.

Writing – review & editing: Vikas Upadhyay, Mohit Anand, Costas D. Maranas.

References

1. Zhang F, Rodriguez S, Keasling JD. Metabolic engineering of microbial pathways for advanced biofuels production. *Curr Opin Biotechnol*. 2011;22(6):775–83. <https://doi.org/10.1016/j.copbio.2011.04.024> PMID: 21620688
2. Turconi J, Griololet F, Guevel R, Oddon G, Villa R, Geatti A, et al. Semisynthetic Artemisinin, the Chemical Path to Industrial Production. *Org Process Res Dev*. 2014;18(3):417–22. <https://doi.org/10.1021/op4003196>
3. Bai W, Geng W, Wang S, Zhang F. Biosynthesis, regulation, and engineering of microbially produced branched biofuels. *Biotechnol Biofuels*. 2019;12(1). <https://doi.org/10.1186/s13068-019-1424-9>
4. Rodriguez GM, Tashiro Y, Atsumi S. Expanding ester biosynthesis in *Escherichia coli*. *Nat Chem Biol*. 2014;10(4):259–65. <https://doi.org/10.1038/nchembio.1476> PMID: 24609358
5. Huang T, Ma Y. Advances in biosynthesis of higher alcohols in *Escherichia coli*. *World J Microbiol Biotechnol*. 2023;39(5):125. <https://doi.org/10.1007/s11274-023-03580-w> PMID: 36941474
6. Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, et al. Lignin valorization: improving lignin processing in the biorefinery. *Science*. 2014;344(6185):1246843. <https://doi.org/10.1126/science.1246843> PMID: 24833396
7. Kumar A, Wang L, Ng CY, Maranas CD. Pathway design using de novo steps through uncharted biochemical spaces. *Nat Commun*. 2018;9(1):184. <https://doi.org/10.1038/s41467-017-02362-x> PMID: 29330441
8. Wang L, Ng CY, Dash S, Maranas CD. Exploring the combinatorial space of complete pathways to chemicals. *Biochem Soc Trans*. 2018;46(3):513–22. <https://doi.org/10.1042/BST20170272> PMID: 29626146
9. Upadhyay V, Boorla VS, Maranas CD. Rank-ordering of known enzymes as starting points for re-engineering novel substrate activity using a convolutional neural network. *Metab Eng*. 2023;78:171–82. <https://doi.org/10.1016/j.ymben.2023.06.001> PMID: 37301359
10. Carbonell P, Koch M, Duigou T, Faulon J-L. Enzyme Discovery: Enzyme Selection and Pathway Design. *Methods Enzymol*. 2018;608:3–27. <https://doi.org/10.1016/bs.mie.2018.04.005> PMID: 30173766

11. Chowdhury R, Grisewood MJ, Boorla VS, Yan Q, Pfleger BF, Maranas CD. IPRO+/-: Computational Protein Design Tool Allowing for Insertions and Deletions. *Structure*. 2020;28(12):1344–1357.e4. <https://doi.org/10.1016/j.str.2020.08.003>
12. Chen W, Yao J, Meng J, Han W, Tao Y, Chen Y, et al. Promiscuous enzymatic activity-aided multiple-pathway network design for metabolic flux rearrangement in hydroxytyrosol biosynthesis. *Nat Commun*. 2019;10(1):960. <https://doi.org/10.1038/s41467-019-08781-2> PMID: 30814511
13. Delépine B, Duigou T, Carbonell P, Faulon J-L. RetroPath2.0: A retrosynthesis workflow for metabolic engineers. *Metab Eng*. 2018;45:158–70. <https://doi.org/10.1016/j.ymben.2017.12.002> PMID: 29233745
14. Carbonell P, Parutto P, Baudier C, Junot C, Faulon J-L. Retropath: automated pipeline for embedded metabolic circuits. *ACS Synth Biol*. 2014;3(8):565–77. <https://doi.org/10.1021/sb4001273> PMID: 24131345
15. Wang L, Dash S, Ng CY, Maranas CD. A review of computational tools for design and reconstruction of metabolic pathways. *Synth Syst Biotechnol*. 2017;2(4):243–52. <https://doi.org/10.1016/j.synbio.2017.11.002> PMID: 29552648
16. Ding S, Tian Y, Cai P, Zhang D, Cheng X, Sun D, et al. novoPathFinder: a webserver of designing novel-pathway with integrating GEM-model. *Nucleic Acids Res*. 2020;48(W1):W477–87. <https://doi.org/10.1093/nar/gkaa230> PMID: 32313937
17. Finnigan W, Hepworth LJ, Flitsch SL, Turner NJ. RetroBioCat as a computer-aided synthesis planning tool for biocatalytic reactions and cascades. *Nat Catal*. 2021;4(2):98–104. <https://doi.org/10.1038/s41929-020-00556-z> PMID: 33604511]
18. Probst D, Manica M, Nana Teukam YG, Castrogiovanni A, Paratore F, Laino T. Biocatalysed synthesis planning using data-driven learning. *Nat Commun*. 2022;13(1). <https://doi.org/10.1038/s41467-022-28536-w>
19. Liu CH, Korablyov M, Jastrzębski S, Włodarczyk-Pruszyński P, Bengio Y, Segler M. RetroGNN: Fast estimation of synthesizability for virtual screening and de novo design by learning from slow retrosynthesis software. *J Chem Inf Model*. 2021;62. <https://doi.org/10.1021/ACS.JCIM.1C01476>
20. Koch M, Duigou T, Faulon J-L. Reinforcement Learning for Bioretrosynthesis. *ACS Synth Biol*. 2020;9(1):157–68. <https://doi.org/10.1021/acssynbio.9b00447> PMID: 31841626
21. Chen B, Li C, Dai H, Song L. Retro*: Learning Retrosynthetic Planning with Neural Guided A* Search. In: *Proceedings of Machine Learning Research*, 2020. 1608–16. <https://proceedings.mlr.press/v119/chen20k.html>
22. Schwaller P, Petraglia R, Zullo V, Nair VH, Haeuselmann RA, Pisoni R, et al. Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chem Sci*. 2020;11(12):3316–25. <https://doi.org/10.1039/c9sc05704h> PMID: 34122839
23. Upadhyay V, Boorla VS, Maranas CD. Rank-ordering of known enzymes as starting points for re-engineering novel substrate activity using a convolutional neural network. *Metabolic Engineering*. 2023;78:171–82. <https://doi.org/10.1016/j.ymben.2023.06.001>
24. Hadadi N, MohammadiPeyhani H, Miskovic L, Seijo M, Hatzimanikatis V. Enzyme annotation for orphan and novel reactions using knowledge of substrate reactive sites. *Proc Natl Acad Sci U S A*. 2019;116(15):7298–307. <https://doi.org/10.1073/pnas.1818877116> PMID: 30910961
25. Carbonell P, Wong J, Swainston N, Takano E, Turner NJ, Scrutton NS, et al. Selenzyme: enzyme selection tool for pathway design. *Bioinformatics*. 2018;34(12):2153–4. <https://doi.org/10.1093/bioinformatics/bty065>
26. Yeh AH-W, Norn C, Kipnis Y, Tischer D, Pellock SJ, Evans D, et al. De novo design of luciferases using deep learning. *Nature*. 2023;614(7949):774–80. <https://doi.org/10.1038/s41586-023-05696-3> PMID: 36813896
27. Beber ME, Gollub MG, Mozaffari D, Shebek KM, Noor E. eEquilibrator 3.0 -- a platform for the estimation of thermodynamic constants. <http://arxiv.org/abs/2103.00621>. 2021.
28. Wang L, Upadhyay V, Maranas CD. dGPredictor: Automated fragmentation method for metabolic reaction free energy prediction and de novo pathway design. *PLoS Comput Biol*. 2021;17(9):e1009448. <https://doi.org/10.1371/journal.pcbi.1009448> PMID: 34570771
29. Chowdhury A, Maranas CD. Designing overall stoichiometric conversions and intervening metabolic reactions. *Sci Rep*. 2015;5:16009. <https://doi.org/10.1038/srep16009> PMID: 26530953
30. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017;45(D1):D353–61. <https://doi.org/10.1093/nar/gkw1092> PMID: 27899662
31. Morgat A, Lombardot T, Axelsen KB, Aimo L, Niknejad A, Hyka-Nouspikel N, et al. Updates in Rhea - an expert curated resource of biochemical reactions. *Nucleic Acids Res*. 2017;45(D1):D415–8. <https://doi.org/10.1093/nar/gkw990> PMID: 27789701
32. Moretti S, Martin O, Van Du Tran T, Bridge A, Morgat A, Pagni M. MetaNetX/MNXref--reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Res*. 2016;44(D1):D523–6. <https://doi.org/10.1093/nar/gkv1117> PMID: 26527720
33. Carbonell P, Wong J, Swainston N, Takano E, Turner NJ, Scrutton NS, et al. Selenzyme: enzyme selection tool for pathway design. *Bioinformatics*. 2018;34(12):2153–4. <https://doi.org/10.1093/bioinformatics/bty065> PMID: 29425325
34. Zheng S, Zeng T, Li C, Chen B, Coley CW, Yang Y, et al. Deep learning driven biosynthetic pathways navigation for natural products with BioNavi-NP. *Nat Commun*. 2022;13(1):3342. <https://doi.org/10.1038/s41467-022-30970-9> PMID: 35688826
35. Moriya Y, Shigemizu D, Hattori M, Tokimatsu T, Kotera M, Goto S, et al. PathPred: an enzyme-catalyzed metabolic pathway prediction server. *Nucleic Acids Res*. 2010;38(Web Server issue):W138–43. <https://doi.org/10.1093/nar/gkq318> PMID: 20435670
36. Carbonell P, Parutto P, Herisson J, Pandit SB, Faulon J-L. XTMS: pathway design in an eXTended metabolic space. *Nucleic Acids Res*. 2014;42(Web Server issue):W389–94. <https://doi.org/10.1093/nar/gku362> PMID: 24792156

37. Wicker J, Lorschbach T, Gütlein M, Schmid E, Latino D, Kramer S, et al. envPath--The environmental contaminant biotransformation pathway resource. *Nucleic Acids Res.* 2016;44(D1):D502–8. <https://doi.org/10.1093/nar/gkv1229> PMID: [26582924](#)
38. Wang L, Upadhyay V, Maranas CD. dGPredictor: Automated fragmentation method for metabolic reaction free energy prediction and de novo pathway design. *PLoS Comput Biol.* 2021;17(9):e1009448. <https://doi.org/10.1371/journal.pcbi.1009448>
39. Burgard AP, Pharkya P, Maranas CD. OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotech & Bioengineering.* 2003;84(6):647–57. <https://doi.org/10.1002/bit.10803>
40. Ranganathan S, Suthers PF, Maranas CD. OptForce: An Optimization Procedure for Identifying All Genetic Manipulations Leading to Targeted Overproductions. *PLoS Comput Biol.* 2010;6(4):e1000744. <https://doi.org/10.1371/journal.pcbi.1000744>
41. Salis HM. The ribosome binding site calculator. *Methods Enzymol.* 2011;498:19–42. <https://doi.org/10.1016/B978-0-12-385120-8.00002-4> PMID: [21601672](#)
42. Pu L, Naderi M, Liu T, Wu HC, Mukhopadhyay S, Brylinski M. EToxPred: A machine learning-based approach to estimate the toxicity of drug candidates. *BMC Pharmacology and Toxicology.* 2019;20:1–15. <https://doi.org/10.1186/S40360-018-0282-6>
43. van der Hoek SA, Borodina I. Transporter engineering in microbial cell factories: the ins, the outs, and the in-betweens. *Curr Opin Biotechnol.* 2020;66:186–94. <https://doi.org/10.1016/j.copbio.2020.08.002> PMID: [32927362](#)
44. Sankar K, Krystek SR Jr, Carl SM, Day T, Maier JKX. AggScore: Prediction of aggregation-prone regions in proteins based on the distribution of surface patches. *Proteins.* 2018;86(11):1147–56. <https://doi.org/10.1002/prot.25594> PMID: [30168197](#)
45. Pujols J, Peña-Díaz S, Ventura S. AGGRESCAN3D: Toward the Prediction of the Aggregation Propensities of Protein Structures. *Methods Mol Biol.* 2018;1762:427–43. https://doi.org/10.1007/978-1-4939-7756-7_21 PMID: [29594784](#)
46. Atsumi S, Liao JC. Directed evolution of *Methanococcus jannaschii* citramalate synthase for biosynthesis of 1-propanol and 1-butanol by *Escherichia coli*. *Appl Environ Microbiol.* 2008;74(24):7802–8. <https://doi.org/10.1128/AEM.02046-08> PMID: [18952866](#)
47. Yang KK, Wu Z, Arnold FH. Machine-learning-guided directed evolution for protein engineering. *Nat Methods.* 2019;16(8):687–94. <https://doi.org/10.1038/s41592-019-0496-6> PMID: [31308553](#)
48. Qiu Y, Wei G-W. CLADE 2.0: Evolution-Driven Cluster Learning-Assisted Directed Evolution. *J Chem Inf Model.* 2022;62(19):4629–41. <https://doi.org/10.1021/acs.jcim.2c01046> PMID: [36154171](#)
49. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583–9. <https://doi.org/10.1038/s41586-021-03819-2> PMID: [34265844](#)