Journal of Computer-Aided Molecular Design (2025) 39:32
https://doi.org/10.1007/s10822-025-00604-5

Thermodynamic free energy map for the non-oxidative glycolysis pathways

Adittya Pal[1]

Received: 28 March 2025 / Accepted: 13 May 2025 / Published online: 16 June 2025

Abstract
Designing reaction pathways that maximize the production of a target compound in a given metabolic network is a fundamental problem in systems biology. In this study, we systematically explore the non-oxidative glycolysis metabolic network, guided by the principle that reactions with negative Gibbs free energy differences are thermodynamically favored. We enumerate alternative pathways that implement the net non-oxidative glycolysis reaction, categorized by their length. Our analysis reveals several alternative thermodynamically favorable pathways beyond the experimentally reported ones. Additionally, we identify molecules within the network, such as 3-hydroxypropionic acid, that may have significant potential for further investigation.

Keywords Non-oxidative glycolysis · Metabolic networks · Thermodynamic analysis of molecular networks · Systems biology · Thermodynamics in biology · Mixed integer linear programming · 3-hydroxypropionic acid

Introduction

Glycolysis is a catabolic pathway that converts carbohydrates to acetylphosphate while releasing energy, catalyzed by enzymes. The pathway that occurs in most organisms does not achieve the theoretically optimal atom efficiency because two out of every six carbon atoms in the carbohydrate are released as carbon dioxide, which is lost to the atmosphere. Synthetically designed pathways, however, have reached the theoretical limit of 100% carbon efficiency, where no carbon atoms are lost as carbon dioxide molecules [1]. These pathways are called non-oxidative glycolysis pathways because no carbon atoms are fully oxidized to carbon dioxide (where the carbon atom has an oxidation number of +4 as compared to +3 in acetylphosphate).

While elaborating on nonoxidative glycolysis pathways, the authors in [1] point out that the reaction catalyzed by the enzyme phosphoketolase is an irreversible step, providing the driving force for the nonoxidative glycolytic pathway. The enzyme phosphoketolase can catalyze three different reactions (all following the same template):

$$Xpk : X5P + P_i \longrightarrow G3P + AcP + H_2O \quad (1)$$

$$Fpk : F6P + P_i \longrightarrow E4P + AcP + H_2O \quad (2)$$

$$Spk : S7P + P_i \longrightarrow R5P + AcP + H_2O \quad (3)$$

This promiscuous activity of the phosphoketolase enzyme allows for variations in the nonoxidative glycolytic pathway. Three experimentally observed pathways are enumerated in 1, 2, and 3. The fourth pathway, described in 4, uses the phosphoketolase enzyme to cleave three sugars-xylulose 5-phosphate, fructose 6-phosphate, and sedoheptulose 7-phosphate-in a single reaction sequence. [1] notes that similar pathways can be derived using sedoheptulose biphosphate instead of fructose biphosphate, combinations of which can generate a large number of pathways for non-oxidative glycolysis. An example of a pathway using sedoheptulose biphosphate was reported in [2].

We aim to evaluate the possible pathways in the non-oxidative glycolysis metabolic network to develop a physical justification for why some pathways are easier to realize under laboratory conditions than others. Furthermore, we aim to explain why certain by-products, such

✉ Adittya Pal
adpal@imada.sdu.dk

[1] Institut for Matematik og Datalogi, Syddansk Universitet, Campusvej 55, 5230 Odense, Denmark

Table 1 The sequence of reactions for the non-oxidative glycolytic pathway using Fpk only from [1] with optimal carbon efficiency

F6P	+	E4P	⟶	G3P	+	S7P		
G3P	+	S7P	⟶	X5P	+	R5P		
X5P	+	E4P	⟶	F6P	+	G3P		
F6P	+	P_i	⟶	E4P	+	AcP	+	H_2O
		R5P	⟶	Ru5P				
		Ru5P	⟶	X5P				
X5P	+	E4P	⟶	F6P	+	G3P		
F6P	+	P_i	⟶	E4P	+	AcP	+	H_2O
		G3P	⟶	DHAP				
G3P	+	DHAP	⟶	FBP				
FBP	+	H_2O	⟶	F6P	+	P_i		
F6P	+	P_i	⟶	E4P	+	AcP	+	H_2O
F6P	+	$2P_i$	⟶	3	AcP	+ 2	H_2O	

Table 2 The sequence of reactions for the non-oxidative glycolytic pathway using Xpk only from [1] with optimal carbon efficiency

F6P	+	E4P	⟶	G3P	+	S7P		
G3P	+	S7P	⟶	X5P	+	R5P		
X5P	+	P_i	⟶	G3P	+	AcP	+	H_2O
		R5P	⟶	Ru5P				
		Ru5P	⟶	X5P				
X5P	+	P_i	⟶	G3P	+	AcP	+	H_2O
		G3P	⟶	DHAP				
G3P	+	DHAP	⟶	FBP				
FBP	+	H_2O	⟶	F6P	+	P_i		
F6P	+	G3P	⟶	E4P	+	X5P		
X5P	+	P_i	⟶	G3P	+	AcP	+	H_2O
F6P	+	$2P_i$	⟶	3	AcP	+ 2	H_2O	

Table 3 The sequence of reactions for the non-oxidative glycolytic pathway from [1] using both Xpk and Fpk with optimal carbon efficiency

F6P	+	E4P	⟶	G3P	+	S7P		
G3P	+	S7P	⟶	X5P	+	R5P		
X5P	+	P_i	⟶	G3P	+	AcP	+	H_2O
		R5P	⟶	Ru5P				
		Ru5P	⟶	X5P				
X5P	+	P_i	⟶	G3P	+	AcP	+	H_2O
		G3P	⟶	DHAP				
G3P	+	DHAP	⟶	FBP				
FBP	+	H_2O	⟶	F6P	+	P_i		
F6P	+	P_i	⟶	E4P	+	AcP	+	H_2O
F6P	+	$2P_i$	⟶ 3	AcP	+ 2	H_2O		

as 3-hydroxypropionate in [4], are observed alongside acetylphosphate in experiments attempting to optimize the nonoxidative glycolytic pathway. Finally, we seek to identify potential bottlenecks in these pathways and propose a few that may be the most feasible under practical conditions.

Methods

Traditionally, constraint-based models have been used to analyze metabolic networks under the steady-state assumption. These constraint-based models can also incorporate kinetic parameters for each (enzymatic) reaction as a further refinement [5, 6]. However, these kinetic parameters

Table 4 The sequence of reactions for the non-oxidative glycolytic pathway using all Xpk, Fpk and Spk from [3] with optimal carbon efficiency

F6P	+	E4P	⟶	G3P	+	S7P		
S7P	+	P_i	⟶	R5P	+	AcP	+	H_2O
		R5P	⟶	Ru5P				
		Ru5P	⟶	X5P				
X5P	+	P_i	⟶	G3P	+	AcP	+	H_2O
		G3P	⟶	DHAP				
G3P	+	DHAP	⟶	FBP				
FBP	+	H_2O	⟶	F6P	+	P_i		
F6P	+	P_i	⟶	E4P	+	AcP	+	H_2O
F6P	+	$2P_i$	⟶ 3	AcP	+ 2	H_2O		

are often not available for all the reactions in the considered network [7, 8]. Conservation laws, such as the conservation of mass, can be easily modeled to ensure that the rate of production of molecules by certain reactions equals the rate at which they are consumed in the system. In addition to this conservation law, thermodynamics is another governing principle for most physical processes in equilibrium. An important thermodynamic variable in metabolic networks, besides kinetic parameters, is the chemical potential of molecules in the network. For a reaction, $\Delta_r G$ represents the difference in the chemical potentials of the product molecules and the reactant molecules involved in the reaction. Thermodynamics favors only those reactions with a negative $\Delta_r G$ value (often termed as reactions that proceed 'downhill').

A critical variable affecting the chemical potential of molecules in a system is their concentration. Modeling the concentrations of molecules in living organisms is challenging because they are actively manipulated by transport across membranes through pumps and channels via active transport, which consumes energy in the form of small molecules such as ATP and NADPH. These driven processes are often difficult to assign free-energy differences to, and hence, challenging to encapsulate within a network. Additionally, different parts of a metabolic network might operate in distinct cellular compartments, with varying concentrations of the same molecule. To avoid these complexities, we assume that the reactions in the analyzed metabolic network occur in a well-stirred bioreactor with the required enzymes present, as in an *in vitro* assessment of a hypothesized pathway before it is engineered into a living cells. It is admitted that under these assumption— not accounting for the free-energy differences of the transport processes and the differential concentration of the same molecule in different cellular compartments— would introduce errors into the modeling.

Constructing the reaction network

The metabolic reaction networks are generated using the graph grammar approach in MØD, as elaborated in [9, 10]. The generic expansion of the molecular space was achieved through the recursive application of reaction templates, represented as graph transformation rules, on molecules represented as graphs. The procedure is illustrated in Algorithm 1 and detailed in [11]. The vertices in these (molecular) graphs correspond to atoms, and the edges represent bonds. Hence, subraphs would correspond to molecular fragments. However, this approach omits stereoinformation for the molecules, which is often critical for metabolic reactions. As a result, stereoisomeric carbohydrates with identical graph representations, such as erythrose and threose, ribose and xylose, or ribulose and xylulose, cannot be distinguished. The graph transformation rules modeling each enzymatic reaction are listed in Table 5 and elaborated in Appendix A. The molecular space was expanded from a set of starting molecules listed in Appendix Table 12 in Appendix A, until no new molecules with fewer than eight carbon atoms could be added. This restriction makes the molecular space finite and reasonable, as carbohydrates with eight or more carbons are typically unstable and rarely observed in metabolic pathways. The expanded molecular space contains 81 molecules, and the reaction networks generated consisted of 414 individual reactions.

Table 5 Graph grammar transformation rules for expanding the molecular space

Abbrev.	Name of the enzyme	Function
AL	Aldolase	performs a generic aldol addition.
AlKe	Aldose-Ketose	isomerizes aldehydes to ketones.
KeAl	Ketose-Aldose	isomerises ketones to aldehydes.
PHL	Phosphohydrolase	hydrolyses the phosphate group.
PK	Phosphoketolase	breaks C–C keto bond, adds phosphate.
TAL	Transaldolase	breaks an aldol bond (generally C3).
TKL	Transketolase	cleaves a keto group (generally C2).

Algorithm 1 Generating the reaction network

Define P_0 : source collection of molecules as graphs
Define $\mathcal{R}$: set reaction templates as graph transformation rules $X \xrightarrow{r} Y$, where $r \in \mathcal{R}$ transforms subgraph X to subgraph Y
Define constraints as right predicates, $c_i : p \to \{0, 1\}$, where constraint c_i decides if graph p (with subgraph Y) is allowed or not.
i = 0 ▷ recursion step counter
while graph $p \in P_i$ satisfies constraint c_i **do**
    i++ ▷ increment step
    $P_i = \left\{ q \in Q : p \xrightarrow{r} q, p \subset \bigcup_{0 \le j < i} P_j, r \in \mathcal{R} \right\}$
(this expands the set of existing molecules P_i at the recursion step i, by producing a collection of new molecules Q through the application of the rules r)
end while
return $\bigcup_{0 \le i} P_i$ ▷ union of sets of molecules from all steps

Enumerating the relevant pathways

Pathways in living organisms metabolize carbohydrates such that two carbon atoms from a six-carbon sugar (e.g., fructose) are lost as carbon dioxide molecules. Our goal is to identify pathways that do not produce carbon dioxide from starting carbohydrates but instead convert all six carbon atoms into three acetylphosphate molecules. Thus, we set a search query that restricts the inflow of fructose 6-phosphate to one and the outflow of acetylphosphate to three. If the number of acetylphosphate molecules produced by the pathway is less than three, the carbon efficiency of the pathway would be suboptimal. The search query is formulated as an integer linear program (ILP) to assign integral flows to the reactions, representing the number of times each reaction must occur in the pathway to produce the target molecules from the given source molecule. The mass conservation law is represented as a flow conservation constraint for all vertices v, where the flow f_e on a hyperedge e is the number of times the reaction represented by the hyperedge occurs in the pathway. For all vertices, the number of copies of the molecule used by the reactions e where v is a reactant (in(v)) must be balanced by the number of copies of the molecule produced by reactions e where v is a product (out(v)) in steady state. Mathematically,

$$\forall v \in V : \sum_{e \in \text{in}(v)} f_e = \sum_{e \in \text{out}(v)} f_e \tag{4}$$

Multiplying the flow f_e by a positive integer also produces a valid flow that satisfies the mass conservation law. Therefore, the inflow is restricted to one fructose 6-phosphate molecule to eliminate superfluous solutions, which are base solutions scaled by a positive integer. An additional constraint ensures that all carbohydrate molecules in the pathway must have a phosphate subgraph, making the returned pathways more realistic. The formulated ILP to search for nonoxidative glycolysis pathways was:

$$\min \left(\sum_{e \in E} f_e \right) \tag{5}$$

$$\forall e \in E : f_e \ge 0 \tag{6}$$

$$\forall v \in V : \sum_{e \in \text{out}(v)} f_e - \sum_{e \in \text{in}(v)} f_e = 0 \tag{4}$$

$$\text{inFlow}[F6P] = 1 \tag{7}$$

$$\text{outFlow}[AcP] = 3 \tag{8}$$

For pathways to 3-hydroxypropionic acid, constraint (8) was replaced by outFlow[HPA] = 3. The formulated ILP had 510 variables, 184 constraints and was solved using a generic ILP solver; in this case, the open-source Coin-OR CBC solver [12] was used. It took the ILP solver ~ 190 seconds to solve for the two pathways using seven reactions described in Section 4.1. The ILP enumerates pathways based on the number of distinct reactions utilized, not the total number of reactions in the pathway. Therefore, if a reaction is used twice (with a flow of two), it is counted only once when ordering the pathways returned by the ILP solver.

Assignment of chemical potentials to vertices

Representing molecules as graphs preserves the connectivity information of the atoms but does not include details about the relative placement of the atoms in three-dimensional space. To address this, we generate a three-dimensional

embedding for the atoms in the molecular graph and use that embedding to assign a chemical potential to the molecule under standard physical conditions. OpenBabel [13] was used to create an initial guess for the three-dimensional embedding of the molecule using a combination of elementary rules and frequently encountered molecular subgraphs. However, these initial guesses often contain substructures with steric clashes or high strain and, therefore, require optimization.

The torsional angles in the molecule are adjusted to locate a lower-energy conformer, where the energy of the resulting conformer is estimated using an elementary force field. In this work, the universal force field (UFF) [14] was used. The Cartesian coordinates of the optimized embedding of the molecule were written out as an `.sdf` file.

The structures generated by OpenBabel cannot be directly used to estimate the chemical potentials because they are based on empirical knowledge of a class of molecules. Semi-empirical methods, such as xTB [15], offer a fast yet accurate way to estimate thermodynamic parameters after optimizing the geometry of the molecule (given a reasonable initial guess). However, as a chemical tool, xTB requires the input file for the molecule to be in one of the standard chemical formats. OpenBabel serves as a bridge, converting graphs with only connectivity information into standard structure data files readable by xTB. xTB iteratively optimizes the geometry of the molecule from the input `.sdf` file until the energy change between iterations falls below an acceptable tolerance. The optimized geometry is then used to assign the chemical potential to the molecule. xTB calculates the thermo-statistical translational, rotational, and vibrational contributions to the electronic energy to determine the chemical potentials, using the rigid-rotor and harmonic-oscillator approximations.

The workflow for assigning chemical potentials to molecules is depicted in Figure 1.

Thermodynamic analyses of pathways

Pathways enumerated by the ILP solver satisfy the constraints of mass conservation, but they must also be realizable under practical reactor-like settings. While practical realizability (encapsulated as whether a reaction is thermodynamically favorable, i.e., has a negative Gibbs free energy difference under certain concentrations of the involved molecules) can be integrated into the ILP modeling as in [16, 17], in this work, we perform a post-processing of the ILP solutions using the calculated chemical potentials. We borrow concepts from probability to understand how energy tends to distribute itself in the reaction network, making some reactions thermodynamically favorable. At first glance, this might seem contradictory, as physical laws are universal and deterministic, not probabilistic.

We consider the following statement for the second law of thermodynamics:

> When a system is composed of a *large number of molecules*, it is *exceedingly likely* to *spontaneously* move to a thermodynamic macrostate that *maximizes* the number of possible microstates.

A caveat is that the system must consist of a large number of molecules, which might not always be the case in cellular compartments in living organisms. The statistics become overwhelming (so that they can be formulated as a statistical law of large numbers) only when the number of molecules in the system is large-hence the justification for using the phrase *exceedingly likely*.

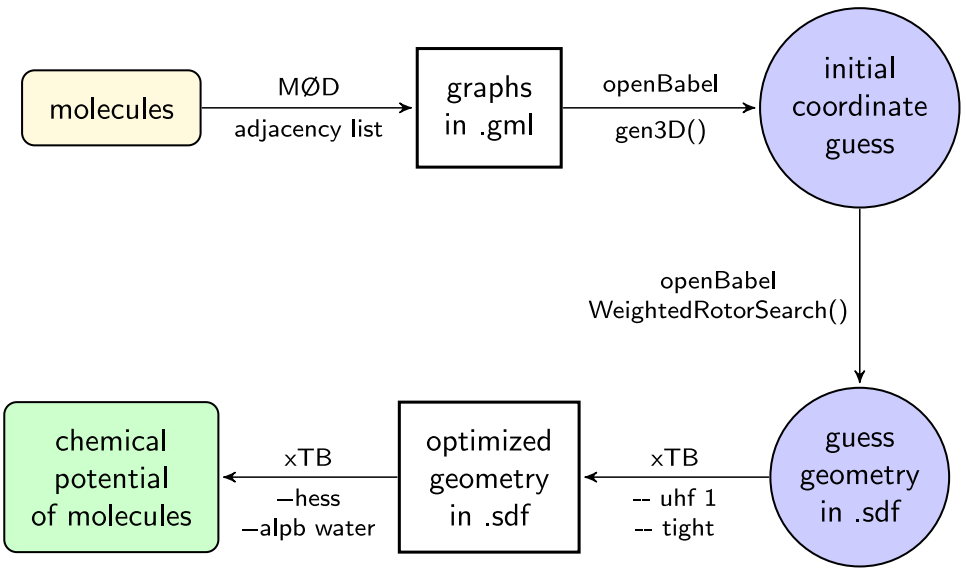


Fig. 1 Sequence of operations to estimate the chemical potential of a molecule at standard physical conditions (273.15 K and 10^5 Pa) from the connectivity information of a graph that represents a molecule

Not all microstates available to the system are equally probable. The probability of a microstate depends on its energy E, through the Boltzmann factor $\sim \exp(-E/k_BT)$. Furthermore, there may be several microstates with the same energy, which is related to the entropy S of the system by the Boltzmann–Planck equation: $S = k_B \ln w$, where w is the number of microstates in the system and k_B is the Boltzmann constant. Therefore, the probability, $P(E, T)$ that a system is in a macrostate with energy E and w microstates at temperature T is given by:

$$P(E,T) \sim w \cdot \exp(-E/k_BT) = \exp(S/k_B) \cdot \exp(-E/k_BT) = \exp\left(-\frac{E-TS}{k_BT}\right) \tag{9}$$

In the context of reactions, we are often concerned with the relative probabilities of two macrostates. The ratio of probabilities of two macrostates with energy E_1 and E_2 is given by:

$$\frac{P(E_1,T)}{P(E_2,T)} = \exp\left(-\frac{(E_1-E_2)-T(S_1-S_2)}{k_BT}\right)$$

Under conditions of constant pressure (commonly encountered in cellular and bioreactor environments), the energy E can be replaced by the enthalpy H. The numerator of the exponent then becomes $(H_1 - H_2) - T(S_1 - S_2) = \Delta H - T\Delta S = \Delta_r G$. The quantity $\Delta_r G$ is the Gibbs free energy difference for the reaction and indicates the relative likelihood of the reaction proceeding in a particular direction.

For a reaction r : LHS $\longrightarrow$ RHS, if $\Delta_r G < 0$, then the ratio $P_{\text{RHS}}/P_{\text{LHS}} > 1$, meaning the system is more likely to be in the state represented by the RHS. Conversely, if $\Delta_r G > 0$, the system prefers the state represented by the LHS.

For pathways, the Gibbs free energy is a state function and thus independent of the sequence of reactions. Therefore, all pathways effecting the net reaction for non-oxidative glycolysis:

frustose 6-phosphate + 2 phosphoric acid $\longrightarrow$ 3 acetyl phosphate + $2H_2O$

will have the same net Gibbs free energy difference, given by $\Delta_r G = 3G(AcP) + 2G(H_2O) - G(F6P) - 2G(P_i)$. However, different pathways use different intermediate reactions, resulting in varying intermediate compositions and thermodynamic free energy profiles. Ideally, all constituent reactions in a pathway should be spontaneous in the required direction. For consecutive reactions with negative Gibbs free energy differences, the molecules on the RHS of the first reaction are more likely to appear, forming a subset that serves as the LHS of the subsequent reaction. This establishes a cascade of reactions, where molecules are consecutively transformed into ones with a high probability of being observed, provided all reactions in the pathway have negative Gibbs free energy differences.

Conversely, a reaction with a positive Gibbs free energy difference represents a bottleneck. A reaction i with $\Delta_r G_i > 0$ does not imply that it will *not* occur. It simply means the reaction is less likely to proceed in the required direction compared to the reverse reaction. In cellular environments, energy-driven processes and transport mechanisms actively manipulating molecule concentrations can render such reactions thermodynamically favorable under those conditions, even if they are not under standard physical conditions. Furthermore, the chemical potentials of the molecules, from which the free energy differences are calculated, are estimated under certain approximations. Errors in these estimations could lead to incorrect conclusions.

Results

We enumerate pathways in the generated reaction network, and the distribution of the pathways by the number of reactions used, the number of unique reactions utilized, and the biphosphate used is shown in Table 6. Intuitively, we hypothesize that pathways utilizing a smaller number of reactions ('shorter paths' in the network) might be favored because it would be simpler to construct a corresponding bioreactor if each reaction takes place in a separate compartment.

Table 6 Count of the different pathways enumerated by the ILP solver

Type of biphosphate used		XBP	FBP	SBP	Total	
7 reactions	7 unique reactions	1	0	1		2
8 reactions	7 unique reactions	6	0	6	12	29
	8 unique reactions	7	2	8	17	
9 reactions	7 unique reactions	3	0	3	6	207
	8 unique reactions	42	12	48	102	
	9 unique reactions	38	17	44	99	

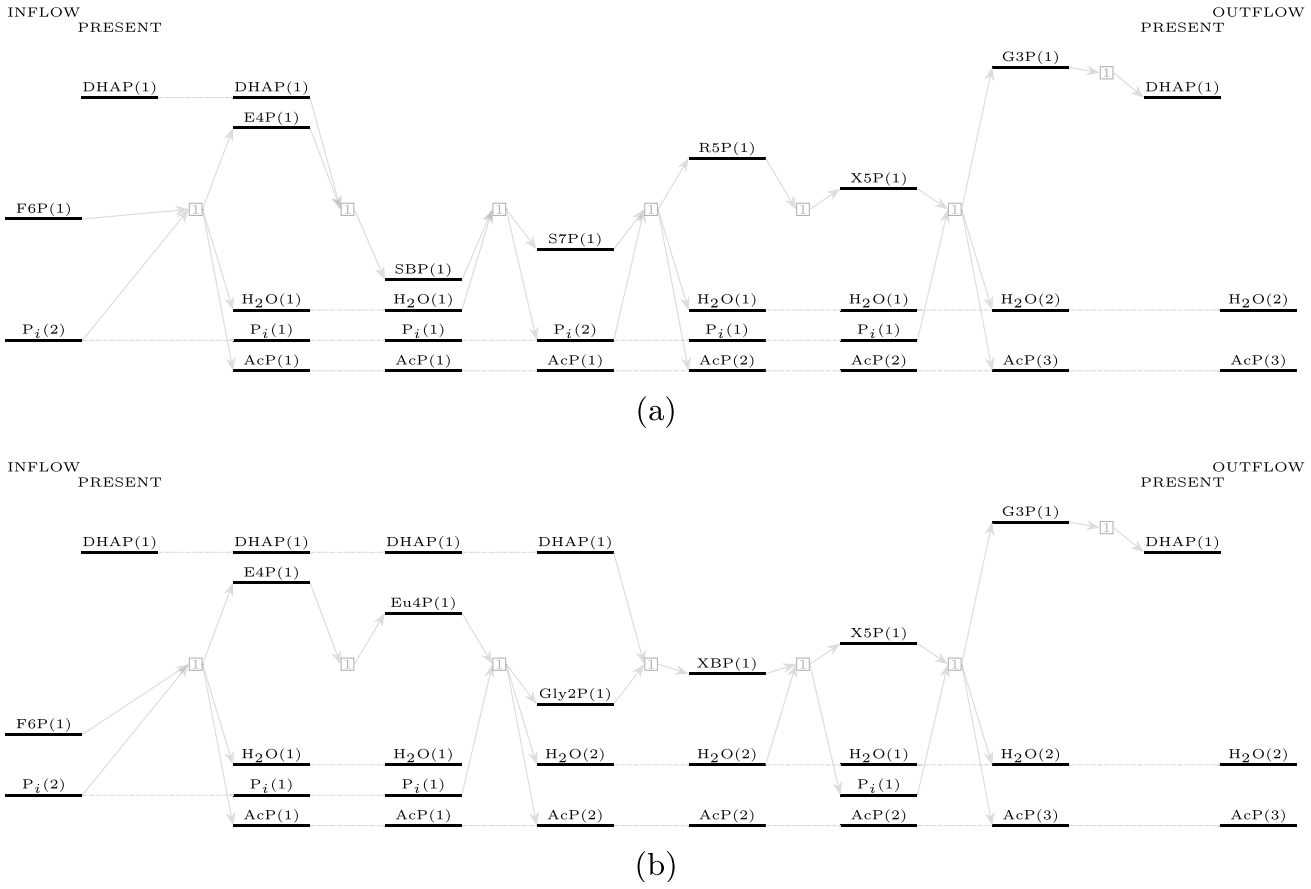


Fig. 2 Two pathways for non-oxidative glycolysis using 7 reactions: 2ausing sedoheptulose-1,7-biphosphate (top panel) and 2busing xylulose-1,5-biphosphate (bottom panel)

Moreover, pathways utilizing fewer reactions would be easier to study and monitor under laboratory conditions.

Pathways with seven reactions

Two non-oxidative glycolytic pathways with seven reactions were found in the expanded reaction network and are shown in Figure 2a and b. The first one uses sedoheptulose-1,7-biphosphate, while the second uses xylulose-1,5-biphosphate. The first pathway has been studied in [2] under laboratory conditions and was reported to "overcome bottlenecks from the previously reported non-oxidative glycolysis cycles." Both pathways use only four of the seven templated enzymatic reactions used to expand the reaction network: aldolase, aldose-ketose isomerase, phosphohydrolase, and phosphoketolase.

We elaborate on the individual reactions in the pathways here. Both pathways have two reactions requiring one substrate molecule, while the remaining five reactions require two substrate molecules. Only one reaction requires two carbohydrate molecules as a substrate, which has been highlighted in both pathways. All the other bimolecular reactions use either water or phosphate molecules, which are present in abundance. Steady-state simulations of the non-oxidative glycolytic pathways by [2] have shown that the intermediate sedoheptulose-7-phosphate starts accumulating over time, limiting the efficiency of the pathway. They addressed this problem by introducing a specific phosphoketolase into the system, which removed the kinetic bottleneck in the pathway. The second pathway with seven reactions, depicted in Table 8 found in the reaction network, does not produce sedoheptulose-7-phosphate at all. However, a drawback of this pathway is that while phosphoketolase has been widely reported to be active on substrates such as fructose-6-phosphate, xylulose-5-phosphate [18], and sedoheptulose-7-phosphate [2], it has not been reported to act on erythrulose-4-phosphate, as required by the third reaction in the pathway shown in 8. However, if one observes the presence of erythrulose-4-phosphate, glycolaldehyde-2-phosphate, and xylulose-1,5-biphosphate in a reactor effecting the non-oxidative glycolytic pathway with optimal carbon efficiency using only these four enzymes, it would act as evidence that the pathway shown in 8 is also practical.

Table 7 The sequence of reactions with the respective enzymes for the non-oxidative glycolytic pathway using 7 reactions using sedoheptulose-1,7-biphosphate

F6P	+	P_i	$\xrightarrow{PK}$	E4P	+	H_2O	+	AcP
E4P	+	DHAP	$\xrightarrow{Al}$		SBP			
SBP	+	H_2O	$\xrightarrow{PHL}$	S7P	+	P_i		
S7P	+	P_i	$\xrightarrow{PK}$	R5P	+	H_2O	+	AcP
		R5P	$\xrightarrow{AlKe}$	X5P				
X5P	+	P_i	$\xrightarrow{PK}$	G3P	+	AcP	+	H_2O
		G3P	$\xrightarrow{AlKe}$	DHAP				
F6P	+	$2P_i$	$\longrightarrow 3$	AcP	+ 2	H_2O		

Table 8 The sequence of reactions with the respective enzymes for the non-oxidative glycolytic pathway using 7 reactions using xylulose-1,5-biphosphate

F6P	+	P_i	$\xrightarrow{PK}$	E4P	+	H_2O	+	AcP
		E4P	$\xrightarrow{AlKe}$	Eu4P				
Eu4P	+	P_i	$\xrightarrow{PK}$	Gly2P	+	H_2O	+	AcP
Gly2P	+	DHAP	$\xrightarrow{Al}$	XBP				
XBP	+	H_2O	$\xrightarrow{PHL}$	X5P	+	P_i		
X5P	+	P_i	$\xrightarrow{PK}$	G3P	+	AcP	+	H_2O
		G3P	$\xrightarrow{AlKe}$	DHAP				
F6P	+	$2P_i$	$\longrightarrow 3$	AcP	+	2	H_2O	

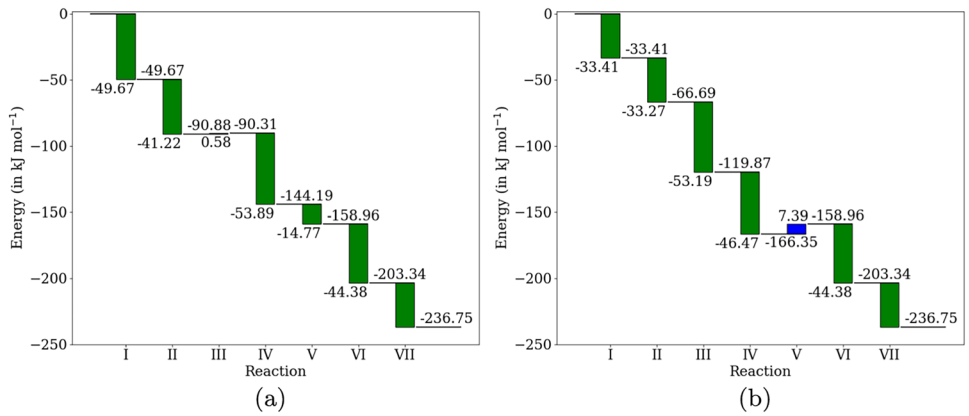


Fig. 3 Thermodynamic free energy profiles for the two pathways for non-oxidative glycolysis using 7 reactions for pathways 7 (left) and 8 (right) respectively

We determined the free energy differences for each reaction used in the two pathways in 7 and 8. The thermodynamic free energy profiles for these two pathways are plotted in Figure 3a and b, respectively. The green bars represent the Gibbs free energy differences for thermodynamically favorable reactions ($\Delta_r G < 0$), while the blue bars represent the Gibbs free energy differences for reactions that are not thermodynamically favorable ($\Delta_r G > 0$) under standard physical conditions. The bars are annotated with the numerical values of the Gibbs free energy differences for the respective reactions. Since free energy is a state function, the net difference in Gibbs free energy for the overall pathway is the same for both pathways: −236.75 kJ mol^{-1}.

In both pathways, the hydrolysis of the biphosphate by the phosphohydrolase is not a thermodynamically favorable reaction. While the hydrolysis of sedoheptulose-1,7-biphosphate has a positive Gibbs free energy difference of 0.58 kJ mol^{-1}, the hydrolysis of xylulose-1,5-biphosphate has a higher value of 7.39 kJ mol^{-1}. The hydrolysis reaction has been observed to be sensitive to the pH of the environment and the activity of water molecules. The hydrolysis of the biphosphate would be expected to be a thermodynamic bottleneck for both pathways. In reality, the cell might adjust the physical conditions in the region where the hydrolysis takes place to make the reaction favorable. Moreover, the free energy of the educts and the products is not simply the sum of those of individual molecules because of the complex intermolecular interactions (hydrogen bonding, polar interactions, hydrophobic effects) that might take place between molecules, especially in solution.

Table 9 Some characteristics of the previously experimentally reported non-oxidative glycolytic pathways

Pathway	1	2	3	4
#reactions[1]	11	10	9	8
#hyperedges[2]	8	8	8	8
#enzymes[3]	7	7	7	6 (no transketolase)
#pk type[4]	(0,3,0)	(3,0,0)	(2,1,0)	(1,1,1)
#bimolecular[5]	9	8	7	6
#two sugars[6]	5	4	3	2

[1] count of the total number of reactions used in the pathway: $\sum_e f_e$

[2] count of the number of unique reactions used in the pathway: $\sum_{e:f_e>0} 1$

[3] count of the different enzymes used in the pathway

[4] 3-tuple representing the count of the number of reactions using (Xpk, Fpk, Spk) respectively

[5] count of the number of bimolecular reactions in the pathway

[6] count of the number of reactions where both the reactant molecules are carbohydrates

Previously reported pathways

In this section, we analyze the four previously reported non-oxidative glycolysis pathways listed in Tables 1, 2, 3, and 4 and compare them to the shortest pathways described in Section 4.1. Note that the count of reactions in Table 9, listing the various characteristics of the pathways, is one less than those in the pathways listed in Tables 1, 2, 3, and 4 because we cannot represent stereo-information in molecular graphs. Hence, the molecules ribulose-5-phosphate and xylulose-5-phosphate have the same representation as graphs. The two reactions $R5P \longrightarrow Ru5P \longrightarrow X5P$ are counted as one. All the pathways use fructose-1,6-biphosphate.

The reactions involving two carbohydrate molecules as substrates (catalyzed by the transaldolase or the transketolase enzymes) might be potential kinetic bottlenecks in the pathway. We generate the thermodynamic profiles for the four pathways to deduce which reactions might act as thermodynamic bottlenecks. The thermodynamic free energy profiles indicate that the first reaction catalyzed by the transaldolase in all four pathways is unfavorable. The hydrolysis of fructose-1,6-biphosphate catalyzed by phosphohydrolase has a marginally negative Gibbs free energy difference. This was expected from the free energy differences for the hydrolysis of the five-carbon and seven-carbon biphosphates. Except for the first pathway 1 (which has three thermodynamically unfavorable reactions), all the pathways have only one thermodynamically unfavorable reaction. Increasing the total number of reactions in the pathway did not have an adverse effect on the free energy differences of the individual reactions.

Two factors might account for the large number of possible pathways for a non-oxidative glycolytic pathway: the promiscuous activity of phosphoketolase and the biphosphate that might be hydrolyzed. The phosphoketolase acting on different substrates, as shown in reactions 1, 2, and 3, have comparable free energy differences: −44.376 kJ mol^{-1}, −49.666 kJ mol^{-1} and −53.887 kJ mol^{-1}, respectively. Regarding the hydrolysis of the biphosphates, the hydrolysis of fructose-1,6-biphosphate has a negative Gibbs free energy difference of 0.0322 kJ mol^{-1}, while

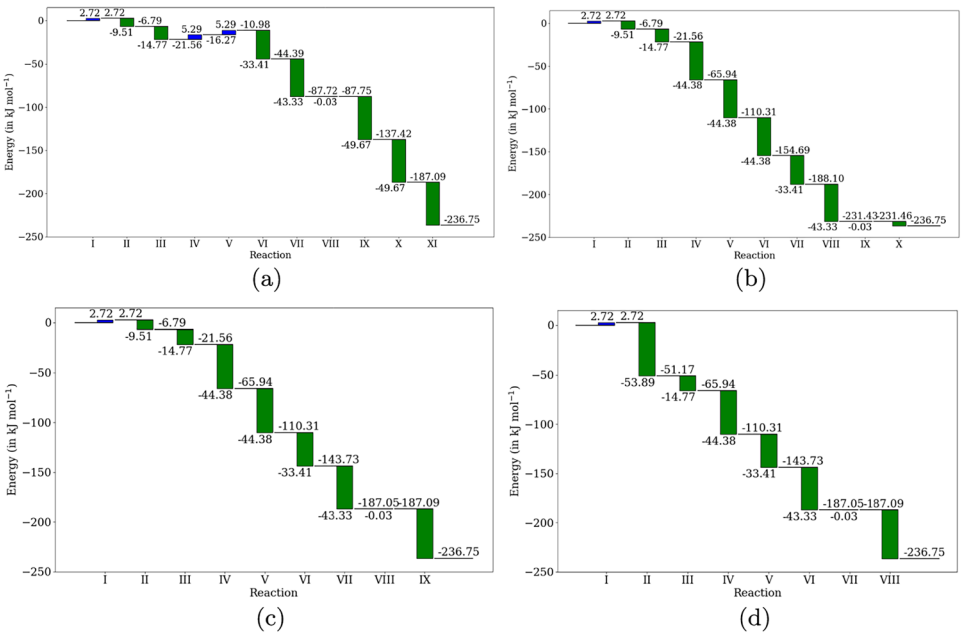


Fig. 4 Thermodynamic free energy profiles for the experimentally reported pathways 1 (top left), 2 (top right), 3 (bottom left) and 4 (bottom right) for non-oxidative glycolysis

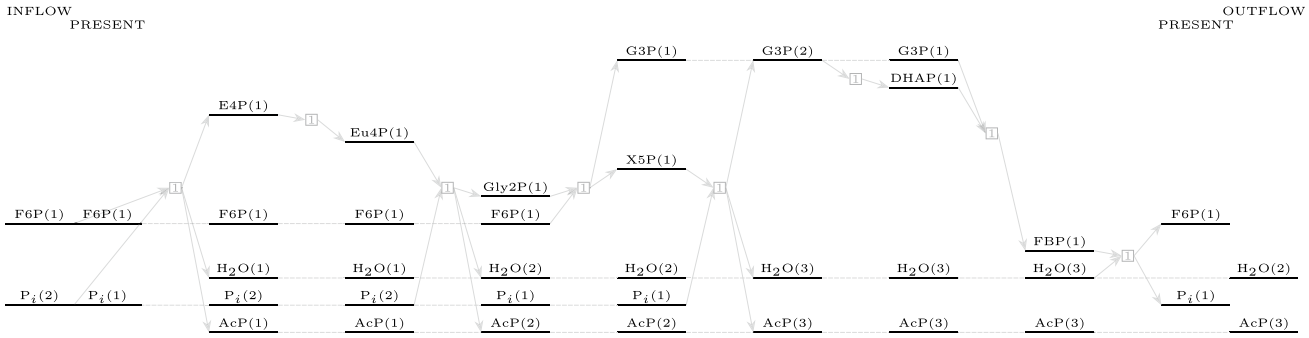


Fig. 5 An alternative non-oxidative glycolytic pathway using fructose-1,6-biphosphate and eight reactions

Table 10 The sequence of reactions in the alternative non-oxidative glycolytic pathway which uses eight reactions and fructose-1,6-biphosphate

F6P	+	P_i	$\xrightarrow{PK}$	E4P	+	AcP	+	H_2O
		E4P	$\xrightarrow{AlKe}$	Eu4P				
Eu4P	+	P_i	$\xrightarrow{PK}$	Gly2P	+	H_2O	+	AcP
Gly2P	+	F6P	$\xrightarrow{PAL}$	X5P	+	G3P		
X5P	+	P_i	$\xrightarrow{PK}$	G3P	+	AcP	+	H_2O
	G3P	+	$\xrightarrow{AL}$	DHAP				
G3P	+	DHAP	$\xrightarrow{AL}$	FBP				
FBP	+	H_2O	$\xrightarrow{PHL}$	F6P	+	P_i		
F6P	+	$2P_i$	$\longrightarrow 3$	ACP	+ 2	H_2O		

the hydrolysis of both xylulose-1,6-biphosphate and sedoheptulose-1,7-biphosphate have positive Gibbs free energy differences: 7.386 kJ mol^{-1} and 0.576 kJ mol^{-1}, respectively. The free energies for reactions can be altered by changing the concentrations of the molecules involved. The hydrolysis of sedoheptulose-1,7-biphosphate is quite close to equilibrium.

$$\Delta_r G = -RT \ln\left(\frac{[\mathrm{S7P}][\mathrm{P}_i]}{[\mathrm{SBP}][H_2O]}\right)$$

A ratio of the concentration of only 1.26 for [SBP]/[S7P] makes the reaction in thermal equilibrium at 300 K, considering unit activities of water and phosphate molecules. The corresponding value for the ratio of [XBP]/[X5P] so that the reaction is in equilibrium is 19.32. These concentration gradients can be maintained by the active transport of the molecules involved in the reactions (Fig. 4).

Alternative equivalent non-oxidative glycolysis pathways

The results from the previous section indicate that the hydrolysis of fructose-1,6-biphosphate is thermodynamically favored compared to the immediate higher or lower

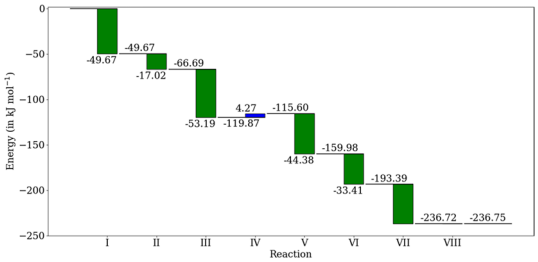


Fig. 6 The thermodynamic free energy profile for the pathway in Table 10

sugars. An exhaustive search for pathways in the generated molecular space indicates that there is an alternative pathway, shown in Figure 5, using eight different reactions (each used once) with the biphosphate of fructose to achieve the non-oxidative glycolytic pathway.

This pathway uses five enzymes: phosphoketolase, aldose-ketose isomerase, transaldolase, aldolase, and phosphohydrolase, as detailed in Table 10. The thermodynamic free energy profile for this pathway is plotted in Figure 6. Only one reaction, the bimolecular aldol addition reaction involving two carbohydrate molecules catalyzed by a transaldolase enzyme, is thermodynamically unfavorable

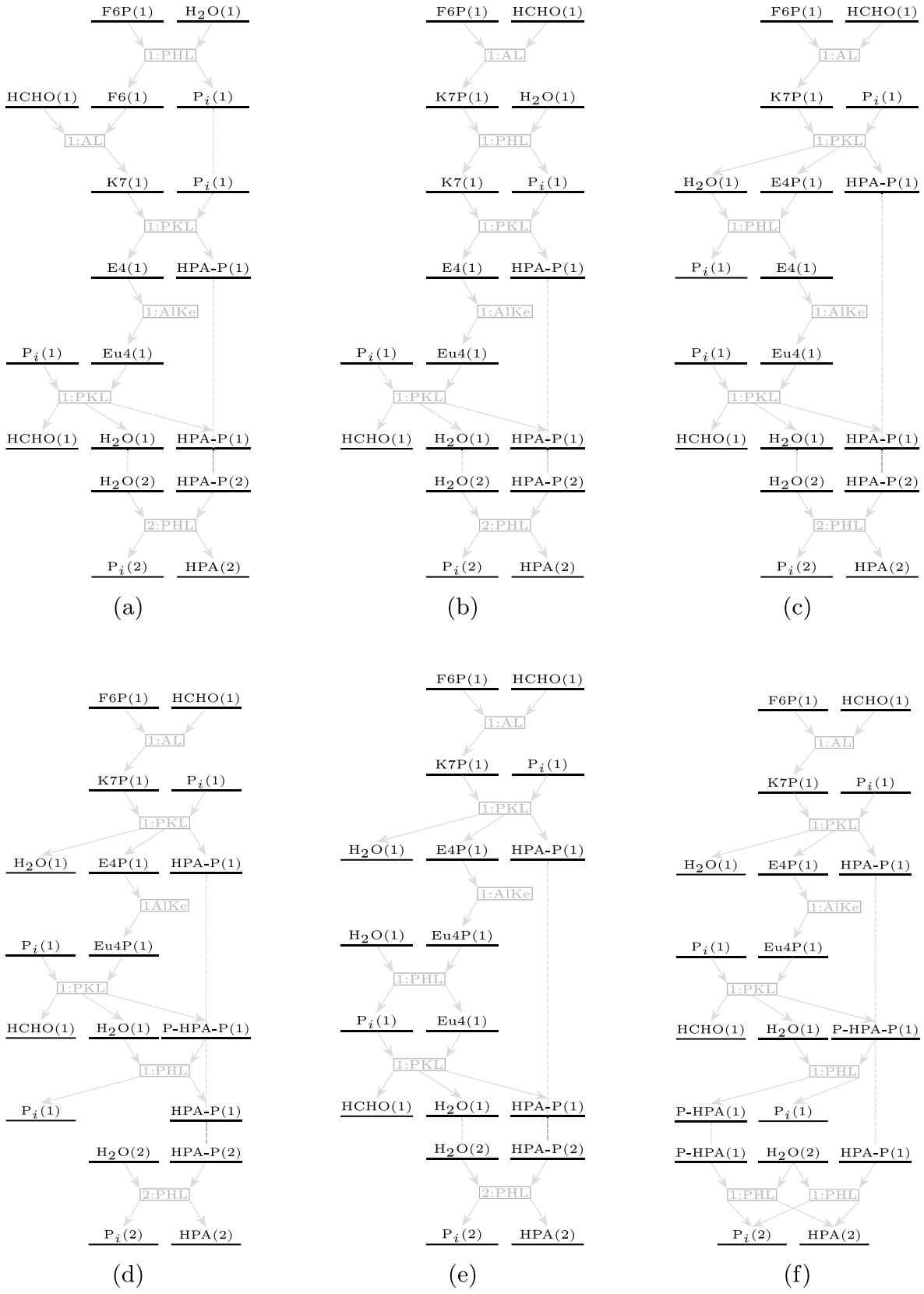


Fig. 7 The six pathways leading to 3-hydroxypropionic acid implementing the net reaction 10. While the first five pathways 7a-7e use six unique reactions (the last reaction is used twice), the last pathway 7f uses seven distinct reactions

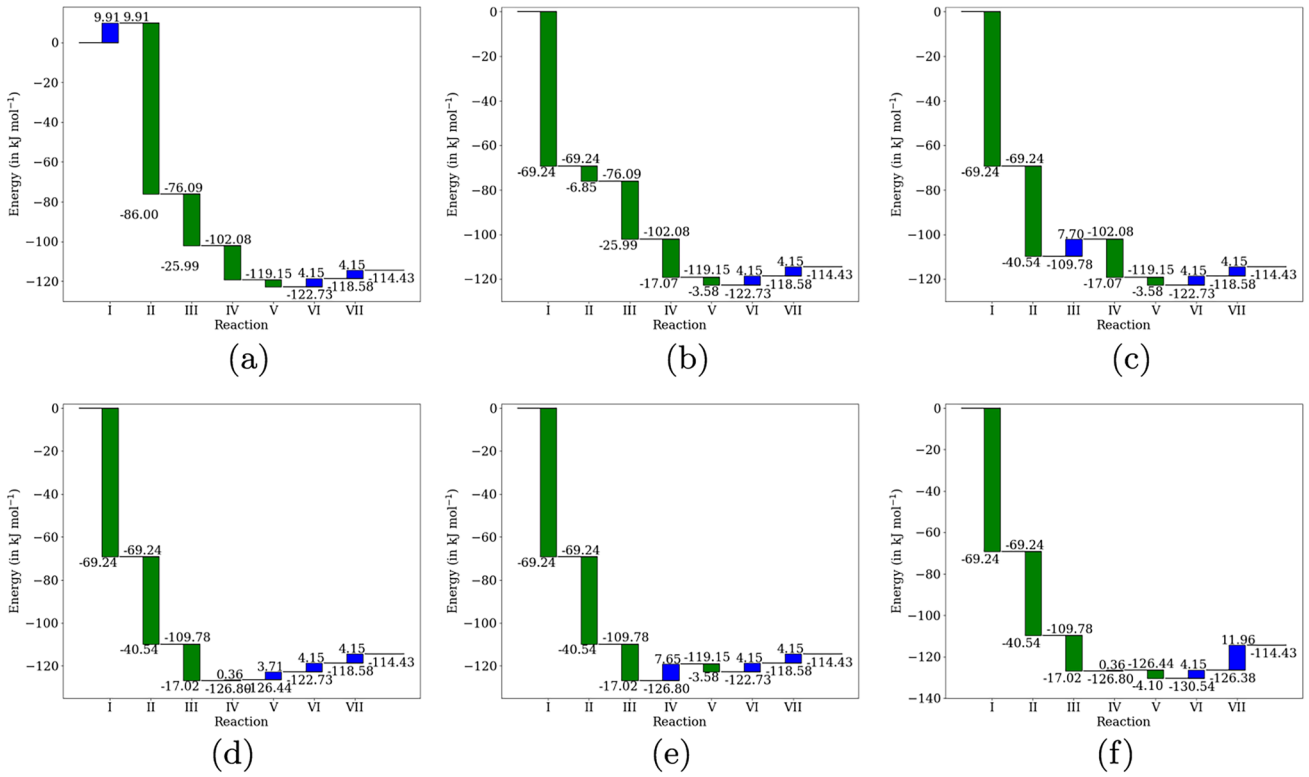


Fig. 8 Thermodynamic profiles for the six corresponding pathways listed in Figure 7 leading to 3-hydroxypropionic acid

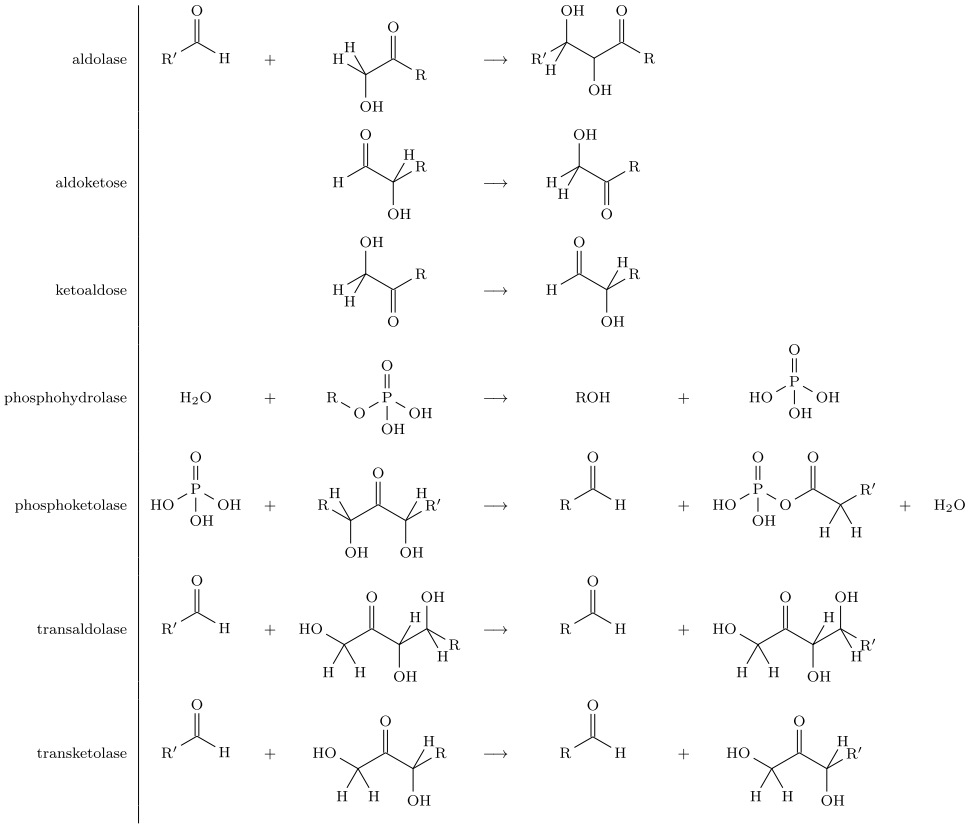
($\Delta_r G = +4.27$ kJ mol^{-1}). The experimentally reported pathway with eight reactions using the same biphosphate also has a thermodynamically unfavorable reaction, albeit one with a smaller positive Gibbs free energy difference ($\Delta_r G = +2.72$ kJ mol^{-1}). Based on this thermodynamic analysis, one might conjecture that both these reactions might proceed in parallel in a bioreactor if a phosphoketolase acting on erythrulose is added to the system.

Formulating the search problem as an ILP query allows us to exhaustively search for non-oxidative glycolytic pathways in the expanded molecular network with the queried length. The results from this search are summarized in Table 6. The length of a pathway is the number of reactions in the pathway. If a reaction is used twice in the pathway, it is counted twice. There are only two pathways of length seven: both of them are described in Section 4.1. There are twelve pathways that use eight reactions in total, but seven distinct reactions, and none of them use fructose-1,6-biphosphate. The shortest pathways utilizing fructose-1,6-biphosphate are of length eight, using all different reactions. They are described in Sections 4.2 and 4.3. As one increases the allowed length of the pathway, the number of solutions returned by the ILP solver increases, as indicated in the last three rows of Table 6. The complete list of the pathways enumerated in the table can be found in the associated GitHub repository.

Other pathways

Expanding the molecular space using the eight rules led to 81 molecules in the reaction network, of which 39 were phosphates of carbohydrates, 14 were biphosphates of carbohydrates, 23 were carbohydrates, and 5 were other small molecules. These other small molecules included phosphoric acid, water, ethanoic acid (isomeric to glycolaldehyde), methanal, and 3-hydroxypropionic acid (isomeric to glyceraldehyde). 3-Hydroxypropionic acid was listed as a "proof of concept" product of biosynthesis. It is identified as a precursor to prop-2-enoic acid (acrylic acid) [19] and the biodegradable polymer poly(3-hydroxypropionic acid), or P(3-HPA) [20, 21]. Moreover, acrylic acid can be used to synthesize acrylates, which are monomers for versatile polymers: poly(methyl methacrylate) (PMMA-lightweight and shatter-resistant alternative to glass) and poly(ethyl cyanoacrylate) (PECA-super glues), to name a few. Both ethanoic acid (acetic acid) and methanal (formaldehyde) have numerous industrial applications apart from being precursors to polymeric units. Therefore, it might be worthwhile to explore pathways leading to these molecules in the network. Pathways producing more than one formaldehyde molecule could not be enumerated in this network. Given that there are multiple alternative economical reaction sequences leading to the production of formaldehyde [22], producing it via this

Table 11 List of the graph transformation rules used as templates for the enzymatic reactions



Reaction conditions: 2-aminophenol and 2-aminothiol (1.2 mmol), aldehyde (1 mmol), K2CO3(1.5 mmol), PEG (3 ml). Isolated yield

biosynthetic method at such a low carbon efficiency might not be justified.

Pathways to 3-hydroxypropioic acid

Since 3-hydroxypropionic acid is a molecule containing three carbon atoms, the optimal efficiency would be achieved when two molecules are produced from one molecule of fructose-6-phosphate, leading to the overall reaction:

$$\mathrm{F6P} + \mathrm{H_2O} \longrightarrow 2\mathrm{HPA} + \mathrm{P}_i \qquad \Delta_r G = -114.428 \text{ kJ mol}^{-1} \tag{10}$$

In the generated reaction network, *six* such pathways were enumerated that carried out the overall reaction 10 in seven steps. Five of these pathways used six unique reactions (one reaction was used twice), while the last pathway used seven distinct reactions, as shown in Figure 7. All the pathways require four of the seven enzymes used to generate the network: aldolase, aldose-ketose isomerase, phosphohydrolase, and phosphoketolase.

The pathway in Fig. 8buses only two reactions with a positive $\Delta_r G$ (it is, in fact, the same reaction implemented twice), while all other pathways have more thermodynamically unfavorable reactions. This might be processed separately, where the phosphate of 3-hydroxypropionic acid is collected as a product and the phosphate group removed separately. However, in a reactor where non-oxidative glycolysis is allowed to execute in parallel, the net reaction 10 would be overshadowed by it, because it has a much lower change in free energy.

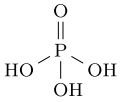
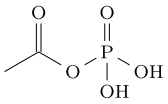
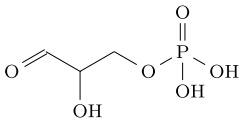
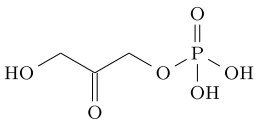
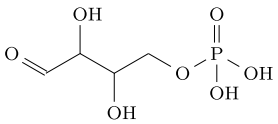
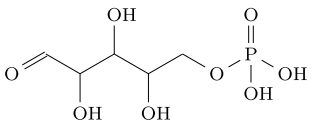
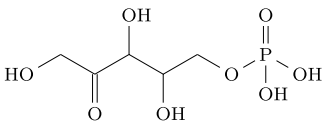
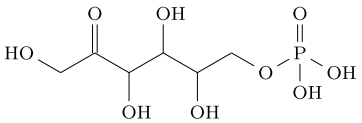
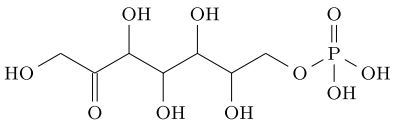
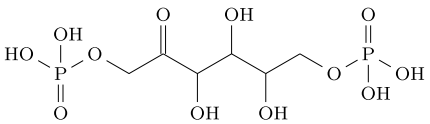
Pathways to acetic acid

Preserving the carbon efficiency would lead to the following overall reaction, with three molecules of acetic acid as the product for every molecule of fructose-3-phosphate:

$$\mathrm{F6P} + \mathrm{H_2O} \longrightarrow 3\mathrm{AcH} + \mathrm{P}_i \qquad \Delta_r G = -286.711 \text{ kJ mol}^{-1} \tag{11}$$

Sixty-one pathways were enumerated effecting the overall reaction 11 in ten steps. Three of these pathways used six different reactions (one reaction was used thrice, and a second reaction was used twice), six pathways used seven different reactions, while the remaining 52 used eight distinct

Table 12 List of the starting molecules used to expand the reaction network

metabolite name	vertex label	molecular structure
water	H_2O	H_2O
phosphoric acid	P_i	
acetylphosphate	AcP	
glyceraldehyde 3-phosphate	G3P	
dihydroxyacetone 3-phosphate	DHAP	
erythrose 4-phosphate	E4P	
ribose 5-phosphate	R5P	
xylulose 5-phosphate	X5P	
fructose 6-phosphate	F6P	
sedoheptulose 7-phosphate	S7P	
fructose 1,6-biphosphate	FBP	

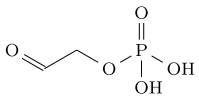
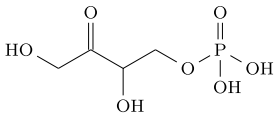
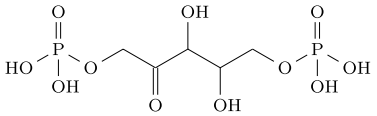
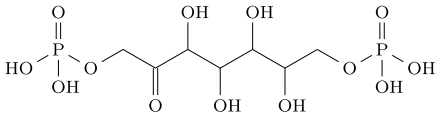
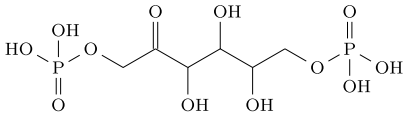
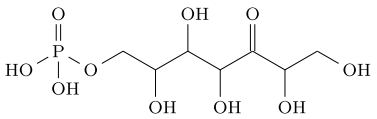
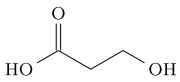
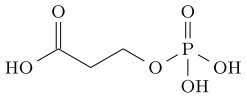
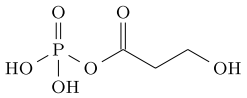
Reaction conditions: 2-aminophenol and 2-aminothiol (1.2 mmol), aldehyde (1 mmol), K2CO3(1.5 mmol), PEG (3 ml). Isolated yield

reactions. Additionally, acetylphosphate on hydrolysis also yields acetic acid. Therefore, in theory, any pathway producing acetylphosphate can also be used to produce acetic acid. Moreover, formaldehyde was also obtained in the expanded molecular space. However, there might be better and more efficient ways to produce formaldehyde than one using enzymes. Hence, such pathways are not being discussed here.

Conclusion

In this work, we attempted to generate the Gibbs free energy profiles for pathways implementing non-oxidative glycolysis in the expanded molecular space. The expansion of the molecular space using reaction templates for the action of each enzyme, along with the ILP-based search for pathways, allows us to explore the molecular space without bias. While

Table 13 List of some molecules encountered in pathways from the reaction network

metabolite name	vertex label	molecular structure
glycolaldehyde 2-phosphate	Gly2P	
erythrulose 4-phosphate	Eu4P	
xylulose 1,5-biphosphate	XBP	
sedoheptulose 1,7-phosphate	SBP	
fructose 1,6-biphosphate	FBP	
phosphate of a ketoheptose	K7P	
3-hydroxypropionic acid	HPA	
3-(phosphonooxy)propanoic acid	HPA-P	
3-hydroxypropionic phospahate	P-HPA	

Reaction conditions: 2-aminophenol and 2-aminothiol (1.2 mmol), aldehyde (1 mmol), K2CO3(1.5 mmol), PEG (3 ml). Isolated yield

the use of Gibbs free energy at standard physical conditions might not always mirror what happens inside a cell, it can assist in elucidating which pathways might be favored by thermodynamics and identifying the bottlenecks in each pathway.

A recurring reaction in the pathways with a positive change in Gibbs free energy was the hydrolysis of a phosphate group from a molecule, whether it involved a carbohydrate biphosphate producing a monophosphate or a carbohydrate monophosphate producing a carbohydrate molecule. Since this reaction involves a water molecule, performing it in an aqueous medium might shift the reaction energetics toward the dephosphorylated products and phosphoric acid.

Another interesting result from the search query for pathways implementing non-oxidative glycolysis was the involvement of phosphoketolase acting on the phosphate of an erythrose phosphate molecule. If one could develop a variant of the phosphoketolase enzyme that is active on this four-carbon ketose, it might be possible to achieve the pathways returned by the ILP solver, provided the required concentration gradients are maintained in the system. Expanding the molecular space with generic rules (irrespective of whether the resulting reactions have been previously reported) provides insights into numerous possibilities for pathways implementing non-oxidative glycolysis-a few of which have been studied in greater detail in this work.

While the existing work focuses on pathways ensuring optimal carbon efficiency during non-oxidative glycolysis, there is a tendency to prioritize obtaining acetyl phosphate as the product because of its metabolic significance. However, other molecules in the generated network might also be of interest, one of them being 3-hydroxypropionic acid. Under laboratory conditions, the products from non-oxidative glycolysis are observed because they have lower Gibbs free energy. The

larger Gibbs free energy difference makes the product, acetyl phosphate, $e^{\frac{-236.75}{-114.28}} \sim 8$ times more likely to be observed than 3-hydroxypropionic acid. However, pathways to 3-hydroxypropionic acid require only four of the seven enzymes needed for the originally proposed non-oxidative glycolysis. While non-oxidative glycolysis might still be achieved using six of these seven enzymes, the optimal carbon efficiency cannot be achieved with only the four enzymes needed to produce 3-hydroxypropionic acid. Therefore, in a system with only four of these enzymes, the pathway producing 3-hydroxypropionic acid might be expected to dominate over some subset of non-oxidative glycolysis.

The method of analyzing pathways satisfying the constraints of mass balance through thermodynamic free energies provides insights into pathways that might be promising under practical conditions. It identifies bottlenecks in the pathways that might require manipulation of the concentrations of the molecules involved in those reactions to make them favorable. Lastly, it highlights potentially useful byproducts in the expanded network and offers directions for further laboratory exploration to synthesize the identified compounds. The results discussed in this work might be found in greater detail at [23].

Some interesting molecules and the graph transformation rules used

Appendix Table 12 lists the starting molecules used to expand the molecular space using the graph transformation rules depicted in Appendix Table 11. These rules represent templates for the reactions effected by the corresponding enzyme. In the rules *R* represents a generic group satisfying the valencies for the formulae that can match any subgraph in the representation of the molecule as a graph. Appendix Table 13 lists some of the interesting molecules which have been used in the pathways studied in this work.

Of the 81 molecules in the expanded network, 6 were small molecules, 22 were carbohydrates, 39 were monophosphates and 14 were biphosphates. There were more monophosphates than carbohydrates because a phosphate group might attach to different hydroxyl positions in the carbohydrate molecule leading to different monophosphate molecules.

Acknowledgements The author would like express their gratitude to the Institut for Matematik og Datalogi, Syddansk Universitet for providing the necessary computational resources for this work.

Funding Open access funding provided by University of Southern Denmark.

References

1. Bogorad IW, Lin T-S, Liao JC (2013) Synthetic non-oxidative glycolysis enables complete carbon conservation. Nature 502:693–697. https://doi.org/10.1038/nature12575
2. Hellgren J, Godina A, Nielsen J, Siewers V (2020) Promiscuous phosphoketolase and metabolic rewiring enables novel non-oxidative glycolysis in yeast for high-yield production of acetyl-coa derived products. Metabolic Eng 62:150–160. https://doi.org/10.1016/j.ymben.2020.09.003
3. Krüsemann JL, Lindner SN, Dempfle M, Widmer J, Arrivault S, Debacker M, He H, Kubis A, Chayot R, Anissimova M, Marlière P, Cotton CAR, Bar-Even A (2018) Artificial pathway emergence in central metabolism from three recursive phosphoketolase reactions. FEBS J 285(23):4367–4377. https://doi.org/10.1111/febs.14682
4. Guo L, Liu M, Bi Y, Qi Q, Xian M, Zhao G (2023) Using a synthetic machinery to improve carbon yield with acetylphosphate as the core. Nature Commun. https://doi.org/10.1038/s41467-023-41135-7
5. Liebermeister W, Klipp E (2006) Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. Theor Biol Med Model. https://doi.org/10.1186/1742-4682-3-41
6. Liebermeister W, Klipp E (2006) Bringing metabolic networks to life: integration of kinetic, metabolic, and proteomic data. Theor Biol Med Model. https://doi.org/10.1186/1742-4682-3-42
7. Steuer R, Gross T, Selbig J, Blasius B (2006) Structural kinetic modeling of metabolic networks. Proc Nat Acad Sci 103(32):11868–11873. https://doi.org/10.1073/pnas.0600013103
8. Tran LM, Rizk ML, Liao JC (2008) Ensemble modeling of metabolic networks. Biophys J 95(12):5606–5617. https://doi.org/10.1529/biophysj.108.135442
9. Andersen JL, Flamm C, Merkle D, Stadler PF (2018) Rule composition in graph transformation models of chemical reactions. MATCH, Commun Math Comput Chem 80(3):661–704
10. Andersen JL, Flamm C, Merkle D, Stadler PF (2013) Inferring chemical reaction patterns using rule composition in graph grammars. J Syst Chem 4(4):661–704. https://doi.org/10.1186/1759-2208-4-4
11. Andersen JL, Flamm C, Merkle D, Stadler PF (2016) In: Echahed, R., Minas, M., (eds.) A Software Package for Chemically Inspired Graph Transformation, pp. 73–88. Springer, Cham. https://doi.org/10.1007/978-3-319-40530-8_5
12. Forrest J, Ralphs T, Vigerske S, Santos HG, Forrest J, Hafer L, Kristjansson B, jpfasano EdwinStraver Lubin M, Jan-Willem rlougee jpgoncal1 Brito S, h-i-gassmann Cristina Saltzman M, tosttost Pitrus B, MATSUSHIMA F (2023) to-st: coin-or/Cbc: Release releases/2.10.11. Zenodo.https://doi.org/10.5281/zenodo.10041724
13. Yoshikawa N, Hutchison GR (2019) Fast, efficient fragment-based coordinate generation for open babel. J Cheminform. https://doi.org/10.1186/s13321-019-0372-5
14. Rappe AK, Casewit CJ, Colwell KS, Goddard WAI, Skiff WM (1992) Uff, a full periodic table force field for molecular

mechanics and molecular dynamics simulations. J Am Chem Soc 114(25):10024–10035. https://doi.org/10.1021/ja00051a040
15. Bannwarth C, Caldeweyher E, Ehlert S, Hansen A, Pracht P, Seibert J, Spicher S, Grimme S (2021) Extended tight-binding quantum chemistry methods. WIREs Comput Mol Sci 11(2):1493. https://doi.org/10.1002/wcms.1493
16. Pal A, Fagerberg R, Andersen JL, Flamm C, Dittrich P, Merkle D (2024) Finding Thermodynamically Favorable Pathways in Reaction Networks using Flows in Hypergraphs and Mixed Integer Linear Programming. https://arxiv.org/abs/2411.15900
17. Pal A, Fagerberg R, Andersen JL, Dittrich P, Merkle D (2025) Finding Pathways in Reaction Networks guided by Energy Barriers using Integer Linear Programming. https://arxiv.org/abs/2504.10609
18. Meadows A, Hawkins K, Tsegaye Y et al (2016) Rewriting yeast central carbon metabolism for industrial isoprenoid production. Nature 537:694–697. https://doi.org/10.1038/nature19769
19. Bozell JJ, Petersen GR (2010) Technology development for the production of biobased products from biorefinery carbohydrates-the US Department of Energy's Top 10 revisited. Green Chem 12:539–554. https://doi.org/10.1039/B922014C
20. Andreeßen B, Taylor N, Steinbüchel A (2014) Poly(3-hydroxypropionate): a promising alternative to fossil fuel-based materials. Appl Environ Microbiol 80(21):6574–6582. https://doi.org/10.1128/AEM.02361-14
21. Vollenweider S, Lacroix C (2004) 3-hydroxypropionaldehyde: applications and perspectives of biotechnological production. Appl Microbiol Biotechnol 64:16–27. https://doi.org/10.1007/s00253-003-1497-y
22. Kamps JJAG, Hopkinson RJ, Schofield CJ, Claridge TDW (2019) How formaldehyde reacts with amino acids. Commun Chem. https://doi.org/10.1038/s42004-019-0224-2
23. Pal A (2024) supplementary data repository on GitHub. https://github.com/AdittyaPal/PathwaysInNOG.git