



OPEN

A mathematically rigorous algorithm to define, compute and assess relevance of the probable dissociation constants in characterizing a biochemical network

Siddhartha Kundu

Metabolism results from enzymatic- and non-enzymatic interactions of several molecules, is easily parameterized with the dissociation constant and occurs via biochemical networks. The dissociation constant is an empirically determined parameter and cannot be used directly to investigate *in silico* models of biochemical networks. Here, we develop and present an algorithm to define, compute and assess the relevance of the probable dissociation constant for every reaction of a biochemical network. The reactants and reactions of this network are modelled by a stoichiometry number matrix. The algorithm computes the null space and then serially generates subspaces by combinatorially summing the spanning vectors that are non-trivial and unique. This is done until the terms of each row either monotonically diverge or form an alternating sequence whose terms can be partitioned into subsets with almost the same number of oppositely signed terms. For a selected null space-generated subspace the algorithm utilizes several statistical and mathematical descriptors to select and bin terms from each row into distinct outcome-specific subsets. The terms of each subset are summed, mapped to the real-valued open interval $(0, \infty)$ and used to populate a reaction-specific outcome vector. The p_1 -norm for this vector is then the probable dissociation constant for this reaction. These steps are continued until every reaction of a modelled network is unambiguously annotated. The assertions presented are complemented by computational studies of a biochemical network for aerobic glycolysis. The fundamental premise of this work is that every row of a null space-generated subspace is a valid reaction and can therefore, be modelled as a reaction-specific sequence vector with a dimension that corresponds to the cardinality of the subspace after excluding all trivial- and redundant-vectors. A major finding of this study is that the row-wise sum or the sum of the terms contained in each reaction-specific sequence vector is mapped unambiguously to a positive real number. This means that the probable dissociation constants, for all reactions, can be directly computed from the stoichiometry number matrix and are suitable indicators of outcome for every reaction of the modelled biochemical network. Additionally, we find that the unambiguous annotation for a biochemical network will require a minimum number of iterations and will determine computational complexity.

The analysis of genome- or large-scale biochemical networks is an important consideration of the “omics”- revolution and a critical step in bridging the genotype and phenotype divide. Despite the availability and accessibility of advanced data analytical tools, true mechanistic insights into the manner in which a biochemical network accomplishes complex function has remained elusive. This is partly due to the absence of a consensus on the choice of parameters needed to characterize a biochemical network. Biochemical networks are a complex interplay of nodes (genes, proteins, metabolites) and stimuli that occur in the cytosol and/or its compartments and

Department of Biochemistry, All India Institute of Medical Sciences, Ansari Nagar, New Delhi 110029, India. email: siddhartha_kundu@yahoo.co.in; siddhartha_kundu@aiims.edu

are regulated by the laws of mass action, diffusion, feedback, partial reactions and thermodynamic parameters^{1,2}. The numerical characterization of a biochemical network is based on simple assumptions in an effort to ensure rigor, maintain computational efficiency and minimize storage. Although informative, insightful and robust, the results that emerge tend to depart significantly from the intracellular milieu and are likely to be of limited biomedical relevance.

The algorithms currently deployed to investigate large biochemical networks include one or more optimization(s) and the enumeration of elementary modes. The former includes flux balance analysis (FBA), dynamic FBA, flux-variable analysis (FVA), regulatory on–off minimization (ROOM) and minimization of metabolic adjustment (MOMA)^{1,3,4}. These algorithms maximize or minimize the biomass of a metabolite of interest and can be used to investigate the effects of deletions and other perturbations on the flux of metabolites through a large network^{1–5}. The algorithms are computationally efficient and can be implemented with ease on an average desktop computer. However, deleting a node as in “gene deletion”-based analyses will alter the network irrevocably, a scenario which is unlikely to occur within the cell. Additionally, user-defined constraints for one or more parameters will introduce a bias into the computations and thence into the set of solutions further mitigating the impact of any analyses on true biochemical and physiological function. The enumeration of elementary flux modes (EFMs), elementary flux vectors (EFVs) and extreme pathway analysis (ExPas) is an alternative strategy to derive meaningful information from biochemical networks^{2,6–9}. This set of algorithms subsumes that a single stoichiometry matrix will lead to one of several states for the “entire”-network of reactions. In these studies a network-state is modelled as a mode (norm, sign) or vector (sign) of “all” rather than individual reactions. Another fundamental assumption is that the solutions of the linear system of equations are strictly positive⁶. Here, too, since each reaction may have several outcomes, this constraint cannot be assumed de facto. The summations and subsequent selection are NP-hard and will result in voluminous run-time data². Despite these limitations the algorithms can identify metabolic hubs and smaller subsets of cooperating reactions^{1,2,6}.

An alternate class of algorithms is concerned with ascribing functionality to each node of a modelled biochemical network by assigning weights to all possible outcomes. The resulting causal networks are probabilistic in construction and the conclusions drawn are inferential¹⁰. The simple architecture notwithstanding, causal networks are able to explain the mechanism of a pharmacophore in the genesis of an adverse drug reaction or reciprocally, its efficacy along with insights into the traits that may be responsible for the same and are powerful aids in precision medicine and data-driven patient management regimes^{11–16}. Operations research, where decisions need to be made on the assignment, allocation and distribution of resources, often, in response to an event is another domain which has benefitted from causal networks and the evaluation of several “what if”-scenarios¹⁷. Biologically relevant data mining for lists of potential genes, proteins and small molecules is yet another clinically relevant application where causal networks have proven to be more than useful^{18–21}. The most important limitation is however, presence of- and access to- previous and existing information. This will dictate the weights for all possible outcomes and determine the overall directionality of the biochemical network. The availability of genome-scale data has permitted investigators to analyse biochemical networks for several organisms, from single cells and with varying time points^{22–24}.

The aforementioned studies (optimization, enumeration, causal networks) are exploratory, discovery-based and focussed on inferring parameters from the data itself, which is in most cases empirical. However, mixed-case scenarios where investigators combine modelling with laboratory data have gained considerable traction in recent years^{24–26}. For example, there is an increased usage of laboratory assays that are based on computationally-derived structure-based molecular descriptors^{27–29}. A generic strategy is to identify the core architecture of a modelled biochemical network and then generate several “synthetic”- or “redundant”-representatives which are progressively trimmed in an effort to reconstruct the originally identified structure³⁰. The parameters that are inferred by this process are then numerically- and/or statistically-validated. Although the utility of inferring network-specific parameters (clustering coefficient, path distance) and data-driven (clinical, epidemiological, pharmacophore) modelling is not debatable, portability/translatibility and widespread usage remain unresolved issues^{25–29}. The inferred parameters, too, will be of limited utility in comprehending the underlying molecular biology and ascribing function to the modelled biochemical network. Selecting a parameter that is simultaneously biochemically and biophysically relevant, from raw datasets, is not trivial and remains a challenge. Certain parameters are truly empirical (reaction rate, rate constant, order of a reaction, turn over number) and are wholly dependent on the experimental design and setup. The inferences and thence the interpretation of these will be of limited clinical and biomedical relevance^{22–26}. Hybrid descriptors such as structurally derived factors for small molecules will combine data points (empirical, computational) with complex theoretical frameworks^{27–29}. It is also entirely possible to derive parameters such as dissociation constant, Michaelis–Menten constant and the Gibbs free energy change directly from theory^{30–34}.

Here, we develop and present an algorithm to compute probable dissociation constants for every reaction of a biochemical network. The algorithm will utilize both, biochemically relevant and theoretically sound constraints to compute these from a stoichiometry number matrix of a biochemical network. Results will be presented, initially for a single reaction and then extended to the whole network. These studies will be complemented with the computation and analysis of a well characterized biochemical network of aerobic glycolysis (AG). The manuscript is organized into an introductory section where some of the principles and definitions that will be used by the algorithm will be defined and explained. A stepwise description of the algorithm along with the necessary mathematical analysis and the supporting computational studies will be presented as part of the “Results and discussion”. The manuscript concludes with a summary of the salient features, limitations and possible future studies which may utilize the probable dissociation constant to characterize a biochemical network. The proofs that underlie this mathematical formalism are described in detail and is presented as part of the Supplementary Material and includes definitions (D), lemmas (L), theorems (T) and corollaries (C) or are stated without a formal proof.

Definitions, parameterization and biomedical relevance of a stoichiometry number matrix-based model for a biochemical network

Metabolism comprises reactions which are enzyme- and non-enzyme (association, disassociation, exchange)-mediated. Whilst enzyme-mediated reactions involve the transformation of one or more substrates into one or more chemically distinct products, non-enzymatic reactions will result in complex formation or disassembly.

Terminology and concepts pertinent to reaction kinetics for use by a stoichiometry number-based model of a biochemical network

The formation of a complex is common to enzyme-mediated catalysis as well as non-catalytic cycles of associations and disassociations. We will utilize the term “reaction” throughout the manuscript to indicate this,

$$\text{Molecular complex} := \left\{ \begin{array}{l} \text{Transformative} \equiv \text{Enzyme mediated} \\ \text{Non - transformative} \end{array} \right\} \equiv \text{“reaction”} \quad \text{Def.(1)}$$

We extend this schema to the encompassing pathway and utilize the phrase “biochemical network” to represent a set of reactants/products that participate in a set of reactions (Def. (2)). We utilize the stoichiometry number for each reactant/product to indicate the presence of a change or lack thereof, that will be brought about by a reaction (Def. (3)). Consider the reaction (r) with j -indexed J -reactants and with rate (R_r). We will define the rate constant (λ_r) as,

$$a_1A_1 + a_2A_2 \dots a_M A_M \xrightarrow{R_r} a_{M+1}A_{M+1} + a_{M+2}A_{M+2} \dots a_{M+N}A_{M+N} \quad \text{Reaction(1)}$$

$$\lambda_r = \frac{R_r}{[A_1]^{l_{A_1}} [A_2]^{l_{A_2}} \dots [A_M]^{l_{A_M}}} \quad (1)$$

where,

$$J = M + N; J, M, N \in \mathbb{N} \quad (2)$$

$$-a_{j \leq M} := \text{Stoichiometry number for reactant } A_{j \leq M} \quad \text{Def.(3a)}$$

$$a_{j > M} := \text{Stoichiometry number for product } A_{j > M} \quad \text{Def.(3b)}$$

$$l_{A_{j \leq M}} := \text{Stoichiometry coefficient for a reactant } (A_j) \quad \text{Def.(3c)}$$

$$\in \mathbb{R}$$

$$\psi_r := \text{Order for a reaction} \quad \text{Def.(3d)}$$

$$= \sum_{j=1}^{j=M} l_{A_{j \leq M}} \in \mathbb{R} \quad (3)$$

The intracellular environment is extremely complex and will support several forms of reactions between the reactants/products whilst precluding others^{28,33–35}. For example, the intracellular milieu or cytosol is regarded as non-Newtonian and imposes constraints on the type of reactions that may occur along with changes that effect reaction rate^{34,35}. These alterations in local viscosity can occur in response to acute stressors such as fluctuations in temperature and hydrogen ion concentration. The modifications have a biochemical (trehalose, glycogen) or biophysical (cytoskeleton) basis and may function to modulate diffusion and diffusion-mediated chemical reactions in vivo or in silico^{34–37}. The physiological manifestations of these will include reduced motion of macromolecules, increased solubility and phagocyte movement towards a chemotactic and noxious stimulus^{36–38}. In order to retain this complexity we will use a reaction vector to model the contribution of each reactant/product. For the set of contributing j -indexed J -reactants or products,

$$\mathbf{r} \sim \mathbf{r} \quad (4)$$

$$\stackrel{\text{def}}{=} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_j \end{bmatrix} \subset \mathbb{Z}^J \quad \text{Def.(4)}$$

$$\text{where } j = 1, 2 \dots J = M + N$$

Definition 5 (D5) The rate constant for a reaction (λ_r) is a strictly positive real scalar,

$$\lambda_r \in \mathbb{R} \cap (0, \infty)$$

Although a reaction may have several outcomes, we will only consider the forward-, reverse- and equivalent-outcomes in this study. Here, we define reaction “equivalence” as the uncertainty with regards to the preferred outcome for a reaction (Def. (6)). This may be due to the paucity of data on the reactants/products and/or the prevailing biophysical state of the intracellular environment where the reaction is expected to occur. In contrast, the reversibility of a reaction (forward, reverse) is a well characterized outcome.

The dissociation constant as the parameter of choice to investigate a biochemical network

Parameterizing a chemical reaction is non-trivial and is done empirically, derived theoretically or inferred from data^{23,29,33,39}. For a non-enzymatic reaction the dissociation (Kd)- and association (Ka)-constants are used to describe reversibility (forward, reverse) at or near equilibrium^{4,5,33,40},

$$Kd \simeq \frac{\lambda_{rf}}{\lambda_{rb}} \quad (5)$$

where,

$$R_{rf} = R_{rb} \quad (6)$$

$$\lambda_{rb} \neq 0 \quad (7)$$

The dissociation constant also allows us to describe a “non-productive” reaction which could occur when two or more reactants/products associate, form a complex and remain associated (Def. (7)). Unlike non-reacting reactants/products, the reaction commences but does not terminate. This can happen when the product of an enzyme forms a covalent linkage with the active site residues and inactivates it irreversibly^{41–43}. A non-productive reaction is also observed when an inhibitor binds preferentially to the bound form of an enzyme, i.e., enzyme–substrate complex, rather than the free enzyme. This unique occurrence is referred to as “uncompetitive”-inhibition and is routinely observed in the tumor protein (TP)53-mediated inhibition of Glucose-6-phosphate dehydrogenase (G6PD; EC1.1.1.49) as a well-supported mechanism for tumor suppression⁴³. For a non-enzymatic reaction this implies an exceptionally high binding affinity between the reactants and/or products with the consequence that these become untenable as independent reactants/products^{41–43}.

The dissociation constant is a non-negative real number ($Kd \in \mathbb{R} \cap (0, \infty)$), can be empirically determined (surface plasma resonance, fluorescence, spectroscopic methods) or theoretically derived, and is a reliable index of reaction outcome^{33,44–47},

$$\text{Forward :} \quad Kd > 1.0 \quad (8)$$

$$\text{Reverse :} \quad Kd \in (0, 1.0) \quad (9)$$

$$\text{Equivalent :} \quad Kd \simeq 1.0 \quad (10)$$

$$\text{Non – productive :} \quad Kd \simeq 0.0 \quad (11)$$

Since the dissociation constant is a ratio, we can use it as a functional metric to compare the rates at which each reaction of a biochemical network takes place as well as the rates of the forward- and reverse-components of a reversible reaction,

Consider the forward reactions:

$$\begin{aligned} & \text{if} \left(\begin{array}{l} Kd_{i-1} > Kd_i > Kd_{i+1} \\ \text{where } Kd_{i-1}, Kd_i, Kd_{i+1} \in (1, \infty) \end{array} \right) \\ & \text{then, } R_{i-1} > R_i > R_{i+1} \end{aligned} \quad (12)$$

Consider the reverse reactions:

$$\begin{aligned} & \text{if} \left(\begin{array}{l} Kd_{i-1} < Kd_i < Kd_{i+1} \\ \text{where } Kd_{i-1}, Kd_i, Kd_{i+1} \in (0, 1) \end{array} \right) \\ & \text{then, } R_{i-1} > R_i > R_{i+1} \end{aligned} \quad (13)$$

Consider the equivalent reactions:

$$\begin{aligned} & \text{if} (Kd_{i-1} \simeq Kd_i \simeq Kd_{i+1} \simeq 1.0) \\ & \text{then, } R_{i-1} \simeq R_i \simeq R_{i+1} \end{aligned} \quad (14)$$

The dissociation constant is versatile, informative and biochemically relevant and is therefore, the parameter of choice to investigate a biochemical network in our study. Here, we will derive a numerical measure directly from the stoichiometry number matrix of a biochemical network which we define as the “probable dissociation constant” (Def. (8)).

Stoichiometry number-based model of a biochemical network

We now model a biochemical network (\mathcal{N}) as the sparse stoichiometry number matrix ($\mathcal{S}_{\mathcal{N}}$) of $J \times \bar{I}$ dimensions ($\mathcal{S}_{\mathcal{N}} \subset \mathbb{Z}^{J \times \bar{I}}$). Briefly, each column of this matrix is an i -indexed ($i = 1, 2, \dots, \bar{I}$) collection of non-trivial reaction vectors ($\mathbf{r}_1 \mathbf{r}_2 \dots \mathbf{r}_{\bar{I}} \subset \mathbb{Z}^J$). Each row of this matrix is contributed to by j -indexed ($j = 1, 2, \dots, J$) m -stoichiometry numbers ($m_{ji} \in \mathbb{Z}$) of J -reactants/products across all \bar{I} ,

$$\mathcal{S} \sim \mathbf{r}_1, \mathbf{r}_2 \dots \mathbf{r}_{i=\bar{i}} \tag{Def.(9)}$$

$$\sim \begin{bmatrix} m_{11} & \dots & m_{1\bar{i}} \\ \vdots & \ddots & \vdots \\ m_{j1} & \dots & m_{j\bar{i}} \end{bmatrix} \tag{Def.(9a)}$$

$$\sim \mathcal{S}(\mathcal{S}) \subset \mathbb{Z}^{j \times \bar{i}} \tag{Def.(9b)}$$

where,

$$\mathbf{r}_i \in \mathcal{R} \subset \mathbb{Z}^j \tag{15}$$

$$\mathbf{r}_i \sim \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_j \end{bmatrix} | m_j \in \mathbb{Z} \tag{16}$$

$$\mathbf{r}_i \neq \vec{0} \tag{17}$$

$$\mathcal{R} := \text{Collection of subsets with non-trivial reaction vectors} \tag{Def.(9c)}$$

The null space-generated subspace can be used to compute the probable dissociation constant for a reaction

For a biochemical network (\mathcal{S}) under the assumption of equilibrium, a subspace of the null space of a stoichiometry number matrix ($\mathcal{V} \subseteq N(\mathcal{S}_\mathcal{S})$) is defined by the null space spanning vectors and those that result from their combinatorial sums (Fig. 1),

$$\{ \mathbf{v}_k \in \mathcal{V} \subset \mathbb{R}^{\bar{i}} | \mathcal{V} \subseteq N(\mathcal{S}_\mathcal{S}); \mathbf{v}_k^T \times \mathcal{S}_\mathcal{S} = \vec{0}^T \} \tag{Def.(10)}$$

$$\mathcal{S}_\mathcal{S} \subset \mathbb{Z}^{j \times \bar{i}} \stackrel{\text{def}}{=} \text{Stoichiometry number matrix for a biochemical network} \tag{18}$$

$$N(\mathcal{S}_\mathcal{S}) \stackrel{\text{def}}{=} \text{Null space of } \mathcal{S}_\mathcal{S} \tag{19}$$

$$\mathcal{V} := \text{Subspace of } N(\mathcal{S}_\mathcal{S}) \text{ with cardinality } \#\mathcal{V} \tag{Def.(11)}$$

$$k = 1, 2, \dots, \#\mathcal{V} \tag{20}$$

Clearly, the cardinality of a null space-generated subspace will increase with the number of cycles of combinatorial summations which we define as u-iterations (Def. (12)). The comprehensive subspace for the uth-iteration (\mathcal{V}_u) is then defined by combinatorially summing the vectors from the (u - 1)th-iteration (\mathcal{V}_{u-1}). Using this notation we can easily see that,

$$\mathcal{V}_{u=0} \sim N(\mathcal{S}_\mathcal{S}) \tag{21}$$

Definition 13 A null space-generated subspace for the uth-iteration is comprehensive and defined as,

$$\mathcal{V}_u \stackrel{\text{def}}{=} \mathcal{V}_{u=0} \cup (\mathcal{V}_{u-1}) \tag{22}$$

$$= N(\mathcal{S}_\mathcal{S}) \cup (\mathcal{V}_{u-1}) \tag{22.1}$$

$$= N(\mathcal{S}_\mathcal{S}) \cup (\mathcal{H}_{u-1} \cup \overline{\mathcal{H}}_{u-1} \cup \mathcal{L}_{u-1}) \tag{22.2}$$

where,

$$N(\mathcal{S}_\mathcal{S}) = \text{Null space spanning vectors for a given stoichiometry number matrix}$$

$$\mathcal{L}_u = \text{Non empty set of trivial vectors} \tag{Def.(13a)}$$

$$\stackrel{\text{def}}{=} \mathbf{v}_{u1} = \mathbf{v}_{u2} \dots = \mathbf{v}_{uK} = \vec{0} | \{ \mathbf{v}_{u1}, \mathbf{v}_{u2}, \dots, \mathbf{v}_{uK} \} \in \mathcal{L}_u \tag{23}$$

$$\text{and } L = \#\mathcal{L}_u \tag{24}$$

$$\mathcal{H}_u = \text{Non empty set of unique vectors} \tag{Def.(13b)}$$

$$\stackrel{\text{def}}{=} \mathbf{v}_{u1} \neq \mathbf{v}_{u2} \dots \neq \mathbf{v}_{uK} | \{ \mathbf{v}_{u1}, \mathbf{v}_{u2}, \dots, \mathbf{v}_{uK} \} \in \mathcal{H}_u \tag{25}$$

$$\text{and } H = \#\mathcal{H}_u \tag{26}$$

$$\overline{\mathcal{H}}_u = \text{Set of identical vectors} \tag{Def.(13c)}$$

$$\text{where } \overline{H} = \#\overline{\mathcal{H}}_u \tag{27}$$

Here,

$$u = 1, 2, \dots, U \in \mathbb{N} \tag{28}$$

In order to partition $\overline{\mathcal{H}}_u$ into subsets we will utilize the following definitions and notation,

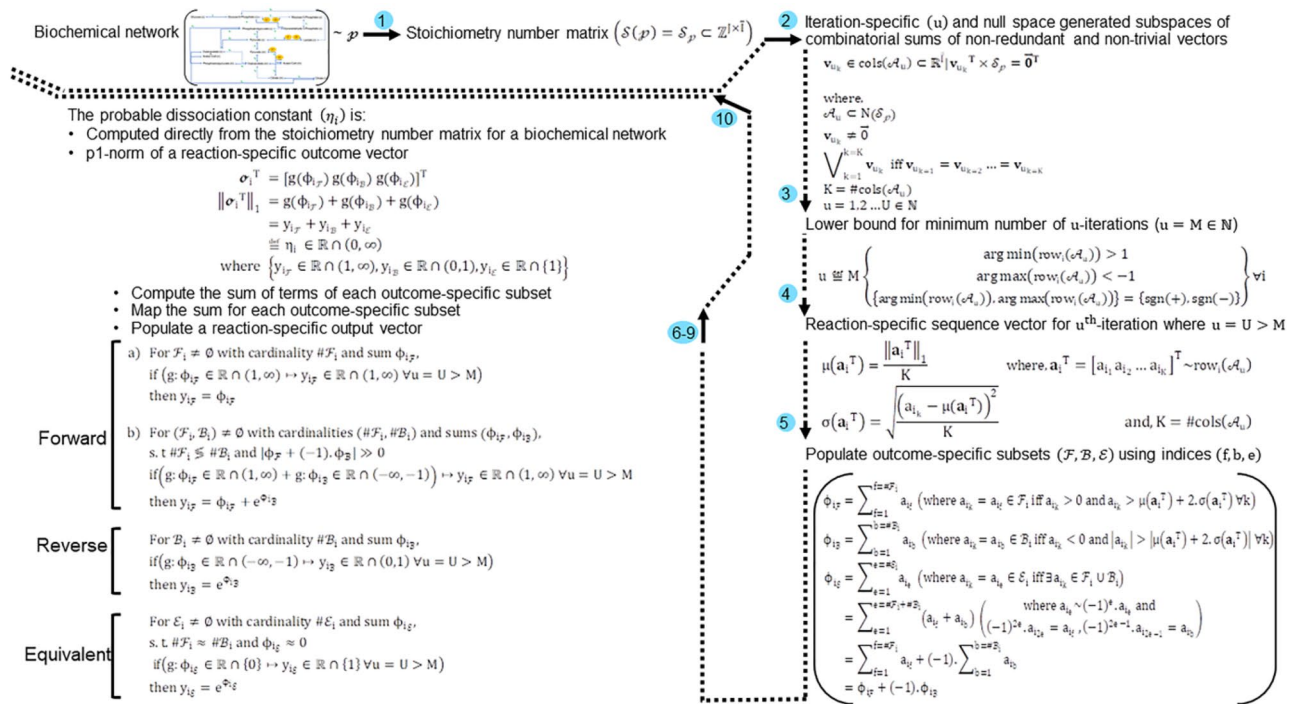


Figure 1. Outline of the algorithm to compute the probable dissociation constant and assign outcome to every reaction of a biochemical network. A biochemical network is modelled in terms of the stoichiometry numbers for a finite number of reactants/products and the accompanying reactions. A null space-generated subspace comprises only unique and non-trivial vectors which are iteratively, recursively and combinatorially summed. The reaction-specific sequence vector that is formed after a finite number of iterations will comprise terms that do not converge and are easily partitioned into distinct subsets. Statistical and mathematical descriptors for these terms (mean, standard deviation, greatest lower bound, least upper bound) are used to select and bin terms for all three predicted outcomes of a reaction. The formed reaction-specific outcome vector has a p1-norm which is the probable dissociation constant for a reaction. This process is repeated until every reaction of the modelled biochemical network can be unambiguously assigned. \bar{I} , Number of reactions for a modelled biochemical network; J , Total number of reactants of modelled biochemical network; S_p , reaction-centric and user-defined stoichiometry number matrix for a biochemical network; A_{u_k} , u -iteration specific subspace of the null space of the stoichiometry number matrix for the modelled biochemical network; $\{\mathcal{F}_i, \mathcal{B}_i, \mathcal{E}_i\}$, Subsets to bin k -th-term of an i -th-reaction-specific sequence vector and ($u = u > M$)th-iteration; $g(\cdot)$, Map to assign the sum of \mathcal{F}_i -, \mathcal{B}_i -, and \mathcal{E}_i -subsets to a strictly positive real number of an i -th-reaction-specific sequence vector and ($u = u > M$)th-iteration; **T**, **C**, **D**; theorem, corollary, definition;

Let $A_{xy} \subset \overline{\overline{\mathcal{H}}_u}$ Def.(14)

where $x = \#\overline{\overline{\mathcal{H}}}$ (29)

and $y = (\alpha, \beta, \gamma)$ (30)

If ($\alpha :=$ Subset with all elements or with a single partition) Def.(14a)

then $(A_{x\alpha} = \overline{\overline{\mathcal{H}}_u})$ (30.1)

If ($\beta :=$ Subset with a single element) Def.(14b)

then $(A_{x\beta} \stackrel{\text{def}}{=} \{.\})$ (30.2)

If ($\gamma :=$ Subset with more than one element) Def.(14c)

then $(A_{x\gamma} \stackrel{\text{def}}{=} \{.,.\}, \{.,.,.\} \dots)$ (30.3)

Lemma 1 (L1) The cardinality of a subset of a finite set of identical vectors ($A_{xy} \subset \overline{\overline{\mathcal{H}}}$) is greater than unity,

$$\#A_{xy} > 1 \quad (31)$$

Proof (L1) By contradiction

Assume $\#A_{xy} = 1$

$$\Rightarrow A_{xy} \subset \mathcal{H} \text{ (unique)} \quad (32)$$

□

Theorem 1 (T1) We can partition a finite set into finite non-overlapping subsets (Θ),

$$\Theta = 1 + \sum_{t=2}^{\#\bar{\mathcal{H}}_u-2} \binom{\#\bar{\mathcal{H}}_u}{t} \quad (33)$$

Interestingly, we can use T1 to compute the lower bounds for a biochemical network in terms of the reactants/products (J) and reaction vectors (\bar{I}). Let us also assume that a single master reaction is an improbable event for a functional biochemical network given the complexity of the intracellular milieu,

$$\begin{aligned} &\text{if } (\Theta \gg 1) \\ &\text{then } \left(\Theta \sim \sum_{t=2}^{t=J-2} \binom{J}{t} \right) \stackrel{\text{def}}{=} \bar{\Theta} \end{aligned} \quad (34)$$

We see that in order to remain biochemically relevant the modelled network must have at least,

$$J \geq 4 \quad (35)$$

The corresponding number of reaction vectors will be,

$$\bar{I} \geq \bar{\Theta} \quad (36)$$

$$\geq 6 \quad (37)$$

Methods

Computational tools to calculate the probable dissociation constant for every reaction of a biochemical network

The biochemical relevance and suitability of the probable dissociation constants for a biochemical network is illustrated by characterizing (constructed, presented, analysed) a biochemical network for AG. The R-package “ReDirection” will be used to assess the contribution of reduced- and oxidized-forms of Nicotinamide adenosine dinucleotide phosphate in regulating the activity of the Pyruvate dehydrogenase complex (PDC)⁴⁸. “ReDirection” is freely available and can be accessed without a login id from the comprehensive R archive network (CRAN; <https://cran.r-project.org/package=ReDirection>)⁴⁸.

Briefly, “ReDirection” will check and modify, if necessary, the stoichiometry number matrix for a biochemical network⁴⁸. The matrix, along with a logical argument (TRUE, FALSE) which indicates whether the reactions are to be regarded as rows or columns, are entered as arguments for the function “*calculate_reaction_vector*”⁴⁸. “ReDirection” will exclude all linear dependent vectors and then recheck the modified matrix for compliance with network-specific indices such as the minimum number of reactants/products, minimum number of reactions and a numerical difference of at least two in favour of the number of reactions⁴⁸. These are necessary arguments and “ReDirection” will not process the stoichiometry number matrix if these are violated. In case there are no deficiencies, “ReDirection” will compute the null space and generate subspaces serially by combinatorially summing all non-redundant and non-trivial vectors⁴⁸. “ReDirection” will continue these steps for a finite number of iterations until a subspace is found where the lower bounds for each sequence of row terms are unambiguous and well-defined. “ReDirection” will then check and bin every selected term to output-specific subsets (forward, reverse, equivalent), compute the linear sum for each subset, map these to strictly positive real numbers and hence populate a reaction-specific output vector⁴⁸. The p1-norm of the latter is denoted as the probable dissociation constant for a reaction⁴⁸. “ReDirection” will repeat this until every reaction is annotated unambiguously⁴⁸. The output will comprise a list of reactions with their computed probable dissociation constants and the predicted outcome (forward, reverse, equivalent)⁴⁸.

Computational studies to highlight and assess biological relevance of the probable dissociation constants

Although the probable dissociation constants that “ReDirection” computes is specific for a biochemical network, there are several generic metrics which can be derived such as the proportion and distribution of matched reactions (equivalent, non-equivalent). These can be used to indicate the flux of reactants/products in a particular direction within a network and compare the modelled biochemical network under variable intracellular conditions.

Aerobic glycolysis occurs when Glucose is converted to Lactate in the presence of non-limiting molecular dioxygen^{49,50}. Although purported initially as a plausible mechanism to explain the atypical metabolism of tumorous cells, the phenomenon is routinely observed in rapidly proliferating non-tumorous (enterocytes, hematopoietic stem cells), quiescent (fibroblasts) and endothelial cells of newly forming vasculature⁵⁰. AG offers plausible explanations for mechanisms for innate immune cell memory, angiogenesis and macrophage polarization^{50–53}. For example, a common observation in pro-inflammatory polarized macrophages is a break in Krebs’ cycle at the level of citrate^{54–57}. However, there is no clarity on the genesis and reversal of this outcome in the presence of an intact and functioning Krebs’ cycle. Conversely, the roles of the bifunctional enzyme phosphofructokinase-2/fructose-2, 6-bisphosphatase 3 (*PFKFBP3*) in the flux of pyruvate prior to its entry into the mitochondria have been implicated in several known AG-pathways^{58,59}. Central to these hypotheses is a role for the PDC and its

regulation. This is a multi-enzyme ($n = 3$; Pyruvate dehydrogenase, EC1.2.4.1; Dihydrolipoyl transacetylase, EC2.3.1.12; Dihydrolipoyl dehydrogenase, EC1.8.1.4) and multi-coenzyme/cofactor ($n = 6$; Magnesium, Thiamine pyrophosphate, lipoic acid, coenzyme-A (CoA), NAD^+ , flavin adenosine dinucleotide (FAD^+)) which decarboxylates mitochondrial Pyruvate into Acetyl-CoA^{60,61}. Regulation of the PDC is complex and occurs via the ratios of $NAD(P)H/NAD(P)$, ATP/ADP and $Acetyl-CoA/CoA$ ^{60,61}.

Results and discussion

The steady-state for a system is described as the absence of apparent change in the numbers of the reactants/products. For a biochemical network this is usually observed when the rates of formation and utilization of the reactants/products that comprise it are assumed to be equal. These conditions, when they occur within the cell are likely to be biochemical in origin and will include the effects of molecular redundancy (alternate splicing, duplicated genes, pseudogenes), feedback (positive, negative), compartmentalization and shared reactants^{22,31}.

Null space generated subspaces with reduced cardinality may improve time to complete annotation

Although the comprehensive null space-generated subspace for a modelled biochemical network is desirable, the vectors that comprise this may be trivial and redundant. This will lead to greater computational complexity and thence delay the time needed to completely annotate every reaction of a biochemical network⁴⁸. It is therefore, imperative that these vectors are identified and excluded. The resulting subspace (\mathcal{A}_u) is of reduced cardinality and is defined as,

$$\left\{ \mathbf{v}_{u_k} \in \mathcal{A}_u \subset \mathcal{V}_u \subset \mathbb{R}^{\bar{I}} \mid \mathcal{V}_u \subseteq N(\mathcal{S}_\rho), \mathbf{v}_{u_k}^T \times \mathcal{S}_\rho = \vec{0}^T \right\} \quad \text{Def. (15)}$$

$$\begin{aligned} \mathcal{S}_\rho &\subset \mathbb{Z}^{I \times \bar{I}} = \text{Stoichiometry number matrix for the biochemical network} \\ N(\mathcal{S}_\rho) &= \text{Null space of } \mathcal{S}_\rho \\ \mathcal{V}_u &= \text{Subspace of } N(\mathcal{S}_\rho) \text{ with cardinality } \#\mathcal{V}_u \\ \mathcal{A}_u &= \text{Null space generated subspace with reduced cardinality} \\ \mathbf{v}_{u_k} \neq \vec{0} &\stackrel{\text{def}}{=} \text{Non-trivial vector of } \mathcal{A}_u & (38) \\ \mathbf{v}_{u_k} &\stackrel{\text{def}}{=} \text{Unique and non-redundant vector of } \mathcal{A}_u & (39) \\ k &= 1, 2 \dots \#\text{cols}(\mathcal{A}_u) & (40) \\ u &= 1, 2 \dots U \end{aligned}$$

The following corollaries about the cardinality of \mathcal{A}_u hold and are easily established.

Corollary 1 (C1) *The expected cardinality of null space generated subspace for the matrix of stoichiometric numbers of the reactants/products of a biochemical network is always greater than the computed cardinality,*

$$\#\mathcal{V}_u > \#\mathcal{A}_u \mid u = 1, 2 \dots U \quad (41)$$

Corollary 2 (C2; without proof) *The cardinality of an arbitrary and recursively generated subspace of the null space of the matrix of stoichiometry numbers of the reactants/products for a biochemical network is always greater than the nullity of the null space,*

$$\#\mathcal{A}_u = \#\mathcal{N}(\mathcal{S}_\rho) + (\#\mathcal{H}_{u-1} + \#\overline{\mathcal{H}}_{u-1} + \#\mathcal{L}_{u-1}) \quad [\text{From T1, C1 and Def. (13)}] \quad (42)$$

$$= \#\mathcal{N}(\mathcal{S}_\rho) + \sum_{q=2}^{\#\mathcal{V}_{u-1}} \binom{\#\mathcal{V}_{u-1}}{q} \quad (42.1)$$

$$\Rightarrow > \#\mathcal{N}(\mathcal{S}_\rho) \quad (43)$$

Corollary 3 (C3; without proof) *The cardinality of recursively generated subspaces of the null space of the matrix of stoichiometric numbers of the reactants/products of a biochemical network forms a monotonically increasing sequence,*

$$(\#\mathcal{A}_u)_{u \geq 1} \mid \#\mathcal{A}_1 < \#\mathcal{A}_2 \dots < \#\mathcal{A}_U \mid u = 1, 2 \dots U \quad (44)$$

Since the time taken to annotate every reaction of a biochemical network will depend on the combinatorial sums of all non-redundant and non-trivial null space generated subspace vectors, the cardinality of this subspace, at each iteration is an important determinant of computational complexity.

Corollary 4 (C4) *The computational complexity per iteration in exponential-time (T_u) is,*

$$T_u \in (0, \mathcal{O}(2^{\#\mathcal{V}_u \cdot \#\mathcal{V}_u}) \mid u = 1, 2 \dots U \quad (45)$$

These results suggest that as the number of null space generated subspace vectors increases there will be an increase in the number of iterations needed to unambiguously annotate a reaction. This in turn will result in an increased cardinality of each null space generated subspace. A corresponding increase in run-time will then be needed to completely annotate every reaction of a biochemical network⁴⁸.

Rationale to utilize the row-sum of a null space-generated subspace as the domain to compute the probable dissociation constant for a reaction

The vectors that comprise each null space-generated subspace represent a biochemical network at equilibrium where each row represents an i th-reaction and each numerical value belongs to $\mathbb{R} \cap (-\infty, \infty)$. However, since the dissociation constant is a non-negative real number we seek a map to the set of strictly positive real numbers such that in the absence of empirical data we can directly compute the probable dissociation constant from the null space of the stoichiometry number matrix for a biochemical network,

$$g : (\cdot) \in \mathbb{R} \cap (-\infty, \infty) \mapsto (\cdot) \in \mathbb{R} \cap (0, \infty) \quad \text{Def.}(16)$$

Let us partition the real part of the numerical values of any non-trivial and non-redundant null space generated subspace vector,

$$v_{ik} \stackrel{\text{def}}{=} v_{ikf} \text{ iff } v_{ik} \in \mathbb{R} \cap (0, \infty) \quad (46)$$

$$v_{ik} \stackrel{\text{def}}{=} v_{ikb} \text{ iff } v_{ik} \in \mathbb{R} \cap (0, -\infty) \quad (47)$$

$$v_{ik} \stackrel{\text{def}}{=} v_{ike} \text{ iff } v_{ik} \approx 0 \quad (48)$$

where,

$$f := \text{Index for a putative forward outcome predicting term } (f \in \mathbb{N}) \quad \text{Def.}(16a)$$

$$b := \text{Index for a putative reverse outcome predicting term } (b \in \mathbb{N}) \quad \text{Def.}(16b)$$

$$e := \text{Index for a putative equivalent outcome predicting term } (e \in \mathbb{N}) \quad \text{Def.}(16c)$$

Combining these,

$$v_{ik} \stackrel{\text{def}}{=} v_{ikf} \vee v_{ikb} \vee v_{ike} \quad (49) \\ \in \mathcal{A} \subset \mathbb{R} \cap (-\infty, \infty) \quad \text{Def.}(17)$$

This means we can only map the following values unambiguously,

$$g(v_{ikf}) \in \mathbb{R} \cap (1, \infty) \quad (50) \\ g(v_{ikb}) \in \mathbb{R} \cap (0, 1) \quad (51)$$

For $v_{ik} \approx 0$ however, this mapping schema for an “equivalent”-reaction will not suffice since,

$$v_{ik} \approx 0 \stackrel{\text{def}}{=} \begin{cases} \text{“Equivalent”}(v_{ike}) \\ \text{“Non-productive”} \end{cases} \quad (52)$$

We can resolve this by treating each row of a null space generated subspace (\mathcal{A}) as a finite series,

$$\phi = \sum_{k=1}^{k=K} a_k \quad (53)$$

where,

$$a_k \in \text{rows}(\mathcal{A}) \subset \mathbb{R}^{1 \times K} \quad \text{Def.}(18)$$

$$K = \#\text{cols}(\mathcal{A}) \quad (54)$$

$$k = 1, 2 \dots K \quad (55)$$

If we do this iteratively (u), recursively and across all columns of a null space-generated subspace ($k = 1, 2 \dots K$) and for all reactions ($\forall i$), we obtain the following expression,

$$\phi \sim \phi_u \sim \phi_{u_i} \quad (56)$$

$$= \sum_{k=1}^{k=K} a_{u_{ik}} \quad (56.1)$$

Theorem 2 (T2) *The numerical values that comprise each row of a null space generated subspace (\mathcal{A}_u) with positive- and negative-terms can be rewritten as an alternating sequence,*

$$\left(a_{u_{ik}} \right)_{k=1,2,\dots,K} \sim \left((-1)^k \cdot a_{u_{ik}} \right)_{k=1,2,\dots,K} \quad (57)$$

where,

$$a_{u_{ik}} \in \text{row}_i(\mathcal{A}_u) \subset \mathbb{R}^{1 \times K} \quad (58)$$

$$K = \#\text{cols}(\mathcal{A}_u) \quad (59)$$

$$k = 1, 2 \dots K \quad (60)$$

$$i = 1, 2 \dots \bar{I} \quad (61)$$

$$u = 1, 2 \dots U \in \mathbb{N} \quad (62)$$

Theorem 3 (T3) *The sum for the finite series formed by each row of a null space generated subspace can be mapped from $\mathbb{R} \cap (-\infty, \infty)$ to $\mathbb{R} \cap (0, \infty)$ for the i th-reaction of the u th-iteration,*

$$g : \phi_{u_i} \in \mathbb{R} \cap (-\infty, \infty) \mapsto y_{u_i} \in \mathbb{R} \cap (0, \infty) \tag{63}$$

where,
 $\phi_{u_i} = \sum_{k=1}^{K} a_{u_{ik}}$
 $a_{u_{ik}} \in \text{row}_i(\mathcal{A}_u) \subset \mathbb{R}^{1 \times K}$
 and,
 $K = \#\text{cols}(\mathcal{A}_u)$
 $k = 1, 2 \dots K$
 $i = 1, 2 \dots \bar{I}$
 $u = 1, 2 \dots U \in \mathbb{N}$

We can now utilize ϕ_{u_i} to unambiguously compute the probable dissociation constant and assign an outcome (forward, reverse, equivalent) to the i^{th} -reaction and thence to every reaction of a biochemical network.

Algorithm to compute the probable dissociation constant for every reaction of a biochemical network

An R-package implementation of the basic ideas that underline the presented algorithm is available for a user-defined biochemical network⁴⁸. However, the mathematical basis for these findings and/or observations have not been addressed. Here, we rigorously investigate some of the several fundamental steps in the computation, usage and utility of the probable dissociation constant for every reaction of a biochemical network.

Schema to screen combinatorially summed vectors for a null space-generated subspace

For each u -iteration the null space-generated subspace will be screened for trivial and redundant vectors and subsequently excluded. The non-redundant and non-trivial vectors that remain will constitute the cardinality for the u -iteration specific null space-generated subspace (Step 1, Fig. 1). The following assumptions will be valid for the presented schema,

$$\left\{ \begin{array}{l} u = 1, 2 \dots U \\ i = 1, 2 \dots \bar{I} \\ \Theta \in \mathbb{N} \\ \#\bar{\mathcal{H}} = \mathbb{K} \end{array} \right\} \tag{64}$$

Rewriting the finite set of identical null space-generated subspace vectors for the u^{th} -iteration ($\bar{\mathcal{H}}_u$) in terms of a finite collection of non-overlapping subsets (Θ),

$$\bar{\mathcal{H}}_u = \{ (\mathbf{v}_{u_1} = \mathbf{v}_{u_2} \dots = \mathbf{v}_{u_A}), (\mathbf{v}_{u_{A+1}} = \mathbf{v}_{u_{A+2}} \dots = \mathbf{v}_{u_B}) \dots (\mathbf{v}_{u_{\mathbb{K}-(\mathbb{K}-1)}} \dots \mathbf{v}_{u_{\mathbb{K}-1}}, \mathbf{v}_{u_{\mathbb{K}}}) \} \tag{65}$$

where $\{ \mathbf{v}_{u_1}, \mathbf{v}_{u_2} \dots, \mathbf{v}_{u_{\mathbb{K}}} \} \in \bar{\mathcal{H}}_u$

A unique representative from each subset is,

$$\left(\bigvee_{k=1}^{k=A} \mathbf{v}_{u_k}, \bigvee_{k=A+1}^{k=B} \mathbf{v}_{u_k} \dots \bigvee_{k=\mathbb{K}-(\mathbb{K}-1)}^{k=\mathbb{K}} \mathbf{v}_{u_k} \right) \tag{66}$$

We will now reassign these vectors to \mathcal{H}_u ,

$$\left(\bigvee_{k=1}^{k=A} \mathbf{v}_{u_k}, \bigvee_{k=A+1}^{k=B} \mathbf{v}_{u_k} \dots \bigvee_{k=\mathbb{K}-(\mathbb{K}-1)}^{k=\mathbb{K}} \mathbf{v}_{u_k} \right) \in \mathcal{H}_u \tag{67}$$

$$= \left(\bigvee_{k=1}^{k=A} \mathbf{v}_{u_k} \in \mathcal{H}_u, \bigvee_{k=A+1}^{k=B} \mathbf{v}_{u_k} \in \mathcal{H}_u \dots \bigvee_{k=\mathbb{K}-(\mathbb{K}-1)}^{k=\mathbb{K}} \mathbf{v}_{u_k} \in \mathcal{H}_u \right) \tag{67.1}$$

and exclude these from $\bar{\mathcal{H}}_u$

$$\left(\bigvee_{k=1}^{k=A} \mathbf{v}_{u_k} \notin \bar{\mathcal{H}}_u, \bigvee_{k=A+1}^{k=B} \mathbf{v}_{u_k} \notin \bar{\mathcal{H}}_u \dots \bigvee_{k=\mathbb{K}-(\mathbb{K}-1)}^{k=\mathbb{K}} \mathbf{v}_{u_k} \notin \bar{\mathcal{H}}_u \right) \tag{68}$$

$$= \left(\bigvee_{k=1}^{k=A} \mathbf{v}_{u_k}, \bigvee_{k=A+1}^{k=B} \mathbf{v}_{u_k} \dots \bigvee_{k=\mathbb{K}-(\mathbb{K}-1)}^{k=\mathbb{K}} \mathbf{v}_{u_k} \right) \notin \bar{\mathcal{H}}_u \tag{68.1}$$

Let us now define a representative vector from each subset as,

$$\begin{array}{l} \zeta_a \sim \bigvee_{k=1}^{k=A} \mathbf{v}_{u_k} \\ \zeta_b \sim \bigvee_{k=A+1}^{k=B} \mathbf{v}_{u_k} \\ \vdots \\ \zeta_{\mathbb{K}} \sim \bigvee_{k=\mathbb{K}-(\mathbb{K}-1)}^{k=\mathbb{K}} \mathbf{v}_{u_k} \end{array} \tag{Def.19}$$

Since there can be one and only one such vector for each subset, we can write,

$$\zeta_1 = \zeta_2 \dots = \zeta_{\Theta} = 1 \tag{69}$$

The finite series formed is then numerically equal to the number of subsets that $\bar{\mathcal{H}}_u$ is partitioned into,

$$\sum_{a=1}^{a=\Theta} \zeta_a = \Theta \tag{70}$$

These vectors will now be assigned to \mathcal{H} which is the subset for all unique null space-generated subspace vectors. The incremented cardinality for the u -th-iteration of this subset ($\overline{\mathcal{H}}_u$) is,

$$\begin{aligned} \#\mathcal{H}_u &= \#\mathcal{H}_u + \sum_{a=1}^{a=\Theta} \zeta_a & (71) \\ &= \#\mathcal{H}_u + \Theta & (71.1) \end{aligned}$$

Reciprocally, $\overline{\mathcal{H}}_u$ will be correspondingly deficient in these vectors and can be written as,

$$\begin{aligned} \overline{\mathcal{H}}_u &= (\mathbf{v}_{u_1} = \mathbf{v}_{u_2} \cdots = \mathbf{v}_{u_{A-1}}), (\mathbf{v}_{u_{A+1}} = \mathbf{v}_{u_{A+2}} \cdots = \mathbf{v}_{u_{B-1}}) \cdots (\mathbf{v}_{u_{\mathbb{K}-(\mathbb{K}-1)}, \dots, \mathbf{v}_{u_{\mathbb{K}-1}}}) & (72) \\ &\text{where, } \{\mathbf{v}_{u_1}, \mathbf{v}_{u_2} \dots, \mathbf{v}_{u_{\mathbb{K}-\Theta}}\} \in \overline{\mathcal{H}}_u \end{aligned}$$

The reduced cardinality for the u^{th} -iteration of $\overline{\mathcal{H}}_u$ is,

$$\#\mathcal{H}_u = A - 1 + B - 1 \dots \mathbb{K} - 1 \tag{73}$$

$$= (A + B \cdots + \mathbb{K}) - (1 + 1 \cdots + 1) \tag{73.1}$$

$$= \#\overline{\mathcal{H}}_u - \sum_{a=1}^{a=\Theta} 1_a \tag{73.2}$$

$$= \#\mathcal{H}_u - \sum_{a=1}^{a=\Theta} \zeta_a \tag{73.3}$$

$$= \#\mathcal{H}_u - \Theta \tag{73.4}$$

Characterizing reaction-specific sequence vectors from the rows of selected null space-generated subspaces

We have already seen that after a finite number of iterations of combinatorial summing and vector exclusion, null space-generated subspaces with reduced cardinalities will result (\mathcal{A}_u). Since the terms from the i -th-row when taken together, i.e., across all columns of the comprising vectors ($\text{cols}(\mathcal{A}_u)$) constitute the i -th-reaction, we observe that $\#\text{cols}(\mathcal{A}_u)$ -terms can be partitioned into distinct subsets. We define $u = M \in \mathbb{N}$ formally as the lower bound for these iterations as (Step 3; Fig. 1).

$$u \stackrel{\text{def}}{=} M \text{ iff } \left\{ \begin{array}{l} \arg \min(\text{row}_i(\mathcal{A}_u)) > 1 \\ \arg \max(\text{row}_i(\mathcal{A}_u)) < -1 \\ \{\arg \min(\text{row}_i(\mathcal{A}_u)), \arg \max(\text{row}_i(\mathcal{A}_u))\} = \{\text{sgn}(+), \text{sgn}(-)\} \end{array} \right\} \forall i \tag{74}$$

where,
 $\text{row}_i(\mathcal{A}_u) \subset \mathbb{R}^{1 \times K}$
 $i = 1, 2 \dots \bar{I}$
 $K = \#\text{cols}(\mathcal{A}_u)$

The numerical values for M are easily computed from the limit of the fractional optimization of two decoupled real-valued functions when evaluated as a single ratio,

$$\underset{x}{\text{minimize}} \left(\frac{A(x)}{B(x)} \right) \geq 0 \tag{75}$$

where

$$A(x) : \mathbb{R}^K \mapsto \mathbb{R}_+ \cap \{0\} \tag{76}$$

$$B(x) : \mathbb{R}^K \mapsto \mathbb{R}_+ \cap \{0\} \tag{77}$$

$$K = \#\text{cols}(\mathcal{A}_u)$$

$$u = 1, 2 \dots U$$

Since the algorithm is based on combinatorial summations, the constituent real-valued terms of a reaction-specific sequence vector will be monotonic and their linear sum will be non-zero. This implies that convergence will occur very quickly (forward, reverse). If the linear sum of the terms of the reaction-specific sequence vector is zero, this will indicate that the sequence has an almost equal number of oppositely signed terms and will not converge (equivalent). However, here too, the use of fractional optimization, decoupling and a single ratio to minimize will result in convergence and the occurrence of a limit,

At $u = M$:

$$\begin{aligned} &\text{if}(\phi_{u_i} \in (0, +\infty)) \\ &\text{then} \left(\lim_{x \rightarrow (+)\infty} \left(1 + \frac{e^{\text{argmin}(\text{row}_i(\mathcal{A}_u))}}{e^{\phi_{u_i}}} \right) = 1 \right) \end{aligned} \tag{78}$$

OR

$$\begin{aligned} &\text{if}(\phi_{u_i} \in (-\infty, 0)) \\ &\text{then} \left(\lim_{x \rightarrow (-)\infty} \left(1 + \frac{e^{\phi_{u_i}}}{e^{\text{argmax}(\text{row}_i(\mathcal{A}_u))}} \right) = 1 \right) \end{aligned} \tag{79}$$

OR

$$\begin{aligned} &\text{if}(\phi_{u_i} \approx 0) \\ &\text{then} \left(\left(\lim_{x \rightarrow (+)\infty} \left(1 + \frac{e^{\text{argmin}(\text{row}_i(\mathcal{A}_u))}}{e^{\phi_{u_i}}} \right) \right) \vee \left(\lim_{x \rightarrow (-)\infty} \left(1 + \frac{e^{\phi_{u_i}}}{e^{\text{argmax}(\text{row}_i(\mathcal{A}_u))}} \right) \right) \right) \\ &\hspace{10em} = (1) \vee (1) \\ &\hspace{10em} = 1 \end{aligned} \tag{80}$$

where

$$\begin{aligned} \phi_{u_i} &= \sum_{k=1}^{k=K} a_{u_i k} \\ a_{u_i k} &\in \text{row}_i(\mathcal{A}_u) \subset \mathbb{R}^{1 \times K} \\ K &= \#\text{cols}(\mathcal{A}_u) \\ k &= 1, 2 \dots K \\ i &= 1, 2 \dots \bar{I} \\ u &= 1, 2 \dots U \in \mathbb{N} \end{aligned}$$

Now that we have established the minimum number of iterations, we will proceed to define, populate and process the reaction-specific sequence vector for every reaction of the modelled biochemical network. We will use a simplified notation hereafter without loss of generality to indicate and annotate,

$$(\cdot)_{u_i} \sim (\cdot)_i \tag{81}$$

$$\begin{aligned} &\text{where,} \\ &u = 1, 2 \dots U > M \\ &i = 1, 2 \dots \bar{I} \end{aligned}$$

We denote an i th-reaction-specific sequence vector sequence (\mathbf{a}_i^T) with real values which for the u th-iteration whose sum is ϕ_i in accordance with T2 (Step 4; Fig. 1),

$$\text{row}_i(\mathcal{A}_u) \sim \mathbf{a}_i^T \tag{82}$$

$$\begin{aligned} &= [a_{i1} \ a_{i2} \ \dots \ a_{iK}]^T \\ \|\mathbf{a}_i^T\|_1 &= \phi_i \end{aligned} \tag{82.1} \tag{83}$$

where,

$$\begin{aligned} a_{i_k} &\in \mathbf{a}_i^T \sim \text{row}_i(\mathcal{A}_u) \subset \mathbb{R}^{1 \times K} \\ i &= 1, 2 \dots \bar{I} \\ K &= \#\text{cols}(\mathcal{A}_u) \\ k &= 1, 2 \dots K \\ u &= 1, 2 \dots U > M \end{aligned}$$

Once the i th-reaction specific sequence vector is defined we will characterize it using standard statistical descriptors (Step 5; Fig. 1),

$$\begin{aligned} \mu(\mathbf{a}_i^T) &:= \text{Arithmetic mean for the } \#\mathcal{A}_u \text{ - terms of } \mathbf{a}_i^T && \text{Def.(20)} \\ &= \frac{\phi_i}{\#\text{cols}(\mathcal{A}_u)} && (84) \\ &= \frac{\|\mathbf{a}_i^T\|_1}{\#\text{cols}(\mathcal{A}_u)} && (84.1) \end{aligned}$$

$$\begin{aligned} \sigma(\mathbf{a}_i^T) &:= \text{Standard deviation for the } \#\mathcal{A}_u \text{ - terms of } \mathbf{a}_i^T && \text{Def.(21)} \\ &= \sqrt{\frac{(\mathbf{a}_{ik} - \mu(\mathbf{a}_i^T))^2}{\#\text{cols}(\mathcal{A}_u)}} && (85) \end{aligned}$$

where,
 $\mathbf{a}_{ik} \in \mathbf{a}_i^T \in \text{row}_i(\mathcal{A}_u) \subset \mathbb{R}^{1 \times K}$
 $K = \#\text{cols}(\mathcal{A}_u)$

and,
 $k = 1, 2 \dots K$
 $u = 1, 2 \dots U > M$
 $i = 1, 2 \dots \bar{I}$

Rule-based population of the outcome-specific subsets for every reaction of a biochemical network

We will first define the criteria to populate the \mathcal{F}_i - and \mathcal{B}_i -subsets from the reaction-specific sequence vector for the i th-reaction of a ($u = U > M$)th-iteration,

$$\left\{ \begin{array}{l} \mathbf{a}_{ik} = \mathbf{a}_{if} \in \mathcal{F}_i \text{ iff } \mathbf{a}_{ik} > \mu(\mathbf{a}_i^T) + 2.\sigma(\mathbf{a}_i^T) \\ \text{and } \mathbf{a}_{ik} > 0 \end{array} \right\} \quad \text{Def.(22)}$$

$$\left\{ \begin{array}{l} \mathbf{a}_{ik} = \mathbf{a}_{ib} \in \mathcal{B}_i \text{ iff } |\mathbf{a}_{ik}| > |\mu(\mathbf{a}_i^T) + 2.\sigma(\mathbf{a}_i^T)| \\ \text{and } \mathbf{a}_{ik} < 0 \end{array} \right\} \quad \text{Def.(23)}$$

where,
 $\mathbf{a}_{ik} \in \mathbf{a}_i^T \in \text{row}_i(\mathcal{A}_u) \subset \mathbb{R}^{1 \times K}$
 $\mu(\mathbf{a}_i^T) \stackrel{\text{def}}{=} \text{Arithmetic mean for the } \#\mathcal{A}_u \text{ - terms of } \mathbf{a}_i^T$
 $\sigma(\mathbf{a}_i^T) \stackrel{\text{def}}{=} \text{Standard deviation for the } \#\mathcal{A}_u \text{ - terms of } \mathbf{a}_i^T$
 $K = \#\text{cols}(\mathcal{A}_u)$

and,
 $k = 1, 2 \dots K$
 $f, g = 1, 2 \dots F, G \in \mathbb{N}$ (86)
 $i = 1, 2 \dots \bar{I}$
 $u = 1, 2 \dots U > M$

Corollary 5 (C5) *If the cardinality of the reaction-specific \mathcal{F}_i - and \mathcal{B}_i -subsets is finite we can identify a unique subset (\mathcal{E}_i) with terms that form an alternating sequence after $u = U > M$ -iterations,*

$$\mathcal{E}_i = (\mathbf{a}_{ie})_{e=1,2,\dots,\#\mathcal{E}_i} \quad (87)$$

$$= ((-1)^e . \mathbf{a}_{ie})_{e=1,2,\dots,\#\mathcal{E}_i} \quad (87.1)$$

$$= ((-1)^{2e} . \mathbf{a}_{i2e} + (-1)^{2e-1} . \mathbf{a}_{i2e-1})_{e=1,2,\dots,\#\mathcal{F}_i+\#\mathcal{B}_i} \quad (87.2)$$

where,

$$(-1)^{2e} . \mathbf{a}_{i2e} = \mathbf{a}_{if} > 0 \quad (87.3)$$

$$(-1)^{2e-1} . \mathbf{a}_{i2e-1} = \mathbf{a}_{ib} < 0 \quad (87.4)$$

$$\{\#\mathcal{F}_i, \#\mathcal{B}_i\} > 0 \quad (87.5)$$

and,
 $e, f, g = 1, 2 \dots E, F, G \in \mathbb{N}$ (87.6)

$i = 1, 2 \dots \bar{I}$
 $u = 1, 2 \dots U > M$

Corollary 6 (C6, without proof) *The sum of the terms of the subset that corresponds to the “equivalent”-outcome is numerically equal to zero after a large $u = U > M$,*

$$\begin{aligned}\phi_{i_{\mathcal{E}}} &= \phi_{i_{\mathcal{F}}} + (-1) \cdot \phi_{i_{\mathcal{B}}} & (88) \\ &= \phi_e + (-1) \cdot \phi_e [\text{From T2, T3, C5}] & (88.1) \\ &\approx 0 & (88.2)\end{aligned}$$

where for a large $u = U > M$

$$\#\mathcal{F}_i \approx \#\mathcal{B}_i \quad (89)$$

$$\#\mathcal{E}_i = \#\mathcal{F}_i + \#\mathcal{B}_i \quad (90)$$

and,

$$e, f, g = 1, 2 \dots E, F, G \in \mathbb{N}$$

$$i = 1, 2 \dots \bar{I}$$

On the basis of these criteria we can populate \mathcal{E}_i for the i th-reaction of the ($u = U > M$)th-iteration,

$$a_{i_e} \sim (-1)^e \cdot a_{i_e} = \left\{ \begin{array}{l} (-1)^{2e} \cdot a_{i_{2e}} = a_{i_f} \\ (-1)^{2e-1} \cdot a_{i_{2e-1}} = a_{i_b} \end{array} \right\} \in \mathcal{E}_i \quad \text{Def. (24)}$$

where,

$$e, f, g = 1, 2 \dots E, F, G \in \mathbb{N}$$

$$i = 1, 2 \dots \bar{I}$$

Mapping the outcome-specific sum to the real-valued open interval $\mathbb{R} \cap (0, \infty)$

We have already identified a generic mapping schema where the lower- and upper-bounds of the reaction-specific sequence vector that comprise the real-valued terms of the null space generated subspace can be mapped to the open interval $\mathbb{R} \cap (0, \infty)$ (T2, T3; C5, C6) (Steps 6–9; Fig. 1). We will now define these maps formally for each outcome-specific subset,

Forward :

a) For the non empty subset \mathcal{F}_i with cardinality $\#\mathcal{F}_i$ and sum $\phi_{i_{\mathcal{F}}}$,

$$g : \phi_{i_{\mathcal{F}}} \in \mathbb{R} \cap (1, \infty) \mapsto y_{i_{\mathcal{F}}} \in \mathbb{R} \cap (1, \infty) \forall u = U > M \quad (91)$$

where,

$$y_{i_{\mathcal{F}}} = \phi_{i_{\mathcal{F}}} \quad (92)$$

b) For the non empty subsets $(\mathcal{F}_i, \mathcal{B}_i)$ with cardinalities $(\#\mathcal{F}_i, \#\mathcal{B}_i)$ and sums $(\phi_{i_{\mathcal{F}}}, \phi_{i_{\mathcal{B}}})$,

$$(g : \phi_{i_{\mathcal{F}}} \in \mathbb{R} \cap (1, \infty) + g : \phi_{i_{\mathcal{B}}} \in \mathbb{R} \cap (-\infty, -1)) \mapsto y_{i_{\mathcal{F}}} \in \mathbb{R} \cap (1, \infty) \forall u = U > M \quad (93)$$

iff,

$$\#\mathcal{F}_i \leq \#\mathcal{B}_i \quad (94)$$

$$|\phi_{i_{\mathcal{F}}} + (-1) \cdot \phi_{i_{\mathcal{B}}}| \gg 0 \quad (95)$$

where,

$$y_{i_{\mathcal{F}}} = \phi_{i_{\mathcal{F}}} + e^{\phi_{i_{\mathcal{B}}}} \quad (96)$$

Reverse :

For the subset \mathcal{B} with cardinality $\#\mathcal{B}_i$ and sum $\phi_{i_{\mathcal{B}}}$,

$$g : \phi_{i_{\mathcal{B}}} \in \mathbb{R} \cap (-\infty, -1) \mapsto y_{i_{\mathcal{B}}} \in \mathbb{R} \cap (0, 1) \forall u = U > M \quad (97)$$

where,

$$y_{i_{\mathcal{B}}} = e^{\phi_{i_{\mathcal{B}}}} \quad (98)$$

Equivalent :

For the non empty subset \mathcal{E} with cardinality $\#\mathcal{E}_i$ and sum $\phi_{i_{\mathcal{E}}}$,

$$g : \phi_{i_{\mathcal{E}}} \in \mathbb{R} \cap \{0\} \mapsto y_{i_{\mathcal{E}}} \in \mathbb{R} \cap \{1\} \forall u = U > M \quad (99)$$

iff,

$$\#\mathcal{F}_i \approx \#\mathcal{B}_i \quad (99.1)$$

$$\phi_{i_{\mathcal{E}}} \approx 0 \quad (99.2)$$

where,

$$y_{i_{\mathcal{E}}} = e^{\phi_{i_{\mathcal{E}}}} \quad (100)$$

The p1-norm of the reaction-specific outcome-vector is the probable dissociation constant for that reaction

Since every reaction can have these possible outcomes, we can define for the i th-reaction of the u th-iteration where $u = U > M$, a reaction-specific outcome vector,

$$g(\phi_i) \in \iota_i^{\mathbf{T}} = [g(\phi_{i_{\mathcal{F}}})g(\phi_{i_{\mathcal{B}}})g(\phi_{i_{\mathcal{E}}})]^{\mathbf{T}} \subset \mathbb{R}_+^{1 \times 3} \quad (101)$$

The p1-norm describes the final outcome of the i th-reaction after a finite number of $u = U > M$ -iterations (Step 10; Fig. 1),

$$\|\alpha^T\|_1 = g(\phi_{i_F}) + g(\phi_{i_B}) + g(\phi_{i_E}) \quad (102)$$

$$= y_{i_F} + y_{i_B} + y_{i_E} \quad (102.1)$$

$$\stackrel{\text{def}}{=} \eta_i \in \mathbb{R} \cap (0, \infty) \quad (103)$$

We can now summarize these assertions and unambiguously annotate every reaction of a biochemical network as,

$$\text{Forward} \quad \eta_i > 1 \quad (104)$$

$$\text{Reverse} \quad \eta_i \in (0, 1) \quad (105)$$

$$\text{Equivalent} \quad \eta_i \simeq 1.0 \quad (106)$$

We can easily see that the probable dissociation constant (η_i) for the i th-reaction of the modelled biochemical network can be equated to the true dissociation constant (Kd_i),

$$\eta_i \sim Kd_i \quad (107)$$

The algorithm is then iteratively applied to all other rows of a null space-generated subspace, i.e., reactions of the modelled biochemical network ($i = 1, 2, \dots, \bar{I}$) and the probable dissociation constants are computed for $\forall i$. If there is ambiguity in the annotation, the combinatorial summation and subsequent steps are recursively repeated⁴⁸. Complete annotation for every reaction of a biochemical network will occur if and only if the annotation is unambiguous. A biochemical network is only regarded as being completely annotated if and only if every reaction that comprises the network is annotated unambiguously⁴⁸.

Biochemical relevance and suitability of the probable dissociation constant to characterize a biochemical network of aerobic glycolysis

The biochemical network for AG ($\bar{I}_{AG} = 20$; $\eta_1 - \eta_{20}$) comprises single ($\bar{I}_{AG} = 14$)- and multi ($\bar{I}_{AG} = 5$)-step reactions in the cytosolic (c) and mitochondrial (m) compartments (Fig. 2) (Supplementary Text 2). Here, the multi-step reactions include the hexose monophosphate shunt (r_2, r_4), glycolysis (r_5), Krebs's tricarboxylic acid (TCA; r_{12}), oxaloacetate-malate shuttle (r_{16}) and Cori's cycle (r_{20}). The modelled biochemical network also includes transport reactions for Pyruvate (r_8), Citrate (r_{14}) and mitochondrial Phosphoenolpyruvate (r_{17}). The computed probable dissociation constants favour the export of oxaloacetate into the cytosol directly or through phosphoenolpyruvate ($\{\eta_{10}, \eta_{13}, \eta_{16}, \eta_{17}\} > 40$) thereby indicating a flux towards gluconeogenesis. The steps towards glucose 6-phosphate are supported by observing the probable dissociation constants for the following conversions: phosphoenolpyruvate \rightarrow glyceraldehyde 3-phosphate ($\eta_5 \approx 0.017$) and glyceraldehyde 3-phosphate \rightarrow glucose 6-phosphate ($\eta_3 \approx 0.03$) (Fig. 2). We can easily infer from the reverse reaction pyruvate \rightarrow acetyl-CoA ($\eta_9 \approx 0.03$) that the TCA-cycle is interrupted (Fig. 2). These findings suggest that contrary to available literature, the TCA-cycle is predisposed to being perturbed and its restoration or "intactness" is actively maintained for metabolic homeostasis or oxidative phosphorylation. These observations are perfectly logical given that the TCA-cycle is anaplerotic with several of its intermediates participating in collateral pathways. For AG to occur it is necessary that the TCA-cycle along with oxidative phosphorylation is interrupted in favour of an increased cell mass via the hexose monophosphate shunt pathway (DNA/RNA synthesis, NADPH) and fatty acid synthesis via acetyl-CoA (Fig. 2).

These data also offer several interesting insights into the metabolic crosstalk that may occur within cells when metabolizing Glucose to Lactate in the presence of adequate molecular dioxygen (Fig. 2). There is a significantly larger fraction (85%; $\bar{I}_{AG} = 17$) of non-equivalent reactions with forward (35%; $\bar{I}_{AG} = 7$)- and reverse (50%; $\bar{I}_{AG} = 10$)-reactions as compared to the equivalent (15%; $\bar{I}_{AG} = 3$) (Fig. 2). This suggests that the prevailing biochemical and physicochemical conditions where AG is expected to occur is, in fact an important determinant of whether AG occurs at all. Conversely, the paucity of equivalent reactions in the biochemical network for AG suggests lack of regulatory reactions. The ratio of NAD(P)H to NAD(P)⁺ along with Acetyl-CoA to CoA and ATP to ADP is an important regulator of the Pyruvate dehydrogenase complex (PDC) and is likely a major determinant for the metabolic shift observed to implement reversible AG.

Limitations of the presented algorithm and relevance of fractional derivatives in modelling and analysing complex biochemical systems

The biomedical relevance in computing the probable dissociation constant notwithstanding, the presented algorithm is dependent on the stoichiometry or integer numbers to model change to the molecular species that comprise the modelled biochemical network. This effect is partially offset by computing the null space, although the assumption that the modelled biochemical network is at equilibrium is yet another presumption. A further limitation of the presented algorithm is that it is enumeration-based and will therefore, be intractable for large biochemical networks. Furthermore, since the computations are entirely dependent on the null space, the initial nullity is also likely to effect the real time needed to annotate every reaction of a modelled biochemical network. This will mean that for smaller networks too, the time to complete annotation may also not be possible⁴⁸.

The probable dissociation constant is a numerical measure that is computed from the null space for a stoichiometry number matrix model of a biochemical network. However, most biochemical systems operate under non-equilibrium conditions. In fact, the true dissociation constant for a reaction is an empirical measure and is the ratio between the exponent (stoichiometry coefficients) forms for each reactant in a perfectly reversible reaction. The advent of large data warrants analytical approaches that depart significantly from previous methods

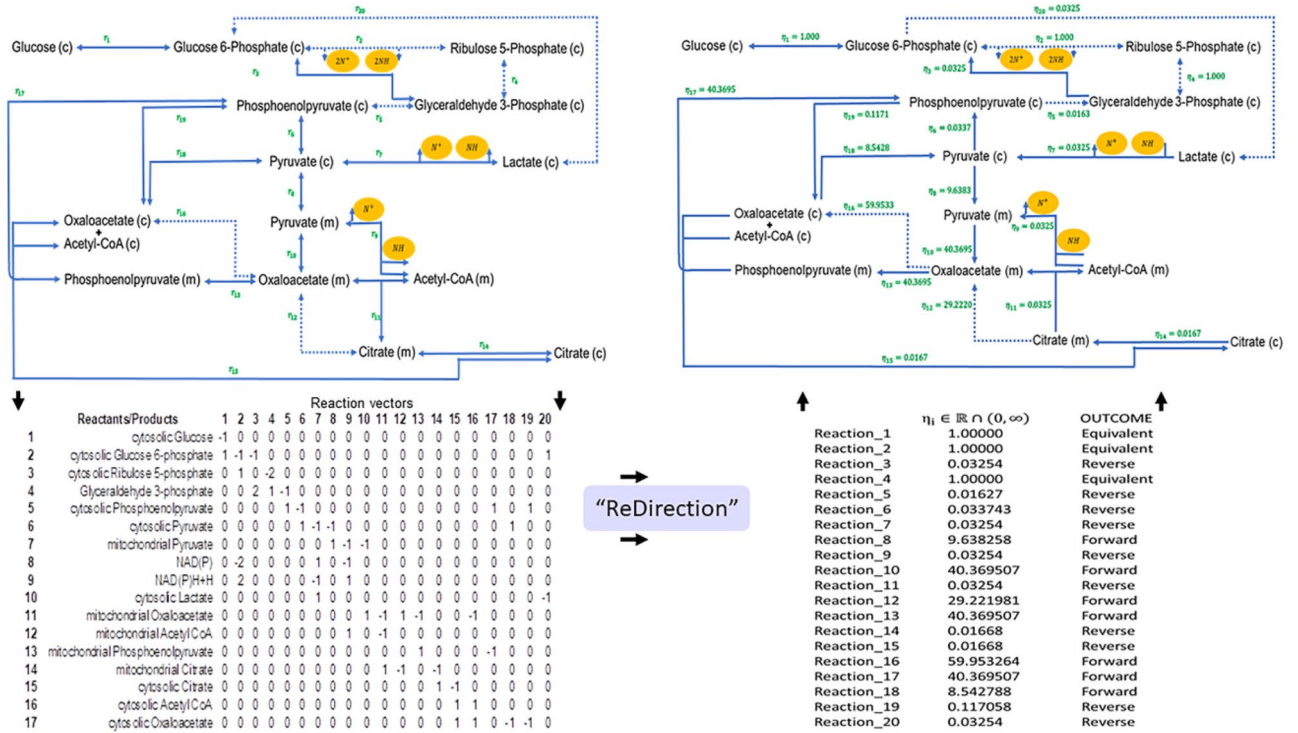


Figure 2. Computational studies with “ReDirection” to demonstrate suitability and biochemical relevance of the probable dissociation constant in characterizing a biochemical network for aerobic glycolysis. The dissociation constant for a reaction is an empirically derived parameter and can reveal valuable information such as the rate and rate constant for a reaction. Theoretically-derived approximations are usually based on user-defined bounds and/or pre-selecting strictly positive null space spanning vectors. Although the computations are rendered more feasible with these assumptions, the biomedical relevance of the data generated with these numeric constraints is likely to be of limited relevance. The algorithm presented computes the probable dissociation constant from first principles and with biochemically relevant constraints. A biochemical network of aerobic glycolysis is modelled and defined in terms of a sparse matrix of stoichiometric numbers of its reactants/products ($J = 17$) and reactions ($\bar{I} = 20$). “ReDirection” checks this user-defined matrix and computes the null space, utilizes mathematically sound (tests of convergence, descriptive statistics, linear maps) constraints to compute the probable dissociation constant and thereby assign a dominant direction to every reaction of the modelled biochemical network of aerobic glycolysis. \bar{I} , Total number of reactions of modelled biochemical network; J , Total number of reactants of modelled biochemical network; $S_{j,i}$, reaction-centric and user-defined stoichiometry number matrix of a biochemical network for aerobic glycolysis; \mathbf{m}, \mathbf{c} ; mitochondrial and cytosolic components of a reactant/product; η_i , probable dissociation constant for the i th-reaction of a biochemical network with forward (\mathcal{F}), reverse (\mathcal{B}), or equivalent (\mathcal{E}) outcomes.

in an effort to optimize computational resources whilst delivering meaningful information about the network. The use of fractional derivatives has gained traction over the last several years in an effort to impute missing data, improve convergence times and model complex systems^{62–66}. For example, fractional derivatives are being used to improve convergence in the presence of multiple local optima and model and/or analyse the complex behaviour of fluids, biomolecules and biochemical systems such as anomalous diffusion in the cytoplasm and across membranes and even telomeres in the nucleoplasm^{62–66}. All these studies have reported significant improvements in predicting an outcome, goodness-of-fit studies and time to convergence.

In general, a fractional derivative of α -order can be approximated with the gamma function using the Riemann–Liouville (RL), Caputo (CAP) or the Grunwald–Letnikov (GL) methods^{67–70}. Consider the Caputo fractional derivative,

$$D^\alpha x^\beta = \left(\frac{\Gamma(\beta+1)}{\Gamma(\beta-\alpha+1)} \right) x^{\beta-\alpha} \quad (108)$$

where,

$$\beta \in \mathbb{Z}_+ \cup \{0\} \quad (109)$$

$$\alpha \in (\beta - 1, \beta) \quad (110)$$

$$\alpha \in \mathbb{R}_+ \quad (111)$$

In the absence of empirical data investigating a biochemical network is challenging. One strategy for interrogating a biochemical network at non-equilibrium or near steady-state conditions is approximating the dissociation constant for a reaction. This equation is fundamental to biochemical analysis and involves a

phenomenological approximation of the integer values for the stoichiometry numbers of the reactants with an expression that involves real number exponents for each reactant of a fully reversible reaction. In accordance with Reaction 1 this is written as,

$$Kd_i = \frac{\prod_{j=1}^{j=M} [A_j]^{a_j}}{\prod_{j=1}^{j=M} [A_j]^{a_j}} \quad (112)$$

where,
 $a_j \in \mathbb{R}$

We can then proceed to compute the fractional gradient descent, post hoc, to a preliminary imputation step to estimate missing or latent data points^{62,63,69,70}. We can combine a small learning rate with a constant (stochastic) or mini-batch update protocol concomitantly with several novel algorithmic approaches to accomplish these steps^{62,63,69,70},

$$\mathbf{r}_{i,x+1} = \mathbf{r}_{i,x} - \gamma \cdot D^\alpha \left((\mathbf{r}_{i,x})^\beta; ((x+1) \vee (x+y)) \right) \quad (113)$$

where,

$$\begin{aligned} \gamma &:= \text{Learning rate} \\ &\in \mathbb{R} \cap (0, 1) \end{aligned} \quad (114)$$

$$\begin{aligned} x &:= \text{Finite number of iterations} \\ &= (0, 1, \dots) \end{aligned} \quad (115)$$

$$\begin{aligned} y &:= \text{Finite number of updates} \\ &\in \mathbb{N} \end{aligned} \quad (116)$$

The imputed values computed earlier will represent an initial approximation to the original stoichiometry numbers matrix for a biochemical network and can be refined iteratively,

$$D^\alpha (\mathbf{r}_i)^\beta \stackrel{\text{def}}{=} D^\alpha \begin{bmatrix} m_1 \\ m_1 \\ \vdots \\ m_J \end{bmatrix}^\beta \quad (117)$$

$$= \begin{bmatrix} \frac{\partial^\alpha}{\partial m_1} (m_1)^\beta \\ \frac{\partial^\alpha}{\partial m_2} (m_2)^\beta \\ \vdots \\ \frac{\partial^\alpha}{\partial m_J} (m_J)^\beta \end{bmatrix} \quad (117.1)$$

$$= \begin{bmatrix} \left(\frac{\Gamma(\beta+1)}{\Gamma(\beta-\alpha+1)} \right) \cdot (m_1)^{\beta-\alpha} \\ \left(\frac{\Gamma(\beta+1)}{\Gamma(\beta-\alpha+1)} \right) \cdot (m_2)^{\beta-\alpha} \\ \vdots \\ \left(\frac{\Gamma(\beta+1)}{\Gamma(\beta-\alpha+1)} \right) \cdot (m_J)^{\beta-\alpha} \end{bmatrix} \quad (117.2)$$

$$\stackrel{\text{def}}{=} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_J \end{bmatrix} \quad (118)$$

$$= \mathbf{w}_i \quad (119)$$

where,

$$\begin{aligned} \mathbf{r}_i &\in \mathbb{Z}^J \\ \omega_j \in \mathbf{w}_i &\in \mathbb{R}^J \end{aligned} \quad (120)$$

$$\begin{aligned} i &\in [1, \bar{I}] \\ \beta &\in \mathbb{Z}_+ \cup \{0\} \\ \alpha &\in (\beta - 1, \beta) \\ \alpha &\in \mathbb{R}_+ \end{aligned}$$

The computed numerical values can be used to assess the contribution of each molecular species to a reaction vector and are easily mapped to the exponents for each reactant. This can then be directly included in the formula to estimate the dissociation constant for a reaction at a particular step. Since this is a strictly positive real number we can use this metric to check whether each reaction vector converges to the expected outcome,

$$\text{Kd}_i = \frac{\prod_{j=1}^{i=M} [A_j]^{a_j}}{\prod_{j=1}^{i=M} [A_j]^{a_j}} \quad (121)$$

$$= \mathbb{R} \cap ((0, 1) \vee \{1\} \vee (1, \infty)) \quad (121.1)$$

Biomedical relevance of the presented algorithm to the study of biochemical networks

The dissociation constant for a reaction is a fundamental parameter in biochemical analysis and is computed predominantly from empirical data. However, this information is sparse and unavailable for several biochemical reactions of interest. For those reactions that laboratory data is available, the absence of corroborating *in vivo* details will render these less likely to have clinical- or biomedical-relevance. Additionally, biochemical assays comprise single or at most coupled reactions which further limits the utility of this information since complex biochemical function is brought about by biochemical networks. Although several algorithms utilize the null space, these are directed towards comprehending possible stable states for a modelled biochemical network. Parameter selection is also left to the discretion of the user further limiting the portability of these studies.

Unlike these investigations the probable dissociation constants are computed for every reaction and directly from a null space-generated subspace of the stoichiometry number matrix for a biochemical network. The algorithm concentrates on using combinatorial summations of unique and non-trivial null space spanning vectors to estimate the probable dissociation constant for every reaction along with mathematically rigorous and statistically relevant term-selection. A significant contribution of this algorithm is that it does not exclude numerical values and in doing so ensures that the computations are all encompassing. The computation of the probable dissociation constants are not only user-independent and computed from first principles, but may also be biochemically relevant. Since the mapping of the probable dissociation constants to the positive real number space is in accordance with established biochemical paradigms, the information gleaned can easily be incorporated into simulation studies for a biochemical network and used to compute the trajectories for a reactant/product^{38,40}. Furthermore, the distribution of the probable dissociation constants across a biochemical network can offer plausible explanations into metabolite flux and usage. These testable hypotheses may be corroborated with detailed site-directed mutagenesis experiments and labelling studies among several others^{44–47}. The probable dissociation constants can also be used to compare individual biochemical reactions of a network. These investigations may also be complemented by detailed theoretical studies such as phylogenetic analysis for enzyme-mediated reactions to ascertain the isoforms that may be deployed by a biochemical network. It may also be possible to perturb these networks and investigate the distribution for each pair of reactions of a biochemical network, *i.e.*, before and after a stimulus.

Conclusions

The probable dissociation constant for a biochemical reaction is a numerical measure which can be computed directly from a null space-generated subspace of the stoichiometry number matrix for a biochemical network. Here, we present a mathematically rigorous algorithm to compute the probable dissociation constant for a reaction and thence for every reaction of a biochemical network. Each step of the algorithm is outlined and supported by the necessary mathematical formalism that is needed to comprehend and utilize this framework. The theoretical assertions presented in the main text are supported by formal proofs wherever applicable and are available and accessible as Supplementary Material. The biological relevance of this work is illustrated with computational studies and analyses of the probable dissociation constants for a well characterized biochemical network. The data generated is in accordance with established kinetic paradigms and is therefore, a suitable index of biochemical function and can be used to parameterize and study a biochemical network. These data are also amenable to detailed analyses and can be used to test various hypotheses for both, baseline- and perturbed-conditions of different cell types and across taxa. Additionally, the resulting data may offer valuable insights into the physiological and thence the pathological basis of diseases such as malignancies, altered immune response, etc.

Data availability

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Received: 21 June 2023; Accepted: 30 January 2024

Published online: 12 February 2024

References

- Kim, O. D., Rocha, M. & Maia, P. A review of dynamic modeling approaches and their application in computational strain optimization for metabolic engineering. *Front. Microbiol.* **9**, 1690. <https://doi.org/10.3389/fmicb.2018.01690> (2018).
- Maarleveld, T. R., Khandelwal, R. A., Olivier, B. G., Teusink, B. & Bruggeman, F. J. Basic concepts and principles of stoichiometric modeling of metabolic networks. *Biotechnol. J.* **8**, 997–1008. <https://doi.org/10.1002/biot.201200291> (2013).
- Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis?. *Nat. Biotechnol.* **28**, 245–248. <https://doi.org/10.1038/nbt.1614> (2010).
- Klamt, S., Muller, S., Regensburger, G. & Zanghellini, J. A mathematical framework for yield (vs. rate) optimization in constraint-based modeling and applications in metabolic engineering. *Metab. Eng.* **47**, 153–169. <https://doi.org/10.1016/j.ymben.2018.02.001> (2018).
- Feist, A. M. & Palsson, B. O. The biomass objective function. *Curr. Opin. Microbiol.* **13**, 344–349. <https://doi.org/10.1016/j.mib.2010.03.003> (2010).
- Klamt, S. *et al.* From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints. *PLoS Comput. Biol.* **13**, e1005409. <https://doi.org/10.1371/journal.pcbi.1005409> (2017).

7. Muller, S. & Regensburger, G. Elementary vectors and conformal sums in polyhedral geometry and their relevance for metabolic pathway analysis. *Front. Genet.* **7**, 90. <https://doi.org/10.3389/fgene.2016.00090> (2016).
8. Urbanczik, R. Enumerating constrained elementary flux vectors of metabolic networks. *IET Syst. Biol.* **1**, 274–279. <https://doi.org/10.1049/iet-syb:20060073> (2007).
9. Wagner, C. & Urbanczik, R. The geometry of the flux cone of a metabolic network. *Biophys. J.* **89**, 3837–3845. <https://doi.org/10.1529/biophysj.104.055129> (2005).
10. Rottman, B. M. & Hastie, R. Reasoning about causal relationships: Inferences on causal networks. *Psychol. Bull.* **140**, 109–139. <https://doi.org/10.1037/a0031903> (2014).
11. Gopalan, L., Sebastian, A., Praul, C. A., Albert, I. & Ramachandran, R. Metformin affects the transcriptomic profile of chicken ovarian cancer cells. *Genes (Basel)* <https://doi.org/10.3390/genes13010030> (2021).
12. Saptarshi, N. *et al.* Epigenetic age acceleration is not associated with age-related macular degeneration. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms222413457> (2021).
13. Salvador, A. C. *et al.* Sex-specific genetic architecture in response to American and ketogenic diets. *Int. J. Obes. (Lond)* **45**, 1284–1297. <https://doi.org/10.1038/s41366-021-00785-7> (2021).
14. Lu, J. *et al.* Causal network inference from gene transcriptional time-series response to glucocorticoids. *PLoS Comput. Biol.* **17**, e1008223. <https://doi.org/10.1371/journal.pcbi.1008223> (2021).
15. Goto, T., Fernandes, A. F. A., Tsudzuki, M. & Rosa, G. J. M. Causal phenotypic networks for egg traits in an F2 chicken population. *Mol. Genet. Genomics* **294**, 1455–1462. <https://doi.org/10.1007/s00438-019-01588-2> (2019).
16. Du, Q. *et al.* Genetic architecture of growth traits in *Populus* revealed by integrated quantitative trait locus (QTL) analysis and association studies. *New Phytol* **209**, 1067–1082. <https://doi.org/10.1111/nph.13695> (2016).
17. Gove, R. P. *et al.* Effects of causal networks on the structure and stability of resource allocation trait correlations. *J. Theor. Biol.* **293**, 1–14. <https://doi.org/10.1016/j.jtbi.2011.09.034> (2012).
18. Seyhan, A. A. & Carini, C. Are innovation and new technologies in precision medicine paving a new era in patients centric care?. *J. Transl. Med.* **17**, 114. <https://doi.org/10.1186/s12967-019-1864-9> (2019).
19. Biane, C. & Delaplace, F. Causal reasoning on Boolean control networks based on abduction: Theory and application to cancer drug discovery. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **16**, 1574–1585. <https://doi.org/10.1109/TCBB.2018.2889102> (2019).
20. Keller, M. P. & Attie, A. D. Physiological insights gained from gene expression analysis in obesity and diabetes. *Annu. Rev. Nutr.* **30**, 341–364. <https://doi.org/10.1146/annurev.nutr.012809.104747> (2010).
21. Ferrara, C. T. *et al.* Genetic networks of liver metabolism revealed by integration of metabolic and transcriptional profiling. *PLoS Genet.* **4**, e1000034. <https://doi.org/10.1371/journal.pgen.1000034> (2008).
22. Haraldsdottir, H. S., Thiele, I. & Fleming, R. M. Quantitative assignment of reaction directionality in a multicompartmental human metabolic reconstruction. *Biophys. J.* **102**, 1703–1711. <https://doi.org/10.1016/j.bpj.2012.02.032> (2012).
23. Shindo, Y., Kondo, Y. & Sako, Y. Inferring a nonlinear biochemical network model from a heterogeneous single-cell time course data. *Sci. Rep.* **8**, 6790. <https://doi.org/10.1038/s41598-018-25064-w> (2018).
24. Wittenstein, T., Leibovich, N. & Hilfinger, A. Quantifying biochemical reaction rates from static population variability within incompletely observed complex networks. *PLoS Comput. Biol.* **18**, e1010183. <https://doi.org/10.1371/journal.pcbi.1010183> (2022).
25. Reinker, S., Altman, R. M. & Timmer, J. Parameter estimation in stochastic biochemical reactions. *Syst. Biol. (Stevenage)* **153**, 168–178. <https://doi.org/10.1049/ip-syb:20050105> (2006).
26. Lecca, P., Palmisano, A., Priami, C. & Sanguinetti, G. In *Proceedings of the 2009 ACM Symposium on Applied Computing—SAC '09* (2009).
27. Koutrouli, M., Karatzas, E., Paez-Espino, D. & Pavlopoulos, G. A. A guide to conquer the biological network era using graph theory. *Front. Bioeng. Biotechnol.* **8**, 34. <https://doi.org/10.3389/fbioe.2020.00034> (2020).
28. Radak, B. K. *et al.* Constant-pH molecular dynamics simulations for large biomolecular systems. *J. Chem. Theory Comput.* **13**, 5933–5944. <https://doi.org/10.1021/acs.jctc.7b00875> (2017).
29. Ruusmann, V., Sild, S. & Maran, U. QSAR DataBank repository: Open and linked qualitative and quantitative structure–activity relationship models. *J. Cheminform.* **7**, 32. <https://doi.org/10.1186/s13321-015-0082-6> (2015).
30. Riva, S. G. *et al.* SMGen: A generator of synthetic models of biochemical reaction networks. *Symmetry* <https://doi.org/10.3390/sym14010119> (2022).
31. Gillespie, D. T. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58**, 35–55. <https://doi.org/10.1146/annurev.physchem.58.032806.104637> (2007).
32. Yu, P. Y. & Craciun, G. Mathematical analysis of chemical reaction systems. *Israel J. Chem.* **58**, 733–741. <https://doi.org/10.1002/ijch.201800003> (2018).
33. Kundu, S. Modeling ligand–macromolecular interactions as eigenvalue-based transition-state dissociation constants may offer insights into biochemical function of the resulting complexes. *Math. Biosci. Eng.* **19**, 13252–13275. <https://doi.org/10.3934/mbe.2022620> (2022).
34. Kim, K. & Guck, J. *The Relative Densities of Cell Cytoplasm, Nucleoplasm, and Nucleoli are Robustly Conserved During Cell Cycle and Drug Perturbations* <https://doi.org/10.1101/2020.04.14.040774> (2020).
35. Persson, L. B., Ambati, V. S. & Brandman, O. Cellular control of viscosity counters changes in temperature and energy availability. *Cell* **183**, 1572–1585.e1516. <https://doi.org/10.1016/j.cell.2020.10.017> (2020).
36. Khan, M., Irfan, M. & Khan, W. A. Impact of nonlinear thermal radiation and gyrotactic microorganisms on the Magneto-Burgers nanofluid. *Int. J. Mech. Sci.* **130**, 375–382. <https://doi.org/10.1016/j.ijmecsci.2017.06.030> (2017).
37. Khan, W. A., Khan, M. & Alshomrani, A. S. Impact of chemical processes on 3D Burgers fluid utilizing Cattaneo-Christov double-diffusion: Applications of non-Fourier's heat and non-Fick's mass flux models. *J. Mol. Liq.* **223**, 1039–1047. <https://doi.org/10.1016/j.molliq.2016.09.027> (2016).
38. Kundu, S. Stochastic modelling suggests that an elevated superoxide anion–hydrogen peroxide ratio can drive extravascular phagocyte transmigration by lamellipodium formation. *J. Theor. Biol.* **407**, 143–154. <https://doi.org/10.1016/j.jtbi.2016.07.002> (2016).
39. Gerstl, M. P., Muller, S., Regensburger, G. & Zanghellini, J. Flux tope analysis: Studying the coordination of reaction directions in metabolic networks. *Bioinformatics* **35**, 266–273. <https://doi.org/10.1093/bioinformatics/bty550> (2019).
40. Kundu, S. Mathematical modeling and stochastic simulations suggest that low-affinity peptides can bisect MHC1-mediated export of high-affinity peptides into “early”- and “late”-phases. *Heliyon* **7**, e07466. <https://doi.org/10.1016/j.heliyon.2021.e07466> (2021).
41. Aurbek, N., Thiermann, H., Szinicz, L., Eyer, P. & Worek, F. Analysis of inhibition, reactivation and aging kinetics of highly toxic organophosphorus compounds with human and pig acetylcholinesterase. *Toxicology* **224**, 91–99. <https://doi.org/10.1016/j.tox.2006.04.030> (2006).
42. Johnson, D. S., Weerapana, E. & Cravatt, B. F. Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future Med. Chem.* **2**, 949–964. <https://doi.org/10.4155/fmc.10.21> (2010).
43. Nyce, J. W. Detection of a novel, primate-specific “kill switch” tumor suppression mechanism that may fundamentally control cancer risk in humans: An unexpected twist in the basic biology of TP53. *Endocr. Relat. Cancer* **25**, R497–R517. <https://doi.org/10.1530/ERC-18-0241> (2018).
44. Furukawa, A., Konuma, T., Yanaka, S. & Sugase, K. Quantitative analysis of protein–ligand interactions by NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **96**, 47–57. <https://doi.org/10.1016/j.pnmrs.2016.02.002> (2016).

45. Sparks, R. P., Jenkins, J. L. & Fratti, R. Use of surface plasmon resonance (SPR) to determine binding affinities and kinetic parameters between components important in fusion machinery. *Methods Mol. Biol.* **199**–210, 2019. https://doi.org/10.1007/978-1-4939-8760-3_12 (1860).
46. Sura, R. & Antalík, M. Determination of proton dissociation constants (pK(a)) of hydroxyl groups of 2,5-dihydroxy-1,4-benzoquinone (DHBQ) by UV-Vis, fluorescence and ATR-FTIR spectroscopy. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **271**, 120863. <https://doi.org/10.1016/j.saa.2022.120863> (2022).
47. Bhowmik, P. & Das Gupta, S. K. Biochemical characterization of a mycobacteriophage derived DnaB ortholog reveals new insight into the evolutionary origin of DnaB helicases. *PLoS One* **10**, e0134762. <https://doi.org/10.1371/journal.pone.0134762> (2015).
48. Kundu, S. ReDirection: An R-package to compute the probable dissociation constant for every reaction of a user-defined biochemical network. *Front. Mol. Biosci.* <https://doi.org/10.3389/fmolb.2023.1206502> (2023).
49. Warburg, O. On the origin of cancer cells. *Science* **123**, 309–314. <https://doi.org/10.1126/science.123.3191.309> (1956).
50. Jones, W. & Bianchi, K. Aerobic glycolysis: Beyond proliferation. *Front. Immunol.* **6**, 227. <https://doi.org/10.3389/fimmu.2015.00227> (2015).
51. Carey, B. W., Finley, L. W. S., Cross, J. R., Allis, C. D. & Thompson, C. B. Intracellular α -ketoglutarate maintains the pluripotency of embryonic stem cells. *Nature* **518**, 413–416. <https://doi.org/10.1038/nature13981> (2014).
52. Goodell, M. A. *et al.* Quiescent fibroblasts exhibit high metabolic activity. *PLoS Biol.* <https://doi.org/10.1371/journal.pbio.1000514> (2010).
53. Chang, C.-H. *et al.* Posttranscriptional control of T cell effector function by aerobic glycolysis. *Cell* **153**, 1239–1251. <https://doi.org/10.1016/j.cell.2013.05.016> (2013).
54. Cheng, S.-C. *et al.* mTOR- and HIF-1 α -mediated aerobic glycolysis as metabolic basis for trained immunity. *Science* <https://doi.org/10.1126/science.1250684> (2014).
55. Suda, T., Takubo, K. & Semenza, G. L. Metabolic regulation of hematopoietic stem cells in the hypoxic niche. *Cell Stem Cell* **9**, 298–310. <https://doi.org/10.1016/j.stem.2011.09.010> (2011).
56. Tannahill, G. M. *et al.* Succinate is an inflammatory signal that induces IL-1 β through HIF-1 α . *Nature* **496**, 238–242. <https://doi.org/10.1038/nature11986> (2013).
57. Colegio, O. R. *et al.* Functional polarization of tumour-associated macrophages by tumour-derived lactic acid. *Nature* **513**, 559–563. <https://doi.org/10.1038/nature13490> (2014).
58. De Bock, K. *et al.* Role of PFKFB3-driven glycolysis in vessel sprouting. *Cell* **154**, 651–663. <https://doi.org/10.1016/j.cell.2013.06.037> (2013).
59. Schoors, S. *et al.* Partial and transient reduction of glycolysis by PFKFB3 blockade reduces pathological angiogenesis. *Cell Metab.* **19**, 37–48. <https://doi.org/10.1016/j.cmet.2013.11.008> (2014).
60. Patel, M. S., Nemeria, N. S., Furey, W. & Jordan, F. The pyruvate dehydrogenase complexes: Structure-based function and regulation. *J. Biol. Chem.* **289**, 16615–16623. <https://doi.org/10.1074/jbc.R114.563148> (2014).
61. Yu, X. *et al.* Structures of the human pyruvate dehydrogenase complex cores: A highly conserved catalytic center with flexible N-terminal domains. *Structure* **16**, 104–114. <https://doi.org/10.1016/j.str.2007.10.024> (2008).
62. Chen, Z. & Wang, S. A review on matrix completion for recommender systems. *Knowl. Inf. Syst.* **64**, 1–34. <https://doi.org/10.1007/s10115-021-01629-6> (2022).
63. Khan, Z. A. *et al.* Enhanced fractional prediction scheme for effective matrix factorization in chaotic feedback recommender systems. *Chaos Solitons Fractals.* <https://doi.org/10.1016/j.chaos.2023.114109> (2023).
64. Couceiro, M. & Sivasundaram, S. Novel fractional order particle swarm optimization. *Appl. Math. Comput.* **283**, 36–54. <https://doi.org/10.1016/j.amc.2016.02.007> (2016).
65. Bronstein, I. *et al.* Transient anomalous diffusion of telomeres in the nucleus of mammalian cells. *Phys. Rev. Lett.* **103**, 018102. <https://doi.org/10.1103/PhysRevLett.103.018102> (2009).
66. Weigel, A. V., Simon, B., Tamkun, M. M. & Krapf, D. Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking. *Proc. Natl. Acad. Sci. USA* **108**, 6438–6443. <https://doi.org/10.1073/pnas.1016325108> (2011).
67. Almeida, R. A Caputo fractional derivative of a function with respect to another function. *Commun. Nonlinear Sci. Numer. Simul.* **44**, 460–481. <https://doi.org/10.1016/j.cnsns.2016.09.006> (2017).
68. de Oliveira, E. C. & Tenreiro Machado, J. A. A review of definitions for fractional derivatives and integral. *Math. Probl. Eng.* **2014**, 1–6. <https://doi.org/10.1155/2014/238459> (2014).
69. Tang, J. & Ben Makhlof, A. Fractional gradient descent-based auxiliary model algorithm for FIR models with missing data. *Complexity* **1–12**, 2023. <https://doi.org/10.1155/2023/7527478> (2023).
70. Chaudhary, N. I., Raja, M. A. Z., Khan, Z. A., Cheema, K. M. & Milyani, A. H. Hierarchical quasi-fractional gradient descent method for parameter estimation of nonlinear ARX systems using key term separation principle. *Mathematics* <https://doi.org/10.3390/math9243302> (2021).

Acknowledgements

This study was supported by an extramural grant from the Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India under the Mathematical Research Impact-Centric Support (MATRICS) scheme to SK (MTR/2021/000290).

Author contributions

S.K., designed the study; formulated and developed the model, developed the algorithms; wrote all proofs; conducted the analysis; collated and analysed data; wrote all the code and manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-53231-9>.

Correspondence and requests for materials should be addressed to S.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024