RESEARCH

SBMLKinetics: A Tool for Annotation Independent Data Driven Classification of Reaction Kinetics for SBML Models

Jin Xu

Full list of author information is available at the end of the article

Abstract

Background: Reaction networks are widely used as mechanistic models in Systems Biology to reveal principles of biological systems. Reactions are governed by kinetic laws that describe reaction rates. Selecting the appropriate kinetic laws is difficult for many modelers. There exist tools that attempt to find the correct kinetics law based on annotations. Here, we develop annotation independent technologies that assist modelers by focusing on finding kinetic laws commonly used for similar reactions.

Results: Recommending kinetic laws and other analyses of reaction networks can be viewed as a classification problem. Existing approaches to determining similar reactions rely heavily on having good annotations of chemical species, a condition that is often unsatisfied in model repositories such as BioModels. We develop an annotation independent *data driven* approach to find similar reactions via reaction classification, and demonstrate our approach using BioModels. Our two dimensional kinetics classification scheme (2DK) analyzes reactions along the dimensions of kinetics type ($\mathbf{K\ type}$) and reaction type ($\mathbf{R\ type}$). We identify approximately 10 mutually exclusive K types, including zeroth order, mass action, Michaelis-Menten, Hill kinetics and others. R types are organized by the number of reactants and the number of products in reactions. We construct a tool, SBMLKinetics, that inputs a collection of SBML models and then calculates reaction classifications as the probability of each 2DK class. We apply our tool to BioModels to compare the kinetics of signaling networks with the kinetics of metabolic networks, and illustrate significant difference of $\mathbf{K\ type}$ distributions.

Conclusions: 2DK has many applications. It provides an annotation independent, *data driven* approach to recommending kinetic laws by using $\mathbf{K\ type}$ common for the kind of model in combination with the $\mathbf{R\ type}$ of the reaction to recommend kinetic laws. Alternatively, 2DK can also be used to alert users that a kinetic law is unusual for the $\mathbf{K\ type}$ and $\mathbf{R\ type}$. Last, 2DK provides a way to analyze groups of models to compare their kinetic laws as we did with signaling and metabolic networks in BioModels.

Keywords: Systems biology; Kinetics; Data analysis; Software; Computational modeling

Background

Reaction networks are widely used as mechanistic models such as in Systems Biology that can reveal governing principles of biological systems. The choice of kinetic laws for reactions (e.g., mass action kinetics, Michaelis-Menten kinetics) is crucial for the development of mechanistic models of biological systems. Computational

studies of Systems Biology often involve the analysis of synthetically generated networks (e.g. [1, 2]). As noted in [3, 4], the choice of kinetic laws for these networks is as important as the set of reactions (the reactants, products, and modifiers for reactions) since reactions alone are not sufficient to determine dynamics [5]. Indeed, many computational studies only consider mass action kinetics, which may be inadequate [1]. In addition, kinetics are crucial considerations in parameter inference and model identification [6].

There is a wide diversity of kinetic laws. For example, allosteric and cooperative kinetics are important considerations in some biological systems [7, 8], and there are generalizations of these laws [9, 10], such as convenience kinetics [11]. Complications arise if there are inhibitors or activators that target the active site (non-allosteric). A further consideration is the order in which elemental reactions occur to form complexes such as compulsory order and ping-pong [12, 13, 14].

A significant motivation for our work is addressing the challenges modelers face with choosing appropriate kinetic laws. To date, the approaches have mostly been *theory driven* approach which is based on the falsifiability of hypothesis, while in data science the *data driven* approach discovers patterns in data without a predefined hypothesis. A detailed example of choosing kinetic laws is given for models of glycolysis [15] and Trypanosoma models [16, 17]. Further, King-Altman [18] is a systematic way to derive Michaelis-Menten kinetics using the steady-state assumption. Kinetic laws are chosen based on the best knowledge of the underlining kinetic mechanism. There are books that describe kinetic laws, such as Segel's enzyme kinetics [13].

Despite the examples, there remains a significant challenge with applying the theory in practice, especially for less experienced modelers. To address this, programs such as COPASI [19] and CellDesigner [20], provide pre-defined lists of kinetic laws. SBMLsqueezer2 [21] goes further by making recommendations for kinetic laws in two ways. One way is a theory driven classification scheme that requires the annotation of all participating reactants, products and regulators. These annotations come from the Systems Biology Ontology (SBO) [22] and Minimal Information Required In the Annotation of Models (MIRIAM) [23]. The second approach also depends on annotations to match reactions in the SBAIO-RK [24] database to obtain kinetic laws. We emphasize that both approaches are limited because of requiring annotations. Further, the authors note that even with the use of annotations there are serious limitations in that "the extraction from SABIO-RK depends on existing biochemical data and might therefore not always yield results." Here, we expand one aspect of the approach taken in SBMLsqueezer2 to finding similar reactions. At the heart of this approach is a way to classify reactions. In SBMLsqueezer2, this is done using the annotations of chemical species, an approach that they report has significant shortcomings. In addition, their classification scheme is theory driven.

In this article, we develop a classification scheme that does not require annotations. Our two dimensional kinetics classification scheme (2DK) analyzes reactions along the dimensions of kinetics type (**K type**) and reaction type (**R type**). We identify approximately 10 mutually exclusive K types, including zeroth order, mass action, Michaelis-Menten, Hill kinetics and others. R types are organized by the number of reactants and the number of products in reactions. We developed SBMLKinetics, a tool that constructs reaction classifications by inputting a collection of

Systems Biology Markup Language (SBML) models and outputting the probability of each 2DK class (combinations of **K type** and **R type**). We demonstrate SBMLKinetics by applying it to the BioModels Database as an example to compare the kinetics of signaling networks with the kinetics of metabolic networks. The results show that the signaling networks are dominated by unidirectional mass action kinetics whereas metabolic networks have a proportionally larger number of regulatory kinetics in the form of kinetic laws that have a fractional representation. This is reasonable because metabolic models are derived heavily from underlying enzyme mechanisms whereas signaling pathways are generally simpler binding/unbinding reactions or gross simplifications of the kinetics because very little is known about the mechanism. We further describe how 2DK can provide a *data driven* approach to recommending kinetic laws.

Methods

We develop a *data driven* classification scheme for kinetic laws based on patterns in existing reaction networks. Our focus is on the standard community of SBML [25], and our approach classifies reactions along two dimensions. The first is a characterization of the algebraic expression of the kinetic law. We refer to this as the kinetics type (**K type**) of the reaction. The second is a simplification of reaction stoichiometry, namely the number of reactants and the number of products. We refer to this as the reaction type (**R type**) of the reaction.

1. Define two dimensional kinetics classification scheme (2DK)

Our two dimensional kinetics classification scheme (2DK) organizes reactions along two dimensions: kinetics type (**K type**) and reaction type (**R type**).

1.1 Define kinetics type (K type)

On one hand, **K type** is defined mainly by the algebraic expressions of kinetic laws. For example, the kinetics with a reaction of $A + B \rightarrow C$ and its kinetic law $K1$ is different from the same reaction but with another kinetic law $K1 * A * B$. On the other hand, **K type** also includes some information from the reactions. For example, it is necessary to consider whether a species inside the kinetic law is a reactant or not. In order to evaluate the kinetics statistically, we have classified the kinetics into commonly used and mutually exclusive ten types, in contrast to SBO [26], with defined categories as below.

- **Zeroth order (ZERO)**: the number of species in the kinetic law is zero;
- **Uni-directional mass action (UNDR)**: the kinetic law is a single product of terms and all species in the kinetic law are reactants;
- **Uni-term with the moderator (UNMO)**: the kinetic law is a single product of terms and NOT UNDR;
- **Bi-directional mass action (BIDR)**: the kinetic law is the difference between two products of terms and the first product of terms with all reactants while the second product of terms with all products;
- **Bi-terms with the moderator (BIMO)**: the kinetic law is the difference between two products of terms and NOT BIDR;

- **Michaelis-Menten kinetics without explicit enzyme (MM)**: the kinetic law is in the format of Michaelis-Menten expressions without an explicit enzyme;
- **Michaelis-Menten kinetics with an explicit enzyme (MMCAT)**: the kinetic law is in the format of Michaelis-Menten expressions with an explicit enzyme;
- **Hill equation (HILL)**: the kinetic law is in the format of Hill equations;
- **Fraction format other than MM, MMCAT and HILL (FR)**: the kinetic law is in the format of fraction with at least one species in the denominator and NOT MM, MMCAT or HILL;
- **Not classified (NA)**: not classified kinetics.

Table 1 Kinetics examples for ten defined kinetics types.

K type	BIOMD	Reaction	Kinetic law
ZERO	5	EmptySet → Y	cell*k1aa
UNDR	5	CP → C2	cell***CP***k9
UNMO	12	→ PX	k_tl***X**
BIDR	1	B → BL	comp1*(kf_0***B**-kr_0***BL**)
BIMO	88	s172 → s135	c1 * g * **s173** * (**s172** - **s135**)
MM	4	MI → M	cell***MI***V1*pow(K1+**MI**,-1)
MMCAT	10	MKK → MKK_P	uVol*k3***MKKK_P***MKK**/(KK3+**MKK**)
HILL	55	→ cLm	compartment * (n1 * pow(**cXn**, a) / (pow(g1, a) + pow(**cXn**, a)))
FR	3	→ M	cell * (1 + -1 * **M**) * V1 * pow(K1 + -1 * **M** + 1, -1)
NA	160	→ timm	wholeCell * ((1 - pow(1 - **prct**, npt)) * tccctimp + pow(1 - **prct**, npt) * tcdvpmt) * **timp**

Table 1 provides an example for each certain **K type**. There are four columns: "K Type", "BIOMD" (BioModel identifier number), "Reaction" and "Kinetic law". In the column of "BIOMD", i.e. 5 means BIOMD0000000005. In the "Kinetic law" column, all the species in the kinetic law are in bold. In particular, the type FR includes generalized kinetics with the kinetic law in the format of fraction based on but beyond Michaelis-Menten and Hill kinetics.

1.2 Define reaction type (R type)

R type is defined mainly by the structure of reactions. It is a simplification of reaction stoichiometry considering the number of reactants and the number of products. For example, there are two reactants and one product in the reaction of $A+B \rightarrow C$. We define **R type** quantitatively with a pair of the number of reactants (R = 0, 1, 2, >2) and products (P = 0, 1, 2, >2). As a comparison, the stoichiometry matrix indicates reactants and products by negative and positive signs respectively.

2. Code structure

We developed a Python package called SBMLKinetics that read a collection of SBML models and then statistically analysis of each 2DK class (combinations of **K type** and **R type**). SBMLKinetics includes mainly four steps implemented by three Python scripts shown in Figure 1.

2.1 Expand the kinetic law formula if applicable

Before doing the analysis of kinetics, we need to expand the kinetic law formula if it is in the format with a function name. In the SBML files, some kinetic laws are expressed via specific function names. Therefore, the function names are supposed to get replaced by certain function bodies. For instance, sometimes the kinetic law is defined by $f(t)$. Then, it is necessary to expand $f(t)$ to its explicit equation $at+b$ based on the definition of $f(x) = ax + b$ in the SBML file.

2.2 Analysis of the kinetic laws and reactions

To classify the K types, we have considered the following features as kinetics properties (K properties).

a. the number of species in the kinetic law;

b. whether the kinetic law is a single product of terms;

c. whether the kinetic law is the difference between two products of terms;

d. whether the first (/second) product of terms with all reactants (/products);

e. whether the kinetic law is in the format of Michaelis-Menten expressions with the single reactant in the numerator;

f. whether the kinetic law is in the format of Michaelis-Menten expressions with the product of the single reactant and another species in the numerator;

g. whether the kinetic law is in the format of Hill equation;

h. whether the kinetic law is in the fraction format with at least one species in the denominator.

For example, ZERO means there are no species in the kinetic law. Therefore, the number of species in the kinetic law should be zero referring to **a**. In addition, the kinetics of UNDR is a single product of terms referring to **b** and the species in the product of terms are all reactants referring to **d**. See Table 2 for details about how K properties determine the K types .

Table 2 Matching kinetics types (K types) with kinetics properties (K properties).

K properties	ZERO	UNDR	UNMO	BIDR	BIMO	MM	MMCAT	HILL	FR
a.	$=0$	>0	>0	>0	>0	1	2	1	>0
b.		Yes	Yes						
c.				Yes	Yes				
d.		Yes	No	Yes	No				
e.						Yes			No
f.							Yes		No
g.								Yes	No
h.						Yes	Yes	Yes	Yes

To classify the R types, we have considered the number of reactants and products. **R type** is quantitatively represented by the number of reactants (R = 0, 1, 2, >2) and products (P= 0, 1, 2, >2).

2.3 Statistics analysis for different kinetics

After analyzing the kinetic laws with the strategies in step 1 and step 2, we need to do a statistics analysis to obtain the following information.

- Detailed Information for each reaction including BioModel identifier number, reaction id, reaction, kinetic law, and **K type**.
- Detailed information for each BioModel including BioModel identifier number, the number of reactions, and the distributions of ten K types per model.

2.4 Application programming interface (API) for users

Finally, users can access the results via APIs in the following perspectives.

- Query distributions of **K type** and **R type**.
- Query elements of (top) **K type** and **R type**.
- Plots for the distributions of **K type** and **R type**.

All the API details are documented and available at GitHub (https://modelengineering.github.io/kinetics_validator/).

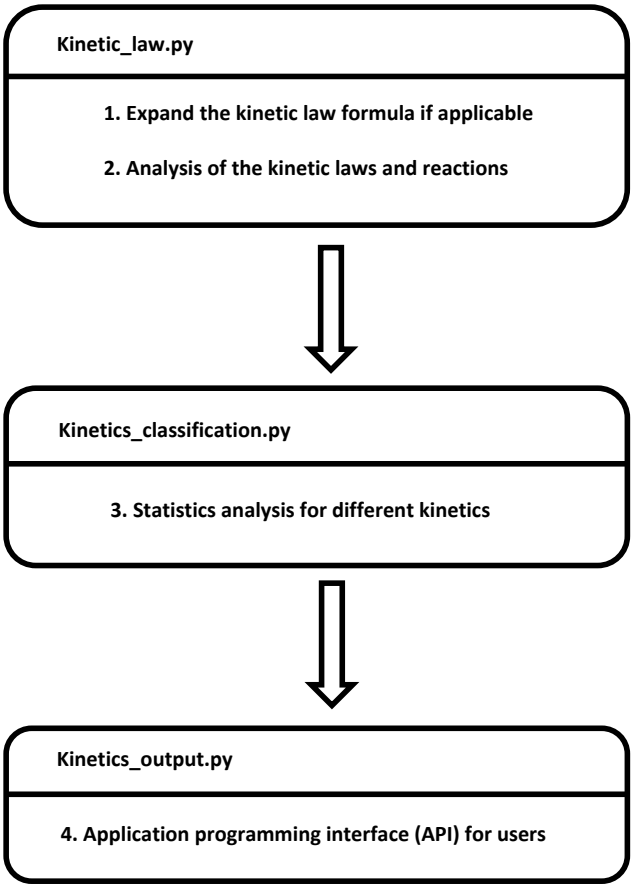

Figure 1 **Code structure and workflow steps of SBMLKinetics.**

Results

We apply our 2DK method and our tool SBMLKinetics to the BioModels Database as an example of analyzing kinetics from different reaction networks, although the

tool can be applied to any repository of SBML kinetics models. The results can illustrate the potential applications of our method and how to use our tool. In the following subsections, we have shown the **K type** distribution and **R type** distribution of all the curated Biomodels. In addition, we have indicated how the variety of **K type** distributions depends on different **R types**. Finally, we compare two subsets of BioModels including signaling networks and metabolic networks.

1. Kinetics type distribution

There are possible kinetics like zeroth order, mass action, Michaelis-Menten, Hill kinetics, and others. See the section Methods in detail. Figure 2 characterizes the kinetics used in the BioModels Database as an example to improve modeling best practices. The classification of the ten types is exclusive. The sum of all the probabilities is 1, which means we have covered all the Biomodels in the selected database from the perspective of **K type**. Figure 2 shows the distribution of K types from 931 curated BioModels with 30592 reactions. The top **K type** is UNDR, and 4.37% reactions are not classified. The running time is 9.2 hours. All computations reported were done using an Intel i7 9700 processor running at 3.00 GHz with 32GB RAM on Windows 10.

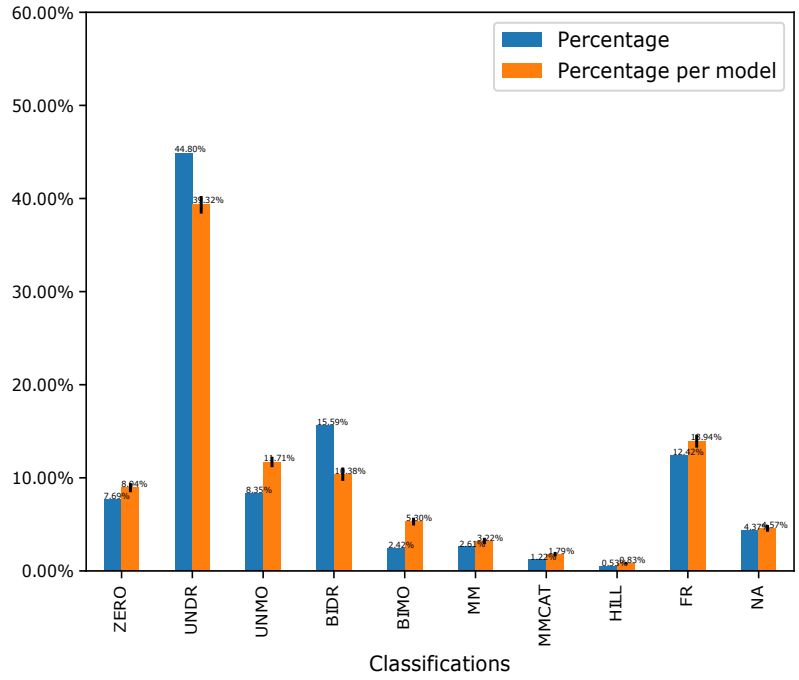

Figure 2 Kinetics type distribution. The blue bars investigate the kinetics type distribution on the average of all the reactions from all the models. The orange bars indicate the distribution on the average of reactions for each model. The trend for the two color bars is qualitatively similar. The error bars on the top of the orange bars represent the uncertainty among different models.

2. Reaction type distribution

To classify the reaction types (R types), we have considered the number of reactants and products. **R type** is quantitatively represented by the number of reactants (R = 0, 1, 2, >2) and products (P= 0, 1, 2, >2). The classification of all the reaction types is exclusive. The sum of all the probabilities is 1, which means we have covered all the curated BioModels from the perspective of **R type**. Figure 3 shows the distribution of R types from 931 curated BioModels with 30592 reactions. The top **R type** is with a single reactant and a single product.

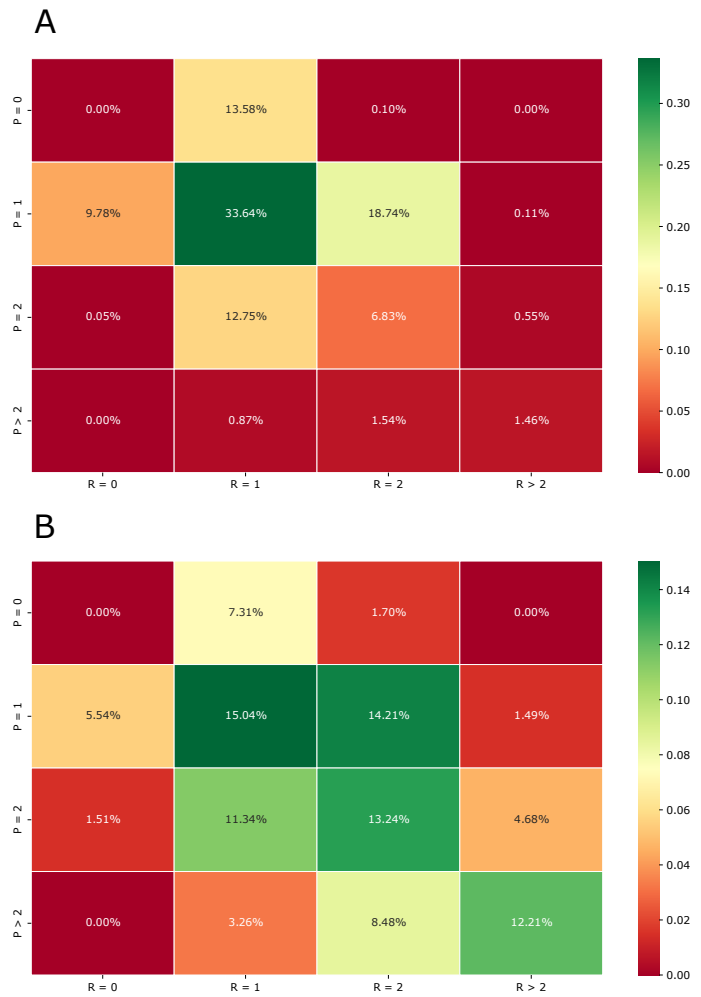

Figure 3 Reaction type distribution. (A) investigates the reaction type distribution on the average of all the reactions from all the models. (B) indicates the distribution on the average of reactions for each model. The trend for the two subplots is qualitatively similar.

3. Kinetics type distributions for different reaction types

We further investigate the distribution of kinetics based on reaction types and find substantial differences among them (Figure 4). Focusing on the four types in the center with the most reactions involved (R=1, P=1; R=2, P=1; R=1, P=2; R=2, P=2), these four subplots are quite different from each other. The top **K type** with

one reactant is UNDR significantly, while the distributions become more evenly with two reactants.

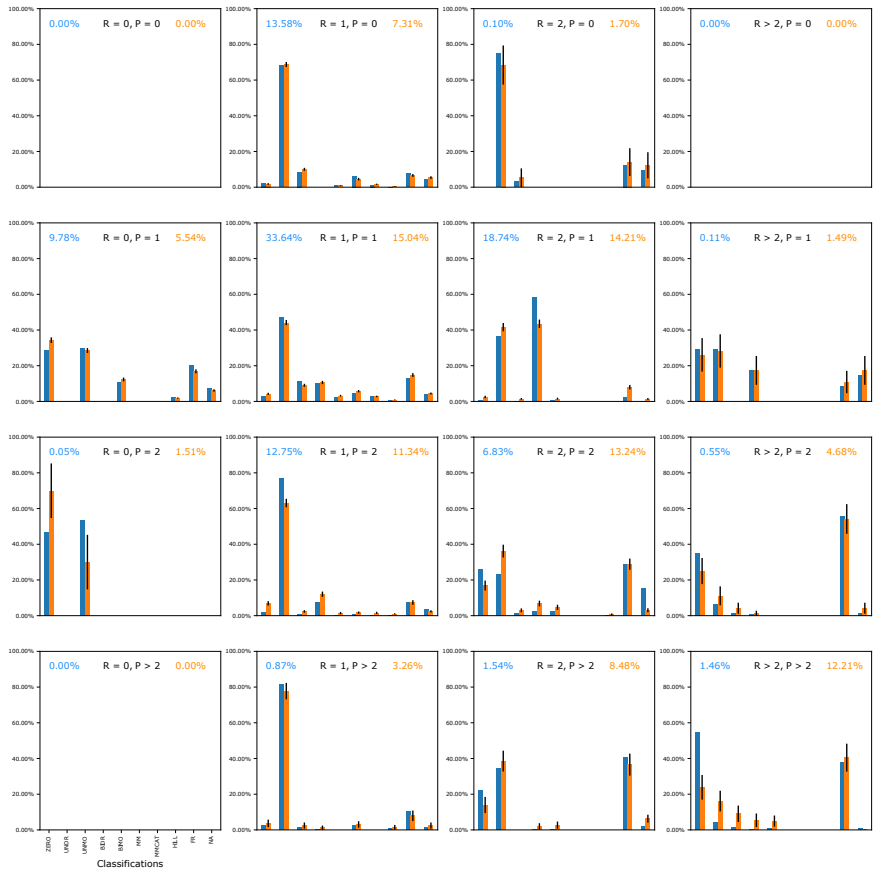

Figure 4 Kinetics type distribution for different reaction types. For each reaction type, represented by the number of reactants and products (marked as the black text in each subplot), a kinetics type distribution is given. The blue text in the top left corner represents the percentage of reactions involved in a certain reaction type. The orange text in the top right corner represents the percentage of reactions per model. The blue bars investigate the kinetics type distribution on the average of all the reactions from all the models. The orange bars indicate the distribution on the average of reactions for each model. The trend for color bars is qualitatively similar for each reaction type. The error bars on the top of the orange bars represent the uncertainty among different models.

4. Comparison between different subsets of BioModels

Our characterization considers the biology being studied as well as the nature of the reactions themselves. For instance, we compare the **K type** distributions for signaling and metabolic networks in the BioModels Database as an example. We find the substantial differences between the two types of networks in Figure 5. Figure 5A shows the distribution of K types from 288 signaling networks with 13572 reactions. The top **K type** is UNDR, and 1.84% reactions are not classified. The running time is 6.9 hours. Figure 5B shows the distribution of K types from 168 metabolic networks with 6904 reactions. The top **K type** is FR, and 9.85%

reactions are not classified. The running time is 8.1 hours. The **K type** distribution of signaling networks is similar to the whole curated BioModels in Figure 2. However, the metabolic networks are quite different with high frequencies of ZERO and FR.

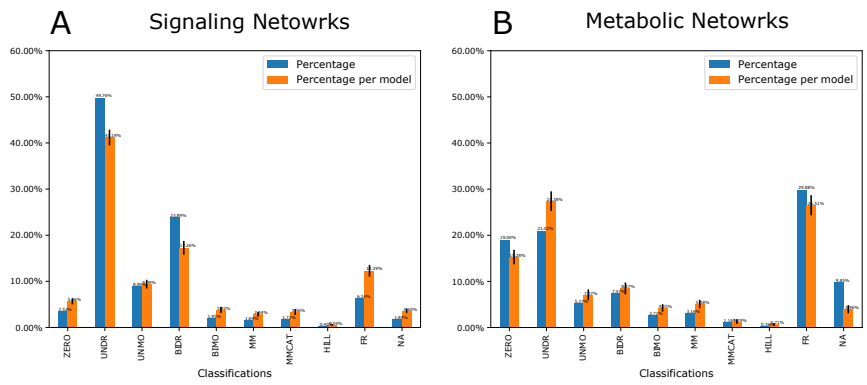

Figure 5 Kinetics type distribution for signaling networks (A) and metabolic networks (B). The blue bars investigate the kinetics type distribution on the average of all the reactions from all the models. The orange bars indicate the distribution on the average of reactions for each model. The error bars on the top of the orange bars represent the uncertainty among different models. The trend for the two color bars is qualitatively similar for each subplot. However, there is a significant difference in the kinetics type distributions between the signaling and metabolic networks.

Discussion

For each **K type** distribution in Figure 2, 4 and 5, The trend for the two color bars is qualitatively similar. The blue bars investigate the **K type** distribution on the average of all the reactions from all the models. The orange bars indicate the distribution on the average of reactions for each model. The error bars on the top of the orange bars represent the uncertainty among different models, which are relatively small. Therefore, the statistics are meaningful for the **K type** distributions. In Figure 3, the trend for the two subplots is qualitatively similar which does not depend on a certain model. Therefore, the **R type** distribution is meaningful too.

Figure 4 shows the significant differences in **K type** distributions among **R types**. The blue text in the top left corner represents the percentage of reactions involved in the certain **R type**, which is more clearly illustrated in Figure 3A. The orange text in the top right corner represents the percentage of reactions per model, which is more clearly indicated in Figure 3B. Following the observation of the blue and orange texts, the most frequent six R types are (R=1, P=0; R=1, P=1; R=2, P=1; R=1, P=2; R=2, P=2; R>2, P>2). Although there are variances between the values in blue and orange, the trend of reaction numbers involved for each **R type** is similar, which means the statistics in meaningful. It is clear to observe this point in Figure 3. The **K type** distributions are non-symmetric depending on different **R types**, which means the effects from the number of reactants and products are not equal. Based on Figure 4, researchers can tell which **R type** has the highest or lowest **K type**. For example, the **R type** with R=2, P=1 has the highest BIDR among all the R types. While there is the highest NA from the **R type** with R>2, P=1. It is also interesting to compare per column (with the same number of reactants) or

per row (with the same number of products). For instance, the third row indicates that the FR happens more frequently as the number of reactants increases with a constant number of product (P=2).

In Figure 5, there is a comparison example between different data sets. As indicated in Figure 5, the **K type** distribution of signaling networks is similar to the whole curated BioModels in Figure 2. However, the metabolic networks are quite different with high frequencies of ZERO and FR. The significant difference between signaling networks and metabolic networks could come from regulatory interactions. The high frequency of ZERO and FR from metabolic networks means the kinetics can be very simple without any species inside the kinetic law or a very complicated kinetic law that have a fractional representation.

The different **K type** distributions among different data sets, if applied to synthetic random reaction networks, can provide more representative kinetic laws for these studies. As an example to indicate how our method can contribute to generating synthetic reaction networks, we make use of the results from the BioModels Database as an example. However, users can generate a distribution based on their own database for further research. Based on the result of Figure 2 from curated BioModels, users can select UNDR with the highest probability as the general mass action kinetics to model synthetic reaction networks. Users can also follow its distribution to assign the probability for each type of kinetics while generating random synthetic reaction networks. In detail, users can generate reaction networks by assigning the probability of the certain kinetics, i.e. ZERO, UNDR, UNMO etc. If users want to model synthetic signaling networks instead of metabolic networks, they can refer to the distributions of Figure 5A instead of Figure 5B. Researchers commonly use simplified mass action kinetics as a first trial while modeling synthetic reaction networks [1], therefore users can choose a certain distribution referring to Figure 4 if they have already known the reaction type information. For example, if there are one reactant and one product in the reaction, UNDR is the most common kinetics to choose. While BIDR is more likely to get chosen for the case with two reactants and one product.

To compare with SBMLsqueezer2 [21] which is the only tool for large-scale biochemical kinetic models, we have summarized its major differences from SBMLKinetics in Table 3. SBMLsqueezer2 uses a *theory driven* classification scheme, called *de novo* creation method, that requires the annotation of all participating reactants, products and regulators. It contains a connection to the kinetics database to recommend kinetic laws, however, this is just a make up to its major *theory driven* classification scheme. In detail, their extraction of kinetic laws from SABIO-RK [24] database is based on the similarity of annotations which can not always yield results [21]. Briefly, SBMLsqueezer uses a *theory driven* of *de novo* creation method to classify kinetics that relies heavily on the information from annotations.

In summary, our work has made use of an annotation independent *data driven* classification scheme to deal with the kinetics analysis based on the SBML database, which can include all the current and future information from the SBML model database and experimental results. *Theory driven* approaches are not always correct or complete. A hypothesis can restrict the propositions deduced from phenomena, thus limiting the amount of knowledge that can be acquired from a rigorous analysis

Table 3 Comparison between SBMLKinetics and SBMLsqueezer2.

Major differences	SBMLKinetics	SBMLsqueezer2
Interface	Platform free	Various frameworks
Annotations	Independent	Dependent
Classification schemes	Data driven 2DK	Theory driven *De novo* creation
Reliability of database	Any SBML database	SABIO-RK [24]

[27]. Although *data driven* approach requires higher computational capacity, it can deal with large amounts of information with novel statistical methods.

Conclusions

SBMLKinetics can analyze any data sets with SBML files as input and generate the analysis of kinetics. Here, we use the BioModels Database as an SBML database example to show the potential applications of 2DK. Therefore, users can use their own experimental data or another model database to look for natural biological properties. See the section Availability and Implementation for detail.

Our work has application to computational studies based on random reaction networks so that more representative kinetic laws are used in these studies. Our methods can make suggestions to choose **K type** distributions depending on different models, for instance, signaling vs metabolic networks or different organisms. As we have discussed the relationship between **K type** and **R type**, this tool can suggest the probabilities of **K type** for each **R type** too. Statistically, our probability distributions of K types can be a prior for kinetic law generations, as Bayesian inference has been discussed about kinetics before [28, 29].

On the contrary, if the kinetic laws inside a reaction model violate the distributions of **K type** and **R type**, it means that there are errors in reaction models from the perspective of kinetics.

It is useful to pursue biological insights from updated models or experimental databases. For example, it is essential to compare the kinetics from a different database, i.e. signaling vs metabolic networks or different organisms. We can also observe how the **K type** distributions vary according to different **R types**.

Declarations

Acknowledgements

J.X. is grateful to Joeseph Hellerstein for introducing her to the problems considered in this paper, many stimulating discussions, and for his comments on the manuscript. J.X. appreciate Herbert M. Sauro's discussions regarding kinetics and synthetic reaction networks cited as references, and for his comments on the early stage of the manuscript. J.X. also thank Lucian Smith for the discussion about the kinetic law formula expansion based on libSBML.

Funding

J.X. who performed the research reported in this was supported by NIBIB of the National Institutes of Health under award number U24EB028887.

Abbreviations

Our two dimensional kinetics classification scheme (2DK) organizes reactions along two dimensions: kinetics type (K type) and reaction type (**R type**).

I. Kinetics type (**K type**)

K type including ten types:

- ZERO: Zeroth order;
- UNDR: Uni-directional mass action;

- UNMO: Uni-term with the moderator;
- BIDR: Bi-directional mass action;
- BIMO: Bi-terms with the moderator;
- MM: Michaelis-Menten kinetics without an explicit enzyme;
- MMCAT: Michaelis-Menten kinetics with an explicit enzyme;
- HILL: Hill kinetics;
- FR: Kinetic law in the fraction format other than MM, MMCAT or HILL;
- NA: not classified.

II. Reaction type ($\mathbf{R\ type}$)
$\mathbf{R\ type}$ is quantitatively represented by the number of reactants (R = 0, 1, 2, >2) and products (P= 0, 1, 2, >2)).

Availability of data and materials
SBMLKinetics is a publicly available Python package (https://pypi.org/project/SBMLKinetics/) and is licensed under the liberal MIT open-source license. The package is available for all major platforms. The source code has been deposited on GitHub (https://github.com/ModelEngineering/kinetics_validator). Users can install the package using the standard pip installation mechanism: pip install SBMLKinetics. The package is fully documented as https://modelengineering.github.io/kinetics_validator/.
The data sets generated and/or analyzed during the current study are available in the GitHub repository, https://github.com/ModelEngineering/kinetics_validator/tree/master/data.

Ethics approval and consent to participate
The data that we used are obtained from the public datasets BioModels Database (https://www.ebi.ac.uk/biomodels/). Therefore, the ethics approval is not applicable for our study.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Authors' contributions
J.X. implemented the project, did the data analysis and was the main developer of the Python package. J.X. wrote the manuscript.

Author details
Department of Bioengineering, University of Washington, Seattle, USA.

References
1. Riva, S., Cazzaniga, P., Nobile, M., Spolaor, S., Rundo, L., Besozzi, D., Tangherloni, A.: Smgen: A generator of synthetic models of biochemical reaction networks. Symmetry **14**, 119 (2022). doi:10.3390/sym14010119
2. Schaffter, T., Marbach, D., Floreano, D.: Genenetweaver: In silico benchmark generation and performance profiling of network inference methods. Bioinformatics (Oxford, England) **27**, 2263–70 (2011). doi:10.1093/bioinformatics/btr373
3. Shen-Orr, S., Milo, R., Mangan, S., Alon, U.: Network motifs in the transcriptional regulation network of escherichia coli. Nature genetics **31**, 64–8 (2002). doi:10.1038/ng881
4. Alon, U.: Alon, u.: Network motifs: theory and experimental approaches. nat. rev. genet. 8, 450. Nature reviews. Genetics **8**, 450–61 (2007). doi:10.1038/nrg2102
5. Ronen, M., Rosenberg, R., Shraiman, B., Alon, U.: Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics. Proceedings of the National Academy of Sciences of the United States of America **99**, 10555–60 (2002). doi:10.1073/pnas.152046799
6. Loskot, P., Atitey, K., Mihaylova, L.: Comprehensive review of models and methods for inferences in bio-chemical reaction networks. Frontiers in Genetics (2019)
7. Christopoulos, A., Lanzafame, A., Mitchelson, F.: Allosteric interactions at muscarinic cholinoceptors. Clinical and Experimental Pharmacology and Physiology **25**(3-4), 185–194 (1998). doi:10.1111/j.1440-1681.1998.t01-4-.x. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1440-1681.1998.t01-4-.x
8. Nussinov, R., Tsai, C.J., Liu, J.: Principles of allosteric interactions in cell signaling. Journal of the American Chemical Society **136** (2014). doi:10.1021/ja510028c
9. Liebermeister, W., Uhlendorf, J., Klipp, E.: Modular rate laws for enzymatic reactions: Thermodynamics, elasticities and implementation. Bioinformatics (Oxford, England) **26**, 1528–34 (2010). doi:10.1093/bioinformatics/btq141
10. Rohwer, J., Hanekom, A., Hofmeyr, J.-H.: A universal rate equation for systems biology, pp. 175–187 (2007)
11. Liebermeister, W., Klipp, E.: Bringing metabolic networks to life: Convenience rate law and thermodynamic constraints. Theoretical biology & medical modelling **3**, 41 (2006). doi:10.1186/1742-4682-3-41
12. Seibert, E., Tracy, T.: Fundamentals of enzyme kinetics. Methods in molecular biology (Clifton, N.J.) **1113**, 9–22 (2014). doi:10.1007/978-1-62703-758-7_2
13. Wagner, M.: Enzyme kinetics, behavior and analysis of rapid equilibrium and steady-state enzyme systems (segel, irwin h.). Journal of Chemical Education **53**(11), 472 (1976). doi:10.1021/ed053pA472. https://doi.org/10.1021/ed053pA472
14. Bisswanger, H.: Enzymkinetik: Theorie und Methoden. Wiley-Blackwell, German (2000)
15. Snoep, J.: The silicon cell initiative: Working towards a detailed kinetic description at the cellular level. Current opinion in biotechnology **16**, 336–43 (2005). doi:10.1016/j.copbio.2005.05.003

16. Bakker, B., Michels, P., Opperdoes, F., Westerhoff, H.: Glycolysis in bloodstream form trypanosoma brucei can be understood in terms of the kinetics of the glycolytic enzymes. The Journal of biological chemistry **272**, 3207–15 (1997). doi:10.1074/jbc.272.6.3207
17. Helfert, S., Estévez, A.M., Bakker, B., Michels, P., Clayton, C.: Roles of triosephosphate isomerase and aerobic metabolism in trypanosoma brucei. The Biochemical journal **357 Pt 1**, 117–25 (2001)
18. Lam, C.F., Priest, D.G.: Enzyme kinetics: Systematic generation of valid king-altman patterns. Biophysical Journal **12**(3), 248–256 (1972). doi:10.1016/S0006-3495(72)86084-3
19. Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., Kummer, U.: Copasi—a complex pathway simulator. Bioinformatics (Oxford, England) **22**, 3067–74 (2007). doi:10.1093/bioinformatics/btl485
20. Funahashi, A., Jouraku, A., Matsuoka, Y., Kitano, H.: Integration of cell designer and sabio-rk. In silico biology **7**, 81–90 (2007)
21. Dräger, A., Zielinski, D., Keller, R., Rall, M., Eichner, J., Palsson, B., Zell, A.: Sbmlsqueezer 2: Context-sensitive creation of kinetic equations in biochemical networks. BMC systems biology **9**, 68 (2015). doi:10.1186/s12918-015-0212-9
22. Courtot, M., Juty, N., Knüpfer, C., Waltemath, D., Zhukova, A., Dräger, A., Dumontier, M., Finney, A., Golebiewski, M., Hastings, J., Hoops, S., Keating, S., Kell, D., Kerrien, S., Lawson, J., Lister, A., Lu, J., Machne, R., Mendes, P., Le Novere, N.: Controlled vocabularies and semantics in systems biology. Molecular systems biology **7**, 543 (2011). doi:10.1038/msb.2011.77
23. Le Novere, N., Finney, A., Hucka, M., Bhalla, U., Campagne, F., Crampin, E., Halstead, M., Klipp, E., Mendes, P., Nielsen, P., Sauro, H., Shapiro, B., Snoep, J., Spence, H., Wanner, B.: Minimum information requested in the annotation of biochemical models (miriam). Nature biotechnology **23**, 1509–15 (2006). doi:10.1038/nbt1156
24. Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., Algaa, E., Weidemann, A., Sauer-Danzwith, H., Mir, S., Krebs, O., Bittkowski, M., Wetsch, E., Rojas, I., Müller, W.: Sabio-rk - database for biochemical reaction kinetics. Nucleic acids research **40**, 790–6 (2011). doi:10.1093/nar/gkr1046
25. Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., *et al.*: The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. Bioinformatics **19**(4), 524–531 (2003)
26. Jupp, S., Burdett, T., Leroy, C., Parkinson, H.E.: A new ontology lookup service at embl-ebi. In: SWAT4LS (2015)
27. Harper, W.: Isaac newton's scientific method: Turning data into evidence about gravity and cosmology. Isaac Newton's Scientific Method: Turning Data into Evidence about Gravity and Cosmology (2012). doi:10.1093/acprof:oso/9780199570409.001.0001
28. Golightly, A., Wilkinson, D.: Bayesian inference for stochastic kinetic models using a diffusion approximation. Biometrics **61**, 781–8 (2005). doi:10.1111/j.1541-0420.2005.00345.x
29. Galagali, N., Marzouk, Y.: Bayesian inference of chemical kinetic models from proposed reactions. Chemical Engineering Science **123**, 170–190 (2015). doi:10.1016/j.ces.2014.10.030