RESEARCH ARTICLE

National Science Review
XX: output, Year
doi: https://doi.org/10.1093/nsr/XXXX

Life Science

Infer metabolic velocities from moment differences of molecular weight distributions

Li Tuobang[1,*]

ABSTRACT

Metabolic pathways are fundamental maps in biochemistry that detail how molecules are transformed through various reactions. The complexity of metabolic network, where a single compound can play a part in multiple pathways, poses a challenge in inferring metabolic balance changes over time or after different treatments. Isotopic labeling experiment is the standard method to infer metabolic flux, which is currently defined as the flow of a metabolite through a given pathway over time. However, there is still no way to accurately infer the metabolic balance changes after different treatments in an experiment. This study introduces a different concept: molecular weight distribution, which is the empirical distribution of the molecular weights of all metabolites of interest. By estimating the differences of the location and scale estimates of these distributions, it becomes possible to infer the metabolic direction changes with magnitudes even without requiring knowledge of the exact chemical structures of these compounds and their related pathways. This research article provides a mathematical framing for a classic biological concept.

Keywords: Metabolism, Moments, Molecular weight distributions

Introduction

Metabolic pathways consist of enzyme-mediated biochemical reactions that are commonly categorized into two main processes within a living organism: biosynthesis (known as anabolism) and breakdown (known as catabolism) of molecules. It is common to compare the concentration changes of compounds in the same metabolic pathway in two groups of samples, i.e., up or down-regulation of a certain pathway. The definitions of these up-regulation and down-regulation are actually completely copied from the principle of chemical equilibrium shift, in which the equilibrium moves from left to right. For example, the overall equation of the urea cycle can be simplified as $2\mathbf{NH_3} + \mathbf{CO_2} + 3\mathbf{ATP} + 3\mathbf{H_2O} \rightarrow \mathbf{urea} + 2\mathbf{ADP} + 4\mathbf{Pi} + \mathbf{AMP}$. Traditionally, if the concentration of urea or ADP or Pi or AMP of samples of experimental group is higher than that of samples in the control group, and the concentration of ammonia or carbon dioxide or ATP is lower than that of samples of the control group, biochemists would say that the urea cycle is up-regulated. This definition is from the irreversible nature of this cycle and analogous to the equilibrium shift in chemistry. Since the urea cycle is a synthetic reaction, it is sometimes said that the anabolic process is dominant. On the contrary, if it is down-regulated and the catabolic process is dominant.

However, this definition is flawed. Even the comparison is within the same individuals over time, as one compound can be part of several pathways, the change in the amount of certain compounds cannot conclusively determine the metabolic direction of a specific pathway. For example, although urea is a product of the urea cycle, it can also be a product of other metabolic pathways, e.g., arginine, as a nitrogen-containing amino acid, can be converted into L-ornithine and urea under the catalysis of L-arginine amidinohydrolase. In addition, urea is also the starting material for many metabolic pathways. For example, it can be directly eliminated from the body, converted into carbon dioxide, and can also be synthesized into allophanic acid. This means that if the urea concentration of the experimental group increases, there are several possibilities, maybe the metabolic pathway from arginine to ornithine is actually up-regulated, or maybe these downstream reactions

*Corresponding authors.
Email:
tuobang@biomathematics.org.

Received: XX XX Year;
Revised: XX XX Year;
Accepted: XX XX Year

are blocked for some reason in the experimental group of samples.

In practice, it is usually necessary to manually compare the concentration changes of multiple compounds before drawing a conclusion about the changes of metabolic balance, but such a conclusion is still unclear. This article aims to introduce a different approach to quantitatively infer the metabolic directions and their associated magnitudes of metabolites of interest without requiring their exact chemical structures and related specific pathways. The concept, metabolic velocity, offers a more accessible and biologically explainable framework, with the potential to significantly advance our understanding of metabolic pathways.

Definitions of metabolic velocities

Traditionally, a synthesis reaction is defined as process in which two or more simple elements or compounds combine to form a more complex product. For a bimolecular reaction, it is often represented as $\mathbf{A}+\mathbf{B} \rightarrow \mathbf{AB}$. Suppose the molecular weights of A and B are a and b respectively. According to Lavoisier's law of conservation of mass, before the reaction there are two molecules with an average molecular mass of $\frac{a+b}{2}$, after the reaction, there is only one molecular with an average molecular mass of $a+b$. Since $a>0$ and $b>0$, $a+b>\frac{a+b}{2}$.

The above inequality reveals that, for a synthesis reaction, the key hallmark is the raise of average molecular weights. The same for decomposition reaction. Based on this principle, the article provides a precise definition of when the anabolic process is dominant and when the catabolic process is dominant.

Suppose, in a biochemical environment, based on biochemical knowledge, there are n molecules of interest that are known to be interrelated through some chemical reactions. Denoting these molecules as $\mathbf{M}_1$, $\mathbf{M}_2$,..., $\mathbf{M}_n$. Their molecule weights are M_1, M_2, ..., M_n. Their molar concentrations are $c_{\mathbf{M}_1}$, $c_{\mathbf{M}_2}$, ..., $c_{\mathbf{M}_n}$. Then, the average molecule weight of these n compounds of interest is $Mn = (c_{\mathbf{M}_1}M_1 + c_{\mathbf{M}_2}M_2 + \ldots + c_{\mathbf{M}_n}M_n) \div (c_{\mathbf{M}_1} + c_{\mathbf{M}_2} + \ldots + c_{\mathbf{M}_n})$. In the same study, denote the average molecule weight of these n molecules of interest in sample A as Mn_A, that in sample B as Mn_B. If $Mn_A < Mn_B$, it is considered that the anabolic process is dominant in sample B compared to sample A with regards to the n molecules of interest. Conversely, the catabolic process is dominant in sample B. In other words, the metabolic balance shift towards catabolism. This provides a mathematical definition for this classic biology concept.

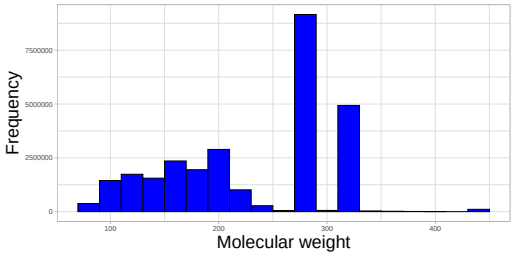

Figure 1. The molecule weight distribution of all GC-MS metabolites in human plasma. Arithmetic mean: 233.318; Hodges-Lehmann estimator: 238; sample median: 290.

Mn is essentially the sample mean of the molecule weight distribution (MWD), where the molecule weights are those of the n metabolites of interest. More generally, the location estimate of MWD of sample A is denoted as $\hat{L}_{n,A}$. The absolute difference of $\hat{L}_{n,A}$ and $\hat{L}_{n,B}$ is the magnitude of this direction change. This magnitude can be further standardized by dividing it by $\frac{1}{2}(\hat{L}_{n,A} + \hat{L}_{n,B})$. Combing this magnitude with the direction, it is called the metabolic velocity of n molecules of interest of sample A to B with regards to location.

Then, further consider a scale estimate of MWD of sample A, denoted as $\hat{S}_{n,A}$. If $\hat{S}_{n,A} > \hat{S}_{n,B}$, i.e., there is a significant decrease in the scale estimates, the metabolic direction of sample B is considered centrabolic compared to sample A for n molecules of interest. Conversely, sample A is considered duobolic compared to sample B for n molecules of interest. This mathematical approach reveals two new metabolic directions, which have clear biological significance. If the metabolic direction of a sample of n molecules of interest is centrabolic compared to that of another sample of the same n molecules of interest, it indicates that for low molecular weight compounds, the related pathways generally shift towards anabolism, while for high molecular weight compounds, the related pathways generally shift towards catabolism. This is often a typical hallmark of certain diseases or stresses (Table 2). $|\hat{S}_{n,A} - \hat{S}_{n,B}|$ is the magnitude of this change, which can be further standardized by dividing it by $\frac{1}{2}(\hat{S}_{n,A} + \hat{S}_{n,B})$. Combing this magnitude with the direction, it is called a metabolic momentum of sample A and B of n molecules of interest with regards to scale. Analogously, higher-order standardized moments of the MWD of sample A of n molecules of interest, can be denoted as $\mathbf{k}\hat{S}M_{n,A}$. However, their biological significance is much weaker. Due to the extreme heterogeneity of mass spectra data, robust statistics are recom-

mended. Here, Hodges-Lehmann estimator (H-L) [1] and Bickel-Lehmann spread (B-L) [2] are used as the location and scale estimators. The overall picture of metabolic velocities of different classes is named as velocitome (Table 2).

Applications: Metabolites Determined by Bioassays

Bioassays are classical ways to determine the concentration of metabolites.

Applications: Metabolites Determined by Mass Spectrometry

The data generated from mass spectroscopy experiments usually consist of two main components: the mass-to-charge ratio (m/z) and its corresponding intensity. The m/z value represents the mass of the ion (when the charge is +1), while the intensity of a peak in the mass spectrum is proportional to the number of ions. Let $C_{1,n}$ represent the first column, which includes the m/z data, and $C_{2,n}$ represent the second column, which includes the corresponding intensity, the molecule weight distribution of the corresponding sample could be generated by rounding the intensity into integer.

It is known that compounds in the same chemical classes are generally interrelated. So, besides metabolic pathways, chemical classes can also be used to classify metabolites of interest.

The study by Yang et al. compares the plasma metabolome of ordinary convalescent patients with antibodies (CA), convalescents with rapidly faded antibodies (CO), and healthy subjects (H) [**?**]. For both CA and CO, purine-related metabolism significantly towards anabolism and duobolism compared to the healthy volunteers (Table **??**), aligned with a previous study that showed purine metabolism is significantly up-regulated after SARS-CoV-2 infection [**?**]. Acylcarnitine-related pathways exhibit a significant inclination towards catabolism and centrabolism (Table **??**). This conclusion, which does not require knowledge of individual compounds within the acylcarnitine class, was also emphasized by Yang et al. [**?**]. It was observed that long-chain acylcarnitines were generally lower in both convalescent groups, while medium-chain acylcarnitines displayed the opposite pattern [**?**]. Bile acid-related pathways leaned towards anabolism and duobolism in CA group, while bile acids have been reported to be immunomodulatory [**?**,**?**]. Organooxygen compounds-related pathways leaned towards catabolism in both convalescent groups. The only accurately annotated compound in this class is kynurenine. This aligns with a previous study that found the kynurenine pathway, which is the primary catabolic pathway of tryptophan, is significantly up-regulated in COVID-19 patients [**?**,**?**]. For both CA and CO, metabolism related to carbohydrates significantly shifts towards anabolism and duabolism compared to that of healthy volunteers (Table **??**). This might be due to the dysregulated glucose metabolism [**?**,**?**].

Ding et al. created a comprehensive metabolome atlas for the wild-type mouse brain [**?**]. Table **??** shows the result of using hydrophilic interaction chromatography (HILIC) to seperate compounds, mainly for amines. During the aging process, in HILIC datasets, mouse brain metabolism generally shifts towards catabolism and centrabolism. This supports their conclusion that the structural degradation of brain matter becomes more pronounced in older age groups, accompanied by increased protein breakdown and elevated levels of amino acids, dipeptides, and tripeptides [**?**].

CONCLUSIONS

Since the discovery of zymase by Buchner and Rapp in 1897 [**?**] and urea cycle by Krebs and Henseleit in 1932 [**?**], a vast body of metabolic pathway knowledge has grown over the last centuries, especially aided by the development of analytical techniques such as chromatography, NMR and mass spectrometry. Metabolomics refers to the large-scale study of small molecules. High-throughput mass spectrometry experiments can collect thousands of mass spectra in just minutes, giving mass spectrometry a unique advantage compared to other analytical methods. The fragmentation pattern of a molecule, or the mass spectrum, can provide valuable structural information about the molecule. However, annotation of these spectra is typically restricted to compounds for which reference spectra are present in libraries or databases [**?**,**?**,**?**,**?**]. Only a small fraction of spectra can be accurately assigned precise chemical structures in nontargeted tandem mass spectrometry studies, a prerequisite for pathway analysis [**?**,**?**]. Moreover, many metabolic pathways are still undiscovered or poorly understood, so in practice, most metabolites cannot be annotated to the related pathways.

Recent developments of in silico methods in class assignment of nontargeted mass spectrometry data can achieve very high prediction per-

Table 1. Sideways Table Caption. For Decimal Alignment Refer Column 4 to 9 in Tabular* Preamble

	col2 head	col3 head	10	20	30	10	20	30
	col2 text	col3 text	0.7568	1.0530	1.2642	0.9919	1.3541	1.6108
		col2 text	12.5701	19.6603	25.6809	18.0689	28.4865	37.3011
3	col2 text	col3 text	0.7426	1.0393	1.2507	0.9095	1.2524	1.4958
		col3 text	12.8008	19.9620	26.0324	16.6347	26.0843	34.0765
	col2 text	col3 text	0.7285	1.0257	1.2374	0.8195	1.1407	1.3691[a]
		col3 text	13.0360	20.2690	26.3895	15.0812	23.4932	30.6060[b]

[a] First sideways table footnote. Sideways table footnote. Sideways table footnote. Sideways table footnote.

[b] Second sideways table footnote. Sideways table footnote. Sideways table footnote. Sideways table footnote.

Table 2. Sideways Table Caption. For Decimal Alignment Refer Column 4 to 9 in Tabular* Preamble

	col2 head	col3 head	10	20	30	10	20	30
	col2 text	col3 text	0.7568	1.0530	1.2642	0.9919	1.3541	1.6108
		col2 text	12.5701	19.6603	25.6809	18.0689	28.4865	37.3011
3	col2 text	col3 text	0.7426	1.0393	1.2507	0.9095	1.2524	1.4958
		col3 text	12.8008	19.9620	26.0324	16.6347	26.0843	34.0765
	col2 text	col3 text	0.7285	1.0257	1.2374	0.8195	1.1407	1.3691[a]
		col3 text	13.0360	20.2690	26.3895	15.0812	23.4932	30.6060[b]

[a] First sideways table footnote. Sideways table footnote. Sideways table footnote. Sideways table footnote.

[b] Second sideways table footnote. Sideways table footnote. Sideways table footnote. Sideways table footnote.

formance [**?**,**?**,**?**,**?**,**?**,**?**,**?**,**?**,**?**,**?**,**?**]. The classification of metabolites can be based on chemical characteristics or spectra characteristics [**?**]. While this approach can provide replicable information about the changes of metabolites in terms of chemical properties, they may not directly reflect their interactions within the cell [**?**]. Moreover, the total amount of certain classes of metabolites may remain relatively constant within groups, even if individual compounds within these classes differ.

Classical view of metabolism mainly focuses on individual reactions, so the metabolic directions are mainly static, anabolic or catabolic. This ariticle provides a statistical view for this classic biology concept and has the potential to bypass the current bottleneck and provide fresh insights into biochemistry studies.

SECTION TITLE OF FIRST APPENDIX

Use `\begin{verbatim}...\end{verbatim}` for program codes without math. Use `\begin{alltt}...\end{alltt}` for program codes with math. Based on the text provided inside the optional argument of `\begin{code}[Psecode|Listing|Box|Code|Specification | Procedure | Sourcecode | Program]` `\end{code}` tag corresponding boxed like floats are generated. Also note that `\begin{code}[Code|Listing]...\end{code}` tag with either Code or Listing text as optional argument text are set with computer modern typewriter font. All other code environments are set with normal text font. Refer below example:

Listing 1. Descriptive caption text

```
for i:=maxint to 0 do
begin
{ do nothing }
end;
Write('Case insensitive ');
WritE('Pascal keywords.');
```

Subsection title of first appendix

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna.

Subsection title of first appendix

Unnumbered figure

Subsection title of second appendix

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus.

Table 3. This is An Example of Appendix Table Showing Food Requirements of Army, Navy and Airforce

col1 head	col2 head	col3 head
col1 text	col2 text	col3 text
col1 text	col2 text	col3 text
col1 text	col2 text	col3 text

Example for an equation inside appendix

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}D_{\mu}\psi - \frac{1}{4}F^{a}_{\mu\nu}F^{a\mu\nu} - m\bar{\psi}\psi. \qquad (1)$$

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetuer at, consectetuer sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

REFERENCES

1. Hodges Jr J and Lehmann E. Estimates of location based on rank tests. *The Annals of Mathematical Statistics* 1963; **34**: 598–611.
2. Bickel PJ and Lehmann EL. Descriptive statistics for nonparametric models iv. spread 2012; 519–526.
3. Jones F and Smith F. Title title title title title title title title title title. *Journal journal journal* 1965; **37**: 126–135.
4. Jones F, Abrams F and Smith F. Title title title title title title title title title title. *Journal journal journal* 1966; **37**: 126–135.
5. Bloggs AJ. title title title title title title title title title title. *Journal journal journal* 1960; **22**: 43–154.

Algorithm 1 Pseudocode for our algorithm

```
for do each frame
    for do water particles f_i
        compute fluid flow [3]
        compute fluid–solid interaction [4]
        apply adhesion and surface tension [5]
    end for
    for do solid particles s_i
        for do neighboring water particles f_j
            compute virtual water film (see Section ??)
        end for
    end for
    for do solid particles s_i
        for do neighboring water particles f_j
            compute growth direction vector (see Section )
        end for
    end for
end for
```