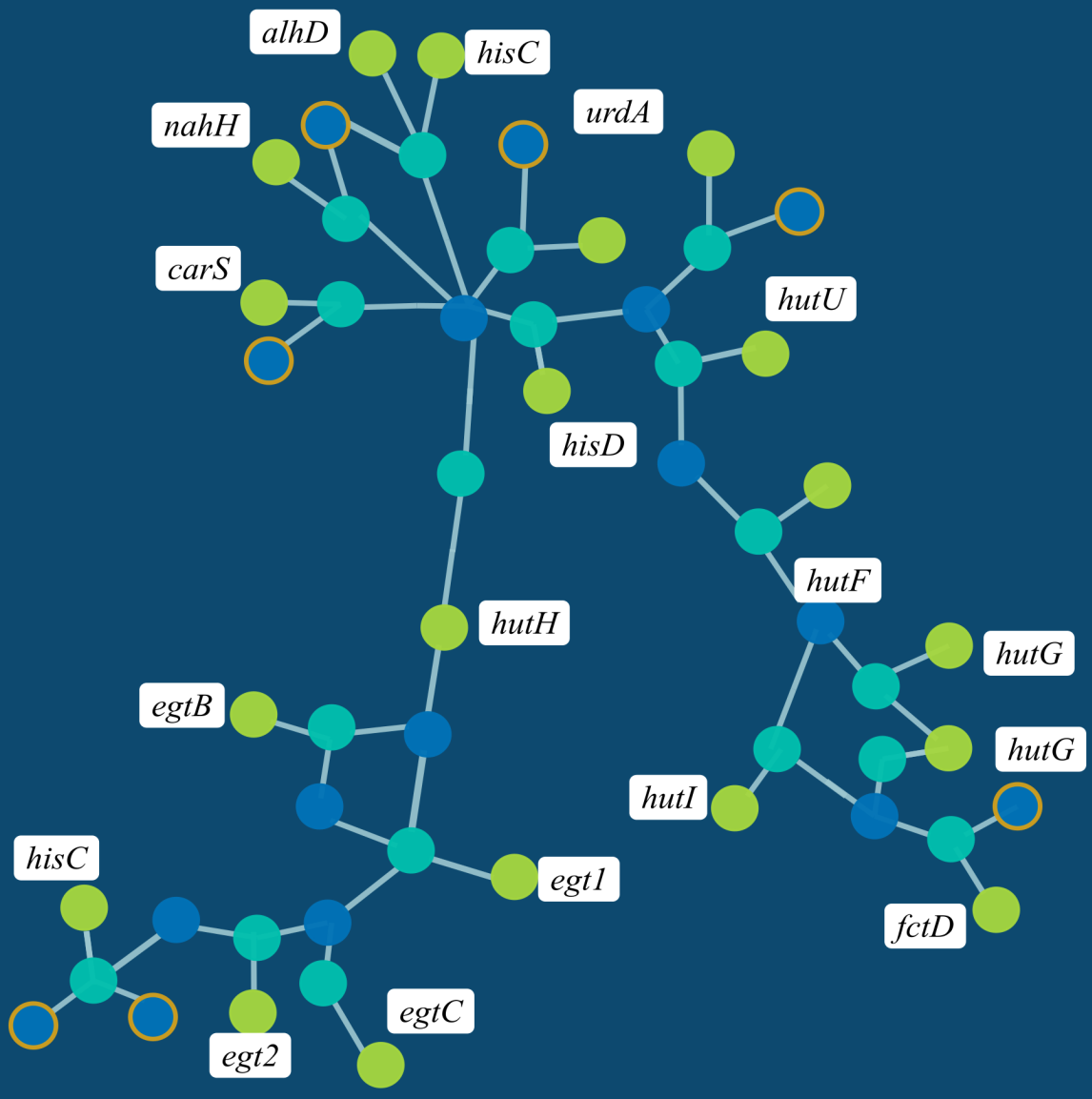
Biochemical Networks

By Simulations in R



Adonis C. Gallardo

Biochemical Networks

By Simulations in R

Adonis Cedeño

This book is available at https://leanpub.com/simbiotech

This version was published on 2025-12-14

This is a Leanpub book. Leanpub empowers authors and publishers with the Lean Publishing process. Lean Publishing is the act of publishing an in-progress ebook using lightweight tools and many iterations to get reader feedback, pivot until you have the right book and build traction once you do.

The author generated this text in part with AI. Upon generating draft language, the author reviewed, edited, and revised the language to their own liking and takes ultimate responsibility for the content of this publication.

Contents

Preface

Biological knowledge grows faster nowaday than never before. In the last decade, people from all the world has lived historical biological events: The Central Dogma Demonstration by IA, wide experimental data release, expansion of molecular techniques and recombinant DNA technologies which were once crazy dreams born in the mind of an intrepid biologist.

But this growth has come with a price: No researcher wants to spend their time, efforts and resources teaching how to use models, tools, and practical techniques. Looking for quality material which goes really deep into the subject and allows yourself to endure as a researcher has become an extremelly difficult task due to the exponential rate of publication asked for institutions and governments, which usually gives a superficial breath of knowledge or even incomplete material.

Acquiring biological knowledge has always needed a recursive process: the more you already know is the less you need to waste looking for complementary information. And this is where this book can help you. Here, we use an intense dry-lab practical approach looking to teach you how to manage data, models and resources to develop biochemical simulations. There is no better method for learning than recreating the system you are working at, and if the simulation is generated before a wet-lab experiment it becomes even better.

In order to use the upcoming material you need to have one biochemical network in your mind. Krebs cycle is a good selection. We use the *Ergothioneine Pathway* to explain fungal metabolic interactions, the GABA *receptors* for signaling and stimuli simulations in humans, and the *TomV Mosaic Infection* for plant viral gene regulation because we truly believe in generate simulations for any soul interested in this book. Be brave, go for the prize and take it all too. This book mainly uses the R language for explaining complex ideas and run calculations fast enough. Many packages are available and are asked to run the code.

Thank you very much for supporting this job.

Adonis.

Foundations on Analytical Biochemistry

Chemical Reactions

This section aims to set an standard over chemical notation which allows us to describe the mechanism behind enzyme-regulated reactions, and explore the association of this relationships to biochemical networks theory. Later sections will explain mechanisms and derivations in terms of these ideas.

Variables describing Change

Kinetics is the study of how fast reactions takes place, the factors controlling its properties and the responsible mechanisms. A chemical *reaction equation* is a representation which depicts a conversion of reactants into products.

$$R \to P$$

Figure 1. Simple Reaction Scheme

Here, we can read the equation as one molecule of reactant generates one molecule of products. We can think on any *number of molecules*, but intermediate values are prohibitive because half molecules are nonsense. We can see this variable as an ideal (theoretical) case.

Hence, lets define the **stoichiometric amount** n to be the number of molecules of any substance involved in a reaction. This is, the number of times a molecule have to be present in a reaction to be able to happen e.g. in the conversion reaction we have $n_r = 1$ and $n_p = 1$, and in a general case for a reactants and b products for species with formula S the reaction

$$\sum_{i=1}^{a} n_i S_i \rightarrow \sum_{j=1}^{b} n_j S_j$$

Figure 2. General Reaction Scheme by Species

holds. Also, let N to be a "*stoichiometric ratio*" of substance, the quantity of reactants (N_r) or products (N_p) actually taking part in a reaction. It is defined as the ratio between the mass of reactant m_r and its molar mass M_r and have its very own unit called *moles*. That's why it is often called *number of moles*.

$$N_r = \frac{m_r}{M_r}$$

Now, we can talk about using $0.5\,[g]$ of reactant, or half a pound, or $3.14[oz]$. Fractional stoichimetric amounts have a one-to-one relation with the number of molecules, so we can read "one molecule of reactant generates one molecule of product" as much as "one mole of reactant generates one mole of product" with the same meaning. For the general reaction, we can find that

$$\sum_{i=1}^{a} N_i S_i \rightarrow \sum_{j=1}^{b} N_j S_j$$

holds too. If the mass-conservation principle is applied, it results that both definitions are related by the property

$$\frac{n_r}{N_r} = \frac{n_p}{N_p}$$

Figure 3. Ratio Between theorethical and observed moles

That is, all ratios between the stoichiometric amounts and the actual amount of substance are *invariant* along and *conservative* between reactants and products. An example is always the best way to explain such an abstract but very practical idea.

> **Example** A combustion reaction occurs when an organic compound burns with oxygen and becomes carbon dioxide and water. How much mass of ethanol do you had if $4.4[g]$ of CO_2 were collected after full combustion?

Solution The reaction is stated as

$$CH_3CH_2OH + O_2 \rightarrow CO_2 + H_2O$$

But it is unbalanced because we must have more than one carbon dioxide produced from a compound with two carbon atoms. Since we have unknown stoichiometric amounts for each substance in the reaction we must use variables to get

$n_eCH_3CH_2OH + n_oO_2 \rightarrow n_cCO_2 + n_wH_2O$

Here we use stoichiometric amounts to balance the reaction until it fit mass-conservation principles. Use any balance method you like, even simple guessing, and the result must become

$CH_3CH_2OH + 3O_2 \rightarrow 2CO_2 + 3H_2O$

with $n_e = 1$, $n_o = 3$, $n_c = 2$ and $n_w = 3$

Now, and since the invariant property could be applied, for this reaction we actually know that 2 amounts of CO_2 are obtained from 1 amount of ethanol. That must be writted as

$\frac{n_e}{N_e} = \frac{n_c}{N_c}$

$N_e = \frac{n_e}{n_c}N_c$

Looking at the periodic table we can establish $M_c = 44[\frac{g}{mol}]$, and since we initially had $m_c = 4.4[g]$ it follows than $N_c = \frac{4.4}{44} = 0.1[mol]$ thus

$N_e = \frac{1}{2}(0.1) = 0.05[mol]$

To compare the big picture, we have used a theoretical reaction as a template, and established the numbers we must fix for our particular case. Our particular reaction state is described as

$$(0.05)CH_3CH_2OH + (0.15)O_2 \rightarrow (0.1)CO_2 + (0.15)H_2O$$

This is the reason of being invariant along the reaction.They are conservative because you can compare any pair or products and reactants and the conclusions must be the same. Finally, and since we are interested in the mass of ethanol, we must derive it from the prior equation and the fractional stoichimetric amount definition

$$m_e = \frac{n_e M_e}{n_c M_c} m_c$$

Or just calculate $M_e = 46.7[\frac{g}{mol}]$ and apply

$$N_e = \frac{m_e}{M_e}$$

$$m_e = N_e M_e$$

$$m_e = 2.34[g] \blacksquare$$

We have established that the stoichiometric amount, as a variable, is defined in a *discrete* form thinking only in integer values, but it can be extended to behave like fractions. The number of moles, by itself, is defined to allow decimals since it is a proportion. This tricky game of using a discrete but obvious variable to define another one continuous but obscure is called *continuization* and it is recalled several times along the book. The inverse is known as discretization.

Reaction yield

When a reaction occurs, some amount of the reactants can avoid the transformation and we get a final mixture of products and reactants, which can be characterized by using fractions. Again, the discrete form of the product yield is easy to relate.

$$y_p = \frac{\text{number of molecules in products}}{\text{number of molecules in reactants}}$$

Figure 4. Yield Discrete Definition

That is, yield is a ratio between the amount of product generated over the amount of initial reactant. If we can have 8 product molecules to be generated from 10 reactant molecules, but only 6 are achieved then the yield is $y_p = 0.6$ instead of a *maximum* at $y_p = 0.8$. This ratio also can be expressed as a percentage but we prefer to keep decimals.

The most simple continuous extension, or continuization, of a *reaction yield* is stated as a function:

$$y(m_p) = \frac{m_p}{m_r}$$

Figure 5. Yield Continuous Definition

Again, if we have 8 grams of product to be generated from 10 grams of reactant, but only $6[g]$ are achieved, then $y(6) = 0.6$. The yield of a reaction can be defined in two ways: ideal (y) and real $(\hat{y})$. The first one is calculated supposing maximal consumption of the reactant, and the highest one is usually called *limit reactant yield*. The later is measured by experimental approaches (yay empirism!). The maximal yield of 1 is theoretical, as it cannot be reached in practice, but is a good reference state for an upper limit of the mass conversion process $0 \leq \text{y} \leq 1$.

For two or more products, each species is assumed to have a fraction of the total yield, called a *partial* yield, and that the sum of every partial is always the total yield. Denominators are always the sum of reactants.

$\sum_i y_{p_i} = y_p$

Example: Valproic acid is initially assimilated into the body by the liver metabolism at a maximum of 15%.

a) How much active principle does a pacient with an intake of $250[mg]$ have? b) How much is available if is $y = 0.94$ c) Calculate yields if plasma can retain up to $87[mg]$ of valproic acid.

Solution

a) The reaction is a decomposition of the intake reactant R as two forms, one active principle A which goes to the brain and the other one metabolized by the liver L.

$R \rightarrow A + L$

As we work with a maximum reaction yield $y = 1$ we can establish than

$y_a + y_l = 1$

And since $y_l = 0.15$ we have $y_a = 0.85$, thus

$y_a = \dfrac{m_a}{m_r}$

$m_a = y_a.m_r$

$m_a = 212.5[mg]$■

b) As the total yield is now $y = 0.94$ we have that a certain amount of valproic acid remains unchanged after reaction, then

$y_a + y_l = 0.94$

$y_a = 0.79$

$m_a = y_a m_r$

$m_a = 197.5[mg]$

c) Since the measurement represents the plasma concentration of active principle:

$m_a = 87$, $m_r = 250$

$y_a = \frac{87}{250}$

$y_a = 0.35$

Assuming $y_l = 0.15$ still, we have

$y = 0.5$

We want the reader to think about the denominator of the yield function. Since m_r is the initial amount of reactant, the reaction step *splits it* into two forms: a free/unchanged m_{r1} and the amount transformed $m_{r2} = m_p$, then for a one:one stoichiometry the yield function could be seen as

$$y(m_p) = \frac{m_p}{m_{r_1} + m_p}$$

This form is a consequence of applying the mass-conservation principle, and has a lot of useful corolaries in chemistry. Note the amount m_{r1} must be a constant for the reaction.

Until now, we have defined what we need to start a reaction and what to measure once it is over. Now, we are going to focus on how to track and forecast the happening of a reaction.

Concentration of chemical species

One of the advantages of the number of moles is that could be combined with other kind of measurements to describe *concentration* of the species inside a region or a space. There are many concentration definitions, but chemists are very used to *molar concentrations* $[\frac{mol}{L}]$. For the reactant formula R we define the concentration variable with lowercase as

$$r = \frac{N_r}{V}$$

Figure 6. Molar Concentration

Now the three quantities described until now relates how much of an species mass (m) do we have, and all they are equivalents under proper conditions. The main difference is on units i.e molecules $[1]$ against grams $[g]$ against moles $[mol]$ against moles by liter $[M]$ or molarity.

But the last one is the standard unit of the entire book. Note that we completely missuses notations like C_r, $c(R)$ or $[R]$ to explain the concentration of a reactant. We want you to get used to express the capital letter R to account for the species inside reactions to replace formulas and the lowercase r for an specific concentration value, a variable, in order to simplify upcoming equations.

Example Hydrolysis is the name for the reaction

$$2H_2O \rightarrow 2H_2 + O_2$$

a) How much water do we need in order to produce $75[g]$ of oxygen gas? b) We need to supply an oxygen cylinder of $40[l]$ with $3000[g]$ of pure oxygen. What is the molar concentration of oxygen in the cylinder?

Solution

To follow notation we rewrite the reaction as

$2R \rightarrow 2S + P$

Thus the problem is to find a certain value m_r for which $m_p = 75$. Using stoichiometric amounts we can see that 2 amounts of R are needed to generate one amount of P. Thus

$N_r = 2N_{p}$

and since mass is related to stoichiometric amount by the molar mass we have

$\frac{m_r}{M_r} = 2\frac{m_p}{M_p}$

$m_r = 2\frac{M_r}{M_p}m_p$

Look at your *periodic table* to find that $M_r = 18[\frac{g}{mol}]$ and $M_p = 32[\frac{g}{mol}]$ so finally we have

$m_r = 84.4[g]$■

This means you ceiling need up to $85[g]$ of water to obtain at least $75[g]$ of oxygen with full theoretical yield.

b)

$p = \frac{N_p}{V} = \frac{m_p}{M_p V}$

$p = \frac{3000}{32(40)}$

$p = 2.35\ [\frac{mol}{l}]$

$p = 2.35[M]$■

These terms can be used for any chemical species like atoms, molecules, inorganics, organics, polymers, etc. Concentration can also be associated with time of reaction.

Concentration Rate of Change

When a reactant is contained inside a flask filling a constant volume V with an initial concentration and a set of reactions applies transformations over time, we can measure the instantaneous rate of change of the concentration e.g. $[M/s]$, as

$$\dot{r} = \frac{\mathrm{d}r}{\mathrm{d}t}$$

Figure 7. Rate of Change Over Time for Concentration

This change describes the consumption rate of R, and we can also measure the formation rate of P in such a similar way. That's why it is called rate of change, being consumption or formation is not considered, yet.

We just need to find an algebraic expression for the concentration of reactant as a function of time, and differentiation allows us to get its rate of change. However, in biochemical modelling, we usually do the opposite: we

found an expression for the rate of change and integrate an expression for the concentration which fits initial values conditions.

Stoichiometric coefficients

This variable measures the *relative* amount of a substance produced or consumed by a reaction. Lets define this coefficient as

$$c_a = n(A)_p - n(A)_r$$

Figure 8. Stoichiometric Coefficients definition

You must read the equation as the difference between stoichiometric amounts of the species A as a product and as a reactant. Thus, it assigns a *sense of direction* for the species. Positive means a product, and negative means it is a reactant.

Example Find the stoichiometric coefficients in the reaction: $3A + 4B \rightarrow 5X + A$

Solution

We have $c_a = 1 - 3 = -2$, $c_b = -4$ and $c_x = 5$■

Reaction rate

The rate (v) of a chemical reaction is described as the ratio between rates of change per stoichiometry coefficients. A generalization for "a$" reactants and "b$" products in the reaction

$$\sum_{i=1}^{a} n_i R_i \rightleftharpoons \sum_{j=1}^{b} n_j P_j$$

becomes

$$v = \frac{1}{c_i}\frac{\partial r_i}{\partial t} = \frac{1}{c_j}\frac{\partial p_j}{\partial t}$$

or more easily rewrited as

$$\dot{r_i} = c_i v \quad \dot{p_j} = c_j v$$

Figure 9. Relation betweenn Rates of Change and Reaction Rate

Example The ergothioneine catabolism could be represented as $R \rightarrow P_1 + P_2 + P_3 + P_4 + P_5$ Where the products are l-glutamate, trimethylamine, hydrogen sulfide, carbon dioxide, and ammonia. a) Find an expression for ergothioneine rate of change if $r(t) = sinh(t) - (t^3 + 2t^2)$. b) Find an expression for the rate of change of every species, if the reaction rate is $v = k\frac{r}{S + r}$ where $S = p_1 + p_2 + p_3 + p_4 + p_5$

Solution

a)

As the rate of change is the time derivative of r it becomes

$\dot{r}(t) = cosh(t) - (3t^2 + 4t)$■

b)

Since the reaction rate is known, we only need to establish the stoichiometric coefficients for products and reactants. Here we demonstrate only for l-glutamate:

$\dot{r} = -v$

$\dot{p_1} = v$■

Be aware of clearly distinguish between the *rate of change* of a species and the *reaction rate*. The reaction rate (v) is an *intensive property* since it is independent on the amount or the species we use to measure it.

As a consequence we must define a change for volume V, as reactions between organelles consider the same amount being passed through different size compartments. This is, if we are going from organelle X to Y, and we have to transport reactants and products across membranes, the volumes V_x and V_y let us consider that $r = \frac{n_r}{V_x}$ and $p = \frac{n_p}{V_x}$ and thus

$$v = \frac{1}{V_y c_j}\frac{\partial N_{pi}}{\partial t} = \frac{1}{V_x c_i}\frac{\partial N_{ri}}{\partial t}$$

Figure 10. Reaction Rate Between Compartments

Rate Kinetics

As the reaction rate is proportional to a time-derivative of the reactants, we must be able to express it as a function of the reactants concentration over time, called *profile* $r(t)$. Empirical studies suggest we can always define the rate of reaction as a power function of a variable ε, called order-variable, and the concentration of the reactant, like

$$v(r) = kr^{\varepsilon}$$

Figure 11. Mass-action Rate Law

This modeling approach is the elementary *mass-action* rate law, where the rate constant k has adaptable units and the order variable could take values in a similar way stoichiometric amount does.

If $\varepsilon = 0$ the reaction is zero-order and the profile becomes $r(t) = k(t - t_0) + r_0$ as integration dictates. If $\varepsilon = 1$ the reaction is first-order and the profile $r(t) = r_0 e^{k(t-t_0)}$ applies, and so on. For the products and reactants generalization scheme we can define the rate in order to account for irreversible reactions as

$$v = k\prod_{i}^{a} r_i^{\varepsilon_i}$$

Figure 12. Irreversible Generalization of The Mass-Action Rate Law

And for a reversible mass-action over the rate law equation we write the difference between the formation rate and the decomposition rate like substance gains minus losses, to generate a net reaction rate which becomes

$$v = k_1\prod_{i}^{a} r_i^{\varepsilon_i} - k_2\prod_{j}^{b} p_j^{\varepsilon_j}$$

Figure 13. Reversible Generalization of The Mass-Action Rate Law

As you can see, this approach allows us to model the kinetics profile of a reaction using a proposed mechanism and ordinary differential equations (ODEs). A big drawback of this model falls in that we must know the full reaction mechanism as a list of *elementary reactions* to set a correct reaction rate in terms of stoichiometric amounts; this could be viewed as a pathway knowledge requirement. This also means any *catalytic effect* is prohibitive, even autocatalysis. Otherwise, non-elementary rate laws appears and must be postulated. This is the case of enzyme kinetics.

Chemical equilibria and Steady State

All reactions are reversible, even if it takes too much time to notice it, and every reversible reaction has a main direction since we are usually interested in generate products over reactants. If we split both consumption and formation reaction in one single equation

$R \leftrightharpoons P$

We can set the net reaction rate as the difference between outlet and inlet rates, or forward and reverse reactions, as

$v = v_{out} - v_{in}$

where $v_{out} = k_{out} r$ and $v_{in} = k_{in} p$, saying we are mainly going from reactants to products. Thus, the *chemical equilibrium* is a description of the final state of this system. We assume both product and reactant are at equilibrium if:

$$v = 0$$

which intrinsically defines an equilibrium constant as a description of all relative amounts of products and reactants at this state.

$K_{eq} = \frac{k_{out}}{k_{in}} = \frac{p_{eq}}{r_{eq}}$

A generalization applies in the form

$$K_{eq} = \frac{\prod_j^b p_{j(eq)}^{\varepsilon_j}}{\prod_i^a r_{i(eq)}^{\varepsilon_i}}$$

Figure 14. Generalized Equilibrium Constant

Where each concentration accounts for a limit at equilibrium. In chemistry, this is called thermodynamic equilibrium and any system which reaches this state is actively interconverting reactants and products. In biology, however, any system which reaches thermodynamic equilibrium is called a *dead system* since the main goal of an organism is to generate active responses to enviromental changes. Organisms have adapted their own, and very similar, pseudoequilibrium state where intermediate concentrations no longer changes but reactions keeps transforming the source into a product. This is called a *steady-state*, and the usual way to define it needs to establish chemical equilibrium between *intermediate reactions*.

Remember, biological systems are *open* since mass and energy exchange are allowed. This condition means we must look for a kinetic state similar to thermodynamical equilibrium but not equal since we need to transform our source reactants into sink nutrients without any chance of source regeneration.

Mass-action and the Disequilibrium ratio

Reactions in closed systems tends to equilibrium, but reactions in living cells are usually far from this state in the wild. The Γ ratio, or mass-action ratio, is a measurement of how far a system is from equilibrium inspired from the equilibrium constant definition. If our reaction is in a state with product concentration p and reactant concentration r we have

$$\Gamma = \frac{\prod_j^b p_j^{\varepsilon_j}}{\prod_i^a r_i^{\varepsilon_i}}$$

Figure 15. Gamma Ratio

Where r_i and p_j are the *in-vivo* concentration of each species. Thus, we can also define the *disequilibrium ratio* $\rho \in [0,1]$ as

$$\rho = \frac{\Gamma}{K_{eq}}$$

Figure 16. Disequilibrium Ratio

as an scaled variable which let us define the effect of the change from an reversible reaction to an irreversible one as a fraction since the net rate becomes

$v = (1 - \rho) k_{out} \prod_i^a r_i^{\varepsilon_i}$

This idea simplifies to write reversible rates of reaction. The ρ variable takes values from zero to one, being next to one at states near from equilibrium. This is one of the main inspiration for the work to be presented in the next chapters: using a fractional quantity to model a complex reversible phenomenon.

Catalysis effect over the neat rate

A catalyst is a substance which applies a certain mechanism in order to facilitate the ocurrence of a reaction while its concentration remains constant changing only the net rate. Enzymes E are highly specific proteic catalysts which degree of acceleration on the reaction depends solely of the catalyst concentration e_0 with, the simplest way, only a linear effect as

$$v = e_0 v_0$$

Figure 17. Linear Rate Effect of Enzymes as Catalyst

In this way we can describe the posterior net-rate as a product of the catalyst concentration and the prior net rate v_0. Also, when we work with enzymes, the main reactant is called *substrate* $(R = S)$, is accounted as a species too and all the catalytic steps are assumed to occur while being attached or exposed to the enzyme.

Matricial methods over Chemical Reactions

Matrices are awesome. Their ability to simplify the behaviour of a system in one equation is part of a handy toolbox to solve high-dimensional problems. Here, we rewrite the prior section with matricial notation and show the use of R functions to solve these kind of problems.

Stoichiometry Balance

We can propose many ways to represent derivations of the generalized equation by species in multivariable notation. Here we show a dimensional array which must solve any stoichiometric balance, but other arrays can be stated thinking over the molecularity of substrates and products as a basis for dimensional analysis.

Single reaction

A simple way to define a vector for stoichimetric amounts using a reactants and b products as species can be derived from the mass-conservation principle keeping the same volume. This allows us to convert the general reaction

$$\sum_i^a n_i R_i \rightleftharpoons \sum_j^b n_j P_j$$

As the sum of the products between stoichiometric amounts n_k and the number of elements α_k inside the formula, as defined by

$$\sum_j^b n_j \alpha_j - \sum_i^a n_i \alpha_i = 0$$

which holds for every atom present in the compounds. This equation must be read as "There exists a set of numbers n_k$ for which the amounts of elements between reactants and products keeps constant". Using this method usually generates infinite solutions, but all of them follows a determined equation.

Example Balance the following reaction

$H_2S_2O_3 + Ag(OH) \rightarrow Ag_2(S_2O_3) + H_2O$

Solution

First, we rewrite the reaction to follow the notation proposed.

$S_1 + S_2 \rightarrow S_3 + S_4$

In this reaction we have $a = b = 2$, $S_1 = H_2S_2O_3$, $\alpha_1 = Ag$ and so on. Therefore, it must exists at least one vector of numbers n_k for which the relation

$n_1 S_1 + n_2 S_2 \rightarrow n_3 S_3 + n_4 S_4$

holds for every atom. Even if simple guess could tell us this vector of stoichimetric amounts is $\{1, 2, 1, 2\}$, we can establish algebraic equations between the elements of every molecule to compare a systemic equation which generalize the method.

Element	Equation
{Ag}$	n_2 = 2n_3$
{S}$	2n_1 = 2n_3$
{O}$	3n_1 + n_2 = 3n_3 + n_4$
{H}$	2n_1 + n_2 = 2n_4$

These equations can be manually operated and reduced onto the set

$$\begin{cases} n_1 = n_3 \\ n_2 = 2n_3 \\ n_4 = 2n_3 \end{cases}$$

which relates every amount to those of the silver compound. Now, we just need to define n_3 and all other amounts must keep the relation. These replacements remains true for fractional stoichiometric amounts too. If $n_3 = 1$ it returns the guess vector. If $N_3 = 0.174$ we have the vector $\{0.174, 0.348, 0.174, 0.348\}$ which fits too.

This problem can be easily stated in matrix notation, and thus solved in one step. Let **A** to be a matrix where the columns represents the chemical species and the rows follows a certain atomic element. The first set of equations would be separated as the null matrix equation $\mathbf{A}\vec{n} = \mathbf{0}$ like

$$\mathbf{A} = \begin{pmatrix} 0 & -1 & 2 & 0 \\ -2 & 0 & 2 & 0 \\ -3 & -1 & 3 & 1 \\ -2 & -1 & 0 & 2 \end{pmatrix}; \quad \vec{n} = \begin{pmatrix} n_1 \\ n_2 \\ n_3 \\ n_4 \end{pmatrix}$$

To define such a solution to the null equation is to find any input $\vec{x} \neq \mathbf{0}$ for which the equation $\mathbf{A}\vec{x} = \mathbf{0}$ is valid, and then use any scalar value k for which $\vec{n} = k\vec{x}$. We find a non-trivial solution, or define the kernel of the matricial transformation, and express the stoichiometric amounts as a linear combination of it. In R, we can do so by using the `null()` function from the pracma package:

```
amounts <- c(0,-2,-3,-2,-1,0,-1,-1,2,2,3,0,0,0,1,2)
A <- matrix(amounts,4,4)
x <- pracma::null(A)
```

$$\vec{x} = \begin{pmatrix} 0.316 \\ 0.632 \\ 0.316 \\ 0.632 \end{pmatrix}_{\blacksquare}$$

This vector, called *normalized kernel*, is a generalization which contains every posible version of the stoichiometric amounts vector $\vec{n}$. Now, any solution which balance the reaction is also a linear combination of the kernel, which characterizes all the infinite solutions for the system. Be aware that any combination duplicates $Ag(OH)$ and `H_2O$ while comparing to the other compounds. *Suggestion*: Try` k=3.16‘$.

However, note that the **A** matrix must define only positive relations since it is equivalent to elemental stoichiometic amounts. Therefore, we should re-arrange the system in order the $\vec{n}$ vector keep the negative signs.

```
amounts <- c(0,2,3,2,1,0,1,1,2,2,3,0,0,0,1,2)
A <- matrix(amounts,4,4)
x <- pracma::null(A)
```

$$\vec{x} = \begin{pmatrix} -0.316 \\ -0.632 \\ 0.316 \\ 0.632 \end{pmatrix}_{\blacksquare}$$

This vector contains stoichiometric coefficients. Now, the first two entries represent reactants and all the others are products. There exists other methods for balance of chemical reactions, but we keep the algebraic method because is one of the best ways to analize reactions as a whole and it could be solved easily by computers.

Set of reactions

We'd like to introduce a matricial method for two or more reactions like a very simple network example. Before, we had implicitely applied charge of

the elements in formulas to do balance. Consider a matrix $\mathbf{A_{m\times p}}$ describing m atomic composition numbers in the rows, and $p = a + b$ compounds associated to these elemental composition. That is

$$\mathbf{A} = [a_{ij}]_{m\times p}$$

Note that this definition avoids to use signs to assign reactant or product nature. Now, if we also consider any generalized set with $q > 1$ ordered reactions between $p = a + b$ substances then this system can be described using an stoichiometric system of equations. We call this array a *stoichiometric matrix* $\mathbf{N}_{p\times q}$ with the form

$$\mathbf{N} = [c_{ij}]_{p\times 1}$$

Where $c_{ij} = n_{pij} - n_{rij}$ are stoichiometric coefficients for the reaction. Therefore, an approach to analyze the stoichiometric relationship between the elements of a system is to define an *atomic matrix* $\mathbf{A}_{m\times p}$, an associated *stoichiometric matrix* with dimensions $\mathbf{N}_{p\times q}$, and verify the product

$$\mathbf{AN} = \mathbf{0}_{m\times q}$$

holds for the system as a whole, with m elements and q reactions.

Example Consider the following pair of reactions which takes place in the same vessel

$$\begin{cases} CH_4 + O_2 \rightarrow CO_2 + H_2O \\ CO_2 + H_2 \rightarrow CH_4 + H_2O \end{cases}$$

Balance the system.

Solution

The first step is to set stoichiometric amounts for every species in the system

$$\begin{cases} n_m CH_4 + n_o O_2 \rightarrow n_d CO_2 + n_w H_2O \\ n_d CO_2 + n_h H_2 \rightarrow n_m CH_4 + n_w H_2O \end{cases}$$

A simple guessing solution is

$$\begin{cases} CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O \\ CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O \end{cases}$$

Now, let's apply the matricial method. The matrix **A** is defined for the elements C, H, O and the compounds CH_4, O_2, CO_2, H_2O and H_2 as

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 4 & 0 & 0 & 2 & 2 \\ 0 & 2 & 2 & 1 & 0 \end{pmatrix}$$

A similar formulation is applied to define **N** as

$$\mathbf{N} = \begin{pmatrix} -n_m & n_m \\ -n_o & 0 \\ n_d & -n_d \\ n_w & n_w \\ 0 & n_h \end{pmatrix}$$

And calculating the kernel of the **A** matrix we have:

```
coefs <- c(1,4,0,0,0,2,1,0,2,0,2,1,0,2,0)
A <- matrix(coefs,3,5)
X <- pracma::null(A)
X
```

$$\mathbf{X} = \begin{pmatrix} -0.218 & -0.345 \\ -0.5727 & -0.282 \\ 0.218 & 0.345 \\ 0.709 & -0.127 \\ -0.272 & 0.817 \end{pmatrix}_{\blacksquare}$$

Where every j-column in the $\mathbf{X} = [x_{i1} \quad x_{i2}]$ matrix is called a *mode* of the system of reactions. As before, these vectors represents linear combinations for all the possible answers which balance this system as $\mathbf{N} = \mathbf{X}\kappa$ which reduces to both guessed vector for the kappa matrix:

$$\kappa = \begin{pmatrix} 3 & 2.07 \\ 1 & -4.21 \end{pmatrix}_{\blacksquare}$$

```
k <- matrix(c(3,1,2.07,-4.21),2,2)
X %*% k # such a good aprox
```

This approach, however, needs to define two matrices **A** and **N** while we can extract the same information only from the stoichiometric matrix. To solve the matrix equation $\mathbf{AN} = \mathbf{0}$ we can also invert the way operations get done by transposing to $\mathbf{N}^T\mathbf{A}^T = \mathbf{0}$ and now we need to find the kernel of the transposed stoichiometric matrix to find the transpose of the *mass conservation modes*.

```
coefs <- c(-1,-2,1,2,0,1,0,-1,2,-4)
N <- matrix(matrix(coefs,5,2))
A <- pracma::null(t(N))
t(A)
```

And now these columns vectors are linear combinations of the elements used to generate chemical species. Try to read the combinations which must be followed in order to balance the system.

Mode analysis

The prior system is quite interesting because postulates all stoichiometric amounts inside a set of reactions are related between them by the *null kernel* for its implicit atomic amounts matrix $\mathbf{AN} = \mathbf{0}$. Finding the kernel of a matrix can be seen as a lineal transformation, or a mapping, and a very clevr question from this analysis could be to ask for the effect of the kernel of the *stoichiometic matrix*.

This is, once we know the stoichiometric relations n_{ij} in our system, but we are interested in their effect over species concentration s_{ij} we can look to solve the system

$\mathbf{NS} = 0$

By finding the $\mathbf{X} = ker(N)$ matrix, where each `X = [x_{ij}]$` are relative ratios, weights within *linear combinations*, or *modes*, allowing us to get no effect on the system. This *redundancy* shows *non-trivial* solutions when there are more reactions than species, that is, when there exists loops from a practical point of view and chemical equilibria is reached.

Once we have found the balance relations between species, we can use the modes to show trends inside the set of reactions.

Example: Find a stoichiometry solution for the following system of coupled reactions

$$\begin{cases} CH_3CH_2OH + 5O_2 \rightarrow 3CO_2 + 4H_2O \\ CH_3OH + 5O_2 \rightarrow 3CO_2 + 4H_2O \\ 4(CH_3)_2CH_2OH + 19O_2 \rightarrow 12CO_2 + 18H_2O \\ 2CH_3OH \rightarrow CH_3CH_2OH + H_2O \end{cases}$$

Solution

First, let's rewrite the system as

$$\begin{cases} S_1 + 5S_2 \rightarrow 3S_3 + 4S_4 \\ S_5 + 5S_2 \rightarrow 3S_3 + 4S_4 \\ 4S_6 + 19S_2 \rightarrow 12S_3 + 18S_4 \\ 2S_5 \rightarrow S_1 + S_4 \end{cases}$$

The substance matrix $\mathbf{S}_{6\times 1}$ takes the form

$$\mathbf{S} = \begin{pmatrix} s_1 \\ s_5 \\ s_6 \\ s_2 \\ s_3 \\ s_4 \end{pmatrix}$$

Remember we use lowercase to denote molar concentration in a fixed volume. Also, the stoichiometric matrix $\mathbf{N}_{4\times 6}$ is defined as:

$$\mathbf{N} = \begin{pmatrix} -1 & 0 & 0 & -5 & 3 & 4 \\ 0 & -1 & 0 & -5 & 3 & 4 \\ 0 & 0 & -4 & -19 & 12 & 18 \\ 1 & -2 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Also remember the conventional way to define the **N** matrix is to fix an order for reactions, then apply it to the rows. We should do so for species which must be in the $\vec{r}$ vector too and apply it to the columns. Later, carefully choose the stoichimetric coefficients. Methods to manually calculate the kernel of a matrix is out of the scope of this book. Again, we just use the `null` function in the `pracma` package.

```
coefs <- c(-1,0,0,1,0,-1,0,-2,0,0,-4,0,-5,-5,-19,0,3,3,12,0,4,4,18,1)
N <- matrix(coefs,4,6)
X <- pracma::null(N)
X
```

What do this means? Hence, we have two options for redundant concentrations of the species vector where the system tends to be nullified.

```
x_1 <- pracma::null(N)[,1]
N %*% x_1
```

Inspecting at the size and sign of the first vector x_1, we can see that it states ethanol, methanol and water have a trend to be inside a loop which adds positive feedback to all the other reactions while its concentrations are equal and small. In this way CO_2 is the most abundant product followed by O_2.

```
x_2 <- pracma::null(N)[,2]
N %*% x_2
```

The second vector indicates there is another state where ethanol, methanol and water got inside a loop but their concentrations are greater. This negative feedback forces CO_2 to act as a reactant since its coefficient is negative and *isopropanol* becomes the most abundant product. If isopropanol production is the aim, we must try to follow the second mode as the answer to our system, and minimize effects of the first one.

Since the kernel has two modes

$$x_1 = \begin{pmatrix} 0.016 \\ 0.016 \\ 0.153 \\ 0.515 \\ 0.843 \\ 0.016 \end{pmatrix} \text{ and}$$

$$x_2 = \begin{pmatrix} 0.41 \\ 0.41 \\ 0.65 \\ 0.12 \\ -0.21 \\ 0.041 \end{pmatrix}$$

We can always state there exists a linear combination $\vec{s} = \kappa_1 x_1 + \kappa_2 x_2$ allowing us to modify the effect of each mode as we need. Both states indicates a condition where concentrations in the system no longer changes but reactions keeps happening. That is, both are conditionated ways to reach chemical equilibrium.

Unlike applying the mass-conservation law to find all the possible concentrations a system can generate, this approach is usually applied in a different way with equilibrium in mind.

Kinetics of several reactions

Once we establish a system of differential equations, we have three options to look for profile solutions.

1. *Find Exact Analytical solutions*. Here, we use properties of the system to propose an integration strategy in order to describe profiles as functions of time and other parameters. We should try to solve ODE system analitically when useful, but this could be the worst approach since many systems have trascendental solutions: there is any way to write profiles as combinations of known algebraic functions.
2. *Solve Numerically*. Applying numerical methods is usually the best approach since we just need to worry over the parameters and initial values. This approach is going to be deeply used throught the book.
3. *Conditionate Solutions*. If we need a profile working on a certain domain or for very-specific conditions, like positive time and concentration, or always been lower than a reference value, we can add these conditions into the ODE system to reduce it into a system of algebraic differential equations. Thermodynamical equilibrium and Kinetical Steady-State are two of the most common conditions to be applied in order to obtain simple profiles.

We have stated the mass-action rate law for reactants and products in one reactions, but as we can imagine there is a generalization for many chemical reactions. Let's use water as an example of conditionated solutions at steady-state.

Example State the reaction rate for water decomposition reaction

$2H_2O \rightarrow 2H_2 + O_2$

which follows the proposed mechanism

$$\begin{cases} H_2O \xrightarrow{k_1} H + OH \\ OH \xrightarrow{k_2} O + H \\ 2H \xrightarrow{k_3} H_2 \\ 2O \xrightarrow{k_4} O_2 \\ 2OH \xrightarrow{k_5} H_2O + O \end{cases}$$

Solution

In this way water decomposition is proposed as a first initiation reaction, followed by three propagation of the trend along the system by the decomposition of its components and a final inhibition equation where water is regenerated. By the proposed notation scheme

$$\begin{cases} S_1 \xrightarrow{k_1} S_2 + S_3 \\ S_3 \xrightarrow{k_2} S_4 + S_2 \\ 2S_2 \xrightarrow{k_3} S_5 \\ 2S_4 \xrightarrow{k_4} S_6 \\ 2S_3 \xrightarrow{k_5} S_1 + S_4 \end{cases}$$

Water as S_1 is the reactant, free hydroxile, hydrogen and oxygen (S_2, S_3, S_4) are intermediates and hydrogen and oxygen are products. The mass-action rate law applied to every single species involved can be grouped as the system

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} s_1 = -k_1 s_1 + k_5 s_3^2 \\ \frac{\mathrm{d}}{\mathrm{d}t} s_2 = k_1 s_1 + k_2 s_3 - 2k_3 s_2^2 \\ \frac{\mathrm{d}}{\mathrm{d}t} s_3 = k_1 s_1 - k_2 s_3 - 2k_5 s_3^2 \\ \frac{\mathrm{d}}{\mathrm{d}t} s_4 = k_2 s_3 - 2k_4 s_4^2 + k_5 s_3^2 \\ \frac{\mathrm{d}}{\mathrm{d}t} s_5 = k_3 s_2^2 \\ \frac{\mathrm{d}}{\mathrm{d}t} s_6 = k_4 s_4^2 \end{cases}$$

As every substrate and every reaction is used to propose this system, we can write it in matrix notation to be similar to the reaction rate equation

$\frac{\mathrm{d}}{dt}\mathbf{S} = \mathbf{N}v(\mathbf{S})$

Thus

$$\dot{\mathbf{S}} = \begin{pmatrix} -1 & 0 & 0 & 0 & 1 \\ 1 & 1 & -2 & 0 & 0 \\ 1 & -1 & 0 & 0 & -2 \\ 0 & 1 & 0 & -2 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} k_1 s_1 \\ k_2 s_3 \\ k_3 s_2^2 \\ k_4 s_4^2 \\ k_5 s_3^2 \end{pmatrix}$$

In order to loook for a solution in chemical equilibrium state, we must solve $\dot{\mathbf{S}} = \mathbf{0}$. This is, to look for concentrations profiles where no more changes in any reaction could happen. Again, using the kernel of the matrix to solve $\mathbf{N}v(\mathbf{S} = 0)$ we found

```
coefs <- c(-1,1,1,0,0,0,0,1,-1,1,0,0,0,-2,0,0,1,0,0,0,0,-2,0,1,1,0,-2,1,0,0)
N <- matrix(coefs,6,5)
X <- pracma::null(N)
X
```

$$X = \begin{pmatrix} 0.58 \\ -0.58 \\ 0 \\ 0 \\ 0.58 \end{pmatrix}$$

All linear combinations of $\mathbf{N}$ columns with weights given by $\mathbf{X}$ are zero, and this equation describes linear dependencies between reactions. Now, for the intermediate reactions

$$\begin{cases} H_2O \xrightarrow{k_1} H + OH \\ OH \xrightarrow{k_2} O + H \\ 2OH \xrightarrow{k_5} H_2O + O \end{cases}$$

We have they establishes an interconversion loop. Water dissociates, then hidroxile dissociates or regenerates water, which again dissociates. We can aware both third and fourth places are zero, which means the reactions

$$\begin{cases} 2H \xrightarrow{k_3} H_2 \\ 2O \xrightarrow{k_4} O_2 \end{cases}$$

Are independent of the others, being final states where reactants cannot be regenerated. If water is the “source”, thus hydrogen and oxygen are some class of “sink” states, where chemical equilibrium has been reached too.

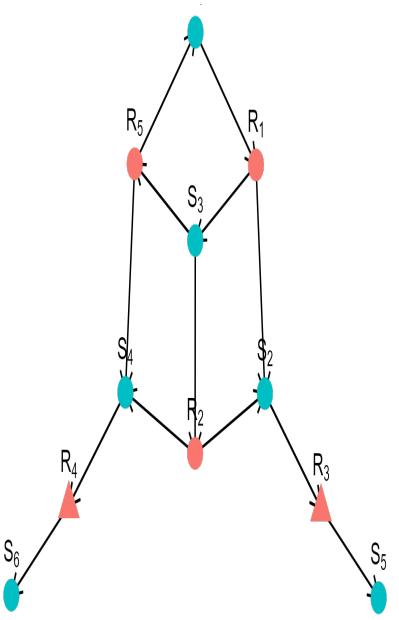


Figure 18. Scheme for Water Decomposition System

As we can see, there is a way to always retain water within a constant concentration, even if other reactions are happening. This is *chemical equilibrium*.

Now we can also approximate a *steady-state* solution for species profiles looking only to retain intermediaries rates to zero. The only reaction for intermediate production is

```
\begin{cases}
OH \overset{k_2}{\rightarrow} O + H
\end{cases}
```

which must have a null derivative in the steady-state

$$\begin{cases} k_1 s_1 + k_2 s_3 - 2k_3 s_2^2 = 0 \\ k_1 s_1 - k_2 s_3 - 2k_5 s_3^2 = 0 \\ k_2 s_3 - 2k_4 s_4^2 + k_5 s_3^2 = 0 \end{cases}$$

We can found the relations $k_1 s_1 = k_2 s_3 + 2k_5 s_3^2$, $2k_4 s_4^2 = k_2 s_3 + k_5 s_3^2$ and $k_3 s_2^2 = k_2 s_3 + k_5 s_3^2$. This equations simplifies the original ordinary differential equations system into an algebraic system of the form

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t} s_1 = -k_1 s_1 + k_5 s_3^2 \\ \dfrac{\mathrm{d}}{\mathrm{d}t} s_5 = k_3 s_2^2 \\ \dfrac{\mathrm{d}}{\mathrm{d}t} s_6 = k_4 s_4^2 \\ k_1 s_1 = k_2 s_3 + 2k_5 s_3^2 \\ 2k_4 s_4^2 = k_2 s_3 + k_5 s_3^2 \\ k_3 s_2^2 = k_2 s_3 + k_5 s_3^2 \end{cases}$$

And reducing terms

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t} s_1 = -k_2 s_3 - k_5 s_3^2 \\ \dfrac{\mathrm{d}}{\mathrm{d}t} s_5 = k_2 s_3 + k_5 s_3^2 \\ \dfrac{\mathrm{d}}{\mathrm{d}t} s_6 = \frac{1}{2} k_2 s_3 + k_5 s_3^2 \end{cases}$$

Negative sign in the first derivative indicates this is a reactant, and its concentration diminishes as the reaction goes on. The hydroxile concentration s_3 determines the net-reaction rate, as this species regulates free hydrogen and oxygen generation k_2 and bound hydrogen formation k_5. Now profiles can be integrated from these equations and the *reaction rate in steady-state* can be derived from these equations as $v = k_2 s_3 + k_5 s_3^2$■.

As you can see, chemical networks theory is such a great and developed field. Our focus into solve problems is only one step more into this fascinating applied world. We want to emphasize the existence of metrics and methods associated to these matricial processes are going to be revisited as become needed. For now, focus on the way to define the stoichiometric matrix from the ordinary differential system generated from rates of change for every reaction.

Kinetic Models

Kinetics is the study of rate equations. As we have seen the equation $\dot{\mathbf{S}} = \mathbf{N}v(\mathbf{S})$ can gave us a lot of useful information. Our main goal is to establish a paradigm to calculate equation rates $v(\mathbf{S})$ for the special case of enzyme-catalyzed reactions, were the mass-action rate law cannot be used.

Until this point we have managed to use the word "reactant" to define the initial state of the mass we are converting to product. But in enzyme kinetics the word "substrate" is preferred since we are using only the portion of mass to which an enzyme catalyzes a reaction with a certain degree of specificity. This synonymy event is common in enzyme kinetics, and it gets more specific to refer other kind of molecules capable to bind the enzyme, as ligands, effectors and modulators. We try to keep these changes as stepped as any reader would need.

Protein - Ligand Model

Before explaining enzyme catalysis, let's establish some basis explaining one of the most simple reaction models we can work with: the protein-ligand model. A *protein* could be any aminoacid oligomere as much as a *ligand* is any molecule which reversibly binds to a protein. No transformation is required, just only a bound. The most simple description for this phenomenon comes from the analysis of the protein-ligand complex.

Binding

If a protein P and a ligand L form a complex PL then the system's mechanism can be described by the reaction:

$$\mathrm{P} + \mathrm{L} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{PL}$$

Where p and l represent the concentration of free protein and ligand, and pl of the protein-ligand complex. However, this is not a function of the free

ligand concentration and we must direct our efforts to do so. Thermodynamic and kinetic definitions are applied to the equilibrium reaction, where:

$v_{-1} = k_{-1}pl$

$v_1 = k_1.l.p$

which indicates how quickly chemical equilibrium is reached using the *mass-action* rate law. At equilibrium, an additional condition holds:

$v_{-1} = v_1$

And we can define a dissociation constant $K_d = \frac{k_{-1}}{k_1}$ which resembles the equilibrium constant as

$K_d = \frac{1}{K_a} = \frac{l.p}{pl}$

$pl = \frac{l}{K_d}p$

By the mass-conservation property

$p_0 = p + pl$

And to relate the amount of the bounded form from the initial amount of protein we use a single ratio called *fractional saturation*:

$$F(l) = \frac{pl}{p_0} = \frac{\frac{l}{K_d}p}{p + \frac{l}{K_d}p} = \frac{l}{K_d + l}$$

Since this reactions do only bind and unbind the ligand, we only need to associate this *fractional amount*. Knowing the free-ligand concentration and its dissociation constant, we can get an estimation of the amount of bounded protein.

Example Chaperonins are proteins which transports old oligopeptides into ubiquitins to be catabolized. If its dissociation constant has been estimated to be $K_d = 2 \times 10^{-3}[\mu M]$. How much oligopeptide can be bounded from a $0.77[\mu M]$ solution in a 0.1 saturated environment?

Solution

$$0.1 = \frac{l}{0.002 + l}$$

$$l = 0.\widehat{2}\,[\mu M]$$

However l accounts for the free-ligand concentration. To find how much oligopeptides does chaperonins can bound we must use the mass-conservation principle:

$$l_0 = l + pl$$

$$pl = 0.77 - 0.22 \approx 0.55[\mu M]$$

Single Enzyme Kinetics

When a reaction actually happens, the most extended model for reaction analysis is the irreversible catalysis scheme converting a substrate into a product as $S \xrightarrow{E} P$, where s, p and e are variables describing the concentration of each species and every stoichiometric amount is one.

Chemical quasiequilibrium approach

We define an irreversible catalysis scheme for an enzyme-regulated reaction converting a substrate into a product as $S \xrightarrow{E} P$. The concentrations we measure are those available in the medium, those which are free, unbounded ones. Consider the formation of an intermediate substrate-enzyme complex which catchs and retains the substrate while applies a finite set of chemical transformations following a *mechanism of reaction*, in order to apply the mass-action rate law, and returns a product which, once generated, is immediately released.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

Where k_i describes rate constants related to the efficiency of the reaction within a some-order kinetics. In fact, these kinetic constants can be used to generate an ordinary differential equations system, the $\dot{\mathbf{S}} = \mathbf{N}v(\mathbf{S})$ system, fully describing a deterministic overview of the concentrations profiles over time.

Problem Formulate the enzyme-substrate ODE system in matrix notation. Why is no elemental solution for this system?

Solution

In order to correctly write the ODE system we must consider enzyme concentration e diminishes with the first reaction since E is consumed to generate ES, thus its rate of change is negative. Also, it gains when the ES complex releases the substrate S or the product P and the corresponding rates of change are positive. The same holds for every other rate of change and we state the system reduces to

$$\begin{cases} \dot{e} = -v_1 + v_{-1} + v_2 \\ \dot{es} = v_1 - v_{-1} - v_2 \\ \dot{s} = -v_1 + v_{-1} \\ \dot{p} = v_2 \end{cases}$$

In matrix notation it becomes

$$\begin{pmatrix} \dot{e} \\ \dot{es} \\ \dot{s} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_{-1} \\ v_2 \end{pmatrix}$$

This system has no unique elemental solution because there are more rows than columns in the stoichiometric matrix. Thus, we must proceed to keep modelling rates as functions of species profiles.

Also, after you complete to read this chapter, please return here and explain the properties of the column vectors in the $ker(\mathbf{N})$ and $ker(\mathbf{N}^T)$ null spaces of this model.

Graphs

Biochemical networks rely on the analysis of entities and its relationships. While modeling, we usually need to rely on those abstract entities to compare reality with our desired analytical system. This chapter introduces *graphs* as a mathematical tool which gives us a way to describe chemical reactions, pathways and networks. Graphs are introduced with the help of `tidygraph, igraph` and `\texttt{ggraph}` packages in the R language, with a deep focus on describe biochemical systems.

Definitions

Biologist have a need to ordering life by some degree of complexity in order to simplify the big picture response as the effect of many individual interactions. The same approach is applied here while describing biochemical reactions in a single cell. The plot below represents the transformation of a *substrate* into a *product* as a graph.

![Conversion of substrate into product reaction as a graph](graphs /g1.png)

A *graph* $\mathcal{G} = (\mathcal{N}, \mathcal{E}, \phi)$ is an ordered triple with two sets of elements called *nodes* $\mathcal{N}$, *edges* $\mathcal{E}$ and a function ϕ mapping every edge to a pair of nodes. The nodes set contains all the elements x_i which are allowed to have interactions between them, and the edges set contains links between those elements represented as pairs. Formally

$$\phi : \mathcal{E} \subseteq (x_i, x_j) | x_i, x_j \in \mathcal{N} \text{ and} i \neq j$$

It is more easy to show you how do this system works using programming. Think in the conversion reaction. We can define

```
nodes <- c("S","P")
edges <- c(1,2)
g <- make_graph(edges)
plot(g)
```

The mapping of the ϕ function is implicit in the $\mathcal{E}$ definition which also uses $\mathcal{N}$ to be defined. Thus, we can use $\mathcal{E}$ to define the whole graph. As you can see, graphs are such a great way to represent chemical reactions. Graphical representations of graphs requires the stablishment of a *layout* criterion which defines rules to show elements and links, like to put them in random order, sort by names or other strategies . We should also add *geometries* in order to add shapes, colours, names, and many more. We are going to use points to represent nodes, and lines for edges.

```
g |> ggraph(layout = "linear")+
    geom_node_point()+
    geom_edge_link()
```

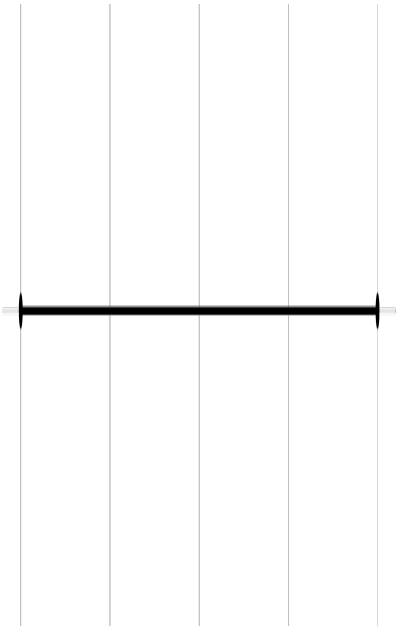


Figure 19. Substrate to Product Graph

Since we have only one edge `(S,P)` both nodes are called *endpoints* and the edge is said to be *incident* on each node by the incidence function ϕ.

Here, we can start to think on different ways to add nodes, link them and generate more complex structures by re-defining ϕ e.g. A *loop* is an edge that joins a node with itself, and it changes the prior ϕ definition because now $i = j$.

Example Generate a graph for a reaction between four elements S_i, where

1. The first three elements reacts in order
2. The third element reacts with itself
3. None element reacts with the fourth node

Solution

We are going to use the *tidy* data scheme suggested for the `tidygraph` package in order to simplify this assignment. We define $\mathcal{N}$ and $\mathcal{E}$ as dataframes, and then relate them as a graph.

```
N <- data.frame(
    vertices = c(1,2,3,4),
    names = c("S1","S2","S3","S4") # Column for additional properties
)

E <- data.frame(
    from = c(1,2,3),
    to = c(2,3,3) # Only do not add edges for S4
)

g <- graph_from_data_frame(E, vertices = N)

g |> ggraph(layout = "linear")+
    geom_node_point()+
    geom_node_label(aes(label = names), repel=TRUE)+
    geom_edge_link()+
    geom_edge_loop0()
```

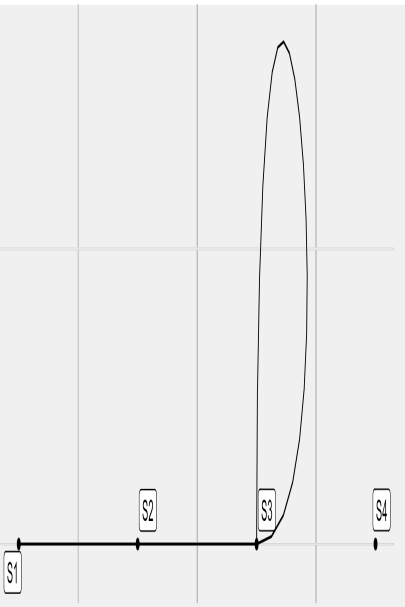


Figure 20. Solution for the empty-edges graph

line 75

Problems

The main goal of these sections is to help you truly understand the ideas exposed in every chapter. In the same way we are unable to become an elite football player by sitting at the couch and seeing others playing football, there is no way anybody can understand science just by reading. If you know any please tell me.

Otherwise, go on an try to solve these problems by your own, in order to develop problem-solving skills focused on pharmaceutical, agricultural, ecological, microbiological and biotechnological issues. Some of them are solved too and may be useful for the upcoming chapters. If you want to discuss any solution feel free to email me.

Variables describing Change

1. Define the following terms

a) Mass m b) Molar mass M b) stoichiometric amount n b) stoichiometric ratio N c) molar concentration r

2. Valproic Acid is an artificial antiepileptic drug derived from *Valeriana officinalis spp.*. Inside liver microsomes, valproic acid (A) can bind an isoform of the P450 cytochrome enzymes (CYP isoform 2C9) which usually works as a monooxigenase. That is, they got involved into the reaction:

$$A - COOH + CYP \rightarrow A - COH + CYP = O$$

Your doctor says you must take $5.5[ml]$ of Valproic acid, and the comercial product instructions indicates it has $250[mg]$ for every $5[ml]$.

a) Define the amount of Valproic acid you must take. b) Since $M = 144.2[\frac{g}{mol}]$, define your intake as an stoichiometric amount. c) Use this information to find the blood molar concentration if an average person has 5 liters of blood in their body. d) If you are going to keep this dosis two times a day for the next two years, how much valproic acid do you need?

Concentration Rate of Change

3. Ergothioneine is a natural secondary metabolite sinthesized from Histidine, an essential aminoacid, in fungi and actinobacterias. Plants and animals use it as a micronutrient degrading it into l-glutamate, trimethylamine, hydrogen sulfide, carbon dioxide, and ammonia as products. Healthy amounts of ergothioneine in adults is up to $1.31[\frac{mg}{kg}]$ of ergothioneine in your body as a healthy upper bound.

a) Define your ergothioneine molar concentration, suppose your volume is round to $67[L]$. b) Find an expression for the X concentration, if $0.5[mM]$ shows a rate of change estimated as $\dot{x} = -1.57x$ c) Find `\dot{x}$ for` x(t) = 0.47 -0.53t + 0.16t^2‘$. Is this a good approximation for the derivative in b)?

Stoichiometry coefficients

4. Late blight is a phenotypical trait of a world-wide spread fungi. It represents a serious treat for agricultural-based economies, and has a very bad history with bananas (Cavendish cultivar), potatoes (Irish famine) and rubber (Ford voragine), as much as for many other plants.

After two years of applied research, you have managed to create an artificial aminoacid never reported before, and you think it could be key in optimization of industrial pesticides production offering unusual specificity against Late blight. Industrial production could be safer and you could become rich with a proper patent.

However, there is a problem. Infection is inhibited by spores inactivation, yes, but you cannot establish if full or partial inhibition, and you run out of founds for more experiments. This is your target reaction and measured stoichiometric amounts are tabulated.

a) Does this reaction can generate complete inhibition? b) How much reacivated spores are expected if there is a ratio of 1000 reactions per spore?

Reaction yield

There are other metrics which describes the transformation steps for a chemical process. We have seen *Yield* as the amount of product obtained. *Conversion*

(x) measures the amount of converted reactant and *Selectivity* (s) is the partial product fraction focused to undesired products. All of this ratios are related by a simple but circumstancial equation

$$y = xs$$

5. Cytochrome P450 is the main pathway for drug metabolism. Specifically, for drugs related to bencene rings and Reactive Oxygen species (ROS). For a $2[mg]$ CYP experimental setup, a valproic acid intake of $250[\mu g]$ was used to react inside microsomes until inactivation. Let the reaction formula get simplified by changing notation

$$A + E \overset{CYP}{\rightarrow} B + P$$

Where all stoichiometric amounts remains one and mass for species are estimated as tabulated below

State	**A**	**E**	**B**	**P**
initial `(m_o)$`	250	2	0	0
final `(m_f)$`	86.5	1.58	163.52	0.4

Calculate the three reaction ratios x, s and y

Solution:

Hence, we can use the definitions for each ratio.

Yields

$$y_b = \frac{m_{(b)f}}{m_{(a+e)o}}$$

$$y_b = \frac{163.52}{252} = 0.6489$$

$$y_p = \frac{m_{(p)f}}{m_{(a+e)o}}$$

$$y_p = \frac{0.4}{252} = 0.001587$$

Conversion

$$x_a = \frac{m_{(a)o} - m_{(a)f}}{m_{(a+e)o}}$$

$$x_a = \frac{163.5}{252} = 0.6488$$

$$x_e = \frac{m_{(e)o} - m_{(e)f}}{m_{(a+e)o}}$$

$$x_e = \frac{0.42}{252} = 0.0017$$

Selectivity

$$s_b = \frac{m_b}{m_b + m_p}$$

$$s_b = \frac{163.42}{163.82} = 0.9976$$

$$s_p = \frac{m_p}{m_b + m_p}$$

$$s_p = \frac{0.4}{163.82} = 0.0024$$

From the previous equation we can estimate the calculations are right because

$$xs = y$$

$$(0.6488 + 0.0017)(0.9976 + 0.0024) = (0.6489 + 0.0016)$$

Reaction rate

6. Found stoichimetry coefficients, rates of change as a function of S_1 and the reaction rate for the following reactions.

a) $S_1 \rightarrow$ b) $S_1 \rightarrow S_2$ c) $c_1 S_1 \rightarrow c_2 S_2$ d) $c_1 S_1 \rightleftharpoons c_2 S_2$ e) $S_2 \rightarrow S_1 \leftarrow S_3$ f) $S_2 \leftarrow S_1 \rightarrow S_3$ g) $N_2O + H_2S \rightarrow N_2 + S + H_2O, \quad S_1 : N_2O$ h) Glucose phosporylation reaction

7. Silver Thiosulphate (STS) is a conservative agent added to different cultivars to increase its useful life and promote flower development throught a cupper-replacement mechanism over senescency ethylene-related signaling pathways, giving enough time to industrial processing of comercial flower bunches. The effect of STS is theorized by the decomposition reaction

$$Ag[(S_2O_3)_2]^{3-} \rightarrow Ag^+ + 2(S_2O_3)^{2-}$$

Where free-silver replaces cupper anions as heavy-metal complex cofactors in ethylene-signaling receptors, supressing the expression of senescency genes.

a) Found an expression for the reaction rate v of this reaction following the mass action rate-law.

b) Using a $0.1[mM]$ solution of STS, we can promote the lifetime from 192 hours to 320 hours, when STS looses all activity. If we increase the concentration to $0.6[mM]$ the lifetime increses to 720 hours. Found the fold change increase in rate constant value.

c) A $4000[L]$ STS vessel is fullfilled everyday to sustain industrial operations. Every day looses up to $10[L]$ of water by environmental changes and manipulation, and must be refilled to the same $0.6[mM]$ concentration. If we use a connected refilling tank with $20[L]$ of water, how much mass of pure reactive we must add?

Rate Kinetics

9. Ozone decomposition is such an interesting case for rate kinetics study, since the mechanism of reaction includes many reactive species which catalyzes even more internal processes. Studies associating only ions and hydrogen from water interaction uses up to 38 reactions involved in the mechanism to explain kinetic data by ions and radicals formation, and this approach avoids the effect of pH, temperature and other factors.

a) Write the reaction rate v using the mass-action principle for the Ozone decomposition reaction

$2O_3 \rightarrow 3O_2$

and find the Ozone profile $r(t)$ for the initial conditions

$k = 2.3x10^{-3}[Lmol^{-1}s^{-1}]$

$r_0 = 0.1[mM]$

$t_0 = 0 \quad t_f = 220[s]$

b) A more complex model for the mechanism of this reaction is stated as follows

$$\begin{cases} O_3 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} O + O_2 \\ O + O_3 \overset{k_2}{\rightarrow} 2O_2 \end{cases}$$

where $k_1 = 10^{-5}[s^{-1}]$, $k_{-1} = 10^6[Lmol^{-1}s^{-1}]$ and $k_2 = 1$. Show that the reaction rate is better explained by $v = ko_3^2/o_2$ and estimate the three profiles.

c) How does the condition $k_1 >> k_{-1}$ affects the general kinetics over the profiles estimation?

d) An ODE system with 38 species and 13 ODEs was proposed and implemented in R. Compare data from the prior models with this one using the `ozone` dataset.

Chemical equilibria

10. Autocatalytic reactions, where a product acts as a catalyzer for itself and accelerates its own production, are such an example that reaching the thermodynamic equilibrium state also keep reactions ongoing. Autocatalytic systems are non-linear, and thus they represent systems with more states where the production is easier respect to the linear version of the reaction. These kinetics usually involves acid-catalyzed hydrolysis of some esters to carboxylic acids and alcohols like in the *Formose reaction*, and were once suggested as a support to the abiogenesis theory on the origin of life.

Consider the following autocatalytic reaction

$S + P \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} 2P$

a) Formulate a two ODEs system using the reversible mass-action rate law b) Find an analytical solution for this system (Hint: Find first a solution for the irreversible case with $k_{-1} = 0$, or try with a logistic function.) c) Find a solution for $s_0 = 0.5[mM]$, $p_0 = 0.01[mM]$, $k_1 = 10^2[Lmol^{-1}s^{-1}]$ and $k_{-1} = 10^{-1}[Lmol^{-1}s^{-1}]$ d) Apply the equilibrium assumption into the system using $p = K_{eq}s$. How does this affect the analytical solution?

Mass-action and Disequilibrium ratio

11. The autocatalytic reaction $S + P \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} 2P$ with net-rate $v = -\dot{s} = \dot{p}$ must be modeled as the system

$$\begin{cases} \dot{s} = k_{-1}p^2 - k_1sp \\ \dot{p} = k_1sp - k_{-1}p^2 \end{cases}$$

a) Find a value for the gamma ratio by dividing both analytical solutions of this system

b) Find a value for the disequilibrium-ratio with $K_{eq} = 1000$

c) The expected molar yield for this reaction roughly reaches 54% in your lab. Write the expected net-rate for this condition.

Catalysis effect over the neat rate

12. Hydrogen Peroxide decomposition rarely occurs once the molecule is generated. This reaction liberates as much energy as heat that has been adapted by insects as a defense mechanism against predators (*Brachinus spp.*) and by plants as an oxidant agent over toxic molecules. Hydrogen peroxide decomposes to water by catalytyc effects following the reaction

$2H_2O_2 \overset{k_1}{\rightarrow} 2H_2O + O_2$

with a rate constant of $k_1 = 1x10^{-6}[s^{-1}]$

a) If Iron $([Fe^{3+}])$ is applied as a linear catalyst, how much Iron do I need to add in order to reduce the net-rate up to one thousand times?

b) Catalase (E) decomposes peroxide hydrogen to reduce toxins, like phenols and ROs. What is the rate constant of the decomposition reaction if an increase of 10^{13} times is expected using the natural reaction as a reference?

Matricial Notation

8. NFKB1 is a critical regulator of the immediate-early response to viral infections. It is known as a *Long Terminal Repeat* (LTR) sequence and its gene encodes a 105 kD protein which can undergo cotranslational processing by the 26S proteasome to produce a 50 kD protein, which works as a DNA binding subunit i.e a transcriptional factor. This sequence is a typical sub-product expected in tests for viral infection in humans.

Suppose a virus infects the genome of a person by insertion of the template DNA string sequence "CCAGCCCTCCGCACCCCCTCCAGCCCCACTCGGACTTCCTCATTCCTGCGCTAACCGCTGGGCTAGACCG" nucleotide by nucleotide.

a) State the balance matrix for the whole insertion reaction.

b) Find and read the null kernel of this reaction. Is there any usefull information you can read from both kernels?

Single reaction

13. Balance methods are useful over organic reactions too. The polimerization reaction

$$A + T + C + G \underset{P,T,C}{\overset{E}{\rightarrow}} C_x H_y N_z O_w P_v$$

generates copies of a template DNA string (T) from single nucleotides using a polymerase (E), a set of primers (P) and mineral cofactors (C) over a set of coupled reactions. We want to replicate the sequence for the whole NFKB1 transcription regulator.

a) How many times does every nucleotide appears in each final sequence under the following conditions? We know that

$A = C_{10}H_{16}N_5O_{12}P_3$

$C = C_9H_{16}N_3O_{14}P_3$

$T = C_{10}H_{17}N_2O_{14}P_3$

$G = C_{10}H_{16}N_5O_{14}P_3$

And a count analysis over the sequence allows us to establish thath

$$x + y + z + w + v = 1137526 + 1877086 + 469946 + 1553444 + 347832 = 5385834$$

$$x + y + z + w + v = 5385834$$

b) Prove your answer is right by loading the `nfkb1.fna` gene file and showing how many nucleotides it does have. That is:

a	c	g	t
34886	21914	21982	37162
34886	21914	21982	37162

c) If we are expecting at least 4Gb (gigabases) of nucleotides to be used, what is the minimun stoichiometric amount of the final sequences which could be produced?

Set of reactions

14. In systems biology, a *stoichiometric matrix* is an array **N** were the matricial equation $\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \mathbf{N}v$ holds for the time derivatives of any set of species **S** and their reaction rate.

Describe the following linear coupled reactions by its stoichiometric matrix and find solutions through the null-kernel approach. How do you believe these reactions can be modeled as "networks"?

a. $S_1 \rightarrow 2S_2 \rightarrow 3S_3 \rightarrow P$

b. $S_1 + S_2 \rightarrow 2S_3 \rightarrow P + S_4$

c. $3S_1 + 2S_2 \rightarrow 4P \leftarrow S_3 + 5S_4$