Bioprocess Engineering

R. Clay Wright

2024-12-12

Contents

1 Introduction

This book has been developed as a companion for the course in Bioprocess Engineering for Biological Systems Engineers at Virginia Tech, BSE3534. It is based on many different resources from the chemical and biological engineering fields (see Additional Resources below) and aims to develop many introductory ideas in kinetics and reactor design through a primarily biological lens.

1.1 Description and Objectives

This text and course introduce engineering concepts for biological conversion of raw materials to food, pharmaceuticals, fuels, and chemicals. These concepts include enzyme kinetics, metabolic and cell growth kinetics, and design and analysis of reactors, bioreactors, and fermenters.

1.1.1 Course Objectives

After reading this text and working associated examples and assignments, students will be able to:

- Model chemical and biochemical reactions using mass balances and kinetic rate equations.
- Analyze and formulate mechanisms and rate equations for enzymatic reactions.
- Estimate kinetic parameters from experimental data. -Understand soluble and immobilized enzyme technologies for the production of industrial and medical products.
- Model microbial growth and metabolism.
- Identify and design bioreactors for various products and organisms.
- Specify required practices and technologies to safely and effectively utilize genetically engineered organisms for biomanufacturing.

1.2 Additional Resources

Chemical Reactions and Chemical Reactors
by George Willard Roberts
Find a copy: http://www.worldcat.org/oclc/441742755

Elements of Chemical Reaction Engineering, 6th Edition
by H Fogler
Available online: https://www.worldcat.org/oclc/1192531500

Fundamentals of Modern Bioprocessing 1st Edition
by Sarfaraz K. Niazi, Justin L. Brown
FREE PDF download: https://www.worldcat.org/oclc/958799491

Bioprocess Engineering: Basic Concepts, 3rd Edition
by Shuler, Kargi & DeLisa
Available online: https://www.worldcat.org/oclc/981256254

Integrated Bioprocess Engineering
by Clemens Posten
Available online: https://www.worldcat.org/oclc/1032686988

Bioprocess Engineering
by Shijie Liu
Available online: http://www.worldcat.org/oclc/1173991723

1.3 Welcome to Bioprocess Engineering (BSE 3534) - Syllabus Rundown

1.3.1 Course Philosophy

This course emphasizes interactive, problem-focused learning. Through group work, in-class problem-solving, and step-by-step examples, you'll develop the skills to approach and solve engineering challenges. Participation, whether live or asynchronous, is key to success.

1.3.2 Community Guidelines

Our collaborative environment values respect and inclusivity. Actively listen, include all voices, and constructively engage with different perspectives. If issues arise, contact the instructor or TAs.

1.3.3 Tools and Platforms

- **Canvas**: All assignments will be submitted through Canvas. We will also keep links to course materials in a Google Drive folder updated on Canvas, but do let us know if any links are broken.
- **Google Drive**: Shared documents for live collaboration. Ask questions directly in documents as comments. Feel free to suggest changes too!

1.3.4 Assessments

- Formative (skill-building) assessments include collaborative homework and active learning credits.
- Summative assessments (skill-evaluating) will take the form of 3 exams.

1.3.5 Modules Overview

1. Modeling chemical reactions: Kinetics and reactors
2. Biochemical/Enzyme kinetics

3. Cell growth kinetics and reactor design

Let's build a supportive and engaging learning community. Reach out anytime via Piazza or email with questions or suggestions. Looking forward to a great semester!

2 Chemical Reactions: Stoichiometry and Rates

Chemical reactions are critical to life and critical to the production of food, fuels, materials, chemicals, and pharmaceuticals. Bioprocess Engineers utilize this intersection to produce these valuable and necessary commodities in a cheaper, less resource and energy intensive, lower risk, and more sustainable way. This is done by harnessing the lifeforms that perform valuable chemical reactions and engineering systems to scale up the production of these valuable chemicals for commercial use.

Before we can delve into engineering chemical reactions we must first develop the ability to quantify chemical reactions.

2.1 Stoichometry

The stoichiometry of a chemical reaction tells us what chemical species are consumed and formed in a reaction and in what proportions. Stoichiometry is the basis of our ability to quantify a chemical reaction

Consider the chemical reaction:

$$Cl_2 + C_3H_6 + 2\,NaOH \longrightarrow C_3H_6O + 2\,NaCl + H_2O$$
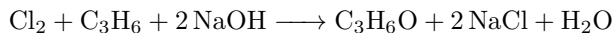
It is important to check that this reaction equation is balanced, and therefore satisfies the law of conservation of matter. We will check that this equation is balanced as a review.

Remembering your general chemistry, in order to write a balanced chemical equation the chemical elements (j) must be conserved, *i.e.* for each chemical compound/species i

$$\sum_i \nu_i N_{ij} = 0$$

where ν_i is the stoichiometric coefficient for each chemical species containing the element and N_{ij} is the number of molecules of the element j in species i.

Within this equation there is a critical convention: *The coefficients of the products of a reaction are positive, and the coefficients of the reactants are negative.*

So for the reaction above we can make a table for each chemical species and element:

$\nu_i N_{ij}$	Cl_2	C_3H_6	NaOH	C_3H_6O	NaCl	H_2O	$\sum_i$
Cl	-1 * 2	0	0	0	2*1	0	0
C	0	-1*3	0	1*3	0	0	0
H	0	-1*6	-2*1	1*6	0	2	0
Na	0	0	-2*1	0	2*1	0	0
O	0	0	-2*1	1	0	1	0

You may have used this stoichiometric notation in thermodynamics to calculate the standard Gibbs free energy or standard enthalpy change of a reaction.

This balanced stoichiometric equation tells us then that for every mole of Cl_2 consumed, one mole of C_3H_6O will be formed, as long as there is sufficient C_3H_6 and NaOH. But how can we figure out if there are sufficient quantities of the other reactants? How can we figure out which reactant is limiting?

2.2 Quantifying the Progress of a Reaction

2.2.1 Extent of Reaction

The **extent of reaction** is a measure of how far a reaction has progressed at any point in time. To define the extent of reaction we will use the example of a simple batch reactor, imagine a beaker containing fixed amounts of chemicals which are reacting. To define how far the reaction has progressed we start by defining the change in each chemical species throughout the reaction.

At the start of the reaction there are n_{i0} moles of each i species in our reaction. At time t the number of moles of species i is defined as n_i. We can then define the change in species i as $\Delta n_i = n_i - n_{i0}$.

The extent of reaction is a property of the whole reaction not each individual species. So to convert the change in moles of any species, to the extent to which the reaction has progressed, we divide the change in moles of species i by its stoichiometric coefficient, *i.e.*

$$\xi = \Delta n_i / \nu_i$$

2.2.1.1 ???Questions??? Is the extent of reaction for reactants positive or negative? Is the extent of reaction for products positive or negative?

The extent of reaction is always positive. Be careful about always sticking to the final − initial convention and check the signs of your stoichiometric coefficients.

The extent of reaction can also be defined for continuous (steady state open) systems

$$\xi = \Delta \dot{n}_i / \nu_i$$

2.2.1.2 Limiting Reactant Extent of reaction is how we calculate the limiting reactant. The limiting reactant is the reactant that is completely consumed in the reaction and prevents the reaction from progressing any further. Therefore the limiting reactant, l, defines the maximum extent of the reaction, because it is completely consumed in the reaction, *i.e.*

$$\xi_{max} = -n_{l0} / \nu_l$$

We can find the limiting reactant by calculating the extent of reaction for each reactant species if it were completely consumed. **The limiting reactant has the lowest extent of reaction, *i.e.* it runs out first if the reaction goes to completion.** In reversible reactions, equilibrium will be reached prior to completion and ξ will always be less than ξ_{max}.

2.2.2 Fractional Conversion

The fraction of the limiting reactant that has been consumed is also frequently used to quantify the progress of a single reaction. **Fractional conversion** is defined as:

$$x_l = \frac{n_{l0} - n_l}{n_{l0}} = 1 - \frac{n_l}{n_l 0}$$

Fractional conversion can also sometimes be specified for certain species in the reaction, *e.g.* x_A. This sometimes done when only one species in the reaction can be measured. For a constant volume/density system conversion can also be defined in terms of concentration.

$$\frac{C_{A0} - C_A}{C_{A0}} = x_A = 1 - C_A / C_{A0}$$

$$\therefore C_A / C_{A0} = 1 - x_A$$

2.2.3 Multiple Reactions, Yield, and Selectivity

If there are multiple reactions happening we can assign each one an extent of reaction and use these to define generation and consumption terms in our material balances. Let's walk through an example to demonstrate this.

Thiophene (C_4H_4S) is a problematic contaminant of both biofuels and petroleum because it contains sulfur. A pilot plant for removal of thiophene by hydrodesulfurization has been run with the following results:

Species	moles fed	moles effluent
C_4H_4S	75.3	5.3
H_2	410.9	145.9
C_4H_{10}	20.1	75.1
H_2S	25.7	95.7

The following reactions are proposed to occur:

$$C_4H_4S + 3\,H_2 \xrightarrow{1} C_4H_8 + H_2S \qquad C_4H_8 + H_2 \xrightarrow{2} C_4H_{10}$$
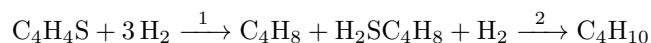
Are the data above consistent with these proposed reactions *i.e.* do the reactive material balances work out?

To answer this we need to write the material balances on each species. We can assume no accumulation in the reactor in this case, so our General Balance simplifies to

$$OUT = IN + GEN - CONS$$

Now we can make a table and fill in each species. Let's skip the symbolic notation here for simplicity

General:	OUT	=	IN	+	GEN	−	CONS
C_4H_4S :	5.3	=	75.3			−	ξ_1
H_2 :	145.9	=	410.9			−	$3\xi_1 - \xi_2$
C_4H_{10} :							
H_2S :							
C_4H_8 :							

OUT and IN values come directly from the table. GEN and CONS values are written in terms of the extents of reactions. It is particularly difficult to write out the balances in terms of conversions for multiple reactions, this is why we introduce extents of reaction. Fill in the last 3 rows of the table for practice.

Now that we have written all of our balances, we can solve for unknowns, the extents of our two reactions. We can do that from the first two species balances above.

$$\xi_1 = 75.3 - 5.3 = 70$$

$$\xi_2 = 410.19 - 3(70) - 145.9 = 55$$

Now to answer the question we need to see if the remaining balances are consistent with these values. If we plug in these values to the other balances are they mathematically correct? If so then the law of conservation of mass is satisfied based on our data and the proposed reactions.

2.2.3.1   ???Questions???   How much butene (C_4H_8) is in the effluent of the reactor?

Butene is formed in the first reaction and consumed in the second, so $N_{Be} = 1 \cdot 70 + (-1) \cdot 55 = 15$ mol.

2.2.3.2   Yield Coefficients   A yield coefficient is a fraction consisting of the amount of product formed over the amount of reactant consumed. Thus yield coefficients are kind of like stoichiometric ratios, but they can be written for multiple reactions. They can even be used to describe the "stoichiometry" of enormous numbers of reactions like all of the reactions required for a cell to consume a sugar and produce more cell mass. Yield coefficients (sometimes just called yields) allow us to do simple calculations of the amounts of products formed from some substrate/reactant, or how much reactant is required to form some amount of product.

What is the yield coefficient of butane (C_4H_{10}) on hydrogen?

We would specify this as

$$Y_{C_4H_{10}/H_2} = \frac{\Delta n_{C_4H_{10}}}{-\Delta n_{H_2}}$$

Yield coefficients are always expressed as positive values.

It is always best practice to carry the species names along with the units when using yields or stoichiometric ratios in calculations. The units of the above yield would be

$$\frac{\text{mol } C_4H_{10}}{\text{mol } H_2}$$

2.2.3.3 Selectivity Another way to quantify the progress of multiple reactions, with a particular focus on the desired product, is called **selectivity**. Selectivity is just a specific yield, the yield of desired product on undesired product, so

$$S_{\mathrm{C_4H_{10}/C_4H_8}} = Y_{\mathrm{C_4H_{10}/C_4H_8}} = \frac{\Delta n_{\mathrm{C_4H_{10}}}}{-\Delta n_{\mathrm{C_4H_8}}}$$

The generation of undesired products is pretty prevalent in in traditional organic chemistry where different isomers are inevitably formed in most reactions. In biological systems, enzymatic reactions are often stereo-specific mostly alleviating the need for selectivity. However, it is certainly still useful as a metric for optimizing systems, particularly in metabolic engineering.

2.3 Reaction Rates: Quantifying the Dynamics of a Reaction

Reaction rates are simply the rate of consumption of reactants or generation of products.

2.3.1 Species-dependent Reaction Rate

To be useful in designing reactors the **reaction rate** must be an intensive variable—independent of the size of the reactor. The reaction rate is typically defined in terms of one of the reactants or products of the reaction. The reaction rate is defined as:

$$r_i \equiv \text{rate of formation of product } i \left[\frac{(\text{moles } i \text{ formed})}{(\text{unit time})(\text{unit reactor})} \right]$$

or

$$-r_i \equiv \text{rate of disappearance of reactant } i \left[\frac{(\text{moles } i \text{ consumed})}{(\text{unit time})(\text{unit reactor})} \right]$$

Reaction rates by convention are always positive. We indicate the reaction rate with respect to a reactant by multiplying by -1. This allows us to use the stoichiometry of the reaction to convert reaction rates between species in a reaction, for a stoichiometrically-simple reaction, *e.g.*

$$r_1/\nu_1 = r_2/\nu_2 = r_3/\nu_3 = ... = r_i/\nu_i$$

If multiple reactions are taking place the rate of each reaction must be known to calculate the total rate of formation/disappearance of a species:

$$r_i = \sum_{k=1}^{R} r_{ki}$$

where R is the number of independent reactions taking place and k is the specific reaction.

So for example, the following 2 reactions are happening in a batch reactor:

$$\mathrm{A} \xrightarrow{\;1\;} 2\,\mathrm{B} \qquad \mathrm{B} \xrightarrow{\;2\;} \mathrm{C}$$

The rate of formation of B, then would be the sum of the rate of formation of B from each reaction, but B is a reactant in reaction 2, so this is a negative rate (rate of consumption). $r_B = r_{1B} - r_{2B}$

2.3.2 Species-independent Reaction Rates

We can make the reaction rate independent of the species for which it was calculated, and referenced to the reaction itself, by dividing the species-dependent rate by the stoichiometric coefficient for that species in the reaction.

$$r \equiv r_i/\nu_i$$

This form of the reaction rate has the advantage of preventing confusion about the rate for one reactant being different from another, but does require that the chemical equation that the rate describes be defined. For example the value of r would be different for the following two equations

$$N_2 + 3\,H_2 \rightleftharpoons 2\,NH_3$$

$$\frac{1}{2}\,N_2 + \frac{3}{2}\,H_2 \rightleftharpoons NH_3$$
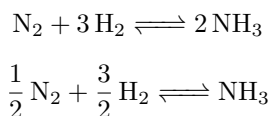
However, r_{N_2} would be the same for the two reactions, and $3r_{N_2} = r_{H_2}$. Fortunately, this species-independent form of the rate equation is rarely used for biochemical reactions. We present it here mostly as a reminder to always check the stoichiometry of the reaction and the units of the rate for consistency, particularly when using data from the literature. This is especially important as for biochemical reactions the symbol v for velocity of reaction instead of r for rate. Velocity is frequently defined in terms of the primary product of the reaction, but always check to make sure, subscripts are rarely used to specify how velocity is defined.

2.3.3 Rate Equations

Now that we have defined reaction rates, how do we quantify them? We might suspect, or know from basic chemistry that a reaction is dependent on temperature and the concentration of all of the species participating in the reaction.

$$r_A \approx f(T, C_i)$$

Based on more than a century of studying chemical kinetics we can lay out some generalizations about chemical reactions to help us understand the general form and properties of reaction rate functions.

2.3.3.1 For single reactions that are *essentially* irreversible, the effects of temperature and concentration are seperable This can be described mathematically as:

$$-r_A = k(T)f(C_i)$$

Here, k, the rate constant is a function of temperature T. Most reactions we will deal with as bioprocess engineers are **not** simple single irreversible reactions, however this generalization is useful in deriving more complicated reaction mechanisms and often provides a useful starting point when designing reactors.

2.3.3.2 The rate constant k follows the Arrhenius equation The Arrhenius equation dictates how typical chemical reaction rates depend on temperature. It has the form

$$k(T) = Ae^{-E/RT}$$

where R is the gas constant, T is the absolute temperature, E is the activation energy, and A is the pre-exponential factor or frequency factor of the reaction. The Arrhenius equation dictates that the rate of a reaction increases exponentially with temperature. On the molecular level temperature reflects the kinetic energy of the molecules within the reactor. The more kinetic energy the molecules in a reactor have, *i.e.* the higher the temperature, the more likely they are to collide, or fluctuate, in the correct orientation to undergo the reaction.

Unfortunately, enzymes—the biological catalysts—are polymers held in their catalytic conformation by non-covalent interactions, which are also sensitive to increases in kinetic energy. At high temperatures enzymes will unfold in a process called thermal denaturation. This leads to a temperature optimum for enzymatic reactions which is typified by the graph below. As temperature increases so does the enzyme activity and rate of reaction, but eventually the enzyme begins to unfold and the activity drops.
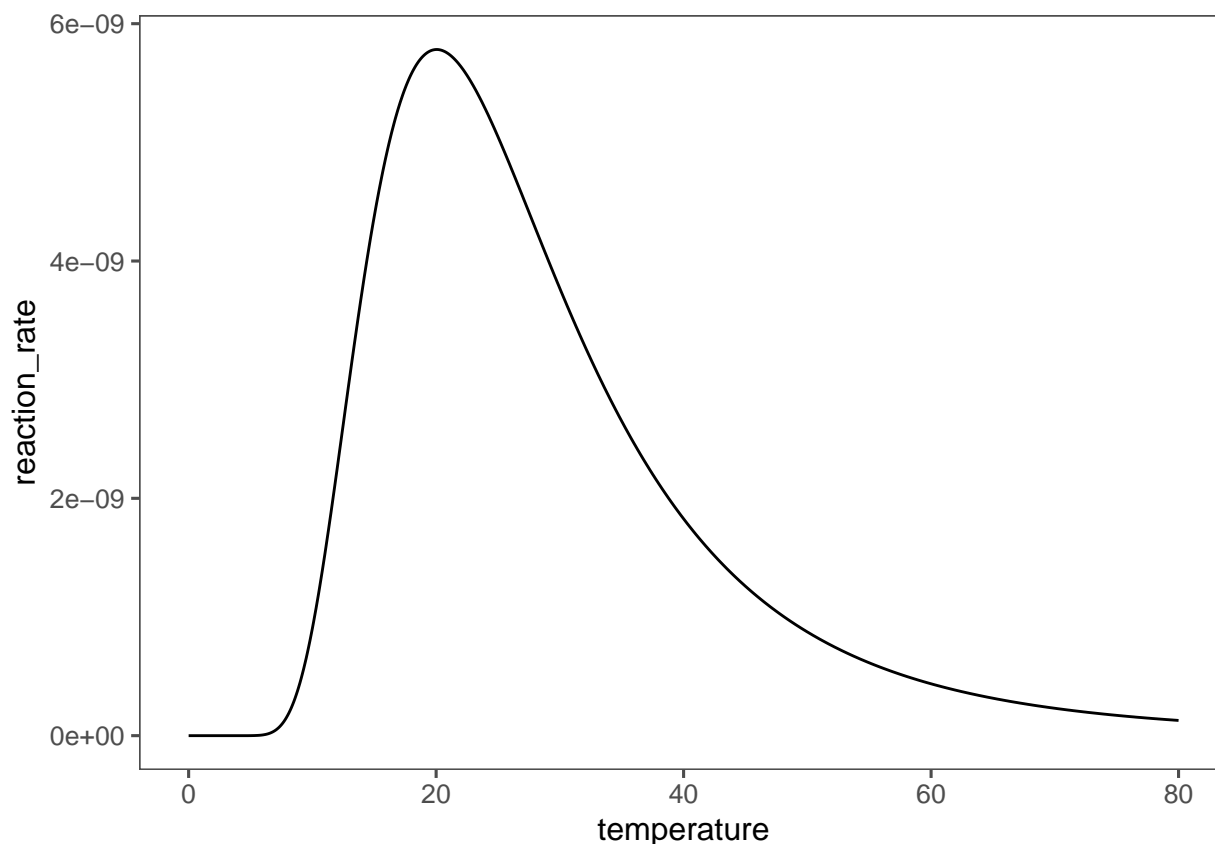
```
temperature <- seq(0, 80, 0.1)
A <- 10000
E <- 1000   #(kJ/mole)
R <- 8.314  #(J/mole/K)
```

```r
K_eq <- 2e-04
reaction_rate <- A * exp(-E/R/temperature)/(1 + K_eq *
    (R * temperature)^6)
data <- tibble(temperature, reaction_rate)
require(ggplot2)
ggplot(data = data, mapping = aes(x = temperature,
    y = reaction_rate)) + geom_line()
```


We will describe this behavior mathematically later on, when we derive the rate equation for enzymes.

2.3.3.3 $f(C_i)$ **decreases as the concentrations of the reactants decrease**   For a given temperature the rate of a reaction will decrease as the concentrations of the reactants decrease. This generalization again is true for most elementary reactions, but is violated in many enzymatic (and other catalytic) reactions, where extremely high substrate concentrations can inhibit some enzymes. We will also discuss this behavior while deriving the rate equations for enzymes.

2.3.3.4 $f(C_i)$ **describes the order of the reaction for each species**   This generalization dictates that $f(C_i)$ can be written as the product of all species raised to their "order" in the reaction.

$$f(C_i) \approx \prod_i C_i^{\alpha_i}$$

where α_i is the order of the reaction for species i. The order of reaction generally ranges from -2 to 2 (fractional values are permissible). The summation $\sum_i \alpha_i$ is referred to as the overall order of reaction.

The order of the reaction for a species reflects the sensitivity of the rate of reaction to a change in concentration of species i. If $\alpha_i = 0$ then the reaction rate is unaffected by the concentration of i. If $\alpha_i < 0$ then the reaction rate is slowed—or inhibited—by i.

Rate equations which obey this behavior are called power-law rate equations. These equations are grounded in thermodynamics governing molecular collisions between reactants and are often a useful starting point for analysis of kinetic data. Using a power law rate equation to describe the reaction:

$$(-\nu_A)A + (-\nu_B)B \longrightarrow \nu_C C + \nu_D D$$
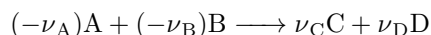
we would propose that

$$f(C_i) = C_A^{\alpha_A} C_B^{\alpha_B} C_C^{\alpha_C} C_D^{\alpha_D}$$

The reaction orders cannot be determined from the stoichiometry of a reaction, *i.e.* $\alpha_i \neq -\nu_i$. There are many ways to write a balanced stoichiometric equation for a reaction, the coefficients can be multiplied by any number. The reaction order reflects the actual behavior and mechanism of the reaction. We cannot assume to know which balanced equation reflects the mechanism, but we can hypothesize a mechanism and check if kinetic data fits. We will do this in a later section.

In the case that the reaction orders match the stoichiometry of the reaction as written, this reaction is called **elementary**. Elementary reactions are critical the development of reaction mechanisms and kinetics. We will discuss elementary reactions more in the next section.

Once again, for enzymatic reactions, this generalization does not hold. We will soon see enzymatic reactions are more representative of multiple reactions. This generalization may hold for the individual reactions comprising the enzymatic mechanism. But for the reaction as a whole, we will soon see, the order of reaction varies with substrate concentration.

2.3.3.5 For reversible reactions, the net rate is the difference between the forward and reverse reactions

$$-r_{A,net} = -r_{A,f} - r_{A,r}$$

Here $-r_{A,net}$ is the total rate at which the reaction proceeds and species A is a reactant as the chemical equation is written. The reaction is moving forward, if $-r_{A,net}$ is positive, because the rate of consumption of A in the forward reaction, $-r_{A,f}$, is greater than the rate of formation of A by the reverse reaction, $r_{A,r}$. Don't forget the convention that rates of reaction are positive and the sign is determined by the stoichiometry of the reaction.

Based on this generalization we can derive the equilibrium expression for a reversible reaction. At equilibrium the net rate of reaction is 0 and we can write out the rates of the forward and reverse reactions according to the order of reaction in each species.

$$0 = -r_{A,f} - r_{A,r} = k_f \prod_i C_i^{\alpha_{f,i}} - k_r \prod_i C_i^{\alpha_{r,i}}$$

Rearranging we define the the the equilibrium constant,

$$K_{eq}^C = \frac{k_f}{k_r} = \prod_i C_i^{(\alpha_{r,i} - \alpha_{f,i})}$$

This was derived completely from the principles of kinetics. An equilibrium constant can also be derived via thermodynamics. As you might remember from thermodynamics:

$$K_{eq} = \prod_i a_i^{\nu_i}$$

where a_i is the activity (or "effective concentration") of species i. This expression for the equilibrium constant depends on the way the stoichiometric coefficients and thus the way the chemical equation is written. For a typical aqueous reaction we can write this equation in terms of concentration

$$K_{eq}^C = \prod_i C_i^{\nu_i}$$

Comparing the thermodynamic and kinetic definitions of K_{eq}^C we can see potential for inconsistency depending on how our chemical equation is written. This implies that in order to be thermodynamically consistent we must define the equilibrium constant based on a single balanced chemical equation. By setting the two versions of the equilibrium constant equal, we can define the condition for this correct equation.

$$\nu_i = \alpha_{r,i} - \alpha_{f,i}$$

Simply, the stoichiometric coefficients must be equal to the order of the total reaction.

2.4 Elementary reactions

The above statement defines an elementary reaction, *i.e.* the order of the forward reaction with respect to species i is equal to the number of molecules of species i participating in each individual reaction. The same is true for the reverse reaction. So for the simple example reaction

$$2\,A \rightleftharpoons B$$
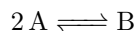
the rate of the forward reaction is given by $-r_{A,f} = k_f C_A^2$, and the rate of the reverse reaction $r_{A,r} = k_r C_B$. Combining the two rate equations we find the net rate of disappearance of A is $-r_{A,net} = k_f C_A^2 - k_r C_B$. In these equation the rate constants are defined in terms of A so the units of k_f are (volume/mole A-time) and the units of k_r are (mole A/mole B-time). Unfortunately, the units first order rate constants are often written just as (time^{-1}).

2.4.0.1 ???Questions??? How would you write the net rate of formation of B?

Using stoichiometry, $r_B = \frac{1}{2}(-r_A)$. Similarly the rate constants can be rewritten in terms of B.

2.4.1 Properties of Elementary Reactions

What assumptions can we make to help us propose elementary reactions?

1. An elementary reaction—a single-step, molecular level event—must be simple. It likely only involves two molecules, as termolecular collisions are very unlikely, and such collisions happening in the correct orientation are even less likely. We can also assume very few bonds will be formed or broken in a single step.

2. Elementary reactions cannot involve fractional molecules.

3. Elementary reactions must obey the principle of microscopic reversibility. Microscopic reversibility means that the reaction must follow the same mechanism in the forward and reverse directions. This means that an elementary reaction must be elementary in both directions.

With this framework for writing elementary reactions we will in Chapter 4 begin deriving mechanisms and rate equations for more complex reactions, which are actually series of multiple elementary reactions.

For now we will assume all of the reactions we are dealing with are elementary, unless we are given a rate equation, and we will begin to examine the kinetics of reactions in ideal reactors.

3 Ideal Reactors

Before jumping into more complex kinetics we will first discuss chemical reactors for simple, elementary reactions. This will give us an understanding of what different reaction orders mean and how reversibility affects dynamics. We will also gain and understanding of different chemical reactors and how reactor choice can effect the reaction outcome.

For any reactor we can consider a small control volume dV, and examine the rate of generation of species i in that control volume:

$$G_i = \iiint_V r_i dV$$

For open systems, e.g. continuous flow reactors we will have some flow rate into and out of this volume element. Remembering your basic material balances:

$$\text{Rate In - Rate Out + Rate of Generation = Rate of Accumulation}$$

$$F_{i0} - F_i + G_i = \frac{dN_i}{dt}$$

Now let's consider three types of ideal reactors that this generalized control volume could be in.

3.1 Batch Reactors

In most cases, examination of a chemical or biochemical reaction will begin with a small scale batch reactor. An ideal batch reactor is defined as a fixed volume, with no flow in or out, that is well mixed such that concentration and temperature are uniform throughout the volume. The concentration and temperature can vary overtime as the reaction proceeds, but typically some temperature-control mechanism will be applied such that the reactor can be considered isothermal.

For batch reactors $F_{i0} = F_i = 0$, such that the rate of generation of species i becomes:

$$G_i = \iiint_V r_i dV = \frac{dN_i}{dt}$$

Since perfect mixing renders r_i independent of position integration is simple yielding

$$r_i = \frac{1}{V} \frac{dN_i}{dt}$$

This equation is called the **design equation** for a batch reactor. We can substitute r_i for the overall rate equation for any reaction or multiple reactions, and if we can separate the variables dependent on species i and time, we can solve this equation to yield an expression for the number of moles of species i versus time.

We can also write this design equation in terms of extent of reaction, as extent of reaction, $\xi = \Delta N_i / \nu_i$.

$$r_i = \frac{\nu_i}{V} \frac{d\xi}{dt}$$

Additionally for single reactions it is sometimes useful to write the design equation in terms of the fraction of the limiting reactant, l that has been consumed, or fractional conversion, x_l.

In this case the design equation is

$$-r_l = \frac{N_{l0}}{V} \frac{dx_l}{dt}$$

If volume is constant, we can also write any of these equations in terms of concentration as $C_i = N_i/V$, *e.g.*

$$r_i = \frac{dC_i}{dt}$$

If there is a catalyst or enzyme in the reactor we can also replace the volume of the reactor with the mass of the catalyst in the reactor in any of these equations.

Solving these differential equations will provide us with an equation for the amount of species i in the reactor at any given time during the reaction. This will then allow us to calculate the amount of time or reactor volume necessary to run a reaction to achieve the desired quantity of product.

Let's do some example calculations.

3.1.1   First Order Reaction Example

Consider a liquid phase reaction A $\longrightarrow$ B that is elementary, therefore the rate equation can be written as $-r_A = r_B = k_1 C_A$. How long must the reactor be run in order for 70% of A to be converted to B if $k_1 = 0.05$ min^{-1} and $C_{A0} = 5$ M?

Our goal is to find the time required for 70% of A to be consumed. This indicates that we should set this problem up in terms of fractional conversion or extent of reaction. So our design equation will simplify to

$$r_A = \frac{dC_A}{dt} = \frac{d[C_{A0}(1 - x_A)]}{dt} = \frac{d(C_{A0} - C_{A0}x_A)}{dt} = 0 - C_{A0}\frac{dx_A}{dt}$$

$$-r_A = C_{A0}\frac{dX_A}{dt}$$

We know what the rate of consumption of A is since the reaction is elementary, so we can plug this relation in for $-r_A$, and also write C_A in terms of fractional conversion, $C_{A0}(1 - x_A)$.

$$k_1 C_{A0}(1 - x_A) = C_{A0}\frac{dx_A}{dt}$$

We can then separate variables

$$k_1 dt = \frac{dx_A}{(1 - x_A)}$$

and finally integrate

$$k_1 \int_0^t dt = \int_0^{0.7} \frac{1}{(1 - x_A)} dx_A$$

to yield

$$t = \frac{-[\ln(1 - 0.7) - \ln(1)]}{k_1}$$

So in this case 24.08 minutes are required to achieve 70% conversion.

What do the dynamics of this reaction look like over time? Let's solve the general form of the integral to find out.

$$k_1 \int_0^t dt = \int_0^{x_A} \frac{1}{(1 - x_A)} dx_A$$

$$t = \frac{-\ln(1 - x_A)}{k_1}$$

$$e^{-k_1 t} = 1 - x_A$$

$$x_A = 1 - e^{-k_1 t}$$

Now we can plot this relationship.

```
k_1 <- 0.05
t <- seq(0, 100, 0.1)
x_A <- 1 - exp(-k_1 * t)
require(tidyverse)
require(magrittr)
data <- tibble(t, x_A)
ggplot(data = data, aes(x = t, y = x_A)) + geom_line() +
    coord_cartesian(ylim = c(0, 1))
```
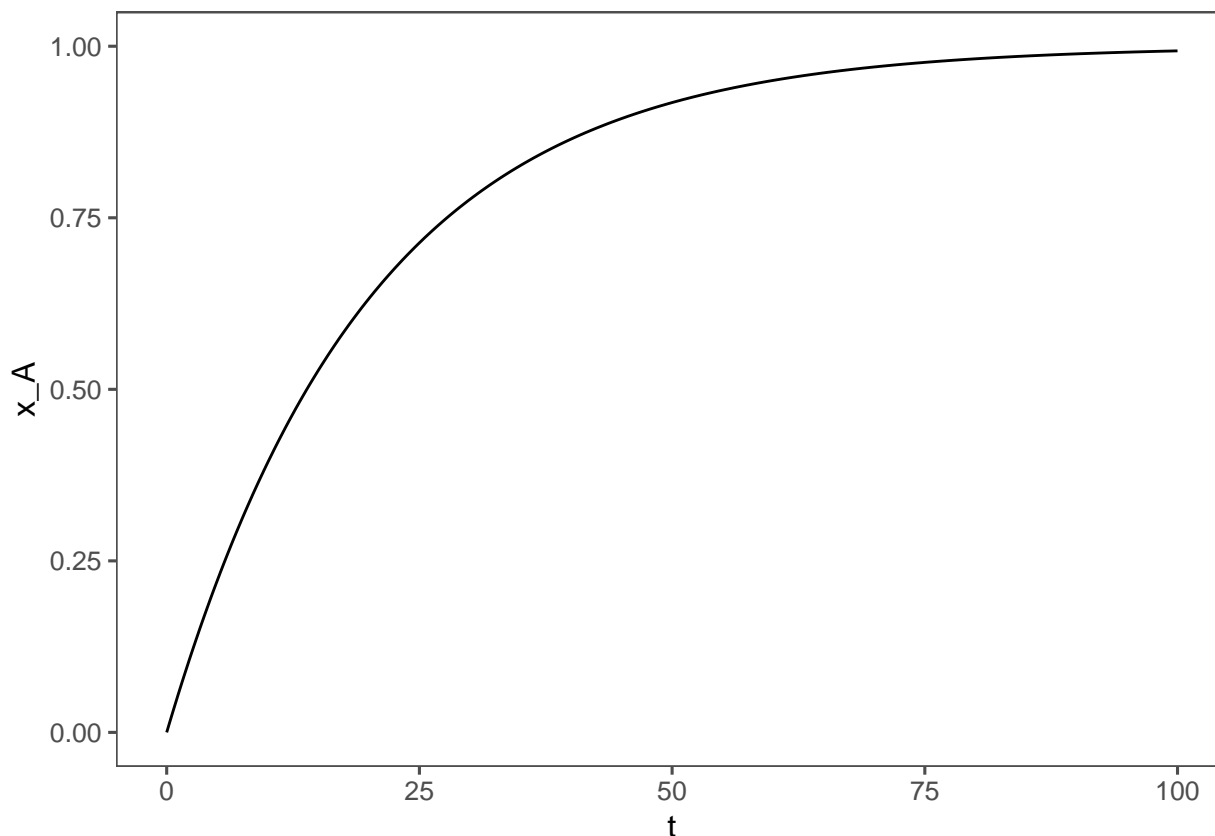
As we calculated we would need to run the reaction for ~24 minutes to achieve 70% conversion.

3.1.1.1   ???Questions???   Does the reaction ever reach 100% conversion?

No, it asymptotically approaches it.


What do the concentrations of A and B look like over time?

For this let's make a stoichiometric table to help us keep everything straight

Species	Initial Concentration	Concentration at t=t
A	C_{A0}	$C_{A0}(1 - x_A)$
B	0	$C_{A0}x_A$

Now we can add these concentrations to our plot.

```
C_A0 <- 5  #M
C_A <- C_A0 * (1 - data$x_A)
C_B <- C_A0 * data$x_A
data <- cbind(data, C_A, C_B)
# we could also use a 'pipe' to send data as the
# first argument data %<>% cbind(C_A, C_B) We
# need to tidy this data up a bit Gather the
# Concentration values into a single column and
# create a new key column Species
```

```
data %<>%
    gather(C_A, C_B, key = "Species", value = "Concentration")
ggplot(data = data, aes(x = t, y = Concentration, color = Species)) +
    geom_line()
```
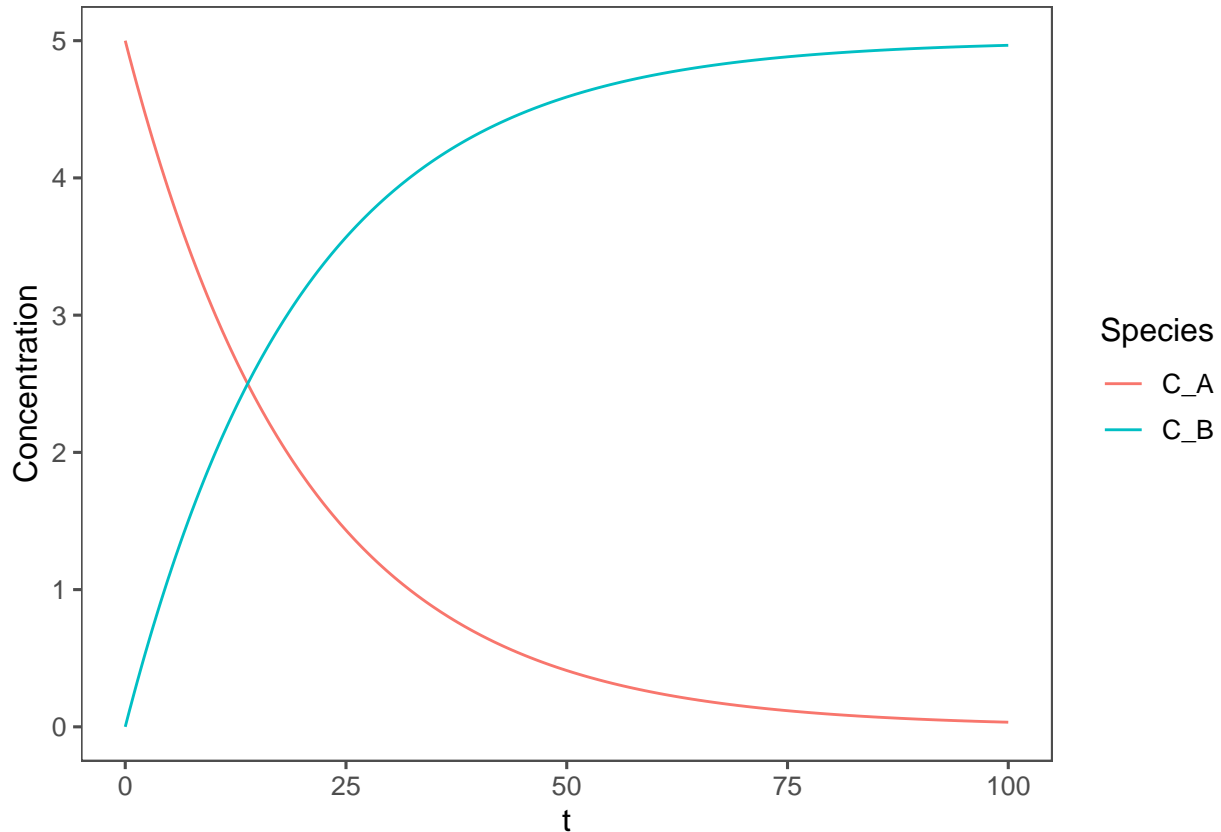

3.1.1.2 ???Questions??? What would the above plot look like if the reaction were zero order? *E.g.* if the reaction was A $\longrightarrow$ B and the rate equation was $-r_A = k_1$

> *The design equation for a zero order batch reaction simplifies to $k_1 = C_{A0}\frac{dx_A}{dt}$. Integrating this yields $x_A = k_1 t$. Graphing this equation as above yields*

```
x_A <- k_1 * t
require(tidyverse)
require(magrittr)
data <- tibble(t, x_A)
C_A <- C_A0 * (1 - data$x_A)
C_B <- C_A0 * data$x_A
data <- cbind(data, C_A, C_B)
data %<>%
    gather(C_A, C_B, key = "Species", value = "Concentration")
ggplot(data = data, aes(x = t, y = Concentration, color = Species)) +
    geom_line() + coord_cartesian(ylim = c(0, 5), xlim = c(0,
    20))
```
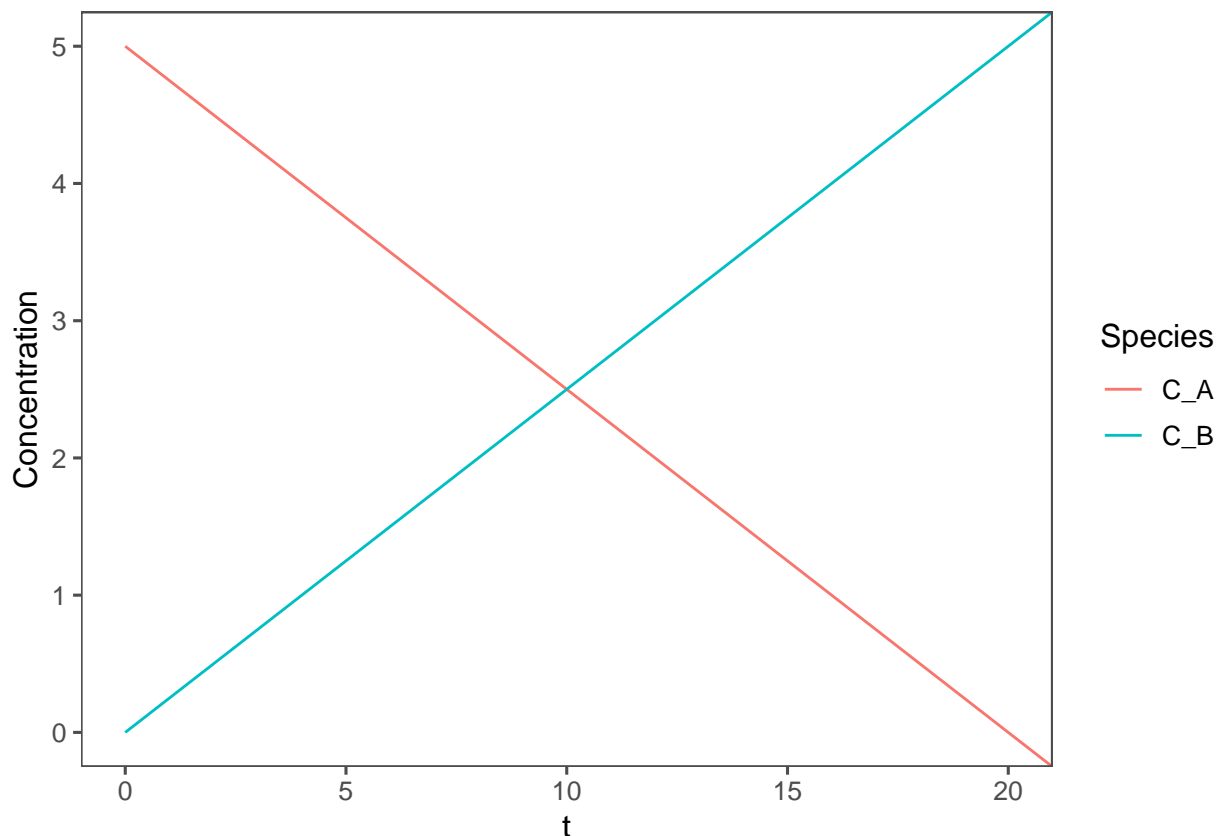
This plot is only valid for the window of time for which C_A is positive, as concentrations cannot be negative. To truly represent zero order reactions across indefinite times a discontinuous function must be used.

3.1.2   Second-order Reaction Example

3.1.2.1   ???Questions???   What would this plot look like if it were second order? *E.g.* if the reaction was $2\,\mathrm{A} \longrightarrow \mathrm{B}$ and the rate equation was $-r_A = k_1 C_A^2$

$$-r_A = C_{A0}\frac{dX_A}{dt}$$

In this case the design equation simplifies to

$$k_1 C_{A0}(1 - x_A)^2 = \frac{dx_A}{dt}$$

$$\int_0^t k_1 C_{A0}\,dt = \int_0^{x_A} \frac{dx_A}{(1 - x_A)^2}$$

$$k_1 C_{A0}t = \frac{1}{1 - x_A} - 1$$

$$1 - x_A = \frac{1}{k_1 C_{A0}t + 1}$$
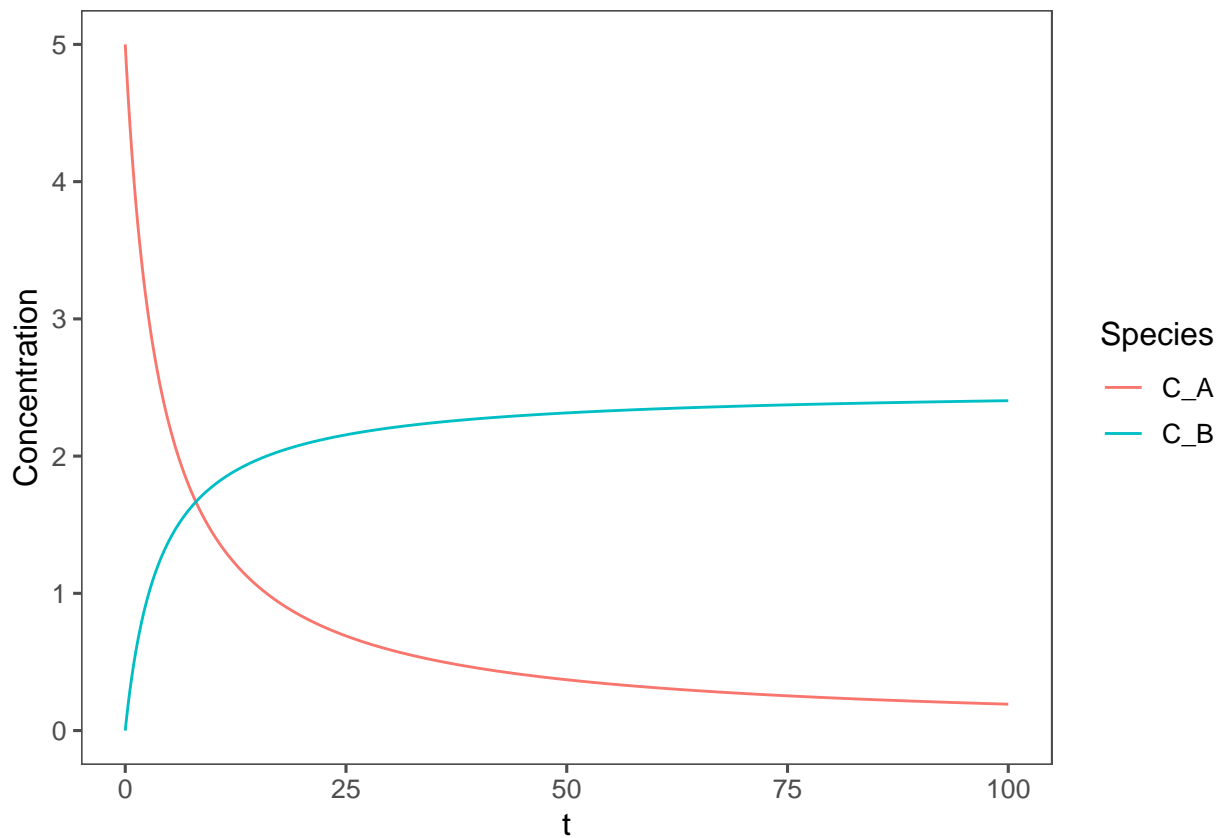
$$x_A = 1 - \frac{1}{k_1 C_{A0}t + 1}$$

Graphing this equation as above yields

```
x_A <- 1 - 1/(k_1 * C_A0 * t + 1)
require(tidyverse)
require(magrittr)
data <- tibble(t, x_A)
C_A <- C_A0 * (1 - data$x_A)
C_B <- C_A0 * data$x_A/2
data <- cbind(data, C_A, C_B)
data %<>%
    gather(C_A, C_B, key = "Species", value = "Concentration")
ggplot(data = data, aes(x = t, y = Concentration, color = Species)) +
    geom_line()
```
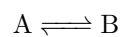

3.1.3 Reversible first-order reaction example

What about if this were first-order and reversible?

$$A \rightleftharpoons B$$

$$-r_A = k_f C_A - k_r C_B$$

Because reversible reactions are actually multiple reactions we cannot use fractional conversion or extent of reaction to measure this reaction, we will just use concentrations.

Let's make a stoichiometric table to start in terms of reactant A.

Species	Initial Concentration	Concentration at t=t
A	C_{A0}	C_A
B	0	$C_{A0} - C_A$

Now we can rewrite the rate equation based on this table.

$$-r_A = k_f C_A - k_r(C_{A0} - C_A)$$

And so the design equation (material balance) is

$$\frac{dC_A}{dt} = k_f C_A - k_r(C_{A0} - C_A)$$

$$\int_0^{C_A} \frac{dC_A}{(k_f + k_r)C_A - k_r C_{A0}} = \int_0^t dt$$

Integrating and solving for C_A we get:

$$C_A = \frac{k_f C_{A0} e^{-(k_f + k_r)t} + C_{A0} k_r}{k_f + k_r}$$

Now we can plot this assuming the same conditions ($k_f = 0.05$ min^{-1} and $C_{A0} = 5$ M) as above, and we can vary the ratio of k_f/k_r.

```
C_A0 <- 5   #M
k_f <- 0.05   #min
kf_kr <- c(10, 1, 0.1)
k_r <- k_f/kf_kr
require(tidyverse)
require(magrittr)
# now we are going to need to apply our C_A
# calculation over each of the k_r values to do
# this we will make C_A a function
C_A <- function(k_f, C_A0, k_r, t) {
    (k_f * C_A0 * exp(-(k_f + k_r) * t) + C_A0 * k_r)/(k_f +
        k_r)
}
# and make a table of all combinations of times
# and k_r's
data <- crossing(k_r, t)
data %<>%
    add_column(k_f, C_A0, .before = TRUE)
# now apply this function over all values of k_r
# and t
data %<>%
    mutate(C_A = unlist(pmap(., C_A)))
data$C_B <- C_A0 - data$C_A
data %<>%
    gather(C_A, C_B, key = "Species", value = "Concentration")
data$kf_kr <- data$k_f/data$k_r
ggplot(data = data, aes(x = t, y = Concentration, color = Species,
    linetype = as.factor(kf_kr))) + geom_line()
```
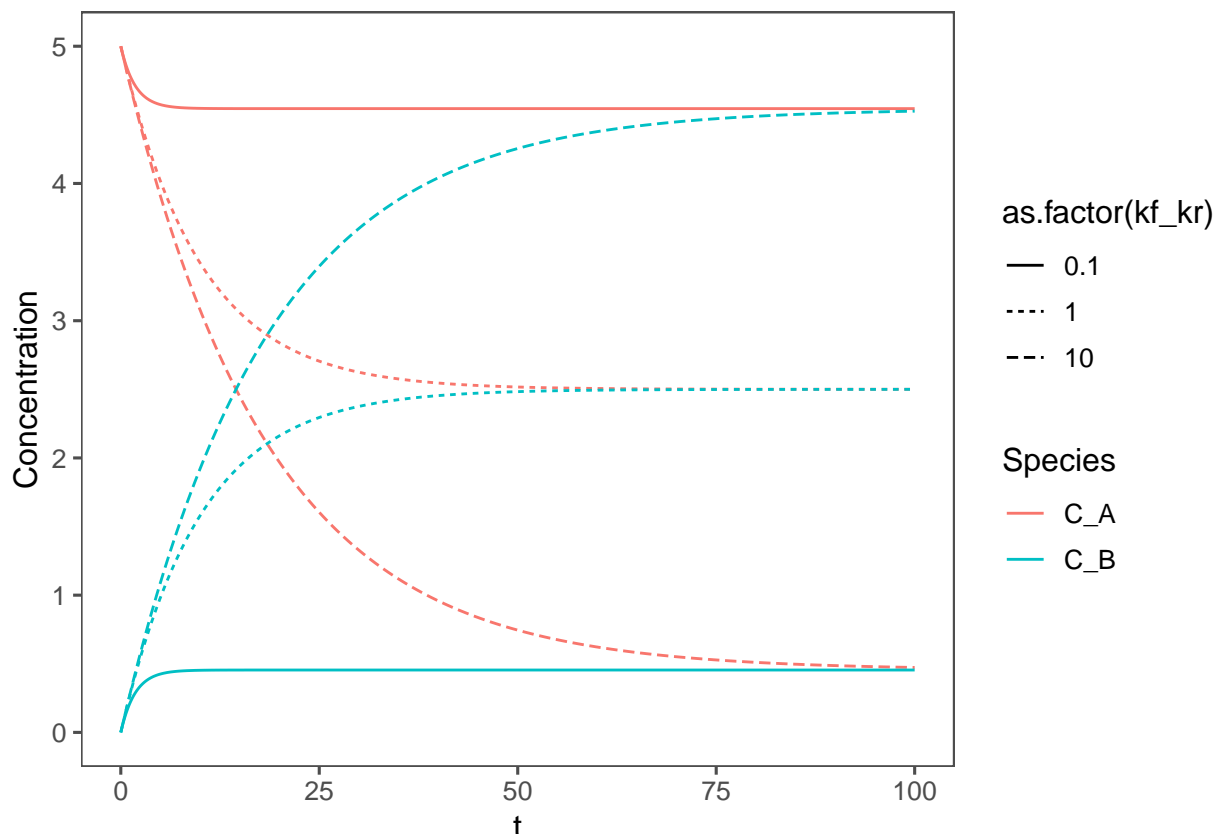
3.2 Continuous-stirred tank reactors

Continuous-stirred tank reactors, or CSTRs, are just like batch reactors except with continuous feed and effluent streams. Ideal mixing is still in effect equalizing temperature and concentration across the contents of the reactor. So to adapt the general equations for a CSTR we simply keep all terms:

$$F_{i0} - F_i + r_i V = \frac{dN_i}{dt}$$

If assumptions such as constant volume or a single reaction are valid we can similarly simplify this equation to be in terms of concentrations or fractional conversion and extent of reaction, respectively. It is always best to start with the general design equation and walk through each assumption to simplify it, instead of jumping straight to a simplified version without fully assessing each assumption.

For CSTRs the rates of reaction are defined at the effluent conditions, which because of perfect mixing are equivalent to the reaction conditions in the entire reactor.

For continuous flow reactors we need to define a quantity that reflects how long the reactants spend in the reactor. Most often a quantity called **space time** is used. Space time is related to the average time that any unit fluid spends in the reactor. Space time at inlet conditions is defined as

$$\tau_0 = V/v_0$$

where v_0 is the inlet volumetric flowrate.

Space time provides an analagous metric to real time in a batch reactor. If a batch reaction is run for a longer time more reactants will be converted to product, and if a continuous reactor is run at higher space time, more reactants will be converted to product in the effluent.

As $F_{i0} = C_{i0}v_0$, for constant volume reactions, that is when fluid density is constant throughout the reaction, the design equation can be written in terms of space time

$$\tau_0 = \frac{C_{i0}x_i}{r_i}$$

Also analagous to batch reactors, catalyst or enzyme weight can also be swapped for reactor volume in the CSTR design equations and definitions of space time. In these cases the catalysts or enzymes are often immobilized on polymeric beads that can easily be filtered from the effluent stream.

3.3 Plug-flow reactors

Alternatively enzyme loaded beads could be packed into a porous bed and reactants could be flowed through this bed. This type of reactor is called a plug-flow reactor. However there does not have to be a packed bed in the reactor at all, you can imagine a pipe with fluid flowing through it and a reaction occuring in that fluid. The further along the length of the pipe the fluid is, the further along the reaction is within that fluid. You might have intuited from that last statement that unlike in batch and continuous stirred tank reactors, in plug-flow reactors the reaction rate varies with position

In order to be an ideal plug flow reactor, two assumptions must be met. 1. There is no mixing in the direction of flow. We can imagine this as little discs or "plugs" of fluid moving through the reactor in a single file line, like miniature batch reactors moving through the reactor. 2. There is no variation in temperature or concentration normal to the direction of flow. *I.e.* each of the plugs is uniform.

For a real reactor to approximate an ideal reactor, the flow must be highly turbulent.

For a plug flow reactor (PFR) our material balance simplifies to

$$F_{i0} - F_i + \iiint_V r_i dV = 0$$

As we said before r_i varies along the length of the reactor and therefore we need to come up with a strategy to solve this integral.

One strategy is to choose a control volume dV such that r_i does not vary within dV. So according to our definition of an ideal plug-flow reactor, each "plug" would be a perfect control volume becasue the concentration and therefore rate of reaction are uniform across this volume perpendicular to flow. Rewriting our material balance across this control volume:

$$F_i - (F_i + dF_i) + r_i dV = 0$$

So our design equation for a PFR is

$$dV = \frac{dF_i}{r_i}$$

Just as with CSTRs, we can define PFRs in terms of space time and concentration.

$$d\tau_0 = C_{A0}\frac{dx_A}{-r_A}$$

## 3.4 Desigining continuous flow reactors

When designing continuous chemical and biochemical reactors we are concerned with the volume of the reactor and flow rate required to meet our production goals. Frequently, a visual representation of the reaction is used to aid in these design decisions for continuous reactors. To derive this visual representation we examine the design equations for a CSTR and PFR.

$$\text{CSTR: } \frac{V}{F_{A0}} = \frac{x_A}{-r_A}$$

$$\text{PFR: } \frac{V}{F_{A0}} = \int_0^{x_A} \frac{dx_A}{-r_A}$$

Now if we plot $1/(-r_A)$ vs x_A, the ratio of the reactor volume to the flow rate for a CSTR is the rectangle where the desired fractional conversion intersects the reaction rate. For a first-order, constant volume, isothermal reaction, this looks like the following:

```
k_1 <- 0.05
C_A0 <- 5   #M
t <- seq(0, 100, 0.1)
x_A <- 1 - exp(-k_1 * t)
require(tidyverse)
require(magrittr)
data <- tibble(t, x_A)
data$nr_A <- k_1 * (C_A0) * (1 - data$x_A)
ggplot(data = data, aes(x = x_A, y = 1/nr_A)) + geom_line() +
    annotate(geom = "rect", xmin = 0, xmax = 0.9, ymin = 0,
        ymax = 1/data$nr_A[min(which(round(data$x_A,
            3) == 0.9))], alpha = 0.5)
```
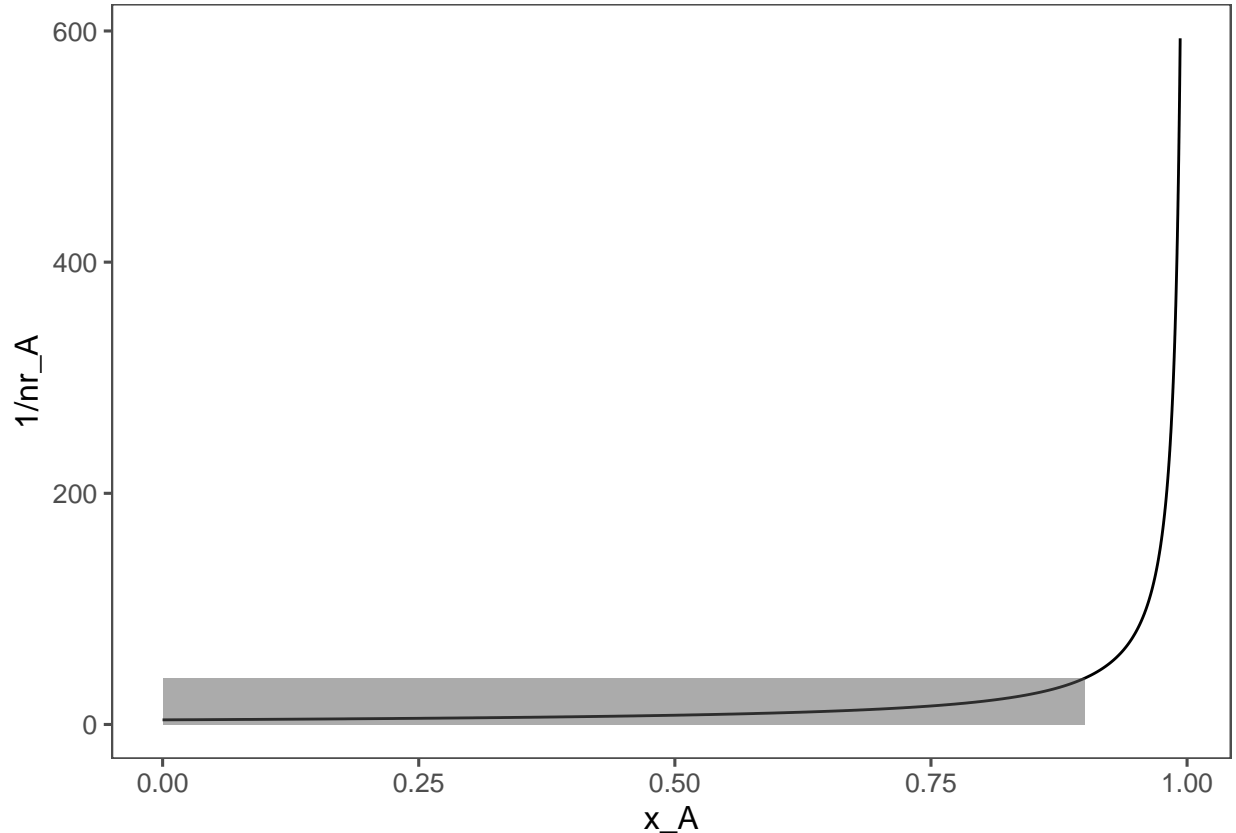

The shaded area in this plot represents the size, V to flow rate F_{A0} ratio of the CSTR required to achieve 90% conversion. For reference, because thinking in terms of the inverse of the rate is tricky, here is what a plot of the rate equation looks like, $-r_A$ vs. x_A:

```r
ggplot(data = data, aes(x = x_A, y = nr_A)) + geom_line()
```
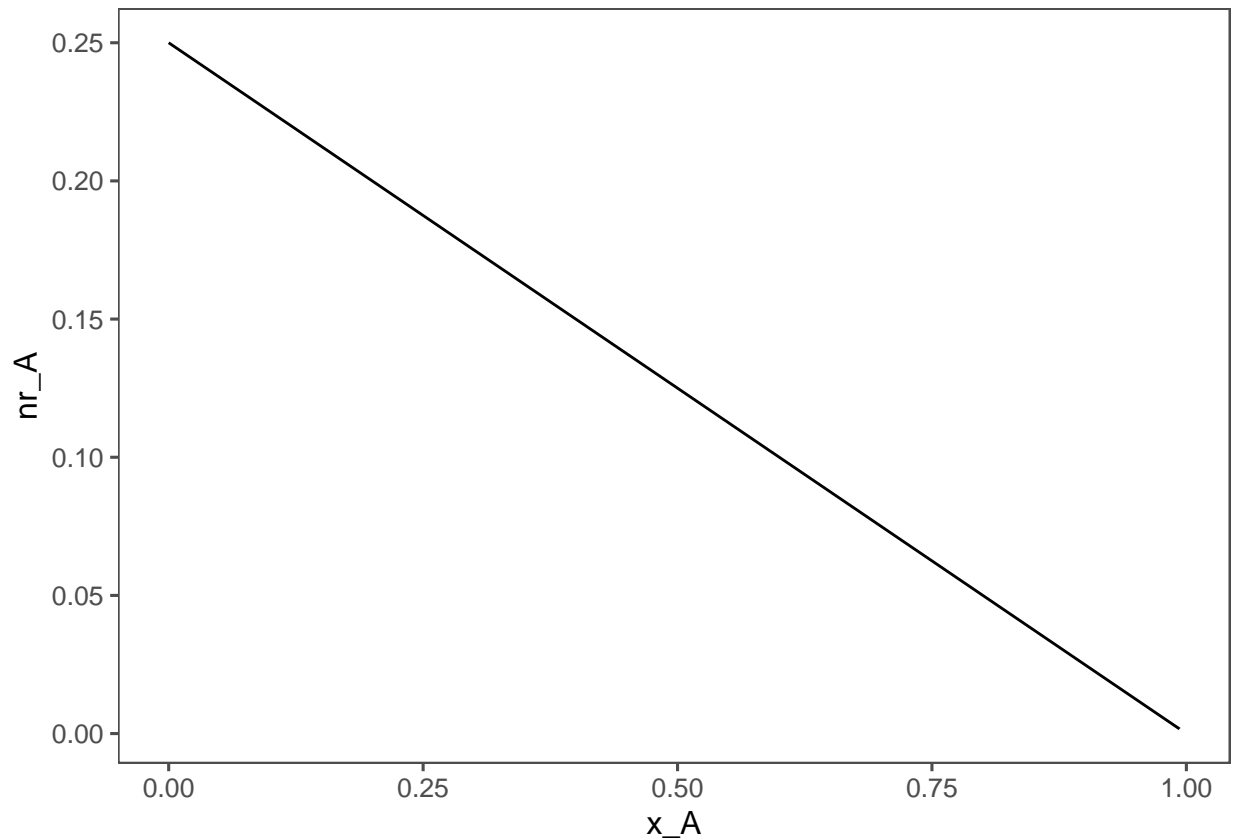

Looking at the PFR equation we can see that instead of a rectangle the V/F_{A0} ratio is represented by the area under the curve.

```r
ggplot(data = data, aes(x = x_A, y = 1/nr_A)) + geom_line() +
    annotate(geom = "rect", xmin = 0, xmax = 0.9, ymin = 0,
        ymax = 1/data$nr_A[min(which(round(data$x_A,
            3) == 0.9))], alpha = 0.5) + geom_area(mapping = aes(x = ifelse(x_A >=
    0 & x_A <= 0.9, x_A, 0)), fill = "red", alpha = 0.5) +
    ylim(c(0, 600))
```
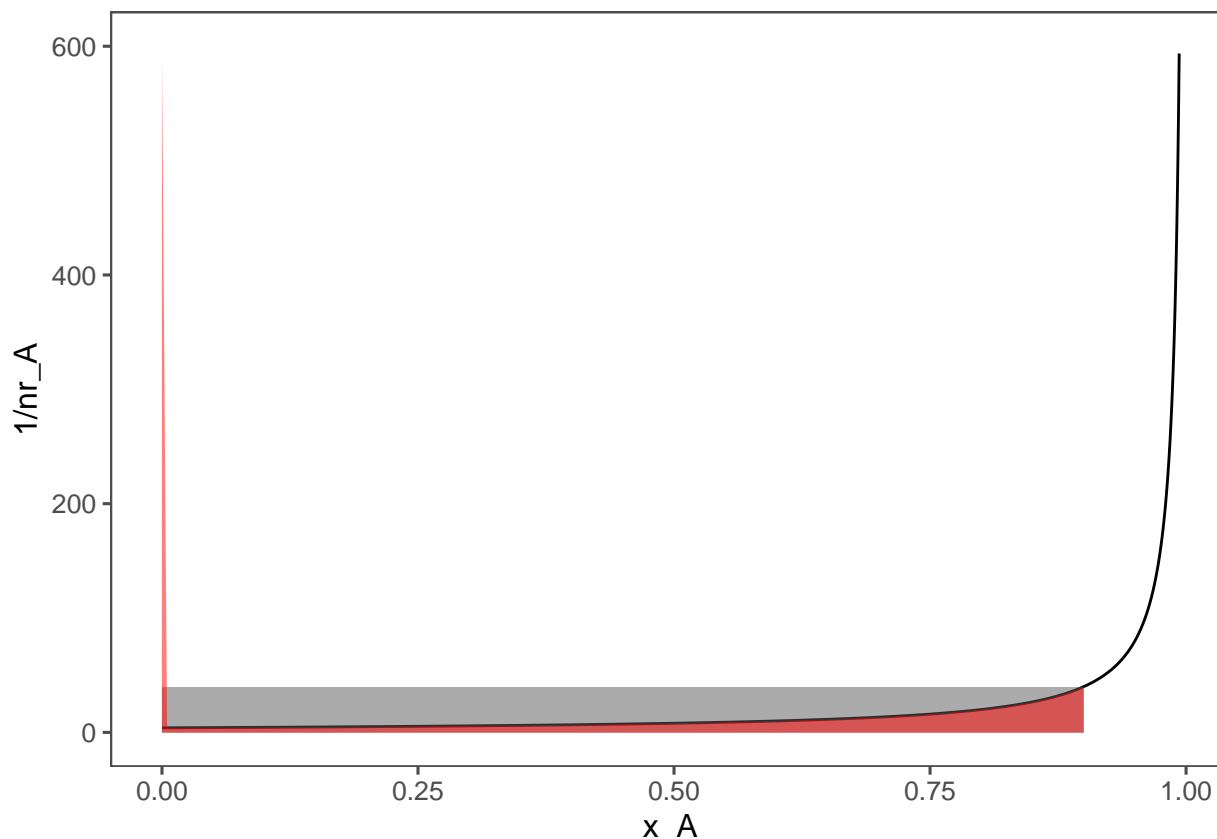
So given a constant flowrate between the two a PFR reactor will be significantly smaller than a CSTR for the same reaction.

3.4.0.1 ???Questions??? Explain qualitatively why a CSTR needs a larger volume to achieve the same conversion? Hint: what is the average reaction rate in a CSTR vs a PFR?

> *In a CSTR the reaction rate is always constant as the entire volume is always well mixed. As soon as reactants enter the reactor they are diluted to the effluent conditions, which determine the reaction rate throughout the whole reactor. This dilution does not happen In a PFR, so reactants can enter the reactor at a much higher concentration and therefore the reaction can occur at a much higher initial rate. As the reactants move through the reactor they gradually approach the concentration at which the same reaction would be conducted in a CSTR, finally reaching this concentration as the reach the end of the reactor.*

---

4 Kinetics: Determining Rate Equations for Chemical and Biochemical Reactions

4.1 Remembering Elementary Reactions

Remember from Chapter 1 that if a reaction is elementary we can write the rate of this reaction based on its stoichiometry.

Here, we will make use of this concept to make mathematical models quantifying the kinetics (or reaction rate equations) of complex chemical reactions as well as simple and complex enzymatic reaction. In later

chapters we will also apply the ideas developed here to model cells growing in a bioreactor. But before we do that let's discuss some properties of elementary reactions.

What assumptions can we make to help us propose elementary reactions? What are the rules/properties of elementary reactions?

1. An elementary reaction must be simple—a single-step, molecular level event. It likely only involves two molecules, as termolecular collisions are very unlikely, and such collisions happening in the correct orientation are even less likely. We can also assume very few bonds will be formed or broken in a single step.

2. Elementary reactions cannot involve fractional molecules.

3. Elementary reactions must obey the principle of microscopic reversibility. Microscopic reversibility means that the reaction must follow the same mechanism in the forward and reverse directions. This means that an elementary reaction must be elementary in both directions.

4.1.1 Example of an elementary reaction

As an example of an elementary reaction let's consider the decomposition of ozone:

$$O_3 \longrightarrow O_2 + O^\bullet$$
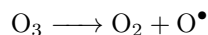
In reality, ozone does not decompose spontaneously, ozone collides with dioxygen and the energy of this collision drives decomposition. Thus the reaction written above is not elementary, because it is not written exactly how the reaction happens at the molecular level. We can write the elementary reaction as:

$$O_3 + O_2 \longrightarrow O_2 + O_2 + O^\bullet$$
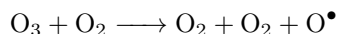
This equation and that above are both stoichiometrically correct, but the later equation describes the elementary reaction. How do we know this is the elementary reaction? From the rate equation. Fitting experimental data on the decomposition of ozone to the equation for the elementary reaction, $-r_{O_3} = kC_{O_3}C_{O_2}$, yields a much better fit than $-r_{O_3} = kC_{O_3}$. We will discuss fitting reaction data to rate equations later on.

This iterative process of hypothesizing elementary reactions, based on plausible reaction mechanisms, and fitting experimental data to rate equations derived from these elementary reactions is the basis of chemical and biochemical kinetics.

With this framework for writing elementary reactions in place we can now begin deriving mechanisms and rate equations for more complex reactions, which are actually series of multiple elementary reactions. Let's look at another more complex example of how a stoichiometrically simple reaction can be broken down into elementary reactions by hypothesizing a mechanism.

4.2 Elementary Reactions Provide Feasible Mechanisms

Most seemingly simple reactions are not elementary as written but can be broken down into a sequence of elementary reactions. Take for example the Haber-Bosch process, by which most nitrogen for fertilizer is fixed from the atmosphere.

$$N_2 + 3\,H_2 \rightleftharpoons 2\,NH_3$$
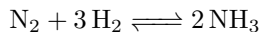
4.2.0.1 ???Questions???
How many molecules are involved in the forward and reverse reactions? How many bonds are formed/broken in the forward and reverse reactions? Is this reaction likely elementary as written? Why or why not?

In the forward reaction 4 species are required to interact simultaneously, which is very unlikely. In the reverse reaction only two species must interact. In the forward reaction 4 bonds must be broken and 6 bonds must be formed, and the opposite is true for the reverse. No, this reaction is not likely elementary as written as the complex interactions and bond rearrangments are highly unlikely in a single step.

So how does this reaction happen? The Haber-Bosch process relies on a iron catalyst, which facilitates much of this bond breaking and forming by allowing hydrogen and nitrogen species to interact with the catalyst. Let's call the catalyst here S for so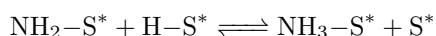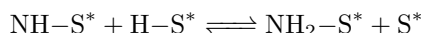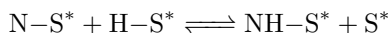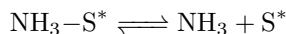lid catalyst. S^* will represent a open binding site on the catalyst surface. Now let's write elementary reactions that may constitute a mechanism for the Haber-Bosch process. Initially H_2 and N_2 will adsorb onto the catalyst surface.

$$H_2 + 2\,S^* \rightleftharpoons 2\,H{-}S^*$$

$$N_2 + 2\,S^* \rightleftharpoons 2\,N{-}S^*$$
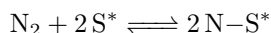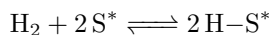
These forward reactions appear to involve three species, however as our catalyst is solid $2\,S^*$ act as essentially a single species. Now to form ammonia these adsorbed species must react.

$$N{-}S^* + H{-}S^* \rightleftharpoons NH{-}S^* + S^*$$

$$NH{-}S^* + H{-}S^* \rightleftharpoons NH_2{-}S^* + S^*$$

$$NH_2{-}S^* + H{-}S^* \rightleftharpoons NH_3{-}S^* + S^*$$

And finally the ammonia must desorb from the surface.

$$NH_3{-}S^* \rightleftharpoons NH_3 + S^*$$

This proposed mechanism is as simple as possible. The final desorbtion step may happen essentially simultaneously with the final $N{-}H$ bond formation. We may be able to test this hypothesis with kinetic data or assess under what conditions this assumption might be fair once we have derived a rate equation. But before we derive a rate equation, we should check that the stoichiometry of this series of reactions is equivalent to the complete Haber-Bosch reaction stoichiometry.

We can also write the above balanced chemical reactions as equations in a similar way to how we made a stoichiometric table which can be helpful in dealing with multiple reactions in parallel or series.

$$\sum_i \nu_i A_i = 0$$

where A_i is each chemical species i participating in the reaction, and ν_i is the stoichiometric coefficient of species i.

We can write the above reactions in this form and then sum these equations to determine the net overall reaction.

$$-\,3\,H_2 - 6\,S^* + 6\,H{-}S^* = 0$$

$$-\,N_2 - 2\,S^* + 2\,N{-}S^* = 0$$

$$-\,2\,N{-}S^* - 2\,H{-}S^* + 2\,NH{-}S^* + 2\,S^* = 0$$

$$-\,2\,NH{-}S^* - 2\,H{-}S^* + 2\,NH_2{-}S^* + 2\,S^* = 0$$

$$-\,2\,NH_2{-}S^* - 2\,H{-}S^* + 2\,NH_3{-}S^* + 2\,S^* = 0$$

$$-\,2\,NH_3{-}S^* + 2\,NH_3 + 2\,S^* = 0$$

We have to multiply these equations to match the stoichiometry of the overall reaction, but in the end you see that the catalyst sites will cancel out along with the other intermediate species yielding the overall net reaction.

$$-\,N_2 - 3\,H_2 + 2\,NH_3 = 0$$

Now that we have a mechanism made of elementary reactions that is consistent with the overall reaction, we can write reaction rates for each species in terms of the formation or consumption of that species in each elementary reaction. Then we can solve this set of equations for the rate of the overall reaction in terms of whichever species we'd like, usually the desired product. We won't do that for this very complicated reaction, but will now jump to simple enzyme kinetics as an example.

4.3 Summary: Kinetics

Given some non-elementary chemical/biochemical reaction we can define a rate equation for this reaction by the following procedure:
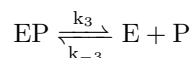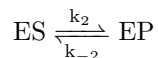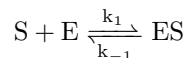
1) Define a series of elementary reactions, which are plausible and sum up to the overall reaction. Assign rate constants to each reaction (and direction).
2) Write out the differential mass balances for each species in the series of reactions. For each species balance look at each reaction and if that species is being generated or consumed by that elementary reaction, write the appropriate rate for that reaction based on the rule of "order = stoichiometry" that applies for elementary reactions.
3) You now have a system of differential equations you can use to simulate the reaction!
4) To derive a rate equation, pick a species that you want to derive a rate for, typically the product or reactant. Identify the intermediate/unknown/unmeasureable concentrations in the mass balance for that species. You need to replace these with equations derived from the other balances in terms of known quantities.
5) Make assumptions to help simplify the system of equations into a single equation (degree of freedom analysis helps here). Is there any step/reaction that might be in equilibrium? What intermediates could you assume to be at steady state? Be sure to note your assumptions and think about when and where they are valid.
6) Simplify the the system of equations you have in terms of the intermediates you need to get rid of in your rate equation and plug them in.
7) Lump together any combinations of parameters in your equation and define them.

4.4 Deriving enzymatic rate equations

So based on our outlined procedure let's write out some elementary reactions describing how a simple enzyme works. Let's assume an enzyme, E converts a substrate, S to product, P:

$$S \xrightarrow{\text{E}} P$$
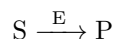
To break this overall reaction up into elementary reactions we need to think about each molecular interaction necessary. First, the enzyme and substrate need to come together in a binding reaction. Then the substrate can be converted to the product in a catalytic step, but still remain bound to the enzyme. Finally the product dissociates from the enzyme. (In the biochemical literature binding and association, and unbinding and dissociation, are used interchangeably). For now all of these reactions could be reversible. So writing out this series as chemical reactions yields the following.

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

$$ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} EP$$

$$EP \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} E + P$$

In this series of equations we have 5 different species—3 of which we probably can't measure, the active centers E (free enzyme), ES, and EP—and 3 independent reactions (reverse reactions are not independent of the forward). We also have 6 unknown parameters, all of the k's. So in terms of degrees of freedom this system is solvable for a rate in terms of S and 4 parameters (these will be combinations of the k's: 6 unknown parameters + 3 unknown variables - 5 equations (one per species) = 4 degrees of freedom. We could write a rate equation for this system if we assume the reactive centers are at steady-state, but the algebra will definitely be tedious. Also, once we get to the final rate equation we may have more parameters than our data can accurately fit. But it is possible to simulate the set of differential equations resulting from this series of reactions relatively easily. We still have many parameters, but we can just give those reasonable and interesting values for demonstration purposes.

So let's write out these equations. We can begin with the general balance for a batch reactor, and write the accumulation term for each species. The for each species balance look at each reaction above and if that species is being generated or consumed by that elementary reaction, write the appropriate rate for that reaction based on the rule of "order = stoichiometry" that applies for elementary reactions. Don't forget the reverse reactions!

$$\text{Accumulation} = \text{Generation} \qquad -\text{Consumption}$$

$$\frac{d[S]}{dt} = k_{-1}[ES] \qquad -k_1[E][S]$$

$$\frac{d[E]}{dt} = k_{-1}[ES] + k_3[EP] \qquad -k_1[E][S] - k_{-3}[E][P]$$

$$\frac{d[ES]}{dt} = k_1[E][S] + k_{-2}[EP] \qquad -k_{-1}[ES] - k_2[ES]$$

$$\frac{d[EP]}{dt} = k_2[ES] + k_{-3}[E][P] \qquad -k_{-2}[EP] - k_3[EP]$$

$$\frac{d[P]}{dt} = k_3[EP] \qquad -k_{-3}[E][P]$$

Now to simulate this set of equations we need to define these equations as a differential function.

```r
## function defining our differential equations
complete_enzyme_model <- function(time, curr_state,
    parms) {
    with(as.list(c(curr_state, parms)), {
        ## The with() function gives access to
        ## the named values of parms and
        ## curr_state within the local function
        ## environment
        dSdt = k_n1 * ES - k_1 * E * S
        dEdt = k_n1 * ES + k_3 * EP - k_1 * E * S -
            k_n3 * E * P
        dESdt = k_1 * E * S + k_n2 * EP - k_n1 * ES -
            k_2 * ES
        dEPdt = k_2 * ES + k_n3 * E * P - k_n2 * EP -
            k_3 * EP
        dPdt = k_3 * EP - k_n3 * E * P
        ## the computed derivatives are returned
        ## as a list order of derivatives needs
        ## to be the same as the order of species
        ## in curr_state
        list(c(dSdt, dEdt, dESdt, dEPdt, dPdt))
    })
}
```

Now we can set the parameter, k values and initial conditions.

```r
## A vector containing the parameter values
parms <- c(k_1 = 0.4, k_n1 = 0.004, k_2 = 40, k_n2 = 0.003,
    k_3 = 1, k_n3 = 0.003)
init_state <- c(S = 5, E = 1, ES = 0, EP = 0, P = 0)
time <- 0

library(tidyverse)
delta.t <- 0.1   ## 0.1 hours
```

```
time.out <- seq(0, 20, by = delta.t)

ts.complete_enzyme_model <- data.frame(deSolve::lsoda(y = init_state,
    times = time.out, func = complete_enzyme_model,
    parms = parms))

ts.complete_enzyme_model <- tidyr::pivot_longer(data = ts.complete_enzyme_model,
    cols = c("S", "E", "ES", "EP", "P"), names_to = "Species",
    values_to = "Concentration")

(complete <- ggplot(data = ts.complete_enzyme_model,
    mapping = aes(x = time, y = Concentration, color = Species)) +
    geom_line())
```
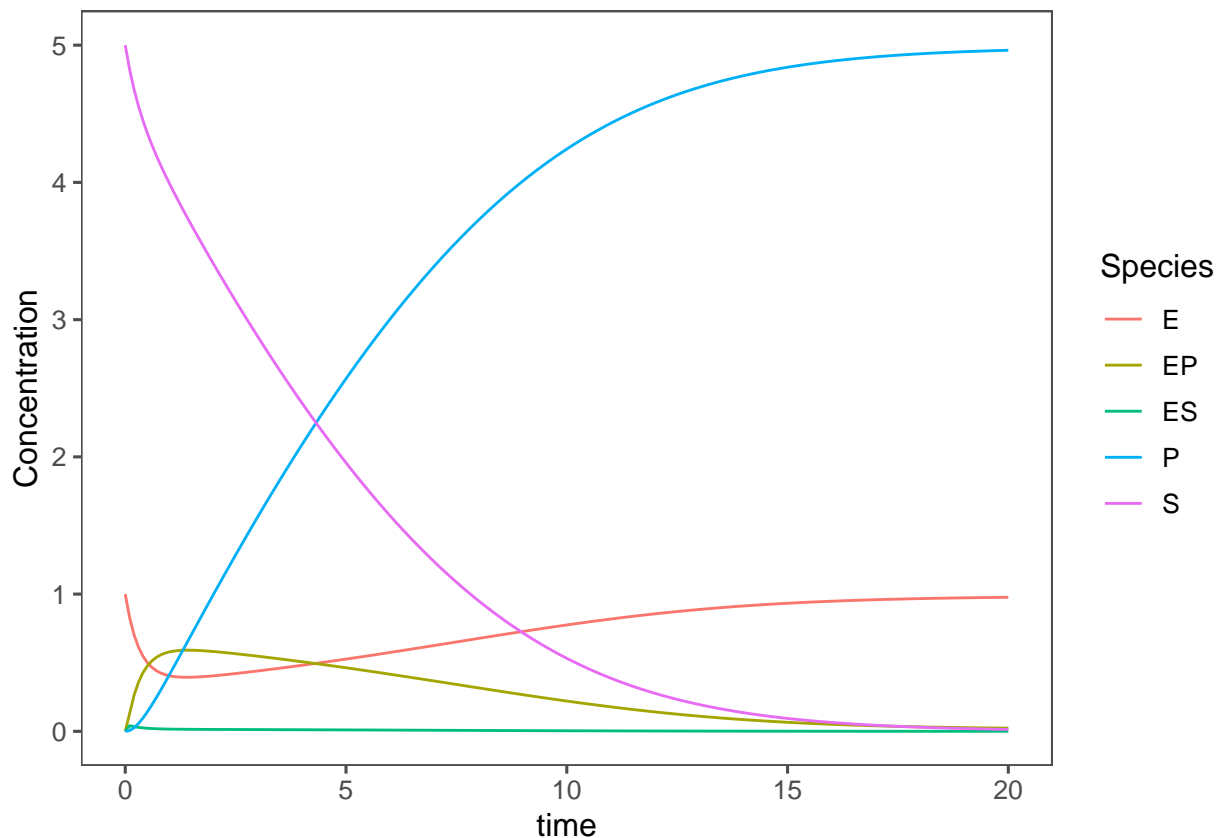

This now gives us a really nice view of each complex and allows us to make some assumptions to simplify our model of enzyme kinetics and check the validity of these assumptions. Bioprocess Engineers and Biochemists typically use a much simpler mechanistic model built on several valid assumptions to describe enzymatic reactions.

To simplify this enzyme reaction mechanism a bit we can make the following assumptions that are likely true for most enzymes: 1) Catalysis is irreversible. This is particularly true in the case inside of cells when products of an enzyme are often further acted upon by other enzymes in a metabolic pathway. It also holds true for the early stages of *in vitro* reactions, when substrate concentrations are high and product concentrations are low. It is also a valid assumption when the ΔG of the reaction is large. 2) EP is short lived, enzyme has low affinity for the product. This is generally true for most natural enzymes as in order to bind more substrate and complete another reaction, the product must first be released. So in general

enzymes will have much lower binding affinity for their products than their substrates.

$$S + E \xrightleftharpoons[k_{-1}]{k_1} ES$$
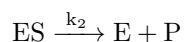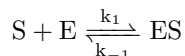
$$ES \xrightarrow{k_2} E + P$$

$$\text{Accumulation} = \text{Generation} \qquad -\text{Consumption}$$

$$\frac{d[S]}{dt} = k_{-1}[ES] \qquad -k_1[E][S]$$

$$\frac{d[E]}{dt} = k_{-1}[ES] + k_2[ES] \quad -k_1[E][S]$$

$$\frac{d[ES]}{dt} = k_1[E][S] \qquad -k_{-1}[ES] - k_2[ES]$$

$$\frac{d[P]}{dt} = k_2[ES]$$

```r
## function defining our differential equations
simplified_enzyme_model <- function(time, curr_state,
    parms) {
    with(as.list(c(curr_state, parms)), {
        ## The with() function gives access to
        ## the named values of parms and
        ## curr_state within the local function
        ## environment
        dSdt = k_n1 * ES - k_1 * E * S
        dEdt = k_n1 * ES + k_2 * ES - k_1 * E * S
        dESdt = k_1 * E * S - k_1 * ES - k_2 * ES
        dPdt = k_2 * ES
        ## the computed derivatives are returned
        ## as a list order of derivatives needs
        ## to be the same as the order of species
        ## in curr_state
        list(c(dSdt, dEdt, dESdt, dPdt))
    })
}
```

```r
## A vector containing the parameter values
parms <- c(k_1 = 0.4, k_n1 = 0.004, k_2 = 40)
init_state <- c(S = 5, E = 1, ES = 0, P = 0)
time <- 0

library(tidyverse)
delta.t <- 0.1  ## 0.1 hours

time.out <- seq(0, 20, by = delta.t)

ts.simplified_enzyme_model <- data.frame(deSolve::lsoda(y = init_state,
    times = time.out, func = simplified_enzyme_model,
    parms = parms))

ts.simplified_enzyme_model <- tidyr::pivot_longer(data = ts.simplified_enzyme_model,
    cols = c("S", "E", "ES", "P"), names_to = "Species",
    values_to = "Concentration")
```

```
(simplified <- ggplot(data = ts.simplified_enzyme_model,
    mapping = aes(x = time, y = Concentration, color = Species)) +
    geom_line())
```
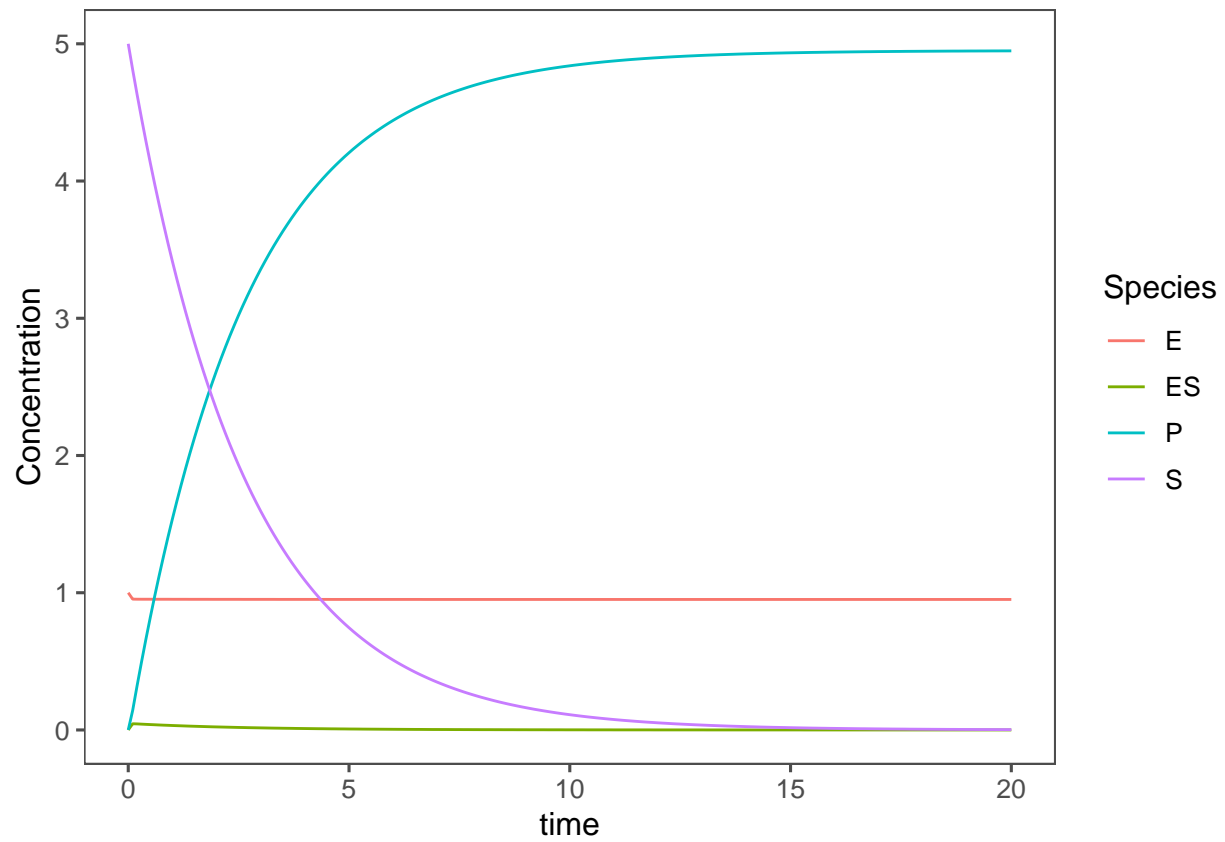

Plotting these next to each other for comparison

```
library(patchwork)
complete + simplified
```
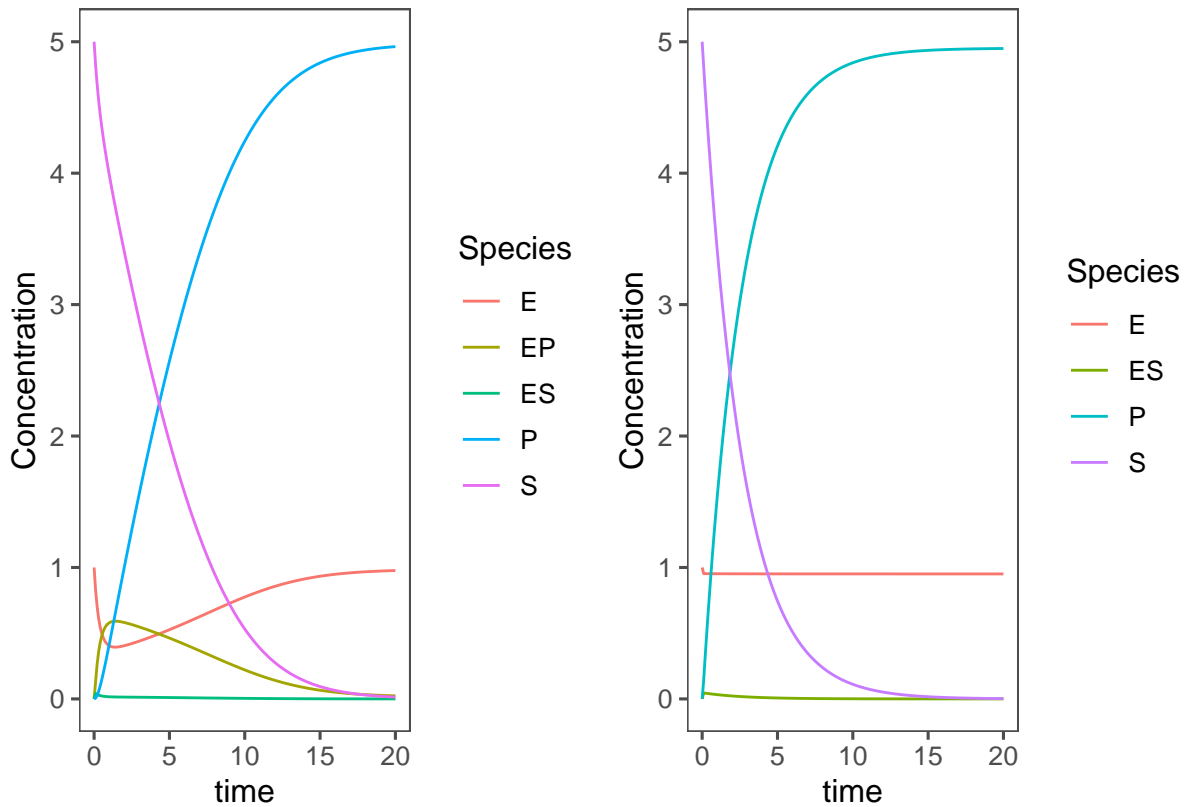
Getting rid of the internal states does, of course, speed up the rate of production of product/consumption of substrate, but we didn't adjust our parameters at all. The k_{-2}, k_3, and k_{-3} values are all now lumped into k_2 or as it is more typically referred to in this Michaelis-Menten model, k_{cat}. We could change these parameters to better match the original complete model.

```r
## A vector containing the parameter values
parms <- c(k_1 = 0.3, k_n1 = 0.004, k_2 = 5)
## And the initial values and time points
init_state <- c(S = 5, E = 1, ES = 0, P = 0)
time <- 0
delta.t <- 0.1   ## 0.1 hours
time.out <- seq(0, 20, by = delta.t)

## Now run the model
ts.simplified_enzyme_model <- data.frame(deSolve::lsoda(y = init_state,
    times = time.out, func = simplified_enzyme_model,
    parms = parms))
## and reformat it for plotting
ts.simplified_enzyme_model <- tidyr::pivot_longer(data = ts.simplified_enzyme_model,
    cols = c("S", "E", "ES", "P"), names_to = "Species",
    values_to = "Concentration")

## and finally plot it
simplified <- ggplot(data = ts.simplified_enzyme_model,
    mapping = aes(x = time, y = Concentration, color = Species)) +
    geom_line()
complete + simplified
```
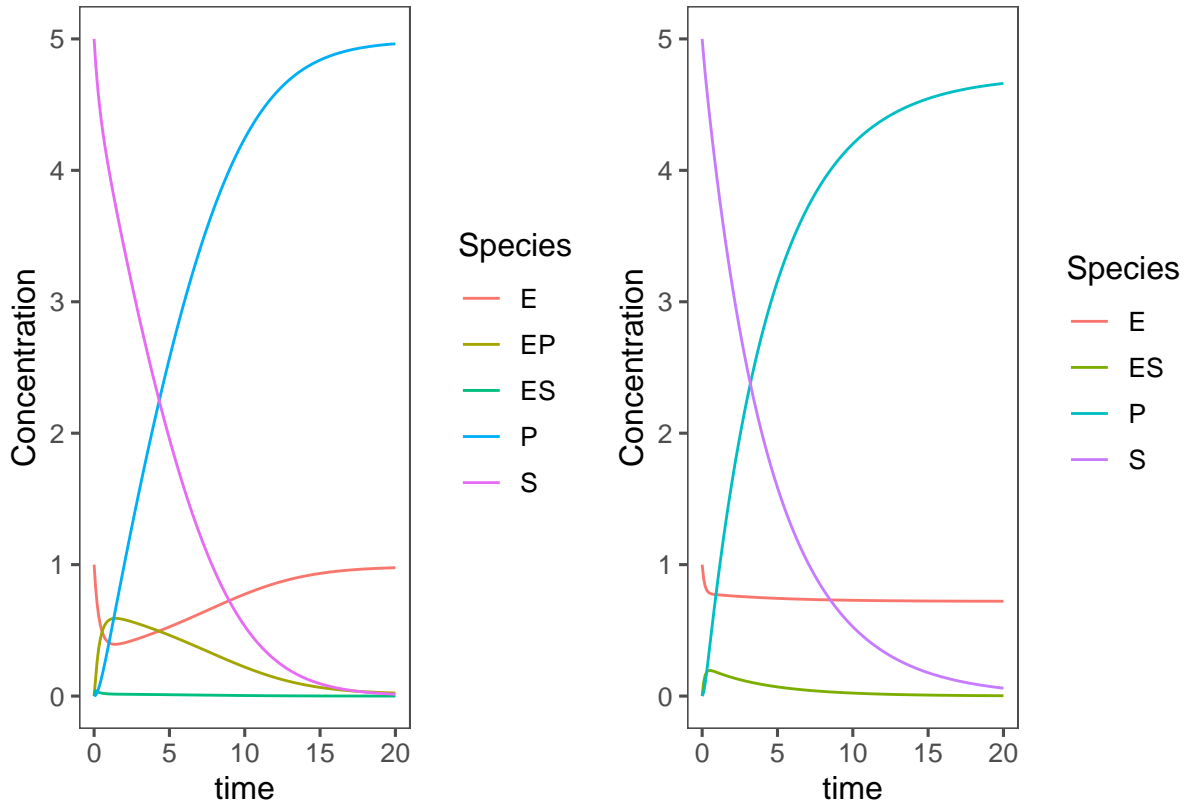
Check out this nice shiny app and play with the sliders to get a feel for what changing each parameter does to the model.

5   Determining Enzyme Rate Parameters

Now that we have derived plausible rate equations for simple enzymes we need to test how these model equations compare to real experimental reaction rates.

Taking a look at our model:

$$\nu = \frac{V_m S}{K_m + S}$$

where $V_m = k_{cat} E_0 = k_2 E_0$, we find two parameters, V_m and K_m which we need to estimate based on experimental data. Although with today's computational power these parameters could be estimated in many possible ways with non-linear regression, historically and still often today linearizations of this enzyme kinetic model are used to fit these parameters to experimental data.

As an aside, we must discuss once again nomenclature issues in the biochemical literature. Unfortunately there is no consensus on what symbol should be used for reaction velocity. In some texts the greek symbol, ν (nu), is used in others V, is used. V can be easily confused with the volume of a reactor, so be sure differentiate your notation when sizing reactors.

5.1 Linearization methods

5.1.1 Lineweaver-Burk plot

One method of linearization would be to invert the entire model equation.

$$\frac{1}{\nu} = \frac{1}{V_m} + \frac{K_m}{V_m}\frac{1}{S}$$

The resulting plot is called a Lineweaver-Burk plot.

5.1.1.1 ???Questions??? What would you plot as x and y, and what would be the slope and y-intercept of the plot be.

A Lineweaver-Burke plot is $1/\nu$ (y) vs $1/S$ (x). The slope is K_m/V_m and the y-intercept is $1/V_m$

This plot is also sometimes called a double-reciprocal plot, because we plot the reciprocal of our independent variable ν versus our dependent variable S.

To use this plot to find V_m and K_m we need to collect reaction velocity data ($\nu = dP/dt = -dS/dt$) at different initial substrate concentrations. The reaction velocity will of course decrease throughout the experiment, but to match the initial substrate concentrations, we want to find the initial velocity, i.e. the initial slope of the curve of product or substrate versus time.

You can perhaps see how determining the instantaneous initial velocity of this curve can be difficult and error prone. We need very rapid and accurate measurements of product or substrate in order to determine this initial slope. Due to the properties of the reciprocal, the error at low substrate concentrations is amplified and has a strong influence on the slope and intercept of the graph. For these reasons the Lineweaver-Burke plot provides better estimates of V_m than K_m.

5.1.2 Eadie-Hofstee Plot

To achieve better estimates of K_m we can use a different linearization and plot of the Michaelis-Menten equation, called the Eadie-Hofstee plot.

$$\nu = V_m - K_m\frac{\nu}{S}$$

In Eadie-Hofstee plot, ν is plotted against ν/S, and $-K_m$ is the slope and V_m is the y-intercept. While there is still potential for errors in the calculation of ν the issues with low substrate concentrations are lessened. Because the error in ν is in x and y coordinates however, this plot does not yield the most accurate measurements of V_m.

5.1.2.1 ???Questions??? Why does this error not affect the slope?

Error still has the potential to affect the slope, but systematic error will affect both the x and y coordinates equally therefore cancelling the effect of this error on the slope.

5.1.3 Hanes-Woolf plot

There is one more linearization that is commonly used for parameter estimation called the Hanes-Woolf plot. The Hanes-Woolf plot has the form

$$\frac{S}{\nu} = \frac{K_m}{V_m} + \frac{1}{V_m}S$$

Therefore S/ν is plotted against S and $1/V_m$ is the slope an K_m/V_m is the y-intercept. The Hanes-Woolf plot accurately estimates V_m, the slope, but again the y-intercept can be innacurate as in this case we cannot measure the reaction velocity at $S = 0$.

Let's look at an some example data and calculate V_m and K_m in several ways. Here we have performed experiments with two different concentrations of the same purified enzyme across a range of substrate concentrations, and calculated the initial velocity of each reaction.

```r
S <- c(20, 10, 6.7, 5, 4, 3.3, 2.9, 2.5)  # substrate concentration (g/L)
v_0.015 <- c(1.14, 0.87, 0.7, 0.59, 0.5, 0.44, 0.39,
    0.35)  # reaction velocity (g/L-min) at E_0 = 0.015 g/L
v_0.00875 <- c(0.67, 0.51, 0.41, 0.34, 0.29, NA, NA,
    NA)  # reaction velocity (g/L-min) at E_0 = 0.00875 g/L

data <- data.frame(S, v_0.015, v_0.00875)
require(knitr)
knitr::kable(data)
```

S	v_0.015	v_0.00875
20.0	1.14	0.67
10.0	0.87	0.51
6.7	0.70	0.41
5.0	0.59	0.34
4.0	0.50	0.29
3.3	0.44	NA
2.9	0.39	NA
2.5	0.35	NA

Sometimes we won't know the concentration of our enzyme, if say we are working with just a cell lysate or if we are unable to completly purify our enzyme. In this case we will just use a *specific activity* of the mixture. Specific activity is just the activity of the enzyme per mass of total protein in the complex mixture. The enzyme activity is measure in units of product formed per time at particular conditions (e.g. temperature, pH, substrate concentration, etc.).

Let's initially look at this data by plotting ν vs S as dictated by the Michaelis-Menten equation.

```r
require(tidyr)
data <- tidyr::gather(data = data, key = E_0, value = v,
    v_0.00875, v_0.015)
require(stringr)
data$E_0 <- as.numeric(stringr::str_remove(string = data$E_0,
    pattern = "v_"))
require(ggplot2)
ggplot2::ggplot(data = data, mapping = aes(x = S, y = v,
    color = as.factor(E_0))) + geom_point() + labs(y = "v (g/L-min)",
    x = "S (g/L)") + expand_limits(x = 0, y = 0)
```

```
## Warning: Removed 3 rows containing missing values or
## values outside the scale range (`geom_point()`).
```
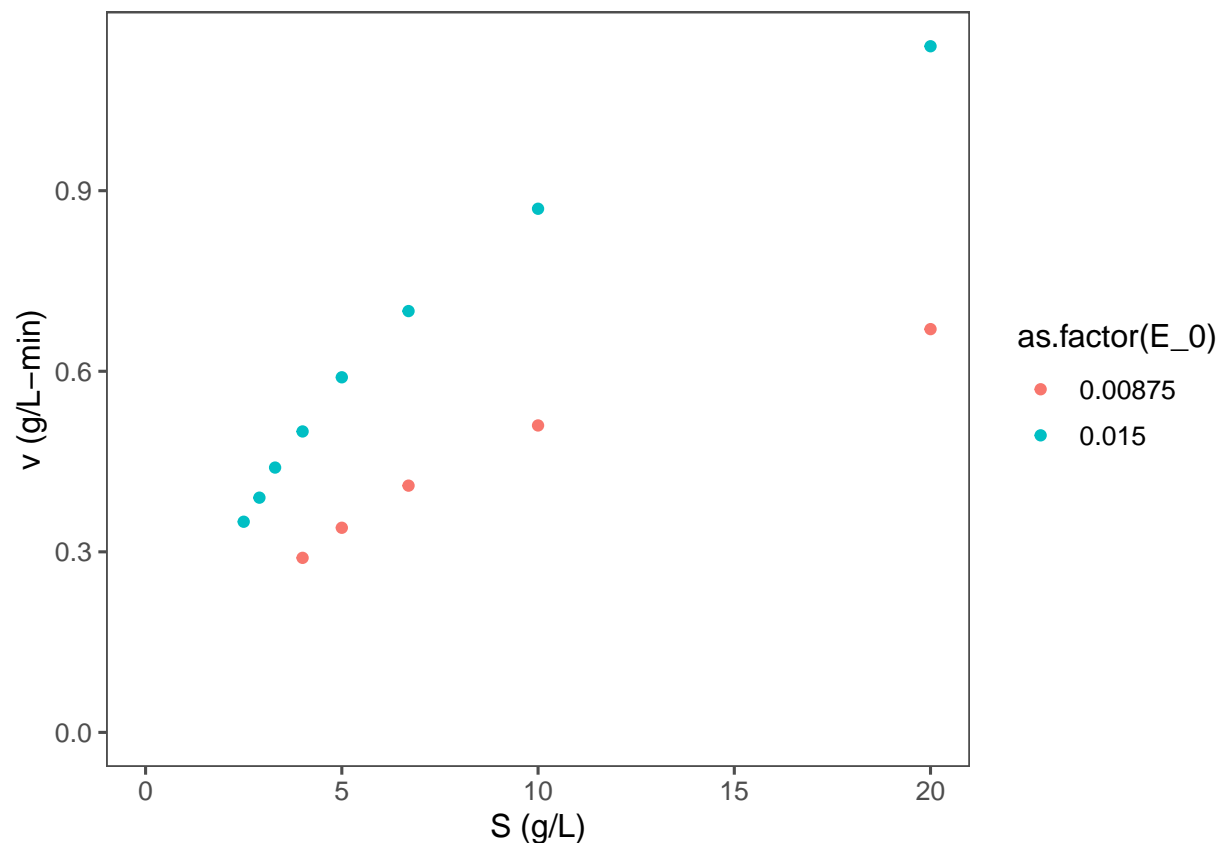
5.1.3.1 ???Questions??? What can you say about this graph?

The higher concentration of enzyme has a higher velocity across all substrate concentrations. The shape of the two curves is similar. As substrate concentration increases the increase in reaction velocity decreases.

Now let's make a Lineweaver-Burk plot.

$$\frac{1}{\nu} = \frac{1}{V_m} + \frac{K_m}{V_m}\frac{1}{S}$$

```
data$LB.y <- 1/data$v
data$LB.x <- 1/data$S
ggplot2::ggplot(data = data, mapping = aes(x = LB.x,
    y = LB.y, color = as.factor(E_0))) + geom_point() +
    labs(y = "1/v (L-min/g)", x = "1/S (L/g)") + expand_limits(x = 0,
    y = 0)
```

```
## Warning: Removed 3 rows containing missing values or
## values outside the scale range (`geom_point()`).
```
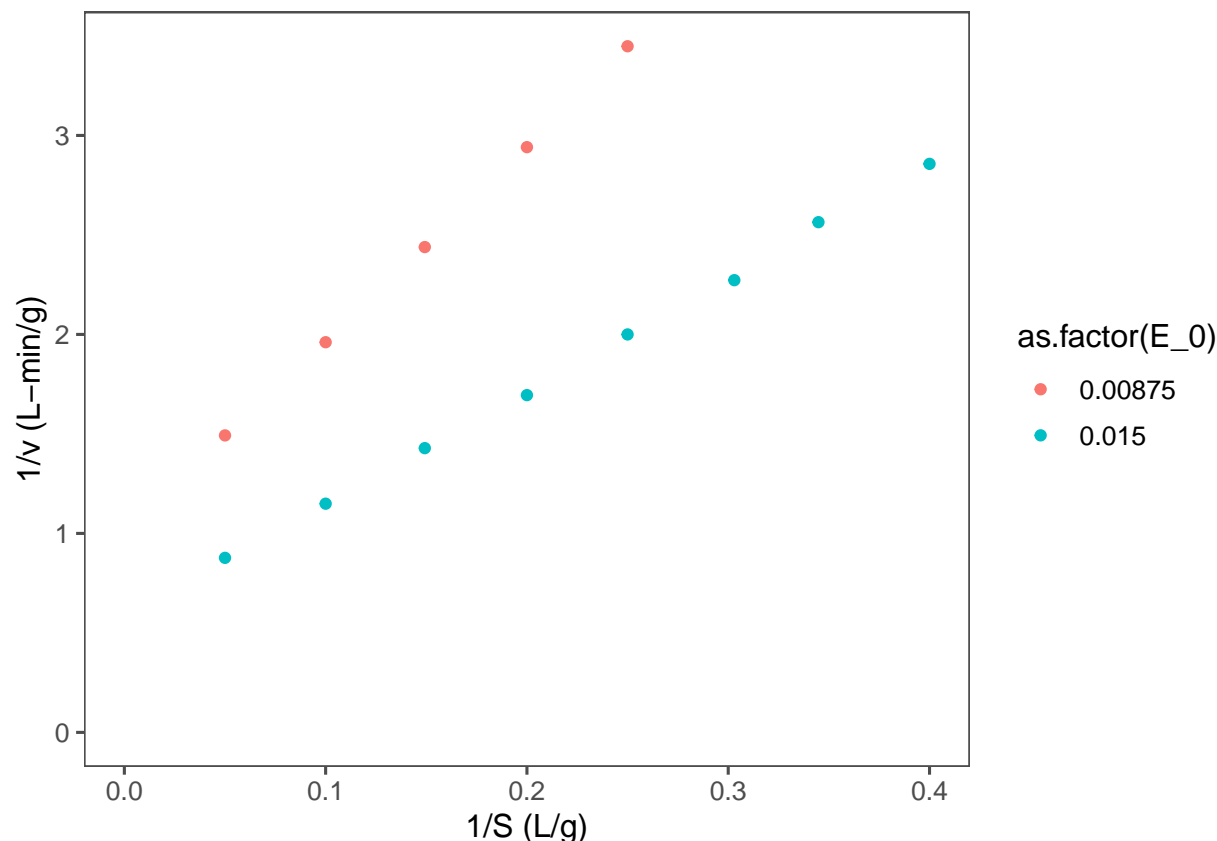
5.1.3.2 ???Questions??? Why are there two lines here?

Do not forget that $V_m = k_{cat}E_0$. So the slope of these lines is really $\frac{K_m}{k_{cat}E_0}$.

There are two ways that we could go about calculating K_m and and V_m or k_{cat} from this data, either find the slope and intercept of both of the above lines and calcuate 2 parameters and then average them, or we could include the enzyme concentration in our linearization. It is best to do the later as it will allow us to more accurately know the uncertainty in our parameter estimates. To do this we will divide the velocity by E_0 to allow us to linearize the right hand side of the Michaelis-Menten equation in terms of k_{cat}.

$$\nu = \frac{k_{cat}E_0 S}{K_m + S}$$

$$\frac{\nu}{E_0} = \frac{k_{cat}S}{K_m + S}$$
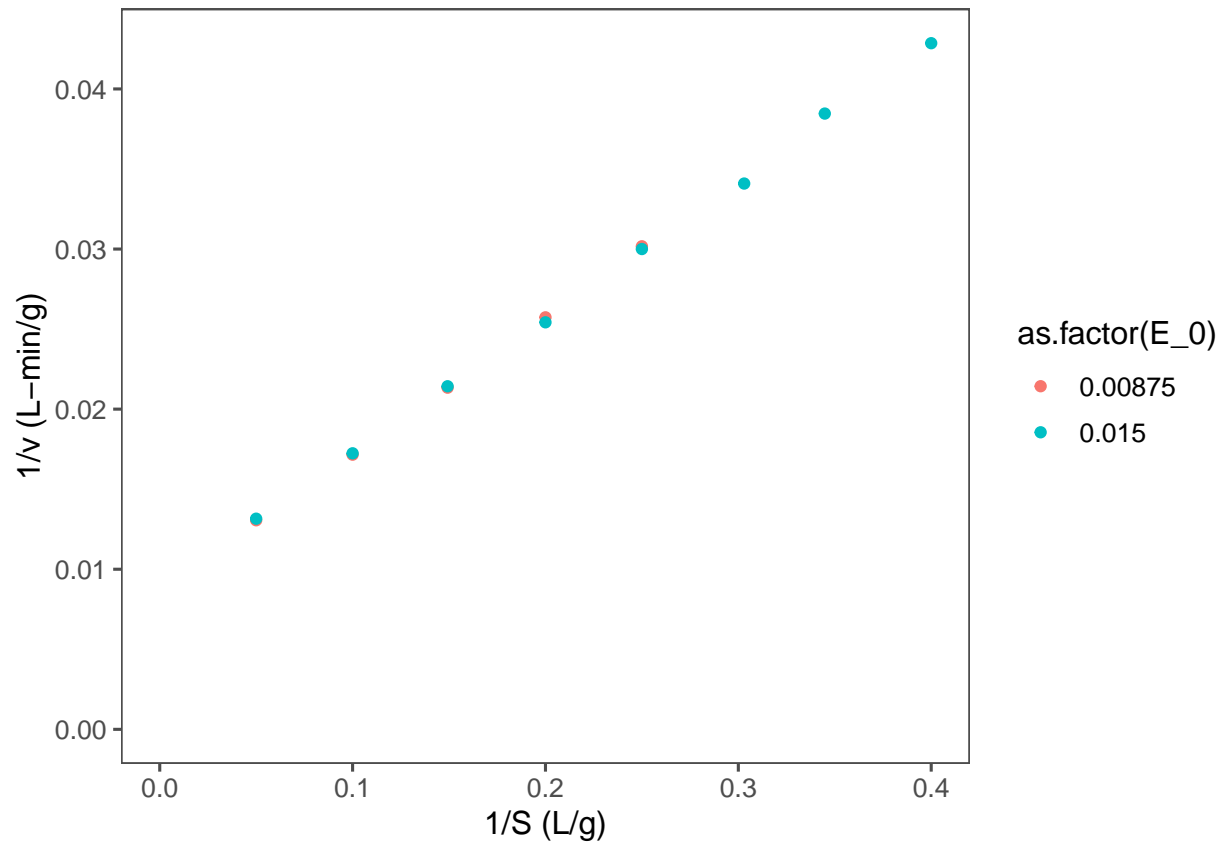
Now invert to linearize

$$\frac{E_0}{\nu} = \frac{K_m}{k_{cat}S} + \frac{1}{k_{cat}}$$

So let's now plot E_0/ν vs $1/(S)$.

```
data$LB.y <- data$E_0/data$v
data$LB.x <- 1/data$S
ggplot2::ggplot(data = data, mapping = aes(x = LB.x,
    y = LB.y, color = as.factor(E_0))) + geom_point() +
    labs(y = "1/v (L-min/g)", x = "1/S (L/g)") + expand_limits(x = 0,
    y = 0)
```

```
## Warning: Removed 3 rows containing missing values or
## values outside the scale range (`geom_point()`).
```


Notice that our two enzyme concentrations now are plotted nearly perfectly overtop of each other. Now we can use linear regression to find the slope and intercept of the line of best fit.

```
LB <- lm(formula = LB.y ~ LB.x, data = data)
summary(LB)
```

```
##
## Call:
## lm(formula = LB.y ~ LB.x, data = data)
##
## Residuals:
##        Min         1Q     Median         3Q
## -4.320e-04 -7.006e-05  5.760e-06  9.224e-05
##        Max
##  3.779e-04
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0087080  0.0001269    68.6 7.84e-16
## LB.x        0.0851892  0.0005698   149.5  < 2e-16
##
## (Intercept) ***
## LB.x        ***
## ---
```

```
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0002181 on 11 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.9995, Adjusted R-squared:  0.9995
## F-statistic: 2.235e+04 on 1 and 11 DF,  p-value: < 2.2e-16
```

Looking back at our linearized model,

$$\frac{E_0}{\nu} = \frac{K_m}{k_{cat}S} + \frac{1}{k_{cat}}$$

we can calculate our estimate of $k_{cat} = 114.8369633 \pm$, and our estimate of $K_m = 9.7828737$. To calculate the uncertainty in these estimates we must propagate the standard error in our slope and intercept, according to the division formula

$$Q = \frac{a}{x}, \delta Q = Q\sqrt{(\frac{\delta a}{a})^2 + (\frac{\delta x}{x})^2}$$

Let's make a table to compare these estimates with other linearizations as well as nonlinear regression methods.

```
require(tibble)
MM_params <- tibble::tibble(method = "LB", k_cat = 1/summary(LB)$coefficients[1,
    1], ek_cat = summary(LB)$coefficients[1, 2], K_m = summary(LB)$coefficients[2,
    1]/summary(LB)$coefficients[1, 1], eK_m = summary(LB)$coefficients[2,
    1]/summary(LB)$coefficients[1, 1] * sqrt((summary(LB)$coefficients[1,
    2]/summary(LB)$coefficients[1, 1])^2 + (summary(LB)$coefficients[2,
    2]/summary(LB)$coefficients[2, 1])^2))
knitr::kable(MM_params)
```

method	k_cat	ek_cat	K_m	eK_m
LB	114.837	0.0001269	9.782874	0.1569057

Looking at the linearization for the Eadie-Hofstee plot

$$\nu = V_m - K_m \frac{\nu}{S}$$

We can simplify V_m to $k_{cat}E_0$ and then do linear regression as above.

```
data$EH.y <- data$v
data$EH.x <- data$v/data$S
EH <- lm(formula = EH.y ~ E_0 + EH.x - 1, data = data)
summary(EH)
```

```
##
## Call:
## lm(formula = EH.y ~ E_0 + EH.x - 1, data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.022251 -0.007251  0.000530  0.004844  0.020436
##
## Coefficients:
##       Estimate Std. Error t value Pr(>|t|)
## E_0   113.0511     1.0901  103.71  < 2e-16 ***
## EH.x   -9.5441     0.1446  -66.02 1.19e-15 ***
## ---
## Signif. codes:
```

```
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01202 on 11 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.9997, Adjusted R-squared:  0.9996
## F-statistic: 1.623e+04 on 2 and 11 DF,  p-value: < 2.2e-16
```

```
MM_params <- rbind(MM_params, c("EH", summary(EH)$coefficients[[1,
    1]], summary(EH)$coefficients[[1, 2]], -summary(EH)$coefficients[[2,
    1]], summary(EH)$coefficients[[2, 2]]))
knitr::kable(MM_params)
```

method	k_cat	ek_cat	K_m	eK_m
LB	114.836963256884	0.000126942716628018	9.78287373281609	0.156905735321379
EH	113.051061858663	1.09005061012357	9.54408258808163	0.144570335085599

Now for the Hanes-Woolf plot

$$\frac{S}{\nu} = \frac{K_m}{V_m} + \frac{1}{V_m} S$$

```
data$HW.y <- data$E_0 * data$S/data$v
HW <- lm(formula = HW.y ~ S, data = data)
summary(HW)
```

```
##
## Call:
## lm(formula = HW.y ~ S, data = data)
##
## Residuals:
##        Min         1Q     Median         3Q
## -0.0014254 -0.0006425 -0.0000589  0.0007844
##        Max
##  0.0016303
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.417e-02  4.921e-04   171.1   <2e-16
## S           8.874e-03  5.127e-05   173.1   <2e-16
##
## (Intercept) ***
## S           ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.001059 on 11 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.9996, Adjusted R-squared:  0.9996
## F-statistic: 2.995e+04 on 1 and 11 DF,  p-value: < 2.2e-16
```

```
MM_params <- rbind(MM_params, c("HW", 1/summary(HW)$coefficients[[2,
    1]], summary(HW)$coefficients[[2, 2]], summary(HW)$coefficients[[1,
    1]]/summary(HW)$coefficients[[2, 1]], summary(HW)$coefficients[2,
    1]/summary(HW)$coefficients[1, 1] * sqrt((summary(HW)$coefficients[1,
    2]/summary(HW)$coefficients[1, 1])^2 + (summary(HW)$coefficients[2,
    2]/summary(HW)$coefficients[2, 1])^2)))
```

```
knitr::kable(MM_params)
```

method	k_cat	ek_cat	K_m	eK_m
LB	114.836963256884	0.000126942716628018	9.78287373281609	0.156905735321379
EH	113.051061858663	1.09005061012357	9.54408258808163	0.144570335085599
HW	112.685492138659	5.12741049599667e-05	9.48505475242782	0.000866570502091196

5.2 Nonlinear least squares regression

With the computational power available today these linearizations are really historical relicts, as we can fit these parameters to our nonlinear model. Nonlinear regression methods essentially sweep through the parameter space, by iteratively changing each parameter and calculating the distance between the model with the current parameter set and the data points. The sum of the squared distance between our model and data is called the residual. The parameter values that minimize the residual are chosen our best fit.

```
# Perform nonlinear least squares regression
nls1 <- nls(v ~ S/(S + Km) * k_cat * E_0, data, list(Km = 1,
    k_cat = 1))
summary(nls1)
```

```
##
## Formula: v ~ S/(S + Km) * k_cat * E_0
##
## Parameters:
##        Estimate Std. Error t value Pr(>|t|)
## Km       9.4042     0.1059    88.8   <2e-16 ***
## k_cat  112.1862     0.6591   170.2   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.003881 on 11 degrees of freedom
##
## Number of iterations to convergence: 11
## Achieved convergence tolerance: 9.964e-06
##   (3 observations deleted due to missingness)
```

```
MM_params <- rbind(MM_params, c("NLS", summary(nls1)$coefficients[[2,
    1]], summary(nls1)$coefficients[[2, 2]], summary(nls1)$coefficients[[1,
    1]], summary(nls1)$coefficients[[1, 2]]))
knitr::kable(MM_params)
```

method	k_cat	ek_cat	K_m	eK_m
LB	114.836963256884	0.000126942716628018	9.78287373281609	0.156905735321379
EH	113.051061858663	1.09005061012357	9.54408258808163	0.144570335085599
HW	112.685492138659	5.12741049599667e-05	9.48505475242782	0.000866570502091196
NLS	112.18622883779	0.659059979310289	9.40416797330833	0.105901382959763

5.3 Comparing linearization vs nonlinear methods

The error in each of these parameters is as we expected, based on our knowledge of the linearizations, but which one best fits our actual data? To assess this we can calculate the residuals for each model and compare the sum of the square of these residuals. Let's do this for only the $E_0 = 0.015$ dataset to keep it simple.

```
# Add the correct E_0 parameter
MM_params$E_0 <- 0.015
```

```r
require(tidyverse)
# make sure that the parameters are numerics
MM_params[, 2:5] <- MM_params[, 2:5] %>%
    map(as.numeric)

# Write a function to calculate velocity based on
# the Michaelis-Menten equation
MM <- function(k_cat, E_0, S, K_m) {
    k_cat * E_0 * S/(K_m + S)
}

# Write a for loop to calculate velocity for each
# S value with each parameter set
resid <- tibble(S, v_0.015)
for (i in 1:nrow(MM_params)) {
    # calculate predicted v values
    v <- with(MM_params[i, ], {
        MM(k_cat, E_0, S, K_m)
    })
    # name the predicted column with the method
    # and add it to resid
    resid <- resid %>%
        add_column(!!(MM_params[[i, "method"]]) :=
            v)
}
resid
```

```
## # A tibble: 8 x 6
##        S v_0.015    LB    EH    HW   NLS
##    <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  20       1.14 1.16  1.15  1.15  1.14
## 2  10       0.87 0.871 0.868 0.867 0.867
## 3   6.7     0.7  0.700 0.699 0.700 0.700
## 4   5       0.59 0.583 0.583 0.583 0.584
## 5   4       0.5  0.500 0.501 0.501 0.502
## 6   3.3     0.44 0.434 0.436 0.436 0.437
## 7   2.9     0.39 0.394 0.395 0.396 0.397
## 8   2.5     0.35 0.351 0.352 0.353 0.353
```

```r
# Calculate the residuals for each method
resid <- resid %>%
    mutate_at(c("LB", "EH", "HW", "NLS"), .funs = funs(. -
        v_0.015))  # . = predicted
```

```
## Warning: `funs()` was deprecated in dplyr 0.8.0.
## i Please use a list of either functions or
##   lambdas:
##
## # Simple named list: list(mean = mean, median =
##   median)
##
## # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
## # Using lambdas list(~ mean(., trim = .2), ~
```

```
##   median(., na.rm = TRUE))
## Call `lifecycle::last_lifecycle_warnings()` to
## see where this warning was generated.
# Calculate the sum of the square of residuals
SSR <- resid %>%
    summarise_at(c("LB", "EH", "HW", "NLS"), .funs = funs(sum(.^2)))

## Warning: `funs()` was deprecated in dplyr 0.8.0.
## i Please use a list of either functions or
##   lambdas:
##
## # Simple named list: list(mean = mean, median =
##   median)
##
## # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
## # Using lambdas list(~ mean(., trim = .2), ~
##   median(., na.rm = TRUE))
## Call `lifecycle::last_lifecycle_warnings()` to
## see where this warning was generated.
# Plot the SSR values
SSR <- gather(SSR, key = "method", value = "SSR")
ggplot(data = SSR, mapping = aes(x = forcats::fct_inorder(method),
    y = SSR)) + geom_bar(stat = "identity") + labs(x = "method",
    y = "Sum of square residuals")
```
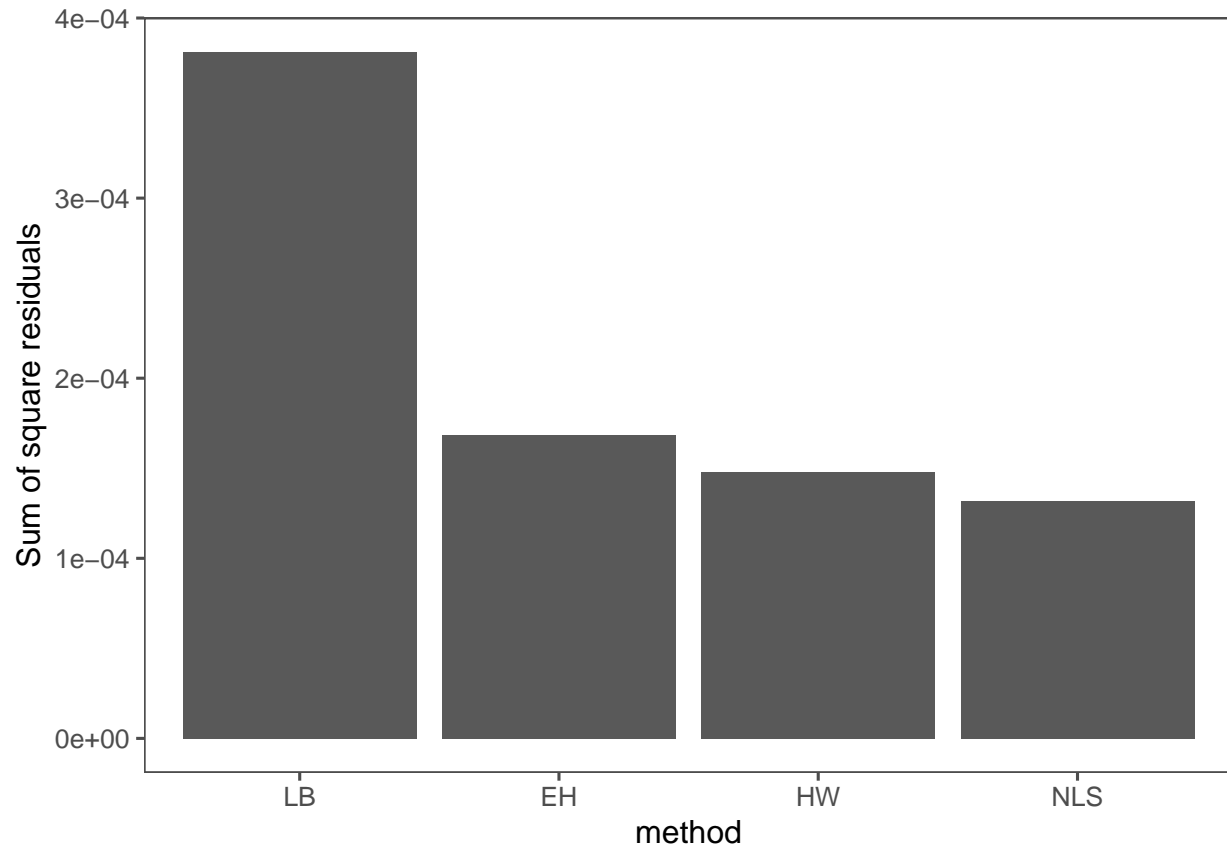
So you can see in this case that the nonlinear least squares regression results in the best fit. In general this will be the case. We could also combine the K_m from Eadie-Hofstee and the V_m from Hanes-Woolf, but we still run into the issue of these linear transformations skewing our uncertainty in or parameters. Nonlinear regression methods yield absolute uncertainties in the parameter estimates that can be easily compared and interpreted.

Just for fun let's see what the difference in these models looks like.

```
S_plot <- seq(0, 25, 0.1)
model_plot <- tibble(S_plot)
for (i in 1:nrow(MM_params)) {
    # calculate predicted v values
    v <- with(MM_params[i, ], {
        MM(k_cat, E_0, S_plot, K_m)
    })
    # name the predicted column with the method
    # and add it to resid
    model_plot <- model_plot %>%
        add_column(!!(MM_params[[i, "method"]]) :=
            v)
}

model_plot <- model_plot %>%
    gather(key = "method", value = "v", 2:5)

ggplot(data = model_plot, mapping = aes(x = S_plot,
    y = v, color = method)) + geom_line() + labs(x = "Velocity g/L-min",
```

```
    y = "Substrate g/L")
```
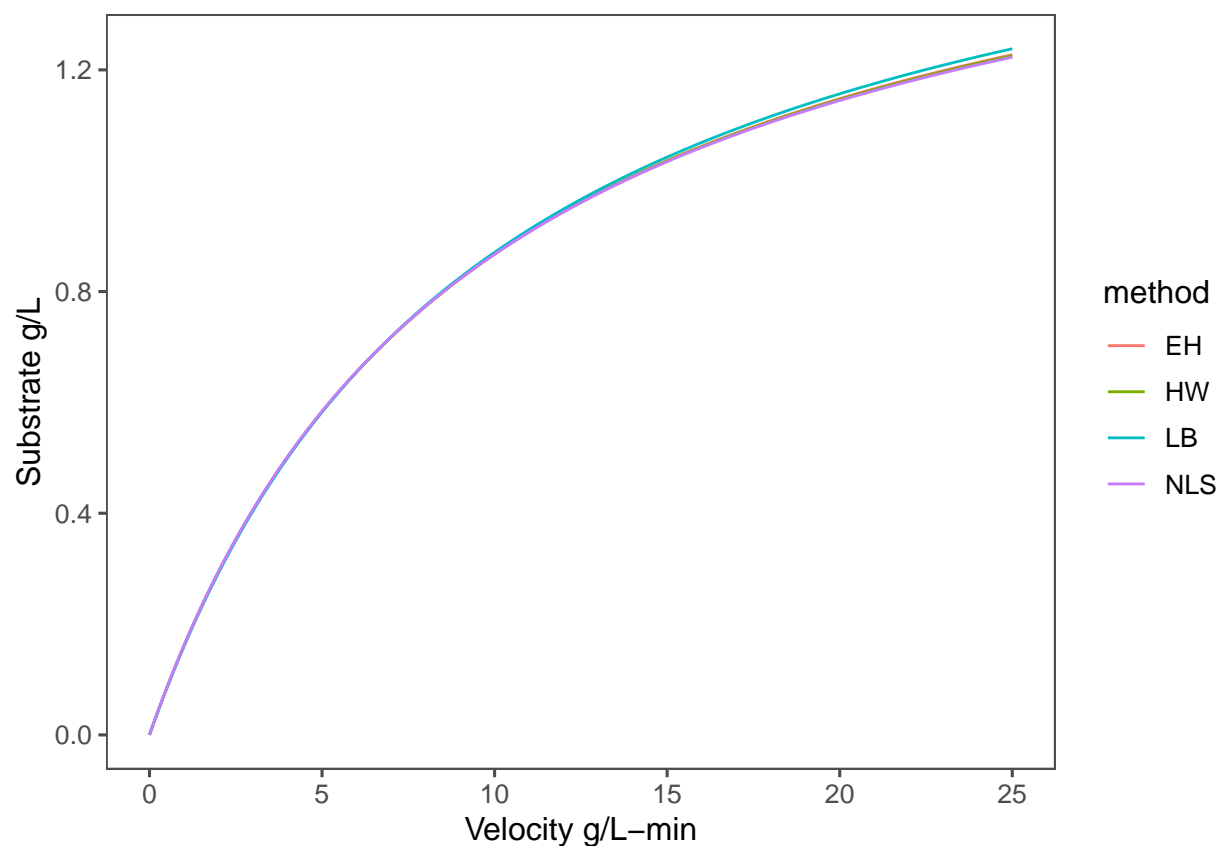

This is a decent shiny app for demonstrating parameter fitting.

Here is a demonstration of linear and nonlinear fitting with MATLAB.

6   Complex Enzyme Kinetics

Many enzymes do not simply convert one substrate to a single product. Enzymes such as amylases which are critical to brewing and distilling convert long chains of sugars (starches) into shorter chains and disaccharides.

6.1   Allosteric Enzymes

Allosteric enzymes have more than one substrate binding site, or an "effector" binding site in addition to a substrate binding site. An effector can be any biomolecule, that when bound to the enzyme, affects the activity of the enzyme.

If the effector or substrate has a positive effect on enzyme activity this is called cooperativity.

6.2   Enzyme Inhibition

6.2.1   Competitive Inhibition

6.2.2   Uncompetitive

6.2.3   Noncompetitive

6.2.4   Substrate Inhibition