Proteins

Overview

Proteins are responsible for most of the chemical and physical "action" in cells:

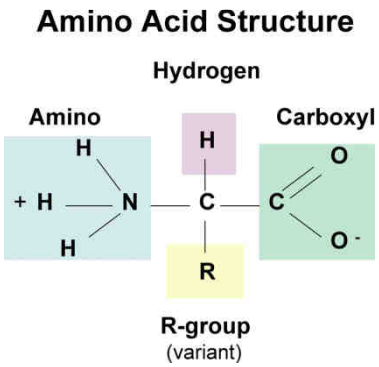
- reactions, enzymes
- structure
- transport
- motoring / mechanical movement
- storage
- signaling
- receptor / sensing
- regulation of protein production

Proteins are polymers of amino acids. They have many properties similar to traditional polymers, however they have two remarkable features that give them a much broader range of functionalities → we will discuss

Protein composition

Polymers of amino acids, made in precise sequences

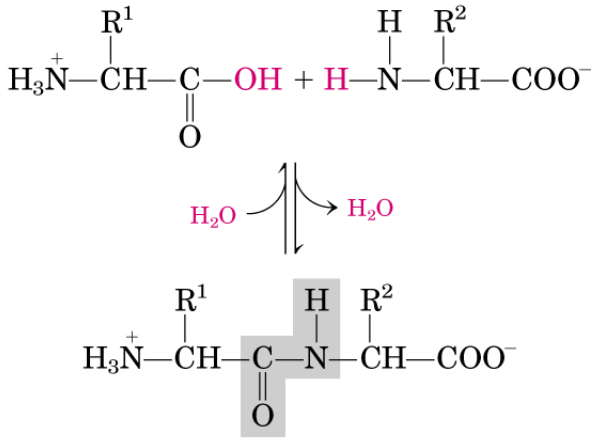
Each amino acid:



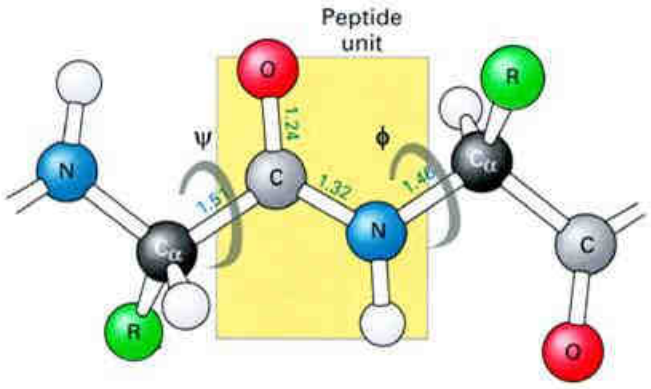
The carbon in the center is the "alpha carbon". It is a chiral center. All amino acids have **L-chirality.**

The R group is called the **side chain.**

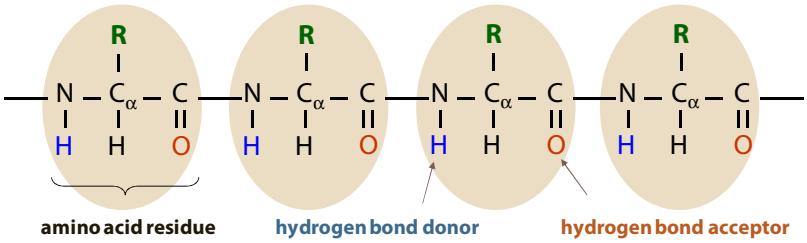
Amino acids join together to form a **peptide bond** by a condensation reaction:



Peptide bonds are between the carbonyl and amino groups of adjacent amino acids and are planar due to partial p character. Rotations can occur along the **phi and psi angles:**

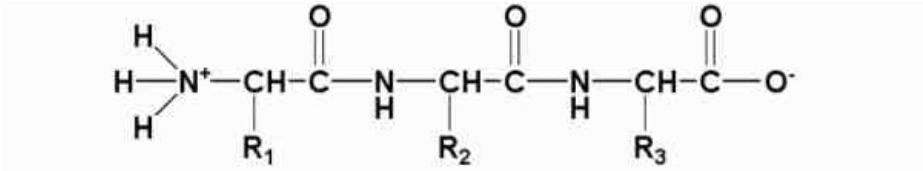


These rotations allow the protein polymer to assume different **conformations**.



The **backbone** of the protein has amide hydrogens and carbonyl oxygens that can interact to form **hydrogen bonds**.

Protein chains have a direction: N and C terminal parts; synthesized starting at the N terminal:



Typically the **termini** of a protein are charged (basic/acidic) in solution.

Amino acids

Nature uses 20 amino acids → more are possible

Often abbreviated with one-letter and three-letter codes

Different chemistries so as to allow different interactions and reactions

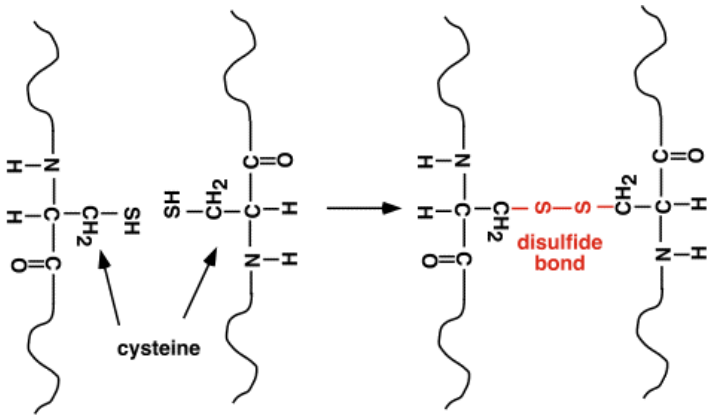
Basic kinds of amino acids

- hydrophobic, of varying sizes
- polar, uncharged
- polar, positively (basic) or negatively (acidic) charged

Special amino acids

- proline – rigid, can't rotate around phi angle, used to make rigid backbone
- glycine – highly flexible, no side chain
- histidine – easily protonated / deprotonated, useful in catalysis of many reactions
- cysteine – can cross link through **disulfide bonds** using sulfur atoms, enables more stable protein structures

Cysteine example:



First amazing property of proteins: are synthesized with *exact* sequences of amino acids, up to long polymers with lengths of ~10000 amino acids. This contrasts synthetic polymers for which it is very hard to make with peaked molecular weight distributions or complex architectures.

Protein sequences ultimately come from DNA sequences → transcription/translation machinery accomplishes this difficult task

Protein folding

Second amazing property of proteins: each protein adopts a unique, three dimensional conformation in aqueous solution

The terminology **folding** is used to describe the process of going from a floppy, unstructured state to the unique, three dimensional **native** state of a protein

The same protein sequence always folds to the same structure → the sequence of a protein **encodes** its structure

→ this implies that protein folding is a **thermodynamic** transition to a lower free energy state, the native state

What interactions drive protein folding?

- hydrophobic interactions (dominant force) – protein adopts a conformation in which hydrophobic amino acids tend to be buried in the interior of the protein
- hydrogen bonding (also important) – atoms along the backbone and sometimes also in side chains try to form hydrogen bonds with each other
- van der Waals interactions and dense packing (also important) – drive to a compact structure from hydrophobic interactions promotes favorable, close-packing between protein atoms

- electrostatic interactions (somewhat important) – basic and acidic residues paired

Evolutionary considerations:

- Will random sequences of amino acids fold? → NO!
- Nature selects for sequence that are able to fold, among other things

Thermodynamics of protein folding

Protein folding can be considered a thermodynamic transition in which the system changes between an unstructured, unfolded state to a folded state.

For a single molecule, we might write

$$U \leftrightarrow F$$

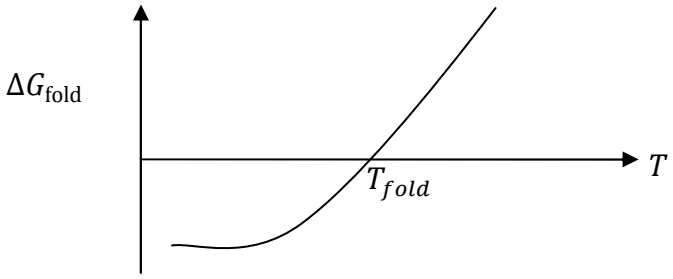
$$\Delta G^0 = G_F^0 - G_U^0$$

$$\Delta G^0 < 0 \quad \text{for spontaneous folding}$$

Let's break down the folding free energy term by term:

$$\Delta G^0 = \Delta H^0 - T\Delta S^0$$

A plot of the folding free energy vs temperature typically looks like:



Notice that we define T_{fold} or T_f:

- below this temperature, the change in free energy upon folding is negative → the protein folds spontaneously
- above this temperature, the change in free energy upon folding is positive → the protein unfolds spontaneously

Therefore, proteins can be **denatured or unfolded** by heating. Example: cooking, eggs

Proteins can also be denatured (unfolded) by:

- chemicals / denaturants. Ex: alcohols, urea, guanidine HCL [question: mechanism?]
- extreme pH
- very high salt concentrations
- reducing agents

By changing the conditions to favor unfolding, one is in effect modifying the free energies so that $\Delta G^0 > 0$, or $G_U^0 < G_F^0$

Unfolded proteins can be **refolded** by returning to normal aqueous conditions

In a solution of a protein, we have to include concentration effects:

$$G_U = G_U^\circ + RT \ln[U]$$

$$G_F = G_F^\circ + RT \ln[F]$$

The total change in free energy is

$$\begin{aligned} \Delta G &= G_F - G_U \\ &= \Delta G^0 + RT \ln \frac{[F]}{[U]} \end{aligned}$$

Define

$$K = \frac{[F]}{[U]} \quad \text{at equilibrium}$$

Then

$$K = e^{-\frac{\Delta G^0}{RT}}$$

Kinetics of protein folding

Proteins "search" for their native structure by "trying out" different conformations. The one with the lowest free energy wins.

Individual proteins fold with rates on the order of microseconds (ultrafast) and milliseconds (typical) to seconds and even longer (big proteins)

Is this fast?

- Consider how many conformations a protein can have. Assume each amino acid residue along the chain can have three conformational states in its backbone torsional angle (phi, psi). This is a gross simplification.

- There would be ~$3^{N_{res}}$ numbers of conformations the protein could adopt for N_{res} amino acid residues.

- Let's say the protein "tried out" a new conformation every 1 ps = 10^{-12} s. How long would it take a protein to sort through all conformations for a 100 residue protein?

$$t = 3^{100} * (10^{-12} s)$$
$$= 10^{35} s$$
$$= 10^{18} \text{ years} \quad (\text{vs } 10^{10} \text{ years for the age of the universe})$$

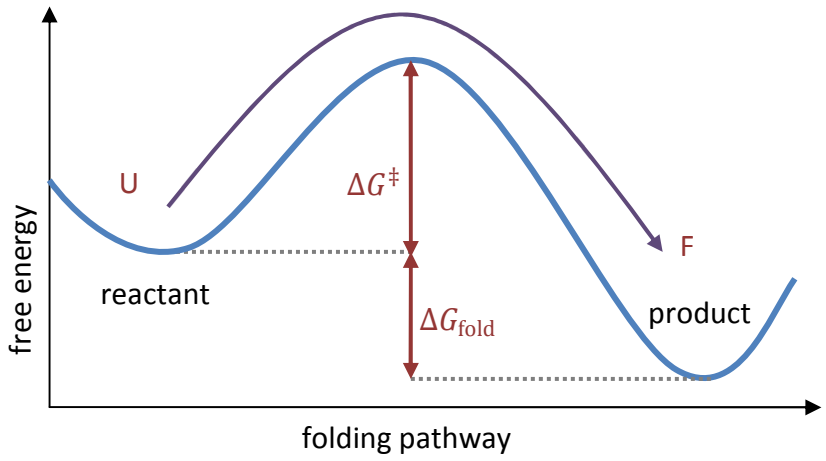
Thus we might think that it would take proteins longer than the age of the universe to fold! This is known as **Levinthal's paradox.**

In reality, proteins do not randomly search through conformations. They perform a guided search. The interactions in a protein guide it along a **folding pathway** that is directed and that results in an efficient search.

Evolutionary considerations:

- Not all amino acid sequences fold fast. Biology has **selected** protein sequences that have efficient folding behavior.

One can model the folding kinetics of many proteins along their folding pathway using a transition state theory approach:



For a solution of many molecules of the same protein, we can write a reaction kind of picture:

$$U \leftrightarrow F$$

$$\frac{d[U]}{dt} = -k_f[U] + k_u[F]$$

$$\frac{d[F]}{dt} = k_f[U] - k_u[F]$$

The rate constant is related to the free energy at the transition state:

$$k_f = k_0 e^{-\frac{\Delta G^{\ddagger}}{RT}}$$

Note that, at equilibrium, we have

$$\frac{d[F]}{dt} = \frac{d[U]}{dt} = 0$$

so that ultimately (time to infinity)

$$\frac{k_f}{k_u} = \frac{[F]}{[U]} = K$$

Substituting free energies in for k_u, k_f, K shows that

$$k_u = k_0 e^{\frac{\Delta G_{fold}}{RT} - \frac{\Delta G^{\ddagger}}{RT}}$$

This picture applies to so-called "**two-state**" folders, that proceed directly from an unfolded to a folded state.

We want to solve this set of coupled first order, differential equations. There are two equations, two unknowns, and both first order, so we need two initial conditions. We will consider the case in which

$$[U]_{t=0} = [U]_0$$

$$[F]_{t=0} = 0$$

Your experience in your math class shows that the one way to solve these equations is using an eigenvector approach:

$$\frac{d}{dt}\begin{bmatrix} U \\ F \end{bmatrix} = \begin{bmatrix} U \\ F \end{bmatrix}^T \begin{bmatrix} -k_f & k_f \\ k_u & k_u \end{bmatrix}$$

Or,

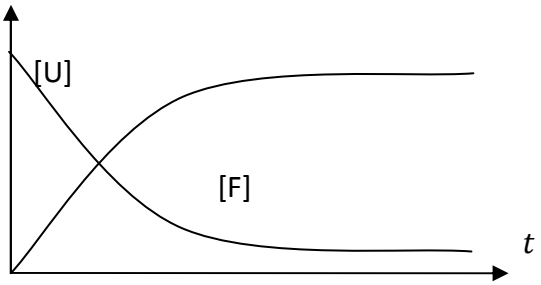
$$\frac{d}{dt}\mathbf{X} = \mathbf{X}^T\mathbf{A}$$

The final solutions are

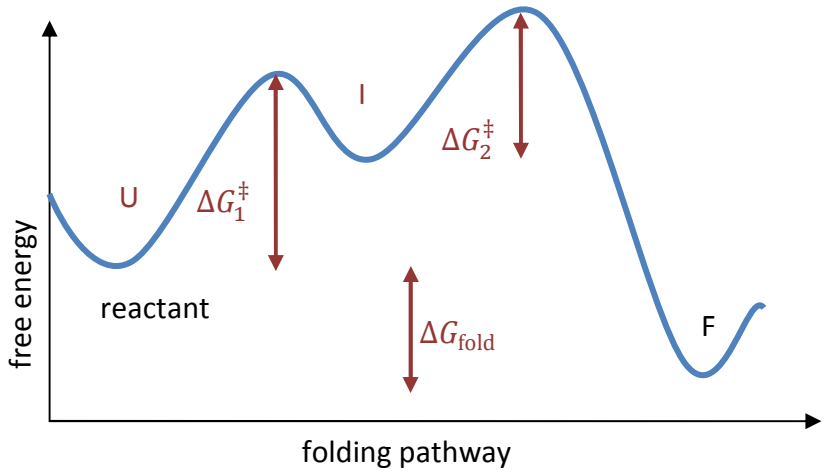
$$[U] = \frac{e^{-(k_f+k_u)t}k_f + k_u}{k_f + k_u}[U]_0$$

$$[F] = [U]_0 - [U] = \frac{k_f\left[1 - e^{-(k_f+k_u)t}\right]}{k_f + k_u}[U]_0$$

The solution looks like:



On the other hand, larger, more complicated proteins can undergo "three-state" folding that contains a stable, detectable intermediate:



For three state folders we have:

$$U \leftrightarrow I \leftrightarrow F$$

$$\frac{d[U]}{dt} = -k_{f1}[U] + k_{u1}[I]$$

$$\frac{d[I]}{dt} = k_{f1}[U] - k_{u1}[I] + k_{u2}[F] - k_{f2}[I]$$

$$\frac{d[F]}{dt} = k_{f2}[I] - k_{u2}[F]$$

The solution to these equations can be represented as two exponential decay processes. The solution is a bit more complicated, so we won't review it further here.

Refolding, misfolding & aggregation

Proteins can often misfold, or fold into incorrect states. This can be the case when there is a long-lived intermediate state.

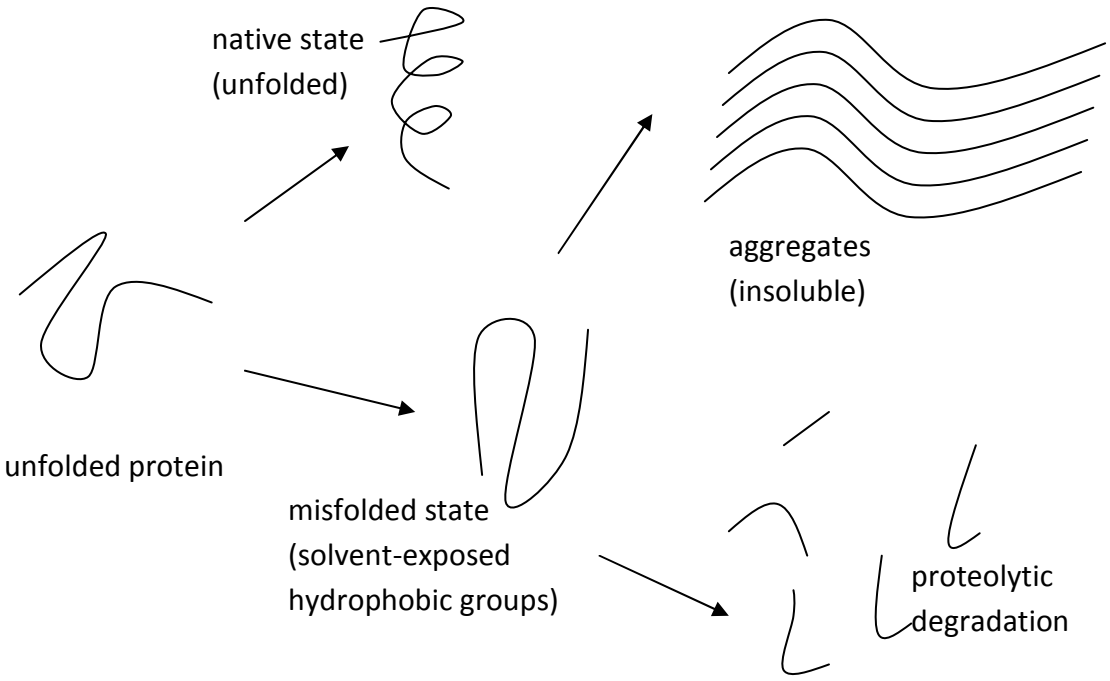
The misfolded conformation is **metastable**, meaning that, with an infinite amount of time, the protein will eventually transform to the correct native structure

If the misfolded state persists for long enough time, there can be problems.

For one, the protein will not function. A significant fraction of proteins in the cell do not fold correctly in a time period required by the cell – these are detected and degraded. There is some energy wastefulness in the production of proteins, but it is needed in order to achieve rapidly functional molecules.

In addition, proteins can aggregate with each other into large, multi-protein structures that are no longer soluble in solution and precipitate out.

Imagine that another way a protein can shield it's hydrophobic groups is to cooperate with many other proteins to hide them from water.



For a solution of proteins, the aggregate phase may be more stable than the phase of solubilized, functional protein molecules in their native state.

Aggregation is a source of disease: Alzheimer's, Prion disease, diabetes

Cells have a number of ways of avoiding aggregation

- Protein sequences are strongly selected against in evolution – fast folding, low misfolding and aggregation propensity
- Rapid degradation of misfolded species: 1/3 of misfolded protein is degraded by the proteasome
- **Chaperones** – large, container-like protein assemblies that sequester a protein and facilitate its rapid folding to the native structure, without allowing interactions with neighbors

Protein structure

Folded proteins can have many levels of structure. We term these primary, secondary, tertiary, and quartenary.

Primary structure simply means the one dimensional sequence of amino acids. Not really a structure, but the simplest level of describing the fold.

Secondary structure refers to the existence of regular structural motifs in the fold. There are three common motifs:

- alpha helices
- beta sheets
- loops and turns

Alpha helices and beta sheets emerge due to hydrogen bonding between different parts of the backbone.

Hydrophobic groups drive the protein to a collapsed, folded state, but proteins do not want to sacrifice hydrogen bonds since these have large energetic penalties [Question: to what are the hydrogen bonding groups in the protein bonded in the unfolded state?]

Alpha helices

- every ith residue hydrogen bonds with the i+4th one
- right-handed almost always

Beta sheets

- hydrogen bonding between two roughly linear strands
- large sheets can be made of many strands
- adjacent strands can be **parallel** or **antiparallel**

Loops and turns

- usually on the exterior of the protein
- anchored by the other secondary structure elements
- frequently the location of catalytic or binding or other activity in the protein

Tertiary structure describes how helices, sheets, and other parts of the structure come together in an overall three-dimensional fold

Quartenary structure describes the association of multiple protein **chains** together into larger, multifunctional units – biology uses protein "toolkits" and assembles different combinations of proteins together to form larger machines; these assemblies are thermodynamically favorable, driven by hydrophobic and recognition interactions

Protein structure determination is a huge enterprise. A protein's structure provides clues as to

- what it does and how it works – function
- what kinds of things it might bind
- how drugs might be designed to agonize (activate) or antagonize (inactivate) it by binding
- what might go wrong in mutants

Protein structures are solved by experimentally intensive methods

- Xray diffraction – requires crystallization of the protein, which can be very challenging
- NMR spectroscopy – relies upon proton coupling in solution

Over 50,000 protein structures have been solved, but more continue to be solved at an increasing pace. Structures are deposited at the **Protein Databank**, www.pdb.org.

Protein structure prediction is also a very active area of computational / theoretical interest. Two kinds of methods:

- physics-based – try to model the true interatomic interactions and folding pathways → limited to the very smallest of proteins
- informatics based – similarities of sequences to proteins with known structures are used to provide "templates" for the structure → very accurate for small proteins, increasingly accurate for medium to modest large proteins

Evolutionary considerations

- Fortunately, proteins tend to be closely related by evolution. That is, we see both similar structures and sequences in proteins because many have evolved from common ancestors over time

Sequence identity: percent of sequence that is exactly the same

Sequence similarity: percent of the sequence that is the same in terms of physiochemical properties (i.e., a hydrophobic amino acid mutated to another hydrophobic one counts as the same)

Example:

KLA**P**TG**I**PP**F**

RLA**G**TG**L**PP**E**

Above shows 60% identity + 20% similarity

It has been found that sequences with > 20% identity have roughly the same folded structure!

[PDB and protein structure tutorial in pymol]

Recognition and ligand binding

By maintaining a specific structure, proteins are able to accomplish a very important task: **binding**

Proteins can bind

- other proteins – **assembly, recognition**
- DNA – **regulation**
- other molecules, especially small molecules, metals, and ions – **ligands**

Binding interactions are **central** to biological function. These interactions underlie:

- assembly of proteins into large functional units – hair, nails, bone, teeth, collagen, silk
- immune response
- synapse function – thought
- catalysis – enzyme reactions
- transport – solutes, ions, sugars

Binding typically involves a **binding pocket**, a cavity on the protein surface that has some **geometric complementarity** with the binding molecule

The term **affinity** is used to describe how tightly a molecule binds to the protein

Importantly, binding can also have very high **specificity** – the protein binds only one kind of molecule

How are affinity and specificity achieved?

- geometry

- paring of hydrophobic groups that shields them from water
- hydrogen bonding
- electrostatic interactions
- van der Waals interactions

By using a combination of many of these interactions, proteins can design binding pockets that have complementarity to specific molecules

What do binding surfaces look like? [show example]

Difficult to identify binding pockets directly, but TYR, TRP, PHE, HIS overrepresented

Highly specific binding is called **recognition.**

Binding thermodynamics

We can describe binding by the reaction

$$P + L \leftrightarrow C$$

$$\Delta G = \Delta G^0 + RT \ln \frac{[C]}{[P][L]}$$

$$\Delta G = 0 \text{ at equilibrium}$$

Solving,

$$\frac{[C]}{[P][L]} = \exp\left(-\frac{\Delta G^0}{RT}\right)$$
$$= \exp\left(-\frac{\Delta H^0 - T\Delta S^0}{RT}\right)$$

We can define two equilibrium constants

$$K_A \equiv \frac{[C]}{[P][L]} \quad \text{association constant}$$

$$K_D \equiv \frac{[P][L]}{[C]} = \frac{1}{K_A} \quad \text{dissociation constant}$$

Note that

$$K_A = \exp\left(-\frac{\Delta G^0}{RT}\right)$$

$$K_D = \exp\left(+\frac{\Delta G^0}{RT}\right)$$

Is ΔG^0 positive or negative for strong binding?

Typically,

$$-\Delta G^0 \sim 5 - 20 \frac{kcal}{mol}$$

How does a change in ΔG affect the dissociation constant K_D?

$$\Delta G' = \Delta G^0 + 1\ kcal/mol$$

$$\frac{K_A'}{K_A} = \exp\left(-\frac{\Delta G' - \Delta G^0}{RT}\right)$$

$$= \exp\left(-\frac{1\frac{kcal}{mol}}{0.002\frac{kcal}{mol\ K} * 300\ K}\right)$$

$$= 0.2 \text{ (factor of 5!)}$$

Kinetics

$$\frac{d[C]}{dt} = k_{on}[P][L] - k_{off}[C]$$

Terms are for association and disassociation. What are the units of k_{on} and k_{off}?

What if all of our protein and ligand are initially bound and we are in a very dilute system?

$$\frac{d[C]}{dt} \approx -k_{off}[C]$$

$$[C]_{t=0} = C_0$$

We can solve this first order differential equation to get:

$$\frac{d\ln[C]}{dt} = -k_{off}$$

$$[C] = C_0 \exp(-k_{off} t)$$

Exponential decay [draw]

Let's define $t_{1/2}$ as the time that it takes half of the bound population to unbind:

$$\frac{1}{2}C_0 = C_0 \exp\left(-k_{off} t_{1/2}\right)$$

$$t_{1/2} = \frac{-\ln\frac{1}{2}}{k_{off}}$$

If $k_{off} = 0.001\ s^{-1}$ then,

$$t = -\frac{\ln\frac{1}{2}}{0.001\ s^{-1}} = 11\ min$$

Typical values:

$$k_{on} \sim 10^5 M^{-1} s^{-1}$$

$$k_{off} \sim 10^{-6} - 1\ s^{-1}$$

Why is k_{on} more or less constant, but k_{off} varies significantly from P/L to P/L?

Considerations for binding:

K_D	$t_{1/2}$	Examples
mM	ms	non-specific
μM	ms – s	multivalent cell surface, intracellular signaling proteins
nM	min-hr	antibodies
pM	hr-days	growth factors / receptors
fM	weeks-months	hydrolase, RNAse / trypsin inhibitors

Temperature dependence of binding

How does the dissociation constant K_A depend on temperature?

$$\ln K_D = \frac{\Delta G}{RT}$$

Let's compare two temperatures T_1 and T_2:

$$\ln\frac{K_D(T_2)}{K_D(T_1)} = \frac{\Delta G(T_2) - \Delta G(T_1)}{RT}$$
$$= \frac{\Delta H(T_2)}{RT_2} - \frac{\Delta H(T_1)}{RT_1} - \frac{\Delta S(T_2) - \Delta S(T_1)}{R}$$

We can use the constant heat capacity model to compute the changes on the RHS [remember? earlier lectures, HW].

$$\frac{d\Delta H}{dT} = \Delta C_P$$

$$\Delta H(T_2) - \Delta H(T_1) = \int_{T_1}^{T_2} \Delta C_P dT$$
$$= \Delta C_P (T_2 - T_1)$$

Solving for $\Delta H(T_2)$:

$$\Delta H(T_2) = \Delta H(T_1) + \Delta C_P (T_2 - T_1)$$

Now for the entropy change:

$$\frac{d\Delta S}{dT} = \frac{\Delta C_P}{T}$$

$$\Delta S(T_2) - \Delta S(T_1) = \int_{T_1}^{T_2} \frac{\Delta C_P}{T} dT$$
$$= \Delta C_P \ln\left(\frac{T_2}{T_1}\right)$$

Plugging these equations into the expression above:

$$\ln \frac{K_D(T_2)}{K_D(T_1)} = \frac{\Delta H(T_1)}{R}\left(\frac{1}{T_2} - \frac{1}{T_1}\right) + \frac{\Delta C_P}{R}\left[1 - \frac{T_1}{T_2} - \ln\left(\frac{T_2}{T_1}\right)\right]$$

Antibodies

There are many, many protein binders in the cell.

Antibodies are biology's universal binders – they are large proteins specifically designed to bind just about anything. This underlies their function in the immune system.

Important in technology!

- Form the basis of many medical diagnostic tests – can bind to and signal very small concentrations of target molecules [ex: Epstein-Barr virus]
- Single largest group of biological therapeutics: 22 approved, 100+ in clinical
- play roles in auto-immune diseases

Antibodies are known as **immunoglobulins.** Often abbreviated Ig.

The basic Ig unit is Y-shaped and has a highly conserved structure:

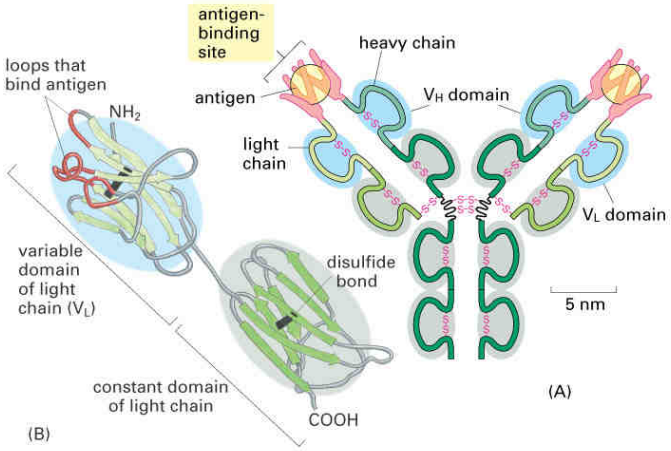


Figure 4-32 Essential Cell Biology, 2/e. (© 2004 Garland Science)

Notice that:

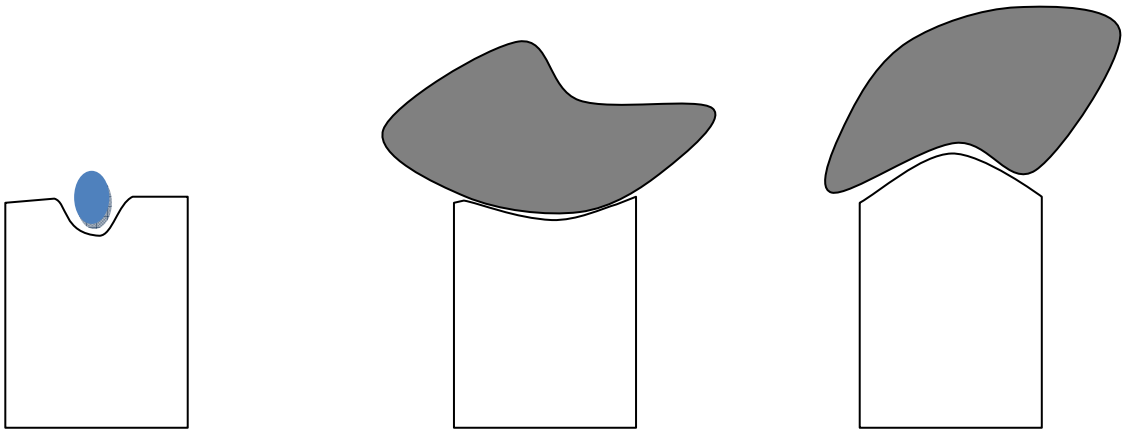
- four protein chains – 2 heavy and 2 light
- the chains form 12 separate protein **domains,** or separately folding units
- the domains are dominated by beta-sheet structures
- disulfide bonds hold the different chains together
- consists of ~1500 amino acids all together
- two **antigen-binding** domains → antigen is the target to which the Ig binds
- top regions are called **FAB (fragment, antigen binding)** and the bottom **Fc (fragment, crystallizable)**

Antibodies have highly variable "loop" regions in the binding site. The immune system works by:

- producing a large number of antibodies with different random binding sites, one per each plasma B cell (a kind of white blood cell) → system for making random mutations in the parts of the genes that produce these proteins corresponding to the binding site
- many (millions or billions) of random antibodies are produced

- binding of an antibody to a foreign agent triggers a cascade of reactions that signal the original B cell producing that antibody to multiply and produce more antibodies

How can antibodies bind? Several modes at the antigen binding site:



small molecules, peptides
deep pocket or groove

proteins
large concave surface

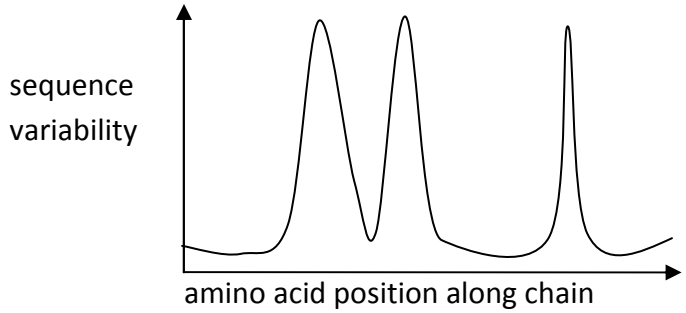
proteins
large convex surface

Antibodies can be broken into smaller fragments:

$$\mathrm{IgG} \xrightarrow{\text{Papain}} 2\mathrm{FAB} + 1\mathrm{FC}$$

$$\mathrm{IgG} \xrightarrow{\text{Pepsin}} (\mathrm{FAB})_2 + \mathrm{pFC\ (fragments)}$$

How did researchers find out where the binding sites were? Analysis of sequence variability (opposite of identity) in multiple antibodies:



Allostery and binding cooperativity

Even thought proteins fold to a unique conformational topology, many can also have fluctuations in the structure:

- atomic vibrations
- rotations, movement of side chains on the surface of the protein
- bending/hinge motions at flexible regions
- more concerted conformational changes throughout the molecule

Oftentimes the binding of a ligand to a protein structure can cause the protein fold to change slightly. Such behavior is called **allostery.**

Why is allostery important? It can change the physiochemical characteristics of other binding pockets on the same protein that modifies the way in which it interacts with other molecules.

Thus, it is a way to use one molecule to "switch" a protein between two conformational states with different binding or recognition properties.

Protein conformational populations dictated by free energies:

$$\frac{\wp(state\ 1)}{\wp(state\ 2)} = \exp\left(-\frac{G_1 - G_2}{RT}\right)$$

Example: cooperativity in hemoglobin

Hemoglobin is a **tetrameric** protein, involving four folded proteins that assemble into a unique, tightly bound quartenary structure

Hemoglobin tightly binds a porphyrin that that holds (chelate) an Fe^{2+}. The iron atom can in turn bind oxygen. Hemoglobin's function is to carry oxygen.

[show structure]

Let's model the binding of oxygen to hemoglobin. Up to four O_2 can bind. What if we assume the binding free energy is the same for each? In this case, we can treat each binding site as independent

$$Hb + O_2 \overset{K_{eq}}{\leftrightarrow} HbO_2$$

$$HbO_2 + O_2 \overset{K_{eq}}{\leftrightarrow} Hb(O_2)_2$$

$$Hb(O_2)_2 + O_2 \overset{K_{eq}}{\leftrightarrow} Hb(O_2)_3$$

$$Hb(O_2)_3 + O_2 \overset{K_{eq}}{\leftrightarrow} Hb(O_2)_4$$

$$K_{eq} = \exp\left(-\frac{\Delta G^{\circ}}{RT}\right)$$

Let's say we impose a bulk solution oxygen concentration → fix $[O_2]$

Then we have:

$$[HbO_2] = K_{eq}[O_2][Hb]$$

$$[Hb(O_2)_2] = K_{eq}[O_2][HbO_2] = K_{eq}^2[O_2]^2[Hb]$$

$$[Hb(O_2)_3] = K_{eq}[O_2][Hb(O_2)_2] = K_{eq}^3[O_2]^3[Hb]$$

$$[Hb(O_2)_4] = K_{eq}[O_2][Hb(O_2)_3] = K_{eq}^4[O_2]^4[Hb]$$

If there is a finite amount of hemoglobin in the solution

$$[Hb] + [HbO_2] + [Hb(O_2)_2] + [Hb(O_2)_3] + [Hb(O_2)_4] = [Hb]_0$$

where $[Hb]_0$ is the total hemoglobin concentration, regardless of binding. Substituting the above equations:

$$[Hb] + K_{eq}[O_2][Hb] + K_{eq}^2[O_2]^2[Hb] + K_{eq}^3[O_2]^3[Hb] + K_{eq}^4[O_2]^4[Hb] = [Hb]_0$$

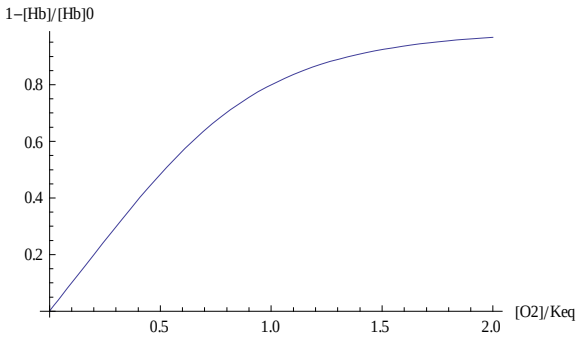
Dividing so as to find the fraction of free hemoglobin,

$$\frac{[Hb]}{[Hb]_0} = \frac{1}{1 + K_{eq}[O_2] + K_{eq}^2[O_2]^2 + K_{eq}^3[O_2]^3 + K_{eq}^4[O_2]^4}$$

The fraction of bound hemoglobin is

$$1 - [Hb]/[Hb]_0$$

What does this look like, as a function of oxygen concentration? **Saturation curve**



When such curves are measured, the shape of oxygen binding to hemoglobin doesn't look like this graph. Instead, it appears more sigmoidal. Something must be wrong with our assumed mechanism!

Let's try a different mechanism, in which we assume the four binding sites are occupied in an all-or-none fashion

$$Hb + 4O_2 \overset{K_{eq}}{\leftrightarrow} Hb(O_2)_4$$

$$[Hb(O_2)_4] = K_{eq}[Hb][O_2]^4$$

Total conservation of Hb:

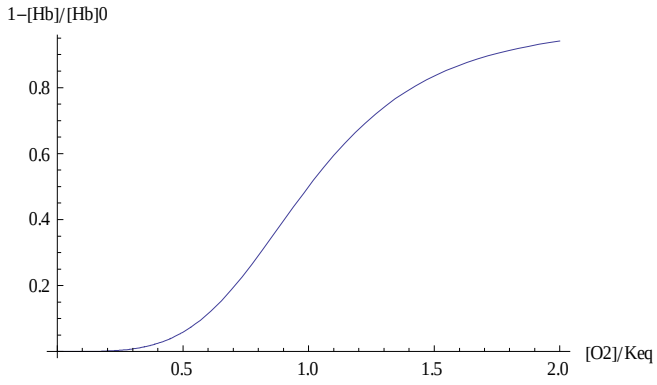
$$[Hb] + [Hb(O_2)_4] = [Hb]_0$$

$$[Hb] + K_{eq}[Hb][O_2]^4 = [Hb]_0$$

Thus

$$\frac{[Hb]}{[Hb]_0} = \frac{1}{1 + K_{eq}[O_2]^4}$$

Now, plotting the fraction that is bound:



Now, we get a signoidal saturation curve… this is closer to true experimental data!

Why might the oxygens bind in an all or none fashion? Why is this mechanism apparently more correct?

It turns out that the binding of oxygen to the porphyrin group requires a **conformational change** in the hemoglobin tetramer [show movie]

When the first oxygen molecule binds, it induces the conformational change such that it makes it then much easier for the other three binding sites to be filled.

This kind of behavior is called **cooperativity** – it is more favorable for multiple binding events to occur in cooperation (all at once) than one at a time. Many biological processes entail cooperativity.

Enzyme catalysis

Almost all biological processes inside the cell need catalysts in order to occur at appreciable rates.

Most catalysts are proteins that bind reactants to facilitate the reaction – these proteins are called **enzymes** → typically end in **"-ase"**

There are also some RNA molecules that catalyze reactions → very few, but very important; will discuss later

Enzymes lower the activation energy to facilitate reactions occurring at fast rates; they do not, however, change the free energy of reaction, since the reactants and products are the same

Unlike synthetic catalysts, enzymes are highly **specific**, each facilitating one particular reaction → also enantio-selective

Specificity is achieved by specific binding

How do enzymes lower the activation energy?

- Holding reactants in a configuration close to that of the transition state – reduces the entropy component of the transition state free energy
- Stabilizing the transition state configuration by creating interactions in the surrounding environment that are favorable with it (e.g., hydrogen bonds, electrostatics) – reduces the enthalpy component of the transition state free energy
- Allowing the formation of intermediate species that utilize chemical groups in the enzyme – alternate pathways

Enzymes only function when folded

Sometimes require **cofactor**

- small molecule (e.g., vitamin)
- metal (e.g., Mn, Zn, Mg, Fe)

Michaelis Menten kinetics

A simple model of enzyme kinetics is due to Leonor Michaelis and Maud Menten. Their model proposes the following:

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

Here

E = enzyme

S = substrate

P = product

Assumptions:

- product formation is **irreversible**, while binding of the substrate to the enzyme is **reversible.** [Are these assumptions justified?]
- the substrate is in large concentrations while the enzyme and product are very dilute. Therefore, the substrate concentration does not change substantially with time $[S] \approx [S]_0$

Write the reaction kinetics equations for this system:

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

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

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

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

Initial conditions:

$$[E] = [E]_0$$

$$[S] = [S]_0$$

$$[ES] = [P] = 0$$

Define the rate of product formation – this is what we are interested in!

$$v \equiv \frac{d[P]}{dt}$$

Can we solve this system of equations? What kinds of equations are they?

- first order – no second order derivatives
- linear – no! there is a coupling between different variables that makes it **nonlinear** due to the $[E][S]$ terms

This makes it very difficult to solve! Instead, we need to make approximations.

First approximation: **quasi steady-state** approximation. We assume that the concentration of the enzyme bound to the substrate, [ES], is approximately constant with time. OR, at least, the rate of its change with time is very small compared to the other species.

What does this assumption mean?

$$0 = k_1[E][S] - k_{-1}[ES] - k_2[ES]$$

Typically we do not know the individual concentration $[E]$ but we only know the total amount of enzyme that we put in to the system, which is constant with time:

$$[E]_0 = [E] + [ES] = constant$$

Therefore,

$$[E] = [E]_0 - [ES]$$

Combining this expression with the quasi steady state approximation

$$0 = k_1([E]_0 - [ES])[S] - k_{-1}[ES] - k_2[ES]$$

Solving for $[ES]$:

$$\begin{aligned}[ES] &= \frac{k_1[E]_0[S]}{k_1[S] + k_{-1} + k_2} \\ &= \frac{[E]_0[S]}{[S] + \frac{k_{-1} + k_2}{k_1}} \\ &= \frac{[E]_0[S]}{[S] + K_M}\end{aligned}$$

where we have defined a constant K_M by

$$K_M \equiv \frac{k_{-1} + k_2}{k_1}$$

Now, we are ready to compute the rate of product formation:

$$\begin{aligned} \frac{d[P]}{dt} = v &= k_2[ES] \\ &= \frac{k_2[E]_0[S]}{[S] + K_M} \\ &= \frac{v_{\max}[S]}{[S] + K_M} \end{aligned}$$

where we defined a second constant,

$$v_{\max} \equiv k_2[E]_0$$

Why the notation $v_{\max}$? Consider the limiting conditions of substrate concentration:

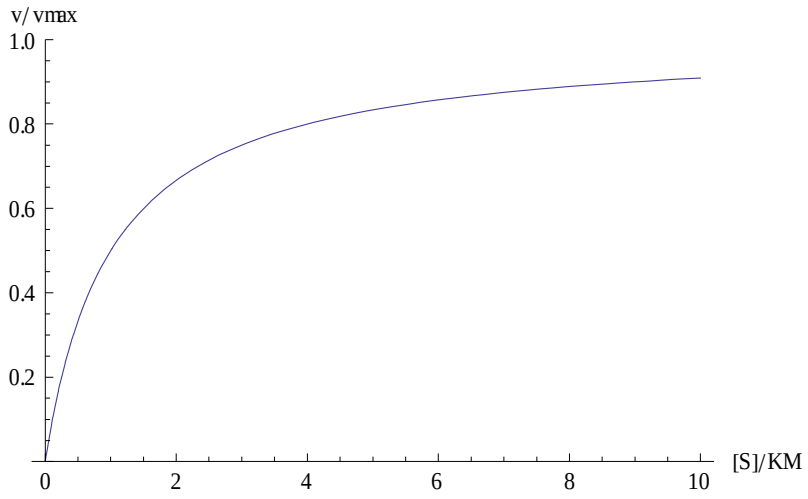
- For very high substrate concentrations, we expect almost all enzyme to be bound to substrate and the limiting rate of the reaction will them be due only to how fast the enzyme can perform the reaction:

$$[S] \text{ large:} \quad v \approx \frac{v_{\max}[S]}{[S]} = v_{\max}$$

- For low substrate concentrations, the limiting rate of reaction also depends on how much substrate can bind to enzyme:

$$[S] \text{ small:} \quad v \approx \frac{v_{\max}[S]}{K_M} = \frac{k_2}{K_M}[E]_0[S]$$

A plot of the rate as a function of substrate concentration looks like:



Such plots can be determined experimentally.

How to determine K_M? Consider:

$$v = \frac{1}{2} v_{\max}$$

then

$$\frac{1}{2} = \frac{[S]}{K_M + [S]}$$

Therefore,

$$K_M = [S]_{v=\frac{1}{2}v_{\max}}$$

Also, consider rearranging the rate equation

$$\frac{1}{v} = \frac{1}{v_{\max}} + \frac{K_M}{v_{\max}} \frac{1}{[S]}$$

Therefore a plot of $1/v$ versus $[S]$ yields a line [draw]:

- y-intercept gives $\frac{1}{v_{\max}}$
- slope gives $K_M / v_{\max}$

Such a plot is called a **Lineweaver-Burk** plot.

Competitive inhibition

What if there is another molecule that can bind to the enzyme, blocking the substrate? This is called **competitive inhibition.**

A general mechanism is:

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

$$E + I \underset{k_{-I}}{\overset{k_I}{\rightleftarrows}} EI$$

where "I" is an inhibitor.

How does this affect the rate? We now need to include an additional steady-state approximation for species EI:

$$\frac{d[ES]}{dt} = k_1[E][S] - k_{-1}[ES] - k_2[ES] \approx 0$$

$$\frac{d[EI]}{dt} = k_I[E][I] - k_{-I}[EI] \approx 0$$

The mass action expression for ES can be solved to yield:

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

The mass action expression for EI can be solved to yield:

$$[EI] = \frac{k_I}{k_{-I}}[E][I]$$
$$= \frac{1}{K_I}[E][I] \quad \text{where } K_I \equiv \frac{k_{-I}}{k_I}$$

Note that K_I is the equilibrium constant for the inhibition reaction.

We also have the condition:

$$[E]_0 = [E] + [ES] + [EI] = constant$$

Plugging in the above expressions for ES and EI, we get

$$[E]_0 = [E] + \frac{1}{K_M}[E][S] + \frac{1}{K_I}[E][I]$$

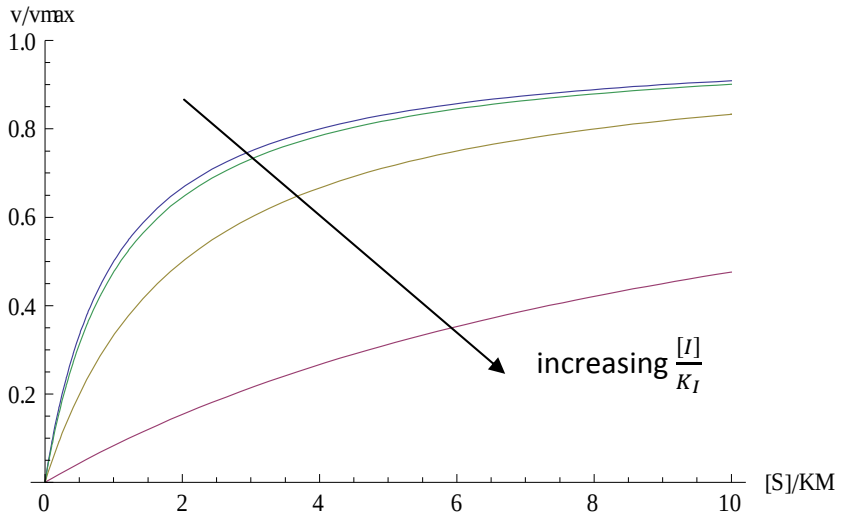
Solving for [E]:

$$[E] = [E]_0 \left(1 + \frac{[S]}{K_M} + \frac{[I]}{K_I}\right)^{-1}$$

Now we want the rate of reaction

$$\begin{aligned} v &= k_2[ES] \\ &= k_2 \frac{1}{K_M}[E][S] \\ &= k_2 \frac{1}{K_M}[E]_0 \left(1 + \frac{[S]}{K_M} + \frac{[I]}{K_I}\right)^{-1}[S] \\ &= (k_2[E]_0) \frac{[S]}{[S] + K_M + \frac{K_M}{K_I}[I]} \\ &= \frac{v_{\max}[S]}{[S] + K_M\left(1 + \frac{[I]}{K_I}\right)} \\ &= \frac{v_{\max}[S]}{[S] + K_{M,\text{app}}} \qquad \text{where } K_{M,\text{app}} \equiv K_M\left(1 + \frac{[I]}{K_I}\right) \end{aligned}$$

Notice that the final form of the equation looks like the Michaelis-Menten expression. The main difference is that we have an *apparent* K_M that now depends on the concentration of the inhibitor.



Consider a Linweaver-Burk plot:

$$\frac{1}{v} = \frac{1}{v_{\max}} + \frac{K_{M,\text{app}}}{v_{\max}} \frac{1}{[S]}$$

How to determine if competitive inhibition is present:

- Fix the concentration of the presumed inhibitor I
- Plot $1/v$ versus $1/[S]$ for different substrate concentrations $[S]$
- Determine $v_{\max}$ and $K_{M,\text{app}}$
- Repeat for different concentrations of inhibitor
- Then,

$$K_{M,\text{app}} = K_M + \frac{K_M}{K_I}[I]$$

A plot of $K_{M,\text{app}}$ versus $[I]$ will allow you to determine the true constant K_M and the inhibitor equilibrium constant K_I.

Product inhibition

What if the inhibitor is the product itself? → Not that surprising, since the product binds to the substrate at some point

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

$$E + P \underset{k_{-P}}{\overset{k_P}{\rightleftarrows}} EP$$

This is similar to the usual competitive inhibition. The equations still apply [try at home]:

Thus the rate expression is:

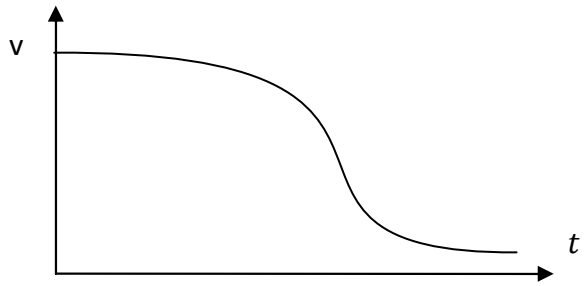
$$v = \frac{v_{\max}[S]}{[S] + K_M\left(1 + \frac{[P]}{K_P}\right)}$$

where

$$K_P \equiv \frac{k_{-P}}{k_P} = \frac{[E][P]}{[EP]}$$

What happens as time progresses?

- Initially $[P]$ very small
- But then it increases, and the reaction velocity *decreases*



Non-competitive inhibition

Another molecule binds the substrate and enzyme simultaneously

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

$$ES + I \underset{k_{-I'}}{\overset{k_{I'}}{\leftrightarrows}} ESI$$

Assume equilibrium for the substrate binding and for the inhibitor binding:

- $K_I' = \frac{[ES][I]}{[ESI]} \rightarrow [ESI] = \frac{[ES][I]}{K_I'}$
- $K_M = \frac{[E][S]}{[ES]} \rightarrow [ES] = \frac{[E][S]}{K_M}$

Consider that the total concentration of enzyme is constant:

$$[E] + [ES] + [ESI] = [E]_0 = const$$

Combining these yields

$$v = \frac{\left(\frac{v_{\max}}{1 + [I]/K_I'}\right)[S]}{\left(\frac{K_M}{1 + [I]/K_I'}\right) + [S]}$$

Therefore we get

$$v_{\max,\text{app}} = \frac{v_{\max}}{1 + \frac{[I]}{K_I'}}$$

$$K_{M,\text{app}} = \frac{K_M}{1 + \frac{[I]}{K_I'}}$$

Plotting Linweaver-Burk style:

$$\frac{1}{v} = \frac{1}{v_{\text{max,app}}} + \frac{K_{M,\text{app}}}{v_{\text{max,app}}}\frac{1}{[S]}$$

Note that in the ratio on the RHS, the factors involving the inhibitor cancel:

$$\frac{1}{v} = \frac{1}{v_{\text{max,app}}} + \frac{K_M}{v_{\text{max}}}\frac{1}{[S]}$$

Therefore, as [I] changes, we would expect only the intercept of the plot to change, not the slope. This can be a signal of non-competitive inhibition.

Substrate inhibition

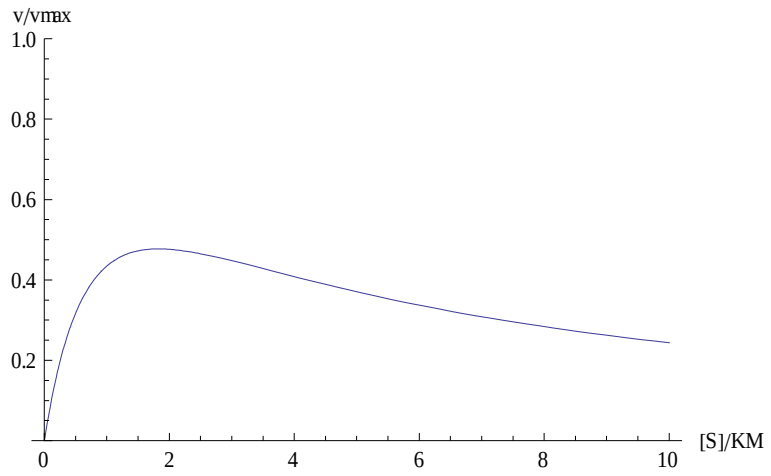
This is a special case of non-competitive inhibition. In this case, the substrate is also the inhibitor that binds with the enzyme:

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

$$ES + S \underset{k_{-I}}{\overset{k_I}{\leftrightarrows}} ES_2$$

In this case,

$$v = \frac{v_{\text{max}}[S]}{K_{M,\text{app}} + [S] + \frac{[S]^2}{K_S}}$$



Mixed inhibition

What if we have both competitive and non-competitive inhibition at the same time?

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

$$E + I \underset{k_{-I}}{\overset{k_I}{\rightleftarrows}} EI$$

$$ES + I \underset{k'_{-I}}{\overset{k'_I}{\rightleftarrows}} ESI$$

Assume equilibrium between all the intermediate species:

$$K_I = \frac{k_{-I}}{k_I} = \frac{[E][I]}{[EI]}$$

$$K'_I = \frac{k'_{-I}}{k'_I} = \frac{[ES][I]}{[ESI]}$$

Consider that the total concentration of enzyme is constant:

$$[E] + [ES] + [ESI] = [E]_0 = const$$

Combining all of these yields

$$v = \frac{v_{\max}[S]}{\left(1 + \frac{[I]}{K_I}\right) K_M + \left(1 + \frac{[I]}{K'_I}\right)[S]}$$
$$= \frac{v_{\max,\text{app}}[S]}{K_{M,\text{app}} + [S]}$$

where

$$v_{\max,\text{app}} = \frac{v_{\max}}{\left(1 + \frac{[I]}{K'_I}\right)}$$

$$K_{M,\text{app}} = K_M \frac{\left(1 + \frac{[I]}{K_I}\right)}{\left(1 + \frac{[I]}{K'_I}\right)}$$

Thus, both $v_{\max,\text{app}}$ and $K_{M,\text{app}}$ will have an $[I]$-dependence when extracted from a Linweaver-Burk plot, however the $[I]$ dependence will *differ*, unlike pure non-competitive inhibition.

Summary of inhibition mechanisms

mechanism	$v_{max,app}$	$K_{M,app}$	special cases
none	$v_{\max}$	K_M	
competitive	$v_{\max}$	$K_M\left(1+\frac{[I]}{K_I}\right)$	*product inhibition*: v decreases with time
non-competitive	$v_{\max}\left(1+\frac{[I]}{K_I'}\right)^{-1}$	$K_M\left(1+\frac{[I]}{K_I'}\right)^{-1}$	*substrate inhibition*: v decreases more than expected with increasing $[S]$
mixed	$v_{\max}\left(1+\frac{[I]}{K_I'}\right)^{-1}$	$K_M\left(1+\frac{[I]}{K_I}\right)\left(1+\frac{[I]}{K_I'}\right)^{-1}$	

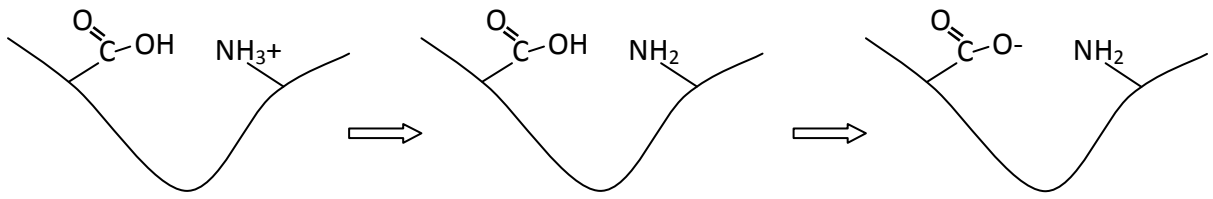
Enzyme regulation

The activity of enzymes can be controlled by a variety of means.

pH effects

Enzyme activity can be very sensitive to pH.

Consider a binding pocket with the following chemistry:



low pH moderate pH high pH

Notice that the interactions in the binding pocket change substantially because the electrostatics are now distinct.

How do we begin to model this? We need a model of which state is catalytically active. For this case, we will assume that it is the second, uncharged state that is active. We might then write:

$$EH + S \underset{k_{-1}}{\overset{k_1}{\longleftrightarrow}} EHS \overset{k_2}{\rightarrow} EH + P \qquad K_M \equiv \frac{[EH][S]}{[EHS]}$$

$$EH \overset{K_3}{\leftrightarrow} E^- + H^+ \qquad K_3 \equiv \frac{[E^-][H^+]}{[EH]}$$

$$EH + H^+ \overset{K_4}{\leftrightarrow} EH_2^+ \qquad K_4 \equiv \frac{[EH][H^+]}{[EH_2^+]}$$

Here, we assume that the protonation and deprotonation reactions are very fast and at equilibrium. This is a good approximation since in reality these processes happen extremely fast. We also assume that the bound species are in equilibrium with the reactants.

Mass balance on E:

$$\begin{aligned}[E]_0 &= [EH] + [E^-] + [EH_2^+] + [EHS] \\ &= [EH] + \frac{[EH]}{[H^+]}K_3 + \frac{[EH][H^+]}{K_4} + \frac{[EH][S]}{K_M}\end{aligned}$$

Solving for $[EH]$ and plugging into the reaction velocity

$$\begin{aligned}v &= \frac{v_{\max}[S]}{K_M\left(1 + \frac{K_3}{[H^+]} + \frac{[H^+]}{K_4}\right) + [S]} \\ &= \frac{v_{\max}[S]}{K_{M,\text{app}} + [S]}\end{aligned}$$

where

$$K_{M,\text{app}} \equiv K_M\left(1 + \frac{K_3}{[H^+]} + \frac{[H^+]}{K_4}\right)$$

Note that

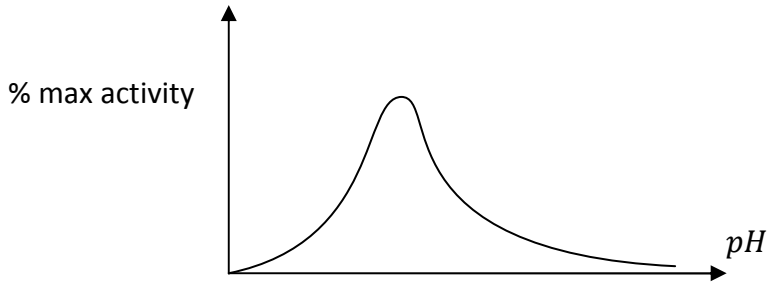
$$pH \sim -\log_{10}[H^+]$$

or

$$[H^+] \sim 10^{-\text{pH}}$$

Thus, we can predict the effect of pH on the reaction velocity.

Typical proteins have optimum conditions for functioning:



Allosteric regulation

Many proteins can be triggered to change conformation slightly by the binding of a small molecule or peptide

These **allosteric** changes can make the protein act as a switch, if one conformation is enzymatically active while another is not. The ligand can either **activate** or **inactivate** the enzyme:

$$\text{activation: } E\ (inactive) + L \rightarrow EL(active)$$

$$\text{inhibition: } E\ (active) + L \rightarrow EL(inactive)$$

There can also be many more complicated mechanisms, whereby a protein may have multiple activators and inhibitors.

These ligands typically act by binding to an **allosteric site** on the enzyme of interest – a second binding cavity away from the primary site involved in catalysis. By binding, the ligands effect a conformational change that affects the shape of the primary binding pocket.

Phosphorylation – the addition of phosphate groups to proteins—is widely used by eukaryotic cells to accomplish this task. Phosphate groups have a significant negative charge and their electrostatic interactions with a protein can easily effect conformational changes:

- **kinases** are protein enzymes that catalyze the transfer of a phosphate group from ATP to a protein
- **phosphatases** are protein enzymes that catalyze the removal of such phosphate groups
- these groups **regulate** the activity of the enzyme
- many interacting kinases and phosphatases form large **signaling cascades or networks** and enable cells to respond to a variety of different environmental conditions by inactivating/activating different enzymes

Allosteric changes not only enable activation & inactivation of enzymes, but can be used to produce movements that provide a mechanical response

The nucleotides **GTP** and **ATP** are widely used by cells in **motor proteins** that produce a net movement or that require concerted motions to accomplish complex tasks

Typically cycles of movement can be produced, according to:

1. ATP/GTP binds to a protein and causes a conformational change that produces some net movement
2. While bound to the protein, ATP/GTP is hydrolyzed to ADP/GDP, losing a phosphate group.
3. The loss of the phosphate weakens the binding affinity and ADP/GDP dissociates from the protein.
4. The protein returns to its original conformational state, ready for the next ATP/GTP to bind.