BIO-212 - Lecture 7
Molecular recognition and binding

Matteo Dal Peraro

Institute of Bioengineering, School of Life Sciences
École Polytechnique Fédérale de Lausanne

29th October 2025

Lecture 6 – Quick Summary

Free energy and chemical potential

Gibbs free energy gives you a quantitative way to understand the spontaneous direction of biological processes

$$G = H - TS$$

The chemical potential determines the equilibrium conditions for these biological processes and depends on their concentration

$$dG = (\mu_{in} - \mu_{out})\, dN_{in} < 0$$

$$\Delta\mu = \mu_2 - \mu_1 = RT \ln\left(\frac{C_2}{C_1}\right)$$

Equilibrium constant

From the chemical potential you can derive the K_{eq}

$$K_{eq} = \frac{[C]_{eq}^{\upsilon_C}\,[D]_{eq}^{\upsilon_D}}{[A]_{eq}^{\upsilon_A}\,[B]_{eq}^{\upsilon_B}}$$

which is related to the free energy (in an exponential way)

$$\Delta G^{o} = -RT \ln K_{eq} \qquad K_{eq} = e^{-\left(\frac{\Delta G^{o}}{RT}\right)}$$

remember the relevant application to protein folding/unfolding

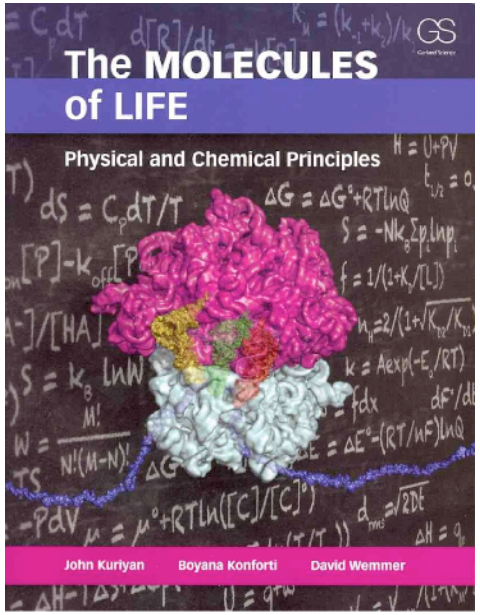
Lecture 7 - Outline

Today:

- Boltzmann distribution continued
- Molecular binding
- Thermodynamics of binding

Reading suggestions:

- The Molecules of Life (Chapters 12-13)



Entropy (statistical definition)

$$S = k_B \ln W$$

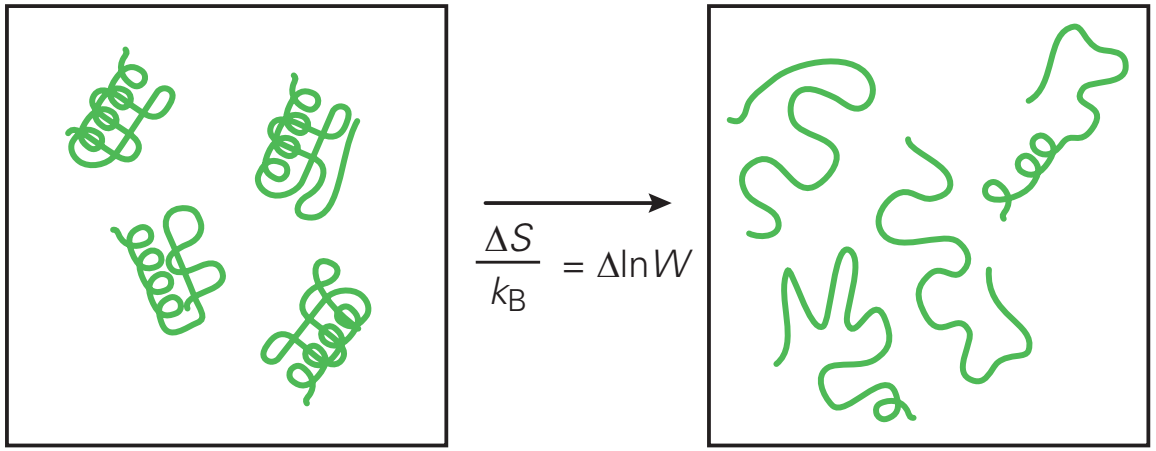
which is therefore extensive and a state function, and has the unit of an energy/temperature, J/K, like the Boltzmann constant.
This is equivalent to the **thermodynamic definition** of S:

$$\Delta S = \frac{q_{rev}}{T}$$

that you have seen derived from the study of heat engines (see page 330 Chapter 7 for a demonstration for ideal gas).
Thus spontaneous processes will increase entropy and at equilibrium S will be maximal

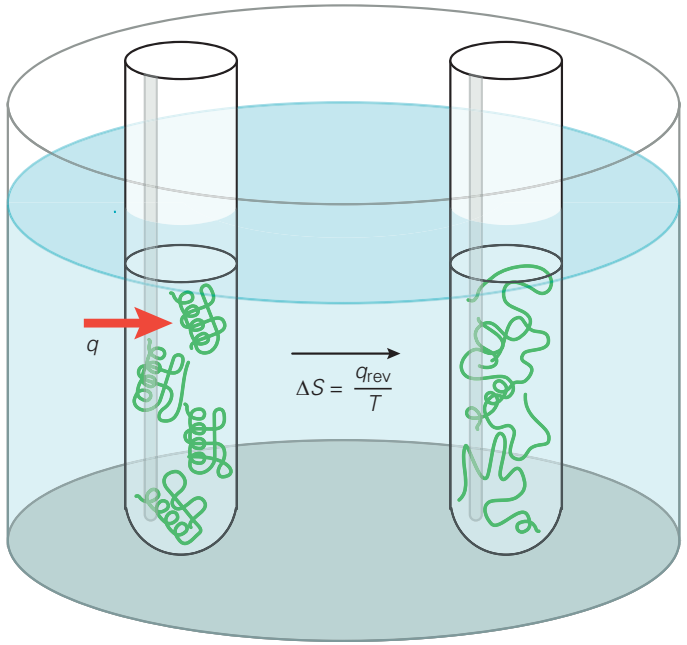
2nd law of thermodynamics - maximal entropy principle

statistical definition of S



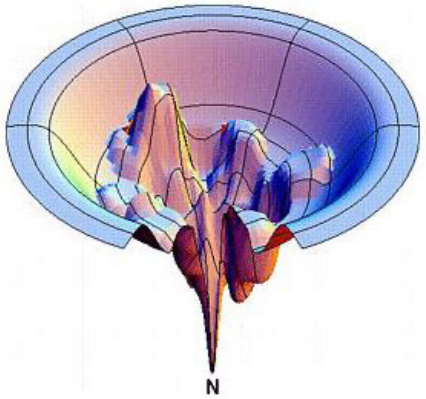
W = number of conformations or configurations

thermodynamic definition of S



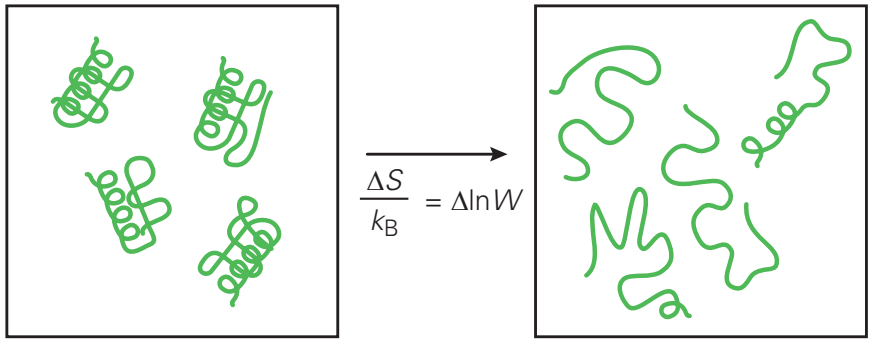
W is difficult to be estimated, but heat can be measured and this equivalence provides a way to reconnect *S* to the molecular features of the system - the link is T, temperature

Boltzmann Distribution



At low T → they all roll into the lowest valley (energy U dominates)

At high T → they spread out, exploring higher hills (entropy S dominates)



W = number of conformations or configurations

Temperature measures how much they care about exploring versus settling

$$P(U) \propto e^{-U/k_B T}$$

mathematical way of saying low energy is good, but higher energy can still happen — less often, by an amount that depends on temperature

consider our folding problem

$$\mathrm{U} \rightleftharpoons \mathrm{F}$$

$$K_{\text{folding}} = \frac{[\mathrm{F}]}{[\mathrm{U}]} \text{ (at equilibrium)} = \frac{P_F}{P_U} = e^{-(E_F - E_U)/k_B T}$$

that's the microscopic explanation of $K = e^{-\Delta G^{\circ}/RT}$

Boltzmann Distribution and Energetic Levels

If you have a system with N molecules and total energy U, when N is large (~N_A), it is difficult to know how the energy is distributed through N atoms. Thus we can only describe in statistical terms the population of a state, i.e. the N_i – number of molecules that will be found in an energy level with energy U_i.

$$P(N_i) = \frac{N_i}{N} = \frac{e^{-U_i/k_B T}}{\sum_i e^{-U_i/k_B T}}$$

thus $$N_i = \frac{N e^{-U_i/k_B T}}{Q}$$

where Q is the **partition function**

$$Q = \sum_i e^{-U_i/k_B T}$$

and k_B is the **Boltzmann constant** (k_B = 1.381 x 10^{-23} J/K)

but Q is constant at a given temperature, thus $$N_i \propto e^{-U_i/k_B T}$$

thus you can estimate the ratios between different populations at different energy levels using the following relation:

$$\frac{N_2}{N_1} = e^{-\Delta U/k_B T}$$

$$\Delta U = U_2 - U_1$$

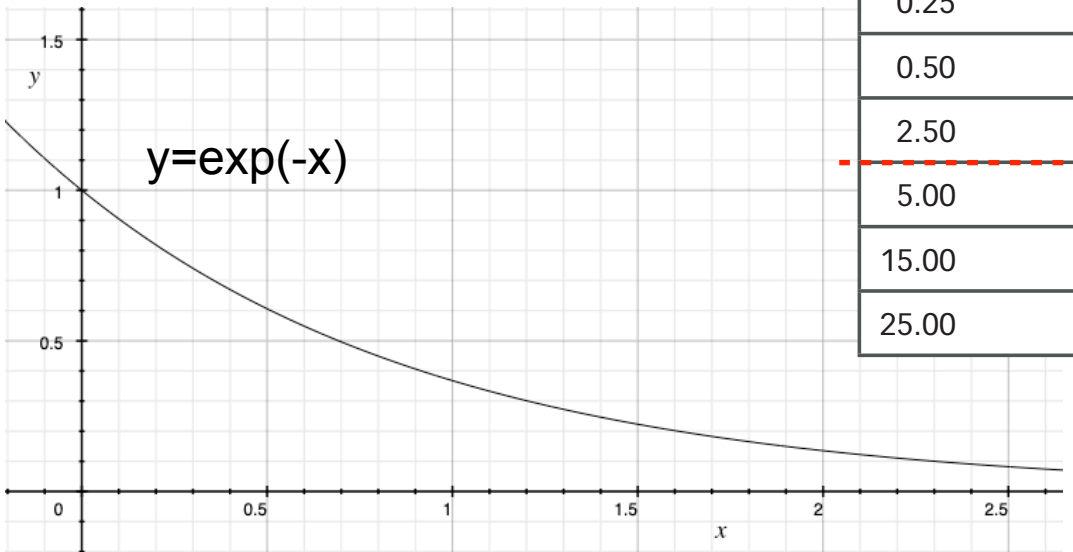
- Remember that the gas constant R is the "molar" form of k_B, in fact:
- $R = N_A k_B$ = 8.3145 J/K*mol, thus if you work with KJ/mol you have to use RT in the Boltzmann distribution

Boltzmann Distribution and Energetic Levels

we can simply access the ratios between populations of different energy levels using the following relation:

$$\frac{N_2}{N_1} = e^{-\Delta U / k_B T} \qquad \Delta U = U_2 - U_1 \qquad \frac{N_2}{N_1} = e^{-\Delta U / 2.529}$$

using as unit kJ/mol



ΔU (kJ•mol^{-1})	$\frac{\Delta U}{k_B T}$ (T = 300 K, $k_B T \approx$ 2.5 kJ•mol^{-1})	$e^{\frac{-\Delta U}{k_B T}}$
0.25	0.1	0.90
0.50	0.5	0.61
2.50	1.0	0.37
5.00	2.0	0.13
15.00	6.0	0.00067
25.00	10.0	0.0000045

$k_B T$

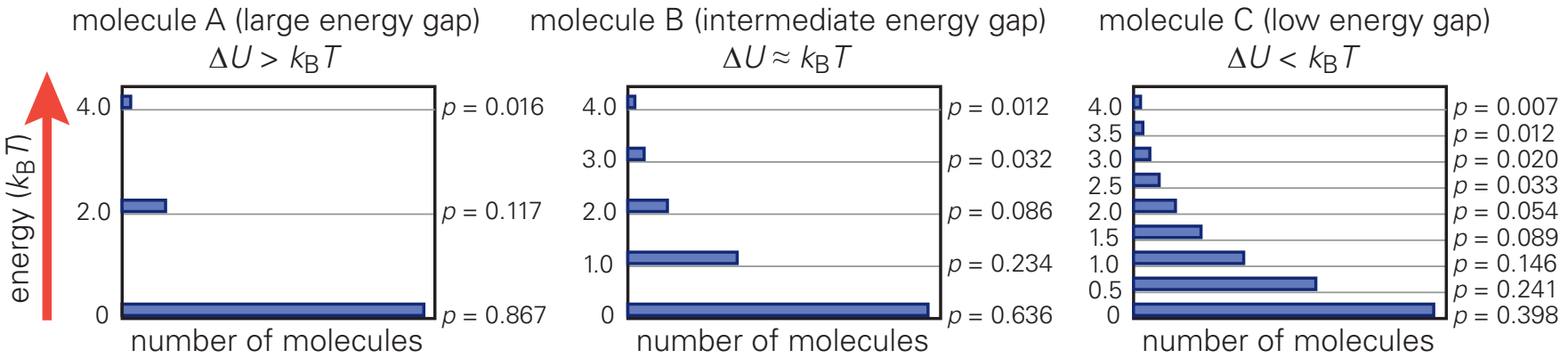
Boltzmann Distribution and Energetic Levels

we can simply access the ratios between populations of different energy levels using the following relation:

$$\frac{N_2}{N_1} = e^{-\Delta U / k_B T} \qquad \Delta U = U_2 - U_1$$

$$\frac{N_2}{N_1} = e^{-\Delta U / 2.529}$$

using as unit kJ/mol



3 molecules with different accessible energy levels - levels that are less spaced (< kT) are more accessible

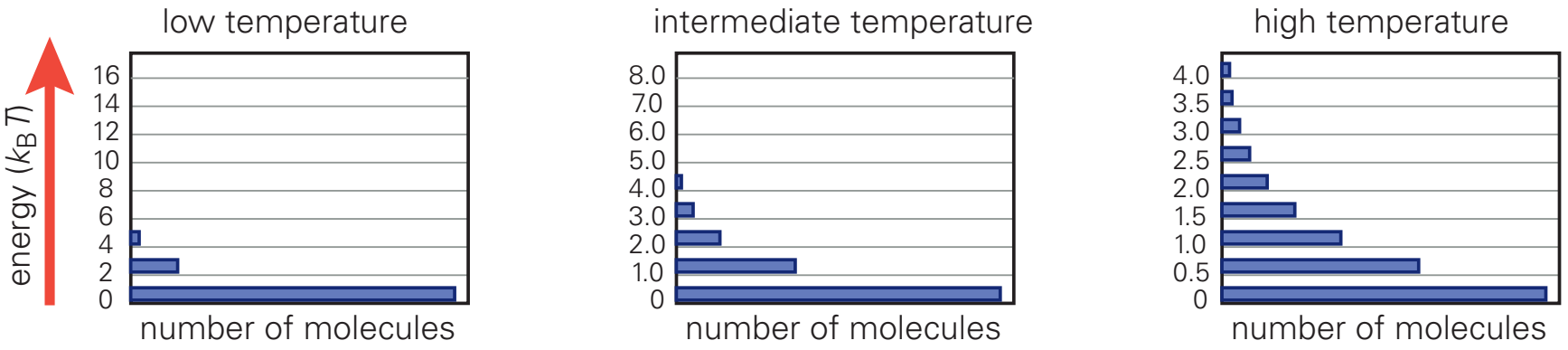
Boltzmann Distribution and Energetic Levels

we can simply access the ratios between populations of different energy levels using the following relation:

$$\frac{N_2}{N_1} = e^{-\Delta U / k_B T} \qquad \Delta U = U_2 - U_1$$

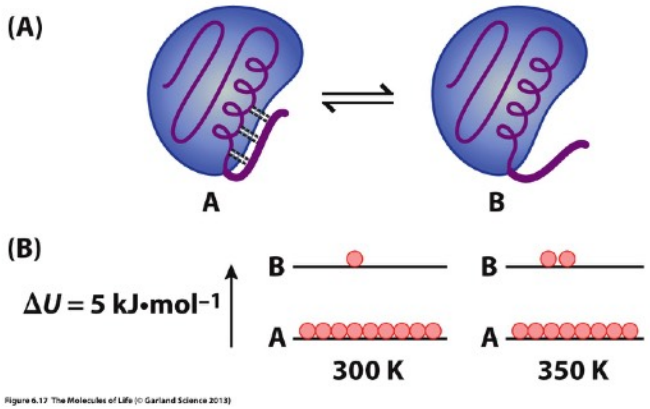
$$\frac{N_2}{N_1} = e^{-\Delta U / 2.529}$$

using as unit kJ/mol



same molecule at different T, the occupancy of energy levels increases with T

Boltzmann Distribution in macromolecules



Protein molecules take up energy as they unfold

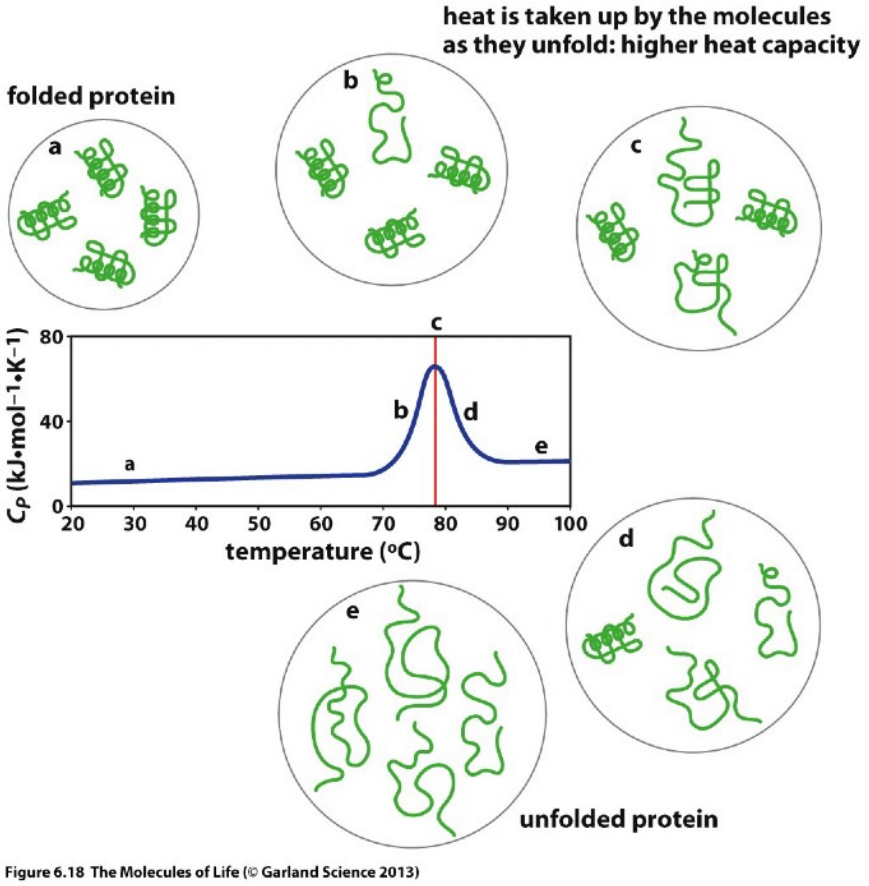
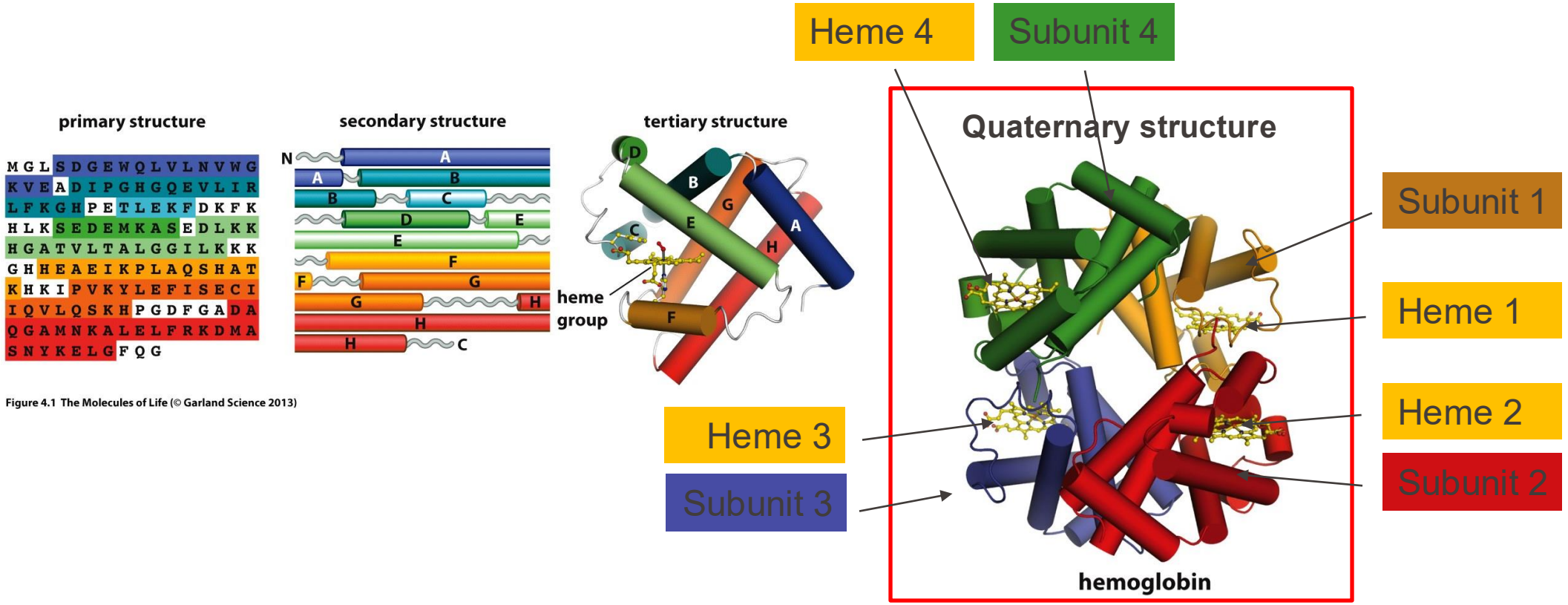


Figure 6.18 The Molecules of Life (© Garland Science 2013)

Shifting the distribution of populations with temperature

One can see how the formalism of the Boltzmann distribution helps us to describe what occurs in proteins and other biomolecules.

Quaternary structure (from Lecture 4)

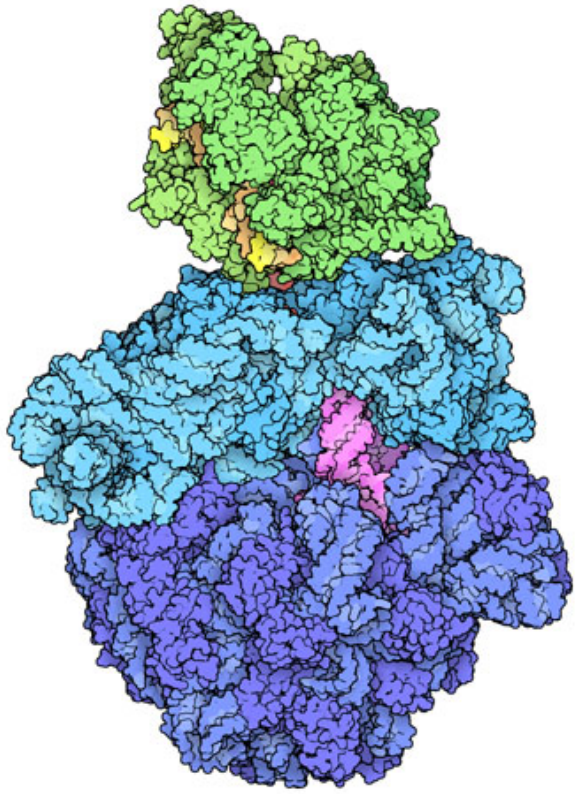


• Quaternary structure refers to the association of different polypeptide chains (subunits) into a multimeric complex with some molecular function.

• For example, hemoglobin is composed of 4 subunits of the same protein (+ heme group). Quaternary structure can also refer to complexes with DNA, RNA, lipids or any other biomolecule.

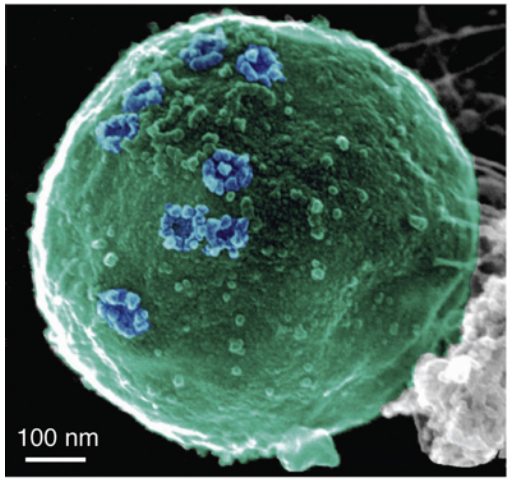
Why binding is that important ?

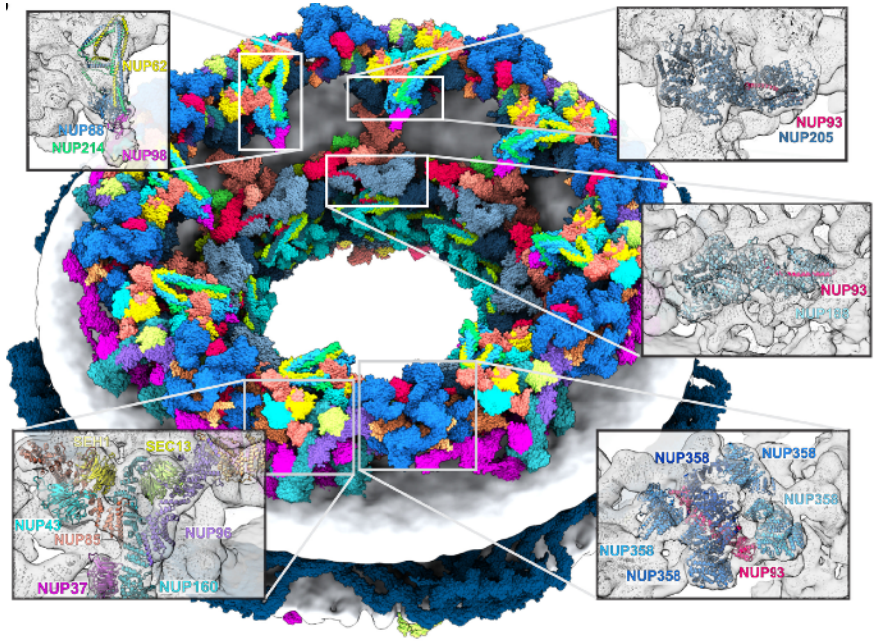
Because all cellular processes are controlled by the way that different molecules interact with each other, for example recognition of proper s**ubstrates by enzymes**, the **transmission of cellular signals**, the recognition of one cell by another, the **control of transcription and translation**, and the fidelity of **DNA replication**.



Expressome (PDB)

e.g., the Nuclear Pore Complex

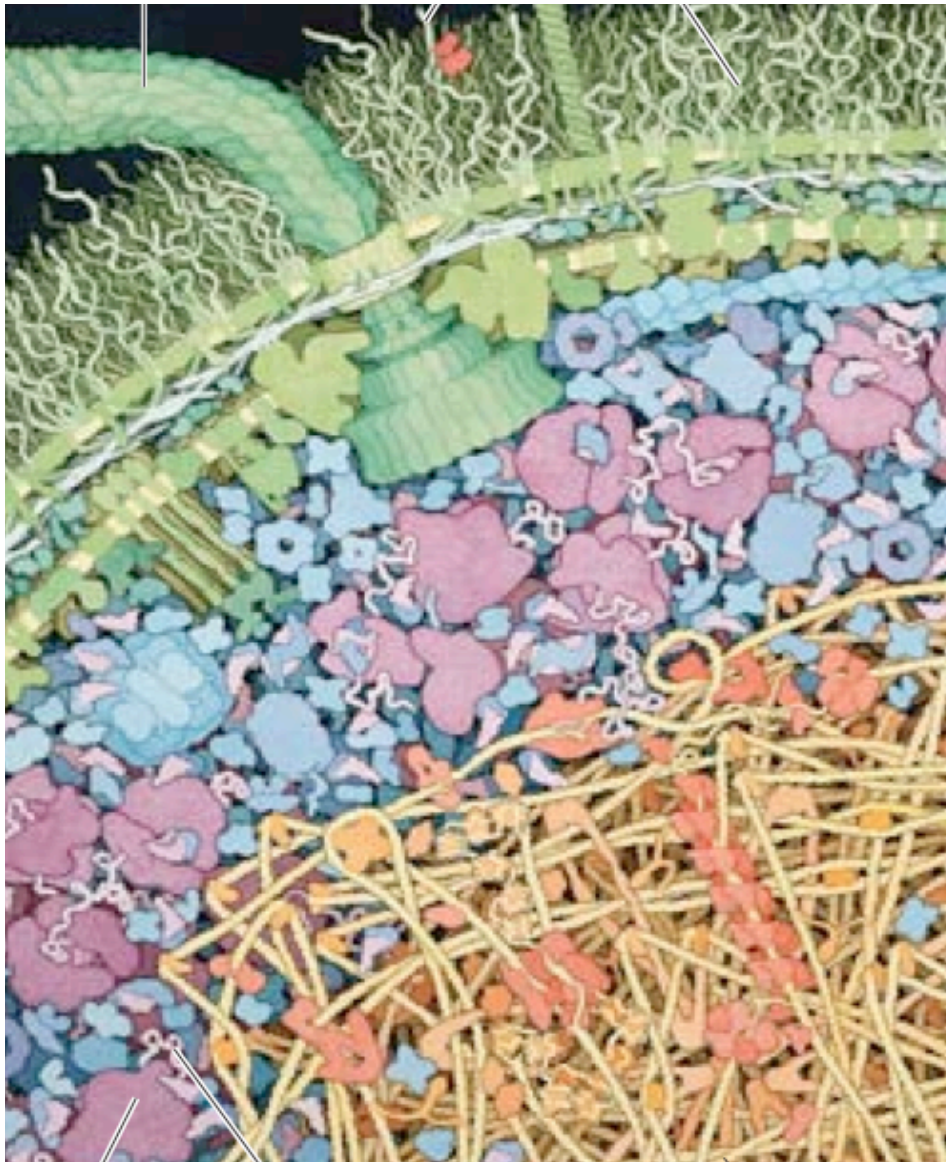




yeast NPC :
~52 MDa complex
~550 protein subunits of ~30 different types

DOI: 10.1126/science.abm9506

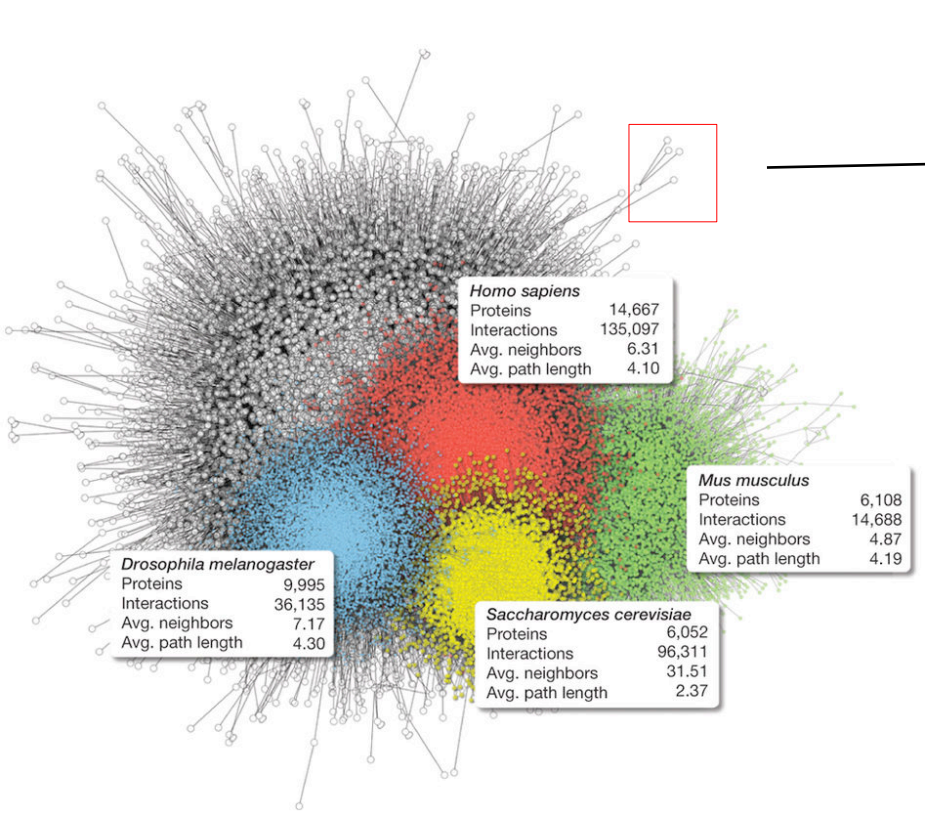
Why binding is that important ?

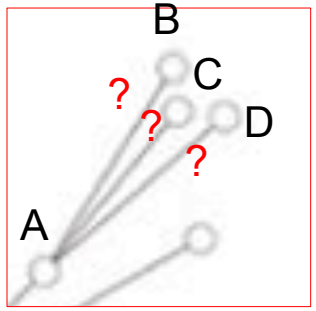


You can imagine that such extensive interaction networks are at least as complex for:

- protein-dna interactions
- protein-small molecule interactions
- protein-lipid interactions as well have started to be recognised as very important

Why binding is that important ?





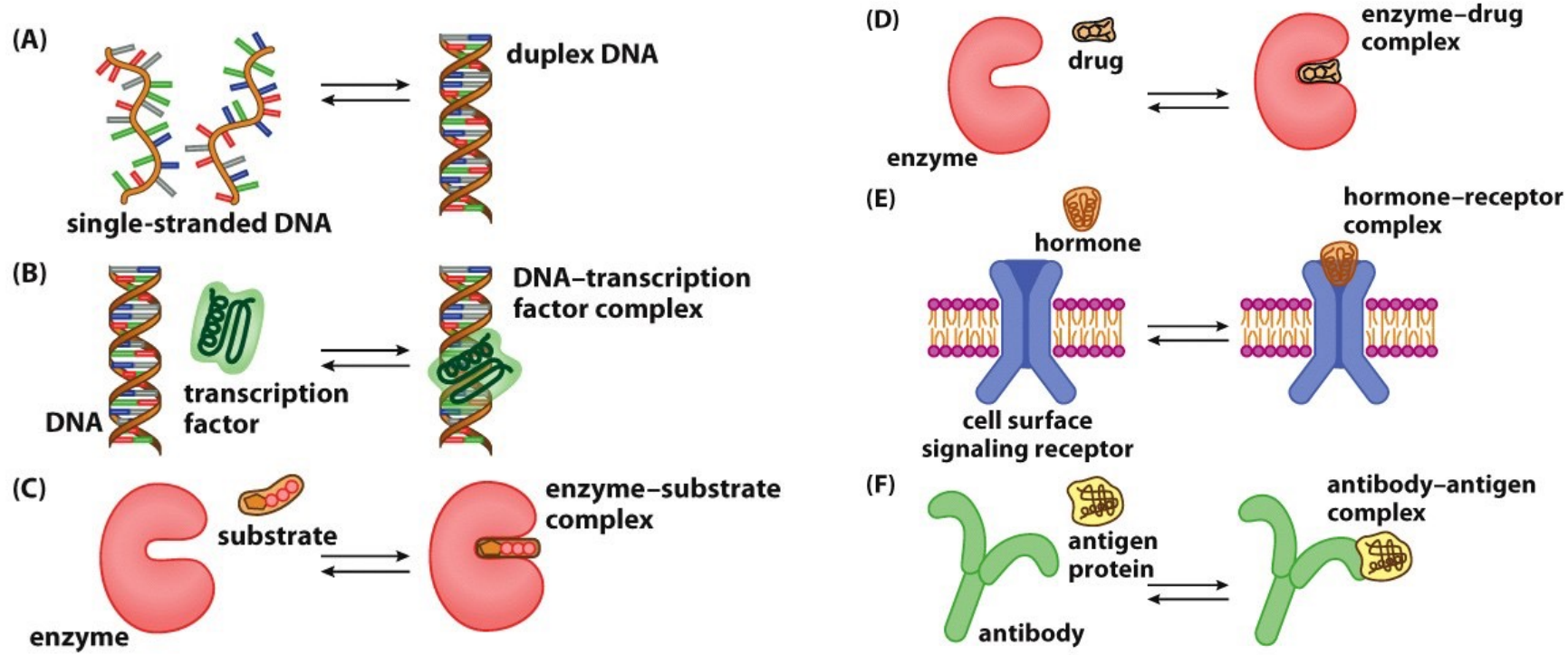
For a **qualitative** description of the this submap its enough to know that A binds B,C and D somehow

However, for a **quantitative** description it is essential to determine measurable quantities to these edges – i.e. binding affinities

Measuring the concentration of the free and associated species at equilibrium, we can calculate the strength of the molecular interactions.

Viewing Binding as a Chemical Reaction

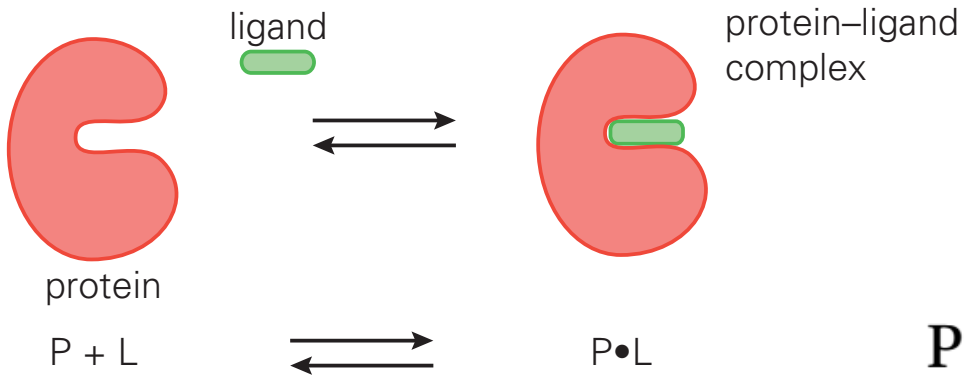
Some examples of important molecular recognition events:



Last lecture we talked about equilibrium constants for chemical reactions, if in general we consider a protein as target **P** and a ligand **L** as the interacting molecule via non covalent interactions the general complex P•L can be considered the product of the following reaction:

$$P + L \rightleftharpoons P \bullet L$$

Looking at the reaction mechanism



$$P + L \rightleftharpoons P \bullet L$$

Equilibrium constant:

$$K = \frac{[P \bullet L]}{[P][L]}$$

Because it is a binding reaction we call it **association constant:**

$$K_A = \frac{[P \bullet L]}{[P][L]}$$

Thus we can define the **binding fee energy** for the association as:

$$\Delta G^{o}_{bind} = -RT \ln K_A$$

which is a measure of the **affinity** of the interaction, that is, how strongly the molecules bind to each other.

Looking at the reaction mechanism

Energetic Landscape of Binding

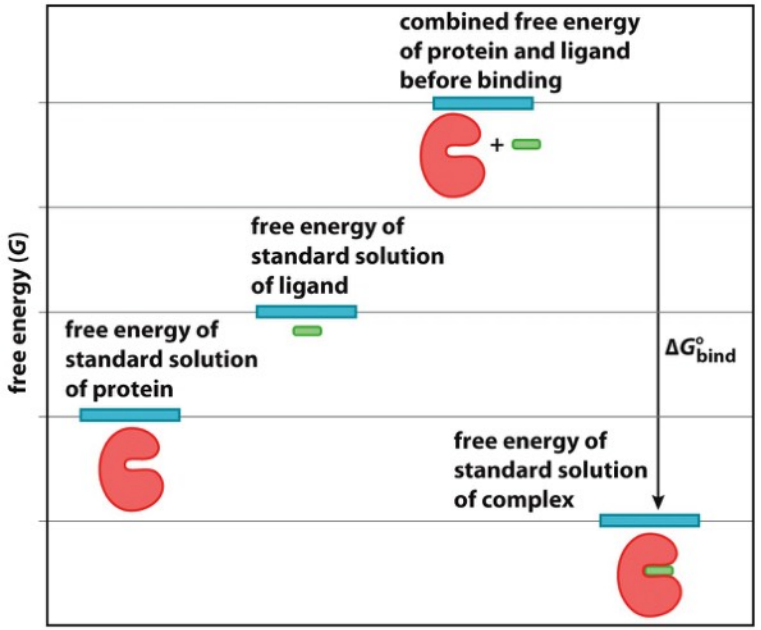


Figure 12.3 The Molecules of Life (© Garland Science 2013)

$$P + L \rightleftharpoons P \bullet L$$

$$K_A = \frac{[P \bullet L]}{[P][L]}$$

$$\Delta G^{o}_{bind} = -RT \ln K_A$$

Definition of Binding Affinity:

- it refers to the strength of a molecular interaction
- the greater the decrease in free energy upon binding the greater the affinity

Energetics of Binding

it is always true that entropy is lost upon binding of the small molecule, but

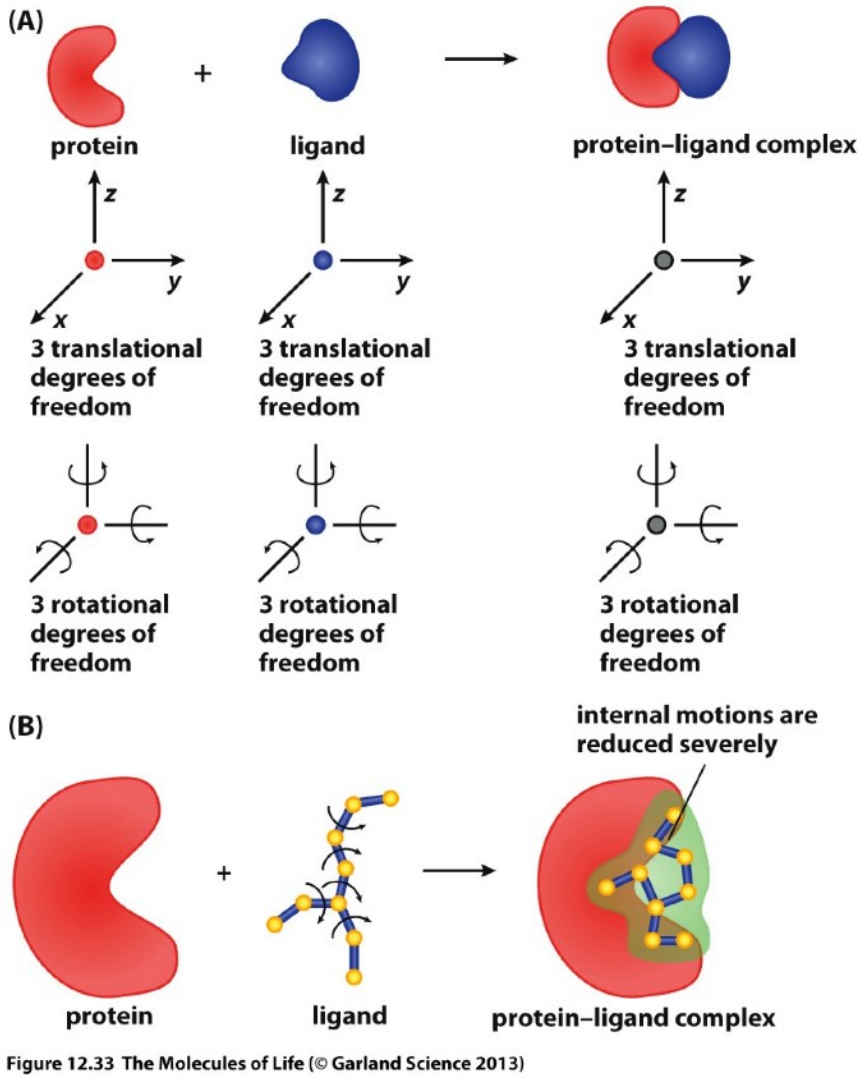


Figure 12.33 The Molecules of Life (© Garland Science 2013)

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

- There are entropic losses on binding which are not energetically favorable

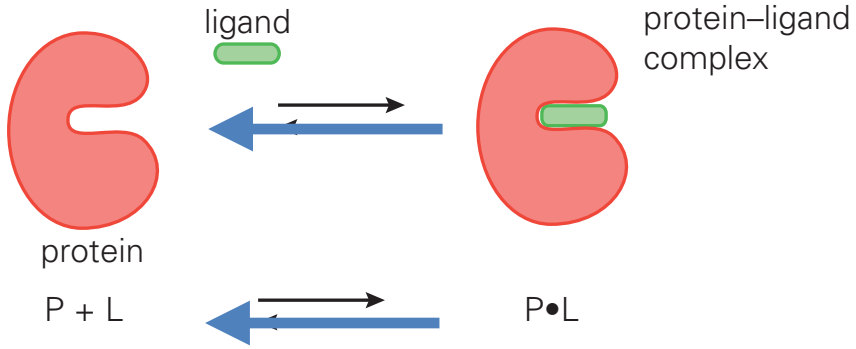
- Enthalpic gains on the interactions of the protein with the ligand

- Entropic gains on the solvent (ie hydrophobic effect)

thus overall $\Delta G < 0$

Looking at the reaction mechanism

It is however common practice to characterize the affinity of a binding interaction in terms of the **dissociation reaction,** which is associated to the **dissociation constant K_D**



$$P \bullet L \rightleftharpoons P + L$$

$$K_D = \frac{[P][L]}{[P \bullet L]} = \frac{1}{K_A}$$

$$\Delta G^{o}_{bind} = +RT \ln K_D$$

K_D is a dimensionless quantity, but usually it is discussed as if it had molar units of concentration. K_Ds that range from picomolar to nanomolar (10^{-12} -10^{-9} M, ~ -50 kJ/mol) are the tightest interactions, if in the order of millimolar (10^{-3} M, ~ -15 kJ/mol) they are the weakest.

Looking at the reaction mechanism

common practice to express K_D in molar terms because concentration at standard conditions are kind of neglected

$$K_D = \left(\frac{[P \bullet L]^o}{[P]^o [L]^o} \right) \frac{[P][L]}{[P \bullet L]} = \left(\frac{[P \bullet L]^o}{[P]^o [L]^o} \right) K_D^*$$

Some examples and affinity ranges :

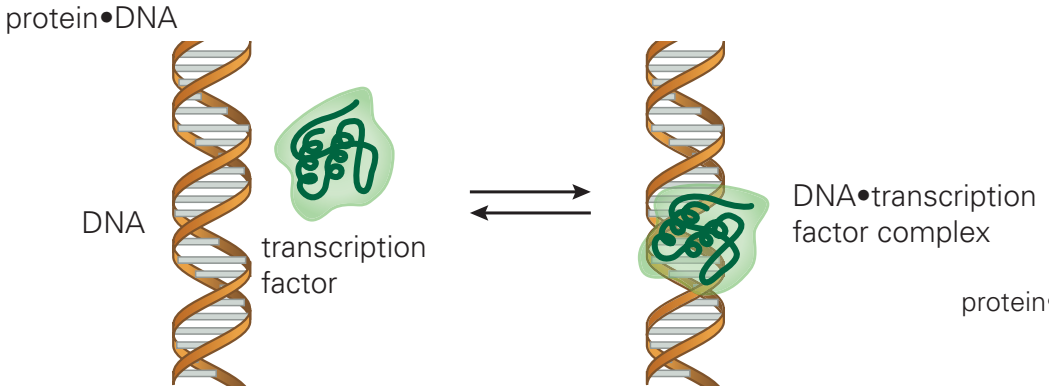
Type of interaction	K_D (molar)	ΔG^o_{bind} (at 300 K) (kJ•mol^{-1})
Enzyme–ATP	~1 × 10^{-3} to ~1 × 10^{-6} (millimolar to micromolar)	–17 to –35
Signaling protein binding to a target	~1 × 10^{-6} (micromolar)	–35
Sequence-specific recognition of DNA by a transcription factor	~1 × 10^{-9} (nanomolar)	–52
Small molecule inhibitors of proteins (drugs)	~1 × 10^{-9} to ~1 × 10^{-12} (nanomolar to picomolar)	–52 to –69
Biotin binding to avidin protein (one of the strongest known noncovalent interactions)	~1 × 10^{-15} (femtomolar)	–86

The trade-off between **affinity** and **specificity** is crucial in many biological processes

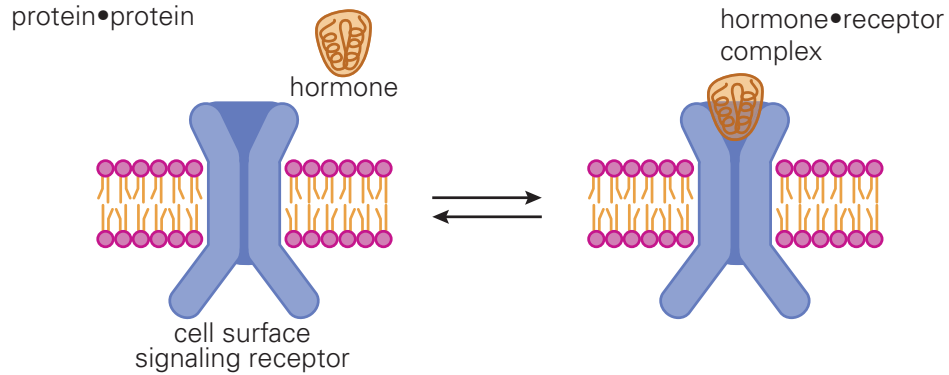
Affinity vs. specificity

Affinity of an interaction defines its strength (ie K_D).
But affinity alone is most of the times not sufficient as if you bind with high affinity to multiple targets then you will engage with targets that are not specific for a given biological function (ie off-target binding)

Specificity thus is the affinity of the ligand for one and only one specific target of interest

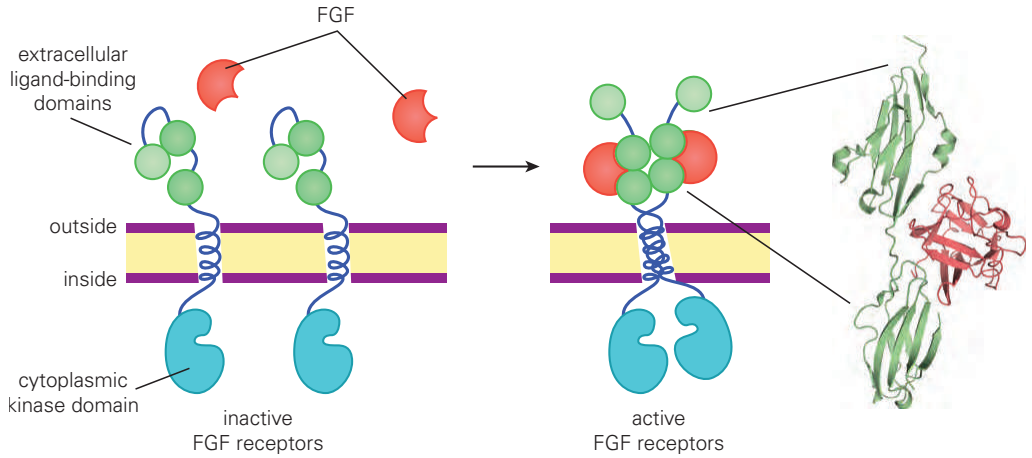


Specificity is mainly achieved by polar, directional interactions (eg H-bonds).

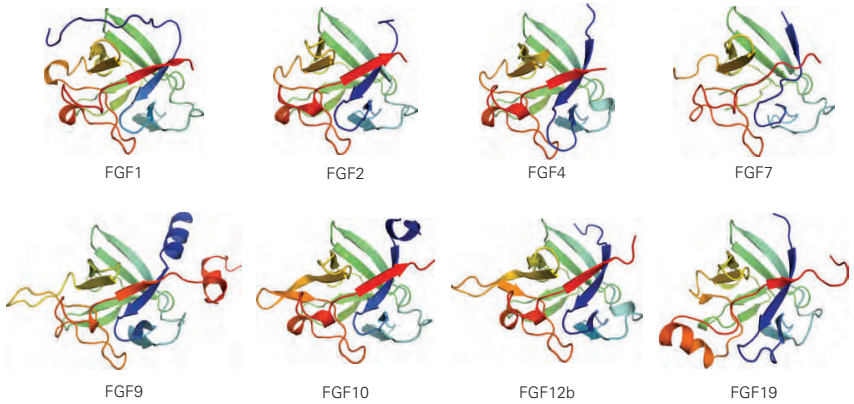


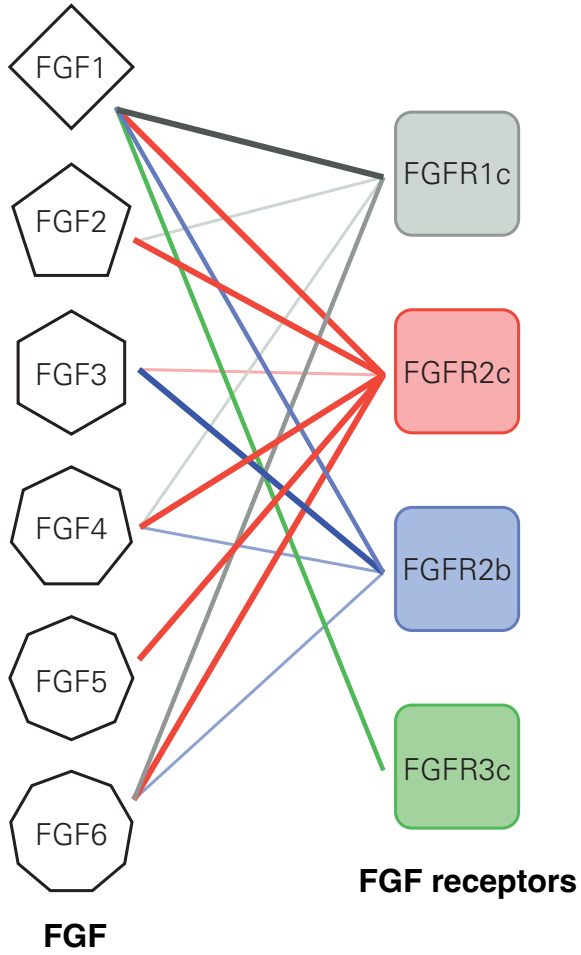
Affinity vs. specificity

Affinity of an interaction defines its strength (ie K_D).
Specificity is the affinity of the ligand for one and only one specific target of interest



Interaction between **fibroblast growth factor** (**FGF**) and its receptor





How to determine K_D

The value of K_D corresponds to the concentration of free ligand at which protein is half saturated

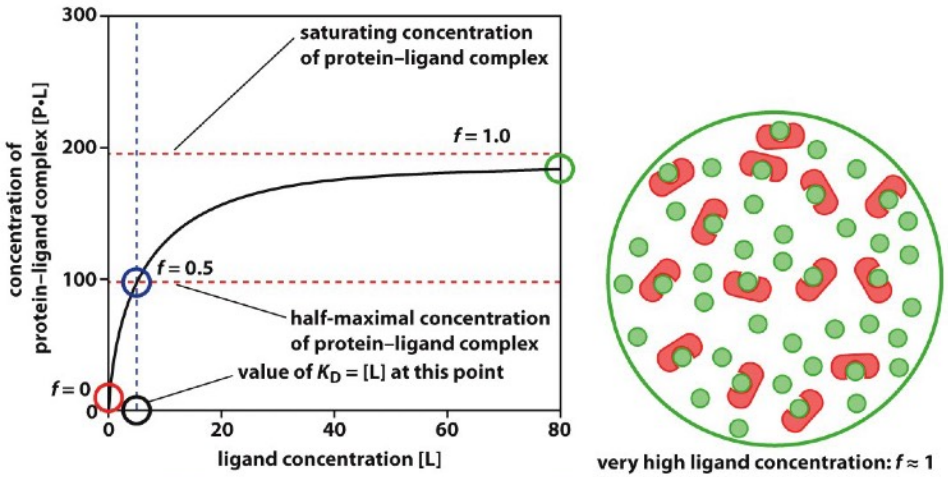


Figure 12.4 The Molecules of Life (© Garland Science 2013)

Let's see why this is true:
we call ***f*** the **fractional saturation** or fractional occupancy of the ligand binding sites

⇩

which is the extent to which the binding sites on a protein are filled with ligand

How to determine K_D

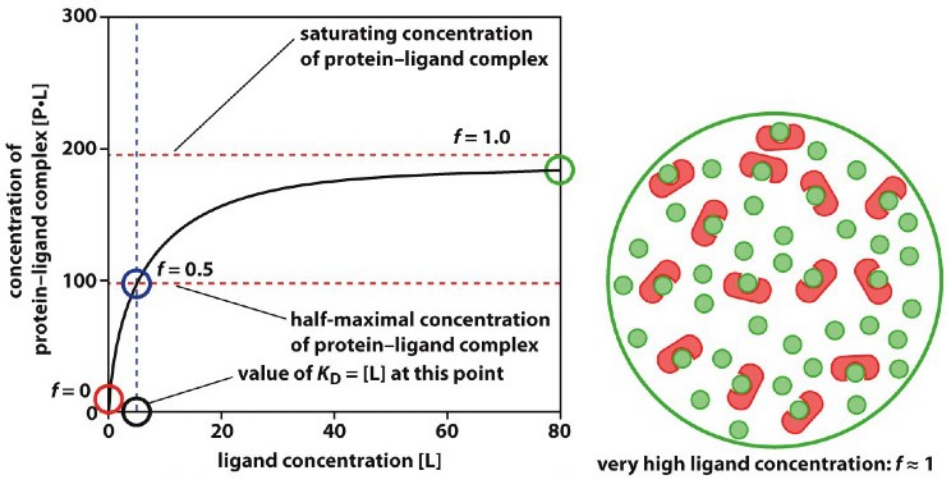


Figure 12.4 The Molecules of Life (© Garland Science 2013)

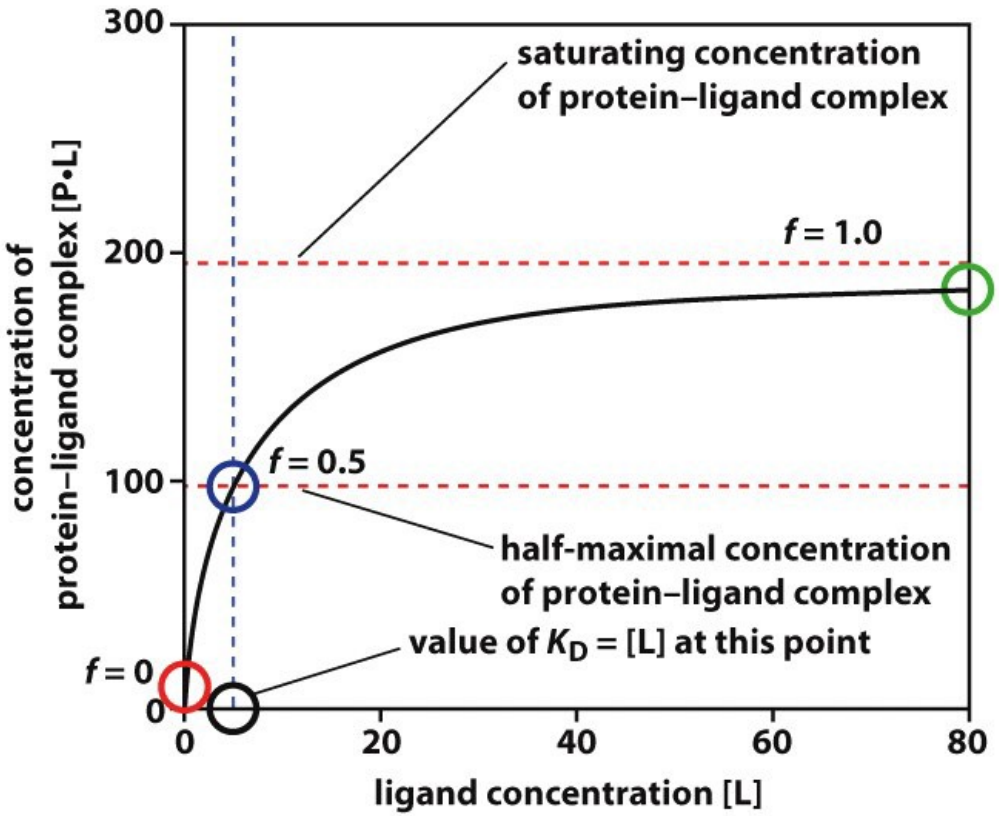
Let's see how this is true:
we call ***f*** the **fractional saturation** or fractional occupancy of the ligand binding sites

$$f = \frac{\text{concentration of protein with ligand bound}}{\text{total protein concentration}} = \frac{[P\bullet L]}{[P]+[P\bullet L]}$$

Recalling that : $[P\bullet L] = \frac{[P][L]}{K_D}$

$$f = \frac{[P][L]}{K_D\left([P]+\frac{[P][L]}{K_D}\right)}$$

How to determine K_D



$$f = \frac{\text{concentration of protein with ligand bound}}{\text{total protein concentration}} = \frac{[P\bullet L]}{[P]+[P\bullet L]}$$

$$\Downarrow$$

$$f = \frac{[P][L]}{K_D\left([P]+\frac{[P][L]}{K_D}\right)}$$

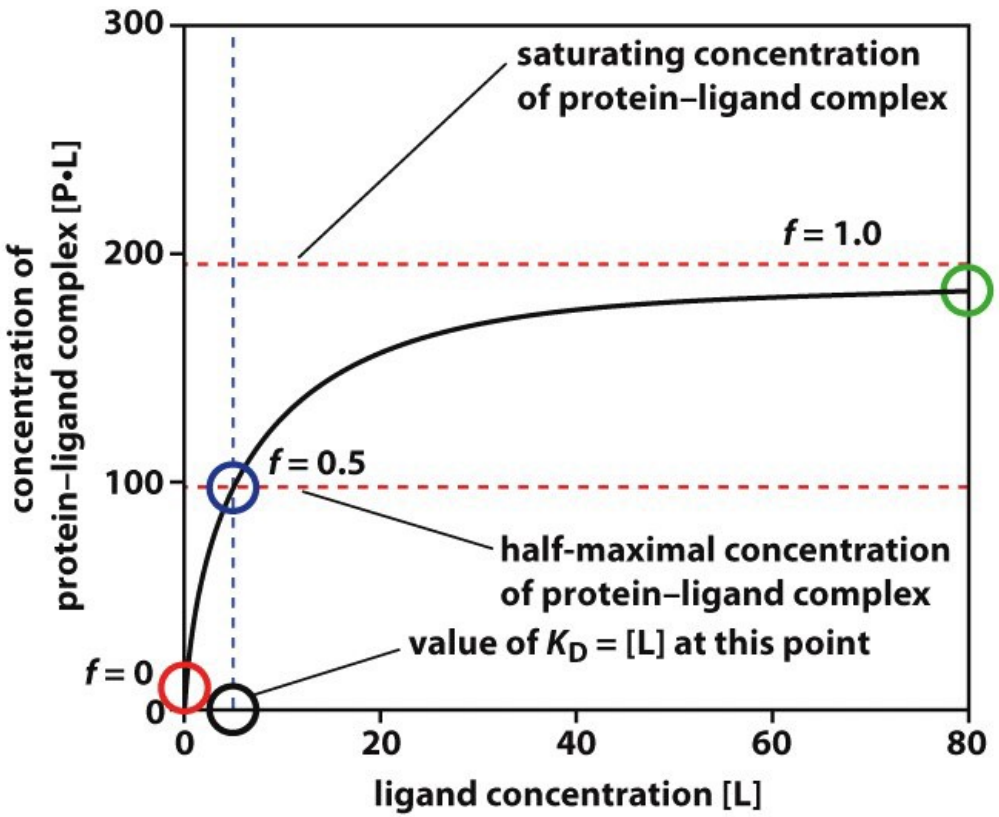
$$\Downarrow$$

$$\Rightarrow f = \frac{[L]}{K_D\left(1+\frac{[L]}{K_D}\right)} = \frac{[L]}{K_D+[L]} = \frac{\frac{[L]}{K_D}}{1+\frac{[L]}{K_D}}$$

(rectangular hyperbolic function)

$$\boxed{f = \frac{[L]}{[L]+K_D}}$$

How to determine K_D



$$f = \frac{[\mathrm{L}]}{[\mathrm{L}] + K_\mathrm{D}}$$

When $[\mathrm{L}] = K_\mathrm{D}$

$$f = \frac{K_\mathrm{D}}{K_\mathrm{D} + K_\mathrm{D}} = \frac{1}{2}$$

Therefore, a plot of fractional saturation as function of ligand concentration is known as a **binding isotherm** or **binding curve**. The K_D value depends on the temperature.

Binding assays

In all cases to estimate dissociation constants we need to come up with some **experimental** procedure able to measure the amount of ligands bound to a protein. These are called **binding assays** - there are many ways to develop a binding assay based on many different techniques and properties of the ligand or proteins. Here is an example based on **radioactivity** applied to steroid hormone receptors

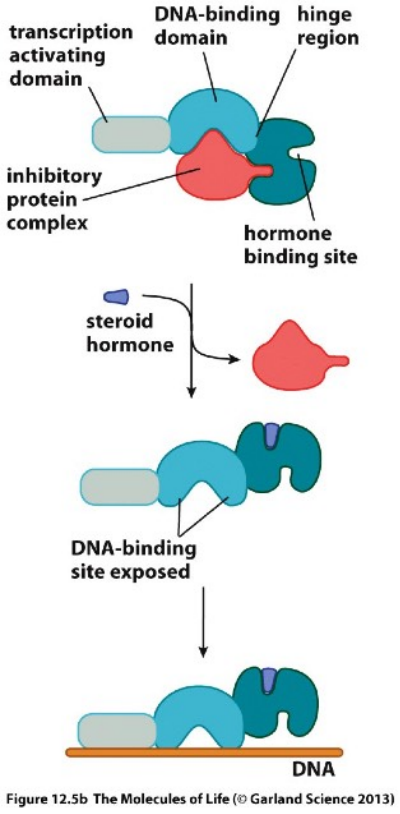
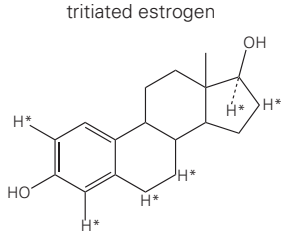
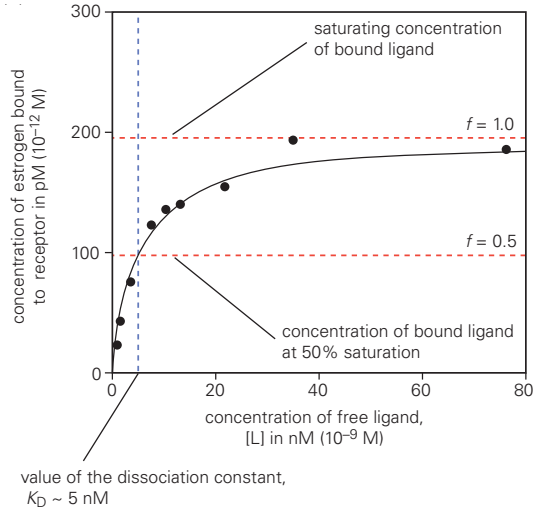
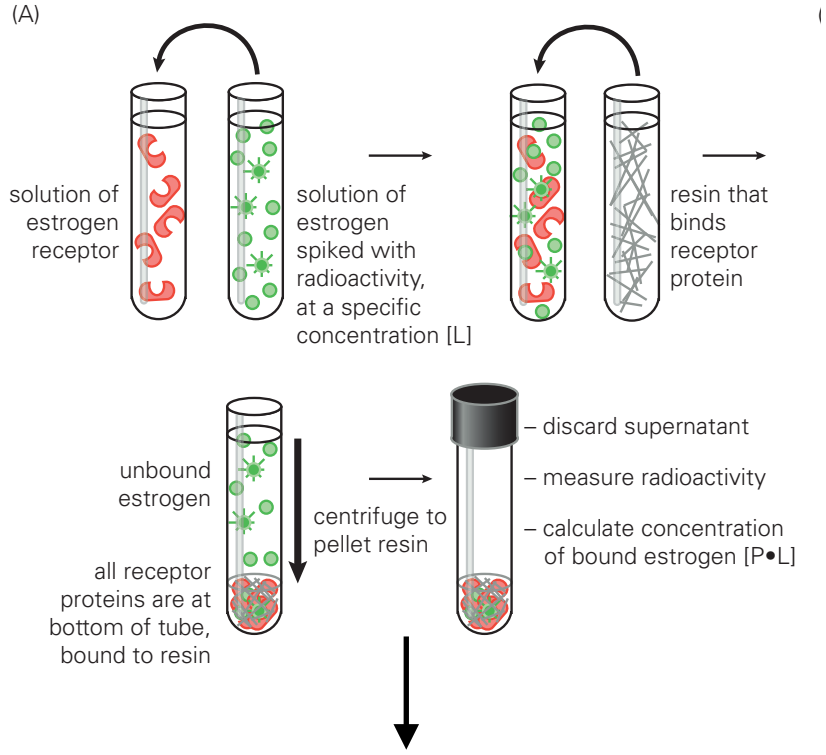


Figure 12.5b The Molecules of Life (© Garland Science 2013)

we use a radioactively labelled ligand

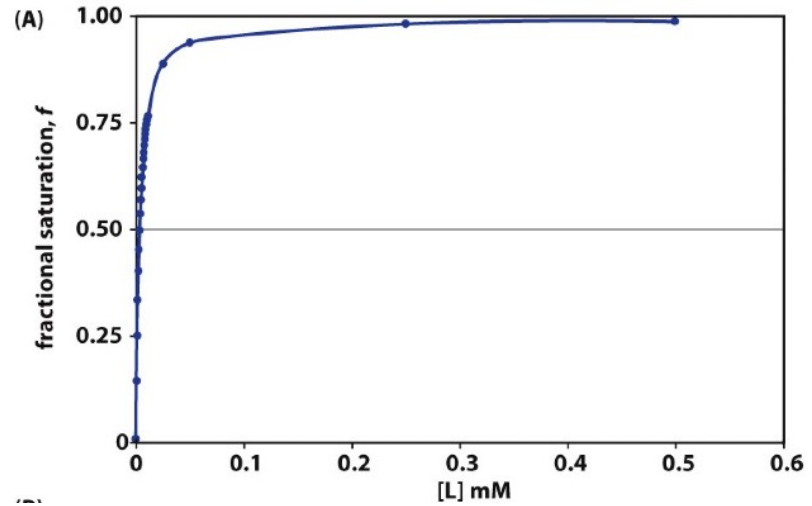


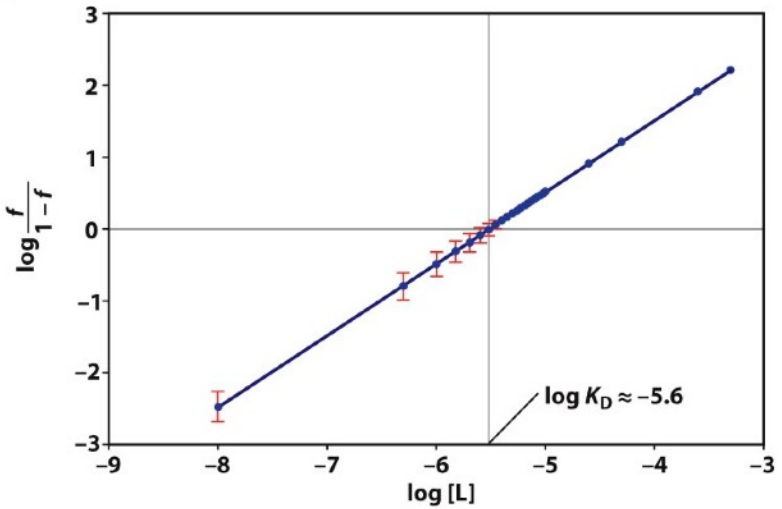




[P●L] is then plotted for various concentration of L and curve fitted to calculate K_D

Binding curves shown in different ways





$$\frac{\text{fraction bound}}{\text{fraction unbound}} = \frac{f}{1-f} = \frac{[\mathrm{L}]}{[\mathrm{L}]+K_\mathrm{D}} \frac{[\mathrm{L}]+K_\mathrm{D}}{K_\mathrm{D}} = \frac{[\mathrm{L}]}{K_\mathrm{D}}$$

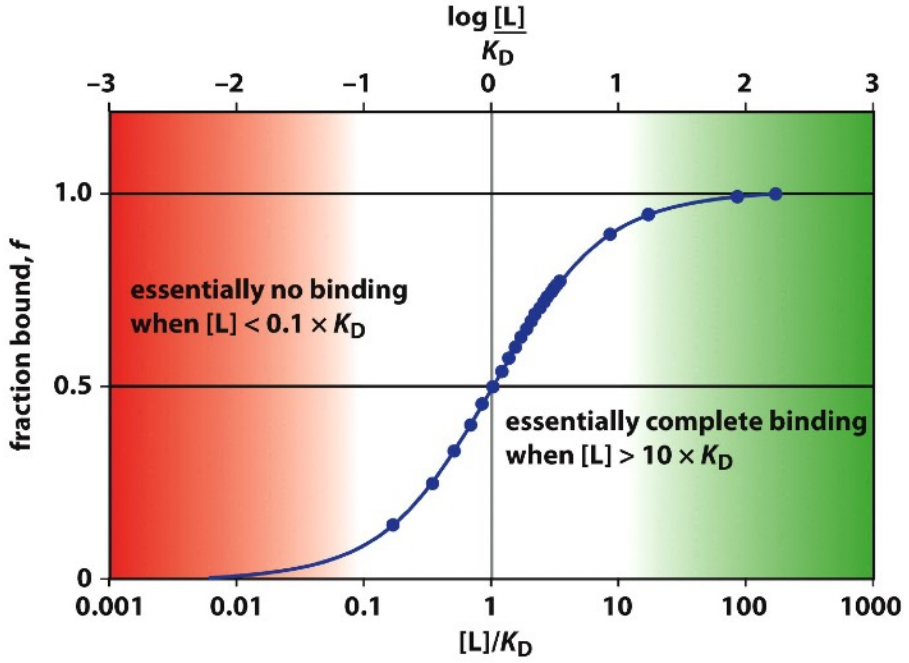
$$\log\left(\frac{f}{1-f}\right) = \log\left(\frac{[\mathrm{L}]}{K_\mathrm{D}}\right) = \log[\mathrm{L}] - \log K_\mathrm{D}$$

When $\left(\frac{f}{1-f}\right) = 1$ then the log is 0 and the intercept of the line on the horizontal axis is equal to the log of K_D

Note: we usually assume that the amount of bound ligand is very small compared to the total amount of ligand available, thus we usually use the free ligand concentration and the total ligand concentration interchangeably

if $[\mathrm{L}]_{\text{bound}} << [\mathrm{L}]_{\text{total}}$ then $[\mathrm{L}]_{\text{total}} = [\mathrm{L}] + [\mathrm{L}]_{\text{bound}} \approx [\mathrm{L}]$

Universal Binding Isotherm



Saturable binding is the hallmark of specific binding interactions, ie a simple binding mode with **one protein associated with one ligand, one binding pocket at the time**

The K_D defines the ligand concentration range over which the protein switches from unbound to bound - using the universal binding curve is a handy way to characterise binding because the **ligand concentration is expressed in terms of dissociation constant** (ie it is a universal curve).

Universal Binding Isotherm

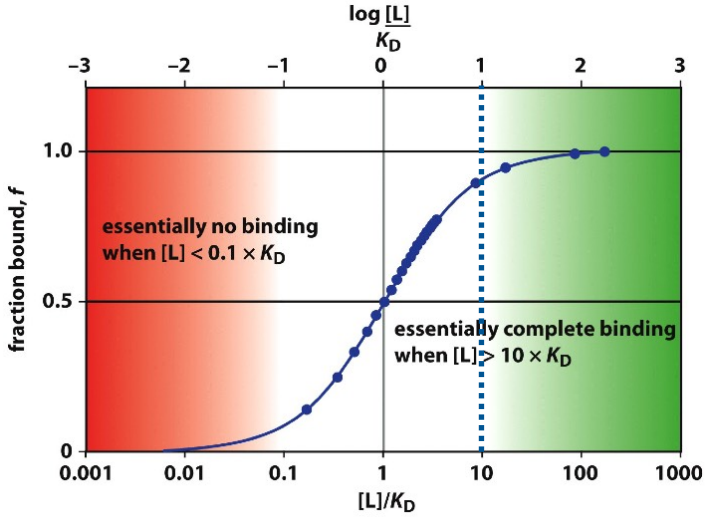


Figure 12.11 The Molecules of Life (© Garland Science 2013)

A common question in biochemistry and pharmacology is how much of a protein is bound to a ligand.

For a concentration where the free ligand is 10 times the value of K_D , the target is almost completely occupied (at 91%), in fact:

$$f = \frac{[\mathrm{L}]}{[\mathrm{L}] + K_\mathrm{D}} = \frac{10K_\mathrm{D}}{10K_\mathrm{D} + K_\mathrm{D}} = \frac{10}{11} = 0.91$$

Universal Binding Isotherm

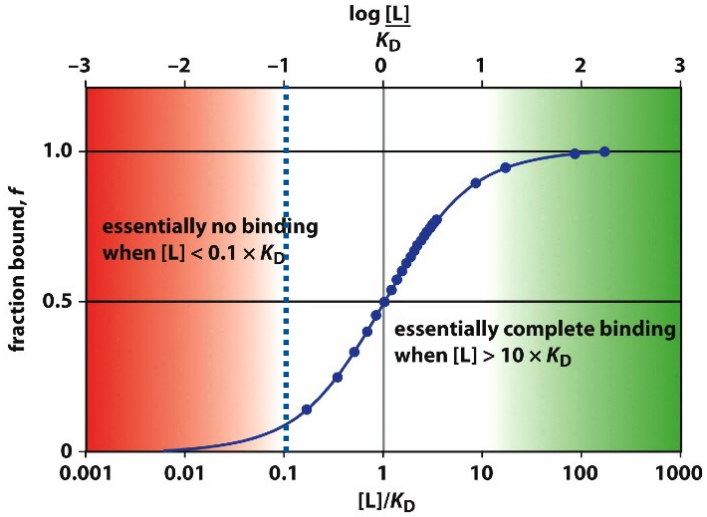


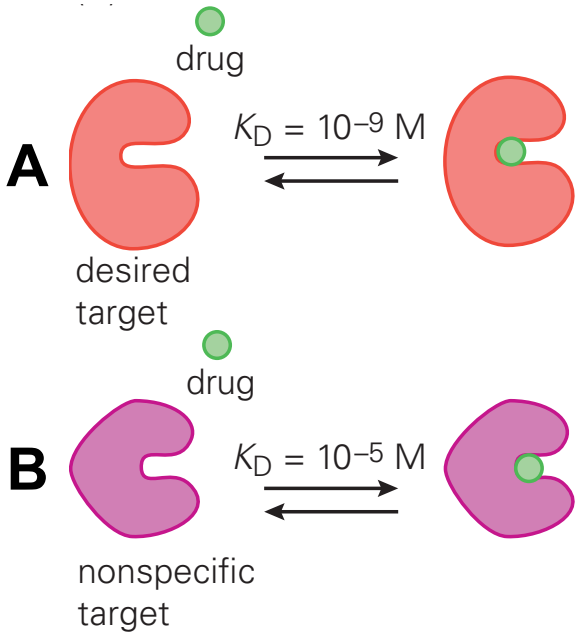
Figure 12.11 The Molecules of Life (© Garland Science 2013)

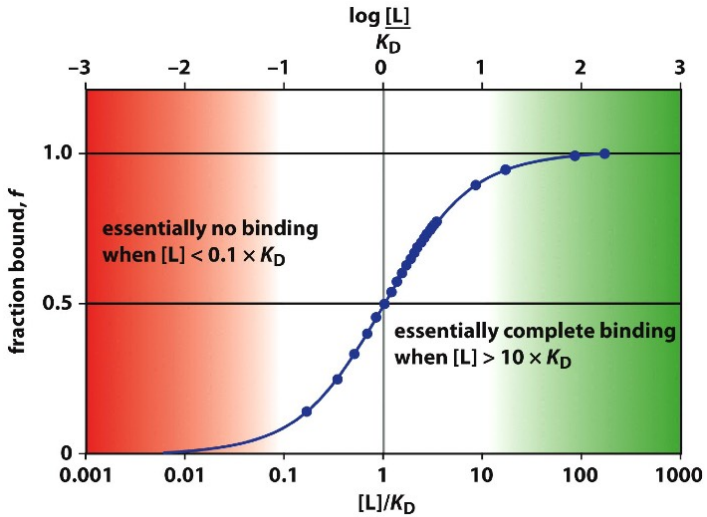
A common question in biochemistry and pharmacology is how much of a protein is bound to a ligand

For a concentration where the free ligand is 0.1 times the value of K_D, the target is occupied only at 9%

$$f = \frac{0.1K_D}{K_D + 0.1K_D} = \frac{0.1}{1.1} = 9\%$$

Thinking again about specificity !





Let's make use of the universal binding curve

If you have this situation in which you have to develop a drug for a given target A, which should also be selective, ie it does not bind a second target B.

What would the optimal K_Ds be in relative terms to achieve this goal?

If the drug is delivered at a 100 nM concentration, can you achieve a proper selectivity?

Energetics of drug binding

Very similar principles to the folding reaction

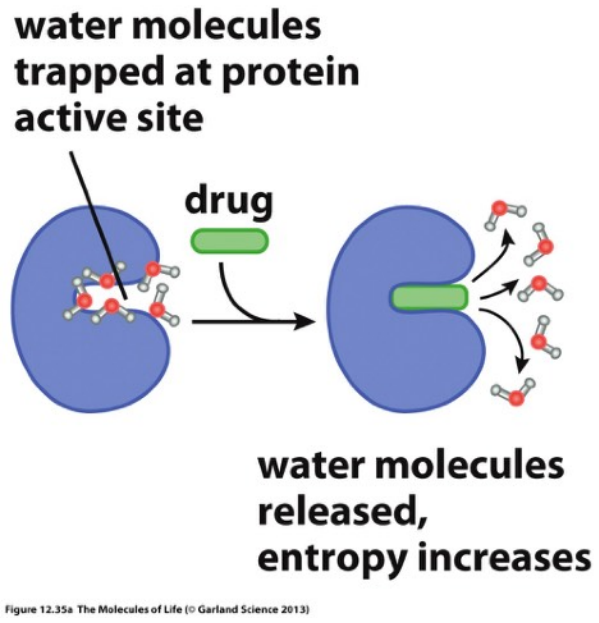


Figure 12.35a The Molecules of Life (© Garland Science 2013)

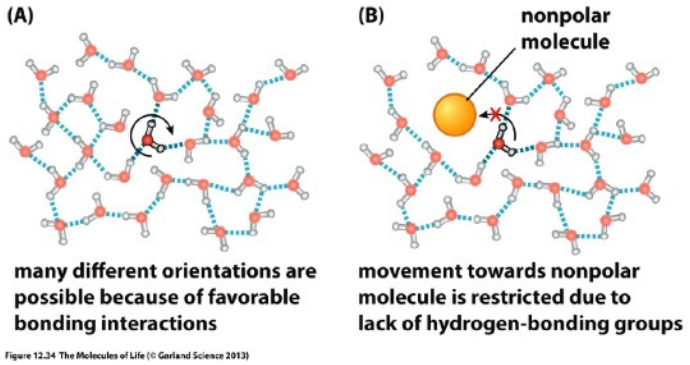
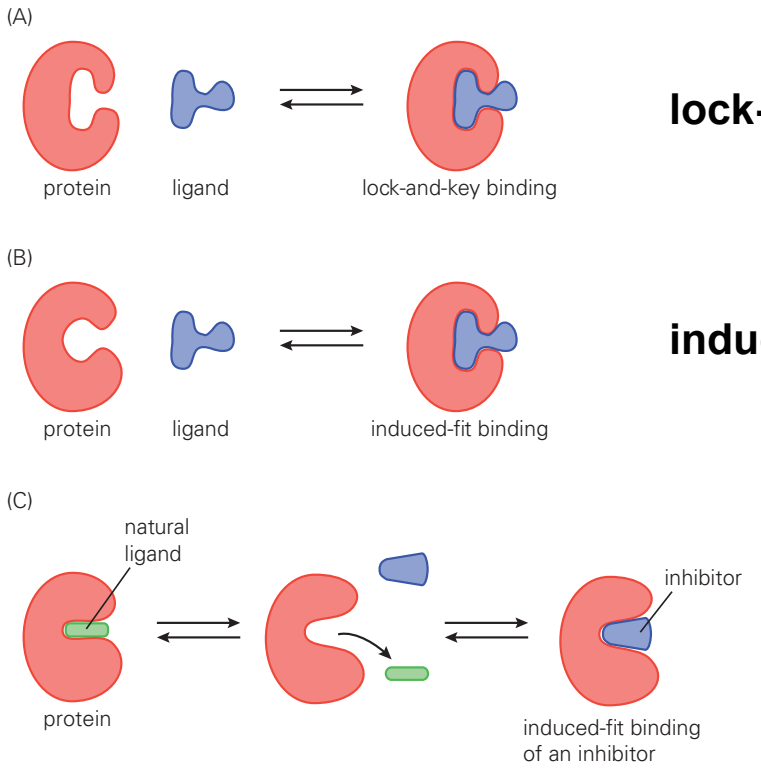


Figure 12.34 The Molecules of Life (© Garland Science 2013)

- Enthalpic gains on the interactions of the protein with the ligand come **mostly from H-bonds** - this contribution is not very large
- Entropic gain in free energy due to the solvent (ie **hydrophobic effect**) is instead quite significant
- As a consequence an optimal drug should **maximise hydrophobic features**
- In practice, drugs cannot be only hydrophobic as they will be **insoluble**, complicating adsorption and administration
- Also hydrophobic interactions tend to be not specific, while polar are more specific
- Thus in the development of a drug one needs to find the **best compromise** between polar and hydrophobic interactions

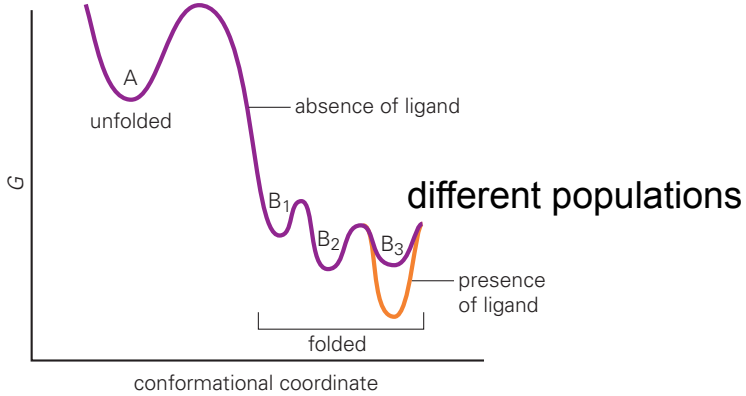
Conformational changes and binding

- first it was thought that binding happened via a **lock-and-key** mechanism, in a pretty rigid way that did not consider the conformational changes of the ligand and the target
- thus the concept of **induced fit** was developed to take this into account - considering the plasticity of the system, **protein are very flexible and dynamic**
- induced fit implies that the conformation change is induced by the ligand, but the different binding states could exist also without the ligand, and when the ligand is present the preferred conformation its selected - this is called **conformational selection** mechanism

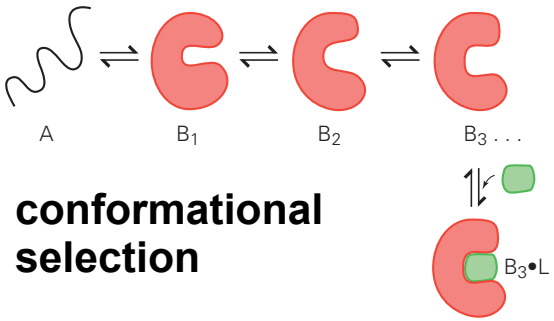


lock-and-key (outdated)

induced fit



different populations



conformational selection

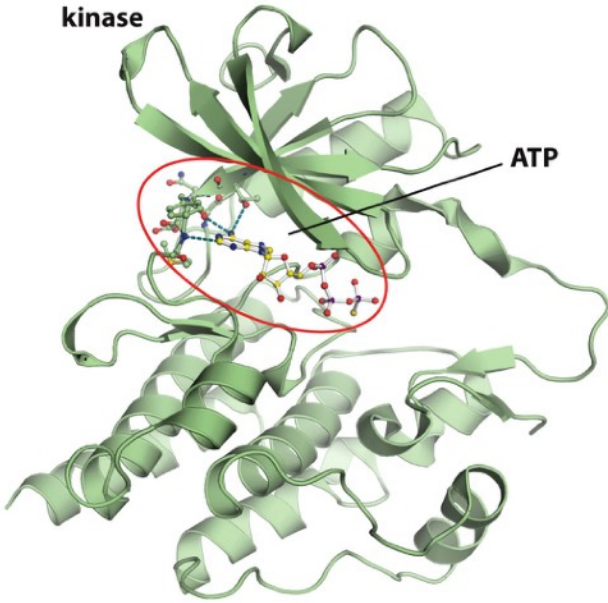
Drug Binding by Proteins

One of the main criteria that drives **drug development (DD)** processes is the binding affinity of the drug to a protein

Generally it starts that have low affinity (ie **lead** compounds), which will then improved by the **lead optimisation** process

Let's look at the example of kinases – key enzymes in signalling and involved in a number of diseases such as cancer

Kinases natively bind ATP and transfer one phosphate group to Ser, Thr or Tyr of protein substrates



Drug Binding by Proteins

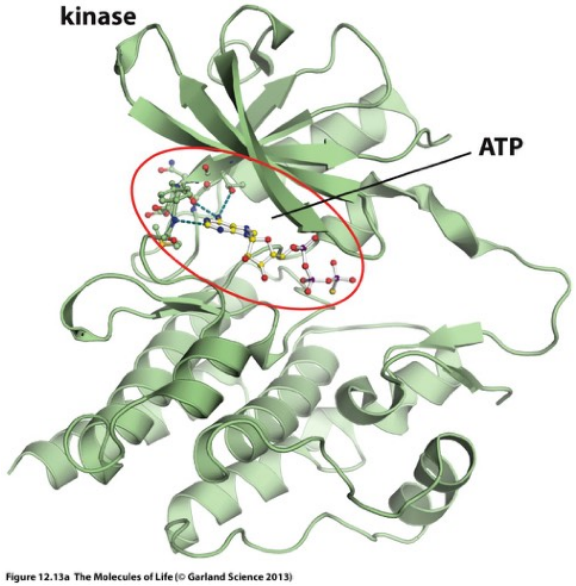


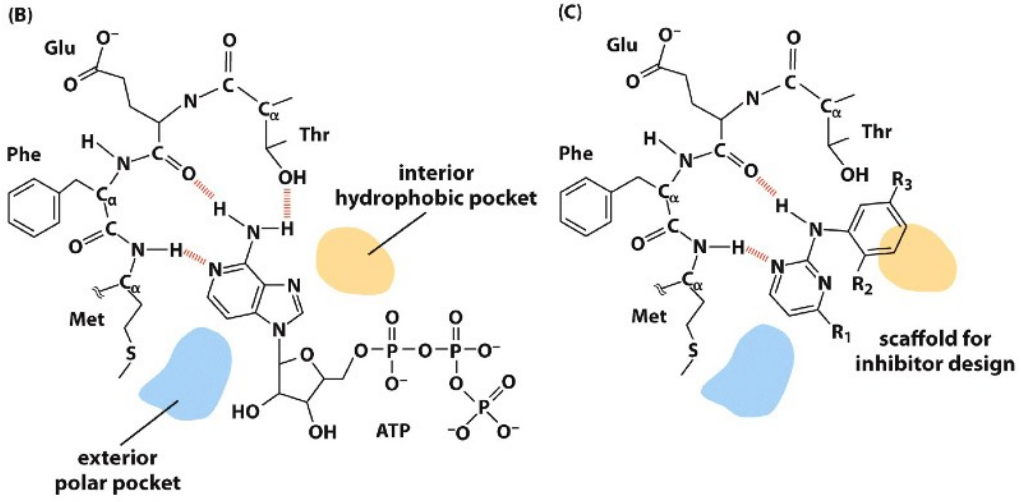
Figure 12.13a The Molecules of Life (© Garland Science 2013)

Structural information does help DD

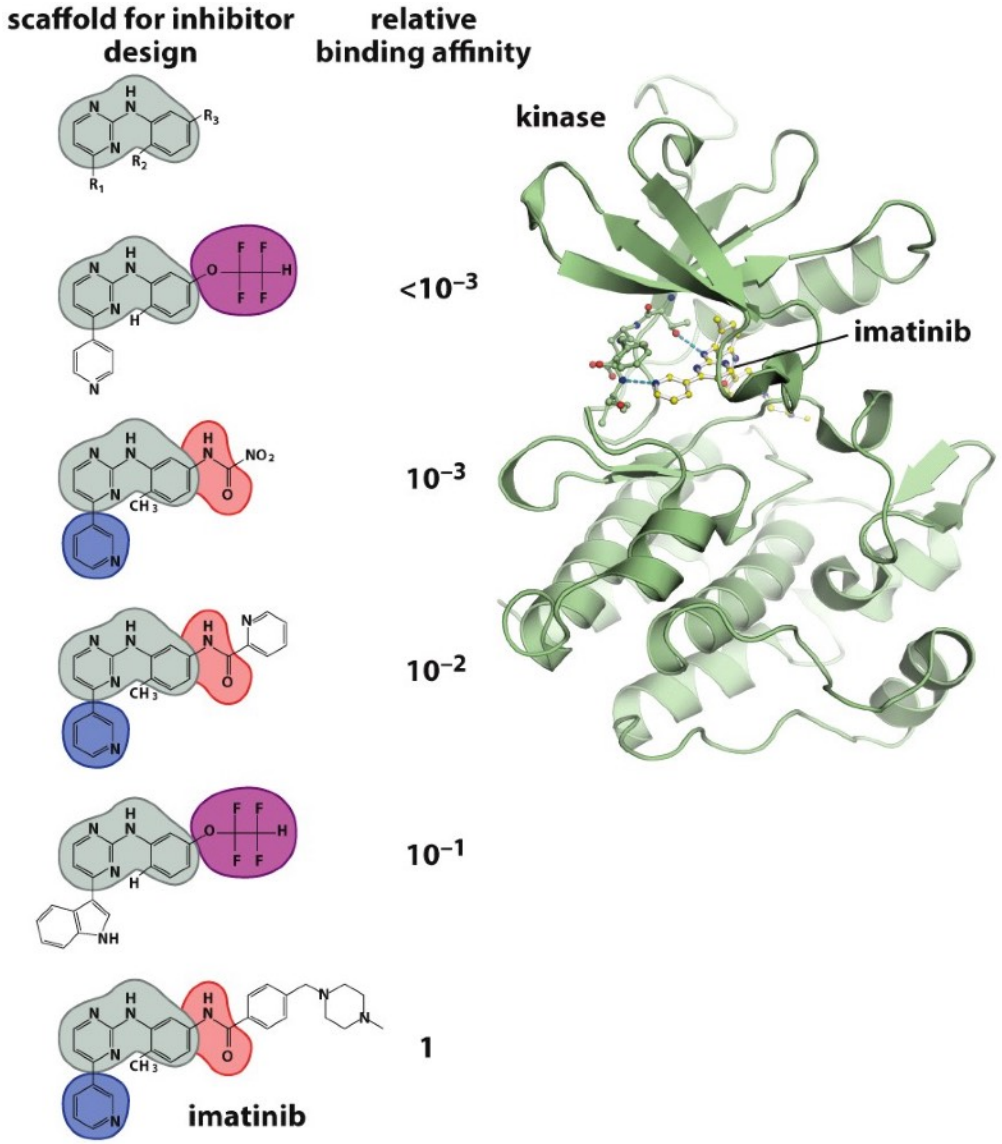
If we look at the active site interactions they can help us to rationally design drug compounds

For example the malfunctioning of the tyrosine-protein kinase Abl causes chronic myelogenous leukemia. Inhibitor of Abl known as imatinib (marketed as Gleevec) blocks the action of Abl and is an effective treatment for the leukemia.

How and where to start with the DD process?



The Drug Development Process



Lead optimization

We can see in this example how different changes in the structure of the small molecule change binding affinities, to reach up to ~10 nM

The Drug Development Process

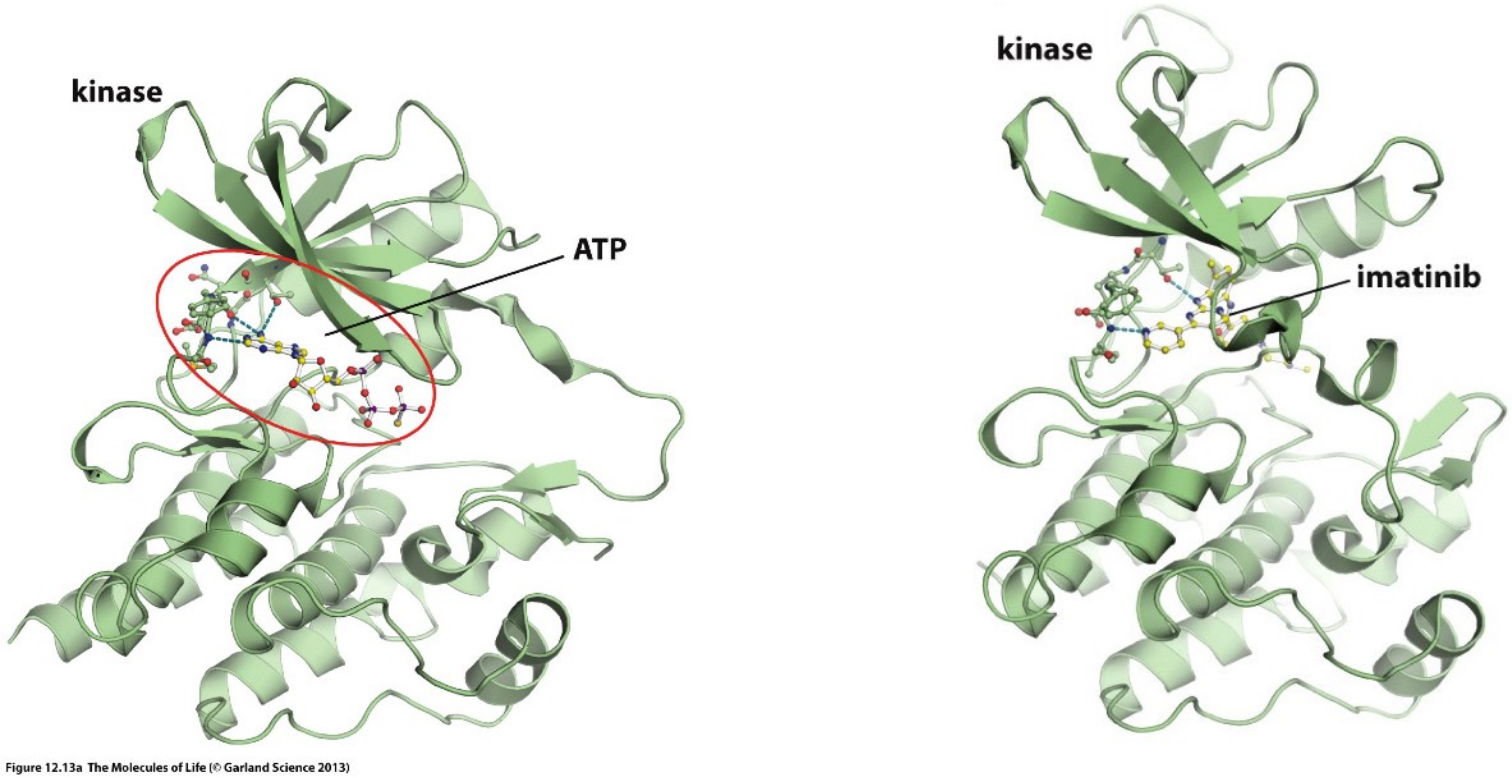


Figure 12.13a The Molecules of Life (© Garland Science 2013)

Imatinib is a **competitive inhibitor** - its mode of function is based on displacing a naturally occurring ligand

A common strategy for most small molecule drugs

Competitive Inhibitors

The affinity of a competitive inhibitor for a protein is reduced by the presence of the natural ligand.
An example with kinases and competitive binding :

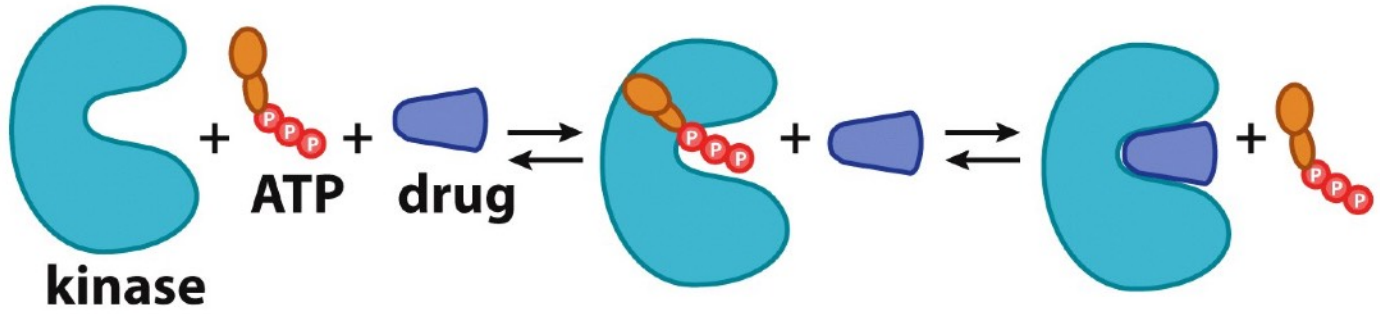


Figure 12.32a The Molecules of Life (© Garland Science 2013)

Often inhibitor activities are expressed in IC_{50}s

IC_{50} is the concentration of the inhibitor that reduce the activity of a protein to the half maximal value

So rather than a strict measurement of affinity one actually obtains a measure of the impact in the activity of the protein

Competitive Inhibitors

How do we relate IC_{50}s value with K_Ds?

$$K_I = (IC_{50})\left(\frac{K_D}{K_D + [L]}\right)$$

Competitor
(e.g. ATP)

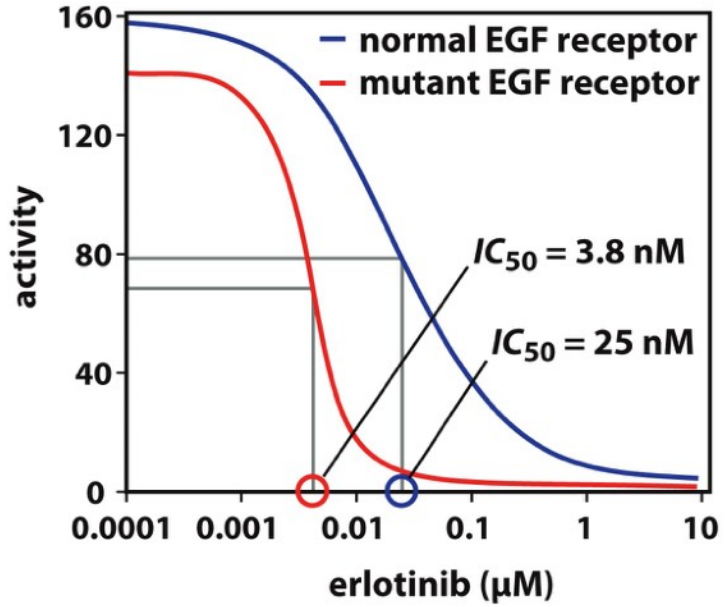


Figure 12.32b The Molecules of Life (© Garland Science 2013)

A few important points considering for the example of a kinase :

- If [ATP] (ie [L]) = 0 then $K_I = IC_{50}$
- To measure concentration of ATP and K_D is required (usually around ~1 mM and 10 μM)
- K_I needs to take into account the K_D and the concentration of the competitor

The full development of the equation is presented in the Molecules of Life

Drug Development

Imatinib is a competitive inhibitor - its mode of function is based on displacing a naturally occurring ligand

A common strategy for most small molecule drugs

Another example:

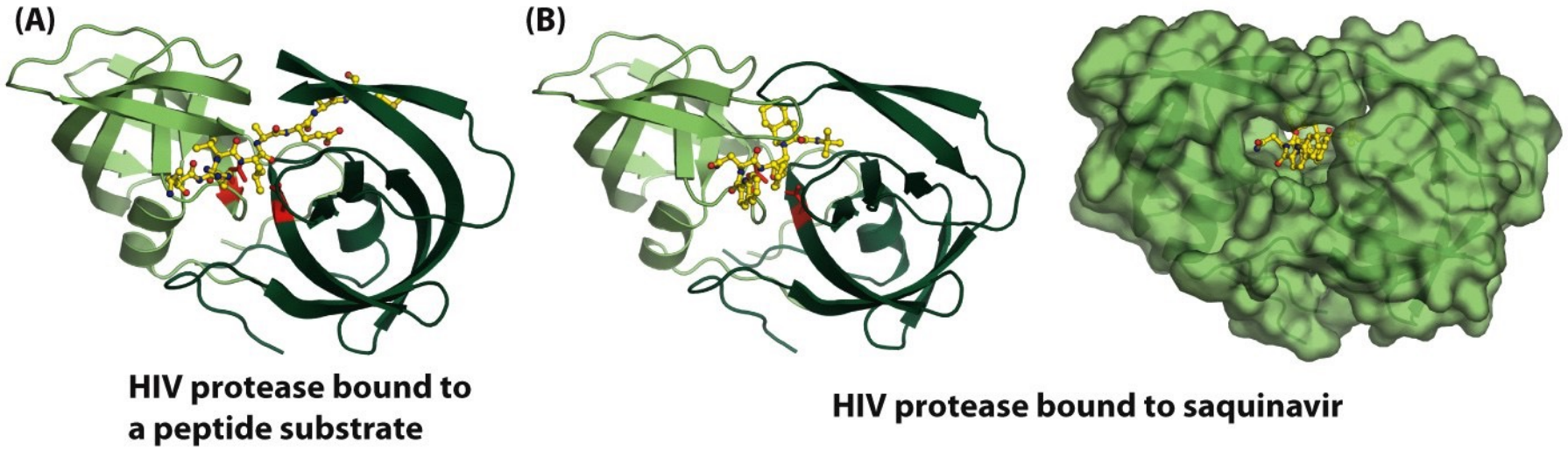
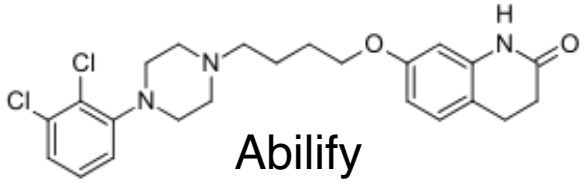
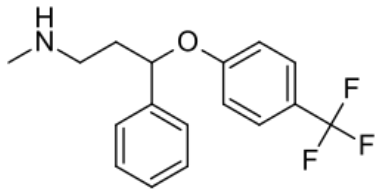


Figure 12.19 The Molecules of Life (© Garland Science 2013)

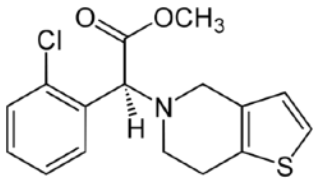
Small molecules drugs



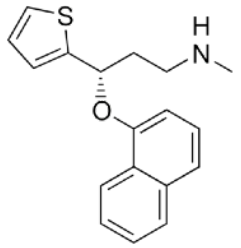
Abilify
(antipsychotic)



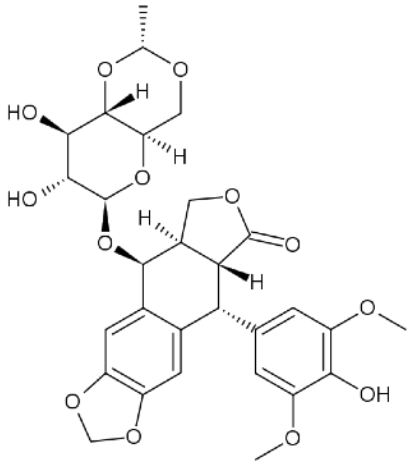
Prozac
(antidepressant)



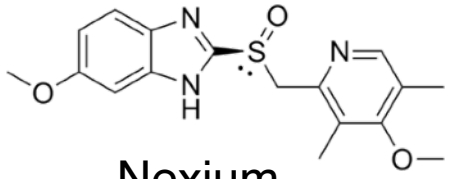
Plavix
(antiplatelet agent)



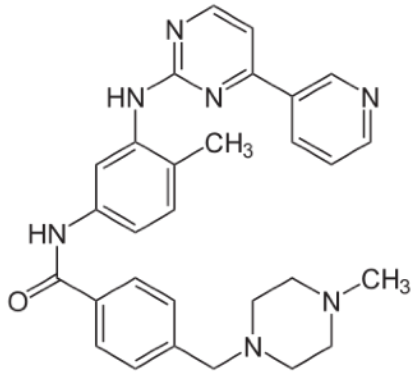
Cymbalta
(pain and anxiety)



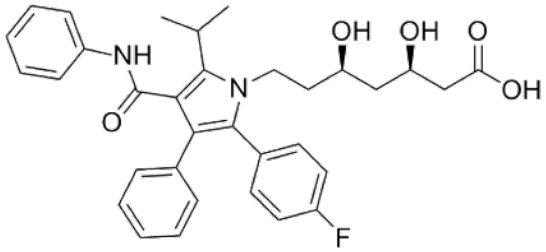
Etoposide
(cancer)



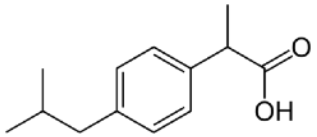
Nexium
(gastric acid)



Gleevec
(cancer)



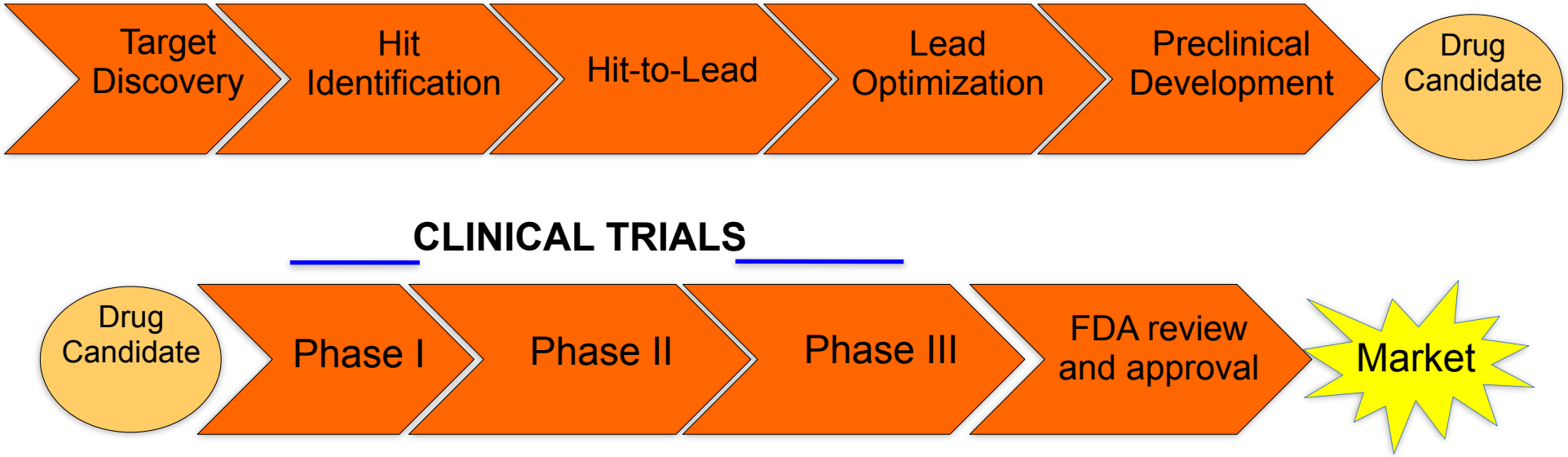
Lipitor
(Cholesterol)



Ibuprofen
(inflammation)

Drug Discovery & Development

Sequence of steps



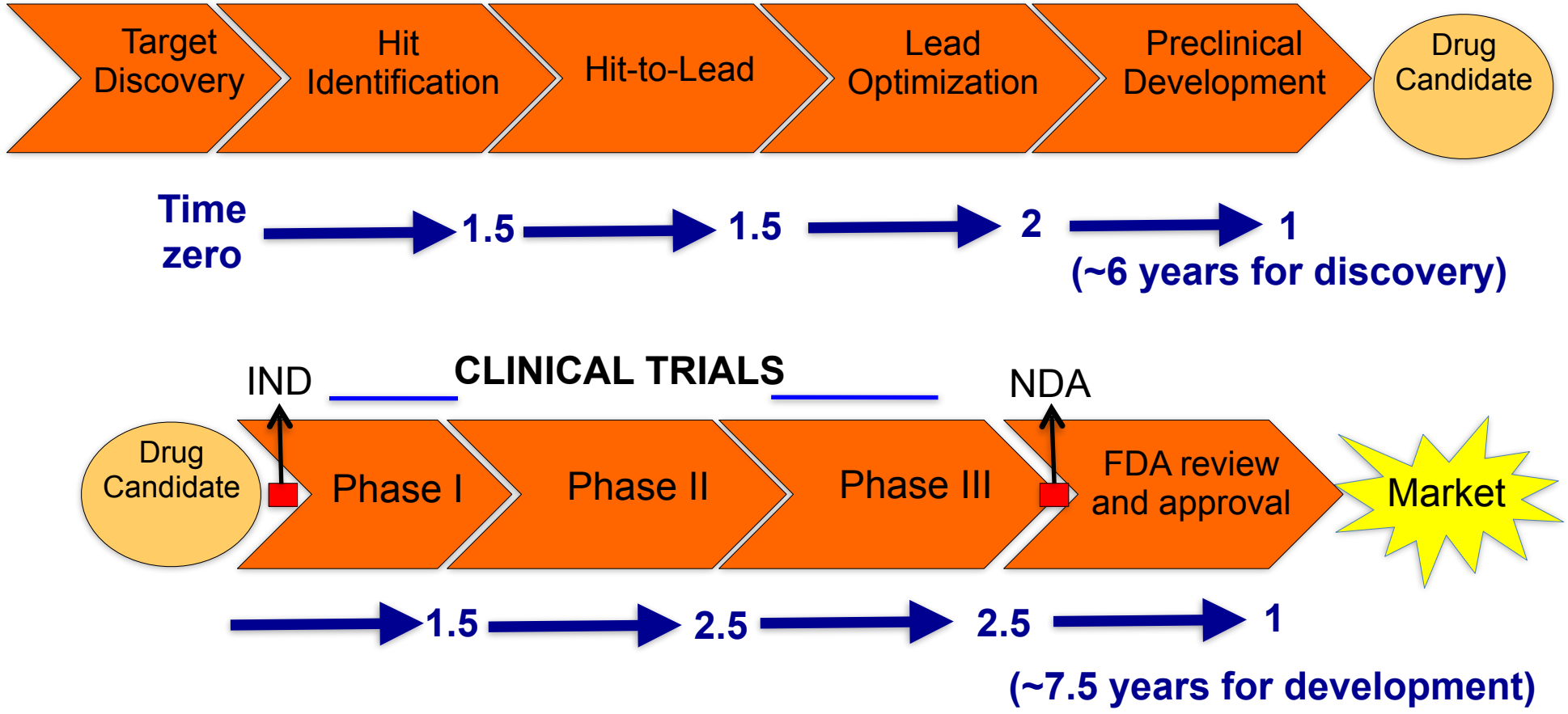
Best leads are optimized in terms of their drug-likeness:
Improved affinity is coupled with a promising **pharmacokinetic (PK)** profile.

PK includes **ADMET** characterization**:**

Absorption (e.g., bioavailability, F)
Distribution (e.g., binding to serum proteins)
Metabolism (metabolites, e.g., cytochrome P450)
Excretion (kidneys system)
Toxicity (e.g., Affinity towards h-ERG)

Drug Discovery & Development

Average time requested for it

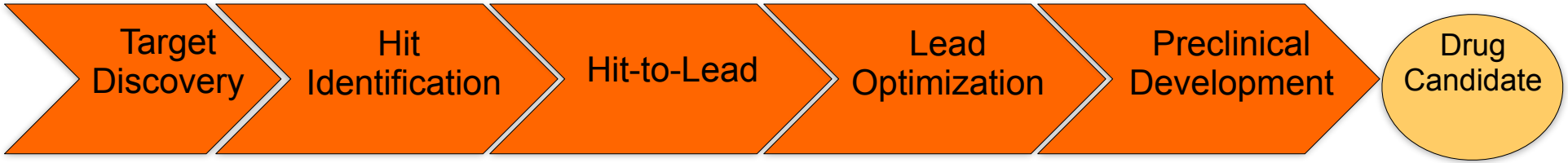


Total time (on average) = ~13.5 years

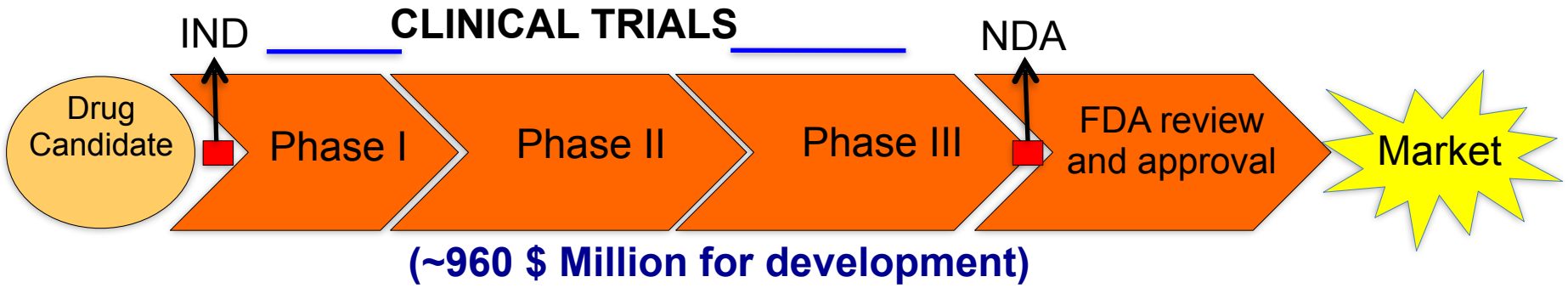
IND: Investigational New Drug application
NDA: Submission of New Drug Application

Drug Discovery & Development

Average cost requested for it



(~820 $ Million for discovery)



(~960 $ Million for development)

Total cost on average = ~1.78 $ Billion for one NME

Source: How to improve R&D productivity: the pharmaceutical industry's grand challenge
Steven M. Paul, *Nat. Review Drug Discovery* March Vol. 9 **2010**

NME: new molecular entity

Drug Discovery & Development

FDA approvals in 2020

53 New Drugs Approved

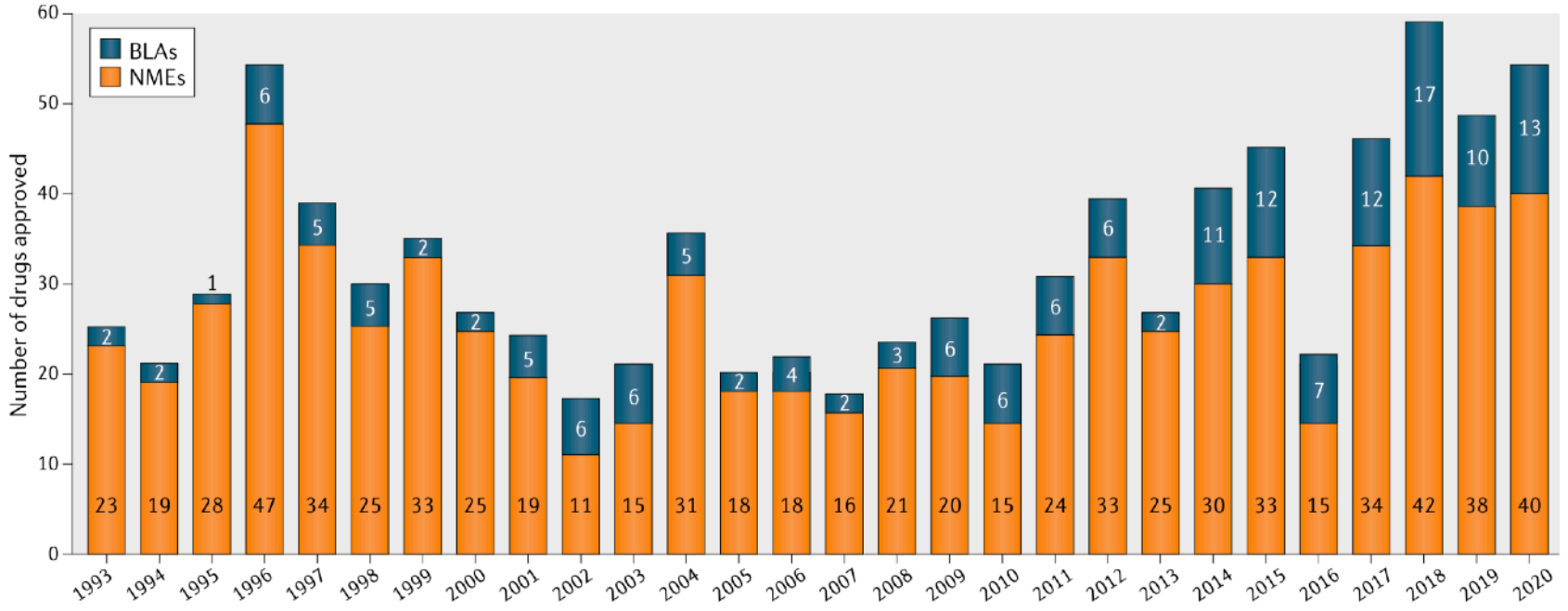


Fig. 1 | **Novel FDA approvals since 1993.** Annual numbers of new molecular entities (NMEs) and biologics license applications (BLAs) approved by the FDA's Center for Drug Evaluation and Research (CDER). See TABLE 1 for new approvals in 2020. Approvals by the Center for Biologics Evaluation and Research (CBER), for products such as vaccines and gene therapies, are not included in this drug count (see TABLE 2). Source: FDA.

Source:
Asher Mullard,
Nature Reviews Drug Discovery, February (**2021**)

Drug Discovery & Development

Approval by therapeutic area 2020

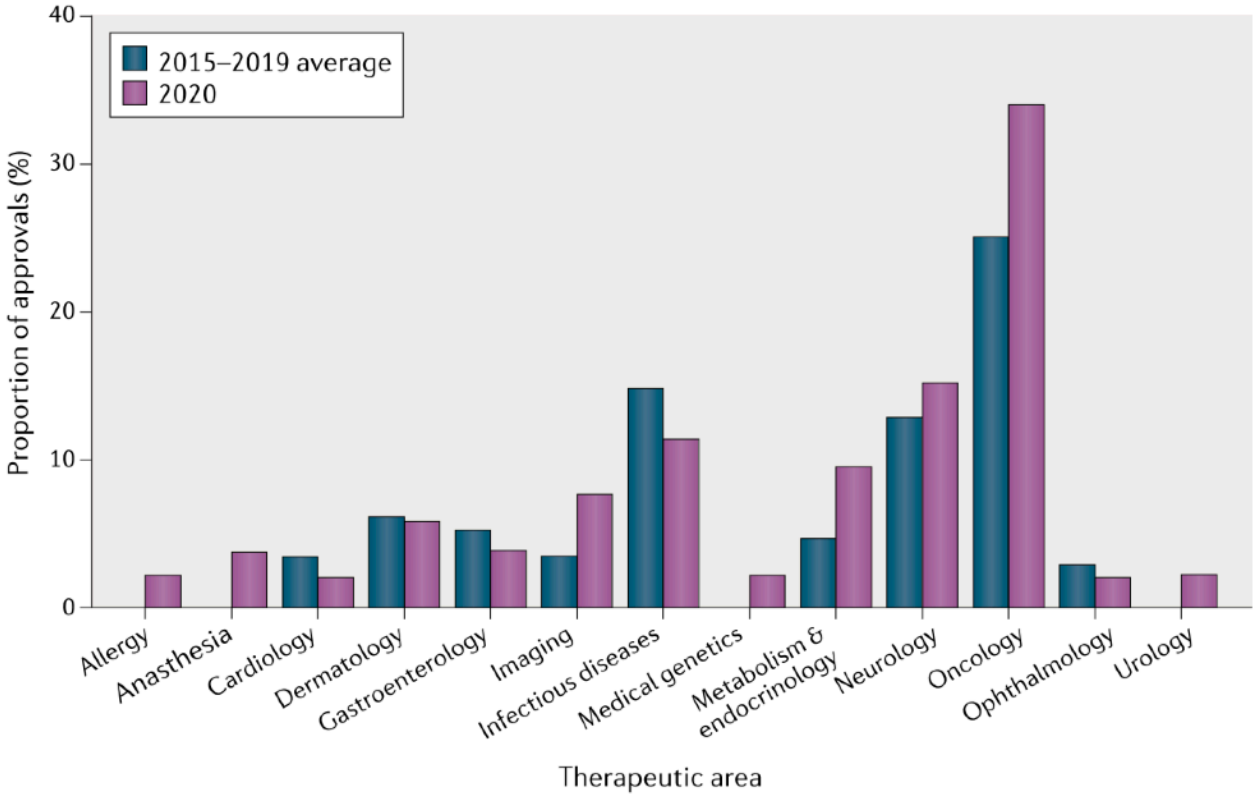


Fig. 2 | **CDER approvals by selected therapeutic areas.** Source: *Nature Reviews Drug Discovery*, FDA.

Source:
Asher Mullard,
Nature Reviews Drug Discovery, February (**2021**)

What to know ...

- The affinity of a protein for a ligand is characterized by the dissociation constant K_D
- The value of the K_D for a binding interaction is the ligand concentration at which half the receptors are bound to ligand
- The value of the K_D determines the concentration range of the ligand over which the receptor switches from unbound to bound
- The K_D for a physiological ligand is usually close to the natural concentration of the ligand, this is the result of evolution
- Key formulas :

$$K_D = \frac{[P][L]}{[P \bullet L]} = \frac{1}{K_A}$$

$$\Delta G^{o}_{bind} = +RT \ln K_D$$

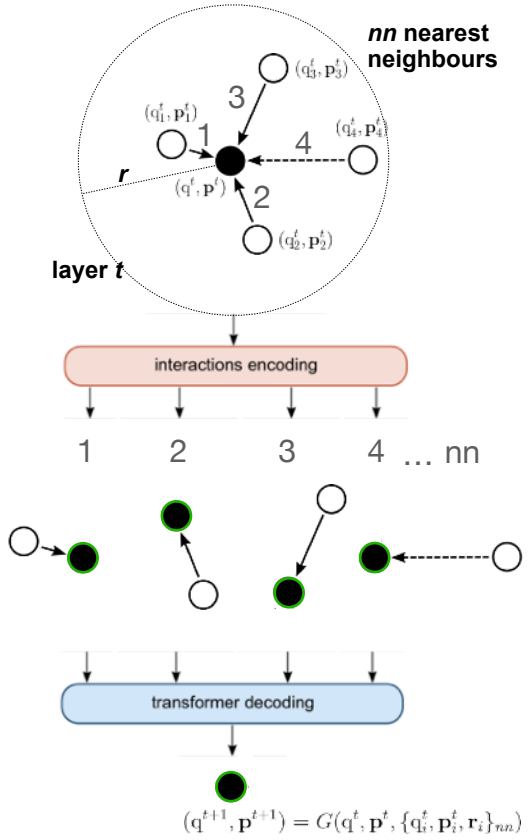
$$f = \frac{[L]}{[L] + K_D}$$

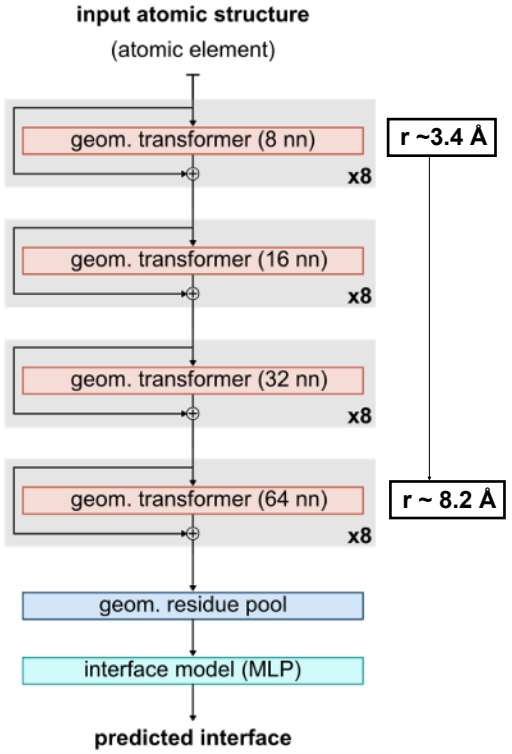
$$\log\left(\frac{f}{1-f}\right) = \log\left(\frac{[L]}{K_D}\right) = \log[L] - \log K_D$$

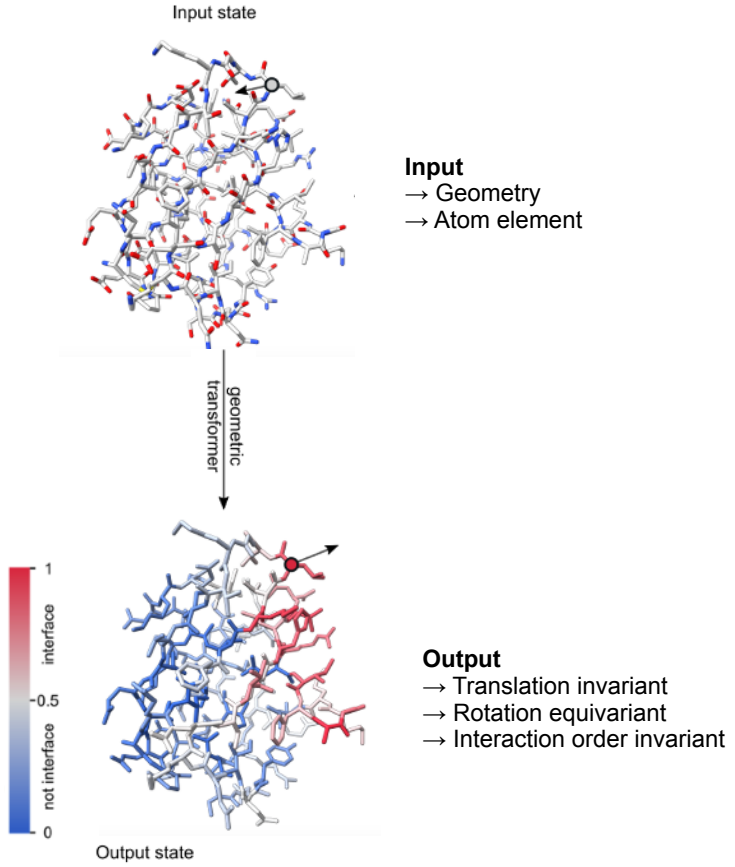
Bonus Slides

We will see that machine learning models are impacting biology in many different ways (Lecture 12)
here is an example developed in my lab for predicting molecular binding interfaces

PeSTo: Protein Structure Transformer







Input
→ Geometry
→ Atom element

Output
→ Translation invariant
→ Rotation equivariant
→ Interaction order invariant

Krapp et al. Nat Comms 2023

@ http://pesto.epfl.ch

PeSTo

PeSTo (Protein Structure Transformer) is a parameter-free geometric deep learning method to predict protein interaction interfaces from a protein structure. It is available for free without registration as an online tool. A manuscript of the method is in preparation and will be available soon.

Learn more about this project in this preprint at Biorxiv.

How to use

Copy-paste your atomic coordinates in PDB format, or upload a PDB file from your drive, or fetch a protein structure/model from:

- The protein data bank by typing a PDB ID. Example: 2CUA
- The AlphaFold-EBI database by typing a Uniprot ID. Example: P27695
- Upload your own PDB formated structure

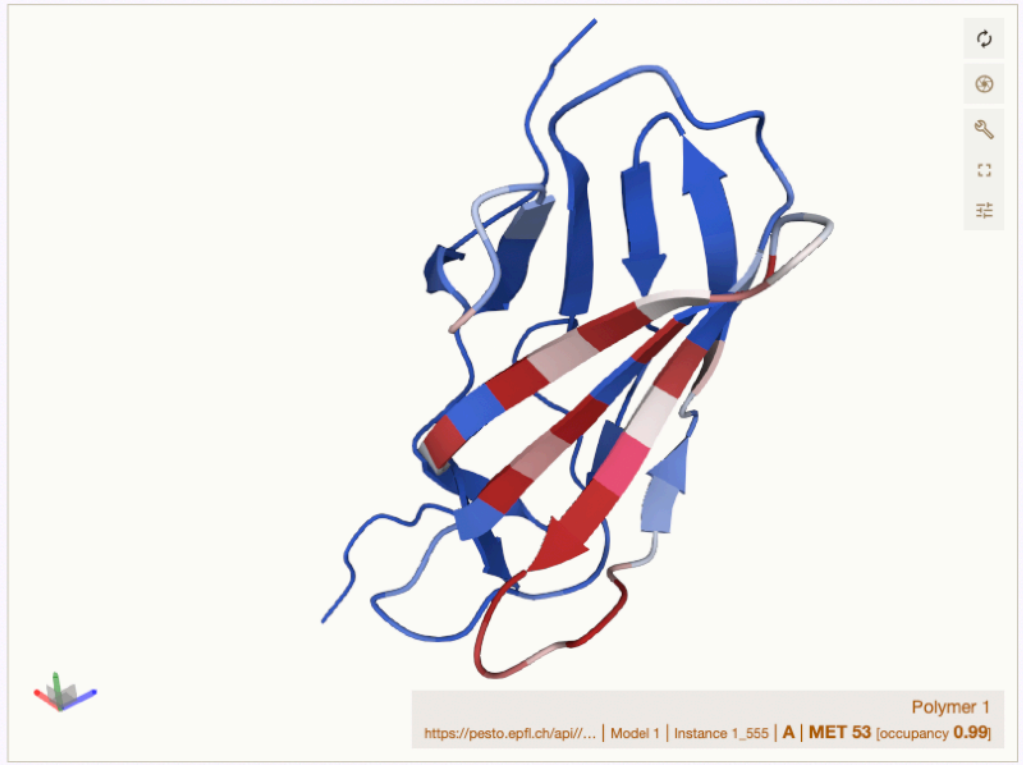
Then click "Detect chains", select one or more, and submit your job to run the prediction. Your results should be available in less than a minute. If an error occurs, the PDB file might be not correctly formated or the input structure is too big

2CUA | Fetch PDB/AF-EBI | Upload PDB

Copy-Paste molecule here

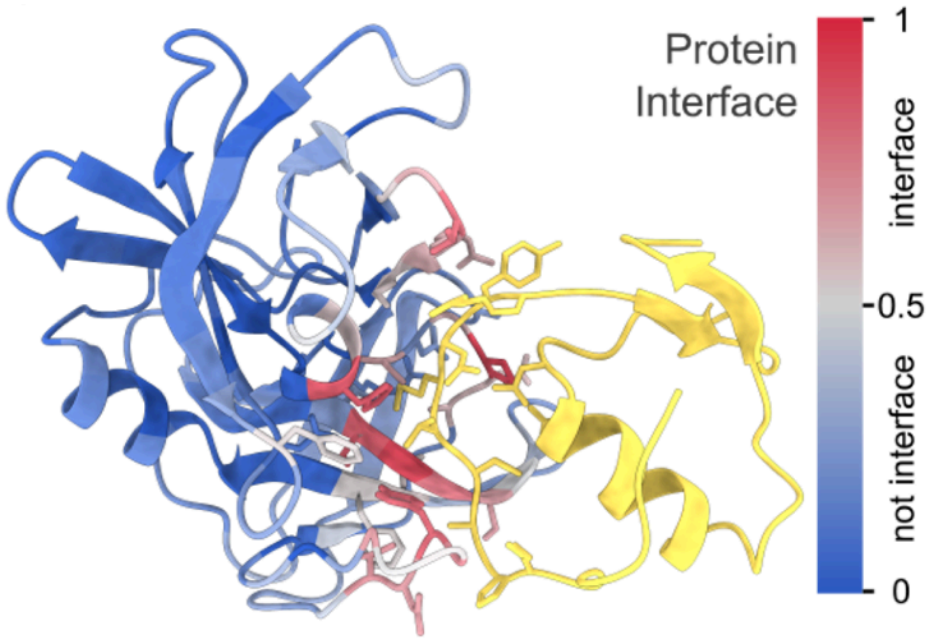
Detect chains

Protein | DNA-RNA | Lipid | Ligand | Ion

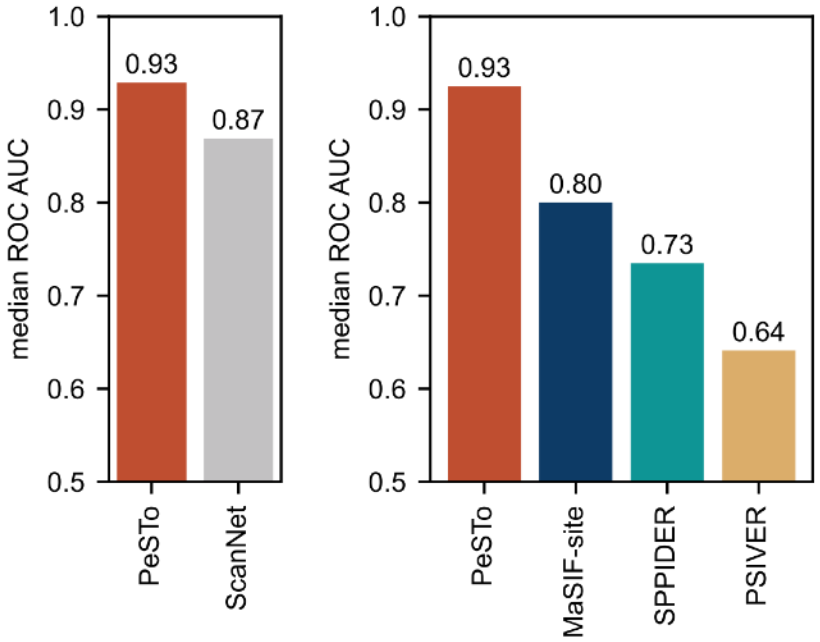


Chain	Res Name	Res ID	Prediction
A:0	ASP	13	0.64
A:0	THR	51	0.91
A:0	VAL	52	0.54
A:0	MET	53	0.99
A:0	ALA	54	0.96
A:0	GLY	55	0.97
A:0	ASN	56	0.98
A:0	ASP	57	0.91
A:0	GLU	58	0.81

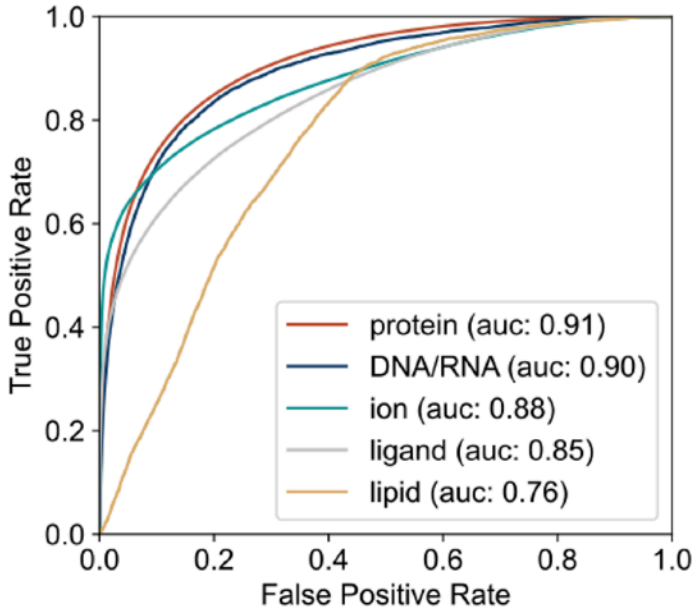
PeSTo for protein-protein interfaces prediction



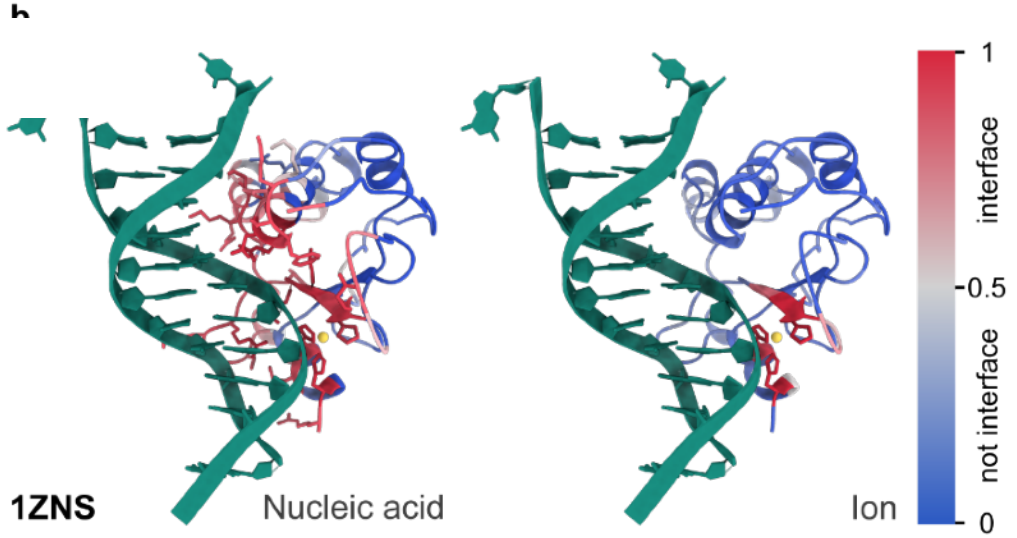
Streptogrisin B with ovomucoid - unbound conformation (0.93 Å RMSD) with a ROC AUC of 96%



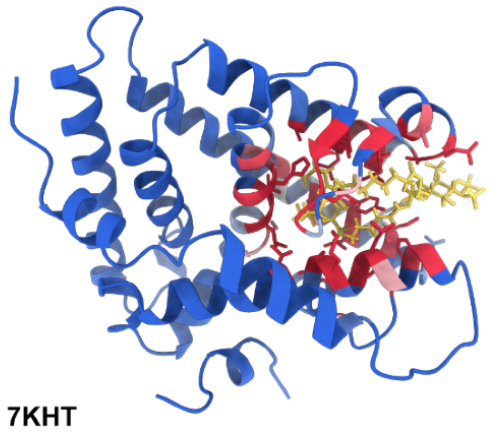
Extending prediction to other interacting interfaces



nucleic acid & ion interface



ColE7 endonuclease domain



Lipid interface

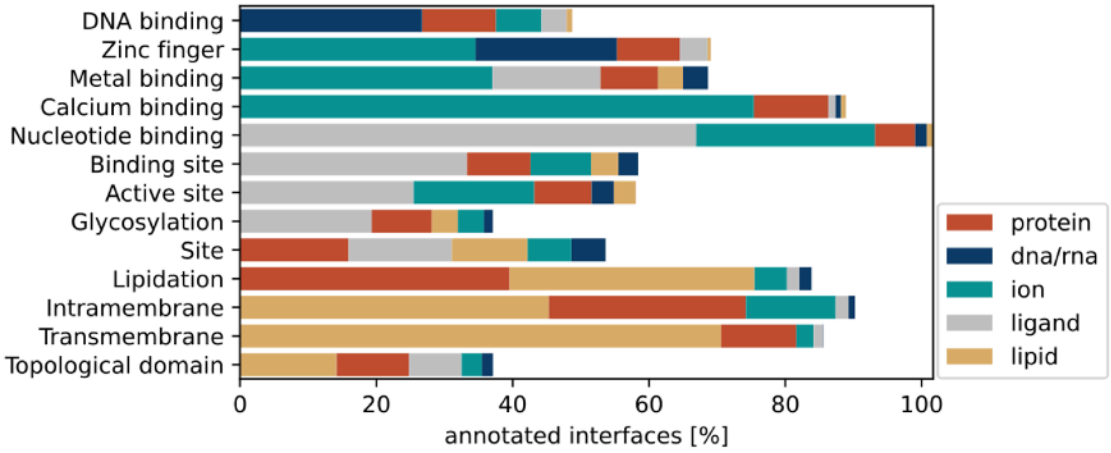
nuclear receptor Steroidogenic Factor-1 (SF-1)

Proteome-wide interface prediction - *interfaceome*

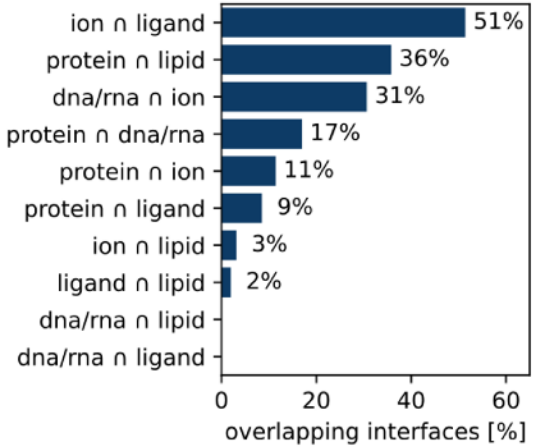
Human proteome prediction using the AF-EBI database

(only high-quality models, 7464 of 20504)

correlation with protein function and features



interface cross-talk



correlation with protein mutations