Using Higher Calculus to Study Biologically Important Molecules

Julie C. Mitchell

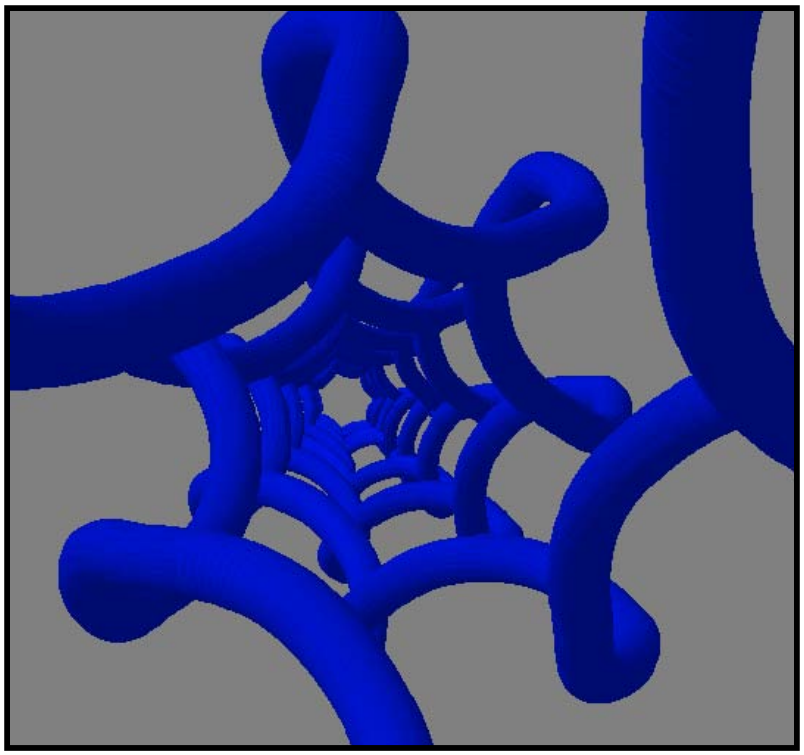
Mathematics and Biochemistry

University of Wisconsin - Madison

There Are Many Kinds Of Proteins

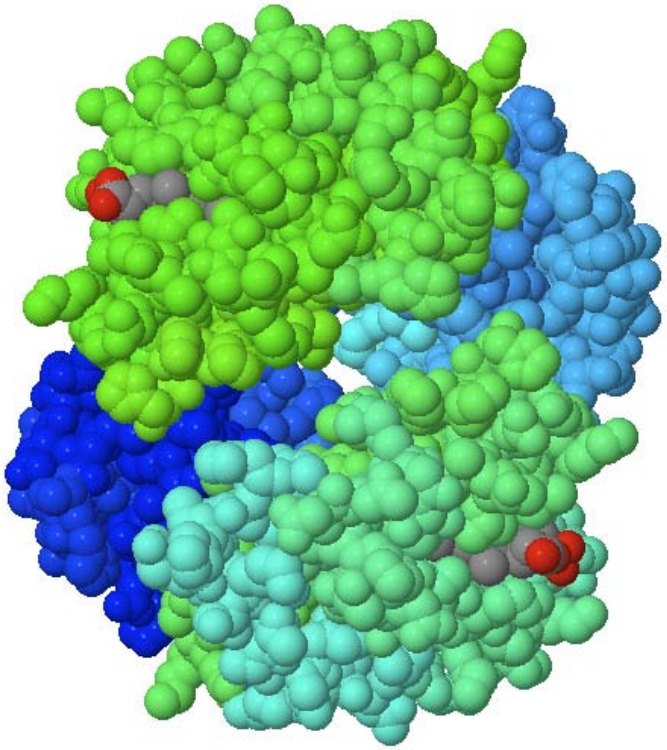
- The word **protein** comes from a Greek word that means “of primary importance”

- Many proteins are familiar, but you may not know they are proteins.
  - Collagen is a protein
  - Hemoglobin is a protein
  - Antibodies are proteins
  - Enzymes are proteins



Proteins And Disease

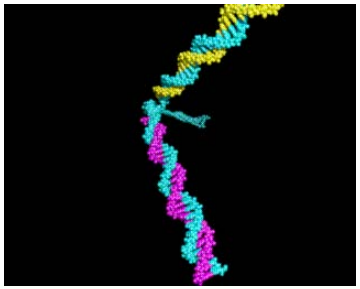
- Mutations and misfolding of proteins can cause disease
  - The cause of **Mad Cow Disease** is a protein that has misfolded
  - The cause of **Sickle Cell Anemia** is a hemoglobin molecule with an unfortunate mutation.



From Genes To Proteins

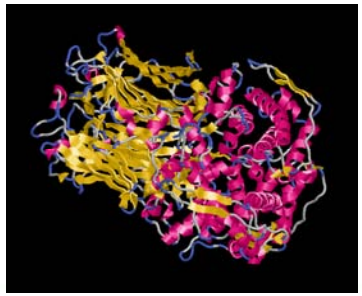
Name	3 Letter Abbreviation	1 Letter Abbreviation	DNA/RNA Codons
Alanine	Ala	A	GCA, GCC, GCG,GCU
Cysteine	Cys	C	UGC, UGU
Aspartic Acid	Asp	D	GAC, GAG
Glutamic Acid	Glu	E	GAA, GAG
Phenylalinine	Phe	F	UUC, UUU
Glycine	Gly	G	GGA, GGC, GGG, GGU
Histidine	His	H	CAC, CAU
Isoleucine	Ile	I	AUA, AUC, AUU
Lysine	Lys	K	AAA, AAG
Leucine	Leu	L	UUA, UUG, CUA, CUC, CUG, CUU
Methionine	Met	M	AUG
Asparagine	Asn	N	AAC, AAU
Proline	Pro	P	CCA, CCC, CCG, CCU
Glutamine	Gln	Q	CAA, CAG
Arginine	Arg	R	CGA, CGC, CGG, CGU
Serine	Ser	S	UCA, UCC, UCG, UCU, AGC, AGU
Threonine	Thr	T	ACA, ACC, ACG, ACU
Valine	Val	V	GUA, GUC, GUG, GUU
Tryptophan	Trp	W	UGG
Tyrosine	Tyr	Y	UAC, UAU
Stop		.	UAA, UAG, UGA

DNA unravels and codes RNA

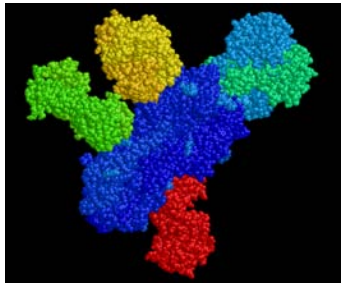


RNA codes proteins -- **the code was dicovered at UW-Madison by Khorana!!!**

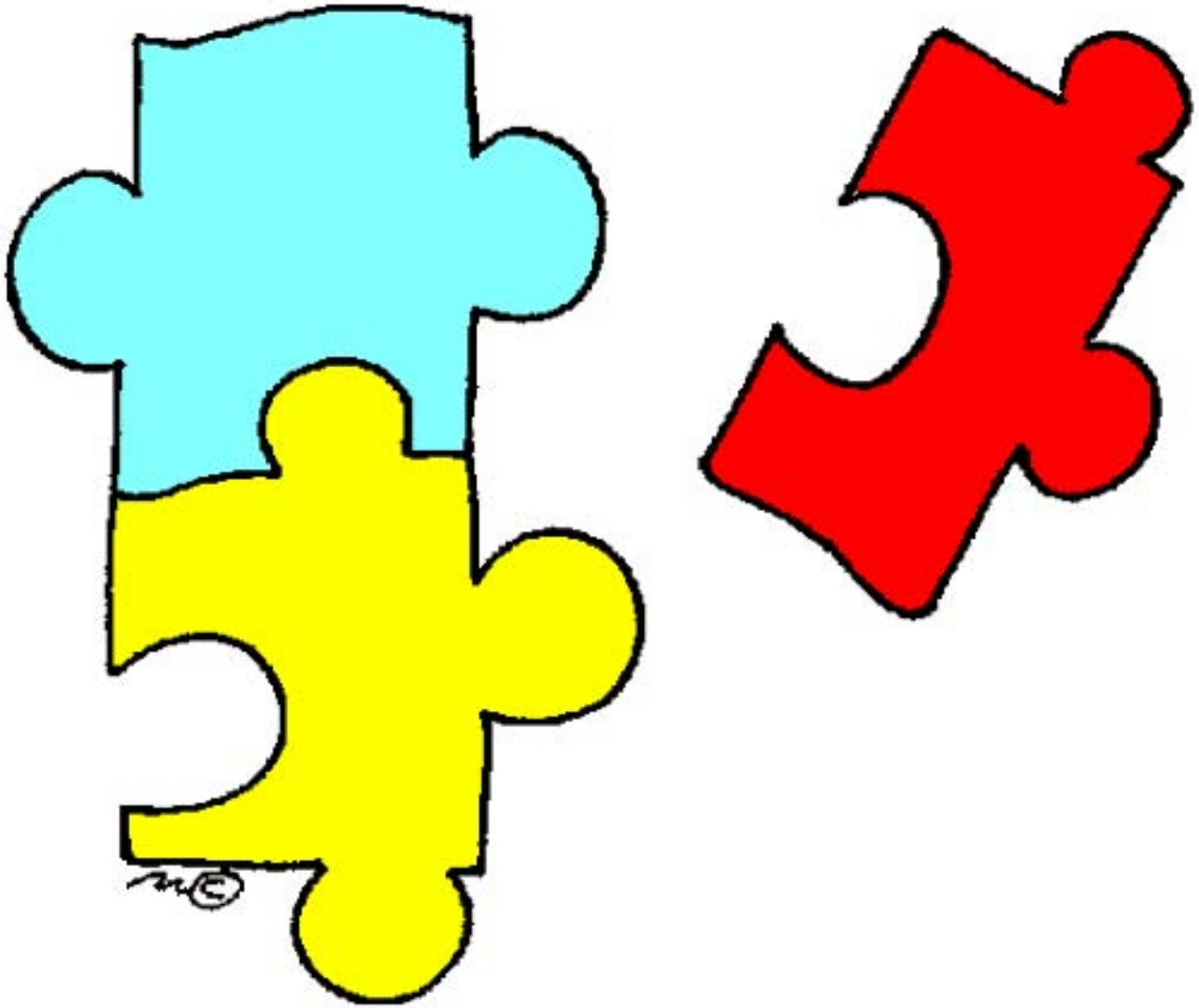
A protein folds into a globular structure



Proteins interact to perform biological functions

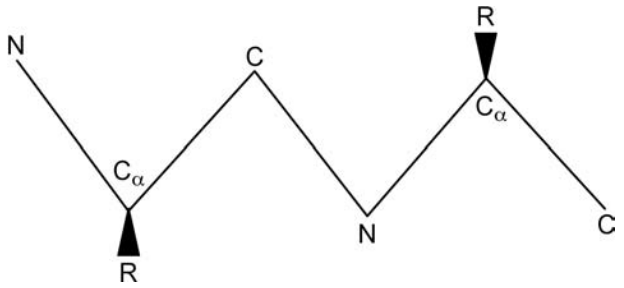


Proteins Are Like Puzzle Pieces

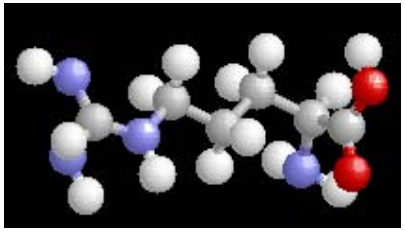


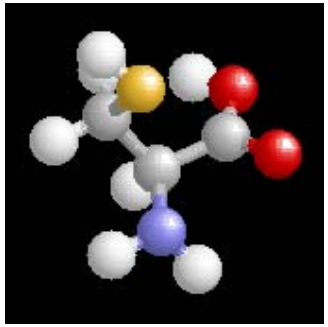
Proteins Built From Amino Acids

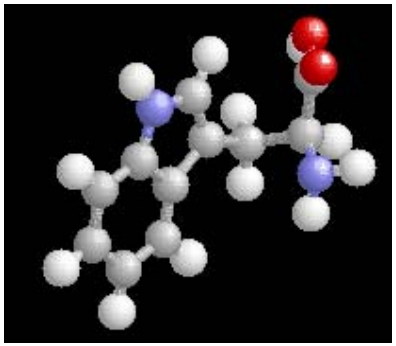
- A protein is a **chain of amino acids**, each of which has three backbone atoms (N-Cα-C) and a sidechain (R)

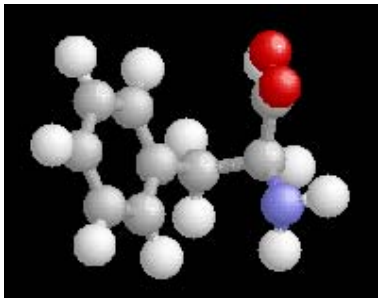


- There are **20 standard amino acids**. Each amino acid has a distinct structure and biochemical properties.


Arginine


Cysteine


Tryptophan


Phenylalanine

C = gray
H = white
O = red
N = blue
S = yellow

Amino Acid Chemical Properties

- **Hydrophobic**
  - Alanine ALA (A)
  - Isoleucine ILE (I)
  - Leucine LEU (L)
  - Methionine MET (M)
  - Phenylalanine PHE (F)
  - Proline PRO (P)
  - Valine VAL (V)

- **Negatively Charged**
  - Aspartic acid ASP (D)
  - Glutamic acid GLU (E)

- **Positively Charged**
  - Arginine ARG (R)
  - Lysine LYS (K)
  - Histidine* HIS (H)

- **Polar Uncharged**
  - Asparagine ASN (N)
  - Cysteine CYS (C)
  - Glutamine GLN (Q)
  - Histidine* HIS (H)
  - Serine SER (S)
  - Threonine THR (T)
  - Tryptophan TRP (W)
  - Tyrosine TYR (Y)

- **Small Neutral**
  - Glycine GLY (G)

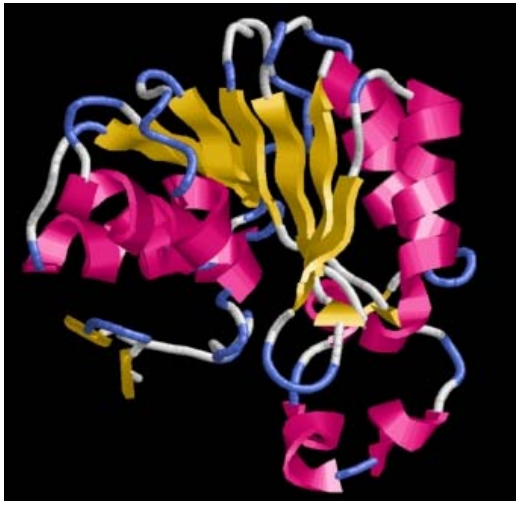
Hydrophobic ↔ Hydrophobic
Negative ↔ Positive
Negative ↔ Polar
Positive ↔ Polar
Polar ↔ Polar

Protein Structure Heirarchy

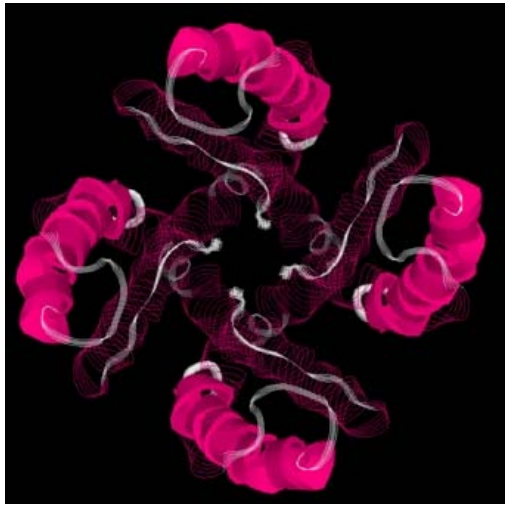
Ala Glu Lys Trp
His Cys Gly Ser
His Pro Cys Gln
Ala Met Arg Asn
Ser His Glu Phe

Primary

Secondary



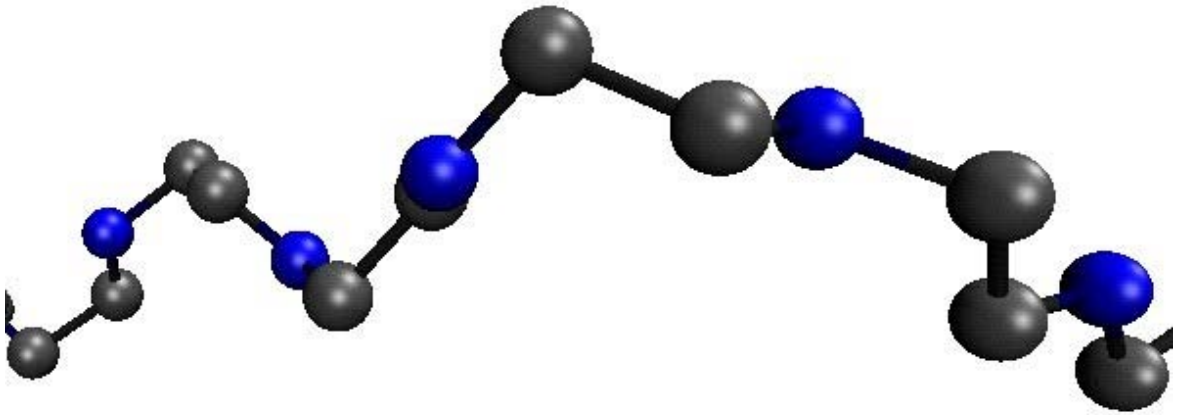
Tertiary



Quaternary

Dihedral (Torsion) Angles

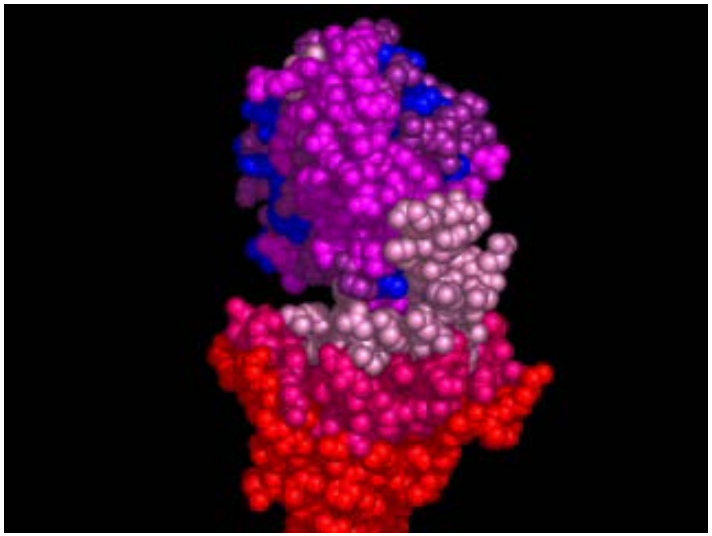
- Torsion angles are formed by **four successive atoms**.
- The ϕ, φ, and ω angles are defined for each sidechain (except for the first ϕ and last ω.)
  - ϕ is the torsion angle between the N and Cα atoms along the backbone
  - φ is the torsion angle between the Cα and C atoms along the backbone
  - ω is the torsion angle between the C from one amino acid and the N of the next – it's usually close to 180˚ and often considered fixed.

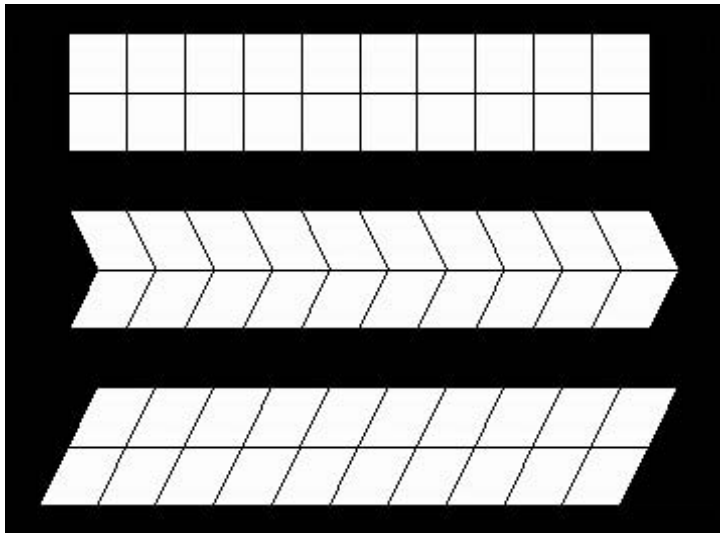


Deforming Molecules

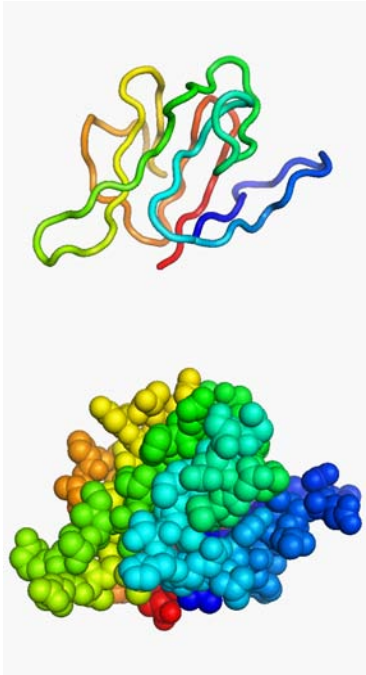
- Several types of **motions are frequently seen** in interacting proteins
  - Sidechain rearrangement, hinge bending and shear

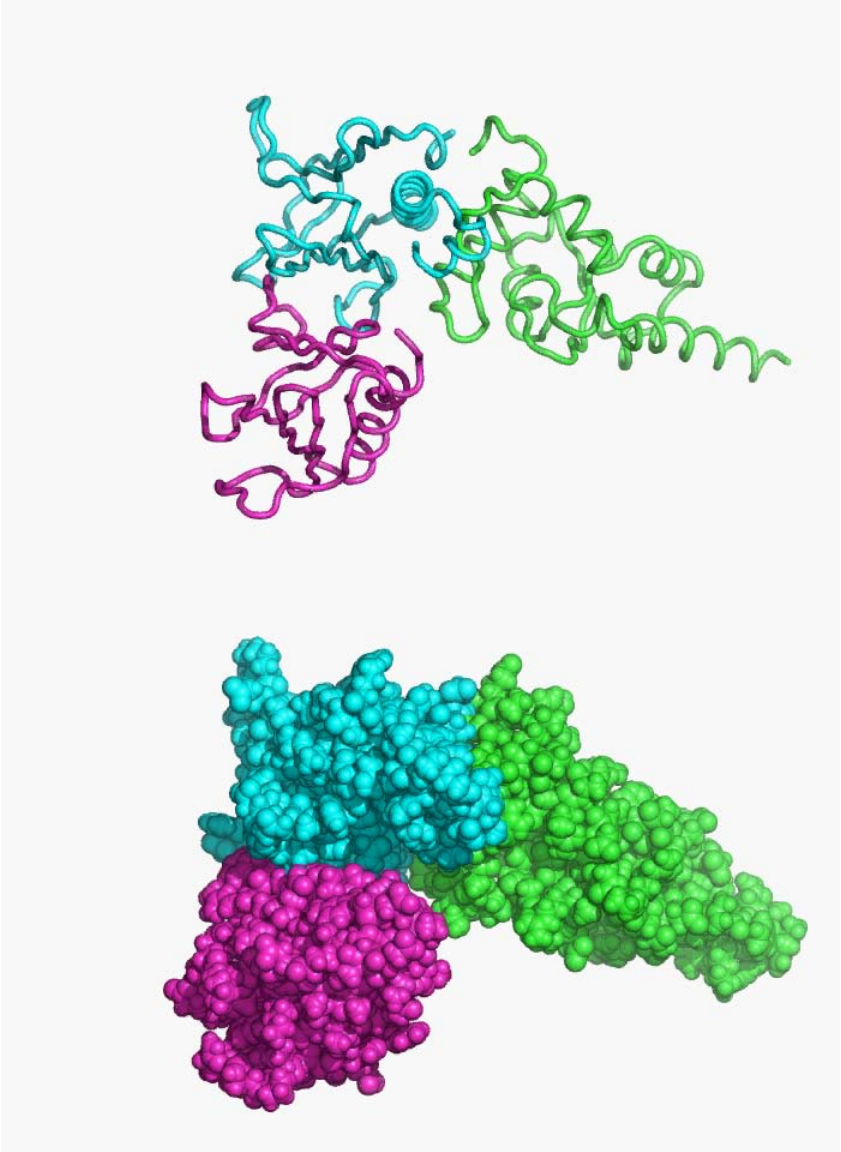
- The goal in modeling is to have **few free variables** and still approximate reality.





Other Deformations

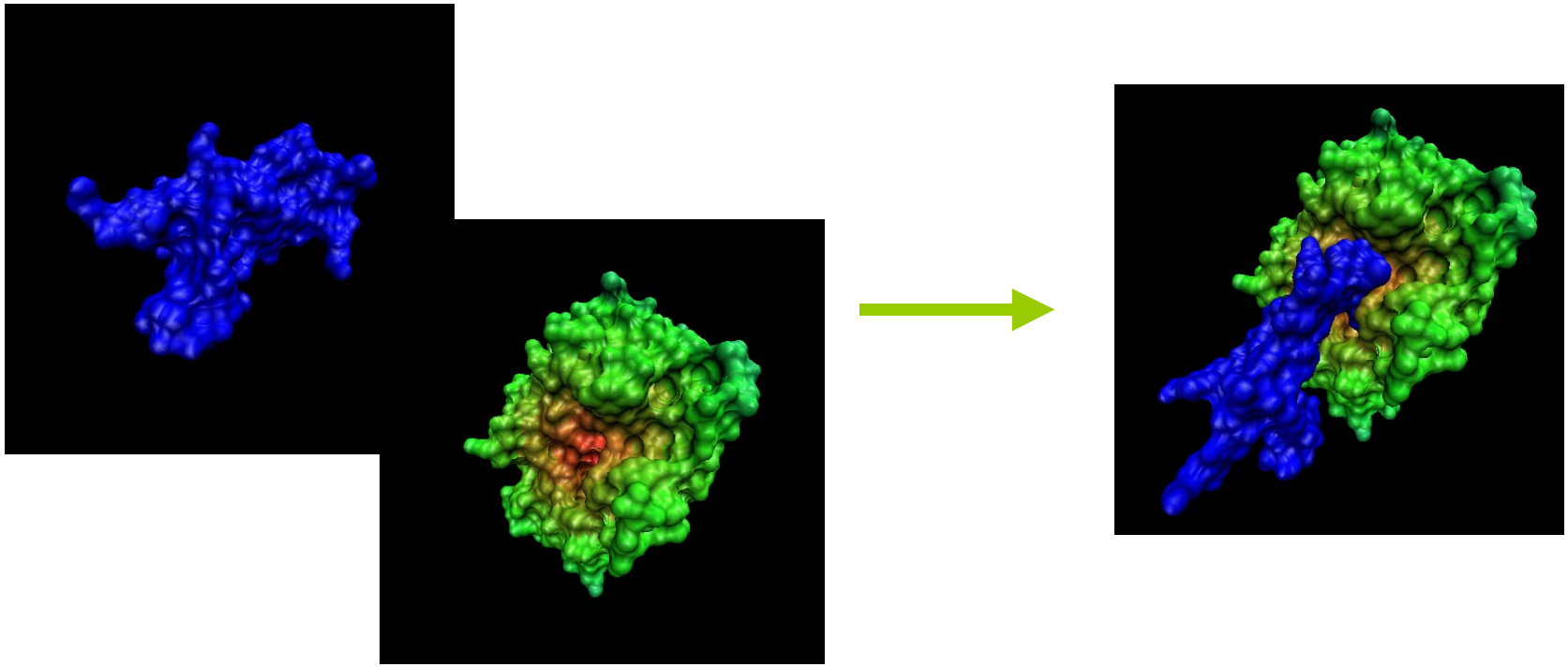




Molecules Bind or “Associate” With One Another

Protein Docking

- We want to determine **most stable** mode of association
- Most stable state corresponds to a **minimum of the energy function**



Energy Functions

- Functions that are used to describe molecular systems are generally based on **physics and chemistry**.
  - Each type of atom (carbon, nitrogen, hydrogen,etc) has unique chemical properties.
- For each pair of atoms, we can use functions to decide if the atoms **attract or repel** one another.

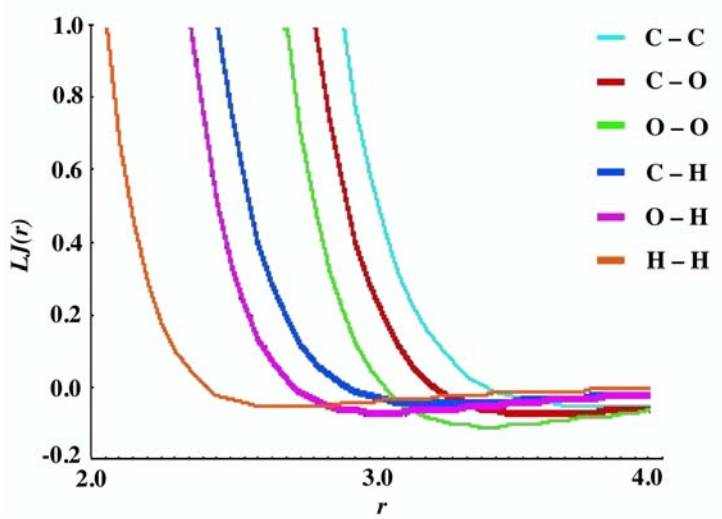
Interaction Energy

- For each **atom-atom pair**, there is an interaction energy we can calculate.
- The **interaction energy** consists of electrostatics, Lennard-Jones, desolvation and hydrogen bonding terms.
- Our goal is to find the relative configuration of molecules with the **least energy**.
- In calculus, this is the same as asking you to **find the global minimum** of a function!!

Lennard-Jones Potentials

- Nearby atoms want to align their electron orbitals, creating a **weak electrostatic force**.
- At the same time, atoms **repel when they get too close**
- The **attraction and repulsion** are often modeled using a single function, a Lennard-Jones potential.

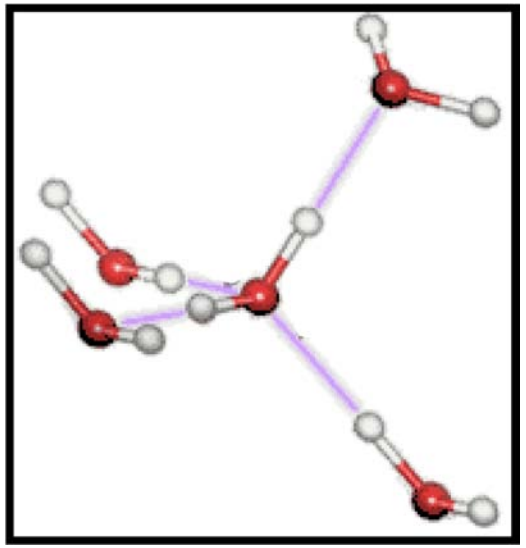
$$P_{LJ}(r) = \frac{A}{r^{12}} - \frac{B}{r^{6}}$$



Hydrogen Bonding

- Hydrogen bonds are the **strongest non-covalent bonds**.
- Hydrogen bonds **join DNA base pairs**, and help proteins fold into their functional forms.
- The formation of hydrogen bonds is also **important to protein interactions**.

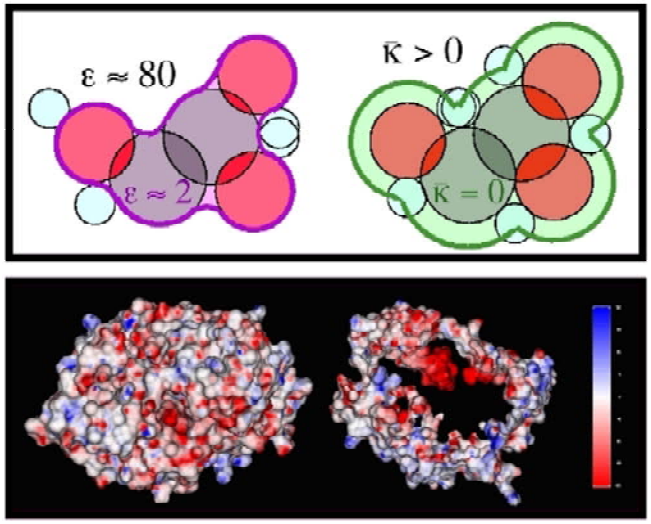
$$P_{hb} = \frac{C \cos \theta}{r^{10}}$$

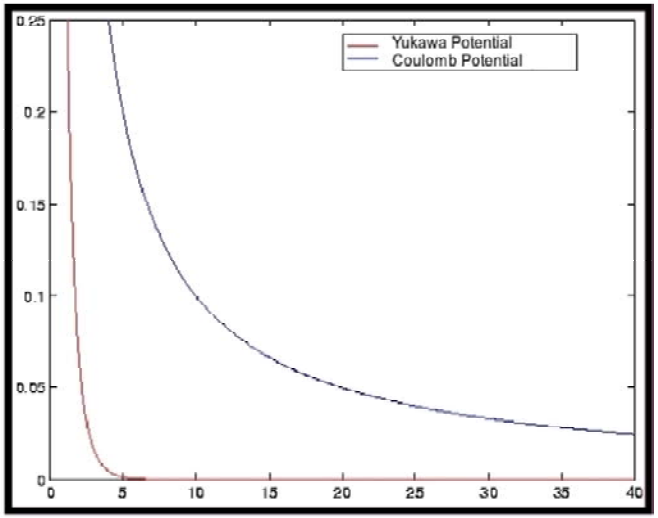


Electrostatics

- Classical energies to describe charge-charge interaction vary as the **reciprocal of distance**. The average effects of **temperature, water and ions**

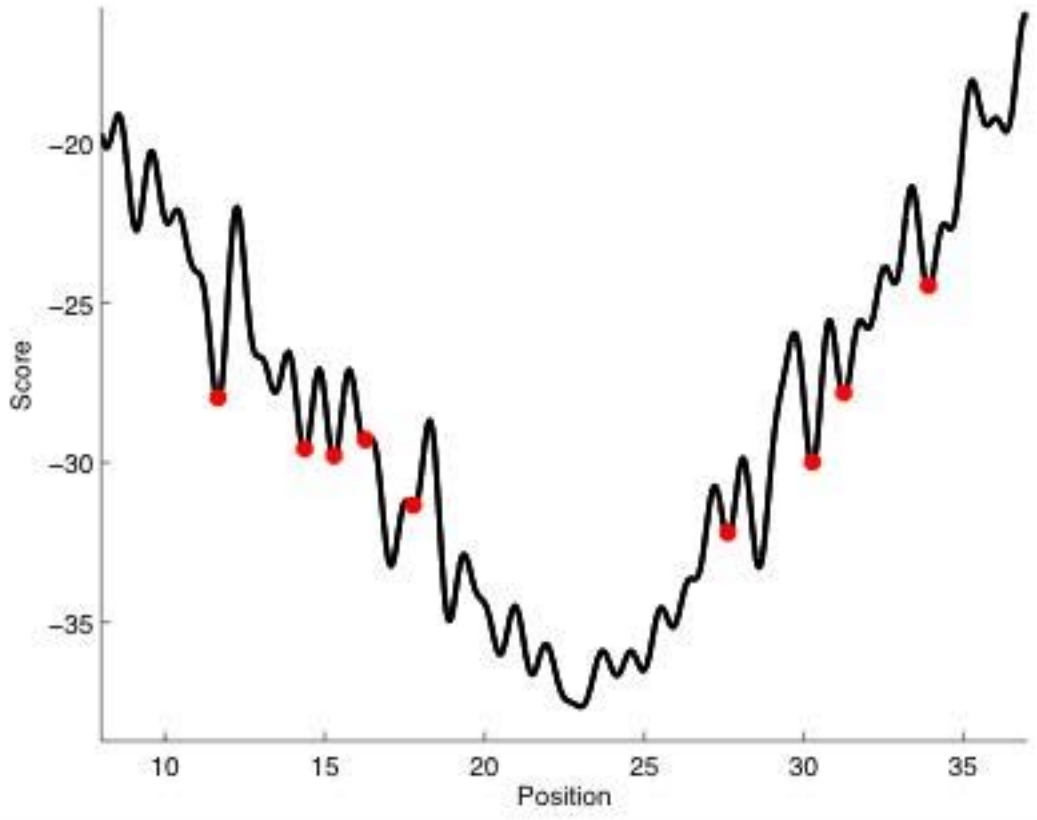
$$P_{elec} = \frac{e^{-\bar{\kappa}r}}{\epsilon r}$$





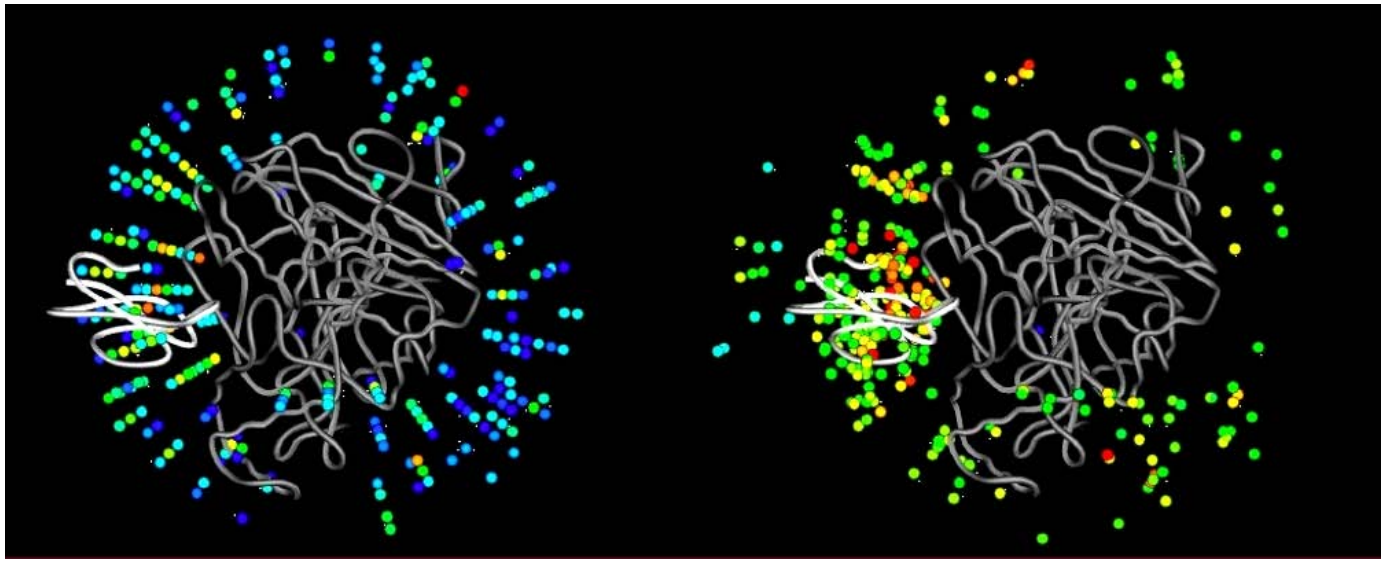
Energy Minimization And Observed Behavior

Local Minima Of Functions



Underestimation Schemes

- **Fact** — Multiple minima = **hard optimization problem**



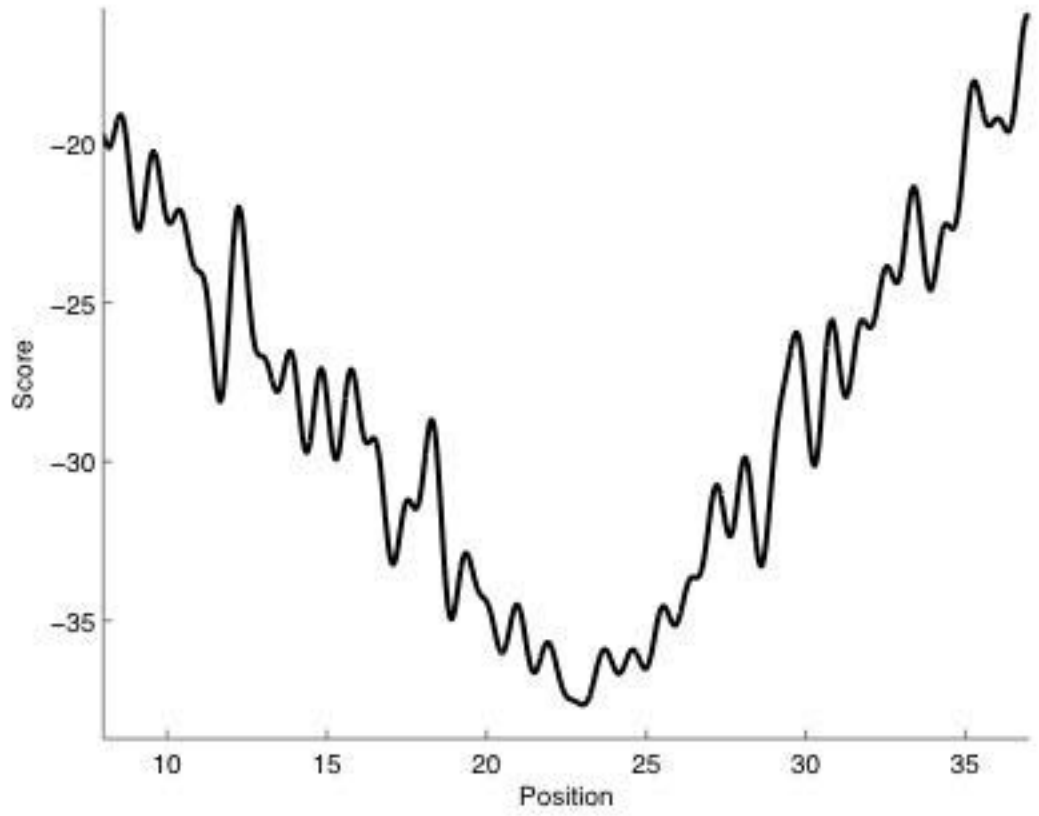
- **Theory** (Onuchic, Wolynes, Dill) — The physics of protein folding and docking suggests an energy "funnel."

- **Idea** (Dill, Rosen, Phillips) — use local minima to construct underestimators. These functions **capture large-scale details** of the energy funnel.

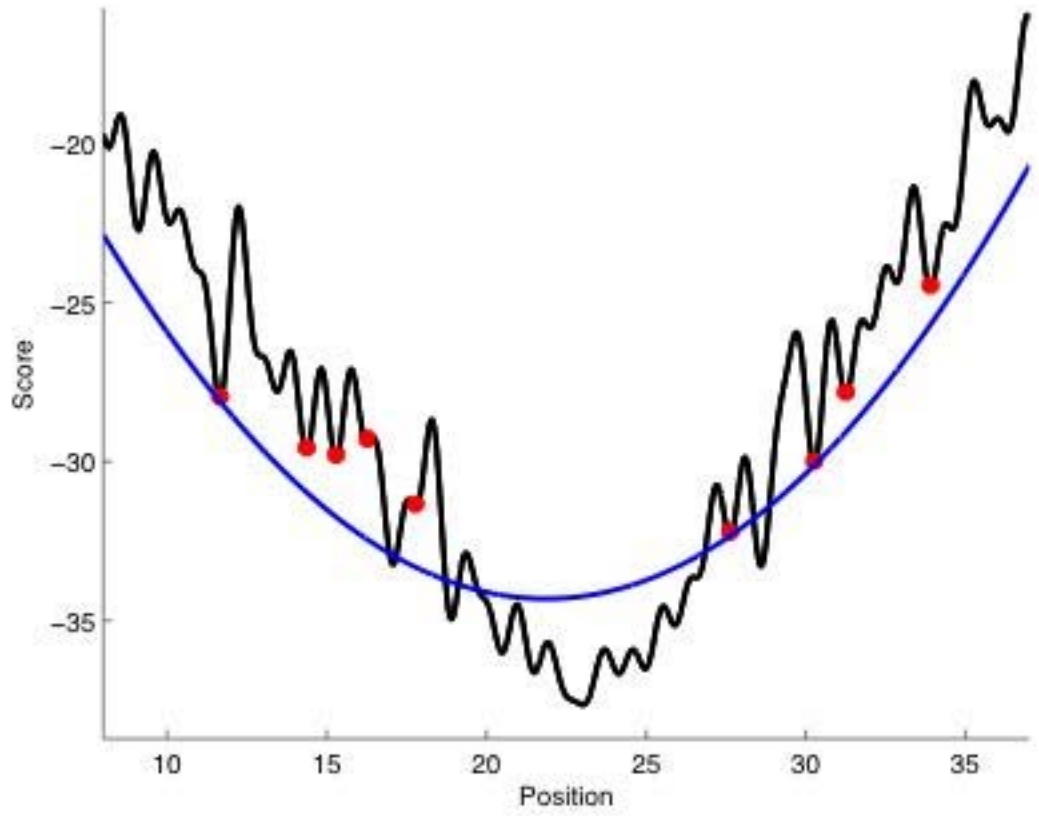
Low-Energy States Are The Norm

- We all understand the concept of the **path of least resistance**
  - This is what molecules do when they have to move between locations, or if they need to change their shape.
  - Molecules will do the **least amount of work possible**.
  - Molecules like to **settle into low-energy states**.
  - Molecules are the couch potatoes of the microscopic world.

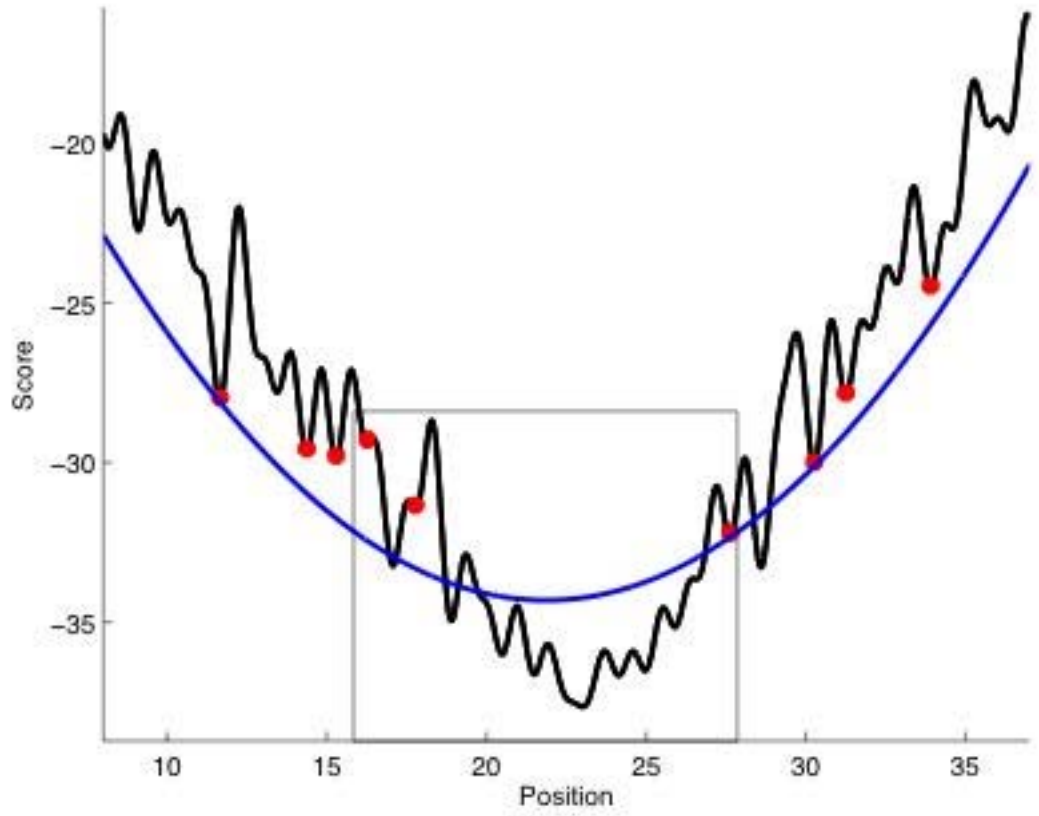
Underestimators



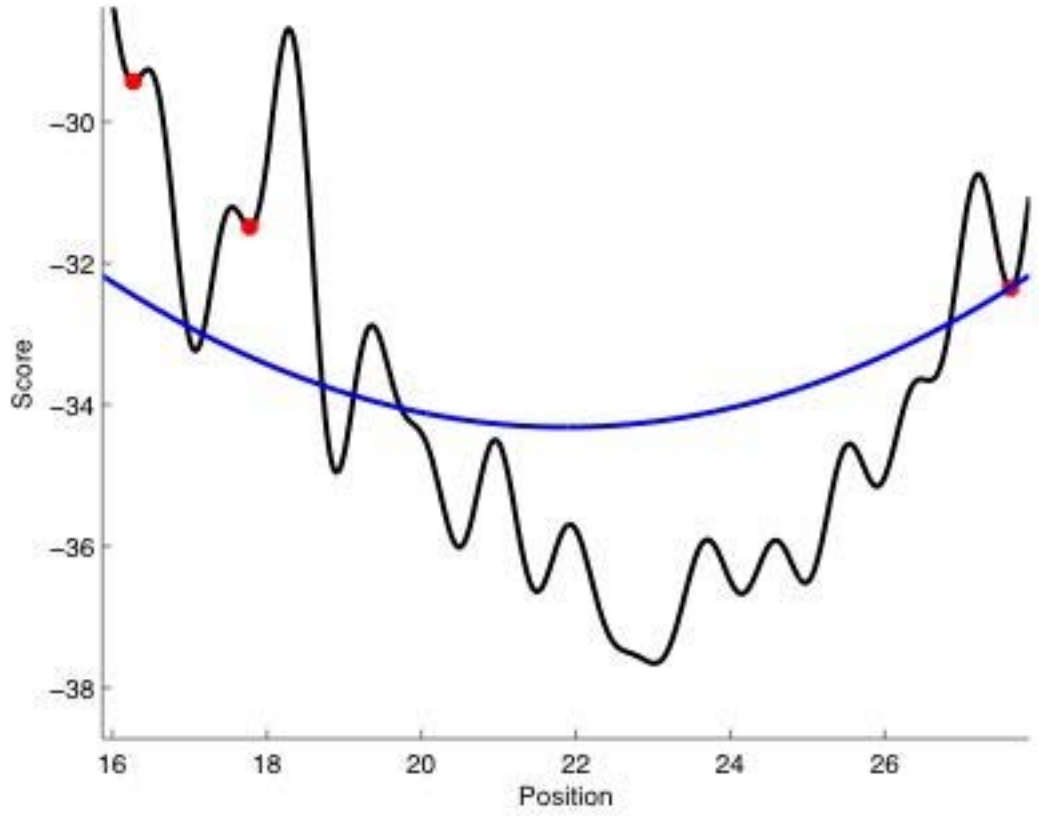
Underestimators



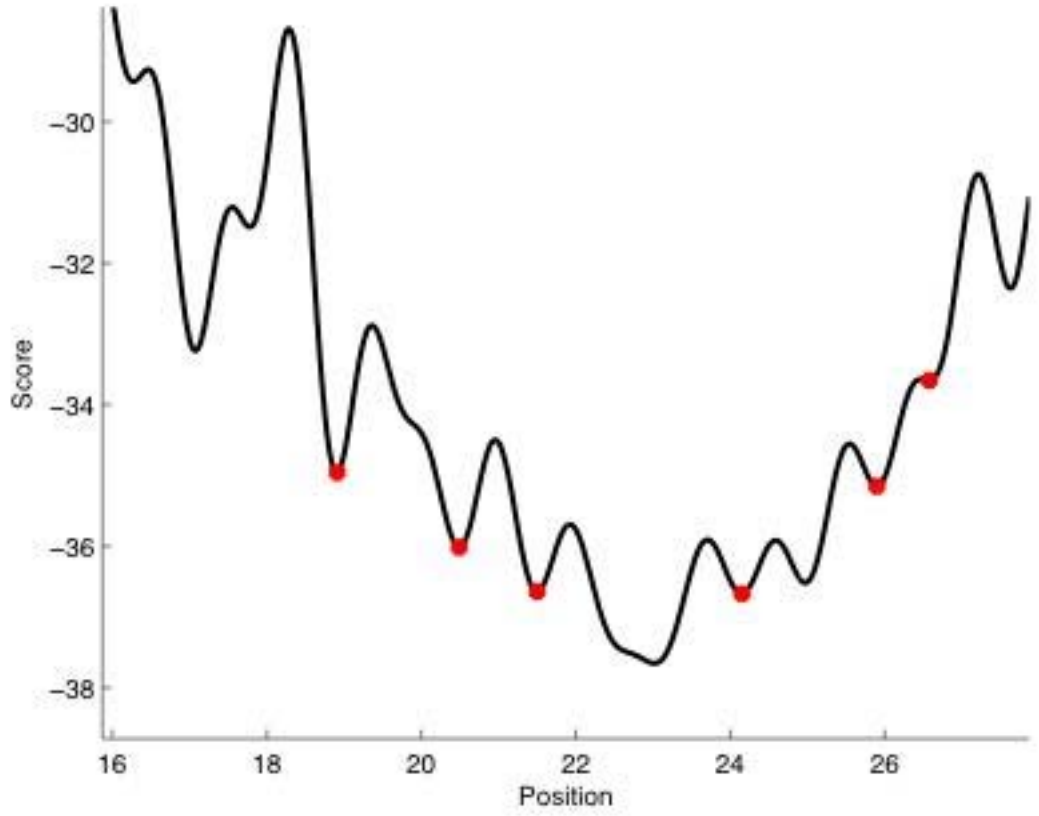
Underestimators



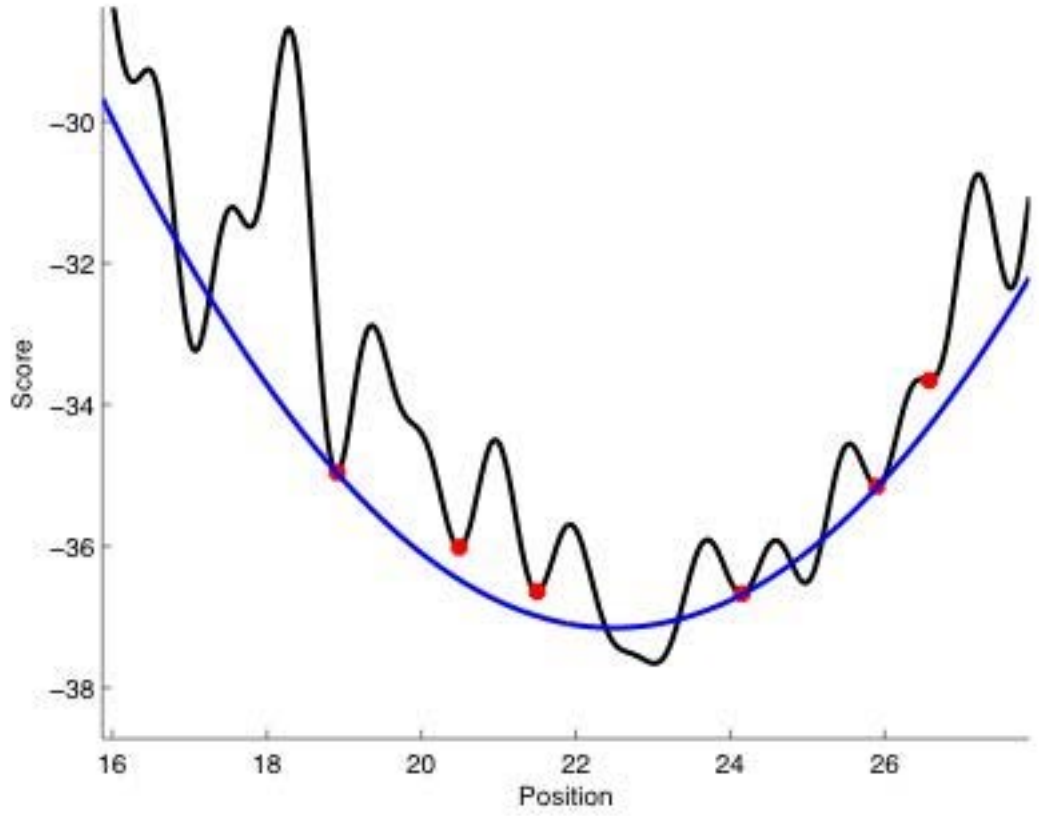
Underestimators



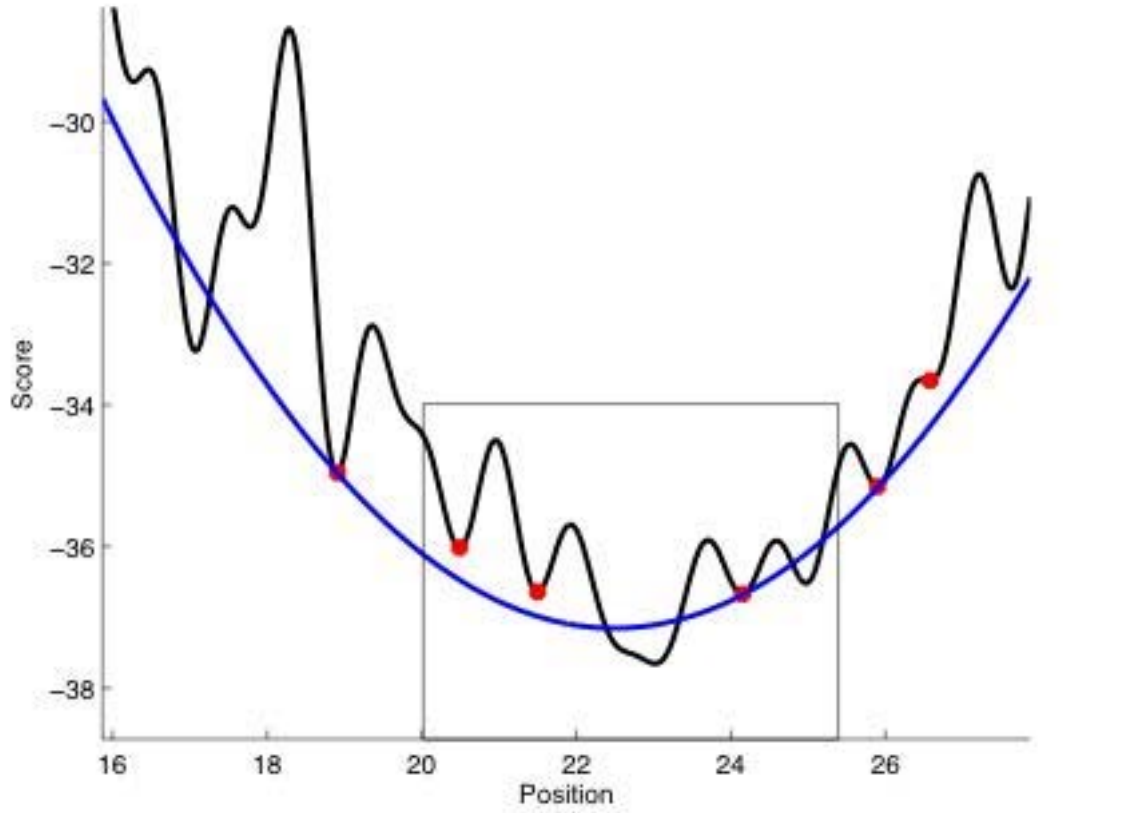
Underestimators



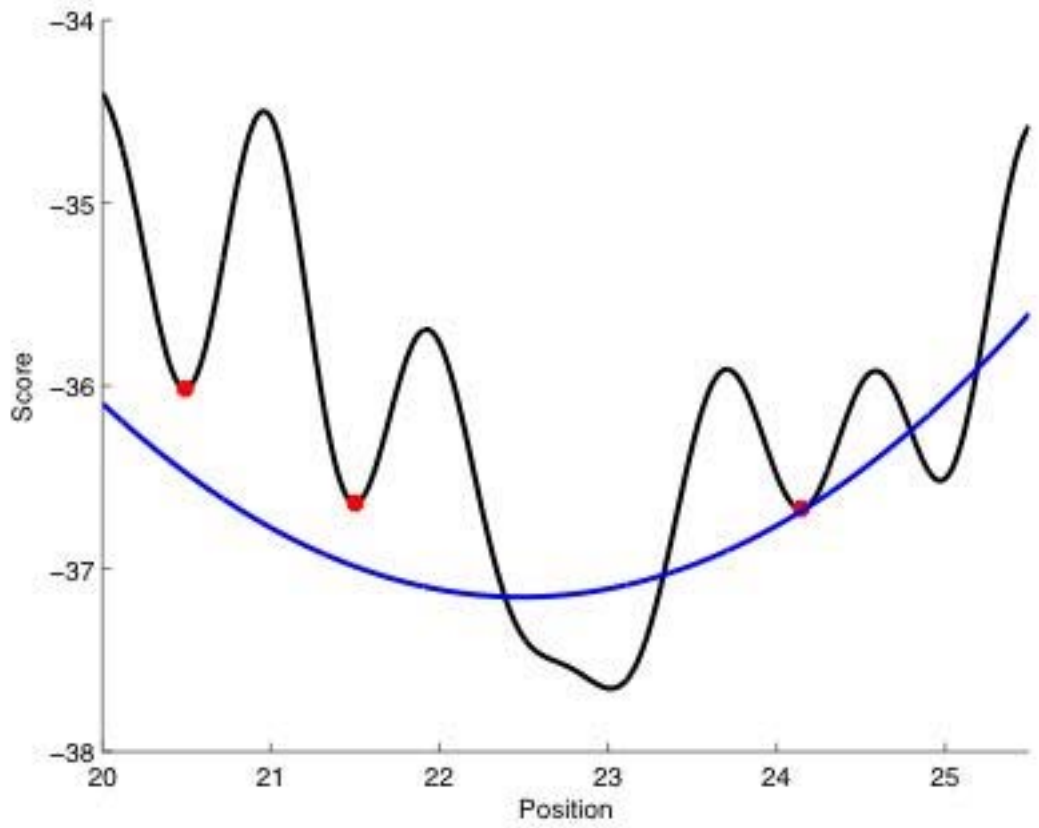
Underestimators



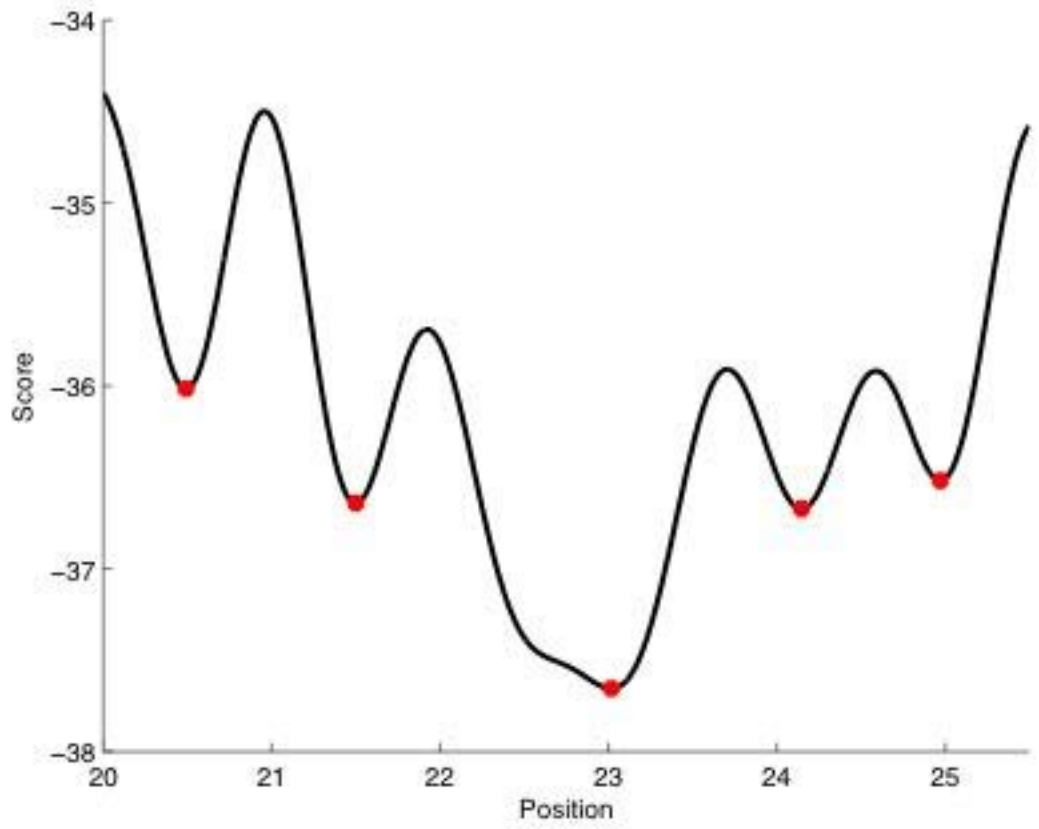
Underestimators



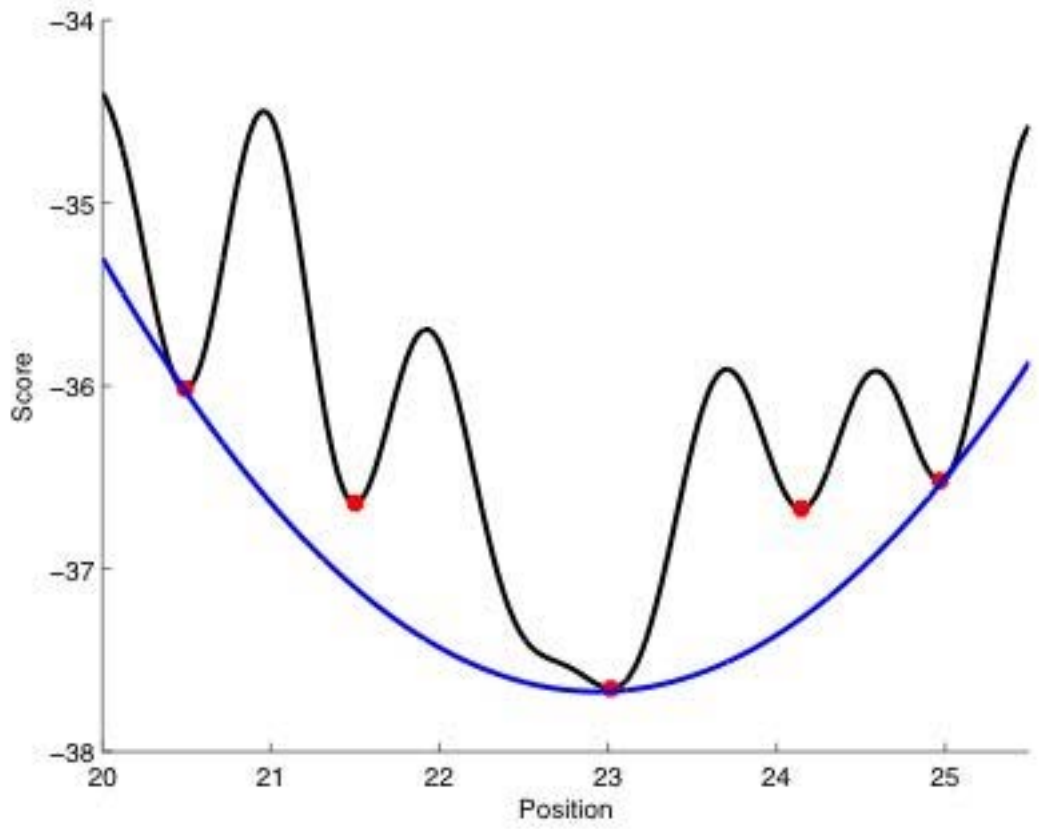
Underestimators



Underestimators



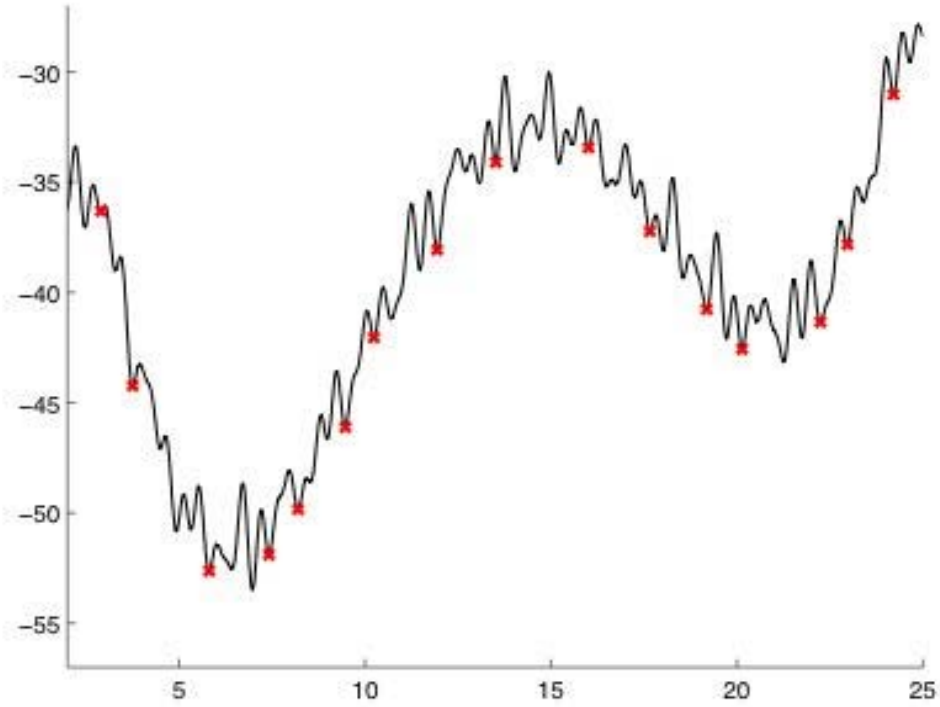
Underestimators



Multi-funnel Functions

- What if our function really looks like this??

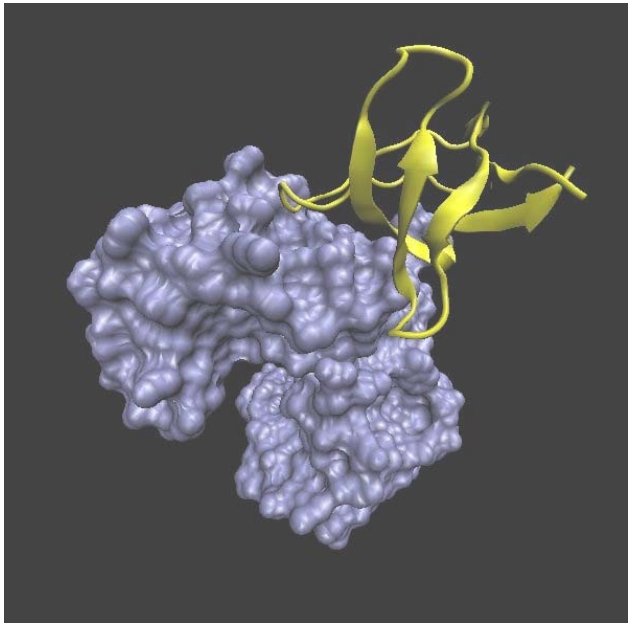
- We will **divide and conquer** our search space, and do underestimation in each region!



1AVZ: Single Underestimator

- 1AVZ: V-1 Nef protein in complex with wild type Fyn SH3 domain

- The interaction ensures long-term survival of infected T cells and the destruction of non-infected T cells by causing these cells to die.

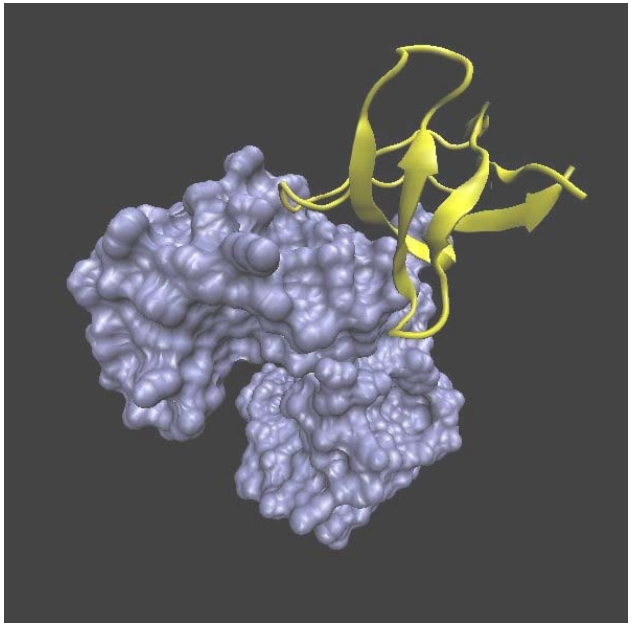


Iter	$E(x^0_{\min})$	$E(x^f_{\min})$	$\|\cdot\|_2$	$\|\cdot\|_\theta$
5	-47.406	-60.195	5.556	166°

1AVZ: Divide And Conquer

- Instead of a single global minimum, **better results are achieved using divide and conquer**

- These “local global minima” define the minima of deep energy basins.

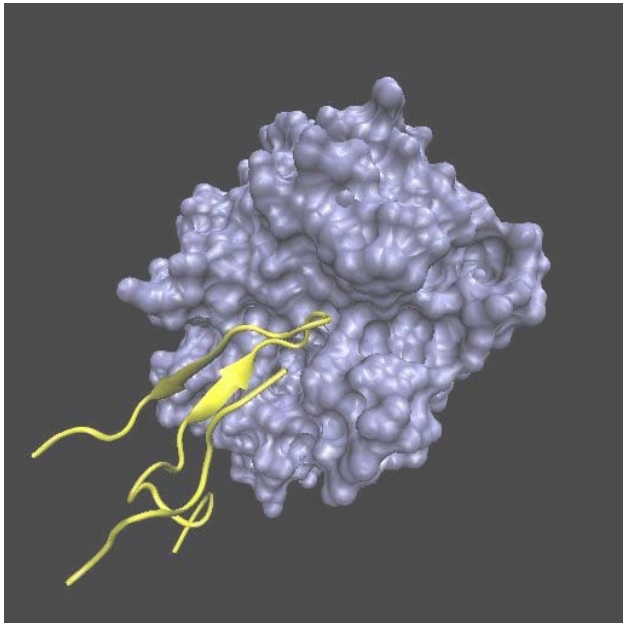


Cluster	Iter	$E(x^0_{\min})$	$E(x^f_{\min})$	$\|\cdot\|_2$	$\|\cdot\|_\theta$
1	3	-31.071	-45.214	24.565	119°
2	4	-42.132	-45.280	24.855	105°
3	4	-45.515	-53.381	6.353	171°
4	4	-47.406	-76.947	0.276	5°
-	5	-47.406	-60.195	5.556	166°

1TAB: Single Underestimator

- 1TAB: Trypsin complex with the Bowman-Birk inhibitor

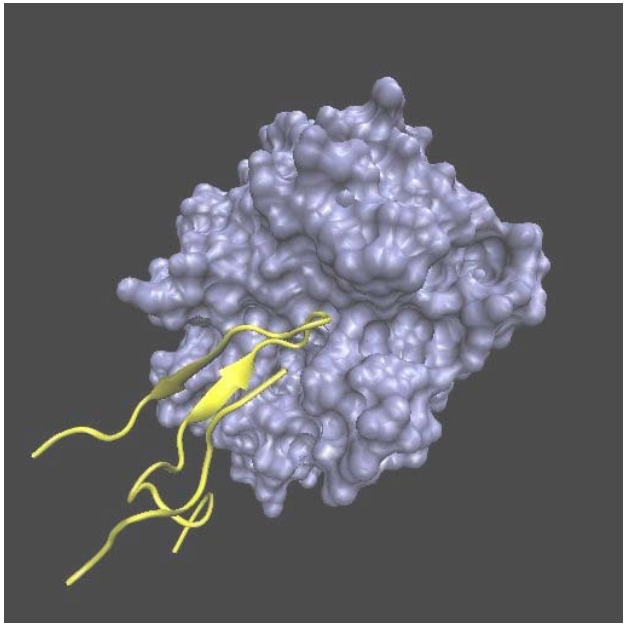
- Elevated levels of trypsin have been found in pancreatic tumors, and BBI has been shown to suppress this type of tumor in various animals.



Iter	$E(x^0_{\min})$	$E(x^f_{\min})$	$\|\cdot\|_2$	$\|\cdot\|_\theta$
5	-52.275	-53.524	36.392	63°

1TAB: Divide And Conquer

- Once again, the divide and conquer optimization scheme led to a much **better predicted global minimum**



Cluster	Iter	$E(x^0_{\min})$	$E(x^f_{\min})$	$\|\cdot\|_2$	$\|\cdot\|_\theta$
1	5	-41.503	-44.450	36.134	159°
2	3	-50.438	-59.761	0.092	5°
3	3	-47.359	-47.587	30.139	32°
4	3	-52.275	-53.503	36.360	63°
-	5	-52.275	-53.524	36.392	63°

Conclusions

- Mathematics can play an **important role in biology and chemistry**
  - At the molecular scale, biology happens because of the **unique shapes of molecules**
  - Using geometric models, we can **understand the behavior of know molecules and design new molecules** to treat deadly diseases