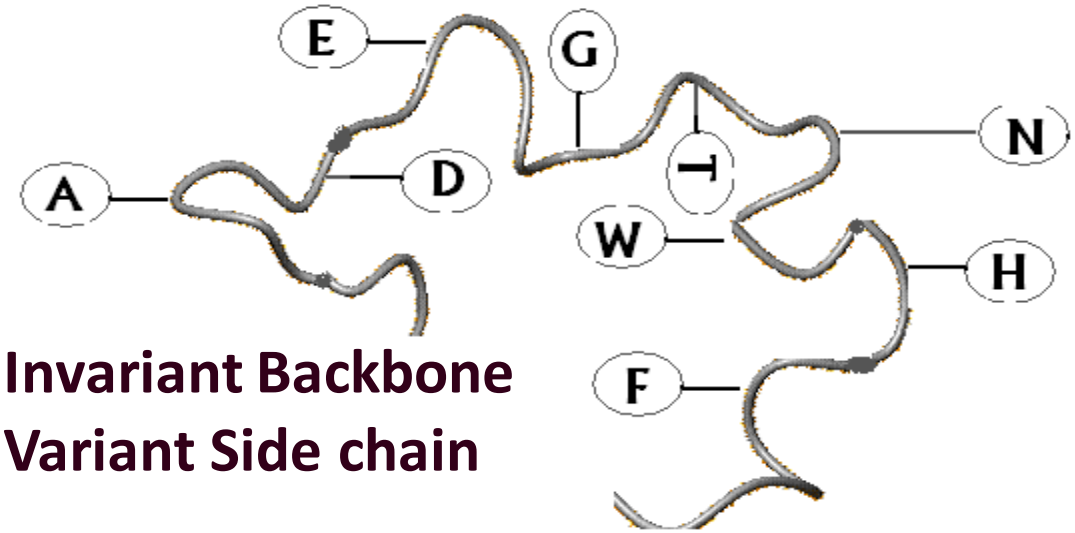
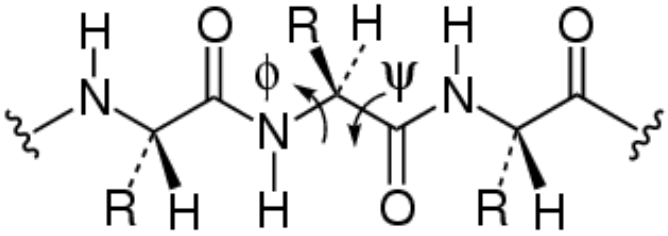
Lecture 5: Protein Folding & Structure Prediction

Polypeptide Chain





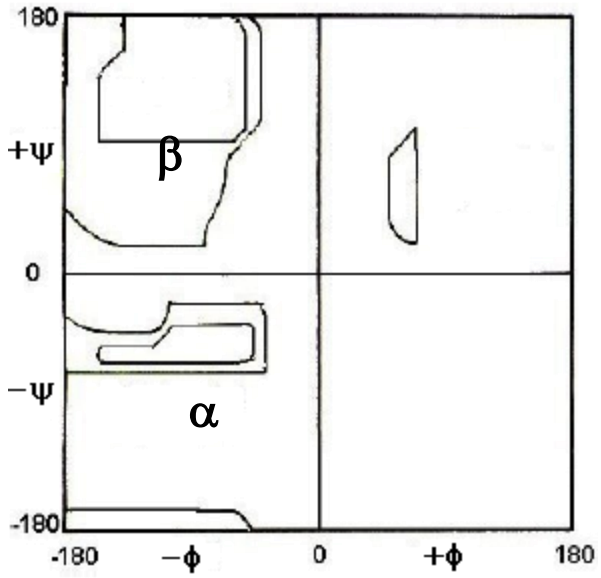
Invariant Backbone
Variant Side chain

Conformationally constrained chain due to the planarity of the peptide bond

Conformational variables: dihedral angles

Local organization restricted to two - alpha helices and beta sheets

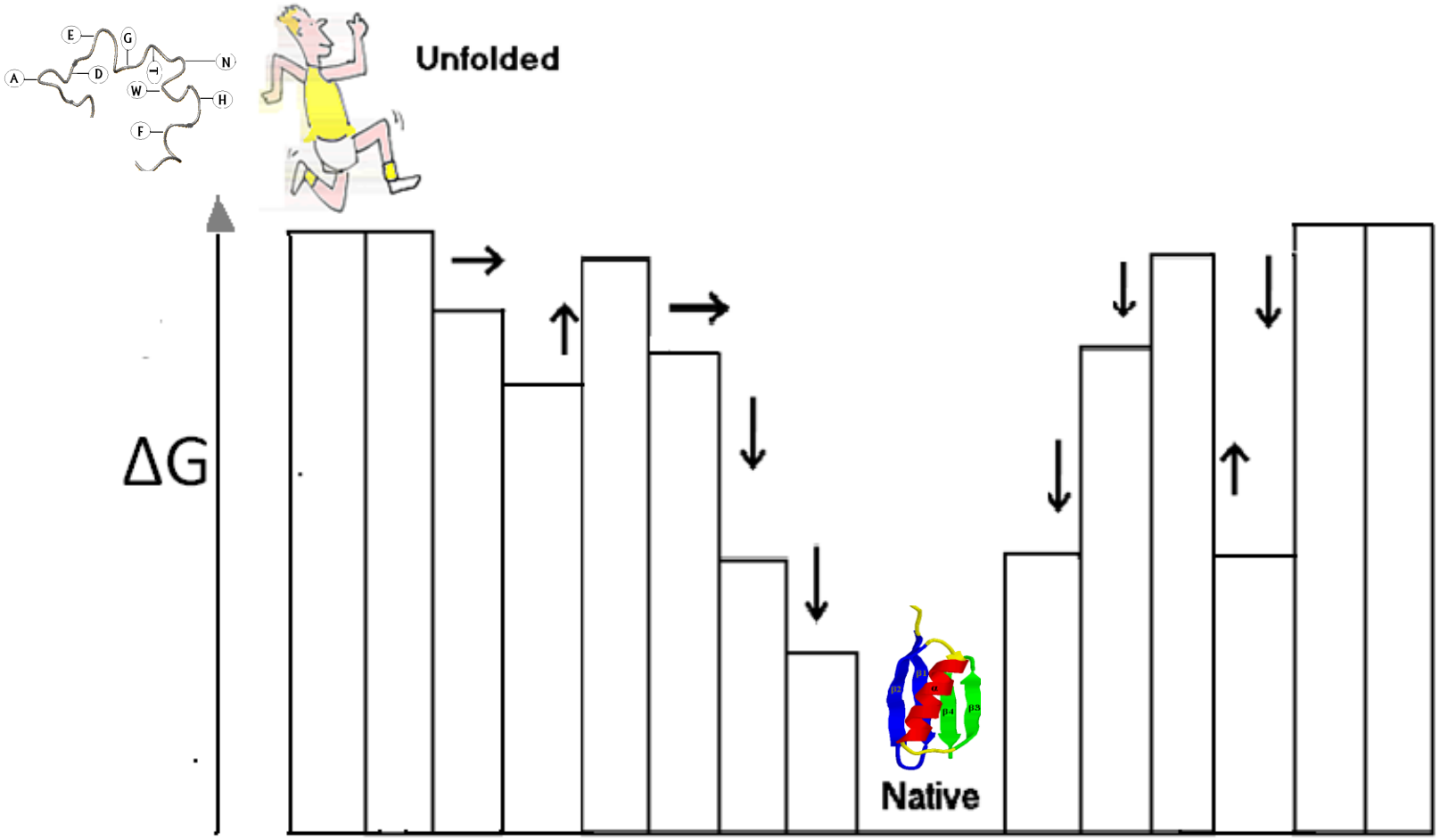
Global organization restricted to only ~1000 variants



Restricted Access (~21%)

Protein Folding Problem

- PF Problem is the difficulty in predicting a three-dimensional structure of a protein from its amino acid sequence.
- How does a protein fold into its prescribed functional structure?



The three dimensional structure of a native protein in its normal physiological conditions is the one in which Gibbs free energy of the whole system is the lowest; that is native conformation is determined by the totality of inter-atomic interactions and hence amino acid sequence

----Thermodynamic hypothesis, Anfinsen 1973

Levinthal's Paradox

We assume that there are three conformations for each amino acid (ex. α-helix, β-sheet and random coil). If a protein is made up of 100 amino acid residues, a total number of conformations is

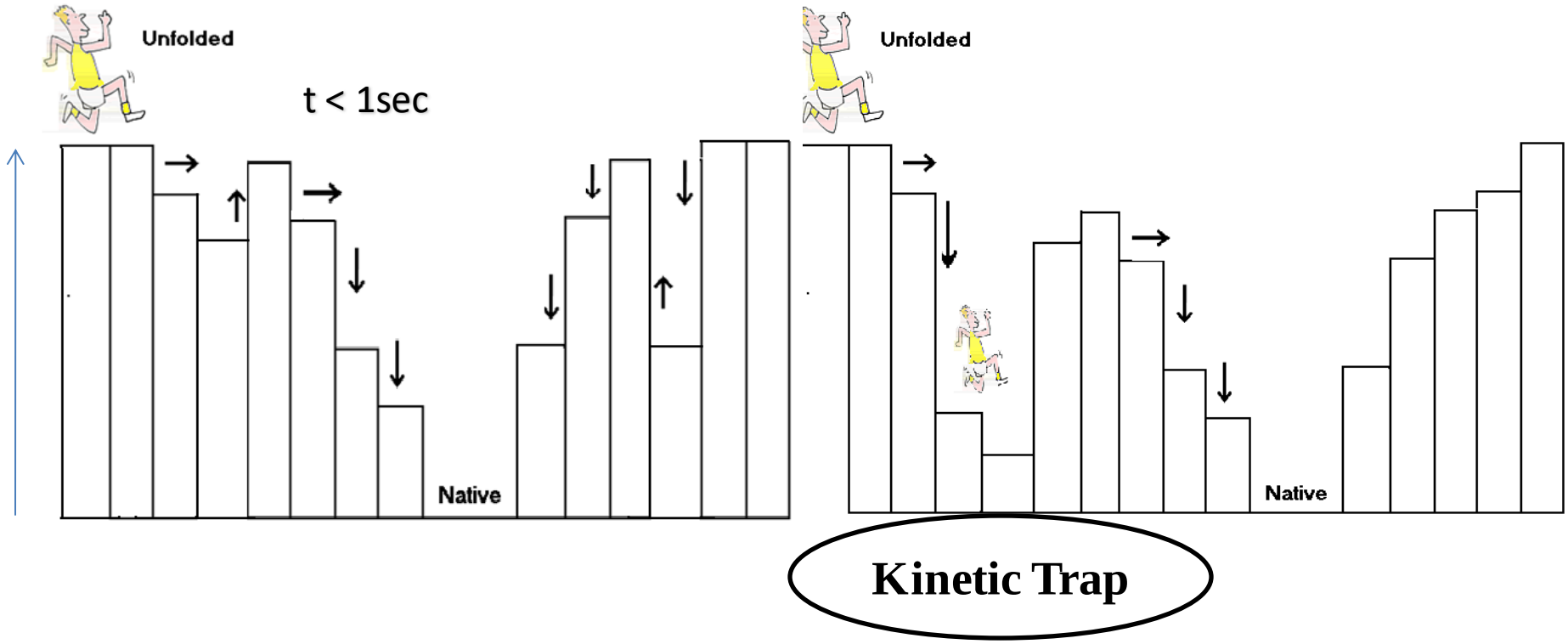
$3^{100} = 515377520732011331036461129765621272702107522001$
$\fallingdotseq 5 \times 10^{47}.$

If 100 psec (10^{-10} sec) were required to convert from a conformation to another one, a random search of all conformations would require

$5 \times 10^{47} \times 10^{-10}$ sec $\fallingdotseq 1.6 \times 10^{30}$ years.

However, folding of proteins takes place in milli second to second order. Therefore, proteins fold not *via* a random search but a more sophisticated search process.

We want to watch the folding process of a protein using molecular simulation techniques.



Only a tiny fraction of total possible conformations available to a polypeptide chain can be sampled during folding; the subset of conformations can be viewed as a kinetic Pathway.

-----Kinetic hypothesis, Levinthal 1968

Kinetic trap along the pathway slows down folding, and if such large TS barrier is for the folded state, such proteins have high degree of kinetic stability

Difficulty in studying Protein Folding:

From the view point of computer simulation,

1. It is difficult to simulate the whole process of protein folding at atomistic level using even state-of-the-art computers.
2. It is uncertain whether the accuracy of current energy functions and parameters are sufficient for protein folding simulation or not.

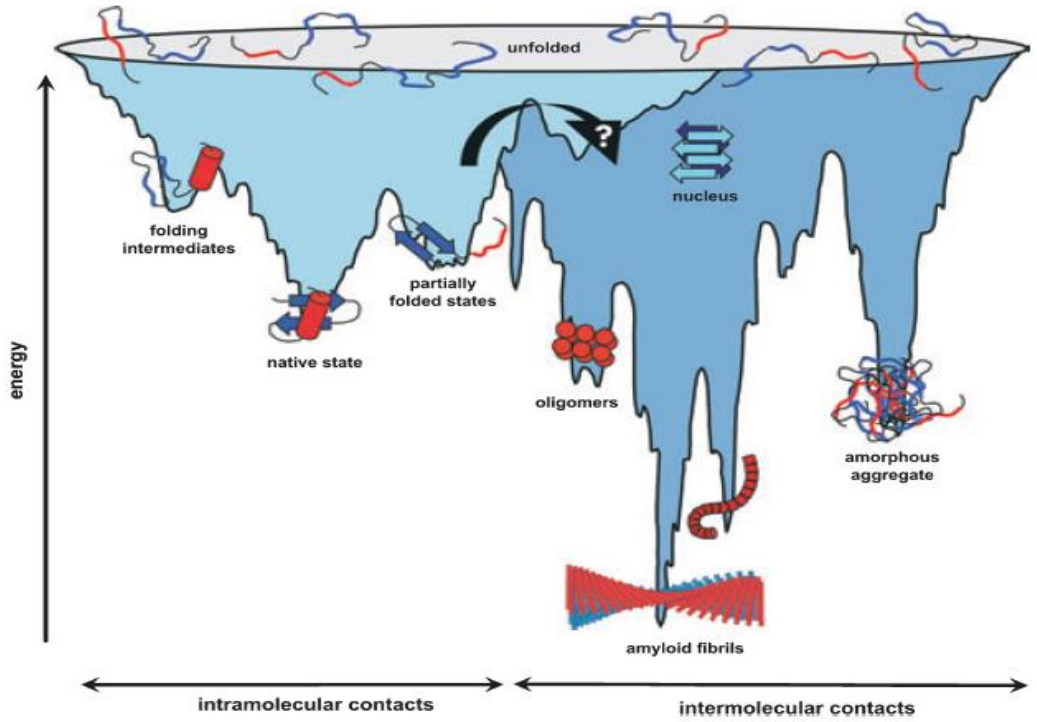
…, let me recount a conversation with Francis in 1975 (who won the Novel prize for discovering the structure of DNA). Crick stated that "it is very difficult to conceive of a scientific problem that would not be solved in the coming twenty years … except for a model of brain function and protein folding". Although Crick was more interested in brain function, he did state that both problems were difficult because they involve many cooperative interactions in three-dimensional space.

(Levitt M, "Through the breach." *Curr*. *Opin*. *Struct*. *Biol*. 1996, **1**, 193-194)

Courtesy: Tokyo University of Science Tadashi Ando

Understanding Protein *mis*-folding pathways

Multiple intermediate states through multiple pathways funneled towards stable native state



Intra-molecular *Vs* Inter-molecular

- **Understanding the kinetic mechanism of misfolding and aggregation.**
- **Explore the effect of mutations (sequence and topological) on folding pathways**

Jahn and Radford, FEBS, 2005

Importance or Protein Folding

Proteins play important roles in living organisms.

Some proteins are deeply related with diseases. And structural information of a protein is necessary to explain and predict its gene function as well as to design molecules that bind to the protein in drug design.

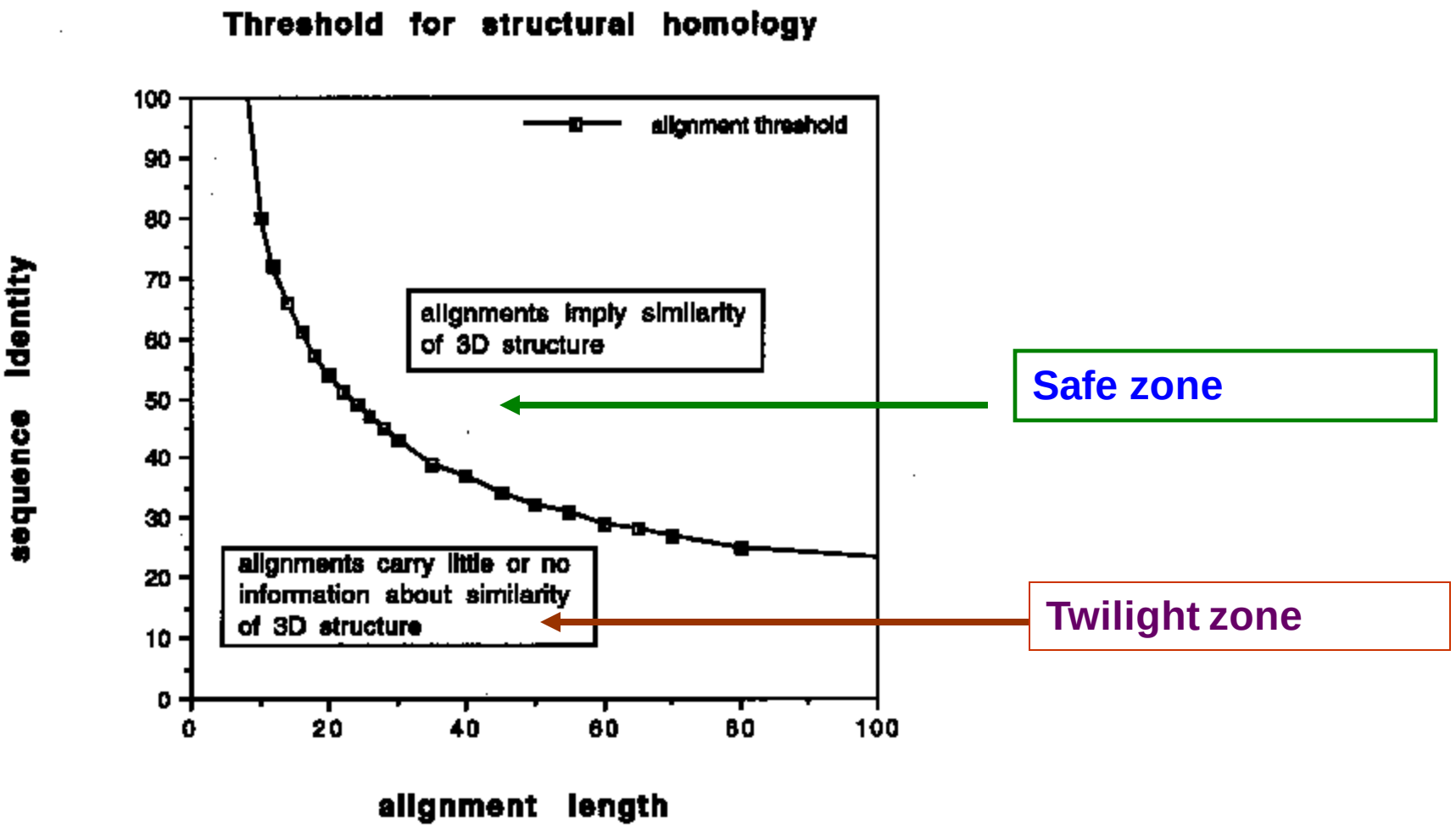
Today, whole genome sequences (the complete set of genes) of various organisms have been deciphered and we realize that functions of many genes are unknown and some are related with diseases.

Therefore, understanding of protein folding helps us to investigate the functions of these genes and to design useful drugs against the diseases efficiently.

In addition to that, the understanding opens the door to designing of proteins having novel functions as new nano machines.

Courtesy: Tokyo University of Science Tadashi Ando

Homology Modeling



Homology Modeling …simple

Template Recognition and initial alignment

V A T T P D K S W L T V

A S T P E R A S W L G T A

V A T T P D K _ S W L T V _

_ A S T P E R A S W L G T A

* Residues with similar proerties in **green**

_ is a gap

Homology Modeling

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

Residue Exchange matrix !!!!

Homology Modeling

Alignment Correction

L T L T L T L T

Y A Y A Y A Y A

Poor possibilities of alignment

L T L T L T L T

T Y T Y T Y T Y

Y A Y A Y A Y A

Concept …of Multiple Sequence Alignment

Homology Modeling

Loop modeling

✓ **Good News**: Conformational change wont happen within regular secondary structural elements.

So its safe to shift all insertions or deletions in the alignment out of helices and strands.

■ **Bad News**: Loops are notoriously difficult to predict.

Possible problems:

❖ Surface loops tend to be involved in crystal contacts…

❖ Exchange of small with bulky sidechains underneath the group pushes it aside

❖ Mutation of loop residue to pro or from ala to pro restrict their respective Ramachandran space considerably

Homology Modeling

Loop modeling approaches:

Knowledge based: Search PDB with endpoints that match the residues between the loop and copy the loop conformation

Energy based: ab initio fold prediction with an energy function, MD, MC etc

- Principles of Molecular Modelling

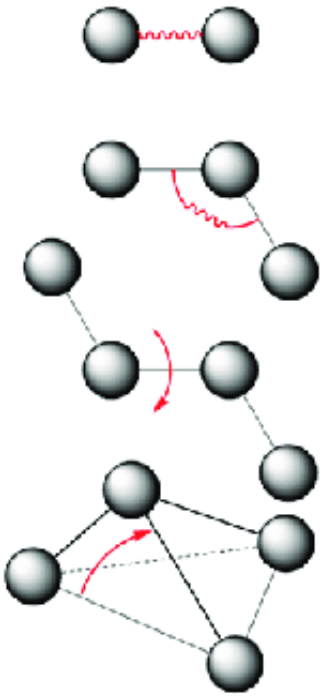
Force Field

$$U(R) = \sum_{bonds} k_r \left(r - r_{eq}\right)^2$$ *bond*

$$+ \sum_{angles} k_\theta \left(\theta - \theta_{eq}\right)^2$$ *angle*

$$+ \sum_{dihedrals} k_\phi \left(1 + \cos\left[n\phi - \gamma\right]\right)$$ *dihedral*

$$+ \sum_{impropers} k_\omega \left(\omega - \omega_{eq}\right)^2$$ *improper*



$$+ \sum_{i<j}^{atoms} \varepsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^{6} \right]$$ *van der Waals*

$$+ \sum_{i<j}^{atoms} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$ *electrostatic*