Introduction to Bioprocess Engineering SQBI2513

Enzymes: Biologists and engineers think differently

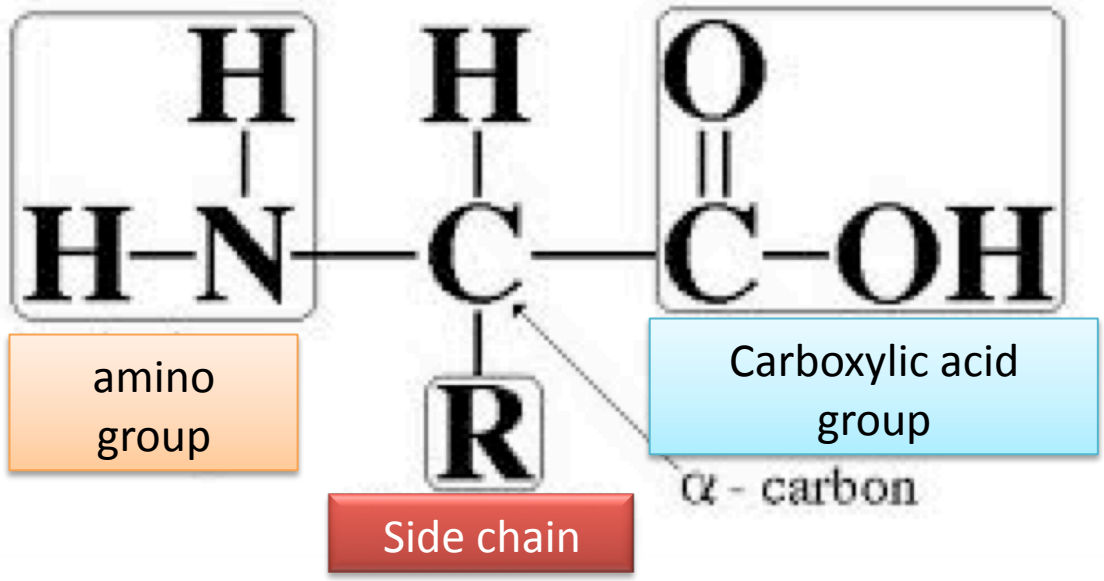
Kian Mau GOH, PhD

Faculty of Biosciences and Bioengineering

http://teknologimalaysia.academia.edu/GohKianMau/CurriculumVitae

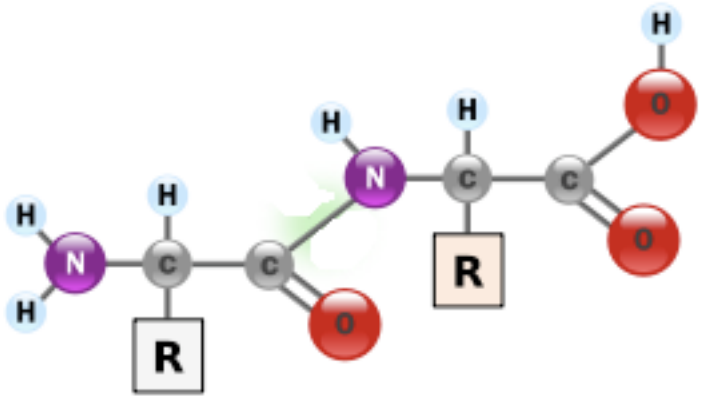
Basic amino acid structure

- All amino acids found in proteins have this basic structure, differing only in the structure of the R-group or the side chain.

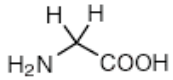


Peptide/protein chain

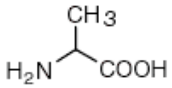
- A peptide or protein is formed by a complex of amino acids linkages
- The flow of link "N-C-C-N-C-C-N-C-C-N-C-C-N-C-C-N-C-C-N-C-C-..." is name as backbone.
- Each aa (shown in different colour) is linked by peptide bond (bold line -)



Small

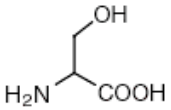


Glycine (Gly, G)
MW: 57.05

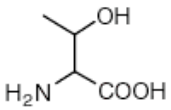


Alanine (Ala, A)
MW: 71.09

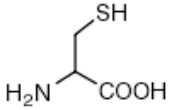
Nucleophilic



Serine (Ser, S)
MW: 87.08, pK_a ~ 16

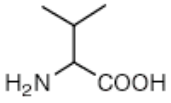


Threonine (Thr, T)
MW: 101.11, pK_a ~ 16

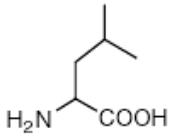


Cysteine (Cys, C)
MW: 103.15, pK_a = 8.35

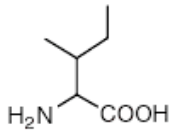
Hydrophobic



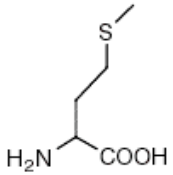
Valine (Val, V)
MW: 99.14



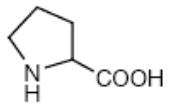
Leucine (Leu, L)
MW: 113.16



Isoleucine (Ile, I)
MW: 113.16

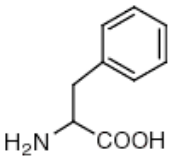


Methionine (Met, M)
MW: 131.19

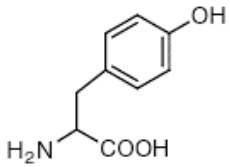


Proline (Pro, P)
MW: 97.12

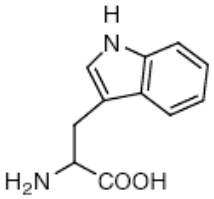
Aromatic



Phenylalanine (Phe, F)
MW: 147.18

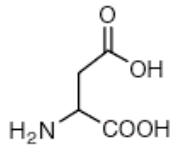


Tyrosine (Tyr, Y)
MW: 163.18

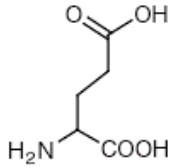


Tryptophan (Trp, W)
MW: 186.21

Acidic

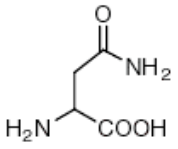


Aspartic Acid (Asp, D)
MW: 115.09, pK_a = 3.9

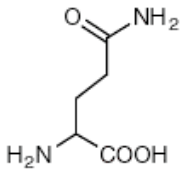


Glutamic Acid (Glu, E)
MW: 129.12, pK_a = 4.07

Amide

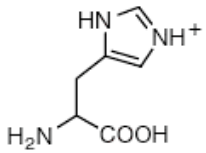


Asparagine (Asn, N)
MW: 114.11

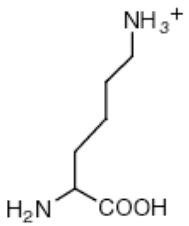


Glutamine (Gln, Q)
MW: 128.14

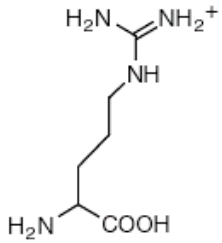
Basic



Histidine (His, H)
MW: 137.14, pK_a = 6.04



Lysine (Lys, K)
MW: 128.17, pK_a = 10.79



Arginine (Arg, R)
MW: 156.19, pK_a = 12.48

Protein primary structure

```
>Anoxybacillus_sp.SK3_4_amylase
  1  MKRVFRALLI FVLLLSVTTP ASAKTERAWQ DERIYFIMVD RFNNGNPKND YDVNVHDPKA
 61 YHGGDLQGII DKLDYIKEMG FTAIWLTPIF ANEKGGYHGY WIEDFYKVEE HFGTLDDFKR
121 LVKEAHKRDM KVILDFVVNH TGYNHPWLND PAKKDWFHEK KDIFNWANQQ EVENGWLFGL
181 PDLAQENPEV KAYLFDVAKW WIKETDVDGY RLDTVKHVPK SFWDEFSKEV KSVKQDFFLL
241 GEVWHDDPRY VAEYGKHGID ALIDFPFYKE ASTIFSNVDQ SLEPLYNVWK RNVTFYDRPY
301 LLGTFLDNHD TVRFTRLALQ NRINPVTRLK LGLTYLFAAP GIPIMYYGTE IALDGGEDPD
361 NRRLMNFRTD KELIDYVTKL GELRETLPSL RRGDFELLYE KDGMALFKRT YGKETTVIAI
421 NNTSKTQKVT LDDELEQGKE LRGLLAGDLV RSKDGKYDII LDRETAEIYV LAPKTGLNIP
481 FIAALLAVYT AFGLFLYFAR KRKASX
```

- Primary structure=amino acid sequence
- The primary sequence is normally directly determined by the sequence of nucleotides in the gene encoding in its primary structure.
- Besides knowing the protein sequence of gene sequencing, small unknown protein sequence can be determined by “puzzle solving” approaches such as Edman Degradation and Mass spectrometers.
- The sequence shown above is a fasta format which starts with a symbol >

Comparing primary structure

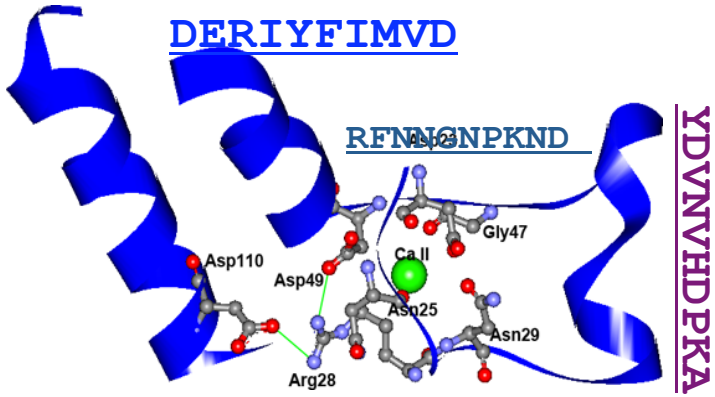
- Expasy database listed a lot of user friendly software that help us to analyze amino acid sequence
- As an example, amino acid sequences from different sources can be compared by sequence alignment using software like ClustalW

```
HSSPALETNPNYVENGAIYDNGTLLGNYSNDQQ-----------NLFHHNGGT-DFSSY 47
HTSPASETNPSYMENGRLYDNGTLIGGYTNDTN-----------SYFHHNGGT-TFSNL 47
HTSPASETDPTYGENGRLYDNGVLLGGYTNDTN-----------GYFHHYGGT-NFSSY 47
HTSPAEVNNPNYAEDGNLYNNGEFVASYSNDLN-----------EIFYHFGGT-DFSTY 47
HSNPA-----TDGEFGALYDNGTLVTNYYEDRKNATRNPYTASLENIYHHNGNINDWFG- 54
HSNPV-----NYGEYGALYDNGTFLTDYFKDTKNAEVNPITGIRENVYHHNGNIYTWSG- 54
HSNPN-----DAGEYGALYDNGTFVIDYPTDANYATVHPITKSLSYIYNHNGGITNWNDR 55
*:.*         * * :*:** :: .*  * :              : * *.   :

EDSIYRNLYDLADYDLN- 64
EDGIYRNLFDLADFNHQ- 64
EDGIYRNLFDLADLDQQ- 64
EDSIYRNLFDLAGLNLN- 64
FQLKYANLFGLADFNQM- 71
IPLKYANLYGLADFNQL- 71
WEVRYKNLFNLADLNQLN 73
    * **:.**. :
```

Protein secondary structure

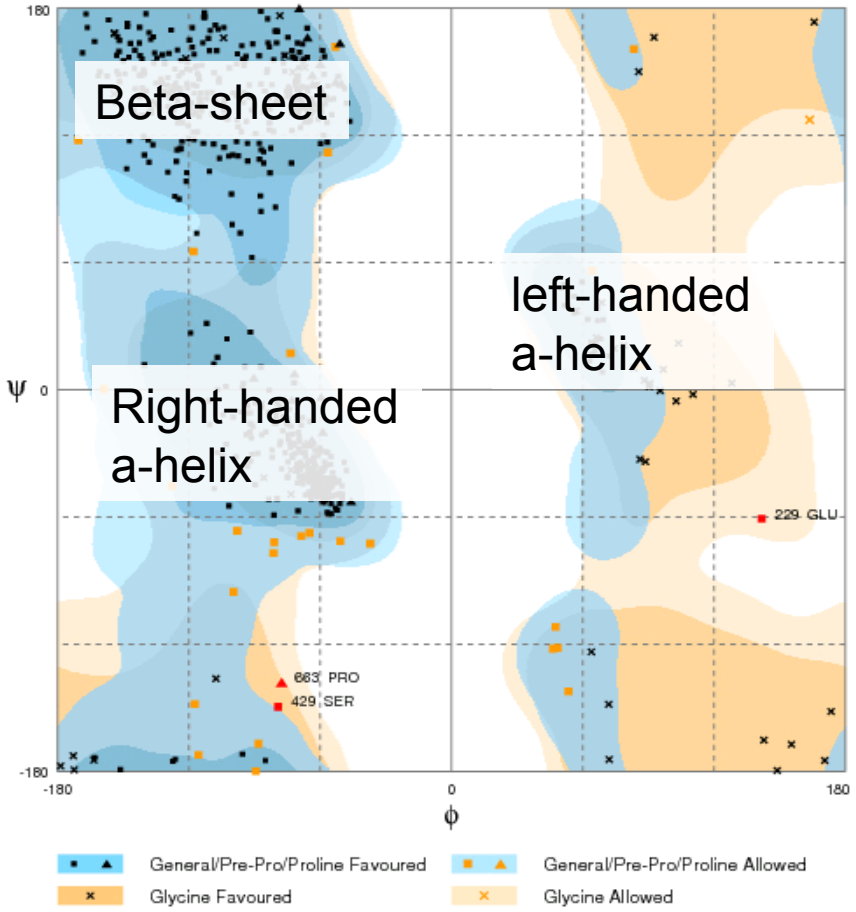
```
>Anoxybacillus_sp.SK3_4_amylase
  1  MKRVFRALLI FVLLLSVTTP ASAKTERAWQ DERIYFIMVD RFNNGNPKND YDVNVHDPKA
 61 YHGGDLQGII DKLDYIKEMG FTAIWLTPIF ANEKGGYHGY WIEDFYKVEE HFGTLDDFKR
121 LVKEAHKRDM KVILDFVVNH TGYNHPWLND PAKKDWFHEK KDIFNWANQQ EVENGWLFGL
181 PDLAQENPEV KAYLFDVAKW WIKETDVDGY RLDTVKHVPK SFWDEFSKEV KSVKQDFFLL
241 GEVWHDDPRY VAEYGKHGID ALIDFPFYKE ASTIFSNVDQ SLEPLYNVWK RNVTFYDRPY
301 LLGTFLDNHD TVRFTRLALQ NRINPVTRLK LGLTYLFAAP GIPIMYYGTE IALDGGEDPD
361 NRRLMNFRTD KELIDYVTKL GELRETLPSL RRGDFELLYE KDGMALFKRT YGKETTVIAI
421 NNTSKTQKVT LDDELEQGKE LRGLLAGDLV RSKDGKYDII LDRETAEIYV LAPKTGLNIP
481 FIAALLAVYT AFGLFLYFAR KRKASX
```



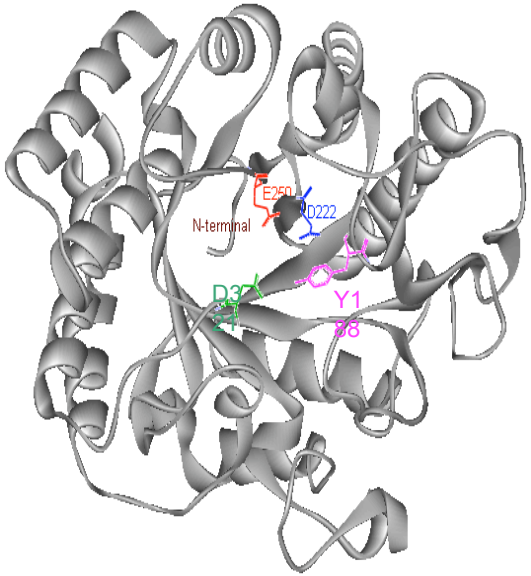
- Stretch of amino acids will fold into two main secondary structure: alpa-helice and beta-sheet
- Stretch of amino acids with no fixed structure is named as loops, turns, or coils

Ramachandran plot

Since the peptide units are rigid groups, the only degrees of freedom they have are rotations around these bonds (N-C_α bond phi Φ; C_α-C bond psi ψ)



Protein tertiary structure



The α-helices, β-strands and loops are linked and folded into a globular form of protein

3-D structure is very important in understanding function of protein

Two major methods to determine 3D structure are

- X-ray crystallography, NMR (Nuclear magnetic resonance)
- Besides that, theoretical structure can be computational model by homology modeling or *Ab initio*- or *de-novo* protein modelling

- Challenges in preparing protein crystallography:

i. Require homogeneous protein in high quantity and purity

ii. High quality crystal difficult to achieve (size, shape, stability)

iii. Crystal does not scatter X-ray well

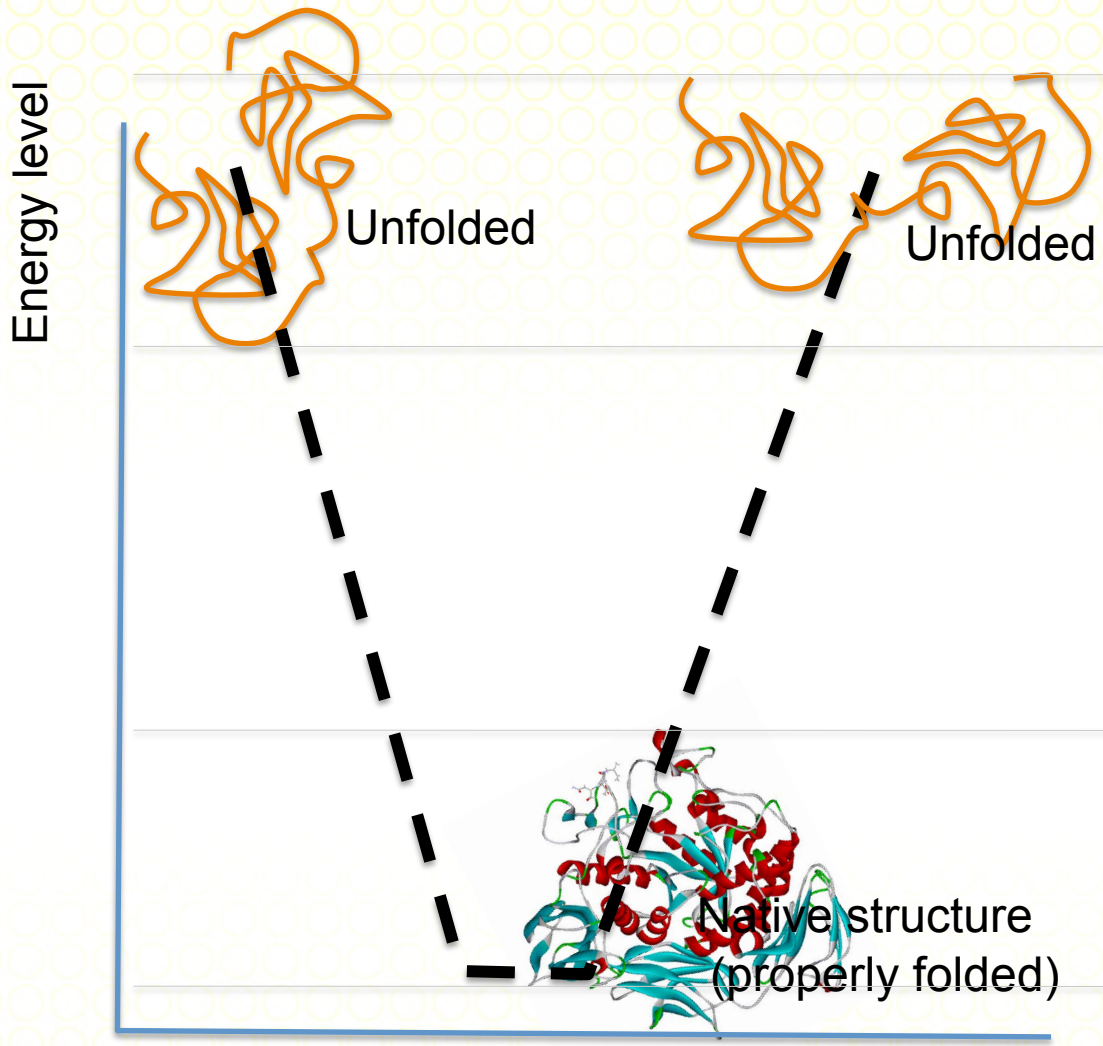
QUATERNARY STRUCTURE

- This is the 4th level of protein structure.
- Some protein hybrid together to form dimer.
- Examples of quaternary structure:

Monomer homodimer heterodimer

- Trimer (3monomers), tetramer (4), pentamer (5),…

Protein Folding



Total unfolded state

Folded into certain secondary structure. Flexibility is high

Low energy structure leads to stable native structure

Tendencies of protein folding

- Hydrophobic effect: Side chain of polar amino acids will tends to clump together *(just like the way the oil droplets do when dispersed in water)*
- This cause the polypeptide to become compact
- Benefits:

  (1) it minimize the total hydrophobic surface area in contact with water

  (2) it brings the polarizable hydrophobic groups close to each other, allowing van der Waals interaction between them.

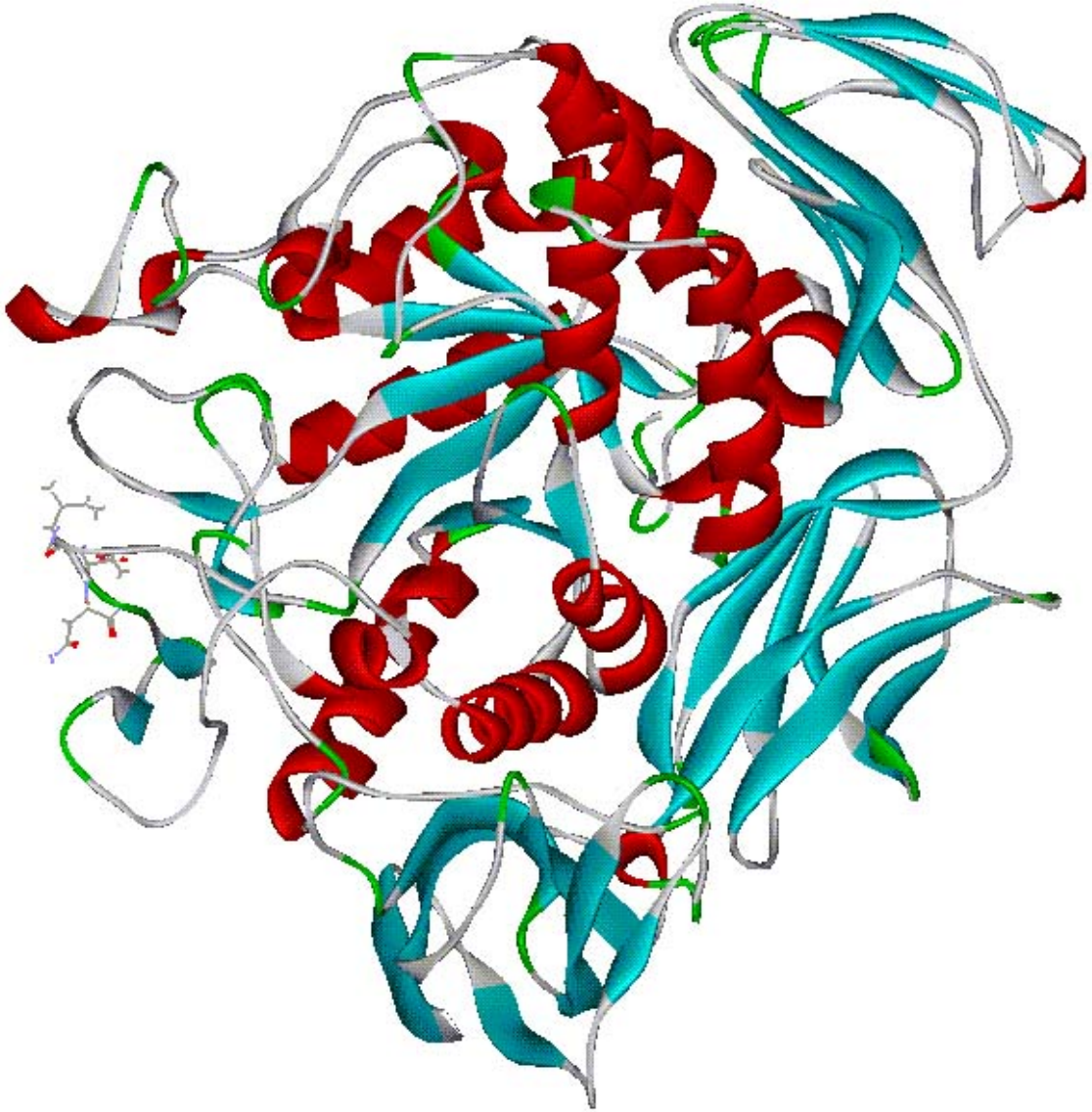
Tendencies of protein folding—cont.

- Although most hydrophobic side chains are buried, some are found on the surface.
- Sometime, when hydrophobic side chains cluster on the surface, they are usually part of a specific binding site for other molecules, or form a patch of mutually interacting nonpolar groups

Loops in protein

- Most proteins are formed by helices or strains, but some amino acids do not adopt regular backbone conformation.
- Such loops are found at the surface of proteins and typically protrude into the solvent.
- Loops provide convenient sites for protein recognition, ligand binding and membrane interaction

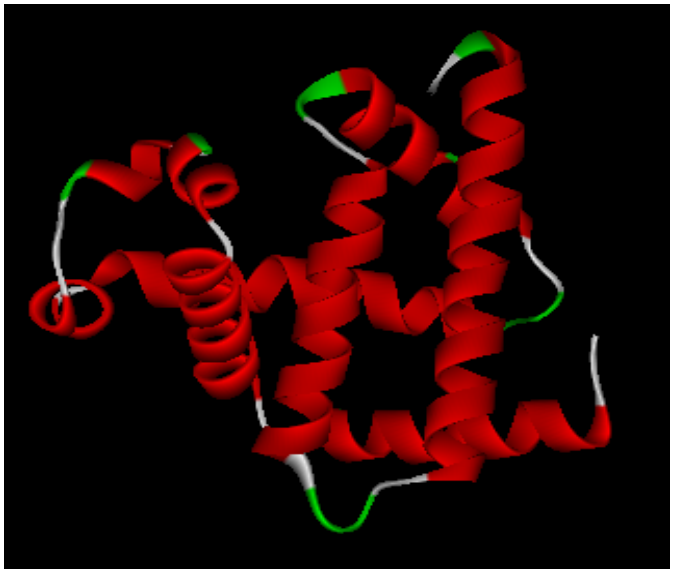
Where is loop?

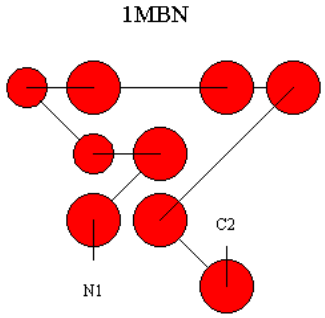


Interactions in a protein – What are the function of these interaction in a protein?

Interaction	Typical distance	Free energy
Covalent bond	1.5 Å	610 kJ/mole for a C=C bond
Disulfide bond	2.2 Å	157 kJ/mole
Ionic interaction/ electrostatic interaction	5-8 Å	Depends on distance and environment
Salt-bridge interaction	< 4 Å	12.5- 17 kJ/mole
Hydrogen bond	3.0 Å	2-6 kJ/mole
Van der Walls interaction	3.5 Å	4 kJ/mole

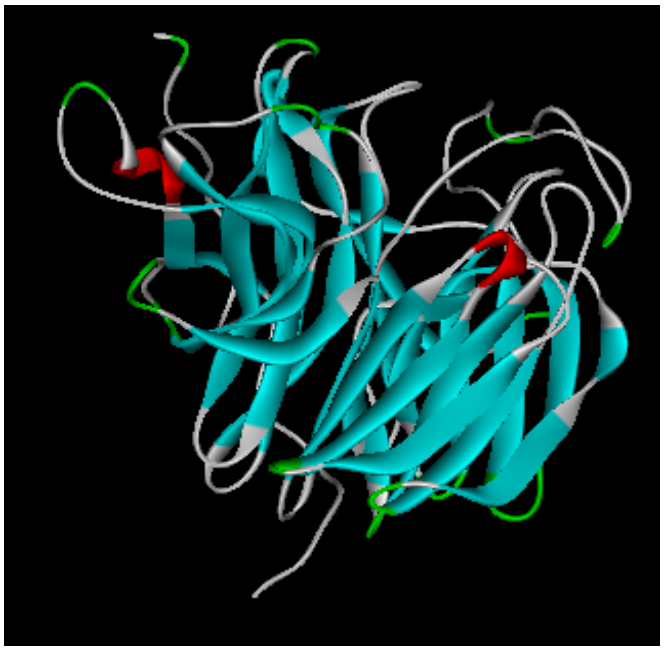
TYPES OF STRUCTURAL ARCHITECTURES

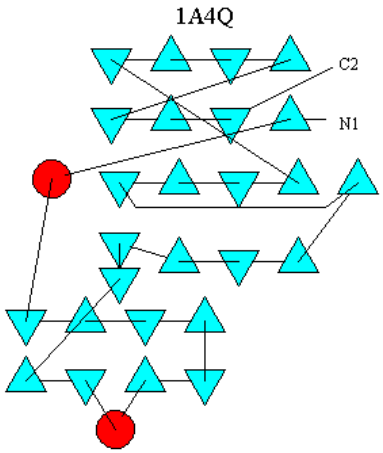




Alpha-domain

- Myoglobulin, 1MBN
- found in muscle cells and gives meat its red color
- Store oxygen in muscle

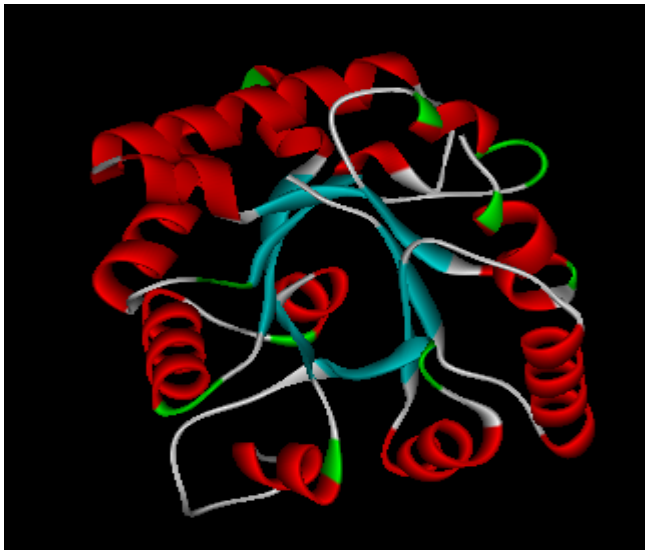


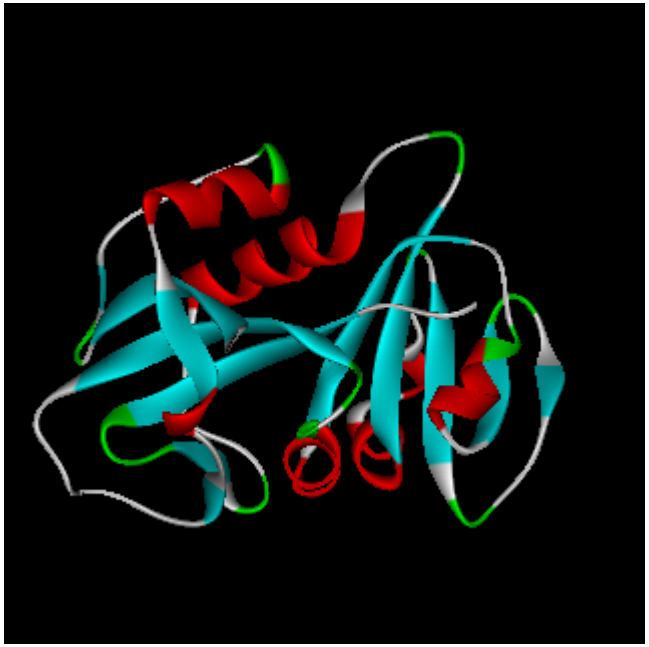


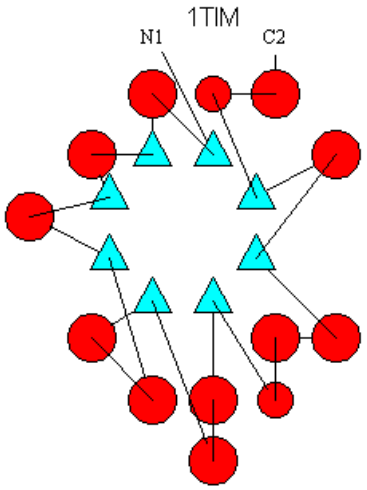
Beta-domain

- Neuraminidase (sialidase), 1A4Q
- Cleave sialic acid from glycoproteins and oligosaccharides.

TYPES OF STRUCTURAL ARCHITECTURES

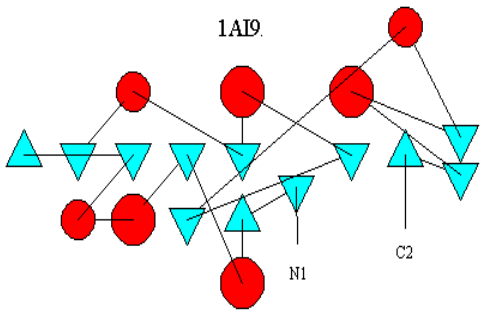






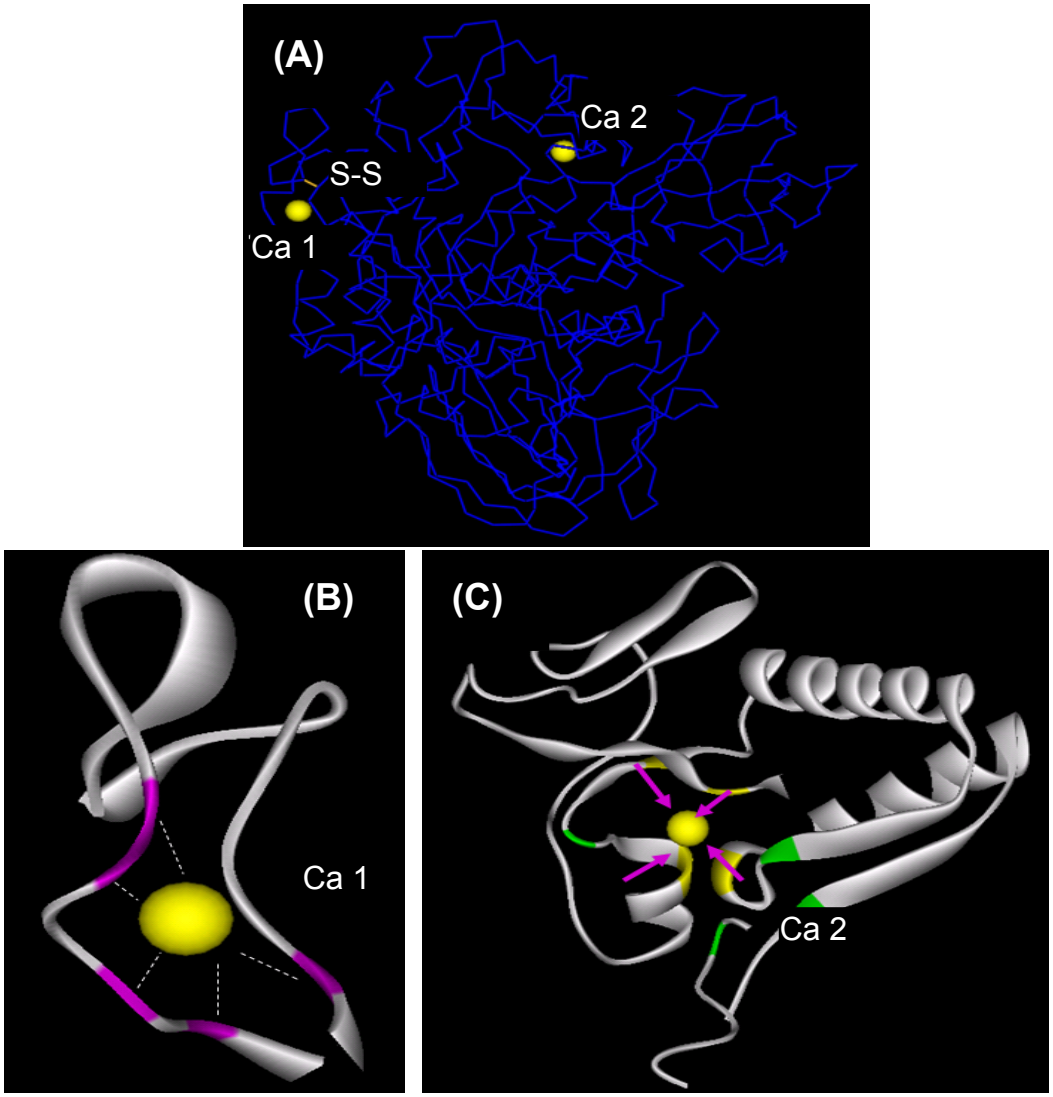
Alpha/beta, Alpha+beta - domain

- Triosephosphate Isomerase, 1TIM
- catalyzed the reversible interconversion of triosephosphates isomers dihydroxyacetone phosphate and D-glyceraldehyde 3-phosphate



- Dihydrofolate reductase/ DHFR, 1AI9
- reduces dihydrofolic acid to tetrahydrofolic acid

Location of S-S and calcium binding sites



(A):Two calcium (yellow circle) ions were able to bind to CGTase G1.

(B): Calcium binding site 1 (Ca 1) involved residue 23, 25, 28, 29, 49 of Domain A

(C): Calcium binding site 2 involved residue 132, 183, 192 and 226. Calcium ion in binding site 2 (Ca 2) was located at the interface of Domains A and B of CGTase G1.

Analysis of ionic interactions in CGTase G1

Row No.	Description	G1	1CIU*	1CYG*	1CXI	1CGT	1PAM
1.	Total residue number in primary sequence	674	683	680	686	686	686
2.	Total numbers of Asp (D)+ Glu (E) residues in primary sequence	74	58	70	58	63	63
3.	Total numbers of D+E involved in ionic interactions	50	48	52	48	41	45
4.	Total numbers of Arg (R) +His (H) + Lysine (K) residues in primary sequence	58	54	60	61	60	60
5.	Total numbers of R+H+K residues involved in ionic interactions	47	45	47	51	48	47
6.	Total ionic pairs	78	73	85	76	70	71
7.	Isolated ionic pairs	12	9	7	11	8	10
8.	Networking ionic pairs	66	64	78	65	62	61
9.	Ionic interactions in domain A only	52	45	52	47	44	45
10.	Ionic interactions in domain B only *(number in bracket refers to cross interaction between Domain B with other domains)*	7 *(+3)*	7 *(+5)*	7 *(+6)*	5 *(+4)*	7 *(+4)*	6 *(+4)*
11.	Ionic interactions in Domain C only	2	3	3	3	4	4
12.	Ionic interactions in Domain D only	2	3	1	4	3	2
13.	Ionic interactions in Domain E only	3	5	4	4	3	3
14.	Cross domains interactions	12	10	18	13	9	11
15.	No. of salt bridges	16	18	15	19	21	19
16.	No. of N-O bridge	12	14	18	14	12	21
17.	No. of long pair	50	41	52	43	37	31

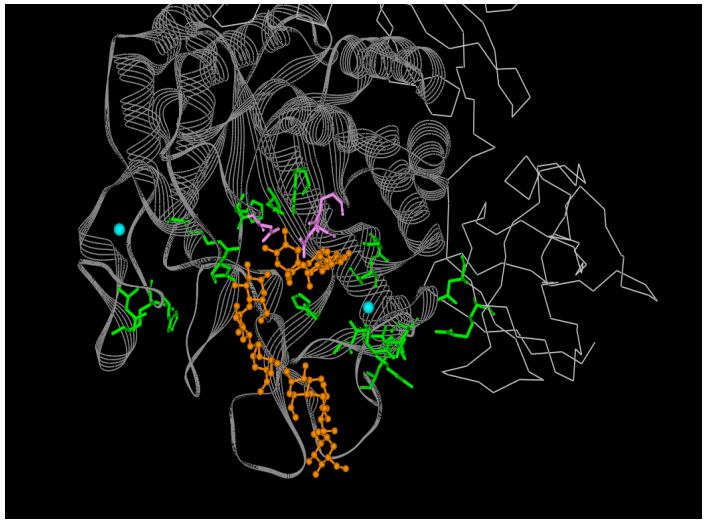
Average:49.7

Average:45.3

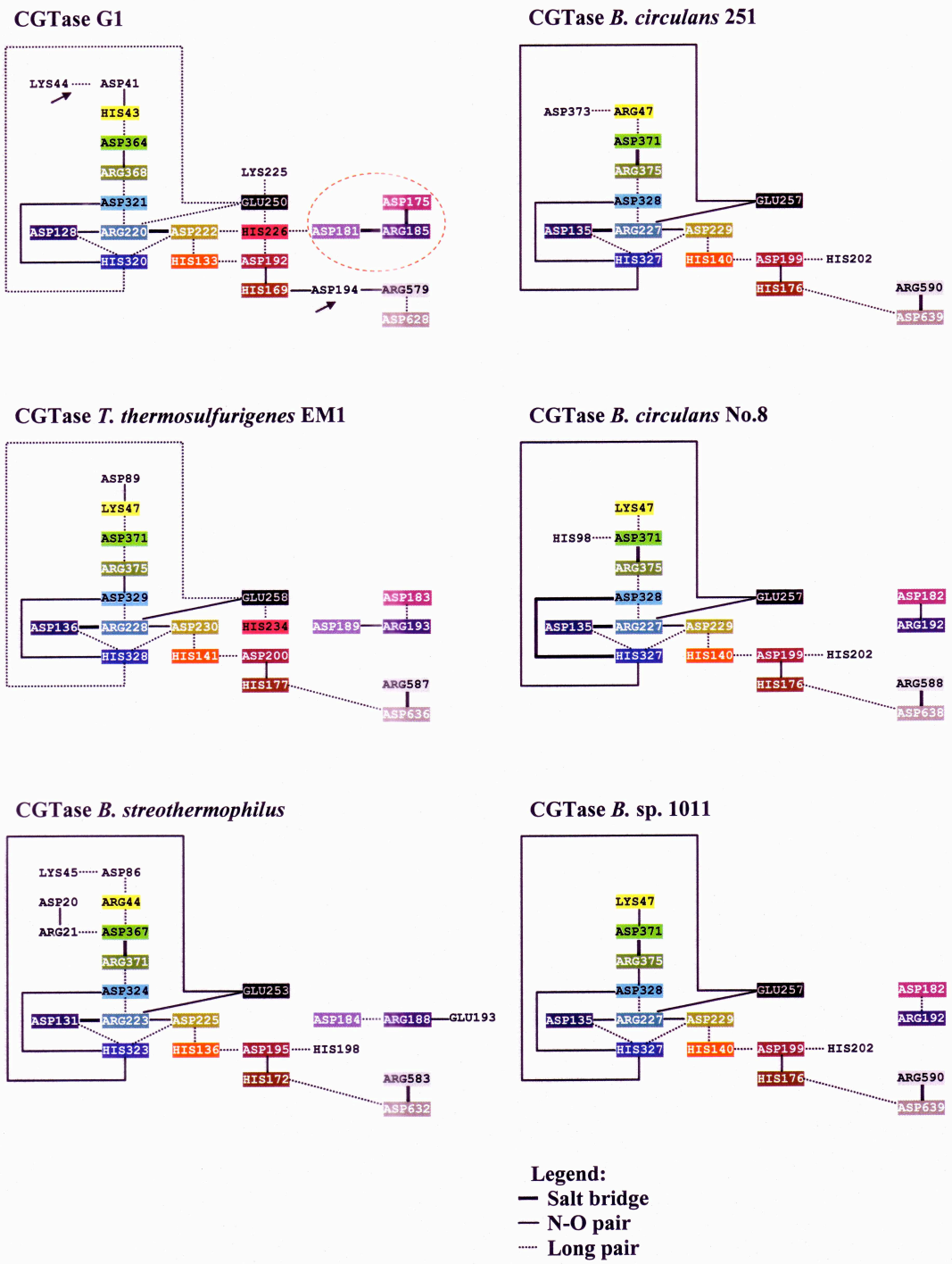
Average:11.7

Average:10

Biggest ionic networking in CGTase



The location of ionic interaction in the protein structure also plays an important factor in deciding the stability of the conformation. The biggest CGTase G1 networking has altogether 22 amino acids. Twenty of the networking residues were actually located in Domain A and B of CGTase G1. The networking covers a huge area (from left to right) and is placed surrounding the active site cleft. The ionic interactions grab hold of a few numbers of secondary structure (α-helice, β-sheet, and loop) strands and create a natural protection for the TIM- barrel structure against heat.



Wait a minute!

- Do you think engineers are interested in all these theories?

BRAINSTORMING

What are the concerns industries are interested of when we talk about enzyme?

Issues	What's the big deal?	Any suggestion to overcome it?
Cost		
Efficiency		