

Keynote Address

The Evolution of Protein Structure Determination: From Column Chromatography to Homology Modeling

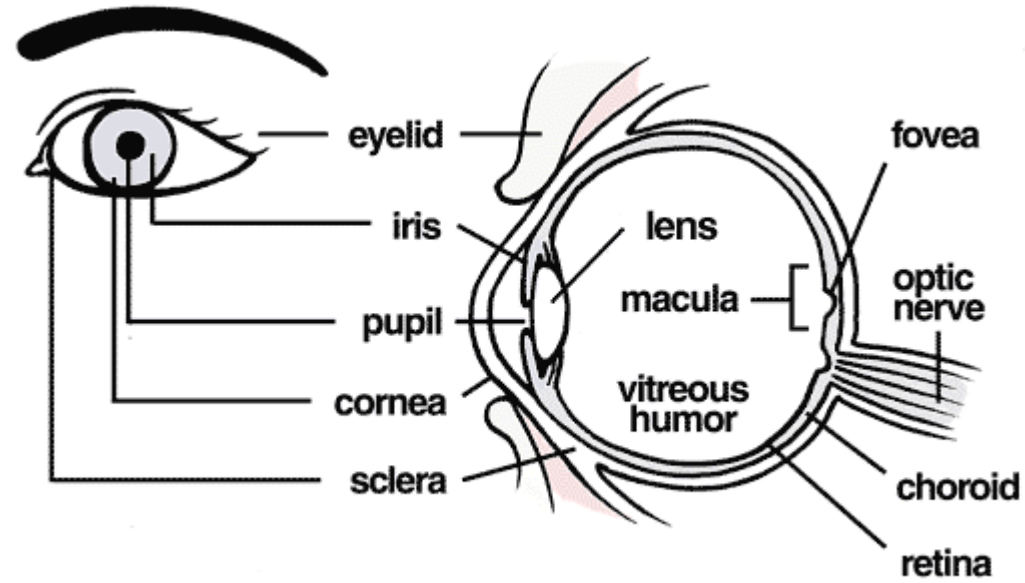
Prof. Hassan M. Khachfe, Ph.D.

Lebanese International University

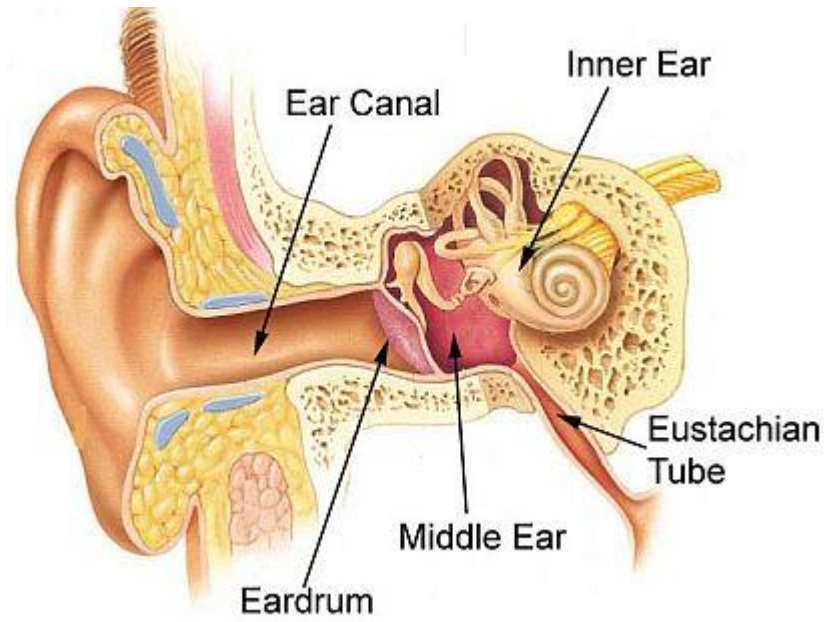


NexTech 2015 International Conference, 19 - 23 July, 2015, Nice, France

Proteins in Action



www.nvrf.org - illustration based upon information from National Eye Institute / National Institutes of Health



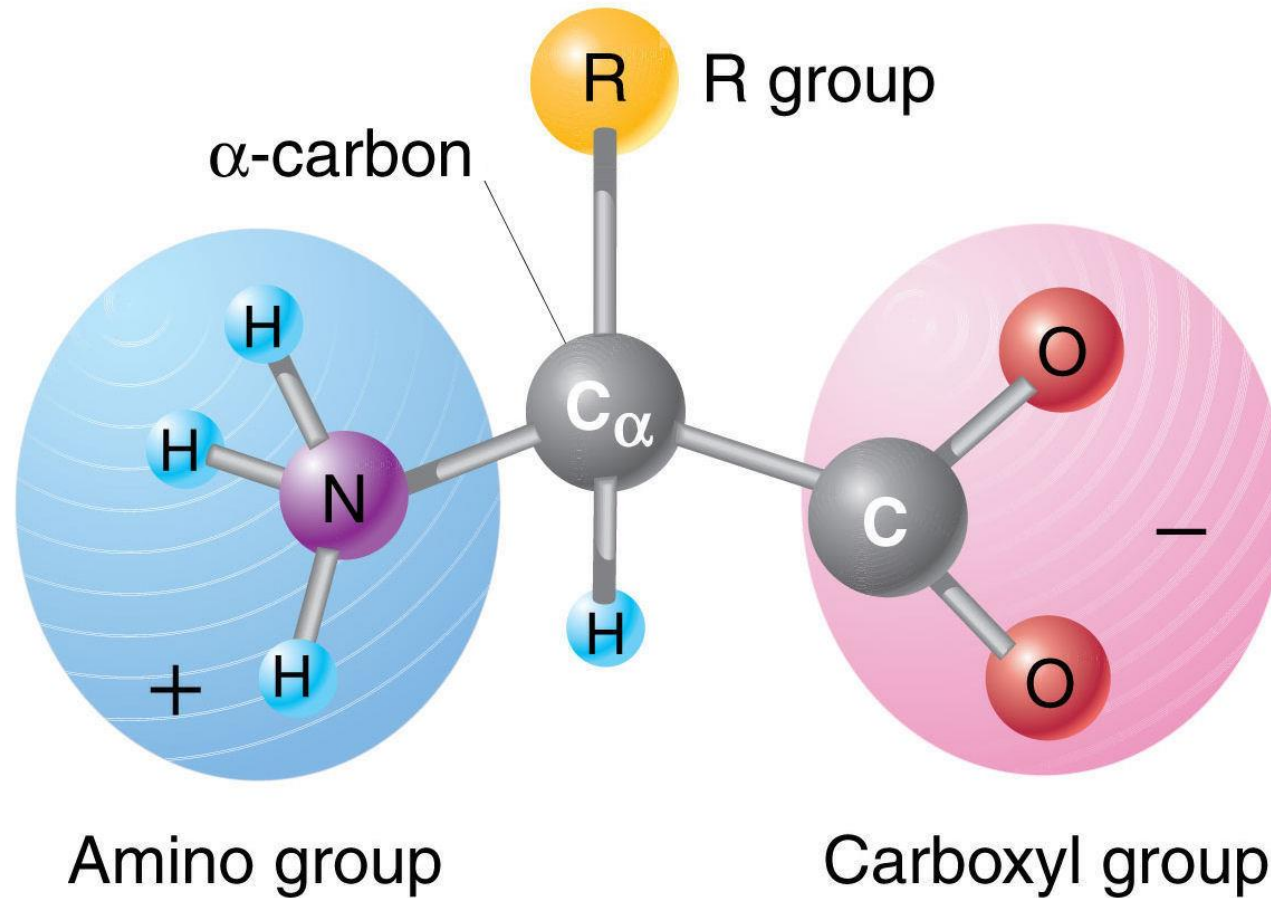
Proteins in Action



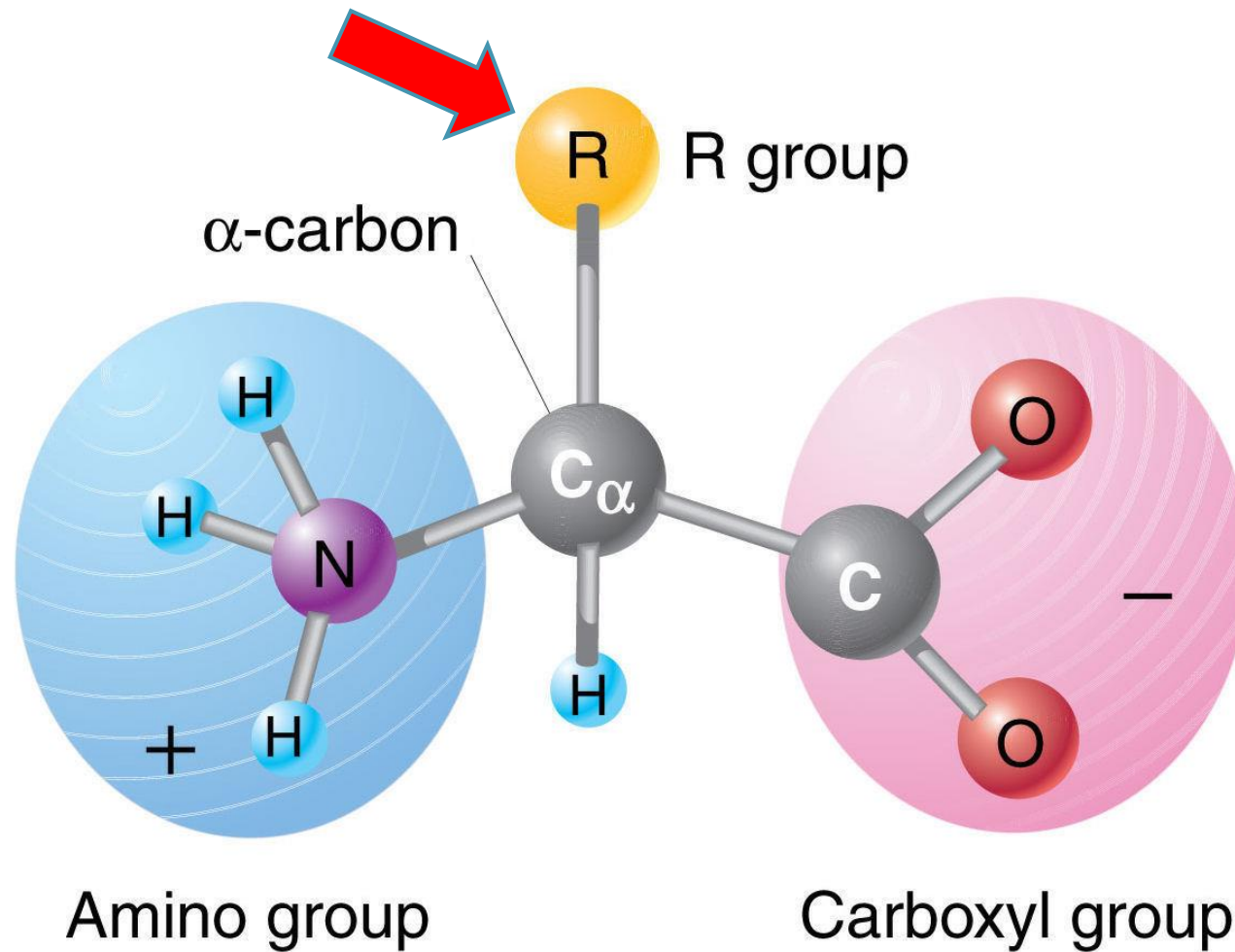
Definition

Macromolecular assemblies composed of basic units (amino acids) connected to each other in a regular format (peptide bond), occupying a specific 3-D shape (fold or structure), which conveys the intended (or faulty) function

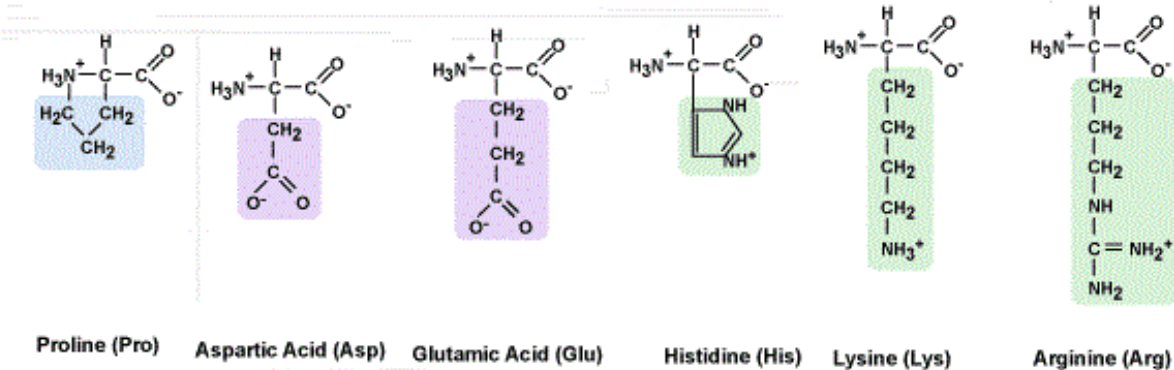
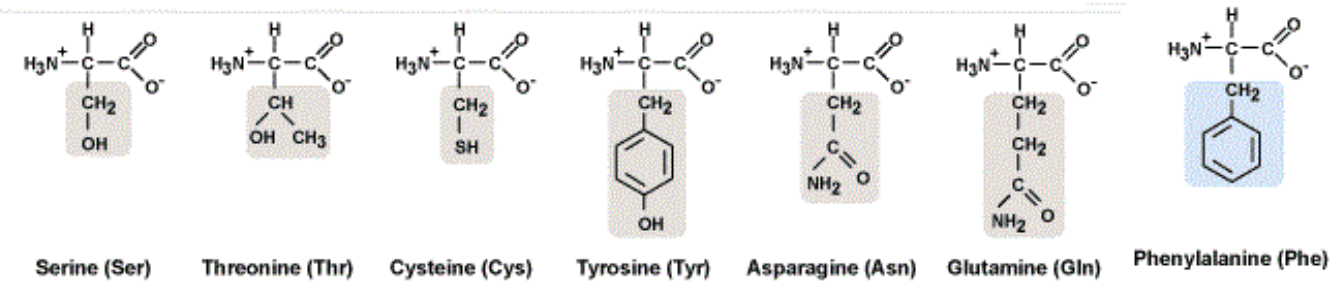
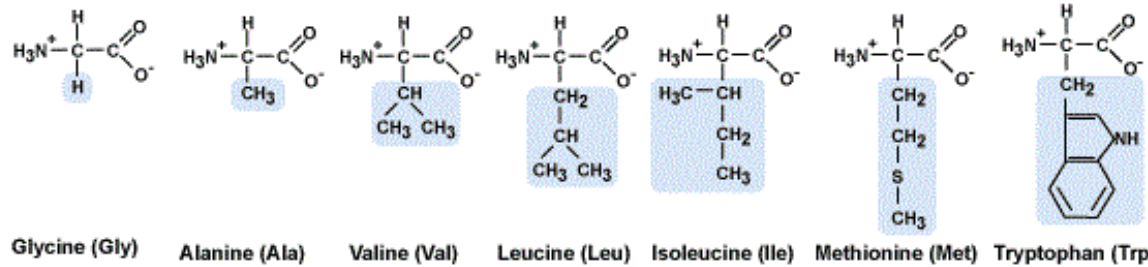
BIOC001 - Biochemistry for Pedestrians



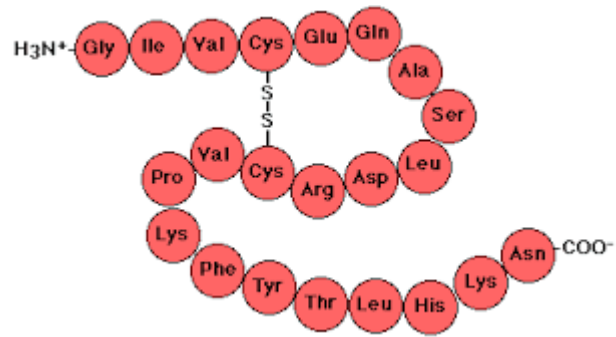
BIOC001 - Biochemistry for Pedestrians



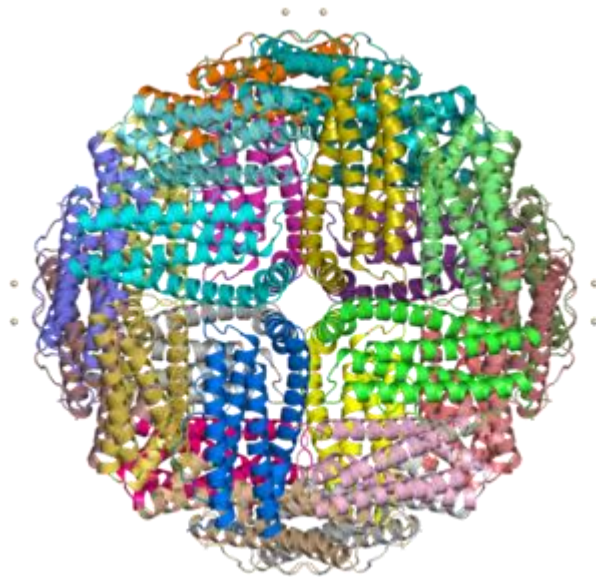
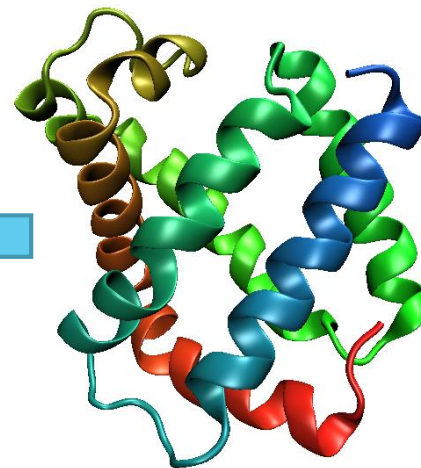
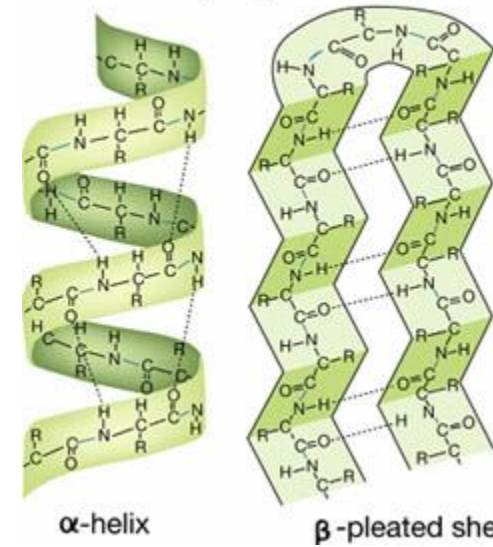
BIOC001 - Biochemistry for Pedestrians



BIOC001 - Biochemistry for Pedestrians



Secondary structure is the result of hydrogen bonding

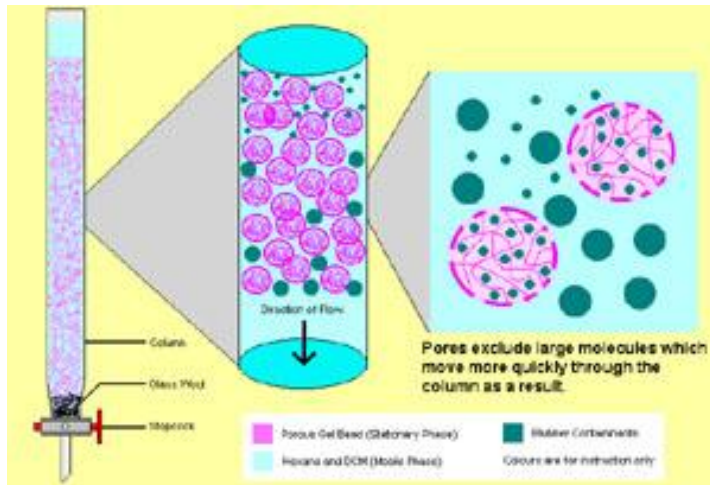


What went wrong?



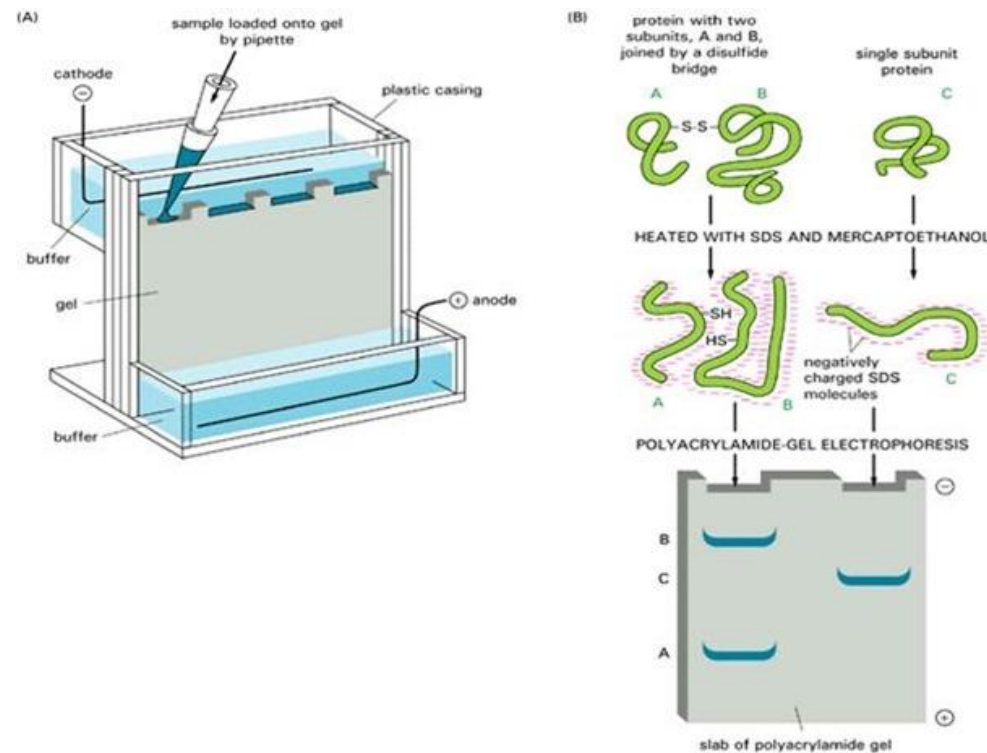
Function ↔ Structure

Initial Methods - very limited

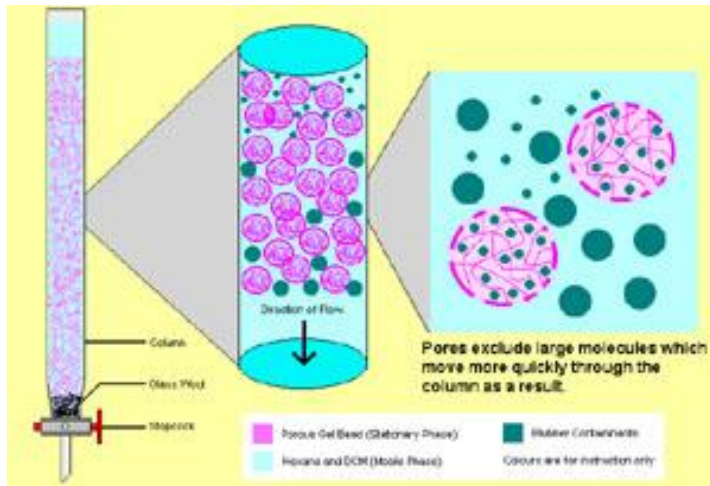


Gel Chromatography

SDS PAGE



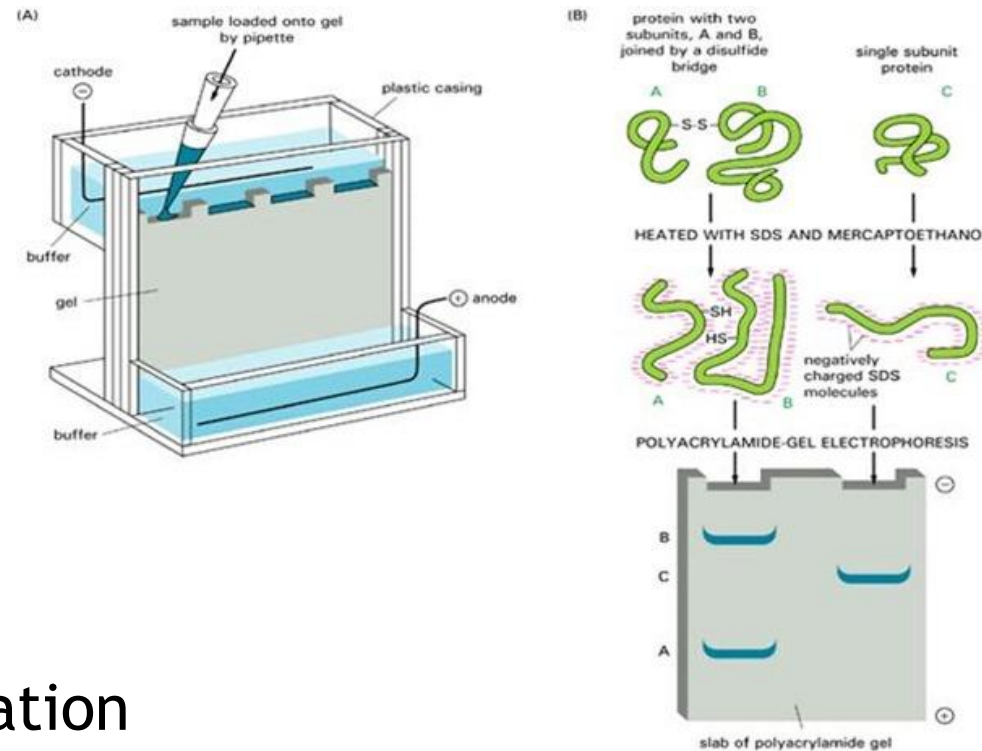
Initial Methods - very limited



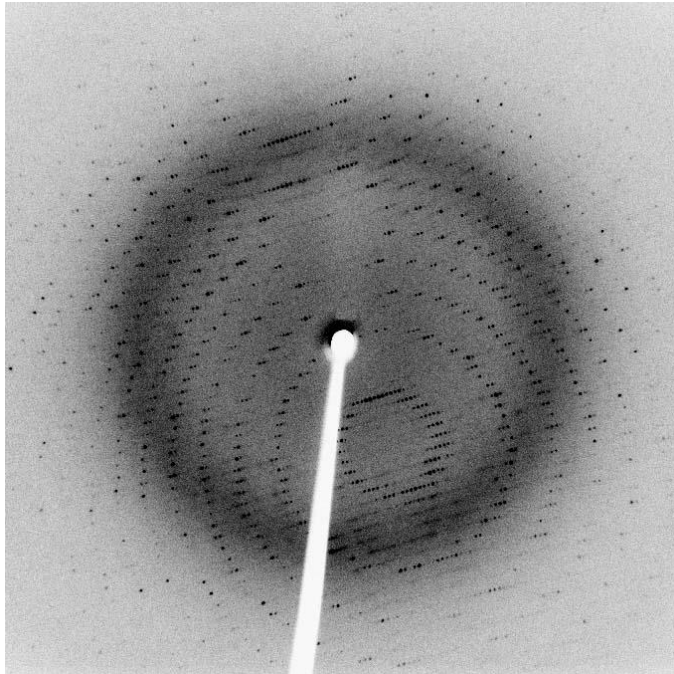
Gel Chromatography

Only SIZE information

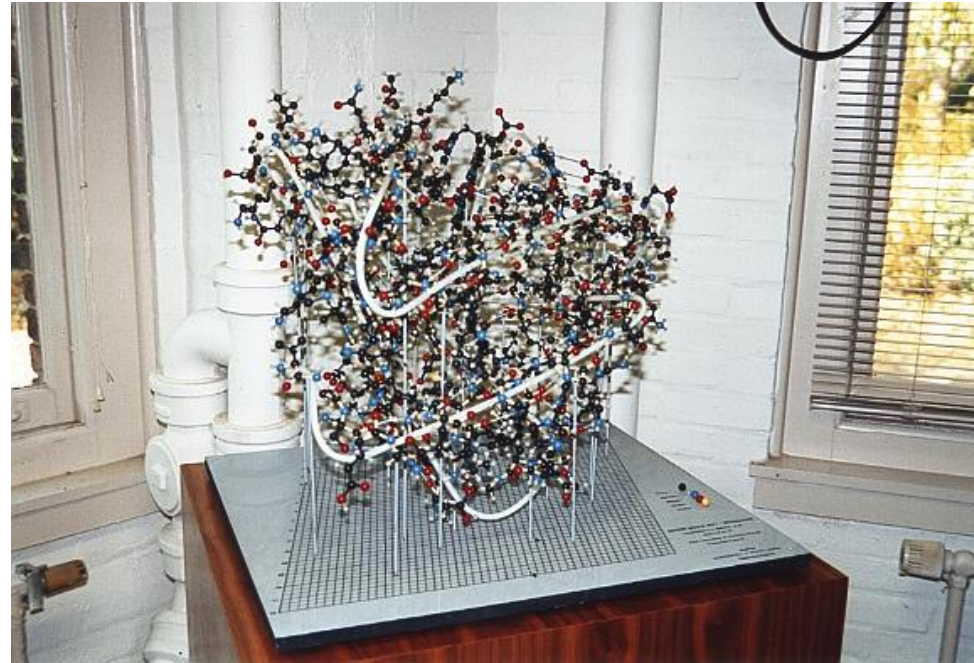
SDS PAGE



The Revolution - 1958



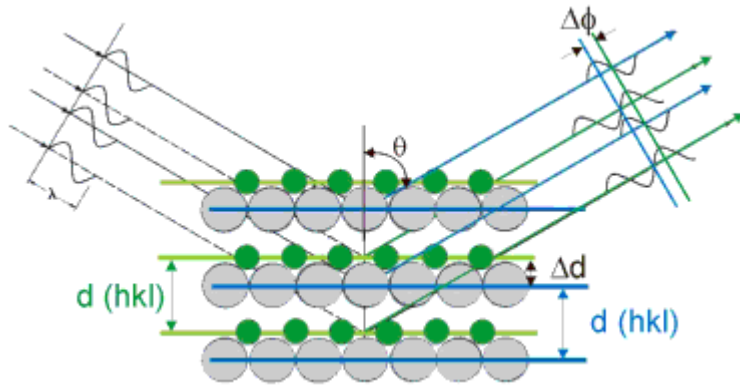
X-ray crystallography



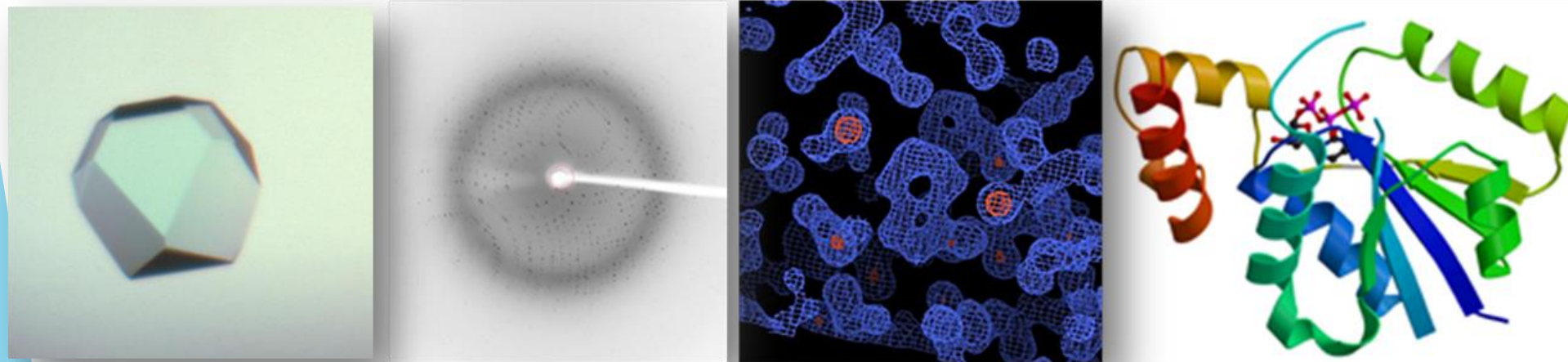
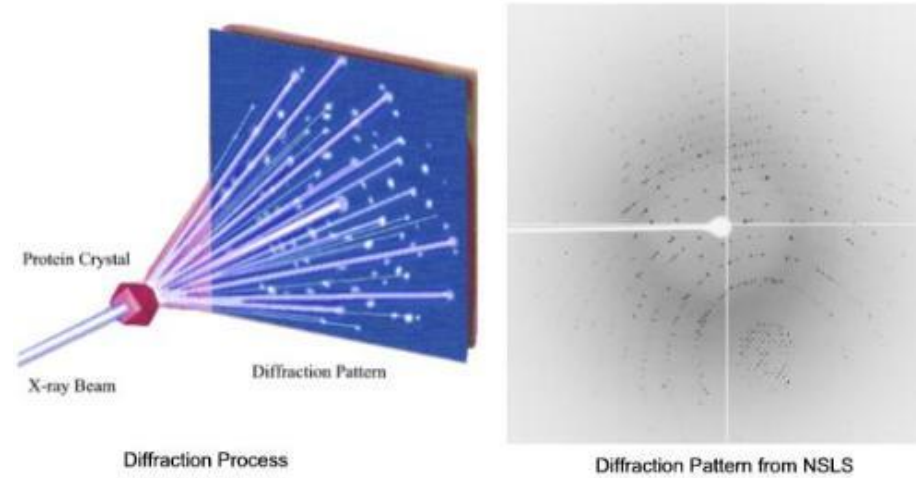
Structure of Myoglobin

1962 - Nobel Prize in Chemistry (Kendrew + Perutz)

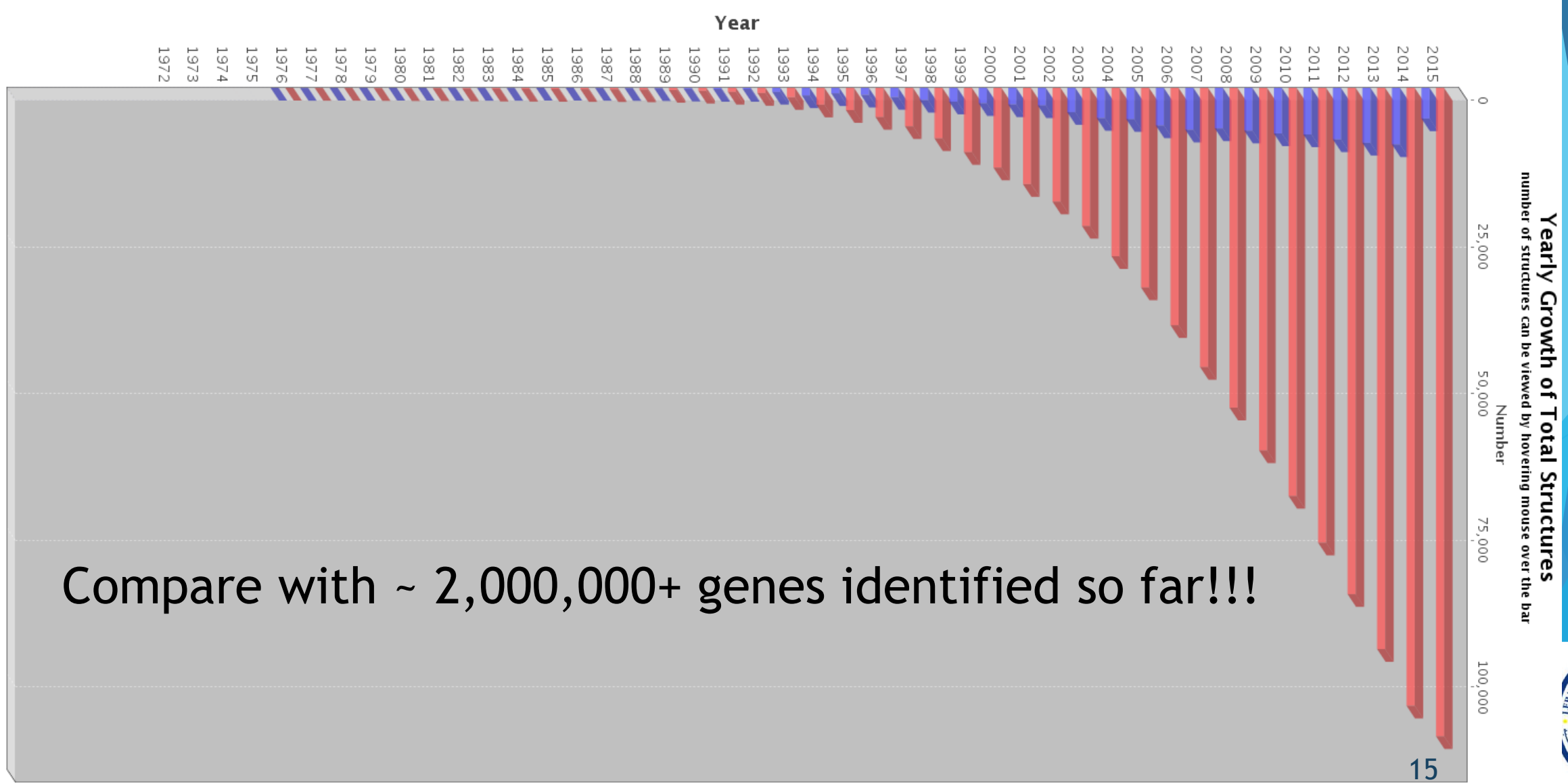
BIOP001 - Biophysics for the Layman



Bragg's Diffraction



Too many sequences, too few structures



Compare with ~ 2,000,000+ genes identified so far!!!



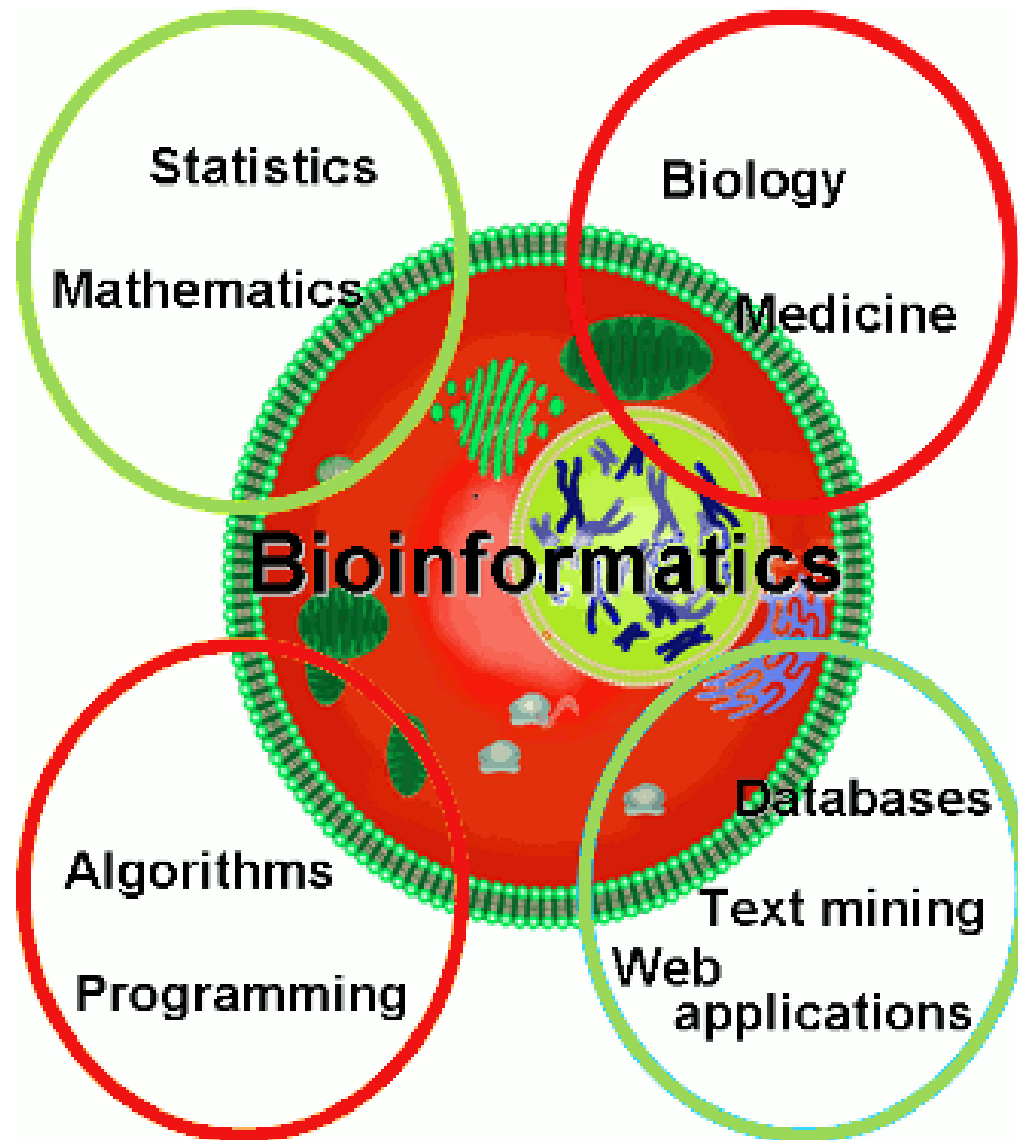
Reasons...technical:

- ▶ X-ray crystallography
 - ▶ Crystals are ordered systems -> defy the laws of nature
- ▶ Structural Electron Microscopy (strEM) - **1983**: Jacque Dubochet
 - ▶ Liquid crystals
- ▶ Structural Nuclear Magnetic Resonance (strNMR) - **1978**: Kurth Wüthrich
 - ▶ Size limitation
- ▶ ..
- ▶ ..
- ▶ Bioinformatics
 - ▶ Still in its infancy

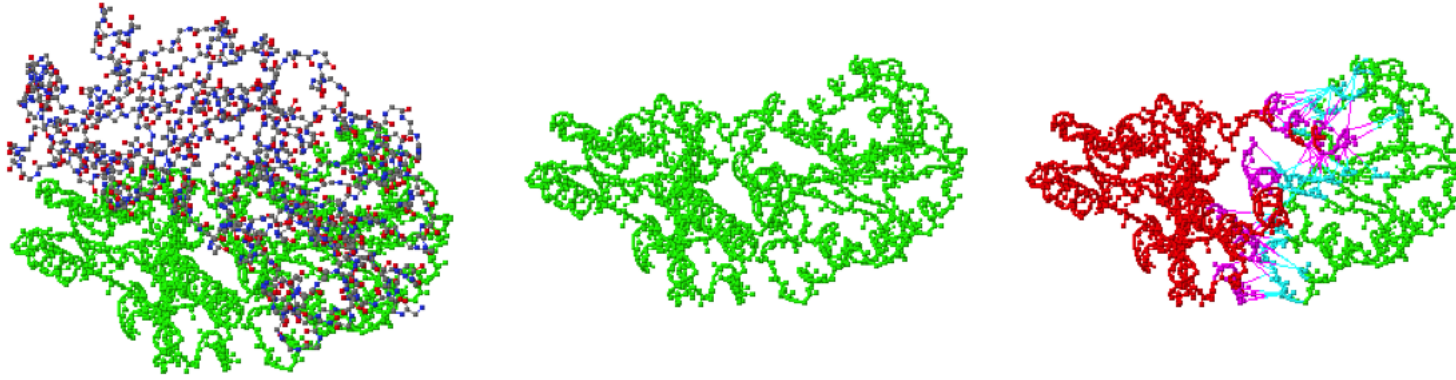
Bioinformatics

- ▶ **1976:** Robert Langridge, credited for developing the first programs to visualize protein structures on a computer screen
- ▶ Advancement in the computer and network technology
- ▶ Success stories in other fields

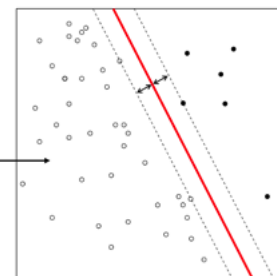
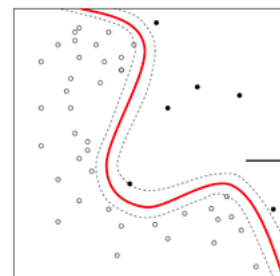
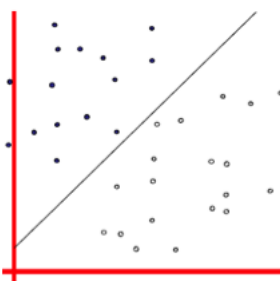
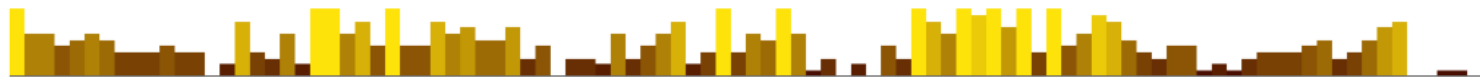
Bioinformatics



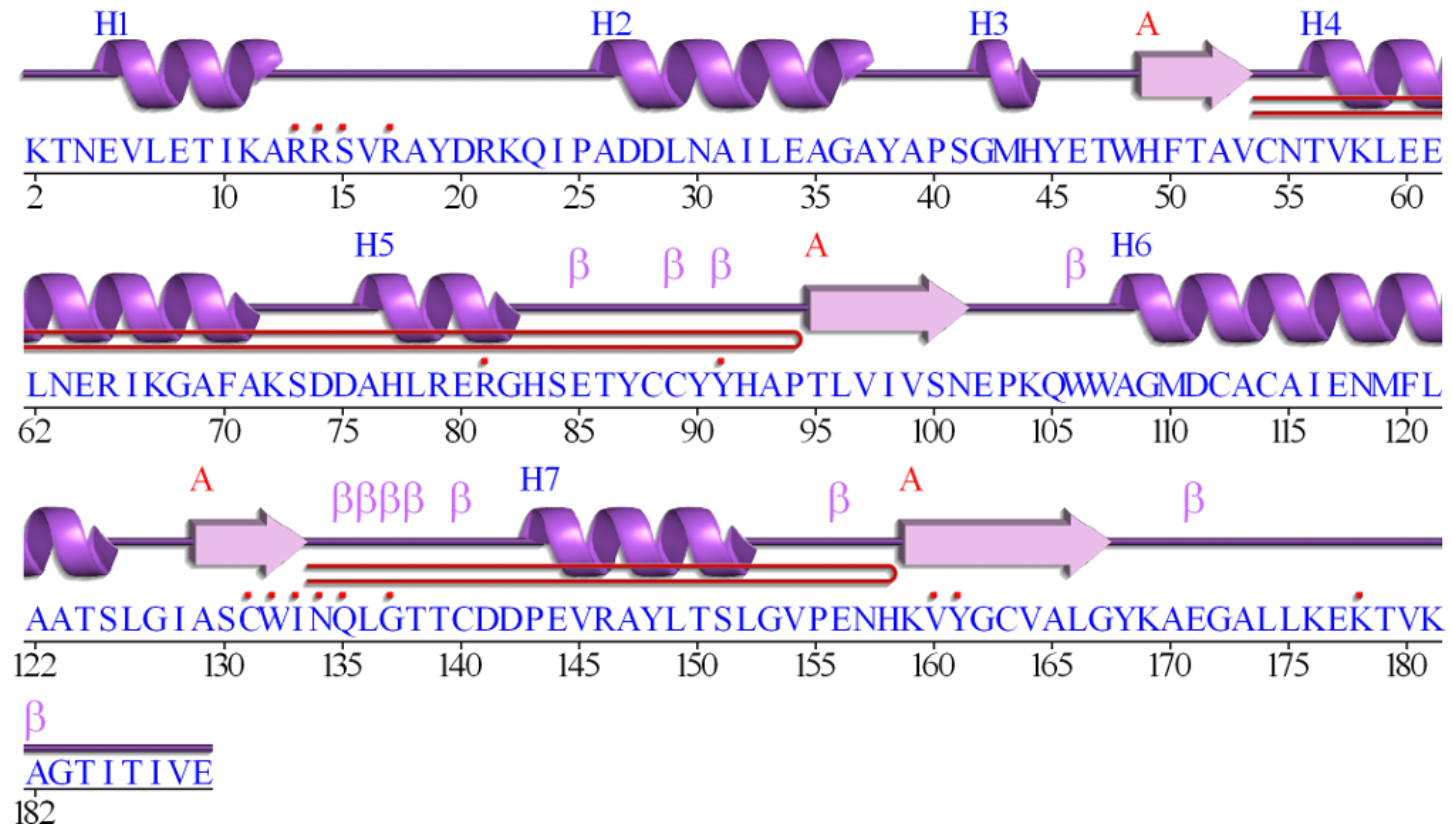
Bioinformatics



10 20 30 40 50 60 70 80 90
 P I I S F L R E I I Q T G D E V E K I E G I F S G T L S Y I F N E F S T S D V K F S D V V K V A K K L G Y T E P D P R D D L N G L D V A R K V T I V G R I S G V E V E S P T S F P V Q S L I P K P L E
 P I I S F L R E I I Q T G D E V E K I E G I F S G T L S Y I F N E F S T S D V K F S D V V K V A K K L G Y T E P D P R D D L N G L D V A R K V T I V G R I S G V E V E S P T S F P V Q S L I P K P L E
 P I I S F L R E I I Q T G D E V E K I E G I F S G T L S Y I F N E F S T S D V K F S D V V K V A K K L G Y T E P D P R D D L N G L D V A R K V T I V G R I S G V E V E S P T S F P V Q S L I P K P L E
 P I I S F L R E I I Q T G D E V E K I E G I F S G T L S Y I F N E F S T S D V K F S D V V K V A K K L G Y T E P D P R D D L N G L D V A R K V T I V G R I S G V E V E S P T S F P V Q S L I P K P L E
 P I I S F L R E I I Q T G D E V E K I E G I F S G T L S Y I F N E F S T S D V K F S D V V K V A K K L G Y T E P D P R D D L N G L D V A R K V T I V G R I S G V E V E S P T S F P V Q S L I P K P L E
 P I I S F L R E I I Q T G D E V E K I E G I F S G T L S Y I F N E F S T S D V K F S D V V K V A K K L G Y T E P D P R D D L N G L D V A R K V T I V G R I S G V E V E S P T S F P V Q S L I P K P L E
 P I I S F L R E I I Q T G D E V E K I E G I F S G T L S Y I F N E F S T S D V K F S D V V K V A K K L G Y T E P D P R D D L N G L D V A R K V T I V G R I S G V E V E S P T S F P V Q S L I P K P L E
 P I I S F L R E I I Q T G D E V E K I E G I F S G T L S Y I F N E F S T S D V K F S D V V K V A K K L G Y T E P D P R D D L N G L D V A R K V T I V G R I S G V E V E S P T S F P V Q S L I P K P L E
 P I I S F L R E I I Q T G D E V E K I E G I F S G T L S Y I F N E F S T S D V K F S D V V K V A K K L G Y T E P D P R D D L N G L D V A R K V T I V G R I S G V E V E S P T S F P V Q S L I P K P L E
 P I I S F L R E I I Q T G D E V E K I E G I F S G T L S Y I F N E F S T S D V K F S D V V K V A K K L G Y T E P D P R D D L N G L D V A R K V T I V G R I S G V E V E S P T S F P V Q S L I P K P L E
 P I I S F L R E I I Q T G D E V E K I E G I F S G T L S Y I F N E F S T S D V K F S D V V K V A K K L G Y T E P D P R D D L N G L D V A R K V T I V G R I S G V E V E S P T S F P V Q S L I P K P L E
 P A L S F L E T L R G S E . . L L E L H G I L N G T T L Y I L Q E M E K . G R T Y A E A L L E A Q R L G Y A E A D P T L D V E G I D A A H K L T L L A R L L V D P G F F A E V E A D G I A R L T P E
 P A L S F L E T L R G S E . . L L E L H G I L N G T T L Y I L Q E M E K . G R T Y A E A L L E A Q R L G Y A E A D P T L D V E G I D A A H K L T L L A R L L V D P G F F A E V E A D G I A R L T P E
 P A L S F L E T L R G S E . . L L E L H G I L N G T T L Y I L Q E M E K . G R T Y A E A L L E A Q R L G Y A E A D P T L D V E G I D A A H K L T L L A R L L V D P G F F A E V E A D G I A R L T P E
 P A L S F L E T L R G S E . . L L E L H G I L N G T T L Y I L Q E M E K . G R T Y A E A L L E A Q R L G Y A E A D P T L D V E G I D A A H K L T L L A R L L V D P G F F A E V E A D G I A R L T P E
 P L S F I D Y S V L P S . R I K F F R G I V S L T I N Y F I R E L A N . K R E F D D V L S E A T L G I V E K N Y K D D L T G L D A A R K S V I L C N H L Y G S S Y R L S D V F Y E G I L Q . . .
 P V V K L A K R Y L A L C . E I E S V K G I F N G T C N Y I L S R M E E E R L P Y E H I L K E A Q E L G Y A E A D P S Y D V E G I D A A L K L V I A N T I G . V K A S Y E D V E M T G I T Q I T P E



Bioinformatics - Stat at its best



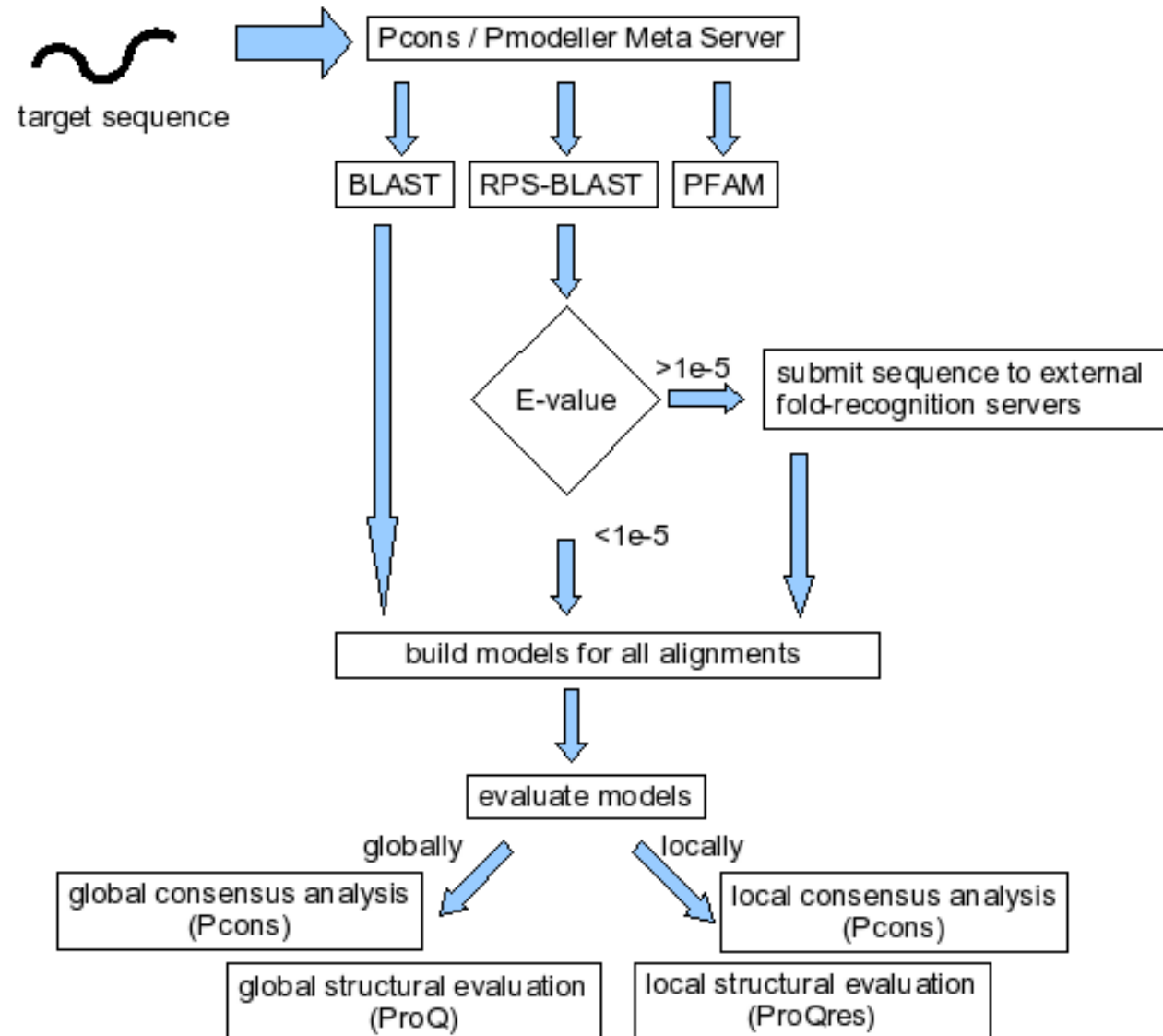
Excellent prediction of Secondary Structural elements



Bioinformatics - 3D Structure?

- ▶ Homology Modeling
- ▶ Ab initio
 - ▶ Molecular Dynamics
 - ▶ Genetic algorithms

Bioinformatics - Homology Modeling



Apo B: 18 RFKHLRKYTYNYEAESSGVPGTADRSATRINCKVELEVPQLCSFILKTSQCTLKEVYGFNPEGKALLK
LipoV: 17 QFQPGKVYRYSYDAFSISGLPEPGVNRAGLSGEMKIEIHGHTHNQATLKITQVNLKYFLGPWP-SDSYFL

Apo B: 89 TKNSEEFAAAMSRVELKLAIPGKQVFLYPEKDEPTYIILNIKRGIISALLVPPETEEAKQVLFLDIVYGN
TAGYDHFIIQQL-EVPEVRETDYSAGRIGDIYAPFQVDTAVNIVRGIILNLFQLSLKKNQQTFFELQETGVEGIC

Apo B: 160 STHFIVKTRKGNVATEISTERDLGQCDFRFPKPIRTGISPLALIKGMTRPLSTLISSSQSCQYTL-DAKRKHV
QITYVVQEGYRTNEMAVVKTKDLNNDH-KVYKIMGTAAYERCPTCQKMNKILRSTAVNYAIFDEPSGYI

Apo B: 230 AEAICKEQHLFLPFSYKNKYGMVAQVITQTLKLEIDTPK-----INSRFFGEGTKKMGLAFES
IKSAHSEEIQQLSVFDIKEGNVVIESRQKLIILEGIQSAPAAASQAASLQNRGGLMYKFPSSAITKMSSLF-V

Apo B: 285 TKSTSPPKQAEAVLKTLOELKLTISEQNIQANLFNKLVTELRLSDEAVTSLLPQLIEVSSPITLQALV
TKGKNLESEIHTVLKHLVENNQLSVHEDAPAK---FLRLTAFLRNVDAGVLQSIWHKLHQQK--YRRWIL

Apo B: 357 QCGQPQCSTHILQWLKRVHANPLLIIDVVITYLVALIPEPSAQQLEIFNMARDQ-----RSRATL-----Y
DAVPAMATSEALLFLKRTLASEQLTSAEATQIVYSTLSNQQATRESLSYARELLHTSFIRNRPILRKTAVL

Apo B: 417 ALSHAVNMYHKTNPFTGTQELLDIANYLMEQIQDDCTGDEDTYLILRVIGNMGQ-----TMEQLTPELKSS
GYGSLVFRYCANVSCPDELLQPLHDLISQSSD--RADEEIVLALKALGNAGQPNSIKKIQRFLPGQKKS

Apo B: 483 IILKCVQSTKPSLMIQKAAIQALRKMEPKD--KDQEVLLQTFLLDDASPGDKRLAAYIMLMRS-PSQADINKI
L-----DEYSTRVQAEAIMALRNIAKRDPRKQEVIVLPIFLNVAIKSELRIRSCIVFFESKPSVALVSMV

Apo B: 550 VQILDFWEQNEQVKNFVASHIANILNSEELDIQDLKKLKVEVLKESQLPTVMDFRKFSRNYQLYKSVSLPSL
AVRLRREPNLQVASFVYSQMRSLSRSSNPEFRDVAACSVAIK--MLGSKLD-RLGCRYSKAVHVDTFNAR

Apo B: 622 DPASAKIEGNLIFDPNNYLPKESMLKTTLTAFGFASADLIEIGLEGKGFEPTEALFGKQGFPPDSVVKAL
TMAGVSADYFRINSPSGPLPRAVAAKIRGQGMGYAS-DIVEFGLRAEGLQELLYRGSQEQDAYGTALDRQT

Apo B: 693 YWVNGQVPDGVSKVLVDHFGYTKDDKHEQDMNGIMLSVEKLIKDLKS--KEVPEARAYLRILGEEELGFAS
LLRSGQARSHVSSI-----HD-----TLRKL-SDWKSVPPEERPLASGYVVKVHGQEVVFAE

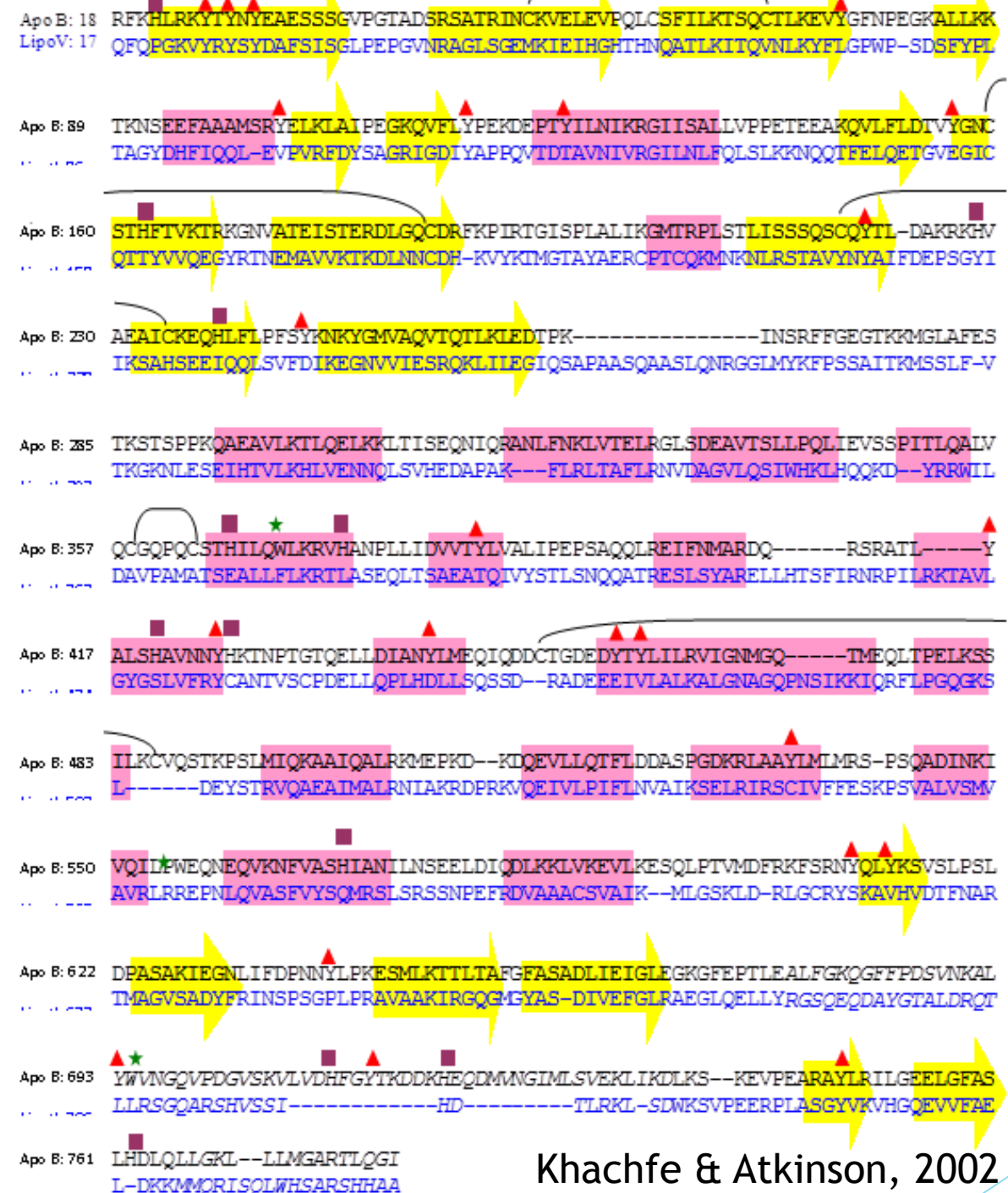
Apo B: 761 LHDLQLLGKL--LLMGARTLQGI
L-DKKMMQRISQLWHSARSHHAA



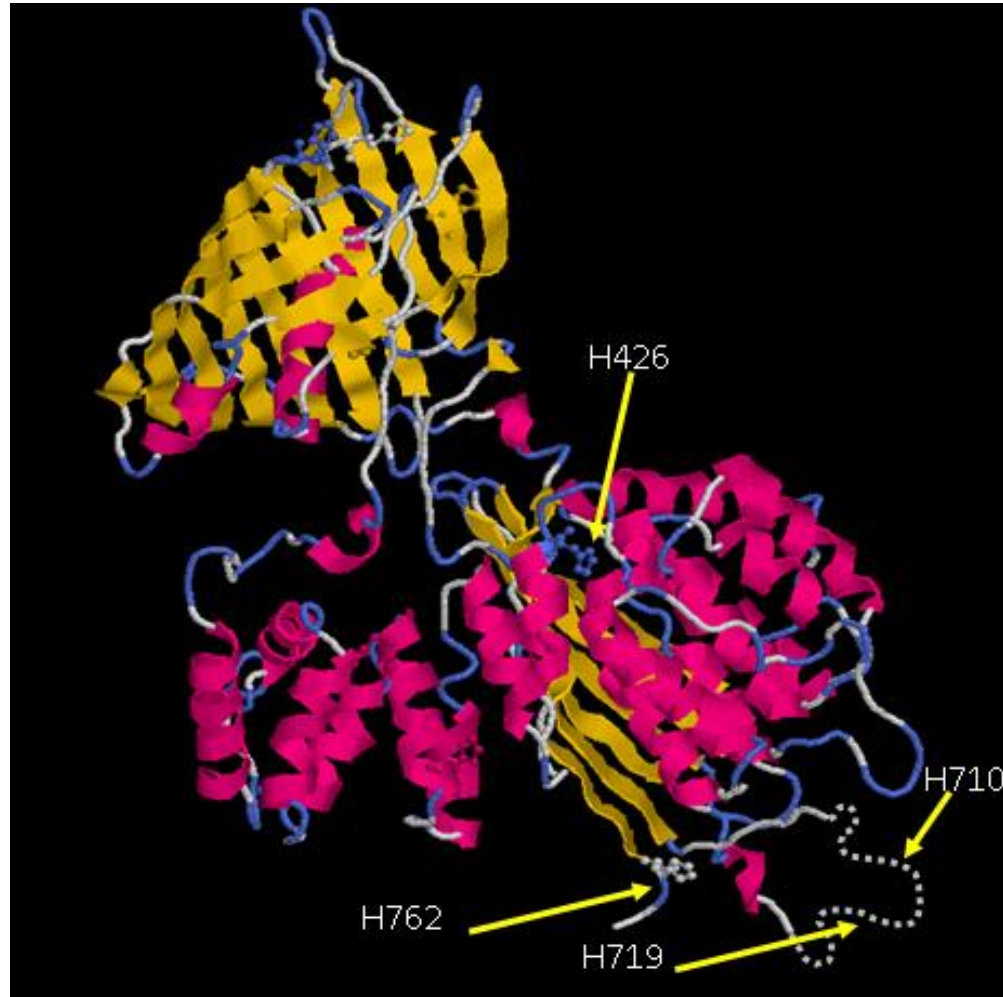
Bioinformatics - Needleman Wunsch alg.

$$S_{ij} = \max \begin{cases} S(a_i, b_j) + S_{i-1, j-1} \\ \delta + S_{i, j-1} \\ \delta + S_{i-1, j} \end{cases}$$

-	-	A	T	C	G	A	C
-	0	-4	-8	-12	-16	-20	-24
C	-4	-3	-7	-3	-7	-11	-15
A	-8	1	-3	-7	-6	-2	-6
T	-12	-3	6	2	-2	-6	-5
A	-16	-7	2	3	-1	3	-1
C	-20	-11	-2	-1	0	-1	8



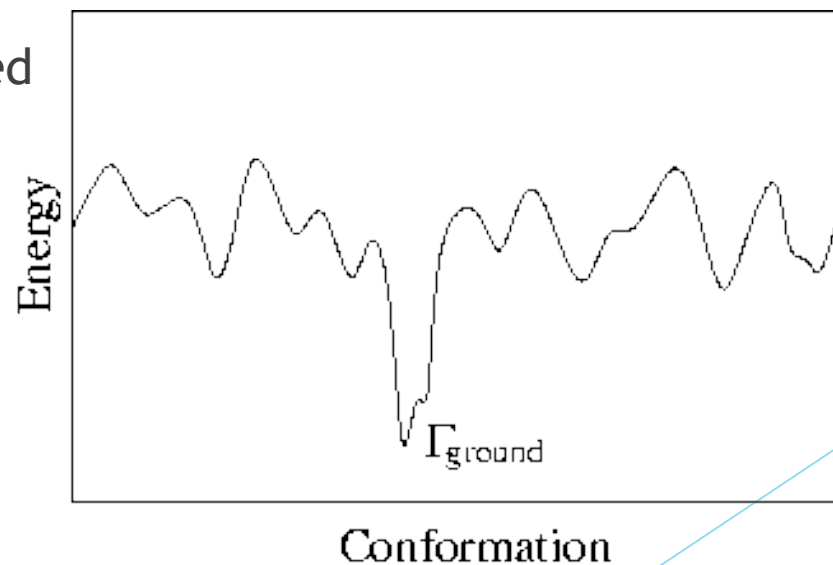
Bioinformatics - Homology Modeling

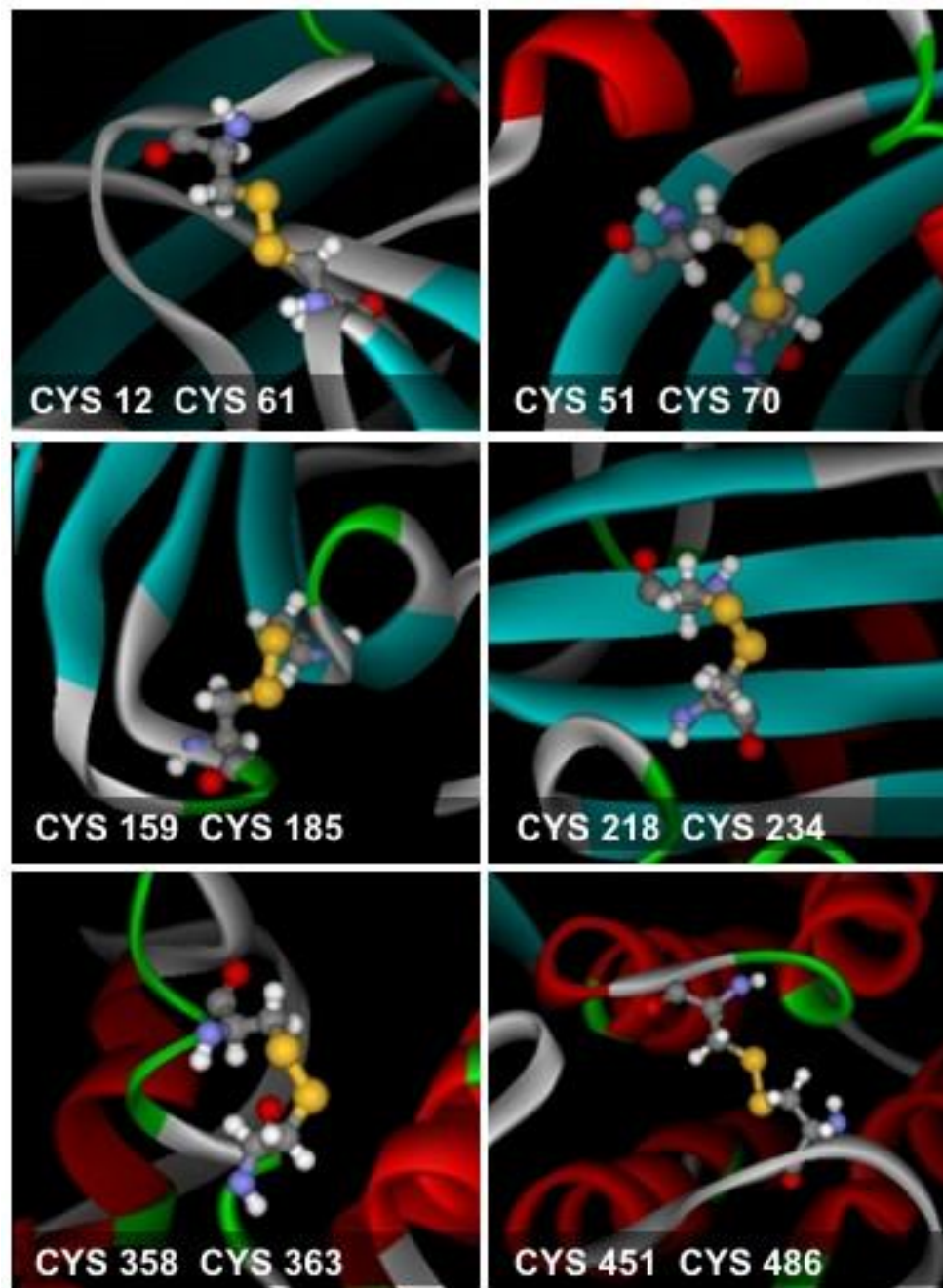


Al-Ali & Khachfe, 2007

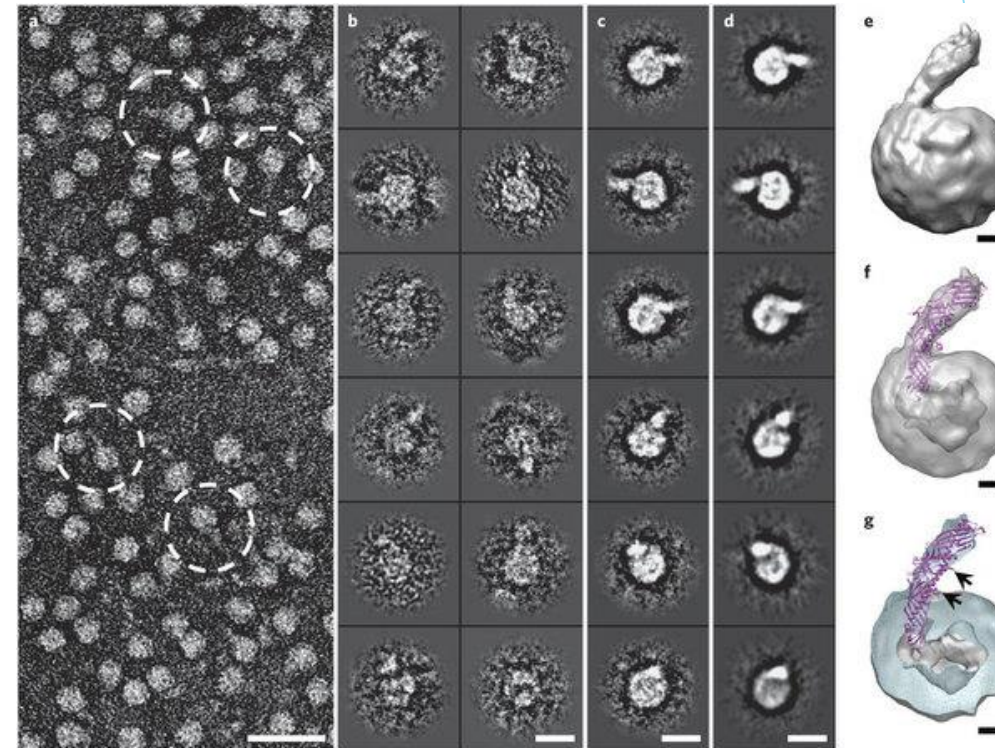
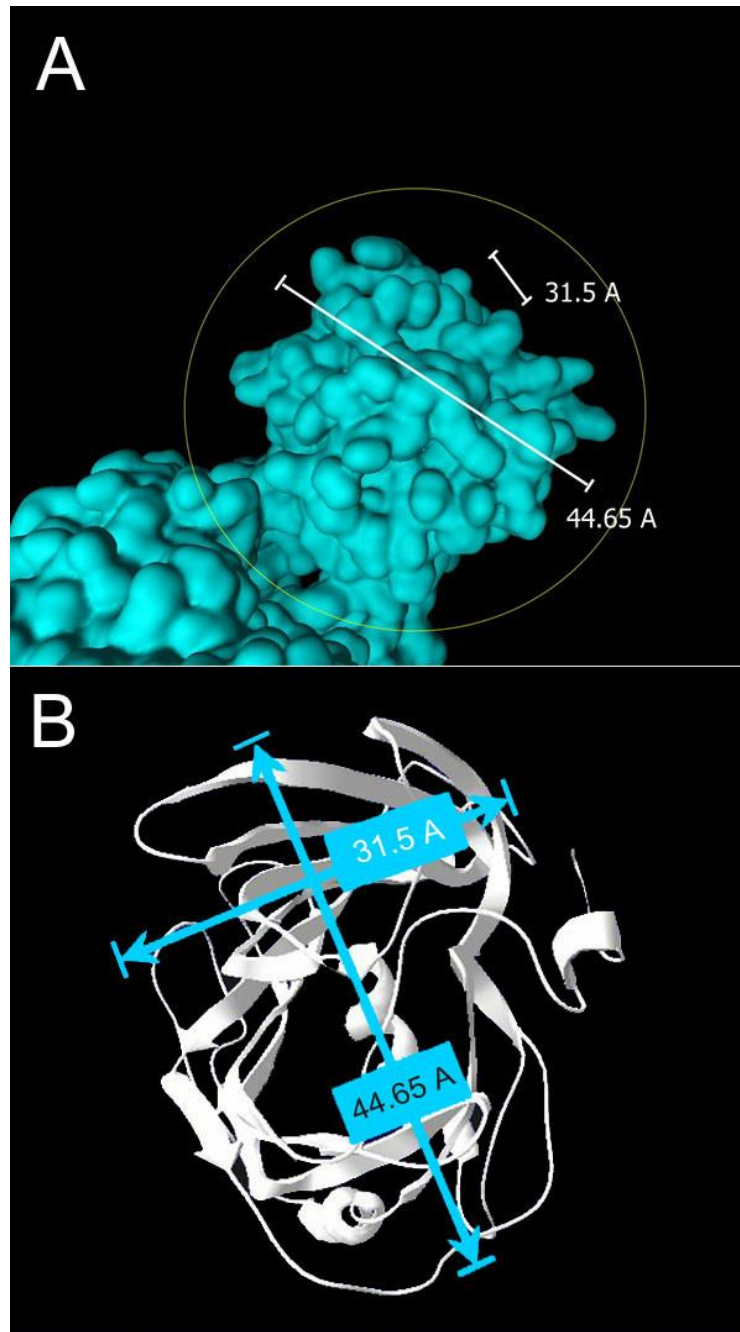
Bioinformatics - Energy minimization

- ▶ Energies are calculated by:
- ▶ $E_{\text{bonded}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}}$
- ▶ $E_{\text{nonbonded}} = E_{\text{electrostatic}} + E_{\text{van der Waals}}$
- ▶ $E_{\text{total}} = E_{\text{bonded}} + E_{\text{nonbonded}}$



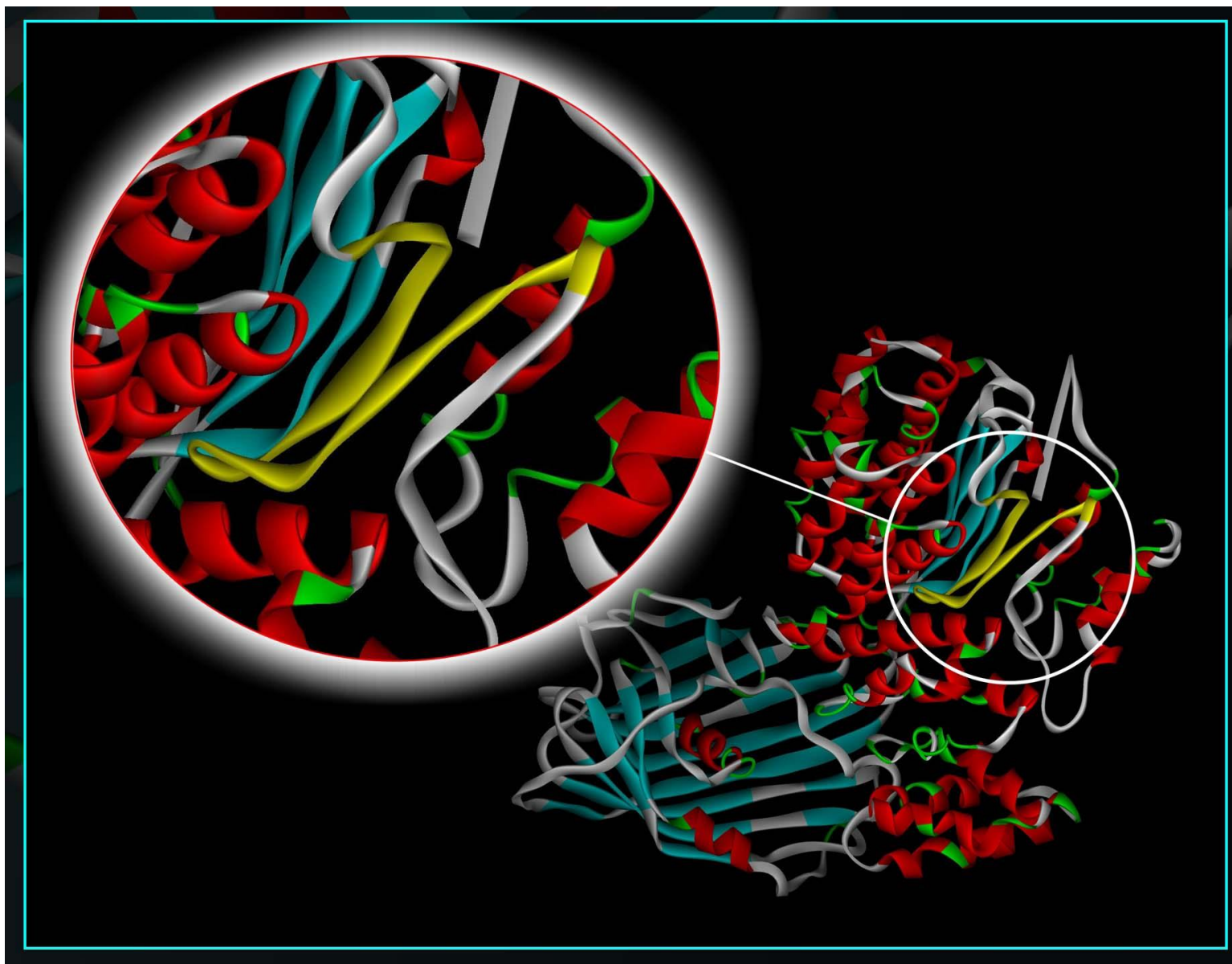


Al-Ali & Khachfe, 2007

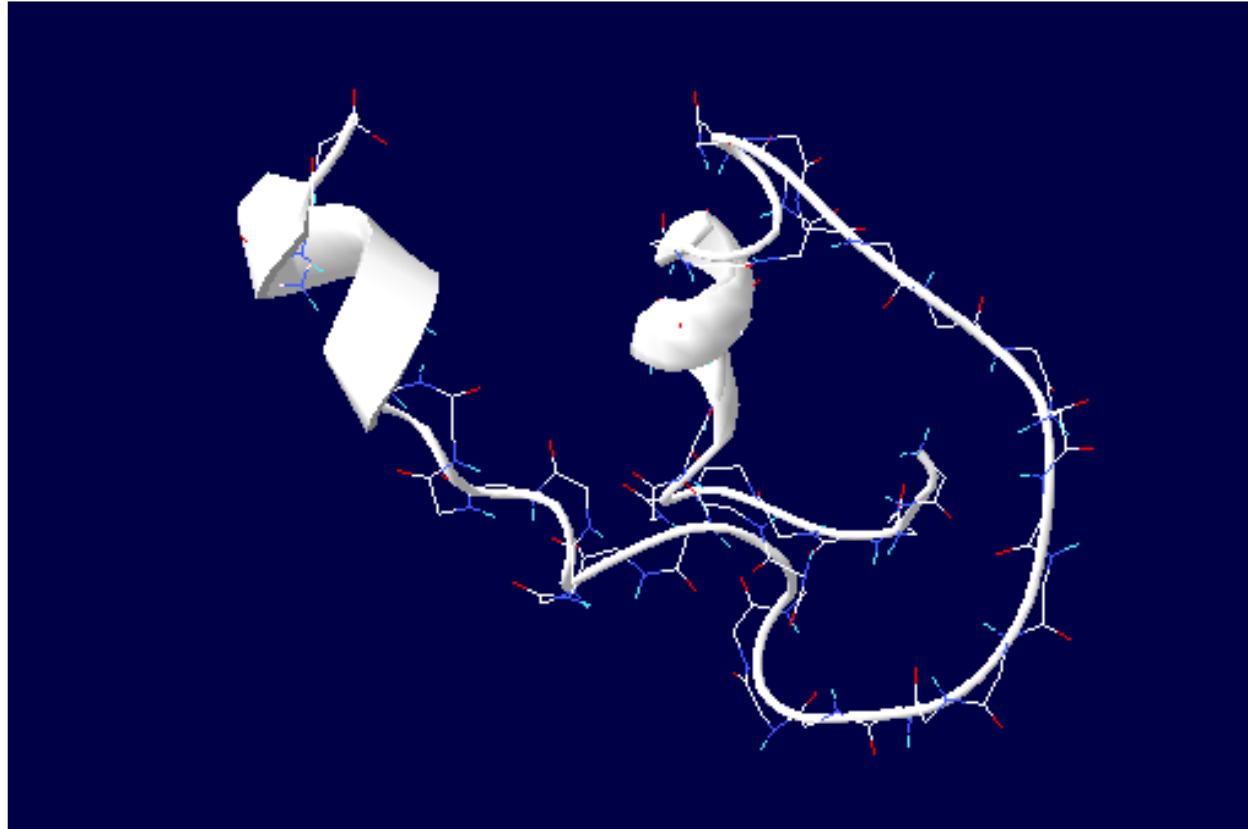


Poulos, 2001

Al-Ali & Khachfe, 2007



Bioinformatics - Success story



- ▶ Total energy before minimization is 205.602 KJ/mol
- ▶ Total energy after minimization is -327.502

Challita & Khachfe, 2015

