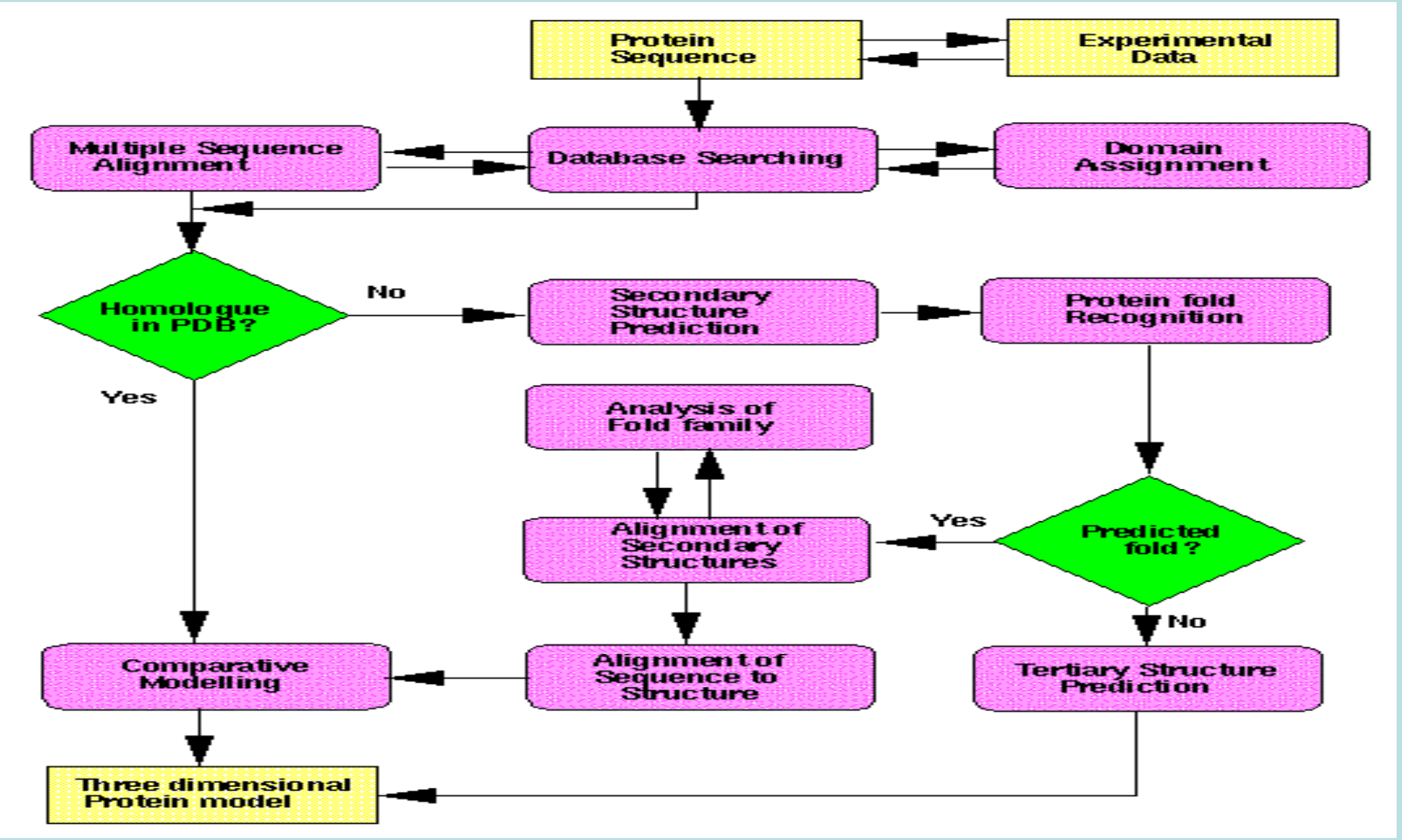
Follow Monty Python's Footsteps

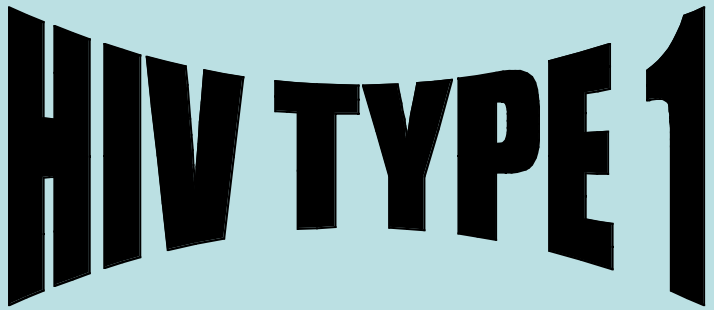

Need A Good Road Map?

General Approach in Protein Structural Prediction Flowchart

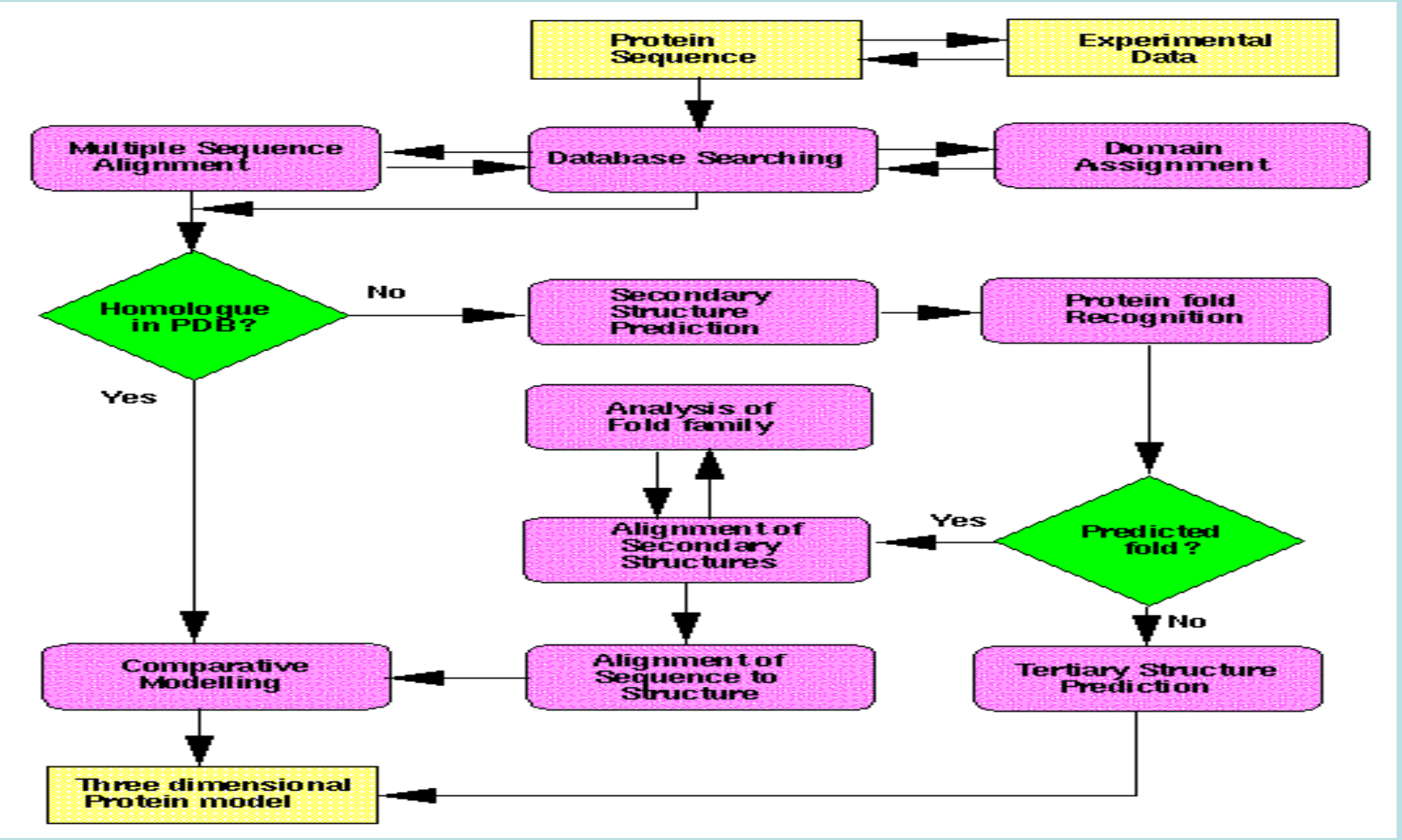


You Got Sequence ?

mgarasiltggkldkwekirlrpggkkhymlkhlvwasrelekfalnpglletsegckqi
ikqlqpalqtgteelrslyntvatlycvhagidvrdtkealdkieeeqnkiqqktqqake
adgkvsqnypivqnlqgqmvhqaisprtlnawvkvieekafspevipm



General Approach in Protein Structural Prediction Flowchart



Experimental Data

Make preliminary analysis of Protein and its Sequence Before Proceeding to Prediction.

If a protein has only qualities, then it is likely to be predictable:

1. Soluble
2. Contains only globular region
3. Comprises a single domain
4. Contains transmemberance segments [TMAP (EMBL)]
5. Contains coiled-coils [COILS server]
6. Contains only regions of low complexity

PDB Search, the First Step!

Purpose: to avoid unnecessary work if there are known protein sequence matching your protein sequence completely or closely.

What does it do?

Compare your protein sequence with other known sequence in PDB to find Homology [at NCBI or Washington University]

Methods:

1. PSI-BLAST
2. eMOTIF --- enlist common characteristics shared by a family of protein sequences.

For example: "H-[FW]-x-[LIVM]-x-G-x(5)-[LV]-H-x(3)-[DE]" describes a family of DNA binding proteins. It can be translated as "histidine, followed by either a phenylalanine or tryptophan, followed by an amino acid (x), followed by leucine, isoleucine, valine or methionine, followed by any amino acid (x), followed by glycine,... [etc.]" (Robert Russell http://www.bmm.icnet.uk/people/rob/CCP11BBS/dbsearch.html).
tools: PROSITE (http://www.expasy.ch/tools/scanprosite/) and Emotif (http://motif.stanford.edu/emotif-search/)

Things to keep in mind

A. Compare your sequence against a database of sequences with known 3D structure(which means that the 3D structure of your protein is readily known if homology is found between your protein and one or more protein of known 3D structure.)

B. Use pre-prepared protein alignment (preferrably hand edited by experts), which likely represents best alignments.

SMART (http://smart.embl-heidelberg.de/)

BLOCKS (http://www.blocks.fhcrc.org/)

Domain Assignment

A. Split a long protein sequence(says comprise of 500 amino acids) into discrete Functional domains and repeat previous PDB search and sequence alignment for each domain.

B. Method of Indentifying domains:

1. Spot the one and only portion of your protein sequence that has homology to a known protein sequence.
2. Search well-curated, pre-defined database of protein domains.
   SMART
3. Regions of your protein containing different protein structural classes (such as alpha helices at one region and beta sheets on the other).

C. Identifying domains seperators:

1. low-complexity region (which are often domain seperator in multiple-domain protein).
   program SEG
2. transmembrane segments(which splits extracecellular from from intracecellular domains).
   TMAP
3. Coilded-coils (sometimes it can indicates where protein splits into different domains).
   COILS SERVER , program COILS

Assigning Domains

Are there partial homologies?

Are there regions of low complexity?

Does prediction suggest domains?

If domains are assigned
do database searches again

Multiple Sequence Alignment

Why Sequence Alignment? It provides information in

protein domain structure

location of residues involved in protein function

states of residues: buried in protein core or exposed

and other useful information for homology modeling and seondary structure prediction

Methods:

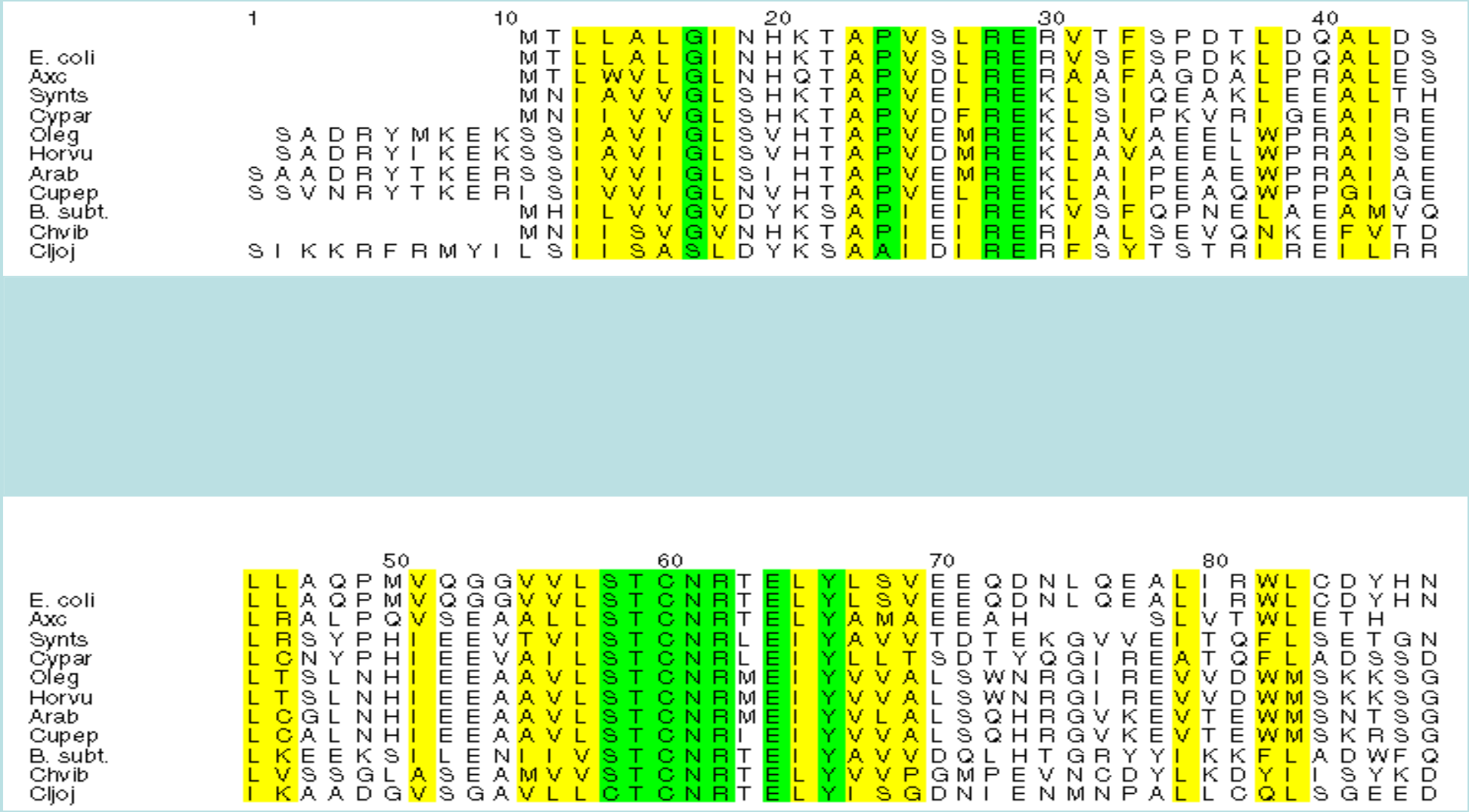
EBI (UK) Clustalw Server

BCM Multiple Sequence Alignment ClustalW Sever (USA)

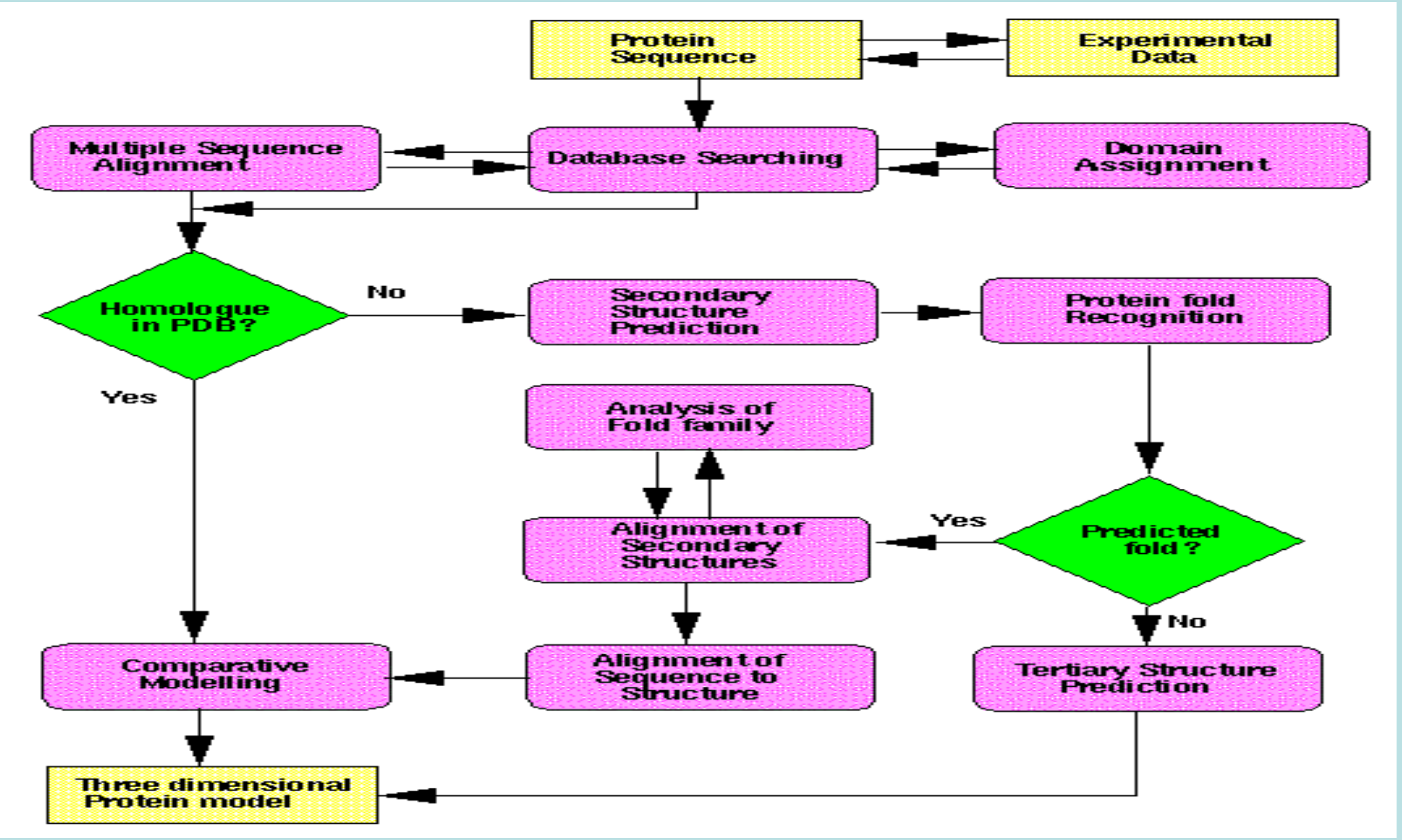
Programs: HMMer (HMM method, Wash U) , MSA (USA)

Things to Keep in Mind:

Align up your protein and its homologues found after throwing out “homologues”, in PDB search results, that are unlikely to be a member of the sequence family of your protein.



General Approach in Protein Structural Prediction Flowchart



Found Homologue?

No! Procede to Secondary Structure Prediction

Secondary Structure Prediction

Goal: to locate alpha helices and beta strands in your protein or your protein family

Methods:

1. Automated prediction(about 70-80% accuracy) : Submit the multiple sequence alignment obtained previously to a server

   PSI-pred

   JPRED

2. Manual Prediction(in some case nearly 100% accurary): Look at residue conservation in your protein for indication of particular secondary protein structure class.

   A. Principle:Different classes of protein structure show different residue conservations.

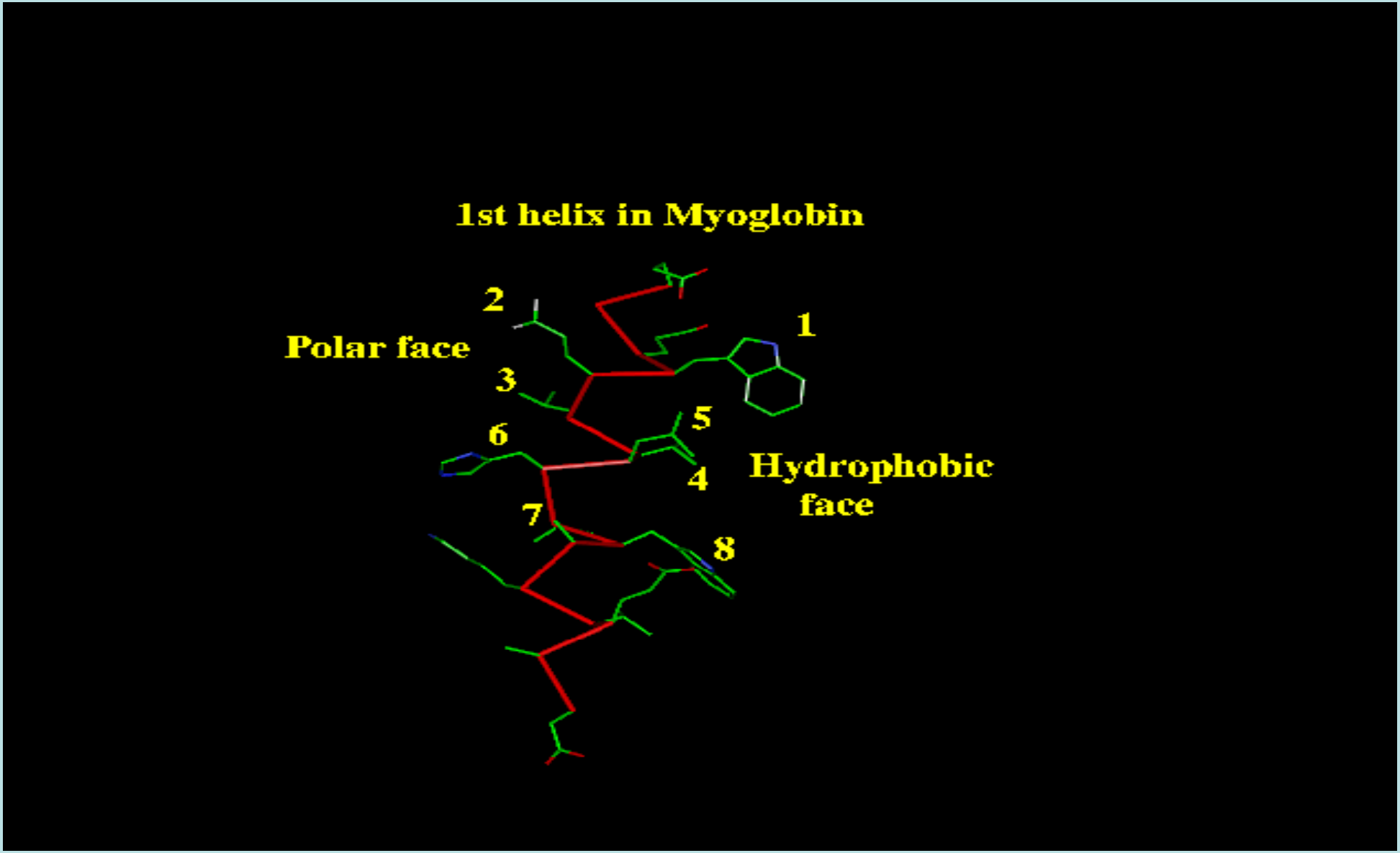
   B. Examples:

   Alpha Helices

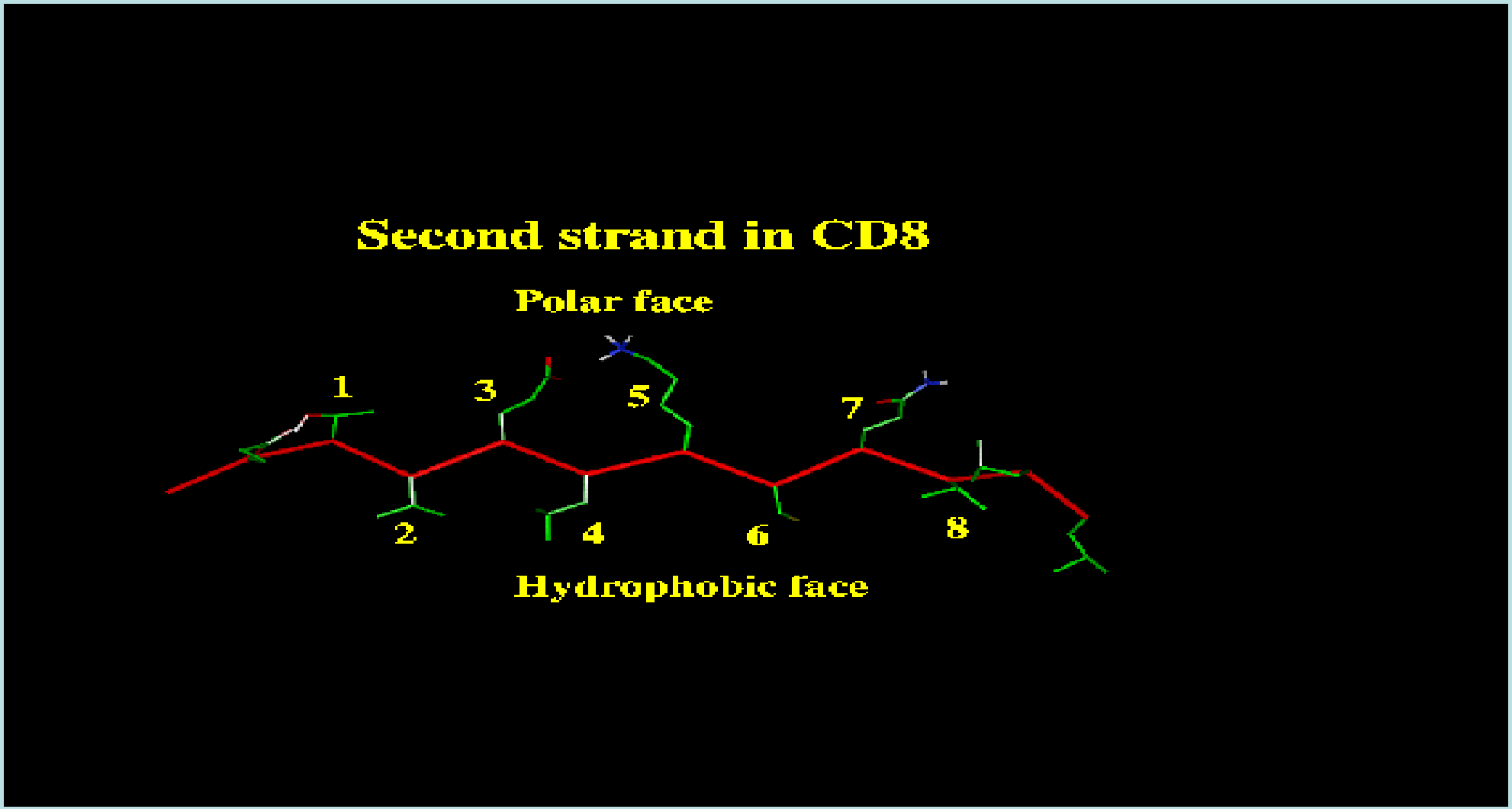
   Beta Strands (half-buried in protein core)

   Beta Strands (total-buried in protein core)

Alpha Helices (with a periodicity of 3.6) --- have residues at positions i, i+3, i+4 & i+7 for helices with one face buried in protein core while the other face exposes to solvent.

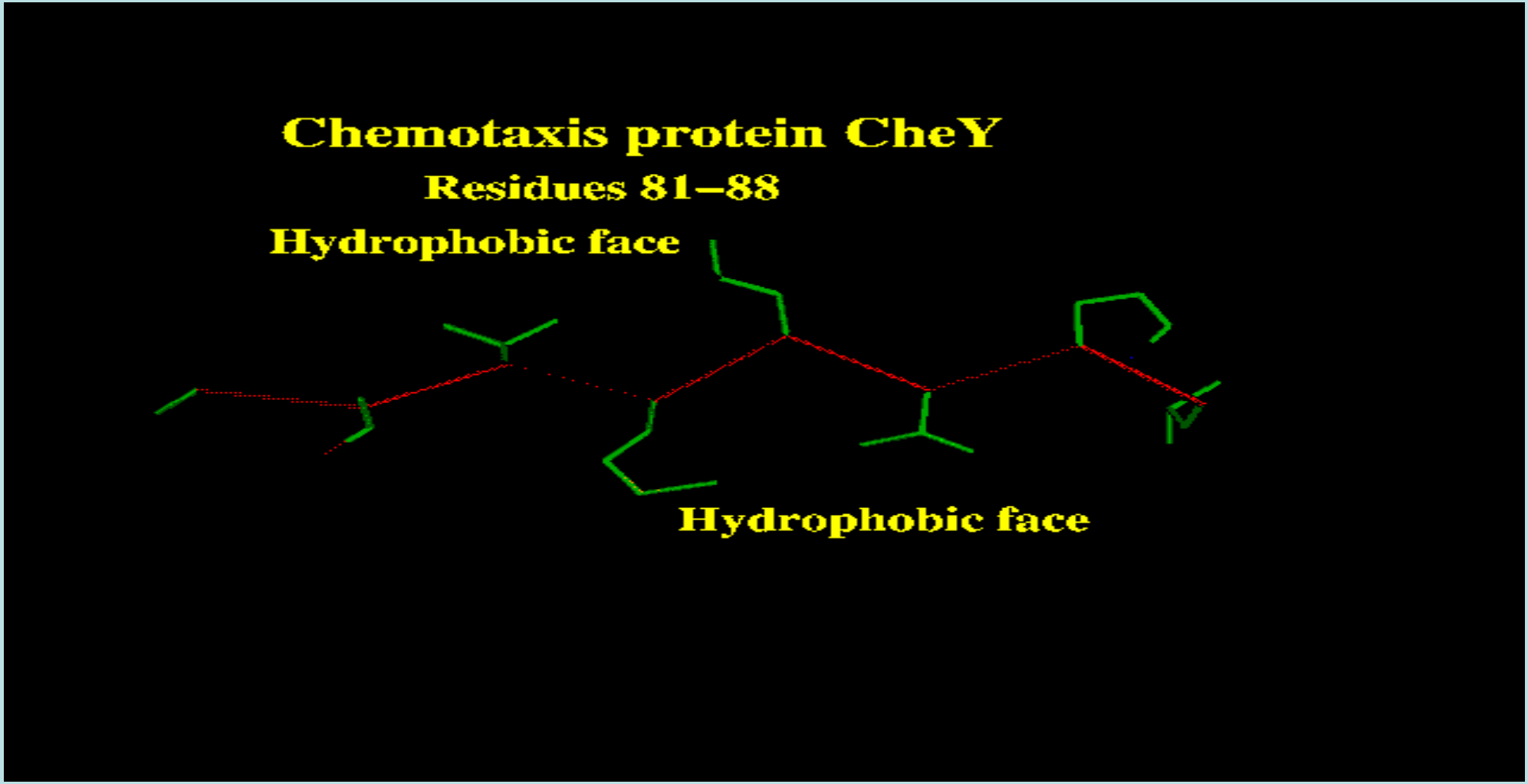


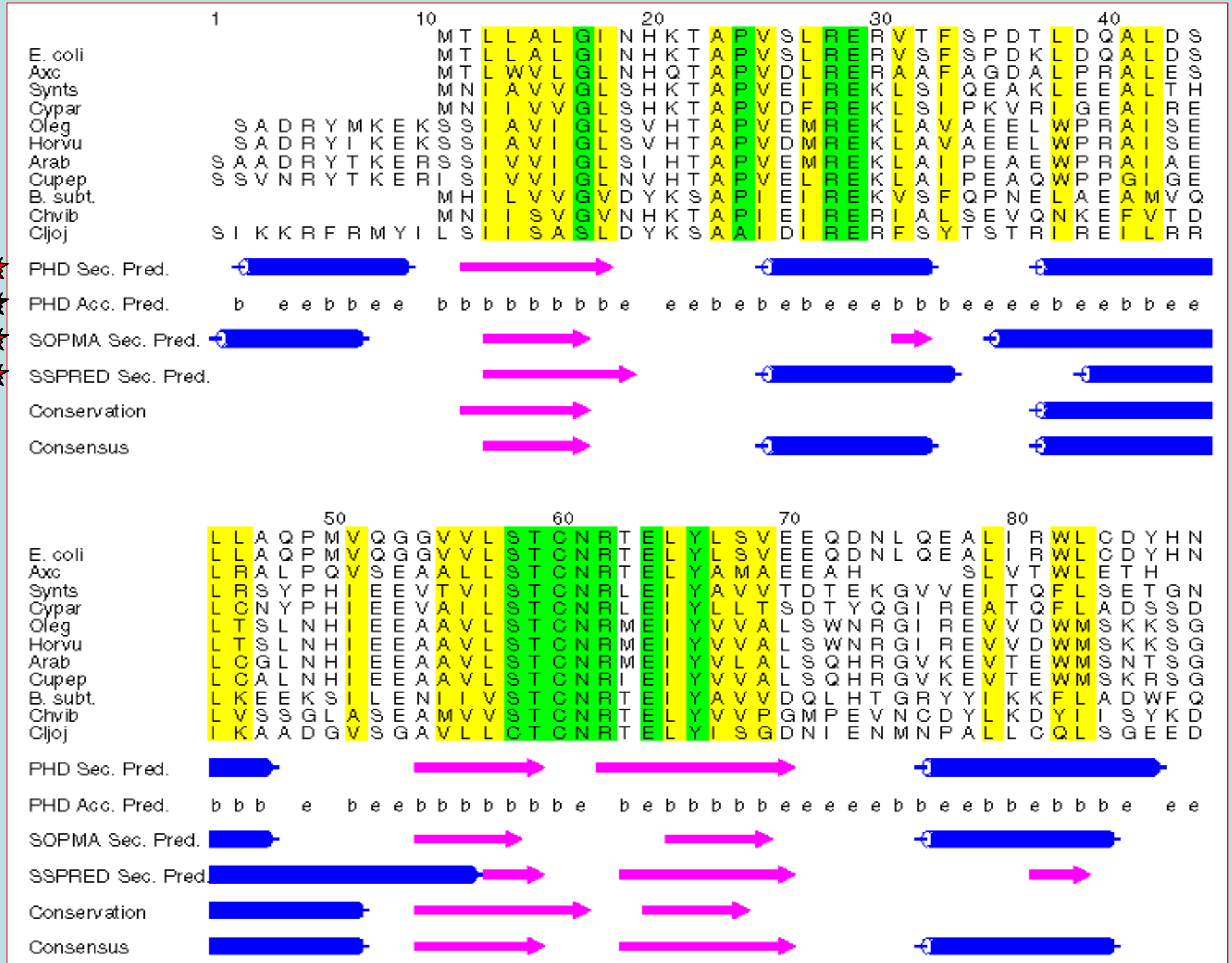
beta strands means that adjacent residues have their side chains pointing in oppposite directions.
Beta strands that are half buried in the protein core will tend to have hydrophobic residues at positions i, i+2, i+4, i+8 etc, and polar residues at positions i+1, i+3, i+5, etc.
For example, this beta strand in CD8 shows this classic pattern:



Beta strands that are completely buried usually contain a run of hydrophobic residues, since both faces are buried in the protein core.

This strand from Chemotaxis protein CheY is a good example:





Protein Fold Recognition

Aim: to discover a 3D structure compatible for a protein by fitting the protein's sequence onto known structures

disclose similarity in 3D structure among proteins that are dissimilar in structure.

Facts:

1. Based on experience, experts knows that proteins with very little similarity in sequences and functions can still have similar 3D structure.
2. There are only a limited number of protein folds in nature.

Methods:

3D-pssm a server

THREADER(Warwick) a downloadable program

ProFIT CAME (Salzburg) a downloadable program

Databases of Protein Structure Classification(According t Robert Russell, the following database can provide a suitable structure to build a 3D model for roughly 70% of all protein):

* SCOP (MRC Cambridge), CATH (University College, London) , FSSP (EBI, Cambridge) 3 Dee (EBI, Cambridge), HOMSTRAD (Biochemistry, Cambridge), VAST (NCBI, USA)

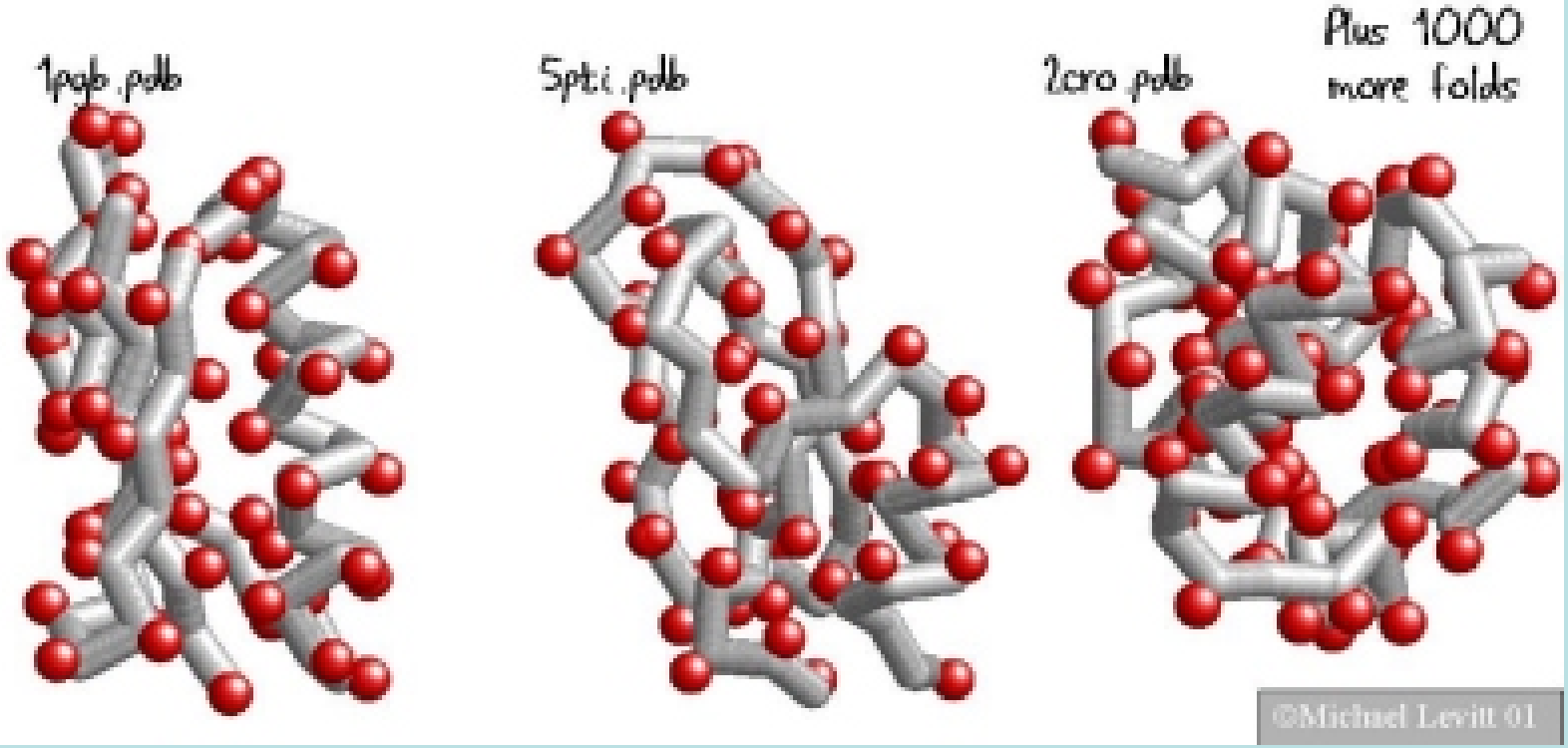
Realities: the Critical Assessment of Structure Predictions (CASP) conferences showed so far the accuracy of fold recognition is no too high.

Limitations:

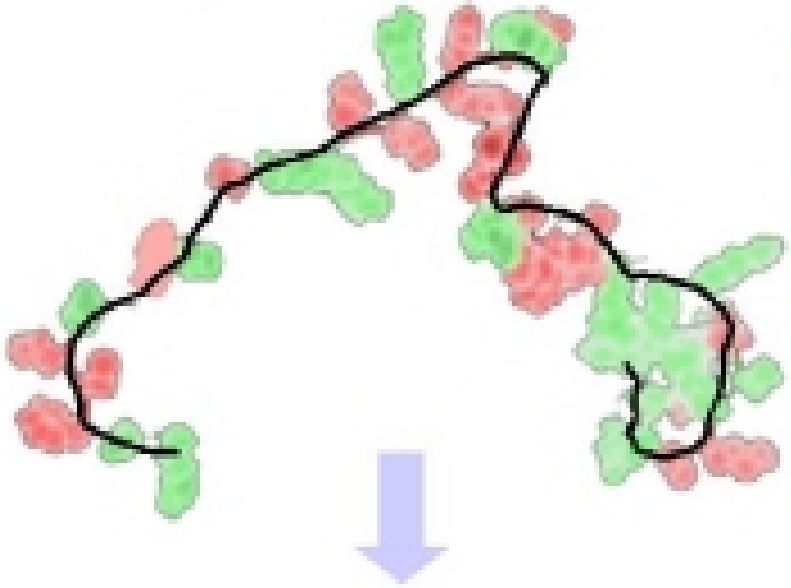
WHAT IS FOLD RECOGNITION?

Find the fold that best fits the query sequence.

Query Sequence: R V L G F I P T W F A L S K Y

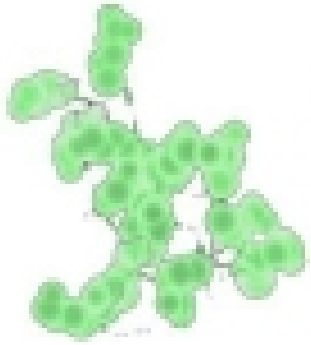


WHAT DRIVES FOLDING?

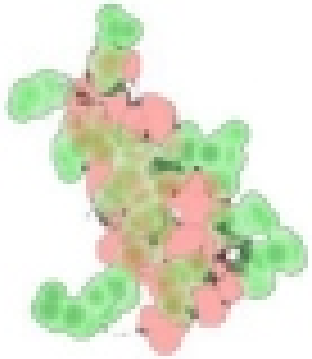


- Protein is a chain.
- Self-avoiding and close packed.
- Residue preferences:
  - Inside/Outside
  - Specific Neighbors

Hydrophobic



Hydrophilic

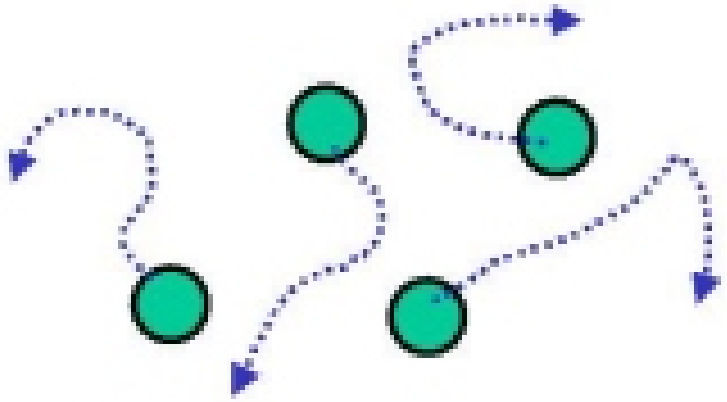


All Residues

Pink are hydrophobic, like to be away from water

Green are hydrophilic, like contact with water

MOLECULAR DYNAMICS THEORY



- Force = $-dU/dx$ (slope of potential, U); acceleration, $m\ a(t)$ = Force.
- All atoms move together so force between atoms change with time.
- Analytical solution for $x(t)$ and $v(t)$ is impossible; numerical solution is trivial.

$$x(t+\Delta t) = x(t) + v(t)\Delta t + \left[4a(t) - a(t-\Delta t)\right]\Delta t^2/6$$

New position — Old position — Old velocity — Acceleration

$$v(t+\Delta t) = v(t) + \left[2a(t+\Delta t) + 5a(t) - a(t-\Delta t)\right]\Delta t/6$$

New velocity — Old velocity — Acceleration

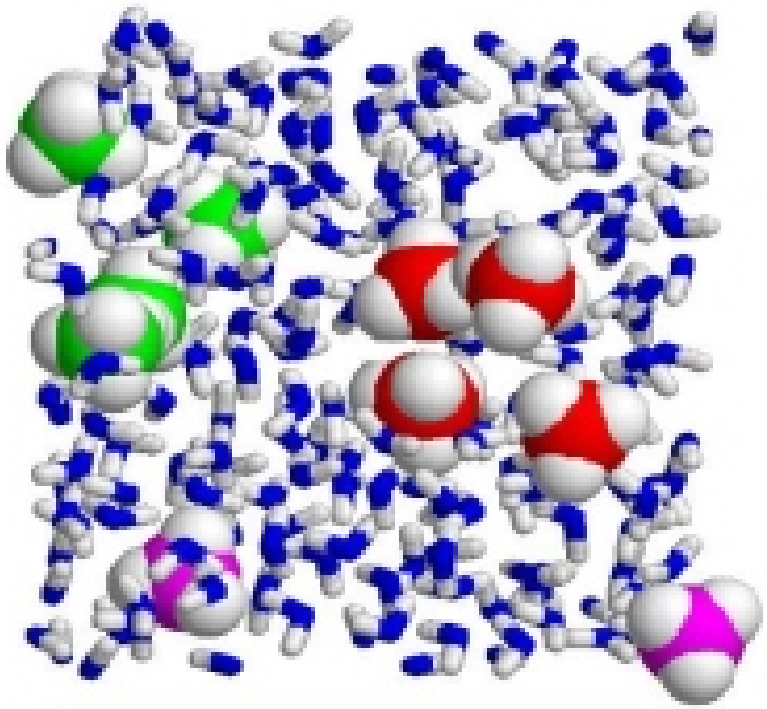
Time step, Δt, must be very small at 10^{-15} seconds or 0.001 ps.

$$U_{kinetic} = \frac{1}{2}\sum m_i v_i(t)^2 = \frac{1}{2} n k_B T$$

Kinetic energy — Atomic masses, velocities — Number of coordinates (not atoms) — Boltzmann's Constant — Temperature

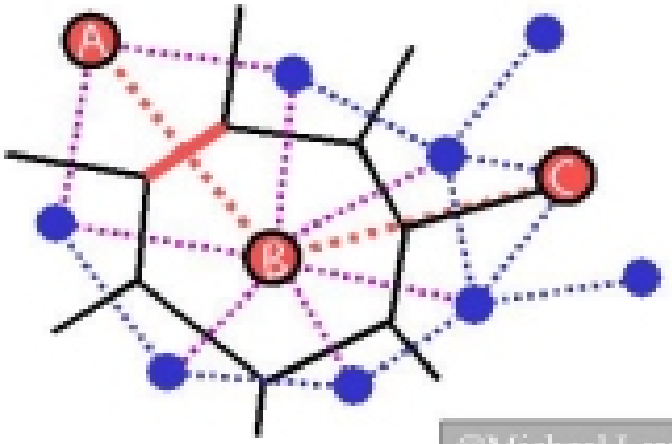
Total energy ($U_{potential} + U_{kinetic}$) must not change with time

HYDROPHOBIC EFFECT

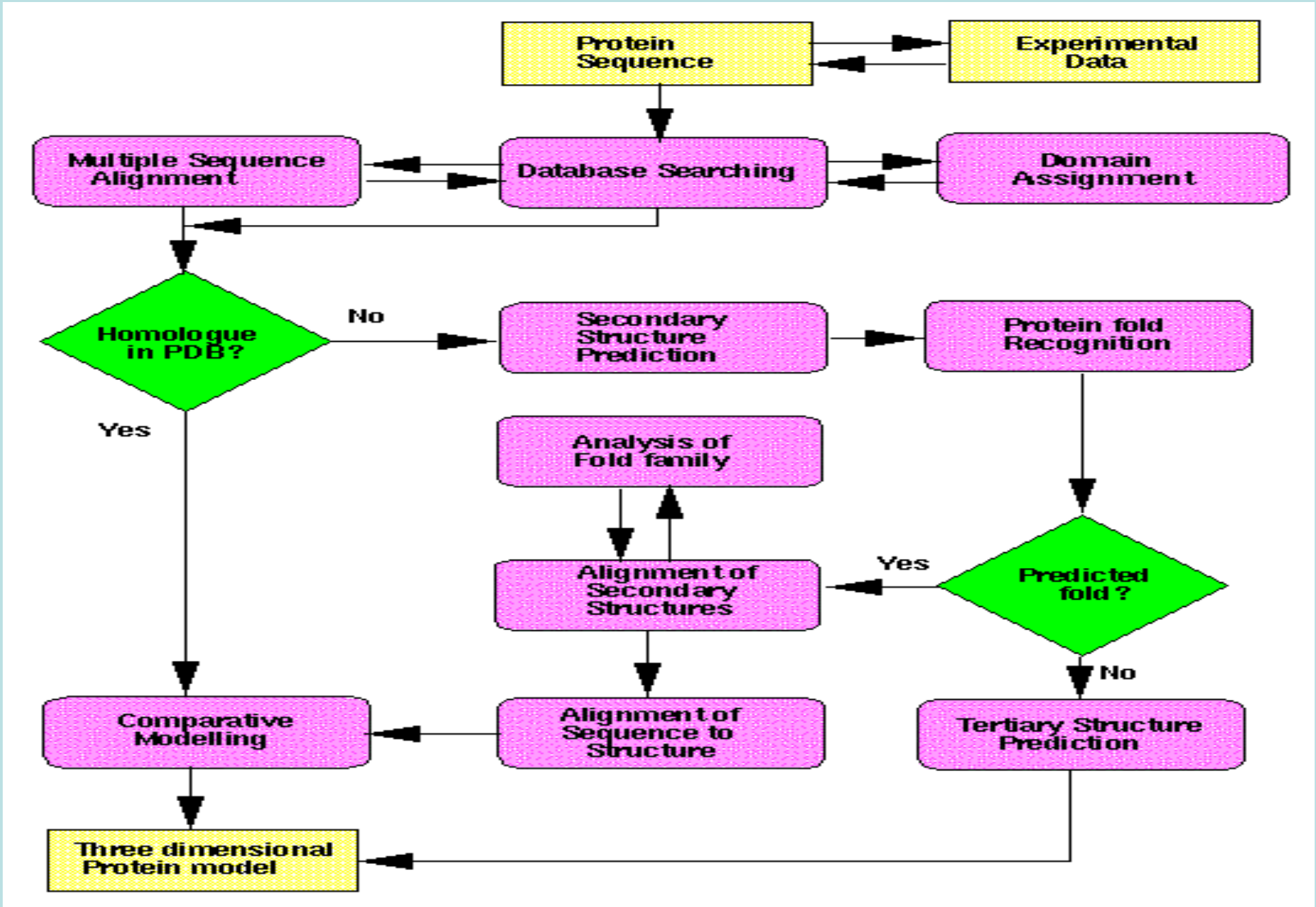


Box with periodic boundaries.

- 1 nanosecond MD simulations in periodic water boxes with from 30mM to 3 Molar hydrocarbon solution.
  Encad with F3C water (1996).
- Measure cluster formation by Voronoi.
  d(AB) = d(BC), but only A, B touch.



General Approach in Protein Structural Prediction Flowchart




I am feeling lucky !

Analysis of protein folds Family

Aim --- to detect what family of folds you protein belong after knowing your

protein adotping a particular fold.

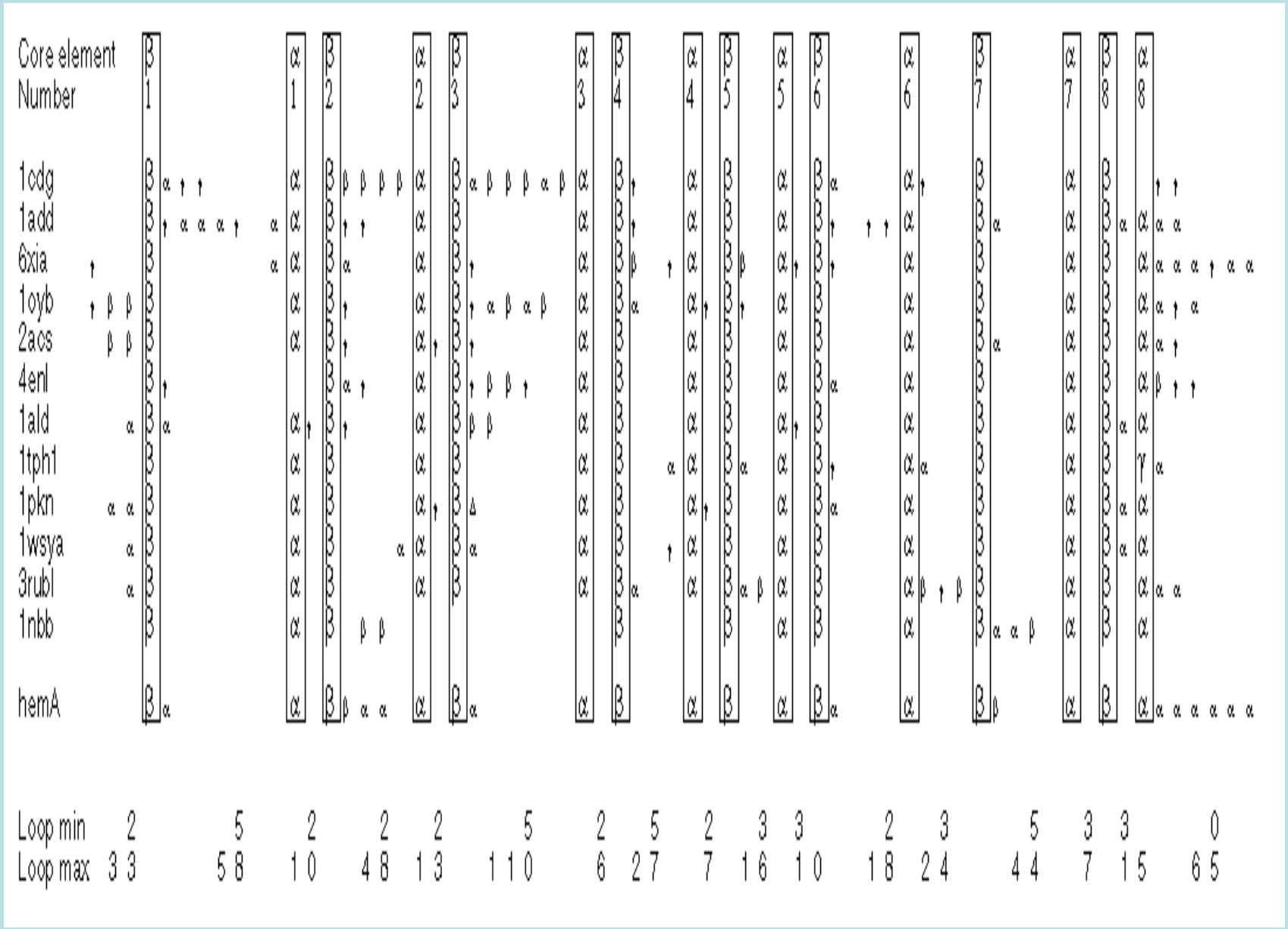
Methods --- Compare your fold to folds in one of the following databases:

* SCOP (MRC Cambridge)
* CATH (University College, London)
* FSSP (EBI, Cambridge)
* 3 Dee (EBI, Cambridge)
* HOMSTRAD (Biochemistry, Cambridge)
* VAST (NCBI, USA)

Analysis of Fold Family

Does the Predicted Fold Family is right family for your protein?
*One or more of its member shares functional similarities with your protein
*Its members also contains core secondary structure elements that are in your
* protein (run your protein and the fold family through a structural alignment program).

Alignment of the secondary structures of hemA to those of the alpha-beta barrel fold

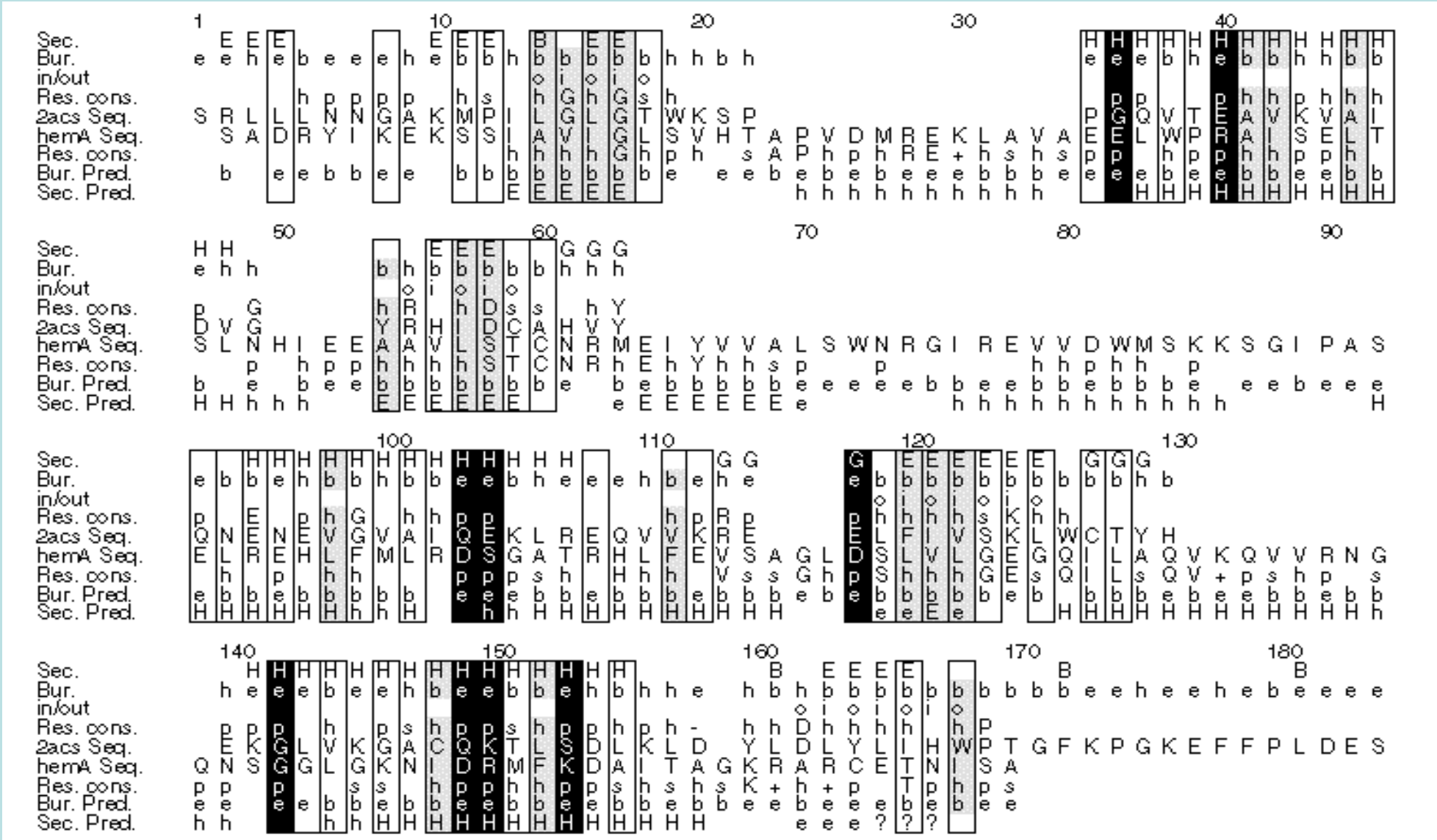


alignments of sequence on to tertiary structure that one gets from fold recognition

Procedures:

1. Aligning residues predicted to be buried/exposed align to those *known* to be buried or exposed in the template structure.(predict residue accessibility manually, or by use of an automated server like PHD.
2. Ensuring no disruption o f critical hydrogen bonding patterns in beta-sheet structures.
3. Conserving residue properties (i.e. size, polarity, hydrophobicity) across known and unknown structure.

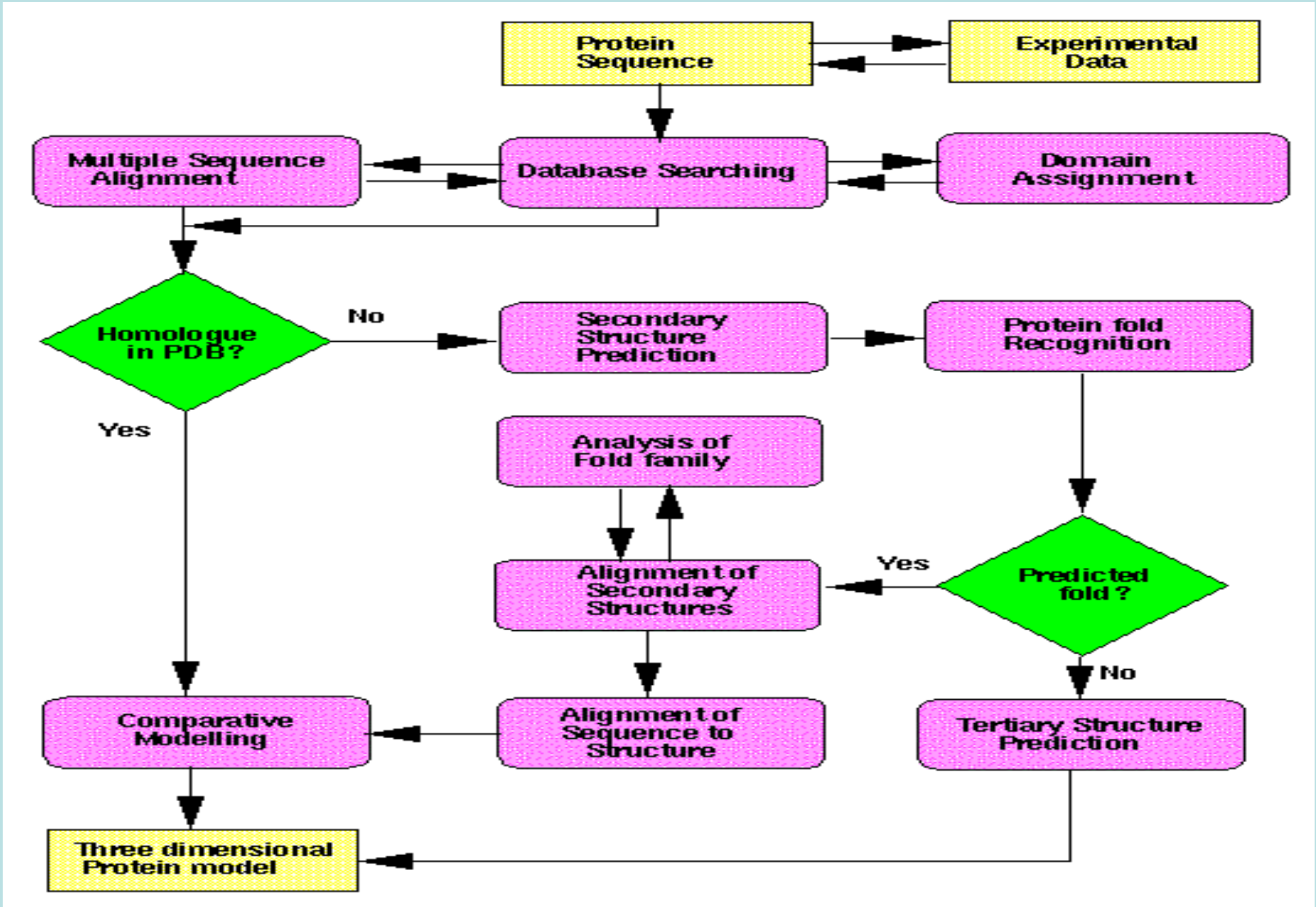
align the prediction of the glutamyl tRNA reductases (hemA) with one alpha/beta barrel structure



Things to keep in mind while constructing the alignment

*The observed residue burial or exposure
*The predicted residue burial or exposure
*The conservation of residue properties in known and unknown structures
*Whether or not the side chains on the core beta-strands pointed in towards the barrel or out towards the helices
*The hydrogen bonding pattern of the beta-strands comprising the core beta-barrel.

General Approach in Protein Structural Prediction Flowchart



Found Homologue?


I am feeling lucky !

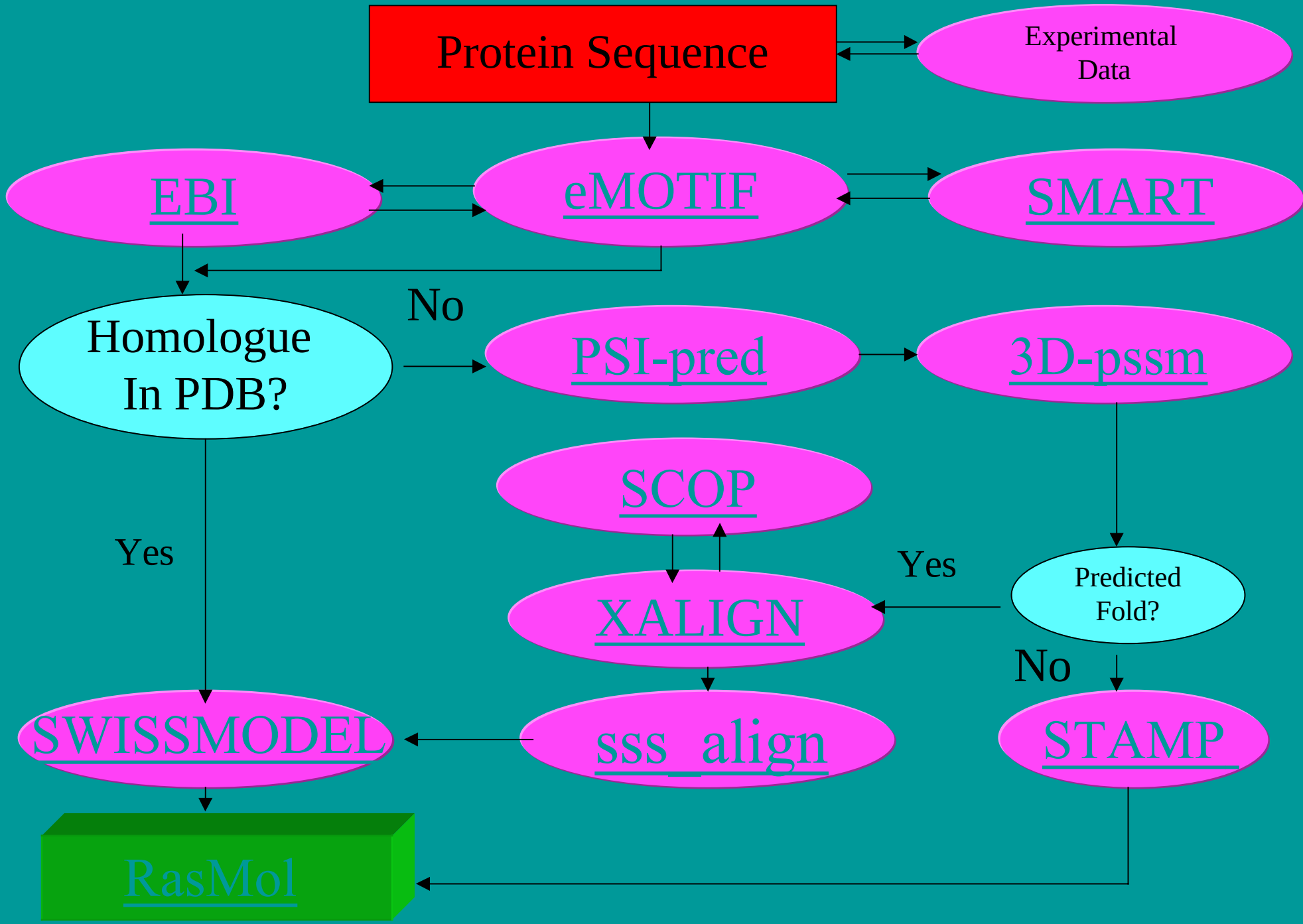
Comparative or Homology Modeling

Build a 3D model for your protein by submitting sequence alignment of your protein and its significant homologues(with homology > 50%) to SWISSMODEL server or WHAT IF (G. Vriend, EMBL, Heidelberg)

Take a look at 3D structure of your protein build upon the 3D model via program Prepi (Suhail Islam, ICRF, U.K) or RasMol Roger Sayle, Glaxo, U.K

Woooo! Holy Grail !

The Site Map to Holy Grail



Specials Thanks to

Doug Brutlag

Robert Russell

Michael Levitt

I hope you I have make you feel a billion dollar richer!