Protein Structure Prediction

Dr Safdar Ali
Department of Biological Sciences
Aliah University, Kolkata
Date: 06/04/2020

Levels of Protein Structure

- Primary
- Secondary
- Tertiary
- Quaternary

- Sequence to Structure

- Structure to Function

- For many proteins sequence identity is extremely limited.
- Example **retinol-binding protein** and **odorant-binding protein**
- Both are **lipocalins** of about **20 kDa** and are abundant, secreted carrier proteins. They share a **GXW motif** that is characteristic of lipocalins.
- However, pairwise alignment the two proteins share less than 20% identity.
- Both structure and function are preserved over evolutionary time more than is sequence identity.
- Thus, the three-dimensional structures of these proteins are extraordinarily similar.

Can it really be generalized?

Can a single priority rule about sequence, structure and function of proteins be forged?

- Even **single amino acid substitutions** can cause a dramatic change in protein structure, as exemplified by **disease-causing mutations.**
- **Many other substitutions have no observable effects on protein structure.**
- Wood and Pearson (1999) examined 36 protein families, each having five or more members with known three-dimensional structures.
- They found a very high correlation between **sequence similarity and structural similarity** for three-quarters of the protein families.
- Wood and Pearson concluded that most amino acid sequence changes cause detectable structural changes. **Also, the amount of structural change is relatively constant within a protein family.**

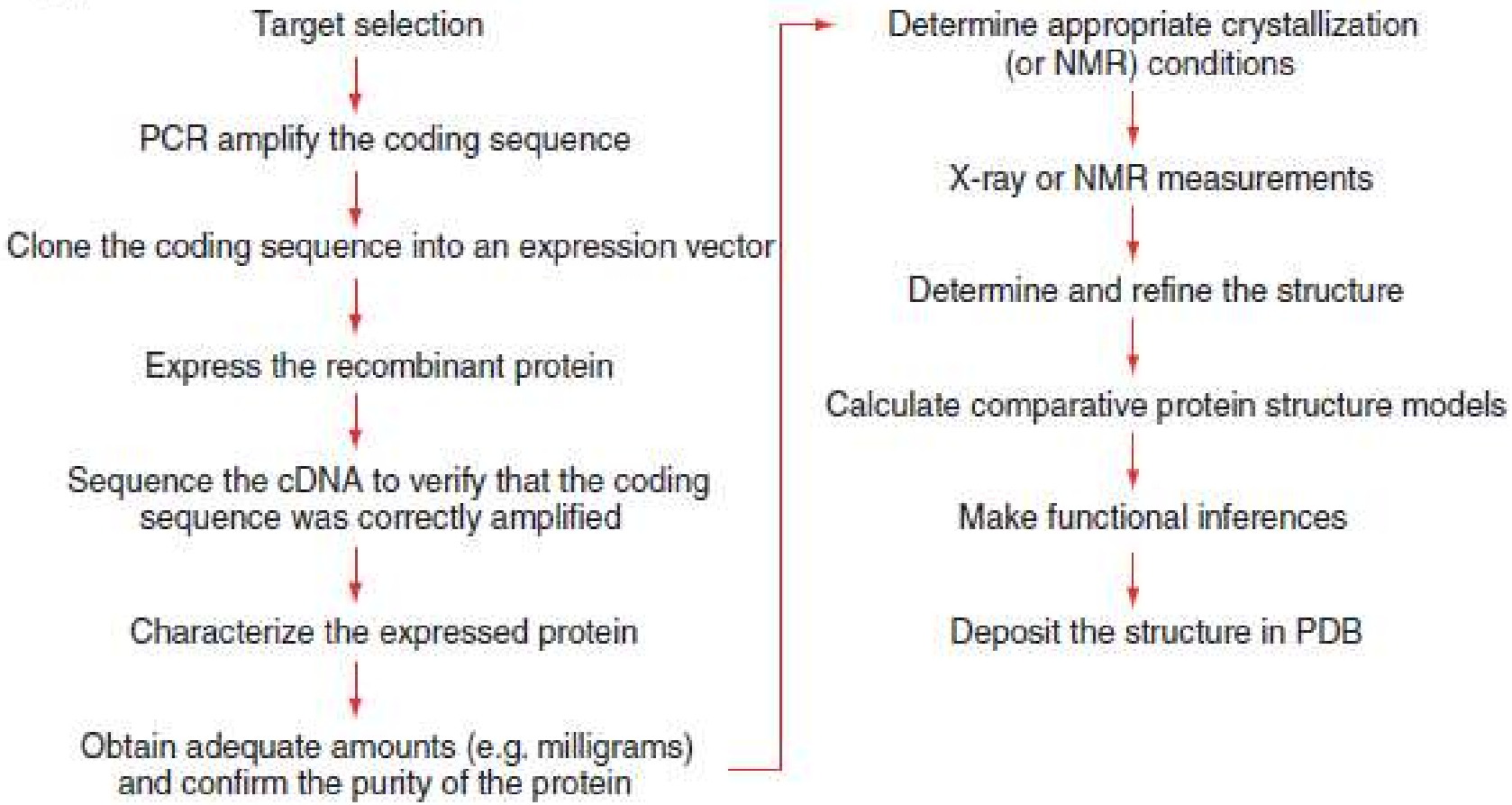
TABLE 11-3 Secondary-Structure Prediction Programs Available on Internet

Program	Comment	URL
APSSP	Based on neural networks	► http://imtech.res.in/raghava/apssp/
DSSP		
GOR4	From the Pole Bio-Informatique Lyonnais	► http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page = npsa_gor4.html
Jpred	From the Barton group (Dundee)	► http://www.compbio.dundee.ac.uk/~ www-jpred/
NNPREDICT	An enhanced neural network approach (from UCSF)	► http://alexander.compbio.ucsf.edu/~ nomi/nnpredict.html
PredictProtein server	Based on neural networks	► http://www.predictprotein.org/
PSIPRED	From the University College London	► http://bioinf.cs.ucl.ac.uk/psipred/
SAM-T02	Uses hidden Markov models (Chapter 6)	► http://www.soe.ucsc.edu/research/compbio/HMM-apps/T02-query.html
Sosui	From the Mitaku Group (Tokyo)	► http://bp.nuap.nagoya-u.ac.jp/sosui/

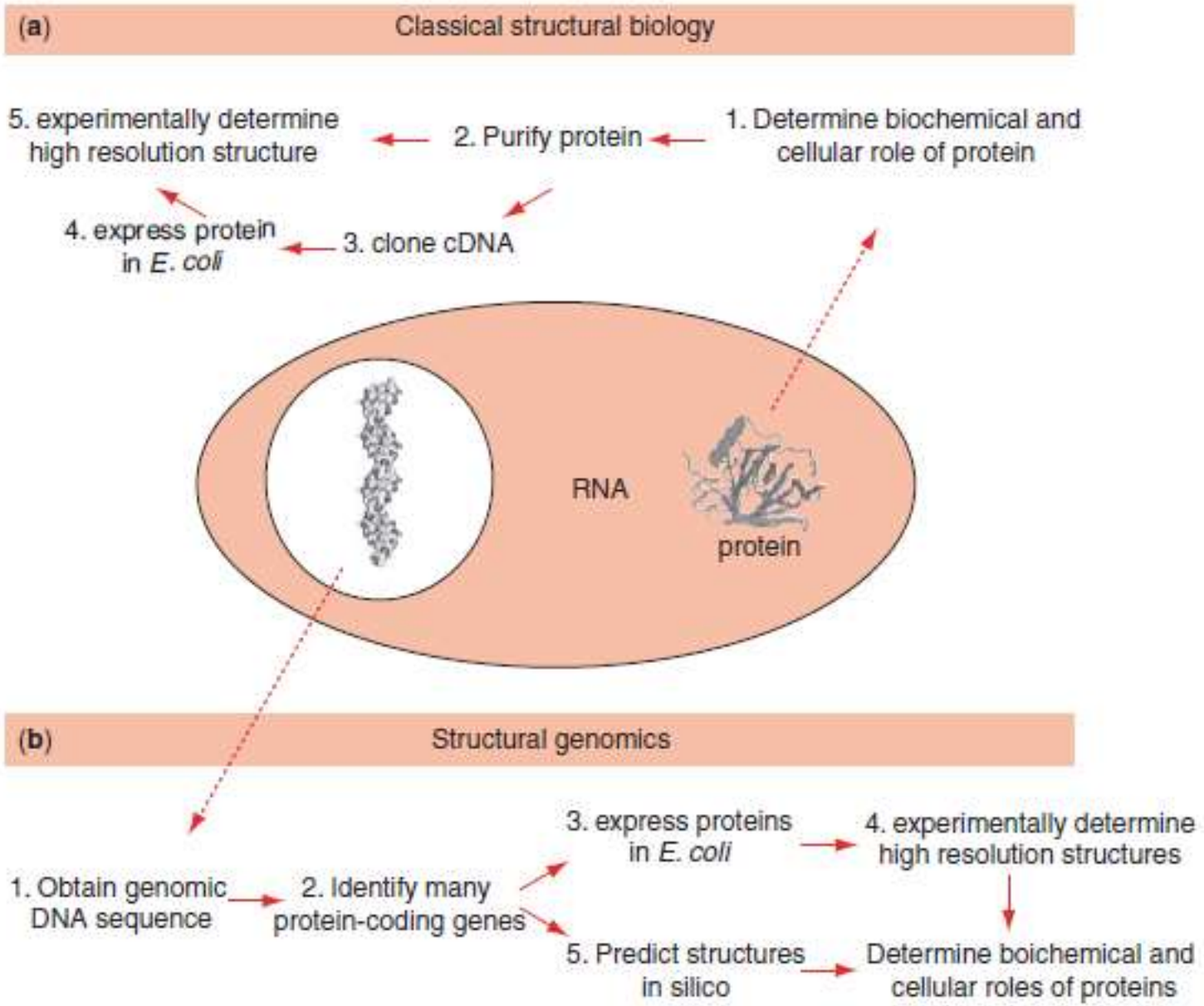
Note: Additional sites are listed at ExPASy (► http://www.expasy.org/tools/#secondary) and PBIL (► http://npsa-pbil.ibcp.fr).

Bioinformatics & Functional Genomics 3rd Edition, 2015, Jonathon Pevsner, Wily and Blackwell Publishing, Chapter 11

- What ligand does each protein transport?
- Can structural studies reveal the binding domain to suggest the identity of the ligand? How much structural information is required in order to predict the ligand from sequence information?
- Mutations in globin genes result in a variety of human diseases, including thalassemias and sickle cell anemia.
- Can we predict the structural and functional consequences of a specific mutation?
- Globins have been divided into subgroups based on phylogenetic analyses and their localization. To what extent do those groupings reflect structural and functional similarities?
- When a genome is sequenced and a gene encoding a putative novel globin is discovered, can we use information about other globins of known structure in order to predict a new structure?



Bioinformatics & Functional Genomics 3rd Edition, 2015, Jonathon Pevsner, Wily and Blackwell Publishing, Chapter 11



Bioinformatics & Functional Genomics 3rd Edition, 2015, Jonathon Pevsner, Wily and Blackwell Publishing, Chapter 11

TABLE 11-5 Interactive Visualization Tools for Protein Structures

Tool	Comment	URL
Chime	Plug-in for a web browser	Instructions at PDB
Cn3D	From NCBI	► http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml
Mage	Reads Kinemages	► http://kinemage.biochem.duke.edu/ and http://www.ncbi.nlm.nih.gov/Structure/CN3D/mage.html
MICE Java applet		Instructions at PDB
RasMol	A stand-alone package	Instructions at PDB
SwissPDB viewer	At ExPASy	► http://www.expasy.org/spdbv/
VMD	Visual Molecular Dynamics; University of Illinois	► http://www.ks.uiuc.edu/Research/vmd/
VRML	Uses MolScript	Instructions at PDB

Note: The Protein Data Bank maintains a list of molecular graphics software links, accessible from the PDB

TABLE 11-7 Partial List of Protein Structure Databases

Database	Comment	URL
3dee	Structural domain definitions	► http://www.compbio.dundee.ac.uk/3Dee/
CASTp	Computed Atlas of Surface Topography of proteins (CASTp)	► http://sts.bioengr.uic.edu/castp/index.php
CE	Complete PDB and representative structure comparison and alignments	► http://cl.sdsc.edu/ce.html
Enzyme Structures Databases	Enzyme classifications and nomenclature	► http://www.ebi.ac.uk/thornton-srv/databases/enzymes/
FATCAT	Flexible structure AlignmenT by Chaining Aligned fragment pairs allowing Twists	► http://fatcat.burnham.org/
FSSP	Structurally similar families	► http://www.sander.ebi.ac.uk/dali/fssp/
HSSP	Homology-derived secondary structures	► http://swift.cmbi.ru.nl/gv/hssp/
JenaLib	Jena Library of Biological Macromolecules (JenaLib)	► http://www.fli-leibniz.de/IMAGE.html
NDB	Database of three-dimensional nucleic acid structures	► http://ndbserver.rutgers.edu/
OCA	Browser-database for protein structure/function	► http://oca.ebi.ac.uk/oca-docs/oca-home.html
PDBSum	Summary information about protein structures	► http://www.ebi.ac.uk/pdbsum/

Bioinformatics & Functional Genomics 3rd Edition, 2015, Jonathon Pevsner, Wily and Blackwell Publishing, Chapter 11

Secondary Structure Prediction Methods

- PHDsec (Rost 1996)
- PROFsec (Rost 2003)
- PSIPRED (McGuffin 2000)
- SAM-T99

Solvent Accessibility Prediction Methods

- PHDacc
- PROFacc
- Jpred (Cuff and Barton 2000)

Transmembrane Segments Prediction Methods

- TopPred
- PHDhtm
- ProfTMB
- SOSUL
- TMHMM
- DAS

Function Predicting Methods

- PROSITE
  - Catalog of biologically significant sites
- Pfam
  - Collection of protein domains
- InterPro
  - Protein families, domains and functions
- BLOCKS
  - Family of proteins

Subcellular Prediction Methods

- PSORT
  - Library of signal peptides
- SUBLOC
  - Amino acid composition
- Target P
  - Signal peptides at N terminal of protein
- LOC3D
  - Nuclear localization signal

Experimental Protein Structure Determination

- X ray crystallography
- NMR

Protein Structure Database

- PDB
- CATH Class, Architecture, Topology, Homology
- SCOP Structural Classification of Proteins

Homology Modeling

- Alignment
- Select backbone segments
- Replace backbone segments
- Replacing side chains
- Energy minimization
- Validating the model

Threading

- Fold recognition method
- Common fold in protein
- No sequence homology
- Approximate models
- Two dimensional threading
  - Prediction based method (PBM)
- Three dimensional threading
  - Distance based Method (DBM)

Ab initio Structure Prediction

- Minimize no of torsional angles disallowed
- Maximize no of hydrogen bonds
- Minimize exposed hydrophobic residues

  - DSSP
  - PROCHECK
  - VADAR
  - Verify3d

Thank You