1. Give a short description of how X-Ray Crystallography is performed?

Protein of interest is produced in high quantity and purified. The purified protein in solution is then crystallized using varying methods.
When we have a good quality crystal, X-ray beam is directed at it from different angles, and refraction pattern and scattering are recorded. From this information, we deduce the electron density and relative position of the residues. This information is further interpreted by assigning observed structure to sequence.

2. Why did researchers develop methods for protein structure prediction?

X-Ray crystallography is expensive and time-consuming. The number of known proteins is growing exponentially, and number of known structures only linearly. Many new proteins already have a homolog (either by sequence or fold) with known structure, which makes predictions somewhat plausible and potentially useful. Scientific interest also lies in developing prediction tools in and of themselves, to test boundaries of our current knowledge and possibly identify unique structures without a homolog which would be prioritized for experimental structure establishment (not sure if that happens but it should).

3. Use the PDB Advanced Search and find out how many PDB entries that have had their structures determined via X-Ray Crystallography?

123125. From Experimental Method: X-Ray, Experimental data: Ignore

4. Search for Azurin in the SCOP database. Select the top hit. What class, fold and family do they belong to?

Class: All beta proteins
Fold: Cupredoxin-like
Family: Plastocyanin/Azurin-like (Copper binding site, mono-domain proteins)

5. Search for Azurin in the CATH database. Select the top hit. What class, architecture and topology do they belong to?

First matching CATH domain 3fpyA00.
Class: Mainly beta
Architecture: Sandwich
Topology: Immunoglobulin-like

6. Search for Azurin in the ECOD database. Select the top hit. What architecture, H-group and topology do they belong to?

Architecture: Beta-sandwiches
H-group: Cupredoxin-related
Topology: Cupredoxin-related

7. To make their web services more user-friendly, ECOD, SCOP and CATH have divided their protein folds into classes. Which classes do they have?

SCOP classes:

All alpha proteins
All beta proteins
Alpha and beta proteins
Mainly parallel beta sheets
Alpha and beta proteins
Mainly antiparallel beta sheets
Multi-domain proteins
Membrane and cell surface proteins and peptides
Small proteins

Coiled coil proteins
Low resolution protein structures
Peptides
Designed proteins

CATH classes:

Mainly alpha
Mainly beta
Alpha Beta
Few secondary structures

ECOD classes:

Alpha arrays
Alpha bundles
Alpha superhelices
Alpha duplicates or obligate multimers
Alpha complex topology
Beta barrels
Beta meanders
Beta sandwiches
Beta duplicates or obligate multimers
Beta complex topology
A+b two layers
A+b three layers
A+b four layers
A+b complex topology
A+b duplicates or obligate multimers
A/b barrels
A/b three-layered sandwiches
Mixed a+b and a/b
Few secondary structure elements
Extended segments
Special

Generally, classification is based on secondary structure arrangements

8. Which of the three databases are most up to date?

Hard to tell which one is more up to date, CATH has very recent snapshots but is updated often with verified proteins that were on pdb for a long time. Both databases do different amount of processing of the data.

CATH (12 days ago) although information might be limited. But latest extensive update 17th of May last year.
ECOD 17th of November 2017, updated with pdb structures up until March 2017.

9. Give a short description of how fold recognition (threading) generally works?

When attempting to predict protein structure, we sometimes do not find very high homology sequences with known 3d structure, which makes homology modeling an unreliable approach. However, we know that protein structure is much more conserved and much less diverse than sequence, and dissimilar sequences can still result in similar structure. Threading is based on evaluating the likelyhood of structure similarity (from statistical observations) instead of basing prediction on sequence similarity.

10. What is a template structure and how does one generally find it?

In threading, multiple template structures (known folds) are tested by aligning the query sequence along the structural templates, and determining the fold that is most likely(through various scoring methods). Known folds and their associated sequences are initially retrieved from PDB.

11. What do you think is a meta-server?

It's a server/service that integrates various tools built for the same purpose, and either shows you the general consensus, or combines the approaches.

12. Which methods did Pcons.net use to perform the homology modeling?

Blast and RPS-blast(reverse-PSSM search) to find homology and conserved motifs. If suitable template is found, STRIDE is used to assign secondary structure(from PDB) in template and PSIPRED to predict 2d structure (my motif) in query sequence. Then Pfrag attempts fragment arrangement into likely tertiary structure, evaluated by RSMD score and L-J potential. Quality of the final model is evaluated by ProQ (based on structure fitness and other model abundance) and Pmodeller (structure and statistics) and Pcons consensus based on scoring from other prediction servers.

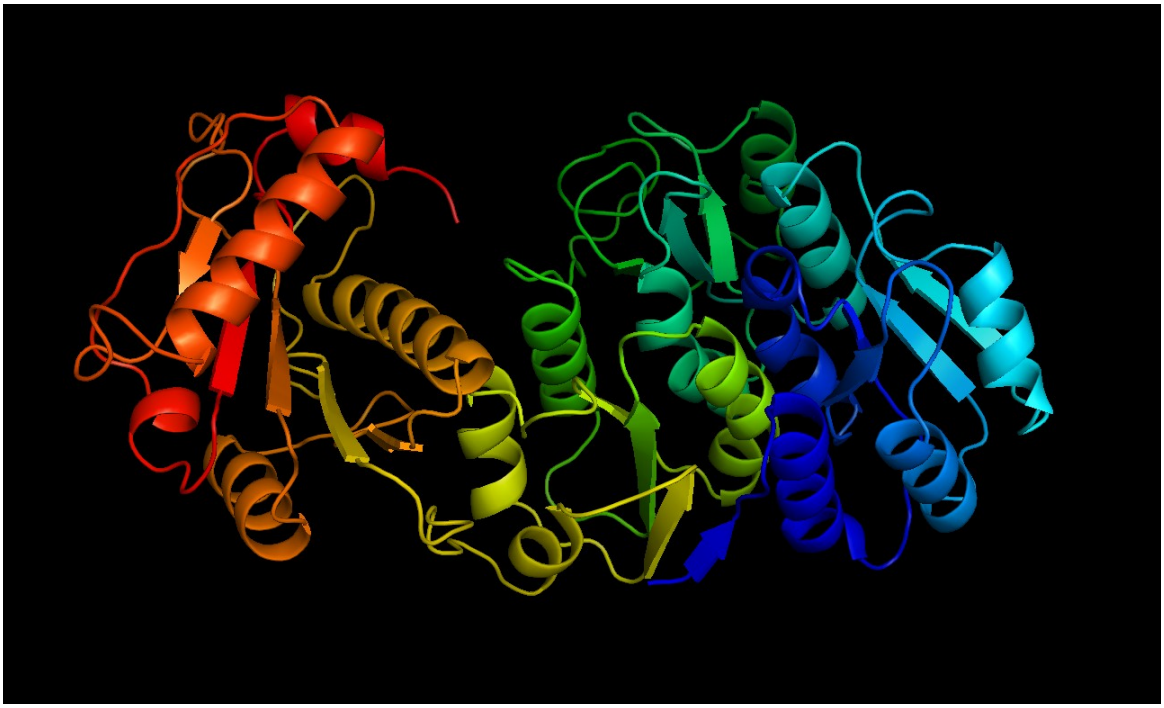
13. Give a short description of how homology modeling generally works?

It is a template-based method of 3D protein structure prediction. BLAST is used to find sequence homologs with known 3D structure which could be a good template candidate.
An alignment is generated with template sequences
Secondary structures are used to scrutinize the fitness of template.
A likely tertiary arrangement (core, side-chain location) of the sequence predicted. The 3D arrangement will be modeled after the conserved sequences of template fold, and the unconserved (loop modeling) ones based on other protein motifs, physical properties of residues, spatial predictions and so on. To improve the model further, different rotamers of side-chain residues might be attempted to find the energetically favorable/more likely prediction.

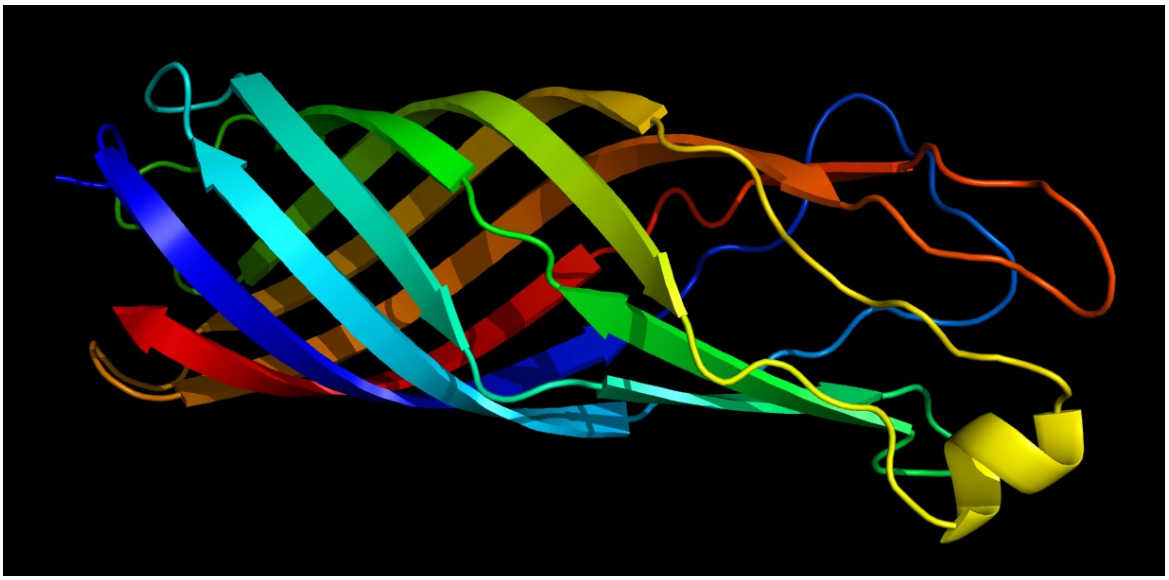
14. Look at the "Quality by SOLVX" tab what does it mean when the blue line is inside the red area?

Purple line – Solvx evaluates solvation profile and the quality of model based on that. If the solvation score of the residue at the same position in the model largely differs(inside the purple zone), it means hydrophobicity/hydrophilicity profile is not similar to template(probably bad model, or needs adjustment)

15. Download the pdb file, open it in Pymol, save an image and attach it to your report.



Pcons: best prediction



Modeller prediction

Attaching both for comparison: personally think the Pcons one is more likely to be correct, considering SOLVX analysis