

## Analyzing protein structures at the molecular level

A series of exercises for Biochemistry 523 course (96 students) by Karsten Theis in 2008

Student responses were graded semi-automatically using python scripts

**Structures for these exercises (each student is assigned one of these):**

1A0I : 1A82 : 1B0U : 1B76 : 1B8A : 1CSN : 1DV2 : 1DY3 : 1E24 : 1E2Q :  
1E4G : 1EE1 : 1FQ1 : 1G5T : 1GN8 : 1GOL : 1GTR : 1H3E : 1HCK : 1HP1 :  
1II0 : 1J09 : 1J7K : 1JI0 : 1JJV : 1JWA : 1KJ9 : 1KO5 : 1KP2 : 1KVK :  
1KXP : 1L2T : 1MAU : 1MB9 : 1MIW : 1MJH : 1N5I : 1NM1 : 1NSF : 1O9T :  
1OBD : 1OL6 : 1PHK : 1PX2 : 1Q97 : 1QHX : 1QRS : 1QRT : 1QRU : 1R8B :  
1RDQ : 1S9J : 1SU2 : 1TC0 : 1U5R : 1U5V : 1VC9 : 1VCI : 1VJD : 1WKL :  
1XDN : 1XEX : 1XKV : 1XNG : 1Y56 : 1Y64 : 1Y8P : 1YFR : 1YTM : 1YUN :  
1ZAO : 1ZYD : 2A42 : 2A84 : 2AQX : 2ARU : 2BIY : 2BTF : 2BU2 : 2BUP :  
2C01 : 2C8V : 2C96 : 2CBZ : 2CJA : 2DDO : 2DRA : 2DTO : 2E5Y : 2EG2 :  
2EWW : 2F02 : 2FAQ : 2FXU : 2GNK : 2HMU : 2IAJ : 2IJM : 2ILY : 2IVP :  
2IXE : 2IYW : 2J3M : 2JAX : 2NT8 : 2O0H : 2OGX : 2OH5 : 2OLR : 2PNN :  
2PQ1 : 2PZE : 2Q0D : 2Q66 : 2Q7G : 2Q97 : 2QB8 : 2QK4 : 2QUI : 2R7L :  
2R86 : 2R9V : 2YW2 : 2YWW : 2YXU : 2Z02 : 2Z08 : 2Z1U : 2ZDQ : 3BLQ :  
3BU5 : 3BUZ : 3C4W : 3C5E : 3C9R : 3CRC : 3CYI : 3PGK :

### Tools for these exercises

- Protein database: <http://www.pdb.org>
- Firstglance in Jmol: <http://molvis.sdsc.edu/fgij>
- Protparam (exercise 5): <http://ca.expasy.org/tools/protparam.html>
- Blast and other tools at NCBI: <http://www.ncbi.nlm.nih.gov/>

### Exercise 1 (warm-up):

- 1) What is the four character identifier (PDB ID#) of "your" structure? (Go to Spark and check your "grade comment" for exercise 1). You will work with this structure for all the Online Homework problems in this course. The example structure is "1D9Z".
- 2) What is the primary citation for that structure?
- 3) What is the last sentence of the abstract?
- 4) How might understanding this protein help society or be relevant to you?
- 5) Were there any technical problems in answering these questions?

## Exercise 2 (tertiary structure)

- 1) What is the PDB ID of your structure?
- 2) What are the chain identifiers of the protein chains?

### For the first Leucine residue after residue 50 in the first chain:

- 3) What is the residue number and chain ID?
- 4) What is the shortest inter-residue contact with the CB atom?  
(For questions 4-7, 11-13 and 15-17, give the residue number and atom type of the contacting atom, and the inter-atomic distance)
- 5) What is the shortest inter-residue contact with the CG atom?
- 6) What is the shortest inter-residue contact with the CD1 atom?
- 7) What is the shortest inter-residue contact with the CD2 atom?
- 8) Are these contacts mainly hydrogen bonds or mainly van der Waals contacts?
- 9) Is the side chain buried or accessible?

### For the first Arginine residue after residue 50 in the first chain:

- 10) What is the shortest inter-residue contact with the NE atom?
- 11) What is the shortest inter-residue contact with the NH1 atom?
- 12) What is the shortest inter-residue contact with the NH2 atom?
- 13) Are the contacts with the nitrogen atoms of Arg mainly hydrogen bonds or mainly van der Waals contacts?
- 14) Is the side chain buried or accessible?

### For the first Asparagine residue after residue 50 in the first chain:

- 15) What is the shortest inter-residue contact with the NE atom?
- 16) What is the shortest inter-residue contact with the NH1 atom?
- 17) What is the shortest inter-residue contact with the NH2 atom?
- 18) Are the contacts with the nitrogen atoms of Arg mainly hydrogen bonds or mainly van der Waals contacts?
- 19) Is the side chain buried or accessible?

### **Excercise 3 (secondary structure)**

For the first Glycine residue after residue 60:

- 1) What are the phi and psi angles?
- 2) What are the main chain hydrogen bonds, if any (use N+2, N+1, N-1 etc. nomenclature)?
- 3) What secondary structure is it in?

For the first non-proline non-glycine residue after residue 60:

- 4) What are the phi and psi angles?
- 5) What are the main chain hydrogen bonds, if any (use N+2, N+1, N-1 etc. nomenclature)?
- 6) What secondary structure is it in?

### **Exercise 4 (ligand binding)**

For the ATP molecule in the structure (if there are multiple ATPs, for the first one):

- 1) What is the PA-O-PB angle?
- 2) What is the O5'-PA-O3A angle?
- 3) Which atom of which amino acid residue comes closest to the adenine base?
- 4) What is the distance and nature of the contact?
- 5) Which atom of which amino acid residue comes closest to the ribose?
- 6) What is the distance and nature of the contact?
- 7) Which atom of which amino acid residue comes closest to the triphosphate?
- 8) What is the distance and nature of the contact?
- 9) Is the ATP buried or accessible?

### **Exercise 5 (protein purification)**

- 1) What organism is the protein from?
- 2) What organism was the protein expressed in?
- 3) Are there any disulfide bridges?
- 4) Are there any affinity tags added to the beginning or end of the sequence?
- 5) What is the estimated net charge of the first chain at pH 9?
- 6) What is the estimated net charge of the first chain at pH 5?
- 7) You receive purified protein dissolved in 1ml buffer. The absorption of this sample at a wavelength of 280nm is 5.6 at a pathlength of 1 cm. Estimate the protein concentration and the total amount of protein

### **Exercise 6 (sequence comparison)**

- 1) What is the closest human sequence (or yeast, if the protein is of human origin)?
- 2) How likely is it that the match found was by chance?
- 3) What structure in the pdb, if any, has the same sequence (or the most similar one)?

## **Alternative assignments 5 – 8: literature research and presentations**

Students worked in groups of 12, explaining the protein structure they were assigned, and picking the protein they were most interested in from another group of 12.

### **Assignment 5:**

Prerequ: Know your group (grade column)

ILE-5: Explain a technical term

- 1) Post your primary citation on RefWorks in the appropriate folder
- 2) Add you pdb ID in the User1 annotation
- 3) Explain a key technical term in the User2 annotation
- 4) Export and post on Spark

### **Assignment 6:** Find papers that mention structure

Prerequ: The populated RefWorks folders

ILE-6: Campaign for a publication

- 1) From a set of 12 proteins/primary citations, choose the one you like best
- 2) Post a reference related to the work on RefWorks
- 2) Write a short note about what the protein does, what the main finding was, and how it affects our lives
- 3) Post on Spark, including citations

### Assignment 7:

Prerequ: A discussion list with the 12 descriptions (anonymous)

ILE-7: Express your interest in a publication

- 1) Browse through 12 descriptions of proteins/citations, and select which you are most interested in
- 2) Formulate why you think your choice is interesting
- 3) Post your choice and your reason on Spark
- 4) If you wish: post your choice/reason on the discussion board
- 5) Be prepared to decide which protein should go into quarterfinals

Model for 3 sentences to introduce protein, structural insight and relevance

"TRPV1 is involved in nociception, leading to the burning sensation due to heat or pH with the potential to damage tissue. The ankyrin repeat domain of TRPV1 contains a binding site that can bind ATP and regulate the sensitivity of the channel. ATP can prevent decreasing response of the channel, and calmodulin can increase the response of the channel due to capsaicin. The mechanism for the regulation of the channel may lead to the ability to prevent the painful burning sensation in humans. "

In class, pick the structures the class is most interested in using 3 rounds of elimination:

First round (In class group work):

- 1) Discuss contenders
- 2) Decide on a quarterfinalist

Quarterfinals (decided by clicker PRS):

- 1) Listen to presentation on the 8 quarterfinalists
- 2) Using PRS, choose your favorite (4 best go forward)

Semifinals (decided by clicker PRS):

- 1) View four structures
- 2) Using PRS, choose your favorite (2 best go forward and get presented in class)

### **Assignment 8: Feedback**

This was the first time we had independent learning exercises, and I would like to get your feedback. As a reminder, here are the ILEs and the associated work in class:

- ILE-1 SPARK and the protein data bank (warmup)
- ILE-2 Secondary structure (jmol)
- ILE-3 Interactions between side chains (jmol)
- ILE-4 Active site: ATP interactions (jmol)
- ILE-5 Online textbooks (intro to reworks)
- ILE-6 Literature search (pubmed and Web of Science)
- ILE-7 Big picture exercise (choose a protein)
- Preliminaries (peer review)
- Quarterfinals (big picture)
- Semifinals (protopedia)
- Final talk 1 (viral DNA packaging)
- Final talk 2 (multi-drug resistance)

- 1) Which exercise was your favorite one? Explain.
- 2) Which one was the most difficult for you? Explain.
- 3) Which exercise did you feel is most relevant for your career? Explain.
- 4) How much time did you spend working on the exercises (average time per ILE)?
- 5) Which aspects did you like about the exercises, and which aspects might be improved?
- 6) Overall, how much did the ILEs enhance your learning experience in the course? Please choose from the following: 5 - greatly 4 - somewhat 3 - neutral 2 - not at all 1 - they were distracting