Andrew Torda
Nils Petersen

Zentrum für Bioinformatik
Übung zur Vorlesung
Grundlagen der Strukturanalyse
Wintersemester 2019/2020

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

16.12.2019

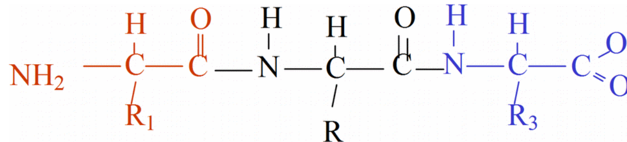
Übung 7: Revision 1

Dies ist die erste von zwei Übungen, welche Ihnen die Prüfungsvorbereitung erleichtern sollen. Auf den folgenden Seiten finden Sie typische Fragen, wie sie in einer Klausur gestellt werden könnten. Dies ist aber kein Fragenkatalog, sondern nur eine kleine Sammlung möglicher Prüfungsfragen. Im Gegensatz zu dieser Übung werden die Prüfungsfragen der Klausur auf Deutsch gestellt.

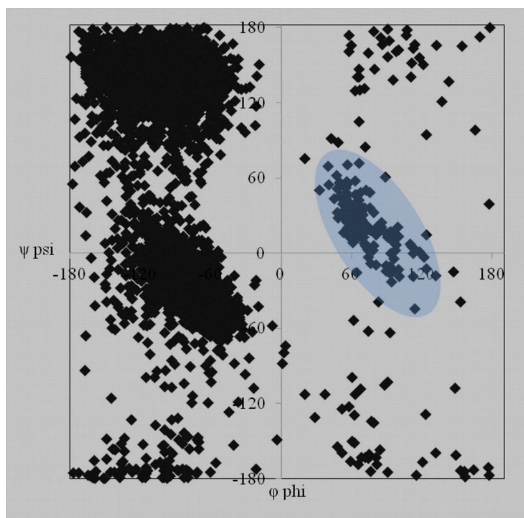
Bitte beantworten Sie mit Hilfe Ihrer Übungs- und Vorlesungsunterlagen alle gestellten Fragen und bringen Sie Ihre Antworten für die gemeinsame Besprechung in der Übung am **16.12.2019** mit.

Fragenblock 1 (Protein Structure):

- What order of magnitude is a chemical bond (in Å)?
- On the diagram, mark the two backbone angles which can rotate in a normal protein. You only need do this for one residue.



- Mark the angle which is nearly planar (flat).
- Why can I not have a short α-helix which is only 2 residues long?
- Name a large hydrophobic amino acid, a small amino acid and a polar (but uncharged) amino acid.
- Name the amino acid which often forms covalent bonds from its side-chain.
- If you consider a Ramachandran plot for a protein, there is a region where only one type of amino acid is found, marked on the diagram by the grey oval.
  Which amino acid is this? Why can it occupy this area?



- Why can proline not be part of a perfect α-helix?
- For length *N*, do all / many / few of the 20^N sequences form specific structures?
- You synthesized a 100 amino acid long protein consisting entirely of aspartate: $(Asp)_{100}$
  What are its properties? Is it soluble? Is it acidic? Is it basic?
  Would it form a compact regular structure?
- If you have a protein of poly-Trp (polytryptophan), would it form a specific structure? How would it behave in solution?
- Why would you want to represent a protein by its surface?
- Why might you draw a protein as a ribbon representation in *Chimera*?
- What is the biggest chain in the protein data bank?
  What is the average size of a protein?

Fragenblock 2 (Distance Geometry):

- Draw three atoms with distances between them, which are not possible in 3-dimensional space.
- Aside from experimental distance information, what information does one add to a metric matrix distance geometry calculation, before applying the triangle inequality (bound smoothing).
- Why is the triangle inequality applied twice during a metric matrix distance geometry calculation?
- In the metric matrix distance geometry method, one generates a trial matrix. Imagine you have no experimental errors. All your distance measurements are correct to 10^{-20} m. Would you expect the trial matrix to correspond to a single set of 3-dimensional coordinates?
- What is the running time of the bound smoothing step in the metric matrix method? Explain in one sentence.
- You use the metric matrix method to calculate the structure of a protein, but you do not have any experimental data.
  What would you expect if you generate 20 structures?
- In a distance geometry calculation, I have a set of atoms *i-j-k-l-m-n*. What stops atoms *i* and *n* ending on top of each other?
- The following matrix contains upper bounds on distances. Draw a graph that corresponds to this distance matrix.

	A	B	C	D	E
A	0		4		
B		0	2	5	3
C			0	2	1
D				0	
E					0

  Add the missing values to the distance matrix and the corresponding edges to your graph. Is there some value in the distance matrix that can be reduced?
  What is the shortest path between points D and E?

- Name an advantage of the variable target function method, compared to the metric matrix method for distance geometry.
- What is the running time of the variable target function method?

Fragenblock 3 (NMR):

- How is uncertainty in protein coordinates from NMR represented?
- Name three elements, with the correct nuclei, which are relevant to biochemistry and NMR.
- When calculating a protein structure based on NMR data, what information does one get from the size of a *J* (spin-spin) coupling constant?
  Which atoms are involved?
- Why are only some values of the coupling constant useful?
- Which experimental phenomenon provides most of the structural information for determining a structure by NMR?

Fragenblock 4 (Crystallography):

- What is the wavelength of an X-ray (order of magnitude / Potenzordnung)?
- Explain why one cannot simply apply a Fourier transform to the measured reflections to obtain electron density.
- When one refines a crystallographic structure
  - What are the variables?
  - What is the cost function?
- What is the difference between static and dynamic disorder in protein crystals?
- In protein crystallography, there is a model for the uncertainty of atomic coordinates. What kind of distribution is it based on?
- This distribution is based on a model for the energy. What is that model?
- A crystal structure represents a space and time average. Explain.
- An R-factor in protein crystallography is given by $R = \frac{\sum_{hkl} \left| |F_{hkl}^{obs}| - |F_{hkl}^{calc}| \right|}{\sum_{hkl} |F_{hkl}^{obs}|}$

  What is R measuring? When is the formula used?
- What is the R-factor for a perfect model?
- What is the difference between R and R_{free} ?
- To determine phases for a crystallographic data set, one might use molecular replacement. Does it require any modification to a protein sample? What does one need?
- What does one need for multiple isomorphous replacement?
- I have a data set with 1.5 Å resolution, but I claim the uncertainty in coordinates is less than ½ Å. Why might this be reasonable ?

Fragenblock 5 (Structure Analysis and Comparison):

- Why is it fundamentally difficult to superimpose two protein structures if they are not the same size?
- I have two models of one protein, but they are rather different.
  Describe an algorithm with pseudo-code to find the more similar regions of the structures.
- I have two proteins and an effective algorithm to find the common region between two protein structures. When I run the program I find the following alignment:

	residues			
protein 1	1-10	11-60		61-90
protein 2		1-50	51-70	71-100

  So, for example, residues 11-60 in protein 1 are aligned to 1 to 50 in protein 2.
  Sketch a diagram of domain structure that would give this alignment.
- You would like to align protein structures of different sizes and you would like to turn the problem into a classic dynamic programming formulation. Describe one method for this.
- Similarity of protein structures is often measured using the root mean square difference of coordinates. Draw an example to show why this may not be a good measure.
- Describe a measure of protein similarity which is quantitative (in Å), but is not the root mean square difference (*rmsd*) of Cartesian coordinates.
  Why may it be better than *rmsd* of Cartesian coordinates.
- Over the course of evolution, which changes faster – protein sequence or structure? Give a reason why this may be the case.
- Some protein structure classifications impose a hierarchy on proteins.
  Why may this be a reasonable thing to do?
- Give an argument why a hierarchical classification may not be appropriate for many proteins.