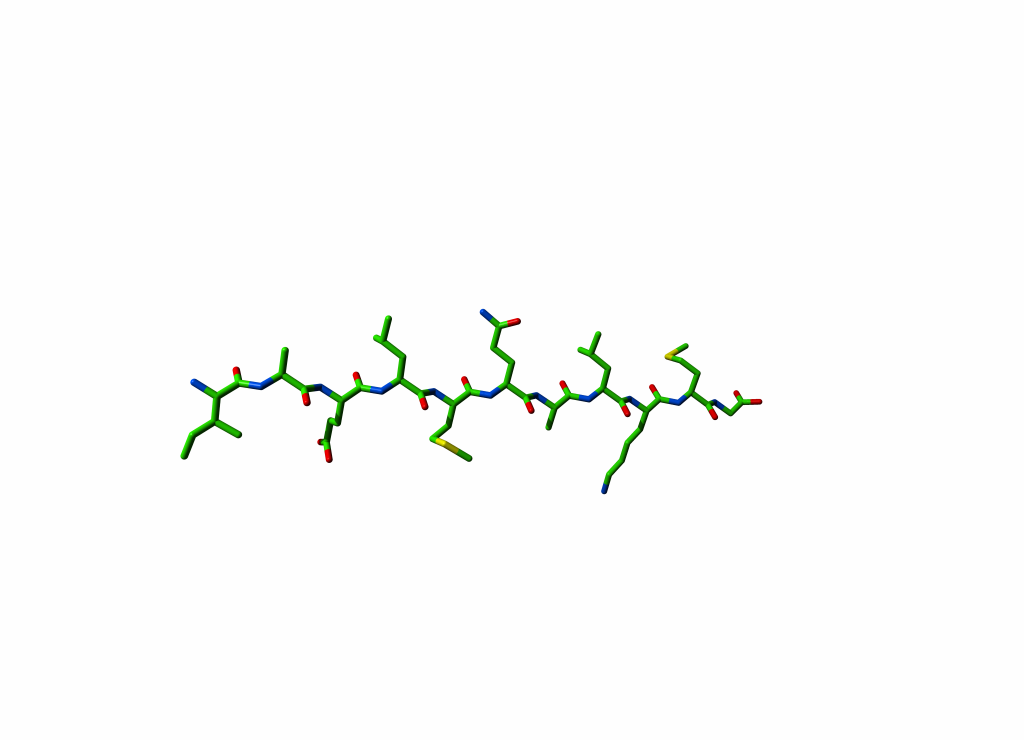
SFB 2014-2105. Name:

Feel free to answer in Dutch, German, or English (or even mixtures of these languages). A dictionary is not allowed but Gert will come to the exam room around two hours into the exam to translate if needed. Please write clearly and legibly.

1) a) Write the one-letter and three-letter code for each amino acid close to its Cα in the figure below (the one atom that is a bit hard to see is red…).



b) What is a torsion angle?

c) Indicate the torsion angles Φ, Ψ and Ω, on the Glutamine in the picture above.

d) This peptide is drawn fully extended. However, when we synthesized it, and measured its CD spectrum we saw that it had a regularly folded structure about 50% of the time. Which structure did we see?

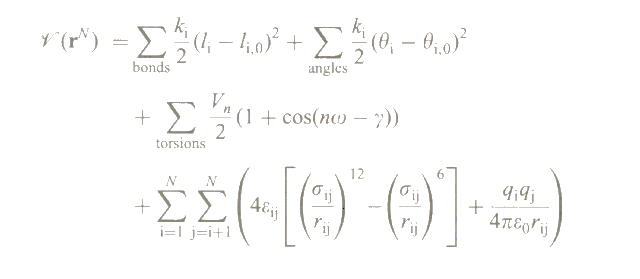
e) Why did we see that regular, folded structure only 50% of the time and not all the time?

f) Suppose we want to see the regular, folded structure 75% of the time instead of 50%, and suppose we can easily mutate residues in this peptide, which mutations do you suggest?

2) When we design mutations we spend a lot of time thinking about the precise geometry of hydrogen bonds, but when we study salt bridges we tend to just look if the charged groups are at least somewhat close together, and if they are we are happy. Why is this? Why does every 0.1Å count for H-bonds while with salt bridges 4 Å is about as good as 5 Å?

3) These formulas you have seen before. Do you remember the meaning of the terms listed below?

The correct answer is of course "Yes", but I guess you won't score well with that answer :-).

ki

rij

qi

θi,0

N

4) How do cysteine bridges stabilize proteins? I.e. where do the δH, and δG come from?

5) A yeast small molecule uptake receptor has a sequence that starts with

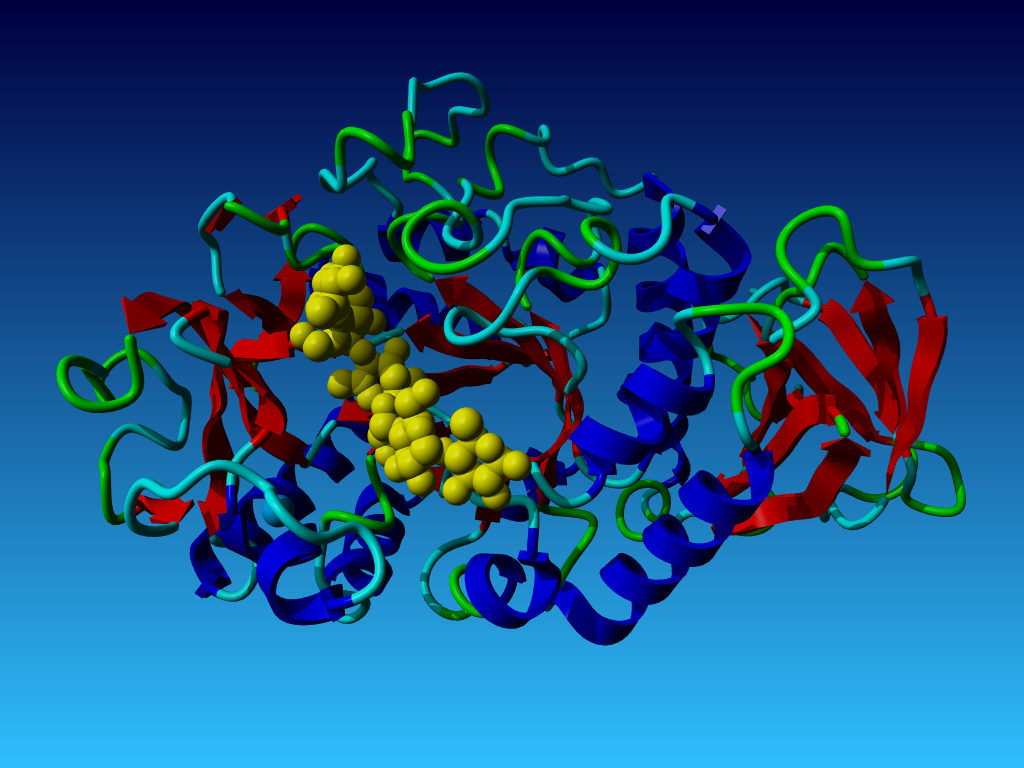
GKNRSKNLLLAILWFLSLLALIMLFFACWLLAINGDSDNG………..

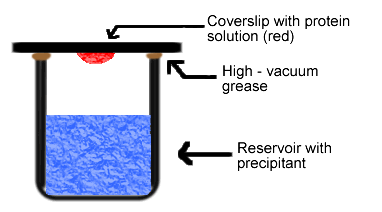
This is also the fragment that is most often found in proteomics experiments. Sometimes a variant of this fragment is found that is about 80 Daltons heavier, from which I conclude that it is phosphorylated. Circle the phosphorylated Serine. Explain your answer.

6) Why is water important for the function of a cell? Mention at least 5 processes that crucially depend on water (and just ‘the entropy of water’ doesn’t count, this time :-)).

7) In the following 4 figures you each time see a yellow ion. These ions are: Calcium, Sodium, Vanadium, and Zinc. Which is which?

|  |  |
| --- | --- |
| ca.png | Na.jpg |
|  | Zn.jpg |

8) During the course, you have seen the picture below. It is taka-amylase, a sugar cleaving enzyme. The yellow 'thing' is a short sugar chain. Taka-amylase likes to cleave just these sugars. But here it nevertheless remained intact for weeks and weeks (the time it takes to grow crystals, travel to the synchrotron, and collect the data).

This second figure comes from my own dark past as crystallographer. The red thing is a drop of water with the protein we wanted to crystallize and some protecting salts and sugars in it. The red drop hangs by cohesion forces on the bottom of a very thin glass plate. Vacuum grease is used because it doesn't let any CO2 go through; the space between the drop and the precipitant fluid contains normal air and is not vacuum, of course. The red drops typically contained 10 micro-liter, and after 3-5 days we would in the best cases find a few (small) crystals in the drop (the drop wasn't red of course, I only coloured it red in this picture).

a) How come we find crystals in the red drop after a few days? To answer this question, think about the precipitant, what is in there? And realize that 'the entropy of water' is an essential part of the answer to this question.

b) What could the crystallographers have done so that taka amylase doesn't cleave this sugar? Give me two suggestions.

b) Mention at least three different roles for sugar(s) in a living system.

c) One of the answers to sub-question b is that sugar chains are connected to the surface of many proteins for recognition and protection. These sugar chains thus seem important and should be studied well. But why do we nevertheless know so little about the structure of sugar chains that are bound to proteins?

d) Can you gamble which residue types make up the active site of taka-amylase?

9) a) The word ‘rotamer’ has multiple meanings. Which?

b) Tools exist to extract from the PDB so-called rotamer distributions. What is that (a rotamer distribution)? Mention at least two application areas for rotamer distributions.

10) At this moment our troops roam around in Mali. Mali has many, many scorpions. So we need a vaccine against scorpion toxin for the poor soldiers who are there trying to protect the locals against the locals. Can you tell me how you would go about designing this vaccine? Especially mention all software used in the whole process, and what each piece of software does (or if you never used it yet, what it should do)?

11) In the previous question you undoubtedly wrote something about force-field-based software that can predict the antigenicity of a peptide or protein. Such software exists already. But let’s assume for a few minutes that it didn’t exist yet, how would you design it? What data would you need? what algorithm(s) would you write (or have somebody write for you)? How would you analyse the performance of your force field based software?

12) The protein listed below is found in the genome of the bacterium horrivus sujeirus. This peptide is the main reason why people can be very allergic to bytes of insects that carry this bacterium in their saliva.

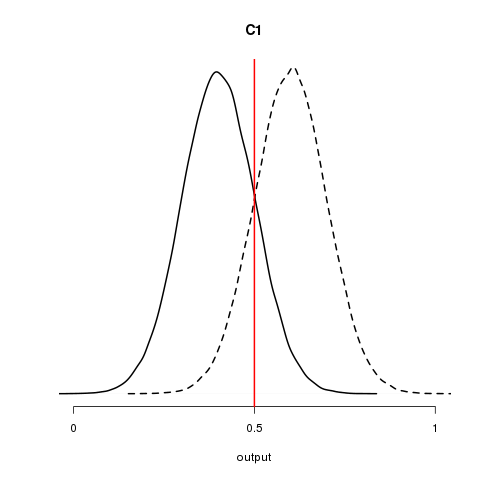
TLSATWCSGLWILAMVLLAMILSLAMVVLAMTSLACGNAWTAQAGGKNRSKTLLLAILWYLSLLALIMLFFACWLLAINGDSD

The department of immunology of the Radboud hospital has made a nicely working system that makes very good antibodies against peptides of length 9 that have a tryptophan at position 5. Can you tell them which nona-peptide to use in this bacterium horrivus sujeirus protein to make antibodies that will reduce the suffering of the allergic people? And why do you suggest just that peptide and not any of the other possibilities?

13) a) Mention a few terms (formulas not really needed, but allowed) that are used in the force field of a molecular dynamics software package. Feel free to look at question 3 first ☺

b) And mention a few terms that generally are not yet in use in such a force field? And explain why they are not used.

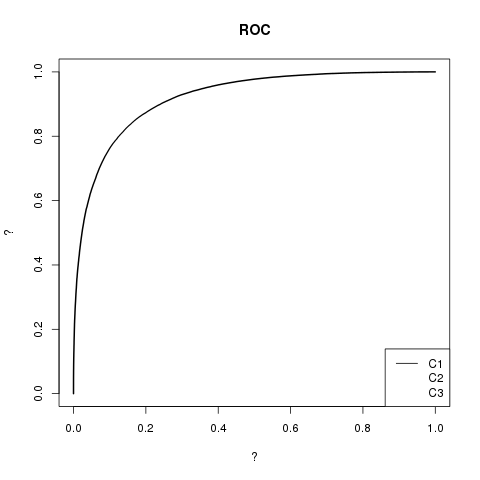
14a) Suppose you have created a two-class classifier for a bioinformatics classification problem. The output of this classifier for the negative (solid line) and positive (dashed lines) classes is shown in the figure below. The red line indicates a decision threshold.



Given the decision threshold, indicate in the figure the true negative, true positive, false negative, and false positive fractions.

b) Macromolecular structure validation software criticizes structures solved by crystallographers and NMR spectroscopists. Suppose the output of a structure checking algorithm is shown in the figure above. Do you think the decision threshold needs to be set differently? Why (not) and how (not)?

c) Suppose you created two other classifiers C2 and C3 in addition to the first classifier C1 mentioned above. C2 performs better than C1 and C3 performs worse than C1. Sketch the possible output of C2 and C3.

The Receiver Operating Characteristic (ROC) curve for C1 is shown in the figure below.

d) Explain what is on the x-axis and what is on the y-axis.

e) Indicate in the ROC plot where the red decision threshold is (approximately).

f) Sketch two possible ROC curves for C2 and C3 in the ROC plot.