Exploring Bioinformatics: A Project-Based Approach

Key for Web Exploration Questions, Chapter 11

1. The HIV protease functions as a dimer. Some enzymes that form dimers then have two active sites. Is this the case for the HIV protease? Briefly describe the relationship of the active site and peptide-binding cleft to the subunits of the enzyme.

   It is not the case for the HIV protease that there are two active sites. In this case, the peptide-binding site is at the interface between the two subunits, so that each subunit contributes essentially half of the "pocket" where the peptide binds and the "flaps" that surround it.

2. What kinds of amino acids do you find in the areas of the protein exposed to the water around it (e.g., when the protein is in solution in the cytoplasm)?

   Almost all of the amino acids that are on the surfaces of the protease structure are hydrophilic amino acids.

3. If you were to design an inhibitor of the HIV protease, where would you want it to bind? What kind of molecule might you use as the prototype to develop the structure of a good inhibitor?

   One good way to inhibit the polymerase would be to block its ability to bind a protein, perhaps by binding an inhibitor in the active site. (Potentially the shape of the active site could also be affected by binding an inhibitory molecule in some slot near the active site that would change the shape of the active site.) Because the protease naturally binds proteins, a peptide that could bind in the active site would be a good "template" for designing a molecule to fit here, even though you probably would not want a protein as your actual inhibitor due to proteins' instability in the body.

4. Using the cartoon or ribbon view, you should be able to identify where a long β-strand on each subunit of the protease makes a hairpin turn, forming flexible flaps that cover the active site cleft. These flaps control access of the substrate to the active site. Which amino acids form the flaps (just give the range of numbers)? Although this region is very important to protease function, why are the flaps not likely to make a good target for rational drug design?

   The "flaps" are amino acids 42-58 or so of each subunit. The problem with using this as a target for a drug is that this long loop is likely to be very flexible, so it is quite unlikely that binding something to it could significantly change the shape of the molecule.

5. What are the numbers of the amino acids on each chain that form the Asp-Thr-Gly (Asn-Thr-Gly in this mutant) aspartate protease motif in 1KJF?

   The protease motif is Asn 25, Thr 26 and Gly 27 on each subunit.

6. How many mutations are there in the mutant protease sequence, as compared to the sequence of the protease you examined in Part I? Use pairwise alignment to find out.

   There are 13 amino-acid changes in the mutant sequence, relative to 1KJF:

```
mutant: PQITLWQRPIVTIKIGGQLKEALLNTGADDTVLEEVNLPGRWKPKLIGG
  1kjf: PQITLWKRPLVTIRIGGQLKEALLNTGADDTVLEEMNLPGKWKPKMIGG

mutant: IGGFVKVRQYDQVPIEICGHKVIGTVLVGPTPANVIGRNLMTQIGCTLNF
  1kjf: IGGFIKVRQYDQIPVEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
```

7. In the regions you highlighted, how would you characterize the effect of the mutations on the structure of the protein, in general?

   Generally, the original protease structure fits the peptide substrate more tightly than the mutant structure. The amino-acid changes appear to have the effect of bending regions near the active site outward, enlarging the binding pocket.

8. How do you think these structural changes would affect the binding of a small inhibitor molecule to the protease active site? Why do you think they have less effect on the binding of the natural substrate?

   Since the binding pocket is enlarged, it is less likely that an inhibitor molecule would bind tightly. The changes would have the same effect on a natural peptide substrate, but this larger substrate likely fills the pocket better and is more difficult to dislodge.

9. If you wanted to design a drug that would inhibit this mutant protease, what characteristics would you want it to have?

   The new drug would need to be shaped so as to interact tightly with the altered regions of the mutant protein.

10. Change the colors of your model so that everything is white except the three amino acids of the aspartyl protease motif (make the substrate gray for contrast). Make these three amino acids blue on the mutant chains, then see what happens when you color them red on the non-mutant chains. Does their position change in the mutant relative to the unmutated protein? Is this what you would expect? Certainly changes in the sequence or structure at these positions could lead to drug resistance; why then do we not observe them among drug-resistant HIV isolates?

    The mutated and unmutated protein structures are extremely close together in this region: so close that in fact they may be indistinguishable depending on the view chosen. This makes sense because the protease will be non-functional if these key active-site amino acids are significantly displaced. We thus don't observe drug-resistant variants that have large differences here because they would also be non-functional and therefore be strongly selected against.

11. How well did PSIPRED predict the secondary structures in the HIV protease? Give specific examples of structures predicted accurately by PSIPRED, predicted

structures not found in the actual structure and actual structures that were not predicted.

The PSIPRED prediction for this protein is really very accurate. As shown in the table at right, nearly all of the structures found by PSIPRED were actually present in the crystal structure—even some unlikely-seeming short β-strands. There were some small differences in where these structures begin and end, however. The major difference is that PSIPRED predicted an α-helix at amino acids 20-25 that does not exist in the crystal structure and is in fact part of an adjacent β-strand. PSIPRED also split a long β-strand from 51-67 into two pieces. On the whole, however, one would have to say that this *de novo* prediction is quite accurate.

structure	PSIPRED	crystal structure
b-strand	3-4	2-5
b-strand	10-15	10-16
b-strand	18-19	17-25
a-helix	20-25	
b-strand	32-35	32-35
b-strand	45-48	42-49
b-strand	52-59	51-67
b-strand	62-67	
b-strand	70-77	69-78
b-strand	84-85	84-85
a-helix	86-93	86-94
b-strand	96-97	96-99

12. PSIPRED uses a prediction algorithm not unlike the Chou-Fasman algorithm we will use in the Programming Project. However, instead of applying its algorithm directly to your input sequence, it first does PSI-BLAST search to get a collection of sequences related to your input. It then applies its prediction algorithm to the results. Why might this method be advantageous in improving the program's ability to identify genuine secondary structure?

    If a given amino acid is part of an actual secondary structure, such as an α-helix, then we would expect that mutations that change this to an amino acid with poor α-helix potential would be selected against and amino acids with good α-helix potential would tend to be conserved at this position. By using a sequence comparison to look at conservation prior to predicting structures, PSIPRED can "filter out" amino acids that don't genuinely contribute to structures and thus improve its accuracy.