Chapter 1 Odd Solutions

1. Asp and Glu are negatively charged; Lys and Arg are positively charged; Ser, Thr, and Tyr are neutral. His has a pK_a close to 7 or just below, so it is more likely to be neutral at pH 7 though it could easily be protonated, depending on its pK_a. At pH 7, only one molecule in 1000 (0.1%) of Glu is protonated.

3. For example, see **Figure 1**, Protein G B1 domain (PDB file 1pga). The protein is color coded from blue at the N terminus to red at the C terminus.

The question is mainly aimed at getting students to gain some familiarity with software for displaying proteins. These three views are popular and useful views, by far the most common being the first. Although the cartoon is by far the easiest to follow and understand, it grossly misrepresents what a protein actually 'looks' like. The other two representations are better representations of this, as stated in the text.

Visible light has a wavelength of about 0.5 μm, which means that objects this size or smaller diffract light rather than block or reflect it. Therefore we cannot 'see' anything as small as a protein. But if we could, the protein would look most like the spacefilling models, and be colorless.

5. (a) Glycine has no side chain; therefore many of the conformations that are impossible to other amino acids, because their side chain would bump into the backbone carbonyl or the adjacent residue, are allowed. Proline has a much more limited conformational space than other amino acids because the side chain is cyclized to the backbone, which restricts the ϕ angle to a region around –65°. (b) At the bottom left, at around $(\phi, \psi) = (-64, -41)$, is the α helix; on the right, at around $(\phi, \psi) = (+60, +60)$, is the 'left-handed α-helix' region, which is populated almost entirely by glycines in turns and to some extent by other small residues in turns, mainly Asn and Asp. A genuine 'left-handed helix' is never seen in real proteins. At the top left is the β-sheet region, which interestingly is divided into two parts.

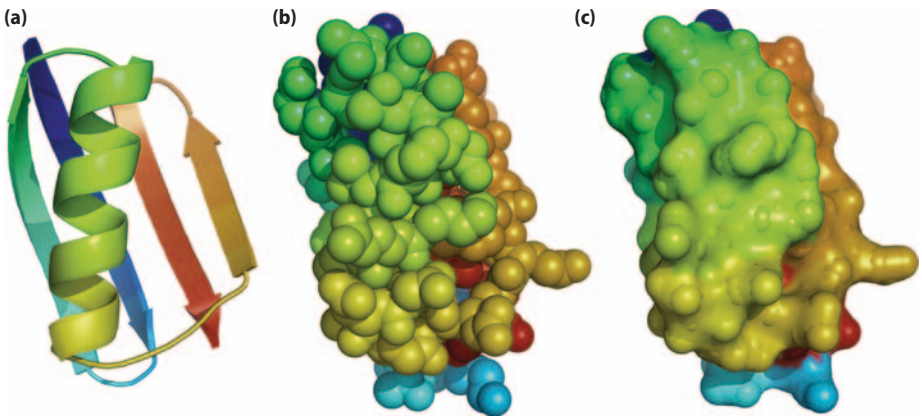

FIGURE 1
Three different views of the same protein: Protein G B1 domain, color coded from blue at the N terminus to red at the C terminus (pdb file 1pga). (a) Cartoon, showing β sheet, α helix and loops. (b) Space-filling representation, in which each heavy atom is shown is shown as a ball with radius corresponding to the van der Waals radius of the atom. (c) Surface representation, showing the solvent accessible surface.

The left part centered at (ϕ, ψ) = (–121, +128) contains mainly residues actually in β sheets, either parallel or antiparallel, and the right part centered at (ϕ, ψ) = (–66, +137) is often called polyproline II and is populated by a high proportion of residues in extended conformation that are not in regular β sheet. [See S. Hovmöller, T. Zhou and T. Ohlson, *Acta Cryst. D* 58:768–776, 2002.]

7. There is a nice section on the mechanism of serpins in A.M. Lesk [Introduction to Protein Science, 2nd ed. Oxford: Oxford University Press, 2010, pages 216–219]. Possible advantages of this major conformational change are that (a) it provides energy to drive the irreversible cleavage, (b) the change sterically hinders cleavage of the link, (c) by changing structure upon binding a substrate, the serpin may function as a signaling molecule, (d) the change is irreversible, which may have some functional benefit, and (e) in some cases (as discussed below) the change occurs in the absence of substrate to deactivate the serpin.

Much of the following is based on an excellent set of Wikipedia pages.

There are now a very large number of serpin structures, in different conformations. Many of them are protease inhibitors, but some have unrelated functions, for example ovalbumin (from egg white). Serpins have three β sheets (A, B, and C; see **Figure 3**) and eight or nine helices. A key part of the structure is the reactive center loop (RCL). In the protease inhibitors, this is where proteolytic cleavage occurs: it is therefore the main site of recognition for specific proteases. The loop can exist in three conformations. The initial state is shown in Figure 3a and is also called Stressed (S) to highlight its metastable nature. The loop can also partly insert into the A β sheet (Figure 3b): these two forms are in equilibrium. After cleavage of the RCL by a protease, the entire RCL can insert as an extra strand into sheet A (**Figure 4**), to form the Relaxed (R) conformation.

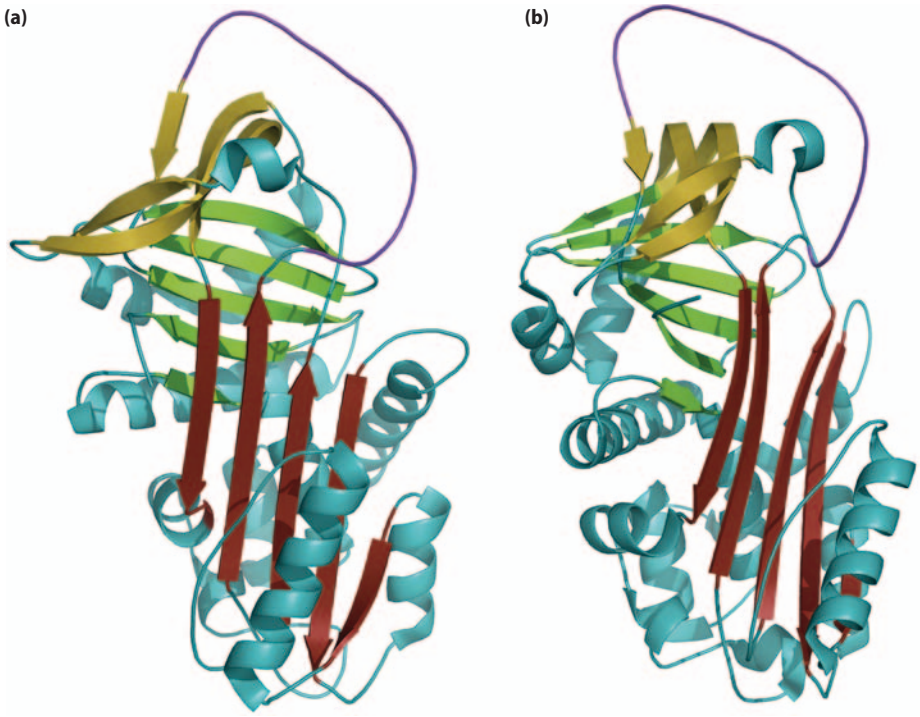

FIGURE 3
The structure of serpins: (a) human antitrypsin (PDB file 1qlp) (b) murine antichymotrypsin (PDB file 1yxa). In both cases β sheet A is in red, B in green and C in yellow. The Reactive Center Loop is in magenta. In (b) two amino acids of the RCL are partially inserted into the top of β sheet A, forcing the adjacent strands apart.

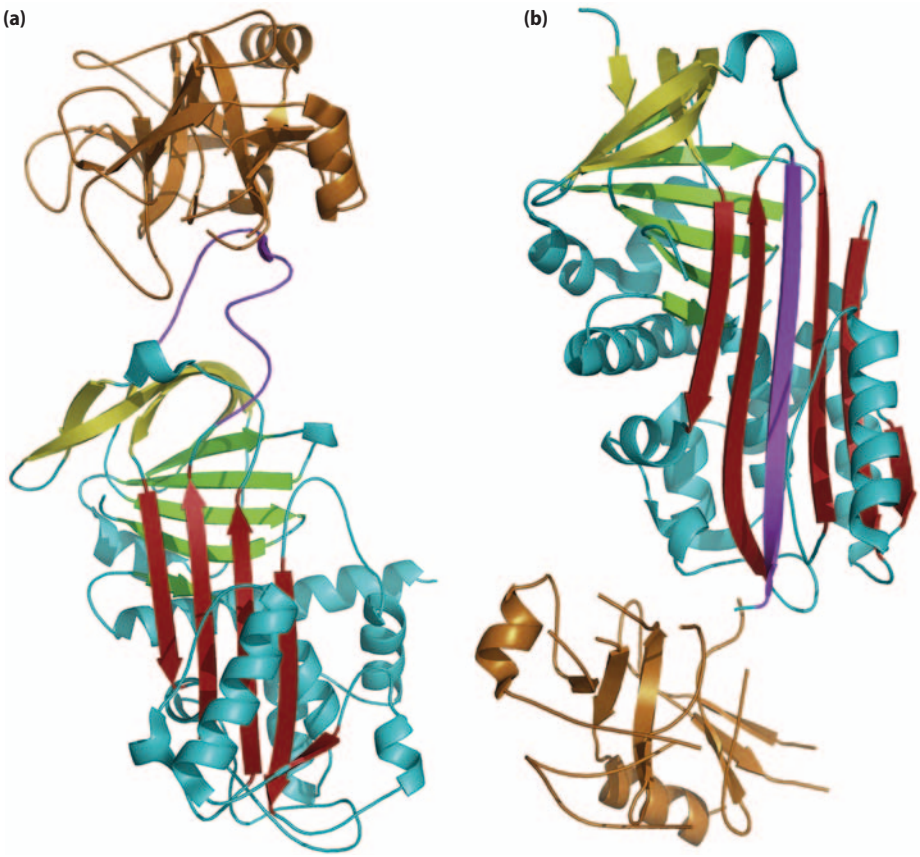

FIGURE 4
Structural rearrangement of serpins following proteolytic cleavage. (a) Structure of the non-covalent complex between insect serpin 1K and an inactivated rat trypsin (brown) (PDB file 1k9o). (b) Structure of the covalent complex between antitrypsin and active trypsin (PDB file 1ezr). Note how the RCL (magenta) has completely altered its conformation and inserted itself in the center of β sheet A, taking trypsin with it.

Serpins therefore work in a fundamentally different way from most protease inhibitors. Most inhibitors merely block the active site of proteases, forming 1:1 noncovalent complexes in a reversible manner. Serpins, however, react to give the acyl intermediate, an irreversible covalent attachment. In normal protease cleavage, the acyl intermediate is rapidly hydrolyzed to regenerate active protease, but when formed with a serpin, the acyl intermediate undergoes the structural rearrangement described above, which moves the protease from one end of the serpin to the other and distorts its conformation, which thereby deactivates the protease.

Some serpins spontaneously undergo the transition from S to R to form a permanently deactivated inhibitor called the latent state. This is a control mechanism for some serpins. For example, PAI-1 will become latent unless it is bound to its cofactor vitronectin; interactions of other proteins with vitronectin therefore inactivate PAI-1 irreversibly.

In several serpins, the equilibrium between S and partly inserted is also used as a control mechanism; this occurs for example in antithrombin and heparin cofactor II, which are both important in blood clotting. Antithrombin is normally found as the partly inserted form, which is less active than the S form. This is because the arginine immediately preceding the cleavage site (P1), which is the key residue for determining protease-specific binding, is buried (**Figure 5**(1)). Interaction with heparin (Figure 5(2)) pulls Arg P1 back out, thereby activating antithrombin. Further activation is provided because heparin acts as a scaffold in also binding to the protease targets of antithrombin, thrombin, and factor Xa (Figure 5(3)), thereby bringing the two proteins together in an appropriate conformation for

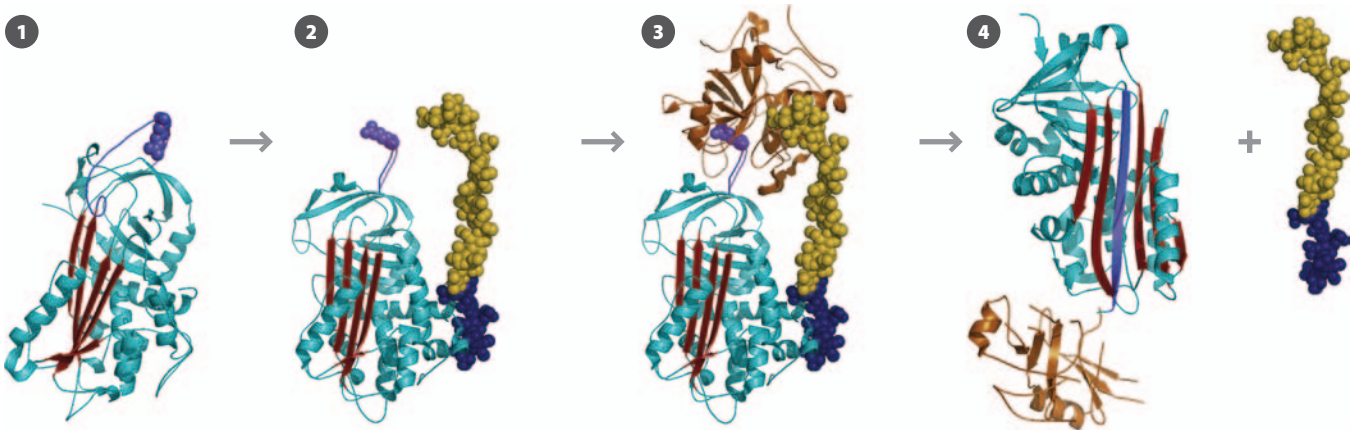

FIGURE 5
Activation of antithrombin by heparin. 1. The partly inserted conformation, in which Arg P1 (magenta spheres) is buried. 2. Heparin (spheres on right) binds to antithrombin mainly by a high-affinity site (blue spheres, the binding site for fondaparinux), and releases Arg P1. 3. Heparin also acts as a scaffold to bind thrombin. 4. This leads to cleavage of antithrombin, rearrangement of the RCL, and inactivation of thrombin. It also leads to the release of heparin.

reaction (Figure 5(4)). An understanding of this process led to the development of a synthetic form of heparin called fondaparinux, used as an anti-clotting drug.

9. Type I synthetases recognize Leu, Ile, Met, Val, Tyr, Gln, Glu, Cys, Trp, and Arg; type II synthetases recognize Phe, Ser, Pro, Thr, Ala, His, Asn, Lys, Asp, and Gly. Thus the smaller amino acids are recognized by type II synthetases, but otherwise there seems to be little logic. Lysine is ligated by either a type I or a type II enzyme in different organisms, the type I enzyme being confined to most archaea and a few bacteria. The two types have very different structures and mechanisms. Type I aminoacylates at the 2′-OH of an adenosine, whereas type II aminoacylates at the 3′-OH (except phenylalanine synthetase, which has a type II structure but aminoacylates at the 2′-OH). They even approach the tRNA from different sides. However, there is no obvious way to rationalize which one binds to which target. This seems to be a frozen accident of evolution. See A.M. Weiner [*Curr. Biol.* 9:R842–R844, 1999] for a fascinating discussion of the evolution of tRNA synthetases.

N1. The Henderson–Hasselbalch equation is pH = pK + ln([salt]/[acid]). Therefore if pH = 7.0 and pK = 6.5, ln([salt]/[acid]) = 0.5 or [salt]/[acid] = exp(0.5) = 1.65. Thus, the ratio of [unprotonated]/[protonated] = 1.65. In other words, [unprotonated] = 1.65 × [protonated]. Therefore the total amount of histidine is the unprotonated plus the protonated, namely [unprotonated] + [protonated] = 2.65 [protonated]. This means that the histidine is protonated a fraction of 1 in 2.65 or roughly 38%. (In other words, at 0.5 pH units above the pK there is still a very significant amount of protonated species present, roughly one-third.)

N3. Questions N3, N4, and N5 are linked. The data and idea for these questions are based on http://www.tiem.utk.edu/bioed/webmodules/aminoacid.htm. The spreadsheet and the resulting graphs for the questions are in the Excel spreadsheet ch1N3.xls. There is a correlation coefficient (R^2) of 0.14, which is not great but is clearly real. This suggests that codon usage and the occurrence of amino acids in proteins are linked. If one makes the reasonable assumption that codon usage crystallized very early, well before protein sequences stabilized, it suggests that the occurrence of amino acids in proteins is determined mainly by the frequency of the codons that code for them (and not, for example, by their usefulness to the cell). In other words, in general one could say that the cell uses amino acids at fairly close to their natural frequency as determined by the codons.

N5. There is a correlation but it is not very strong, suggesting that metabolic cost is not a major factor in determining which amino acid is used. In fact, the correlation is much weaker than that with codon frequency, implying that cost is relatively unimportant compared with the 'evolutionary accident' of how many codons code for the amino acid. However, it is possibly significant that for example the amino acid that gets used much more frequently than one might expect based on its energetic cost is leucine, a very useful hydrophobic acid because of its versatile shape, whereas the two that are used least often compared with what one might expect (by a small margin) are cysteine and histidine, both somewhat 'ambiguous' amino acids that can cause problems if located in the wrong place.