

chapter

4

The Three-Dimensional Structure of Proteins

- 1. Properties of the Peptide Bond** In x-ray studies of crystalline peptides, Linus Pauling and Robert Corey found that the C—N bond in the peptide link is intermediate in length (1.32 Å) between a typical C—N single bond (1.49 Å) and a C=N double bond (1.27 Å). They also found that the peptide bond is planar (all four atoms attached to the C—N group are located in the same plane) and that the two α -carbon atoms attached to the C—N are always trans to each other (on opposite sides of the peptide bond).
- (a) What does the length of the C—N bond in the peptide linkage indicate about its strength and its bond order (i.e., whether it is single, double, or triple)?
- (b) What do the observations of Pauling and Corey tell us about the ease of rotation about the C—N peptide bond?

Answer

- (a) The higher the bond order (double or triple vs. single), the shorter and stronger are the bonds. Thus, bond length is an indication of bond order. For example, the C=N bond is shorter (1.27 Å) and has a higher order ($n = 2.0$) than a typical C—N bond (length = 1.49 Å, $n = 1.0$). The length of the C—N bond of the peptide link (1.32 Å) indicates that it is intermediate in strength and bond order between a single and double bond.
- (b) Rotation about a double bond is generally impossible at physiological temperatures, and the steric relationship of the groups attached to the two atoms involved in the double bond is spatially “fixed.” Since the peptide bond has considerable double-bond character, there is essentially no rotation, and the —C=O and —N—H groups are fixed in the trans configuration.
- 2. Structural and Functional Relationships in Fibrous Proteins** William Astbury discovered that the x-ray diffraction pattern of wool shows a repeating structural unit spaced about 5.2 Å along the length of the wool fiber. When he steamed and stretched the wool, the x-ray pattern showed a new repeating structural unit at a spacing of 7.0 Å. Steaming and stretching the wool and then letting it shrink gave an x-ray pattern consistent with the original spacing of about 5.2 Å. Although these observations provided important clues to the molecular structure of wool, Astbury was unable to interpret them at the time.
- (a) Given our current understanding of the structure of wool, interpret Astbury’s observations.
- (b) When wool sweaters or socks are washed in hot water or heated in a dryer, they shrink. Silk, on the other hand, does not shrink under the same conditions. Explain.

Answer

- (a) The principal structural units in the wool fiber polypeptide, α -keratin, are successive turns of the α helix, which are spaced at 5.4 Å intervals; two α -keratin strands twisted into a coiled coil produce the 5.2 Å spacing. The intrinsic stability of the helix (and thus the fiber) results from *intra* chain hydrogen bonds (see Fig. 4–4a). Steaming and stretching the fiber yields an extended polypeptide chain with the β conformation, in which the distance between adjacent R groups is about 7.0 Å. Upon resteamng, the polypeptide chains again assume the less-extended α -helix conformation.

(b) Freshly sheared wool is primarily in its α -keratin (α -helical coiled coil) form (see Fig. 4-11). Because raw wool is crimped or curly, it is combed and stretched to straighten it before being spun into fibers for clothing. This processing converts the wool from its native α -helical conformation to a more extended β form. Moist heat triggers a conformational change back to the native α -helical structure, which shrinks both the fiber and the clothing. Under conditions of mechanical tension and moist heat, wool can be stretched back to a fully extended form. In silk, by contrast, the polypeptide chains have a very stable β -pleated sheet structure, fully extended along the axis of the fiber (see Fig. 4-6), and have small, closely packed amino acid side chains (see Fig. 4-14). These characteristics make silk resistant to stretching and shrinking.

3. Rate of Synthesis of Hair α -Keratin Hair grows at a rate of 15 to 20 cm/yr. All this growth is concentrated at the base of the hair fiber, where α -keratin filaments are synthesized inside living epidermal cells and assembled into ropelike structures (see Fig. 4-11). The fundamental structural element of α -keratin is the α helix, which has 3.6 amino acid residues per turn and a rise of 5.4 Å per turn (see Fig. 4-4a). Assuming that the biosynthesis of α -helical keratin chains is the rate-limiting factor in the growth of hair, calculate the rate at which peptide bonds of α -keratin chains must be synthesized (peptide bonds per second) to account for the observed yearly growth of hair.

Answer Because there are 3.6 amino acids (AAs) per turn and the rise is 5.4 Å/turn, the length per AA of the α helix is

$$\frac{5.4 \text{ Å/turn}}{3.6 \text{ AA/turn}} = 1.5 \text{ Å/AA} = 1.5 \times 10^{-10} \text{ m/AA}$$

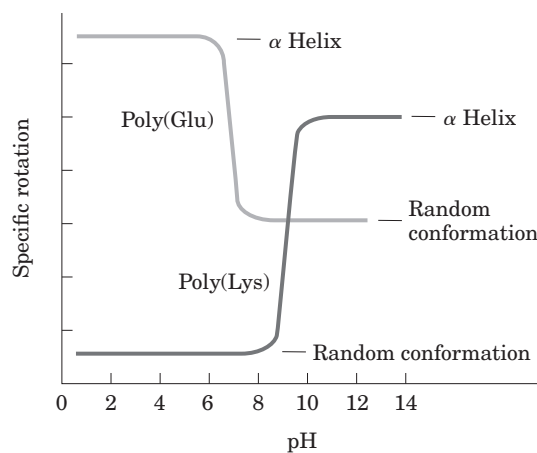
A growth rate of 20 cm/yr is equivalent to

$$\frac{20 \text{ cm/year}}{(365 \text{ days/yr})(24 \text{ h/day})(60 \text{ min/h})(60 \text{ s/min})} = 6.3 \times 10^{-7} \text{ cm/s} = 6.3 \times 10^{-9} \text{ m/s}$$

Thus, the rate at which amino acids are added is

$$\frac{6.3 \times 10^{-9} \text{ m/s}}{1.5 \times 10^{-10} \text{ m/AA}} = 42 \text{ AA/s} = 42 \text{ peptide bonds per second}$$

4. Effect of pH on the Conformation of α -Helical Secondary Structures The unfolding of the α helix of a polypeptide to a randomly coiled conformation is accompanied by a large decrease in a property called specific rotation, a measure of a solution's capacity to rotate circularly polarized light. Polyglutamate, a polypeptide made up of only L-Glu residues, has the α -helical conformation at pH 3. When the pH is raised to 7, there is a large decrease in the specific rotation of the solution. Similarly, polylysine (L-Lys residues) is an α helix at pH 10, but when the pH is lowered to 7 the specific rotation also decreases, as shown by the following graph.



S-46 Chapter 4 The Three-Dimensional Structure of Proteins

What is the explanation for the effect of the pH changes on the conformations of poly(Glu) and poly(Lys)? Why does the transition occur over such a narrow range of pH?

Answer At pH values above 6, deprotonation of the carboxylate side chains of poly(Glu) leads to repulsion between adjacent negatively charged groups, which destabilizes the α helix and results in unfolding. Similarly, at pH 7 protonation of the amino-group side chains of poly(Lys) causes repulsion between positively charged groups, which leads to unfolding.

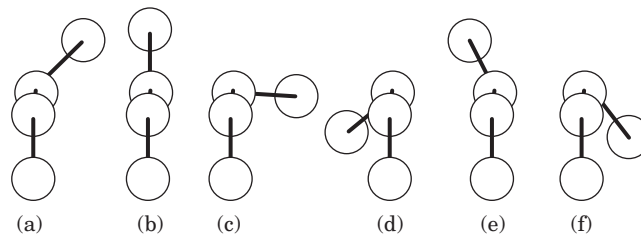
5. Disulfide Bonds Determine the Properties of Many Proteins Some natural proteins are rich in disulfide bonds, and their mechanical properties (tensile strength, viscosity, hardness, etc.) are correlated with the degree of disulfide bonding.

- (a) Glutenin, a wheat protein rich in disulfide bonds, is responsible for the cohesive and elastic character of dough made from wheat flour. Similarly, the hard, tough nature of tortoise shell is due to the extensive disulfide bonding in its α -keratin. What is the molecular basis for the correlation between disulfide-bond content and mechanical properties of the protein?
- (b) Most globular proteins are denatured and lose their activity when briefly heated to 65 °C. However, globular proteins that contain multiple disulfide bonds often must be heated longer at higher temperatures to denature them. One such protein is bovine pancreatic trypsin inhibitor (BPTI), which has 58 amino acid residues in a single chain and contains three disulfide bonds. On cooling a solution of denatured BPTI, the activity of the protein is restored. What is the molecular basis for this property?

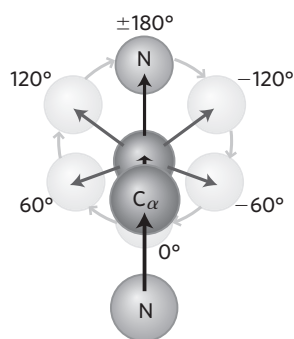
Answer

- (a) Disulfide bonds are covalent bonds, which are much stronger than the noncovalent interactions (hydrogen bonds, hydrophobic interactions, van der Waals interactions) that stabilize the three-dimensional structure of most proteins. Disulfide bonds serve to cross-link protein chains, increasing stiffness, hardness, and mechanical strength.
- (b) As the temperature is raised, the increased thermal motion of the polypeptide chains and the vibrational motions of hydrogen bonds ultimately lead to thermal denaturation (unfolding) of a protein. Cystine residues (disulfide bonds) can, depending on their location in the protein structure, prevent or restrict the movement of folded protein domains, block access of solvent water to the interior of the protein, and prevent the complete unfolding of the protein. Refolding to the native structure from a random conformation is seldom spontaneous, owing to the very large number of conformations possible. Disulfide bonds limit the number of conformations by allowing only a few minimally unfolded structures, and hence the protein returns to its native conformation more easily upon cooling.

6. Dihedral Angles A series of torsion angles, ϕ and ψ , that might be taken up by the peptide backbone is shown below. Which of these closely correspond to ϕ and ψ for an idealized collagen triple helix? Refer to Figure 4-9 as a guide.



Answer Figure 4-9a shows the values of ϕ and ψ for various secondary structures, including the collagen triple helix (in the upper-left quadrant). By inspection, the ϕ angle is approximately -50° and the ψ angle is approximately $+160^\circ$. By convention, ϕ and ψ are 180° (or -180°) in the conformation in which the peptide bond is fully extended. By referring to the angles shown in Figure 4-2d (shown below), we can see that the conformation shown in (f) is closest to the ϕ value of approximately -50° and that the conformation in (e) is closest to the ψ value of approximately $+160^\circ$.



7. Amino Acid Sequence and Protein Structure Our growing understanding of how proteins fold allows researchers to make predictions about protein structure based on primary amino acid sequence data. Consider the following amino acid sequence.

Ile-Ala-His-Thr-Tyr-Gly-Pro-Phe-Glu-Ala-Ala-Met-Cys-Lys-Trp-Glu-Ala-Gln-Pro-Asp-Gly-Met-Glu-Cys-Ala-Phe-His-Arg
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

- Where might bends or β turns occur?
- Where might intrachain disulfide cross-linkages be formed?
- Assuming that this sequence is part of a larger globular protein, indicate the probable location (the external surface or interior of the protein) of the following amino acid residues: Asp, Ile, Thr, Ala, Gln, Lys. Explain your reasoning. (Hint: See the hydropathy index in Table 3-1.)

Answer

- Bends or turns are most likely to occur at residues 7 and 19 because Pro residues are often (but not always) found at bends in globular folded proteins. A bend may also occur at the Thr residue (residue 4) and, assuming that this is a portion of a larger polypeptide, at the Ile residue (residue 1).
- Intrachain disulfide cross-linkages can form only between residues 13 and 24 (Cys residues).
- Amino acids with ionic (charged) or strongly polar neutral groups (e.g., Asp, Gln, and Lys in this protein) are located on the external surface, where they interact optimally with solvent water. Residues with nonpolar side chains (such as Ala and Ile) are situated in the interior, where they escape the polar environment. Thr is of intermediate polarity and could be found either in the interior or on the exterior surface (see Table 3-1).

8. Bacteriorhodopsin in Purple Membrane Proteins Under the proper environmental conditions, the salt-loving archaeon *Halobacterium halobium* synthesizes a membrane protein (M_r 26,000) known as bacteriorhodopsin, which is purple because it contains retinal (see Fig. 10-21). Molecules of this protein aggregate into “purple patches” in the cell membrane. Bacteriorhodopsin acts as a light-activated proton pump that provides energy for cell functions. X-ray analysis of this protein reveals that it consists of seven parallel α -helical segments, each of which traverses the bacterial cell membrane (thickness 45 \AA). Calculate the minimum number of amino acids necessary for one segment of α helix to traverse the membrane completely. Estimate the fraction of the bacteriorhodopsin protein that is involved in membrane-spanning helices. (Use an average amino acid residue weight of 110.)

S-48 Chapter 4 The Three-Dimensional Structure of Proteins

Answer Using the parameters from Problem 3 (3.6 AA/turn, 5.4 Å/turn), we can calculate that there are 0.67 AA/Å along the axis of a helix. Thus, a helix of length 45 Å (sufficient to span the membrane) requires a minimum of $(45 \text{ Å})(0.67 \text{ AA/Å}) = 30$ amino acid residues.

The membrane protein has an M_r of 26,000 and average AA M_r of 110. Thus the protein contains $26,000/110 = 240$ AA. Of these, $(30 \text{ AA/helix})(7 \text{ helices}) = 210$ AA are involved in membrane-spanning helices, which is $210/240 = 0.87$, or 87%, of the protein.

9. Protein Structure Terminology Is myoglobin a motif, a domain, or a complete three-dimensional structure?

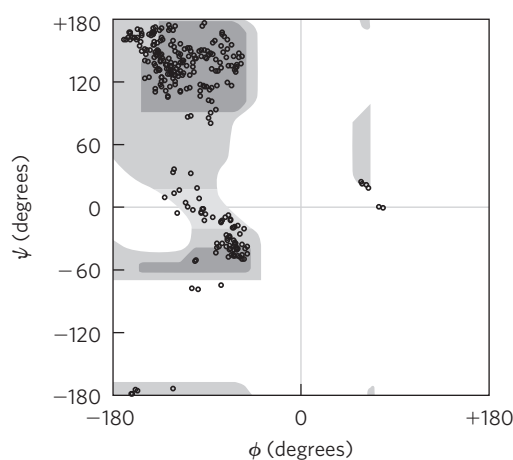
Answer Myoglobin is all three. The folded structure, the “globin fold,” is a motif found in all globins. The polypeptide folds into a single domain, which for this protein represents the entire three-dimensional structure.

10. Interpreting Ramachandran Plots Examine the two proteins labeled **(a)** and **(b)** below. Which of the two Ramachandran plots, labeled **(c)** and **(d)**, is more likely to be derived from which protein? Why?

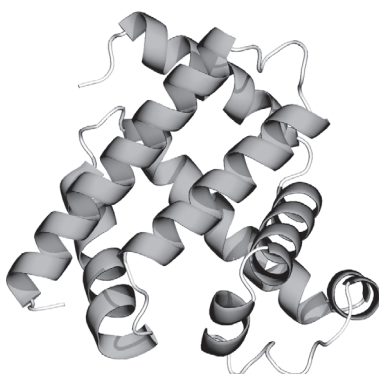
(a)



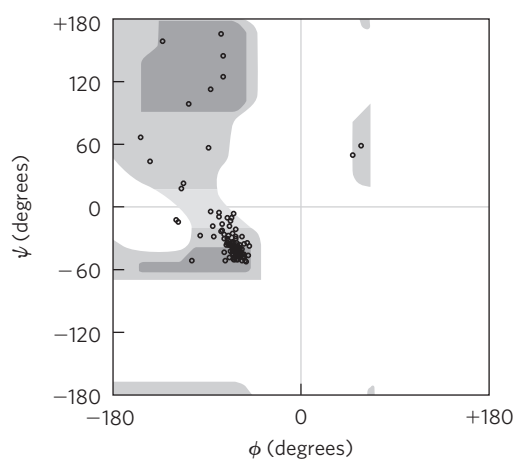
(c)



(b)

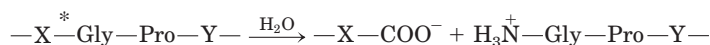


(d)



Answer The protein in **(a)** has primarily β structure (represented by flat arrows) with a small amount of α -helical structure (represented by coils). The protein in **(b)** has only α -helical structure. Figure 4-9 shows that β structures fall in the upper-left quadrant of the Ramachandran plot. Plot **(c)** has many points in the upper-left quadrant, whereas part **(d)** has few; so the plot that corresponds to **(a)** would be **(c)**. Figure 4-9 also shows that right-handed α -helices have ϕ and ψ values that put it into the lower-left quadrant of the Ramachandran plot, so protein **(b)** must correspond to graph **(d)**.

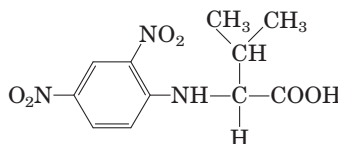
- 11. Pathogenic Action of Bacteria That Cause Gas Gangrene** The highly pathogenic anaerobic bacterium *Clostridium perfringens* is responsible for gas gangrene, a condition in which animal tissue structure is destroyed. This bacterium secretes an enzyme that efficiently catalyzes the hydrolysis of the peptide bond indicated by an asterisk:



where X and Y are any of the 20 common amino acids. How does the secretion of this enzyme contribute to the invasiveness of this bacterium in human tissues? Why does this enzyme not affect the bacterium itself?

Answer Collagen is distinctive in its amino acid composition, having a very high proportion of Gly (35%) and Pro residues. The enzyme secreted by the bacterium is a collagenase, which breaks down collagen at the X-Gly bonds and damages the connective-tissue barrier (skin, hide, etc.) of the host; this allows the bacterium to invade the host tissues. Bacteria do not contain collagen and thus are unaffected by collagenase.

- 12. Number of Polypeptide Chains in a Multisubunit Protein** A sample (660 mg) of an oligomeric protein of M_r 132,000 was treated with an excess of 1-fluoro-2,4-dinitrobenzene (Sanger's reagent) under slightly alkaline conditions until the chemical reaction was complete. The peptide bonds of the protein were then completely hydrolyzed by heating it with concentrated HCl. The hydrolysate was found to contain 5.5 mg of the following compound:



2,4-Dinitrophenyl derivatives of the α -amino groups of other amino acids could not be found.

- Explain how this information can be used to determine the number of polypeptide chains in an oligomeric protein.
- Calculate the number of polypeptide chains in this protein.
- What other protein analysis technique could you employ to determine whether the polypeptide chains in this protein are similar or different?

Answer

- Because only a single 2,4-dinitrophenyl (DNP) amino acid derivative is found, there is only one kind of amino acid at the amino terminus (i.e., all the polypeptide chains have the same amino-terminal residue). Comparing the number of moles of this derivative to the number of moles of protein gives the number of polypeptide chains.
- The amount of protein = $(0.66 \text{ g}) / (132,000 \text{ g/mol}) = 5 \times 10^{-6} \text{ mol}$.
Because M_r for DNP-Val ($\text{C}_{11}\text{H}_{13}\text{O}_6\text{N}_3$) = 283, the amount of DNP-Val = $(0.0055 \text{ g}) / (283 \text{ g/mol}) = 1.9 \times 10^{-5} \text{ mol}$.

The ratio of moles of DNP-Val to moles of protein gives the number of amino-terminal residues and thus the number of chains per oligomer:

$$\frac{1.9 \times 10^{-5} \text{ mol DNP-Val}}{5 \times 10^{-6} \text{ mol protein}} = 4 \text{ polypeptide chains}$$

S-50 Chapter 4 The Three-Dimensional Structure of Proteins

An alternative approach to the problem is through the proportionality (n = number of polypeptide chains):

$$\frac{n(283 \text{ g/mol})}{132,000 \text{ g/mol}} = \frac{5.5 \text{ mg}}{660 \text{ mg}}$$

$$n = \frac{(5.5 \text{ mg})(132,000 \text{ g/mol})}{(660 \text{ mg})(283 \text{ g/mol})} = 3.9 \approx 4$$

- (c) Polyacrylamide gel electrophoresis in the presence of a detergent (such as sodium dodecylsulfate [SDS]) and an agent that prevents the formation of disulfide bonds (such as β -mercaptoethanol) would provide information on subunit structure of a protein. In the example here, an oligomeric protein of M_r 132,000 that had four *identical* subunits would produce a single band on the electrophoretic gel, with apparent M_r \sim 33,000 ($132,000/33,000 = 4$). If the protein were made up of different polypeptide subunits, they would likely appear as multiple discrete bands on the gel.

13. Predicting Secondary Structure Which of the following peptides is more likely to take up an α -helical structure, and why?

- (a) LKAENDEAARAMSEA
(b) CRAGGFWDQPGTSN

Answer By cursory inspection, peptide (a) has five Ala residues (most likely to take up an α -helical conformation), and peptide (b) has five Pro and Gly residues (least often found in an α helix). This suggests that (a) is more likely than (b) to form an α helix. Referring to Table 4-1, (a) has 15 residues with a total $\Delta\Delta G^\circ$ of 13 kJ/mol, and (b) has 15 residues with a total $\Delta\Delta G^\circ$ of 41 kJ/mol. Given that a lower $\Delta\Delta G^\circ$ indicates a greater tendency to take up an α -helical structure, this confirms that peptide (a) is much more likely to form an α helix.

14. Amyloid Fibers in Disease Several small aromatic molecules, such as phenol red (used as a non-toxic drug model), have been shown to inhibit the formation of amyloid in laboratory model systems. A goal of the research on these small aromatic compounds is to find a drug that would efficiently inhibit the formation of amyloid in the brain in people with incipient Alzheimer disease.

- (a) Suggest why molecules with aromatic substituents would disrupt the formation of amyloid.
(b) Some researchers have suggested that a drug used to treat Alzheimer disease may also be effective in treating type 2 (non-insulin-dependent) diabetes mellitus. Why might a single drug be effective in treating these two different conditions?

Answer

- (a) Aromatic residues seem to play an important role in stabilizing amyloid fibrils. Thus, molecules with aromatic substituents may inhibit amyloid formation by interfering with the stacking or association of the aromatic side chains.
(b) Amyloid is formed in the pancreas in association with type 2 diabetes, as it is in the brain in Alzheimer's disease. Although the amyloid fibrils in the two diseases involve different proteins, the fundamental structure of the amyloid is similar and similarly stabilized in both, and thus they are potential targets for similar drugs designed to disrupt this structure.

Using the Web

15. Protein Modeling on the Internet A group of patients with Crohn disease (an inflammatory bowel disease) underwent biopsies of their intestinal mucosa in an attempt to identify the causative agent. Researchers identified a protein that was present at higher levels in patients with Crohn disease than

in patients with an unrelated inflammatory bowel disease or in unaffected controls. The protein was isolated, and the following *partial* amino acid sequence was obtained (reads left to right):

```
EAELCPDRCI HSFQNLGIQC VKKRDLEQAI SQRIQTNNNP FQVPIEEQRG
DYDLNAVRLC FQVTVRDPSG RPLRLPPVLP HPIFDNRAPN TAEKICRVN
RNSGSCGGD EIFLLCDKVQ KEDIEVYFTG PGWEARGSFS QADVHRQVAI
VFRTPPYADP SLQAPVRVSM QLRRPSDREL SEPMEFQYLP DTDDRHRIEE
KRRRTYETFK SIMKKSPFSG PT DPRPPPRR IAVPSRSSAS VPKPAPQYP
```

- (a) You can identify this protein using a protein database on the Internet. Some good places to start include Protein Information Resource (PIR; <http://pir.georgetown.edu>), Structural Classification of Proteins (SCOP; <http://scop.mrc-lmb.cam.ac.uk/scop>), and Prosite (<http://prosite.expasy.org>). At your selected database site, follow links to the sequence comparison engine. Enter about 30 residues from the protein sequence in the appropriate search field and submit it for analysis. What does this analysis tell you about the identity of the protein?
- (b) Try using different portions of the amino acid sequence. Do you always get the same result?
- (c) A variety of websites provide information about the three-dimensional structure of proteins. Find information about the protein's secondary, tertiary, and quaternary structures using database sites, such as the Protein Data Bank (PDB; www.pdb.org) or SCOP.
- (d) In the course of your Web searches, what did you learn about the cellular function of the protein?

Answer

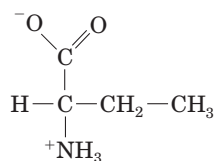
- (a) At the PIR—International Protein Sequence Database (<http://pir.georgetown.edu>), click on “Search/Analysis” and choose “BLAST search.” Paste the first 30 amino acid residues of the sequence into the search box and submit the sequence for comparison. The table that returns includes many proteins that have 100% sequence identity with these 30 residues. Among the human proteins are RelA and the transcription factor NF κ B. Proteins from other species match as well. Click on the “Help” button for explanations of the various options and table items.
- (b) As more proteins are sequenced, the number of hits returned from a 30-residue sequence increases. Sequence matching based on the first 30 residues brings up several proteins that contain this sequence (identity = 100%). Using sequence segments from different parts of the protein will return some different results, but the proteins with high sequence identities will likely be similar. Even when the *entire* sequence is entered in the search field, similar proteins from cattle, mouse, and rat match with very high scores, and several hundred hits are returned. When the entire sequence is used, the human protein with the best match is the p65 subunit of nuclear transcription factor kappa B (NF κ B). A synonym for this protein is RelA transforming protein.
- (c) At the PDB (rcsb.org) search on “NF-kappa-B p65.” You will get more than a dozen hits. Adding “human” to the search limits the results further. Go back to the more general search on “NF-kappa-B p65” and scan through the returned items. NF κ B has two subunits. There are multiple variants of the subunits, with the best-characterized being 50, 52, and 65 kDa (p50, p52, and p65, respectively). These pair with each other to form a variety of homodimers and heterodimers.
- (d) The various proteins that predominate in this search are eukaryotic transcription factors, which stimulate transcription of genes involved in development and some immune responses. The proteins have two distinct domains, including an amino-terminal Rel homology domain 300 amino acid residues long and a carboxyl-terminal domain involved in gene activation. A search of the various links in the databases will reveal much additional information about the proteins' structure and function.

Data Analysis Problem

16. Mirror-Image Proteins As noted in Chapter 3, “The amino acid residues in protein molecules are exclusively L stereoisomers.” It is not clear whether this selectivity is necessary for proper protein function or is an accident of evolution. To explore this question, Milton and colleagues (1992) published a study of an enzyme made entirely of D stereoisomers. The enzyme they chose was HIV protease, a proteolytic enzyme made by HIV that converts inactive viral preproteins to their active forms.

Previously, Wlodawer and coworkers (1989) had reported the complete chemical synthesis of HIV protease from L-amino acids (the L-enzyme), using the process shown in Figure 3–32. Normal HIV protease contains two Cys residues at positions 67 and 95. Because chemical synthesis of proteins containing Cys is technically difficult, Wlodawer and colleagues substituted the synthetic amino acid L- α -amino-*n*-butyric acid (Aba) for the two Cys residues in the protein. In the authors’ words, this was done to “reduce synthetic difficulties associated with Cys deprotection and ease product handling.”

- (a) The structure of Aba is shown below. Why was this a suitable substitution for a Cys residue? Under what circumstances would it not be suitable?



L- α -Amino-*n*-butyric acid

Wlodawer and coworkers denatured the newly synthesized protein by dissolving it in 6 M guanidine HCl, and then allowed it to fold slowly by dialyzing away the guanidine against a neutral buffer (10% glycerol, 25 mM NaPO₄, pH 7).

- (b) There are many reasons to predict that a protein synthesized, denatured, and folded in this manner would not be active. Give three such reasons.
- (c) Interestingly, the resulting L-protease was active. What does this finding tell you about the role of disulfide bonds in the native HIV protease molecule?

In their new study, Milton and coworkers synthesized HIV protease from D-amino acids, using the same protocol as the earlier study (Wlodawer et al.). Formally, there are three possibilities for the folding of the D-protease: it would give (1) the same shape as the L-protease, (2) the mirror image of the L-protease, or (3) something else, possibly inactive.

- (d) For each possibility, decide whether or not it is a likely outcome and defend your position.

In fact, the D-protease was active: it cleaved a particular synthetic substrate and was inhibited by specific inhibitors. To examine the structure of the D- and L-enzymes, Milton and coworkers tested both forms for activity with D and L forms of a chiral peptide substrate and for inhibition by D and L forms of a chiral peptide-analog inhibitor. Both forms were also tested for inhibition by the achiral inhibitor Evans blue. The findings are given in the table.

HIV protease	Substrate hydrolysis		Inhibition		
			Peptide inhibitor		Evans blue (achiral)
	D-substrate	L-protease	D-inhibitor	L-inhibitor	
L-protease	–	+	–	+	+
D-protease	+	–	+	–	+

- (e) Which of the three models proposed above is supported by these data? Explain your reasoning.
- (f) Why does Evans blue inhibit both forms of the protease?
- (g) Would you expect chymotrypsin to digest the D-protease? Explain your reasoning.
- (h) Would you expect total synthesis from D-amino acids followed by renaturation to yield active enzyme for any enzyme? Explain your reasoning.

Answer

- (a) Aba is a suitable replacement because Aba and Cys have approximately the same sized side chain and are similarly hydrophobic. However, Aba cannot form disulfide bonds so it will not be a suitable replacement if these are required.
- (b) There are many important differences between the synthesized protein and HIV protease produced by a human cell, any of which could result in an inactive synthetic enzyme: (1) Although Aba and Cys have similar size and hydrophobicity, Aba may not be similar enough for the protein to fold properly. (2) HIV protease may require disulfide bonds for proper functioning. (3) Many proteins synthesized by ribosomes fold as they are produced; the protein in this study folded only after the chain was complete. (4) Proteins synthesized by ribosomes may interact with the ribosomes as they fold; this is not possible for the protein in the study. (5) Cytosol is a more complex solution than the buffer used in the study; some proteins may require specific, unknown proteins for proper folding. (6) Proteins synthesized in cells often require chaperones for proper folding; these are not present in the study buffer. (7) In cells, HIV protease is synthesized as part of a larger chain that is then proteolytically processed; the protein in the study was synthesized as a single molecule.
- (c) Because the enzyme is functional with Aba substituted for Cys, disulfide bonds do not play an important role in the structure of HIV protease.
- (d) *Model 1*: it would fold like the L-protease. *Argument for*: the covalent structure is the same (except for chirality), so it should fold like the L-protease. *Argument against*: chirality is not a trivial detail; three-dimensional shape is a key feature of biological molecules. The synthetic enzyme will not fold like the L-protease. *Model 2*: it would fold to the mirror image of the L-protease. *For*: because the individual components are mirror images of those in the biological protein, it will fold in the mirror-image shape. *Against*: the interactions involved in protein folding are very complex, so the synthetic protein will most likely fold in another form. *Model 3*: it would fold to something else. *For*: the interactions involved in protein folding are very complex, so the synthetic protein will most likely fold in another form. *Against*: because the individual components are mirror images of those in the biological protein, it will fold in the mirror-image shape.
- (e) Model 1. The enzyme is active, but with the enantiomeric form of the biological substrate, and it is inhibited by the enantiomeric form of the biological inhibitor. This is consistent with the D-protease being the mirror image of the L-protease.
- (f) Evans blue is achiral; it binds to both forms of the enzyme.
- (g) No. Because proteases contain only L-amino acids and recognize only L-peptides, chymotrypsin would not digest the D-protease.
- (h) Not necessarily. Depending on the individual enzyme, any of the problems listed in (b) could result in an inactive enzyme.

References

- Milton, R. C., Milton, S. C., & Kent, S. B. (1992) Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show demonstration of reciprocal chiral substrate specificity. *Science* **256**, 1445–1448.
- Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., & Kent, S. B. (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **245**, 616–621.