Chapter 3. Weaver, 2/e. Mol Biol X107A.

Ch 3. An introduction to gene function.

Reading note. We will skip the sections titled Transcription, Translation, Replication. These sections preview topics that will be discussed at length in later chapters. If you are unfamiliar with the basics, you may want to read these now. In any case, browse them, and refer to them as needed.

Clark & Russell: Ch 7, on proteins.

A. Introduction

Highlights:

- Using DNA - an introduction. The Central Dogma of molecular biology summarizes the overall process. The genetic code is the translation table.
- mRNA as the message, a transient copy of the gene.
- Weak bonds: types, energy range, roles.

- Proteins:
  - Types of amino acids. (But there are also post-translational modifications.)
  - Peptide bonds (amide linkage)
  - Conformation (secondary and tertiary structure).
  - Folding; role of chaperone proteins.
- The concept of the "gene".
- Mutations and DNA.

B. Central Dogma; Genetic code

The Central Dogma of molecular biology summarizes how DNA functions (Fig 3.1).

Simple statement: DNA → RNA → protein. However, a more rigorous statement is that information cannot flow out of proteins. *DNA and RNA are informationally equivalent*, but the form of information in proteins is quite different. Information can be translated from the nucleic acid form to the protein form. However, we do not know (and cannot conceive ??) the reverse process. Crick (1970).

Coding information is stored in the sequence of bases in the DNA chain. Consecutive triplets (called codons) code for consecutive amino acids in the protein product. The genetic code is the translation table from the nucleotide language to the amino acid language.

⇒ The genetic code is shown in Fig 18.6, p 573.

It is inherent in this story that one strand of DNA contains information and the other strand contains the complement, or "anti-information". DNA itself does not know which is which, and it can vary along the chromosome. That is, one DNA strand may carry coding information for one gene, and the other strand may carry coding information for another gene. RNA polymerase, the enzyme for transcription (making RNA), "knows" which strand to choose at any site; Ch 6.

Thieffry & Sarkar (1998) review the Central Dogma.

C. Terms (strands)

The strand names, in the context of mRNA corresponding to one strand of the DS DNA…

The message sequence (or the corresponding DNA sequence) is called sense. The complementary strand is called antisense.

The sense strand may also be called the coding strand; however, unfortunately the latter term is used by some to mean the strand that codes for the mRNA. Hence the terms coding and noncoding (or anti-coding) strands have become ambiguous (p 39).

The antisense strand is called the template strand, in that it serves as the template for making RNA.

> Weaver prefers the terms template and nontemplate (p 39). I tend to prefer sense and antisense, which clearly refer to coding for protein.

D. Messenger RNA

The messenger RNA (mRNA) is a transient copy of the gene. The mRNA is the form of the gene actually used for translation.

The need for such an intermediate is "obvious" in eukaryotes, where the genes and the ribosomes are in separate cellular compartments. However, the basic role of the mRNA is the same in all organisms.

The first convincing experimental demonstration of mRNA was the 1961 experiment of Brenner et al (Fig 3.12). In the context of that time, this experiment showed that "new genes" function via the old ribosomes, thus disproving that ribosomes are specific for the protein being made. The experiment also provided direct evidence for a class of transient RNA involved in gene function.

E. Weak bonds

(Discussed little by Weaver.)

Weak bonds are responsible for the shapes of macromolecules, and for the interactions between molecules. Specifically, weak bonds are responsible for the transfer of genetic information and for the action of enzymes.

Roles of weak bonds

- in determining the secondary, tertiary and quaternary structure of proteins;
- in determining the general structure of DNA and the specifics of base pairing; (Fig 2.13)
- in the binding of small molecules (substrates, effectors) to proteins;
- in the binding of proteins to nucleic acids;
- in maintaining a structure, yet allowing it to breathe;

- in mediating the effect of ligand binding on the structure of a protein.

Nature of weak bonds:

- Types of weak bonds: ionic bonds, hydrogen bonds, London (van der Waals) forces; hydrophobic interactions.

- The energy range of weak bonds (0.1 to 10 kcal/mol), as compared to strong bonds (covalent; 100 kcal/mol) and to thermal fluctuations (0.6 kcal/mol).

- Individual weak bonds are rapidly turned over.

- The key role of the solvent (water).

Recall that chemical bonds are explained in terms of interactions between electron clouds and nuclei. You should be generally familiar with ionic and covalent bonds, in terms of electrons.

Weak bonds can be understood within this framework. They involve attraction between opposite charges. The charges may be the obvious charges of ions or the partial charges from polar covalent bonds or induced dipoles.

Hydrogen bonds are a special subclass of dipole interactions involving polar bonds. We recognize them as a special class because they are so strong -- and so important, in water and in biological systems.

Dougherty (1996) discusses a new part of this story. Aromatic rings, which we normally think of as hydrophobic, may be involved in hydrogen bonding.

Hydrophobic interaction. We know that oil and water do not mix. (More generally, nonpolar substances tend to leave water.) The oil would disrupt the relatively strong water-water interactions if it were in the water, so it gets out. It will serve our purposes here if you understand the hydrophobic interaction at that level. This is sufficient to explain the "hydrophobic collapse", which is undoubtedly an early stage of protein folding. Silverstein (1998) explains the basic hydrophobic effect in more detail, emphasizing the key role of entropy changes. Cheng & Rossky (1998) discuss the greater complexity of the hydrophobic effect in proteins.

The strength of polar interactions in aqueous solution is actually rather low. Thus hydrophobic interactions are extremely important in determining the general shape of proteins. Xu et al (1997) compare the solvent effect -- and hence the relative importance of dipole interactions -- in folding of a single chain vs inter-chain interactions.

Davis et al (1999) discuss predicting the solubility of proteins, based on amino acid content. They use their model to predict some proteins of unusually high solubility, suitable for use as fusion partners for cloned proteins.

F. Proteins

Proteins are the most common molecules in cells (recall Ch 1 handout Sect D). They play several roles in the cell, including enzymatic, structural, hormonal, and regulatory (pp 44-45). In discussing protein function here, Weaver introduces the historical evidence of genes → proteins.

The monomer subunits of proteins are L-α-amino acids. You should understand the general nature of these monomers, and the types of side chains. Amino acids are polymerized into peptides or protein chains, via peptide bonds, which are ordinary amide linkages (Fig 3.3).

Post-translational modifications. Proteins can be modified after primary synthesis. Thus the number of amino acids found in proteins is greater than the 20 that are genetically encoded. Further, non-amino acids, such as sugars and lipids, can be added to the protein. See Kelly et al (2000) and Hurtley et al (2001). McBride & Silver (2001) discuss protein methylation.

G. Protein structure

The terms primary structure, secondary structure, etc, describe levels of protein structure (pp 39-43).

The primary structure of proteins is due to covalent bonds; the higher order structures are (mainly) due to the various weak bonds discussed above. Secondary and tertiary structures are due to interactions within a single chain; quaternary structure refers to interactions between chains.

A useful view is that secondary structure is due to the backbone, specifically interactions between the NH and CO parts of the amide linkage. 2° structure is largely due to hydrogen bonding. The α-helix and β-sheet are examples of common secondary structures. Note how the side chains are arranged in these 2° structures.

Tertiary (and quaternary) structures involve interactions between amino acid side chains. All of the kinds of weak bonds are involved, depending on the nature of the side chains. The solvent plays a major role in determining the strength of these interactions (Sect E, above).

> [Disulfide bonds are an example of a covalent bond involved in protein tertiary structure. However, many proteins have no disulfide bonds; this includes cytoplasmic proteins in general. In any case, covalent bonds in general are not a major part of the story of higher order structure. Tu et al (2000) and Tsai et al (2001) discuss parts of the disulfide bond story.]
>
> Zhou et al (2000) discuss peptide bond isomerization at prolines. Pappenberger et al (2001) discuss isomerization for other amino acids.

The term domain (p 42) is used loosely, but consistent with normal English usage of the word. For example, visual inspection of Fig 3.7 reasonably suggests that there are two domains, connected by a less ordered region. However, this has no particular technical significance. It is quite plausible that large proteins fold on a domain-by-domain basis.

Some FR deal with unusual proteins, or unusual aspects or uses of protein structure. As with the collection of "unusual uses of DNA" items in Ch 2 FR, some of these are "for fun". These FR include: Elowitz et al (1999); Imai et al (1999); Kröger et al (1999); Goodsell (2000); Whaley et al (2000); Guo et al (2000).

Deng et al (2000) explore similar structural issues for a non-protein molecule. Work on ribozymes is included in Sect I, below.

H. Self-assembly; chaperones

Molecules interact because it is energetically and kinetically favorable to do so, not in order to achieve a goal. The principle of self-assembly fits into this perspective.

Examples at a small scale: A protein folds to give the proper secondary and tertiary structure; subunits aggregate into the proper quaternary structure.

At a larger scale: some viruses can be re-assembled from their parts, the nucleic acid and the protein coat. Ribosomes, containing 3 kinds of RNA and some 50 kinds of protein, spontaneously assemble in the test tube and -- maybe -- inside the cell.

Not all assembly fits this simple view of self-assembly. In fact, the cases mentioned above may not be true in vivo. We now realize that sometimes an accessory protein is needed to assist in assembly.

On p 42 Weaver hints that proteins may not fold spontaneously in vivo. "Chaperone proteins" aid protein folding. Their role may be primarily to prevent incorrect folding or incorrect quaternary interactions (i.e., aggregation). For example, one type of chaperone provides a protected environment in which folding -- but not aggregation -- can occur. The protein concentration in the cell is much higher than a biochemist would normally use in the lab; aggregation is more likely to be a problem in the cell than in the test tube.

Thus the role of a chaperone is kinetic. It does not specify the stable structure, but rather improves the kinetics of getting there. (How it improves kinetics may be somewhat different from how ordinary enzymes act.)

Similar proteins play a role in keeping proteins unfolded so they can be translocated through the membrane. Further, "heat shock proteins", which protect the cell against high levels of denatured proteins, are also similar (Sect 8.5).

Baker (1999) discusses how chaperones may be involved in preparing proteins to be degraded. Tsai et al (2001) describe a chaperone who shape cycle is driven by redox changes, rather than

by ATP-ADP. Other FR on how proteins fold (or unfold) include: Cordes et al (1999), Oesterhelt et al (2000), Baker (2000), Brooks et al (2001). Dobson (1999) discusses the role of defective protein folding in human disease. Altamirano et al (2000) try to design new proteins.

I. Ribozymes

[brief preview; not in book at this point]

Enzymes are proteins. Always. Or so we thought, until it was discovered that some enzymes are RNA molecules, or "ribozymes". RNA molecules can fold up into specific 3D shapes, and some RNA molecules have catalytic activity.

A few ribozymes are found in nature; we will see examples of them in some later chapters: Ribosomes, Ch 17-19; RNA processing and splicing, Ch 14-16. Further, studying catalysis by RNA molecules (and more recently, even DNA molecules) has become an active field of study. In the lab, studying the activity of something which can actually replicate itself is a great convenience. FR on this include: Hermann & Patel (2000), Schultes & Bartel (2000), Tanabe et al (2000). Ridanpaa et al (2001) may have found a ribozyme mutation that causes human disease. Breaker (2000) is an introduction to DNA enzymes.

An implication of realizing that RNA can catalyze reactions is that RNA can carry out all the basic life functions of the three modern informational macromolecules: DNA, RNA, and protein. It thus becomes a plausible story that RNA might have been the primordial informational macromolecule, carrying out all these functions. Over time, RNA "farmed out" gene storage to DNA and catalysis to proteins. This story of the "RNA world" now dominates much discussion of early life. Johnston et al (2001) is an example.

J. The "gene"

A gene is a formal unit of heredity. (Mendel)

A gene is a part of the chromosome (DNA) which produces one protein. One gene-one enzyme (or one gene-one polypeptide). (pp 44-45)

A gene consists of a sequence of bases in the DNA, which is translated sequentially, three bases at a time via the genetic code, into a protein product. Colinearity: a gene is colinear with its product.

A gene is a sequence of bases from a start codon to a stop codon.

A gene is a sequence of bases from an mRNA start signal to an mRNA stop signal.

A gene is a sequence of bases from an mRNA start signal to an mRNA stop signal, plus associated regulatory sequences.

A gene is "the entire nucleic acid sequence that is necessary for the synthesis of a functional polypeptide or RNA molecule." (from a previous textbook)

Phenomena such as polycistronic messages and introns are all accommodated within the broader meaning of gene.

Griffiths & Neumann-Held (1999) discuss the "gene". Labrador et al (2001) present an example that doesn't quite fit the rules. Each such case tests our understanding of what is "really fundamental". This one can be accommodated. Lewis & Palevitz (2001) tackle the problems raised by the recent finding that humans make far more proteins than they have genes.

K. Mutations

A change in the DNA is called a mutation. Much of the story of how mutations occur is intertwined with the story of repair of damage to DNA (Ch 20).

Eyre-Walker & Keightley (1999) discuss your mutations.

L. Further reading

F Crick, Central dogma of molecular biology. Nature 227:561, 8/8/70. Available online at http://www.euchromatin.org/Crick01.htm
This is the article in which Crick addresses attacks on the central dogma that were stimulated by the discovery of reverse transcriptase (RNA $\rightarrow$ DNA). He clarifies what he really meant, and how reverse transcription fits right in. I had occasion to re-read this classic recently; I heartily recommend it! The Crick article that Weaver lists for Ch 3 is the original presentation of the Central Dogma.

D A Dougherty, Cation-π interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. Science 271:163, 1/12/96. The π electron cloud of an aromatic ring can interact with positive charges. This may include interaction with alkali metal cations (e.g., K^+), or hydrogen bonding-type interactions with polar hydrogens, as in water ($—H^{\delta+}$). The bonding discussed here is probably relevant to protein structure and ligand bonding.

D Xu et al, Protein binding versus protein folding: the role of hydrophilic bridges in protein associations. J Mol Biol 265:68, 1/10/97. They show that polar interactions are more important in association of proteins with each other than in folding of an individual protein. They argue that this is due to the different role of water in the two types of transitions.

T P Silverstein, The real reason why oil and water don't mix. J Chem Ed 75:116, 1/98. A discussion -- in chemical terms, with thermodynamic data -- of the "hydrophobic" interaction. The key point is that the effect is driven by entropy, with both solute and solvent being restricted when a hydrocarbon dissolves in water. (Enthalpy changes for dissolving hydrocarbons in water are actually favorable, in many cases.) Silverstein also discusses why

methane is quite soluble in water at low temperature with high pressure -- leading to the storage of considerable amounts of hydrocarbons in cold oceans at great depths.

S Tan & T J Richmond, Crystal structure of the yeast MATα2/MCM1/DNA ternary complex. Nature 391:660, 2/12/98. See hw.

Y-K Cheng & P J Rossky, Surface topography dependence of biomolecular hydrophobic hydration. Nature 392:696, 4/16/98. The reason that small hydrocarbon molecules do not dissolve in water is understood rather well (Silverstein, 1998, above). However, with larger hydrophobic surfaces, the hydrophobic effect is considerably more complex -- as illustrated here with a small membrane protein.

D Thieffry & S Sarkar, Forty years under the central dogma. TIBS 23:312, 8/98. Discussion of history (with extensive list of historical references). They try to carefully delineate what the Central Dogma does and does not mean, by discussing its limitations and challenges.

A Eyre-Walker & P D Keightley, High genomic deleterious mutations rates in hominids. Nature 397:344, 1/28/99. (+ News, Crow, p 293.) What is our "genetic load"? That is, how many deleterious mutations do we accumulate? This study suggests "too many".

M B Elowitz et al, Protein mobility in the cytoplasm of Escherichia coli. J Bact 181:197, 1/99. Measurement of protein mobility suggests that it cannot be simply explained by assuming a uniform viscosity of the cytoplasm.

E Imai et al, Elongation of oligopeptides in a simulated submarine hydrothermal system. Science 283:831, 2/5/99. A simple amino acid can form oligomers under conditions which some think may be similar to those at the origin of life.

M H J Cordes et al, Evolution of a protein fold in vitro. Science 284:325, 4/9/99. They switch two amino acids in a protein, and find that a β-region is changed to a region of two α-helices. Interestingly, the new protein is well structured and stable, and probably fairly similar to the original (not entirely clear). This paper has evolutionary implications, for how modest changes in primary structure can make substantial changes in the type of secondary structure. Cf Schultes & Bartel (2000), for ribozymes.

P E Griffiths & E M Neumann-Held, The many faces of the gene. BioScience 49:656, 8/99. They explore the evolution of the term "gene". I'm not sure I particularly agree with where they go with this, but they raise many good issues.

T A Baker, Protein unfolding: Trapped in the act. Nature 401: 29, 9/2/99. News. Discusses how chaperone proteins are involved in unfolding proteins so they can be degraded. The interrelationships of folding/unfolding and synthesis/degradation are emerging ideas.

C M Dobson, Protein misfolding, evolution and disease. TIBS 24:329, 9/99. Discusses several diseases in which protein misfolding seems to be relevant. A central focus is "amyloid", a generic term for the inappropriate protein aggregates found in Alzheimer's disease and prion diseases. Also see Li et al (2000).

N Kröger et al, Polycationic peptides from diatom biosilica that direct silica nanosphere formation. Science 286:1129, 11/5/99. (+ News, Amato, p 1059.) What is responsible for the intricate silicaceous skeletons of diatoms? Proteins, called silaffins, with modified lysines.

G D Davis et al, New fusion protein systems designed to give soluble expression in Escherichia coli. Biotechnol Bioeng 65(4):382, 11/20/99. A common problem when expressing foreign proteins at high levels in bacteria is that they are often insoluble. The resulting product is often in the form of "inclusion bodies". One approach is to make a fusion protein, with the desired product being fused to a very soluble protein. Here, Davis et al develop a way to predict the solubility of proteins, based on their amino acid content. Then, they show that a protein that is predicted to be extremely soluble is, in fact, a very good choice as a fusion partner.

Y Deng et al, Direct observation of the "pac-man" effect from dibenzofuran-bridged cofacial bisporphyrins. J Am Chem Soc 122:410, 1/19/00. (+ Note in Science, with key Fig, 287:769, 2/4/00.) They design a synthetic chemical that changes shape upon binding a "ligand".

T Hermann & D J Patel, Adaptive recognition by nucleic acid aptamers. Science 287:820, 2/4/00. Review. We now understand that nucleic acids (especially ones that are nominally SS) can have specific shapes, and in fact we recognize the role of ribozymes in Nature. Because nucleic acids are self-replicating, it is easier to "evolve" them in the lab for specific properties. For example, one might replicate a population of random RNA sequences, and then see what binds to some specific ligand. Take what binds, mutate it, replicate again, and repeat. The overall process is a selection for an RNA that binds to a specific ligand. Such RNA molecules are known as aptamers. Much is now known about these RNAs -- and is reviewed here.

M M Altamirano et al, Directed evolution of new catalytic activity using the α/β-barrel scaffold. Nature 403:617, 2/10/00. (+ News, Pestko, p 606.) One way to test our understanding of protein structure or enzyme function is to see if we can design novel proteins or enzymes. Here, they do a combination of "rational" design followed by some natural selection to develop a new enzyme activity. The new article is particularly good as an overview of the broad topic of designing proteins. *[This paper has been formally retracted. See Nature 417:468, 5/23/02.]*

R Kelly et al, Geranylgeranyltransferase I of Candida albicans: Null mutants or enzyme inhibitors produce unexpected phenotypes. J Bact 182:704, 2/00. This work deals with a posttranslational modification of proteins in which lipids are attached, thus promoting membrane anchoring of the protein. They had suspected that a somewhat unusual lipid-adding enzyme found in the pathogenic yeast Candida might be a good target for drug therapy. However they show here that genetic disruption of the gene for this enzyme is not lethal, suggesting that the fungus somehow compensates for the absence.

F Oesterhelt et al, Unfolding pathways of individual bacteriorhodopsins. Science 288:143-146, 4/7/00. (+ News, Forbes & Lorimer, p 63.) They measure the force required to unfold individual protein molecules, using an atomic force microscope.

D Baker, A surprising simplicity to protein folding. Nature 405:39, 5/4/00. Review. A surprisingly optimistic, but simplified, view of protein folding. A key point is that proteins

with many local interactions fold faster than proteins with many interactions between distant parts.

D S Goodsell, Biomolecules and nanotechnology. Amer Sci 88:230, 5/00. An overview of proteins, with an eye on what they tell us about how machines are designed. Lots of good pictures, as usual for Goodsell.

S R Whaley et al, Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly. Nature 405:665, 6/8/00. (+ News, Mirkin & Tatun, p 626.) They screen combinatorial peptide libraries for peptides that bind to inorganic materials -- semiconductors -- with specificity. This is part of the story of attempts to use biological materials for technological applications.

Y-M Li et al, Photoactivated γ-secretase inhibitors directed to the active site covalently label presenilin 1. Nature 405:689, 6/8/00. (+ News, De Strooper, p 627.) Part of exciting developments in Alzheimer's disease. Attention is focused on the production of an amyloid-forming peptide. This peptide is thought -- but not proven -- to be involved in the pathology. Now, proteases that cleave the peptide from its precursor have been identified; one is a novel type of protease, called presenilin or γ-secretase. The development of drugs against such proteases has implications for research, and maybe in the long run for therapy. Recall Dobson (1999).

Z Guo et al, Designing small-molecule switches for protein-protein interactions. Science 288:2042, 6/16/00. They start with two proteins that bind each other. They then introduce mutations that reduce the binding -- unless a particular small molecule is also bound. Thus the small molecule serves as a switch.

E A Schultes & D P Bartel, One sequence, two ribozymes: implications for the emergence of new ribozyme folds. Science 289:448, 7/21/00. (+ News, Joyce, p 401.) Work on developing ribozymes by in vitro selection. They show that one particular sequence has two stable structures, both with catalytic activities (different ones). Small mutations can stabilize one or the other activity. Cf Cordes et al (1999), for proteins.

T Tanabe et al, Oncogene inactivation in a mouse model: Tissue invasion by leukaemic cells is stalled by loading them with a designer ribozyme. Nature 406:473, 8/3/00. Example of attempt to develop a therapeutic use of a ribozyme, to cleave an RNA. What is novel here is that the ribozyme is dimeric, with two binding sites. This allows it to distinguish a more complex target. I should caution that the pharmacology of RNA is not very favorable, and the immediate practical implications of such work are questionable.

X Z Zhou et al, Pin1-dependent prolyl isomerization regulates dephosphorylation of Cdc25C and Tau proteins. Molecular Cell 6(4):873, 10/00. Peptidyl prolyl isomerase (PPI) enzymes catalyze the cis-trans interconversion of peptide bonds involving proline. The biological significance of this reaction has long been mysterious; loss of PPI generally does not lead to any observable phenotype. Here they show that dephosphorylation of a regulatory protein acts only on the trans isomer (of a peptide bond very near the phosphate). They show that PPI is essential for the regulatory function, and in this case essential for cell division. Also see Pappenberger et al (2001), below.

B P Tu et al, Biochemical basis of oxidative protein folding in the endoplasmic reticulum. Science 290:1571, 11/24/00. A study of the details of disulfide bond formation, in yeast.

R R Breaker, Molecular biology: Making catalytic DNAs. Science 290:2095, 12/15/00. News. An introduction to the development of DNA enzymes. So far, none have been found in nature, but this is an interesting area of work.

M Ridanpaa et al, Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia. Cell 104(2):195-203, 1/26/01. (Also News item by Clayton, Nature 410:29, 3/1/01.) This RNase is involved in rRNA processing, and also in mitochondrial DNA replication. It is likely, but not certain, that the RNA studied here is a catalytic RNA.

M Labrador et al, Molecular biology: Protein encoding by both DNA strands. Nature 409:1000, 2/22/01. This brief note reports finding a Drosophila gene that violates some of our principles. The protein-coding region comes from both DNA strands -- though apparently from different DNA regions. The two transcripts are partially complementary to each other, and are presumably spliced together to form one mRNA for translation.

B Tsai et al, Protein disulfide isomerase acts as a redox-dependent chaperone to unfold cholera toxin. Cell 104(6):937-948, 3/23/01. An unusual story of disulfide bonds, but of interest because it involves a disease process. It is also a story of chaperones. Host PDI is involved in activation of cholera toxin. It does not affect the disulfide bonds of the toxin. The redox cycle of the PDI affects its shape, and its role as a chaperone to unfold the toxin, thus allowing its transport to the cytosol. That is, instead of the chaperone being driven by an ATP-ADP cycle, it is driven by a redox cycle.

S Hurtley et al, Cinderella's coach is ready. Science 291:2337, 3/23/01. Introduction to special issue: Carbohydrates and Glycobiology. Several articles on glycobiology, including glycoproteins.

W K Johnston et al, RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. Science 292:1319, 5/18/01. (+ News, Davenport, p 1278.) If there was a primitive RNA world, then RNA enzymes must have replicated RNA. This work shows that is possible.

G Pappenberger et al, Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics. Nature Structural Biology, 8(5):452-458, 5/01. (+ News, p 380.) Cis-trans isomerism of amides is a formal possibility, but usually not considered a big issue, because of rapid interconversion. That prolyl amides interconvert slowly has long been known, and there are enzymes to mediate the isomerization of prolyl peptide bonds (Zhou et al, 2000, above). This work suggests that, surprisingly, slow isomerization of some non-prolyl peptide bonds may be relevant to the rate of protein folding.

R Lewis & B A Palevitz, Genome economy. The Scientist 6/11/01, p 19. A discussion of how we humans may make 100-200,000 proteins from only 35,000 genes. Alternative splicing is undoubtedly one big part of the story -- but only part. Nice article.

A E McBride & P A Silver, State of the Arg: Protein methylation at arginine comes of age. Cell 106(1):5-8, 7/13/01. Another example of a post-translational modification of proteins. They discuss several examples, including myosin. Recent work suggests that histone methylation may be relevant to gene control; we may mention this in Ch 10.

C L Brooks et al, Statistical thermodynamics: Taking a walk on a landscape. Science 293:612, 7/27/01. News. Discussion of current ideas on protein folding. The broad point is that folding pathways follow generally downward paths on a complex energy landscape.

M. Computer resources

The Crick (1970) article is available online, as listed above.

If you have not already done so, I strongly encourage you to get and install RasMol, for viewing molecular structures. This was introduced in the Computer resources section of the Ch 2 handout, and details are at the web site.

You can use RasMol to see good examples of secondary structure; see my RasMol page.

The following site was recommended to me as excellent for exploring protein structure. I have had only a brief time to try it. Please let me know how it works for you. Perhaps you can offer some specific suggestions on using it.
http://info.bio.cmu.edu/Courses/BiochemMols/ProtG/ProtGMain.htm

I have posted a page at my site on the general properties of Amides. It was written for my introductory organic chemistry class, but you may find it useful. Remember that the peptide bond is an amide linkage. The resonance structures of an amide lead to it not behaving as you might expect.

N. Computer resources -- Ch 4-5

> There is no handout for Ch 4-5 (Molecular Cloning Methods; Molecular Tools for Studying Genes and Gene Activity). The following item is listed under Ch 4-5 on the X107 Internet Resources page.

The Biology Animations Library, DNA Learning Center, Cold Spring Harbor Laboratory:
http://vector.cshl.org/resources/BiologyAnimationLibrary.htm
Among the topics listed there are PCR (Ch 4) and Southern blotting (Ch 5). Each topic contains a tutorial, with animations.

Other methods-related sites are listed in the various topic-oriented sections at the end of the X107 Internet Resources page. See list of sections at the top.

O. Homework

Do the questions at end of chapter. In doing #7-15 note that most of these are on material that is introduced here, but which will be expanded later. Thus your analysis of these questions may be rather simple for now.

My questions…

1. What effect would you expect glycosylation of a protein to have on its solubility, based on chemical principles? ("Glycosylation" refers to the attachment of sugars to the protein. This is an example of post-translational modification of proteins.)

2. What effect on the functionality of a protein would you expect if one amino acid is deleted in a β-sheet region? Explain. (Assume that the β-sheet actually forms in the mutant protein. Look at where the side chains are.)

3. Weaver provides a table of the standard amino acids (aa), Fig 3.2. However, he does not classify them by the type of side chain. So let's do that here. Classify the 20 aa into the following groups, based on the side chain: non-polar; polar with acidic side chain; polar with basic side chain; polar with neutral side chain. (Caution: Some of the aa are not easily classified in this way. Don't worry if you aren't sure of a few of them, or disagree with the classification I show. In fact, different books will show them classified differently. Maybe you should even make a list of ambiguous amino acids.)

4. Consider the following two peptides:

1. Arg Pro Ile Val Thr Thr Lys Lys Ser Ser Val Pro Met Asn Trp

2. Arg Pro Ile Leu Ser Thr Arg Glu Tyr Ser Val Phe Met Gln Trp

a. We define "homology" as the percentage of identity in two amino acid sequences. What is the homology between peptides 1 and 2?

b. "Similarity" is the same idea as homology, but also includes amino acids that are chemically similar -- presumably also functionally similar. What is the similarity between peptides 1 and 2? Use Fig 3.2 as a guide to determining similarity.

c. Consider also peptide 3: Arg Pro Ile Val Thr Thr Lys Ser Ser Val Pro Met Asn Trp. What is the homology between peptides 1 and 3?

5. You are studying the interaction between two proteins. You find that the stability of the dimer between them changes dramatically around pH 5. Based on this information, what can you conclude about the kind of amino acid that is involved in this interaction? Explain.

6. Estimate the width of the α-helix, Fig 3.4, based on your general understanding of atomic size. (Anything within a factor of 2 or 3 is probably reasonable for most students at this point.) As a context for making this estimate… Recent work (which we will look at in Ch 18) suggests that a newly made protein exits the ribosome through a channel that is only 10 Å (1 nm) wide. Is it possible for α-helical secondary structure to form in the exit channel?

7. Is it reasonable that a particular sequence of (say) 10 amino acids might be found in an α-helix in one situation and in a β-sheet in another situation? Explain. Would you answer differently if we focused on a particular 10 amino acid region of a particular protein?

8. Let's go back to DNA, now that we have discussed weak bonds more systematically. We know that base pair specificity is due to hydrogen bonds. But another contributor to the general double helical structure of DS DNA is hydrophobic forces, as the aromatic base rings stack on top of each other; this effect is often called "stacking forces". Consider the following simple view of DNA renaturation (Ch 2 handout Sect K): a base moves from water into the interior of a double helix. How does this move affect (each of) these two main contributors to the energy of the double helix?

9. Weaver sneaked an English lesson into Ch 3. What is it?

10. Improve the chapter summary.

P. Partial answers

1. It should increase solubility, due to the increased number of polar groups (-OH) on the sugars.

> Pharmaceutical companies are experimenting with attaching polyethylene glycol (PEG) to proteins to increase their solubility. Same principle.

2. The key issue is that alternate side chains in a β-sheet point in opposite directions (Fig 3.5). Thus a one amino acid deletion will result in a complete change in which amino acids are on which side (beyond the point of the mutation). Of course, the phenotypic effect of this depends on the functional role of the particular β-sheet. But unless it is more or less filler sequence, with no specific contacts, it's likely to be highly disruptive.

3. Here is one scheme, from a chemistry book (Seager & Slabaugh, Chemistry for Today, 3/e, 1997, Table 19.1, p 561).

non-polar: Gly, Ala, Val, Leu, Ile, Phe, Pro, Met
polar acidic: Asp, Glu
polar basic: His, Lys, Arg
polar neutral: Ser, Thr, Tyr, Trp, Cys, Asn, Gln

As noted, some of these are questionable. In particular:

Gly. The side chain is so small that it has minimal effect on the polarity at all.
Cys. The -SH group is so very slightly polar that it is often called non-polar. The -SH group is also slightly acidic, but so slightly that it is not usually of concern. Both of these issues detract from the main feature of the Cys side chain, its ability to be oxidized to the disulfide.

Others may come up for discussion.

4. a. 8 of the 15 are the same. 53% homology.

b. Look at each change. Is it to a "similar" aa??? Here's one way of looking at it:

Yes: Val-Leu, Thr-Ser, Lys-Arg, Asn-Gln

No: Lys-Glu (opposite charge), Ser-Tyr (very different size, one is aromatic), Pro-Phe (Pro is bent, Phe is aromatic)

By this classification, 4 of the changes are "conservative"; along with 8 identical aa, that gives 12/15 = 80% similarity.

c. 93% (14/15). #3 is missing one in the middle, but otherwise lines up perfectly. Do you think this is fair? Is it a relevant way to consider homology in this case?

5. A transition at pH 5 suggests that the ionization of carboxyl groups is relevant to the interaction. It thus implicates acidic amino acids (Asp, Glu) in the interaction.

6. Atoms are about 1 Å across. Since the helix is a few atoms wide, one might estimate the width of the helix as a few Å. If you estimated 5-10 Å, that is very good. Obviously, the width varies, depending on the specific amino acids. The point is that α-helical structure may form within the exit pore -- barely. It is unlikely that any more complex structure would form.

7. The big issue is that the conformation of a protein is the lowest energy conformation -- in that environment. Thus the same peptide sequence may form different conformations under different situations -- different conditions or different neighbors. Might this extend to as dramatic a difference as α-helix vs β-sheet? The evidence says yes. Artificial (model) examples were first made in 1996. More recently, natural examples have been found. For example, Tan & Richmond (1998) studied the structure of a regulatory complex of DNA and protein in yeast. They found that one key region was in an α-helix in one copy of the MATα2 protein, and in a β-sheet in another copy. It may also be a useful perspective that some amyloid proteins, such as that in Alzheimer's disease, are due to α-helix → β-sheet changes, converting a normal protein to a toxic protein in this case.

8. The base is hydrogen bonded in either case, either to water or to its paired base. However, the base stacking energy is truly a net gain. The sides of the bases go from the unfavorable environment of water to the favorable environment of stacking with the sides of other bases, which are similar. Thus the stacking interaction is the main driving force behind formation of the double helix. Within that environment, the bases prefer correct pairing to incorrect, because of favorable H-bonds. The work with poorly H-bonding bases (e.g., Matray & Kool, 1999, Ch 2 FR) reflects the importance of stacking interactions.

9. Need a hint? Look carefully at p 39.

x107a\wv3h
9/8/01 (minor updates 4/1/02, 6/4/02, 12/3/14)