بیوشیمی پروتئین‌ها و اسیدهای نوکلئیک

Biochemistry of Proteins and Nucleic Acids

هدف کلی: آشنایی با ساختار، خصوصیات و عملکرد پروتئین‌ها و اسیدهای نوکلئیک

General Goal: Familiar with structure, properties and function of proteins and nucleic acids

سرفصل (Syllabus)

پروتئین‌ها

- واحدهای سازنده پروتئین
- میان‌کنش‌های بین و درون مولکولی در ساختار ماکرومولکول‌ها
- سطوح مختلف ساختاری در پروتئین‌ها
- تاخوردگی پروتئین و ارتباط آن با پایداری
- نقش حلال در ساختار و فعالیت پروتئین
- رابطه ساختار و عملکرد پروتئین‌ها

اسیدهای نوکلئیک

- واحدهای سازنده اسیدهای نوکلئیک
- تشکیل جفت باز و Stacking در اسیدهای نوکلئیک
- پارامترهای ساختمانی در اسیدهای نوکلئیک
- آرایش فضایی بازها و قندها در انواع ساختارهای اسیدهای نوکلئیک
- انواع آرایش‌های فضایی اسیدهای نوکلئیک
- ساختارهای خاص در اسیدهای نوکلئیک (ساختارهای سه رشته‌ای، چهاررشته‌ای و...)
- نقش حلال در ساختار اسیدهای نوکلئیک

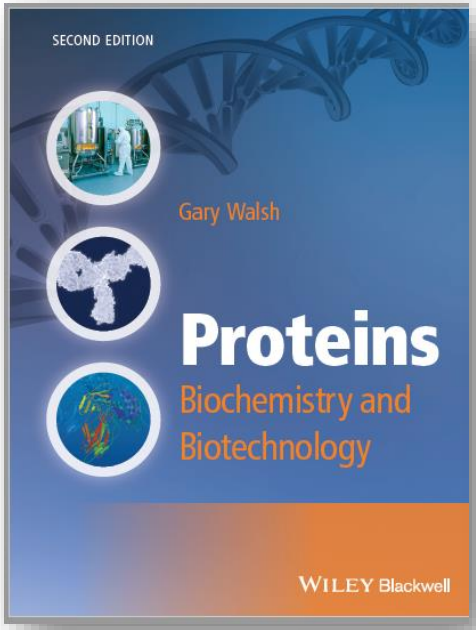
Proteins

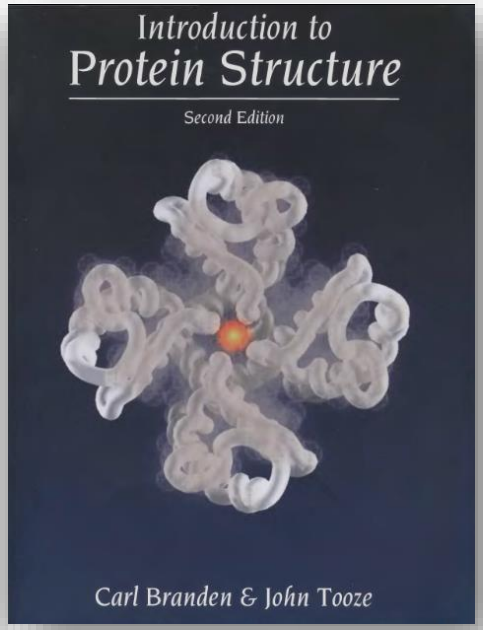
- Building blocks of proteins
- Inter and intra-molecular interactions in macromolecules structure
- Different levels of proteins structure
- Protein folding and its relation to stability
- The role of solvent in structure and activity of protein
- Protein structure-function relationships

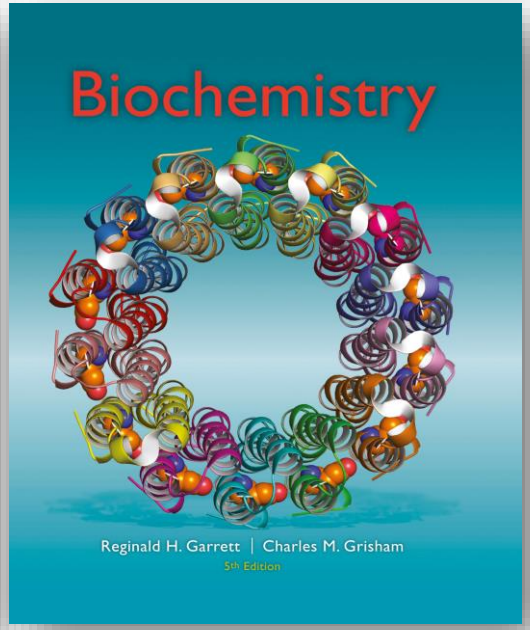
Nucleic Acids

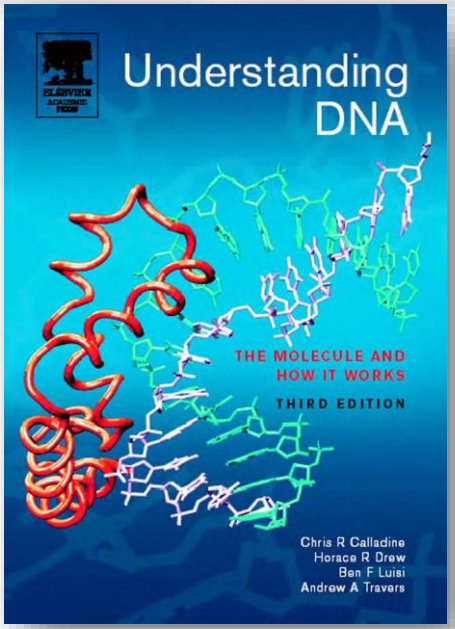
- Building blocks of nucleic acids
- Base pairing and stacking in nucleic acids
- Structural parameters in nucleic acids
- Spatial arrangement of sugar and bases in structural variants of nucleic acids
- Unusual nucleic acids structures (triplex, tetraplex and ….)
- The role of solvent in structure of nucleic acids

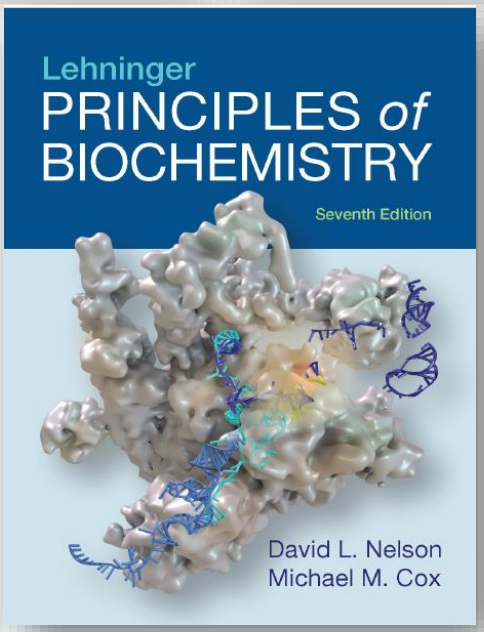
منابع (References)











Proteins, an introduction

Each polypeptide consists of a chain of amino acids linked together by peptide (amide) bonds.

The exact amino acid sequence is determined by the gene coding for that specific polypeptide.

When synthesized, a polypeptide chain folds up, assuming a specific three-dimensional shape (i.e. a specific conformation) that is unique to the protein. The conformation adopted depends on the polypeptide's amino acid sequence, and this Conformation is largely stabilized by multiple, weak interactions.

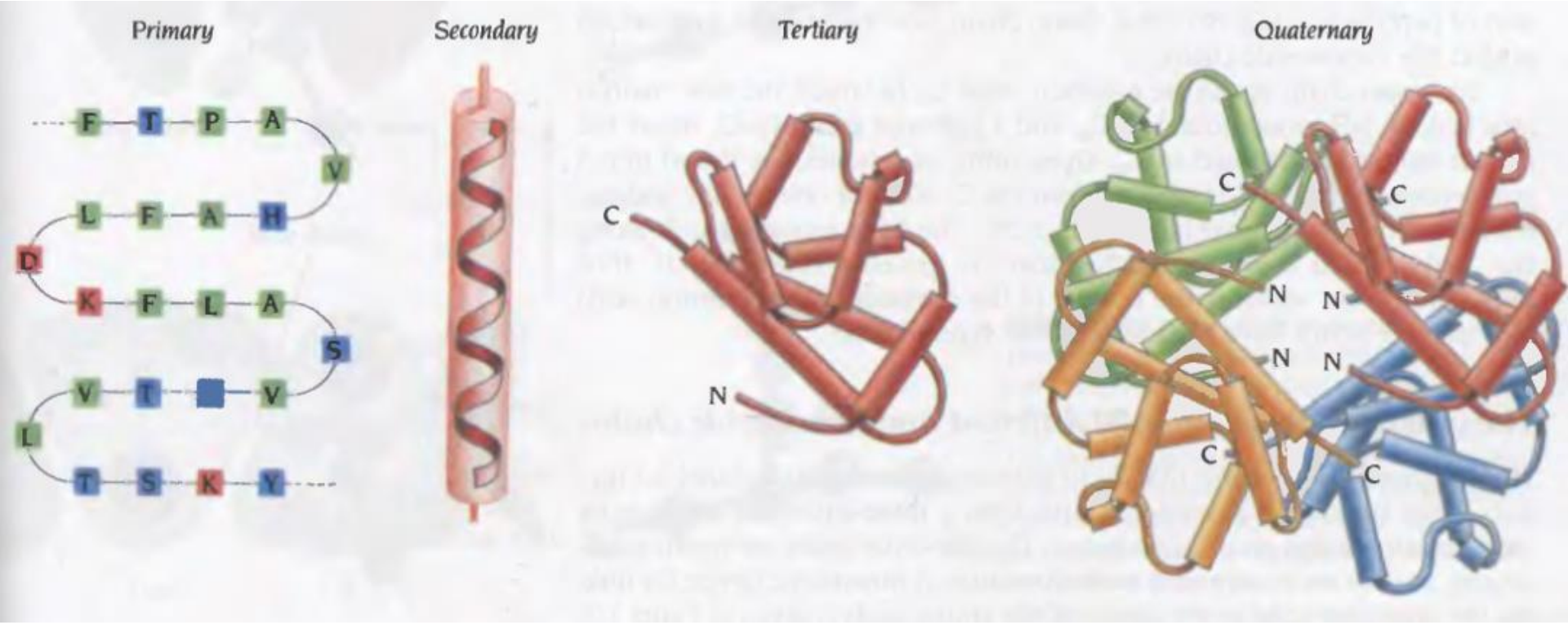
Overall, a protein's structure can described at up to four different levels.

- Primary structure: the specific amino acid sequence of its polypeptide chain(s), along with the exact positioning of any disulfide bonds present.
- Secondary structure: regular recurring arrangements of adjacent amino acid residues, often over relatively short contiguous sequences within the protein backbone. The common secondary structures are the α-helix and β-strands.
- Tertiary structure: the three-dimensional arrangement of all the atoms which contribute to

the polypeptide. In other words, the overall three-dimensional structure (conformation) of a polypeptide chain, which usually contains several stretches of secondary structure interrupted by less ordered regions such as bends/loops.

- Quaternary structure: the overall spatial arrangement of polypeptide subunits within a protein composed of two or more polypeptides.

A protein's structure can described at up to **four** different levels.



PRIMARY STRUCTURE

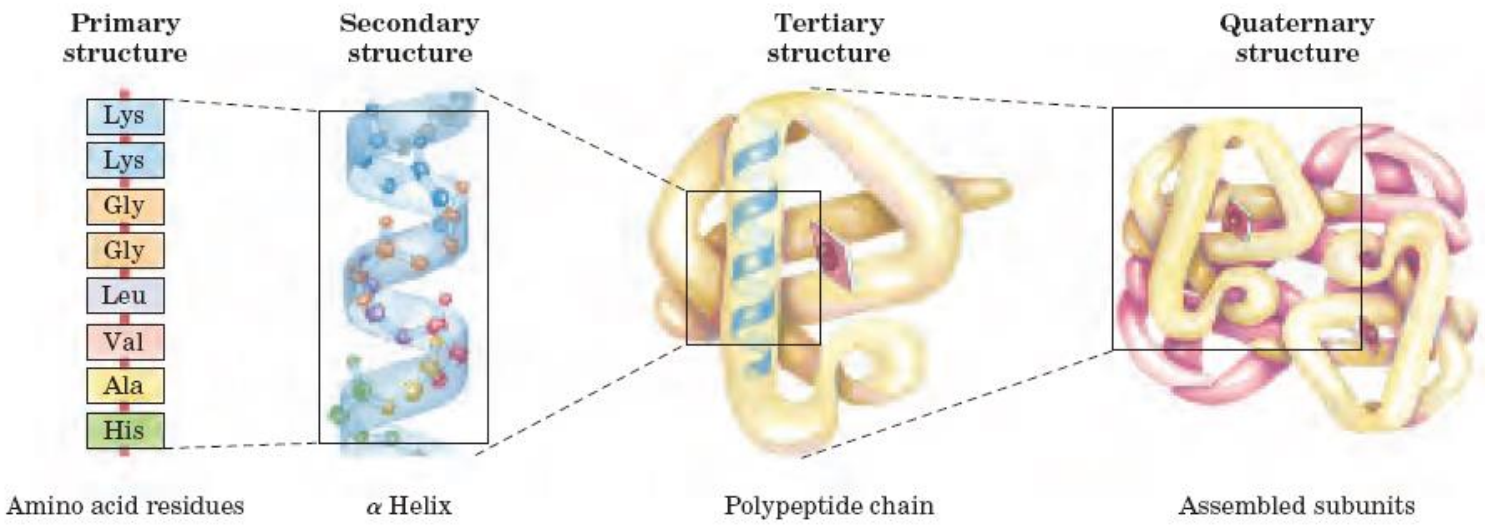


FIGURE 3–16 Levels of structure in proteins. The *primary structure* consists of a sequence of amino acids linked together by peptide bonds and includes any disulfide bonds. The resulting polypeptide can be coiled into units of *secondary structure,* such as an α helix. The helix is a part of the *tertiary structure* of the folded polypeptide, which is itself one of the subunits that make up the *quaternary structure* of the multisubunit protein, in this case hemoglobin.

Table 2.1 The 20 commonly occurring amino acids. They may be subdivided into five groups on the basis of side-chain structure. Their three- and one-letter abbreviations are also listed (one-letter abbreviations are generally used only when compiling extended sequence data, mainly to minimize writing space and effort). In addition to their individual molecular masses, the per cent occurrence of each amino acid in an 'average' protein is also presented. This data was generated from sequence analysis of over 1000 different proteins.

R group classification	Amino acid	Abbreviated name (3 letter)	Abbreviated name (1 letter)	Molecular mass (Da)	Per cent occurrence in 'average' protein
Non-polar, aliphatic	Glycine	Gly	G	75	7.2
	Alanine	Ala	A	89	8.3
	Valine	Val	V	117	6.6
	Leucine	Leu	L	131	9.0
	Isoleucine	Ile	I	131	5.2
	Proline	Pro	P	115	5.1
Aromatic	Tyrosine	Tyr	Y	181	3.2
	Phenylalanine	Phe	F	165	3.9
	Tryptophan	Trp	W	204	1.3
Polar but uncharged	Cysteine	Cys	C	121	1.7
	Serine	Ser	S	105	6.0
	Methionine	Met	M	149	2.4
	Threonine	Thr	T	119	5.8
	Asparagine	Asn	N	132	4.4
	Glutamine	Gln	Q	146	4.0
Positively charged	Arginine	Arg	R	174	5.7
	Lysine	Lys	K	146	5.7
	Histidine	His	H	155	2.2
Negatively charged	Aspartic acid	Asp	D	133	5.3
	Glutamic acid	Glu	E	147	6.2

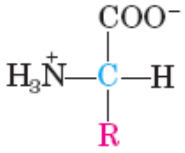


FIGURE 3–2 **General structure of an amino acid.** This structure is common to all but one of the α-amino acids. (Proline, a cyclic amino acid, is the exception.) The R group or side chain (red) attached to the α carbon (blue) is different in each amino acid.

The Amino Acid Residues in Proteins Are L Stereoisomers

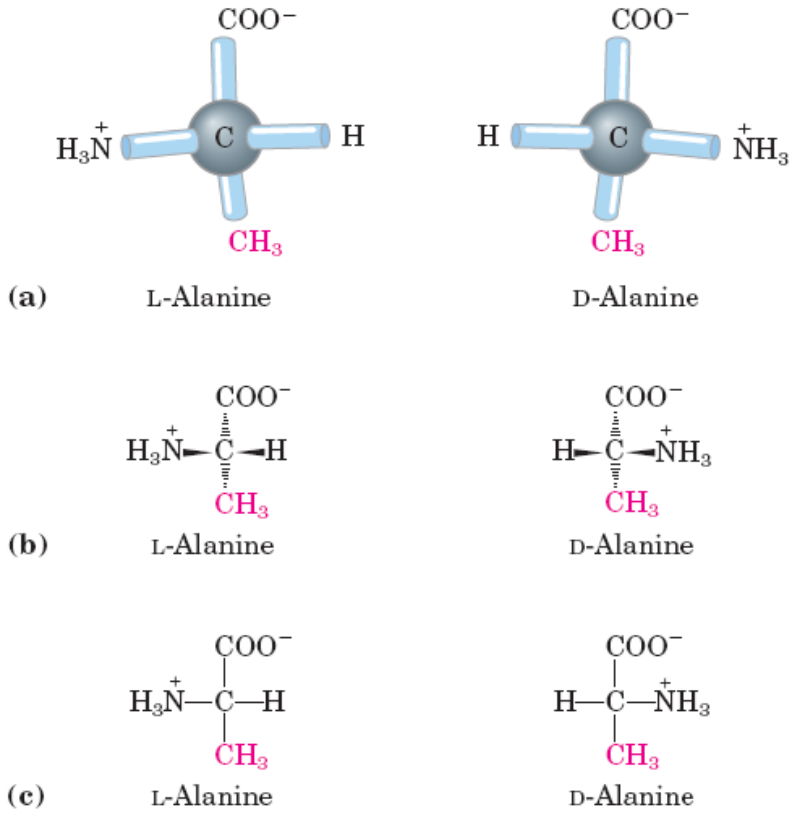


FIGURE 3–3 **Stereoisomerism in α-amino acids.** (a) The two stereoisomers of alanine, L- and D-alanine, are nonsuperimposable mirror images of each other (enantiomers). (b, c) Two different conventions for showing the configurations in space of stereoisomers. In perspective formulas (b) the solid wedge-shaped bonds project out of the plane of the paper, the dashed bonds behind it. In projection formulas (c) the horizontal bonds are assumed to project out of the plane of the paper, the vertical bonds behind. However, projection formulas are often used casually and are not always intended to portray a specific stereochemical configuration.

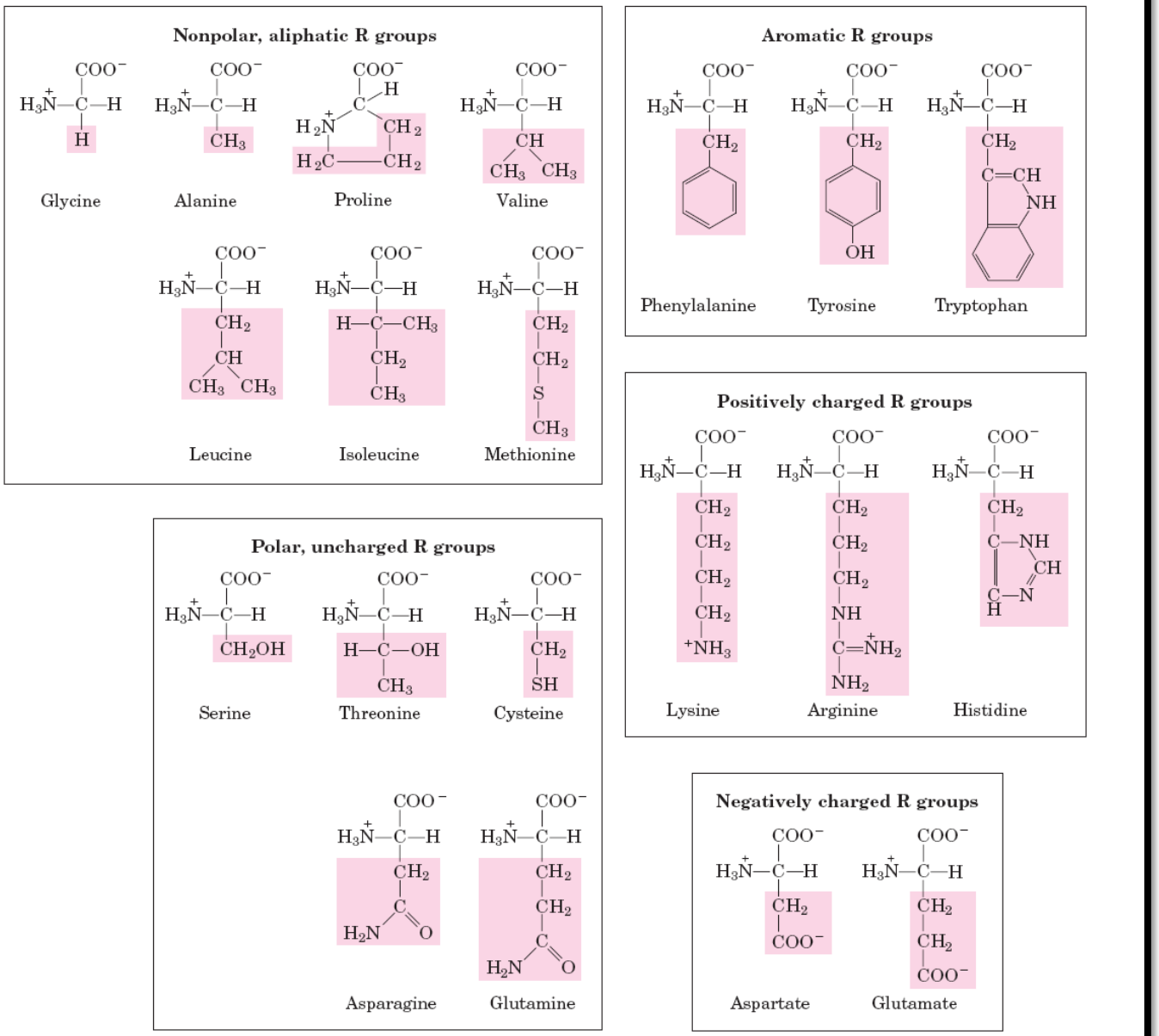
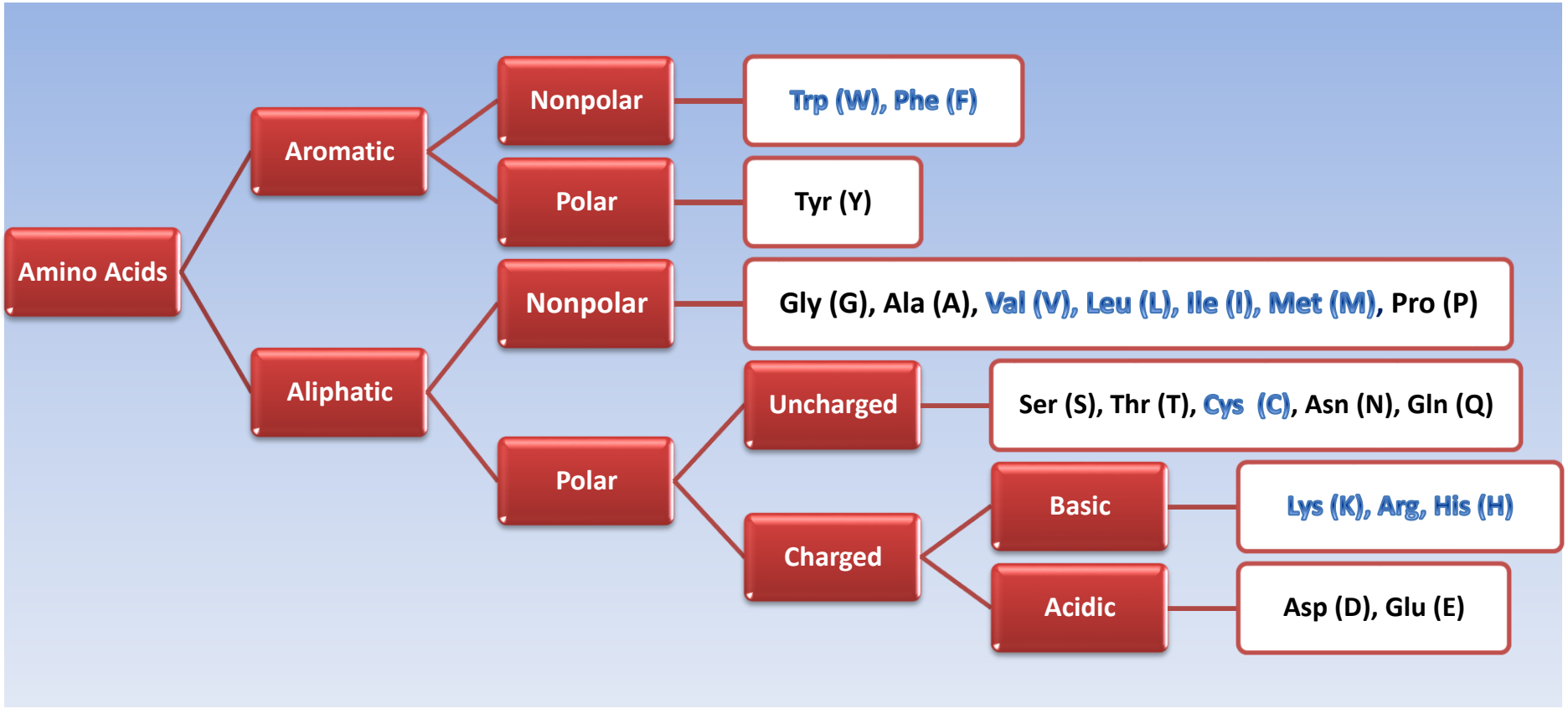
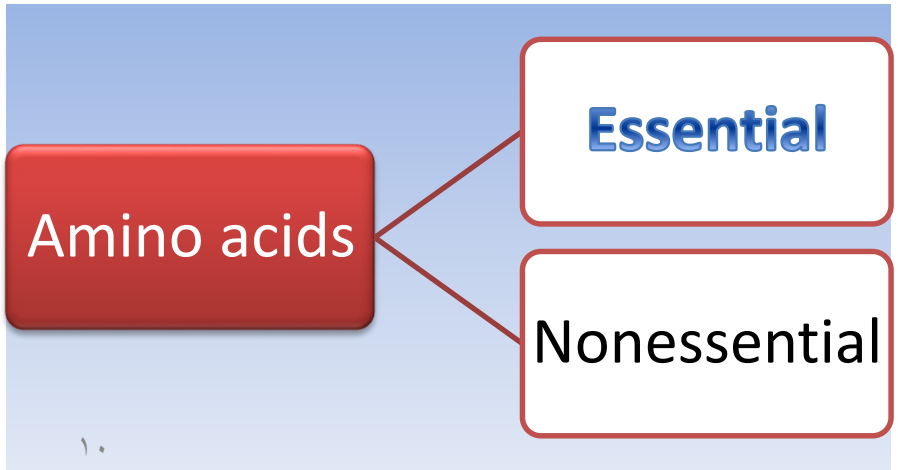
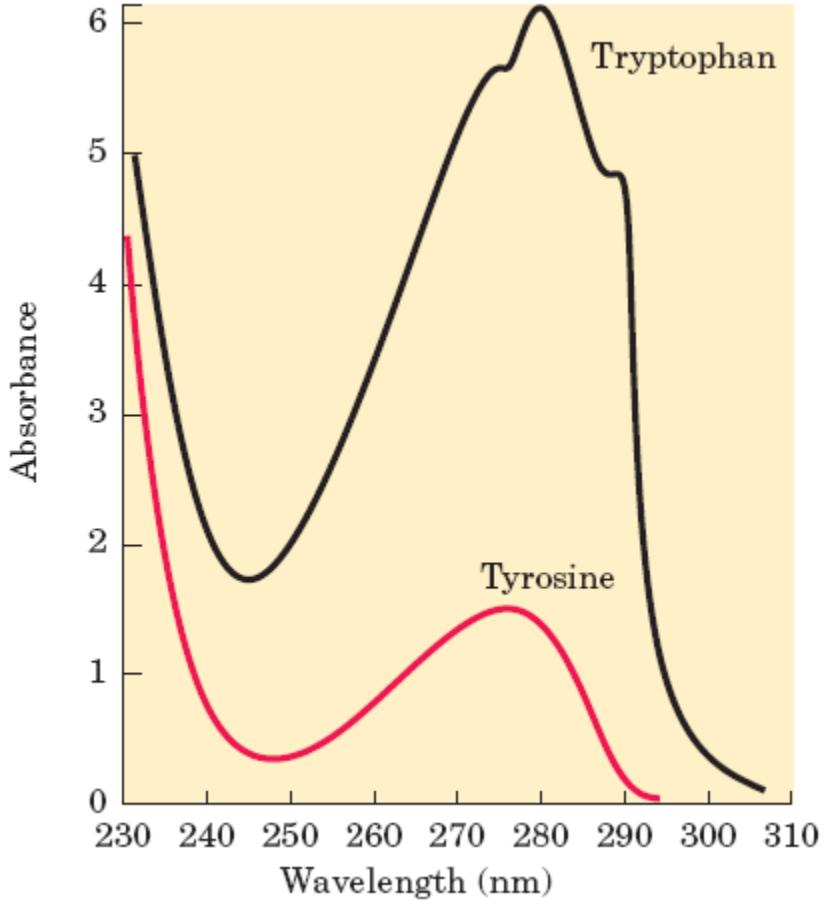


FIGURE 3–5 **The 20 common amino acids of proteins.** The structural formulas show the state of ionization that would predominate at pH 7.0. The unshaded portions are those common to all the amino acids; the portions shaded in red are the R groups. Although the R group of histidine is shown uncharged, its pK_a (see Table 3–1) is such that a small but significant fraction of these groups are positively charged at pH 7.0.





- Standard amino acids (20)
- Prolin (P) → imino acid
- Ile and Thr → Two chiral center
- Gly → without chiral center



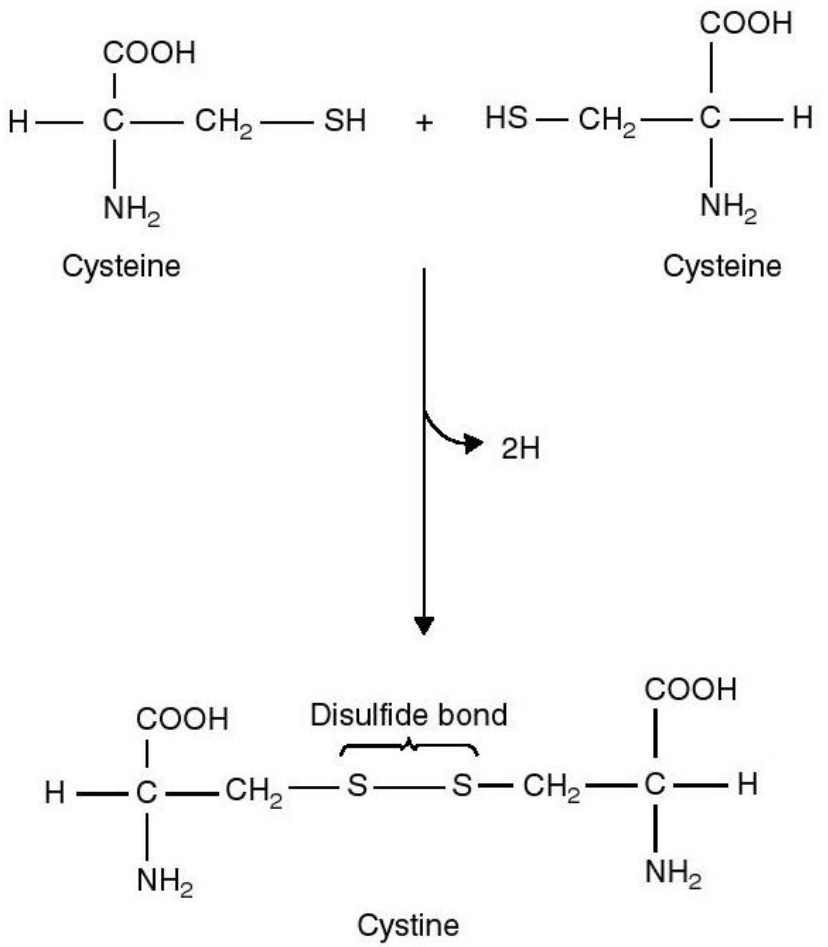
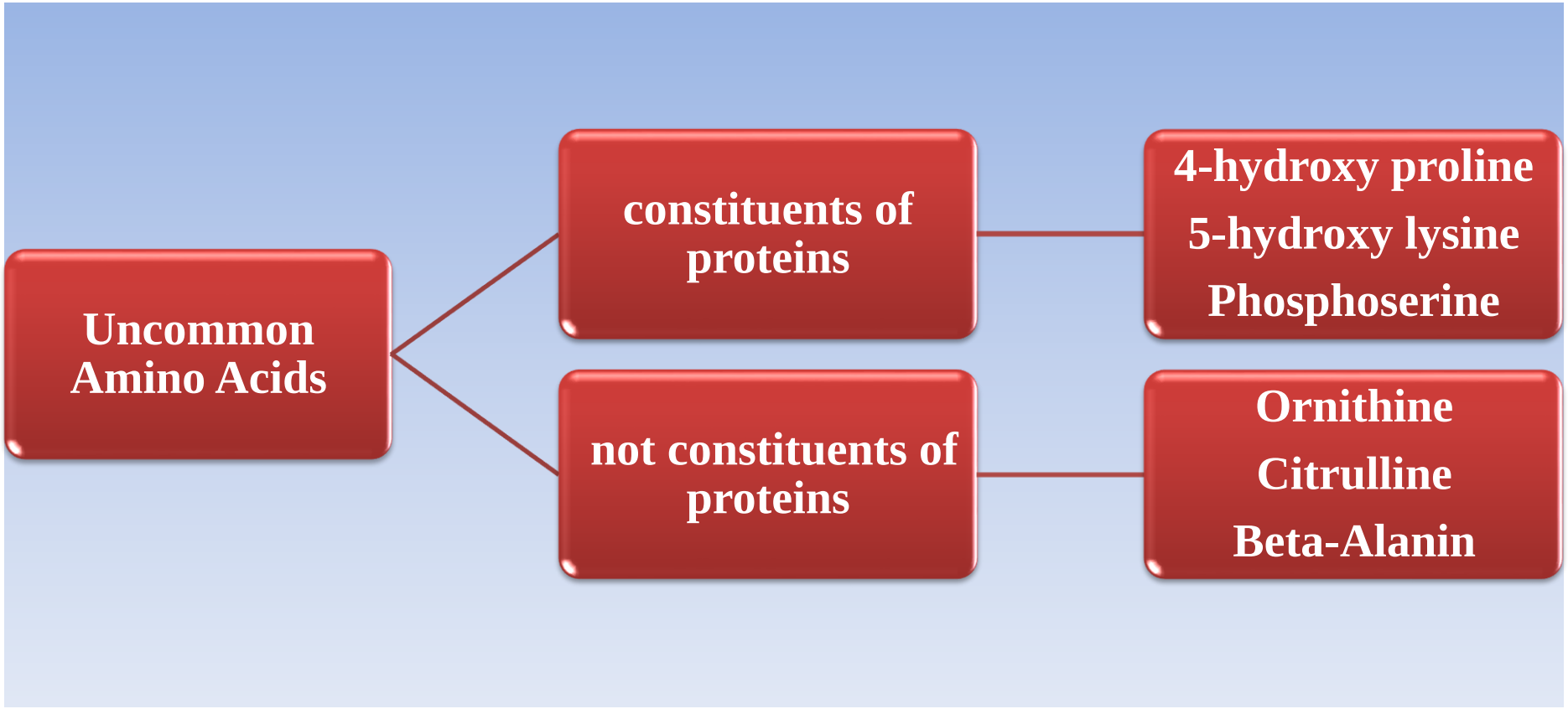
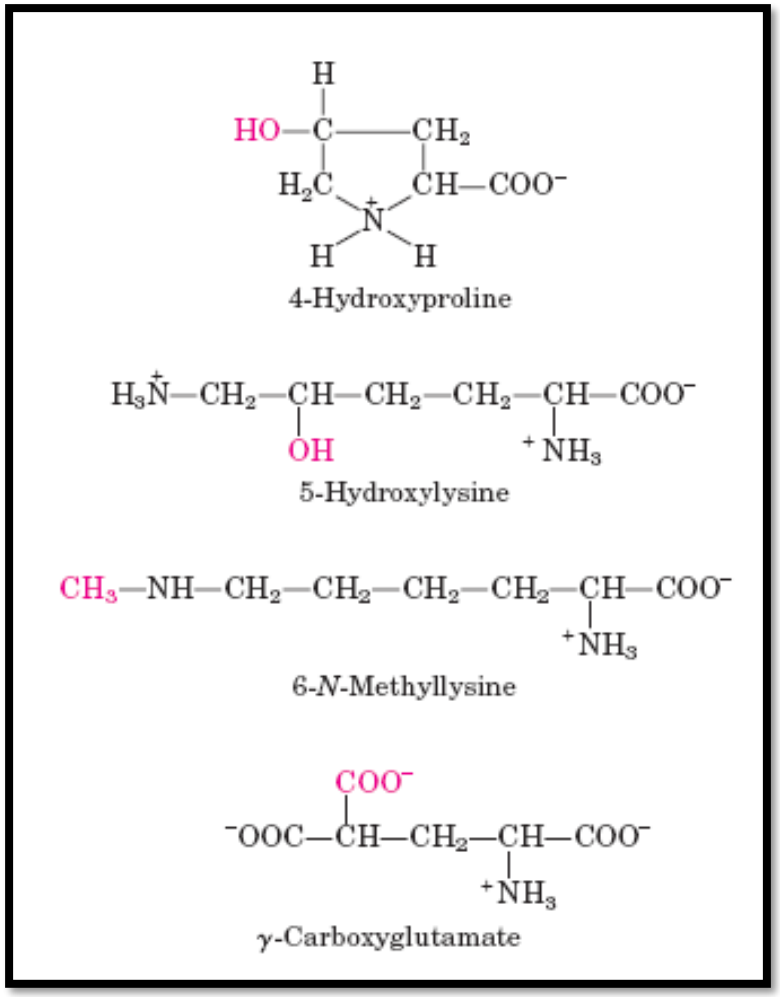


Figure 2.2 The formation of cystine via disulfide bond formation between two cysteines.



Uncommon amino acids created by modification of common residues already incorporated into a polypeptide.



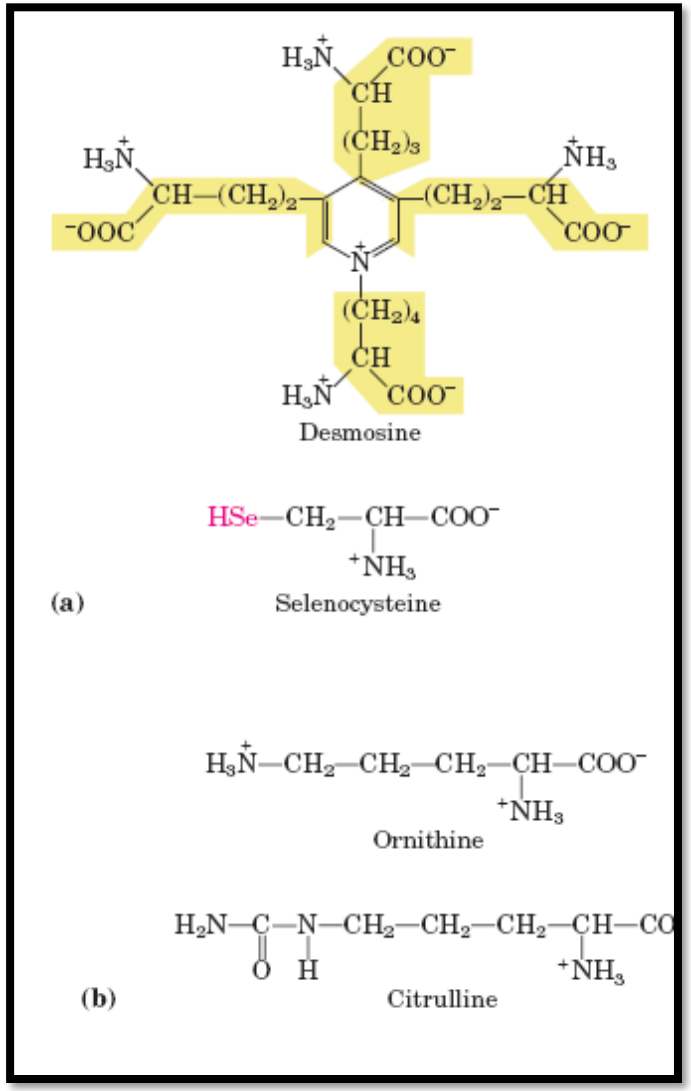


FIGURE 3–8 **Uncommon amino acids.** (a) Some uncommon amino acids found in proteins. All are derived from common amino acids. Extra functional groups added by modification reactions are shown in red. Desmosine is formed from four Lys residues (the four carbon backbones are shaded in yellow). Note the use of either numbers or Greek letters to identify the carbon atoms in these structures. (b) Ornithine and citrulline, which are not found in proteins, are intermediates in the biosynthesis of arginine and in the urea cycle.

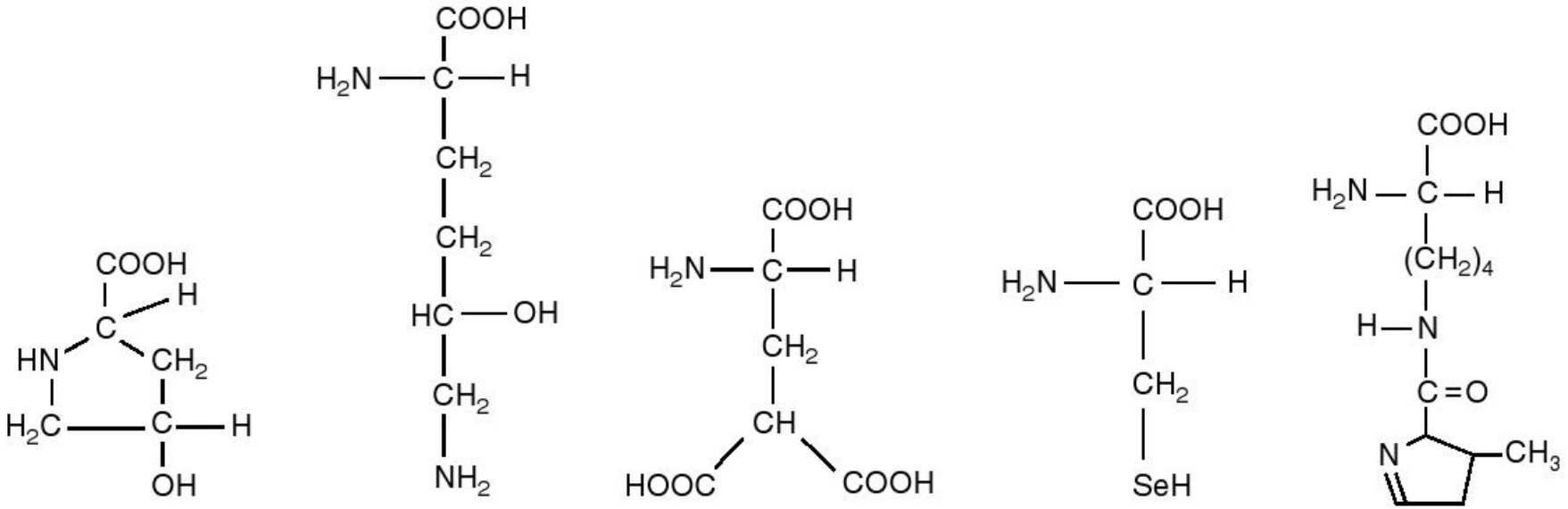
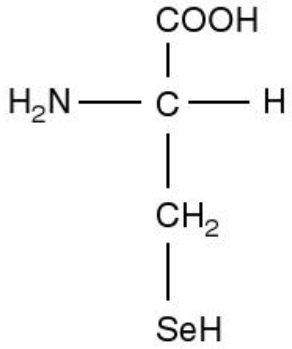


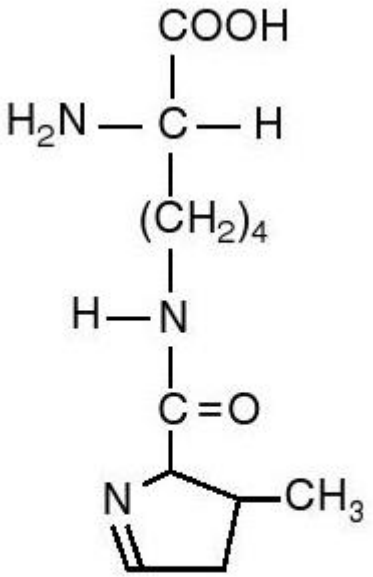
Figure 2.3 Structure of some modified amino acids.

In addition to the 20 common amino acids, some modified amino acids are also found in several proteins. In most instances these modified amino acids are formed by PTM reactions, as discussed later in this chapter. However, two amino acids (selenocysteine and pyrrolysine; Figure 2.3) exist as a preformed amino acid in their own right and are hence sometimes called the 21st and 22nd proteinogenic amino acids.



Selenocysteine

Selenium in the form of selenocysteine (Sec or U) is an essential component of a small number of enzymes in some species (including glutathione peroxidase, thioredoxin reductases and some hydrogenases). The nucleotide sequence of the genes coding for such enzymes contains a UGA codon, which codes for selenocysteine. In non-selenocysteine proteins, UGA normally functions as a termination codon. The reading of UGA as selenocysteine rather than the more usual stop codon is apparently dependent on the presence of a so-called *cis*-acting selenocysteine insertion sequence element.



Pyrrolysine

Pyrrolysine (Pyl or O) displays a side chain similar to lysine, with the presence of an added pyrroline ring at the end of the lysine side chain. Similarly to Sec, Pyl is encoded by a codon which normally functions as a stop signal (UAG), with Pyl insertion likely requiring a pyrrolysine insertion sequence element. Its presence appears to be restricted to a small number of methanogenic, mainly archael, microorganisms, where it appears to reside within the active site of several methyltransferase enzymes, playing a direct catalytic role therein.

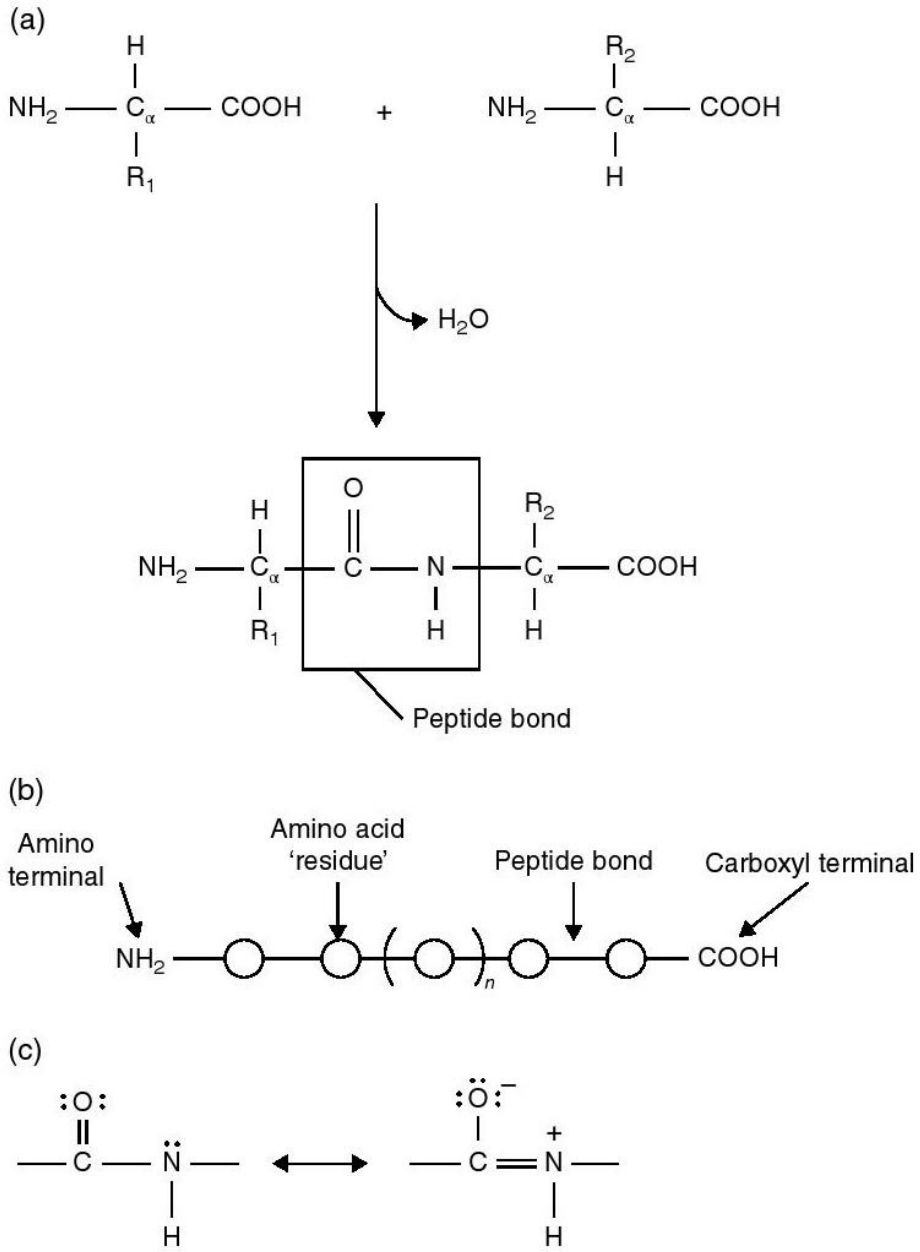


Figure 2.4 (a) Peptide bond formation. (b) Polypeptides consist of a linear chain of amino acids successively linked via peptide bonds. (c) The peptide bond displays partial double-bonded character.

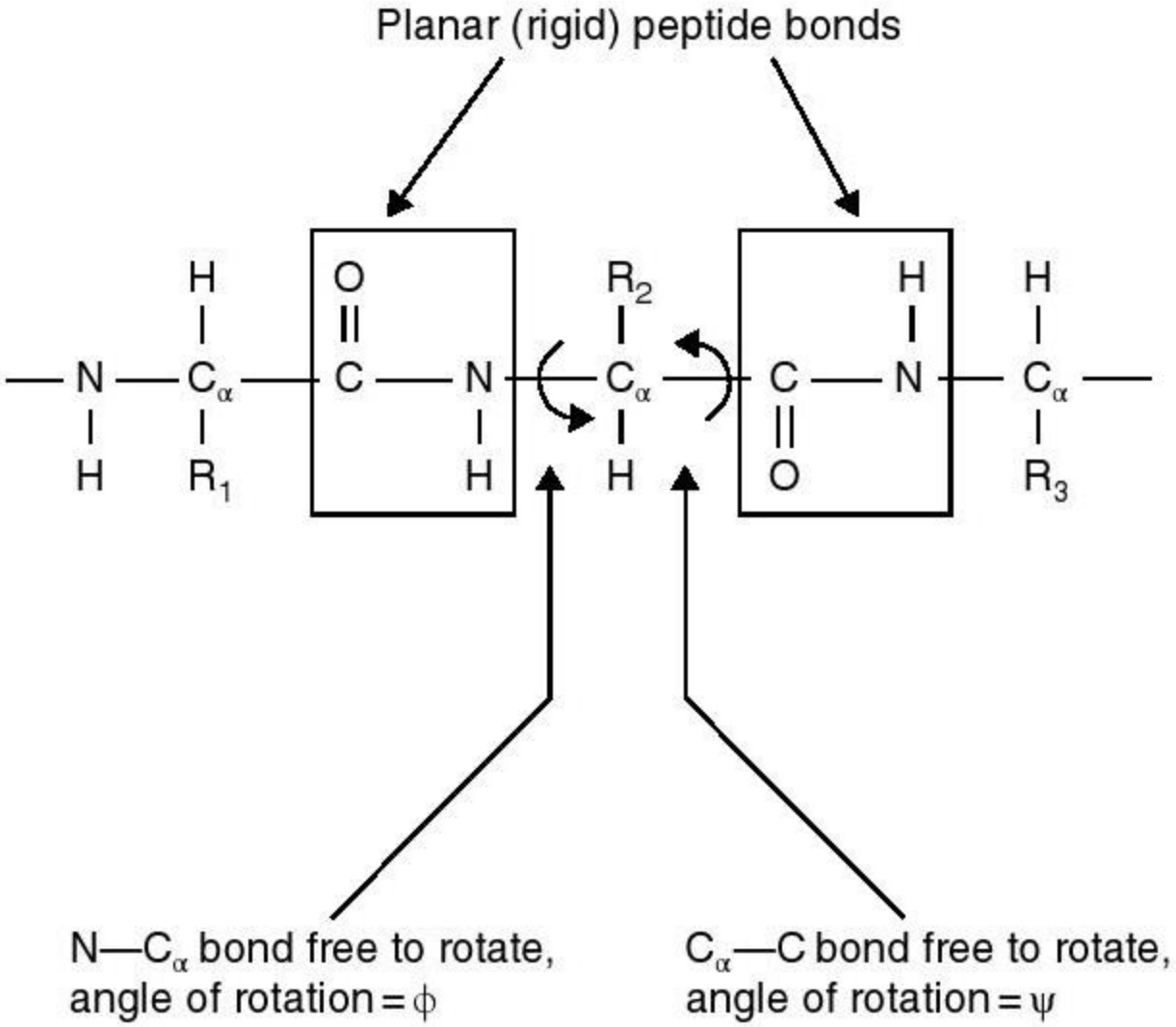


Figure 2.5 Fragment of polypeptide chain backbone illustrating rigid peptide bonds and the intervening N—Cα and Cα—C backbone linkages, which are free to rotate.

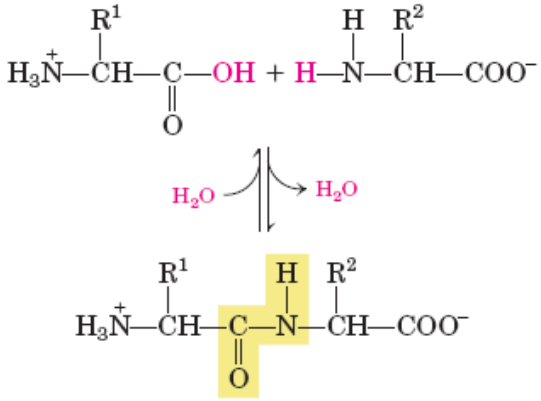


FIGURE 3–13 **Formation of a peptide bond by condensation.** The α-amino group of one amino acid (with R^2 group) acts as a nucleophile to displace the hydroxyl group of another amino acid (with R^1 group), forming a peptide bond (shaded in yellow). Amino groups are good nucleophiles, but the hydroxyl group is a poor leaving group and is not readily displaced. At physiological pH, the reaction shown does not occur to any appreciable extent.

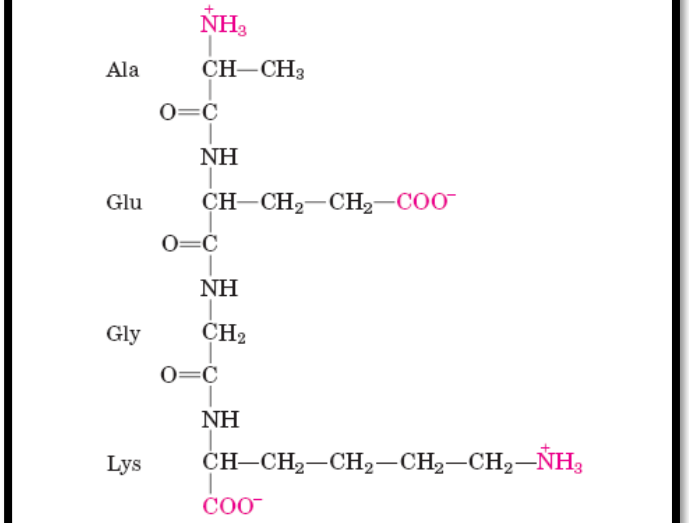


FIGURE 3–15 **Alanylglutamylglycyllysine.** This tetrapeptide has one free α-amino group, one free α-carboxyl group, and two ionizable R groups. The groups ionized at pH 7.0 are in red.

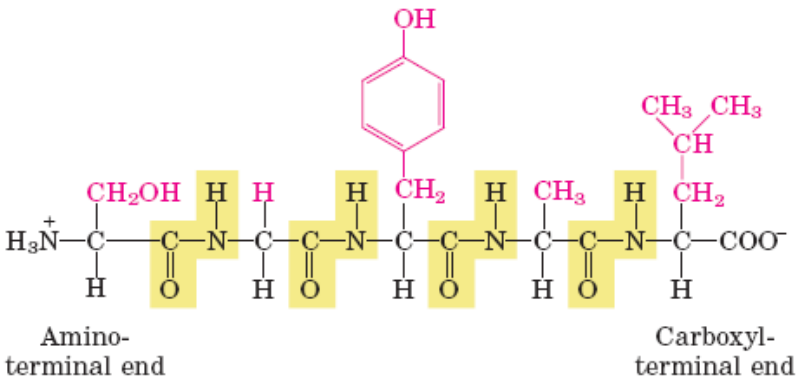
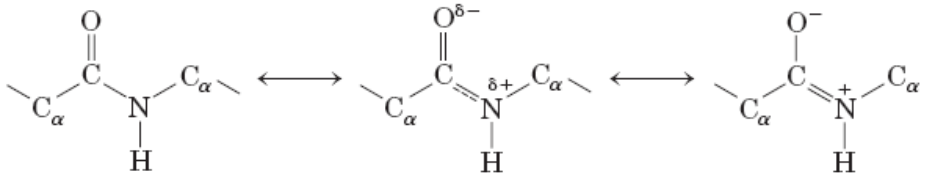
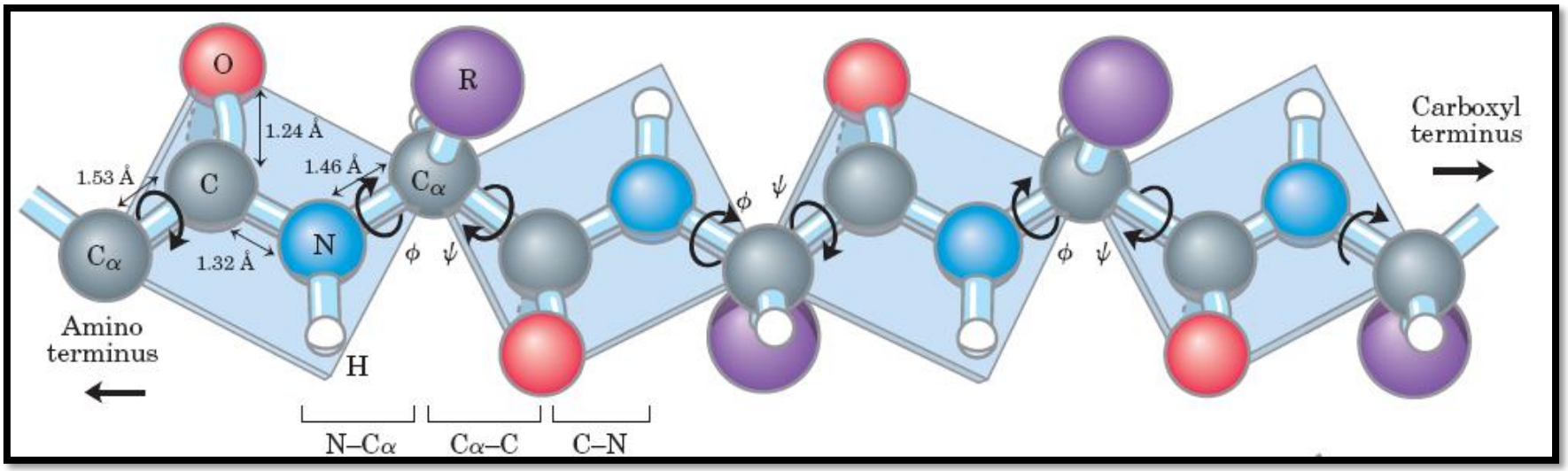


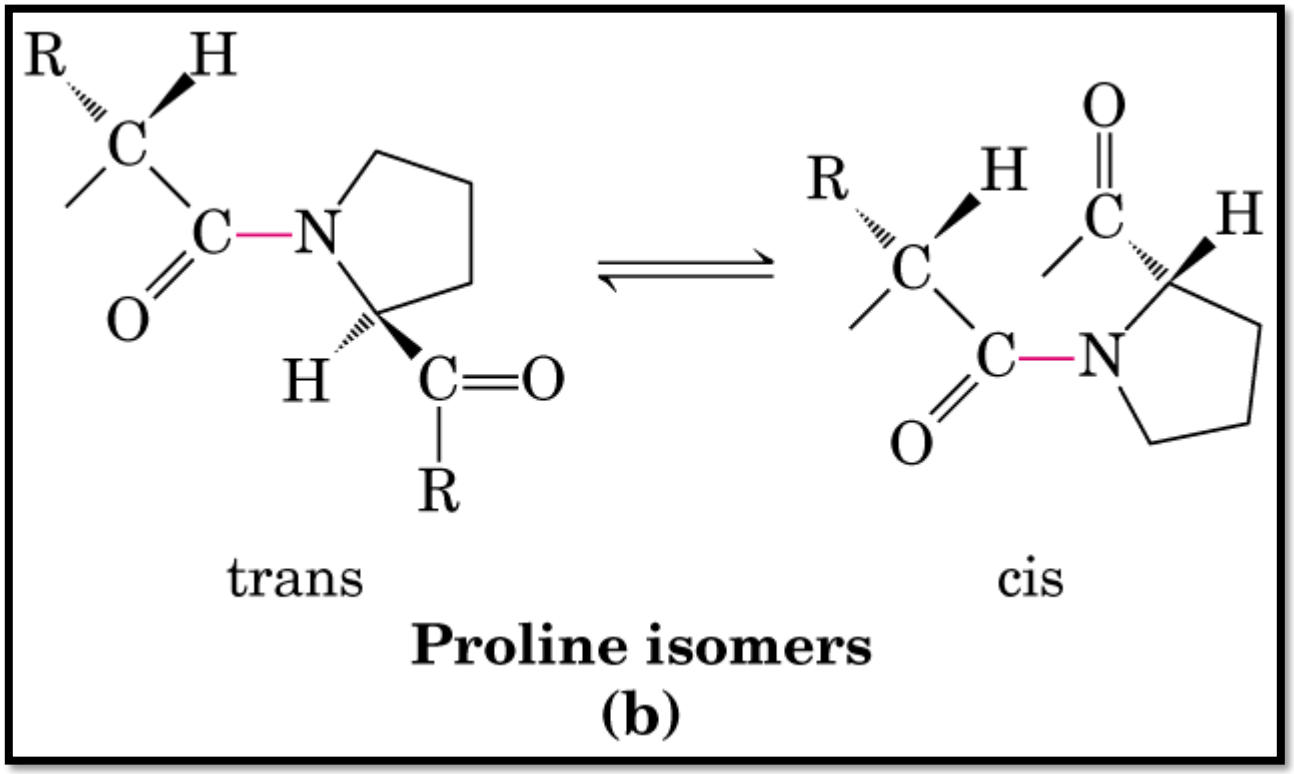
FIGURE 3–14 The pentapeptide serylglycyltyrosylalanylleucine, or Ser–Gly–Tyr–Ala–Leu. Peptides are named beginning with the amino-terminal residue, which by convention is placed at the left. The peptide bonds are shaded in yellow; the R groups are in red.

The Peptide Bond Is Rigid



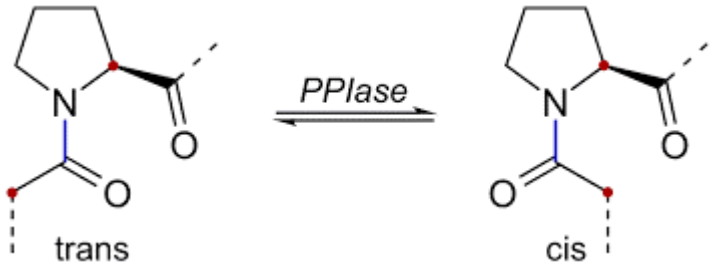
The carbonyl oxygen has a partial negative charge and the amide nitrogen a partial positive charge, setting up a small electric dipole. Virtually all peptide bonds in proteins occur in this trans configuration; an exception is noted in Figure 4–8b.





Proline isomers

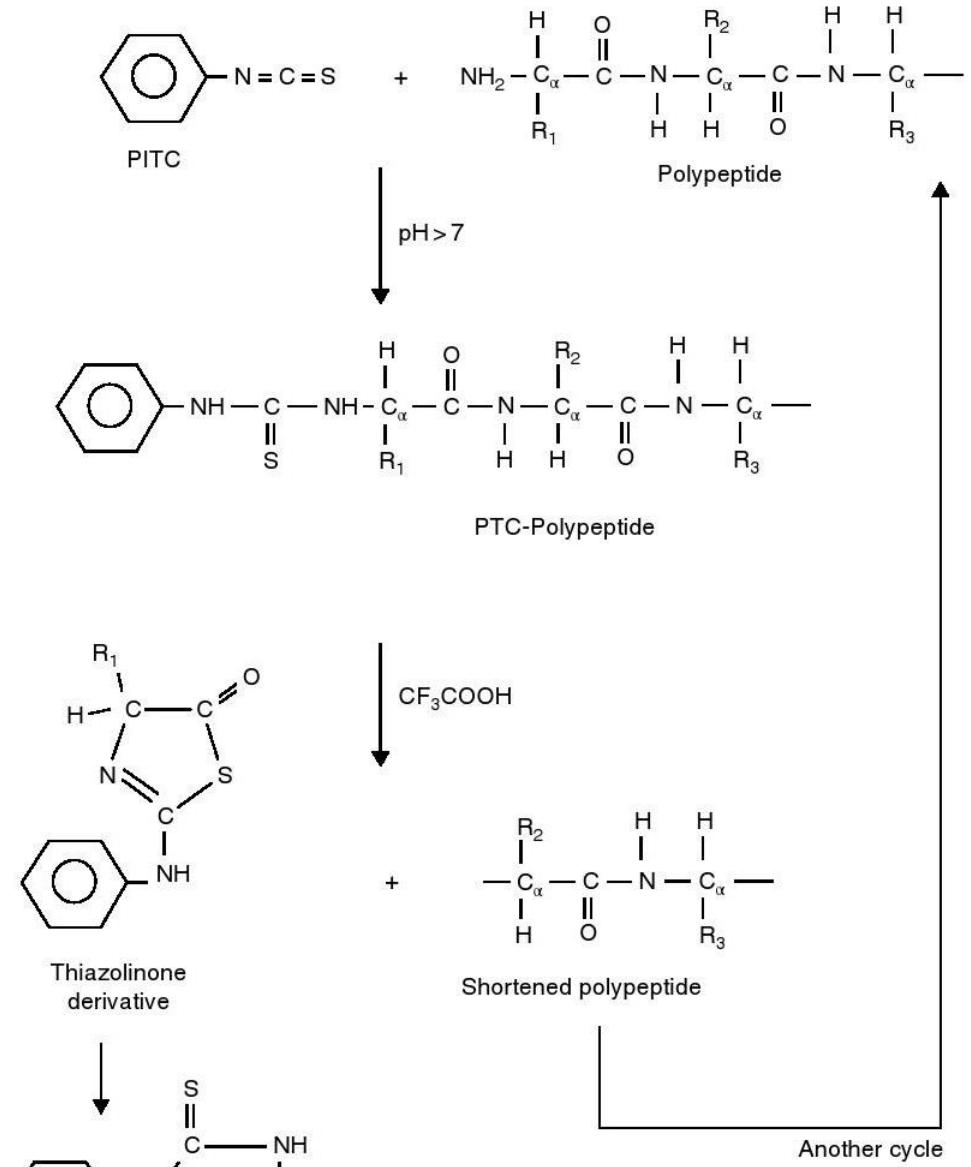
(b)

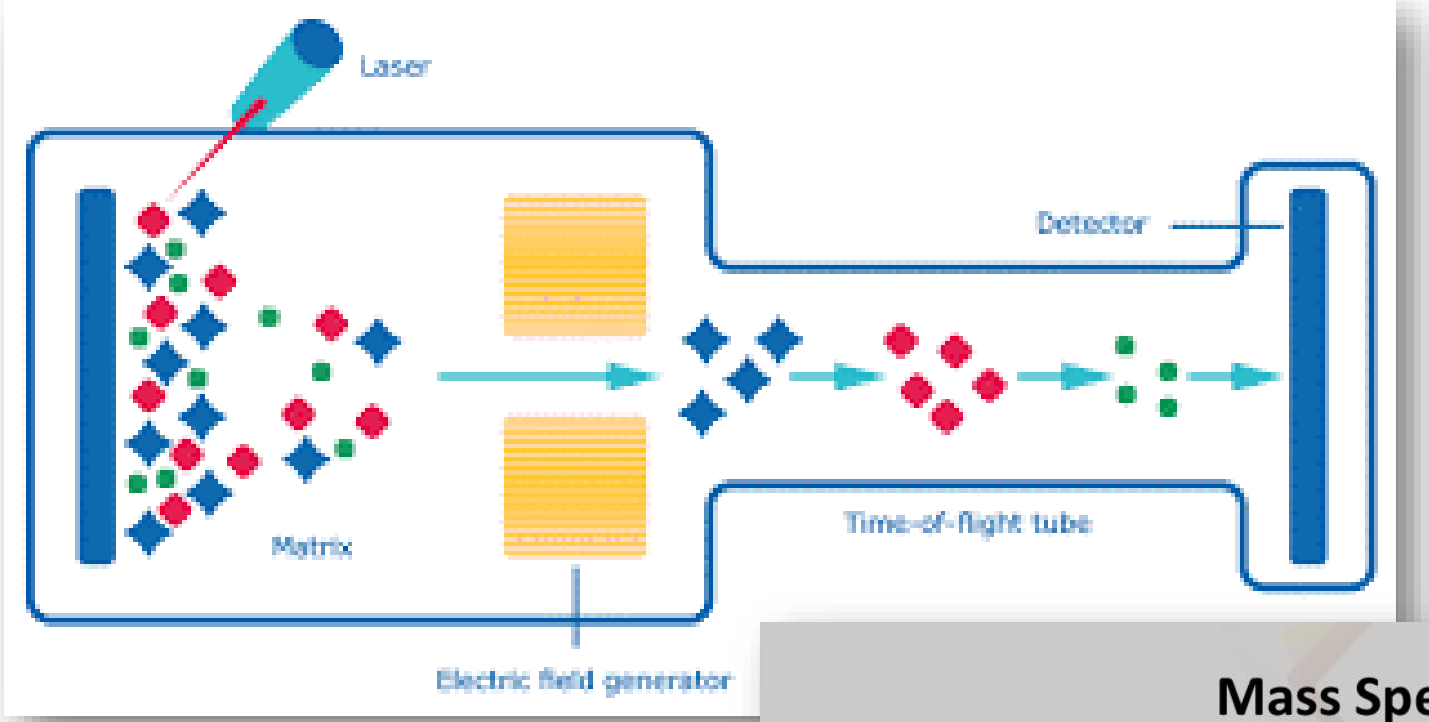


Amino Acid Sequence Determination

- ❑ **Edman degradation**
- ❑ **MS (Mass Spectrometry)**

- MS-based approaches are faster and more convenient than Edman degradation.
- Unlike the Edman approach, MS-based approaches are amenable to high-throughput analyses and therefore generally more useful for proteomics.
- MS-based approaches are more sensitive: the Edman technique, though sensitive, usually requires 1–10 pmol ($1–10 \times 10^{-12}$ mol) of protein sample, whereas MS requires only a few femtomoles (10^{-15} mol) of protein, making MS between 10 and 1000 times more sensitive (see Chapter 1).
- MS-based approaches can provide sequence information from blocked/modified peptides.

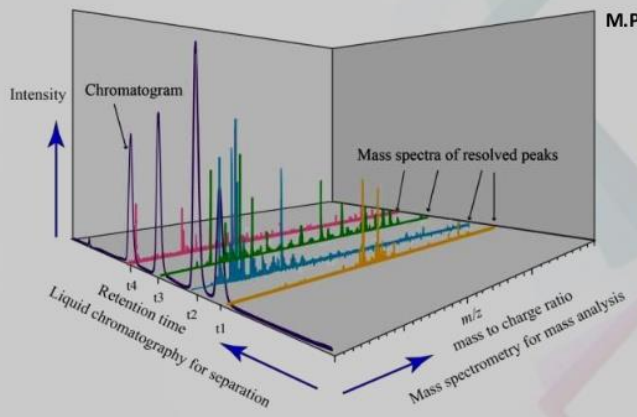

PITC
Polypeptide
pH > 7
PTC-Polypeptide
CF3COOH
Thiazolinone derivative
Shortened polypeptide
Another cycle
PTH–amino acid



Mass Spectrometry

Submitted to :
Rani Mansuri

Submitted by:
Surbhi
M.Pharma 1st sem



Pappin DJ, Hojrup P, Bleasby JA. Curr biol 3 (1993) 327–332.
The techniques and its background will be described in detail in the method chapters.

Peptide mass fingerprinting The easiest and fastest way to identify proteins is shown in figure 2: peptide mass fingerprinting (PMF), which was introduced by four independent groups, including Pappin *et al.* (1993). The gel plug containing the protein of interest is cut out of the gel slab, the protein is digested inside the gel plug with a proteolytic enzyme, mostly trypsin. The cleavage products, the peptides, are eluted from the plug and submitted to mass spectrometry analysis. Mostly MALDI ToF instruments are employed, because they are easier to handle than electrospray systems. The mass spectrum with the accurately measured peptide masses is matched with theoretical peptide spectra in various databases using adequate bioinformatics tools. When no match is found in peptide and protein databases, genomic databases can be searched. The DNA sequence in the open reading frames can be theoretically translated into the amino acid sequence we have to remove this because it is not very practical to search DNA with MALDI data, it is not specific enough. You can do it easily with MS/MS though. Since the cleavage sites of trypsin are known, theoretical tryptic peptide masses can be generated and compared with the experimentally determined masses. If a sufficient number of experimental peptide masses match with the theoretical peptides within a protein, then protein identification with high confidence can be achieved.

This procedure works very well for protein identification. However, the method can be compromised for a number of reasons. In these circumstances, more specific information is needed for unambiguous protein identification, specifically peptide sequence information.

Peptide Mass Fingerprinting (PMF)

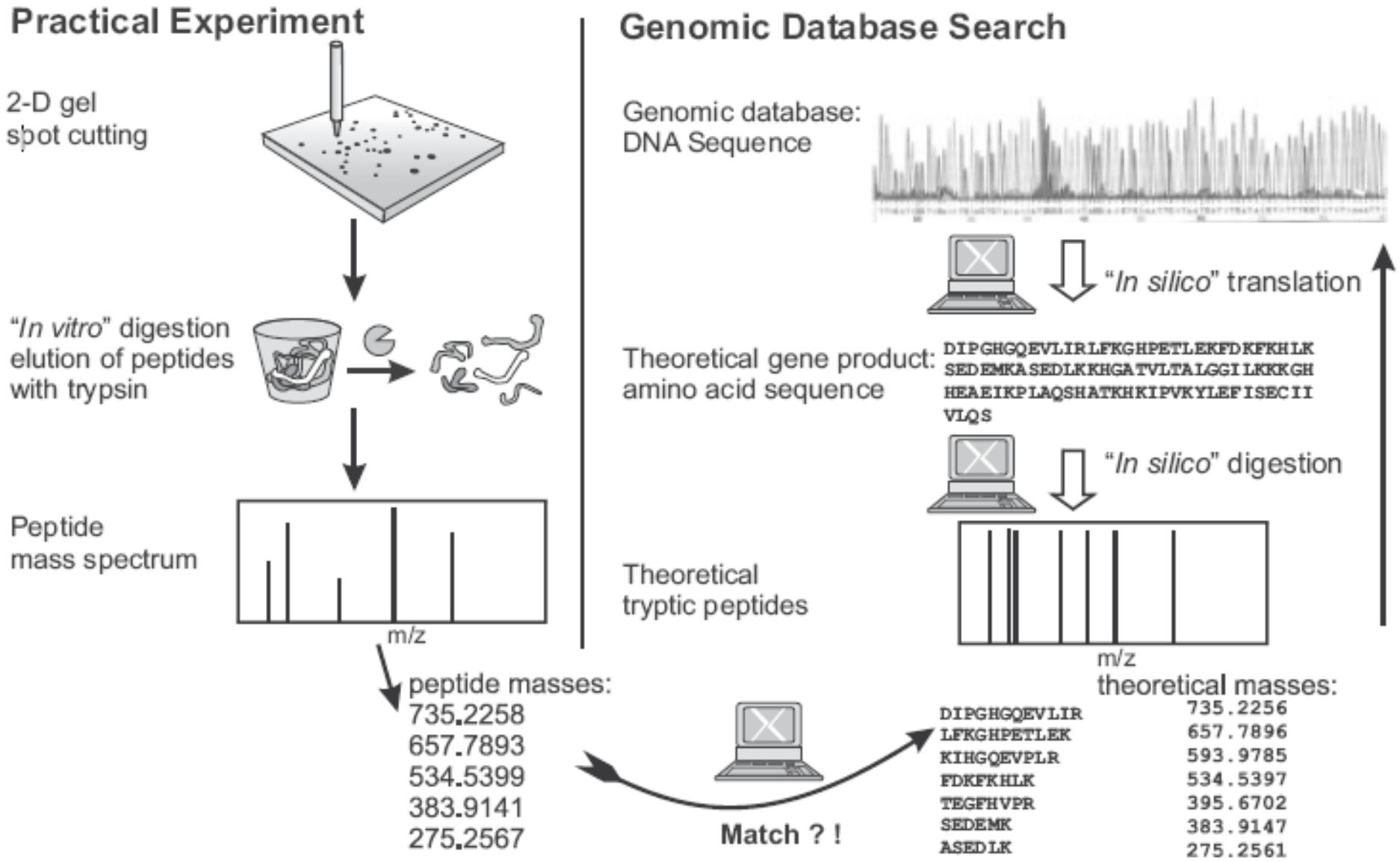
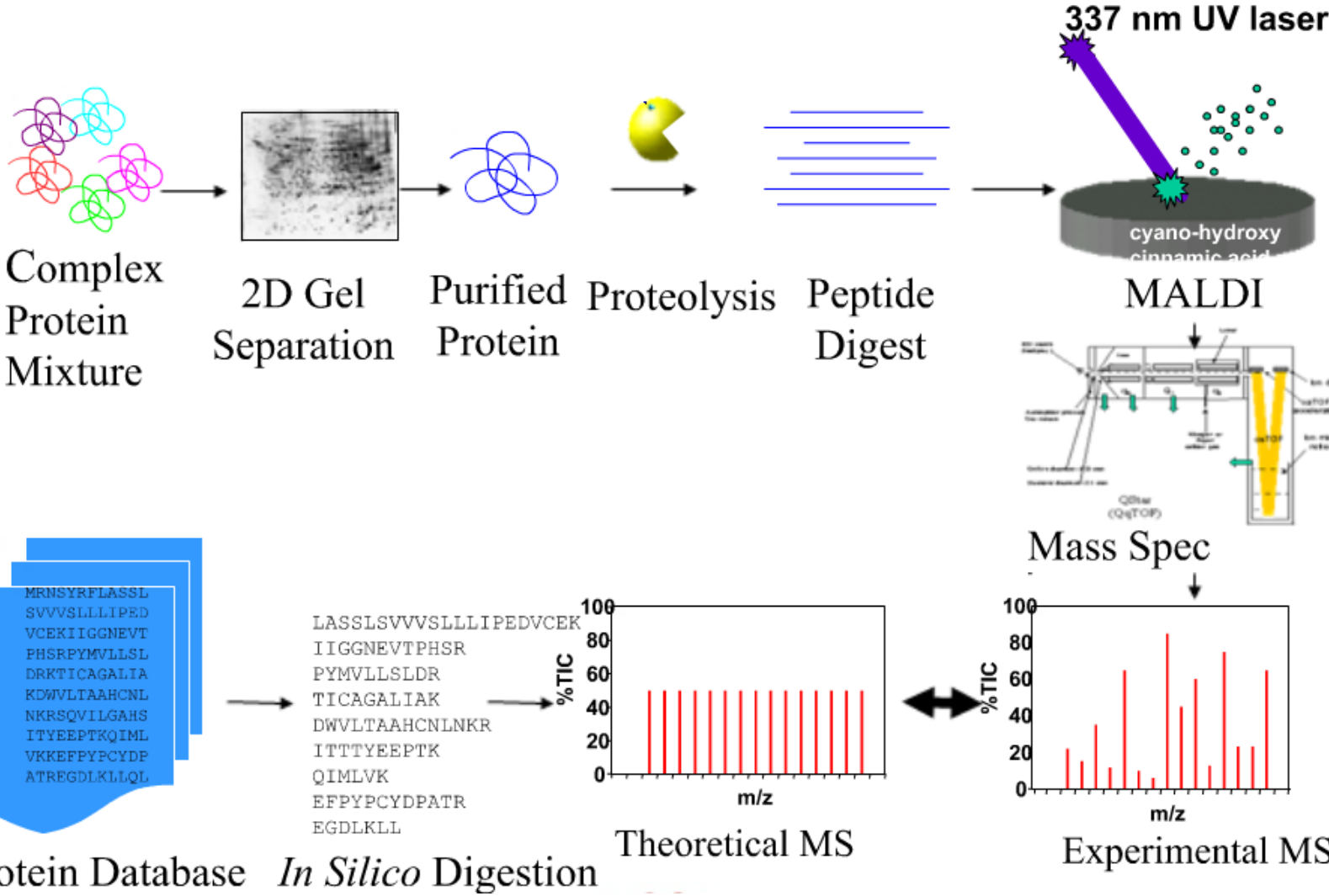


Fig. 2: Protein identification with peptide mass fingerprinting. The peptide masses of the digested protein are matched with a list of theoretical masses of peptides, which are mathematically derived from the open reading frames of the genome database of a certain organism.

Review: Peptide Mass Fingerprinting



www.matrixscience.com

MATRIX SCIENCE

Search this site

Home | Mascot database search | Products | Technical support | Training | News | Blog | Newsletter | Contact

Access Mascot Server | Database search help

Welcome to the home of Mascot software, the benchmark for identification, characterisation and quantitation of proteins using mass spectrometry data. Here, you can learn more about the tools developed by Matrix Science to get the best out of your data, whatever your chosen instrument.

Blog

October 19, 2016

Although retention time is not part of the Mascot scoring algorithm, it can be used by Percolator to improve the [...]

Subscribe

Mascot Server

Mascot Server is live on this website for both Peptide Mass Fingerprint and MS/MS database searches. A selection of popular sequence databases are online, including SwissProt, NCBInr, and

- **FREE search**
- **Help topics**
- **Training course**
- **Technical support**

Mascot Distiller

Mascot Distiller offers a single, intuitive interface to native (binary) data files from Agilent, AB Sciex, Bruker, Shimadzu, Thermo and Waters. Raw data can be processed into high quality,

- **Download**
- **Try a 30 day evaluation**
- **Technical support**

www.matrixscience.com/search_form_select.html

MATRIX SCIENCE

Search this site

Home | Mascot database search | Products | Technical support | Training | News | Blog | Newsletter | Contact

Access Mascot Server | Database search help

Mascot database search > Access Mascot Server

Access Mascot Server

You are welcome to submit searches to this free Mascot Server. Searches of MS/MS data are limited to 1200 spectra and some functions, such as no enzyme searches, are unavailable. Automated searching of batches of files is not permitted. If you want to automate search submission, perform large searches, search additional sequence databases, or customise the modifications, quantitation methods, etc., you'll need to license your own, in-house copy of Mascot Server.

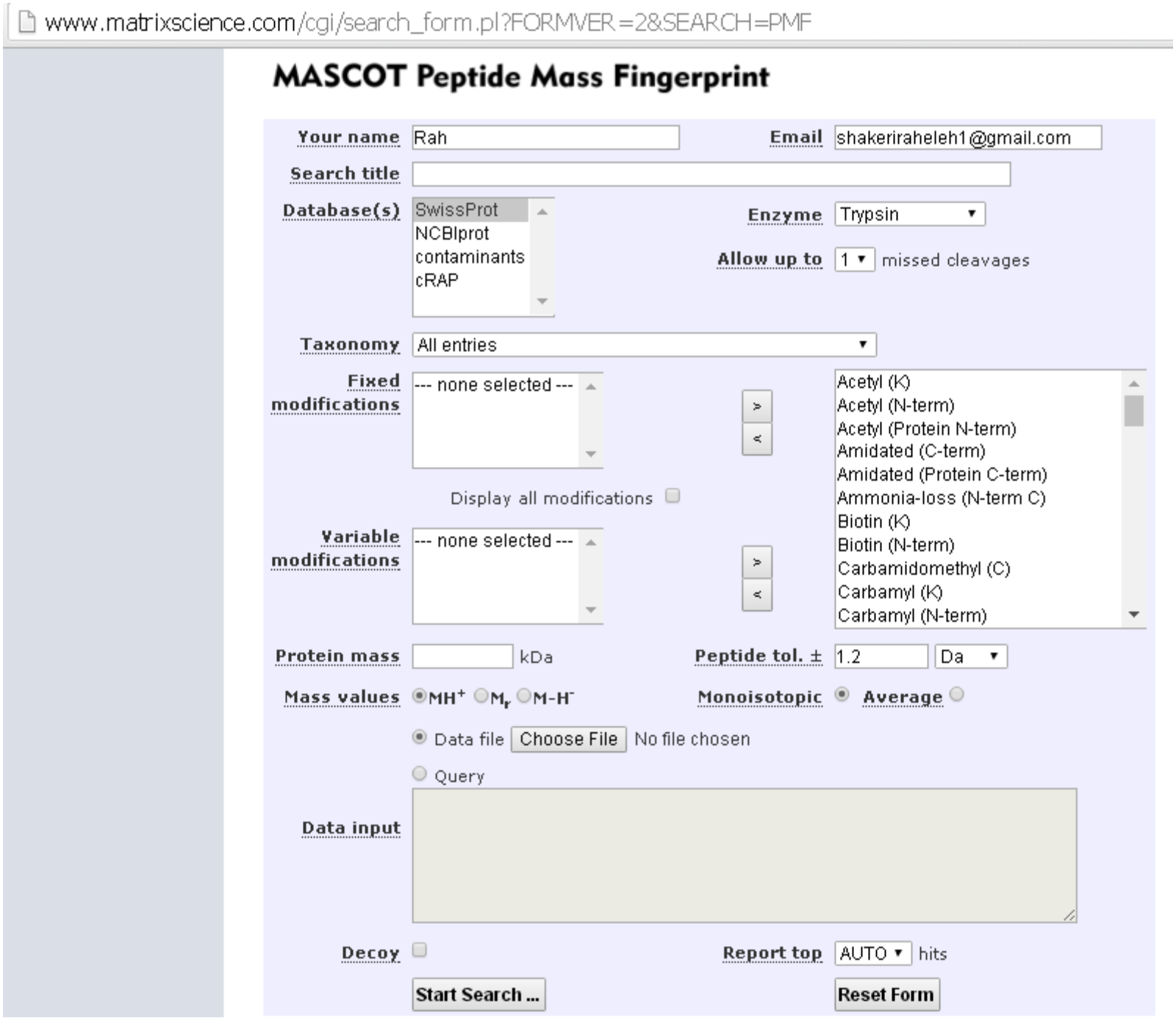
Peptide Mass Fingerprint

The experimental data are a list of peptide mass values from the digestion of a protein by a specific enzyme such as trypsin.

Perform search | Example of results report | Tutorial

More info

- Mascot overview
- Search parameter reference
- Data file format
- Results report overview

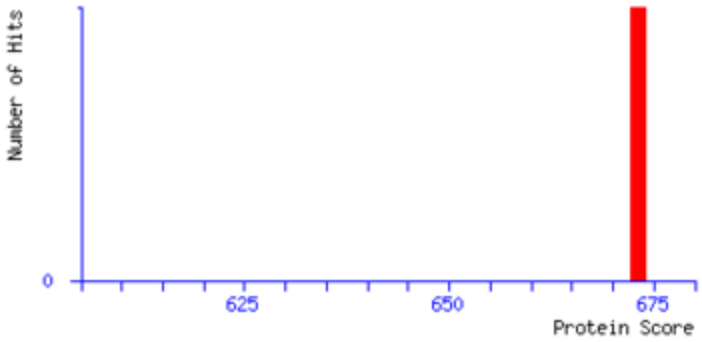

www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF
MASCOT Peptide Mass Fingerprint
Your name
Rah
Email
shakeriraheleh1@gmail.com
Search title
Database(s)
SwissProt
NCBIprot
contaminants
cRAP
Enzyme
Trypsin
Allow up to
1
missed cleavages
Taxonomy
All entries
Fixed modifications
--- none selected ---
Display all modifications
Variable modifications
--- none selected ---
Acetyl (K)
Acetyl (N-term)
Acetyl (Protein N-term)
Amidated (C-term)
Amidated (Protein C-term)
Ammonia-loss (N-term C)
Biotin (K)
Biotin (N-term)
Carbamidomethyl (C)
Carbamyl (K)
Carbamyl (N-term)
Protein mass
kDa
Peptide tol. ±
1.2
Da
Mass values
MH+
Mr
M-H-
Monoisotopic
Average
Data file
Choose File
No file chosen
Query
Data input
Decoy
Report top
AUTO
hits
Start Search ...
Reset Form

MATRIX SCIENCE Mascot Search Results

User : **Rah**
Email : **shakeriraheleh1@gmail.com**
Search title : **Apaf-1**
Database : **SwissProt 2016_09 (552259 sequences; 197423140 residues)**
Timestamp : **27 Oct 2016 at 17:34:19 GMT**
Top Score : **673 for APAF_HUMAN, Apoptotic protease-activating factor 1 OS=Homo sapiens GN=APAF1 PE=1 SV=2**

Mascot Score Histogram

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 70 are significant (p<0.05).



Concise Protein Summary Report

Format As: Concise Protein Summary ▾ Help
Significance threshold p< 0.05 Max. number of hits AUTO
Preferred taxonomy: All entries ▾

Re-Search All | Search Unmatched

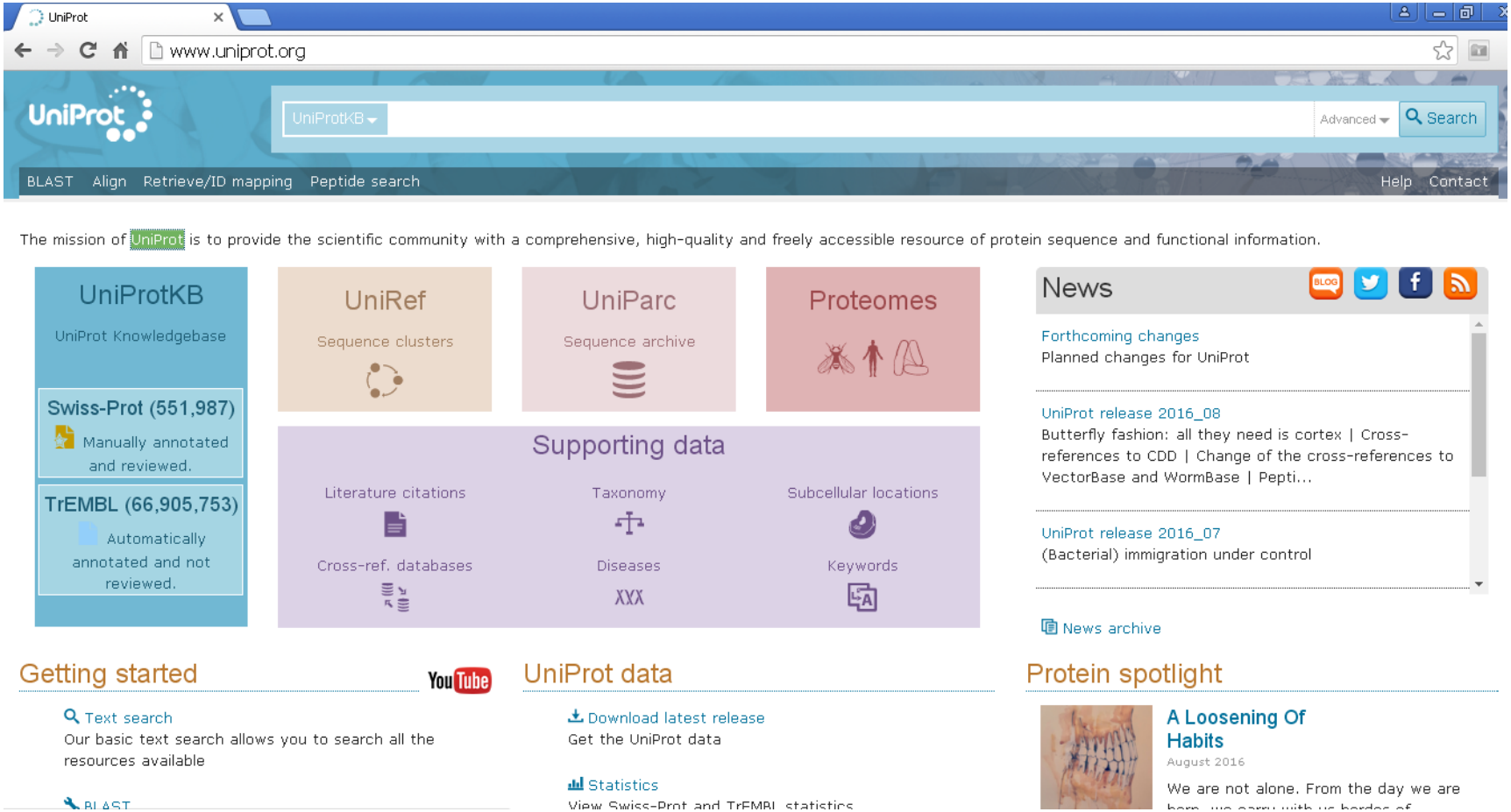
1. APAF_HUMAN **Mass:** 141749 **Score:** 673 **Expect:** 2.8e-62 **Matches:** 88
Apoptotic protease-activating factor 1 OS=Homo sapiens GN=APAF1 PE=1 SV=2
LPTD_CHRSD **Mass:** 93592 **Score:** 54 **Expect:** 2.2 **Matches:** 29

The last point in particular has always been a complicating factor when applying the Edman approach to eukaryotic-derived proteins. Up to 80% of such proteins display chemically altered N-terminal amino acid residues, which do not react with the Edman PITC reagent (Box 2.1). The most common N-terminal chemical alteration observed is acetylation (see section 2.9.4), but blocking may also be the result of glycosylation and formylation for example.

Today, however, the vast majority of protein sequences are obtained/predicted indirectly via nucleotide sequence data generated from genome sequencing projects (Chapter 1), which now means that amino acid sequence data for several tens of millions of different proteins are available and may be accessed and interrogated through databases such as the Uniprot database (www.uniprot.org; Box 2.2).

Despite the central importance of the genomic approach, direct sequencing methods remain important/essential for a number of applications. For example, direct sequencing (full-length or at least partial sequencing of the first 10–20 amino acids at the N-terminus of a protein) can be used to:

- design polymerase chain reaction (PCR) primers to assist in the ultimate cloning of the gene coding for the protein if the protein has been purified directly from, for example, a source for which no genome sequence data is available;
- serve as a quality control tool to directly verify the identity/sequence of protein products such as biopharmaceuticals.



سایت UniProt با همکاری سه موسسه شامل موسسه بیوانفورماتیک اروپا(EMBL-EBI)، موسسه بیوانفورماتیک سوئیس (SIB) و منبع اطلاعات پروتئینی (PIR) شکل گرفته است. با همکاری سه موسسه بیش از ۱۰۰ نفر با عناوین مسئول پایگاه داده، توسعه نرم افزار و پشتیبانی مشغول به کار هستند.

هرکدام از موسسات مذکور وظایف مختلفی را بر عهده دارد. بدین صورت که EMBL-EBI و SIB باهم به تولید محتوای Swiss-Prot و TrEMBL (کتابخانه و مرکز داده توالی های نوکلئوتیدی ترجمه شده) می پردازند. همچنین موسسه PIR مسئول تهیه بانک اطلاعاتی توالی پروتئین (PIR-PSD) می باشد. مجموع داده های تهیه شده توسط این موسسات که مربوط به توالی های مختلف پروتئینی می باشند، بخش اعظمی را پوشش می دهند.

TrEMBL با همکاری Swiss-Prot با سرعت بالایی به تولید محتوا می پردازند. مجموعه PIR نیز در همین حال مجموعه بانک اطلاعاتی توالی پروتئین ها را تهیه و نگهداری می کند. در سال ۲۰۰۲ سه موسسه مذکور منابع خو را با هم ادغام کرده و UniProt را شکل دادند.

UniProt (Universal Protein Resource) is a comprehensive web-based resource (www.uniprot.org) housing information on proteins, particularly protein sequence and function. It is a collaboration between three bioinformatic-based institutes: the European Bioinformatics Institute, the Swiss Institute of Bioinformatics, and the Protein Information Resource institute.

Virtually all the protein sequences provided by UniProtKB are derived from the translation of coding sequences submitted to public nucleic acid databases (EMBL, GenBank and DDBJ

Bioinformatic analysis of sequence data

a major goal, and indeed achievement, of bioinformatics has been the development of computer programs/software tools which can interrogate and analyse raw protein sequence information in order to generate additional information.

Various and often multiple different bioinformatic programs/tools are available that interrogate protein sequence information/databases in order to:

- identify proteins containing similar amino acid sequences (i.e. run similarity searches) and assess how closely related two (or more) proteins are, or if there is a high probability that they undertake similar functions (see next section);
- calculate a theoretical molecular mass, isoelectric point (see Chapter 4) or other physicochemical property of a protein;
- predict elements of a protein's higher-order structure (secondary and tertiary structure, or for example protein domains, as discussed in section 2.2.2);
- predict if a protein is likely to undergo PTMs (see section 2.9), and at what point(s) along the protein backbone this is likely to occur;
- predict where in the cell the protein is likely to function (or if it is likely exported from the cell).

Sequence similarity and sequence alignment analysis

Table 2.2 Top matches obtained from a BLAST search using the human erythropoietin (EPO) amino acid sequence as a query sequence against the 42 million sequence entries present in the UniProtKB database (Box 2.2). A total of 121 hits were obtained, the top 26 of which are presented here. Unsurprisingly, the highest matches were to the human EPO sequence entries already present in the database. Many of the additional hits are EPO sequences from other species. An outline of how similarity is graded is presented in the main text.

Accession	Entry name	0Query hit193	0Match hit (sqrt scale) 24531	Name (organism)
Query	2013072970Q0V94AU2			
G9JKG7	G9JKG7_HUMAN			Erythropoietin (*Homo sapiens*)
P01588	EPO_HUMAN			Erythropoietin (*Homo sapiens*)
H2QV42	H2QV42_PANTR			Uncharacterized protein (*Pan troglodytes*)
G3RS27	G3RS27_GORGO			Uncharacterized protein (*Gorilla gorilla gorilla*)
B7ZKK5	B7ZKK5_HUMAN			EPO protein (*Homo sapiens*)
G1RMP4	G1RMP4_NOMLE			Uncharacterized protein (*Nomascus leucogenys*)
G3RPR5	G3RPR5_GORGO			Uncharacterized protein (*Gorilla gorilla gorilla*)
P07865	EPO_MACFA			Erythropoietin (*Macaca fascicularis*)
Q28513	EPO_MACMU			Erythropoietin (*Macaca mulatta*)
G7P0D4	G7P0D4_MACFA			Putative uncharacterized protein (*Macaca fascicularis*)
F6WN92	F6WN92_MACMU			Erythropoietin (*Macaca mulatta*)
F7DTH0	F7DTH0_CALJA			Uncharacterized protein (*Callithrix jacchus*)
Q867B1	EPO_HORSE			Erythropoietin (*Equus caballus*)
17AKF2	17AKF2_FELCA			Erythropoietin (*Felis catus*)
13MLF9	13MLF9_SPETR			Uncharacterized protein (*Spermophilus tridecemlineatus*)
F7DQY8	F7DQY8_HORSE			Erythropoietin (*Equus caballus*)
P33708	EPO_FELCA			Erythropoietin (*Felis catus*)
D2HX05	D2HX05_AILME			Putative uncharacterized protein (*Ailuropoda melanoleuca*)
G1M830	G1M830_AILME			Uncharacterized protein (*Ailuropoda melanoleuca*)
G3UDT5	G3UDT5_LOXAF			Uncharacterized protein (*Loxodonta africana*)
K4Q170	K4Q170_CANFA			Erythropoietin (*Canis familiaris*)
M3YWD4	M3YWD4_MUSPF			Uncharacterized protein (*Mustela putorius furo*)
H0Y1U0	H0Y1U0_OTOGA			Uncharacterized protein (*Otolemur garnettii*)
L5K6F9	L5K6F9_PTEAL			Erythropoietin (*Pteropus alecto*)
F1PPB9	F1PPB9_CANFA			Erythropoietin (*Canis familiaris*)
J9NYY7	J9NYY7_CANFA			Erythropoietin (*Canis familiaris*)

BLAST
(Basic Local Alignment Search Tool): UniProt or NCBI

Alignment
Pairwaise alignment
Multiple alignment

Homology
Similarity
Identity

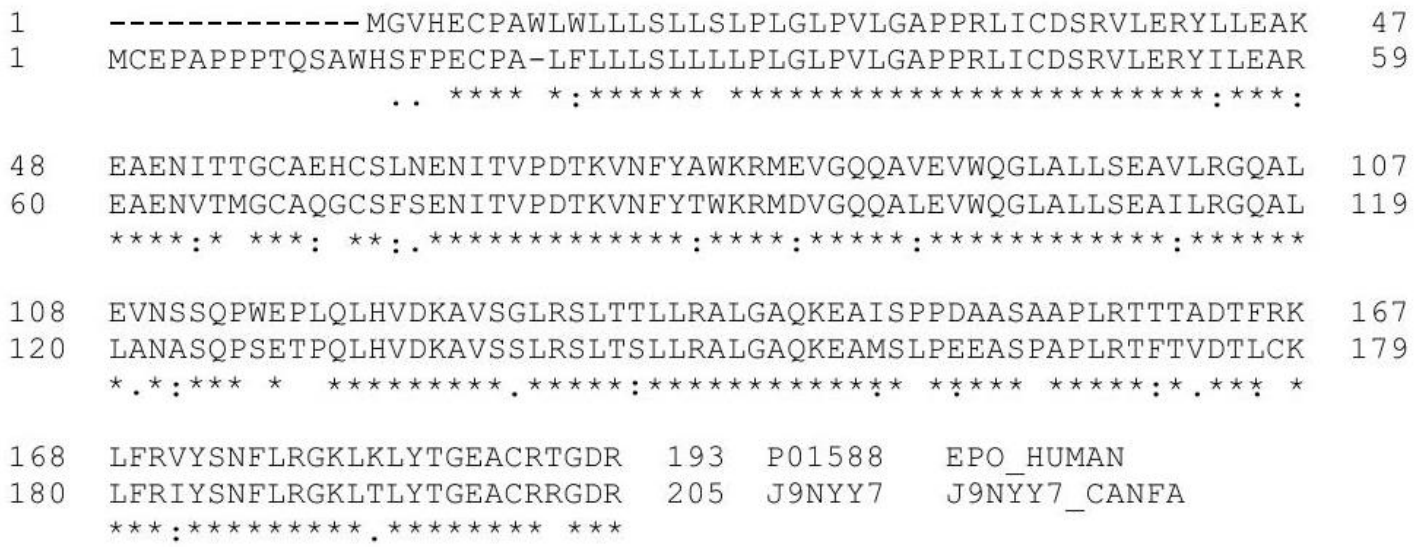


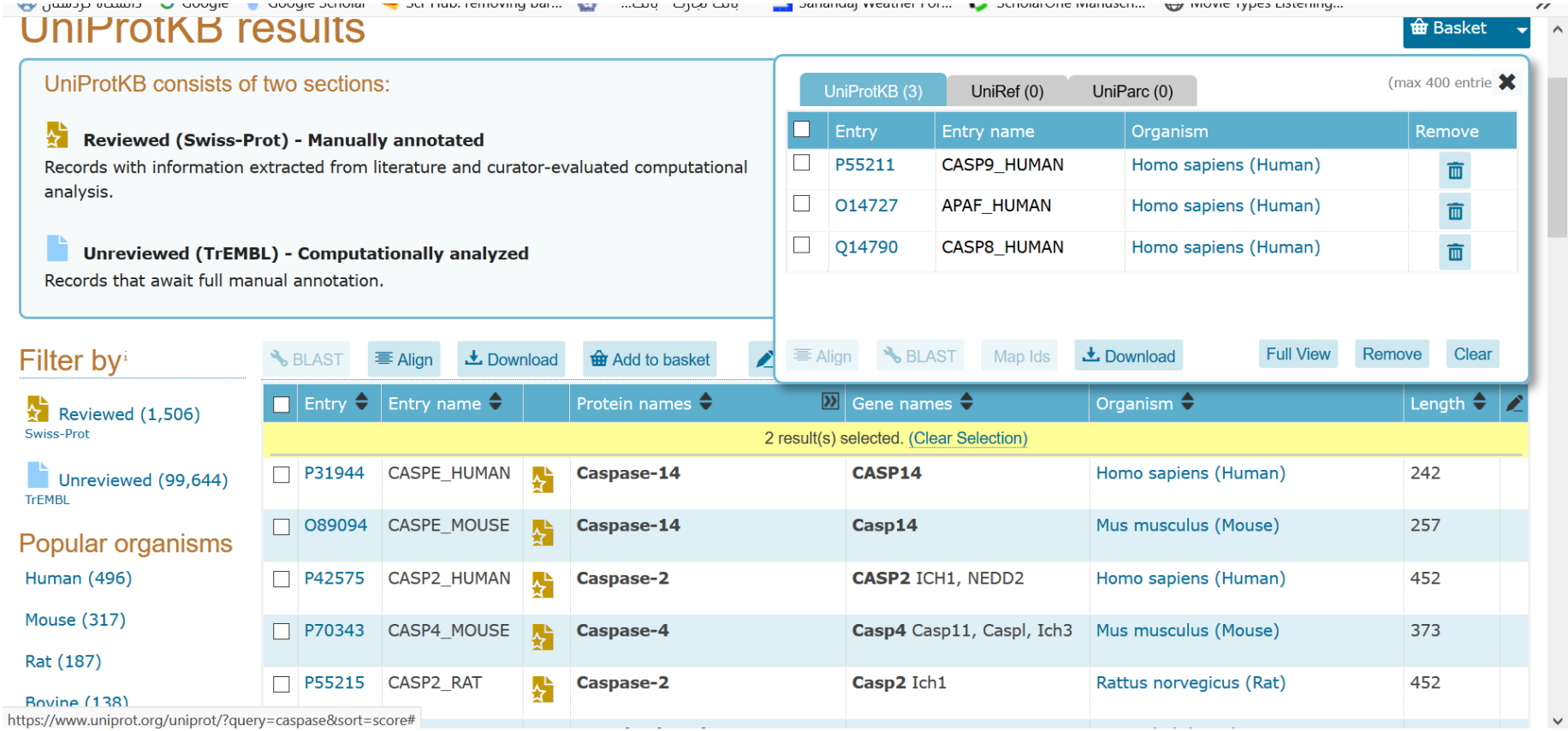
Figure 2.6 A pairwise sequence alignment between the amino acid sequence of human erythropoietin (EPO, top line of each twin sequence) and canine EPO (bottom line of each twin sequence) (a). The sequence alignment was undertaken via the UniProt website. Asterisks are automatically placed underneath sequence positions housing identical amino acid residues while double or single dots (i.e. a colon or a period) appear underneath residue positions which display strongly or weakly similar properties, respectively. Thus, human and canine EPOs contain identical residues at 155 positions (i.e. they display approximately 75% identity) and similar residues at a further 24 positions. The software also facilitates the generation of additional information such as the positioning of amino acid residues with particular properties.

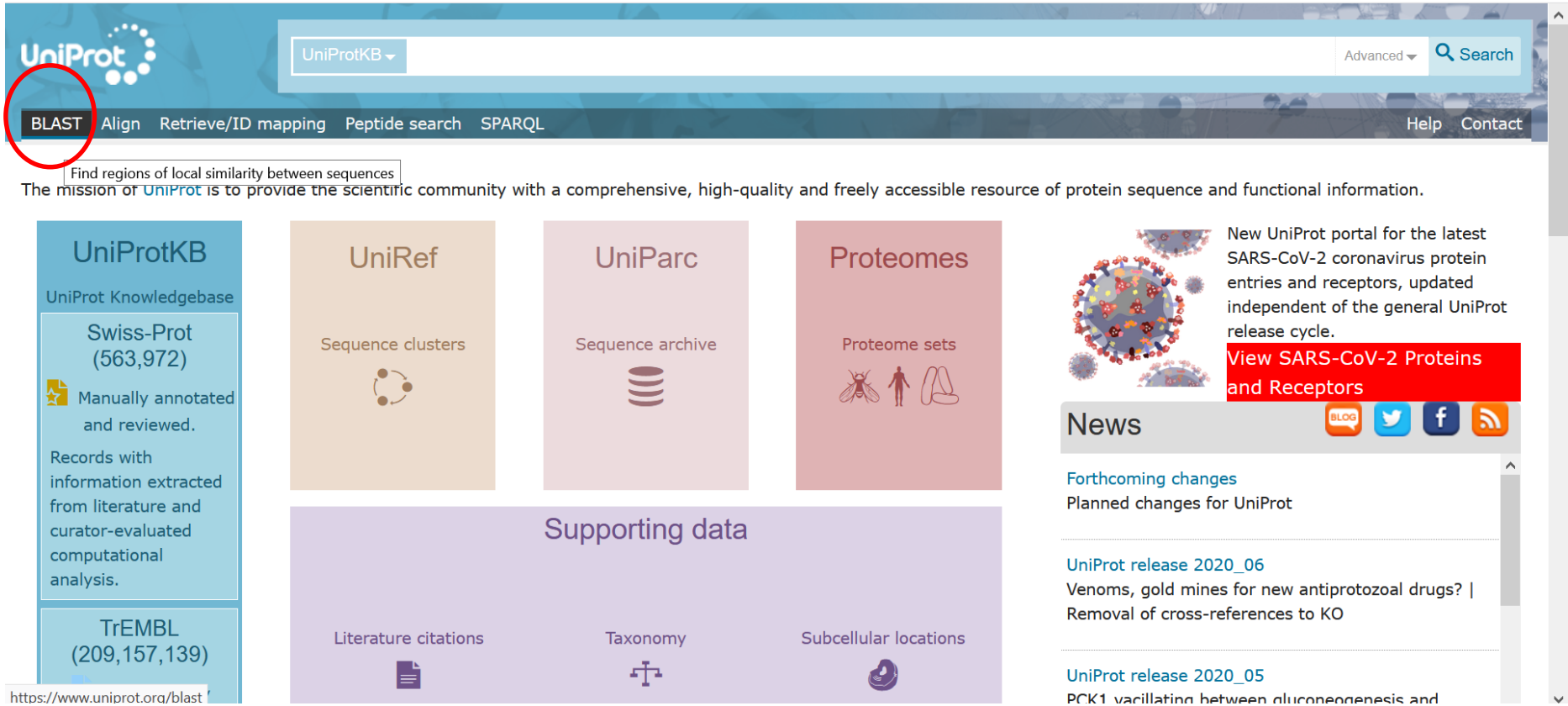
57 similar amino acides
37 full match
20 +

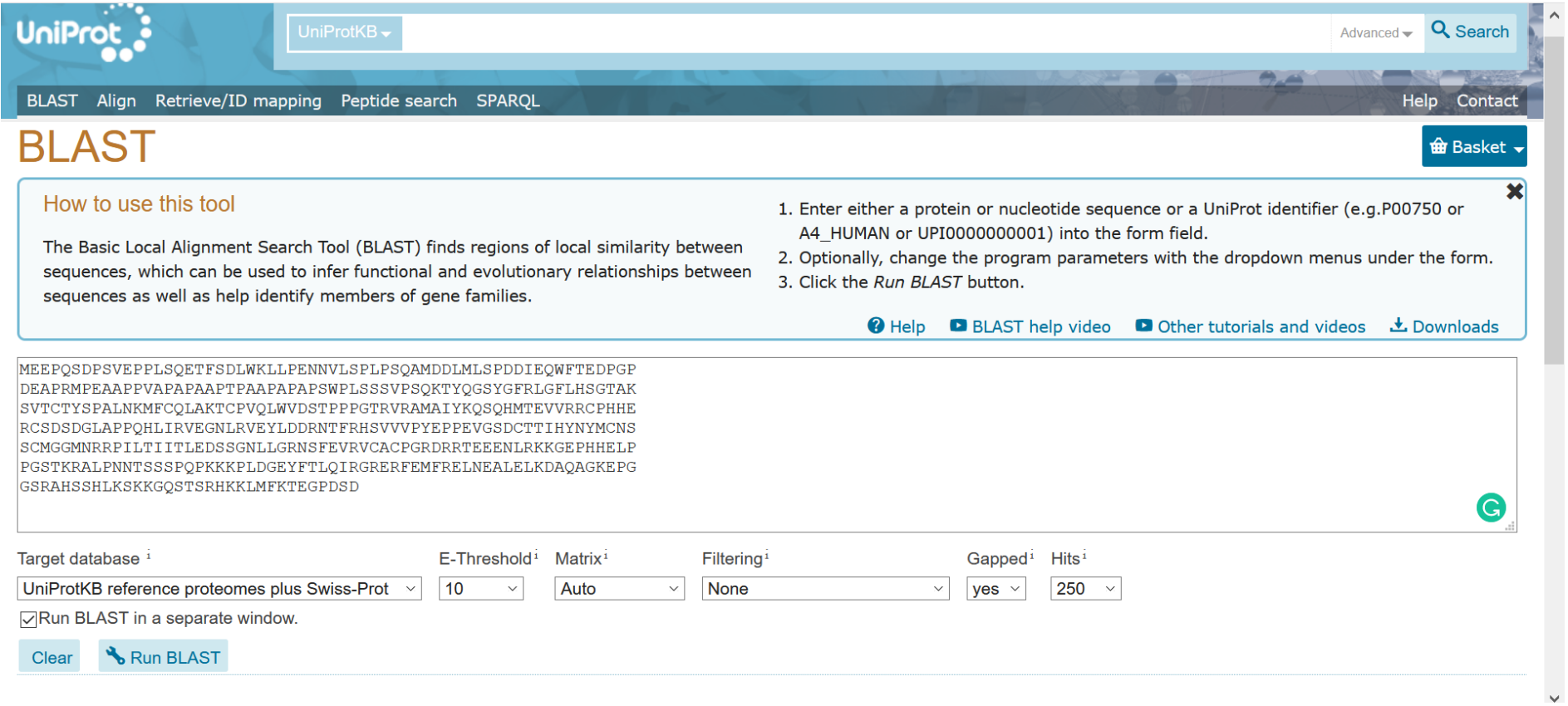
```
 Score = 43.9 bits (102),  Expect = 1e-09, Method: Composition-based stats.
 Identities = 37/145 (25%), Positives = 57/145 (39%), Gaps = 2/145 (1%)

Query  4    LTPEEKSAVTALWGKVNVD--EVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV  61
            L+  E   V  +WGKV  D    G E L RL   +P T   F+ F  L + D +  +  +
Sbjct  3    LSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDL  62

Query  62   KAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGK  121
            K HG  VL A    L    + +     L++ H  K  +  +    +   ++ VL
Sbjct  63   KKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPG  122

Query  122  EFTPPVQAAYQKVVAGVANALAHKY  146
            +F    Q A  K +       +A  Y
Sbjct  123  DFGADAQGAMNKALELFRKDMASNY  147
```

% identity = (طول ناحیه انطباق /تعداد ریشه های یکسان) ×۱۰۰

% Similarity = (طول ناحیه انطباق /تعداد ریشه های مشابه + تعداد ریشه های یکسان) ×۱۰۰


UniProtKB results
Basket
UniProtKB consists of two sections:
Reviewed (Swiss-Prot) - Manually annotated
Records with information extracted from literature and curator-evaluated computational analysis.
Unreviewed (TrEMBL) - Computationally analyzed
Records that await full manual annotation.
UniProtKB (3)
UniRef (0)
UniParc (0)
(max 400 entrie
Entry
Entry name
Organism
Remove
P55211 CASP9_HUMAN Homo sapiens (Human)
O14727 APAF_HUMAN Homo sapiens (Human)
Q14790 CASP8_HUMAN Homo sapiens (Human)
Align
BLAST
Map Ids
Download
Full View
Remove
Clear
Filter by
BLAST
Align
Download
Add to basket
Reviewed (1,506)
Swiss-Prot
Unreviewed (99,644)
TrEMBL
Popular organisms
Human (496)
Mouse (317)
Rat (187)
Bovine (138)
Entry
Entry name
Protein names
Gene names
Organism
Length
2 result(s) selected. (Clear Selection)
P31944 CASPE_HUMAN Caspase-14 CASP14 Homo sapiens (Human) 242
O89094 CASPE_MOUSE Caspase-14 Casp14 Mus musculus (Mouse) 257
P42575 CASP2_HUMAN Caspase-2 CASP2 ICH1, NEDD2 Homo sapiens (Human) 452
P70343 CASP4_MOUSE Caspase-4 Casp4 Casp11, Caspl, Ich3 Mus musculus (Mouse) 373
P55215 CASP2_RAT Caspase-2 Casp2 Ich1 Rattus norvegicus (Rat) 452
https://www.uniprot.org/uniprot/?query=caspase&sort=score#


UniProt
UniProtKB
Advanced
Search
BLAST
Align
Retrieve/ID mapping
Peptide search
SPARQL
Help
Contact
Find regions of local similarity between sequences
The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.
UniProtKB
UniProt Knowledgebase
Swiss-Prot (563,972)
Manually annotated and reviewed.
Records with information extracted from literature and curator-evaluated computational analysis.
TrEMBL (209,157,139)
UniRef
Sequence clusters
UniParc
Sequence archive
Proteomes
Proteome sets
Supporting data
Literature citations
Taxonomy
Subcellular locations
New UniProt portal for the latest SARS-CoV-2 coronavirus protein entries and receptors, updated independent of the general UniProt release cycle.
View SARS-CoV-2 Proteins and Receptors
News
Forthcoming changes
Planned changes for UniProt
UniProt release 2020_06
Venoms, gold mines for new antiprotozoal drugs? | Removal of cross-references to KO
UniProt release 2020_05
https://www.uniprot.org/blast


UniProt
UniProtKB
Advanced
Search
BLAST
Align
Retrieve/ID mapping
Peptide search
SPARQL
Help
Contact
BLAST
Basket
How to use this tool
The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences, which can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.
1. Enter either a protein or nucleotide sequence or a UniProt identifier (e.g.P00750 or A4_HUMAN or UPI0000000001) into the form field.
2. Optionally, change the program parameters with the dropdown menus under the form.
3. Click the Run BLAST button.
Help
BLAST help video
Other tutorials and videos
Downloads
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP
DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK
SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS
SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP
PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG
GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD
Target database
E-Threshold
Matrix
Filtering
Gapped
Hits
UniProtKB reference proteomes plus Swiss-Prot
10
Auto
None
yes
250
Run BLAST in a separate window.
Clear
Run BLAST

UniProtKB | Advanced | Search

BLAST | Align | Retrieve/ID mapping | Peptide search | SPARQL | Help | Contact

BLAST

About BLAST | Basket

Filter by

Reviewed (28)
Swiss-Prot

Unreviewed (214)
TrEMBL

With 3D structure (4)

Popular organisms

Human (9)

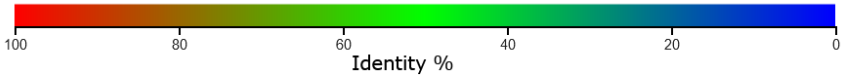
Mouse (4)

Zebrafish (4)

Rat (3)

Bovine (1)

All (241)



Edit and resubmit | Order by: Score

Overview

Show all 241

Entry	Protein names	Match hit	Identity
P04637	**Cellular tumor antigen p53** (Homo sapiens)		100.0%
A0A2R9A5P4	**Cellular tumor antigen p53** (Pan paniscus)		100.0%
H2QC53	**Cellular tumor antigen p53** (Pan troglodytes)		100.0%
G3R2U9	**Cellular tumor antigen p53** (Gorilla gorilla gorilla)		99.7%

- [ ] Helix
- [ ] Metal binding
- [ ] Modified residue
- [ ] Motif
- [ ] Mutagenesis
- [ ] Natural variant
- [ ] Region
- [ ] Site
- [ ] Turn

Amino acid properties

- [ ] Similarity
- [ ] Hydrophobic
- [ ] Negative
- [ ] Positive
- [ ] Aliphatic
- [ ] Tiny
- [ ] Aromatic
- [ ] Charged
- [ ] Small
- [ ] Polar
- [ ] Big
- [ ] Serine Threonine

Query Length: 393
Match Length: 393

```
Query                  1     MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP       60
                             MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP
P04637 P53_HUMAN       1     MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP       60

Query                 61     DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK      120
                             DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK
P04637 P53_HUMAN       61     DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK      120

Query                121     SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE      180
                             SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE
P04637 P53_HUMAN       121     SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE      180

Query                181     RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS      240
                             RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS
P04637 P53_HUMAN       181     RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS      240

Query                241     SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP      300
                             SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP
P04637 P53_HUMAN       241     SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP      300

Query                301     PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG      360
                             PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG
P04637 P53_HUMAN       301     PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG      360

Query                361     GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD                                  393
                             GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD
P04637 P53_HUMAN       361     GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD                                  393
```

- [ ] Helix
- [ ] Metal binding
- [ ] Modified residue
- [ ] Motif
- [ ] Mutagenesis
- [ ] Natural variant
- [ ] Region
- [ ] Site
- [ ] Turn

Amino acid properties

- [ ] Similarity
- [x] Hydrophobic
- [ ] Negative
- [ ] Positive
- [ ] Aliphatic
- [ ] Tiny
- [ ] Aromatic
- [ ] Charged
- [ ] Small
- [ ] Polar
- [ ] Big
- [ ] Serine Threonine

Query Length: 393
Match Length: 393

```
Query                1    MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP   60
                          MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP
P04637 P53_HUMAN     1    MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP   60

Query               61    DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK  120
                          DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK
P04637 P53_HUMAN     61    DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK  120

Query               121    SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE  180
                          SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE
P04637 P53_HUMAN     121    SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE  180

Query               181    RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS  240
                          RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS
P04637 P53_HUMAN     181    RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS  240

Query               241    SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP  300
                          SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP
P04637 P53_HUMAN     241    SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP  300

Query               301    PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG  360
                          PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG
P04637 P53_HUMAN     301    PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG  360

Query               361    GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD                            393
                          GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD
P04637 P53_HUMAN     361    GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD                            393
```

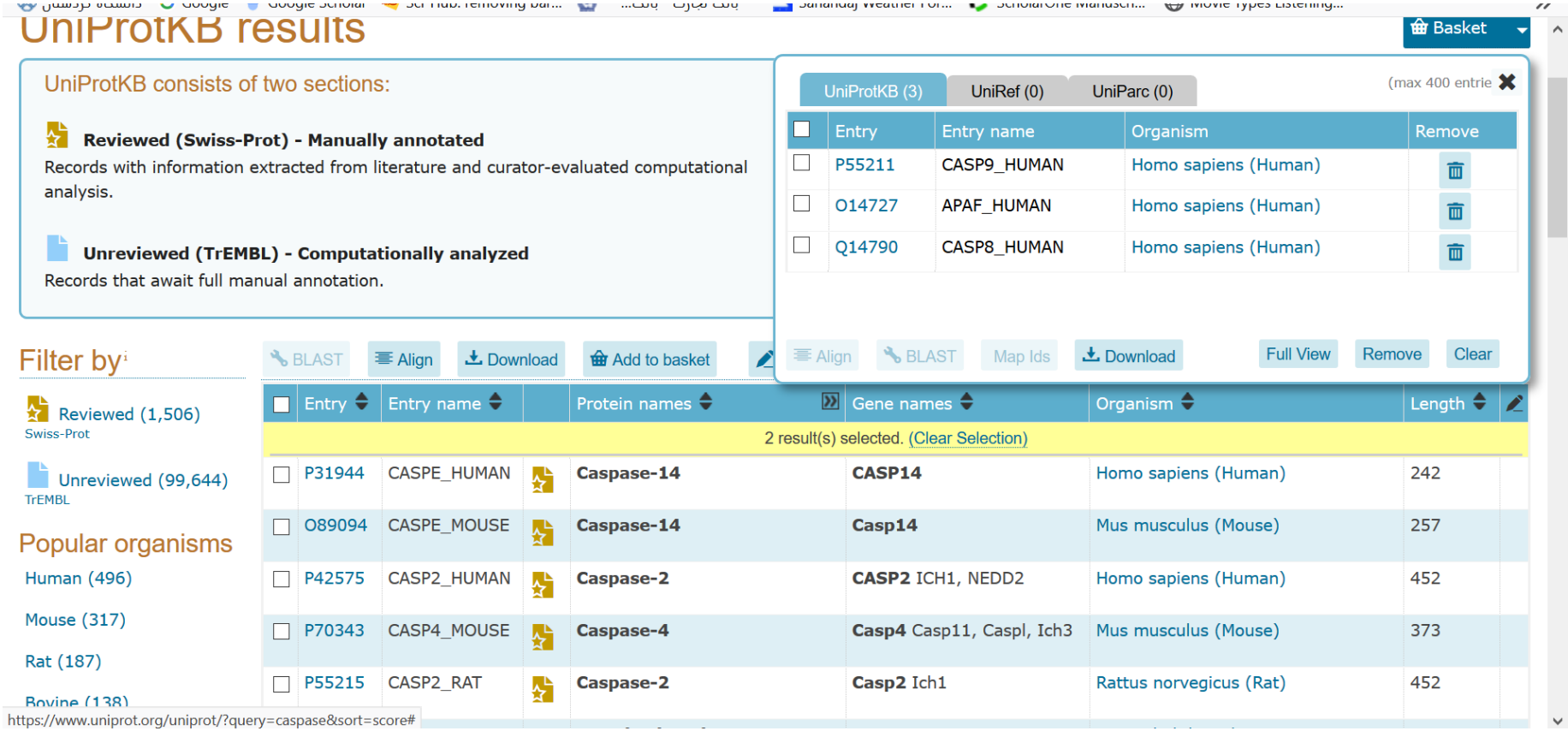
- [ ] Motif
- [ ] Region
- [ ] Site

Amino acid properties

- [x] Similarity
- [ ] Hydrophobic
- [ ] Negative
- [ ] Positive
- [ ] Aliphatic
- [ ] Tiny
- [ ] Aromatic
- [ ] Charged
- [ ] Small
- [ ] Polar
- [ ] Big
- [ ] Serine Threonine

Query Length: 393
Match Length: 393

```
Query                1    MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP   60
                          MEEPQSDPS+EPPLSQETFSDLWKLLPENNVLSPLPSQA+DDLMLSPDD+ QW TEDPGP
P61260 P53_MACFU     1    MEEPQSDPSIEPPLSQETFSDLWKLLPENNVLSPLPSQAVDDLMLSPDDLAQWLTEDPGP   60

Query               61    DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK  120
                          DEAPRM EAAPP+AP PAAPTPAAPAPAPSWPLSSSVPSQKTY GSYGFRLGFLHSGTAK
P61260 P53_MACFU     61    DEAPRMSEAAPPMAPTPAAPTPAAPAPAPSWPLSSSVPSQKTYHGSYGFRLGFLHSGTAK  120

Query              121    SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE  180
                          SVTCTYSP LNKMFCQLAKTCPVQLWVDSTPPPG+RVRAMAIYKQSQHMTEVVRRCPHHE
P61260 P53_MACFU    121    SVTCTYSPDLNKMFCQLAKTCPVQLWVDSTPPPGSRVRAMAIYKQSQHMTEVVRRCPHHE  180

Query              181    RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS  240
                          RCSDSDGLAPPQHLIRVEGNLRVEY DDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS
P61260 P53_MACFU    181    RCSDSDGLAPPQHLIRVEGNLRVEYSDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS  240

Query              241    SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP  300
                          SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEEN RKKGEP H+LP
P61260 P53_MACFU    241    SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENFRKKGEPCHQLP  300

Query              301    PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG  360
                          PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEP
P61260 P53_MACFU    301    PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPA  360

Query              361    GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD                            393
                          GSRAHSSHLKSKKGQSTSRHKK MFKTEGPDSD
P61260 P53_MACFU    361    GSRAHSSHLKSKKGQSTSRHKKFMFKTEGPDSD                            393
```


UniProtKB results
Basket
UniProtKB consists of two sections:
Reviewed (Swiss-Prot) - Manually annotated
Records with information extracted from literature and curator-evaluated computational analysis.
Unreviewed (TrEMBL) - Computationally analyzed
Records that await full manual annotation.
UniProtKB (3)
UniRef (0)
UniParc (0)
(max 400 entrie
Entry
Entry name
Organism
Remove
P55211
CASP9_HUMAN
Homo sapiens (Human)
O14727
APAF_HUMAN
Homo sapiens (Human)
Q14790
CASP8_HUMAN
Homo sapiens (Human)
Align
BLAST
Map Ids
Download
Full View
Remove
Clear
Filter by
BLAST
Align
Download
Add to basket
Reviewed (1,506)
Swiss-Prot
Unreviewed (99,644)
TrEMBL
Popular organisms
Human (496)
Mouse (317)
Rat (187)
Bovine (138)
Entry
Entry name
Protein names
Gene names
Organism
Length
2 result(s) selected. (Clear Selection)
P31944 CASPE_HUMAN Caspase-14 CASP14 Homo sapiens (Human) 242
O89094 CASPE_MOUSE Caspase-14 Casp14 Mus musculus (Mouse) 257
P42575 CASP2_HUMAN Caspase-2 CASP2 ICH1, NEDD2 Homo sapiens (Human) 452
P70343 CASP4_MOUSE Caspase-4 Casp4 Casp11, Caspl, Ich3 Mus musculus (Mouse) 373
P55215 CASP2_RAT Caspase-2 Casp2 Ich1 Rattus norvegicus (Rat) 452
https://www.uniprot.org/uniprot/?query=caspase&sort=score#

SECONDARY STRUCTURE

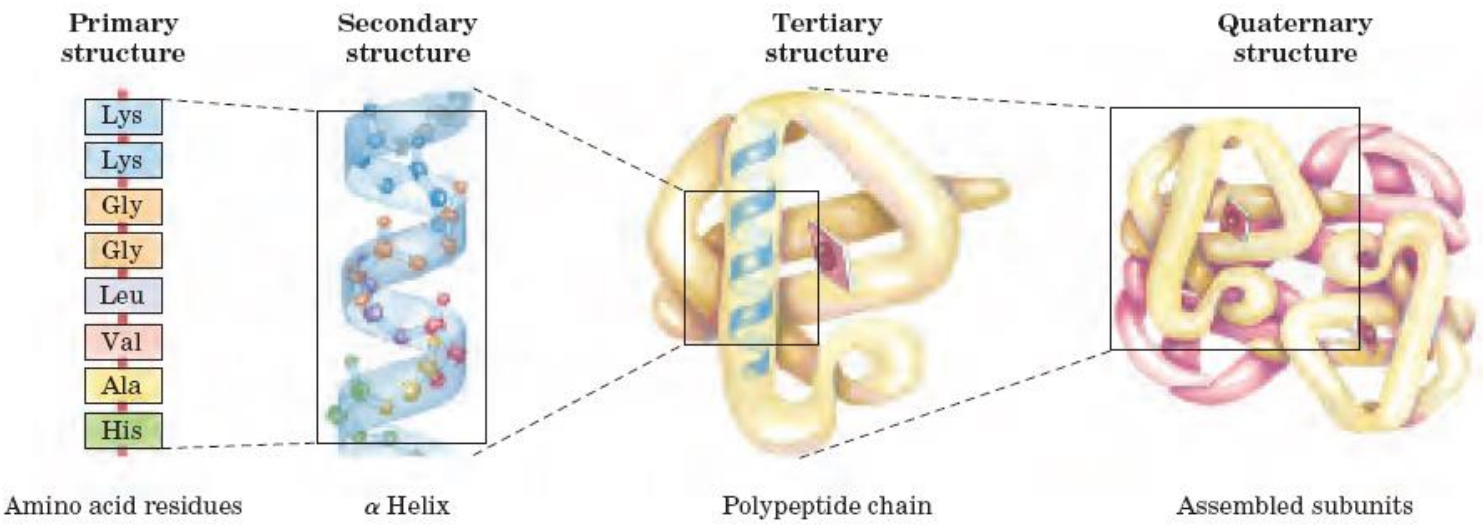


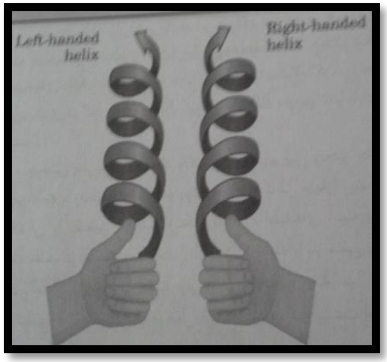
FIGURE 3–16 Levels of structure in proteins. The *primary structure* consists of a sequence of amino acids linked together by peptide bonds and includes any disulfide bonds. The resulting polypeptide can be coiled into units of *secondary structure,* such as an α helix. The helix is a part of the *tertiary structure* of the folded polypeptide, which is itself one of the subunits that make up the *quaternary structure* of the multisubunit protein, in this case hemoglobin.

Higher-level structure

Secondary structure: α-helix, β-strand and β-Turn
Tertiary structure
Quaternary structure

- Fibrous proteins versus Globular proteins
- Why Secondary structures are formed?
- α-helix and β-strand properties
- Loops such as β-turn
- Types of β-sheets: Parallel, Antiparallel and Mixed
- Detection of secondary structures by???

The α Helix Is a Common Protein Secondary Structure



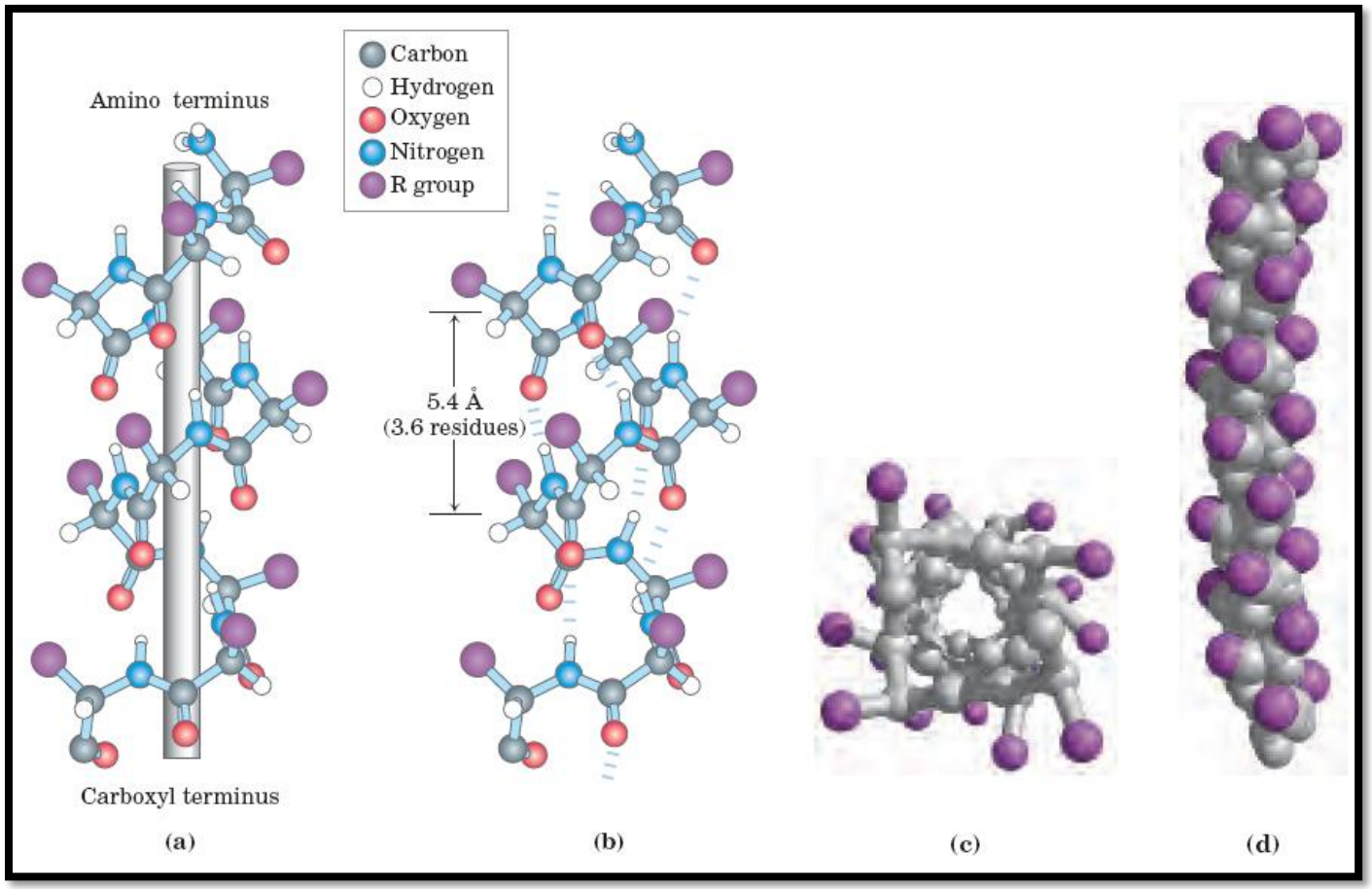
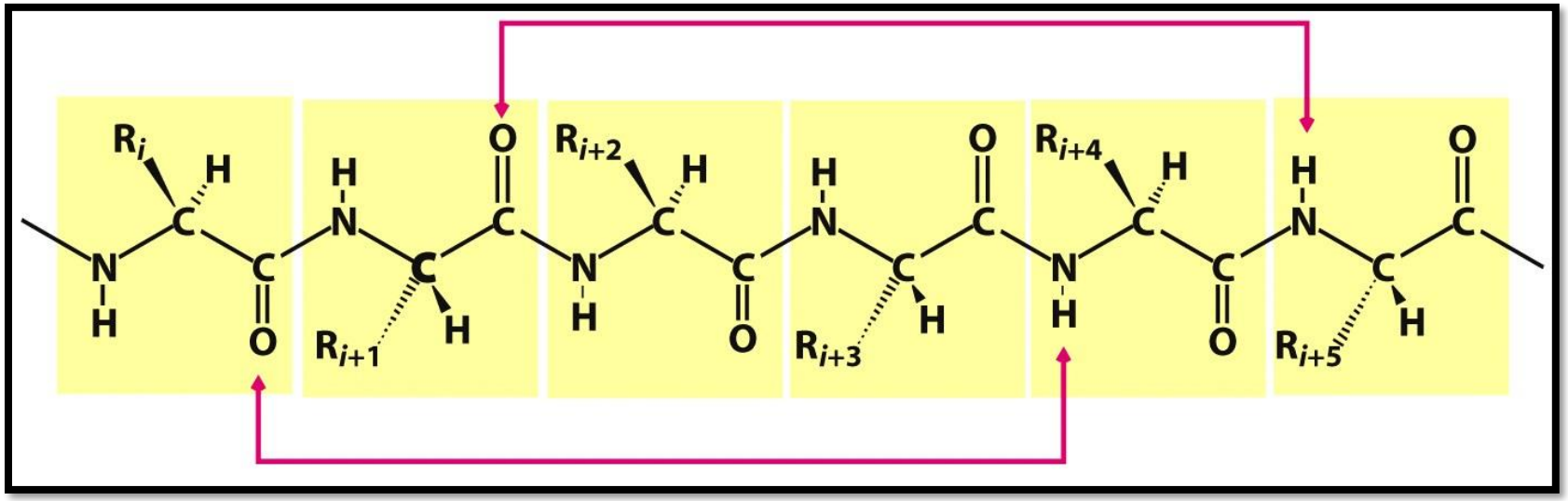
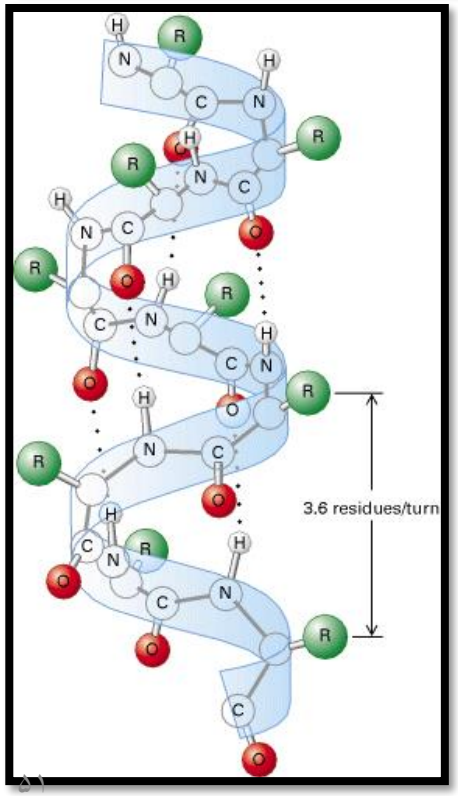
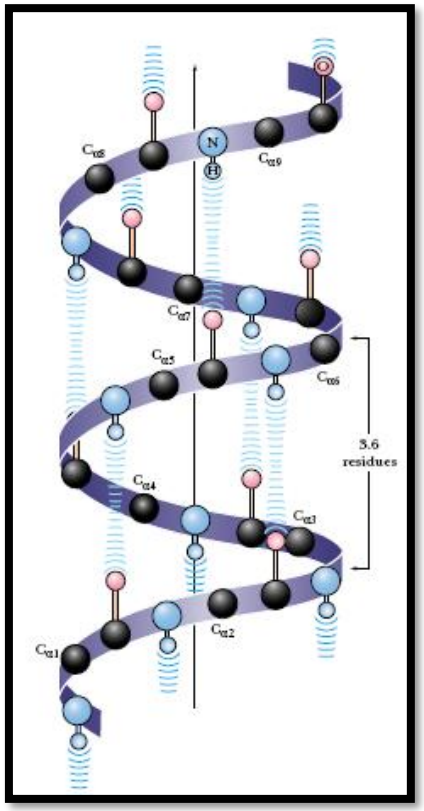


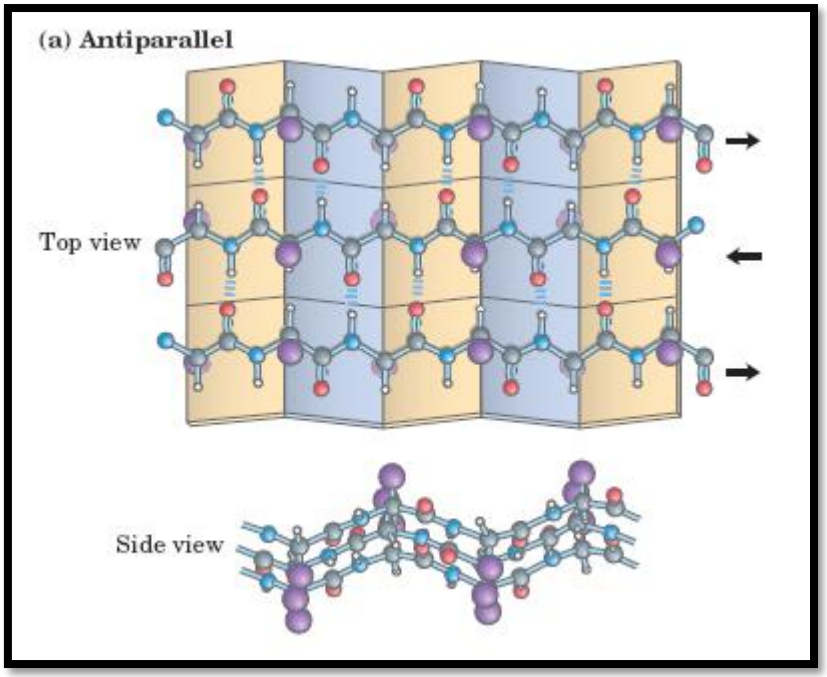
FIGURE 4–4 Four models of the α helix, showing different aspects of its structure. **(a)** Formation of a right-handed α helix. The planes of the rigid peptide bonds are parallel to the long axis of the helix, depicted here as a vertical rod. **(b)** Ball-and-stick model of a right-handed α helix, showing the intrachain hydrogen bonds. The repeat unit is a single turn of the helix, 3.6 residues. **(c)** The α helix as viewed from one end, looking down the longitudinal axis (derived from PDB ID 4TNC). Note the positions of the R groups, represented by purple spheres. This ball-and-stick model, used to emphasize the helical arrangement, gives the false impression that the helix is hollow, because the balls do not represent the van der Waals radii of the individual atoms. As the space-filling model **(d)** shows, the atoms in the center of the α helix are in very close contact.


R_i
R_{i+1}
R_{i+2}
R_{i+3}
R_{i+4}
R_{i+5}


3.6 residues/turn


3.6 residues

The β Conformation Organizes Polypeptide Chains into Sheets



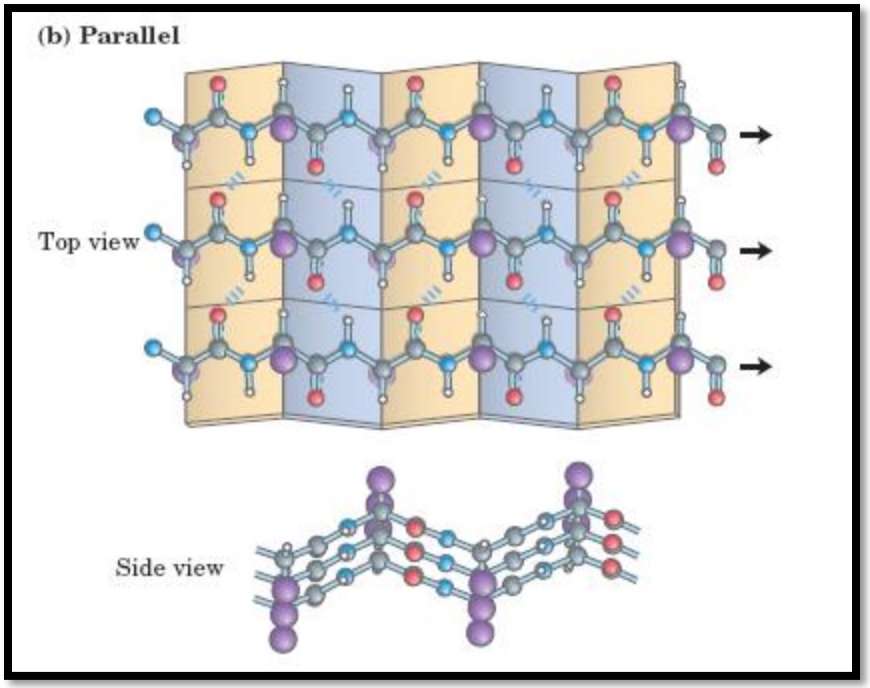
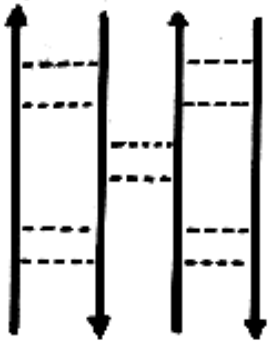
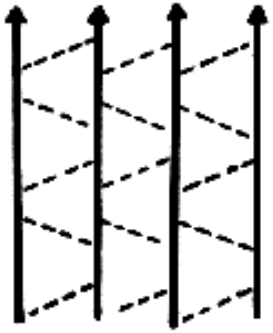


FIGURE 4–7 The β conformation of polypeptide chains. These top and side views reveal the R groups extending out from the β sheet and emphasize the pleated shape described by the planes of the peptide bonds. (An alternative name for this structure is β-pleated sheet.) Hydrogen-bond cross-links between adjacent chains are also shown. **(a)** Antiparallel β sheet, in which the amino-terminal to carboxyl-terminal orientation of adjacent chains (arrows) is inverse. **(b)** Parallel β sheet.

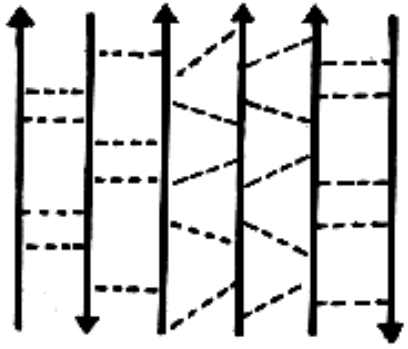
Antiparallel beta-sheet



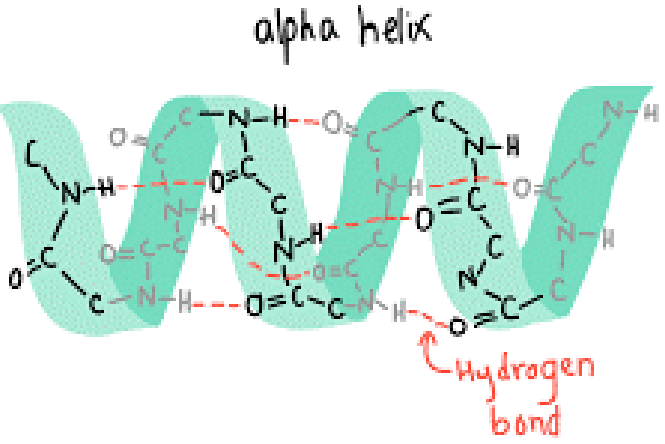


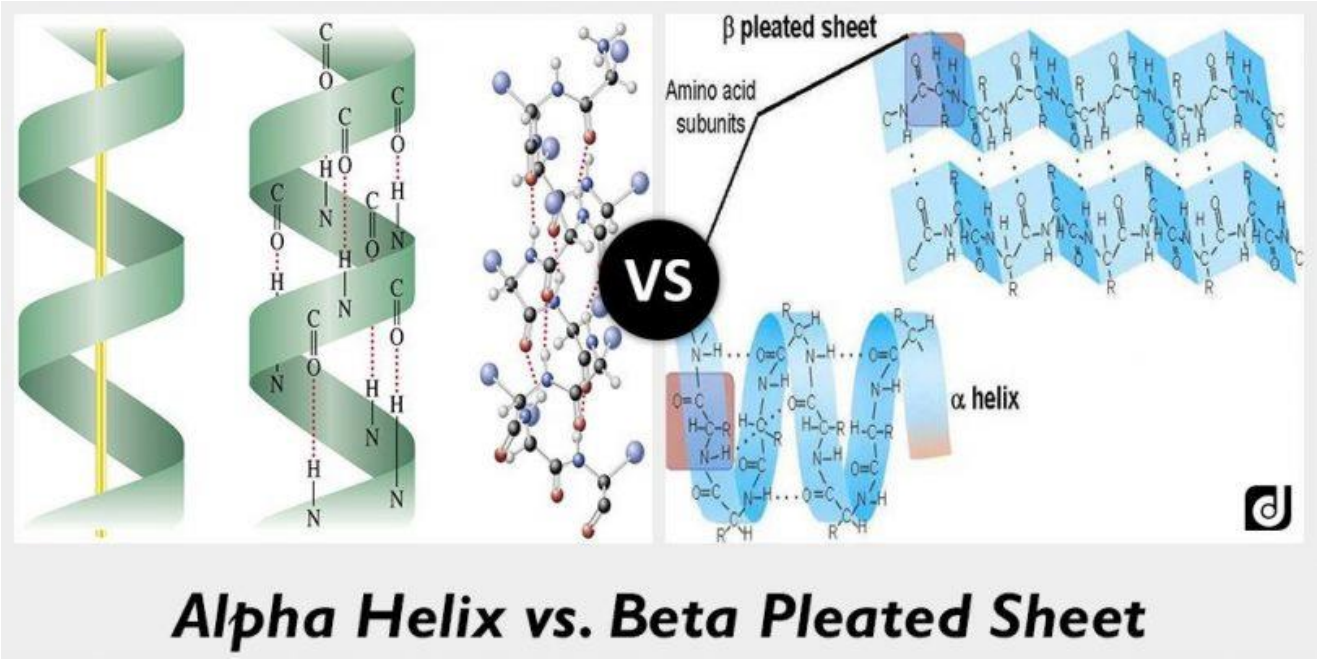
Parallel beta-sheet

The different types of beta-sheet. Dashed lines indicate main chain hydrogen bonds.



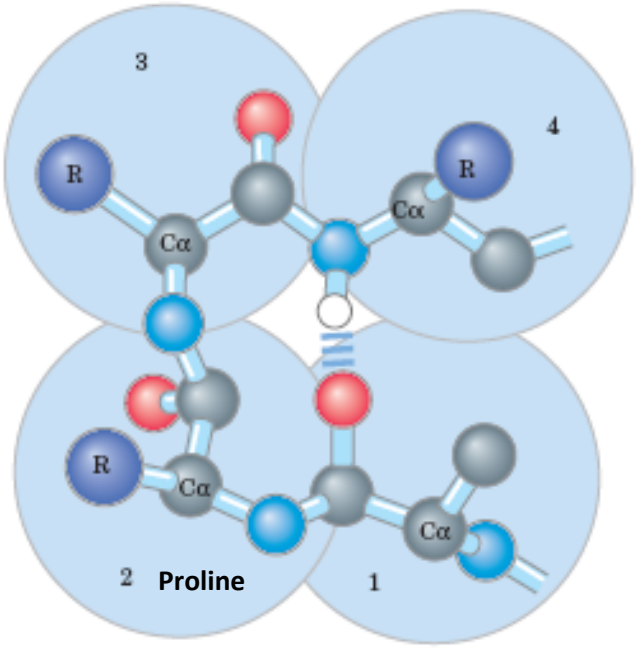
Mixed beta-sheet



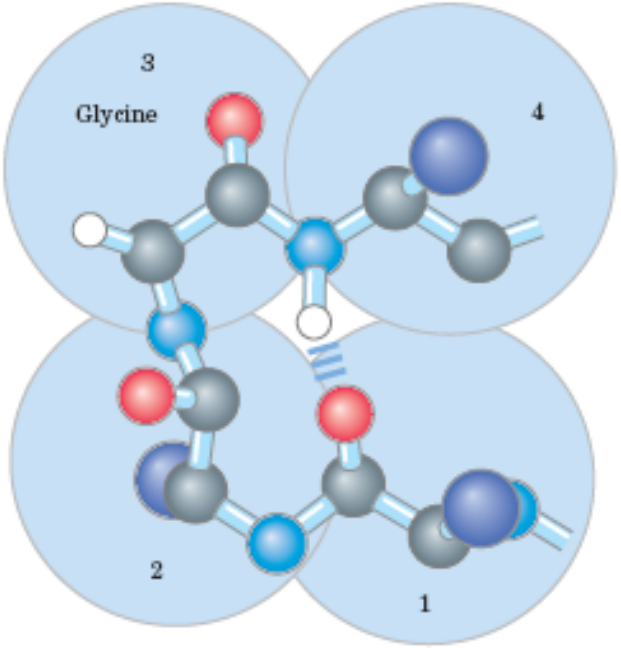


β Turns Are Common in Proteins

(a) β Turns



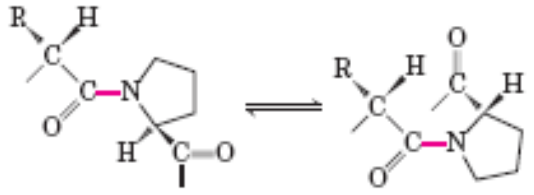
Type I



Type II

FIGURE 4–8 Structures of β turns. (a) Type I and type II β turns are most common; type I turns occur more than twice as frequently as type II. Type II β turns always have Gly as the third residue. Note the hydrogen bond between the peptide groups of the first and fourth residues of the bends. (Individual amino acid residues are framed by large blue circles.) (b) The trans and cis isomers of a peptide bond involving the imino nitrogen of proline. Of the peptide bonds between amino acid residues other than Pro, over 99.95% are in the trans configuration. For peptide bonds involving the imino nitrogen of proline, however, about 6% are in the cis configuration; many of these occur at β turns.

(b) Proline isomers



trans

cis

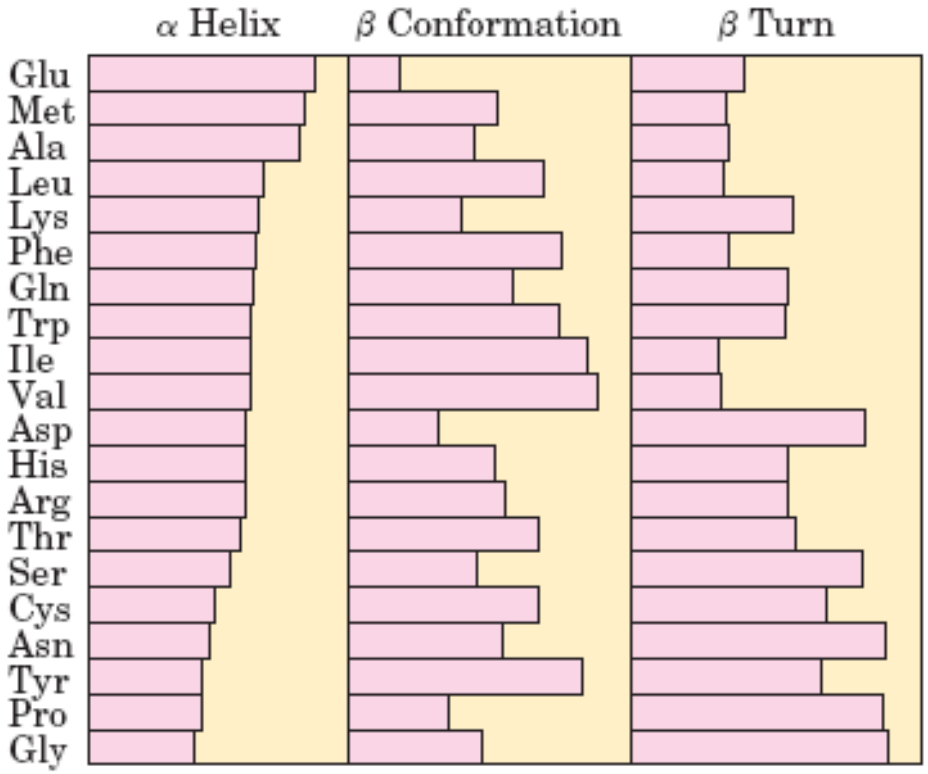
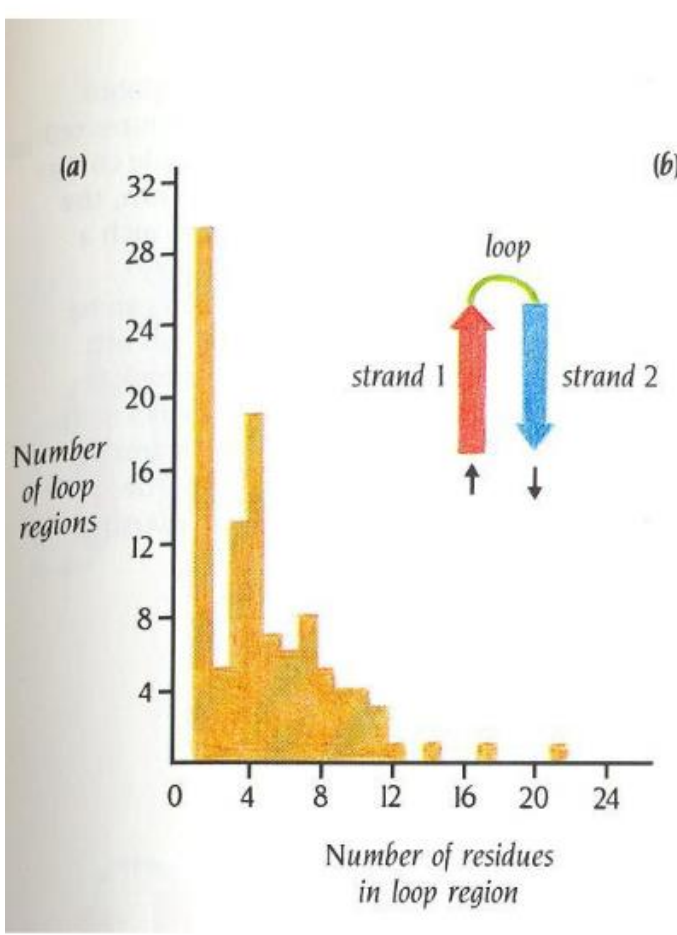


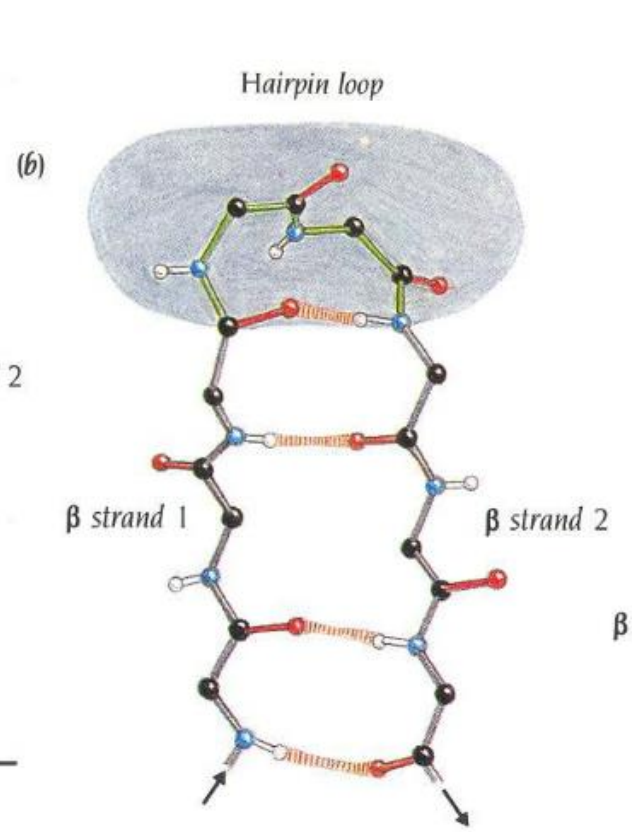
FIGURE 4–10 Relative probabilities that a given amino acid will occur in the three common types of secondary structure.

Loops

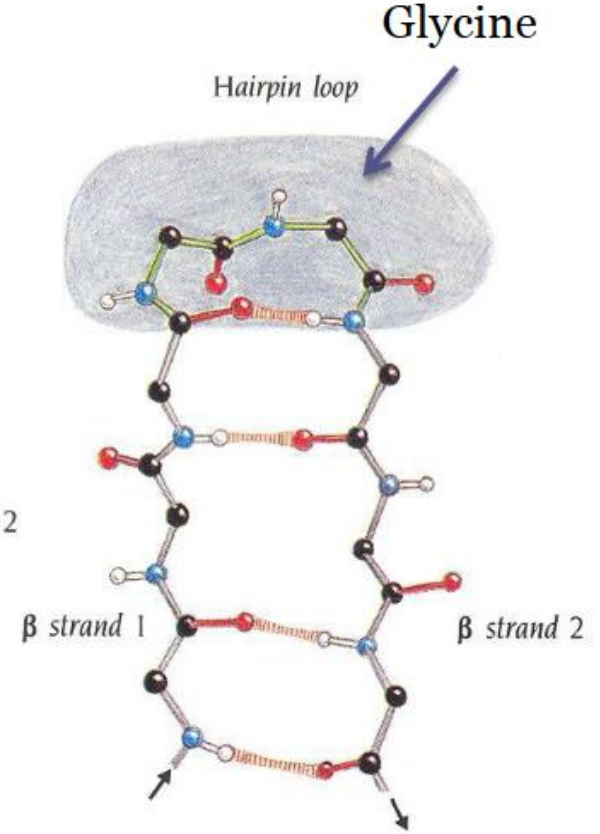
- Connected by loops secondary structure elements- reverse direction
- Various length and irregular shape
- A combination of secondary structure elements form the **stable hydrophobic core**
- The **loop** are at the **surface**
- The structure is a 180 turn involving **four** amino acid residues, the carbonyl oxygen of the 1^{st} residue forming a hydrogen bond with the amino-group hydrogen of the 4^{th} .
- The peptide groups of the central two residues **do not participate** in any inter residue hydrogen bonding.(exposed to solvent)
- Rich in charged and polar residue
- Participate in forming binding sites, enzyme active site
- For homologous amino acid sequences from different species insertions and deletions are mostly found in loop regions.
- Homologous sequence proteins show similar core structure which are not affected with various loop regions.

Loop regions are at the surface of protein molecules





Glycine



Type I

Type II

TERTIARY STRUCTURE

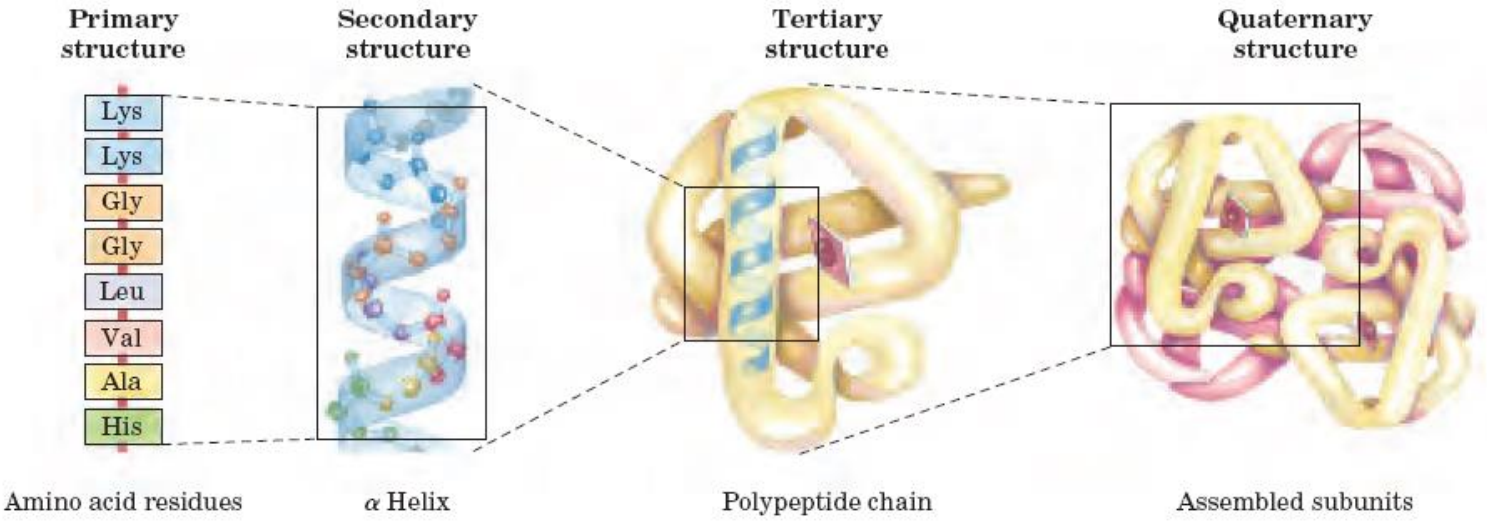
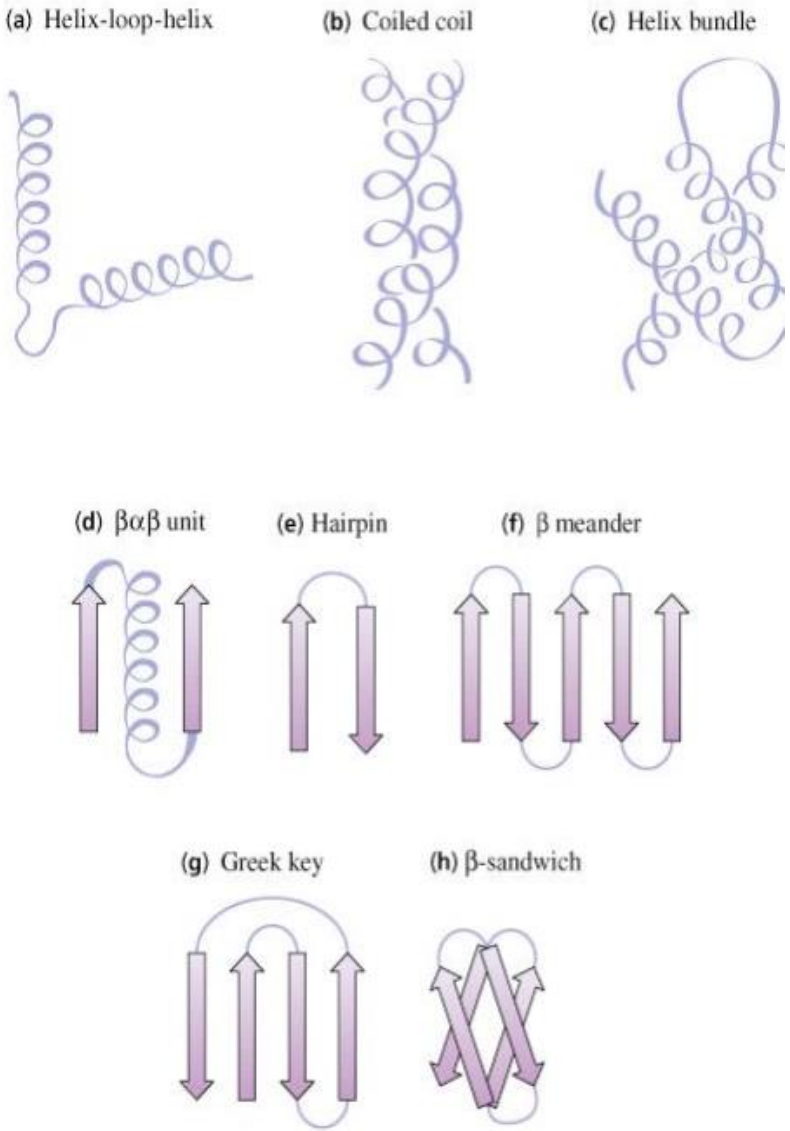
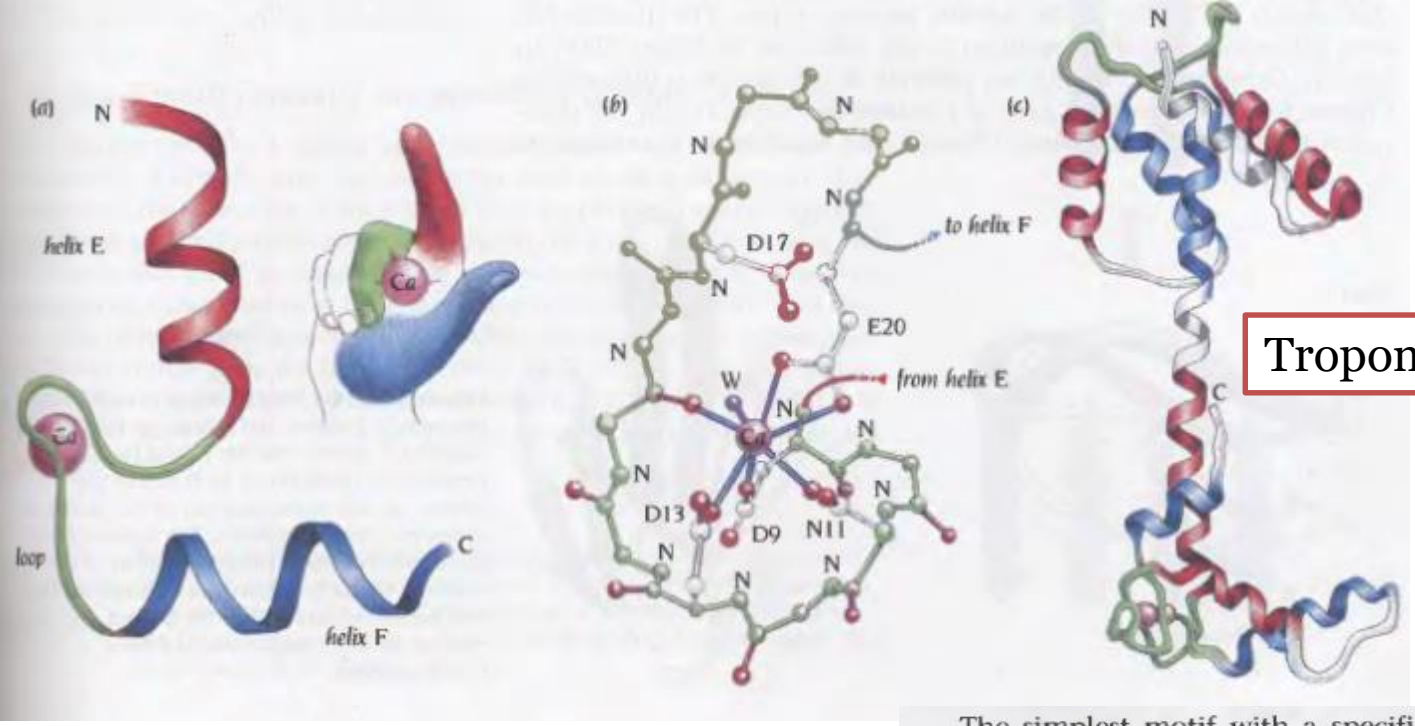


FIGURE 3–16 Levels of structure in proteins. The *primary structure* consists of a sequence of amino acids linked together by peptide bonds and includes any disulfide bonds. The resulting polypeptide can be coiled into units of *secondary structure,* such as an α helix. The helix is a part of the *tertiary structure* of the folded polypeptide, which is itself one of the subunits that make up the *quaternary structure* of the multisubunit protein, in this case hemoglobin.

Secondary structure elements are connected to form simple motifs



- Simple combination of a few secondary structure elements with a specific geometric arrangement
- Might be associated with specific function or not have biological function.
- BUT they are part of larger structural and functional assemblies



Troponin C has four EF hands

EF hand

- Helix-turn-Helix is specific for DNA binding.
- Helix-Loop-Helix is specific for calcium binding.

The simplest motif with a specific function consists of two α helices joined by a loop region. Two such motifs, each with its own characteristic geometry and amino acid sequence requirements, have been observed as parts of many protein structures (Figure 2.12).

One of these motifs, called the helix-turn-helix motif, is specific for DNA binding and is described in detail in Chapters 8 and 9. The second motif is specific for calcium binding and is present in parvalbumin, calmodulin, troponin-C, and other proteins that bind calcium and thereby regulate cellular activities. This calcium-binding motif was first found in 1973 by Robert Kretsinger, University of Virginia, when he determined the structure of parvalbumin to 1.8 Å resolution.

Parvalbumin is a muscle protein with a single polypeptide chain of 109 amino acids. Its function is uncertain, but calcium binding to this protein probably plays a role in muscle relaxation. The helix-loop-helix motif appears three times in this structure, in two of the cases there is a calcium-binding site. Figure 2.13 shows this motif which is called an **EF hand** because the fifth and sixth helices from the amino terminus in the structure of parvalbumin, which were labeled E and F, are the parts of the structure that were originally used to illustrate calcium binding by this motif. Despite this trivial origin, the name has remained in the literature.

Domain and Motif

- Several motifs usually combine to form compact globular structures, which are called **domains** (fundamental functional and structural units).

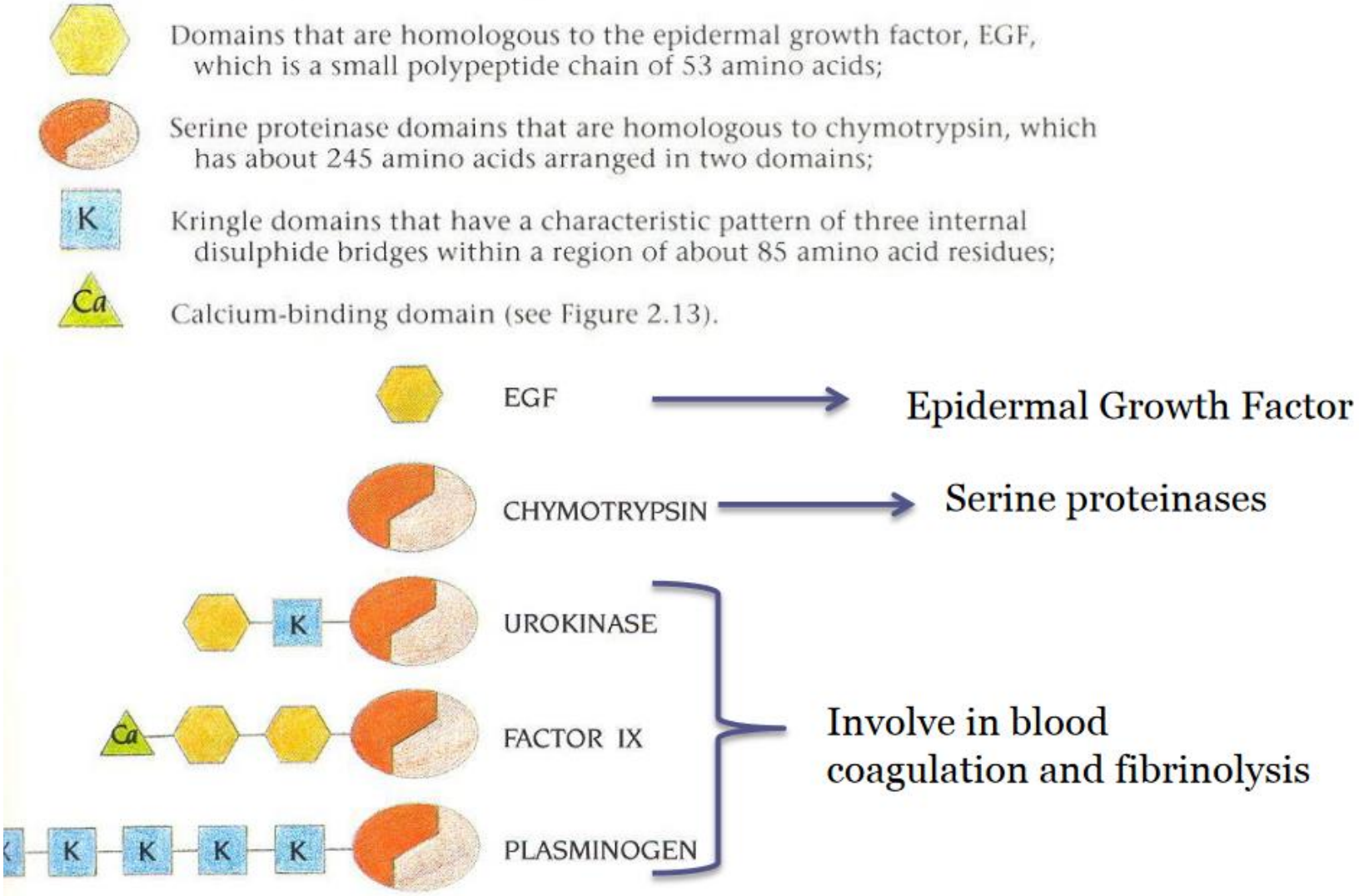
- **Tertiary structure**: the way motifs are arranged into domain structures and for the way a single polypeptide chain folds into one or several domains.
- Large polypeptide chains fold into several domains.

- There are many known examples where several biological functions that are carried out by separate polypeptide chains in one species are performed by domains of a single protein in another species.

- Sequences → Structural Motifs → Domain → Tertiary structure

  The number of such combinations is limited.

Large polypeptide chains fold into several domains



3)Tertiary structure

- The term *tertiary structure refers to the* ***unique three-dimensional*** conformations that globular proteins assume as they fold into their native (biologically active) structures.

1) amino acid residues that are distant from each other in the primary structure come into close proximity.

2) Because of efficient packing as the polypeptide chain folds, globular proteins are compact. Most water molecules are excluded from the protein's interior making interactions between both polar and nonpolar groups possible.

3) Large globular proteins often contain several compact units called **domains**.

Tertiary structure

- Domain, Motif (structural motif, sequence motif, functional motif) and Fold

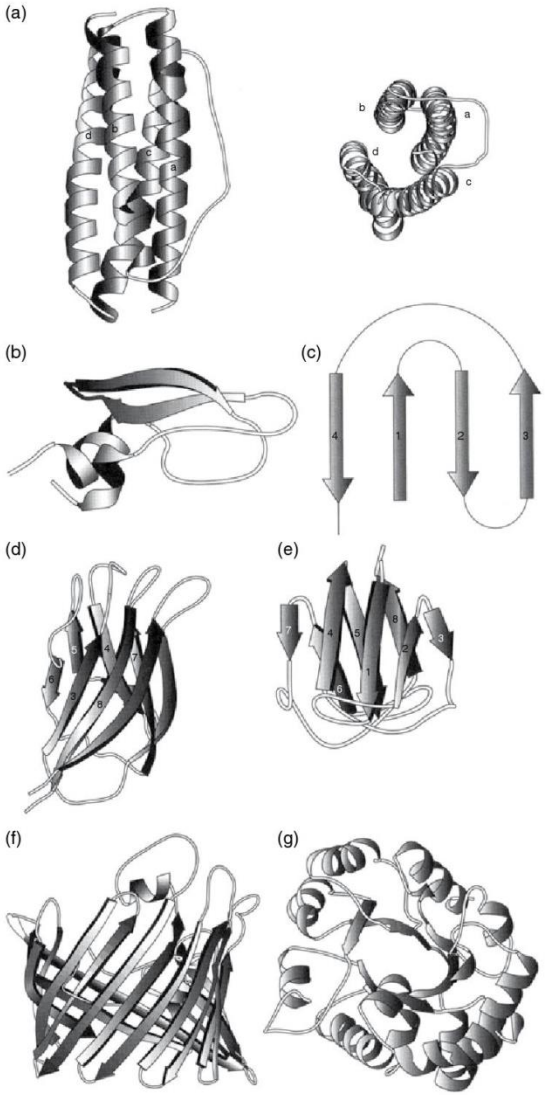


Figure 2.11 Some structural motifs commonly associated with (globular) polypeptides: (a) a four-helix bundle (b) a hairpin structure (c) a β sheet with a Greek key topology (d) a jelly roll motif (e) a β sandwich (f) a β barrel (g) an α/β barrel. Refer to text for further details. Reproduced from *Current Protocols in Protein Science* by kind permission of the publisher, John Wiley & Sons, Ltd.

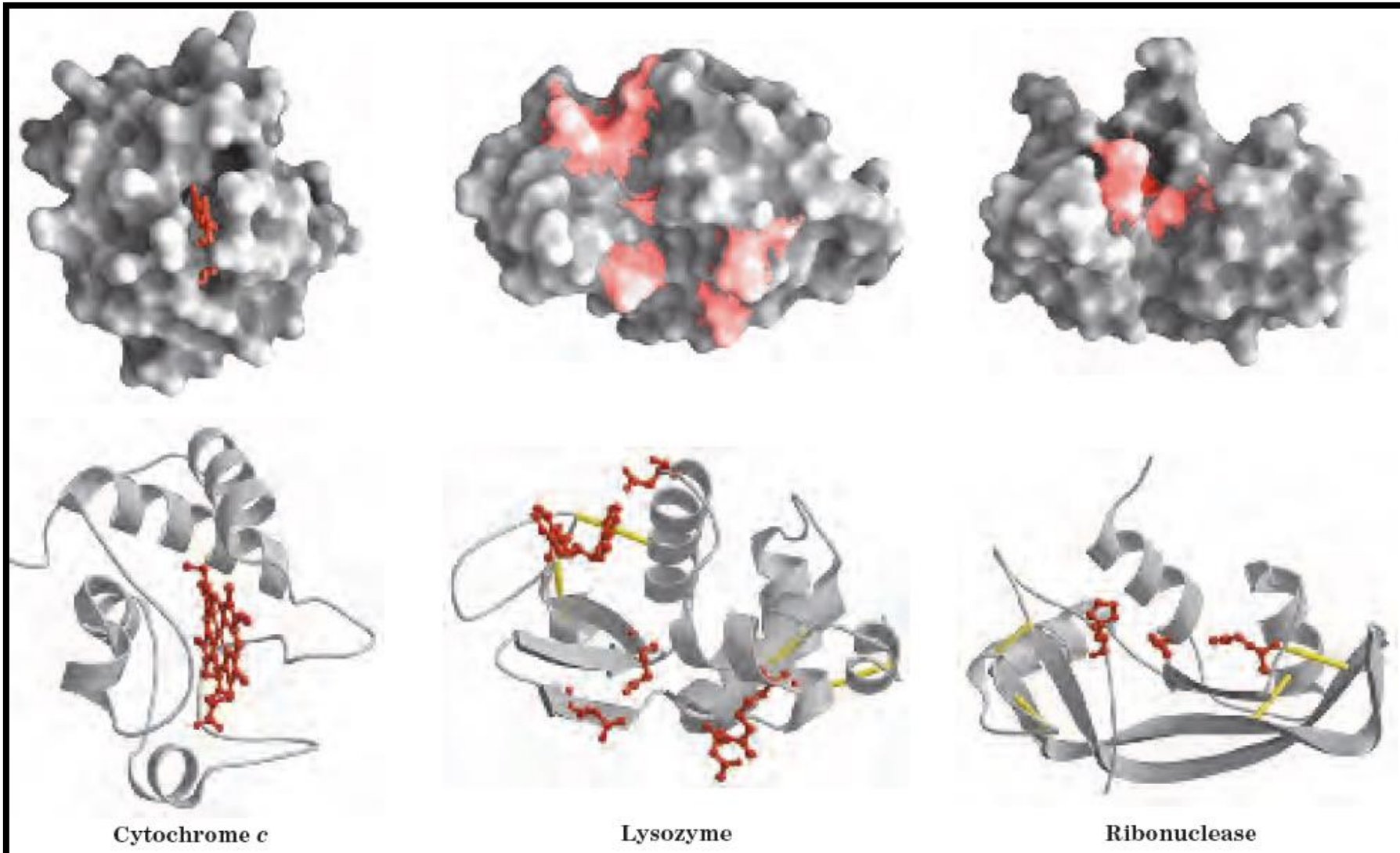


FIGURE 4–18 Three-dimensional structures of some small proteins. Shown here are cytochrome *c* (PDB ID 1CCR), lysozyme (PDB ID 3LYM), and ribonuclease (PDB ID 3RN3). Each protein is shown in surface contour and in a ribbon representation, in the same orientation. In the ribbon depictions, regions in the β conformation are represented by flat arrows and the α helices are represented by spiral ribbons. Key functional groups (the heme in cytochrome *c;* amino acid side chains in the active site of lysozyme and ribonuclease) are shown in red. Disulfide bonds are shown (in the ribbon representations) in yellow.

QUATERNARY STRUCTURE

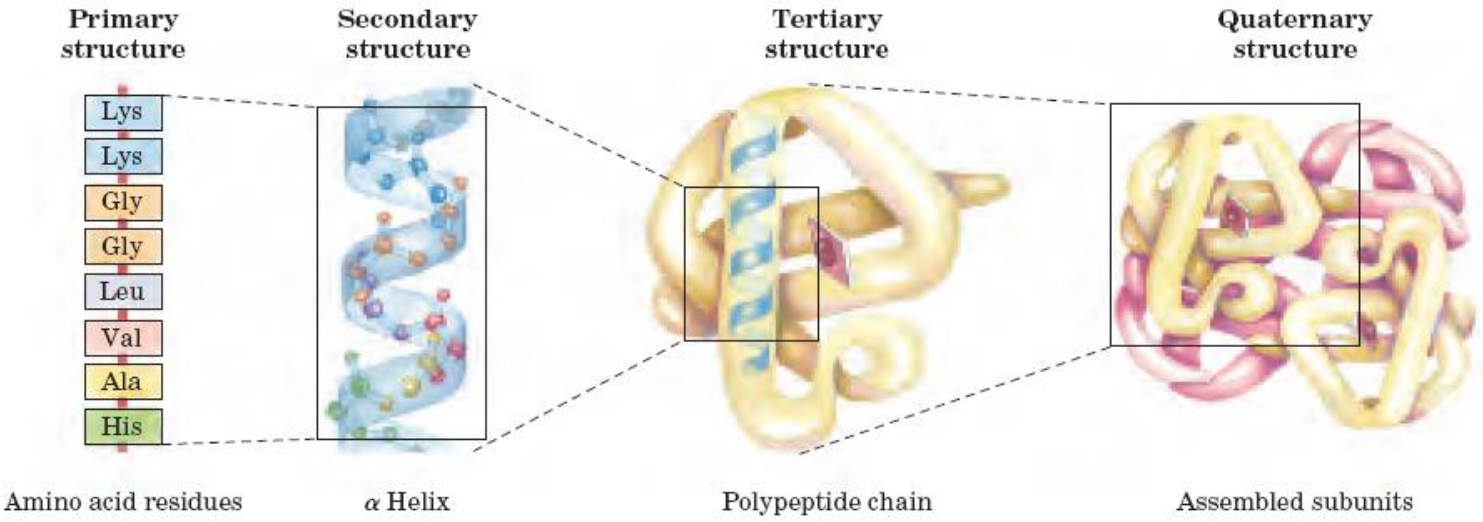
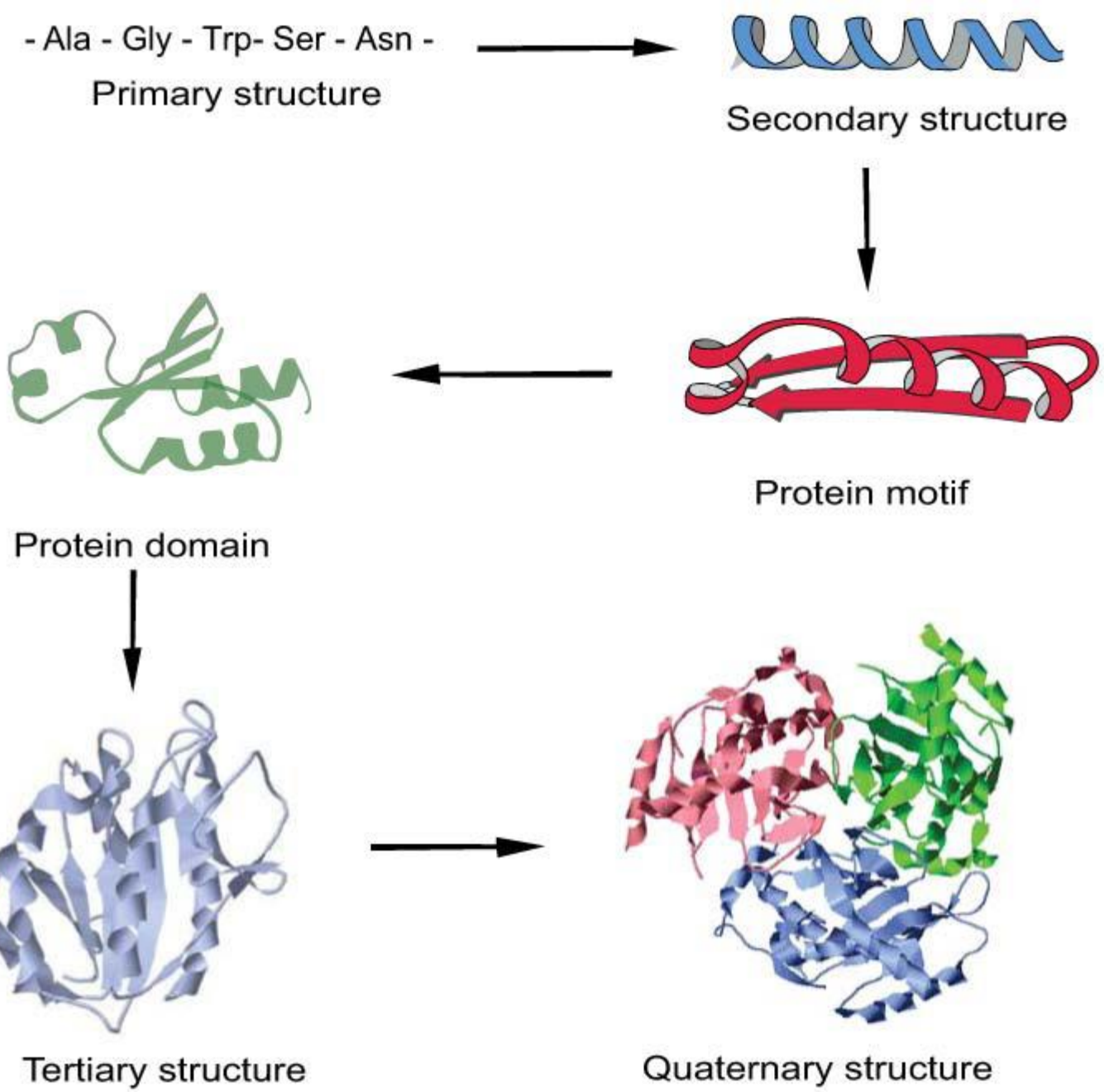


FIGURE 3–16 Levels of structure in proteins. The *primary structure* consists of a sequence of amino acids linked together by peptide bonds and includes any disulfide bonds. The resulting polypeptide can be coiled into units of *secondary structure,* such as an α helix. The helix is a part of the *tertiary structure* of the folded polypeptide, which is itself one of the subunits that make up the *quaternary structure* of the multisubunit protein, in this case hemoglobin.


- Ala - Gly - Trp- Ser - Asn -
Primary structure
Secondary structure
Protein motif
Protein domain
Tertiary structure
Quaternary structure

A Protein's Conformation Can Be Described as Its Overall Three-Dimensional Structure

The overall three-dimensional architecture of a protein is generally referred to as its **conformation.** This term is not to be confused with **configuration,** which denotes the geometric possibilities for a particular set of atoms (Figure 5.6). In going from one configuration to another, covalent bonds must be broken and rearranged. In contrast, the *conformational possibilities* of a molecule are achieved without breaking any covalent bonds. In proteins, rotations about each of the single bonds along the peptide backbone have the potential to alter the course of the polypeptide chain in three-dimensional space. These rotational possibilities create many possible orientations for the protein chain, referred to as its conformational possibilities. Of the great number of theoretical conformations a given protein might adopt, only a very few are favored energetically under physiological conditions. At this time, the rules that direct the folding of protein chains into energetically favorable conformations are still not entirely clear; accordingly, they are the subject of intensive contemporary research.

Difference between Conformation and Configuration

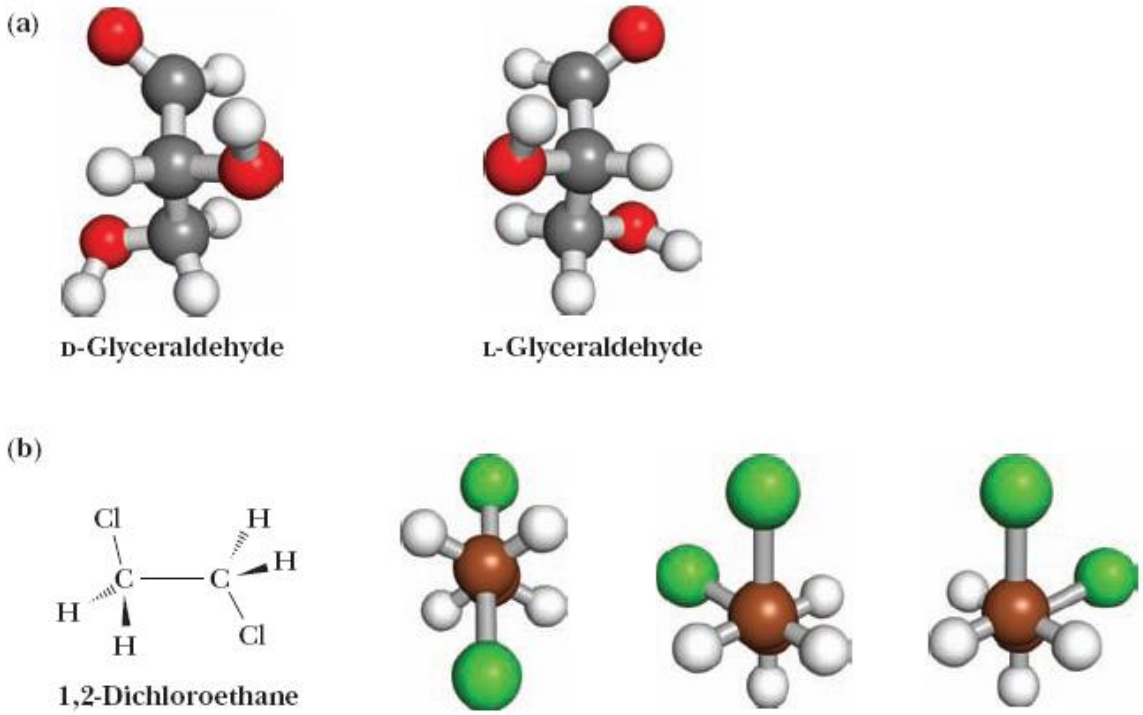


FIGURE 5.6 Configuration and conformation are *not* synonymous. **(a)** Rearrangements between configurational alternatives of a molecule can be achieved only by breaking and remaking bonds, as in the transformation between the D- and L-configurations of glyceraldehyde. **(b)** The intrinsic free rotation around single covalent bonds creates a great variety of three-dimensional conformations, even for relatively simple molecules, such as 1,2-dichloroethane. **(c)** Imagine the conformational possibilities for a protein in which two of every three bonds along its backbone are freely rotating single bonds. (Illustration: Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)

تفاوت کانفورماسیون و کانفیگوراسیون