
1

INORGANIC CHEMISTRY AND BIOCHEMISTRY ESSENTIALS

1.1 INTRODUCTION

Bioinorganic chemistry involves the study of metal species in biological systems. As an introduction to the basic inorganic chemistry is needed for understanding bioinorganic topics, this chapter will discuss the essential chemical elements, the occurrences and purposes of metal centers in biological species, the geometries of ligand fields surrounding these metal centers and ionic states preferred by the metals. The occurrence of organometallic complexes and clusters in metalloproteins will be discussed briefly and an introduction to electron transfer in coordination complexes will be presented. Since the metal centers under consideration are found in a biochemical milieu, basic biochemical concepts, including a discussion of proteins and nucleic acids, are presented later in this chapter.

1.2 ESSENTIAL CHEMICAL ELEMENTS

Chemical elements essential to life forms can be broken down into four major categories: (i) bulk elements (H/H^+ , C, N, $\text{O}^{2-}/\text{O}_2^-/\text{O}_2^{2-}$, P, and S/S^{2-}); (ii) macrominerals and ions (Na/Na^+ , K/K^+ , Mg/Mg^{2+} , Ca/Ca^{2+} , Cl^- , PO_4^{3-} , and SO_4^{2-}); (iii) trace elements ($\text{Fe}/\text{Fe}^{\text{II}}/\text{Fe}^{\text{III}}/\text{Fe}^{\text{IV}}$, $\text{Zn}/\text{Zn}^{\text{II}}$, and $\text{Cu}/\text{Cu}^{\text{I}}/\text{Cu}^{\text{II}}/\text{Cu}^{\text{III}}$); and (iv) ultratrace elements, that comprise nonmetals (F/F^- , I/I^- , Se/Se^{2-} , $\text{Si}/\text{Si}^{\text{IV}}$, As, and B) and metals ($\text{Mn}/\text{Mn}^{\text{II}}/\text{Mn}^{\text{III}}/\text{Mn}^{\text{IV}}$, $\text{Mo}/\text{Mo}^{\text{IV}}/\text{Mo}^{\text{V}}/\text{Mo}^{\text{VI}}$, $\text{Co}/\text{Co}^{\text{II}}/$

Co^{III}, Cr/Cr^{III}/Cr^{VI}, V/V^{III}/V^{IV}/V^V, Ni^I/Ni^{II}/Ni^{III}, Cd/Cd²⁺, Sn/Sn^{II}/Sn^{IV}, Pb/Pb²⁺, and Li/Li⁺). In the preceding classification, only the common biologically active ion oxidation states are indicated (see references [1, 2d] for more information). If no charge is shown, the element predominately bonds covalently with its partners in biological compounds, although elements such as carbon (C), sulfur (S), phosphorus (P), arsenic (As), boron (B), and selenium (Se) have positive formal oxidation states in ions containing oxygen atoms; i.e. S = +6 in the SO₄²⁻ ion or P = +5 in the PO₄³⁻ ion. The identities of essential elements are based on historical work done by Klaus Schwarz in the 1970s [3]. Other essential elements may be present in various biological species. Essentiality has been defined by certain criteria: (i) a physiological deficiency appears when the element is removed from the diet; (ii) the deficiency is relieved by the addition of that element to the diet; and (iii) a specific biological function is associated with the element [4]. Table 1.1 indicates the approximate percentages by weight of selected essential elements for an adult human.

Every essential element follows a dose-response curve, shown in Figure 1.1, as adapted from reference [4]. At lowest dosages, the organism does not survive whereas in deficiency regions the organism exists with less than optimal function. After the concentration plateau of the optimal dosage region, higher dosages cause toxic effects in the organism eventually leading to lethality. Specific daily requirements of essential elements may range from microgram to gram quantities.

Considering the content of earth's contemporary waters and atmospheres, many questions arise as to the choice of essential elements at the time of life's origins 3.5 billion or more years ago. Certainly sufficient quantities of the bulk elements were available in primordial oceans and at shorelines. However, the concentrations of essential trace metals in modern oceans may differ considerably from those found in prebiotic times. Iron's current approximate 10⁻⁴ mM concentration in seawater, for instance, may not reflect accurately its prelife-forms availability. If one assumes a mostly reducing atmosphere contemporary with the beginnings of biological life, the availability of the more soluble iron(II) ion in primordial oceans must have been much higher. Thus, the essentiality of iron(II) at a concentration of 0.02 mM in the blood plasma heme (hemoglobin) and muscle tissue heme (myoglobin) may be explained. Beside the availability factor, many chemical and physical properties of elements and

TABLE 1.1 Percentage Composition of Selected Elements in the Human Body

Element	Percentage (By Weight)	Element	Percentage (By Weight)
Oxygen	53.6	Silicon and Mg	0.04
Carbon	16.0	Iron and fluorine	0.005
Hydrogen	13.4	Zinc	0.003
Nitrogen	2.4	Copper and bromine	2 × 10 ⁻⁴
Sodium, potassium, and sulfur	0.10	Selenium, manganese, arsenic, and nickel	2 × 10 ⁻⁵
Chlorine	0.09	Lead and cobalt	9 × 10 ⁻⁶

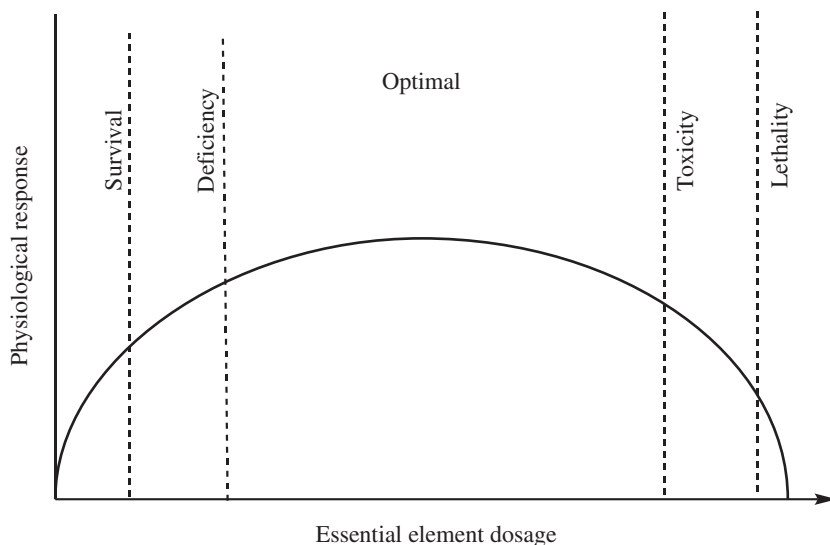


Figure 1.1 Dose-response curve for elements. *Source:* adapted from Kaim et al. [4].

their ions are responsible for their inclusion in biological systems. These include ionic charge, ionic radius, ligand preferences, preferred coordination geometries, spin pairings, systemic kinetic control, and the chemical reactivity of the ions in solution. These factors are discussed in detail by daSilva and Williams [1].

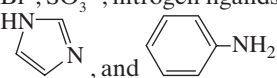
1.3 INORGANIC CHEMISTRY BASICS

Ligand preference and possible coordination geometries of the metal center are important bioinorganic principles. Metal ligand preference is closely related to the hard-soft acid-base nature of metals and their preferred ligands. These are listed in Table 1.2.

In general, hard metal cations form their most stable compounds with hard ligands and soft metal cations with soft ligands. Hard cations can be thought of as small dense cores of positive charge whereas hard ligands are usually the small highly electronegative elements or ligand atoms within a hard polyatomic ion, i.e. oxygen ligands in $(\text{RO})_2\text{PO}_2^-$ or CH_3CO_2^- .

It is possible to modify a hard nitrogen ligand towards an intermediate softness by increasing the polarizability of its substituents or the π electron cloud about it. The imidazole nitrogen of the amino acid histidine, a ubiquitous ligand in biological proteins, is an example. Increasing the softness of phosphate ion substituents can transform the hard oxygen ligand of $(\text{RO})_2\text{PO}_2^-$ to a soft state in $(\text{RS})_2\text{PO}_2^-$. Soft cations and anions are those with highly polarizable, large electron clouds – Hg^{2+} , sulfur ligands as sulfides or thiolates, and iodide ions. Also, note that metal ions can overlap into different categories. Lead as Pb^{2+} , for instance, appears in both the intermediate and soft categories. The Fe^{3+} ion, classified as a hard cation, coordinates to histidine (imidazole) ligands in

TABLE 1.2 Hard–soft Acid–base Classification of Metal Ions and Ligands

Metals, Ions, and Molecules				Ligands
Hard				Hard
High charge density				Low polarizability
Small ionic radius				High electronegativity
				Hard to oxidize
Hard				Hard
H ⁺	Mg ²⁺	Al ³⁺	SO ₃	Oxygen ligands in H ₂ O, CO ₃ ²⁻ , NO ₃ ⁻ , and PO ₄ ³⁻ ,
Na ⁺	Ca ²⁺	Co ³⁺	CO ₂	ROPO ₃ ²⁻ , (RO) ₂ PO ₂ ⁻ , and CH ₃ COO ⁻ ,
K ⁺	Mn ²⁺	Cr ³⁺		OH ⁻ , RO ⁻ , R ₂ O, and crown ethers
	VO ²⁺	Ga ³⁺		Nitrogen ligands in NH ₃ , N ₂ H ₄ , RNH ₂ , or Cl ⁻
		Fe ³⁺		Nitrogen ligands in NH ₃ , N ₂ H ₄ , RNH ₂ , or Cl ⁻
Intermediate				Intermediate
Fe ²⁺ , Ni ²⁺ , Zn ²⁺ , Co ²⁺ , Cu ²⁺ , Pb ²⁺ , Sn ²⁺ , Ru ²⁺ , Au ³⁺ , SO ₂ , and NO ⁺				Br ⁻ , SO ₃ ²⁻ , nitrogen ligands in NO ₂ ⁻ , N ₃ ⁻ , N ₂ , 
Soft				Soft
Low charge density				High polarizability
Large ionic radius				Low electronegativity
Easily excited outer shell electrons				Low energy vacant orbitals
				Easily oxidized
Soft				Soft
Cu ⁺	Pt ²⁺	Pt ⁴⁺		Sulfur ligands in RSH, RS ⁻ , R ₂ S, and R ₃ P
Au ⁺	Pb ²⁺			RNC, CN ⁻ , CO, R ⁻ , H ⁻ , I ⁻ , and S ₂ O ₃ ²⁻
Tl ⁺	Hg ²⁺			(RS) ₂ PO ₂ ⁻ and (RO) ₂ P(O)S ⁻
Ag ⁺	Cd ²⁺			
Hg ₂ ²⁺	Pd ²⁺			

biological systems and Fe²⁺, classified as intermediate, can coordinate to sulfur ligands and the carbon atom of CO (see Sections 3.1–3.3, 3.6, and 4.1).

In biological systems, many factors affect metal–ligand complex formation. Hard–soft acid–base considerations have already been mentioned. Concentrations of the metal and ligand at the site of complexation are determined locally through concentration gradients, membrane permeability to metals and ligands, and other factors. Various competing equilibria – solubility products, complexation, and/or acid–base equilibrium constants – sometimes referred to as “metal ion speciation,” all affect complex formation. Ion size and charge, preferred metal coordination geometry and ligand chelation effects, all affect metal uptake.

1.4 ELECTRONIC AND GEOMETRIC STRUCTURES OF METALS IN BIOLOGICAL SYSTEMS

A survey of metal ions in biological systems is given in Tables 1.3–1.7 including idealized geometry and references to additional information.

TABLE 1.3 Metals in Biological Systems: Charge Carriers

Metal	Coordination Number, Geometry	Preferred Ligands	Functions and Examples	References [#] Chapter.Section
Sodium, Na ⁺	6, octahedral	O-ether, hydroxyl, carboxylate	Charge carrier, osmotic balance, nerve impulses, sodium channels, sodium pump, and Na ⁺ /K ⁺ ATPase	[4] 13 [1] 8 [5] 5 [6b] 10.2 [6a] 9 [7] 6.4
Potassium, K ⁺	6–8, flexible	O-ether, hydroxyl, and carboxylate	Charge carrier, osmotic balance, nerve impulses, potassium channels, and Na ⁺ / K ⁺ ATPase	[4] 13 [1] 8 [5] 5 [6b] 10.2 [6a] 9 [7] 6.4

Common geometries adopted by transition metal ions that will be of most concern to readers of this text are illustrated in Figure 1.2. It is important to remember that in biological systems these geometries are usually distorted in both bond length and bond angle.

Transition metal ions play special roles in biological systems, with all elements from the first transition series except titanium (Ti) and scandium (Sc) occurring in thousands of diverse metalloproteins. Metal ions determine the geometry of enzymatic active sites, act as centers for enzyme reactivity, and act as biological oxidation–reduction facilitators. Molybdenum (Mo) appears to be the only transition element in the second transition series with a similar role. Vanadium (V), technetium (Tc), platinum (Pt), ruthenium (Ru), and gold (Au) compounds, as well as gadolinium (Gd) and other lanthanide complexes, are extremely important in medicinal chemistry.

Tables 1.3–1.7 list the d electron configuration for transition metal ions common to biological systems. To find the number of d electrons for any transition metal ion, the following is a useful formula:

Number of d electrons = atomic number for the element (Z) – oxidation state of the element’s ion – Z for the preceding noble-gas element.

Examples:	Fe(II): 26 – 2 – 18 (Argon) = 6 Mo(V): 42 – 5 – 36 (Krypton) = 1.
-----------	--

Note that there are a number of different methods for indicating the oxidation state of a metal ion, especially transition metal ions that have variable oxidation states. As an example, the iron ion in its +2 oxidation state may be written as Fe²⁺, Fe(II), Fe^{II}, or iron(II). In this text, the methods are used interchangeably.

As a consequence of their partially filled d orbitals, transition metals exhibit variable oxidation states and a rich variety of coordination geometries and

TABLE 1.4 Metals in Biological Systems: Structural and Triggers

Metal	Coordination Number, Geometry	Preferred Ligands	Functions and Examples	Reference [#] Chapter.Section <i>This work.Section</i>
Magnesium Mg^{2+}	6, octahedral	O-carboxylate, phosphate	Structure in hydrolases, isomerases, kinases, ATPases, phosphate transfer, and trigger reactions	[4] 13, 14 [1] 9 [5] 6.2 [6b] 4.2 [6a] 10 [7] 7.2, 10.4 [8] IX.1.2, IX.2.2, 5.5, 5.6, 5.7
Calcium Ca^{2+}	6–8, flexible	O-carboxylate, carbonyl, and phosphate	Structural/storage proteins, charge carrier, Ca-ATPases, phosphate transfer, calmodulin, troponin, and trigger reactions	[4] 13, 14, 15 [1] 10 [5] 6.3 [6b] 10.2 [6a] 11 [7] 7.1.c [8] IX.2.3, 5.5, 5.6, 5.7, XIV.3
Zinc Zn^{2+} (d^{10})	4, tetrahedral	O-carboxylate, carbonyl, S-thiolate, and N-imidazole	Structure in zinc fingers, gene regulation, anhydrases, dehydrogenases, hydrolases, carbonylpeptidase, and carbonic anhydrase	[4] 12, 15 [1] 11.4.4 [5] 2.4.1 [6b] 4.3, 10.1.3 [6a] 7.8, 12 [7] 7.1.b, 8.1 [8] XIV.1.1, 1.2, 1.5
Zinc Zn^{2+} (d^{10})	5, square pyramid	O-carboxylate, carbonyl, and N-imidazole	Structure in hydrolases and peptidases	4.2 [4] 12, 15 [1] 11.4.4 [6b] 4.3, 10.1.3 [6a] 7.8, 12 [7] 7.1.b, 8.1 [8] XIV.1.1, 1.2, 1.5
Manganese Mn^{2+} (d^5)	6, octahedral	O-carboxylate, phosphate, and N-imidazole	Structure in oxidases and photosynthesis	4.2 [4] 4 [1] 14 [5] 6.2.4 [6b] 4.4.2, 5.5 [6a] 16 [7] 10.4 [8] XIV.1.1, 1.2, 1.5
Manganese Mn^{2+} (d^4)	6, tetragonal	O-carboxylate, phosphate, and hydroxide	Structure in oxidases and photosynthesis	[4] 4 [1] 14 [5] 6.2.4 [6b] 4.4.2, 5.5 [6a] 16 [7] 10.4 [8] XIV.1.1, 1.2, 1.5

TABLE 1.5 Metals in Biological Systems: Electron Transfer

Metal	Coordination Number, Geometry	Preferred Ligands	Functions and Examples	Reference [#] <i>Work.Section</i>	Chapter.Section <i>This</i> <i>Work.Section</i>
Iron Fe ²⁺ (d ⁶) Fe ³⁺ (d ⁵)	4, tetrahedral	S-thiolate	Oxidoreductases Nitrogen fixation in nitrogenases and iron-sulfur clusters	[4] 73, 74, 11.2 [1] 12, [5] 3.6.4 [6b] 8 [6a] 13 [7] 5.3, 5.4.d, 9.1.a [8] X.1.3, XII 3.3	
Iron Fe ²⁺ (d ⁶) Fe ³⁺ (d ⁵)	6, octahedral	Porphyrins, N-imidazoles O-carboxylate, alkoxide, oxide, and phenolate	Oxidoreductases, Cytochrome hemes Soluble methane monooxygenase sMMO	[4] 6.1–6.3 [1] 13.9 [5] 7.3–7.8 [6b] 5.1.1.1, 5.2.1, 5.2.3 [6a] 13 [8] X.1.4 3.2 [4] 76.2 [1] 12, [5] 79.3 [6b] 6.3.3 [7] 11.2.b [8] XI.5.1.2 3.5	
Copper Cu ⁺ (d ¹⁰) Cu ²⁺ (d ⁹)	3, trigonal planar 4, tetrahedral	Cu _B N-Imidazole Cu _A N-imidazole S-thiolate O-carboxylate	Oxidoreductases Type III heme-copper oxidases, Type III Cu _A and Type II Cu _B in cytochrome c oxidase	[4] 10.4 [1] 13.9.1, 15.3.2 [5] 7.8 [6a] 13 [8] X.3.7, XI.8 3.2	
Copper Cu ⁺ (d ¹⁰) Cu ²⁺ (d ⁹)	4, tetrahedral	S-Thiolate, thioether and N-imidazole	Particulate methane monooxygenase pMMO Oxidoreductases Type I blue copper proteins plastocyanin, azurin, and nitrite reductase	[4] 76.2 [1] 12 [5] 79 [7] 11.2.b [8] XI.5.1.2 3.5 [4] 10.4 [1] 13.9.1, 15.3.2 [6b] 5.3 [6a] 13 [8] X.3.7, XI.8	
	Distorted tetrahedral		Type III copper oxidases and Cu _A in cytochrome c oxidase	[5] 7.8 [6b] 5.3	
	Distorted tetrahedral or trigonal planar	N-imidazole O-carbonyl	Particulate methane monooxygenase pMMO	[4] 76.2 [1] 12 [5] 79 [7] 11.2.b [8] XI.5.1.2 3.5	

TABLE 1.6 Metals in Biological Systems: Dioxygen Transport

Metal	Coordination Number, Geometry	Preferred Ligands	Functions and Examples	Reference [#] Chapter.Section <i>This Work.Section</i>
Copper Cu ²⁺ (d ⁹) Cu ⁺ (d ¹⁰)	5, square pyramid 6, tetragonal	O-carboxylate, N-imidazole	Type II copper oxidases and hydroxylases Type III copper hydroxylases and dioxygen transport in hemocyanin	[4] 5.3, 10.1–10.3 [1] 15 [6b] 5.3, 10.1.2 [6a] 14 [7] 11.1.c [8] XI 4.4.2 [9] 5.2.4
Iron Fe ²⁺ (d ⁶) Fe ³⁺ (d ⁵)	6, octahedral	N-imidazole and porphyrin	Dioxygen transport in hemoglobin and myoglobin	[4] 5.2 [1] 13 [5] 7.2 [6a] 13 [7] 11.1.a [8] XI 4.5–4.7 [9] 4 3.1

ligand spheres. Although a free metal ion would exhibit degenerate d-electron energy levels, ligand field theory describes the observed splitting of these d electrons for metal ions in various ligand environments. In all cases, the amount of stabilization or destabilization of d-electron energy levels centers about the so-called barycenter of unsplit d-electron energy levels. Important splittings for bioinorganic applications are shown in Figure 1.3

The t_{2g} (d_{xy} , d_{yz} , and d_{zx}) and e_g ($d_{x^2-y^2}$ and d_{z^2}) energy level designations identify symmetry properties of the d orbitals and are often used to indicate the degenerate energy levels under discussion. Generally, the energy gap between stabilized and destabilized d-electron energy levels for tetrahedral fields (Δ_t) is approximately one half that for octahedral fields (Δ_{oh}) and that for square planar fields is approximately $1.2 \Delta_{oh}$. Many thermodynamic and kinetic properties of transition metal coordination complexes can be predicted by knowing the magnitude of Δ . Measurement of ultraviolet and visible absorption spectra of transition metal complexes that arise from these quantum mechanically forbidden (but observed) d–d transitions provide a measure of Δ .

For octahedral (O_h) geometry, the repelling effect of like charge approach of the ligand electrons towards regions of high d-electron density along the x, y, and z axes elevates the energy of the e_g ($d_{x^2-y^2}$ and d_{z^2}) orbitals while the t_{2g} (d_{xy} , d_{yz} , and d_{zx}) orbitals are proportionally lowered in energy. For the tetrahedral (T_d) case ligands approach between the x, y, and z axes, stabilizing $d_{x^2-y^2}$ and d_{z^2} and destabilizing d_{xy} , d_{yz} , and d_{zx} orbital energy levels. For the square planar case, ligands will approach along the x and y axes.

Distorted octahedral and tetrahedral geometries are quite common in biological systems. Octahedral geometries are found for iron ions in heme ligand systems to be discussed in Section 3.1, for instance, while copper ions occur in distorted pyramidal, tetrahedral, or even trigonal bipyramidal forms.

TABLE 1.7 Metals in Biological Systems: Enzyme Catalysis

Metal	Coordination Geometry	Preferred Ligands	Functions and Examples	Reference [#] Chapter.Section <i>This Work</i> .Section
Copper and zinc Cu ²⁺ (d ⁹) Zn ²⁺ (d ¹⁰)	Cu 4–5, square planar to square pyramidal Zn 4-distorted tetrahedral	N-imidazole N-imidazole O-carboxylate	Oxidoreductases, dismutases, and Type II copper in superoxide dismutase Anhydrases and carbonic anhydrase	[4] 10.5 [1] 15.5 [6b] 5.3.6 [6a] 14 [7] 11.4.a [8] XI.2 [9] 5.2.3 3.4 [4] 11.1–11.3 [1] 17 [6b] 4.3.1 [6a] 12 [7] 5.4.d [8] XII.6 4.2
Cobalt Co ²⁺ (d ⁷) substitution for zinc	4, tetrahedral 5, trigonal bipyramidal with water or solvent ligands	S-thiolate, thioether, N-imidazole	Transferases, oxidases, carbonic anhydrase (CA), and alcohol dehydrogenase (AD)	[10] [11]
Cobalt Co ³⁺ (d ⁶) Co ²⁺ (d ⁷) Co ⁺ (d ⁸)	6, octahedral Co ⁺ (d ⁸) usually missing the 6th ligand	O-carboxylate and N-imidazole	Transferases Alkyl group transfer in vitamin B ₁₂ (cyanocobalamin)	[4] 3 [1] 16.4 [6b] 7.4 [6a] 15 [7] 1.3.d, 3.3.b [8] XIII.2
Nickel Ni ²⁺ (d ⁸)	4, square planar-tetrahedral mix, called seesaw	S-Thiolate, thioether, N-imidazole, and polypyrrole	Hydrogenases and hydrolases	[4] 9.3 [1] 16.3 [6b] 5.6.1 [6a] 15, [7] 3.3.b [8] XII.1 4.1
Nickel Ni ⁺ (d ⁹) Ni ³⁺ (d ⁷)	6, octahedral		Hydrolases Urease	[4] 9.2 [1] 16.6 [6a] 15, [8] IX.3.2

continued

TABLE 1.7 Continued

Metal	Coordination Number, Geometry	Preferred Ligands	Functions and Examples	Reference [#] Chapter.Section <i>This Work.Section</i>
Fe ²⁺ (d ⁶) Fe ³⁺ (d ⁵)	4, tetrahedral	S-thiolate Thioether	Oxidoreductases Ferredoxin and rubredoxin Nitrogenases	[4] 72, 73 [1] 12 [6a] 13 [7] 5.3.a [8] X.1.3 6 [9] 6 1.6 [4] 11.1–11.3 [1] 17 [6b] 8 [7] 5.4.d [8] XII.6 [9] 6 3.3
Fe ²⁺ (d ⁶) Fe ³⁺ (d ⁵)	6, octahedral	S-thiolate N-imidazole	Oxygenases Nitric oxide synthase Reductases Nitrite reductase Nitrogen fixation in nitrogenases	4.3.1 4.3.2 [4] 11.1–11.3 [1] 17.5.1 [6b] 8 [7] 5.4.d [6a] 17 [8] XII.6 [9] 6 3.3
Molybdenum Mo ⁴⁺ (d ²) Mo ⁵⁺ (d ¹) Mo ⁶⁺ (d ⁰)	6, octahedral	O-Oxide, carboxylate, phenolate, S-sulfide, and thiolate	Oxo transfer in oxidases and xanthine oxidase	[4] 11.1, 11.1.2 [1] 17.4.4 [6b] 5.4.1

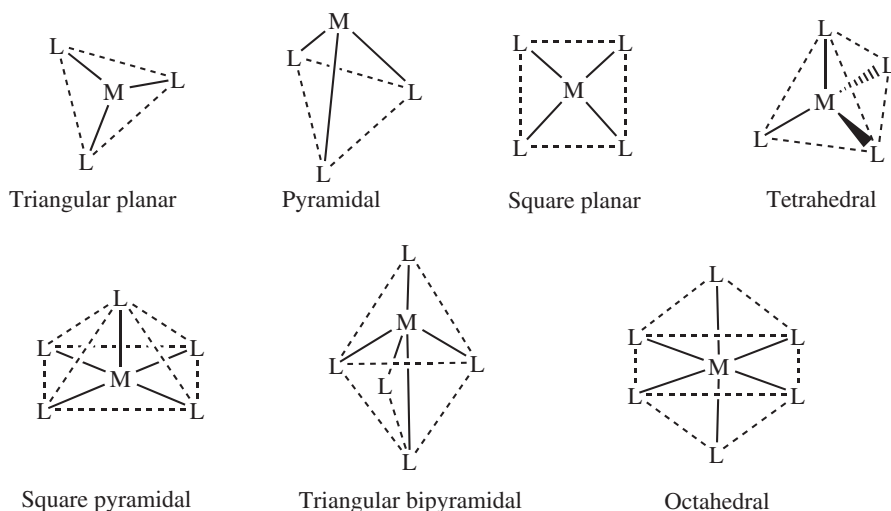


Figure 1.2 Common transition metal coordination geometries.

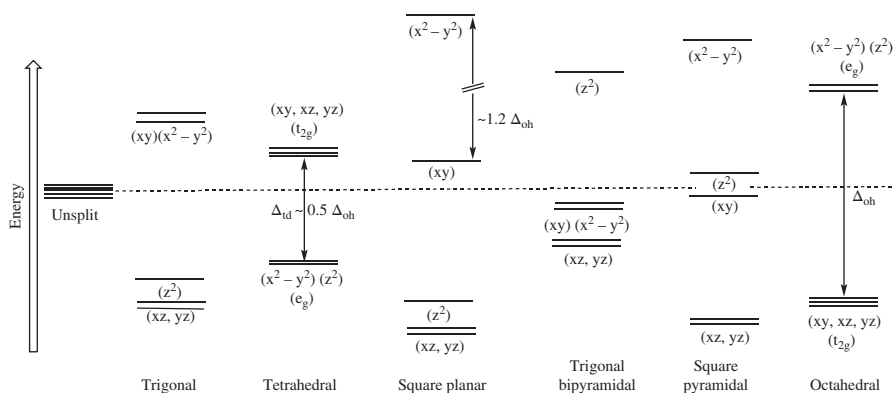


Figure 1.3 Ligand field splitting for d electrons in various ligand fields.

(see Sections 3.4 and 3.5.3). Square planar geometries, less common, are found for d^8 transition metal ions especially for gold(III), iridium(I), palladium(II), and platinum(II) and for nickel(II) species in strong ligand fields. The platinum anticancer agent, *cis*-dichlorodiammineplatinum(II), has a square planar geometry all important for its utilization as an antitumor agent. While the other geometries shown in Figure 1.2 might be less common for metal ions in biological species, they do occur (also with distorted bond distances and angles) and will be described in discussions of the metal center in the specific protein or enzyme.

The strength of the ligand field at a metal center is strongly dependent on the character of the ligand's electronic field and leads to the classification of ligands according to a "spectrochemical series" arranged below in order from weak field (halides, sulfides, and hydroxides) to strong field (cyanide and carbon monoxide):

$I^- < Br^- < S^{2-} < Cl^- < NO_3^- < OH^- \sim RCOO^- < H_2O \sim RS^- < NH_3 \sim \text{imidazole} < \text{en}$ (ethylenediamine or diaminoethane) $< \text{bpy}$ (2,2'-bipyridine) $< CN^- < CO$.

The iron–sulfur clusters of the nitrogenase enzyme with weak-field sulfur ligands are discussed in Section 3.3.

Ligand field strength may determine coordination geometry. For example, $NiCl_4^{2-}$ occurs as a tetrahedral complex (small splitting – small Δ_{td}) whereas $Ni(CN)_4^{2-}$ occurs in the square planar geometry (large energy gap – large Δ_{sp}). In octahedral fields, ligand field strength can determine the magnetic properties of metal ions since for d^4 through d^7 electronic configurations both high-spin (h.s., maximum unpairing of electron spins) and low-spin (l.s., maximum pairing of electron spins) complexes are possible. Possible configurations are shown in Figure 1.4.

In general, weak-field ligands form high spin (h.s.) complexes (small Δ_{oh}) and strong field ligands form low spin (l.s.) complexes (large Δ_{oh}). Tetrahedral complexes will be high spin (small Δ_{td}) and paramagnetic. Square planar complexes, usually found for metal ions having the electron configuration d^8 , will be diamagnetic – all electrons paired – since a large energy gap occurs between the last filled orbital (d_{xy}) and the $d_{x^2-y^2}$ orbital (see Figure 1.3). Detection of paramagnetism (unpaired electrons) and diamagnetism (all electrons paired) in bioinorganic ligand fields can help determine coordination geometry at active enzymes sites using electron paramagnetic resonance (EPR) spectroscopy. In the case of hemoglobin, for example, the d^6 iron(II) center cycles between high spin (paramagnetic) and low spin (diamagnetic) configurations. The change is evident in the presence or absence of an EPR spectrum. Section 3.5 of reference [5] discusses more detail on EPR. Placement of d electrons also affects the placement of the iron center in or out of the plane of its porphyrin ligand in hemoglobin or myoglobin – high spin systems

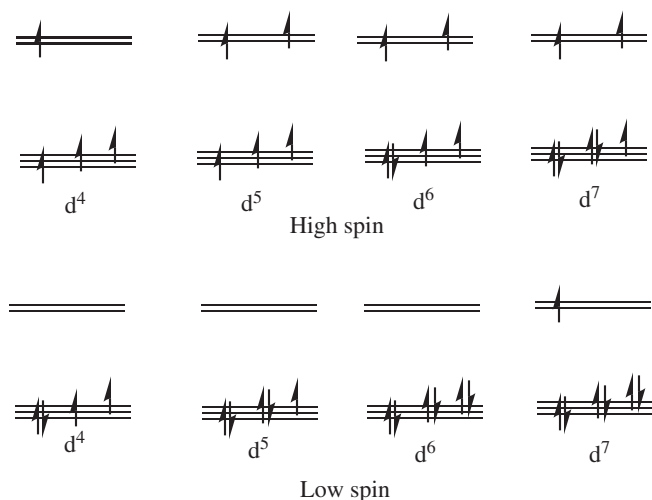


Figure 1.4 High-spin and low-spin electron configurations in an octahedral field.

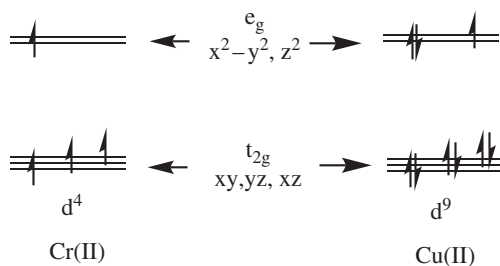


Figure 1.5 Electron configurations for high-spin Cr(II) and Cu(II).

require more room so that an h.s. Fe(II) ion will move out of the porphyrin ligand's planar coordination sphere (see Section 3.1 for further discussion with respect to this phenomenon in myoglobin and hemoglobin). In Type III copper enzymes, two d^9 copper(II) centers become antiferromagnetically coupled resulting in a loss of the expected paramagnetism (see Section 3.2 for a discussion of binuclear copper centers in cytochrome c oxidase).

The Jahn Teller effect arises in cases where removal of degeneracy of a d-electron energy level is caused by partial occupation of a degenerate level. Two common examples are those of Cu(II), d^9 and high spin Cr(II), d^4 , as shown in Figure 1.5. Electrons in the e_g level could be placed in either the $d_{x^2-y^2}$ or d_{z^2} orbitals. Placing the odd electron in either orbital destroys the degeneracy of the e_g orbitals and usually has the effect of moving the ligands on one axis in or out. For Cu(II) complexes, this effect is very common, resulting in longer bond lengths on what is usually taken as the complex's z axis. This will affect Cu(II)'s structure in enzymes such as cytochrome c oxidase (Section 3.2), superoxide dismutase (Section 3.4), and particulate methane monooxygenase (Section 3.5). The effect is also seen for high spin d^4 Mn(III) and for low spin d^7 Co(II) and Ni(III) complexes.

1.5 THERMODYNAMICS AND KINETICS

Thermodynamic equations apply to biological systems and are used by bioinorganic chemists to detail enzyme catalytic mechanisms especially using computational methods such as molecular dynamics (MD) and density functional theory (DFT) methods. These methods will be discussed in Sections 2.4 and 2.5.

The basic thermodynamic equations of interest are

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 = -nFE^0 = -RT\ln K \quad (1.1)$$

where ΔG^0 = the Gibbs free energy, ΔH^0 = the enthalpy change, ΔS^0 = the entropy change, n = the number of moles of electrons reducing 1 mol of the species being reduced, F = Faraday constant, E^0 = the standard reduction potential, R = the gas constant, T = temperature in degrees kelvin, and K = the

equilibrium constant = [products]/[reactants]. All conditions are standard: 1 atm pressure, 1 molar concentrations, $T = 298$ (25°C). The reduction potential for an oxidation–reduction reaction under other conditions is given by the Nernst equation:

$$E = E^0 - \left(\frac{RT}{nF} \right) \ln \left(\frac{[\text{reduced form}]}{[\text{oxidized form}]} \right) \quad (1.2)$$

Equation (1.2) assumes that the half reaction is written as a reduction, that is



When considering two half reactions combined into a cell, one must decide on the oxidant (the substance reduced) and the reductant (the substance oxidized) based on their relative E^0 values – the substance with the highest reduction potential will be reduced. The resulting reaction combining the two half reactions will be written:



The Nernst equation is then written in a convenient form by calculating out all the constants, converting to $\log_{10}$ form, and concentrations of products/reactants as

$$E = E^0 - \frac{0.0592 \text{ V}}{n} \log_{10} \frac{[B^{y+}][A^x]}{[B^y][A^{x+}]} \quad (1.5)$$

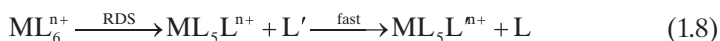
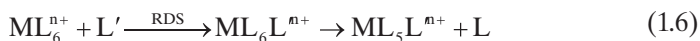
where V = volts = units of electromotive force.

In practice, the reduction potential of transition metal ions in biological systems will depend on a number of factors. For the Fe(III)/Fe(II) pair, strong field ligands such as CN^- or CO will cause the iron ions to adopt a low spin conformation preferentially, stabilizing the higher oxidation state and decreasing the systems reduction potential. Weak-field ligands such as H_2O or S^{2-} will cause the high spin state, stabilizing the lower oxidation state and increasing the reduction potential. The subject of electron transfer is addressed in section 1.8 of reference [5] including a brief discussion of Marcus theory. Rudolf Marcus' theory explains the rates of electron transfer reactions – the rate at which an electron can move from the electron donor to the electron acceptor. Section X.2.2.1 of reference [8] gives a more complete explanation. Electron-transfer proteins, including determinants of reduction potentials in proteins, are discussed in section X.1 of reference [8]. In this volume, electron transfer is discussed for iron–sulfur proteins (Section 3.3) and cytochromes (Section 3.2).

As students learned in their introductory chemistry courses, rates of reaction are divided into several classes depending on the number of reactants involved in rate determination. These are (i) zero order – the reaction rate is

independent of the concentration of that reactant; (ii) first order – the reaction rate is dependent on the concentration of one reactant; (iii) second order – the reaction rate is dependent on the concentration of two reactants; and (iv) higher order – the reaction rate is dependent on more than two reactants. Higher-order reaction rates are very rare as the possibility of bringing more than two reactants together productively is very small. Bioinorganic kineticists, studying the reaction rates of complex enzymatic reactions, often simplify matters to isolate a reaction of interest and relate it to a proposed mechanism for the enzyme's catalytic activity. For instance, in a pseudo-zero-order reaction – that is, one that would be first order under normal circumstances – the concentration of the enzyme may be held constant while a particular substrate's concentration is varied but does not affect the reaction rate. This condition may apply when the enzyme is saturated with substrate over the range of substrate concentration studied. In a pseudo-first-order reaction – that is, one that would normally be second order – the concentration of one reactant is held constant while the other is varied so that the reaction rate is directly proportional to the reactant whose concentration is varied. This is the most commonly used experimental technique used by enzyme kineticists.

In biological systems, as in all others, metal ions exist in an inner coordination sphere, an ordered array of ligands binding directly to the metal. Surrounding this is the outer coordination sphere consisting of other ligands, counterions, and solvent molecules. In stoichiometric mechanisms where one can distinguish an intermediate, substitution within the metals inner coordination sphere may take place through an associative (A), S_N^2 process as shown in Eqs. (1.6) (for six-coordinate complexes) and (1.7) (for four-coordinate complexes) or a dissociative (D), S_N^1 mechanism as shown in Eq. (1.8).



Associative mechanisms for metals in octahedral fields are difficult stereochemically (due to ligand crowding) and therefore rare for all but the largest metal ion centers. The associative mechanism is well known and preferred for four coordinate square planar complexes. Pure dissociative mechanisms are rare as well. When an intermediate cannot be detected by kinetic, stereochemical, or product distribution studies, the so-called interchange mechanisms (I) are invoked. Associative interchange (I_A) mechanisms have rates dependent on the nature of the entering group whereas dissociative interchange (I_D) mechanisms do not.

As the above discussion indicates, assigning mechanisms to simple anation reactions of transition metal complexes is not straight forward. The situation

becomes even more difficult for a complex enzyme system containing a metal cofactor at an active site. Methods developed to study the kinetics of enzymatic reactions according to the Michaelis–Menten model will be discussed in Section 1.9.3. Since enzyme-catalyzed reactions are usually very fast, experimentalists have developed rapid kinetic techniques to study them. Techniques used by bioinorganic chemists to study reaction rates will be further detailed in sections discussing application to specific enzymatic systems.

Transition state theory makes a connection between the thermodynamics and kinetics of the system under study. One develops the relationship between energy and reaction progress over an energy surface that changes as the reaction progresses. The highest point along the energy axis is called the energy of activation (E_A), the enthalpy of activation (ΔH^*), or the free energy of activation (ΔG^*). Figure I.17 of reference [6b] shows a graphical representation. Figure 3.20 shows the reaction progress along a pathway adding successive H atoms or ions to dinitrogen (N_2) and forming ammonia (NH_3) catalyzed by the nitrogenase enzyme. The Eyring–Polanyi and Arrhenius equations make the connection:

$$k = \frac{kT K_B}{h} e^{-\frac{\Delta G^*}{RT}} \quad (1.9)$$

$$k = \frac{kT K_B}{h} e^{-\frac{\Delta H^*}{RT}} e^{\frac{\Delta S^*}{RT}} \quad (1.10)$$

$$\ln \frac{k}{T} = \frac{-\Delta H^*}{R} \left(\frac{1}{T} \right) + \ln \frac{k K_B}{h} + \frac{\Delta S^*}{RT} \quad (1.11)$$

where k = the reaction rate constant, k = a coefficient often given the value unity, T = temperature in kelvin, R = the gas constant, K_B = the Boltzmann constant, h = Planck's constant, ΔH^* = the enthalpy of activation, and ΔS^* = the entropy of activation. If one now assumes that ΔH^* and ΔS^* are constant over the course of the reaction, the plot of $\ln(k/T)$ versus $1/T$ (Eq. 1.11) should yield a straight line with slope = $-\Delta H^*/R$, and y intercept = $\ln K_B/h + \Delta S^*/R$ allowing the experimentalist to find ΔH^* and ΔS^* .

Sections 1.5.1 and 1.5.2 of reference [5] and chapter 2 of reference [7] discuss the thermodynamics and kinetics of biological systems in more detail.

1.6 BIOORGANOMETALLIC CHEMISTRY

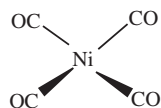
Organometallic complexes obey the so-called eighteen (18-e) and sixteen (16-e) electron rules as described below. These rules may also be applied to bioinorganic – bioorganometallic – systems. According to 16- or 18-e rules, the valence electrons of transition metals are considered to be filling the 4s, 3d, 4p or 5s, 4d, 5p shells. The most stabilized filled shell is determined to be 18 electrons – s^2 , d^{10} , p^6 , differing by the 10 electrons of the filled d shell from main

group element compounds stabilized by electron octets. Compounds or complexes fulfill the 18-e rule by addition of metal valence electrons and electron contributions from ligands. Metal valence shell electrons may be counted as if the metal, in its 0, +1, +2, +3, and +4 oxidation states, combines with ligand electrons counted according to Table 1.8 that lists common ligands and their electron contributions. Counting electrons in Table 1.8, if one gives the total number of valence electrons to the metal – sometimes called the covalent or neutral system – Fe would have eight electrons. Another electron counting system classifies ligands according to an ionic contribution – Fe's oxidation state depends on contributions from the ligands. Both these systems are more completely described in a website found at <http://www.ilpi.com/organomet/electroncount.html>.

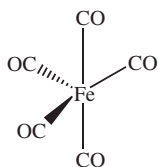
Many stable coordination complexes can be counted as having 16 electrons (16-e rule), especially those having square planar geometry and those bonded to aromatic rings through their π electronic systems. Some of these complexes, belonging to a group of compounds called metallocenes, bind to DNA and have antitumor properties. Examples discussed in the literature usually contain two cyclopentadienyl ligands, η^5 -cp, two chloride, Cl^- , ligands, and the metals titanium, Ti, vanadium, V, molybdenum, Mo, or niobium, Nb. The complexes have the overall formula Cp_2MCl_2 and are named, for example, as titanocene, vanadocene, or molybdocene complexes. The titanium antitumor complex, bis- η^5 -cyclopentadienyldichlorotitanium(IV), Cp_2TiCl_2 , is shown in Figure 1.6. This complex is the first metallocene, and the first non-platinum metal complex, to have undergone clinical trials as an anticancer agent. Structure–activity relationship studies have shown that the complex forms stable adducts with DNA [12]. Two unsubstituted cyclopentadienyl ligands

TABLE 1.8 Ligand Contributions to the 16- or 18-electron Rule

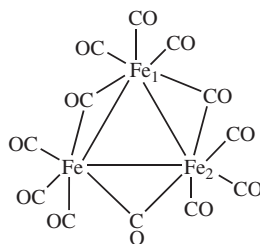
Ligand	Electrons Neutral/Covalent	Electrons Ox State/Ionic
Hydrogen H [•] and chloride radical $\text{Cl}^\bullet$	1	1
Alkyl (CH_3 , CH_3CH_2 , etc.) or acyl ($\text{RC}=\text{O}$) groups	1	2
NO (bent)	1	2
Carbonyl groups (RCOO^-), CO, CN^- , RCN or RNC, ethers (ROR), sulfides (R_2S), and ketones ($\text{R}(\text{C}=\text{O})\text{R}$)	2	2
Lewis bases Cl^- , O^{2-} , S^{2-} , ammonia (NH_3), amines (NR_3), phosphines (PR_3), CO, N_2 , R_2S , RCN, and RNC	2	2
Alkenes ($\text{RCH}=\text{CH}_2$)	2 per double bond	2 per double bond
Nitrosyl group (NO) linear	3	3
Cyclopentadienyl, cp, (C_5H_5) or (C_5H_5^-)	5 per ring	6
Benzene (C_6H_6)	6 per ring (π donation)	6



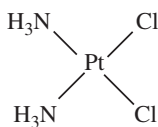
Covalent method
tetracarbonylnickel(0)
 $\text{Ni}(0) = 10 \text{ e } (4s^2 3d^8)$
 $4 \text{ CO} = 8 \text{ e }$
Total = 18 e



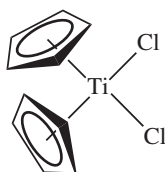
Covalent method
pentacarbonyliron(0)
 $\text{Fe}(0) = 8 \text{ e } (4s^2 3d^6)$
 $5 \text{ CO} = 10 \text{ e }$
Total = 18 e



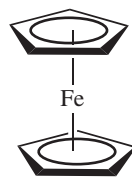
Covalent method
 $\text{Fe}_3(\text{CO})_{12}$ cluster
 $\text{Fe}_1(0) = 8 \text{ e }$
 $3 \text{ terminal CO} = 6 \text{ e }$
 $2 \text{ bridging CO} = 2 \text{ e }$
 $2 \text{ Fe-Fe bonds} = 2 \text{ e }$
Total = 18 e



Ionic method
Cis-dichlorodiammine-
platinum(II)
 $\text{Pt(II)} = 8 \text{ e }$
 $2 \text{ NH}_3 = 4 \text{ e }$
 $2 \text{ Cl}^- = 4 \text{ e }$
Total = 16 e



Ionic method
bis- η^5 -cyclopentadienyl-
dichlorotitanium(IV)
 $\text{Ti(IV)} = 0$
 $2 \text{ cp}^- = 12 \text{ e }$
 $2 \text{ Cl}^- = 4 \text{ e }$
Total = 16 e



Covalent method
ferrocene
 $\text{Fe}(0) = 8 \text{ e }$
 $2 \text{ cp} = 10 \text{ e }$
Total = 18 e

Ionic method
ferrocene
 $\text{Fe(II)} = 6 \text{ e }$
 $2 \text{ cp}^- = 12 \text{ e }$
Total = 18 e

Figure 1.6 Complexes obeying the 16-e and 18-e rule.

appear to be important for activity as the compound $(\text{MeCp})_2\text{TiCl}_2$ has been found to be biologically inactive. The corresponding molybdenum-containing metallocene, Cp_2MoCl_2 , appears to have a different mode of cytotoxic (antitumor) activity as it does not form stable DNA adducts under the same conditions as the titanocene. Although the molybdocene readily forms soluble adducts with the amino acid cysteine and physiologically prevalent sulfur-containing compounds such as glutathione, thiol derivatives of a molybdocene dichloride have been shown to lack antitumor activity. Researchers in the field have concluded that the biological chemistry of each of the metallocene dihalides is unique. The ongoing study of organometallic metallocenes as anti-tumor agents have been reported in various articles as referenced here [13]. The anticancer agent cisplatin, *cis*-dichlorodiammineplatinum(II), also obeys the 16-e rule.

Many organometallic complexes are clusters involving multiple metals that feature metal-metal bonds. The electrons in Me-Me bonds are counted by contributing one electron to each metal connected. Bridging ligands contribute

one half of their electrons to each metal center. Some simple examples in Figure 1.6 illustrate application of the rules.

Iron–sulfur clusters, such as those found in the enzyme nitrogenase (see Section 3.3), cannot be treated using the 16-e or 18-e rules. Other frameworks exist to treat large metal clusters and these have some utility in treating $[\text{Fe}_x\text{S}_y]^{n+}$ clusters. One method treats the number of metal atoms and the metal–metal bonds in a cluster according to the following formula [14]:

$$\Sigma \text{Valence electrons} = \# \text{ of cluster Me atoms} \times 18 - (\text{no. of metal – metal bonds} \times 2)$$

Applying this formula to the cubane $[(\text{Fe(II)})_4(\text{S-cys})_4(\mu_3\text{-S})_4]$ shown in Figure 1.7c results in the following electron count:

$$\Sigma \text{Valence electrons} = 4 \times 18 - (6 \times 2) = 60 \text{ electrons}$$

the so-called “magic number” for four metal atoms in a cluster.

Cluster magic numbers would occur as 48 e for a triangular clusters, 60 e for tetrahedral, 72 e for trigonal bipyramidal, 74 e for square pyramidal, 84 e

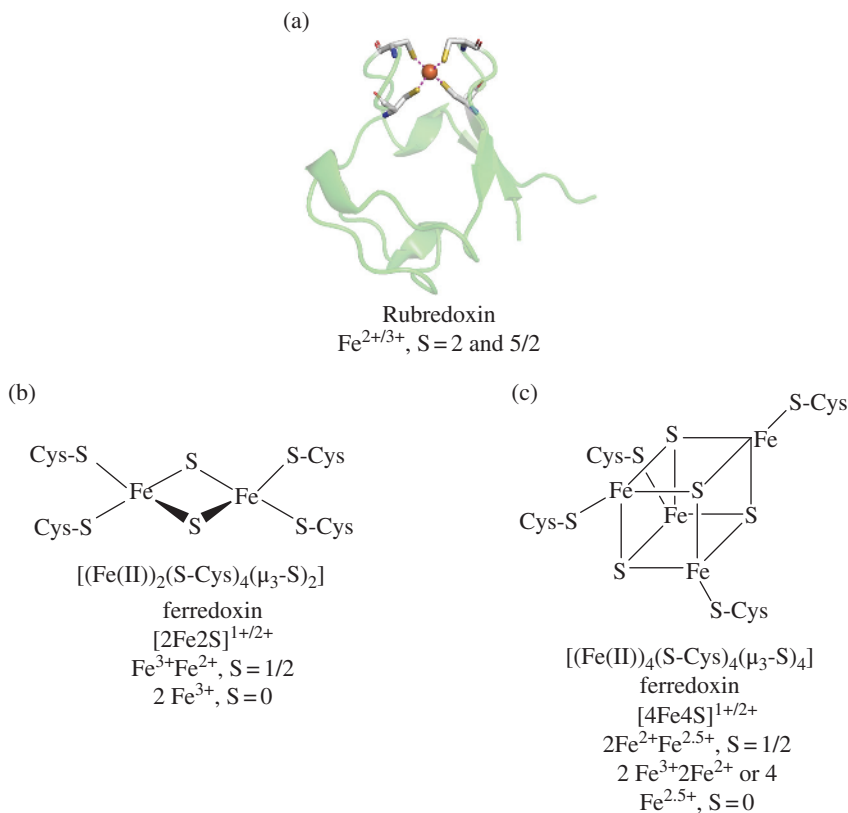


Figure 1.7 (a) Rubredoxin. (b) $[2\text{Fe}2\text{S}]$ cluster in ferredoxins. (c) $[4\text{Fe}4\text{S}]$ cluster in ferredoxins.

for octahedral complexes like $\text{Zr}_6\text{I}_{14}\text{C}$ or $[\text{Mo}_6\text{Cl}_{14}]^{2-}$ or 86 e for octahedral complexes such as $\text{Rh}_6(\text{CO})_{16}$ and $[\text{Os}_6(\text{CO})_{18}]^{2-}$, 90 e for trigonal prisms, and 120 e for cubic structures. Reference [14a] contains a more complete discussion of cluster valence electron counting.

For biological systems such as ferredoxins, problems arise when counting electrons by the valence electron method. This system assumes six Fe–Fe bonds within the tetrahedral iron–sulfur clusters but Fe–Fe bond distances within biological iron–sulfur clusters as found by X-ray crystallography often do not indicate Fe–Fe bonds. It is known that for the Fe_4S_4 cubane found in biological systems, oxidations are accompanied by increasing distortion of the cubane frame. Nevertheless, ^{57}Fe Mössbauer spectra indicate that the four iron atoms remain equivalent suggesting delocalization within the Fe–S framework. Most biological iron–sulfur clusters deviate substantially from the electron-counting rules for iron–sulfur clusters discussed here.

A 2014 Chemical Reviews publication by Yi Lu and coworkers includes a comprehensive study of iron–sulfur cofactors in proteins [15]. Iron–sulfur cluster-containing proteins occur in three broad categories: electron carriers, enzymatic catalysts, and gene regulators. As electron carriers iron sulfur proteins are divided into major groups:

- Rubredoxins, Rds, $[\text{1Fe4S}]$
- Low-potential ferredoxins, Fds, $[\text{2Fe2S}]$, $[\text{4Fe4S}]$, $[\text{3Fe4S}]$, $[\text{3Fe4S}][\text{4Fe4S}]$, and $[\text{4Fe4S}][\text{4Fe4S}]$
- High-potential Rieske proteins $[\text{2Fe2S}]$
- High-potential iron–sulfur proteins, HiPIPs, $[\text{4Fe4S}]$

In these systems, the Fe ions cycle between +2 and +3 oxidation states. Within the cluster, the ions exhibit Fe–Fe bonds and antiferromagnetic coupling between iron ions. Low and high electromotive potentials depend on the local conditions at the cluster's site within the protein, on the iron ion's oxidation state, pH, temperature, and other factors. Two iron–sulfur clusters found in ferredoxins are shown in Figure 1.7b and c. The reduction potentials of iron–sulfur clusters found in proteins vary greatly among different protein classes and within the classes themselves. Figure X.1.2 of reference [8] gives a graphical representation of the reduction potentials for various protein classes between –600 and +600 mV. Those for iron–sulfur clusters include

- $[\text{4Fe4S}]$ high-potential iron proteins (HiPIP) 93–458 mV
- $[\text{4Fe4S}]$ ferredoxins (Fdx) –645 to –410 mV
- $[\text{3Fe4S}]$ ferredoxins (Fdx) –130 mV
- $[\text{2Fe2S}]$ ferredoxins (Fdx) –428 to –239 mV

Rubredoxin is the simplest iron sulfur protein of 45–54 amino acids and with a molar mass of 6–7 kDa, mainly found in bacteria, archaea, and anaerobes. It

contains one mono-iron center in distorted tetrahedral geometry coordinated by four cysteines in two conserved protein segments with the sequence Cys-(Xxx)₂-Cys-Gly. In Figure 1.7a, the rubredoxin protein from *Pyrococcus furiosus* is visualized using protein data bank (PDB) 1BRF X-ray crystallographic data [16].

Ferredoxins mediate electron transfer in a great variety of organisms from bacteria to plants to humans. In plants, a ferredoxin is part of the Photosystem I electron cascade and a Rieske iron-sulfur ferredoxin plays a similar role in Photosystem II. In humans, ferredoxin-1 is involved in thyroid hormone synthesis. Bacterial proteins contain a great variety of iron-sulfur clusters that are evolutionarily ancient.

The iron-sulfur protein (ISP, [2Fe2S]) component of the cytochrome b(6)f and cytochrome bc₁ complexes, now called the “Rieske” ISP, was first discovered and isolated by John S. Rieske and co-workers in 1964 (in the cytochrome bc₁ complex). More information about the Rieske iron-sulfur protein (RISP) is found in section 7.5 and figure 7.26 (cytochrome b(6)f), section 7.6 and figure 7.30 (cytochrome bc₁), and section 7.9.2.1 and figure 7.50 (aconitase, [4Fe4S]) of reference [5]. Also in reference [5], see section 3.6 for a discussion of Mössbauer spectroscopy, table 3.2 for experimental findings, and section 3.6.4 for a descriptive example of nitrogenase enzyme’s FeMo cofactor studied by Mössbauer. Section 6.3 of reference [9] discusses iron-sulfur clusters in nitrogenase. Section 3.3 of this text discusses the complex iron-sulfur cofactors found in nitrogenase. Section 4.4.3.2 discusses the [4Fe4S] cluster found in the nitrite reductase enzyme that reduces nitrite ion to the ammonium ion.

Many systems in organometallic chemistry involve the activation of hard-to-oxidize alkanes or other organic moieties. Several metalloenzymes discussed in reference [5] are efficient alkane oxidizers. Section 7.4 of reference [5] discusses the monooxygenase cytochrome P450 that inserts one atom of a dioxygen molecule into a wide variety of organic substrates. The other oxygen atom of the dioxygen molecule is converted to water during the P450 catalytic cycle. Methane monooxygenase (section 7.9.3.1 of reference [5] and Section 3.5 of this text) catalyzes the conversion of the hardest-of-all-to-oxidize alkanes, methane, CH₄, to CH₃OH. Biomimetic bioinorganic researchers model the active centers of these metalloenzymes as small molecules, to learn more about the metalloenzyme’s catalytic cycle, or to design efficient organometallic catalysts for industrial processes.

Researchers studying the metalloenzyme hydrogenase would like to design small compounds that mimic this enzyme’s ability to reversibly reduce protons to H₂ and H₂ to 2H⁺, using an active center that contains iron and nickel (see Section 4.1). Cobalamins (vitamin B₁₂ and its derivatives) contain an easily activated Co-C bond that has a number of biological functions, one of which is as a methyl transferase, 5-methyltetrahydrofolate-homocysteine methyltransferase (MTR). This enzyme converts homocysteine (an amino acid that has one more CH₂ group in its alkyl side chain than cysteine) to methionine as methylcobalamin is converted to cobalamin.

1.7 INORGANIC CHEMISTRY CONCLUSIONS

The preceding brief review of inorganic chemistry has been oriented towards questions that will arise in the discussion of bioinorganic systems which follow. Reference [2] lists a number of inorganic chemistry texts that give much more detail than this brief introduction. The bioinorganic chemistry texts referenced earlier in this chapter and here are good sources for answering the additional questions sure to arise in studying the behavior of metals in biological systems [1, 4–9]. It is important to keep in mind that metal behavior in the biological milieu will be influenced greatly by the surroundings. Metal–ligand systems existing in thermodynamic equilibrium and slow to react to changing cellular or noncellular dynamics will not long endure. Therefore, most of the metallo-enzyme systems to be described in following chapters contain metals in distorted and changeable ligand fields. These systems will continue to challenge the ingenuity of inorganic and bioinorganic chemists attempting to understand, modify, model, or design synthetic substitutes for them.

1.8 INTRODUCTION TO BIOCHEMISTRY

Biochemistry concerns itself with the study of life at the molecular level. In introductory biochemistry, one becomes familiar with amino acids (aa) and how they combine (polymerize) to become peptides and proteins. Proteins fold into three-dimensional shapes and become enzymes, the catalysts of biochemical reactions, structural materials in muscles and tendons, or small molecule carriers such as the dioxygen-carrying metalloproteins myoglobin and hemoglobin. At least one-third of all proteins and enzymes contain metals and are known as metalloproteins and metalloenzymes. Some of these will be the bioinorganic chemistry topics treated in Chapters 3 and 4 of this text.

A second major area of biochemical importance concerns study of nucleotide polymerization to produce ribonucleic acids (RNA) and deoxyribonucleic acids (DNA). Genes, the basis for inherited characteristics, are contained in DNA double-helical sections incorporated into coiled and supercoiled DNA structures. Genomics, the study of the total genetic assemblage of any species, is now well-known topic to all, especially with the announcement of the sequencing of the human genome in 2001. More information on this topic is given in Section 1.10.5.

The goal of this chapter is to provide a basic understanding of the structure of and biochemical processes undertaken by proteins and nucleic acids. The field of molecular biology integrates biochemical and genetic functions in an attempt to understand the workings of the cell at the most basic level. This knowledge will be needed by students, teachers, and researchers wishing to understand how genes operate (or misfire in disease states) to replicate, transcribe, and translate themselves into proteins and enzymes and how these entities integrate themselves into cells that function to sustain life. Goals of particular interest to bioinorganic chemists are to use biochemistry and

molecular biology to replicate the catalytic activity of metalloenzymes or to design compounds (many of them containing inorganic atoms) that will be able to detect, diagnose, and treat disease.

1.9 PROTEINS

1.9.1 Amino Acid Building Blocks

Polypeptides are formed through the polymerization of any combination of the 20 naturally occurring amino acids (aa). In humans, 10 of these amino acids are essential (cannot be synthesized by the body and must be ingested in the diet). The 10 essential amino acids are arginine (arg, R), histidine (his, H), isoleucine (ile, I), leucine (leu, L), lysine (lys, K), methionine (met, M), phenylalanine (phe, F), threonine (thr, T), tryptophan (trp, W), and valine (val, V). Relatively short amino acid chains, called polypeptides, have important hormonal (control) functions in biological species. Proteins are classified as polypeptide chains exceeding 50 amino acids in length while enzyme molecules usually contain more than 100 amino acid residues. Amino acids contain a central carbon, called the α carbon, to which four substituent groups are attached:

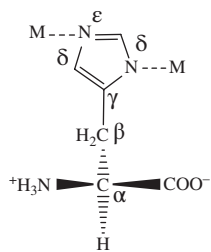
- The amine group ($-\text{NH}_2$)
- A carboxylic acid group ($-\text{COOH}$)
- A hydrogen atom ($-\text{H}$)
- A side-chain ($-\text{R}$) group unique to each amino acid

The amino acids (aa) structures in Figure 1.8 are those most commonly found as metal ligands in biological species. Points of metal attachment are shown in the figure. Many other aa residues near the metalloprotein's active site and actively involved in structure or catalytic mechanism will be discussed as they occur in future sections. The three- and one-letter common abbreviations for amino acids are shown in the figure. The amino acids are shown in their zwitterion form at pH 7 in which they have their COOH group ($\text{pK}_a = 2.35$) in the COO^- form and their NH_2 group ($\text{pK}_a = 9.69$) in the NH_3^+ form [17]. Side-chain R groups will be protonated or deprotonated at physiological pH (~ 7.4) based on the pK_a valued of the side chain. The Henderson–Hasselbalch equation (1.16) defines the relationship between the pH of a system and the pK_a of any acid in the system:

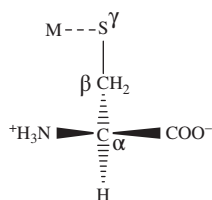


For a weak acid such as those found in amino acids, the equilibrium will lie to the left.

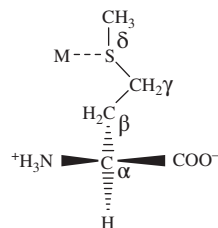
$$\text{K}_a = \frac{[\text{R} - \text{COO}^-][\text{H}^+]}{\text{R} - \text{COOH}} \quad (1.13)$$



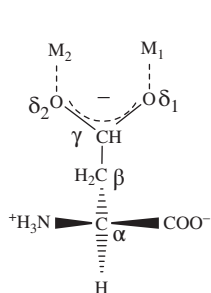
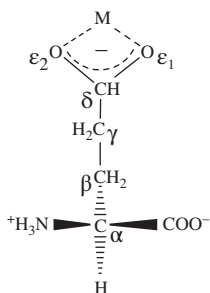
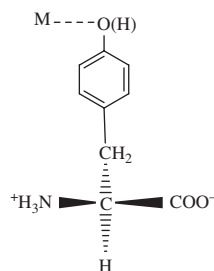
Histidine, His, H



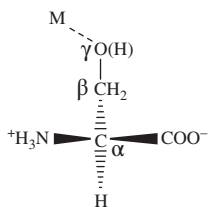
Cysteine, cys, C



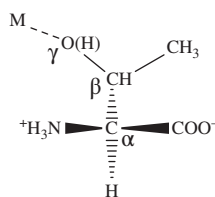
Methionine, Met, M

Aspartic acid, aspartate,
Asp, DGlutamic acid, glutamate,
Glu, E

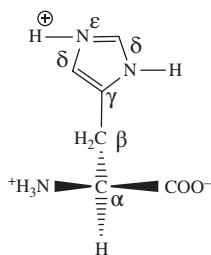
Tyrosine, Tyr, Y



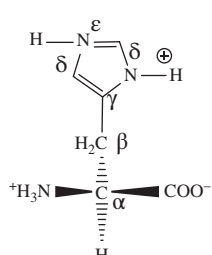
Serine, Ser, S



Threonine, Thr, T



or

Histidine conjugate acid $pK_a \sim 6$ **Figure 1.8** Possible amino acid-metal complexation structures found in metalloproteins.

For aspartic acid or glutamic acid, $K_a \approx 10^{-4}$ and for histidine $K_a \approx 10^{-6}$.

$$pK_a = -\log K_a \quad (1.14)$$

For any acid, where HA is the acid and A^- is the conjugate base:



and the Henderson–Hasselbalch equation will be

$$pH = pK_a + \log \frac{[A^-]}{[HA]} \quad (1.16)$$

If one applies Eq. (1.16) to aspartic acid, one finds that the ratio of conjugate base to acid is 10^3 , that is most aspartic acid molecules are deprotonated at physiological pH and ready to complex with positive metal ions. Considering histidine ($pK_a \sim 6$) as a ligand, many N atoms with two unpaired electrons will be available at two positions for bonding to metal ions. Tyrosine, serine, and threonine with $pK_a \sim 9$ will be much less deprotonated at physiological pH. Conjugate acid–conjugate base chemistry is not the only consideration for amino acid–metal bonding in metalloproteins. One must also consider hard–soft acid–base bonding effects (see Table 1.2). Group I and II metals in general will prefer hard base ligands – alcoholic, phenolic, carbonyl, or carboxyl – intermediate metal ions such as Fe^{2+} will prefer intermediate softness as found in amines and imidazoles, but even soft bases such as the sulfur atoms of cysteine and methionine.

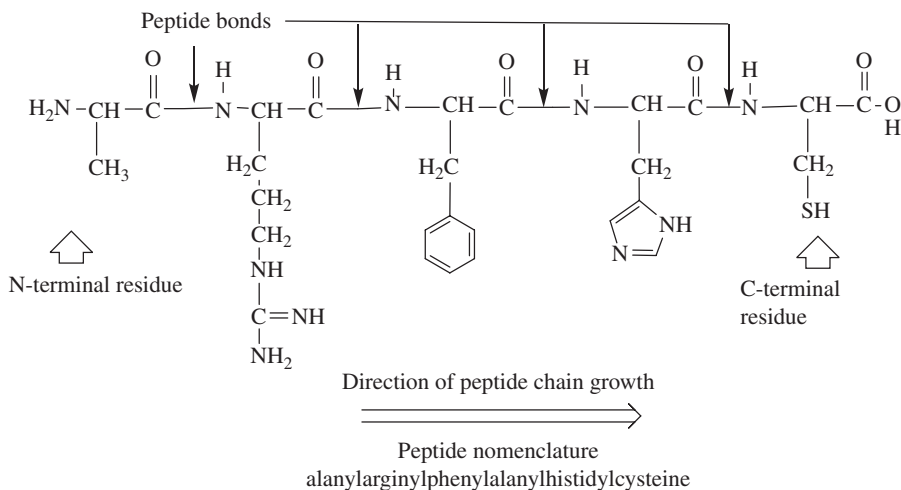
All amino acids found in proteins are called α amino acids because the amine group is bonded to the α carbon. All amino acids found in proteins are L-stereoisomers with respect to the α carbon (except glycine whose R group is H) although D-stereoisomers are found in bacterial cell walls and some peptide antibiotics.

In nomenclature for amino acids as ligands, the R group atoms are labeled with Greek letters starting with β for the first atom attached to the α carbon followed by γ , δ , and ϵ . The common bioinorganic ligands from amino acid side chains have their atoms labeled in this manner in Figure 1.8. The labeling becomes important in identifying the ligand atom that may be the metal-binding site in metalloproteins. In histidine, metals bind at either the δ or ϵ positions. In superoxide dismutase, one his ligand bridges two metal ions with Zn(II) coordinated at the δ nitrogen and Cu(II) at the ϵ nitrogen. For carboxylic acid side chains, metals may bind in monodentate mode (Asp-M), in bidentate chelate mode, or with the carboxylic acid oxygens bonding to two different metal ions (Asp- M_1M_2). Some examples are shown in Figure 1.8, however these modes are not all inclusive. Metal ions exhibit a great variety of binding partners and bonding modes in biological species, not only to amino acid side- and main-chain atoms, but also to water, sugars, atoms of DNA and RNA, and other convenient atoms or ions in the physiological milieu.

1.9.2 Protein Structure

All proteins have at least three levels of structure – primary, secondary, and tertiary. Proteins with more than one polypeptide chain – hemoglobin and nitrogenase are examples – also possess quaternary structure. The primary, secondary, tertiary, and quaternary structure of proteins controls their three-dimensional shape that in turn affects their activity. The primary structure of proteins is formed by a condensation reaction resulting in a peptide bond as shown, along with the direction of peptide chain growth, in Figure 1.9a. In this example, the zwitterions of cysteine and isoleucine combine to form the peptide bond via a condensation reaction between the α -carboxyl group of one amino acid and the α -amino group of another amino acid. A water molecule is eliminated as each successive peptide bond is formed. The peptide bond is rigidly planar and has a bond length intermediate between a single and a double bond. This rigid bond forms the backbone of the protein. Other single bonds in a polypeptide chain are flexible and can and do rotate along the chain as shown in Figure 1.9b.

(a)



(b)

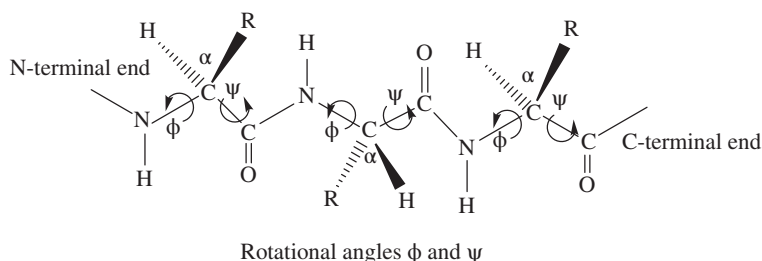


Figure 1.9 (a) Primary structure of a peptide chain. (b) Possible rotation angles along a peptide.

The N-terminal end of a polypeptide chain, with its free amino group, is conventionally known as beginning of the chain, while the last amino acid, with free carboxyl group, is the C-terminal end of the chain (see Figure 1.9b).

A protein's secondary structure arises from the formation of intra- and intermolecular hydrogen bonds. All carboxyl group oxygens and amine hydrogens of a polypeptide participate in H-bonding. Protein secondary structure also derives from the fact that while all C—N bonds in peptides have some double bond character and cannot rotate, rotation about the C_α—N and C_α—C bonds takes place. These rotations are affected by peptide R groups that have differing space and charge constraints. The illustration in Figure 1.9b indicates the C_α—N bond angle ϕ (phi) and C_α—C bond angle ψ (psi) about which rotation is possible.

Two types of protein secondary structure are the α helix and β -pleated sheet, and their representations can be seen in Figure 1.10

The right-handed α helix occurs in globular proteins and is formed by intramolecular hydrogen bonds between the carboxyl group oxygen of one amino acid and the amine hydrogen of the fourth amino acid away. This helix completes one turn every 3.6 residues and rises approximately 5.4 Å with each turn. Other known types of protein helices are the π helix and the 3–10 helices. The left-handed α helix is known but not found in protein structures. The β -pleated sheet structure occurs in fibrous as well as globular proteins and is formed by intermolecular hydrogen bonds between a carboxyl group oxygen of one amino acid and an amine hydrogen of an adjacent polypeptide chain. Parallel β -pleated sheets form when the adjacent polypeptide chains are oriented in one direction (from N-terminal to C-terminal end or vice versa). Antiparallel β -pleated sheets form when the adjacent polypeptide chains are traveling in opposing directions (one chain N-terminal to C-terminal and the other C-terminal to N-terminal). Representations are shown in Figure 1.10. The secondary structure motifs of α helix and β sheet are joined by unstructured areas called loops or coils, also shown in Figure 1.10. Many websites feature information and graphic illustrations on proteins and peptides. Some examples are

<https://www.khanacademy.org/science/biology/macromolecules/proteins-and-amino-acids/a/orders-of-protein-structure>

https://www.particlesciences.com/docs/technical_briefs/TB_8.pdf

https://www.expasy.org/proteomics/protein_structure

The last website (expasy.org) is the Swiss Institute of Bioinformatics portal that provides access to a comprehensive set of life science databases and software tools.

The tertiary structure of proteins arises from the interactions of the various R groups along the polypeptide chain. Some of the forces responsible for the tertiary structures include van der Waals forces, ionic bonds, hydrophobic bonds, and hydrogen bonds. Usually hydrophilic R groups arrange themselves on the exterior of the tertiary structure, and interact with the aqueous environment while hydrophobic R groups usually orient themselves on the interior of the protein's tertiary structure, and exclude water. Most proteins are defined as

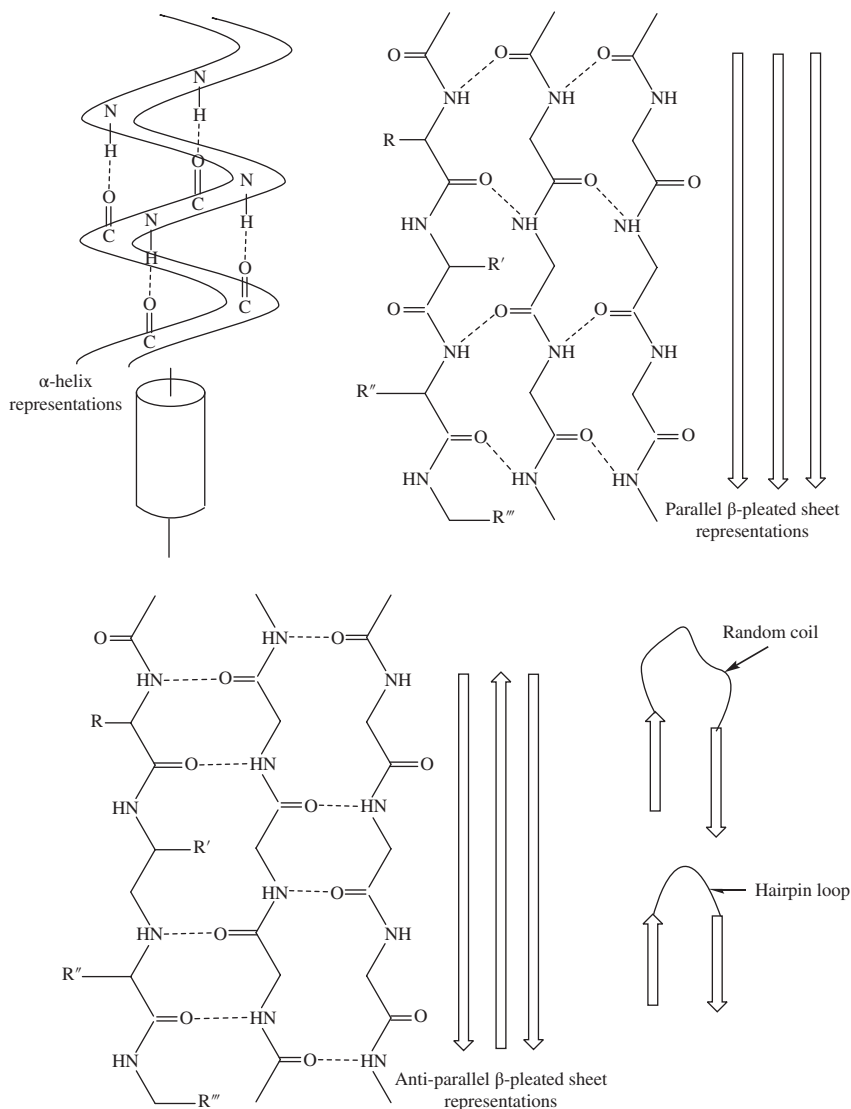


Figure 1.10 Hydrogen bonding to form protein α -helical and β -pleated sheet secondary structures.

being amphipathic – that is, they contain both nonpolar and polar side chains and environments. The nonpolar groups or areas will align themselves with nonpolar or hydrophobic molecules while the polar groups or areas will align themselves with polar or hydrophilic molecules. One bond that often helps establish tertiary structure is a disulfide bond between two cysteine side-chain groups.

Most proteins contain more than one polypeptide chain. The manner in which these chains associate determines quaternary structure. Binding involves

the same types of noncovalent forces mentioned for tertiary structure – van der Waals forces, hydrophobic and hydrophilic attractions, and hydrogen bonding. However, the interactions are now interchain rather than intrachain (tertiary structure determination). The quaternary structure of hemoglobin (four almost identical subunits) will be discussed in Section 3.1.

Secondary structure, as well as tertiary and quaternary structure of proteins, is intimately dependent on the primary sequence of amino acids in the chain. In fact, the manner in which proteins fold into their ultimate structures in biological species is a subject of much research and continued uncertainty even though much is known. Predicting protein folding is an enormous challenge. Most proteins contain dozens or hundreds of amino acids, so there is an astronomical number of ways in which these might be arranged into a compact, folded structure. While only a tiny fraction of the possible folds – perhaps a 1 000–10 000 – are found in natural proteins, the challenge is to deduce the best fit of a particular protein sequence to one of these folds. This is called the protein-threading problem. Traditionally, the problem is tackled by assuming that each amino acid prefers to be surrounded by others of a specific kind, and then to look for the best compromise between the needs of all the amino acids. Success using this approach depends on how well we know what the amino acids prefer. Some publications with reference to protein folding are included here [18]. A 2017 *Chemical and Engineering News* article reviewed protein folding history and its players and practitioners. <https://cen.acs.org/articles/95/i31/Protein-folding-Much-intricate-thought.html>. One facet of the protein folding problem involves protein chaperones – proteins that are intimately involved in a protein's life cycle. The research surrounding this aspect of protein science has been given the name proteostasis (combining the words protein and homeostasis – the maintenance of a stable environment in physiological systems). Proteostasis means regulation of all cellular proteins to maintain the health of both the cellular proteome and the whole organism. The field of proteomics can be defined as the analysis of *all* the proteins expressed at a given time or in a given location rather than the study of one isolated protein. In describing the behavior of numbers of proteins functioning together in an organelle, the proteome becomes defined as the protein complement of the organism's genome. One can think of the human proteome as the universe of proteins encoded by gene sequences in the human genome.

The amino acid composition of a protein can be determined by cleaving all peptide bonds and identifying the constituent amino acids. The sequence of a given protein is determined by using various methods to cleave only selected peptide bonds and then assembling the information to deduce the amino acid sequence. Proteins with multiple subunits are usually broken down into individual subunits (denatured with heat or chemical reagents) before composition and sequencing analyses are carried out. Characterization of proteins on the basis of size and/or charge can be accomplished by a number of methods. These methods are described in section 2.2.3 of reference [5].

1.9.3 Protein Function, Enzymes, and Enzyme Kinetics

Protein functions and interactions are infinitely varied in biological species – one of the major problems associated with complete classification of any proteome. Proteins may transport substances – myoglobin and hemoglobin, discussed in Section 3.1 – and transport oxygen and carbon dioxide in mammalian blood. Proteins called enzymes catalyze necessary biochemical reactions. The active site of an enzyme contains those amino acids that come in direct contact with the substrate (substance acted upon) and bind it. Some enzymes require nonprotein molecules or groups, called cofactors, for their activity. Cofactors may be metal ions such as $\text{Fe}^{2+,3+}$, $\text{Cu}^{1+,2+}$, $\text{Ni}^{1+,2+,3+}$ or $\text{Co}^{1+,2+,3+}$, etc., or be organic molecules such as hemes in myoglobin, hemoglobin (Section 3.1, Figure 3.2), a nitric oxide synthase (Figures 4.26–4.28), or cytochromes (Figures 3.7, 3.8, and 3.12), and Figure 3.17 shows the FeMo cofactor (M-cluster) of nitrogenase.

These cofactor molecules are sometimes called *coenzymes*. An enzyme without its cofactor, called an *apoenzyme*, usually will be inactive in its catalytic role or exhibits greatly reduced catalytic activity. An active enzyme with its cofactor is called a *holoenzyme*. Enzymes that require metal ions as cofactors are called *metalloproteins* or *metalloenzymes*. Essential cofactors and coenzymes that must be ingested in the diet include water-soluble vitamins such as the cobalt-containing vitamin B_{12} . *Isozymes* may be structurally different forms of the same enzyme or oligomeric proteins with differing sets of subunits. Isozymes may also be defined as multiple forms of an enzyme whose synthesis is controlled by more than one gene. Usually isozymes catalyze similar types of reactions. *Allosteric* enzymes have their activity modulated by the binding of a second molecule, whereas nonallosteric enzymes do not. Proteins maintain biological structures; that is, the protein collagen constitutes part of fibrous connective tissues in skin, bone, tendon, cartilage, blood vessels, and teeth. Proteins may facilitate movement; that is, the proteins actin and myosin mediate muscle contraction. Proteins serve as storage compounds; that is, the protein ferritin stores iron in the human body. *Immunoglobulins* are proteins that react with and neutralize foreign compounds (antigens) in the body. The protein hormone insulin regulates glucose levels in the blood.

Many enzymes are named in classes by attaching “ase” to the type of reaction they catalyze. ATPase enzymes catalyze the hydrolysis of adenosine triphosphate (ATP) to adenosine diphosphate (ADP). Oxidoreductases catalyze oxidation–reduction reactions – alcohol dehydrogenase catalyzes the oxidation of an alcohol to an aldehyde. Transferases catalyze the transfer of groups from one molecule to another. Hydrolases catalyze the breaking of covalent bonds using water, e.g. peptidase hydrolyzes a peptide bond. Lyases either remove a group by splitting a bond and forming a double bond or add a group to a double bond to form a single bond; for example, decarboxylase removes a carboxyl group to form carbon dioxide. Isomerases catalyze internal atom rearrangements in a molecule. Ligases catalyze the formation of covalent bonds.

The general theory of enzyme kinetics is based on work of L. Michaelis and M. L. Menten later extended by G. E. Briggs and J. B. S. Haldane [17a]. The basic reactions (E = enzyme, S = substrate, and P = product) are shown in Eq. (1.17)



Assuming that the reactions are reversible and that a one-substrate enzyme-catalyzed reaction is being studied, one can derive the Michaelis–Menten rate Eq. (1.18):

$$V = \frac{V_{\max} [S]}{[S] + K_m} \quad (1.18)$$

where V is the initial rate for first-order breakdown of the enzyme–substrate ($[ES]$) complex into enzyme (E) and product (P); $V_{\max}$ is the maximum reaction rate for a given concentration of enzyme in the presence of saturating levels of substrate; $[S]$ is the substrate concentration; and K_m is the Michaelis constant, the concentration of substrate required to achieve one half of the enzyme's maximal velocity. Equation (1.18) applies to single-substrate reactions at a constant enzyme concentration. Only nonallosteric enzymes – those not dependent on binding of a second molecule in addition to the substrate – are treated by the Michaelis–Menten rate equation. When the substrate concentration is low, the rate of reaction (V) increases in direct proportion to the substrate concentration (first-order kinetics). When the substrate concentration becomes high enough to saturate all available enzyme active sites, the reaction rate becomes constant (zero-order kinetics). The graphical representation of this behavior is shown in Figure 1.11.

The Lineweaver–Burk plot uses the reciprocal of the Michaelis–Menten equation in the form of the equation of a straight line, $y = mx + b$, having the form shown in Eq. (1.19).

$$\frac{1}{V} = \left(\frac{K_m}{V_{\max}} \right) \frac{1}{[S]} + \frac{1}{V_{\max}} \quad (1.19)$$

Plotting $1/V$ versus $1/[S]$, one obtains a straight line having a slope of $K_m/V_{\max}$ with a y-axis intercept of $1/V_{\max}$ and an x-intercept of $-1/K_m$ as shown in Figure 1.12.

Lineweaver–Burk plots of enzyme activity in the presence of an inhibitor can distinguish the type of inhibitor. Competitive inhibitors have a molecular structure similar to the substrate and will alter K_m but not $V_{\max}$ because they compete with the substrate for binding at the enzyme's active site but do not change the enzyme's affinity for substrate. Noncompetitive inhibitors bear no structural similarity to the substrate but bind the free enzyme or enzyme–substrate complex at a site other than the active site. They reduce

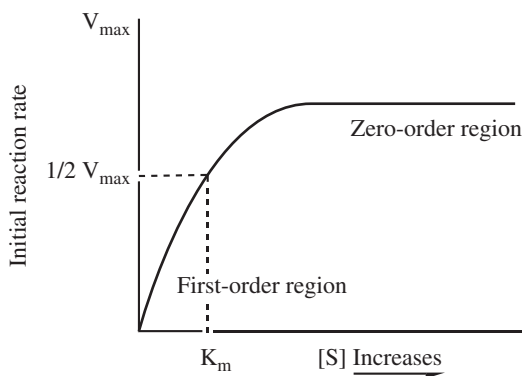


Figure 1.11 Graphical representation of the Michaelis–Menten equation for nonallosteric enzymes.

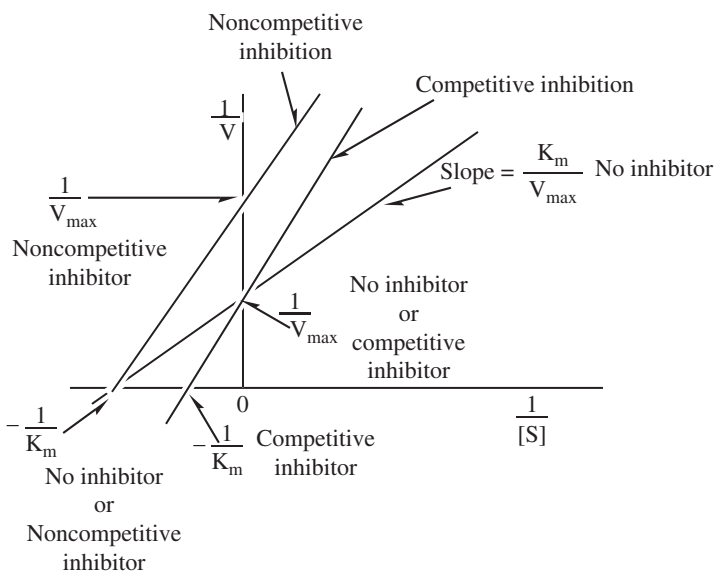


Figure 1.12 Lineweaver–Burk plot for no inhibitor, competitive and noncompetitive inhibition.

the enzyme's affinity for substrate and decrease $V_{\max}$ while not changing K_m . The lines indicating competitive and noncompetitive inhibition are included in Figure 1.12.

1.10 DNA AND RNA BUILDING BLOCKS

The sequence of amino acids in a given protein determines its structure. The amino acid sequence is controlled ultimately by sequences of genes that are made up of deoxyribonucleic acid (DNA). DNA is composed of repeating

units of the nitrogenous bases adenine and guanine (purines) and thymine and cytosine (pyrimidines) linked to the five-carbon cyclic sugar β -D-2-deoxyribose. In turn, the sugars are connected by 3'-5' phosphodiester bonds forming the linkages in the DNA polymeric chain. In nucleic acid numbering, positions on the sugar ring carry a prime (') to distinguish them from the numbering of nitrogenous base positions. DNA chains form the well-recognized double helix through hydrogen bonding of adenine to thymine (A-T pairs) and guanine to cytosine (G-C pairs). More about the forms of the double helix is found in the next section. The complete nitrogenous base, sugar, and phosphate unit is called a *nucleotide*. Nitrogenous base plus sugar moiety are called *nucleosides*. Ribonucleic acids (RNA) resemble DNA in that nucleoside monophosphates are joined through phosphodiester bonds. RNAs differ in that the sugars are β -D-ribose units and the pyrimidine uracil is found in place of thymine. Molecular structures and nomenclature for nitrogenous bases, nucleosides, and nucleotides are delineated in Table 1.9.

Common metal-binding sites on nucleobases are indicated by arrows in Table 1.9. Many of these are endocyclic (in the ring) nitrogen atoms that have more available lone pair characteristics than exocyclic (exterior to the ring) amino groups whose lone pairs are delocalized into the ring by resonance. Experimental and theoretical studies indicate that the purine N7 sites are the best nucleophiles among all the possible metal-binding sites [7]. They are exposed to solvent in the B DNA major groove and are not involved in Watson-Crick base-pairing. The N7 sites of adenine and especially guanine are the preferred platinum ion coordination location for the platinum-containing anticancer drugs. Hard metal ions such as Na^+ , K^+ , and Mg^{2+} prefer the hard negatively charged oxygen atoms of the phosphodiester groups with the result that enzymes requiring ATP often require Mg^{2+} as well [7]. Various metal chelates of the α , β , and γ phosphate groups have been identified by X-ray crystallography and ^{31}P nuclear magnetic resonance (NMR).

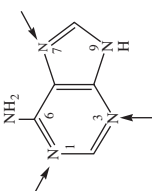
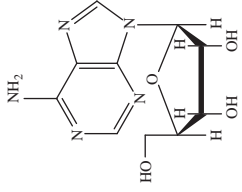
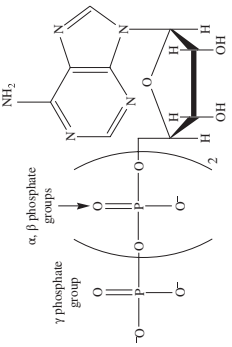
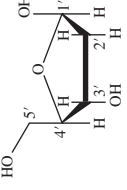
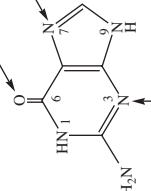
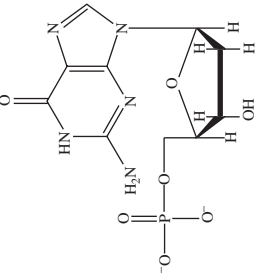
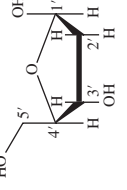
1.10.1 DNA and RNA Molecular Structures

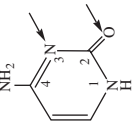
The nitrogenous base pairs in DNA (A-T and C-G) link through the formation of hydrogen bonds within the double-stranded (ds) structure for DNA. One type of hydrogen bonding is shown in Figure 1.13. Watson-Crick pairing is shown, however other hydrogen bonding conformations are found. Some of these are known as Hoogsteen pairs, reversed Hoogsteen pairs, and reversed Watson-Crick pairs. Note that in Watson-Crick base-pairing A-T pairs form two hydrogen bonds while C-G pairs form three. RNA usually exists in the single-stranded (ss) form but may fold into secondary and tertiary structures through the formation of specific base-pairings (A-U and C-G). Double-stranded RNA is also known.

Double stranded DNA exhibits complementarity in forming the double helix. The complementary sequences have opposite polarity; that is, the two chains run in opposite directions as in the following illustration:



TABLE 1.9 Nitrogenous Bases, Nucleosides, Nucleotides, and Sugars Found in DNA and RNA

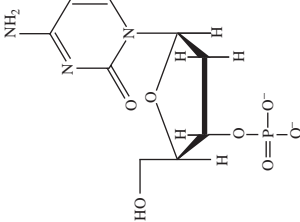
Base ^a	Nucleoside	Nucleotide Example	Sugar
 Adenine, A	Adenosine (RNA) or deoxyadenosine (DNA) 	Adenosine triphosphate (ATP) 	β-D-2-deoxyribose^b in DNA or β-D-ribose^c in RNA 
 Guanine, G	Guanosine (RNA) or deoxyguanosine (DNA) 	Deoxyguanosine 5'-monophosphate (5'-dGMP) 	β-D-2-deoxyribose^b in DNA or β-D-ribose^c in RNA 



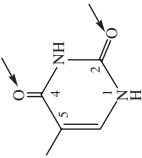
Cytosine, C

Cytidine (RNA) or
deoxycytidine (DNA)

Deoxycytidine
3'-monophosphate
(3'-dCMP)



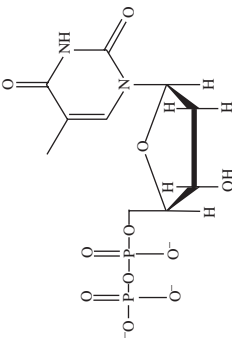
β -D-2-deoxyribose^b in
DNA
or
 β -D-ribose^c in RNA



Thymine, T
5-methyl uracil

Deoxythymidine (DNA)

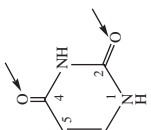
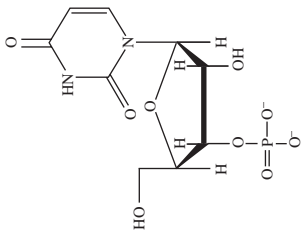
Deoxythymidine 5'-diphosphate
(5'-dTDP)



β -D-2-deoxyribose^b in
DNA

continued

TABLE 1.9 Continued

Base ^a	Nucleoside	Nucleotide Example	Sugar
 Uracil, U	Uridine	Uridine 3'-monophosphate (3'-UMP)	β -D-ribose ^c in RNA
			

^a Arrows indicate common metal-binding sites.^b H at 2' position.^c OH at 2' position.

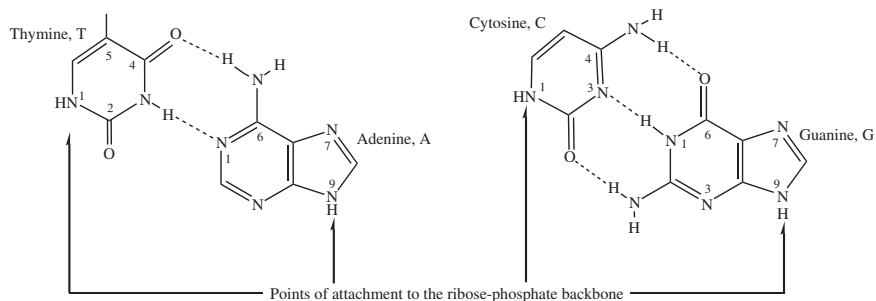


Figure 1.13 Watson-Crick base-pairing in DNA.

This is described as an antiparallel arrangement. This arrangement allows the two chains to fit together better than if they ran in the same direction (parallel arrangement). A sequence illustrating complementarity is shown in Figure 1.14.

In forming the double-helix polymeric DNA structure, the two sugar-phosphate backbones twist around the central stack of base pairs, generating a major and minor groove. Several conformations, known as DNA polymorphs, are possible. The classical form, B DNA, exhibits well-defined major and minor grooves; C_2' -endo deoxyribose sugar pucker (see Figure 1.15) and parallel stacked base pairs perpendicular to the helix axis. A and B DNA form right-handed helices, while Z DNA forms a left-handed helix. Usually, DNA adopts the B conformation; but in high salt or organic solvent conditions, the Z conformation may be found. Sequences containing alternating GC nucleotides favor the Z DNA conformation. Several different base-sequence-dependent DNA polymorphs may exist within a given genome. Double stranded RNA or RNA-DNA hybrids normally form the A DNA conformer. DNA conformations are dependent on a number of parameters, including the type of pucker found in the nonplanar ribose sugar ring, the syn or anti conformation of the nucleobase relative to the sugar moiety and the orientation about the C_4' - C_5' bond (Figure 1.15, as adapted from references [19a, b]).

The differing conformations of the ribose have been named with respect to that ring atom which puckers out of the plane given by the other ring atoms. The most prominent conformations are C_2' -endo in B-helices and C_3' -endo in A-helices. At room temperature, both conformers are in a dynamic equilibrium. Intermediates between C_2' -endo and C_3' -endo are found in several (time-averaged) structures obtained from X-ray crystallography or NMR. When changing the pucker phase angle from 0 to 360° , one steps through all possible conformations of the ribose ring. This pseudo-rotation cycle can be depicted as a conformational wheel such as that found in figure 2.8 of reference [19a]. Figure 9.3 of reference [19a] shows a view of C_2' -endo and C_3' -endo conformations observed in A- and B-type polynucleotides.

The distance between two subsequent base pairs along the helical axis is called helical rise (h). The pitch (P) is the length of the helix axis for one complete helix turn. The turn angle per nucleotide or twist angle (t) is given by

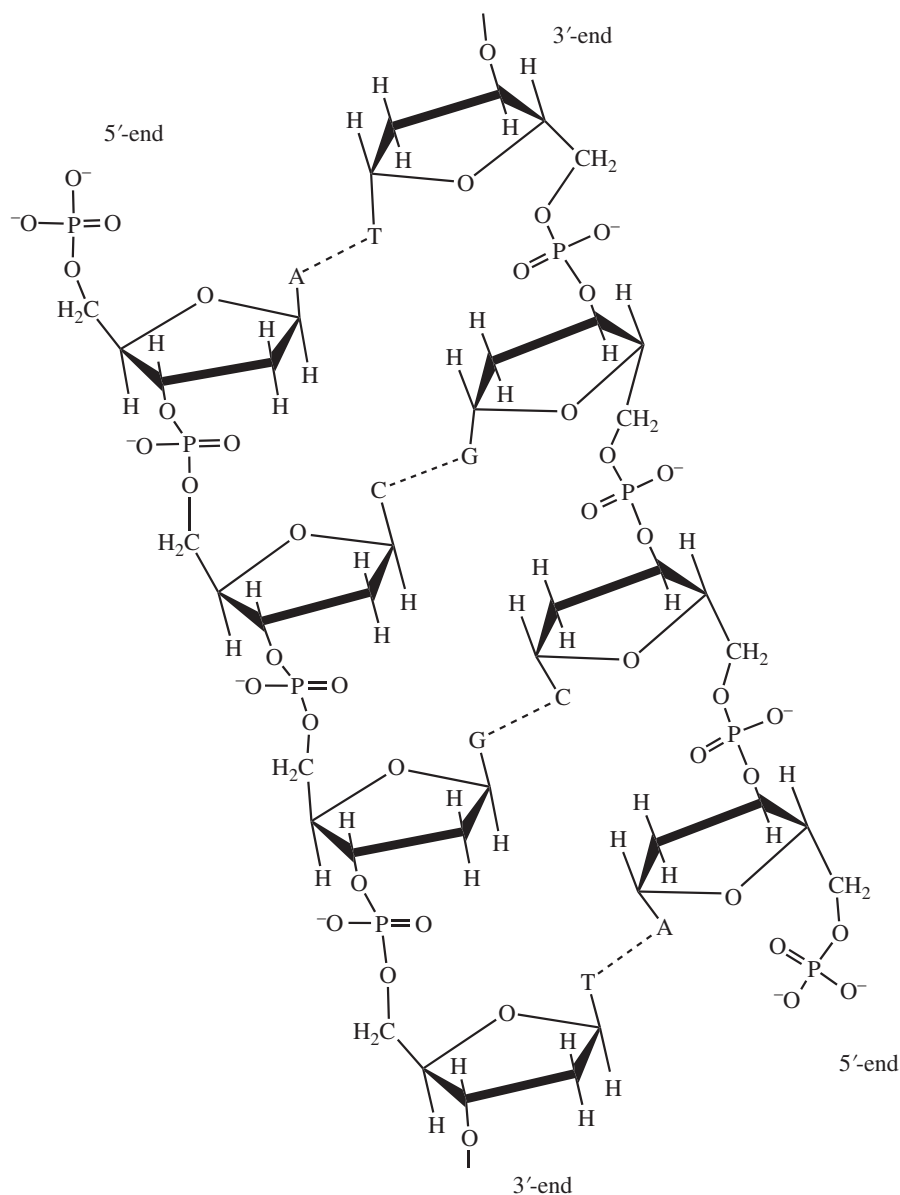
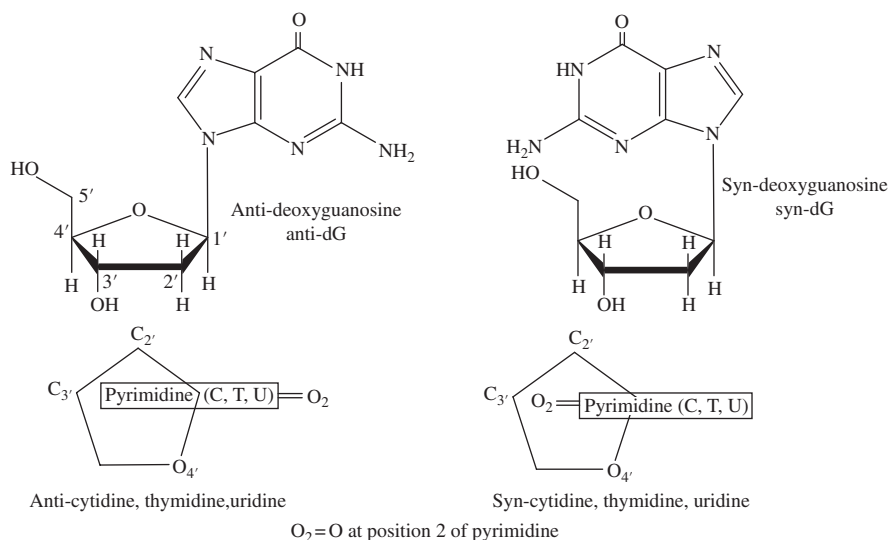


Figure 1.14 Complimentary antiparallel double-stranded (ds) DNA (dsDNA).

360° /number of nucleotides per turn. Data describing some properties of A, B, and Z DNA structure is found in Table 1.10 as adapted from table 1.10 of reference [19b].

In the nucleus of eukaryotic cells, a compacted DNA is formed by winding in a shallow, left-handed superhelix around a group of eight histone proteins to form nucleosome core particles. The core particles are further organized into



Definition of torsion angles about the $C_4'-C_5'$ bond looking in the direction $C_5' \rightarrow C_4'$

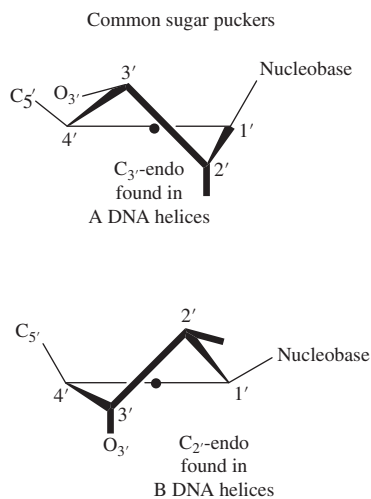
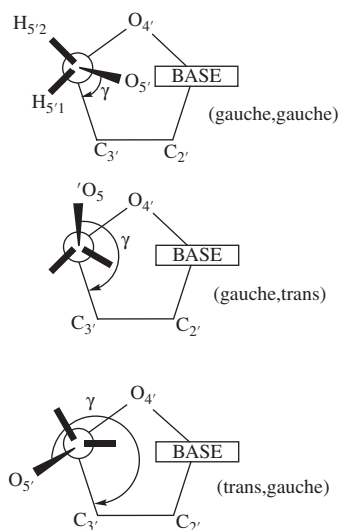


Figure 1.15 Orientations found in DNA helices. *Source:* adapted with permission from figure 2.11 of Saenger [19a]; copyright 1984 Springer-Verlag New York; and figure 1.22 A, B, C of Cowan [19b]. Copyright 1997, Wiley-VCH.

chromatin through interaction with additional histone proteins. Chromatin is anchored to a scaffold of proteins within the chromosomes. In order for replication, transcription, or translation to take place, supercoiled DNA must unravel and the DNA strands separate, at least temporarily. It is thought that the unraveling process, or some version of it, may be necessary before platinum

TABLE 1.10 Some Properties of A DNA, B DNA, and Z DNA

Property	A DNA	B DNA	Z DNA (The Repeat Unit is the Dimer – G C-)
Helix handedness	Right	Right	Left
Sugar pucker	C _{3'} - <i>endo</i>	C _{2'} - <i>endo</i>	C _{3'} - <i>endo</i>
Number of nucleotides per pitch or base pairs per turn	11	10.4	12
Turn angle per nucleotide or twist angle (t , °)	32.7	34.6	30
Helical rise (h, Å)	2.56	3.38	3.7
Pitch (P, Å)	24.6	34.0	45.6
Diameter (Å)	25.5	23.7	18.4
Conformation of glycosidic bond	Anti	Anti	Anti at C Syn at G
Major (minor) groove width (Å)	2.7 (11.0)	11.7 (5.7)	
Major (minor) groove depth (Å)	13.5 (2.8)	8.5 (7.5)	

antitumor agents such as *cis*-dichlorodiammineplatinum(II), *cis*DDP, can bind to DNA, their major target *in vitro* and *in vivo*.

1.10.2 Transmission of Genetic Information

Three major components in the transmission of genetic information are deoxy-ribonucleic acids (DNA), ribonucleic acids (RNA), and proteins. The genetic code expressed through DNA ultimately determines which proteins a cell will produce. Coiled and supercoiled DNA molecules contain numerous sequences of nucleotides that may be transcribed as RNAs and translated to many different proteins. DNA molecules also contain long sequences of nucleotides not coding for protein and whose purpose is not completely understood. A gene is a specific sequence of DNA that encodes a sequence of messenger RNA (mRNA) codons needed to synthesize a complete protein. Genes may be collected together in a chromosome, the DNA packaging unit in eukaryotes. The human genome is an ordered sequence of almost three billion adenine, guanine, cytosine, and thymine bases found on 46 chromosomes. Most cells transfer genetic information DNA → RNA but retroviruses (such as the human immunodeficiency virus, HIV) transmit their information RNA → DNA. Cells have three major types of RNA. Messenger RNA (mRNA) is produced directly from a DNA template (the sequence of nitrogenous bases along the DNA chain) in a process called transcription. Transfer RNA (tRNA), found in the cytosol, transfers an amino acid to the ribosome for protein synthesis in the translation process. Ribosomal RNA (rRNA) is an essential component of the ribosome, the organelle in which particles of RNA and protein come together to synthesize proteins. The process of RNA interference, RNAi, prevents genes from being translated into proteins. This significant “gene-silencing” discovery, described experimentally in 1998 by biologists Andrew Fire and Craig Mello

[20], revolutionized genetics, transformed understanding of cellular behavior and has spurred development of new medical treatments. Fire, of Stanford University School of Medicine, and Mello, of the University of Massachusetts Medical School, shared the 2006 Nobel Prize in Physiology or Medicine for their work.

The processes of replication, transcription, and translation are illustrated schematically in Figure 1.16 and further explained in the following paragraphs.

When a cell divides, DNA duplicates itself in a multistep process called *replication*. DNA replication takes place in a semiconservative manner meaning that the DNA double strand unwinds and each single strand acts as a template for a new complementary strand. Thus, each daughter DNA strand half comprises molecules from the old strand. DNA replication is not a passive nor spontaneous process. Many enzymes are required to unwind the double helix and synthesize a new DNA strand. The enzyme topoisomerase, for instance, is responsible for the initiation of DNA unwinding. Nicking a single DNA strand releases the tension holding the helix in its coiled and supercoiled structure. Another enzyme, helicase, requires energy in the form of ATP to unwind the original double strand, held together by their A–T and G–C hydrogen bonds, finally forming the single-stranded DNA (ssDNA) to be replicated. DNA polymerase holoenzyme, a complex aggregate of several different protein subunits, then proceeds along the single DNA strand, bringing free deoxynucleotide triphosphates (dNTPs) to hydrogen bond with their complementary nucleotide on the single strand. Energy stored in the incoming dNTP is used to covalently bind each deoxynucleotide to the growing second strand assisted by DNA polymerase III. Primase is an enzyme component of the protein aggregate called the primeosome. It attaches a small RNA primer to the single stranded DNA to act as a substitute 3'-OH for DNA polymerase to begin synthesis of the second strand. The RNA primer is eventually removed by RNase H and the gap filled in by DNA polymerase I. The enzyme ligase can catalyze the formation of a phosphodiester bond given adjacent but unattached 3'-OH and 5'-phosphate groups. Platinum-containing anticancer agents attach themselves preferentially to adjacent guanines in double stranded DNA. The DNA platination causes kinking in the ds DNA and this is believed to prevent DNA replication by the mechanism just discussed. There are many repair mechanisms that can remove the DNA platination lesion and other mechanisms for skipping over the platinated portion of ds DNA. These repair mechanisms are believed responsible for some of the acquired resistance that cancerous cells develop, eventually rendering the platinum drugs ineffective.

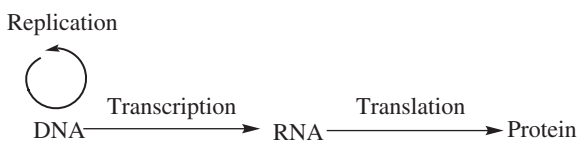


Figure 1.16 Schematic representation of replication, transcription, and translation.

Transcription is the process that transfers sequence information from the gene regions of DNA to messenger RNA (mRNA) so that it can be carried to the ribosomes in the cytoplasm. The primary protein aggregate responsible for effecting this process is RNA polymerase holoenzyme. This enzyme aggregate directs the synthesis of mRNA on a DNA template. To begin the process, RNA polymerase must be able to recognize a particular DNA sequence at the beginning of genes called the promoter. The promoter is a unidirectional sequence on one strand of the DNA that tells RNA polymerase both where to start and in which direction (on which strand) to continue synthesis. The DNA strand from which RNA polymerase copies is called the antisense or template strand, and the other DNA strand the sense or coding strand. The RNA polymerase gathers ribonucleic nucleotide triphosphates (NTPs or rNTPs) proceeding to synthesize the single RNA strand in the 5' to 3' direction. Transcription terminates when the RNA polymerase reaches a stop signal on the gene.

In the *translation* process, proteins are synthesized in the cellular factory called the ribosome. The ribosome consists of structural RNA and about 80 different proteins. When inactive, the ribosome contains a large subunit and a small subunit. When the small subunit encounters mRNA, the process of translation to protein begins. Amino acids necessary for protein synthesis bind at two sites on the large subunit; these sites are close enough together to facilitate the formation of a peptide bond. One site (A) accepts a new tRNA carrying an amino acid while the second site (P) bears the tRNA attached to the growing protein chain. Each tRNA has a specific anticodon and acceptor site. The genetic code is composed of 64 triplet *codons* (4^3 , 4 being the nucleobases A, U/T, C, G with the triplet code raising 4 to the third power) so that each of the 20 amino acids has multiple codons. Some examples are as follows: CAT/U and CAC code for histidine (his); GG T/U, GGC, GGA, and GGG code for glycine (gly); and T/U AA, T/U AG, and T/U GA are STOP codons that end the protein sequence. Each tRNA also has a specific charger protein, called an aminoacyl tRNA synthetase, that binds to a particular tRNA and attaches the correct amino acid to the acceptor site. The start signal for translation is the AU/TG codon, coding for methionine (met). Not every protein starts with met so that often this amino acid must be removed during post-translational protein processing. This so-called posttranslational modification (PTM) is one of a huge number of protein modifications that take place following protein synthesis. For example, newly synthesized proteins may have certain tyrosine residues phosphorylated or asparagine residues glycosylated. PTMs are important research targets in proteomics as they have far-reaching implications for cellular behavior (and misbehavior in disease processes). The ribosomal large subunit binds to the mRNA and the ribosomal small subunit and elongation begins. Entering tRNAs with attached amino acids attach to the A site matching their codon to that on the mRNA. A peptide bond is formed with the amino acid attached to the tRNA at the P site followed by a shift of the new peptide to the P site freeing the A site for the next tRNA. When the ribosome reaches a stop codon, no aminoacyl tRNA binds to the

empty A site, the ribosome breaks into its large and small subunits and the newly synthesized protein and mRNA are released.

1.10.3 Genetic Mutations and Site-Directed Mutagenesis

Two major types of genetic mutation are possible: *chromosomal mutations* and *point mutations*. Chromosomal mutations can alter or shift large sections of chromosomes (collections of genes) leading to changes in the way their genes are expressed. In *translocations*, large DNA segments are interchanged between two different chromosomes causing possible changes in gene expression. *Inversions* occur when a region of DNA flips its orientation with respect to the rest of the chromosome, also leading to changes in gene expression. Sometimes large sections of a chromosome may be deleted, leading to loss of important genes. Chromosome nondisjunction arises when one daughter cell receives more or less DNA than it should during cell division. Point mutations arise from single base pair changes. A *nonsense mutation* creates a stop codon where none previously existed, possibly removing essential regions in the resulting protein – that is, U/TAA (stop) instead of U/TAC (tyr). A *missense mutation* changes the mRNA code, placing a different amino acid into the primary sequence and possibly changing the shape or the metal coordination sphere of the resulting protein. For example, AAC (asparagine, asn, N) could be substituted for CAC (histidine, his, N) by a missense point mutation. A *silent mutation* codes for the same amino acid in the protein primary sequence, i.e. CAC and CAU/T both code for histidine. Within a gene, insertion of a small number of bases not divisible by three will result in a *frame shift mutation*, resulting in an incorrect series of amino acids following the frame shift.

Organisms have elaborate proofreading and repair mechanisms that recognize false base-pairings and other types of DNA damage and repair these. Genetic mutations that occur in spite of proofreading and repair mechanisms cause genetic diseases. Recent advances in genetic manipulation have allowed the source of genetic diseases to be discovered, and in some cases treated. Publication of the human genome sequence in 2001 constituted a large step forward in the information available for scientists studying genetic disease and treatment.

Scientists use mutations to study proteins by introducing amino acid substitutions in crucial locations to determine regulatory functions, protein tertiary and quaternary structure, catalytic mechanisms, and substrate-binding sites. Researchers often target conserved amino acid residues – those aa residues that appear in the same enzyme in different biological species and are usually evolutionarily conserved as well. In order to carry out these mutations, a process called *site-directed mutagenesis* is invoked, during which the coding gene must be isolated, cloned, sequenced, and then changed in a specific manner. These processes of DNA isolation and cloning, called recombinant DNA techniques, will be discussed further in Section 1.10.4. Site-directed mutagenesis enables the scientist to change any residue in any protein to any other residue by changing the DNA code and then transcribing and translating the mutated

DNA. The essential process involves extracting the plasmid containing the gene of interest from bacterial cells (usually *Escherichia coli* cells) and then using a series of enzymatic reactions to switch codons. The plasmid is then reintroduced to the *E. coli* cells for expression – transcription and translation – into the mutated protein. The three-dimensional structure of the mutated protein is studied by many methods, one of the most popular being X-ray crystallography. Examples using mutation to understand the behavior of proteins and enzymes are discussed in the following sections:

- Section 3.2.4 where the flow of electrons to the enzyme's active site is affected by amino acid mutations in the so-called K channel while the flow of protons is affected by mutations in the D channel (see Figure 3.12).
- Section 3.3.3 where mutation of histidine or valine residues makes the hydrogenation of dinitrogen less efficient.
- Section 3.4 where hundreds of different mutations of the CuZnSOD enzyme have been found to cause diseases such as amyotrophic lateral sclerosis (ALS) and trisomy 21 – Down syndrome (see Figure 3.34).
- Section 4.2.3 where a ser320ala (amino acid residue 320 replaced by alanine) in the [FeFe]-hydrogenase enzyme under study shows a dramatic decrease in dihydrogen production that is thought to be related to ser320's proton donation capability. Figure 4.11 visualizes the enzyme using X-ray crystallographic data deposited in the PDB. Section 2.7 explains how to access and use the PDB.
- Section 4.4.3.3 where copper-containing nitrite ion reductases are discussed. Mutation of aspartic acid and histidine residues near the active site markedly reduces enzyme activity.

1.10.4 Genes and Cloning

Much of the information contained in this section and that following has been obtained from the website http://www.lsic.ucla.edu/ls3/tutorials/gene_cloning.html and references [7, 17a, 19b]. Eukaryotes, species whose cells contain a nucleus and other intracellular structure, have more complex gene expression than prokaryotes, which are simpler life forms whose cells have no intracellular compartments. In eukaryotes, DNA is segmented into exons (regions that encode for protein) and introns (regions that do not). When messenger RNA (mRNA) is initially transcribed, both exon and intron regions are included. Subsequently, the mRNA is spliced together at points so that only the exon regions will be translated into protein in the ribosome. DNA molecules are large and complex to work with in their native state. Usually scientists wish to work with only the DNA exons, and particularly the exons that code for the specific protein(s) of interest. *Recombinant DNA* technology (also called *cloning* and *genetic engineering*) combines a number of different techniques that allow scientists to manipulate and replicate DNA as well as study its structure and function. Recombinant DNA is DNA that has been created artificially by

techniques to be described in the following paragraphs. The recombinant DNA molecule must be replicated many times to become useful for laboratory research, this process is called cloning.

The first step in the cloning process is to break the large DNA molecule down into fragments. Sanger sequencing (also called the chain termination, first generation, or dideoxy method) uses an enzymatic procedure to produce DNA chains of varying length in four different reactions, stopping DNA replication at positions occupied by one of the four bases and then determining the resulting fragment lengths. This method uses naturally occurring enzymes called *restriction endonucleases*. Restriction endonucleases are a class of enzymes, generally isolated from bacteria, that recognize and cut specific sequences – restriction sites – in DNA. One restriction enzyme, BamHI, locates and cuts any occurrence of



The enzyme clips after the first G in each strand. Both strands contain the sequence GGATCC but in antiparallel orientation. This type of recognition site is called *palindromic*. Another sequencing method, Maxam–Gilbert (also called the chemical degradation method), uses chemicals to cleave DNA at specific bases, resulting in fragments of different lengths. The resulting DNA restriction fragments are separated by electrophoresis on an agarose gel. Even fragments that differ in size by a single nucleotide can be resolved. The gel-separated fragments are converted to single stranded DNA (ssDNA) by treatment with strong base. In the Southern blot technique, the gel is then covered with nitrocellulose paper and compressed with a heavy plate. The nitrocellulose paper preferentially takes up the ssDNA producing a pattern of DNA identical to that on the gel. A DNA probe (either an ssDNA or RNA fragment with a complementary base sequence to the DNA of interest) then identifies the desired single strands of DNA through use of a ^{32}P radioactively tagged nucleotide probe. For eukaryotic DNA, one usually begins with an mRNA template that will synthesize the desired DNA by treatment with reverse transcriptase, an enzyme that transcribes RNA to DNA, the reverse of the normal transcription process. The synthesized DNA is called cDNA as it carries a complementary base sequence to the mRNA template. In this cloning method, only the DNA exon portion that codes for the desired protein is transcribed, and the intron portion that does not code for protein is not transcribed.

Next (or second)-generation sequencing (NGS) methods began to be developed in the early 2000s. These methods are highly automated, allowing the entire genome to be sequenced at once. The genome is fragmented into small pieces, the fragments are sampled randomly, and multiple fragments are sequenced simultaneously using a variety of automated technologies.

Depending on the sequencing technology to be used, DNA or RNA extraction samples require further preparation. For Sanger sequencing, one must use cloning or polymerase chain reaction (PCR) procedures prior to sequencing. The sequence of steps for the PCR process is further described in section 2.3.5 of reference [5].

The PCR method has found applications in clinical medicine, genetic disease diagnostics, forensic science, and evolutionary biology. For example, very small samples of DNA, perhaps from a crime scene or from an ancient mummy, can produce sufficient copies to carry out forensic tests such as DNA profiling.

For NGS methods, library preparation is required. An overview of NGS library preparation methods and challenges can be found in reference [21]. For all methods, assessing the quantity and quality of nucleic acids is necessary and interpretation of the ultimate findings adds another layer of uncertainty. In 2019, several companies that provide sequencing technology instrumentation and computing solutions are found at these websites: <https://www.illumina.com> (Illumina), <https://www.pacb.com> (Pacific Biosciences), and https://lifescience.roche.com/en_us.html (Roche).

1.10.5 Genomics and the Human Genome

Each cell in an organism contains its complete genome but depending on the cell type, only the genes necessary for conducting the work of that cell type are expressed. The human genome sequence consists of an ordered listing of the adenine, cytosine, guanine, and thymine bases found on the 46 human chromosomes. Only about 1% of the genome sequence codes for proteins are necessary for human life. Most of the rest of the genome consists of large repetitive noncoding regions whose function is not well understood. It is known, however, that critical clues to diseases such as cancer, diabetes, and osteoporosis lie in areas of the genome that do not code for protein. About one-fourth of the genome contains long, gene-free segments while other regions contain much higher gene concentrations. The number of protein coding regions (genes) in the human genome, estimated at 30 000–40 000, is surprisingly small, given that the fruit fly has 13 000 genes and the thale cress plant 26 000 [22d].

Francis Collins, director of The National Human Genome Research Institute and Ari Patrinos and head of the Office of Biological and Environmental Research at the Department of Energy's (DOE) Office of Science, managed and coordinated the sequencing of the human genome program, initiated in 1990. A substantially complete version of the 2.9 billion base-pair human genome sequence was published in 2001 [22]. Reference [22a] reported the results of an international collaboration to produce and make freely available a draft sequence of the human genome, presented an initial analysis of the draft sequence data, and described some of the insights that could be gleaned from the sequence. In addition to the draft sequence, pp. 745–964 of this February 2001 issue of *Nature* contained accompanying discussions from many scientists on diverse topics in evolutionary, medical, and biological sciences that would benefit from this accomplishment.

The Human Genome Project used the shotgun sequencing method in which enzymes cut DNA into hundreds or thousands of random bits that were then sent to automated sequencing machines capable of handling DNA fragments up to 500 bases long. After sequencing, the fragments were pieced back together to become part of the sequenced genome. The shotgun approach was

applied to cloned DNA fragments that had already been mapped; that is, the fragment's location on the genome was already known.

The Human Genome Project delivered a complete human genome sequence available to scientists in a freely accessible database when it released the first reference sequence for the human genome in April 2003. The National Institutes of Health's website address for current human genome sequence data and information is <http://www.ncbi.nlm.nih.gov/genome/guide/human> and the National Human Genome Research Institute's researcher resources website is <http://www.genome.gov>.

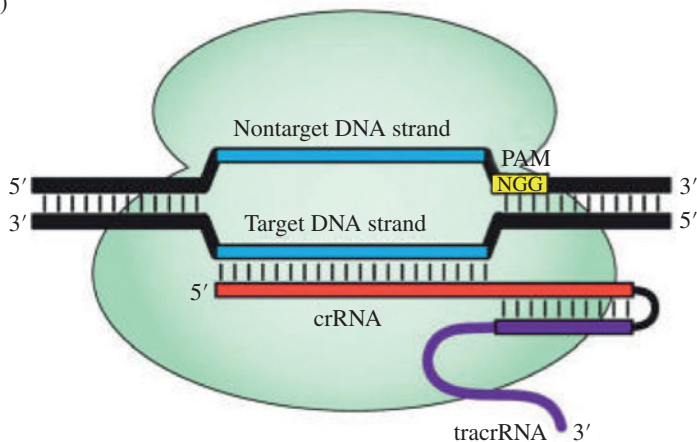
The Human Genome Project's goal was to produce high-quality, accurate, finished DNA sequences according to the following standards: (i) The DNA sequence is 99.99% accurate; (ii) the sequence must be assembled; that is, the smaller lengths of sequenced DNA have been incorporated into much longer regions reflecting the original piece of genomic DNA; (iii) the task must be affordable (the project funds technology development to reduce costs as much as possible); and (iv) the data must be accessible. To this end, verified DNA sequencing data is deposited in public databases on a daily basis (see the website <http://www.ncbi.nlm.nih.gov/genome/guide/human>).

1.10.6 CRISPR

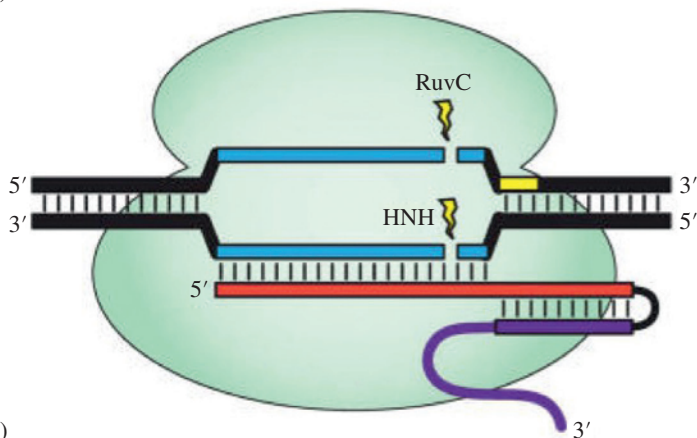
Clustered regularly interspaced short palindromic repeats (CRISPR) families of DNA sequences are found in the genomes of prokaryotic organisms such as bacteria and archaea. CRISPR sequences have been studied since the 1980s. Since 2012, the class 2 Type II Cas9 (CRISPR-associated protein 9), an RNA-guided DNA endonuclease, has been studied because it targets sites in the genome through simple DNA–RNA base-pairing [23]. As an endonuclease, the Cas9 enzyme cleaves the phosphodiester bond along a polynucleotide chain. It uses CRISPR sequences to recognize and cleave specific strands of DNA that are complementary to the CRISPR sequence. The bacterial Type II CRISPR/Cas9 system may be used to edit genes within an organism fostering basic biological research, developing a myriad of biotechnological products, and possibly treating various genetic diseases. See reference [23f] for an updated classification of CRISPR-Cas systems. In reference [23h], David Scott and coworkers discuss functionally diverse class 2 type V CRISPR-Cas systems that can cleave RNA strands and ssDNA and cleave or nick dsDNA depending on their size (720–1093 amino acid residues) and other characteristics such as having an intact RuvC domain. In 2019, CRISPR-Cas technology has been used in cultured cells and animals to edit genes responsible for a number of diseases. A 2016 *Science* article [23i] reported using an adeno-associated virus to deliver a CRISPR-Cas9 system to a mouse model of Duchenne muscular dystrophy (DMD). The system removed mutated exon 23 from the dystrophin gene and expressed a modified dystrophin gene. This resulted in improved muscle biochemistry and muscle force.

In reference [23b], Jennifer Doudna and colleagues write a summary of the CRISPR/Cas9 technology summarizing it clearly in their figure 1. See reference [23b] and Figure 1.17 for a visualization of the system taken from PDB 4UN3

(a)



(b)



(c)

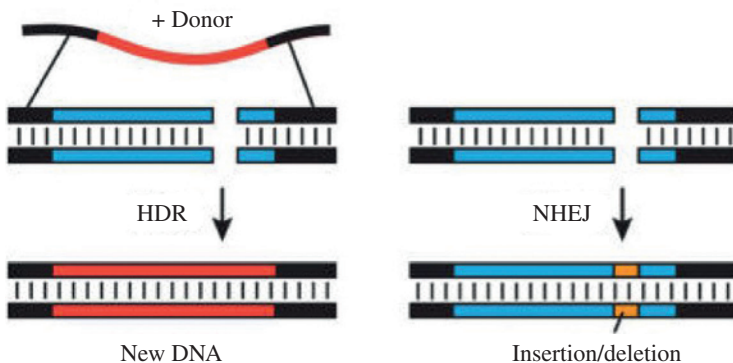


Figure 1.17 (a) Cas9 (green); crRNA (orange); tracrRNA (purple); PAM sequence (yellow). (b) RuvC domain cuts non-target strand; HNH domain cleaves target strand. (c) Two repair pathways activated after DNA cleavage. *Source:* figure 1 of reference [23b]. Copyright 2016 American Chemical Society.

data [23b]. The Cas9 protein, shown as a green cloud, requires a CRISPR RNA (crRNA) and a transactivating RNA (tracrRNA) to target the DNA sequence with 20 base pair complementarity to the crRNA. The two RNAs may be spliced together to create a single-guide RNA (sgRNA). Additionally, a protospacer adjacent motif (PAM) is required for target binding and DNA cleavage. The PAM resides on the nontarget DNA strand directly adjacent to the crRNA-matching DNA sequence. PAMs usually have the sequence 5'-NGG-3': in the 4UN3 system, the sequence is 5'-TGG-3'. The Cas9 recognition lobe (Rec) is shown in magenta in Figure 1.18. Figure 1 of reference [24] shows detailed representations of the DNA–RNA interactions. Two DNA cleavage active sites are shown:

- The RuvC domain cuts the nontarget DNA strand. RuvC is a small endonuclease protein requiring a magnesium ion. Shown in cyan in Figure 1.18.
- The HNH domain endonuclease cleaves the target DNA strand. It also requires one or more magnesium ions and potassium ions. It is shown in yellow in Figure 1.18.

In reference [23b], figure 1 also indicates two repair pathways:

- Homology-directed repair (HDR) introduces a donor DNA strand template that has matching sequences at the site of delivery. The added homologous DNA molecule serves as a recombination-mediated repair.

PAM by element spheres

RNA orange

Target DNA blue

Nontarget DNA black

HNH-Cas-9 yellow

RuvC cyan

Rec magenta

Mg²⁺ green spheres

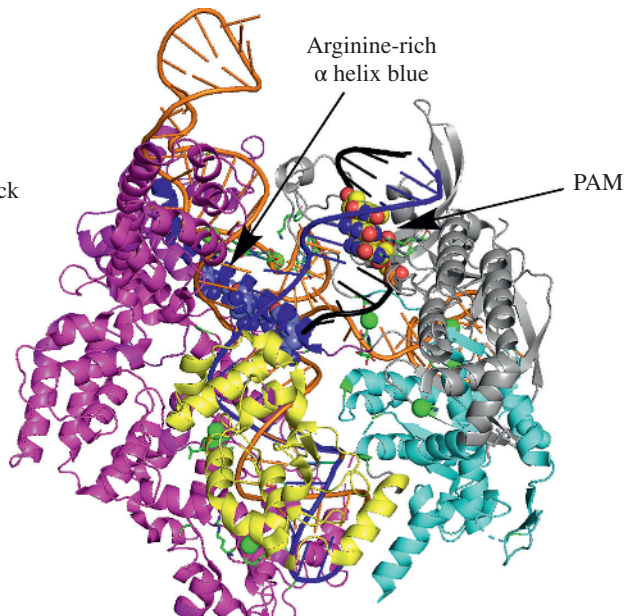


Figure 1.18 Visualization of Cas9 ribonucleoprotein (RNP) from PDB 4UNS [24].

- Nonhomologous end joining (NHEJ) stitches the cleavage site together resulting in an insertion or deletion mutation.

In addition, the [23b] authors describe light and chemically inducible Cas9 as well as non-inducible split-Cas9 technologies. The latter technology takes advantage of Cas9 protein's ability to split into two lobes – an α -helical recognition (REC) lobe (magenta in Figure 1.18) and a nucleobase lobe (yellow, cyan, and gray in Figure 1.18) – that can be reassembled *in vivo*. In one example shown in figure 2 of reference [23b], the two lobes are fused to inteins – small protein sequences that recognize each other and are spliced out when the REC and nucleobase lobes reassemble. The described technologies are Cas9 “on switches” that will prevent “off target” reactions and help target the CRISP/Cas9 to its desired location in cultured cells or eventually in organisms.

In Figure 1.18, green sphere magnesium ion ligands – either from amino acid residues or nucleotide bases – are shown in green stick color-by-element format. The magnesium ion near the PAM is coordinated by a phosphate oxygen atom of adenine65 and the O _{γ} atom of thr1103. The magnesium ion in the HNH domain (yellow) is coordinated by the oxygen atoms of glu584 and arg586. In the cyan RuvC domain, the O _{$\delta 2$} atom of asp10 provides one Mg²⁺ ligand. The arginine-rich helical bridge region (blue) includes arg residues 63, 66, 70, 71, 74, and 78 in sphere format. The PAM region includes nucleotides dG6, dG7, and dT8 in yellow colored-by-element sphere format. The important arg1333 and arg1335 residues reaching into the PAM are shown in green stick color-by-element format. An informative schematic diagram of the Cas9 protein can be found at <http://sites.tufts.edu/crispr/crispr-mechanism>.

Selectivity and specific delivery of agents to cellular target DNA for gene editing remains the major concern for CRISPR/Cas9 technology. A consortium of researchers recently has reported on cell receptor-mediated delivery of CRISPR/Cas9 for specific gene editing [25]. This delivery method would allow cell-selective gene editing without concerns about delivery of foreign genetic material using viruses or nanoparticles. The researchers concentrated on a hepatocyte-selective delivery system using engineered CRISPR/Cas9 RNA-protein complex (RNP) that featured a ligand for a particular receptor on liver-derived cells. The system was shown to be internalized by endocytosis. Subsequently endosomal escape – still a bottleneck in this approach – followed by nuclear transport ultimately placed the system in the nucleus and resulted in selective gene editing (see figure 1 of reference [25]). The researchers are optimistic that the “receptor-guided” approach will eventually allow targeting of specific cell types *in vivo*.

Ethical questions arising from the use of the technology abound. In late 2018, a Chinese scientist claimed to have used CRISPR/Cas9 gene editing to create embryos resistant to HIV, implanted the embryos, and reported the birth of twin girls resulting from the implantation. Critical reaction to this report was immediate with scientists decrying the use of the technology for human gene editing. In early 2019, the scientist had been fired from his Chinese institution and was facing investigation and possible prosecution. The

technology has created controversy internationally and in the United States as laboratories and start-up companies vie for patent rights and technology access [26]. (see reports in Science magazine for updates, for example, https://www.sciencemag.org/news/2017/02/how-battle-lines-over-crispr-were-drawn?r3f_986=https://www.google.com).

1.11 A DESCRIPTIVE EXAMPLE: ELECTRON TRANSPORT THROUGH DNA

The laboratory of Jacqueline K. Barton, a professor at the California Institute of Technology, studies charge transport (CT) in DNA attempting to elucidate fundamental mechanisms and kinetics of electron transport through DNA. The researchers study how DNA becomes damaged in the genome and how this damage may be sensed and repaired. Experimental evidence indicates that DNA base excision repair proteins may use DNA charge transport to localize repair enzymes near the damage site. The research has been in progress since the mid-1990s [27].

One branch of the research uses metal complexes with bulky planar ligands that can intercalate between the DNA-stacked bases. Two such compounds often used in this regard are the ruthenium complexes:

- $\Delta\text{-}[\text{Ru}(\text{bpy})_2\text{dppz}]^{2+}$ (bpy = 2,2'-bipyridine and dppz = dipyrldophenazine)
- $\text{Ru}(\text{phen})_2\text{dppz}]^{2+}$ (phen = phenanthroline and dppz = dipyrldophenazine)

The former complex, luminescent when intercalated into a duplex DNA strand, was used in one instance to identify DNA mismatches – locations that must be repaired efficiently. Deficiencies in DNA-mismatch repair (MMR) are implicated in many cancers and cells that are MMR deficient show much higher frequency of mismatches than normal cells [28a]. The researchers were able to obtain the X-ray crystallographic structure of the $\Delta\text{-}[\text{Ru}(\text{bpy})_2\text{dppz}]^{2+}$ complex interacting with the oligonucleotide 5'-(dCGGAAATTACCG)₂. The data is deposited in the PDB with accession number 4E1U. See Section 2.7 for information on how to use the PDB (see Figure 1.19). The planar dppz ligand inserts into the DNA duplex through the minor groove, ejecting both mismatched adenine nucleotides (shown as yellow sticks in Figure 1.19). This process is named metalloinsertion. At the mismatched sites, the bpy ligands have stacking interactions with the extruded adenosines. Other dppz ligands intercalate between matched base pairs in the process called metallointercalation. In Figure 1.19, ruthenium ions are shown as metallic gray spheres, attendant barium ions as red spheres, and the bpy and dppz ligands in green stick format (nitrogen-coordinating atoms in blue). The PhD thesis of Adam Boynton gives much more detail on the synthesis of these ruthenium complexes, their selective probing of DNA mismatches. The complexes operate as a “light switch” with considerably brighter and longer light emission in DNA with mismatches relative to well-matched DNA. The behavior has

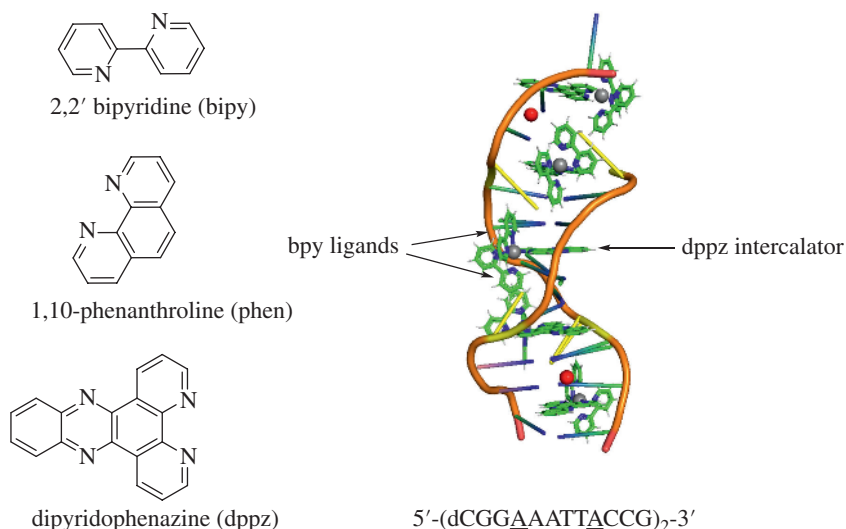


Figure 1.19 A ruthenium complex that can detect a DNA mismatch. PDB 4E1U data visualized using PyMOL. *Source:* Barton et al. [28a] and The PyMOL Molecular Graphics System Version 2.3.2 Copyright (C) Schrodinger LLC. For more information: <https://pymol.org> sales@schrodinger.com [28b].

been characterized in live cells [29]. The researchers believe that ruthenium complexes that luminesce more brightly in areas of DNA mismatches could become a diagnostic probe for detecting mismatches in (MMR) mismatch repair-deficient cancer cells.

DNA damage may be caused by reactive oxygen species such as superoxide ion radicals, UV irradiation, and environmental mutagens. See Section 3.4 for a discussion of superoxide dismutases, which are enzymes whose task is to remove physiologically formed superoxide radicals. Millions of cells per day in the genome may experience lesions, DNA mismatches, or base-pair modifications. A multitude of repair mechanisms exist to counteract DNA damage. One mechanism involves DNA-processing enzymes, many of which have been found to contain [4Fe4S] clusters. See Section 1.6, Figure 1.7, and Section 3.3 for more discussion of iron-sulfur clusters.

J. K. Barton and coworkers have proposed that the DNA-processing enzymes containing [4Fe4S] clusters become efficient at finding DNA damage by signaling each other via charge transport (CT) through dsDNA between [4Fe4S] sites [30]. The proteins of interest, *E. coli* endonuclease III (EndoIII) and MutY are DNA glycosylases known to stabilize the genome after oxidative stress.

The researchers had shown previously that CT-active proteins containing [4Fe4S] clusters (like EndoIII) locate themselves in the vicinity of DNA damage [31]. The [4Fe4S] clusters within the proteins alternate between the [4Fe4S]²⁺ and [4Fe4S]³⁺ states by CT through the dsDNA helix. Efficiency of this process declines in the presence of DNA mismatches located between the

protein and a DNA-modified electrode. Reference [30] explains the preparation of DNA-modified electrodes for the electrochemical studies. Electrochemical studies previously have shown that DNA binding by EndoIII leads to a large negative shift in the midpoint potential of the $[4\text{Fe}4\text{S}]^{3+/2+}$ cluster consistent with the stabilization of the oxidized $[4\text{Fe}4\text{S}]^{3+}$ form [32]. This appears to be the case for all $[4\text{Fe}4\text{S}]$ cluster-containing proteins that bind to DNA.

In reference [30], the researchers find that binding to DNA shifts the $[4\text{Fe}4\text{S}]^{2+/3+}$ couple negatively ~ 200 to $+80$ mV versus NHE (normal hydrogen electrode). Thermodynamically, this would indicate a stabilization of the $[4\text{Fe}4\text{S}]^{3+}$ state by >2 orders of magnitude compared to $[4\text{Fe}4\text{S}]^{2+}$. To indicate how the $[4\text{Fe}4\text{S}]^{3+/2+}$ cluster oxidation state modulates DNA binding and consequently controls DNA damage detection activity of repair proteins, Endonuclease III (EndoIII) and DinG are bound to a 3 kilobase pair (kbp) DNA duplex with a single-base mismatch. EndoIII is a base excision repair (BER) glycosylase that repairs oxidized pyrimidines in *E. coli*. DinG is a DNA helicase that unwinds R-loops caused by nascent mRNA strand invasion of DNA duplexes. EndoIII and DinG are known to contain high-potential $[4\text{Fe}4\text{S}]$ clusters that cycle between $[4\text{Fe}4\text{S}]^{3+}$ and $[4\text{Fe}4\text{S}]^{2+}$ states when bound to dsDNA. They are also known to act synergistically to localize on mismatch DNA strands. However, neither EndoIII nor DinG binds preferentially to the mismatch site, so the preference for binding to the mismatch-containing strand is instead attributed to the mismatch's ability to inhibit DNA charge transport (CTs).

Cyclic voltammetry of DNA-modified electrodes with and without added EndoIII was performed. See Section 1.11.1 for a description of this electrochemical method. Addition of EndoIII exhibited a single redox couple with a DNA-bound midpoint of approximately $+80$ mV versus NHE, consistent with previous measurements of EndoIII's $[4\text{Fe}4\text{S}]^{2+/3+}$ couple associated with DNA [32]. The researchers used Atomic Force Microscopy (AFM) to find the percentage of $[4\text{Fe}4\text{S}]^{3+}$ -containing protein bound on a DNA strand. Section 3.7.1 of reference [5] and Section 5.2.1.6 contain discussions of AFM. Briefly, AFM offers the magnification range of both light and electron microscopy and allows imaging of materials under the "natural" conditions usually associated only with the light microscope. Scientists can produce high-resolution images of biological material, images of molecules, and their interactions even under physiological conditions and study molecular processes in living systems. Images can be obtained for samples in air, water, or under vacuum with typical resolutions on the order of 10 nm. In the present case, the researchers are visualizing protein–nucleic acid complexes. Some other applications include the study of DNA, RNA, peptides, biopolymers and biomaterials, ligand–receptor binding, and chromosomal material.

Figure 2(b) of reference [30] indicates that for 33% oxidized EndoIII (67% $[4\text{Fe}4\text{S}]^{2+}$), there were 1.51 $[4\text{Fe}4\text{S}]^{3+}$ clusters per DNA strand and for $>99\%$ oxidized EndoIII ($<1\%$ $[4\text{Fe}4\text{S}]^{2+}$) there were 2.14 $[4\text{Fe}4\text{S}]^{3+}$ clusters per strand. This result indicates that as the ratio of oxidized EndoIII $[4\text{Fe}4\text{S}]^{3+}$ clusters

increases, the number of EndoIII protein bound on dsDNA with a single-base mismatch increases.

Published X-ray crystallographic structures of EndoIII complexed with dsDNA constructs show many interactions between the protein and the DNA. These include interactions between positively charged amino acid residues – arginine for one example – and the DNA phosphate oxygen groups. Additional contacts between nucleobases and EndoIII amino acid residues are evident. The protein data bank (PDB) entry 1ORN shows the structure of an EndoIII-DNA intermediate with a so-called “estranged” guanine [33]. This guanine does not have a partner cytosine residue but is held in place by covalent bonds between the carbonyl oxygen of glutamine residue and the N1 nitrogen atom of guanine (in the PyMOL parlance gln “O” bonded to dG’6 “N1”). A second aa residue, isoleucine (ile80), binds to the N2 atom of guanine G’d6 and could be written as ile80 O–N2 dG’6. See Table 1.9 for guanine numbering system – the N2 atom is the un-numbered exocyclic nitrogen atom.

In conclusion, the reference [30] authors state that DNA-mediated charge transport (CT) enhances the search for DNA mismatches. If a reduced [4Fe4S] cluster-containing protein encounters a lesion (or mismatch), it will be released to continue searching. If there is a lesion or mismatch disrupting DNA CT, the encountering oxidized [4Fe4S] cluster-containing protein remains attached. Through DNA-mediated redox signaling (CT), these proteins become very efficient in finding and repairing DNA damage.

In 2019, the Barton group published a study of the effective distance for DNA-mediated charge transport (CT) between associated repair proteins [34]. The researchers use the [4Fe4S]-containing repair protein EndoIII, incubating it with different lengths (1.5–9 kb) of duplex DNA. Atomic force microscopy (AFM) was used to quantify binding of the protein to DNA and to determine how the protein affinity for DNA changes with increasing DNA length. A sharp change in binding slope is observed at 3509 base pairs, a DNA length of about 1.2 μm , indicating a range for DNA CT. The change in slope was seen by plotting the number of base pairs versus the number of proteins bound to DNA (as found by AFM). Inside the DNA CT range, a reduced and bound [4Fe4S]²⁺-containing EndoIII protein could transfer an electron to an oxidized and bound [4Fe4S]³⁺-containing EndoIII protein via DNA CT. Outside the DNA CT range, electron transport did not take place. An EndoIII mutant Y82A (Y = tyr = tyrosine substituted by A = ala = alanine), known to not carry out DNA CT, incubated with duplex DNA exhibited no change in binding slope over a range from approximately 1000 to 8000 base pairs. Decreasing the temperature from 25 to 4 °C, which should increase base stacking and decrease DNA movement, increased the length of DNA CT to 5383 base pairs (~1.8 μm). Adding 50 μM of the metallointercalator [Ru(phen)₂dppz]Cl₂, known to stabilize DNA base-pair stacking, increased the length of DNA CT to 5399 base pairs (see Figure 1.19).

The experiments are explained in the following manner. When a [4Fe4S]²⁺-containing EndoIII protein is bound within DNA CT distance of a [4Fe4S]³⁺-containing EndoIII protein, the oxidation states are averaged and one slope is

attained. In another regime where $[4\text{Fe}4\text{S}]^{3+}$ -containing EndoIII protein is outside the range for effective DNA CT, more $[4\text{Fe}4\text{S}]^{3+}$ -containing EndoIII protein is bound and a different slope is attained. The researchers conclude that effective DNA CT length can be controlled and that effective DNA CT length as long as seen here enables faster genome lesion or mismatch searching via this long range signaling.

1.11.1 Cyclic Voltammetry

The description of the cyclic voltammetry (CV) electrochemical method is taken from a 2018 *Journal of Chemical Education* publication [35]. Cyclic voltammetry is used to investigate the reduction and oxidation behaviors of species in solution. For the biological or bioinorganic chemist, this usually means aqueous solutions with buffers to maintain pH and electrolytic agents to provide charge carriers. A cyclic voltammogram is shown in Figure 1.20.

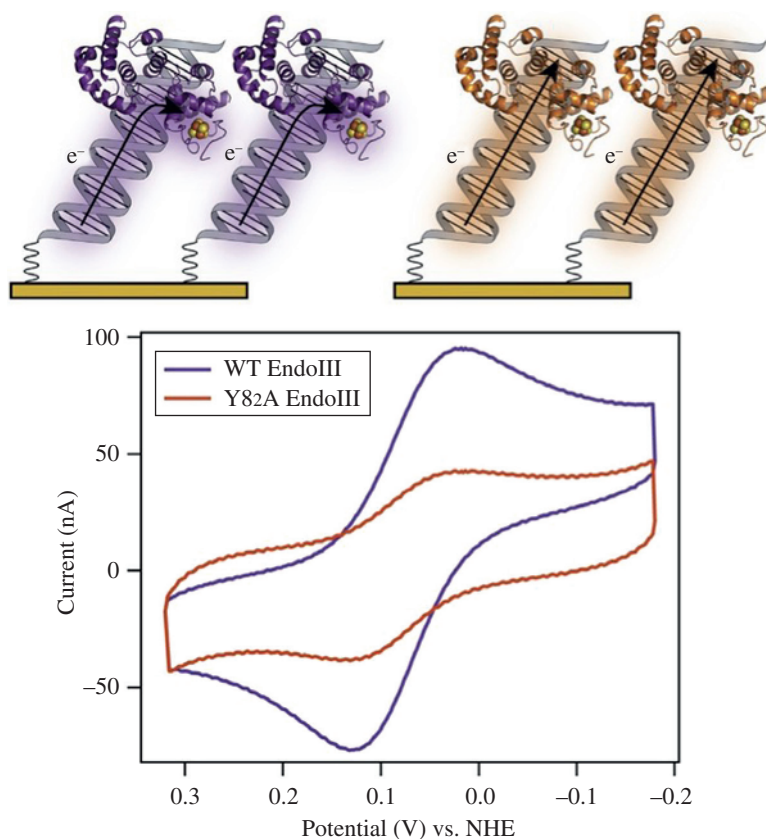


Figure 1.20 Cyclic voltammetry of wild-type and mutant EndoIII from figure 6 of reference [36]. *Source:* reproduced with permission of Elsevier.

The reduction potential for an oxidation–reduction reaction under other conditions is given by the Nernst equation as developed in Section 1.5:

$$E = E^0 - \left(\frac{RT}{nF} \right) \ln \left(\frac{[\text{reduced form}]}{[\text{oxidized form}]} \right) \quad (1.2)$$

Using the property of logarithms which states that $-\ln x = \ln(1/x)$, one can rewrite Eq. (1.2) as

$$E = E^0 + \left(\frac{RT}{nF} \right) \ln \left(\frac{[\text{oxidized form}]}{[\text{reduced form}]} \right) \quad (1.2a)$$

which is the form the equation takes in equation (1) of reference [35].

When a current is applied to the solution of interest, the plot of current (A) versus potential (V) takes the form of Figure 1.20 if the oxidation–reduction reaction is reversible. In this diagram, the experimentalists are assuming that one electron is transferred to and from one species in solution – in this case the $[4\text{Fe}4\text{S}]^{3+/2+}$ couple discussed in Section 1.11. By the convention used in the United States (opposite to that used by the International Union of Pure and Applied Chemistry – IUPAC), the potential is scanned from more positive to less positive. As the species of interest, $[4\text{Fe}4\text{S}]^{3+}$, starts to be reduced to $[4\text{Fe}4\text{S}]^{2+}$ at the working electrode (the electrode recording the oxidation/reduction event of interest), the cathodic wave increases to a peak cathodic current ($i_{p,c}$). As the rate of diffusion of $[4\text{Fe}4\text{S}]^{3+}$ slows the current decreases until the chosen switching point is reached. In the reverse scan, $[4\text{Fe}4\text{S}]^{2+}$ is oxidized to a peak anodic current ($i_{p,a}$) which slowly decreases as $[4\text{Fe}4\text{S}]^{2+}$ diffusion to the working electrode decreases. The two peaks are separated due to the diffusion of the $[4\text{Fe}4\text{S}]^{3+/2+}$ couple to and from the electrode. In an ideal reversible chemically and electrochemically reduction process, the difference between the cathodic and anodic peaks (ΔE_p) under standard conditions of temperature and pressure should be approximately 57 mV. Measurement of ΔE_p in Figure 1.20 shows a larger number indicating possible slow diffusion of $[4\text{Fe}4\text{S}]^{3+/2+}$ species and some irreversibility of the reaction being studied.

Many factors change the behavior of the species of interest in a cyclic voltammetry experiment. Solvents, buffers, electrolytes, scan rate in millivolt per second (mV/s), and type of electrodes used all are crucial in obtaining good results. In the CV experiments carried out in reference [30], the researchers report these conditions: scan rate = 100 mV/s, pH 7, 5 mM NaH_2PO_4 , and 50 mM NaCl (DNA buffer and electrolyte). To assess the effect of WT EndoIII and Y82A EndoIII on DNA duplexes, the authors use a DNA-modified electrode. Preparation of this electrode is described in reference [30]. Briefly, dsDNA is modified with a disulfide linker so that the dsDNA attaches to a gold electrode. This is shown schematically in Figure 1.20.

Reference [36] reviews the J. K. Barton laboratory's research into DNA charge transport (CT) providing a history of the work that has taken place from the mid-1990s until the article's publication in 2016. The data in Figure 1.20 reveals the difference in the behavior of wild-type EndoIII versus a mutated

version of the enzyme – a glycosylase. The mutant is less proficient in DNA charge transfer electrochemically resulting in less charge passing from the DNA through the EndoIII protein to its $[4\text{Fe}4\text{S}]^{3+/2+}$ cluster. The result is the lower signal intensity seen in the cyclic voltammetry experiment as shown in Figure 1.20.

1.12 SUMMARY AND CONCLUSIONS

The material introduced in Chapter 1 Inorganic and Biochemistry Essentials was chosen to reflect the emphases presented in the rest of this text. Analysis of metalloproteins and metalloenzymes discussed in Chapters 3 and 4 provides important information not only on structures but on functions of these bioinorganic systems. Knowledge of the protein composition and structural analyses of these metalloproteins and metalloenzymes will be assumed on the reader's part from the background presented in this chapter. In Chapter 6, familiarity with protein and nucleic acid structure and function is essential for understanding of the mechanisms of activity of inorganic drugs. The student is invited to refer to Chapter 1 materials often for greater understanding of the bioinorganic systems presented in the following chapters.

1.13 QUESTIONS AND THOUGHT PROBLEMS

1. Write out electron configurations for first row transition metals from Ti to Zn. Decide what oxidation states for the positive metal ions are most likely. Where low spin and high spin possibilities exist, write them out for octahedral, tetrahedral, and trigonal pyramidal cases.
2. Use the 16- and 18-electron rule to count electrons for the following (Figure 1.21):

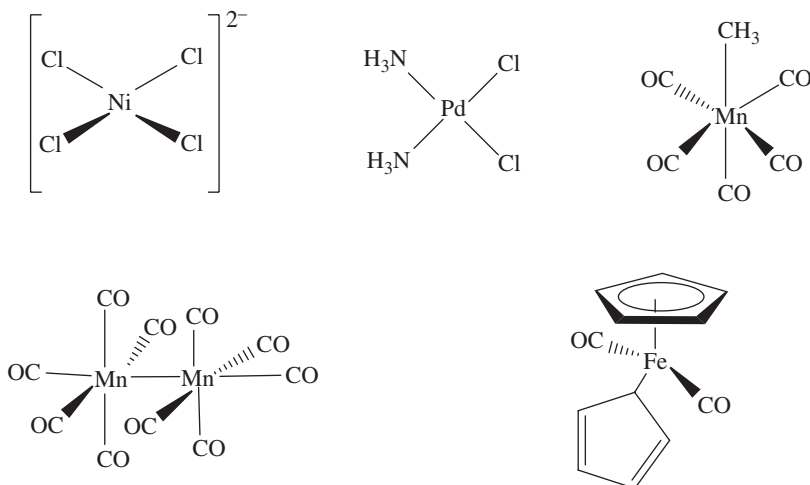


Figure 1.21 Structures for counting electrons using 16- and 18-electron rule.

3. Add five metalloenzymes to Table 1.7.
4. Assemble a peptide from the following amino acid residues: ala, his, lys, tyr, asp, phe, and thr. Attach metal ions in monodentate and bidentate coordination to appropriate locations based on actual proteins and enzymes.
5. Explain primary, secondary, tertiary, and quaternary protein structure and draw or name an example of each.
6. Look up the three letter codons that translate to specific amino acids via tRNA for the 10 essential amino acids.
7. Scan the metalloenzymes of Chapters 3 and 4 to decide whether conjugate acid–conjugate base or hard–soft acid–base characteristics determine the ligand field surrounding the metal ions.
8. Find a recent publication on CRISPR technology or CRISPR-Cas systems more current than any reference listed in Section 1.10.6. Identify a different theory or experiment that is not discussed in Section 1.10.6. Discuss the ethical ramifications of CRISPR technology with a friend or with classmates.
9. CRISPR-Cas systems contain magnesium and potassium ions in their published Protein Data Bank (PDB) structures. Are these required for genome editing? Are apo CRISPR-Cas systems active? Can other metals be substituted for magnesium and potassium ions? How about calcium? Or sodium?

REFERENCES

1. Silva, J.R.R., Frausto, d., and Williams, R.J.P. (2001). *The Biological Chemistry of the Elements: The Inorganic Chemistry of Life*, 2e. Oxford, UK: Oxford University Press.
2. (a) Housecroft, C.E. and Sharpe, A.G. (2018). *Inorganic Chemistry*, 5e. Essex UK: Pearson Education Ltd ISBN-13: 9781292134147. (b) Cotton, F.A., Wilkinson, G., Murillo, C.A., and Bochmann, M. (1999). *Advanced Inorganic Chemistry*, 6e. New York: John Wiley & Sons. (c) Cotton, F.A., Wilkinson, G., and Gaus, P.L. (1995). *Basic Inorganic Chemistry*, 3e. New York: John Wiley & Sons, Inc. (d) Wulfsberg, G. (2000). *Inorganic Chemistry*. New York: University Science Books.
3. (a) Schwarz, K. (1974). Recent dietary trace element research exemplified by tin, fluorine, and silicon. *Ged. Proc.* 33: 1748–1757. (b) Prasad, A.S. (ed.) (1976). *Trace Elements in Human Health and Disease: Vol II, Essential and Toxic Elements*. New York: Academic Press.
4. Kaim, W., Schwederski, B., and Klein, A. (2013). *Bioinorganic Chemistry: Inorganic Elements in the Chemistry of Life*, 2e. Chichester: John Wiley and Sons, Ltd.
5. Roat-Malone, R.M. (2007). *Bioinorganic Chemistry: A Short Course*, 2e. Hoboken, New Jersey: John Wiley & Sons, Inc.
6. (a) Crichton, R.R. (2008). *Biological Inorganic Chemistry An Introduction*. Oxford, UK: Elsevier. (b) Ochiai, E. (2008). *Bioinorganic Chemistry A Survey*. Oxford UK: Academic Press Elsevier.

7. Lippard, S.J. and Berg, J.M. (1994). *Principles of Bioinorganic Chemistry*. Mill Valley, CA: University Science Books.
8. Bertini, I., Gray, H.B., Stiefel, E.I., and Valentine, J.S. (2007). *Biological Inorganic Chemistry Structure and Reactivity*. Sausalito, CA: University Science Books.
9. Roat-Malone, R.M. (2002). *Bioinorganic Chemistry: A Short Course*. Hoboken, New Jersey: John Wiley & Sons Inc.
10. (a) Bearinger, J.P., Aines, R.D., Lightstone, F.C. et al. (2013). Comparison and analysis of zinc and cobalt-based systems as catalytic entities for the hydration of carbon dioxide. *PLoS ONE* 8 (6): e66187. <https://doi.org/10.1371/journal.pone.0066187>. (b) Tanner, D.B., Silverman, D.N., McKenna, R. et al. (2010). Comparison of solution and crystal properties of Co(II)-substituted human carbonic anhydrase II. *Arch. Biochem. Biophys.* 502: 53–59. PDB 3KOI, 3KON, 3KOK.
11. Vitale, A., D'Auria, S., Auld, D.S. et al. (2013). Physicochemical characterization of a thermostable alcohol dehydrogenase from *Pyrobaculum aerophilum*. *PLoS ONE* 8 (6): e63828. <https://doi.org/10.1371/journal.pone.0063828>.
12. Waern, J.B. and Harding, M.M. (2004). *Inorg. Chem.* 43 (1): 206–213.
13. (a) Waern, J.B., Harding, M.M., and Dillon, C.T. (2005). *J. Biol. Inorg. Chem.* 10 (5): 443–452. (b) Waern, J.B., Dillon, C.T., and Harding, M.M.J. (2005). *Med. Chem.* 48 (6): 2093–2099. (c) Caruso, F. and Rossi, M. (2004). *Met. Ions Biol. Syst.* 42: 353–384.
14. (a) Hoffmann, R.A. and Wheeler, R.A. (1986). A new magic cluster electron count and metal-metal multiple bonding. *J. Am. Chem. Soc.* 108 (21): 6605–6610. (b) Mingos, D.M.P. (1984). *Acc. Chem. Res.* 17: 311–319. (c) Mingos, D.M.P. (1981). *Comprehensive Organometallic Chemistry* (eds. G. Wilkinson, F.G.A. Stone and E.W. Abel). Oxford: Pergamon; (d) Elschenbroich, C. (1992). *Organometallics: A Concise Introduction*. New York: VCH.
15. Liu, J., Yu, Y., Lu, Y. et al. (2014). Metalloproteins containing cytochrome, iron–sulfur, or copper redox centers. *Chem. Rev.* 114: 4366–4469.
16. Bau, R., Eidsness, M.K., Rees, D.C. et al. (1998). Crystal structure of Rubredoxin from *Pyrococcus Furiosus* at 0.95 Ångstroms Resolution, and the structures of N-terminal methionine and formylmethionine variants of Pf Rd. Contributions of N-terminal interactions to thermostability. *J. Biol. Inorg. Chem.* 3: 484–493.
17. (a) Nelson, D.L. and Cox, M.M. (2005). *Lehninger Principles of Biochemistry*, 4e. New York: W.H. Freeman. (b) Berg, J.M., Tymoczko, J.L., and Stryer, L. (2006). *Biochemistry*, 6e. New York: W.H. Freeman. (c) Lehninger, A.L., Nelson, D.L., and Cox, M.M. (2000). *Principles of Biochemistry*, 3e. New York: Worth Publishers.
18. Sanders, C.R., Schleich, J.P., Capra, J.A. et al. (2019). Folding and Misfolding of Human Membrane Proteins in Health and Disease: From Single Molecules to Cellular Proteostasis. *Chem. Rev.* 119 (9): 5537–5606. (b) Zhou, H.-X. and Pang, X. (2018). Electrostatic interactions in protein structure, folding, and condensation. *Chem. Rev.* 118 (4): 1691–1741. (c) Dyson, H.J. and Wright, P.E. (2017). How does your protein fold? Elucidating the apomyoglobin folding pathway. *Acc. Chem. Res.* 50 (1): 105–111. (d) Luchinat, E. and Banci, L. (2018). In-cell NMR in human cells: direct protein expression allows structural studies of protein folding and maturation. *Acc. Chem. Res.* 51 (6): 1550–1557.
19. (a) Saenger, W. (1984). *Principles of Nucleic Acid Structure*. New York: Springer-Verlag, Inc. (b) Cowan, J.A. (1997). *Inorganic Biochemistry, An Introduction*, 2e. New York: Wiley-VCH.

20. Fire, A., Driver, S.E., Mello, C. et al. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391 (6669): 806–811.
21. Head, S.R., LaMere, S.A., Ordoukhanian, P. et al. (2014). Library construction for next-generation sequencing: overviews and challenges. *BioTechniques* 56: 61–77. <https://doi.org/10.2144/000114133>.
22. (a) Lander, E.S., Szustakowki, J., International Human Genome Sequencing Consortium et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409: 860–921. (b) Venter, J.C., Adams, M.D., Myers, E.W. et al. (2001). The sequence of the human genome. *Science* 291: 1304–1351. (c) Roberts, L. (2001). Controversial from the start. *Science* 291: 1182–1188. (d) Borman, S. (2001). Human genome sequence in print. *C&E News*, 12 February 2001: 9.
23. (a) Jinek, M., Doudna, J.A., Charpentier, E. et al. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816–821. (b) Doudna, J.A., Nunez, J.K., and Harrington, L.B. (2016). Chemical and biophysical modulation of Cas9 for tunable genome engineering. *ACS Chem. Biol.* 11: 681–688. <https://doi.org/10.1021/acscchembio>. (c) Jinek, M., Doudna, J.A., Charpentier, E. et al. (2012). A programmable dual-RNA – guided DNA endonuclease in adaptive bacterial immunity. *Science (Washington, DC, US)* 337: 816–822. (d) Doudna, J.A. and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science (Washington, DC, U.S.)* 346: 1258096–1258096. (e) van der Oost, J., Westra, E.R., Wiedenheft, B. et al. (2014). Unravelling the structural and mechanistic basis of CRISPR – Cas systems. *Nat. Rev. Microbiol.* 12: 479–492. (f) Makarova, K.S., Wolf, Y.I., Koonin, E.V. et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13: 722–736. (g) Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature* 526: 55–61. (h) Yan, W.X., Hunnewell, P., Scott, D.A. et al. (2019). Functionally diverse type V CRISPR-Cas systems. *Science* 363: 88–91. (i) Olson, E.N. (2016). Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science* 351 (6271): 400–403.
24. Jinek, M., Anders, C., Duerst, A. et al. (2014). Structural basis of Pam-dependent target DNA recognition by the Cas9 endonuclease. *Nature* (513): 569–573. PDB 4UNS.
25. Liras, S., Mascitti, V., Doudna, J.A. et al. (2018). mReceptor-mediated delivery of CRISPR-Cas9 endonuclease for cell- type-specific gene editing. *J. Am. Chem.* 140: 6596–6603.
26. (a) Cross, R. (2019). Scientist behind birth of gene-edited babies will be punished, Chinese investigators say. *C&ENews* 97 (4): 7. (b) Cross, R., Mullin, R., Satyanarayana, M., and Tremblay, J.-F. (2018). As claims of CRISPR use in first gene-edited babies emerge, scientists and ethicists respond. *C&ENews* 96 (48). (c) Cross, R., Mullin, R., and Satyanarayana, M. (2018). First gene-edited babies born, scientist claims. *C&ENews* 96 (48): 6–7.
27. Barton, J.K. and Barbara, P.F. (1996). Rates of DNA-mediated electron transfer between metallointercalators. *Science* 273 (5274): 475–480.
28. (a) Barton, J.K., Song, H., and Kaiser, J.T. (2013). Crystal structure of Δ -[Ru(bpy)₂dppz]²⁺ bound to mismatched DNA reveals side-by-side metalloinsertion and intercalation. *Nat. Chem.* 4 (8): 615–620. PDB 4E1U; (b) The PyMOL Molecular Graphics System Version 2.3.2 Copyright (C) Schrodinger LLC. <https://pymol.org/sales@schrodinger.com>.

29. Boynton, A.N. (2017) Targeting DNA mismatches with luminescent ruthenium complexes. PhD thesis. https://thesis.library.caltech.edu/10329/37/Adam_Boynton_thesis_July%202017.pdf
30. (a) Barton, J.K., Tse, E.C.M., and Zwang, T.J. (2017). The oxidation state of [4Fe4S] clusters modulates the DNA-binding affinity of DNA repair proteins. *J. Am. Chem. Soc.* 139: 12784–12792. (b) Furst, A.L., Hill, M.G., and Barton, J.K. (2013). DNA-modified electrodes fabricated using copper-free click chemistry for enhanced protein detection. *Langmuir* 29 (52): 16141–16149.
31. (a) O'Brien, E., Silva, R.M.B., and Barton, J.K. (2016). Redox signaling through DNA. *Isr. J. Chem.* 56: 705–723. (b) Sontz, P.A., Tainer, J.A., Barton, J.K. et al. (2012). DNA charge transport as a first step in coordinating the detection of lesions by repair proteins. *Proc. Natl. Acad. Sci. U. S. A.* 109: 1856–1861. (c) Slinker, J.D., Muren, N.B., Barton, J.K. et al. (2011). DNA charge transport over 34 nm. *Nat. Chem.* 3: 230–235.
32. (a) Gorodetsky, A.A., Boal, A.K., and Barton, J.K. (2006). Electrochemistry of endonuclease III in the presence and absence of DNA. *J. Am. Chem. Soc.* 128: 12082–12083. (b) Arnold, A.R., David, S.S., Barton, J.K. et al. (2017). Electrochemistry of the [4Fe4S] cluster in base excision repair proteins: tuning the redox potential with DNA. *Langmuir* 33 (10): 2523–2530.
33. Fromme, J.C. and Verdine, G.L. (2003). Structure of a trapped endonuclease III–DNA covalent intermediate. *EMBO J.* 22 (13): 3461–3471.
34. Tse, E.C.M., Zwang, T.J., Barton, J.K. et al. (2019). Effective distance for DNA-mediated charge transport between repair proteins. *ACS Cent. Sci.* 5: 65–72. <https://doi.org/10.1021/acscentsci.8b00566>.
35. Elgrishi, N., Rountree, K.J., Dempsey, J.L. et al. (2018). A practical beginner's guide to cyclic voltammetry. *J. Chem. Educ.* 95: 197–206. <https://doi.org/10.1021/acs.jchemed.7b00361>.
36. Arnold, A.R., Grodick, M.A., and Barton, J.K. (2016). DNA charge transport: from chemical principles to the cell. *Cell Chemical Biology* 23: 183–197. doi: 10.1016/j.chembiol.2015.11.010 <http://dx.doi.org/>.

