Chapter 11
Subcellular Systems

11.1 Protein Folding and "Info-statistical Mechanics"

The field of protein folding appears to have gone through a paradigm shift around 1995, largely due to the work of Wolynes and his group (Wolynes et al. 1995; Dill and Chan 1997; Harrison and Durbin 1985). The paradigm shift involves replacing the idea of *folding pathways* with the so-called folding funnel. In other words, the earlier notion of a denatured protein folding to its final native conformation through a series of distinct intermediate conformational states has been replaced by a new view, according to which an ensemble of conformational isomers (often called "conformers," not to be confused with "conformons"; a conformer can carry many conformons in it; see Sect. 11.3.2) of a denatured protein undergoes a transition to a final native conformation through a series of "ensembles" of conformational intermediates, each intermediate following a unique folding path to the final common native structure. In short, the paradigm shift is from *individual intermediate conformational isomers* of a protein to an *ensemble of the conformational isomers*, on the one hand, and from a *single folding pathway* to an *ensemble of folding pathways* (down the folding funnel), on the other.

Leopold et al. (1992) characterize the "protein folding funnel" as follows:

> . . . a kinetic mechanism for understanding the self-organizing principle of the sequence-structure relationship. This concept follows from a few general considerations. (i) Proteins fold from a random state by collapsing and reconfiguring [i.e., mainly conformationally rearranging polypeptides without breaking or forming covalent bonds: *my addition*], (ii) reconfiguration occurs diffusively [i.e., as a consequence of Brownian motions of proteins: *my addition*] and follows a general drift from higher energy to lower energy conformations, and (iii) reconfiguration occurs between conformations that are geometrically similar – i.e., global interconversions are energetically prohibitive after collapse – so local interconversions alone are considered. We define the folding funnel as a collection of geometrically similar collapsed structures, one of which is thermodynamically stable with respect to the rest, though not necessarily with respect to the whole conformation space . . ..

Just as water flows down a funnel, higher energy conformers (i.e., conformational isomers) of a denatured protein are thought to "flow" down the folding funnel

S. Ji, *Molecular Theory of the Living Cell: Concepts, Molecular Mechanisms, and Biomedical Applications*, DOI 10.1007/978-1-4614-2152-8_11,

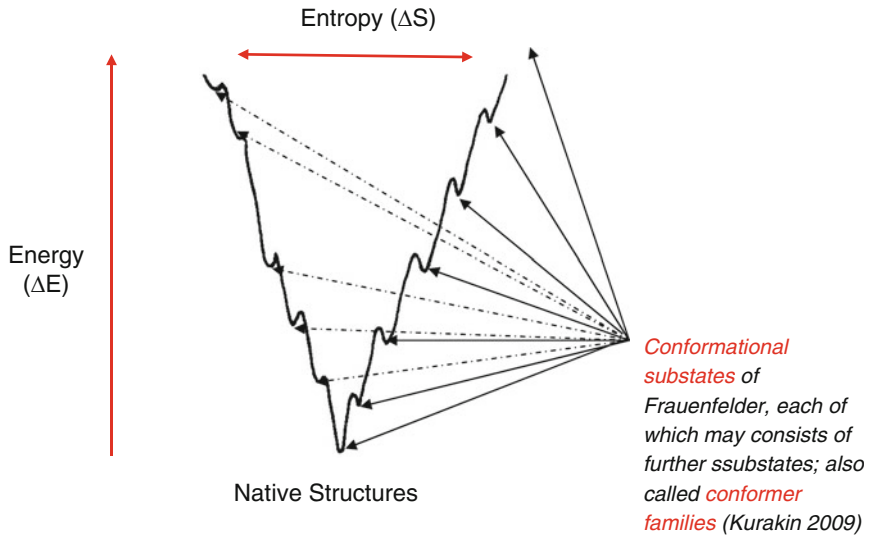


Fig. 11.1 The free energy landscape for protein folding (adopted from Brooks et al. 2009). The *vertical axis* encodes the energy changes, ΔE, accompanying a protein folding process, and the *horizontal axis* encodes the associated entropy changes, ΔS. The unfolded proteins have high potential energy and high entropy content, whereas the folded proteins have low energy and low entropy. Therefore the unfolded-to-folded transition leads to the so-called enthalpy-entropy compensation (i.e., the mutual cancellation between ΔE and TΔS) (Lumry 1974; Lumry and Gregory 1986) due to the mathematics of the Gibbs free energy change i.e., $\Delta G = \Delta E - T\Delta S$, when volume changes, ΔV, are negligible (see Eq. 2.1)

toward lower free energy conformers through several conformational states ("molten globular states," "transition state," "glass transition," "discrete folding intermediates," etc.) to the final native structure. The movement of protein conformers down the folding funnel is accompanied by two kinds of thermodynamic changes: (a) *energy* (i.e., Gibbs free energy under most conditions) decrease due to downward movement and (b) *entropy* decrease due to the narrowing of the funnel width, leading to increased conformational constraints (i.e., as conformations of a protein become more compact to minimize energy, the conformational motions of proteins become confined to an increasingly smaller volume, leading to a decrease in entropy).

Since protein folding is ultimately driven by Gibbs free energy changes under constant T and P conditions, $\Delta G = \Delta E + P\Delta V - T\Delta S$ (see Eq. 2.1 in Sect. 2.1.1), which becomes $\Delta G = \Delta E - T\Delta S$, if the pressure-volume work is negligible in protein folding, it would follow that, at some point along the vertical axis of the folding funnel, the free energy decrease ($-\Delta G$), due to energy decrease, $-\Delta E$, should exactly cancel out the free energy gain ($+\Delta G$) due to entropy decrease, $-\Delta S$, resulting in $\Delta G = 0$. At this point, spontaneous protein folding process would cease (except thermal fluctuations) and an equilibrium state established (see Fig. 11.1).

In April 2004, I had the privilege of attending a lecture on protein folding given by P. Wolynes at Rutgers. After his lecture, it occurred to me that the folding funnel theory as now formulated might lack a "biological dimension," because the theory seems to be based on the fundamental assumption that protein folding is driven by the tendency of proteins to minimize Gibbs free energy (cf., the "principle of minimal frustrations" [Bryngelson and Wolynes 1987]) in contrast to the alternative possibility that proteins in living cells have been selected by evolution not based on *free energy minimization* but rather based on their *biological functions*, regardless of their free energy levels. Their biological functions in turn would depend on their three-dimensional molecular shapes (Ji and Ciobanu 2003). When I asked Dr. Wolynes whether it would be possible to expand his two-dimensional folding funnel diagram (similar to the one shown in Fig. 11.1, wherein the *y*-axis encodes energy, E, and the *x*-axis encodes entropy, S) by erecting a *z*-axis perpendicular to the *xy*-plane to encode genetic information, I, the effects of biological evolution on protein folds, he did neither object nor explicitly endorse the idea. However, he did acknowledge the importance of taking into account biological evolution in theorizing about protein folding. Such an extension of the protein folding funnel model would bring protein folding processes within the purview of what was referred to as the *info-statistical mechanics* defined in Sect. 4.9.

The "folding funnel" model is also called "energy landscape" model. One way to incorporate biological evolution (and hence the genetic information) into the energy landscape theory of protein folding may be to identify the topology (i.e., surface shape) of the energy landscape as the extra dimension for encoding the effects of biological evolution. Although no proof is yet available, it seems that there may be a good correlation between the degree of the bumpiness or *frustrations* (Bryngelson and Wolynes 1987) of the energy landscape and the genetic information encoded in amino acid sequence of proteins, if the bumpiness somehow contributed to the fitness of the cell under given environmental condition and thus affected the I value (see below). The bumpier the surface of the energy landscape of a protein, the higher would be its information content of the Shannon type (Klir 1993). Thus, the notion of "bumpy folding funnel" may embody the following three elements

1. E, energy encoded in the depth of the funnel
2. S, entropy encoded in its width
3. I, genetic information encoded in the "bumpiness" or "ruggedness" of the funnel surface

The protein folding theory incorporating these three elements, E, S, and I, as described here may be referred to as the "information-energy landscape" theory of protein folding ("entropy" being included as a part of "energy," an abbreviation for "free energy") to contrast with the now widely accepted "energy landscape" theory of protein folding. It is my opinion that the "energy landscape theory" of protein folding is a physical theory and not a biological one, since there is no role (or room) for genetic information and, hence, biological evolution in it. To transform the energy landscape theory into a biological theory, it may be necessary to combine it

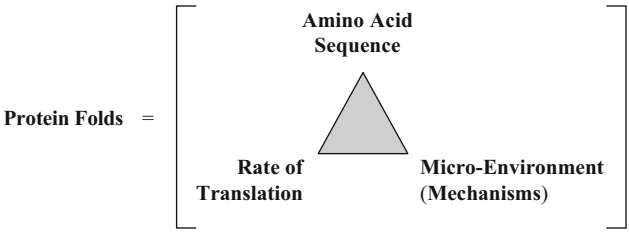


Fig. 11.2 A triadic model of protein folding inside the cell. According to this model, protein folds are dissipative structures (Sect. 3.1)

with a theory of biological evolution (in the form of say a cell model). One such biological theory is the "information-energy" landscape theory outlined here, which can be viewed as another manifestation of the *information-energy complementarity principle* presented in Sect. 2.3.2.

There are two contrasting views on the genotype-phenotype relation in protein folding: (a) the so-called Anfinsen's dogma that a protein's 3-D structure is completely determined by its amino acid sequence (but see Sect. 6.1.1) and (b) the opposite view that genotype has no *predictable* influence on three-dimensional protein folds. Belonging to the latter group is the *translation-dependent folding* (TDF) hypothesis recently proposed by Newman and Bhat (2007), according to which the differences in the rates of the translation of the mRNAs on the ribosomes can lead to different protein folds. In contrast to the TDF hypothesis of Newman and Bhat, Anfinsen's dogma may be referred to as the *sequence-dependent folding* (SDF) hypothesis. Although I support TDF hypothesis, it is deemed unnecessary to view the SDF hypothesis as completely irrelevant. Rather it appears to me that both SDF and TDF hypotheses are needed to formulate a coherent molecular theory of protein folding. One such effort is shown below which utilizes the definition of the function given in Sect. 6.2.11, the justification of which being that protein functions are determined by their folds, i.e., molecular shapes (Ji and Ciobanu 2003).

Workers in the field of protein folding have tended to think in term of thermodynamic principles only as exemplified by the *energy landscape model* of protein folding (see above) but Newman and Bhat (2007) suggest an alternative approach and emphasize the role of kinetics as being dominant in protein folding. As indicated in Fig. 11.2, I advocate a third view, the view that protein folding is an example of dissipative structure encompassing the following three irreducible elements:

1. Amino acid sequence (or the nucleotide sequence of mRNA)
2. Thermodynamic and kinetic factors determining the rates of translation of mRNA on the ribosome
3. Microenvironmental factors inside the cell that select a subset of translation rates compatible with the needs of the cell

It may be that these three aspects are inseparably "fused" together in the phenomenon of protein folding. We can only prescind (Sect. 6.2.12) one of them

at a time for the convenience of thought, but in reality they may all represent different aspects of the one and the same phenomenon we refer to as protein folding.

It is interesting to note that the *triadic model* of protein folding depicted in Fig. 11.2 is consistent with the protein folding mechanism deduced from an entirely different direction, namely, based on the *principle of constrained freedom*, the molecular version of the linguistic principle known as the *rule-governed creativity* (see Sect. 6.1.4).

11.2 What Is a Gene?

11.2.1 Historical Background

Prior to 2007, when the results of an international research effort known as the ENCODE (Encyclopedia of DNA Elements) Project was announced, the definition of gene was simple: *DNA segments encoding RNAs leading to protein synthesis* (Gerstein et al. 2007). But the ENCODE project has revealed numerous findings that cannot be readily accommodated by such a simple conception of a gene, and a new definition of a gene is called for. The failure of the pre-ENCODE conception of a gene may be traced ultimately to the following fact: *Biologists have been measuring the functions of genes (which belong to the class P of processes) and reduced the results to nucleotide sequences of DNA (which belongs to the class S of static structures) without specifying requisite mechanisms (which belongs to the class M of mechanisms)*. To stimulate discussions, such approaches in biological research where P is erroneously inferred from S alone may be referred to as the *P-to-S reduction error*. One way to resolve the problems revealed by the ENCODE project is to postulate that there are two equally important classes of genes – the S-genes and P-genes. The former is identified with the pre-ENCODE conception of genes (also called the Watson-Crick genes [Ji 1988]) and the latter is a new class of genes called the Prigoginian genes (Ji 1988). S-genes are analogous to *sheet music* (or written language) and P-genes are analogous to *audio music* (or spoken language) (see Fig. 11.3). Just as the sheet music is converted into audio music by a pianist, so are the Watson-Crick genes postulated to be transduced into Prigoginian genes by molecular motions driven by conformons, the sequence-specific conformational strains of enzymes (Chap. 8). Thus, conformons in biology may be analogous to the de Broglie equation in quantum physics, since mechanisms based on conformons can convert structure (S-genes) to processes (P-genes) in cells, just as the de Broglie equation can convert the particle-like properties of moving objects to their wave-like properties (Herbert 1987; Morrison 1990). If this analogy is true, it may be predicted that *conformons will resolve the structure-process paradox in molecular cell biology in the twenty-first century just as the de Broglie equation resolved the wave-particle paradox in particle physics in the twentieth century*.

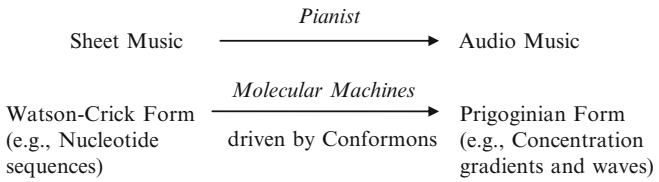


Fig. 11.3 The hypothesis that there are two forms of genes in the cell in analogy to music – (1) the Watson-Crick form analogous to *sheet music* and (2) the Prigoginian form analogous to *audio music*. The examples of the former include static nucleotide sequences of DNA and RNA, while the examples of the latter include the concentration gradients or waves of diffusible molecules such as RNA and mechanical deformations of biopolymers such as SIDDs in supercoiled DNA duplexes (see Sect. 8.3)

The history of the concept of the gene is at least one and a half centuries old (see Table 11.1) and yet its definition is still in flux. The concept of a gene as a heritable trait was established through the experiments with pea plants performed by the Austrian monk G. Mendel in the mid-nineteenth century. The term "gene" itself was coined by W. Johannsen in 1909 in analogy to the term "pangene" used by Darwin in formulating his ill-fated pangenesis hypothesis which Johannsen opposed (see Table 11.1). The modern idea of a gene as a sequence of nucleotides in DNA was firmly established by three main discoveries: (a) Avery, MacLeod, and McCarthy found that DNA is the carrier of genetic information, (b) Watson and Crick discovered that DNA is a double helix formed from two complementary strands of deoxyribonucleic acid intertwining with each other, and (c) the discovery by Crick, Brenner, Watts-Tobin, Nirenberg, Matthaei, and Khorana that triplets of nucleotides code for an amino acid. These and other historical facts about the evolution of the gene concept have been reviewed by Gerstein et al. (2007).

Unlike Gerstein et al. (2007), the history of the gene concept reviewed in Table 11.1 includes several additional facts that have to do with the energetic (E) aspect of the gene in addition to its information (I) aspect, as indicated by the last two columns labeled I and E. The decision to list both these aspects of the gene is based on the fundamental postulate that *a complete understanding of the nature of the gene on the molecular and cellular levels is impossible without taking into account both the informational and energetic aspects of the genetic processes* in line with the gnergy theory of self-organization (Sects. 2.3.2 and 4.13).

The idea that there must exist some material entity that is responsible for transmitting physical characteristics from one generation to the next (the process known as *inheritance*) must have occurred to the human mind from prehistoric times. This inference is based on the well-known facts that (1) children resemble their parents in appearance and (2) father's sperm must enter mother's womb for conception to occur. It would not have taken ancient humans too long to realize the next logical conclusion that the material entity mediating the transfer of traits from parents to offspring must be contributed not only by father's sperm but also by some material entity from mother later identified as eggs. Thus the concept of *heritable traits* must have been established in the human mind long before the Austrian

Table 11.1 A brief history of the gene concept. The following abbreviations are used: *I* informational aspect; *E* energetic aspect; *WC* Watson-Crick; *IDSs* intracellular dissipative structures (e.g., the Ca^{++} ion gradient in the cytosol); "=>" signifies "leads to," "indicates," or "causes"; + explained; − unexplained

Year	Author(s)	Key events and ideas	I	E
1. 1865	G. Mendel (1822–1884)	A *factor* that conveys traits from parent to offspring was hypothesized to exist in pea plant seeds	+	−
2. 1868	C. Darwin (1809–1882)	The *pangenesis* hypothesis was proposed as a possible mechanism for inheritance. Body cells were postulated to shed *gemmule* which collected in the reproductive organs before fertilization. Through gemmule, every cell's influence was thought to be transmitted from the parent to the offspring	+	−
3. 1886	H. de Vries (1848–1935)	The smallest particle representing one hereditary characteristic was named *pangene*	+	−
4. 1903	W. Johannsen	The terms *genotype* and *phenotype* were introduced	+	−
1909	(1857–1927)	The term *gene* was coined in opposition to *pangene* derived from Darwin's pangenesis hypothesis		
5. 1910	T. H. Morgan (1866–1945)	Genes occupy specific locations on chromosomes. The *beads on a string* model of genes	+	−
6. 1941	G. W. Beadle E. L Tatum	Mutations in genes were found to cause errors in specific steps in metabolic pathways, which led to the *one-gene-one-enzyme hypothesis*	+	−
7. 1943	E. Schrödinger (1887–1961)	The gene is a molecule capable of encoding genetic information and of executing it to form an organism	+	+
8. 1944	O. Avery C. M. MacLeod M. McCarthy	*DNA* carries genetic information	+	−
9. 1953	J. D. Watson F. Crick	The *DNA* double helix was discovered, which suggested possible molecular mechanisms underlying inheritance	+	−
10. 1958	F. Crick	The central dogma was proposed: DNA=> RNA => proteins => function	+	−
11. 1961	F. Crick S. Brenner Others	The *triplet genetic code* was discovered	+	−

(continued)

Table 11.1 (continued)

Year	Author(s)	Key events and ideas	I	E
12. 1969–1975	R. Lumry C. McClare W. Jencks	Enzymes can store mechanical energy as conformational trains to be utilized to drive molecular work processes (Ji 2000; Astumian 2001; Lumry 2009)	–	+
13. 1972	D. E. Green S. Ji	*Conformons*, defined as the mechanical energy stored in conformational strains of enzymes, were postulated to mediate energy transfer from respiration to ATP synthesis in mitochondria	–	+
14. 1972	W. Fries et al.	The first *nucleotide sequence* of a gene, *COAT_BPMS2*, was determined	+	–
15. 1976	M. Gilbert K. Mizuuchi M. H. O'Dea H. A. Nash	*DNA gyrase* was discovered that introduces *negative supercoils* into circular double-stranded DNA. DNA supercoils were later found to be essential for DNA functions, including transcription, replication, and recombination (Cozzarelli 1980; Reece and Maxwell 1991)	-	+
16. 1977	R. J. Roberts P. Sharp	*Exons* and *introns* were discovered. Genes could be split into segments, and one gene could make several proteins through alternative splicing	+	–
17. 1985 1988	S. Ji	*Conformons* were redefined as packets *of mechanical energy* and *genetic information* localized in sequence-specific sites within biopolymers and postulated to drive all goal-directed molecular movements inside the cell Two forms of genetic information are postulated to exist; *sequence information* (called Watson-Crick genes) and *time-varying patterns of chemical concentrations* (called *Prigoginian genes*) *Watson-Crick genes* = Equilibrium structures Sheet music Information transmission in time *Prigoginian genes* = Dissipative structures Audio music Information transmission in space	+	+

18. 1992	C. J. Benham	Superhelical coil–induced strand separations (called Stress-Induced DNA duplex Destabilizations, or *SIDDs*) occur in DNA duplexes predominantly at 5′ and 3′ *regulatory sites*	+	+
19. 1994	Stanford University scientists	The first genome-wide analysis of RNA functions was made possible through the invention of cDNA microarrays	+	+
20. 2007	ENCODE Project Consortium	Regions of the DNA coding for distinct proteins may overlap = > *genes are one long continuum* (ENCODE Project Consortium 2007; Gerstein et al. 2007)	+	−
21. 2009	S. Ji	(1) Genes = Dissipatons = Functions (2) Dissipatons = Equilibrons + Processes + Mechanisms (3) Prigoginian genes = WC genes + Free Energy (4) *Conformons* are *Prigoginian genes* residing in *biopolymers* (sequence information being WC genes) (Chap. 8) (5) IDSs are *Prigoginian genes* residing in the *cytosol* (molecular structures of solutes being WC genes) (Chap. 9)	+	+

monk, G. Mendel (1822–1884), established it as a scientific fact based on his experimental studies on heritable traits in pea plants in the mid-nineteenth century (See Row 1, Table 11.1). Mendel referred to the material entity or agent responsible for transmitting heritable traits as a "factor," which was later renamed as the "gene" in 1909 (Row 4, Table 11.1). The term "gene" was coined in opposition to "pangene," the smallest particle representing one hereditary characteristic within the framework of Darwin's ill-fated *pangenesis hypothesis* of inheritance (Row 3, Table 11.1), as already mentioned above. Darwin himself referred to the inherited material entity as *gemmule*, the material factor thought to be shed by every cell in the body that collects in the reproductive organs before fertilization (Row 2, Table 11.1).

Before proceeding further, it is important to stop and recall that, according to Prigogine's theory of irreversible thermodynamics (Babloyantz 1986; Kondepudi and Prigogine 1998; Kondepudi 2008), there are two and only two classes of structures in the Universe – *equilibrium* and *dissipative structures* (Sects. 3.1 and 3.1.5). The main difference between these two classes of structures is that *equilibrium structures* (or *equilibrons*) are stable within the time window of observations, requiring no expenditure of free energy, while *dissipative structures* (or *dissipatons*) are unstable and dynamic, absolutely dependent on a continuous utilization of free energy to be maintained. It is often the case that one equilibrium structure (e.g., candle) is transformed into another (e.g., CO_2 and water vapor) spontaneously, generating unstable intermediates in the process (e.g., reactive carbon-centered radicals) that self-organize into concentration patterns recognized by human eyes as dissipative structures. Thus it can be stated that dissipative structures presuppose equilibrium structures, that dissipative structures enclose equilibrium structures, or that equilibrium structures can exist without dissipative structures but dissipative structures cannot exist without equilibrium structures. These relations are embodied in the *triadic diagram* shown in Fig. 3.3.

The concept of the gene as the *material agent responsible for transmitting heritable traits from parents to offspring* as established by Mendel by the mid-nineteenth century is here claimed to belong to the class of *dissipative structures* because transmitting anything is a form of work requiring dissipation of free energy (Row 21, Table 11.1) and not to that of *equilibrium structures* as has been widely assumed by molecular biologists in recent decades. It is further suggested that the turning point when the original concept of the gene of Mendel was reduced to static sequences of nucleotides occurred around the mid-twentieth century when in agreement with Schrödinger's prediction, DNA was found to be the molecule encoding heritable traits (see Rows 7–11, Table 11.1).

The transmission of heritable traits from one generation to the next requires not only *information* (to be transmitted) but also *energy* dissipation to drive the information transmission as mandated by the principle of the *minimum energy requirement for information transfer* (see Shannon's channel capacity equation in Sect. 4.8). Of the 21 items listed in Table 11.1, 15 are concerned mainly with the informational aspect of the gene, 3 items (i.e., Rows 11, 12, and 15) are related to the energetic aspect, and only 3 items (Rows 17, 18, and 21) address both the energy

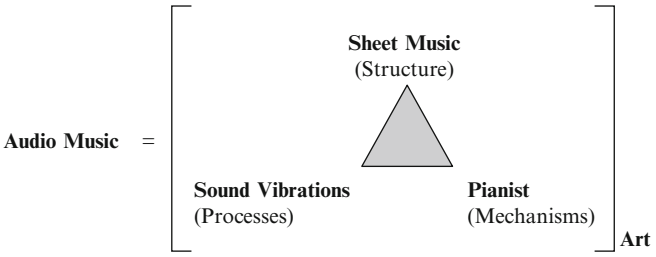


Fig. 11.4 Music as a Peircean sign. The piano music heard in a concert hall is an irreducible triad of sheet music, audio music, and pianist's performance

and information aspects simultaneously, consonant with the principle of information-energy complementarity discussed in Sects. 2.3.2 and 4.9.

One of the most significant conclusions given in Table 11.1 is that a gene is a dissipative structure selected by evolution for its functional value (see Row 21). This conclusion appears reasonable in view of the fact that functions in general are dissipative structures (Fig. 6.9) and *genes can be viewed as heritable functions*. The new conception of a gene and its relation to the traditional conception of genes as nucleotide sequences (or Watson-Crick form of genes) is discussed in Sect. 11.2.2 using Fig. 11.3. If correct, this will have many important consequences in our understanding of how the living cell works (see Items 2–5 in Row 21, Table 11.1).

*11.2.2 The Watson-Crick (*Sheet Music*) and Prigoginian Forms (*Audio Music*) of Genetic Information*

In Ji (1988), the concepts of *equilibrium* and *dissipative* structures were utilized to distinguish between two forms of genetic information: (a) the Watson-Crick form whose function was postulated to be to transmit information in *time* and (b) the Prigoginian form whose function was postulated to be to transmit information in *space*. It was then suggested that the former is akin to *musical scores* (i.e., sheet music) and the latter to *musical sounds* (i.e., audio music). Just as converting a *sheet music* to *audio music* requires a *pianist*, converting the genetic information encoded in *nucleotide sequences* to the *concentration gradients*, *waves*, or *trajectories* of chemicals inside the cell (e.g., RNA concentration trajectories shown in Fig. 11.6) requires *enzymes* acting as decoders and molecular motors such as RNA polymerase and topoisomerases:

Music can be viewed as a sign to human mind standing for human mood and emotion in the context of art. As such music can be represented using the same triadic template employed to represent the Peircean sign (Fig. 6.2) as shown in Fig. 11.4.

In analogy to music, a gene in the context of the cellular boundary conditions may be viewed as an irreducible triad of the Watson-Crick and the Prigoginian

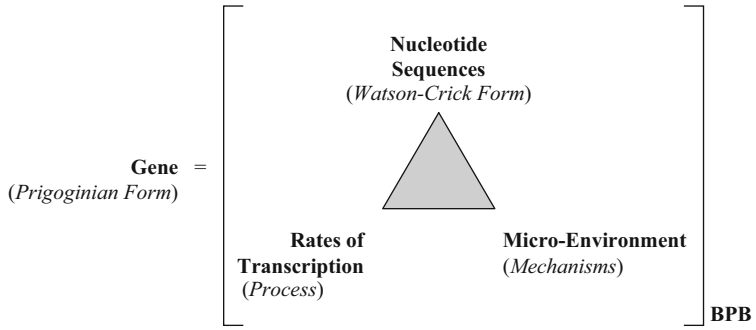


Fig. 11.5 The triadic definition of the gene based on the triadic model of function viewed as an irreducible triad of *structure*, *process*, and *mechanism* (Sect. 6.2.11). The subscript attached to the *lower right-hand side* of the bracket, BPB, stands for the "Bernstein-Polanyi boundaries" which control the mechanisms of selecting the functional processes out of all the processes allowed for by the laws of physics and chemistry

forms of genetic information and the complex mechanisms of transcription, translation, and associated processes in microregions within the cell (Fig. 11.5).

The Watson-Crick and Prigoginian forms of genes can be illustrated using Fig. 11.6 (see also Sect. 12.2) which depicts the kinetics of RNA levels (i.e., also called *ribbons*, RNA waves, or RNA trajectories) measured using DNA arrays (Sect. 12.2) in budding yeast undergoing the glucose-galactose shift (Garcia-Martinez et al. 2004; Ji et al. 2009a). The kinetic traces shown here contain two types of information: (a) the *sequence information* of individual RNA molecules, i.e., their open reading frames (ORFs), listed in the box on the right-hand side, and (b) the time-dependent *concentration levels* of the RNA molecules (i.e., ribbons, *RNA trajectories,* or *RNA waves*). The former can be identified with the *Watson-Crick form* (i.e., sheet music, equilibrium structures, or *equilibrons*) and the latter with the *Prigoginian form* (i.e., audio music, dissipative structures, or *dissipatons*) of genetic information, since the former is unaffected while that latter disappears when free energy supply to yeast cells is interrupted.

Since *dissipatons* encompass or embed *equilibrons* (see Fig. 3.4), it would follow that the Prigoginian form of genetic information embeds (or encompasses) the Watson-Crick form. This "encompassing" or "embedding" relation between the Watson-Crick form of genes and the Prigoginian form are represented using the *triadic diagram* derived from Peircean semiotics (Sect. 6.2), as shown in Fig. 11.5.

It may be significant that the definition of the gene given in Fig. 11.5 utilizes the same *triadic topological template* as used in defining Peircean signs (Fig. 6.2), *dissipative structures* (Fig. 3.4), functions (Fig. 6.10), and the model of protein folding (Fig. 11.2), indicating that *genes*, *dissipative structures*, *functions*, and *protein folding* all belong to the same class or category of entities, namely, the Peircean sign, which agrees with *Sebeok's doctrine of signs* described in Sect. 6.2. If the definition of the gene given in Fig. 11.5 is valid, genes can be described as follows:

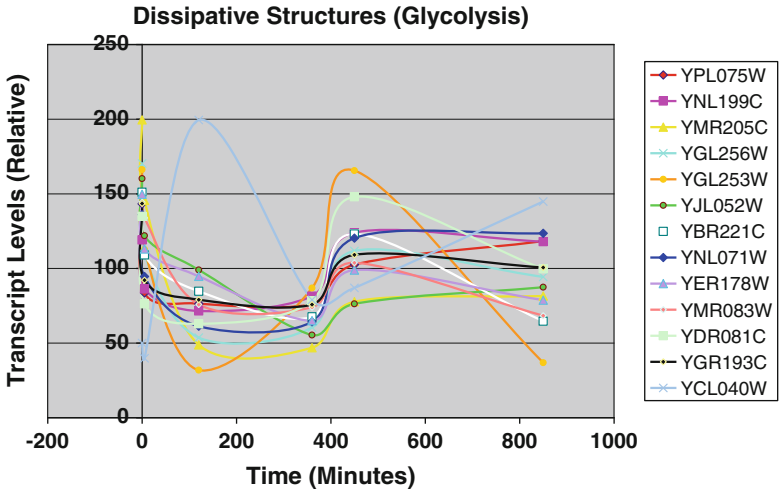


Fig. 11.6 The time-dependent RNA levels (i.e., RNA trajectories, *ribbons* or RNA waves) of the glycolytic genes of budding yeast measured with DNA arrays after switching glucose to galactose at t = 0 (Garcia-Martinez et al. 2004). RNA levels were measured in triplicates at six time points; t = 0, 5, 120, 360, 450, and 850 min, and the data points are the averages over the triplicates with coefficients of variations less than 50%. Of the 13 *ribbons* shown in this figure, the ribbon labeled YCL040W (*light blue*) exhibits an unusual behavior of increasing (rather than decreasing) its concentration between 5 and 120 min. This is most likely because the transcription of YCL040W gene is suppressed by glucose under normal physiological condition and the removal of glucose at t = 0 leads to glucose derepression of the gene (DeRisi et al. 1997; Johnston 1999; Ashe et al. 2000; Jona et al 2000; Kuhn et al. 2001)

> *Genes* are *functions* encoded in static *structures* (e.g., nucleotide sequences) which are activated into *processes* (e.g., transcription) by molecular *machines* that have been selected by evolution for their ability to fulfill the needs of the cell under a given environmental condition. (11.1)

For convenience, Statement 11.1 will be referred to as the *triadic model of the gene* (TMG) which can be logically derived from combining the concepts and theories embodied in Figs. 3.4, 6.10, and 11.3.

The concepts of *genotype* and *phenotype* were introduced by Johannsen in 1903 (see Row 4, Table 11.1). Johannsen was motivated to coin these terms to account for his finding that genetically identical common beans revealed normally distributed seed sizes (http://en.wikipedia.org/wiki/Wilhelm_Johannsen). Genotypes are heritable information stored in genes, while phenotypes are the observable manifestations of genes, on either macroscopic (e.g., morphology, behavior, biochemical properties of blood) or microscopic (e.g., amino acid sequences of proteins, RNA trajectories in cells) scale. Referring to Fig. 11.5, we can identify genotypes with the Watson-Crick form of genetic information and phenotypes with the Prigoginian form.

Phenotypes are commonly associated with stable structures such as the shape of seeds, height of plants, etc., while dissipatons (of which the Prigoginian form of genetic information is an example) are associated with unstable, dynamic structures

such as the flame of a candle and ion gradients across cell membranes. However, upon a closer examination, it is clear that neither of these initial impressions can be valid, since phenotypes can be unstable structures with short lifetimes such as RNA levels inside the cell (Fig. 11.6) and dissipatons can have stable structures with long lifetimes such as us human beings. So it appears necessary to specify lifetimes whenever we discuss dissipatons (or the Prigoginian form of genetic information). If we denote the lifetime of a dissipaton t_D, depending on the level of observations, t_D may range from nanoseconds (10^{-9} s) to years and centuries. Examples of the former include substrate levels/concentrations of enzymic catalysis (e.g., RNA levels) and neuronal firing patterns in the brain, and the examples of the latter include centuries (10^9 s)-old trees. Therefore, t_D can span 18 orders of magnitudes in time. Similarly the physical size of a dissipaton can range from about a micron (10^{-6} m) to tens of meters (10 m), thus spanning 7 orders of magnitude in linear dimension or 21 orders of magnitude in volume and mass. In other words, *dissipatons* can vary over about 20 orders of magnitude in both time and mass, the same order of magnitude spanned by the number of molecules in a macroscopic volume, i.e., the Avogadro's number, 6×10^{23}. Thus dissipatons and the Prigoginian form of genetic information can cover the whole range of space, time, and mass from the microscopic to the macroscopic, whereas the Watson-Crick form of genetic information always remains at the microscopic level (within the volume of about 10^{-20} m^3), most likely because self-replication of material systems is possible only at the molecular level where thermal fluctuations and the associated fluctuation-dissipation theorem (Ji 1991, pp. 50–51) become effective and operative (see Sect. 8.2 and Fig. 8.1 for the essential role played by thermal fluctuations in generating virtual conformons which are subsequently converted into real conforms by chemical reactions). In Sect. 5.2.3, snowflakes were compared with RNA trajectories in yeast cells which are microscopic dissipatons. This analogy can now be extended to the process of growth – just as snowflakes represent the process of the growth (i.e., crystallization) of an *equilibron* (i.e., a water molecule) from the microscopic to the macroscopic dimensions (see the left-hand side of Fig. 5.3), the human body may represent (or be the sign of) the process of the growth of a *dissipaton* from the microscopic level (i.e., cells) to the macroscopic level. In other words, *equilibrons* can grow from the microscopic level to the macroscopic one as exemplified by snowflake formation, so can *dissipatons* as exemplified by the development (or morphogenesis) of a mature human body from a fertilized egg.

11.2.3 The Iconic, Indexical, and Symbolic Aspects of the Gene

Although the birth of molecular biology was four decades away when Peirce (1839–1914) (Sect. 6.2.1) passed away, it is here assumed that the basic theory of signs he developed applies to molecular biology as well. According to Peirce

A sign (1) stands for something (2) to someone (3) in some context (4). (11.2)

We can conveniently represent Statement 11.2 diagrammatically as shown in Fig. 11.7. As pointed out in Sect. 6.2.1, the term "sign" has a dual meaning: (a) the

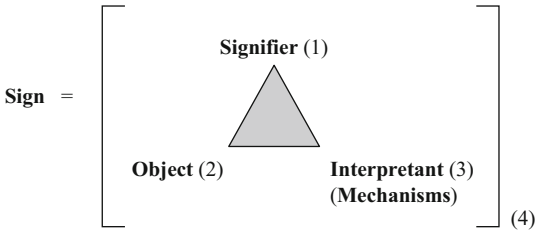


Fig. 11.7 A diagrammatic representation of the sign as a signifier, distinct from the sign as a function (see *left*). Unless otherwise noted, the term "sign" will be used in this book in the sense of signifier, following the common practice in the semiotics literature

Table 11.2 Three kinds of signs and their examples

Sign	Sign is related to object through	Examples
1. Iconic	Similarity (including structural complementarity)	Statues Portraits
2. Indexical	Causality (deterministic, one-to-one)	Smoke Weather vane
3. Symbolic	Conventions or codes (nondeterministic, one-to-many, arbitrary, creative, environment-sensitive)	Words Sentences Texts

sign as a representamen or sign vehicle (also called signifier) and (b) the sign as a function (i.e., semiosis) which is triadic as discussed in Sect. 6.2.11 (see Fig. 6.9). In Statement 11.2, as is the common practice in the semiotics literature, the term "sign" is used in the first sense.

Statement 11.2 is essentially the same statement as Statement 6.19 of Peirce (Buchler 1955, p. 99) and Statement 6.20 of Houser et al. (1998). Figure 11.7 is also essentially the same as Fig. 6.2.

Peirce distinguishes three kinds of signs as pointed out in Sect. 6.2.5 and summarized in Table 11.2.

It is assumed that Peirce's triadic division of signs is applicable to microscopic signs as well, lending support to the notion that semiotics of Peirce can be divided into two branches – *macrosemiotics* dealing with signs on the macroscopic level (e.g., words, texts) and *microsemiotics* concerned with signs on the microscopic level (e.g., molecules, DNA) (Ji 2001, 2002a). Some examples of signs, objects and sign processing mechanisms (or interpretants) at these three levels of material organization are listed in Table 11.3.

Row (1) is concerned with macrosemiotics which is a further elaboration of Table 11.2. Row (2) is about microsemiotics which is here identified with cell language (Sect. 6.1.2). Cell language (conveniently called *cellese*) is a hierarchically organized system of four sublanguages – DNA language (*DNese*), RNA language (*RNese*), protein language (*proteinese*), and biochemical language (*biochemicalese*). The question as to why the *cellese* consists of these multiple

Table 11.3 Sign processes (semiosis) at the macroscopic and microscopic levels. (1) = Sign, signifier, representamen, or sign vehicle; (2) = Object, signified, or referent; (3) = Sign processing mechanisms or interpretants; (4) = Environment or context

Levels		Sign (1)	Object (2)	Interpretant (3)	Context (4)
1. *Human language* (macrosemiotics)	Iconic	Statute	Person	Visual processing	Human brain
	Indexical	Smoke	Fire	Visual processing and reasoning	Human brain
	Symbolic	Words	Object signified	Language processing	Brain and society
2. *Cell language* (*cellese*[a]) (microsemitoics)					
(a) *DNA language* (*DNese*[a])	Iconic	DNA shape	RNA shape	Cell	Cell state[b]
	Indexical	DNA supercoil	Conformons[c]	Cell	Cell state
	Symbolic	Regulatory genes	Structural genes	Cell	Cell state
(b) *RNA language* (*RNese*[a])	Iconic	mRNA shape	tRNA	Cell	Cell state
	Indexical	RNA hairpins (?)	Conformons[c]	Cell	Cell state
	Symbolic	tRNA anti-codons	Amino acids	Cell	Cell state
(c) *Protein language* (*proteinese*[a])	Iconic	Active site shape	Ligand shape	Cell	Cell state
	Indexical	Virtual conformons[d]	Conformons[c]	Cell	Cell state
	Symbolic	Amino acid sequence	3-D protein folds	Cell	Cell state
(d) *Biochemical language* (*biochemicalese*[a])	Iconic	First messengers	Receptors	Cell	Cell state
	Indexical	Binding affinity	Virtual conformons of receptors	Cell	Cell state
	Symbolic	Acetyl-choline	Contraction	Smooth muscle cell	Cell state
			Relaxation	Cardiac muscle cell	Cell state
		Histamine	Contraction	Bronchial muscle cell	Cell state
			Relaxation	Vascular smooth muscle cell	Cell state

[a] The terms, *DNese*, *RNese* and *proteinese* were coined by a young American biochemist whom I met at the International Workshop on the Linguistics of Biology and the Biology of Language held in Cuernavaca, Mexico, in 1998, where I had presented the cell language ("cellese") theory, prior to the young biochemist's lecture in which he announced these neologisms. I am responsible for the coinage of *cellese* and *biochemicalese*

[b] The study of which is called "cytosociology" in Smith and Welch (1991)

[c] Conformational or mechanical energy stored in sequence-specific sites in DNA, RNA, or proteins (Chap. 8)

[d] *Virtual conformons* are the transient local conformational strains induced at a binding packet by thermal fluctuations of a protein which can be converted into real conformons upon exergonic binding of a ligand (Ji 2000) (Sect. 8.2)

sublanguages is not entirely clear to me but may be for the same reason that the computer language is hierarchically organized into multiple layers starting from the machine language level to (a) the micro architecture level, to (b) the instruction set architecture (ISA) level, to (c) the operating system machine level, to (d) the assembly language level, and finally to (5) the problem-oriented language level (Tanenbaum 2003). Just as all these computer languages are necessary in the computer architecture to facilitate the communication between humans and the computer, it is here postulated that the multiple sublanguages in the cell are necessary to couple the molecular (i.e., microscopic) processes (studied in molecular biology and enzymology) to the mesoscopic or macroscopic processes (studied in cell biology and animal physiology, respectively) that span spatial and temporal dimensions differing in scale by 5–15 orders of magnitude (see Statement 11.3). According to the *spatiotemporal scaling hypothesis* (Ji 1991, p. 56) quoted below, the generalized Franck-Condon principle may be ultimately responsible for the number of levels into which cell metabolism is organized:

> . . . Cellular metabolism implicates spatial structures ranging in size from 10^{-10} cm (diameter of the proton) to 10^{-5} cm (diameter of a cell), spanning 5 orders of magnitude, and in time from 10^{-9} s (electron transfer reactions) to 10^{6} s (cellular differentiation), spanning 15 orders of magnitude. To organize the intracellular processes that span such wide ranges of space and time, these processes may have to be "chunked" into manageable functional units, to each of which the generalized Franck-Condon Principle may apply. Various intracellular structures from DNA to enzymes to subcellular organelles may be viewed as a part of the cell's tactics for subdividing the spatial and temporal scales into optimal sizes for efficient control and regulation. (11.3)

We will refer to Statement 11.3 as the "Spatiotemporal Scaling Hypothesis of Biology."

Within each sublanguage, we can recognize *iconic*, *indexical*, and *symbolic* signs as indicated in Rows (2a)–(2d). When DNA is viewed as a molecular sign, its immediate object can be identified with either RNA for transcription or DNA itself for self-replication. The relation between DNA and its objects is *iconic* due to the structural similarity resulting from the Watson-Crick base paring, *indexical* due to the causal role postulated to be played by the mechanical energy stored in DNA supercoils in effectuating transcription factor binding (see the *TF-conformon collision hypothesis* discussed in Sect. 8.3), and *symbolic* due to the arbitrariness of the relation between the structural genes and the DNA segments regulating their expression (Ji et al. 2009b).

Similarly, within the RNese, mRNA is *iconically* related to tRNA through Watson-Crick base paring, *indexically* through the postulated causal role of conformons during the translation step catalyzed by ribosomes. The symbolicity in the interaction between mRNA and tRNA is probably not present but the relation between the anticodons located in the middle of a tRNA molecule and the corresponding amino acyl group located at the amino acid attachment site at the 3′ end of tRNA is clearly arbitrary and thus *symbolic*.

When proteins act as molecular signs, their objects can be identified with their cognate ligands which can be small molecular-weight organic or inorganic species or biopolymers such as DNA, RNA, carbohydrates, glycoproteins, or other proteins.

Proteins are related to their objects *iconically* due to their complementary molecular shapes and *indexically* due to the postulated requirement for converting virtual conformons to real conformons during binding processes, as required by the pre-fit hypothesis (Sect. 7.1.3). The symbolicity of *proteinese* may be best illustrated by the arbitrary relation between the primary and tertiary structure of a protein as summarized in the *postulate of the unpredictability of the 3-D protein fold* discussed in Sect. 6.1.1. Another example of the symbolicity of the proteinese is provided by the arbitrariness of the rate constants of the chemical reaction catalyzed by an enzyme, as exemplified by the single-molecule enzyme kinetics of cholesterol oxidase discussed in Sect. 11.3 (see Fig. 11.18).

In the *biochemicalese*, biochemicals are viewed as molecular signs (also called first messengers) and their objects can be receptors, second messengers, or gene-directed molecular processes that they trigger. The iconicity and the indexicality of the relations between first messengers and their cognate receptors are clear from the previous examples given in connection with the proteinese. The symbolicity of the relation between first messengers and the gene-directed intracellular processes is exemplified by the two opposite processes triggered by the same ligands, depending on cell types, as shown in the bottom four rows in Table 11.3.

When, say, a smooth muscle cell chooses to contract rather than relax when acetylcholine binds to its cell membrane receptors, what dictate the choice is not any thermodynamic changes resulting from ligand-receptor interactions but rather the cell states; smooth muscle cells contract while cardiac muscle cells relax, which reflects the history of biological evolution. Thus physical systems follow the laws of physics and chemistry (including thermodynamics) but living systems follow in addition the rules of biological evolution, most likely to increase the probability of reproduction. This is consistent with what H. Pattee (2001, 2008) refers to as the principle of matter-symbol complementarity.

Table 11.3 exemplifies the utility of Peircean semiotics in differentiating between physical systems and biological ones:

1. Molecules in physical systems act as iconic and/or indexical signs, whereas those in living systems act as symbolic signs in addition to iconic and indexical signs.
2. Iconic and indexical signs obey the laws of physics and chemistry.
3. Symbolic signs obey the rules (or codes) forged by evolution, which allows arbitrariness or *freedom within* (not beyond) the constraints of the laws of physics and chemistry (e.g., see "phenotypic freedom with genotypic constraints" discussed in Sect. 12.10).

11.2.4 Three Kinds of Genes: drp-, dr-, and d-Genes

The new concept of genes (as dissipatons and functions as defined in Fig. 11.5 and Statement 11.1) views the traditional conception of genes (as protein-coding DNA segments) as projections of the functional genes onto the three-dimensional Euclidean space. Since DNA can serve as the template for self-replication, the DNA

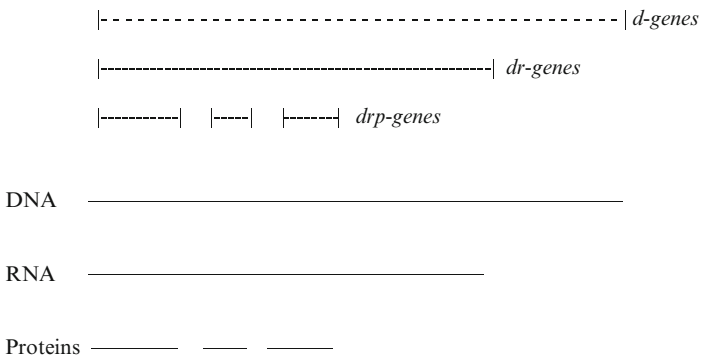


Fig. 11.8 The "overlapping triadic gene" (OTG) model of DNA

Table 11.4 The postulated hierarchical genetic architecture of DNA

New names	Traditional names
1. d-genes	Junk DNA, noncoding DNA, regulatory regions, silencers, enhancers
2. dr-genes	Introns, RNA-coding regions
3. drp-genes	Structural genes, exons

molecules as a whole can be defined as a gene, to be denoted as the *d-gene*. It is now known that about 30% of the human DNA molecule serves as the template for RNA, leading to the concept of dr-gene, namely, the DNA segments encoding not only DNA but also pre-mRNA. As is also well known, only less than 3% of the total DNA mass of the human genome codes for proteins and such DNA segments can be referred to as the drp-genes, since they code for not only proteins (p) but also DNA (d) and RNA (r). So the genetic architecture of DNA can be schematically depicted as shown in Fig. 11.8.

According to Fig. 11.8, exons are drp-genes and introns are dr- or d-genes. The parts of the DNA molecule not included in the dr-gene category are denoted as d-genes, which can be identified with the so-called noncoding DNA, including promoters, silencers, and enhancers. It was suggested elsewhere that noncoding DNA, i.e., d-genes, play a role in chromatin remodeling and hence in timing of gene expressions, driven by conformons stored in DNA double helix (Ji 1991, 1999b, 2002b). So I am suggesting here a completely new way of classifying genes in DNA structure as summarized in Table 11.4.

The DNA of the cell is often referred to as the *book of life*. This metaphor was analyzed in detail elsewhere (Table 8 in Ji 2002a), which is reproduced as Table 11.5, with the last row (starting with drp-genes) added anew.

A book contains more than just a set of words. In addition, it contains the *information* about the *order* in which the words are arranged in a linear series from the opening to the last page. The literary information and skills that were needed to select and order the words in a given pattern (i.e., a network topology [Barabasi 2002]) in a book came from the brain of the author who is physically

Table 11.5 An analysis of the DNA-book-of-life metaphor. The symbols p-, r-, and d- in the last row stand for protein, RNA, and DNA, respectively

	Words	Ordering information	Information processor
Book	Dictionary	Author's brain	Author/reader
DNA	Coding regions: "Lexical Genes" (Ji 1999b, 2002a) *drp-genes* (this book)	Noncoding regions: Semantic genes (Ji 1999b, 2002a) *dr-* and *d-genes* (this book)	The cell

absent in the book he produced. In other words, what are necessary to produce a book are *words* and the author's *literary skills*, since words alone cannot self-arrange into a book. The interesting question that arises is: What in DNA corresponds to the author's literary skills? Based on then-available data it was conjectured in Ji (1991, 1997a, b, 1999b, 2002b) that the information required to order gene expression in *time* and *space* is encoded in the so-called noncoding (or silent) regions of DNA that do not code for proteins. It should be pointed out that space and time are postulated to be inseparably linked on the molecular level in cells due to Brownian motions and the operation of the generalized Franck-Condon principle discussed in Sect. 2.2.3 and in Ji (1991, pp. 52–56). These and related ideas are summarized in Table 11.5.

The so-called regulatory RNA molecules recently reviewed by Mattick (2003, 2004) can be viewed as indirectly supporting the concept of *spatiotemporal genes* that was postulated in 1991 to be located in non-protein-coding DNA (Ji 1991), although I was not aware then that at least a part of such non-protein-coding DNA could code for regulatory RNAs (cf. dr-genes). The *spatiotemporal gene hypothesis* formulated in Ji (1991) can now be updated as follows:

> DNA molecules contain drp-, dr-, and d-genes that are located, respectively, in:
>
> (1) Protein coding regions (1–3% of the total DNA mass in humans [Mattick 2004])
> (2) RNA coding regions (~30% of the total human DNA mass)
> (3) DNA-coding regions (the total human DNA mass)
>
> The term "DNA-coding genes," or d-genes, is self-referential, since DNA itself, without being first transduced to RNA, is postulated to transmit genetic information as during the replication step of the cell cycle. The d-genes include cis-regulatory regions, enhancers, and silencers as already indicated above (Tjian 1995). It is further postulated that drp- and dr-genes primarily encode *equilibrium* structures and d-genes primarily encode dissipative structures consisting of conformationally deformed DNA. (11.4)

Statement 11.4 will be referred to as the postulate of the "triadic overlapping genomic architecture TOGA)," the triadicity stemming from the trichotomy of the drp-, dr-, and d-genes and "overlapping" referring to the fact that drp genes code for DNA, RNA, and proteins and dr genes code for DNA and RNA. Statement 11.4 clearly is a "spatiotemporal gene" hypothesis, because the drp- and dr-genes carry spatial (or geometric) information specifying protein and RNA structures,

whereas d-genes are thought to carry *timing* information about *when* and *for how long* target genes are to be expressed in the nucleus, which will lead to the production of various dissipative structures inside the cell. The letter "d" in d-genes can be viewed as "dual" in the sense that it can be interpreted either as "dissipative structure-forming" or "DNA," which may be viewed as being consistent with the *information-energy complementarity principle* (Sect. 2.3.2).

An interesting difference between the *spatiotemporal gene hypothesis* (Ji 1991) updated in Statement 11.4 and the *regulatory RNA hypothesis* of Mattick (2003, 2004) is that the former endows the non-RNA-coding regions of DNA with a full gene status (i.e., as d-genes) storing the conformons with timing information about gene expression, whereas the Mattick hypothesis endows the timing information only to RNA-coding DNA segments and does not explicitly specify any biological role for non-RNA-coding DNA regions which accounts for more than a half of the total DNA mass in the human genome.

It is also possible that d-regions of DNA carry information required to control or effectuate *long-range correlations* within DNA (and hence indirectly within RNA and proteins as well), leading to the couplings between *the very small* (e.g., ions, atoms) and *the very large* (e.g., cell behaviors including shape changes, and cell migration), and *the very fast* (e.g., time constant in the range of 10^{-12} s characteristic of covalent bond rearrangement events occurring locally on DNA) and *the very slow* (e.g., time constants in the rage of years or decades, $10^8 \sim 10^9$ s, associated with DNA sequence evolution by genetic drift) (Ji 1991, pp. 52–56). Thus, it is predicted that the d-genes will be found to play a critical role in effectuating *long-range spatiotemporal correlations* both within DNA molecules and within cells generally in order to transduce genetic information to various intracellular dissipative structures (IDSs), the final form of gene expression according to the Bhopalator model of the cell (Fig. 2.11).

Fink et al. (2007) recently analyzed the variations of the coding (C) and noncoding (N) DNAs for 800 prokaryotic and eukaryotic species (see Fig. 11.9). The double-logarithmic plot of N against C of these species produced a straight line passing through mostly prokaryotic species (67) with a slope of 1.07 and a set of lines that can be drawn through eukaryotic species (733) with an average slope of 4.33 (individual slopes ranging from 2 to about 10). Fink et al. (2007) proposed that eukaryotes require a certain minimum amount of noncoding DNA which increases with the amount of coding DNA. This allowed them to account for the slope of about 2 in Fig. 11.9 (see the slightly curved solid line) but left unexplained the wide upward divergences of data points from this lower bound.

Based on the spatiotemporal gene hypothesis, Statement 11.4, we can provide alternative explanations for the distribution patterns of the data points shown in Fig. 11.9:

There exists a general power law relation between N and C:

$$\mathrm{N} = \alpha \mathrm{C}^{w} \tag{11.5}$$

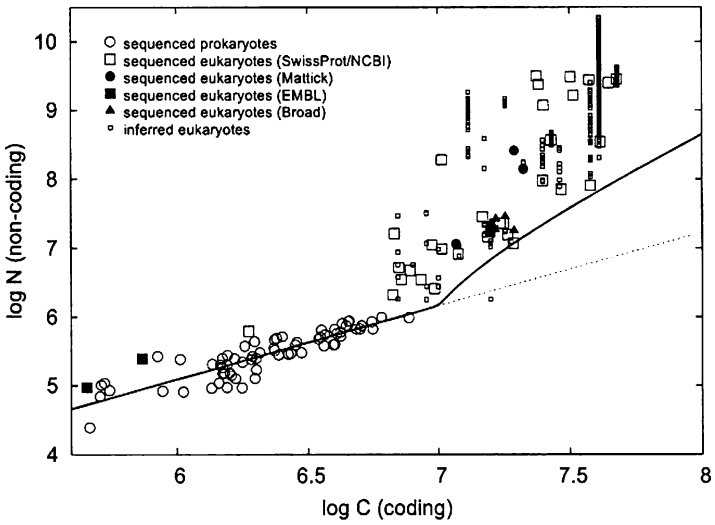


Fig. 11.9 The relation between the amounts of coding (C) versus noncoding (N) DNA of 67 prokaryotic and 733 eukaryotic species in units of base pairs (Reproduced from Fink et al. 2007)

where α is proportionality constant and w is a "critical exponent" postulated to be unique to organisms, reflecting the functional characteristics of their genomes.

For the prokaryotic genomes which consist of mostly drp genes and are devoid of any dr- and d-genes:

$$w = 1 \tag{11.6}$$

For eukaryotic species with only drp- and dr-genes:

$$w = 2 \tag{11.7}$$

For eukaryotic species with drp-, dr-, and d-genes:

$$2 < w < 10 \tag{11.8}$$

It is interesting to note that drp- and dr-genes each contribute one unit to the numerical value of w, leaving d-genes to contribute one to about eight units to the numerical value of w. It is suggested here that the numerical values of w greater than 2 be referred to as "excess critical exponent (ECE)." Although the physical meaning of ECE defined here is not yet clear, ECE may be related to the *number of subunits in multisubunit complexes of enzymes* such as those participating in signal transduction networks, translation, protein degradation, transcription (cf. transcriptosomes), transcript degradation (cf. degradosomes), and nuclear pore complexes. The number of the subunits constituting an enzyme complex ranges from 30 to 100. If this interpretation of ECE is correct, there may be a profound difference between

a *simple enzyme* and an *enzyme complex* besides their physical sizes, indirectly supporting the conclusion drawn in Table 4.7, where the difference between enzymes and enzyme complexes was compared to the difference between atoms and quantum dots.

In addition, the explanations suggested here are consistent with the general hypothesis proposed in (Ji and Zinovyev 2007b; Ji 2007a) that *life is a critical phenomenon*, wherein long-range interactions among microscopic entities or particles lead to macroscopic behaviors of organisms (see also Sect. 16.7). If this hypothesis is correct, the numerical values of w ranging from 1 to 10 revealed in Fig. 11.9 may be analogous to the variable numerical values for the critical exponents measured in condensed matter physics which vary from 0.1 to about 5 (Domb 1996; Landau and Lifshitz 1990).

The DNA-book analogy described in Table 11.5 can be expanded by replacing "book" with "survival manual." That is, DNA can be more accurately viewed as the *molecular record* of all the instructions found useful for cell survival throughout the evolutionary history of a lineage. In other words, DNA can be viewed as *survival programs* or *algorithms* written in atoms and hence may be referred to as the "atomic survival programs (ASP)" or "atomic survival algorithms (ASA)." The maximum number of ASPs stored in the human genome may be estimated as shown in Eqs. 11.9 and 11.10, if we can assume that (a) protein-coding genes (whose number will be denoted as a, which is about 30,000, in the human genome) are equivalent to words in human language; (b) the average number of words, b, in a cell-linguistic sentence is about 10; and (c) the average number, c, of cell linguistic sentences in an ASP is 5.

$$\begin{aligned}\text{Log (Maximum Number of ASPs)} &= \text{c log } a^{b} \\ &= 5 \log (3 \times 10^{4})^{10} \\ &= 5 \times 40 \log 3 \\ &= 95.42 \end{aligned} \tag{11.9}$$

$$\text{Maximum Number of ASPs} = 10^{95.42} \tag{11.10}$$

The information, I, required to retrieve one of these ASPs from the nucleus may be estimated as in Eq. 11.11:

$$\begin{aligned} \mathrm{I} &= \log_2\left(10^{95.42}\right) \\ &= \left(\log_{10} 10^{95.42}\right)/(\log_{10} 2) \\ &= (95.42)/0.303 \\ &= 315 \text{ bits} \end{aligned} \tag{11.11}$$

Equation 11.11 indicates that the cell will need 315 bits of information to be able to select one ASP out of all possible ASPs stored in DNA. With this information the cell will be able to make 315 binary choices or decisions. In order for

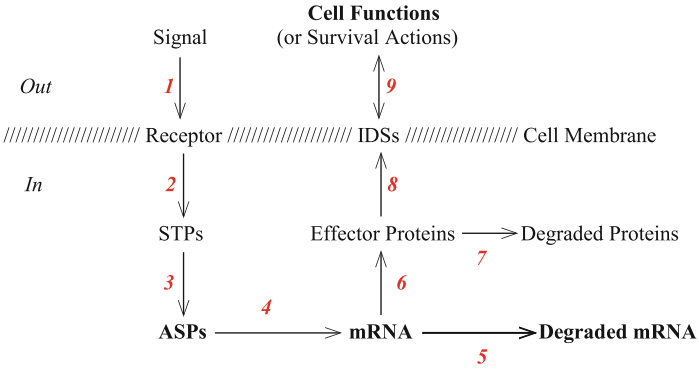


Fig. 11.10 The cell viewed as a system (or network) of *molecular machines* that has evolved to carry out two main functions: (1) *signal transduction* (see Steps 1, 2, and 3) to retrieve select "atomic survival programs" (ASPs) stored in DNA,and (2) *gene expression* (see Steps 4, 6, and 8) to transduce ASPs into survival actions or mechanisms, otherwise known as cell functions. See Figs. 12.34 and 12.35 for more detailed signal transduction pathways

the cell to be able to make such complex decisions, it would need similarly complex internal states, according to the Law of Requisite Variety (see Sect. 5.3.2), and we can identify the signal transduction pathways as the cellular machinery carrying out the required selective actions. So the well-recognized complexities of signal transduction pathways can now be understood as resulting from the cell's need to retrieve select ASPs from the nucleus and to execute them into actions in order to survive under very complex environmental conditions that are constantly changing. We may represent this idea schematically as shown in Fig. 11.10:

Step 1 is where the external signal is recognized through the receptors embedded in the cell membrane or nuclear receptors present in the cytosol and the nucleus (not shown), and in Step 2 the original signal is transformed to retrieve appropriate information from DNA in the nucleus in Step 3. In Step 4, selected ASPs are retrieved from chromosomes and transcribed into mRNA expressed at right times and for right durations. It is important to note here that mRNA levels are functions of not only the rates of transcription (Step 4) but also of the rates of transcript degradation (Step 5). Ignoring this simple fact has led to many false positive and false negative conclusions in the field of microarray technology as pointed out in Ji et al. (2009a). Similarly, effector protein levels are functions of both translation (Step 6) and protein degradation (Step 7). Step 8 represents catalysis, the all-important step where chemical reactions are catalyzed to generate free energy that drives all intracellular molecular processes. Abbreviations are as follows: STP = signal transducing proteins; IDSs = intracellular dissipative structures such as intracellular ion gradients (Chap. 9) and mechanical stress gradients of cytoskeletons (Ingber 1998). Effector proteins include transcription initiation/elongation complexes (acting on Step 4 thus constituting a feedback loop), molecular motors and pumps, kinases, phosphatases, synthetic enzymes, cytoskeletons, etc.

The double-headed arrow in Step 9 indicates that IDSs and survival actions are synonymous or identical. This idea has been referred to as the *IDS-cell function identity hypothesis* (Sect. 10.2).

11.2.5 Two-Dimensional Genes: The Quality-Quantity Complementarity of Genes

The currently most widely accepted definition of a gene is "a segment of DNA that encodes an RNA or a protein molecule." This definition was probably established around 1961 when the *triplet genetic code* was discovered (see Table 11.1 and Crick et al. 1961). Such a definition of a gene is no longer adequate to completely account for what we can now measure about a gene. For example, the DNA microarray technique invented in the mid-1990s (see Sect. 12.1) measures two aspects of a gene simultaneously: (a) the nucleotide sequence and (b) the copy number (also called abundance, levels, or concentrations) (see Fig. 11.6). The former reflects the qualitative aspect of a gene and the latter reflects the quantitative aspect, and *quality* and *quantity* are complementary to each other in the Bohrian sense (see Sect. 2.3.1). Thus a gene may be viewed as the complementary union of the two irreconcilably opposite properties, sequence, and copy number, just as light can be regarded as the complementary union of irreconcilably opposite waves and particles (see Sect. 2.3.5). We may represent this idea algebraically as shown in Eq. 11.12:

$$\textbf{Gene} = [\textbf{S},\ \textbf{CN}] \tag{11.12}$$

where S is the nucleotide *sequence* and CN is copy number. In Fig. 11.6, Ss are given in the column on the right-hand side of the figure, and CNs (in relative units) are indicated on the *y*-axis. Eq. 11.12 also indicates that a gene is two-dimensional, since a gene is a function of two independent variables, S and CN. Eq. 11.12 also applies to gene products, RNAs, and proteins.

$$\textbf{RNA} = [\textbf{S},\ \textbf{CN}] \tag{11.12a}$$

$$\textbf{Protein} = [\textbf{S},\ \textbf{CN}] \tag{11.12b}$$

The microarray data measured in budding yeast undergoing glucose-galactose shift (described in Sect. 12.3) demonstrate that there is a good correlation between metabolic functions of the cell and the kinetic behavior (or trajectory) of an mRNA molecule (Sect. 10.2), leading to the following triple correlation:

$$\textbf{Gene X} \sim \textbf{mRNA X} \sim \textbf{CellFunction X} \tag{11.12c}$$

where X is the name of S such as "ORF (open reading frame)" or "ATPase," etc., and the symbol "A ~ B" reads "A is associated with B" or more specifically "A is the necessary condition for B." To emphasize the latter sense which is "hierarchical" or "ordinal" (i.e., without Gene X, no mRNA X; without mRNA X, no Cell Function X), Eq. 11.12c can be rewritten as:

$$\textbf{Gene X} < \textbf{mRNA X} < \textbf{CellFunction X} \tag{11.12d}$$

where the symbol "<" indicates the hierarchical or ordinal relation. Because of the importance of Inequality (11.12d) in interpreting microarray data, it may be justified to refer to Inequality (11.12d) as the *Principle of the Ordinal Relation among Genes, RNAs, and Cell Functions* or *PORAGRF*. One practical consequence of PORAGRF is this: Although genes, RNAs, and cell functions are often given the same name (e.g., "glycolysis gene," "glycolysis mRNA," "glyclysis proteins"), they cannot be *equated* or *interchanged*, ultimately because a gene and its products are not one dimensional entities but two-dimensional so that, although they have the same S, they can have different CNs (see Eqs. 11.12, 11.12a, 11.12b).

Most of the papers on microarray data experiments that have been published since the beginning of the microarray era in the mid-1990s have made false positive and false negative conclusions (Ji et al. 2009a). The reason for such errors may be traced ultimately to the following two erroneous assumptions which are interconnected:

1. "Gene X" can be replaced by "mRNA X" (violating PORAGRF, Inequality (11.12d)).
2. Microarrays measure the expression of "gene X" because "gene X" is synonymous with "mRNA X."

11.2.6 DNA and RNA as Secondary and Primary Memories of the Cell

To understand the role of RNAs in cell functions, it may be instructive to use the digital computer as a model, in agreement with the so-called Simpson thesis discussed in the Preface that biology is the study of phenomena to which all principles apply, including most likely the architectural principles of modern-day digital computers. It is interesting in this connection to point out that Wang and Gribskov's (2005) made a theoretical comparison between the digital computer and the living cell in a complementary way to the comparison made in Ji (1999a). According to Wang and Gribskov, DNA is analogous to the *secondary* memory and RNA to the *primary* memory of computers (Fig. 11.11).

One of the crucial differences between the primary and secondary memories of the computer is that the former is dynamic and disappears when the computer is turned off, whereas the secondary memory (e.g., stored in hard drives, CDs, etc.)

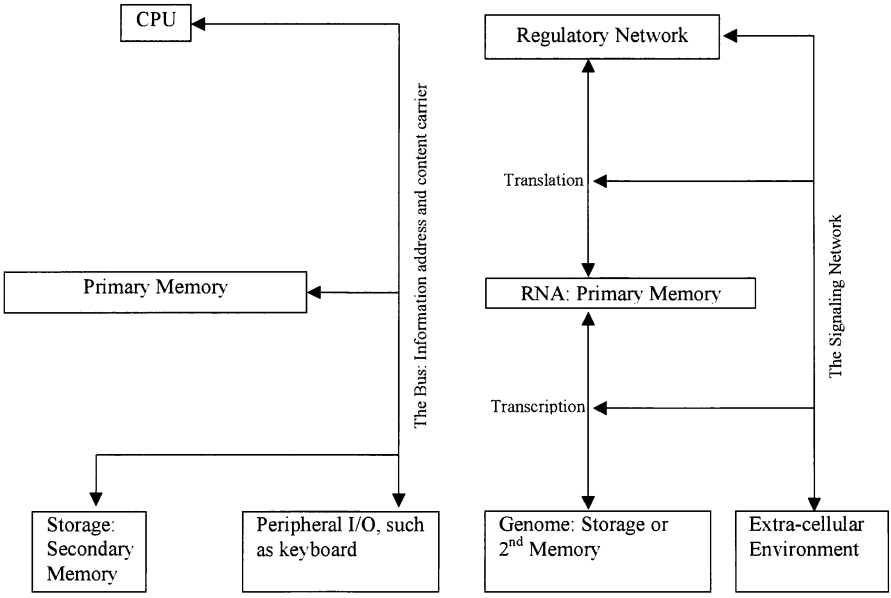


Fig. 11.11 A comparison between the architecture of the computer and the living cell (Reproduced from Wang and Gribskov 2005). *CPU* central processing unit, *I/O* input/output device

is stable even when the computer is turned off. This is reminiscent of the metabolic difference between RNA and DNA in cell biology; the former levels are changing with cell functions, disappearing when the free energy sources are turned off (Garcia-Martinez et al. 2004; Ji et al. 2009a), while the latter is relatively stable in cells, maintaining its sequence information even when cells are deprived of free energy. The dynamic nature of RNA levels in cells is the consequence of the fact that RNA is both produced via the transcription process and degraded via the transcript degradation process simultaneously and with comparable kinetic constants (see Figs. 11.6, 12.1, 12.2). Another interesting similarity between the computer and the cell is that, just as the CPU of the computer cannot utilize the information stored in the secondary memory without first converting it to the primary one, the cell cannot utilize the genetic information encoded in DNA without first converting it to RNA and then to proteins, since the utilization of DNA information requires *free energy dissipation* which in turn requires proteins acting as catalysts for free-energy supplying chemical reactions (Fig. 11.12).

Another way to describe the fundamental difference between DNA and RNA is in terms of the concepts of *equilibrium* and *dissipative* structures (Babloyantz 1986; Kondepudi and Prigogine 1998; Prigogine 1977, 1980; Kondepudi 2008). DNA is an equilibrium structure and RNA levels are dissipative structures, for the obvious reason that the former remains and the latter disappears when free energy input into the cell is blocked as mentioned above (Sect. 3.1). The concentration of DNA in cells remain more or less constant except during the S phase in the cell cycle,

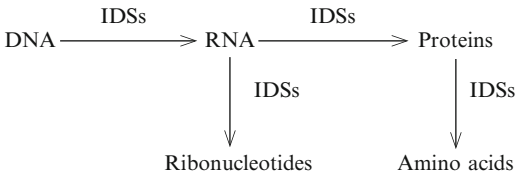


Fig. 11.12 A simplified representation of the Bhopalator model of the living cell, highlighting the relationship among the key structural components – DNA, RNA, proteins, and IDSs, and the last including biochemical concentration gradients

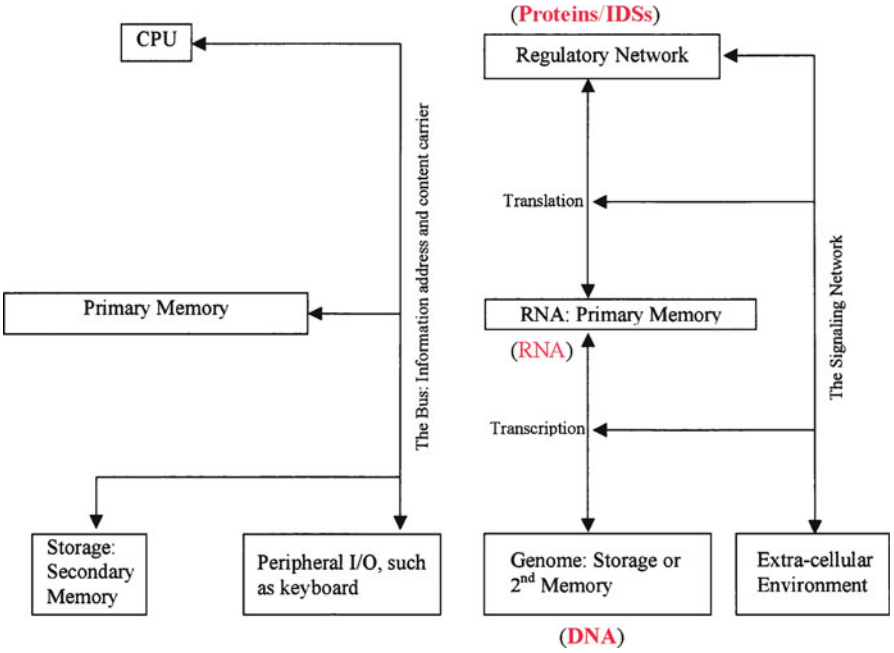


Fig. 11.13 A comparison between the Wang-Gribskov model (*left*) of the cell and the Bhopalator (*right*). The main difference is the concept of IDSs which is absent in the former model of the cell. See text for more details

whereas the concentrations of RNA molecules, both large such as mRNA and small such as siRNA (small interfering RNA), can change dramatically while cells carry out their functions in interaction with their environment throughout the cell cycle.

There is one important difference between the Wang-Gribskov model of the cell shown on the right-hand side of Fig. 11.11 and the Bhopalator model of the cell shown in Figs. 2.11 and 11.13. In the former, what corresponds to CPU of the computer is "regulatory network" and proteins (Wang and Gribskov 2005), whereas what corresponds to CPU in the Bhopalator model is the combination of proteins and IDSs, the final form of gene expression. One of the most interesting points of

the Wang-Gribskov model of the cell is the importance placed on RNA as the primary memory relative to DNA viewed as the secondary memory, which may seem to go against the DNA-centered paradigm in cell biology that has been dominating the field of molecular biology since the discovery of the DNA double helix in 1953. But this shift in emphasis from DNA to RNA embodied in the Wang-Gribskov model of the cell is consistent with the emerging prominence of RNA not only as the reliable indicator of biological complexity (Mattick 2004) but also as the regulators of cell metabolism and functions (Mattick 2003, 2004; Storz 2002; Zamore 2002). One of the reasons for the prominence of RNA over DNA in cell biology may derive from the fact that RNA can serve as an enzyme but DNA cannot. Thus RNA may have played an essential role in the origin of life because of its ability to manipulate both *energy* and *information* and thus control organization (see Fig. 4.8).

11.2.7 Cell Architectonics

The study of the principles and theories of the architecture of the living cell may be referred to as the *cell architectonics*. Just as a human architect designs a building to achieve a set of functions desired by the occupant, the biological evolution has designed the living cell to accomplish a set of functions that are essential for self-reproduction or self-replication. It is important to recognize the fundamental difference between *artificial buildings* and *natural cells*, however. The former is *other-designed* and the latter is *self-designed* and *self-organizing* (Sect. 3.1).

If we designate the *d*iameter of a *c*ell as DC and the average *d*iameter of the *p*articles inside the cell as DP, the *n*umber of *p*articles inside the *c*ell, NPC, can be approximated as $\mathrm{NPC} = (\mathrm{DC}/\mathrm{DP})^3$. Typically, a eukaryotic cell has a diameter of 10^{-5} m, and the average diameter of cell components (e.g., protons, oxygen, water, metal ions, metabolites, proteins, RNAs, and DNA) can be estimated to be 10^{-9} m, leading to 10^{12} as the approximate total number of particles inside the cell. These particles are not randomly distributed inside the cell but organized in space and time to accomplish a set of functions that are beyond individual component particles. In other words, to produce the properties essential for cell survival that are beyond the capabilities of the components of the cell, these components must be organized into higher-order structures according to the principle referred to as the Principle of Emergent Properties (PEP) (Ji 1991). PEP states that:

> Whenever a complex system S is constructed out of n sub-systems, s_1, s_2, s_3, …, s_n, according to a set R of m rules, r_1, r_2, r_3, …, r_m, then a set P of k properties, p_1, p_2, p_3, …, p_k can emerge that is not found in less complex systems composed of any subset of S. (11.13)

To organize the large number (~10^{12}) of particles inside the cell, the following two requirements must be met.

Thermodynamic Requirement: Most, if not all, molecular motions of the intracellular particles must be driven away from random directions and random durations, utilizing the free energy supplied by exergonic chemical reactions (e.g., ATP hydrolysis), since random motions are incompatible with life. (11.14)

Control Information Requirement: The non-random molecular motions of the intracellular particles must be constrained in space and time by the *boundary conditions* that embody both the genetic information (or "internal constraints") and environmental information (or "external constraints"). (11.15)

Requirement 11.14 is met by the *conformon theory of molecular machines* (Ji 1974a, b, 2000, 2004a), according to which all molecular machines (e.g., molecular motors, ion pumps, and enzymes) inside the cell are driven by sequence-specific conformational strains called *conformons* that are generated from exergonic chemical reactions or ligand-binding processes based on the generalized Franck-Condon mechanisms (Chaps. 7, 8).

The control information essential for cell functions may be transferred in two distinct ways – through (a) covalent interactions (e.g., via forming equilibrium structures such as phosphorylated proteins, RNAs, etc.), and (b) noncovalent interactions (e.g., via forming dissipative structures such as transient protein complexes, cytosolic ion gradients, etc.). The former may act as *internal* constraints that transmit control information through time, i.e., from one moment to the next during the lifetime of a cell or from one cell generation to the next, while the latter may act as external constraints on molecular machines (e.g., membrane potentials, cytosolic ion gradients, ATP levels in the cytosol, etc.) transmitting control information through space, e.g., between the nucleus and the cytosol and between the cytosol and the extracellular space (Ji 1988). Through these two mechanisms, the cell can control its molecular processes or events in space and time.

More specifically, the cell must control its molecular processes (conformational motions, ATP hydrolysis, etc.) that can occur on the subpicosecond (10^{-12} s) timescales and the subnanometer (10^{-9} m) length scales (referred to as the *microscopic* level) to drive the molecular processes that occur on the time- and length scales of 10^{-3} s and 10^{-5} m, respectively (referred to as the *mesoscopic* level). Thus, the micro- and mesoscopic levels are separated from each other by approximately 10 orders of magnitude in time- and mass scales. The coupling of events separated by such divergent temporal and spatial scales will be referred to as the *micro-meso coupling*. The fundamental significance of the *micro-meso coupling* in biology stems from the fact that the free energy needed to drive all living processes inside the cell ultimately derives from exergonic chemical reactions (e.g., oxidation of glucose, ATP hydrolysis) that occur only on the microscopic level. The question as to how the cell accomplishes the micro-meso couplings across the spatiotemporal gaps separated by 10 orders of magnitude is one of the most challenging problems facing the contemporary biology, since they underlie most of the unsolved problems in molecular and cell biology, including the mechanisms of force generation in muscle, gene expression in the nucleus, and morphogenesis of living tissues (Sect. 15.1).

The possible mechanism employed by the cell to accomplish the *micro-meso coupling* on the mass scale is suggested by the molecular tactics used by the cell to control its DNA molecule during cell division, namely, the *chunk-and-control* mechanism discussed in Sect. 2.4.2. The figure showing the chunking of DNA double helix into a chromosome is reproduced from Fig. 2.9 on the right-hand side of Fig. 11.14. As indicated in Sect. 2.4.2, the chunking operation not only compactifies DNA by a factor of 10^9 but also *constrains* the motional degrees of freedom of DNA components such as atoms, nucleotides, and DNA segments. In addition, since the frequency of an oscillator is inversely proportional to the square root of its reduced mass (http://en.wikipedia.org/wiki/Molecular_vibration), it is here postulated that the chunking operation (which increases mass) applied to biopolymers can lead to *slowing down* of their global oscillatory motions. For convenience, this idea will be referred to as the principle of "slowing down oscillation by increasing mass" (SDOBIM). Hence, the principle of SDOBIM defined here can provide the theoretical foundation for the principle of "chunk-and-control" (C&C) (Sect. 2.4.2), and these two principles may be viewed as the two sides of the same coin. As will be discussed in Sect. 11.3, the principle of SDOBIM appears to apply to single-molecule enzyme mechanics, i.e., to the low-frequency oscillations of cholesterol oxidase measured with a single-molecule manipulation technique (Lu et al. 1998).

It is interesting to note that the chunk-and-control operation depicted in Fig. 11.14 involves a single molecule consisting of a set of monomeric units *covalently* linked in a linear chain. It is postulated here that the chunk-and-control operation can be applied to a set of molecules that are not covalently linked to one another but nevertheless interact through *noncovalent bonds* such as electrostatic bond, hydrophobic bond, and the van der Waals force. Thus it may be necessary to distinguish between two kinds of chunk-and-control operations: (a) the *covalent-chunk-and-control* (CC&C) operation and (b) the *non-covalent-chunk-and-control* (NC&C) operation. The DNA packaging shown in Fig. 11.14 embodies both CC&C (e.g., covalent linking of thousands of structural genes and regulatory regions into a chromatin) and NC&C (e.g., beads-on-a-string to form a chromatin).

11.3 Single-Molecule Enzymology

Single-molecule measurements can provide mechanistic information on enzymic catalysis that is not available through conventional ensemble-averaged enzyme kinetic studies (Xie 2001; Ishii and Yanagida 2007; Deniz et al. 2008). A good example of single-molecule enzyme experiments is provided by Lu et al. (1998) who monitored the fluorescence emission from a single molecule of the flavin adenine dinucleotide (FAD) bound to the active site of cholesterol oxidase (COx) as it went through catalytic cycles (Figs. 11.15, 11.17). The principle underlying the phenomenon of fluorescence emission is explained in Fig. 11.16.

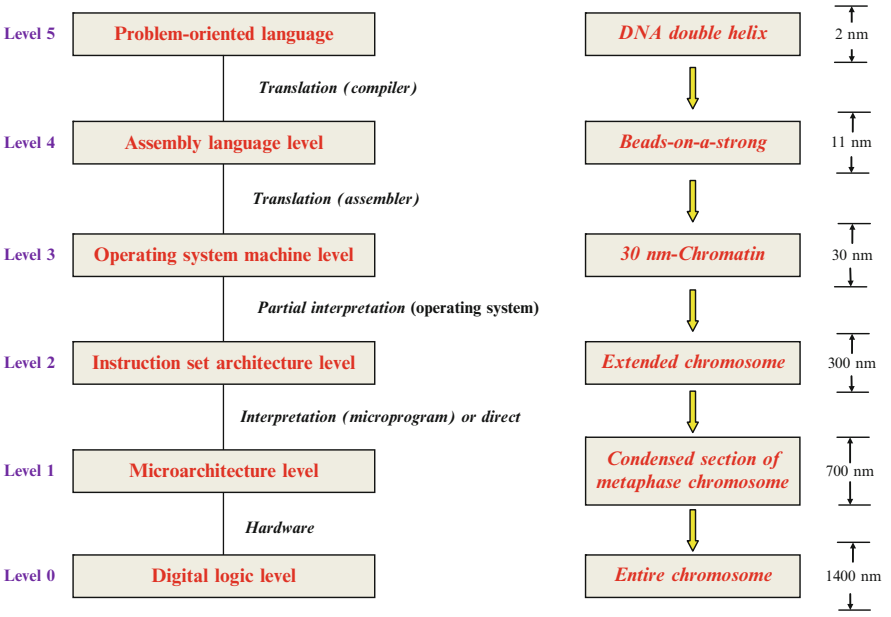


Fig. 11.14 (*Left*) The multilevel structural organization of the computer (Adapted from Tanenbaum 2003). Digital logic => Microarchitecture => Instruction set architecture => Operating system machine => Assembly language => Problem-oriented language. (*Right*) Multistep chunking of DNA into chromosome. DNA double helix => beads-on-a-string form of chromatin => 30-nm chromatin fiber => 300-nm section chromosome => 700-nm section chromosome => 1,400-nm section metaphase chromosome

Many mathematical models have been proposed to account for the results of the single-molecule enzymological measurements on COx (Lu et al. 1998; Kurzynski 2006; Qian and Xie 2006; Prakash and Marcus 2007), but they are all rooted in the conventional Michaelis-Menten mechanisms as applied to individual molecules of enzyme and couched in the language of physics- and chemistry-based continuous mathematical functions such as Eq. 11.25 shown in Sect. 11.3.3. The main objective of this section is to propose an alternative theory of single-molecule enzymology that is based on the concept of *conformons* (i.e., the packets of conformational *energy* and genetic *information* stored at sequence-specific sites within biopolymers that are postulated to drive all goal-directed molecular processes in the cell including catalysis; see Chap. 8). The concept of conformons was originally developed to account for the energy-coupled processes of the living cell such as oxidative phosphorylation, muscle contraction, and active transport (Green and Ji 1972a, b; Ji 1974b, 1985a, b, 2000) but has also been found to provide a reasonable explanation for the mechanistic isomorphism (or similarity) between *blackbody radiation* and *enzymic catalysis* (see Fig. 11.24) as found in Ji (2008b). If the conformon-based explanation for single-molecule enzymological data on COx proves to be correct, it would be possible to conclude that the single-molecule enzymological

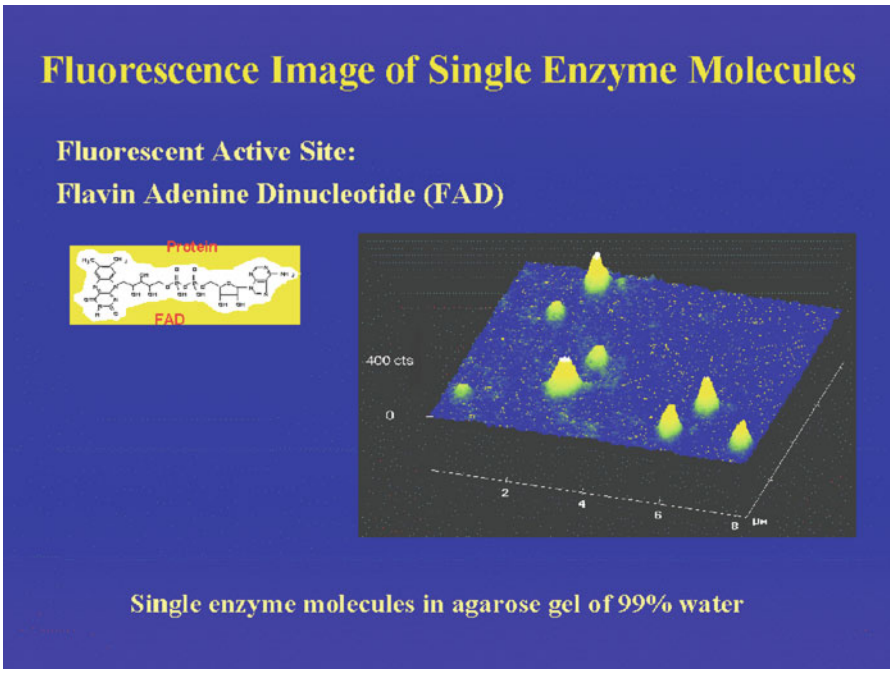
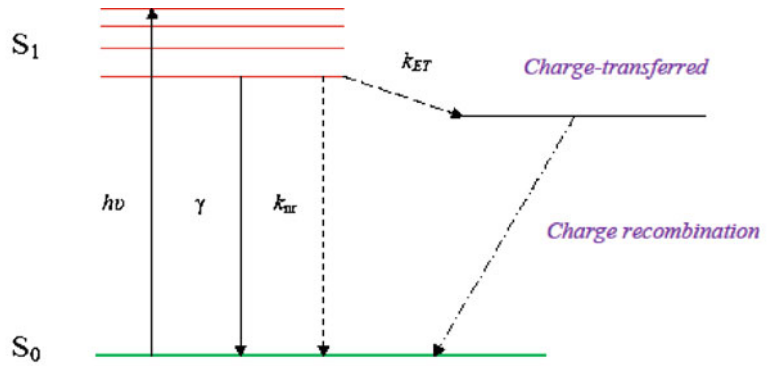


Fig. 11.15 The fluorescence image of single molecules of cholesterol oxidase (COx) immobilized in agarose gel. When FAD is illuminated at 442 nm (see the *upward arrow* in Fig. 11.16), the prosthetic group absorbs this photon and emits fluorescence at 530 nm (see the *solid downward arrow* in Fig. 11.16). This image was taken in 4 min with an inverted fluorescence microscope by raster-scanning the sample with a focused laser beam of 500 nW at the excitation wavelength. Each individual fluorescent spot indicates the presence of a single molecule of COx. The intensity variations are due to different longitudinal positions of COx molecules in the gel (Reproduced from http://www.nigms.nih.gov/News/Reports/single_molecules.htm)

data on COx cannot be fully explained unless the traditional physics- and chemistry-based mathematical formalisms are supplemented with the evolution-derived *catalytic information* of the kind discussed by the Ranganathan group (Poole and Ranganathan 2006; Socolich et al. 2005; Süel et al. 2003) and others. One possible way by which "catalytic information" may affect the probability of the occurrence of a given waiting time (or a rate constant) was constructed using as a metaphor the potential energy transfer between the *garage door* and its *spring* as explained in Table 11.12 in Sect. 11.3.3.

11.3.1 Waiting Time Distribution of Cholesterol Oxidase

Cholesterol oxidase is a 53,000 Da protein with 504 amino acid residues that catalyzes the oxidation of cholesterol by oxygen to form cholesterone. The active



S_0 = ground electronic state of the fluorophore, i.e., the dye molecule emitting fluorescence
S_1 = first electronic excited state of the fluorophore with 4 vibrational states
γ = the wavelength of fluorescence emitted
k_{nr} = rate of non-radiative (nr) decay
k_{ET} = rate of electron transfer from the fluorophore to a to a nearby electron acceptor.
$h\nu$ = the absorption of the excitation light which promotes the electron from the ground electronic energy level, S_0, to the first excited electronic energy level, S_1, with excess vibrational energies, which are quickly lost to fall to the ground vibrational level, from which fluorescence emission

Fig. 11.16 The diagram showing the submolecular processes underlying photon-molecule interactions

site of the enzyme contains one FAD molecule, which is fluorescent in its oxidized form. FAD is reduced by a cholesterol molecule to $FADH_2$, which is then oxidized back to FAD by molecular oxygen (Fig. 11.17). The fluorescence of FAD is turned on and off (giving rise to the so-called blinking phenomenon) as the redox state of the FAD undergoes transitions between the oxidized and reduced states, each on-off cycle corresponding to one turnover of the enzyme.

The electronically excited state S_1 in Fig. 11.16 can decay back to the ground state S_0 through three different mechanisms:

1. The radiative pathway (i.e., the γ step)
2. The nonradiative (nr) pathway (i.e., the k_{nr} step)
3. The electron transfer (ET) pathway (i.e., the k_{ET} step)

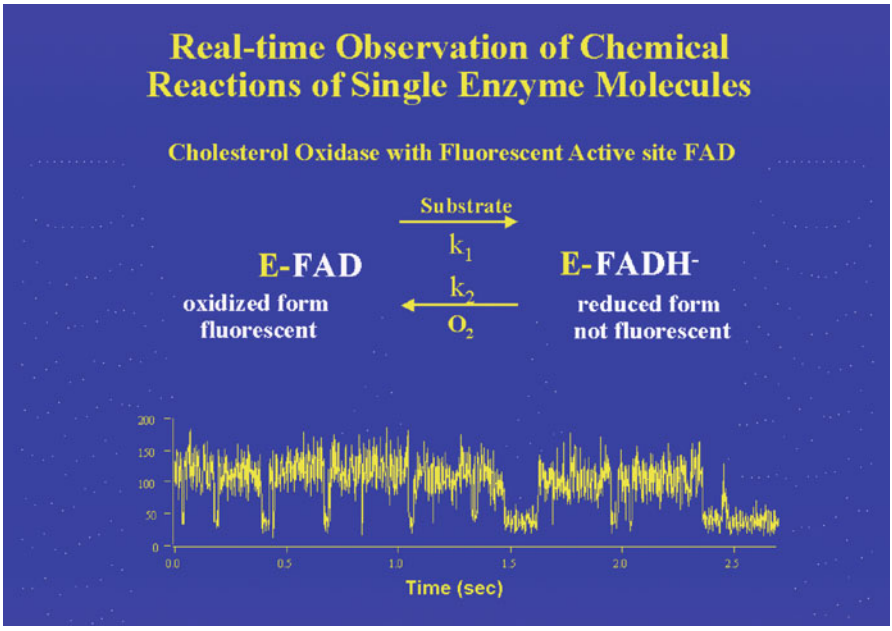


Fig. 11.17 The measurement of the turnover of a cholesterol oxidase (COx) molecule in the presence of cholesterol (0.20 mM) and oxygen (0.25 mM). The prosthetic group, FAD, is fluorescent when in its oxidized state with an average relative intensity of about 130 units (which is referred to as the "on" state) and nonfluorescent when in its reduced state with an average intensity of about 40 units (which is referred to as the "off" state) (Reproduced from http://www.nigms.nih.gov/News/Reports/single_molecules.htm)

The relative importance of these pathways during a given cycle of photon absorption and emission probably depends on the conformation of FAD molecule which is in turn most likely affected by the local conformational structure of the FAD binding pocket of COx. It is for this reason that, as the conformation of COx fluctuates, the fluorescence efficiency, defined as $k_\gamma/(k_\gamma + k_{nr} + k_{ET})$, also fluctuates, thus accounting for the fluorescence fluctuations observed during the on- or off-times as shown in Fig. 11.17, or between A and B and between C and D in Fig. 11.20.

Lu et al. (1998) used a single-molecule manipulation technique (Xie 2001; Ishii and Yanagida 2007) to measure the cycling of cholesterol oxidase between its oxidized ("on") and reduced ("off") states (Fig. 11.17). Two of the most significant findings Lu et al. (1998) made are (a) that the enzyme molecule spends variable times in on- or off-states and (b) that the on- and off-times are not distributed randomly (or normally) but have a long tail. As can be seen on the lower left corner of Fig. 11.18, the on-times varied from about 70 ms (milliseconds) to 1,300 ms with most probable on-times lying between 100 and 200 ms. Estimating from the area under the "on-time" distribution histogram in Fig. 11.18, it may be concluded that

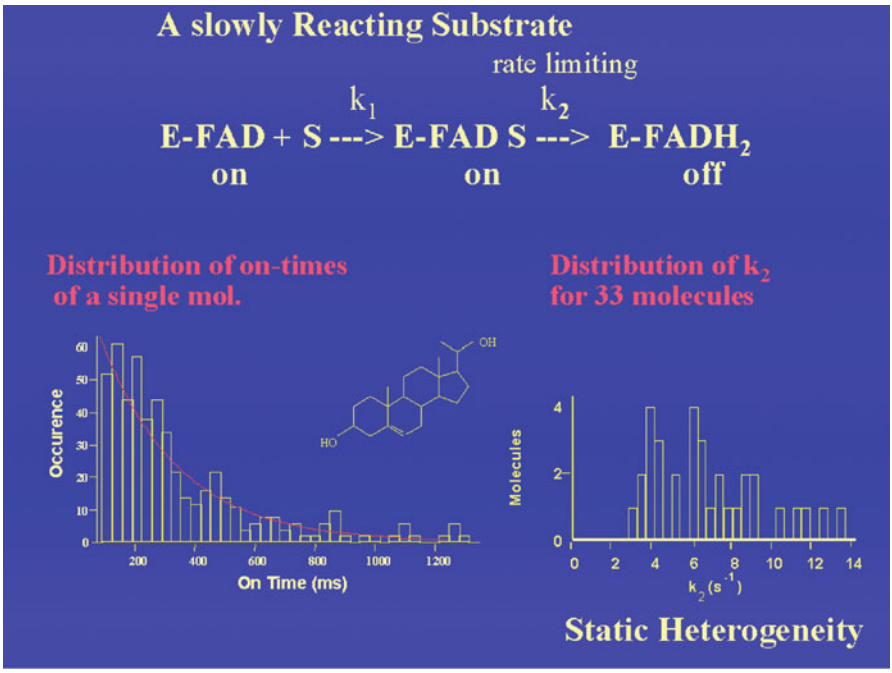


Fig. 11.18 The distribution of the kinetic parameters (on-times) measured from COx molecules. (*Left*) The distribution of the on-times (i.e., the duration of the times that the FAD molecule remains in its oxidized state and fluorescent). This phenomenon of varying on-times is known as "dynamic disorder" or "dynamic heterogeneity" (Zwanzig 1990). Cholesterol concentration = 0.2 mM. (*Right*) The distribution of the rate constants, k_2, for the reduction of FAD to $FADH_2$ measured from 33 COx molecules. At any given time, the k_2 values measured from different molecules of COx vary by a factor of up to 5 (Reproduced from http://www.nigms.nih.gov/News/Reports/single_molecules.htm)

the probability of the oxidized form of COx to be reduced by cholesterol (present at 0.2 mM) within less than 500 ms is about 80%. Since the off-times in Fig. 11.17 appear shorter than the on-times on average, the probability of reduced COx being reoxidized by molecular oxygen is faster than the rate of the reduction of COx by cholesterol. Also, since the concentration of cholesterol and oxygen molecules were approximately the same (0.20 vs. 0.25 mM), it may be concluded that the rate constant for the reduction of COx, i.e., k_1 in Fig. 11.17, is smaller than the rate constant, k_2, for the reoxidation of COx.

Lu et al. (1998) proposed a simple mechanism, Scheme (11.16), to account for the distribution of the off-times shown in Fig. 11.18:

$$\underset{a}{\text{E-FAD}} + \text{S} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \underset{b}{\text{E-FAD} \bullet \text{S}} \xrightarrow{k_2} \underset{c}{\text{E-FADH}_2} + \text{P} \tag{11.16}$$

where k's are rate constants. A similar mechanism, Scheme (11.17), was proposed for the on-time distribution.

$$\underset{d}{\text{E-FADH}_2} + \text{O}_2 \rightleftarrows \underset{e}{\text{E-FAD} \bullet \text{O}_2} \rightarrow \underset{f}{\text{E-FAD}} + \text{H}_2\text{O}_2 \tag{11.17}$$

Schemes (11.16) and (11.17) do not obey the *Principle of Microscopic Reversibility* (PMR) (discussed in Sect. 3.3). In other words, the transitions from *b* to *c* or from *e* to *f* are not symmetric as demanded by PMR. An alternative mechanism that obeys PMR is proposed in Fig. 11.19. Since this mechanism is based on the generalized Franck-Condon principle discussed in Sect. 2.2.3, we may refer to the mechanisms proposed in Fig. 11.19 as the *Franck-Condon mechanism* of the action of cholesterol oxidase. The Franck-Condon mechanism entails expanding the number of the states involved in one cycle of the enzymatic turnover from the original 6 to a total of 16, as explained in the legend to Fig. 11.19a, b. The unique features of the Franck-Condon mechanism are:

1. The enzyme can exist in two states – the ground state (see *a, h, i* and *p*) and the thermally *activated/excited* state (see *b, g, j* and *o*).
2. The enzyme binds its substrate or product only when in a thermally activated/excited state (see *c, f, k,* and *n*). *It should be noted that "energized state" is synonymous with "thermally activated/excited state stabilized by ligand binding." Without such ligand-induced stabilization, thermally activated/excited states are thought to relax rapidly back to ground states.*
3. In the Franck-Condon state, the distinction between *substrate* and *product* disappears due to the highly unusual microenvironment of the enzyme active site prevailing in this state. This is tantamount to asserting that $d = e$, and $l = m$ at the Franck-Condon state, within the Heisenberg uncertainty principle (Reynolds and Lumry 1966; Ji 1974a).

The Franck-Condon mechanism proposed in Fig. 11.19 provides a novel set of explanations for the single-molecule fluorescence data reported by Lu et al. (1998). Their data are reproduced in Fig. 11.20 with a particular attention given to the variations in the amplitudes of the fluorescence fluctuations recorded in Fig. 1b of their paper. The features of their data that are of special interest are summarized in Table 11.6 along with the corresponding explanations offered by the Franck-Condon mechanism shown in Fig. 11.19.

One of the most significant outcomes of analyzing the single-molecule fluorescence data measured by Lu et al. (1998) as shown in Table 11.6 is that there may exist two high-energy states – one relatively long-lasting after stabilization by ligand binding (denoted by the superscript *) and capable of performing external work and the other short-lived (denoted by the superscript ‡). Thus we will refer to the former as the *energized state* distinct from the Franck-Condon state. Since energized states are characterized by their *stored energy* in the sense of McClare (1971, 1974) (see also Sect. 2.1.4), the enzymes in energized states are capable of

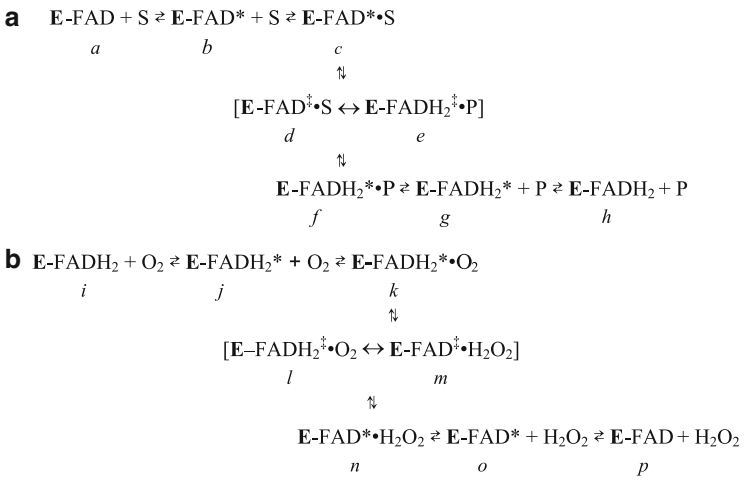


Fig. 11.19 A complete mechanism of the cholesterol redox cycle based on the *generalized Franck-Condon principle* (Sect. 2.2.3) and *the principle of microscopic reversibility* (Sect. 3.3). The *hyphen* and the *dot* symbolize noncovalent binding interactions. The superscript * indicates thermally activated/excited (*b, g, j,* and *o*) or energized (*c, f, k,* and *n*) states of the enzyme-ligand complex and the superscript ‡ denotes the transition state (also called the Franck-Condon state; Reynolds and Lumry 1966; Ji 1979). (**a**) **The reduction of the cholesterol oxidase molecule by cholesterol (S).** *a* = the oxidized enzyme before binding cholesterol; *b* = the oxidized enzyme in a *thermally activated/excited state* (or a *thermally activated conformer*; see Sect. 11.3.2); *c* = the energized state of the oxidized enzyme that has bound substrate S (thermal energy is thought to be transduced to mechanical energy upon binding S; see Sect. 11.3.2); *d* = the Franck-Condon state of the oxidized enzyme-substrate complex accessible from *c* through another round of thermal fluctuations; *e* = the Franck-Condon state of the reduced enzyme-product complex thermally accessible from *f*; *f* = the energized state of the reduced enzyme binding product P; *g* = the reduced enzyme in a thermally activated state; *h* = after the product dissociates from reduced COx. (**b**) **The oxidation of reduced cholesterol oxidase by molecular oxygen.** *i* = the COx in the reduced state before binding oxygen; *j* = the reduced COx in a thermally activated state; *k* = the energized COx in its reduced state binding oxygen; *l* = the Franck-Condon state of the reduced COx-oxygen complex; *m* = the Franck-Condon state of the oxidized COx-hydrogen peroxide complex; *n* = the energized COx in its oxidized state binding hydrogen peroxide; *o* = the oxidized COx in a thermally activated state; *p* = after hydrogen peroxide dissociates from the ground-state COx molecule

performing work either on their environment (as in ion pumping or movement of actin filament) or internally (e.g., modulation of the activation free energy barrier for catalysis, or k_2, in Scheme 11.17), the latter contributing to the dynamic disorder of the COx enzymic activity.

In constructing Table 11.7, the following assumptions have been made:

1. Both FAD and $FADH_2$ fluoresce, although the former is much more fluorescent than the latter probably by a factor of about 100.
2. Both FAD and $FADH_2$ molecules undergo many cycles of the photon absorption-radiative decay process during the lifetime of any of the four conformational states (or conformers) of COx labeled A, B, C, and D (see Fig. 11.20).

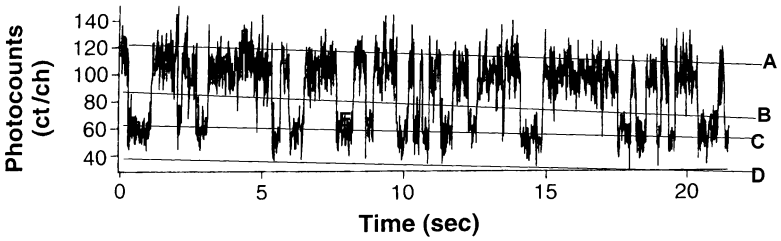


Fig. 11.20 The fluctuations of the fluorescence intensity of the cholesterol oxidase (COx) molecule in its oxidized and reduced states adopted from Fig. 1b in Lu et al. (1998). It is postulated here that the oxidized form of COx fluctuates between the fluorescence levels A and B, while the reduced form of the enzyme fluctuates between the florescence levels C and D. The four *lines* decline with time more or less in parallel, most likely due to light-induced structural damages suffered by the protein surrounding the fluorophore, which in turn alter the energy levels of the fluorophore, affecting the fluorescence efficiencies of both the oxidized and reduced form of COx. The average fluorescence intensity of the oxidized form of COx decreases from about 120 ct/ch (counts per channel) to 90 ct/ch in 22 s, while that of the reduced form decreases from about 60 to 40 ct/ch (in the same time period) which is near the background fluorescence. In addition to the fluorescence fluctuations occurring between A and B and between C and D, there are other fluorescence intensity transitions taking place from B to C and from C to B. See the region indicated by E (which is partially hidden) between B and C at around 8 s. Since there are four fluorescence states of COx, there are $4 \times 4 = 16$ possible transitions as shown in Table 11.7. Due to the diagonal symmetry of the table, there are only six distinct transitions whose probabilities of occurrences need not be equal. The relative transition probabilities predicted on the basis of the generalized Franck-Condon principle (Sect. 2.2.3) are given in the table

Table 11.6 The generalized Franck-Condon-principle-based explanations for the single-molecule measurements of the cholesterol oxidase enzymic activities reported by Lu et al. (1998). A, B, C, and D refer to the fluorescence levels defined in Fig. 11.20

Single-molecule measurements on cholesterol oxidase (see Figs. 11.17 and 11.18)	Explanations offered by the Franck-Condon-principle-based mechanism of action of COx (see Fig. 11.19)
1. There are two kinds of fluorescence fluctuations – one averaging around 100 ct/ch and the other around 50 ct/ch	The fluorescence intensity of both oxidized and reduced forms of COx fluctuates. The time-averaged fluorescence intensity of the oxidized form of the COx molecule is approximately 100 ct/ch, while the corresponding value for the reduced form is around 50 ct/ch
2. Each form of the COx molecule fluctuates between a high and low fluorescence level, giving rise to four fluorescence levels denoted as A, B, C, and D in Fig. 11.20	The molecular species responsible for the various fluorescence levels are postulated to be as follows: A = **E**-FAD B = **E**-FAD* and **E**-FAD*•S C = **E**-$FADH_2$*•P and **E**-$FADH_2$* D = **E**-$FADH_2$
3. There are two kinds of fluorescence transitions – (1) from level B to level C, and (2) from level C to level B, both taking place almost instantaneously compared to the on- and off-times (Fig. 11.17)	(1) The transition from B to C passes through and is rate-determined by the free energy level of the Franck-Condon state [**E** -$FAD^{\ddagger}$•S ↔ **E**-$FADH_2^{\ddagger}$•P] (2) The transition from C to B passes through and is rate-determined by the free energy level of the Franck-Condon state [**E**–$FADH_2^{\ddagger}$•O_2 ↔ **E**-$FAD^{\ddagger}$•H_2O_2]

Table 11.7 The relative transition probabilities among the four conformational states of COx-FAD complex predicted on the basis of the generalized Franck-Condon principle (Sect. 2.2.3). For the definition of the conformational states, A–D, see Figs. 11.17 and 11.20. The Arabic numerals in the table refer to the relative probabilities for the X to Y transition, where X and Y represent rows and columns, respectively, and the relative probabilities are in the order of $1 > 2 > 3 > 4 > 5$

	A	B	C	D
A	1	2	4	5
B	2	1	3	4
C	4	3	1	2
D	5	4	2	1

3. The generalized Franck-Condon principle is postulated to apply to the COx-FAD and COx-$FADH_2$ complexes in the sense that the probability of the fluorescence transitions of these complexes are inversely proportional to the Euclidean distances between the corresponding conformational states of COx indicated in Fig. 11.20.

Although the current state of development of single-molecule mechanics may not allow measurements to be made of these six transitions predicted in Table 11.7, it may be possible to detect them in the future when the single-molecule mechanics techniques improve.

11.3.2 Molecules, Conformers, and Conformons

In order to rigorously analyze single-molecule enzymological data such as shown in Figs. 11.18 and 11.24, it may be necessary to utilize some of the concepts, theories, and principles that have been developed in molecular enzymology and biology by various investigators since the mid-twentieth century, including Widom (1965), Volkenstein (1972, 1986), Green and Ji (1972a, b), Ji (1974a, b, 1990, 2000), Lumry (1974, 2009), Lumry and Gregory (1986), Lumry and Biltonen (1969), Northrup and Hynes (1980), Anderson (1983, 1987), Frauenfelder (1987), Frauenfelder et al. (2001), Welch and Kell (1986), Benham (1992, 1996a, b), Kurzynski (1993, 1997, 2006), and Eisenmesser et al. (2002).

In physical organic chemistry, the terms *configuration* and *conformation* are carefully differentiated (see Fig. 3.5) (Sect. 3.2) unlike in physics and molecular biology where they are often used interchangeably (Ji 1997a, see Table 4). Strictly speaking, not distinguishing *configurations* and *conformations* in chemistry is equivalent to conflating *electrons* and *protons* in physics, since *configurations* involve the movement of *electrons* while *conformations* entail *proton* displacement in molecules secondary to breaking and making H-bonds, the study of which being referred to

as the "electronic-conformational interactions" (Volkenstein 1986). A molecule has a unique *configuration* (as defined by the set of covalent bonds it possesses) and many *conformations* (as defined by the three-dimensional arrangements of the covalently bound atoms in a molecule which can be altered without breaking or forming any covalent bonds) (Sect. 3.2). Some of the simplest examples of molecules having different configurations are provided by isomers. Thus, cis- and trans-1-chloro-2-bromo-ehtylene have an identical set of atoms, i.e., C_2H_4BrCl, and yet have two different *configurations* (or arrangements of these atoms in the Euclidean space), namely, cis- and trans-isomers which cannot be interconverted without breaking at least one of the covalent bonds of C=C, C–H, C–Cl, and C–Br.

A given *configuration* of a set of atoms can assume numerous *conformations* (also called *conformers*). For example, 1-chloro-2-bromo-ethane, the product of reducing the C = C double bond in 1-chloro-2-bromo-ethylene, can exist in two conformations, one in which the chlorine and bromine atoms are located farthest apart and the other in which they are located nearest to each other, and these two conformations can be interconverted through rotations around covalent bonds. One crucial difference between *configurations* and *conformations* is that the activation free energy barrier separating one *configuration* from another is much higher (*50–100 kcal/mol*) than that separating one *conformation* from another within a given configuration (*1–3 kcal/mol*). Therefore, the thermal energy available under physiological conditions (about 0.6 kcal/mol) is usually not large enough to change configurations but sufficient to cause appreciable conformational changes, leading to the following general statement:

> *Configurations* of molecules are too robust to be altered but *conformations* of a molecule are labile enough to be interconverted, through thermal fluctuations. (11.18)

According to the cell language theory (Ji 1997a, 1999b), Statement 11.18 is ultimately responsible for the phenomenon known as "rule-governed creativity" (Sect. 6.1.4), namely, the ability of biopolymers (and cells) to generate indefinitely large numbers of states within the constraints of a finite number of components and syntactic rules. The relationship between *configurations* and *conformations* can be diagrammatically represented as shown in the first two columns in Fig. 11.21. Note that the energy separations between configurations are greater than those between conformations.

In discussing biochemistry, molecular biology and biophysics, it is useful to differentiate between *small molecules* (which may be referred to as "micromolecules" (e.g., glucose, ATP, NADH, and FAD) and *large molecules* (*or macromolecules*) including proteins, RNA, and DNA. What characterizes *biological macromolecules vis-à-vis micromolecules* of organic chemistry is that biological macromlecules (or *biomacromolecules*) are the product of a long biological evolution and hence carry biological (also called evolutionary or genetic) information, whereas micromolecules do not. In other words, biomacromolecules possess an extra degree of freedom not available to micromolecules, i.e., the dimension of *genetic* or *evolutionary information* (Fig. 4.2). This extra degree of

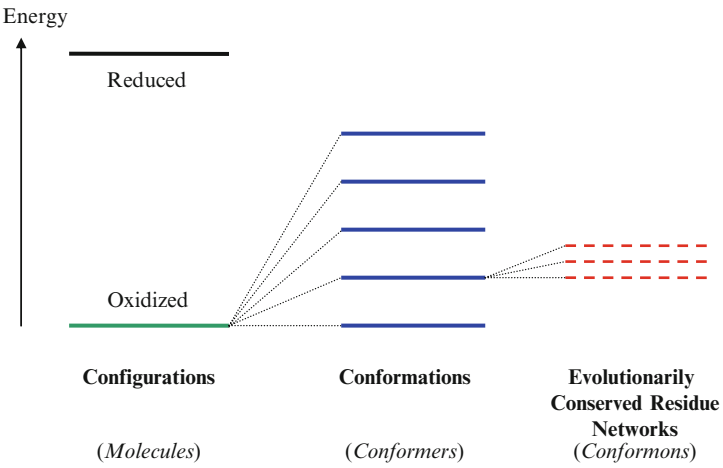


Fig. 11.21 A diagrammatic representation of the relations among *configurations* and *conformations* on the one hand and between *conformers* and *conformons* on the other. Multiple conformational energy levels are also available for the reduced configuration of a molecule but are not shown for brevity. The energy scale is approximate

freedom (or the *bioinformation or bioinformatic* dimension) is thought to be manifested in the form of evolutionarily conserved amino acid residues as indicated on the right-hand side of Fig. 11.21. That there indeed exists such an internal degree of freedom in enzymes was strongly suggested by the finding that many families of enzymes, receptors, and DNA-binding proteins are characterized by unique networks of a small number (10–25% of the total) of amino acid residues that are evolutionarily conserved and coevolved (Lockless and Ranganathan 1999; Süel et al. 2003; Poole and Ranganathan 2006).

Another distinguishing feature between these two classes of molecules is that micromolecules are too small to harbor any long-lived internal conformational strains or kinks, whereas biomacromolecules are large and complex enough to retain relatively stable internal *conformational strains* produced either during their syntheses on the ribosomes (Klonowski and Klonowska 1982) or during their catalytic cycles. Such conformational strains have been variously referred to as *conformons* (Green and Ji 1972a, b; Ji 1974b, 2000, 2004a), *frustrations* (Anderson 1983, 1987), *mobile defects* (Lumry 1974; Lumry and Gregory 1986), or *SIDDs* (Stress-Induced Duplex Destabilizations; Benham 1992, 1996a, b). It is here suggested that the concept of the conformational *gates* that are postulated to control the rates of enzymatic reactions in the *stochastic model of enzymic catalysis* proposed by Kurzynski (1997, 2006) can also be viewed as equivalent to *conformons*, since no *gate* can be opened or closed at right times for right durations without utilizing *mechanical energy* and *control information* both stored in local conformational strains in proteins. There are other interesting commonalities between the *conformon theory of molecular machines* (Ji 1974a, b, 2000, 2004a) and *the stochastic model of protein machines* proposed by Kurzynski (1993, 1997, 2006).

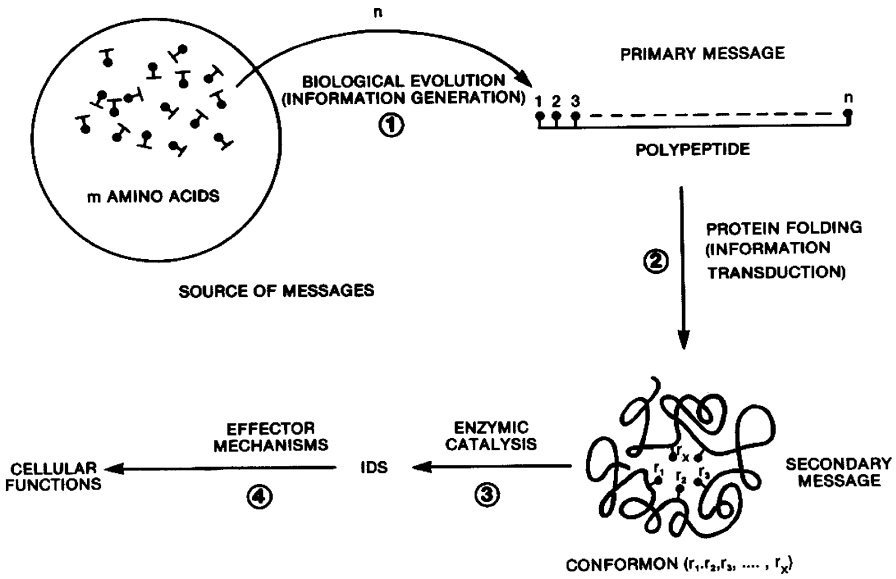


Fig. 11.22 A theoretical model of the communication system responsible for generating biological information (Ji 1990, 2000). The message source = biological evolution; the channel = conformons and IDSs (intracellular dissipative structures such as ion gradients); the receiver = the cell. Conformons are the packets of mechanical energy or conformational strains entrapped in local sites within biopolymers, thus carrying free energy and genetic information that are both necessary and sufficient to drive all goal-directed molecular processes inside the cell (Ji 1974a, b, 2000) (Sect. 8.1). The key postulate underlying the model is that the information generated in Step 1 is utilized to produce *n* conformons in an enzyme, each of which being capable of driving one elementary step in enzymic catalysis, Step 3, leading to the production of IDSs, Intracellular Dissipative Structures (Chap. 9), which are postulated to be the immediate causes (or driving forces) for all cell functions

It was postulated in Ji (1985a, b) that the genetic information of an enzyme is not *homogeneously* distributed within the enzyme but *heterogeneously* distributed as a part of *conformons*, defined as the discrete units of *mechanical energy* and *genetic information* (i.e., the information that is transmitted from one cell generation to the next). Conformons are postulated to be responsible for all goal-directed molecular motions (e.g., chromosome remodeling) and processes (e.g., enzymic catalysis, active transport, muscle contraction) in the cell (Chap. 8). Based on a protein-centered model of biological evolution schematized in Fig. 11.22, it was possible to derive Eq. 11.19 that can be used to estimate the number of amino acid residues constituting one conformon within an enzyme (Ji 1990, pp. 198–200) (see Table 11.8 for more details):

$$n!/(n-x)! = 20^{n/p} \tag{11.19}$$

Table 11.8 The components of the communication system modeling biological evolution

System components	Identification (see Fig. 11.22)
Source of messages	Biological evolution regulating the linear sequence of amino acids of enzymes through DNA and RNA (see Step 1)
Channel	Polypeptides, conformons, IDS (see Steps 2 and 3)
Receiver	The living cell (see Step 4)

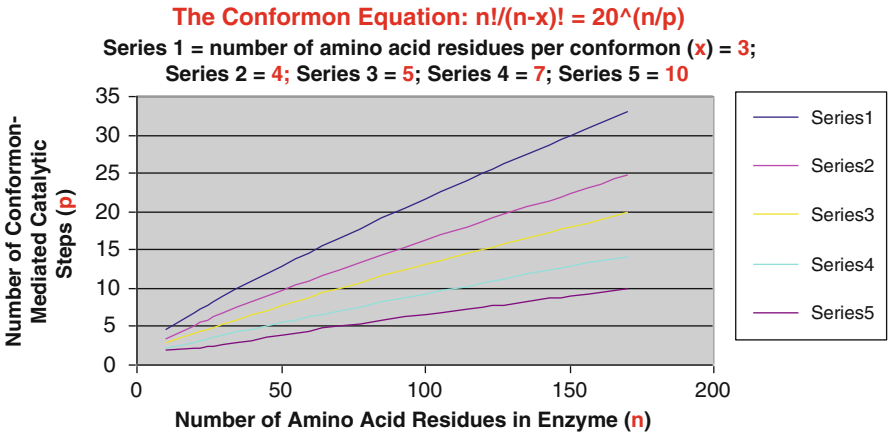


Fig. 11.23 A numerical simulation of the *conformon equation* derived from the model of biological evolution shown in Fig. 11.22. See text for more details

where n = the number of the amino acid residues of an enzyme; x = the number of the amino acid residues constituting a conformon that participates in (or are essential for) a catalytic act such as binding, de-binding, covalent rearrangement, free energy storage, and free energy transfer; and p = the maximum number of the conformon-mediated catalytic steps within an enzyme molecule. For convenience, Eq. 11.19 will be referred to as the "conformon equation." Inserting into the *conformon equation* a set of reasonable numerical values for a typical enzyme, i.e., n = 150 and p = 10, it is found that the average number of the amino acid residues constituting one conformon is approximately nine (see Table 2 in Ji 2000). This appears to agree with the number of the evolutionarily conserved residues involved in numerous enzymic functions described in Lockless and Ranganathan (1999), Süel et al. (2003), Poole and Ranganathan (2006). The results of a more systematic calculation based on Eq. 11.19 are depicted in Fig. 11.23, from which it is clear that the following generalization can be made:

> The number of conformon-mediated catalytic steps are directly proportional to the number of the amino acid residues constituting an enzyme and inversely proportional to the number of the amino acid residues constituting a conformon. (11.20)

According to Fig. 11.23, a polypeptide with 40 amino acid residues catalyzing three catalytic processes will have three *conformons*, each comprising 10 amino acid residues (see the lower left region of Series 5 in Fig. 11.23). If there is no overlap among the amino acid residues that constitute two different conformons, the maximum number of amino acid residues required to generate three conformons would be 30. If an overlap is allowed, the maximum number would be less than this number, leading to the prediction that at most 75% of the amino acid residue of the 40-mer polypeptide should be evolutionarily conserved. This prediction seems to be in agreement with the observation reported by Socolich et al. (2005): Multiple sequence alignments of the 120 members of the WW domain family revealed that about 50% (i.e., 17 out of 36) of the amino acid residues of WW domains is evolutionarily conserved. These conserved residues exhibited the SCA (statistical coupling analysis) conservation scores greater than 0.5, the background level (see Fig. 1b in Socolich et al. 2005). The discrepancy between the predicted value of 75% and the observed one of 50% may simply indicate that

> An amino acid residue at a given locus on a polypeptide chain can participate in producing more than one conformons. (11.21)

Statement 11.21 is reminiscent of the piano keys (*amino acid residues*) which can be struck (at different times) to produce more than one musical sounds or melodies (analogous to *conformons*). We may refer to Statement 11.21 as the *conformon composition rule*, which can be more generally stated as follows:

> Conformons are generated in a biopolymer from a set of evolutionarily conserved elements (amino acid residues or nucleotides) that are combinatorially arranged in space and time, just as musical melodies are generated from the combinatorial arrangements of musical notes in time. (11.22)

Just as producing melodies requires a pianist's expending energy by striking a select set of right keys in a right temporal order, it is clear that generating a conformon in a conformer of an enzyme by selecting a set of right amino acid residues arranged in space and time must be paid for by some exergonic processes such as ligand binding/de-binding and electronic rearrangements (known as chemical reactions). From the bioenergetics point of view, the production of a conformon requires coupling two partial processes – one endergonic (free energy-consuming) and the other exergonic (free energy supplying) (see Fig. 11.30 for more details).

The fundamental assumption made in deriving the *conformon equation* simulated in Fig. 11.23 can be stated as follows:

> The information (secondary message) required to order amino acid residues in space and time to produce conformons within an enzyme cannot be greater than the amount of the genetic information (primary message) encoded in the primary structure of the enzyme. (11.23)

Statement 11.23, if proven to be true, may be referred to as the *Principle of Information Conservation Principle* (PIC), in analogy to the *Principle* of *Energy Conservation* (PEC) in thermodynamics.

In view of the importance of the concept of the *conformon* in interpreting single-molecule data of Lu et al. (1998) to be presented in Sect. 11.3.3, the step-by-step derivation of the *conformon equation*, Eq. 11.19, is thought to be important enough to be reproduced below from Ji (1990, pp. 197–200) (replacing the original figure numbers with the corresponding ones from this chapter and adding a new table):

> To estimate the information content of conformons, it is necessary to postulate a communication system that links the source of message to the receiver through a channel (Pierce 1980). The communication system of our interest is schematically shown Fig. 11.22, and the various components of the systems are identified in Table 11.8.
>
> To calculate the Hartely information content of any message, it is necessary to know the number of all possible messages (W_0) and the number of messages actually selected (W). Then the average information content (I) of a message is given by
>
> $$I = \log_2(W_0/W) \tag{1}$$
>
> Equation (1) can be derived from Shannon's formula (Pierce 1980) by assuming that all messages have equal probability of selection. I is maximum when W = 1;
>
> $$I = \log_2 W_0 \tag{2}$$
>
> Applying Eq. (2) to Step 1 in Fig. 11.22, it is clear that the maximum information content of the primary message (I_p) is
>
> $$I_p = \log_2 m^n \tag{3}$$
>
> where m = the number of different amino acids and n = the number of amino acid residues constituting an enzyme. We assume that all of the information contained in the primary message, I_p, is transduced into the information content of conformons, each conformon consisting of an alignment of x amino acid residues (out of n) into a transient (i.e., kinetically labile and metastable) structure at the active site of an enzyme (see Step 2). The maximum information content of one conformon can then be estimated on the basis of two further assumptions: (1) all conformons are unique (i.e., no degeneracy) and (2) the maximum number of conformons consisting of x amino acid residues out of a polypeptide chain of n amino acid residues can be calculated as
>
> $$W_0 = n!/(n - x)! \tag{4}$$
>
> Inserting Eq. (4) into Eq. (2) leads to the maximum information content of one conformon:
>
> $$I_{conformon} = \log_2[n!/(n - x)!] \tag{5}$$
>
> If all the primary information is transduced into p conformons, we have
>
> $$\begin{aligned} I_p &= p \cdot I_{conformon} \\ &= p \cdot \log_2[n!/(n - x)!] \end{aligned} \tag{6}$$
>
> From Eqs. (3) and (6), one obtains
>
> $$[n!/(n - x)!]^p = m^n \tag{7}$$

If the value of the number of amino acids has been stable at 20 during most of the biological evolution, Eq. (7) can be rewritten as

$$[n!/(n-x)!]^p = 20^n \tag{8}$$

Using Eq. (8), the maximum number of conformons, p, that can be generated within a typical enzyme of 150 amino acid residues can be estimated, if the value of x is taken to be typically 5–10, in agreement with the recent findings of A. Fehrst (1985) that 12 amino acid residues participate in the active site of tyrosyl t-RNA synthase.

$$[150!/(150-10)!]^p = 20^{150} \tag{9}$$

Eq. (9) is satisfied when $p = 9.02$, or 9. Therefore, the maximum information content of one conformon, from Eq. (6), is

$$\begin{aligned} I_{conformon} &= I_p/p \\ &= \left(\log_2 20^{150}\right)/9 \\ &= 653/9 = 73 \text{ bits} \end{aligned} \tag{10}$$

If we assume that the value of x is 5 instead, the same calculation leads to

$$I_{conformon} = 36 \text{ bits} \tag{11}$$

Therefore, we conclude that the information content of one conformon may be in the range of 40–80 bits.

Equation (8) is the same as Eq. 11.19, the *conformon equation*, which has allowed us to estimate the maximum information content of a conformon to be 40–80 bits.

The maximum *free energy* content of a conformon can also be estimated. Since the synthesis of one molecule of ATP in mitochondria is known to require about 16 kcal/mol of free energy and since it takes at least one conformon to couple the redox reaction of the respiratory chain and the phosphorylation reaction of ADP catalyzed by ATP synthase, it would follow that the maximum free energy content of one conformon is 16 kcal/mol (Ji 2000). If the number of steps coupling respiration and phosphorylation is x, then the average free energy content of one conformon would be reduced to 16/x kcal/mol. With $x = 10$, the free energy content of one conformon would be 1.6 kcal/mol. Hence it may be reasonable to conclude that the free energy content of one conformon ranges between 2 and 16 kcal/mol (see Table 2 in Ji 2000).

11.3.3 Isomorphism Between Blackbody Radiation and Enzymic Catalysis

At a first glance, there seems to be little connection between *blackbody radiation* (i.e., the study of the relation between the intensity of the light emitted from an object as a function of wavelength and temperature) and *enzymic catalysis*. The distance between these two topics is so vast that it is not surprising that more than one

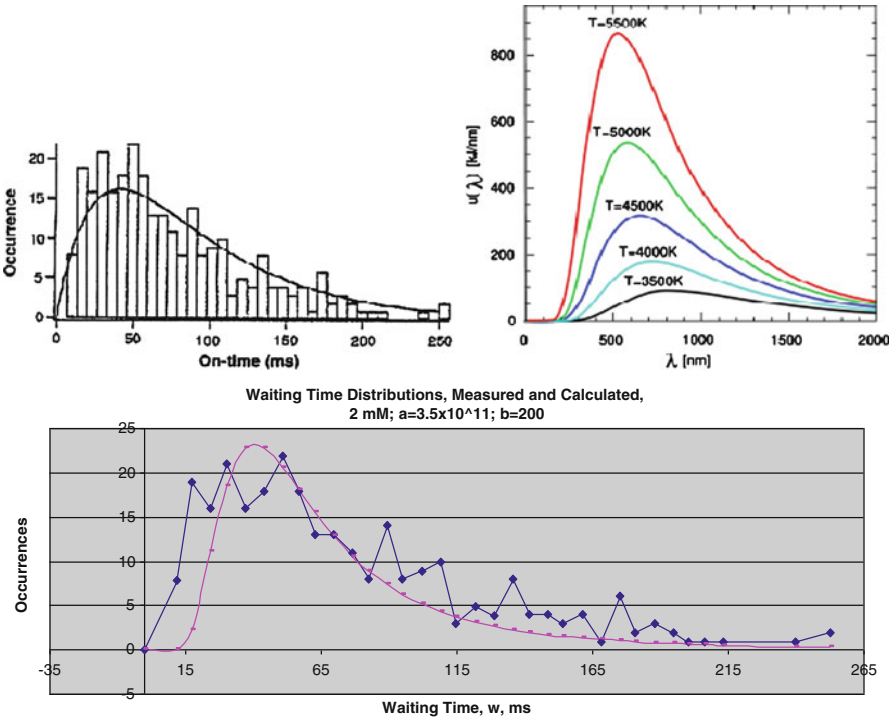


Fig. 11.24 The waiting time distribution (or histogram) of one molecule of cholesterol oxidase (COx) at the substrate concentration of 2 mM (Adapted from Fig. 1d in Lu et al. 1998). (*Upper*) The *solid line* (*left panel*) was derived from Eq. 11.25 with $k_1 = 33 \pm 6\ s^{-1}$, $k_2 = 17 \pm 2\ s^{-1}$, and $k_1 = 0$. The blackbody spectrum (*right panel*) was reproduced from *http://schools-wikipedia.org/images/705/70552.png.htm*. (*Lower*) The *smooth curve* derived from Eq. 11.27, with $a = 3.5 \times 10^5$, $b = 200$, and $X(w) = 0$. [Cholesterol] = 2.0 mM

statistical mechanicians with whom I had occasions to discuss this topic at the 100th Statistical Mechanics Conference held at Rutgers in 2008 (Ji 2008b) showed either no interest or even negative reactions. The purpose of this section is to present theoretical and experimental evidence that supports the following two conclusions:

1. Both blackbody radiation and enzymic catalysis can be viewed as resulting from *thermal excitation* of systems of *molecular oscillators*. Viewing an enzyme as a collection of oscillators is in agreement with the prediction made in Ji (1974a):

   Given all the vibrational frequencies of the individual bonds in an enzyme, as well as their three-dimensional arrangements, we can in principle deduce the thermodynamic and catalytic properties of the enzyme under any conditions. (11.24)

2. The waiting time distribution data of Lu et al. (1998), Figs. 11.18 and 11.24, can be accounted for in terms of a *quasi-deterministic* equation, Eq. 11.27, consisting of (a) a deterministic (or synchronic; see Table 4.2 for the definition of "synchronic" and "diachronic") term isomorphic with the Planck's radiation

formula, Eq. 11.26, and (b) a nondeterministic (or diachronic) term reflecting the genetic information embodied in *conformons* (discussed in Sect. 11.3.2).

The key observations reported by Lu et al. (1998) on cholesterol oxidase (COx) are as follows:

(A) The rate constant (or turnover time) of a COx molecule is not constant (as was generally believed prior to the Lu et al. 1998 experiments) but changes after each cycle of catalysis more or less randomly, ranging from tens of milliseconds to seconds (Figs. 11.18, 11.24). This phenomenon is known as "dynamic disorder" (Zwanzig 1990) or "dynamic heterogeneity." The histogram of these waiting times is *asymmetric* and *rugged* (uneven, zigzag, saw tooth-shaped) as evident in Figs. 11.18 and 11.24. The *ruggedness* of the waiting time histogram (defined as the difference between the measured waiting times and the theoretical waiting times calculated from a *smooth curve* that best fits the histogram) can be either due to *experimental error* or , at least in part, to *biological causes*. Lu et al. (1998) found one *smooth* mathematical function called the probability distribution function, p(t), i.e., Eq. 11.25, that fits the waiting time histogram reasonably well (see the upper left-hand panel in Fig. 11.24) but cannot model the ruggedness of the histogram. The conformon-based theory of single-molecule enzymology to be described below can account for both the *smooth* and *rugged* portions of the waiting time histograms.

(B) Increasing the cholesterol concentration by tenfold from 0.2 to 2.0 mM decreased the longest waiting times measured by tenfold and the mode (i.e., the ordinate value under the peak of the distribution curve) of the waiting time by about fourfold (from 150 to 40 ms).

(C) Pairs of waiting times separated by m turnovers are correlated as long as m is less than about 10 (see Fig. 3 in Lu et al. 1998). This observation is known as the "memory effect."

(D) The spectral mean (i.e., the average wave number or wavelength of the fluorescence emission of FAD) of a COx molecule fluctuates by about 1.5% around the mean (see Fig. 5a in Lu et al. 1998).

(E) The autocorrelation function of *waiting times* and that of the *spectral means* are similar (see Fig. 4a, b in Lu et al. 1998).

The clue to explaining these observations mechanistically was provided by the unexpected finding that the waiting time distributions of cholesterol oxidase, Observations **A** and **B**, could be modeled using an equation similar in form to *Planck's radiation formula* derived from the blackbody spectrum (see the lower panel of Fig. 11.24). A detailed analysis of this finding (as outlined in Table 11.9 below) led me to conclude that the *conformon theory of enzymic catalysis* developed in Ji (1974a, b, 1979, 2000, 2004a) applies to cholesterol oxidase. In other words, the single-molecule enzymological data on cholesterol oxidase reported by Lu et al. (1998) can be accounted for by the conformon theory of molecular machines as detailed in Table 11.10. Again, conformons are defined as the sequence-specific conformational strains of biopolymers (proteins, RNA, and

Table 11.9 A comparison between blackbody radiation and enzymic catalysis. For a more detailed discussion of this table, see Footnote 5 in Table 11.10

	Blackbody radiation	Enzymic catalysis
1. Experimental observation	Asymmetric bell-shaped distribution of radiation intensity as a function of wavelengths, $\lambda \sim 1{,}860$	*Asymmetric* bell-shaped and *rugged* histograms of the frequency of catalytic rate constants as a function of waiting times, w. (Lu et al. 1998)
2. System of	$\sim 10^{27}$ oscillators	$\sim 10^{3}$ oscillators
3. System dimension	~1 m	~1 nm
4. Temperature, K	3,000 ~ 5,000	~300
5. Difference	Heat energy, E, absorbed (or emitted) by blackbody under measurement	Rate constant, k, of the enzyme molecule inferred from the measured lifetimes of the fluorescence of the COx coenzyme, FAD
6. Variables[a]	$\lambda \sim 1/E$, or $E = hc/\lambda$	$k \sim 1/w$, or $w \sim 1/k$
7. Formula[a]	$u(\lambda) = (8\pi hc\lambda^{-5})/(e^{hc/\lambda k\_BT} - 1)$	$f(w) = (aw^{-5})/(e^{b/w} - 1)$ $a = 3.5 \times 10^{11}$(or 4.0×10^{14}) $b = 2.0 \times 10^{2}$ (or 6.5×10^{2})
8. Theory[a]	$n(E)\Delta E = g(E)f(E)\Delta E = g(E)e^{-E/k\_BT}\Delta E$	$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger}$ $k \approx e^{-\Delta G\ddagger/k\_BT} = e^{\Delta S\ddagger/k\_B}e^{-\Delta H\ddagger/k\_BT}$
9. Density of states determined by	External constraints (i.e., the geometry of the blackbody cavity, i.e., boundary conditions)	Internal constraints (i.e., the primary, secondary and tertiary structure of the COx molecule)
10. Isomorphism	λ E g(E) e^{-E/k_BT}	w k $e^{\Delta S\ddagger/k\_B}$ $e^{-\Delta H\ddagger/k\_BT}$
11. Conclusion	*Quantization* of radiation energy (*photons*)	*Selection/quantization* of conformational strains, called *conformons* (Ji 1974a, b), based on their ability to control both (a) the size of rate constants of the enzymic reactions they catalyze and (b) the probability of the occurrence of the rate constants. It is assumed that (a) is determined by the *energy content* and (b) by the catalytic *information content* of *conformons*

[a]Symbols are defined as follows: λ = wavelength, E energy, h = Planck constant, c = the speed of light, k = rate constant, w = waiting time, u = a function, k_B or k_B = the Boltzmann constant, f = another function, a, b = constants, T = temperature, $\Delta S^{\ddagger}$ = the entropy of activation, $\Delta H^{\ddagger}$ = the enthalpy of activation, $\Delta G^{\ddagger}$ = the Gibbs free energy of activation, and g = a function

Table 11.10 The suggested explanations of the single-molecule enzymological observations on cholesterol oxidase based on the conformon theory of molecular machines. Extensive footnotes are provided below, numbered 1–10

Observations of Lu et al. 1998	Explanations based on the Conformon Theory of Molecular Machines (Ji 1974b, 2000)
A *Asymmetric waiting time distribution, or dynamic disorder* (Zwanzig 1990)	1. The enzymic activity of COx is supported in part by the heat absorbed by its covalent and non-covalent bonds acting as molecular oscillators just as the blackbody radiation is supported by a system of molecular oscillators, both obeying the Bose-Einstein statistics rather than the Boltzmann statistics (see 5) 2. The *deterministic* shape (or the smooth portion) of the waiting time distribution is due to different content of the *mechanical energy* stored in the conformons of the COx molecule in its ground state (Fig. 11.28) 3. *The nondeterministic* shape (or the rugged portion) of the waiting time distribution (see Fig. 11.24) is attributed to the different *catalytic negentropy* (Ji 1974a) encoded in the evolutionarily conserved amino acid residues constituting conformons (see 2)
B *Concentration dependence of waiting times*	4. The formation of the *real* conformons in COx at the Franck-Condon state depends on the substrate binding–induced stabilization of *virtual* conformons, an example of the *Circe* effect (Jencks 1975) (see Fig. 11.30), and binding interactions are concentration dependent
C *Autocorrelation among waiting times*	5. Within the lifetime of a conformer, more than one cycle of catalysis can take place, each cycle involving the thermal activation of a ground-state (or *static*) conformon to *dynamic* conformons at the Franck-Condon (FC) state and the relaxation of the *dynamic* conformon back to its original or closely related *static* conformons or to another conformer. "Static" and "dynamic" conformons are defined in (**7**) below
D *Fluctuations of the spectral mean*	6. The fluorescence efficiencies of FAD and $FADH_2$ (Fig. 11.16) depend on the conformational states of FAD and $FADH_2$ which are in turn influenced by the conformational state of the binding pocket of the COx molecule
E *Similarity between the autocorrelation functions of waiting times and spectral means*	7. The conformational state of the binding pocket of the COx molecule affects not only the spectral means of the fluorophores but also the cycling between *static* and *dynamic* conformons at the active site of the COx molecule

DNA) deemed *necessary* and *sufficient* to drive all goal-directed molecular motions inside the cell (see Chap. 8 and Sect. 11.3.2).

1. Lu et al. (1998) derived a mathematical equation, Eq. 11.25, to fit the waiting time distribution data of cholesterol oxidase shown in Figs. 11.17, 11.18, and 11.24. More recently Prakash and Marcus (2007) attempted to provide possible molecular mechanisms underlying the *dynamic disorder* phenomenon and the similarity

between the autocorrelation functions of waiting times and spectral means utilizing the concepts of "electrostatic interaction energy" and "solvatochromism." The theory of the COx waiting time distribution proposed here based on the *analogy between blackbody radiation and enzymic catalysis* (Ji 2008a) can provide qualitative molecular mechanistic explanations for all of the observations listed in Table 11.10.

Both the Xie and the Marcus groups agree that the waiting time fluctuations of COx are the result of the conformational fluctuations of the COx molecule. The Xie group made the assumption that different conformations of the COx molecule somehow give rise to different waiting times, without proposing any realistic molecular mechanisms responsible for coupling COx conformational states and the waiting time variations of its coenzyme, FAD. Given this assumption and the additional assumption that the Michaelis-Menten mechanisms shown in Schemes (11.16) and (11.17) hold for single-molecule enzymes, Xie and his coworkers derived the *probability distribution function* for waiting times:

$$p(t) = k_1 k_2/(k_2 - k_1)\left(e^{-k\_1t} - e^{-k\_2t}\right) \tag{11.25}$$

where p(t) is the probability of observing turnover time t and k_1 (or k_1) and k_2 (or k_2) are the time-dependent rate constants appearing in Scheme (11.16). The fitting of Eq. 11.25 to measured waiting time distribution is shown in the upper left panel of Fig. 11.24 (Lu et al. 1998).

Prakash and Marcus (2007) assumed that the *conformations* (i.e., the three-dimensional structure of a protein that can be altered without breaking or forming covalent bonds, Sect. 11.3.2) of a COx molecule, denoted as X(t), can fluctuate on the milliseconds to seconds timescale, despite the fact that the dynamics of proteins in general occurs on timescales ranging from tens of femtoseconds (10^{-15} s) to seconds (Prakash and Marcus 2007, p. 15984; Kurzynski 1997, Fig. 1). Their reasoning proceeds as follows:

(a) Fluctuations in X(t)

⇩

(b) Fluctuations in the "local electrostatic interaction energy at the active site, E(t)"

⇩

(c) Dynamic disorder and fluctuations in spectral means,

where the symbol "⇩" reads "leads to" or "causes." Given these assumptions, Prakash and Marcus (2007) derived the autocorrelation functions for waiting time distributions and for spectral mean fluctuations that are similar, in agreement with observation **E** (Table 11.10). However, it should be pointed out that Prakash and Marcus (2007) did not provide any explanation as to (a) how the slow conformational fluctuations of COx, namely, X(T), can arise in the first place in

view of the rapid oscillatory motions of peptide bonds, and (b) how X(t) mediates the reduction of FAD to $FADH_2$, which probably takes place on the femtosecond timescale, thus implicating the coupling between two events whose time constants differ by a factor of about 10^{10}. In contrast, the theory based on (a) the analogy between blackbody radiation and enzymic catalysis and (b) the generalized Franck-Condon principle (discussed in Sect. 2.2.3) can provide qualitative answers to both these questions (see **D** and **E** in Table 11.10 and (5) below).

2. In 2008, I noticed the similarity between the waiting time distribution of COx enzymic activity reported by Lu et al. (1998) and the blackbody spectrum (see the upper two panels in Fig. 11.24). (http://www.nationmaster.com/encyclopedia/Planck%27 s-law-of-blackbody-radiation). The similarity between the histogram given on the upper left-hand corner and the blackbody spectrum measured at 4,500 K shown in the upper right-hand corner of Fig. 11.24 is particularly striking. This observation motivated me to use Planck's radiation formula (Nave 2009), Eq. 11.26, generalized in the form of Eq. 11.27 (with the X(w) term set to zero), to model the waiting time distribution of Lu et al. (1998), leading to the result shown in the lower portion of Fig. 11.24 (see the smooth curve marked with squares).

The Planck radiation formula which successfully accounted for the blackbody spectrum in 1900 is given in Eq. 11.26 (Nave 2009):

$$u(\lambda, T) = \left(8\pi hc/\lambda^5\right)/\left(e^{hc/\lambda k_B T} - 1\right) \tag{11.26}$$

where $u(\lambda, T)$ is the *spectral energy density*, i.e., the intensity of radiation emitted or absorbed at wavelength λ by the blackbody wall when heated to T K; h is the Planck constant; c is the speed of light; and k_B is the Boltzmann constant.

The equation derived in Ji (2008b) on the basis of the analogy between blackbody radiation and enzymic catalysis is given in Eq. 11.27:

$$p(w) = \left(aw^{-5}\right)/\left(e^{b/w} - 1\right) + X(w) \tag{11.27}$$

where p(w) is the frequency (or probability) of the occurrence of waiting time w, a and b are constants with numerical values of 3.5×10^5 and 2×10^2, respectively, and X(w) is a *nondeterministic* function of w.

It should be noted that Eq. 11.26, which is based on the Bose-Einstein statistics, reduces to Eq. 11.28, known as the Wien's law (Kragh 2000), which is based on the Boltzmann statistics, if the exponential term is much greater than unity:

$$u(\lambda, T) = \left(8\pi hc/\lambda^5\right)e^{-hc/\lambda k_B T} \tag{11.28}$$

Equation 11.27 can also be reduced to a form similar to Eq. 11.28, if $X(w) = 0$ and the $e^{b/w}$ term is much greater than unity, but calculations showed that the exponential term was not much greater than 1 and hence Eq. 11.27 could not be simplified. As explained by Kragh (2000), the Wien's law fitted the short-wavelength portion

of the blackbody spectrum perfectly but underestimated the emission intensities at the long-wavelength regions. It was in an attempt to remedy this shortcoming that Planck was led to invoke the concept of "energy quanta," which allowed him to derive his radiation law, Eq. 11.26. In effect, Planck demonstrated that the blackbody radiation data are better explained in terms of the Bose-Einstein statistics than in terms of the Boltzmann statistics.

Equation 11.27 consists of two terms, referred to as the "deterministic" and "nondeterministic" terms. These could be equally well referred to as "synchronic" and "diachronic" terms, respectively (for the definitions of "synchronic" versus "diachronic" information, see Table 4.1). The deterministic term is isomorphic with Planck's radiation formula, Eq. 11.26. It is important to point out that neither Lu et al. (1998) nor Prakash and Marcus (2007) discussed the possibility of including any *nondeterministic* (also called *diachronic* or *arbitrary*) term, X(w), in their equations, probably because they assumed, as most biological theorists do, that all biological data, including the waiting time distribution histogram, should fit *deterministic equations* as long as they are noise free. Under such an assumption, any experimental data that do not fit deterministic equations such as Eqs. 11.25 and 11.26 would be *logically* regarded as noise and hence automatically excluded from any theoretical considerations. In contrast, it is here assumed that:

(a) The nondeterministic (or *diachronic* or *arbitrary*) term, X(w), which is defined by Eq. 11.27 as the difference between the measured and the predicted values of w, is too large to be discounted as noise (as seems evident in the lower panel of Fig. 11.24).
(b) The nondeterministic term, X(w), in fact carries biological information (yet to be determined and hence the symbol X), most likely encoded in the *evolutionarily conserved set of amino acid residues* constituting conformons. By "evolutionarily conserved set of amino acid residues," I mean something similar to the "evolutionarily coevolving amino acid residues" found in the WW domains by Lockless and Ranganathan (1999), Süel et al (2003), Socolich et al. (2005), and Poole and Ranganathan (2006).

3. One possible mechanism by which the *evolutionary information* encoded in a conformon can influence the probability of the occurrence of a given waiting time w is described in Table 11.11. This table is constructed on the basis of the relations among *configurations*, *conformers*, and *conformons* described in Fig. 11.21. Waiting time, w, is postulated to be determined by *conformers* denoted as $(\ldots)_A$, $(\ldots)_B$, and $(\ldots)_C$ in Column (2), which is consistent with the postulated mechanism of enzymic catalysis described in the right-hand panel of Fig. 11.28. Please note that three different conformers, A, B, and C, are associated with three different waiting times, 51, 56, and 63 ms, as shown in Column (2), and Column (4) lists the numerical value of the deterministic component of the probability of the occurrence of w calculated from Eq. 11.27 with X(w) set to zero. Column (5) lists the numerical values of the nondeterministic component of Eq. 11.27 calculated as the difference between the measured probability value minus the deterministic component, and these differences are attributed to the differences in (or the arbitrariness of) the *evolutionary*

Table 11.11 The catalytic function of the *evolutionary information* carried by conformons generated in different conformers of an enzyme consisting of *n* amino acids. The numerical data appearing in Columns (2), (4), and (5) are experimental values read off from the *x*- and *y*-coordinates of the three pairs of p(w) values centered at the *x*-coordinate of 56 in the lower panel of Fig. 11.24, except the numbers in parentheses in Column (5).N = the number of amino acid residues in Cox. The numbers in Column (3) indicate the position of the (arbitrarily chosen) amino acids in a protein that are thought to constitute a conformon. The number in green indicates the amino acid residue thought to be common to the conformons belonging to a set of functionally related conformers. The numbers in the parentheses in Column (5) are hypothetical

1. Configuration[a]	2. Conformers (w)	3. Conformons[a]	4. Deterministic component of p(w)[b]	5. Nondeterministic component of p(w)
1, 2, 3, . . ., *n*	$(1, 2, 3, \ldots, n)_A$ (51)	(. . . 12, 35, 60, 100, 120, . . .)$_A$	20	(+1.9)
		(. . . 10, 35, 60, 100, 120, . . .)$_A$	20	+2.0
		(. . . 10, 35, 60, 100, 125, . . .)$_A$	20	(+1.8)
	$(1,2, 3, \ldots, n)_B$ (56)	(. . . 45, 55, 60, 111, 128, . . .)$_B$	18	(0)
		(. . . 45, 55, 60, 111, 155, . . .)$_B$	18	0
		(. . . 50, 55, 60, 111, 155, . . .)$_B$	18	(0)
	$(1, 2, 3, \ldots, n)_C$ (63)	(. . . 6, 10, 60, 125, 142, . . .)$_C$	15	(−1.3)
		(. . . 6, 10, 60, 125, 110, . . .)$_C$	15	−2.0
		(. . . 9, 10, 60, 125, 110, . . .)$_C$	15	(−1.9)

[a] See Sect. 3.2 for the definitions of "configurations" and "conformations"
[b] Read off from Fig. 11.24

information carried by conformons depicted as different patterns of amino acid sequences organized in the active site of an enzyme in a given conformation, as illustrated in Column (3). The following points are important to note in connection with Column (3):

(a) A protein consisting of n amino acid residues can generate more than one conformers, designated as $(1, 2, 3, \ldots, n)_A$, $(1, 2, 3, \ldots, n)_B$, etc., in Columns (2) and (3) in Table 11.11, where values within parentheses symbolize the three-dimensional arrangement of the amino acid residues.

(b) *A conformer can generate more than one conformon*, each generated by activating or spatiotemporally ordering a set of, say, five amino acid residues at the active site (see the different sets of five numbers listed within the parentheses in Column 3). All the conformons belonging to a conformer share a subset of say three amino acid residues in common (residues 35, 60, and 100 for Conformer A; residues 55, 60, and 111 for Conformer B; etc.). The conformons belonging to the set of functionally related conformers given in Column (2) share one amino acid residue in common, i.e., residue 60.

(c) A *conformer* is associated with one waiting time as indicated in Column (2) and hence with one value of the probability of occurrence, p(w), as shown in Column (4), thus giving rise to the deterministic component of Eq. 11.27. In contrast, *conformons*, depending on their amino acid residues shown in Column (3), can modulate the probability of the occurrence of w either positively (Rows 2 and 4 in Column 5), negatively (Rows 8 and 10, Column 5), or not at all (Rows 5 and 7, Column 5), thus engendering the nondeterministic term in Eq. 11.27. That is, it is suggested here that *conformers* account for the deterministic term while the genetic information determines the non-deterministic term. This introduces an *extra degree of freedom* for an enzyme to regulate its activity and this extra degree of freedom is identified as a part of the "evolutionary information" discussed by Socolich et al. (2005), "catalytic information," or "catalytic negentropy" discussed in Ji (1974a) (more at (7) below).

(d) The similarities among the numbers in Column (4) of Table 11.11 reflect the similarities among the amino acid sequences of the conformons in Column (3). Thus the amino acid sequence constituting the core of conformons belonging to Conformer A (i.e., 35, 60, 100) is more similar in structure (or closer in what may be referred to as the *conformon space*, an abstract space wherein each point represents one conformon within an enzyme) to Conformon B (i.e., 55, 60, 111) than is Conformon C (i.e., 10, 60, 125), and so are the corresponding p(w) values given in Column (4), i.e., 20 is closer to 18 than is 15. Similar correlations hold between Column (3) and Column (5): The amino acid residues constituting conformons belonging to Conformer A are more tightly centered on (35, 60, 100) than are those constituting the conformons belonging to Conformer C around (10, 60, 125). Consequently, the values of the nondeterministic component of Eq. 11.27 due to the conformons belonging to Conformer A are more tightly centered around +2.0 than are those due to the conformons belonging to Conformer C centered around −2.0.

(e) In calculating the numerical values given in Columns (4) and (5) in Table 11.11, it is assumed that

The closer the *conformers* are in the *conformon space*, so are the values of the *deterministic component* of p(w); the closer the *conformons* are in the conformon space, so are the values of the *non-deterministic* component of p(w). (11.29)

The *conformon space* is characterized by both *energy* (expressed in terms of the familiar Gibbs free energy, $G = E + PV - TS$) and *information* (expressed in terms of the "evolutionary information" of the kind discussed by Socolich et al. (2005), i.e., "sequence information", or "genetic information") and hence belong to the general class of what was referred to as Gnergy Space, gnergy being defined as a complementary union of information and energy (Sects. 2.3.2 and 4.9).

(f) The amino acid sequences given in Column (3) specify the conformons belonging to a conformer, C_i, of an enzyme at its ground state, which undergoes thermal excitation/activation in the time span of t_{ie} (e standing for excitation or

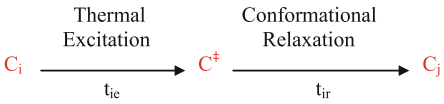


Fig. 11.25 The thermal excitation of ground-state and the relaxation of the transition-state conformations of an enzyme during a catalytic cycle. C_i and C_j are the ith and jth conformational states (or conformers), respectively, and $\mathbf{C}^{\ddagger}$ is the *common* transition state of the enzyme (Fig. 11.28). The symbols t_{ie} and t_{jr} stand for, respectively, "the ith excitation time" or the time required for the conformational transition from C_i to $C^{\ddagger}$ and "the jth relaxation time" or the time taken for the conformational transition from $C^{\ddagger}$ to C_j

activation) to reach the common transition state, $C^{\ddagger}$ (see Fig. 11.28). After catalysis, $C^{\ddagger}$ must relax, in the time span of t_{jr} (r standing for relaxation), back to a ground state, C_j, where j can be equal to or different from C_i (see Fig. 11.25). If the lifetime t_i of C_i is longer than the sum of t_{ie} and t_{jr}, i.e.,

$$t_i = a\left(t_{ie} + t_{jr}\right) \tag{11.30}$$

where a is a positive constant greater than unity, there is a high probability that $C^{\ddagger}$ will relax back to C_i or to C_j where j is in the neighborhood of i. This would provide one possible mechanism to account for the "memory effect," Observation C, in Table 11.10, the observation that long waiting times are likely to be followed by long waiting times and short waiting times are likely to be followed by short waiting times, since $C^{\ddagger}$ is more likely to relax back to C_i than to C_j.

4. Assumption (a) in (2) is experimentally testable. If this assumption is proven to be invalid, the waiting time distribution of Lu et al. (1998) can be completely accounted for *deterministically*, i.e., based on the laws of physics (see Row 3, Table 4.1), either by Eqs. 11.25 or 11.27 with X(w) set to zero. If Assumption (a) is proven to be valid, then the waiting time distribution of Lu et al. (1998) cannot be completely accounted for based on the laws of physics alone (i.e., in terms of *synchronic* laws alone as defined in Table 4.1) but entails invoking additional laws or principles (e.g., *diachronic rules* in Table 4.1) reflecting the evolutionary history of living systems. In other words, if Assumption (a) can be validated by further experiments, it would be possible to conclude that the waiting time distribution of Lu et al. (1998) embodies two orthogonal sets of regularities referred to in Table 4.1 as *synchronic laws* and *diachronic rules*. Generalizing Eq. 11.27, it may be possible to make the following statements:

$$\text{Biological Phenomena} = \text{Synchronic Laws} + \text{Diachronic Rules} \tag{11.31}$$

Biological phenomena embody synchronic laws and diachronic rules that are orthogonal to each other. (11.32)

Evidently, Eq. 11.31 and Statement 11.32 combine both *inexorable* laws of physics and the *arbitrary* rules of biological evolution, reminiscent of the theory

of the *matter-symbol complementarity* advocated by Pattee (1996, 2001, 2008), which has alternatively been referred to as the *von Neumann-Pattee principle of matter-sign complementarity* (Ji 1999b). It is tempting to suggest that Statement 11.32 and similar generalizations be referred to as the *First Law of Biology* (FLB) in analogy to *First Law of Thermodynamics*. The so-called PCS (physics, chemistry, and semiotics) paradigm that is emerging in biology as discussed in Ji (2002a) and Barbieri (2003, 2008a, b, c) may also be viewed as an expression of FLB.

If FLB is true, it may provide a possible answer to the question raised by Frauenfelder in "Plenary Debate: Quantum Effects in Biology: Trivial or Not?" (Abbott et al. 2009):

> If we find a general law that determines or explains life, is it quantum mechanical or classical? (11.33)

If we assume that both quantum mechanics and classical mechanics have *synchronic* and *diachronic* aspects and that FLB indeed represents a general law of biology, applying Eq. 11.31 to the Question 11.33 would lead to the following answer:

> The general laws of life are compatible with both quantum mechanics and classical mechanics. (11.34)

In the thought-provoking debate reported in (Abbott et al. 2009), several participants raised questions of the same type as (11.33), namely, "Is it A or B?," when the most likely answer is "Both A and B." For convenience, we may refer to such questions as the *Frauenfelder questions*. In a less obvious way, the title question itself of the Gran Canaria Debate "Quantum Effects in Biology: Trivial or Not?" may be viewed as a *Frauenfelderian question*, and one possible answer to it may be stated as follows:

> Quantum effects in biology are both trivial and non-trivial, depending on the time and spatial scales involved. (11.35)

More specifically, since the spatiotemporal scales involved here are most likely related to quantum mechanical tunneling, the following generalization may be made:

> Quantum effects in biology are fundamental below the critical threshold of the space and time scales where quantum mechanical tunneling occurs and trivial above this threshold. (11.36)

A specific example of quantum mechanical tunneling playing an essential role in enzymology is provided by the electron transfer reaction between an electron donor, AH_2, and an electron acceptor, B, when they are brought close enough to each other through thermal fluctuations for the electron tunneling to occur (see b in Fig. 8.1).

Statement 11.36 can be viewed as an extension of the concept of the *thermal barrier* that separates macroscopic and molecular machines (Ji 1991, pp. 29–35) to the concept of what is here called the *quantum barrier* that separates molecule machines and quons (Herbert 1987) (i.e., the material entities that exhibit the

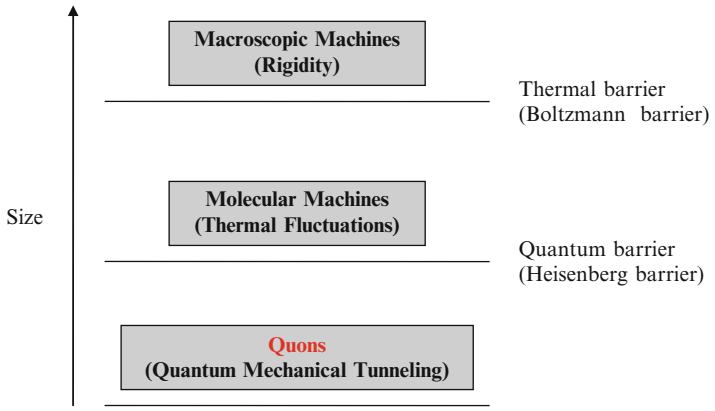


Fig. 11.26 The importance of the size of particles or material systems in determining their physical properties under physiological conditions. Macroscopic machines (e.g., computer) must be large enough to resist the thermal fluctuations of their component parts (to avoid short circuiting) but molecular machines (e.g., enzymes) must be (1) small enough to undergo thermal fluctuations under physiological conditions and yet (2) large enough to prevent, when necessary, the quantum mechanical tunneling of quons (e.g., electrons, protons, etc.) at their active sites. Thus *molecular machines* can be said to be separated from macroscopic machines by what has been referred to as the *thermal (or Boltzmann) barrier* (Ji 1991, pp. 29–35) and from quons by what may be referred to as the *quantum* (or *Heisenberg*) *barrier*

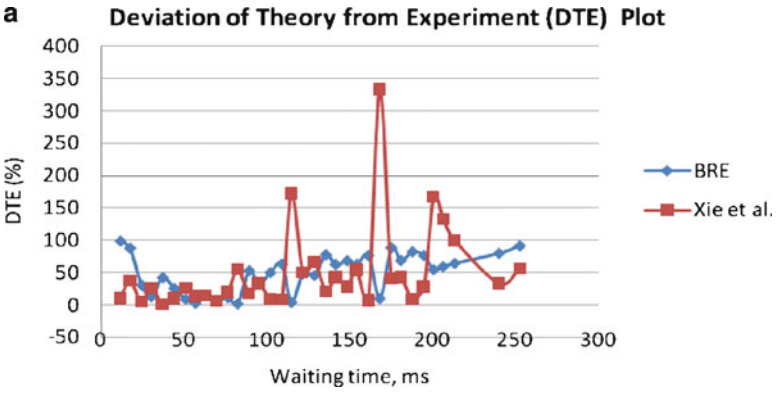
wave-particle duality and the quantum mechanical tunneling, including molecules, atoms, and subatomic particles). This idea is diagrammatically represented as shown in Fig. 11.26.

5. Having discussed the meaning of the nondeterministic term of Eq. 11.27 in some detail, we are now ready to tackle the meaning of the deterministic term of the equation, which will reveal, among other things, the fundamental role of *thermal fluctuations* in enzymic catalysis. As already pointed out, Eq. 11.27 with X(w) set to 0 fits the waiting time distribution fairly well, at least as well as Eq. 11.25 derived by Lu et al. (1998). To quantitatively compare the capabilities of Eqs. 11.25 and 11.27 to fit the experimentally measured waiting time distribution, the following quantity was defined as a measure of the *deviation of the theoretical predictions from measured data* called DTE (*Deviation of Theory from Experiment)*:

$$\mathrm{DTE} = \left(\left(\left(\text{Calculated w} - \text{Measured w}\right)/\text{Measured w}\right)^2\right)^{1/2}\Big) \times 100 \qquad (11.37)$$

DTE represents the absolute value of the deviation of the theoretically predicted w from the measured w expressed as a fraction of the measured w. Figure 11.27 shows the results of applying Eq. 11.37 to Eqs. 11.25 and 11.27. The average DTE values are about the same for these two equations but the standard deviations and the coefficients of variations are about twice as large for Eq. 11.25 as

a



b

Parameters	BRE	Xie et al.
Mean DTE (%)	49.0	49.5
Standard Deviations	30.5	65.9
Coefficient of variations (CV)	62.2	133.2

Fig. 11.27 The deviation of theory from experiment (DTE), the distances between the waiting times measured and those predicted as defined by Eq. 11.35. (**a**) (*Diamond*) Calculations were performed using the blackbody radiation-like equation (BRE) given in Eq. 11.27 with X(w) = 0. (*Square*) Calculations based on the distances between the measured and predicted values read off from Fig. 1d in Xie and Lu (1999). (**b**) A statistical comparison between the BRE and Xie et al. models

for 11.27 (see the table in Fig. 11.27b), indicating that BRE fits the single-molecule enzymological data better than the bi-exponential function of Lu et al. (1998).

The results shown in Fig. 11.27 indicate that Eq. 11.27 derived on the basis of the analogy between blackbody radiation and enzymic catalysis is superior to Eq. 11.25 derived on the basis of the Michaelis-Menten mechanism (Lu et al. 1998).

The interesting question that now arises is:

> What is the theoretical rationale, if any, for the apparent analogy between blackbody radiation and enzymic catalysis? (11.38)

The most important reason for the similarity between *blackbody spectrum* and the *waiting time distribution* of COx may be traced to the fact that both the blackbody wall and the enzyme molecule consist of *systems of molecular oscillators* (see Row 2 in Table 11.9). A blackbody is a system of approximately 10^{27} oscillators that are vibrationally and electronically excited from their common ground-state energy level, E_0, to various excited states denoted as E_i, where i runs from 0 to n, the total number of energy levels available to each oscillator (see the left panel in Fig. 11.28). On the other hand, the COx molecule is a system of about 10^5 atoms linked to form a three-dimensional network of covalent bonds forming

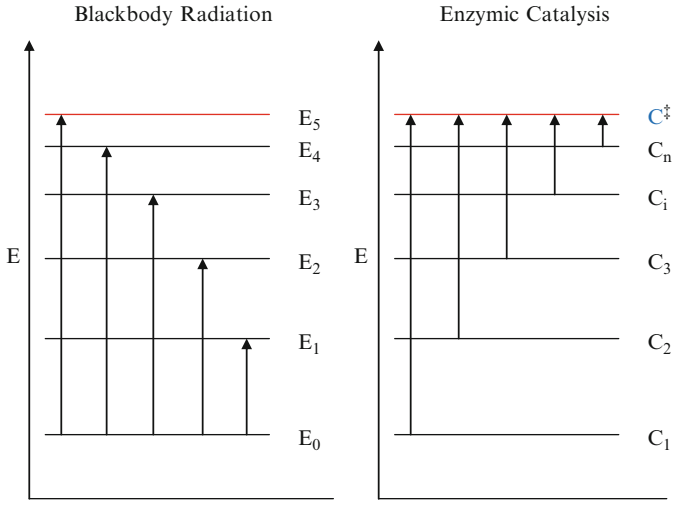


Fig. 11.28 A comparison between *blackbody radiation* and *enzymic catalysis*. (*Left*) Blackbody radiation involves promoting the energy levels (vibrational, electronic, or vibronic, i.e., both vibrational and electronic) of oscillators from their ground state E_0 to higher energy levels, E_1–E_5. The wavelength of the radiation (or quantum) absorbed or emitted is given by $\Delta E = E_i - E_0 = hf$, where E_i is the ith excited-state energy level, h is the Planck constant, f is the frequency, and ΔE is the energy absorbed when an oscillator is excited from its ground state to the ith energy level. Alternatively, blackbody radiation can be thought of as resulting from the transitions of electrons from one energy level to another within matter, e.g., from E_1 to E_0, from E_2 to E_0, etc. (*Right*) A single molecule of cholesterol oxidase (COx) is postulated to exist in n different conformational states (i.e., conformers, also called *conformational substates* by Frauenfelder et al. 2001), denoted here as C_i with i running from 1 to n. Each *conformational state* (or *conformer*) is thought to carry a set of sequence-specific conformational strains, or *conformons*, as explained in Fig. 11.20 and can be excited to a common transition state (denoted as $C^{\ddagger}$) by thermal fluctuations

the skeleton of a *globular protein*, each covalent bond acting as an oscillator (with vibrational frequencies in the range of 10^{14}/s [Kurzynski 1997, 2006]) which, when coupled properly, can lead to low-frequency *collective modes* of oscillations with frequencies as low as 10^3/s or less (according to the Fourier theorem [Herbert 1987]), thus accounting for the genesis of the so-called the slow protein coordinate, X(t), of Prakash and Marcus (2007).

One crucial difference between *blackbody radiation* and *enzymic catalysis* is thought to be this: Although both the *blackbody* and the *enzyme molecule* absorb heat or energy from their environment (see the upward arrows in Fig. 11.28) and reemit it to their environment in equal amounts at equilibrium, what is experimentally measured from these two systems is different: From the blackbody, the energy reemitted is measured (see the blackbody spectrum in the upper right-hand corner of Fig. 11.24), while from COx, the disappearance of the fluorescence emission is measured that results from enzymic catalysis, the *precondition of which being energy absorption* (see the histogram in the upper left-hand corner of Fig. 11.24).

In other words, what is measured from the blackbody is what it absorbs from its environment, but what is measured from the COx molecule is not what it absorbs from its environment but what *that absorbed energy helps the enzyme molecule to accomplish (e.g., a → b transition in Fig. 11.19), namely, the removal of the fluorescence from its coenzyme, FAD, by catalyzing its reduction to the non-fluorescent* $FADH_2$. It is important to point out that the heat absorbed by an enzyme cannot perform any molecular work, including catalysis, *since no thermal energy can drive any work under isothermal conditions*, according to the Second Law of thermodynamics (see Sect. 2.1.4). However, according to the new version of the Second Law as formulated by McClare in 1971 (see Statement 2.5 in Chap. 2), an enzyme can utilize thermal energy if it can be paid back or returned to its environment within its turnover time, leading to the following generalization:

> An enzyme can utilize thermal energy to facilitate catalysis without violating the Second Law, if the chemical reaction being catalyzed can release heat to the environment within the times shorter than τ, the turnover time of the enzyme. (11.39)

We may refer to Statement 11.39 as the *First Law of Enzymic Catalysis*, to emphasize the fundamental role that thermal energy, as manifested in molecular motions (or Brownian motions), plays in enzymic catalysis, the foundational process of the phenomenon of life.

The blackbody spectrum measures all the thermal photons absorbed by the blackbody wall, but the waiting time distribution of COx measures only a small fraction of the thermal photons absorbed by the COx molecule that helps COx to reach its transition state, $C^{\ddagger}$. This difference is visualized in Fig. 11.28 in terms of the *multiple levels* (labeled E_1 though E_5) of thermal excitations for blackbody radiation on the one hand and the *fixed level* of thermally activated transition state in the COx molecule labeled $C^{\ddagger}$ that triggers catalytic event on the other.

The reason that Planck's radiation formula, Eq. 11.26, can account for the waiting time distribution of COx, in the form of Eq. 11.27 with X(w) set to zero, is probably because of the mechanistic isomorphism between blackbody radiation and enzymic catalysis as indicated by the similar distributions of the upward arrows in the two panels in Fig. 11.28: A set of arrows starting from a common ground state reaching different activated states on the one hand and a similar set of arrows starting from different ground states arriving at the common activated state, $C^{\ddagger}$, on the other. The rationale for invoking different ground states for the thermal activation in COx is given in (6) below. Due to the varying levels of conformational energy associated with C_i, the thermal energies required for C_i to reach the common transition state, $C^{\ddagger}$, differs as indicated by the varying lengths of the upward arrows in the right-hand panel of Fig. 11.28.

6. It is well known that a protein molecule contains many internal mechanical (i.e., conformational) strains variously referred to as "mobile defects" (Lumry 1974; Lumry and Gregory 1986), "frustrations" (Anderson 1983, 1987), or conformons (Green and Ji 1972a, b; Ji 1974b, 2000) as already indicated in Sect. 11.3.2. The number and locations of these mechanical strains within an enzyme probably vary from one conformer to another (Fig. 11.21). The conformers of a protein

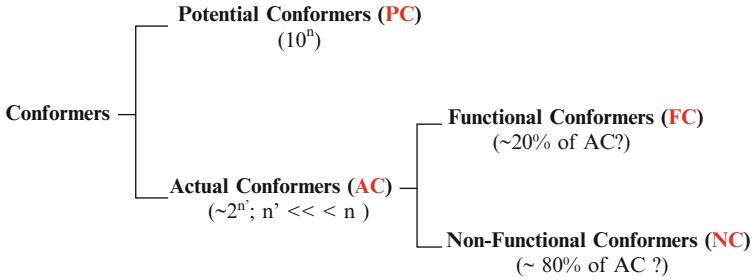


Fig. 11.29 The postulated hierarchical organization of the conformers of a protein. The number of PC is estimated to be 2^n, where n is the number of the amino acid residues, and the number of AC is thought to be far less than that of PC, i.e., AC <<< PC, and those of FC and NC are probably comparable: FC ~ NC. The number n appears in both Figs. 11.23 and 11.28, but Fig. 11.23 deals with the number of all possible linear sequences of *n nucleotides* (i.e., configurations) whereas Fig. 11.28 is concerned with the number of all possible three-dimensional arrangements of a given linear sequence of n amino acid residues (i.e., conformations or conformers). The number of potential n-nucleotide sequences is probably smaller than the number of the conformers theoretically possible for a protein with n amino acid residues, because the former is 4^n whereas the latter is x^n where x is the number of possible orientations that an amino acid residue can assume as it is being added to the growing polypeptide chain during protein synthesis, x being most likely larger than 4

are denoted as C_i in Fig. 11.28, where i runs from 1 to n, the total number of conformers belonging to a molecule of COx which can be astronomically large. Since COx has 504 amino acid residues, n could be as large as 2^{504} or about 10^{152}, assuming that each amino acid residue added during protein synthesis on the ribosome leads to at least two new conformational states for the resulting polypeptide. Of these theoretically predicted large number of conformers of COx, it is expected that only a small number, n′, would be selected by evolution to be realized in COx under the intracellular environmental conditions: i.e., n′ <<< n. It is difficult to estimate the precise magnitude of n′. Whatever the actual size of n′ will turn out to be, it is probably reasonable to assume that most, if not all, of these n′ conformers serve some biological functions in order for them to be evolutionarily conserved. Thus n′ conformers may be divided into two classes – *functional* and *nonfunctional* under a given environmental condition. The n conformers that are theoretically predictable, whether realized or not in cells, represents a third class to be referred to as *potential* conformers in contrast to *actual* conformers which divide into *functional* and *nonfunctional* ones. Thus, conformers can be classified as shown in Fig. 11.29.

As discussed in Sect. 11.3.2, the three terms *molecules, conformers,* and *conformons* are distinct and their mutual relations may be simply summarized as follows:

A molecule is a set of confomers; a conformer is a set of conformons. (11.40)

A conformer of an enzyme may contain one or more conformons belonging to two different classes: (a) the conformons (i.e., conformational kinks) introduced into an enzyme during its biosynthesis on ribosomes and (b) the conformons

generated from exergonic processes (i.e., binding, de-binding, covalent bond rearrangements, etc.) catalyzed by the enzyme itself such as the myosin head during muscle contraction (see (d) (3) in Fig. 11.33). We will refer to the former as *static* (or *intrinsic*) *conformons* (also called the Klonowski-Klonowska conformons in [Ji 2000]) and the latter as *dynamic* (or *extrinsic*) *conformons*. These terms were already employed in Rows (C) and (E) in Table 11.10 and are further discussed in Sect. 11.4.2. It is postulated here that the nonzero ground energy levels of the conformers postulated in Fig. 11.28 result from the presence of one or more *static conformons* in each conformer.

A functional *conformer* is postulated to carry as many *conformons* as the number of elementary processes it catalyzes (see Figs. 11.23, 11.28), which may be at least two, namely, binding and de-binding processes. In order for a conformer to carry out its catalytic act (be it binding, de-binding, or electronic rearrangement, singly or in combinations), it must satisfy the following two requirements: (a) a conformer must be thermally activated/excited to reach the transitions state $C^{\ddagger}$ (as indicated in Fig. 11.28) and (b) a conformer must recruit the right set of x out of its n amino acid residues obeying the conformon equation, Eq. 11.19.

Satisfying these requirements is tantamount to generating a set of *dynamic* conformons by a conformer within its lifetime.

It is natural to associate with COx the so-called slow protein coordinate, X(t), that plays a central role in the theory of waiting time distribution of COx proposed by Prakash and Marcus (2007). COx can be viewed as a system of atoms linked by covalent bonds, each of which acts as an oscillator with vibrational periods in the 10^{-14} s range (or vibrational frequencies in the 10^{14}/s range), as already indicated. According to Fourier's theorem (Herbert 1987, pp. 79–92; http://en.wikipedia.org/wiki/Fourier_series), it is possible to generate low-frequency collective vibrational modes by appropriately coupling these high-frequency primary oscillators. Nature may *slow down* molecular oscillations by increasing the effective mass of oscillators through compactifying or "chunking" lower-level molecular systems to higher level ones in steps of about 5 (Ji 1991, pp. 52–56), as exemplified by the chunking of a DNA double helix into a chromosome (Fig. 2.9) reducing the linear size of DNA by a factor of about 10^{10} and therefore its effective mass by a factor of 10^{30}. This in turn should decrease oscillatory frequency by a factor of $(10^{30})^{1/2}$ or 10^{15}, since frequency, f, is inversely proportional to the square root of the mass, m, according to the equation for simple harmonic oscillators,

$$f = (1/2\pi)(k/m)^{1/2} \tag{11.41}$$

where k is the spring constant. Eq. 11.41 also suggests that the oscillatory frequency can be reduced by reducing the spring constant, k. Thus, when large domains of a protein oscillate with respect to one another, the frequency of oscillation can be reduced not only by increasing effective mass but also by weakening the spring constant k by, say, changing local electric fields by protonation-deprotonation, phosphorylation-dephosphorylation, and acetylation-deacetylation reactions involving critical amino acid residues.

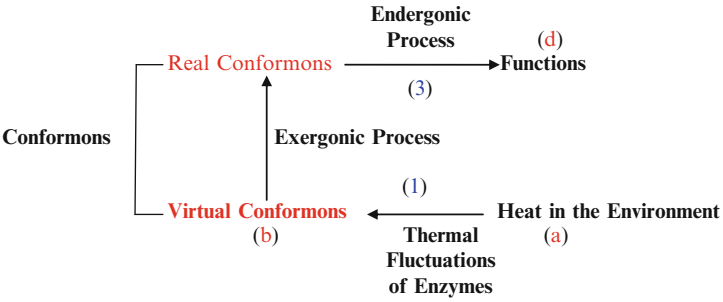


Fig. 11.30 A network representation of *the architectonics of the conformon*. The network consists of four nodes, (a)–(d), and three edges, (1)–(3), all of which are indispensible in defining the structure and function of the conformon, *sequence-specific conformational strains of biopolymers that are postulated to drive all goal-directed molecular motions in the living cell* (Ji 1974b, 1985b). Thermal fluctuations are alternatively referred to as *Brownian motions*

7. Indirect evidence gleaned from the studies carried out by molecular biologists, e.g., Socolich et al. (2005), Lockless and Ranganathan (1999), and by enzyme kineticists, e.g., Northrup and Hynes (1980) and Kurzynski (1993, 1997, 2006), suggests that within the lifetime of a conformer, more than one conformons are likely to be produced. The relation among a *conformer,* a *conformon*, and *amino acid residues* is triadic in that the production of a conformon requires (a) a set of amino acid residues, (b) the mechanical energy associated with (or stored in) their nonequilibrium arrangements in space and time, and (c) the function associated with the conformer harboring the conformon under consideration.

We may summarize the relation among these three entities as follows:

A conformer (at the $(n + 1)^{th}$ level) provides the biological meaning of the conformons (at the n^{th} level) that are generated within it utilizing a subset of its amino acid residues (at the $(n-1)^{th}$ level) through their interactions with an exergonic process such as ligand binding or chemical reactions. (11.42)

We may refer to Statement 11.42 as the *triadic architectonics of the conformon* (TAC), the term "architectonics" referring to the *design principles of an architecture or a building* as already alluded to previously. The molecular mechanisms implementing TAC require introducing two more concepts – *virtual conformons* and *thermal fluctuations* (see (1) and (b) in Fig. 11.30):

The heat or thermal energies supplied by the environment of living systems (see (a) in Fig. 11.30) are essential for thermal fluctuations (also known as Brownian motions) (1) which are in turn prerequisite for all living processes, which leads to the following generalization:

No thermal fluctuations, no life. (11.43)

To emphasize the fundamental significance of *thermal fluctuations* in understanding the phenomenon of life on the molecular level, Statement 11.43 may be referred to as the *Second Law of Biology,* the First Law of Biology being given in Statement 11.32.

Virtual conformons (b) are those conformational strains that are produced spontaneously and transiently as the result of the thermal fluctuations of a conformer of an enzyme. Because of the constraints imposed by the Second Law of Thermodynamics, *virtual conformons* cannot be utilized to do any molecular work, and this is ultimately because they are produced under isothermal environment without any sustained thermal energy gradient. However, if, during their transient life times, virtual conformons can participate in some exergonic processes such as ligand bindings and electronic rearrangements in substrates, *virtual conformons* can be converted to *real conformons* (see Process 2 in Fig. 11.30) as long as the following conditions are satisfied:

(1) The rate of free energy release from the exergonic process (e.g., ligand binding or electronic rearrangement) is fast relative to the conformational relaxation times of the enzyme so that the *generalized Franck-Condon principle* is not violated (see Sect. 2.2.3)
(2) A part of the energy released from the exergonic process pays back to the environment the thermal energy borrowed by virtual conformons during their thermal genesis.
(3) The exergonic process produces effects (e.g., reaction products) that stabilize virtual conformons (Ji 1979).

If Step (2) involves a ligand binding, the rate of the formation of real conformons (c) would depend on the concentration of the ligand. Observation B in Table 11.10 that a tenfold decrease in cholesterol concentration led to a tenfold increase in the largest waiting time supports Step (2). *Real conformons*, once formed, can persist for arbitrarily long times until they are utilized to drive some *goal-directed endergonic process* (3) otherwise known as *functions* (d). In contrast, the lifetimes of *virtual conformons* are limited (due to the Second Law of Thermodynamics) to times shorter than the enzyme turnover times (Sect. 2.1.4) (McClare 1971, 1974; Welch and Kell 1986).

It is clear from Fig. 11.30 that the so-called Second Law of Biology, Statement 11.43, is only one of the many necessary conditions for the production and utilization of the conformon, which was postulated to be the quantum of life (Ji 2000). Another necessary condition may be stated as follows:

No coupling between exergonic and endergonic processes, no life. (11.44)

Examples of the coupling between exergonic (i.e., free energy supplying) and endergonic (i.e., free energy consuming) processes include the coupling between the exergonic respiration and endergonic phosphorylation of ADP to produce ATP (Ji 1974b, 2000). Statement 11.44 may well deserve to be referred to as the *Third Law of Biology*. Finally, since all of the first three laws of biology are necessary to generate conformons, according to Fig. 11.30, it may be logical to refer to the following statement as the *Fourth Law of Biology*:

No conformons, no life. (11.45)

8. A conformer of COx can contain or harbor one or more conformons (whose number can be estimated from the *conformon equation* discussed in Sect. 11.3.2

and Fig. 11.21; see also Column (3) in Table 11.11). Each conformon is thought to catalyze one elementary step of a goal-directed process for which a conformer of COx is postulated to have been selected by evolution. It is also assumed in Fig. 11.21 that each conformer of COx must be thermally activated/excited to reach the transition state, $C^{\ddagger}$, which is essential for catalyzing the overall process of the electron transfer from cholesterol to FAD to produce $FADH_2$. Because different conformers are at different distances from the transition state $C^{\ddagger}$ (Fig. 11.28) and hence associated with different activation free energy, $\Delta G^{\ddagger}$, different conformers are associated with different *waiting times, w,* for catalyzing the FAD reduction by cholesterol. (For the relation between $\Delta G^{\ddagger}$ and w, see Rows 6 and 8 in Table 11.9 and Eq. 11.46.)

According to Eq. 11.27, waiting time is determined by a deterministic component which is the function of $\Delta G^{\ddagger}$ as indicated in Rows (6) and (8) in Table 11.9 and a nondeterministic component which is thought to result from the genetic information of the conformons (also called "catalytic negentropy" in Ji 1974a) or "evolutionary information" in Socolich et al. (2005) and Poole and Ranganathan (2006) (see further explanation given below).

The probability of a photon with wavelength λ to be absorbed or emitted by the blackbody at equilibrium is given by the product of two terms: (a) the *Boltzmann factor*, $e^{-E/kT}$, giving the probability of an energy level, E, to be occupied, and (b) the statistical weight of the energy level E (i.e., the *density of states*, n(E) in Row 9, Table 11.9), which is determined by the number of modes of the standing waves that can be accommodated by the blackbody cavity (Nave 2009).

Similarly, it is postulated here that the deterministic portion of Eq. 11.27 specifying the frequency of the occurrence of waiting time w is given by the product of two factors:

(a) The *Bose-Einstein factor* (which applies to conformons assumed to be members of the boson family of quons), $1/(e^{b/w} - 1)$, where b is a constant and w is related to activation free energy as shown in Eq. 11.46
(b) The density of states determined by the aw^{-5} term, where a is another constant reflecting the geometrical properties of COx including the "internal constraints" of the primary, secondary, and tertiary structure of COx (see Row 9 in Table 11.9) which differs from the "external constraints" such as size of the reaction vessel

Based on the transition state theory of reaction rates (Frost and Pearson 1965, pp. 97–102; Kurzynski 2006, pp. 169–171), the waiting time, w, can be related to the rate constant k, which in turn is related to the thermodynamic variables as shown in Eq. 11.46:

$$w = 1/k = (h/k_B T)e^{\Delta G^{\ddagger}/RT} = (h/k_B T)(e^{-\Delta S^{\ddagger}/R})(e^{\Delta H^{\ddagger}/RT}) \qquad (11.46)$$

where $\Delta G^{\ddagger}$, $\Delta S^{\ddagger}$, and $\Delta H^{\ddagger}$ are the activation free energy, activation entropy, and activation enthalpy, respectively; k_B is the Boltzmann constant; T is the absolute

temperature; h is the Planck constant; and R is the gas constant. Since $\Delta H^{\ddagger} = \Delta E^{\ddagger} + P\Delta V^{\ddagger}$, where P is the pressure and $\Delta V^{\ddagger}$ is the activation volume of COx, $\Delta H^{\ddagger} = \Delta E^{\ddagger}$ if $\Delta V^{\ddagger}$ is zero, which is generally assumed to be true, thus transforming Eq. 11.46 to 11.47:

$$w = 1/k = (h/k_B T)(e^{-\Delta S^{\ddagger}/R})(e^{\Delta E^{\ddagger}/RT}) \tag{11.47}$$

Equation (11.47) is equivalent to Eq. 2 in Ji (1974a, p. 420), based on which the following postulate was made in Ji (1974a, pp. 434–435):

> . . .the enzyme can control the rate of chemical reactions by affecting either $\Delta E^{\ddagger}$ or $\Delta S^{\ddagger}$. The enzyme can alter the magnitude of $\Delta E^{\ddagger}$ by changing the curvature of the potential energy hypersurface of the substrate, which it does by undergoing appropriate conformational transitions . . .. I now postulate that the catalytic negentropy stored in the enzyme can regulate the rate of enzymic reactions by modulating the magnitude of $\Delta S^{\ddagger}$. I regard the enzyme as a negentropy reservoir from which the substrate can borrow negentropy (i.e., negative entropy) change [Sect. 2.1.5]to reach the transitions state. At the end of an elementary chemical reaction, the negentropy withdrawn from the enzyme can be returned to the protein, so that no net changes in the negentropy content of the enzyme may result. Although the vast kinetic energies stored in the enzyme cannot be localized on the substrate as a means of catalysis, because of the second law, I believe that the transfer of negentropy from the enzyme to the substrate can be invoked for the purpose of catalysis without violating the second law. The molecular mechanism of negentropy transfer from the enzyme to the substrate may be identified with the ordering of the catalytic groups in the reaction cavity in the Franck-Condon state and the disordering of structures elsewhere in the enzyme. (11.48)

The *negentropy transfer* from domain A to domain B, the active site, in an enzyme to reach the transition state can affect $\Delta S^{\ddagger}$ (and hence alter the associated waiting time, w; see Eq. 11.46) in three ways" (a) *positively*, increasing $\Delta S^{\ddagger}$; (b) *negatively*, decreasing $\Delta S^{\ddagger}$; and (c) *neutrally*, having no effect on $\Delta S^{\ddagger}$. Although some negentropy transfer from A to B may not affect $\Delta S^{\ddagger}$ and associated w, it can still affect the probability, p(w), of the occurrence of w, again in three ways – (d) *positively*, increasing p(w); (e) *negatively*, decreasing p(w); and (f) *neutrally*, having no effect on p(w) – depending on the amino acid sequence of the conformons involved, as illustrated in Column (5) in Table 11.11.

The term "negentropy" should be interpreted with caution. This term is paradoxical in the sense that when defined as "negative entropy change (NEC)," negentropy signifies or is associated with "order" and "information," but, when defined as "negative entropy (NE)," it violates the Third Law of thermodynamics, according to which entropy of a thermodynamic system can never be less than zero (or negative). This situation was referred to as the *Schrödinger's paradox* in Sect. 2.1.5. Please note that NEC is a *differential* concept whereas *NE* is an absolute one.

Therefore the phrase "negentropy reservoir" appearing in Statement 11.48 can be interpreted as those regions or domains of an enzyme serving as storage sites for negentropy in the form of *ordered* or *low-entropy conformational structures*. The key message of Statement 11.48 is that an enzyme can utilize the negentropy stored in one region of an enzyme to organize the catalytic residues at the active site to reach the transition state, and this process can be viewed as a form of *negentropy*

Table 11.12 The *garage-door metaphor* of enzymic catalysis

	Garage door	Enzyme
1. Storage device	Mechanical spring	Nonactive site domains (domain A or negentropy reservoir, NR)
2. What is stored	Mechanical energy	Negentropy (low-entropy local structure)
3. Mechanism of storage	*Gravitational potential energy* of the door is transduced into the *mechanical energy* of the spring as the door closes	Negentropy is generated at domain A allosterically (in the form of a *conformon*) as the substrate binds to domain B, the active site
4. Mechanism of utilization	*The mechanical energy* of the spring is transferred to the raised door as its *gravitational potential energy*	The negentropy stored in domain A is transferred to domain B, the active site, as a low-entropy conformation of the catalytic residues needed for catalysis, without *necessarily* changing the overall activation entropy, $\Delta S^{\ddagger}$ of the enzyme-substrate complex
5. Net result	Less energy is required to open the garage door	The observed probability of the occurrence of waiting times can deviate from the deterministic values (due to the ruggedness of the histogram) either positively, negatively, or neutrally, thereby increasing the variety of the probabilities, p(w), of the occurrence of w

transfer from domain A (i.e., the negentropy reservoir) to domain B (i.e., active site), leading to the disordering of A and ordering of B without changing the total amount of the negentropy stored in A and B during the transfer process. This idea can be conveniently expressed using the *garage door* as a metaphor:

> Just as a part of the gravitational potential energy of the garage door is stored in mechanical springs as the door closes and the stored energy is subsequently re-utilized when the garage door opens, so the negentropy generated (allosterically) within domain A in an enzyme when a substrate binds to domain B, or the active site, can be transferred to B as "catalytic negentropy" which organize the amino acid residues to effectuate catalysis, returning the catalytic negentropy back to A as the activated enzyme relaxes to its ground state. (11.49)

Statement 11.49 will be referred to as the *garage-door mechanism* of enzymic catalysis or the *garage-door postulate of enzymic catalysis*, and its salient features are summarized in Table 11.12.

Two features of the garage-door postulate deserve special attention:

(a) The existence of a group of amino acid residues that is postulated to function as a *negentropy reservoir* (NR) (Row 1, Table 11.12) at domain A. Domain A need not be located at the active site (domain B) making direct contact with a bound substrate but can involve a part or the whole of the enzyme other than the

active site, as long as domain A is allosteric (i.e., mechanically coupled) to domain B. Because of their postulated role in catalysis, it can be predicted that these *NR residues* in domain A will be evolutionarily conserved and coevolve with the *catalytic residues* in domain B, similar to the critical residues found in the WW domain proteins (Lockless and Ranganathan 1999; Süel et al 2003; Socolich et al. 2005, and Poole and Ranganathan 2006). It is postulated here that the NR residues are responsible for the *ruggedness* of the waiting time histograms (Fig. 11.24). Therefore *it can be predicted that mutating one or more of the NR residues of COx will remove the ruggedness of the waiting time histograms* (see the lower panel of Fig. 11.24). (This prediction was suggested to me on July 28, 2009, by one of my undergraduate students at Rutgers, Ms. Julie Bianchini.)

(b) The *iso-entropic nature* (i.e., no change in $\Delta S^{\ddagger}$) postulated for the negentropy transfer process from NR to the active site of the enzyme (Row 4, Table 11.12) is the most salient feature of the *garage-door postulate*. Lumry and Gregory (1986, p. 23) discussed enzymic mechanisms similar to the garage-door mechanism described here (Lumry 1974; Lumry and Gregory 1986).

If these conjectures are correct, we can conclude that *blackbody radiation* and *the waiting time distribution of COx* are isomorphic with each other as summarized in Row (10) of Table 11.9.

9. Armed with the conclusions drawn above, we can return to Eq. 11.7 with a new perspective, namely, the view that the X(w) factor does not merely reflect experimental noise but also carries biological (i.e., genetic or evolutionary) information. We will refer to this view as the "X factor hypothesis," which may be expressed as follows, among others:

   The waiting time distribution of COx contains both deterministic and non-deterministic (or stochastic) components. (11.50)

   The waiting time distribution of COx contains the deterministic and non-deterministic components, the former being constrained by, and the latter harnessing, the laws of physics and chemistry. (11.51)

   The waiting time distribution of COx contains the deterministic and non-deterministic components, the former being constrained by the *laws of physics and chemistry* and the latter embodying the *evolutionary information* that harness the laws of physics and chemistry. (11.52)

   Enzymic catalysis embodies two complementary aspects – the *energy/matter aspect* obeying the laws of physics and chemistry and the *bioinformatic aspect* embodying the evolutionarily selected amino acid residues capable of harnessing the laws of physics and chemistry. (11.53)

It should be noted that the adjectives "deterministic" and "nondeterministic" appearing in the above statements can be replaced by their synonyms, "synchronic" and "diachronic," respectively, as already indicated (see Row 3 in Table 4.1).

10. The *waiting time distribution* of COx measured by Xie and his colleagues in Lu et al. (1998) may be comparable to the *blackbody spectrum* measured by Otto Lummer and Ernst Pringsheim in 1899 (Kragh 2000). Also the probability

Table 11.13 A suggested historical analogy between blackbody radiation and single-molecule enzymology. COx = cholesterol oxidase

	Blackbody radiation	Single-molecule enzymology
1. Data	Blackbody spectrum	Waiting time distribution of COx
2. Early theory	Wien's law Rayleigh-Jeans law	Probability distribution function for waiting times derived on the basis of the *Michaelis-Menten kinetics* (Lu et al. 1998; Kurzynski 2006; Qian and Xie 2006; Prakash and Marcus 2007)
3. Final theory	Planck's law	Probability distribution function for waiting times derived from the *blackbody radiation-enzymic catalysis analogy* (Ji 2008b)
4. Improvement	*Better fit* in the low frequency region of the blackbody spectrum	Better fit for the "noisy" or "rugged" portions of the waiting time distribution by introducing the *X(w) factor as a measure of evolutionary information carried by catalytic residues*
5. New concept	Energy quantization (photons)	Conformons as carriers of (a) quantized energy and (b) catalytic information or catalytic negentropy (Ji 1974a)

distribution for waiting times derived by Lu et al. (1998), Eq. 11.25, may be comparable to Wien's law and Rayleigh-Jeans law (Kragh 2000; Nave 2009), just as that derived on the basis of the analogy between blackbody radiation and enzymic catalysis, Eq. 11.27 (Ji 2008b), is comparable to Planck's law, Eq. 11.26. These and other analogies between *blackbody radiation* and *enzymic catalysis* are summarized in Table 11.13.

If the *blackbody radiation-enzymic catalysis* analogy turns out to be true, the single-molecule waiting time distribution measured by Lu et al. (1998) may be viewed as the first direct experimental evidence for demonstrating that an enzyme molecule must absorb heat before it can carry out catalysis, *thus establishing the fundamental role of heat in molecular biology*, consistent with the "thermal barrier" hypothesis of molecular machines (Ji 1991, pp. 29–31) and the Second Law of Molecular Biology, Statement 11.43. Although biochemists have known for a long time that raising temperature leads to an increase in catalytic rates (the so-called Q_{10} value of an enzymic reaction being in the 2–4 range), these rates are ensemble averages that are affected by many physicochemical steps, making it difficult to pin down the precise catalytic process affected by heat. This difficulty is largely overcome in single-molecule kinetic experiments where one elementary step (e.g., the electron transfer from cholesterol to FAD, Scheme 11.16) can be studied.

Unlike in chemical reactions where heat provides all of the energy required to overcome the activation energy barrier, an enzyme molecule appears to supplement the heat energy absorbed from its environment with its *intrinsic* ground-state potential energy stored in local conformational strains (called *static* or *intrinsic conformons*; see (6) above) in activating its substrate to form the product. A similar

conclusion was reached by Rufus Lumry in his "The Protein Primer" available online (Lumry 2009).

If the content of Table 11.9 is true, there may be an interesting historical analogy between *blackbody radiation* in physics and *enzymic catalysis* in biology as summarized in Table 11.13. Just as the blackbody spectrum led Max Plank in 1900 to postulate the existence of *quantized energy packets* (later called "photons"), so the single-molecule kinetic data of COx suggest that:

(a) There exists a minimal unit of *c*atalytically *e*ffective *c*onformational *s*trains (CECS) of an enzyme molecule
(b) The minimal unit of CECS is characterized by its mechanical energy and *evolutionarily conserved amino acid residues* encoding genetic information (Row 5, Table 11.13)

The CECSs invoked above can be identified with *conformons* since conformons are defined as the conformational strains localized at sequence-specific sites within biopolymers carrying both *mechanical energy* and *genetic information* to drive goal-directed molecular motions in the cell (Chap. 8) (Ji 1974a, b, 2000). In other words, the *blackbody radiation-enzymic catalysis analogy* entails invoking two complementary factors underlying catalysis" (a) the *conformational energy* stored in an enzyme (see Fig. 11.28) and (b) the *catalytic information* (or negentropy Ji 1974a) encoded in the select set of the amino acid residues constituting the active site of an enzyme which may be similar to the *evolutionarily coevolving amino acid residues* of enzymes (Poole and Ranganathan 2006), and Socolich et al. (2005). The former is necessary to account for the deterministic component of Eq. 11.27, and the latter is necessary to account for the nondeterministic, stochastic, and arbitrary component, X(w). Together, these two factors – *conformational energy* and *genetic information* – completely account for the waiting time distribution of COx measured by Lu et al. (1998).

In conclusion, the *conformon theory of molecular machines* first invoked to account for the molecular mechanisms underlying the phenomenon of oxidative phosphorylation in mitochondria (Green and Ji 1972a, b; Ji 1974a, b, 2000) appears to provide a most comprehensive and mechanistically and evolutionarily realistic explanation to date for the single-molecule enzymological data on cholesterol oxidase measured by Lu et al. (1998).

11.4 The Conformon Model of Molecular Machines

The living cell is a system (or "renormalizable" network; see Sect. 2.4) of molecular machines that self-organizes to carry out its varied functions using the free energy supplied by exergonic chemical reactions or light absorption. Molecular machines playing essential roles in the cell include ordinary enzymes, molecular motors (e.g., myosin, kinesins, dyneins), ion pumps and channels, signal transducing proteins, DNA polymerase, RNA polymerase, and chaperones. Like macroscopic machines

such as cars, molecular machines must be driven by free energy derived from chemical reactions, but the mechanism by which these processes manage to drive molecular machines is not yet fully understood. One possibility is suggested by the *conformon theory*, according to which all molecular machines are driven by chemical reaction–derived or ligand binding/de-binding-induced *mechanical strains* stored in sequence-specific sites in biopolymers known as the *conformons* (see Sects. 8.2 and 11.4.1 for the mechanisms of conformon generation).

11.4.1 The Conformon Model of "Biomotrons"

There are several related terms used in the fields of molecular biology and the emerging field of single-molecule enzymology (Xie 2001; Deniz et al. 2008) such as "molecular energy machines" (McClare 1971, 1974; Welch and Kell 1986), "molecular machines," "molecular motors" (Astumian 2000, 2001), "molecular rotors," "molecular switches," "Brownian ratchets," "molecular catalysts," and "protein machines" (Kurzynski 2006). We can regard all these terms as representing different *species* (or *tokens*) of the same *class* (or *type*) of objects which may conveniently be referred to as "biomotrons," a term coined by one of the pioneers of the single-molecule mechanics, T. Yanagida (http://www.wtec.org/loyola/word/erato/pendixbf.doc) (Douglas 1995). *Biomotrons* stands for "*bio*logical *mo*lecular mo*tor*s." The main purpose of this section is to present the following two assertions:

1. All *biomotrons* are driven by *conformons*.
2. Conformons are generated in *biomotrons* from exergonic chemical reactions or exergonic ligand-binding and de-binding processes, based on the generalized Franck-Condon mechanisms (Sects. 2.2.3 and 8.2).

Conformons are the mechanical energies stored in biopolymers in the form of conformational strains (Chap. 8 and Sect. 11.3.2). Since work or energy is defined as (*energy*) = (*force*)(*displacement*), force is the rate of change of energy with respect to displacement. In other words, *energy* and *force* are intimately related so that one can be used to derive the other, given the numerical value of the displacement and the associated potential energy function. It is for this reason that *conformons can be used to produce molecular forces inside biopolymers*. In addition, conformons can be generated from chemical reactions through the generalized Franck-Condon mechanisms as shown in Fig. 8.1, thus making *conformons* as the realistic molecular mechanisms for transducing *chemical energy* to *mechanical energy* in muscle contraction and other molecular motions or movements (Astumian 2000, 2001).

The conformon mechanism was first applied to muscle contraction in Ji (1974b), Fig. 6 on p. 223, reproduced in Fig. 11.34). The essential content of this mechanism is depicted in Fig. 11.31b in terms of symbols rather than pictures. In Fig. 6 of Ji (1974b) and Fig. 11.34, *conformons* are represented as a stretched spring attached to the myosin head (also called subfragment-1 of myosin, or S-1).

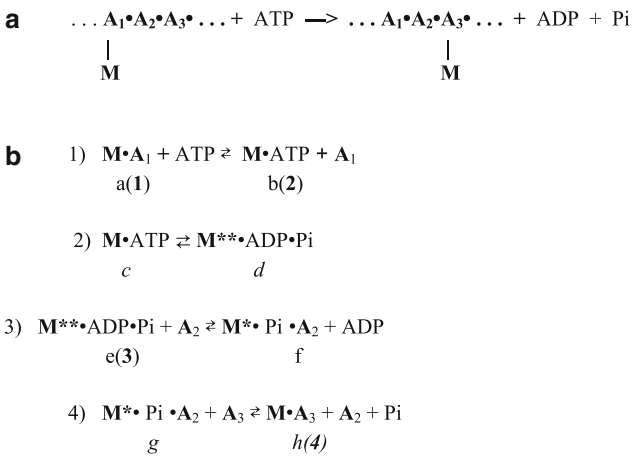


Fig. 11.31 A simplified version of the *conformon mechanism of muscle contraction* first proposed in Ji (1974b) on the basis of the generalized Franck-Condon principle (Sect. 2.2.3). (**a**) A schematic representation of one turnover of the actomyosin system. $\mathbf{A_1}$, $\mathbf{A_2}$, and $\mathbf{A_3}$ are the actin monomers that are linearly arranged, the center-to-center distance between adjacent actin monomers being about 5 nm. **M** stands for the myosin head (also called the S-1 subfragment of myosin) bound to an actin (**A**) through non-covalent bond indicated by a vertical bar, |. The figure indicates that the hydrolysis of one molecule of ATP to ADP and Pi results in the movement of the myosin head from $\mathbf{A_1}$ to $\mathbf{A_3}$ across a distance of about 10 nm or 100 Å. (**b**) The muscle contraction mechanism based on the conformon theory (see Sect. 8.2) consists of four main steps. *The numbers in the parentheses refer to the states of the actomyosin system that have been actually measured (see Fig. 11.33d) using the single-molecule manipulation techniques described in Fig. 11.32.* Step 1 = The actin ($\mathbf{A_1}$) bound to myosin (see State *a*) is displaced by ATP to produce the myosin-ATP complex (see State *b*). Step 2 = The myosin-actin complex fluctuates between two conformational states – the ground state, *c*, in which the potential energy is stored in ATP, and the energized state, *d,* in which the potential energy is stored in mechanical strains or conformons (denoted as two superscripts *, each symbol indicating one conformon). In other words, it is assumed that the hydrolysis of one molecule of the bound ATP to ADP and Pi generates two conformons in this step. Step 3 = One conformon is used to translocate the myosin head from $\mathbf{A_1}$ to $\mathbf{A_2}$, with a concomitant release of ADP into the medium. Step 4 = The second conformon is used to translocate the myosin head from A_2 to A_3, releasing Pi into the medium

Each of the steps shown in Fig. 11.31b can be expanded using the *pre-fit* (or generalized Franck-Condon) *hypothesis* discussed in Sect. 7.1.3. For example, Step (B) (1) may implicate the following processes: $\mathbf{M{\bullet}A_1}$ + ATP ⇄ $\mathbf{M'}$ + $\mathbf{A_1}$ + ATP ⇄ $\mathbf{M'{\bullet}}$ATP + $\mathbf{A_1}$, where $\mathbf{M'}$ represents a high-energy (or thermally excited) conformation whose shape is complementary to that of ATP. To account for some of the unexpected observations reported by Ishijima et al. (1998), e.g., the negative time values in the histogram of the temporal relation between actin displacement and the release of nucleotides from myonsin (see Fig. 7 in the above reference), it may be necessary to invoke a role of the pre-fit mechanisms in the operation of the actomyosin motor. More specifically, $\mathbf{M'}$ may have a long enough lifetime (without violating the Second Law of thermodynamics) to begin to exert force on the actin filament before receiving free energy input from ATP hydrolysis, thus accounting for the negative time values in Fig. 7 of Ishijima et al. (1998).

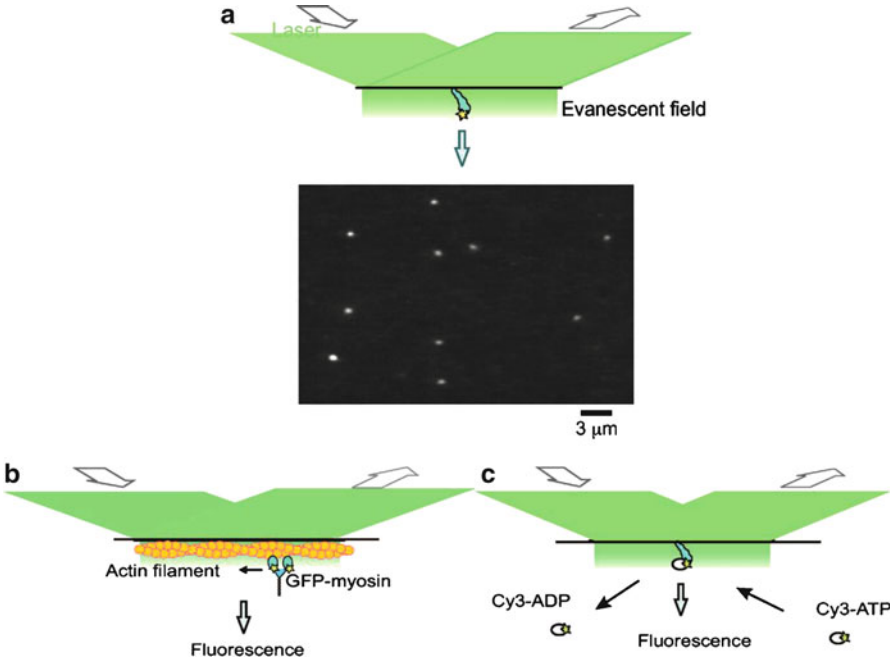


Fig. 11.32 Single-molecule fluorescence measurements using the total internal reflection fluorescence (TIRF) microscopy (Axelrod 1989). (**a**) The basic system of illuminating single fluorescent molecules on the glass surface (see the *black horizontal line*) using the *total internal reflection fluorescence (TIRF) microscopy*. Typical images of single molecules measured with TIRF microscopy are shown (see the *white dots* in the *dark background*). (**b**) Visualizing the sliding movement of a GFP (*Green* Fluorescent Protein)-tagged myosin molecule along an actin filament immobilized on the glass surface. (**c**) Visualizing the turnover of ATP hydrolysis by a single myosin molecule attached to a glass surface. Cy3-ATP and Cy3-ADP are labeled with a fluorophore (i.e., a portion of a molecule that absorbs light at one wavelength and emits a part of it as a longer-wavelength fluorescence; see Fig. 11.16). Cy3-ATP binds to the myosin head and stays bound until myosin head cleaves Cy3-ATP to produce Cy3-ADP and inorganic phosphate, Pi. Cy3-ADP then de-binds from the myosin head due to its low binding affinity and the fluorescent spot disappears. Thus, by using the methods employed in (**b**) and (**c**), it is possible to measure both single myosin molecular motions and the kinetics of ATP hydrolysis accompanying the myosin motion (see Fig. 11.33) (Reproduced from Ihsii and Yanagida 2007))

The conformon mechanism of muscle contraction proposed in Fig. 11.31 is supported by the data obtained from the single-molecule measurements of myosin moving along the actin filament in the presence of ATP. Such a single-molecule experiment was made possible because of the development of the optical (or laser) tweezers and the total internal reflection fluorescence (TIRF) microscopy (see Figs. 11.31 and 11.32). A 'laser tweezer', or 'optical tweezer' is a laser beam focused down to a diffraction-limited spot of about 1 μm in diameter. The laser beam provides an electric field with a gradient in every direction such that there is one point of maximum field strength. Due to the polarizing effect of the focused field,

any dielectric object feels a force proportional to the magnitude of the gradient that pulls the object into the region of maximum field strength. The laser beam therefore can be used to apply a force to any dielectric particle to manipulate its position (Ishii and Yanagida 2007). Employing optical tweezers, biophysicists during the past decade were able to measure simultaneously both the translational motion of the myosin head (which has the ATPase activity) along actin filament and the hydrolysis of ATP that powers the myosin movement (Ishijima et al 1998; Ishii and Yanagida 2007). A typical example of such experiments is shown in Fig. 11.33.

The single-molecule measurements revealed three distinct states for the ATP-actomyosin system as schematized in Fig. 11.33d. In State (1), actin is displaced from its equilibrium position as indicated by the upper trace in Fig. 11.33d due to myosin, free of ATP (as indicated by the low fluorescence level in the lower panel of Fig. 11.33d), exerting a force on it. In State (2), myosin rapidly dissociates from actin which causes a rapid return of actin to its equilibrium position and ATP begins to bind slowly to the dissociated myosin (which is not too far away from actin) causing an increase in fluorescence level. In the transient State (3), ATP is rapidly hydrolyzed by myosin and ADP leaves (as indicated by the precipitous drop in ATP fluorescence intensity), making the ATP-free myosin exert force on actin. In State (4), ATP-free myosin keeps actin displaced from its equilibrium position as in State (1).

In the lower portion of Fig. 11.33d, Ishii and Yanagida (2007) propose a four-state (labeled 1–4) mechanism of the myosin movement along the actin filament driven by ATP hydrolysis. In contrast, the *conformon-based mechanism* of muscle contraction shown in Fig. 11.31 contains eight steps (labeled *a*–*h*), some of which correspond to the four states invoked by Ishii and Yanagida (see the numbers in parentheses in Fig. 11.31). The Ishii-Yanagida and conformon mechanisms are compared in Table 11.14.

As evident in Table 11.14, the conformon-based mechanism of myosin-actin interactions can explain every observation accounted for by the mechanism proposed by Ishii and Yanagida (2000, 2007) and, in addition, provides reasonable explanations (i.e., two conformons generated per ATP hydrolysis event) as to why there are two 5.5-nm steps per turnover of the actomyosin system. The single-molecule mechanical measurements of the actomyosin system presented in Fig. 11.33 seem to support the conformon model of muscle contraction proposed in Ji (1974b). To evaluate the validity of this conclusion, the pictorial version of the conformon-based model of muscle contraction is reproduced in Fig. 11.34.

According to the conformon model, two processes are crucial in muscle contraction: (a) the transduction of the chemical free energy of ATP to conformons stored in myosin. In State a, one molecule of ATP is bound to S-1 and myosin is in its ground state (as symbolized by a relaxed spring). Brownian motions (also called thermal fluctuations) bring S-1 close to the myosin-binding site on the thin filament (see the upper bar with two indentations) that is located on the Z-line side of myosin (see a → b). Upon binding actin, myosin catalyzes the phosphoryl group transfer from the bound ATP to a hypothetical phosphoryl group acceptor X located in S-1 (see b). The exergonic nature of this reaction enables the following two events

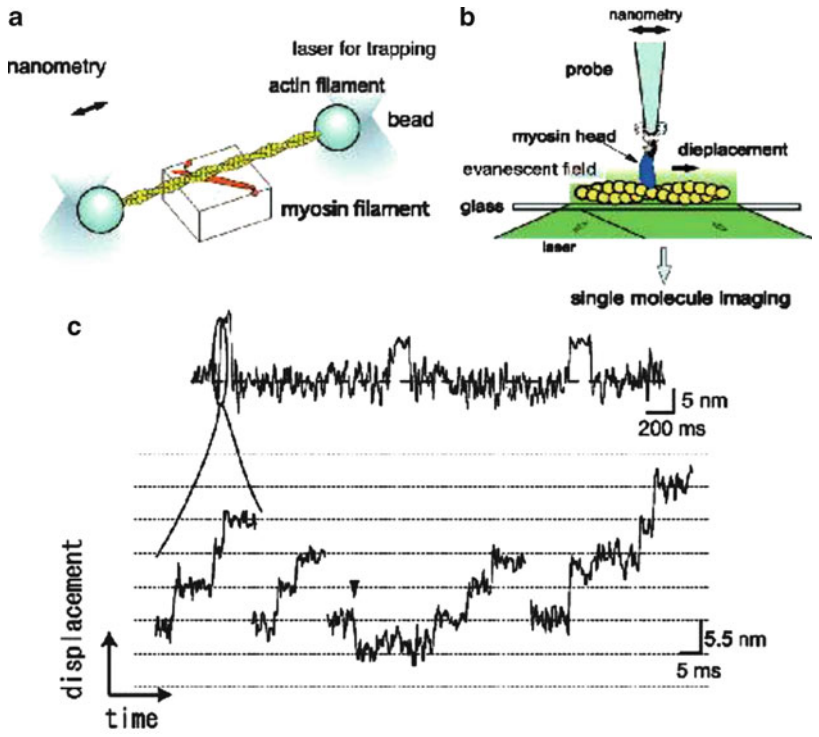


Fig. 11.33 Simultaneous measurements of the mechanical movement of the myosin head and the associated kinetics of ATP hydrolysis (Reproduced from Ishii and Yanagida 2007) with modification: I added the star symbol, *, to myosin in State (3) to indicate the presence of *conformational strains* or *the conformons postulated to be stored* in myosin. (**a**) In the presence of ATP, a molecule of myosin head immobilized on a block exerts a force on an actin filament attached to two beads on its ends. One of these beads is fixed with a laser tweezers and the movement of the other is measured with another optical tweezers. (**b**) An alternative way of measuring the myosin's mechanical activity. The myosin molecule attached to the tip of a micro-needle is brought to an actin filament fixed on a glass surface so that it can touch actin molecules. Given ATP in the medium, the myosin head exerts a force on the actin filament thereby producing an equal and opposite force on the tip of the micro-needle (following the Third Law of Newtonian mechanics), which causes the micro-needle to undergo displacements as recorded in (**c**). (**c**) Expanding the rising phase of the step movement in (**c**) revealed substeps of 5.5 nm, the diameter of the actin monomer. (**d**) The results of a simultaneous measurement of the mechanical movement of the myosin head and the hydrolysis of one molecule of a fluorescent ATP analogue. The upper trace records the movement of myosin along the actin filament, and the lower trace shows the binding activity of ATP to myosin. The high level of fluorescence indicates the binding of a molecule of ATP to the myosin head and the low level of fluorescence signals the desorption of ATP and ADP from myosin. It should be noted here that the superscript * added to the myosin head in State (3) (to denote the conformationally strained and hence mechanically energized state of the myosin head) is a modification proposed in this book for the first time based on the conformon theory of molecular machines (Green and Ji 1972a, b; Ji 1974a, b, 1979, 1991, 2000)

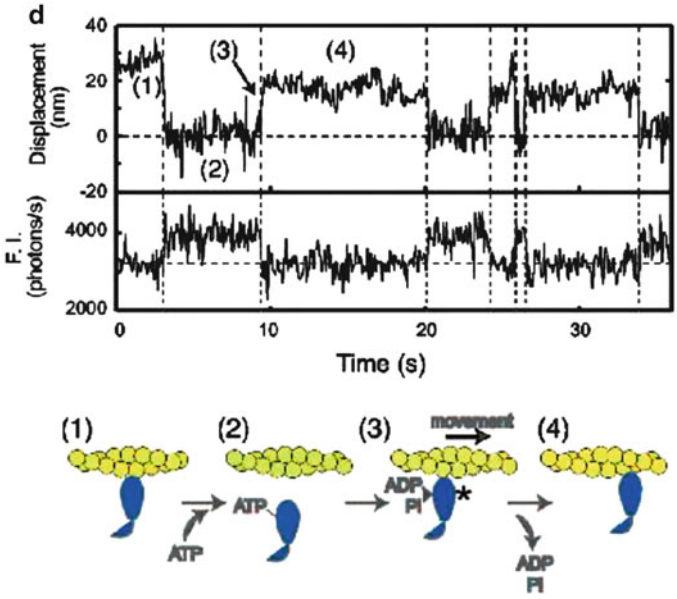


Fig. 11.33 (continued)

to take place: (a) the generation of a charge on myosin which increases the actin-binding affinity and (b) the paying back of the thermal energy borrowed from the environment to extend the S-2 subfragment in going from State a to State b. Actin and myosin are now tightly coupled electrostatically and mechanical energy is stored in myosin (which corresponds to State 3 in Fig. 11.33d). As S-2 relaxes, the thin filament is pushed toward left as indicated by the arrow in State b. When S-2 contracts to a critical distance, through allosteric interactions, the phosphoryl group in the myosin head (i.e., S-1) is thought to be transferred from X to Y (which could well be bound H_2O) (see State c) and the actin-binding affinity is drastically reduced so that myosin becomes detached from the thin filament (see c $\rightarrow$ d), thus completing one machine cycle.

It is known that one ATP split is capable of moving the thin filament by a maximum of about 100 Å or 10 nm (Huxley and Hanson 1960). This finding was the basis for the assumption that one turnover of ATP hydrolysis causes myosin to be displaced by about 100 Å (or 10 nm) in two steps, from States b to c (accompanied by the release of ADP from myosin) and from States c to d (associated with the release of Pi from the same). The sequential releases of ADP and Pi were postulated on the basis of the analogy drawn between the *electron* and the highly unstable phosphoryl group, $PO3^{-}$, which was conveniently termed the *phosphoron* (Ji 1974b). Just as the electron flow from carrier A to carrier B in Fig. 8.1 led to the generation of conformons (see the cocked spring stabilized by two opposite charges at State c), it was thought plausible to generate conformons in myosin by transferring the phosphoron from ATP to a hypothetical *phosphoron* carrier X and then to another *phosphoron* carrier, Y, thereby generating two conformons, each

Table 11.14 A comparison between the two explanations for the single-molecule measurement data on the myosin-actin interactions coupled to ATP hydrolysis (Fig. 11.33)

States in Fig. 11.33d	Mechanism proposed by Ishii and Yanagida (2007)	Mechanism based on the conformon theory (Green and Ji 1972a, b; Ji 1974b, 1991, 2000, 2004a, b; see Sect. 8.2)
1	*Myosin strongly interacts with actin and dissociates upon binding ATP*	The conformon stored in myosin exerts force on actin
2	*Myosin weakly interacts with actin*	Myosin dissociates from actin upon binding ATP
3	*Myosin generates mechanical motion when Pi and ADP dissociates from myosin*	1. ATP is hydrolyzed and a part of the chemical free energy supplied thereby is stored in myosin as mechanical energy or *conformons* (as denoted by the superscript * on the myosin molecule) 2. The mechanical energy stored in myosin is used to exert forces on actin in two steps, each step involving a 5.5-nm displacement (see Fig. 11.33c) 3. Each 5.5-nm displacement of actin is allowed if and only if one of the two products of ATP hydrolysis dissociates from myosin. For example, ADP can dissociate from myosin and the other product, Pi, remains bound to myosin to prevent its slippage from actin filament, thereby preventing the futile cycling (i.e., hydrolyzing ATP without producing any processive movement of actin), in agreement with the mechanism of muscle contraction proposed in Ji (1974b) (see Fig. 11.34)
4	*Myosin interacts with actin strongly*	After both ADP and Pi dissociate from myosin, myosin binds and exerts force on actin as in State (1)

carrying about 8 kcal/mol of free energy. Thus it was assumed that in States b and c in Fig. 11.34, myosin was phosphorylated at the hypothetical carriers X and Y, respectively (Ji 1974b, p. 221). *Thus, the conformon-based mechanism of muscle contraction proposed in Ji (*1974b*) predicted the two, 5 nm-step motion of the myosin along the actin filament which was confirmed by the single-molecule measurements of Ishijima et al. (*1998*) shown in Fig. 11.33c.*

From the point of view of the conformon hypothesis, the most important steps in Fig. 11.33d are States (2) and (3), since these steps represent the first *direct experimental evidence* for the conversion of *chemical energy* to *mechanical energy* stored in myosin, i.e., conformons. In State (2), ATP is bound to myosin and actin is relaxed as indicated by the two related traces labeled (2) in Fig. 11.33d. In State (3), ATP is

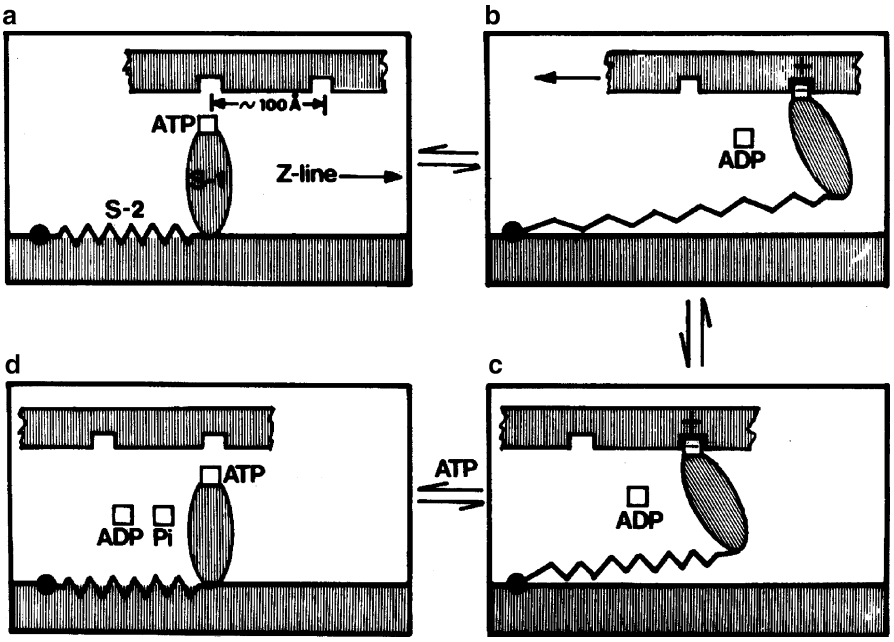


Fig. 11.34 A pictorial representation of the conformon model of muscle contraction discussed in Fig. 11.31 (Reproduced from Ji 1974b). The S-1 subfragment of myosin is depicted as an ellipsoid and the S-2 subfragment as a mechanical spring, but both are thought to be involved in storing mechanical energy, i.e., *conformons*. The *upper bar* represents the thin filament composed of a linear aggregate of actin monomers and the *lower bar* is the thick filament, an intertwined chain of myosin subfragments S-2 with subfragments S-1 protruding from the body of the thick filament. (**a**) ATP bound to the myosin head which is detached from the actin filament. (**b**) ATP is split into ADP and Pi and the Pi-bund myosin head attaches to the actin filament. (**c**) The S-2 subfragment contracts and pushes the actin filament to the left. (**d**) The myosin head releases Pi and detaches from the actin filament to restart the cycle. In some publications, the order of the release of ADP and Pi from the myosin head is the reverse of what is given in this figure, which will not substantially alter the basic mechanisms underlying the conformon-mediated coupling between chemical reactions and mechanical processes

hydrolyzed and the resulting ADP dissociates from myosin as indicated by the rapid decrease in the fluorescence signal at t = 10 s in Fig. 11.33d which is associated with the $a \rightarrow b$ transition postulated in Fig. 11.34, and the ADP-free myosin then exerts a mechanical force on actin as shown by the relatively slow displacement of actin around t = 10 s in Fig. 11.33d, supporting the c → d transition in Fig. 11.34. *The fact that the fluorescence drop in State (3) is faster than the velocity of the associated displacement of actin is consistent with the proposed mechanism where ATP hydrolysis leading to conformon generation in myosin precedes the displacement of actin by myosin.* Therefore, it may be concluded that *the experimental data presented in Fig. 11.33d provide the first direct experimental evidence to validate the following mechanism of free energy transduction in molecular biology that was proposed more than three decades ago (Ji* 1974b*):*

ATP Hydrolysis ⟶ Conformon Generation ⟶ Force Generation

Fig. 11.35 The conformon mechanism of chemical-to-mechanical energy transduction in biology proposed in Ji (1974b) which was supported by the single-molecule mechanical measurements made by Ishijima et al. (1998; Ishii and Yanagida 2007)

Figure 11.35 supported by the empirical data in Fig. 11.34 thus substantiates the following generalizations:

1. Motor enzymes are molecular machines that carry out chemical-to-mechanical (i.e., "chemo-mechanical") energy conversion/transduction at the microscopic level at physiological temperatures, similar to combustion engines at the macroscopic level at high temperature.
2. Motor enzymes convert chemical energy first to the intermediate form of conformational energy known as *conformons* before they are converted to mechanical forces acting on their environment to do work.

If this analysis turns out to be correct, it could be concluded that the conformon concept embodies the most fundamental characteristics of enzymes (both simple enzymes and molecular motors) at the microscopic level providing a theoretical foundation for single-molecule mechanics (Xie 2001; Ishii and Yanagida 2007; Deniz et al. 2008).

11.4.2 Static/Stationary *versus* Dynamic/Mobile *Conformons*

In Sect. 11.3, the single-molecule enzymological data of Lu et al. (1998), namely, the waiting time distribution of cholesterol oxidase (see Fig. 11.24) were analyzed on the basis of the postulated analogy between the *blackbody radiation* and *enzymic catalysis*. The net result of this analysis was the conclusion that the ground-state cholesterol oxidase (and presumably those of enzymes in general) contains a set of *conformons* with different free energy contents (See Fig. 11.28). In Sect. 11.4.1, the single-molecule mechanics data on the actomyosin system measured by Ishijima et al. (1998) were found to be consistent with the predictions made by the *conformon* model of muscle contraction proposed in Ji (1974b). In both analyses, the concept of the *conformon* played a key role, but with some difference. In cholesterol oxidase, the conformons are postulated to be present even before the catalysis takes place, most likely due to the *endogenous conformational strains* introduced into the protein as it is synthesized on the ribosome one amino acid at a time and folded into a globule before the last amino acid is added (see Klonowski and Klonowska 1982 for a related discussion). In contrast, the conformons postulated to be generated within the myosin molecule during muscle contraction are the result of ATP hydrolysis and are not thought to remain stored in myosin long, before they are used up in doing

molecular work on the actin filament. In other words, the conformons present in cholesterol oxidase are present from the birth of a protein while those generated from ATP hydrolysis (and other exergonic chemical reactions such as ligand binding, methylation, oxidation, reduction, etc.) are introduced later in the life cycle of a protein. For the lack of better terms, we will refer to the former as *static* (or *intrinsic*/*endogenous*) and the latter as *dynamic* (or *extrinsic*/*exogenous*) conformons. As pointed out in (6) in Sect. 11.3.3, "static" conformons are closely related to what is known as the Klonowski-Klonowska conformons (Ji 2000) and to "frustrations" of Anderson (Ji 2000).

In conclusion, the single-molecule enzymological and mechanical data measured by two independent groups in 1998 have been rationally accounted for in terms of the concept of the *conformon* introduced into molecular biology in 1972 (Green and Ji 1972a, b; Ji 1974b, 2000). If the explanations proposed in Sects. 11.3 and 11.4 turn out to be correct upon further investigation, we will be able to conclude that *it took a quarter of a century for the theoretical concept of the conformon to be experimentally confirmed with reasonable certainty*.

11.4.3 Stochastic Mechanics of Molecular Machines

In Sect. 4.9, the concept of "info-statistical mechanics" was introduced based on the *information-energy complementarity* as applied to statistical mechanics. In the present section, a related term, "stochastic mechanics," is introduced, motivated by recent emergence of single-molecule enzymology and mechanics (Xie and Lu 1999; Xie 2001; Ishii and Yanagida 2007; Deniz et al. 2008). Both these terms are closely related as forest and trees or as global and local views. In other words, we may regard "info-statistical mechanics" and "stochastic mechanics" as global (or forest) and local (or trees) views, respectively, of the same phenomenon of life on the microscopic level.

We can divide machines into *deterministic* and *stochastic* machines. *Deterministic machines* (e.g., computers, cars, washing machines) are macroscopic in size, robust against thermal fluctuations (i.e., machine component configurations are not destroyed/rearranged by Brownian motions) and obey deterministic rules. *Stochastic machines* in contrast are microscopic in size (e.g., enzymes, molecular motors, cells), depend critically on Brownian motions for their functions, and exhibit stochastic behaviors that can be represented as time-dependent functions of some random variables (e.g., enzymic activity).

As already mentioned, during the past decade a new field in molecular biology has emerged variously referred to as "single-molecule mechanics" (SMM) or "single-molecule enzymology" (SME) (see Sects. 11.3 and 11.4) as a result of the development of new experimental techniques such as optical tweezers and Foerster (or fluorescence) resonance energy transfer (FRET) methods that have enabled biophysicists to visualize and manipulate single biopolymer molecules (proteins, DNA, and RNA) and measure their motions in real time, involving forces and displacements in the ranges of piconewtons and nanometers, respectively (Ishii and

Yanagida 2000, 2007; Deniz et al. 2008). SME differs from the conventional ensemble-averaged enzymology (EAE) in one important respect: Whereas EAE studies time-dependent *concentrations* of materials involved in chemical reactions, SME investigates time-dependent *probabilities* of single-molecule events such as the activation or deactivation of an enzyme (Fig. 11.17), stochastic movements of molecular motors along DNA, actin filament (Fig. 11.33), or microtubules. It is here suggested that SMM and SME obey the same set of the principles and rules (described below) on which what is here called the "single-molecule stochastic mechanics" (SMSM) is grounded and hence are synonymous with it: i.e., SMM = SME = SMSM.

There appears to be only a small number of the physical principles and rules that underlie SMSM:

1. All molecular machines are driven by mechanical energy packets known as *conformons* that are stored in biopolymers as sequence-specific and mobile conformational deformations (Green and Ji 1972a, b; Ji 1974b, 1991, 2000, 2004a).
2. Conformons can be generated from chemical reactions in three steps (see Fig. 8.1 in Sect. 8.2 and Fig. 11.30):
   (a) Enzymes borrow thermal energies from their environment to produce *virtual conformons* (as a result of *thermal fluctuations*) lasting for less than the time, τ, required for enzymes to complete their machine cycles.
   (b) Virtual conformons mediate the catalysis of an exergonic (i.e., free-energy-supplying) physicochemical processes within their lifetimes.
   (c) Enzymes avoid violating the Second Law by "paying back" the thermal energy borrowed in Step (a) by letting the free energy released from the physicochemical processes equilibrate with its environment within times less than τ and also by synchronously stabilizing the virtual conformons with one or more products generated from the exergonic chemical reactions (Ji 1979).
3. All molecular machines perform work on their environment (e.g., actin filament in the case of myosin, ions in the case of ion pumps) through energy transfer during the coupled phase of the machine-environment interaction cycle, preventing slippage (Ji 1974b).

Items (2)(a) and (2)(b) may appear to violate the traditional formulation of the Second Law of thermodynamics, according to which no *thermal energy can be utilized to do work in the absence of temperature gradients*, but this is not the case, because these mechanisms obey the "molecularized" Second Law of thermodynamics (MSLT) formulated by McClare (1971) (see Statement 2.5 in Sect. 2.1.4).

Rule (2)(b) appears reasonable in view of the fact that *virtual conformons* can last for times much longer than the time required for electronic transitions (or covalent bond rearrangements) and hence the generalized Franck-Condon principle (GFCP) can be applied to them (Sect. 2.2.3). Single-molecule measurements indicate that conformational changes attending enzymic actions are slower than the

electronic rearrangements entailed by chemical reactions by a factor of 10^3–10^6 (see data in Ishii and Yanagida 2000, 2007; Xie and Lu 1999; Xie 2001). Therefore, Rules (1), (2), and (3) presented above appear to provide a sound theoretical framework for grounding the newly emerging *single-molecule stochastic mechanics* (SMSM).

11.4.4 Biopolymers as Molecular Machines*: Three Classes of Molecular Machines and Three Classes of Their Mechanisms of Action*

All biopolymers (proteins, RNA, and DNA) have two properties in common: (a) *sequence information* and (b) *sequence-preserving mechanical deformability* (also called "conformational changes"), which enables biopolymers to store mechanical energy as conformational strains. These properties were first clearly recognized in enzymes and were postulated to play essential roles in the molecular mechanisms underlying enzymic catalysis (Lumry 1974; Ji 1974a, b, 1991, 2000, 2012) as embodied in the concept of the *conformon*, the conformational strains storing mechanical energy and genetic information to drive all goal-directed molecular motions in living cells (see Chap. 8). The main objective of this section is to propose that

> The principles and mechanisms of molecular machines discovered in proteins are universally applicable to all biopolymers, including RNA and DNA to varying degrees. (11.54)

We may refer to Statement 11.54 as the *Principle of the Universality of Molecular Machines* (PUMM) and represent its content in a tabular form as shown in Table 11.15.

Molecular machines can be divided into *passive* and *active* machines depending on whether their outputs approach to or move away from equilibria, respectively. Examples of the former include voltage- and ligand-gated ion channels and those of the latter include active transporters such as the Na^+/K^+ ATPase, H^+ pumps and Ca^{++} pumps. The active machines in turn divide into two classes – primary active machines and secondary active machines – depending on whether the form of the free energy driving the machine is chemical (e.g., Na^+/K^+ ATPase) or nonchemical, i.e., osmotic (e.g., Na^+/Ca^{++} antiporter) or mechanical (e.g., DNA supercoil-driven regulation of gene expression; see Sect. 8.3), respectively. Thus these dual dichotomous divisions lead to three distinct types of molecular machines (a) primary active (or Type I) machines, (b) secondary active (or Type II) machines, and (c) passive (Type III) machines and each type is divided into three classes based on molecular types, i.e., proteins, RNAs and DNAs, on the one hand, and based on whether the processes involved are Victoria or scalar, thus resulting in a total of 18 classes of molecular machines as summarized in Table 11.15. The 9×2 matrix constituting the content of Table 11.15 may be viewed as the *periodic table of molecular machines* (PTMM), and, as was the case with its chemical counterpart, many cells or blocks are left vacant

Table 11.15 The Principle of the Universality of Molecular Machines (PUMM): three classes of molecular machines (proteins, RNA and DNA) and three classes of their mechanisms of actions (Types I, II, and III)

Molecular machines (principles and mechanisms)		Vectorial process (v)	Scalar process (s)
Type I *Primary active machines (conformons Ji* 1974a, b, 2000, *2011)*	Protein	Na^+/K^+ *ATPase*	*RNA polymerase*
		H^+ *pump*	*DNA polymerase*
		Ca^{++} *pump*	*Glycogen synthase*
		Actomyosin system	*ATP synthase*[a]
	RNA		*Riboswitch-ribozymes*, e.g., glmS riboswitch (Winkler et al. 2004)
	DNA		
Type II *Secondary active machines (generalized Franck-Condon principle, pre-fit hypothesis Ji* 1974a, b, 1991, *2011)*	Protein	Na^+/Ca^{++} *antiporter*	*Receptor tyrosine kinase*
		Na^+*/amino acid symporter*	*kinases*
	RNA		*Ribozymes*, e.g., RNA component of RNAse-P, and many introns of RNA (Kruger et al. 1982; Wochner et al. 2011; Tang and Breaker 2000)
	DNA		*Deoxyribozymes*
			DNA supercoils
			Self-regulatory genes (Ji et al. 2009c)
Type III *Passive machines (pre-fit hypothesis (Ji* 1974a, b, 1991, *2011)*	Protein	Voltage- and ligand-gated ion channels	*Esterases*
			Phosphatases
			G-protein-coupled receptors
			Cytokine receptors
			Other simple receptors
	RNA		*Riboswitches*, e.g., FMN riboswitch, glycine riboswitch, and TPP riboswitch (Nudler and Mironov 2004; Tucker and Breaker 2005; Vitreschak et al. 2004; Batey 2006; Nahvi et al. 2002)
	DNA		*Noncoding DNA* as regulators of gene expression (Mattick 2003, 2004)
			Deoxyriboswitches (?)

[a]According to the *chemiosmotic* hypothesis (Sect. 11.6), the respiratory ATP synthesis in mitochondria is driven by the proton gradient across the inner mitochondrial membrane. In contrast, the *conformon hypothesis*, the respiration-driven ATP synthesis is not mediated by any proton gradient but by *conformons*, i.e., the conformational strains of enzymes induced by respiration (Ji 1974b, 1976, 1977, 1979, 1991, 2011; Sect. 11.5). In other words, according to the chemiosmotic hypothesis, ATP synthase is a Type I(v) machine whereas, according to the conformon hypothesis, it is a Type I(s) machine

which may be filled in the future as our knowledge on molecular machines advance, especially with respect to the RNA and DNA varieties.

11.5 The Conformon Theory of Oxidative Phosphorylation

The molecular mechanism of oxidative phosphorylation described in Fig. 11.36 is reproduced from (Ji 1974b, 1979). To the best of my knowledge, it is the most comprehensive model of oxidation phosphorylation that accommodates most, if not all, of the key experimental findings on mitochondria, including the phenomenon of

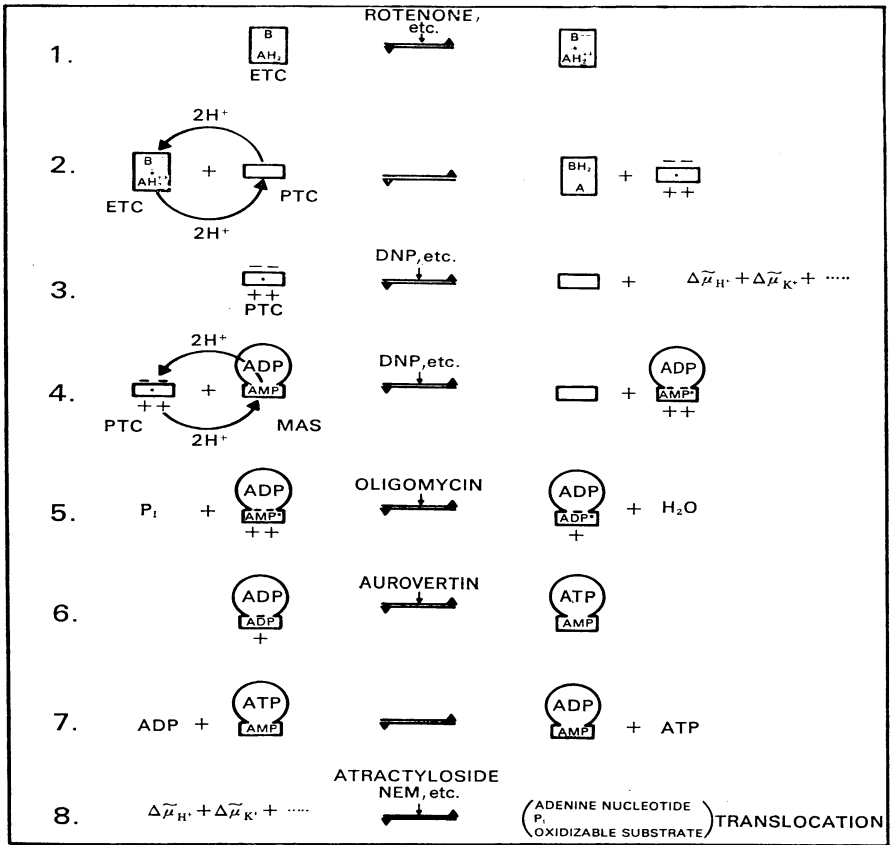


Fig. 11.36 *The Madisonator* – A molecular model of oxidative phosphorylation based on the theory of conformons as energy and information carriers in bioploymers (Reproduced from Ji 1974b, 1979). For detailed explanations for individual steps, see text. The symbol * appearing in Steps 1–5, e.g., AMP*, denotes conformons, i.e., the conformational strains of biopolymers (or their ligands) carrying mechanical strains at sequence- or stereo-specific sites (Chap. 8)

chemiosmosis (see Steps 1–3) (Mitchell 1961, 1968). What distinguishes the *conformon model* and the *chemiosmotic model* of oxidative phosphorylation (see Sect. 11.6) is that the former is rooted in the generalized Franck-Condon principle (Sect. 2.2.3) which has been found to account for not only oxidative phosphorylation but also enzymic catalysis (Sect. 7.2) and muscle contraction (Fig. 11.34). It is clear that the chemiosmotic hypothesis is inapplicable to both enzymic catalysis and muscle contraction, because these two processes can occur without any osmotic barrier, the essential requirement of chemiosmosis – the conversion of chemical energy to osmotic energy. In other words, the chemiosmotic hypothesis has little to say concerning the fundamental molecular mechanisms underlying enzymic catalysis and muscle contraction.

In Step 1, the electron transfer complex (ETC) catalyzes the separation of electrons and protons, storing a part of the free energy released from the redox reaction as conformons (denoted by the symbol *). In Step 2, the energized ETC collides with a hypothetical intra-membrane protein acting as a proton pump (and hence called proton transfer complex, PTC), and two protons are postulated to be donated to the matrix (or the lower) side of PTC and two protons are thought to be abstracted from the cytosolic (or the upper) side, resulting in a depolarized and de-energized ETC and a polarized and energized PTC. In Step 3, the polarized PTC utilizes its conformons to actively pump protons out from the matrix space to the cytosolic space to create a pH gradient and a membrane potential (as in the chemiosmotic hypothesis). Alternatively, the energized PTC can transfer its conformons (via asymmetric protonation-deprotonation reactions as in Step 2) to the mitochondrial ATP synthase (MAS), leading to the de-energization of PTC and energization of MAS as shown in Step 4. In Step 5, a part of the conformons is used to phosphorylate the AMP bound to the basepiece of MAS, and a second conformon is postulated to be used to transfer the phosphoryl group from the ADP bound to the basepiece of MAS to the ADP bound to the F_1 subunit (denoted as a circle) of MAS, thus generating one ATP bound to MAS (see Step 6). In Step 7, this ATP is exchanged for the ADP in the matrix space of mitochondria. Finally, in Step 8, the ATP in the matrix compartment of mitochondria is actively transported out into the cytosol, driven by the proton electrochemical gradient and membrane potential generated in Step 3. All the steps included in Fig. 11.36 are supported by experimental data on the actions of the inhibitors and uncouplers specific for them, except Steps 2 and 7 whose inhibitors and uncouplers appear not to have been discovered yet to the best of my knowledge.

11.6 Deconstructing the Chemiosmotic Hypothesis

The British biochemist, P. Mitchell (1920–1992), proposed the concept of *chemiosmosis* in 1960 (Mitchell 1961, 1968) in an attempt to explain how mitochondria, *the powerhouse of the cell,* synthesize ATP from ADP and inorganic phosphate, P_i, utilizing the free energy supplied by the oxidation of substrates such as

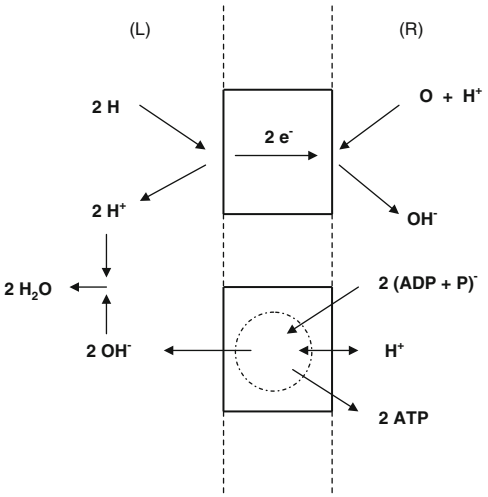


Fig. 11.37 The chemiosmotic hypothesis proposed by P. Mitchell in 1961 (adopted from Mitchell 1961)

NADH and succinate during respiration. His basic idea is that the *chemical* energy of, say, NADH is first converted into the *osmotic* energy (hence the adjective "chemiosmotic") of the proton gradient across the mitochondrial inner membrane (inside high pH and outside low pH) and associated membrane potential (inside negative and outside positive) which subsequently drives the synthesis of ATP (Scheme (11.55)):

$$(\textbf{Chemical Energy of NADH}) \xrightarrow{1} (\textbf{Proton Gradient}) \xrightarrow{2} (\textbf{ChemicalEnergy of ATP}) \tag{11.55}$$

where *Process 1* indicates the translocation of protons across the mitochondrial inner membrane driven by respiration (see the upper box in Fig. 11.37 below), and *Process 2* indicates the proton gradient-driven phosphorylation of ADP to ATP (see the lower box in Fig. 11.37). The key postulates of the Mitchell hypothesis are as follows:

1. The membrane-embedded respiratory enzymes (symbolized by the upper box in Fig. 11.37) *somehow* separate the electron (indicated by e^- the encircled negative charge) and the proton (H^+) from the hydrogen atom (H) and move the former across the membrane (from the left side, L, to the right side, R), leading to the generation of a transmembrane proton gradient and a membrane potential (not shown) and attendant acidification of the L compartment and alkalinization of the R compartment.
2. The osmotic energy stored in the proton gradient (also called the electrochemical gradient of protons or the "proton-motive force", PMF) then drives the abstraction of the hydroxyl ion (OH^-) from the L compartment and the proton from the R compartment to effectuate the synthesis of ATP from ADP and P_i at the reaction center embedded inside the M phase (see the lower box in

Fig. 11.37). But the realistic molecular mechanisms accomplishing such complex molecular processes were left undefined.

3. The respiratory enzyme system catalyzing Process 1 and the reversible ATPase system catalyzing Process 2 are coupled through the mediation of the common chemical species, i.e., *protons*, which are generated from respiration (upper box) and consumed by the ATP synthase (lower box) through as yet unknown mechanisms.

I have long expressed my doubt about the validity of the chemiosmotic hypothesis (despite the seemingly universal acceptance of the chemiosmotic theory by biochemistry textbook writers around the world), because it does not provide any realistic theoretical insights into the possible molecular mechanisms underlying oxidative phosphorylation (Ji 1974b, 1979, 1991). Even if the chemiosmotic hypothesis proves to be theoretically correct, it cannot represent a *universal* principle of *biological energy coupling* because there are membrane-independent (and hence *non-osmotic*) energy-coupled processes in biology, including *muscle contraction*, *molecular motors* moving cargoes along cytoskeletal tracks in the cytosol, and the *tracking of RNA polymerase* along DNA during transcription. In Ji (1991, pp. 60–61), I wrote:

> . . . The chemiosmotic hypothesis of oxidative phosphorylation proposed by P. Mitchell in 1961 (Nicholls 1982; Skulachev and Hinkle 1981) postulates that ATP synthesis is driven by the electrochemical gradient of protons across the inner mitochondrial membrane which is set up by respiration; in other words, in Mitchell's model the generation of the transmembrane electrochemical gradient of protons driven by respiration precedes the phosphorylation reaction. However, according to the Madisonator [i.e., Fig. 11.36], the respiration-driven generation of the transmembrane proton gradient (step 3 in Fig. 1.3) [which is Fig. 11.36 here] and the phosphorylation reaction (steps 5 and 6) are parallel events that are driven by a common free energy precursor generated from respiration, namely conformons (see steps 1 and 2). The primary biological role of the active transport in mitochondria is thought to be not the synthesis of ATP as P. Mitchell assumes but most likely the *communication between mitochondria and metabolic events* going on in the cytosol. If this interpretation turns out to be correct, then the phenomenon of the proton gradient-driven ATP synthesis, well-known in the literature, may have no general biological significance in mitochondria, except perhaps that such a process may contribute to the survival of cells under anoxic conditions when the cytosolic pH drops due to the accumulation of lactate produced by anaerobic glycolysis and that the resulting transmembrane proton gradient may drive ATP synthesis according to the chemiosmotic mechanism of P. Mitchell. (emphasis added)

R. J. P. Williams (1969) is another critic of the chemiosmotic coupling concept. His criticism, aired from the very beginning of the chemiosmotic conception, is based on the consideration of thermodynamic efficiency, which I find persuasive as discussed in (2) below. The following are my specific criticisms against the chemiosmotic hypothesis:

1. Mitchell's proposed mechanism for effectuating respiration-driven proton translocation across the mitochondrial membranes is based on what he calls *vectorial metabolism* or *anisotropy of membrane protein organization* (Mitchell 1961). This idea seems insufficient to account for oxidative phosphorylation, because structural organization alone, no matter how asymmetric, cannot cause an asymmetric transport of the products of chemical reactions without dissipating requisite free

energy. Without any *enzymologically realistic mechanisms* for coupling a down-hill reaction (e.g., oxidation of NADH to NAD^+) to an up-hill chemical reaction or physical process (e.g., the vectorial movement of protons and asymmetric removal of water or the hydroxyl group from the ATP synthesis center), structural asymmetry alone cannot achieve or cause asymmetric metabolism. In fact to cause a symmetry breaking in molecular processes without dissipating requisite free energy is tantamount to violating the First Law of thermodynamics, because the resulting gradients could be harnessed to do work, thereby producing energy.

2. According to the chemiosmotic coupling scheme, the protons generated from respiration are extruded from the membrane phase (with volume V_M) to the bulk phase (with volume V_B). As Williams (1969) correctly pointed out, this is an energy-dissipating process because V_B is much greater than V_M. (The situation is analogous to perfume molecules diffusing out of a perfume bottle into the vast space of a dancing hall.) In addition, at least the same amount of free energy must be dissipated during the proton-driven ATP synthetic step, because in order to bring the protons back into the ATP synthesis center in the M phase (Fig. 11.37), free energy must be dissipated in the amount proportional to the volume ratio, V_B/V_M, which would probably be greater than 10^9. This ratio can be estimated by measuring the areas occupied by the mitochondrial inner membrane relative to the cytoplasmic area in an electron micrograph of the cell and raising the resulting ratio to the power of 3/2.
3. As indicated in (1), structural asymmetry is necessary but not sufficient for effectuating asymmetric process or metabolism. In addition, it is critical for the chemiosmotic hypothesis that the ATP synthesizing reaction center (see the dotted circle in the lower box in Fig. 11.37) be located *within* the M phase, not in the R phase, since, in the latter case, the ATP synthesis cannot be driven by any osmotic energy of the transmembrane proton gradient. However, the recent X-ray crystallographic findings (Aksimentiev et al. 2004; Junge et al. 1997) clearly demonstrate that the ATP synthase is located not in the M phase as Mitchell assumed but outside the membrane phase attached to the proton pumping structure (i.e., F_0) in the M phase through a set of long polypeptide chains (designated as γ and ε subunits).
4. It is generally accepted that, when ATP synthase catalyzes the formation of ATP from ADP and P_i driven by the proton gradient, the electrochemical energy of proton gradient is first converted into the mechanical energy (in the form of "torque," i.e., the energy producing a rotatory motion) within F_0, which is then transmitted to F_1 (through the rotatory motion of the shaft composed of the γ and ε subunits) where the energy is utilized to release (or de-bind) ATP from F_1 (Aksimentiev et al. 2004). Thus, the sequence of events involved in the proton gradient–driven synthesis of ATP can be depicted as follows (Scheme (11.56)):

$$\begin{array}{c}(\textbf{Proton Gradient}) \overset{1}{\leftrightarrow} (\textbf{Mechanical Energy of the } \boldsymbol{\gamma} \textbf{ and } \boldsymbol{\varepsilon} \textbf{ Subunits}) \\ \updownarrow 2 \\ (\textbf{Chemical Energy of ATP})\end{array} \tag{11.56}$$

Scheme (11.56) is known to be reversible so that protons can be pumped across the mitochondrial inner membrane (producing osmotic energy) driven by the chemical energy of ATP hydrolysis. On the phenomenological level, therefore, the concept of chemiosmotic coupling proposed by Mitchell may appear validated since *chemical* and *osmotic* energies are indeed interconverted. But this way of looking at the problem is superficial. The heart of the problem concerns not so much *whether or not* the process of chemiosmosis occurs in mitochondria (which was known to take place in living systems long before 1961 when the chemiosmotic hypothesis was formulated) but exactly *how* such a process occurs on the molecular level. In other words, we must distinguish between the *phenomenon* of chemiosmosis and the *molecular mechanisms* underlying the phenomenon. *On the phenomenological level, the Mitchell hypothesis cannot be faulted. But it is on the level of molecular mechanisms of chemiosmosis that the Mitchell hypothesis fails as I have been pointing out over the past three decades* (Ji 1979, pp. 34–35; Ji 1991, pp. 60–61; Williams 1969).

Any mechanical (i.e., conformational) energy stored in biopolymers can be viewed as examples of *conformons*. Therefore we can rewrite Scheme (11.56) as follows:

$$\begin{array}{c}(\textbf{Proton Gradient}) \overset{1}{--}> (\textbf{Conformons Stored in the } \boldsymbol{\gamma} \textbf{ and } \boldsymbol{\varepsilon} \textbf{ Subunits}) \\ \downarrow 2 \\ (\textbf{Conformons Stored in } \mathbf{F_1} \leftrightarrow \textbf{Chemical Energy of ATP})\end{array} \tag{11.57}$$

Process 1 above is the step where conformons are generated from proton gradients, most likely by reversing the molecular steps postulated for the conformon-driven active transport described in Fig. 2 in Ji (1979) or Fig. 8.1 in this book. Process 2 involves conformon transfer from F_0 to F_1 through the γ and ε subunits, which probably occurs through the mechanism of conformon transfer proposed in Fig. 4 in Ji (1974b).

All the problems encountered by the chemiosmotic hypothesis as indicated above can be resolved simply by invoking the concept of *conformons* which can drive either active transport or ATP synthesis as indicated in Fig. 11.36, depending on the metabolic needs of the cell. The conformon theory of molecular machines accounts for not only membrane-dependent oxidative phosphorylation and active transport but also membrane-independent processes such as muscle contraction and DNA transcription and replication by RNA and DNA polymerases, respectively, DNA supercoiling, and cytoplasmic molecular motor movements – all through the common agency of the energy and information carried by conformons (Ji 1974b, 1979, 2000, 2004a). Thus the conformon concept provides a bioenergetic mechanism that can be applied *universally* to all energy-coupled processes in living systems, including chemiosmosis, but the chemiosmotic approach is limited to explaining membrane-dependent energy-coupled processes such as proton gradient–driven ATP synthesis – all on the phenomenological level.

A direct experimental evidence for the production of conformons from ATP hydrolysis in the F1-ATPase was recently reported by Uchihashi et al. (2011; Junge and Müller 2011). Using the high-speed AFM (atomic force microscopy), these authors succeeded in visualizing the propagation of the *conformational waves* of the β subunits in the counterclockwise direction around the isolated F_1 stator ring without the central γε subunits (see Fig. 2a in Uchihashi et al. 2011). These *conformational waves* are identical to what Green and I defined as *conformons* in 1972 (Green and Ji 1972a, b; Ji 1974b, 2000; also see Chap. 8) and thus provide *the best experimental evidence reported do date that verifies the theoretical concept of the conformon*.

11.7 Molecular Machines as Maxwell's "Angels"

In 1867, in a letter to Peter G. Tait, James C. Maxwell (1831–1879) described a microscopic "being" capable of separating fast moving molecules from slow moving ones into two compartments, thereby decreasing the entropy of the system, "without expenditure of work". Such a "being" was named "Maxwell's demon" in 1874 by William Thompson (1824–1907), implying the mediating, rather than the malevolent, connotation of the word. Maxwell describes his intelligent, microscopic "being" as follows:

> . . . if we conceive of a being whose faculties are so sharpened that he can follow every molecule in its course, such a being, whose attributes are as essentially finite as our own, would be able to do what is impossible to us. For we have seen that molecules in a vessel full of air at uniform temperature are moving with velocities by no means uniform, though the mean velocity of any great number of them, arbitrarily selected, is almost exactly uniform. Now let us suppose that such a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole, so as to allow only the swifter molecules to pass from A to B, and only the slower molecules to pass from B to A. He will thus, without expenditure of work, raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics. . . . (http://en.wikipedia.org/wiki/Maxwell's_demon)

To me, Maxwell's "being" violates the second law of thermodynamics basically because he "'opens' and 'closes' this hole . . . without expenditure of work". According to the second law of thermodynamics, no work of any kind can be performed without dissipating energy into heat, i.e., without increasing the entropy of the system which, in this case, consists of both gas molecules and Maxwell's "being". A more rigorous proof of a similar kind was formulated by Léo Szilárd in 1929. Many other proofs are now available to show that Maxwell's "being" violates the second law of thermodynamics (Leff and Rex 1962; Maruyama et al. 2009).

Maxwell's "being" can carry out his assigned task without violating the second law of thermodynamics *if and only if* he dissipates the requisite energy into heat – in other words, if and only if he receives energy from and dissipates heat into his environment while performing his molecular work. That is, Maxwell's "being" can

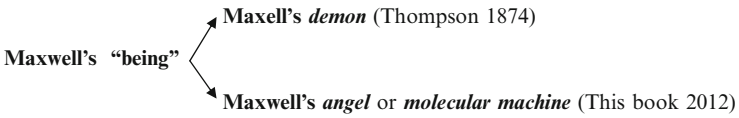


Fig. 11.38 The duality of Maxwell's "being." Maxwell's demon cannot exist because it violates the second law of thermodynamics, but Maxwell's angels exist because they do not violate the second law and act as molecular machines in all living systems

carry out his assigned task without violating the second law if he acts as or is a molecular machine, the concept well established in the molecular biology of the twentieth century (Alberts 1998). Thus, it seems logical to distinguish two kinds of Maxwell's beings – the original one that violates the second law, i.e., Maxwell's demon, and the other that obeys the second law, which I suggest here to be referred to as "Maxwell's angel" (see Fig. 11.38).

Maxwell's demon is of particular interest to us because it acts as the focal point of interaction between information (I) and energy (E), the two fundamental entities underlying the operation of all molecular machines. Therefore, in order to understand the I–E relation, it may be helpful to understand how Maxwell's demon works or does not work. According to Proposition I of the "dualogy" of Sabah Karam (2011), "In order to describe any single event, action, state of being, thought, or any other real or abstract concept it is necessary to identify two distinct complementary co-dependent entities that define the state (conjugate variables) and behavior (conjugate transformations) of the phenomenon under consideration." If Karam is right, the reason for the century-long debates about Maxwell's demon and the associated I-E relation (Leff and Rex 1962) may be traced to the failure on the part of the debaters "to identify two distinct complementary co-dependent entities" that define Maxwell's "being". In this book, I am suggesting that the "two distinct complementary co-dependent entities" that characterize Maxwell's "being" are *Maxwell's demon*, named by William Thompson in 1874, and *Maxwell's angel* that was born on December 1, 2011, during one of my lectures on the information-energy complementarity given to a group of interdisciplinary honors seminar students at Rutgers. Since *Maxwell's demon* and *Maxwell's angel* are conjugate entities, understanding one will inevitably lead to understanding the other, again according to the dualogy of Karam (2011). By identifying Maxwell's angel with the molecular machine of the twentieth-century molecular biology (Alberts 1998), it is claimed here that we will understand Maxwell's angel and its conjugate, Maxwell's demon, and hence the I-E relation if and when we understand how molecular machines work, the Holy Grail of the twenty-first-century molecular biology. Just as Maxwell himself succeeded in integrating *electricity* and *magnetism* into *electromagnetism* in the mid-nineteenth century, Maxwell's angles (or molecular machines) may contribute to integrating *information* and *energy* into *information-energy* (also called *gnergy*; see Sect. 2.3.2) in the twenty-first century. Figure 11.39 represents these ideas diagrammatically. In formulating these diagrams, two ideas emerged unexpectedly: (1) There may exists triadic categories composed of three

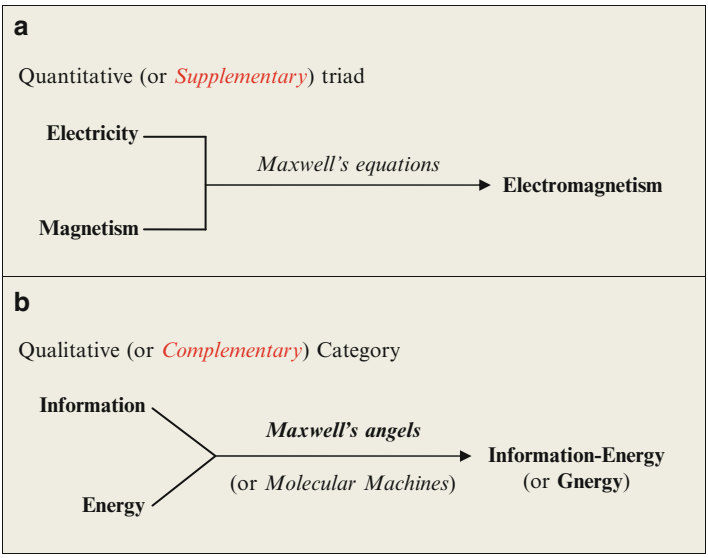


Fig. 11.39 Two kinds of "triadic" categories suggested by *Maxwell's equations* and *molecular machines* (or Maxwell's angels). The concepts of *supplementarity* and *complementarity* are defined in Sect. 2.3.1

nodes rather than the usual two and (2) there may be two kinds of triadic categories – the *quantitative* (or *supplementary*) category where the mapping can be expressed mathematically (e.g., Maxwell's equations), and the *qualitative* (or *complementary*) category where the mapping cannot be expressed in terms of mathematical equations. These two kinds of mappings are depicted with two different symbols in Fig. 11.39 (see Table 2.6 for related symbols).