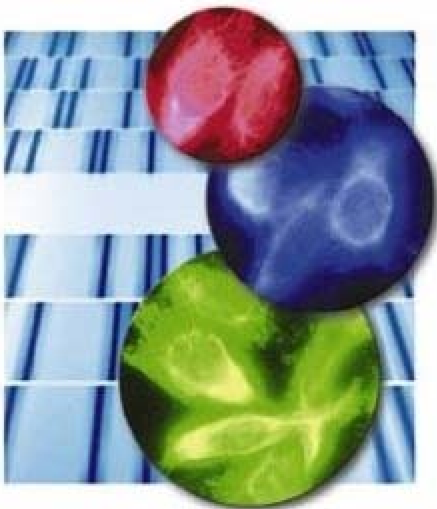
WILEY-VCH

Encyclopedia of Molecular Cell Biology and Molecular Medicine

Edited by Robert A. Meyers



Adipocytes

Todd Leff[1] *and James G. Granneman*[1,2]
[1]*Department of Pathology, Wayne State University School of Medicine, Detroit, MI, USA*
[2]*Department of Psychiatry, Wayne State University School of Medicine, Detroit, MI, USA*

1 Introduction – Adipocytes and the Regulation of Metabolism 3

2 Physiology of the Adipocyte 4
2.1 Absorption and Storage of Circulating Lipid 4
2.2 Mobilization and Release of Stored Lipid into the Circulation 5
2.3 Regulatory Mechanisms that Balance Lipid Storage and Utilization 6
2.4 Adipose Tissue as an Endocrine Organ 6

3 Developmental Origin of Adipocytes 9
3.1 Adipose Tissue Development and Plasticity 9
3.2 Differentiation of Adipocytes from Precursor Cells 11

4 Adipose Dysfunction and Metabolic Disease 13
4.1 Obesity – Too Much Fat 14
4.2 Lipodystrophy – Too Little Fat 15

5 The Adipocyte as a Therapeutic Target for Metabolic Disease 15
5.1 Antidiabetic PPARγ Ligands Act on Adipocytes 16

Bibliography 18
Books and Reviews 18
Primary Literature 19

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Keywords

Metabolism
The chemical processes that occur within a living organism in order to maintain life; the interconnected sequences of mostly enzyme-catalyzed chemical reactions by which a cell, tissue, organ, etc., sustains energy production, and synthesizes and breaks down complex molecules.

Lipids
Any of the large group of fats and fatlike compounds which occur in living organisms and are characteristically soluble in certain organic solvents but only sparingly soluble in water.

Fatty Acids
Long-chain aliphatic compounds that contain a carboxylic acid group.

Gene Expression
The process by which a gene's coded information is converted into the structures present and operating in the cell. Expressed genes include those that are transcribed into mRNA and then translated into protein and those that are transcribed into RNA but not translated into protein (e.g., transfer and ribosomal RNAs).

Transcription Factors
The DNA binding proteins that carry out the organic process whereby the DNA sequence in a a gene is copied into mRNA.

Nuclear Receptors
Transcription factors that are activated by the binding of specific ligans that are usually lipid soluble, such as fatty acids or steroids.

Adipose Tissue
Body tissue which is able to store high amounts of neutral fats.

Adipocytes
The cell type found in adipose tissue that is specialized for storage of neutral lipids.

The adipocyte, or fat cell, is a unique, highly specialized cell whose primary function is to provide a depot for storage of excess fat, derived mainly from the diet, which can be mobilized and distributed to other tissues as dictated by the energetic demands of the organism. Although long thought of as a simple passive storage depot for dietary lipid, it is now clear that adipose tissue plays a very active role in regulating energy balance and not only stores and secretes lipid as needed, but also produces hormonal signals that have multiple effects on energy metabolism in many tissue

and organs throughout the body. This new view of adipose tissue (and the adipocyte) as a key regulator of fuel metabolism and partitioning has dramatically increased the interest in the role that this tissue plays not only in normal physiology but also in the etiology and treatment of important metabolic diseases such as diabetes and obesity.

1 Introduction – Adipocytes and the Regulation of Metabolism

Although the physiology of adipose tissue and the molecular and cellular biology of the adipocyte have been subjects of research interest for many years, new work in this area has transformed our understanding of the adipocyte and its role in regulating human physiology. We now recognize that in addition to its long appreciated function as a lipid storage depot, the adipocyte plays a much more active role in regulating whole body fuel partitioning and metabolism. The growing interest in the physiology of adipose tissue and the role that it plays in metabolic regulation is fueled to a large degree by the realization that we are in the midst of a growing epidemic of the metabolic disorders of diabetes and obesity. In most developed countries, the incidence of both of these diseases has increased dramatically in the past few decades. Currently (as of 2003), there are 16 million diabetics in the United States, and the incidence of the disease has increased by 49% from 1990 to 2000, with projections indicating a 165% increase by 2050. There has also been a dramatic increase in the incidence of obesity, with over one-third of the population in the United States now classified as obese. Although the causal relationship between diabetes and obesity is not fully understood, a likely common link is the adipocyte.

In healthy individuals, excess fat is stored in adipocytes while only low amounts of triglyceride are maintained in nonadipocytes. It is thought that in obese individuals, the capacity for adipose tissue to accommodate excess lipid is exceeded, resulting in the abnormal accumulation of lipid in other tissues. This elevation in intracellular triglyceride content has been associated with physiological dysfunction (lipotoxicity) that contributes to the development of obesity-related type 2 diabetes. This pathological accumulation of lipid in nonadipose tissue may be the result of a physiological dysfunction of the adipocyte that is induced by the obese state. In this chapter, we will review the current state of knowledge about fat cell lipid metabolism, and how adipocytes function to balance lipid storage and mobilization to meet the energetic demands of the organism without exposing nonadipose tissues to deleterious fat accumulation.

In addition to its function as an energy-storage depot, we now understand that adipose tissue is also a bona fide endocrine organ, secreting hormones that regulate fat metabolism in other tissues throughout the body. The list of biologically active peptides known to be secreted by fat cells has grown significantly in recent years, and although the physiological function of most of these adipocyte-derived hormones (adipokines) is not fully understood, it is clear that they are important components of the physiological system

that controls lipid storage, distribution, and utilization throughout the body. Our current knowledge of the regulation of adipokine production and their downstream metabolic effects is reviewed below.

Our increased understanding of the active role that adipocytes play in regulating metabolism has stimulated a growing interest in adipose tissue as a therapeutic target for new agents to treat diabetes, obesity, and other metabolic diseases. The underlying assumption for this effort is that drugs acting on adipocyte lipid metabolism parameters or on hormone production pathways could have beneficial effects on metabolic abnormalities of diabetes and obesity. In support of this possibility are recent findings, which are reviewed below, demonstrating that an important family of antidiabetic drugs acts, at least in part, by modulating adipocyte physiology. A complete understanding of the medical significance of the adipocyte will require a comprehensive knowledge of the development of adipocytes and adipose tissue, of molecular biology and physiology of mature adipocyte, as well as the interaction of adipose tissue with the broader regulatory systems that control the whole body energy balance and fuel partitioning.

2
Physiology of the Adipocyte

Perhaps the most well-established function of adipose tissue in man is as a reservoir for the storage and mobilization of energy. Cells of the body utilize energy continuously, yet fuel is obtained from external sources intermittently. Thus, animals have a need to be able to efficiently store energy when food is available and access that energy during periods of fasting. Adipose tissue is highly specialized for the storage of lipid energy in the form of triglycerides and the mobilization of that energy in the form of free fatty acids. Of the energy substrates utilized by tissues, lipid is most efficiently stored and provides greater than twice the energy per gram than either carbohydrate (glycogen) or protein. The significance of this energy reserve is readily appreciated when considering that a man of normal body weight (70 kg, 10% body fat) can survive 40 days on energy reserves stored in adipose tissue.

2.1
Absorption and Storage of Circulating Lipid

Energy is stored in adipocytes in the form of triglycerides and is mainly derived from dietary fat, and *de novo* biogenesis from liver and within adipose tissue. Although human adipose tissue is capable of synthesizing lipid *de novo*, most triglycerides that accumulate in the tissue are derived directly or indirectly from dietary fat. Dietary fat is digested in the gut and repackaged as chylomicrons that are delivered to the systemic circulation via the lymphatics. The liver also plays a major role in the production and packaging of lipids in the form of various lipoproteins that are targeted for utilization in various target tissues, including adipose tissue.

A major mechanism for clearing chylomicrons and very low density lipoproteins (VLDL) from the circulation is through the action of lipoprotein lipase (LPL). As mentioned above, adipocytes express LPL, which is released and targeted to the capillary endothelium (Fig. 1). There, LPL acts on circulating chylomicrons and VLDL, hydrolyzing the neutral lipid core and releasing free fatty acids. The free fatty acids that are released can cross cell membranes and enter

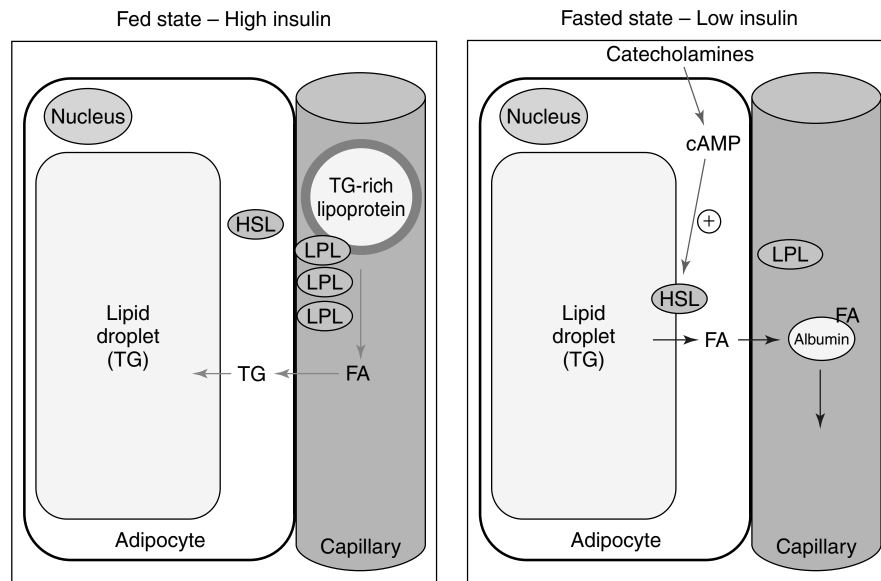


Fig. 1 Lipid storage and mobilization in adipocytes during fed and fasted states. See text for details. Flux of dietary lipid from circulating lipoprotein particles to the adipocyte lipid droplet is shown in green. Flux of stored lipid from adipocyte back into circulation shown in blue. TG, triglyceride; FA, fatty acid; HSL, hormone-sensitive lipase; LPL, lipoprotein lipase. (See color plate p. xxii).

adipocytes directly. Additionally, fatty acids can enter adipocytes via transport proteins, such as CD36 and FATP1. The flow of free fatty acids is governed by its concentration gradient. In the case of adipocytes, free fatty acid (FFA) that enter cells quickly encounter proteins that bind FFA, such as members of the fatty acid binding protein (FABP) family involved in cytoplasmic transport, or enzymes involved in metabolic conversion into triglyceride, such as long chain acyl-CoA synthase, thereby maintaining a significant concentration gradient.

2.2 Mobilization and Release of Stored Lipid into the Circulation

Triglycerides are typically stored within the adipocyte as a single lipid droplet. Recent work indicates that the lipid droplet is an extension of the endoplasmic reticulum surrounded by a single phospholipid layer. Closely associated with the lipid droplet are ''coat proteins'', especially perilipin isoforms and caveolin 2. Perilipin is the most abundant target for phosphorylation by cyclic AMP-dependent protein kinase (PK) in fat cells, and plays a key role in the retention and mobilization of energy from lipid droplets. The energy contained in the droplet is liberated when the triglyceride is hydrolyzed into free fatty acids and glycerol. The rate-limiting step in this process is the activity of hormone-sensitive lipase (HSL) at the surface of the lipid droplet. Although several protein kinases can influence lipolytic rate, the most significant of these is cyclic AMP-dependent protein kinase (PKA). Overall, the rate of lipolysis is

closely governed by the phosphorylation state of perilipin and HSL. Recent work indicates that phosphorylation of perilipin on multiple residues allows recruitment of phosphorylated HSL to the surface of the lipid droplet where triglyceride hydrolysis can take place. Phosphorylation of HSL has only a modest effect on the activity of the enzyme; rather, the dramatic increase in lipolysis produced by PKA-activation reflects the translocation and accessibility of the enzyme to its substrate within the cell. HSL appears to be physically associated with the cytosolic FABP4 (aP2). The significance of this association is not known, but could involve regulation of HSL activity or efflux of mobilized fatty acids. In this regard, mice lacking FABP4 show reduced rates of lipolysis.

Efflux of fatty acids from the fat cell and their transport across the capillary endothelium are thought to occur through passive diffusion across a concentration gradient, although this process has not been studied in detail. Long-chain fatty acids are essentially insoluble in aqueous solution and thus rely on carrier proteins for transport. In plasma, fatty acids are immediately bound with high affinity to serum albumin, which serves as a carrier of these substrates to the sites of oxidation. During fasting, mobilized fatty acids are quickly removed from the circulation, with a half-life of about two minutes, and provide an important source of energy for heart and skeletal muscle.

2.3 Regulatory Mechanisms that Balance Lipid Storage and Utilization

The pathways that promote FFA uptake and esterification to triglyceride are strongly regulated by hormones, particularly insulin. Insulin strongly promotes the synthesis of LPL whose expression coincides with the peak of triglyceride absorption. Insulin and chylomicrons also stimulate the secretion of acylation stimulation protein – a protein made by adipocytes that strongly promotes triglyceride synthesis. Insulin also promotes *de novo* fatty acid synthesis from glucose by stimulating glucose uptake and the expression of lipogenic enzymes.

Fatty acids, derived from adipose tissues, are an important energy source during fasting, exercise, and stress. As mentioned above, the overall rate of lipolysis is largely governed by the relative activation of PKA. The activity of PKA is governed by cAMP levels, which in turn are controlled by receptors coupled positively (beta adrenergic, glucagon, ACTH) or negatively (adenosine, niacin, alpha2, adrenergic) to adenylyl cyclase. In addition, insulin regulates adipocyte cAMP levels by activating phosphodiesterase PDE3b that degrades cAMP. The nature of the pathways controlling lipolysis depends on physiological circumstances. Lipolysis, in response to stress and exercise, is highly dependent on the activity of the sympathetic innervation of adipose tissue as well as the release of epinephrine from the adrenal gland. During fasting, lipolysis does not depend on neural activity, but rather appears to result from decline in insulin that provides tonic activation of PDE3b.

2.4 Adipose Tissue as an Endocrine Organ

As described above, the primary metabolic role of the adipocyte is to absorb and store excess lipid in the form of triglyceride, and to make it available to other tissues in the body as energy needs dictate, by measured release of fatty acids into the circulation. Proper functioning of

Tab. 1 Metabolically active proteins secreted by adipocytes.

Protein/hormone	***Physiological effects***
Leptin	Appetite, autonomic nervous activity
Adiponectin	Insulin sensitivity, fatty acid oxidation
Resistin	Insulin sensitivity
TNF-α	Insulin sensitivity, adipocyte differentiation, inflammation
ANG II	Lipogenesis, blood pressure
ASP	Lipogenesis
IGF	Lipogenesis, adipocyte differentiation
Interleukin-6	Lipolysis in adipocytes, inflammation
Kinins	Insulin sensitivity, tissue remodeling
PAI-1	Insulin sensitivity, blood clotting, atherosclerosis
TGF-β	Lipolysis, angiogenesis

Notes: ANG: angiotensin; ASP: acylation-stimulating protein; IGF: insulin-like growth factor; PAI-1: plasminogen activator inhibitor; TNF: tumor necrosis factor; TGF: transforming growth factor. Table adapted from Schling, P., Loffler, G. (2002) Cross talk between adipose tissue cells: impact on pathophysiology, *News Physiol. Sci.* **17**, 99–104.

this system requires communication between adipose tissue and essentially all the organ systems in the body. In addition to the hormonal and neural signals that regulate lipid uptake and storage, or induce lipolysis and release of fatty acids into the circulation during a fast (discussed above), there are signals that originate in adipose tissue that act to modify various physiological activities in tissues and organs throughout the body. For example, to insure that dietary intake is sufficient to maintain an adequate level of adiposity, there must be communication between adipose tissue and the centers in the brain that control appetite. Likewise, there is communication between adipose tissue and the organs, and the tissues that utilize fat for energy to insure that fatty acid delivered to nonadipose tissues, such as muscle and liver, are handled properly and do not accumulate to abnormal levels. One of the major advances in metabolic research in the last few years has been the discovery that adipocytes secrete hormones that act at specific sites in the body and have important effects on many aspects of energy metabolism. This new understanding of adipose tissue as an endocrine organ has dramatically changed our understanding of the significance of adipocytes in the regulation of metabolism. It is now believed that the adipose-derived hormones (referred to as adipokines) are important components of the integrated system of hormonal and neural signaling pathways that function to regulate the storage and use of metabolic energy. The following is a brief summary of the biology of four of the adipokines that have relatively clear effects on metabolism. Table 1 presents a more complete list of hormone like proteins produced by adipocytes.

The seminal contribution to the concept that adipose tissue produces hormones with important metabolic effects was made by Friedman and colleagues in 1994. These investigators identified the protein product of the *obese* (*ob*) gene that causes

severe obesity in mice when mutated. Leptin is secreted from adipose tissue and regulates body weight by acting directly in the CNS to inhibit feeding behavior. The control of leptin synthesis and secretion is still poorly understood. In general, leptin production and secretion are promoted under conditions of positive energy balance (fed state, high insulin) and suppressed by conditions of net energy deficit (e.g. fasting, catabolic hormone stimulation). As such, plasma leptin levels correlate strongly with total adipose tissue mass, and thereby provide an integrated, long-term signal indicating the status of lipid reserves. The actions of leptin are mediated through specific cell surface receptors, which are located in key central and peripheral target cells. Activation of leptin receptors in diverse brain regions signal a state of positive energy balance. Leptin-sensitive neural systems regulate the activity of the autonomic nervous system involved in energy storage and mobilization, feeding behavior, reproductive physiology, and sexual behavior. Leptin may also have direct effects on energy metabolism in peripheral tissues such as muscle, where it has been reported to cause an increase in fatty acid oxidation rates. Although leptin behaves as an antiobesity hormone in certain animal models, common human obesity does not appear to be due to abnormally low leptin levels.

Another recently identified adipocyte-secreted hormone that may play a role in both obesity and diabetes is adiponectin (also called ACRP30 or adipoQ). Originally identified as a secreted fat-specific protein whose expression was induced following adipogenesis, adiponectin levels were found to be reduced in obesity and increased by weight loss. In addition, the *adiponectin* gene maps to a region on chromosome 3 that is associated with diabetes and metabolic syndrome. Treatment of rodents with adiponectin was found to increase muscle fatty acid oxidation, reverse insulin resistance and improve hepatic insulin action. Together, these observations suggest that the physiological role of adiponectin may be to promote lipid oxidation in nonadipose tissues; in essence it may be a signal from fat indicating to the rest of the body that lipid energy is available and should be used.

In contrast to adiponectin, an adipokine that has recently been identified called *resistin* appears to have diabetes-promoting effects on metabolism. While adiponectin clearly promotes fatty acid oxidation and appears to have insulin-sensitizing effects throughout the body, resistin (also known as adipocyte secreted factor, ADSF or FIZZ3) was found to be over-expressed in rodent models of diet-induced obesity and to induce insulin resistance and glucose intolerance in normal mice. These data suggest that resistin acts in a converse manner to adiponectin, increasing insulin resistance and promoting the development of diabetes. However, this relationship between resistin and diabetes was not observed in all models of the disease and additional work will need to be carried out to fully clarify the role of resistin as another potential link between obesity and diabetes. Another potentially prodiabetic adipokine is the inflammatory cytokine tumor necrosis factor alpha (TNFα), which is secreted by adipocytes under some circumstances. TNFα production by adipocytes is elevated in obese rodents and humans and positively correlates with insulin resistance and in some studies inactivation of TNFα using antibody treatment improved insulin action. As with resistin, the combination of elevated expression in obesity

and insulin resistance promoting activity of these adipokines raises the possibility that it contributes to the functional link between obesity and diabetes. Although a great deal more work needs to be done in this area before we can fully appreciate the multiple roles that adipokine hormones play in the regulation of metabolism, it is clear that they are a crucial component of the physiological system that regulates energy balance and fuel partitioning.

3 Developmental Origin of Adipocytes

3.1 Adipose Tissue Development and Plasticity

The basic functional unit of adipose tissue is the adipocyte. Nevertheless, adipose tissue is complex and contains several cell types in addition to adipocytes, such as endothelial cells, interstitial cells, undifferentiated mesenchymal cells, pericytes, and "very small adipocytes". Indeed, adipocytes constitute less than 20% of the cells residing in typical adult fat tissue. Moreover, there are very important interactions among the various cell types that are critical to the proper functioning of the tissue. In view of the huge increase in obesity rates in the United States, and its negative impact on health, new attention has been focused on the development, maintenance, and plasticity of this important tissue.

Analysis of adipose tissue histogenesis and remodeling has relied mainly upon descriptive approaches to define cell phenotypes and deduce their transition to mature cells. In humans, adipose tissue appears as distinct lobules during the second trimester of fetal development. The specific timing of adipose tissue histogenesis and fat cell differentiation varies according to location in the body. Adipocytes within fat tissue are thought to derive initially from mesenchymal progenitors capable of differentiating into bone, muscle, as well as fat. Mesenchymal cells that are highly committed to the adipocyte lineage first appear closely associated with vessel formation, and there appears to be a close reciprocal association of developing fat cells with angiogenesis. This is not surprising since early committed cells express lipoprotein lipase that is targeted at the capillary lumen, and provides the mechanism for the transport of dietary triglyceride to fill the developing fat cells. As fat cells develop, triglycerides coalesce into small lipid droplets (nearly all of which are triglycerides) within the cytoplasm that eventually fuse to form a large single lipid droplet. The typical mature adipocyte is relatively large (30–50 micron diameter) and can reach a size of greater than 120 μm under certain conditions (Fig. 2).

Experimental investigations of adipose tissue have mainly utilized rodent models, although it is important to note that the ontogeny of adipose tissue varies widely among species, and even among fat depots within a given species. In rodent models, white adipose tissue generally appears late in embryonic development and continues to expand and differentiate during the neonatal period prior to weaning. Classic "flash" labeling experiments with an ^{3}H thymidine have shown that most proliferation of cells that are destined to become adipocytes occurs in the first postnatal week. Mitoses are mostly found in poorly differentiated mesenchymal cells that are closely associated with developing capillaries. The transition of cells from mesenchymal progenitors to mature cells can be deduced by evaluating the cellular distribution of ^{3}H label over time following

Fig. 2 Histological appearance of normal mouse adipose tissue. Shown is a 6-μm paraffin section stained with hematoxylin and eosin. Note the single large lipid droplet with numerous interstitial cells.

flash labeling. Over time, the percentage of labeled mesenchymal cells declines as the label appears in cells that have accumulated lipid. In contrast, labeling of endothelial cells remains relatively constant. These data suggest a dynamic process in which mature adipocytes are derived from committed mesenchymal progenitors that divide and develop into adipocytes. Interestingly, nearly 90% of the initial label is lost by five months, strongly indicating that cellular renewal occurs throughout life. In this regard, it is well established that cells can be isolated from human and rodent adipose tissue that readily differentiate into mature adipocytes *in vitro*. Indeed, pluripotent progenitors derived from adult adipose tissue may have numerous therapeutic applications. These observations indicate that adipose tissue contains a significant population of committed progenitors that are capable of contributing to tissue renewal and remodeling under appropriate conditions.

Under normal laboratory conditions, cellular proliferation in rat adipose tissue drops to very low levels after weaning. Nonetheless, a variety of physiological and pharmacological conditions reveal dynamic regulation of adipose tissue. For example, Hirsch and colleagues demonstrated in the 1970s that the obesity produced by high fat diets in rats involves both fat cell hyperplasia as well as hypertrophy. Fat cell renewal has also been observed after partial lipectomy, and elevated fat cell turnover has been observed in models of hypothalamic obesity. One of the best examples of physiological adipose tissue plasticity occurs in seasonal fat deposition of hibernators. Although the mechanisms involved in fat cell proliferation are largely unknown, adipose tissue itself is a rich source of growth factors and cytokines

that could trigger tissue expansion. As discussed below, certain pharmacological agents that exhibit antidiabetes properties in rodents and man target receptors that are enriched in fat tissue, and produce pronounced tissue remodeling that is likely related to the therapeutic actions of these agents.

3.2 Differentiation of Adipocytes from Precursor Cells

Together with adipocyte size, the number of adipocytes in the body is an important determinant of obesity and of multiple parameters of energy metabolism. The number of adipocytes present in an organism is determined to a large degree by the adipocyte differentiation process that generates mature adipocytes from fibroblast-like preadipocytes. Many of the molecular details of this process are now known, and the following section summarizes our current understanding of the molecular control of adipogenesis. It is important to note that our understanding of how adipocytes are generated from precursor cells is based primarily on cell culture models of adipogenesis such as the mouse 3T3L1 cell line. While these cell lines are very amenable to experimentation, they produce adipocytes that are strikingly different in some respects than native adipocytes found in adipose tissue *in vivo*. For example, fully differentiated 3T3L1 adipocytes are multilocular (contain multiple lipid droplets), while native adipocytes in white fat (the predominant type of adipose tissue in humans) display a unilocular distribution of lipid (compare Figs. 2 and 3). While we know that many of the characteristics of adipocyte differentiation in cultured cell lines are also important features of *in vivo* adipogenesis, it is important to bear in mind that some aspects of adipogenesis that have been learned from cell culture systems, as described below, may differ from the process as it occurs *in vivo*.

When cultured preadipocyte cultures are grown to confluence and cease cellular division (growth arrest), they can be induced to differentiate into adipocytes by treatment with an adipogenic hormonal cocktail containing insulin, dexamethasone, and an inducer of intracellular cAMP concentration. One of the first steps in the process of adipogenesis is the reentry of growth-arrested preadipocytes into the cell cycle and the completion of several rounds of clonal expansion. Multiple genes involved in the cell cycle control are required for this step to proceed, including the tumor suppressor retinoblastoma protein (Rb) and several cyclin-dependent kinases and their inhibitors (p18, p21, and p27). This and the subsequent steps of the program of adipogenesis are controlled, to a large degree, by a cascade of gene expression events regulated by a small set of transcription factors. Two families of transcription factors have emerged as the key determinants of this process: the three CCAAT/enhancer-binding proteins C/EBP α, β and δ, and the two-peroxisome proliferator-activated receptors gamma-1 and gamma-2 (PPARγ1 and PPARγ2).

One of the initial steps in the transcriptional cascade in response to adipogenic signals is the rapid induction of C/EBP β and δ expression. These transcription factors orchestrate cell cycle reentry by stimulating the expression of the CDK inhibitor p21, which acts to inhibit the Rb protein and relieve its block on cell cycle progression. C/EBP δ and β have also been shown to induce expression of the gene for the PPARγ transcription factor that plays

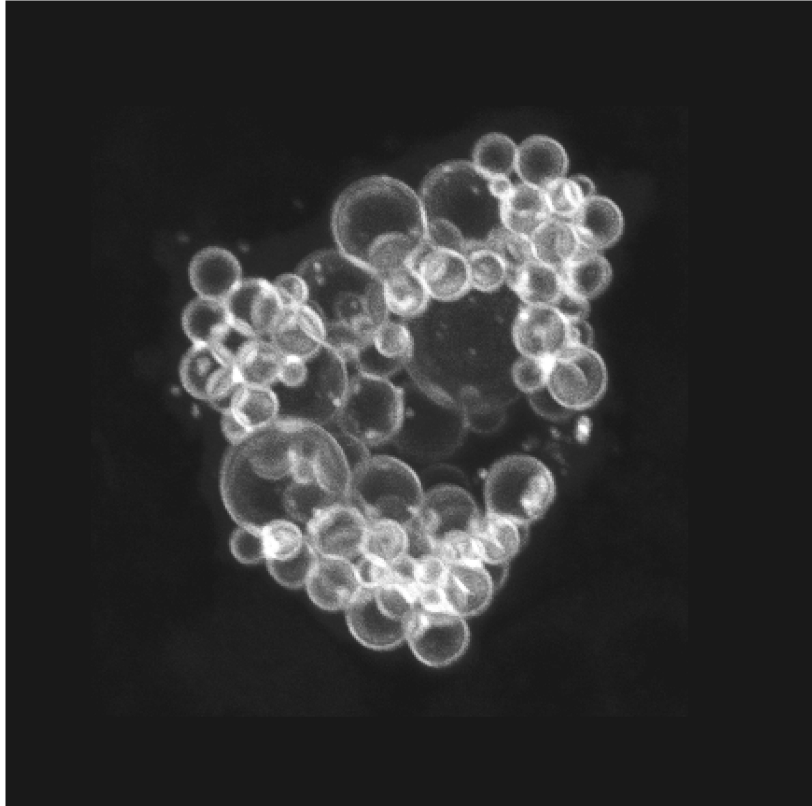


Fig. 3 Fluorescence micrograph of a cultured adipocyte expressing EGFP–perilipin fusion protein. A C3H 10T1/2 mesenchymal cell was differentiated in culture and transfected with an expression vector encoding perilipin fused to enhanced green fluorescent protein. Note the localization of the fluorescence to surface of the numerous lipid droplets (see color plate p. xxi).

a key role in the terminal differentiation of adipocytes (discussed in more detail below). The importance of C/EBP β and δ for adipogenesis was clearly demonstrated by loss-of-function and gain-of-function genetic studies in mice. Overexpression of either C/EBP β or δ in preadipocytes enhanced adipogenesis, while embryonic fibroblast cells derived from mice lacking either C/EBP β or δ had reduced levels of adipogenesis compared with wild type.

The induction of C/EBP β and δ is immediately followed by an increase in PPARγ and C/EBPα expression. PPARγ is a member of the nuclear hormone receptor family of ligand-activated transcription factors. It is absolutely required for adipocyte differentiation, as a genetic knockout of the PPARγ gene in mice prevents the development of all fat tissue. In addition to its crucial role in adipocyte differentiation PPARγ is the receptor for the thiazolidinedione (TZD) class of antidiabetic drugs, indicating that it is also important in metabolic regulation in adult organisms. In mice and humans there are two isoforms, PPARγ1 and PPARγ2, which are derived from the same gene by

alternative promoter usage and RNA splicing. While the expression of PPARγ2 is restricted almost exclusively to adipocytes, PPARγ1 has a broader pattern of expression although it is still most abundant in adipocytes. Although PPARγ2 is identical to PPARγ1 except that it contains an additional 28 amino acids on its N-terminus, it appears that the two proteins have distinct activities with regard to adipocyte differentiation. When the expression of the PPARγ2 isoform was blocked, adipogenesis was more strongly inhibited than when the PPARγ1 isoform was blocked. In addition, exogenous delivery of PPARγ2 into PPARγ deficient cells was able to completely restore the adipogenesis, whereas overexpression of PPARγ1 had little effect. It may be that PPARγ1, which is already expressed in preadipocytes, behaves as a priming factor (along with C/EBP β and δ) for the induction of PPARγ2 or for the generation of endogenous PPARγ ligands that play a role in later stages of adipogenesis.

As the program of differentiation proceeds, the expression of C/EBPα rises immediately after the increase in PPARγ2 expression. Like PPARγ, C/EBPα also plays an essential role in adipose development as targeted gene knockout in mice results in embryonic lethality and failure to develop normal adipose tissue. There has been an intense research effort to understand the relationship between these two transcription factors and the role they play in adipogenesis. Several studies have demonstrated that PPARγ2 and C/EBPα coregulate each other's expression. Mice with reduced PPARγ expression due to heterozygous gene knockout displayed a drastically reduced level of C/EBPα, and mice with disrupted C/EBPα expression showed a reduced level of PPARγ. Introduction of either PPARγ or C/EBPα into NIH3T3 cells is sufficient to convert these normally nonadipogenic cells from fibroblasts into adipocytes. However, it is unclear if either of the transcription factors, completely on its own, could induce adipogenesis. Taken together, most of the recent evidence supports the model that while both of the transcription factors work coordinately to carry out adipogenesis, PPARγ2 probably plays the primary role, while C/EBPα may act mostly by inducing and maintaining PPARγ2 expression. C/EBPα may also function to regulate the transcription of genes involved in the metabolic actions of insulin such as glucose transporter 4 (Glut 4). Clearly, PPARγ and C/EBPα are key transcription factors in adipogenesis, acting synergistically to generate fully differentiated, insulin-responsive adipocytes.

Although our understanding of adipocyte differentiation, as described above, is derived from work in cultured cell lines, it is likely that many of the pathways and key components also play an important role in generating adipocytes from precursor cells *in vivo* and in the remodeling of adipose tissue as it occurs during certain metabolic stress (see Sect. 3, above) or treatment with specific pharmacological agents (see Sect. 5, below).

4
Adipose Dysfunction and Metabolic Disease

Just as our appreciation of how active a role the adipose tissue plays in the metabolic regulation has grown in recent years, so too has the interest in the role that adipocytes play in the development (and treatment) of metabolic diseases such as obesity and diabetes. The investigation of the role of adipose tissue in these diseases has taken on added significance in recent years

because of the dramatic rise in the prevalence of these related disorders. Although the causal relationship between diabetes and obesity is not fully understood, a likely common link is the adipocyte. Given that adipose tissue serves to buffer excess lipid, and to confer insulin resistance that is associated with inappropriate accumulation of lipid in nonadipose cells, it is not surprising that the dysfunction in adipose tissue could profoundly affect diabetes susceptibility. The inability of adipose tissue to buffer circulating lipid levels and concurrent accumulation of lipid in muscle and liver may, in fact, be an early event in the development of diabetes. This potential link between adipose tissue function and diabetes raises the new and exciting possibilities for the development of therapeutic agents to treat diabetes. Drugs targeted to very specific metabolic or hormonal functions in the adipocyte could potentially have a profound beneficial effects on metabolism throughout the body. The following is a summary of the known defects in adipose function and their effects on energy metabolism and glucose homeostasis. We will also review what is known about the antidiabetic thiazolidinedione drugs that are thought to have their effects by modifying adipocyte physiology.

4.1
Obesity – Too Much Fat

A long-standing observation in the field of diabetes research is that obese individuals have an increased risk of developing insulin resistance and type 2 diabetes. The current epidemic of diabetes in western cultures began about thirty years ago and correlates precisely with the concurrent explosion of obesity. This has led to the assumption that there is a functional, or causative relationship between these two disorders; that the presence of obesity increases susceptibility to diabetes. Although the relation between these two metabolic disorders is unmistakable, it is not understood how obesity precipitates the multiple functional defects that precede overt diabetes. However, an emerging candidate is the acquired inability of adipocyte to protect important tissues and cells throughout the body from exposure to excessive lipids.

Normally, excess calories are stored in the body in the form of triglycerides, mainly in adipocytes of white fat tissue. However, nonadipocyte cells also store low levels of triglycerides to meet internal metabolic requirements. In these cells, the amount of stored triglyceride is maintained within a very narrow range. A fatty acid homeostatic system functions in healthy individuals to direct excess fat to adipocytes while maintaining normal amounts of triglyceride in nonadipocytes. Under conditions of chronic positive energy balance and ensuing obesity, however, the capacity for adipose tissue to buffer excess lipid energy may be exceeded, resulting in the pathological accumulation of lipid within key metabolic tissues.

The elevation of free fatty acid (FFA) in the blood is predictive of conversion from impaired glucose tolerance to diabetes, suggesting that FFA themselves could contribute to organ defects that precede type 2 diabetes, such as insulin resistance of skeletal muscle, and reduced pancreatic insulin secretion. More recently, it has been shown that intracellular accumulation of lipid within myocytes strongly predicts insulin resistance *independent of the magnitude of obesity*. Thus, it is not the accumulation of lipid in fat cells *per se* that is problematic, rather it is the redirection of lipid to other key cells that occurs once the capacity of the adipocyte to handle

lipid has been exceeded. This condition is perhaps most dramatically illustrated by patients with lipodystrophy, who are severely insulin-resistant and develop diabetes at an early age. These individuals lack certain adipose tissue depots and therefore have reduced capacity to buffer excess lipid (presented in more detail below). Viewed from this perspective, obesity leads to diabetes when abnormal amounts of triglyceride accumulate in nonadipose tissues such as skeletal muscle and pancreatic islets. Excess triglyceride deposition in nonadipocytes might generate a lipid environment that interferes with cellular physiology and gene expression in ways that contribute to insulin resistance and β-cell failure associated with type 2 diabetes.

4.2 Lipodystrophy – Too Little Fat

The model described above proposes that obesity promotes the development of type 2 diabetes because adipocytes from obese individuals have reduced capacity for further lipid storage, which leads to inappropriate fat accumulation in nonadipose tissue. As mentioned above, the effects of aberrant amounts of adipose tissue on physiology are best illustrated in the rare, but informative, lipodystrophy syndromes. These inherited and acquired syndromes are characterized by the partial or complete loss of adipose tissue, insulin resistance, and diabetes. Several mutations that cause inherited forms of lipodystrophy, such as partial familial lipodystrophy (PFLD), have been identified. One of the most common genetic defects in PFLD are mutations in the gene encoding the nuclear lamin a/c protein. Although the mechanism by which mutant lamin a/c protein causes lipodystrophy is not known at this time, it seems likely that the protein, which plays a fundamental role in the function of the nucleus, is required for normal adipocyte differentiation of maintenance. Patients carrying PFLD mutations exhibit an absence of adipose tissue in their extremities, elevated circulating lipids, abnormally high levels of intracellular triglyceride in muscle and liver, and diabetes. Interestingly, mutations causing PFLD have recently been identified in the PPARγ gene, underscoring the key role of this transcription factor in the development and maintenance of adipose tissue.

The conclusions derived from observations of human lipodystrophy syndromes are strongly supported by animal models of lipodystrophy. Several different genetic methodologies have been used to generate mice that have reduced amounts of adipose tissue. Like their human counterparts, these animals invariably exhibit elevated circulating lipids, abnormally high levels of intracellular triglyceride levels in multiple tissues including muscle and liver, severe insulin resistance, and diabetes. Interestingly, these metabolic defects can be improved by transplanting adipose tissue back into the fatless mice. These findings clearly illustrate the vital role that adipose tissue plays in protecting the organism from abnormal lipid accumulation and its deleterious effects on metabolism.

5 The Adipocyte as a Therapeutic Target for Metabolic Disease

As our understanding of the importance of adipose tissue in controlling metabolism has grown, so too has the realization that therapeutic agents designed to modify adipocyte physiology could provide new avenues for the treatment of diabetes, obesity, and other metabolic

diseases. As described above, our current understanding of the role of adipose in protecting other tissue from deleterious lipid accumulation suggests that drugs acting on this aspect of adipocyte function might have beneficial effects. Likewise, the observations (also described above) on adipocyte derived hormones and their metabolic activities, suggests that therapeutic agents modulating the production of key adipokines could also improve certain metabolic parameters. The view that drugs acting on adipocytes could be useful is reinforced by the relatively recent observation that a major class of antidiabetic drugs, the thiazolidinediones, have direct effects on adipocytes. Although this is an active area of research, a generally accepted model of TZD action has emerged in which the antidiabetic effect of these drugs is mediated by activation of the PPARγ transcription factor in adipocytes. The following is a review of our current understanding of how activation of PPARγ in adipocytes might improve insulin sensitivity and other metabolic parameters. This model can serve as a general example of how modification of adipocyte physiology could improve whole body metabolic status.

5.1 Antidiabetic PPARγ Ligands Act on Adipocytes

In patients with type 2 diabetes, treatment with TZD compounds results in a dramatic improvement in peripheral insulin sensitivity and a reduction in plasma glucose concentrations. The TZD compounds were first identified in the 1980s as antidiabetic agents in rodents, well before the discovery of the PPARγ receptor and in the absence of any knowledge of their mechanism of action. In the 1990s, it was discovered that TZDs could activate PPARγ and cause the differentiation of preadipocytes into adipocytes and it is now generally accepted that the antidiabetic activities of the TZDs are mediated by activation of PPARγ. Perhaps the clearest evidence that PPARγ activation mediates the antidiabetic effects of these drugs are the recently discovered synthetic ligands for PPARγ that have been selected exclusively for their ability to activate the receptor. These non-TZD PPARγ agonists show very similar antidiabetic activity to the TZDs.

Although the exact mechanism by which these drugs improve peripheral insulin sensitivity and reduce plasma glucose concentration is not fully understood, several general possibilities have emerged. First, TZDs may have a beneficial effect on metabolism by increasing the fat cell number and size, leading to greater lipid storage capacity and increased protection of nonadipose tissues from the deleterious effects of excess lipid accumulation. Another scenario is that, PPARγ agonists act on the mature adipocyte to alter the production of adipose-derived hormones or metabolic signals that function to improve metabolic parameters in other tissues and organs such as muscle, liver and pancreas. Finally, it is also possible that TZDs exert some of their metabolic effects through PPARγ present in nonadipose tissues such as skeletal muscle.

One feature of TZD treatment that could be central to the therapeutic actions of these compounds is a remodeling of adipose tissue. TZD treatment induces the appearance of clusters of small multilocular adipocytes and loss of large unilocular adipocytes in Zucker diabetic rats. This effect is also observed in mice treated with TZD compounds as seen in Fig. 4. It has been hypothesized that the newly

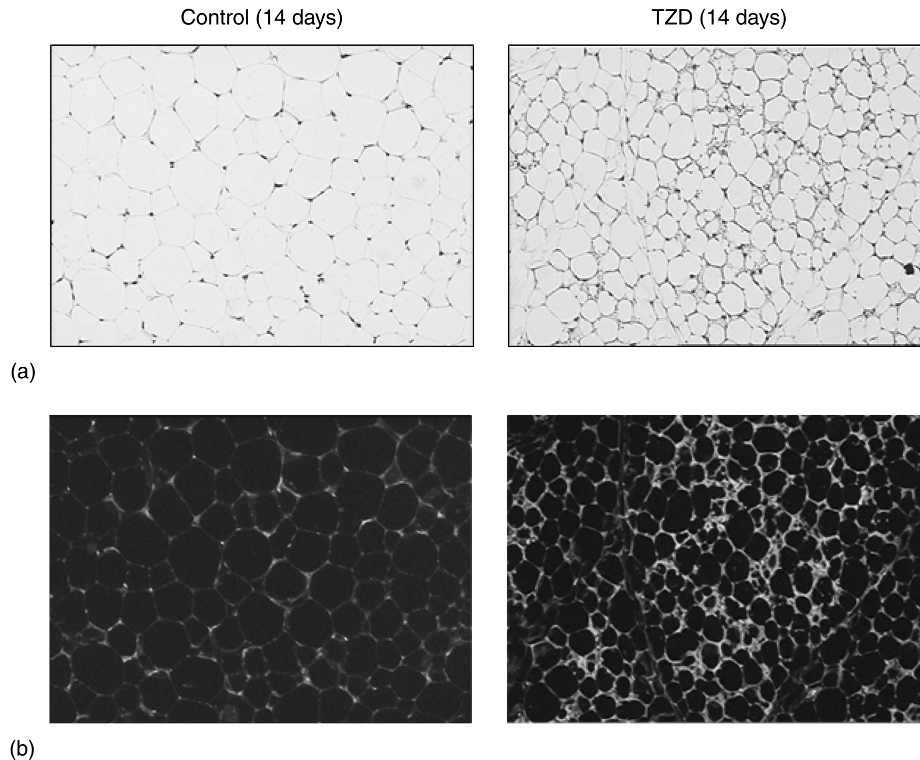


Fig. 4 Thiazolidinedione treatment remodels white adipose tissue. Mice were treated for 2 weeks with the TZD T174, and subcutaneous white fat was processed by (a) hematoxylin and eosin staining and (b) tissue autofluorescence. Note the appearance of numerous clusters of small multilocular adipocytes in TZD-treated tissue.

appearing small adipocytes might be more insulin sensitive and/or secrete lower levels of prodiabetes hormones and thereby contribute to the insulin-sensitizing effects of the drug. It is also possible that the increased number of smaller adipocytes, especially in the appropriate adipose beds, may improve the ability of adipose tissue to store excess lipid and reduce deleterious accumulation of triglyceride in muscle, liver, and pancreatic islets. In support of this possibility is the observation that many of the known target genes for PPARγ, whose expression would presumably be activated upon TZD treatment, are involved in lipogenesis. The increased rates of lipogenesis resulting from gene activation may increase the capacity of adipocytes to store lipid, thereby preventing triglyceride accumulation (and lipotoxicity) in nonadipose tissue. Whether the small adipocytes are derived from stem cell mitosis, recruitment of committed preadipocytes, or possibly by division of mature cells is not known. The loss of large fat cells was attributed to cellular apoptosis; however, the impact of TZD treatment on the fate and turnover of mature adipocytes has not been investigated directly.

Consistent with the observations of adipose tissue remodeling is the increased subcutaneous fat mass, and reduced

visceral fat mass seen in diabetic patients treated long-term with TZDs. Visceral fat is known to be more lipolytic in response to catecholamine stimulation than subcutaneous fat, and to efficiently deliver free fatty acids and other secreted factors to insulin-sensitive tissues such as liver and muscle, possibly causing an increase in insulin resistance. Although intrinsic metabolic differences between subcutaneous and visceral fat are not completely understood, current evidence suggests that subjects with increased visceral fat are at considerably higher risk for diabetes and cardiovascular complications than those with increased subcutaneous fat. These observations, plus the demonstration that PPARγ levels are higher in subcutaneous than in visceral fat, raise the possibility that PPARγ activation by TZDs is fat depot specific, and that differential activation of PPARγ in subcutaneous fat leads to a beneficial reproportioning of key metabolically active adipose beds.

The other possible mechanism by which activation of PPARγ in adipocytes could have effects on metabolism throughout the body, is by modulation of adipokine production. Adiponectin is an excellent candidate for a fat-derived hormone mediating the antidiabetic effects of PPARγ ligands. As described above, adiponectin has antidiabetic and antiobesity activity when introduced into rodents. Importantly, it has recently been demonstrated that levels of adiponectin are increased in patients treated with TZDs and that its expression in adipocytes is induced by PPARγ agonists. Another possibility is that the adipokine, resistin is negatively regulated by PPARγ. It has been reported that in some models, treatment with TZDs results in a reduction of resistin synthesis, which would have beneficial effects on insulin sensitivity. Finally, TZDs, could suppress expression of genes, such as TNFα and PAI-1, that might also contribute to systemic insulin resistance. Interestingly a mutual antagonism exists between TNFα and PPARγ; TNFα inhibits PPARγ expression in adipocytes whereas PPARγ activation by TZDs can partially overcome the diabetogenic effects of TNFα, potentially explaining at least some of the insulin-sensitizing activity of PPARγ ligands.

While there is still obviously a great deal to learn about which of these many possible mechanisms actually mediate the effects of PPARγ activation in adipose tissue, it is clear that an understanding of how these drugs affect adipose physiology will provide many clues into the complex pathways by which the adipocyte influences metabolism throughout the organism.

See also Diabetes Insipidus, Molecular Biology of.

Bibliography

Books and Reviews

Frayn, K.N., Karpe, F., Fielding, B.A., Macdonald, I.A., Coppack, S.W. (2003) Integrative physiology of human adipose tissue, *Int. J. Obes.* **27**, 875–888.

Friedman, J.M. (2002) The function of leptin in nutrition, weight, and physiology, *Nutr. Rev.* **60**, S1–14; discussion S68–84, 5–7.

Londos, C., Brasaemle, D.L., Schultz, C.J., Adler-Wailes, D.C., Levin, D.M., et al. (1999) On the control of lipolysis in adipocytes, *Ann. N. Y. Acad. Sci.* **892**, 155–168.

Picard, F., Auwerx, J. (2002) PPAR(gamma) and glucose homeostasis, *Annu. Rev. Nutr.* **22**, 167–197.

Renold, A.E., Cahill, J.F. (1965) Adipose Tissue, *Handbook of Physiology, A Critical, Comprehensive Presentation of Physiological Knowledge and Concepts*, Section 5, American Physiological Society, Washington, DC.

Rosen, E.D., Spiegelman, B.M. (2000) Molecular regulation of adipogenesis, *Annu. Rev. Cell. Dev. Biol.* **16**, 145–171.

Primary Literature

Agarwal, A.K., Garg, A. (2002) A novel heterozygous mutation in peroxisome proliferator-activated receptor-gamma gene in a patient with familial partial lipodystrophy, *J. Clin. Endocrinol. Metab.* **87**, 408–411.

Akazawa, S., Sun, F., Ito, M., Kawasaki, E., Eguchi, K. (2000) Efficacy of troglitazone on body fat distribution in type 2 diabetes, *Diabetes Care* **23**, 1067–1071.

Atanassova, P. (1998) Histochemical and ultrastructural criteria for early differentiation of human subcutaneous adipocytes, *Folia Med. (Plovdiv)* **40**, 24–27.

Atanassova, P.K. (2002) Cytochemical study of in situ differentiation of white adipose cells in rat, *Folia Med. (Plovdiv)* **44**, 32–36.

Bartness, T.J., Wade, G.N. (1985) Photoperiodic control of seasonal body weight cycles in hamsters, *Neurosci. Biobehav. Rev.* **9**, 599–612.

Berg, A.H., Combs, T.P., Du, X., Brownlee, M., Scherer, P.E. (2001) The adipocyte-secreted protein Acrp30 enhances hepatic insulin action, *Nat. Med.* **7**, 947–953.

Brodows, R.G., Campbell, R.G., Al-Aziz, A.J. (1976) Lack of central autonomic regulation of substrate during early fasting in man, *Metabolism* **25**, 803–807.

Brown, D.A. (2001) Lipid droplets: proteins floating on a pool of fat, *Curr. Biol.* **11**, R446–R449.

Cianflone, K., Xia, Z., Chen, L.Y. (2003) Critical review of acylation-stimulating protein physiology in humans and rodents, *Biochim. Biophys. Acta* **1609**, 127–143.

Cobb, J.E., Blanchard, S.G., Boswell, E.G., Brown, K.K., Charifson, P.S., et al. (1998) N-(2-Benzoylphenyl)-L-tyrosine PPARgamma agonists. 3. Structure-activity relationship and optimization of the N-aryl substituent, *J. Med. Chem.* **41**, 5055–5069.

Darlington, G.J., Ross, S.E., MacDougald, O.A. (1998) The role of C/EBP genes in adipocyte differentiation, *J. Biol. Chem.* **273**, 30057–30060.

de Souza, C.J., Eckhardt, M., Gagen, K., Dong, M., Chen, W., et al. (2001) Effects of pioglitazone on adipose tissue remodeling within the setting of obesity and insulin resistance, *Diabetes* **50**, 1863–1871.

De Ugarte, D.A., Ashjian, P.H., Elbarbary, A., Hedrick, M.H. (2003) Future of fat as raw material for tissue regeneration, *Ann. Plast. Surg.* **50**, 215–219.

Faust, I.M., Johnson, P.R., Hirsch, J. (1979) Adipose tissue regeneration in adult rats, *Proc. Soc. Exp. Biol. Med.* **161**, 111–114.

Faust, I.M., Johnson, P.R., Stern, J.S., Hirsch, J. (1978) Diet-induced adipocyte number increase in adult rats: a new model of obesity, *Am. J. Physiol.* **235**, E279–E286.

Freake, H.C., Moon, Y.K. (2003) Hormonal and nutritional regulation of lipogenic enzyme mRNA levels in rat primary white and brown adipocytes, *J. Nutr. Sci. Vitaminol. (Tokyo)* **49**, 40–46.

Friedman, J.M. (2002) The function of leptin in nutrition, weight, and physiology, *Nutr. Rev.* **60**, S1–S14; discussion S68–84, 5–7.

Fruebis, J., Tsao, T.S., Javorschi, S., Ebbets-Reed, D., Erickson, M.R., et al. (2001) Proteolytic cleavage product of 30-kDa adipocyte complement-related protein increases fatty acid oxidation in muscle and causes weight loss in mice, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2005–2010.

Gimble, J.M. (2003) Adipose tissue-derived therapeutics, *Expert Opin. Biol. Ther.* **3**, 705–713.

Haemmerle, G., Zimmermann, R., Hayn, M., Theussl, C., Waeg, G., et al. (2002) Hormone-sensitive lipase deficiency in mice causes diglyceride accumulation in adipose tissue, muscle, and testis, *J. Biol. Chem.* **277**, 4806–4815.

Hausberger, F.X. (1966) Pathological changes in adipose tissue of obese mice, *Anat. Rec.* **154**, 651–660.

Hausman, G.J., Kauffman, R.G. (1986) Mitotic activity in fetal and early postnatal porcine adipose tissue, *J. Anim. Sci.* **63**, 659–673.

Hausman, G.J., Thomas, G.B. (1984) Enzyme histochemical differentiation of white adipose tissue in the rat, *Am. J. Anat.* **169**, 315–326.

Hegele, R.A. (2001) Molecular basis of partial lipodystrophy and prospects for therapy, *Trends Mol. Med.* **7**, 121–126.

Hegele, R.A., Cao, H., Frankowski, C., Mathews, S.T., Leff, T. (2002) PPARG F388L, a transactivation-deficient mutant, in familial partial lipodystrophy, *Diabetes* **51**, 3586–3590.

Hellerstein, M.K. (1999) De novo lipogenesis in humans: metabolic and regulatory aspects, *Eur. J. Clin. Nutr.* **53**(Suppl. 1), S53–S65.

Hellerstrom, B.H.A.C. (1961) Cell renewal in the white and brown fat tissue of the rat, *Acta Pathol.* **51**, 347–353.

Hotamisligil, G.S., Arner, P., Caro, J.F., Atkinson, R.L., Spiegelman, B.M. (1995) Increased adipose tissue expression of tumor necrosis factor-alpha in human obesity and insulin resistance, *J. Clin. Invest.* **95**, 2409–2415.

Kelley, D.E., Goodpaster, B.H., Storlien, L. (2002) Muscle triglyceride and insulin resistance, *Annu. Rev. Nutr.* **22**, 325–346.

Londos, C., Brasaemle, D.L., Schultz, C.J., Adler-Wailes, D.C., Levin, D.M., et al. (1999) On the control of lipolysis in adipocytes, *Ann. N. Y. Acad. Sci.* **892**, 155–168.

Merkel, M., Eckel, R.H., Goldberg, I.J. (2002) Lipoprotein lipase: genetics, lipid uptake, and regulation, *J. Lipid Res.* **43**, 1997–2006.

Miller, W.H., Jr, Faust, I.M., Hirsch, J. (1984) Demonstration of de novo production of adipocytes in adult rats by biochemical and radioautographic techniques, *J. Lipid Res.* **25**, 336–347.

Minokoshi, Y., Kim, Y.B., Peroni, O.D., Fryer, L.G., Muller, C., et al. (2002) Leptin stimulates fatty-acid oxidation by activating AMP-activated protein kinase, *Nature* **415**, 339–343.

Mokdad, A.H., Bowman, B.A., Ford, E.S., Vinicor, F., Marks, J.S., Koplan, J.P. (2001) The continuing epidemics of obesity and diabetes in the United States, *JAMA* **286**, 1195–1200.

Moller, D.E. (2000) Potential role of TNF-alpha in the pathogenesis of insulin resistance and type 2 diabetes, *Trends Endocrinol. Metab.* **11**, 212–217.

Mori, Y., Murakawa, Y., Okada, K., Horikoshi, H., Yokoyama, J., et al. (1999) Effect of troglitazone on body fat distribution in type 2 diabetic patients, *Diabetes Care* **22**, 908–912.

Ochi, M., Furukawa, H., Yoshioka, H., Sawada, T., Kusunoki, T., Hattori, T. (1991) Adipocyte dynamics in hypothalamic obese mice during food deprivation and refeeding, *J. Nutr. Sci. Vitaminol. (Tokyo)* **37**, 479–491.

Ochi, M., Sawada, T., Hattori, T. (1987) Tritiated thymidine autoradiographic study on postnatal development of epididymal adipose tissue in the normal mouse, *Anat. Embryol. (Berl)* **177**, 139–145.

Okuno, A., Tamemoto, H., Tobe, K., Ueki, K., Mori, Y., et al. (1998) Troglitazone increases the number of small adipocytes without the change of white adipose tissue mass in obese Zucker rats, *J. Clin. Invest.* **101**, 1354–1361.

Paolisso, G., Howard, B.V. (1998) Role of non-esterified fatty acids in the pathogenesis of type 2 diabetes mellitus, *Diabet. Med.* **15**, 360–366.

Picard, F., Auwerx, J. (2002) PPAR(gamma) and glucose homeostasis, *Annu. Rev. Nutr.* **22**, 167–197.

Pilgrim, C. (1971) DNA synthesis and differentiation in developing white adipose tissue, *Dev. Biol.* **26**, 69–76.

Poissonnet, C.M., Burdi, A.R., Bookstein, F.L. (1983) Growth and development of human adipose tissue during early gestation, *Early Hum. Dev.* **8**, 1–11.

Poissonnet, C.M., Burdi, A.R., Garn, S.M. (1984) The chronology of adipose tissue appearance and distribution in the human fetus, *Early Hum. Dev.* **10**, 1–11.

Ranallo, R.F., Rhodes, E.C. (1998) Lipid metabolism during exercise, *Sports Med.* **26**, 29–42.

Reitman, M.L. (2002) Metabolic lessons from genetically lean mice, *Annu. Rev. Nutr.* **22**, 459–482.

Ren, D., Collingwood, T.N., Rebar, E.J., Wolffe, A.P., Camp, H.S. (2002) PPARgamma knockdown by engineered transcription factors: exogenous PPARgamma2 but not PPARgamma1 reactivates adipogenesis, *Genes Dev.* **16**, 27–32.

Rosen, E.D., Spiegelman, B.M. (2000) Molecular regulation of adipogenesis, *Annu. Rev. Cell. Dev. Biol.* **16**, 145–171.

Scheja, L., Makowski, L., Uysal, K.T., Wiesbrock, S.M., Shimshek, D.R., et al. (1999) Altered insulin secretion associated with reduced lipolytic efficiency in aP2-/- mice, *Diabetes* **48**, 1987–1994.

Schling, P., Loffler, G. (2002) Cross talk between adipose tissue cells: impact on pathophysiology, *News Physiol. Sci.* **17**, 99–104.

Scow, R.O., Blanchette-Mackie, E.J. (1991) Transport of fatty acids and monoacylglycerols in white and brown adipose tissues, *Brain Res. Bull.* **27**, 487–491.

Shen, W.J., Sridhar, K., Bernlohr, D.A., Kraemer, F.B. (1999) Interaction of rat hormone-sensitive lipase with adipocyte lipid-binding protein, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5528–5532.

Spiegelman, B.M. (1998) PPAR-gamma: adipogenic regulator and thiazolidinedione receptor, *Diabetes* **47**, 507–514.

Stahl, A., Gimeno, R.E., Tartaglia, L.A., Lodish, H.F. (2001) Fatty acid transport proteins: a current view of a growing family, *Trends Endocrinol. Metab.* **12**, 266–273.

Steppan, C.M., Lazar, M.A. (2002) Resistin and obesity-associated insulin resistance, *Trends Endocrinol. Metab.* **13**, 18–23.

Tanaka, T., Yoshida, N., Kishimoto, T., Akira, S. (1997) Defective adipocyte differentiation in mice lacking the C/EBPbeta and/or C/EBPdelta gene, *EMBO J.* **16**, 7432–7443.

Tauchi-Sato, K., Ozeki, S., Houjou, T., Taguchi, R., Fujimoto, T. (2002) The surface of lipid droplets is a phospholipid monolayer with a unique fatty acid composition, *J. Biol. Chem.* **277**, 44507–44512.

Tsao, T.S., Lodish, H.F., Fruebis, J. (2002) ACRP30, a new hormone controlling fat and glucose metabolism, *Eur. J. Pharmacol.* **440**, 213–221.

Unger, R.H. (2002) Lipotoxic diseases, *Annu. Rev. Med.* **53**, 319–336.

Wang, S.P., Laurin, N., Himms-Hagen, J., Rudnicki, M.A., Levy, E., et al. (2001) The adipose tissue phenotype of hormone-sensitive lipase deficiency in mice, *Obes. Res.* **9**, 119–128.

Way, J.M., Gorgun, C.Z., Tong, Q., Uysal, K.T., Brown, K.K., et al. (2001) Adipose tissue resistin expression is severely suppressed in obesity and stimulated by peroxisome proliferator-activated receptor gamma agonists, *J. Biol. Chem.* **276**, 25651–25653.

Yamauchi, T., Kamon, J., Waki, H., Murakami, K., Motojima, K., et al. (2001) The mechanisms by which both heterozygous peroxisome proliferator- activated receptor gamma (PPARgamma) deficiency and PPARgamma agonist improve insulin resistance, *J. Biol. Chem.* **276**, 41245–41254.

Yamauchi, T., Kamon, J., Waki, H., Terauchi, Y., Kubota, N., et al. (2001) The fat-derived hormone adiponectin reverses insulin resistance associated with both lipoatrophy and obesity, *Nat. Med.* **7**, 941–946.

Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L., Friedman, J.M. (1994) Positional cloning of the mouse obese gene and its human homologue, *Nature* **372**, 425–432.

Aggregation, Protein

Jeannine M. Yon
Université de Paris-sud, Orsay, France

1 Introduction 25

2 Protein Folding, Misfolding, and Aggregation 25
2.1 The New View of Protein Folding 25
2.2 Detection of Aggregates during the Refolding Process 27
2.2.1 Transient Aggregation 27
2.2.2 Irreversible Aggregation 29
2.3 Mechanisms of Protein Aggregation 30

3 Protein Folding in the Cellular Environment 34
3.1 Molecular Crowding in the Cells 34
3.2 The Role of Molecular Chaperones 34

4 Protein Aggregation in the Cellular Environment 40
4.1 The Formation of Inclusion Bodies 40
4.1.1 Occurrence of Inclusion Bodies 40
4.1.2 Characteristics of Inclusion Bodies 41
4.1.3 Strategies for Refolding Inclusion Body Proteins 43
4.2 The Formation of Amyloid Fibrils and its Pathological Consequences 45

Bibliography 49
Books and Reviews 49
References of Primary Literature 49

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Keywords

Aggregates
The association of nonnative protein molecules through intermolecular hydrophobic interactions.

Amyloid Fibrils
Ordered aggregates.

Inclusion Bodies
Insoluble, amorphous, disordered aggregates.

Molecular Chaperones
Proteins that assist protein folding within cells.

Protein Folding
The process by which polypeptide chains acquire their three-dimensional and functional structure.

Three dimensional–Domain Swapping
An aggregation mechanism in which one domain in a multidomain protein is swapped with the same domain of another molecule.

Protein misfolding and aggregation are frequent phenomena that occur under different conditions *in vivo* as well as *in vitro*. Aggregation is a serious problem affecting both the production of proteins in the biotechnology and pharmaceutical industries and human health. The aggregates are formed from nonnative proteins through intermolecular interactions that compete with intramolecular interactions. There is thus a kinetic competition between proper folding and misfolding, which can generate aggregates.

Recent evidence for transient association of intermediates during *in vitro* refolding has been obtained for several monomeric proteins. Irreversible and insoluble aggregates are formed in an off-pathway folding process; their formation is concentration dependent and could be prevented by using very small protein concentrations. These aggregates can dissociate and dissolve only in the presence of high concentrations of denaturant. The mechanisms involved in these aggregation processes will be discussed in light of the so-called *new view* of protein folding.

The environmental conditions within cells are markedly different from those used in *in vitro* refolding studies. In the production of recombinant proteins in foreign hosts, the formation of disordered aggregates, that is, inclusion bodies, is often observed. However, aggregation can also result in the formation of amyloid fibrils, which are ordered aggregates. These amyloid formations are at the origin of serious diseases.

1 Introduction

Protein misfolding and aggregation have been recognized for many years as common processes. Aggregation can occur under various conditions. The aggregation of which we are speaking is very different from the precipitation of a native protein at the isoelectric point or upon salting out, which can be reversed under appropriate conditions. In the precipitate, the protein remains in a native conformation. The aggregates, however, are formed from partially folded intermediates and result from intermolecular interactions, which compete with intramolecular interactions. Thermal denaturation of proteins is frequently accompanied by the formation of aggregates leading to the irreversibility of the process. As early as 1931, Wu, in a review on protein denaturation, distinguished between aggregation and precipitation. The aggregated species are not in equilibrium with the soluble species, complicating experimental approaches.

Aggregation has been reported to occur during the *in vitro* refolding of monomeric as well as oligomeric proteins, lowering the refolding yield. As mentioned above, the use of very low protein concentrations could prevent protein aggregation. However, during the folding of nascent polypeptide chains biosynthesized within prokaryotic and eukaryotic cells, aggregates can accumulate. The overexpression of genes in foreign hosts often result in aggregated nonnative proteins called *inclusion bodies*, which are disordered aggregates, leading to serious limitation in the production of recombinant proteins. It is a real problem needing a lot of effort to fully exploit the sequence information contained in the genome projects. Ordered aggregates resulting in amyloid fibrils lead to a number of serious human diseases such as Alzheimer's disease and the transmissible spongiform encephalopathies. The formation of amyloid aggregates has also been reported in *in vitro* experiments.

Experimental and theoretical studies together have provided significant insights into the mechanisms of protein folding, also allowing a better understanding of the aggregation processes.

The following different aspects of protein aggregation must be considered:

1. Theoretical and methodological aspects of protein folding, misfolding, and aggregation including the detection of aggregates and the mechanisms of aggregation processes.
2. Protein aggregation in the cellular environment including the folding into the cell, the role of molecular chaperones, and the formation of different aggregate morphologies, as well as the pathological consequences.

2 Protein Folding, Misfolding, and Aggregation

2.1 The New View of Protein Folding

The question of the mechanisms of protein folding has intrigued scientists for many decades. As early as the 1930s, attempts to refold denatured proteins were published, but significant progress began to be made when Anfinsen successfully refolded, denatured, and reduced ribonuclease into the fully active enzyme. In 1973, he stated the fundamental principle of protein folding referred to as the Anfinsen postulate: ‘‘all the information necessary

to achieve the native conformation of a protein in a given environment is contained in its amino acid sequence.'' The thermodynamic control of protein folding was considered to be a corollary of the Anfinsen postulate, meaning that the native structure is at a minimum of the Gibbs free energy. This statement was discussed by Levinthal in a consideration of the short time required for the folding process *in vitro* as well as *in vivo*. It was concluded that a random search of the native conformation among all possible ones would require an astronomic time and is therefore unrealistic. Thus, it is clear that evolution has found an effective solution to this combinatorial problem. This is referred to as the Levinthal paradox and has dominated discussions for the last three decades.

In order to understand how the polypeptide chain could overcome the Levinthal paradox, different folding models were proposed and submitted to experimental tests. Kinetic studies were carried out to follow the folding pathway. A considerable number of experiments were performed to detect and characterize the folding intermediates. A stepwise sequential and hierarchical folding process in which several stretches of structure are formed and assembled at different levels following a unique route was supported by a majority of scientists for many years. According to this view, misfolded species could be formed from folding intermediates leading to the formation of aggregates in a kinetic competition with the correct folding.

Progressively, with the development of computers, theoretical studies have approached the folding problem, using simplified models to take into account the computational limitations in simulations of the folding from the random coil to the native structure. Different methods were developed using either lattice models or molecular dynamics simulations. In the lattice model, the polypeptide chain is represented as a string of beads on a two-dimensional square lattice or on a three-dimensional cubic lattice. The interactions between residues (the beads) provide the energy function for Monte Carlo simulations. In such simplified models, the essential features of proteins, that is, the heterogeneous character (hydrophobic or polar) of the interactions and the existence of long-range interactions, were included to explore the general characteristics of the possible folds. Lattice models were first applied to protein folding by Go and coworkers while simple exact models were initiated by Dill and his group, and have been used by several theoreticians. From the lattice simulations, insights into possible folding scenarios have been obtained, providing a basis for exploring the general characteristics of folding for real proteins. The exploration of such models supplies useful information that can be submitted to experimental tests.

The so-called ''new view'' has evolved during the past 10 years from both experiment and theory with the use of simplified models. It is illustrated by the metaphor of the folding funnel introduced in 1995 by Wolynes and coworkers. The model is represented in terms of an energy landscape and describes the thermodynamic and kinetic behavior of the transformation of an ensemble of unfolded molecules to a predominantly native state as illustrated in Fig. 1. According to this model, there are several micropathways, each individual polypeptide chain following its own route. Toward the bottom of the funnel, the number of protein conformations decreases as does the protein entropy. The steeper the slope, the faster the folding. As written by Wolynes et al., ''To fold, a protein navigates with remarkable ease

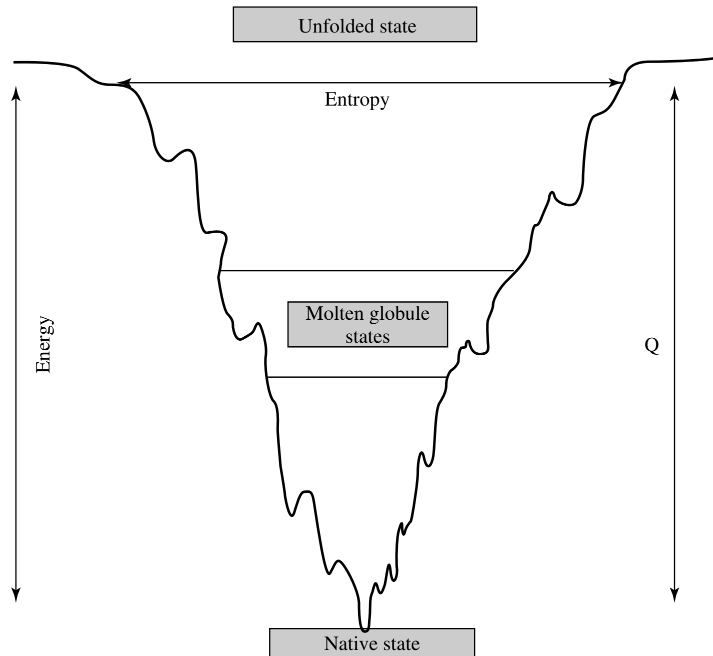


Fig. 1 Schematic representation of the folding funnel. Q is the number of native interactions.

through a complicated energy landscape.'' Thus, a wide variety of folding behaviors emerge from the energy landscape, depending on the energetic parameters and conditions. The folding rate could be slowed by ripples in the energy landscape corresponding to local minima populated by transiently stable intermediates. In a rugged energy landscape with kinetic traps formed by energy barriers, the folding will be even slower. When local energy barriers are high enough, protein molecules could be trapped and possibly aggregate.

The new view has progressively replaced the classical one of a unique sequential pathway and is now quite generally accepted. It is similar to the jigsaw puzzle model proposed in 1986 by Harrison and Durbin, suggesting the possibility of multiple folding routes to reach a unique solution. Many experimental results are consistent with this view. There is an increasing amount of evidence showing that the extended polypeptide chain folds through a heterogeneous population of partially folded intermediates in fluctuating equilibrium. Several alternative folding pathways have been observed for different proteins. From the convergence of theoretical and experimental studies, a unified view of the folding process has progressively emerged, also providing an explanation for the aggregation processes.

2.2 **Detection of Aggregates during the Refolding Process**

2.2.1 **Transient Aggregation**

Several observations indicate that transient aggregation could occur during *in vitro* protein refolding. Direct evidence for the

transient association of intermediates has been obtained from small angle X-ray scattering, in the case of apomyoglobin by Doniach and his group, and in the case of carbonic anhydrase by Semisotnov and Kuwajima, and by Silow et al. During the refolding of phosphoglycerate kinase, rapidly transient multimeric species (dimers, trimers, and tetramers) yielding to the native monomeric protein have been detected by Pecorari et al. These species are not in equilibrium, but are formed rapidly and disappear in the slow folding step. Unlike classical aggregates, their distribution does not depend on protein concentration, and they are produced at concentrations as low as 0.05 μM. The distribution of the oligomers is completely established at the end of the fast refolding step. To take into account all these observations, a model, which is formally similar to a reaction of copolymerization between two types of monomers, has been proposed. In this model, the refolding of the protein produces two types of intermediate conformers that can associate with the same or the other type. In the latter case, the association cannot be extended further (Fig. 2). Transient multimeric species have also been observed during the refolding of the isolated N-terminal domain under conditions in which neither the whole native protein nor the folded isolated N-domain associate. However, they cannot transform to the native form in the absence of the interactions with the complementary

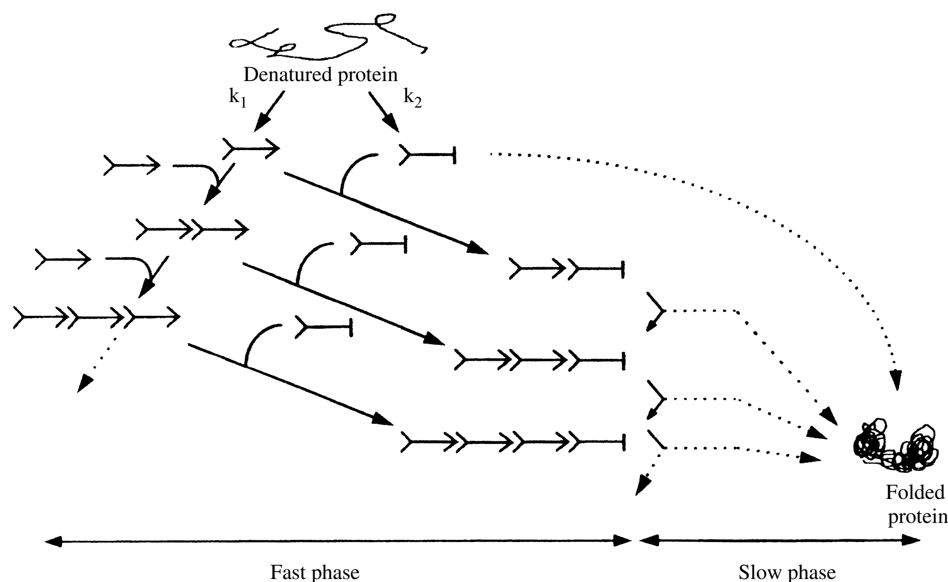


Fig. 2 Model proposed for the formation of transient multimeric species during the refolding of yeast phosphoglycerate kinase. Two types of conformers are produced in the early step of folding. One of these can be directionally extended by association with either the same conformer or another type of conformer. In this last case, the association cannot be further extended. The distribution of species results from a kinetic competition between two kinetic processes. (Reproduced from Pecorari, F., Minard, P., Desmadril, M.,Yon, J.M. (1996) Occurrence of transient multimeric species during the refolding of a monomeric protein, *J. Biol. Chem.* **271**, 5270–5276.)

domain indicating the importance of long-range interactions in directing the correct folding. Such species have not been observed with the C-terminal fragment. Thus, the occurrence of transient multimeric species arising from partially folded intermediates through hydrophobic interactions does not prevent the correct folding of a monomeric protein.

2.2.2 Irreversible Aggregation

Thermal unfolding of proteins is frequently accompanied by the formation of aggregates and therefore behaves as an irreversible process. It occurs at temperatures that vary widely according to the protein, since the temperature of optimum stability depends on the balance between hydrogen bonds and hydrophobic interactions. Generally, the products of thermal denaturation are not completely unfolded and retain some structured regions. At the end of the thermal transition, the addition of a denaturant such as urea or GdnHCl frequently induces further unfolding.

An apparent irreversibility at a critical concentration of denaturant has been observed during the refolding of monomeric as well as oligomeric proteins. It was reported for the first time by M.Goldberg and coworkers for the refolding of β-galactosidase, and for tryptophanase. It was also observed for a two-domain protein, horse muscle phosphoglycerate kinase by Yon and coworkers. In the latter study, when the enzyme activity was used as a conformational probe of the native structure, an irreversibility was observed for a critical concentration of denaturant equal to 0.7 M $\pm$ 0.1 M GdnHCl, a concentration very close to the end of the transition curve. Such irreversibility was found to be concentration dependent. For protein concentrations higher than 30 μM, restoration of enzyme activity was practically null. The formation of irreversible nonnative species was found to be temperature dependent; it was practically abolished at 4 °C, suggesting that aggregation occurs through hydrophobic interactions. The aggregation also depends on the time of exposure of the protein to the denaturant. When the unfolding–refolding process was observed using structural signals such as fluorescence or circular dichroism, it appeared completely reversible whatever the final denaturant concentration.

Another example is provided by rhodanese, a two-domain monomeric protein. During refolding at low denaturant concentration, an intermediate accumulates with partially structured domains and apolar surfaces exposed to the solvent, leading to the formation of aggregates. The aggregation can be prevented by refolding the protein in the presence of lauryl maltoside.

Most of the examples discussed above are related to multidomain proteins. Another degree of complexity appears in the folding of oligomeric proteins. It is generally accepted that the early steps of the process are practically identical to the folding of monomeric proteins. In the last step, subunit association and subsequent conformational readjustments yield the native and functional oligomeric protein. The correct recognition of subunit interfaces is required to achieve the process. The overall process of the folding of oligomeric proteins was extensively studied by Jaenicke and his coworkers for several enzymes and described in reviews. As with monomeric proteins, the formation of aggregates is concentration dependent. The kinetics of aggregation are complex and multiphasic, indicating that several rate-limiting reactions are involved in the process. In an attempt to characterize these aggregates, it was shown that noncovalent interactions occur between monomeric species with

partially restored secondary structures. The aggregates formed by either heat or pH denaturation can be disrupted in 6 M GdnHCl into monomeric unfolded species and then renatured under optimal conditions to yield an active enzyme. Only strong denaturants such as high concentrations of guanidine hydrochloride are efficient in this disruption process.

The presence of covalent cross-links such as disulfide bridges in a protein molecule can complicate the refolding of the denatured and reduced protein resulting in the formation of incorrect and intramolecular disulfide bridges leading to further aggregation. The first well-documented studies were performed by Anfinsen and his group on the refolding of reduced ribonuclease. The authors showed that the reoxidation of the enzyme produces a great number of species with incorrectly paired disulfide bonds. This scrambled ribonuclease is capable of regaining its native structure in a slow step, a process that is accelerated by the addition of a small quantity of reducing reagent such as β-mercaptoethanol yielding about 100% of active enzyme. The reshuffling of a protein's disulfide bonds takes place through a series of redox equilibria according to either an intramolecular or an intermolecular exchange. To prevent a wrong pairing of half-cystine and further aggregation, the addition of small amounts of reducing reagents or redox mixture is frequently used as investigated by Wetlaufer.

The detection and characterization of aggregates represent an important aspect of folding studies. The aggregation phenomenon can occur without precipitation. Indeed, the degree of association of protein intermediates during folding might be small, depending on the intermolecular interactions, and does not necessarily lead to a visible insolubility. The association state may be determined in several ways. The most common methods, available in any biochemistry laboratory, are gel permeation and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), used both with and without cross-linking. The detection of aggregates can also be monitored by other hydrodynamic methods such as analytical ultracentrifugation or classical light scattering. The latter method also gives information on the size of the aggregates. Quasi-elastic light scattering is a dynamic technique that can be used to determine macromolecule diffusion coefficients as a function of time, that is, to follow the kinetics of aggregation. Neutron scattering can also be used to detect protein aggregates, and mass spectrometry has become a useful tool as well.

2.3 Mechanisms of Protein Aggregation

A substantial body of information supports the idea that protein aggregation arises from partially folded intermediates through hydrophobic interactions. The formation of aggregates has often been considered as a trivial phenomenon, a nonspecific association of partially folded polypeptide chains to form a disordered precipitate. However, several analyses indicate that aggregation occurs by specific intramolecular associations involving the recognition of a sequence partner in another molecule rather than in the same molecule during the folding process. Analyses of the aggregation mechanisms of various proteins, such as bovine growth hormone and phosphoglycerate kinase, has permitted the identification of specific sites that are critical in the association.

An elegant demonstration of the specificity of aggregation was provided by King

and coworkers. The authors showed that during the *in vitro* refolding of a mixture of two proteins, tailspike endorhamnidase and coat protein from phage P22, no heterogeneous aggregates were formed. Tailspike endorhamnidase is a thermostable trimer whose folding intermediates are thermolabile and either undergo productive folding or form multimeric aggregates (Fig. 3). The P22 coat protein, which comprises the capsid shell of phage P22, yields either a correct fold or "off-pathway" aggregates upon refolding. Both proteins were intensively studied by King and coworkers who first denatured the two proteins in urea and then chose refolding conditions such that aggregation competes with correct folding. Folding and soluble aggregates of the two proteins were characterized either separately or mixed together. No heterogeneous aggregates were found, clearly indicating that only self-association of transient refolding molecules occurs in the formation of soluble multimers.

One mechanism that accounts for the formation of aggregates during refolding of multidomain proteins is domain swapping. This was first suggested by Monod and later proposed by Goldberg and colleagues to account for the formation of aggregates during the refolding of tryptophanase. The concept was foreshadowed by the results of Crestfield and coworkers in 1962. From their experiments based on chemical modification of bovine pancreatic ribonuclease, the authors proposed that the dimer is formed by exchanging the N-terminal fragments. The term *3D domain swapping* was introduced in 1994 by Bennett and coworkers to describe the structure of a diphtheria toxin dimer. The mechanism involves the replacement of one domain of a monomeric protein by the same domain of an identical neighboring

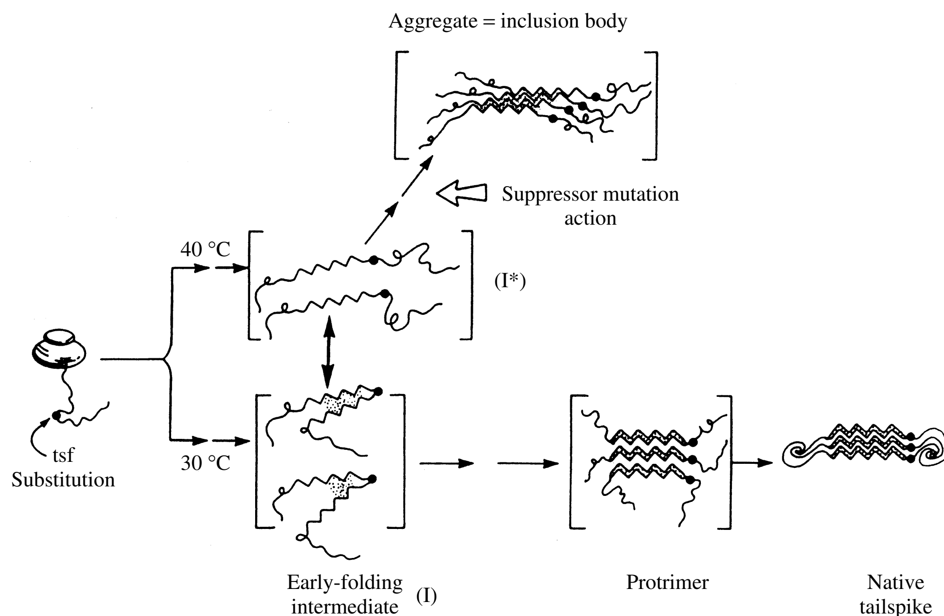


Fig. 3 The folding pathway of the P22 tailspike protein. (From Mitraki, A., King, J. (1992) *FEBS Lett.* **307**, 20–25; reproduced with permission.)

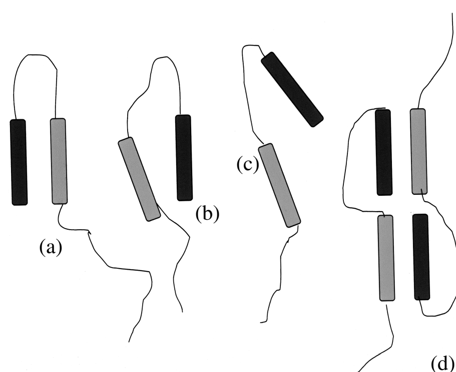


Fig. 4 Schematic representation of domain swapping. (a) monomeric protein, (b) and (c) partially unfolded monomers, (d) domain-swapped dimer.

molecule, thus resulting in an intertwined dimer or oligomer, as defined by Eisenberg and colleagues (Fig. 4). When the exchange is reciprocated, domain-swapped dimers are formed. However, if the exchange is not reciprocated but propagated along multiple polypeptide chains, higher order assemblies or aggregates may form. Domain-swapped oligomers are divided into two types, open and closed. The open oligomers are linear and have one closed interface (closed in the monomer) exposed to the solvent, whereas closed oligomers are cyclic and do not expose a closed interface. Eisenberg and coworkers have defined the structure of the monomer as the "closed monomer" and the conformation of the polypeptide chain in the domain-swapped oligomer as the "open monomer."

The ability of monomeric proteins to swap structural elements requires the presence of a hinge or linker region that permits the protein to attain the native fold with parts of two polypeptide chains. In fact, domain-swapped structures reveal regions of protein structure that are flexible. Bergdoll and coworkers have suggested that a proline in the linker region, by rigidifying the hinge region in intermediate states, might facilitate domain swapping. Baker and colleagues proposed that strain in a hairpin loop might predispose a protein to domain swapping. The possible role of 3D domain swapping in the evolution of oligomeric proteins has been discussed in several reviews. In the past years, the number of known domain-swapped proteins has increased and today about 40 such structures are solved. One common feature of these proteins is that all the swapped domains are from either the N-terminus or the C-terminus of the polypeptide chain. In this regard, an interesting example arises from the work of Eisenberg and his group on the dimerization of ribonuclease A. This protein forms two types of dimers upon concentration in mild acid. The minor dimer is formed by swapping of its N-terminal α-helix with that of an identical molecule. The major dimer results from the swapping of its C-terminal β-strand. RNase A was also reported to form trimers. On the basis of the structure of the N- and C-terminal swapped dimers, a model was proposed (Fig. 5). This indicates that two types of swapping can occur simultaneously in the same oligomer. Further biochemical studies have supported this model. A less abundant trimer in which only the C-terminal β-strand is swapped and exhibits a cyclic structure was also found. RNase represents the

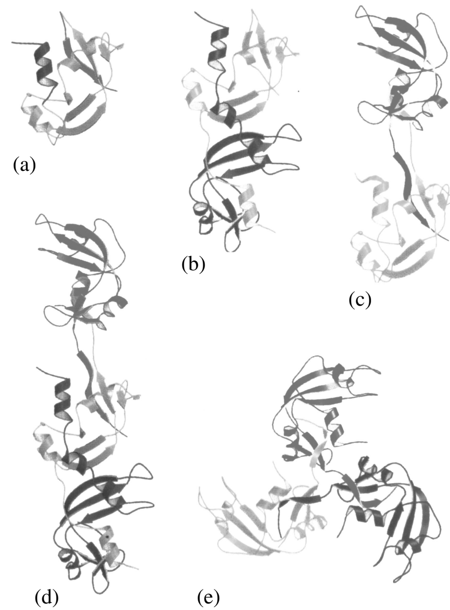


Fig. 5 Domain swapping in ribonuclease. Ribbon diagram of the structures of (a) the ribonuclease A monomer (2.0 Å), (b) the N-terminal swapped dimer (2.1 Å), (c) the C-terminal swapped dimer (1.75 Å), (d) the N- and C-terminal trimer model, and (e) the cyclic C-terminal swapped trimer (2.2 Å) (reproduced from Liu et al. Prot. Sci. **11**, 371, 2002 with permission).

first protein found to form both linear and cyclic domain-swapped oligomers. This protein also was described to form tetramers. Models based on the structures of dimers and trimers were proposed for these tetramers. Two linear models exhibit both types of swapping that occur in one molecule, and a cyclic tetramer shows the swapping of the C-terminal β-strand only. A trimeric domain-swapped barnase was obtained at low pH and high protein concentration. Crystallographic studies revealed a structure suggesting a probable folding intermediate. Domain swapping was described for the cell cycle regulatory protein p13suc1, a small protein of 113 amino acids.

Folding studies as well as molecular dynamics simulations have shown that domain swapping occurs in the unfolded state. Eisenberg and his colleagues have proposed a free energy diagram for the pathway of domain swapping. The free energy difference between the closed monomer and domain-swapped oligomer is small since they share the same structures except at the hinge loop, but the energy barrier can be reduced under certain conditions making domain swapping more favorable. Several molecular or environmental events may favor the formation of extended domain-swapped polymers. Genetic mutations introducing a deletion in the hinge loop can destabilize the monomeric form of a protein. The replacement of only one amino acid can also favor the polymerization of the mutated protein. Three-dimensional domain-swapped oligomers are expected to be increasingly favored as the protein concentration increases. Thus, a metabolic change that increases the concentration of a protein will favor aggregation. Charge effects, caused either by mutations or by pH change or salt concentration, can induce domain swapping; for example, in RNase A, a decrease in pH, by protonating the residues involved in hydrogen bonds and in salt bridges, lowers the energy barrier of the formation of the open monomer, hence inducing domain swapping.

There is great diversity of swapped domains, with different sizes and sequences. They can consist of entire tertiary domains or smaller structural elements made of several residues. No specific sequence motif seems to be involved among the swapped domains. Three-dimensional domain swapping has also been proposed as a mechanism for amyloid formation. This aspect will be discussed in Sect. 4.2.

As can be seen here, several mechanisms exist, which lead to the formation of aggregates. It is recognized that aggregation results from the association of incompletely or incorrectly folded intermediates through hydrophobic interactions. In the energy landscape of protein folding, the presence of local minima separated by an energy barrier allows the accumulation of intermediates. If the barrier is high enough, these intermediates cannot easily reach the native state, and kinetic competition thus favors the formation of aggregates.

3
Protein Folding in the Cellular Environment

3.1
Molecular Crowding in the Cells

The main rules of protein folding have been deduced from a considerable body of *in vitro* and *in silico* studies. It has been accepted that the same mechanisms are involved in *in vitro* refolding and in the folding of a nascent polypeptide chain in the cell. However, the intracellular environment differs markedly from that of the test tube where low protein concentrations are used. The interior of a cell is highly crowded with macromolecules. The concentration is so high that a significant proportion of the volume is occupied. As mentioned by Ellis, in general, 20 to 30% volume of the interior of the cells are occupied by macromolecules; for example, the concentration of total protein inside cells ranges from 200 to 300 g L^{-1}. The total concentration of proteins and RNA inside *Escherichia coli* ranges from 300 to 400 g L^{-1} depending on the growth phase. Polysaccharides also contribute to the crowding. It can be predicted practically that diffusion coefficients will be reduced by factors up to 10-fold due to crowding. Since the average time for a molecule to move a certain distance varies by D^{-2}, D being the diffusion coefficient, it will take 100 times longer to move this distance in the cell as would be necessary under low concentration conditions. Another prediction indicates that equilibrium constants for macromolecular associations may be increased by two to three orders of magnitude.

Molecular crowding inside cells also has consequences for protein folding, favoring the association of partly folded polypeptide chains into aggregates. This could explain why cells contain molecular chaperones, even though most denatured proteins refold spontaneously in the test tube.

3.2
The Role of Molecular Chaperones

The discovery of a ubiquitous class of proteins mediating the correct folding in cellular environment has led to a reconsideration of the mechanism of protein folding *in vivo*. Historically, the term *molecular chaperone* was introduced by Laskyard and coworkers in 1987 to describe the function of nucleoplasmin, which mediates the *in vitro* assembly of nucleosomes from separated histones and DNA. The concept was further extended by Ellis to define a class of proteins whose function is to ensure the correct folding and assembly of proteins through a transient association with the nascent polypeptide chain. Studies on heat-shock proteins have widely contributed to the development of this concept.

Today, more than 20 protein families have been identified as molecular chaperones. Molecular chaperones comprise

several highly conserved families of related proteins. They can be divided into two classes according to their size. Small chaperones are less than 200 kDa, whereas large chaperones are more than 800 kDa. During the past few years, a large amount of biochemical, biophysical, and low- and high-resolution structural data have provided mechanistic insights into the machinery of protein folding as assisted by molecular chaperones.

Molecular chaperones are involved in diverse cellular functions. The constitutive members of the heat-shock protein family (Hsp70) can stabilize nascent polypeptide chains during their elongation in ribosomes. The large cylindrical chaperonins GroEL in bacteria, mitochondria, and chloroplasts and the corresponding TriC in eukaryotes and archaebacteria provide a sequestered environment for productive folding. Several chaperones are stress-dependent; their expression is induced under conditions such as high temperatures, which provoke protein unfolding and aggregation. The members of the Hsp90 and Hsp100 families, as well as small Hsp, play a role in preventing protein aggregation under stress. Chaperone interactions are also important for the translocation of polypeptide chains into membranes.

Within cells, the nascent polypeptide chain is synthesized sequentially on the ribosome by a vectorial process. For many proteins, the rate of this process is slower than the rate of folding. Synthesis times range from 20 s for a 400 residue–polypeptide chain in *E. coli* at 37 °C to 10 times as long for such a chain in an eukaryotic cell. Many unfolded proteins refold completely in 20 s under the same conditions. Thus, there is the possibility for the elongating polypeptide either to misfold before completion or to be degraded by proteolytic enzymes. Chaperones prevent such unfavorable events by protecting the nascent chain. Hsp70 and its prokaryotic homolog DnaK recognize extended hydrophobic regions of the elongating polypeptides. These interactions are not specific. Hsp70 and DnaK interact with most unfolded polypeptide chains that expose hydrophobic residues. They do not recognize folded proteins. Binding and release of unfolded proteins from Hsp70 are ATP-dependent and require the presence of various cochaperones such as DnaJ and GrpE. The basic mechanism of Hsp70 (DnaK in *E. coli*) is represented in Fig. 6. In *E. coli*, DnaJ binds the nascent unfolded polypeptide, U; then the complex binds to the ATP-bound state of DnaK. ATP is hydrolyzed in the ternary complex

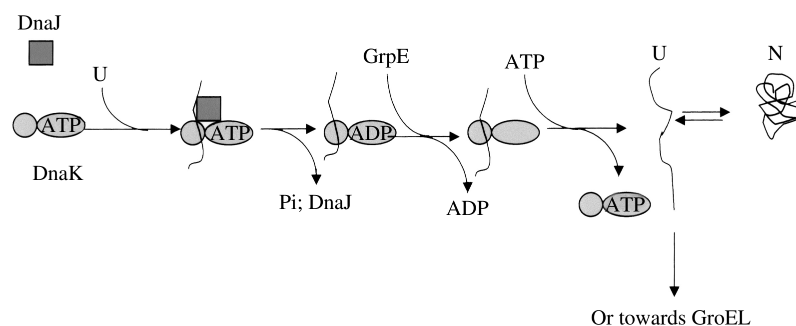


Fig. 6 Schematic representation of the basic mechanism of DnaK (see text).

allowing the release of DnaJ and P_i. In the following step, GrpE acts as an exchange factor to regenerate the ATP-bound state of DnaK. The unfolded polypeptide chain is released into the bulk solution. Thus, Hsp70 systems bind and release the polypeptide in an unfolded conformation. The unfolded protein has the possibility either to fold or to be transferred to the GroEL system, as illustrated in Fig. 6. Significant insights into this mechanism were obtained from structural data. The three-dimensional structures of Hsp70 and DnaJ as well as those of a complex DnaK–polypeptide and a complex of GrpE with the ATP binding domain of DnaK are known. DnaK and its homologs are composed of two domains, a C-terminal domain that binds ATP and an N-terminal domain that binds peptides. GrpE is a tight homodimer associated along two long helices. It binds DnaK–ATPase domain through its proximal monomer. DnaJ activates the ATP hydrolysis by DnaK. It was shown that a conformational change may occur upon ATP binding, opening the polypeptide binding cleft in the polypeptide binding domain of DnaK. The closed state may correspond to the ADP-bound conformation. The ADP-bound state of DnaK binds the peptide tightly. Peptide release requires the dissociation of ADP, which is mediated by GrpE. DnaK then rebinds ATP.

The GroEL–GroES system acts by a different mechanism in which the unfolded protein is sequestered. The chaperonins are large cylindrical protein complexes. The crystal structure of *E. coli* chaperonin GroEL was determined in 1994 and that of the asymmetric GroEL–GroES–(ADP)7 complex in 1997 by Sigler and his group. GroEL consists of two heptameric rings of 58-kDa subunits stacked back to back with a dyad symmetry and forming a porous cylinder (Fig. 7). Each subunit is organized in three structural domains. A large equatorial domain forms the foundation of the assembly and holds the rings together. It contains the nucleotide binding site. A large apical domain forms the end of the cylinder. The apical domain contains a number of hydrophobic

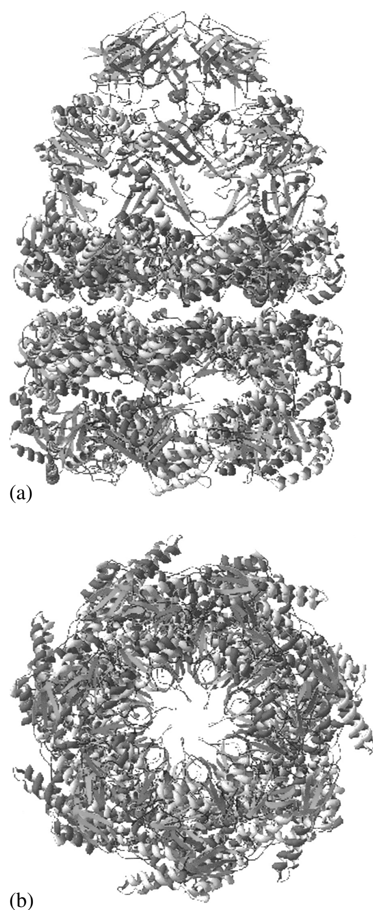


Fig. 7 Crystal structure of GroEL–GroES–(ADP)7 complex determined by Sigler et al. (a) view along the axis and (b) view from the top of the complex. (Reproduced from the PDB web site.)

residues exposed to the solvent. A small intermediate domain connects the two large domains. The intermediate segments have some flexibility allowing a hinge-like opening of the apical domains, which occurs upon nucleotide binding. These movements are large and have been visualized by three-dimensional reconstruction from cryoelectron microscopy by Sebil and her group.

GroES is a heptamer of 10 kDa subunits forming a flexible dome-shaped structure with an internal cavity large enough to accommodate proteins up to 70 kDa. Each subunit is folded into a single domain containing β-sheets and flexible loop regions. The loop regions are critical for the interactions between GroEL and GroES. It was deduced from electron microscopy studies that GroES binding to GroEL induces large movements in the apical GroEL domains. This provokes a significant increase in the volume of the central cavity in which protein folding proceeds. NMR coupled with the study of hydrogen-exchange techniques has indicated that small proteins are essentially unfolded in their GroEL-bound states. Mass spectroscopy has revealed the presence of fluctuating elements of secondary structure for several proteins. In a way, the GroEL–GroES system recognizes nonnative proteins.

The reaction cycle of the GroEL–GroES system is represented in Fig. 8. The nonnative protein binds to the apical domain of the upper ring of GroEL through hydrophobic interactions. Then, the equatorial domain of the same ring binds ATP, and GroES caps the upper ring, sequestering the protein inside the internal chamber in which the protein folding proceeds. The binding of GroES induces a conformational change in GroEL and ATP hydrolysis, which is a cooperative process that produces a conformational change in the lower ring, allowing it to bind a nonnative protein molecule. This promotes subsequent binding of ATP and GroES in the lower ring, and the dissociation of the upper complex, releasing the protein and ejecting GroES. If the protein has not reached the native state, it is subjected to a new cycle.

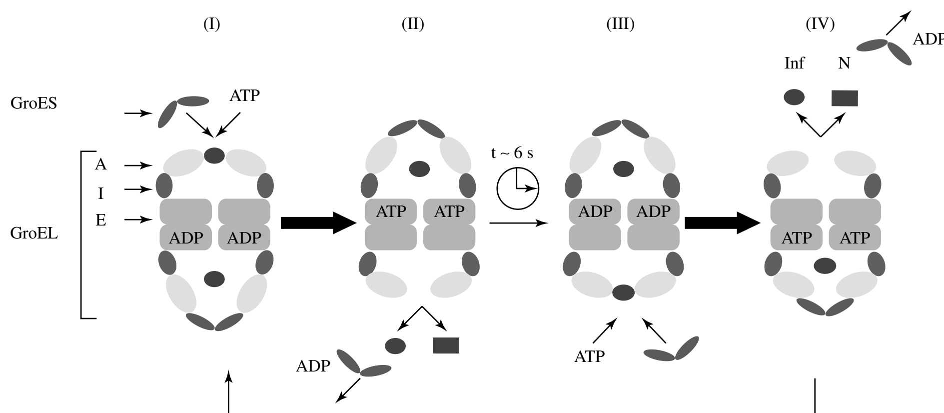


Fig. 8 The reaction cycle of GroEL–GroES. Inf is the unfolded protein, N the folded one, A is the apical domain, (in blue), I the intermediate domain (in red) and E the equatorial domain (in magenta). (Reproduced from Wang & Weissman (1999) *Nat. Struct. Biol.* **6**, 597, with permission.) (See color plate p. xxii).

The hydrolysis of ATP by GroEL is used only to induce conformational changes of the chaperone, which permits the release of the folded protein. The molecular chaperones, by their transient association through hydrophobic interactions with nascent, stress-destabilized, or translocated proteins, have a role in preventing improper folding and subsequent aggregation. They do not interact with folded proteins. They do not carry information capable of directing a protein to assume a structure different from that dictated by its amino acid sequence. Therefore, molecular chaperones assist the folding in the cells without violation of the Anfinsen postulate. They increase the yield but not the rate of folding reactions; in this respect they do not act as catalysts. Furthermore, the majority of newly synthesized polypeptide chains in both bacterial and eukaryotic cells fold spontaneously without the assistance of molecular chaperones.

Many proteins from prokaryotic and eukaryotic organisms are produced with an amino-terminal propeptide, which is removed by limited proteolysis during the activation process. Several of these propeptides consist of a long polypeptide chain; for example, there are 174 amino acids in the propeptide of pro-α-lytic protease, 91 in that of procarboxypeptidase Y, and 77 in that of prosubtilisin. Several studies have shown that the propeptide is required for proper folding of these proteins. The mature enzymes are not able to refold correctly. They seem to have kinetic stability only, whereas the proenzymes have thermodynamic stability. Since propeptides perform the function of mediating protein folding, they have been classified as intramolecular chaperones. However, this terminology is not appropriate since the nascent protein is the proenzyme, not the enzyme that has undergone proteolytic cleavage. Thus, it is not surprising that the proenzyme refolds spontaneously, whereas the mature protein does not. Indeed, the information is contained in the totality of the proenzyme sequence.

Two other classes of proteins play the role of helpers during protein folding *in vivo*: protein disulfide isomerases (PDIs) and peptidyl–prolyl cis – trans isomerases. Protein disulfide isomerase is an abundant component of the lumen of the endoplasmic reticulum in secretory cells. The enzyme was discovered independently in 1963 by two research groups: in rat and ox by Anfinsen and coworkers, and in chicken and pigeon pancreas by Straub and coworkers. Proteins destined to be secreted enter the endoplasmic reticulum in an unfolded state. In this environment, the folding process is associated with the formation of disulfide bonds, which is catalyzed by PDI through thiol–disulfide interchange. The first PDI cDNA was sequenced in 1985 by Edman et al. It displays sequence homologies implying a multidomain architecture. PDI consists of four structural domains arranged in the order a, b, b′, a′, with the b′ and a′ domains being connected by a linker region. Furthermore, it possesses an acidic C-terminal extension. The a and a′ domains contain the active site motif – W-C-G-H-C-. They display significant sequence identity to thioredoxin, a small cytoplasmic protein involved in several redox functions, and they have a similar active site sequence.

Recombinants of the a and b domains have been obtained and studied by high-resolution NMR. The a domain has the same overall fold as thioredoxin, an α/β fold with a central core made up of a five-stranded β-sheet surrounded by four helices. As in thioredoxin, the active site is located at the N-terminus of helix

α_2. Preliminary NMR data of the a′ domain confirm its structural similarity to the a domain. The b and b′ domains have significant sequence similarity to each other, but no similarity with the a domain. Nevertheless, NMR studies of the b domain have indicated a similar overall fold. From its sequence, it could be inferred that b′ also has the same fold. Neither b nor b′ contain the active site.

The folding pathway of disulfide-bound proteins involves isomerizations between a number of species containing disulfide bonds. *In vitro* experimental studies were performed using the isolated a and a′ domains, and the results were compared with those obtained with the holoenzyme. It was concluded that the activity of long length PDI is not simply the sum of the activities of the isolated a and a′ domains. Using a series of constructs including nearly every linear combination of domains, the contribution of each domain was investigated. It was determined that the thiol-disulfide chemistry requires only the a and a′ domains, and that simple isomerization requires one of these in a linear combination including b′, whereas complex isomerization involving large conformational changes requires all the PDI domains except the C-terminal extension. Thus, it appears that the b′ domain is the principal peptide binding site, but all domains contribute to the binding of larger polypeptide chains holding them in a partially unfolded conformation while the catalytic sites acts synergistically to perform the thiol-disulfide exchange. Since PDI has binding properties, it has been proposed that it acts as a molecular chaperone. However, as underlined by Freedman and coworkers, this property does not represent a chaperone activity and instead reflects its role as a catalyst to accelerate the formation of native disulfide bridges during protein folding.

Several gene products with similarity to PDI have been identified in higher eukaryotes. All are probably localized in the endoplasmic reticulum and have thiol-disulfide exchange activity.

In prokaryotes, the disulfide formation occurs in the periplasm and is catalyzed by a protein called DsbA, which exchanges its Cys30–Cys33 to a pair of thiols in the target protein, leaving DsbA in its reduced state. The crystal structure of oxidized DsbA displays a domain with a thioredoxin-like fold and another domain, which caps the thioredoxin-like active site C30-P31-H32-C33, located at the domain interface. Reoxidation of DsbA is catalyzed by a cytoplasmic membrane protein called DsbB, which contains four cysteine residues essential for catalysis. DsbB transfers the electrons from the reduced DsbA to membrane embedded quinones. The reduced quinones are then oxidized enzymatically either aerobically or anaerobically. Thus, DsbA is found in normal cells in its oxidized state. *E. coli* also has a complex reductive system including another periplasmic protein DsbC, which is a homodimer. The molecule consists of two thioredoxin-like domains with a CxxC motif, joined via hinged linker helices to an N-terminal dimerization domain. The hinge regions allow movement of the active site, and a broad hydrophobic cleft between the two domains may bind the polypeptide chain. Its function consists of reducing proteins with incorrect disulfide bonds. DsbC is maintained in its reduced form by a membrane protein called DsbD, which contains six essential cysteine residues. Then, the electrons are transferred to thioredoxin and ultimately to NADPH by thioredoxin reductase.

All these enzymes, which catalyze the pairing of cysteine residues in disulfide-bridged proteins, have functional domains pertaining to the thioredoxin superstructure.

Another type of enzyme, peptidyl–prolyl cis–trans isomerases, facilitates the folding of some proteins by catalyzing the cis–trans isomerization of X-Pro peptide bonds. Two classes of unrelated proteins demonstrate this activity, those that bind cyclosporin, which are known as *cyclophilins*, and those that bind FK506. The cellular function of these enzymes is important, since cyclosporin and FK506 are potent immunosuppressors that regulate T-cells activation. Both classes of peptidyl–prolyl isomerases are ubiquitous, and abundant in prokaryotes and eukaryotes. The sequences of several members of each family are known, and the three-dimensional structures of at least one member of each family have been elucidated by X-ray crystallography and multidimensional NMR. Their role is to accelerate the cis-trans isomerization of X-pro peptide bonds when this process is the rate-limiting step in protein folding. Although they do not present structural similarity, both exhibit a hydrophobic binding cleft favoring the rotamase activity by excluding water molecules.

4
Protein Aggregation in the Cellular Environment

4.1
The Formation of Inclusion Bodies

The overexpression of genes introduced in foreign hosts frequently results in aggregated nonnative proteins called *inclusion bodies*. In cells, inclusion bodies appear as unordered amorphous aggregates clearly separated from the rest of the cytoplasm; they form a highly refractive area when observed microscopically. A great variety of experimental studies indicates that the formation of inclusion bodies results from partially folded intermediates in the intracellular folding pathway and not from either totally unfolded or native proteins.

4.1.1 Occurrence of Inclusion Bodies

Inclusion bodies were first identified in the blood cells of patients with abnormal hemoglobins, the resulting pathology being anemia. Pathological point mutants of hemoglobin aggregate into inclusion bodies; this is the case for hemoglobin Köln (Val98Met on the β chain) and hemoglobin Sabine (Leu91Pro on the β chain). Similar deposits have been described in studies on the metabolism of abnormal proteins subjected to covalent modification in *E. coli*. The formation of aggregates also occurs when cells are subjected to heat shock.

The *in vivo* folding pathway of tailspike endorhamnosidase of Salmonella phage 22 is a well-documented system studied by J.King's group. Furthermore, it is one of the few systems in which the *in vivo* folding pathway has been compared with the *in vitro* refolding pathway. The protein is a trimer of 666 amino acids. The secondary structure is predominantly β-sheet. Newly synthesized polypeptide chains released from the ribosome generate an early partially folded intermediate. This intermediate further evolves into a species sufficiently structured for chain–chain recognition. In the following step, an incompletely folded trimer is formed upon close association with the latter species. The protrimer is then transformed into the native tailspike. A clear difference between the physicochemical properties of the intermediates and the native state has

allowed their identification. Figure 3 illustrates the folding pathway of the protein. The native protein is highly thermostable with a T_m of 88 °C; it is also resistant to detergents and proteases. During the *in vivo* folding process, the intermediates are sensitive to these factors, allowing their identification. At low temperature, almost 100% of the newly synthesized chains reach the native trimer conformation. When the temperature increases in the cells, the number of polypeptide chains achieving the native state decreases. At 39 °C, the maturation proceeds with 30% efficiency, while the remainder aggregates into inclusion bodies. It has been shown that the aggregation does not result from an intracellular denaturation of the native protein, but is generated from an early thermolabile intermediate. The aggregated chains cannot recover their proper folding by lowering the temperature. But when polypeptide chains that have been synthesized at high temperatures are shifted to low temperature early enough, they can refold correctly.

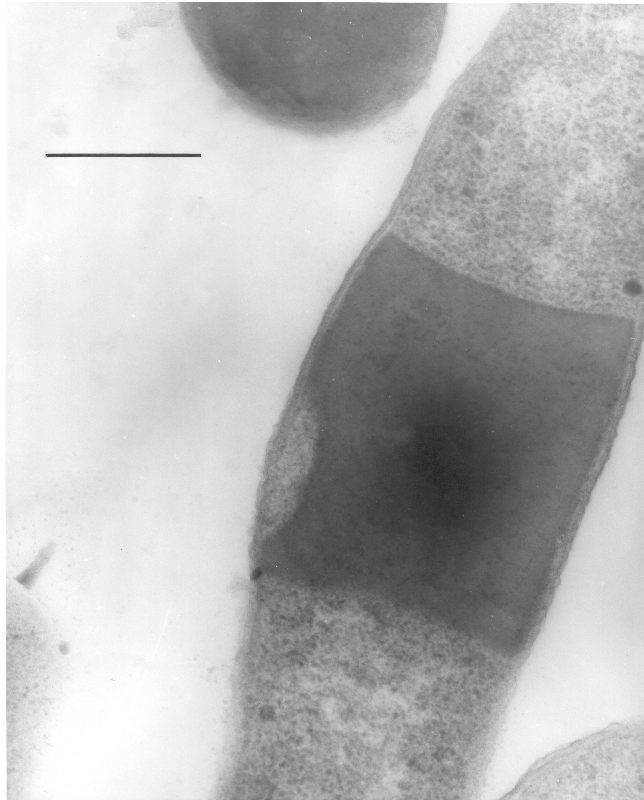
A set of mutations that alter protein folding without modifying the properties and stability of native P22 tailspike has been identified; they are referred to as temperature-sensitive folding (tsf) mutants. These mutations have been supposed to destabilize the already thermolabile intermediate and are located at more than 30 sites in the central region of the polypeptide chain. Starting from mutants kinetically blocked in their folding, a second set of mutants capable of correcting the folding defects was selected, and the sequences surrounding the suppressor mutations were identified. Only two substitution positions on the 666 amino acids of the polypeptide chain were sufficient to prevent inclusion body formation. Thus, single temperature mutations that affect the folding pathway but not the native conformation of a protein are efficient in preventing off-pathway and subsequent aggregation. A similar result has been found for heterodimeric luciferase. For recombinant proteins such as interferon-γ and interleukin 1β, as well as for P22 tailspike, amino acid substitutions that can decrease or increase the formation of inclusion bodies without alteration of the functional structure were found by Wetzel and coworkers.

The formation of inclusion bodies is frequently observed in the production of recombinant proteins. High levels of expression of these proteins result in the formation of inactive amorphous aggregates, and has been reported for proteins expressed in *E. coli* and also in several host cells, gram-negative as well as gram-positive bacteria, and eukaryotic cells such as *Saccharomyces cerevisiae*, insect cells, and even animal cells. The production of recombinant proteins, among them human insulin, interferon-γ, interleukin 1β, β-lactamase, prochymosin, tissue plasminogen activator, basic fibroblast growth hormone, and somatotropin, gives rise to inclusion bodies.

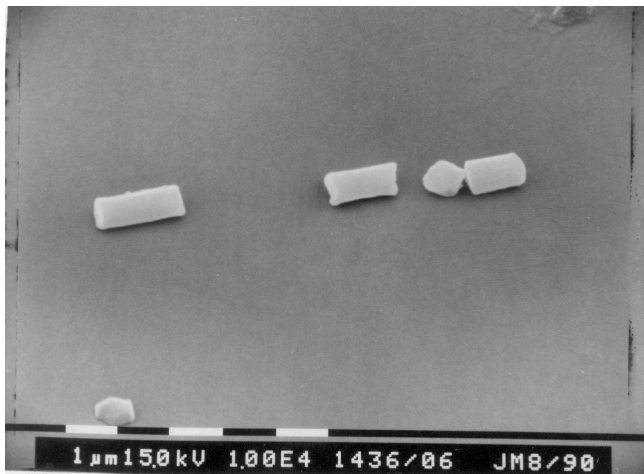
4.1.2 **Characteristics of Inclusion Bodies**

Inclusion bodies can form in the cytoplasm and in the periplasmic space of *E. coli*. Wild-type β-lactamase expressed in *E. coli* results in the formation of inclusion bodies in the periplasm, whereas the protein expressed without its signal sequence aggregates in the cytoplasm.

The characteristics of the aggregates depend on how the protein is expressed. Different sizes and morphologies have been observed. Generally, inclusion bodies appear as dense isomorphous aggregates of nonnative proteins separated from the rest of the cytoplasm, but not surrounded



(a)



(b)

Fig. 9 Electron micrographs of (a) cytoplasmic β-lactamase inclusion bodies in *E. coli* RB791(pGB1) and (b) purified inclusion bodies from the same origin (courtesy of G.A. Bowden, A.M. Paredes & G. Georgiou).

by a membrane (Fig. 9). They look like refractile inclusions, which can be easily recognized by phase contrast microscopy when large enough. For prochymosin expressed in *E. coli*, the lack of birefringence indicates that inclusion bodies are not crystalline. The size distribution of inclusion bodies has been studied for prochymosin and interferon-γ, and Marston reported the mean size of particles to be 0.81 and 1.28 μm respectively, with a relatively high void fraction. The void volume was about 70% of the total volume for interferon-γ and 85% for prochymosin. Structural characterization studies using ATR-FTIR (attenuated total reflectance Fourier transformed infrared spectroscopy) have shown that the insoluble nature of inclusion bodies may be due to their increased levels of nonnative intramolecular β-sheet content.

Inclusion bodies consist mostly of the overexpressed recombinant protein, and can contain little contaminating molecules. Thus, they can be used as a source of relatively pure misfolded protein when refolding yields the active protein. However, some amorphous bodies incorporate other molecules, for example, inclusion bodies from *E. coli* cells overexpressing β-lactamase contain only between 35 and 95% intact β-lactamase. The rest consists of a variety of intracellular proteins, some lipids, and a small amount of nucleic acids. Homogeneous inclusion bodies were obtained by expressing β-lactamase without its leader peptide. Under these conditions, aggregation occurs within the cytoplasm. The extent of incorporation of other macromolecules in inclusion bodies depends upon the overexpressed protein.

The formation of inclusion bodies generally appears to be a disadvantage, since it requires the dissolving of the aggregates in denaturant and subsequent refolding of the protein. However, when the recovery of the active product can be obtained with a sufficient yield, certain advantages may accrue. Indeed, aggregation generally prevents proteolytic attack, except when the protein coaggregates with a protease. The formation of inclusion bodies is also an advantage for the production of proteins that are toxic for the host cells. Furthermore, these aggregates contain a great quantity of the overexpressed protein.

4.1.3 Strategies for Refolding Inclusion Body Proteins

The recovery of the active protein from inclusion bodies is crucial for industrial purposes. In structural proteomics today, efficient production of genetically engineered proteins is a prerequisite for exploiting the information contained in the genome sequences. The strategy to recover active proteins involves several steps of purification. The first step, the separation of the inclusion bodies from the cell, consists of cell lysis monitored either by high-pressure homogeneization, or by a combination of mechanical, chemical, and enzymatic techniques such as the use of EDTA and lysozyme. The lysates are then treated by low-speed centrifugation or filtration to remove the soluble fraction from the pellet containing inclusion bodies and cell debris. The most difficult task is to remove the contaminants; this is achieved by the washing steps, which commonly utilize EDTA and low concentrations of denaturants or detergents such as Triton X-100, deoxycholate, or octylglucoside. Using centrifugation in a sucrose gradient, it is generally possible to remove cell debris and membrane proteins. When the accumulation levels of aggregates are very high, inclusion bodies may be directly solubilized by treatment in a high concentration of denaturant, eliminating the need for gradient centrifugation. In this

case, the costs of production are considerably reduced.

A variety of techniques are available to solubilize purified inclusion bodies. The most commonly used solubilizing reagents are strong denaturants such as guanidine hydrochloride and urea. Generally, high denaturant concentrations are employed, 4 to 6 M for guanidine hydrochloride, and 5 to 10 M for urea to allow the disruption of noncovalent intermolecular interactions. Conditions may differ somewhat according to the denaturant and the protein. Lower denaturant concentrations have been used to solubilize cytokines from *E. coli* inclusion bodies. The purity of the solubilized protein was much higher at 1.5 to 2 M guanidinium chloride than at 4 to 6 M guanidinium chloride. At higher denaturant concentrations, contaminating proteins were also released from the particulate fractions.

Extremes of pH have also been used to solubilize inclusion bodies and for growth hormone, proinsulin, and some antifungal recombinant peptides. However, exposure to very low or very high pH may not be applicable to many proteins and may cause irreversible chemical modifications.

Detergents such as sodium dodecylsulfate (SDS) and n-cetyl trimethylammonium bromide (CTAB), have also been used to solubilize inclusion bodies. Extensive washing may then be needed to remove the solubilizing detergents. They also may be extracted from the refolding mixture by using cyclodextrins, linear dextrins, or cycloamylose. Recent developments include the use of high hydrostatic pressure (1–2 kbar) for solubilization and renaturation. For proteins with disulfide bonds, the addition of a reducing reagent such as dithiothreitol or β-mercaptoethanol is necessary to disrupt the incorrectly paired disulfide bonds. The concentrations generally used are 0.1 M for dithiothreitol and 0.1 to 0.3 M for β-mercaptoethanol.

When expression levels are very high, an *in situ* solubilization method can be used. It consists of adding the solubilizing reagent directly to the cells at the end of the fermentation process. The main disadvantage of this technique concerns the release of contaminants.

The last step is the recovery of the active protein. When inclusion bodies have been solubilized, the refolding is achieved by removal of the denaturant. This can be done by different techniques including dilution, dialysis, diafiltration, gel filtration, chromatography, or immobilization on a solid support. Dilution has been extensively used. It considerably reduces concentrations of both denaturant and protein. This procedure, however, cannot be applied to the commercial scale refolding of recombinant proteins, because large downstream processing volumes increase the cost of products. Although dialysis through semipermeable membranes has been used successfully to refold several proteins, it is not employed in large-scale processes. This is because it requires very long processing times, and there is the risk that during dialysis, the protein will remain too long at a critical concentration of denaturant and aggregate. The removal of the denaturant may be accomplished through gel filtration. However, here again, a possible aggregation could lead to flow restriction within the column. Dialfiltration through a semipermeable membrane allows the removal of denaturant and other small molecules and retains the protein. This procedure has been used for large-scale processing and was particularly efficient in the refolding of prorennin and interferon-β.

During the refolding process, the formation of incorrectly folded species and aggregates usually decreases the refolding yield. For disulfide-bridged proteins, the renaturation buffer must contain redox-shuffling mixture to allow the formation of correctly paired disulfide bridges. Stabilizing reagents may be added to improve the refolding yield. An efficient strategy is the addition of small molecules to suppress intermolecular interactions leading to aggregation. Sugar, alcohols, polyols (including sucrose, glycerol, polyethylene glycol, isopropanol), cyclodextrin, lauryl-maltoside, sulfobetains, L-arginine, and low concentrations of denaturants and detergents, have been used to increase the refolding yield. L-arginine at a concentration ranging from 0.4 M to 0.8 M is the most widely used additive today.

Another important factor in the refolding process is the rate of removal of the denaturant. Since there is kinetic competition between the correct folding and the formation of aggregates from a folding intermediate, conditions that favor folding over the accumulation of aggregates must be found. To optimize this selection, Vilick and de Bernadez–Clark developed a strategy for achieving high protein refolding yields. They start from a model of refolding, develop the equations of refolding kinetics, characterize the rate-limiting step of the process, determine the influence of various environmental parameters, and finally optimize the system of equations in a scheme involving diafiltration to remove the denaturant. The approach was evaluated in the refolding of carbonic anhydrase from 8 M urea. The yield obtained after three diafiltration experiments was 69% whereas the model predicted a yield of 73%.

The properties of molecular chaperones have also been utilized to increase the refolding yield. Altamiro and coworkers have developed a system for refolding chromatography that utilizes GroEL, DsbA, and peptidyl–prolyl isomerase immobilized on an agarose gel. Kohler and coworkers have built a chaperone-assisted bioreactor; however, it could only be used for three cycles of refolding and needs to be improved. Another strategy consists of the co-overproduction of the DnaK–DnaJ or GroEL–GroES chaperones with the desired protein; this can greatly increase the soluble yield of aggregation-prone proteins. Fusion proteins have also been used to minimize aggregation.

The recovery of active proteins from inclusion bodies is a rather complex process. Although some general strategies have been developed, optimal conditions have to be determined for each protein. Recently, genetic strategies to improve recovery processes for recombinant proteins have been introduced. They consist of the introduction of combinatorial protein engineering to generate molecules highly specific to a particular ligand. Such methods, which allow efficient recovery of a recombinant protein, will be increasingly used in industrial scale bioprocesses as well.

4.2 The Formation of Amyloid Fibrils and its Pathological Consequences

The formation of amyloid fibrils plays a key role in the origin of several neurodegenerative pathologies, such as spongiform encephalopathies and Alzheimer's disease. Historically, the term amyloid was introduced to describe fibrillar protein deposits associated with diseases known as *amyloidoses* that involve the extracellular deposition of amyloid fibrils and plaques with the aspect of starch. For many of these diseases, the major fibrillar

protein component has been identified. In the 1970s, it was demonstrated that lysosomal proteins under acidic conditions could form amyloid fibrils. It was generally accepted at this time that proteolysis was the amyloidogenic determinant. Twenty years later, it was shown that purified transthyretin is converted into amyloid fibrils via an acid-induced conformational change *in vitro*, demonstrating that conformational changes alone were responsible for producing an intermediate generating amyloid structure. These aberrant protein self-assemblies are at the origin of more than hundred human amyloid diseases, some of them being lethal.

Twenty unrelated protein precursors are known to form amyloid fibrils, among them transthyretin, lysozyme, immunoglobulin light chain, β_2 microglobulin, Alzheimer Aβ1–40 and Aβ1–42 peptides, the mammalian prion protein, and the yeast prion-like proteins (Table 1). Since they are subjects of another chapter, prion proteins will not be discussed here. Although they have no homology in sequence and structure, all form amyloid fibrils with a similar overall structure, suggesting a common self-assembly

Tab. 1 Amyloidogenic proteins and the corresponding diseases.

Clinical syndrome	***Precursor protein***	***Fibril component***
Alzeimer's disease	APP	β-peptide 1–40 to 1–43
Primary systemic amyloidosis	Immunoglobulin light chain	Intact light chain or fragments
Secondary systemic amyloidosis	Serum amyloid A	Amyloid A (76-residue fragment)
Senile systemic amyloidosis	Transthyretin	Transthyretin or fragments
Familial amyloid polyneuropathy I	Transthyretin	Over 45 transthyretin variants
Hereditary cerebral amyloid angiopathy	Cystatin C	Cystatin C minus 10 residues
Hemodialysis-related	β_2-microglobulin	β_2-microglobulin
amyloidosis	Apolipoprotein A1	Fragments of Apolipoprotein A1
Familial amyloid polyneuropathy III	Gelsosin	71-amino acid fragment of gelsosin
Finnish hereditory systemic amyloidosis	Islet amyloid polypeptide (IAPP)	Fragment of IAPP
Type II diabetes	Calcitonin	Fragments of calcitonin
Medullary carcinoma of the thyroid		
Spongiform encephalopathies	Prion	Prion or fragments thereof
Atrial amyloidosis	Atrial natriuretic factor (ANF)	ANF
Hereditary nonneuropathic systemic amyloidosis	Lysozyme	Lysozyme or fragments thereof
Injection-localized amyloidosis	Insulin	Insulin
Hereditary renal amyloidosis	Fibrinogen	Fibrinogen fragments
Parkinson disease	α-synuclein*	

Source: (According to Kelly, J.W. (1996) Alternative conformations of amyloidogenic proteins govern their behavior, *Curr. Opin. Struct. Biol.* **6**, 11–17); *From J.C. Rochet & P.T. Lansbury (2000) *Curr. Opin. Struct. Biol.* **10**, 60–68.

pathway. In all proteins known to form amyloid fibrils, there is a conversion of α- to β-structure. Amyloid fibrils are abnormal, insoluble, and generally protease-resistant structures. They were first recognized by their staining properties. The most commonly used method to detect amyloid is staining by Congo Red, which exhibits a green birefringence. Amyloid fibrils are generally 60 to 100 Å in diameter and of variable length. X-ray diffraction data on fibrils, solid-state NMR studies, cryoelectron microscopy, and infrared Fourier transform experiments have shown that amyloid fibrils are made of two or more β-sheet filaments wound around one another. They have a characteristic cross-β repeat structure, the individual β-strands being oriented perpendicular to the long axis of the fibril.

Recently, progress has been made in the knowledge of the mechanisms involved in the formation of amyloid fibrils. Oligomeric prefibrillar intermediates have been extensively characterized with respect to their structure and temporal evolution. A well-documented example is provided by the studies on transthyretin. The biological role of this protein is the transport of thyroxin by direct binding and the transport of retinol via the retinol binding protein. The wild-type protein is very stable at neutral pH. In certain individuals, however, it is converted into amyloid fibrils, and this is associated with the disease, senile systemic amyloidosis. Several variants are associated with familial polyneuropathies. *In vitro* biophysical studies have identified conditions leading to amyloid formation. The three-dimensional structure of the protein is known. The wild-type protein is a tetramer at pH ranging between 5 and 7; the tetramer dissociates into a monomer when the pH decreases. The dissociation is the rate-limiting step of the process. The monomer exhibits an altered tertiary structure, which aggregates in amyloid

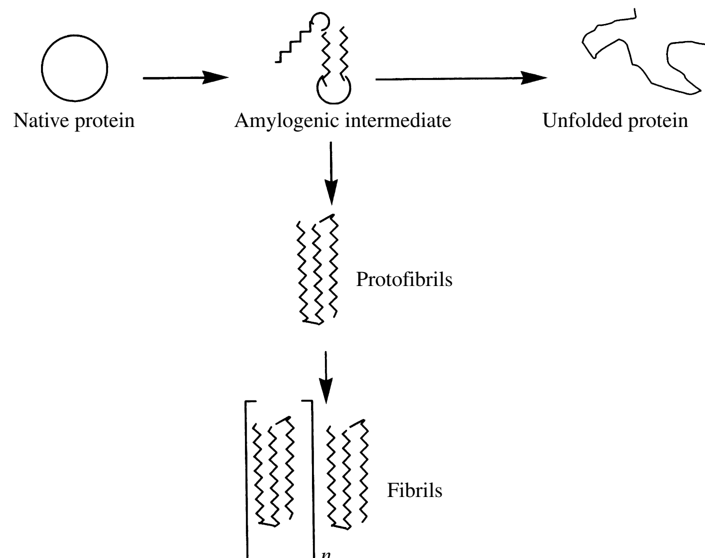
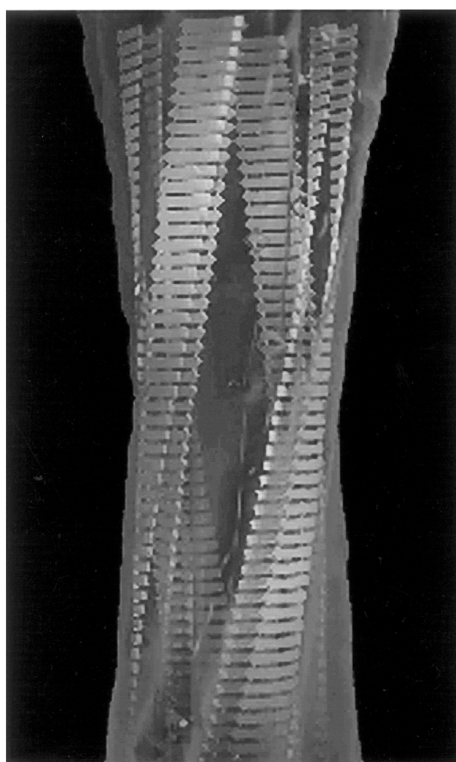


Fig. 10 Schematic representation of the formation of amyloid fibrils from a partially folded intermediate.

protofilaments, and then forms amyloid fibrils. This formation is at its maximum at pH 4.4. Using deuterium–proton exchange monitored by two-dimensional NMR spectroscopy on transthyretin at pH 5.75 and 4.5, Liu et al. have shown a selective destabilization of one half of the β-sandwich structure of the protein, increasing the mobility of this region. These studies have identified the residues that undergo increased conformational fluctuations under amyloidogenic conditions. The mutations in the pathological variants responsible for familial amyloid polyneuropathies are localized in this region. A strategy to delay the formation of amyloid fibrils proposed by Saccheti & Kelly was to develop molecules capable of stabilizing the tetramer.



Two variants of human lysozyme, Ile56Thr and Asp67His have been reported to be amyloidogenic; they are responsible for fatal amyloidoses. Pepys and colleagues have determined the precise structures and properties of these mutants. The native fold of the two amyloidogenic variants, as resolved by X-ray crystallography, is similar to that of the wild-type protein. Both variants are enzymatically active, but have been shown to be unstable. The replacement of an aspartate by a histidine suppresses a hydrogen bond formed in the wild-type protein with a tyrosine in a neighboring β-strand. This rupture opens a large gap between two β-strands. In the other variant, the replacement of an isoleucine by a threonine suppresses a van der Waals contact with a neighboring helix. Consequently, changes in the interface between the α- and β-domains occur in both variants, destabilizing the molecule.

The mutations leading to amyloid fibril formation are observed to result in a decreased stability of the native state. In all cases, the formation of fibrils occurs from a partially structured molecule via nucleation-dependent oligomerization. It was observed for several proteins that fibrillation takes place only after a lag phase, which is abolished upon seeding. Nucleation is followed by the formation of protofibrils whose characteristics have been determined (Fig. 10). Atomic force microscopy and fluorescence correlation spectroscopy have been used to monitor transitions among the different types of assemblies.

Fig. 11 Molecular model of an amyloid fibril derived from cryoelectron microscopy analysis of fibrils grown from an SH3 domain by incubation of a solution containing the protein at low pH (reproduced from Dobson, C. (1999) *TIBS* **24**, 331, with permission).

Recent observations from Dobson and his group have shown that several proteins unrelated to amyloid diseases are able to aggregate *in vitro* into amyloid fibrils when exposed to mild denaturing conditions. These fibrils are indistinguishable from those found in pathological conditions. It was demonstrated for different proteins such as normal lysozyme, an SH_3 domain of a phosphatidyl inositol protein kinase (Fig. 11), an acyl phosphatase, and an α-helical protein, myoglobin suggesting a common mechanism for the formation of amyloid. These findings clearly indicate that amyloid formation is a general property of polypeptide chains rather than one restricted to definite sequences as occurs with chameleon sequences capable of adopting either a β- or an α-helicoidal structure depending on their environment. Furthermore, these aggregates exhibited an inherent toxicity when incubated with mouse fibroblasts. Several groups suggest that oligomeric intermediates rather than fibrils themselves are responsible for pathogenicity.

Significant progress has been made in understanding the mechanisms involved in the formation of amyloid fibrils. This is an important step in guiding research into the discovery of molecules with therapeutic efficiency.

See also Circular Dichroism in Protein Analysis.

Bibliography

Books and Reviews

Dobson, C.M., Šali, A., Karplus, M. (1998) Protein folding. A perspective from theory and experiment, *Angew. Chem. Int. Ed.* **37**, 868–893.

Ellis, R.J. (1991) Molecular chaperones, *Ann. Rev. Biochem.* **60**, 321–347.

Ellis, R.J., Hartl, F.U. (1999) Principles of protein folding in the cellular environment, *Curr. Opin. Struct. Biol.* **9**, 102–110.

Georgiou, G., de Bernadez-Clark, E. (Eds.) (1991) *Protein Refolding*, ACS Symposium Series 470, American Chemical Society, Washington, DC.

Ghélis, C., Yon, J.M. (1982) *Protein folding*, Academic Press, New York.

Jaenicke, R. (1987) Folding and association of proteins, *Prog. Biophys. Mol. Biol.* **49**, 117–237.

Kim, P.S., Baldwin, R.L. (1990) Intermediates in protein folding of small proteins, *Ann. Rev. Biochem.* **59**, 631–660.

Schlunegger, M.P., Bennett, M.J., Eisenberg, D. (1997) Oligomer formation by 3D domain swapping: a model for protein assembly and disassembly, *Adv. Protein Chem.* **50**, 61–122.

Wetlaufer, D.B., Ristow, S. (1973) Acquisition of the three-dimensional structure of proteins, *Ann. Rev. Biochem.* **42**, 135–158.

References of Primary Literature

Altamiro, M.M., Garcia, C., Possani, L.D., Fersht, A.R. (1999) Oxidative refolding chromatography: folding of the scorpion toxin Cn5, *Nat. Biotechnol.* **17**, 187–191.

Anfinsen, C.B. (1973) Principles that govern the folding of polypeptide chains, *Science* **181**, 223–230.

Anfinsen, C.B., Haber, E., Sela, M., White, F.H. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain, *Proc. Natl. Acad. Sci. USA* **47**, 1309–1314.

Baneyx, F. (1999) Recombinant protein expression in Escherichia coli, *Curr. Opin. Biotechnol.* **10**, 411–421.

Bennett, M.J., Choe, S., Eisenberg, D. (1994) Domain swapping: Entangling alliance between proteins, *Proc. Natl. Acad. Sci. USA* **91**, 3127–3131.

Bergdoll, M., Remy, M.H., Capron, C., Masson, J.M., Dumas, P. (1997) Proline-dependent oligomerization with arm exchange, *Structure* **5**, 391–401.

de Bernadez-Clark, E. (2001) Protein refolding for industrial processes, *Curr. Opin. Biotechnol.* **12**, 202–207.

de Bernadez-Clark, E., Schwarz, E., Rudolph, R. (1999) Inhibition of aggregation side-reactions during in vitro protein folding, *Methods Enzymol.* **309**, 217–236.

Booth, D.R., Suide, M., Belotti, V., Robinson, C.V., Hutchinson, W.L., Fraser, P.E., Hawkins, P.N., Dobson, C.M. Radford, S.E., Blake, C.C.F., Pepys, M.B. (1997) Instability, unfolding, and aggregation of human lysozyme variants underlying amyloid fibrillogenesis, *Nature* **385**, 787–793.

Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C.M., Stefani, M. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases, *Nature* **416**, 507–511.

Chiti, F., Wabster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., Dobson, C.M. (1999) Designing conditions for in vitro formation of amyloid protofilaments and fibrils, *Proc. Natl. Acad. Sci. USA* **96**, 3342–3344.

Crestfield, A.M., Stein, W.H., Moore, S. (1962) On the aggregation of bovine pancreatic ribonuclease, *Arch. Biochem. Biophys.* **1** (Suppl.), 217–222.

Dijkstra, K., Karvonen, P., Pirneskoski, A., Koivunen, P., Kivirikko, K.I., Darby, N.J., van Staaden, M., Sheek, R.M., Kemmink, J. (1999) Assignment of ^{1}H, ^{13}C and ^{15}N resonances of the a domain of protein disulfide isomerase, *J. Biomol. NMR* **14**, 194–196.

Dill, K.A. (1985) Theory for folding and stability of globular proteins, *Biochemistry* **24**, 1501–1509.

Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.H., Thomas, P., Chan, H.S. (1995) Principles of protein folding: a perspective from simple exact models, *Protein Sci.* **4**, 561–602.

van Duyne, G.D., Standaert, R.F., Karplus, P.A., Stuart, L., Shreiber, S.L., Clardy, J. (1993) Atomic structure of the human immunophilin FKBP12 complexes with FK506 and rapamycin, *J. Mol. Biol.* **229**, 105–124.

Edman, J.C., Ellis, L., Blacher, R.W., Roth, R.A., Rutter, W.J. (1985) Sequence of protein disulfide isomerase: implications of its relationship to thioredoxin, *Nature* **317**, 267–270.

Edwards, K.J., Ollis, D.L., Dixon, N.E. (1997) Crystal structure of cytoplasmic Escherichia coli peptidyl-prolyl isomerase: evidence for decrease mobility of loops upon complexation, *J. Mol. Biol.* **271**, 258–265.

Eliezer, D., Chiba, K., Tsurata, H., Doniach, S., Hodgson, K.O., Kihara, H. (1993) Evidence of an associative intermediate on the myoglobin refolding pathway, *Biophys. J.* **65**, 912–917.

Eliezer, D., Jennings, P.A., Wright, P.E.H. Doniach, S., Hodgson, K.O., Tsurata, H. (1995) The radius of gyration of an apomyoglobin folding intermediate, *Science* **270**, 487–488.

Ellis, R.J. (2001) Molecular crowding: an important but neglected aspect of the cellular environment, *Curr. Opin. Struct. Biol.* **11**, 114–119.

Ellis, R.J., Hartl, F.U. (1999) Principles of protein folding in the cellular environment, *Curr. Opin. Struct. Biol.* **9**, 102–110.

Fandrich, M., Flechter, M.A., Dobson, C.M. (2001) Amyloid fibrils from muscle myoglobin, *Nature* **410**, 163–166.

Fink, A.L. (1998) Protein aggregation, folding aggregates, inclusion bodies and amyloid, *Fold. Des.* **3**, R9–R23.

Freedman, R.B., Klappa, P., Ruddock, L.W. (2002) Protein disulfide isomerases exploit synergy between catalytic and specific binding domains, *EMBO Rep.* **31**, 136–140.

Georgiou, G., Valax, P. (1999) Isolating inclusion bodies from bacteria, *Methods Enzymol.* **309**, 48–58.

Goldberg, M.E. (1973) L'état natif est-il l'état fondamental? in: Sadron, C. (Ed.) *Dynamic Aspect of Conformational Changes in Biological Macromolecules*, Reidel, Dordrecht.

Guijarro, J.L., Sunde, M., Jones, J.A., Campbell, I.D., Dobson, C.M. (1998) Amyloid fibril by an SH3 domain, *Proc. Natl. Acad. Sci. USA* **95**, 4224–4228.

Harrison, S.C., Durbin, R. (1985) Is there a single pathway for the folding of a polypeptide chain? *Proc. Natl. Acad. Sci. USA* **82**, 4028–4030.

Horowitz, P. Criscimagna, N.L. (1986) Low concentrations of guanidinium chloride expose apolar surfaces and cause differential perturbation in catalytic intermediates of rhodanese, *J. Biol. Chem.* **261**, 15652–15658.

Jaenicke, R. (1991) Protein folding: local structures, domains, subunits and assemblies, *Biochemistry* **30**, 3147–3161.

Kadokura, H., Beckwith, J. (2002) Four cysteines of the membrane protein DsbB act in concert to oxidize its substrate DsbA, *EMBO J.* **21**, 2354–2363.

Karplus, M. Šali, A. (1996) Theoretical studies of protein folding and unfolding, *Curr. Opin. Struct. Biol.* **5**, 58–73.

Kelly, J.W. (1996) Alternative conformations of amyloidogenic proteins govern their behavior, *Curr. Opin. Struct. Biol.* **6**, 11–17.

Kelly, J.W. (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways, *Curr. Opin. Struct. Biol.* **8**, 101–106.

Kelly, J.W. (2002) Towards an understanding of amyloidogenesis, *Nat. Struct. Biol.* **9**, 323–325.

Kohler, R.J., Preuss, M., Miller, A.D. (2000) Design of a molecular chaperone-assisted protein folding bioreactor, *Biotechnol. Prog.* **16**, 671–675.

Kuhlman, B., O'Neill, J.M., Kim, D.E., Zhang K.Y., Baker, D. (2001) Conversion of a monomeric protein L to an obligate dimer by computational protein design, *Proc. Natl. Acad. Sci. USA* **98**, 10687–10691.

Langer, T., Lu, C., Flanagan, J., Hayer, A.K., Hartl, F.U. (1994) Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding, *Nature* **356**, 683–689.

Levinthal, C. (1968) Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44–45.

Liu, K. Cho, H.S., Lashuel, H.A., Kelly, J.W., Wemmer, D.E. (2000) A glimpse of a possible amyloidogenic intermediate of transthyretin, *Nat. Struct. Biol.* **7**, 754–757.

Liu, Y., Gotte, G. Libonati, M., Eisenberg, D. (2001) A domain-swapped ribonuclease A dimer with implication for amyloid formation, *Nat. Struct. Biol.* **8**, 282–284.

Liu, Y., Gotte, G. Libonati, M., Eisenberg, D. (2002) Structures of the two 3D domain-swapped ribonuclease A trimers, *Protein Sci.* **11**, 371–380.

London, J., Skrzynia C., Goldberg, M.E. (1974) Renaturation of Escherichia coli tryptophanase after exposure to 8 M urea. Evidence for the existence of nucleation centers, *Eur. J. Biochem.* **47**, 409–415.

Marston, F.A.O. (1986) The purification of eukaryotic polypeptides synthesized in Escherichia coli, *Biochem. J.* **240**, 1–12.

Martin, J.L., Bardwell, J.C., Kuriyan, J. (1993) Crystal structure of DsbA protein required for disulfide bond formation, *Nature* **365**, 464–468.

McCarthy, A.A., Haebel, P.W., Torronen, A., Rybin, V., Baker, E.N., Metcalf, P. (2000) Crystal structure of the protein disulfide bond isomerase, DsbC, from Escherichia coli, *Nat. Struct. Biol.* **7**, 196–199.

Mitraki, A., Betton, J.M., Desmadril, M., Yon, J.M. (1987) Quasi-irreversibility in the unfolding-refolding transition of phosphoglycerate kinase induced by guanidine hydrochloride, *Eur. J. Biochem.* **163**, 29–34.

Morozova-Roche, L.A., Zurdo, J., Spencer, A., Noppe, W., Receveur, V., Archer, D.B., Joniau, M., Dobson, C.M. (2000) Amyloid fibril formation and seeding by wild-type lysozyme and its disease-related mutational variants, *J. Struct. Biol.* **130**, 339–351.

Pecorari, F., Minard P., Desmadril, M., Yon, J.M. (1996) Occurrence of transient multimeric species during the refolding of a monomeric protein, *J. Biol. Chem.* **271**, 5270–5276.

Rochet, J.C., Lansbury, P.T. (2000) Amyloid fibrillogenesis: themes and variations, *Curr. Opin. Struct. Biol.* **10**, 60–68.

Roseman, A.M., Chen, S., White, H., Sebil, H.R. (1996) The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate binding domains in GroEL, *Cell* **87**, 241–251.

Saccheti, J.C., Kelly, J.W. (2002) Therapeutic strategies for human amyloid diseases, *Nat. Rev. Drug Disc.* **1**, 267–275.

Semisotnov, G.V., Kuwajima, K. (1996) Protein globularization during folding- a study by synchrotron small-angle X-ray scattering, *J. Mol. Biol.* **262**, 559–574.

Silow, M., Tan, Y.J., Fersht, A.N., Oliveberg, M. (1999) Formation of short-lived protein aggregates directly from the coil in two-state folding, *Biochemistry* **38**, 13006–13012.

Speed, M.A., Wang, D.I.C., King, J. (1996) Specific aggregation of partially folded polypeptide chains. The molecular basis of inclusion body composition, *Nat. Biotech.* **14**, 1283–1287.

Taketomi, H., Kano, F., Gô, N. (1988) The effect of amino acid substitutions on protein folding and unfolding studied by computer simulation, *Biopolymers* **27**, 527–560.

Wetzel, R., Perry, L.J., Veilleux, C. (1991) Mutations in human interferon γ affecting inclusion body formation by a general immunochemical screen, *Biotechnology* **9**, 731–737.

Wolynes, P.G., Onuchic, J.N., Thirumalai, D. (1995) Navigating the folding routes, *Science* **267**, 1618–1620.

Xu, Z., Horwich, A.L., Sigler, P.B. (1997) Crystal structure of the asymmetric GroEL-GroES-$(ADP)_7$ chaperonin complex, *Nature* **388**, 741–750.

Yon, J.M. (1996) The specificity of protein aggregation, *Nat. Biotech.* **14**, 1231.

Zhang, X., Beuron, F., Freemont, P.S. (2002) Machinery of protein folding and unfolding, *Curr. Opin. Struct. Biol.* **12**, 231–238.

Aging and Sex, DNA Repair in

Carol Bernstein and Harris Bernstein
University of Arizona, Tucson, AZ, USA

1 The DNA Damage Theory of Aging 56
1.1 Occurrence of DNA Damage and Pathways of DNA Repair 56
1.2 Consequences of Unrepaired DNA Damage 56
1.3 Life Span Extension by Genetic Alterations that Increase DNA Repair, Reduce Oxidative Damage, or Reduce Cell Suicide (Apoptosis) due to DNA Damage 57
1.4 Premature Aging Associated with Defects in DNA Repair or Increased Oxidant Status 63
1.5 Normal Aging in spite of Certain Defects in DNA Repair or Increases in Antioxidant Enzyme Production 64
1.6 Negative Correlation between Mitochondrial ROS Production and Life Span 65
1.7 Other Work Indicating the Central Role of DNA Damage and DNA Repair in Aging 67
1.8 Calorie Restriction and Aging 67
1.9 General Strategies for Coping with DNA Damage and Some Consequences 68
1.10 Potential Immortality of the Germ Line 69

2 DNA-repair Pathways and Their Relation to Aging 70
2.1 NER (Nucleotide Excision Repair) 70
2.2 BER (Base Excision Repair) 70
2.3 HRR (Homologous Recombinational Repair) 73
2.4 NHEJ (Nonhomologous End Joining) 75
2.5 MGMT (O^6-Methylguanine-DNA Methyltransferase) 76
2.6 Enzymes of DNA Repair Pathways and Enzymes Regulating DNA Damage–inducing ROS Contribute to Determination of Aging 77

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

3 The DNA Repair (and Complementation) Theory of Sex 77
3.1 Meiosis, an Adaptation for Repairing Germ Line DNA 77
3.2 Frequency of Sexual Reproduction 78
3.3 Costs of Sex 78
3.4 Benefits of Sex 79
3.5 Repair of DNA Damage of the Germ Line in Nonmeiotic Cell Divisions 80
3.6 Meiotic Recombination in Eukaryotes Probably Evolved from Recombination in Prokaryotes 81
3.7 Homologs of the Bacterial RecA Protein Have a Key Role in Eukaryotic Meiotic Recombinational Repair 81
3.8 The Adaptive Function of Recombination Appears to Be Removal of DNA Damage 82
3.9 Other Expectations of the DNA Repair and Complementation Theory of Sex 83

4 Vegetative Survival Strategies 84
4.1 Survival of Vegetative Cell Populations 84
4.2 Vegetative Survival Strategy for Mitochondria (and Chloroplasts) in the Germ Line 84
4.3 Dolly the Cloned Lamb, Cumulina the Cloned Mouse, and Low Success of Cloned Mammals 85

5 Three Levels of Sexual Communication Reflect (1) DNA Repair, (2) Complementation, and (3) Selection for Fitness 86
5.1 Sexual Communication in Bacteria, Primarily for DNA Repair 87
5.2 Sexual Communication in Fungi, Primarily for DNA Repair, but Also for Limited Complementation 88
5.3 Sexual Communication in Higher Eukaryotes, Primarily for Complementation while Repair Is an Automatic Concomitant of Sexual Reproduction 89
5.4 "Level 3" of Sexual Communication Is Used to Select for Fitness among Potential Mates 91

6 Overview 91

Bibliography 92
Books and Reviews 92
Primary Literature 92

Keywords

Aging
The progressive impairments of functions experienced by many organisms throughout their life span.

DNA Damage
A DNA alteration that has an abnormal structure, which cannot itself be replicated when the DNA is replicated, but which may be repaired.

DNA Repair
The process of removing damage from DNA and restoring the DNA structure.

Mutation
A change in the sequence of DNA base pairs, which may be replicated and thus inherited.

Sex
The process by which genetic material (usually DNA) from two separate parents is brought together in a common cytoplasm where recombination of the genetic material ordinarily occurs, followed by the passage of the recombined genome(s) to progeny.

Complementation
The masking of the expression of mutant genes by corresponding wild-type genes when two homologous chromosomes share a common cytoplasm.

A number of theories have been proposed to account for the biological phenomena of aging and sexual reproduction (sex). An emerging unified theory that accounts for a considerable amount of the data relating to both aging and sex is presented here.

Aging appears to be a consequence of DNA damage, while sexual reproduction (sex) appears to be an adaptation for coping with both DNA damage and mutation. DNA, the genetic material of most organisms, is composed of molecular subunits that are not endowed with any peculiar chemical stability. Thus, DNA is subject to a wide variety of chemical reactions that might be expected of any such molecule in a warm aqueous medium. DNA damages are known to occur very frequently, and organisms have evolved enzyme-mediated repair processes to cope with them. In any cell, however, some DNA damage may remain unrepaired despite repair processes. Aging appears to be due to the accumulation of unrepaired DNA damage in somatic cells, especially in nondividing cells such as those in mammalian brain and muscle.

On the other hand, the primary function of sex appears to be the repair of damages in germ cell DNA through efficient recombinational repair when chromosomes pair during the sexual process. This allows a relatively undamaged genome to initiate the next generation. In addition, in diploid organisms, sex allows chromosomes from genetically unrelated individuals (parents) to come together in a common cytoplasm (that of progeny). Since genetically unrelated parents ordinarily would not have common mutations, the chromosomes present in the progeny should complement each other, masking expression of any deleterious mutations that might be present.

Thus, aging and sex appear to be two sides of the same coin. Aging reflects the accumulation of DNA damage and sex reflects the removal of DNA damage, and in diploid organisms, the masking of mutations by complementation.

1 The DNA Damage Theory of Aging

1.1 Occurrence of DNA Damage and Pathways of DNA Repair

Except for certain viruses with an RNA genome, the genomes of most organisms are composed of DNA. If DNA damage is the cause of aging, then DNA damage is expected to occur frequently in multicellular organisms. Table 1 lists some important types of DNA damage caused by normal metabolic processes in mammals. These data suggest, for instance, that in the rat at least 95,000 DNA damages of various types occur, averaged over all cell types, per cell per day. The majority of these damages alter the structure of only a single DNA strand, so the redundant information in the complementary strand can usually be used to repair the damage. The damages shown in Table 1 are the newly occurring damages, most being rapidly repaired.

Five major DNA repair pathways known to be utilized by cells to repair the damages indicated in Table 1 are as follows:

- Nucleotide excision repair (NER) [with two subpathways, largely using the same enzymes: transcription coupled repair (TCR) and global genomic repair (GGR)]
- Base excision repair (BER)
- Nonhomologous end joining (NHEJ)
- Homologous recombinational repair (HRR)
- O^6-methylguanine-DNA methyltransferase (MGMT)

1.2 Consequences of Unrepaired DNA Damage

If accumulated DNA damages are the cause of aging, then repair processes would be less than 100% efficient; some types of unrepaired damages left each day would gradually build up in nondividing or slowly dividing cells. Most investigators examining the presence of DNA damages in tissues of young versus old mammals (usually rodents) have found an accumulation of damaged bases or single- or double-strand breaks with age. The tissues where accumulation of DNA damage has been shown include liver, kidney, heart, muscle, and brain.

A number of different types of DNA damage have been tested for their effects on transcription and DNA replication. It was found that transcription is blocked by UV-induced damages (mainly pyrimidine dimers) by adducts produced by derivatives of benzo[a]pyrene, *N*-acetoxy-2-fluorenylacetamide, or aflatoxin B_1 and also by the oxidized base, thymine glycol. UV-induced DNA damages and thymine glycol have also been shown to block DNA replication. These findings suggest that many types of DNA damage inhibit transcription and replication.

A reduction in the ability to transcribe mRNA should lead to a decline in the function of the cells. In fact, in mammalian

Tab. 1 Endogenous DNA damages in mammalian cells.

Type of damage	*Approximate average incidence (DNA damages/cell/day)*
Oxidative	500,000 (young mouse brain)
	2,000,000 (old mouse brain)
	86,000 (rats, all tissues)
	10,000 (humans, all tissues)
Depurinations	9,000 (humans and rats)
Single-strand break	7,200 (*in vitro*)
O^6-methylguanine	2,000 (*in vitro*)
Double-strand break	>40 (rats)[a]
	>3 (humans)[a]
DNA cross-link	>37 (rats)[a]
	>3 (humans)[a]
Glucose 6-phosphate adduct	3 (humans)

[a]These numbers were calculated from the values in the references by methods indicated in the literature.

brain, it has been shown that as single-strand DNA damages accumulate with age, mRNA synthesis and protein synthesis decline, neuron loss occurs, tissue function is reduced, and functional impairments directly related to the central processes of aging (e.g. cognitive dysfunction and decline in homeostatic regulation) occur. Similarly, it has been shown in muscle cells that as single-strand DNA damages accumulate, mRNA and protein synthesis decline, cellular structures deteriorate, cells die, and this is accompanied by a reduction in muscle strength and speed of contraction. Thus, for brain and muscle, accumulation of DNA damage is paralleled by declines in function, suggesting a direct cause-and-effect relationship between the accumulation of DNA damage and major features of aging. In other cells, including those of liver and lymphocytes, evidence for an increase in DNA damage paralleled by a decline in gene expression and cellular function has also been observed. In general, it appears that tissues composed of nondividing or slowly dividing cells accumulate DNA damage and experience functional declines with age.

1.3 Life Span Extension by Genetic Alterations that Increase DNA Repair, Reduce Oxidative Damage, or Reduce Cell Suicide (Apoptosis) due to DNA Damage

Table 2 lists alterations in genes controlling DNA repair, oxidant status, or apoptosis that result in increased life span. The increases in life span found with the genetic alterations in Table 2 are usually an increase in the maximum life span (not just the mean life span) by about 30 to 40%. Mean life span can be extended by reductions in tumorigenesis or acute and sporadic diseases, not generally regarded as a cause of aging. The organisms with increased maximum life span reported here showed longer spans of normal vigorous activity (not merely slowed metabolism, which can also extend life span). The cellular roles

Tab. 2 Life span extension: increased life span from alterations in genes controlling DNA repair, apoptosis, or oxidant status.

Organism	*Genetic alteration*	*Pathway*	*Aging phenotype*	*Fertility phenotype*	*Spontaneous cancer phenotype*	*Cellular ROS*	*Effect on*		
							Induced DNA damage	*Spont. mutation*	*Induced apoptosis*
Mouse	100-fold excess copies MGMT gene	O^6-meG DNA repair	Life span extension	n.t.	Reduced	n.t.	Reduced O^6-meG	No effect at *H-ras* locus	n.t.
Fruit fly	Excess SOD in neurons or all tissues or with catalase in all tissues	Removes ROS	Life span extension	No change	n.t.	Reduced occurrence of oxidized proteins	n.t.	n.t.	n.t.
Fruit fly	MsrA excess	Methionine sulfoxide reductase	Life span extension	Extended	n.t.	Repaired protein oxidation	n.t.	n.t.	n.t.
Mouse	$p66^{shc}$ defect	Blocks oxidant and apoptosis parts of p53 pathways	Life span extension	n.t.	n.t.	Reduced	Reduced oxidative damage	n.t.	Reduced
Human	Higher specific activity of PARP	BER	Life span extension	n.t.	n.t.	n.t.	n.t.	n.t.	n.t.

Notes: ROS: reactive oxygen species; Spont.: spontaneous; MGMT: O^6-methylguanine-DNA methyltransferase; n.t.: not tested; SOD: superoxide dismutase.

of these genetic alterations are described below.

1. *MGMT*. One frequent type of DNA damage (see Table 1) is O^6-methylguanine, caused by low levels of alkylating agents present in food, water, air, and tobacco smoke, and formed by normal processes in the body mediated by gastric bacteria and macrophages. O^6-methylguanine is specifically repaired by a DNA repair enzyme called *O^6-methylguanine-DNA methyltransferase* (MGMT). MGMT transfers the extra methyl group from guanine in DNA to a particular amino acid within itself and becomes "used up" after the transfer occurs. The *MGMT* gene codes for one of the five DNA repair mechanisms listed in Section 1.1. As indicated in Table 2, when 100 copies of the *MGMT* gene were inserted into the mouse genome, these mice (under the usual conditions of mouse maintenance) had their life span extended and died at a considerably slower rate than wild-type mice.

2. *SOD*. Another important type of metabolically caused DNA damage is oxidative damage, the most frequent damage identified (Table 1). An apparently unavoidable by-product of normal respiratory metabolism is the production of reactive oxygen species (ROS) from molecular oxygen, and ROS cause oxidative damage. ROS include free radicals (where the symbol $^\bullet$ indicates an unpaired electron): the superoxide radical ($O_2^{\bullet -}$) and the hydroxyl radical ($OH^\bullet$). Another oxygen respiration by-product is hydrogen peroxide (H_2O_2). H_2O_2, if not removed, it diffuses fairly easily through the cell, and when it encounters Fe^{2+} (the ferrous ion), it can undergo the Fenton reaction and produce $OH^\bullet$ and other ROS. ROS produce a number of lesions in DNA, including base lesions, sugar lesions (the deoxyribose sugar is in the backbone of DNA), DNA–protein cross-links, single-strand breaks, double-strand breaks, and abasic sites.

The major ROS produced by the cell is $O_2^{\bullet -}$, formed in the mitochondria (the energy-producing organelles of the cell). Superoxide dismutase (SOD) occurs in two forms, manganese SOD (MnSOD) and copper/zinc SOD (Cu/ZnSOD). Both forms of SOD convert $O_2^{\bullet -}$ to the less damaging H_2O_2, and then another enzyme, catalase, converts H_2O_2 to molecular oxygen and water. MnSOD occurs in the mitochondria and Cu/ZnSOD occurs in the cytoplasm. As shown in Table 2, a number of investigators have found that inserting genes producing higher than normal levels of superoxide dismutase into the fruit fly (*Drosophila melanogaster*) genome results in life span extension. Insertion of genes producing either MnSOD or Cu/ZnSOD caused life span extension, although the artificially inserted Cu/ZnSOD only produced life span extension when its expression was restricted to the motor neurons, or solely to the adult phase of the fruit fly life cycle.

Aging has been found to correlate with increased levels of oxidative products, such as protein carbonyls and 8-oxo-guanine in DNA, and fruit flies lacking either catalase or Cu/ZnSOD have a reduced life span. Further, selection of a population of fruit flies for increased life span correlates with strongly increased expression of both MnSOD and Cu/ZnSOD. Reverse selection of these long life span flies to a shorter life span resulted in reduced expression of Cu/ZnSOD.

3. *MsrA*. In addition to DNA damage, free radicals damage proteins, lipids, and carbohydrates. Most proteins have a short half-life (averaging about three days in mouse liver). Oxidatively damaged proteins and lipids are subject to both

degradation and some repair reactions. If cellular genes that code for enzymes involved in the replacement of damaged proteins are themselves damaged, then damaged proteins may not turn over as rapidly, and protein damages may become important as they accumulate with age. Table 2 shows that insertion of an extra gene encoding bovine methionine sulfoxide reductase (MsrA) in the fruit fly genome, which helps repair oxidatively damaged proteins, leads to life span extension. Consistent with this, MsrA, when defective in the mouse, results in early aging (Table 3).

4. *p66Shc*. The *p53* gene has a central role in response to DNA damage. The p53 protein is directly active in three forms of DNA repair (NER, BER, and HRR). When there is no externally induced DNA damage, p53 has a half-life of only 5 to 40 minutes since specific enzymes target p53 for degradation. Thus, p53 is kept at a low level when there is no DNA damage. However, upon exposure of a cell to DNA-damaging agents, p53 becomes metabolically stable and, in addition, more copies of it are produced in the cell. In the presence of various types of DNA damage, p53 undergoes modifications at some of the 18 different sites within the protein. Some of these modifications [phosphorylations, acetylations, poly(ADP-ribosyl)ations, or sumoylations (covalent attachments of small ubiquitin-like proteins) allow the p53 protein to act as a regulatory agent, activating numerous other genes, carrying out different responses to different kinds or levels of DNA damage. The p53 protein can regulate or act in at least four major types of responses to DNA damage (acting as a "master switch"), and which action or transactivation (regulating the induction of other genes) it performs depends on the level and type of DNA damage. p53 can (1) send the cell into cell cycle arrest (to allow extra time for repair of DNA damage); (2) act directly in DNA repair (see Fig. 1 for where p53 acts in NER); (3) cause the cell to switch into a cell suicide mode (apoptosis); or (4) cause the cell to produce higher levels of ROS (apparently as a preliminary to entering the cell suicide mode of apoptosis). When acting to increase the internal level of ROS and entry into apoptosis, p53 acts through another gene it controls, *p66Shc*.

When a mouse embryo is produced with both copies of its *p66Shc* gene inactive (a p66Shc "knockout"), mouse embryo fibroblast cells derived from it have intracellular levels of ROS reduced by about 40%. Consistent with this reduction in ROS, there is also greatly reduced oxidative damage accumulation in both nuclear and mitochondrial DNA of these cells. A similar reduction in nuclear and mitochondrial DNA damage is seen *in vivo* in the tissues of lung, spleen, liver, and skin in 3- and 24-month-old *p66Shc* knockout mice, although there is no reduction in the brain, where *p66Shc* is not normally expressed. Cells of these mice are inhibited from undergoing apoptosis after cellular oxidative damage (when challenged with externally applied H_2O_2). Knockout mice without *p66Shc* show life span extension without any notable increase in cancer or other pathological defects (Table 2). Mice with a type of overactive p53 (an increase in some p53 functions) and intact *p66Shc* show early aging (Table 3). On the other hand, removal of all p53 functions (some of which are protective in DNA repair) also results in early aging (Table 3).

5. *PARP*. DNA damages caused by alkylating agents (such as those that methylate guanine, discussed above),

Tab. 3 Early aging: decreased life span from alterations in genes controlling DNA repair or protein oxidation.

Organism	*Genetic alteration*	*Pathway*	*Aging phenotype*	*Fertility*	*Spontaneous cancer*	*Effect on*			
						Cellular ROS	*Induced DNA damage*	*Spont. mutation*	*Induced apoptosis*
Human	RECQ3 helicase and exonuclease defect (Werner syndrome)	HRR and NHEJ	Early aging	Reduced	Increased	n.t.	Increased	Increased	n.t.
Human	RECQ2 helicase defect (Bloom syndrome)	HRR and NHEJ	Early aging	Reduced	Increased	n.t.	Increased	Increased	n.t.
Human	RECQ4 helicase defect (Rothmund–Thomson syndrome)	DNA repair pathway, unknown type	Early aging	Reduced	Increased	n.t.	n.t.	Increased	n.t.
Human and Mouse	XPD helicase defective at certain sites (Trichothiodystrophy)	NER, also alters transcription initiation	Early aging	Reduced	No change	n.t.	n.t.	n.t.	Increased
Human	CSB defect at 2 helicase motifs or ATPase motif (Cockayne syndrome)	BER if defective at helicase motif V or VI, TCR if defective in ATPase function	Early aging	n.t.	No change	n.t.	Increased	n.t.	Increased
Mouse	Ku-80 (activator of Ku-70 helicase) defect	NHEJ	Early aging	n.t.	Increased	n.t.	n.t.	n.t.	
Mouse	Topoisomerase IIIβ defect	Unknown, but probably DNA repair, replication, or recombination	Early aging	n.t.	n.t.	n.t.	n.t.	n.t.	

(*continued overleaf*)

Tab. 3 (*continued*)

Organism	***Genetic alteration***	***Pathway***	***Aging phenotype***	***Fertility***	***Spontaneous cancer***	***Effect on***			
						Cellular ROS	***Induced DNA damage***	***Spont. mutation***	***Induced apoptosis***
Mouse	ERCC1 defect	NER and interstrand cross-link repair (HRR)	Early aging	Infertile	No change	n.t.	Increased	Increased	
Mouse	p53 overactivated	Increases some p53 functions	Early aging	n.t.	Reduced	n.t.	n.t.	n.t.	
Mouse	p53 defect	Blocks all p53 functions, including NER, BER and HRR	Early aging	Reduced	Increased	n.t.	Increased	Increased	
Mouse	MsrA Defect	Methionine sulfoxide reductase	Early aging	n.t.	n.t.	Increased protein oxidation	n.t.	n.t.	

Notes: ROS: reactive oxygen species; Spont.: spontaneous; n.t.: not tested.

ionizing radiation (which produces DNA single- and double-strand breaks and oxidative damages), and ROS result in rapid activation of an enzyme called *poly(ADP-ribose) polymerase*, or PARP. PARP, similar to p53 discussed above, has a role as a "master switch". PARP can (1) act directly in one form of DNA repair, BER, (2) control the function of many other proteins by catalyzing the addition of ADP-ribose branched polymers onto them (either activating or repressing their function), and (3) trigger apoptosis (cell suicide). In addition, PARP controls new transcription or activities of a number of genes affecting survival or apoptosis, including p53. It was found that centenarians (humans who have lived for more than 100 years) have a modified form of PARP, which is more efficiently activated than the PARP of noncentenarians (Table 2), thereby apparently causing life span extension. In addition, the maximal poly(ADP-ribosyl)ation capacity (efficiency of activation of PARP) in leukocytes of 13 mammalian species of different life span was measured. There was a strong correlation of PARP efficiency of activation with species-specific life span.

1.4 Premature Aging Associated with Defects in DNA Repair or Increased Oxidant Status

DNA damages are so frequent (Table 1) that total absence of DNA repair of a common damage is likely to be incompatible with life. If a DNA repair pathway lacks an essential enzyme, but the missing enzyme can be, at least, partially compensated for by a similar enzyme, then repair may be adequate to allow sufficient survival and growth to show premature aging. This will also occur if a particular DNA damage is preferentially repaired by one pathway, but another repair pathway, with less efficiency, also repairs that damage.

1. *Helicase*. There are at least 31 human enzymes that are helicases or contain helicase-motif domains. Helicases are enzymes that unwind and separate the strands of DNA, usually using the hydrolysis of ATP to provide the necessary energy. Some enzymes with multiple helicase-motif domains only act as ATPases, providing energy to DNA-related processes. Helicases or enzymes with helicase-motif domains participate in DNA repair, DNA replication, and DNA recombination. Usually, the helicase activity is specific for a particular DNA configuration. Some helicases involved in particular DNA-repair pathways may be partially replaceable, at least at a low level, by other helicases. That may be why five genes, which code for enzymes with helicase functions, or helicase motifs plus an ATPase function, and which are required in different DNA repair pathways, when genetically defective, cause syndromes characterized by early aging in humans (Table 3). These syndromes are Werner syndrome, Bloom syndrome, Rothmund–Thomson syndrome, Trichothiodystrophy and Cockayne syndrome (Table 3). Similarly, in the mouse, a defect in the *Ku-80* gene, which normally activates the Ku-70 helicase function, results in an early aging phenotype (Table 3). The different helicases listed in Table 3 have specificities for HRR, NHEJ, NER, TCR or BER, so that defects in each of these DNA repair pathways may allow accumulation of different types of DNA damage, each type being able to contribute to premature aging.

2. *Topoisomerase*. Topoisomerases interact with helicases in DNA repair, recombination, and replication. When a helicase

unwinds the two DNA strands of the double helix, this introduces supercoiling of the associated DNA. Topoisomerases introduce controlled breaks plus reattachments in DNA to relieve supercoiling. There are a number of topoisomerases in mouse and human cells. The different topoisomerases interact specifically with different helicases. However, some topoisomerases may be partially replaceable by another topoisomerase at a low level. In the mouse, a mutant lacking topoisomerase IIIβ develops to maturity but shows early aging (Table 3). Topoisomerase IIIβ interacts with human RecQ5β helicase and is thought to act in DNA repair, replication, or recombination (Table 3).

3. *ERCC1.* Excision Repair Cross Complementing 1 (*ERCC1*), when defective, is another gene whose absence or truncation causes an early aging phenotype in the mouse (Table 3). ERCC1 functions in both NER and interstrand cross-link repair (in a step prior to HRR). ERCC1 has homology with an endonuclease active in NER in yeast, and that yeast endonuclease can compensate for the loss of a topoisomerase or a helicase. Thus, ERCC1 may have some functional similarity to topoisomerase or helicase in DNA repair. Conversely, loss of ERCC1 may be partially compensated for by a helicase or topoisomerase, or by another endonuclease in mouse, so that a defect in ERCC1 is not lethal but causes early aging. ERCC1 primarily functions in NER as an endonuclease as illustrated in Fig. 1.

4. *p53.* Similar to helicase and topoisomerase, p53 occurs as one of a family of enzymes, p53, p63, and p73 (and both p73 and p63 have multiple isoforms), which share significant homology and have similar functions. In particular, p73 has a role in activating DNA repair enzymes and in carrying out apoptosis in the face of excess DNA damage (see below in Fig. 3). Thus, loss of p53 may, in part, be compensated for by functions of p73 and/or p63. In Section 1.3.4, we briefly discussed an overactive form of p53 that causes early aging. This mutant form of p53 has its effect in the presence of a wild-type p53 (a heterozygous situation) where it may increase some functions detrimental to the cell. Although it is not known which functions it increases, an increase of p66Shc under p53 control could reasonably be expected to cause early aging, since it would increase DNA damage through increases in ROS. In addition, a p53 knockout mouse, lacking all functions of p53, including its functionality in three DNA repair pathways (NER, BER, and HRR), is also viable but ages prematurely (Table 3).

5. *MsrA.* As discussed in Section 1.3.3, if cellular genes, which code for activity in the replacement of damaged proteins, are themselves damaged, then damaged proteins may not turn over as rapidly, and protein damages may become important as they accumulate with age. Added activity of MsrA in the fruit fly gave greater longevity. Defective MsrA in the mouse caused early aging (Table 3).

1.5 Normal Aging in spite of Certain Defects in DNA Repair or Increases in Antioxidant Enzyme Production

If a DNA repair pathway lacks an enzyme, but the missing enzyme is partially compensated for by a similar enzyme, then repair may be sufficient to allow survival, growth, and normal aging. While such fairly good compensation may allow normal aging, the repair of DNA damages would still be less than if the repair pathway were intact, and that could lead to

increased carcinogenesis. Such fairly good compensation for DNA repair enzyme defects may be the basis for normal aging, but increased carcinogenesis, shown by mice with the DNA repair mutations listed in Table 4. As shown in Fig. 1, XPA and XPC proteins each occur as one member of a pair of complexed proteins whose function is recognition of DNA damage to be repaired by an NER pathway. It is possible that the other complexes that can recognize DNA damage can compensate, at some level, when one of the recognition complexes is absent.

Further, in some instances, genetic alterations caused increased SOD production in fruit flies but did not affect aging. However, the inserted *SOD* gene may have been turned on under the control of promoters expressing in tissues where it may not have been useful, or at rather low levels.

1.6
Negative Correlation between Mitochondrial ROS Production and Life Span

If DNA damage is the major cause of aging, ROS are a major source of DNA damage, and mitochondria are a major source of ROS, then animals with mitochondria producing higher levels of ROS may have a shorter life span, other factors being equal. Comparisons were made between long-lived birds and short-lived mammals. Pigeons, with a maximum life span of 35 years were compared with rats (of similar body size) with a maximum life span of 4 years, and parakeets and canaries (maximum life spans of 21 and 24 years, respectively) were compared with mice (maximum life span of 3.5 years). Mitochondrial ROS production was lower in the longer-lived avian species. However, in addition, pigeons were shown to have higher levels of SOD in brain, heart, and kidney than the levels shown by rats, so there was also higher antioxidant enzymatic protection in the longer-lived pigeon. This indicates that antioxidant enzymes, which confer resistance to an externally added source of ROS may be of comparable importance to longevity as endogenous rates of ROS production. The values of one measured DNA-damaged base, 8-oxodeoxyguanine, were lower in canary brain and parakeet heart nuclear DNA than in the comparison tissues of the mouse, while in the other comparisons, the level of this one damaged base was not significantly different in nuclear DNA. In another experiment quoted by Herrero and Barja, other workers showed that nuclei of starlings (another long-lived bird, maximum life span of 20 years) have less DNA breaks and abasic sites after exposure to H_2O_2 than those of mice.

Mice heterozygous for a MnSOD defect ($MnSOD^{+/-}$ mice) have higher levels of oxidative damage to DNA, protein, and lipids in their mitochondria, but no increased damage to nuclear DNA or cytoplasmic proteins. The $MnSOD^{+/-}$ mice live as long as wild-type mice, showing that mitochondrial DNA damage (as distinct from nuclear DNA damage) may not be central to longevity. This may be because there are on the order of 1,000 mitochondria per cell, and mitochondria with excess damage may be replaced by replication of less damaged mitochondria.

In Section 1.3.2, we mentioned that the long-lived strains of fruit flies have higher levels of antioxidant defense enzymes. The long-lived flies also had mitochondria that had lower levels of ROS leakage. These less leaky mitochondria, when transferred to short-lived flies through maternal inheritance (only maternal mitochondria are passed on to and maintained in progeny

Tab. 4 Normal aging: alterations in genes controlling DNA repair or oxidant status but which do not alter life span.

Organism	*Genetic alteration*	*Pathway*	*Aging phenotype*	*Fertility*	*Spontaneous cancer*	*Effect On*			
						Cellular ROS	*Induced DNA damage*	*Spont. mutation*	*Induced apoptosis*
Mouse	XPA defect	NER	Unchanged[a]	Unchanged	Increased liver cancer	n.t.	n.t.	Increased in liver, no change in brain	n.t.
Mouse	XPC defect	GGR (NER)	Unchanged[a]	Unchanged	Unchanged (when not exposed to UV light)	n.t.	n.t.	Increased	n.t.
Fruit fly	Excess SOD	Removes ROS	Unchanged	Unchanged	n.t.	n.t.	n.t.	n.t.	n.t.

Notes: ROS: reactive oxygen species; Spont.: spontaneous; n.t.: not tested.
[a] Based on observations during the first half of life span.

flies), also transferred the ability to live about 25% longer.

1.7 Other Work Indicating the Central Role of DNA Damage and DNA Repair in Aging

Numerous studies have been performed in mammals on the correlation between the ability of cells to repair DNA and the life span of the species from which the cells were taken. The life spans of the species varied from 1.5 years for the shrew to 95 years for man. Almost all of the studies showed a positive correlation between DNA repair capacity and life span. Many experiments have been performed on the effect of adding antioxidants to the diets of organisms upon the organism's life span. Although the results of such experiments are not entirely consistent, certain antioxidants have been found to generally increase life span. Vitamin E, for example, has been found to increase the life span of rat, insects, rotifers, nematodes, and paramecium.

More than 50 studies have been performed to examine the possible experimental acceleration of aging by externally applied DNA-damaging agents. Overall, it has been found that sublethal doses of ionizing radiation or DNA-damaging chemicals in the diet shorten life span, but many specific aspects of normal aging are not accelerated. Several authors have noted that the distribution (over time and in different tissues) of DNA damages induced by external agents does not closely mimic that of natural damages. This difference could explain why the life-shortening effects induced by external agents do not closely conform to natural aging. In particular, natural damages probably accumulate gradually, so they would tend to build up in nondividing cells, while they would be diluted out in dividing cells. Exposure to an external agent over a brief period, on the other hand, could cause equally large numbers of damages to nondividing and rapidly dividing cells. The effect on rapidly dividing cells could be very large by interfering with DNA replication. In addition, if oxidative damages are important in normal aging, then brain cells, which have a high level of oxidative metabolism should have more damages than most other cell types. Externally applied damages would not be expected to produce this particular type of bias. Thus, the general finding that sublethal exposure to DNA-damaging agents shortens life span, while not uniformly accelerating the natural aging process, is consistent with the DNA damage theory of aging.

1.8 Calorie Restriction and Aging

Numerous studies have shown that calorie restriction in yeast, nematode worms, fruit flies, and rodents can give life span extension. The transcription of a large number of genes is altered with calorie restriction. In rat muscle, 34 of 800 genes tested were altered in their transcription under calorie restriction, with some of the largest changes being upregulation of Cu/ZnSOD and MnSOD. In fruit flies, where 14,028 genes were assessed, 2,188 showed significant up- or downregulation after calorie restriction. In particular, DNA repair genes were downregulated (which may reflect reduced DNA damage). Evaluation of upregulated genes was problematic, however, because expression measures are proportional to the total mRNA pool, and there was massive downregulation of a large number of genes related to cell growth, and

other targeted areas. Thus, genes whose absolute transcription abundance remains unchanged will appear to be upregulated.

Overall, calorie restriction results in slower accrual of oxidative damage. The steady state oxidative damage measured represents the equilibrium between oxidant generation, oxidant scavenging, repair, and protein and lipid turnover (DNA does not turn over, although it is repaired, with damaged bases and nearby bases being replaced). The consensus is that the primary reason for lower oxidative damage observed after calorie restriction is a reduction in the generation of ROS.

1.9 General Strategies for Coping with DNA Damage and Some Consequences

Different organisms, or even different tissues within the same organism, appear to use different strategies for dealing with DNA damage. The three major strategies are described below.

1. *Cell replacement strategy*. Bone marrow and hemopoietic cells of man, guinea pig, and mouse seem to maintain their population numbers by a cell-replacement strategy. For instance, mouse bone marrow cells have a turnover time of about 1 to 2 days, and there appear to be no significant differences in the erythrocyte production from marrow stem cell lines in old and young adult mice, suggesting that DNA damages do not accumulate in this cell population. However, rapidly dividing cells are vulnerable to accumulation of mutations, which arise by errors of replication. Hematopoietic stem cells accumulate mutations (not damages) as people age from birth to 96 years. The accumulation of mutations in replicating somatic cells is widely regarded as the cause of cancer.

2. *DNA repair in nondividing somatic cell populations*. The organs including brain, muscle, and liver consist, largely, of nondividing cells, and these cells carry out DNA repair as a strategy of coping with DNA damages. However, DNA repair is less than 100% efficient, and DNA damages accumulate with time. Brain, muscle, and liver are subject to some of the more conspicuous progressive declines in function characteristic of human aging. Investigations were reviewed in 1992 on the accumulation of DNA damage in mammalian muscle, brain, and liver. In that review, 4 studies on muscle, 9 on brain, and 14 on liver reported accumulation of DNA damage with age. In most of these investigations, the type of damage measured was single-strand breaks. In 1996, an increase in DNA single- and double-strand breaks in neurons of the rat cerebral cortex with age was found. Neurons in young 4-day-old rats had about 3,000 single-strand breaks, which increased to 7,400 in neurons of old rats more than 2 years of age. Double-strand breaks increased from about 156 in young rats to about 600 in old rats. It was suggested that gradual accumulation of DNA damage with age could be a primary reason for the breakdown of the metabolic machinery leading to the eventual senescence and death of the neuron. It was found in the same year that DNA adducts, a type of DNA damage, increase in rat brain with age. Some of these were identified as malondialdehyde adducts of dGMP. It was suggested that this accumulation of DNA damages may contribute to cerebral aging. A significant increase in single-strand breaks/alkali-labile sites with age in rat liver hepatocytes were reported in 1994. Thus, numerous studies

in mammals indicate that long-lived, nondividing, differentiated cells accumulate DNA damage with time. This damage may account for many of the progressive declines in functions that define aging.

3. *Cellular redundancy*. Like repair, cellular redundancy may be another strategy for coping with DNA damage in nondividing cells. The brain has an unusually high level of oxidative metabolism compared to other organs. Brain neurons are nondividing and there is evidence for DNA damage accumulation in the brain. There is a clear loss of neurons with age. Compared to young rats, old rats have about a 50% loss of neurons in many regions of their brains. Numerous studies in humans have also shown a loss of neurons with age. The brain appears to use a strategy of cellular redundancy to compensate for the loss of neurons with age. It has been estimated that the brain is twofold larger than necessary for short-term survival. Comparisons of different mammalian species indicate that the maximum life span in mammals is directly proportional to brain size. Thus, the brain appears to be protected from loss of neuronal function by cellular redundancy, and this type of redundancy may be significant in determining life span.

1.10 Potential Immortality of the Germ Line

Cells of the germ line are capable of avoiding aging. While multicellular organisms ordinarily age and die, their germ line is potentially immortal. The germ line is distinguished from other cell lines by periodic meiosis. HRR is especially promoted during meiosis, when the homologous chromosomes are closely paired along their length. Meiosis appears to be an adaptation for removing DNA damages through HRR, and the potential immortality of the germ line may be related to this special repair capability. This idea has been tested, using single-celled paramecia that can undergo either asexual or sexual reproduction. When they grow asexually, clones of *Paramecium tetraurelia* age (show reduced vigor) and then die. These paramecia have a macronucleus containing 800 to 1,500 copies of the genome that expresses cellular functions, and a micronucleus that contains the germ line DNA.

If the macronuclei of clonally young paramecia are injected into old paramecia, the old paramecia have their life span prolonged. In contrast, cytoplasmic transfer from young to old paramecia does not prolong the life span of the old paramecia. This suggests that the macronucleus, rather than the cytoplasm, determines clonal aging. Asexually growing clones of paramecia have been found to accumulate DNA damage in the macronucleus over successive generations of clonal growth. Upon sexual reproduction (conjugation) or self-fertilization (automixis), a new macronucleus develops from the micronucleus and the old macronucleus disintegrates. Both of these processes (conjugation and automixis) include meiosis, which involves pairing of homologous chromosomes and the opportunity for homologous recombinational repair (HRR) of DNA (see Section 2.3). It was found that at a few clonal generations after meiosis, the level of DNA damage in macronuclear DNA is low but then increases as the cells undergo clonal aging. Thus, accumulation of DNA damage in the macronucleus may account for clonal aging and DNA repair during meiosis (principally HRR) may account, in large part, for the potential immortality of the germ line.

2 DNA-repair Pathways and Their Relation to Aging

2.1 NER (Nucleotide Excision Repair)

NER repairs DNA with helix-distorting single-strand DNA damages, including some oxidative damages such as the cyclopurine 8,5′-(S)-cyclo-2′-deoxyadenosine. About 30 proteins are involved in NER. The key steps in the process (Fig. 1) are (1) recognition of a DNA defect; (2) recruitment of an initial repair complex; (3) preparation of the DNA for repair through the action of helicases; (4) incision of the damaged strand on each side of the damage with release of the damage in a single-strand fragment about 24 to 32 nucleotides long; (5) filling in of the gap by repair synthesis; and (6) ligation to form the final phosphodiester bond. The two subpathways of NER, TCR, and GGR, are initiated somewhat differently, but after initiation most enzymatic steps are the same.

Individuals with the inherited disease xeroderma pigmentosum (XP) are defective in NER and are sensitive to UV radiation. About 25% of such patients also have neurodegeneration attributed to neural oxidative damage. XP patients have defects in one of seven *XP genes*, *XPA* through *XPG*, with some of their roles indicated in Fig. 1. A number of specific mutations resulting in defects in XPD, a helicase, cause a milder disease than XP, called *trichothiodystrophy*, and these mutations cause early aging (Table 3). XPA or XPC, both involved in the recognition of DNA damage (Fig. 1), when defective, do not appear to affect aging (Table 4), perhaps because if either is defective, compensation can occur to some extent. However, a double-mutant mouse, defective with a trichothiodystrophy-causing XPD mutation plus an XPA mutation, ages more rapidly than a mouse with a single XPD defect.

ERCC1, an endonuclease, associates with XPF, and then this pair associates with the initial repair complex to carry out an incision on a strand carrying the DNA damage. Defects in ERCC1 cause early aging (Table 3). CSB is an enzyme with seven helicase motifs and an ATPase function. The ATPase function is required in TCR of NER. CSB interacts with the TFIIH complex [a transcription complex (transcription factor II H)] required in NER (Fig. 1), as well as with XPA, another NER protein (Fig. 1). Patients with alterations in CSB show early aging (Table 3). p53 affects the helicase step of NER and is also required for apoptosis when damage requiring NER is excessive (Fig. 1). Defects in p53 cause early aging (Table 3). Thus, at least four NER enzymes (XPD, CSB, p53, ERCC1), when defective, may allow helix-distorting DNA damages to accumulate more rapidly than normal, leading to premature aging. In addition, the p53-regulated pathway of apoptosis (Fig. 1), perhaps triggered by helix-distorting DNA damages, if defective in p66Shc, results in life span extension (Table 2).

2.2 BER (Base Excision Repair)

BER protects mammalian cells against single-base DNA damage by methylating agents, most oxidative damages, and a large number (about 9,000 per cell per day – see Table 1) of spontaneous depurinations. BER is mediated through at least two subpathways, one involving removal and replacement of a single nucleoside and the other involving a longer patch repair of 2 to 15 nucleotides. Repair

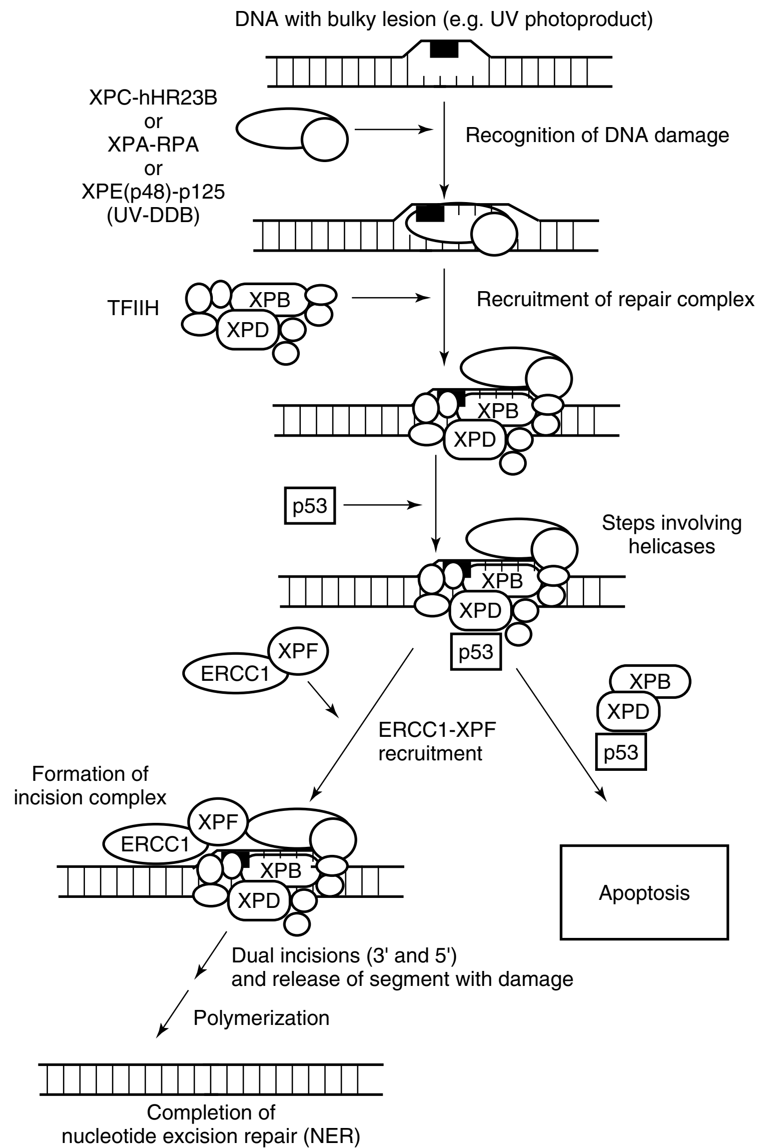


Fig. 1 Nucleotide excision repair (NER).

can be initiated by removal of a damaged base by a DNA glycosylase, which binds the altered deoxynucleoside in an extrahelical position and catalyzes cleavage of the base–sugar bond, or can occur at a site of spontaneous depurination [an apurinic (AP) site] (Fig. 2). The glycosylases that remove damaged bases are specific to the particular damage. Ref-1 (also called apurinic/apyrimidinic endonuclease or APE) then makes a 5′ nick in the DNA backbone, followed by PARP, 1 acting as a nick surveillance protein. PARP locates at the site of the single chain

nick. The nick then activates PARP, which catalyzes formation of poly(ADP-ribose) (polyAR) branched chain polymers, using the adenosine diphosphate ribose part of nicotinamide adenosine dinucleotide (NAD^+) with the release of nicotinamide (NAM) (Fig. 2). PolyAR is attached to numerous nearby proteins, including PARP itself and p53. Proteins modified by polyAR gain or lose functions. Fig. 2 indicates alteration in the transactivation role of p53 and its role in cell cycle arrest and apoptosis after it is modified by polyAR. Subsequently, a multiprotein complex (including XRCC1 and possibly DNA polymerase β and DNA ligase III) is recruited to the site by PARP, and repair patch synthesis and DNA ligation complete the process of BER (Fig. 2).

Individuals with Cockayne Syndrome show premature aging. Mutant cells with a defect in helicase motifs V or VI of Cockayne syndrome B (CSB) are defective in BER of one common oxidized base, 8-hydroxyguanine, but not defective in BER of two other oxidized bases, thymine glycol or 5-hydroxy-dCytosine. CSB is involved in the specific glycosylase step for BER of 8-hydroxyguanine (Fig. 2).

Centenarians (individuals who have lived for more than 100 years) appear to have an altered PARP with higher specific

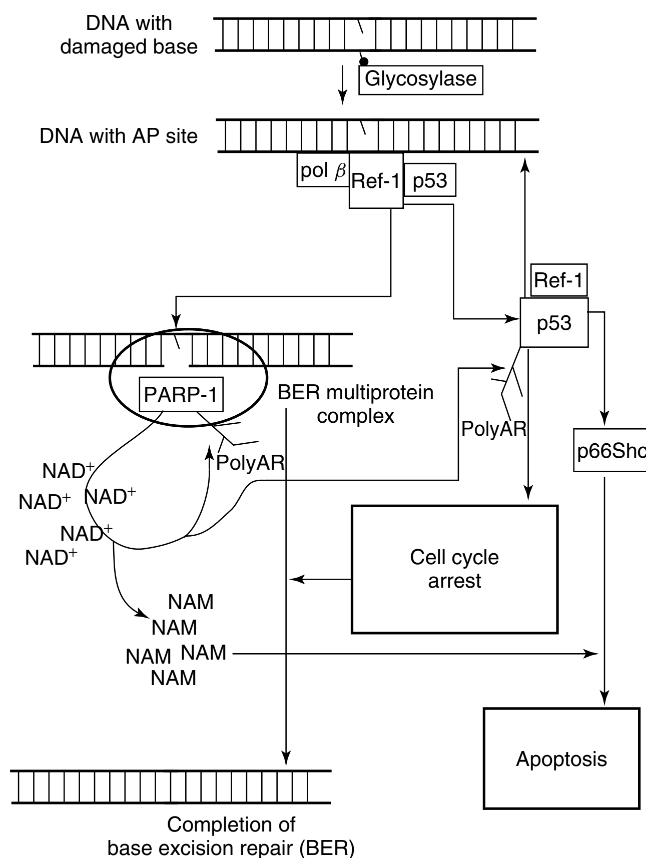


Fig. 2 Base excision repair (BER).

activity (Table 2). PARP has a central role in recruitment of the BER multiprotein complex (Fig. 2). p53 stimulates BER through interaction with ref-1 and DNA polymerase β (Fig. 2). Mice defective in p53 show early aging (Table 3).

p66Shc, when absent, causes life span extension of mice raised under protected laboratory conditions, and also longer survival of mice challenged by the oxidatively damaging compound paraquat. Mouse embryo fibroblast cells that are defective in p66Shc, when treated with H_2O_2, have much less apoptosis than wild-type cells treated with H_2O_2. This would be expected from the known role of p66Shc in the p53-controlled apoptosis pathway shown in Fig. 2. In addition, a defect in p66Shc causes reduced intracellular constitutive levels of ROS in lung, spleen, liver, and skin cells, where p66Shc is normally active. These reduced ROS levels are reflected in lower levels of oxidative DNA damage both in nuclear DNA and mitochondrial DNA of these tissues (both types of DNA damage normally repaired by BER).

Thus, two gene products involved in carrying out BER (CSB and p53), when defective, cause early aging. One gene product essential for BER, PARP, when more active, allows longer life span. One gene product, p66Shc, causes more DNA damages, needing BER when it is active. When p66Shc is absent, there are fewer DNA damages and there is life span extension.

2.3 HRR (Homologous Recombinational Repair)

Double-strand breaks, interstrand cross-links, and DNA damages blocking a replication fork can be repaired by one of two pathways, HRR or NHEJ. The selection of pathway depends on whether the cell has replicated its DNA (the S/G_2 phase of the cell cycle), so that sister chromatids are available (which favors HRR), or whether the cell is in G_0/G_1 (which favors NHEJ). The choice of HRR or NHEJ also depends on whether the cell is from embryonic tissue (which favors HRR) or from adult tissue (which, for certain damages, favors NHEJ). While HRR is especially promoted during meiosis (where there is ordinarily pairing and recombination between homologous chromosomes), HRR is also important in somatic cells. Defects in enzymes of HRR, including WRN (Werner syndrome), the RECQ3 helicase, and BLM (Bloom syndrome), the RECQ2 helicase, cause genetic instability and cancer (Table 3).

HRR is initiated through a multimeric complex, the BRCA1-associated genome surveillance complex (BASC). BASC includes the DNA repair proteins BRCA1, MSH2, MSH6, MLH1, ATM, BLM, and the RAD50-MRE11-NBS1 protein complex, some of which (BRCA1, ATM, and BLM) are indicated in Fig. 3. BASC is thought to act as a sensor for DNA damage, to which it binds. Depending on the type or amount of double-strand damage, BASC signals for (1) initiation of further steps of HRR, (2) interaction with the "master switch p53" to turn on cell cycle arrest (to allow more time for HRR), (3) apoptosis through the p53/p66Shc pathway, or (4) apoptosis through the p73 pathway (Fig. 3). As indicated in Fig. 3, BRCA1 and p53 have a type of feedback loop whereby BRCA1 may activate p53, and p53 may, in turn, inhibit BRCA1.

Further steps of HRR, shown in Fig. 3, involve (1) recruitment of a second homologous chromosome, (2) Rad51-dependent

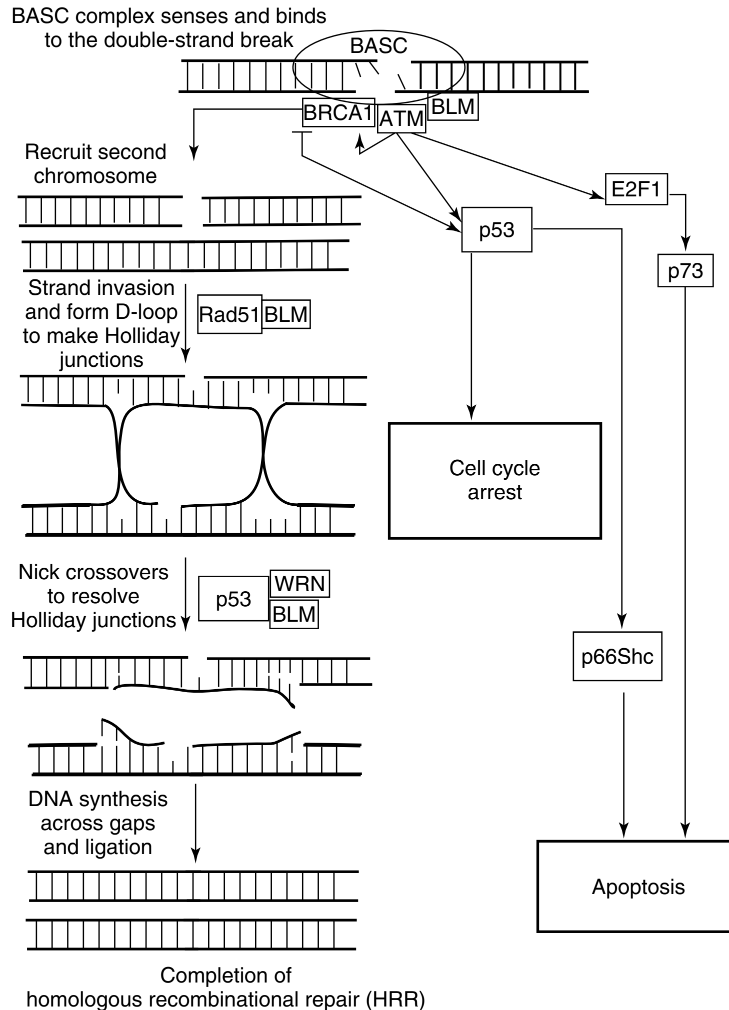


Fig. 3 Homologous recombinational repair (HRR).

strand invasion from the ends of the double-strand break plus formation of a "D-loop" in the invaded chromosome by strand migration facilitated by the helicase action of BLM, (3) nicking of the crossover strands to resolve the Holliday junctions, carried out by BLM and WRN helicases in association with p53 (which has a 3′ to 5′ exonuclease function and a strand transfer function that may act at this step), and (4) DNA synthesis across single-strand gaps and a final ligation step. An interstrand cross-link will be converted to a double-strand break, in part by action of ERCC1 (Table 3), and thus can also be repaired by this pathway.

Modified forms of HRR have been suggested for replicating somatic cells. In these models, the "D-loop" (the strand of DNA from the intact chromosome

that has been unwound and looped out to lend its information to cover the gap in the broken chromosome) is not cut to resolve the crossovers. These models, called the *synthesis-dependent strand annealing* or *migrating D-loop models*, were proposed to account for the prominence of mitotic gene conversion without reciprocal exchange observed in replicating cells. The standard model of HRR is shown in Fig. 3, however, since it applies to nonreplicating somatic cells, the major cells showing decreased function with age, and to germ line cells, as discussed below.

As indicated in Table 3, four genes that act directly in HRR (*BLM*, *WRN*, *p53*, and *ERCC1*), when defective, cause early aging. Oxidative damage to DNA includes double-strand breaks. Thus the gene product, *p66Shc*, which increases the level of intracellular ROS, may cause more DNA damage, needing HRR when it is active. When *p66Shc* is absent, there are fewer such DNA damages and there is life span extension.

2.4
NHEJ (Nonhomologous End Joining)

As noted in Section 2.3, double-strand breaks, interstrand cross-links, and DNA damages blocking a replication fork can be repaired by one of two pathways, HRR or NHEJ. HRR is an accurate repair pathway, obtaining information missing within a damaged chromosome from a homologous, undamaged chromosome. NHEJ, alone, however, is inaccurate at the nucleotide level because it involves end-joining reactions between single-stranded ends to form junctions containing regions of microhomology of 1 to 10 base pairs within 20 base pairs of the ends. However, coupled homologous and nonhomologous repair can occur (using short homologous regions on a nearby heterologous chromosome) to employ NHEJ in a more accurate process, with short regions of gene conversion from the undamaged chromosome. In addition, it was recently shown that NHEJ, when the number of double-strand breaks is small, serves to properly reconnect the two broken ends of a given chromosome. In the absence of NHEJ, more misrejoined chromosomes (chromosomal aberrations) are produced.

NHEJ of a double-strand break is illustrated in Fig. 4. Two Ku heterodimers (Ku consists of Ku70 complexed with Ku86) attach to two broken ends of a chromosome. Ku then recruits DNA-PK$_{CS}$, which becomes activated by the interaction with Ku and a DNA end. Activated DNA-PK$_{CS}$ phosphorylates a wide range of DNA binding proteins and also phosphorylates Werner syndrome protein (WRN). WRN displaces DNA-PK$_{CS}$ and, through interaction with Ku, activates its WRN exonuclease function, trimming back single strands on the chromosome ends. The complex of Mre11, Rad50, and NBS1 (MRN), plus BRCA1, may interact at this point with Ku to keep the end joining fairly accurate. In the absence of WRN, frequent gross chromosomal rearrangements occur. The Bloom syndrome protein, BLM, is also needed to keep the NHEJ fairly accurate. BLM may also assist with the alignment of two nearby ends for fairly accurate NHEJ (BLM is drawn with a dashed oval in Fig. 4 to indicate that its role in the NHEJ repair process is not yet clearly established). Pol μ then associates with small gaps in the aligned and stabilized joined ends, and pol μ, in association with Ku, recruits ligase IV and XRCC4 (X-ray repair cross complementing protein 4). Pol μ fills in

the gaps and ligase IV and XRCC4 complete the NHEJ repair of the double-strand break in a chromosome.

As indicated in Table 3, three genes that act directly in NHEJ [*Ku80* (also called *Ku86*), *WRN* and *BLM*], when defective, cause early aging. *Ku80* is only active in NHEJ, while *WRN* and *BLM* also have roles in HRR.

2.5
MGMT (O^6-Methylguanine-DNA Methyltransferase)

As indicated in Table 1, O^6-methylguanine is a frequent DNA damage. O^6-methylguanine is specifically repaired by a DNA repair enzyme called *O^6-methylguanine-DNA methyltransferase* (MGMT). As

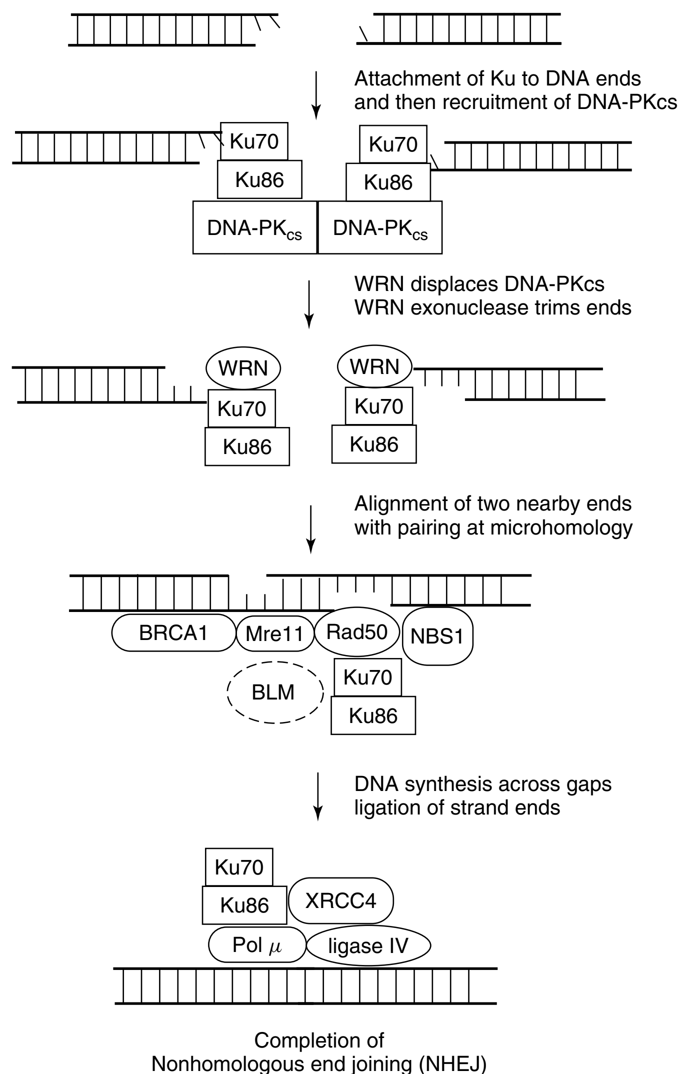


Fig. 4 Nonhomologous end joining (NHEJ).

indicated in Table 2, excess MGMT results in decreased aging (life span extension).

2.6 Enzymes of DNA Repair Pathways and Enzymes Regulating DNA Damage–inducing ROS Contribute to Determination of Aging

Eleven specific DNA repair enzymes distributed among NER, BER, HRR, NHEJ and MGMT were found to increase lifespan when elevated or decrease life span when defective (Tables 2 and 3). In addition, alterations of four enzymes causing reduced levels of DNA-damaging ROS (increases of Cu/ZnSOD, MnSOD or catalase, or loss of p66Shc) also increase life span. These findings strongly support the DNA damage theory of aging.

3 The DNA Repair (and Complementation) Theory of Sex

3.1 Meiosis, an Adaptation for Repairing Germ Line DNA

It is useful for further discussion to define sexual reproduction and to clarify its key components. Such a definition, to be general, should encompass sexual processes in all organisms, both eukaryotes and prokaryotes. *Sexual reproduction is the process by which genetic material (usually DNA) from two separate parents is brought together in a common cytoplasm where recombination of the genetic material ordinarily occurs, followed by the passage of the recombined genome(s) to progeny.* Thus, sexual reproduction has two basic elements: (1) recombination, in the sense of the exchange of genetic material between two homologous chromosomes, and (2) outcrossing, in the sense that the homologous chromosomes from two different individuals come together in the same cell.

In eukaryotes, the germ line is the cellular lineage that connects successive meioses. The germ line is a key feature of sexual reproduction. In obligate sexual organisms such as humans, the germ line is potentially immortal (see Section 1.9) in contrast with the somatic line. This feature of the germ line presumably is due to adaptations present in the germ line that are lacking in the somatic line. Meiosis is a process unique to the germ line, and it appears to be an adaptation specifically for promoting recombinational repair (HRR), since a major characteristic of meiosis is the close pairing of homologous chromosomes.

Some diploid organisms form gametes (such as sperm and egg cells) by meiosis, but then undergo self-fertilization or automixis (a process in which two haploid products of meiosis fuse to form a diploid zygote). Such processes are common in plants and invertebrates. Self-fertilization and automixis are not strictly sexual processes since they lack the outcrossing feature of sex. However, many of the organisms that undergo these asexual processes are also facultatively sexual. In the following discussion, we assume that meiosis serves essentially the same function in all organisms in which it occurs and that the germ line in facultative or obligate self-fertilizing and automictic eukaryotes is the cell lineage that connects successive meioses.

Tab. 5 Modes of reproduction among higher plant species.

Reproduction strategies	*Number*	*Percent*
Principally cross-fertilized	830	55
Self-incompatible, dichogamous, or dioecious		
Dichogamy: male and female parts mature at different times;		
Dioecious: unisexual (male or female) flowers on different plants		
Partially self-fertilized, partially cross-fertilized	105	7
Principally self-fertilized (autogamous)	229	15
Facultative apomicts (can be cross-fertilized)	16	1
Apomixis: vegetative methods of reproduction including propagation by runners or bulbs and agamospermy		
Agamospermy: embryos and seeds formed by asexual means		
Facultative apomicts, can be self- or cross-fertilized	3	0.2
Facultative apomicts, can be self-fertilized	2	0.1
Apomicts, not known if also self- or cross-fertilized	199	13
Obligate apomicts	121	8

3.2 Frequency of Sexual Reproduction

Sexual reproduction is a widespread strategy for reproduction. About 99.9% of the approximately one million known animal species are sexual. Among higher plants, the majority of species are sexual. As shown in Table 5, only 8% of higher plants are known to be obligate apomicts that reproduce only by vegetative means such as by runners, bulbs, or by asexual formation of seeds. Sex is also common among the simple eukaryotes, including fungi, algae, and protozoa. Sex is found among bacterial species as well, and is common in bacterial viruses and animal viruses.

3.3 Costs of Sex

Sex, while widespread, is very costly to the organism using it. For example, a sexual female lizard passes only 50% of her genes to a particular egg, while a comparable nonsexual (parthenogenetic) female lizard passes 100% of her genes to each egg. Thus, a sexual female is only half as efficient in propagating her genes as a nonsexual female (all other factors being equal). In addition, when two individuals must find each other to mate, there is a cost of searching out the other party.

Another cost of sex arises from the randomization of genetic information during meiosis. A parent organism, which has met the test of survival, has by definition, a well-adapted combination of genes. The process of meiosis, which includes recombination, generates untested new combinations of genes to be passed on to progeny. These new combinations, on average, should be less successful than the parental combinations of genes because random changes in successful genetic information are more likely to be deleterious than beneficial.

The noted biologist E.O. Wilson describes sex in humans as a "gratuitously consuming and risky activity". Reproductive organs are anatomically

complex in humans and these are vulnerable to such problems as ectopic pregnancy and venereal diseases. Further, courtship activities are costly. Even at the microscopic level in humans, genetic processes for determining sexual development are easily perturbed with one sex chromosome too few or too many, or a shift in the hormone balance of a fetus, causing abnormalities in physiology and behavior.

Clearly sex has large costs.

3.4 Benefits of Sex

Sex must have a large benefit to make up for its large costs. A major function of sex is to counteract two types of "noise" in the transmission of genetic information from parent to progeny: DNA damage and mutation.

As pointed out in Section 1.1, about 95,000 DNA damages occur, on average, per day per cell in the rat. Most of these are single-strand damages that can be removed by excision repair or other repair processes that only need the redundant information present on the opposite strand of DNA. However, excision repair of single-strand damages is not 100% efficient, and double-strand damages also occur at significant frequencies (Table 1). Such DNA damages remaining in germ cells (e.g. egg and sperm, in mammals) would cause the death of zygotes and loss of potential progeny. During the meiotic stage of the sexual process, however, HRR is strongly enhanced by the systematic homologous pairing of chromosomes, the major feature of meiosis. This enhanced HRR repair is one major benefit that can compensate for the costs of sex, since clearing the germ line of lingering damages can greatly increase viability of progeny.

The other type of noise in the transmission of genetic information from parent to progeny is mutation. Mutation is quite different from DNA damage. A mutation is a change in the DNA sequence rather than a change to a deformed DNA structure (DNA damage). Mutation and DNA damage have distinctly different consequences. Mutations can be replicated (when DNA replicates) and thus can be inherited, while a damage cannot be replicated. However, damages can be recognized by enzymes and repaired, whereas mutation cannot be recognized and thus cannot be repaired.

Genes carrying mutations often code for nonfunctional proteins. If one of a pair of chromosomes in a diploid cell carries a mutation in a given gene, and the second chromosome carries the homologous gene in a functional form, this second gene can usually provide an adequate level of gene expression for the organism to function normally. This masking of the expression of mutant genes by wild-type genes is called *complementation*. Complementation is available in the diploid phase of the life cycle of organisms. However, complementation is most beneficial when an organism undergoes outcrossing.

To see why mutation makes it beneficial to have outcrossing, consider a hypothetical population of diploid organisms that is strictly inbreeding (all of the organisms are self-fertilizing) and assume that the population has been long established. In such a population, the rate at which new mutations arise will be balanced by the rate at which they are lost from the population by natural selection. In this self-fertilizing population, each mutation present in an individual will have a one in four chance of being paired with

the same mutation (becoming homozygous recessive) in each progeny. If an average of one to several deleterious mutations are present in each individual, the cost of inbreeding in terms of defective progeny would be high. If a hypothetical outcrossing individual should arise in such an otherwise inbreeding population, any mutations in this outcrosser would very likely be complemented by the wild-type alleles from its mating partner and thus defective progeny would be avoided. Complementation would occur because the homologous chromosomes from the two parents are not likely to carry the same mutations. Thus, loss of progeny due to expression of mutations would be greatly reduced. This gives a strong immediate selective advantage to switching to outcrossing from inbreeding.

This advantage would not last indefinitely in our hypothetical example. Because of the ability of outcrossers to mask deleterious mutations by complementation, mutations would not be weeded out by natural selection as efficiently as in inbreeding individuals. Eventually, the mutations that build up in the population will cause as much lethality to progeny of the outcrosser as that in the original inbreeding population. However, if the outcrosser tried to switch back to inbreeding, there would be a great loss of progeny due to the larger number of mutations now present. In summary, there is a large immediate benefit in switching from inbreeding to outcrossing and a large immediate disadvantage in switching back. Therefore, mutation provides a selective pressure to maintain the outcrossing feature of sexual reproduction among diploids.

Overall, meiosis, with its promotion of recombinational repair, may be the only way to efficiently correct endogenous double-strand damages and leftover single-strand damages in the diploid cells that produce germ cells. On this view, the recombination aspect of sex is an adaptation for dealing with DNA damage, a major type of genetic noise. Furthermore, outcrossing allows the masking of mutations through complementation. The outcrossing aspect of sex deals with mutation, the second major type of genetic noise.

3.5
Repair of DNA Damage of the Germ Line in Nonmeiotic Cell Divisions

The germ line is characterized by periodic events of meiosis, but during the intervals between such events, cell divisions are ordinarily by mitosis. Cells of the germ line are presumably capable of the same types of repair processes that occur in somatic cells, discussed in Section 2. For instance, there is a relatively high level of PARP activity in premeiotic and meiotic spermatocytes, an indication of BER. Fraga and coworkers in 1990 measured the accumulation of one type of oxidatively damaged DNA base, 8-hydroxy-2′-deoxyguanosine, in various tissues of the rat. Although the average accumulation in the rat kidney was 80 residues per cell per day, there was no detectable accumulation in the testes. This lack of accumulation in the testes of normal individuals can be interpreted as a reflection of efficient BER in the germ line of normal individuals.

It is notable, in Table 2, that where life span is extended, fertility is either normal or extended as well, where it has been tested. In Table 3, where defects in DNA repair cause early aging, fertility is always reduced, where it has been tested.

3.6
Meiotic Recombination in Eukaryotes Probably Evolved from Recombination in Prokaryotes

Several workers have suggested that the processes of recombination in prokaryotes and eukaryotes share a common ancestry. Dougherty in 1955 may have been the first to conclude that the evolution of sexuality as it exists today was the result of a single phylogenetic sequence. He based this conclusion on the fact that recombination in bacteriophage and in bacteria, on the one hand, and meiotic recombination in eukaryotes, on the other hand, seemed to share fundamental similarities. Later Stahl in 1979 also concluded that despite numerous differences in detail, the "similarities in recombination in creatures as diverse as the phage and fungi are impressive."

In recent years, there has been considerable work on the biochemistry of recombination, with an emphasis on the RecA protein of the bacterium *Escherichia coli*. This protein catalyzes the key steps in recombination of homologous DNA pairing and strand exchange. Homologs of the *E. coli RecA* gene have been identified in over 60 bacterial species and in bacteriophage T4, suggesting that RecA-catalyzed recombination is very common in the prokaryotic world. Since about 1992, there has been much work indicating that RecA homologs play a key role in meiotic recombination in fungi and vertebrates. A RecA homolog in humans shows 30% amino acid sequence identity with the *E. coli* RecA protein. RecA homologs in yeast and humans form helical filaments with DNA, such as those formed by *E. coli* RecA, leading to the conclusion that the RecA protein has been conserved from bacteria to man. These findings suggest that eukaryotic meiotic recombination and recombination processes in extant bacteria are probably both derived from a common ancestor that existed before the divergence of prokaryotes and eukaryotes, at least 1.8 billion years ago.

3.7
Homologs of the Bacterial RecA Protein Have a Key Role in Eukaryotic Meiotic Recombinational Repair

Substantial evidence indicates that RecA homologs have a central role in recombinational repair during meiosis in yeast. The yeast *rad*51 and *dmc1* genes are homologs of the *recA* gene of *E. coli*. The Rad51 and Dmc1 proteins probably share redundant functions, since recombination is reduced by only a few fold in *dmc1* and *rad51* single mutants, but *dmc1* and *rad51* double mutants are profoundly defective in meiotic recombination. The Rad51 protein acts during both mitosis and meiosis, whereas Dmc1 protein acts only during meiosis. Both Rad51 and Dmc1 proteins function in repair of double-strand breaks. Sung in 1994 demonstrated that Rad51 protein, like *E. coli* RecA protein, catalyzes ATP-dependent homologous DNA pairing and strand exchange. The Dmc1 protein has an overall similarity to *E. coli* RecA protein in tertiary structure. These findings suggest that the yeast Rad51 and the bacterial RecA proteins have similar functions in recombinational repair. Morita and coworkers showed in 1993 that in the mouse a *recA* homolog is expressed at high level in the testes. Shinohara and coworkers in 1993 also showed that the mouse *recA* homolog is expressed at a high level in the testis and ovary, and suggested that the protein product of this gene is involved in meiotic recombination. The expression of a *recA* gene homolog has also been demonstrated in chicken testis and ovary and in human

testes. The human RecA homolog carries out the distinctive reactions of *E. coli* RecA protein, including DNA-dependent hydrolysis of ATP, renaturation of complementary strands, homologous pairing of a single strand with duplex DNA, and strand exchange. A pair of *recA* gene homologs of the lily, *lim15* and *rad51*, which are also homologs of the yeast *dmc1* and *rad51* genes, were found by Terasawa and coworkers in 1995. The *lim15* gene in lily is specifically expressed in meiotic prophase during microsporogenesis. Thus, in both animals and plants, homologs of bacterial *recA* appear to have an important role in meiotic recombination.

3.8
The Adaptive Function of Recombination Appears to Be Removal of DNA Damage

DNA repair is probably the principal adaptive function of the RecA protein and its homologs. RecA protein binding is largely limited to regions in the DNA containing suitable nucleation sites, especially single-strand gaps. A variety of DNA damages cause structural perturbations that provide favorable nucleation sites.

About 100 ATPs are hydrolyzed for every base pair of heteroduplex DNA generated by RecA-mediated strand exchange. In 1993, Cox reviewed evidence that the energy released by ATP hydrolysis is used specifically to allow the strand-exchange process to traverse damaged regions of DNA. He argued that this use of ATP is readily understood as an adaptation of repair of DNA damage. Thus, the evidence reviewed by Cox indicates that the adaptive function of RecA homologs, acting during meiosis, is DNA repair.

In humans, HRR is carried out by a group of interacting proteins in which the RecA homolog Rad51 has a central role. Other proteins involved in HRR in humans include the products of genes *BRCA1*, *BRCA2*, *ATM*, *ATR*, *FANCD2*, *HMLH1*, and *p53*.

These proteins appear to be responsible for HRR during meiosis and for HRR between sister DNA homologs in somatic cells. HRR is employed in removing a variety of DNA damages, particularly double-strand damages such as double-strand breaks and DNA cross-links. The proteins named just above, involved in HRR, also appear to use their DNA damage recognition capability to induce apoptosis (a form of programmed cell death) when the number of DNA damages in a cell is higher than those that can be repaired. In somatic cells, apoptosis is a protective mechanism for the whole organism since it eliminates cells that might otherwise survive despite having unrepaired DNA damages, and upon replication acquire mutations.

In fact, germ line mutations in genes *BRCA1*, *BRCA2*, *ATM*, *ATR*, *FANCD2*, *HMLH1*, and *p53* predispose individuals to cancer. Cancer predisposition results from an inadequate response to DNA damage in somatic cells leading to increased mutations, some of which cause progression to malignancy. Thus, the central function of these genes is to cope with DNA damages, either to repair the damages, or if their number in a cell is unmanageable, to induce apoptosis.

Males that are defective in *ATM* or *p53* are unable to respond appropriately to DNA damage during meiosis, and, as a likely result, they suffer from low sperm quantity and quality. The high effectiveness of DNA damage removal processes during normal meiosis is reflected in mice, where there is a one-third lower rate of mutation in germ cells compared to that in somatic cells.

Overall, the evidence reviewed in this section lends strong support to the idea that the primary general function of recombination enzymes, whether in somatic cells or during meiosis, is to remove DNA damage, either by repair or apoptosis.

3.9
Other Expectations of the DNA Repair and Complementation Theory of Sex

If DNA damage and its repair are important in maintaining sexual reproduction, then certain expectations follow. If recombination during meiosis and sexual reproduction reflect recombinational repair of germ line DNA, then there should be other evidence of avoidance of DNA damage in the germ line as well.

As indicated in Table 1, the largest known source of DNA damage is oxidative damage. Such damage occurs due to endogenous cellular metabolism. Presumably, to avoid DNA damage, germ line cells should have evolved ways to avoid high levels of metabolism. Eggs would seem, at first sight, to be poor candidates for avoiding metabolism. They are, in general, much larger than somatic cells of the organism (e.g. egg cells have about a 1,000-fold greater mass than somatic cells in humans). It requires considerable metabolism to produce large egg cells. However, much of the cytoplasmic material within an egg cell is generated by the activity of other cells. Some insects have nurse cells around each egg cell. These nurse cells are connected to the egg cell by cytoplasmic bridges and provide most of the ribosomes, mRNA, and proteins of the egg cell. The nurse cells themselves contain hundreds to thousands of copies of their genomes, presumably to protect the nurse cells themselves from losing function from the oxidative damage they may suffer while providing large amounts of metabolic products for the egg cell. Vertebrate egg cells are surrounded by follicle cells rather than nurse cells. The follicle cells do not have cytoplasmic bridges to the egg cell, but rather have small gap junctions connecting them to the egg. While these gap junctions are not large enough to transmit bulky macromolecules, they do transmit precursor molecules to the egg. In addition, for chickens, amphibians, and insects, the yolk proteins accumulated by the egg are made in liver or liver-type cells. These mechanisms allow eggs to be protected from oxidative damage while they store up material to sustain the zygote in its initial growth.

Sperm or pollen cells, in contrast to egg cells, are usually the smallest cells of an animal or plant. This allows a different strategy for effective protection against oxidative damage to their DNA. Because of their very small size, minimal metabolism would have been used in their formation. Thus, both egg and sperm production appears to have been adapted to circumvent the production of DNA damage in their especially important germ line DNA.

Another way to avoid production of sperm with damaged DNA is for spermatogenic cells with DNA damage to undergo cell cycle arrest to allow more time for meiotic repair, and then, if this fails, to undergo apoptosis. The p53 protein plays a key role in mediating cell cycle arrest in response to DNA damage. Low-level irradiation was found to activate a p53-dependent premeiotic delay, allowing time for increased DNA repair leading to increased motile spermatozoa. Higher levels of γ-irradiation induced p53-independent apoptosis during meiosis.

If complementation of mutations is important in maintaining the outcrossing

aspect of sex, this should be consistent with general biological observations as well. Indeed, both in animals and in plants, it is usually seen that when hybrids are formed from the crossing of two genetically distinct inbred lines, these hybrids are more vigorous than either of their two parental lines. This hybrid vigor is responsible for much of the crop improvement that has been achieved in modern agriculture. The opposite side of this observation is the fact that consanguineous marriages, in humans, result in an increased frequency of impaired offspring. Observations in other animals and in plants suggest that close inbreeding results in the production of less vigorous progeny. This inbreeding depression appears to be due largely to expression of deleterious recessive mutations and reflects in part the cumulative effect of numerous mildly deleterious mutations.

4
Vegetative Survival Strategies

4.1
Survival of Vegetative Cell Populations

Vegetatively growing populations of bacteria can be regarded as potentially immortal as long as nutrient resources are abundant. However, even in such populations there appears to be constant attrition due to DNA damage and deleterious mutation.

Flowering plants generally reproduce sexually, with gamete formation by a meiotic process followed by gamete fusion as the prelude to embryogenesis and seed formation. Plants, unlike most animals, are also able to generate complete new individuals of similar genetic constitution from vegetative parts. Meristematic buds or excised pieces of tissue can propagate in the appropriate environment. Apparently, plant vegetative cell lines can be maintained indefinitely under appropriate conditions. These lines probably maintain themselves by a strategy of replacement, where cells with lethal unrepaired DNA damage or expressed deleterious mutations die and are replaced by replication of nondefective cells. Nevertheless, in some plant tissues DNA damage may accumulate. For instance, in dry seeds, fragmentation of nuclear DNA occurs with time.

Most forest trees live for at least 100 years, many of them for more than 300 years, and a few survive for more than 1,000 years. Clonal tree species may occupy a location for several thousand years. Most of the tree is dead, and only a thin shell of dividing cells (cambium) occurs around the trunk and in the leaves. A tree actually represents a free-living clone of cells in which selective removal of cells with irreversible accumulated damage is constantly occurring. One would not expect to find old cells in a tree any more than one would find old cells in a growing culture of bacteria. The evidence discussed in this section suggests that some proliferating cell populations can cope with unrepaired DNA damage by a replacement strategy, which can be maintained indefinitely as long as nutrient resources are abundant and the level of unrepaired DNA damage is not excessive.

4.2
Vegetative Survival Strategy for Mitochondria (and Chloroplasts) in the Germ Line

An analog of cellular replacement at the molecular level (“molecular replacement”) facilitates the purging of both DNA damage and deleterious mutations in cytoplasmic genomes (mitochondria and

chloroplasts) of germ cell lines. While nuclear genes occur as single copies per gamete, there are thousands of mitochondrial DNA (mtDNA) molecules in most cells, and several hundred thousand may occur in a mature oocyte. The many mtDNAs in oocytes appear to stem from a vastly smaller pool of mtDNA molecules that must have survived a process of replicative segregation in earlier cytokinetic divisions of the germ line lineage, since most heterogeneity of mtDNA is distributed among, rather than within, individuals. This implies that there are mtDNA population bottlenecks in germ lines. That is, the mtDNA, which is generally solely transmitted from one generation to the next through the oocyte, is of only one or a few genotypes. Avise concluded that the mtDNA molecules that survive and replicate to populate a mature oocyte probably have been scrupulously screened by natural selection for replicative capacity and functional competence in the germ cell lineages they inhabit. This strategy, at the molecular level, is equivalent to a vegetative cell replacement strategy for cell lineages that do not have a sexual cycle. While the mitochondria and chloroplasts within cells follow a vegetative replacement strategy, they, like vegetatively replicating bacteria and yeast, utilize homologous DNA recombinational repair to remove DNA damages due to oxidation or environmental stresses.

4.3
Dolly the Cloned Lamb, Cumulina the Cloned Mouse, and Low Success of Cloned Mammals

The evidence reviewed above indicates that germ cells, having undergone meiosis, are relatively free of DNA damage, and thus should be able to give rise to viable offsprings with high probability. At lower probability, a given somatic cell nucleus may also be sufficiently free of damage to be able to produce viable progeny if transferred to an enucleated egg cell.

Eight viable lambs were derived in 1997 – one (named Dolly) from donor cells of a mature sheep, four from donor embryo-derived cells, and three from donor fetal fibroblast cells. For Dolly, the lamb derived from a somatic cell of a mature sheep, they first made 277 fused couplets (enucleated oocytes fused to donor cells), using donor cells from their 3rd to 6th passage, cultured from the mammary gland of a six-year-old pregnant ewe. A morula or blastocyst was able to form from only 11.7% of the cultured fused couplets from the mature ewe. By comparison, when couplets were derived from embryo or fetal fibroblast cells, a morula/blastocyst was formed 27 to 39% of the time. Some of the morula/blastocysts implanted in recipient ewes formed fetuses detectable by ultrasound at 50 to 60 days. Subsequently, 62% of these fetuses were lost, a much greater proportion than the estimated 6% after natural (meiosis based) mating in sheep.

Similar observations were made with mice. Ten healthy mice were cloned, the first of which was named Cumulina, from donor nuclei of differentiated, nonreplicating granulosa cells of mature mice injected into enucleated recipient mouse oocytes. Cumulina and the other nine cloned healthy newborn mice were the only successful progeny produced from 800 injected oocytes, which had formed embryos and were transplanted into foster mother recipient mice. In multiple cloning experiments by these authors, the rate of successful implantation of embryos was 57

to 71%, formation of fetuses was 5 to 16%, and full-term development was 2 to 3%.

The low survival rate to full development in both sheep and mice could have three possible explanations. First, injuries introduced by the experimental manipulation of the embryos before implantation might be deleterious to further development. Even unreconstructed embryos experience some increased prenatal loss after manipulation or culture. Second, the differentiated donor nuclei, transplanted into the recipient oocytes, have to reprogram their developmental clock to zero, and errors in this process may have deleterious effects on fetal development. Third, the low survival of fetuses may be a consequence of the greater amount of DNA damage in donor somatic cells (than in meiotically produced gametes), which can give rise to deleterious mutations when replicating after transfer to an oocyte.

The gene expression profile of about 10,000 genes in the placentas and the livers of surviving cloned mice derived by nuclear transfer were obtained in 2002. The transferred nucleus came from either an embryonic stem (ES) cell or from a mature, differentiated cumulus cell. They compared those gene expression profiles (derived from ES nucleus or cumulus cell nucleus) with the gene expression profile in the same tissues (placenta, liver) of healthy mice derived from normal mating. These comparisons showed that several hundred (at least 4%) of the expressed genes had pronounced dysregulation in the cloned mice, probably accounting for the altered phenotype of cloned individuals (many cloned mice were obese, etc.). Their evidence pointed to difficulty in reprogramming the developmental clock back to zero, in the low percentage of surviving mice.

In humans, about 50 to 80% of all natural meiosis-based conceptions fail to result in live birth. Cytogenetic studies of spontaneous and induced abortions and on perinatal deaths indicate that many types of chromosome abnormalities are present in these failed conceptions. Chromosome abnormalities often derive from DNA damages.

Thus, DNA damage may be a serious problem for germ cells despite available mechanisms for avoiding and repairing such damage, and this problem may be considerably greater in nonmeiosis-derived conceptions.

5 Three Levels of Sexual Communication Reflect (1) DNA Repair, (2) Complementation, and (3) Selection for Fitness

Sexual communication occurs when signals are used to promote or modulate sexual interaction between individuals. Sexual communication is prevalent among organisms from bacteria to man. Sexual communication occurs at three levels. Level 1: the first level of sexual communication includes signals that increase the likelihood that two organisms will come together for sexual interaction. Level 2: the second level involves signals that modulate the sexual interactions to inhibit inbreeding or facilitate outbreeding. Level 3: the third level includes signals that further modulate the sexual interactions to promote selection among potential mating partners based on relative fitness. Evidence indicates that the selective advantages of the three levels of sexual communication are, respectively, (1) the repair of DNA damage, (2) the masking of mutation, and (3) the choice of a fit mating partner.

5.1 Sexual Communication in Bacteria, Primarily for DNA Repair

The simplest organisms in which sexual communication occurs are bacteria. Bacterial transformation is a form of sexual interaction in bacteria and has been shown in several bacterial species to involve sexual communication. Transformation involves the transfer of naked DNA from one member of a bacterial population to another through the surrounding medium. Transformation occurs naturally in a wide range of bacterial species. The bacteria that take up transforming DNA ordinarily incorporate this DNA into their genomes by recombination. Transformation is an evolved, rather than an incidental, trait since it results from a complex, energy-requiring, developmental process. For a bacterium to bind, take up, and recombine exogenous DNA into its genome, it must first enter a special physiological state referred to as *competence*. Transformation involves expression of genes required for competence and for recombination. In the completely sequenced genome of *Haemophilus influenzae*, transformation is promoted by at least 15 genes of the 1007 for which role assignments can be made. In another bacterium, *Bacillus subtilis*, about 40 genes necessary for competence have been identified.

Transformation in a number of bacterial species depends on the production of an extracellular factor (referred to as *competence factor*), which, upon release into the surrounding medium, induces the competent state in neighboring cells in the population. These extracellular factors are examples of pheromones. (Pheromones are molecules released by an organism into the external medium to influence other individuals of the same species.) These extracellular factors have been studied in *B. subtilis, Streptococcus pneumoniae,* and *B. cereus*. In *B. subtilis,* two different extracellular peptide pheromones are used for competence development. In *S. pneumoniae,* the competence pheromone is a 17 amino acid peptide.

The adaptive function of bacterial transformation has been studied in *B. subtilis* as a model system for understanding the adaptive function of bacterial transformation generally. The results indicated that transformation provides external DNA template for recombinational repair of DNA damage in the recipient. Transformation and recombinational repair in *B. subtilis* require the RecE protein, which is a homolog of the well-studied RecA protein of *E. coli*. DNA damages arising from such sources as UV irradiation, endogenous oxidative free radicals, and desiccation are a pervasive problem for bacteria. The RecE protein of *B. subtilis* probably catalyzes recombinational repair of these prevalent DNA damages. Thus, sexual communication via pheromones in bacteria is probably an evolved mechanism to promote sexual interaction for recombinational repair of prevalent DNA damages. This is ''Level 1'', the primary level of sexual communication.

Transformation in bacteria involves partial diploidy as an intermediate stage. Even though the initial advantage of this process, we think, was recombinational repair of DNA damage, diploidy (or partial diploidy) may also provide the benefit of masking deleterious recessive mutations through complementation. Thus, recombinational repair in bacteria may have set the stage, or may have been a precursor to, natural selection for diploidy as a distinct phase of the sexual cycle in eukaryotes. Further, recombinational variation is produced as a by-product

of recombinational repair in bacterial transformation as well as in other sexual processes. The infrequent beneficial new traits generated by recombination presumably promote evolutionary success, just as infrequent beneficial mutations do.

5.2 Sexual Communication in Fungi, Primarily for DNA Repair, but Also for Limited Complementation

1. *Sexual communication in fungi.* Among eukaryotic microorganisms, pheromones promote sexual interaction in many species. These simple eukaryotes use a variety of molecules, such as steroids, other lipids, peptides, and derivatives of organic acids, as well as large molecules such as glycoproteins as sex pheromones. Although these sex pheromones are usually transmitted through an aqueous medium, they may also be transmitted through the air as in the fungus *Mucos mucedo*.

2. *Sexual communication in yeast.* The sexual cycle and communication have been especially well described in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. Mating in these yeasts occurs between two cells of opposite mating type and is facilitated by the reciprocal action of pheromones. Cells of each mating type release pheromones that induce mating through promotion of specific modifications in cells of the opposite mating type. The two cells then conjugate, forming a diploid zygote that can undergo meiosis. This leads to sporulation and release of ascospores that can germinate to form haploid vegetative cells, thus completing the life cycle. The pheromones produced by these yeasts are short peptides.

The results of several studies bear on the adaptive advantage of the meiotic events promoted by these sex pheromones. Recombinational repair of DNA double-strand breaks occurs during meiosis in *S. cerevisiae* and *S. pombe*. The *rad51* and *dmc1* genes of *S. cerevisiae* are homologs of the *recA* gene of *E. coli*. A homolog of *E. coli* RecA also occurs in *S. pombe*.

It was found that treatment of *S. pombe* with H_2O_2, a DNA-damaging agent, caused increased mating. This stimulation of mating can be interpreted as an adaptation to promote recombinational repair of the introduced DNA damages. Thus, in *S. cerevisiae* and *S. pombe*, production of sex pheromones appears to be an adaptation for promoting recombinational repair of DNA, "Level 1" of sexual communication, as in bacteria.

In *S. pombe*, the mating pheromones M-factor (produced by the "Minus" mating type) and P-factor (produced by the "Plus" mating type) ensure sexual interaction of cells of opposite mating type. However, during vegetative growth, a cell of one mating type switches to the other mating type in about every two generations. Similarly, in *S. cerevisiae*, haploid descendants from spores of one mating type can change their mating type frequently so that after several cell divisions a diploid population develops from a single haploid spore. These observations indicate that inbreeding is avoided to some extent in *S. cerevisiae* and *S. pombe*, since mating with an individual of the opposite mating type is required, but that it is not strongly avoided since a cell of one mating type can give rise to a cell of the opposite mating type frequently, and then sibling cells can mate.

In *S. cerevisiae* and *S. pombe*, the diploid stage of the sexual cycle, formed by mating, probably expresses functions necessary for meiosis and sporulation. Recessive mutations in the genes encoding these functions may explain the observation that

inbreeding is avoided to some extent. We discussed above (Section 3.4) that the diploid stage of a life cycle serves the adaptive function of complementation, or the masking of deleterious recessive mutations. Optimal masking is achieved when the diploid zygote is formed from the union of haploid cells of two genetically unrelated individuals, so that the recessive mutations carried by each genome are most likely to be different. However, relatively few functions are probably expressed specifically in the diploid phase of *S. cerevisiae* and *S. pombe*, and this might explain why the masking of deleterious recessive mutations is only weakly promoted in these organisms.

In these yeasts, the pheromones operate at ''Level 2'' as well, to promote ''outcrossing,'' in addition to acting at ''Level 1'' to promote mating between individuals.

3. *Sexual communication in neurospora. Neurospora crassa* is an ascomycete, like the yeasts *S. cerevisiae* and *S. pombe*, but its sexual phase is more elaborate than those of yeasts. *Neurospora crassa* encodes a gene *mei3*, which is homologous to *E. coli recA* and *S. cerevisiae dmc1* and *rad51*. A mutant defective in *mei3* has reduced repair of DNA damage and, when homozygous, is blocked in the zygotene stage of meiotic prophase. This implies that *N. crassa mei3*, like *S. cerevisiae dmc1* and *rad51*, is required for recombinational repair of DNA damage during meiosis.

Neurospora crassa has two stable mating types which produce sex pheromones. The sex pheromone of one mating type induces the formation of the female sexual structure, the differentiated protoperithecium, in the opposite mating type. Once a protoperithecium is formed, the vegetative conidiospores or mycelia of the opposite mating type can act as the male partner. Upon mating, a diploid nucleus is formed by fusion of male and female nuclei of opposite mating type. This diploid nucleus then undergoes meiosis, followed by mitosis, to produce eight ascospores. Each set of eight spores is enclosed within a sac, referred to as an *ascus*, and many asci are contained within the mature fruiting body, termed a *perithecium*. Genes promoting ascus and ascospore maturation, as well as perithecium development are expressed in the diploid stage. It was found that 74 of the 99 wild-collected *N. crassa* isolates from 26 populations carried one or more recessive mutations in genes that expressed in the diploid stage. They estimated that the number of genes expressed in the diploid stage is at least 435. Mutations in these genes, when homozygous in the diploid stage, cause formation of aborted asci, ascospores with maturation defects, or barren fruiting bodies with few sexual spores. The evolution of two distinct mating types, which communicate via pheromones, inhibits inbreeding in *N. crassa*, and this provides the advantage of masking recessive mutations in the numerous genes that express in the diploid stage.

Overall, in fungi, it appears that pheromones act at both the primary level of sexual communication to promote sexual interaction for meiotic recombinational repair and the secondary level to promote outcrossing, which allows masking of deleterious mutations.

5.3
Sexual Communication in Higher Eukaryotes, Primarily for Complementation while Repair Is an Automatic Concomitant of Sexual Reproduction

1. *Flowering plants.* Since the diploid stage is the most prominent stage in the life cycle of flowering plants, the advantages of complementation should be large. Thus,

the investment in sexual communication to promote cross-fertilization, as the most common mode of reproduction among flowering plants, is probably due to the prominence of the diploid stage and the benefit of masking deleterious mutations. Instances in which outcrossing has been abandoned in favor of self-fertilization or parthenogenesis may be explained by the need of some plants to survive in sparse populations in which the costs of sexual communication would be prohibitively high.

Sexual signals in plants promote mating, but the subsequent meiosis and recombinational repair happen in the succeeding generation when the progeny produce germ cells. Among those flowering plants that depend on cross-fertilization to reproduce, sexual communication acts at both the primary and the secondary level. However, in plants that can undergo self-fertilization, sexual communication is unnecessary when this option is used.

2. *Vertebrates.* In bacteria, fungi, and protozoa, sexual reproduction is generally facultative and favored only under certain conditions. In nearly all vertebrates, however, the sexual cycle is obligatory, and thus sexual communication is ordinarily necessary for reproduction.

In vertebrates, meiosis and concommitant recombinational repair ordinarily occur at a time separate from sexual signaling. However, sexual signaling to attract a mate is essential to continue the sexual cycle, which includes fertilization, production of progeny, and further meioses. Thus, sexual signals promote mating, but meiosis and recombinational repair occur subsequently in the succeeding generation, when progeny form germ cells.

Among vertebrates, sexual communication for promotion of outcrossing has been studied in many systems. In toads, advertisement vocalizations given by males apparently serve as cues by which females recognize their kin and thus avoid inbreeding.

In mouse, mating preference is strongly influenced by the major histocompatibility complex (MHC) genotype. A mouse can distinguish close relatives from more distantly related mice on the basis of MHC genotype through their sense of smell. Using this device, mice tend to outcross, avoiding mating with close relatives. Inbreeding of mice derived from wild populations has a significant detrimental effect on survivorship when the mice are reintroduced into a natural habitat. This effect is even more severe than that observed in laboratory studies of the population. In the other vertebrates, to be described below, the communication processes used to promote outcrossing have not been as well defined as in the mouse. Nevertheless, we infer that analogous processes also exist in these cases.

Outcrossing is promoted among birds. The great tit (*Parus major*) is a monogamous woodland bird. In natural populations, outcrossing is promoted by the dispersal of daughters, but not sons. Furthermore, among infrequent incestuous matings, nestling mortality was nearly double that of outbreeding pairs.

Among primates, avoidance of inbreeding occurs in rhesus monkeys, chimpanzees, and gorillas. Juvenile rhesus monkey males leave the troop that they were born into when they mature. This pattern is similar to that in many mammals in which male progeny leave the social group into which they are born and find a mate in another group. Chimpanzees and gorillas follow a less common pattern in which the female, rather than the males, leave their natal group and transfer to other groups.

Inbreeding ordinarily is avoided in humans. In humans, body odor preferences (determined by men and women scoring the odors of T-shirts previously worn by men or women) would serve to increase heterozygosity in the progeny. Prohibitions against marriage and sexual relations between close relatives are a near universal feature of human social behavior in many different cultures. This general prohibition is known as the incest taboo.

The basis for the incest taboo appears to be the avoidance of inbreeding because of the effect of the expression of deleterious mutations. There is a significantly higher incidence of major congenital malformations and postnatal mortality among the progeny of consanguineous marriages than among the progeny of nonconsanguinous marriages.

We have reviewed the evidence that "Level 2" of sexual communication, to avoid inbreeding and promote outcrossing, is common in vertebrates. Since sexual interaction in vertebrates is usually essential for reproduction, sexual communication at "Level 1" for sexual interaction is ubiquitous.

5.4 "Level 3" of Sexual Communication Is Used to Select for Fitness among Potential Mates

In addition to "Level 1" and "Level 2" functions of sexual communication, the first to promote homologous recombinational repair and the second to promote outcrossing, a third aspect of sexual communication involves courtship behavior. The adaptive function of courtship probably is selection of a mate with the characteristics that optimize reproductive success (selection for fitness). This leads to competition among individuals of the same gender. Such competition is the basis for sexual selection, regarded as having molded many features of animal behavior.

As an example, females of the guppy fish, *Poecilia reticulata*, respond more strongly to the mating displays of male guppies which have brighter orange spots. The brightness of the spots correlates positively with health and vigor and negatively with previous parasitic infection. Thus female guppies tend to select healthier mates that are more resistant to parasites.

Substantial evidence has been presented on human courtship behavior suggest that the basic biologic function of human courtship is to optimize reproductive success by the selection of a mate of the highest possible fitness.

6 Overview

The occurrence of high levels of endogenous DNA damage in mammals is now well established. DNA of mammalian somatic cells is the master informational molecule, and accumulated damages in this molecule probably cause the progressive irreversible deterioration of cell, tissue, and organ function that defines mammalian aging. When an organism forms progeny via germ cells, it is important that these cells be free of DNA damages, since such damages cause inviability. To facilitate efficient DNA repair, the redundant information available in the diploid cell can be used to replace damaged information through recombinational repair. Sexual reproduction appears to be an adaptation to promote pairing and exchange between homologous chromosomes for the purpose of efficient repair of the DNA, which is passed on to germ cells. In eukaryotes this occurs during meiosis,

while in bacteria and viruses it occurs during less complex, but similar, processes.

Mutations are another type of error in DNA in addition to damage. Unlike damages, mutations cannot be recognized by repair enzymes. Mutations, however, can be masked when information from two unrelated individuals (parents) is brought together through fertilization to form the progeny zygote. Thus, the outcrossing aspect of the sexual cycle in diploid organisms appears to be maintained by the advantage of masking mutations. Overall, aging appears to be a consequence of the accumulation of DNA damage, and sex appears to be an adaptation for the removal of damage through enhanced recombinational repair and the masking of mutations through outcrossing.

Sexual communication probably arose in bacteria as an adaptation to promote DNA transfer in order to allow recombinational repair of damages in DNA. Sexual communication is probably maintained in organisms with a distinct diploid stage of their life cycle by (1) the advantages of recombinational repair during meiosis, (2) outcrossing that facilitates complementation of recessive deleterious mutations, and (3) mate selection for fitness.

See also DNA Repair in Yeast; Female Reproductive System; Genetics, Molecular Basis of; Male Reproductive System.

Bibliography

Books and Reviews

Alexander, P. (1967) The role of DNA lesions in processes leading to aging in mice, *Symp. Soc. Exp. Biol.* **21**, 29–50.

Ames, B.N., Gold, L.S. (1991) Endogenous mutagens and the causes of aging and cancer, *Mutat. Res.* **250**, 3–16.

Arking, R. (1998) *Biology of Aging; Observations and Principles*, Sinauer Associates, Sunderland, MA.

Bernstein, C., Bernstein, H. (1991) *Aging, Sex and DNA Repair*, Academic Press, San Diego.

Bernstein, C., Bernstein, H., Payne, C. (1999) Cell Immortality: Maintenance of Cell Division Potential, in: Macieira-Coelho, A. (Ed.) *Progress in Molecular and Subcellular Biology Vol. 24; Cell Immortalization*, Springer-Verlag, Heidelberg, Germany, pp. 23–50.

Bernstein, H., Hopf, F.A., Michod, R.E. (1987) The molecular basis of the evolution of sex, *Adv. Genet.* **24**, 323–370.

Darwin, C. (1889) *The Effects of Cross and Self Fertilization in the Vegetable Kingdom*, D. Appleton & Co., New York.

Michod, R.E., Levin, B.R. (Eds.) (1988) *The Evolution of Sex: An Examination of Current Ideas*, Sinauer Associates, Sunderland, MA.

Primary Literature

Anson, R.M., Bohr, V.A. (2000) Mitochondria, oxidative DNA damage, and aging, *J. Am. Aging Assoc.* **23**, 199–218.

Arking, R., Buck, S., Novoseltev, V.N., Hwangbo, D.S., Lane, M. (2002) Genomic plasticity, energy allocations, and the extended longevity phenotypes of *Drosophila*, *Ageing Res. Rev.* **1**, 209–228.

Arking, R., Burde, V., Graves, K., Hari, R., Feldman, E., Zeevi, A., Soliman, S., Saraiya, A., Buck, S., Vettraino, J., Sathrasala, K., Wehr, N., Levine, R.L. (2000) Forward and reverse selection for longevity in *Drosophila* is characterized by alteration of antioxidant gene expression and oxidative damage patterns, *Exp. Gerontol.* **35**, 167–185.

Aufderhide, K.J. (1987) Clonal Aging in *Paramecium tetraurelia*. II Evidence of functional changes in the macronucleus with age, *Mech. Ageing Dev.* **37**, 265–279.

Avise, J.C. (1993) The evolutionary biology of aging, sexual reproduction, and DNA repair, *Evolution* **47**, 1293–1301.

Barrows, L.R., Magee, P.N. (1982) Nonenzymatic methylation of DNA by S-adenosylmethionine in vitro, *Carcinogenesis* **3**, 349–351.

Bernstein, C. (1998) Sex as a Response to Oxidative DNA Damage, in: Okefie, I.A., Halliwell, B. (Eds.) *Free Radicals and DNA: Techniques, Mechanisms and Applications*, OICA International, London, UK, pp. 99–129.

Bernstein, C., Bernstein, H. (1991) *Aging, Sex and DNA Repair*, Academic Press, San Diego.

Bernstein, C., Bernstein H. (1997) Sexual communication, *J. Theor. Biol.* **188**, 69–78.

Bernstein, C., Bernstein, H., Payne, C.M., Garewal, H. (2002) DNA repair/pro-apoptotic dual-role proteins in five major DNA repair pathways: fail-safe protection against carcinogenesis, *Mutat. Res.* **511**, 145–178.

Bernstein, C., Johns, V. (1989) Sexual reproduction as a response to H_2O_2, damage in *Schizosaccharomyces pombe*, *J. Bacteriol.* **171**, 1893–1897.

Bernstein, H., Byerly, H.C., Hopf, F.A., Michod, R.E. (1985) Genetic damage, mutation and the evolution of sex, *Science* **229**, 1277–1281.

Brooks, P.J., Wise, D.S., Berry, D.A., Kosmoski, J.V., Smerdon, M.J., Somers, R.L., Mackie, H., Spoonde, A.Y., Ackerman, E.J., Coleman, K., Tarone, R.E., Robbins, J.H. (2000) The oxidative DNA lesion 8,5′-(S)-cyclo-2′-deoxyadenosine is repaired by the nucleotide excision repair pathway and blocks gene expression in mammalian cells, *J. Biol. Chem.* **275**, 22355–22362.

Brosh, R.M. Jr., Bohr, V.A. (2002) Roles of the Werner syndrome protein in pathways required for maintenance of genetic stability, *Exp. Gerontol.* **37**, 491–506.

Bucala, R., Model, P., Cerami, A. (1984) Modification of DNA by reducing sugars: a possible mechanism for nucleic acid aging and age-related dysfunction in gene expression, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 105–109.

Bucala, R., Model, P., Russell, M., Cerami, A. (1985) Modification of DNA by glucose-6-phosphate induces DNA rearrangements in an *Escherichia coli* plasmid, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 8439–8442.

Buss, D.M. (1994) What do people desire in a mate? The evolution of human sexual strategies, *J. NIH Res.* **6**, 37–41.

Cai, Q., Tian, L, Wei, H. (1996) Age-dependent increase of indigenous DNA adducts in rat brain is associated with a lipid peroxidation product, *Exp. Gerontol.* **31**, 387–392.

Camerini-Otero, R.D., Hsieh, P. (1995) Homologous recombination proteins in prokaryotes and eukaryotes, *Annu. Rev. Genet.* **29**, 509–552.

Cathcart, R., Schwiers, E., Saul, R.L., Ames, B.N. (1984) Thymine glycol and thymidine glycol in human and rat urine: a possible assay for oxidative DNA damage, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5633–5637.

Cerutti, H., Johnson, A.M., Boynton, J.E., Gillham, N.W. (1995) Inhibition of chloroplast DNA recombination and repair by dominant negative mutants of *Escherichia coli* RecA, *Mol. Cell. Biol.* **15**, 3003–3011.

Chiarugi, A. (2002) Poly(ADP-ribose) polymerase: killer or conspirator? The 'suicide hypothesis' revisited, *Trends Pharmacol. Sci.* **23**, 122–129.

Cox, M.M. (1991) The recA protein as a recombinational repair system, *Mol. Microbiol.* **5**, 1295–1299.

Cox, M.M. (1993) Relating biochemistry to biology: how the recombinational repair function of RecA protein is manifested in its molecular properties, *Bioessays* **15**, 617–623.

Darwin, C. (1871) *The Descent of Man, and Selection in Relation to Sex*, Princeton, NJ, Reprinted 1981 by Princeton University Press.

de Boer, J., Andressoo, J.O., de Wit, J., Huijmans, J., Beems, R.B., van Steeg, H., Weeda, G., van der Horst, G.T., van Leeuwen, W., Themmen, A.P., Meradji, M., Hoeijmakers, J.H. (2002) Premature aging in mice deficient in DNA repair and transcription, *Science* **296**, 1276–1279.

Degler, C.N. (1991) *In Search of Human Nature*, Oxford University Press, New York.

Dizdaroglu, M. (1998) Mechanisms of Free Radical Damage to DNA, in: Okefie, I.A., Halliwell, B. (Eds.) *Free Radicals and DNA: Techniques, Mechanisms and Applications*, OICA International, London, UK, pp. 3–26.

Dougherty, E.C. (1955) Comparative evolution and the origin of sexuality, *Syst. Zool.* **4**, 145–190.

Driver, C., Tawadros, N. (2000) Cytoplasmic genomes that confer additional longevity in *Drosophila melanogaster*, *Biogerontology* **1**, 255–260.

Dudas, S.P., Arking, R. (1995) A coordinate upregulation of antioxidant gene activities is associated with the delayed onset of senescence in a long-lived strain of *Drosophila*, *J. Gerontol. Biol. Sci.* **50A**, B117–B127.

Fraga, C.G., Shigenaga, M.K., Park, J.-W., Degan, P., Ames, B.N. (1990) Oxidative damage to DNA during aging: 8-hydroxy-2′-deoxyguanosine in rat organ DNA and urine, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4533–4537.

Fryxell, P.A. (1957) Mode of reproduction in higher plants, *Botanical Rev.* **23**, 135–233.

Furuichi, Y. (2001) Premature aging and predisposition to cancers caused by mutations in RecQ family helicases, *Ann. N.Y. Acad. Sci.* **928**, 121–131.

Gaiddon, C., Lokshin, M., Ahn, J., Zhang, T., Prives, C. (2001) A subset of tumor-derived mutant forms of p53 down-regulate p63 and p73 through a direct interaction with the p53 core domain, *Mol. Cell. Biol.* **21**, 1874–1887.

Giese, H., Dolle, M.E., Hezel, A., van Steeg, H., Vijg, J. (1999) Accelerated accumulation of somatic mutations in mice deficient in the nucleotide excision repair gene XPA, *Oncogene* **18**, 1257–1260.

Gilley, D., Blackburn, E.H. (1994) Lack of telomere shortening during senescence in *Paramecium*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1955–1958.

Grammer, K. (1989) Human Courtship Behavior: Biological Basis and Cognitive Processing, in: Rasa, A.E., Vogel, C., Voland, E. (Eds.) *The Sociobiology of Sexual and Reproductive Strategies*, Chapman & Hall, New York, pp. 147–169.

Graymes, T.J., North, P.S., Brady, N., Hickson, I.D., Mufti, G.J., Rassool, F.V. (2002) Increased error-prone non homologous DNA end-joining – a proposed mechanism of chromosomal instability in Bloom's syndrome, *Oncogene* **21**, 2525–2533.

Grube, K., Burkle, A. (1992) Poly(ADP-ribose) polymerase activity in mononuclear leukocytes of 13 mammalian species correlates with species-specific life span, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2169–2173.

Gupta, R.C., Bazemore, L.R., Golub, E.I., Radding, C.M. (1997) Activities of human recombination protein Rad51, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 463–468.

Hamilton, M.L., Van Remmen, H., Drake, J.A., Yang, H., Guo, Z.M., Kewitt, K., Walter, C.A., Richardson, A. (2001) Does oxidative damage to DNA increase with age? *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10469–10474.

Hanawalt, C. (1994) Transcription-coupled repair and human disease, *Science* **266**, 1957, 1958.

Herrero, A., Barja, G. (1999) 8-oxo-deoxyguanosine levels in heart and brain mitochondrial and nuclear DNA of two mammals and three birds in relation to their different rates of aging, *Aging (Milano)* **11**, 294–300.

Higami, Y., Shimokawa, I., Okimoto, T., Ikeda, T. (1994) An age-related increase in the basal level of DNA damage and DNA vulnerability to oxygen radicals in the individual hepatocytes of male F344 rats, *Mutat. Res.* **316**, 59–67.

Holmes, G.E., Bernstein, C., Bernstein, H. (1992) Oxidative and other DNA damages as the basis of aging: a review, *Mutat. Res.* **275**, 305–315.

Holmes, G.E., Holmes, N.R. (1986) Accumulation of DNA damages in aging *Paramecium tetraurelia*, *Mol. Gen. Genet.* **204**, 108–114.

Houde, A.E., Torio, A.J. (1992) Effect of parasitic infection on male color pattern and female choice in guppies, *Behav. Ecol.* **3**, 346–351.

Huang, J., Dynan, W.S. (2002) Reconstitution of the mammalian DNA double-strand break end-joining reaction reveals a requirement for an Mre11/Rad50/NBS1-containing fraction, *Nucleic Acids Res.* **30**, 667–674.

Humphreys, D., Eggan, K., Akutsu, H., Friedman, A., Hochedlinger, K., Yanagimachi, R., Lander, E.S., Golub, T.R., Jaenisch, R. (2002) Abnormal gene expression in cloned mice derived from embryonic stem cell and cumulus cell nuclei, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12889–12894.

Imamura, O., Fujita, K., Shimamoto, A., Tanabe, H., Takeda, S., Furuichi, Y., Matsumoto, T. (2001) Bloom helicase is involved in DNA surveillance in early S phase in vertebrate cells, *Oncogene* **20**, 1143–1151.

Jimenez, J.A., Hughes, K.A., Alaks, G., Graham, L., Lacy, R.C. (1994) An experimental study of inbreeding depression in a natural habitat, *Science* **266**, 271–273.

Johnson, R.D., Jasin, M. (2000) Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells, *EMBO J.* **19**, 3398–3407.

Kaliraman, V., Mullen, J.R., Fricke, W.M., Bastin-Shanower, S.A., Brill, S.J. (2001) Functional overlap between Sgs1-Top3 and the Mms4-Mus81 endonuclease, *Genes Dev.* **15**, 2730–2740.

Kokoszka, J.E., Coskun, P., Esposito, L.A., Wallace, D.C. (2001) Increased mitochondrial oxidative stress in the Sod2 (+/−) mouse results

in the age-related decline of mitochondrial function culminating in increased apoptosis, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2278–2283.

Ku, H.H., Sohal, R.S. (1993) Comparison of mitochondrial pro-oxidant generation and anti-oxidant defenses between rat and pigeon: possible basis of variation in longevity and metabolic potential, *Mech. Ageing Dev.* **72**, 67–76.

Kuhne, M., Rothkamm, K., Lobrich, M. (2002) Physical and biological parameters affecting DNA double strand break misrejoining in mammalian cells, *Radiat. Prot. Dosim.* **99**, 129–132.

Kwan, K.Y., Wang, J.C. (2001) Mice lacking DNA topoisomerase IIIβ develop to maturity but show a reduced mean lifespan, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5717–5721.

Lehmann, A.R. (1998) Dual functions of DNA repair genes: molecular, cellular, and clinical implications, *BioEssays* **20**, 146–155.

Leslie, J.F., Raju, N.B. (1985) Recessive mutations from natural populations of *Neurospora crassa* that are expressed in the sexual diplophase, *Genetics* **111**, 759–777.

Li, B., Comai, L. (2002) Displacement of DNA-PKcs from DNA ends by the Werner syndrome protein, *Nucleic Acids Res.* **30**, 3653–3661.

Lindahl, T. (1977) DNA Repair Enzymes Acting on Spontaneous Lesions in DNA, in: Nichols, W.W., Murphy, D.G. (Eds.) *DNA Repair Processes*, Symposia Specialists Inc., Miami, FL, pp. 225–240.

Lindnor, N.M., Furuichi, Y., Kitao, S., Shimamoto, A., Arndt, C., Jalal, S. (2000) Rothmund-Thomson syndrome due to *RECQ4* helicase mutations: report and clinical and molecular comparisons with Bloom syndrome and Werner syndrome, *Am. J. Med. Genet.* **90**, 223–228.

Mahajan, K.N., McElhinny, S.A.N., Mitchell, B.S., Ramsden, D.A. (2002) Association of DNA polymerase μ (pol μ) with Ku and ligase IV: role for pol μ in end-joining double-strand break repair, *Mol. Cell Biol.* **22**, 5195–5202.

Mandavilli, B.S., Rao, K.S. (1996) Accumulation of DNA damage in aging neurons occurs through a mechanism other than apoptosis, *J. Neurochem.* **67**, 1559–1565.

Massie, H.R., Samis, H.V., Baird, M.B. (1972) The kinetics of degradation of DNA and RNA by H_2O_2, *Biochim. Biophys. Acta* **272**, 539–548.

Merry, B.J. (2002) Molecular mechanisms linking calorie restriction and longevity, *Int. J. Biochem. Cell Biol.* **34**, 1340–1354.

Michod, R.E. (1993) Genetic error, sex, and diploidy, *J. Heredity* **84**, 360–371.

Migliaccio, E., Giorgio, M., Mele, S., Pelicci, G., Reboldi, P., Pandolfi, P.P., Lanfrancone, L., Pelicci, P.G. (1999) The p66^{shc} adaptor protein controls oxidative stress response and life span in mammals, *Nature* **402**, 309–313.

Mohaghegh, P., Hickson, I.D. (2001) DNA helicase deficiencies associated with cancer predisposition and premature aging disorders, *Hum. Mol. Genet.* **10**, 741–746.

Morita, T., Yoshimura, Y., Yamamoto, A., Murata, K., Mori, M., Yamamoto, H., Matsushiro, A. (1993) A mouse homolog of the *Escherichia coli recA* and *Saccharomyces cerevisiae RAD51*genes, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6577–6580.

Moskovitz, J., Bar-Noy, S., Williams, W.S., Requena, J., Berlett, B.S., Stadtman, E.R. (2001) Methionine sulfoxide reductase (MsrA) is a regulator of antioxidant defense and lifespan in mammals, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12920–12925.

Muiras, M.L., Muller, M., Schachter, F., Burkle, A. (1998) Increased poly(ADP-ribose) polymerase activity in lymphoblastoid cell lines from centenarians, *J. Mol. Med.* **76**, 346–354.

Nakamura, J., Swenberg, J.A. (1999) Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues, *Cancer Res.* **59**, 2522–2526.

Nemoto, S., Finkel, T. (2002) Redox regulation of forkhead proteins through a p66shc-dependent signaling pathway, *Science* **295**, 2450–2452.

Ochem, A.E., Skooac, D., Costa, M., Rabilloud, T., Vuillard, L., Simoncsits, A., Giacca, M., Falaschi, A. (1997) Functional properties of the separate subunits of human DNA helicase II/Ku autoantigen, *J. Biol. Chem.* **272**, 29919–29926.

Orr, W.C., Sohal, R.S. (1994) Extension of lifespan by overexpression of superoxide dismutase and catalase in *Drosophila melanogaster*, *Science* **263**, 1128–1130.

Parkes, T.L., Elia, A.J., Dickinson, D., Hilliker, A.J., Phillips, J.P., Boulianne, G.L. (1998) Extension of *Drosophila* lifespan by overexpression of human *SOD1* in motor neurons, *Nat. Genet.* **19**, 171–174.

Pastink, A., Eeken, J.C., Lohman, P.H. (2001) Genomic integrity and the repair of double-strand DNA breaks, *Mutat. Res.* **480–481**, 37–50.

Pletcher, S.D., Macdonald, S.J., Marguerie, R., Certa, U., Stearns, S.C., Goldstein, D.B., Partridge, L. (2002) *Curr. Biol.* **12**, 712–723.

Qin, X., Zhang, S., Matsukuma, S., Zarkovic, M., Shimizu, S., Ishikawa, T., Nakatsuru, Y. (2000) Protection against malignant progression of spontaneously developing liver tumors in transgenic mice expressing O^6-methylguanine-DNA methyltransferase, *Jpn. J. Cancer Res.* **91**, 1085–1089.

Queille, S., Drougard, C., Sarasin, A., Daya-Grosjean, L. (2001) Effects of XPD mutations on ultraviolet-induced apoptosis in relation to skin cancer-proneness in repair-deficient syndromes, *J. Invest. Dermatol.* **117**, 1162–1170.

Richardson, C., Jasin, M. (2000) Coupled homologous and nonhomologous repair of a double-strand break preserves genomic integrity in mammalian cells, *Mol. Cell Biol.* **20**, 9068–9075.

Ruan, H., Tang, X.D., Chen, M.L., Joiner, M.L., Sun, G., Brot, N., Weissbach, H., Heinemann, S.H., Iverson, L., Wu, C.F., Hoshi, T. (2002) High-quality life extension by the enzyme peptide methionine sulfoxide reductase, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2748–2753.

Sancar, A. (1994) Mechanisms of DNA excision repair, *Science* **266**, 1954–1956.

Sands, A.T., Abuin, A., Sanchez, A., Conti, C.J., Bradley, A. (1995) High susceptibility to ultraviolet-induced carcinogenesis in mice lacking XPC, *Nature* **377**, 162–165.

Satoh, M.S., Jones, C.J., Wood, R.D., Lindahl, T. (1993) DNA excision-repair defect of xeroderma pigmentosum prevents removal of a class of oxygen free radical-induced base lesions, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6335–6339.

Schwartz, D., Goldfinger, N., Kam, Z., Rotter, V. (1999) p53 controls low DNA damage-dependent premeiotic checkpoint and facilitates DNA repair during spermatogenesis, *Cell Growth Differ.* **10**, 665–675.

Selzer, R.R., Nyaga, S., Tuo, J., May, A., Muftuoglu, M., Christiansen, M., Citterio, E., Brosh, R.M. Jr., Bohr, V.A. (2002) Differential requirement for the ATPase domain of the Cockayne syndrome group B gene in the processing of UV-induced DNA damage and 8-oxoguanine lesions in human cells, *Nucleic Acids Res.* **30**, 782–793.

Shen, J., Loeb, L.A. (2001) Unwinding the molecular basis of the Werner syndrome, *Mech. Ageing Dev.* **122**, 921–944.

Shapiro, R. (1981) Damage to DNA Caused by Hydrolysis, in: Seeberg, E., Kleppe, K. (Eds.) *Chromosome Damage and Repair*, Plenum Publishing, New York, pp. 3–18.

Shimamoto, A., Nishikawa, K., Kitao, S., Furuichi, Y. (2000) Human RecQ5β, a large isomer of RecQ5 DNA helicase, localizes in the nucleoplasm and interacts with topoisomerases 3α and 3β, *Nucleic Acids Res.* **28**, 1647–1655.

Shinohara, A., Ogawa, H., Matsuda, Y., Ushio, N., Ikeo, K., Ogawa, T. (1993) Cloning of human, mouse and fission yeast recombination genes homologous to RAD51 and recA, *Nat. Genet.* **4**, 239–243.

Sohal, R.S., Agarwal, A., Agarwal, S., Orr, W.C. (1995) Simultaneous overexpression of copper and zinc-containing superoxide dismutase and catalase retards age-related oxidative damage and increases metabolic potential in *Drosophila melanogaster*, *J. Biol. Chem.* **270**, 15671–15674.

Sreekumar, R., Unnikrishnan, J., Fu, A., Nygren, J., Short, K.R., Schimke, J., Barazzoni, R., Nair, K.S. (2002) Effects of caloric restriction on mitochondrial function and gene transcripts in rat muscle, *Am. J. Physiol. Endocrinol. Metab.* **283**, E38–E43.

Stahl, F.W. (1979) *Genetic Recombination: Thinking About it in Phage and Fungi*, Freeman & Co., San Francisco.

Sun, J., Folk, D., Bradley, T.J., Tower, J. (2002) Induced overexpression of mitochondrial Mn-superoxide dismutase extends the life span of adult *Drosophila melanogaster*, *Genetics* **161**, 661–672.

Sung, P. (1994) Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein, *Science* **265**, 1241–1243.

Terasawa, M., Shinohara, A., Hotta, Y., Ogawa, H., Ogawa, T. (1995) Localization of RecA-like recombination protein on chromosomes of the lily at various meiotic stages, *Genes Dev.* **9**, 925–934.

Trinei, M., Giorgio, M., Cicalese, A., Barozzi, S., Ventura, A., Migliaccio, E., Milia, E., Padura, I.M., Raker, V.A., Maccarana, M., Petronilli, V., Minucci, S., Bernardi, P., Lanfrancone, L., Pelicci, P.G. (2002)

A p53-p66Shc signalling pathway controls intracellular redox status, levels of oxidation-damaged DNA and oxidative stress-induced apoptosis, *Oncogene* **21**, 3872–3878.

Tuo, J., Muftuoglu, M., Chen, C., Jaruga, P., Selzer, R.R., Brosh, R.M. Jr., Rodriguez, H., Dizdaroglu, M., Bohr, V.A. (2001) The Cockayne syndrome group B gene product is involved in general genome base excision repair of 8-hydroxyguanine in DNA, *J. Biol. Chem.* **276**, 45772–45779.

Tyner, S.D., Venkatachalam, S., Choi, J., Jones, S., Ghebranious, N., Igelmann, H., Lu, X., Soron, G., Cooper, B., Brayton, C., Hee Park, S., Thompson, T., Karsenty, G., Bradley, A., Donehower, L.A. (2002) p53 mutant mice that display early ageing-associated phenotypes, *Nature* **415**, 45–53.

van Brabant, A.J., Ye, T., Sanz, M., German, J.L. Jr., Ellis, N.A., Holloman, W.K. (2000) Binding and melting of D-loops by the Bloom syndrome helicase, *Biochemistry* **39**, 14617–14625.

van Steeg, H., Mullenders, L.H.F., Vijg, J. (2000) Mutagenesis and carcinogenesis in nucleotide excision repair-deficient XPA knock out mice, *Mutat. Res.* **450**, 167–180.

Vikhanskaya, F., Marchini, S., Marabese, M., Galliera, E., Broggini, M. (2001) p73α overexpression is associated with resistance to treatment with DNA-damaging agents in a human ovarian cancer cell line, *Cancer Res.* **61**, 935–938.

Vogel, H., Lim, D.S., Karsenty, G., Finegold, M., Hasty, P. (1999) Deletion of Ku86 causes early onset of senescence in mice, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10770–10775.

Wakayama, T., Perry, A.C.F., Zucotti, M., Johnson, K.R., Yanagimachi, R. (1998) Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei, *Nature* **394**, 369–374.

Wang, X., Peterson, C.A., Zheng, H., Nairn, R.S., Legerski, R.J., Li, L. (2001) Involvement of nucleotide excision repair in a recombination-independent and error-prone pathway of DNA interstrand cross-link repair, *Mol. Cell. Biol.* **21**, 713–720.

Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S.J., Qin, J. (2002) BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures, *Genes Dev.* **14**, 927–939.

Wedekind, C., Furi, S. (1997) Body odor preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity? *Proc. R. Soc. London B Biol. Sci.* **264**, 1471–1479.

Weeda, G., Donker, I., de Wit, J., Morreau, H., Janssens, R., Vissers, C.J., Nigg, A., van Steeg, H., Bootsma, D., Hoeijmakers, J.H.J. (1997) Disruption of mouse *ERCC1* results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence, *Curr. Biol.* **7**, 427–439.

Wijnhoven, S.W., Kool, H.J., van Oostrom, C.T., Beems, R.B., Mullenders, L.H., van Zeeland, A.A., van der Horst, G.T., Vrieling, H., van Steeg, H. (2000) The relationship between benzo[a]pyrene-induced mutagenesis and carcinogenesis in repair-deficient Cockayne syndrome group B mice, *Cancer Res.* **60**, 5681–5687.

Williams, M.D., Van Remmen, H., Conrad, C.C., Huang, T.T., Epstein, C.J., Richardson, A. (1998) Increased oxidative damage is correlated to altered mitochondrial function in heterozygous manganese superoxide dismutase knockout mice, *J. Biol. Chem.* **272**, 28510–28515.

Willis, J.H. (1999) Inbreeding load, average dominance and mutation rate for mildly deleterious alleles in *Nimulus guttatus, Genetics* **153**, 1885–1898.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., Campbell, K.H.S. (1997) Viable offspring derived from fetal and adult mammalian cells, *Nature* **385**, 810–813.

Wilson, E.O. (1978) *On Human Nature*, Harvard University Press: Cambridge, MA.

Wu, L., Davies, S.L., Levitt, N.C., Hickson, I.D. (2001) Potential role for the BLM helicase in recombinational repair via a conserved interaction with RAD51, *J. Biol. Chem.* **276**, 19375–19381.

Yang, Q., Zhang, R., Wang, X.W., Spillare, E.A., Linke, S.P., Subramanian, D., Griffith, J.D., Li, J.L., Hickson, I.D., Shen, J.C., Loeb, L.A., Mazur, S.J., Appella, E., Brosh R.M. Jr., Karmakar, P., Bohr, V., Harris, C.C. (2002) The processing of Holliday junctions by BLM and WRN helicases is regulated by p53, *J. Biol. Chem.* **277**, 31980–31987.

Yannone, S.M., Roy, S., Chan, D.W., Murphy, M.B., Huang, S., Campisi, J., Chen, D.J.

(2001) Werner syndrome protein is regulated and phosphorylated by DNA-dependent protein kinase, *J. Biol. Chem.* **276**, 38242–38248.

Zhong, Q., Chen, C.F., Chen, P.L., Lee, W.H. (2002) BRCA1 facilitates microhomology-mediated end joining of DNA double strand breaks, *J. Biol. Chem.* **277**, 28641–28647.

AIDS/HIV, Molecular and Cell Biology

Andrew M. L. Lever
University of Cambridge, Cambridge, UK

1 Origins of HIV 100

2 The Molecular Biology of HIV 102
2.1 Virus Structure 102
2.2 Life cycle 103
2.3 Additional Accessory Proteins 108
2.3.1 Virion Associated Proteins 108
2.3.2 Nonvirion Associated Proteins 108
2.4 Envelope Variants 109

3 Routes of Transmission 110

4 Natural History of Infection 110

5 Immune Response to HIV 112

6 Drug Therapy of HIV 114

7 Viral Escape 115

8 Vaccines 116

9 Summary 117

Acknowledgments 117

Bibliography 117
Books and Reviews 117
Primary Literature 118

Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition. Edited by Robert A. Meyers.
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
ISBN: 3-527-30543-2

Keywords

HIV (Human Immunodeficiency Virus)
A retrovirus of the lentivirus family responsible for AIDS.

AIDS (Acquired Immune Deficiency Syndrome)
A clinical state of profound susceptibility to infection with low pathogenicity (opportunistic) infectious agent and unusual malignancies.

Retrovirus
A virus characterized by a diploid RNA genome converted to a DNA provirus by its reverse transcriptase enzyme and integrated into the target cell DNA.

Provirus
The integrated DNA form of the virus.

Antiretroviral Drugs
Pharmacological agents targeting virus-specific processes used in the treatment of AIDS.

This article describes the basic molecular biology of the viruses causing AIDS and the nature of the disease in causes, summarizing the immune response, antiviral therapy and varrine prospects.

1 Origins of HIV

HIV is a retrovirus of the lentivirus family. Closely related viruses have been discovered in many groups of African primates and sequence comparison would suggest that there has been more than one transspecies transmission of these viruses from monkeys into humans within the last 80 years. In the simian population, these viruses are transmitted by blood contact during biting and fighting. The butchering of monkeys for sale as bush meat is a highly plausible route of transmission of these viruses into humans through monkey blood contamination of cuts and scratches on human hands. Hypotheses surrounding contaminated polio vaccine have been comprehensively disproved.

There are two major divisions of HIV that have infected humans, HIV-1 and HIV-2. HIV-1 is phylogenetically closely related to SIV_{cpz}, a lentivirus of chimpanzees. HIV-2 is different in sequence from both of these and more closely related to the sooty mangabey lentivirus, SIV_{smm}. HIV-1 is divided into a number of different groups, the main (M) and the out (O) and new (N) groups within each of which are clades. The M group is estimated to have entered the human

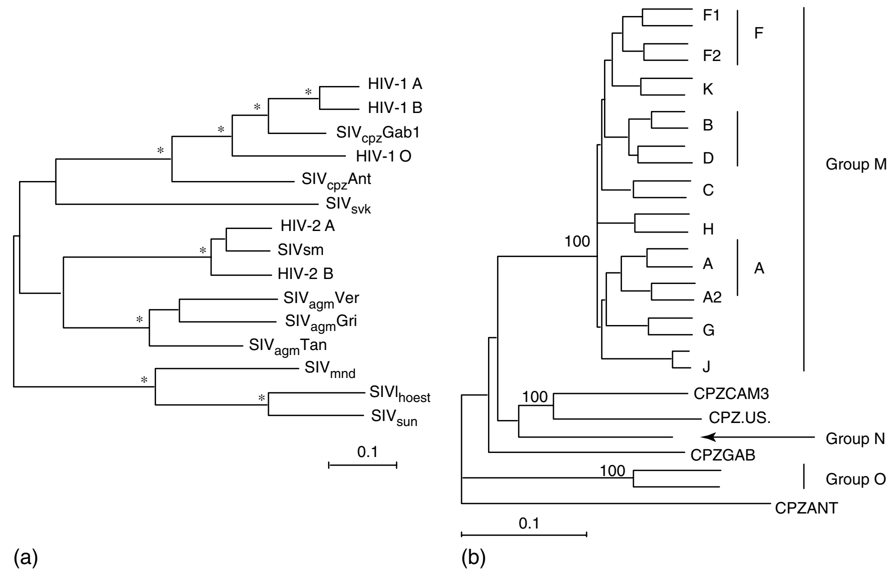


Fig. 1 (a) Relationships of HIV and SIV based on *pol* sequence comparison and (b) clades of HIV.

race in the 1930s. Within the M group, nine clades are recognized (Fig. 1a and b). In addition to clades, many recombinant or "mosaic" subtypes have also been identified. HIV-1 and HIV-2 have a very similar genetic structure except as detailed below. The sequence of the virus is, however, very variable indeed and two clades within the M group may differ in certain regions of the genome at the amino acid level by 20 to 40%. This is an astonishing level of variation and contributes significantly to the virus' success in establishing infection in the human population.

In their native hosts, simian lentiviruses appear to cause little disease. SIV from the African green monkey, SIV_{agm}, is asymptomatic in its natural host but produces profound immunodeficiency when introduced into rhesus macaques or cynomologus macaques. This may give us clues to the reason why HIV is so catastrophic in humans given the short time (<100 years) that it has been infecting humans and the consequent lack of host or pathogen adaptation.

The name lentivirus (lenti = slow) is derived from the archetypal member of this family, Maedi-Visna virus, which infects sheep and causes a very slow progressive disease involving pneumonic change and wasting, sometimes associated with central nervous system defects. In HIV, there may be symptoms associated with the initial viremia but thereafter, the speed of onset of clinical disease following infection is variable and infected individuals may remain asymptomatic for a number of years before the immunodeficiency becomes severe enough to cause illness.

Both HIV-1 and HIV-2 infection lead to a progressive decline in immune competence to the point where it becomes incapable of protecting the individual from otherwise very low pathogenicity infections, which are easily dealt with by a normal individual. This progressive loss of immune competence has given

the disease its name, *acquired immune deficiency syndrome* (*AIDS*). In general, the tempo of the disease is slower with HIV-2, possibly related to its replicating to a lower titer. This may also explain why it is less easily transmitted vertically from mother to child than HIV-1.

2
The Molecular Biology of HIV

Understanding the interactions of the virus with the immune system, its ability to cause disease, and potential therapies requires an understanding of the retroviral life cycle. The genetic structure of the virus is shown in Fig. 2. It has the conventional *gag, pol,* and *env* open reading frames of all retroviruses together with at least seven additional accessory and regulatory genes.

2.1
Virus Structure

HIV is a conventionally structured retrovirus (Fig. 3) with a C-type assembly mechanism. On the exterior, it has a lipid envelope derived from the surface plasma membrane of the cell from which it budded and in this are the envelope glycoproteins consisting of trimers of a transmembrane (TM) protein, each of which is noncovalently linked to an external surface (SU) glycoprotein. TM and SU are often referred to by their molecular weights as gp41 and gp120 respectively (in HIV-1).

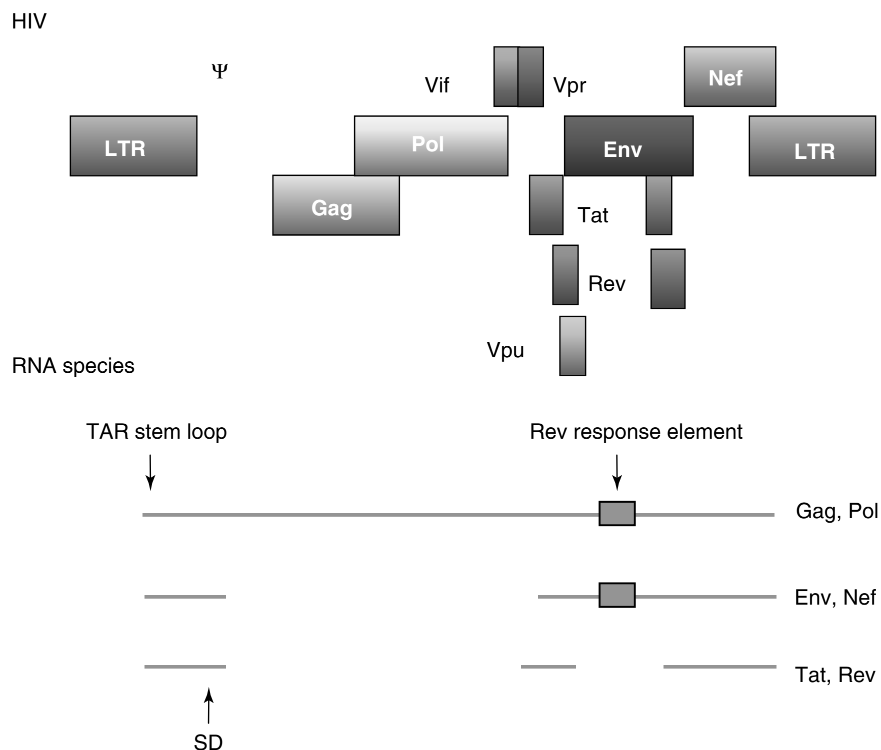


Fig. 2 Genetic structure of lentivirus DNA provirus and families of RNAs transcribed.

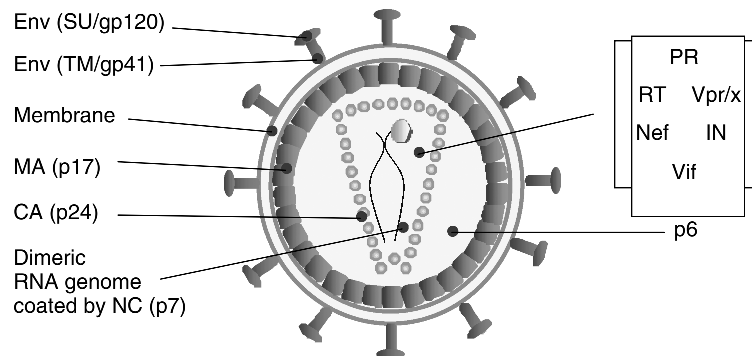


Fig. 3 Diagram of HIV 1 virus particle.

Within the mature virus particle, the inner surface of the envelope is lined by the matrix protein (MA), a cleavage derivative of the major core polyprotein, Gag. Within this is a pyramidal core structure made up of capsid (CA) protein subunits. This encloses the two copies of the RNA genome that are coated with the nucleocapsid (NC) protein. The two RNAs are tightly linked at a site known as the *dimer linkage site* (DLS). Each has a cell-derived $tRNA_{lys}$ annealed to a complementary sequence at the 5′ end of the genome, the primer binding site. Within the particle are a number of viral accessory proteins including Vif and Vpr. The virus also captures a number of cellular proteins including glycoproteins from the plasma membrane of the cell in its envelope and a number of cytoplasmic proteins including cyclophilin that binds specifically to the capsid protein in the core. The viral gene products of the *pol* gene are also incorporated into the virus. These include the protease (Pro) that cleaves the Gag and Pol precursor polyproteins to generate Gag products (p6, p7, p17, and p24) and Pol products including the reverse transcriptase enzyme (RT), which is responsible for generating a DNA copy from the RNA genome and the integrase enzyme (IN) that inserts the viral DNA into the chromosomes of the target cell.

2.2
Life cycle

Infection of a cell by HIV occurs following binding of the virus through its SU glycoproteins to the cell surface molecule CD4 and coreceptors (Fig. 4). The CD4 glycoprotein is found on a variety of cells, predominantly in those of the immune system. HIV tropism is broadly divided into viruses that can infect T lymphocytes (T-cell tropic) and those that infect cells of the macrophage/monocyte lineage (M tropic) (see Sect. 2.4).

Binding to the cell is followed by a conformational change in TM in which its hydrophobic terminus penetrates the target cell membrane and approximates it to the viral envelope lipids, fusing the two membranes together. Cellular lipid rafts appear to play a role in virus entry. The capsid and its contents enter the cell cytoplasm and the processes of disassembly and reverse transcription begin in which the viral RNA is copied into a DNA form. Reverse transcription is shown in Fig. 5. More detailed accounts can be found in the literature. Two important features of the

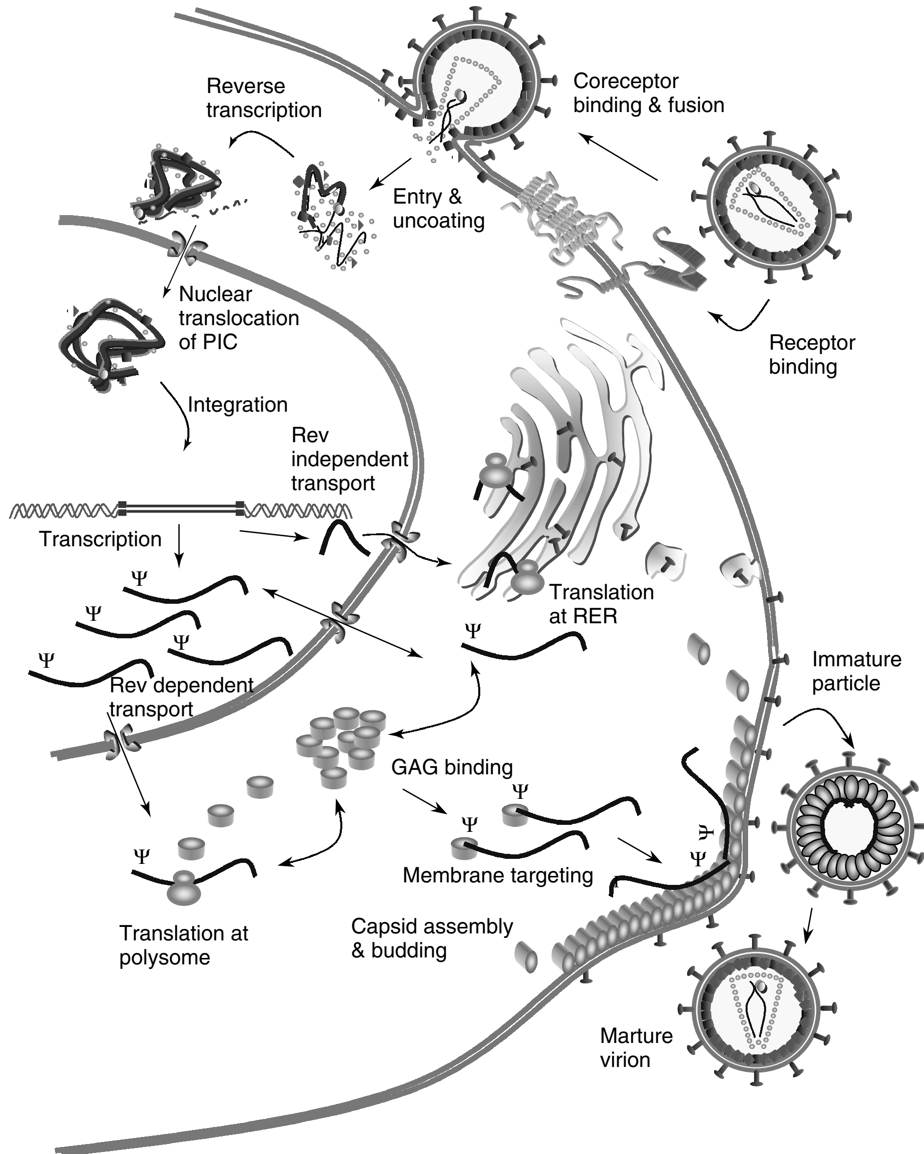


Fig. 4 Retroviral life cycle.

reverse transcription process are that both the RNA genomes are used as a template to produce a DNA copy in which template switching can occur such that the resultant DNA may be a combination of copies of segments from each of the RNA strands. Secondly, the resulting double-stranded DNA terminal sequences have both been templated by the 3′ long terminal repeat. It is during this reverse transcription step that the combination of template switching and nucleotide misincorporation, due

Fig. 5 Reverse transcription.

to the lack of RT proof-reading, produces the remarkable sequence diversity seen in HIV.

The DNA provirus is transported into the nucleus as a preintegration complex (PIC) using karyophilic signals found on proteins with the complex including the matrix protein, the integrase, and the Vpr accessory protein. There is some functional redundancy between these signals. In lentiviruses, the PIC can cross an intact nuclear membrane, meaning that these viruses have the unique capability among retroviruses of being able to integrate into cells that are not undergoing mitosis. The integration step involves the integrase enzyme cleaving the cellular DNA and making a 3′ to 5′ recess in the two ends of the proviral DNA. The chromosomal DNA undergoes a similar staggered

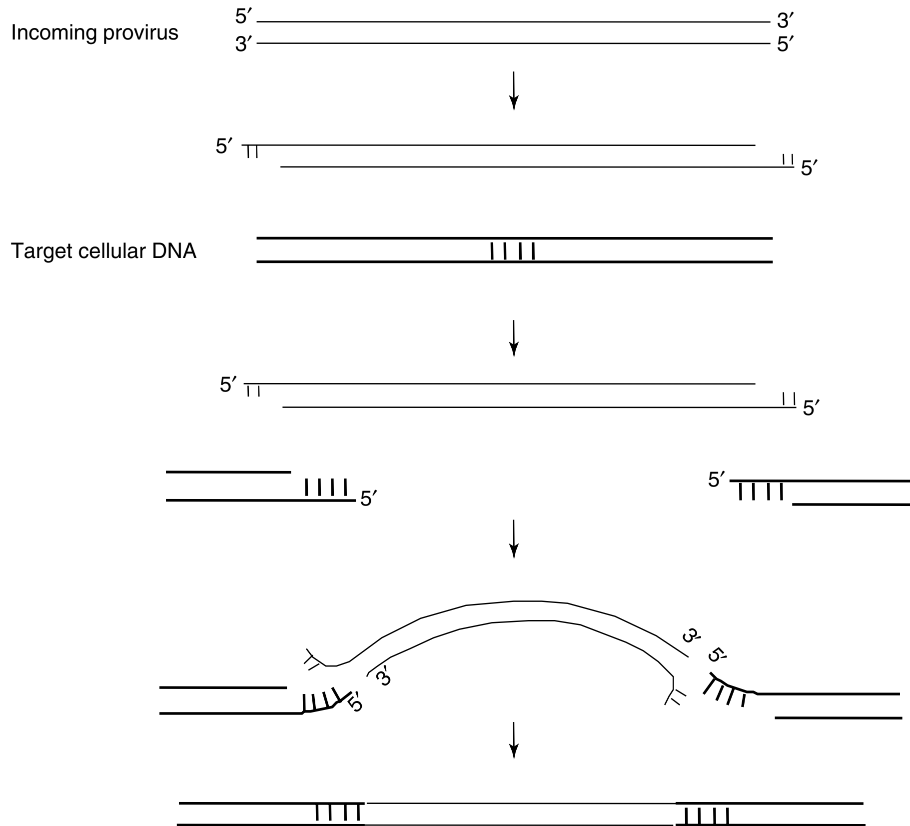


Fig. 6 Integration process.

cleavage and the viral and cellular DNA ends are linked. Cellular enzymes fill in the defect (Fig. 6). The resulting sequence shows a reduplication of the cellular nucleotides at either end of the provirus.

The integrated provirus may remain transcriptionally silent. Transcriptional activation depends in part on the degree of stimulation of the cell. In T cells, for example, mitotic stimulation is associated with activation of the retroviral promoter. Viral latency is a useful strategy for this pathogen to escape immune recognition and studies suggest that, at any one time, probably less than 10% of infected cells have a transcriptionally active virus.

The 5′ long terminal repeat of the virus contains a promoter element with a conventional TATA box. When activated, this is transcribed by RNA polymerase II to produce a genome length transcript. Multiple small abortive transcripts are seen early on but sufficient full-length RNAs are produced to initiate gene expression. Instability sequences are found in the full-length RNA, which default the RNA transcript to be spliced, and intronic sequences encompassing the Gag and Pol regions and the central part of envelope are spliced out to produce multiply spliced RNAs that are then exported into the cytoplasm and translated on free cytoplasmic

ribosomes. The earliest transcripts code for the regulatory proteins, Tat, Rev, and Nef (Fig. 2). Tat is a transcriptional activator that shuttles back into the nucleus and interacts with an RNA stem loop structure, the Tat Response Element (TAR) at the 5′ end of all viral transcripts (Fig. 2). Tat bound to TAR, recruits cyclin T and CDK9, which cause hyperphosphorylation of the C-terminal domain of the RNA polymerase II, massively enhancing its transcriptional efficiency. The sequence of transcription, export, and translation is replicated and amplified and more Tat protein is produced in what amounts to a positive feedback loop. The spliced RNA species encoding Rev is also translated and like the Tat protein, it shuttles back into the nucleus. It interacts with an RNA structure called the *Rev Responsive Element* (*RRE*) in the envelope intron (Fig. 2), enhancing export of both the full-length genomic RNA and the singly spliced RNA that encodes the envelope glycoprotein. Thus, the Tat protein causes an early amplification of transcription, and Rev subsequently leads to an early to late shift in the species of RNA, which are exported from the nucleus from those encoding regulatory factors to mRNAs for structural and enzymatic proteins. Unspliced full-length RNAs are translated on free cytoplasmic ribosomes to produce the Gag and Pol proteins. The singly spliced envelope encoding RNA binds to a ribosome and with the translation of its signal peptide, the complex is transported to the rough endoplasmic reticulum where envelope glycoproteins are translated and glycosylated using the conventional cellular glycoprotein secretory route.

The Gag and Pol proteins are encoded on the same transcript but in different reading frames. Translation of the polyprotein is interrupted by a frameshift sequence at the Gag/Pol overlap with an efficiency of about 5%. The ribosome shifts into the −1 frame, so that for approximately every 19 Gag proteins that are produced, 1 Gag/Pol fusion protein is synthesized. Gag and Gag/Pol proteins assemble at the plasma membrane in approximately the same stoichiometric ratio. At some stage, following transcription, the Gag protein has interacted with full-length genomic RNA transcripts and captured these through binding to a complex folded region of the RNA termed the *packaging signal* or Ψ region. This RNA is then transported into the assembling virion where it binds other Gag proteins through the zinc fingers of their nucleocapsid subfragments, the RNA acting to some extent as a scaffold for viral assembly. RNA–protein and protein–protein interactions contribute to the generation of a convex curved raft of protein underneath the cell membrane. Assembly continues with the viral bud bulging out of the cell and the Gag and Gag/Pol proteins accumulating to form a spherical viral core. This mode of assembly at the membrane defines lentiviruses as having a C-type morphology.

The viral core that forms is composed of Gag and Gag/Pol polyproteins, all aligned with the matrix region outermost and the NC region toward the center of the particle. At the extreme C-terminal of the Gag protein is a small region termed p6, which is intimately involved in the budding process. Mutations in p6 have been shown to lead to arrested budding in which stalked but otherwise intact particles remain attached to the plasma membrane of the cell. p6 has recently been shown to interact with a cellular protein, TSG101. In the normal cell, TSG101 is involved in the budding process of endosomes into multivesicular endosomes. It appears that the virus has

hijacked this budding mechanism for an external budding process and uses the same cellular factors to facilitate this. The complete array of proteins involved is not yet elucidated.

After or during the process of budding, the protease component of the Gag/Pol polyprotein becomes activated. This enzyme is inactive in the Gag/Pol monomer and functions as a homodimeric molecule. An as yet unknown process appears to approximate two protease monomers leading to autolytic cleavage of the first protease enzyme, following which cleavage of further Gag/Pol molecules will lead to release of further protease subunits and a cascade of proteolytic cleavage of Gag and Gag/Pol can occur. This cleavage is accompanied by a morphological change in which the particle changes from a spherical one containing a "doughnut" ring of structural proteins into a typical retrovirus particle with matrix protein underlying the envelope and an apparent space between this and the pyramidal core, which is composed of capsid proteins.

2.3 Additional Accessory Proteins

These can be divided into those that are or are not incorporated into the viral particle.

2.3.1 Virion Associated Proteins

As noted previously, the Vpr protein is involved in the preintegration complex targeting the nucleus of the newly infected cell. There is evidence that Vpr binds to the nucleoporin hCG1. Vpr also appears to have other functions. It may be a transcriptional transactivator in its own right through binding to p300/CBP coactivators. It has also been implicated in G2 phase cell cycle arrest. This is suggested to be the phase of the cell cycle at which Tat transactivation is most efficient. Vpr may also have a role in pathogenesis of AIDS since it has clearly been associated with apoptosis in a number of different cell types including neurons. It appears to gain entry to the virus by binding to the p6 region of Gag.

In HIV-2 and SIV, a gene duplication appears to have occurred to produce a second regulatory protein, Vpx, with homology to Vpr. There is evidence that in these viruses, Vpx may have subsumed the specific nuclear entry role of Vpr.

Vif is another small accessory protein that is incorporated into the virus particle. It has been clear for many years that Vif is an essential protein for viral replication in some cell lines but not others. The Vif phenotype depends on virion associated Vif since expression of Vif in a target cell does not appear to allow replication of an incoming Vif negative virus. Recent work suggests that Vif may be interacting with an inhibitory cellular protein that has sequence homology to proteins of the cytidine deaminase family and it may be that the actions of this nucleotide-modifying enzyme must be neutralized for successful reverse transcription and integration to occur.

2.3.2 Nonvirion Associated Proteins

Vpu is a small protein found only in HIV-1 and SIV_{cpz}. It has a number of roles that enhance the efficiency of budding of the virus. The envelope glycoprotein, gp120, which on the virion particle interacts with the CD4 protein, is synthesized and exported in the infected cell through the same pathway as the cell's own CD4. Complexing of these two molecules in the ER has been documented. Vpu appears to facilitate disaggregation of these two by binding to and degrading CD4, enhancing the export of the gp120 SU protein.

Vpu also appears to have additional envelope-independent effect, enhancing virus export.

The Nef protein is found throughout the primate lentiviruses. It is a multifunctional molecule. It has been documented as being capable of downmodulating cell surface expression of certain cellular proteins including CD4 and the MHC proteins. Recent evidence suggests that Nef may be involved in recruiting lymphocytes to HIV-infected macrophage/monocytes. Specific effects on migration inhibition of lymphocytes have been documented and these properties seem to be a powerful functional argument for Nef having important effects on virus spread and infectivity. This is further supported by the observation that Nef-deleted or Nef-mutated viruses have been shown to be associated with a severely attenuated pathogenicity. Recipients of HIV-infected blood from a single donor in Australia remained well for many years and the virus was identified as having a defective Nef protein. In SIV infection, Nef mutant viruses show an attenuated phenotype and reversion of the Nef mutation to wild type is associated with a regain of virulence.

2.4
Envelope Variants

The envelope gene of HIV-1 encodes a polyprotein comprising the two components of the external viral receptor ligand. At the N-terminus is the surface (SU) or gp120 protein. At the C-terminal, the TM is found. This latter sequence contains a hydrophobic stretch of amino acids, which comprise the membrane anchor and a second hydrophobic region involved in binding to the target cell membrane and fusing it with the virus envelope. The SU region is divided into five constant (C1-5) and five variable (V1-5) regions. The SU protein of HIV is responsible for binding to target cells. The V3 loop region of SU has attracted much attention because of its critical importance in receptor binding. The CD4 protein found on lymphocytes and cells of the monocyte/macrophage series was identified early in the AIDS epidemic as being a major receptor for the virus. At the same time, it was clear that binding to CD4 was necessary but not sufficient for infection. Observations that activated lymphocytes released substances that block infection by HIV and the identification of these as chemokines and their receptors as belonging to the family of 7-transmembrane spanning chemokine receptors revealed the existence of a family of coreceptor molecules, which are required for successful entry and infection. The two major families of these are the so-called *CC chemokine receptors* and the *CXC receptors*. Two of these are the principal HIV receptors. The CC chemokine receptor CCR5 responds to the cytokines MIP1α, MIP1β, and RANTES. The CXC chemokine receptor CXCR4 is bound by SDF1. There are a number of other receptors that have been identified as being involved in binding of HIV-1, HIV-2, and SIV. The identification of these receptors neatly mapped on the long-standing observation that there were at least two types of tropism identifiable in HIV. Some viruses appeared to be able to infect cells of the monocyte/macrophage series very efficiently but were very poor at infecting T lymphocytes. Others appear to be extremely efficient at infecting T lymphocyte cell lines *in vitro* but were very poor at infecting macrophages. These were broadly divided into ''macrophage tropic'' and ''lymphocyte tropic'' viruses (although this is a simplification). The chemokine receptor data link these tropisms with the

fact that the lymphocyte tropic viruses have a surface envelope protein that preferentially binds to the CXCR4 receptor and the macrophage tropic to the CCR5 receptor. These tropisms are commonly abbreviated by describing viruses as X4 or R5, respectively. More recently, a mutation has been identified, which affects about 1% of Caucasians who are homozygous for a stop codon causing a truncation of the CCR5 protein, and renders these individuals uninfectable by macrophage tropic virus. This explained another observation that there were some individuals who appeared to be repeatedly exposed to HIV through high-risk sexual practices and yet did not become infected.

3 Routes of Transmission

Infection can occur via cell-free virus or through contact between infected cells and those of the recipient. SIV in monkeys is probably primarily transferred from host to host by blood contact. HIV is also infectious by blood contact but the predominant mode of transmission is by the sexual route. HIV positive mothers can transmit the virus to their newborn child at birth or through breast feeding. The fact that cell-free virus is infectious means that HIV was transmitted very efficiently in clotting factor concentrates in the 1980s, resulting in the infection of large cohorts of hemophiliac patients. In the developed world, blood donations are screened for HIV as they are in many countries in the developing world. There is a very small but finite risk that a blood donor in the "window" period before seroconversion may transmit an infection through blood donation but this is in the order of less than one in a million. Blood transmission occurs quite efficiently in the case of intravenous drug users who share needles. Sexual transmission can occur from both male to female and female to male with an estimated risk of 1 : 200. Mucous membrane transmission probably begins with the infection of sessile cells of the macrophage/monocyte system in the mucous membrane tissues with subsequent replication and spread of virus to other cells of the immune system. Recent evidence implicates Langerhans cells as the first target. These may bind the virus using the DC-sign cell surface lectin, which has a high affinity for gp120. Whether replication occurs within these cells or whether surface-bound virus infects migrating T cells and macrophages is not established.

4 Natural History of Infection

The characteristic natural history of infection is illustrated in Fig. 7. Following early replication in lymph nodes and within a period of approximately four to eight weeks after exposure to virus, viremia becomes detectable in the form of ELISA positivity to the p24 capsid antigen and detection of viral RNA in the serum. This viremia rises rapidly to a peak that may be around five million copies of the viruses per milliliter of blood. At this time, the beginnings of an immune response against the virus become detectable, including special antibodies directed against the viral proteins. Subsequently, the virus level in the blood falls to a "set point" and this correlates most closely with detection of a rise in the cytotoxic T lymphocyte cell mediated immunity. At the time of this seroconversion, between a third and a

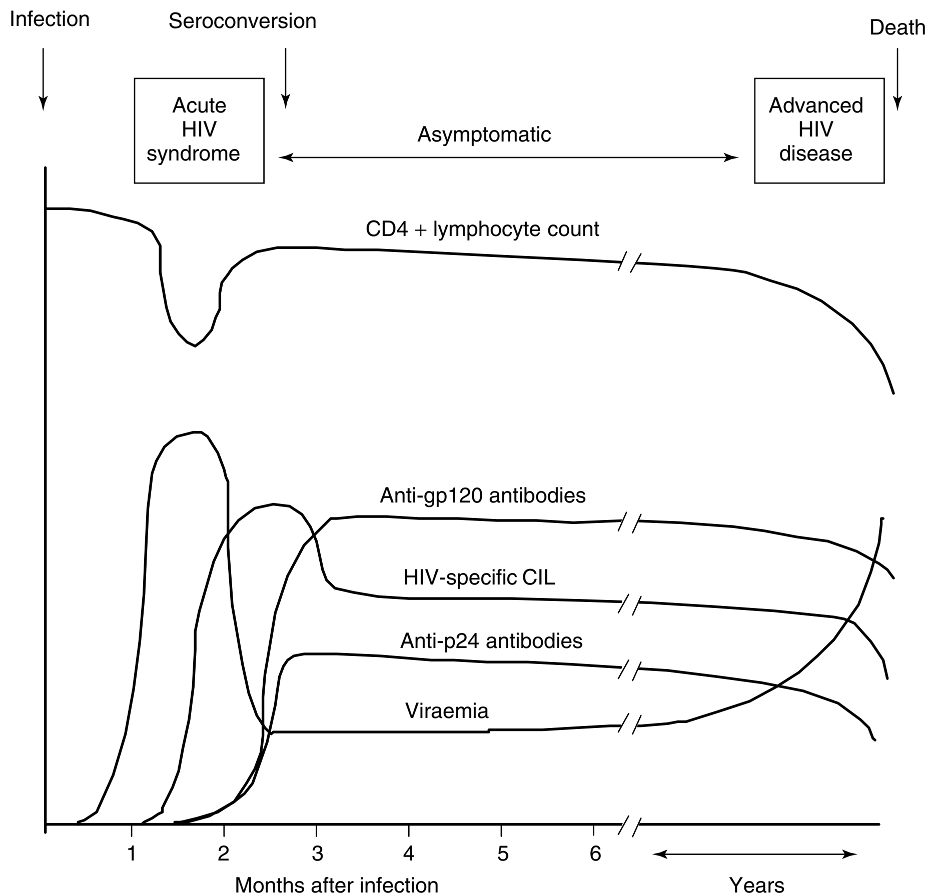


Fig. 7 Natural history of HIV infection.

half of infected patients will have symptoms resembling a glandular feverlike illness with lymphadenopathy, transient petechial rashes, arthropathy, evidence of peripheral neuropathy, and occasionally a brief meningoencephalitic illness. However, virtually without exception, these symptoms resolve. There then follows a period of apparent stability in which the viral level or "viral load" remains stable and there is ongoing evidence of an immune response. In some individuals, persistent enlargement of lymph nodes occurs associated with weight loss and fevers. This so-called *persistent generalized lymphadenopathy* appears to be associated with a worse prognosis in terms of the speed of onset of immunodeficiency. Despite the asymptomatic nature of infection in most individuals at this stage, measurements of the CD4 lymphocyte count show that during the initial viral peak, there is a corresponding trough in lymphocyte numbers, implying a significant amount of direct virus mediated lymphocyte destruction. This rises again as the virus load is brought under control but does not achieve the preinfection level and

from then on there is a very slow but progressive decline.

Calculations of the quantity of virus being produced during this time and the rate of lymphocyte destruction/regeneration have generated the remarkable figures that around 10^9–10^{10} viruses are produced and destroyed each day. This is probably an underestimate. The lymphocyte turnover is also around 10^9 per day, but while the viral load may be constant or may rise, the lymphocyte count slowly falls such that there is a net loss of around 10^6 lymphocytes per day. The fall in lymphocyte count correlates with a progressive decline in immunological competence and is a marker, if not the cause, of the eventual breakdown in immune competence. At this stage, the number of infected cells in the blood is very low, at around 1 in 10^5. This might suggest that virus mediated lymphocyte destruction is less important at this stage, although localized destruction in lymph nodes would be difficult to detect. Immune clearance of infected cells, however, may be the dominant force affecting lymphocyte numbers. Depending on the success of the immune response in controlling the virus, the length of time between seroconversion and immunodeficiency may be anything from 2 up to 12+ years. However, as yet, although there are examples of very slow progressors, there is no evidence of complete control of viral infection by the immune system once the virus has established itself. At the late stages of the disease, the predominant type of virus appears to change with the emergence of faster replicating T-cell tropic (X4) variants.

Symptomatic immunodeficiency becomes apparent when the lymphocyte count falls below 200×10^9 L^{-1}, at which point the individual is at very high risk of opportunistic infection and the development of unusual malignancies. The fall in lymphocyte count correlates with a loss of control over viral replication and a rise in detectable virus in the circulation in the absence of antiretroviral therapy. Death from an overwhelming infection or an uncontrolled malignancy usually occurs within two to three years of the onset of the severe immunodeficiency even with treatment of the individual infections and cancers.

As will be described later, the advent of antiretroviral therapy has transformed the natural history of disease in the western world. Pharmacological control of virus replication leads to a fall in viral load and a corresponding rise in the CD4 lymphocyte count. This leads to a reversal of the loss of immunocompetence, which, although it may not be complete, is enough to significantly reduce the risk of opportunistic infection.

5 Immune Response to HIV

Following the initial peak of virus in the circulation after infection, the viral load declines to a set point, which, on average, is around 30 000 copies per milliliter. If it is higher than this, the onset of AIDS is quicker and vice versa. It appears that qualitative and quantative parameters of the immune response determine this set point. A broad antibody response develops within the first three months of infection with specificity for some of the structural proteins such as Gag. Antibodies to Envelope and Polymerase proteins appear swiftly thereafter, and the regulatory and accessory proteins also trigger an antibody response.

Antibodies interfering with interaction between the envelope protein and the

cellular receptor may be neutralizing and the site at which they bind on the viral protein provides strong clues as to the important regions of SU involved in viral entry. In particular, the V3 loop is a major target for neutralizing antibody. A second region on gp120, which is involved in both CD4 and chemokine receptor binding, is another neutralization target. However, V3 loop neutralizing antibodies often appear to be more effective against laboratory strains of HIV than primary clinical isolates. It is possible that this relates to the different conformation of the envelope in the *in vitro* adapted strains in which the lack of immune selection pressure has allowed a more open structure that favors rapid cellular entry. *In vivo*, such a structure would predispose to the blockade by neutralizing antibodies and may be disadvantageous.

Other evidence suggesting that antibodies are of relatively low importance is the fact that some individuals in whom the illness appears to be held in check by the immune response do not have high levels of neutralizing antibody. In addition, passive protection using antibodies in animal models is only effective when they are present in very high titer. There is some evidence that high titers of neutralizing antibody may be associated with a lower "set point." The immense variability of the viral envelope facilitates immunological escape of the virus. This may occur either by amino acid variability or by modification of envelope glycosylation.

Cytotoxic T-cell responses can be detected before the antibody response and have long been considered one of the major immunological defenses against HIV. CD8+ T cells can kill virus infected cells by specific recognition of the HIV peptides presented on class I MHC proteins on the surface of the infected cell. *In vitro*, CTLs can eliminate infected cells even before new viruses have been produced. A second effect involves the release of β-chemokine mediators, which, as previously mentioned, are themselves able to bind to their natural ligand CCR5 and block the use of this protein by HIV for infection of the cell. The strong correlation between appearance of CD8+ T cells and the decline in viral replication in the early stages together with data on adoptive transfer in humans and CD8 depletion experiments in monkeys, which leads to a sustained increase in viral load, confirms the critical role played by this arm of the immune response in combating infection.

The signature of HIV infection is the decline in CD4 lymphocytes. The mechanisms of destruction of CD4 lymphocytes are probably many and varied as described earlier. Production of HIV with its gp120 within a cell that is already making CD4 may lead to protein aggregation within the cell and dysfunction leading to apoptosis. This can be observed *in vitro* as single cell killing. Expression of gp120 on the surface of the cell may induce fusion of that cell with a CD4+ cell. This latter phenomenon is also seen *in vitro* and leads to killing by syncitia formation. Passive binding of shed gp120 on to the CD4 proteins of immune competent cells may lead to them being a target for immunological clearance and the gp120 itself may have adverse effects. The cytotoxic T-cell response, clearing infected cells, is probably responsible for a very large proportion of the destruction of CD4+ infected cells. Thus, a number of direct and indirect mechanisms contribute to the loss of the CD4 cell population.

Individuals whose disease progresses very slowly (long-term nonprogressors – LTNPs) have clearly been shown to have powerful T-cell responses against HIV with the CD4 lymphoproliferative

response correlating well with successful suppression of the viral load. Whether these cells act independently or through enhancement of CD8+ CTL killing is not clear and it may be a combination of both. Apart from specific genetic variants such as the CCR5 truncation mutation described earlier, there does appear to be some genetic influences in that certain HLA alleles HLAB35 and CW4 are associated with more rapid disease progression, and certain other polymorphisms in chemokines and chemokine receptor genes may influence responses to therapeutic strategies and vaccination.

The immune response is capable of a measure of control of HIV replication in all patients, and in some, there is very significant delay in the onset of AIDS. Recent evidence suggests that not only does the ability to mount an immune response decline as disease progresses but the ability to regenerate this when antiretroviral treatment is started also declines the longer the initiation of treatment is delayed. Indeed, treatment during the early stages of infection at seroconversion has been correlated with a much better long-term CD4 lymphoproliferative response. Thus, early treatment might lead to the ability to maintain a strong CD4 cell response and this might provide the possibility of powerful long-term control of viral replication. Against this is the risk of generating drug resistance early in infection and the loss of utility of the limited range of chemotherapeutic agents when they may be more useful at a later stage.

6 Drug Therapy of HIV

Antiviral drugs are commonly targeted at viral enzymes. This is the case with HIV in which the first two types of highly successful pharmacotherapies have been directed to inhibit the actions of the viral reverse transcriptase gene and the protease gene. Reverse transcriptase is clearly a virus-specific process since the cell does not have any functional RNA-dependent DNA polymerase activity. The reverse transcriptase enzyme does not have proof-reading capability and while, as mentioned later, this may be an advantage to the virus in generating variants, it also is an Achilles' heel for the virus in that nucleoside analogs that mimic the bases that the enzyme is attempting to incorporate into the DNA chain can substitute for these and once inserted cannot be excised. Pyrimidine analogs have been particularly effective and they appear to act both by inhibiting the enzyme and by chain termination. They require phosphorylation within the cell to form triphosphates. Another class of drug that inhibits RT is the nonnucleoside reverse transcriptase inhibitors (NNRTIs), which bind to reverse transcriptase away from the active site but are not incorporated into the transcript. A third major class of antiretroviral drugs are those inhibiting the viral protease. These take advantage of the target peptide sequence of the aspartyl protease of HIV, being slightly different from cellular homologs of this enzyme. Predictably, however, the specificity of these drugs is somewhat lower than RT inhibitors and they are associated with a greater number of side effects, probably relating to effects on cellular proteases. Newer classes of drugs are being developed including those that may inhibit viral entry, viral integration, and viral export. Injection of a peptide, which interferes with the formation of the fusogenic envelope of the virus and blocks cell entry, has recently been introduced into clinical practice.

Structured treatment interruptions (STI) is being tested as a strategy to reduce drug toxicity. As yet, the effect of this on overall prognosis and immune regeneration is not fully established.

7
Viral Escape

The previously described infidelity of the reverse transcriptase enzyme is responsible for considerable sequence change in the virus. Undoubtedly, many of the incorporated nucleotides will cause lethal mutations and a nonviable integrated provirus will result. The huge numbers of viruses being produced everyday, however, will ensure that a large number of viable variants will be produced, which may have minor changes in their amino acid sequence. Thus, it is easy to see that a virus entering a macrophage may mutate slightly during its reverse transcription and integration process such that the progeny virions from that cell may have a slightly greater predisposition to infect lymphocytes than the parental version. The other major mechanism of sequence variation comes from recombination. When a cell is infected with two different viruses, the diploid nature of the genome means that it is possible for an RNA strand from two different integrated proviruses to be copackaged. During reverse transcription in the target cell, the reverse transcriptase enzyme is known to dissociate from its template and may reassociate with the partner template. Thus, exchange of large blocks of genetic information between the two strands is quite common. Indeed, it is estimated that the enzyme skips between strands at least four times during every replication cycle. In this way, mixing and matching of genomes can occur with ease and, in this case, large blocks of functionally intact genome will be changed. Recombination is, therefore, believed to be as significant if not more significant than RT errors in generating diversity. Some of the major worldwide clades have clearly been formed from recombination. Genotype E in Thailand, for example, is a recombinant between a clade A Gag and a clade E envelope sequence.

Variation allows the virus to introduce point mutations, which will render the binding of antiviral drugs much less efficient, and strategies to combat viral resistance have become an integral part of clinical management. This is particularly the case since antiviral therapy in the western world became more widespread. The chances of acquiring a virus, which has already generated some antiviral resistance, has increased greatly. For this reason, where possible, it is recommended that antiviral resistance testing is undertaken before initiating antiretroviral therapy so that the combination of drugs chosen is one in which the virus is sensitive to each component. As well as this, it is vital that combinations of three or more drugs are used to prevent emergence of resistance. Both of these strategies are designed to avoid presenting the virus with a single drug against which it can develop resistance.

The genetic variation also contributes very largely to the difficulty the immune system has in neutralizing this virus. There are clearly documented examples of immunological escape where viral proteins, which have cytotoxic T-cell epitopes, mutate and thus escape cytotoxic T-cell recognition. In terms of envelope variation, the virus has the additional weapon of variable glycosylation. HIV envelope is extremely heavily glycosylated compared to a number of its close viral relatives

and it is clear that glycosylation may significantly inhibit the function of neutralizing antibodies.

8 Vaccines

From the above, it is clear that the development of a vaccine against HIV is a monumental task compared to many other pathogens. The features that make it particularly challenging are as follows. Firstly, there is no evidence that complete sterilizing immunity occurs in the human race after infection. Thus, even the best immune response in the world does not appear to be capable of indefinitely checking virus replication. Secondly, the huge variability of the virus means that unlike, for example, polio or smallpox, vaccines containing a restricted number of variants will not successfully prevent infection and disease caused by HIV since the virus has such a vast repertoire of viable variants. Thirdly, the virus integrates into the genome of the cell. A number of the features of the life cycle of this and other retroviruses have been designed to specifically maintain cellular viability since death of the cell equates with death of the provirus. The action of many of the accessory genes of HIV in downmodulating cell surface proteins makes even a good immune response incapable of easily clearing infection. Over and above that, however, the virus has the capacity to become latent. Complete transcriptional arrest may occur (or a level of basal transcription insufficient to generate Tat), with transcriptional reactivation occurring under appropriate conditions such as triggering of the cell into mitosis. A transcriptionally inactive virus is completely invisible to the immune system and persistence of a small number of latently infected cells leaves the possibility of reactivation when the immune response wanes. Reactivation of other viruses such as *Varicella zoster* provides a comparable parallel. It is, thus, not surprising that progress toward a vaccine has been slow. To date, vaccines have been generated that are capable of preventing infection in SIV models. Most of these appear to be effective against homologous strains but with, as yet, relatively little evidence that effective cross-neutralization of heterologous strains occurs. Use of a live attenuated vaccine such as one deleted in *nef* and some of the other regulatory genes has been associated with protection against subsequent challenge. Attenuation, however, appears to be a finite phenomenon and mutation back to the wild-type sequence and full pathogenicity is a feature. Recent data using prime boost techniques in which carrier vaccine strains are used such as modified vaccinia Ankara have shown some promising evidence of good immune responses against HIV proteins. As yet, the duration of immune response using these techniques is not clear. A cohort of prostitutes in western Africa who were repeatedly exposed to HIV but remained uninfected originally, generated hope that a fully neutralizing immune response was possible with frequent low level viral challenge. The follow-up of these individuals showed that when sexual exposure to HIV ceases, the immunity wanes and the individuals become susceptible to infection again. This, again, is rather a concerning issue in terms of vaccine development since it would suggest that repeated boosts of vaccination will be required to maintain a level of immunity, which would otherwise be lost. There are strong parallels here with immunity to malaria.

9 Summary

The HIV epidemic is a global problem that has been controlled but not eliminated in the western world. The vast majority of infected individuals live in sub-Saharan Africa where access to treatment for the virus or its complications is extremely limited. Significant loss of life and disruption of family and society has occurred in this part of the world due to HIV. Recurrent suggestions that HIV may not be the causative agent of AIDS have seriously set back educational programs in a number of countries and are unfounded and irresponsible. Countries that have taken advice and help and approached the problem with a vigorous education campaign such as Uganda have reaped the rewards with the HIV positivity rate in the antenatal population halving over a period of less than 10 years. Until vaccines are developed and antiretroviral therapy becomes more widely available, these sort of approaches should be powerfully supported. In the western world, antiretroviral therapy has transformed the prognosis, and conceivably, a middle aged person acquiring HIV might now have a normal lifespan given the additive effects of the slow onset of clinical disease, the availability of antiretroviral drugs, and the combined time window these provide for the appearance of new therapies. While the death rate has fallen in the west, the incidence of HIV infection has continued to rise. This is partly because of increased longevity of infected individuals and also from increasing numbers of migrants from highly endemic areas who may seek access to western health resources. This trend is increasing and will provide an increasing burden on western healthcare services. There is still no reason why an uninfected person cannot be protected against HIV since transmission is preventable by barrier methods of contraception, vaginal virucidal agents, and so on. However, the key to the control of HIV in the near future in the developing and developed world is education.

Acknowledgments

I would like to acknowledge S. Griffin for contributions to the figures.

See also Genetic Engineering of Vaccines; Genetics, Molecular Basis of; RNA Virus Genome Packaging.

Bibliography

Books and Reviews

British HIV Association Guidelines for Antiretroviral Treatment of HIV Seropositive Individuals. BHIVA Guidelines Co-ordinating Committee (1997), *Lancet* **349**, 1086–1092.

Carpenter, C.C., Cooper, D.A., Fischl, M.A., Gatell, J.M., Gazzard, B.G., Hammer, S.M., Hirsch, M.S., Jacobsen, D.M., Katzenstein, D.A., Montaner, J.S., Richman, D.D., Saag, M.S., Schechter, M., Schooley, R.T., Thompson, M.A., Vella, S., Yeni, P.G., Volberding, P.A. (2000) Antiretroviral therapy in adults: updated recommendations of the international AIDS society-USA panel, *JAMA* **283**, 381–390.

Coffin, J.M., Hughes, S.H., Varmus, H.E. (Eds.) (1997) *Retroviruses*, Cold Spring Harbor Laboratory Press, New York.

Cohen, O.J., Fauci, A.S. (2001) Pathogenesis and Medical Aspects of HIV-1 Infection, in: Knipe, D.M., Howley, P.M. (Eds.) *Fields Virology*, 4th edition, Lippincott Williams & Wilkins, Philadelphia, pp. 2043–2094.

Fackler, O.T., Peterlin, B.M., Weis, K. (2001) Lessons from HIV: movement of macromolecules inside the cell, *Curr. Mol. Med.* **1**, 1–7.

Freed, E.O., Martin, M.A. (2001) HIVs and their Replication, in: Knipe, D.M., Howley, P.M. (Eds.) *Fields Virology*, 4th edition, Lippincott Williams & Wilkins, Philadelphia, pp. 1971–2042.

Jeang, K.-T. (Ed.) (2000) *Advances in Pharmacology: HIV-1: Molecular Biology and Pathogenesis – Viral Mechanisms (Vol 48)* & *Clinical Applications (Vol. 49)*, Academic Press, San Diego.

Persaud, D., Zhou, Y., Siliciano, J.M., Siliciano, R.F., (2003) Latency in human immunodeficiency virus type 1 infection: no easy answers, *J. Virol.* **77**, 1659–1665.

Weiss, R.A., Adler, M.W., Rowland-Jones, S.L. (Eds.) (2001) The changing face of HIV and AIDS, *Br. Med. Bull.* **58**(1).

Wyatt, R., Sodroski, J. (1998) The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens, *Science* **280**, 1884–1888.

Primary Literature

Addo, M.M., Altfeld, M., Rosenberg, E.S., Eldridge, R.L., Philips, M.N., Habeeb, K., Khatri, A., Brander, C., Robbins, G.K., Mazzara, G.P., Goulder, P.J., Walker, B.D. (2001) The HIV-1 regulatory proteins Tat and Rev are frequently targeted by cytotoxic T lymphocytes derived from HIV-1-infected individuals, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1781–1786.

Adjorlolo-Johnson, G., De Cock, K.M., Ekpini, E., Vetter, K.M., Sibailly, T., Brattegaard, K., Yavo, D., Doorly, R., Whitaker, J.P., Kestens, L., Ou, C.Y., George, J.R., Gayle, H.D. (1994) Prospective comparison of mother-to-child transmission of HIV-1 and HIV-2 in Abidjan, Ivory Coast, *JAMA* **272**, 462–466.

Amara, R.R., Villinger, F., Altman, J.D., Lydy, S.L., O'Neil, S.P., Staprans, S.I., Montefiori, D.C., Xu, Y., Herndon, J.G., Wyatt, L.S., Candido, M.A., Kozyr, N.L., Earl, P.L., Smith, J.M., Ma, H.L., Grimm, B.D., Hulsey, M.L., Miller, J., McClure, H.M., McNicholl, J.M., Moss, B., Robinson, H.L. (2001) Control of a mucosal challenge and prevention of AIDS by a multiprotein DNA/MVA vaccine, *Science* **292**, 69–74.

Andreasson, P.A., Dias, F., Naucler, A., Andersson, S., Biberfeld, G. (1993) A prospective study of vertical transmission of HIV-2 in Bissau, Guinea-Bissau, *AIDS* **7**, 989–993.

Asjo, B., Morfeldt-Manson, L., Albert, J., Biberfeld, G., Karlsson, A., Lidman, K., Fenyo, E.M. (1986) Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection, *Lancet* **2**, 660–662.

Baltimore, D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses, *Nature* **226**, 1209–1211.

Barouch, D.H., Kunstman, J., Kuroda, M.J., Schmitz, J.E., Santra, S., Peyerl, F.W., Krivulka, G.R., Beaudry, K., Lifton, M.A., Gorgone, D.A., Montefiori, D.C., Lewis, M.G., Wolinsky, S.M., Letvin, N.L. (2002) Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes, *Nature* **415**, 335–339.

Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., Montagnier, L. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS), *Science* **220**, 868–871.

Beaumont, T., van Nuenen, A., Broersen, S., Blattner, W.A., Lukashov, V.V., Schuitemaker, H. (2001) Reversal of human immunodeficiency virus type 1 IIIB to a neutralization-resistant phenotype in an accidentally infected laboratory worker with a progressive clinical course, *J. Virol.* **75**, 2246–2252.

Berry, N., Davis, C., Jenkins, A., Wood, D., Minor, P., Schild, G., Bottiger, M., Holmes, H., Almond, N. (2001) Vaccine safety. Analysis of oral polio vaccine CHAT stocks, *Nature* **410**, 1046–1047.

Borrow, P., Lewicki, H., Hahn, B.H., Shaw, G.M., Oldstone, M.B. (1994) Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection, *J. Virol.* **68**, 6103–6110.

Brodie, S.J., Patterson, B.K., Lewinsohn, D.A., Diem, K., Spach, D., Greenberg, P.D., Riddell, S.R., Corey, L. (2000) HIV-specific cytotoxic T lymphocytes traffic to lymph nodes and localize at sites of HIV replication and cell death, *J. Clin. Invest.* **105**, 1407–1417.

Bukrinsky, M.I., Haggerty, S., Dempsey, M.P., Sharova, N., Adzhubel, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M., Stevenson, M. (1993) A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells, *Nature* **365**, 666–669.

Carrington, M., Nelson, G.W., Martin, M.P., Kissner, T., Vlahov, D., Goedert, J.J., Kaslow, R., Buchbinder, S., Hoots, K., O'Brien, S.J. (1999) HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage, *Science* **283**, 1748–1752.

Chackerian, B., Rudensey, L.M., Overbaugh, J. (1997) Specific N-linked and O-linked glycosylation modifications in the envelope V1 domain of simian immunodeficiency virus variants that evolve in the host alter recognition by neutralizing antibodies, *J. Virol.* **71**, 7719–7727.

Champagne, P., Ogg, G.S., King, A.S., Knabenhans, C., Ellefsen, K., Nobile, M., Appay, V., Rizzardi, G.P., Fleury, S., Lipp, M., Forster, R., Rowland-Jones, S., Sekaly, R.P., McMichael, A.J., Pantaleo, G. (2001) Skewed maturation of memory HIV-specific CD8 T lymphocytes, *Nature* **410**, 106–111.

Chen, M.Y., Maldarelli, F., Karczewski, M.K., Willey, R.L., Strebel, K. (1993) Human immunodeficiency virus type 1 Vpu protein induces degradation of CD4 in vitro: the cytoplasmic domain of CD4 contributes to Vpu sensitivity, *J. Virol.* **67**, 3877–3884.

Cheng-Mayer, C., Seto, D., Levy, J. (1988) Biologic features of HIV-1 that correlate with virulence in the host, *Science* **240**, 80–82.

Cho, M.W., Kim, Y.B., Lee, M.K., Gupta, K.C., Ross, W., Plishka, R., Buckler-White, A., Igarashi, T., Theodore, T., Byrum, R., Kemp, C., Montefiori, D.C., Martin, M.A. (2001) Polyvalent envelope glycoprotein vaccine elicits a broader neutralizing antibody response but is unable to provide sterilizing protection against heterologous Simian/human immunodeficiency virus infection in pigtailed macaques, *J. Virol.* **75**, 2224–2234.

Chun, R.F., Jeang, K.T. (1996) Requirements for RNA polymerase II carboxyl-terminal domain for activated transcription of human retroviruses human T-cell lymphotropic virus I and HIV-1, *J. Biol. Chem.* **271**, 27888–27894.

Cochrane, A.W., Perkins, A., Rosen, C.A. (1990) Identification of sequences important in the nucleolar localization of human immunodeficiency virus Rev: relevance of nucleolar localization to function, *J. Virol.* **64**, 881–885.

Dalgleish, A.G., Beverley, P.C., Clapham, P.R., Crawford, D.H., Greaves, M.F., Weiss, R.A. (1984) The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus, *Nature* **312**, 763–767.

Deacon, N.J., Tsykin, A., Solomon, A., Smith, K., Ludford-Menting, M., Hooker, D.J., McPhee, D.A., Greenway, A.L., Ellett, A., Chatfield, C. (1995) Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients, *Science* **270**, 988–991.

De Cock, K.M., Zadi, F., Adjorlolo, G., Diallo, M.O., Sassan-Morokro, M., Ekpini, E., Sibailly, T., Doorly, R., Batter, V., Brattegaard, K. (1994) Retrospective study of maternal HIV-1 and HIV-2 infections and child survival in Abidjan, Cote d'Ivoire, *BMJ* **308**, 441–443.

Dittmar, M.T., Simmons, G., Hibbitts, S., O'Hare, M., Louisirirotchanakul, S., Beddows, S., Weber, J., Clapham, P.R., Weiss, R.A. (1997) Langerhans cell tropism of human immunodeficiency virus type 1 subtype A through F isolates derived from different transmission groups, *J. Virol.* **71**, 8008–8013.

Eckert, D.M., Kim, P.S. (2001) Mechanisms of viral membrane fusion and its inhibition, *Annu. Rev. Biochem.* **70**, 777–810.

Felber, B.K., Drysdale, C.M., Pavlakis, G.N. (1990) Feedback regulation of human immunodeficiency virus type 1 expression by the Rev protein, *J. Virol.* **64**, 3734–3741.

Finzi, D., Hermankova, M., Pierson, T., Carruth, L.M., Buck, C., Chaisson, R.E., Quinn, T.C., Chadwick, K., Margolick, J., Brookmeyer, R., Gallant, J., Markowitz, M., Ho, D.D., Richman, D.D., Siliciano, R.F. (1997) Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy, *Science* **278**, 1295–1300.

Fischer, U., Huber, J., Boelens, W.C., Mattaj, I.W., Luhrmann, R. (1995) The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs, *Cell* **82**, 475–483.

Fouchier, R.A., Groenink, M., Kootstra, N.A., Tersmette, M., Huisman, H.G., Miedema, F., Schuitemaker, H. (1992) Phenotype-associated sequence variation in the third variable domain