4 Worlds Apart

It was already evident that multicellular eukaryotes are orders of magnitude more complex than bacteria. Humans, for example, have ~30–40 trillion cells[1,2] that are precisely assembled during embryonic and post-natal development into a myriad of different and precisely sculpted muscles, bone and other organs, and a brain with over 85 billion neurons and a trillion synaptic connections.[3–5]

Eukaryotic cells are also generally much larger[a] and have more complex organization than bacterial cells. They also have larger genomes,[b] especially in multicellular organisms, split between linear chromosomes, with variable but usually substantial amounts of 'repetitive' sequences (Chapters 5 and 10).

Eukaryotic cells have a membrane-bound nucleus where the chromosomes are located, an important consequence of which is the separation of transcription from translation[c] (Chapter 7). They also contain other internal membranous structures and membrane-bound 'organelles'[11] including caveolae (surface pits for endocytosis of external material);[12] endosomes (which traffic proteins, lipids and other components within the cell);[13] peroxisomes (where specialized oxidative metabolism takes place);[14] lysosomes (which degrade engulfed particles and intracellular components);[15–17] the endoplasmic reticulum (ER, the 'rough' form of which is studded with ribosomes);[18,19] the Golgi apparatus (a distribution network wherein proteins are imported from the ER, tagged with carbohydrates, sorted and packaged into endosomal vesicles destined for lysosomes, the cell surface or export,[20,21] named after their discoverer Camillo Golgi in 1898[22]); mitochondria (which generate energy by oxidation of carbohydrates and fatty acids);[23] and (in plants and algae) chloroplasts (photosynthetic energy capturing factories that produce sugars, sometimes called 'plastids')[24] (Figure 4.1). These organelles display an intricate degree of interaction and coordination.[25–30]

Eukaryotic cells also have many non-membrane-bound compartments, such as the nucleolus (the site of ribosomal biogenesis within the nucleus), which was first observed in 1835.[31] These compartments are phase-separated domains nucleated by RNAs and proteins containing intrinsically disordered regions (Chapter 16). Prokaryotic (bacterial and archaeal) cells have no internal structures as obvious as those in eukaryotes, although there is spatial organization and compartmentalization.[32–34] Phase-separated domains occur in prokaryotes and may predate cellular life (Chapter 16).

Recent advances in scanning electron microscopy and cryo-electron tomography have also enabled high-resolution imaging of cellular organelles and subcellular structures.[35–39]

THE ORIGIN OF CELLS

The origin of life involved the evolution of macromolecules capable of transmitting information and catalyzing biosynthetic reactions, as well as their encapsulation and the harnessing of energy – the reversal of entropy to create ordered systems, as first argued by Schrödinger in 1944.[40,41] There are two leading hypotheses about where and how this might have occurred: in deep ocean hydrothermal vents where proton gradients could form, proposed by Michael Russell, Nick Lane, Crispin Little and colleagues;[42–46] or in terrestrial hot springs where hydrothermal pools undergo wet-dry cycles (reprising Darwin's "warm little ponds"[47]) that favor the synthesis of organic polymers, lipids, peptides and nucleic acids, put forward by David Deamer, Martin Van Kranendonk, Armen Mulkidjanian, Eugene Koonin, Steve Benner and others,[48–55] with evidence favoring the latter.[47,49,50,56] There is also evidence that

---

[a] There are exceptions.[6]

[b] Bacterial genomes have a maximum size of around 10 Mb,[7] are usually circular and replicated bidirectionally from a single origin of replication, first shown by John Cairns in 1963.[8] Bacteria can also contain additional circular DNAs called plasmids, a term introduced in 1952 by Joshua Lederberg to refer to "any extrachromosomal hereditary determinant".[9] Plasmids often carry antibiotic resistance genes or others that confer selective advantage and may replicate autonomously or become integrated into the chromosome.

[c] Transcription and translation are coupled processes in bacteria. Translational stalling can result in transcription termination, as in the *trp* operon, where it is used to attenuate the production of tryptophan biosynthetic enzymes, shown by Charles Yanofsky in the late 1970s.[10]
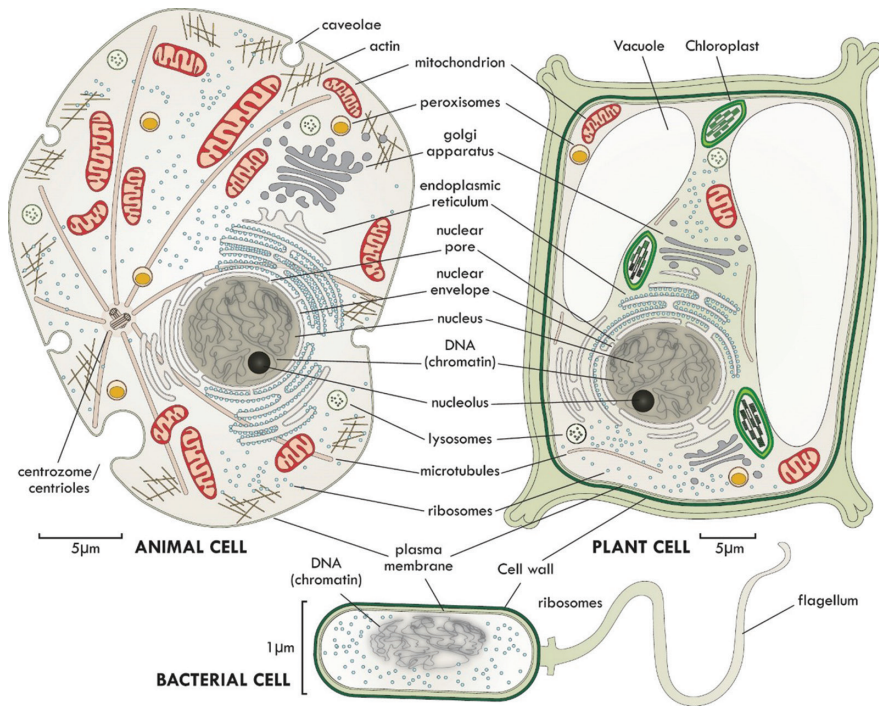
FIGURE 4.1   Comparison of prokaryotic and eukaryotic cell structures, original illustration by Heidi Cartwright.

the ubiquitous presence of ATP is due, in part, to its ability to aid protein solubility.[57]

In 1977, Carl Woese[d] and George Fox[59] made the unexpected discovery by ribosomal RNA gene sequencing that there are not two but three distinctive domains of life on Earth:[60] the unicellular Bacteria and the superficially similar Archaea, collectively called Prokarya, and the unicellular and multicellular Eukarya – protozoa, algae, fungi, plants and animals.[e]

The eukaryotes – the last common ancestor of which is estimated to have emerged around two billion years ago[63] – appear to have arisen from an archaeal progenitor that fused with a bacterium, as they possess many nuclear genes derived from each, but those involved in the core processes of DNA replication, transcription and translation, including the histones that are used to package eukaryotic

chromatin (Chapter 14), are clearly archaeal in origin.[64–74] Whether there are two (original) or (now) three primary branches of life[f] is a moot and almost semantic point.[76]

Mitochondria and chloroplasts are also descended from bacterial ancestors[g] captured by endosymbiosis,[24,67] proposed controversially by Lynn Margulis (Sagan) in 1967[81] but later confirmed by Robert Schwartz and Margaret Dayhoff.[82] Mitochondria and chloroplasts contain remnant small circular genomes[80,83–86] and bacterial-like translation systems for key hydrophobic proteins that must be made *in situ*.[87,88]

A plausible theory advanced by Tom Cavalier-Smith is that eukaryotes initially made their living as cellular scavengers and predators (think amoebae), which required the development of a flexible external membrane for phagocytosis.[89,90] It also required internal membranes for the protection of the genome and compartmentalization of lysosomes and other

---

[d] Woese was also one of the originators of the RNA World Hypothesis in 1967 (Chapter 9), although not by that name.[58] The 3-domain theory took many years to be accepted, although the data was clear.

[e] The terms 'eukaryote' and 'prokaryote' were first coined by Édouard Chatton in 1925, and the distinction formalized in 1962 by Roger Stanier and C. B. (Cornelius) van Niel, based on the presence or absence of a nucleus.[61,62]

[f] It has also been suggested that the giant viruses (Megavirales), discovered in amoebae in 2002, comprise a fourth super-kingdom of life.[75]

[g] Mitochondria are descended from a hydrogen-producing α-proteobacterium;[77,78] chloroplasts from a cyanobacterium.[79,80]

organelles, which is consistent with the flexible membranes and microvesicles observed in the extant lineage of the proposed archaeal ancestor of the eukaryotes.[69,72]

GENETIC RECOMBINATION

Whereas prokaryotes only have one genome copy (in addition to self-replicating extrachromosomal plasmids), eukaryotic cells are (usually) 'diploid',[h] having obtained one nuclear genome copy from each parent,[i] produced by the process of meiosis to form 'haploid' 'gametes' (sperm and ova). This and the elaboration of two sexes in eukaryotes may have arisen (although there are many possible explanations[92–95]) to allow recombinational exchange of larger genomes in complex cells and especially between multicellular organisms.[j]

In (most) eukaryotes, having two genome copies means having two copies of each of the genes therein, which are referred to as 'alleles' if they differ in any demonstrable way. These alleles may be dominant or recessive with respect to the other in their impact on phenotype, as demonstrated by Mendel. This arrangement provides the advantage that defective alleles can be tolerated, as having one functional version is usually enough to compensate,[k] which enables flexibility. Regulatory variation may be co-dominant, which allows more complex dynamics in the evolution and expression of quantitative traits.

DNA exchange occurs *ad hoc* in prokaryotes and at meiosis in eukaryotes, where it involves chromosomal pairing, formation of a 4-stranded cruciform structure between homologous DNA duplexes[l] ('crossing-over') and strand exchange.[100] Homologous recombination was the basis of genetic mapping by Benzer, Morgan and in the first half of the 20th century (Chapter 2), and later used to track down the protein-coding genes that are damaged in disorders such as cystic fibrosis (Chapter 11).

Recombinational exchange is an evolvability strategy that appears to have arisen at the dawn of life to enable genetic variations to be separated and discriminated by selection.[101] Gene assortment and fault tolerance were major considerations in the development of evolutionary theory and mathematical models of population genetics.

The evolution of the pathways and infrastructure for genetic recombination is an example of second-order Darwinian selection, where an accidental innovation has no particular or immediate phenotypic consequence but confers long-term advantages. There are almost certainly other evolutionary search optimization strategies that have not yet been recognized, because of the emphasis on phenotypic selection and the belief that mutation occurs randomly (Chapter 18).

It is also worth noting the growing appreciation of transposons (Chapters 5 and 10) and viruses,[m] the most abundant biological entities on Earth (which may have predated cellular life), as wider currencies for genetic exchange and dissemination, with central roles in the early evolution of cells, the 'invention' of DNA and DNA replication, the formation of the three domains of life and the diversification of multicellular organisms.[105–110]

THE EMERGENCE OF COMPLEX ORGANISMS

Multicellular plants appeared around 1–1.2 billion years ago, initially in the oceans, then in freshwater environments.[111,112] They colonized the land around 850 million years (Myr) ago[113] and diversified into more complex vascular forms with roots, leaves

---

[h] Some are 'polyploid', such as wheat, which has six copies of each chromosome (hexaploid), and whose gametes have three.

[i] Mitochondrial genomes are maternally transmitted in ova, the source in part of the cytoplasmic inheritance of some characteristics first observed by the Boveris around 1900 (Chapter 2). Chloroplast genomes are usually also maternally inherited.[91]

[j] Linear chromosomes may be required for meiotic segregation.[96] They are copied via multiple replication origins and an overlapping series of bidirectional replication bubbles, a highly complex process that is tightly controlled during differentiation and development[97] (Chapter 15).

[k] Some defective genes can be dominant for mechanistic reasons, such as those encoding proteins involved in multi-component complexes, because only one copy is expressed (as in parental imprinting; Chapter 5) or because they only occur on sex chromosomes, as exemplified by the higher frequency of color blindness in males (the genes are located on the X chromosome). Some 'heterozygous' combinations of functional and defective alleles produce intermediate effects, called 'haploinsufficiency', which may be more common than appreciated and have benefits in some circumstances.

[l] Called a 'Holliday junction', after Robin Holliday who proposed it in 1964.[98,99]

[m] Norton Zinder and Joshua Lederberg showed in 1952 that bacterial viruses can integrate into the bacterial genome,[102] providing an explanation for the lysogeny phenomenon earlier described by Eugene Wollman and André Lwoff, which became an important genetic tool for gene mapping.[103] Animal retroviruses were described by multiple groups in the 1960s and 1970s.[104]

and seeds between 500 and 200 Myr ago, when the angiosperms (flowering plants) emerged following an ancient genome duplication.[114,115] The phylogenetic tree from algae to angiosperms is now being constructed from DNA and RNA sequence data.[116] Interestingly, improvements in the efficiency of CO_2 fixation, known as the C4 pathway, appeared just 25–32 Myr ago, likely as an adaptation to declining CO_2 concentrations due to biological sequestration into calcium carbonate and carbonaceous deposits.[117]

Sponges, the most primitive of the animal phyla, existed around 890 Myr.[118] The first animals with complex body plans, the Ediacaran fauna,[n] appear in the fossil record between 620 and 550 Myr ago in an evolutionary radiation called the Avalon explosion,[120] with antecedents up to 800 Myr ago.[111] The Ediacaran fauna were soft-bodied organisms, ranging in size from 1 cm to over 1 m, likely making a living as scavengers – like fungi, with which animals share a common ancestor.[121–124] Ediacarans had radial or bilateral symmetry and segmented tube-, quilt- or frond-like structures, some with similarities to modern worms and jellyfish,[125–129] which may be their descendants.[130–133]

The Ediacarans were largely supplanted around 540–500 Myr ago by a second and more spectacular large-scale radiation. It is known as the Cambrian explosion, first revealed by fossils in the Burgess Shale of the Canadian Rocky Mountains, discovered by Richard McConnell in 1886 and characterized in detail in the early 20th century by Charles Walcott and others since, notably Simon Conway Morris,[134] in many Cambrian 'Lagerstätten'. In one strata of rock, and an estimated time window of ~10–20 Myr, recognizable ancestors of all extant metazoan phyla appear, including arthropods and chordates, with hard skeletons, advanced predation and locomotive capacity, along with other bizarre forms[135–139] (Figure 4.2), likely driven by the evolution of macrophagy.[140]

Soon after, fish were swimming in the oceans.[141] Similar rapid phenotypic diversifications also occurred after later mass extinction events,[142] including that following the meteorite strike in the Gulf of Mexico 66 Myr ago, which wiped out the non-avian dinosaurs and allowed the rise of mammals into vacated ecological niches.[143,144]

The initial appearance and rapid evolution of animals has commonly been thought to have been potentiated by the increase in atmospheric oxygen from photosynthesis and the advantages of aerobic energy generation by mitochondrial electron transport[o] in eukaryotes.[132,146–148] However, there was substantial atmospheric oxygen long before the evolution of animals,[149,150] and it seems that sufficient oxygen may have been an enabler but not a direct cause of their emergence.[151,152] Rather, the transition from unicellular to developmentally complex organisms with highly organized assemblages of specialized cell types was likely achieved by advances in genome organization and regulatory systems (Chapters 14–16). Later transitions, undoubtedly also requiring genetic innovations, occurred in the colonization of the land, to enable physiological adaptability to a more variable environment and more complex structures for terrestrial mobility.

CHROMATIN

The most obvious molecular genetic difference between prokaryotes and eukaryotes is that the much larger genomes of the latter are not only sequestered in a nucleus but also segmented into chromosomes and packaged into complex chromatin structures.

Eukaryotic chromatin is not homogeneous, and contains regions with different properties, broadly divided into open 'euchromatin' (gene rich, transcriptionally active, lightly stained) and compacted 'heterochromatin' (transcriptionally quieter, densely stained), first documented in the late 1920s by Emil Heitz,[153] a pioneer of cytogenetics,[154] notably in the giant 'polyploid' or 'polytene' chromosomes[p] in the salivary glands of insects.[157–159]

Dynamic changes in chromatin were observed in the appearance and disappearance at different developmental stages of 'facultative' heterochromatin by Heitz and others,[153] and 'puffs' in polytene chromosomes formed by localized decondensation of small

---

[n] Named after the Australian site where they were found in abundance by Reginald Sprigg in 1947.[119]

[o] There are conflicting views.[145]

[p] Polytene chromosomes were first observed by Édouard-Gérard Balbiani in 1881, who is remembered in the term 'Balbiani ring' which refers to the large chromosomal puffs where transcription occurs. They contain multiple DNA molecules in parallel, generated by multiple rounds of DNA replication without an intervening cell division.[155] Their banded structure corresponds to topologically associated domains (Chapter 14), which are preserved between polytene and diploid cells.[156]
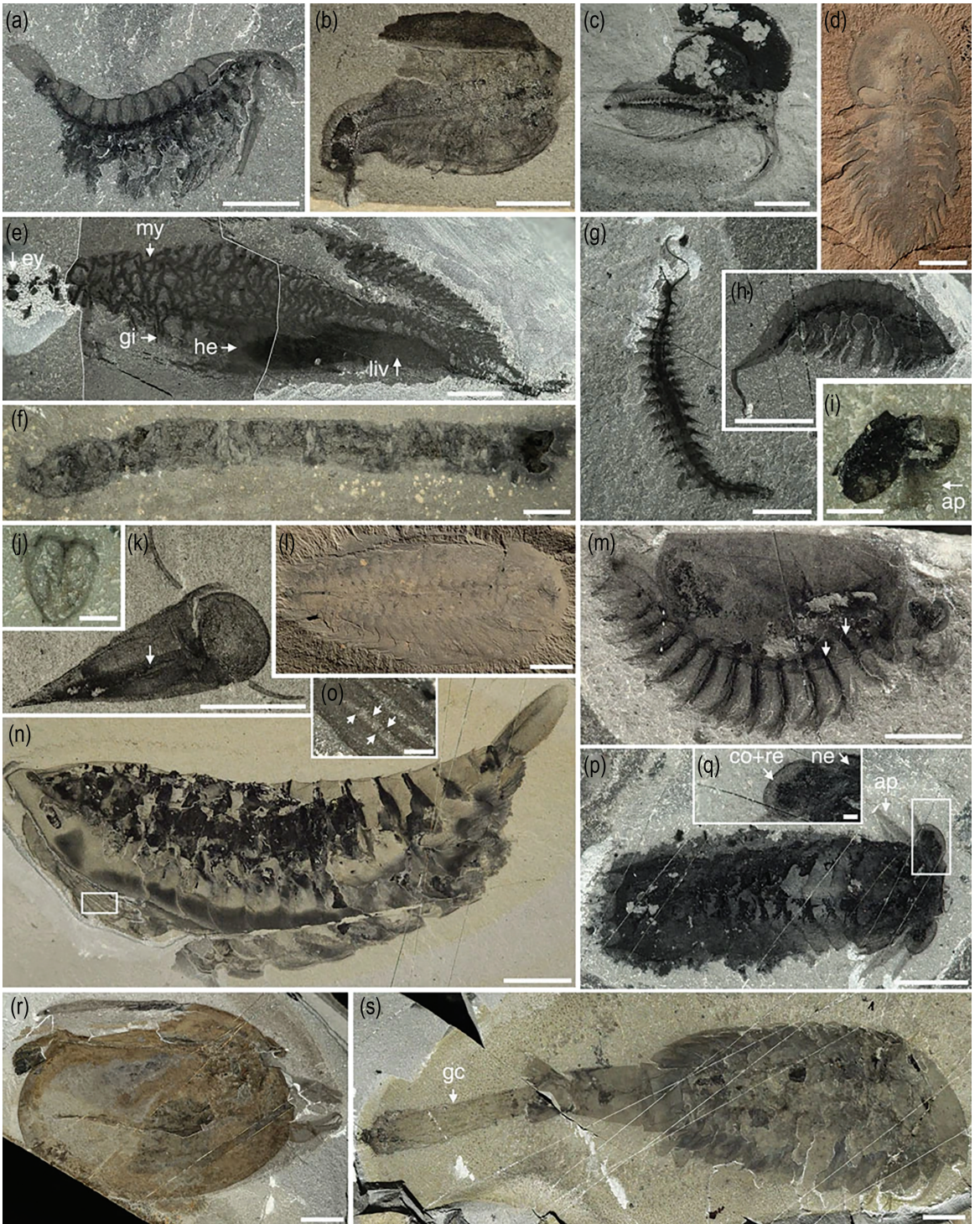
FIGURE 4.2 Cambrian fossils from the Burgess Shales, including ancestral arthropods (a–d, h–j, l–s), primitive chordates (e, f), annelid (g) and mollusk (k). (Reprinted from Caron et al.[139] by permission of Springer Nature.)
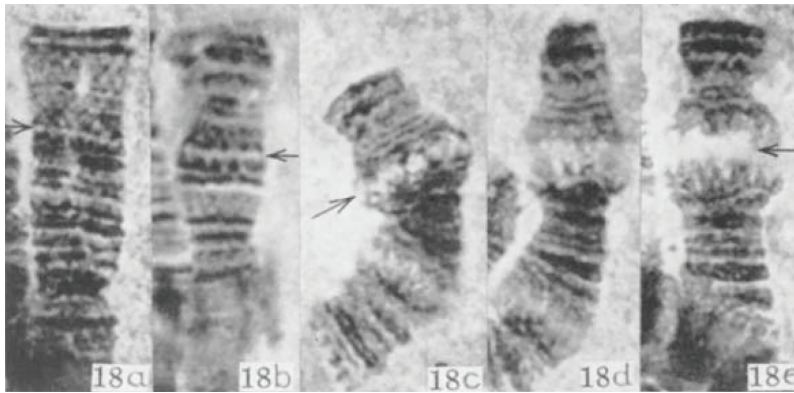
FIGURE 4.3 Micrographs of banded polytene chromosomes, arrows indicate 'puffs'. (Reprinted from Poulson and Metz [160] by permission of John Wiley and Sons.)

chromosomal segments, described by Donald Poulson and Charles Metz in 1938 (Figure 4.3), and by others in the 1950s and 1960s.[160–162] Puffs exhibit developmental stage- and tissue-specific patterns,[163,164] can be induced by heat shock[165] and hormones such as ecdysone,[166] and are sites of RNA synthesis.[167–172] In 1973, it was shown by Adolf and Monika Graessman that puff induction involves RNA,[173] and later by Subhash Lakhotia and colleagues that the product of a puff induced by heat shock is an RNA that does not encode a protein[174,175] (Chapter 9).

Specific heterochromatic regions of eukaryotic chromosomes form centromeres,[176,177] which act as organizing centers of cell division, described first by Edouard van Beneden and then by Boveri (who coined the name) in the 1870s and 1880s.[178,179] Centromeres contain internal granules (called 'centrioles' by Boveri) and attach to kinetochores for spindle formation and chromatid pairing and separation to daughter cells during mitosis and meiosis[180] (see Chapter 15).

In 1959, Susumu Ohno showed that one of the two X-chromosomes in female mammals is heterochromatic[181] (called 'nucleolar satellite' and later the 'Barr body' after its discoverer, Murray Barr[182]). In 1961, Mary Lyon demonstrated that X-chromosome inactivation occurs randomly in early embryogenesis:[183,184] females are mosaics of active X-chromosomes inherited from either parent,[q] a 'dosage compensation'

mechanism to equalize with males, who only have one X-chromosome – a traditional system in genetic and cytological studies[r] (Chapter 2) later shown to be controlled by RNA[s] (Chapter 9). It was also known that in some insects one entire set of chromosomes becomes heterochromatic during male early embryonic development,[189] and that chromosomes in embryonic cells often have a different morphology than those in adult cells.[190]

The existence of 'facultative' heterochromatin, position effect variegation, chromosomal puffs and 'lampbrush' chromosomes[t] (described by Alexander Flemming in 1882[193]), which occur in the oocytes of all animals except, curiously, mammals,[194] suggested that there are higher-order genomic arrangements and additional modes of gene regulation during plant and animal development.[153]

It was also found that eukaryotic DNA is wrapped

---

[q] The mosaic pattern of X-chromosome inactivation can be observed in variegated coat colors, such as in 'tortoise shell' cats, which are almost invariably female (except when chromosomal aberrations such as XXY occur[185]), and in the mosaic pattern of sweat glands in women.

[r] X-linked (sex-linked) traits have also been of great value to human genetics (Chapter 11), given that recessive mutations are exposed in males, classic examples being red-green color-blindness and Duchene muscular dystrophy,[186] with variable intermediate phenotypes (severity of effect) in females because of mosaic expression.[187]

[s] In *Drosophila*, dosage compensation is achieved not by inactivation of one of the X-chromosomes in females, but by global upregulation of the activity of the single X-chromosome in males,[188] which also centrally involves non-coding RNAs (Chapter 9).

[t] In the late 1950s, electron microscopic visualization of elongating transcripts on lampbrush chromosomes first suggested rapid packaging of nascent RNAs with proteins.[191] In the 1970s, several laboratories began focusing on biochemical purification and compositional/structural analysis of non-ribosomal ribonucleoproteins, which led to identification of mRNA cap–binding proteins, polyA-binding protein, pre-mRNA splicing proteins (Chapter 8) and mRNA transport proteins, among others.[192]
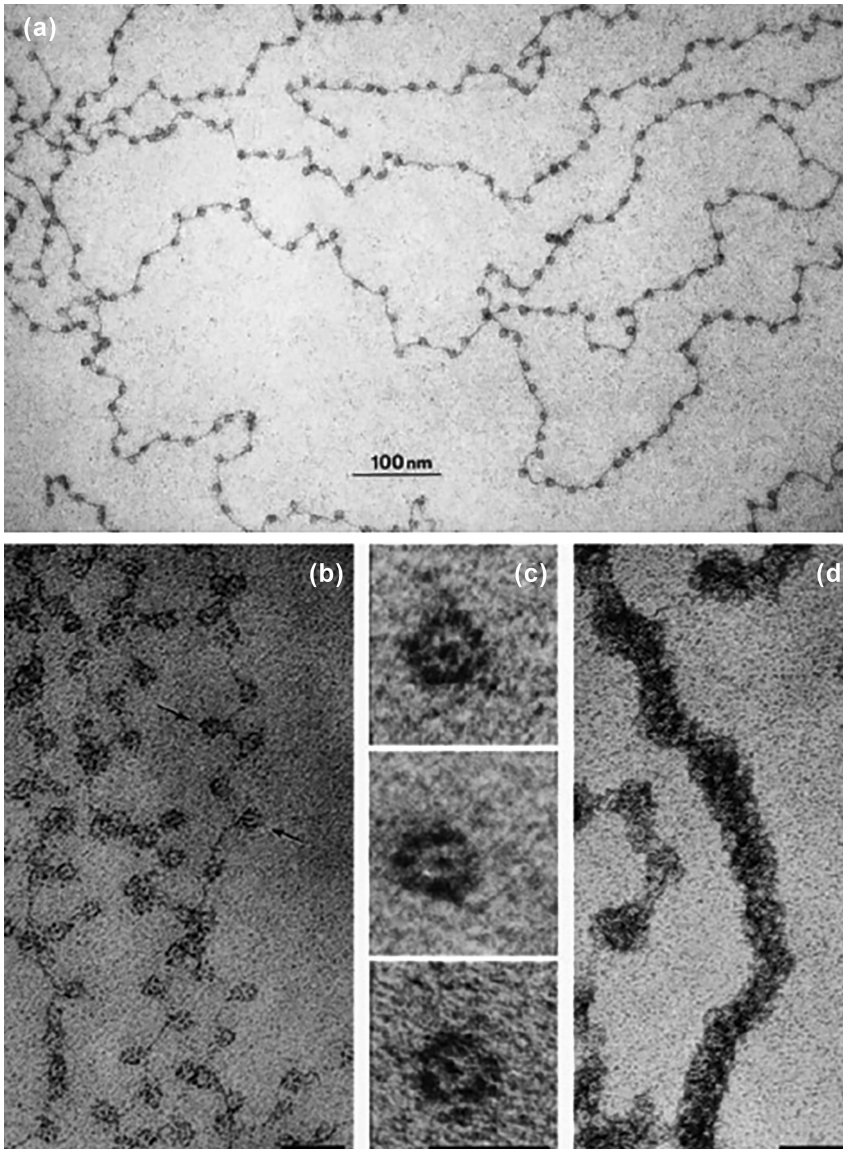
FIGURE 4.4 Electron micrographs taken by Donald and Ada Olins of (a) and (b, size marker 30 nm) low ionic strength chromatin spreads showing the 'beads on a string'; (c) nucleosomes derived from nuclease-digested chromatin (size marker 10 nm); (d) chromatin spread at a moderate ionic strength showing a 30 nm higher-order structure (size marker 50 nm). (Reproduced from Olins and Olins[201,202] with permission of American Scientist (a) and Springer Nature (b–d).)

around proteins called histones,[u] like cotton around a spool in a repeating structure, called 'nucleosomes'.[198] Histones were identified by Kossel in 1894,[199] but it was another 80 years before nucleosomes were visualized in the electron microscope by Ada and Donald Olins[200–202] (Figure 4.4) and their octameric histone complement defined by Roger Kornberg and colleagues.[203,204] It was even longer before it became evident that histones are the major repositories of epigenetic information (Chapter 14), although hints of a role in gene regulation were emerging.

[u] Eukaryotic histones have ancestral orthologs in archaea,[70,71,74] where they regulate gene activity in response to environmental circumstances.[195] Histones have been shown to have copper reductase activity,[196] suggesting their original role was to protect against oxygen toxicity.[197]

In 1950, Ellen and Edgar Stedman[v] proposed that histones are repressors that could inactivate genes in a tissue-specific manner, based on the quantities of histones in growing and non-growing tissues.[205,206] Their proposal was supported by work of Ru-chih Huang and James Bonner, and Vincent Allfrey, Alfred Mirsky and colleagues, who found that histones inhibit the transcription of DNA *in vitro*.[207,208]

However, histones, superficially at least, displayed uniformity between tissues and species, as well as between 'repressed' and active chromatin (see below), which led John Frenster, Allfrey and Mirsky to conclude in 1963 that they were unlikely to be gene-specific regulators.[209] For decades thereafter, nucleosomes were considered primarily a mechanism for compacting large genomes,[210–212] given the widespread conviction that transcription factors are the primary means of gene regulation.[213]

In the mid-1960s, Allfrey and Mirsky proposed that post-translational modifications of histones (acetylation and methylation) have regulatory functions.[214,215] They showed that lymphocyte activation triggers massive acetylation of chromatin[216] and that histone acetylation also occurs in insects.[217] A decade on DNA methylation was also suggested as a mechanism to regulate gene activity,[218,219] although these ideas would only be tested and confirmed much later[220–223] (Chapter 14).

Beyond this, how the structure of chromatin was organized and how it affected gene expression in eukaryotes was unknown; progress was slow because of the sheer size and complexity of the genomes and chromosomes, and the difficulties of working with all but unicellular models such as yeast.

CHROMATIN-ASSOCIATED RNAs

By this time, it was also clear that RNA is the third component of chromatin. In his early histological experiments on the distribution of DNA and RNA, Brachet showed that RNA is present not only in the cytoplasm but also in chromatin,[224,225] subsequently found by Mirsky and Hans Ris to reside in a NaCl insoluble fraction, which comprised only ~10% of the total but retained all the usual features

of chromatin.[226] RNA was also reported to remain attached to chromosomes during cell division.[227]

Indeed, while largely overlooked during the heady days of the genetic code, a number of publications in the 1960s and early 1970s reported the presence of RNA in chromatin fractions,[228] some of which were proposed to be structural and regulatory agents. These included Frenster's 1965 model of 'De-repressor RNAs', based on the observation that RNA added to heterochromatin fractions increased the level of transcription, which was most pronounced when nuclear RNAs were added (compared to cytoplasmic and non-specific RNAs such as rRNA or yeast RNA), posited to involve RNA hybridization to complementary sequences in the repressed DNA.[229,230]

In 1965 also, Huang and Bonner reported the presence of low molecular weight RNAs in chromatin. These "chromosomal RNAs" (cRNAs) were protected from RNase degradation and corresponded to ~8% of the total nucleic acid mass present in nucleohistones.[231] This was the first in a series of reports of the existence of tissue-specific short RNAs in chromatin in plants and animals that associate with non-histone chromatin proteins and can hybridize to homologous DNA,[232–236] leading to the hypothesis that cRNAs had a role in regulating gene expression.[235,237–239]

These short RNAs are not precursors for any cytoplasmic product[240] and some were distinguished by a high content of methylated or dihydropyrimidine nucleotides,[233,240,241] a signature of small nucleolar and small spliceosomal RNAs (Chapter 8).

Interestingly, cRNAs were found to hybridize extensively to "middle repetitive" DNA sequences, which Bonner proffered as evidence that repetitive sequences may be regulatory elements.[237,242,243] These observations would have impact on models of genome regulation in the higher organisms (Chapter 5), but were sidelined later by the widespread assumption that much of the genomes of higher organisms is junk, partly and ironically because they contained so many 'repetitive' sequences (Chapters 7 and 10).

Soon after the Frenster and Bonner publications, William Benjamin and colleagues reported that RNA isolated from a rat liver nucleoprotein fraction co-sedimented with histones.[244] This RNA had high adenine and uridine content and had heterogeneous sizes by sucrose gradient analysis, adding to the complexity of the types of RNAs of unknown functions found in the eukaryotic nucleus. Although

---

[v] The Stedmans had earlier suggested that non-histone chromosomal proteins, which they called "chromosomins", represent the "basis of inheritance" and are also involved in gene regulation, predicting that the physical association of chromosomins and nucleic acids was required for synthesis of specific proteins.[205]

recognizing that these RNAs might represent an intermediate in the synthesis of mRNA, Benjamin et al. also speculated that these RNAs could play a role in the control of gene expression, invoking Paul Sypherd and Norman Strauss' 1963 suggestion that regulatory systems involve both protein and RNA,[245] such that associated RNAs might confer specificity to repressive histones by base-pairing with the target gene[244] (Chapter 16).

During the 1970s, nuclear RNAs began to be better characterized. Some were visualized with chromatin at specific stages of the cell cycle and proposed to act as "programmers" of "chromosomal information" and gene regulation.[246] *In vitro* experiments by Takeharu Kanehisa and colleagues with purified chromatin indicated that specific short chromatin-associated RNAs could "modify" chromatin structure and stimulate RNA synthesis, particularly in chromatin isolated from the same tissue, suggesting a tissue-specific effect.[247–249]

In 1973, Isaac Bekhor showed that "chromosomal RNA-protein complexes" can interact with DNA *in vitro* and increase its melting temperature, indicating a stabilizing effect, leading him to postulate that cRNAs present in chromosomal RNA-protein complexes constituted a structural component of chromatin, rather than regulatory molecules,[250] an idea favored by other studies reporting RNAs associated with heterochromatin.[251] In 1978 Sheldon Penman's group showed that stable species of high molecular weight RNAs also associate with nuclear complexes, from which it was again hypothesized that "RNA networks" had structural roles in the nucleus.[252]

Thoru Pederson and Jaswant Bhorjee reported in 1979 that three short RNAs are associated with chromatin and favored the hypothesis that these "DNA-linked RNAs" are involved in the control of the tertiary structure of chromatin.[253] These RNAs were ~130–200-nt long, highly abundant, relatively stable, and were designated small nuclear RNAs D, C and G'[253] (later small nuclear RNAs U1, U2 and U5, respectively, Chapter 8). They showed that only a fraction (<10%) of these RNAs is associated with chromatin, supporting the earlier results of Mirsky and Ris, while the remainder was nucleoplasmic.[253] Others reported that separated fractions contained between 6 and 11 size classes of small RNAs, most of which seemed to be reversibly bound to chromatin proteins,[233,254,255] but some of which did not dissociate in high salt concentrations, possibly reflecting RNA-DNA hybrids.[256]

Later studies using hybridization and cytogenetic techniques showed that cRNAs from human placenta displayed a widespread pattern of hybridization to metaphase chromosomes, preferentially in telomeric regions and heterochromatic short arms of acrocentric chromosomes, as well as regions with a high content of repetitive DNA.[257] Chromatin fractionation studies using Frenster's method had also indicated that heterochromatin contains a more stable fraction of chromatin-associated RNAs compared to euchromatin.[251]

Other studies focused on the effects of high molecular weight RNAs and nascent transcripts in chromatin and in other nuclear structures. Jean-Pierre Bachellerie and colleagues used subnuclear fractionations, autoradiographic and ultrastructural techniques to show that "perichromatin fibrils represent the morphological state of newly formed heterogeneous nuclear RNA (hnRNA)".[258]

Much of this would make sense later (Chapters 7, 8 and 16), but at the time there was a fierce debate over the existence, biological relevance and specificity of chromatin-associated RNAs. The concerns ranged from the reproducibility of the findings, the questionable purity of cellular and chromatin fractions (complicated by a plethora of fractionation procedures), the presence of nucleases that could lead to contamination with degradation products of tRNAs, rRNAs and heterogeneous nuclear RNAs (see below), and the feeling that the models proposed were too speculative.[259–264]

On the other hand, a number of lines of evidence were proffered against the criticisms, including that cRNAs have characteristic elution properties, their complexity and hybridization kinetics differed from common RNAs such as tRNA and rRNA, they had a different stability, and chromatin preparations using stringent methods contained a reproducible RNA fraction.[233,236,239,243,256,265,266]

By the end of the 1970s, different groups that had analyzed the composition of total chromatin estimated that the RNA/DNA ratio of the chromatin was approximately ~0.05–0.2 in different eukaryotes, and that distinct classes of RNAs with specific properties were chromatin-associated, including species of varying stability and molecular weight,[253,256,267–269] which would include new classes of infrastructural RNAs involved in rRNA biogenesis and splicing (Chapter 8).

Nevertheless, uncertainty about chromatin-associated RNA remained because the characteristics

of the reported RNA profiles varied by tissue and the method of chromatin isolation. The exact nature of these RNAs was unknown, given that the methods of identification were rudimentary at that time. It was a complicated muddle.

EARLY MODELS OF RNAs IN NUCLEAR ARCHITECTURE

There was also emerging evidence that RNA is involved in the organization of the nuclear 'matrix', a fibrous structure first reported in 1948,[270] although its composition, stability and function has been the subject of ongoing conjecture.[271]

Penman recognized that chromosomes are not randomly distributed in the nucleus, much later described in detail (Chapter 14), and that the nuclear matrix played a central role in its three-dimensional architecture. His group used a chromatin-depletion strategy to show that ribonucleoprotein networks extend throughout a nuclear structural lattice and that the integrity of nuclear and chromatin architecture is dependent on RNA, as indicated by the treatment of cells with transcription inhibitors or extracted nuclei with RNase.[w] Penman concluded that RNA is not only a structural component of the nuclear matrix (an "RNA-dependent nuclear matrix") but also organizer of higher-order structures of chromatin ("architectural RNA"),[273,274] an idea that would reemerge powerfully much later when it was discovered that RNA and transcription modulate chromatin territories and nucleate subcellular domains (Chapter 16).

Similar ideas were elaborated by others, including the 'Unified Matrix Hypothesis' postulated in 1989 by Klaus Scherrer, in which the transcribed and non-transcribed part of non-coding DNA would have a direct morphogenic function.[275] According to this hypothesis, these regions have an intrinsic role in "the tridimensional network of chromatin and nuclear topological organization".[275] This was posited to explain phenomena such as the 'Chromosome Fields' with co-localization of linked genetic loci within chromosome regions and the specificity of sites of chromosome recombination in cancer.[275] Scherrer also suggested that RNA processing may play a role in nuclear architecture by organizing the selective transport and control of individual

transcripts, which would then act as signals for specific proteins and in combination define the nuclear matrix.[276]

HETEROGENEOUS NUCLEAR RNA

Radioactive labeling kinetics, sucrose gradient sedimentation and hybridization studies in the early 1960s by Scherrer, James Darnell, Georgii Georgiev and colleagues indicated that rapidly labeled transcripts were formed in the nucleus of mammalian cells, which exhibited high molecular weight and heterogeneous sedimentation profiles (Figure 4.5), AU-rich composition and unstable character.[277–282] These unexpected 'giant' nuclear transcripts were only broadly defined and described variously as DNA-like RNA ('dRNA'),[278] nascent messenger-like RNA (nascent 'mlRNAs')[281] and heterogeneous (or heterodisperse) nuclear RNA ('hnRNA').[282]

In 1963, Scherrer and Darnell showed that ribosomal RNAs in human cells are initially produced as large precursor molecules and subsequently processed into mature rRNAs,[279,283] confirmed by Penman and Guiseppe and Barbara Attardi,[284,285] and shown to take place in the nucleolus.[286,287] This may have been an idiosyncratic feature of the ribosomal operons, but they also reported that other "giant" RNAs with "messenger RNA properties" also exist in the nucleus.[283,284,288]

The biological significance of the large heterogeneous transcripts was puzzling. The difference in base composition between hnRNA and cytoplasmic mRNA, as well as differences in their half-lives, did not suggest a simple relationship.[289] Henry Harris noted in 1965 that "only a small proportion of the RNA made in the nucleus of animal and higher plant cells serves as a template for the synthesis of protein", and considered that "most of the nuclear RNA, however, is made on parts of the DNA which do not contain information for the synthesis of specific proteins. This RNA does not assume the configuration necessary for protection from degradation and is eliminated" (quoted in [290]). Harris later noted that "pulse-labeled RNA was almost universally misdiagnosed as messenger RNA" and that other suggestions were considered profoundly heretical at the time.[291]

In 1966, Scherrer was the first to propose that hnRNAs are precursors of mRNAs[288] but, despite his intense efforts, his proposal was not widely entertained.[276] In 1968, Penman showed that

---

[w] RNase treatment had been used previously (in 1974) to show that RNA is involved in the maintenance of condensation of dinoflagellate chromosomes.[272]
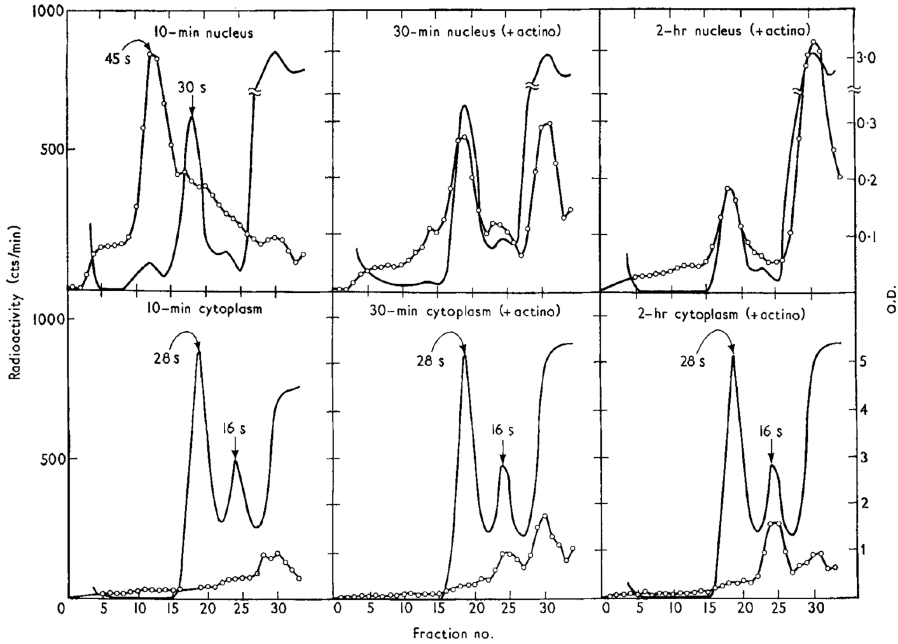
FIGURE 4.5 Sucrose gradient profiles of pulse-labeled DNase-treated nuclear and cytoplasmic preparations before and following actinomycin D treatment (which blocks RNA synthesis) showing the presence of high molecular weight RNAs in the nucleus but not the cytoplasm. Open circles, radioactivity; continuous line optical density at 260 nm, which detects nucleic acids. (Reproduced from Penman[286] with permission of Elsevier.)

"heterogeneous nucleoplasmic RNA" is produced in the absence of ribosomal RNA synthesis (blocked by specific concentrations of actinomycin D) and is turned over rapidly, with a mean life of approximately 1 hour, although he did not think that this might be a precursor to mRNA.[292,293] Penman also showed that the size of hnRNA increased with increased genome size,[294] all of which is consistent with the later discovery of introns and large pre-mRNA primary transcripts that are spliced to produce mRNAs (Chapter 7).

Hybridization studies by Allfrey and Mirsky indicated that while ~80% of the DNA was not accessible for transcription, transcription products covered up to 20% of the DNA in a given mammalian cell.[295] This and similar observations were potentially relevant for gene regulation because, given that the majority of the genome DNA was inactive in a particular cell type, it implied a mechanism of genome repression and allowed the suggestion of specialization of chromosomal regions for the control of growth and development of differentiated tissues.[208,295]

In 1967, Ruth Shearer and Brian McCarthy showed that while all the sequences of cytoplasmic

RNA were present in the nucleus, the latter contains a much greater fraction of sequences that are not exported to the cytoplasm, but rather are rapidly turned over,[296] confirmed by others.[297,298] They went on to say: "The existence of RNA molecules specific to the nucleus suggests a role as mediators of the regulation of gene transcription … although these functions are entirely speculative, the finding that the majority of the active genome codes for short-lived RNA molecules which are restricted to the nucleus opens up exciting possibilities for the study of the regulation of gene action in mammals."[296]

HEROES OR FOOLS?

Some working in animal genetics at the time, such as Ed Lewis (Chapter 5), drew the obvious conclusion that that "the 'genome' of phage and bacteria may be structurally organized in manner different from the chromosomes of higher forms".[299] However, not much notice was taken. The *lac* operon dominated the models of gene organization and the regulation of gene expression in eukaryotes; it was the basis of the ruminations by Jacob and Monod,[300] and others such as Ernst Mayr[301] on "genetic programmes"

and gene networks (or "nets of interacting genes"), in what developed into the belief that the combinatorial action of 'transcription factors' is sufficient to execute complex developmental programs[302–304] (Chapter 15).

In a 1969 paper entitled 'On the Structural Organization of Operon and the Regulation of RNA Synthesis in Animal Cells', Georgiev assumed that the principles of regulation of transcription defined in bacteria are retained in multicellular organisms.[304] He defined an operon as an "elementary unit of transcription" and proposed that operons in higher organisms consisted of a promoter-proximal regulatory "zone" and a structural zone that contained the coding sequences of several mRNAs with related functions, as in the *lac* operon. In his view, the entire operon would be transcribed as a "giant D-RNA", with regulatory sequences at the 5' end degraded in the nucleus and the mRNA transferred into the cytoplasm. Being aware of Roy Britten's findings regarding the abundance of repeat regions in the genome,[305] he also posited that repetitive sequences could be present in many operons and be targets for regulatory proteins, thereby assigning a functional role for the repetitive sequences scattered throughout the genome, similar to that proposed in the same year by Britten and Eric Davidson (Chapter 5).

Scherrer also proposed (in 1968) that hnRNAs are polycistronic precursors of mRNAs and be subject to both transcriptional and post-transcriptional regulation, the "cascade regulation" model.[281] Accordingly, a polycistronic hnRNA may be cleaved sequentially in the nucleus or in the cytoplasm to generate multiple mRNAs in a regulated fashion. Indeed, this logic assumed that "as a consequence of the central dogma, postulating that gene activity leads to phenotypic expression of genetic information through the mediation of mRNA, localized genes related to particular phenotypic characteristics should produce mRNA".[281]

In the following years many observations pointed to a relationship between hnRNA and mRNA.[306] In 1971, Darnell, George Brawerman, Mary Edmonds and colleagues found that eukaryotic mRNAs[x] and hnRNAs both contain an extended sequence of adenines ('polyA tails')[y] at their 3' end,[314–316] confirmed by others,[316–318] and that eukaryotic mRNAs are derived from longer precursors.[319–321] In 1974, Kin-Ichiro Miura, Yasuhiro Furuichi, Aaron Shatkin, Fritz Rottman and colleagues showed that eukaryotic mRNAs and pre-mRNAs both also contain an inverted modified (methylated) nucleotide ('cap') structure[z] at their 5' end[324–329] (m7G, later shown to play a role in RNA splicing and translational control[330,331]), also confirmed,[332–334] along with a contemporary report of widespread methylation of mRNA[335] (Chapter 17).

These findings supported the conclusion that hnRNAs are, in fact, precursors to mRNAs,[316] with a suggestion that mRNA might be comprised of fragments from each end of the hnRNA,[306,318] but the idea that hnRNAs are mRNA precursors was only accepted and understood after the discovery of introns and splicing in 1977[276] (Chapter 7).

On the other hand, it was gradually recognized by the late 1960s that the multigenic operon was an unlikely mode of eukaryotic genome organization and regulation since, for example, related genes such as those encoding alpha and beta hemoglobins or the enzymes for galactose metabolism were not co-localized in the genome, and the length of hnRNAs was much larger than that which would be expected for reasonably sized polycistronic mRNAs.[281]

Because the relation of hnRNAs and mRNA was not yet established, Scherrer and Lise Marcaud contemplated that hnRNAs might contain sequences "other than those carrying structural cistrons", such as regions for interaction with proteins through secondary and tertiary structures, thus conferring specificity to the "functional RNA".[281] Alternatively,

---

[x] Prokaryotic mRNAs can also contain a shorter polyA sequence at their 3' end.[307]

[y] The existence of 3' polyA tails in mRNAs was exploited and remains widely used in cloning and sequencing protocols (Chapter 6), to separate mRNA from the large amounts of ribosomal, transfer and other RNAs in cells by its hybridization to oligomers of U or T ('oligo dT') affixed to solid or colloidal surfaces.[308]

Later the 5' cap would also be exploited for mRNA purification,[309] although it turned out that many other transcripts that do not encode proteins are also capped and polyadenylated (Chapter 13). Moreover, while widely overlooked, a large fraction of the RNAs detected in human cells are not polyadenylated, although they are often capped.[310–313]

[z] In fact, there are at least 25 different types of 5' caps in eukaryotic cells, at least some of which are cell- and tissue-specific, with roles in the initiation of protein synthesis, protection from exonuclease cleavage and as identifiers for recruiting protein factors for pre-mRNA splicing, polyadenylation and nuclear export.[322,323]

it was suggested that these RNAs might have independent regulatory roles, such as interacting with other regulatory molecules (inducers and repressors) or regulating allosteric proteins.[281]

In one way or another, many of these speculations would prove true, but the technology of the time was still too limited to test them, and animal and plant systems were much too complicated. At that time only a handful of laboratories tried, valiantly, to understand genetic information and gene expression in eukaryotes. As recalled by Scherrer,

> only a few investigators were interested in the molecular biology of animal cells; the 'serious' research was with *E. coli* and bacteriophages. James Watson visited MIT frequently, and would discuss our strange results. One day, he told me 'To work with animal cells, you've got to be a hero or a fool!'[276]

FURTHER READING

Brown S.W. (1966) Heterochromatin. *Science* 151: 417–25.

Cooper G.M. (2000) *The Cell, A Molecular Approach* (Sinauer Associates, Sunderland, MA).

Darnell J.E. (2011) *RNA: Life's Indispensable Molecule* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

Lane N. (2015) *The Vital Question: Why is Life the Way It Is?* (Profile Books, London).

Pederson T. (2009) The discovery of eukaryotic genome design and its forgotten corollary--the postulate of gene regulation by nuclear RNA. *FASEB Journal* 23: 2019–21.

Quammen D. (2018) *The Tangled Tree: A Radical New History of Life* (Simon Schuster, New York).

Smith J.M. (1978) *The Evolution of Sex* (Cambridge University Press, Cambridge).

Valentine J.W. (1978) The evolution of multicellular plants and animals. *Scientific American* 239: 140–58

Van Kranendonk M.J., Deamer D.W. and Djokic T. (2017) Life springs. *Scientific American* 317: 28–35.