Annual Plant Reviews (2013) **46**, 1–18
doi: 10.1002/9781118472507.ch1

http://onlinelibrary.wiley.com

Chapter 1

INTRODUCTION: MYSTERIES, MOLECULES AND MECHANISMS

John A. Bryant

Biosciences, University of Exeter, Exeter, UK

Abstract: This brief chapter mentions the main structural and functional features of plant nuclei and in doing so, provides a very general introduction to other chapters in the book. It also covers aspects that are not featured elsewhere, especially the replication of nuclear DNA and the import of the replication proteins. Throughout the chapter there is an underlying theme of evolution, relating both to the similarities to and differences from the Archaea and to the possible evolutionary origins of the nucleus.

Keywords: Archaea; DNA replication; evolution; nuclear envelope; nuclear localization signal; origin; protein import

1.1 Darwin and Margulis revisited

In a famous letter sent in July 1879 to Joseph Hooker, the Director of Kew Gardens, Charles Darwin described the origin of the flowering plants as 'an abominable mystery'. Over 130 years later, the mystery seems to be solved, if not in detail, at least in general terms. It is now thought that flowering plants diverged from a lineage of seed ferns (now a totally extinct group) in the late Jurassic or early Cretaceous period (Doyle, 2006, 2008). Based on extensive phylogenetic analysis, the living plant that most resembles the earliest angiosperms (i.e. which is at the base of the angiosperm phylogenetic tree) is *Amborella trichopoda*, a semi-climbing shrub only found in the rain forests of New Caledonia. So, while a solution to that mystery has been found, a further, and perhaps more fundamental mystery remains. It is a

Annual Plant Reviews Volume 46: Plant Nuclear Structure, Genome Architecture and Gene Regulation, First Edition. Edited by David E. Evans, Katja Graumann and John A. Bryant.

mystery that involves not just flowering plants but all eukaryotes and at the beginning of the 21st century it is still not completely solved. That mystery is the origin of the nucleus, the organelle that is the subject of this book. As is evident in subsequent chapters, we have extensive knowledge of its structure and activities. It is a truly beautiful organelle – one that induces in many of us a sense of wonder. However, we are not at all sure where it came from although, as will become clear later in the chapter, a few hypotheses are beginning to emerge as front runners.

On the quest to solve the puzzle, one factor to consider is the origin of eukaryotes. It is now accepted that the two other major membrane-bound organelles, mitochondria and chloroplasts, have evolved from bacterial symbionts that invaded or were engulfed by what we could call proto-eukaryotes (as originally proposed by Margulis, 1971a, b, 1981). This idea has been extensively confirmed by genomic and proteomic studies, which also suggest strongly that those proto-eukaryotic host cells were derived from the Archaea and, in terms of energy metabolism, were using a form of glycolysis[1]. Further, it is clear that following the endosymbiotic events, transfer of genes from both the non-photosynthetic (i.e. mitochondrial) and the photosynthetic (chloroplastic) endosymbionts to the host's genome occurred on a large scale. Indeed, that the process is still going on (Huang *et al.*, 2004, Rousseau-Gueutin *et al.*, 2011, Wang *et al.*, 2012). But where, and in what state were the genomes of those proto-eukaryotic host organisms?

It was thought for several years that relevant information could be obtained by study of amitochondrial eukaryotes, eukaryotes presumed to date back to before the first endosymbiotic event. However, it is now known that these are secondarily amitochondrial, as revealed by the presence of endosymbiont-derived genes in the nucleus and the vestiges of a mitochondrion (e.g. van der Giezen and Tozar, 2005; Minge *et al.*, 2009). So, these cells cannot tell us what the proto-eukaryote looked like. Nevertheless, it is clear that in more recent instances of gene transfer (as mentioned above), the organelle gene has been integrated into a typical eukaryotic nuclear genome located in a typical eukaryotic nucleus. These structures are no hindrance to gene transfer. Further, the use of bioinformatics coupled with comparative cell physiology and biochemistry in attempts to 'root' the eukaryotic phylogenetic tree all lead to the conclusion that most of the approximately 60 differences between eukaryotes and prokaryotes were developed or developing before the first symbiotic event, the acquisition of mitochondria (de Duve, 2007; Margulis *et al.*, 2007; Cavalier-Smith, 2009).

The eukaryotic features possessed by the proto-eukaryotes are thought to have included the possession of a nucleus, nucleoskeleton and cytoskeleton (Margulis *et al.*, 2007; de Duve, 2007; Cavalier-Smith, 2009). Looking at the

[1]But note that in modern Archaea there are several variants of the 'conventional' glycolysis pathway (Sato and Atomi, 2011).

first two of these, these data do not provide any clear clues about where the nucleus came from and there are also questions about the nature of the nucleoskeleton in the earliest eukaryotes. Focussing specifically on this problem, we note that after the first symbiotic event (acquisition of mitochondria), the eukaryotic lineage split into two major branches (Cavalier-Smith, 2002), the unikonts (with one flagellum) that gave rise to, amongst other things, fungi and Metazoa, and the bikonts (with two flagella), one lineage of which became plants by the acquisition of chloroplasts (as mentioned above; see also Keeling, 2010).

Turning now to look at extant lineages, as is shown in Chapters 2 and 4, part of the nucleoskeleton in animals is the prominent lamina, consisting mainly of proteins known as lamins. However, plants lack lamins but do possess a lamina-like structure that has been called the 'plamina' (Fiserova *et al*., 2009), consisting of plant-specific proteins that are functional analogues of lamins. Finally, in fungi, at least as represented by yeasts, the nucleoskeleton does not have any form of lamina. So, based on the origins of these groups, it is suggested that the proto-eukaryotic nucleoskeleton lacked a lamina and that this has developed subsequent to the uni-kont/bikont split. This gives us a little more information on the early nucleus, but the question of its origin remains.

At this point further specific discussion of the origin of nucleus is deferred to the end of the chapter, although it will appear more indirectly from time to time in the next three sections. Attention is now turned to the genome itself. Particular focus will be placed on the general structure of the genome, on its replication and on the implications for the latter process of enclosing the genome in an organelle.

1.2 Nuclei – general features

In plant cells that are not extensively vacuolated, the nucleus is the largest and usually the most obvious organelle. Even in mature cells with large vacuoles, the nucleus is usually clearly visible within the cytoplasm. It is the organelle that contains the bulk of the cell's DNA, the nuclear genome. Indeed, chromatin (Chapters 5 and 6), consisting mainly of a complex of DNA and proteins, is usually the most obvious component of the nucleus. The chromatin is attached *via* scaffold- or matrix-associated regions (SARs/MARs) to the nuclear matrix/scaffold/nucleoskeleton (Chapters 4 to 6). Within chromatin, the highly repeated genes encoding the major ribosomal RNAs (rRNAs) are looped out in structures called nucleoli. The fibrillar centres of the nucleoli are the sites of transcription of these genes and the transcripts are processed in the outer regions of the nucleoli.

The nucleus is bounded by the nuclear envelope or NE (Chapters 2 and 3), which consists, in effect, of three membranous components (shown diagrammatically in the cartoon in Figure 1.1). Firstly, the outer envelope is connected

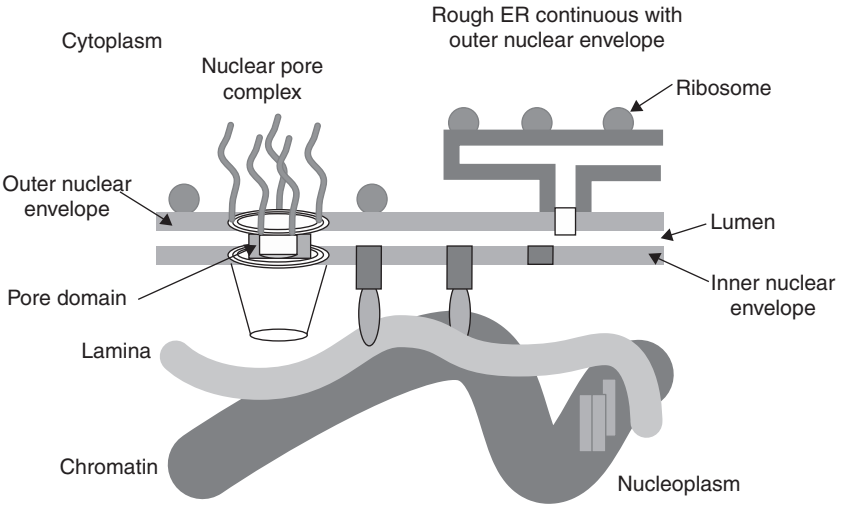

Figure 1.1 Diagrammatic cartoon of the nuclear envelope and nuclear pore complex. (From Evans *et al.*, 2004.) Reproduced by permission of the Society for Experimental Biology.

to the ER and the lipids and proteins of the outer NE are similar to those of the rough ER. Further, as with the rough ER, ribosomes are often present on the outer NE. So, the outer NE may be a site of protein synthesis and is certainly a part of the cell's endomembrane system. Secondly, there is the inner NE separated from the outer NE by the lumen, which is about 30 nm across. The inner surface of the inner NE is closely associated with the nuclear lamina, a structure consisting of filamentous proteins and which forms the main component of the nuclear matrix or nucleo-skeleton. Thirdly there is the pore membrane, which links the inner and outer NEs and forms part of the nuclear pore complex or NPC (Chapters 2, 4 and 8).

The containment of chromatin within its own membrane-bound organelle has major implications for the life of the cell. Amongst other things, it permits precise and complex regulation of gene activity and DNA replication 'protected' from more general aspects of cellular metabolism. However, it also imposes constraints. The nucleus does not contain protein-synthesizing machinery, even though proteins may be made on the surface of the outer NE. All the enzymes, together with structural and regulatory proteins necessary for the activities and components of the nucleus, over 1000 proteins in all (Nuclear Protein Database: http://npd.hgu.mrc.ac.uk/), must be able to get in from the outside. At the same time, several thousand more proteins, those that are not involved in the life of the nucleus, are kept out. There are also proteins that shuttle between the nucleus and the cytosol. Finally, all the different RNAs that function in the cytosol must leave the nucleus (in the form of nucleoprotein complexes). The NPCs have a major role in the

control of entry into and exit from the nucleus (Chapters 4 and 8), along with specific signalling and transport mechanisms. This provides one more level of regulation of chromatin-associated biochemical activity.

1.3 The plant nuclear genome

1.3.1 General features

A recent review by Heslop-Harrison and Schwarzacher (2011) gives a wealth of information about plant nuclear genomes, while Chapters 5 and 6 in this volume deal with specific aspects of chromatin organization. In this chapter, the focus is on those features related to genome evolution and replication. Higher plant genomes vary enormously in overall size, ranging over three orders of magnitude. There is some correlation between genome size and nuclear size so that, in general, plants with large genomes have larger nuclei than plants with small genomes. Some of the differences in genome size have arisen by duplication of individual genes and of whole genomes (polyploidy). Within individual genomes, much of the DNA does not code for proteins or RNA. Comparison between closely related species (see, e.g. Bryant and Hughes, 2011) that have differing amounts of nuclear DNA show that most of the variation can be accounted for by repeated DNA sequences. Some of the variation is in the number of copies of repeated genes, such as those coding for rRNA but most of it is accounted for by variations in non-coding DNA. This includes highly repeated 'satellite' DNA of around 180 base pairs per repeat (Sibson *et al.*, 1991; Round *et al.*, 1997), simple sequence repeats and retrotransposons. 'Satellite' DNA sequences are concentrated at the centromeres where they appear to be essential for centromere function (Nagaki *et al.*, 2003 and Chapter 6, this volume). Retrotransposons or retro-elements include LINES (long interspersed sequences) and various highly repetitive sequences of different sizes and copy number. Taken together these retro-elements can make a very large proportion of the genome (for details, see Bryant and Hughes, 2011; Heslop-Harrison and Schwarzacher, 2011). There are also 'fossil' coding sequences or pseudogenes, some of which seem to have been derived from other species. All these types of sequence are present to some extent in all eukaryotes and have possibly been features of chromatin since the first appearance of the Eukarya. However, over the period in which higher plants have been evolving, their nuclear genomes have been and continue to be amongst the most dynamic.

As in all eukaryotes, the DNA itself exists as linear molecules, one long double helix per chromosome. In the chromosomes, the DNA is complexed with proteins, mainly the basic proteins known as histones, to form chromatin, as described in more detail in Chapters 5 and 6. Some chromatin (heterochromatin) remains condensed and therefore clearly visible throughout the cell cycle. As described in much more detail in Chapter 6, much of the

heterochromatin is located at the centromeres (and thus involves 'satellite' DNA, mentioned above) and at the telomeres (ends of chromosomes). By contrast with heterochromatin, the majority of the chromatin, known as euchromatin, decondenses as mitosis is completed. The significance of these two patterns of behaviour and of the distribution heterochromatin is discussed in Chapter 6.

The linear structure of eukaryotic DNA molecules has caused some to question the origins of eukaryotic genomes (see Section 1.5). The consensus remains that the original proto-eukaryotic host cell was derived from the archaean lineage and yet amongst extant members of the Archaea, we do not yet know of any that have linear chromosomes. Nevertheless, DNA in many Archaea is complexed with histone-like proteins to form features that resemble the nucleosomes of eukaryotic chromatin (Pereira and Reeve, 1998; Sandman *et al.*, 2001), albeit that these nucleosomes contain only 80 base pairs of DNA wrapped round four, not eight histone molecules (see Chapter 5 for a detailed description of eukaryotic chromatin)[2]. Further, several eubacteria are known that have linear chromosomes, and some species are able to maintain both linear and circular DNA molecules (Volff and Altenbuchner, 2000; Lin and Moret, 2011), a situation that has also been described for mitochondria in some lower and higher plants and fungi (Borza *et al.*, 2009; Lo *et al.*, 2011). The absence of linear DNA molecules does not therefore rule out Archaea as being progenitors of or a sister group to the proto-eukaryotes.

1.3.2 Replication of the nuclear genome

The general features of eukaryotic genomes raise several problems for DNA replication. These have been discussed more fully in an earlier publication (Bryant, 2010) but need to be mentioned briefly here. The first is that the complex of DNA and protein in chromatin (which is of course common to all eukaryotes) means that copying the DNA is slower than in prokaryotes. Taking this together with the length of eukaryotic DNA molecules, especially in some plant genomes, has led to the evolution of multiple replication origins (points at which replication may start) along the axis of each DNA molecule. The nature of these replication origins in relation to DNA structure has been a matter for debate for many years (see e.g. Hernández *et al.*, 1988; Van't Hof and Lamm, 1992; Bryant and Francis, 2008; Bryant, 2010; Lee *et al.*, 2010; see also Berbenetz *et al.*, 2010) and it is still not clear whether or not specific sequences are involved. What is clear, however, is that origins are AT-rich and are therefore more prone to transient, localized short-range strand separation

[2]It must also be noted that many, if not all, Archaea possess a different type of DNA-binding protein, known as Alba, which is also able to generate a form of chromatin in which the DNA is inside the protein (Tanaka *et al.*, 2012) – i.e. very different from the nucleosome structure.

known as 'breathing'. It is also clear that the timing and order in which the origins are active ensures that the DNA is replicated within an S-phase that is completed within a few hours. There are again links with the Archaea in that several species have more than one replication origin (usually two or three: Lundgren and Bernander, 2005; Robinson and Bell, 2005), which seem to be attached to specific locations at the cell's periphery (Gristwood *et al.*, 2012). Like the replication origins of plants (and other eukaryotes), archaean origins are AT-rich and, in this group, specific sequence is important for correct function (Majernik and Chong, 2008).

Finally, the uni-directional (5′ to 3′) nature of DNA replication, coupled with the inability of DNA polymerases to initiate replication without a primer (see Bryant, 2010) causes problems at the ends of molecules. This has led to the development during evolution of specialized structures called telomeres at the ends of chromosomal DNA molecules, with an associated enzyme, telomerase[3]. As discussed in Chapter 7, the ends are protected from degradation and from being inappropriately targeted by the DNA repair machinery (see Chapter 5) because of the telomere/telomerase combination.

The enzymes and other proteins which carry out replication of nuclear DNA in plants have been described in some detail in two recent papers (Schultz *et al.*, 2007; Bryant, 2010). Here the focus is on a selection of those aspects that provide clues about evolution. Looking first at pre-replication events (see e.g. Aves, 2009; Bryant and Aves, 2011), it is clear that in plants, as in other eukaryotes, replication origins are bound and therefore 'marked' by a complex of six proteins, the Origin-Recognition Complex (ORC) (Collinge, *et al.*, 2004; Mori *et al.*, 2005; Shultz *et al.*, 2007; Bryant 2010). A protein known as CDC6, along with CDT1, then facilitates the loading of the CMG complex consisting of the GINS hetero-tetramer, MCMs 2-7 (the helicase that separates the two strands of DNA at the replication fork) and the protein known as CDC45, which will later facilitate loading of the initiating DNA polymerase. Looking at the Archaea, it is clear that both recognition of replication origins and the first stage in their activation are carried out by a single protein that is both similar to and fulfils the functions of the ORC and CDC6 (reviewed by Bryant and Aves, 2011). The function of the GINS complex is carried out either by a homo-tetramer or tetramer consisting of two different homo-dimers (Yoshimochi *et al.*, 2008). Sequence similarity to eukaryotic GINS proteins is limited but the proteins interact to form a complex of similar architecture to the eukaryotic complex. With respect to MCMs, most Archaea have just one, which forms a homohexamer as compared with the eukaryotic heterohexamer. Hints of multiple MCMs are seen in *Thermococcus kodakarensis* (Pan *et al.*, 2011), which has three. Both MCMs 2 and 3 can form homohexamers but only MCM3 is essential for DNA replication.

[3]See also the discussion on linear DNA molecules in mitochondria (e.g. Valach *et al.*, 2011).

Table 1.1 Comparison between Archaea and Eukarya in respect of proteins involved in the initiation of DNA replication

Function	Proteins in Archaea	Proteins in Eukarya
Origin recognition	Single Origin-Recognition Protein/CDC6	Complex of six different Origin-Recognition Proteins
Loading of pre-replicative complex onto origin	The same ORP/CDC6	CDC6 and CDT1
Helicase accessory proteins (GINS complex)	Either a homo-tetramer or two homo-dimers make up the complex	A hetero-tetramer
Replicative DNA helicase	A homo-hexamer of one type of MCM protein	A hetero-hexamer of six different MCM proteins
Polymerase loading factor	Member of the Rec-J protein family	CDC45

Finally, CDC45 is represented in Archaea by a member of the Rec J protein family, many of which have nuclease activity. What emerges from this is that heteromeric protein complexes involved in plant DNA replication are represented by homo-polymers in Archaea, or, as expressed by Bryant and Aves (2011), '*The proteins themselves represent the essential core, compared with the eukaryotic plenitude*.' Thus, individual proteins in Archaea are now represented by multiple versions in plants and other eukaryotes. Using the MCMs as an example, a detailed bioinformatic study (Lui *et al.*, 2009) indicates that the different MCMs have arisen by a series of gene duplication events and, by analogy, we can envisage a similar series of gene duplications giving rise to six ORC proteins plus CDC6 (Table 1.1). Further, the analysis carried out by Lui *et al.* (2009) suggests that these gene duplications are very ancient and were already present in the last common ancestor of all eukaryotes. Further, without going into detail here, similar conclusions have been reached from a study of transcription factors (Bell *et al.*, 2001).

After preparation of the DNA for replication, synthesis itself is initiated by DNA primase and the primers are then extended by DNA polymerase-α. This is the only DNA polymerase that can work with primase and, indeed, the two enymes form a complex, as described in previous publications (Bryant *et al.*, 1992; Bryant *et al.*, 2001; Bryant, 2010). Replication within each replicon (i.e. the tract of DNA replicated from one origin) is then completed by DNA polymerase-δ and DNA polymerase-ε as described in much greater detail in Bryant (2010). This also involves a nuclease to remove the primers (either a 'flap nuclease' such as Fen 1 or ribonuclease-H) and DNA ligase to join pieces of newly synthesized DNA (Bryant, 2010). Focussing specifically on the three replicative polymerases, these are all members of the B family of DNA polymerases and perhaps by now it is not surprising to learn that the replicative DNA polymerase of Archaea is also a member of this family.

For a much fuller description of DNA replication, of the enzymes and other proteins involved in replication and of the regulation of replication, the reader is referred to the comprehensive papers mentioned above. At this point in this chapter, it is time to move on to consider the implications for DNA replication and gene function of enclosing the genetic material in a sub-cellular compartment.

1.4 DNA inside, ribosomes outside

The enclosure of the genetic material inside a membrane-bound organelle has many advantages. It allows localized control mechanisms to operate inside the organelle, with that control being partly exerted by the inside-outside segregation itself. It frees the cell's gene expression and gene regulation machinery from the possibility of unwanted cross-reactivity with other aspects of metabolism. The division of labour between the sub-cellular organelles, including the nucleus, is held to have made multi-cellularity possible. However, this compartmentation (and here, obviously, focus is on just one of the organelles, the nucleus) also raises clear difficulties and especially that the enzymes and other proteins required for chromatin structure, DNA replication, gene expression and so on are made on ribosomes located in the cytosol. Further, once they have been synthesized by transcription from the relevant genes, RNA molecules must be able to fulfil their various functions in protein synthesis. These RNA molecules do not just diffuse out of the nucleus but are transported out in the form of nucleo-protein complexes. The proteins involved in this are thus able to both enter and leave the nucleus (and this is true of some proteins involved in DNA replication, in signalling and in control of gene expression).

The need to enable specific proteins to enter and leave the nucleus has led to the evolution of a sorting mechanism in which proteins destined for the inside of the nucleus contain within their amino acid sequence a label, the nuclear localization signal or NLS. In proteins that are required to come out again there is also a nuclear export signal or NES. In nuclei as they are now constituted, the gatekeeper between the cytosol and the nucleus is the nuclear pore complex (NPC) with import machinery based on importins and a RAN-GTPase (as described in Chapters 3 and 4). We have no clear picture of what a 'proto-nucleus' might have looked like in respect of NPCs. However, as indicated earlier, it is widely held that the engulfment/symbiosis that led to the development of mitochondria involved a host cell with many of the special eukaryotic features, including some form of nucleo-skeleton. This is somewhat confirmed by results from a genomic analysis of 19 (out of a total of about 30) nuclear pore complex proteins (Nups) across 60 different eukaryotes from a wide range of groups. The analysis indicates that all the major sub-complexes of proteins in the NPC are traceable as far back as the Last Eukaryotic Common Ancestor or LECA (Neumann *et al.*, 2010).

Clues to the development of these, perhaps in a lineage leading to the LECA, comes from a study of protein architecture and folding in Nups, leading to the suggestion that extensive gene duplication and motif duplication within genes has led to the development of the range of Nups from a small number of precursor proteins (Devos *et al.*, 2006).

At this point, attention is focussed purely on proteins imported into the nucleus (the same principles apply to those that must later leave the nucleus). In a wide-ranging study of proteins that are required to enter nuclei (Dingwall and Laskey, 1991), several different sequences were identified that were involved in nuclear uptake. Three of these are found in plants, namely:

- the SV-40 virus-type monopartite NLS, which consists of a run of five basic amino acids (named for the virus in which it was first found);
- the yeast Matα2 type of NLS in which a run of four basic amino acids is interrupted by three hydrophobic amino acids (named after a yeast mating type protein);
- the bi-partite NLS, consisting of two short regions containing basic amino acids (*i.e.* arginine, histidine or lysine) separated by a spacer of up to 10 amino acids (Figure 1.2).

The last of these is by far the commonest in plants. In different proteins, the NLS occurs at different places within the amino acid sequence although many of them are at or near the N-terminus. Presumably what matters is that, in the folded protein, the NLS is 'visible' to the import machinery. The efficacy of a bi-partite NLS is illustrated in Figure 1.3.

Two more things need to be said about nuclear import. Firstly, there are some nuclear proteins that appear to lack completely an NLS (Stiedl *et al.*, 2004) and are thus 'piggy-backed' into the nucleus on an NLS-containing protein (Stiedl *et al.*, 2004; see also Galstyan *et al.*, 2012).

Secondly, there is an increasing list of proteins that appear to have roles in the cytosol and in the nucleus, for example in both glycolysis and in DNA replication or repair. Such dual-function proteins are known as 'moonlighting' proteins. A typical example is phosphoglycerate kinase (PGK), with a primary function in glycolysis but which also enters the nucleus (Anderson *et al.*, 2004; Brice *et al.*, 2004). *In vitro* it stimulates the activity of DNA polymerase-α on poorly primed templates (Burton *et al.*, 1997; Bryant *et al.*, 2000; Bryant, 2008) and it has been suggested that it acts as an accessory protein for the polymerase (Bryant *et al.*, 2000; Bryant, 2008). Concomitant with a

MAT**KRSVGTLKEADLKGKRV**FVR

Figure 1.2 A typical plant bi-partite nuclear localization, situated near the N-terminus. The NLS itself is shown in bold type with the basic residues that are part of the signal sequence underlined (note that in this example there are also two basic amino acids in the intervening sequence between the two halves of the signal).

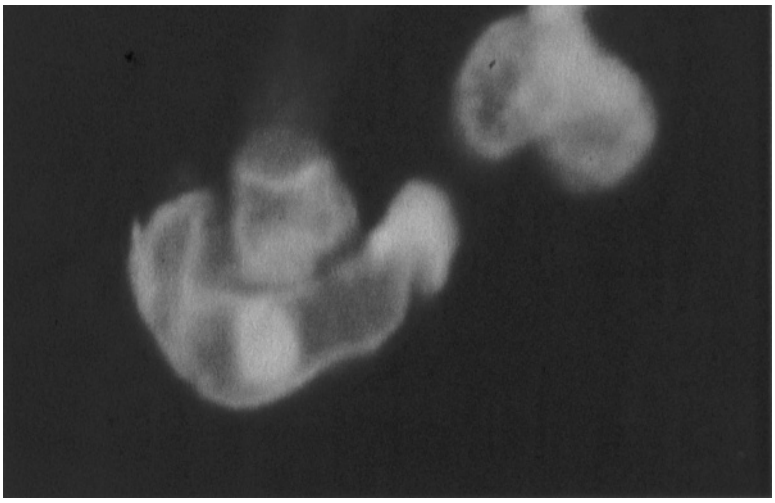

Figure 1.3 A transcriptional fusion was made between the coding sequences for Green Fluorescent Protein (GFP) and the moonlighting protein PGK which has an NLS. The hybrid protein was transported into the nucleus of tobacco cells, as is shown by the bright fluorescence (from Brice *et al.*, 2004).

role in the nucleus, PGK possesses an NLS located at the N-terminus (Brice *et al.*, 2004 and Figure 1.3) and an important question relating to this and, indeed, to all proteins with dual cytosolic/nuclear locations is how they are actually partitioned between the two sites. For PGK, submitting the sequence of the protein to the PSORTII program suggests that *in vivo*, 12% of the protein population will be transported into the nucleus (Bryant, 2008; see also Weis, 2003; Terry *et al.*, 2007; Meier and Somers, 2011). Another question concerns the evolution of moonlighting proteins. Do their dual roles represent an earlier stage in evolution in which a given protein might carry out more than one function? If so, it is interesting that the apparently minor role (as with PGK in the nucleus) has led to the acquisition of an NLS.

1.5 Concluding comments on the evolution of the nucleus

At the beginning of the chapter, reference was made to the mystery surrounding the evolutionary origin of the nucleus. The picture we have of the proto-eukaryote that became the host for the mitochondrial symbiont is certainly of a cell with a nucleus. Further, that picture is becoming clearer and clearer as a result of studies in various aspects of bio-informatics, some of which have been mentioned in this chapter. In analysis of genes and proteins involved in the function of the nucleus, for example, in DNA replication, the Archaea continue to dominate: the archaeal origin of many eukaryotic

nuclear enzymes seems very clear (reviewed by Bryant and Aves, 2011). The same picture emerges from analyses based on histones (Pereira and Reeve, 1998), on ribosomal RNA (Xie *et al.*, 2012) and from wider-ranging genomic analyses (Saruhashi *et al.*, 2008). But that still does not tell us how an archaeal cell became a 'proto-eukaryote' with a nucleus. Currently there are three main theories.

Firstly, it has been suggested that an archaeal cell was invaded by an enveloped virus with a linear genome (Bell, 2001, 2006; Villareal and Witzany, 2010). The authors cite the similarity between replicative DNA polymerases encoded by certain DNA viruses and the B-family of replicative DNA polymerases in Archaea and Eukarya (see above). Even so, this may seem far fetched because it is generally held that viruses are a later addition to the rich variety of biology. However, a recent paper in which the ancestry of viruses was investigated through analysis of protein architecture and folding suggests that 'giant' viruses appeared earlier than previously thought (Nasir *et al.*, 2012). Indeed, the latter authors suggest rooting a virus clade alongside the base of the Eukarya. However, we have already seen that linear genomes may arise from circular genomes, so it is not necessary to invoke a virus in order to provide a linear genome. Further, there is the question of the integration of the viral genome into that of the host. Although, based on examples of extant viruses, this could certainly happen, bio-informatic analysis of eukaryotic and archaeal genes (e.g. Lui *et al.*, 2009) gives no indication of a viral ancestry for the proteins involved in the initiation of DNA replication.

Secondly it has been proposed that the nucleus arose through a symbiosis between an archaeal cell and a eubacterial cell, with the eubacterial cell as host and the archaeal cell as invading symbiont (Gupta and Golding, 1996; Ohyanagi *et al.*, 2008). This is clearly different from the view that that the progenitor of the Eukarya was an actual archaean but it does allow for the presence of archaean features. According to this view, the nucleus represents an archaeal symbiont that either took over the genetic function of the host or that contributed to a larger genome derived from both cells. The genomic and rRNA analyses mentioned above are consistent with a dominant archaeal genome in such a symbiosis although analysis of ribosome export factors suggests the possibility of a significant contribution from a eubacterial genome (Ohyanagi *et al.*, 2008).

Finally, some authors propose that there is no need to invoke any sort of invasion or symbiosis to explain the origin of the nucleus. This was firmly stated by Martin (1999); more recent evidence provides some support for this view. The discovery that replication origins in Archaea appear to be attached at the cell's periphery (Gristwood *et al.*, 2012) is consistent with the idea that the archaeal chromatin first became attached to the cell membrane and then was enveloped by invaginations of the membrane, possibly in connection with phagocytosis (Cavalier-Smith, 2009; see also Cavalier-Smith, 2010). Development of a nuclear skeleton is then presumed to have

followed. An extension of this view is that surrounding the genetic material with a membrane was a protective measure, evolved in response to the presence of reactive oxygen species that became abundant after the evolution of photosynthesis and the 'great oxidation event.' (Gross and Bhattacharya, 2010). Construction of the nucleus from within an archaeal cell is actually the simplest of the three hypotheses and fits better with the majority of the genomic analyses. It is the hypothesis that is favoured by the present author. Hopefully, time will tell.

References

Anderson, L.E., Bryant, J.A. and Carol, A.A. (2004) Both chloroplastic and cytosolic phosphoglycerate kinase isozymes are present in the pea leaf nucleus. *Protoplasma* **223**,103–110.

Aves, S.J. (2009) DNA replication initiation. *Methods in Molecular Biology* **521**, 3–17.

Bell, P.J.L. (2001) Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *Journal of Molecular Evolution* **53**, 251–256.

Bell, P.J.L. (2006) Sex and the eukaryotic cell cycle is consistent with a viral ancestry for the eukaryotic nucleus. *Journal of Theoretical Biology* **243**, 54–63.

Bell, S.D, Magill, C.P and Jackson, S.P (2001) Basal and regulated transcription in Archaea. *Biochemical Society Transactions* **29**, 392–395.

Berbenetz, N.M., Nislow, C. and Brown, G.W. (2010) Diversity of eukaryotic DNA replication origins revealed by genome-wide analysis of chromatin structure. *PLOS Genetics* **6**. doi: 10.1371/journal.pgen.1001092.

Borza, T., Redmond, E.K., Laflamme, M. and Lee, R.W. (2009) Mitochondrial DNA in the Oogamochlamys clade (Chlorophyceae): high GC content and unique genome architecture for green algae. *Journal of Phycology* **45**, 1323–1334.

Brice, D.C., Bryant, J.A., Dambrauskas, D. *et al.* (2004) Cloning and expression of cytosolic phosphoglycerate kinase from pea (*Pisum sativum* L.). *Journal of Experimental Botany* **55**, 955–956.

Bryant, J.A. (2008) Copying the template: with a little help from my friends? In: Bryant, J.A. and Francis, D. (eds) *The Eukaryotic Cell Cycle*. Taylor & Francis, Abingdon, pp. 71–80.

Bryant, J.A. (2010) Replication of Nuclear DNA. *Progress in Botany* **71**, 25–60.

Bryant, J.A. and Aves, S.J. (2011) Initiation of DNA replication: functional and evolutionary aspects. *Annals of Botany* **107**, 1119–1126.

Bryant, J.A., Brice, D.C., Fitchett, P.N. and Anderson, L.E. (2000) Novel DNA-binding protein associated with DNA polymerase-α in pea stimulates polymerase activity on infrequently primed templates. *Journal of Experimental Botany* **51**, 1945–1947.

Bryant, J.A., Fitchett, P.N., Hughes, S.G. and Sibson, D.R. (1992) DNA polymerase-α in pea is part of a large multiprotein complex. *Journal of Experimental Botany* **43**, 31–40.

Bryant, J.A. and Francis, D. (2008) Initiation of DNA replication. In: Bryant, J.A. and Francis, D. (eds) *The Eukaryotic Cell Cycle*. Taylor & Francis, Abingdon, pp 29–44.

Bryant, J.A. and Hughes, S.G. (2011) Vicia. In: Cole, C. (ed.) *Wild Crop Relatives: Genomic and Breeding resources. Legume Crops and Forages*, Springer, Berlin, pp 273–289.

Bryant, J.A., Moore, K.A. and Aves, S.J. (2001) Origins and complexes: the initiation of DNA replication. *Journal of Experimental Botany* **52**,193–202.

Burton, S.K., Bryant, J.A. and Van't Hof, J (1997) Novel DNA-binding characteristics of a protein associated with DNA polymerase-α in pea. *Plant Journal* **12**, 357–365.

Cavalier-Smith, T. (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of protozoa. *International Journal of Systematic and Evolutionary Microbiology* **52**, 297–354.

Cavalier-Smith, T. (2009) Predation and eukaryote cell origins: A coevolutionary perspective. *International Journal of Biochemistry and Cell biology* **41**, 307–322.

Cavalier-Smith, T. (2010) Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. *Biology Direct* **5**, Article Number: 7 doi: 10.1186/1745-6150-5-7.

Collinge, M.A., Spillane, C., Kohler, C. *et al.* (2004) Genetic interaction of an origin recognition complex subunit and the Polycomb group gene MEDEA during seed development. *Plant Cell* **16**, 1035–1046.

de Duve, C. (2007) The origin of eukaryotes – a reappraisal. *Nature Reviews Genetics* **8**, 395–403.

Devos, D., Dokudovskaya, S., Williams, R. *et al.* (2006). Simple fold composition and modular architecture of the nuclear pore complex. *Proceedings of the National Academy of Sciences, USA.* **103**, 2172–2177.

Dingwall, C. and Laskey, R.A. (1991) Nuclear targeting sequences – a consensus. *Trends in Biochemical Sciences* **16**, 478–481.

Doyle, J.A. (2006) Seed ferns and the origin of angiosperms. *Journal of the Torrey Botanical Society* **133**, 169–209.

Doyle, J.A. (2008) Integrating molecular phylogenetic and paleobotanical evidence on origin of the flower. *International Journal of Plant Sciences* **169**, 816–843.

Evans, D.E., Bryant, J.A. and Hutchison, C.J. (2004) The nuclear envelope: a comparative overview. In Evans, D.E., Hutchison, C.J. and Bryant, J.A. (eds), *The Nuclear Envelope*, BIOS, Oxford, pp. 1–8.

Fiserova, J., Kiseleva, E. and Goldberg, M.W. (2009) Nuclear envelope and nuclear pore complex structure and organization in tobacco BY-2 cells. *Plant Journal* **59**, 243–255.

Galstyan, A., Bou-Torrent, J., Roig-Villanova, I. and Martinez-Garcia, J.F. (2012) A dual mechanism ontrols nuclear localization in the atypical basic-helix-loop-helix protein PAR1 of *Arabidopsis thaliana*. *Molecular Plant* **5**, 669–677.

Gristwood, T., Duggin, I.G., Wagner, M. *et al.* (2012) The sub-cellular localization of *Sulfolobus* DNA replication. *Nucleic Acids Research* **40**, 5487–5496.

Gross, J. and Bhattacharya, D. (2010) Uniting sex and eukaryote origins in an emerging oxygenic world. *Biology Direct* **5** Article Number: 53, doi: 10.1186/1745-6150-5-53.

Gupta, R.S. and Golding, G.B. (1996) The origin of the eukaryotic cell. *Trends in Biochemical Sciences* **21**, 166–171.

Hernández, P., Bjerknes, C.A., Lamm, S.S. and Van't Hof, J. (1988) Proximity of an *ARS* consensus sequence to a replication origin of pea (*Pisum sativum*). *Plant Molecular Biology* **10**, 413–422.

Heslop-Harrison, J.S. and Schwarzacher, T. (2011) Organisation of the plant genome in chromosomes. *Plant Journal* **66**, 18–33.

Huang, C.Y., Ayliffe, M.A. and Timmis, J.N. (2004) Simple and complex nuclear loci created by newly transferred chloroplast DNA in tobacco. *Proceedings of the National Academy of Sciences,USA* **101**, 9710–9715.

Keeling, P.J. (2010) The endosymbiotic origin, diversification and fate of plastids. *Philosophical Transactions of the Royal Society, Biological Sciences* **365**, 729–748.

Lee, T.-L., Pascuzzi, P.E., Settlage, S.B. *et al.* (2010) Arabidopsis thaliana chromosome 4 replicates in two phases that correlate with chromatin state. *PLOS Genetics* **6** Article Number: e1000982 doi: 10.1371/journal.pgen.1000982.

Lin, Y. and Moret, B.M.E. (2011) A new genomic evolutionary model for rearrangements, duplications, and losses that applies across eukaryotes and prokaryotes. *Journal of Computational Biology* **18**, 1055–1064.

Lo, Y.-S., Hsiao, L.-J., Cheng, N. *et al.* (2011) Characterization of the structure and DNA complexity of mung bean mitochondrial nucleoids. *Molecules and Cells* **31**, 217–224.

Lui, Y., Richards,T.A. and Aves, S.J. (2009) Ancient diversification of eukaryotic MCM DNA replication proteins. *BMC Evolutionary Biology* **9** Article Number: 60 doi: 10.1186/1471-2148-9-60.

Lundgren, M. and Bernander, R. (2005) Archaeal cell cycle progress. *Current Opinion in Microbiology* **8**, 662–668.

Majernik, A.I. and Chong, J.P.J. (2008) A conserved mechanism for replication origin recognition and binding in Archaea. *Biochemical Journal* **409**, 511–518.

Margulis, L. (1971a) Symbiosis and evolution. *Scientific American* **225**, 48–57.

Margulis, L. (1971b) Origin of plant and animal cells. *American Scientist* **59**, 230–235.

Margulis, L. (1981) *Symbiosis in Cell Evolution: Life and Its Environment on the Early Earth*. W.H. Freeman, New York.

Margulis, L., Chapman, M.J. and Dolan, M.F. (2007) Eukaryosis: phagocytosis and hydrogenases. *Symbiosis* **43**, 161–163.

Martin, W. (1999) A briefly argued case that mitochondria and plastids are descendants of endosymbionts, but that the nuclear compartment is not. *Proceedings of the Royal Society, Series B* **266**, 1387–1395.

Meier, I. and Somers, D.E. (2011) Regulation of nucleocytoplasmic trafficking in plants. *Current Opinion in Plant Biology* **14**, 538–546.

Minge, M.A., Silberman, J.D., Orr, R.J.S. *et al.* (2009) Evolutionary position of breviate amoebae and the primary eukaryote divergence. *Proceedings of the Royal Society B-Biological Sciences* **276**, 597–604.

Mori, Y., Yamamoto, T., Sakaguchi, N. *et al.* (2005) Characterization of the origin-recognition complex (ORC) from a higher plant, rice (*Oryza sativa* L.) *Gene* **353**, 23–30.

Nagaki, K., Talbert, P.B., Zhong, C.X. *et al.* (2003) Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. *Genetics* **163**, 1221–1225.

Nasir, A., Kim, K.M. and Caetano-Anolles, G. (2012) Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with super-kingdoms Archaea, Bacteria and Eukarya. *BMC Evolutionary Biology* **12**, 156. Epub ahead of print PMID:22920653.

Neumann, N., Lundin, D. and Poole, A.M. (2010) Comparative genomic evidence for a complete nuclear pore complex in the Last Eukaryotic Common Ancestor. *PLOS ONE* **5**, Article Number: e13241 doi: 10.1371/journal.pone.0013241.

Ohyanagi, H., Ikeo, K. and Gojobori, T. (2008) The origin of nucleus: Rebuild from the prokaryotic ancestors of ribosome export factors. *Gene* **423**, 149–152.

Pan, M., Santangelo, T.J., Li, Z. *et al.* (2011) *Thermococcus kodakarensis* encodes three MCM homologs but only one is essential. *Nucleic Acids Research* **39**, 9671–9680.

Pereira, S.L. and Reeve, J.N. (1998) Histones and nucleosomes in Archaea and Eukarya: a comparative analysis. *Extremophiles* **2**, 141–148.

Robinson, N.P. and Bell, S.D. (2005) Origins of DNA replication in the three domains of life. *FEBS Journal* **272**, 3757–3766.

Round, E.K., Flowers, S.K. and Richards, E.J. (1997) *Arabidopsis thaliana* centromere regions: Genetic map positions and repetitive DNA structure. *Genome Research* **7**, 1045–1053.

Rousseau-Gueutin, M., Ayliffe, M.A. and Timis, J.N. (2011) Conservation of plastid sequences in the plant nuclear genome for millions of years facilitates endosymbiotic evolution. *Plant Physiology*, **157**, 2181–2193.

Sandman, K., Soares, D. and Reeve, J.N. (2001) Molecular components of the archaeal nucleosome. *Biochimie* **83**, 277–281.

Saruhashi, S., Hamada, K., Miyata, D. *et al.* (2008) Comprehensive analysis of the origin of eukaryotic genomes. *Genes and Genetic Systems* **83**, 285–291.

Sato, T. and Atomi, H. (2011) Novel metabolic pathways in Archaea. *Current Opinion in Microbiology* **14**, 307–314.

Shultz, R.W., Tatineni, V.M., Hanley-Bowdoin, L. and Thompson, W.F. (2007) Genome-wide analysis of the core DNA replication machinery in the higher plants Arabidopsis and rice. *Plant Physiology* **144**, 1697–1714.

Sibson, D.R., Hughes, S.G., Bryant, J.A. and Fitchett, P.N. (1991) Sequence organization of simple, highly repetitive DNA elements in *Brassica* species. *Journal of Experimental Botany* **42**, 243–249.

Stiedl, S., Tuncher, A., Goda, H. *et al.* (2004) A single subunit of a heterotrimeric CCAAT-binding complex carries a nuclear localization signal: Piggy back transport of the pre-assembled complex to the nucleus. *Journal of Molecular Biology* **342**, 515–524.

Tanaka, T., Padavattan, S. and Kumarevel, T. (2012) Crystal structure of Archaeal chromatin protein Alba2-double-stranded DNA complex from *Aeropyrum pernix* K1. *Journal of Biological Bhemistry* **287**, 10394–10402.

Terry, L.J., Shows, E.B. and Wente, S.R. (2007) Crossing the nuclear envelope: hierarchical regulation of nucleocytoplasmic transport. *Science* **318**, 1412–1416.

Valach, M., Farkas, Z., Fricova, D. *et al.* (2011) Evolution of linear chromosomes and multipartite genomes in yeast mitochondria. *Nucleic Acids Research* **39**, 4202–4219.

van der Giezen, M. and Tozar, J. (2005) Degenerate mitochondria. *EMBO Reports* **6**, 525–530.

Van't Hof, J. and Lamm, S.S. (1992) Site of initiation of replication of the ribosomal RNA genes of pea (*Pisum sativum*) detected by 2-dimensional gel electrophoresis. *Plant Molecular Biology* **20**, 377–382.

Villarreal, L.P. and Witzany, G. (2010) Viruses are essential agents within the roots and stem of the tree of life. *Journal of Theoretical Biology* **262**, 698–710.

Volff, J.N. and Altenbuchner, J. (2000) A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution. *FEMS Microbiology Letters* **186**, 143–150.

Wang, D., Lloyd, A.H. and Timmis, J.N. (2012) Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants. *Proceedings of the National Academy of Sciences, USA* **109**, 2444–2448.

Weis, K. (2003) Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. *Cell* **112**, 441–451.

Xie, Q., Wang, Y.H., Lin, J.Z. *et al.* (2012) Potential key bases of ribosomal RNA to kingdom-specific spectra of antibiotic susceptibility and the possible archaeal origin of eukaryotes. *PLOS ONE* **7** Article Number: e29468 doi: 10.1371/journal.pone.0029468.

Yoshimochi, T., Fujikane, R., Kawanami, M., Matsunaga, F. and Ishino, Y. (2008) The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. *Journal of Biological Chemistry* **283**, 1601–1609.