# Chloroplast Origin and Integration[1]

**Geoffrey I. McFadden***

Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Victoria 3010, Australia

The origin of oxygenic photosynthesis must rank just after the origin of life itself as one of the most significant events in the history of life. The early photosynthesizers, which Bill Schopf has shown were similar to modern cyanobacteria (32), made earth a nicer place to live by quietly cranking out oxygen over hundreds of millions of years. Having established an ozone UV shield, the cyanobacteria then entered into an extraordinary partnership with eukaryotic cells: They became chloroplasts. Equipped with chloroplasts, the early plants were ready to colonize the land and green the planet. This review recounts major leaps in our understanding of chloroplast evolution from the preceding 25 years.

## ENDOSYMBIOSIS TRIUMPHS

In 1975 the theory of endosymbiosis (which describes the origins of mitochondria and plastids from eubacteria-like cells living within eukaryotic hosts) was still hanging in the balance. Major players like Lawrence Bogorad (4) and Tom Cavalier-Smith (7) were weighing the merits of the autogenous (nonendosymbiotic) alternatives for chloroplasts against the early compelling scenarios of Mereschkowsky and later Lynn Margulis. However, in the same year, an innovative paper from Linda Bonen and Ford Doolittle (5; which provided the first quantitative measure of similarities between T1 rRNA catalogues of cyanobacteria and chloroplasts) is now recognizable as the first ripple in a tidal wave of chloroplast molecular data that swept the autogenous origin hypothesis away.

Hans Kössel was one of the first to apply the recently developed Maxan/Gilbert DNA sequencing technology to chloroplasts. Primary and secondary structure of maize chloroplast rRNA was revealed to be more closely related to the bacterium *Escherichia coli* than to equivalent genes in the nuclei of eukaryotes (the yeast sequence only became available to Kössel at the proof stage!), corroborating the endosymbiotic hypothesis (34).

Flügge and Heldt recognized that integration of the endosymbiont required a transporter for export of photosynthate and characterized a family of phosphate translocators located in the chloroplast envelope (12, 13). It has subsequently emerged that plants (and probably algae) have learned to operate these transporters in reverse to "feed" their chloroplasts at night or in organs underground. Norman Weeden then recognized that the endosymbiont, in addition to the gift of photosynthesis, was responsible for the introgression of several other metabolic pathways into the host (38). Weeden's insight was the spur for the ongoing characterization of the bacterial-type chloroplast pathways for amino acid biosynthesis, carotenoid biosynthesis, non-Shemin tetrapyrole biosynthesis, non-mevalonate isoprenoid biosynthesis, and nitrate and sulfate assimilation. Weeden also recognized that many of the genes for these processes must have undergone intracellular transfer from plastid to nucleus.

## ONE RING TO RULE THEM. . . AND IN THE DARKNESS BIND THEM

The advent of Sanger-style DNA sequencing opened the way to more ambitious projects. Oyhama's team (28) was first to sequence a chloroplast genome determining the 121,024 bp of the liverwort *Marchantia polymorpha* chloroplast. The Japanese group's work inspired numerous other chloroplast genome sequencing projects from which various patterns, and a few surprises, have emerged. The *Marchantia* genome has proven to be a good general model. Typical chloroplast DNAs, including those of plants and most algae, are circular with genes organized in operons. Plastid gene content varies from about 70 in some non-photosynthetic plastids such as that of the malaria parasite (39) up to about 200 in red algae (30); plants typically have around 100 genes in the chloroplast. Dinoflagellate algal chloroplasts are the only example that so far seriously defy this trend, having single gene mini-circles that encode only a handful of genes (40). One ring no longer binds them.

The next major leap in the application of sequencing was the complete genome of the cyanobacterium of *Synechocystis* (20). The Japanese team identified 3,229 genes, which provided a reference point for the gene content of the chloroplast endosymbiont at the outset of endosymbiosis. For the first time we held a list of genes with which the endosymbiont probably started. Comparisons of this list with the present content of plastid genomes issues some weighty challenges.

The genome sequences also provided an opportunity to compare gene synteny between chloroplasts

[1] G.I.M. is supported by the Australian Research Council and is an International Scholar of the Howard Hughes Medical Institute.

* E-mail g.mcfadden@botany.unimelb.edu.au; fax 61–3–9347–1071.

and bacteria. Conservation of large operons proved that chloroplast DNAs are reduced bacterial genomes (28). In a converse manner, unique gene assemblages shared among all chloroplasts suggest that all chloroplasts arose from a single primary endosymbiosis (30), as first argued by Tom Cavalier-Smith (8). Many other studies now concur, but the idea is not universally accepted (23). Large amounts of sequence data have also allowed meta-analyses in which multiple genes are used in determining more robust phylogenies (25). These trees are consistent with the belief that the glaucophyte alga *Cyanophora*, which has a blue-green chloroplast with a peptidoglycan wall, is one of the earliest diverging lines to contain a chloroplast. Meta-analysis also confirms morphological clues suggesting that the prasinophyte *Mesostigma* is the closest living relative to the unicellular alga from which the land plants are descended (22).

## INTRACELLULAR GENE TRANSFER AND PROTEIN TARGETING

An important outcome of comparing the gene content of a cyanobacterium with that of chloroplasts is the detailed reconstruction of genome reduction. When it was first established that nuclear genes encode some types of chloroplast proteins it became clear that chloroplasts were genetically only semiautonomous and had relinquished genes to their hosts. The genome maps in conjunction with the phylogeny tell us that this process occurred in a somewhat ad hoc manner, although certain classes of genes are apparently less amenable to transfer (25). Two intriguing questions are how many nuclear genes service the chloroplast and are they all derived from the endosymbiont cyanobacterium. The Arabidopsis genome project is beginning to allow us to address these questions; early surveys indicate that as many as 2,000 nuclear genes encode proteins with plastid functions (1). The tagged mutagenesis program also provides a means to assay the functions of those genes whose products are predicted to be chloroplast targeted, but for which no obvious function is apparent.

The transfer of genes from chloroplast to nucleus typically requires return passage of the gene product to its place of function. This mechanism is central to the establishment and refinement of endosymbiosis. The first vagrant gene to be studied was the Rubisco small subunit (RbcS), and it was established early that a precursor form was produced in the cytosol and targeted. It was subsequently established that the precursor bears an N-terminal extension, the so-called "transit peptide," which mediates transport. There were several key developments in the understanding of this process. Dobberstein et al. (10) established that RbcS began as a precursor with a small (3.5 kD) extension that is proteolytically removed within the chloroplast prior to assembly of the holoenzyme. Gregory Schmidt and colleagues (31) determined the leader sequence and the site of processing. Chua and Schmidt (9) and Highfield and Ellis (19) subsequently demonstrated that the translocation into the chloroplast occurs post-translationally.

These experiments were important for two key reasons: first, they showed that chloroplast targeting was different to the "signal hypothesis" for cotranslational insertion into the endomembrane system, and second, they were the germ of a system to dissect the import process. By using in vitro translated precursors in conjunction with isolated plastids it has been possible to manipulate the system experimentally to identify conditions required for transport and components of the transport machinery. Mishkind et al. (27) demonstrated that the transit peptide was sufficient and necessary for import of proteins. Schrier et al. (33) and van den Broek et al. (37) established the use of transit peptides to direct foreign proteins into chloroplasts in transgenic plants, which paved the way for targeting of reporter proteins such as the jellyfish green fluorescent protein.

A key development in understanding the mechanisms of import was the identification of ATP as the essential requirement for transport (16). This dependence was exploited to interrupt transfer, which enabled the partially transferred precursor to be used as a tag (often through innovative use of cross linkers or affinity motifs attached to the targeted protein) to recover components of the transport apparatus. Approaches based on this strategy have identified numerous transport-related proteins (known as Tocs and Tics) located in the inner and outer chloroplast membranes such as the receptors (36) and the channel protein (24).

The targeting of proteins within the confines of the chloroplast was first studied by Smeekens et al. (35). They demonstrated that plastocyanin was targeted into the chloroplast stroma by a typical transit peptide, but that a second signal, this time homologous to bacterial secretion peptides, directed the protein across the thylakoid membranes. In other words, the original bacterial mechanisms for targeting proteins across the inner membrane have been retained in chloroplasts, a phenomenon now termed "conservative sorting." The transit peptide was thus perceived as a prefix appended to the protein to return it to the organelle for internal distribution.

The intense study of Rubisco allowed plant biologists to discover a major unifying principle for the assembly of oligomeric protein structures. John Ellis and colleagues (3) identified a 60-kD protein essential for Rubisco assembly. Sequencing of the so-called "Rubisco-binding protein" and the GroEL heat shock proteins of *E. coli* demonstrated that these proteins, which Ellis dubbed chaperones, are fundamental and ubiquitous components of protein folding and protein/protein interactions (18). This discovery is

one of the rare instances where plant science forged the way into a unifying concept with application across all cell biology.

## SECOND-HAND CHLOROPLASTS

Plant chloroplasts have two bounding membranes, but chloroplasts of many algae have three or four membranes. Sally Gibbs (15) first articulated the idea that these multi-membraned, or complex, chloroplasts were acquired indirectly, not by the classic mechanism of endosymbiosis of a cyanobacterial-like prokaryote. Gibbs proposed that the *Euglena* chloroplast, for instance, derived from a eukaryotic alga that had been engulfed by a eukaryotic phagotroph (15). The engulfed cell underwent drastic reduction such that in most cases the only residues are the chloroplast and the extra membranes created by the engulfment. Acquisition of chloroplasts in this manner occurred multiple times (the number of acquisitions is argued hotly) and thus explains the patchy distribution of chloroplasts across the eukaryotic tree. The process was first verified in cryptomonads where a minute residue of the endosymbiont nucleus has been identified (11). This process of acquisition is known as secondary endosymbiosis, as distinct from primary endosymbiosis between a eukaryote host and a prokaryote endosymbiont. Secondary endosymbiosis allowed lateral transfers of chloroplasts into non-photosynthetic lineages, including unexpected lineages such as the malaria parasite (26, 39). Gibbs also predicted (correctly) how proteins could be targeted across the multiple bounding membranes by initial utilization of the secretory pathway to cross the outermost membrane.

## REFORGING THE RING

During its hundreds of millions of years tenure within the host, the chloroplast genome has undergone substantial modification. Human intervention is now beginning to make even more drastic modifications to chloroplast DNA. Transgenics is an immensely powerful tool for biological understanding. Chloroplast transformation was first achieved in the green alga *Chlamydomonas* using a "gene gun" approach in which selectable markers were literally blasted into the cells (6). However, engineering of chloroplast genomes has not really achieved widespread application. Nevertheless, chloroplast transformation has opened doors to some exciting developments. Abrogation of the chloroplast's indigenous transcription system (knockouts of a cyanobacterial-like *rpo* gene in chloroplast DNA) provided definitive evidence for the existence of an elusive alternate polymerase (2). A single subunit, phage-type polymerase, encoded by the nucleus and similar to that used for mitochondrial transcription, was subsequently shown to be responsible for a major fraction of chloroplast gene transcription (17). The dual transcription systems are thought to allow regulation of chloroplast development—a hypothesis whose testing will require further transgenics.

The second major development stemming from chloroplast transformation was the laboratory reconstruction of intracellular gene relocation. Pal Maliga's group was able to delete the chloroplast gene for RbcL and complement the deletion by inserting a nuclear copy, complete with a motif to target the product, into the chloroplast (21). Perhaps the greatest promise for chloroplast transformation lies in commercial applications (14). The chloroplast genome has many features that make it an ideal site for insertion of useful genes. Chloroplast DNA is multicopy per cell, which ensures a high gene dosage and generally high expression levels for transgenes. Unlike nuclear plant transformation, genes can be targeted to specific sites in chloroplasts. This precision avoids the poorly understood "position effect" whereby the site of insertion of transgenes influences transgene behavior in an unpredictable manner. Chloroplast transgenics makes genetic engineering of plants more controlled. Last, because chloroplasts are maternally inherited in angiosperms, the risk of transgene spread into the environment through pollen is greatly reduced.

## THE FUTURE

The breakthroughs outlined here position us to tackle some fundamental questions in the next 25 years. We will soon have full gene complements for the host and endosymbiont and the ability to manipulate both genomes and target foreign proteins from the host to the endosymbiont (there are no leads on mechanisms to do the reverse). These approaches will be central in developing our understanding of still mysterious processes such as chloroplast division (29), the molecular signals regulating plastid differentiation, and the mechanisms of cross talk between the plant cell and its little green slaves. Armed with this information we may be in a position to undertake some very bold experiments. We may even be able to reconstruct endosymbiosis in the laboratory, putting chloroplasts into non-photosynthetic hosts.

## ACKNOWLEDGMENTS

Apologies to those whose work I could not cite. Thanks to Bill Martin who shared his literature knowledge so enthusiastically.

## LITERATURE CITED

1. **Abdallah F, Salamini F, Leister D** (2000) Trends Plant Sci **5:** 141–142
2. **Allison LA, Simon LD, Maliga P** (1996) EMBO J **15:** 2802–2809

3. **Barraclough R, Ellis R** (1980) Biochim Biophys Acta **608:** 19–31
4. **Bogorad L** (1975) Science **188:** 891–898
5. **Bonen L, Doolittle W** (1975) Proc Natl Acad Sci USA **72:** 2310–2314
6. **Boynton JE, Gillham N, Harris E, Hosler JP, Johnson AM, Jones AR, Randolph-Anderson BL, Robertson D, Klein TM, Shark KB** (1988) Science **240:** 1534–1538
7. **Cavalier-Smith T** (1975) Nature **256:** 463–468
8. **Cavalier-Smith T** (1982) Biol J Linn Soc **17:** 289–306
9. **Chua N-H, Schmidt GW** (1978) Proc Natl Acad Sci USA **75:** 6110–6114
10. **Dobberstein B, Blobel G, Chua N-H** (1977) Proc Natl Acad Sci USA **74:** 1082–1085
11. **Douglas SE, Murphy CA, Spencer DF, Gray MW** (1991) Nature **350:** 148–151
12. **Fliege R, Flügge U, Werdan K, Heldt H** (1978) Biochim Biophys Acta **502:** 232–247
13. **Flügge U, Fischer K, Gross A, Sebald W, Lottspeich F, Eckerskorn C** (1989) EMBO J **8:** 39–46
14. **Garcia JS, Graves J, Hajdukiewicz P, Hunter P, Nehra N, Paradkar V, Schlittler M, Carroll J, Spatola L, Ward D, Ye G, Russell D** (2000) Nat Biotechnol **18:** 333–338
15. **Gibbs SP** (1978) Can J Bot **56:** 2882–2889
16. **Grossman A, Bartlett S, Chua N-H** (1980) Nature **285:** 625–628
17. **Hedtke B, Borner T, Weihe A** (1997) Science **277:** 809–811
18. **Hemmingsen S, Woolford C, Vries Svd, Tilly K, Dennis D, Georgopoulos C, hendrix R, Ellis R** (1988) Nature 330–333
19. **Highfield PE, Ellis RJ** (1978) Nature **271:** 420–424
20. **Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S** (1996) DNA Res **3:** 109–116, 185–209
21. **Kanevski I, Maliga P** (1994) Proc Natl Acad Sci USA **91:** 1969–1973
22. **Lemieux C, Otis C, Turmel M** (2000) Nature **403:** 649–652
23. **Lockhart PJ, Steel MA, Hendy MD, Penny D** (1994) Mol Biol Evol **11:** 605–612
24. **Ma Y, Kouranov A, LaSalsa SE, Schnell DJ** (1996) J Cell Biol **134:** 315–327
25. **Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik K** (1998) Nature **393:** 162–165
26. **McFadden GI, Reith M, Munholland J, Lang-Unnasch N** (1996) Nature **381:** 482
27. **Mishkind M, Wessler S, Schmidt G** (1985) J Cell Biol **100:** 226–234
28. **Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H** (1986) Nature **322:** 572–574
29. **Osteryoung KW, Vierling E** (1995) Nature **376:** 473–474
30. **Reith M, Munholland J** (1995) Plant Mol Biol Rep **13:** 333–335
31. **Schmidt G, Devillers-Thiery A, Desruisseaux H, Blobel G, Chua N-H** (1979) J Cell Biol **83:** 615–622
32. **Schopf JW** (1993) Science **260:** 640–646
33. **Schreier PH, Sefotr EA, Schell J, Bohnert HJ** (1985) EMBO J **4:** 25–32
34. **Schwarz Z, Kössel H** (1980) Nature **283:** 739–742
35. **Smeekens S, Bauerie C, Hageman J, Keegstra K, Weisbeek P** (1986) Cell **46:** 365–375
36. **Sveshnikova N, Soll J, Schleiff E** (2000) Proc Natl Acad Sci U.S.A **97:** 4973–4978
37. **van den Broek G, Timko M, Kausch A, Cahmore A, Montagu M, Herrera-Estrella L** (1985) Nature **313:** 358–363
38. **Weeden N** (1981) J Mol Evol **17:** 133–139
39. **Wilson RJM, Denny PW, Preiser PR, Rangachari K, Roberts K, Roy A, Whyte A, Strath M, Moore DJ, Moore PW, Williamson DH** (1996) J Mol Biol **261:** 155–172
40. **Zhang Z, Green B, Cavalier-Smith T** (1999) Nature **400:** 155–159