Chapter 7

Reconstructing the Evolution of the Endocytic System:
Insights from Genomics and Molecular Cell Biology

Mark C. Field*, Carme Gabernet-Castello and Joel B. Dacks

Endocytosis is an essential process undertaken by most eukaryotic cells. At its most general, the term refers to the uptake of material from the cell milieu.[1] Cell biologists, however, have come to recognise a number of distinct modes of endocytic transport that are accompanied by differences in their underlying molecular mechanisms. Multiple modes can coexist in the same cell type and are frequently ongoing concurrently. Broadly, endocytic mechanisms can be subdivided based on the size of the ingested particle or cargo. Phagocytosis, or cell eating, is the uptake of large particles, including whole cells, and is accompanied by transport through large vesicular structures (>250nm in diameter). Pinocytosis, or cell drinking, involves uptake of rather smaller cargo, typically macromolecules and complexes. The study of endocytic pathways has, for very good technical reasons, focused on a small number of taxa, principally metazoa, yeast and a restricted number of protists. This has served well and has allowed the definition of a number of pathways in part by virtue of the molecules that are required for their operation.

A more complete understanding of eukaryotic diversity, both with respect to the evolutionary relationships between the major groups, and the availability of an increasingly representative taxonomic sampling of genomes now allows for a meaningful survey of the endocytic potential of eukaryotes and the reconstruction of major events in the evolution of the endocytic system. From a combination of single and multi-gene phylogenetic studies, along with morphological data, there are now recognized six major "super-groups" of eukaryotes,[2,3] as shown in Figure 1. The use of rare genetic characters, such as gene losses, innovations, and especially gene fusions, have also been key in establishing these groups and in rooting the eukaryotic tree. The presence of a derived gene fusion between dihydrofolate reductase and thymidilate synthase shared by the plantae, chromalveolata, excavata and rhizaria, and a unique myosin gene fusion in amoebae and opisthokonts have divided the eukaryotes into two groups; the bikonts on one side and the unikonts on the other.[4,5] Although this is still contentious, being based on only few characters and limited taxon sampling, ciliary root patterns and other diverse evidence[2] appear to support this root (see also the chapter by Brinkmann and Philippe).

These relationships are important as they allow us to make evolutionary deductions about traits within eukaryotes. If a trait (morphological or molecular) is found in representatives of most of the super-groups, and in particular in groups on both sides of the eukaryotic root (see Fig. 1), then the trait is ancient and any absences due to derived loss. If, on the other hand, a trait is found in a single group, then it is likely recently derived.

*Corresponding Author—Mark C. Field, The Molteno Building, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, U.K., Email: mcf34@cam.ac.uk

Eukaryotic Membranes and Cytoskeleton: Origins and Evolution, edited by Gáspár Jékely.

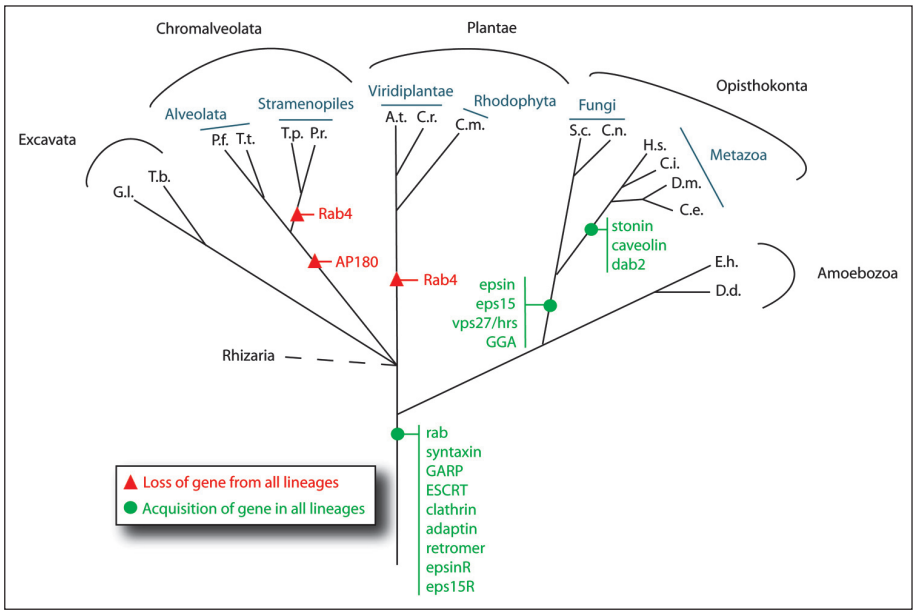

Figure 1. Model for major acquisition and loss of selected intracellular trafficking pathways during eukaryote evolution. The evolutionary relationships of various eukaryotic taxa are shown following the scheme of Simpson and Roger and references therein,[3] as well as evidence from Richards and Cavalier-Smith.[5] Positions of acquisitions and losses of gene families are indicated (green dots and red triangles respectively). Most lineages above a point of acquisition possess the gene, indicating that the last common ancestor of that clade would have had the respective gene, whilst all taxa above a red triangle lack this gene. A number of taxa-specific secondary losses are likely also present, but these have been omitted for simplicity (see Fig. 3). Taxa represented are H.s.; Homo sapiens, D.m.; *Drosophila melanogaster*, C.e.; *Caenorhabditis elegans*, C.i.; *Ciona intestinalis*, S.c; *Saccharomyces cerevisiae*, C.n; *Cryptococcus neoformans*, E.h.,; *Entamoeba histolytica*, D.d.; *Dictyostelium discoidium*, A.t.; *Arabidopsis thaliana*, C.r.; *Chlamydomonas reinhardtii*, C.m.; *Cyanidioschyzon merolae*, P.f.; *Plasmodium falciparum*, T.t.; *Tetrahymena thermophila*, G.l.; *Giardia lamblia*, T.p.; *Thalassiosira pseudonana*, P.r.; *Phytophthora ramorum* and T.b.; *Trypanosoma brucei*. A color version of this figure is available online at www.eurekah.com.

In this article we first consider the overall structures of the endocytic systems of selected model eukaryotes and then survey selected genomes for the presence of key factors involved in endocytosis and associated trafficking pathways. This latter aspect assumes that the mechanisms subtending various transport routes are conserved, and implies that the presence or absence of a particular factor mirrors the presence/absence of a pathway or process. This approach obviously falls short of 100% accuracy, but given high degrees of conservation we have observed in experimental studies of the deeply divergent trypanosomes[6] as well as detailed in silico analysis of restricted gene families across many taxa,[7] we consider this to be an informative strategy.

Defining Endocytosis

Phagocytosis has been observed in organisms on both sides of the eukaryotic root and can therefore be considered an ancient mechanism. Indeed some authors speculate that phagocytosis was the founding innovation and the driving evolutionary force for the evolution of the eukaryotic state.[8] In organisms such as *Paramecium* and amoebae (e.g., *Entamoeba* and *Dictyostellium*) the process serves to supply the cell with nutrients via the ingestion of bacteria

and other organic material. Phagocytic mechanisms are retained in multicellular eukaryotes, including metazoans, where the ability to ingest whole cells has become coopted for specialised functions, including defence against infectious agents as well as for management of programmed cell death or apoptosis; in both of these examples specialised phagocytes, macrophages, are responsible. At the molecular level, phagocytosis is characterised by a dependence on actin and also small GTPases of the Rho subfamily.[9] Beyond noting its obviously critical and ancient nature, however, we will not treat phagocytosis further here, in order to focus on the various and better characterized molecular components of pinocytosis.

Pinocytosis has been reported in the majority of eukaryote taxa where directly investigated and is also therefore an ancient mechanism. There are several types and multiple functions of pinocytotic endocytosis, which includes fluid-phase uptake of the media (which may include dissolved solutes) and receptor-mediated endocytosis (RME). Functions include nutrient uptake, environmental sampling, turnover of surface components and also cell signalling, whilst the various mechanisms can be divided into clathrin-dependant and -independent, reflecting a requirement for the conserved heterodimeric clathrin protein. Further, there are several modes of clathrin-independent endocytosis, at least one of these involves a cholesterol-binding protein caveolin, which can also be differentiated from clathrin-dependant endocytosis based on the morphology of endocytic structures associated with the cell surface membrane. Specifically caveolin is associated with caveolae, whilst clathrin-mediated mechanisms are associated with clathrin-coated pits and vesicles. Additional pathways that require neither clathrin nor caveolin are also present in some cells, but the lack of a marker molecule specific to these modes has precluded detailed investigation of these systems, and they will not be considered further here.

The General Structure and Morphological Evolution of Endocytic Systems

The basic architecture of the endocytic system is shown in Figure 2A. The principle features are (i) multiple routes from the surface, (ii) several functionally differentiated endosomal structures including the early endosome, the recycling endosome and late endosomes, (iii) integration with a degradative pathway, variously termed the lysosome, vacuole or reservosome in different systems, and (iv) close integration with the Golgi complex and the trans-Golgi network in particular, and hence exocytic traffic.[1] In general most systems retain these features, but there are examples of organisms where one or more aspect has been lost.

In some eukaryotic supergroups, endocytic specialization is absent,[10] whether due to a photosynthetic lifestyle and thus presumably reduced endocytic activity (Plantae) or due to the possible selective advantage of amoebic versatility in being able to phagocytose wherever prey may be available (Amoebae). Cercozoans also feed phagocytically by filose pseudopodia and lack a truly specialized endocytic region.[10] Numerous organisms however, demonstrate restriction in the spatial location of their endocytic apparatus. More extreme examples include *Paramecium*, where the phagocytic system is associated with a cytopharynx[11] and trypanosomatids where all endocytic activity is restricted to the flagellar pocket,[12] but restriction of cell surface sites where endocytic activity may occur is also observed in multicellular systems, including metazoa. The selective pressures that underpin such polarisation are likely multiple, and include segregation of function by control of membrane protein/lipid composition (polarised cells in metazoa), feeding efficiency in many excavates, and Chromalveolates, including the model ciliate *Paramecium* and immune evasion (the flagellar pocket in trypanosomes).

In Opisthokonts (i.e., many metazoan cells and in some fungi including *Saccharomyces cerevisiae*), there is limited differentiation of the plasma membrane, with endocytic activity being initiated from most regions of the membrane (Fig. 2D). Microdomains may exist whereby small areas of the membrane are marked for preferential endocytic activity, but specialised membranous structures are, by and large, absent. Polarised cells, for example epithelia and neuronal cell types, are an exception where specific endocytic activity is derived from distinct membrane microdomains.[13] In these examples, barriers to diffusion of surface proteins and

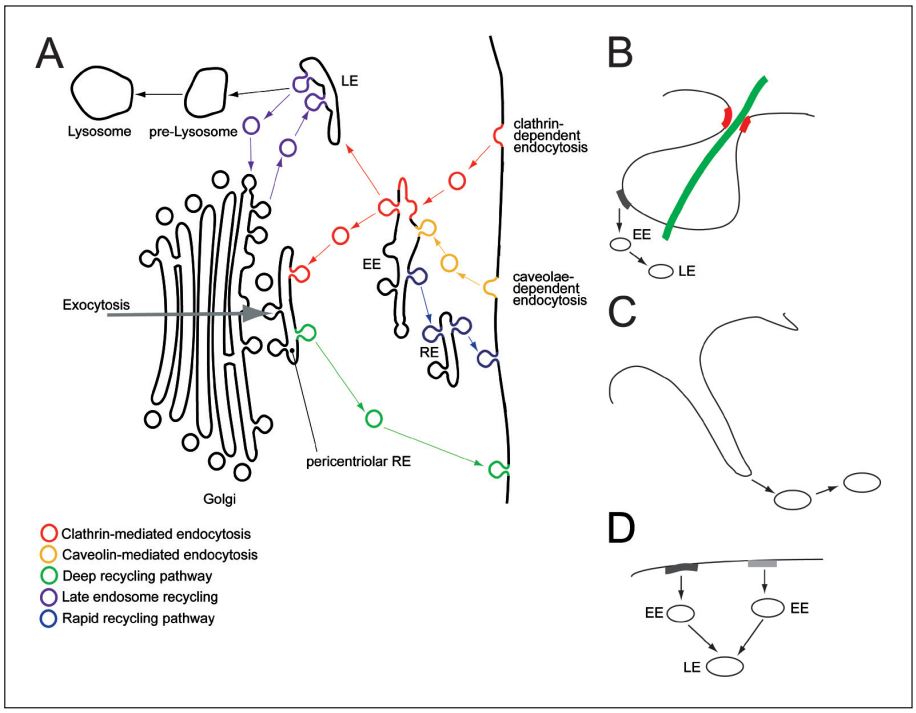

Figure 2. Schematics of arrangements of endocytic systems at the cell surface. A) general endocytic pathway showing the relationship between the Golgi, endosomes, cell surface and the lysosome. EE; early endosome, LE; late endosome, RE; recycling endosome. B) Strictly gated configuration. One (or more) endocytic mode originates from a sequestered area of the membrane. In this example the area of active endocytosis is shown as an invagination, with barriers to diffusion designated in red. Only clathrin mediated mechanisms are shown. The structure in green is a flagellum. Examples of this configuration can be found amongst the trypanosomes, although similar invaginations have been observed in many protists. C) Funnel configuration. A specialised invagination where endocytosis is restricted. In this example it is not clear if there is a strict barrier to diffusion of membrane proteins. A good example of this is the cytopharynx of *Paramecium*. D) Open access configuration. Multiple endocytic mechanisms, as denoted by the dark (clathrin) and light (caveolin) gray bars at the plasma membrane, originate from an essentially undifferentiated membrane. The presence of specific tags on membrane microdomains directing assembly of endocytic structures may serve to differentiate the membrane, but at the time of writing the presence of such tagged sites has yet to be rigorously established. The two distinct endocytic routes are shown converging later within the system, consistent with current knowledge from the mammalian system. Examples of this type of system are found in the metazoa.

lipids serve to maintain a functional division. These latter features are clearly restricted to multicellular organisms, and therefore likely represent specific specialisations that have arisen during evolution of given multicellular lineages.

In *Paramecium*, a well-studied representative of the Chromalveolata, the architecture of the cell is much more specialised. From the perspective of membrane transport, two organelles deserve specific comment; the cytopharynx and the cytoproct. The former is a specialised invagination in the plasma membrane that forms a funnel-like structure into which particles are guided by the beating of cell surface cilia; from the distal end of this organelle rapid phagocytosis takes place (Fig. 2C). Phagocytosis is followed by rapid acidification, presumably to aid both in digestion of prey and also in killing; acidification is achieved by an unusual process that

involves fusion with preexisting acidisomes derived from the Golgi complex.[14] This stands in stark contrast to the acidification of endosomes and phagosomes in the Opisthokonta which involves a bafilomycin-sensitive membrane H+ pump. Further, *Paramecium* phagosomes are cycled through the cell, and material is ultimately expelled from the cell via the cytoproct, an unusual example of exocytosis. In addition, the presence of a large number of trichocysts, essentially dense granule vesicles involved in rapid regulated exocytosis, also requires a rapid endocytic mechanism to recapture exocytosed membrane at the cell surface.[11] Subtending these surface events are several endosomal compartments, many of which appear to have rather unusual morphologies.[11]

In the kinetoplastida, which include free living and parasitic forms, the cell is highly polarised, with an invagination (the flagellar pocket) at one end of the cell where the flagellum enters the cell body and crosses the membrane (Fig. 2B). In some systems, e.g., *Trypanosoma brucei,* all endocytic and exocytic transport is directed towards the flagellar pocket membrane, which functions as a unique membrane domain. Further subdivision of the structure into flagellar pocket and cytostome is seen in *Trypanosoma cruzi*; and may serve to further differentiate the surface as the cytostome appears to be dedicated to endocytic activity.[15] It is speculated that the flagellar pocket plays a role in immune evasion in parasitic protozoa by sequestering immunogenic determinants.[6]

The sister group to kinetoplastids, the euglenids, also have a flagellar pocket-like structure. This, however, is used as an ingestion tubule for feeding, suggesting a more standard endocytic role for the flagellar pocket in the past. Strong molecular sequence data links the euglenozoa (euglenids plus kinetoplastids) to the heterolobosean amoeboflagellates,[16] which have a much more complex ventral feeding groove.[17] The presence of conserved homologous cytoskeletal features underlying the ventral groove links these three groups to other ventral groove possessing taxa, including *Malawimonas*, *Trimastix*, "core" jakobids, *Carpediemonas* and retortamonads. The scenario of a nongroove-possessing organism being linked to one possessing the suite of homologous ultrastructural characters is repeated several times for the excavate taxa.[17] Indeed, the molecular link within the groove containing taxa is the contentious point for this super-group, with no analysis uniting all ten proposed excavate groups into a single clade. However, the nodes separating the groups are generally poorly supported and rates of evolution for the genes used in the analyses vary quite strongly between excavate taxa, possibly explaining the failure to resolve them as a group.[2] Nonetheless, based on the molecular evidence linking kinetoplastids with feeding groove possessing taxa, as well as on the strong morphological evidence, it seems clear that the flagellar pocket is a highly diverged version of the ventral feeding groove. Overall, the specialization of endocytic machinery seems to be a wide-spread, if not always homologous, feature of eukaryotic cells being found in three of the six major eukaryotic super-groups.

Key Factors Involved in Endocytic Systems and their Evolutionary Distribution

Vesicle transport is controlled by a large number of proteins and lipids that interact dynamically to assemble and disassemble specific complexes in a time and location-dependent manner. Some of these factors are general, e.g., NEM-sensitive factor, an ATPase that plays a role in the majority of vesicle fusion events. However, many are specific to individual pathways, including Rab family GTPases, proteins of the SNARE superfamily and many others. Here we consider some of the key factors that are required for endocytosis and the associated recycling pathways, and by probing 17 genomes covering as wide a range of the eukaryotic tree of life as possible attempt to infer the evolutionary distribution of these factors (see results in Fig. 1).

Rabs and Syntaxins

The major endocytic and recycling pathways are modulated by one of four Rab proteins; Rab5 mediates endocytosis itself, Rab4 and Rab11 control rapid and deep recycling pathways respectively, whilst Rab7 regulates delivery to the lysosome/vacuole (Fig. 2A -see figure legend).

Figure 3, viewed on following page. Evolutionary distribution of selected major protein players in endocytosis. The presence or absence of endocytic factors, as assessed by BLAST are shown. Databases were searched in mid-2005 for the presence of a range of diagnostic protein factors with known major roles in a number of important endocytic pathways using translated BLAST. Over 550 independent BLAST analyses were performed, involving genome specific BLAST using authentic query sequences (typically from *H. sapiens* or fungi), followed by a reverse BLAST to the nr database. Whilst such analysis needs to be dealt with the appropriate caution, the presence of a factor is a good indicator that a specific pathway is present. Conversely, absence indicates that a pathway is likely to be absent. For most databases coverage of the relevant genome is of sufficient depth to make prediction reliable; however, several of the taxa discussed here have genome datasets that are not sufficient for confident prediction of absence, whilst even completed genome projects may lack ~1% of ORFs. Note that all databases are considered complete except *Tetrahymena thermophila, Giardia lamblia, Thalassiosira pseudonana, Chlamydomonas reinhardtii* and *Phytophthora ramorum*. Also *Entamoeba histolytica* and *Dictyostelium discoidium* genomes are nominally completed but depth of coverage may not be sufficient to ensure all ORFs are included in the current data sets. Filled symbols: protein:protein BLAST routines retrieve significant hit ($\sim 10^{-10}$), most relevant domains are present, predicted protein is of size consistent with orthology, and reverse BLAST is successful, open symbols: not returned from database searches, significant hit not obtained by BLAST, or inspection of sequence returned indicates nonorthology. We prefer the term "not returned" to "not found" as the possibility remains that an orthologue has been missed by our analysis. Large taxon groupings as are commonly recognised are subdivided into segments for each species within that grouping (if appropriate). Colours used are arbitrary and are for clarity only. Taxa represented are H.s.; *Homo sapiens*, D.m.; *Drosophila melanogaster*, C.e.; *Caenorhabditis elegans*, C.i.; *Ciona intestinalis*, S.c; *Saccharomyces cerevisiae*, C.n; *Cryptococcus neoformans*, E.h.,; *Entamoeba histolytica*, D.d.; *Dictyostelium discoidium*, A.t.; *Arabidopsis thaliana*, C.r.; *Chlamydomonas reinhardtii*, C.m.; *Cyanidioschyzon merolae*, P.f.; *Plasmodium falciparum*, T.t.; *Tetrahymena thermophila*, G.l.; *Giardia lamblia*, T.p.; *Thalassiosira pseudonana*, P.r.; *Phytophthora ramorum* and T.b.; *Trypanosoma brucei*. A color version of this figure is available online at www.eurekah.com.

Rab6 controls retrograde traffic through the Golgi complex. Rab proteins have several roles, including modulation of SNARE function and recruitment of effector proteins to specific vesicles, which in turn serve to propel transport forward - effectors include lipid kinases, components of the cytoskeleton and tethering factors. Syntaxins are coiled-coil transmembrane proteins, of the SNARE superfamily, that can form complexes in trans, i.e., between membranes. They are functionally involved in various stages of the fusion process.[18]

Extending conclusions from previous work,[19-21] we found that both the Rab and syntaxin families are deeply conserved (Fig. 3), and multiple members are present throughout evolution. Rab11 and Rab7 are universal, being found in all taxa sampled; whilst the conserved presence of a terminal endosomal compartment, i.e., lysosome/vacuole, was expected, the complete retention of Rab11, and hence a deep recycling system, was not so obvious. In addition, Rab6 was recovered from all taxa except *G. lamblia*; therefore Rab6, 7 and 11 represent core components of the endocytic system that are likely essential for eukaryotic life. Rab5 was also recovered from most taxa, with the only exceptions being *C. merolae* and again *G. lamblia*. *C. merolae* has a reduced genome and also lives at high acidity (pH1.5). It likely has minimal endocytic activity, albeit with retention of the Rab11/Rab7 pathways.[22] Although the *Giardia* genome database is incomplete and *Giardia* gene sequences are known to be divergent and thus vulnerable to mis-classification by BLAST, our failure to identify various *Giardia* components may be due to true absence. *Giardia* appears to have a minimal endomembrane system, with true early/late endosomes lacking, and a fused endosomal compartment (peripheral vacuole) instead.[23,24] It is clear from multigenome analysis and placement of the eukaryotic root,[2,25] that whilst this configuration has been suggested to represent a basal state,[23] it has likely arisen by secondary simplification. Rab4 is less well retained overall, being lost from representatives of the fungi, metazoa, chromalveolates and also all plantae. Hence the Rab4 pathway, whilst ancient, has been subjected to secondary loss from multiple taxa and the function of Rab4 in recycling is likely, in some circumstances, overshadowed by the Rab11 pathway.[26] The syntaxins

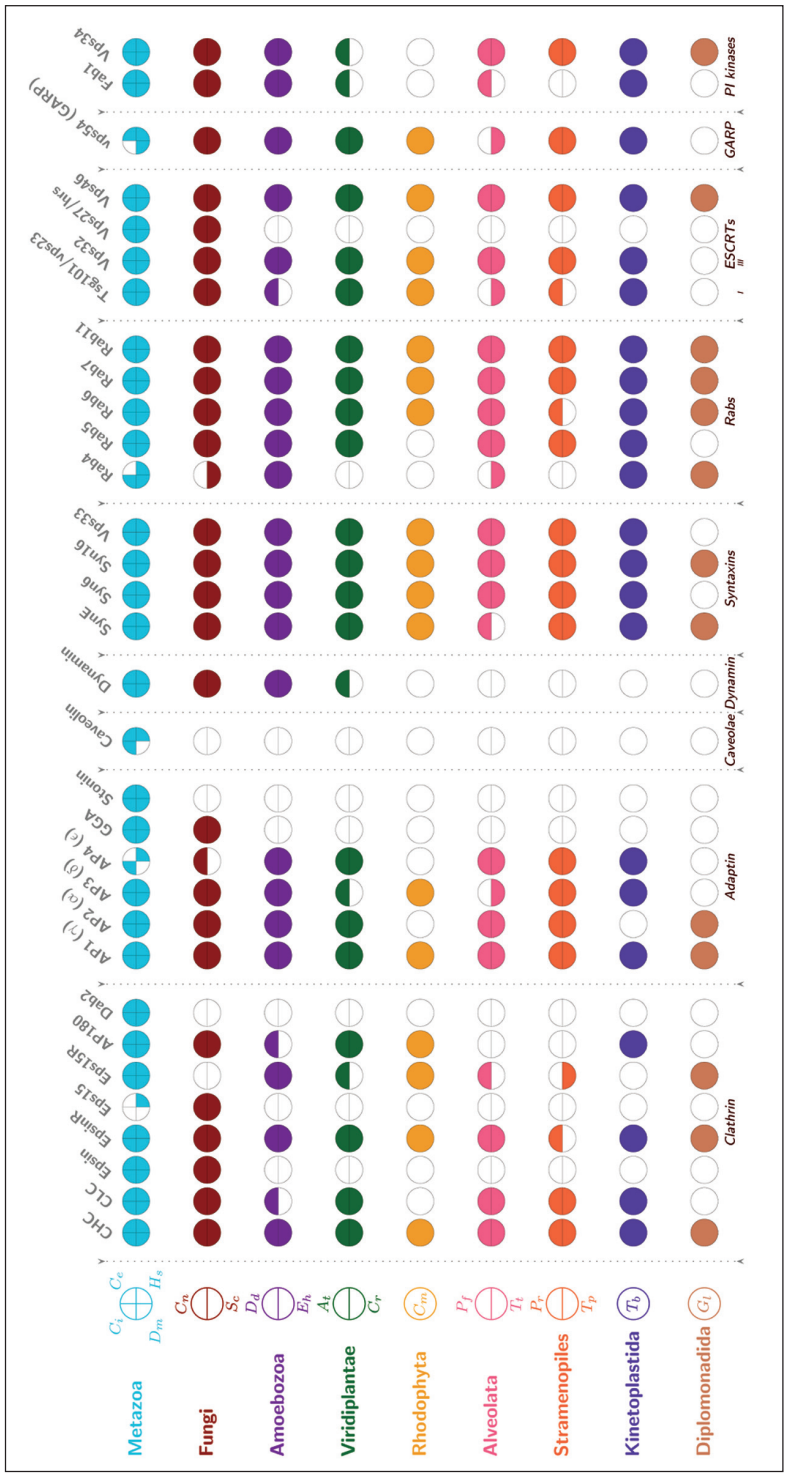

Figure 3. See figure legend on previous page.

are also highly conserved (Fig. 3), with the representatives of endosomal syntaxins (SynE), Syn6 and Syn16 being near universal. Absences of Syn6 in *G. lamblia* and SynE in *T. thermophila*, *C. reinhardtii* and *C. merolae* are explainable by artefact or secondary loss.

Two tethering factor complexes that interact with both Rab and SNARE proteins are worthy of consideration here. The Golgi-associated retrograde protein (GARP) complex is composed of four vps gene products (vps51, 52, 53 and 54) and acts as a tethering complex for Rab6. GARP also links with Syntaxin 10. Its major function appears to be mediating a retrograde pathway from endosomes to the Golgi complex.[27] A second recently described tethering complex, the HOPS complex, is involved in vacuole biogenesis, is an effector for Rab7 in yeast, and contains a vacuole SNARE protein.[28] Therefore, GARP and HOPS appear to share similar mechanisms in their function.

Both of these complexes are ancient (Fig. 1), but retention of GARP through evolution is less strong than for HOPS. Unusual aspects of the Golgi complex structure and function,[29,23] or the unusually high levels of sequence divergence in both *Plasmodium* and *Giardia* could equally account for their apparent lack of GARP. The absence of GARP from *C. intestinalis* was confirmed by unsuccessful searches for additional subunits of the complex; the functional implications of GARP loss in a metazoan genome are unclear. Vps33, a Syntaxin-binding protein and component of the HOPS complex, is universal, except in *Giardia*, where it could have been lost from the fused peripheral vacuole.

Endocytic Coats

The major endocytic coat proteins, clathrin and caveolin function in independent pathways but both operate mechanistically via a cooperative assembly process that subtends the plasma membrane and also includes the concentration of cargo in a local region or domain.[30] Searches for both the light and heavy chains showed that clathrin is a universal component of the eukaryotes; by contrast caveolin, which is primarily involved in endocytosis of lipid-anchored proteins, is only found in metazoa (Fig. 3). Therefore the latter pathway is a recent acquisition and indicates that the trafficking of lipid-anchored proteins is fundamentally different between divergent systems. Significantly this indicates that model systems may not always produce universally applicable generalizations regarding trafficking processes, in contrast to recent suggestions (e.g., 31).

Dynamin and Adaptins

Assembly of clathrin-coated pits at the plasma membrane results in the recruitment of a large number of cofactors, including the large GTPase dynamin and members of the adaptin complex family. The dynamin family is extensive, but not all members are components of membrane transport systems. Those involved in the endocytic system are responsible for vesicle scission from a donor membrane and appear to be defined by the presence of at least three domains, the GTPase, the "middle" region and the GTPase effector domain.[32] A ubiquitous requirement for dynamin in clathrin-dependent endocytosis has been challenged by the discovery of dynamin-independent endocytosis in *T. brucei*.[12] Further, we find the presence of dynamins, retaining the same domain configuration as those proteins documented to have a role in endocytosis in higher eukaryotes, being restricted to the metazoa, fungi, amoebozoa and higher plants and absent from the excavata and chromalveolata (Fig. 3). Based on the current understanding of relationships (Fig. 1), instead of invoking secondary loss, it is slightly more parsimonious to propose that dynamin is a later addition to the endosomal system, whilst a role in mitochondrial membrane biogenesis may be the more primitive function.

The adaptins are heterotetramers that in part are responsible for cargo recognition and retention in clathrin-coated membrane regions. There are four known adaptin complexes, with differing roles; AP-2 is the principal player in endocytosis, but AP-3 and 4 likely also have a role in the recycling and lysosome targeting portions of the endocytic system.[33] We found all adaptins to be widely distributed, indicating an ancient origin for all four complexes (Fig. 3).

Both AP1 and AP2 are universally represented, with the sole absence of AP2 from *T. brucei*, likely due to some novel aspects of endocytosis in that system as the complex is present in related kinetoplastids.[6] AP3 and AP4 are less well retained, and their absence is noted from a wide range of taxa. This is likely due to a dispensable function, dependant on the niche the organism occupies, with such conditions having been met on multiple occasions. Evidence from genetic studies in yeast and other systems support the idea that these adaptins are nonessential and not part of the mechanistic machinery.[34]

Distantly related to the adaptins are the GGAs; these modular proteins contain a domain related to the "ear' of γ-adaptin, function in a pathway modulated by ARF-like GTPases, and also recognise cargo as well as bind to components of the Rab5 system. A further adaptin-related coat complex are the stonins. In metazoa and yeast, the GGAs deliver post-Golgi cargo to the late endosome.[35] Our survey shows GGAs and stonins to be recent acquisitions, with GGAs found in Opisthokonts whilst the stonins are restricted to metazoan systems only.

Epsin and Associated Proteins

Four further proteins of special note participate in the early steps of endocytosis; epsin, eps15, dab2 and AP180. Epsin may be able to function to deform membrane in both the presence and absence of clathrin. It has a modular structure: comprising an epsin N-terminal homology domain, which binds phosphoinositides; ubiquitin-interaction motifs (UIMs); and a flexible region that includes binding sites for clathrin, AP-2 and epsin-homology domain-containing proteins such as Eps15, another UIM-containing endocytic factor. Epsin is implicated in mediating a clathrin-independent endocytic pathway as well as conducting interactions with multiple proteins and membrane lipids involved in the clathrin-dependent route.[36] Eps15 is another clathrin-binding protein that forms part of the major network of proteins that subtend the clathrin coat;[37] eps15 interacts with a host of other factors including the actin cytoskeleton and ubiquitin via UIMs, and data suggests is important for coordination of the endocytic system through interaction with the vesicle uncoating system. Both epsin and esp15 are restricted to metazoa and fungi, but the closely related epsinR (R for "related") and esp15R are widely distributed (Fig. 3). The major difference between epsin/eps15 and epsinR/eps15R is the presence of UIMs in epsin and eps15, but not in epsinR/eps15R, suggesting the recent acquisition of a ubiquitin-dependent aspect to endocytosis by the Opisthokonta.

A further component of this system may also be Dab2; this protein also likely functions as an adaptor, recognising cargo molecules via the presence of specific peptide signals and is present as part of the clathrin coat;[35] Dab2 is metazoa specific (Fig. 3). The final adaptor molecule, AP180, is also implicated in the recognition of membranes in clathrin-dependent transport processes.[38] In contrast to Dab2, we found AP180 to be an ancient component of the endocytic system, but which appears to have been lost from the Chromalveolate lineage (Fig. 1).

PI-Kinases

The control of the endocytic system is mediated in part by the Rab proteins, but is also integrated with the protein and lipid kinase system. Specifically, kinases act as effectors to transmit information through a pathway, frequently this information is derived ultimately from a GTPase such as a Rab protein. In the case of endosomal trafficking, considerable evidence indicates a specific role for phosphatidylinositol (PI) lipid kinases. The most important of these appear to be PI-3 and PI-5 kinases, including the vps34 and Fab1 gene products. Vps34 interacts with Rab7, likely controlling vacuolar/lysosomal targeting,[39] and is widely distributed, but with secondary losses within the plantae (Fig. 2). Fab1, also important for vacuolar transport, is clearly ancient but has been lost from lineages of the Plantae, Chromalveolates and the Excavata (Fig. 2), suggesting a more dispensable function. There is good evidence that, amongst other roles, formation of phosphatidylinositol phosphates generates specific binding sites for a number of other endocytic factors, including epsin and components of the ESCRT system.

ESCRTs

The ESCRT group of complexes mediate delivery of endocytic cargo to the multivesicular body and through to the lysosome. Three ESCRT complexes, each consisting of several orthologues of the yeast vps class E complementation group gene products, have been implicated in lysosomal delivery of receptors, viral proteins and other factors.[40] As shown in Figure 1, the three complexes are also ancient, with some secondary loss of the Tsg101/vps23 gene, a component of ESCRT I. Importantly, associated complexes are responsible for processing and delivery of endocytosed proteins that have been modified by ubiquitin to the multivesicular body (MVB), a late stage in the pathway to the lysosome, via ESCRT-dependant pathways. One factor important in MVB targeting, Vps27/Hrs, is in fact a recent development and is restricted to the fungi and metazoa (Fig. 2). This poses the important question of how proteins are targeted to MVBs in the majority of taxa.

Perspectives

There is ample evidence for the presence of conserved core vesicle fusion machinery throughout the eukaryotic lineage,[7] together with suggestions for retained coat complexes and other highly conserved factors.[6,19-21,41,42] However, many of these factors are involved in diverse pathways. Hence whilst it is no real surprise that these proteins are represented, their presence does serve to underscore the ancient origin of the majority of membrane-trafficking systems. An even more ancient origin of the overall system is suggested by the demonstration that Sec13p in *S. cerevisiae* is a component of both the COP II complex that mediates exit from the endoplasmic reticulum and also of the nuclear pore complex (NPC), implying the presence of a primitive membrane-binding and deforming complex that predates the endomembrane system.[43] The model was further supported by demonstration of considerable secondary structural similarity between several core proteins of the vesicle transport system (clathrin, adaptin, COP I and II) and members of the NPC Nup84p subcomplex.[44] The NPC, which also binds membrane and is responsible for increasing bilayer curvature in a manner akin to vesicle budding and membrane trafficking systems may have a common origin predating the emergence of a true eukaryotic cell. Hence establishment of the basic mechanisms of membrane trafficking is very ancient indeed, and the last common eukaryotic ancestor (LCEA) would have possessed the machinery to carry out such processes. The question then is not how ancient is the general system but, how elaborate had the endocytic system become in the LCEA?

The presence of multiple Rab and syntaxin family members throughout the eukaryota indicates that the LCEA likely possessed a differentiated endosomal system, which included at least two recycling systems depending on Rab4 and Rab11. Secondary losses are common, likely reflecting significant selective pressures on multiple lineages, and indeed Rab4 appears lost on at least two occasions. Hence the endocytic system displays a degree of flexibility and redundancy, but the core components are highly conserved, with the majority of organisms retaining nearly all of the factors we have analysed. Figure 2 illustrates the complex endocytic machinery present at the base of eukaryotes. In addition to the factors surveyed here, we have evidence for the ancient presence of the retrograde endosomal recycling coat complex, retromer (JBD et al, unpublished). The identification of individual components establishes the genomic presence of these genes early on. However, the fact that we can identify multiple members of complexes (e.g., The three ESCRTs, or the GARP plus syntaxin 6) suggests conserved functional interactions and pathways in diverse eukaryotes retained from the LCEA.

Nonetheless, a number of significant transitions are apparent, particularly associated with emergence of the Opisthokonta. For example, the appearance of the epsin/eps15 proteins may be explained by the transfer of the UIM into these proteins, facilitating exploitation of ubiquitination as a mechanism in endocytosis. As well as informing the evolution of the endocytic system, these traits also iterate back into understanding of the relationships amongst eukaryotic groups, representing novel examples of rare genetic characters. The Opisthokont innovations (Epsin, EPS15, GGA and Vps27/hrs) help to further cement the monophyly of this

supergroup, while the loss of AP180 (although weaker due to its being negative data) supports the, sometimes contentious, Chromalveolata.[45] Exploring the evolutionary details of these endocytic components, especially amongst unrepresented members of the unicellular relatives of Opisthokonts such as nuclearid amoebae[46] or other chromalveolate lines such as haptophytes and cryptophytes[45] may help to resolve relationships within the super-groups.

The great range of trafficking pathways in some of the protozoan systems perhaps suggests that there may be truly novel mechanisms at work for membrane transport in these systems. However, the comparatively poor experimental tractability of many of these organisms means that molecular studies are not very advanced and all evidence so far in fact points to a conservation of mechanism despite great novelty at the morphological level. The lack of acquisition of novel endocytic factors in the Excavata, Chromalveolata and Plantae could reflect a true paucity of innovation in these taxa, but an attractive alternative is that the poor current state of understanding of endocytosis in these taxa could explain this absence, and further experimental examination of tractable systems is clearly called for. These observations also impose limits on the validity of a model system, and highlight the need to sample endocytic systems across the eukaryota. For example, experimental data coupled with in silico analysis indicates that tryanosomes are indeed divergent from fungi and mammals and as such cannot be viewed as a model for processes within the Opisthokonta.[6,46] To understand the endocytic system in diverse taxa requires direct experimental work on representative organisms and cannot be inferred. The identification of truly novel factors in trypanosomes, *Giardia*, or any other system, will be an exciting and significant challenge.

Acknowledgements

We are especially grateful to Richard Coulson (EBI) for designing Figure 3. MCF's laboratory is supported by project and program grants from the Wellcome Trust. JD is a Wellcome Trust Postdoctoral Fellow. We are also grateful to the following for open access to genome sequence data that was used in our analysis; the Sanger Institute, The Institute for Genome Research, the Department of Energy (United States) and the *Cyanidioschyzon merolae* genome project.

References

1. Alberts B, Johnson A, Lewis J et al. Molecular biology of the cell. 4th ed. New York: Garland Publishing, 2002.
2. Simpson AG, Roger AJ. Excavata and the origin of amitochondriate eukaryotes. In: Hirt RP, Horner DS, eds. Organelles, Genomes and Eukaryote Phylogeny: An Evolutionary Synthesis in the Age of Genomics. Boca Raton, Florida, USA: CRC Press, 2004.
3. Simpson AG, Roger AJ. The real 'kingdoms' of eukaryotes. Curr Biol 2004; 14:R693-6.
4. Stechmann A, Cavalier-Smith T. Rooting the eukaryote tree by using a derived gene fusion. Science 2002; 297(5578):89-91.
5. Richards TA, Cavalier-Smith T. Myosin domain evolution and the primary divergence of eukaryotes. Nature 2005; 436:1113-1118.
6. Field MC, Carrington M. Intracellular membrane transport systems in Trypanosoma brucei. Traffic 2004; 5(12):905-13.
7. Dacks JB, Field MC. Eukaryotic cell evolution from a genomic perspective: The endomembrane system. In: Hirt RP, Horner DS, eds. Organelles, Genomes and Eukaryote Phylogeny: An Evolutionary Synthesis in the Age of Genomics. London: CRC Press, 2004:309-334.
8. Cavalier-Smith T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int J Syst Evol Microbiol 2002; 52(Pt 2):297-354.
9. Bokoch GM. Regulation of innate immunity by Rho GTPases. Trends Cell Biol 2005; 15:163-71.
10. In: Lee JJ, Leedale GF et al, eds. The Illustrated Guide to the Protozoa. Lawrence, Kansas: Society of Protozoologists, 2002.
11. Plattner H, Kissmehl R. Molecular aspects of membrane trafficking in paramecium. Int Rev Cytol 2003; 232:185-216.
12. Morgan GW, Hall BS, Denny PW et al. The kinetoplastida endocytic apparatus. Part I: A dynamic system for nutrition and evasion of host defences. Trends Parasitol 2002; 18(11):491-6.

13. Wang E, Pennington JG, Goldenring JR et al. Brefeldin A rapidly disrupts plasma membrane polarity by blocking polar sorting in common endosomes of MDCK cells. J Cell Sci 2001; 114:3309-21.
14. Allen RD, Ma L, Fok AK. Acidosomes: Recipients of multiple sources of membrane and cargo during development and maturation. J Cell Sci 1993; 106(Pt 1):411-22.
15. Porto-Carreiro I, Attias M, Miranda K et al. Trypanosoma cruzi epimastigote endocytic pathway: Cargo enters the cytostome and passes through an early endosomal network before storage in reservosomes. Eur J Cell Biol 2000; 79:858-69.
16. Baldauf SL, Roger AJ, Wenk-Siefert I et al. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 2000; 290(5493):972-7.
17. Simpson AG. Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). Int J Syst Evol Microbiol 2003; 53(Pt 6):1759-77.
18. Teng FY, Wang Y, Tang B. The syntaxins. Genome Biology 2001-2, (reviews 3012.1-3012.7).
19. Dacks JB, Doolittle WF. Novel syntaxin gene sequences from Giardia, Trypanosoma and algae: Implications for the ancient evolution of the eukaryotic endomembrane system. J Cell Sci 2002; 115(Pt 8):1635-42.
20. Lal K, Field MC, Carlton J et al. Identification of a very large Rab GTPase family in the parasitic protozoan Trichomonas vaginalis. Mol Biochem Parasitol; 2005; 143:226-235.
21. Dacks JB, Doolittle WF. Molecular and phylogenetic characterization of syntaxin genes from parasitic protozoa. Mol Biochem Parasitol 2004; 136(2):123-36.
22. DeLuca P, Taddei R, Varano L. Cyanidioschyzon merolae: A new alga of thermal acidic environments. Webbia 1978; 33:37-44.
23. Lujan HD, Touz MC. Protein trafficking in Giardia lamblia. Cell Microbiol 2003; 5:427-34.
24. Hehl AB, Marti M. Secretory protein trafficking in Giardia intestinalis. Mol Microbiol 2004; 53:19-28.
25. Tovar J, Leon-Avila G, Sanchez LB et al. Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. Nature 2003; 426(6963):172-6.
26. Hall BS, Pal A, Goulding D et al. Rab4 is an essential regulator of lysosomal trafficking in trypanosomes. J Biol Chem 2004; 279:45047-56.
27. Conibear E, Cleck JN, Stevens TH. Vps51p mediates the association of the GARP (Vps52/53/54) complex with the late Golgi t-SNARE Tlg1p. Mol Biol Cell 2003; 14:1610-23.
28. Collins KM, Thorngren NL, Fratti RA et al. Sec17p and HOPS, in distinct SNARE complexes, mediate SNARE complex disruption or assembly for fusion. EMBO J 2005; 24:1775-1786.
29. Van Wye J, Ghori N, Webster P et al. Identification and localization of rab6, separation of rab6 from ERD2 and implications for an 'unstacked' Golgi, in Plasmodium falciparum. Mol Biochem Parasitol 1996; 83:107-20.
30. Pelkmans L, Burli T, Zerial M et al. Caveolin-stabilized membrane domains as multifunctional transport and sorting devices in endocytic membrane traffic. Cell 2004; 118:767-80.
31. Overath P, Engstler M. Endocytosis, membrane recycling and sorting of GPI-anchored proteins: Trypanosoma brucei as a model system. Mol Microbiol 2004; 53:735-44.
32. Praefcke GJ, McMahon HT. The dynamin superfamily: Universal membrane tubulation and fission molecules? Nat Rev Mol Cell Biol 2004; 5:133-47.
33. Robinson MS. Adaptable adaptors for coated vesicles. Trends Cell Biol 2004; 14(4):167-74.
34. Motley A, Bright NA, Seaman MN et al. Clathrin-mediated endocytosis in AP-2-depleted cells. J Cell Biol 2003; 162(5):909-18.
35. Bonifacino JS, Traub LM. Signals for sorting of transmembrane proteins to endosomes and lysosomes. Annu Rev Biochem 2003; 72:395-447.
36. Ford MG, Mills IG, Peter BJ et al. Curvature of clathrin-coated pits driven by epsin. Nature 2002; 419:361-6.
37. Dupre S, Urban-Grimal D, Haguenauer-Tsapis R. Ubiquitin and endocytic internalization in yeast and animal cells. Biochim Biophys Acta 2004; 1695:89-111.
38. Legendre-Guillemin V, Wasiak S, Hussain NK et al. ENTH/ANTH proteins and clathrin-mediated membrane budding. J Cell Sci 2004; 117:9-18.
39. Stein MP, Feng Y, Cooper KL et al. Human VPS34 and p150 are Rab7 interacting partners. Traffic 2003; 4:754-71.
40. Bowers K, Lottridge J, Helliwell SB et al. Protein-protein interactions of ESCRT complexes in the yeast Saccharomyces cerevisiae. Traffic 2004; 5:194-210.
41. Robibaro B, Hoppe HC, Yang M et al. Endocytosis in different lifestyles of protozoan parasitism: Role in nutrient uptake with special reference to Toxoplasma gondii. Int J Parasitol 2001; 31:1343-53.

42. Jurgens G. Membrane trafficking in plants. Annu Rev Cell Dev Biol 2004; 20:481-504.
43. Siniossoglou S, Lutzmann M, Santos-Rosa H et al. Structure and assembly of the Nup84p complex. J Cell Biol 2000; 149:41-54.
44. Devos D, Dokudovskaya S, Alber F et al. Components of coated vesicles and nuclear pore complexes share a common molecular architecture. PLoS Biol 2004; 2:e380.
45. Harper JT, Waanders E, Keeling PJ. On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. Int J Syst Evol Microbiol 2005; 55(Pt 1):487-96.
46. Berriman M, Ghedin E, Hertz-Fowler C et al. The genome of the African trypanosome Trypanosoma brucei. Science 2005; 309(5733):416-22.
47. Lang BF, O'Kelly C, Nerad T et al. The closest unicellular relatives of animals. Curr Biol 2002; 12(20):1773-8.