Introduction

Sandra Siehler and Graeme Milligan

This book provides a comprehensive overview of recent discoveries and the current understanding in the G protein-coupled receptor (GPCR) field.

A plethora of distinct GPCRs exist on the cell surface of every cell type and generate signals inside cells to regulate key physiological events. The human genome contains between 720 and 800 GPCRs with specific tissue and subcellular expression profiles. Chapter 1 of this volume illustrates the evolutionary history of GPCRs based on genomic information available from distinct species and ancient genomic information. Many GPCRs are involved in olfactory/sensory mechanisms. Three hundred sixty-seven non-sensory human GPCRs are known or predicted to be activated by native ligands; endogenous ligands for 224 human GPCRs are described currently, but remain to be identified for 143 orphan receptors. Three hundred sixty-seven ligand-activated non-sensory GPCRs consist of 284 class A (rhodopsin-like) receptors, 50 class B (secretin-like) receptors, 17 class C (metabotropic receptor-like) receptors, and 11 belong to the atypical class of frizzled-/smoothened receptors. Polymorphisms (e.g., of β adrenoceptors, see Chapter 15) and alternative splicing (e.g., of metabotropic glutamate receptors, see Chapter 16) further increase the variety of GPCR proteins. Posttranslational modifications such as N-linked glycosylation or carboxyterminal palmitoylation can influence their function.

GPCRs are integral membrane proteins containing an extracellular amino terminus of widely varying length, seven transmembrane α-helical stretches, and an intracellular carboxy terminus. The molecular understanding of GPCRs developed with the cloning of the β_2 adrenoceptor in 1986 and appreciation that it was related to the photon receptor rhodopsin. The majority of signaling events originate at the inner face of the plasma membrane and involve transactivation of one or more members of the four G protein families (G_s, $G_{i/o}$, $G_{q/11}$, $G_{12/13}$), which link GPCRs to effector cascades. Chapter 7 explains functions of mammalian G proteins elucidated using subunit- and tissue-specific gene targeting. Besides effector cascades involving G proteins, non-G protein-mediated signaling has been described for various GPCRs. Moreover, the activity of G proteins can be regulated by non-GPCR proteins such as receptor tyrosine kinases. The activity of GPCRs is further modulated by cellular signals in an auto- and trans-regulatory fashion. GPCRs form intra- and juxtamembrane signaling complexes

comprising not only G proteins, but also other GPCRs, ion channels, membrane and cytosolic kinases and other enzymes, G protein-modulatory proteins, and interact with elements of the cell cytoskeleton. Chapters 3–6 describe homo- and hetero-oligomerization features of GPCRs including receptors for glutamate, $GABA_B$, dopamine, and chemokines. Dopamine receptors can hetero-dimerize not only with other subtypes in the same receptor family, but also with less-related GPCR members and ion channels such as NMDA or $GABA_A$ receptors. For class C receptors, which contain a large extracellular domain, oligomerization is mandatory for receptor function. For other GPCRs, oligomerization may result in altered and/or novel ligand pharmacology. Methods applied to measure GPCR complexes and oligomer signaling comprise GPCR-Gα protein fusion constructs containing either a mutated receptor or Gα mutant, and time-resolved fluorescence resonance energy transfer (TR-FRET).

Downstream of the cellular plasma membrane, the complexity of intracellular communication controlled by GPCRs increases dramatically. Ligand-activated GPCRs often internalize, which mostly causes desensitization of signaling events, although both prolonged signaling and even signaling initiated following receptor internalization have been described. Receptor hetero-oligomers can co-internalize, and activation and internalization of one partner can therefore silence the other interaction partner. Chapters 8–11 describe key signaling features of GPCRs better understood because of significant recent advancements. These include understanding of kinetics of receptor activation and signaling events studied using FRET and bioluminescent RET (BRET). Multiple related proteins control GPCR-mediated cell signaling processes. For example four RhoGTPase nucleotide exchange factors (Rho-GEFs) link $G_{12/13}$ to pathways controlling, for example, contractile complexes of the cytoskeleton, whereas nine mammalian adenylyl cyclases (ACs) are regulated by GPCRs in a receptor- and tissue-specific manner. These enzymes are integral membrane proteins directly regulated by G_s and $G_{i/o}$ proteins, although $G_{q/11}$-coupled GPCRs also influence AC activities via calcium and protein kinase C, and $G_{12/13}$ proteins were recently found to regulate AC activity as well. Arrestins are known to bind to agonist-stimulated phosphorylated GPCRs and promote endocytosis. Novel functions of arrestins include interactions with non-GPCR receptors or direct interaction with signaling proteins including, for example, the ERK MAP kinases. Modern assay technologies to assess GPCR signaling and ligand pharmacology are described in Chapter 12. Multiplexing subcellular readouts using high content screening allows the simultaneous capture of multiple signals, in both temporal and spatial fashion. The pharmacological complexity of orthosteric and allosteric GPCR ligands in the context of both receptor-G protein complexes and activation state models, is illustrated in Chapters 13 and 14. Functional selectivity of GPCR ligands due to receptor allosterism toward intracellular effector pathways contributes to the complex pharmacological nature.

Dysregulated ligand concentration, GPCR protein level, coupling, and/or signaling are implicated in and often causative for many pathophysiological conditions including central nervous system (CNS) disorders, cardiovascular and

metabolic diseases, respiratory malfunctions, gastrointestinal disorders, immune diseases, cancer, musculoskeletal pathologies, and eye illnesses. Targeting of GPCRs is hence widely utilized for therapeutic intervention using small molecule weight ligands and, increasingly, therapeutic antibodies. About 30 percent of marketed drugs target GPCRs. Pathophysiological aspects of β-adrenoceptors in cardiovascular and respiratory diseases, of metabotropic glutamate receptors in CNS disorders, of sphingosine 1-phosphate (S1P) receptors in the immune system, and of Wnt/Frizzled receptors in osteoporosis are described in Chapters 15–18. Frizzled receptors possess a GPCR-like architecture, however, their coupling to G proteins remains controversial. Drugability of GPCRs is generally high since ligand binding pockets are found in the extracellular facing segments of GPCRs, meaning that cell permeability is not a requirement. Exceptions exist regarding drugability (e.g., for many chemokine receptors as elaborated in Chapter 6), and a few unique examples for intracellular binding sites for drugs have emerged.

Despite the high drugability and importance of this target class, drug discovery technologies for GPCRs remained limited for a long time when compared to other target classes such as kinases. Integrated lead finding strategies for cytosolic kinases and intracellular parts of membrane kinases comprise biochemical, biophysical, structural, and cellular approaches, which enable a detailed understanding of mechanisms of actions of compounds. Lead finding for GPCRs, on the other hand, was so far solely based on cellular approaches using recombinant and native systems, and either intact cells or cell membranes. Reasons included the challenges of purifying GPCRs in sufficient quantities, the stability of these as isolated membrane proteins, and the lack of structural knowledge. All three issues have been tackled, and recent successes become prominent. Expression, solubilization, and purification methods of GPCRs using eukaryotic insect or mammalian cells, prokaryotic bacterial cells, or in vitro expression systems have been significantly improved. New methods are being applied to stabilize isolated membrane proteins in semi-native lipid environments like, for example, recombinant high density lipoprotein (rHDL)-membrane discs. Functional studies of isolated GPCR-G protein complexes reconstituted in rHDLs are described in Chapter 2 and deliver novel insights that cannot be obtained from cellular systems.

The first crystal structures of a non-rhodopsin GPCR were published for the human β_2 adrenoceptor in 2007 using either a T4 lysozyme fusion replacing the third intracellular loop or a Fab antibody fragment binding to the third intracellular loop, and with the receptor in complex with an inverse agonist and stabilized in a lipid environment. The T4 lysozyme approach also facilitated the identification of the crystal structure of the human A_{2A} adenosine receptor in complex with an antagonist one year later. A novel approach for receptor stabilization uses targeted amino acid mutations in order to thermostabilize the receptor, and enabled crystal structure determination of the turkey β_1 adrenoceptor in complex with an antagonist in 2008. All GPCR structures available to date are derived from class A GPCRs and resemble inactive receptor conformations. More GPCR structures are expected to become public soon and will

enable structural drug discovery approaches including fragment-based screening and ligand co-crystallizations. Stabilized purified GPCRs reconstituted in a lipid environment facilitate not only biochemical, but also biophysical methods such as surface plasmon resonance (SPR) or back-scattering interferometry (BSI) measurements. These novel advances allow confirmation of direct binding of a ligand – whether of competitive or allosteric nature – to a GPCR, and to directly study mechanisms of actions of ligands and G protein activation to determine pharmacological textures of GPCRs. This will boost further understanding of GPCR biology, biomedical research, and ultimately translation of new therapies into the clinic.

We thank all the authors for their comprehensive and professional contributions, and Amanda Smith, Katherine Tengco, Joy Mizan, Allan Ross and Monica Finley from Cambridge University Press and Newgen for assistance, final editing and formatting of the chapters, and printing of the book. From planning the outline of the book to final printing, it has been a rewarding experience. We hope the book will be exciting to read for both newcomers and professionals in the GPCR field.

PART I: ADVANCES IN GPCR PROTEIN RESEARCH

1 The evolution of the repertoire and structure of G protein-coupled receptors

Torsten Schöneberg, Kristin Schröck, Claudia Stäubert, and Andreas Russ

INTRODUCTION

With the advent of large, publicly available genomic data sets and the completion of numerous invertebrate and vertebrate genome sequences, there has been much effort to identify, count, and categorize G protein-coupled receptor (GPCR) genes.[1,2] This valuable source of large-scale genomic information also initiated attempts to identify the origin(s) and to follow the evolutionary history of these receptor genes and families. Since all recent genomes have been shaped by selective forces over millions of years, understanding structure-function relationships and the physiological relevance of individual GPCRs makes sense only in the light of evolution. Until recently, the study of natural selection has largely been restricted to comparing individual candidate genes to theoretical expectations. Genome-wide sequence and single nucleotide polymorphism (SNP) data now bring fundamental new tools to the study of natural selection. There has been much success in producing lists of candidate genes, which have potentially been under selection in vertebrate species or in specific human populations.[3–9]

Less effort has gone into a detailed characterization of the candidate genes, which comprises the elucidation of functional differences between selected and nonselected alleles, as well as their phenotypic consequences, and ultimately the identification of the nature of the selective force that produced the footprint of selection. Such further characterization creates a profound understanding of the role and consequences of selection in shaping genetic variation, thus verifying the signature of selection obtained from genome-wide data. Since GPCRs control almost every physiological process, several receptor variants are involved in adaptation to environmental changes and niches. Consistently, genomic scans for signatures of selection revealed a number of such loci containing GPCR genes. This chapter sheds light on the origin(s), rise, and fall of GPCR genes and functions, and focuses on recent advantages in elucidating selective mechanisms (still) driving this process.

GAIN AND LOSS OF GPCRs

The origin of GPCR genes

The GPCR superfamily comprises at least five structurally distinct families/subfamilies (GRAFS classification) named: **G**lutamate, **R**hodopsin, **A**dhesion, **F**rizzled/Taste2, and **S**ecretin receptor families.[2] Because there is very little sequence homology among the five families, the evolutionary origin of GPCRs and their ancestry remain a matter of debate.

The evolutionary success of the GPCR superfamily is reflected by both its presence in almost every eukaryotic organism and by its abundance in mammals, but proteins that display a seven transmembrane (7TM) topology are already present in prokaryotes. The prokaryotic light-sensitive 7TM proteins, such as proteo-, halo-, and bacteriorhodopsins, facilitate light energy harvesting in the oceans, coupled to the carbon cycle via a non-chlorophyll-based pathway. Further, there are prokaryotic sensory rhodopsins for phototaxis in halobacteria, which control the cell's swimming behavior in response to light. As in rhodopsins of bilateral animals, prokaryotic rhodopsins contain retinal covalently bound to 7TM. Moreover, 7TM proteins with a structural similarity to prokaryotic sensory rhodopsins are found in eukaryotes.[10,11] These structural and functional features shared by pro- and eukaryotic rhodopsins suggest a common ancestry. However, despite these similarities, sequence comparisons provide no convincing evidence of an evolutionary linkage between prokaryotic rhodopsins and eukaryotic G protein-coupled rhodopsins.[12] Therefore, the question about the evolutionary origin of eukaryotic GPCRs remains open. Currently, all our insights into their evolutionary history are based on the analysis of the GPCR repertoire of distantly related extant species.

Structural and functional data clearly show that **G-protein signaling via GPCRs** is present in yeast/fungi,[13] plants,[14] and primitive unicellular eukaryotes, such as the slime mold *Dictyostelium discoideum*.[15] This receptor-signaling

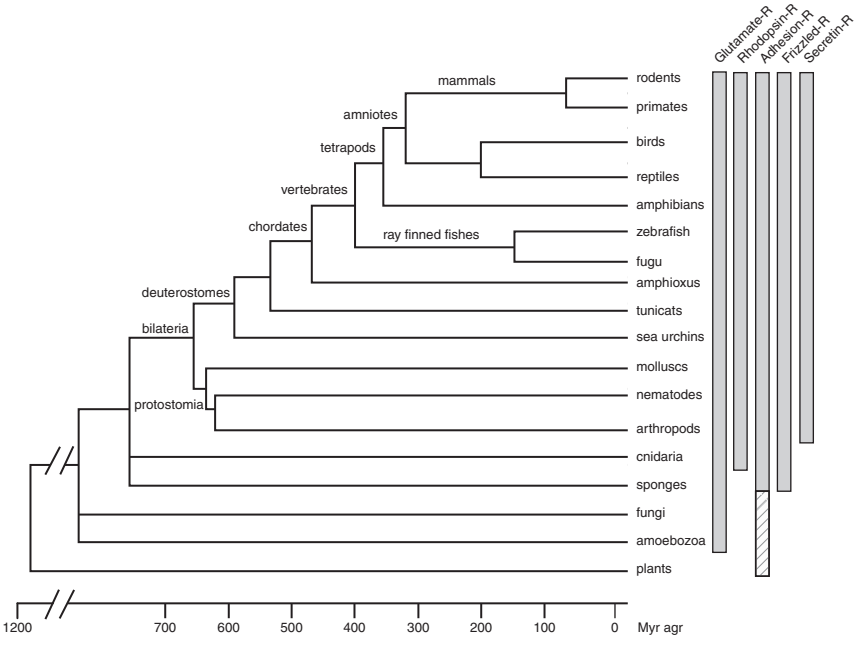

Figure 1-1: Evolutionary occurrence of the different GPCR families in eukaryotes. GPCRs and their signal transduction probably evolved ~1.2 billion years ago, before plant/fungi/animal split. Genomes of extant plants and fungi usually contain less than ten GPCR genes. The first rhodopsin-like GPCRs, which compose the main GPCR family in vertebrates, appeared ~570–700 Myr ago. Expansion of rhodopsin-like GPCRs started ~500 Myr ago, giving rise to over 1,000 members in some mammalian genomes. The relationships of some major lineages are controversially discussed, hence a very simplified phylogenetic tree of eukaryotes together with a raw time scale are shown. There is some sequence relation between adhesion receptors and GPCRs in plants and fungi, but key features of adhesion receptors, such as the GPS domain in the N terminus, are not present in plant and fungi GPCRs ([23,24]).

complex must have evolved before the plant/fungi/animal split about 1.2 billion years ago (Figure 1.1). Signal transduction through G proteins is the most prominent and eponymous feature of GPCRs. However, one has to consider that GPCRs signal not only via G proteins but also via alternative, non-G-protein-linked signaling pathways.[16] Therefore, it remains open whether G proteins were involved in GPCR signaling from the very evolutionary beginning or if the prototypes of what we now call GPCRs initially fulfilled other functions.

In contrast to GPCR signaling as such, it is more difficult to ascertain the deep evolutionary origin of the five prototypical receptor structures we know today. Genomic data and functional evidence indicate that **glutamate-receptor-like receptors** are present in *D. discoideum*[17,18] and the sponge *Geodia cydonium*,[19,20] which diverged more than 600 million years (Myr) ago (Figure 1.1). The ligand-binding domain of glutamate-receptor-like receptors, also known as the "Venus fly trap" domain, is distantly related to the prokaryotic periplasmic-binding proteins involved in amino acid and nutrient transport in bacteria.[21] Free amino acids act at glutamate-like receptors as either direct-acting orthosteric agonists or allosteric modulators of receptor activity. In contrast to *Dictyostelium*

glutamate-like receptors, the sponge receptor did show weak activation by millimolar concentrations of glutamate. This suggests that glutamate activation of glutamate-like receptors may have arisen early in metazoan evolution, with the high glutamate affinity seen in the resurrected ancestral receptor fully present by the time of the bilaterian ancestor.[22] The experimental support for glutamate-receptor-like proteins in *Dictyostelium* suggests that a prototypical receptor structure might predate the origin of metazoa.

Comparative genomic analysis also indicates that precursors of the **adhesion-GPCR** subfamily were present before the onset of metazoan evolution. Sequences distantly related to adhesion-GPCRs are found in fungi and plant genomes;[23,24] however, the homology is modest and mainly based on alignment of putative 7TM regions. Clear evidence for the ancient origin of the adhesion-GPCR subfamily comes from the analysis of the genome of a single-cell eukaryote considered to be the closest relative to metazoans, the choanoflagellate *Monosiga brevicollis*.[25,26] The *Monosiga* genome encodes proteins with the GPS (**G**PCR **p**roteolytic **s**ite)-7TM domain architecture characteristic for adhesion-GPCRs, but no clearly defined homologs of **frizzled** receptors or other elements of the *wnt* pathway. Thus, like glutamate-receptor-like proteins, the signaling module used in adhesion-GPCRs might predate the origin of metazoa.

Frizzled-like receptors are identified in sponges,[27,28] jellyfishes (Cnidaria),[29,30] and placozoa (*Trichoplax adherens*)[31] and can be interpreted as a true innovation of metazoan evolution.[30] They predate the origin of *bilateria*.

A recent phylogenetic study suggests that secretin-like GPCRs descended from the family of adhesion GPCRs.[32] Several members of the ***secretin-like receptor family*** (corticotropin-releasing factor receptor, calcitonin/calcitonin gene-related peptide receptor) are found in both deuterostomia and protostomia[33] but not in Cnidaria, *Monosiga brevicollis*, and *Dictyostelium*. This suggests an evolutionary age of more than 550 Myr, concurrent with the evolution of bilaterial animals.

The structural signatures of **rhodopsin-like GPCRs** have been found in protostome Bilateria (insects, mollusks, nematodes, vertebrates, etc.) and in jellyfishes (*Cnidaria*), which suggests that rhodopsin-like receptors appeared ~570–700 Myr ago[34–37] (Figure 1.1). Within the rhodopsin-like GPCRs, glycoprotein hormone receptors and serotonin receptors appear to be among the oldest, as suggested by their presence in sea anemone (*Cnidaria*), planarians, and nematodes.[38–40] Rhodopsin is the name-giving GPCR in this family. Recent reports suggested that opsins diverged at least before the deuterostome-protostome split about 550 Myr ago.[41] There is growing evidence that opsins are abundant in jellyfishes, indicating that prototypical opsins may have existed before divergence of Cnidaria and Bilateria about 570 to 700 Myr ago,[42,43] although still long after GPCR and G-protein signaling evolved. Given that G-protein coupling to 7TM proteins evolved before the plant/fungi/animal split (about 1.2 billion years ago), and that the first rhodopsin-like receptors appeared early in metazoan evolution, one must consider the retinal-based photosensory system to be a "reinvention" (convergent evolutionary model).

Expansion of GPCR genes

The recent completion of many vertebrate and nonvertebrate genome projects has enabled us to obtain a complete inventory of GPCRs in these species and, by a comparative genomics approach, to analyze the evolution of the GPCR subfamilies. Comparison of the repertoires of GPCRs in insects (fly, mosquito, beetle) and protochordate (Ciona) to that in vertebrates (mammals, birds, fish) reveals a high level of orthology. This indicates that nonvertebrates contain the basic ancestral complement of vertebrate GPCR genes.[44–50] However, the number of GPCRs in most sequences of nonvertebrate genomes (exceptions are the chemokine receptors in worms) is substantially lower than that in vertebrate genomes.[2] This is especially true for rhodopsin-like GPCRs that constitute the most abundant GPCR family in vertebrates when compared to nonvertebrates. Most modern rhodopsin-like GPCR subfamilies expanded in the very early vertebrate evolution. Still, many nonvertebrate GPCR clusters evolved about 500 Myr ago during a time called the "Cambrian Explosion." There are interesting theories about what triggered the enormous gain of species, functionalities, and genes. It was proposed that vision triggered the Cambrian Explosion by creating a new world of organismal interactions, the evolutionary consequence of which was a race in the invention of attracting, attacking, and defending mechanisms.[43] Concomitantly, duplication events of gene and genomic blocks, including several genes of the phototransduction, were traced to the very early vertebrate evolution.

Processes of creating new genes using preexisting genes as raw materials are well characterized, such as exon shuffling, gene duplication, retroposition, gene fusion, and fission. GPCR gene expansion in vertebrates is mainly the result of a combination of species-specific gene duplications and gene or genome duplication events. Two rounds of whole-genome duplications are thought to have played an important role in the establishment of gene repertoires in vertebrates.[51] These events occurred during chordate evolution, after the split of the urochordate and cephalochordate lineages but before the radiation of extant jawed vertebrates.[52] Whole-genome duplications can contribute to prompt gene multiplication and may trigger evolutionary adaptation. One copy or even both members of a gene pair may mutate and acquire unique functionality without risking the fitness of the organism, which is ensured by the homolog. Further, gene duplications often retained overlapping expression patterns and preserved partial-to-complete redundancy consistent with a role in boosting robustness or gene doses. On the other hand, if not advantageous, continuous accumulation of mutations (neutral evolution) will eliminate one of the duplicated genes. As for other genes, disadvantageous mutations in GPCRs are removed from a population through purifying selection. Therefore, many evolutionarily old GPCR genes, including the rhodopsins, display strong features of purifying selection.[5,53,54] It is of interest that specifically GPCRs were retained in vertebrate genomes after genome duplications.[55]

Current evidence suggests that an additional whole-genome duplication occurred in the teleost lineage after it split from the tetrapod lineage, and that only a subset of the duplicates have been retained in modern teleost genomes.[56] Support of these findings comes from sequence analysis of coelacanth, one of the nearest living relatives of tetrapods. The two modern coelacanth species that are known, *Latimeria chalumnae* and *Latimeria menadoensis*, are remarkably similar to their fossil relatives, showing little morphological change over 360 Myr. Genomic sequence analyses show no evidence of whole-genome duplication, consistent with the explanation that the coelacanth genome has not experienced a recent polyploidization event.[57] Therefore, whole-genome duplications did not contribute to GPCR expansion in tetrapode evolution. However, polyploidy is common in fishes and has been determined in sturgeons up to ploidy levels of 8n.[58] GPCR duplications due to polyploidization in ancient teleost fishes have been suggested, for example, trace amine-associated (TAAR) and (purinergic) $P2Y_{12}$-like receptors.[53,59,60]

Whole-chromosome duplications have been made responsible for parallel duplications of more related GPCR. For example, parts of chromosome 4 and 5 show a very similar order of paralog receptor genes. It is assumed that a chromosomal duplication gave rise to dopaminergic receptor paralogs, DRD5 and DRD1, and adrenoceptor paralogs, ADRA2C and ADRA1, at chromosomes 4 and 5, respectively.[61]

Several GPCR subfamilies, such as olfactory,[62–64] chemokine,[65] aminergic,[66] TAAR,[59,67], vomeronasal,[68] and nucleotide receptor-like receptors,[60] cluster in vertebrate genomes and are often arranged in a tandem-like fashion. The numbers of functional receptor genes and pseudogenes of these GPCR subfamilies vary enormously among the genomes of different vertebrate species. Much of the variation in these receptor repertoires can probably be explained by the adaptation of species to different environments. For example, the platypus, a semiaquatic monotreme, has the largest repertoire of vomeronasal receptors in all vertebrates surveyed to date, with more than 300 intact genes and 600 pseudogenes in this chemosensory receptor family.[68] However, it seems that a substantial portion of variety is generated by genomic drift, which probably also has an important role in both adaptive and nonadaptive evolution.[69,70]

The molecular mechanisms of gene amplification and genomic clustering are extensively studied in prokaryotes.[71] Here, gene duplication can occur during genome replication due to an unequal sister strand exchange, producing two adjacent identical copies of a region (the amplicon) that can undergo homologous recombination. Another mechanism, termed circle-excision and reinsertion, involves creation of a circular DNA molecule and its subsequent recombination with another DNA molecule to form the duplication. Alternatively, a rolling-circle mechanism may account for instances of very rapid gene amplification. There is evidence that similar mechanisms are responsible for gene amplifications in eukaryotes.[72] However, the precise molecular mechanisms for odorant GPCR cluster formation, for example, are not yet known.