Deducing the origin of soluble adenylyl cyclase, a gene lost in multiple lineages

Roelofs, Jeroen; Haastert, Peter J.M. van

Deducing the Origin of Soluble Adenylyl Cyclase, a Gene Lost in Multiple Lineages

Jeroen Roelofs, and Peter J. M. Van Haastert

GBB, Department of Biochemistry, University of Groningen, Nijenborgh, The Netherlands

The family of eukaryotic adenylyl cyclases consists of a very large group of 12 transmembrane adenylyl cyclases and a very small group of soluble adenylyl cyclase (sAC). Orthologs of human sAC are present in rat *Dictyostelium* and bacteria but absent from the completely sequenced genomes of *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, and *Saccharomyces cereviciae*. sAC consists of two cyclase domains and a long ~1,000 amino acid C-terminal (sCKH) region. This sCKH region and one cyclase domain have been found in only four bacterial genes; the sCKH region was also detected in bacterial Lux-transcription factors and in complex bacterial and fungal kinases. The phylogenies of the kinase and cyclase domains are identical to the phylogeny of the corresponding sCKH domain, suggesting that the sCKH region fused with the other domains early during evolution in bacteria. The amino acid sequences of sAC proteins yield divergence times from the human lineage for rat and *Dictyostelium* that are close to the reported divergence times of many other proteins in these species. The combined results suggest that the sCKH region was fused with one cyclase domain in bacteria, and a second cyclase domain was added in bacteria or early eukaryotes. The sAC was retained in a few bacteria and throughout the entire evolution of the human lineage but lost independently from many bacteria and from the lineages of plants, yeast, worms, and flies. We conclude that within the family of adenylyl cyclases, soluble AC was poorly fixed during evolution, whereas membrane-bound AC has expanded to form the subgroups of prevailing adenylyl and guanylyl cyclases.

Introduction

New genes may appear in a genome either by horizontal transfer from other organisms or by reshuffling and amplification of parts of the endogenous genome. Such a new gene may become fixed in the organism or the gene may get lost somewhere during evolution. The phylogeny of many genes is relatively straightforward because they are well fixed and represented in many species. For genes that are present in only a few species, it is a priori more difficult to discriminate between horizontal gene transfer to multiple species and extensive loss of the gene in many lineages.

The completion of sequencing of the genomes of an expanding number of organisms, ranging from bacteria to human, has resulted in a better insight into the role of horizontal gene transfer and gene loss during evolution (Koonin et al. 1997; Aravind et al. 1998; Andersson and Andersson 1999; Nelson et al. 1999; Aravind et al. 2000). Among prokaryotes, horizontal gene transfer as well as gene loss seem to have occurred frequently and are expected to be major contributors to the diversity among prokaryotes (Ochman and Moran 2001). For multicellular eukaryotes, the chance of a horizontal gene transfer event seems low because this transfer should occur in the germ line to be transferred to a future generation. For these organisms, gene duplication and the combination of protein domain–encoding sequences seems to be the main path to obtain new genes.

From an analysis of the recently published human genome, 113 genes were identified that were proposed to be most likely derived from bacteria by horizontal gene transfer to the vertebrate lineage (Lander et al. 2001) because orthologs are present in several bacteria but absent from *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, and *Saccharomyces cereviciae*. However, a more comprehensive analysis suggests that a significant portion of the cases can be explained by gene loss (Roelofs and Van Haastert 2001*c*; Salzberg et al. 2001; Stanhope et al. 2001). One of these genes is a soluble adenylyl cyclase (sAC) (Buck et al. 1999) that belongs to the class-III ACs (Danchin 1993). These ACs form an interesting phylogenetic family of genes because in eukaryotes it consists of two main groups, a very large group of prevailing cyclases present in nearly all eukaryotes encoding membrane-bound ACs and GCs and soluble GCs and a very small group of genes related to sAC present in only a few bacteria, *Dictyostelium*, and several vertebrates. In the present study, we investigate the origin of the small family of genes related to sAC.

The three-dimensional structure of mammalian AC (Tesmer et al. 1997; Zhang et al. 1997) reveals that the catalytic core is formed by two cyclase domains that are associated in an antiparallel manner. This dimer can be formed by two identical as well as two different domains. For the heterodimer, these domains can be derived from one polypeptide or two different polypeptides. In metazoa, most AC enzymes consist of 12 transmembrane segments and two cyclase domains, C1 and C2; these enzymes are regulated by G-proteins (Hanoune and Defer 2001). Two forms of GC enzymes are present; a soluble enzyme composed of two different subunits and a membrane-bound enzyme with one transmembrane segment and one catalytic domain that functions as the homodimer (Wedel and Garbers 2001). sAC

is present in human and rat and contains two cyclase domains and a ~1,000 amino acid long C-terminal region (Buck et al. 1999). The enzyme is regulated by bicarbonate and involved in sperm maturation (Chen et al. 2000). The cyclase domains of sAC share the highest degree of identity with the cyclase domains of type-III bacterial cyclases and far less identity with the cyclase domains of other vertebrate adenylyl or guanylyl cyclases. Recently we identified a putative ortholog of sAC in the eukaryotic microorganism *Dictyostelium* (Roelofs et al. 2001*b*), which is, however, a guanylyl cyclase (sGC). It shares significant sequence identity with the corresponding cyclase domains and the ~1,000 amino acid C-terminal region of mammalian sAC, indicating common ancestry. To trace the possible evolutionary origin of human sAC and *Dictyostelium* sGC, we have analyzed the protein and nucleotide databases for genes that share significant identity with segments of human sAC and *Dictyostelium* sGC. The results suggest that in bacteria a protein segment of ~1,000 amino acids was fused with a cyclase domain. In bacteria, the 1,000 amino acid segment was shortened, whereas in the eukaryotic lineage a second cyclase domain was added. The gene was retained in the eukaryotic lineage up to human and rat but was apparently lost independently in several other lineages.

Materials and Methods

The database searches were carried out by using the BLASTP or TBLASTN program in the nonredundant protein and nucleotide sequence database at NCBI (http://www.ncbi.nlm.nih.gov), the EST_others databases at NCBI, the microbial genomes at NCBI, the other eukaryotes at NCBI, and several specific databases for *Drosophila* (http://www.fruitfly.org), *C. elegans* (http://www.wormbase.org), human (http://www.ncbi.nlm.nih.gov/genome/seq/HsBlast.html and http://publication.celera.com.), yeast (http://www.ncbi.nlm.nih.gov), *Plasmodium* (http://www.ncbi.nlm.nih.gov/Malaria/plasmodiumbl.html), *Dictyostelium* (http://www.sdsc.edu/mpr/dicty/), and *Arabidopsis* (http://www.arabidopsis.org). The sequences and analyses reported here are based on searches performed in June 2000. After the submission of the manuscript, searches were repeated in October 2001 using the databases described previously and in the database for *Giardia lamblia* (http://www.mbl.edu/Giardia/blast.html). No additional sequences were retrieved, except for a fragment of sAC1 in *Sus scrofa* (pig) and a sCKH region in bacteria (*Nostoc punctiforme*, *Chloroflexus aurantiacus*, and *Rhodopseudomonas palustris*).

The SMART program (Schultz et al. 1998, 2000) was used to analyze the domain structure of the retrieved sequences. Multiple sequence alignments were constructed using the CLUSTAL W program (Thompson, Higgins, and Gibson 1994), followed by manual optimization. In DdsGC, three short (~18 amino acids) and three long (39–114 amino acids) stretches of repetitive sequence within the sCKH region were deleted; repetitive sequences are often found in *Dictyostelium* proteins, even within very conserved domains. Distance matrices

were constructed from the alignments with the PROTDIST program of the PHYLIP package, which uses the Dayhoff's PAM 001 matrix for the calculation of evolutionary distances (PHYLIP 3.5, J. Felsenstein 1993 [Felsenstein 1996]). Phylogenetic trees were generated by using the FITCH program of the PHYLIP package, with 250 bootstrap replications to assess the reliability of the nodes. Tree topologies obtained by FITCH were confirmed using Neighbor-Joining and Protein Parsimony from the same PHYLIP package.

For the calculation of the divergence time (Nei, Xu, and Glazko 2001), linearized forms of the trees, as presented in figures 2 and 3, were used for the C1 and C2 cyclase domains and the first and second half of the sCKH region; bacterial sequences were AsCYAA for cyclase domains and AsKin for sCKH regions. The evolutionary distance between human and bacterial orthologs of a protein should be between the origin of eubacteria, about 3,500 Myr, and endosymbiotic acquisition of organelles, about 2,000 Myr (Doolittle et al. 1996; Feng, Cho, and Doolittle 1997). The absolute divergence time of the bacteria-eukaryote separation is the subject of substantial debate because fossil-based records are missing. Therefore, we use a relative time for the origin of eukaryotes. The divergence times of *Dictyostelium* and rat from the human sAC were calculated as described (Nei, Xu, and Glazko 2001) by using a relative distance of 1.00 from bacteria to human.

The amino acid sequence of human sAC2 was deduced from the genomic contigs AL133404 and AL031778 and from cDNA EST sequences 58262 and 58263. Gene finder programs provided about 65% of the expected ORF using sAC1 as reference. Manual inspection of the translated contigs allowed further identification of about 80% of the putative protein sequence.

Sequences used are: RnACV, NP072122; RnACII, P26769; RnGCE, P51840; RnGCα2, NP076446; HssAC1, NP060887; RnsAC1, NP067716; DdsGC, AF361947; AsCYAA, BAA13997; MlsAC, CAA19149; SmsAC, S60684; MlosAC1, AP003001; MlosAC2, AP002995; AsKin, AF230361; BrKin, AF222754; SpKin1, CAA20836; SpKin2, CAB11683; CaKin, AAC39451; ReKin, BAA34261; ScLux1, CAB37582; ScLux2, CAB93733; ReLux, AAD28307. Abbreviations of species used in sequence names are: Hs, *Homo sapiens*; Rn, *Rattus norvegicus*; Dd, *Dictyostelium discoideum*; As, *Anabaena* sp.; Ml, *Mycobacterium leprae*; Sm, *Sinorhizobium meliloti*; Mlo, *Mesorhizobium loti*; Sp, *Schizosaccharomyces pombe*; Ca, *Candida albicans*; Sc, *Streptomyces coelicolor*; Br, *Bradyrhizobium* sp.; and Re, *Rhodococcus erythropolis*.

Results
Signatures of sAC in Other Organisms

Human sAC1 is composed of two cyclase domains of ~225 amino acids each and a long C-terminal region of about 1,000 amino acids (fig. 1); a sequence in rat shares a high degree of identity (77%), indicating that it encodes the rat ortholog of human sAC1 (Buck et al. 1999). Using the sequence of human sAC1 as an input,
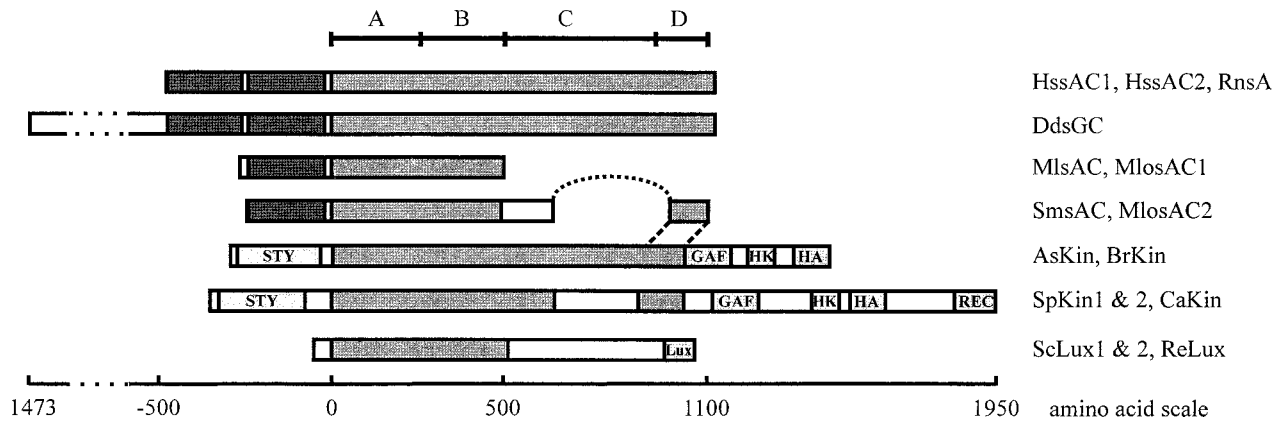
Fig. 1.—Schematic representation of domain composition of the different proteins used in this study. Sequence alignment suggests that the sCKH region may be divided in four segments, as indicated by the letters A–D above the alignment. The dark gray region represents the cyclase domains, intermediate gray refers to the sCKH region, and light gray region shows the other domains, STY: serine-threonine-tyrosine kinase domain; GAF: GAF domain; HK: Histidine kinase domain; HA: H-type ATPase; Rec: receiver domain.

a second sAC gene was identified in human sequence databases. From genomic and cDNA sequences, a partial amino acid sequence could be deduced showing about 65% similarity with sAC1 in the catalytic domains and 55% in the C-terminal segment; the gene was addressed as sAC2.

Recently, we identified in the eukaryotic microorganism *Dictyostelium*, the gene DdsGC that encodes a guanylyl cyclase and is an ortholog of sAC (Roelofs et al. 2001*b*). The protein has an N-terminal extension of about 1,000 amino acids when compared with sAC (fig. 1). Searching several eukaryotic databases (see *Material and Methods*) with sAC1, sAC2, and DdsGC protein sequences reveals neither another gene encoding a cyclase domain and a C-terminal region nor another gene encoding a cyclase domain with BLAST expectation values lower than the values for bacterial cyclase domains. Bacterial databases reveal four genes that posses one cyclase domain and a C-terminal extension with significant sequence identity with the C-terminal region of sAC: MlsAC from *Mycobacterium leprae*, SmsAC from *Sinorhizobium meliloti*, and two genes from *Mesorhizobium loti*, MlosAC1 and MlosAC2 (fig. 1*A*). The C-terminal regions of these bacterial cyclases are shorter than that of human sAC (see subsequently).

The amino acid sequences of the C-terminal region of HssAC1, DdsGC, and MlsAC were used to search for genes that could provide information on the origin of this C-terminal region. Two bacterial genes, AsKin and BrKin, that show significant sequence identity over the entire length of the cyclase C-terminal region were identified (fig. 1). These bacterial proteins are complex kinases with an N-terminal serine-threonine-tyrosine kinase and a C-terminal extension with a GAF domain, a histidine kinase, and a H-ATPase. We therefore addressed the C-terminal region as soluble Cyclase-Kinase Homology (sCKH) region. Further database searches with the sCKH region of HssAC and AsKin uncovered several bacterial sequences with an N-terminal kinase and a part of the sCKH region (e.g., ReKin; data not shown) and complex yeast kinases with a similar topology as AsKin, except for an additional C-terminal his-

tidine kinase receiver (Rec) domain (fig. 1). Finally, a group of bacterial transcription factors containing the sCKH region and a C-terminal Lux-domain were identified. Subsequent database searches with complete or partial sequences of the Lux-containing proteins or the yeast kinases did not reveal new sequences in bacteria or eukaryotes that share significant sequence identity with soluble cyclases.

The sCKH Region

On the basis of sequence alignment, the sCKH region can be divided into four tentative segments A–D (fig. 1). When no indications for a specific domain identity can be given, we prefer to use the neutral terms "region" and "segment" for longer and shorter sequences, respectively. The first segment, A, of the sCKH region is about 260 amino acids in length and clearly encodes an AAA ATPase domain with a P-loop motif found in many ATP and GTP-binding proteins. This domain is observed in all proteins with the sCKH region. The second segment, B, also present in all the proteins, is about 250 amino acids in size but has no domain identity in databases. The third segment, C, is the least conserved, even in relative closely related proteins such as HssAC1 and RnsAC1. The C and D segments of the sCKH region, about 400 and 150 amino acids, respectively, are present in a subset of the proteins. MlsAC and MlosAC1 only comprise segments A and B, whereas SmsAC and MlosAC2 also comprise segment D, but lack most of segment C. The yeast kinases and the bacterial Lux-transcription factors do have amino acid sequence corresponding to the length of segments C and D, but these show little sequence identity with the sCKH region of sAC1 (with the exception of segment D of SpKin2).

Phylogeny of the sCKH Region

To deduce the evolutionary origin of human sAC, we first analyzed the phylogeny of the sCKH regions from the different groups of proteins and then compared the observed phylogenetic relations with those deduced
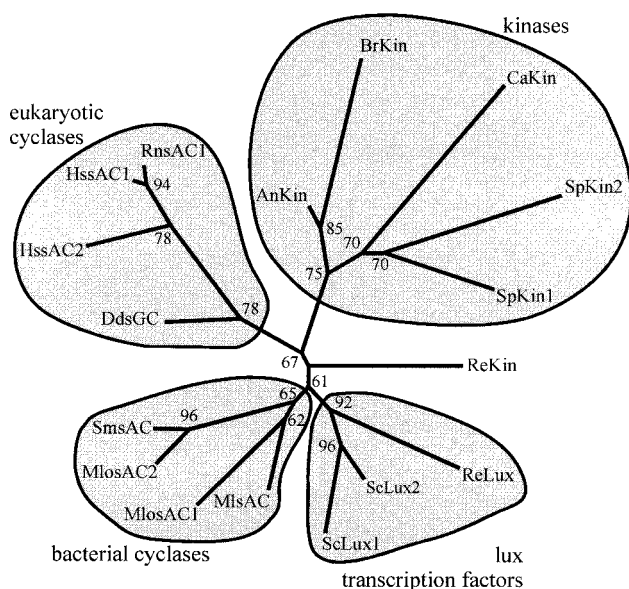
FIG. 2.—Phylogeny of the sCKH region. The segments A and B of the sCKH region (see fig. 1) were aligned using CLUSTAL W and manual optimization. The Fitch program of the PHYLIP package was used to make the phylogenetic tree. The numbers indicate bootstrap values using ReKin as outgroup. See *Materials and Methods* for protein abbreviations and sequence accession numbers.
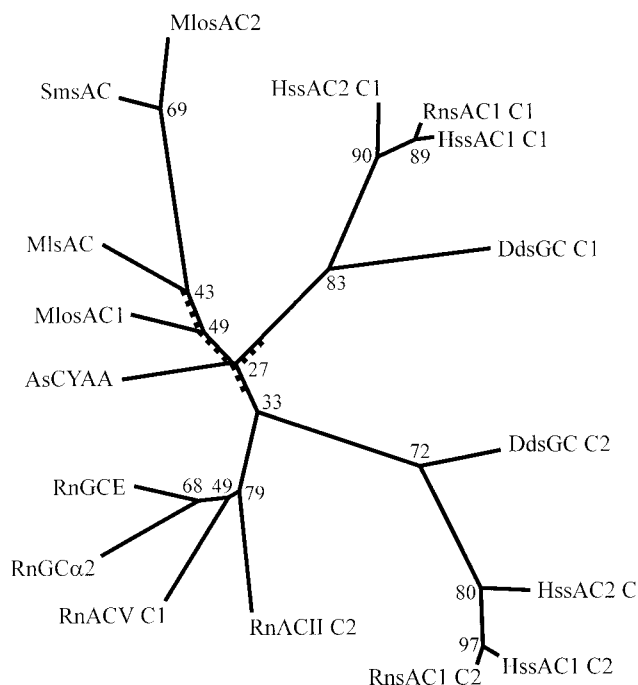

FIG. 3.—Phylogeny of the cyclase domains. The cyclase domains of bacterial and eukaryotic sAC and the prevailing eukaryotic cyclases were aligned using CLUSTAL W and manual optimization. The Fitch program of the PHYLIP package was used to make the phylogenetic tree. The numbers indicate bootstrap values using AsCYAA as outgroup. The dotted line indicates the region of the tree where other class-III bacterial cyclases branched in analysis where they were included. This indicates that the origin is not well defined as is also indicated by the low bootstrap values in this area of the tree.

from the cyclase and kinase domains. Because only the first half of the sCKH region is well conserved among all proteins, this segment was used to construct a phylogenetic tree (fig. 2). Bootstrap analyses suggest four sister groups: bacterial cyclases, eukaryotic cyclases, transcription factors, and kinases. Within these sister groups branching is strongly supported (bootstrap values above 85); the positions of the four sister groups relative to each other is less well resolved (bootstrap above 50). In the group of sCKH regions from bacterial cyclases, MlosAC1 groups with MlsAC, and MlosAC2 with SmsAC, which is in agreement with the classification based upon the overall topology of these four proteins (fig. 1). The sCKH regions from eukaryotic cyclases form a monophyletic group, with human and rat sAC1 in the crown; human sAC2 apparently arose before the division of human and rat and the *Dictyostelium* gene from a branching halfway between vertebrates and the presumed bacterial origin. The third sister group of sCKH regions is formed by the Lux-containing transcription factors, whereas the fourth sister group consists of all kinases, with the yeast kinases in the crown and the bacterial kinases together closer to the root. In this analysis, the cyclase and kinase domains were not included; nevertheless, the phylogeny of the first half of the sCKH region places all sequences in accordance with the domains found outside the sCKH region.

The phylogeny of the N-terminal kinase domain and the C-terminal GAF/H-ATPase region were analyzed separately by bootstrap analysis to investigate whether the evolutionary relations seen in the sCKH region are also present in other domains of these proteins. The phylogeny of the STY-kinase domains of the yeast and bacterial proteins is weakly resolved (some boot-

strap values are below 40; data not shown). In contrast, the GAF/H-ATPase regions are well resolved (all bootstap values above 98), revealing coevolution of sCKH region and the GAF/H-ATPase domains (data not shown). This strongly suggests that the sCKH region was combined in bacteria with the N-terminal STY kinase domain and the C-terminal GAF, H kinase, and H-ATPase domains, and that the Rec domain was added to this sequence in the yeast lineage.

Phylogeny of the Cyclase Domain

The phylogeny of the cyclase domains (fig. 3) suggests the presence of four sister groups: one group of bacterial sAC enzymes and three eukaryotic cyclase groups. Other bacterial cyclases (only AsCYAA is included in the fig. 3) are positioned close to the region of the tree where the four sister groups come together. The three groups of eukaryotic cyclases are characterized by the C1 domains of sAC (including the C1 of DdsGC), the C2 domains of sAC (including the C2 of DdsGC), and all other vertebrate cyclases; this third group contains many members (all 12 transmembrane AC and all animal guanylyl cyclases) and will be referred to in this article as the prevailing eukaryotic cyclases. Bootstrap analysis demonstrates that the three eukaryotic sister groups are well resolved. The positioning of the bacterial sAC cyclase domains into one group is not well supported, which is also the case for the posi-

Table 1
Estimates of Relative Divergence Times of *Dictyostelium* sGC and Rat sAC1 from the Human Lineage

	RELATIVE DIVERGENCE TIME	
	Dictyostelium (DdsGC)	Rat (RnsAC1)
C1 domain...................................	0.523	0.022
C2 domain...................................	0.454	0.022
sCKH region a + b	0.554	0.057
sCKH region c + d	0.583	0.037
Average ± SD	0.528 ± 0.055	0.035 ± 0.016
Reported distance (Nei, Xu, and Glazko 2001).....	0.464 ± 0.084	0.040 ± 0.012

NOTE.—The divergence times are shown as values relative to the bacteria-human divergence time, which is set at 1.00. Linearized forms of the trees, as presented in figures 2 and 3, were used for the C1 and C2 cyclase domains and the first and second half of sCKH region; bacterial sequences were AsCYAA for cyclase domains and AsKin for sCKH regions. The reported distance for *Dictyostelium* was deduced from the phylogeny of many proteins. This revealed that *Dictyostelium* separated from the human lineage after the plants but before the fungi (Baldauf et al., 2000); the relative divergence time of plants and fungi was estimated based on data in Nei, Xu and Glazko (2001).

tioning of this group relative to the eukaryotic cyclases. In some analysis, members of the group of bacterial sAC associate with the C1 domains of eukaryotic soluble cyclases and in other experiments with the C2 domains. Within the group of bacterial cyclases, the cyclase domains of MlosAC1 and MlsAc are always positioned close together and that of MlosAC2 with SmsAC, as was observed for the sCKH regions.

Evolutionary Age of Human and *Dictyostelium* Soluble Cyclase

The results of figures 2 and 3 indicate that the C1 domain, the C2 domain, and the sCKH region of eu-
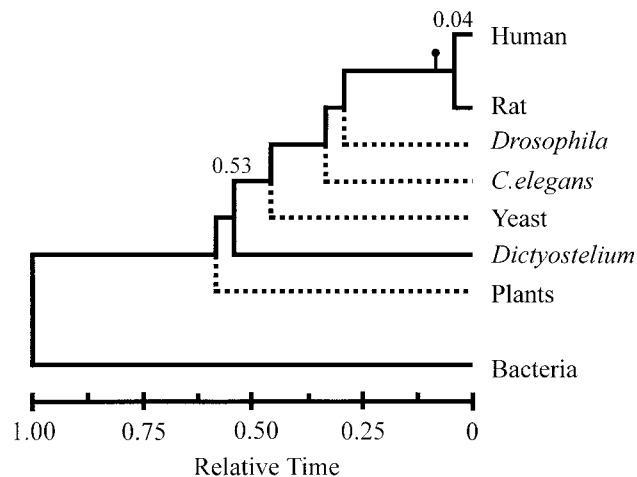

FIG. 4.—Schematic representation of the phylogeny of species (dashed line) and sAC protein (solid line). The phylogeny of species was derived from the estimated diversion time of several proteins in human, rat, *Drosophila*, *C. elegans*, yeast, plants, and bacteria (Nei, Xu, and Glazko 2001). The relative distance from human to bacteria is set at 1.00. The position of *Dictyostelium* relative the other species was derived from a large phylogenetic study of four proteins in the eukaryotic kingdom (Baldauf et al. 2000). The divergence time of sGC in *Dictyostelium* and sAC1 in rat were derived from table 1. The tentative divergence time of human sAC2 (0.095 ± 0.005) is indicated (🍄). The data suggest the vertical gene transfer of human sAC1 and consequently the loss of the sAC gene in plants, yeast, *C. elegans*, and *Drosophila*.

karyotic sAC enzymes all have the same monophyletic origin, with *Dictyostelium* as early diverging taxon. To further investigate the evolution of these genes, we may estimate the divergence times of *Dictyostelium* sGC and rat sAC1 from the human sAC1 and compare these with the reported divergence times of other proteins in these species. Because the absolute divergence time of eukaryotes is not known, we use a relative distance from human to bacteria sAC of 1.00 and calculate the divergence times of *Dictyostelium* sGC and the rat sAC1 relative to this distance. These relative divergence times are then compared to the divergence times of many other proteins in these species (Baldauf et al. 2000; Nei, Xu, and Glazko 2001). For the C1 and C2 cyclase domains (with AsCYAA as bacterial root), we used a linear form of the tree presented in figure 2 and for the sCKH region (with AnKin as bacterial root), a linearized tree of figure 3. The results (table 1) suggest a relative divergence time of about 0.53 ± 0.06 for *Dictyostelium* sGC and a divergence time of 0.04 ± 0.02 for rat sAC1, relative to the distance of sAC from human to bacteria. This relative divergence time for *Dictyostelium* sGC corresponds very well with the reported position of *Dictyostelium* in the phylogenetic tree of eukaryotes (see fig. 4), with divergence from the animal lineage before the fungi but after the plants; using several proteins, this plant-fungi diversion time has been estimated to be about 0.46 ± 0.08 on the relative scale from human to bacteria (Nei, Xu, and Glazko 2001). Also the rat-human divergence time for sAC1 corresponds well with the reported time for the divergence of the species (Nei, Xu, and Glazko 2001). Thus, the divergence times of *Dictyostelium* sGC and rat sAC1 from the human sAC is essentially identical to the divergence times of several other proteins in these species, indicating that sAC-sGC follows the same evolutionary trait as many other proteins in these organisms.

Discussion

It has been proposed that human sAC is an example of horizontal gene transfer from bacteria to the verte-

brate lineage. No orthologs of sAC have been found in the completely sequenced genomes of *Drosophila*, *C. elegans*, *Arabidopsis*, and *S. cereviciae*. Because bacterial cyclases are the closest homologs of sAC, the hypothesis of a recent horizontal gene transfer seemed a more likely explanation for the absence of the gene in several eukaryotes than the normal generation-to-generation vertical-transmission route in which case sAC would have to be lost or mutated beyond recognition in multiple lineages (yeast, worm, flies, and plants) (Lander et al. 2001; Ponting 2001). However, the unambiguous finding of the ortholog of human sAC in *Dictyostelium* both demands and allows for a more refined analysis. *Dictyostelium* diverged from the vertebrate lineage before yeast but after plants, which is about halfway the evolutionary distance from bacteria to mammals (Baldauf et al. 2000; Nei, Xu, and Glazko 2001).

Deducing the origin of a complex protein such as sAC with sequence information from only a few organisms is difficult but might be possible when information on amino acid sequence comparison and domain composition are combined. Proteins of the class of sAC enzymes contain one or two cyclase domains and a long C-terminal sCKH region that is found in a limited set of other proteins. The phylogeny of the sCKH region points to four sister groups: bacterial cyclases, eukaryotic soluble cyclases, transcription factors, and complex kinases. For all proteins, the phylogeny of the sCKH region is convergent with the phylogeny of the other domains. This suggests that an sCKH region was fused with other domains early during evolution, leading to the four groups of proteins, as we recognize them presently (figs. 1 and 2).

The coevolution of the sCKH region and cyclase domains strongly suggests that the sCKH region and at least one cyclase domain were combined in bacteria. From this ancestor sAC enzyme, the four presently known bacterial sAC enzymes were likely derived by deletion of the second half of the sCKH region in MlsAC and MlosAC1 and by the deletion of an internal segment of sCKH in SmsAC and MlosAC2 (fig. 1). In eukaryotic sAC enzymes, a second cyclase domain has been added. This occurred very early during evolution, possibly in bacteria or early eukaryotes, because no common phylogenetic trait can be identified for the soluble cyclase C1 and C2 domains in eukaryotes. In the group of prevailing eukaryotic ACs, which also has two cyclase domains, the C1 and C2 domains appear to have diverged much later during eukaryotic evolution (see fig. 3).

The transition of the bacterial ancestor sAC to the eukaryotic lineage could be either by generation-to-generation gene transfer (vertical) or by species-to-species gene transfer (horizontal), with important implications for each hypothesis. Figure 4 presents our current knowledge on the phylogeny of eukaryotic species (dotted lines) combined with phylogeny of sAC enzymes (thick lines). The normal vertical gene transfer of sAC implies that the gene was present during all the lineage deviations up to human and thus has been lost independently in several lineages after their separation from the

lineage leading to human. On the other hand, horizontal gene transfer should explain how *Dictyostelium* and the vertebrate lineage received the gene. One explanation would be two events of horizontal gene transfer from bacteria to *Dictyostelium* and to vertebrates, respectively. Another explanation would be normal evolution to the *Dictyostelium* lineage (which implies loss in plants), loss in the fungi-animal clade, and horizontal gene transfer from *Dictyostelium* or bacteria to the vertebrate lineage.

The hypothesis that genes have entered the eukaryotic cell by horizontal gene transfer in the period of the earliest eukaryotes is widely accepted (Doolittle 1998; Gray 1999). Two arguments are used for proposing a more recent transfer of a gene from bacteria to vertebrates (Wolf, Kondrashov, and Koonin 2000; Lander et al. 2001; Ponting 2001). First, the amino acid sequence encoded by a eukaryotic candidate gene shows considerably higher sequence similarities with bacterial proteins than with proteins in closer related organisms. A second argument is a limited presence of the gene in eukaryotes and a widespread presence of the bacterial counterpart over different bacterial lineages.

The following lines of evidence suggest the vertical gene transfer of human sAC. First, sAC1, sAC2, and DdsGC show a long monophyletic origin. The phylogenetic distance of human sAC to bacteria is even longer than the distance of the large group of prevailing human cyclases to bacterial cyclases. Thus, there are no indications of a recent or even ancient horizontal transfer of the gene. Second, the estimated relative divergence times of *Dictyostelium* sGC and rat sAC on a relative bacteria-human time scale are not significantly different from the proposed divergence times of the species, as deduced from the phylogeny of many proteins (table 1 [Baldauf et al. 2000; Nei, Xu, and Glazko 2001]). This strongly suggests a normal evolutionary trait of sAC from bacteria via eukaryotic microorganisms to vertebrates, which implies the loss of the gene in multiple lineages. The family of sAC genes is an example of a dozen cases in which *Dictyostelium* forms the only phylogenetic connection identified so far between bacteria and vertebrates (Roelofs and Van Haastert 2001*c*).

The difference in evolutionary success between the subclasses of cyclases is intriguing. Although the sAC branch is almost completely lost in many species, the group of prevailing eukaryotic cyclases is extensively used in several organisms. Vertebrates have nine subtypes of the 12 transmembrane ACs (Hanoune and Defer 2001). Also membrane GCs and soluble GCs have expanded in animals. In *C. elegans* the membrane GC family contains even more than 29 members (Yu et al. 1997). Possibly the genes of the sAC group have been fixed only in organisms where it received an important function, such as chemotaxis in *Dictyostelium* and sperm maturation in vertebrates. The imbalance in the number of sequences present in the sAC group compared with the prevailing eukaryotic cyclases makes it tempting to refer to sAC as closely related to bacterial ACs (Buck et al. 1999; Hanoune and Defer 2001; Roelofs et al. 2001*a*, 2001*b*). When performing blast searches using

sAC cyclase domains as input (e.g., C1 domain of HssAC1), the first bacterial sequence appears at position 6, whereas for a cyclase domain of the prevailing group (e.g., RnGCE), the first bacterial sequence does not come before position 260. However, the expectation values obtained are 6×10^{-7} and 5×10^{-15} for HssAC1 and RnGCE, respectively, suggesting that the evolutionary distances toward bacterial cyclases is slightly shorter for RnGCE than for HssAC1, which was also observed in the phylogenetic analysis (fig. 3). The phylogenetic analysis of the family of cyclase domains in eukaryotes suggests that both the small group of sACs and the large group of prevailing cyclase are derived from the cyclase domain in bacteria. Whereas the prevailing cyclases became very successful during evolution, the soluble cyclases have become fixed in only a few eukaryotes and bacteria but have not become ubiquitous among them.

Acknowledgments

LITERATURE CITED

ANDERSSON, J. O., and S. G. ANDERSSON. 1999. Insights into the evolutionary process of genome degradation. Curr. Opin. Genet. Dev. **9**:664–671.

ARAVIND, L., R. L. TATUSOV, Y. I. WOLF, D. R. WALKER, and E. V. KOONIN. 1998. Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. Trends. Genet. **14**:442–444.

ARAVIND, L., H. WATANABE, D. J. LIPMAN, and E. V. KOONIN. 2000. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. Proc. Natl. Acad. Sci. USA **97**:11319–11324.

BALDAUF, S. L., A. J. ROGER, I. WENK-SIEFERT, and W. F. DOOLITTLE. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science **290**:972–977.

BRAUN, E. L., A. L. HALPERN, M. A. NELSON, and D. O. NATVIG. 2000. Large-scale comparison of fungal sequence information: mechanisms of innovation in *Neurospora crassa* and gene loss in *Saccharomyces cerevisiae*. Genome Res. **10**:416–430.

BUCK, J., M. L. SINCLAIR, L. SCHAPAL, M. J. CANN, and L. R. LEVIN. 1999. Cytosolic adenylyl cyclase defines a unique signaling molecule in mammals. Proc. Natl. Acad. Sci. USA **96**:79–84.

CHEN, Y., M. J. CANN, T. N. LITVIN, V. IOURGENKO, M. L. SINCLAIR, L. R. LEVIN, and J. BUCK. 2000. Soluble adenylyl cyclase as an evolutionarily conserved bicarbonate sensor. Science **289**:625–628.

DANCHIN, A. 1993. Phylogeny of adenylyl cyclases. Adv. Second Messenger Phosphoprotein Res. **27**:109–162.

DOOLITTLE, R. F., D. F. FENG, S. TSANG, G. CHO, and E. LITTLE. 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science **271**:470–477.

FELSENSTEIN, J. 1996. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. Methods Enzymol. **266**:418–427.

FENG, D. F., G. CHO, and R. F. DOOLITTLE. 1997. Determining divergence times with a protein clock: update and reevaluation. Proc. Natl. Acad. Sci. USA **94**:13028–13033.

GRAY, M. W. 1999. Evolution of organellar genomes. Curr. Opin. Genet Dev. **9**:678–687.

HANOUNE, J., and N. DEFER. 2001. Regulation and role of adenylyl cyclase isoforms. Annu. Rev. Pharmacol. Toxicol. **41**:145–174.

KOONIN, E. V., A. R. MUSHEGIAN, M. Y. GALPERIN, and D. R. WALKER. 1997. Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. Mol. Microbiol. **25**:619–637.

LANDER, E. S., L. M. LINTON, BIRREN ET AL. (251 co-authors). 2001. Initial sequencing and analysis of the human genome. Nature **409**:860–921.

NEI, M., P. XU, and G. GLAZKO. 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. Proc. Natl. Acad. Sci. USA **98**:2497–2502.

NELSON, K. E., R. A. CLAYTON, S. R. GILL et al. (25 co-authors). 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature **399**:323–329.

OCHMAN, H., and N. A. MORAN. 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science **292**:1096–1099.

PONTING, C. P. 2001. Plagiarized bacterial genes in the human book of life. Trends Genet. **17**:235–237.

ROELOFS, J., H. SNIPPE, R. G. KLEINEIDAM, and P. J. M. VAN HAASTERT. 2001a. Guanylate cyclase in *Dictyostelium discoideum* with the topology of mammalian adenylate cyclase. Biochem. J. **354**:697–706.

ROELOFS, J., M. MEIMA, P. SCHAAP, and P. J. VAN HAASTERT. 2001b. The Dictyostelium homologue of mammalian soluble adenylyl cyclase encodes a guanylyl cyclase. EMBO J. **20**:4341–4348.

ROELOFS, J., and P. J. VAN HAASTERT. 2001c. Genes lost during evolution. Nature **411**:1013–1014.

SALZBERG, S. L., O. WHITE, J. PETERSON, and J. A. EISEN. 2001. Microbial genes in the human genome: lateral transfer or gene loss? Science **292**:1903–1906.

SCHULTZ, J., R. R. COPLEY, T. DOERKS, C. P. PONTING, and P. BORK. 2000. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res. **28**:231–234.

SCHULTZ, J., F. MILPETZ, P. BORK, and C. P. PONTING. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA **95**:5857–5864.

STANHOPE, M. J., A. LUPAS, M. J. ITALIA, K. K. KORETKE, C. VOLKER, and J. R. BROWN. 2001. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. Nature **411**:940–944.

TESMER, J. J., R. K. SUNAHARA, A. G. GILMAN, and S. R. SPRANG. 1997. Crystal structure of the catalytic domains of adenylyl cyclase in a complex with $G_{s\alpha} \cdot GTP\gamma S$. Science **278**:1907–1916.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

WEDEL, B., and D. GARBERS. 2001. The guanylyl cyclase family at y2k. Annu. Rev. Physiol. **63**:215–233.

WOLF, Y. I., F. A. KONDRASHOV, and E. V. KOONIN. 2000. No footprints of primordial introns in a eukaryotic genome. Trends Genet. **16**:333–334.

YU, S., L. AVERY, E. BAUDE, and D. L. GARBERS. 1997. Guanylyl cyclase expression in specific sensory neurons: a new family of chemosensory receptors. Proc. Natl. Acad. Sci. USA **94**:3384–3387.

ZHANG, G., Y. LIU, A. E. RUOHO, and J. H. HURLEY. 1997. Structure of the adenylyl cyclase catalytic core. Nature **386**: 247–253.