cGMP signalling
Roelofs, Jeroen; Smith, Janet L.; Haastert, Peter J.M. van

Published in: Trends in Genetics

DOI: 10.1016/S0168-9525(02)00044-6

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date: 2003

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
translated, but where the entire primary sequence of the protein is not the information that is under selection. In many cases, the number of amino acids under selection could be few or none, which can lead to the appearance of a high $K_s/K_a$ ratio, even though there is selection for expression of the small protein sequence. For example, leader peptides are small proteins that have critical roles in gene expression through translational control [2]. Here, pausing of the ribosome during translation of a leader protein may allow for the formation of an anti-terminator RNA structure, thereby allowing transcription of the downstream genes in an operon. If the ribosome does not pause, a rho-independent terminator will form and the operon is not expressed. Leader peptides in the Escherichia coli genome control the thr, leu, trp and phe operons, and their $K_s/K_a$ ratios are far greater than expected for protein-coding genes (three of them are even greater than Ochman’s conservative $\mu + 2\sigma$ threshold, Fig. 1a), which would seem to indicate that these small ORFs do not encode proteins. In reality, only amino acids responsible for ribosome pausing when a charged tRNA becomes depleted (e.g. two tryptophan codons in the trpL gene) might be under selection for amino acid identity. The additional residues of the leader peptide are not under selection for the function of the protein (although they can participate in mRNA secondary structure formation). Similar leader peptides are found upstream of the tnaA tryptophanase gene [3,4] and chloramphenicol resistance genes [5] and operate by different methods. Similar small upstream ORFs occur in eukaryotic genomes (e.g. the 25-codon arginine-sensing peptide cotranscribed with the Saccharomyces cerevisiae CPA1 gene [6]). But the end result is the same: the short length, poor conservation and often unusual composition of these leader peptides can easily lead one to dismiss the coding potential of their genes.

In some cases, the majority of a protein could be disposable, also leading to high $K_s/K_a$ ratios. Although small disposable portions are commonly seen among leader sequences for secreted proteins, or pro-proteins made in inactive states (e.g. Bacillus pro-$\alpha$ factors); the most extreme case might be the pqqA peptide, which is overproduced relative to other proteins in the Klebsiella pqq operon and may serve as the substrate for the synthesis of the cofactor PQQ [7,8]. Here, the amino acids glutamate and tyrosine may be cleaved from the peptide backbone and serve as the substrate for cofactor biosynthesis, and the remaining residues may serve as a scaffold. Some transcribed regions may not encode polypeptides at all. The $K_s/K_a$ test is useless in the identification of small functional RNAs, which can have important roles in cellular metabolism, and may be great in number [9]. Some of the regions designated as small ORFs – even those with genetic evidence for their importance – may act through an RNA product.

The manifold ways small protein products affect cellular metabolism make their identification onerous, and sometimes even comparisons with closely related genomes cannot aid in their unambiguous identification. In these cases, hands-on experimentation could be the only route towards gene discovery and the potentially fascinating insights in molecular biology that can result.

References

0168-9525/03 $ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0168-9525(02)00038-0

---

cGMP signalling: different ways to create a pathway

Jeroen Roelofs¹, Janet L. Smith² and Peter J.M. Van Haastert³

¹Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115-5730, USA
²Boston Biomedical Research Institute, 64 Grove Street, Watertown, Massachusetts 02472-2829, USA
³Dept of Biochemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Recently, a novel cGMP signalling cascade was uncovered in Dictyostelium, a eukaryote that diverged from the lineage leading to metazoans after plants and before yeast. In both Dictyostelium and metazoans, the ancient cAMP-binding (cNB) motif of bacterial CAP has been modified and assembled with other domains into cGMP-target proteins. The domain structures of these cGMP targets, as well as the enzymes responsible for cGMP synthesis and degradation, are entirely different between Dictyostelium and the ancient cNB motif. These findings provide new insights into the evolution of cGMP signalling. The discovery of novel small RNAs using comparative genomics and microarrays demonstrates the potential for identifying novel small RNA pathways in model organisms and may lead to the identification of new regulatory mechanisms in complex eukaryotic systems. This work highlights the importance of comparative genomics in understanding the evolution of signalling pathways. The use of microarrays allows for the rapid identification of new small RNA pathways, providing valuable insights into the regulation of gene expression in complex eukaryotic systems.
and metazoa, suggesting that different cGMP-signalling pathways evolved in these two lineages.

In metazoa, cGMP is synthesized by two guanylyl cyclases, one membrane-bound and the other soluble, that have a common phylogenetic predecessor (Fig. 1). Close homologues of these proteins are not found in Dictyostelium. Instead, the two Dictyostelium guanylyl cyclases, GCA and sGC, resemble adenylyl cyclases: GCA has the topology of the ubiquitous twelve membrane-spanning metazoan adenylyl cyclases, whereas sGC is homologous to a small family of soluble adenylyl cyclases present in vertebrates and bacteria [1,2]. Cyclic GMP is degraded in metazoa by class I phosphodiesterases, several of which are activated by cGMP-binding to GAF domains [3]. Dictyostelium also expresses cGMP-activated phosphodiesterases (PDE5 and PDE6; original names GbpA and GbpB, respectively), but these enzymes have a common evolutionary origin that is entirely different from the origin of metazoan enzymes [4–6]. Instead of a class I phosphodiesterase they contain the unrelated class II phosphodiesterase found only in bacteria and primitive eukaryotes, and cGMP stimulation is mediated by cNBD domains instead of GAF domains, two structurally and evolutionarily unrelated cyclic nucleotide binding domains.

In metazoa, cGMP signalling is mediated by cGMP-regulated protein kinase (PKG), ion channels and possibly Ras guanine nucleotide exchange factors (CNRasGEF), all of which possess cNBD domains [7,8]. Cyclic GMP signalling in Dictyostelium is mediated by Gbpc and D [4,5]. Although the combination of cNBD and RasGEF domains in Gbpc and D is reminiscent of CNRasGEF, the domains are shuffled, and in phylogenetic analysis the cNBD domains from CNRasGEF group with those in protein kinases, whereas the cNBD domains of Gbpc and D are more closely related to bacterial CAP. The protein kinase domain of Gbpc is homologous to mitogen-activated protein kinases, GCA and sGC, resemble adenylyl cyclases: GCA has the topology of the ubiquitous twelve membrane-spanning metazoan adenylyl cyclases, whereas sGC is homologous to a small family of soluble adenylyl cyclases present in vertebrates and bacteria [1,2].

**Fig. 1.** cGMP signalling in Dictyostelium and metazoa. Domains are colour-coded by biochemical functions, and have the same shape if they are evolutionarily closely related. cGMP is produced by cyclases (blue, with a black dot for the catalytic site) and degraded by phosphodiesterases (PDE, orange). The domains that bind cGMP are red, and the domains that mediate the downstream effects of cGMP are shades of green; N refers to the N-terminus of the proteins. The order of domains in Dictyostelium is generally the opposite of that in metazoa. PDE5 and PDE6 are phosphodiesterases with identical domain composition, that are encoded by the gbpA and gbpB genes, respectively. GbpD is homologous to the C-terminal part of GbpC, including the GEF and cNBD domains.

**Fig. 2.** Species tree. The phylogeny of species derived from the estimated divergence time of several proteins in human, Drosophila, yeast, plants, Plasmodium and bacteria [11]. The phylogenetic position of Dictyostelium has been determined in two independent studies [12,13]. Dictyostelium is a member of a group of early eukaryotes including Entamoebae and Mastigamoeba, called Canosa [13,14]. The presence of cGMP signalling the other Canosa species is unknown. cGMP signalling is absent in yeast, plants and most bacteria, but is present in Plasmodium and the related species Paramecium and Tetrahymena. The scale indicates the estimated species divergence time before present. Myr, million years.
kinase kinase kinase (MAPKKK), placing it in an entirely different branch of the protein kinase superfamily from PKG. The Ras and MAPKKK region of GbpC is similar to the human protein KIAA1790, suggesting that GbpC might have arisen by fusion of a KIAA1790-like protein with a protein containing RasGEF and cNB domains.

Cyclic GMP signalling does not appear to occur in eubacteria, yeast and plants, whereas Drosophila and Caenorhabditis elegans have cGMP signalling proteins that are similar to those found in vertebrates. There is also evidence for guanylyl cyclases in cyanobacteria [9], Paramecium, Tetrahymena and Plasmodium [10]. As the last common ancestor of Dictyostelium and metazoa evolved after Plasmodium (Fig. 2), it will be interesting to see how other cGMP signalling proteins deduced from the completed genome of Plasmodium compare with the systems in Dictyostelium and metazoa.

At this juncture it is clear that phylogenetically distinct cGMP signalling components have evolved in the ancestor of Dictyostelium and in the ancestor of metazoa, to create two different but in many ways analogous cGMP signalling pathways.

Acknowledgements

We are indebted to all of the teams involved in the Dictyostelium sequencing projects and to Leonard Bosgraaf for preparing the figures.

References