Nucleic Acids Research, 1998, Vol. 26, No. 16 **3619–3625**

SURVEY AND SUMMARY

Survey, analysis and genetic organization of genes encoding eukaryotic-like signaling proteins on a cyanobacterial genome

Cheng-Cai Zhang*, Leticia Gonzalez and Vincent Phalip

Unité d'Immunotechnologie et Microbiologie Moléculaires, Ecole Supérieure de Biotechnologie de Strasbourg, Université Louis Pasteur de Strasbourg, Boulevard Sébastien Brandt, F-67400 Illkirch, France

Received March 18, 1998; Revised and Accepted June 4, 1998

ABSTRACT

Bacteria usually use two-component systems for signal transduction, while eukaryotic organisms employ Ser/Thr and Tyr kinases and phosphatases for the same purpose. Many prokaryotes turn out to harbor Ser/Thr and Tyr kinases, Ser/Thr and Tyr phosphatases, and their accessory components as well. The sequence determination of the genome of the cyanobacterium *Synechocystis* sp. strain PCC 6803 offers the possibility to survey the extent of such molecules in a prokaryotic organism. This cyanobacterium possesses seven Ser/Thr kinases, seven Ser/Thr and Tyr phosphatases, one protein kinase interacting protein, one protein kinase regulatory subunit and several WD40-repeat-containing proteins. The majority of the protein phosphatases presented in this study were previously reported as hypothetical proteins. We analyze here the structure and genetic organization of these ORFs in the hope of providing a guidance for their functional analysis. Unlike their eukaryotic counterparts, many of these genes are clustered on the chromosome, and this genetic organization offers the opportunity to study their possible interaction. In several cases, genes of two-component transducers are found within the same cluster as those encoding a Ser/Thr kinase or a Ser/Thr phosphatase; the implication for signal transduction mechanism will be discussed.

INTRODUCTION

Bacteria are able to sense a variety of internal and external factors in order to respond to environmental changes. The molecular mechanism underlying the signal transduction process, which involves the so-called 'two-component systems', is now well understood (1,2). A simple prototype of two-component systems includes two proteins, a histidine kinase and a response regulator. The histidine kinase autophosphorylates on a conserved histidine residue in response to a stimulus and then transfers the phosphate group to an aspartate residue of a cognate response regulator which is often also a transcription factor (1,2). Two-component systems are ubiquitously present among prokaryotes, and similar components are also reported in several eukaryotic organisms (2). In the cyanobacterium *Synechocystis* sp. strain PCC 6803 alone, more than 80 open reading frames (ORFs), representing 2.5% of its total coding capacity, are found to encode proteins of two-component systems (3,4).

Bacteria may use eukaryotic-like components for signal transduction as well. Indeed, during the last few years, several bacteria have been shown to harbor Ser/Thr and Tyr kinases and phosphatases (for reviews see 5,6). Such enzymes are very abundant in all eukaryotic organisms. In the yeast *Saccharomyces cerevisiae* for example, 2% of its total genes encode protein kinases and a similar amount of genes encode protein phosphatases (7–9). These enzymes, together with their regulatory proteins, form signaling cascades and networks in order to regulate a variety of cellular activities. Eukaryotic-type protein kinases can be divided into two classes based on their substrate specificity: Ser/Thr kinases and Tyr kinases (10–12). Only a few dual-specificity protein kinases can phosphorylate on both Ser/Thr and Tyr residues. Both classes of protein kinases belong to a single enzyme superfamily, as they share a homologous catalytic domain of ~260 amino acids with some conserved signatures indicative of their belonging to either of the two classes (11,12). On the other hand, at least three families of Ser/Thr and Tyr phosphatases can be distinguished based on their sequence comparison (8,13,14). Members of the PPP family, such as the mammalian PP1, 2A and 2B, are Ser/Thr phosphatases (13). The second family of Ser/Thr phosphatases is the PPM family, represented by the mammalian PP2C (13). No significant sequence similarity can be found between these two families of Ser/Thr phosphatases. The third family of protein phosphatases, the PTP family, includes low-molecular-weight PTPs, and PTPs with dual specificity because of their ability to dephosphorylate both phospho-Ser/Thr and phospho-Tyr residues (14). The catalytic domains of low-molecular-weight PTPs show a similar three dimensional structure to those of the dual specificity protein

*To whom correspondence should be addressed. Tel: +33 3 8865 5290; Fax: +33 3 8865 5330; Email: cheng-cai.zhang@esbs.u-strasbg.fr

phosphatases, but these two classes of protein Tyr phosphatases are not related in primary sequence and share only the CXXXXXR catalytic motif (14,16).

How many eukaryotic-type signaling molecules could be found in a bacterium? The whole genome of the unicellular cyanobacterium *Synechocystis* sp. PCC 6803 was recently sequenced (3,15). This offers a possibility to answer this question and to assess the extent and importance of such molecules in bacterial signal transduction besides the classical two-component systems. In this report, different classes of eukaryotic-type signaling molecules are compiled and analyzed, along with their surrounding sequences and genetic organization.

MATERIALS AND METHODS

Sequences or conserved motifs of major families of eukaryotic signaling proteins were used to screen for similar molecules in the CyanoBase (http://www.kazusa.or.jp/cyano/cyano.html), a databank with the entire sequence of the cyanobacterium *Synechocystis* sp. PCC 6803. All positive scores after this screening were again used to search for similar sequences in the CyanoBase, in order to ascertain that sequences distantly related to eukaryotic proteins could also be found. Multiple sequence alignment was carried out using the Cluster W program (17).

To check the DNA sequence at the junction between sll1574 and sll1575 (see below for more details), two oligonucleotide primers were used to amplify the corresponding genomic region by polymerase chain reaction (PCR) using the high-fidelity Vent DNA polymerase (Biolabs). The sequences of the two PCR primers are: primer 1, ACTATTTCGGCCCTAC; primer 2, TGGGCACAATCCAAGC. The 204 bp long PCR product was purified from agarose gel using a GeneClean II kit (Bio101) and sequenced with the Sequenase II (Amersham).

Major classes of eukaryotic-type signal transducers find their putative representatives in *Synechocystis* sp. PCC 6803

A family of seven putative Ser/Thr and Tyr kinases has been previously identified (3; Table 1), and their detailed analysis is shown below. A thorough survey of the genome of *Synechocystis* sp. PCC 6803 has also identified seven putative protein phosphatases (Table 1), one belonging to the PPP family (sll1387), five to the PPM family (slr1860, slr1983, slr0114, slr2031 and sll1365) and one to the PTP family (slr0328). Most of these protein phosphatases were previously considered as hypothetical (3). Several ORFs encode potential accessory signaling molecules; these include slr1234 encoding a protein kinase C interacting protein, slr0593 encoding a protein kinase regulatory subunit, and several WD40-repeat-containing proteins (3,18).

Ser/Thr and Tyr kinases and their structural analysis

The following ORFs encode Ser/Thr and Tyr kinases: sll0776, sll1574, sll1575, slr0152, slr0599, slr1225, slr1443 and slr1697 (Table 1). Several observations suggest that sll1574 and sll1575 form one single ORF and encode two different parts of the same Ser/Thr kinase. Indeed the peptide deduced from sll1574 shows sequence similarity to the catalytic domain of Ser/Thr kinases (12), but only from subdomains I–V, while subdomains VI–XI are followed immediately on the polypeptide deduced from sll1575 (Fig. 1). We first suspected a sequencing error at the junction between sll1574 and sll1575. Therefore, a genomic region of 204 bp covering this junction was amplified by PCR and sequenced. The result of this experiment, however, confirms the original data. Dr S.Tabata's group at the Kazusa DNA Research Institute, where the genome of *Synechocystis* sp. PCC6803 was sequenced, could not find any sequencing error in this region either (S.Tabata, personal communication). A 1.57 kb DNA fragment covering sll1574 and sll1575 was inserted into an *Escherichia coli* overexpression vector, and the size of the produced polypeptide corresponds well with that of only sll1574 instead of a sll1574–sll1575 fusion product (data not shown). These results further indicate that a stop codon indeed interrupts sll1574 and sll1575 (Fig. 1). For the purpose of convenience, these two ORFs will be jointly referred to as sll1574–75 in the text since neither of them alone is likely to be functional as a protein kinase. Several possibilities may explain this observation. *Synechocystis* sp. PCC 6803 has been cultured in the laboratory since three decades ago, therefore, sll1574–75 may no longer be required for cell growth under laboratory culture conditions. Another possibility would be that the stop codon interrupting these two ORFs is suppressed during the translation or restored to a non-stop codon post-transcriptionally by a mechanism such as the RNA editing.

Table 1. Eukaryotic-type signal transducers in *Synechocystis* sp. PCC 6803 analyzed in this study

ORF	Similarity to
sll0776	protein Ser/Thr kinase
sll1574–75	protein Ser/Thr kinase
slr0152	protein Ser/Thr kinase
slr0599	protein Ser/Thr kinase
slr1225	protein Ser/Thr kinase
slr1443	protein Ser/Thr kinase
slr1697	protein Ser/Thr kinase
sll1365	protein Ser/Thr phosphatase (PPM family)
sll1387	protein Ser/Thr phosphatase (PPP family)
slr0114	protein Ser/Thr phosphatase (PPM family)
slr0328	protein Tyr phosphatase (PTP family)
slr1860	protein Ser/Thr phosphatase (PPM family)
slr1983	protein Ser/Thr phosphatase (PPM family)
slr2031	protein Ser/Thr phosphatase (PPM family)
slr1234	protein kinase C interacting protein
slr0593	protein kinase A regulatory chain

The seven Ser/Thr and Tyr kinases in *Synechocystis* sp. PCC 6803 show various degrees of sequence identity (23.8–58.4%) to PknA (19), the first Ser/Thr kinase found among cyanobacteria (Fig. 1). The protein kinase catalytic domains of these proteins are always located at the N-terminal, as is the case for all known eukaryotic-type protein kinases isolated so far from the filamentous cyanobacterium *Anabaena* sp. PCC 7120 in our laboratory (5,19,20,26). Protein sequence analysis suggests that they are more likely to be Ser/Thr kinases than Tyr kinases (11). In addition, four out of the seven protein kinases have one (slr0599, sll0776 and slr1443) or several (slr1225) putative transmembrane segments at their C-terminal regions. Those protein kinases with

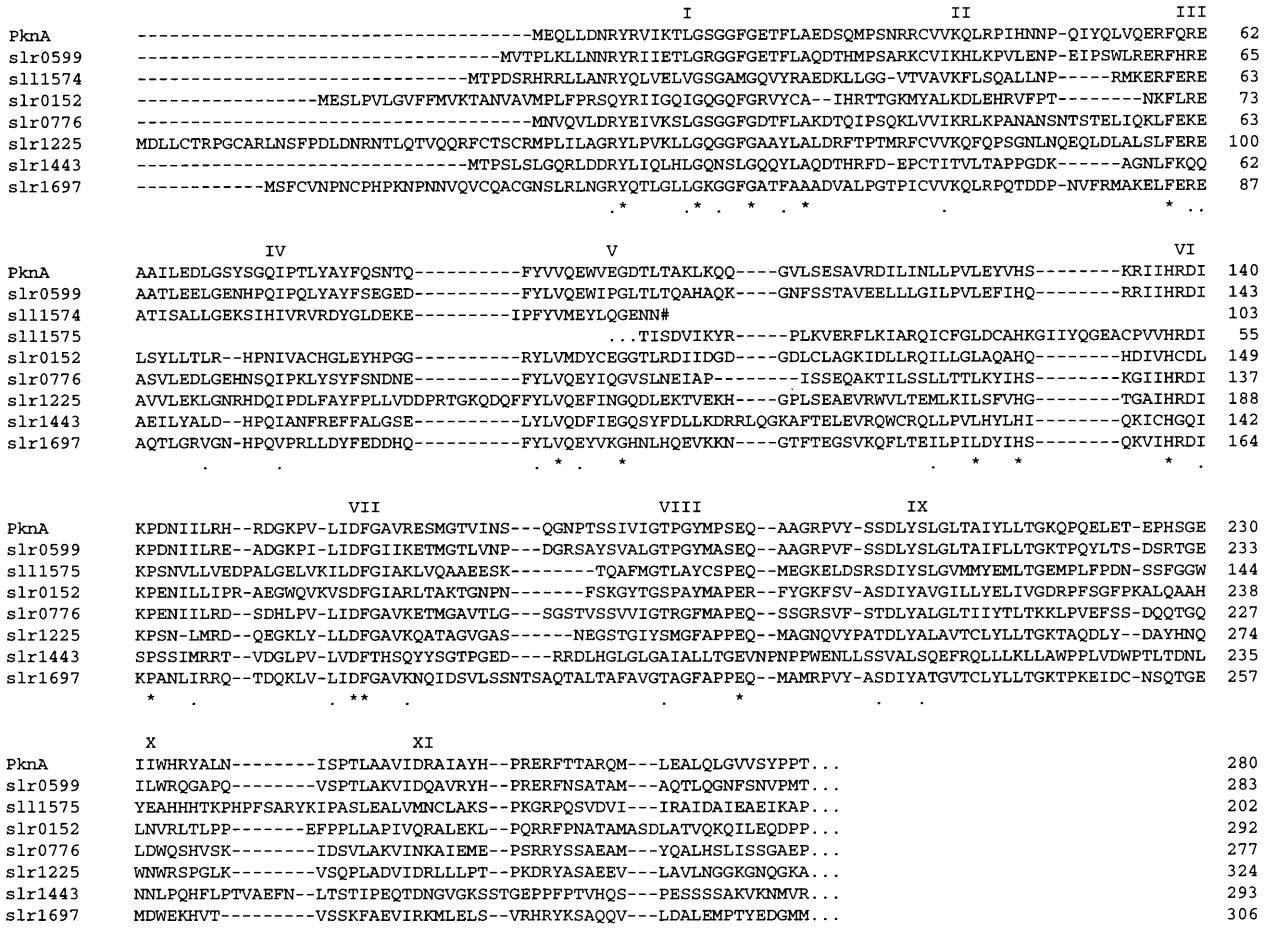

Figure 1. Sequence comparison of the catalytic domains of all protein Ser/Thr kinases from *Synechocystis* sp. strain PCC 6803. PknA is from the nitrogen-fixing cyanobacterium *Anabaena* sp. PCC 7120 (19). sll1574 and sll1575 are jointly referred to as sll1574–75 elsewhere in the text (see text for detailed discussion). Identical residues are marked by (*) and conserved residues are identified by a period (.). # indicates a stop codon. Residues are numbered on the right. Subdomains I–XI are also indicated as proposed (12).

one C-terminal transmembrane segment could be localized on the membrane, as in the case of the vertebrate protein kinase Myt1, in which the C-terminal hydrophobic region serves as an internal signal sequence (21). Other protein kinases with several putative transmembrane segments could act as membrane receptors in signal transduction.

Regions other than the catalytic domain of Ser/Thr and Tyr kinases often play regulatory functions in regard to their catalytic domain. Some of the regulatory domains of Ser/Thr kinases, for example, exert an inhibitory effect that is crucial for the maintenance of a protein kinase in an inactive form in the absence of signals, while others are required for the proper cellular localization of a given protein kinase (22,32). To gain a better understanding of the function of these enzymes in *Synechocystis* sp. PCC 6803, the C-terminal regions following their catalytic domains were compared against protein sequences in various databanks. The C-terminal region of slr1697 shows significant sequence similarity to regions of several proteins such as the C-terminal region of HglK from the heterocyst-forming and nitrogen-fixing cyanobacterium *Anabaena* sp. PCC 7120 (23). This conserved domain is characterized by repeats of a short motif of five residues (A D/N L S/T X), as pointed out in the HglK sequence by Bauer *et al*. (23). A particular feature of this domain is the highly conserved leucine residue found by every five residues (Fig. 2). For this reason, this conserved domain is termed as a leucine-repeat (LR) domain. The length of this domain is variable, depending on the number of pentapeptidic repeats present (Fig. 2). The only characterized proteins containing a LR domain are HglK and McbG. The former is believed to be involved in the localization of heterocyst-specific glycolipids in *Anabaena* sp. PCC 7120 (23) and the latter required for the export of a peptide antibiotic, microcin B17 (24). While a dozen proteins from *Synechocystis* sp. PCC 6803 possess a LR domain, no or few such proteins have been found in other organisms. As already suggested by Bauer *et al*. (23), the LR domain may interact with glycolipids in *Anabaena* sp. PCC 7120. Other possible roles of this domain would include the localization of the protein kinase slr1697 to its proper cellular compartment (22) or the transport of molecules from their sites of synthesis to the thylakoids in cyanobacteria. The latter possibility would explain the scarcity of such proteins in other non-photosynthetic prokaryotes whose genomes have also been sequenced.

The 235 residue long C-terminal region of slr0599 contains 27.2% of proline residues. Proline-rich sequences have diverse

```
slr1697   NLVGIVLAKAFVPGINCYQANLTNANFEQAELTRADFGKARLKNVIFKGANLSDAYFGYADLRGADLRGANLNGVNFKYANLQGANFSGADLGSAKVSPEQL  256
pSW200    DLTNANLNQTDLPNVNLSGANLAHANLTMAYLSEADLSNANLSNADLKRADLSNANLSGADLTNANLNQTDLPNVNLSGANLAHANLTMAYLSEADLSNANL  234
slr0516   NLAGADLREFNLENARLNRSDLSGANLSGVNLRRALLDRANLTGANLSETDLTEAALTEANLAGADLSGANLERSFLRDVDLTGANLKGANLAWANLTAANL  145
slr1152   NLQNADLSGFTLISVDFERTNLIGSNLQRTFLTKARLGHCQMNWADLTYAKLNQADLSHADLTKASLYGAFAVKTNFKGAKLSGATLAHANLRGANLEQTNL  152
slr1851   DLGGANLTRAQLDSATLKNANLALANMTEVCLIYADLSNADLSGANLVGADLTNADLSGAKLGGADLRKANLSEASLRGADLRGVNLIEANLTNTDFSEADL  145
slr1819   DLIGIVLNEADLRGANLLFCYLNRANLGQANLVAANLSGASLNQADLAGADLRSANFHGAMLQGAILRDSDMTLATLQDTNLIGADLRGADLSGATLTGACL  125
HglK      DLSQAQMKQANFTDANLSRVLMTRSDLSRATLNRANLSNARLIGANLSSAQLVGADLRGTVLENASLTGADLGDAKLQEANLYGARLSRVIAIGAQLSFANL  603
McbG      LIDCEFKDCLLQGVNAADIMFPCTFSLVNCDLRFVDFISLRLQKSIFLSCRFRDCLFEETDLRKSDFTGSEFNNTEFRHSDLSHCDFSMTEGLDINPEINRI  170
```

Figure 2. Multiple sequence alignment of the LR domain from various proteins. pSW200 (accession no L42525) is a plasmid from *Erwinia stewartii*. HglK is from the nitrogen-fixing cyanobacterium *Anabaena* sp. PCC 7120 (23). McbG is from *E.coli* (24). All other sequences are from *Synechocystis* sp. strain PCC 6803. The size of the LR domain in a given protein can be shorter or longer than that shown here, depending on the number of the pentapeptidic repeat present. Residues are numbered on the right, and the most conserved leucine (L) and alanine (A) residues are highlighted with bold characters. The consensus sequence of the pentapeptidic repeat is A D/N L S/T X.

```
sll1387   ------------------------MPN-----PRRIVFGDVHGHFDALTALFEAIAPNERDGVYFVGDLIDRGPESAKVV--DFVMENKYHC----LLGNHE   67
PP13-AT   TKQVQLTEAEIKHLCSTAKQIFLTQPNLLELEAPIKICGDTHGQFSDLLRLFEYGGYPPAANYLFLGDYVDRGKQSVETICLLLAYKIKYKENFFLLRGNHE  123
PP2A1-SC  SKCEPLSEDDVARLCKMAVDVLQFEENVKPINVPVTICGDVHGQFHDLLELFKIGGPCPDTNYLFMGDYVDRGYYSVETVSYLVAMKVRYPHRITILRGNHE  179
PP2A-Hu   NECKQLSESQVKSLCEKAKEILTKESNVQEVRCPVTVCGDVHGQFHDLMELFRIGGKSPDTNYLFMGDYVDRGYYSVETVTLLVALKVRYRERITILRGNHE  119
                                        *          . ** **.*  *  **          .  *.** .***  *   .         .*      * ****

sll1387   QMMLDAVGGF--NFSPQL--LHAWIYSGGKSTLESYEHQIPQSHVDWMRNLPLYLDLGDVWLVHAGVDPRLPIEEQG-EAQFCWIRDEFHRYPYPYFANKLI  164
PP13-AT   CASINRIYGFYDECKKRY-SVRVWKIFTDCFNCLPVAALIDEKILCMHGGLSPELKHLD--EIRN-IPRPADIPDHGLLCDLLWSDPDKDIEGWGENDRGVS  221
PP2A1-SC  SRQITQVYGFYDECLRKYGSANVWKMFTDLFDYFPITALVDNKIFCLHGGLSPMIETID--QVRE-LNRIQEVPHEGPMCDLLWSDPD-DRGGWGISPRGAG  277
PP2A-Hu   SRQITQVYGFYDECLRKYGNANVWKYFTDLFDYLPLTALVDGQIFCLHGGLSPSIDTLD--HIRA-LDRLQEVPHEGPMCDLLWSDPD-DRGGWGISPRGAG  217
             .  . **      .       *                .          .*    .   *   ..  .      .   *  .   *    .

sll1387   IT-GHTITFTFANVEPGKLVSGPGWLDIDTGAYHPKSGWLTALELNNQEVYQAHIFTNEVRRLPLEDAVVPLPP---QN-----LHRSRQKGKKRGLLG#    254
PP13-AT   YTFGADKVEEFLQTHDLDLICRAHQVVEDGYEFFANRQLVTIFSAPNYCGEFDNAGAMMSVDDSLTCSFQILKASEKKGNFGFGKNAGRRGTPPRKGGGKG#  322
PP2A1-SC  FTFGQDVSEQFNHTNDLSLIARAHQLVMEGYAWSHQQNVVTIFSAPNYCYRCGNQAAIMEVDENHNRQFLQYDPSVRPG-----EPSVSRKTPDYFL#     369
PP2A-Hu   YTFGQDISETFNHANGLTLVSRAHQLVMEGYNWCHDRNVVTIFSAPNYCYRCGNQAAIMELDDTLKYSFLQFDPAPRRG-----EPHVTRRTPDYFL#     309
           * *      *       *..    .  .          .*     *       .  .                          .          .
```

Figure 3. Sequence comparison between sll1387 and some PPP-family Ser/Thr phosphatases. PP13-AT, PP1 isoform 3 from *A.thaliana* (databank accession number P48483); PP2A1-SC, PP2A-1 from *S.cerevisiae* (P23594); PP2A-Hu, PP2A α catalytic subunit from human (P13197). Identical residues are marked by (*) and conserved residues are identified by a period (.). # indicates the end of a polypeptide chain. Residues are numbered on the right.

binding functions (25). This region indeed shows sequence similarity with fibronectin-binding proteins (A32192, U53585) from *Staphylococcus aureus* and *Mycobacterium avium*, an actin-associated protein from *Schizosaccharomyces pombe* (Z98980), and a SH3-domain binding protein (U25281) from *Rattus norvegicus*. The majority of the identical residues in these aligned sequences are proline residues (data not shown).

Ser/Thr and Tyr phosphatases; a new class of membrane sensors?

The ORF sll1387 encodes a protein of 255 amino acid residues showing significant sequence similarity to the catalytic domains of PPP-family Ser/Thr phosphatases from eukaryotes (Fig. 3). It is also similar to the catalytic domain of a putative protein phosphatase from *Anabaena* sp. PCC 7120 characterized in our laboratory (26). Another ORF, slr0328, has been previously described as a gene encoding a homolog of low-molecular weight Tyr phosphatases (3, see also Fig. 4). A polypeptide similar to slr0328, encoded by a partially-sequenced ORF adjacent to genes encoding phycobiliproteins from another cyanobacterium *Synechococcus* sp. WH8020, was also reported (27).

Five ORFs, sll1365, slr0114, slr1860 (*icfG*), slr1983 and slr2031, encode proteins similar to PPM-family Ser/Thr phosphatases (13). They all possess a conserved catalytic domain at their C-terminal end (Fig. 5), with signatures shown to be critical for the activity of PPM-type protein phosphatases (28). Among these genes, only *icfG* was previously investigated (29) and will be analyzed further along with its surrounding sequences below. The other four ORFs were labeled as hypothetical in the CyanoBase databank (3).

The N-terminal regions of these ORFs, preceding their protein-phosphatase catalytic domain, were also analyzed. The N-terminal region of slr1983 shows strong sequence similarity to response regulators of bacterial two-component signaling systems. Response regulators are the second member of bacterial two-component signal transduction systems (1,2). A response regulator domain can form either a single polypeptide, or a part of a multi-domain protein such as in the case of the hybrid His kinases which possess in addition a His kinase domain (1,2). Because of its particular structural features, slr1983 represents a new class of regulatory proteins which can be termed as hybrid protein phosphatases equivalent to hybrid His kinases. Interestingly, an ORF (slr1982) upstream of slr1983 encodes a His kinase protein. The possibility that slr1982 and slr1983 working together as a two-component system has been previously postulated by Mizuno *et al.* (4). It would be interesting to know the role of the protein phosphatase domain of slr1983 in signal transduction within this putative two-component system.

The N-terminal regions of both sll1365 and slr0114 possess two putative transmembrane domains. In addition, a region of 64 amino acids of sll1365 shows 23% sequence identity (49% sequence similarity) with the serine chemoreceptor protein of *E.coli* (30). sll1365 and slr0114 could function as membrane receptors (sensors). Typically, a sensor domain is found at the N-terminal

```
slr0328      -----MKLLFVCLGNICRSPAAENIMNAQIDQAGLGAKI-VCDSAGTSSYHVGDSPDRRMTESLKKRGYRVQ-GRARQFFPEDFAEFDLILAMDGDNYRN   93
YfkJ-BS      ----MISVLFVCLGNICRSPMAEAIFRDLAAKKGLEGKI-KADSAGIGGWHIGNPPHEGTQEILRREGISFDGMLARQVSEQDLDDFDYIIAMDAENIGS   95
PtpA-SC      ---MTYRVCFVCTGNICRSPMAEAVFRARVEDAGLGHLV-EADSAGTGGWHEGEGADPRTEAVLADHGYGLD-HAARQFQQSWFSRLDLVVALDAGHLRA   95
STP1-SP      -MTKNIQVLFVCLGNICRSPMAEAVFRNEVEKAGLEARFDTIDSCGTGAWHVGNRPDPRTLEVLKKNGIHTK-HLARKLSTSDFKNFDYIFAMDSSNLRN   98
ACP1-Bovin   AEQVTKSVLFVCLGNICRSPIAEAVFRKLVTDQNISDNW-VIDSGAVSDWNVGRSPDPRAVSCLRNHGINTA-HKARQVTKEDFVTFDYILCMDESNLRD   98
PTP-Hu       --QATKSVLFVCLGNICRSPIAEAVFRKLVTDQNISENW-VIDSGAVSDWNVGRSPDPRAVSCLRNHGIHTA-HKARQITKEDFATFDYILCMDESNLRD   96
                    . *** ******* ** .        ..       **      . *          *   *       **.         * . ..*  .

slr0328      ILAQDP-AGQYHHKVKMICDYT--EKFGDREVPDPYYGGQAGFEHVIDLLEDACGNLLTSLGKELVN#                                  157
YfkJ-BS      LRSMAG--FKNTSHIKRLLDYVE--DSDLADVPDPYYTGN--FEEVCQLIKTGCEQLLASIQKEKQL#                                 156
PtpA-SC      LRRLAP-TERDAAKVRLLRSYDPAVAGGDLDVPDPYYGGRDGFEECLEMVEAASTGLLAAVREQVEGRAA#                               164
STP1-SP      INRVKP--QGSRAKVMLFGEYAS--PGVSKIVDDPYYGGSDGFGDCYIQLVDFSQNFLKSIA#                                       156
ACP1-Bovin   LNRKSNQVKNCRAKIELLGSYD---PQKQLIIEDPYYGNDADFETVYQQCVRCCRAFLEKVR#                                      157
PTP-Hu       LNRKSNQVKTCKAKIELLGSYD---PQKQLIIEDPYYGNDSDFETVYQQCVRCCRAFLEKAH#                                      155
             .            ..     *          . **** .   *          .   *
```

Figure 4. Sequence comparison of slr0328 to some low-molecular-weight protein Tyr phosphatases. YfkJ-BS, YfkJ from *B.subtilis* (accession number D83967); PtpA-SC, PtpA from *Streptomyces coelicolor* (P53433); STP1-SP, STP1 from *S.pombe* (P41893); ACP1-Bovin, bovine ACP1 (P11064); PTP-Hu, human red-cell type PTP (M83654). Identical residues are marked by (*) and conserved residues are identified by a period (.). Underlined sequences indicate the highly conserved active site (16). # indicates the end of a polypeptide chain. Residues are numbered on the right.

```
slr0114   LEIKVAERTADLAKANEAITGLNAQLKQENMRLGTELDVARKLQMMVLPKHEELSEITDLDIACYMASATEVGGDYYDVVKIGN----------------  316
sll1365   TQDEMAEVVHAFNTVADALCRLTEKLQDENLRLEAEIDITRQLQQMLMPSEQELAAVAGLEIAGFMESATEVGGDYYDVLQTNG----------------  439
slr2031   TRRKLLQLISDQTAVAIANSDLNQKLRSR-ESQDRELEIASEIQNQLLPRC--CPQINGLDIAAQCKTASRVGGDYYDFIPANYDQLRQGDWLCRNTSHL  219
slr1860   RQDEIGILGNSFIYMKNQI-KTLIAQEVKDGVNRLELEKGRQIQQNFLPIS--LPDLQQWQINAVFEPARSVSGDFYDAFLLGDD---------------  419
slr1983   LKARIGSSLEKKRLRDQEK-LYTQQVEGLSEMMAKELEKGRQMQKNFLPAH--LLTRSGWEFSAYFSPAQQLAGDFYDLFELPGD---------------  380
              .                                *..    .*   .                      *  . **.**

slr0114   --RIFISIGDVTGHGLESGVMAIMAQTAIQTL---------------LTAE---------------ITEAVQFHNLLNQVLYKNTQRMELSKNMTLVMLD  384
sll1365   --QVRISIGDVTGHDLESSLVMIMAQTALRTL---------------LEHG---------------VTDPVKLLSAMNRTIYENTRRMGSSKNMTLSVLQ  507
slr2031   GVPWSIVIGDVMGKGVPAGLIMTMTRGMLRAE---------------VLNR---------------HSPAQILNHLNRVMYADLENSHRFVTLFYSEYN  288
slr1860   --YLAIVIGDVCDKGVGAAMFMGLFRSLLRVFSGETMPGDTCIRDVNYKCS--ANDGNGK---KKVIVQFLNAVRLTNDYIATEHGDMAMFATLFFGVID  512
slr1983   --RLGIVVADVCDKGVGAALFMGLFRSLIRIFSGQAALDGLINPSFNLPSGQGLLDVNLDNLLNNINFEPLESIKLINNYVAINHGEASMFATIFFGILE  477
              * . **  . . . .  . .  ..                                                 *  .          ..

slr0114   YQSG-ELILSGQHESIFVVRADG--AIEEIDTIDLGFPLALEKDITTFIARTQINLHSGDVVVLYTDGIPEAENGEKQFYGLA--RFQEVLIANHQSTAL  479
sll1365   YEAG-LIRLSGQHEEVLIIRSNG--ATEQIDTFELGFPLGLEQDISAFVTEREITLNSGDLAVLYTDGITEAMNSQRQFYGIE--RLQTVLCQHRHQAPA  602
slr2031   PETSILSYSNAAHNPPLLWRAGSDSPQCLIPLDTEGALIGLESDS--NYRDAQIQLVPGDVVLYYTDGLTDAGNAKGDRFDDKNLRIAFQRACETAQTAQ  386
slr1860   ISNGNLSYINAGHEPVFILNS----EGIKHRLKSTGPAVGMMPNS-TFTID-SLKIDPGEMLIGYTDGVTDARSPTKEFFGRQ--RLMETLTAN-FSAKT  603
slr1983   PSSGKLSYINGGHEPVFIVDSN---HQLKTKLTSTGPAVGMLPDL-QFKTA-EIILQPGDLLLSYTDGVTEAKSPTGNFFGKE--KLLNALGSS-FNSVD  570
                     *    .  .              *  . .            . .  *.. . ****. .*       .     .           .

slr0114   EISQAVVEDLMAYIG-DS-------RVFDDITLVIFKKI#                                                              510
sll1365   EIRHQAIADLRNFLG-DQ-------PAEDDVTLLILKQK#                                                              633
slr2031   GILTEIFTSVEAFVGSDNSHQTDKIPARDDMTLVVLRVKSTE#                                                           428
slr1860   EILDIIKQELIGHID-GS-------IQFDDITMIAVYRN#                                                              634
slr1983   QLMLNIKESLEEHMG-EA-------QQFDDITLLAIKFQTT#                                                           603
           .       .   .              **.*..
```

Figure 5. Sequence comparison of the catalytic domains of all PPM-family Ser/Thr phosphatases from *Synechocystis* sp. strain PCC 6803. Identical residues are marked by (*) and conserved residues are identified by a period (.). # indicates the end of a polypeptide chain. Residues are numbered on the right.

part of a sensory His kinase in bacterial two-component systems (1,2). sll1365 and slr0114, if confirmed, would thus represent a new class of membrane sensors with an original molecular architecture containing a sensory domain linked to a protein phosphatase domain instead of a protein kinase domain. This situation can be compared to that encountered in many eukaryotic receptors which contain either a protein kinase domain or a protein phosphatase domain, although the early examples were all receptor protein-tyrosine kinases (9,10).

The N-terminal region of slr2031 contains a GAF domain believed to play a role in light and redox sensing (31).

Accessory signaling molecules

The proper functioning of Ser/Thr and Tyr kinases and phosphatases requires different classes of regulatory proteins acting as scaffolds or adaptors for signaling complex formation and enzyme localization, or as effectors of protein kinases or phosphatases (25,41,42). Several WD40-repeat containing proteins (18) have already been previously pointed out (3). In addition, two other polypeptides (slr0593, slr1234) in *Synechocystis* sp. PCC 6803 may also be regulatory signaling molecules (Table 1). slr0593 is a protein of 434 residues which contains a segment similar to the regulatory subunit of the cAMP-dependent protein kinase (protein kinase A). The same conserved domain is also found in a protein of unknown function, U21853, present in another cyanobacterium *Anabaena* sp. PCC 7120. In eukaryotes, the regulatory subunit inhibits the catalytic subunit of protein kinase A in the absence of a signal, and the catalytic subunit is subsequently released and activated once cAMP is bound to the regulatory subunit (32).

slr1234 is similar to the protein kinase C interacting protein 1 from mammals (33). A similar ORF, ORF1, has also been found in the cyanobacterium *Synechococcus* sp. strain PCC 7942 in which it is co-transcribed with *psbAII* encoding form II of the D1

protein in the photosystem II reaction center (34). The expression level of ORF1-*psbAII* increases after exposure to high light intensity and an ORF1-disrupted mutant displays slower growth rate than the wild type (34). The protein kinase C interacting protein 1, also called protein kinase C inhibitor, is a zinc-binding protein belonging to the histidine triad (HIT) protein family (33). The zinc-binding motif, His-X-His-X-His, is conserved in slr1234 and ORF1 of *Synechococcus* sp. strain PCC 7942. The role of the protein kinase C interacting protein 1 in regulating protein kinase C activity remains unclear. Recent evidence from structural analysis suggests that these proteins act as nucleotidyl hydrolases, transferases, or both (35). In this case, it seems unlikely that they act as protein kinase C regulators.

Because the conserved signatures of accessory signaling proteins are often short and poorly characterized (36), the number of such molecules in *Synechocystis* sp. PCC 6803 is probably very much underestimated in this study.

Genetic organization

In order to gain a better insight into the function of eukaryotic-type signaling molecules in *Synechocystis* sp. PCC 6803, the genetic organization of a few gene clusters of major interest is also analyzed. Genes included within a cluster are those in close proximity to, and transcribed in the same direction as, a gene encoding either a protein kinase, a protein phosphatase, or a regulatory protein of these enzymes.

One cluster with potentially important regulatory function is that containing the protein phosphatase gene *icfG* (slr1860). This cluster contains 10 genes (slr1852–slr1862) packed in a region of ~9.5 kb. *icfG* was shown to be inducible by glucose and required for cell growth under conditions of low concentration of inorganic carbon in the presence of glucose (29). One ORF in this cluster, slr1857, encodes a protein highly similar to the glycogen-debranching enzyme, the second enzyme required for glycogen catabolism (37). One possible function of this gene cluster would thus be the degradation of glycogen, in accordance with the presence of slr1857 and the results of genetic analysis of *icfG* (29). Another interesting aspect of the *icfG* cluster is the presence of several genes similar to those in the *rsb* gene cluster in *Bacillus subtilis* (38). The *rsb* gene cluster in *B.subtilis* contains two serine kinase genes (*rsbT* and *rsbW*), two protein phosphatase genes (*rsbU* and *rsbX*) and two genes (*rsbS* and *rsbV*) encoding substrates of serine kinases and phosphatases. The interaction among these molecules, through a partner-switching mechanism, is required for cell response to environmental stress (38). IcfG shares sequence similarity with RsbU and RsbX, slr1861 is similar to RsbT and RsbW, and slr1856 and slr1859 are similar to RsbS and RsbV. The genetic organization of the *icfG* cluster is thus reminiscent of the *rsb* gene cluster in *B.subtilis*, and a similar mechanism could be postulated for the regulation of glycogen catabolism by the *icfG* cluster in *Synechocystis* sp. PCC 6803.

The gene cluster (slr0144 to slr0152) containing the Ser/Thr kinase gene slr0152 encodes several proteins of functional importance in cyanobacteria. slr0148 and slr0151 encode proteins similar to ferredoxin II (*fdxB*) and I (*petF1*), respectively. slr0149 is similar to the allophycocyanin alpha chain from the same strain (slr2067, ApcA) as well as that from other cyanobacteria. slr0149 and slr2067 are similar in size, but are only distantly related, with 20% identity and 41% similarity.

Another interesting gene cluster is the one involving the Ser/Thr kinase gene sll0776. This cluster contains 8 ORFs (sll0775–sll0782) in a region of ~10 kb. The polypeptides encoded by this gene cluster include an ABC transporter subunit (sll0778), a hybrid histidine kinase (sll0779), and a transcription factor with a helix–turn–helix DNA binding motif (sll0782).

The protein Ser/Thr kinase gene slr1225 is separated by ~13 kb from slr1234 encoding a homolog of the protein kinase C interacting protein 1. Thirty five kb further downstream of slr1234 is another Ser/Thr kinase gene slr1443. The close linkage between slr1982 and slr1983 has already been discussed above. About 13 kb downstream of slr1983 is sll1387 which encodes another protein phosphatase (see above).

Clustering of genes encoding Ser/Thr kinases or Ser/Thr phosphatases with those encoding two-component systems

Although eukaryotic-type protein kinases and phosphatases are found in several prokaryotes, in most cases it is still unclear how these molecules are involved in bacterial signal transduction. One possibility is that they participate in signal transduction through a cascade of Ser/Thr and Tyr phosphorylation/dephosphorylation, in a way similar to that which takes place in eukaryotes (9,10). The missing links in this possibility are protein Tyr kinase receptors and G-protein coupled receptors for signal transmission across the membrane. Several protein kinases and phosphatases analyzed here possess transmembrane segments, and thus may fulfil the task of membrane receptors, although how they transmit signals downstream remains unknown.

Another possibility would be the coupling of some Ser/Thr kinases and phosphatases to two-component systems, with sensors of two-component systems acting as membrane receptors. This is the case for ethylene response in the plant *Arabidopsis thaliana* (39) and for high osmolarity adaptation in the yeast *S.cerevisiae* (40). In both cases, a two-component system acts upstream of a cascade of Ser/Thr kinases in the same signal transduction pathway. It is thus very interesting to notice that several genes encoding Ser/Thr kinases or phosphatases in *Synechocystis* sp. PCC 6803 are found in the same cluster as those encoding members of two-component systems; or as in the case of slr1983, the same protein contains both a response regulator domain and a protein Ser/Thr phosphatase domain. The slr0114 cluster, for example, contains slr0114 encoding a protein Ser/Thr phosphatase and slr0115 encoding a DNA-binding response regulator. Similarly, the Ser/Thr kinase gene slr1697 is followed immediately by sll0921 which encodes a response regulator, although these two genes are transcribed in opposite directions. The sll0776 cluster is also in the same situation, since it contains one Ser/Thr kinase gene (sll0776) and one hybrid His kinase gene (sll0779).

In prokaryotes, genes involved in the same cellular process are frequently clustered or form an operon. It could thus be expected that at least some Ser/Thr kinases or phosphatases may interact with two-component regulatory proteins encoded by the same gene cluster. The elucidation of such molecular interaction will provide a new mechanism of signal transduction in bacteria. It will also advance our understanding of signal transduction in eukaryotes as well since more and more two-component systems are also being discovered in various eukaryotic organisms (2). Even in *A.thaliana* and *S.cerevisiae* in which the coupling between a two-component system and Ser/Thr kinases has already been

suggested (2,39,40), how this coupling is accomplished at the molecular level still remains unclear.

DISCUSSION

The discovery of eukaryotic-type protein kinases or phosphatases in many bacterial strains raises one fundamental question about the origin of these enzymes in evolution. Genes encoding such enzymes are either genuine prokaryotic ones, or they were recruited during evolution from eukaryotic organisms through horizontal gene transfer. The second possibility could account for the origin of at least one bacterial Ser/Thr kinase (5), namely the protein kinase *YpkA* from *Yersinia pseudotuberculosis* (43). However, most of those Ser/Thr kinases found so far in the two cyanobacterial strains *Anabaena* sp. PCC 7120 and *Synechocystis* sp. PCC 6803 are likely to be true prokaryotic enzymes. Genes encoding Ser/Thr kinases or phosphatases in *Synechocystis* sp. PCC 6803 are, in most cases, scattered over the entire chromosome, which is difficult to ascribe to one or a few horizontal gene transfer events. In addition, most of the protein kinases are more related to those from *Anabaena* sp. PCC 7120 or from other bacteria such as *Myxococcus xanthus* and *Mycobacteria leprae*, than those from eukaryotic organisms (data not shown, see also 44). These observations suggest that there is a bacterial lineage of evolution for most of the Ser/Thr kinases found in bacteria so far. These arguments, together, are in favor of the possibility that Ser/Thr kinases or phosphatases existed before the divergence of eukaryotes and prokaryotes in evolution. However, since the origin and evolution of the different kingdoms of living organisms are still in much heated debate (45), such conclusions should be treated with caution at the moment.

Given that eukaryotic-type signaling proteins are only recently found in some bacterial species (5,6), it will be of considerable interest to study their function in *Synechocystis* sp. strain PCC 6803 by taking the advantage of the availability of the whole genome sequence information. In addition, *Synechocystis* sp. strain PCC 6803 represents a unique genetic model for the study of photosynthesis among all organisms whose genomes have been sequenced so far, as this strain is easy to manipulate genetically and its photosynthetic systems are very similar to those of higher plants.

ACKNOWLEDGEMENTS

We are indebted to Dr S.Tabata's team at the Kazusa DNA Research Institute for the sequence analysis of the genome of *Synechocystis* sp. which made this study possible and for their timely effort to recheck the sll1574–75 region. We would like to thank Professor Françoise Joset and Dr Sylvie Bédu (CNRS, Marseille) for their critical reading of the manuscript, and Professor Jean-François Lefèvre (ESBS) for his constant support of our work. We also thank one anonymous reviewer for very constructive suggestions. The work at our laboratory was supported by the Ministère de l'Education Nationale, de la Recherche et de la Technologie, and the ESBS at the Université Louis Pasteur de Strasbourg.

REFERENCES

1 Stock,J.B., Stock,A.M. and Motten,J.M. (1990) *Nature*, **344**, 395–400.
2 Alex,L.A. and Simon,M.I. (1994) *Trends Genet.*, **10**, 133–138.
3 Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y., Miyajima,N., Hirosawa,M., Sugiura,M., Sasamoto,S., Kimura,T., Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K., Okumura,S., Shimpo,S., Takeuchi,C., Wada,T., Watanabe,A., Yamada,M., Yasuda,M. and Tabata,S. (1996). *DNA Res.*, **3**, 109–136.
4 Mizuno,T., Kaneko,T. and Tabat,S. (1996) *DNA Res.*, **3**, 407–414.
5 Zhang,C.-C. (1996) *Mol. Microbiol.*, **20**, 9–15.
6 Kennelly,P.J. and Potts,M. (1996) *J. Bacteriol.*, **178**, 4759–4764.
7 Hunter,T. and Plowman,G.D. (1997) *Trends Biochem. Sci.*, **22**, 18–22.
8 Zolnierowicz,S. and Hemmings,B.A. (1996) *Trends Cell Biol.*, **6**, 359–362.
9 Hunter,T. (1995) *Cell*, **80**, 225–236.
10 Hanks,S.K. and Hunter,H. (1995) *FASEB J.*, **9**, 576–596.
11 Taylor,S.S., Radzio-Andelm,E. and Hunter,T. (1995) *FASEB J.*, **9**, 1255–1266.
12 Hanks,S.K., Quinn,A.M. and Hunter,T. (1988) *Science*, **241**, 42–52.
13 Barford,D. (1996) *Trends Biochem. Sci.*, **21**, 407–412.
14 Fauman,E.B. and Saper,M.A. (1996) *Trends Biochem. Sci.*, **21**, 413–417.
15 Nakamura,Y., Kaneko,T., Hirosawa,M., Miyajima,N. and Tabata,S. (1998) *Nucleic Acids Res.*, **26**, 63–67.
16 Zhang,M., Van Etten,R.L. and Stauffacher,C.V. (1994) *Biochemistry*, **31**, 11097–11105.
17 Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) *Nucleic Acids Res.*, **22**, 4673–4680.
18 Neer,E.J., Schmidt,C.J., Nambudripad,R. and Smith,T.F. (1994) *Nature*, **371**, 297–300.
19 Zhang,C.-C. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 11840–11844.
20 Zhang,C.-C. and Libs,L. (1998) *Mol. Gen. Genet.*, **258**, 26–33.
21 Mueller,P.R., Coleman,T.R., Kumagai,A. and Dunphy,W.G. (1995) *Science*, **270**, 86–90.
22 Johnson,L.N., Noble,M.E.M. and Owen,D.J. (1996) *Cell*, **85**, 149–158.
23 Black,K., Buikema,W.J. and Haselkorn,R. (1995) *J. Bacteriol.*, **177**, 6440–6448.
24 Garrido,M.C., Herrero,M., Kolter,R. and Moreno,F. (1988) *EMBO J.*, **7**, 1853–1862.
25 Williamson,M.P. (1994) *Biochem. J.*, **297**, 249–260.
26 Zhang,C.-C., Friry,A. and Peng,L. (1998) *J. Bacteriol.*, **180**, 2616–2622.
27 Wilbank,S.M. and Glazer,A.N. (1993) *J. Biol. Chem.*, **268**, 1226–1235.
28 Adler,E., Donella-Deana,A., Arigoni,F., Pinna,L.A. and Stragier,P. (1997) *Mol. Microbiol.*, **23**, 57–62.
29 Beuf,L., Bédu,S., Durand,M.C. and Joset,F. (1994) *Plant Mol. Biol.*, **25**, 855–864.
30 Boyd,A.W., Kendall,K. and Simon,M.I. (1983) *Nature*, **301**, 623–626.
31 Aravind,L. and Ponting,C.P. (1997) *Trends Biochem. Sci.*, **22**, 458–459.
32 Clegg,C.H., Cadd,G.G. and McKnight,G.S. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 3703–3707.
33 Lima,C.D., Klein,M.G., Weinstein,I.B. and Hendrickson,W.A. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 5357–5362.
34 Bustos,S.A., Schaefer,M.R. and Golden,S.S. (1990) *J. Bacteriol.*, **172**, 1998–2004.
35 Lima,C.D., Klein,M.G. and Hendrickson,W.A. (1997) *Science*, **278**, 286–290.
36 Ponting,C., Schultz,J. and Bork,P. (1997) *Trends Biochem. Sci.*, **22**, 296–298.
37 Yang,H., Liu,M.Y. and Romeo,T. (1996) *J. Bacteriol.*, **178**, 1012–1017.
38 Yang,X., Kang,C.M., Brody,M.S. and Price,C.W. (1996) *Genes Dev.*, **10**, 2265–2275.
39 Chang,C., Kwok,S.F., Bleecker,A.B. and Meyerowitz,E.M. (1993) *Science*, **262**, 539–544.
40 Maeda,T., Wurgler-Murphy,S.M. and Saito,H. (1994) *Nature*, **369**, 242–245.
41 Faux,M.C. and Scott,J.D. (1996) *Cell*, **85**, 9–12.
42 Mochly-Rosen,D. and Gordon,A.S. (1998) *FASEB J.*, **12**, 35–42.
43 Galyov,E.E., Hakansson,S., Forsberg,A. and Wolf-Watz,H. (1993) *Nature*, **361**, 730–732.
44 Peirs,P., De Wit,L., Braibant,M., Huygen,K. and Content,J. (1997) *Eur. J. Biochem.*, **244**, 604–612.
45 Pennisi,E. (1998) *Science*, **280**, 672–674.