Eukaryotic-type protein kinases in *Streptomyces coelicolor*: variations on a common theme

Kateřina Petříčková and Miroslav Petříček

Laboratory of Physiology and Genetics of Actinomycetes, Institute of Microbiology ASCR, Vídeňská 1083, 14220 Prague, Czech Republic

The increasing number of genes encoding eukaryotic-type Ser/Thr protein kinases (ESTPKs) in prokaryotes, identified mostly due to genome-sequencing projects, suggests that these enzymes play an indispensable role in many bacterial species. Some prokaryotes, such as *Streptomyces coelicolor*, carry numerous genes of this type. Though the regulatory pathways have been intensively studied in the organism, experimental proof of the physiological function of ESTPKs is scarce. This review presents a family portrait of the genes identified in the sequence of the *S. coelicolor A3(2)* genome. Based on the available experimental data on ESTPKs in streptomyces and related bacteria, and on computer-assisted sequence analyses, possible roles of these enzymes in the regulation of cellular processes in streptomyces are suggested.

Overview

Ser/Thr/Tyr-specific protein phosphorylation represents one of the fundamental regulatory mechanisms in eukaryotes. Though, generally, His/Asn phosphorylation plays a more important role in prokaryotes, some groups of bacteria were recently shown to employ both systems. The first Ser/Thr protein kinase (ESTPK; stands for eukaryotic-type Ser/Thr protein kinase) gene was isolated from *Myxococcus xanthus* by Munoz-Dorado *et al.* (1991). Since then, numerous genes have been identified in diverse bacterial species thanks to genome sequencing projects and other studies. Recent data show that even though ESTPKs are not as widely and universally utilized in bacteria as in eukaryotes, their more or less conserved homologues may be traced across the prokaryotic world (Kennelly, 2002). Certain bacteria (*Mycobacterium, Streptomyces* and some cyanobacteria) harbour numerous representatives of these enzymes. Their common feature seems to be a complex life cycle, including morphological and physiological differentiation, sophisticated cell communication, or interactions with host cells.

The genome of *Streptomyces coelicolor* A3(2) (Bentley *et al.*, 2002) was screened for the presence of ESTPK genes in the present study. This organism is an extensively studied model of bacterial differentiation that to some extent resembles eukaryotes (mycelial growth, sporulation and secondary metabolite production). Though 34 putative ESTPK genes can be revealed in the genome sequence and some of them have already been reported, their roles remain quite unclear. Only two examples, *afsK* and *ramC*, have been shown to be involved in the antibiotic production regulatory cascade (Matsumoto *et al.*, 1994; Umeyama & Horinouchi, 2001; Lee *et al.*, 2002) or aerial hyphae formation (Hudson *et al.*, 2002; O’Connor *et al.*, 2002), respectively. The role of some others in differentiation of *S. coelicolor* still remains speculative (Umeyama *et al.*, 2002; Petrickova *et al.*, 2000). The functions of some ESTPKs have been investigated in other streptomycete species: Pkg2 is required for aerial mycelium formation in *Streptomyces granaticolor* (Nadvornik *et al.*, 1999) and StoPK-1 of *Streptomyces toyocaensis* NRRL 15009 influences the oxidative stress response in connection with glucose metabolism (Neu *et al.*, 2002).

The aim of this review is to compile all available ESTPK-related experimental data in actinomycetes with the results of computer-assisted analysis of the *S. coelicolor* genome and also other bacterial genomes. Based on this, we suggest possible roles for these enzymes in the regulation of cellular processes in streptomyces.

The *S. coelicolor* A3(2) genome carries 34 putative ESTPK genes

Based on the screening of the chromosome sequence data (Bentley *et al.*, 2002) using the standard BLAST program (Altschul *et al.*, 1990) with some already identified streptomycete ESTPKs (*AfsK, PkaA* and *PkaC*) as queries, we identified 37 suspected ESTPK genes with *P*(N) values below 0·1. Their predicted amino acid sequences were first analysed with reciprocal BLAST searches; these excluded the SCK13.20 gene, which was assigned as adenylate cyclase. Secondly, the remaining 36 were aligned with the eukaryotic consensus (Hardie & Hanks, 1995). As a criterion, conservation of all catalytic subdomains in the range common for the eukaryotic protein kinases, except subdomain I, was applied (see Fig. 1 and the supplementary figure in the online version of this paper at http://mic.sgmjournals.org). According to recent studies, subdomain I, which serves as...
the ATP-binding site (P-loop), may be substituted with different ATP-binding motifs (Shi et al., 1998). After the sequence inspections, two more candidates were excluded from the study: SCC24.21, with well-conserved subdomains I–VIB, but with a speculative following section; and SCC24.15c, with a totally missing subdomain VIII. The remaining 34 genes were numbered according to their order on the chromosome (PK01–PK34), disregarding their previous characterization (Table 1).

The arrangement of the ESTPK genes in the *S. coelicolor* chromosome is not even: the central part of the chromosome, encoding the essential life functions, contains the majority (PK01–PK34), disregarding their previous characterization (Table 1).

The length of the putative ESTPKs varies from 380 to 1557 amino acid residues. The catalytic domain, of typical length of 260–275 amino acid residues, is usually N-terminal. The catalytic domain of the PK23 kinase is interrupted between the VII and VIII subdomains by an Ala-rich insert (372 residues) of unknown function. The structure of PK30, the longest kinase, is exceptional; it contains two ESTPK catalytic domains in its N-terminal half and a putative C-terminal basic helix–loop–helix motif, which usually serves as a dimerization domain. A similar structure was reported in eukaryotes (Jones et al., 1988), but there is no evidence of a double ESTPK catalytic domain in a single protein in prokaryotes. However, a somewhat comparable multi-domain structure is common for some regulatory proteins of cyanobacteria. They combine features of ESTPKs and two-component system proteins in a single molecule (Ohmori et al., 2001).

**Fig. 1.** Consensus amino acid sequence of the catalytic domains in *S. coelicolor* ESTPKs. The first line shows the Hardie & Hanks’ eukaryotic consensus with essential amino acid residues (Hardie & Hanks, 1995). The second line represents the consensus amino acid sequence of the *S. coelicolor* ESTPKs. Numbers in parentheses within the *S. coelicolor* consensus show the length range of spacers between particular subdomains. The strength of the consensus (i.e. fraction of sequences conformable with the consensus) is indicated graphically below. The Hardie & Hanks’ consensus sequences are given according to the following code: uppercase letters, invariant residues; lowercase, nearly invariant residues; o, nonpolar residues; *, polar residues; +, small residues with near neutral polarity. Catalytic subdomains are indicated with roman numerals.
### Table 1. Summary of the properties of *Streptomyces coelicolor* A3(2) ESTPKs

<table>
<thead>
<tr>
<th>Gene no.</th>
<th>Gene code</th>
<th>Name*</th>
<th>Location (kbp)</th>
<th>Size (aa)</th>
<th>TM†</th>
<th>Extra domains</th>
<th>Nearby genes/encoded proteins‡</th>
<th>Proposed role</th>
</tr>
</thead>
<tbody>
<tr>
<td>PK01</td>
<td>SCL6.25c, SCO1468</td>
<td>PK01</td>
<td>1568</td>
<td>774</td>
<td>–</td>
<td>–</td>
<td>nusB, fmt, rpe, guaB2</td>
<td>RNA metabolism</td>
</tr>
<tr>
<td>PK02</td>
<td>SCL11.07, SCO1551</td>
<td>PK02</td>
<td>1659</td>
<td>493</td>
<td>T</td>
<td>–</td>
<td>coba, cobe, cbe, trpE</td>
<td>Cobalamin biosynthesis, amino acid biosynthesis</td>
</tr>
<tr>
<td>PK03</td>
<td>SCI11.13, SCO1724</td>
<td>PK03</td>
<td>1844</td>
<td>550</td>
<td>D</td>
<td>–</td>
<td>ABC transporter, sigma factor, exocellular hydrolases</td>
<td>Hydrolytic enzyme regulation</td>
</tr>
<tr>
<td>PK04</td>
<td>SC6E10.04, SCO2110</td>
<td>PK04</td>
<td>2267</td>
<td>667</td>
<td>T</td>
<td>–</td>
<td>PASTA Oxidoreductases, DNA repair endonucleases</td>
<td>DNA repair</td>
</tr>
<tr>
<td>PK05</td>
<td>SC1G2.06c, SCO2244</td>
<td>PK05</td>
<td>2415</td>
<td>686</td>
<td>D</td>
<td>WD-40</td>
<td>glnA</td>
<td>Metabolic regulation?</td>
</tr>
<tr>
<td>PK06</td>
<td>SC6D10.09, SCO2666</td>
<td>PK06</td>
<td>2900</td>
<td>903</td>
<td>–</td>
<td>–</td>
<td>Sugar transport system</td>
<td>Sugar uptake and degradation</td>
</tr>
<tr>
<td>PK07</td>
<td>SCE59.32c, SCO2973</td>
<td>PK07</td>
<td>3234</td>
<td>417</td>
<td>T</td>
<td>–</td>
<td>ftsX, ftsE, prfB, sbmB</td>
<td>Translation, cell-division regulation</td>
</tr>
<tr>
<td>PK08</td>
<td>SCE50.02c, SCO2974</td>
<td>PK08</td>
<td>3236</td>
<td>543</td>
<td>D</td>
<td>–</td>
<td>ftsX, ftsE, prfB, sbmB</td>
<td>Translation, cell-division regulation</td>
</tr>
<tr>
<td>PK09</td>
<td>SCE41.11c, SCO3102</td>
<td>PK09</td>
<td>3399</td>
<td>510</td>
<td>–</td>
<td>–</td>
<td>TRCF, cytochrome P450, Rpf-like growth factor</td>
<td>DNA repair, catabolism of glycans, growth regulation</td>
</tr>
<tr>
<td>PK10</td>
<td>SCE7.11, SCO3344</td>
<td>PK10</td>
<td>3700</td>
<td>720</td>
<td>D</td>
<td>PQQ</td>
<td>ilvD, proC</td>
<td>Amino acid biosynthesis</td>
</tr>
<tr>
<td>PK11</td>
<td>SC66T3.32c, SCO3621</td>
<td>PK11</td>
<td>4001</td>
<td>783</td>
<td>T</td>
<td>–</td>
<td>recR, sigma factors, gabT, asd2, ask</td>
<td>DNA recombination, nitrogen metabolism</td>
</tr>
<tr>
<td>PK12</td>
<td>SCGD3.21c, SCO3820</td>
<td>PK12</td>
<td>4200</td>
<td>522</td>
<td>T</td>
<td>–</td>
<td>ABC transporter, pyruvate dehydrogenase complex, quinone oxidoreductase</td>
<td>Energy metabolism: pyruvate/succinate dehydrogenase, electron transport</td>
</tr>
<tr>
<td>PK13</td>
<td>SCGD3.22, SCO3821</td>
<td>PK13</td>
<td>4201</td>
<td>556</td>
<td>D</td>
<td>PASTA</td>
<td>ABC transporter, pyruvate dehydrogenase complex, quinone oxidoreductase</td>
<td>Energy metabolism: pyruvate/succinate dehydrogenase, electron transport</td>
</tr>
<tr>
<td>PK14</td>
<td>SCH69.18, SCO3848</td>
<td>PK14</td>
<td>4234</td>
<td>673</td>
<td>T</td>
<td>–</td>
<td>ftsW, PBP, PP2C</td>
<td>Cell division</td>
</tr>
<tr>
<td>PK15</td>
<td>SCH69.30, SCO3860</td>
<td>PK15</td>
<td>4245</td>
<td>576</td>
<td>T</td>
<td>–</td>
<td>DNA-binding proteins</td>
<td>Transcription regulation?</td>
</tr>
<tr>
<td>PK16</td>
<td>SCD10.09, SCO4377</td>
<td>PK16</td>
<td>4790</td>
<td>580</td>
<td>T</td>
<td>–</td>
<td>mutT, FA synthase components</td>
<td>DNA repair, FA metabolism</td>
</tr>
<tr>
<td>PK17</td>
<td>SC6F11.21, SCO4423</td>
<td>PK17</td>
<td>4839</td>
<td>799</td>
<td>–</td>
<td>PQQ</td>
<td>afsR, afsS, kbpA</td>
<td>Differentiation, secondary metabolism</td>
</tr>
<tr>
<td>PK18</td>
<td>SC6D9.01, SCO4481</td>
<td>PK18</td>
<td>4900</td>
<td>632</td>
<td>D</td>
<td>Sugar-b.</td>
<td>ccaA</td>
<td>Receptor kinase (sugar signals), energy metabolism</td>
</tr>
<tr>
<td>PK19</td>
<td>SCD69.07, SCO4487</td>
<td>PK19</td>
<td>4905</td>
<td>592</td>
<td>D</td>
<td>LamGL</td>
<td>ubiDAX operon</td>
<td>Energy metabolism</td>
</tr>
<tr>
<td>PK20</td>
<td>SCD69.08, SCO4488</td>
<td>PK20</td>
<td>4907</td>
<td>626</td>
<td>T</td>
<td>Sugar-b.</td>
<td>ubiDAX operon</td>
<td>Receptor kinase (sugar signals), energy metabolism</td>
</tr>
<tr>
<td>PK21</td>
<td>SCD35.14, SCO4507</td>
<td>PK21</td>
<td>4927</td>
<td>586</td>
<td>D</td>
<td>–</td>
<td>FA synthesis, scoF2 cold shock, ftsK</td>
<td>FA synthesis, cell division, cold shock</td>
</tr>
<tr>
<td>PK22</td>
<td>SCD63.07, SCO4775</td>
<td>PK22</td>
<td>5189</td>
<td>717</td>
<td>T</td>
<td>–</td>
<td>Metabolic genes, sigma factor</td>
<td>Nucleotide, sugar metabolism</td>
</tr>
<tr>
<td>PK23</td>
<td>SCD63.08, SCO4776</td>
<td>PK23</td>
<td>5192</td>
<td>979</td>
<td>T</td>
<td>–</td>
<td>Metabolic genes, sigma factor</td>
<td>Nucleotide, sugar metabolism</td>
</tr>
<tr>
<td>PK24</td>
<td>SCD63.09, SCO4777</td>
<td>PK24</td>
<td>5195</td>
<td>599</td>
<td>T</td>
<td>–</td>
<td>Metabolic genes, sigma factor</td>
<td>Nucleotide, sugar metabolism</td>
</tr>
<tr>
<td>PK25</td>
<td>SCD63.10, SCO4778</td>
<td>PK25</td>
<td>5196</td>
<td>380</td>
<td>T</td>
<td>–</td>
<td>Metabolic genes, sigma factor</td>
<td>Nucleotide, sugar metabolism</td>
</tr>
<tr>
<td>PK26</td>
<td>SCD63.11, SCO4779</td>
<td>PK26</td>
<td>5198</td>
<td>548</td>
<td>D</td>
<td>–</td>
<td>Metabolic genes, sigma factor</td>
<td>Nucleotide, sugar metabolism</td>
</tr>
<tr>
<td>PK27</td>
<td>SC2A6.02c, SCO4817</td>
<td>PK27</td>
<td>5248</td>
<td>452</td>
<td>D</td>
<td>PG-binding</td>
<td>gbsA, mdh, folD</td>
<td>Receptor kinase (peptidoglycan signals), osmotic adaptation</td>
</tr>
<tr>
<td>PK28</td>
<td>SC2A6.05c, SCO4820</td>
<td>PK28</td>
<td>5251</td>
<td>712</td>
<td>T</td>
<td>SLT</td>
<td>gbsA, mdh, folD</td>
<td>Osmotic adaptation</td>
</tr>
<tr>
<td>PK29</td>
<td>SCK13.03, SCO4911</td>
<td>PK29</td>
<td>5344</td>
<td>670</td>
<td>D</td>
<td>Solute-binding</td>
<td>Sigma factor, afsQ operon</td>
<td>Differentiation, secondary metabolism</td>
</tr>
</tbody>
</table>
The amino acid sequences of the catalytic domains suggest in most of the proteins their Ser/Thr specificity (see the supplementary figure in the online version of this paper at http://mic.sgmjournals.org), though for some (AfsK, PK17 in the study) dual specificity, i.e. Ser/Thr together with Tyr, has been reported (Matsumoto et al., 1994). Only two of the S. coelicolor ESTPKs, PK30 (in its second ESTPK domain) and PK31, resemble the eukaryotic Tyr-specific consensus to some extent, mainly in subdomain VI. PK23 and the first domain of PK30 deviate from both groups (see the supplementary figure at http://mic.sgmjournals.org).

The catalytic domains of the S. coelicolor putative ESTPKs were aligned together with those of Mycobacterium tuberculosis, the most related bacterium whose entire genome has been sequenced, and several representatives of eukaryotic ESTPKs. The phylogenetic tree based on the alignment is shown in Fig. 3. Generally, the presented kinases tend to group in an origin-specific manner (shown in different colours), though only some groups (in dashed ellipses) are well defined by the bootstrap test (i.e. bootstrap values >70 %). The tree revealed two putative pairs of homologues from S. coelicolor and M. tuberculosis, PknB–PK14 and PknL–PK04. Further inspection of their domain architectures and relevant gene regions showed the striking similarity of the first pair, PknB and PK14, which is discussed later in the review.

Table 1. cont.

<table>
<thead>
<tr>
<th>Gene no.</th>
<th>Gene code</th>
<th>Name*</th>
<th>Location (kbp)</th>
<th>Size (aa)</th>
<th>TM</th>
<th>D</th>
<th>Extra domains</th>
<th>Proposed role†</th>
<th>Nearby genes/encoded proteins‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>PK30</td>
<td>SCO1F2.23, SC0606b</td>
<td>7348</td>
<td>1557 D</td>
<td></td>
<td></td>
<td></td>
<td>ATP/GTP-binding proteins</td>
<td>Secondary metabolism</td>
<td>abA-like locus, IS oxoeroductase</td>
</tr>
<tr>
<td>PK31</td>
<td>SCO2A3.31, SCO0681</td>
<td>7420</td>
<td>930 D</td>
<td></td>
<td></td>
<td></td>
<td>RamC</td>
<td>Differentiation, energy metabolism</td>
<td>Possible bacterial regulator</td>
</tr>
<tr>
<td>PK32</td>
<td>SCO2F9.13, SCO07240</td>
<td>7631</td>
<td>565 D</td>
<td></td>
<td></td>
<td></td>
<td>PknA</td>
<td>Respiratory genes</td>
<td>KLIC</td>
</tr>
<tr>
<td>PK33</td>
<td>SCO2A12.07, SCO07240</td>
<td>8047</td>
<td>745 D</td>
<td></td>
<td></td>
<td></td>
<td>RamC</td>
<td>Stress genes, hydrolases</td>
<td>Keratin-like protein, IS oxidoreductase</td>
</tr>
<tr>
<td>PK34</td>
<td>SCO2A12.07, SCO07240</td>
<td>8097</td>
<td>538 T</td>
<td></td>
<td></td>
<td></td>
<td>RamC</td>
<td>Stress genes, hydrolases</td>
<td>Keratin-like protein, IS oxidoreductase</td>
</tr>
</tbody>
</table>

*Previously reported genes: afsK (Matsumoto et al., 1994; Ueda et al., 1996; afsA and pkaB (Urabe & Ogawara, 1995); pkaD–pkaG (Ogawara et al., 1999); pkaH–pkaI (Ogawara et al., 1999); pkaH–pkaI (Bentley et al., 2002); et al., 1994); afsL (Umeyama et al., 2002); ramC (O'Connor et al., 2002); PK30 (Petrikovska et al., 2000).
†Prediction of transmembrane (TM) helices: T, TMHMM-predicted (http://www.cbs.dtu.dk/services/TMHMM-2.0/); D, additional, DAS-predicted (http://www.sbc.su.se/~miklos/DAS); H, no transmembrane helices predicted.
‡FA, fatty acid; TRCF, transcription-repair coupling factor; PBP, penicillin-binding protein; PP2C, protein phosphatase, 2C-type; IS oxidoreductase, iron–sulphur (2Fe–2S) oxidoreductase. For abbreviations of the extra protein domains see the Fig. 4 legend.

Fig. 2. Arrangement of the putative ESTPK genes along the S. coelicolor A3(2) linear chromosome. The central part of the chromosome, shown in black, has been reported to carry mainly ‘house-keeping’ genes, whereas both arms, in grey, encode mostly adaptive functions. The size scale (in Mbp) and position of the oriC origin of replication are indicated. The chromosome scheme is adapted from the work of Bentley (2002).
The majority of the kinases contain additional functional domains, mostly non-enzymic

In addition to the ESTPK catalytic domain, several other domains can be predicted in the ESTPK sequences. Thus 27 contain possible membrane-spanning regions, suggesting their membrane localization. Several ESTPKs carry binding domains for various substrates, such as the ricin B sugar-binding domain (PF00652), the bacterial peptidoglycan-binding domain 1 (PF01470), the bacterial extracellular solute-binding domain 3 (PF00497) and the PASTA (PBP and serine/threonine kinase associated) \( \beta \)-lactam binding domain (PF03793). Several kinases encompass repetitive domains, some of them with predicted protein-interactive roles, e.g. WD-40 (PF00400) and PQQ (PF01011). PK28 is
Fig. 4. Domain structure analysis of ESTPKs from *S. coelicolor* A3(2). The programs SMART at http://smart.embl-heidelberg.de (Schultz et al., 1998), BLOCKS at http://www.blocks.fhcrc.org (Henikoff et al., 1999), InterProScan at http://www.ebi.ac.uk/interproscan.html (Mulder et al., 2002) and SBASE at http://hydra.icgeb.trieste.it/~kristian/SBASE (Vlahovicek et al., 2002) were used for domain predictions. Transmembrane helices predicted with TMHMM (version 2.0) software at http://www.cbs.dtu.dk/services/TMHMM-2.0/ (Sonnhammer et al., 1998) are shown as black dots; additional, predicted by the DAS prokaryotic protein topology server at http://www.sbc.su.se/~miklos/DAS (Cserzo et al., 1997), are shown as grey dots. The following domains are presented (domain accession nos in the PROSITE, Pfam, ProDom or SMART databases are indicated in parentheses): (a) ESTPK domains, black rectangles; (b) predicted transmembrane regions, black or grey dots; (c) enzymic SLT transglycosylase domain, a yellow rectangle; (d) binding domains, green rectangles or pentagons in the case of repetitive domains, including: ‘Sugar-b.’, ricin B sugar-binding domain (PF00652); ‘PG-b.’, bacterial peptidoglycan-binding domain 1 (PF01470); ‘EC solute-b.’, bacterial extracellular solute-binding domain 3 (PF00497); ‘PASTA’, PASTA β-lactam binding domain (PF03793); ‘ATP-b.’, ATP/GTP-binding site, P-loop (PS00017); (e) repetitive domains with putative protein–protein interaction roles (pentagons): ‘WD-40’, β-transducin repeat (PF00400); ‘PQQ’, bacterial PQQ repeat (PF01011); ‘KLC’, kinesin light-chain repeat (PD148673); (f) repetitive domains of unknown function, white rectangles together with the repetitive motifs; (g) other domains: ‘HLH’, dimerization domain of the eukaryotic basic helix–loop–helix type (PS00038); ‘LamGL’, LamG-like jellyroll fold domain (SM0560); (h) regions rich in a particular amino acid: proline (blue rectangles), and alanine, ‘Ala-rich’; and (i) regions of unknown function shared by some ESTPKs: SR1, SR2 and SR3.
the only enzyme that carries a recognizable additional enzymic domain, the SLT transglycosylation domain (PF01464), putatively involved in murein degradation. In addition to previously characterized protein domains, we have identified three types of protein regions, shared by some of the ESTPKs, designated SR1, SR2 and SR3 (discussed later). A schematic representation of the domain structures of the proteins is shown in Fig. 4.

**ESTPK function predictions**

**ESTPK topology – putative receptor kinases**

Two-thirds of the *S. coelicolor* ESTPKs are putative membrane proteins resembling the structure of eukaryotic receptor kinases. One of the membrane-associated ESTPKs, RamC, was recently reported to be required for aerial hyphae formation. Transcription of *ramC* is developmentally regulated and probably activated by the RamR response regulator (O’Connor *et al.*, 2002). RamC protein kinase activity is weak in *vitro*. It has been speculated that as a primary role it might phosphorylate different targets, e.g. lipopolysaccharides (Hudson *et al.*, 2002).

For the topology predictions, two algorithms were generally used (TMHMM and prokaryote-specific DAS), the results of which differ in some cases (for details see Fig. 4). A receptive function of some ESTPKs is supported by the presence of C-terminal ligand-binding domains, probably located outside the cell. Three membrane kinases (PK04, PK13 and PK14) carry a C-terminal PASTA domain, a small globular fold composed of three β-sheets and an α-helix. The domain is common to several penicillin-binding proteins and bacterial ESTPKs and is responsible for binding of β-lactam antibiotics and their peptidoglycan analogues. It is probable that it may act in sensing free peptidoglycan units as signals for cell wall biosynthesis (Yeats *et al.*, 2002). We can find support for the theory in the case of PK14 and its close homologue PknB of *M. tuberculosis* (Av-Gay *et al.*, 1999): both relevant genes are clustered with genes implicated in cell wall biosynthesis and with other regulatory genes involved in protein Ser/Thr phosphorylation/dephosphorylation (Fig. 5). The *pknB* gene is pertaining expressed in *M. tuberculosis* cells in *vitro* as well as in animal host cells (Av-Gay *et al.*, 1999). A highly similar architecture of the chromosomal regions together with the lack of a comparable cluster in the genomes of other, less related, prokaryotes (*Escherichia coli*, *Helicobacter pylori* and *Bacillus subtilis*) suggest the presence of a specific, cell-division-associated regulatory pathway common to the high-GC branch of the Gram-positive bacteria.

The other two PASTA-containing ESTPKs, PK04 and PK13, do not have close mycobacterial homologues. Considering their gene neighbours, they might be involved in DNA repair and energy metabolism regulation, respectively, in response to peptidoglycan compounds.

The PK27 kinase may be involved in the transmission of peptidoglycan signals too, since it carries a C-terminal peptidoglycan-binding domain I (PF01471). The domain probably has a general peptidoglycan-binding function and is found in a variety of enzymes implicated in bacterial cell wall degradation. The PK27-encoding gene is clustered with several genes encoding hypothetical secreted proteins and with the PK28-encoding gene. PK28 is a multifunctional protein kinase with a C-terminal SLT transglycosylase (PF01464). The SLT domain degrades murein by cleaving the 1,4-β-glycosidic bond between N-acetylmuramic acid and N-acetylglucosamine. Moreover, nearby genes are responsible for the biosynthesis of the osmoprotectant glycine betaine (Boch *et al.*, 1997). Thus in response to changes of the peptidoglycan structure caused by osmotic shock, PK27 may transmit a signal to the enzymes involved in osmotic adaptation. As possible results of the pathway activation, glycine betaine is synthesized and the cell wall is modified by the SLT transglycosylase activity of the activated PK28 (Fig. 6).

Two ESTPKs, PK18 and PK20, carry C-terminal, probably extracellular, carbohydrate-binding domains of the ricin type. Both kinases are also highly similar in their catalytic domains and their genes are situated close to each other in the chromosome (Fig. 7). The carbohydrate-binding domain, originally found in the legume lectin ricin, is present in many carbohydrate recognition proteins where it binds simple sugars, such as galactose or lactose (Hazes, 1996). Both kinases may thus respond to sugar signals. The third kinase-encoding gene (PK19) in the region may also be associated with sugar signals; its additional LamGL domain seems often to be linked with enzymes involved in carbohydrate metabolism. However, its exact function is not clear. Products of other surrounding genes, putatively responsible for cytochrome and quinone biosynthesis, may serve as possible targets of regulation by Ser/Thr phosphorylation. Moreover, expression of the PK18 gene is possibly developmentally regulated in a bldA-dependent manner, since it contains the TTA rare leucine codon (Leskiw *et al.*, 1991). The PK18, PK19 and PK20 kinases might be involved in the control of the respiratory chain in response to simple sugar availability in a developmental-stage-specific manner.

A bacterial extracellular solute-binding domain of the family 3 (PF00497) is present in PK29. In Gram-positive bacteria, such domains are part of receptor proteins that trigger or initiate translocation of the solute through the membrane by interaction with a specific transport system. In some cases, they also initiate sensory transduction pathways. The family 3 members seem to be specific for amino acids and opines (Tam & Saier, 1993). In the case of PK29, the domain is most similar (25% identity) to that of the GluR0 receptor of the glutamate-gated potassium channel of *Synechocystis* (Chen *et al.*, 1999). The surrounding genes encode enzymes responsible for degradation of carbon compounds and nucleotide interconversion. In addition, the afsQ1–Q2 two-component regulatory system genes, involved in antibiotic production, and a possible σ factor gene are located nearby. Though no transport systems that could
putatively be regulated by the PK29 in a glutamate-specific manner are encoded in the region, implication of the kinase in glutamate/other solute-mediated signal transduction in *S. coelicolor* is a subject for further investigation.

There are more putative ESTPKs in *S. coelicolor* that seem to be integrated in the membrane, but their predicted extracellular parts do not share any homology with known proteins. In some cases possible targets of their regulatory actions may be suggested based on the composition of regions around their genes (Table 1). As an example, PK07 (PkaB) and PK08 (PkaA), the genes of which lie in an operon followed by genes involved in translation (*prfB* encoding the chain release factor 2, and *smpB* encoding the small protein B) and cell division (*ftsX, ftsE*), may be involved in translation and cell division control in response to environmental stimuli.

**ESTPKs as organizers of signal transmission complexes**

Within the set of *S. coelicolor* ESTPKs, four contain repetitive domains, which are generally responsible for protein–protein interactions and assembly of multi-protein complexes: PK10 and PK17, bacterial PQQ repeat, PF01011; PK05, β-transducin WD-40 repeat, PF00400; and PK33, kinesin light-chain repeat, PD148673. All the repetitive motifs share a typical core repeat length (about 40 amino acid residues); in addition, PQQ and WD-40 form the same β-propeller structure. PQQ occurs in enzymes with pyrroloquinoline quinone as a cofactor, in Ser/Thr kinases, and in prokaryotic dehydrogenases (Oubrie et al., 1999). Though the PQQ-containing protein kinases are present in both prokaryotes and eukaryotes, their exact domain architecture varies according to the origin: in all prokaryotic
representatives the PQQ domain is situated in the C-terminus of the kinase molecule, whereas in eukaryotes it is always N-terminal (based on the SMART protein architecture database).

Some of the PQQ-containing ESTPKs in actinomycetes have already been studied. In S. coelicolor, the best-characterized kinase, AfsK (PK17 in the study), is necessary for antibiotic production, whereas in Streptomyces griseus (AfsK-g), it is conditionally needed for morphological differentiation (Umeyama et al., 1999). AfsK activity is regulated by the KbpA protein, which binds to the catalytic domain of AfsK and inhibits its autophosphorylation (Umeyama & Horinouchi, 2001). The active form of AfsK phosphorylates the AfsR global regulator, which is needed for the biosynthesis of actinorhodin and undecylprodigiosin. Recent work has shed more light on the mechanism of regulation. The phosphorylated form of AfsR binds efficiently to a promoter region of afsS and initiates its transcription (Lee et al., 2002). The product of the gene then, directly or indirectly, activates transcription of actII-ORF4, the actinorhodin-pathway-specific transcriptional activator (Umeyama et al., 2002).

In S. granaticolor, three ESTPKs containing a PQQ domain have been characterized, but none of them seems to be a close homologue of those found in S. coelicolor. Pkg2 is probably involved in aerial mycelium development (Nadvornik et al., 1999). The genes encoding the next two, Pkg3 and Pkg4, are organized in an operon. Their physiological role is not clear (Vomastek et al., 1998).

Similarly to the PQQ repeats, the WD-40 repetitive domain forms a propeller structure (Smith et al., 1999). The domain is characteristic of the β-subunits of trimeric G-proteins and many other eukaryotic regulatory proteins; its presence in prokaryotic proteins is quite rare. According to the gene region content, the only S. coelicolor gene encoding an ESTPK containing a WD-40 domain (PK05) may be involved in metabolic regulation.

The last of the protein-interactive domains, the KLC repeats (present in PK33), were first identified in the light chain of the eukaryotic kinesin motor protein. In the multimeric kinesin molecule they are involved in the coupling of a cargo to the heavy-chain subunits and in the modulation of its Fig. 6. Tentative model of the roles of PK27 and PK28 in osmoprotection. Osmotic changes alter the peptidoglycan (PG) architecture of the cell wall, CW (1), which serves as a signal received by the peptidoglycan-binding domain (PGb) of PK27 (2). The signal reception causes activation of the PK27 ESTPK catalytic domain (K27). As a result, the genes responsible for glycine betaine biosynthesis (gbs) are expressed (3), probably via activation of specific transcription factors, and the osmo-protective glycine betaine is produced (4). Next, the active PK27 activates (by phosphorylation?) the kinase domain of PK28 (K28), which subsequently causes activation of the extracellular SLT domain (5). The SLT transglycosylase modifies the peptidoglycan in a way that improves cell wall resistance to osmotic pressure (6).

Fig. 7. PK18–19–20 chromosomal region. Genes encoding PK18 and PK20 are shown by patterned arrows. The region contains genes encoding the following products: ccdA, similar to cytochrome-c-type biogenesis protein; ccsA, possible cytochrome c assembly protein; MCM, possible transferase, MCM family signature (PS00847); SH, possible secreted hydrolase; ubiD, probable 3-octaprenyl-4-hydroxybenzoate carboxy-lyase; ubiA, probable 4-hydroxybenzoate octaprenyl-transferase; ubiX, probable octaprenyl carboxylase; DPR, putative DNA polymerase related protein; AcT, putative acetyltransferase; glutP, putative proton transport protein – glutamate transporter. Genes encoding transcription regulators are shown by bold grey arrows, transmembrane proteins by small grey arrows, and unknown (similar to hypothetical) proteins by small black arrows. Genes containing the TTA codon are indicated with asterisks.
ATPase activity (Gauger & Goldstein, 1993). It was also shown that the protein-binding abilities of KLC are regulated by phosphorylation (Ichimura et al., 2002). Considering the gene organization around the gene encoding the KLC-containing PK33, the kinase might be involved in the regulation of respiration and electron transport. Phosphorylation of the KLC repetitive domain by the kinase domain may modulate its interactions with protein ligands.

What the exact physiological functions of the protein-interactive domains in ESTPKs are remains to be elucidated. Some hints were given in the case of the gene pkwA from the thermophilic actinomycete Thermomonospora curvata. The gene encodes the WD-40-containing ESTPK, which is probably membrane-associated. It was shown that the WD-40 domain of PkwA itself could be phosphorylated by an unknown kinase activity in the T. curvata cell-free extracts and also by the Pkg2 kinase of S. granaticolor. However, the presence of the PkwA catalytic domain prevents phosphorylation in vitro (Joshi et al., 2000). How the phosphorylation of the WD-40 domain affects kinase function has not been investigated yet. Our latest experiments have revealed that pkwA gene expression is developmentally regulated and is strictly associated with the early exponential growth phase. In young T. curvata mycelium, PkwA is predominantly present in the form of high-molecular-mass protein complexes (unpublished). Among the genes near pkwA are those putatively encoding subunits of DNA polymerase III and DNA helicases. Both enzyme families are certainly involved in replication and may thus be required in the fast-growing cells in the early exponential phase.

Other domains that may give ESTPKs extra functions

Many ESTPKs contain regions that are particularly rich in proline residues (over 30% of all amino acid residues) (see Fig. 4). Proline-rich domains are often discussed in connection with their protein-binding abilities. Specific proline-rich motifs interact with the eukaryotic WW, SH3 and other protein-interactive domains (Einbond & Sudol, 1996; Kay et al., 2000). They are frequently involved in signalling pathways of eukaryotes. It is possible that the proline-rich regions may also specify functions of the bacterial ESTPKs by coupling them to their targets or regulators.

Some of the identified ESTPKs share regions that do not fit to any protein pattern and motif in the databases. In the set of the S. coelicolor ESTPKs we have identified three types, SR1 (over 200 amino acid residues), SR2 (about 60 residues) and SR3 (about 100 residues), seen in Fig. 4. Their functions are clearly speculative. SR1 may serve as an extracellular (signal?)-binding domain. Three of four SR1-encoding ESTPK genes (PK22, PK23 and PK24) lie in a group at the chromosome and might also be products of gene duplication. SR2, present in PK11 and PK21, closely follows the ESTPK domain, so it may be involved in substrate binding or kinase activity regulation.

Implications of the content of the ESTPK gene region

As was mentioned before, the character of surrounding genes may provide clues to a predicted possible role of an ESTPK. Considering their possible targets of regulation, they might control membrane transport systems (ABC transporters and other permeases), biosynthetic pathways (primary or secondary), energy metabolism, cell division, stress response and differentiation. While conveying their signals, they may interact with other ESTPKs, protein phosphatases, members of two-component systems, transcription regulators and $\sigma$ factors (see also Table 1).

PK04 (PkaF) may be given as an example. Just downstream of pkaF we found a gene similar to many DNA-repair endonuclease genes. Its protein product follows the consensus of the AP endonuclease class, i.e. apurinic or apyrimidinic site-specific DNA lyases. Based on it we hypothesize that the kinase might be involved in the DNA-repair or stress response. Moreover, PkaF contains the extracellular PASTA peptidoglycan-binding domain that may sense stress-induced peptidoglycan changes of the cell wall as was discussed before in the case of PK14. Association of cell-wall-related genes with oxidative stress defence has been discussed in other bacteria (Thibassard et al., 2002). Recently, the PkaF gene homologue StoPK-1 was characterized in S. toyocaensis. Disruption of the gene causes increased sensitivity to oxidative stress and the kinase activity is needed for the wt phenotype. It was also shown that PkaF of S. coelicolor could fully substitute StoPK-1 (Neu et al., 2002). Taking the information together, it is conceivable that the oxidative-stress-protective role of PkaF/StoPK-1 is mediated by the control of the AP endonuclease and the signal transduction pathway affects DNA-repair control.

For details on other ESTPK genes see Table 1.

Evolutionary insights into the presence of ESTPKs in prokaryotes

In 1991, Munoz-Dorado and coworkers first reported a eukaryotic-type Ser/Thr protein kinase, Pkn2, in bacteria and broke the traditional classification of protein kinases as eukaryotic Ser/Thr/Tyr-specific and prokaryotic His/Asn-specific (Munoz-Dorado et al., 1991). Since then, many bacterial Pkn2-type kinases have been characterized, and over a hundred others have been identified thanks to genome-sequencing projects in many bacterial species, including archaea. However, it is important to note that these enzymes are not universally spread within the prokaryotes and their occurrence is limited to particular bacterial groups. It seems that they are dispensable for the accomplishment of the ‘basic’ prokaryotic way of life, represented by E. coli, but they are required for cell differentiation, a multicellular lifestyle, complex secondary metabolism, circadian cycling, pathogenicity, etc. (Ogawara et al., 1999). Perhaps, derived from a common ancestor, they
became involved in the prokaryotic regulatory pathways only when the more complex lifestyle demanded them. According to recently available data, two prokaryotic groups exhibit high numbers of Pkn2-type ESTPK genes in their genomes: cyanobacteria and actinomycetes. Other bacterial species usually do not contain more than just two or three such genes (e.g. Bacillus, Chlamydia, Salmonella, Yersinia, etc.) or do not have them at all (Escherichia, Neisseria, Borrelia, etc.). It should be taken into account that the choice of organisms for genome-sequencing projects affects the overall image of ESTPK distribution in prokaryotes.

Recently, new groups of putative ESTPK homologues, ABC1, RIO1, piD261, AQ578 and others, were assigned in bacterial genomes and eukaryotes (Leonard et al., 1998; Shi et al., 1998; Ponting et al., 1999; Kennelly, 2002). Though they are evidently more distant from the Hardie & Hanks' consensus sequence, they seem to share all the structural features with eukaryotic protein kinases. In the case of piD261 of Saccharomyces cerevisiae, the protein kinase activity was clearly shown (Stoccheto et al., 1997); other cases remain hypothetical. Recently, RIO1 group members were shown to act as lipopolysaccharide kinases in Gram-negative bacteria (Krupa & Srinivasan, 2002). Thus it is quite probable that some of these kinases predominantly phosphorylate non-protein targets. Consistently, aminoglycoside phosphotransferases, responsible for inactivation of aminoglycoside-type antibiotics, also show striking structure similarity with ESTPKs, though the sequence match is extremely low (Wright, 1999). As well as their primary enzymic activity they exhibit Ser protein kinase activity, too (Daigle et al., 1999). None of the putative ESTPKs that we identified in S. coelicolor fit to the new ESTPK groups as characterized by Leonard (1998). It seems probable that with further accumulation of genome data other atypical kinase classes will be recognized.

Interestingly, it seems that even closely related organisms, such as different species of a single genus, do not contain identical sets of ESTPK genes. Available genome sequence data in Mycobacterium (avium, bovis, tuberculosis, paratuberculosis and leprae) reveal 9–11 ESTPK genes. Of these, about half are common to all species, but the rest are species-specific. Similarly, three PQQ-containing ESTPKs (Pkg2–4) identified in S. granaticolor do not have close homologues in S. coelicolor. S. coelicolor contains only two PQQ-containing ESTPKs, and their catalytic domain sequences, exact domain structures and gene area arrangements are different. Apparently, ESTPK involvement in prokaryotic regulatory pathways has evolved in a species-specific manner, presumably to satisfy individual lifestyle demands.

Comparison of sets of ESTPK genes found in the genomes of S. coelicolor and the cyanobacterium Nostoc supports the idea of species-specific design of ESTPKs. Both prokaryotic organisms have comparably large genomes (8·7 and 10 Mbp, respectively), containing numerous ESTPKs, and both have incredibly sophisticated life cycles. Nostoc is a nitrogen-fixing filamentous cyanobacterium which produces several specialized cell types and often establishes endosymbiosis with the Geosiphon pyriforme fungus (Castenholz & Waterbury, 1989). Streptomycetes also exhibit hyphal growth, differentiate into several cell types, and produce a vast number of secondary metabolites, extracellular signals and enzymes. The genome of Nostoc carries over 50 putative ESTPK genes. As in S. coelicolor, about 70% of them probably encode transmembrane enzymes. However, their functional domain architecture differs remarkably. None of the additional functional domains found in S. coelicolor kinases is present in the set of Nostoc kinases. The PQP and WD-40 protein interactive domains of S. coelicolor are substituted with TPR repeats (PF00515) in Nostoc. Many of the Nostoc kinases are huge proteins with a multidomain structure typical of cyanobacterial regulatory proteins and carry GAF (PF01590), His kinase A phosphoacceptor (PF00512), histidine-kinase-like ATPase (PF02518), PAS (PF00989) and PAC (PF00785) domains in a single molecule (Ohmori et al., 2001). Thus the overall characteristics of the ESTPK group are different in each species, as are the demands for regulatory circuit functions.

Concluding remarks

This review collates experimental data on ESTPKs of S. coelicolor and other related actinomycetes with computer analysis of the S. coelicolor genome sequence data. The experimental evidence indicates that Ser/Thr kinases are involved in cell differentiation, antibiotic production and stress-response regulation (Umeyama et al., 2002). Computer analysis of the S. coelicolor genome data led us to assign putative roles to previously uncharacterized ESTPKs encoded by the genome.

Comparing the data with those available in M. tuberculosis (Av-Gay & Everett, 2001), we found only one pair of almost identical genes. This finding was done with respect to the level of identity of the ESTPK domains, the presence of additional functional domains and the relevant gene regions. Both PknB of M. tuberculosis and PK14 of S. coelicolor are probably involved in cell growth and division regulation and both relevant genes lie close to the origin of chromosome replication. Other S. coelicolor ESTPKs are more distant from mycobacterial representatives, though some common characteristics can be found, such as the presence of PQP domains and proline-rich regions.

It is difficult to ascribe exact functions to particular ESTPK genes without analyses of relevant mutants. Interestingly, no ESTPK genes were detected by classical genetic screening techniques, including mutagenesis. This may be explained by either their indispensability for life or their putative functional redundancy. The latter cause was in some cases supported by the fact that the disruption of particular ESTPK genes does not cause any distinct phenotype changes (Petrickova et al., 2000; Vomastek et al., 1998). In spite of that, when we consider their abundance in the genome, their
frequent involvement in the regulation of cell processes in Streptomyces is quite likely. The presence of 34 putative genes in the genome, representing about 0.5% of coding sequences, cannot be explained just by incidental horizontal gene transfer. As a comparison, one of the smallest eukaryotic genomes, that of Saccharomyces cerevisiae, contains 113 ESTPK genes, representing 2% of all its coding sequences (Hunter & Plowman, 1997).

Another question may rise: do the ESTPKs functionally replace ordinary bacterial two-component systems in some prokaryotes? Table 2 compares the numbers of Ser/Thr- and His-specific protein-kinase-encoding genes in the genomes of some selected bacteria: S. coelicolor, M. tuberculosis, Nostoc punctiforme, Anabaena sp. PCC 7120, Pseudomonas aeruginosa and Bacillus subtilis. The data do not show any obvious correlation between the genome size and the number of two-component systems or ESTPKs. We did not find any decrease in the number of two-component system His kinases in those organisms possessing numerous ESTPKs. In general, we suppose that Ser/Thr phosphorylation has not developed in certain bacteria just to replace two-component systems. It provides new regulatory circuits to control various cell processes based on the particular needs of a species. Moreover, both systems are probably tightly connected, as was shown in eukaryotes employing ESTPK pathways in osmotic regulation and the plant hormone response (Loomis et al., 1997).

Summarizing recent findings, it has become evident that ESTPKs play as indispensable a role in some prokaryotic organisms as they do in eukaryotic cells. Newly emerging experimental and genomic data provide clues to their functions, which help us to suggest the probable areas of action of ESTPKs in streptomycetes. Nevertheless, the complete picture of their exact physiological roles remains to be revealed.

Acknowledgements

We wish to thank Dr Petr Kotlik for the assistance in phylogenetic tree constructions. The work is supported by the Grant Agency of the Czech Republic (204/02/D120) and the Grant Agency of the ASCR (IAA5020207).

References


Eukaryotic-type protein kinases in S. coelicolor


