Sequences and evolutionary analyses of eukaryotic-type protein kinases from *Streptomyces coelicolor* A3(2)

Hiroshi Ogawara, Narumi Aoyagi, Mami Watanabe and Hiroaki Urabe

Author for correspondence: Hiroshi Ogawara. Tel: +81 424 95 8474. Fax: +81 424 95 8474.
e-mail: hogawara@my-pharm.ac.jp

Department of
Biochemistry, Meiji
Pharmaceutical University,
Noshio-2, Kiyose, Tokyo
204-8588, Japan

Four eukaryotic-type protein serine/threonine kinases from *Streptomyces coelicolor* A3(2) were cloned and sequenced. To explore evolutionary relationships between these and other protein kinases, the distribution of protein serine/threonine kinase genes in prokaryotes was examined with the TFASTA program. Genes of this type were detected in only a few species of prokaryotes and their distribution was uneven; *Streptomyces*, *Mycobacterium*, *Synechocystis* and *Myxococcus* each contained more than three such genes. Homology analyses by GAP and Rdf2 programs suggested that some kinases from one species were closely related, whilst others were only remotely related. This was confirmed by examining phylogenetic trees constructed by the neighbour-joining and other methods. For each species, analysis of the coding regions indicated that the G+C content of protein kinase genes was similar to that of other genes. Considered with the fact that in phylogenetic trees the amino acid sequences of STPK from *Aquifex aeolicus* and some other eukaryotic-type protein kinases in prokaryotes form a cluster with protein kinases from eukaryotes, this suggests that the eukaryotic-type protein kinases were present originally in both prokaryotes and eukaryotes, but that most of these genes have been lost during the evolutionary process in prokaryotes because they are not needed. This conclusion is supported by the observation that the prokaryotes retaining several of these kinases undergo complicated morphological and/or biochemical differentiation.

Keywords: eukaryotic-type protein kinase, *Streptomyces*, protein serine/threonine kinase, evolution, phylogenetic tree

INTRODUCTION

Protein phosphorylation is one of the most important regulatory mechanisms in the signal transduction systems of eukaryotes as well as prokaryotes. However, the main phosphorylation sites in prokaryotes are on histidine and aspartic acid residues (Parkinson & Kofoed, 1992), whereas in eukaryotes they are on serine, threonine and/or tyrosine (Hunter & Cooper, 1985; Hanks *et al.*, 1988). Numerous eukaryotic protein kinase genes have been cloned. Although these kinases are classified into several groups on the basis of their amino acid sequences, a series of 11 subdomains (I–XI) have

been conserved (<http://www.sdsc.edu/kinases/>). Recently, eukaryotic-type protein kinases have been detected in a small number of prokaryotes, including *Streptomyces* (Matsumoto *et al.*, 1994; Urabe & Ogawara, 1995; Hirakata *et al.*, 1998), *Myxococcus* (Munoz-Dorado *et al.*, 1991; Zhang *et al.*, 1992) and *Mycobacterium* (Fsihi *et al.*, 1996; Cole *et al.*, 1998).

Since complete nucleotide sequences of the chromosomal DNA of over 10 bacterial species have now been determined (Doolittle, 1998), it would be interesting to know whether these contain eukaryotic-type protein kinases, and if so, how such kinases are distributed. This paper describes the distribution of eukaryotic-type protein kinases in prokaryotes and analyses evolutionary relationships among these kinases in prokaryotes and eukaryotes. It also reports the nucleotide sequences of four newly isolated protein kinase genes from *Streptomyces coelicolor* A3(2).

Abbreviations: cAMPPK, cAMP-dependent protein kinase; CaMPK, calmodulin-dependent protein kinase; S6-2PK, S6-2 protein kinase.

The GenBank accession numbers for the sequences determined in this work are AB016932, AB018799, AB019394 and AB021679.

METHODS

Bacterial strains, plasmids and growth conditions. *Escherichia coli* K-12 JM109 was used as a host for pBluescript II SK(+) and its derivatives. *E. coli* and pUC119 were purchased from Takara Shuzo. *E. coli* was cultivated in 2 × YT medium (Sambrook *et al.*, 1989). *S. coelicolor* A3(2) M145 was used as a DNA donor. It was cultured in liquid YEME medium (Hopwood *et al.*, 1985).

DNA manipulations. Preparation of plasmid DNA and *in vitro* DNA manipulations were carried out as described by Sam-

brook *et al.* (1989). Enzymes for DNA manipulations were obtained from Takara Shuzo.

Nucleotide sequence determination. Nucleotide sequences were determined by the method of Sanger *et al.* (1977) with the Thermosequenase II kit (Amersham Pharmacia Biotech).

Search for protein kinases. Eukaryotic-type protein kinases were investigated by two methods. First, the word 'kinase' was searched for in the GenBank data available for the genomes of 16 species (Table 1) for which nucleotide sequences have already been determined (<http://www.genome.ad.jp/kegg/java/org-list.html>). The amino acid sequences of the

Table 1. Distribution of eukaryotic-type protein kinases in bacteria

Based on detection of subdomains I–XI using the GAP program; + indicates a similarity value greater than 40 and – indicates a similarity value less than 29.

Category	Species	Genome size (nt)	Protein kinase	PkaA	PkaB	Pkn2	cAMPPK	CaMPK	S6-2PK
Proteobacteria	<i>Escherichia coli</i>	4639221	–	–	–	–	–	–	–
	<i>Haemophilus influenzae</i>	1830135	–	–	–	–	–	–	–
	<i>Helicobacter pylori</i>	1667867	HPO432	–	–	+	–	–	–
Gram-positive bacteria	<i>Bacillus subtilis</i>	4214814	YloP	+	+	+	+	+	+
			YbdM	–	–	+	+	+	–
	<i>Mycoplasma genitalium</i>	580073	MG109	–	+	+	+	+	+
	<i>Mycoplasma pneumoniae</i>	816394	ORF389	–	+	+	+	+	+
	<i>Mycobacterium tuberculosis</i>	4411529	PknA	+	+	+	+	+	+
			PknB	+	+	+	+	–	+
			PknD	+	+	+	+	+	+
			PknE	+	+	+	+	+	+
			PknF	+	+	+	+	+	+
			PknG	–	–	–	+	+	+
			PknH	+	+	+	+	–	+
			PknI	+	+	+	–	–	–
			PknJ	+	+	+	+	–	+
			PknK	+	–	+	+	+	+
			PknL	+	+	+	+	+	+
			PknM	–	–	–	–	–	–
Chlamydia	<i>Chlamydia trachomatis</i>	1042519	Pkn5	+	+	+	+	–	+
			PknD	+	+	+	+	–	+
Spirochetes	<i>Borrelia burgdorferi</i>	910724	–	–	–	–	–	–	–
	<i>Treponema pallidum</i>	1138011	–	–	–	–	–	–	–
O ₂ -reducing bacteria	<i>Aquifex aeolicus</i>	1551335	STPK	+	–	+	+	+	+
Cyanobacteria	<i>Synechocystis</i> sp.	3573470	SLL1575	+	+	+	+	+	–
			SLR1225	+	–	–	–	–	+
			SLR1443	–	–	+	–	–	+
			SLR1697	+	–	+	–	+	+
			SLR0152	+	+	+	+	+	+
			SLL0782	–	–	–	–	–	–
			SLL0776	–	–	–	+	–	–
			SLR0599	+	+	+	+	+	+
Archaea	<i>Methanococcus jannaschii</i>	1664987	–	–	–	–	–	–	–
	<i>Methanobacterium thermoautotrophicum</i>	1751377	–	–	–	–	–	–	–
	<i>Archaeoglobus fulgidus</i>	2178400	–	–	–	–	–	–	–
	<i>Pyrococcus horikoshii</i>	1738505	PHCJ009	–	–	+	+	–	–

Table 2. Protein kinases used in this work

Protein kinase	Species	Number of amino acids	Accession number
PkaA	<i>Streptomyces coelicolor</i>	543	D86821
PkaB	<i>Streptomyces coelicolor</i>	417	D86821
PkaD	<i>Streptomyces coelicolor</i>	598	AB016932
PkaE	<i>Streptomyces coelicolor</i>	487	AB018799
PkaF	<i>Streptomyces coelicolor</i>	667	AB019394
PkaG	<i>Streptomyces coelicolor</i>	592	AB021679
AfsK	<i>Streptomyces coelicolor</i>	799	D45382
Pks3	<i>Streptomyces coelicolor</i>	565	U19250
Pkn2	<i>Myxococcus xanthus</i>	830	M94857
Pkn5	<i>Myxococcus xanthus</i>	380	U40656
Pkn6	<i>Myxococcus xanthus</i>	710	U40656
cAMPPK	<i>Saccharomyces cerevisiae</i>	380	M17073
CaMPK	<i>Mus musculus</i>	469	X58995
S6-2PK	<i>Homo sapiens</i>	735	L07597
HPO432	<i>Helicobacter pylori</i>	296	AE000511
YloP	<i>Bacillus subtilis</i>	648	AL009126
YbdM	<i>Bacillus subtilis</i>	256	AL009126
MG109	<i>Mycoplasma genitalium</i>	362	L43967
ORF389	<i>Mycoplasma pneumoniae</i>	389	U00089
PknA	<i>Mycobacterium tuberculosis</i>	431	AL123456
PknB	<i>Mycobacterium tuberculosis</i>	626	AL123456
PknD	<i>Mycobacterium tuberculosis</i>	664	AL123456
PknE	<i>Mycobacterium tuberculosis</i>	566	AL123456
PknF	<i>Mycobacterium tuberculosis</i>	476	AL123456
PknG	<i>Mycobacterium tuberculosis</i>	750	AL123456
PknH	<i>Mycobacterium tuberculosis</i>	626	AL123456
PknI	<i>Mycobacterium tuberculosis</i>	585	AL123456
PknJ	<i>Mycobacterium tuberculosis</i>	589	AL123456
PknK	<i>Mycobacterium tuberculosis</i>	1110	AL123456
PknL	<i>Mycobacterium tuberculosis</i>	399	AL123456
CTPkn5	<i>Chlamydia trachomatis</i>	490	AE001273
CTPknD	<i>Chlamydia trachomatis</i>	934	AE001273
STPK	<i>Aquifex aeolicus</i>	271	AE000657
SLL1575	<i>Synechocystis</i> sp.	416	AB001339
SLR1225	<i>Synechocystis</i> sp.	495	AB001339
SLR1443	<i>Synechocystis</i> sp.	614	AB001339
SLR1697	<i>Synechocystis</i> sp.	574	AB001339
SLR0152	<i>Synechocystis</i> sp.	631	AB001339
SLL0776	<i>Synechocystis</i> sp.	505	AB001339
SLR0599	<i>Synechocystis</i> sp.	535	AB001339
PHCJ009	<i>Pyrococcus horikoshii</i>	219	Pyro_h

kinases detected were compared to determine whether they really contained eukaryotic-type protein kinase domains. Second, the 16 nucleotide sequences and the amino acid sequences of subdomains I–IX of PkaA, PkaB, Pkn2 and cAMP-dependent protein kinase (cAMPPK) from *Saccharomyces cerevisiae*, calmodulin-dependent protein kinase (CaMPK) from *Mus musculus* and S6-2 protein kinase (S6-2PK) from *Homo sapiens* were inspected thoroughly by the TFasta program for the presence of protein kinases (Pearson & Lipman, 1988). The protein kinases investigated in

this paper are listed in Table 2, which includes some protein kinases other than those detected by the above procedure.

Homology analyses. GAP (Needleman & Wunch, 1970; Lipman & Pearson, 1985) and Rdf2 (Pearson & Lipman, 1988) were used to compare pairs of proteins. FASTA (Pearson & Lipman, 1988) was used for comparisons of more than three proteins.

Phylogenetic trees. Phylogenetic trees were constructed mainly by a neighbour-joining method (CLUSTAL W;

Thompson *et al.*, 1994). MOLPHY (a maximum-likelihood method; Adachi & Hasegawa, 1992), PHYLIP (a maximum-parsimony method; Felsenstein, 1995) and ODEN (a neighbour-joining method; Ina, 1991) were also used.

RESULTS

Nucleotide sequences

In a previous paper, we reported the cloning of seven DNA fragments containing putative protein serine/threonine kinase genes from *S. coelicolor* A3(2) (Hirakata *et al.*, 1998). The nucleotide sequences of four of them, *pkaD*, *pkaE*, *pkaF* and *pkaG*, were determined in

this work and are available in DDBJ/GenBank/EMBL, accession numbers AB016932, AB018799, AB019394 and AB021679, respectively. Preceding the translation start sites of these genes, Shine-Dalgarno-like sequences were detected (GCGGGG, AGGGCG, ACGCGG and GG-GAGG, respectively). Each of the kinases contained the 11 protein serine/threonine kinase subdomains I–XI predicted for eukaryotic-type enzymes. Kinase PkaD, corresponding to Sch5 in the previous paper, consisted of 598 amino acids; PkaE, corresponding to Sch1, consisted of 487 amino acids; PkaF, corresponding to Sch3, consisted of 667 amino acids; and PkaG, corresponding to Sch4, consisted of 592 amino acids. At

Table 3. Relationship between pairs of eukaryotic-type protein kinase sequences analysed by the Rdf2 program

Values represent standard deviations above the mean score of all the comparisons. The analysis was carried out with a ktup of 2 and a cutoff of 28.

	PkaB	PkaD	PkaE	AfaK	Pks3	Pkn2	Pkn5	Pkn6	cAMPPK	CaM PK	S6 2PK	HPO432	YloP	YbdM	MG109	ORF389	PknA	PknB	PknD	PknE
PkaA	53.64	29.99	25.03	36.32	24.61	36.72	25.56	32.93	22.21	32.41	33.78	4.91	35.16	14.14	12.01	15.17	34.51	43.01	21.74	44.39
PkaB		27.11	26.16	24.56	28.59	36.14	15.48	29.61	36.51	37.88	38.88	4.21	58.29	12.98	9.26	9.26	31.29	35.71	27.79	53.61
PkaD			28.23	30.54	42.78	31.25	31.33	25.78	28.94	33.18	26.92	-0.54	47.73	11.19	5.8	6.21	31.02	34.82	38.31	36.66
PkaE				36.82	35.03	40.61	39.17	30.61	23.72	25.79	19.41	1.25	77.57	14.73	10.92	9.92	41.28	45.31	48.77	66.47
AfaK					21.52	43.46	21.44	23.85	15.31	14.27	26.15	5.47	48.04	11.28	9.74	14.83	31.53	39.32	23.01	24.69
Pks3						31.75	25.97	40.13	22.11	29.66	13.71	0.18	54.94	13.51	14.36	15.72	27.48	27.57	34.26	51.04
Pkn2							22.48	21.15	29.28	25.37	20.41	8.32	46.48	20.75	18.71	25.71	35.47	41.82	35.21	37.52
Pkn5								21.38	14.61	27.16	19.48	5.39	25.39	3.23	16.14	18.06	34.15	31.71	19.28	20.91
Pkn6									21.88	14.09	23.71	3.14	61.77	6.11	14.36	23.71	40.23	47.56	45.69	46.31
cAMPPK										60.21	75.11	2.21	30.81	14.73	13.09	5.02	35.12	37.48	18.35	39.92
CaM PK											33.01	4.63	27.69	15.38	11.93	14.02	19.48	21.91	14.74	17.34
S6 2PK												3.61	39.72	11.65	16.37	2.33	17.14	25.08	18.65	22.43
HPO432													6.11	2.97	1.01	-1.41	4.61	4.26	0.66	1.37
YloP														16.58	9.67	10.59	63.57	69.59	45.11	49.61
YbdM															5.55	0.15	13.43	12.84	13.81	16.06
MG109																89.53	3.99	11.31	13.48	10.11
ORF389																	4.88	14.95	17.02	6.87
PknA																		55.51	46.01	62.21
PknB																			57.91	54.42
PknD																				84.69
PknE																				
PknF																				
PknG																				
PknH																				
PknI																				
PknJ																				
PknK																				
PknL																				
PknM																				
CTPkn5																				
CTPknD																				
STPK																				
SLL1575																				
SLR1225																				
SLR1443																				
SLR1697																				
SLL0152																				
SLL0782																				
SLL0776																				
SLR0599																				

least two of these enzymes, PkaE and PkaG, were shown to have protein serine/threonine and/or tyrosine kinase activity (unpublished data).

Distribution of eukaryotic-type protein kinases in bacteria

Investigation of the protein kinase sequences in the complete nucleotide and/or amino acid sequences of 16 bacterial genomes determined at the present time detected the eukaryotic-type enzyme in only a few species; however, in some species, there were more than three protein kinases present (Tables 1 and 2). Among the proteobacteria, YjbH (accession no. U00096) in *E. coli* is reported to have a protein kinase ATP-binding signature. However, none of the subdomains of protein kinases could be detected, so we conclude that there is

no eukaryotic-type protein kinase gene in *E. coli*. From similar inspections, *Haemophilus influenzae* does not possess any protein kinase, and the one (HPO432) in *Helicobacter pylori* is of the eukaryotic-type.

Myxococcus xanthus, a Gram-negative soil bacterium, has more than five protein kinases (Zhang *et al.*, 1996), and among Gram-positive bacteria, *Bacillus subtilis* has two (YloP and YbdM), whilst *Mycoplasma genitalium* and *Mycoplasma pneumoniae* each have one (MG109 and ORF389, respectively). However, within the high G + C group of Gram-positive bacteria, *Mycobacterium tuberculosis* and *S. coelicolor* A3(2) each have more than five eukaryotic-type protein kinases. *Mycobacterium leprae* is also reported to have more than three kinases (Fsihi *et al.*, 1996); *Chlamydia trachomatis* has two (Pkn5 and PknD), whereas spirochaetes (*Borrelia burg-*

Table 3 (cont.)

PknF	PknG	PknH	PknI	PknJ	PknK	PknL	PknM	CTPkn5	CTPknD	STPK	SLL1575	SLR1225	SLR1443	SLR1697	SLR0152	SLL0782	SLL0776	SLR0599	PHCJ009
30.33	9.61	33.89	16.31	40.31	29.02	30.62	0.37	16.24	11.68	17.02	27.56	19.22	7.23	22.31	44.22	0.51	14.27	16.85	4.95
30.47	5.85	55.47	20.03	33.09	22.28	25.89	-0.39	10.39	17.71	17.88	16.59	11.25	6.78	15.56	26.75	-2.02	14.26	17.46	9.05
34.64	10.22	42.65	19.94	51.12	43.98	30.02	-0.73	10.81	10.94	9.04	16.89	20.34	0.11	1.27	32.06	1.27	18.43	15.4	6.67
29.21	17.51	72.38	31.61	25.02	37.46	51.88	-0.64	22.17	27.82	15.04	23.54	21.22	8.01	30.92	30.61	1.06	18.3	31.45	18.23
28.26	17.68	36.16	10.37	30.14	21.93	46.37	-0.91	12.98	17.36	18.14	16.26	20.35	-0.11	15.94	21.71	0.6	18.23	17.87	5.84
28.57	18.54	38.37	14.94	25.11	35.52	29.57	0.21	20.62	10.61	20.99	19.19	20.66	0.16	16.72	33.71	0.32	16.51	32.31	10.99
44.96	13.08	44.51	24.28	31.55	28.44	30.71	-0.26	26.69	22.18	38.18	21.13	17.52	10.17	29.47	35.39	-0.58	19.23	23.67	10.67
24.59	8.82	24.37	10.11	37.43	20.06	24.41	-0.52	19.68	9.21	18.21	29.46	13.07	1.43	11.79	16.55	-0.51	7.98	19.71	4.41
27.76	12.79	52.39	25.13	44.19	30.06	23.08	-0.23	12.31	24.31	18.42	20.36	10.35	13.82	25.29	27.81	0.05	8.88	18.39	0.41
16.79	17.82	27.28	7.16	22.35	19.71	35.71	0.08	31.17	10.33	14.72	14.35	12.51	9.91	19.93	20.17	-0.95	15.82	17.71	8.07
20.57	10.65	10.62	9.14	23.68	24.14	21.69	-0.31	14.95	9.49	21.89	15.41	11.78	9.69	33.37	23.93	0.13	13.94	13.68	5.91
12.81	22.91	43.14	-0.57	39.08	32.31	19.17	-0.86	21.86	9.53	14.45	17.02	12.45	11.82	16.39	33.25	-1.08	18.65	30.38	8.03
-0.14	4.61	-0.11	0.64	0.38	11.72	4.72	-0.92	2.81	9.57	4.96	2.46	-0.4	-0.68	0.15	1.79	-1.54	3.81	-0.28	6.58
58.12	23.03	43.98	26.51	57.61	63.96	53.57	-0.14	17.96	34.65	26.88	24.57	21.16	7.54	27.01	46.73	-1.06	17.16	30.07	12.55
20.65	11.72	18.71	17.65	22.75	1.86	13.12	-1.24	7.14	12.81	4.18	4.11	2.15	26.74	15.07	5.31	-0.52	9.69	0.44	8.52
10.18	12.32	9.98	8.83	19.18	10.96	8.55	0.02	8.21	9.56	12.18	16.95	6.99	2.6	10.45	8.65	-0.21	7.35	13.01	5.23
12.43	8.81	10.43	11.24	11.41	5.51	5.13	-1.08	4.55	13.71	10.26	7.39	9.94	4.14	15.15	7.47	-0.65	9.15	10.13	0.38
34.92	15.03	56.72	12.11	31.18	60.28	38.72	-0.25	27.57	21.45	16.81	22.37	16.57	7.62	22.39	33.99	-0.58	11.59	45.91	8.46
46.67	15.89	54.28	28.35	60.36	38.17	46.95	1.26	34.15	26.18	28.33	35.11	25.81	0.77	21.66	33.05	-0.65	12.37	19.56	10.26
43.41	4.44	88.24	19.95	41.14	32.31	56.99	-0.16	22.85	20.63	31.43	12.54	16.74	10.56	13.76	27.68	-0.54	13.73	17.72	-0.66
52.76	6.48	88.25	24.91	58.01	38.24	44.62	-0.45	16.84	26.78	29.32	20.73	10.16	9.81	17.42	27.03	-0.21	37.81	18.77	-0.93
	6.61	62.06	41.12	123.38	41.51	49.54	0.81	17.17	32.17	15.52	17.56	21.71	5.76	22.98	21.18	-0.53	12.48	29.74	0.31
		10.91	8.79	14.96	16.38	15.23	0.81	17.23	8.45	12.09	2.89	18.28	3.71	18.89	11.44	0.16	8.87	10.21	8.95
			27.41	49.21	42.05	41.81	-0.53	22.49	24.89	19.77	16.12	16.07	-0.97	26.31	24.91	0.64	9.75	32.71	14.46
			36.25	17.36	18.03	3.91	14.47	18.59	15.14	5.83	12.05	0.01	17.96	13.99	1.06	2.05	16.04	1.31	
				47.08	30.62	-0.92	16.41	16.18	26.43	30.16	22.25	1.75	16.81	35.06	0.08	13.27	25.17	2.74	
					43.46	0.92	21.83	21.01	15.12	10.88	11.71	0.75	4.25	18.44	0.24	11.34	19.29	6.65	
						0.66	45.36	16.17	30.24	30.13	25.02	4.92	15.44	23.97	-0.58	13.41	24.99	9.51	
							1.75	2.14	0.68	1.53	-0.39	-0.38	-1.35	0.28	3.37	-1.17	-0.89	1.94	
								13.93	18.21	7.56	8.48	4.76	13.11	20.59	3.08	10.58	-0.53	10.59	
									3.89	10.83	4.84	7.68	14.69	10.43	-0.91	9.83	18.83	10.36	
										12.84	14.55	9.62	27.69	32.28	-1.22	20.21	16.51	6.86	
											11.67	1.86	12.34	23.73	0.69	7.13	15.92	0.28	
												11.13	60.07	0.04	0.71	37.27	63.15	5.71	
													33.61	5.79	-0.61	23.36	24.09	5.55	
															13.82	-0.21	65.68	63.15	0.72
																-0.93	18.66	15.23	8.68
																	-0.61	-0.13	0.25
																		94.43	-1.31
																			5.69

Fig. 1. Phylogenetic tree constructed on the basis of amino acid sequences of the conserved subdomains I–XI in 41 eukaryotic-type protein kinases using a neighbour-joining method (CLUSTAL W). The numbers at the branchpoints are bootstrap probabilities. The scale-bar indicates 0.1 substitutions per site.

Fig. 2. Phylogenetic tree constructed on the basis of amino acid sequences of the conserved subdomains I–XI in eight eukaryotic-type protein kinases from *S. coelicolor* A3(2) using a neighbour-joining method (CLUSTAL W). The numbers at the branchpoints are bootstrap probabilities. The scale-bar indicates 0.1 substitutions per site.

dorferi and *Treponema pallidum*) do not have any kinase gene. In the O_2 -reducing bacterium, *Aquifex aeolicus*, one kinase (STPK) was detected.

Among cyanobacteria, *Synechocystis* sp. and *Anabaena* sp. have more than five protein kinases (Zhang & Libs,

1998). Among archaea, only *Pyrococcus horikoshii* has one kinase-like protein (PHCJ009), although this shows very low homology with other protein kinases. Recently, the complete nucleotide sequence of chromosomal DNA in *Rickettsia prowazekii* has been determined (Andersson *et al.*, 1998), but no eukaryotic-type protein kinase gene was detected within it, although four histidine kinase genes were found.

Homology analyses

Table 1 shows the result of analysis by the GAP program using the amino acid sequences of subdomains I–XI. Protein kinases HPO432 from *Helicobacter pylori* and SLL0776 from *Synechocystis* sp. exhibited high similarity values with only one of the six protein kinases in the comparison, including those from eukaryotes; and kinase PknM from *Mycobacterium tuberculosis* showed no homology at all (Table 1). These proteins, except for SLL0776, lacked kinase subdomains VIa, VII and VIII. The Rdf2 analysis (Table 3) indicates that PknM from *Mycobacterium tuberculosis* scored very low for homology with 40 other protein kinases. Therefore, it is uncertain whether PknM really has kinase activity. Indeed, the corrected version of the PknM sequence excluded this protein from the list of eukaryotic-type protein kinases in *Mycobacterium tuberculosis* (Cole *et al.*, 1998). On the other hand, HPO432, SLL0776 and PHCJ009 gave high scores for homology with several kinases: HPO432 with PknK, CTPknD and PH009; SLL0776 with PkaD, AfsK, PknE, CTPknD, SLR1225, SLR1443, SLR1697 and SLR0152; and PHCJ009 with PkaE, YloP, PknH, PknL, CTPkn5 and CTPknD (Table 3). Interestingly, in FASTA analyses, STPK from *Aquifex aeolicus* showed higher homology scores with protein kinases from eukaryotes such as a serine/threonine kinase of *Caenorhabditis elegans* (*par-1*, accession no. U40858), human protein p78 (accession no. M80359) and a serine/threonine kinase of *Rattus norvegicus* (accession no. Z83868) than with protein kinases from prokaryotes. *par-1*, human protein p78 and the serine/threonine kinase of *Rattus norvegicus* showed 31.51% identity in a 238 amino acid overlap (Smith–Waterman score of 412), 30.83% identity in a 240 amino acid

Table 4. Distance matrix for amino acid sequences of eukaryotic-type protein kinases from *Streptomyces coelicolor* A3(2) computed by the PROTDIST program using the Dayhoff PAM model

The distance is scaled in units of expected fraction of amino acids changed.

	PkaA	PkaB	PkaG	AfsK	PkaE	PkaF	PkaD
PkaB	0.82972						
PkaG	1.39321	1.60986					
AfsK	1.38722	1.66052	0.72440				
PkaE	1.23367	1.24283	1.31829	1.45375			
PkaF	1.21327	1.30989	1.44339	1.56899	1.19901		
PkaD	1.43179	1.56128	1.51208	1.60317	1.53002	1.46337	
Pks3	1.43914	1.51712	1.83070	1.73382	1.39806	1.54905	1.4363

Table 5. G + C contents of eukaryotic-type protein kinases in various species

Species	Protein kinase	G + C content of coding region (mol %)*
<i>Streptomyces coelicolor</i>		(71.67)
	PkaA	75.51
	PkaB	77.54
	PkaD	74.81
	PkaE	75.36
	PkaF	73.26
	PkaG	77.20
	AfsK	76.31
<i>Myxococcus xanthus</i>	Pks3	65.31
		(68.58)
	Pkn2	71.45
	Pkn5	67.75
<i>Saccharomyces cerevisiae</i>	Pkn6	68.17
	cAMPPK	(39.73)
<i>Mus musculus</i>		40.96
	CaMPK	(52.76)
<i>Homo sapiens</i>		51.31
	S6-2PK	(52.85)
<i>Helicobacter pylori</i>		57.14
	HPO432	(39.65)
<i>Bacillus subtilis</i>		33.91
		(44.32)
	YloP	44.24
<i>Mycoplasma genitalium</i>	YbdM	41.93
		(31.74)
	MG109	32.61
<i>Mycoplasma pneumoniae</i>		(41.03)
	ORF389	40.79
<i>Mycobacterium tuberculosis</i>		(65.81)
	PknA	67.75
	PknB	63.31
	PknD	65.31
	PknE	65.72
	PknF	67.58
	PknG	66.13
	PknH	66.93
	PknI	67.35
	PknJ	65.68
	PknK	67.93
	PknL	65.16
<i>Chlamydia trachomatis</i>		(41.55)
	CTPkn5	42.99
	CTPknD	40.22
<i>Aquifex aeolicus</i>		(43.58)
	STPK	44.83
<i>Synechocystis</i> sp.		(48.66)
	SLL1575	49.76
	SLR1225	49.29
	SLR1443	50.98
	SLR1697	50.64
	SLR0152	49.34
	SLL0776	39.61
<i>Pyrococcus horikoshii</i>	SLR0599	54.45
		(42.31)
	PHCJ009	45.81

* Mean G + C contents of all the genes analysed to date in each species are shown in parentheses.

overlap (Smith–Waterman score of 392) and 29.83 % identity in a 238 amino acid overlap (Smith–Waterman score of 388) with STPK, respectively. The protein with the highest score in prokaryotes was SLR0152 of *Synechocystis* sp., but it showed 30.83 % identity in a 240 amino acid overlap (Smith–Waterman score of 363) and ranked only 35th in similarity scores. Consistent with this, STPK forms a cluster with eukaryotic protein kinases such as cAMPPK, S6-2PK and CaMPK in the phylogenetic tree. This cluster includes some other eukaryotic-type protein kinases from various species of prokaryotes, as described below.

Phylogenetic trees

Fig. 1 shows a phylogenetic tree constructed on the basis of 41 amino acid sequences of the conserved subdomains I–XI (excluding PknM) by using a neighbour-joining method (the CLUSTAL W program). Generally, protein kinases from the same species cluster in the phylogenetic tree, but this is not true of all them. For example, PkaA, PkaB, PkaE, PkaG and AfsK from *S. coelicolor* form a branch, but PkaD and Pks3 from the same species belong to different branches. Similarly, YbdM and YloP from *Bacillus* and CTPknD and CTPkn5 from *Chlamydia trachomatis* are located in different branches. Comparable results were obtained in phylogenetic trees constructed with other programs, including PROTML (a maximum-likelihood method) in the MOLPHY package, TREENJ (a neighbour-joining method) in the ODEN package and PROTPAR (a maximum-parsimony method) in the PHYLIP package (data not shown). As described above, STPK from *Aquifex aeolicus*, PknG from *Mycobacterium tuberculosis* and HPO432 from *Helicobacter pylori* were closely related to CaMPK, which is of eukaryotic origin. MG109 and ORF389 from *Mycoplasma* form a single branch, although they come from different species. The close relationship between MG109 and ORF389 is confirmed by the high value obtained from an Rdf2 analysis (Table 3).

Fig. 2 shows a phylogenetic tree of protein kinases from *S. coelicolor* A3(2). Although these kinases come from the same species, they are only distantly related. They show long distances after branching, indicating that they are now related only remotely having separated early and evolved independently of each other. This situation was confirmed by Rdf2 analysis (Table 3) and examination with the PROTDIST program using the Dayhoff PAM model (Table 4). Similar relationships were observed in the kinases from *Mycobacterium* and those from cyanobacteria. Phylogenetic trees constructed by other programs supported the conclusion reached above (data not shown).

Codon usage

Table 5 shows the mean G + C content of the genes analysed to date in various species (obtained from <http://www.dna.affrc.go.jp/>) and those of the coding region of the protein kinases. It indicates that almost all

the protein kinase genes showed G + C contents similar to those of other genes in the same species. In addition, the G + C contents in the first, second and third bases of codons of STPK are 47.15, 28.23 and 55.31%, respectively. These values are closer to the mean values of all the genes analysed so far in *Aquifex aeolicus* than to those in the genes in eukaryotes such as *Saccharomyces cerevisiae* and *Homo sapiens*. The mean values in *Aquifex aeolicus* are 50.44, 32.37 and 47.93%, respectively. Similar relationships were observed for other protein kinases. Furthermore, the G + C content in the third letter of codons in the protein kinase genes from *S. coelicolor* was over 90%, whilst the mean value of the genes analysed in the present work in *Streptomyces* is 92.82%. These results indicate that codon usage for these kinases follows that of other proteins in each species.

DISCUSSION

The nucleotide sequences of four newly isolated putative eukaryotic-type protein serine/threonine kinase genes from *S. coelicolor* A3(2) were determined. All of them contained 11 conserved subdomains, and at least two of them showed protein serine/threonine and/or tyrosine kinase activity (unpublished data). Although the number of amino acids in these and the other kinases from *S. coelicolor* differed appreciably (from 417 to 799), the number of amino acids in the subdomains was limited to 233 (PkaG and AfsK) to 249 (PkaA). On this basis, at least eight eukaryotic-type protein kinases were detected in *S. coelicolor* A3(2).

We were interested in determining whether these protein kinases are distributed ubiquitously in bacteria and archaea, and if so, what evolutionary relationship exists between those in prokaryotes and eukaryotes. Our search showed that the eukaryotic-type protein kinases have a restricted distribution in prokaryotes. On the other hand, almost all of the prokaryotes analysed here have histidine kinase genes. There are 30, 4, 4, 36, 14, 4, 3, 4, 32, 15, 24, 1 and 4 histidine kinase genes in *E. coli*, *Haemophilus influenzae*, *Helicobacter pylori*, *Bacillus subtilis*, *Mycobacterium tuberculosis*, *Borrelia burgdorferi*, *T. pallidum*, *Aquifex aeolicus*, *Synechocystis*, *Methanobacterium thermoautotrophicum*, *Archaeoglobus fulgidus*, *P. horikoshii* and *Rickettsia prowazekii*, respectively. The results suggest, therefore, that the eukaryotic-type protein kinases are not essential for primary metabolism in prokaryotes. In addition, the presence of eukaryotic-type protein kinases is not directly related to the classification of prokaryotes. In fact, it is hard to correlate the distribution of these kinase genes in prokaryotes with the classification of prokaryotes on the basis of 16S rRNA. For example, in archaea, which are believed to be more closely related to eukaryotes, only one eukaryotic-type protein kinase could be detected (PHCJ009 in *P. horikoshii*). Both *Myxococcus xanthus* and *E. coli* are Gram-negative bacteria, but while *Myxococcus xanthus* has more than five eukaryotic-type protein kinases, no such enzyme could be detected in *E. coli*. The common feature among

the bacteria that contain more than three kinase genes is that they exhibit various morphological forms and/or carry out biochemical reactions in response to changes in their environment. For example, *Myxococcus xanthus* forms a spore-filled fruiting body and migrates on semi-solid surfaces by gliding. *Streptomyces* undergoes complicated morphological differentiation into aerial mycelium and chain-like spores and produces secondary metabolites such as antibiotics. *Mycobacterium tuberculosis* and *Mycobacterium leprae* grow very slowly and form rudimentary mycelium under some conditions. *Synechocystis* has thylakoids thought to be the origin of chloroplasts in eukaryotes. Furthermore, among the proteins listed in Table 2, three from *Myxococcus xanthus* and five from *S. coelicolor* have protein serine/threonine and/or tyrosine kinase activity and engage in morphological and/or biochemical differentiation (Matsumoto *et al.*, 1994; Zhang *et al.*, 1996; Udo *et al.*, 1997; H. Ogawara & H. Urabe, unpublished data). Thus, we speculate that the eukaryotic-type protein kinases have been preserved only in these specific species during the course of evolution.

The phylogenetic tree (Fig. 1) constructed by comparing the amino acid sequences of consensus regions shows that the protein kinases from one species do form clusters as suggested by Zhang (1996), but not always. CTPknD and CTPkn5 from *Chlamydia trachomatis* belong to different branches and the protein kinases from *Mycobacterium tuberculosis* scatter divergently in the tree. Similar situations were observed for protein kinases from other bacteria, such as YbdM and YloP from *Bacillus subtilis* and those from *Synechocystis*. Furthermore, the substitution frequencies are very high after branching. This is more clearly shown in the tree constructed on the basis of amino acid sequences of protein kinases from a single species, *S. coelicolor* (Fig. 2); Rdf2 and PROTDIST analyses confirmed the high substitution frequencies within species (Tables 3 and 4). We suggest, therefore, that these eukaryotic-type protein kinases from prokaryotes are only remotely related to each other, even if they come from a single species. However, critical amino acid residues involved in the kinase reaction are almost always conserved.

The G + C contents of the genes of the protein kinases resemble the mean values of the genes from the same species, indicating that codon usage is retained. The only exceptions are Pks3 from *S. coelicolor* and SLL0776 from *Synechocystis*. The reason for the discrepancy is not yet clear. These genes might come from other species by horizontal transfer. In both cases, it is the third letter that differs most: the G + C contents of the first, second and third letters in Pks3 and SLL0776 are 68.67, 50.62 and 76.46; and 51.29, 40.41 and 27.13, respectively. The mean G + C contents are 71.75, 50.45 and 92.82 in *Streptomyces* and 56.08, 39.92 and 49.98 in *Synechocystis*, respectively. These results strongly suggest that the eukaryotic-type protein kinases were already present before the branching of eukaryotes from prokaryotes and that these kinases disappeared in some prokaryotes

because the enzymes were unnecessary. The alternative explanation, that the kinases were acquired in prokaryotes after the divergence of prokaryotes from eukaryotes, is less plausible, although a few kinases might be the result of horizontal transfer.

ACKNOWLEDGEMENTS

We thank M. Bibb of the John Innes Institute for providing us with *S. coelicolor* A3(2), M. Hasegawa and Y. Cao of the Institute of Statistical Mathematics for their valuable advice on MOLPHY analysis, and the National Institute of Genetics and the Institute of Medical Sciences, University of Tokyo, for permission to use their computers for genetic analyses. This work was supported in part by the special fund of the Meiji Pharmaceutical University.

REFERENCES

- Adachi, J. & Hasegawa, M. (1992). MOLPHY: programs for molecular phylogenetics. I. PROTML: maximum likelihood inference of protein phylogeny (Computer Science Monographs no. 27), Institute of Statistical Mathematics, Tokyo.
- Andersson, S. G. E., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponten, T., Alsmark, U. C. M., Podowski, R. M., Naslund, A. K., Erissson, A. S., Winkler, H. H. & Kurland, C. G. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133–140.
- Cole, S. T., Brosch, R., Parkhill, J. & 39 other authors (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **396**, 537–544.
- Doolittle, R. F. (1998). Microbial genomes opened up. *Nature* **392**, 339–342.
- Felsenstein, J. (1995). PHYLIP, version 3.5c, University of Washington, USA.
- Fsihi, H., De Rossi, E., Salazar, L. & 7 other authors (1996). Gene arrangement and organization in an ~ 76 kb fragment encompassing the *oriC* region of the chromosome of *Mycobacterium leprae*. *Microbiology* **142**, 3147–3161.
- Hanks, S. K., Quinn, M. & Hunter, T. (1988). The kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42–52.
- Hirakata, T., Kieser, H., Hopwood, D., Urabe, H. & Ogawara, H. (1998). Putative protein serine/threonine kinase genes are located in several positions on the chromosome of *Streptomyces coelicolor* A3(2). *FEMS Microbiol Lett* **159**, 1–5.
- Hopwood, D. A., Bibb, M. J., Chater, K. F. & 7 other authors (1985). *Genetic Manipulation of Streptomyces: a Laboratory Manual*. Norwich: John Innes Foundation.
- Hunter, T. & Cooper, J. A. (1985). Protein-tyrosine kinases. *Annu Rev Biochem* **54**, 897–930.
- Ina, Y. (1991). ODEN, molecular evolutionary analysis system for DNA and amino acid sequences. DDBJ, National Institute of Genetics, Japan.
- Lipman, D. J. & Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science* **227**, 1435–1441.
- Matsumoto, A., Hong, S. K., Ishizuka, H., Horinouchi, S. & Beppu, T. (1994). Phosphorylation of the AfsR protein involved in secondary metabolism in *Streptomyces* species by a eukaryotic-type protein kinase. *Gene* **146**, 47–56.
- Munoz-Dorado, J., Inouye, S. & Inouye, M. (1991). A gene encoding a protein serine/threonine kinase is required for normal development of *M. xanthus*, a gram-negative bacterium. *Cell* **67**, 995–1006.
- Needleman, S. B. & Wunch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**, 443–453.
- Parkinson, J. S. & Kofoed, E. (1992). Communication modules in bacterial signalling proteins. *Annu Rev Genet* **26**, 71–112.
- Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* **85**, 2444–2448.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463–5467.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific GAP penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.
- Udo, H., Inouye, M. & Inouye, S. (1997). Biochemical characterization of Pkn2, a protein Ser/Thr kinase from *Myxococcus xanthus*, a gram-negative developmental bacterium. *FEBS Lett* **400**, 188–192.
- Urabe, H. & Ogawara, H. (1995). Cloning, sequencing and expression of serine/threonine kinase-encoding genes from *Streptomyces coelicolor* A3(2). *Gene* **153**, 99–104.
- Zhang, C.-C. (1996). Bacterial signalling eukaryotic-type protein kinases. *Mol Microbiol* **20**, 9–15.
- Zhang, C. C. & Libs, L. (1998). Cloning and characterization of the *pknD* gene encoding an eukaryotic-type protein kinase in the cyanobacterium *Anabaena* sp. PCC7120. *Mol Gen Genet* **258**, 26–33.
- Zhang, W., Munoz-Dorado, J., Inouye, M. & Inouye, S. (1992). Identification of a putative eukaryotic-like protein kinase family in the developmental bacterium *Myxococcus xanthus*. *J Bacteriol* **174**, 5450–5453.
- Zhang, W., Inouye, M. & Inouye, S. (1996). Reciprocal regulation of the differentiation of *Myxococcus xanthus* by Pkn5 and Pkn6, eukaryotic-like Ser/Thr protein kinases. *Mol Microbiol* **20**, 435–447.

Received 17 February 1999; revised 6 July 1999; accepted 9 August 1999.