

A Deduced *Thermomonospora curvata* Protein Containing Serine/Threonine Protein Kinase and WD-Repeat Domains

LUBOMÍR JANDA, PAVEL TICHÝ, JAROSLAV SPÍŽEK, AND MIROSLAV PETŘÍČEK*

Institute of Microbiology, Academy of Sciences of the Czech Republic, Videňská 1083, 142 20 Prague 4, Czech Republic

Received 11 September 1995/Accepted 19 December 1995

The gene *pkwA* coding for a typical WD-repeat protein was found in the chromosome of the bacterium *Thermomonospora curvata* CCM 3352. Until now WD-repeat proteins were thought to be confined to eukaryotes.

Neer et al. (6) have recently reviewed and defined the family of WD-repeat proteins, so called because they carry four to eight repeats of a conserved motif which usually ends in the amino acids tryptophan and aspartate (W and D in the single-letter code). The members of the family (insofar as their functions are known) are regulatory proteins, and WD repeats are assumed to function in protein-protein interactions. Until now these proteins were thought to be confined to eukaryotes.

During work on the isolation and cloning of genes coding for the components of the maltose regulon in *Thermomonospora curvata* CCM 3352 with the help of degenerate oligonucleotide probes (4a), we isolated a DNA fragment carrying a part of a gene that was apparently not a member of this regulon. However, its deduced amino acid sequence displayed interesting features, predicting a role for this gene product in global regulatory mechanisms of this organism. We have sequenced the whole *pkwA* gene (*pkwA* stands for protein kinase with WD-repeat domain) (Fig. 1) from this facultatively thermophilic actinomycete. The gene was cloned on a 3.2-kb *SalI-BamHI* fragment of chromosomal DNA and is predicted to code for a 742-amino-acid polypeptide. The C terminus of the putative protein has a typical WD structure (Fig. 2): the number of repeats is seven (the typical number of repeats is between four and eight); the spacing between repeats is 11 amino acids (the typical number of amino acids is between 6 and 94); the consensus N-terminal amino acids (GH), C-terminal amino acids (WD), and length of the constant-length core of 26 to 27 amino acids (the typical number is between 23 and 41) are all typical features; and the amino acid composition is also in agreement with the consensus sequence published by Neer et al. (6). Thus, there seems to be no doubt that *pkwA* encodes a typical WD-repeat protein.

An inspection of the complete amino acid sequence of PkwA revealed the presence of three distinct protein regions. The N-terminal sequence is highly similar to protein kinases of a serine/threonine type and encompasses all 11 domains conserved among this class of enzymes (3). The central spacer region of PkwA is formed by a stretch rich in proline and glutamic acid. In this protein section confined by Glu-301 and Pro-394 there are 46 residues of either Pro or Glu. The C-terminal portion of PkwA is composed of seven recurring WD repeats.

The amino acid sequence of the N-terminal region of the PkwA protein (267 amino acids) shows the highest similarity

with the N-terminal portion of the protein kinase AfsK sequence (268 amino acids) from *Streptomyces coelicolor* A3(2) (5). These regions, which contain all 11 protein kinase conserved domains, are identical in 50% of residues. Both sequences are also similar with respect to their organization into three distinct domains in the protein molecule. AfsK also possesses a spacer domain of a length similar to that of PkwA that contains a number of Pro residues which are accompanied by Ala. The C-terminal portion of AfsK, whose function has not yet been assessed, does not contain any WD motifs; however, we found that a clearly observable recurring stretch of amino acids containing the consensus sequence (L/V)X(A/T)L(D/E/K)(A/V)XXGXX(R/K)W occurs here five times and that the less conserved motif with L/W at the fourth position even occurs seven times.

Thermomonospora and *Streptomyces* are related genera (2, 7, 8) characterized by complex morphologies resembling those of filamentous fungi and by their ability to produce a wide variety of secondary metabolites (1). AfsK protein kinase is a part of a control cascade involved in the control of secondary metabolite production. It is likely that PkwA also plays a regulatory role during the organism's complex growth cycle and in secondary metabolite production.

Two of three domains (kinase and WD-repeat domains) linked in the PkwA protein also resemble the complex of two eukaryotic proteins (protein kinase [PKC] and RACK1) (9). In this complex, the activated protein kinase C is bound to the receptor protein RACK1, composed almost entirely of seven WD repeats. While bound to RACKs, the activated PKCs are translocated from cytosolic into particulate (membrane) fractions. Therefore, the function of the putative PkwA protein could be to interact with a subset of proteins and mediate signal relay by phosphorylation.

The data presented here should be supported by results from the isolation and characterization of the PkwA protein; however, they strongly indicate that WD proteins are also present in prokaryotes. So far only a single example of a protein with a WD-repeat domain and a catalytic domain (protein kinase) in one polypeptide has been described (the Vps15p protein found in *Saccharomyces cerevisiae* [4]). The resemblance of the domain organization of Pkw1 protein with that of AfsK from *Streptomyces coelicolor* A3(2) indicates that these proteins could represent a novel family of bacterial regulators similar to those found so far only in eukaryotes.

While this paper was being reviewed, another example of a prokaryotic WD-repeat protein gene from *Synechocystis* spp. was released by GenBank under accession number D63999 as the 44th CDS in this entry.

* Corresponding author. Phone: 42 2 475 2426. Fax: 42 2 471 5743. Electronic mail address: petras@biomed.cas.cz.

Sal I
 1 GTCGACGACACCAGCACCCCGTGCAGCCATCCTCGCCGACCAAGGCACCCACCCG 60
 V D D T S T P V R A I L A A G W L A A L S R P
 61 CACCTGGAACCTCCTGACAACCTGCACGCCCGGATGGCTCGCCGACTCAGCCGCCCC 120
 H L E L L D N L H A A G W L A A L S R P
 121 GTGCCCGGACGCGACTCGCCGACGCCGACCAAGAAGCCGCATGCTTCTCATCCGCGCC 180
 V P G S A L A D A D Q E A A M L L I R A
 181 CAGGCCACCGGCGCCCGCTCCTGCCCCGAAACCCAGCCGACCCCTCGAAACCCCTCCTC 240
 H A T G R P L L P E T O P D P L E T L L
 241 GAAACCGACGAAGCCCGCGCCGACGCCGACCCCTCCTCGGCCCGCTCGCCGGAACGAC 300
 E T D E A R A A A R T L L G P L A G N D
 301 AGCACCTCCCGACCTCGCCGACCCCTCGCCGCTGGCTCGCCACACGCGCAGCTGG 360
 S T S R T L R T T L R A W L A H H G S W
 361 GACCGGACCGCCGCCGACCTCGCCGCCACCGCAACAGCGTCCGCTACCGCATCAGCCGC 420
 D R T A A D L G A H R N S V R Y R I S R
 421 ATCGAACCGACCTCGGCATCGACCTGTCGCCACCGGAACAGCGAATGCGCATGTGGTT 480
 I E R D L G I D L S D P E Q R M R M W F
 481 GCGCTGTCCGATTCGCCGACACCCACCCACCCGACACAAAGGACATACGCA 540
 A L S R F P D T P T Q R D I S R
 541 TGAGGTGAAGAAATTCACCGCATCGTGGGAAGAAAGTTCTTGACATAATGCTGTGT 600
 +
 601 GGGCAGCATCGTGGCGGTACGCGGGTACGGCGAGGAGGAATGATAGAGCGTTGCA 660
 1 AGGAGGGAATGATAGAGCGTTGCA 660
Pvu I RBS *fm I E P L Q 6*
 661 GCCCGGCGACCCCGGGCGGATCGGTCCTACCGGCTCGTACGCCGACTAGGCGCGGAGG 720
 7 P G D P G R I G P Y R L V S R L G A G G 26
 721 AATGGGCGAGTATTCCTCGCCCGTCCCTGGAGGGGACCGGTAGTGGTCAAAGTCAT 780
 27 M G Q V F L A R S P G G R P V V V K V I 46
 781 CCTGCCCGAATACGCCAACGACGACGAATACCGCATTCGTTTCGCGGTGAAGTCGAAGC 840
 47 L P E Y A N D D E Y R I R F A R E V E A 66
 841 TGCCCGCGCGTCCGCGGCTTCCACACCGCCAAAGTCATCGACGCCGACCCACCGCCGA 900
 67 A R R V G G F H T A Q V I D A D P T A D 86
 901 CCCGCGTGGATGGCCACCGCTACATCCCGGCCCTCCTCGCGAAAGCGTCAACCGA 960
 7 P P W M A T A Y I P G P S L R K A V T E 106
 961 ACGGCGCCCTCTACGGAACAACTCGCCGACCCCTCGTCCGCGACTCGTCAAGGCCCT 1020
 107 R G P L Y G N N L R T L A A G L V E G L 126
 1021 CGCGCCATCCACGCTCGCGGCTCGTCCACCGGACTTCAAACCCAGCAACATCGTCT 1080
 127 A A I H A C G L V H R D F K P S N I V L 146
 1081 CGCGCCGACGGCCCCCGTGTATCGACTTCGGCGTCCGCCGCCCTCGACAGCAGCGT 1140
 147 A A D G P R V I D F G V A R P L D S S V 166
 1141 CATGACCCAAAGCGCGCCGCTATCGGACCCCTCGCTACATGTCCCGAACAACCGA 1200
 167 M T Q S G A V I G T L A Y M S P E Q T D 186
 1201 CGGCAGCAAGTCGGGCCCGCCAGCGACGTGTTCTCCCTCGGCACCGTGCCTCCGTCG 1260
 187 G S Q V G P A S D V F S L G T V L A F A 206
 1261 CGCCACCGGACGTTCCCGTTCATGGCCGACTCCATCGCGAAATCATCGCGCATCTC 1320
 207 A T G R S P F M A D S I G E I I A R I S 226
 1321 CGGACCGCTCCCGAATCCCGAATCCCGGACGACCTCGGGGAACCTCGTACGCATG 1380
 227 G P P P E L P E L P D D L R E L V Y A C 246
 1381 CTGGGAGCAACCTGACTTGGGCCACCCAGCGCCGAACCTCCTCGCCAGCTCAGCAC 1440
 247 W E Q N P D L R P T T A E L L A Q L S T 266
 1441 CGACCACCCGGCGACGACTGGCCCCCGCCCACTTGAGCGACCTGATCGGCTCAATGCT 1500
 267 D H T G D D W P P P H L S D L I G S M L 286
 1501 CCCACTCGGGGCAACCACTCCCGCAACCCGTCGCTCGCATCGAACGCCACCCCGCTC 1560
 287 P L G A T T S P N P S L A I E P P P S 306
 1561 CCACGGCCCCCGCGCGCTCGCAACCGTGGCCGACCCCGGGGACGCGCCGACGAACC 1620
 307 H G P P R P S E P L P D P G D D A D E P 326
 1621 CTCTCGGGAAACCAAGTCGGACCCCTCCGGAACCGGAGCCGCCGGAACGGAAGAAA 1680
 327 S A E K P S R T L P E P E P P E L E E K 346
 1681 ACCCATCCAGGTCATCCACGAAACCGGACCCCTGCACCCACCCCGCCCGCCCGCGGA 1740
 347 P I Q V I H E P E R P A P T P P R P R E 366
 1741 ACCGCCAGAGGGCGATAAAACCGAAAAATCCCGCCCGGAGCGCCCAACCCCGCTG 1800
 367 P A R G A I K P K N P R P A A P Q P P W 386
 1801 GTCCCCGCCCGGGTCCAGCCACCCCGTGGAAACAACATCATCAGAAACCCGCTGGC 1860
 387 S P P R V Q P P R W K Q L I T K K P V A 406
 1861 CGGCATCTCACCAGTCCGACGCGCCGCGCTCGTGTCTCTCCTCGTGTGGCAGTG 1920
 407 G I L T A V A T A G L V V S F L V W Q W 426
 1921 GACACTCCCGAGACCCCGTCCGCCCGGACGACGACCCCGCCCTCAGAGTCCGCTG 1980
 427 T L P E T P L R P D S S T A P S E S A D 446
 1981 CCGCAGCAACTCAACGAGCTCGCATCTCACCACCGACCGGGAAGCGGTTGCCGTCG 2040
 447 P H E L N E P R I L T T D R E A V A V A 466
 2041 ATTCAGCCCGGCGGATCCCTCCTCGCCGGCGGACGCGGACAACTCATCCACGTTG 2100
 467 F S P G G S L L A G G S G D K L I H V W 486
 2101 GGACGTGGCAAGCGGCGAACCTCCACCCCTGGAAGGCCACCCGACTGGTGGTGC 2160
 487 D V A S G D E L H T L E G H T D W V R A 506
 2161 CGTCGCGTTCAGCCCGACGGCGCCCTCCTCGCCAGCGGACGCGGACGCCACCGTGC 2220
 507 V A F S P D G A L L A S G S D A T V R 526
 2221 GCTGTGGGAGTGGCCGCCGAGAAGAAGCGGCTTTTTCGAAGGACACCCACTACGT 2280
 527 L W D V A A A E E R A V F E G H T H Y V 546
 2281 GCTGGACATCGGTCACCCCGACGGCTCGATGGTGGCCAGCGGACGAGAGCGGCAC 2340
 547 L D I A F S P D G S M V A S G S R D G T 566
 2341 CGCCCGGCTGGAACGTCGCCACCGGGACAGAACCAGCCCTCAAGGGACACCCGA 2400
 567 A R L W N V A T G T E H A V L K G H T D 586
 2401 CTACGTCACGCTGTCGCGTTCAGCCCGACGGCTCGATGGTGGCCAGCGGACGAGAGA 2460
 587 Y V Y A V A F S P D G S M V A S G S R D 606
 2461 CGCCACCATCCGCTGTGGGATGTCGCCACCGGAAAGAAGCGGACGCTGCAAGCACC 2520
 607 G T I R L W D V A T G K E R D V L Q A P 626
 2521 GCGCGAATGTCGCTCCTCGCGTTCAGCCCGGACGGGACGATGCTCGTCCACGGCAG 2580
 627 A E N V V S L A F S P D G S M L V H G S 646
 2581 CGACAGCACCCTGTCGGGACGTCAGCGGCGGCAAGCACTGCACACCTTCAAGG 2640
 647 D R L W H L W D V A S G E A L H T F E G 666
 2641 CCACACCGACTGGTGGTGTGTCGCGTTCAGCCCGGACGGCGCCCTCCTCGCCAGCG 2700
 667 H T D W V R A V A F S P D G A L L A S G 686
 2701 CAGCGACGACCGACCATCCGCTGTGGGATGTCGCCGCCAGGAAGAACAACAACCT 2760
 687 S D D R T I R L W D V A A Q E E H T T L 706
 2761 GGAAGGGACACCCAGCCGCTCCACTCCGTAGCTTCCACCCGGAAGGACCAACGCTAGC 2820
 707 E G H T E P V H S V A F H P E G T T L A 726
 2821 CAGCGCAAGCGAAGCGGACTATCCGTATCTGGCCATCGCCACGGAATAACCCCGCG 2880
 727 S A S E D G T I R I W P I A T E * 742
 2881 CTGCTCCCGGACGCTCGTCCGCGGCTTTCGTCGCCCGCCCTCGGAACG 2940
 2941 GCAGGGAACACTCTGGTGCCTGAGTCTCGACAGCCCTCTCTACTGAGGAGAGCACA 3000
 3001 CCTCAAGACAGTTCAGTTCGGTGTGTCTGTCCCGTCAAGCAGCGGTCACCCACGAG 3060
 3061 AGAGAGTCAGCGCGTCCCGGCGCCCTCCTCCGTAGCGACCTCAGGGTACATATCCG 3120
 3121 GCGGCTCCACTGGACCGGAACCTTCCCGAGCGGATCTCAGCCACCGCATGGCG 3180
 3181 ATGCTCCAGGATGTGCAAAACCTTCGGATCC BamH I 3213

FIG. 1. Nucleotide and amino acid sequences of the *T. curvata* *pkwA* gene. The deduced amino acids are aligned below the second nucleotide of each codon of the nucleotide sequence. A putative ribosomal binding site (RBS) is underlined. Restriction endonuclease sites are indicated.

	I. domain	II. domain		
1	MIEPLQPGDPRIGTPYRLVSR	L GAGG MGVFLARS	PGGRP VVKVILPEYANDDEYR	IRF 60
	III. domain	IV. domain	V. domain	
61	AREVEAARRVGGFHTAQ	VIDADPTADP	PMMATAYIPGSLRKA	VTGRLYGNLRLTAA 120
	VI. domain	VII. domain	VIII. domain	
121	GLVEGLAAITHACGLVHRDFKPS	MIVLAADGPRVIDFGVAR	PLDSSVMTOSGAVIGTLAYM	180
	IX. domain	X. domain		
181	SPEQTDGSGVGPASOVFSLG	TVLAFATGRSPFMAD	SIGETIARISGPPPELPEL	PDDL R 240
	XI. domain			
241	ELWACWEQNPDRLRPTTAEL	LQALST		266
267	DHTGDDWPPPHLSDLIGSML	PLGATTSNPNSLATEPP	PPPS	306
307	HGPPRPSEPLPDGDDA	DEPSAEKPSRTLPEPEP	ELEEK	346
347	PIQVIHEPERPAPT	PPRPREPARGAIKPKN	PRPAAPPPW	386
387	SPPRVQPRNKQLITK	KPVAGILTA VATAGLV	VSFLVQW	426
427	TLPETPLRPSDSTAP	SESADPHLNEPR		464
455	IL TT DRE AVAV A F S	PGG S LLA GGS G D	KL I H V W D	VASGD E L H T 496
497	LE GH TDW VRAV A F S	PDG A LLA SGS D D	ATVR L W D	VAAAEERAV 538
539	FE GH THY VLDI A F S	PDG S MVA SGS R D	GTAR L WN	VATGTEHAV 580
581	LK GH TDY VYAV A F S	PDG S MVA SGS R D	GTIR L WD	VATGKERDV 621
622	LQ AP AEN VVSL A F S	PDG S MLV HGS D	STVH L WD	VASGEALHT 663
664	FE GH TDW VRAV A F S	PDG A LLA SGS D D	RTIR L WD	VAAQEEHTT 705
706	LE GH TEP VHSV A F H	PEG T TLA SAS E D	GTIR I WP	IATE 742

KINASE DOMAIN
SPACER DOMAIN
WD-REPEAT DOMAIN



FIG. 2. Map of the PkwA protein from *T. curvata*. Residues of the N-terminal part conserved among protein kinases of the serine/threonine type are shown in italic type; the typical 11 domains are indicated. The alignment of the seven repeats within the C-terminal portion of the polypeptide, designated the WD-40 repeat domain, was performed according to the suggestions of Neer et al. based on the regular expression of the gene (6). The plot at the bottom showing the units of organization of *pkwA* (black areas, no deviations, or "misses," from the consensus sequence; stippled area, one miss; hatched area, two misses; white area, three misses) is also based on Neer et al. (6).

Nucleotide sequence accession number. The nucleotide sequence of the *pkwA* gene is recorded in GenBank under accession number U23820.

We thank J. Cullum for helpful comments during preparation of the manuscript.

This work was supported by the Grant Agency of the Czech Republic (grant 301/93/0600 to J.S.) and the Grant Agency of the Academy of Sciences of the Czech Republic (grant 52045 to M.P.).

REFERENCES

- Chater, K. F. 1993. Differentiation in *Streptomyces*. *Annu. Rev. Microbiol.* **47**:685-713.
- Goodfellow, M., and T. Gross. 1984. Classification, p. 7-144. *In* M. Goodfellow, M. Mondarski, and S. T. Williams (ed.), *The biology of Actinomycetes*. Academic Press, London.
- Hanks, S. K., M. Quinn, and T. Hunter. 1988. The kinase family conserved features and deduced phylogeny. *Science* **241**:42-52.
- Herman, P. K., J. H. Stack, J. A. DeModena, and S. D. Emr. 1991. A novel protein kinase homolog essential for protein sorting to the yeast lysosome-like vacuole. *Cell* **64**:425-437.
- Janda, L., P. Tichý, J. Spížek, and M. Petříček. Unpublished data.
- Matsumoto, A., S.-K. Hong, H. Ishizuka, and S. Horinouchi. 1994. Phosphorylation of the AfsR protein involved in secondary metabolism in *Streptomyces* species by a eukaryotic-type protein kinase. *Gene* **140**:47-56.
- Neer, E. J., C. J. Schmidt, R. Nambudripad, and T. F. Smith. 1994. The ancient regulatory protein family of WD-repeat protein. *Nature (London)* **371**:297-300.
- Petříček, M., K. Stajner, and P. Tichý. 1989. Cloning of a thermostable α -amylase gene from *Thermomonospora curvata* and its expression in *Streptomyces lividans*. *J. Gen. Microbiol.* **135**:3303-3309.
- Petříček, M., P. Tichý, and M. Kuncová. 1992. Characterization of α -amylase gene from *Thermomonospora curvata*. *Gene* **112**:77-83.
- Ron, D., C.-H. Chen, J. Caldwell, L. Jamieson, E. Orr, and D. Mochely-Rosen. 1994. Cloning of an intracellular receptor for protein kinase C: a homolog of the β -subunit of G proteins. *Proc. Natl. Acad. Sci. USA* **91**:839-843.