ELSEVIER

FEBS Letters 337 (1994) 289–292

FEBS 13546

FEBS LETTERS

The novel hexapeptide motif found in the acyltransferases LpxA and LpxD of lipid A biosynthesis is conserved in various bacteria

Riitta Vuorio[a,*], Taina Härkönen[a], Martti Tolvanen[b], Martti Vaara[a]

[a]*Department of Bacteriology and Immunology, and* [b]*Department of Biochemistry, P.O. Box 21 (Haartmaninkatu 3), SF-00014 University of Helsinki, Helsinki, Finland*

Received 1 December 1993

Abstract

Two bacterial acyltransferases (LpxA of *Escherichia coli*, LpxD of *E. coli* and *Salmonella typhimurium*) have previously been shown to consist of a very unusual tandem-repeat structure with tens of repeating hexapeptides (24 hexapeptides in LpxA, 26 in LpxD). By sequencing LpxD of *Yersinia enterocolitica* (a distant relative of *E. coli* and *S. typhimurium* within Enterobacteriaceae) as well as LpxA of *S. typhimurium* and *Y. enterocolitica*, and by analyzing the existing data on these enzymes of *Ricketsia rickettsii*, it was now shown that the hexapeptide repeat pattern is a very conservative property of these enzymes. Even though the overall homology (allowing equivalent amino acids) between the four proteins was only 59% in LpxA and 58% in LpxD, the homology in the first residue of each hexapeptide was 87% in LpxA and 100% in LpxD. Secondary structure prediction by PredictProtein server suggested a very strong beta strand dominance in all the hexad regions. Accordingly, LpxA and LpxD of various bacterial origins can now be regarded as structurally very unusual enzymes, largely consisting of hexad repeats.

Key words: LpxA; LpxD; Acyltransferase; Hexapeptide repeat theme; Proteobacteria

1. Introduction

We have previously found a novel hexapeptide repeat theme in eight proteins [1,2]. These include LpxA [3], LacA [4], CysE [5] and DapD [6] of *Escherichia coli*, LpxD of *E. coli* [7] and *Salmonella typhimurium* (gene formerly named as *firA* or *ssc*) [8], NodL from *Rhizobium leguminosarum* [9], Yglm of *E. coli* [10] and its analogue Tms of *Bacillus subtilis* [11], as well as Yerm of *B. sphaericus* [12]. Of these proteins, Yglm/Tms and Yerm are hypothetical and the others are enzymes. LacA functions as a thiogalactoside acetyltransferase, CysE as a serine acetyltransferase and the nodulation protein, NodL, is probably also an acetyltransferase. DapD is a succinyl-amino-pimelate aminotransferase. The two remaining enzymes, LpxA and LpxD, have more of their structures made with this periodically repeating hexapeptide motif than do the other proteins mentioned above. The repeats cover 55% of the total lenght of LpxA and 44% of that of LpxD [2]. LpxA and LpxD are essential, remarkably homologous [1,3,8,13] and the self-comparison shows that they both contain regions homologous with other parts of the same protein. Mutations in either of these enzymes cause the bacterial cell to become thermosensitive and to quickly die after a raise of temperature and also sensitize the bacteria to hydrophobic antibiotics [13–16]. The functions of LpxA and LpxD are related as they are both involved in the synthesis of lipopolysaccharide (= LPS). LpxA is the first (UDP-*N*-acetylglucosamine acyltransferase) [3] and LpxD the second acyltransferase (UDP-3-*O*-[3-hydroxymyristoyl] glucosamine *N*-acyltransferase) [17,18] in the biosynthesis of lipid A, the hydrophobic part anchoring LPS to the outer leaflet of the outer membrane. The codon usage for *E. coli lpxA* and *lpxD* is typical for a lowly expressed gene.

Because about half of the LpxA and LpxD proteins are made of hexad repeats, we were wondering whether this motif is essential for the function of these proteins. If it is, the theme is expected to be conserved in various bacteria. To test this, we sequenced *lpxA* and *lpxD* genes from *Yersinia enterocolitica* and *lpxA* from *S. typhimurium* and compared their deduced amino acid sequences with those already known (i.e. *lpxA* and *lpxD* of *E. coli* (=*firA*) and *lpxD* (=*ssc*) of *S. typhimurium*). We also modified the original consensus sequence which we have described before [1,2], to find the minimal backbone covering this theme, and tested whether it could be able to find similar hexapeptide repeats in additional bacterial proteins.

2. Materials and methods

2.1. Bacterial strains, plasmids and phage M13

The bacterial strains were rough derivatives of *S. typhimurium* LT2 (SH5014) [8] and *Y. enterocolitica* (EH902) [19]. *E. coli* strains JM105

*Corresponding author. Fax: (358) (0) 434 6382.

SSDI 0014-5793(93)E1463-V

[20] and TG1 [21] were used as cloning hosts and pUC19 was used as the plasmid vector. *lpxD* of *Y. enterocolitica* was sequenced from pUCHS115 ([19] and from its derivative pUCHS114, which carries a smaller genomic insert (the *Pst*I–*Bam*HI fragment of pUCHS115) in pUC18. Besides plasmid sequencing, also M13 mp8 and mp9 clones [22] were used in sequencing *lpxA* of *S. typhimurium*. Bacteria were grown under aeration at 30 or 37°C in LB broth with tetracycline (12.5 μg/ml) or ampicillin (100 μg/ml) when necessary.

2.2 General DNA techniques, cloning of lpxA

Isolation of chromosomal DNA, preparation of hybridization probes, their radiolabeling, Southern blotting, cloning, sequencing and other DNA manipulations were carried out as described in [19]. After nick translation, free label was separated from the labeling mixture with Nuctrap push columns (Stratagene, La Jolla, CA). To clone *lpxA* of *S. typhimurium* and *Y. enterocolitica*, probes were prepared from pUCHS16 [8] and pUCHS115, respectively. The labeled probes detected a 5.0 kb *Pst*I-fragment from the chromosome of SH5014 and a 2.5 kb *Pst*I–*Bgl*II-fragment from the chromosome of EH902. These fragments were ligated to pUC19 and the resultant plasmids were named pRV40 and pTH2, respectively.

2 3 Computer analyses

The conserved motif $[(I,V,L)GXXXX]_3(I,V,L)$ was used to screen all bacterial amino acid sequences in SwissProt (release 25.0, April 1993) with the Findpatterns program allowing two mismatches. The multiple alignments were made with the PileUp program using default parameters (both programs provided by the Genetics Computer Group, Program Manual for the GCG Package, version 7, 1991). The equivalent amino acids in these alignments were: A,G,S,T,P; E,D,N,Q; H,K,R; I,L,M,V; F,Y,W; C. The ProSite [23] database (ver. 10.1) was searched for other consensus patterns of the proteins containing this hexapeptide repeat theme. The absence of three-dimensional structures for any of these proteins was confirmed in the Brookhaven Protein Data Bank (PDB, release of Jan 1993). Secondary structure predictions were performed with the Heidelberg neural network prediction server [24] (data and software as on Oct 7, 1993). The PredictProtein server was given three different sequence alignments: LpxA and LpxD sequences with four aligned sequences in each group, and all eight sequences together.

2 4. Nucleotide sequence accession numbers

The EMBL accession numbers are: Z25462 for the *lpxA* gene from *S typhimurium* SH5014 and Z25463 for the sequence containing *lpxD* and *lpxA* from *Y. enterocolitica* EH902.

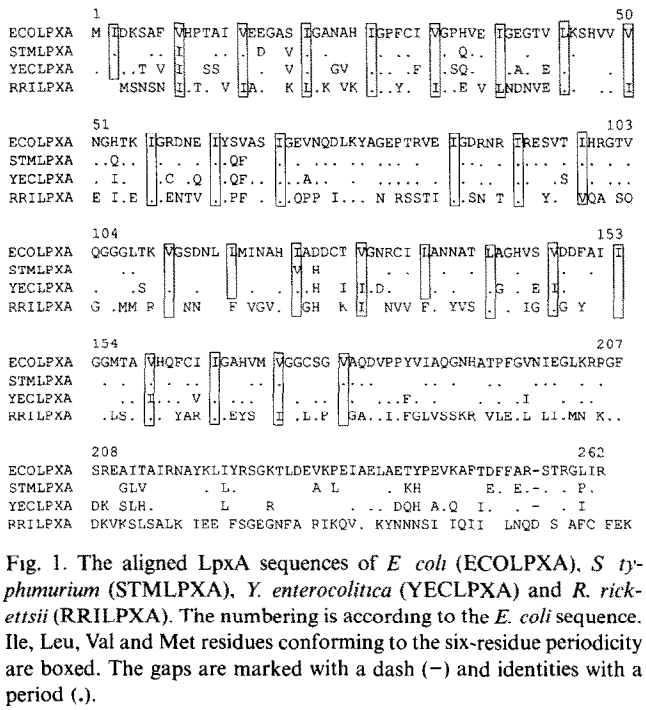

Fig. 1. The aligned LpxA sequences of *E coli* (ECOLPXA), *S typhimurium* (STMLPXA), *Y. enterocolitica* (YECLPXA) and *R. rickettsii* (RRILPXA). The numbering is according to the *E. coli* sequence. Ile, Leu, Val and Met residues conforming to the six-residue periodicity are boxed. The gaps are marked with a dash (−) and identities with a period (.).

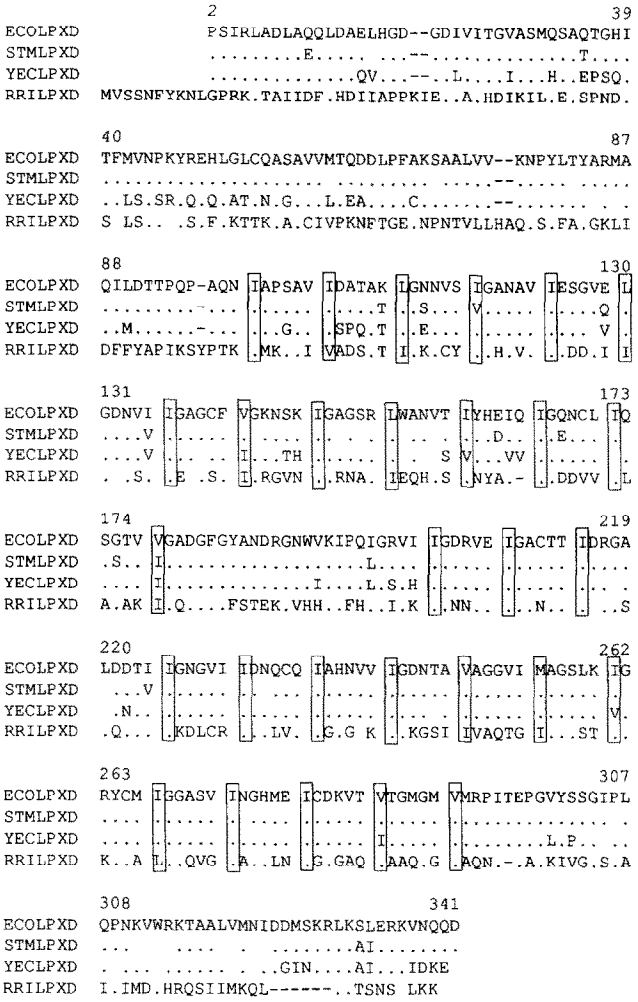

Fig. 2. The aligned LpxD sequences of *E coli* (ECOLPXD), *S. typhimurium* (STMLPXD), *Y. enterocolitica* (YECLPXD) and *R. rickettsii* (RRILPXD). The six-residue periodicity and gaps are marked as in Fig. 1. The mature *E coli* protein does not contain the initiation methionine [7]; this residue is omitted from the other enterobacterial sequences, too.

3. Results and discussion

We have been impressed by our recent finding that the acyltransferases LpxA and LpxD of *E. coli* and LpxD of *S. typhimurium* are, to a remarkable extent, built from repeating hexapeptide blocks. To see how universal a property this is for enterobacterial LpxA and LpxD proteins, we determined the sequence of LpxA of *S. typhimurium* and those of LpxA and LpxD of *Y. enterocolitica*, a phylogenetically distant relative of *Escherichiae* and *Salmonellae*. Also the sequences of *lpxA* and *lpxD* of *Ricketsia rickettsii* have been determined recently [25], which enhances the available information. These four species all belong to the eubacterial group proteobacteria (rickettsiae to the α subdivision, enteric bacteria to the γ subdivision) [26].

Each of the LpxA (Fig. 1) and LpxD (Fig. 2) proteins

now known display a hexapeptide repeat pattern. Each hexad starts with Ile, Leu, Val (or their equivalent [27] Met), and, as also noted previously [2], the second position is often occupied by Gly, the fourth by Gly, Asn or Asp and the fifth by Val or Ala. There are 24 hexads in LpxA and 26 hexads in LpxD of the three enterobacterial species; the corresponding numbers for *R. rickettsii* are 19 and 25, respectively. The identity between the enterobacterial LpxA proteins is 80% and between all four LpxA proteins, 35%. The corresponding similarities (allowing equivalent amino acids) are 89% and 60%, respectively. The identity percentages between the LpxD proteins are 82% (enteric bacteria) or 30% (all four bacteria); the corresponding similarities are 94% or 60%, respectively.

The degree of conservatism within different regions of LpxA and LpxD is shown in Table 1. Regarding LpxA, it is obvious that the only notably conserved amino acids in the four bacteria are placed as the first and second residue of each hexad. As for LpxD, the preserved location is the first residue of the hexads (there is also a short, 10 residues long, highly conserved region from position 215 to 224 between two hexapeptide regions). The other parts of LpxA or LpxD do not show any preferred conservation.

Accordingly, it is now clear that the hexapeptide repeat pattern is very characteristic for the LpxA and LpxD proteins of all the studied bacteria. Even though the overall sequences of these proteins from different bacteria do not appear to be unexceptionally conserved, the hexad arrangement and the first residue (in LpxA, two first residues) of each hexad are extremely conserved. These residues can be regarded as backbone residues.

We then simplified our original consensus sequence [1] to make it cover the hexad repeat motif as a universal theme. The final motif is $[(I,V,L)GXXXX]_5(I,V,L)$. By this sequence, using the Findpatterns program and allowing two mismatches, we found most of the proteins which we have previously reported to contain a homologous theme. Only Yglm and DapD were left outside. The search found all the members of the suggested acetyltransferase family [23], i.e. CysE (from *E. coli* and *S. typhimurium*), LacA, NifP (a putative serine acetyltransferase from *Azotobacter chroococcum*) [28], and NodL (from *Rh. leguminosarum* and *Rh. meliloti*) [29]. The search also detected the chloramphenicol acetyltransferases (CATs) of the *E. coli* transposon Tn2424 [30] and *Agrobacterium tumefaciens* [31]. These two CATs have no similarity with other chloramphenicol acetyltransferases. Finally a β-glucosidase from *B. polymyxa* ([32] gene name: *bglA*) as well as a 24.3 kDa hypothetical protein from enterobacterial antibiotic resistance plasmid R1 [33] was found to contain this motif. The β-glucosidase active site and the hydrolase consensus pattern are situated outside the hexad theme. To summarize, the hexapeptide motif can now be stated to be characteristic of certain families of acyl- and acetyltransferases, as well as of some other proteins.

Periodicity repeat themes are very rare in enzymes (for known examples, see [2]). It is plausible to expect that the hexad regions of LpxA and LpxD form a structural (non-catalytic) part of the protein. We employed PredictProtein program to predict the secondary struc-

Table 1
Conservatism in different regions (hexapeptide vs. non-hexapeptide) of the known LpxA and LpxD proteins, shown as the missmatch percentage in the aligned amino acid sequences

				Missmatch %[b] in	
		LpxA[a]	LpxD[a]	LpxA	LpxD
Entire protein				41	42
Non-hexapeptide regions					
Region	A	(0– 1)	(2– 99)	100	47
	B	(68– 85)	(178–202)	44	36
	C	(98–110)	(215–224)	46	0
	D	(177–262)	(291–341)	52	55
Total	(A–D)	118 residues	184 residues	51	45
Hexapeptides (abcdef)					
Residues in position	a			13	0
	b			17	38
	c			38	35
	d			54	58
	e			33	46
	f			46	58
Total	(a–f)	26 × 6 = 144	26 × 6 = 156	33	39

[a] The amino acids in the aligned LpxA and LpxD (see Figs. 1 and 2) are numbered according to the *E. coli* sequences. The locations of non-hexapeptide regions (residue positions) are shown in parentheses; the rest of the proteins is built from repeated hexads. The non-aligned regions in the amino (LpxD) and carboxy (LpxA) termini of *R. rickettsii* sequences, as compared to the *E. coli* proteins, are omitted.

[b] Indicates the percentage of non-conservative amino acid substitutions from the total number of amino acid residues. Conservative substitutions are defined in Section 2.

ture of these proteins and found a clearcut predominance of beta structure in the hexapeptide region. In case of LpxD the PredictProtein server judged this region to consist solely of beta strands and of loops connecting them. The prediction for LpxA shows four short, less certainly predicted alpha-helical segments in addition to the dominating beta structure, whereas the combined prediction of the eight sequences is all-beta in this region. In the C-terminal region both LpxA and LpxD show alpha helices predicted with a high reliability. In the N-terminal 110 amino acids of LpxD there are two strongly predicted alpha-helices as well.

Interestingly, all the known defective alleles of *lpxD* (i.e. *ssc*-1, *omsA*, *firA200*, *firA201*) have point mutations in the hexad region or (in *ssc*-1) in the residue immediately flanking the hexad region [8,13,14]. Unfortunately, there is no crystallographic data on the structure of any of the proteins with the hexapeptide repeat theme. It would now be very interesting to determine the three-dimensional structures of LpxA and LpxD, including the mutant proteins of the latter.

Acknowledgements: We thank B. Kuusela for excellent technical assistance. This study was supported by Grant 1011749 from the Academy of Finland and by the Sigrid Juselius Foundation.

References

[1] Vuorio, R., Hirvas, L. and Vaara, M. (1991) FEBS Lett. 292, 90–94.
[2] Vaara, M. (1992) FEMS Microbiol. Lett. 97, 249–254.
[3] Coleman, J. and Raetz, C.R.H. (1988) J. Bacteriol. 170, 1268–1274.
[4] Fowler, A.V., Hediger, M.A., Musso, R.E. and Zabin, I. (1985) Biochimie 67, 101–108.
[5] Tei, H., Murata, K. and Kimura, A. (1990) Biochem. Biophys. Res. Commun. 167, 948–955.
[6] Richaud, C., Richaud, F., Martin, C., Haziza, C. and Patte, J.-C. (1984) J. Biol. Chem. 259, 14824–14828.
[7] Dicker, I.B. and Seetharam, S. (1991) J. Bacteriol. 173, 334–344.
[8] Hirvas, L., Koski, P. and Vaara, M. (1991) EMBO J. 10, 1017–1023.
[9] Hayo, C.J., Cremers, C., Spaink, H.P., Wijfjes, A.H.M., Pees, E., Wijffelman, C.A., Okker, R.J.H. and Lugtenberg, B.J.J. (1989) Plant Mol. Biol. 13, 163–174.
[10] Walker, J.E., Gay, N.J., Saraste, M. and Eberle, A.N (1984) Biochem. J. 224, 799–815.
[11] Nilsson, D., Hove-Jensen, B. and Arnvig, K. (1989) Mol. Gen. Genet. 218, 565–571.
[12] Monod, M., Mohan, S. and Dubnau, D. (1987) J. Bacteriol. 169, 340–350.
[13] Dicker, I.B. and Seetharam, S. (1992) Mol. Microbiol. 6, 817–823.
[14] Vuorio, R. and Vaara, M. (1992) J. Bacteriol. 174, 7090–7097.
[15] Galloway, S. and Raetz, C.R.H. (1990) J. Biol. Chem. 265, 6394–6402.
[16] Vuorio, R. and Vaara, M. (1992) Antimicrob. Agents Chemother. 36, 826–829.
[17] Raetz, C.R.H. (1993) J. Bacteriol. 175, 5745–5753.
[18] Kelly, T.M., Stachula, S.A., Raetz, C.R.H. and Anderson, M.S. (1993) J. Biol. Chem. 268, 19866–19874.
[19] Hirvas, L., Koski, P. and Vaara, M. (1991) J. Bacteriol. 173, 1223–1229.
[20] Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103–119.
[21] Gibson, T.J. (1984) Ph.D. thesis, Cambridge University, Cambridge, UK.
[22] Messing, J. and Vieira, J. (1982) Gene 19, 269–276.
[23] Bairoch, A. (1993) Nucleic Acids Res. 21, 3097–3103.
[24] Rost, B. and Sander, C. Nature 360, 540.
[25] Shaw, E.I. and Wood, D.O. (1993) Genbank Accession No. L22690.
[26] Woese, C.R (1987) Microbiol. Rev. 51, 221–271.
[27] Higgins, D.G. and Sharp, P.M. (1988) Gene 73, 237–244.
[28] Evans, D.J., Jones, R., Woodley, P.R., Wilborn, J.R. and Robson, R.L. (1991) J. Bacteriol. 173, 5457–5469.
[29] Nedelcho, B. and Kondorosi, A. (1992) Plant Mol. Biol. 18, 843–846.
[30] Parent, R. and Roy, P.H. (1992) J. Bacteriol. 174, 2891–2897.
[31] Tennigkeit, J. and Matzura, H. (1991) Gene 98, 113–116
[32] Gonzalez-Candelas, L., Ramon, D. and Polaina, J. (1990) Gene 95, 31–38.
[33] Bravo, A., de Torrontegui, G. and Diaz, R. (1987) Mol. Gen. Genet. 210, 101–110.