Protein Engineering, Design & Selection vol. 17 no. 10 pp. 721–729, 2004
Published online November 19, 2004 doi:10.1093/protein/gzh087

Bacterial lipid modification of proteins for novel protein engineering applications

S.Kamalakkannan[1], V.Murugan[1], M.V.Jagannadham[2], R.Nagaraj[2] and K.Sankaran[1,3]

[1]Centre for Biotechnology, Anna University, Chennai 600 025 and [2]Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad 500 007, India

[3]To whom correspondence should be addressed.
E-mail: ksankran@yahoo.com

Functioning of proteins efficiently at the solid–liquid interface is critical to not only biological but also modern man-made systems such as ELISA, liposomes and biosensors. Anchoring hydrophilic proteins poses a major challenge in this regard. Lipid modification, *N*-acyl-*S*-diacylglyceryl-Cys, providing an N-terminal hydrophobic membrane anchor is a viable solution that bacteria have successfully evolved but remains unexploited. Based on the current understanding of this ubiquitous and unique bacterial lipid modification it is possible to use *Escherichia coli*, the popular recombinant protein expression host, for converting a non-lipoprotein to a lipoprotein with a hydrophobic anchor at the N-terminal end. We report two strategies applicable to non-lipoproteins (with or without signal sequences) employing minimal sequence change. Taking periplasmic *Shigella* apyrase as an example, its signal sequence was engineered to include a lipobox, an essential determinant for lipid modification, or its mature sequence was fused to the signal sequence of abundant outer membrane lipoprotein, Lpp. Lipid modification was proved by membrane localization, electrophoretic mobility shift and mass spectrometric analysis. Substrate specificity and specific activity measurements indicated functional integrity after modification. In conclusion, a convenient protein engineering strategy for converting non-lipoprotein to lipoprotein for commercial application has been devised and tested successfully.

Keywords: apyrase/bacterial lipid modification/signal sequence

Introduction

Biological systems and biotechnological applications depend on amphipathic surface proteins that have to perform effectively and efficiently in the aqueous or solid–aqueous interface, while firmly anchored to a hydrophobic surface. For example, nutrient uptake by organisms, signal transduction and response, growth and development require many types of membrane-associated proteins. Similarly, man-made applications such as ELISA and biosensors require a variety of functional proteins immobilized to hydrophobic surfaces such as plastics. Biological systems have resorted to modifying proteins with lipids or having hydrophobic transmembrane regions to anchor such proteins to membranes. Further, attachment of proteins to membranes has also been shown to reduce diffusion in three dimensions to two dimensions, resulting in effective concentration of ~1000-fold (Stanley *et al*., 1998), an important factor in initial signal transduction events.

Chemical modification of proteins with lipid was first attempted for exploiting lipoproteins as superior antigens owing to the action of lipid as a powerful adjuvant (Cote-Sierra *et al*., 2002; Pappalardo, 2003). However, such chemical modifications suffer from lack of specific, quantitative modification (Shigematsu *et al*., 1998). Therefore, a biological alternative is an attractive proposition owing to the possibility of accurate and precise site-directed lipid modification. Among different types of biological lipid modification of proteins, the bacterial approach is the most suitable for the following reasons. In all known bacteria, more than 770 reported lipoproteins (Madan Babu and Sankaran, 2002) with a variety of functions are attached to the membrane having the same *N*-acyl-*S*-diacylglyceryl-Cys as their N-terminal amino acid (Figure 1) (Sankaran and Wu, 1994), clearly indicating that a variety of functional proteins can be lipid modified.

The covalent lipid modification of protein was first demonstrated in *Escherichia coli* murein lipoprotein (Braun and Rehn, 1969) and all three enzymes constituting the pathway have been identified and even partly characterized (Sankaran and Wu, 1994). The consensus amino acids sequence [LVI][ASTVI][GAS] C for lipid modification is present in the C-region of the signal sequence (Hayashi and Wu, 1990). Bacterial lipoproteins are synthesized as precursors containing a tripartite signal sequence. The N-region contains 5–7 residues with two positively charged Lys or Arg residues. The uncharged hydrophobic region contains 7–22 residues, with a modal value of 12 residues. These features are common even to other non-lipoproteins that cross the inner membrane. The distinguishing feature between the two is the C-region, which has a consensus sequence of [LVI][ASTVI][GAS] C, C being the lipid-modified Cys, the first amino acid in all bacterial lipoproteins (Von Heijne, 1983; Gennity *et al*., 1990; Madan Babu and Sankaran, 2002). The first enzyme of

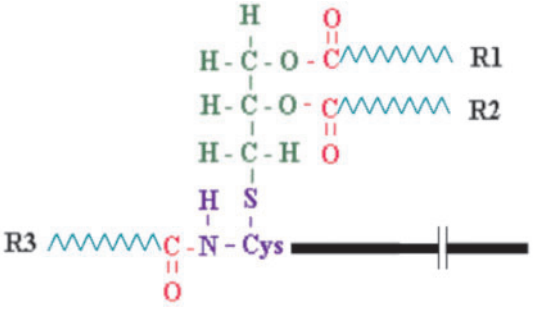

Fig. 1. Structure of *N*-acyl-*S*-diacylglycerylcysteine. The diacylglyceryl moiety is attached to the cysteine SH group by a thioether linkage and the amino group is fatty acylated. R1 and R3 are palmitoyl residues. R2 can be either a palmitoyl or an oleoyl residue. This lipid moiety anchors lipoproteins to membranes.

the biosynthetic pathway, diacylglyceryl transferase (*lgt*), recognizes the consensus sequence at the C-region.

In the absence of a complete understanding of all the determinants needed for bacterial lipid modification, fusion with prototype lipoproteins was used as a standard strategy for imparting lipoprotein character to target proteins. The *lpp*–OmpA (amino acids 46–159)–PhoA and *lpp*–OmpA (46–159)–Bla triple fusions were expressed on an *E.coli* cell surface (Georgiou *et al*., 1996; Stathopoulos *et al*., 1996) as a means to achieve surface expression of soluble proteins. *Staphylococcus aureus* protein-A, whose beta domain was fused with the signal sequence, and nine N-terminal amino acid residues of mature sequence of *lpp*, incorporated stably into proteoliposome and was shown to exhibit higher IgG binding activity than the non-lipoprotein counterpart (Shigematsu *et al*., 1998). Leishmanial antigen gp63, fused to OprI lipoprotein of *Pseudomonas aeruginosa*, was found to elicit a better immune response than native gp63 (Cote-Sierra *et al*., 2002).

In triple fusions, the adverse effect on folding of the target protein and *E.coli* outer membrane integrity were highlighted while expressing such fused proteins (Georgiou *et al*., 1996; Stathopoulos *et al*., 1996). It is absolutely essential in protein engineering to restrict structural alterations to the minimum by avoiding unnecessary fusions for applications such as biosensors, ELISA, liposomal integration and enhanced antigenicity. Moreover, there is no report on systematic studies evaluating the effect of lipid modification on the folding and functioning of sensitive molecules such as enzymes. Therefore, it is worthwhile to develop enzymatic reporter lipoprotein with minimum alterations to study structure–function aspects and convert non-lipoproteins into lipoprotein form without fusion.

An easily assayable *Shigella* apyrase (NA) that acts on both organic and inorganic pyrophosphates to release phosphate (Bhargava *et al*., 1995) was chosen as a model protein for lipid modification. Based on the current knowledge and recent analysis of structural determinants for lipid modification, two strategies were designed without any significant fusion or alteration of sequence in the final product. In the first strategy, the periplasmic signal sequence of apyrase was modified at the C-region with consensus sequence for lipid modification. In the second strategy, the outer membrane Braun's lipoprotein (*lpp*) signal sequence was fused with the apyrase mature sequence. Both the modified apyrase inserts were ligated into pRSET-B vector, expressed in *E.coli* GJ1158 and characterized.

Materials and methods

Bacterial strains, growth conditions and plasmid

The expression host *E.coli* BL21 (DE3) and plasmid maintenance host *E.coli* DH5α were obtained from Invitrogen. *Escherichia coli* BL21 (DE3) has T7 RNA polymerase gene under the control of IPTG inducible *lac* promoter in chromosomal DNA. Another commercially available expression host, *E.coli* GJ1158, has T7 RNA polymerase gene under the control of salt inducible *proU* promoter in chromosomal DNA. Sodium chloride being inexpensive, the large-scale production of engineered lipoprotein using *E.coli* GJ1158 will be cost-effective (Poonam Bhandari and Gowri Shankar, 1997).

Native apyrase gene originally in clone pARC251 was subcloned into plasmid pSK, purchased from Stratagene. The *Pvu*II–*Hin*dIII 0.9 kb fragment (Bhargava *et al*., 1995) served as the template for amplification of native apyrase and the 738 bp *Nde*I–*Hin*dIII fragment containing only apyrase ORF cloned into pRSETB (Invitrogen) served as a template for engineering lipid modification forms (see below). The gene of Braun's lipoprotein (*lpp*), an abundant prototype outer membrane murein-bound lipoprotein, was obtained from the plasmid pKEN111 (Zwiebel *et al*., 1981).

Escherichia coli BL21 (DE3) and *E.coli* DH5α as host were grown at 37°C in Luria Bertani (LB) medium and their recombinants were grown in the presence of 100 μg/ml of ampicillin. *Escherichia coli* GJ1158 was grown in LB medium without NaCl and its recombinant was grown in the presence of 100 μg/ml of ampicillin.

Lipid modification strategy

Strategy I. As depicted in Figure 2, the signal sequence of apyrase (residues 1–19) was amplified by PCR using forward primer (apy-1, 5′GGGAATTCCATATGAAAACCAAAAA3′) with restriction site *Nde*I and reverse primer (apy-2, 5′CGCGGATCCACAACCAGCCAGGGGGATAAAAATCATATTTG3′) containing sequences for introducing lipobox and *Bam*HI site that codes for Gly and Ser. Mature sequence of apyrase (residues 24–246) was amplified by forward primer (apy-3) with *Bam*HI site (5′CGCGGATCCCTGAAGGCAGAAGGTTTTC3′) and reverse primer (apy-4) with *Hin*dIII site (5′CCCAAGCTTTTATGGGGTCAGTTCATT3′). The PCR products of the engineered signal sequence and mature apyrase sequences were restricted with *Bam*HI followed by ligation. The gene for LMA1 was then amplified with forward primer, apy1 and reverse primer, apy4. Amplified LMA1 gene and pRSET-B vector were restricted with *Nde*I and *Hin*dIII and ligated (Figure 2) to yield recombinant plasmid pLMA1.

Strategy II. Braun's lipoprotein signal sequence containing lipobox and the next residue Ser (residues 1–22) and two additional amino acids Gly and Ser due to the *Bam*HI site in the gene was amplified from *lpp* by PCR using forward primer (apy-5) with *Nde*I site (5′GGGAATTCCATATGAAAGCTACTAAACTG3′) and reverse primer (apy-6) containing *Bam*HI site (5′CGCGGATCCGGAGCAACCTGCCAG3′). Apyrase mature sequence (residues 24–246) was amplified as described in strategy I and ligated to *lpp* signal sequence. LMAII gene was again amplified using forward primer (apy-5) of *lpp* signal sequence and reverse primer (apy-4) of mature apyrase sequence. Amplified LMA II sequence and pRSET-B vector were restricted with *Nde*I and *Hin*dIII and ligated to give pLMAII plasmid (Figure 2). All constructs were sequenced using an Applied Biosystems 310 DNA sequencer to confirm their sequence.

Whole-cell assay for apyrase expression and screening for consistent hyper producer

Apyrase clones were screened for expression by simple colorimetric assay (Madan Babu *et al*., 2002). At 0.6 $OD_{600\ nm}$, *E.coli* BL21 (DE3) clones were induced with 1 mM IPTG and *E.coli* GJ1158 clones were induced with 0.3 M NaCl for 3 h at 37°C. A 1 ml volume of induced cultures was spun down and the cells were washed three times with 0.9% saline, suspended in 500 μl of Tris (50 mM)–EDTA (10 mM), pH 7.5 (TE), and left on ice for 20 min; 20 μl aliquots were added to TE containing ATP to give a final assay volume of 150 μl and a final ATP concentration of 5 mM. After 15 min of incubation at 37°C, cells were pelleted using a microfuge, the supernatant was transferred to a

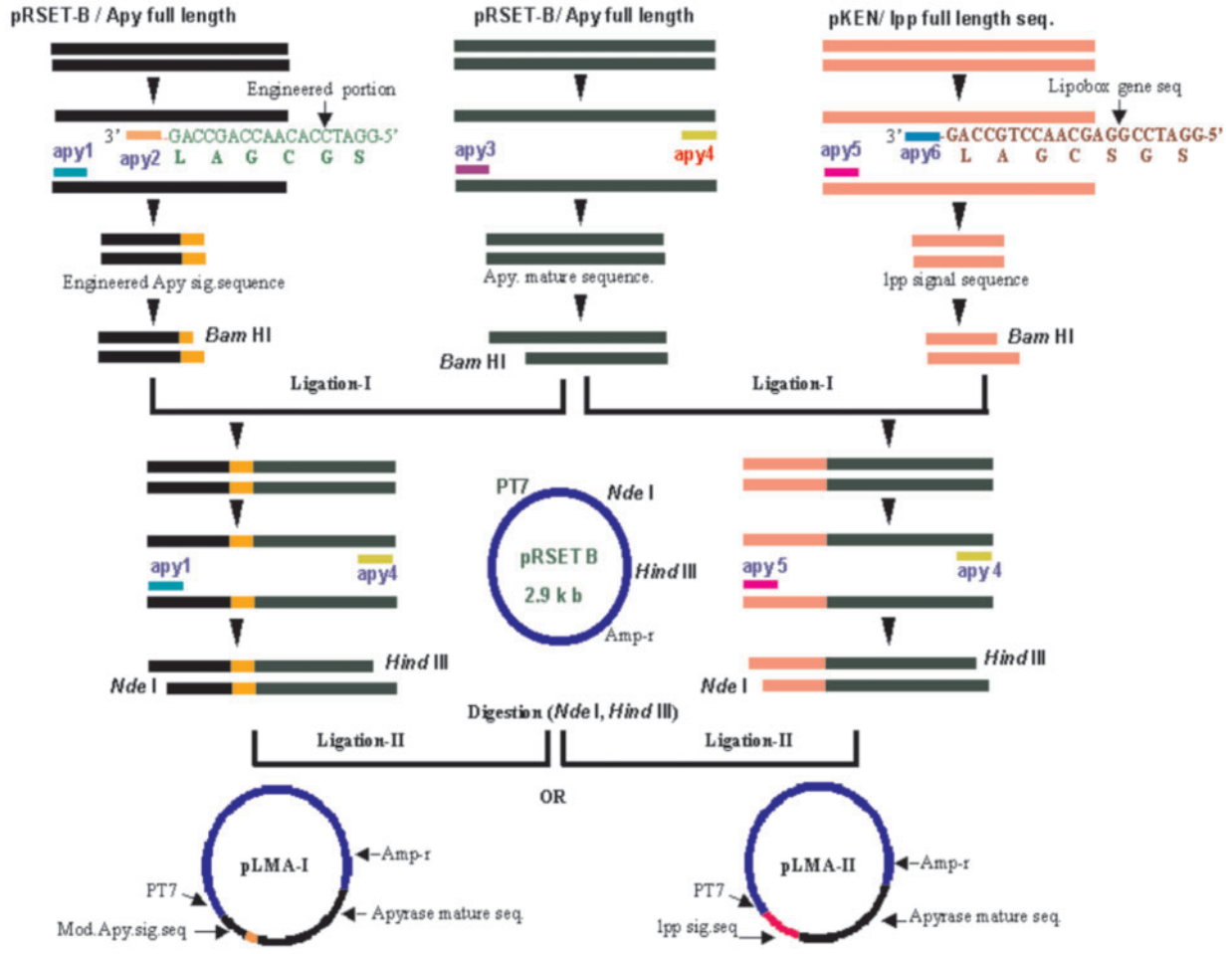

Fig. 2. Cloning strategies for lipid modification of apyrase. LMA-I: four amino acid residues at the C-terminal end of the apyrase signal sequence were replaced by a lipobox sequence (LAGC) by PCR and ligated with apyrase mature sequence (see Materials and methods, Strategy I, for details). LMA-II: Braun's lipoprotein signal sequence is fused with apyrase mature sequence (see Materials and methods, Strategy II, for details). LMA-I and LMA-II inserts were ligated to pRSET-B vector and the two recombinant plasmids pLMAI and pLMAII were expressed in *E.coli* BL21 (IPTG induction) and in *E.coli* GJ1158 (salt induction).

microplate well and 100 µl of Chen's reagent were added. After 15 min at room temperature, the absorbance was measured at 650 nm in a microplate reader and the amount of phosphate released was read from a phosphate calibration graph.

Routinely after every transformation, 20 colonies were randomly selected and screened by the above method to choose hyper-expressing clones. Normally we obtained 4–5 hyper-expressing clones. These were plated for single colonies and, from each set, five colonies were selected and assayed again. Only those which were consistently over-producing (activity in 1 ml of 1 O.D. culture >0.5 µmol/min) were retained as stable hyper-expressing clones. Glycerol stocks were prepared and frozen at −80°C for storage. Stab cultures were prepared and stored at 4°C. Hyper-expressing clones were confirmed by PCR for apyrase insert.

Cell fractionation

Volumes of 50 ml of induced native and lipid modified apyrase cultures were pelleted at 2500 *g* for 10 min at 4°C. The pellets were washed three times with 0.9% saline and suspended in 5 ml of TE containing 20% sucrose, 100 µg of lysozyme/ml final, pH 7.5. The cell suspensions were incubated for 20 min at room temperature and pelleted at 1000 *g* for 15 min at 4°C. The supernatant was used as the periplasmic fraction and stored at −20°C for further use. The pellets were suspended in 5 ml of TE. The cells were lysed using a French press at 16 000 psi and centrifuged at 2500 *g* for 10 min at 4°C to remove debris. Supernatants were centrifuged at 200 000 *g* for 30 min at 4°C to pellet total membrane, which was then suspended in 500 µl of 1% sodium lauryl sarcosine (which gives a detergent to protein ratio of ~5) and incubated for 1 h on ice for solubilization (Filip *et al*., 1973).

The sample was centrifuged at 200 000 *g* for 45 min at 4°C. Sarcosyl-solubilized supernatant was retained as the inner membrane fraction and stored at −20°C. The pellet, sarcosyl-insoluble portion, was suspended in 500 µl of TE and used as the outer membrane fraction. Apyrase activity was assayed in these fractions essentially as described above except that the assay was conducted in microplate wells with appropriate dilutions of the enzyme source added in place of cells.

Purification of apyrases by electroelution

Total protein amounting to 5 mg from each of inner membrane preparations of LMA-I, LMA-II and periplasmic fractions of native apyrase were resolved separately using 12% preparative SDS–PAGE. The gels were immersed in 1 M ice-cold KCl for 15 min. Owing to over-expression apyrases appeared as thick,

white bands, whose middle portion was easily excised from the gel. Following soaking of the gel pieces in 1 M NaCl for 15 min and then washing in water, apyrases were electroeluted in 25 mM Tris, 190 mM glycine, pH 8.9 at 100 mA for 3 h (Bhargava *et al*., 1995). The purity and molecular weight of the electroeluted samples were confirmed in 12% SDS–PAGE stained with Coomassie Brilliant Blue R-250 and also tricine SDS–PAGE (see below). The protein concentrations were measured by Lowry's method (Lowry *et al*., 1951).

Mobility shift in tricine SDS–PAGE

Tricine SDS–PAGE has been regularly used in the demonstration of mobility shift after bacterial lipid modification, especially in case of Braun's lipoprotein (Schlesinger, 1993). In fact, this method has been successfully used to identify the intermediates during *in vivo* and *in vitro* lipid modification (Ichihara *et al*., 1981). Therefore, it was used as first-line evidence of lipid modification of apyrase and also to establish the homogeneity of the preparations by electroelution. Samples of 5 μg of purified native and lipid-modified apyrases, as estimated by Lowry's method, were loaded on to 15% tricine SDS–PAGE gel. Tricine SDS–PAGE (15% T, 6% C gel and 18 h run at 100 V) was performed as described previously (Strom *et al*., 1993) and protein was stained with Coomassie Brilliant Blue R-250.

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry

Protein bands (NA, LMA-I and LMA-II) were excised from Coomassie Brilliant Blue G-250-stained 12% SDS–PAGE gel and washed three times in 1 ml of 50% acetonitrile containing 25 mM ammonium bicarbonate pH 8.0 for 30 min. The gel pieces were soaked in acetonitrile for 5 min. Acetonitrile was aspirated out and the gel pieces were dried in a Speed-Vac and re-swollen in 6 μl of 25 mM ammonium bicarbonate pH 8 containing 15 ng/μl of TPCK-treated Trypsin (Sigma). The swollen gel pieces were incubated overnight at 37°C. The gel slices were soaked in 100 μl of 50% acetonitrile containing 5% trifluoroacetic acid for 1 h to extract the peptides. The extraction was repeated and the two extracts were pooled and dried completely. The dried samples were dissolved in 6 μl of 50% acetonitrile containing 0.1% trifluoroacetic acid. The tryptic digests of the protein were analyzed using α-cyano-4-hydroxycinnamic acid (HCCA) or 2,5-dihydroxybenzoic acid (DHB) as matrix. In a modification for lipopeptides, 2 μl of sample were mixed with 2 μl of octyl glucoside (OG) (1% w/v) in 50% acetonitrile containing 0.1% trifluoroacetic acid. A 1 μl volume of sample was mixed with matrix solution (3-hydroxybenzoic acid 10 mg/ml in 50% acetonitrile containing 0.1% trifluoroacetic acid) on a MALDI target plate. MALDI-TOF mass spectra were acquired using a PerSeptive Biosystems Voyager-DE STR Biospectrometry workstation.

The mass spectrometer is fitted with a nitrogen laser that operates at 337 nm for ionizing the samples. The spectra were recorded in reflectron mode with delayed extraction in the positive ion mode. The accelerating voltage was set at 20 kV, the grid voltage at 72% and the delayed extraction was varied between 160 and 230 ns. The laser intensity was adjusted to obtain optimal signals and spectra were accumulated from 100 shots in the mass range m/z 800–4000. The mass accuracies using the calibrating peptides were 100 p.p.m. The mass spectra were processed using Data Explorer software supplied by the manufacturer and the list of monoisotopic peaks was also verified from the spectra manually. The proteins were identified by using MALDI fingerprint datasets with Mascot as described earlier (Chamrad *et al*., 2004). The search parameters used were partial methionine oxidation, allowed up to two missed cleavages for trypsin, by selecting NCBI database and taxonomy Eubacteria. With these experimental conditions, the tryptic digest of apyrase is expected to give 58 peptides excluding the N-terminal peptide. The N-terminal peptide could be identified only by using DHB as matrix along with OG. Detection of lipid-modified peptides by MALDI mass spectrometry will be considered in detail in another study. External calibration of spectra was effected using standard peptides des-Arg′-bradykinin, angiotensin, Glu′-fibrinopeptide-B and neurotensin (Gharahdaghi *et al*., 1999).

Comparison of specific activities of native and lipid-modified apyrases

Five units (1 unit is defined as the amount of enzyme required to release 1 μmol of phosphate per minute) each of native (periplasmic) and lipid-modified apyrases (inner membrane) and known concentrations of standard proteins were separated by 12% SDS–PAGE. The gel was stained with Coomassie Brilliant Blue G-250 and destained thoroughly. The protein bands corresponding to apyrases were cut from the gel and dye was eluted with 200 μl of 50% acetonitrile containing 25 mM ammonium bicarbonate. The blue color was read at 650 nm. The amount of native and lipid-modified apyrase was determined by comparison with the standard protein marker trypsin loaded alongside.

Haloperoxidase, peroxidase and catalase activities

Shigella apyrase is a member of the bacterial acid phosphatases and resembles very closely in sequence and 3D structure *E.blattae* acid phosphatase. Bacterial acid phosphatases, apart from phosphatase activity, exhibit haloperoxidase activity. Therefore, phosphatase, chloroperoxidase, peroxidase and catalase activities for native and lipid-modified apyrase were assayed as reported previously (Madan Babu *et al*., 2002).

Results

Twin strategies for engineering lipid modification

In the first strategy (lipid-modified apyrase-I, LMA-I), the *Shigella* apyrase signal sequence was modified using PCR to include a typical lipobox sequence, –LAGC–, seen in about 75% of bacterial lipoproteins (Von Heijne, 1983; Gennity *et al*., 1990) and two extra amino acids G and S were introduced after the cysteine because of the *Bam*HI site. In the second strategy, Braun lipoprotein's signal sequence (containing lipobox: LAGC) was directly fused to a mature sequence of apyrase, with an extra three new amino acids, S, G and S, following the C in the lipobox. As *Shigella* apyrase's N-terminus is an extended coil in the predicted 3D structure having no significant role in structure–function relationships (Madan Babu and Sankaran, 2002), these extra amino acids at the N-terminal end were presumed not to affect the activity. On the other hand, they could be useful in probing the role of the mature sequence adjacent to N-terminal Cys in lipoprotein targeting.

Table I. Whole-cell apyrase activity (μmol/min) corresponding to 1 ml of 1 O.D. culture

Protein	*E.coli* BL21DE3	*E.coli* GJ1158
NA	1.30	0.94
LMA-I	0.83	0.53
LMA-II	0.91	0.64
Vector control	ND[a]	ND

[a]ND, not detectable.

Expression of engineered lipoproteins in E.coli hosts

Apyrases, both native and lipid-modified, were expressed in both *E.coli* BL21 (DE3) and *E.coli* GJ1158. As can be seen from Table I, which contains data from a representative set, the expression was greater in *E.coli* BL21 (DE3) than in *E.coli* GJ1158. Even in the case of native apyrase the expression in *E.coli* GJ1158 was only 75% compared with *E.coli* BL21 (DE3). LMA-I and -II were expressed to the extent of 64 and 69%, respectively, compared with *E.coli* BL21 (DE3). However, a noteworthy observation was the lack of stable and consistent expression in *E.coli* BL21 (DE3). When screened after 7 days, one or two out of 10 colonies from *E.coli* BL21 (DE3) but nine out of 10 colonies from *E.coli* GJ1158 consistently over-expressed apyrase. Therefore, based on consistency, apyrase was expressed in *E.coli* GJ1158 for further experimentation.

Both LMA-I and LMA-II constructs resulted in similar expression levels in both *E.coli* expression systems (salt and IPTG inducible), but distinctly lower than that of native apyrase. This may perhaps be due to additional processing of lipoproteins through lipoprotein biosynthetic machinery even though both lipoproteins and non-lipoproteins share the same Sec machinery for transport. The salt inducible expression was lower by 30% for native and lipid-modified apyrases, indicating that the lower expression in *E.coli* GJ1158 was nothing to do with lipid modification. The reasons for instability in *E.coli* BL21 (DE3) upon storage are not clear.

Membrane localization of periplasmic apyrase after lipid modification

Lipoproteins are generally membrane bound, either to inner or outer membrane, normally requiring detergents to solubilize them. So far no predictive rule for targeting has been formulated for want of a clear understanding of the signals and factors involved. Therefore, the apyrase activity of native and lipid-modified forms was assayed in the periplasm, cytoplasm, inner membrane and outer membrane fractions of *E.coli* GJ1158 clones. As can be seen from Table II, native apyrase was specifically localized to the periplasm; a small amount (<10%) of cytoplasmic activity could presumably be due to contamination by the periplasmic fraction. In contrast, lipid-modified apyrases (LMA-I and -II) are localized to the sarcosyl-soluble membrane fraction normally considered to be inner membrane fraction (Filip *et al*., 1973).

Since it is possible that apparently the same activity could have resulted from fortuitous combinations of expression levels and specific activities, SDS–PAGE was performed to locate the protein in various fractions and correlate expression levels with activity. It should be clear from Figure 3A, in LMA-1 and -II (lanes 5 and 6), only inner membrane fractions

Table II. Activity (μmol/min) in whole cell and corresponding fractions from 1 ml of 1 O.D. culture

Clone	Whole cell	Cytoplasm	Periplasm	Total membrane	Inner membrane	Outer membrane
NA	0.89	0.07	0.87	ND[a]	ND	ND
LMA-I	0.51	ND	ND	0.53	0.49	ND
LMA-II	0.61	ND	ND	0.62	0.57	ND
Vector control	ND	ND	ND	ND	ND	ND

[a]ND, not detectable.

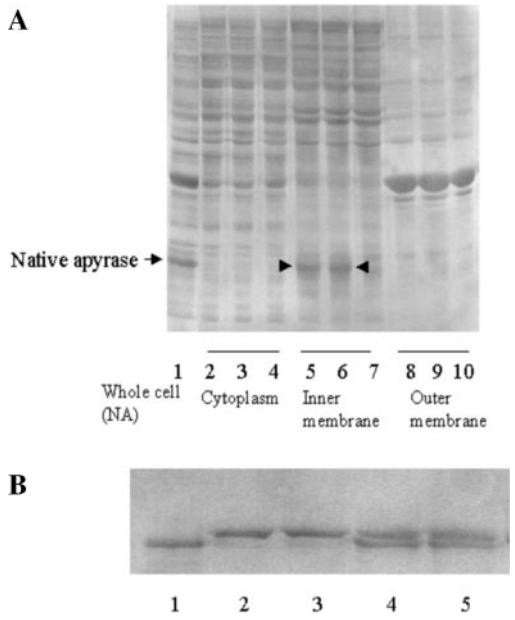

Fig. 3. (**A**) Localization of apyrase in the membrane fraction by SDS–PAGE. Equal amounts, 100 mg, of total protein from different fractions were loaded on to SDS–PAGE gel. Lane 1, NA whole cell lysate; lanes 2, 3 and 4, cytoplasmic fraction of NA, LMA-I and LMA-II, respectively; lanes 5, 6 and 7, inner membrane fraction of LMA-I, LMA-II and NA, respectively; lanes 8, 9 and 10, outer membrane fraction of LMA-I, LMA-II and NA, respectively. Lipid-modified apyrases LMA-I and -II are seen at 25 kDa in the inner membrane fraction only (lanes 5 and 6). (**B**) Mobility shift of lipid-modified apyrase in tricine SDS–PAGE. Both LMA-I and LMA-II (lanes 2 and 3) show identical mobility and move more slowly than the native apyrase (lane 1). When native apyrase was mixed with LMA-I and -II and ran separately in lanes 4 and 5, respectively, a doublet was formed corresponding to their individual mobility (lanes 1, 2 and 3).

contained lipid-modified forms as bands with comparable intensity at 25 kDa.

Functional *Shigella* apyrase has been purified to apparent homogeneity by electroelution from sarcosyl gels (Bhargava *et al*., 1995). In this study, we electroeluted the native and lipid-modified apyrases in periplasmic and membrane fractions respectively from SDS–PAGE after cold KCl staining. Subsequent SDS–PAGE showed a single band for lipid-modified apyrases migrating along with pure native apyrase (data not shown). The apparent homogeneity can be seen in tricine SDS–PAGE also (see below).

Demonstration of lipid modification by mobility shift in tricine SDS–PAGE and mass spectrometric analysis

Tricine SDS–PAGE has been shown to differentiate between lipid-modified forms of bacterial lipoproteins and it has in fact

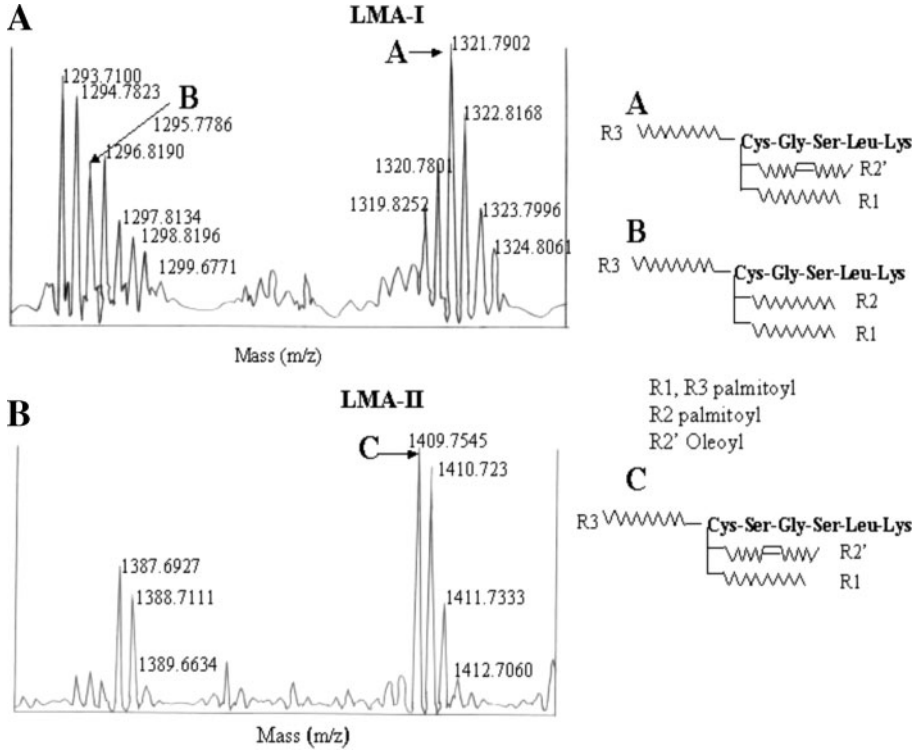

Fig. 4. (**A**) Mass spectrum of LMA-I. Two peaks shown by arrows with m/z 1321.7902 and 1295.7786 indicate the monoisotopic mass of *N*-palmitoyl-*S*-(1-palmitoyl-2-oleoyl)glyceryl-Cys-Gly-Ser-Leu-Lys and *N*-palmitoyl-*S*-(1,2-dipalmitoyl)glyceryl-Cys-Gly-Ser-Leu-Lys, respectively. (**B**) Mass spectrum of LMA-II. A single peak shown by an arrow with m/z value of 1409.7545 indicates the monoisotopic mass of *N*-palmitoyl-*S*-(1-palmitoyl-2-oleoyl)glyceryl-Cys-Ser-Gly-Ser-Leu-Lys.

been routinely used to identify partially modified forms of Braun's lipoprotein in gel-based assays for biosynthetic enzymes (Schlesinger, 1993; Sankaran and Wu, 1994). Therefore, as a first level of indication of bacterial lipid modification, it was exploited to differentiate the lipid-modified and native apyrases by virtue of their mobility shift. As expected, lipid-modified apyrases moved more slowly than native apyrase (Figure 3B). The doublet in lanes 4 and 5, corresponding to native and lipid-modified apyrases, ruled out electrophoretic artifacts.

Although the first-level of evidence from the membrane localization and mobility shift in tricine SDS–PAGE indicated that our products were perhaps lipid modified, it is possible that the signal sequences were not recognized properly by both lipoprotein and non-lipoprotein signal peptidases and therefore they were localized in the membrane and moved as higher molecular weight forms. To prove the point conclusively, we resorted to mass analysis of the N-terminal fragment.

Mass/charge analysis of tryptic digests of native and lipid modified apyrases (LMA-I, LMA-II) by MALDI-TOF mass spectrometry was performed using an improvised method described in the methodology. Previously mass spectrometric analysis was used to demonstrate lipid modification of surface lipoprotein from *S.pneumoniae* and outer membrane lipoproteins MlpA from *P.multocida* (Cullen *et al*., 2003).

The tryptic digest of LMA-I in addition to other peaks showed peaks at *m*/*z* 1321.7902 and 1295.7786 that corresponded to N-terminal lipidated peptides *N*-palmitoyl-*S*-(1-palmitoyl-2-oleoyl)glyceryl-Cys-Gly-Ser-Leu-Lys and *N*-palmitoyl-*S*-(1,2-dipalmitoyl)glyceryl-Cys-Gly-Ser-Leu-Lys, respectively. LMA-II showed a peak at *m*/*z* 1409.7545 that corresponded to its N-terminal lipidated peptide, *N*-palmitoyl-*S*-1-palmitoyl-2-oleoylglyceryl-Cys-Ser-Gly-Ser-Leu-Lys (Figure 4A and B). Owing to the high resolving power, isotopic resolution of the molecular ions is observed. These *m*/*z* values determined by mass spectrometry are in good agreement with the theoretically calculated *m*/*z* values of N-terminal lipidated peptides. Native apyrase, which was used as a control, did not show a peak that corresponded to the *m*/*z* of N-terminal lipidated peptides. When a list of peptide monoisotopic peaks arising from tryptic digests of lipid modified apyrase was submitted to Mascot (http://www.matrixscience.com) by selecting NCBI database and taxonomy Eubacteria for a similarity search, it showed similarity with *Shigella flexneri* apyrase with acceptable scores, which is significant for identification of the protein (Table IV). This confirms that the N-terminus of apyrase is lipid modified in the same way as normal bacterial lipoproteins (Gharahdaghi *et al*., 1999; Cullen *et al*., 2003).

The mass spectral analysis revealed another established fact characteristic of lipoproteins. The lipid acyl composition of diacylglyceryl group (Figure 1) is actually the same as that of phospholipids, especially its donor phosphatidylglycerol. Although the relative proportions could vary depending on the growth stage and metabolic status, R1 is always palmitate and R2 can be oleoyl or palmitoyl, the former being predominant (Cronan and Gelmann, 1975; Raetz, 1978, 1982; Schlesinger, 1993). LMA-I showed peaks at *m*/*z* 1321.7902 and 1295.7786, matching the peaks predicted based on oleoyl (predicted mass 1321.97) or palmitoyl residues (predicted mass 1295.94) at the second position. The peak for LMA-II is at *m*/*z* 1409.7545, corresponding to the oleoyl form (predicted mass 1409.06).

Table III. Comparison among specific activities of native and lipid-modified apyrases

Clone	Units (μmol/min)	Amount (mg)	Specific activity (units/mg)
NA	5	0.014	357
LMA-I	5	0.015	333
LMA-II	5	0.013	385

Table IV. Analysis of *m/z* of peptides by Mascot software

Protein sample	Score	No. of peptides matched	Sequence coverage (%)
NA	97	11	43
LMA-I	49	4	9
LMA-II	109	10	44

Demonstration of functional integrity of lipid-modified apyrases

All previous attempts to modify non-lipoproteins to lipoproteins addressed the mechanism of lipid modification for increasing the antigenicity, owing to the lipid acting as an adjuvant (Zwiebel *et al*., 1981; Cote-Sierra *et al*., 2002). However, the major concern in protein engineering is that the modification should not adversely affect the functional properties of the proteins to be modified. No workers, to the best of our knowledge, have deliberately looked into this aspect, even though it is known that lipoproteins come in a variety of structures and functions. Therefore, we employed an easily assayable heterologous non-lipoprotein enzyme with good turnover number (reported k_{cat} = 420) as a reporter lipoprotein (Bhargava *et al*., 1995). Apyrase structure has been predicted to have a major α-helical structure and rather flexible backbone giving rise to a variety of related activities; a number of known enzymatic variants under the super-family of bacterial acid phosphatases are known. We surmised that the N-terminal modification should not perturb the enzyme and, if it did, then we should be able to monitor it not only through loss of its activity but also in the alteration of its substrate specificity. The variation from phosphatase and haloperoxidase activities to pyrophosphatase activity seems to be subtle, as the active site residues are identical and the active site conformation is apparently the same.

The specific activity of an enzyme is a sensitive measure of integrity and proper folding of the protein to attain functional conformation. As can be seen from Table III, the specific activity of LMA-I and LMA-II remains essentially the same as that of native apyrase. As expected, there was no variation in specific activity after bacterial lipid modification. Apyrase does not possess any significant phosphatase or chloroperoxidase activity and therefore it has been proposed to form a new member of the bacterial acid phosphatase family. Our assays for chloroperoxidase, peroxidase, catalase and phosphatase were negative and the pyrophosphatase activity of lipid-modified forms matched that of native apyrase as reported previously (Madan Babu and Sankaran, 2002).

Discussion

Although bacterial lipoproteins originally contain respective signal sequences with appropriate structural attributes for lipid modification, prospective candidates for engineered lipid modification can come with or without their own signal sequences, depending on whether they are cytoplasmic or secretory proteins. Therefore, the applicability of different strategies for these proteins, the former requiring fusion of a typical lipoprotein signal sequence and the latter requiring mere modification of the signal sequence to contain predicted attributes for lipoprotein, was tested in this study. Since using the same protein for both the strategies makes better sense for comparison and evaluation, *Shigella* apyrase, a periplasmic protein from another organism, was engineered using either site-directed mutagenesis of residues at the C-terminal end of the original signal sequence of apyrase to introduce the typical lipobox (–LAGC–) present in about 75% of the lipoproteins (Von Heijne, 1983; Gennity *et al*., 1990) or fusion of the signal sequence of the abundant outer membrane lipoprotein, Lpp (Braun and Rehn, 1969), to the apyrase mature sequence. The fact that both led to bacterial lipid modification of this heterologous protein and anchored this periplasmic enzyme to the membrane suggests that both the strategies work well for this type of protein engineering and therefore its general applicability to a variety of proteins. Previously the beta domain of Staph-A protein, alkaline phosphatase and β-lactamase were lipid modified and expressed as the final fusion product (Stathopoulos *et al*., 1996). The fused part may alter the structure and function of a target protein. In the developed strategies, modified or fused signal peptide is cleaved off by signal peptidase-II and finally lipoprotein is expressed without fusion.

The membrane localization of both LMA-I and LMA-II indicated that all the attributes and signals necessary for lipid modification can be adequately provided in the signal sequence according to the predicted generalization (Von Heijne, 1983; Gennity *et al*., 1990; Madan Babu and Sankaran, 2002). Moreover, comparable levels of both modified proteins (even though LMA-II actually contained the intact signal sequence of the most abundant Braun's lipoprotein in the cell) indicate that the signal sequence does not significantly influence the levels of expression, at least as long as the same lipobox, –LAGC–, is used. In bacteria it is known that lipoproteins with the same lipobox could be expressed to varying extents. Therefore, the necessary post-translational attributes such as recognition by the biosynthetic machinery, especially the first enzyme, catalyzing the committed step, for maximum expression of these lipoproteins and the roles of different lipoboxes will be an interesting study. In this context, this engineered lipoprotein reporter enzyme will be of great help in identifying such factors, which have not been adequately probed into.

The indication for lipid modification was apparent from the retardation normally seen for lipoproteins and their biosynthetic intermediates in tricine SDS–PAGE. Mass spectrometry was used to establish lipid modification conclusively by analyzing the tryptic fragments of native and lipid-modified apyrase molecules. The mass derived from the spectral data could be accounted for only by assuming lipid modification and not by fragmentation of unmodified signal peptide as LMA-I and LMA-II had entirely different signal peptide sequences and the mass would be much higher than that observed for the lipopeptide fragment.

The mass spectrometric analysis revealed another established fact characteristic of lipoproteins. The lipid acyl

composition of the diacylglyceryl group (Figure 1) is actually the same as that of phospholipids, especially its donor phosphatidylglycerol. Although the relative proportions could vary depending on the growth stage and metabolic status, R1 is always palmitate and R2 can be oleoyl or palmitoyl, the former being predominant (Cronan and Gelmann, 1975; Raetz, 1978, 1982; Schlesinger, 1993). LMA-I showed peaks at *m/z* 1321.7902 and 1295.7786, matching with peaks predicted based on oleoyl or palmitoyl residues at the second position. The peak for LMA-II is at *m/z* 1409.7545, corresponding to the oleoyl form.

This observation further confirms that the engineered lipoprotein signal sequences have been properly recognized and modified in the same way as that of a typical lipoprotein by the lipid modification pathway enzymes, namely diacylglyceryl transferase (*lgt*), signal peptidase II and *N*-acyl transferase (Sankaran and Wu, 1994).

Recently, the sorting of lipoproteins to inner or outer membrane has been studied in some detail and it was shown to depend on amino acid residues at position 2 and 3, especially in its role to associate lipoprotein with lipoprotein outer membrane localization complex (Lol CDE) in the inner membrane. The Lol CDE complex in the inner membrane releases outer membrane-directed lipoproteins from the inner membrane in an ATP-dependent manner (Yakushi *et al*., 2000; Narita *et al*., 2002), leading to the formation of a lipoprotein–Lol A complex in the periplasm (Matsuyama *et al*., 1995; Tajima *et al*., 1998). This complex then interacts with outer membrane receptor Lol B, which anchors lipoproteins to the outer membrane (Matsuyama *et al*., 1997; Tanaka *et al*., 2001). The strong inner membrane retention or Lol CDE avoidance function occurs with Asp at position 2 (Masuda *et al*., 2002) and Asp, Glu, Gln or Asn at position 3 (Hara *et al*., 2003). His, Lys, Val, Ile, Ala, Cys or Thr at position 3 significantly decreases the inner membrane retention or Lol CDE avoidance functions of Asp at position 2 (Hara *et al*., 2003). It was reported that the distance between Cα of the main chain and the negative charge of the side chain makes only Asp suitable for this retention function.

Lipoproteins, LMA-I and LMA-II, have either Gly or Ser at position 2, respectively (Figure 2), but still retained in the inner membrane. This indicates that these small and uncharged amino acids also help in inner membrane retention by preventing complex formation with Lol CDE. In fact, LMA-II has Ser of Lpp, the major outer membrane lipoprotein, and it can be considered for all practical purposes identical with Lpp up to this point and apyrase can be considered as a fusion protein to the first 22 amino acids of Lpp prolipoprotein. Interestingly, both the LMAs have been retained in the inner membrane, as judged from their solubilization by sarcosyl and the activity seen in inverted vesicles.

The fact that both specificity (as judged from ATP hydrolysis, pyrophosphatase, peroxidase, catalase, chloroperoxidase and phosphatase) (Madan Babu *et al*., 2002) and specific activity were the same among the native and the lipid-modified forms supports the general belief that lipid modification at the N-terminal end as occurs in bacteria does not affect the folding of the rest of the protein and therefore retains its function, when still anchored to the membrane. It is inferred from our study that N-terminal lipid modification does not affect function and the same kind of lipid modification at the N-terminal end is possible for other proteins, perhaps except those in which the N-terminus might be important for its function.

In conclusion, employing a heterologous, secretory and sensitive enzymatic protein, we have demonstrated that bacterial lipid modification can be an effective strategy to introduce a hydrophobic anchor at the N-terminal end without perturbing the structure and function. Such a modification can be achieved using *E.coli* by either fusing the lipoprotein signal sequence or modifying an existing secretory non-lipoprotein signal sequence to contain lipobox. These strategies can be employed to lipid modify other commercially important hydrophilic proteins to acquire the benefits of lipoproteins. Lipoprotein apyrase can also be employed to understand the lipid modification mechanism operative in bacteria, such as membrane localization and lipobox variations.

Acknowledgements

The authors thank the Director, Centre for Biotechnology, the Director, Centre for Cellular and Molecular Biology and Professor Paul Rick, Chairman, Department of Biochemistry, Uniformed Services University of the Health Sciences, USA, for their support and M.Madan Babu, MRC Laboratory of Molecular Biology, Cambridge, UK, for critical reading and suggestions. The LG foundation, Lady Tata Memorial Trust (LTMT) and CSIR New Delhi are acknowledged for their financial support to S.Kamalakkannan and the All India Council for Technical Education (AICTE) for a financial grant to initiate this work.

References

Bhargava,T., Datta,S., Ramachandran,V., Ramakrishnan,R., Roy,R.K., Sankaran,K. and Subramanyam,Y.V.B.K. (1995) *Curr. Sci.*, **68**, 293–300.
Braun,V. and Rehn,K. (1969) *Eur. J. Biochem.*, **10**, 426–438.
Chamrad,D.C., Korting,G., Stuhler,K., Meyer,H.E., Klose,J. and Bluggel,M. (2004) *Proteomics*, **4**, 619–628.
Cote-Sierra,J., Bredan,A., Toldos,C.M., Stiijlemans,B., Brys,L., Cornelis,P., Segovia,M., Baetselier,P.D. and Revets,H. (2002) *Infect. Immun.*, **70**, 240–248.
Cronan,J.E.,Jr and Gelmann,E.P. (1975) *Bacteriol. Rev.*, **39**, 232–256.
Cullen,P.A., Miranda,L., Bulach,D.M., Cordwell,S.J. and Adler,B. (2003) *Plasmid*, **49**, 18–29.
Filip,C, Fletcher,G., Wulff,J.L. and Earhart,C.F. (1973) *J. Bacteriol.*, **115**, 717–722.
Gennity,J., Goldstein,J. and Inouye,M. (1990) *J. Bioenerg. Biomembr.*, **22**, 233–269.
Georgiou,G., Stephens,D.L., Stathopoulos,C., Poetschke,H.L., Mendenhall,J. and Earhart,C.F. (1996) *Protein Eng.*, **9**, 239–247.
Gharahdaghi,F., Weinberg,C.R., Meagher,D.A., Imai,B.S. and Mische,S.M. (1999) *Electrophoresis*, **20**, 601–605.
Hara,T., Matsuyama,S. and Tokuda,H. (2003) *J. Biol. Chem.*, **278**, 40408–40414.
Hayashi,S. and Wu,H.C. (1990) *J. Bioenerg. Biomembr.*, **22**, 451–471.
Ichihara,S., Hussain,M. and Mizushima,S. (1981) *J. Biol. Chem.*, **256**, 3125–3129.
Lowry,O.H., Rosbrough,N.J., Farr,A.L. and Randall,R.J. (1951) *J. Biol. Chem.*, **193**, 265–275.
Madan Babu,M. and Sankaran,K. (2002) *Bioinformatics*, **18**, 641–643.
Madan Babu,M., Kamalakkannan,S., Subramanyam,Y.V.B.K. and Sankaran,K. (2002) *FEBS Lett.*, **512**, 8–12.
Masuda,K., Matsuyama,S. and Tokuda,H. (2002) *Proc. Natl Acad. Sci. USA*, **99**, 7390–7395.
Matsuyama,S., Tajima,T. and Tokuda,H. (1995) *EMBO J.*, **14**, 3365–3372.
Matsuyama,S., Yokota,N. and Tokuda,H. (1997) *EMBO J.*, **16**, 6947–6955.
Narita,S., Tanaka,K., Matsuyama,S. and Tokuda,H. (2002) *J. Bacteriol.*, **184**, 1417–1422.
Pappalardo,M., Bousquet,E., Annino,A., Lombardo,G., Bernardini,R. and Ronsisvalle,G. (2003) *Farmaco*, **58**, 329–336.
Bhandari,P. and Gowri Shankar,J. (1997) *J. Bacteriol.*, **179**, 4403–4406.
Raetz,C.R.H. (1978) *Microb. Rev.*, **42**, 614–659.
Raetz,C.R.H. (1982) In Hawthorne,J.N. and Ansell,G.B. (eds), *Phospholipids*. Elsevier Biomedical Press, Amsterdam, pp. 435–477.
Sankaran,K. and Wu,H.C. (1994) *J. Biol. Chem.*, **269**, 19701–19706.
Schlesinger,M.J. (1993) *Lipid Modifications of Proteins.* CRC Press, Boca Raton.
Shigematsu,H., Ebihara,T., Yanahida,Y., Haruyama,T., Kobatake,E. and Aizawa,M. (1998) *J. Biotechnol.*, **74**, 23–31.

Stanley,P., Koronakis,V. and Hughes,C. (1998) *Microb. Mol. Biol. Rev.*, **62**, 309–333.
Stathopoulos,C., Georgiou,G. and Earhart,C.F. (1996) *Appl. Microb. Biotechnol.*, **45**, 112–119.
Strom,M.S., Nunn,D.N. and Lory,S. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 2404–2408.
Tajima,T., Yokota,N., Matsuyama,S. and Tokuda,H. (1998) *FEBS Lett.*, **439**, 51–54.
Tanaka,K., Matsuyama,S. and Tokuda,H. (2001) *J. Bacteriol.*, **183**, 6538–6542.
Von Heijne,G. (1983) *Eur. J. Biochem.*, **133**, 17–21.
Yakushi,T., Masuda,K., Narita,S., Matsuyama,S. and Tokuda,H. (2000) *Nat. Cell Biol.*, **2**, 212–218.
Zwiebel,L.J., Minukai,M., Nakamura,K. and Inouye,M. (1981) *J. Bacteriol.*, **145**, 654–656.

Received April 22, 2004; revised September 22, 2004; accepted October 21, 2004

Edited by Roger Williams