Multidomain organization of eukaryotic guanine nucleotide exchange translation initiation factor eIF-2B subunits revealed by analysis of conserved sequence motifs

EUGENE V. KOONIN

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894

(RECEIVED March 8, 1995; ACCEPTED May 16, 1995)

Abstract

Computer-assisted analysis of amino acid sequences using methods for database screening with individual sequences and with multiple alignment blocks reveals a complex multidomain organization of yeast proteins GCD6 and GCD1, and mammalian homolog of GCD6—subunits of the eukaryotic translation initiation factor eIF-2B involved in GDP/GTP exchange on eIF-2. It is shown that these proteins contain a putative nucleotide-binding domain related to a variety of nucleotidyltransferases, most of which are involved in nucleoside diphosphate-sugar formation in bacteria. Three conserved motifs, one of which appears to be a variant of the phosphate-binding site (P-loop) and another that may be considered a specific version of the Mg^{2+} -binding site of NTP-utilizing enzymes, were identified in the nucleotidyltransferase-related domain. Together with the third unique motif adjacent to the P-loop, these motifs comprise the signature of a new superfamily of nucleotide-binding domains. A domain consisting of hexapeptide amino acid repeats with a periodic distribution of bulky hydrophobic residues (isoleucine patch), which previously have been identified in bacterial acetyltransferases, is located toward the C-terminus from the nucleotidyltransferase-related domain. Finally, at the very C-termini of GCD6, eIF-2B ϵ , and two other eukaryotic translation initiation factors, eIF-4 γ and eIF-5, there is a previously undetected, conserved domain. It is hypothesized that the nucleotidyltransferase-related domain is directly involved in the GDP/GTP exchange, whereas the C-terminal conserved domain may be involved in the interaction of eIF-2B, eIF-4 γ , and eIF-5 with eIF-2.

Keywords: acetyltransferase; eukaryotic initiation factors; multidomain proteins; nucleotidyltransferase

Eukaryotic mRNA translation is mediated by numerous protein factors that form multisubunit complexes and are subject to intricate regulatory networks (Hershey, 1991; Merrick, 1992). cDNA sequences have been determined for most of these factors. Some of the deduced protein sequences contain distinct, conserved domains typical of functionally characterized protein families, allowing prediction of a specific activity. In several cases, these predictions are compatible with the results of biochemical experiments or have been tested in specifically designed experiments, whereas other predictions still await confirmation. Generally, attribution of a particular activity to an individual

polypeptide has been difficult because translation factors mostly function as heteromeric complexes. It has been shown that eIF-4A is an RNA helicase (Rozen et al., 1990; Pause & Sonenberg, 1992), eIF-4B, which contains the conserved eukaryotic RNA-binding domain, is an mRNA-binding protein (Naranda et al., 1994), and EF-1 α and EF-2 are GTPases (Lauer et al., 1984). eIF-5 also has been shown to possess GTPase activity (Trachsel & Staehelin, 1978; Brown-Luedi et al., 1982), even though the sequence of this protein does not contain the classical GTP-binding motifs (Das et al., 1993; Chakravarti & Maitra, 1993; E.V. Koonin, unpubl. obs.). In contrast, eIF-2 γ is predicted to be a GTPase based on the presence of the conserved GTP-binding domain (Hannig et al., 1992), but the activity remains to be demonstrated. Other amino acid sequence-based predictions awaiting experimental corroboration include RNA-binding properties of eIF-2 α , which is related to the bacterial ribosome protein S1 (Gribnikov, 1992), and unexpectedly, glutathione S-transferase activity of EF-1 γ (Koonin et al., 1994). For many

Reprint requests to: Eugene V. Koonin, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894; e-mail: koonin@ncbi.nlm.nih.gov.

Abbreviations: eIF, eukaryotic initiation factor; EF, elongation factor (Merrick, 1992); NDP, nucleoside diphosphate; NR, nonredundant; ORF, open reading frame.

other translation factors, analysis of the amino acid sequence so far has not provided functional clues, and their activity and/or mechanism of action remain uncertain.

eIF-2B is the guanine nucleotide exchange factor for eIF-2 (Konieczny & Safer, 1983; Hershey, 1991; Merrick, 1992). eIF-2B consists of five distinct subunits that in yeast are designated GCD1, GCD2, GCD6, GCD7, and GCN3 (Cigan et al., 1993). The cDNAs for each of these yeast proteins have been cloned and sequenced (Hannig & Hinnebusch, 1988; Hill & Struhl, 1988; Paddon et al., 1989; Bushman et al., 1993a). There is good correspondence between the experimentally determined M_r values of the eIF-2B in yeast and mammals (Pain, 1986), and, in addition, the human homolog of GCD6 (eIF-2Be) has recently been cloned and sequenced (Bushman et al., 1993a). Thus, it is thought that eIF-2B is conserved throughout the evolution of eukaryotes (Cigan et al., 1993). eIF-2B promotes the rapid recycling of eIF-2·GTP from GTP and eIF-2·GDP; the latter complex is released from the 40 S subunit after each delivery of a charged initiator tRNA^{Met} (Hershey, 1991; Bushman et al., 1993b; Hinnebusch, 1994). The GTP/GDP exchange catalyzed by eIF-2B is the key regulation point of eukaryotic translation initiation. This regulation occurs via phosphorylation of eIF-2 α at a specific serine residue that inhibits the GTP/GDP exchange activity of eIF-2B (Matts et al., 1983; Siekierka et al., 1984; Rowlands et al., 1988; Bushman et al., 1993b; Hinnebusch, 1993, 1994). Even though this mechanism of translation initiation regulation has been studied in considerable detail, specific activities of the individual subunits of eIF-2B are not known.

Amino acid sequence comparisons of the eIF-2B subunits have shown that GCD2, GCD7, and GCN3 form a cluster of related proteins, whereas GCD1 and GCD6 are related to each other (Paddon et al., 1989; Bushman et al., 1993a). Other than these observations and the similarity between GCD1, GCD6, and eIF-2Be, the mammalian homolog of GCD6, no similarities between the eIF-2B subunits and other proteins that could shed light on their specific activities have been reported.

Here I describe the results of amino acid sequence comparisons that reveal multidomain organization of GCD6 and GCD1 and allow one to predict the nucleotide-binding and protein-binding domains in these proteins.

Results

The N-terminal domain of GCD6 and GCD1 is related to nucleotidyltransferases: A new superfamily of nucleotide-binding domains

Comparison of the amino acid sequences of GCD6, eIF-2Be, and GCD1 with the NR sequence database using the BLASTP program (see Materials and methods) revealed, in addition to the previously reported similarity to GCD1, a statistically significant similarity (probability of occurring by chance, $P < 10^{-4}$) to a variety of NDP-sugar pyrophosphorylases (nucleotidyltransferases). The sequence most closely related to GCD6 ($P < 10^{-16}$) was that of an uncharacterized, partially sequenced ORF from archaeon *Sulfolobus acidocaldarius*, whose similarity to nucleotidyltransferases has been reported previously (Klenk et al., 1993). The most significant similarity between GCD6 and a known nucleotidyltransferase—glucose-1-phosphate thymidyltransferase from *Streptomyces griseus*—was associated with a P value of about 5×10^{-7} . A somewhat lower similar-

ity with the same set of proteins was observed for GCD1. The conserved region was located near the N-termini of GCD6, GCD1, and the related nucleotidyltransferases. Further analysis that included successive rounds of database screening using position-dependent weight matrices and additional BLAST searches (Koonin & Tatusov, 1994; Tatusov et al., 1994; see Materials and methods) resulted in the delineation of a new superfamily of (predicted) nucleotide-binding domains that is characterized by three conserved motifs located within a region consisting of 120–150 amino acid residues (Fig. 1A). Each pair of the conserved motifs could serve as a unique fingerprint of the superfamily, with the probability of both motifs being found in the same sequence by chance below 10^{-5} when the entire NR database was searched.

The N-terminal motif 1 and a part of motif 2 appear to comprise a specific version of the P-loop that is part of the substrate-binding site in a variety of nucleotide-utilizing enzymes (Saraste et al., 1990; Schulz, 1992). The motif typically consists of a glycine-rich loop preceded by a strongly predicted, hydrophobic β -strand and succeeded by an α -helix. In the new superfamily, the loop appears to vary with respect to both length and sequence. The classical form of the motif is found in the NDP-sugar pyrophosphorylases that cleave a β - γ bond in the respective NTP (Fig. 1A; Table 1). On the other hand, the motif is modified in the eIF-2B subunits, in cytidyltransferases that cleave an α - β bond in NDP, and in the GlgD protein (Fig. 1A; Table 1). In these sequences, even though the conformation typical of the P-loop is predicted to be preserved, the glycines in the loop are replaced by other residues (Fig. 1A). According to the secondary structure prediction results, the nucleotide-binding domain in the new superfamily is unusual in that the β -strand preceding the P-loop appears to be separated from the next β -strand by two rather than one α -helix (Fig. 1A). The first helix, the loop separating the helices, and the second β -strand contain the conserved residues comprising motif 2, in particular two positively charged residues that are found in the great majority of sequences (Fig. 1A). The C-terminal motif 3, separated from motif 2 by a variable region, corresponds to the B site of the nucleotide-binding motif and is predicted to have a β -strand-turn- α -helix structure, with an invariant aspartic acid located in the turn (Fig. 1A). The invariant acidic residue is predicted to chelate Mg^{2+} (Schulz, 1992). This motif was easily detectable in most of the proteins in the new superfamily, including the eIF-2B subunits, but could not be identified in NeuA and related proteins (Fig. 1A). Thus, 2 subunits of eIF-2B contain an N-terminal domain belonging to a previously uncharacterized superfamily of nucleotide-binding domains. It appears likely that this domain is directly involved in the GDP/GTP exchange mediated by eIF-2B.

The isoleucine patch hexamer repeat domain

Another region of moderate conservation, located C-terminally of the putative nucleotide-binding domain, was identified by the initial BLAST searches in GCD6, eIF-2Be, some of the nucleotidyltransferases in the new superfamily described above, and a variety of bacterial acyltransferases (Fig. 1B). The probability of finding this conserved region by chance in GCD6 and LpxA protein from *Escherichia coli* was about 4×10^{-5} . This region consists of multiple, imperfect repeats with a unit length

B

sec. struct.	ELLEEEELLEEELLEEELLEE	
GCD6_YEAST	333: LAQSKIGKCTAIGSGTKIGEGTKI	P32501
GCD6_YEAST	362: IGRNCGIGENIRIKNSF-IWDDCI	
GCD6_YEAST	396: IASNATLGSNVRLNDGCIIGFNVI	NA
eIFBe/RABIT	336: LGHGSILEENVLLGSGTVIGSNCSI	
eIFBe/RABIT	365: IGPCCIGDENVVLDRAV-LWKGQV	
eIFBe/RABIT	399: LCDHAEVKEQVTLKPHCVLTSQVV	
GCD1_YEAST	423: I-QSAAIGADAIVDPKCQISAHNV	P09032
GCD1_YEAST	452: LGTQANIGSRCRVAGSL-LFPGVHL	
GCD1_YEAST	480: ILENCIIGPMKSVSKCKLSNICYIV	
PSA1_YEAST	263: ISSTAKIGPDVVIGPNVTIGDGVRI	P41940
PPPI1_CAEL	331: IGEKTKLKES-IIAKGVVIGNGCLH	P42004
LPXD_RICRI	126: IGKNCYIGHNVVIEDDVIIGDNSII	P32202
THGA_ECOLI	134: IGNNVWIGSHVINPGVTIGDNSVI	P07464
NODL_RHIME	134: IGKHVWIGGGAILPGVTIGDHAVV	P28266
SATA_ENTFA	116: IGDNVWIGKDVVIMPGVIGDGAIV	L12033
CAPG/STAAU	85: IGNNVFIGINSIILPGVTIGNNVVV	U10927
FIRA/PASMU	120: IGANAVIEDGVTLGDHVVIGANCFV	S47342
HRA196/YEAST	133: IGDNVWIGGGVSIIPGVNIGKNSVI	S45154
YR43_BACSU	133: IGDQVWIGGRAVINPGVTIGDNAV	P37515
ORF/COXBU	120: IGDHVSIAIGAIVMNAIIGKNCII	S38220
RFB/SALCH	152: IGNCWIGDNAVILAGSEICDGCVI	S22616
CAT4_BACSH	74: FKGDVTIGNDVWIGQNVITMPGVII	P26840
LPXD_YEREN	111: LGENVSIGANAVIESGVVLGDNVVI	P32203
LPXD_ECOLI	224: IINGVVIDNQCIHNVVIGDNTAV	P21645
o256/ECOLI	155: IGEDVTVGHKVMLH-GCTIGNRVLV	U18997
Fbp/PSEAE	96: VGINAVILNGAKIGKYCIIGANALI	M82832
LPXA_ECOLI	20: IGANAHIGPPCIVGPHVEIGETVL	P10440
CAT/MORMO	113: IGDNVWIGSEAMIMPGIKIGDGAIV	X82455
SAT/ARATH	241: IGAGSCILGNITIGEGAKIGSGSVV	L34076
LPXA_RICRI	12: IAGAGLKGKGVKIGPYCIIGPEVVL	P32199
GLMU_ECOLI	274: IDTNVIEGNTVLGHRVKIGTGCVI	P17114
YEFH_ECOLI	135: IGDNVWIGENVTVLPGTIIGNGVVV	P37750
TMS_BACSU	326: IGPFAHIRPDSVIGNEVKIGNFEVI	P14192
THGA_ECOLI	128: YSFPITIGNNVWIGSHVINPGVTI	P07464
CAIE_ECOLI	98: IGRDALVGMNSVIMDGAIVIGESIV	P39206
LPXA_YEREN	2: IDKTAVIHPSSIVEEGAVIGAGVHI	P32201
CAT4_ECOLI	119: IGSEAMIMPGIKIGHGAVIGSRALV	P26838
CYSE_ECOLI	196: IREGVMIGAGAKILGNIEVGRGAKI	P05796
GLMU_ECOLI	268: HGRDVEIDTNVIEGNTVLGHRVKI	P17114
NIFP_AZOCH	83: HGACVVIGETAIEIGRDVTLHYGVTL	P23145
GLGC_ECOLI	373: VGRSCLRLRC-VIDRACVIEGMVI	P00584
GLGC_ANASP	368: IGPDTIIRRA-IIDKNARIGHDVKI	P30521
YEFH_ECOLI	129: ESSAVVIGQVWLGENVTVLPGTII	P37750
DAPD_ECOLI	134: VDEGTMVDTWATVGSQAQIGKNVHL	P03948
GLGC_BACSU	323: TVTSSVIMPDVTIGEHVVI-ENAVI	P39122

C

sec. struct.	hhhhh?????l111111HHHHHHH	hhhhhhhhhh11111111	
GCD6_YEAST	654 LFSALVSLYDNDIIEEDVIYKWDNV	10 KLTVKVWVWLQNADEESSSE	P32501
eIFBe/RABIT	709 IAKVLMGFYQLEILAETILSWFGQR	13 QQLQRFIQLWLKEAEEESSE	NA
GK1/HUMAN	358 FQKIVVLFYKAEVLSEEPILKYKDA	10 EQMKKFVWLKNAEESESE	D13630
eIF4g/RABIT	1343 LRMFFDALYDEDDVVEAFYSWESSK	12 KSVTAFFKWLREVEEESDH	L22090
eIF4g/HUMAN	1342 LRMFFDCLYDEEVEDAFYKWESEK	12 KSVTAFFTWLREAEESSEDN	Z34918
IF5_RAT	330 IPHILKEMYDADLLEEEVVISWSEKA	14 VKAEPFIKWLKEAEEESGG	Q07205
IF5_YEAST	342 LPKILVQLYNDIIEEEIMRFGTKS	14 RAAKPFITWLETAESDDDEE	P38431
consensus	U..UU..UY...UU.EB.UU.W....	FU.WL..ABEB....	

Fig. 1. Continued. eIF-2B subunits are multidomain proteins with three distinct, conserved domains. **A:** New superfamily of nucleotide-binding domains including the eIF-2B subunits and a variety of nucleotidyltransferases. The three conserved motifs are overlined. Initial alignment was generated by scanning the NR with two alignment blocks derived from BLAST outputs using the MoST2 program. The first block included motifs 1 and 2, and the second block consisted of motif 3. Positions of gaps between motifs 1 and 2 was determined using the MACAW program. Several sequences closely related to those shown are omitted. The P-loop (putative phosphate-binding site) and the putative Mg^{2+} -binding site are indicated by rectangles. The secondary structure predicted for all of the sequences (upper case) or the majority of the sequences (lower case) is shown as E, extended (β -sheet) conformation; h, α -helix; and l, loop (turn). The consensus line shows amino acid residues conserved in at least 90% of the sequences (upper case) or in the majority of the sequences (lower case); U, bulky hydrophobic residues (I, L, V, M, F, Y, and W); O, small residues (G, A, and S); J, positively charged residues (K and R); S, S or T; and dot, any residue. The distance from the protein N-terminus and the distance between the two alignment blocks are indicated by numbers for each sequence. Sequences are split into groups, with those within each group connected by BLAST scores above 100 (see Materials and methods). In NeuA and related proteins, a counterpart to motif 3 could not be identified. In GlgD, the sequence aligned with this motif is shown in lower case because it did not contain the conserved aspartic acid and is unlikely to be a functional Mg^{2+} -binding site. The accession number from SWISS-PROT (starting with P or Q), PIR (starting with C or S), or GenBank is given for each sequence. The eIF-2B ϵ sequence was from Bushman et al. (1993a). **B:** Repetitive "isoleucine patch" domain conserved in eIF-2B subunits, some nucleotidyltransferases, and acetyltransferases. The alignment was generated by screening the NR with an alignment block derived from the BLASTP output for GCD6 and consisting of four hexad repeats, which are shown by arrows, using the MoST program. Segments from the rabbit eIF-2B ϵ were added using the alignment with GCD6. For each of the eIF-2B subunits, three best-scoring four-repeat units are shown, whereas for each of the other sequences, only the best-scoring unit was included. Bulky hydrophobic residues recurring in every sixth position and the adjacent glycines are highlighted by bold type. Gaps were introduced manually in some of the sequences in order to restore the periodicity. Other designations are as in A. **C:** Uncharacterized C-terminal domain conserved in different eukaryotic translation initiation factors. The alignment was generated by screening the NR with two alignment blocks derived from the BLASTP output for the GK1 protein, using the MoST2 program. The consensus shows amino acid residues that are conserved in all sequences or in all but one. Other designations are as in A.

Typically, acyltransferases contain 20–30 hexamer repeats, which in some of them (e.g., *E. coli* LpxA) comprise most of the protein length. I found that GCD6 and GCD1, as well as GlmU, a nucleotidyltransferase whose substrate is an acetyl-sugar (Table 1), also contained more than 20 repeats each. In contrast, GlgC, a protein not known to bind acyl groups, contained only eight clearly distinguishable repeats.

An all- β -sheet structure was strongly predicted for the I-patch domains, in agreement with the previous observations (Fig. 1B; Vuorio et al., 1994). The exact function of the I-patch domain

is not known, but given that some acetyltransferases consist almost entirely of the hexads, it seems likely that this domain may be involved in binding acyl groups.

A conserved C-terminal domain in otherwise unrelated translation initiation factors

The C-terminal region of GCD6 and eIF-2 ϵ showed moderate similarity to the sequences of an uncharacterized human ORF product, designated GK1 ($P < 10^{-4}$), and eukaryotic transla-

Table 1. *eIF-2B* subunits and related proteins – Functional information

Protein/gene ^a	Organism(s)	Number of aa/quaternary structure ^b	Activity	Reaction	Metabolic pathway/physiological function	References
Eukaryotic translation initiation factors						
GCD6	Yeast	712; subunit of eIF-2B, which consists of GCN3, GCD1, GCD2, GCD6, and GCD7	GTP/GDP exchange factor subunit	eIF-2-GDP + GTP = eIF-2-GTP + GDP	Regeneration of eIF-2-GTP; translation regulation	Bushman et al., 1993a, 1993b; Cigan et al., 1993
eIF-2B ϵ	Rabbit	707; subunit eIF-2B	GTP/GDP exchange factor subunit	eIF-2-GDP + GTP = eIF-2-GTP + GDP	Regeneration of eIF-2-GTP; translation regulation	Bushman et al., 1993a, 1993b; Cigan et al., 1993
GCD1	Yeast	511; subunit of eIF-2B, which consists of GCN3, GCD1, GCD2, GCD6, and GCD7	GTP/GDP exchange factor subunit	eIF-2-GDP + GTP = eIF-2-GTP + GDP	Regeneration of eIF-2-GTP; translation regulation	Bushman et al., 1993a, 1993b; Cigan et al., 1993
eIF-5	Yeast, mammals	405; 429; monomer	GTPase; binds to the 40S-met-tRNAf-eIF-2-GTP ternary complex via eIF-2	40S-met-tRNAf-eIF-2-GTP = 80S-met-tRNAf-eIF-2-GTP + Pi	Formation of a functional 80S translation initiation complex	Das et al., 1993; Chakravarti and Maitra, 1993; Chaudhuri et al., 1994
eIF-4 γ	Mammals	1,396; subunit of eIF-4, which consists of eIF-4A, eIF-4B, eIF-4E, and eIF-4 γ	Component of the 43S and 48S preinitiation complexes	?	Formation of a functional 43S translation initiation complex	Joshi et al., 1994
Nucleotidyltransferases						
GlmU/ <i>glmU</i>	Bacteria	456	UDP-N-acetylglucosamine pyrophosphorylase	UTP + N-acetyl- α -D-glucosamine-1-P = PP _i + UDP-N-acetyl- α -D-glucosamine	Cell wall peptidoglycan biosynthesis	Mengin-Lecreux and Van Heijenoort, 1993
GlgC/ <i>glgC</i>	Bacteria, plants	431; homotetramer	ADP-glucose pyrophosphorylase	ATP + α -D-glucose-1-P = PP _i + ADP-glucose	First step in glycogen biosynthesis	Smith-White and Preiss, 1992
GalU/ <i>galU</i>	Bacteria	302	UDP-glucose pyrophosphorylase	UTP + α -D-glucose-1-P = PP _i + UDP-glucose	Lipopolysaccharide and capsular polysaccharide biosynthesis	Weissborn et al., 1994
RfbA/ <i>rfaA</i>	Bacteria	293	TDP-glucose pyrophosphorylase	dTTP + α -D-glucose-1-P = PP _i + dTDP-glucose	dTDP-L-rhamnose formation in O antigen biosynthesis pathway of lipopolysaccharide biosynthesis	Stevenson et al., 1994
CpsB/ <i>cpsB</i>	Bacteria	478	GDP-mannose pyrophosphorylase	GDP + D-mannose-1-P = PP _i + GDP-mannose	Slime polysaccharide colanic acid biosynthesis	Stevenson et al., 1991

KdsB/ <i>kdsB</i>	Bacteria	247	3-Deoxy-manno-octulosonate cytidyllyltransferase	CTP + 3-deoxy-manno- octulosonate = Ppi + CMP-3-deoxy-manno- octulosonate	Capsular lipopolysaccharide biosynthesis	Pazzani et al., 1993; Schnaitman and Klena, 1993
KpsU/ <i>kpsU</i>	Bacteria	246	3-Deoxy-manno-octulosonate cytidyllyltransferase	CTP + 3-deoxy-manno- octulosonate = Ppi + CMP-3-deoxy-manno- octulosonate	Capsular lipopolysaccharide biosynthesis	Pazzani et al., 1993; Schnaitman and Klena, 1993
NeuA/ <i>neuA</i>	Bacteria	419; probably homodimer	Acetylneuraminate cytidyl- transferase	CTP + N-acetylneuraminate = Ppi + CMP-N-acetylneur- aminate	Capsular lipopolysaccharide biosynthesis	Zapata et al., 1989
GlgD/ <i>glgD</i>	<i>B. subtilis</i>	343	?	?	Glycogen biosynthesis	Kiel et al., 1994
Mob/ <i>mob</i>	<i>E. coli</i>	194	Molybdopterin guanine nucleotide synthetase	Molybdopterin guanine dinucleotide formation	Mo cofactor biosynthesis	Johnson et al., 1991
Acyltransferases						
LpxA/ <i>lpxA</i>	Bacteria	262	Acyl-[acyl-carrier protein]- UDP-N-acetylglucosamine O-acyltransferase	(R)-3-hydroxydecanoyl- [acyl-carrier protein] + UDP-N-acetylglucos- amine = acyl-carrier protein + UDP-3-O- (3-hydroxydecanoyl)- N-acetylglucosamine	Lipid A biosynthesis	Coleman and Raetz, 1988; Kelly et al., 1993
FirA/ <i>firA</i> (<i>lpxD</i>)	Bacteria	340	UDP-3-O-[3-hydroxymyristoyl]- glucosamine N-acyltransferase	(R)-3-hydroxydecanoyl- [acyl-carrier protein] + UDP-3-O-(3-hydroxyde- canoyl)-N-acetylglucosa- mine = UDP-2,3-di(3-hy- droxydecanoyl)-N-acetyl- glucosamine + acyl-carrier protein	Lipid A biosynthesis	Dicker and Seetharam, 1992; Helander et al., 1993; Kelly et al., 1993
LacA(<i>ThgA</i>)/ <i>lacA</i>	Bacteria	203	Thiogalactoside acetyltransferase	Acetyl-CoA + β -D-galacto- side = CoA + 6-acetyl- β -D-galactoside	Lactose biosynthesis	Hediger et al., 1985
CysE/ <i>cysE</i>	Bacteria	273	Serine acetyltransferase	Acetyl-CoA + L-serine = CoA + O-acetyl-L-serine	Cysteine biosynthesis	Wigley et al., 1990
CatE/ <i>catE</i>	Bacteria	203	Acetyltransferase ?	?	Carnitine biosynthesis	Eichler et al., 1994
Cat/ <i>cat</i>	Inc plasmids	210 (INCFII NR79)	Chloramphenicol acetyltrans- ferase	Acetyl-CoA + chloramphen- icol = CoA + chloram- phenicol 3-acetate	Effector of chloramphenicol resistance	Parent and Roy, 1992

^a The protein/gene name is for *E. coli* unless otherwise indicated.^b The information on subunit structure is included where available.

tion initiation factors eIF-5 ($P \sim 0.1$) and eIF-4 γ (not significant statistically). In all these proteins, the conserved region was located at the very C-terminus; two conserved motifs that together comprised a unique fingerprint of this previously uncharacterized domain were identified (Fig. 1C). The original BLASTP values for this domain were not fully convincing by themselves, primarily because of its small size; nevertheless, using the MoST2 program, I observed that the probability of finding both motifs in the same sequence by chance was below 10^{-14} .

Both motifs were characterized by the conservation of acidic residues, particularly at the very C-terminus, and two conserved tryptophan residues (Fig. 1C). Secondary structure predictions indicated an α -helical structure for the conserved C-terminal domain. The sequence conservation was strictly limited to the C-terminal domain about 70 amino acid residues in length; no similarity could be detected between the rest of the proteins. GCD1 contained no detectable counterpart to the conserved C-terminal domain. The presence of this domain may indicate a common function in different translation initiation factor and an as yet uncharacterized human protein, perhaps binding to the same component of the translation machinery.

Discussion

The sequence analysis described above indicated that subunits of the GDP/GTP exchange factor eIF-2B have a complex, multidomain organization (Fig. 2). The N-terminal domain of these proteins belongs to a superfamily of nucleotide-binding domains that has not been described previously. It includes mostly nucleotidyltransferases (NDP-sugar pyrophosphorylases) that are involved in the biosynthesis of glycogen or bacterial capsular lipopolysaccharide, as well as recently sequenced eukaryotic homologs of these proteins. Interestingly, the *E. coli* Mob protein involved in molybdenum cofactor biosynthesis also belongs to the superfamily (Table 1). The domains comprising the new superfamily contain the two typical elements of many other nucleotide-binding domains—the P-loop and a putative Mg^{2+} -binding motif, as well as an additional unique motif. In spite of the common structural fold, there are many versions of the P-loop that are distinguishable by motif-searching methods (e.g., Bork & Koonin, 1994; Bork et al., 1995).

Two basically different mechanisms of action of a guanine nucleotide exchange factor such as eIF-2B are imaginable, namely a direct mechanism that involves binding of GDP and/or GTP by eIF-2B facilitating nucleotide exchange on eIF-2 or an indirect mechanism that would involve a modification of the eIF-2 conformation leading to dissociation of the GDP from the complex with eIF-2. The finding of a nucleotidyltransferase-related domain in eIF-2B is in favor of the direct mechanism, which involves GDP(GTP) binding. It has been shown that eIF-2B binds GTP and ATP (Dholakia & Wahba, 1989; Dholakia et al., 1989). Using photoaffinity labeling with 8-azido-GTP, the GTP-binding site has been tentatively mapped to the 40-kDa subunit of the mammalian eIF-2B, which corresponds to the yeast GCD7 (Dholakia et al., 1989). This conclusion, however, has to be treated with much caution because the same series of experiments has identified eIF-2 β as the GTP-binding subunit of eIF-2, whereas the conserved motifs typical of GTPases have been found in eIF-2 γ (Hannig et al., 1992).

Fig. 2. Domain organization of eIF-2B subunits and related proteins. The scheme is only roughly to scale. The C-terminal domain that is conserved in translation factors is indicated by an oval.

The eIF-2B subunits contain an altered P-loop, with no glycine residues remaining, and it is tempting to relate these deviations with the fact that the GDP/GTP exchange requires only binding, rather than hydrolysis, of a nucleotide. Caution is due with this interpretation, however, as KdtB, a bona fide NTP-hydrolyzing enzyme (Table 1), contains an equally dramatic modification of the P-loop (Fig. 1).

The second domain in the eIF-2B subunits is the I-patch repeat domain that previously has been described only in bacterial acyltransferases (Vaara, 1992; Vuorio et al., 1994). The identification of the I-patch domains in the eIF-2B subunits, as well as in two recently sequenced putative nucleotidyltransferases from yeast and nematode (PSA1 and PPP1, respectively, in Fig. 1A,B), shows that this repetitive motif antedates the eukaryotic/prokaryotic radiation. The function of this peculiar β -sheet domain apparently has to do with binding acyl groups. In support of this hypothesis, this domain is fully conserved in *E. coli* GlmU, a nucleotidyltransferase whose substrate is an acetylated sugar (Table 1). In contrast, GlgC, a nucleotidyltransferase whose substrate does not contain an acyl group, contains only a small number of I-patch repeats, which may be an evolutionary vestige. The function of the I-patch domain in the eIF-2B subunits containing the full complement of repeats typical of acyltransferases is unclear. One may speculate that either these proteins have an as yet unknown activity including binding of acyl groups, or the I-patch domain has been adopted to be involved in a different function, perhaps protein-protein interaction.

Finally, it was unexpected that one of the eIF-2B subunits shared a small C-terminal domain with two other translation ini-

tiation factors, as well as an uncharacterized human protein that perhaps also may be involved in translation. Given the remarkable conservation of two distinct motifs in this domain (Fig. 1C), one would expect a common binding site for all these proteins. The nature of this site remains unknown. One plausible speculation would be that the conserved C-terminal domain may bind eIF-2. eIF-2B has been isolated from eukaryotic cells mostly as a complex with eIF-2 (Konieczny & Safer, 1983; Matts et al., 1983). Very recently, it has been shown that eIF-5 also forms a complex with eIF-2, and it has been proposed that this binding is essential for the GTP hydrolysis within the ternary initiation complex 40S-met-tRNAf-eIF-2-GTP by eIF-5 (Chaudhuri et al., 1994). eIF-4 γ is a component of the 43S preinitiation complex that binds met-tRNAf-eIF-2 and eIF-4E-mRNA complexes to form the 48S complex (Joshi et al., 1994). One may hypothesize that binding of met-tRNAf-eIF-2 to the 43S complex may involve the interaction between eIF-2 and the conserved C-terminal domain of eIF-4 γ . Thus, the experimentally testable predictions are that: (1) GCD6 (eIF-2B ϵ) is the subunit of eIF-2B responsible for the interaction with eIF-2; (2) disruption of the C-terminal conserved domain will abolish binding of eIF-2B and eIF-5 with eIF-2; and (3) disruption of the C-terminal domain of eIF-4 γ will affect binding of met-tRNAf-eIF-2 to the 43S preinitiation complex. GK1 may be yet another eIF-2-binding protein. Interestingly, the C-terminal conserved domain was not found in the sequences of eIF-4 γ homologs from plants and yeast, perhaps suggesting a difference in the protein-protein interactions involved in the formation of translation initiation complexes.

In spite of the considerable long-term experimental effort, relatively little is known about the specific roles of the individual subunits of eIF-2B in the GDP/GTP exchange on eIF-2, and generally, in translation regulation. The experimentally testable predictions produced by the present computer analysis may help elucidate the functions of each of these important proteins.

Interestingly, the putative nucleotidyltransferase-related domain in the eIF-2B subunits is the second unexpected enzymatic domain discovered in guanine nucleotide exchange factors involved in translation by means of computer analysis of amino acid sequences. Previously, we found that EF-1 γ contains a glutathione S-transferase-related domain (Koonin et al., 1994). Perhaps a systematic analysis of the amino acid sequences of translation factors using sensitive methods for detection of conserved motifs may bring more unexpected findings.

Materials and methods

Sequences and databases

Amino acid and nucleotide sequences were extracted from the amino acid or nucleotide versions of the NR sequence database maintained at the National Center for Biotechnology Information (NIH). The amino acid sequence of the rabbit eIF-2B ϵ was from Bushman et al. (1993a).

Computer-assisted sequence analysis

Amino acid sequences were compared with the NR using programs based on the BLAST algorithm, version 1.4 (Altschul et al., 1990). The BLASTP program was used to screen the amino acid sequence database and the TBLASTN program was used to

screen the conceptual translation of the nucleotide sequence database in all six reading frames. BLAST computes the probability of each ungapped pairwise alignment produced in the database search to occur by chance, based on the statistical theory of asymptotic extremal distribution for high-scoring segments (Karlin & Altschul, 1990, 1993; Altschul et al., 1994). For all database searches, the amino acid substitution matrix BLOSUM62 (Henikoff & Henikoff, 1992, 1993) was used. Compositionally biased segments in the query sequences that may produce spurious hits in database searches were detected and masked using the SEG program (Wootton & Federhen, 1993; Altschul et al., 1994; Wootton, 1994).

Database search for conserved segments similar to a conserved motif presented as a multiple alignment block was performed using the MoST program (Tatusov et al., 1994). Briefly, the multiple alignment blocks are initially constructed by combining mutually consistent segments from the pairwise alignments produced by BLAST, using the CAP program (Tatusov et al., 1994). These blocks are converted into position-dependent weight matrices using the Dirichlet prior distribution mixture method that combines the observed amino acid residue frequencies for each column with a priori knowledge of amino acid relationships. These matrices are used for iterative database scanning, with the matrix recomputed at each iteration after including new sequence segments retrieved from the database. The cut-off for the motif searches is defined either as the probability of finding a sequence segment with a given score found in a database of a given size by chance (P value), or as the ratio of the expected number of sequence segments with a given score to the observed number (R value). Recently, the MoST method was generalized to search the database for two motifs separated by spacer sequences of varying length and to approximately calculate the associated P values; this procedure was implemented in the MoST2 program (S.F. Altschul, R.L. Tatusov, & E.V. Koonin, unpubl.).

For delineating protein superfamilies, a previously described iterative strategy including multiple, alternating rounds of database searches using BLAST and MoST was employed (Koonin & Tatusov, 1994). The two methods for database search are at least partially complementary because the motif search reveals many subtle similarities not detected by BLAST, but in some cases, BLAST finds related sequences with deviations in the conserved motifs that are missed by MoST (Koonin & Tatusov, 1994).

Single-linkage clustering of protein sequences on the basis of BLAST scores was performed using the CLUS program (R.L. Tatusov & E.V. Koonin, unpubl.).

Multiple amino acid sequence alignments were constructed using the MACAW program (Schuler et al., 1991).

Protein secondary structure was predicted using the multiple alignment-based neural network method, for which a 70–80% accuracy has been reported, depending on the number of sequences in the database that are related to the query (Rost & Sander, 1993).

Acknowledgments

I thank Dr. Roman L. Tatusov for providing unpublished computer programs. Bobby Baum has independently found and explored the similarity between the eIF-2B subunits and nucleotidyltransferases in the putative nucleotide-binding domain. I am grateful to him for sharing his unpublished results.

References

- Altschul SF, Boguski MS, Gish W, Wootton JC. 1994. Issues in searching molecular sequence databases. *Nature Genet* 6:119–129.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.
- Bairoch A, Bucher P. 1994. PROSITE: Recent developments. *Nucleic Acids Res* 22:3583–3589.
- Bork P, Holm L, Koonin EV, Sander C. 1995. The cytidyltransferase superfamily: Identification of the nucleotide-binding site and topology prediction. *Proteins Struct Funct Genet*. Forthcoming.
- Bork P, Koonin EV. 1994. A P-loop-like motif in a widespread ATP pyrophosphatase domain – Implications for the evolution of sequence motifs and enzyme activity. *Proteins Struct Funct Genet* 20:347–355.
- Brown-Luedi ML, Meyer LJ, Milburn SC, Yau PMP, Corbett S, Hershey JWB. 1982. Protein synthesis initiation factors from human HeLa cells and rabbit reticulocytes are similar: Comparison of protein structure, activities, and immunochemical properties. *Biochemistry* 21:4202–4206.
- Bushman JL, Asuru AI, Matts RL, Hinnebusch AG. 1993a. A protein complex of translational regulators of GCN4 mRNA is the guanine nucleotide-exchange factor for translation initiation factor 2 in yeast. *Mol Cell Biol* 13:1920–1932.
- Bushman JL, Foiani M, Cigan AM, Paddon CJ, Hinnebusch AG. 1993b. Guanine nucleotide exchange factor for eukaryotic translation initiation factor 2 in *Saccharomyces cerevisiae*: Interactions between the essential subunits GCD2, GCD6, and GCD7 and the regulatory subunit GCN3. *Mol Cell Biol* 13:4618–4631.
- Chakravarti D, Maitra U. 1993. Eukaryotic translation initiation factor 5 from *Saccharomyces cerevisiae*. Cloning, characterization, and expression of the gene encoding the 45,346-Da protein. *J Biol Chem* 268:10524–10533.
- Chaudhuri J, Das K, Maitra U. 1994. Purification and characterization of bacterially expressed mammalian translation factor 5 (eIF-5): Demonstration that eIF-5 forms a specific complex with eIF-2. *Biochemistry* 33:4794–4799.
- Cigan AM, Bushman JL, Boal TR, Hinnebusch AG. 1993. Evidence that GCD6 and GCD7, translational regulators of GCN4, are subunits of the guanine nucleotide exchange factor for eIF-2 in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 90:5350–5354.
- Coleman J, Raetz CR. 1988. First committed step of lipid A biosynthesis in *Escherichia coli*: Sequence of the *lpxA* gene. *J Bacteriol* 170:1268–1274.
- Das K, Cheshevich J, Maitra U. 1993. Molecular cloning and expression of cDNA for mammalian translation initiation factor 5. *Proc Natl Acad Sci USA* 90:3058–3062.
- Dholakia JN, Francis BR, Haley BE, Wahba AJ. 1989. Photoaffinity labeling of the rabbit reticulocyte guanine nucleotide exchange factor and eukaryotic initiation factor 2 with 8-azidopurine nucleotides. Identification of GTP- and ATP-binding domains. *J Biol Chem* 264:20638–20642.
- Dholakia JN, Wahba AJ. 1989. Mechanism of the nucleotide exchange reaction in eukaryotic polypeptide chain initiation. Characterization of the guanine nucleotide exchange factor as a GTP-binding protein. *J Biol Chem* 264:546–550.
- Dicker IB, Seetharam S. 1992. What is known about the structure and function of the *Escherichia coli* protein Fira? *Mol Microbiol* 6:817–823.
- Eichler K, Schunck WH, Kleber HP, Mandrand-Berthelot MA. 1994. Cloning, nucleotide sequence, and expression of the *Escherichia coli* gene encoding carnitine dehydratase. *J Bacteriol* 176:2970–2975.
- Gribskov M. 1992. Translational initiation factors IF-1 and eIF-2 alpha share an RNA-binding motif with prokaryotic ribosomal protein S1 and polynucleotide phosphorylase. *Gene* 119:107–111.
- Hannig EM, Cigan AM, Freeman BA, Kinzy TG. 1992. GCD11, a negative regulator of GCN4 expression, encodes the γ subunit of eIF-2 in *Saccharomyces cerevisiae*. *Mol Cell Biol* 13:506–520.
- Hannig EM, Hinnebusch AG. 1988. Molecular analysis of GCN3, a translational activator of GCN4: Evidence for posttranslational control of GCN3 regulatory function. *Mol Cell Biol* 8:4808–4820.
- Hediger MA, Johnson DF, Nierlich DP, Zabin I. 1985. DNA sequence of the lactose operon: The *lacA* gene and the transcriptional termination region. *Proc Natl Acad Sci USA* 82:6414–6418.
- Helander IM, Lindner B, Seydel U, Vaara M. 1993. Defective biosynthesis of the lipid A component of temperature-sensitive *firA* (*omsA*) mutant of *Escherichia coli*. *Eur J Biochem* 212:363–369.
- Henikoff S, Henikoff J. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* 89:10915–10919.
- Henikoff S, Henikoff J. 1993. Performance evaluation of amino acid substitution matrices. *Proteins Struct Funct Genet* 17:49–61.
- Hershey JW. 1991. Translational control in mammalian cells. *Annu Rev Biochem* 60:717–755.
- Hill DE, Struhl K. 1988. Molecular characterization of GCD1, a yeast gene required for general control of amino acid biosynthesis and cell-cycle initiation. *Nucleic Acids Res* 16:9253–9265.
- Hinnebusch AG. 1993. Gene-specific translational control of the yeast GCN4 gene by phosphorylation of eukaryotic initiation factor 2. *Mol Microbiol* 10:215–223.
- Hinnebusch AG. 1994. Translational control of GCN4: An in vivo barometer of initiation factor activity. *Trends Biochem Sci* 20:409–414.
- Johnson JL, Indermaur LW, Rajagopalan KV. 1991. Molybdenum cofactor biosynthesis in *Escherichia coli*. Requirement of the *chlB* gene product for the formation of molybdopterin guanine dinucleotide. *J Biol Chem* 266:12140–12145.
- Joshi B, Yan R, Rhoads RE. 1994. In vitro synthesis of human protein synthesis initiation factor 4 gamma and its localization on 43 and 48 S initiation complexes. *J Biol Chem* 269:2048–2055.
- Karlin S, Altschul SF. 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc Natl Acad Sci USA* 87:2264–2268.
- Karlin S, Altschul SF. 1993. Applications and statistics for multiple high-scoring segments in molecular sequences. *Proc Natl Acad Sci USA* 90:5873–5877.
- Kelly TM, Stachula SA, Raetz CR, Anderson MS. 1993. The *firA* gene of *Escherichia coli* encodes UDP-3-O-(R-3-hydroxymyristoyl)-glucosamine N-acyltransferase. The third step of endotoxin biosynthesis. *J Biol Chem* 268:19866–19874.
- Kiel JA, Boels JM, Beldman G, Venema G. 1994. Glycogen in *Bacillus subtilis*: Molecular characterization of an operon encoding enzymes involved in glycogen biosynthesis and degradation. *Mol Microbiol* 11:203–218.
- Klenk HP, Schleper C, Schwass V, Brudler R. 1993. Nucleotide sequence transcription and phylogeny of the gene encoding the superoxide dismutase of *Sulfolobus acidocaldarius*. *Biochim Biophys Acta* 1174:95–98.
- Konieczny A, Safer B. 1983. Purification of the eukaryotic initiation factor 2-eukaryotic initiation factor 2B complex and characterization of its guanine nucleotide exchange activity during protein synthesis initiation. *J Biol Chem* 258:3402–3408.
- Koonin EV, Mushegian AR, Tatusov RL, Altschul SF, Bryant SH, Bork P, Valencia A. 1994. Eukaryotic translation elongation factor 1 γ contains a glutathione transferase domain – Study of a diverse, ancient protein superfamily using motif search and structural modeling. *Protein Sci* 3:2045–2054.
- Koonin EV, Tatusov RL. 1994. Computer analysis of bacterial dehalogenases defines a large superfamily of hydrolases with diverse specificity. *J Mol Biol* 245:125–132.
- Lauer SJ, Burks E, Irvin JD, Ravel JM. 1984. Purification and characterization of three elongation factors, EF-1 α , EF-1 β , and EF-2, from wheat germ. *J Biol Chem* 259:1644–1648.
- Matts RL, Levin DH, London IM. 1983. Effect of phosphorylation of the alpha-subunit of eukaryotic initiation factor 2 on the function of reversing factor in the initiation of protein synthesis. *Proc Natl Acad Sci USA* 80:2559–2563.
- Mengin-Lecreux D, van Heijenoort J. 1993. Identification of the *glmU* gene encoding N-acetylglucosamine-1-phosphate uridylyltransferase in *Escherichia coli*. *J Bacteriol* 175:6150–6157.
- Merrick WC. 1992. Mechanism and regulation of eukaryotic protein synthesis. *Microbiol Rev* 56:291–315.
- Naranda T, Strong WB, Menaya J, Fabbri J, Hershey JW. 1994. Two structural domains of initiation factor eIF-4B are involved in binding to RNA. *J Biol Chem* 269:14465–14472.
- Paddon CJ, Hannig EM, Hinnebusch AG. 1989. Amino acid sequence similarity between GCN3 and GCD2, positive and negative translational regulators of GCN4: Evidence for antagonism by competition. *Genetics* 122:551–559.
- Pain VM. 1986. Initiation of protein synthesis in mammalian cells. *Biochem J* 235:625–637.
- Parent R, Roy PH. 1992. The chloramphenicol acetyltransferase gene of Tu2424: A new breed of cat. *J Bacteriol* 174:2891–2897.
- Pause A, Sonenberg N. 1992. Mutational analysis of a DEAD box RNA helicase: The mammalian translation initiation factor eIF-4A. *EMBO J* 11:2643–2654.
- Pazzani C, Rosenow C, Boulnois GJ, Bronner D, Jann K, Roberts IS. 1993. Molecular analysis of region 1 of the *Escherichia coli* K5 antigen gene cluster: A region encoding proteins involved in cell surface expression of capsular polysaccharide. *J Bacteriol* 175:5978–5983.
- Rost B, Sander C. 1993. Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol* 232:584–599.
- Rowlands AG, Panniers R, Henshaw EC. 1988. The catalytic mechanism of guanine nucleotide exchange factor action and competitive inhibition by phosphorylated eukaryotic initiation factor 2. *J Biol Chem* 263:5526–5533.

- Rozen F, Edery I, Meerovitch K, Dever TE, Merrick WC, Sonenberg N. 1990. Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. *Mol Cell Biol* 10:1134–1144.
- Saraste M, Sibbald PR, Wittinghofer A. 1990. The P-loop—A common motif in ATP- and GTP-binding proteins. *Trends Biochem Sci* 15:430–434.
- Schnaitman CA, Klena JD. 1993. Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol Rev* 57:655–682.
- Schuler GD, Altschul SF, Lipman DJ. 1991. A workbench for multiple alignment construction and analysis. *Proteins Struct Funct Genet* 9:180–190.
- Schulz GE. 1992. Binding of nucleotides by proteins. *Curr Opin Struct Biol* 2:61–67.
- Siekierka J, Manne V, Ochoa S. 1984. Mechanism of translational control by partial phosphorylation of the alpha subunit of eukaryotic initiation factor 2. *Proc Natl Acad Sci USA* 81:352–356.
- Smith-White BJ, Preiss J. 1992. Comparison of proteins of ADP-glucose pyrophosphorylase from diverse sources. *J Mol Evol* 34:449–464.
- Stevenson G, Lee SJ, Romana LK, Reeves PR. 1991. The *cps* gene cluster of *Salmonella* strain LT2 includes a second mannose pathway: Sequence of two genes and relationship to genes in the *rfb* gene cluster. *Mol Gen Genet* 227:173–180.
- Stevenson G, Neal B, Liu D, Hobbs M, Packer NH, Batley M, Redmond JW, Lindquist L, Reeves P. 1994. Structure of the O antigen of *Escherichia coli* K-12 and the sequence of its *rfb* gene cluster. *J Bacteriol* 176:4144–4156.
- Tatusov RL, Altschul SF, Koonin EV. 1994. Detection of conserved segments in proteins: Iterative scanning of sequence databases with alignment blocks. *Proc Natl Acad Sci USA* 91:12091–12095.
- Trachsel H, Staehelin T. 1978. Binding and release of eukaryotic translation initiation factor eIF-2 and GTP during protein synthesis initiation. *Proc Natl Acad Sci USA* 75:204–208.
- Vaara M. 1992. Eight bacterial proteins, including UDP-N-acetylglucosamine acyltransferase (LpxA) and three other transferases of *Escherichia coli*, consist of a six-residue periodicity theme. *FEMS Microbiol Lett* 76:249–254.
- Vuorio R, Harkonen T, Tolvanen M, Vaara M. 1994. The novel hexapeptide motif found in the acyltransferases LpxA and LpxD of lipid A biosynthesis is conserved in various bacteria. *FEBS Lett* 337:289–292.
- Weissborn AC, Liu Q, Rumley MK, Kennedy EP. 1994. UTP: Alpha-D-glucose-1-phosphate uridylyltransferase of *Escherichia coli*: Isolation and DNA sequence of the *galU* gene and purification of the enzyme. *J Bacteriol* 176:2611–2618.
- Wigley DB, Derrick JP, Shaw WV. 1990. The serine acetyltransferase from *Escherichia coli*. Over-expression, purification and preliminary crystallographic analysis. *FEBS Lett* 277:267–271.
- Wootton JC. 1994. Non-globular domains in protein sequences: Automated segmentation using complexity measures. *Comput Chem* 18:269–285.
- Wootton JC, Federhen S. 1993. Statistics of local complexity in amino acid sequences and sequence databases. *Comput Chem* 17:149–163.
- Zapata G, Vann WF, Aaronson W, Lewis MS, Moos M. 1989. Sequence of the cloned *Escherichia coli* K1 CMP-N-acetylneuraminic acid synthetase gene. *J Biol Chem* 264:14769–14774.