Protein Science (1994), *3*:2045–2054. Cambridge University Press. Printed in the USA.

Eukaryotic translation elongation factor 1γ contains a glutathione transferase domain—Study of a diverse, ancient protein superfamily using motif search and structural modeling

EUGENE V. KOONIN,[1] ARCADY R. MUSHEGIAN,[2,4] ROMAN L. TATUSOV,[1] STEPHEN F. ALTSCHUL,[1] STEPHEN H. BRYANT,[1] PEER BORK,[3] AND ALFONSO VALENCIA[3]

[1] National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894
[2] Department of Plant Pathology, University of Kentucky, Lexington, Kentucky 40546-0091
[3] European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900, Heidelberg, Germany

(RECEIVED June 23, 1994; ACCEPTED August 9, 1994)

Abstract

Using computer methods for multiple alignment, sequence motif search, and tertiary structure modeling, we show that eukaryotic translation elongation factor 1γ (EF1γ) contains an N-terminal domain related to class θ glutathione *S*-transferases (GST). GST-like proteins related to class θ comprise a large group including, in addition to typical GSTs and EF1γ, stress-induced proteins from bacteria and plants, bacterial reductive dehalogenases and β-etherases, and several uncharacterized proteins. These proteins share 2 conserved sequence motifs with GSTs of other classes (α, μ, and π). Tertiary structure modeling showed that in spite of the relatively low sequence similarity, the GST-related domain of EF1γ is likely to form a fold very similar to that in the known structures of class α, μ, and π GSTs. One of the conserved motifs is implicated in glutathione binding, whereas the other motif probably is involved in maintaining the proper conformation of the GST domain. We predict that the GST-like domain in EF1γ is enzymatically active and that to exhibit GST activity, EF1γ has to form homodimers. The GST activity may be involved in the regulation of the assembly of multisubunit complexes containing EF1 and aminoacyl-tRNA synthetases by shifting the balance between glutathione, disulfide glutathione, thiol groups of cysteines, and protein disulfide bonds. The GST domain is a widespread, conserved enzymatic module that may be covalently or noncovalently complexed with other proteins. Regulation of protein assembly and folding may be 1 of the functions of GST.

Keywords: conserved sequence motifs; glutathione *S*-transferase domain; motif search; structure modeling; translation elongation factor 1γ

Combination of functionally distinct domains in a single polypeptide is one of the general principles in the build-up of complex biochemical systems (Bork, 1992; Doolittle, 1992; Doolittle & Bork, 1993). In particular, several widespread enzymatic domains are known that may be combined with a variety of other domains and provide a function that is common to different processes. Examples of such universal domains include ATPase (Gorbalenya & Koonin, 1990; Milner-White et al., 1991), protein kinase (Hanks et al., 1988), and serine protease (Neurath, 1986).

We show here that glutathione *S*-transferase (GST) may be another "portable" enzymatic domain. GSTs are dimeric proteins that catalyze the conjugation of glutathione (GSH) with a variety of electrophiles, according to the equation GSH + EN = GS-E + NH, where EN is an electrophilic substrate (reviewed in Pickett & Lu, 1989; Fahey & Sundquist, 1991; Pemble & Taylor, 1992; Rushmore & Pickett, 1993; Dirr et al., 1994; Wilce & Parker, 1994). The best studied electrophilic substrates include chlorinated compounds and other xenobiotics, epoxides, and peroxides, e.g., hydrogen peroxide. It has been suggested that protection of cells against oxygen toxicity may be the pri-

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.
[4] Present address: Department of Microbiology, University of Washington, Seattle, Washington 98195.

mary function of GSTs (Fahey & Sundquist, 1991). In eukaryotes, GSTs are encoded by multiple genes. Comparison of amino acid sequences of mammalian GSTs has revealed 4 classes, designated α (GSTA), μ (GSTM), π (GSTP), and θ (GSTT). Recent studies have identified numerous, sometimes unexpected GST-related proteins, including bacterial reductive dehalogenases (La Roche & Leisinger, 1990; Orser et al., 1993) and β-etherases (Masai et al., 1993), plant stress-induced proteins (Czarnecka et al., 1988; Takahashi et al., 1989; Dominov et al., 1992), bacterial stringent starvation proteins (Toung & Tu, 1992), yeast nitrogen metabolism regulator URE2 (Coschigano & Magasanik, 1991), S-crystallins from cephalopod eye lens (Doolittle, 1988; Tomarev & Zinovieva, 1988; Tomarev et al., 1992), and several uncharacterized proteins from different sources with significant sequence similarity to GSTs (e.g., Zhao et al., 1993). Most of these sequences, with the exception of the crystallins, are related to GSTTs, resulting in a broad class of "GSTT-like" proteins (Pemble & Taylor, 1992; E.V. Koonin, unpubl. obs.). GSTA, GSTM, and GSTP are compact groups closely related to one another, whereas the GSTT-like class is much more heterogeneous, with many of the sequences showing only remote similarity to the other 3 classes of GSTs (Pemble & Taylor, 1992).

Using computer methods for database search, sequence motif analysis, multiple alignment, and tertiary structure modeling, we show here that the γ subunit of eukaryotic translation elongation factor 1 (EF1γ), which together with the β and δ subunits (EF1β and EF1δ) forms the guanine nucleotide exchange factor (Riis et al., 1990; Van Damme et al., 1990), contains an N-terminal GST-related domain. We propose that this domain may be involved in the regulation of the formation of multisubunit complexes containing EF1 through the balance between disulfide bonds and free thiol groups of cysteines. More generally, GST may be the key element of a novel system of regulation of protein folding and assembly.

Results

The N-terminal domain of EF1γ is related to GSTT-like proteins

In order to explore the relationships among GST-related proteins, we performed database searches with all relevant amino acid sequences. In the course of this analysis, moderate, but in many cases statistically significant similarity was revealed between the sequences of GSTT-like proteins and EF1γ. The highest scoring alignment was observed between GSTT from yeast *Issatchenkia orientalis* and the putative EF1γ from the fungus *Emericella* (*Aspergillus*) *nidulans*, with the probability of matching by chance (P) 7×10^{-12}. The same GST gave the P value 6×10^{-7} with EF1γ from yeast *Schizosaccharomyces pombe* and 3.1×10^{-5} with EF1γ from *Saccharomyces cerevisiae*. Many other GSTT/EF1γ pairs had P values between 10^{-4} and 10^{-2}. Thus, the similarity between EF1γ and GSTT-like proteins was obviously more significant than the similarity between the latter and other GST classes (typically, $P > 0.1$).

A multiple alignment was generated for 6 EF1γ sequences from evolutionarily diverse eukaryotes and 5 GSTT-like protein sequences (Fig. 1). The alignment was highly significant statistically, with adjusted score of over 20 standard deviations (computed using the OPTAL program), which is indicative of a genuine relationship (Gorbalenya et al., 1989). Analysis using the MACAW program (Schuler et al., 1991) revealed 2 strongly conserved alignment blocks (Fig. 1), with the probability of occurring by chance below 10^{-19}. The similarity between EF1γ and GSTT spanned almost the whole length of the latter set of proteins, i.e., about 200 amino acid residues. In EF1γ, this corresponded to the N-terminal domain comprising about one-half of the polypeptide. Strikingly, however, the *Emericella* protein that showed the highest similarity to GSTT among the EF1γ species consists of only 215 amino acid residues, a size typical of GSTs (Fig. 1). Thus, this protein appears to represent a stand-alone version of the GST-related domain of EF1γ.

The 2 conserved motifs define the entire superfamily of GST-related proteins

As indicated above, in BLAST searches, most of the class GSTT-like proteins showed very limited similarity to the GSTs of classes α, μ, and π. Therefore, we used a block search approach to probe the sequence conservation among all GST-related proteins. Analysis of the BLAST outputs for consistent alignments using the CAP program (Tatusov et al., 1994) showed that in all known GST-related proteins, the equivalents of the 2 motifs shown in Figure 1 are the only highly conserved blocks. When these blocks, derived from the BLAST alignments for GSTs belonging to each of the classes, were used to scan the database iteratively by the MoST procedure (Tatusov et al., 1994), the homologous segments from the great majority of GST-related proteins were selected without false positives (Fig. 2). In particular, the blocks derived from the sequences of GSTP or GSTM identified the GSTT-related proteins including EF1γ (curve I in Fig. 2 and data not shown); conversely, the GSTT-specific block selected GSTP and GSTM (curve II in Fig. 2). Very similar results were obtained with motifs I and II (compare curves I and II in Fig. 2), with the exception that motif II was detected in several additional sequences, with scores suggesting a genuine relationship. Interestingly, these included human valyl-tRNA-synthetase, glutaminyl-tRNA-synthetases from man and *Drosophila*, and translation elongation factor EF1β (Fig. 3). The similarity between the N-terminal domains of these aminoacyl-tRNA synthetases (aaRSs) and EF1γ has been noticed previously (Fett & Knippers, 1991; Hsieh & Campbell, 1991). Motif I appeared to be dramatically modified in ValRS and lacking in GlnRS and EF1β, in which motif II was located close to the N-terminus (Fig. 3).

These findings show that each of the 2 individual conserved motifs is a specific determinant of the GST superfamily, in spite of the fact that only 1 amino acid residue, namely P5 in motif I, is strictly conserved, with the D11 in motif II being replaced by glutamate in only 1 putative GST (Fig. 3).

3D model building and functional implications

The tertiary structures of EF1γ and GSTT are not yet known, but given the conservation of motifs I and II, approximate models may be based on the known structures of GSTM, GSTP, and GSTA. Our first approach to such knowledge-based modeling involved "threading" (Bryant & Lawrence, 1993) the sequences of EF1γ and other GSTT-related proteins containing motifs I and II through the core structure of GSTM (entry 1GST in the Protein Data Bank; Reinemer et al., 1991). This core contains the loops and adjacent secondary structural elements corre-

```
                    b1b!       aaaaaaaaaaaa   bb2bbb
EF1G  hum    3   GTLYT--Y-PENWRAFKALIAAQYSGAQVRVLSAPPHFHFGQTNRTPEFLRK-F
EF1G  As     3   GKLYT--Y-PENFRAFKALIAAQYSGAKLEIA---KSFVFGETNKSDAFLKS-F
EF1G  Tc     2   LTLWS-GVNPENARTHKLLAAAALANVAVTLK----ACEYGRENETAEYCRNCS
EF1G  En     3   GTLYT---RPFNPRSLAILAIAKANNLPLKIK-----TITSFKDATEEYLQL-N
EF1G  yeast  3   GTLYA-NF-RIRTWVPRGLVKALKLDVKVVTP----------DAAAEQFARD-F
EF1G  Sp     3   GTVYG-KI--GSPRVLFCVSVAAVAGVEVEHV------DVQPHNFPADLAAK-F
                 :***  *         *  ::     *:*:  *          :*
GST   yeast  3   ATVYI-KPHTPRGDWLASLGQYV--GLEIKTV----DYKSAEASKFEELF----
GST1  maize  3   MKLYGAVMSWNLTRCATALEEAG-SDYEIVPI----NFATAEHKSPEHLVR--N
GST32 maize  3   LKLYGMPLSPNVVRVATVLNEKG-LDFEIVPV----DLTTGAHKQPDFLAL--N
GST1  rat    2   LELYLDLLS-QPCRAIYIFAKKNNIPFQMHTV----ELRKGEHLSDAFAQV--N
ARP1  tobac  7   GFWYSP----FSRRVEWALKIKG-VKYEYIEE-------DRDNKSSLLLQS--N
consensus        ..&Y.............&.......h.&.....................h.....

                 I  glutathione binding                domain boundary
                      bb4bb bb5bb! aaaaaaaaaaaaa        ↓
EF1G  hum    PA-GKVPAFEGDDGFCVFESNAIAYYVSNEELRGSTP----------EAAAQVVQWVS
EF1G  As     PL-GKVPAFESADGHCIAESNAIAYYVANETLRGSSD----------LEKAQIIQWMT
EF1G  Tc     PC-GRYPVLQTEEG-CVFESNAILRHIARLDRSGGFL-----YGRTPLEGSQVDMWLD
EF1G  En     PL-GKIPTFVGADAYVLTESIAIALYDSNTTLLGTTG----------QEYASIIRWMA
EF1G  yeast  PL-KKVPAFVGPKGYKLTEAMAINYYLVKLSQDDKMKTQLLGADDDLNAQAQIIRWQS
EF1G  Sp     PL-QKMPVFVGKDGFPLSETLAIAFYLASLNKTRALN------GTTAEEKYETIQWVF
             **  ::* :*  :** *:* :**  *: :       *       * * **:     *:
GST   yeast  PL-KRVPALVTPNGFQLTELIAIVEYIVAKGSKPEL------SGKTTEERATNTRWLS
GST1  maize  PF-GQVPALQDG-DLYLFESRAICKYA-ARKNKPELL-----REGNLEEAAMVDVWIE
GST32 maize  PF-GQIPALVDG-DEVLFESRAINRYIASKYASEGTD-----LLPATASAAKLEVWLE
GST1  rat    PM-KKVPAMKDG-GFTLCESVAILLYLAHKYKVPDHW-----YPQDLQARARVDEYLA
ARP1  tobac  PIHKKVPVLIHN-GKRIVESMVILEYIDETFEGPSIL------PKDPYDRALARFWAK
consensus    Ph...&P.&.......U.E..hI.............................@..

EF1G  hum    FADSDIVPP-ASTWVFPTLGIM----HHNKQATENAKEEVRRILGLLDAYLKTRT-F
EF1G  As     FADTEILPA-SCTWVFPVLGIM----QFNKQATARAKEDIDKALQALDDHLLTRT-Y
EF1G  Tc     FSATELDAA-SEPFVHHAFRGE----PL----PANAMDRVHEVLRLFAWLETRT-F
EF1G  En     FGITEILPA-LGGWFNPLIGRA----NFNADNIYQSKDDTLARLKILDNHLCGRE-Y
EF1G  yeast  LANSDLCIQ-IANTIVPLKGGA----PYNKKSVDSAMDAVDKIVDIFENRLKNYT-Y
EF1G  Sp     FQMAAVGPM-FGQIIAPRVFGA----PYDEQAAKEAETAIALIFARFDEELASKT-Y
             %  : :     *  :   ::*       :::   ::   :  ::  :*  *:    %
GST   yeast  FFNSDFVQA-AGGYFM----GP------NDEIKQQSLQTMLSLLEYIDKHLSQSK-Y
GST1  maize  VEANQYTAA-LNPILFQVLISPMLGGTTDQKVVDENLEKLKKVLEVYEARLTKCK-Y
GST32 maize  VESHHFHPN-ASPLVFQLLVRPLLGGAPDAAVVEKHAEQLAKVLDVYEAHLARNK-Y
GST1  rat    WQHTTLRRSCLRTLWHKVMFPVFLGEQIRPEMLAATLADLDVNVQVLEDQFLQDKDF
ARP1  tobac  FLDDKVPAV-VKTFL------------RKGEEQEKDKEEVCEMLKVLDNELKDKK-F
consensus    &....&........&.............................&..&E..&.....@
                                                          D
                  II
                  aaaaaaaaaaaa
EF1G  hum    LVGERVTLADITVVCTLLWLY----KQVLEPSFRQAFPNTNRWFLTCINQPQFRAV 234
EF1G  As     LVGERITLADIVVTCTLLHLY----QHVLDEAFRKSYVNTNRWFITLINQKQVKAV 230
EF1G  Tc     LVGERMTVADVAVAFALQWHY--RLNGAEGEALTKKYRNAYRMYNTVMQQPKTVEV 208
EF1G  En     LVGETLSLADLFVLGIVQGAF----RFFLDKRWRDEHRNLSTWFERVHALPIVVDV  17
EF1G  yeast  LATENISLADLVAASIFTRYF----ESLFGTEWRAQHPAIVRCFNTVRASPFLKDE 211
EF1G  Sp     LVGSRLTLADIFFTCFLKFGA----TYVLTKSYLAKYTHIYRYYQTIYHQAKLDAI 208
             :    :  ****   :: :         :   :   *  : *:   :      :  :
GST   yeast  FTNNTILTADIFAFQIFAM------AKQFGVDF-THYPNVERFTGEVSQHPIIKNM   ?
GST1  maize  LAGDFLSLADLNHVSVTLCLF----ATPYASVL-DAYPHVKAWWSGLMERPSVQKV   8
GST32 maize  LAGDEFTLADANHALLPALTS---ARPPRPGCV-AARPHVKAWWEAIAARPAFQKT  14
GST1  rat    LVGPHISLADVVAITELMHPV----GGGCP-VF-EGRPRLAAWYRRV-EAAVGKDL  34
ARP1  tobac  FVGDKFGFADIAANLVAFWLGVFEEASGVVLVTSEKFPNFCKWRGEYINCSQIKES  26
consensus    &....&.&ADh.............................h...h.........
```

Fig. 1. Multiple alignment of the amino acid sequences of EF1γ and GSTT-related proteins. Asterisks designate identical residues and colons designate similar residues in the sequences of *Scizosaccharomyces pombe* (Sp) EF1γ and GST from *Issatchenkia orientalis* (designated yeast). Consensus shows amino acid residues that are conserved in all of the 11 aligned sequences (U designates a bulky aliphatic residue, namely I, L, V, or M; & designates a bulky hydrophobic residue, namely I, L, V, M, F, Y, or W; h designates any hydrophobic residue, namely I, L, V, M, F, Y, W, C, or A; and dot designates any residue). The amino acid residues conforming to the consensus are highlighted by bold type. The exclamation marks designate residues directly interacting with GSH. Secondary structure elements of the structural core of GSTs that are thought to be conserved in EF1γ are indicated (a designates α-helix and b designates β-strand). Additional abbreviations: As, *Artemia salina*; Tc, *Trypanosoma cruzi*; En, *Emericella nidulans*.

sponding to motifs I and II, plus the 10 other α-helices and β-strands, which constitute the GST fold, and which are conserved among GSTM, GSTP, and GSTA (Rushmore & Pickett, 1993; Sinning et al., 1993; Dirr et al., 1994). The result of this threading experiment is thus a set of models based on the GST core and the lowest-energy mapping of the EF1γ and GSTT sequences onto this structure (Bryant & Lawrence, 1993). The threading energies and the associated statistical scores, obtained under the constraint on the alignment of motifs I and II, are shown in Table 1. The sequence-structure matching scores for GSTT and EF1γ were highly significant, as were positive controls based on GSTM and GSTA and, in contrast to negative controls based on the sequences of glutathione reductase and glutathione peroxidase, 2 other GSH-binding proteins whose known structures are dissimilar to GST. "Energy scaffolds" (Nicholls et al., 1991; Bryant & Lawrence, 1993) for the native structure of GSTM and for the model of EF1γ are shown in Figure 4. The pattern of favorable pairwise residue interactions is very similar in the 2 scaffold models, offering further evidence that the tertiary structures of GSTT and EF1γ are indeed closely related to those of GSTA, GSTM, or GSTP. Threading without constraint on the alignment of motifs I and II yielded simi-

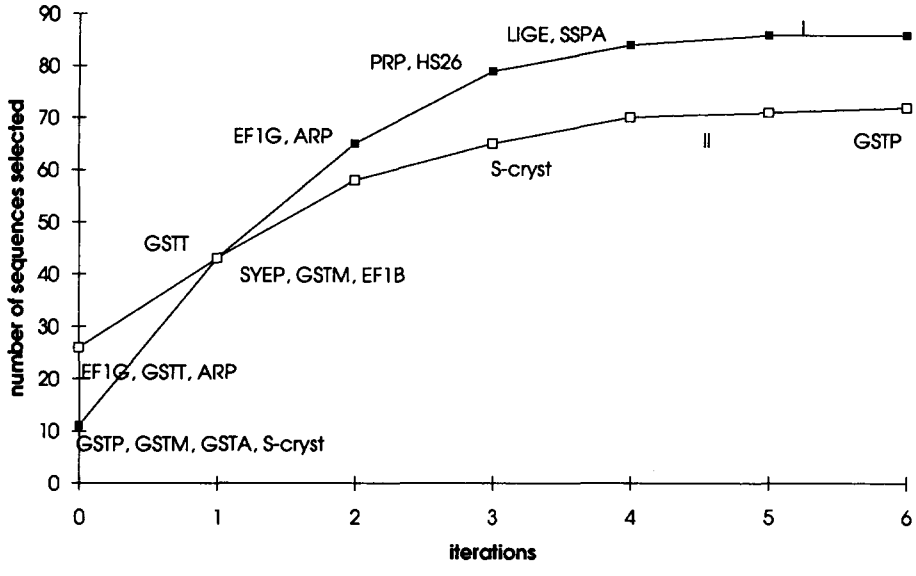

Fig. 2. Detection of different groups of GST-related proteins in an amino acid sequence database by iterative profile search. Progressive retrieval of sequence segments from the SWISS-PROT database in the course of iterative search with position-dependent weight matrices is illustrated. The search was initiated with conserved blocks derived from BLAST outputs. Each group of sequences is associated with the iteration at which at least 1 sequence from this group was first detected. The indicated number of selected sequences includes unique segments; identical segments were omitted. The searches were run with the expected/observed segments ratio of 0.001 as the cutoff. I (filled squares), search with the block containing motif I and derived from the BLAST output for human GSTP (GTP_HUMAN). II (open squares), search with the block containing motif II and derived from the BLAST output for yeast EF1γ (EF1G_YEAST); in this search, no GSTAs were detected.

Table 1. *Scores for selected GST-related proteins threaded through the GSTM structure*[a]

Protein	PI	ΔG	Z	P
GTB1_RAT	100	−379.05	9.12	3.81e − 10
GTP_HUMAN	30	−326.62	8.36	5.86e − 10
GTH1_HUMAN	19	−363.63	8.39	1.39e − 08
EF1G Tc	14	−329.36	8.74	5.89e − 08
GTY2_ISSOR	11	−300.09	8.02	1.62e − 07
SSPA_ECOLI	16	−311.43	7.98	7.36e − 07
GTT1_RAT	15	−298.19	6.85	1.05e − 4
EF1G_YEAST	14	−292.57	6.94	1.28e − 04
EF1G_HUMAN	11	−276.76	7.21	9.23e − 04
GSHC_BOVIN (1GP1)	5	−232.34	6.49	6.00e − 02
GSHR_HUMAN (4GR1)	4	−256.66	5.17	1.00e + 00

[a] PI is the percentage of amino acid residue identity within the core substructure, ΔG is the threading energy in nominal kT units, Z is Z score in standard deviation units, i.e., the departure of ΔG from the mean of the distribution for random sequences; P is the probability that this energy would be observed by chance, in threading random sequences with the same composition ($e - n = 10^{-n}$). GTB1_RAT, GTP_HUMAN, and GTH1_HUMAN, which represent the μ, π, and α classes of GST, respectively, are positive controls. GSHC_BOVIN (glutathione peroxidase, PDB entry 1GP1; Epp et al., 1983) and GSHR_HUMAN (glutathione reductase; PDB entry 4GR1; Janes & Schulz, 1990), 2 GSH-binding proteins with structures unrelated to that of GST, are negative controls. Threading scores were calculated as described previously (Bryant & Lawrence, 1993), except that the optimal alignments were identified by a fast heuristic procedure (S.H. Bryant, unpubl.), and P values were determined empirically, by taking the rank order of the indicated sequence versus threading of 100 random permutations. In threading GSTT and EF1γ sequences, the alignments of subsequences containing motifs I and II were constrained to their known positions in GSTM. The negative controls provide a conservative estimate of the values expected by chance; they are strictly comparable to the positive controls and test sequences only in their P values, which account for the larger number of accessible alignments due to the lack of the constraint on the conserved motifs.

lar models, which means that the similarity between the GSTM, GSTT, and EF1γ structures is suggested independently by both threading and motif analysis (not shown).

Our second approach involved prediction of the secondary structure for EF1γ and GSTT (Fig. 1). A tentative alignment with GSTM, GSTP, and GSTA was generated by superposition of conserved motifs and secondary structure elements (not shown). Homology-based modeling using the WHATIF program (Vriend, 1990) indicated that despite the low sequence similarity, the N-terminal domain of EF1γ is likely to assume a GST-like structure. The reliability of the model is much higher in the conserved core, but the overall atomic contact quality index was within the range typical of protein structures modeled by homology (Vriend & Sander, 1993). Thus, the model may be useful in mapping conserved sequence motifs to the 3-dimensional structures (Fig. 5).

All GSTs are dimers, with each subunit binding 1 GSH molecule. The subunit is further divided into the N-terminal domain that consists of a 5-stranded β-sheet and the α-helical C-terminal domain. The GSH-binding site is formed by residues in the N-terminal domain, whereas both the N-terminal and the C-terminal domain contribute to the binding site for the electrophilic GSH acceptor (Ji et al., 1992, 1994; Liu et al., 1992; Reinemer et al., 1992; Rushmore & Pickett, 1993; Sinning et al., 1993; Dirr et al., 1994). The GST-related domain of EF1γ is predicted to have a similar organization (Fig. 5). The essential tyrosine near the N-terminus and motifs I and II belong to the core that appears to be conserved in all GST-related proteins, whereas there are significant structural variations in some of the loops. The conserved tyrosine is at the end of β1, where it is able to contact the bound GSH. Motif I consists of β4, β5, and an α-helix, with the latter extending to the interdomain hinge (Figs. 1, 4). The conserved glutamic acid and proline in motif I (E21 and P5, respectively, in Fig. 3) directly contact GSH, whereas the preceding hydrophobic residues contribute to the hydrophobic binding pocket (Fig. 5). Motif II is in the C-terminal domain and includes a long, conserved loop and a subsequent

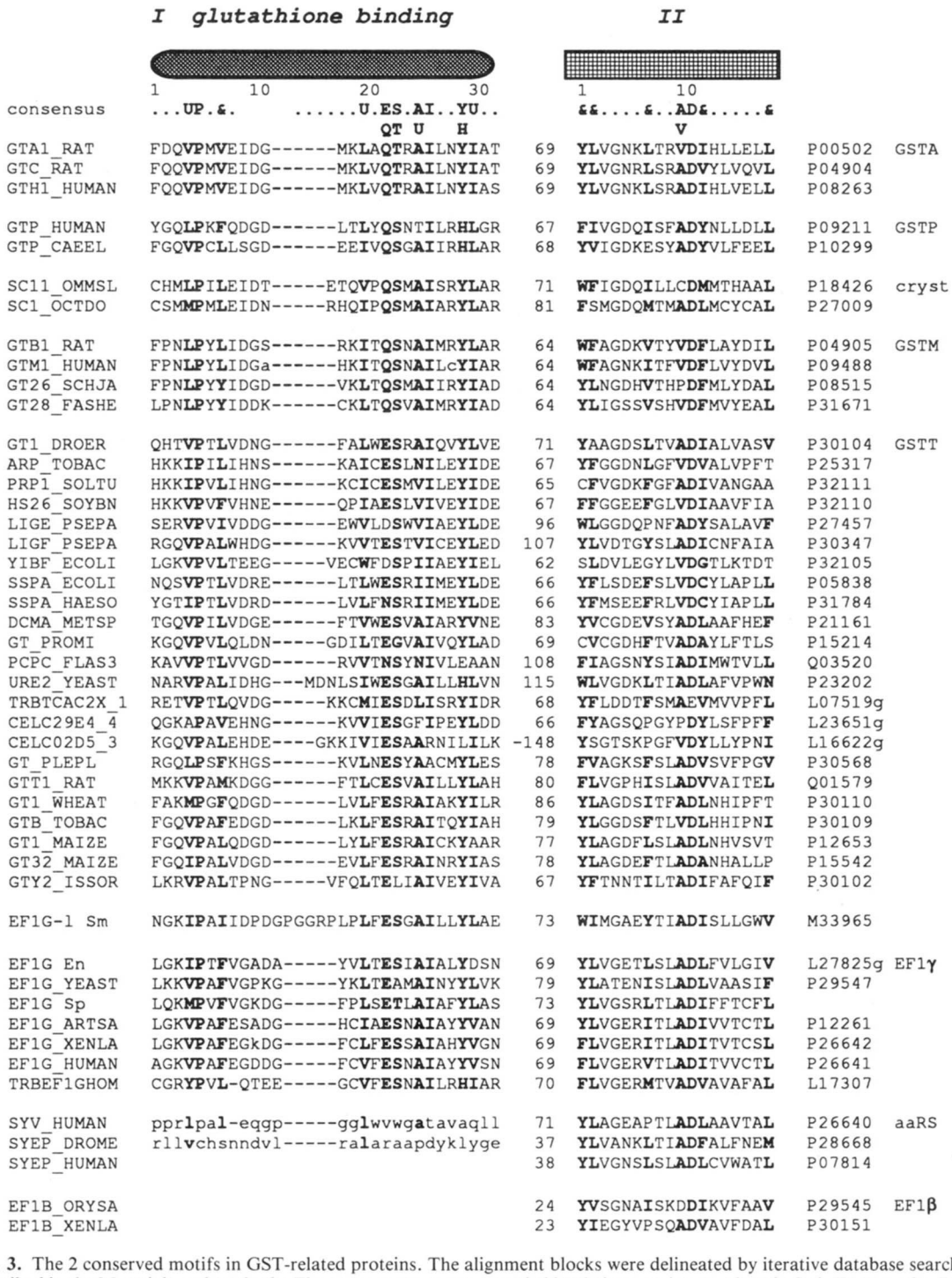

Fig. 3. The 2 conserved motifs in GST-related proteins. The alignment blocks were delineated by iterative database search as described in the Materials and methods. The sequences are accompanied by their accession numbers in SwissProt or GenBank (g). The *S. pombe* EF1γ sequence was from Momoi et al. (1993). Sequence segments belonging to different groups are separated by blank lines. GSTT designates the large group of related proteins, of which only some have been identified as actual GSTs. The consensus shows amino acid residues that are found in over 50% of the sequences in each of the groups. Residues conforming to the consensus are highlighted by bold type. The distance between the 2 conserved blocks is indicated for each sequence. EF1G-l (after EF1γ-like) is a previously uncharacterized putative protein from *Serratia marcescens* (Sm) that showed approximately the same level of similarity to EF1γ and GSTTs. This protein has been identified by nucleotide database search using TBLASTN and is encoded by nucleotides 1–532 of the GenBank entry SMAPHOAA (alkaline phosphatase gene). SYV is valyl-aaRS and SYEP is glutaminyl-aaRS. In the sequences of SYV_HUMAN and SYEP_DROME, the sequence corresponding to motif I in alignments with EF1γ (not shown) is in lowercase to indicate the low level of conservation. In the sequences of SYEP_HUMAN and EF1β, the counterpart to motif I could not be detected. In the sequence of the putative nematode protein CELC02D5_3, the 2 conserved motifs could be unequivocally identified but were swapped. Where available, the sequence names were directly from SwissProt. Other abbreviations: Fb, *Flavobacterium*; TRB, *Trypanosoma brucei*; CEL, *Caenorhabditis elegans*.

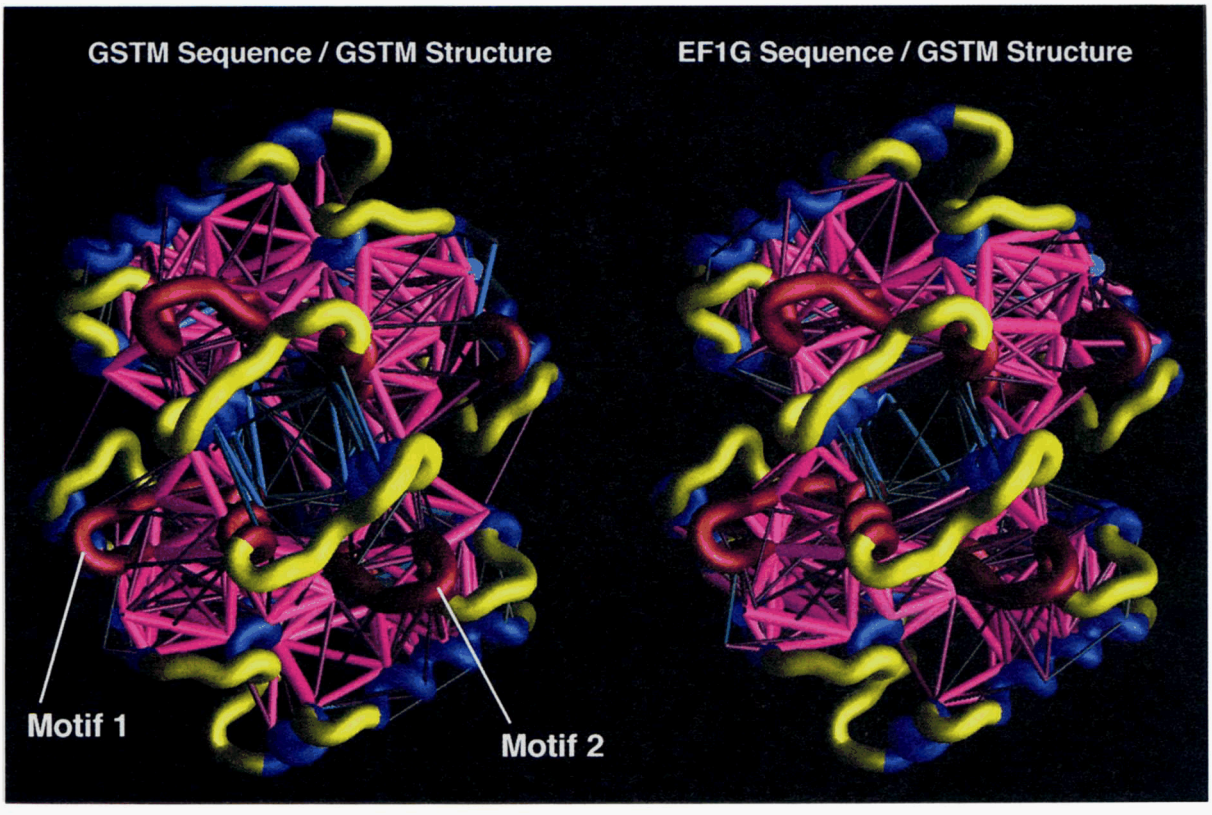

Fig. 4. "Energy scaffolds" for the native structure of GSTM and a model of the N-terminal domain of EF1γ. Sequences are for GTB1_RAT and trypanosomal EF1γ, respectively. Each model shows the dimer of the respective protein/domain. Colored rods indicate the strength of pairwise residue interactions via their thickness, with magenta rods indicating favorable interactions, and cyan rods indicating unfavorable interactions. The locations of the elements comprising the GST core substructure are indicated by blue coloring of the backbone "worm." These correspond to residue positions 1–6, 13–20, 27–32, 43–51, 60–81, 93–115, 120–131, 133–140, 146–168, 171–175, 178–189, and 193–198 in the 1GST structure. Elements 5 and 9, colored red, correspond to motifs I and II. The figure was prepared using the GRASP program.

α-helix (Figs. 1, 4). The conserved aspartic acid (D11 in Fig. 3) forms 2 internal hydrogen bonds and appears to be important for stabilization of the loop (Fig. 5). Motif II is located far from the active site of GST (Figs. 4, 5); recent structural studies did not implicate any of the residues in this motif in binding of either of the substrates (Garcia-Saez et al., 1994; Ji et al., 1994). Rather, these experiments have shown that, unlike the GSH-binding site, the electrophile-binding site is formed by variable segments of the GSTs. The function of motif II therefore remains uncertain. The fact that it contains a conserved charged residue (aspartic acid), which is buried in the protein globule and forms internal hydrogen bonds, may suggest that this motif is a key structural element in the conserved core of GST. On the other hand, there is a contact between the α-helices in motifs I and II (Fig. 5), suggesting that motif II still may affect substrate binding in a more direct fashion.

Possible evolutionary relationships among GST-related proteins

Conventional phylogenetic analysis is difficult for distant sequences like those of the entire set of GST-related proteins, as the number of unequivocally aligned amino acid residues is too small for deriving a reliable tree topology. Accordingly, we used an alternative approach, namely grouping by BLAST scores. The clustering obtained by this method showed a clear distinction between GSTT-related proteins and GSTs α, π, and μ (Fig. 6). Strikingly, the EF1γ sequences were within the GSTT division, which also included a variety of other functionally diverse GST derivatives (Fig. 6). The grouping shown in Figure 6 is essentially phenetic, and evolutionary implications require much caution. Nevertheless, these results clearly confirm that the N-terminal domain of EF1γ belongs to the GSTT-related proteins.

Discussion

We found that the γ subunit of eukaryotic translation elongation factor 1 contains a GST-related domain. Two pronounced sequence motifs and a third, very short motif around the essential tyrosine are conserved in all GST-related proteins. Taken together, the results of motif analysis and 3-dimensional modeling strongly suggest that EF1γ contains a catalytically active GST domain. The related domain in ValRS, in which motif I is significantly changed, whereas the N-terminal conserved tyrosine is lacking, may have lost the GST activity.

Extensive site-directed mutagenesis studies have revealed the role of the N-terminal tyrosine in GSH binding and stabilization of the thiolate anion (Stenberg et al., 1991; Manoharan

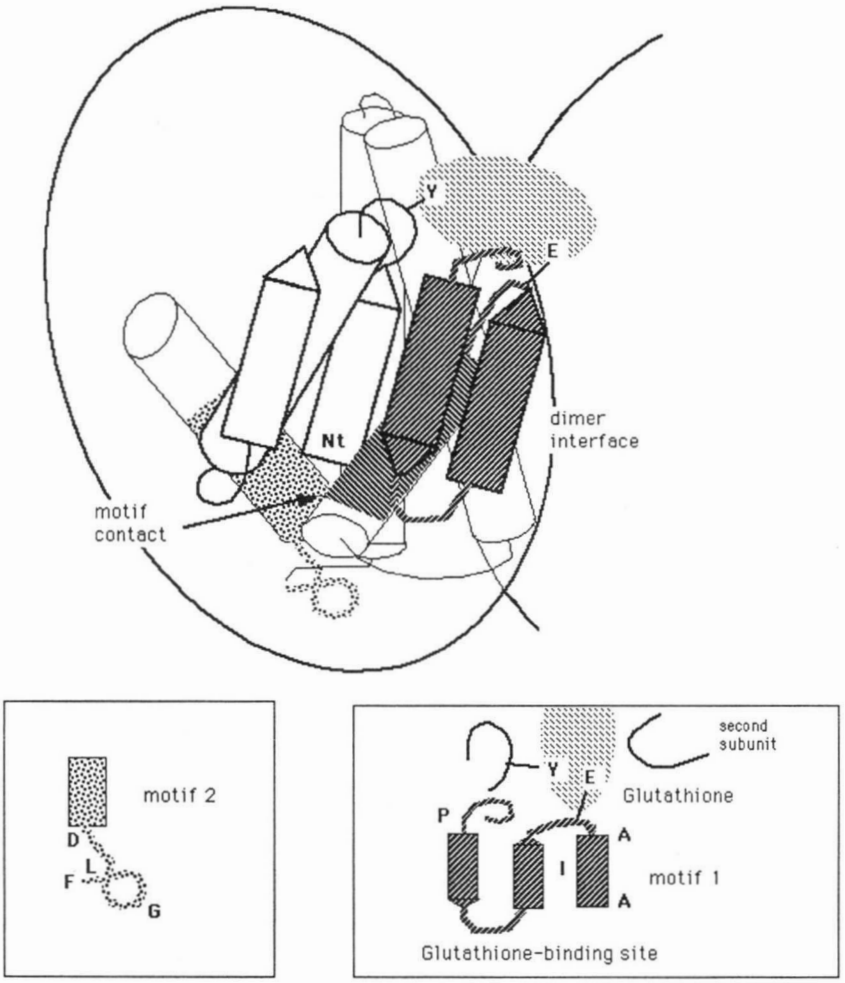

Fig. 5. Crude model of the 3-dimensional structure of the GST domain core in EF1γ. The main scheme shows the folding in the conserved structural core, with α-helices designated by cylinders and β-strands designated by arrowheaded rectangles. The shaded oval shows the bound glutathione. Nt, N-terminus. The boxes show the predicted topology of motif I and II, with some of the conserved residues discussed in the text indicated (see also Figs. 1, 3).

et al., 1992; Wang et al., 1992). Replacement of amino acid residues in motif I, including the conserved glutamine, has resulted in loss of enzymatic activity and impairment of GSH binding (Kong et al., 1992; Manoharan et al., 1992). The results of mutagenesis of the conserved aspartic acid in motif II have been somewhat contradictory because one study has reported GST inactivation (Wang et al., 1992), whereas others have found only decrease in the thermal stability of the enzyme (Kong et al., 1993).

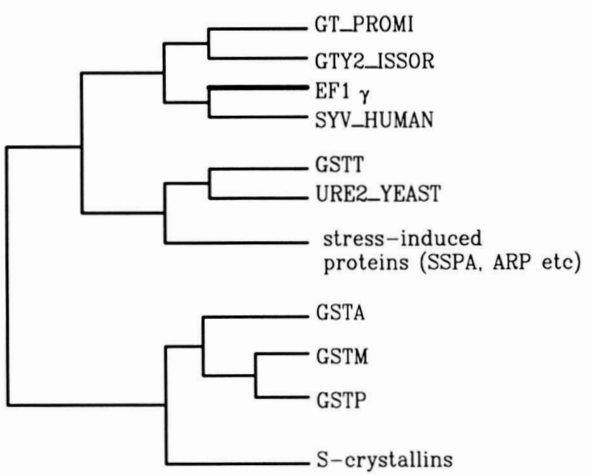

Fig. 6. Cluster dendrogram for GST-related proteins. The dendrogram was constructed as described in the Materials and methods and redrawn to show only the major divisions. The branch lengths are arbitrary.

What may be the role of the GST domain in the components of the translation machinery? It has been shown that the interaction between EF1γ and EF1β involves the N-terminal portion of each of these proteins, which according to our findings, is related to GST (Van Damme et al., 1991). Human ValRS is present in the cell almost entirely as a complex with EF1, which accounts for about 25–50% of the total activity of this factor (Motorin et al., 1987, 1991; Bec et al., 1989; Venema et al., 1991). Recent experiments have shown that the N-terminal domain of ValRS, which may be an inactive homologue of the GST domain of EF1γ (see above), is responsible for the formation of the complex with EF1 (Beck et al., 1994). Conversely, although it is the δ subunit of EF1 that directly interacts with ValRS, the binary complex of EF1γ and EF1β is required for the formation of a complex of defined quaternary structure rather than high molecular weight aggregates that are formed in the presence of EF1δ alone (Bec et al., 1994). Along a different line of investigation, data have been presented indicating that tightly bound GSH or disulfide GSH (GSSG) may be required for the activity of yeast ValRS in a high molecular weight complex (Black,

1986, 1993). All these observations are compatible with a role of the GST domain of EF1γ in the regulation of specific, multisubunit protein complex assembly.

The equilibrium between GSH and GSSG is an important regulator of the state of thiol groups in proteins, according to the reaction Cys-S-S-Cys + 2GSH = 2Cys-SH + GSSG (Gilbert, 1990; Hwang et al., 1992; Zapun et al., 1993, and references therein). In turn, the balance between free thiol groups of cysteines and disulfide bonds is important for protein folding. Generally, due to the high reducing potential in the cytoplasm, most cytoplasmic proteins do not contain disulfide bonds (Gilbert, 1990; Branden and Tooze, 1991). However, prevention of disulfide bond formation in these proteins appears to be an active process involving at least 1 enzyme, namely thioredoxin reductase (Derman et al., 1993). GST activity also may be involved in this process. More specifically, it is possible to speculate that disulfide bond exchange between endogenous GSSG and protein catalyzed by the GST domain of EF1γ may mediate the assembly of EF1 and/or the interconversion between different physical and functional forms of high molecular weight complexes of EF1 and aaRSs.

Utilization of compartmentalized GSH and GSSG for interconversion between free SH groups and disulfide bonds, facilitated by GST domains, may be a novel regulatory mechanism of protein folding and assembly of multisubunit complexes. It seems likely that this mechanism is not restricted to EF1 and that built-in GST domains eventually will be discovered in other proteins. In yet other macromolecular ensembles, GST-related domains may perform this regulatory function through tight, but not covalent association with other subunits. The finding that the EF1γ homologue in *Emericella* contains only the GST domain (see above) is a striking illustration of such a possibility. It remains to be elucidated whether or not this organism encodes a separate protein equivalent to the C-terminal domain of EF1γ. The association of the *E. coli* GST-related protein SSPA with the RNA polymerase holoenzyme (Ishihama & Saitoh, 1979) may be another example of a noncovalent, GST-containing complex.

In a very general sense, it may be conjectured that GST domains facilitate protein folding and assembly in a chaperone-like manner (Gething & Sambrook, 1992). A recent dramatic example of the possible chaperone-like activity in a GSTT-related protein includes yeast protein URE2 that appears to undergo an autocatalytic, inheritable conformational change, in analogy to mammalian prion proteins (Weissmann, 1994; Wickner, 1994).

Materials and methods

Sequences

Amino acid and nucleotide sequences were from the SwissProt, PIR, and GenBank databases that are combined in the nonredundant sequence database (NR) at the National Center for Biotechnology Information (NIH). Protein X-ray structures were from the Brookhaven Protein Data Bank (Bernstein et al., 1977).

Computer-assisted sequence analysis

Amino acid sequences were compared with NR using programs based on the BLAST algorithm (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 6 reading frames (Altschul et al., 1994). Compositionally biased protein sequence segments that tend to produce artifactual high scores in database searches were masked in all query sequences using the SEG program (Wootton & Federhen, 1993; Altschul et al., 1994).

Database search for conserved segments similar to multiple alignment blocks was performed using a recently developed iterative procedure, called MoST (Motif Search Tool), a full description of which is presented elsewhere (Tatusov et al., 1994). Briefly, the multiple alignment blocks are initially constructed by parsing consistent segments from the ungapped pairwise alignments produced by a BLAST search. These blocks are converted into position-dependent weight matrices using a method that combines the observed amino acid residue frequencies for each column with a priori knowledge of amino acid relationships (Brown et al., 1993). Using these weight matrices, scores are computed for all segments of the corresponding length in the amino acid sequence database, and the observed distribution of scores is compared with the theoretical distribution. The ratio of the expected to the observed number of sequence segments with a given score is used as the cutoff in database searches.

Multiple alignments were generated using the programs MACAW (Schuler et al., 1991) and OPTAL (Gorbalenya et al., 1989).

For classifying distantly related protein sequences, a clustering procedure was developed that used the alignment scores produced by BLAST searches as the measure of sequence similarity. A cluster was defined as a group of sequences that formed a connected graph component, with each edge corresponding to a BLAST alignment with a score higher than the chosen cutoff. Using progressively lower cutoff scores, a cluster dendrogram is constructed for the given set of sequences. This procedure was implemented in a program called CLUS.

Protein secondary structure was predicted using the PHD program that implements a recently developed neural network method (Rost & Sander, 1993). Structure modeling based on the known structure of related proteins was performed using 2 methods: (1) "threading" of sequences through known structures (Bryant & Lawrence, 1993)—energy scaffold diagrams based on the threading results were prepared using the GRASP program (Nicholls et al., 1991); (2) homology-based modeling according to the procedure implemented in the WHATIF program (Vriend, 1990) and quality control for the resulting model using directional atomic contact analysis (Vriend & Sander, 1993).

The programs CAP and MoST used in this study for motif search (running under the Unix operating system) are available upon request from tatusov@ncbi.nlm.nih.gov, and the complete model of the N-terminal domain of EF1γ is available via ftp from the fileserver EMBL-heidelberg-de (file EF1G_on_1GSR.model in the directory /pub/databases/protein_extras/models).

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.

Bec G, Kerjan P, Waller JP. 1994. Reconstitution in vitro of the valyl-tRNA synthetase-elongation factor (EF) 1βγδ complex. Essential roles of the

NH2-terminal extension of valyl-tRNA synthetase and of the EF-1δ subunit in complex formation. *J Biol Chem 269*:2086–2092.
Bec G, Kerjan P, Zha XD, Waller JP. 1989. Valyl-tRNA synthetase from rabbit liver. I. Purification as a heterotypic complex in association with elongation factor 1. *J Biol Chem 264*:21131–21137.
Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structure. *J Mol Biol 112*:535–542.
Black S. 1986. Reversible interconversion of two forms of a valyl-tRNA synthetase-containing protein complex. *Science 234*:1111–1114.
Black S. 1993. A provisional mechanism for regulating the aminoacyl-tRNA synthetases. *Biochem Biophys Res Commun 191*:95–102.
Bork P. 1992. Mobile modules and motifs. *Curr Opin Struct Biol 2*:413–421.
Branden C, Tooze J. 1991. *Introduction to protein structure*. New York/London: Garland.
Brown M, Hughey R, Krogh A, Mian IS, Sjölander K, Haussler D. 1993. Using Dirichlet mixture priors to derive hidden Markov models for protein families. In: Hunter L, Searls D, Shavlik J, eds. *Proc. First International Conference on Intelligent Systems for Molecular Biology*. Menlo Park, California: AAAI Press. pp 47–55.
Bryant SH, Lawrence CE. 1993. An empirical energy function for threading protein sequence through the folding motif. *Proteins Struct Funct Genet 16*:92–112.
Coschigano PW, Magasanik B. 1991. The URE2 gene product of *Saccharomyces cerevisiae* plays an important role in the cellular response to the nitrogen source and has homology to glutathione S-transferase. *Mol Cell Biol 11*:822–832.
Czarnecka E, Nagao RT, Key JL, Gupley WB. 1988. Characterization of Gmhsp-26-A, a stress gene encoding a divergent heat shock protein of soybean: Heavy-metal-induced inhibition of intron processing. *Mol Cell Biol 8*:1113–1122.
Derman AI, Prinz WA, Belin D, Beckwith J. 1993. Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science 262*:1744–1747.
Dirr H, Reinemer P, Huber R. 1994. X-ray crystal structures of cytosolic glutathione *S*-transferases. Implications for protein structure, substrate recognition and catalytic function. *Eur J Biochem 220*:645–661.
Dominov JA, Stenzler L, Lee S, Schwartz JJ, Leisner S, Howell SH. 1992. Cytokinins and auxins control the expression of a gene in *Nicotiana plumbaginifolia* cells by feedback regulation. *Plant Cell 4*:451–461.
Doolittle RF. 1988. Lens proteins. More molecular opportunities. *Nature (Lond) 336*:18.
Doolittle RF. 1992. Stein and Moore Award address. Reconstructing history with amino acid sequences. *Protein Sci 1*:191–200.
Doolittle RF, Bork P. 1993. Evolutionary mobile modules in proteins. *Sci Am 269*:50–56.
Epp O, Ladenstein R, Wendel A. 1983. The refined structure of the selenoenzyme glutathione peroxidase at 0.2 nm resolution. *Eur J Biochem 133*:51–69.
Fahey RC, Sundquist AR. 1991. Evolution of glutathione metabolism. *Adv Enzymol Relat Areas Mol Biol 64*:1–53.
Fett R, Knippers R. 1991. The primary structure of human glutaminyl-tRNA synthetase. *J Biol Chem 266*:1448–1455.
Garcia-Saez I, Parraga A, Phillips MF, Mantle TJ, Coll M. 1994. Molecular structure at 1.8 Å of mouse liver class pi glutathione *S*-transferase complexed with *S*-(*p*-nitrobenzyl)glutathione and other inhibitors. *J Mol Biol 237*:298–314.
Gething MJ, Sambrook J. 1992. Protein folding in the cell. *Nature 355*:33–45.
Gilbert H. 1990. Molecular and cellular aspects of thiol-disulfide exchange. *Adv Enzymol Relat Areas Mol Biol 63*:69–172.
Gorbalenya AE, Blinov VM, Donchenko AP, Koonin EV. 1989. An NTP-binding motif is the most conserved sequence in a highly diverged group of proteins involved in positive strand RNA viral replication. *J Mol Evol 28*:256–268.
Gorbalenya AE, Koonin EV. 1990. Superfamily of UvrA-related NTP-binding proteins. Implications for rational classification of recombination/repair systems. *J Mol Biol 213*:583–591.
Hanks SK, Quinn AM, Hunter T. 1988. The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. *Science 241*:42–52.
Hsieh SL, Campbell RD. 1991. Evidence that gene G7a in human major histocompatibility complex encodes valyl-tRNA synthetase. *Biochem J 278*:809–816.
Hwang C, Sinskey AJ, Lodish HF. 1992. Oxidized redox state of glutathione in the endoplasmic reticulum. *Science 257*:1496–1501.
Ishihama A, Saitoh T. 1979. Subunits of RNA polymerase in function and structure. IX. Regulation of RNA polymerase activity by stringent starvation protein (SST). *J Mol Biol 129*:517–530.
Janes W, Schulz G. 1990. The binding of the retro-analogue of glutathione disulfide to glutathione reductase. *J Biol Chem 265*:10443–10445.
Ji X, Johnson WW, Sesay MA, Dickert L, Prasad SM, Ammon HL, Armstrong RN, Gilliland GL. 1994. Structure and function of the xenobiotic substrate binding site of a glutathione *S*-transferase as revealed by X-ray crystallographic analysis of product complexes with the diastereomers of 9-(*S*-glutathionyl)-10-hydroxy-9,10-dihydrophenanthrene. *Biochemistry 33*:1043–1052.
Ji X, Zhang P, Armstrong RN, Gilliland GL. 1992. The three-dimensional structure of a glutathione *S*-transferase from the Mu gene class. Structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2 Å resolution. *Biochemistry 31*:10169–10184.
Kong KH, Inoue H, Takahashi K. 1992. Site-directed mutagenesis of amino acid residues involved in the glutathione binding of human glutathione *S*-transferase P1-1. *J Biochem 112*:725–728.
Kong KH, Inoue H, Takahashi K. 1993. Site-directed mutagenesis study on the roles of evolutionarily conserved aspartic acid residues in human glutathione *S*-transferase P1-1. *Protein Eng 6*:93–99.
La Roche SD, Leisinger T. 1990. Sequence analysis and expression of the bacterial dichloromethane dehalogenase structural gene, a member of the glutathione *S*-transferase supergene family. *J Bacteriol 172*:164–171.
Liu S, Zhang P, Ji X, Johnson WW, Gilliland GL, Armstrong RN. 1992. Contribution of tyrosine 6 to the catalytic mechanism of isoenzyme 3-3 of glutathione *S*-transferase. *J Biol Chem 267*:4296–4299.
Manoharan TH, Gulick AM, Reinemer P, Dirr HW, Huber R, Fahl WE. 1992. Mutational substitution of residues implicated by crystal structure in binding the substrate glutathione to human glutathione *S*-transferase pi. *J Mol Biol 226*:319–322.
Masai E, Katayama Y, Kubota S, Kawai S, Yamasaki S, Morohoshi N. 1993. A bacterial enzyme degrading the model lignin is a member of the glutathione *S*-transferase superfamily. *FEBS Lett 323*:135–140.
Milner-White EJ, Coggins JR, Anton IA. 1991. Evidence for an ancestral core structure in nucleotide-binding proteins with the A type motif. *J Mol Biol 221*:751–754.
Momoi H, Yamada H, Ueguchi C, Mizuno T. 1993. Sequence of a fission yeast gene encoding a protein with extensive homology to eukaryotic elongation factor-1 γ. *Gene 134*:119–122.
Motorin YA, Wolfson AD, Orlovsky AF, Gladilin KL. 1987. Purification of valyl-tRNA synthetase high-molecular-mass complex from rabbit liver. *FEBS Lett 220*:363–365.
Motorin YA, Wolfson AD, Rohr D, Orlovsky AF, Gladilin KL. 1991. Purification and properties of a high-molecular-mass complex between ValtRNA synthetase and the heavy form of elongation factor 1 from mammalian cells. *Eur J Biochem 201*:325–331.
Neurath H. 1986. The versatility of proteolytic enzymes. *J Cell Biochem 32*:35–49.
Nicholls A, Sharp KA, Honig B. 1991. Protein folding and association: Insights from the thermodynamic properties of hydrocarbons. *Proteins Struct Funct Genet 11*:281–296.
Orser CS, Dutton JD, Lange C, Jablonski P, Xun L, Hargis M. 1993. Characterization of a *Flavobacterium* glutathione *S*-transferase gene involved in reductive dechlorination. *J Bacteriol 175*:2640–2644.
Pemble SE, Taylor JB. 1992. An evolutionary perspective on glutathione transferases inferred from class-theta glutathione transferase cDNA sequences. *Biochem J 287*:957–963.
Pickett CB, Lu AYH. 1989. Glutathione *S*-transferases: Gene structure, regulation and biological function. *Annu Rev Biochem 58*:743–764.
Reinemer P, Dirr HW, Ladenstein R, Huber R, LoBello ML, Federici G, Parker MW. 1992. Three-dimensional structure of class pi glutathione *S*-transferase from human placenta in complex with *S*-hexylglutathione at 2.8 Å resolution. *J Mol Biol 227*:214–226.
Reinemer P, Dirr HW, Ladenstein R, Schaffer J, Gallay O, Huber R. 1991. The three-dimensional structure of class pi glutathione *S*-transferase in complex with glutathione sulfonate at 2.3 Å resolution. *EMBO J 10*:1997–2005.
Riis B, Rattan SIS, Clark BFC, Merrick WC. 1990. Eukaryotic protein elongation factors. *Trends Biochem Sci 15*:420–424.
Rost B, Sander C. 1993. Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol 232*:584–599.
Rushmore TM, Pickett CB. 1993. Glutathione *S*-transferases, structure, regulation, and therapeutic implications. *J Biol Chem 268*:11475–11478.
Schuler GD, Altschul SF, Lipman DJ. 1991. A workbench for multiple alignment construction and analysis. *Proteins Struct Funct Genet 9*:180–190
Sinning I, Kleywegt GJ, Cowan SW, Reinemer P, Dirr HW, Huber R, Gilliland G, Armstrong RN, Ji X, Board PG, Olin B, Mannervik B, Jones TA. 1993. Structure determination and refinement of human alpha class glutathione *S*-transferase A1-1, and a comparison with the Mu and Pi class enzymes. *J Mol Biol 232*:192–212.
Stenberg G, Board PG, Mannervik B. 1991. Mutation of an evolutionarily

conserved tyrosine residue in the active site of a human class alpha glutathione transferase. *FEBS Lett 293*:153–155.

Takahashi Y, Kuroda H, Tanaka T, Machida Y, Takebe I, Nagata T. 1989. Isolation of an auxin-regulated gene cDNA expressed during the transition from G0 to S phase in tobacco mesophyll protoplasts. *Proc Natl Acad Sci USA 86*:9279–9283.

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*. Forthcoming.

Tomarev SI, Zinovieva RD. 1988. Squid major lens polypeptides are homologous to glutathione *S*-transferase subunits. *Nature (Lond) 336*:86–88.

Tomarev SI, Zinovieva RD, Piatigorsky J. 1992. Characterization of squid crystallin genes. Comparison with mammalian glutathione *S*-transferase genes. *J Biol Chem 267*:8604–8612.

Toung YPS, Tu CPD. 1992. *Drosophila* glutathione *S*-transferases have sequence homology to the stringent starvation protein of *Escherichia coli*. *Biochem Biophys Res Commun 182*:355–360.

Van Damme H, Amons R, Janssen G, Möller W. 1991. Mapping the functional domains of the eukaryotic elongation factor 1$\beta\gamma$. *Eur J Biochem 197*:505–511.

Van Damme H, Amons R, Karssies R, Timmers CJ, Janssen GM, Möller W. 1990. Elongation factor 1β of *Artemia*: Localization of functional sites and homology to elongation factor 1δ. *Biochim Biophys Acta 1050*:241–247.

Venema RC, Peters HI, Traugh JA. 1991. Phosphorylation of elongation factor 1 (EF-1) and valyl-tRNA synthetase by protein kinase C and stimulation of EF-1 activity. *J Biol Chem 266*:11993–11998.

Vriend G. 1990. WHATIF: A molecular modelling and drug design program. *J Mol Graphics 8*:52–55.

Vriend G, Sander C. 1993. Quality control of protein models: Directional atomic contact studies. *J Appl Crystallogr 26*:47–60.

Wang RW, Newton DJ, Huskey SE, McKever BM, Pickett CB, Lu AY. 1992. Site-directed mutagenesis of glutathione *S*-transferase YaYa. Important roles of tyrosine 9 and aspartic acid 101 in catalysis. *J Biol Chem 267*: 19866–19871.

Weissmann C. 1994. The prion connection: Now in yeast? *Science 264*: 528–530.

Wickner RB. 1994. [URE3] as an altered URE2 protein: Evidence for a prion analog in *Saccharomyces cerevisiae*. *Science 264*:566–569.

Wilce MC, Parker MW. 1994. Structure and function of glutathione *S*-transferases. *Biochim Biophys Acta 1205*:1–18.

Wootton JC, Federhen S. 1993. Statistics of local complexity in amino acid sequences and sequence databases. *Comput Chem 17*:149–163.

Zapun A, Bardwell JCA, Creighton TE. 1993. The reactive and destabilizing disulfide bond of DsbA, a protein required for protein disulfide bond formation in vivo. *Biochemistry 32*:5083–5092.

Zhao S, Sandt CH, Feulner G, Vlazny DA, Gray JA, Hill CW. 1993. Rhs elements of *Escherichia coli* K-12: Complex composite of shared and unique components that have different evolutionary histories. *J Bacteriol 175*:2789–2808.