Protein Science (1999), *8*:1714–1719. Cambridge University Press. Printed in the USA.

FOR THE RECORD

A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases

KIRA S. MAKAROVA,[1,2,4] L. ARAVIND,[2,3] AND EUGENE V. KOONIN[2]

[1]Department of Pathology, F.E. Hebert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814-4799
[2]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894
[3]Department of Biology, Texas A&M University, College Station, Texas 77843

(RECEIVED February 3, 1999; ACCEPTED April 9, 1999)

Abstract: Computer analysis using profiles generated by the PSI-BLAST program identified a superfamily of proteins homologous to eukaryotic transglutaminases. The members of the new protein superfamily are found in all archaea, show a sporadic distribution among bacteria, and were detected also in eukaryotes, such as two yeast species and the nematode *Caenorhabditis elegans*. Sequence conservation in this superfamily primarily involves three motifs that center around conserved cysteine, histidine, and aspartate residues that form the catalytic triad in the structurally characterized transglutaminase, the human blood clotting factor XIIIa′. On the basis of the experimentally demonstrated activity of the *Methanobacterium* phage pseudomurein endoisopeptidase, it is proposed that many, if not all, microbial homologs of the transglutaminases are proteases and that the eukaryotic transglutaminases have evolved from an ancestral protease.

Keywords: catalytic triad; evolution of enzymes; iterative database search; thiol protease; transglutaminase

Transglutaminases catalyze the calcium-dependent acyl-transfer reaction between a γ-carboxyamide group of glutamine and ϵ-amino group of lysine or other primary amines, which results in the formation of γ-glutamyl-ϵ-lysine peptide chains bridges (Lorand & Conrad, 1984). These enzymes thus establish either intramolecular or intermolecular cross-links in proteins. The animal transglutaminases possess a catalytic triad of three amino acids, namely a cysteine, a histidine, and an aspartate (asparagine), and the reaction proceeds via an intermediate linked to the nucleophilic cysteine of the enzyme (Hettasch & Greenberg, 1994; Micanovic et al., 1994). This reaction is the reversion of the proteolysis reaction catalyzed by the thiol proteases that possess the same catalytic triad. The crystal structure of a transglutaminase precursor, the human blood clotting factor XIII, has been solved (Yee et al., 1994). Not surprisingly, given the similarities in the catalytic triad and the reaction mechanism (Pedersen et al., 1994), it turned out that the transglutaminases share the core structural fold with the papain-like thiol proteases. Accordingly, transglutaminases and papain-like proteases have been classified within the same superfamily in the Structural Classification of Proteins (SCOP) database (Hubbard et al., 1999).

The involvement of a transglutaminase, namely the catalytic domain of factor XIII, in blood clotting in vertebrates has triggered a number of functional studies of these enzymes. To date, seven functional types of transglutaminases that differ in terms of their specificity toward target proteins have been characterized in humans (Aeschlimann et al., 1995; Kim et al., 1995; Steinert & Marekov, 1995). They are involved in a variety of protein modifications associated with animal development and pathology (Muszbek et al., 1996). One of them, namely band 4.2 protein, has lost its enzymatic activity and plays a structural role as a cytoskeleton component (Cohen et al., 1993).

Genes for members of the transglutaminase family typified by factor XIII have been characterized in a wide range of vertebrates (Weraarchakul-Boonmark et al., 1992; Nakaoka et al., 1994) and invertebrates (Tokunaga et al., 1993; Cariello et al., 1997). Transglutaminase activity has been detected also in plants but the respective genes so far have not been cloned (Serafini-Fracassini et al., 1995). Cloning of transglutaminases from two bacteria has been reported but the sequences of these proteins do not resemble any known families of enzymes or each other (Washizu et al., 1994). Recently, it has been shown that the Cytotoxic Necrotizing Factor 1 (CNF1) from *Escherichia coli* and the homologous dermonecrotic toxin from *Bordetella pertussis*, which act by deamidation of a specific glutamine residue in animal Rho GTPases, also possess transglutaminase activity (Horiguchi et al., 1997; Schmidt et al., 1998). Again, however, although the catalytic cysteine and histidine residues of CNF1 have been identified (Schmidt et al., 1998), no similarity to animal transglutaminases or papain-like proteases is detectable in the sequences of these proteins.

Reprint requests to: Eugene V. Koonin, National Center for Biotechnology Information, National Library of Medicine, NIH, Bethesda, Maryland 20894; e-mail: koonin@ncbi.nih.gov.
[4]Permanent address: Institute of Cytology and Genetics, Russian Academy of Sciences, Novosibirsk 630090, Russia.

In the course of our comparative analysis of completely sequenced genomes, we detected proteins homologous to the classical animal transglutaminases in a variety of archael and bacterial genomes as well as in some eukaryotes, such as the yeasts *Saccharomyces cerevisae*, *Schizosaccharomyces pombe,* in which they have not been previously identified.

Here, we describe this new protein superfamily typified by animal transglutaminases. Most of these proteins have not been functionally characterized. Only the protein from an archaeal phage has been studied and shown to possess protease activity (Pfister et al., 1998), which might indicate that many of the prokaryotic transglutaminase homologs are in fact proteases. The discovery of this superfamily suggests an evolutionary scenario for the origin of eukaryotic transglutaminases from ancient proteases.

The transglutaminase-like superfamily of enzymes—Sequence and structure conservation: During our comparative analysis of the archaeal genomes (Makarova et al., 1999), we identified a family of large multidomain archaeal proteins that encompassed predicted signal peptides and a globular domain (as predicted using the SEG program (Wootton, 1994; Wootton & Federhen, 1996)). The latter domain contains conserved cysteine, histidine, and aspartate residues, which is reminiscent of the catalytic triad of a variety of thiol hydrolases. Further iterative searches of the nonredundant database (NR) using the PSI-BLAST program (Altschul et al., 1997) with a number of different query sequences not only detected homologous protein sequences in bacterial genomes but also demonstrated a statistically significant similarity to animal transglutaminases. For example, a search initiated with the *Sulfolobus solfataricus* protein c02013 (GenBank gi:1707725) detects the human transglutaminase 4(gi: 1082749) with an e-value of 10^{-4} in the 3rd iteration and factor XIII with a similar e-value in the 4th iteration. Conversely, searches with the human transglutaminases as queries detected bacterial and archaeal proteins from the same set with e-values below 0.001 in the 2nd and subsequent iterations. All sequences of tranglutaminase homologs detectable in the current nonredundant database were identified via transitive PSI-BLAST searches, and a representative set of them was used to generate a multiple alignment with the ALITRE program (Seledtsov et al., 1995). The alignment was then manually adjusted on the basis of alignments generated by PSI-BLAST and the structural elements from the crystal structure of factor XIII. Figure 1 shows the multiple alignment of the transglutaminase-like superfamily with the secondary structure elements assigned using the factor XIII structure.

The sequence conservation in the superfamily of transglutaminase-like enzymes clearly centers around the three (predicted) active residues of the catalytic triad (Fig. 1). Motif I contains the catalytic cysteine and encompasses the two strands and an α-helix whereas motifs II and III include the two strands associated with the active histidine and aspartate, respectively. Some of the characteristic features of the animal transglutaminases, namely the tiny residue located two positions upstream of the catalytic cysteine, the aromatic residues two positions downstream from the catalytic histidine and the aromatic residue flanking the catalytic aspartate from the N-terminal side, are well conserved in the microbial proteins. The main difference between the archaeal and bacterial transglutaminase homologs and classical animal enzymes is the variability of the two insert regions between the conserved motifs (Fig. 1). Iterative database searches fail to detect any similarity between the proteins of the transglutaminase-like superfamily and thiol proteases. When fragments of papain-like protease sequences surrounding the catalytic residues are manually fit into the multiple alignment, some local conservation can be seen, however, which is compatible with the presence of a common structural core (Fig. 1).

Using the multiple alignment shown in Figure 1, the structure of factor XIII (Protein Data Bank (PDB) code: 1fie) as a template and the ProModII program (Peitsch, 1996), we generated a homology model for the microbial transglutaminase-like proteins (Fig. 2). This model shows that the insert I in the animal transglutaminases (between Motif I and Motif II) is a large extended structure, which interacts with other β-strands that are conserved in the animal transglutaminases but not in the microbial homologs. The second insert also appears to be positioned for an interaction with a specific C-terminal domain present in the eukaryotic proteins. By contrast, most bacterial and archaeal proteins as well as the newly detected members of this superfamily from yeast and the nematode have short inserts, which is consistent with the lack of counterparts to these additional conserved elements of the transglutaminases (Fig. 2).

Phyletic distribution, family classification, and domain organization of the transglutaminase-like superfamily: All the completely sequenced euryarchaeal genomes encode at least one protein of the transglutaminase-like superfamily (Fig. 1), and one member was already detected in the available sequence from the crenarchaeon *Sulfolobus*. Clustering of the transglutaminase-like domains by sequence similarity (see the caption to Fig. 1) produced seven distinct families, along with some unclassified members (Fig. 1). Family 1 represents a set of seven different transglutaminases from human and functionally unusual enzyme from rat. Family 2 includes eukaryotic protein proteins from *Caenorhabditis elegans,* two yeast species, and the cyanobacterium *Synechocystis sp.* (Family 2 in Fig. 1). The eukaryotic members of this family lack the normal catalytic cysteine but have another conserved cysteine downstream (Fig. 1), which potentially could possess a catalytic role; the position of this cysteine in the three-dimensional (3D) model is not inconsistent with this proposal (Fig. 2). The *C. elegans* protein, however, is likely to be inactive given the disruption of the other motifs (Fig. 1). Inactivation of the enzymatic domain, indicated by the elimination of two of the catalytic residues, is also seen in proteins from the cyanobacterium *Synechocystis sp*. and the archaeon *Archaeoglobus fulgidus* that belong to Family 7 of microbial tranglutaminase homologs and also in the human protein band 4.2 (Fig. 1). Of further interest is the replacement of the catalytic cysteine by a serine in a transglutaminase homolog from *Aquifex* (Fig. 1). It seems likely, given the conservation of the other two motifs, that this serine functions as the catalytic nucleophile. Substitution of serine for the active cysteine has been reported for a family of highly conserved papain-like protease homologs from the malaria plasmodium (Higgins et al., 1989; Gardner et al., 1998).

The remaining five families consist primarily of archaeal and bacterial proteins. Examination of the taxonomic distribution in the families of transglutaminase homologs, together with a comparison of the domain architectures of archaeal transglutaminase homologs, indicates that multiple lineage specific duplication events must have been involved in the generation of the gene families seen in *A. fulgidus* and *Methanobacterium thermoautotrophicum* (Families 3 and 6 in Fig. 1). Among the bacteria, a similar, lineage-specific expansion was detected in the Mycobacteria (Family 5

in Fig. 1), while scattered representatives were found in diverse bacterial lineages, such as the Cyanobacteria, *Bacillus subtilis*, *Haemophilus influenzae*, and among the unfinished genomes, in *B. pertussis*, *Pseudomonas aeruginosa*, and *Deinococcus radiodurans*. Families 4, 5, and 7 include representatives from diverse bacterial lineages and archaea, which suggests dissemination by horizontal gene transfers amid prokaryotes. Some of the horizontal transfer events might have involved also eukaryotes as seen in

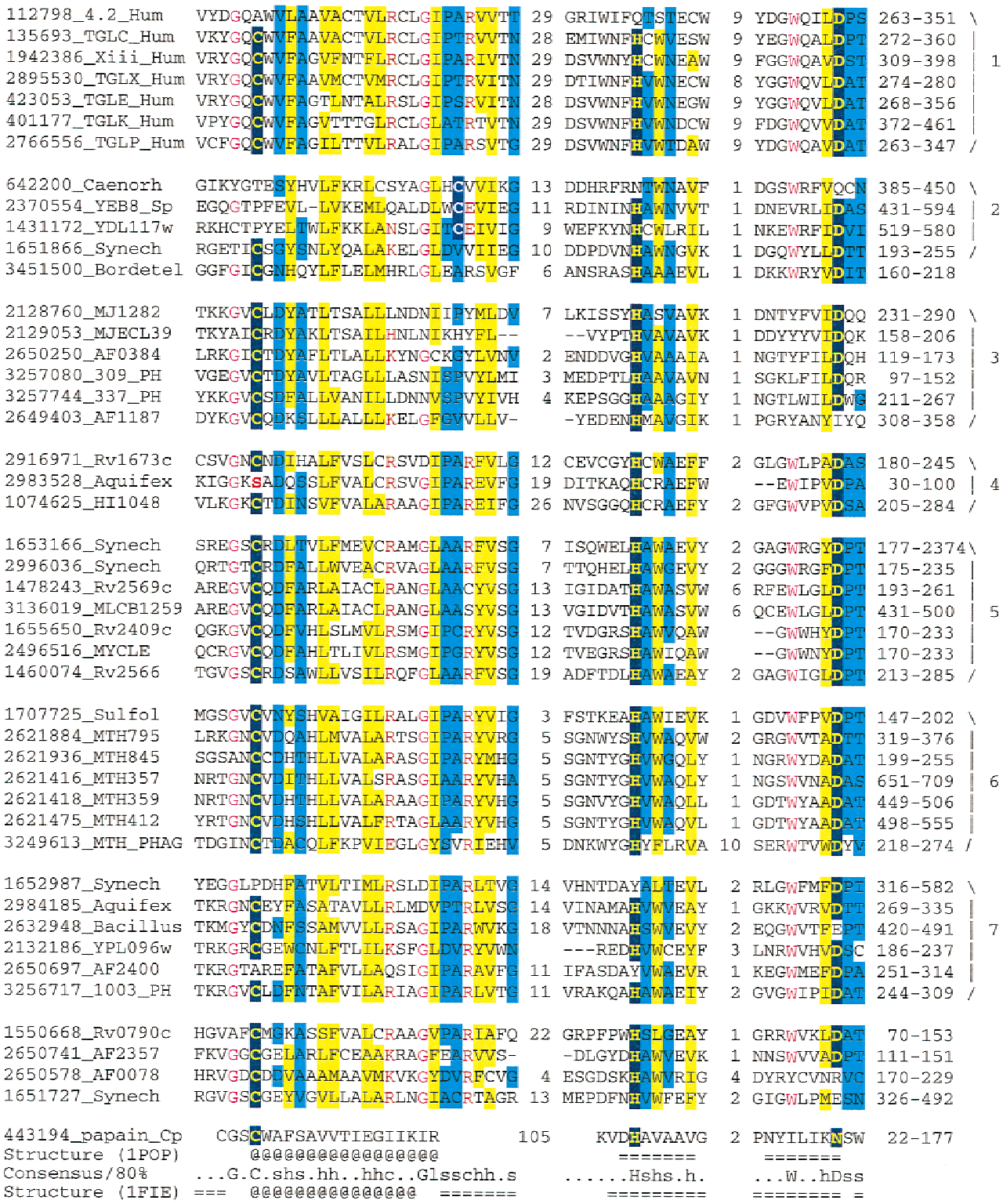

Fig. 1. *See caption on facing page.*

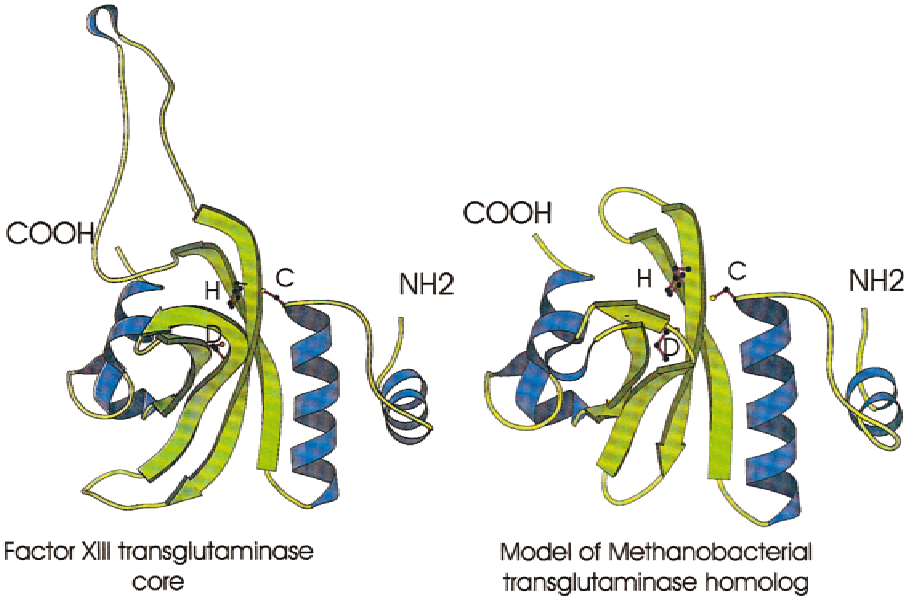

Fig. 2. The transglutaminase structure and a structural model for microbial transglutaminase homologs. The methanobacterial protein MTH795 was aligned with the sequence of the human factor XIII (PDB code 1fie). The alignment was adjusted to minimize energy based on a sudo-Sippl field and submitted to the PROMODII modeling program using the SWISS-PDB viewer. The figure, which was generated using the MOLSCRIPT program (Kraulis, 1991), shows the core structure of factor XIII and the model for the methanobacterial protein. The catalytic triads are shown by ball-and-stick models.

Families 2 and 6 (Fig. 1). Among the eukaryotes, transglutaminase homologs were detected, for the first time, in the yeasts *S. pombe* and *S. cerevisiae* (Fig. 1). Interestingly, in the (nearly) complete proteome of *C. elegans* (The *C. elegans* Sequencing Consortium, 1998), we detected only one, most likely inactive transglutaminase-like enzyme (see above), in a sharp contrast to at least six active and one inactive transglutaminases in humans. Transglutaminase activity has been described in *C. elegans* (Madi et al., 1998), but the above protein cannot account for it; thus either there is a transglutaminase homolog among the products of the still unsequenced *C. elegans* genes or the nematode has a completely different transglutaminase that is unrelated to the protein superfamily described here.

We analyzed the other domains that are detectable in proteins containing the transglutaminase-like core to shed light on their possible functions and interactions (Fig. 3). On the basis of their domain composition, the transglutaminase-like proteins could be divided into three groups; this grouping is likely to have direct

Fig. 1 (*on facing page*)**.** Multiple alignment of the conserved core of the transglutaminase-like protein superfamily. The alignment was constructed using the ALITRE program and adjusted manually on the basis of the PSI-BLAST search results. The transglutaminase-like domains were clustered by sequence similarity on the basis of a detailed examination of the PSI-BLAST search results obtained, for which portions of the respective proteins that include the core domain together with flanking region were used as queries. Family 1 includes seven distinct functional types of mammalian transglutaminases represented here by the respective human proteins; the other seven families include previously undetected transglutaminase homologs. The numbers between aligned blocks indicate the lengths of variable inserts that are not shown; the numbers at the end of each sequence indicate the distances from the protein termini to the proximal and distal aligned blocks. The shading of conserved residues is according to the consensus and includes residues conserved in at least 80% of the aligned sequences. The three residues of the catalytic triad are shown in inverse shading (yellow against a dark blue background); the putative alternative catalytic cysteine in Family 2 is shown in white against a light blue background; the putative catalytic serine in an *Aquifex* protein is shown in red (Family 4). The fragments of the papain sequence surrounding the catalytic residues were incorporated in the alignment manually, on the basis of the published structural comparisons (Pedersen et al., 1994). In the consensus line, **h** indicates hydrophobic residues (A,C,F,L,I,MV,W,Y; yellow background); **s** indicates small residues (A,C,S,T,D, N,V,G,P; blue background); **c** indicates charged residues (R,K,E,D,H; brown coloring); separately conserved residues are colored in magenta. PDB record 1fie. The GeneBank gene identifier and name (following an underline) or an abbreviated species names where a gene name is not available are shown at the beginning of each sequence. Gene names that start with AF are from *A. fulgidus*; MJ, *Methanococcus jannaschii*; MTH, *M. thermoautotrophicum*; PH, *Pyrococcus horikoshii*; Rv, *Mycobacterium tuberculosis*; YDL117w and YPL096w are from the yeast *S. cerevisiae* and YEB8 is from the yeast *S. pombe*. Other species abbreviations: Aquifex, *Aquifex aeolicus*; Bacillus, *B. subtilis*; Caenorh, *C. elegans*; Hum, *Homo sapiens*; Rat, *Rattus norvegicus*; Synech, *Synechocystis sp.* 112798_4.2 is human band 4.2 protein (an inactivated transglutaminase); TGLC-P are various human transglutaminase isoenzymes; 1942386_XIII is human blood-clotting factor XIII; and 3249613_MTH_PHAG is the pseudomurein endoisopeptidase from the *Methanobacterium* bacteriophage psiM2. The secondary structure elements extracted from the crystal structure of the human blood-clotting factor XIII (Yee et al., 1994) and papain (Schroder et al., 1993) are shown underneath the alignment; == indicates β-strands and @@ indicates α-helix.

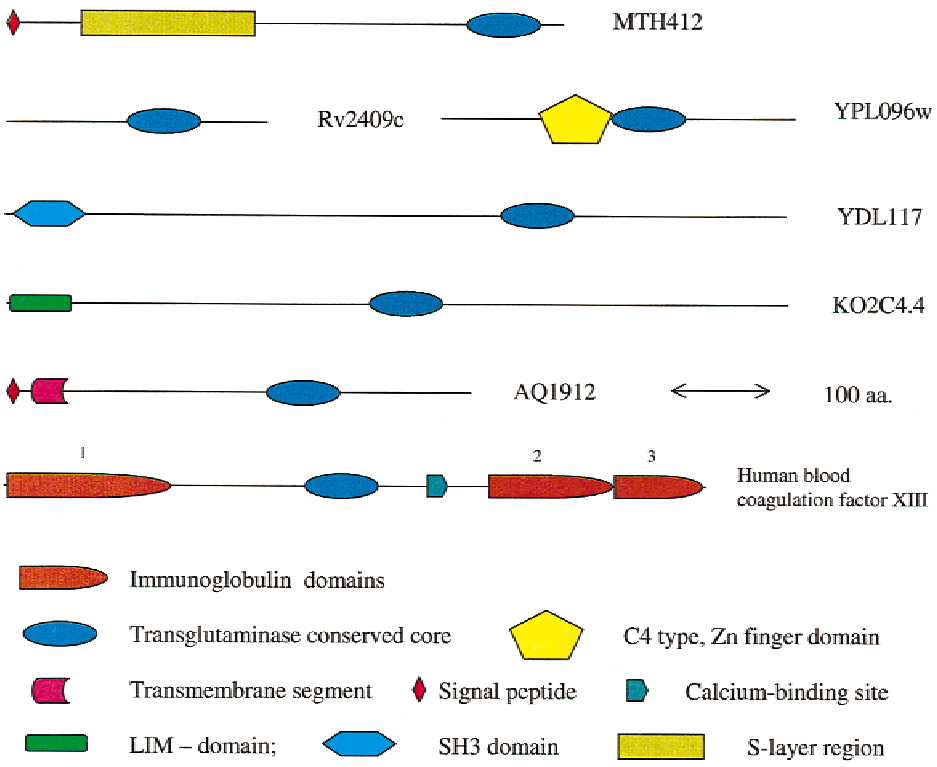

Fig. 3. Distinct domain architectures of proteins containing the transglutaminase-like catalytic core.

functional implications. The first group consists of membrane or secreted proteins as indicated by the presence of a predicted signal peptide and/or membrane-spanning regions. Notably, several transglutaminase-like proteins that are predicted to be secreted are encoded by pathogenic bacteria, such as *Bordetella* and *H. influenzae*. The *Aquifex* transglutaminase-like protein and its relatives (Family 4) contain both signal peptides and predicted membrane-spanning regions (Fig. 3), which is again compatible with action at the cell surface. Some of the archaeal proteins contain a signal peptide and S-layer repeats and are involved in the formation of the surface structures typical of the archaea. Members of the second group do not contain any detectable domains other than the transglutaminase-like one and probably are intracellular enzymes (Family 5). The third group also includes intracellular proteins that contain, in addition to the core transglutaminase domain, different types of protein–protein interaction domains, so they are likely to function as parts of protein complexes. The interaction domains that combine with the transglutaminase core include SH3 domains in yeast proteins, a LIM domain in a *C. elegans* protein, and C4 Zn-finger domains in a yeast protein and an *A. fulgidus* protein (Fig. 3).

Some potential functions for the newly identified transglutaminase homologs: The only prokaryotic transglutaminase homolog with an experimentally demonstrated function is the pseudomurein endoisopeptidase of the *Methanobacterium* phage psiM2—a protease that is involved in the host cell wall lysis (Pfister et al., 1998). It seems likely that many of the prokaryotic members of this superfamily are proteases, although it cannot be ruled out that some of them actually possess a transglutaminase activity. As already mentioned, a specific function for some of the archaeal transglutaminase homologs is suggested by the presence of the secreted S-layer protein repeats in the methanobacterial proteins (Fig. 3). The S-layers are crystalline surface layers with different functions that are present in a variety of bacteria and archaea (Moens & Vanderleyden, 1997; Sleytr & Sara, 1997). The methanobacterial transglutaminase homologs catalyze the necessary cleavage steps during the assembly of these layers, whereas the methanobacterial phage seems to employ the protein scavenged from the host for a similar function. The presence of multiple transglutaminase homologs in Mycobacteria implies a possible role for these proteins in the development of the unusual surface structures found in these bacteria (Brennan & Nikaido, 1995). Another likely function for the secreted transglutaminase homologs from Mycobacteria and *B. pertussis* is cleavage of host proteins as part of the intracellular life cycle of these pathogens or in the induction of apoptosis. Thus, these proteins might be potential new therapeutic targets. Finally, the proteins that contain inactivated transglutaminase-like domains may perform structural and protein-binding functions as shown for the 4.2 protein (Cohen et al., 1993), or alternatively, might function as dominant-negative regulators of active enzymes.

This analysis showed that a class of enzymes hitherto only found in animals and plants in fact is widespread in prokaryotes and apparently is of an ancient origin. Previous structural comparisons have led to the suggestion that transglutaminases have an evolutionary relationship with papain-like thiol proteases; this is compatible with the local sequence similarities seen in the vicinity of the catalytic residues. Here we demonstrated a statistically significant sequence similarity between the animal transglutaminases

and a distinct class of microbial proteins whose only functionally characterized representative is a protease. This suggests that animal transglutaminases have evolved from ancestral proteases.

Acknowledgments: K.S.M. is supported by DOE OBER grant DE-FG02-98ER62583. We thank Nick Grishin for helpful discussions.

References

Aeschlimann D, Kaupp O, Paulsson M. 1995. Transglutaminase-catalyzed matrix cross-linking in differentiating cartilage: Identification of osteonectin as a major glutaminyl substrate. *J Cell Biol 129*:881–892.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Mille W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res 25*:3389–3402.

Brennan PJ, Nikaido H. 1995. The envelope of mycobacteria. *Annu Rev Biochem 64*:29–63.

The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science 282*:2012–2028.

Cariello L, Ristoratore F, Zanetti L. 1997. A new transglutaminase-like from ascidian Ciona intestinalis. *FEBS Lett 408*:171–176.

Cohen CM, Dotimas E, Korsgren C. 1993. Human erythrocyte membrane protein band 4.2 (pallidin). *Semin Hematol 30*:119–137.

Gardner MJ, Tettelin H, Carucci DJ, Cummings LM, Aravind L, Koonin EV, Shallom S, Mason T, Yu K, Fujii C, et al. 1998. Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science 282*:1126–1132.

Hettasch JM, Greenberg CS. 1994. Analysis of the catalytic activity of human factor XIIIa by site-directed mutagenesis. *J Biol Chem 269*:28309–28313.

Higgins DG, McConnell DJ, Sharp PM. 1989. Malarial proteinase? [letter]. *Nature 340*:604.

Horiguchi Y, Inoue N, Masuda M, Kashimoto T, Katahira J, Sugimoto N, Matsuda M. 1997. *Bordetella bronchiseptica* dermonecrotizing toxin induces reorganization of actin stress fibers through deamidation of Gln-63 of the GTP-binding protein Rho. *Proc Natl Acad Sci USA 94*:11623–11626.

Hubbard TJP, Ailey B, Brenner SE, Murzin AG, Chothia C. 1999. SCOP: A structural classification of proteins database. *Nucleic Acids Res 27*:254–256.

Kim SY, Chung SI, Steinert PM. 1995. Highly active soluble processed forms of the transglutaminase 1 enzyme in epidermal keratinocytes. *J Biol Chem 270*:18026–18035.

Kraulis P. 1991. A program to produce both detailed and schematic plots of proteins. *J Appl Crystallogr 24*:946–950.

Lorand L, Conrad SM. 1984. Transglutaminases. *Mol Cell Biochem 58*:9–35.

Madi A, Punyiczki M, di Rao M, Piacentini M, Fesus L. 1998. Biochemical characterization and localization of transglutaminase in wild-type and cell-death mutants of the nematode *Caenorhabditis elegans*. *Eur J Biochem 253*:583–590.

Makarova KS, Aravind L, Galperin MY, Grishin NV, Tatusov RL, Wolf YI, Koonin EV. 1999. Comparative genomics of the archaea: Evolution of conserved protein families, the stable core and the variable shell. *Genome Research*. Forthcoming.

Micanovic R, Procyk R, Lin W, Matsueda CR. 1994. Role of histidine 373 in catalytic activity of coagulation factor XIII. *J Biol Chem 269*:9190–9194.

Moens S, Vanderleyden J. 1997. Glycoproteins in prokaryotes. *Arch Microbiol 168*:169–175.

Muszbek L, Adany R, Mikkola H. 1996. Novel aspects of blood coagulation factor XIII. I. Structure, distribution, activation, and function. *Crit Rev Clin Lab Sci 33*:357–421.

Nakaoka H, Perez DM, Baek KJ, Das T, Husain A, Misono K, Im MJ, Graham RM. 1994. Gh: A GTP-binding protein with transglutaminase activity and receptor signaling junction. *Science 264*:1593–1596.

Pedersen LC, Yee VC, Bishop PD, Le Trong I, Teller DC, Stenkamp RE. 1994. Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules. *Protein Sci 3*:1131–1135.

Peitsch MC. 1996. ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochem Soc Trans 24*:274–279.

Pfister P, Wasserfallen A, Stettler R, Leisinger T. 1998. Molecular analysis of Methanobacterium phage psiM2. *Mol Microbiol 30*:233–244.

Schmidt G, Selzer J, Lerm M, Aktories K. 1998. Thr Rho-deamidating Cytotoxic Necrotizing Factor 1 from *Escherichia coli* possesses transglutaminase activity. *J. Biol Chem 273*:13669–13674.

Schroder E, Phillips C, Garman E, Harlos K, Crawford C. 1993. X-ray crystallographic structure of papain-leupeptin complex. *FEBS Lett 315*:38–42.

Seledtsov IA, Vul'f IuI, Makarova KS. 1995. Multiple alignment of biopolymer sequences, based on the search for statistically significant common segments. *Mol Biol (Mosk) 29*:1023–1039.

Serafini-Fracassini D, Del Duca S, Beninati S. 1995. Plant transglutaminases. *Phytochemistry 40*:355–365.

Sleytr UB, Sara M. 1997. Bacterial and archaeal S-layer proteins: Structure-function relationships and their biotechnological applications. *Trends Biotechnol 15*:20–26.

Steinert PM, Marekov LN. 1995. The proteins elafin, filaggrin, keratin intermediate filaments, loricrin, and small proline-rich proteins 1 and 2 are isodipeptide cross-linked components of the human epidermal cornified cell envelope. *J Biol Chem 270*:17702–17711.

Tokunaga F, Muta T, Iwanaga S, Ichinose A, Davie EW, Kuma K, Miyata T. 1993. Limulus hemocyte transglutaminase. cDNA cloning, amino acid sequence, and tissue localization. *J Biol Chem 268*:262–268.

Washizu K, Ando K, Koikeda S, Hirose S, Matsuura A, Takagi H, Motoki M, Takeuchi K. 1994. Molecular cloning of the gene for microbial transglutaminase from Streptoverticillium and its expression in *Streptomyces lividans*. *Biosci Biotechnol Biochem 58*:82–87.

Weraarchakul-Boonmark N, Jeong JM, Murthy SN, Engel JD, Lorand L. 1992. Cloning and expression of chicken erythrocyte transglutaminase. *Proc Natl Acad USA 89*:9804–9808.

Wootton JC. 1994. Non-globular domains in protein sequences: Automated segmentation using complexity measures. *Comput Chem 18*:269–285.

Wootton JC, Federhen S. 1996. Analysis of compositionally biased regions in sequence databases. *Methods Enzymol 266*:554–571.

Yee VC, Pedersen LC, Le Trong I, Bishop PD, Stenkamp RE, Teller DC. 1994. Three-dimensional structure of a transglutaminase: Human blood coagulation factor XIII. *Proc Natl Acad Sci USA 91*:7296–7300.