FOR THE RECORD

Non-animal origin of animal thioredoxin reductases: Implications for selenocysteine evolution and evolution of protein function through carboxy-terminal extensions

SERGEY V. NOVOSELOV AND VADIM N. GLADYSHEV

Department of Biochemistry, University of Nebraska, Lincoln, Nebraska 68588–0664, USA

(RECEIVED August 1, 2002; FINAL REVISION November 1, 2002; ACCEPTED November 4, 2002)

Abstract

Thioredoxin reductase (TR) and thioredoxin constitute a major cellular redox system present in all organisms. In contrast to a single form of thioredoxin, there are two TR types: One (bacterial type or small TR) is present in bacteria, archaea, plants, and most unicellular eukaryotes, whereas the second (animal or large TR) is only found in animals and typically contains a carboxy-terminal penultimate selenocysteine encoded by TGA. Surprisingly, we detected sequences of large TRs in various unicellular eukaryotes. Moreover, green algae *Chlamydomonas reinhardtii* had both small and large TRs, with the latter being a selenoprotein, but no examples of horizontal gene transfer from animals to the green algae could be detected. In addition, phylogenetic analyses revealed that large TRs formed a subgroup of lower eukaryotic glutathione reductases (GRs). The data suggest that the large TR evolved in a lower eukaryote capable of selenocysteine insertion rather than in an animal. The enzyme appeared to evolve by a carboxy-terminal extension of GR such that the resulting carboxy-terminal glutathionelike peptide became an intramolecular substrate for GR and a reductant for thioredoxin. Subsequently, small TRs were lost in an organism that gave rise to animals, large TRs were lost in plants and fungi, and selenocysteine/cysteine replacements took place in some large TRs. Our data implicate carboxy-terminal extension of proteins as a general mechanism of evolution of new protein function.

Keywords: Evolution; thioredoxin reductase; selenocysteine; cysteine; carboxy-terminal extension

The thioredoxin (Trx) system constitutes a major cellular redox pathway and is present in all living organisms (Arner and Holmgren 2000; Powis and Monfort 2001). Thioredoxin is a small thiol/disulfide oxidoreductase that uses a CxxC motif (two cysteines separated by two other residues) to maintain the reduced state of cysteines in proteins (Martin 1995). This $\alpha\beta$ protein is conserved in every organism from bacteria to humans and defines a Trx-fold family of proteins, with other members of the family being peroxiredoxins, protein disulfide oxidoreductases, glutaredoxins,

glutathione peroxidases, glutathione-S-transferases, and other proteins (Martin 1995).

Reducing equivalents for Trx reduction are provided by NADPH-dependent thioredoxin reductase (TR), a member of the pyridine nucleotide disulfide oxidoreductase family (Arner and Holmgren 2000). In contrast to a single form of Trx, two forms of TR are known (Williams et al. 2000). The so-called bacterial-type TR is a homodimer of ~35-kD subunits and is present in bacteria, archaea, and previously characterized lower eukaryotes, including plants and fungi (Fig. 1A). This enzyme is a member of the pyridine nucleotide disulfide oxidoreductase class II family. For substrate reduction, this TR uses a CxxC motif, which can accept electrons from a protein-bound flavin adenine dinucleotide (FAD) and further transfer them to oxidized Trx using ex-

Reprint requests to: Vadim N. Gladyshev, Department of Biochemistry, University of Nebraska, Lincoln, NE 68588–0664; e-mail: vgladyshev1@unl.edu; fax: (402) 472-7842.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.0226503>.

Figure 1. Schematic representation of occurrence of small and large TRs. (A) Previously, large TRs were thought to be specific to animals, whereas small TRs specific to non-animal eukaryotes and prokaryotes. (B) Identification of large TRs in non-animal eukaryotes modified the view of occurrence of large TRs. Small TRs are illustrated by smaller green circles—each circle is a subunit of a homodimeric enzyme—and the large TRs, by larger red subunits (also as a homodimer). Eukaryotic kingdom-level phylogeny is shown as in Baldauf et al. (2000), except that the location of *Dictyostelium* is not defined.

tensive conformational changes and domain rearrangements. The bacterial-type TR is also called a small TR, which distinguishes this enzyme from the second TR form, designated large TR.

Large TRs are highly homologous to glutathione reductases (GR), trypanothione reductases, and lipoamide dehydrogenases rather than to small TRs and are members of the pyridine nucleotide disulfide oxidoreductase class I family (Williams et al. 2000). Active sites of large TRs as well as of their close homologs use a CxxxxC motif (two cysteines separated by four other residues) that forms a reversible disulfide bond during the catalytic cycle. Large TRs had been found only in animals and are often called animal TRs (Fig. 1A). Most of these enzymes contain a selenocysteine (Sec) residue present in a carboxy-terminal Gly–Cys–Sec–Gly motif (Gladyshev et al. 1996), which is reduced by an amino-terminal disulfide active center and in turn can reduce Trx (Sandalova et al. 2001; Sun et al. 2001). Because of the presence of Sec, animal TRs have broad substrate specificity and can reduce a variety of proteins and low molecular weight compounds (Arner and Holmgren 2000). Some members of the large TR family, such as two *Drosophila* TRs (Kanzok et al. 2001) and one of two *Caenorhabditis elegans* TRs (Gladyshev et al. 1999) have cysteine in place of Sec.

Small TRs and the enzymes of the glutathione reductase family (including large TRs) apparently evolved by convergent evolution of similar function (Kuriyan et al. 1991). In other words, these enzymes independently acquired thiol/disulfide oxidoreductase activity. It is thought that small TRs are ancient enzymes, which were replaced during evolution with large TRs. However, important questions regarding TR evolution have yet to be answered. For example, when were small TRs replaced with large TRs? What trig-

gered these events? Did small and large TRs coexist during evolution? If evolution of large TRs is a recent event, what was the prototype enzyme and how did these enzymes evolve? Why do insect and some nematode TRs have the carboxy-terminal penultimate cysteine, but mammals contain Sec in this position? Which of these forms was the original form? What was the mechanism for replacing Sec with Cys (or vice versa)? Unexplained are also evolutionary relationships among animal TRs. For example, two known *Drosophila* TRs are most closely related to a mammalian mitochondrial TR3, whereas no close homologs of mammalian cytosolic TR1 could be found in fruit flies.

One previously characterized enzyme, a *Plasmodium falciparum* TR, did not fit to the observation of exclusive occurrence of large TRs in animals. Although this enzyme uses a carboxy-terminal Cys–Gly–Gly–Lys–Gly–Cys–Gly motif, it belongs to a family of large (animal) TRs (Becker et al. 2001). Because *Plasmodium* is a unicellular eukaryotic parasite (a human pathogen and a major cause of malaria), the possibility remained that *Plasmodium* acquired this protein from an animal host.

In this study, we analyzed evolutionary relationships among TRs, which provided an explanation for events that led to the evolution of large TRs. These results are discussed in regard to selenocysteine evolution and implicate carboxy-terminal extensions and deletions in proteins as a general mechanism of evolution of protein function.

Results and discussion

Large TRs were previously thought to exist exclusively in animals, whereas bacteria, archaea, plants, and unicellular eukaryotes contained small TR homologs. To study the occurrence of large and small TRs, nonredundant, microbial

genomic and EST eukaryotic databases were analyzed by TBLASTN against human TR1, TR2, and TR3 sequences. Surprisingly, we detected homologs of these large TRs in a ciliate *Tetrahymena* (EST sequences BM395971, BM395814, BM395972) and a parasite *Toxoplasma* (EST sequences BI997644, BG852257, BG852256, and BG852259), both unicellular eukaryotic organisms, but not in any prokaryotic genomes. We also identified and cloned (Novoselov et al. 2002) a large TR from *Chlamydomonas*

reinhardtii, a green algae and a member of the plant kingdom (Fig. 2B). Because plants are known to contain exclusively small TR, we tested, by homology searches, the presence of small TRs in *Chlamydomonas* and identified three such proteins in this green algae (accession nos. for EST sequences, first: BF863161, AV633771, and AV624607; second: AV387919, AW758306, BE122071, and AV628812; and third: AV628023, AV631626, AV628019, and AV629759). The finding of large TRs in lower eukary-

Figure 2. Phylogenetic analysis of pyridine nucleotide disulfide oxidoreductase family. (A) Schematic representation of the tree that shows enzyme origins and consists of both class I (including large TRs) and class II (small TRs) pyridine nucleotide disulfide oxidoreductases. (B) Pyridine nucleotide disulfide oxidoreductase class I tree. Separation of prokaryotic and eukaryotic glutathione reductases (GR) and the origin of large TRs are indicated on the tree. Accession numbers are shown for each protein sequence, and those large TRs that are selenoproteins are indicated by stars. (C) Small TR tree. The trees were constructed using sequences with indicated accession numbers and a PileUP program (part of Wisconsin Package GCG) as described in Novoselov et al. (2002) and Kryukov and Gladyshev (2000). Support values are bootstrap proportion calculated using PHYLIP package; 100 bootstrap replicates (SEQBOOT program), Kimura distance correction method (PROTDIST program), UPGMA algorithm (NEIGHBOR program), and Majority Rule type consensus tree (CONSENSE program) were used. TR, thioredoxin reductase; GR, glutathione reductase; LADH, lipoamide dehydrogenase; MerA, mercuric ion reductase.

otes was also consistent with the previously reported occurrence of a large TR in *Plasmodium falciparum* (Becker et al. 2000). However, in contrast to *Plasmodium*, *Chlamydomonas* is not a parasite, and currently is the only known organism that has both TR types.

The *Chlamydomonas* large TR was a Sec-containing enzyme and initial BLAST homology analyses revealed that its closest homolog was human TR1, whereas several other animal TRs, such as both TRs from *Drosophila* and mammalian TR3, were more distantly related to the algal TR. These observations suggested that either large TRs evolved early in the evolution of eukaryotes (e.g., before formation of multicellular organisms), or that the genes for large TRs have been transferred to certain lower eukaryotes from animals by horizontal gene transfer. To distinguish between these two possibilities, we first constructed a phylogenetic tree for the pyridine nucleotide disulfide oxidoreductase family (Fig. 2).

Within this tree, most large TRs could be divided into two groups. One group (TR1/2) was represented by mammalian TR1 and TR2 and consisted exclusively of selenoproteins, whereas the second group (TR3 group) was represented by mammalian mitochondrial TR3 and had both selenoproteins and cysteine homologs (Fig. 2A,B). Surprisingly, both *Drosophila* TRs were present in the TR3 group, whereas this organism lacked TR1 group members. *Chlamydomonas* TR1 was a member of the TR1/2 group. Because *Toxoplasma* and *Tetrahymena* TRs were represented by only partial sequences and these were derived from ESTs (therefore, sequence errors could have influenced the results of phylogenetic analyses), we could not reliably determine the place of these proteins in the phylogenetic tree.

The overall composition of the pyridine nucleotide disulfide oxidoreductase tree (Fig. 2A) confirmed the independent origin of small and large TRs, as the latter clustered with lipoamide dehydrogenases, mercuric ion reductases, glutathione reductases, and trypanothione reductases (Fig. 2B). Small TRs were present in all prokaryotes and most non-animal eukaryotes and formed a separate cluster (Fig. 2A,C).

Interestingly, large TRs formed a clear subgroup of the eukaryotic GR family (Fig. 2B) and the two enzymes appeared to diverge in early eukaryotes. Moreover, the location of the *Chlamydomonas* TR1 in the TR1/2 group suggested that TR1/2 and TR3 groups could have separated before divergence of plants and animals.

To further distinguish whether large TRs evolved in an early eukaryote or evolved in animals and were transferred to some lower eukaryotes by horizontal gene transfer, we tested whether a lower eukaryote that expresses a large TR has genes transferred from animals. To search for possible examples of horizontal gene transfer, we obtained all available *Chlamydomonas* protein sequences (608 nonredundant sequences) from SWISS-PROT and TrEMBL and con-

structed phylogenetic trees for all these proteins (data not shown). As expected, most *Chlamydomonas* proteins clustered with plant sequences. In addition, we detected 12 proteins, which clustered with animal sequences. All these proteins were involved in flagella formation: 13-kD deflagellation-inducible protein (Q9XF62); 78-kD dynein, intermediate chain, flagellar outer arm (Q39578); 14-kD light chain, dynein, flagellar outer arm (Q39591); 8-kD light chain, dynein, flagellar outer arm (Q39580); dynein, light chain Tctex1 (T07930); 19-kD outer arm dynein, light chain (O04355); 28-kD inner arm dynein, light chain (Q39604); kinesin-like protein FLA10 (P46869); kinesin-like protein KLP1 (P46870); radial spoke protein 3 (P12759); flagellar radial spoke protein 4 (Q01656); and outer arm dynein, light chain 1 (Q9XHH). However, no plant orthologs of these proteins were detected in either nonredundant database or the genome of *Arabidopsis thaliana*, which precluded characterization of their phylogenetic relationships. Thus, no examples of horizontal gene transfer from animals were detected by phylogenetic profiles of *Chlamydomonas* proteins.

Our study suggests a scenario for evolution of large TRs that is consistent with the phylogenetic trees shown in Figure 2. Early eukaryotes appeared to evolve large TR, and this protein coexisted with small TR. Subsequently, most organisms, including fungi and plants, lost large TRs and retained small TRs. In contrast, *Plasmodium* and animals only retained large TR, but lost small TRs. In addition, large TRs formed two groups before divergence of animals and plants. These two groups of large TRs were retained in vertebrates (TR1 gene has further duplicated to generate mammalian TR1s and TR2s). However, *Drosophila* lost the TR1 gene and duplicated the TR3 gene, whereas *Chlamydomonas* lost the TR3 gene and retained TR1 (as well as small TRs).

Because large TRs occur as both Sec-containing and Cys-containing proteins, it was not clear which of these forms was the prototypic large TR. It appears that the TR1 group consists exclusively of Sec-containing proteins, whereas the TR3 group contains both selenoproteins and Cys-containing proteins (Fig. 2B). Because Sec is used by both groups, we consider the initial occurrence of a Sec-containing large TR as a more likely scenario. A less likely possibility is that a Cys-containing form evolved first; Cys was replaced with Sec and the Sec-encoding gene duplicated to generate TR1/2 and TR3 families. Independently of the initial redox residue, large TR likely evolved in an organism capable of Sec insertion as Sec insertion evolved before the formation of the three major domains of life (bacteria, archaea, and eukaryotes) (Gladyshev and Kryukov 2001) and has been conserved in all organisms in which large TR sequences can be detected.

The carboxy-terminal Gly-Cys-Sec-Gly tetrapeptide of large TRs is an intramolecular substrate for the amino-ter-

minimal thiol/disulfide active site in the enzymes. It resembles oxidized glutathione (GSSG), the GR substrate (Sandalova et al. 2001; Sun et al. 2001). Both contain reactive cysteines (or Sec) and glycines and may be viewed as small substrates. GSSG and the tetrapeptide differ in that the disulfide bond in GSSG is intermolecular and that this compound has an additional residue, γ -glutamate. If the prototype large TR was a Sec-containing enzyme, the Sec-encoding UGA should have evolved in parallel with the Sec insertion sequence (SECIS) element. The SECIS elements reside in 3'-UTRs of eukaryotic selenoprotein genes and are required for recognition of UGA as Sec codons (Low and Berry 1996). The carboxy-terminal sequences of GR and other pyridine nucleotide disulfide oxidoreductases exhibit low conservation and variable length (Fig. 3), suggesting that extension and shortening of carboxy-terminal sequences are easily achieved during evolution. Interestingly, in this en-

zyme superfamily, only large TRs and mercuric ion reductases (MerA) have essential carboxy-terminal sequences. In the latter protein, two cysteines bind Hg^{2+} and deliver the metal for reduction to the amino-terminal thiol/disulfide active center. Sequence alignments (Fig. 3) and phylogenetic trees (Fig. 2B) of pyridine nucleotide disulfide oxidoreductases suggest that functional groups in carboxy-terminal extensions of large TRs and MerAs were generated by independent evolutionary events.

Little is known about the mechanism of evolution of Sec-containing proteins (Gladyshev and Kryukov 2001). However, the majority of selenoproteins have homologs in other species that contain Cys in place of Sec. Catalytic functions of selenoproteins are generally superior to those of Cys homologs, but the latter proteins do not need to be dependent on supply of selenium, a trace element (Hatfield and Gladyshev 2002). Thus, it appears that depending on the

Large TR1 <i>H. sapiens</i>	GLTKKQLDST	IGIHPTVCAEV	FTTLSVTKRS	GASIL.QAGG	UG-----
Large TR1 <i>M. musculus</i>	GLTKQQLDST	IGIHPTVCAEI	FTTLSVTKRS	GGDIL.QSGG	UG-----
Large TR2 <i>H. sapiens</i>	GLTKQLLDDT	IGIHPTCGEV	FTTLEITKSS	GLDI.TQKGC	UG-----
Large TR2 <i>M. musculus</i>	GLTKQLLDDT	IGIHPTCGEV	FTTLEITKSS	GLDI.TQKGC	UG-----
Large TR2 <i>X. leavis</i>	GLTKEKLDST	IGIHPTCAEI	FTTMDTSKSS	GGDI.SQKGC	UG-----
Large TR2 <i>T. rubripes</i>	GATKEQLDGT	IGIHPTCAEC	RTIF.AKHC	WDTC.QEKRC	UG-----
Large TR1 <i>S. mansoni</i>	GATKADFDR	IGIHPTCSET	FTTLHVTKKS	GVSP.IVSQC	UG-----
Large TR1 <i>C. elegans</i>	AAKADFDR	IGIHPTVAEN	FTTLTLEKKE	GDEELQASG	UG-----
Large TR1 <i>C. reinhardtii</i>	NATKADFDDC	IGIHPTVAEE	FTTLEVTKRS	GKSA.LKKGC	UG-----
Large TR1 <i>M. domestica</i>	GLTMKILLNT	VGIHPTTAAE	FTRLAITKRS	GLDP.TPASC	CS-----
Large TR1 <i>D. melanogaster</i>	GLTINTLLNT	VGIHPTTAAE	FTRLAITKRS	GLDP.TPASC	CS-----
Large TR2 <i>D. melanogaster</i>	GLTVKTLTNT	VGIHPTTAAE	FTRLITKRS	GRDP.TPASC	CS-----
Large TR3 <i>H. sapiens</i>	CASYAQVMRT	VGIHPTCSSE	VVKLRISKRS	GLDP.TVTGC	UG-----
Large TR3 <i>R. norvegicus</i>	CASYAQVMQT	VGIHPTCSSE	VVKLRISKRS	GLDP.TVTGC	UG-----
Large TR3 <i>T. rubripes</i>	GATYSHLLQT	VGIHPTCAEE	VIKVNITKRS	GLDA.TVTGC	UG-----
Large TR <i>P. falciparum</i>	KVKKKDFDNC	IGIHPTDAES	FMNLFVTISS	GLSYAAKGGC	GGGKCG-----
GR <i>N. tabacum</i>	GLTKADFDT	VGIHPTSAEE	FVTMRTPTRK	VR.SSPSEGG	AEHDIKAAAGV
GR <i>V. vinifera</i>	GLTKADFDT	VGIHPTAAEE	LVTMRTPTRK	IRNSSPPEGK	TDPEVKAAAGV
GR <i>A. thaliana</i>	GLTKADFDT	VGIVPTAAEE	FVTMRAPTRK	FR.KDSSEGG	ASPEAKTAAGV
GR <i>P. sativum</i>	GATKAQFDST	VGIHPTSSAE	FVTMRSETRR	VTGGVKPKTN	L-----
GR <i>H. sapiens</i>	GATKADFNT	VAIHPTSSSE	LVTILR-----	-----	-----
GR <i>M. musculus</i>	GATKADFNT	VAIHPTSSSE	LVTILR-----	-----	-----
GR <i>S. pombe</i>	GATKSDFDSC	VAIHPTSAEE	LVTILV-----	-----	-----
GR <i>M. mazei</i>	GLRASSIKKM	AFTYTTCSD	IRYML-----	-----	-----
GR <i>S. aureus</i>	GISTKELKQM	IFAYPTAASD	IAHML-----	-----	-----
LADH <i>Halobacterium</i> sp.	GATLEDVAST	IHTPTLSA	TMEAAEHALG	HAVHTLNR--	-----
LADH <i>H. volcanii</i>	GATLEDVAST	IHTPTLSA	VMEAAENALG	QAIHTLNR--	-----
LADH <i>H. sapiens</i>	GASCEDIARV	CHAHPTLSA	FREANLAASF	GKSPINF----	-----
LADH <i>M. musculus</i>	GASCEDIARV	CHAHPTLSA	FREANLAASF	GKSPINF----	-----
LADH <i>A. gambiae</i>	GASAEVARV	CHAHPTCAEA	LREAHTAASF	GKSPINF----	-----
LADH <i>D. melanogaster</i>	GAAAEVARV	CHAHPTCSEA	LR-----	-----	-----
LADH <i>A. thaliana</i>	DASSEDIAEV	CHAHPTMSEA	LKEAAMATYD	KPIHI-----	-----
LADH <i>B. gymnorrhiza</i>	DASSEDIAEV	CHAHPTMSEA	LKEAAMATYD	KPIHI-----	-----
LADH <i>O. sativa</i>	GASSEDIAEV	CHAHPTVSEA	LKEAALQFTT	KAIHI-----	-----
LADH <i>R. capsulatus</i>	GASAQDLALT	CHAHPTYSEA	VREAALACGD	GATHA-----	-----
LADH <i>S. pombe</i>	GASAEVARV	CHAHPTLSA	TKEAMMAAWC	GKSIHF-----	-----
LADH <i>P. carbinolicus</i>	GGTAHDIGAM	FHGHPTLSA	VKEAALDVG	AAVHC-----	-----
LADH <i>X. campestris</i>	SGSADDLARI	CHAHPTLSA	IHDAAAVSK	RAIHKAN---	-----
MerA <i>E. cloacae</i>	RMTVQELADQ	LFPYLLTMVG	LKLAQTFSK	DVKQLSCAG	-----
MerA <i>T. acidophilum</i>	GMTVGDVDT	PHVFPTVNG	IKLAAQSFLR	DISMMSGME	-----

Figure 3. Alignment of carboxy-terminal sequences in pyridine nucleotide disulfide oxidoreductase class I family. Enzymes were aligned with BLASTP, but only carboxy-terminal regions are shown in the figure. Accession numbers for sequences are given in Figure 2B. Conserved histidine, proline, and glutamate residues are highlighted in dark gray. These residues participate in the catalytic mechanism common to all pyridine nucleotide disulfide oxidoreductase class I enzymes. In addition, conserved cysteines and selenocysteines are highlighted in dark gray. These residues evolved independently in large TRs and MerAs through carboxy-terminal extensions and deletions. Glycine residues that may assist in the catalytic reaction are shown in light gray. TR, thioredoxin reductase; GR, glutathione reductase; LADH, lipoamide dehydrogenase; MerA, mercuric ion reductase.

need of the organism and availability of selenium, both Sec can replace Cys and be replaced by Cys. An attractive possibility is that a carboxy-terminal extension in GR (in an organism capable of Sec insertion) generated an in-frame TGA codon (or TGA was a stop codon at the time). Concurrent with this process, a SECIS element evolved in the 3'-UTR, which allowed the TGA to be recognized as a Sec codon.

Conclusions

We found that animal TRs evolved in lower eukaryotes, much earlier than bacterial-type TRs were lost in species that gave rise to animals. It appears that the two TR types coexisted during evolution, and we identified an organism that contains both enzyme types (Fig. 1B). Interestingly, large TRs became a predominant TR form only in animals, whereas small TRs remained in plants and fungi. The origin of large TRs was traced to GRs of lower eukaryotes. Animal TRs and GRs currently share ~30% identity in amino acid sequences. Large TR evolved by extension of a GR carboxy-terminal sequence to generate a new redox center, which replaced GSSG as a transient substrate for the enzyme.

Initial extension or subsequent modification of the carboxyl terminus appeared to generate Sec, encoded by TGA. Perhaps, the catalytic advantage of Sec was the crucial factor for evolution of large TRs. However, for TGA to code for Sec, an additional mRNA structure was required. Thus, Sec origin was concurrent with the evolution of a SECIS element, and this process could have occurred only in organisms with active Sec insertion systems.

Generally speaking, we described a mechanism by which new functions may evolve in proteins. Carboxy-terminal amino acids are rarely conserved in proteins and are unlikely to participate in enzyme active sites. Extensions of carboxy-terminal sequences may occasionally generate sequences that could interact (or mediate interactions of other biological molecules) with upstream structures (e.g., carboxy-terminal extensions in large TRs and MerAs) or modify, in some useful ways, protein functions (e.g., carboxy-terminal extensions in certain glyceraldehyde-3-phosphate dehydrogenases and crystallins) (Pasta et al. 2002; Sparla et al. 2002). The build-up of carboxy-terminal sequences could potentially result in formation of new domains in proteins and possibly in generation of new protein folds. Carboxy-terminal extensions have a potential to change location of proteins, for example, by generating carboxy-terminal endoplasmic reticulum retention signals in secreted proteins and nuclear location signals in cytosolic proteins or by disrupting such signals for some ER-resident and nuclear proteins.

Carboxy-terminal extensions are likely more common than those at amino termini, because a single nucleotide

change (replacement, deletion, or insertion) in an ORF stop codon would generally extend the ORF, whereas an amino-terminal extension would require generation of an entire new initiator codon upstream of the actual ATG, and it should be present in the right context to allow alternative upstream initiation of protein synthesis. Nevertheless, amino-terminal extensions may also contribute to evolution of protein function, particularly by modifying cellular locations of proteins through amino-terminal signal peptides.

One could also envision carboxy-terminal shortenings (deletions) of unnecessary extensions, which may occur by occasional generation of new stop codons upstream of actual terminator signals. Thus, carboxy-terminal extensions and deletions may be viewed as a general mechanism of protein evolution. It would be interesting to determine the relative contribution of carboxy-terminal extensions/deletions to protein evolution.

Acknowledgments

We thank Gregory Kryukov and Dolph Hatfield for helpful comments. This study is supported by NIH grant GM065204 (to V.N.G.).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Arner, E.S. and Holmgren, A. 2000. Physiological functions of thioredoxin and thioredoxin reductase. *Eur. J. Biochem.* **267**: 6102–6109.
- Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., and Doolittle, W.F. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**: 972–977.
- Becker, K., Gromer, S., Schirmer, R.H., and Muller, S. 2000. Thioredoxin reductase as a pathophysiological factor and drug target. *Eur. J. Biochem.* **267**: 6118–6125.
- Gladyshev, V.N. and Kryukov, G.V. 2001. Evolution of selenocysteine-containing proteins: Significance of identification and functional characterization of selenoproteins. *Biofactors* **14**: 87–92.
- Gladyshev, V.N., Jeang, K.T., and Stadtman, T.C. 1996. Selenocysteine, identified as the penultimate C-terminal residue in human T-cell thioredoxin reductase, corresponds to TGA in the human placental gene. *Proc. Natl. Acad. Sci.* **93**: 6146–6151.
- Gladyshev, V.N., Krause, M., Xu, X.M., Korotkov, K.V., Kryukov, G.V., Sun, Q.A., Lee, B.J., Wootton, J.C., and Hatfield, D.L. 1999. Selenocysteine-containing thioredoxin reductase in *C. elegans*. *Biochem. Biophys. Res. Commun.* **259**: 244–249.
- Hatfield, D.L. and Gladyshev, V.N. 2002. How selenium has altered our understanding of the genetic code. *Mol. Cell. Biol.* **22**: 3565–3576.
- Kanzok, S.M., Fechner, A., Bauer, H., Ulschmid, J.K., Muller, H.M., Botella-Munoz, J., Schneuwly, S., Schirmer, R., and Becker, K. 2001. Substitution of the thioredoxin system for glutathione reductase in *Drosophila melanogaster*. *Science* **291**: 643–646.
- Kryukov, G.V. and Gladyshev, V.N. 2000. Selenium metabolism in zebrafish: Multiplicity of selenoprotein genes and expression of a protein containing 17 selenocysteine residues. *Genes Cells* **5**: 1049–1060.
- Kuriyan, J., Krishna, T.S., Wong, L., Guenther, B., Pahlner, A., Williams, Jr., C.H., and Model, P. 1991. Convergent evolution of similar function in two structurally divergent enzymes. *Nature* **352**: 172–174.
- Low, S.C. and Berry, M.J. 1996. Knowing when not to stop: Selenocysteine incorporation in eukaryotes. *Trends Biochem. Sci.* **21**: 203–208.

- Martin, J.L. 1995. Thioredoxin—A fold for all reasons. *Structure* **3**: 245–250.
- Novoselov, S.V., Rao, M., Onoshko, N.V., Zhi, H., Kryukov, G.V., Xiang, Y., Weeks, D.P., Hatfield, D.L., and Gladyshev, V.N. 2002. Selenoproteins and selenocysteine insertion system in the model plant cell system, *Chlamydomonas reinhardtii*. *EMBO J.* **21**: 3681–3693.
- Pasta, S.Y., Bakthisaran, R., Tangirala, R., and Mohan Rao, C. 2002. Role of the C-terminal extensions of α -crystallins: Swapping the C-terminal extension of α A-crystallin to α B-crystallin results in enhanced chaperone activity. *J. Biol. Chem.* Sept. 15.
- Powis, G. and Montfort, W.R. 2001. Properties and biological activities of thioredoxins. *Annu. Rev. Biophys. Biomol. Struct.* **30**: 421–455.
- Sandalova, T., Zhong, L., Lindqvist, Y., Holmgren, A., and Schneider, G. 2001. Three-dimensional structure of a mammalian thioredoxin reductase: Implications for mechanism and evolution of a selenocysteine-dependent enzyme. *Proc. Natl. Acad. Sci.* **98**: 9533–9538.
- Sparla, F., Pupillo, P., and Trost, P. 2002. The C-terminal extension of glyceraldehyde-3-phosphate dehydrogenase subunit B acts as an autoinhibitory domain regulated by thioredoxins and nicotinamide adenine dinucleotide. *J. Biol. Chem.* **47**: 44946–44952.
- Sun, Q.A., Kirnarsky, L., Sherman, S., and Gladyshev, V.N. 2001. Selenoprotein oxidoreductase with specificity for thioredoxin and glutathione systems. *Proc. Natl. Acad. Sci.* **98**: 3673–3678.
- Williams, C.H., Arscott, L.D., Muller, S., Lennon, B.W., Ludwig, M.L., Wang, P.F., Veine, D.M., Becker, K., and Schirmer, R.H. 2000. Thioredoxin reductase two modes of catalysis have evolved. *Eur. J. Biochem.* **267**: 6110–6117.