Biosci. Rep. (2013) / **33** / art:e00015 / doi 10.1042/BSR20120112

OPEN ACCESS

Original Paper

Characterization of the bipartite degron that regulates ubiquitin-independent degradation of thymidylate synthase

Karen W. BARBOUR*†, Yang-Yang XING*†, Edsel A. PEÑA*‡ and Franklin G. BERGER*†[1]

*Center for Colon Cancer Research, University of South Carolina, Columbia, SC 29208, U.S.A., †Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, U.S.A., and ‡Department of Statistics, University of South Carolina, Columbia, SC 29208, U.S.A.

Synopsis

TS (thymidylate synthase) is a key enzyme in the *de novo* biosynthesis of dTMP, and is indispensable for DNA replication. Previous studies have shown that intracellular degradation of the human enzyme [hTS (human thymidylate synthase)] is mediated by the 26S proteasome, and occurs in a ubiquitin-independent manner. Degradation of hTS is governed by a degron that is located at the polypeptide's N-terminus that is capable of promoting the destabilization of heterologous proteins to which it is attached. The hTS degron is bipartite, consisting of two subdomains: an IDR (intrinsically disordered region) that is highly divergent among mammalian species, followed by a conserved amphipathic α-helix (designated hA). In the present report, we have characterized the structure and function of the hTS degron in more detail. We have conducted a bioinformatic analysis of interspecies sequence variation exhibited by the IDR, and find that its hypervariability is not due to diversifying (or positive) selection; rather, it has been subjected to purifying (or negative) selection, although the intensity of such selection is relaxed or weakened compared with that exerted on the rest of the molecule. In addition, we have verified that both subdomains of the hTS degron are required for full activity. Furthermore, their co-operation does not necessitate that they are juxtaposed, but is maintained when they are physically separated. Finally, we have identified a 'cryptic' degron at the C-terminus of hTS, which is activated by the N-terminal degron and appears to function only under certain circumstances; its role in TS metabolism is not known.

Key words: proteasome, relaxed purifying selection, thymidylate synthase, ubiquitin-independent

Cite this article as: Barbour, K.W. Xing, Y.-Y., Peña, E.A. and Berger, F.G. (2013) Characterization of the bipartite degron that regulates ubiquitin-independent degradation of thymidylate synthase. Biosci. Rep. **33**(1), art:e00015.doi:10.1042/BSR20120112

INTRODUCTION

The pyrimidine biosynthetic enzyme TS (thymidylate synthase; EC 2.1.1.45) catalyses the reductive methylation of dUMP by N^5N^{10}-methylenetetrahydrofolate, to form TMP and DHF (dihydrofolate) [1,2]. In being the sole *de novo* source of TMP for cellular DNA replication and repair, the enzyme has been an effective target for chemotherapeutic drugs used in the treatment of cancer [1,2]. Anti-metabolites such as 5-fluorouracil and raltitrexed form TS substrate analogues that bind to and inhibit the enzyme, leading to reduced thymidylate pools and apoptosis. X-ray crystallographic studies of the TS polypeptide from an evolutionarily wide range of eukaryotic and prokaryotic species have shown that the enzyme is highly conserved in terms of the location of secondary structural elements and overall three-dimensional (3D) conformation [3–6].

In previous work, we demonstrated that the intracellular stability of TS is increased by the binding of inhibitory ligands [7], a phenomenon that potentially constrains the effectiveness of TS-directed anti-cancer agents. Subsequent studies showed that the human enzyme [designated hTS (human thymidylate synthase)] is degraded by the 26S proteasome in a ubiquitin-independent manner [8–11]. Most proteasomal substrates require ligation of ubiquitin moieties to be recognized and degraded by the proteasome [12]. However, an increasing number of substrates

Abbreviations used: CHX, cycloheximide; eGFP, enhanced green fluorescent protein; FEL, fixed effects likelihood; GFP, green fluorescent protein; hA, α-helix; HLH, helix-loop-helix; hTS, human thymidylate synthase; IDR, intrinsically disordered region; ODC, ornithine decarboxylase; PUP, prokaryotic ubiquitin-like protein; REL, random effects likelihood; SNAP, Synonymous Non-synonymous Analysis Program; TS, thymidylate synthase.

[1] To whom correspondence should be addressed (e-mail fgberger@mailbox.sc.edu).

do not require ubiquitinylation, including ODC (ornithine decarboxylase) [13,14], tumour suppressor $p21^{Waf1/Cip1}$ [15,16], the NF-κB (nuclear factor κB) regulator IκBα (inhibitor of κBα) [17,18], apomyoglobin [19] and the yeast transcription factor Rpn4 [20], among others. Thus, hTS is a model for ubiquitin-independent protein turnover.

Studies in our laboratory have shown that an N-terminal segment of hTS spanning the first 45 residues is required for degradation, defining the region as a degron [9–11]. The hTS degron is portable, in that it is capable of destabilizing a heterologous protein to which it is attached [9–11]. Furthermore, it is bipartite in nature, comprising of two co-operating subdomains: an IDR (intrinsically disordered region) spanning residues 1–27, and an amphipathic α-helix (designated hA) at residues 31–42 [9]. The IDR is hypervariable in amino acid sequence among mammalian species [11], although its disordered nature has been conserved [4–6,11]. A free, unmodified N-terminal amino group and an arginine–arginine motif at residues 10–11 are required for IDR activity in promoting proteasomal degradation [10,11]. The hA segment, which is much more highly conserved among species [11], requires maintenance of an α-helical conformation for its degradative function [9]. Interestingly, the hTS degron is active when placed at either the N- or C-terminus of the reporter substrate [9]. The specific roles of the two subdomains in regulating hTS proteolysis have yet to be determined.

In the present report, we have carried out further analysis of the structure and function of the hTS degron. We have examined the origin of inter-species hypervariability within the IDR, and find that its tolerance of amino acid substitution derives from relaxed purifying selection, as opposed to positive Darwinian selection. We also show that co-operation between the IDR and hA does not require that the two elements be juxtaposed, but is maintained when they are physically separated. Finally, we have identified a 'cryptic' degron at the C-terminus of the hTS polypeptide that appears to be active only under particular circumstances.

EXPERIMENTAL

Bioinformatics

Sequences corresponding to the coding regions of TS mRNAs from 16 mammalian species were downloaded from the Ensembl database [21] and aligned using ClustalW2 [22,23]. Species included human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), orang-utan (*Pongo abelii*), marmoset (*Callithrix jacchus*), tarsier (*Tarsius syrichta*), rhesus monkey (*Macaca mulatta*), mouse (*Mus musculus*), rat (*Rattus rattus*), guinea pig (*Cavia porcellus*), cat (*Felis catus*), dog (*Canis lupus*), cow (*Bos taurus*), horse (*Equus caballus*), pika (*Ochotona princeps*), rabbit (*Oryctolagus cuniculus*) and dolphin (*Tursiops truncatus*). A 40-codon insertion within the cow and dog sequences, corresponding to an additional exon and likely representing a cDNA copy of an incompletely processed transcript, were removed from the analysis.

Extents of synonymous and non-synonymous substitutions within the coding regions of mammalian TS transcripts were determined for all pairwise comparisons among the 16 species using SNAP (Synonymous Non-synonymous Analysis Program), as available on the HIV database website. SNAP estimates the numbers of synonymous substitutions per synonymous site (d_S) and non-synonymous substitutions per non-synonymous site (d_N), corrected for multiple hits [24,25]. Cumulative behaviour plots for both types of substitutions along the coding regions were constructed using the 'xyplot' option of SNAP.

Identification of codons where $d_S \neq d_N$ was carried out by the FEL (fixed effects likelihood) and REL (random effects likelihood) methods, as provided on the Datamonkey webserver [26]. A P-value of 0.1 was set as the cut-off for FEL.

Prediction of secondary structure was done using PSIPRED, as provided on the Bioinformatics Group Server of University College London. PSIPRED is a two-stage neural network that predicts protein secondary structure based on position-specific scoring matrices [27].

Plasmid constructs

All constructs for mutant analysis (shown in Figure 1) were generated using standard techniques and were verified directly by DNA sequencing. The parental expression plasmid was pMP610, containing a full-length eGFP [enhanced GFP (green fluorescent protein)] cDNA under the control of a CMV (cytomegalovirus) promoter. For selection of transfectants, an *Escherichia coli* TS cDNA under the control of a murine stem cell virus LTR (long terminal repeat) promoter, optimized for expression in mammalian cells by modification of codon-usage, was included in the plasmid [28]. Mutagenesis was carried out by standard PCR-based protocols. Details of all plasmid constructions are available upon request.

Cell culture and transfection

Cell line RJK88.13, a TS-deficient derivative of V79 Chinese hamster lung cells [29], was maintained in DMEM (Dulbecco's modified Eagle's medium) (Cellgro) containing 4.5 g/l glucose and supplemented with 10 % heat-inactivated fetal bovine serum (Atlanta Biologicals) and 10 μM thymidine. Cells were cultured at 37 °C in a humidified 5 % CO_2 atmosphere.

For generation of stable transfectants, RJK88.13 cells were transfected with the indicated expression plasmids using Lipofectamine™ 2000 according to the manufacturer's instructions. Transfectants were selected in thymidine-free medium containing the nucleoside transport inhibitor dipyridamole (5 μM; Sigma–Aldrich), pooled and maintained in mass culture.

Determination of protein half-lives

Half-lives were assessed following the addition of 50 μg/ml CHX (cycloheximide; Acros Organic) to the growth medium for the indicated times. Cells were harvested by scraping, and lysed in M-Per® buffer (Pierce) containing 10 mM DTT (dithiothreitol), 2 mM 2-mercaptoethanol, 5 mM PMSF, 200 μg/ml aprotinin,

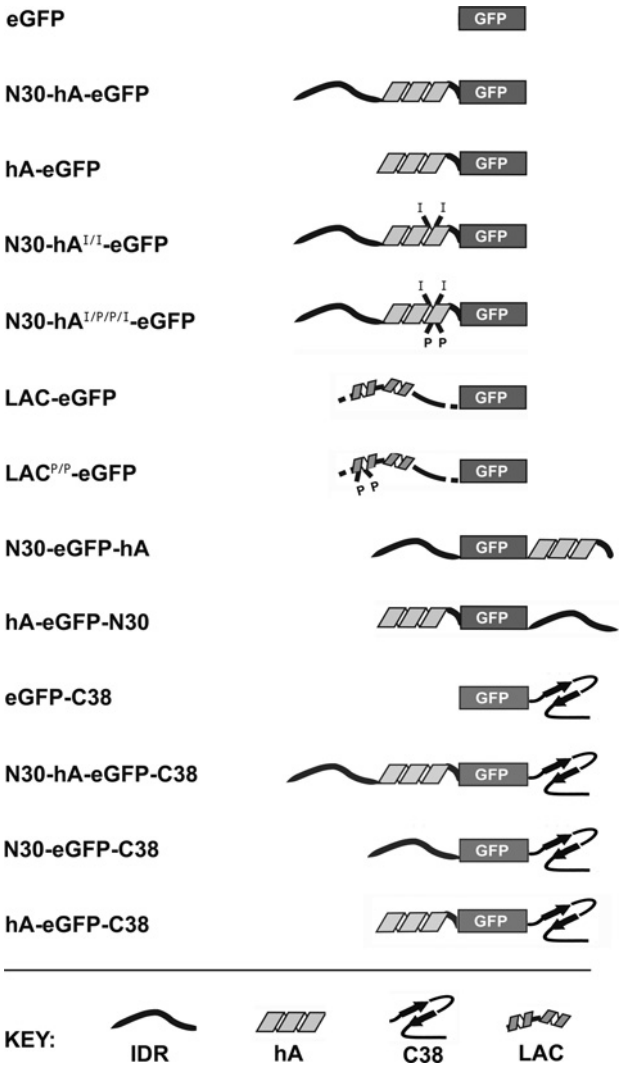

Figure 1 Constructs used in the present study

The eGFP fusion constructs utilized in the current study are shown. Schematic depictions of the various segments are indicated in the KEY at the bottom, including the disordered region (IDR), the amphipathic α-helix (hA) and the C-terminal element (C38) of hTS, as well as the N-terminal region of the *E. coli lac* repressor (LAC).

100 μg/ml pepstatin and 50 μg/ml leupeptin. Crude lysates were centrifuged at 15000 ***g*** for 1 h at 4 °C, and protein concentrations in the resulting extracts were quantified using the Bio-Rad assay reagent with BSA as a standard.

Immunoblotting was performed by standard techniques. Following SDS/PAGE, fractionated proteins were transferred to nylon membranes, and probed with a monoclonal antibody to GFP (Santa Cruz Biotechnology; catalogue no. sc-9996). To control for equal loading, blots were reprobed with an anti-actin monoclonal antibody (Sigma–Aldrich, Clone AC-40). The antigen–antibody complexes were visualized using appropriate secondary antibodies with the ECL® (enhanced chemiluminescence) kit (Amersham Biosciences). Densitometry was carried out using ImageJ software maintained by the National Institutes of Health. A 2-fold dilution series of each extract was included on the blots for calibration and to correct for film exposure times. All values were normalized to actin concentrations on the same blots. Data were plotted according to first-order decay kinetics, and half-lives were determined from the slopes of the resulting lines. Each half-life determination was carried out at least twice with independently prepared cells.

RESULTS

Origin of amino acid sequence divergence within the IDR

The amino acid sequence of the IDR within the TS degron is highly divergent among mammalian species [11]. Despite such divergence, the region's role in degradation is maintained, so long as a free, unblocked N-terminus and an arginine–arginine motif at residues 10–11 are present [10]. This, along with the observation that both the IDR and hA can be replaced with segments from unrelated proteins, led us to suggest that it is the overall structure, rather than the specific primary sequence, that is critical to degron function [9].

In order to gain a deeper understanding of IDR hypervariability and its origin, we aligned and compared the coding regions of the TS mRNAs from 16 mammalian species (Supplementary Figure S1 at http://www.bioscirep.org/bsr/033/bsr033e015add.htm). The IDR (encoded by nucleotides 1–81 of the human transcript) spans a maximum of 31 codons, including small insertions/deletions in some species. The remainder of the transcript (encoded by nucleotides 82–939 of the human, which includes hA) spans 286 codons and contains no insertions/deletions. Within the IDR, only 16 nucleotide sites (17.8 %, not including the initiator Met codon) are completely conserved among the species; most of these are located within the region encompassing residues 5–8 (Supplementary Figure S1). In contrast, 556 sites (64.8 %) within the body of the polypeptide are conserved. The difference in degree of nucleotide conservation between the IDR and the remainder of the molecule is highly significant ($P < 10^{-16}$), and indicates a dramatically higher level of divergence for the IDR-encoding portion of the mammalian TS transcript.

Cumulative behaviour plots (Figure 2) indicated that synonymous substitutions are distributed equally throughout the length of the mRNA (Figure 2). They occur at a rate of 0.227 ± 0.0079 per codon within the region encoding the IDR ($R^2 = 96.96\,\%$; 95 % CI = 0.211–0.243) and 0.233 ± 0.00073 per codon within the region encoding the body of the polypeptide ($R^2 = 99.72\,\%$; 95 % CI = 0.231–0.234). Thus, no significant differences were observed between the two domains ($P = 0.40$). In contrast, non-synonymous substitutions are distributed in a biphasic manner (Figure 2), with the rate being 0.395 ± 0.011 per codon within the IDR ($R^2 = 97.97\,\%$; 95 % CI = 0.372–0.418), as compared with 0.0660 ± 0.0005 per codon within the body of the molecule

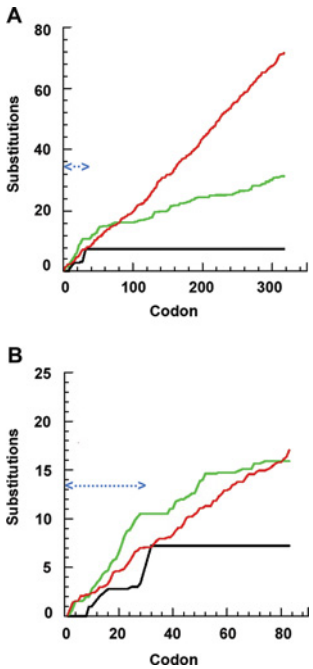

Figure 2 Cumulative behaviour of synonymous and non-synonymous substitutions along the coding region of TS mRNA
Accumulation of amino acid substitutions within the TS polypeptide for the 16 mammalian species, as determined by SNAP, is plotted along the coding region of the mRNA (see the Experimental section). Codon numbering refers to hTS. The entire coding region (313 codons) is shown in (**A**), whereas a closer view of the first 80 codons is shown in (**B**). The plot for synonymous substitutions is in red, while that for non-synonymous substitutions is in green. The black line shows accumulation of insertions/deletions. The IDR is indicated by a double-arrowed dashed line (<–>).

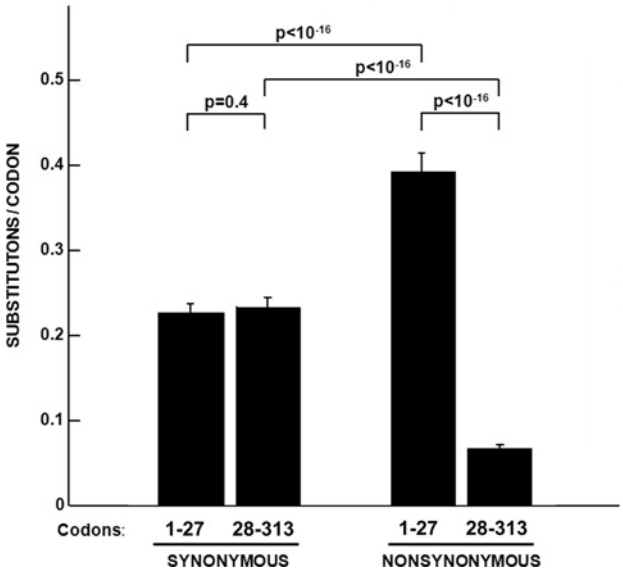

Figure 3 Rates of accumulation of synonymous and non-synonymous substitutions along the TS mRNA
Slopes of the plots in Figure 2 were determined by linear regression. Codons 1–27 represent the IDR of hTS, while codons 28-313 represent the rest of the polypeptide. Bars indicate values ± S.E.M. *P* values are shown for various comparisons, as indicated by the brackets.

($R^2 = 98.09\%$; 95 % CI = 0.065–0.067). Thus, a 6–7-fold higher rate of accumulation of non-synonymous substitutions is observed for the IDR ($P < 10^{-16}$).

Within the body of the polypeptide, the occurrence of non-synonymous substitutions is 3.5-fold lower than for synonymous substitutions (i.e. 0.0660 compared with 0.233 per codon, respectively; $P < 10^{-16}$) (Figure 3). This indicates strong negative selection constraining amino acid substitutions throughout most of the TS polypeptide. In contrast, within the IDR-encoding region, the rate for non-synonymous substitutions is 1.7-fold higher than for synonymous substitutions (0.395 compared with 0.227 per codon, respectively; $P < 10^{-16}$) (Figure 3). This may reflect positive Darwinian selection, whereby amino acid substitutions within the IDR are beneficial or adaptive, and are therefore favoured. Alternatively, it could be a consequence of relaxed or weakened intensity of negative (or purifying) selection, indicating a greater tolerance of amino acid changes.

To distinguish between these two possibilities, we determined the proportions of non-synonymous substitutions per non-synonymous site (d_N) and synonymous substitutions per synonymous site (d_S), and calculated d_N/d_S ratios. Purifying, or negative, selection is indicated when such ratios are <1, while positive selection is indicated when they are >1; d_N/d_S ratios equal to 1 indicate neutrality, or an absence of selection. The d_N and d_S values for each of the 120 pairwise combinations among the 16 mammalian species were determined for both the IDR and the body of the polypeptide (see Supplementary Table S1 at http://www.bioscirep.org/bsr/033/bsr033e015add.htm). Average d_N/d_S ratios were 0.5953 ± 0.4433 for the IDR and 0.0745 ± 0.0334 for the body of the molecule. Thus, the IDR exhibited an 8-fold higher ratio ($P < 10^{-16}$) that was still significantly less than 1.0 ($P < 10^{-16}$). This indicates that amino acid substitutions within the disordered segment are constrained by negative selection, albeit at a relaxed or milder intensity as compared with the remainder of the polypeptide. No evidence for positive Darwinian selection was detected.

As an independent test of this conclusion, we utilized an approach that identifies individual codons under positive or negative selection. REL is a maximum likelihood strategy that determines if d_N and d_S are significantly different at each codon along a mRNA [26]. The approach failed to detect any positively selected sites in either the IDR or the body of the TS polypeptide. In contrast, 10.0 % of sites within the IDR (3 of 30, not including the initiator Met codon) and 36.4 % of sites within the body of the polypeptide (104 of 286) were identified as being under negative selection. The difference in the fraction of negatively selected codons between the IDR and the body of the polypeptide is statistically significant ($P < 10^{-4}$), and indicates that the intensity of negative selection is diminished in the former.

In all, our analysis indicates that positive Darwinian selection is not a major contributor to high levels of sequence divergence

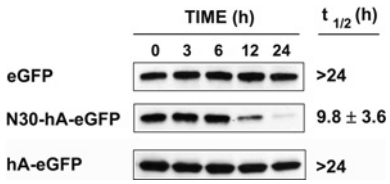

Figure 4 hA is inactive on its own as a degron

The indicated plasmid constructs (see the text and Figure 1) were stably transfected into cell line RJK88.13 and treated with CHX. Decreases in protein concentrations were monitored over time by Western blotting (see the Experimental section for details). A representative blot is shown for each construct. The half-lives ($t_{\frac{1}{2}}$) are shown to the right of the corresponding blots, and are presented as means ± S.D.

within the IDR of mammalian TS. Rather, the region is constrained by negative selection, although such selection is relaxed as compared with that exerted upon the rest of the polypeptide.

The IDR and hA are both required for degron activity

In previous investigations [11], we found that a fusion protein containing the IDR of hTS (residues 1–30, denoted N30) appended to an eGFP reporter is resistant to degradation, as is the parental eGFP reporter itself. Inserting hA into N30–eGFP between the IDR and the reporter results in a protein (N30–hA–eGFP; see Figure 1) that is significantly destabilized (Figure 4 and [11]). On the basis of these findings, we concluded that both the IDR and hA are required for degradation. However, the possibility that hA on its own functions as a degron was not formally tested. To do this, we measured the half-life of a polypeptide, hA–eGFP, in which hA alone was ligated to eGFP (Figure 1). A stably transfected cell line expressing hA–eGFP was generated (see the Experimental section), and its half-life was measured by a CHX chase assay. As seen in Figure 4, hA–eGFP, like the parental eGFP substrate, is quite stable, having a half-life >24 h. Thus, hA is inactive on its own as a degron. We conclude that neither the IDR nor hA is capable of acting alone in promoting degradation; both segments are required, and must co-operate for full degron function.

A helical conformation is required for degron function

Earlier, we observed that insertion of helix-disrupting proline substitutions at residues 40 and 41 within the hA segment of fusion protein N30–hA–eGFP stabilized the molecule, indicating that a helical conformation is required for maximal degron function [9]. This was the case regardless of whether the degron was at the N- or C-terminal end of the reporter [9]. As a more rigorous test of this requirement, we examined an eGFP-fusion protein containing a very strong degron derived from a mutant hTS molecule. Polypeptide N30–hA$^{I/I}$–eGFP (see Figure 1) contains isoleucine substitutions at positions 39 and 42 within hA, rendering the protein very unstable, with a half-life of about 0.8 h [9]. We used the secondary structure prediction tool PSIPRED [27] to assess the potential effect of proline substitutions on the

Figure 5 Predicted impact of amino acid substitutions on the helical propensity of hA segment

The helical propensity of the region spanning residues 24–50 of wild-type hTS and mutants I39/I42, P40/P41 and I39/P40/P41/I42 were determined using PSIPRED [27]. For each molecule, the amino acid sequence is shown; targeted substitutions within mutated segments are indicated in red. The predicted secondary structure is shown below each sequence; black lines indicate coiled regions, magenta cylinders indicate α-helices and yellow arrows indicate β-strands. The blue bars at the bottom show the relative confidence for the predictions at each position along the sequence, with the height and depth of blue indicating greater confidence. The predictor accurately identified hA of wild-type hTS, known from X-ray crystallographic studies [4,5].

helical propensity of the hA segment in both N30–hA$^{I/I}$–eGFP and in its wild-type counterpart N30–hA–eGFP. As seen in Figure 5, for both the wild-type and mutant segments, an α-helical domain spanning residues 31–42, which is the precise location of hA as detected by X-ray crystallography [4,5], was predicted with high confidence for both segments. In contrast, loss of the helical conformation was predicted when proline residues were introduced at positions 40 and 41 (Figure 5). Thus, the secondary structure of hA is likely to be significantly disrupted by insertion of proline substitutions at positions 40 and 41.

With this information, we measured the half-life of polypeptide N30–hA$^{I/P/P/I}$–eGFP, which is a derivative of N30–hA$^{I/I}$–eGFP containing proline substitutions at sites 40 and 41 within hA (see Figure 1). As shown in Figure 6(A), N30-hA$^{I/P/P/I}$-eGFP had a half-life >24 h, indicating it to be significantly more stable than its parent, N30–hA$^{I/I}$–eGFP. This provides strong verification of the notion that promotion of degradation by hA segment requires a helical conformation.

To assess the generality of this conclusion, we tested the requirement for α-helix in a completely different system. We have shown that the N-terminal domain of the *Escherichia coli lac* repressor, which contains an HLH (helix-loop-helix) motif spanning residues 9–29, functions as a degron [9]. Deletion of the first helix of the HLH motif (residues 9–17) stabilizes the polypeptide [9], whereas deletion of the second (residues 22–29) has no effect (K. Barbour, unpublished work); thus, the first helix drives the region's degron activity. To test if an α-helical conformation is

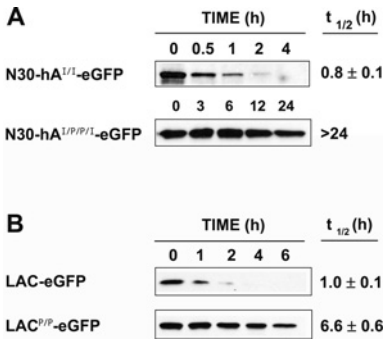

Figure 6 Degron activity requires a helical segment
The indicated plasmid constructs were stably transfected into cell line RJK88.13 and treated with CHX (see the text and Figure 1). Decreases in protein concentrations over time were monitored by Western blotting (see the Experimental section for details). A representative blot is shown for each construct. The half-lives ($t_{\frac{1}{2}}$) are shown to the right of the corresponding blots, and are presented as means ± S.D.

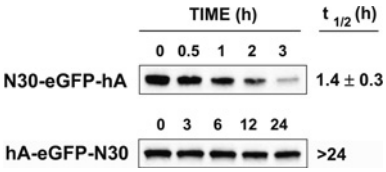

Figure 7 The IDR and hA co-operate when physically separated
Plasmid constructs were stably transfected into cell line RJK88.13 and treated with CHX (see the text and Figure 1). Decreases in protein concentrations over time were monitored by Western blotting (see the Experimental section for details). A representative blot is shown for each construct. The half-lives ($t_{\frac{1}{2}}$) are shown to the right of the corresponding blots, and are presented as means ± S.D.

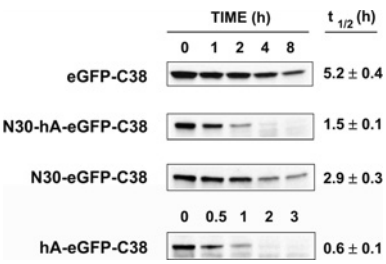

Figure 8 A 'cryptic' degron at the C-terminus of hTS
Plasmids expressing (see the text and Figure 1) were stably transfected into cell line RJK88.13 and treated with CHX. Decreases in protein concentrations over time were monitored by Western blotting (see the Experimental section for details). A representative blot is shown for each construct. The half-lives ($t_{\frac{1}{2}}$) are shown to the right of the corresponding blots, and are presented as means ± S.D.

required, we introduced proline substitutions at residues 13 and 14 of the first helix, resulting in LAC$^{P/P}$–eGFP (see Figure 1). PSIPRED correctly predicted the wild-type LAC region to have an α-helical conformation spanning residues 9–17, which was abolished by insertion of proline at sites 13 and 14 (results not shown). Figure 6(B) shows that LAC$^{P/P}$–eGFP exhibited a half-life of ~7 h, which is significantly longer than the parental LAC–eGFP molecule. Thus, as with hA of hTS, full degron function requires maintenance of a helical conformation.

Juxtaposition of the IDR and hA is not required for co-operative degron function

The hTS degron exerts its degradative function when placed at either the N-terminal or C-terminal end of the substrate [9]. Given its bipartite nature, it was of interest to ask if the two co-operating segments, i.e. the IDR and hA, must be juxtaposed. We generated and tested a fusion construct (denoted N30–eGFP–hA) having the IDR at the N-terminus and hA at the C-terminus (see Figure 1). As shown in Figure 7, N30–eGFP–hA had a half-life of 1.4 h, which is considerably shorter than that exhibited by the N30–hA–eGFP polypeptide. Thus, the two segments of the hTS degron co-operate even when placed at opposite ends of the target substrate. We also tested a protein (termed hA–eGFP–IDR; see Figure 1) having the opposite orientation, i.e. with hA at the N-terminus and the IDR at the C-terminus. Interestingly, hA–eGFP–N30 was very stable, exhibiting a half-life >24 h (Figure 7). Thus, it may be that the ability of the two degron components to co-operate when separated is position-dependent, although further studies are necessary to verify this interpretation.

The C-terminal end of TS contains a 'cryptic' degron

Previously, we observed that blocking the N-terminal amino group of hTS, either with a histidine-tag or by mutations that promote N-α-acetylation, results in a very stable polypeptide [8]. We interpreted this as indicating that initiation of proteasomal degradation occurs at the N-terminal end, and proceeds in the $N \rightarrow C$ direction [8]. Of interest was the finding that certain mutant hTS molecules (e.g. those with a P303L substitution) are stabilized by blocking the C-terminal end, suggesting that degradation of these mutant molecules may occur from the C-terminal end and proceed in a $C \rightarrow N$ direction [8,10]. Thus, the C-terminal region of hTS may contain a 'cryptic' degron that is activated in certain mutant polypeptides.

To test whether the C-terminal domain of hTS functions as a degron, we generated a polypeptide, termed eGFP–C38, containing the C-terminal 88 amino acids of hTS (residues 276–313) appended to eGFP (see Figure 1). The region includes a short, two-stranded β-sheet formed from residues 277–282 and 295–300. Figure 8 shows that the half-life of this molecule was 5.2 h, indicating that the hTS domain promotes degradation of the reporter, perhaps to a greater extent than the N-terminal degron (see Figure 4). Ligation of the N-terminal degron (i.e. residues 1–45) to eGFP–C38 results in a polypeptide, denoted N30–hA–eGFP–C38 (see Figure 1), that has a half-life of 1.5 h, indicating it to be even more unstable than eGFP–C38 (Figure 8). Thus, the C-terminal region of hTS functions as a degron, the activity of which is enhanced by the presence of the N-terminal degron at the other end of the substrate.

To determine if one or the other of the two subdomains comprising the N-terminal degron enhances the activity of the C38

element, we analysed polypeptides containing either the IDR or hA, neither of which is active on its own, appended to eGFP–C38. Polypeptide N30–eGFP–C38 (see Figure 1) exhibited a half-life of 2.9 h (Figure 8), indicating that the IDR modestly destabilizes eGFP–C38. Polypeptide hA–eGFP–C38 (see Figure 1) had a very short half-life of 0.6 h (Figure 8), indicating that hA induces potent destabilization of the substrate. We conclude that both the IDR and hA element, which by themselves are inactive in promoting degradation, promote the degron activity of the C38 element.

DISCUSSION

In earlier studies, we showed that the intracellular stability of TS is increased by the binding of inhibitory ligands [7], a finding that prompted an examination of the mechanism of TS degradation. We showed that proteolysis of the human enzyme (designated hTS) is carried out by the 26S proteasome, and occurs in a ubiquitin-independent manner. Furthermore, it is mediated by a portable, bipartite degron comprised of an IDR and an amphipathic α-helix located at the enzyme's N-terminal end [8–11]. A striking characteristic of the IDR is the occurrence of extensive sequence divergence among mammalian species [11]. Despite a high rate of amino acid substitution during evolution, the IDR has maintained its disordered character and its role in degron function. Thus, it is overall structure (or lack thereof) that is critical to degradation [9]. The specific amino acid sequence is less important, so long as it does not promote order.

It is well-recognized that IDRs typically evolve more rapidly than ordered regions, resulting in greater extents of amino acid substitution [30–32]; however, the forces driving such high substitution rates, specifically the role of positive against negative selection, are not well defined. Our analysis showed that the mean ratio of rates of non-synonymous to synonymous substitutions among pairwise comparisons of TS polypeptides from a variety of mammalian species is about 8-fold higher for the IDR as compared with the rest of the polypeptide; however, the ratio is still significantly <1 (Supplementary Table 1). Thus, although the disordered domain is more tolerant of amino acid substitutions than is the remainder of the molecule, it is still subject to negative selection. This conclusion is corroborated by the use of codon-based methods of analysis, which showed that the fraction of negatively selected codons is significantly reduced in the IDR as compared with the body of the polypeptide, and that few, if any, codons within the IDR are positively selected. In all, positive Darwinian selection appears to play little if any role in the high extents of amino acid variation within the IDR of mammalian TS.

Neither the IDR nor hA on its own is capable of promoting degradation (Figure 4). Interestingly, degron activity does not require that the two components be juxtaposed, in that they maintain the ability to co-operate when the IDR is at the N-terminus and hA is at the C-terminus (Figure 7). It is of interest that degron activity is lost when the two elements are in the opposite orientation, i.e. when the IDR is at the C-terminus and hA is at the N-terminus (Figure 7). This may indicate a position-dependency between the two segments when physically separated. However, other interpretations are possible. For example, it may be that when placed at the N-terminus, hA is blocked or modified in some fashion that inhibits its degradative role. The fact that the helical element promotes degradation of the newly identified 'cryptic' degron at the C-terminal end of TS (see Figure 8) suggests that it is functional. However, additional studies will be required to delineate whether or not the TS degron components do exhibit position-dependency when separated.

The 'cryptic' degron spanning residues 276–313 at the C-terminus of hTS (Figure 8) was predicted from the properties of certain mutant hTS molecules analysed in earlier studies [8,10]. This degron appears to be active only when large deletions or certain amino acid substitutions are introduced into the hTS polypeptide [8,10]. The region has a small β-sheet composed of two strands spanning residues 277–282 and 295–300. The importance of this structural element is unknown. The 'cryptic' degron may play a role in regulating hTS degradation under circumstances where the N-terminal degron is either blocked or inactivated. Definition of such circumstances and their physiological implications will require further studies.

It is formally possible that some of the unstable chimaeric proteins produced and examined in the present study have acquired a ubiquitin-dependent mode of degradation, resulting in shortened half-lives. We consider this unlikely, since we find no evidence of high-molecular-mass ubiquitinylated isoforms in any of the Western blots we have analysed. This holds true even when cells were grown in the presence of the proteasome inhibitor MG132, which generally causes accumulation of ubiquitinylated forms.

The requirement for disordered domains (and, in some cases, co-operating helical elements) may be a general theme among ubiquitin-independent substrates. Degradation of mammalian ODC is governed by an intrinsically disordered 37-residue region at the C-terminal end of the polypeptide [13,33]. Yeast ODC contains a disordered 45-residue domain at its N-terminal end that functions as a ubiquitin-independent degron and that is followed by a predicted α-helix [14]. The ubiquitin-independent degron of Rpn4, which is 80 amino acids in length, is located at the molecule's N-terminus and is predicted to consist of a 15–20-residue disordered domain followed by a structured region [20]. Finally, analysis of apomyoglobin degradation in a purified system *in vitro* has shown that a disordered element in the middle of the polypeptide, in co-operation with helical segments located at the N-terminus, is required for proteolytic breakdown [19]. In all, these studies indicate that ubiquitin-independent degradation requires flexible, disordered regions.

Our finding that TS proteolysis requires both a disordered and an ordered segment is consistent with the so-called 'two-component' model of proteasomal function [34,35]. In this model, degradation is viewed as occurring in two phases: proteasome recognition and binding followed by initiation of polypeptide insertion into the proteolytic chamber. For ubiquitin-dependent substrates, the first phase (recognition/binding) is

mediated by covalently attached polyubiquitin moieties that recognize one or more subunits of the proteasomal complex [35–37]. For ubiquitin-independent substrates, residues within the target polypeptide chain are utilized [13]. The second phase (initiation) involves unfolding of the proteasome-bound substrate, 'threading' of the substrate through the pore that leads to the proteolytic chamber, and entry into the chamber itself, where proteolysis ensues [33,35]. These steps are typically governed by disordered segments that are often (although not always) located at one or the other end of the protein [33,37]. With regard to TS, the helical hA segment may be responsible for proteasome binding, whereas the IDR guides substrate unfolding and chamber entry. Specific amino acids within each domain (e.g. Arg^{10}/Arg^{11} within the IDR, His^{39}/Arg^{42} within hA) represent critical residues that mediate the respective functions of the degron sub-elements.

Co-operation between disordered and helical segments, such as in TS, has been observed for the PUP (prokaryotic ubiquitin-like protein) of actinobacteria. Similar to ubiquitin, PUP ligation to proteins targets them for degradation by a proteasome-like complex [38]. Detailed structural studies have shown that upon ligation to a target substrate, PUP forms what is essentially a 'two-component' degron in which a helical C-terminal region binds regulatory subunits of the proteasome, whereas a disordered N-terminal region governs initiation of chamber entry [39]. Sequence alignments indicate that the helix-forming region of the PUP degron is more conserved among bacterial species than is the disordered region [40]. Such differential rates of evolution for disordered against the ordered degron sub-regions was also observed for TS in the present study.

AUTHOR CONTRIBUTION

Karen Barbour and Franklin Berger designed the research plan and analysed data; Karen Barbour and Yang-Yang Xing carried out the experiments; Edsel Peña carried out the statistical analysis of the data; and Franklin Berger wrote and prepared the paper with the assistance of Karen Barbour.

FUNDING

This work was supported by the National Institutes of Health [grant number CA44013, RR017698].

REFERENCES

1 Berger, F. G. and Berger, S. H. (2006) Thymidylate synthase as a chemotherapeutic drug target: where are we after fifty years? Cancer Biol. Ther. **5**, 1238–1241
2 Carreras, C. W. and Santi, D. V. (1995) The catalytic mechanism and structure of thymidylate synthase. Annu. Rev. Biochem. **64**, 721–762
3 Hardy, L. W., Finer-Moore, J. S., Montfort, W. R., Jones, M. O., Santi, D. V. and Stroud, R. M. (1987) Atomic structure of thymidylate synthase: target for rational drug design. Science **235**, 448–455
4 Phan, J., Steadman, D. J., Koli, S., Ding, W. C., Minor, W., Dunlap, R. B., Berger, S. H. and Lebioda, L. (2001) Structure of human thymidylate synthase suggests advantages of chemotherapy with noncompetitive inhibitors. J. Biol. Chem. **276**, 14170–14177
5 Schiffer, C. A., Clifton, I. J., Davisson, V. J., Santi, D. V. and Stroud, R. M. (1995) Crystal structure of human thymidylate synthase: a structural mechanism for guiding substrates into the active site. Biochemistry **34**, 16279–16287
6 Sotelo-Mundo, R. R., Ciesla, J., Dzik, J. M., Rode, W., Maley, F., Maley, G. F., Hardy, L. W. and Montfort, W. R. (1999) Crystal structures of rat thymidylate synthase inhibited by Tomudex, a potent anticancer drug. Biochemistry **38**, 1087–1094
7 Kitchens, M. E., Forsthoefel, A. M., Rafique, Z., Spencer, H. T. and Berger, F. G. (1999) Ligand-mediated induction of thymidylate synthase occurs by enzyme stabilization. Implications for autoregulation of translation. J. Biol. Chem. **274**, 12544–12547
8 Pena, M. M., Xing, Y. Y., Koli, S. and Berger, F. G. (2006) Role of N-terminal residues in the ubiquitin-independent degradation of human thymidylate synthase. Biochem. J. **394**, 355–363
9 Melo, S. P., Barbour, K. W. and Berger, F. G. (2011) Cooperation between an intrinsically disordered region and a helical segment is required for ubiquitin-independent degradation by the proteasome. J. Biol. Chem. **286**, 36559–36567
10 Melo, S. P., Yoshida, A. and Berger, F. G. (2010) Functional dissection of the N-terminal degron of human thymidylate synthase. Biochem. J. **432**, 217–226
11 Pena, M. M., Melo, S. P., Xing, Y. Y., White, K., Barbour, K. W. and Berger, F. G. (2009) The intrinsically disordered N-terminal domain of thymidylate synthase targets the enzyme to the ubiquitin-independent proteasomal degradation pathway. J. Biol. Chem. **284**, 31597–31607
12 Glickman, M. H. and Ciechanover, A. (2002) The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. Physiol. Rev. **82**, 373–428
13 Takeuchi, J., Chen, H., Hoyt, M. A. and Coffino, P. (2008) Structural elements of the ubiquitin-independent proteasome degron of ornithine decarboxylase. Biochem. J. **410**, 401–407
14 Godderz, D., Schafer, E., Palanimurugan, R. and Dohmen, R. J. (2011) The N-terminal unstructured domain of yeast ODC functions as a transplantable and replaceable ubiquitin-independent degron. J. Mol. Biol. **407**, 354–367
15 Chen, X., Barton, L. F., Chi, Y., Clurman, B. E. and Roberts, J. M. (2007) Ubiquitin-independent degradation of cell-cycle inhibitors by the REGgamma proteasome. Mol. Cell **26**, 843–852
16 Li, X., Amazit, L., Long, W., Lonard, D. M., Monaco, J. J. and O'Malley, B. W. (2007) Ubiquitin- and ATP-independent proteolytic turnover of p21 by the RE Ggamma-proteasome pathway. Mol. Cell **26**, 831–842
17 Truhlar, S. M., Mathes, E., Cervantes, C. F., Ghosh, G. and Komives, E. A. (2008) Pre-folding IkappaBalpha alters control of NF-κB signaling. J. Mol. Biol. **380**, 67–82
18 Mathes, E., Wang, L., Komives, E. and Ghosh, G. (2010) Flexible regions within IκBα create the ubiquitin-independent degradation signal. J. Biol. Chem. **285**, 32927–32936
19 Singh Gautam, A. K., Balakrishnan, S. and Venkatraman, P. (2012) Direct ubiquitin independent recognition and degradation of a folded protein by the eukaryotic proteasomes-origin of intrinsic degradation signals. PLoS ONE **7**, e34864
20 Ha, S. W., Ju, D. and Xie, Y. (2012) The N-terminal domain of Rpn4 serves as a portable ubiquitin-independent degron and is recognized by specific 19S RP subunits. Biochem. Biophys. Res. Commun. **419**, 226–231
21 Flicek, P., Amode, M. R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S. et al. (2012) Ensembl 2012. Nucleic Acids Res. **40**, D84–D90
22 Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. and Lopez, R. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. **38**, W695–W699

23 Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R. et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics **23**, 2947–2948
24 Nei, M. and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**, 418–426
25 Ota, T. and Nei, M. (1994) Variance and covariances of the numbers of synonymous and nonsynonymous substitutions per site. Mol. Biol. Evol. **11**, 613–619
26 Kosakovsky, Pond, S. L. and Frost, S. D. (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. **22**, 1208–1222
27 Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. **292**, 195–202
28 Shaw, D., Berger, F. G. and Spencer, H. T. (2001) Retroviral expression of *Escherichia coli* thymidylate synthase cDNA confers high-level antifolate resistance to hematopoietic cells. Hum. Gene Ther. **12**, 51–59
29 Nussbaum, R. L., Walmsley, R. M., Lesko, J. G., Airhart, S. D. and Ledbetter, D. H. (1985) Thymidylate synthase-deficient Chinese hamster cells: a selection system for human chromosome 18 and experimental system for the study of thymidylate synthase regulation and fragile X expression. Am. J. Hum. Genet. **37**, 1192–1205
30 Brown, C. J., Johnson, A. K. and Daughdrill, G. W. (2010) Comparing models of evolution for ordered and disordered proteins. Mol. Biol. Evol. **27**, 609–621
31 Huntley, M. and Golding, G. B. (2000) Evolution of simple sequence in proteins. J. Mol. Evol. **51**, 131–140
32 Tompa, P. (2003) Intrinsically unstructured proteins evolve by repeat expansion. Bioessays **25**, 847–855
33 Takeuchi, J., Chen, H. and Coffino, P. (2007) Proteasome substrate degradation requires association plus extended peptide. EMBO J. **26**, 123–131
34 Inobe, T., Fishbain, S., Prakash, S. and Matouschek, A. (2011) Defining the geometry of the two-component proteasome degron. Nat. Chem. Biol. **7**, 161–167
35 Schrader, E. K., Harstad, K. G. and Matouschek, A. (2009) Targeting proteins for degradation. Nat. Chem. Biol. **5**, 815–822
36 Elsasser, S. and Finley, D. (2005) Delivery of ubiquitinated substrates to protein-unfolding machines. Nat. Cell Biol. **7**, 742–749
37 Prakash, S., Tian, L., Ratliff, K. S., Lehotzky, R. E. and Matouschek, A. (2004) An unstructured initiation site is required for efficient proteasome-mediated degradation. Nat. Struct. Mol. Biol. **11**, 830–837
38 Burns, K. E. and Darwin, K. H. (2010) Pupylation: a signal for proteasomal degradation in mycobacterium tuberculosis. Subcell. Biochem. **54**, 149–157
39 Wang, T., Darwin, K. H. and Li, H. (2010) Binding-induced folding of prokaryotic ubiquitin-like protein on the *Mycobacterium* proteasomal ATPase targets substrates for degradation. Nat. Struct. Mol. Biol. **17**, 1352–1357
40 Sutter, M., Striebel, F., Damberger, F. F., Allain, F. H. and Weber-Ban, E. (2009) A distinct structural region of the prokaryotic ubiquitin-like protein (PUP) is recognized by the N-terminal domain of the proteasomal ATPase Mpa. FEBS Lett. **583**, 3151–3157

Received 30 October 2012/20 November 2012; accepted 22 November 2012

Published as Immediate Publication 26 November 2012, doi 10.1042/BSR20120112

Biosci. Rep. (2013) / **33** / art:e00015. / doi 10.1042/BSR20120112

OPEN ACCESS

Original Paper

SUPPLEMENTARY DATA

Characterization of the bipartite degron that regulates ubiquitin-independent degradation of thymidylate synthase

Karen W. BARBOUR*†, Yang-Yang XING*†, Edsel A. PEÑA*‡ and Franklin G. BERGER*†[1]

*Center for Colon Cancer Research, University of South Carolina, Columbia, SC 29208, U.S.A., †Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, U.S.A., and ‡Department of Statistics, University of South Carolina, Columbia, SC 29208, U.S.A.

See the following pages for Supplementary Figure S1 and Supplementary Table S1.

[1] To whom correspondence should be addressed (e-mail fgberger@mailbox.sc.edu).

```
                  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
HUMAN           1 ATGCCTGTGGCCGGCTCGGA-GCTG---CCGCGCCGGCCCTTGCCCCCCGCCGCACAGGA 56
CHIMPANZEE        ATGCCTGTGGCCGGCTCGGA-GCTG---CCGCGCCGGCCCTTGCCGCCCGCCGCACAGGA
ORANGUTAN         ATGCCCGTCGCCGGCTCCGA-GCTG---CCGCGCCGGCCCTTGCCGCCCTCCGCACAGGA
MACAQUE           ATGCCCGTCGCCGGCTCAGA-GCTG---CCGCGCCCGCCCTTGCAGCCCGCCTCACAGGA
MARMOSET          ATGCCCGTCGCCGGCTCCGA-GCTG---CCACGCCAGCCCTCGCAGCCCGCCGCCCAGGA
TARSIER           ATGCCCGCCGCCGGCTCCGA-GCTG---CAGCGCCCACCCTCGCCGCCCGCCGCGCAGGA
MOUSE             ATGCTGGTGGTTGGCTCCGA-GCTG---CAGT---------------CCGATGCTCAGCA
RAT               ATGCTGGTTGAAGGCTCTGA-GCTG---CAGT---------------CCGGTGCTCAACA
RABBIT            ATGCCCGCCGCCGGCTCTGA-GCTG---CCG---------TCGCCGCCCACGGCGCAGGA
PIKA              ATGCCGGCTGCCGGCTCCGA-GCTG---CCGCTCCCGCCTTCGCCGCCCACAGCACAGCA
COW               ATGCCAGCGGCCGGCTCCGA-GCCG---TCGCGCCCGCCGTCGCCGCCCGGCGTGCAGGA
DOLPHIN           ATGCCCGCCGCCGGCTCCGA-GCTG---CCGCGCCCGATGTCGCCGCCCGCCGCGCAGGA
HORSE             ATGCCAGCCTTCGGCTCCGA-GCTG---CAGCGCCCGCCGTCGCCGCCCGCCGCGCGGGA
DOG               ATGCCCGCCCCCGGCTCCGA-GCTG---CAGCGCCCGCCGTCGCCGCCCGCCGCGCAGAA
CAT               ATGCCCGCTCCCGGCTCCGA-GCTG---CAGCGCCCGCCGGCGCAGCCTGCCGAGCAGAA
GUINEA_PIG        ATGCCAGTTGCCGGCTCCGATGCCGGCTCCG-ACCTTCTGCCGCCGACTGCAGCACAGGA
                  ****  *     ***** ** ** *                      *       *   *


                  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
HUMAN          57 GCGGGACGCCGAGCCGCGTCCGCCG---------CACGGGGAGCTGCAGTACCTGGGGCA 107
CHIMPANZEE        GCGGGACGCCGAGCCGCGTCCGCCG---------CACGGGGAGCTGCAGTACCTGGGGCA
ORANGUTAN         GCGGGACGTCGAGCCGCGTCCGCCG---------CACGGGGAGCTGCAGTACCTGGGGCA
MACAQUE           GCGGGACGCCGAGCCACGGCCGCCG---------CACGGAGAACTGCAGTACCTGGGGCA
MARMOSET          GCTGGACGCCCAGCCGCGGCCGCCT---------CACGGGGAGCTGCAGTACCTGGGGCA
TARSIER           GCAGGGCGCCGAGCCGCGGCCGCCC---------CACGGGGAGCTGCAGTACCTGGAGCA
MOUSE             GCTGAGCGCGGAAGCCCCACGG------------CATGGAGAACTCCAGTACCTGAGGCA
RAT               GCCACGCACGGAAGCCCCGCAG------------CATGGAGAACTCCAGTACCTGAGGCA
RABBIT            GCAGGGCGCCGAGGCGCGGCCGCCC---------CACGGGGAGCTGCAGTACCTGGGGCA
PIKA              GCAGGGTACCGAGGCACGGCCGGCC---------CACGGGGAGCTGCAGTACCTGGGGCA
COW               GCAGAGCGCCGAGCCGCGGCCGCCGCCGCCGCCGCACGGGGAGCTGCAGTACCTGGGGCA
DOLPHIN           GCAGGGCGCCGAGTCGCGGCCGCCGCCGCCG---CACGGGGAGCTGCAGTACTTGGGGCA
HORSE             GCGGGGCTCGGAACCGCGGCCT------------CACGGGGAGCTGCAGTACCTGGGGCA
DOG               GCCGCGCGCC------CGTCCGCAGCCGCCTCCCCACGGGGAGCTGCAGTACCTGAGGCA
CAT               GCGGGGCGCCGAACCGCAGCCGCAGCCGCCACCCCACGGGGAGCTGCAGTACCTGGGGCA
GUINEA_PIG        CCCGGGCCCGGAGCCCTTGGCGCCG---------CATGGGGAGCTGCAGTACCTGCGGCA
                   *                                ** ** ** ** ****** **  ***
```

Figure S1 **Alignment of coding regions of mammalian TS mRNAs**

The amino acid coding region of the TS mRNA from each of 16 species was downloaded and aligned (see the Experimental section for details). Numbers refer to the human transcript, with the initiation codon set as 1. Completely conserved nucleotides are marked by stars (*). The region encoding the IDR, corresponding to nucleotides 1–81 of the hTS transcript, is indicated by a tilde (~).

```
HUMAN       108 GATCCAACACATCCTCCGCTGCGGCGTCAGGAAGGACGACCGCACGGGCACCGGCACCCT 167
CHIMPANZEE      GATCGAACACATCCTCCGCTGCGGCGTCAGGAAGGACGACCGCACGGGCACCGGCACCCT
ORANGUTAN       GATCGAGCACATCCTCCGCTGCGGCGTCAGGAAGGACGACCGCACGGGCACCGGCACCCT
MACAQUE         GATCGAGCACATCCTCCGCTGCGGCGTCAGGAAGGACGACCGCACGGGCACCGGCACGCT
MARMOSET        GATCGAGCACATCCTTCGCTGCGGCGTCAGGAAGGATGACCGCACCGGCACCGGGACCCT
TARSIER         GGTCGAGCACATCCTGCGCTGCGGCGTCCGGAAGGACGACCGCACAGGCGCCGGGACCCT
MOUSE           GGTGGAGCACATTTTGCGCTGCGGCTTCAAGAAGGAGGACCGCACGGGCACTGGCACCCT
RAT             GGTGGAGCACATTATGCGCTGCGGCTTCAAGAAGGAGGACCGCACGGGCACTGGCACCCT
RABBIT          GATCGAGCACATCCTGCGCTGCGGCTTCAGGAAGGAGGACCGCACGGGCACCGGCACCCT
PIKA            AATCGAGCAGATCCTGCGCTGCGGCTTTCGGAAGGAGGACCGCACGGGCACCGGCACCCT
COW             GATAGAACACATCCTCCGCTGCGGCTTCCGGAGGGATGACCGCACCGGCACTGGCACCCT
DOLPHIN         GATCGAGCACATTCTCCGCTGCGGCTTTCGGAAGGATGACCGCACAGGCACCGGCACCCT
HORSE           GATTGAGCACATCCTGCGCTCCGGCTTCCGGAAGGAGGACCGCACGGGCACCGGCACCCT
DOG             GGTGGAGCACATCCTGAGCTGCGGGGCCCGGAAGCACGACCGCACGGGCACGGGCACGCT
CAT             AGTGGAGCACATCCTGCGCTGCGGGTTCCAGAAGGATGACCGCACCGGCACCGGCACGCT
GUINEA_PIG      GGTGGAGCACATTCTGCGTAGCGGCTTCCACAAGGAGGACCGCACCGGCACCGGCACGCT
                  *  * ** **  *  *   ***       * * * ******** *** * ** ** **

HUMAN       168 GTCGGTATTCGGCATGCAGGCGCGCTACAGCCTGAGAGATGAATTCCCTCTGCTGACAAC 227
CHIMPANZEE      GTCGGTATTCGGCATGCAGGCGCGCTACAGCCTGAGAGATGAATTTCCTCTGCTGACAAC
ORANGUTAN       GTCGGTATTCGGCATGCAGGCGCGCTACAGCCTGAGAGATGAATTTCCTCTGCTGACAAC
MACAQUE         GTCGGTATTCGGCATGCAGGCGCGCTACAGCCTGAGAGATGAATTTCCTCTGCTGACAAC
MARMOSET        GTCGGTGTTTGGCATGCAGGCGCGCTACAGCCTGAGAGATGAATTTCCTCTGCTGACAAC
TARSIER         GTCGGTGTTCTGCATGCGGGCCCGCTACGACCTGAGAGGTGAATTTCCTCTGCTGACAAC
MOUSE           GTCGGTGTTCGGCATGCAGGCACGATACAGCCTGAGAGATGAATTTCCTCTGCTCACAAC
RAT             GTCGGTGTTCGGCATGCAGGCACGGTACAGCCTGAGAGATGAATTTCCTCTGCTCACAAC
RABBIT          GTCGGTGTTCGGCATGCAGGCGCGCTACAGCCTGAGAGATGAATTTCCTCTGCTGACAAC
PIKA            GTCGGTGTTCGGCATGCAGGCACGCTACAACCTGAGAGATGAATTTCCTCTGCTGACAAC
COW             GTCGGTGTTCGGGATGCAGGCGCGGTACAACTTGAGAGATGAATTTCCTCTGCTGACAAC
DOLPHIN         GTCGGTGTTCGGCATGCAGGCGCGCTACAGCCTGAGAGATGAATTTCCTCTGCTGACAAC
HORSE           GTCGGTGTTCGGCCTGCAGGCGCGCTACAACCTGAGAGATGAATTTCCTCTGCTGACAAC
DOG             GTCGGTGTTCGGCATGCAGGCGCGCTACAGCCTGAGAGATGAATTTCCTCTGCTGACAAC
CAT             GTCGGTGTTCGGCATGCAGGCGCGCTACAGCCTGAGAGATGAATTTCCTTTACTGACAAC
GUINEA_PIG      GTCTGTGTTCGGTATGCAGGCGCGCTACAGCCTACGAGATCAATTTCCCCTGCTGACGAC
                *** ** **  *  *** *** ** ***  * *  *** * **** **  * ** ** **

HUMAN       228 CAAACGTGTGTTCTGGAAGGGTGTTTTGGAGGAGTTGCTGTGGTTTATCAAGGGATCCAC 287
CHIMPANZEE      CAAACGTGTGTTCTGGAAGGGTGTTTTGGAGGAGTTGCTGTGGTTTATCAAGGGATCCAC
ORANGUTAN       CAAACGTGTGTTCTGGAAGGGTGTTTTGGAGGAGTTGCTGTGGTTTATCAAGGGATCCAC
MACAQUE         CAAACGTGTGTTCTGGAAGGGTGTTTTGGAGGAGTTGCTGTGGTTTATCAAGGGATCTAC
MARMOSET        CAAACGTGTGTTCTGGAAGGGTGTTTTGGAGGAGTTGCTGTGGTTTATCAAGGGATCCAC
TARSIER         CAAGCGTGTATTCTGGAAGGGTGTTTTGGAGGAGTTACTGTGGTTTATCAAGGGATCCAC
MOUSE           CAAACGAGTGTTCTGGAAGGGTGTTTTGGAGGAGTTGTTGTGGTTTATCAAGGGATCCAC
RAT             CAAACGAGTATTCTGGAAGGGTGTTTTGGAGGAGTTGTTGTGGTTTATTAAGGGCTCCAC
RABBIT          CAAACGTGTATTCTGGAAGGGTGTCTTGGAGGAGTTGCTGTGGTTTATCAAGGGATCCAC
PIKA            CAAACGTGTATTCTGGAAGGGTGTCTTGGAGGAGCTGCTGTGGTTTATCAAGGGATCCAC
COW             CAAACGTGTTTTCTGGAAAGGTGTTTTGGAGGAGTTGCTGTGGTTTATCAAGGGATCCAC
DOLPHIN         CAAACGTGTTTTCTGGAAGGGTGTTTTGGAGGAGTTGCTGTGGTTTATCAAGGGATCCAC
HORSE           CAAACGTGTTTTCTGGAAGGGTGTCTTGGAAGAGTTGCTGTGGTTTATCAAGGGATCCAC
DOG             CAAACGTGTATTTTGGAAGGGTGTTTTGGAGGAGTTGCTGTGGTTTATCAAGGGATCCAC
CAT             CAAGCGTGTATTCTGGAAGGGTGTTTTGGAGGAGTTGCTGTGGTTTATCAAGGGATCTAC
GUINEA_PIG      CAAGCGTGTGTTTTGGAAGGGTATTCTGGAGGAGCTGCTGTGGTTTATTAAGGGATCTAC
                *** ** ** ** ***** *** *  **** *** *  ********** ***** ** **
```

Figure S1 Continued

```
HUMAN        288 AAATGCTAAAGAGCTGTCTTCCAAGGGAGTGAAAATCTGGGATGCCAATGGATCCCGAGA 347
CHIMPANZEE       AAATGCTAAAGAGCTGTCTTCCAAGGGAGTGAAAATCTGGGATGCCAATGGATCCCGAGA
ORANGUTAN        AAATGCTAAAGAGCTGTCTTCCAAGGGAGTGAAAATCTGGGATGCCAATGGATCCCGAGA
MACAQUE          AAATGCTAAAGAGCTGTCTTCCAAGGGAGTGAAAATCTGGGATGCCAATGGATCCCGAGA
MARMOSET         AAATGCTAAAGAGCTGTCTTCCAAGGGAGTGAAAATCTGGGATGCCAATGGATCCCGAGA
TARSIER          AAATGCTAAAGAACTGTCTTCTAAGGGAGTGAAAATCTGGGATGCCAATGGATCCCGAGA
MOUSE            AAATGCTAAAGAATTGTCCTCAAAGGGAGTGAGAATCTGGGATGCCAATGGATCCCGAGA
RAT              AAATGCTAAAGAACTGTCCTCCAAGGGAGTGAGAATTTGGGATGCCAATGGGTCCCGAGA
RABBIT           AAATGCTAAAGAACTGTCTTCCAAGGGAGTGAAAATCTGGGATGCCAATGCATCCCGAGA
PIKA             AAATGCTAAAGAACTGTCTGCCAAGGGAGTGAAAATCTGGGATGCGAACGCTTCCCGAGA
COW              CAACGCTAAGGAACTCTCTTCCAAGGGAGTGAAAATTTGGGATGCCAATGGGTCCCGAGA
DOLPHIN          AAACGCTAAGGAGCTGTCTTCCAAGGGAGTGAAAATCTGGGATGCTAATGGGTCCCGAGA
HORSE            AAATGCTAATGAACTGTCTTGCAAGGGAGTGAAAATCTGGGATGCCAATGGGTCCCGAGA
DOG              AAACGCTAAGGAACTGTCTTCCAGGGGAGTGAAAATCTGGGATGCCAATGGGTCTCGAGA
CAT              AAACGCTAAGGAACTGTCATCCAAGGGAGTGAAAATCTGGGATGCCAATGGGTCCCGAAA
GUINEA_PIG       AAATGCCAAAGAACTGTCATCCAAGGGCGTGAAAATTTGGGATGCCAATGGGTCTCGAGA
                  ** ** ** **  * **    * *** **** *** ******** ** *  ** *** *

HUMAN        348 CTTTTTGGACAGCCTGGGATTCTCCACCAGAGAAGAAGGGGACTTGGGCCCAGTTTATGG 407
CHIMPANZEE       CTTTTTGGACAGCCTGGGATTCTCCACCAGAGAAGAAGGGGACTTGGGCCCAGTTTATGG
ORANGUTAN        CTTTTTGGACAGCCTGGGATTCTCCACCAGAGAAGAAGGGGACTTGGGCCCAGTTTATGG
MACAQUE          CTTTTTGGACAGCCTGGGGTTCTCCGCCAGAGAAGAAGGGGACTTGGGTCCAGTTTATGG
MARMOSET         CTTTCTGGATAGCTTGGGATTTTCCGCCAGAGAAGAAGGGGATTTGGGCCCAGTTTATGG
TARSIER          CTTTTTGGACTGCCTAGGATTCTCCACCAGAGAAGAAGGAGATTTGGGCCCTGTCTATGG
MOUSE            TTTTCTGGACAGCTTGGGATTTTCTGCCCGACAGGAAGGGGACCTGGGCCCAGTTTATGG
RAT              CTTTTTGGACAGCTTGGGATTCTCTGCCCGACAGGAAGGAGACCTGGGCCCAGTTTATGG
RABBIT           CTTTTTGGACAGCCTGGGATTCTCCACCAGACAAGAAGGGGACCTGGGTCCAGTTTATGG
PIKA             ATTTCTGGACAGCCTTGGATTCTCAACCAGAGAGGAGGGGGACCTGGGTCCAGTTTATGG
COW              CTTCTTGGATGGCCTGGGCTTCTCCGACAGAGCTGAAGGGGATTTAGGCCCAGTTTATGG
DOLPHIN          CTTCTTGGACAGCCTGGGATTCTCCACCAGAGCAGAAGGGGATTTAGGCCCAGTTTATGG
HORSE            CTTCTTGGACAGCCTGGGATTCTCCGCCAGAGAAGAAGGGGATTTAGGCCCAGTTTACGG
DOG              CTTTTTGGACAGCCTGGGATTCGCCAACCGAGAAGAAGGGGATTTAGGCCCAATTTATGG
CAT              CTTCTTGGACAGCCTAGGATTCTCCAACAGAGAAGAAGGGGATTTAGGCCCAGTTTATGG
GUINEA_PIG       CTTCCTGGACAGCCTGGGGTTTTCTAGCCGCCAGGAAGGGGACCTGGGCCCCATTTATGG
                  **  ****  ** * ** **  *   * *    ** ** **  * ** **  * ** **

HUMAN        408 CTTCCAGTGGAGGCATTTTGGGGCAGAATACAGAGATATGGAATCAGATTATTCAGGACA 467
CHIMPANZEE       CTTCCAGTGGAGGCATTTTGGGGCAGAATACAGAGATATGGAATCAGATTATTCAGGACA
ORANGUTAN        CTTCCAGTGGAGGCATTTTGGGGCAGAATACAGAGATATGGAATCAGATTATTCAGGACA
MACAQUE          CTTCCAGTGGAGGCATTTTGGGGCAGAATACAGAGATATGGAATCAGATTATTCAGGACA
MARMOSET         CTTCCAGTGGAGGCATTTTGGGGCAGAATACAGAGATATGGATTCAGATTATTCAGGACA
TARSIER          CTTCCAATGGAGGCATTTTGGGGCAGAATACAAAGATATGAATTCAGATTATTCAGGTCA
MOUSE            TTTCCAATGGAGGCATTTTGGAGCAGAGTACAAAGATATGGATTCAGATTACTCGGGACA
RAT              ATTCCAGTGGAGACATTTTGGAGCAGACTACAAAGATATGGATTCAGATTACTCGGGTCA
RABBIT           CTTCCAGTGGAGGCATTTTGGTGCAGAATACAAAGATAAGGACTCAGATTATTCAGGTCA
PIKA             CTTCCAGTGGAGGCATTTTGGGGCAGAATACAAAGATAAAGATTCTGATTACTCCGGCCA
COW              CTTCCAGTGGAGGCATTTTGGGGCTGAATACAAAGATATGGATTCAGAGTATTCAGGTCA
DOLPHIN          CTTTCAGTGGAGGCATTTTGGGGCCGAATACAAAGAGATGGATTCAGATTATTCAGATAA
HORSE            CTTTCAGTGGAGGCATTTTGGGGCAGAATACAAAGATATGGATTCAGATTATTCAGGTCA
DOG              CTTTCAGTGGAGGCATTTTGGGGCAGAATACAAAGATAAGGATTCAGATTATTCAGGTCA
CAT              CTTTCAGTGGAGGCATTTTGGGGCAGAATACAAAGATAAGGATTCAGATTATTCAGGTCA
GUINEA_PIG       CTTCCAGTGGAGGCATTTTGGGGCTGAGTACAAAGATATGGACTCAGATTACTCAGGTCA
                  ** ** ***** ******** ** ** **** *** *   * ** ** ** ** *   *
```

Figure S1 Continued

```
HUMAN        468 GGGAGTTGACCAACTGCAAAGAGTGATTGACACCATCAAAACCAACCCTGACGACAGAAG 527
CHIMPANZEE       GGGAGTTGACCAGCTGCAAAGAGTGATTGACACCATCAAATCCAACCCTGACGACAGAAG
ORANGUTAN        GGGAGTTGACCAACTGCAAAGAGTGATTGACACCATCAAAACGAACCCTGACGACAGAAG
MACAQUE          GGGAGTTGACCAACTACAAAGAGTGATTGACACCATCAAAACCAACCCTGACGACAGAAG
MARMOSET         AGGAGTTGACCAACTGCAAAGAGTGATTGACACCATCAAAACCAACCCTGATGACAGAAG
TARSIER          AGGAGTTGACCAGCTCCAAAAGGTGATTGACACCATCAAAACCAACCCTGATGACAGAAG
MOUSE            AGGAGTAGACCAGCTGCAAAAAGTGATTGACACCATCAAAACCAACCCTGATGACAGAAG
RAT              AGGAGTAGACCAGCTGCAAAAAGTGATTGACACCATCAAAACCAACCCCGATGACAGAAG
RABBIT           AGGAGTAGACCAACTGCAGAAGGTGATCGATACCATCAAAACCAACCCTGATGACAGAAG
PIKA             AGGAGTAGACCAGCTGCAAAAAGTGATCGATACCATCAAAACCAACCCTGATGACCGCAG
COW              AGGAGTAGATCAACTGCAAAAAGTGATCGACACAATCAAAACCAACCCTAACGACAGAAG
DOLPHIN          AGGAGTAGATCAACTGCAAAAAGTGATTGACACAATCAAAACCAACCCTGACGACAGAAG
HORSE            AGGAGTAGACCAACTGCAAAAAGTGATTAACACAATCAAAACCAACCCTGACGACAGAAG
DOG              AGGAGTCGACCAGCTGCAAAAAGTGATTGACACAATCAAAACCAACCCTGACGACAGAAG
CAT              AGGAGTAGATCAACTGCAAAAGGTGATTGACACAATCAAAACCAACCCCAATGACAGAAG
GUINEA_PIG       AGGAGTTGACCAGCTGCAGAAAGTGATTGACACCATCAAAACCAACCCTGATGACAGAAG
                  ***** ** ** ** ** *  *****  * ** ****** * *****  * *** * **

HUMAN        528 AATCATCATGTGCGCTTGGAATCCAAGAGATCTTCCTCTGATGGCGCTGCCTCCATGCCA 587
CHIMPANZEE       AATCATCATGTGCGCTTGGAATCCAAGAGATCTTCCTCTGATGGCGCTGCCTCCATGCCA
ORANGUTAN        AATCATCATGTGCGCTTGGAATCCAAGAGATCTTCCTCTGATGGCGCTGCCTCCATGCCA
MACAQUE          AATCATCATGTGTGCTTGGAATCCAAGAGATCTTCCTCTGATGGCACTGCCTCCATGCCA
MARMOSET         AATCATCATGTGCGCTTGGAATCCAAGAGATCTTCCTCTGATGGCGCTGCCTCCATGCCA
TARSIER          GATCATCATGTGTGCTTGGAATCCAAAAGATCTTCCTCTGATGGCGCTGCCTCCATGCCA
MOUSE            AATCATCATGTGTGCCTGGAACCCAAAAGATCTTCCCCTGATGGCACTGCCTCCTTGCCA
RAT              AATCATCATGTGTGCCTGGAACCCAAAAGATCTTCCCCTGATGGCACTGCCTCCTTGCCA
RABBIT           AATCATCATGTGTGCCTGGAATCCCAAAGACCTTCCTCTGATGGCGCTGCCCCCGTGCCA
PIKA             AATCATCATGTGTGCCTGGAATCCTAAAGACCTTCCTCAGATGGCCCTGCCCCCATGCCA
COW              AATCATCCTGTGTGCTTGGAATCCAAAAGATCTGCCTCTCATGGCCCTCCCCCCATGCCA
DOLPHIN          AATCATCCTGTGTGCTTGGAATCCAAAAGATGTCCCTCTCATGGCCCTACCCCCGTGCCA
HORSE            AATCATCCTGTGTGCTTGGAATCCAAAAGATCTTCCTCTCATGGCTCTGCCTCCCTGCCA
DOG              AATTATTCTGTGTGGTTGGAATCCAAAAGATCTTCCTCTGATGGCCCTACCTCCGTGCCA
CAT              AATCATCCTGTGTGGTTGGAATCCAAAAGATCTTCCTCTGATGGCCCTGCCTCCGTGCCA
GUINEA_PIG       AATTATCATGTGTGCCTGGAACCCGAAAGATCTTCCTCTGATGGCGCTGCCTCCCTGTCA
                  ** **  **** *  ***** ** * ***  * ** *  ***** ** ** ** ** **

HUMAN        588 TGCCCTCTGCCAGTTCTATGTGGTGAACAGTGAGCTGTCCTGCCAGCTGTACCAGAGATC 647
CHIMPANZEE       TGCCCTCTGCCAGTTCTATGTGGTGAACAGTGAGCTGTCCTGCCAGCTGTACCAGAGATC
ORANGUTAN        TGCCCTCTGCCAGTTCTATGTGGTGAACAGTGAGCTGTCCTGCCAGCTGTACCAGAGATC
MACAQUE          TGCTCTCTGCCAGTTCTATGTGGTGAACAGTGAGCTGTCCTGCCAGCTGTACCAGAGATC
MARMOSET         TGCCCTCTGCCAGTTCTATGTGGTGAACGGTGAGCTGTCCTGCCAGCTGTACCAGAGATC
TARSIER          TGCCCTCTGCCAGTTCTACGTGGTGAACGGTGAGCTGTCCTGCCAGCTGTACCAGAGGTC
MOUSE            TGCCCTCTGTCAGTTCTATGTGGTGAATGGGGAACTGTCTTGCCAGCTTTACCAGAGGTC
RAT              TGCCCTCTGTCAATTTTATGTGGTGAATGGGGAGCTGTCTTGCCAGCTTTACCAGCGGTC
RABBIT           TGCCCTCTGCCAGTTCTACGTGGTGAATGGGGAGCTGTCCTGCCAGCTGTACCAGAGGTC
PIKA             TGCTCTCTGCCAGTTCTACGTAGTGAACGGCGAGCTGTCCTGCCAGTTGTACCAGAGGTC
COW              CGCCCTCTGCCAGTTCTACGTGGTGAATGGGGAGTTGTCCTGCCAGTTGTACCAGCGGTC
DOLPHIN          TGCCCTTTGCCAGTTCTACGTGGTGAACGGTGAGCTGTCCTGCCAGCTGTACCAGAGGTC
HORSE            TGCCCTCTGCCAGTTCTACGTGGTGAACGGTGAGCTGTCCTGCCAGCTGTACCAGAGATC
DOG              TGCCCTCTGCCAGTTCTATGTGGTGAACGGTGAGCTGTCCTGCCAGCTATACCAGAGGTC
CAT              TGCTCTTTGCCAGTTCTACGTGGTGAATGGTGAACTGTCCTGCCAGCTGTACCAGAGGTC
GUINEA_PIG       CACCCTCTGCCAGTTCTACGTGGTGAACGGGGAATTGTCCTGTCAGCTGTACCAGCGCTC
                   * ** ** ** ** ** ** *****  * **  **** ** *** * ****** * **
```

Figure S1 Continued

```
HUMAN        648 GGGAGACATGGGCCTCGGTGTGCCTTTCAACATCGCCAGCTACGCCCTGCTCACGTACAT 707
CHIMPANZEE       GGGAGACATGGGCCTCGGTGTGCCTTTCAACATCGCCAGCTACGCCCTGCTCACGTACAT
ORANGUTAN        GGGAGACATGGGCCTAGGTGTGCCTTTCAACATCGCCAGCTACGCCCTGCTCACGTACAT
MACAQUE          GGGAGACATGGGCCTAGGTGTGCCTTTCAACATCGCCAGCTACGCCCTGCTCACCTACAT
MARMOSET         GGGAGACATGGGCCTAGGTGTGCCTTTCAACATCGCCAGCTACGCCCTGCTCACATACAT
TARSIER          AGGAGACATGGGCCTGGGTGTACCTTTCAACATCGCCAGCTATGCCCTGCTCACATACAT
COW              GGGGGACATGGGCCTGGGTGTGCCCTTCAACATTGCCAGCTACGCCCTGCTCACCTACAT
MOUSE            AGGAGATATGGGTCTGGGCGTGCCCTTCAACATTGCCAGCTATGCTCTGCTCACCTACAT
RAT              AGGAGATATGGGTCTGGGTGTGCCCTTCAACATTGCCAGCTATGCTCTGCTGACCTACAT
RABBIT           GGGAGACATGGGCCTGGGCGTGCCCTTCAACATCGCCAGCTATGCCCTGCTCACCTACAT
PIKA             GGGAGACATGGGCCTGGGCGTGCCCTTCAACATTGCCAGCTACGCCCTGCTCACCTACAT
DOLPHIN          AGGAGACATGGGCCTGGGTGTGCCCTTTAACATCGCCAGCTACGCCCTGCTCACCTACAT
HORSE            GGGAGACATGGGCCTGGGTGTGCCCTTCAACATTGCCAGCTACGCCCTGCTCACCTACAT
DOG              AGGAGACATGGGGCTGGGCGTGCCCTTCAACATCGCCAGCTACGCCCTGCTCACCTACAT
CAT              AGGAGACATGGGCCTGGGTGTACCCTTCAACATCGCCAGCTACTCCCTGCTCACCTACAT
GUINEA_PIG       TGGGGACATGGGCCTGGGCGTGCCCTTCAACATTGCCAGCTACGCCCTGCTCACCTACAT
                  ** ** ***** ** ** ** ** ** ***** ********  * ***** ** *****

HUMAN        708 GATTGCGCACATCACGGGCCTGAAGCCAGGTGACTTTATACACACTTTGGGAGATGCACA 767
CHIMPANZEE       GATTGCGCACATCACGGGCCTGAAGCCAGGTGACTTTATACACACTTTGGGAGATGCACA
ORANGUTAN        GATTGCGCACATCACGGGCCTGAAGCCAGGTGACTTTATACACACTTTGGGAGATGCACA
MACAQUE          GATTGCGCACATCACGGGCCTGAAGCCAGGTGATTTTGTACACACTTTGGGAGATGCACA
MARMOSET         GATCGCGCACATCACGGGCCTGAAGCCAGGTGACTTTGTACATACTTTGGGAGATGCACA
TARSIER          GATCGCACACATCACAGGCCTGAAGCCAGGTGACTTTGTACATACTTTGGGAGATGCACA
MOUSE            GATTGCACATATCACAGGCCTGCAGCCAGGTGATTTTGTCCACACTTTGGGAGATGCACA
RAT              GATTGCACATATCACGGGCCTGCAGCCGGGTGATTTTGTCCATACTTTGGGAGATGCACA
RABBIT           GATCGCGCACGTCACCGGCCTGAAGCCAGGTGATTTTGTACATACTTTGGGAGATGCTCA
PIKA             GATTGCGCATGTCACTGGACTGAAGCCAGGTGACTTTGTACACACTCTGGGAGATGCACA
COW              GATCGCACACATCACGGACCTGAAGCCAGGTGACTTCGTGCACACTCTGGGAGATGCACA
DOLPHIN          GATCGCACACATCACGGGCCTGAAGCCAGGTGACTTTGTACACACTTTGGGAGATGCACA
HORSE            GATCGCACACATCACGGGCCTGAAGCCGGGTGACTTTGTGCACACTTTGGGAGATGCACA
DOG              GATCGCACACATCACAGGCCTGAAGCCAGGTGACTTCATACATACTTTGGGAGATGCGCA
CAT              GATCGCACACATCACAGGCCTGAAGCCAGGTGACTTCATACATACTTTGGGAGATGCGCA
GUINEA_PIG       GATTGCGCACGTCACGGGCCTGCAGCCAGGTGACTTTGTGCATACACTGGGAGACGCACA
                 *** ** **  **** *  *** **** ***** **  * ** **  ******* ** **

HUMAN        768 TATTTACCTGAATCACATCGAGCCACTGAAAATTCAGCTTCAGCGAGAACCCAGACCTTT 827
CHIMPANZEE       TATTTACCTGAATCACATCGAGCCACTGAAAATTCAGCTTCAGCGAGAACCCAGACCTTT
ORANGUTAN        TATTTACCTGAATCACATTGAGCCACTGAAAATTCAGCTTCAGCGAGAACCCAGACCTTT
MACAQUE          TATTTACCTGAATCACATCGAGCCACTGAAAATTCAGCTTCAGCGAGAACCCAGACCTTT
MARMOSET         TATTTACCTGAATCACATCGAGATTTTAAAAGTTACTCTTCAGCGAGAACCCAGACCTTT
TARSIER          TATTTACCTGAATCACATTGAGCCACTGAAAACTCAGCTTCAGAGAGAACCAAGACCCTT
MOUSE            TATTTACCTGAATCATATAGAGCCGCTGAAAATTCAGCTACAGCGAGAACCAAGACCTTT
RAT              CATTTATCTGAATCATATTGAGCCACTGAAAATTCAGCTACAGCGAGAACCAAGACCTTT
RABBIT           TATTTACCTGAACCACATTGAGCCTCTGAAAACTCAGCTTCAGCGAGAACCAAGACCTTT
PIKA             TGTTTACCTGAATCACATCGAGCCACTCAAAGTGCAGCTTCAACGAGAGCCAAGACCTTT
COW              CATTTACCTGAATCACATCGAGCCACTGAAAACTCAGCTGCAGCGAGAACCAAGGCCTTT
DOLPHIN          CATTTACCTGAATCACATTGAGCCGCTGAAAATTCAGCTTCAGCGAGAACCAAGGCCTTT
HORSE            TATTTACCTGAATCACATTGACCCGCTGAAAACGCAGCTTCAGCGAGAACCAAGGCCTTT
DOG              TATTTACCTGAATCACATTGAGCCGCTGAAAATACAGCTGCAGCGAGAACCAAGGCCTTT
CAT              TATTTACCTGAATCACATTGAACCGCTAAAAATGCAGCTCCAGCGAGAACCAAGGCCTTT
GUINEA_PIG       CATTTACCTGAATCACATTGAGCCGCTGAAAACTCAGCTTCAACGAGAACCAAGACCTTT
                   **** ***** ** ** **     * ***      ** **  **** ** ** ** **
```

Figure S1 Continued

```
HUMAN       828 CCCAAAGCTCAGGATTCTTCGAAAAGTTGAGAAAATTGATGACTTCAAAGCTGAAGACTT 887
CHIMPANZEE      CCCAAAGCTCAAGATTCTTCGAAAAGTTGAGAAAATTGATGACTTCAAAGCTGAAGACTT
ORANGUTAN       CCCAAAGCTCAAGATTCTTCGAAAAGTTGAGAAAATTGATGACTTCAAAGCTGAAGACTT
MACAQUE         CCCAAAGCTCAAAATTCTCCGAAAAGTTGAGAAAATTGATGACTTCAAAGCTGAAGACTT
MARMOSET        TCCAAAGCTCAAGATTCTTCGAAAAGTTGAGAAAATTGATGACTTCAAAGCTGAAGACTT
TARSIER         CCCGAAGCTCAAAATCCTTCGAAAAGTTGAGAAAATTGATGATTTCAAAGCTGAAGACTT
MOUSE           CCCAAAGCTCAAAATCCTTCGAAAAGTTGAGACAATCGATGATTTCAAAGTTGAAGACTT
RAT             CCCAAAGCTCAGAATCCTCCGAAAAGTTGAGACAATCGATGATTTCAAAGTTGAAGACTT
RABBIT          CCCAAAGCTCAAAATTCTTCGAAAAGTTGAGACGATTGATGATTTCAAAGCTGAAGATTT
PIKA            CCCAAAGCTCAAAATCCTTCGAAAAGTTGAGACGATTGACGATTTCAAAGCTGAAGATTT
COW             CCCCAAGCTCAAAATCCTTCGAAAAGTTGAGACAATCGATGACTTCCAGGCCGAAGACTT
DOLPHIN         CCCCAAGCTCAAAATACTTCGAAAAGTTGAGAAAATTGATGACTTCAAGGCTGAAGACTT
HORSE           CCCGAAACTCAAAATCCTTCGAAAAGTTGAGACAATTGATGACTTCAAGGCTGAAGACTT
DOG             CCCGAAGCTCAGAATCCTTCGAAGAGTTGAGAAAATTGATGACTTCAAGGCTGAAGACTT
CAT             CCCAAAGCTCAGAATCCTTCGAAAAGTTGAGACAATTGATGACTTCAAGGCTGAAGACTT
GUINEA_PIG      CCCAAAGCTCAAAATCGTTCGAAAAGTTGAGGCAATTGATGACTTCACAGCTGACGACTT
                 ** ** ****  **  * **** *******   ** ** ** ***   *  ** ** **

HUMAN       888 TCAGATTGAAGGGTACAATCCGCATCCAACTATTAAAATGGAAATGGCTGTT 939
CHIMPANZEE      TCAGATTGAAGGGTACAATCCACATCCAACTATTAAAATGGAAATGGCTGTT
ORANGUTAN       TCAGATTGAAGGGTACAATCCACATCCAACTATTAAAATGGAAATGGCTGTT
MACAQUE         TCAGATTGAAGGGTACAATCCGCATCCAACTATTAAAATGGAAATGGCTGTT
MARMOSET        TAAGGTTGAAGGGTACAATCCGCATCCAACTATTAAAATGGAAATGGCTGTT
TARSIER         CCAGATTGAAGGGTACAATCCGCATCCAACTATTAAAATGGAAATGGCTGTT
MOUSE           TCAGATTGAAGGGTATAATCCACATCCAACGATTAAAATGGAAATGGCTGTT
RAT             TCAGATTGAAGGGTATAATCCACATCCAACGATTAAAATGGAAATGGCTGTT
RABBIT          TCAGCTTGAAGGGTACAACCCACATCCAACTATTAAAATGGAGATGGCTGTT
PIKA            TCAGATTGAAGGGTATAACCCACATCCCACCATTAAAATGGAAATGGCTGTT
COW             TCAGATTGAAGGCTACAATCCTCATCCGACTATTAAAATGGAAATGGCTGTC
DOLPHIN         TCAGATTGAAGGCTACAATCCTCACCCGACCATTAAAATGGAAATGGCTGTT
HORSE           TCAGCTTGAAGGGTACCATCCACACCCGACCATCAGAATGGAAATGGCTGTT
DOG             TCAGATCGAGGGGTACAATCCTCACCCAACTATTAAAATGGAAATGGCTGTG
CAT             TCAGATTGAGGGGTACAATCCTCACCCCACTATTAAAATGGATATGGCTGTT
GUINEA_PIG      CCGGCTGGAAGGGTACAATCCACATCCTGCCATTAAGATGGAAATGGCTGTG
                   * * ** ** **  * ** ** **  * ** *  ***** ********
```

Figure S1 Continued

Table S1 Synonymous (d_S) and nonsynonymous (d_N) nucleotide substitutions for the IDR and the body of the TS polypeptide
Values for each of the 120 pairwise comparisons among 16 mammalian species were determined, and d_N/d_S ratios were calculated.

		IDR			Body		
Species 1	**Species 2**	d_N	d_S	d_N/d_S	d_N	d_S	d_N/d_S
Human	Chimpanzee	0.0000	0.0441	0.0000	0.0045	0.0159	0.2830
	Orang-utan	0.0355	0.1946	0.1824	0.0030	0.0321	0.0935
	Macaque	0.0629	0.2880	0.2184	0.0060	0.0948	0.0633
	Marmoset	0.1011	0.4321	0.2340	0.0205	0.1341	0.1529
	Tarsier	0.1113	0.4674	0.2381	0.0290	0.2979	0.0973
	Rabbit	0.1257	0.5646	0.2226	0.0267	0.3083	0.0866
	Pika	0.2201	0.6355	0.3463	0.0354	0.4987	0.0710
	Cow	0.1953	0.3235	0.6037	0.0361	0.4854	0.0744
	Dolphin	0.1590	0.3699	0.4298	0.0236	0.3595	0.0656
	Horse	0.1833	0.6655	0.2754	0.0386	0.2948	0.1309
	Dog	0.2150	0.3490	0.6160	0.0299	0.3998	0.0748
	Cat	0.2609	0.4975	0.5244	0.0306	0.4741	0.0645
	Mouse	0.3041	1.0397	0.2925	0.0321	0.5807	0.0553
	Rat	0.3536	2.8876	0.1225	0.0329	0.6567	0.0501
	Guinea pig	0.3908	1.1119	0.3515	0.0571	0.6529	0.0875
Chimpanzee	Orang-utan	0.0355	0.1410	0.2518	0.0015	0.0267	0.0562
	Macaque	0.0536	0.2567	0.2088	0.0045	0.1009	0.0446
	Marmoset	0.0913	0.3941	0.2317	0.0190	0.1405	0.1352
	Tarsier	0.1113	0.3904	0.2851	0.0275	0.2905	0.0947
	Rabbit	0.1257	0.4661	0.2697	0.0252	0.3008	0.0838
	Pika	0.2201	0.5416	0.4064	0.0338	0.4695	0.0720
	Cow	0.1953	0.2590	0.7541	0.0345	0.4860	0.0710
	Dolphin	0.1590	0.3014	0.5275	0.0220	0.3599	0.0611
	Horse	0.1833	0.5653	0.3243	0.0370	0.2874	0.1287
	Dog	0.2150	0.2784	0.7723	0.0314	0.3827	0.0820
	Cat	0.2490	0.4546	0.5477	0.0321	0.4747	0.0676
	Mouse	0.3041	1.0397	0.2925	0.0306	0.5480	0.0558
	Rat	0.3536	2.8876	0.1225	0.0345	0.6208	0.0556
	Guinea pig	0.3908	0.9414	0.4151	0.0554	0.6176	0.0897
Orang-utan	Macaque	0.0916	0.1433	0.6392	0.0030	0.0948	0.0316
	Marmoset	0.1312	0.1995	0.6576	0.0175	0.1341	0.1305
	Tarsier	0.1524	0.1979	0.7701	0.0259	0.2902	0.0892
	Rabbit	0.1489	0.3010	0.4947	0.0236	0.2851	0.0828
	Pika	0.2437	0.4582	0.5319	0.0322	0.4886	0.0659
	Cow	0.2413	0.2590	0.9317	0.0330	0.5056	0.0653
	Dolphin	0.2027	0.1270	1.5961	0.0205	0.3427	0.0598
	Horse	0.2351	0.3856	0.6097	0.0354	0.2721	0.1301
	Dog	0.2665	0.0993	2.6838	0.0299	0.3823	0.0782
	Cat	0.2978	0.3041	0.9793	0.0306	0.4546	0.0673
	Mouse	0.3756	1.0397	0.3613	0.0290	0.5582	0.0520
	Rat	0.4363	2.5045	0.1742	0.0329	0.6199	0.0531
	Guinea pig	0.4513	0.8025	0.5624	0.0538	0.6168	0.0872
Macaque	Marmoset	0.0907	0.2635	0.3442	0.0144	0.1992	0.0723
	Tarsier	0.1308	0.2612	0.5008	0.0259	0.3457	0.0749
	Rabbit	0.1722	0.3041	0.5663	0.0236	0.3239	0.0729
	Pika	0.2421	0.3904	0.6201	0.0322	0.5083	0.0633
	Cow	0.2165	0.3295	0.6571	0.0291	0.5525	0.0527
	Dolphin	0.1794	0.1837	0.9766	0.0205	0.4026	0.0509

Table S1 Continued

		IDR			Body		
Species 1	**Species 2**	d_N	d_S	d_N/d_S	d_N	d_S	d_N/d_S
	Horse	0.2054	0.4870	0.4218	0.0315	0.3379	0.0932
	Dog	0.2386	0.2177	1.0960	0.0330	0.4264	0.0774
	Cat	0.2130	0.4226	0.5040	0.0337	0.4637	0.0727
	Mouse	0.3406	1.3061	0.2608	0.0259	0.5575	0.0465
	Rat	0.3926	1.6170	0.2428	0.0298	0.6072	0.0491
	Guinea pig	0.4272	0.9178	0.4655	0.0530	0.6581	0.0805
Marmoset	Tarsier	0.1304	0.2029	0.6427	0.0346	0.3005	0.1151
	Rabbit	0.1489	0.3010	0.4947	0.0299	0.3472	0.0861
	Pika	0.2413	0.5592	0.4315	0.0401	0.5483	0.0731
	Cow	0.2158	0.3326	0.6488	0.0369	0.5504	0.0670
	Dolphin	0.1788	0.1853	0.9649	0.0296	0.4027	0.0735
	Horse	0.2286	0.4099	0.5577	0.0386	0.3490	0.1106
	Dog	0.2251	0.2524	0.8918	0.0423	0.4450	0.0951
	Cat	0.2583	0.3490	0.7401	0.0422	0.4492	0.0939
	Mouse	0.3055	1.3061	0.2339	0.0343	0.6084	0.0564
	Rat	0.3926	2.5914	0.1515	0.0377	0.7319	0.0515
	Guinea pig	0.4256	0.9313	0.4570	0.0585	0.6890	0.0849
Tarsier	Rabbit	0.0600	0.1073	0.5592	0.0275	0.3435	0.0801
	Pika	0.1316	0.4674	0.2816	0.0364	0.5635	0.0646
	Cow	0.1103	0.2635	0.4186	0.0322	0.6074	0.0530
	Dolphin	0.0780	0.1290	0.6047	0.0275	0.4243	0.0648
	Horse	0.0957	0.3326	0.2877	0.0354	0.3577	0.0990
	Dog	0.1098	0.1868	0.5878	0.0370	0.3853	0.0960
	Cat	0.1590	0.3099	0.5131	0.0361	0.3844	0.0939
	Mouse	0.2683	1.0892	0.2463	0.0314	0.6311	0.0498
	Rat	0.3154	1.8318	0.1722	0.0361	0.6540	0.0552
	Guinea pig	0.3975	0.9178	0.4331	0.0514	0.7208	0.0713
Rabbit	Pika	0.0604	0.4608	0.1311	0.0170	0.3563	0.0477
	Cow	0.1517	0.2947	0.5148	0.0268	0.5869	0.0457
	Dolphin	0.0397	0.1643	0.2416	0.0213	0.4503	0.0473
	Horse	0.1827	0.3916	0.4665	0.0229	0.3729	0.0614
	Dog	0.1773	0.2180	0.8133	0.0346	0.4419	0.0783
	Cat	0.2561	0.3658	0.7001	0.0283	0.4367	0.0648
	Mouse	0.3213	1.1765	0.2731	0.0190	0.5587	0.0340
	Rat	0.3597	1.1861	0.3033	0.0236	0.5908	0.0399
	Guinea pig	0.3501	1.0714	0.3268	0.0339	0.6528	0.0519
Pika	Cow	0.2218	0.4586	0.4836	0.0373	0.7123	0.0524
	Dolphin	0.1590	0.4452	0.3571	0.0314	0.6124	0.0513
	Horse	0.2068	0.6655	0.3107	0.0355	0.5008	0.0709
	Dog	0.2275	0.4693	0.4848	0.0434	0.6735	0.0644
	Cat	0.2978	0.5429	0.5485	0.0346	0.6620	0.0523
	Mouse	0.3213	0.9241	0.3477	0.0354	0.6200	0.0571
	Rat	0.2901	1.1861	0.2446	0.0401	0.7198	0.0557
	Guinea pig	0.4780	0.7037	0.6793	0.0518	0.7532	0.0688
Cow	Dolphin	0.1505	0.0954	1.5776	0.0229	0.3044	0.0752
	Horse	0.2286	0.2064	1.1076	0.0284	0.3442	0.0825
	Dog	0.2127	0.1585	1.3420	0.0409	0.4518	0.0905
	Cat	0.2583	0.2138	1.2081	0.0306	0.4513	0.0678
	Mouse	0.3701	0.8740	0.4235	0.0361	0.8683	0.0416
	Rat	0.4232	1.9237	0.2200	0.0385	0.7946	0.0485

Table S1 Continued

		IDR			Body		
Species 1	**Species 2**	d_N	d_S	d_N/d_S	d_N	d_S	d_N/d_S
	Guinea pig	0.3817	0.6208	0.6149	0.0515	0.7589	0.0679
Dolphin	Horse	0.1815	0.2064	0.8794	0.0299	0.1974	0.1515
	Dog	0.1650	0.0492	3.3537	0.0314	0.2922	0.1075
	Cat	0.2349	0.1544	1.5214	0.0283	0.2793	0.1013
	Mouse	0.3041	1.0397	0.2925	0.0282	0.7807	0.0361
	Rat	0.3536	1.6805	0.2104	0.0314	0.7932	0.0396
	Guinea pig	0.3817	0.7301	0.5228	0.0490	0.7717	0.0635
Horse	Dog	0.1521	0.2177	0.6987	0.0474	0.2588	0.1832
	Cat	0.1797	0.2758	0.6516	0.0386	0.2795	0.1381
	Mouse	0.3423	0.7940	0.4311	0.0322	0.6407	0.0503
	Rat	0.4273	0.9607	0.4448	0.0362	0.6450	0.0561
	Guinea pig	0.5244	0.6816	0.7694	0.0500	0.6355	0.0787
Dog	Cat	0.1290	0.1355	0.9520	0.0244	0.2369	0.1030
	Mouse	0.3089	0.7739	0.3991	0.0346	0.6765	0.0511
	Rat	0.2870	1.6841	0.1704	0.0362	0.7140	0.0507
	Guinea pig	0.5769	0.6649	0.8676	0.0523	0.6959	0.0752
Cat	Mouse	0.3503	0.7920	0.4423	0.0290	0.7893	0.0367
	Rat	0.4019	0.9313	0.4315	0.0306	0.8020	0.0382
	Guinea pig	0.5456	0.6053	0.9014	0.0506	0.7445	0.0680
Mouse	Rat	0.1618	0.4879	0.3316	0.0045	0.1833	0.0245
	Guinea pig	0.6132	0.9435	0.6499	0.0331	0.7597	0.0436
Rat	Guinea pig	0.5573	1.2745	0.4373	0.0363	0.8011	0.0453

Received 30 October 2012/20 November 2012; accepted 22 November 2012

Published as Immediate Publication 26 November 2012, doi 10.1042/BSR20120112