12 Makalowski, W. and Boguski, M.S. (1998) Evolutionary parameters of the transcribed mammalian genome: an analysis of 2820 orthologous rodent and human sequences. *Proc. Natl Acad. Sci. USA* 95, 9407–9412
13 Galtier, N. and Gouy, M. (1998) Inferring pattern and process: maximum likelihood implementation of a non-homogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol. Biol. Evol.* 18, 871–879
14 Bernardi, G. (1993) The vertebrate genome: isochores and evolution. *Mol. Biol. Evol.* 10, 186–204
15 Hughes, S. *et al.* (1999) Warm-blooded isochore structure in Nile crocodile and turtle. *Mol. Biol. Evol.* 16, 1521–1527
16 Goncalves, I. *et al.* (2000) Nature and structure of human genes that generate retropseudogenes. *Genome Res.* 10, 672–678
17 Sharp, P.M. *et al.* (1995) DNA sequence evolution: the sounds of silence. *Phil. Trans. R. Soc. Lond. Ser. B* 349, 241–247
18 Eyre-Walker, A. (1999) Evidence of selection on silent site base composition in mammals: potential implications for the evolution of isochores and junk DNA. *Genetics* 152, 675–683
19 Smith, N.G.C. and Eyre-Walker, A. (2001) Synonymous codon bias is not caused by mutation bias in human. *Mol. Biol. Evol.* 18, 982–986
20 Nagylaki, T. (1983) Evolution of a finite population under gene conversion. *Proc. Natl Acad. Sci. USA* 80, 6278–6281

0168-9525/03/$ - see front matter
PII: S0168-9525(02)00002-1

Origin of a substantial fraction of human regulatory sequences from transposable elements

I. King Jordan[1], Igor B. Rogozin[1], Galina V. Glazko[2] and Eugene V. Koonin[1]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A/Room N511M, 8600 Rockville Pike, Bethesda, MD 20894, USA
[2]Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University, 328 Mueller Lab, University Park, PA 16802, USA

Transposable elements (TEs) are abundant in mammalian genomes and have potentially contributed to their hosts' evolution by providing novel regulatory or coding sequences. We surveyed different classes of regulatory region in the human genome to assess systematically the potential contribution of TEs to gene regulation. Almost 25% of the analyzed promoter regions contain TE-derived sequences, including many experimentally characterized *cis*-regulatory elements. Scaffold/matrix attachment regions (S/MARs) and locus control regions (LCRs) that are involved in the simultaneous regulation of multiple genes also contain numerous TE-derived sequences. Thus, TEs have probably contributed substantially to the evolution of both gene-specific and global patterns of human gene regulation.

Fully 45% of the human genome draft sequence is composed of transposable element (TE) derived sequences (compared with ~1% dedicated to protein-coding sequences), and this figure is certainly an underestimate, because many TE sequences in the genome will have evolved beyond recognition [1]. TEs have been found in every genome so far examined, and are similarly abundant in the genomes of many other eukaryotic species as they are in humans [2]. The staggering evolutionary success of TEs is attributed mostly to their ability to out-replicate the host genomes in which they reside, as opposed to any selective advantage that they might provide to their hosts. Indeed, TEs can spread within and among genomes, even when there is a selective cost to their hosts [3]. These ideas form the core of the 'selfish DNA' concept of TEs, which focuses on the parasitic nature of the elements, and emphasizes the deleterious effects of transposition and the negligible evolutionary benefit that TEs provide to their hosts [4,5].

However, the sheer abundance of TEs in the genome, as well as the variety of mutation effects induced by their mobility, suggests that they might, in some cases, be 'domesticated' [6] to serve the evolutionary interests of their hosts. Indeed, recent accumulation of evidence shows that the presence of TE sequences can result in positive and creative evolutionary changes for the host [2,7]. In particular, TE sequences seem to contribute substantially to the evolution of human protein-coding sequences [1,7,8], and there are many cases of TE-induced changes in host gene regulation [9,10]. In light of the regulatory effects that some TEs exert on host genes, we sought to examine systematically the contribution of TE-derived sequences to regulatory regions in the human genome. Gene expression is regulated at a number of different levels, including transcriptional regulation by promoters and *cis*-regulatory sequences, posttranscriptional regulation affected by untranslated mRNA regions, and higher-order regulation that is influenced by chromatin formation and nuclear architecture. Various classes of human regulatory regions mediate these distinct modes of gene regulation, and we surveyed them for the presence of TE-derived sequences using the RepeatMasker (http://ftp.genome.washington.edu/RM/RepeatMasker.html) program. This program employs a scoring function to ensure that only statistically significant (i.e. not expected to occur by chance) alignments are reported. In all cases, only experimentally

Corresponding author: I. King Jordan (jordan@ncbi.nlm.nih.gov).

Table 1. Transposable element (TE)-derived sequences in human regulatory regions[a]

Element class	Number elements	Length (bp) occupied	Percent of sequence
Human 5′ promoter regions (2004 sequences)[b]			
LINE	100	15 607	1.6
SINE	378	52 895	5.3
LTR	21	3992	0.4
DNA	50	7026	0.7
Total	549	79 520	7.9
Human S/MARs[c]			
LINE	36	24 535	39.4
SINE	25	5451	8.7
LTR	13	4344	7.0
DNA	1	379	0.6
Total	75	34 709	55.7

[a]Abbreviations: DNA, DNA-type transposons; LINE, long intersperded nuclear elements, also known as non-LTR retrotransposons; LTR, long terminal repeats (LTRs) of endogenous retroviruses and mammalian apparent LTR-retrotransposons (MaLRs); SINE, short interspersed nuclear elements, mainly Alu elments.
[b]Human promoter sequences (500 bp 5′ to the transcription start site) taken from the Human Promoter Database [11].
[c]Scaffold/matrix attachment regions (S/MARs) taken from the scaffold/matrix attachment region (S/MAR) transaction database [18].

characterized (as opposed to predicted) regulatory regions were evaluated.

5′ promoter regions

Promoters can be defined as the sequence regions that are located directly 5′ of transcription initiation sites and that regulate their 3′ adjacent genes. The Human Promoter Database (HPD; http://zlab.bu.edu/~mfrith/HPD.html) is a repository of >2000 such human promoter sequences, each of ~500 bp, that were identified by their location 5′ of experimentally characterized transcription initiation sites [11]. We analyzed these promoters to assess the extent to which they are derived from TEs. Of the 2004 sequences analyzed, 475 (~24%) contain TE-derived sequences, making up ~8% of the total nucleotides in all of the promoters (Table 1). Promoter regions contain all types of common human TE (Table 1), and the relative abundances of the different classes of element are slightly different than those for the entire genome. LINE elements are most abundant in the genome as a whole, whereas SINEs predominate among promoters. This might be due to the fact that LINEs tend to be found in AT-rich DNA characteristic of intergenic regions, as opposed to SINEs (Alus in particular), which are more often found in GC-rich regions where genes also tend to reside. The presence of SINEs in GC-rich DNA does not seem to be a function of insertion site preference, but rather appears to be due to differential retention, presumably mediated by selection after insertion [1]. A positive role for SINEs in the transcriptional regulation of human genes might be partly responsible for this effect. Together with the examples of TE-exerted regulatory effects on host genes, the presence of TE sequences in almost a quarter of human promoter sequences analyzed suggests the potential of TEs to influence the regulation of human genes substantially.

To ascertain whether there is any relationship between the fraction of TE-derived promoter sequences and the distance to the transcription initiation site, the human promoter regions were broken down into 100-bp segments. Six non-overlapping 100-bp regions ranging from positions −500 to +100 bp with respect to the start site of transcription were surveyed independently for the presence of TEs. Beginning at the most-distal 100-bp region, there is a consistent decrease in the fraction of each segment that is derived from TE sequences (Fig. 1a). This suggests that TE insertions proximal to the transcription initiation site are, on average, potentially more deleterious with respect to gene function and, by extension, host fitness, and are removed by selection more often than insertions that are further away.

cis-regulatory elements

To demonstrate unequivocally an effect of TEs on the regulation of host genes, it is necessary to show that experimentally characterized *cis*-regulatory elements that bind nuclear transcription factors have been derived from TE sequences. We searched systematically for such cases using the Transcription Factor Database (TRANSFAC; http://transfac.gbf.de/TRANSFAC/) [12]. A total of 846 experimentally characterized human *cis*-regulatory sites from 288 genes, along with their coordinates in GenBank nucleotide sequence entries, were taken from TRANSFAC. We surveyed the GenBank sequences for the presence of TEs, and recorded the cases where the TEs overlapped

Fig. 1. Transposable elements (TEs) in human regulatory sequences. (a) Percent of TE-derived sequences in 100-bp segments from human 5′ promoter regions. For each segment, the most 5′ position, with respect to the start site of transcription, is shown. (b) TE-derived sequences in three regions of human mRNA sequences: 5′ untranslated regions (5′UTR), protein-coding regions (CDS) and 3′ untranslated regions (3′UTR). Blue bars show the percent of the total nucleotides that are derived from TEs. Red bars show the percent of sequences that have at least one TE insertion.

Table 2. Transposable element (TE)-derived *cis*-regulatory elements

Gi[a]	AC[b]	Medline[c]	Element[d]	Alignment coords[e]	*cis* coords[f]	From tc start[g]	Binding factor[h,i]	Gene	Regulation[i,j]
187844	R08265	94019377	AluY	4402–4736	4434–4464	+4434	GATA-3	CD8 α	T-cell specific expression
187844	R08266	94019377	AluY	4402–4736	4523–4551	+4523	LyF-1	CD8 α	T-cell specific expression
187844	R08267	94019377	AluY	4402–4736	4681–4709	+4681	LyF-1	CD8 α	T-cell specific expression
180261	R04505	93054520	MER113	35–168	115–130	+246	C/EBPalpha	Cholesteryl ester transfer protein (CETP)	Liver specific expression
187211	R04232	93024407	L1MB8	1–55	39–63	−692	HNF-3-like	Lipoprotein lipase (LPL)	Gradual activation of the LPL gene during adipocyte development
927063	R05061	98145936 96201628	AluSx	758–1067	1047–1059	−742	HFH-8	Nitric oxide synthase	na
31759 455025 183870 182961 182964	R02826	91081330	MLT1A2 (LTR)	9407–9743	9468–9480	+7342	GATA-1	γ-globin (A)	Involved in the human gamma globin promoter–enhancer interaction
31759 182961 182964 455025 183870	R02830	91081330	MLT1A2 (LTR)	9407–9743	9495–9509	+7369	na	γ-globin (A)	Involved in the human gamma globin promoter–enhancer interaction
31759 182961 182964 455025 183870	R02831	91081330	MLT1A2 (LTR)	9407–9743	9542–9564	+7416	na	γ-globin (A)	Involved in the human gamma globin promoter–enhancer interaction
31759 182961 182964 455025 183870	R02832	91081330	MLT1A2 (LTR)	9407–9743	9612–9631	+7486	na	γ-globin (A)	Involved in the human gamma globin promoter–enhancer interaction
455025	R08238	96081893	L1MA5	63023–63118	63028–63059	+841	Multiple	β-Globin	Erythroid regulation
288311	R02756	91006005	AluSq	476–550	524–534	−510	na	CD2	Lymphocyte-specific transcription
288311	R02757	91006005	AluSq	476–550	541–548	−493	na	CD2	Lymphocyte-specific transcription
183448	R03305	91224990	L2	874–1069	888–900	−366	na	Platelet glycoprotein IIb (ITGA2B)	Bind megakaryocyte-specific nuclear proteins acting as positive transcription factors; tissue-specific expression
183448	R03306	91224990	L2	874–1069	1015–1025	−239	na	Platelet glycoprotein IIb (ITGA2B)	Bind megakaryocyte-specific nuclear proteins acting as positive transcription factors; tissue-specific expression
183448	R08205	94012737	L2	874–1069	873–882	−381	GATA-1	Platelet glycoprotein IIb (ITGA2B)	Bind megakaryocyte-specific nuclear proteins acting as positive transcription factors; tissue-specific expression
183448	R08206	94012737	L2	874–1069	1005–1014	−249	GATA-1	Platelet glycoprotein IIb (ITGA2B)	Bind megakaryocyte-specific nuclear proteins acting as positive transcription factors; tissue-specific expression
32671	R00917	88311092 89091081 89354547	AluSp	14507–14814	14749–14754	−14749	IRF-1 IRF-2	Interferon α/βreceptor	Mediates virus-induced transcription of the gene
185967	R02751	91006003	LTR39	1–359	9–16	−1038	kappaY factor	IgK subgroup I germline gene	Necessary for expression
537351	R08639	99145597	AluJb	120–339	231–250	−646	TCF-4	Fra-1 (fos-related antigen)	Development of colorectal cancer; part of beta-catenin tc complex
183793	R02887	92351709	AluSg	12469–12728	12614–12623	na	AP-1 NF-E2	α-globin	Erythroid-specific transcriptional regulation; induction of human alpha-globin genes following erythroid differentiation
28630	R05105	97184488	L1MB3	7–490	184–195	−953	HNF-3alpha HNF-3B	Alkaline phosphatase	na

[a]GenBank identifier number for the nucleotide accessions where overlapping TE sequences and *cis*-regulatory elements are found.
[b]Accesion number for the TRANSFAC transcription sites entries that describe the *cis*-regulatory elements.
[c]Unique medline identifier number for the references that describe the experimental results that support each TRANSFAC transcription site entry.
[d]Names of the elements, taken from the RepeatMasker program, that overlap *cis*-regulatory sites.
[e]Coordinates of the region in the GenBank nucleotide sequence entry (corresponding to the first Gi) that is homologous to the TE sequence.
[f]Coordinates of the *cis*-regulatory elements, taken from TRANSFAC, in the GenBank nucleotide sequence files (correspond to the first Gi).
[g]Distance from the transcription start site to the *cis*-regulatory elements for each gene.
[h]Name of the nuclear transcription binding factors, taken from TRANSFAC, that have been shown to interact with the *cis*-regulatory elements.
[i]na indicates data not available.
[j]Brief description of the regulatory phenotypes mediated by the TE-derived *cis*-regulatory elements.

the *cis*-regulatory elements (Table 2). A total of 21 *cis*-regulatory sequence elements (~2.5% of those analyzed) from 13 genes (~4.5%) were derived from TE sequences. Extrapolating from a human gene number of at least 30 000, it could well be the case that more than 1000 human genes are regulated by *cis*-elements generated by the insertion of TEs.

Several cases of TE-derived *cis*-regulatory sequences were found in the β-globin locus on human chromosome 11, which encodes a tightly regulated cluster of four globin genes. For example, there is a γ^{A}-globin enhancer immediately 3′ of the protein-coding region, where four *cis*-regulatory elements that interact with nuclear proteins [13] map to a long terminal repeat (LTR) sequence from a

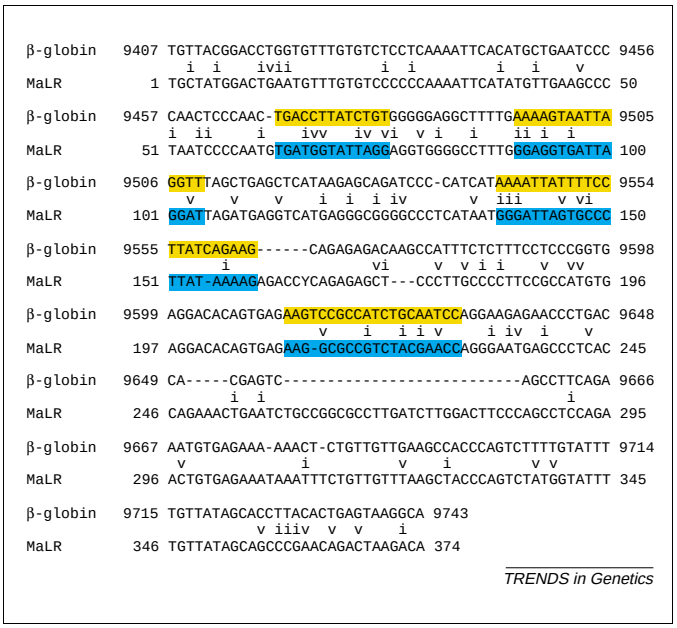

Fig. 2. Sequence alignment of the β-globin locus and a mammalian apparent long terminal repeat (LTR)-retrotransposon (MaLR). The β-globin locus sequence is shown in the top line of the alignment and the sequence coordinates are from the GenBank nucleotide file (gi# 31759). The consensus sequence of the MaLR (MLT1A2#LTR) is shown beneath the β-globin sequence; this sequence and its coordinates are taken from the libraries provided with the RepeatMakser program. Experimentally characterized *cis*-regulatory sequences from the β-globin locus are highlighted in yellow and the homologous nucleotides from the MaLR sequence are highlighted in blue. Transitions (49) are indicated with an 'i' and transversions (32) are indicated with a 'v'. Ten gaps of, on average, 4.7 nucleotides were introduced to improve the sequence alignment. The alignment (excluding gaps) is 332 residues long and there is 75.6% identity between the sequences.

mammalian apparent LTR-retrotransposon (MaLR, Fig. 2). Thus, the MaLR sequence helps to mediate the interaction between the enhancer and the promoter, resulting in both tissue- and developmental-stage-specific expression of the γ^A-globin gene. MaLRs are fairly ancient elements, and sequence comparison of the γ^A-globin enhancer MaLR with a consensus sequence suggests that this particular insertion is ~120–180 million years old. Interestingly, the MaLR insertion pre-dates the diversification of the human and mouse evolutionary lineages, but no such orthologous insertion exists in the mouse genome. Here, TE-driven regulation might be responsible for differences in the regulatory mechanisms at the β-globin locus between the two species.

In addition to *cis*-element driven regulation, transcription of the β-globin locus is also controlled by a conserved ~20-kb sequence region upstream of the first gene in the cluster. This is the 'locus control region' (LCR), which facilitates tissue-specific expression of the locus by opening up the chromatin in a locus-specific manner while insulating against effects of surrounding chromatin [14]. The β-globin LCR also contains abundant TE-derived sequences (~30%). Within the LCR, there are five evolutionarily conserved DNaseI-hypersensitive cores (HSs) that contain clusters of *cis*-regulatory binding sites. The conserved HS3 region includes a SINE element that overlaps several blocks of sequence that are thought to have a role in LCR regulation because they are conserved among all mammalian LCRs and they encode known protein-binding sites [15].

Untranslated regions of mRNA

Both 5′ and 3′ untranslated regions (UTRs) of mRNA sequences often encode important *cis*-elements that function to regulate either transcription or translation. Human mRNA sequences taken from the Mammalian Gene Collection (http://mgc.nci.nih.gov/) [16], a database of experimentally characterized full-length mRNA sequences, were surveyed for the presence of TE-derived sequences. mRNA sequences were partitioned into 5′UTRs, protein-coding sequences (CDSs) and 3′UTRs for comparison. Not surprisingly, CDSs had the lowest TE content, whereas 3′UTRs had far more TEs than did either the CDSs or 5′UTRs (Fig. 1b). The fraction of TE-containing CDSs detected here is comparable to (falls between) previous estimates reported by others for human protein-coding genes [7,8]. It remains a formal possibility that some of the TE-CDS similarity results from mis-annotation of mRNA sequences or is due to the presence of nonfunctional mRNAs. 3′UTRs are substantially longer, on average (592 bp), than 5′UTRs (171 bp). The relative abundance of TEs in the longer 3′UTRs might simply reflect a lack of selection against insertion. However, there are several cases where TE sequences have donated regulatory sequences to the 3′UTRs of host genes [7,10].

Scaffold/matrix attachment regions (S/MARs)

Another mode of transcription regulation in eukaryotes involves the formation of distinct chromatin loops mediated by attachment of specific DNA regions to the nuclear scaffold or matrix [17]. The S/MAR transaction database (S/MARt DB, http://transfac.gbf.de/SMARtDB/) includes a collection of experimentally characterized S/MAR sequences compiled from original publications [18]. We surveyed these sequences for the presence of TEs, and found they are enriched in TE-derived sequences (Table 1). The proportion of TEs in human S/MARs is even greater than in the genome as a whole. LINE elements in particular are over-represented by almost twofold among S/MAR sequences relative to their overall abundance in the human genome sequence. In addition, 98 LINE1 consensus sequences were found to contain 14 distinct S/MAR recognition signatures [19]; these sequence motifs are found in many S/MARs and faithfully identify DNA regions that bind to the nuclear scaffold/matrix. The abundance of TEs in human S/MARs is consistent with the evidence that TEs in yeast [20], *Drosophila* [21] and plants [22] provide matrix attachment regions. Thus, in addition to providing *cis*-regulatory sequences, TEs also appear to have a substantial role in gene regulation by facilitating the partitioning of the human genome into distinct transcriptional foci.

Conclusion

In addition to their well-documented parasitic properties, TE insertions could result in evolutionary changes that are beneficial to the host, particularly by the donation of regulatory sequences. Here, we demonstrate the potential of TEs to affect substantially the regulation of thousands of human genes by donating *cis*-regulatory sites. In addition to these gene-specific regulatory effects, TEs appear to

affect regulation of the human genome in a more global manner by creating S/MARs that form chromatin loops, and by shaping the sequence evolution of LCRs.

Acknowledgements

Galina V. Glazko was supported by research grants from NIH (GM-20293) and NASA (NCC2-1057) awarded to Masatoshi Nei. We thank Nathan J. Bowen and Wolfgang J. Miller for discussions on the relationship between TEs and S/MARs.

References

1 The Human Genome Sequencing Consortium, (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
2 Kidwell, M.G. and Lisch, D.R. (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution Int. J. Org. Evolution* 55, 1–24
3 Hickey, D.A. (1982) Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics* 101, 519–531
4 Doolittle, W.F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284, 601–603
5 Orgel, L.E. and Crick, F.H. (1980) Selfish DNA: the ultimate parasite. *Nature* 284, 604–607
6 Miller, W.J. *et al.* (1999) Molecular domestication – more than a sporadic episode in evolution. *Genetica* 107, 197–207
7 Makalowski, W. (2000) Genomic scrap yard: how genomes utilize all that junk. *Gene* 259, 61–67
8 Nekrutenko, A. and Li, W.H. (2001) Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* 17, 619–621
9 Britten, R.J. (1996) DNA sequence insertion and evolutionary variation in gene regulation. *Proc. Natl Acad. Sci. U.S.A.* 93, 9374–9377
10 Brosius, J. (1999) Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica* 107, 209–238
11 Frith, M.C. *et al.* (2002) Statistical significance of clusters of motifs represented by position specific scoring matrices in nucleotide sequences. *Nucleic Acids Res.* 30, 3214–3224
12 Wingender, E. *et al.* (2001) The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.* 29, 281–283
13 Purucker, M. *et al.* (1990) Structure and function of the enhancer 3′ to the human A gamma globin gene. *Nucleic Acids Res.* 18, 7407–7415
14 Li, Q. *et al.* (1999) Locus control regions: coming of age at a decade plus. *Trends Genet.* 15, 403–408
15 Hardison, R. *et al.* (1997) Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* 205, 73–94
16 Strausberg, R. *et al.* (1999) The mammalian gene collection. *Science* 286, 455–457
17 Bode, J. *et al.* (1996) Scaffold/matrix-attached regions: topological switches with multiple regulatory functions. *Crit. Rev. Eukaryot. Gene Expr.* 6, 115–138
18 Liebich, I. *et al.* (2002) S/MARt DB: a database on scaffold/matrix attached regions. *Nucleic Acids Res.* 30, 372–374
19 van Drunen, C.M. *et al.* (1999) A bipartite sequence element associated with matrix/scaffold attachment regions. *Nucleic Acids Res.* 27, 2924–2930
20 Wyrick, J.J. *et al.* (2001) Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science* 294, 2357–2360
21 Nabirochkin, S. *et al.* (1998) A nuclear matrix/scaffold attachment region co-localizes with the gypsy retrotransposon insulator sequence. *J. Biol. Chem.* 273, 2473–2479
22 Avramova, Z. *et al.* (1998) Matrix attachment regions and structural colinearity in the genomes of two grass species. *Nucleic Acids Res.* 26, 761–767

0168-9525/03/$ - see front matter. Published by Elsevier Science Ltd.
PII: S0168-9525(02)00006-9

An evolutionary basis for scrapie disease: identification of a fish prion mRNA

Eric Rivera-Milla, Claudia A.O. Stuermer and Edward Málaga-Trillo

Department of Biology, University of Konstanz, 78457 Konstanz, Germany

Infectious prion proteins cause neurodegenerative disease in mammals owing to the acquisition of an aberrant conformation. We cloned a *Fugu rubripes* gene that encodes a structurally conserved prion protein, and found rapid rates of molecular divergence among prions from different vertebrate classes, along with molecular stasis within each class. We propose that a directional trend in the evolution of prion sequence motifs associated with pathogenesis and infectivity could account for the origin of scrapie in mammals.

Prion proteins (PrP) are membrane-anchored glycoproteins of unknown function, with the unique ability to change their structure irreversibly from a normal α-helix-rich isoform (PrP^{C}) to a pathological β-sheet-rich isoform known as scrapie (PrP^{Sc}) [1]. This transformation can occur in an autocatalytic manner or with the aid of a hypothetical 'Protein X' [2], leading to the accumulation of insoluble PrP^{Sc} aggregates in the brain. These aggregates cause transmissible spongiform encephalopathies (TSE), a group of lethal, neurodegenerative diseases described only in mammals (e.g. kuru and Creutzfeldt–Jacob in humans, scrapie in sheep, and BSE or 'mad cow' disease in cattle) [3]. Transmission of prion disease between species depends on the degree of sequence similarity at specific amino acids required for interaction between the infectious PrP^{Sc} and the host's PrP^{C} molecules [4,5]. Variability at these sites can create 'host barriers', even between related species [5–7], although infection between distant species can also occur after long exposure times [4]. In fact, human fatalities during the 'mad cow crisis' resulted from the consumption of meat from cows that had been fed dietary supplements contaminated with sheep PrP^{Sc} [8]. Thus, the prion's success in infecting different host species along a food chain is an evolutionary puzzle.

Corresponding author: Edward Málaga-Trillo (edward.malaga@uni-konstanz.de).