

Evolutionary relationships among Rel domains indicate functional diversification by recombination

Isabella A. Graef, Julie M. Gastier, Uta Francke, and Gerald R. Crabtree*

Departments of Genetics, Pathology, and Developmental Biology, Howard Hughes Medical Institute, Stanford University Medical School, Stanford, CA 94305-5323

Contributed by Gerald R. Crabtree, December 20, 2000

The recent sequencing of several complete genomes has made it possible to track the evolution of large gene families by their genomic structure. Following the large-scale association of exons encoding domains with well defined functions in invertebrates could be useful in predicting the function of complex multidomain proteins in mammals produced by accretion of domains. With this objective, we have determined the genomic structure of the 14 genes in invertebrates and vertebrates that contain rel domains. The sequence encoding the rel domain is defined by intronic boundaries and has been recombined with at least three structurally and functionally distinct genomic sequences to generate coding sequences for: (i) the rel/Dorsal/NF κ B proteins that are retained in the cytoplasm by I κ B-like proteins; (ii) the NFATc proteins that sense calcium signals and undergo cytoplasmic-to-nuclear translocation in response to dephosphorylation by calcineurin; and (iii) the TonEBP tonicity-responsive proteins. Remarkably, a single exon in each NFATc family member encodes the entire Ca²⁺/calcineurin sensing region, including nuclear import/export, calcineurin-binding, and substrate regions. The Rel/Dorsal proteins and the TonEBP proteins are present in *Drosophila* but not *Caenorhabditis elegans*. On the other hand, the calcium-responsive NFATc proteins are present only in vertebrates, suggesting that the NFATc family is dedicated to functions specific to vertebrates such as a recombinational immune response, cardiovascular development, and vertebrate-specific aspects of the development and function of the nervous system.

The rel DNA-binding domain is found in a variety of proteins with diverse functions and mechanisms of action. It was first recognized in the transforming gene of the avian reticuloendotheliosis virus (1), the *Drosophila* developmental control gene *Dorsal* (2), and the NF κ B proteins (3). The NF κ B, rel, and dorsal proteins are mechanistically related in that they are retained in the cytoplasm by anchoring proteins related to cactus or I κ B (4). The I κ B and cactus proteins mask nuclear localization signals within the rel domain and are under the control of signaling pathways that regulate their degradation (for review, see ref. 5). In addition, sequences outside the rel domains of dorsal, p65, and c-rel are conserved, indicating that they are likely to be true homologues in the strict sense (6).

A poorly conserved rel domain is also present in the four members of the NFATc family, which encode the cytoplasmic calcium/calcineurin-responsive subunits of NFAT transcription complexes (7, 8). In these proteins, the DNA-binding domain is about 18–20% identical to the rel domain of NF κ B p50 (7). The sequence similarity of the rel and NFATc DNA-binding domains is supported by structural analysis, showing that the topology of the DNA-binding domains of NFATc1(c) and NFATc2(p) is related to the NF κ B p50 rel domain. The rel domain in the NFATc proteins lacks a critical loop that makes contact with DNA in NF κ B (8). These structural features of the NFATc rel domain explain the observation that this group of proteins requires a partner (NFATn) for DNA binding and transcriptional activation. Hence the NFATc proteins act as signal

integrators and coincidence detectors (9). However, because no detectable similarity exists outside of the DNA-binding domains of the NFATc and rel/Dorsal/NF κ B families, they are not homologues or orthologues but simply contain a conserved rel domain. Furthermore, although the NF κ B proteins are retained in the cytoplasm by the I κ B proteins, the NFATc proteins undergo nuclear translocation in response to dephosphorylation by calcineurin, which is regulated by calcium signals (10, 11).

A rel domain is also found in the TonEBP (NFAT5) (12). This protein was discovered in a search for the transcriptional regulators of the tonicity response (13, 14). The TonEBP protein binds to the regulatory regions of several genes that are essential for adaptation to osmotic stress. Although the function of this protein is not completely clear, changes in tonicity activate its transcription, and it in turn activates genes that enhance biosynthesis of solutes that oppose osmotic stress. Although the rel domain of TonEBP is most closely related to the rel domain of NFATc family members, there is no similarity outside the rel domain, and the protein contains none of the distinctive structural and functional features of the NFATc family, such as the SP repeats and the calcineurin-binding domain (12, 15). The TonEBP protein also lacks similarity to the rel/dorsal/NF κ B proteins outside the rel domain and hence is not a homologue or orthologue of rel/dorsal/NF κ B proteins but simply contains a conserved domain.

To help understand the evolutionary origins of the rel domain proteins, we made use of recent genomic sequence data and, where these data were not available, determined experimentally the sequence, genomic structure, and chromosomal map positions for each of the 14 genes for proteins containing rel domains in insects and vertebrates. Our analysis indicates that during the course of evolution, the rel domain has gained function by: (i) recombination with a locus encoding cytoplasmic retention domains containing ankyrin repeats to generate p105, relish, and p100; (ii) an independent recombination with an exon encoding a Ca²⁺/calcineurin-responsive domain to generate the Ca²⁺-dependent NFATc proteins; and (iii) recombination with a regulatory region responsive to tonicity signals to generate the TonEBP or NFAT5 genes.

Methods

Genomic Mapping Studies. The human genes were mapped by genotyping two radiation hybrid mapping panels: Genebridge 4 (93 hybrid lines with resolution of 270 kb/cR) and Stanford G3 (83 hybrid lines and conversion of 30 kb/cR) (16, 17). Two-point maximum likelihood analysis results were obtained by submitting the scores to www-genome.wi.mit.edu/cgi-bin/contig/

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF309388 and AF309389).

*To whom reprint requests should be addressed. E-mail: crabtree@cmgm.stanford.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

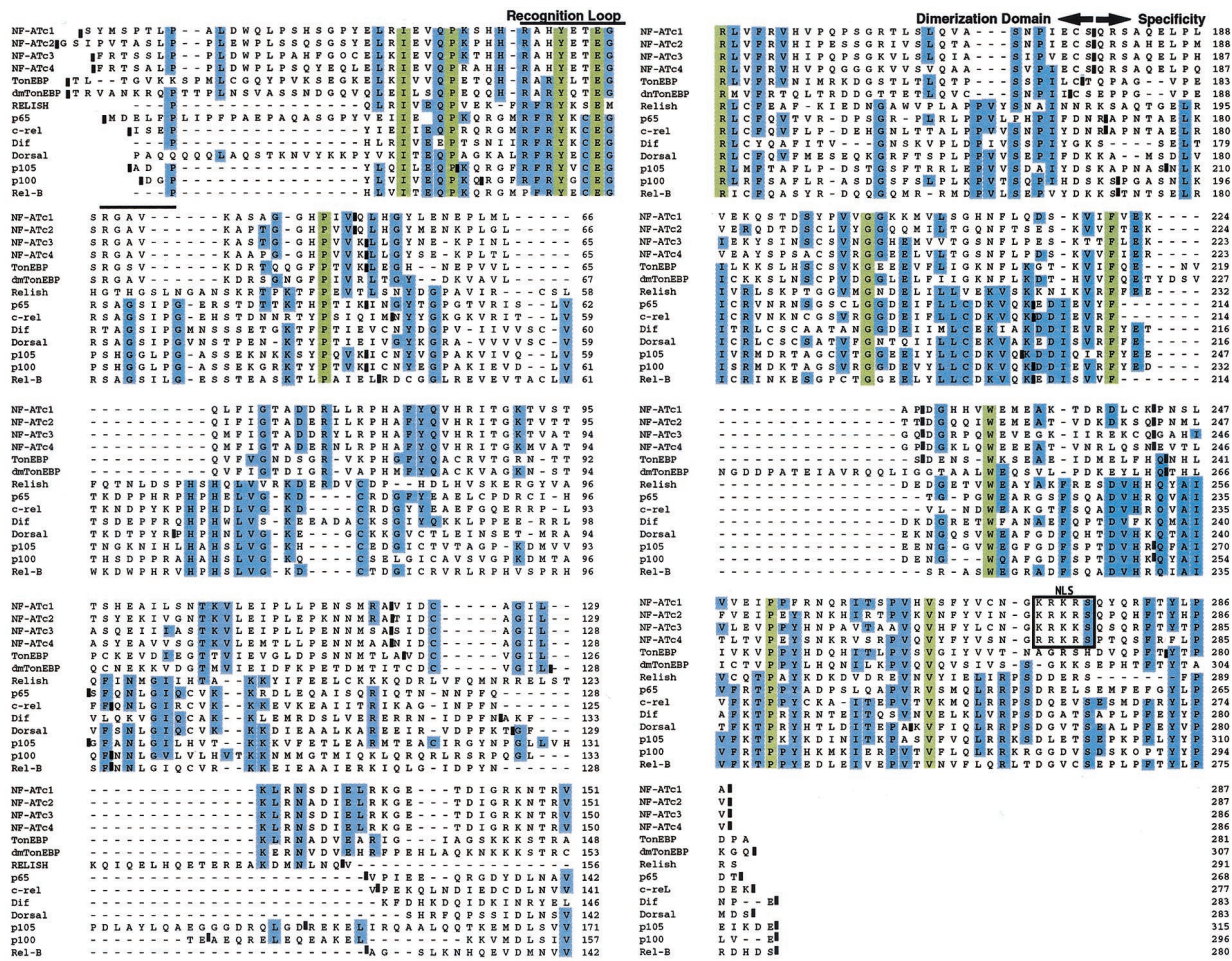


Fig. 1. Placement of introns within the aligned rel domain protein sequences. Only the rel domain is shown. Intron positions in the protein sequence are indicated by a black bar. No attempt has been made to assign the positions of the introns in codons. The alignment shown is that of Brocchieri and Karlin (50). The recognition, dimerization, and specificity regions are indicated above the aligned sequences. For the relish protein, a gap was shifted 18 amino acids to produce the alignment shown, allowing the common intron insertion site to align. The realignment resulted in the loss of only one amino acid identity. NLS, nuclear localization sequence.

rhmapper.pl. The mouse NFATc4 gene was mapped on The Jackson Laboratory interspecific backcross panel BSS (C57BL/6Jei × SPRET/Ei)F1 × SPRET/Ei. For SSCP (120 mM NaCl/5 mM sodium citrate/20 mM sodium phosphate, pH 6.8) analysis, PCR products were separated on 10% nondenaturing polyacrylamide gels at 200 V for 2 h and were visualized by silver staining. Uniquely migrating conformers were identified for the B and S parental strains and used to genotype the 93 backcross animals.

Sequence Determination. The sequences of the genes for murine NFATc4 were determined by conventional methods (18) and have been given GenBank numbers AF309388–AF309389. The sequences for the full-length murine NFATc4 cDNA were determined and entered as GenBank number AF283284.

Results

Because intronic boundaries are some of the most stable features of a gene family (for review, see refs. 19 and 20), we determined the positions of introns in each of the 14 known genes encoding proteins with rel domains. The positions of introns in these genes are highly conserved, with introns positioned to either side of the sequence encoding the rel domain (Fig. 1). The exceptions to this are informative: the sequences encoding the rel domain in relish, Dif, Dorsal, and Rel B lack an intron 5' to the coding region. For the ancestral

gene contained an intron demarcating the N-terminal coding region in these genes, this intron must have been lost before the formation of Rel B, Dorsal, Dif, and Relish, because the other vertebrate genes all have retained this intron. Alternatively, if the ancestral gene lacked an intron demarcating the N-terminal coding region of the rel domain, it must have been inserted after the Relb, dif, dorsal, and relish genes had originated from the ancestral gene. By either scenario, Rel B is the closest vertebrate relative of Dorsal, Dif, and Relish. Introns could not have been randomly lost or inserted, because a number of studies have shown that their positions are highly conserved within gene families (21, 22). The sequence encoding the C terminus of the rel domain is also bounded by introns for each of the proteins except relish (Fig. 1). Indeed, the conserved proline codon at the C terminus of all rel domains occurs within five amino acid codons of the C-terminal intronic insertion (Fig. 1).

The most distinctive structural feature of the rel domain is the division of the dimerization and specificity domains (23, 24). Remarkably, in all vertebrate rel domain-containing genes, an intron precisely separates the sequences encoding the dimerization and the DNA specificity domains within the rel domain (Figs. 1 and 2). Again, the exceptions are informative, in that no insect gene other than *Drosophila* TonEBP has this intron insertion site between the recognition and dimerization do-

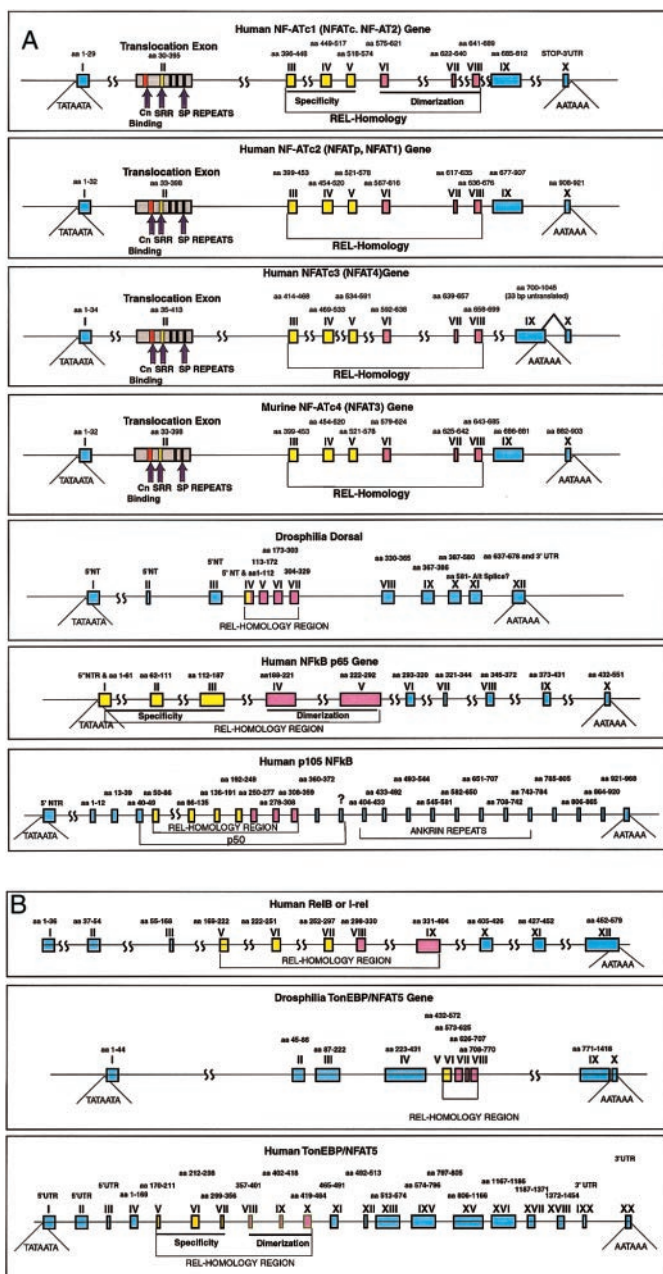


Fig. 2. Intronic domains predict functional domains within the proteins containing rel domains. (A) Genomic structure of the NFATc family of proteins; Roman numerals designate exons. The exon encoding the cytoplasmic-to-nuclear shuttling motifs or the “translocation exon” contains the serine-rich region and SP repeats (33) that are phosphorylation sites by GSK3 and perhaps other kinases, as well as substrates for calcineurin (32). In addition, this domain contains the nuclear localization signal and nuclear export sequence in NFATc1 required for cytoplasmic-to-nuclear shuttling (31). The rel domains are shown according to Muller *et al.*, with the yellow exons indicating the specificity domain and pink exons containing the dimerization domain (23). (B) Genomic structures of four genes encoding rel proteins that are mechanistically distinct from the NFATc proteins by virtue of being retained in the cytoplasm by IκB-like proteins. The specificity and dimerization regions of the rel domain of the Dorsal protein are encoded by a common exon shown in pink and yellow.

mains. One possible explanation is that the ancestral gene contained an intron at this position, which was lost. However, several lines of evidence bode against intron loss (25, 26), particularly because there is no evidence of processing and

reinsertion of the insect rel domains. A more likely scenario is that the ancestral gene gained an intron separating the sequences encoding the dimerization and specificity domains, which then allowed the rel domain to successfully recombine and disseminate in vertebrates.

To determine the time of origin of the structural genomic difference between insect and human rel domains, we examined the rel domains of other vertebrates, including zebrafish, chickens, mice, and hamsters. In all *Drosophila* genes except *relish*, the sequence encoding the rel domain itself is divided by four introns, but in vertebrate genes, it is divided by five to seven conserved introns in all vertebrate genes encoding proteins with rel domains. Although complete information is not available about these other vertebrate species, all vertebrate rel domains from which we have been able to obtain sequence data have the more complex structures, consistent with a gain of introns near the origin of vertebrates, as predicted by studies with other genes (25). This possibility is reinforced by the fact that we were unable to find a protein containing a rel domain or either of the two subdomains of the rel domain in the *Caenorhabditis elegans* genome. However, because structural and topological features of proteins may be similar, with no apparent similarity at the amino acid sequence level, it is still possible that a homologous domain is present in the *C. elegans* genome.

Table 1 shows the positions of conserved introns in all of the rel domains and gives a picture of the likely evolutionary history of the rel domain. Those rel domains sharing intronic positions are almost certainly the most closely related. All rel domains encoding sequences have an intron at position 6 (Table 1). A comparison to the structures of NFκB p50 (23), NFATc1, and NFATc2 (8, 27, 28) indicates this position has no clear relation to the structural domains within the rel protein. Most likely this intron appeared in the ancestral gene for all rel domains and hence represents a useful mark for the evolution of this gene family. Several other relationships are apparent from data presented in Table 1. The shared intronic position five in c-rel, rel B, p65, p105, and p100 indicates they are most related and distinct from the NFATc proteins. Conserved intronic positions at three and nine indicate that the NFATc proteins are most closely related to p105, p100, and rel B.

The exons encoding the Rel domain have been shuffled into other genomic loci, resulting in functional diversification. The demarcation of the rel domain-encoding sequence in vertebrates by introns suggests that after its appearance, it was shuttled about the genome as a functional unit. In the case of the four genes that encode the cytoplasmic components of the NFAT transcription complex, a recombination of the rel domain-coding region has occurred, with an exon encoding those sequences necessary and sufficient for cytoplasmic-nuclear shuttling of the NFATc proteins (11). The proteins of the NFATc family reside in the cytoplasm and translocate to the nucleus after dephosphorylation by the Ca²⁺-dependent phosphatase calcineurin (10, 11), where they combine with nuclear subunits (NFATn) to generate the NFAT transcription complex (29). These proteins are also rapidly exported from the cytoplasm to the nucleus in a manner that allows them to discern specific types of calcium signals, for example rendering them dependent on the CRAC Ca²⁺ channel in lymphocytes (30). This shuttling mechanism requires nuclear import and export sequences (31) as well as phosphorylation sites provided by the serine-proline repeats, which are phosphorylated by GSK3 and dephosphorylated by calcineurin (32). Remarkably, the region of the protein both necessary and sufficient for cytoplasmic-to-nuclear shuttling is encoded by a single exon of NFATc1, c2, c3, and c4 (Fig. 2), which we will refer to as the translocation exon. Because there is no homologue of the NFATc genes in *Drosophila* or *C. elegans*, it appears that recombination between the translocation exon

Table 1. Conserved intron positions within the Rel domains

Gene	N terminal	2	3	4	5	6	7	8	9	10	C terminal
NFATc1	x		x			x	X	x	x		x
NFATc2	x		x			x	X	x	x		x
NFATc3	x		x			x	X	x	x		x
NFATc4	x		x			x	X	x	x		x
TonEBP	x		x			x	X	x	x		x
D Ton	x					x	x		x		x
Relish						x					
P65	x		x		x	x	x	x			x
c-rel	x		x		x	x	x	x			x
Dif						x					x
Dorsal				x		x				x	x
P105	x	x	x		x	x	x	x	x		x
P100	x	x	x		x	x	x	x	x		x
Relb					x	x	x	x	x		x

and the rel domain-coding region occurred near the origin of vertebrates.

Sequences similar to those encoding the translocation domain were not found in extensive searches of available databases. However, the motifs within the exon are present in a number of

other proteins (Table 2). Most notably, the SP repeats (33) appear in a number of proteins that are calcineurin-regulated, including the yeast Rcn1 gene that represses calcineurin activity and is also responsive to calcineurin. The SP repeat is present twice in the heat-shock factor 1 (HSF1) transcription factor of all species. One of the residues in the SP repeat (S-303) of HSF1 is an essential site of phosphorylation by GSK3 (34, 35), indicating that this motif is a substrate for GSK3, as has been demonstrated for mammalian NFATc1 (32) and c4 (36). The Rcn1 gene is homologous to the mammalian Down's Syndrome Critical Region 1 gene (DSCR1) and its homologues, the products of which are inhibitors of calcineurin (37–40). Interestingly, the DSCR1 gene contains a sequence similar to the calcineurin-binding site in the NFATc family members (41, 42). The second motif in the translocation domain of the NFATc proteins is the serine-rich region (SRR), a site of phosphorylation by GSK3 and dephosphorylation by calcineurin, which is essential for cytoplasmic-to-nuclear translocation and nuclear export (11). A number of proteins have this motif, but perhaps most interesting is the occurrence in the protein Crz1p/Tcn1p identified by Cyert, Cunningham, and colleagues (43, 44). Crz1/tan1p contains sequences similar to the phosphoserine-rich region (SRR) originally defined in NFATc1 (Table 2). Crz1 undergoes calcineurin-dependent shuttling similar to NFATc

Table 2. Motifs within the translocation exon of NFATc genes

SP repeat SPxxSP(x)2-5(D/E)(D/E)	Cn-binding domain PxlxlT	Serine-rich region S-P-x-S-S-x-S-x-S-
DSCR1 SPPAASPPVGVKQVED	DSCR1 219-225 PKIIQT	CRZ1 186-208 SRSCNSEASSYESN
Heat shock factor 1 (HSF1) 275-286 SPMASPGGSIDE 303-312 SPPQSPRVEE	African swine fever virus (ASFV) 216-221 PEINIT	Spore coat protein SP96 (DM) 1 486-495 SPSSASSSSS 2 496-505 SPSSASSSSS
SRF Accessory protein (M) 209-218 SPSPSPSSEE	Cadherin γA3 345-350 PEITIT	HSV CP4 Marek's disease 157-166 SPSSSSSSSS
T-type Ca²⁺ channel 28-38 SPPSPPGLEE	Acetylcholine E receptor B 255-260 PCILIT	Hairless (DM) 639-648 SPGSSSSSTS
PAX-9 215-224 SPYHSPKVEE	LAR (M) 5-10 PPIPIT	PPAR-α @ 73-82 SPASSPSSVS
Enhancer of Split M8 (DM) 166-175 SPAPSPMPEE	Synpase-110 (H) 125-130 PGIFIT	HNF-3/Forkhead 261-273 SPDSSSSSLSSGS
E2F1 (Ch) 284-296 SPVKSPFKAPAE	γ-Receptor β-3 (H) 253-258 PSILIT	M-phase inducer phosphatase 264-273 SPCSSTSSCS
PTP ζ (H) 58-69 SPKQSPINIED	JAK1 728-733 PGIPIT	HCF136 (A) 43-52 SPSSSSSSLS
Zygote-specific protein 447-461 SPSPSPSPATDDD	Patched (CE) 241-246 PYISIT	AF-9 protein (H) 380-389 SPASSSSSSS
Hyperplastic disc (DM) 337-348 SPMLSPIWISE	PI3-kinase DM 658-663 PEIYIT	Transformer (CE) 711-720 SPRSSASSGS
Taurine transporter 21-33 SPGKSPGTRPEDE	Stoned-B (DM) 403-408 PDIEIT	Spore coat protein SP70 180-189 SPSSSSSSSS
HSV, Kaposi's ORF 64 2515-2531 SPESPPTSPQPIRVDD	SAP102 (H) 158-163 PGIFIT	
Rcn1p (SC) SPPASPHSEHDD	Importin α (H) 210-215 PSIPIT	
	GABA-A receptor β 2 (Ch) 252-257 PSILIT	
	HSV (type 2) 76-81 PSIPIT	
	JAK2 700-705 PGISIT	
	Protein A4 Vaccinia virus 74-79 PTIHIT	

H, human; M, mouse; @, rat; CE, *C. elegans*; DM, *Drosophila melanogaster*; Ch, chicken; C.S., yeast. If no species is indicated, the motif occurs in multiple species.

proteins. Also of interest is the fact that a number of viral proteins contain the conserved NFATc motifs, including Herpes virus, African Swine Fever, and others. The African Swine Fever virus encodes an inhibitor of calcineurin that allows it to evade the immune response by inhibiting calcineurin (45). Of the different possible evolutionary origins for the translocation exon of NFATc genes, only DSCR1 and its homologues in worms, *Drosophila*, and yeast appear to have more than a single motif that is involved in Ca^{2+} sensing and translocation of the NFATc family. At present, there is no evidence that DSCR1 proteins translocate in response to calcineurin activity, and at this point an evolutionary origin for the translocation exon is uncertain.

In the p100, p105, and relish proteins, a cytoplasmic retention domain is a distinct region in each protein and is characterized by the presence of ankyrin repeats. This region is processed and eventually degraded to allow translocation to the nucleus (46). This cis-acting cytoplasmic retention function in relish is encoded by a single exon, which in vertebrate p105 is divided into 13 different exons (Fig. 2b) and a large but as yet undetermined number of exons in the p100 gene. Cytoplasmic retention can also be provided by the cactus or I κ B proteins, which have sequence similarity to p105, p100, and relish outside the rel domain.

A rel domain related to the one found in the NFATc proteins was recently reported in the mammalian TonEBP or NFAT5 (12, 47). This protein is encoded by a single mammalian gene and is transcriptionally regulated by osmotic stress (14). We found a gene related to mammalian TonEBP in *Drosophila*, the structure of which is shown in Fig. 2c. This gene also has a large exon 5' to the coding sequence for the rel domain, but the protein contains neither the ankyrin repeats of the p105/relish proteins nor the translocation domain of the NFATc family. The *Drosophila* protein shares some features of the human TonEBP protein outside the rel domain, including the glutamine-rich regions. The mammalian gene has been partially sequenced and found to encode a rel domain with its sequence divided by introns at sites that correspond to those present in the NFATc genes. However, outside of the rel domain, the genomic structure of TonEBP is unrelated to NFATc family members. Most definitively, TonEBP lacks the translocation exon, indicating that it is not functionally related to the NFATc proteins.

Genomic Dispersion of Proteins Containing rel Domains. To further discern relationships between the different proteins containing rel domains, we determined the genomic localization of the remaining unmapped genes. We mapped the human NFATC4 gene and TonEBP or NFAT5 gene with two radiation hybrid mapping panels. On the Genebridge 4 panel, NFATC4 was located 15.67 cM from WI-4204, whereas on the Stanford G3 panel, NFATC4 was concordant with SHGC-31972. On the basis of the location of surrounding markers as reported in the GenBank database (D14S72, D14S264, D14S275, and D14S80), NFATC4 was assigned to bands 14a11.2-q12. In the mouse, the NFATc4 gene was assigned to chromosome 14 in a region of known conserved synteny with human region 14q11.2-q12 (Table 3). On the Genebridge 4 panel, TonEBP is located 4.92 cR

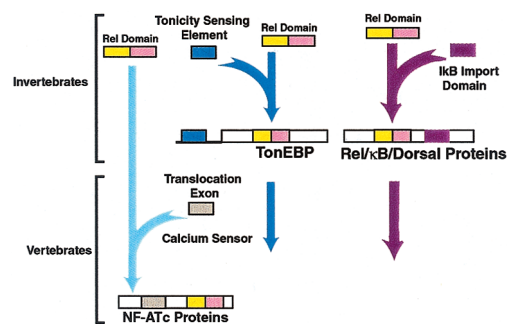


Fig. 3. Evolutionary diversification of the function of the rel domain by recombination.

distal to D16S496, near SHGC-34961. On the Stanford G3 panel, TonEBP is completely concordant with SHGC-34961, located 74 cR proximal to D16S496. The map locations of surrounding markers (D16S400, D16S496) from GenBank and published sources (48) suggest a chromosomal assignment of 16q22-q23 for TonEBP.

Because NFATC3 was previously mapped to 16q21-q22 by fluorescence *in situ* hybridization (49), we determined its distance from TonEBP on the radiation hybrid panels. By using conversion factors of 30 kb/cR and 270 kb/cR for the G3 and Genebridge 4 panels, respectively, NFATC3 was estimated to be 174 kb (G3) to 570 kb (Genebridge 4) proximal to TonEBP on 16q and hence unlikely to arise by tandem duplication. Consistent with this degree of separation, TonEBP and NFATC3 have no sequence similarity outside of the rel domain and hence probably arose by recombination of the rel domain into a distinct locus rather than duplication of the NFATC3 gene.

Discussion

Our analysis of genes encoding proteins with rel domains demonstrates that the rel domain first appeared in invertebrates such as the fly but apparently was not widely disseminated to other species, including worms (Fig. 3). In flies, three types of rel domains are apparent: one, such as relish, linked with exons encoding cytoplasmic retention sequences; a second, such as dorsal, in which the cytoplasmic retention sequences are separated, possibly resulting in the origin of cactus-like proteins; and a third, exemplified by TonEBP, in which the rel domain is linked to an osmotic sensor. At the time of the appearance of vertebrates and clearly by the time of the appearance of bony fish, the NFATc family appeared by recombination of an exon encoding a Ca^{2+} /calcineurin-sensing domain with an exon encoding a rel domain (Fig. 3). The NFATc family contains a rel domain similar to the TonEBP proteins found in insects but is otherwise distinct from the TonEBP proteins. These observations predict that the NFATc family will fulfill Ca^{2+} -dependent signaling functions specific to vertebrates, such as the signaling needs of the recombinational immune system, advanced neurologic functions, and/or the development of a complex cardiovascular system.

- Chen, I. S., Mak, T. W., O'Rear, J. J. & Temin, H. M. (1981) *J. Virol.* **40**, 800–811.
- Steward, R. (1987) *Science* **238**, 692–694.
- Nolan, G. P., Ghosh, S., Liou, H.-C., Tempst, P. & Baltimore, D. (1991) *Cell* **64**, 961–969.
- Baeuerle, P. A. & Baltimore, D. (1988) *Science* **242**, 540–546.
- Baeuerle, P. A. & Baltimore, D. (1996) *Cell* **87**, 13–20.
- Reeck, G. R., de Haen, C., Teller, D. C., Doolittle, R. F., Fitch, W. M., Dickerson, R. E., Chambon, P., McLachlan, A. D., Margoliash, E. & Jukes, T. H. (1987) *Cell* **50**, 667.
- Northrop, J. P., Ho, S. N., Chen, L., Thomas, D. J., Timmerman, L. A., Nolan, G. P., Admon, A. & Crabtree, G. R. (1994) *Nature (London)* **369**, 497–502.

- Wolfe, S. A., Zhou, P., Dotsch, V., Chen, L., You, A., Ho, S. N., Crabtree, G. R., Wagner, G. & Verdine, G. L. (1997) *Nature (London)* **385**, 172–176.
- Crabtree, G. R. (1989) *Science* **243**, 355–361.
- Flanagan, W. M., Corthesy, B., Bram, R. J. & Crabtree, G. R. (1991) *Nature (London)* **352**, 803–807.
- Beals, C. R., Clipstone, N. A., Ho, S. N. & Crabtree, G. R. (1997) *Genes Dev.* **11**, 824–834.
- Miyakawa, H., Woo, S. K., Dahl, S. C., Handler, J. S. & Kwon, H. M. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 2538–2542.
- Rim, J. S., Atta, M. G., Dahl, S. C., Berry, G. T., Handler, J. S. & Kwon, H. M. (1998) *J. Biol. Chem.* **273**, 20615–20621.

14. Takenaka, M., Preston, A. S., Kwon, H. M. & Handler, J. S. (1994) *J. Biol. Chem.* **269**, 29379–29381.
15. Crabtree, G. R. (1999) *Cell* **96**, 611–614.
16. Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega-Czarny, N., Spillet, D., Muselet, D., Prud'Homme, J. F., Dib, C., Auffray, C., *et al.* (1996) *Hum. Mol. Genet.* **5**, 339–346.
17. Stewart, E. A., McKusick, K. B., Aggarwal, A., Bajorek, E., Brady, S., Chu, A., Fang, N., Hadley, D., Harris, M., Hussain, S., *et al.* (1997) *Genome Res.* **7**, 422–433.
18. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
19. Long, M., de Souza, S. J. & Gilbert, W. (1995) *Curr. Opin. Genet. Dev.* **5**, 774–778.
20. Gilbert, W., de Souza, S. J. & Long, M. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 7698–7703.
21. Saxonov, S., Daizadeh, I., Fedorov, A. & Gilbert, W. (2000) *Nucleic Acids Res.* **28**, 185–190.
22. Crabtree, G. R., Comeau, C. M., Fowlkes, D. M., Fornace, A. J., Malley, J. D. & Kant, J. A. (1985) *J. Mol. Biol.* **185**, 1–19.
23. Muller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L. & Harrison, S. C. (1995) *Nature (London)* **373**, 311–317.
24. Ghosh, G., van Duyne, G., Ghosh, S. & Sigler, P. B. (1995) *Nature (London)* **373**, 303–310.
25. Cho, G. & Doolittle, R. F. (1997) *J. Mol. Evol.* **44**, 573–584.
26. Crabtree, G. R., Comeau, C. M., Fowlkes, D. M., Fornace, A. J., Jr., Malley, J. D. & Kant, J. A. (1985) *J. Mol. Biol.* **185**, 1–19.
27. Chen, L., Glover, J. N., Hogan, P. G., Rao, A. & Harrison, S. C. (1998) *Nature (London)* **392**, 42–48.
28. Zhou, P., Sun, L. J., Dotsch, V., Wagner, G. & Verdine, G. L. (1998) *Cell* **92**, 687–696.
29. Shaw, J.-P., Utz, P. J., Durand, D. B., Toole, J. J., Emmel, E. A. & Crabtree, G. R. (1988) *Science* **241**, 202–205.
30. Timmerman, L. A., Clipstone, N. A., Ho, S. N., Northrop, J. P. & Crabtree, G. R. (1996) *Nature (London)* **383**, 837–840.
31. Klemm, J. D., Beals, C. R. & Crabtree, G. R. (1997) *Curr. Biol.* **7**, 638–644.
32. Beals, C. R., Sheridan, C. M., Turck, C. W., Gardner, P. & Crabtree, G. R. (1997) *Science* **275**, 1930–1934.
33. Ho, S. N., Thomas, D. J., Timmerman, L. A., Li, X., Francke, U. & Crabtree, G. R. (1995) *J. Biol. Chem.* **270**, 19898–19907.
34. Xavier, I. J., Mercier, P. A., McLoughlin, C. L., Ali, A., Woodgett, J. R. & Ovsenek, N. (2000) *J. Biol. Chem.* **275**, 29147–29152.
35. Chu, B., Soncin, F., Price, B. D., Stevenson, M. A. & Calderwood, S. K. (1996) *J. Biol. Chem.* **271**, 30847–30857.
36. Graef, I. A., Mermelstein, P. G., Stankunas, K., Neilson, J. R., Deisseroth, K., Tsien, R. W. & Crabtree, G. R. (1999) *Nature (London)* **401**, 703–708.
37. Fuentes, J. J., Pritchard, M. A., Planas, A. M., Bosch, A., Ferrer, I. & Estivill, X. (1995) *Hum. Mol. Genet.* **4**, 1935–1944.
38. Kingsbury, T. J. & Cunningham, K. W. (2000) *Genes Dev.* **14**, 1595–1604.
39. Grolach, J., Fox, D. S., Cutler, N. S., Cox, G. M., Perfect, J. R. & Heitman, J. (2000) *EMBO J.* **19**, 3618–3629.
40. Rothermel, B., Vega, R. B., Yang, J., Wu, H., Bassel-Duby, R. & Williams, R. S. (2000) *J. Biol. Chem.* **275**, 8719–8725.
41. Wesselborg, S., Fruman, D. A., Sagoo, J. K., Bierer, B. E. & Burakoff, S. J. (1996) *J. Biol. Chem.* **271**, 1274–1277.
42. Luo, C., Shaw, K. T., Raghavan, A., Aramburu, J., Garcia-Cozar, F., Perrino, B. A., Hogan, P. G. & Rao, A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 8907–8912.
43. Stathopoulos-Gerontides, A., Guo, J. J. & Cyert, M. S. (1999) *Genes Dev.* **13**, 798–803.
44. Matheos, D. P., Kingsbury, T. J., Ahsan, U. S. & Cunningham, K. W. (1997) *Genes Dev.* **11**, 3445–3458.
45. Miskin, J. E., Abrams, C. C., Goatley, L. C. & Dixon, L. K. (1998) *Science* **281**, 562–565.
46. Lin, L., DeMartino, G. N. & Greene, W. C. (1998) *Cell* **92**, 819–828.
47. Lopez-Rodriguez, C., Aramburu, J., Rakeman, A. S. & Rao, A. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 7214–7219.
48. Bray-Ward, P., Menninger, J., Lieman, J., Desai, T., Mokady, N., Banks, A. & Ward, D. C. (1996) *Genomics* **32**, 1–14.
49. Masuda, E. S., Naito, Y., Tokumitsu, H., Campbell, D., Saito, F., Hannum, C., Arai, K. & Arai, N. (1995) *Mol. Cell Biol.* **15**, 2697–2706.
50. Brocchieri, L. & Karlin, S. (1998) *J. Mol. Biol.* **276**, 249–264.