residue 152 (one-letter amino-acid code), which is consistent with a casein kinase II (CK II) phosphorylation motif[9]. CK II is a ubiquitous enzyme, localized in both the nucleus and the cytoplasm, which is thought to regulate expression of a number of growth factors and proto-oncogenes such as the SV40 large T antigen, adenovirus E1a protein and c-Myc[10].

As CK II has a preference for β-turns near its recognition motif[11] the secondary structure of CREB, as predicted by the Chou-Fasman rules[12], may contain β-turns directly N terminal to the serine residue. Indeed, preliminary experiments using purified CK II suggest that CREB is phosphorylated by this enzyme (G.A.G. and M.R.M., unpublished observations). The proximity of this CK II motif to protein kinase A and C sites indicates that they may interact to regulate CREB activity. Moreover, the preference of CK II for acidic residues is in marked contrast to protein kinase A and C sites, which are highly basic. Phosphorylation by CK II may therefore serve to inhibit protein kinase A and C phosphorylation and consequently cAMP-responsive gene transcription.

At a cellular level, CK II activation seems to be associated with serum-stimulated cell growth[10]. By contrast, cAMP-dependent protein kinase activation enhances cellular differentiation and loss of proliferation[13]. The cellular phenotype may therefore depend, in part, on alternate phosphorylation of CREB by these kinases. *In vitro* mutagenesis of the cloned CREB cDNA will help to clarify the role of CK II and protein kinases A and C in regulating CREB activity.

To verify that the CREB protein encoded by this cDNA contains DNA-binding activity, we used DNAse I protection assays with a truncated fragment of the protein expressed in *Escherichia coli*. To express this protein, we constructed a bacterial expression plasmid that encodes the first 10 amino acids of the β-galactosidase protein, followed by the C-terminal 210 amino acids of the CREB protein. After preparing crude lysates from *E. coli* containing this plasmid, we performed footprinting assays and observed that the fusion protein contained CRE-footprinting activity (Fig. 4*b*, lane 2) which was indistinguishable from native CREB protein (Fig. 2*c*). Lysates prepared from *E. coli* containing pUC 18 vector alone did not contain CRE-binding activity (data not shown).

We observed a stretch of basic amino acids characteristic of a DNA-binding domain near the predicted C terminus of the protein. This putative DNA-binding region shares sequence homology with transcription factor c-Jun/AP-1[14,15] and the proto-oncogene c-fos product (Fig. 4*c*). The close similarity of CREB and c-Jun/AP1-binding sites to one another had prompted us to speculate earlier that these two factors might be structurally related[3]. Moreover, the high-affinity binding of the AP-1 protein to the somatostatin CRE (unpublished data) indicated that CREB and AP-1 shared overlapping DNA-sequence specificities.

Alignment of the three protein sequences reveals a discreet block of 50 amino acids with substantial homology. Closer inspection of this region also reveals a conserved series of regularly spaced leucine residues, interspersed with charged amino acids. Such regions can potentially form amphipathic α-helices, also termed leucine zippers[16], which may interact with other proteins through their hydrophobic surfaces. Dimerization of CREB could therefore occur, in part, through monomer-monomer interactions in this region. Alternatively, other factors like c-Fos may bind to CREB at this site and regulate its activity.

We thank S. Gould for advice on the PCR technique, Joan Vaughan for preparation of CREB antiserum, P. Sawchenko and V. Roberts for assistance with immunocytochemistry and J. Rivier for synthesis of peptide 239. We also thank W. Boyle and T. Hunter for help with phosphorylation experiments, and R. Evans for advice and support. This work was supported by the NIH, and conducted in part by the Clayton Foundation for Research, California Division. W.W.V. and M.R.M. are Clayton Foundation Investigators.

Note added in proof: After submission of this manuscript, Hoeffler *et al.*[22] reported the structure of a human placental cAMP-responsive DNA binding protein which is homologous to rat CREB.

Received 28 November 1988; accepted 16 January 1989.

1. Montminy, M. R., Sevarino, K. A., Wagner, J. A., Mandel, G. & Goodman, R. H. *Proc. natn. Acad. Sci. U.S.A.* **83,** 6682-6686 (1986).
2. Riabowol, K. T. *et al. Nature* **336,** 83-86 (1988).
3. Montminy, M. R. & Bilezikjian, L. M. *Nature* **328,** 175-178 (1987).
4. Yamamoto, K. K., Gonzalez, G. A., Biggs III, W. H. & Montminy, M. R. *Nature* **334,** 494-498 (1988).
5. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. & Kent, S. B. *Proc. natn. Acad. Sci. U.S.A.* **84,** 6970-6974 (1987).
6. Edelman, A. M., Blumenthal, D. K. & Krebs, E. G. *A. Rev. Biochem.* **56,** 567-613 (1987).
7. Lee, C. C. *et al. Science* **239,** 1288-1290 (1987).
8. Saiki, R. K. *et al. Science* **230,** 1350-1352 (1985).
9. Kuenzel, E. A., Mulligan, J. A., Sommercorn, J. & Krebs, E. B. *J. biol. Chem.* **262,** 9136-9140 (1987).
10. Figge, J., Webster, T., Smith, T. & Paucha, E. *J. Virol.* **62,** 1814-1818 (1988).
11. Hathaway, G. M. & Traugh, J. A. *Curr. Topics cell. Regul.* **21,** 101-127 (1982).
12. Chou, P. & Fasman, G. D. *Biochemistry* **13,** 222-245 (1974).
13. Trepel, J. B. & Neckers, L. M. *Molec. cell. Biol.* **7,** 2644-2648 (1987).
14. Bohmann, D. *et al. Science* **238,** 1386-1392 (1987).
15. Angel, P. *et al. Nature* **332,** 166-171 (1988).
16. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. *Science* **240,** 1759-1763 (1988).
17. Vaughan, J. M. *et al. Meth. Enzym.* **168,** 588-617 (1988).
18. Sawchenko, P. E. & Swanson, L. W. *Brain Res.* **210,** 31-51 (1981).
19. Chiu, R., Boyle, W., Meek, J., Smeal, T., Hunter, T. & Karin, M. *Cell* **54,** 541-552 (1988).
20. Montminy, M. R., Goodman, R. H., Horovitch, S. J. & Habener, J. F. *Proc. natn. Acad. Sci. U.S.A.* **81,** 7161-7165 (1984).
21. Van Straaten, F., Muller, R., Curran, T., van Beveren, C. & Verma, I. M. *Proc. natn. Acad. Sci. U.S.A.* **80,** 3183-3187 (1983).
22. Hoeffler, J. P., Meyer, T. E., Yun, Y., Jameson, J. L. & Habener, J. F. *Science* **242,** 1430-1432 (1988).

A pseudo-exon in the functional human αA-crystallin gene

Cynthia J. Jaworski & Joram Piatigorsky

Laboratory of Molecular and Developmental Biology, National Eye Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

The frequent correspondence of exons to structural or functional domains in proteins has suggested that many proteins have evolved by modular assembly[1]. This idea is supported by examples of apparent exon duplication and by shared domains among both alternatively spliced and completely separate genes[1-4]. During this process it is probable that some combinations of exons would not prove advantageous and would therefore be lost. Here we report that within the active single-copy human gene for αA-crystallin there is a 'pseudo-exon' in the early stages of being extinguished, perhaps the result of a failed experiment in the evolution of this specialized, lens-specific protein.

The protein and gene structures of αA-crystallin, a major structural protein in the ocular lenses of all vertebrates, are highly conserved throughout evolution[5,6]. The lenses of rodents[7-9] and certain other mammals (hedgehog, bat and pika)[10] contain an additional form called αA^{ins}, which is identical to αA except for the insertion of 23 amino acids, the result of alternative RNA splicing of an 'insert exon' (IE) located in the first intron[11,12]. No evidence for expression of αA^{ins}-crystallin has been found in lens proteins from primates (including humans) or from many other mammalian classes, birds or other vertebrates[10].

We have now found an element in the first intron of the human αA-crystallin gene[13] closely resembling the alternatively spliced exon of the rodent gene. The DNA sequence of clone pαA-T2, a fragment of the human αA-crystallin gene, is shown in Fig. 1. Comparison with the homologous mouse[11] and chicken[14] sequences shows striking similarity in exons 1 and 2. Surprisingly, a third region of similarity between human and mouse (but not chicken) is apparent in the first intron and

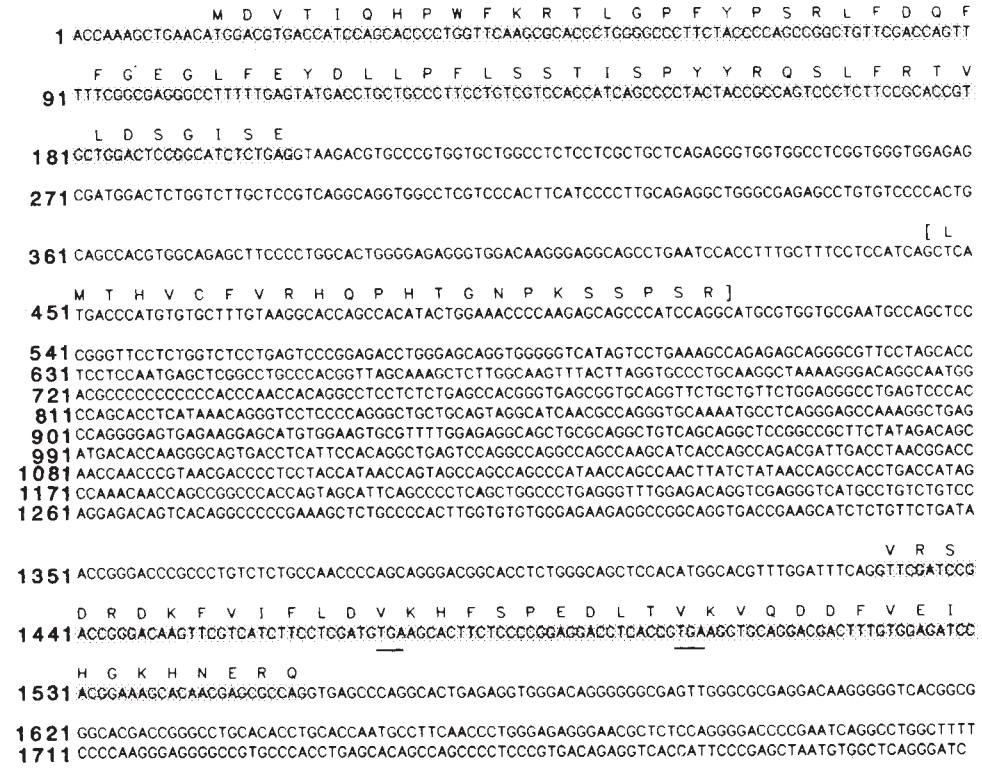

Fig. 1 Nucleotide and deduced amino-acid sequence for the genomic fragment in pαA-T2. Clone pαA-T2 extends from 15 nucleotides upstream of the initiator methionine of the protein-coding region in exon 1 through to 240 nucleotides of the second intron. The DNA sequence of this fragment is presented with the exons displayed against a stippled background and the deduced amino-acid sequence of the exons is shown above the single letter code. Exon junctions are at the same position as in all other species examined, with exon 1 containing amino-acid residues 1 through to 63 and exon 2 corresponding to amino-acid residues 64 through to 104 of the αA polypeptide. The deduced amino-acid sequence of the gene fragment in pαA-T2 confirms the assigned order of residues in the tryptic fragments of the human αA-crystallin protein[19,20]. A region within the first intron corresponding to the alternatively spliced exon of rodent αA-crystallin is in brackets. The relative position of this element is similar in both the human and rodent genes. Underlined codons in the second exon are stop codons that would be brought into frame by an upstream deletion (see text).

Methods. Clone pαA-T2 was obtained by subcloning into bacteriophage M13 mp19 a *Kpn*I/*Bam*H1 fragment from λ phage LS77 which was isolated from a human spleen genomic library[13]; LS77 contains the coding region for all three exons of the human αA-crystallin gene. DNA sequencing was performed on both strands by the dideoxy chain-termination method[21] using oligonucleotide primers made on an Applied Biosystems Model 380B oligonucleotide synthesizer.

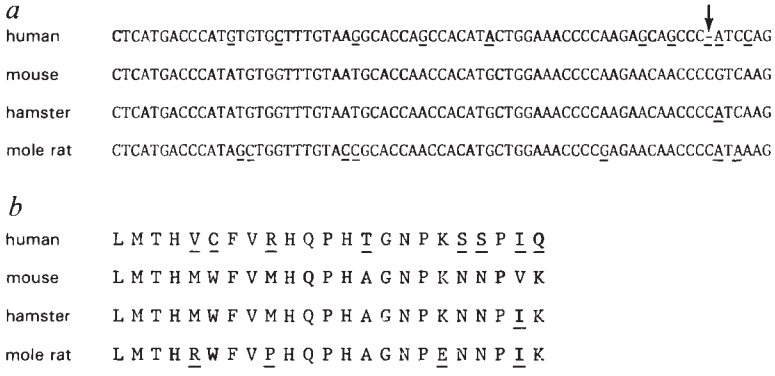

Fig. 2 Comparison of the insert exon from various species. *a*, The nucleotide sequence of the IE of the mouse, hamster and blind mole rat aligned with the analogous element from the human αA-crystallin gene. *b*, The deduced amino-acid sequence of the human IE-like element is shown with the IE protein sequences expressed by mouse, hamster and blind mole rat. A deletion in the human DNA sequence is marked by an arrow. Underlined residues in both panels indicate mismatches relative to the mouse sequence.

corresponds to the coding region for the rodent IE, revealing the presence of an IE-like element within the first intron of the human gene.

The nucleotide and deduced amino-acid sequences of the human IE-like sequence were aligned with those of the mouse[11], hamster[12] and blind mole rat[15] (Fig. 2). The best alignment, preserving the splice junctions found in the rodent sequences, introduces a deletion in the human sequence (see arrow in Fig. 2). The resulting frameshift brings two stop codons into the reading frame in exon 2 (see underlined codons in Fig. 1). Analysis of the lens proteins of humans and other primates reveals no detectable levels of an αA-crystallin with an insert peptide[10], nor of a truncated αA-crystallin polypeptide (C.J.J., unpublished result; J. Horwitz, personal communication). Apparently, the IE-like element in the human αA-crystallin gene is not a functional protein-coding sequence.

The IE-like element in the human gene is more conserved than the introns but less well conserved than the coding sequences when compared with the rodent genes (see Table 1). The exon 1 coding sequences in the mouse and human genes are identical in 94% of the nucleotide positions. All but one difference is in the third nucleotide of a codon; only one amino acid out of 63 is different (position 13). In contrast, the human IE-like element is 86% identical in nucleotide sequence to the mouse IE, and has differences nearly equally distributed among the codon positions; for 9 nucleotide changes there are 8 amino-acid differences. The hamster and mouse IE sequences differ by only one amino acid, whereas the blind mole rat, whose eyes are degenerate, has 4 alterations. This suggests that the human IE-like element is not under the same selective pressure as the other exons. By analogy with pseudogenes, we consider the human IE-like sequence to be a silent pseudo-exon.

The presence of the IE-like pseudo-exon in the human αA gene suggests that αA IE arose in early ancestors of many mammals; although useful enough to be retained in some species, it has been silenced in most others. Using the method of Perler *et al.*[16] for calculating diversity between two sequences, and assuming the sequence of the human IE-like element is

Table 1 The divergence of the αA-crystallin genes of the human, hamster and blind mole rat relative to that of the mouse gene

	Exon 1	Intron 1	Insert exon
Human	94	47	86
Hamster	95	59	99
Mole rat	90	49	93

Divergence is shown by comparing nucleotide sequences for the coding region of the first exon, the IE and the intervening sequence between them. Values given for coding regions represent the percentage of positions at which the nucleotides are identical; values for intron sequences represent the percentage of matched residues after alignment by the NUCALN program[22] using default values for all parameters.

accumulating changes at the same rate as pseudogenes (4.5×10^{-9} substitutions per site per year)[17], we estimate that the human gene stopped using the IE some 30 to 40 million years ago. Interestingly, the human genome contains pseudogenes for two other lens proteins, both γ-crystallins with functional counterparts in rats[18]. This silencing of genes and an exon may have been triggered by similar environmental changes resulting in altered requirements for lens proteins.

The IE was acquired by an extremely well conserved, tissue-specific gene, either by exon shuffling or by the development of functional splicing signals within an intron. Its transformation into a pseudo-exon shows that part of a functional gene can be released from evolutionary pressures while the rest remains stringently conserved. This process may have occurred many times in the evolutionary history of modern proteins, but is only rarely glimpsed now.

We thank Dr Graeme Wistow, Luke Pallansch and Thomas Lietman for discussion and Dawn Chicchirichi for typing the manuscript.

Received 19 October 1988; accepted 6 January 1989.

1. Gilbert, W. *Nature* **271**, 501-502 (1978).
2. Blake, C. C. F. *Int. Rev. Cytol.* **93**, 149-185 (1985).
3. Holland, S. K. & Blake, C. C. F. *Biosystems* **20**, 181-206 (1987).
4. Breitbart, R. E., Andreadis, A. & Nadal-Ginard, B. *A. Rev. Biochem.* **56**, 467-495 (1987).
5. de Jong, W. W., van Amelsvoort, A. M., van der Ouderaa, F. J. & Bloemendal, H. *Nature new Biol.* **246**, 233-236 (1973).
6. de Jong, W. W., Zweers, A., Versteeg, M. & Nuy-Terwindt, E. C. *Eur. J. Biochem.* **141**, 131-140 (1984).
7. Cohen, L. H., Westerhuis, L. W., de Jong, W. W. & Bloemendal, H. *Eur. J. Biochem.* **89**, 259-266 (1978).
8. Cohen, L. H., Westerhuis, L. W., Smits, D. P. & Bloemendal, H. *Eur. J. Biochem.* **89**, 251-258 (1978).
9. de Jong, W. W., Cohen, L. H., Leunissen, A. M. & Zweers, A. *Biochem. biophys. Res. Commun.* **96**, 648-655 (1980).
10. Hendriks, W., Sanders, J., de Leij, L., Ramaekers, F., Bloemendal, H. & de Jong, W. W. *Eur. J. Biochem.* **174**, 133-137 (1988).
11. King, C. R. & Piatigorsky, J. *Cell* **32**, 707-712 (1983).
12. van den Heuvel, R., Hendriks, W., Quax, W. & Bloemendal, H. *J. molec. Biol.* **185**, 273-284 (1985).
13. McDevitt, D. S., Hawkins, J. W., Jaworski, C. J. & Piatigorsky *J. expl. Eye Res.* **43**, 285-291 (1986).
14. Thompson, M. A., Hawkins, J. W. & Piatigorsky, J. *Gene* **56**, 173-184 (1987).
15. Hendriks, W., Leunissen, J., Nevo, E., Bloemendal, H. & de Jong, W. W. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5320-5324 (1987).
16. Perler, F. *et al. Cell* **20**, 555-566 (1980).
17. Li, W.-H., Gojobori, T. & Nei, M. *Nature* **292**, 237-239 (1981).
18. Meakin, S. O., Du, R. P., Tsui, L. C. & Breitman, M. L. *Molec. cell. Biol.* **7**, 2671-2679 (1987).
19. de Jong, W. W. & Goodman, M. *Sonderdruck Z. Saugetierkunde* **47**, 257-276 (1982).
20. de Jong, W. W., Terwindt, E. C. & Bloemendal, H. *FEBS Lett.* **58**, 310-313 (1975).
21. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
22. Wilbur, W. J. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 726-730 (1983).

Dynamical transition of myoglobin revealed by inelastic neutron scattering

Wolfgang Doster*, Stephen Cusack† & Winfried Petry‡

* Physik Department E13, Technischen Universität München, D-8046 Garching, FRG
† EMBL Grenoble Outstation, Institut Laue-Langevin, 156X, 38042 Grenoble Cedex, France
‡ Institut Laue-Langevin, 156X, 38042 Grenoble Cedex, France

Structural fluctuations in proteins on the picosecond timescale have been studied in considerable detail by theoretical methods such as molecular dynamics simulation[1,2], but there exist very few experimental data with which to test the conclusions. We have used the technique of inelastic neutron scattering to investigate atomic motion in hydrated myoglobin over the temperature range 4–350 K and on the molecular dynamics timescale 0.1–100 ps. At temperatures below 180 K myglobin behaves as a harmonic solid, with essentially only vibrational motion. Above 180 K there is a striking dynamic transition arising from the excitation of non-vibrational motion, which we interpret as corresponding to torsional jumps between states of different energy, with a mean energy asymmetry of 12 kJ mol^{-1}. This extra mobility is reflected in a strong temperature dependence of the mean-square atomic displacements, a phenomenon previously observed specifically for the heme iron by Mössbauer spectroscopy[3–5], but on a much slower timescale (10^{-7} s). It also correlates with a glass-like transition in the hydration shell of myoglobin[6] and with the temperature-dependence of ligand-binding rates at the heme iron, as monitored by flash photolysis[7]. In contrast, the crystal structure of myoglobin determined down to 80 K shows no significant structural transition[8–10]. The dynamical behaviour we find for myoglobin (and other globular proteins) suggests a coupling of fast local motions to slower collective motions, which is a characteristic feature of other dense glass-forming systems.

Inelastic neutron scattering is a spectroscopic technique which can be used to study protein motions on exactly the same timescale (0.1–100 ps) as is now widely accessible by computer simulation[11]. Because of the anomalously large incoherent neutron cross-section of the ^{1}H nucleus, the method specifically probes the motion of hydrogen atoms. As hydrogens are abundant and are uniformly distributed in proteins, the method gives a global view of protein dynamics. The quantity measured experimentally is the incoherent dynamic structure factor $S(\mathbf{q}, \omega)$, where $\hbar\mathbf{q}$ and $\hbar\omega$ are respectively the momentum and energy transfers between system and incident neutron. $S(\mathbf{q}, \omega)$ is the Fourier transform of the time-correlation function of the density fluctuations in a system and can be directly calculated from the results of a molecular dynamics simulation. For samples of myoglobin hydrated with D_2O, we have measured as a function of temperature both the elastic intensity $S(q, \omega \approx 0)$, which gives information on the geometry of motions, and $S(q, \omega > 0)$, which gives the timescale (or spectrum) of the corresponding diffusive and vibrational motion[12,13].

Between 4 K and 180 K the elastic intensity of myoglobin has the gaussian form expected for a harmonic solid whose vibrational motion can be described by a Debye-Waller factor, that is, $S(q, \omega \approx 0) = \exp(-q^2\langle\Delta x^2\rangle)$, where the mean-square displacement $\langle\Delta x^2\rangle$ is proportional to temperature (except for quantum effects at very low temperatures). But near 200 K an extra decrease in the elastic intensity at low q is observed, indicating the excitation of new degrees of freedom. The geometry of these motions gives rise to a transition to a non-gaussian elastic intensity (Fig. 1). As the deviation from gaussian behaviour increases with temperature, we can conclude that the motion involves jumping of hydrogens to distinct sites of different energy. The simplest model accounting for these observations involves jumps between two sites separated by a distance d and free energy ΔG (Fig. 2). The powder-averaged normalized elastic