The structure of the mammalian AcChR has ramifications with regard to human pathology. In myasthenia gravis the neuromuscular AcChR is destroyed mainly by specific autoantibodies, which accelerate the catabolism of AcChR molecules as a result of cross-linking (*1*). Since mammalian AcChR is formed by similar subunits, one of which is present in two copies, repetitive homologous antigenic determinants should be present on its surface. Antibodies directed against such determinants would be inefficient in accelerating the degradation of the AcChR because, instead of cross-linking adjacent AcChR molecules, they should preferentially bind to two homologous determinants within the same AcChR molecule (*19*). In addition, both in myasthenia gravis (*20*) and in EAMG, its experimental model (*5*), a significant fraction of the antibodies present in the serum are directed against antigenic determinants on the "α" subunits. Such considerations would explain the existing lack of correlation between titers of antibody to AcChR and the severity of symptoms both in myasthenia gravis and in EAMG (*21, 22*), because only particular antibody subpopulations (that is, those directed against nonrepetitive antigenic determinants) should be able to induce acceleration of AcChR catabolism.

The highly conserved pentameric subunit structure of the AcChR complex strongly suggests that each of the subunits evolved to perform discrete, crucial functions in the physiological actions of the AcChR. In this respect, one could speculate about the existence of multiple binding sites for cholinergic ligands on the homologous domains of the AcChR molecules, whose binding could trigger different functions, such as activation, inactivation, or desensitization. The existence of an ancestral gene for this acetylcholine-binding protein raises the possibility of a shared ancestry with other acetylcholine-binding proteins, such as the muscarinic AcChR and cholinesterases, at least regarding the recognition domains of these molecules.

BIANCA M. CONTI-TRONCONI*
CECILIA M. GOTTI*
Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena 91125
MICHAEL W. HUNKAPILLER
Division of Biology, California Institute of Technology, Pasadena
MICHAEL A. RAFTERY
Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena

References and Notes

1. Reviewed in D. B. Drachman, *N. Engl. J. Med.* **298,** 135, 186 (1978).
2. Reviewed in M. B. Conti-Tronconi and M. A. Raftery, *Annu. Rev. Biochem.* **51,** 491 (1982).
3. J. Patrick and J. Lindstrom, *Science* **180,** 871 (1973).
4. J. Lindstrom, B. Einarson, J. Merlie, *Proc. Natl. Acad. Sci. U.S.A.* **75,** 769 (1978).
5. S. Tzartos and J. Lindstrom, *ibid.* **77,** 755 (1980).
6. C. Gotti, B. M. Conti-Tronconi, M. A. Raftery, *Biochemistry* **21,** 3148 (1982).
7. J. A. P. Rostas, P. G. Kelly, C. W. Cotman, *Anal. Biochem.* **80,** 366 (1977).
8. R. B. Adelstein and W. M. Keuhl, *Biochemistry* **9,** 1355 (1979); R. R. Weihing and E. D. Koni, *ibid.* **10,** 590 (1979); F. M. Booyse, T. P. Movelle, M. E. Rafelson, *J. Biochem.* **248,** 4083 (1973); J. D. Pardee and J. R. Bamburg, *Biochemistry* **18,** 2245 (1979).
9. E. Lazarides, *J. Cell Biol.* **68,** 202 (1976).
10. J. M. Wolosin, A. Lyddiatt, J. O. Dolly, E. A. Barnard, *Eur. J. Biochem.* **109,** 495 (1980).
11. Polypeptides with molecular weights of 42, 44, 49, 53, 55, and 58 K were isolated from SDS gels by electroelution [B. M. Conti-Tronconi, M. W. Hunkapiller, J. Lindstrom, M. A. Raftery, *Biochem. Biophys. Res. Commun.* **106,** 312 (1982)] and submitted to amino terminal sequence analysis by automated Edman degradation in a gas-phase microsequenator [R. M. Hewick, M. W. Hunkapiller, L. E. Hood, W. J. Dreyer, *J. Biol. Chem.* **256,** 7990 (1981)]. Sequences were obtained directly for the polypeptides with molecular weights of 42, 49 and 53 K, whereas those with M_r 44, 55, and 58 K did not yield any discernible sequence, presumably due to blocked amino terminal residues.
12. One-letter abbreviations for amino acids: E, glutamic acid; H, histidines; N, asparagine; K, lysine; L, leucine; Q, glutamine; A, alanine; F, phenylalanine; D, aspartic acid; Y, tyrosine; S, serine; P, proline; and R, arginine.
13. J. Vandekerckhone and K. Weber, *Proc. Natl. Acad. Sci. U.S.A.* **75,** 1106 (1978).
14. M. A. Raftery, M. W. Hunkapiller, C. D. Strader, L. E. Hood, *Science* **208,** 1454 (1980); C. D. Strader, M. Hunkapiller, L. Hood, M. A. Raftery, in *Psychopharmacology and Biochemistry of Neurotransmitter Receptors*, H. Yamamura, R. Olsen, E. Usdin, Eds. (Elsevier/North-Holland, New York, 1980), pp. 35–46.
15. B. M. Conti-Tronconi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79,** 6489 (1982).
16. Purified AcChR (approximately 0.25 nmole) was electrodialyzed against 0.05*M* NH_4HCO_3, 0.1 percent SDS and subjected to amino terminal sequence analysis in a gas-phase microsequenator; the phenylthiohydantoin amino acids released at each cycle were quantified by reverse-phase high-performance liquid chromatography. The stoichiometry of the AcChR polypeptides was determined from the relative amounts of the amino acids released at selected steps, with sequence assignments for the individual amino acids to particular polypeptides made as described.
17. It is not possible at present to define the precise subunit composition of calf AcChR since, although it is clear that four homologous subunits comprise the AcChR, five polypeptides were present upon SDS gel electrophoresis that could give rise to these sequences in a variety of ways. There is no doubt that the peptides with molecular weights of 42, 49, and 53 K are AcChR subunits. However, the 53 K component could either be a subunit particularly susceptible to degradation or arise from either the 55 K or 58 K polypeptides as the result of either variation in the extent of glycosylation or by degradation. Alternatively, it is possible that the 55 K and 58 K components could represent different degrees of glycosylation of the same peptide.
18. B. Einarson, W. Gullick, B. M. Conti-Tronconi, M. Ellismann, J. Lindstrom, *Biochemistry*, in press.
19. B. Conti-Tronconi, S. Tzartos, J. Lindstrom, *ibid.* **20,** 2181 (1981).
20. S. Tzartos, M. E. Seybold, J. Lindstrom, *Proc. Natl. Acad. Sci. U.S.A.* **79,** 188 (1982).
21. R. R. Almond and S. H. Appel, *Ann. N.Y. Acad. Sci.* **274,** 235 (1976); J. Lindstrom, M. E. Seybold, V. A. Lennon, S. Whittingham, D. D. Drake, *Neurology (Minneapolis)* **26,** 1054 (1976); S. H. Appel and S. B. Elias, *Plasmapheresis and the Immunobiology of Myasthenia Gravis*, P. Dau, Ed. (Houghton Mifflin, Boston, 1979), pp. 52–58.
22. P. W. Berman and J. Patrick, *J. Exp. Med.* **151,** 204 (1980).
23. This research was supported by USPHS NS-10294 and GM-06965 and by grants from the Muscular Dystrophy Association of America, the Myasthenia Gravis Foundation (Los Angeles Chapter), the Pew Charitable Trust, and the Weingart Foundation. We thank John Racs and Laurie Jutzi for technical assistance.
* Permanent address: Department of Pharmacology, University of Milano, via Vanvitelli 32, Milano, Italy

24 May 1982; revised 24 September 1982

DNA Sequence of the Gene Encoding the E_α Ia Polypeptide of the BALB/c Mouse

Abstract. *A 3.4-kilobase DNA fragment containing the gene coding for the E_α chain of an Ia (I region–associated) antigen from the BALB/c mouse has been sequenced. It contains at least three exons, which correlate with the major structural domains of the E_α chain—the two external domains α1 and α2, and the transmembrane-cytoplasmic domain. The coding sequence of the mouse E_α gene shows striking homology to its human counterpart at the DNA and protein levels. The translated α2 exon demonstrates significant similarity to β_2-microglobulin, to immunoglobulin constant region domains, and to certain domains of transplantation antigens. These observations and those of others suggest that the Ia antigen, transplantation antigen, and immunoglobulin gene families share a common ancestor.*

The major histocompatibility complex (MHC) of the mouse encodes several families of cell-surface glycoproteins which regulate various aspects of immune responsiveness (*1, 2*). Certain products of the class I genes, the transplantation antigens, serve as restricting elements in T cell immunosurveillance. The products of the class II genes, the I region–associated (Ia) antigens, play a fundamental role in determining the effectiveness of cell-cell interactions between regulatory T cells, B cells, and macrophages. A variety of recent data from genetic, functional, and biochemical studies suggest that the Ia antigens are the products of the Ir (immune response) genes, which control the ability of an animal to respond to synthetic and naturally occurring antigenic determinants (*3–7*).

Two Ia antigens, I-A and I-E, have

0036-8075/82/1217-1229$01.00/0

been defined serologically and biochemically [reviewed in (*8*)]. Each is composed of an α and a β polypeptide chain, denoted A_α and A_β for I-A antigens and E_α and E_β for I-E antigens. The class II β chains appear to be quite polymorphic, whereas their α chain counterparts are less so. The set of alleles at the MHC loci of a particular inbred strain of mouse is denoted its haplotype. For example, the BALB/c mouse, whose genes we are studying, is of the d haplotype, and its class II genes are denoted A^d_α, A^d_β, E^d_α, and E^d_β.

Protease digestion studies of human α and β chains, protein sequence of a human β chain, as well as DNA sequence analyses of human class II complementary DNA (cDNA) clones, suggest that both the α and β polypeptides are divided into two external domains each of approximately 90 residues (α1 and α2 and β1 and β2), a transmembrane region of about 30 residues, and a cytoplasmic region of about 15 residues (*9–13*). Studies on the genomic organization of the class II genes are in progress in several laboratories. We now report the DNA sequence of the E_α gene of the BALB/c mouse.

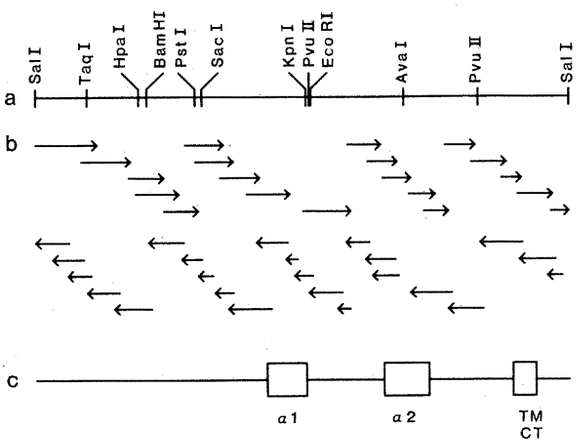

Fig. 1. The organization and sequencing strategy for the E_α gene. (a) A restriction map of the 3.4-kb Sal I fragment containing exons of the E_α gene which hybridized to the human DR_α probe (*14, 17*). The 3.4-kb fragment was subcloned into the Sal I site of pBR325 for mapping by single and double digests with the restriction enzymes indicated. This map was confirmed by computer analysis of restriction enzyme sites in the DNA sequence of this fragment shown in Fig. 2. (b) Sequence strategy for the E^d_α gene. Each arrow represents the sequence of an M13 clone. The 3.4-kb fragment from cosmid 32.1 (*14*) was cloned in both orientations into M13 mp8 to give two parental subclones, mp8-32.11 and mp8-32.12. Each parental subclone was then used to generate a series of overlapping subclones by a deoxyribonuclease deletion technique (*16*). (c) Organization of the E^d_α gene. Exons are represented by boxes and introns by lines. The 3.4-kb fragment contains one exon encoding most of the first protein domain, α1 (amino acids 3 to 84), a second exon encoding the second protein domain, α2 (amino acids 85 to 178), and a third exon that contains coding sequence for amino acids 179 to 230 and includes the transmembrane and cytoplasmic protein domains (*TM-CT*).

We isolated the mouse E_α gene by screening a cosmid library constructed from BALB/c sperm DNA with a human DR_α (D-related) chain cDNA clone (*14*). A 3.4-kilobase (kb) Sal I fragment that hybridized to the human cDNA probe was isolated from cosmid 32.1 and cloned in both orientations into M13mp8; its nucleotide sequence was determined by the dideoxy chain termination method. A novel sequencing strategy adapted from the method of Frischauf *et al.* (*15*) was employed to generate an ordered series of subclones with increasing lengths of deletions starting at one end of the insert by deoxyribonuclease I digestion (*16*) (Fig. 1). The DNA sequence of the 3.4-kb fragment containing the E_α gene is given in Fig. 2.

The exons of the mouse E_α gene were identified on the basis of their homology to cDNA sequences of human DR_α clones (*13, 17*) as well as to the amino acid sequences of human DR_α chains and mouse E_α sequences (*18*). Three exons were identified. These exons encode the α1 domain (codons 3 to 84), the α2 domain (codons 85 to 178), and a domain which includes both the transmembrane and cytoplasmic regions (codons 179 to 230). A termination codon in phase with this reading frame is found at codon 231. Intervening sequences split the codons 3, 85, and 179 between base positions 1 and 2. Intervening sequences also split the first and second bases of codons in the genes encoding class I molecules, immunoglobulins, and β_2-microglobulin (*19–21*). The bases GT and AG are found at the 5′ and 3′ ends of each intron, respectively, as in virtually all other eukaryotic genes.

The three exons of the E_α gene contain the entire coding sequence except for the first two amino acids and a presumed leader peptide. Thus the E_α gene is split into at least four exons. Translation of the DNA sequence of 1480 bases 5′ to the α1 exon in all three reading frames did not reveal a hydrophobic stretch of amino acids beginning with a methionine and ending with isoleucine and lysine, which are the first two amino acids of the α1 domain of the E_α polypeptide from a mouse of the k haplotype (*18*). It appears

Table 1. Homology comparisons of the α2 exon of the E^d_α gene with class II, class I, immunoglobulin, β_2-microglobulin, and Thy-1 sequences. Sequences used to determine percent homology with the α2 exon of E^d_α were as follows: the α2 exon of the human DR_α gene (*13, 22*); the β_2 exon of the human pDR-β-1 cDNA clone (*12*); the β_2 domain of a human DR_β chain (*11*); the α3 exons of the mouse H-2L^d gene (*28*), H-2K^b gene (*29*), the Qa (27.1) pseudogene (*23*) and a human HLA gene (*30*); immunoglobulin constant regions (*31–34*); β_2-microglobulin (*21, 35*), and Thy-1 (*36*). Percent homology is based on alignments of the compared sequences given by best-fit matrix analysis (*23*). Alignments are based solely on amino acid sequence comparisons. Each percentage involves the full length of the α2 domain or exon of E_α (94 amino acids and 282 bases, respectively) and a comparable length sequence from the other protein or gene being compared. Insertions and deletions for alignment were minimized, and percentages given are only the minimum values. Percentages given are statistically significant. Each was compared to an expected percent homology and standard deviation for any random alignment of sequence fragments the length of α2 from E_α and the test sequence. The expected value is the mean homology of all possible fragments the appropriate length from E_α to all possible fragments of the same length from either the test sequence or some other unrelated sequence. Each comparison is four or more standard deviations from the expected homology.

Gene name	Percent homology		Insertions and deletions*
	Protein	DNA	
DR_α	81	82	0
pDR-β-1	32	46	0
DR_β chain†	32	N.A.‡	1
H-2L^d	24	37	0
H-2K^b	23	38	0
Qa (27.1)	23	37	0
HLA	26	38	0
$C_\mu(C_H2)$	21	38	1
$C_\mu(C_H4)$	22	44	1
$C_{\delta 2b}(C_H1)$	27	40	2
C_κ	26	43	3
$C_{\lambda 1}$	30	40	2
β_2-microglobulin (mouse)	32	42	1
β^2-microglobulin (human)	31	43	1
Thy-1 chain†	20	N.A.‡	3

*Indicates the fewest number of insertions, deletions, or both, needed to align sequences for comparison. †These data are from protein sequences; no DNA sequences are available. ‡Sequence not available.

that the exon encoding the leader sequence and first two residues of E_α is separated from the α1 exon by an intervening sequence longer than ~ 1.5 kb.

About 250 base pairs (bp) of DNA has been sequenced 3′ to the third exon, and no polyadenylation signal has yet been observed. Indeed, we have compared the DNA sequence downstream of the termination codon of the E_α gene to the sequence of the 3′ untranslated region of the human DR_α chain cDNA clone (*13, 17*) using a computer homology search program. Except for 13 bp immediately adjacent to the termination codon, no significant homology was found for about 250 bp of sequence that could be compared up to the end of the 3.4-kb Sal I fragment. The gene for DR_α, the human homolog of E_α, contains at least four exons for the α1 domain, α2 domain, the transmembrane together with the cytoplasmic domain, and the 3′ untranslated region. The presumptive exon encoding the leader peptide has not been identified and is located more than 0.4 kb upstream of the α1 exon (*22*). Thus, the structure of the mouse E_α gene is in complete agreement with the structure of the human DR_α gene.

The E_α gene appears, by sequence analysis, to be a functional gene in that it lacks any obvious elements that would render it a pseudogene, such as termination codons or inappropriate reading frameshifts. In addition, as mentioned, all of the exon-intron boundaries have the consensus upstream or downstream RNA-splicing signals.

Several lines of evidence indicate that the class II gene we have described is the E_α gene. First, the translated sequence of the E_α gene is identical to the NH_2-terminal 34 residues of the E_α polypeptide isolated from C3H mice except for one residue, which may reflect genetic polymorphism (*18*). Of course, the NH_2-terminal two residues are missing from our sequence. Second, the E_α coding region shows a 79 percent DNA sequence homology and 75 percent protein sequence homology to its human counterpart, DR_α (*13, 14, 22*). A comparison, exon by exon, indicates that the α1 exons of E_α and DR_α are as similar to each other as are the α2 exons to each other, 81 percent and 82 percent, respectively, at the DNA level and 79 percent and 81 percent, respectively, at the protein level. The transmembrane cytoplasmic exons show less homology (73 percent at the DNA level and 57 percent at the protein level). However, a comparison at the level of amino acid functional types (nonpolar, polar, basic, and acidic) gives a match of 88 percent, indicating strong

```
GTCGACAACCAAGGAAGGAAGACAAGCAGAAAGGAAATGAGAAGAGTAGGGAAGCCAGCTCCAGTCACACAAGAGCAAGTCTGGGCTTCTATTGCCCAAGAAGGTTTTAGCGAGGCAGAT   120

GTTTTTCCTGTGCAGCAGAGCCAGAGGAAGAAGGAATATAGAAGATACGTTTGTGCATAAAGACAATCACTAAAATAAATCCCTTTGACTAAACATTAAAATCGGTCACCTTATTCCTAA   240

AGACAACACAGAAGCCCTCATCTCATCGTCCTTGTTGTTGGTTATCCTCCCTGTCAATCATAACAGACCATGTCCCACAAAAGTAGTCCTAGTCGAAGTAGTGGTTAGATAAAGGTCTAG   360

TTTCAAGGTCAACCCAGCTATGCCCTTCTAAAGACGTAGTTTGGGACGAGGATAAGCGATCCAGAGCATTGCAAGGTCTCAGTGTCAATCTCAGGTGCCTTGGGACGCTCATGTCTCTAA   480

GGGTTACTGGTTAAAGGAATTTGTAACCAAAAATTGTCAGCAGAGTTTGTTAGGTTTGGTGAGGGTACTGCAAGTTCTTGTCTCAAACCTATCCTGATAGAATGGGACTCAGCAGAGGTA   600

TGACAGGAAACTCCCTTATTAAGTAGACAGTAGTGATATGCTCAGTTAACAGGGTGAGTGTCACAATTGTAATCAGTTCCCAAGTTCCCAAAGGATCCTTTCAAATAACACTTCTCTACT   720

CGGAGATACCAGTGCCAGGATAAAAAGAAGCCCCTGTGTGCTGACCCAAGCTATTTAATTCCCTGCACAACAGGGAACAAGGGATGCTTTTTTCTGATGCACTGTAGCCAAATTTCAAAA   840

GTCATGAAGTCATTAAATACCCACTCAAATATGTTTTCTGAATCAACCTGCCACTCCAAGGCCAAAGGGACAGTGTAGGAGAGGGAACAGAAAGAACATGTGAGCCAGAGGATGGGGGTG   960

GGGGGACGACGCCATGAGAGATACTGTCTTCCAGACATGACATGACATTGCACACATGACTGAACTTGCAGTGGCTGCAGATATATGCACAAGACCTTCCCAGGCCGGCTAAACGATGTA  1080

GCATGGATGGGGAAGGAGCTCTTGAGAGACTGTTTACAGTTGAAGGTTGCTGATGGAGAGTATGTGTTTCAGGGGGTGGCCACTGGTTGGTGGTCATGCTGGAGTGCACATATGGACAAC  1200

AGTTGACTTGGTGCTAATAAAGAAGGCTATGTTGAAGGCTGTGTCTTGCAGCCCCATCCATTCATGTTCTAACACAAACAGCTCCCTTCTTTCTCACCTGCCTGCTTTACTCAGTATGAA  1320

TGTCCATCACCCAGACACTGTTTCTTCCCTAACTCCCATTCTATGCTCTTCCATCCCGATTGTCCCAGTCACCGTCCATCCCCGCCCACCCAAATGCCTGTGGGTTTTTCATCATCATCT  1440

                                                lu Glu His Thr Ile Ile Gln Ala Glu Phe Tyr Leu Leu Pro Asp Lys Arg Gly Glu Phe    22
TCTTCCATCTCCCTCCCTAATCCCCTTTCCTTTCTTTTCAG AG GAA CAC ACC ATC ATC CAG GCA GAG TTC TAT CTT TTA CCA GAC AAA CGT GGA GAG TTT  1540

Met Phe Asp Phe Asp Gly Asp Glu Ile Phe His Val Asp Ile Glu Lys Ser Glu Thr Ile Trp Arg Leu Glu Glu Phe Ala Lys Phe Ala    52
ATG TTT GAC TTT GAC GGC GAT GAG ATT TTC CAT GTA GAC ATT GAA AAG TCA GAG ACC ATC TGG AGA CTT GAA GAA TTT GCA AAG TTT GCC  1630

Ser Phe Glu Ala Gln Gly Ala Leu Ala Asn Ile Ala Val Asp Lys Ala Asn Leu Asp Val Met Lys Glu Arg Ser Asn Asn Thr Pro Asp    82
AGC TTT GAG GCT CAG GGT GCA CTG GCT AAT ATA GCT GTG GAC AAA GCT AAC CTG GAT GTC ATG AAA GAG CGT TCC AAC AAC ACT CCA GAT  1720

Ala Asn V                                                                                                                    83
GCC AAC G GTACCTGCCTCTCCTCCCTATCCCCTCCTCCCAGATGTGGGAACGCAGCTGTAAATAGATACTTTGGCGAATTCTATAAGGGTGTAAGGAGGTCTTGCTTGCCCTTCTAA  1837

GACCTAAGGTTTGCTCAGCACAGCAAATCTAATTCTCAGGCCACACAGGCCCAGGGATTTTAGAAGTTTTTTCCTTTTTTTTTTTTTTTTTTTTTTTTTGTAGTCTTGGTTTATTCATC  1957

AATTAACATTTATGAAAGCACTGACTTGAGTCTATACAGTGGGAGTGAGAAACTGGGATGGGGGCTTGGGTGGAATCGCGTGAGTGGTTGGATACCTCATACCTCAATACTGGCTCCAAA  2077

ATGTCAGATCTGGGACTAGGGTCACATGGACTGGACTCCTCAGGGAGCCGTTTCAGAGTCCTAGGTAGCATTGTACCACAGGACAGCATGGGCAGAGAGAAGCTGAAGGAATGAAAGCTA  2197

                al Ala Pro Glu Val Thr Val Leu Ser Arg Ser Pro Val Asn Leu Gly Glu Pro Asn Ile Leu Ile Cys Phe I    109
ACTTTATGCCATGTCCCCCACAG TG GCC CCA GAG GTG ACT GTA CTC TCC AGA AGC CCT GTG AAC CTG GGA GAG CCC AAC ATC CTC ATC TGT TTC A  2292

le Asp Lys Phe Ser Pro Pro Val Val Asn Val Thr Trp Leu Arg Asn Gly Arg Pro Val Thr Thr Gly Val Ser Glu Thr Val Phe Leu P    139
TT GAC AAG TTC TCC CCT CCA GTG GTC AAT GTC ACG TGG CTT CGG AAT GGA AGG CCT GTC ACG ACA GGC GTG TCA GAG ACA GTG TTT CTC C  2382

ro Arg Asp Asp His Leu Phe Arg Lys Phe His Tyr Leu Thr Phe Leu Pro Ser Thr Asp Asp Phe Tyr Asp Cys Glu Val Asp His Trp G    169
CG AGG GAC GAT CAC CTC TTC CGC AAA TTC CAC TAT CTG ACC TTC CTG CCC TCC ACA GAT GAT TTC TAT GAC TGT GAG GTG GAT CAC TGG G  2472

ly Leu Glu Glu Pro Leu Arg Lys Ala Trp G                                                                                      177
GC TTG GAG GAG CCT CTG CGG AAA GCC TGG G GTAGGGTGCAGTCATGACTTCACTTGCAGGCTCCCAGACTGCTTCAATGCTGTATTCTGGACACGTAGCATTTAACAGT  2581

ACAGGGATGGGAGCACAGTACTGTTAGAGTAATGGCCGTATAGGTGGTGGGGCTTCAAGCTCAATTACCAGCATCACAAACTAAGAGTTCACAGTGTTTCGTAACATTGACTACAATATA  2701

TACACCTTGGTTATCTTGTAAGGCACATTCATTCATTTTTGAAAAATTATTATGCATGTGGGTGTTTTGCCAGTGTGTATGACTATGCACCATGTAAGTGCCTGGTGCCCTCAGAGGTCA  2821

GAACAAGGTGTCCTATCCCCTGCACCTGGAGCTGTGTGCAGTTGTCAGCTGCCATATGGATACTGGAAATGAAACCTAGGTTCTCTGCAACAGCAGCCAGTGTTCTTAACGAATGAGTTG  2941

                                                                                lu Phe Glu Glu Lys Thr    184
TCCCTCCAGTCCCAGGACACACTCTCTTTCTTATATTGGCTACCCATGTTATTTCCTACAACATCAACTGACATCTCCTCTGCTTTATTTTCCCCAG AG TTT GAA GAG AAA ACC  3055

Leu Leu Pro Glu Thr Lys Glu Asn Val Met Cys Ala Leu Gly Leu Phe Val Gly Leu Val Gly Ile Val Val Gly Ile Ile Leu Ile Met    214
CTC CTC CCA GAA ACT AAA GAG AAT GTC ATG TGT GCT CTT GGG TTG TTT GTG GGT CTG GTG GGC ATC GTT GTG GGG ATT ATC CTC ATC ATG  3145

Lys Gly Ile Lys Lys Arg Asn Val Val Glu Arg Arg Gln Gly Ala Leu Trm                                                          230
AAG GGT ATT AAA AAA CGC AAT GTT GTA GAA CGC CGA CAA GGA GCC CTG TGA GATACCTGGAGGTGCGTTAAATGTGCTCAGAGACTGACAGATGTGTGAATGT  3248

CTGAGGGAGGAAAGCACGAGTGTGGTGCCTTTAAGAGAAGGGTAGGGGTAGTGGTGTCTCTTAATTCCTTTTGTTGGAAAAGTTGAGCTTTGAGGTTCAGATGCTTCCCAAACCTTCAGG  3368

ATCTGTGATCCTTCCTAGGTGTTCCTGGACCAGTTGTGAGTCTTGGAAATTTTCTTCAGTTCCCAAGACTGTCGAC                                              3444
```

Fig. 2. Nucleotide sequence of the 3.4-kb Sal I fragment containing the E^d_α gene. The amino acid translation of each exon is given above the nucleotide sequence. Certain amino acid residues are encoded by two exons and are so indicated by splits in the three-letter amino acid code. Abbreviations: A, adenine; C, cytosine; G, guanine; T, thymine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; Cys, cysteine; Gln, glutamine; Glu, glutamic acid; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; and Val, valine.

conservation for function. The direct confirmation that this class II gene is indeed the E_α gene will await gene transfer and expression studies.

We have used a graphically displayed computer routine termed the best-fit matrix analysis (*23*) to analyze possible similarities between the DNA and protein sequences of the E_α gene exons and between these exons and other class II genes, class I genes, immunoglobulin genes, and Thy-1 antigen. Such analyses showed no significant similarities between the different domains of E_α or between the α1 and transmembrane-cytoplasmic domains of E_α and anything other than the same regions of the DR_α cDNA. However, similarity alignment was possible between an area of each of the tested sequences and the α2 domain of E_α. Table 1 lists the sequences compared and the percent homology of the aligned regions to the α2 domain of E_α at both the protein and DNA levels.

The α2 domain of E_α has significant similarity to "homology unit" (*9*) sequences of the genes listed, a sequence associated with the "antibody fold" tertiary structure of antibody domains (Table 1). This observation has been made by several other groups analyzing cDNA (*13, 17*) or genomic (*22*) clones. Of the comparisons made, perhaps the most interesting is that to β_2-microglobulin. Not only is β_2-microglobulin as similar in sequence to any of the class II α2 and β2 domains as these are to each other, but it is strikingly similar in genomic organization to the E_α and DR_α genes. Like these two genes, β_2-microglobulin has its leader peptide and first two codons separated from the main protein coding sequence by a very large intervening sequence (2.8 kb). Of the non–class II exons compared in Table 1, only those of β_2-microglobulin align precisely end to end with those of E_α, employing the same split codon rule. The β_2-microglobulin gene also has the bulk of its 3′ untranslated sequence isolated as a distinct exon some distance 3′ to the last coding sequence (1.1 kb for β_2-microglobulin, 0.8 kb for DR_α).

Though Table 1 suggests that the domains compared diverged from a common anecestor, the evolutionary relationships between the entire genes is unclear. Non-α2-like sequences might have been under much different selective constraints and simply have diverged beyond recognition. It is also conceivable that the α2-like domain has been placed in different genomic contexts through the evolutionary process of exon shuffling (*24–27*). The organizational similarities between E_α and β_2-microglobulin suggest a more direct evolutionary relationship. Regardless, there appear to be fundamental evolutionary relationships among the three classes of genes that regulate and mediate immune responsiveness—Ia antigens, transplantation antigens, and immunoglobulins.

JANET MCNICHOLAS
Department of Biological Sciences, Stanford University, Stanford, California 94305
MICHAEL STEINMETZ
TIM HUNKAPILLER
Division of Biology, California Institute of Technology, Pasadena 91125
PATRICIA JONES
Department of Biological Sciences, Stanford University
LEROY HOOD
Division of Biology, California Institute of Technology

References and Notes

1. J. Klein, *Biology of the Mouse Histocompatibility Complex* (Springer-Verlag, Berlin, 1975).
2. G. D. Snell, J. Dausset, S. Nathenson, *Histocompatibility* (Academic Press, New York, 1976).
3. J. A. Frelinger, J. Niederhuber, D. C. Shreffler, *Science* **188**, 268 (1975).
4. R. H. Schwartz, C. S. David, D. H. Sachs, W. E. Paul, *J. Immunol.* **117**, 531 (1976).
5. E. A. Lerner, L. A. Matis, C. A. Janeway, Jr., P. P. Jones, R. H. Schwartz, D. B. Murphy, *J. Exp. Med.* **152**, 1085 (1980).
6. J. M. McNicholas, D. B. Murphy, L. A. Matis, R. H. Schwartz, E. A. Lerner, C. A. Janeway, Jr., P. P. Jones, *ibid.* **155**, 490 (1982).
7. L. A. Matis *et al.*, *ibid.*, p. 508.
8. D. B. Murphy, in *The Role of the Major Histocompatibility Complex in Immunobiology*, M. E. Dorf, Ed. (Garland, New York, 1981), pp. 1–32.
9. J. L. Strominger *et al.*, in *ibid.*, pp. 115–172.
10. J. Kaufman and J. L. Strominger, *Nature (London)* **297**, 694 (1982).
11. H. Kratzin *et al.*, *Hoppe-Seyler's Z. Physiol. Chem.* **362**, 1665 (1981).
12. D. Larhammar, L. Schenning, K. Gustafsson, K. Wiman, L. Claesson, L. Rask, P. A. Peterson, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3687 (1982).
13. D. Larhammar *et al.*, *Cell* **40**, 153 (1982).
14. M. Steinmetz *et al.*, *Nature (London)* **300**, 35 (1982).
15. A. M. Frischauf, H. Garoff, H. Lehrach, *Nucleic Acids Res.* **8**, 5541 (1980).
16. D. Fisher, M. Pecht, L. Hood, in preparation.
17. C. T. Wake, E. O. Long, M. Strubin, N. Gross, R. Accolla, S. Carrel, B. Mach, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
18. E. Sung, M. Hunkapiller, L. Hood, P. Jones, in preparation.
19. M. Steinmetz, K. W. Moore, J. G. Frelinger, B. T. Sher, F. W. Sher, E. A. Boyse, L. Hood, *Cell* **25**, 683 (1981).
20. P. A. Sharp, *ibid.* **23**, 643 (1981).
21. J. R. Parnes and J. G. Seidman, *ibid.* **29**, 661 (1982).
22. A. J. Korman, C. Auffray, A. Schamboeck, J. L. Strominger, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6013 (1982).
23. M. Steinmetz *et al.*, *Cell* **24**, 125 (1981).
24. W. F. Doolittle, *Nature (London)* **271**, 581 (1978).
25. J. E. Darnell, *Science* **202**, 1257 (1978).
26. W. Gilbert, *Nature (London)* **271**, 501 (1978).
27. F. Crick, *Science* **204**, 264 (1979).
28. K. W. Moore, B. T. Sher, Y. H. Sun, K. A. Eakle, L. Hood, *ibid.* **215**, 679 (1982).
29. A. A. Reyes, M. Schöld, K. Itakura, R. B. Wallace, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3270 (1982).
30. M. Malissen, B. Malissen, B. R. Jordan, *ibid.*, p. 893.
31. T. Kawakami, T. Takahashi, T. Honjo, *Nucleic Acids Res.* **8**, 3933 (1980).
32. Y. Yamawaki-Kataoka, T. Kataoka, N. Takahashi, M. Obata, T. Honjo, *Nature (London)* **283**, 786 (1980).
33. P. Hieter, E. Max, J. Seidman, J. Maizel, P. Leder, *Cell* **22**, 197 (1980).
34. A. Bothwell, M. Paskind, M. Reth, T. Imanishi-Kari, K. Rajewsky, D. Baltimore, *Nature (London)* **298**, 380 (1982).
35. S. V. Suggs, R. B. Wallace, T. Hirose, E. H. Kawashima, K. Itakura, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6613 (1981).
36. A. F. Williams and J. Gagnon, *Science* **216**, 696 (1982).
37. Supported by grants from the National Institutes of Health. We thank D. Fisher for his guidance throughout this work, especially for help with the deoxyribonuclease deletion subcloning and sequencing strategy. We thank A. Korman and J. Strominger, D. Larhammar, and P. Peterson for making the class II DNA sequences and preprints available to us prior to publication and for helpful discussions. We also thank K. Minard for preparation of sequencing gels and B. Larsh for preparation of the manuscript. J.M. is a postdoctoral fellow of the Arthritis Foundation. M.S. was supported by a Senior Lievre fellowship from the California Division of the American Cancer Society.

23 August 1982; revised 20 September 1982

Ban of DDT and Subsequent Recovery of Reproduction in Bald Eagles

Abstract. *Reproduction of bald eagles in northwestern Ontario declined from 1.26 young per breeding area in 1966 to a low of 0.46 in 1974 and then increased to 1.12 in 1981. Residues of DDE in addled eggs showed a significant inverse relation, confirming the effects of this toxicant on bald eagle reproduction at the population level and the effectiveness of the ban on DDT. The recovery from DDE contamination in bald eagles appears to be occurring much more rapidly than predicted.*

Low rates of reproduction in bald eagle (*Haliaeetus leucocephalus*) populations have caused concern for many years (*1, 2*) and led, in part, to declaring the species endangered. A variety of toxicants, particularly dichlorodiphenyldichloroethylene (DDE), have been implicated in the lowered productivity (*2, 3*). DDE and dichlorodiphenyldichloroethane (DDD) are metabolites of the insecticide dichlorodiphenyltrichloroethane (DDT). Of the three, DDE poses the greatest physiological threat to birds of prey and is the most persistent contaminant in the environment (*4, 5*) and in the bodies of birds (*6*). DDE is so stable in the environment that, in a controlled 11-year study, Beyer and Gish (*4*) were unable to calculate a half-life for the chemical. Evidence of the environmental problems associated with DDT and its metabolites led the Environmental Pro-

0036-8075/82/1217-1232$01.00/0