THE NUCLEOTIDE SEQUENCE OF A cDNA

FOR THE RAT Ia-A a CHAIN

By

ANNE ELIZABETH WALLIS


A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Department of Biochemistry)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

July 1983

© Anne Elizabeth Wallis, 1983
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of **Biochemistry**

The University of British Columbia  
1956 Main Mall  
Vancouver, Canada  
V6T 1Y3  

Date **Aug. 23, 1983**
ABSTRACT

The Ir genes of the Major Histocompatibility Complex have been shown to play an important role in the ability of an individual to produce an immune response. The products of the Ir genes, the Ia antigens, are highly homologous between species. To understand the structure and function of the Ia antigens it is necessary to examine the Ia molecules and the Ir genes which code for them both within and between species. The nucleotide sequence of a cDNA which corresponded to the 3' end of the Rat Ia-A α chain was determined. Comparison of this cDNA sequence with cDNA sequences from other species (mouse and human) showed that there was a much higher degree of sequence identity between the Ir gene products of equivalent loci between species than that observed between the Ir gene products of homologous loci within a species. This indicates that the duplication which gave rise to the homologous Ir gene loci occurred before the divergence of rodents and mammals.
### TABLE OF CONTENTS

**Introduction**  
Cellular Immunology of the MHC  
1. Discovery of the Immune Response (Ir) Genes  
2. Linkage of the Ir Genes to the MHC  
3. Function of the Ir Genes  
   (a) Role of the Ir Genes in T Helper Cell - Macrophage Interactions  
   (b) Role of the Ir Genes in T Helper Cell - B Cell Interactions  
   (c) The Ia Antigens  
4. The Ia Antigens are the Ir Gene Products**  

**Biochemistry and Molecular Genetics of the MHC**  
1. The Class I Molecules  
2. The Ia Molecules

**Materials and Methods**  
1. Plasmid pPRIa.2  
2. Plasmid DNA Preparation  
3. Isolation of cDNA Inserts  
4. Restriction Enzyme Mapping of cDNA Cloned Inserts  
5. Nick Translation  
6. Colony Hybridization  
7. Maxam and Gilbert Sequencing

**Results and Discussion**  
I DNA Sequence Analysis of the cDNA Insert of pPRIa.2  
II Predicted Amino Acid Sequence of the Carboxy Terminal End of Rat Ia-α Chain  
III Search for the 5' End of the Rat Ia-α Chain  
IV Conclusions  
Acknowledgement  
References
LIST OF TABLES

Table I  Maxam and Gilbert Modification Reactions  37

Table II  DNA Sequencing Gels  39

Table III  Comparison of the Sequence Identity Between the \( \alpha \) Chains of Rat Ia-A, Mouse H-2 I-A and Human HLA-DC1  53

Table IV  Comparison of the Amino Acid Sequences of the Transmembrane Regions of the Ia \( \alpha \) Chains of Different Species  58

Table V  Percent Sequence Identity Between Ig-Like Domains of Different Immune System Molecules  67
LIST OF FIGURES

Figure 1  MHC Restriction of Cellular Interactions in the Immune System  11

Figure 2  The Restriction Map and Sequencing Strategy of the cDNA Insert of pRIa.2  42

Figure 3  The Nucleotide Sequence of the cDNA Insert of pRIa.2  44

Figure 4  Comparison of the Nucleotide Sequences of the cDNA Encoding the \( \alpha \) Chains of Rat Ia-A, Mouse H-2 I-A and Human HLA-DC1  47

Figure 5  Translation of the Coding Sequence of the cDNA Insert of pRIa.2  49

Figure 6  Comparison of the Amino Acid Sequences of the \( \alpha \) Chains of Rat Ia-A, Mouse H-2 I-A and Human HLA-DC1  52

Figure 7  Diagram of an Immunoglobulin Constant Region Domain  62
Figure 8  Comparison of the Amino Acid Sequences of Ig-Like Domains  64

Figure 9  Schematic Representations of Some Members of the Immunoglobulin Superfamily  69

Figure 10  Alignment of the Partial Nucleotide Sequence of the cDNA Insert of pRIa.3 with the Sequence of the cDNA Insert of pRIa.2  72
ACKNOWLEDGEMENT

I would like to gratefully acknowledge the support and advice I received from Rob McMaster, Sarah Eccles and Heather Wallis.
INTRODUCTION

The Major Histocompatibility Complex

During early studies on tumour transplantation and skin grafting it became apparent that there was a genetic basis to transplantation or grafting. This phenomenon of tumour or graft rejection was found to be regulated by a number of independently segregating genes. Snell, in the 1940's, developed congenic mouse strains which differed only at regions of the genome which apparently regulated tumour or graft rejection. At the same time, Gorer, using antisera developed against these congenic strains defined a group of antigens which segregated with the genes responsible for the susceptibility and resistance to tumour transplantation. These transplantation or histocompatibility antigens correlated with the same area on the genome as the tumour susceptibility genes. The region of the mouse genome to which both the tumour susceptibility genes and the histocompatibility antigens mapped was called the H-2 complex. The H-2 complex was found to be the major region of the genome which determined the ability of an animal to reject grafts or transplants and because of this it was termed the Major Histocompatibility Complex (MHC).

Regions homologous to the mouse H-2 complex have been identified in other species including Rats (RT-1), humans (HLA), chickens (B) and pigs (SLA) (Gotze, 1977).
1. Discovery of the Immune Response (Ir) Genes

The existence of a genetically transmissible variable which resulted in unresponsiveness to certain antigens in animals usually responsive to this antigen, was noted from the 1940's onward (Carlinfanti, 1948; Sang and Sobey, 1954; Ipsen, 1959; Sobey et al., 1966). Early studies to examine this genetically controlled immune responsiveness, were done in guinea pig strains using large foreign protein molecules as the challenging antigen. The use of these large multi-determinant proteins as antigen gave results that were indicative of a trend in immune responsiveness but which were often not clear cut.

To demonstrate the strict genetic control of the immune responsiveness to limited or even single antigenic determinants, Levine et al. (1963) studied the response of two inbred guinea pig strains (strain 2 and strain 13) to synthetic antigens composed of a specific hapten conjugated to the carrier poly-L-lysine (PLL). A carrier is defined as a molecule which can be recognized as immunogenic by the immune system. A hapten is defined as a molecule, usually a chemical group, which cannot provoke an immune response when injected by itself but can act as an immunogen when linked to a carrier. By simplifying the antigen used, it was hoped that a more clear cut set of data would result.

In this study, the hapten-carrier conjugate 2,4-dinitrophenol (DNP) - PLL was used to immunize guinea pigs of either strain 2 or strain 13. Strain 2 guinea pigs showed a
response as measured by the presence of serum anti-DNP antibodies and by the occurrence of a delayed hypersensitivity response. Strain 13 guinea pigs, on the other hand, were non-responders exhibiting neither a delayed hypersensitivity response nor any production of anti-DNP antibodies. Breeding experiments were done with the two guinea pig strains. The first filial generation (F1) of strain 2 guinea pigs (responders) and strain 13 guinea pigs (non-responders) resulted in progeny which were all responders. Crossing two F1 guinea pigs resulted in a F2 generation which were 75% responders and 25% non-responders. This indicated that the gene controlling the immune response to DNP-PLL was a single dominant gene. Work by McDevitt and Sela (1965) using similar techniques in mice supported these observations.

In a further study Benacerraf et al. (1967) showed that different haptens conjugated to PLL showed a similar trend of response or non-response and that immune responsiveness to PLL was linked to the strain of guinea pig used. The gene controlling this immune response was, therefore, named the PLL gene. By linking the DNP hapten to another carrier, such as bovine serum albumin (BSA) or ovalbumin (OVA), it was shown that the non-responder strains could recognize and respond immunologically to the hapten DNP when it was linked to BSA or OVA (Benacerraf et al., 1967). This demonstrated that the non-responder animals had the capability of producing antibodies to DNP if it was presented to the immune system on an immunogenic carrier. This observation also indicated that the critical recognition step was the recognition of carrier. The antibody
response which resulted, however, was not primarily directed against the carrier itself, but was primarily specific for the hapten conjugated to it. This argued that although the PLL gene was responsible in some way for the process of antigen recognition, it was not directly involved in identification of the specific determinants to which the antibody produced would recognize. The PLL gene and other genes shown to control the ability of an animal to mount an immune response to a specific antigen were termed immune response or Ir genes.

2. Linkage of the Ir Genes to the MHC

The availability of inbred mouse and guinea pig strains enabled experiments to be carried out to analyse the location of the Ir genes on the genome.

McDevitt and Chinitz (1969), showed that the Ir genes were linked to the H-2 complex. Linkage mapping studies in inbred mouse strains localized the site of one of the Ir genes, the Ir-1 gene, to the middle of the mouse H-2. The region of the MHC to which the Ir genes mapped was called the I region (McDevitt et al., 1972). Evidence from studies in guinea pigs also showed that the Ir genes were linked to the MHC (Ellman et al., 1970). Studies of the Ir genes in other species such as the rat (Amerding et al., 1974) and the rhesus monkey (Dorf et al., 1975) established that the Ir genes were linked to the MHC of these species as well. Subsequent studies indicated that different mouse Ir genes mapped to different areas of the I region, these were called the I-A and the I-E subregions (Klein, 1981). Once it had been shown that the loci for the Ir genes were linked to the
MHC many investigations were done searching for a correlation between the Ir genes and the histocompatibility antigens coded by the MHC.

3. Function of the Ir Genes

The Ir genes were known to determine the ability of an organism to respond to an antigen. The mechanism by which this regulation occurred, however, was unclear. In order to understand further the role of the Ir genes in the immune response, experiments were done to determine the pattern of expression of the Ir genes in cells of the immune system.

At this time it was known that for a humoral immune response to occur an antigen-specific interaction between T helper cells and B cells must occur. This went partway towards explaining the dichotomy of the hapten-carrier conjugate responses, in which the immune response was specific for the carrier, even though the antibody response that was stimulated was directed against the hapten. If the two lymphocyte subpopulations involved recognized different parts of the same antigen this apparent contradiction would be explained. If it was the T helper cell which was responsible for recognition of the carrier part of the antigen as immunogenic then it was the T helper cell which regulated the B cell response or lack of response to hapten. The ability of the T helper cell to recognize a carrier as immunogenic appeared to be determined by the Ir genes the animal possessed.

Since it appeared that it was the T helper cell which regulated immune responsiveness, it seemed probable that the
expression of Ir genes occurred at the T helper cell level and many experiments were done aimed at validating this supposition.

(a) Role of the Ir Genes in T helper cell - Macrophage Interactions

Experiments had shown that the response of lymphocytes to antigen required the presence of macrophages. To clarify further the exact nature of the role of macrophages, Lipsky and Rosenthal (1973) used guinea pigs to investigate the interaction between macrophages and lymphocytes in the absence of antigen. They demonstrated that a physical cell-cell contact occurred and that there was a specificity for the cell types involved: macrophages bound only to thymocytes or lymphocytes and no other cells of the body and thymocytes from mice did not bind to guinea pig macrophages. There appeared to be an unique cellular recognition mechanism between thymocytes and macrophages. In the presence of antigen the interactions between macrophages and lymphocytes increased and it seemed that a direct cell-cell contact was necessary for cellular cooperation leading to antigen-specific lymphocyte stimulation. Further studies showed that for the antigen to be in an immunogenic form for T cell recognition, the macrophage was required to process the antigen in some way that involved the use of metabolic energy.

The existence of mouse and guinea pig strains that were congenic at the I region of the MHC allowed the production of antisera directed against different I region specificities. This antisera was produced by cross immunization of these congenic strains with lymphoid tissue. Studies done by Shevach et al.
(1972) looked at the relationship between Ir genes and the mechanism of immune responsiveness. Lymphocyte proliferation was used to assay T helper cell response. The effect of alloantibody directed against the I region of the two inbred guinea pig strains, 2 and 13, on T helper cell function was tested. The results of this study showed that the alloantibody specifically blocked the activation of T lymphocytes by antigens that were under Ir gene control.

To see if the addition of anti-I region alloantibody was simply blocking or inhibiting the antigen uptake by macrophages, Shevach et al. (1973) incubated antigen plus macrophages with and without alloantibody before assaying for Ir gene control by in vitro T cell proliferation. This experiment showed that the alloantibody had no effect on the ability of macrophages to take up and process antigen for presentation to T helper cells. The most acceptable explanation for such results was that the alloantibody inhibited antigen-induced proliferation of T helper cells by blocking recognition of antigen which has been bound to or processed by macrophages. It was also apparent that the Ir genes coded for a cell surface associated product and that this product was somehow involved in the mechanism of antigen recognition by T helper cells.

The immune response seemed to involve a multi-step pathway involving the two subsets of lymphocytes, the T and B cells, macrophages and antigen. The general theme of proposals put forward to explain the pathway of an immune response went as follows: foreign antigen enters the body and reacts with B cells which express immunoglobulin specific for that antigen on their
cell surface. In order for B cells to proliferate and produce antigen-specific antibody a second signal is necessary. Antigen also binds to or is in some way processed by macrophages. Once antigen has been appropriately processed it is presented to T cells. Depending on the ability of T cells to recognize the macrophage-processed antigen, or the ability of macrophages to process antigen, T cells will proliferate and send a second signal to B cells leading to the induction of specific antibody synthesis.

The evidence of Shevach and Rosenthal (1973) established that although the functional role of the Ir gene product is in the process or mechanism of antigen recognition by T lymphocytes, this function did not necessarily require the presence of the Ir gene product on the surface of the T cell. They offered the alternative explanation that the Ir genes might be expressed on the surface of a cell that controls T cell activation, for example the macrophage. To investigate this proposal Shevach and Rosenthal examined whether (1) macrophages from non-responder animals would stimulate T lymphocytes of (responder X non-responder) F1 animals and (2) whether anti-I region alloantibody inhibited by blocking macrophage-T cell interaction nonspecifically, by specifically blocking Ir gene products or by inhibiting both functions. The results of this study, showed that it was necessary for T cells and macrophages to be compatible at some I region loci to enable macrophages and antigen to interact with immune T helper cells.

Further studies showed that T helper cells exhibited a dual specificity in that recognition of antigen on the macrophage
surface was specific for that antigen but only when the antigen was presented on the surface of macrophages which were I region compatible with the T helper cells (Sprent, 1978a). The requirement that T helper cells and macrophages could interact only when they shared MHC identity at the I region and that antigen could be responded to only in the context of this shared identity was termed MHC restriction.

(b) Role of the Ir Genes in T Helper Cell - B Cell Interactions

MHC restriction of the antigen-specific interaction between T helper cells and macrophages was also shown to apply to the interaction between T helper cells and B cells. Experiments by Kindred and Shreffler (1972) and Katz et al. (1973), demonstrated that histoincompatible T lymphocytes and B cells were unable to cooperate and produce an antibody response. Carrier-primed T helper cells could not produce the appropriate stimulus for induction of antibody synthesis in hapten-primed B cells unless the T and B cells were fully or semi-histocompatible. Therefore, there was a failure in antigen specific T cell-B cell cooperation across MHC barriers. Kappler and Marrack (1976, 1978) and Sprent (1978a,b) confirmed these results using different experimental techniques.
(c) The Ia antigens

Shreffler and David (1975) in mice and Schwartz et al. (1976) in guinea pigs found that anti-I region antisera reacted with a group of cell surface glycoprotein molecules expressed on B lymphocytes and macrophages. Because of the unclear relationship of these alloantigens to the Ir genes these alloantigens were termed immune-associated or Ia antigens. The Ia antigens were shown to be highly polymorphic (Shreffler and David, 1975) and to consist of two non-covalently associated glycoprotein chains (Cullen et al., 1976). The Ia antigens will be discussed in detail in a further section.

In Figure 1 the mechanism of MHC restriction in T helper cell-macrophage and T helper cell-B cell interactions are diagrammed.
In this diagram the dual specificity of the T helper cell (Th) is depicted. In order for an immune response to occur antigen must be presented to the T helper cell in association with immune associated antigens (Ia) on the macrophage cell surface. A T helper cell which has been able to interact with a histocompatible macrophage becomes stimulated and can then interact antigen-specifically with a B cell (B) which presents the antigen in association with the same histocompatible Ia antigen. The immune response to antigen is said to be MHC restricted because of the dual specificity of the T helper cell.
4. The Ia antigens are the Ir gene Products

Dorf and Benacerraf (1975) showed, in mice, that the immune response to some antigens was under the control of two dominant genes. When two non-responders animals were bred the resulting progeny were responders. The non-responder alleles, therefore, appeared to complement one another resulting in offspring with a responder phenotype. One gene was shown to map to the I-A region of the H-2 complex (see section 2). It seemed that for an immune response to some antigens both alleles had to be present. Dorf and Benacerraf further showed that if the I region haplotype of an animal was known it was possible to predict whether that animal would be a responder to a specific antigen.

Using the same strains as Dorf and Benacerraf, Jones et al. (1978) found evidence that the expression of one set of Ia antigens on the cell surface was under the control of two genes. There appeared to be a correlation between the Ir genes and two I region gene products which controlled Ia expression. Jones et al. found that one component of the Ia antigen mapped to the I-A region and that a second locus which mapped to the I-E region appeared to control whether this I-A encoded molecule was expressed on the cell surface. They also showed that anti-I-E antisera coprecipitated a molecule which consisted of two non-covalently associated polypeptide chains in the F1 and that this Ia antigen was not present in either of the parental strains.

Taken together these results indicated that the Ia I-E antigen in the mouse, was composed of two non-covalently associated polypeptide chains and that one subunit (β) was encoded by the I-A subregion and the other subunit (α) by the I-
E subregion. Both the $\alpha$ and the $\beta$ chain were required to produce a functional Ia antigen and concommitent with expression of a complete Ia I-E antigen the immune responsiveness to some antigens appeared.

Additional evidence that the Ia antigens were the Ir gene products came from experiments in which monoclonal anti-Ia antibody was found to specifically block T helper cell-macrophage interactions \textit{in vitro}. The T cell response to antigens which were under the control of one Ir locus were inhibited only by antibody directed against that locus and this same antibody had no effect on the T cell response to antigen under the control of other Ir loci (Lerner et al., 1980; Baxevanis et al., 1980; Sredni et al. 1981).

MHC restriction of T helper cell interactions in the immune response results from the dual specificity of T helper cells. This dual specificity requires that a specific antigen be recognized only in the context of an Ia antigen.
Serological data had defined several types of molecules which were encoded by the MHC. Genetic studies, done by serologically analysing lymphocytes from mice strains where recombination events occurred within the MHC, divided the mouse MHC into five subregions: the K, I, S, G, and D regions. The D region was subsequently split into the D and L regions based on serological findings by Hansen and Levy 1978. The G region was shown to be identical to that of the S region which encoded the fourth component of the complement pathway (C4) (Huang and Klein, 1980; Ferriera et al., 1980). The K, D and L regions encode the Class I molecules of the MHC. The I region encodes the Ia antigens as previously discussed.

THE CLASS I MOLECULES

The earliest studies on the K and D region gene products were done by Shimada and Nathanson (1969). Specific antibodies were used to immunoprecipitate either H-2K or H-2D molecules after solublization by papain treatment of spleen cells to release the molecules from the cell surface. The proteins identified in this way were found to be of 37,000 molecular weight. Later studies (Schwartz et al., 1973) using radioactive labelling techniques to increase the sensitivity and overcome the problem of limited amounts of protein, showed that the K and D molecules consisted of two noncovalently linked chains. The heavy chain was a glycoprotein of 37,000 molecular weight when
papain treatment was used, and of 44,000 molecular weight when the membrane was solublized by detergent. The determinant recognized by anti-H-2K or anti-H-2D antibodies was found to reside within the larger 44,000 molecular weight polypeptide chain. The light chain associated with the K or D region encoded polypeptide was shown to be $\beta_2$-microglobulin (Natori et al., 1974). This polypeptide had a molecular weight of 12,000 and was not encoded by the MHC. The heavy chain was shown to contain two carbohydrate chains (Muramatsu and Nathenson, 1970). Studies to determine whether the location of the antigenic determinant(s) recognized by the antibodies were on the protein or carbohydrate moieties of the heavy chain, showed that the antibodies reacted with the polypeptide moiety (Nathenson and Cullen, 1974). The H-2K or D antigens are now called the Class I molecules of the MHC. The use of antibodies directed against K and D region haplotype differences showed that K and D loci were polymorphic. The number of different alleles at each of the K and D loci is now thought to be in the range of 30-60. The Class I molecules are found on most tissues.

The Class I molecules in humans are called HLA-A, HLA-B and HLA-C antigens. Orr et al. (1979) reported the complete amino acid sequence of a papain solublized HLA-B7 molecule. This molecule was 271 residues in length, had a single carbohydrate moiety attached at amino acid 86 and had 2 disulphide loops. A search for internal homologies by computer analysis suggested homology between the amino terminal 90 amino acids and the region of the first disulphide loop (residues 91-180). Significant homology between the second disulphide loop
(residues 182-271) and immunoglobin (Ig) constant domains and $\beta_2$ microglobulin was also shown.

Coligan et al. (1981) reported the first complete protein sequence of a Class I antigen. Using radiochemical techniques, the murine H-2K$^b$ chain was completely sequenced. The protein was shown to be 346 amino acids in length. The H-2K antigen could be divided into 3 functional segments; the extracellular region, the transmembrane region and the cytoplasmic region. A transmembrane segment of 25 uncharged hydrophobic amino acids was identified starting at amino acid 282 and extending approximately to amino acid 307. This agreed with sequence data which had been derived from papain treated human Class I molecules (Martinko, 1980) which were 281 amino acids in length. The region from amino acid 308 to the carboxy terminus at amino acid 346 represented a 38 amino acid long intracellular or cytoplasmic tail. The H-2K$^b$ molecule contained two sites for carbohydrate attachment and had two intra-chain disulphide loops. By analogy with the human HLA-B7 molecule, which was approximately 70% homologous to H-2K$^b$, the extracellular segment could be divided into 3 domains; the amino-terminal domain $\alpha_1$, the first disulphide loop domain $\alpha_2$ and the second disulphide loop domain $\alpha_3$ which was adjacent to the membrane. Studies on the Class I antigens suggested that the 3 external domains were organized into $\beta$-pleated sheet structures (Uehara, 1980a,b; Martinko, 1980). Lack of sufficient amounts of protein, however, limited the data which could be derived from these studies.

Recombinant plasmids containing cDNA inserts corresponding to Class I antigen mRNAs were isolated for both human
transplantation antigens (Ploegh et al., 1980; Sood et al., 1981) and for mouse transplantation antigens (Kvist et al., 1981; Steinmetz et al., 1981a). Kvist et al. isolated a recombinant plasmid containing an approximately 1000 base pair cDNA insert, which corresponded to the carboxy-terminal half of a murine Class I antigen. The cDNA isolated by Kvist et al. was sequenced (Bregegere et al., 1981) and used as a probe to isolate other Class I specific clones (Jordan, 1981). Sood et al. (1981) using oligodeoxyribonucleotide primers isolated a cDNA clone for an HLA-B Class I antigen.

Several genomic clones were isolated by the use of cDNA probes (Steinmetz et al., 1981b; Jordan et al., 1981; Singer et al., 1982). Because of the high degree of homology between Class I antigens of mice and humans, it was found that cDNA clones for Class I antigens of one species could identify Class I antigens of other species (Singer et al., 1982). The first sequence of genomic DNA which was reported encoded what appeared to be a murine pseudogene homologous to the Class I antigens (Steinmetz et al., 1981b). This pseudogene consisted of 8 exons, 7 of which correlated with the domain structure of the protein. The amino acid sequence from exon 1 was thought to represent a leader sequence. Exons 2, 3 and 4 correlated with the domain structures predicted from early protein and cDNA sequence data (Coligan et al., 1981; Bregegere et al., 1981). Exon 5 corresponded to the transmembrane region. The last 3 exons coded for the cytoplasmic portion of the protein and the 3' untranslated region. The need for the presence of 3 domains for the cytoplasmic region of the protein was unclear, as the cytoplasmic tail appeared to
comprise only one functional region.

One interesting aspect of the sequences which have been elucidated is the similarity between products of homologous loci for the Class I antigens i.e. it is possible to determine, from sequence alone, whether a sequence is from the human HLA or mouse H-2 complex. It is not possible to determine from which loci within a species a sequence is derived. Therefore, as well as a great deal of diversity at each Class I locus there appears to be a 'homogeneity' which renders each allelic product equally like or unlike another allelic or gene product.

Steinmetz et al. (1981a) carried out a Southern blot analysis of mouse DNA using a known Class I cDNA probe. Approximately fifteen bands were identified which hybridized to the probe DNA. This indicated that the Class I antigens comprised a multigene family. An elegant technique to identify a cloned H-2 Class I gene was used by Moore et al.(1982) and Goodenow et al. (1983). Moore et al. used DNA-mediated gene transfer to transform a mouse cell line, haplotype H-2^k, with a genomic clone of 27.5kb which hybridized strongly to a Class I cDNA probe of haplotype d. The resulting positive transformants were detected by radioimmunoassay using anti-H-2^d monoclonal antibodies. The results indicated that the genomic clone contained a functional H-2L gene and that the H-2L^d molecule was expressed correctly on the recipient cell-surface. This method, therefore, not only allows unambiguous assignment of cloned genes but may provide a system to study the function of the Class I antigens. This technique was also used by Singer et al. in 1982 to examine expression of a genomic clone for the
porcine Class I or SLA antigens, and by Mellor et al. (1982) to look at expression of an H-2K\(^b\) Class I antigen.

THE Ia MOLECULES

The I region of the mouse H-2 was originally subdivided into five subregions: A,B,J,E, and C (Shreffler and David, 1975). At present however, it appears that there may actually be only two I subregions: the A and E (Klein et al., 1981; Robertson, 1982). The Ia antigens in mice are called the I-A and I-E and in rats the Ia-A and Ia-E antigens. These correspond respectively to the human HLA-DC\(^1\) (Bono and Strominger, 1982; Goyert et al., 1982) and HLA-DR (Allison et al., 1978) antigens.

Similar techniques to those described for the Class I molecules were used to purify and characterize the molecules coded for by the I region. The Ia molecules were shown (Silver et al., 1977) to consist of two non-covalently linked polypeptide chains, both of which were glycosylated. The mouse Ia heavy chain (\(\alpha\)) had a molecular weight of 33,000 and the light chain (\(\beta\)) a molecular weight of 26,000. The antigenic determinants recognized by antibodies were shown to be on the polypeptide portions of the molecules (Cullen et al., 1975). Immunoprecipitation studies also showed that a single Ia molecule expressed more than one determinant (Cullen et al., 1976).

Due to the small amount of Ia protein available only limited protein sequencing could be done. Limited amino-terminal sequence data showed that in the mouse there were two distinct \(\alpha\)
chains and two distinct $\beta$ chains (Silver et al., 1975, 1977). Antigenic specificites determined by I-A and I-E subregions were found on independent molecules (Jones et al., 1978). Thus, each subregion appeared to code for a distinct cell-surface molecule. Crossreaction using alloantibodies for the different subregions was noted (Murphy et al., 1975; Murphy and Shreffler, 1975). This indicated that a degree of homology between the different A region and E region molecules existed. In contrast to the Class I antigens, the Ia molecules had a restricted tissue distribution, they are found primarily on B cells, macrophages, thymic epithelium and some T helper cells.

Larhammar et al. (1981), described the cloning of a cDNA clone of approximately 1000 base pairs corresponding to an HLA-DR $\beta$-like chain of approximately 230 amino acids. The amino acid sequence which was predicted from this almost full length cloned cDNA displayed sequence homology to the human Class I antigen heavy chains, $\beta_2$ microglobulin and immunoglobulin constant region domains. This was especially interesting because the original protein data from amino-terminal sequencing had revealed no homology to the molecules listed above.

Isolation and sequencing of HLA-DR $\alpha$ and $\beta$ chain DNA was reported by Lee et al. (1982a) Wiman et al. (1982), Korman et al. (1982) and Larhammar et al. (1982a,b). The mature HLA-DR-like $\beta$ chain was shown to be 229 amino acids in length, containing two putative disulphide loops, a 21 amino acid transmembrane segment and a 10 amino acid cytoplasmic tail. There was a single putative site for carbohydrate attachment within the amino-terminal Ig-like region. This amino-terminal
domain exhibited homology to the corresponding regions of the HLA-A, B and C antigen heavy chains. The membrane proximal Ig-like region was shown to be homologous to β2-microglobulin, Ig constant region domains and the Ig-like domains of the Class I antigens (Larhammar, 1982a).

The mature HLA-DR α chain was also shown to be 229 amino acids in length. The transmembrane region was 23 amino acids long and there was a 15 residue cytoplasmic tail. The remaining 191 amino acids were exposed on the cell surface. The HLA-DR α chain was shown to be composed of two extracellular domains; α1 the amino-terminal domain (residues 1-84) and α2, an Ig-like disulphide containing domain (residues 85-179). The connecting peptide, transmembrane region and cytoplasmic tail comprise another third domain.

A cDNA corresponding to the HLA-DC1 α chain was described by Auffray et al. (1982). The cDNA for the HLA-DC1 α chain was isolated by using an HLA-DR α chain cDNA probe. The protein encoded by this isolated cDNA was shown to be 232 amino acids in length and to correspond to the HLA-DC1 α chain by comparison with the known amino-terminus protein sequence data for HLA-DC1 α chain. The HLA-DC1 α chain and the HLA-DR α chain exhibit approximately 60% homology. By analogy to the domain structure of the HLA-DR α chain protein the HLA-DC1 α chain protein was divided into two extracellular domains; α1(amino acids 1-87) and α2(amino acids 195-217), a connecting peptide (amino acid 182-194) a transmembrane region (amino acids 195-217) and an intracytoplasmic region (amino acids 218-232). The HLA-DC1 α chain had two sites for carbohydrate attachment one in each of
the a1 and a2 domains. Like the HLA-DR a2 domain, the HLA-DC1 a2 domain contains the only disulphide loop and exhibits homology to the Ig constant region domains. Benoist et al. (1983), described cDNA's corresponding to the murine I-A and I-E a chains.

The highly polymorphic nature of the MHC encoded molecules is unique and it is of interest to try to understand the evolution of this polymorphism. One way in which this can be done is to compare the sequences of MHC molecules of other species such as the rat, to the equivalent sequences of mouse and human molecules.
MATERIALS AND METHODS

1. Plasmid pRIa.2

Plasmid pRIa.2 (supplied by Dr. W.R. McMaster) was constructed in the following manner. Total RNA was prepared from spleens of Wistar strain rats (haplotype RT1\(^v\)), using 7.5M guanidium-HCl and the procedures of Chirgwin et al. (1979). Poly-adenylated RNA (Poly(A)-RNA) was purified by chromatography on oligo (dT)-cellulose (Aviv and Leder, 1972). Spleen poly(A)-RNA was fractionated by sucrose density gradient centrifugation. Fractions containing mRNA were translated in vitro in the presence of radioactive amino acids using a cell-free rabbit reticulocyte lysate system. The Ia polypeptides were immunoprecipitated with rabbit antibodies prepared against rat Ia-A\(a\) chain and analysed by SDS-PAGE electrophoresis. Fractions of mRNA coding for rat Ia-A\(a\) chain were used to prepare double stranded cDNA as described by Land et al. (1981). cDNA was inserted into the Pst I site of the bacterial plasmid pBR322 (Bolivar and Backman, 1979) by G/C tailing (Land et al., 1981) and the resulting recombinant plasmids used to transform E.coli strain RR1 (Bolivar and Backman, 1979). Specific cDNA clones containing sequence corresponding to Ia-\(A\\)\(a\) chains were detected by using a positive mRNA selection assay (Parnes et al., 1981). The assay involved immobilising pools of plasmid DNA onto nitrocellulose filters and hybridising total mRNA to the immobilised DNA. mRNA which hybridised was eluted and the mRNA coding for Ia polypeptides was detected by cell-free translation.
and immunoprecipitation as described above. Pools of plasmid DNA containing cDNA inserts which hybridised to mRNA for Ia polypeptides could then be screened individually. This strategy resulted in the identification of a recombinant plasmid, pRIa.1, which contained a cDNA insert of approximately 600 base pairs, coding for a rat Ia-A α chain.

A cDNA library was prepared from total rat poly(A)-RNA as described above (R. McMaster, unpublished). This cDNA library was screened for the presence of cDNA inserts containing sequences which hybridised to the cDNA insert of plasmid pRIa.1. The plasmid pRIa.2, which contained a 800 base pair cDNA insert coding for a rat Ia I-A α chain was identified in this way.

2. Plasmid DNA Preparation

(Bolivar and Backman, 1979)

Media and Solutions:

Luria-Bertani(LB) Medium (1.0g Bacto dextrose, 5.0g Bacto yeast extract, 10.0g Bacto tryptone, 10.0g NaCl made up to 1 litre with distilled water (dH₂O) and the pH adjusted to pH 7.2. All Bacto products were from Difco Laboratories, Detroit, Mich.)

TEN-8 (100mM NaCl, 10mM Tris(hydroxymethyl)aminomethane-HCl (Tris-HCl) pH 8.0, 1mM disodiummethylenediaminetetra-acetic acid (EDTA). pH to 8.0 )

Lysis Solution (20% Sucrose, 50mM Tris-HCl pH 8.0)

Agarose Gel Electrophoresis Buffer 40mM Tris-HCl, 20mM acetic acid, 2mM EDTA. pH to 8.0 )
**Method:** One litre of LB medium containing tetracycline (10µg/ml) was inoculated with 10ml of fresh overnight cultures of *E.coli* strain RR1 cells containing plasmid pRIa.2 and grown at 37°C in a shaking water bath. At an absorbance at 600nm of 0.7, 1.2ml of freshly made chloramphenicol (160mg/ml in 95% ethanol) was added and the cells were shaken for 12 to 16 hours.

Cells were centrifuged at 5000 rpm for 15 minutes in a GSA rotor in a Sorvall RC-5B centrifuge. Each pellet was resuspended in 25.0ml sterile TEN-8, transferred to 50ml polycarbonate tubes and centrifuged in a SS-34 rotor at 5000 rpm for 15 minutes.

Each pellet was resuspended in 20.0ml of lysis solution and 1.0ml freshly dissolved lysozyme (10mg/ml in 150mM Tris-HCl pH 8.0) and 1.0ml 0.25M EDTA pH 8.0 were added. The mixture was swirled and left on ice for 15 minutes. To each tube was added 1.0ml ribonuclease-A (Sigma type I-A, St. Louis, Mo., 2 mg/ml in dH₂O, previously heated to 100°C for 5 minutes) and 10.0ml 10% v/v Triton X-100 (BioRad, Richmond, Calif.) in dH₂O. The mixture was swirled and left at room temperature for 5 minutes. The lysed cells were centrifuged at 15000 rpm for 30 minutes in a SS-34 rotor. The resulting supernatants were extracted with 1/2 volume phenol and 1/2 volume chloroform:isoamylalcohol (24:1 v/v). The DNA in the aqueous layer was precipitated by adding 1/10 volume 2M NaCl and one volume 100% isopropanol and being placed at -20°C for a minimum of 1 hour. Precipitated DNA was collected by centrifugation at 2700 rpm in a Beckman TJ-6 benchtop centrifuge for 30 minutes, air dried for 10 minutes and dissolved in 1.0ml TEN-8. Plasmid DNA was reprecipitated by the addition of 1/10 volume 2M NaCl and three volumes 95% ethanol.
and storing at -20°C for 4 to 16 hours. The resulting precipitates were washed with 95% ethanol, dried under vacuum and dissolved in 1.0ml TEN-8. Plasmid DNA was then chromatographed on a Sephacryl-1000 (Pharmacia) column (100 cm x 1.6 cm) run in TEN-8 to separate tRNA from plasmid DNA. The absorbance at 260nm of each fraction (2ml) was recorded and aliquots from fractions containing plasmid DNA were analysed by electrophoresis on a horizontal 1% agarose gel (170x120x5mm) run at 150 volts for 40 minutes in agarose gel electrophoresis buffer. Fractions containing plasmid DNA were precipitated with ethanol as described above. The precipitate was dissolved in 1.0ml dH₂O and the concentration of plasmid DNA calculated. At an absorbance of 260 nm, 20 OD units was equivalent to a concentration of 1mg/ml DNA.

3. Isolation of cDNA Inserts

Plasmid DNA, 200μg, was digested to completion with the restriction enzyme Pst I. Digested DNA was electrophoresed on a 1% agarose gel containing 1μg/ml ethidium bromide and run at 250 volts for 40 to 60 minutes (section 2). DNA bands corresponding to linearized pBR322 and inserted fragment DNA were visualized with long wave ultra violet (u.v.) light. To elute the inserted fragment DNA a thin slot, 3 to 5 mm in width, was cut with a razor blade directly in front of and parallel to the u.v. visible band representing the inserted fragment. The gel was electrophoresed such that the buffer touched the gel at either
end but did not submerge it. The slot which had been cut out of the gel was filled with fresh electrophoresis buffer. Progress of the DNA into the buffer filled slot was followed using the long wave u.v. light. The current was stopped before the DNA traversed to the other side of the slot and the buffer containing DNA was transferred to a sterile plastic tube. The slot was refilled with fresh buffer and this process repeated 4 times to allow for maximal DNA recovery. DNA was ethanol precipitated at -20°C overnight as described in section 2. The resulting DNA pellet was dissolved in dH₂O to a concentration of 1mg/ml.

4. Restriction Enzyme Mapping of cDNA Cloned Inserts

cDNA inserts were analysed by single and double restriction enzyme digests. A 2μg aliquot of purified cDNA insert was incubated with 2 units of each restriction enzyme for 60 to 90 minutes at 37°C or 65°C depending on the enzyme used. Enzymes used were: Alu I, Ava I, Ava II, Bam HI, Bgl I, Dde I, Eco RI, Hae III, Hha I, Hin FI, Mbo II, Nci I, Pvu II, Sal I, Sau IIIA, Sau 96I, Sma I, Sst I, Taq I, Tha I, Xba I, Xho I. Enzymes were from Bethesda Research Laboratories Inc. (BRL), Gaithersburg, MD. Enzyme digests were carried out under reaction conditions recomended by BRL.

The digested DNA was analysed by electrophoresis on a 5% non-denaturing polyacrylamide gel (acrylamide:bis acrylamide 29:1(w/w); BioRad, Richmond, Calif.) size 180x160x1.5mm. The
gel was run in 1/2 X TBE (40mM Tris-HCl, 40mM boric acid, 1mM EDTA, pH 8.3) at 100 volts for 2 hours. Gels were stained with ethidium bromide (1μg/ml in dH₂O) and photographed under u.v. light.

To increase the sensitivity the cDNA insert was radioactively labelled. Fragments of the cDNA insert of pRIa.2 generated by restriction enzyme digestion were labelled by the extension of a recessed 3' end. The restriction enzymes used were Hpa II and Hin FI. Each of these enzymes recognizes a specific four base pair sequence and cleaves the DNA to give fragments with a 3' recessed end. The cDNA (2μg) was digested in the appropriate buffer in a final volume of 15μl by incubation at 37°C for 90 minutes. To label the 3' recessed ends, the digested DNA (in 15μl) was heated to 68°C for 5 minutes and allowed to cool to room temperature. To this 1μl dithiothreitol (50mM), 1μl DNA Polymerase I (Klenow Fragment) (2 units/μl, New England Nuclear, Boston, Mass.) and 15μCi of a-³²P-deoxyribonucleotide triphosphate (a-³²P-dNTP) (10mCi/ml in 10mM Tricine, 2500 Ci/m mole; New England Nuclear, Boston, Mass.) were added and the reaction carried out at room temperature for 15 minutes. cDNA digested with Hpa II was labelled with a-³²P-deoxyribocytidine triphosphate (a-³²P-dCTP) and the cDNA digested with Hin FI was labelled with a-³²P-deoxyriboadenosine triphosphate (a-³²P-dATP).

After incubation the reaction mix was phenol extracted (as described in section 2) and chromatographed on a 9ml column of Sephadex-G50 Fine (Pharmacia) in TEN-8. Fractions of 400μl were collected and those containing the radioactive DNA were pooled.
and ethanol precipitated overnight at -20°C (as described in section 2). The precipitated cDNA was washed once in 95% ethanol, dried under vacuum and dissolved in 25μl dH$_2$O. An aliquot of the labelled cDNA was subjected to digestion by a second restriction enzyme. Single and double digested DNA was run on a 5% non-denaturing polyacrylamide gel as described above. The gel was dried down under vacuum at 60°C for two to three hours and then exposed to X-ray film (Kodak X-Omat XAR-2) which was developed after 48 hours.

5. Nick Translation

(Melgar and Goldwaith, 1968; Maniatis et al., 1975; Rigby et al., 1977)

Nick translation of cDNA fragments was carried out under the following conditions:

(a) Deoxyribonuclease (DNase) Activation: DNase I (1mg/ml in dH$_2$O, Boehringer Mannheim, Indianapolis, Ind.) was diluted with DNase Activation Buffer (10mM Tris pH 7.5, 5mM MgCl$_2$, 1mg/ml nuclease free bovine serum albumin (BSA) (Enzo Biochemicals Inc., New York, NY.)) to give a final concentration of 100μg DNase I/ml. The activated DNase I solution was left on ice for between 5 and 120 minutes. Prior to addition to the nick translation reaction, the DNase I was further diluted by 2 serial dilutions of 1/100 and 1/10 into DNase Activation Buffer. This gave an activated DNase I solution of 100ng/ml.

(b) Nick Translation Reaction: To a 1.5ml microfuge tube the following were added on ice: 5μl 10X Nick Translation Buffer
(500mM Tris-HCl pH 7.5, 50mM MgCl₂), 1µl dithiothreitol (50mM), 1µl BSA (2mg/ml), 2µl deoxyriboguanosine triphosphate (dGTP) (500µM), 2µl deoxyribothymidine triphosphate (dTTP) (500µM), 2µl deoxyriboadenosine triphosphate (dATP) (35µM), 2µl deoxyribocytidine triphosphate (dCTP) (35µM), 4µl α⁻³²P-dATP (10mCi/ml, 2000 Ci/mmole, New England Nuclear, Boston, Mass.) 4µl α⁻³²P-dCTP (10mCi/ml, 2000 Ci/mmole, New England Nuclear, Boston, Mass.), 1µl CaCl₂ (10mM), 2µl cDNA fragment (250 µg/ml) 19.5µl dH₂O, 2.5µl Activated DNase I Solution (100ng/ml), and 1µl DNA Polymerase I (8 units/µl, New England Nuclear, Boston, Mass.). The reaction mix was gently mixed, centrifuged in a microfuge and incubated at 15°C for 90 minutes.

After incubation 3µl carrier tRNA (10mg/ml, Sigma type X, St. Louis, MO.) was added. The mix was heated to 68°C for 5 minutes and chromatographed on a Sephadex G-50 Fine column (see section 4). Fractions containing radioactive DNA were pooled, ethanol precipitated overnight at -20°C as before and the precipitate dissolved in 100µl dH₂O. The average incorporation was 60 X 10⁶ cpm/µg DNA.

6. Colony Hybridisation

(Grunstein and Hogness, 1975)

E.coli strain RR1 cells containing a cDNA library prepared from total rat spleen poly(A)-RNA (see section 1) were plated out on LB medium plus 1.5% agar containing tetracycline (10µg/µl) to give approximately 300 to 400 colonies/plate and grown overnight at 37°C. Circular 82mm diameter nitrocellulose
filters (BA 85 0.45μm, Schleicher and Schuell Inc., Keene, NH.) were autoclaved, numbered with soft lead pencil and then placed carefully, number side down, on top of the colonies. Filters were gently pushed onto plates and the number from the filter traced onto the bottom of the plate. After 5 minutes the filters were gently peeled off and placed colony side up onto freshly made LB medium plus 1.5% agar containing chloramphenicol (100μg/μl). Filters were incubated on the plates overnight at 37°C to amplify the plasmid DNA contained in the colonies. The original plates from which the colonies had been removed were incubated to regrow the colonies that had been pulled off.

(a) Immobilisation of DNA on nitrocellulose filters

The filters were removed from the chloramphenicol plates and placed colonies up on Whatman 3MM paper which had been saturated with 0.5M NaOH, 1.5M NaCl. Filters were left for 20 minutes and then transferred to filter paper and air dried for 5 minutes. This step was then repeated. Filters were transferred to Whatman 3MM paper saturated with 1.0M Tris-HCl pH 7.5 for 20 minutes, air dried for 5 minutes and then placed on Whatman 3MM paper saturated with 0.5 M Tris-HCl (pH 7.5), 1.5 M NaCl. Filters were air dried for 30 minutes and baked at 68°C for 3 hours.
(b) Hybridisation

Solutions

1 X SSC (0.15M NaCl, 0.015M Na Citrate, pH 7.0)

1 X Denhardt Solution (0.04% Ficoll, 0.04% polyvinylpyrrolidone, 0.04% BSA all from Sigma, St. Louis, MO. (Denhardt, 1966)

Pre Hybridisation Solution (6 X SSC, 1 X Denhardt Solution)

Hybridisation Solution (6 X SSC, 1 X Denhardt Solution, 0.5% sodium dodecylsulphate (SDS), 1mM EDTA, 0.5mg/ml carrier E.coli DNA (Sigma type VIII, St. Louis, MO.)

Method. Filters were placed in 500ml prehybridisation solution at 68°C for a minimum of 2 hours in a pyrex dish (20X20cm).

Just before use, the nick-translated probe DNA (section 5) was denatured. The following were incubated at 37°C for 30 minutes: 100μl radioactive probe DNA (0.2X10⁶cpm/μl), 15μl carrier pBR322 DNA (800μg/mg) and 3μl NaOH (4M) to give a final volume of 120μl. The denaturation mixture was neutralized by addition of 1.5M NaH₂PO₄ to a final concentration of 0.15M. The denatured probe (approximately 10⁶cpm/filter) was added to preheated hybridisation solution (2ml/filter).

Hybridisation solution containing the radioactive DNA was placed in a petri dish. Filters were then carefully placed in the solution (10 filters/plate maximum) such that there were no air bubbles in between the filters. An exposed piece of X-ray film cut to the dimensions of a petri dish was placed on top of the filters to ensure the filters were kept immersed in the hybridisation solution and out of contact with air. The petri
dish was then placed in an air tight container and incubated at 68°C for a minimum of 6 hours.

(c) Washing

After hybridisation the filters were rinsed twice in 2 X SSC to remove the majority of non-hybridised probe DNA. Filters were then washed three times for 2 hours in 1 X SSC, 0.5% SDS at 68°C (500ml per wash), rinsed in 2 X SSC and air dried on filter paper (approximately 60 minutes). Filters were exposed to X-ray film (Kodak X-Omat XAR-2) for 2 days. Colonies which were identified as positives were isolated from the original regrown plates and the plasmid DNA they contained analysed for the presence of a cDNA insert.

7. Maxam and Gilbert Sequencing

(Maxam and Gilbert, 1980, as modified by Dr. C. Astell, Dept. of Biochemistry, University of BC)

(a) End Labelling Of Fragments With 3' Recessed Ends

cDNA was labelled at the 3' end exactly as described in section 4. In addition to the restriction enzymes Hpa II and Hinf I, the enzymes Eco RI and Sau 3A were also used to digest cDNA fragments to be used for sequencing. The Eco RI and Sau 3A fragments were labelled with α-32P-dATP. To the reaction labelling the Eco RI fragments, 25μM unlabelled dATP was added and incubation was carried out for an additional minute. After the labelling reaction, the DNA was isopropanol precipitated at
-20°C for a minimum of 30 minutes as described in section 2.

(b) Gel Electrophoresis And Isolation Of Labelled Fragments

Precipitated DNA was washed in 95% ethanol, dried under vacuum for 10 minutes, dissolved in 5μl dye mix (1.5% ficoll, 1X TBE, 0.05% xylene cyanol, 0.05% bromophenol blue) and loaded onto a 5% nondenaturing polyacrylamide gel (see section 4). The gel was run at 200 V for 60 to 90 minutes depending on the size of fragments and the separation required. The gel was then exposed to X-ray film (Kodak X Omat XAR-2) for 10 minutes and the film developed to visualize the locations of the DNA bands on the gel. Labelled fragments were cut out of the gel and placed in dialysis tubing with 400μl elution buffer (1/2 X electrophoresis buffer). The DNA was electroeluted from the gel by placing the dialysis bags perpendicular to the current and applying a potential difference of 230 V for approximately 40 minutes. The elution buffer containing radioactive DNA was transferred to siliconized, sterile 1.5ml microfuge tubes and the tubing rinsed with a further 100μl of elution buffer. The eluate was centrifuged in a microfuge for 5 minutes to pellet any polyacrylamide and the supernatant was transferred to a new tube. Labelled fragments were then ethanol precipitated at -20°C overnight as described in section 2.

(c) Double Labelled Fragments

Cleavage of the DNA with restriction enzymes which had more than one recognition site within the cloned cDNA fragment,
resulted in fragments which were labelled at both ends. In order to obtain a single labelled fragment necessary for DNA sequencing, double labelled fragments were either digested with an enzyme (as described in section 4) which cut within the labelled fragment or the fragment was strand separated (see below). Either method resulted in two single labelled fragments which were electrophoresed on a 5% non-denaturing polyacrylamide gel and isolated as described above.

(d) Strand Separation

To strand separate double-labelled fragments, precipitated DNA was dissolved in 40μl of Strand Separation Solution (30% v/v dimethylsulphoxide (DMSO), 1mM EDTA, 0.05% w/v bromophenol blue) and heated to 90°C for 2 minutes. The DNA was quick-cooled in ice water and loaded immediately onto a prerun 5% non-denaturing polyacrylamide gel. The gel was run, with cooling, at 400 volts for 60 to 90 minutes depending on the size of the fragments being separated. It is important to note that when strand separated DNA is electrophoresed on polyacrylamide gels the double stranded DNA migrates faster than the single stranded species and the two single stranded DNAs usually migrate at different speeds thus allowing isolation of the two different strands as single labelled fragments.

The DNA was electroeluted and precipitated as previously described in section 7(b).

(e) End Labelling Of Fragments With 3' Extended Ends
To label fragments with 3' extended ends, such as those resulting from Pst I cleavage, α-^32^P-cordycepin triphosphate and terminal deoxynucleotidyl transferase (TdT) were used. The cDNA fragment, 15μl (600ng/μl), was incubated at 37°C for 30 minutes with 10μl 5X TdT buffer (700mM potassium cacodylate, 150mM Tris-HCl pH 7.0), 5μl CoCl₂ (10mM), 2μl dithiothreitol (2mM), 14μl α-^32^P-cordycepin triphosphate (7mCi/ml in 10mM Tricine, 5700 Ci/m mole, New England Nuclear) and 2μl TdT (20 units/μl, New England Nuclear, Boston, Mass.).

After end labelling the DNA was phenol extracted and precipitated with ethanol as described in section 2. The DNA was then digested with Hpa II, which cut only once within the cDNA fragment, and the 2 single labelled fragments were isolated from a 5% non-denaturing polyacrylamide gel as previously described.

(f) Maxam and Gilbert Modification Reactions

SOLUTIONS:

Cacodylate Buffer (0.05M cacodylate, 0.1mM EDTA pH 8.0)

G-Stop Mix (3M sodium acetate pH 6.0, 2.5M 2-mercaptoethanol, 500μg/ml tRNA (Sigma type X, St. Louis, MO.))

Pyrimidine-Stop Mix (0.3M sodium acetate, 0.1mM EDTA, 83μg/ml tRNA)

A-Stop Mix (0.3M sodium acetate, 0.1mM EDTA, 0.5mM ATP, 83μg/ml tRNA).

The single labelled DNA fragments were taken up in 32μl dH₂O and aliquoted for the 4 base-specific modification reactions. The modifications were carried out as in Table I.
Table I  Maxam and Gilbert Modification Reactions

<table>
<thead>
<tr>
<th>BASES MODIFIED</th>
<th>C + T</th>
<th>G</th>
<th>T</th>
<th>A + G</th>
</tr>
</thead>
<tbody>
<tr>
<td>REACTION CONDITIONS</td>
<td>20 ul 5M NaCl</td>
<td>300ul Cacodylate</td>
<td>15ul dH2O</td>
<td>10ul dH2O</td>
</tr>
<tr>
<td></td>
<td>1ul E.coli DNA</td>
<td>1ul E.coli DNA</td>
<td>1ul E.coli DNA</td>
<td>1ul E.coli DNA</td>
</tr>
<tr>
<td></td>
<td>5ul P-DNA</td>
<td>5ul P-DNA</td>
<td>10ul P-DNA</td>
<td>10ul P-DNA</td>
</tr>
<tr>
<td>START REACTION AT</td>
<td>0 min.</td>
<td>1 min.</td>
<td>2 min.</td>
<td>3 min.</td>
</tr>
<tr>
<td>BASE MODIFYING SOLUTION</td>
<td>30ul hydrazine</td>
<td>2ul DNS</td>
<td>30ul hydrazine</td>
<td>3ul 10% formic acid</td>
</tr>
<tr>
<td>INCUBATE AT</td>
<td>R.T.</td>
<td>R.T.</td>
<td>R.T.</td>
<td>37'C</td>
</tr>
<tr>
<td>STOP REACTION AT</td>
<td>10 min.</td>
<td>4 min.</td>
<td>7 min.</td>
<td>13 min.</td>
</tr>
<tr>
<td>STOP MIX</td>
<td>300ul Pyr-STOP</td>
<td>50ul G-STOP</td>
<td>300ul Pyr-STOP</td>
<td>300ul A-STOP</td>
</tr>
<tr>
<td></td>
<td>1.0ml EtOH</td>
<td>1.0ml EtOH</td>
<td>1.0ml EtOH</td>
<td>1.0ml EtOH</td>
</tr>
</tbody>
</table>
After the addition of 1.0ml 95% ethanol (kept on dry ice), the tubes were placed immediately into a -70°C dry ice/ethanol bath to precipitate the DNA and to ensure the reactions were stopped. After fifteen minutes, the tubes were centrifuged in a microfuge for 10 minutes and the supernatant removed with a drawn-out Pasteur pipet. The DNA was dissolved in 250μl 0.3M sodium acetate (pH 6.3) and precipitated with 1.0ml 95% ethanol at -70°C. The DNA was then dissolved in 10μl dH_{2}O and precipitated with 1.0ml 95% ethanol as above, washed once with 1.0ml 95% ethanol and dried in a vacuum dessicator.

(g) Piperidine Cleavage Reaction

DNA pellets were dissolved in 100μl of freshly diluted piperidine 1/10 (v/v) in dH_{2}O. Caps on the tubes were lined with teflon tape to provide a tight seal and prevent liquid from escaping or entering the tube. The DNA was heated to 90°C for 30 minutes, freeze dried under vacuum (approximately 2 to 3 hours), resuspended in 20μl dH_{2}O and freeze dried (approximately 1 hour), resuspended in a further 20μl dH_{2}O and freeze dried again to ensure that the piperidine was removed.

(h) DNA Sequencing Gels

Polyacrylamide gels were prepared as described in Table II. The gel size was 360 X 200 X 0.35 mm and all gels were run in 1/2 X TBE (see section 4).
Table II  DNA Sequencing Gels

<table>
<thead>
<tr>
<th></th>
<th>20X</th>
<th>8%</th>
<th>6%</th>
</tr>
</thead>
<tbody>
<tr>
<td>UREA</td>
<td>25g</td>
<td>25g</td>
<td>25g</td>
</tr>
<tr>
<td>10x TBE</td>
<td>2.5ml</td>
<td>2.5ml</td>
<td>2.5ml</td>
</tr>
<tr>
<td>40% ACRYL-AMIDE</td>
<td>9.5g ACRYL-AMIDE</td>
<td>10ml</td>
<td>7.5ml</td>
</tr>
<tr>
<td>(19:1)</td>
<td>0.5g BIS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH2O</td>
<td>27.2ml</td>
<td>20ml</td>
<td>21.7ml</td>
</tr>
</tbody>
</table>

Warm to 42°C to dissolve UREA Degas

<table>
<thead>
<tr>
<th></th>
<th>10% AMPS</th>
<th>10% AMPS</th>
<th>10% AMPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEMED</td>
<td>15ul</td>
<td>15ul</td>
<td>15ul</td>
</tr>
</tbody>
</table>

TOTAL VOL         50ml    50ml    50ml
(i) Gel Electrophoresis

After the final freeze drying, the radioactivity of each tube was determined and the DNA dissolved in a volume of dye mix (80% formamide (deionized), 0.1% xylene cyanol, 0.1% bromophenol blue, 10mM NaOH, 1mM EDTA), such that there were twice as many counts/μl in the C + T and A + G tubes as there were in the T and A tubes. The optimal volume of dye mix loaded into a gel slot was 4μl, therefore the DNA in the tube containing the least radioactivity was dissolved in 4μl of dye mix per gel and the DNA in the other tubes was dissolved accordingly.

DNA was heated to 90°C for 3 minutes, quick-cooled in ice-water, immediately loaded onto prerun gels (see above) and electrophoresed at 37 watts. The 20% gels were electrophoresed until the bromophenol blue had migrated 20cm from the loading point (approximately 75 minutes), the top plate of the gel was then removed and the gel was covered in Saran Wrap. 8% gels were electrophoresed for 90 and 180 minutes. The gels were then transferred to Whatman 3MM filter paper, covered with Saran Wrap and dried at 80°C under vacuum for 20 minutes. The dried gel was then exposed to X-ray film (Kodak X-OMAT RP) for one to three days depending on the number of counts loaded onto the gel. In some cases intensifying screens were used (Dupont Cronex Xtra Life, Wilmington, DE.).

(j) DNA Sequence Analysis

A computer program written by Dr. A. Delaney (1982) was used to store and analyse DNA sequence data.
RESULTS AND DISCUSSION

I DNA SEQUENCE ANALYSIS OF THE cDNA INSERT OF pRIa.2

(a) Sequencing strategy

In Figure 2 the strategy used to sequence the cDNA insert in pRIa.2 is shown. The sequencing was done by the methods of Maxam and Gilbert (1980), as described in Materials and Methods. The cDNA insert was digested with restriction enzymes which gave a staggered cut at a 4 base pair recognition sequence to give 3' recessed ends. The ends were then labelled with the appropriate \( \alpha^{-32}P\text{-dNTP} \) using DNA Polymerase I (Klenow fragment) and the fragments were then recut or strand separated. The 3' extended ends produced by Pst I cleavage were labelled with \( \alpha^{-32}P\text{-cordycepin triphosphate} \) using terminal transferase. Both strands of the cDNA insert were sequenced using overlapping restriction enzyme fragments.
The restriction map shows the restriction enzyme recognition sites used for sequencing the cDNA insert of pRIa2. The numbers indicate the 5' nucleotide of the recognition sequence. The numbering starts following the poly(G) tract at the 5' end of the insert. The horizontal arrows beneath the cDNA insert indicate the direction and extent of sequencing.
(b) DNA Sequence of the cDNA Insert of pRIa.2

The nucleotide sequence of the cDNA insert of pRIa.2 (see Figure 3) was 779 base pairs in length. At one end of the insert there was a tract of 18 guanine (G) nucleotides and at the other end there was a tract of 17 cytosine (C) nucleotides. Immediately preceding the poly(C) tract was a tract of poly(A), 63 nucleotides in length. This poly(A) tract corresponded to the poly(A) tail of the mRNA and defined the 3' end of the cDNA insert. There was a putative polyadenylation signal, AATAAA, (Proudfoot and Brownlee, 1976) 15 nucleotides 5' to the start of the poly(A) tail. Following the 5' poly(G) tract there was an open-reading frame of 388 nucleotides ending in the stop codon TGA. The open reading frame represented the coding region for the carboxy terminal 129 amino acids of the Ia a chain. The 290 nucleotides following the stop codon corresponded to a 3' untranslated region of the mRNA.
Figure 3 The Nucleotide Sequence of the cDNA Insert of pRIa.2

(G) TCAGCCCAACACCCTCATCTGCTTTTGAGACAACATCTTTTCTGCTGTGAATCAATA
116 TCACATGGTTGAGAAACAGCAAGCCAGTCACAGAAGGCGTTATGAGACCAGCTTT
176 CCAAACCTGACCATTCCTTCCACAAGATGGCTTACCTCACCTTCATCCCTCCAACGACG
236 ACATTATGACTGCAAGGTGGAGCACTGGGGCCTGGACGAGCCGGTTTAAAACACTGGG
296 AACCTGAGGTTTCCAGCCCCCCTGTCAGAGCTGACAGAGACTGTGGTCTGTGCCCTGGGGTT
356 TGCTGTCGCGCCTCTGGGCACTCGTGTTGGGACACCATCTCTCATCTCAGGGCTCGGAT
416 CAGATGCCCCCTCCAGACACCCAGGGCCCTTTGAGTCAACACCCTGGGAAAGAAGGTGCG
476 TGCCCTCCTACAGGGCAAAGTGTAGTGTAGGCTGGAGCTGGCAACGTGTGGTTTCTGCC
536 CAATTCATCGGTCTTCTCTTCTTCTCCTTGGTGTCTCCCATCTTGCTCTTCCCTGGCCCC
596 CAGGCTGTCCACCTCATGGCTCTCACGCCCTTGGAATTCCCTGACCTGAGTTTCACTT
656 TTTGGCATCTTCAGAATGTAATAGTATGACTCCGGAGACCTGATGATGCTCCACCA
681 AACCAATCCCTCTCATAGTGGG(C) (A)
63 17

The poly(G) and poly(C) tracts at either end of the insert are represented by a bracketed G/C with a subscript representing the number of nucleotides. The poly(A) tail is similarly represented. The TGA stop codon at 389 is underlined as is the putative polyadenylation signal, AATTTA, at nucleotide 661. The numbers above each line represent the nucleotide number, starting at the first nucleotide following the 5' poly(G) tract and ending at the last nucleotide preceding the poly(A) tail at number 681.
To gain a better understanding of the evolution of the Ia molecules encoded by the MHC, it is useful to study the equivalent genes and their gene products in species which are closely related such as mice and rats, which diverged only 8 to 10 million years ago, and far apart such as rodents and humans, which diverged approximately 70 million years ago (Young, 1950).

In Figure 4 the nucleotide sequence of the cDNA insert of pRIa.2 is compared with the equivalent sequences of cDNA coding for mouse H-2 I-A and the human HLA-DC1 a chains. There is 91% DNA sequence identity between rat and mouse, and 85% DNA sequence identity between rat and human when coding sequences only are compared. When the 3' untranslated regions are compared, however, the rat and mouse sequences show 82% identity, whereas the rat and human sequences are only 27% identical.

This difference in sequence identity between coding and noncoding regions is consistent with the proposal that noncoding regions are much more free from restraints and diverge from one another in evolution much more rapidly than coding regions (Crick, 1981). Coding regions are maintained because of the evolutionary restraints or pressures which act on the gene products coded by the DNA. Non-coding regions are able to mutate and thus evolve at a much higher rate because the selection
pressures of evolution are able to act only indirectly on these regions. The function of non-coding or untranslated regions of mRNA is still unknown. It is thought that the poly(A) tail may act to stabilize mRNA (Wilson et al., 1978; Zeeri et al., 1981). If this is so, perhaps the 3' untranslated region of mRNA also functions to stabilize the message. It is interesting to note that the human HLA-DC1 mRNA has only a 123 nucleotide 3' untranslated region (not including the poly(A) tail), while the rat has a 290 nucleotide 3' untranslated region. The mouse 3' untranslated region is at least 184 nucleotides in length but the full extent of the 3' end has not been reported.
Comparison of the Nucleotide Sequences of the cDNA encoding the \( \alpha \) Chains of Rat Ia-A, Mouse H-2 I-A and Human HLA-DC1.

Asterisks show positions of sequence identity. The stop codons are boxed. Data from Benoist et al., 1983 (H-2 I-A) and Auffray et al., 1982 (HLA-DC1).
II PREDICTED AMINO ACID SEQUENCE OF THE CARBOXY TERMINAL END OF RAT Ia-A α CHAIN

(a) Translation of the cDNA Sequence

The predicted amino acid sequence of the cDNA insert of pRIa.2 is shown in Figure 5. The translation starts at the second nucleotide after the G tail and corresponds to the carboxy terminal 129 amino acids. By analogy with the known domain organization of the HLA-DR α chain (Korman et al., 1982b), the predicted amino acid sequence of the cDNA insert of pRIa.2 can be divided into several different structural regions. The first 91 amino acids correspond to the majority of the first extracellular domain (a2). The next 13 amino acids define the connecting peptide, a region which joins the a2 domain to the transmembrane region of the protein. The transmembrane region itself, consists of 23 mainly hydrophobic amino acids. The remaining 15 amino acids are thought to be located intracellularly and define the cytoplasmic region.
Figure 5  Translation of the Coding Sequence of the cDNA Insert of pRIa.2.

<table>
<thead>
<tr>
<th>16</th>
<th>31</th>
<th>46</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAG CCC AAC ACC CTG ATC TGG TTT GTA GAC AAC ATC TTT CCT CCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLN PRO ASN THR LEU ILE CYS PHE VAL ASP ASN ILE PHE PRO PRO</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

○

<table>
<thead>
<tr>
<th>61</th>
<th>76</th>
<th>91</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTG ATC AAT ATC ACA TGG TGG AGA AAC AGC AAG CCA GTC ACA GAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAL ILE ASN ILE THR TRP LEU ARG ASN SER LYS PRO VAL THR GLU</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>106</th>
<th>121</th>
<th>136</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCC GTT TAT GAG ACC AGC TTC CCT ACC AAC GCT GAC CAT TCC TTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLY VAL TYR GLU THR SER PHE LEU SER ASN PRO ASP HIS SER PHE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>151</th>
<th>166</th>
<th>181</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC AAG ATG GCT TAC CTC ACC TTC GAC CAT GAC ATT HIS LYS MET ALA TYR LEU THR PHE ILE PRO SER ASN ASP ASP ILE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

X

<table>
<thead>
<tr>
<th>196</th>
<th>211</th>
<th>226</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAT GAC TGG AAG GTG GAG CAC TGG GGC GTG GAC GAG CCG GTT CTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYR ASP CYS LYS VAL GLU HIS TRP GLY LEU ASP GLU PRO VAL LEU</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>241</th>
<th>256</th>
<th>271</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA CAC TGG GAA CCT GAG GTT CCA GCC CCC ATG TCA GAG CTG ACA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYS HIS TRP GLU PRO GLU VAL PRO ALA PRO MET SER GLU LEU THR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>286</th>
<th>301</th>
<th>316</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAG ACT GTG GTC TGT GCC CTG GGG TTG TCT GTG GGC CTC GTG GGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLU THR VAL VAL CYS ALA LEU GLY LEU SER VAL GLY LEU VAL GLY</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>331</th>
<th>346</th>
<th>361</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATC GTG GTG GGC ACC ATC TTC ATC ATT CAA GCC CTG CGA TCA GAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILE VAL VAL GLY THR ILE PHE ILE ILE GLN GLY LEU ARG SER ASP</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>376</th>
<th>391</th>
<th>406</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCC CCC TCC AGA CAC CCA GGG CCC CTT TGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLY PRO SER ARG HIS PRO GLY PRO LEU</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The predicted amino acid sequence of the cDNA insert of pRIa.2 is shown. Translation starts at the second nucleotide after the G tail. The numbers correspond to the nucleotide numbering of Figure 4. Cysteine residues which may be involved in disulphide bonding are marked (X). The putative glycosylation site is marked (○).
(b) The Amino Acid Sequence Predicted From The cDNA Insert Of pRIa.2 Compared With The Corresponding Mouse H-2 I-A And Human HLA-DC1 \(a\) Chain Sequences

A comparison of the amino acid sequences predicted from cDNA corresponding to the rat Ia-A, mouse H-2 I-A and human HLA-DC1 \(a\) chains is shown in Figure 6. The data from Figure 6 is summarised in Table III. The overall sequence identity for the regions of the \(a\) chains available for comparison is 90.7% when rat Ia-A and mouse H-2 I-A \(a\) chains are compared. The overall sequence identity drops to 81.4%, however, when rat Ia-A (or mouse H-2 I-A) are compared to the human HLA-DC1 \(a\) chain. Not surprisingly, the rat and mouse proteins are more similar to one another then either is to the human protein. It is interesting that both rodent \(a\) chains differ to the same extent when compared to the human \(a\) chain, even though both rodent proteins have diverged from one another.
The amino acid sequences for the equivalent regions of the rat Ia-A\(\alpha\), mouse H-2 I-A\(\alpha\) (Benoist et al., 1983) and human HLA-DC1\(\alpha\) (Auffray et al., 1982) are compared. The Ig-like conserved residues are marked with a cross (+), and the MHC conserved residues by an open circle (O). The numbers below the lines refer to the numbering of the human HLA-DC1\(\alpha\) chain (Auffray et al., 1982). The amino acids defining the borders of the protein domains are also numbered.
Figure 6 Comparison of the Amino Acid Sequences of the α Chains of Rat Ia-A, Mouse H-2 I-A and Human HLA-DC1.

<table>
<thead>
<tr>
<th></th>
<th>H-2 I-A</th>
<th>RAT Ia-A</th>
<th>HLA-DC1</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-2 I-A</td>
<td>GLN PRO ASN THR LEU ILE CYS PHE VAL ASP ASN ILE PHE PRO PRO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAT Ia-A</td>
<td>GLN PRO ASN THR LEU ILE CYS PHE VAL ASP ASN ILE PHE PRO PRO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-DC1</td>
<td>GLN PRO ASN THR LEU ILE CYS LEU VAL ASP ASN ILE PHE PRO PRO</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>104</td>
</tr>
<tr>
<td>H-2 I-A</td>
<td>VAL ILE ASN ILE THR TRP LEU ARG ASN SER LYS SER VAL THR ASP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAT Ia-A</td>
<td>VAL ILE ASN ILE THR TRP LEU ARG ASN SER LYS PRO VAL THR GLU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-DC1</td>
<td>VAL VAL ASN ILE THR TRP LEU SER ASN GLY HIS SER VAL THR GLU</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>133</td>
</tr>
<tr>
<td>H-2 I-A</td>
<td>GLY VAL TYR GLU THR SER PHE PHE VAL ASN ARG ASP TYR SER PHE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAT Ia-A</td>
<td>GLY VAL TYR GLU THR SER PHE LEU SER ASN PRO ASP HIS SER PHE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-DC1</td>
<td>GLY VAL SER GLU THR SER PHE LEU SER LYS SER ASP HIS SER PHE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>148</td>
</tr>
<tr>
<td>H-2 I-A</td>
<td>HIS LYS LEU SER TYR LEU THR PHE ILE PRO SER ASP ASP ASP ILE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAT Ia-A</td>
<td>HIS LYS MET ALA TYR LEU THR PHE ILE PRO ASN ASP ASP ILE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-DC1</td>
<td>PHE LYS ILE SER TYR LEU THR PHE LEU PRO SER ALA ASP GLU ILE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>163</td>
</tr>
<tr>
<td>H-2 I-A</td>
<td>TYR ASP CYS LYS VAL GLU HIS TRP GLU LEU GLU GLU PRO VAL LEU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAT Ia-A</td>
<td>TYR ASP CYS LYS VAL GLU HIS TRP GLU LEU ASP GLU PRO VAL LEU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-DC1</td>
<td>TYR ASP CYS LYS VAL GLU HIS TRP GLU LEU ASP GLU PRO VAL LEU</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>178</td>
</tr>
<tr>
<td>H-2 I-A</td>
<td>LYS HIS TRP GLU PRO GLU ILE PRO ALA PRO MET SER GLU LEU THR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAT Ia-A</td>
<td>LYS HIS TRP GLU PRO GLU VAL PRO ALA PRO MET SER GLU LEU THR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-DC1</td>
<td>LYS HIS TRP GLU PRO GLU ILE PRO THR PRO MET SER GLU LEU THR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>193</td>
</tr>
<tr>
<td>H-2 I-A</td>
<td>GLU THR VAL VAL CYS ALA LEU GLY LEU SER VAL GLY LEU VAL GLY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAT Ia-A</td>
<td>GLU THR VAL VAL CYS ALA LEU GLY LEU SER VAL GLY LEU VAL GLY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-DC1</td>
<td>GLU THR VAL VAL CYS ALA LEU GLY LEU SER VAL GLY LEU VAL GLY</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>194</td>
</tr>
<tr>
<td>H-2 I-A</td>
<td>ILE VAL VAL GLY THR ILE PHE ILE ILE GLN GLY LEU ARG SER GLY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAT Ia-A</td>
<td>ILE VAL VAL GLY THR ILE PHE ILE ILE GLN GLY LEU ARG SER ASP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-DC1</td>
<td>ILE VAL VAL GLY THR VAL LEU ILE ARG GLY LEU ARG SER VAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>217</td>
</tr>
<tr>
<td>H-2 I-A</td>
<td>GLY THR SER ARG HIS PRO GLY PRO LEU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAT Ia-A</td>
<td>GLY PRO SER ARG HIS PRO GLY PRO LEU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-DC1</td>
<td>GLY ALA SER ARG HIS GLN GLY PRO LEU</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table III  Comparison of the sequence identity between the \( \alpha \) chains of Rat Ia-\( \alpha \), Mouse H-2 I-A and Human HLA-DC1.

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>RAT Ia-( \alpha ) /MOUSE H-2 I-A</th>
<th>RAT Ia-( \alpha ) /HUMAN HLA-DC1</th>
<th>MOUSE H-2/HUMAN I-A HLA-DC1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% SEQUENCE IDENTITY</td>
<td>NUMBER OF AMINO ACIDS COMPARED</td>
<td>% SEQUENCE IDENTITY</td>
</tr>
<tr>
<td>2 (104-181)</td>
<td>88.5%</td>
<td>69/78</td>
<td>79.5%</td>
</tr>
<tr>
<td>CP (182-194)</td>
<td>92.3%</td>
<td>12/13</td>
<td>84.6%</td>
</tr>
<tr>
<td>TM (195-217)</td>
<td>100%</td>
<td>23/23</td>
<td>91.3%</td>
</tr>
<tr>
<td>CYT (218-232)</td>
<td>86.7%</td>
<td>13/15</td>
<td>73.3%</td>
</tr>
<tr>
<td>OVERALL</td>
<td>90.7%</td>
<td>117/129</td>
<td>81.4%</td>
</tr>
</tbody>
</table>

References are the same as for Figure 6.
Although the overall homology, in terms of sequence identity, between these proteins is high there are quite different levels of homology for the different domains. In Table III the sequence is divided into domains or regions of structural difference and the percent sequence identity of the rat a chain is compared to the mouse and human a chains. Although the homology between rat and mouse is higher than that between rat and human, those domains showing the least conservation between mouse and rat are the same domains with the least homology between human and rat.

The different functional regions are discussed in detail below.

The Cytoplasmic Tail The least conserved region is the cytoplasmic tail. The homology between mouse and rat is 86.7% for this region and between human and rat is 73.3%. The cytoplasmic tail is 15 amino acids in length and is approximately 50% hydrophilic in nature. This region is thought to be important in anchoring the protein in the membrane. There are 3 basic residues within the cytoplasmic tail. It is thought that the presence of such residues may act to stabilize the protein in the membrane. This may occur by interactions between negatively charged phosphate groups of the membrane phospholipids and the positively charged basic amino acids (Bretscher, 1975). Clusters of basic amino acids have been found in the cytoplasmic tails of glycophorin (Tomita et al., 1978), HLA-A2 and HLA-B7 (Robb et al., 1978) and membrane IgM (Rogers et al., 1980). The presence of proline residues also appears to be a characteristic of cytoplasmic regions of some proteins. In
glycophorin there are 5 proline residues in the 24 amino acid cytoplasmic tail (Tomita et al., 1978). In the rat Ia-A a chain cytoplasmic tail there are 3 prolines. Proline residues are, however, not characteristic of human HLA heavy chain cytoplasmic regions (Robb et al., 1978). There is some evidence that the Ia antigens may be phosphorylated at a serine residue within the cytoplasmic tail (Korman et al., 1982a). There are two serine residues in the rat Ia-A a chain and both are adjacent to arginine residues. The cAMP-dependent protein kinase which may catalyse the phosphorylation reaction requires that the arginine be on the amino-terminal side of the serine (Kemp et al., 1977). Therefore, only the serine at amino acid 222 would be available for phosphorylation.

It is thought that another function of the cytoplasmic tail is to provide a means of communicating information from outside the cell to the cytoplasm. The exact nature of the mechanism by which this communication occurs is not known but it may be that the cytoplasmic tail interacts with molecules in the cytoplasm which then travel to the nucleus. If this was so then it might be expected that the cytoplasmic regions would be highly conserved within and between species. It is surprising then, to find that the cytoplasmic region is the least conserved region of the molecules over the regions compared. The hydrophilic nature of this region is conserved and the presence of the adjacent arginine-serine residues is notable but there are marked differences. Perhaps this indicates that these Ia a chain molecules do not communicate by interaction with cytoplasmic molecules after all. Alternatively, the residues which are
important for these interactions are conserved and the only constraint on the other residues is that they retain their hydrophilic nature. It is interesting that comparison of the cytoplasmic regions of the murine, H-2 I-A alpha and H-2 I-E alpha chains shows only 47% sequence identity and the human, HLA-DC1 alpha and HLA-DR alpha chains show only 45% sequence identity. One interpretation of this is that if the I-A and I-E (DC1 and DR) alpha chains interact with molecules within the cytoplasm that they may interact with different molecules. It appears therefore, that the cytoplasmic tails of Ia products of closely related loci within a species are less homologous than Ia products of equivalent loci in different species.

The Transmembrane Region  
The transmembrane region of proteins which span the membrane is thought to be characterized by the presence of a mainly hydrophobic stretch of amino acids long enough to and capable of taking on an alpha-helical structure which will stretch across the membrane. The 23 amino acid transmembrane regions of the three alpha chains compared appear to have these characteristics. If this were the only function of a transmembrane region, however, it may be expected that there would be a low degree of sequence conservation as long as the general hydrophobic nature of the region was maintained. Therefore it is surprising to find that this region exhibits not the lowest but the highest degree of sequence identity of all the regions compared. Between rat and mouse there is 100% sequence identity and between rat and human a 91.3% sequence identity is observed. This unexpected degree of conservation seems to indicate that there must be a functional constraint
maintaining the sequence and selecting against any nucleotide changes which would result in an amino acid change in this region.

When the transmembrane regions of two different murine Class I antigens of the same haplotype, H-2K\textsuperscript{b} and H-2D\textsuperscript{b}, were compared they showed a protein sequence identity of 75\% whereas overall the protein sequences showed a higher identity of 83\% (Reyes et al., 1982). A comparison of Class I antigens from equivalent loci of different species, the murine H-2K\textsuperscript{b} and the human HLA-B7, showed an even more striking disparity between the protein identities of the extracellular domains, 70\%, and the transmembrane regions, 30\% (Coligan et al., 1981). The high degree of sequence identity between the transmembrane regions of the \(\alpha\) chains described seems unusual and may suggest that these proteins are involved in a quaternary interaction with other membrane proteins (Korman et al., 1982b). It has been proposed (Benoist et al., 1983) that the \(\alpha\) and \(\beta\) chains interact within the membrane. This would explain the unexpected sequence constraint on this region and is further supported by data from the murine H-2 I-E \(\alpha\) and human HLA-DR \(\alpha\) chain sequences which also have similar transmembrane region sequences (Benoist et al., 1983, Auffray et al., 1982). In Table IV the transmembrane regions of Ia \(\alpha\) chains of different species are compared.

It is also of interest to note the presence of a cysteine residue within the transmembrane region of all three \(\alpha\) chains. There is evidence to suggest that this may be a site for post-translational modification (Kaufman and Strominger, 1979).
Table IV Comparison of the Amino Acid Sequences of the Transmembrane Regions of the Ia α Chains of Different Species

<table>
<thead>
<tr>
<th></th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-DR</td>
<td>ASN-VAL-VAL-CYS-ALA-LEU-GLY-LEU-THR-VAL-GLY-LEU</td>
</tr>
<tr>
<td>HLA-DC1</td>
<td>THR-VAL-VAL-CYS-ALA-LEU-GLY-LEU-SER-VAL-GLY-LEU</td>
</tr>
<tr>
<td>H-2 I-E</td>
<td>ASN-VAL-VAL-CYS-ALA-LEU-GLY-LEU-PHE-VAL-GLY-LEU</td>
</tr>
<tr>
<td>H-2 I-A</td>
<td>THR-VAL-VAL-CYS-ALA-LEU-GLY-LEU-SER-VAL-GLY-LEU</td>
</tr>
<tr>
<td>Rat Ia-A</td>
<td>THR-VAL-VAL-CYS-ALA-LEU-GLY-LEU-SER-VAL-GLY-LEU</td>
</tr>
<tr>
<td>HLA-DR</td>
<td>VAL-GLY-ILE-ILE-ILE-ILE-GLY-THR-ILE-PHE-ILE-ILE</td>
</tr>
<tr>
<td>HLA-DC1</td>
<td>VAL-GLY-ILE-VAL-VAL-GLY-THR-VAL-LEU-ILE-ILE</td>
</tr>
<tr>
<td>H-2 I-E</td>
<td>VAL-GLY-ILE-VAL-VAL-GLY-ILE-ILE-LEU-ILE-MET</td>
</tr>
<tr>
<td>H-2 I-A</td>
<td>VAL-GLY-ILE-VAL-VAL-GLY-THR-ILE-PHE-ILE-ILE</td>
</tr>
<tr>
<td>RAT Ia-A</td>
<td>VAL-GLY-ILE-VAL-VAL-GLY-THR-ILE-PHE-ILE-ILE</td>
</tr>
</tbody>
</table>

Data from Lee et al.,1982a (HLA-DR), Auffray et al.,1982 (HLA-DC1) and Benoist et al.,1983 (H-2 I-A and I-E).
The Connecting Peptide  The 13 amino acids between the \( a2 \) domain and the transmembrane region define the connecting peptide. The connecting peptide is less conserved than the transmembrane region and only slightly more conserved than the cytoplasmic region. There is 92.3% homology between rat and mouse and 84.6% homology between rat and human sequences.

The high degree of homology between the transmembrane regions of the rat, mouse and human \( a \) chains is especially notable because of the much less conserved nature of the cytoplasmic tail and the connecting peptide. If the gene organization of the Ia-A \( a \) chain is the same as that of the human HLA-DR \( a \) chain gene then all three regions would be encoded by the same exon. It appears that only the transmembrane region is strictly conserved and that the two regions surrounding it are more free to diverge. One explanation is that the HLA-DR \( a \) chain gene organization is not the same as the gene organization of the HLA-DC1, mouse H-2 I-A and rat Ia-A \( a \) chain genes at least for the exon(s) coding for the connecting peptide, transmembrane and cytoplasmic regions.

The extracellular domain (\( a2 \)) The 91 amino acids of the \( a2 \) domain are 88.5% homologous between rat Ia-A \( a \) and mouse H-2 I-A \( a \) and 79.5% homologous between rat Ia-A \( a \) and human HLA-DC1 \( a \).

The Ia molecules are glycoproteins and there is a potential glycosylation site at the asparagine residue at position 121, approximately 70 amino acids from the transmembrane portion of the protein. The sequence at this point asparagine-isoleucine-threonine (Asn-Ile-Thr) conforms to the consensus sequence for carbohydrate attachment, Asn-X-(Thr or Ser) (Spiro, 1974; Wagh...
and Bahl, 1981). There are two cysteine residues which may form a disulphide loop within the a2 domain.

**Homology to Immunoglobulin (Ig) Domains**

Several immune system molecules such as β₂microglobulin, and the Class I antigen heavy chains have been shown to contain amino acid sequences which are homologous to the amino acid sequences of the constant region domains of the immunoglobulin (Ig) molecules. Analysis of the amino acid sequence predicted from the cDNA insert of pRIa.2 shows that it too exhibits homology to the sequences of the Ig constant region domains.

The first studies on Ig molecules showed that they were four chain structures (two light chains and two heavy chains) (Porter, 1962). Characterization of immunoglobulin molecules by amino acid sequence analysis indicated that the immunoglobulin molecules could be divided into a series of globular domains (Edelman et al., 1969, 1970). This was further supported by X-ray crystallographic experiments (Poljak et al., 1973). There are several features which characterize immunoglobulin domains, notably a disulphide loop of approximately 55-65 amino acids and several typically conserved amino acids which are thought to be involved in maintenance of the three-dimensional structure of these domains. When the amino acid sequences of immunoglobulin domains are aligned, these Ig-conserved residues are found at the same positions.

In Figure 6 the majority of the a2 domain sequences (amino acids 104-181) of rat Ia-A, mouse H-2 I-A, and human HLA-DC1 a chains are compared. There are 13 amino acids within each a2
domain which are characteristic of Ig-conserved residues. The two cysteine residues thought to form a disulphide loop are included in these Ig-conserved residues. There are a few Ig-conserved residues in the middle of the putative loop structure but the majority are clustered around the two cysteine residues. As well as the 13 Ig-conserved residues there are three residues which are found in the Ig-like domains of all MHC encoded molecules which have been compared. These MHC-conserved residues may play an important role in some MHC specific function of this domain.

Despite differences in amino acid sequence between the constant region domains, X-ray crystallographic experiments indicated that the different constant regions all have a similar structure (Amzel and Poljak, 1979). The similarity in amino acid sequence of these domains argued that they represented a similar pattern of tertiary folding (Beale and Feinstein, 1976). A typical constant region domain structure defined from X-ray crystallographic techniques is shown in Figure 7.
Diagram showing the basic immunoglobulin fold of an immunoglobulin constant region (Poljak et al., 1973).
The constant region domain consists of 2 faces, the x face (fx) and the y face (fy). There are stretches of β-pleated sheet within each face and most of the amino acid homologies between constant region domains are within the β-pleated sheets. Connecting the β-pleated sheets are non-β-pleated segments or bends; within the bends homology between constant region domains is lowest. The bends and β-pleated segments are indicated in Figure 8. There is a characteristic pattern of alternating hydrophobic and hydrophilic amino acids around the cysteine residues of constant region domains. The side chains of these hydrophobic amino acids fill the internal spaces between the x and y faces of the constant region domains. The Ig-conserved residues which are clustered around the cysteines of the α2 domains of the Ia α chains conform to the pattern of alternating hydrophobic residues. For example in most constant region domains the sequence of Tyr (or Phe)-X-Cys-X-Val-X-His is highly conserved in the second stretch of β pleated sheet of the y face (fy2) (Beale and Feinstein, 1976). This sequence is found in Ia molecules as shown in Figure 8.
Figure 8  
Comparison of the Amino Acid Sequences of  
Ig-Like Domains

<table>
<thead>
<tr>
<th>Rat Ia-A</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>H-2 I-E</td>
<td>V A P E V - T V L S R S P V N L G E P N I L C F I D K F S P</td>
<td></td>
</tr>
<tr>
<td>HLA-DC1</td>
<td>V P P E V - T V L T N S P V E L R E D N I L C F I D K F S P</td>
<td></td>
</tr>
<tr>
<td>HLA-DR</td>
<td>E V P E V - T V F S K S P V T L G Q P N V L I C F I D K F T P</td>
<td></td>
</tr>
<tr>
<td>CH3</td>
<td>(\text{fx1})</td>
<td>b</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rat Ia-A</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-DC1</td>
<td>P V V N V T W L S N - G H S V T E G V S E T S F L S K S - D H</td>
<td></td>
</tr>
<tr>
<td>HLA-DR</td>
<td>P V V N V T W L R N - G K P V T T G V S E T V F L P R E - D H</td>
<td></td>
</tr>
<tr>
<td>HLA-B7</td>
<td>A E I T L T W Q R - D G - E D Q T Q D T E L V E T R P A G D R</td>
<td></td>
</tr>
<tr>
<td>(\beta_2^M)</td>
<td>S D I E V D L L K - D G - E R - I E K V E H S D L S F S K D W</td>
<td></td>
</tr>
<tr>
<td>CH3</td>
<td>S D I A V E W E S N O G - E - P E N Y K T T P P V L D S D G</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rat Ia-A</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>H-2 I-A</td>
<td>S F H K M A Y L T F I P S N - - D D - I Y D C K V E H W G L D</td>
<td></td>
</tr>
<tr>
<td>HLA-DC1</td>
<td>S F F K I S Y L T F L P S A - - D E - I Y D C K V E H W G L D</td>
<td></td>
</tr>
<tr>
<td>HLA-DR</td>
<td>L F R K F H Y L P F L P S T - - D E - V Y D C R V E H W G L D</td>
<td></td>
</tr>
<tr>
<td>HLA-B7</td>
<td>T F E K W A A V Y V - P S - - E E O R Y T C H V O H E G L P</td>
<td></td>
</tr>
<tr>
<td>(\beta_2^M)</td>
<td>S F Y L L Y T E F T P T - - - E K D E Y A C R V N H V T L S</td>
<td></td>
</tr>
<tr>
<td>CH3</td>
<td>S F F L Y S K L T V D K S R W O E G N V F S C S V M H E A L H</td>
<td></td>
</tr>
</tbody>
</table>

The \(\beta\) pleated sheet segments of the \(x\) face (\(\text{fx}\)) and the \(y\) face (\(\text{fy}\)) are numbered as are the bends (b). Positions where spaces have been left to give maximal alignment are denoted by a (-). References as for Table IV plus data from Orr et al., 1979 (HLA-B7), Peterson et al., 1972 (\(\beta_2\)microglobulin) and Edelman et al., 1969 (Ig constant region domain, CH3)
\( \beta_2 \) microglobulin, the Class I antigen light chain, is 100 amino acids in length and contains a 57 residue disulphide loop (Peterson et al., 1972; Cunningham et al., 1973). When the amino acid sequence of \( \beta_2 \) microglobulin was compared to Ig sequences a high degree of homology was found. The highest degree of sequence identity was observed when \( \beta_2 \) microglobulin was compared to the CH3 domain. Because of its homology to immunoglobulin constant region domains \( \beta_2 \) microglobulin was termed a "free" immunoglobulin domain (Peterson et al., 1972).

Sequence analysis of the Class I antigen heavy chains, showed that two of the three extracellular domains contained disulphide loop structures of the same size as the characteristic Ig domain disulphide loop. Comparison of the heavy chain sequences of human Class I antigens (HLA-A and B) with immunoglobulin constant region domains and \( \beta_2 \) microglobulin showed that only the Class I antigen extracellular disulphide loop adjacent to the membrane exhibited homology to immunoglobulins and \( \beta_2 \) microglobulin (Tragardh et al., 1978; Orr et al., 1979b; Wiman et al., 1979). This so-called immunoglobulin-like (Ig-like) domain of the Class I antigens (\( \alpha_3 \)) was shown to be as homologous to \( \beta_2 \) microglobulin and constant immunoglobulin domains as \( \beta_2 \) microglobulin and immunoglobulin constant region domains were to one another. From this data the proposal was made that the immunoglobulins, \( \beta_2 \) microglobulin, and the Class I antigen heavy chains had a related evolution and that the Class I antigens may have arisen by a gene duplication involving the same ancestral gene as that which gave rise to immunoglobulin chains (Wiman et al., 1979).
The rat Thy-I glycoprotein is a major cell surface molecule which is found on thymocytes, neuronal and some other cell types. The amino acid sequence of Thy-I was determined by Campbell et al. (1981). Thy-I was shown to be 111 amino acids in length and to contain 4 cysteine residues which defined a single disulphide loop. This disulphide loop was shown to be homologous to immunoglobulin domains but unlike the Class I antigen heavy chains and $\beta_2$-microglobulin, Thy-I was more homologous to the variable region domains (Cohen et al., 1981; Williams and Gagnon, 1982).

Comparison of a human Ia antigen $\beta$ chain sequence (Larhammar et al., 1981, 1982a) with the other Ig-like domains previously identified showed that the $\beta_3$ domain (the domain adjacent to the membrane) was also an Ig-like domain. The human HLA-DR $\alpha$ chain was also shown to contain an extracellular Ig-like domain adjacent to the membrane, the $\alpha_2$ domain (Larhammar et al., 1982b). The $\alpha_2$ domains of both mouse H-2 I-A and human HLA-DC1 $\alpha$ chains were also shown to be homologous to $\beta_2$-microglobulin, Class I $\alpha_3$ domains and immunoglobulin constant region domains.

In Table V the Ig-like domains of several different immune system molecules are compared.
### Table V  Percent Sequence Identity Between Ig-Like Domains of Different Immune System Molecules

<table>
<thead>
<tr>
<th></th>
<th>RAT Ia-Aa</th>
<th>MOUSE H-2 I-Aa</th>
<th>MOUSE H-2 I-Ea</th>
<th>HUMAN HLA-DC1a</th>
<th>HUMAN HLA-DRα</th>
<th>HUMAN HLA-A, B, C</th>
<th>β₂MICROGLOBULIN</th>
<th>CH3</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOUSE H-2 I-Aa</td>
<td>89</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOUSE H-2 I-Ea</td>
<td>67</td>
<td>60</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUMAN HLA-DC1a</td>
<td>80</td>
<td>67</td>
<td>70</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUMAN HLA-DRα</td>
<td>68</td>
<td>65</td>
<td>80</td>
<td>65</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUMAN HLA-A, B, C</td>
<td>29</td>
<td>27</td>
<td>32</td>
<td>31</td>
<td>31</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>β₂MICROGLOBULIN</td>
<td>32</td>
<td>29</td>
<td>29</td>
<td>32</td>
<td>29</td>
<td>25</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CH3</td>
<td>24</td>
<td>26</td>
<td>22</td>
<td>24</td>
<td>23</td>
<td>28</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

References as for Figure 8.
The finding that the Ia and Class I antigens, as well as \( \beta_2 \) microglobulin and Thy-1 are homologous to immunoglobulin domains led to the proposal of an immunoglobulin superfamily. Implicit in this proposal, is the hypothesis that all the molecules within the family originated from a single ancestral gene which would be equivalent to a single exon at the DNA level or a single domain at the protein level. The immunoglobulin constant domains to which the Class I antigen heavy chains, the Ia antigens and \( \beta_2 \) microglobulin had been likened, always occurred in pairwise combination. It was suggested, therefore, that the Class I heavy chain and \( \beta_2 \) microglobulin and the Ia \( \alpha \) and \( \beta \) chains associate in a manner similar to that of the immunoglobulin constant region domains (Steinmetz et al., 1981b; Larhammar et al., 1982b). The proposed structures for the Class I and Ia antigens are diagrammed in Figure 9 along with the structure of a membrane immunoglobulin molecule.
The immunoglobulin-related molecules Thy-1, MHC Class I and Ia antigens are shown diagrammatically along with a membrane IgM molecule. The drawings are approximately to scale. (Jensenius, J.C. and Williams, A.F., 1982).
III SEARCH FOR THE AMINO TERMINAL END
OF THE RAT Ia-A \(\alpha\) CHAIN

In order to isolate a cDNA clone corresponding to the 5' end of mRNA encoding a rat Ia-A \(\alpha\) chain, two cDNA libraries made from total spleen poly(A)-RNA (approximately 4000 independent transformants) were screened (see Materials and Methods). The cDNA insert of plasmid pRIa.2 was nick translated and used as a probe to screen for bacterial transformants containing rat Ia-A \(\alpha\) chain sequences (see Materials and Methods).

Fifteen clones which hybridized to the pRIa.2 cDNA insert were isolated. Restriction enzyme analysis indicated that there were only two independently arising clones. The clone with the larger insert (named pRIa.3) was amplified and plasmid DNA was prepared. The cDNA insert of the plasmid was approximately 800 base pairs in length and was bounded at either end by a PstI site. The restriction map of the cDNA insert of pRIa.3 indicated that it might be the 5' end of a cDNA coding for the rat Ia-A \(\alpha\) chain. The EcoR1 site, situated at position 570 of the cDNA insert in pRIa.2, was missing and since the inserts were approximately the same length this seemed to imply that the cDNA insert of pRIa.3 would extend at least another 200 nucleotides 5' to that of the cDNA insert of pRIa.2. This extra 200 nucleotides of coding region could code for another 70 amino acids and would, therefore, extend well into the \(\alpha1\) domain of the rat Ia-A \(\alpha\) chain. The cDNA insert was isolated as described (see Materials and Methods) and the cDNA was partially sequenced.
according to the methods of Maxam and Gilbert (see Materials and Methods).

The partial sequence of the cDNA insert of pRIa.3

Approximately one half of the sequence of the cDNA insert of pRIa.3 was determined and was compared to the sequence of the pRIa.2 cDNA insert to identify any regions of homology. The comparison is shown in Figure 10.

Although the partial restriction map indicated that the pRIa.3 cDNA insert might contain DNA sequences corresponding to the 5' end of the protein, comparison of the DNA sequences of the cDNA inserts of pRIa.2 and pRIa.3 showed that this was not the case. Instead of the expected 5' sequences, the cDNA insert of pRIa.3 appeared to contain a different 3' untranslated region, 600 nucleotides in length. The 5' end of the cDNA insert of pRIa.3 was homologous to the open reading frame of the cDNA insert of pRIa.2. This homology initiated at nucleotide 252 of the pRIa.2 insert sequence and extended into the 3' untranslated region. The first 198 nucleotides of the pRIa.3 insert were nearly identical to the sequence of the pRIa.2 insert. At nucleotide 451 of the pRIa.2 insert sequence, however, the sequences of the two inserts became totally different. The pRIa.3 insert sequence extends approximately another 550 nucleotides from this point in the presumed 3' direction and is entirely different in sequence from the remaining 130 nucleotides of the 3' untranslated region of pRIa.2.
Figure 10 Alignment of the Partial Nucleotide Sequence of the cDNA insert of pRIa.3 with the Sequence of the cDNA insert of pRIa.2

pRIa.2 (G)^TCAGCCCAACACCTCATCTGCTTTGTAGACAACATCTTTCTCTCTGTGTACATA
pRIa.2 TCACATGGTTGAGAAACAGCAAGCCGAGTCAGAAGGCGTTTATGAGACCAAGCTTTCTTT
pRIa.2 CCAACCCCTGACCATTCTCTCCAAACAGATGGCTTACCTACCTTACCCCTTCACAAGACG
pRIa.2 ACATTTATGACTGCAAGGTGGAGCATTGGGCTGGACGAGCCGGTTTCATCTAAAACACTGGG
pRIa.2 AACCTGAGGTCCACCCACCATGTCAAGCAGATCGACAGACTCGTGTGTCGCCCTTGGGG
pRIa.3 GGGGGGGGGGTTCC--CCCCATGTCAGAAGCTGACAAAGACTGTGGTCTGTCGCTTGGGGT
pRIa.2 TGTCTGGGCGTCGTCGGGATGTCATCGGCTCGCCACCATCTTCATCTCAAGGCTGCGAT
pRIa.3 TGTCTGGGCGTCGTCGGGATGTCATCGGCTCGCCACCATCTTCATCTCAAGGCTGCGAT
pRIa.2 CAGATGGCACCCTCCAGACACCAGGCCCCCCTTGTAGTCACACCCTTGGGAAAGAAGGTGCCG
pRIa.3 CAGGGGGGGCCCTTGTAGTCACACCCTTGGGAAAGAAGGTGCCG
pRIa.2 TGGGCTTCAAGGCAAGATGTAGTGAGGGGTTGACCTGGACAGGCTGAGTTTCTGCCCC
pRIa.3 TGGGCTTCAAGGCAAGATGTAGTGAGGGGTTTACAACCTGACGAGGTGTTTTCTGCCCC
pRIa.2 CAATTCATCGGTGGTCTTCTTCTTCTTCTGTCGGTGCTTTCATCTTCCTCTTGGGCCC
pRIa.3 GTTCTCTGAGTAGTTGAGAAACAGAAGTTATTTTTAGTAGGAGAGTTTTGGGGCTTCTGCCC
pRIa.2 CAGGCTGTCCACCTCATGGCTCTACGCCCTTGGAATTCTCCCCTGACCTGAGTTTCTGCTATT
pRIa.3 --AGCACTTAAGACTATCATACTCTCTA----CTCTTCTCTCCTACGAGACTTT--TTCTCTATCCTATCACGACGAG
pRIa.2 TTGGGACTCTTCCAGAATCTACTATAGATTTCCGAGACCTGTGTGGATGCTCCACCTCT
pRIa.3 --T-T-CTCAGAATCTACTATAGATTTCCGAGACCTGTGTGGATGCTCCACCTCT
pRIa.2 AACCAATAAACCTCTCATAAGTGG(\textit{A})_{n}(\textit{C})_{n}
pRIa.3 CAACATTTGAGAAAGGACCGAGACGAGGTGAGTTTGAGTCTCGTCCGG

The arrow indicates a possible splice site (see text).
The presumed 3' end of the cDNA insert of pRIa.3 has not been sequenced, but due to the method of synthesis of the cDNA there should be a poly(A) tail at the 3' end. If this is true and the cDNA insert of pRIa.3 represents a functional mRNA and not a cloning artifact, then the 3' untranslated region of this mRNA is approximately 600 nucleotides in length.

The unusual properties of the cDNA insert of pRIa.3 are puzzling. If it is assumed that this cDNA represents a functional mRNA which can be translated into the rat Ia-A protein, then it seems surprising that there should be two mRNA species differing only in their 3' untranslated regions which code for the same protein. There are many possible explanations for this finding but it is not possible to determine which if any are correct from the data available.

One explanation is that the mRNA from which the cDNA insert of pRIA.3 was derived was not completely processed i.e. it still contained nucleotide sequences which corresponds to intron DNA. This would have been spliced out of the mRNA from which the cDNA insert of pRIa.2 was derived. The last two nucleotides of shared sequence identity between the two cDNA inserts are GT. It is possible that this corresponds to the 5' end of the presumed intron sequence. Although this correlates with the consensus sequence for intron borders (GT at the 5' end and AG at the 3' end; Lewin, 1980) to explain why this dinucleotide is also found in the sequence of the cDNA insert of pRIa.2 it must be proposed that the dinucleotide GT also occurs at the 5' end of the following exon. This explanation is acceptable but it is unclear why there would be an intron within the region of the DNA
corresponding to the 3' untranslated region of the mRNA. There are, however, several findings which would indicate that the 3' ends of some MHC encoded proteins have unusual properties.

The gene organization of the human HLA-DR α gene is known (Lee et al., 1982b, Korman et al., 1982). The 3' untranslated region is encoded by two exons. Exon three contains the connecting peptide, the transmembrane region, the cytoplasmic tail and extends 10 nucleotides downstream of the TGA stop codon. Exon four contains the remainder of the 3' untranslated region. Analysis of other genes has indicated that the stop codon and the 3'untranslated regions are usually found within the same exon (Korman et al., 1982). For the human HLA-DR α chain gene this is not the case.

One isolated cDNA for the human HLA-DR α chain (Lee et al., 1982a) had 2 polyadenylation signals within the 3'untranslated sequence; one was 28 nucleotides upstream of the poly(A) tail and the other was 130 nucleotides upstream of the poly(A) tail. It appeared that both polyadenylation signals may be used. If the first polyadenylation signal were recognized and signalled the end of transcription the mRNA would have an approximately 300 nucleotide untranslated region. If the second polyadenylation signal were used, as was the case for the cDNA which was isolated, then the mRNA would have an approximately 400 nucleotide untranslated region. If the presence of two polyadenylation signals is a characteristic of Ia α chains then perhaps this indicates that there is some requirement for mRNA's with different lengths of 3' untranslated sequences. The mRNA species with different 3'untranslated regions may be synthesised
in response to different conditions or requirements in the cell.

The most likely explanation of the cDNA insert of pRIa.3 is that it represents an incompletely processed mRNA. To further investigate the insert of pRIa.3 experiments such as isolation of genomic DNA corresponding to rat Ia-A α genes and characterisation of this DNA by sequence analysis are needed.
IV CONCLUSIONS

The sequence of a cDNA corresponding to a rat Ia-A α chain and the amino acid sequence predicted from this cDNA were presented. The rat Ia-A α chain sequence shows significant homology to Ig constant region domains and to Ig-like domains of other immune system molecules. The Ia-A α chain appears to belong to the proposed immunoglobulin superfamily and has probably evolved from the same ancestral gene from which the other molecules of this superfamily are derived.

Comparison of the rat Ia-A α chain sequence to sequences from the equivalent loci of mouse and human shows that a high degree of sequence identity exists across the species barrier. The level of homology observed between the Ia α chain sequences from different loci within any single species is significantly lower. This indicates that the two Ia loci, Ia-A and Ia-E, duplicated prior to the speciation of rats, mice and humans.
ACKNOWLEDGEMENT

I would like to gratefully acknowledge the support and advice I received from Rob McMaster, Sarah Eccles and Heather Wallis.
REFERENCES


Baxevanis, C.N., Wernet, D., Nagy, Z.A., Maurer, P.H. and Klein, J. (1980) Genetic Control of T-Cell Proliferative Responses to Poly (Glu⁰Ala⁶⁰) and Poly (Glu₅¹Lys₃¹Tyr₁₅): Subregion-Specific Inhibition of the Responses with Monoclonal Ia Antibodies. Immunogenetics 11, 617-628.


Merryman, C.F. and Maurer, P.H. (1976) Genetic Control of Immune Response Against Random Copolymers of Glutamic Acid and Alanine (GA) and Tyrosine (GT) in Inbred Mice. J. Imm. 116, 739-742.


