[Frontiers in Bioscience 6, d904-913, August 1, 2001]

GENETICS OF HUMAN COMPLEMENT COMPONENT C4 AND EVOLUTION THE CENTRAL MHC

O. Patricia Martinez,[1] Natalie Longman-Jacobsen,[1] Richard Davies,[1] Erwin K. Chung,[2] Yan Yang,[2] Silvana Gaudieri,[1] Roger L. Dawkins,[1] and C. Yung Yu[2]

[1] Centre for Molecular Immunology and Instrumentation, University of Western Australia, PO Box 5100, Canning Vale WA 6155, Australia, [2] Children's Research Institute, Columbus, Ohio, and Department of Pediatrics, Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University, 700 Children's Drive, Columbus Ohio 43205

TABLE OF CONTENTS

1. ABSTRACT

The two classes of human complement component C4 proteins C4A and C4B manifest differential chemical reactivities and binding affinities towards target surfaces and complement receptor CR1. There are multiple, polymorphic allotypes of C4A and C4B proteins. A complex multiplication pattern of C4A and C4B genes with variations in gene size, gene dosage and flanking genes exists in the population. This is probably driven by the selection pressure to respond to a great variety of parasites efficiently and effectively, which the bony fish achieved through the multiplication and diversification of the related complement C3 proteins. Complement C4, C3 and C5 belong to the α_2 macroglobulin protein family but acquired specific features that include an anaphylatoxin domain, a netrin (NTR) domain, and stretches of basic residues for proteolytic processings to form multiple chain structures. Complement C3 and C4 are important in the innate immune response as they opsonize parasites for phagocytosis. The emergence of complement C3 predates proteins involved in the adaptive immune response as C3 is present in deuterostome invertebrates such as echinoderms. The human C4 genes are located in the central MHC at chromosome 6p21.3. C3 and C5 are located at chromosome 19 and 9, respectively, with representatives of the other groups of genes paralogous to the MHC at 19p13.1-p13.3, 1q21-25, and 9q33-34. The central MHC also contains genes for complement components C2 and Bf. These genes appear to have similar evolutionary histories to C3/C4/C5 and are used here to illustrate stepwise processes resulting in co-location of diverse domains, chromosomal duplication, local segmental duplication and divergence of sequence and function. This model of evolution is useful in the investigation of innate and acquired immunity and in seeking explanations for diseases associated with MHC ancestral haplotypes.

2. THE GENETIC DIVERSITY AND THE NOMENCLATURE OF HUMAN C4A AND C4B

The human complement component C4 is one of the most complex and polymorphic molecules. The activated product of C4, C4b, is a non-catalytic subunit C3 and C5 convertase in the classical and lectin pathways (reviewed in refs. 1, 2). Downstream of C4, the complement pathway includes the generation of anaphylatoxins C3a and C5a, and the assembly of the membrane attack complexes. Through the intra-chain thioester bond formed between Cys-991 and Gln-994, C4b forms a covalent ester or amide bond with targets or any nearby surfaces. Therefore, it localizes the activation of complement in a spatially specific manner. Binding of C4b opsonizes antigens or immune complexes for phagocytosis. It also helps solubilization, prevents immunoprecipitation, and enhances the clearance of immune complexes through the interaction of C4b with the erythrocyte complement receptor CR1.

A wealth of knowledge on human C4 genetics has accumulated recently. Experimental evidence is pointing to a dynamic (or drifting) 1-2-3 loci concept of the human C4 genes, rather than a rigid two-locus model that would imply the single locus or three-locus haplotypes to be infrequent or aberrations (3, 4). While the two-locus configurations are quite prevalent with a frequency of 69%, the single and three locus haplotypes of C4 genes account for most of the remaining 31% of the Caucasian population.

Structurally, the C4 gene is defined by five specific nucleotide changes in exon 26. These changes code for the isotypic residues PCPVLD (C4A) and LSPVIH (C4B) at positions 1101-1106 (5, 6). Functionally, the activated form of C4A is characterized by lower hemolytic activity, a longer half-life against hydrolysis, a high binding affinity to amino-group containing substrates and to

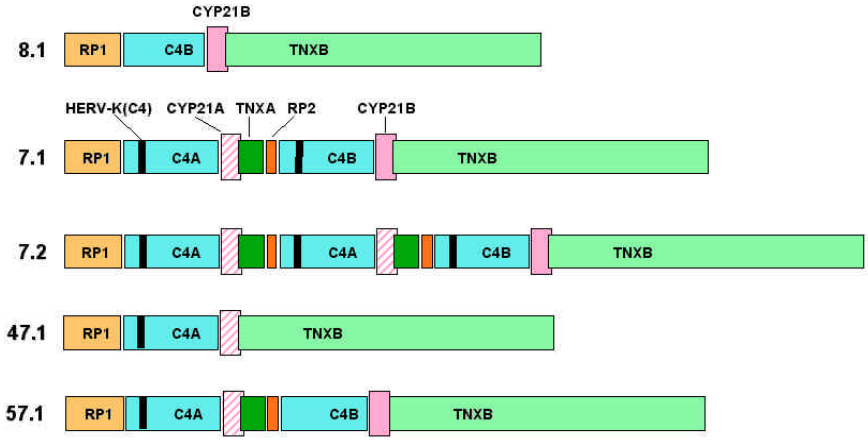

Figure 1. Characterized MHC ancestral haplotypes. C4 and adjacent genes form modules that exhibit substantial polymorphism due to variation in the number of modules, the presence of the HERV-K(C4) sequence and the segments that accompany C4 duplications. AH 47.1 is a carrier of congenital adrenal hyperplasia. (3, 4, 20, 73-77).

complement receptor CR1, and a greater ability to prevent immunoprecipitation. In comparison with C4A, C4B has higher hemolytic activity. The activated C4B is very reactive because one of its isotypic residues, His-1106, is catalytic to the transacylation reaction of the thioester bond to form covalent ester or amide bonds. However, the half-life of activated C4B is transient and less than 1 second (7, 8). Therefore, C4B enables the activation of the classical complement pathway specifically at the target site and at the right time. C4A, however, probably evolved to play a larger role in opsonization and immunoclearance (7 – 12).

Typically, C4 loci may code for C4A or for C4B. The number of C4A and C4B genes is a function of the ancestral haplotype of the major histocompatibility complex (MHC). As shown in Tables 1 and 2, there is substantial polymorphism. Two additional aspects of variation are shown in Figure 1. First, each of the C4 genes may be long or short, depending on the presence of the endogenous retrovirus, HERV-K(C4), in intron 9 of the C4 gene (13, 14). Second, the duplication of a C4 gene is always accompanied by a CYP21 gene and two adjacent gene segments TNXA and RP2, implying that the process is segmental or modular (15 – 17). The steroid 21-hydroxylase gene in the duplicated module may be the pseudogene designated CYP21A, or the functional gene designated CYP21B (3, 18 – 21).

Polymorphism of the C4A and C4B proteins is most conveniently demonstrated by immunofixation of EDTA-plasma using high voltage agarose gel electrophoresis (HVAGE), a technique known as allotyping (Figure 2) (22, 23). The separation of C4A and C4B allotypes is based on differences in electrophoretic migration as C4A allotypes travel faster towards the anode (24, 25). The C4A and C4B proteins may be further identified by a hemolytic overlay as most allotypes of C4B lyse sensitized sheep red blood cells four times faster than those of C4A. Most C4A allotypes are associated with the Rodgers (Rg) blood antigen and C4B with Chido (Ch) blood group antigen; an immunoblot analysis of the C4 protein separated by HVAGE and reacted with Rg1 and Ch1 monoclonal antibodies allows further resolution of the C4A and C4B allotypes. Altogether more than 41 allotypes have been documented by these phenotyping techniques. However, there is accumulating evidence suggesting that the allotypes in each group are heterogeneous. As shown in Table 2, the most common allotype of C4A is A3 and the most common allotype of C4B is B1. Other common allotypes include A2, A4 and A6 for C4A; and B2, B3 and B5 for C4B. It is still not completely clear at this stage how each allotype differs structurally and at the sequence level but some sequences are available (2, 5, 26). The current nomenclature of C4A and C4B has many limitations (27, 28), which include the following:

a. the electrophoretic variants are likely to be heterogeneous;
b. the association of C4A and C4B with Rg or Ch antigens is not indicated;
c. the variation in the functional activities is not included;
d. the C4 gene may be long or short, e.g, C4B1(L) as in the 44.3 *ancestral haplotype* (AH) and C4B1(S) as in AH 8.1;
e. the number of C4 loci in the haplotype is not described;
f. the HLA ancestral haplotypes, complotypes, and RCCX haplotypes are not implicated.

A systematic revision of the nomenclature of C4A and C4B will be necessary, perhaps describing the entire segment. However, it may be necessary to await more structural information of the C4A and C4B alleles, at least in the polymorphic C4d region (Figure 4).

3. EVOLUTION OF THE MHC-COMPLEMENT PROTEINS AND THE CENTRAL MHC

The standard view of the complement system is based on a cascade that can be initiated by three different

Table 1. Number of Copies of Genes in the RCCX Modules and C4 Gene Size Variation in the MHC Ancestral Haplotypes

AH	RP1	C4A	CYP21A	TNXA	RP2	C4B	CYP21B	TNXB	C4 GENE SIZE
7.1	1	1	1	1	1	1	1	1	L, L
7.2	1	2	2	2	2	1	1	1	L, L, L
8.1	1	0	0	0	0	1	1	1	S
18.1	1	1	1	1	1	1	1	1	L, S
18.2	1	1	0	0	0	0	1	1	L
44.3	1	0	0	0	0	1	1	1	L
47.1	1	1	1	0	0	0	0	1	L
52.1	1	2	1	1	1	0	1	1	L, L
57.1	1	1	1	1	1	1	1	1	L, S

There is substantial polymorphism in the region occupied by the C4 genes and other components of the RCCX modules, as illustrated here. Only some ancestral haplotypes (AH) are shown. AHs have a characteristic number of copies of each gene. C4 genes may be classified as being long (L) or short (S) depending on the presence or absence of HERV-K(C4), respectively (20, 73 – 77).

Table 2. A List of the HLA Class I, Class II , Complement Components Bf, C4A and C4B Alleles in Ancestral Haplotypes (AHs) of the Human MHC

AH	HLA-A	HLA-B	Bf	C4A	C4B	HLA-DR	HLA-DQ
7.1	3	7	S	3	1	15	6
7.2	24	7	S	3, 3	1	1	5
8.1	1	8	S	Q0	1	3	2
13.1	30	13	S	3	1	7	2
18.1	25	18	S	4	2	15	6
18.2	30	18	F1	3	Q0	3	2
18.3		18	S	3	1	11	7
35.1		35	S	3	1	11	7
35.2	3	35	F	3, 2	Q0	1	5
35.3	11	35	S	3	Q0	1	5
37.1	1	37	F	3	1	10	
38.1	26	38	S	2	1	4	8
42.1		42	F	12, 91	Q0	3	4
44.1	2	44	S	3, 3	Q0	4	7
44.2	29	44	F	3	1	7	2
44.3	29	44	S	Q0	1	7	2
44.4	33	44	F	3	1	13	6
46.1	2	46	S	4	2	9	9
46.2	2	46	S	4	2	8	6
47.1	3	47	F	1	Q0	7	2
50.1		50	S07	2	1, 12	7	
51.1		51	F	3	Q0	4	3
52.1	24	52	S	3, 2	Q0	15	6
54.1		54	S	3	5	4	4
55.1		55	S	4	5	14	
57.1	1	57	S	6	1	7	9
58.1	33	58	S	3	Q0	3	2
58.2	33	58	F	Q0	1	13	
59.1		59	S	3	5	9	9
60.1		60	S	3	1	4	3
60.2		60	S	3	Q0	8	4
60.3	2	60	S	Q0	2	13	6
61.1	26	61	S	3	1	9	9
62.1	2	62	S	3	3	4	8
62.2		62	S	4	2	4	8
62.3		62	F	3	1	13	6
62.4	11	62	S	3	1	4	
64.1		64	S	3	1	7	2
65.1		65	S	2	1, 2	1	5

AHs contain characteristic alleles at each locus. The number of copies of each C4 gene has been determined by the C4A : C4B gene dosage ratios and C4 *Taq* I RFLP product sizes (3, 15, 75 – 77).

Figure 2. Human C4A and C4B allotypes. Electrophoretic migration of C4A and C4B proteins indicates their highly polymorphic nature. This technique has been used to determine C4A and C4B allotypes. C4A3 and C4B1 (*red* boxes) are the most frequent Caucasian C4 allotypes. *Dotted* lines are used to indicate the relative positions of some allotypes (25, 28, 74). An *arrow* indicates the direction of migration. (After 28, 74.)

pathways. The classical pathway is initiated via interactions of antibodies with the C1 complex that activates C4 and C2 to form the classical pathway C3 convertase (29). Alternatively, the binding of mannose binding lectin to bacterial cell walls may activate the mannan associated serine proteases (MASP) that in turn activate C4 and C2 (30). The microbial products may also activate the alternative pathway as the assembly of activated C3 (through the tick-over mechanism) and Bf forms the alternative pathway C3 convertase. Irrespective of the initial events, complement C5 is activated and that leads to the spontaneous, sequential assembly of the membrane attack complex for the lytic pathway: C5b, C6, C7, C8, and C9 (31). Cascades of this type are normally precisely controlled and require inhibitors to prevent deleterious inflammation and self-injuries (1, 32). These include a family of receptors known as Complement Control Proteins (CCP) that are characterized by the presence of Short Consensus Repeats (SCR) (33). Most SCR-containing CCPs are located on 1q32 within a gene cluster known as the Regulators of Complement Activation (RCA) (34) but SCRs within C2 and Bf are also present within the central MHC (*vide infra*). Each CCP has a specific number and specific sequence of SCRs and binds particular ligands in a species specific manner (33, 35). Therefore, complement components and CCP have co-evolved together to permit a more extensive but finely regulated cascade. SCR structural motifs are also present in *C. elegans* and *Drosophila melanogaster* (36). It is of interest to determine if the SCR-containing proteins in these lower animals are involved in the host-parasite discrimination.

Analysis of the primary structures of C2 and factor B (Bf) revealed a composite structure with a CCP, a von Willebrand type A repeat (vWRA) and a serine protease domain (SP) (37, 38). The 3 SCRs of the CCP region in each gene are very similar, in keeping with a common origin of C2 and Bf, undoubtedly, through a process of segmental duplication (Figure 3). In each case, SCR1 is degenerate but SCR 2 and 3 are similar to the f and h subfamilies of CR1.

In humans, the major ligands of the CCP are the activated or inactivated products of complement C3 and C4. Together with C5, C3 and C4 belong to the α_2 macroglobulin protein (A2M) family (39, 40). A2M is also present in worms and flies. It is a proteinase inhibitor that binds to all classes of proteases and conveys them to a receptor-mediated endocytosis clearance pathway. Thus A2M may potentially contribute to immunity by inactivating the proteinase virulence factors of pathogens (41). Structurally, A2M proteins are characterized by the

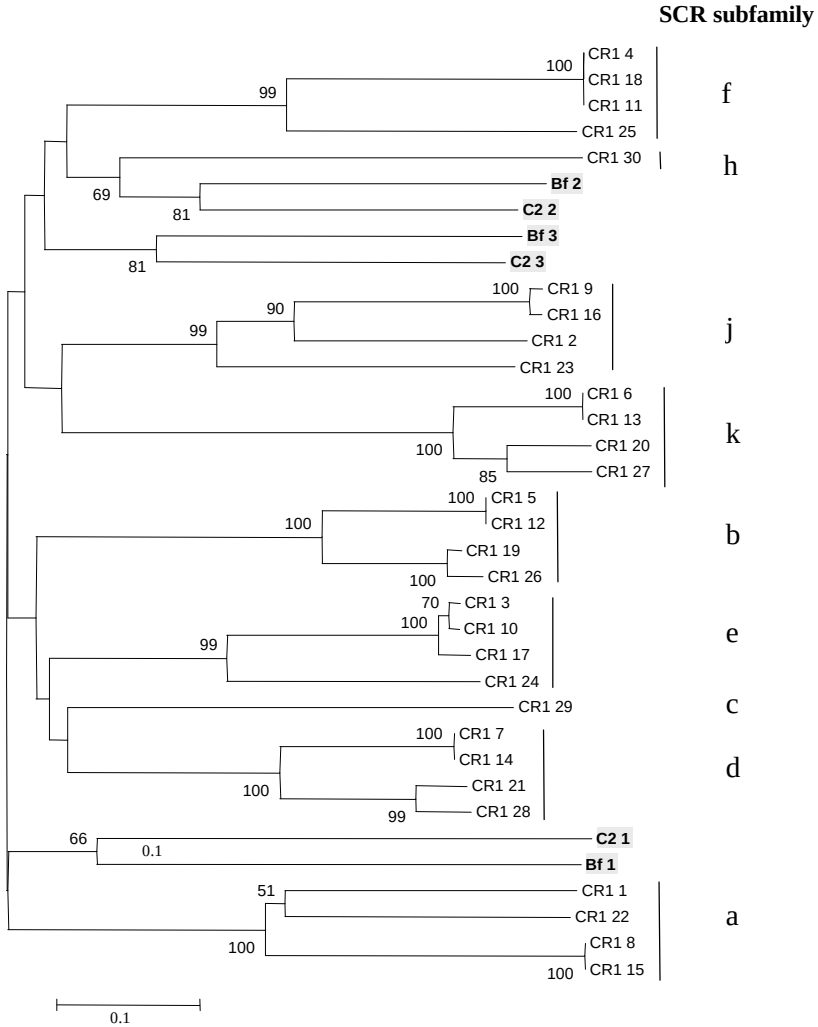

Figure 3. The relationships among the SCRs of human complement component C2, factor B (Bf), and complement receptor type 1 (CR1). The amino acids of each domain were aligned (ClustalW, version 1.8) and consensus distance trees generated using the MEGA package (Molecular Evolutionary Genetics Analysis version 1.03). Evolutionary distances were estimated using p-distance and the phylogenetic tree was constructed by neighbor-joining. The tree under went 1000 bootstrap replications and is not rooted. The accession numbers for the amino acid sequences are: Hosa CR1, P17927; Hosa Factor B, AAA16820; and Hosa C2, P06681.

presence of an intra-chain thioester bond that is able to form covalent amide or ester linkage with targets. In humans there are *seven* members in the A2M protein family (42, 43), which are α_2 macroglobulin, pregnancy zone protein (PZP) (44), KIAA1283 (45), and complement components C4A, C4B, C3 and C5. The human α_2 macroglobulin and pregnancy zone protein (PZP) are closely linked and located on chromosome 12, KIAA1283 and complement C3 are located at chromosome 19p13.1-13.3, C4A and C4B are present in the MHC, and complement C5 is located on chromosome 9q33-34.

The C4/C3/C5 proteins are more closely related as they acquire an anaphylatoxin (ANA) domain of 74 -77 amino acids close to the middle of the protein, and a netrin domain (NTR) about 152 residues in size at the carboxyl terminus of each protein (46). Preceding the ANA domain there are four basic residues which are subjected to proteolysis and therefore give rise to the β and α chains of the proteins. Further distinct structural features added to complement C4 include the basic residues between the α–γ chain junction, and a cluster of three tyrosine-sulfation sites close to the carboxyl terminus of the α-chain (47). Complement C5 is an unusual member of the A2M protein family as it has lost the thioester structure.

The anaphylatoxins C5a and C3a (and C4a) are potent inflammatory agents that promote contraction of smooth muscles, increase vascular permeability and platelet aggregation. They also induce the release of histamine by degranulation of mast cells and stimulate the secretion of

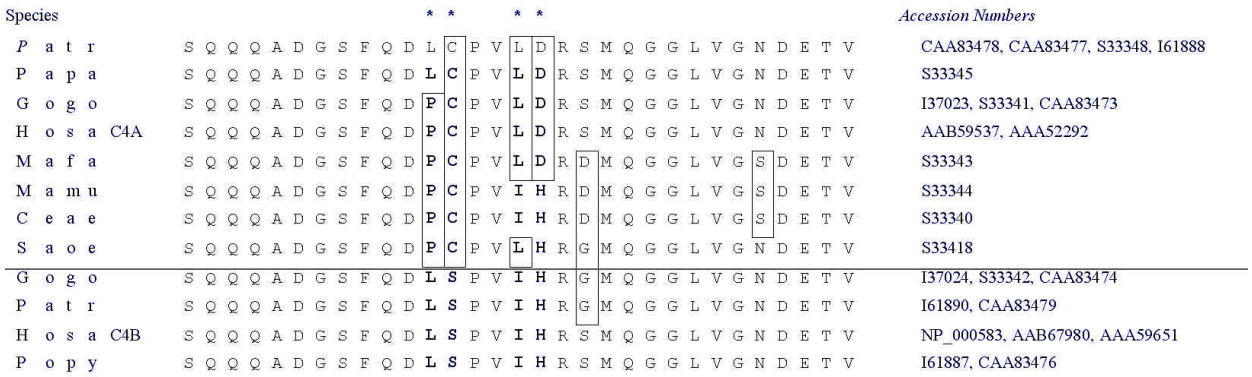

Figure 4. Alignment of the amino acid sequences around the catalytic residues that may modulate the thioester activities of the primate complement component C4A and C4B proteins. Amino acid sequence positions correspond to 1090 – 1120 with respect to human C4. Positions 1101, 1102, 1105 and 1106 are shown by asterisks. Sequence variants at each position are boxed. Species abbreviations are: Patr, *Pan troglodytes* (common chimpanzee); Papa, *Pan paniscus* (pygmy chimpanzee); Gogo, *Gorilla gorilla*; Hosa, *Homo sapiens*; Mafa, *Macaca fascicularis* (crab-eating macaque); Mamu, *Macaca mulatta* (Rhesus macaque); Ceae, *Cercopithecus aethiops* (African green monkey); Saoe, *Saguinus oedipus* (cotton top tamarin); Popy, *Pongo pygmaeus* (orangutan) (5, 61, 62, 78, 79).

interleukin-1 from monocytes. C5a is also a powerful chemotactic agent for phagocytes (48, 49). Besides the anaphylatoxins, ANA motifs are also present in fibulin-1 (50) and fibulin-2 (51), which are extracellular matrix proteins in close association with microfibrils containing fibronectin or fibrillin.

The NTR module at the carboxyl region of complement C5 has been shown by alanine- scanning site-directed mutagenesis to be the distal binding site of the *classical* pathway C5 convertase, C4b2a3b (52). The physiological roles of the NTR modules in complement C4 and C3 are unknown and it will be worthwhile to investigate if they form binding sites for the multi-component proteases in the complement activation pathway such as the C1 complex, the MBL/MASP complex, and the C3 convertase. The NTR modules are also present in type I collagen C-proteinase enhancer protein and tissue inhibitors of metalloproteases (53).

The elucidation of the evolutionary history leading to the emergence of the composite structures with A2MN+ANA+A2MC+NTR in C3, C4 and C5, and CCPs+vWRA+SP as in Bf and C2 should shed light on the mechanism of segmental duplication and the evolution of the innate and the adaptive immune responses (54) and the MHC. Specifically, it will be interesting to determine the selective advantage of having CCP domains within the central MHC.

Although components of the classical pathway and the lytic pathway, and the regulatory proteins of the complement system are vertebrate specific proteins, complement proteins of the lectin pathway and the alternative pathway appeared much earlier in evolution. In *C. elegans* and *Drosophila* there are one and five A2M-like genes, respectively. Four of the *Drosophila* A2M-like proteins, TEP1 – TEP4, have the *t*hio*e*ster *p*rotein signature sequences. The expression levels of these *Tep* genes are markedly up-regulated after an immune challenge. Moreover, the increased expression level of *Tep1* is dependent on the Janus kinase gene *hopscotch* (55), implying that the TEP proteins could be involved in the innate immune response in the insect (55, 56), although the mechanism is yet to be determined. Our sequence analyses suggested that the five A2M-like proteins in *Drosophila* are closely related to α_2 macroglobulin and to KIAA1283.

The presence of complement C3/C4/C5-like protein can be traced back to echinoderms and ascidians (57 – 59). These complement proteins probably play important roles in the discrimination of the host and parasites through the opsonization of parasites for phagocytosis. In different species of bony fishes, 2-5 genes for complement C3 are present. In a phenomenon similar to that of complement C4A and C4B proteins in old world primates (Figure 4) (60 – 62) and in cattle and sheep (39), the teleost C3 proteins diversified to react to substrates with differential affinities (54, 58). In particular, one of the two C3 genes in the Japanese medaka fish has the catalytic residue corresponding to His-1106 in human C4B changed to Ala (54). The common carp has the five C3 genes, C3-1 to C3-5, which have Ser, His, His, Gln and His residues at the described catalytic residue, respectively. Interestingly, the C3-1 isoform with Ser at the putative catalytic position has the highest hemolytic activity in the artificial assay system (63). This phenomenon underscores the importance of C3 and C4 in innate immune response, and the selection pressure on C3 or C4 to react with a variety of parasite cell surfaces efficiently or effectively. The presence of multiple C3 genes in the common carp and multiple C4 genes in humans, old world primates, sheep and cattle is an excellent example of divergent evolution resulting in functional diversities to defend against different parasites. In this context, it is not surprising that complement control proteins also diverged.

4. PARALOGY MAPPING COULD HELP IDENTIFY CANDIDATE GENES FOR MHC – ASSOCIATED DISEASES: PROSPECTS FOR THE CENTRAL MHC

Some of the MHC non-class I and II genes such as RXR, TNX, PBX and Notch are physically linked even in invertebrates such as the worm and fly (64 – 66). This leads to the hypothesis that the MHC probably began at the center and has been augmented progressively with genes such as the complement genes, the TNF and the HSP gene clusters. Sub-chromosomal or segmental duplications and inter-chromosomal translocations occurred. Humans have three derivatives of the MHC-like structures on chromosomes 1, 9 and 19 as well as the MHC on chromosome 6 and these contain *inter alia* the paralogous members of the complement family such as C3 and C5, and of the MHC class I family such as CD1 (67) and MR1 (68). If this evolutionary history is correct, it should be possible to deduce the architecture of the MHC as it existed in early vertebrates. For example, there would be a cluster corresponding to those which occur on the four distinct chromosomes in *Homo sapiens* (Hosa). According to this strategy, a complement-like gene would be present in the putative "invertebrate MHC". Class I genes, which are represented twice in Hosa (chromosomes 1 and 6), would have been added somewhat later. Class II genes would appear to have been introduced after these major events since, depending upon definition, there may be no paralogues outside the MHC.

There are differences in the precise arrangement of genes within the paralogous regions apparently because of transpositions and indels but some clusters have been substantially maintained and might be regarded as potentially functional units. However, once the paralogous regions migrated apart, they seem to have been free to diverge apparently permitting diversification of function. An interesting example is provided by the complement genes. Paralogy provides a mechanism for the development of cascades such as the C4-C3-C5, i.e. the common ancestor of these components diverged enough for each to play a somewhat different role towards a similar end. At the same time, local duplication events were able to generate polymorphisms within each paralogue.

The MHC is well-known for its association with numerous diseases, particularly autoimmune diseases such as type 1 diabetes, systemic lupus erythematosus, multiple sclerosis and rheumatoid arthritis (69 – 72). Autoimmunity is probably due to immune dysregulation and/or defects in self-nonself discrimination irrespective of the antigen specific reactions which determine which disease results. We have shown that there are common chromosomal locations for gene clusters that confer generic susceptibility to multiple autoimmune diseases (72, 73). We postulate that the chromosomes carrying paralogues of the MHC also carry genes involved in self-nonself or host-parasite discriminations, immunoclearance and immunoregulation. Polymorphisms of these genes are responsible for generic susceptibility to autoimmune disease in mammals and possibly other vertebrates. The ancestral genes in the invertebrate MHC could subserve similar regulatory functions in relation to innate immunity. If correct, it could be predicted that searches in other species would reveal associations with those regions which are syntenic to the paralogous clusters shown in Hosa. The approach can be further evaluated by examining other species with autoimmunity, such as the mouse, rat and dog, and by comparing the locations of potentially regulatory clusters in all species. The complement genes in the central MHC are prime candidates.

5. ACKNOWLEDGMENTS

OPM, NLJ, RD, SG and RLD at CMII are supported by NHMRC, IRF and C.Y. O'Connor Institute. CYY was supported by NIAMS (R01 AR43969), Columbus Children's Research Institute (299400), and Pittsburgh Supercomputing Center through NIH Center for Research Resources Cooperative Agreement grant 1P41 RR06009. Manuscript 0106 of the Centre for Molecular Immunology and Instrumentation, University of Western Australia.

6. REFERENCES

1. K Rother, GO Till, & GM Hansch: The Complement System. Springer, Berlin (1997)

2. Blanchong C. A, E. K. Chung, K. L. Rupert, Y. Yang, Z. Yang, B. Zhou & C. Y. Yu: Genetic, structural and functional diversities of human complement components *C4A* and C4B and their mouse homologs, Slp and C4. *Int Immunopharmacol* 1, 365-392 (2001)

3. Blanchong C. A, B. Zhou, K. L. Rupert, E. K. Chung, K. N. Jones, J. F. Sotos, R. M. Rennebohm & C. Y. Yu: Deficiencies of human complement component C4A and C4B and heterozygosity in length variants of RP-C4-CYP21-TNX (RCCX) modules in Caucasians: the load of RCCX genetic diversity on MHC-associated disease. *J Exp Med* 191, 2183-2196 (2000)

4. Yang Z, A. R. Mendoza, T. R. Welch, W. B. Zipf & C. Y. Yu: Modular variations of HLA class III genes for serine/threonine kinase RP, complement C4, steroid 21-hydroxylase CYP21 and tenascin TNX (RCCX): a mechanism for gene deletions and disease associations. *J Biol Chem* 274, 12147-12156 (1999)

5. Yu C. Y, K. T. Belt, C. M. Giles, R. D. Campbell & R. R. Porter: Structural basis of the polymorphism of human complement component C4A and C4B: gene size, reactivity and antigenicity. *EMBO J* 5, 2873-2881 (1986)

6. Belt K. T, C. Y. Yu, M. C. Carroll & R. R. Porter: Polymorphism of human complement component C4. *Immunogenetics* 21, 173-180 (1985)

7. Dodds A. W, X.-D. Ren, A. C. Willis & S. K. A. Law: The reaction mechanism of the internal thioester in the human complement component C4. *Nature* 379, 177-179 (1996)

8. Sepp A, A. W. Dodds, M. J. Anderson, R. D. Campbell, A. C. Willis & S. K. A. Law: Covalent binding properties of the human complement protein C4 and hydrolysis rate of the internal thioester upon activation. *Protein Science* 2, 706-716 (1993)

9. Carroll M. C, D. M. Fathallah, L. Bergamaschini, E. M. Alicot & D. E. Isenman: Substitution of a single amino acid (aspartic acid for histidine) converts the functional activity of human complement C4B to C4A. *Proc Natl Acad Sci USA* 87, 6868-6872 (1990)

10. Schifferli J. A, G. Steiger, J.-P. Paccaud & A. G. Sjoholm: Difference in the biological properties of the two forms of the fourth component of human complemetn (C4). *Clin Exp Immunol* 63, 473-477 (1986)

11. Gatenby P. A, J. E. Barbosa & P. J. Lachmann: Differences between C4A and C4B in the handling of immune complexes: the enhancement of CR1 binding is more important than the inhibition of immunoprecipitation. *Clin exp Immunol* 79, 158-163 (1990)

12. Isenman D. E. & J. R. Young: The molecular basis for the differences in immune hemolysis activity of the Chido and Rodgers isotypes of human complement component C4. *J Immunol* 132, 3019-3027 (1984)

13. Dangel A. W, A. R. Mendoza, B. J. Baker, C. M. Daniel, M. C. Carroll, L.-C. Wu & C. Y. Yu: The dichotomous size variation of human complement C4 gene is mediated by a novel family of endogenous retroviruses which also establishes species-specific genomic patterns among Old World primates. *Immunogenetics* 40, 425-436 (1994)

14. Schneider P. M, K. Witzel-Schlomp, C. Rittner & L. Zhang: The endogenous retroviral insertion in the numan complement C4 gene modulates the expression of homologous genes by antisense inhibition. *Immunogenetics* 53, 1-9 (2001)

15. Zhang W. J, F. T. Christiansen, X. Wu, L. J. Abraham, M. Giphart & R. L. Dawkins: Organization and evolution of C4 and CYP21 genes in primates: importance of genomic segments. *Immunogenetics* 37, 170-176 (1993)

16. Shen L. M, L. C. Wu, S. Sanlioglu, R. Chen, A. R. Mendoza, A. Dangel, M. C. Carroll, W. Zipf & C. Y. Yu: Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and C4B genes in the HLA class III region: Molecular cloning, exon-intron structure, composite retroposon and breakpoint of gene duplication. *J Biol Chem* 269, 8466-8476 (1994)

17. Yu C. Y, Z. Yang, C. A. Blanchong & W. Miller: The human and mouse MHC class III region: a parade of the centromeric segment with 21 genes. *Immunol Today* 21, 320-328 (2000)

18. White P. C. and P. W. Speiser: Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Endocrine Rev* 21, 245-291 (2000)

19. Carroll M. C, K. T. Belt, A. Palsdottir & Y. Yu: Molecular genetics of the fourth component of human complement and the steroid 21-hydroxylase. *Immunol Rev* 87, 39-60 (1985)

20. White P. C, A. Vitek, B. DuPont & M. I. New: Characterization of frequent deletions causing steroid 21-hydroxylase deficiency. *Proc Natl Acad Sci U S A* 85, 4436-4440 (1988)

21. Carroll M. C, R. D. Campbell & R. R. Porter: Mapping of steriod 21-hydroxylase genes adjacent to complement C4 genes in HLA, the major histocompatibility complex in man. *Proc Natl Acad Sci U S A* 82, 521-525 (1985)

22. Mauff G, B. Luther, P. M. Schneider, C. Rittner, B. Strandmann-Bellinghausen, R. Dawkins & J. M. Moulds: Reference typing report for complement component C4. *Exp Clin Immunogenet* 15, 249-260 (1998)

23. Yu C Y, C. A. Blanchong, E. K. Chung, K. L. Rupert, Y. Yang, Z. Yang, B. Zhou, and J. M. Moulds: Molecular genetic analyses of human complement components C4A and C4B. In: Manuals of Clinical Laboratory Immunology, 6th Edition. American Microbiology Press, San Diego, in press (2001)

24. Awdeh Z. L. and C. A. Alper: Inherited structural polymorphism of the fourth component of human complement. *Proc Natl Acad Sci USA* 77, 3576-3580 (1980)

25. Sim E. and S. Cross: Phenotyping of human complement component C4, a class III HLA antigen. *Biochem J* 239, 763-767 (1986)

26. Martinez-Quiles N, E. Paz-Artal, M. A. Moreno-Pelayo, J. Longas, S. Ferre-Lopez, M. Rosal & A. Arnaiz-Villena: C4d DNA sequencesof twoi infrequent human allotypes (C4A13 and C4B12) and the presence of signal sequences enhancing recombination. *J Immunol* 161, 3438-3443 (1998)

27. Schneider P. M, B. Stradmann-Bellinghausen & C. Rittner: Genetic polymorphism of the fourth component of human complement: population study and proposal for a revised nomenclature based on genomic PCR typing of Rodgers and Chido determinants. *Eur J Immunogenet* 23, 335-344 (1996)

28. Mauff G, C. A. Alper, R. Dawkins, G. Doxiadis, C. M. Giles, G. Hauptmann, C. Rittner & P. M. Schneider: C4 nonmenclature statement (1990). *Complement Inflamm* 7, 261-268 (1990)

29. Reid K. B. M: Activation and control of the complement system. *Essays Biochem* 22, 27-68 (1986)

30. Turner M. W: Mannose-binding lectin: the pluripotent molecule of the innate immune system. *Immunol Today* 17, 532-540 (1996)

31. Muller-Eberhard H. J.: Molecular organization and function of the complement system. *Ann Rev Biochem* 57, 321-347 (1988)

32. BP Morgan & CL Harris: Complement Regulatory Proteins. Academic Press, San Diego (1999)

33. Reid K. B. M, D. R. Bentley, R. D. Campbell, L. P. Chung, R. B. Sim, T. Kristensen & B. F. Tack: Complement system proteins which interact with C3b or C4b. *Immunol Today* 7, 230-234 (1986)

34. Rey-Campos J, P. Rubinstein & S. Rodriguez de Cordoba: A physical map of the human regulator of complement activation gene-cluster linking the complement genes CR-1, CR-2, CR-1, CR-2, DAF, and C4BP. *J Exp Med* 166, 246-252 (1987)

35. Atkinson J. P. and T. Farries: The separation of self and non-self in the complement system. *Immunol Today* 8, 212-215 (1987)

36. Adams M. D, S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. *et al*, *Science* 287, 2185-2195 (2000)

37. Campbell R. D, D. R. Bentley & B. J. Morley: The factor B and C2 genes. *Phil Trans R Soc Lond* B306, 367-378 (1984)

38. Colombatti A. and P. Bonaldo: The superfamily of proteins with von Willebrand factor type A-like domains: one theme common to components of extracellular matrix, hemostasis, cellular adhesion, and defense mechanisms. *Blood* 77, 2305-2315 (1991)

39. Dodds A. W & S. K. Law: the phylogeny and evolution of the thioester bond-containing proteins C3, C4 and α2 macroglobulin. *Immunol Rev* 166, 15-26 (1998)

40. Hughes A. L. and M. Yeager: Molecular evolution of the vertebrate immune system. *Bioessays* 19, 777-786 (1997)

41. Armstrong P. B.: The contribution of proteinase inhibitors to immune defense. *Trends Immunol* 22, 47-52 (2001)

42. Venter J. C, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton & *et al*.: The sequence of the human genome. *Science* 291, 1304-1351 (2001)

43. International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome. *Nature* 409, 860-921 (2001)

44. Devriendt K, H. Van den Berghe, J. J. Cassiman & P. Marynen: Primary structure of pregnancy zone protein. Molecular cloning of a full-length PZP cDNA clone by the polymerase chain reaction. *Biochim Biophys Acta* 1088, 95-103 (1991)

45. Nagase T, K. Ishikawa, R. Kikuno, M. Hirosawa, N. Nomura & O. Ohara: Prediction of the coding sequences of unidentified human genes. XV. The complete sequences of 100 new cDNA clones from brain which code for large proteins *in vitro*. *DNA Research* 7, 337-345 (1999)

46. Banyai L. and L. Patthy: The NTR module: domains of netrins, secreted frizzled related proteins, and type I procollagen C-proteinase enhancer protein are homologous with tissue inhibitors of metalloproteases. *Protein Sci* 8, 1636-1642 (1999)

47. Hortin G, A. C. Chan, K. F. Rok, A. W. Strause & J. P. Atkinson: Sequence analysis of the COOH terminus of the α-chain of the fourth component of human complement. *J Biol Chem* 261, 9065-9069 (1986)

48. SKA Law & KBM Reid: Complement. IRL Press at Oxford University Press, Oxford (1995)

49. Hugli T. E: Biochemistry and biology of anaphylatoxins. *Complement* 3, 111-127 (1986)

50. Argraves W. S, H. Tran, W. H. Burges & H. Dickerson: Fibulin is an extracellular matrix and plasma glycoprotein with repeated domain structure. *J Cell Biol* 111, 3155-3146 (1990)

51. Zhang R.-Z, T.-C. Pan, Z.-Y. Zhang, M.-G. Mattei, R. Timpl & M.-L. Chu: Fibulin (FBLN2): human cDNA sequence, mRNA expression, and mapping of the gene on human and mouse chromosomes. *Genomics* 22, 425-430 (1994)

52. Sandoval A, R. Ai, J. M. Ostresh & R. T. Ogata: Distal recognition site for the classical pathway convertase located in the C345C/Netrin module of complement component C5. *J Immunol* 165, 1066-1073 (2000)

53. Banyai L. and L. Patthy: The NTR module: domains of netrins, secreted frizzled related proteins, and type I procollagen C-proteinase enhancer protein are homologous with tissue inhibitors of metalloproteases. *Protein Sci* 8, 1636-1642 (1999)

54. Kuroda N, K. Naruse, A. Shima, M. Nonaka & M. Sasaki: Molecular cloning and linkage analysis of complement C3 and C4 genes of the Japanese medaka fish. *Immunogenetics* 51, 117-128 (2000)

55. Lagueus M, E. Perrodou, E. A. levashina, M. Capovilla & J. A. Hoffmann: Constitutive expression of a complement-like protein in Toll and JAK gain-of -function mutants of *Drosophila*. *Proc Natl Acad Sci USA* 21, 11427-11432 (2001)

56. Kimbrell D. A & B. Beutler: The evolution and genetics of innate immunity. *Nature Rev Genet* 2, 256-267 (2001)

57. Nonaka M, K. Azumi, X. Ji, C. Namikawa-Yamada, M. SAasaki, H. Saiga, A. W. Dodds, H. Sekine, M. K. Homma, Y. Matsushita, T. Fujita &: Opsonic complement component C3 in the solitary ascidian *Halocynthia roretzi*. *J Immunol* 162, 387-391 (1999)

58. Sunyer J. O, I. K. Zarkadis & J. D. Lambris: Complement diversity: a mechanism for generating immune diversity. *Immunol Today* 19, 519-523 (1998)

59. Nonaka M: Origin of the complement system. In: Origin and Evolution of the Vertebrate System. Eds: Du Pasquier L, Litman G W, Springer, Berlin, 37-50 (2000)

60. Leelayuwat C, W. J. Zhang, L. J. Abraham, D. Townend, S. Gaudieri & R. L. Dawkins: Differences in the central major histocompatibility complex between humans and chimpanzees. *Hum Immunol* 38, 30-41 (1993)

61. Dangel A. W, B. J. Baker, A. R. Mendoza & C. Y. Yu: Complement component C4 gene intron 9 as a phylogenetic marker for primates: Long terminal repeats of the endogenous retrovirus ERV-K(C4) are a molecular clock of evolution. *Immunogenetics* 42, 41-52 (1995)

62. Paz-Artal E, A. Corell, M. Alvarez, P. Varela, L. Allende, A. Mardrono, M. Rosal & A. Arnaiz-Villena: C4 gene polymorphism in primates: evolution, generation, and Chido and Rodgers antigenicity. *Immunogenetics* 40, 381-396 (1994)

63. Nakao M, J. Mutsuro, R. Obo, K. Fujiki, M. Nonaka & T. Yano: Molecular cloning and protein analysis of divergent forms of the complement component C3 from a bony fish, the common carp (Cyprinus carpio): presence of variants lacking the catalytic histidine. *Eur J Immunol* 30, 858-866 (2000)

64. Trachtulec Z, R. M. J. Hamvas, J. Forejt, H. R. Lehrach, V. Vinvek & J. Klein: Linkage of TATA-binding protein and proteasome subunit C5 genes in mice and humans reveals synteny conserved between mammals and invertebrates. *Genomics* 44, 1-7 (1997)

65. Kasahara M: The chromosomal duplication model of the major histocompatibility complex. *Immunol Rev* 167:17-32, 17-32 (1999)

66. Rached L. A, M. F. McDermott & P. Pontarotti: The MHC big bang. *Immunol Rev* 167:33-44, 33-44 (1999)

67. Yu C. Y & C. Milstein: A physical map linking the five CD1 human *thymocyte* differentiation antigen genes. *EMBO J* 8, 3727-3732 (1989)

68. Hashimoto K, M. Hirai & Y. Kurosawa: A gene outside the human MHC related to classical HLA class I genes. *Science* 269, 693-695 (1995)

69. JL Tiwari & PI Terasaki: HLA and Disease Associations. Springer-Verlag, New York (1985)

70. R Lechler & A Warrens: HLA in Health and Disease. Academic Press, San Diego (2000)

71. Dawkins R. L, F. T. Christiansen, P. H. Kay, M. Garlepp, J. McCluskey, P. N. Hollingsworth & P. J. Zilko: Disease associations with complotypes, supratypes and haplotypes. *Immunol Rev* 70, 5-22 (1983)

72. Dawkins R, C. Leelayuwat, S. Gaudieri, G. Tay, J. Hui, S. Cattley, P. Martinez & J. Kulski: Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunol Rev* 167, 275-304 (1999)

73. Dawkins R, J. Berry, P. Martinez, S. Gaudieri, J. Hui, N. Longman, J. Kulski and P. Carnegie: Potential for paralogous mapping to simplify the genetics of diseases and functions associated with MHC haplotypes. In: Major Histocompatibility Complex. Evolution, Structure and Function. Eds: Kasahara M, Springer-Verlag, Tokyo, 146-157 (2000)

74. Mauff G, C. A. Alper, Z. Awdeh, J. R. Batchelor, T. Bertrams, G. Bruun-Petersen, R. L. Dawkins, P. Demant, J. Edwards, H. Grosse-Wilde, G. Hauptmann, P. Klouda, L. lamn, E. Mollenhauer, C. Nerl, B. Olaisen, G. O'Neill, C. Rittner, M. H. Ross, V. Skanes, N. Teisberg & L. Wells: Statement on the nomenclature of human C4 allotypes. *Immunobiology* 164, 184-191 (1983)

75. Schneider P. M, M. C. Carroll, C. A. Alper, C. Rittner, A. S. Whitehead, E. J. Yunis & H. R. Colten: Polymorphism of human complement C4 and steroid 21-hydroxylase genes. Restriction fragment length polymorphisms revealing structural deletions, homoduplications, and size variants. *J Clin Invest* 78, 650-657 (1986)

76. Yu C. Y. and R. D. Campbell: Definitive RFLPs to distinguish between the human complement C4A/C4B isotypes and the major Rodgers/Chido determinants: application to the study of C4 null alleles. *Immunogenetics* 25, 383-390 (1987)

77. Zhang W. J, M. A. Degli-Esposti, T. J. Cobain, P. U. Cameron, F. T. Christiansen & R. L. Dawkins: Differences in gene copy number carried by different MHC ancestral haplotypes. Quantitation after physical separation of haplotypes by pulsed field gel electrophoresis. *J Exp Med* 171, 2101-2114 (1990)

78. Kawaguchi H, Z. Zaleska-Rutczynska, F. Figueroa, C. O'hUigin & J. Klein: C4 genes of the chimpanzee, gorilla, and orang-utan: evidence for extensive homogenization. *Immunogenetics* 35, 16-23 (1992)

79. Gaudieri S, J. K. Kulski, and R. L. Dawkins: MHC and disease associations in nonhuman primates. In: Molecular Biology and Evolution of Blood Group and MHC Antigens in Primates. Eds: Blancher A, Klein J, Socha W W, Springer-Verlag, *Berlin*, 464-490 (1997)

Key Words: α_2-Macroglobulin Protein Family; Bf/C2, C3/C4/C5, Complement, Genetics, Evolution, Central MHC, Segmental Duplication, Review

Send correspondence to: C. Yung Yu, D.Phil, Room W208, Children's Research Institute, 700 Children's Drive, Columbus, Ohio 43205-2696, USA, Tel: 614-722-2821, Fax: 614-722-2774, E-mail: cyu@chi.osu.edu

Roger L. Dawkins, M.D., Centre for Molecular Immunology and Instrumentation, University of Western Australia, PO Box 5100, Canning Vale WA 6155, Australia, Tel: 618-9397-1556, Fax: 618-9397-1559, E-mail: cmii@cyllene.uwa.edu.au

O. Patricia Martinez, Department of Pathology, University of Western Australia, Nedlands, WA 6009, Australia, E-mail: Patricia.Martinez@health.wa.gov.au