CHAPTER 2

PHYLOGENY OF THIRD COMPONENT OF COMPLEMENT, C3

John D. Lambris, Manolis Mavroidis and J. Oriol Sunyer

Of the thirty distinct complement proteins recognized to date, C3 is probably the most versatile and multifunctional molecule known. It interacts with numerous sera, cell surface, and foreign origin proteins and plays an important role in immune surveillance and immune response pathways. Due to the pivotal role C3 plays in complement activation and immune processes, the analysis of its structure and functions has been the subject of intense investigation (for detailed reviews see ref. 1). The aim of this review is to summarize recent findings pertaining to the phylogeny of C3, an area in which the phylogenetic analysis assisted us in understanding C3 structural/functional relationships. Because human C3 is the best characterized complement protein, for the purpose of this review, generalizations are made to human C3 and the numbering of amino acid residues is based on the deduced sequence of human C3.2

CHEMISTRY AND BIOLOGY

Human C3 is the most abundant complement component in serum (1-2 mg/ml). It has been characterized as a glycoprotein comprised of a 115 kDa α chain linked to a 75 kDa β chain by a single disulfide bond and noncovalent forces (Fig. 2.1). The primary structure, deduced from the cDNA sequence2 consists of 1,663 amino acids, including a 22-amino acid signal peptide. Molecular modeling of C3, based on data derived from X-ray scattering studies depicts it as a two-domain shape, with a flat ellipsoid associated with a smaller flat domain. These two domains move closer together following proteolytic activation of C3 and removal of C3a.3 Carbohydrate analysis revealed that human C3 possesses 2 N-linked carbohydrate moieties, positioned at residues 63 of the β chain

(Man$_5$GlcNac$_2$ + Man$_5$GlcNac$_3$) and 917 of the α chain (Man$_5$GlcNac$_2$ + Man$_6$GlcNac$_3$), which together account for 1.5% of the molecular weight of C3 (see Fig. 2.1).$^4$5

One of the distinguishing characteristics of C3 is its ability to bind covalently to acceptor molecules on cell surfaces$^6$ via ester or amide linkages.$^7$ This feature has been attributed to the thioester bond present within the C3d region of C3 which is sensitive to nucleophilic attack. The thioester bond is the product of an intramolecular transacylation between the thiol group of cysteine and the γ-amide group of the glutamine within the C3 sequence Gly-Cys$^{998}$-Gly-Glu-Gln$^{991}$-Asn.$^8$ This thioester moiety is also found in C4 and α$_2$-macroglobulin (α$_2$-M), two plasma proteins homologous to C3. In native C3, the thioester group appears to be protected within a hydrophobic pocket and is exposed in the C3b fragment upon cleavage of C3 by C3 convertase. Thus, the transiently expressed thioester group can then participate in a transacylation reaction with nucleophilic groups present on cell surfaces, complex carbohydrates, or immune complexes.$^8$ The deposition of C3b to surface structures is important for initiation of the membrane attack complex (MAC), for phagocytosis of foreign particles, and for enhancement of effector cell-target cell contact.

Cleavage of C3 between residues 726 and 727 (Arg-Ser) by either the classical (C4b,2a) or the alternative (C3b,Bb) pathway C3 convertases leads to the generation of C3b (185,000 kDa) and C3a (9,000 kDa). In contrast to native C3, C3b expresses multiple binding sites for other complement components, including C5, properdin (P), factors H, B, and I, C4 binding protein (C4bp), CR1 (C3b-receptor), and the membrane cofactor protein (MCP).$^9$ Binding of these proteins to C3b leads either to amplification of the C3 convertase (by B and P in the presence of factor D) and initiation of the membrane attack complex (C5), or to the inactivation of C3b (by factor I).$^{10}$ Whether amplification or inactivation occurs depends on the nature of the surface to which C3b is fixed.

The inactivation of C3b by factor I, an event by which complement activation is downregulated, proceeds in three steps$^{11-13}$ and requires one of several cofactor molecules (MCP, CR1, CR2, H, or C4bp). The cleavage of the α’ chain of C3b first between residues 1281-1282 (Arg-Ser) and then between residues 1298-1299 (Arg-Ser) of C3 liberates the C3f fragment (Mr 2,000) and yields iC3b.$^{11-14}$ A third factor I cleavage site, with CR1 or factor H serving as cofactors,$^{11-13}$ has been reported to exist at residues 932-933 (Arg-Glu) of the α’ chain of C3, generating the C3c and C3dg fragments.$^{14}$

The C3 fragments, both soluble and/or surface bound, generated during comple-

---

**Fig. 2.1. Schematic representation of the C3 molecule.** Human C3 consists of an α chain linked to a β chain by a disulfide bond. N-linked high-mannose carbohydrate moieties are present at residues 917 (α chain) and 63 (β chain) of C3. An internal thioester bond is located within the C3d region. Binding sites for receptors (CR1, CR2) and regulatory proteins (H, B, properdin) as well as conglutinin are shown. The factor I (I$_1$, I$_2$, I$_3$) and convertase cleavage sites are shown by arrows.
ment activation have the potential to bind specifically to several cell surface receptors, known as CR1, CR2, CR3, CR4, CR5, and C3a receptor. These C3-receptor interactions lead to various biological responses (for detailed reviews see also refs. 16-22).

C3 is synthesized as a single chain pre-pro molecule with α and β chains linked by a tetra-arginine sequence which is removed by a furin-like enzyme23 during post-translational modification. Following translocation through the endoplasmic reticulum to the Golgi, N-linked high-mannose type carbohydrate moieties are attached at the α and β chain (Fig. 2.1).45 The vast majority (more than 90%) of C3 biosynthesis occurs in the liver,25 but synthesis at numerous other sites (reviewed in ref. 25) may be involved in localized inflammatory processes. C3 expression is regulated by cytokines such as interferon (IFN), interleukin-1 (IL-1), and tumor necrosis factor (TNF).25 Regulation of C3 synthesis appears to be tissue-specific as in the cases of estrogen regulation in the uterus26 or vitamin D regulation in osteoblasts.27

Studies on allelic variants of C3 have demonstrated the existence of 22 different variants, the most common of which are C3F and C3S.28 The 41-kb gene for human C3 has been localized to chromosome 19.29 Of the 41 exons (ranging in size from 52-213 bases) contained in the gene, 16 are in the β chain and 25 are in the α chain.30,31 Each of the major binding sites in the α chain of C3 appear to be encoded by single exons (reviewed in ref. 25). The highly conserved thioester domain is encoded by class 1-1 exons, the type most commonly reshuffled and duplicated.32 Coding regions of C4 and C5 also contain approximately 40 exons. The exons flanking the thioester site in C4 are similar to those of C3.

Rarely occurring, inherited C3 deficiencies are characterized by a heightened susceptibility to bacterial infections that lead to purulent lesions. In addition, several C3-deficient patients have shown impaired chemotactic activity, sluggish re-

sponse of neutrophils to infectious agents, and development of immune complex disease, systemic lupus erythematosus, and membrane proliferative glomerulonephritis.33 Although the antibody response to routine immunizations is normal, an impaired switch from IgM to IgG was observed in two C3-deficient patients immunized with limited doses of the T-dependent antigen, bacteriophage ΦX174.34 A similar defective antibody response was observed in C3-deficient guinea pigs35 and in dogs.35 These findings and those made using pharmacologically C3-depleted animals suggest that C3 may play an important role in the generation of a normal immune response (reviewed in ref. 33).

**PHYLOGENY**

Although complement and other nonspecific lytic systems have been described in lower vertebrates and invertebrates, relatively little is known about the individual components that make up these systems. The main immune features of the major phyla of living species are summarized in Figure 2.2. The best phylogenetically characterized component of complement is C3, no doubt due to its prominent role in both pathways of complement system. C3-like activity has been reported in a variety of species, including invertebrates; yet, thus far, C3 has been purified only from chordates, and has been found to be present in representatives of each of the seven living classes of vertebrates as described below. Although a number of studies on the presence of various complement components in different species have been reported (for review see ref. 36) our focus here is on the C3 molecule, and the inclusion of the other complement components is beyond the scope of this review. Features of C3 from different species are summarized and compared in Table 2.2.

**MAMMALIA**

We have focused primarily on features of human C3, about which the bulk of information is known, but the structure and function of C3 in all mammals so far
investigated is very similar (Table 2.1), with a few notable exceptions. C3 has also been studied from mouse,\textsuperscript{37} rat,\textsuperscript{37,38} rabbit,\textsuperscript{39,40} pig,\textsuperscript{41,42} monkey,\textsuperscript{43} guinea pig,\textsuperscript{44,45} and cat.\textsuperscript{46}

The complete primary structures have been deduced for C3 of human,\textsuperscript{2} mouse,\textsuperscript{47,48} guinea pig\textsuperscript{49} and rat,\textsuperscript{50} and partial primary structures have been determined for C3 of rabbit.\textsuperscript{31} All of these C3s show high degrees of homology with human C3 (Table 2.2) and have a two-chain structure with the thioester site present in the α chain.

Both the classical and alternative pathways are present in all mammals as is the binding of different mammalian C3s to human CR1, CR2, H, B and properdin.\textsuperscript{42} The presence of carbohydrates, as determined by binding to \textsuperscript{125}I-labeled ConA, varies among species and may be present in either the α or β chain or both (Table 2.2).\textsuperscript{42} The difference in glycosylation is interesting because the α chain carbohydrate moiety on human C3 has been found to be involved in the binding of bovine conglutinin, an analog of which has recently been found in human serum.\textsuperscript{52}

**AVES**

C3 of the chicken and Japanese quail consists of two chain (α and β) structures with a methyamine-sensitive thioester bond as in mammals.\textsuperscript{53,54} Chicken C3 exists in multiple molecular forms, yet genetic polymorphism has not been demonstrated.\textsuperscript{53} The complete cDNA sequence of chicken C3 was obtained in our laboratory. It was found that chicken pro-C3 consists of an 18-amino acid signal peptide, a 640-amino acid β-chain (70 kDa), a 989-amino acid α-chain (111 kDa) and a RKRR linker region.\textsuperscript{55} The deduced amino acid sequence contains an internal thioester sequence and three potential N-glycosylation sites, all of which make up the α-chain (one on C3d and two on the C-terminal 40 kDa α chain fragment); ConA binding studies demonstrated that chicken C3 is indeed

---

**Fig. 2.2.** A simplified scheme of the evolutionary relationships of the major phyla of living organisms and their immune system features. ACP, alternative complement pathway; CCP, classical complement pathway; MAC, membrane attack complex; Ab, antibodies; C3, complement component C3; MHC, Major histocompatibility complex; T, thymus; S, spleen.
### Table 2.1. Summary of structural and functional properties of C3 from different species

<table>
<thead>
<tr>
<th>Species</th>
<th>Hu</th>
<th>Po</th>
<th>Rb</th>
<th>Ra</th>
<th>GP</th>
<th>Mo</th>
<th>Ch</th>
<th>Xe</th>
<th>Co</th>
<th>Ax</th>
<th>SB</th>
<th>Tr</th>
<th>Ha</th>
<th>La</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of residues</strong></td>
<td>1663</td>
<td>nd</td>
<td>nd</td>
<td>1663</td>
<td>1666</td>
<td>1661</td>
<td>1652</td>
<td>nd</td>
<td>1651</td>
<td>nd</td>
<td>nd</td>
<td>1640&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1620</td>
<td>1660</td>
</tr>
<tr>
<td><strong>% Identity/Similarity</strong></td>
<td>100/100</td>
<td>nd</td>
<td>80/87&lt;sup&gt;a&lt;/sup&gt;</td>
<td>78/86</td>
<td>78/86</td>
<td>77/85</td>
<td>54/70</td>
<td>52/66&lt;sup&gt;a&lt;/sup&gt;</td>
<td>51/66</td>
<td>nd</td>
<td>nd</td>
<td>44/59</td>
<td>31/47</td>
<td>32/48</td>
</tr>
<tr>
<td><strong>Chain structure</strong></td>
<td>α/β</td>
<td>α/β</td>
<td>α/β</td>
<td>α/β</td>
<td>α/β</td>
<td>α/β</td>
<td>α/β</td>
<td>α/β</td>
<td>α/β</td>
<td>α/β</td>
<td>α/β</td>
<td>α/β</td>
<td>α/β</td>
<td>α/β</td>
</tr>
<tr>
<td><strong>Mr α/β X 103</strong></td>
<td>115/75</td>
<td>112/68</td>
<td>119/72</td>
<td>115/65</td>
<td>115/65</td>
<td>113/66</td>
<td>118/68</td>
<td>112/83</td>
<td>107/64</td>
<td>110/73</td>
<td>112/70</td>
<td>112/70</td>
<td>115/77</td>
<td>84/74</td>
</tr>
<tr>
<td><strong>Thioester in</strong></td>
<td>α</td>
<td>α</td>
<td>α</td>
<td>α</td>
<td>α</td>
<td>α</td>
<td>α</td>
<td>α</td>
<td>α</td>
<td>α</td>
<td>α</td>
<td>α</td>
<td>α</td>
<td>α</td>
</tr>
<tr>
<td><strong>ConA binding to</strong></td>
<td>α/β</td>
<td>α</td>
<td>α/β</td>
<td>α</td>
<td>α</td>
<td>α</td>
<td>α</td>
<td>β</td>
<td>-</td>
<td>β</td>
<td>α</td>
<td>β</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td><strong>Human CR1 binding</strong></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>nd</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>nd</td>
<td>-</td>
<td>nd</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td><strong>Human CR2 binding</strong></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>nd</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>nd</td>
<td>-</td>
<td>nd</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td><strong>Human C5 binding</strong></td>
<td>+</td>
<td>nd</td>
<td>+</td>
<td>nd</td>
<td>+</td>
<td>+</td>
<td>nd</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td><strong>Human H binding</strong></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Human B binding</strong></td>
<td>+</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td><strong>Human P binding</strong></td>
<td>+</td>
<td>nd</td>
<td>+</td>
<td>nd</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td><strong>Human MCP binding</strong></td>
<td>+</td>
<td>nd</td>
<td>+</td>
<td>nd</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td><strong>Classical pathway</strong></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><strong>Alternative pathway</strong></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** Hu, human; Mo, mouse; Ra, rat; Rb, rabbit; Xe, Xenopus, Ax, axolotl; Co, cobra; GP, guinea pig; Tr, trout; Po, pig; Ch, chicken; SB, sea bream; La, lamprey; Ha, hagfish; n.d., not detected.

<sup>a</sup> Based on partial sequence; <sup>b</sup> Functioning only in opsonization; <sup>c</sup> Without the signal peptide.

### Table 2.2. Amino acid sequence conservation between C3 and other related proteins

<table>
<thead>
<tr>
<th></th>
<th>Hu C3</th>
<th>Rb C3&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Ra C3</th>
<th>Mo C3</th>
<th>GP C3</th>
<th>Ch C3</th>
<th>Xe C3&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Co C3</th>
<th>Tr C3</th>
<th>La C3</th>
<th>Ha C3</th>
<th>Hu C4</th>
<th>Mo C5</th>
<th>Hu α&lt;sub&gt;2&lt;/sub&gt;M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human C3</td>
<td>100</td>
<td>80/87</td>
<td>78/85</td>
<td>77/85</td>
<td>78/86</td>
<td>54/70</td>
<td>52/66</td>
<td>51/66</td>
<td>44/59</td>
<td>32/48</td>
<td>31/47</td>
<td>28/43</td>
<td>28/45</td>
<td>21/35</td>
</tr>
<tr>
<td>Rabbit C3&lt;sup&gt;a&lt;/sup&gt;</td>
<td>100</td>
<td>80/87</td>
<td>79/86</td>
<td>78/89</td>
<td>52/67</td>
<td>51/65</td>
<td>50/66*</td>
<td>43/58</td>
<td>33/48</td>
<td>28/45</td>
<td>27/42</td>
<td>26/45</td>
<td>18/35</td>
<td></td>
</tr>
<tr>
<td>Rat C3</td>
<td>100</td>
<td>90/94</td>
<td>81/88</td>
<td>54/70</td>
<td>53/68</td>
<td>52/67</td>
<td>44/58</td>
<td>33/48</td>
<td>29/45*</td>
<td>27/42</td>
<td>29/45</td>
<td>21/36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse C3</td>
<td>100</td>
<td>80/87</td>
<td>54/70</td>
<td>53/68</td>
<td>51/66</td>
<td>43/60</td>
<td>32/48</td>
<td>29/45*</td>
<td>27/43</td>
<td>28/45</td>
<td>21/35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guinea Pig C3</td>
<td>100</td>
<td>53/70</td>
<td>53/67</td>
<td>52/67</td>
<td>45/60</td>
<td>33/49*</td>
<td>30/46</td>
<td>30/46</td>
<td>29/46</td>
<td>25/40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicken C3</td>
<td>100</td>
<td>56/69</td>
<td>58/70</td>
<td>46/61</td>
<td>34/52</td>
<td>32/48</td>
<td>31/46</td>
<td>28/45</td>
<td>23/38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xenopus C3&lt;sup&gt;a&lt;/sup&gt;</td>
<td>100</td>
<td>55/68</td>
<td>47/60</td>
<td>35/52</td>
<td>32/49</td>
<td>30/46</td>
<td>28/45</td>
<td>24/39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cobra C3</td>
<td>100</td>
<td>44/60</td>
<td>32/47</td>
<td>30/46</td>
<td></td>
<td>27/42</td>
<td>27/43</td>
<td>21/35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trout C3</td>
<td>100</td>
<td>33/50</td>
<td>30/46</td>
<td>27/42*</td>
<td>28/45</td>
<td>23/37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lamprey C3</td>
<td>100</td>
<td>33/48</td>
<td>29/45</td>
<td>29/45</td>
<td>22/35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hagfish C3</td>
<td>100</td>
<td>28/42</td>
<td>26/43</td>
<td>21/35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human C4</td>
<td>100</td>
<td>27/42</td>
<td>22/35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human C5</td>
<td>100</td>
<td>11/27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human α&lt;sub&gt;2&lt;/sub&gt;M</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: Hu, human; Mo, mouse; Ra, rat; Rb, rabbit; Xe, Xenopus; Co, cobra; GP, guinea pig; Tr, trout; Po, pig; Ch, chicken; La, lamprey; Ha, hagfish; % Identity/Identity plus Similarity.

All alignments were performed using Clustal program of PcGENE with open gap cost 7 and unit gap cost 2 except the ones indicated with * where open gap cost 15. All numbers are rounded to nearest integer.

<sup>a</sup> Partial amino acid sequence.

Phylogeny of Third Component of Complement, C3

glycosylated at sites (unpublished observations). Of all the C3 so far analyzed, chicken C3 is the only one which is glycosylated in the C3d fragment. The convertase and second factor I cleavage site, predicted from the deduced amino acid sequence to be Arg-Ser (Fig. 2.3), were confirmed by sequencing the zymosan eluted C3 fragments fixed on it upon complement activation. N-terminal amino acid sequence of C3 chains showed that: (a) pro-C3 is indeed cleaved at the RKRR linkase sequence to generate the mature two chain molecule and (b) the β chain of chicken C3 is blocked.

Quail complement can be activated by zymosan, but not by CVF as can that of the chicken. These data suggest that the alternative pathway is present in quail, although it differs from the pathway found in the chicken and in mammals. The chicken alternative pathway, however, appears to be similar to that of mammals. One of its components, factor B, recently purified and partially sequenced (200 residues), is thought to function in both classical and alternative pathways. This hypothesis is based on the high amino acid sequence similarity to both mammalian C2 and factor B and the apparent lack of a protein with exclusive C2 activity in chicken plasma. To determine whether this is indeed the case will require further investigations. Koppenheffer has found that the terminal components in chicken serum can be activated directly by C1 with interaction of an intermediate component through a Ca<sup>2+</sup>-dependent mechanism. The physiologic role of this pathway is not known; it may represent a vestigial activation pathway.

**REPTILIA**

The cobra complement system is activated via either the classical pathway or an alternative pathway and functions similarly to that of mammals. The recent cloning of cDNA sequencing of cobra (Naja naja kaouthia) C3 showed that it is a two-chain (α/β) polypeptide of 1,651-amino acid residues with an internal thioester bond in the α chain that has a high degree of homology with C3 of other species. In agreement with the absence of glycosylation sites in its deduced protein sequence, cobra C3 does not have any Con A binding carbohydrates. A complement activating protein that is structurally and functionally related to C3 is cobra venom factor (CVF), which has been isolated from the venom of several species of cobra. It is a three-chain molecule, that resembles C3c and contains three oligosaccharide chains with composition Galα1-3Galβ1-4(Fucα1-3)GlcnAcβ1. When exposed to B and D, CVF forms a stable enzyme complex mimicking the C3bBb convertase, but it cannot be disassembled by H, thereby resulting in a much longer half-life for the convertase. Hence, CVF continuously

---

<table>
<thead>
<tr>
<th>C3_CONVERTASE</th>
<th>K17</th>
<th>E,17</th>
<th>183</th>
<th>17</th>
<th>181</th>
<th>182</th>
<th>cef</th>
<th>1W0</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUMAN C3</td>
<td>GLARSNLDDE ...II</td>
<td>KNTVARSTLD PRLGREGVQ KEDIPP</td>
<td>SSXKTHRI ZWESAALRS EEKT</td>
<td><strong><strong>PVEH</strong> Y</strong>******</td>
<td><strong><strong>PVEH</strong> Y</strong>******</td>
<td><strong><strong>PVEH</strong> Y</strong>******</td>
<td>1W0</td>
<td></td>
</tr>
<tr>
<td>RABIT C3</td>
<td><em><em><strong>DV</strong></em> ...</em>*</td>
<td>********* H<strong>NQ</strong> R**VHA</td>
<td><strong><strong>PVEH</strong> Y</strong>******</td>
<td>1W0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RABIT C3</td>
<td><em><em><strong>DV</strong></em> ...</em>*</td>
<td>********* H<strong>NQ</strong> R**VHA</td>
<td><strong><strong>PVEH</strong> Y</strong>******</td>
<td>1W0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOUSE C3</td>
<td><em><em><strong>EVE</strong></em> ...</em>*</td>
<td>********* H<strong>NQ</strong> R**VHA</td>
<td><strong><strong>PVEH</strong> Y</strong>******</td>
<td>1W0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRAB C3</td>
<td>E**<em>EV</em>DA ...<em>L</em></td>
<td>E<strong>KIVE</strong> K<strong>TN</strong> EKCVK</td>
<td><strong><strong>PVEH</strong> Y</strong>******</td>
<td>1W0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRAB C3</td>
<td>E**<em>EV</em>DA ...<em>L</em></td>
<td>E<strong>KIVE</strong> K<strong>TN</strong> EKCVK</td>
<td><strong><strong>PVEH</strong> Y</strong>******</td>
<td>1W0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COBRA C3</td>
<td>F***DFDCH ...<em>L</em></td>
<td>KNT<strong>PIE</strong> SVFQG** G<strong>T</strong></td>
<td>K<strong>EV</strong>EY** M<strong>X</strong></td>
<td><em><strong><strong>PVEH</strong> Y</strong></em>*****</td>
<td><strong><strong>PVEH</strong> Y</strong>******</td>
<td>1W0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TROUT C3</td>
<td>E<em><strong>EVE</strong></em> ...<em>E</em></td>
<td>R<strong>BTH</strong>E** U<strong>F</strong></td>
<td>G<strong>Y</strong>E<strong>Y</strong> M<strong>X</strong></td>
<td><em><strong><strong>PVEH</strong> Y</strong></em>*****</td>
<td><strong><strong>PVEH</strong> Y</strong>******</td>
<td>1W0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TROUT C3</td>
<td>E<em><strong>EVE</strong></em> ...<em>E</em></td>
<td>R<strong>BTH</strong>E** U<strong>F</strong></td>
<td>G<strong>Y</strong>E<strong>Y</strong> M<strong>X</strong></td>
<td><em><strong><strong>PVEH</strong> Y</strong></em>*****</td>
<td><strong><strong>PVEH</strong> Y</strong>******</td>
<td>1W0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAGFISH C3</td>
<td>E*<strong>D</strong>O<strong>E</strong> ...<em>E</em></td>
<td>ENH<strong>E</strong>E<strong>E</strong> E<strong>V</strong>E**</td>
<td>E<strong>N</strong>F<strong>E</strong>E<strong>E</strong> E<strong>V</strong>E**</td>
<td><em><strong><strong>PVEH</strong> Y</strong></em>*****</td>
<td><strong><strong>PVEH</strong> Y</strong>******</td>
<td>1W0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAGFISH C3</td>
<td>E*<strong>D</strong>O<strong>E</strong> ...<em>E</em></td>
<td>ENH<strong>E</strong>E<strong>E</strong> E<strong>V</strong>E**</td>
<td>E<strong>N</strong>F<strong>E</strong>E<strong>E</strong> E<strong>V</strong>E**</td>
<td><em><strong><strong>PVEH</strong> Y</strong></em>*****</td>
<td><strong><strong>PVEH</strong> Y</strong>******</td>
<td>1W0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2.3. Amino acid sequence comparison of C3 convertase and factor I cleavage sites between C3 from different species and human. Numbering of the human C3 residues is adapted from DeBruijn and Fey after the signal peptide is subtracted. Asterisks indicate identical residues and periods indicate gaps introduced for maximal sequence alignment. The factor I (I1, I2, I3), kallikrein, and convertase cleavage sites are shown by arrows. If indicates putative factor I cleavage sites.
cleaves C3 and C5 in species other than cobra, leading to complement depletion in animals injected with CVF.\textsuperscript{60} Two mRNAs that have 90\% sequence identity have been identified for this protein. They both also show 90\% sequence identity to cobra C3,\textsuperscript{61} which implies that there are at least three different C3-related genes in cobra genome. Although the sequence of the thioester site is conserved in these isoforms, the site is absent from the mature CVF protein due to removal of the C3d peptide during processing. The physiologic role of CVF is still unclear.

**AMPHIBIA**

C3 has been purified from the African clawed frog, *Xenopus*,\textsuperscript{62} and the salamander, axolotl, (*Ambystoma mexicanum*).\textsuperscript{63} It has been found to resemble mammalian C3 with the characteristic two-chain (\(\alpha, \beta\)) structure and a thioester bond in the \(\alpha\) chain. Both the classical and alternative activation pathways are present in *Xenopus laevis* and presumably in axolotl as well, although this has yet to be determined. Among the frog complement proteins identified and/or characterized are factor I,\textsuperscript{64} C4,\textsuperscript{65} C1\textsuperscript{66} and complement receptors.\textsuperscript{67} Our laboratory has evidence for the presence of proteins with factor H, B, I and D activities in *Xenopus laevis* serum. The cDNA sequence of the C-terminal \(\alpha\) chain region (residues 1299-1641 based on human C3 numbering) of *Xenopus laevis* C3\textsuperscript{68} and 95\% of the C3 cDNA sequence of a *Xenopus laevis*/*Xenopus gilli* (*Xenopus LG*) hybrid have been obtained.\textsuperscript{69} The obtained sequence contains one potential N-glycosylation site in the \(\beta\) chain, which confirms the ConA binding to \(\beta\) but not to \(\alpha\) chain of C3. The deduced amino acid sequence shows that the C3 convertase and second factor I cleavage sites (Arg-Ser) are conserved in *Xenopus C3*; protein sequencing of the *Xenopus C3* fragments fixed on zymosan during complement activation proved that indeed *Xenopus C3* is cleaved by *Xenopus C3* convertase and factor I at these sites. When the *Xenopus LG* liver cDNA library was screened with different C3 oligonucleotide probes, we identified a C3 clone containing a deletion of 2,502 bases, in addition to full-length C3. The former suggests the presence of a novel C3 transcript in *Xenopus LG* liver.\textsuperscript{69} The presence of this C3 transcript was confirmed by RT-PCR using *Xenopus* liver mRNA and specific oligonucleotide probes. The molecule coded by this novel C3 transcript contains the \(\beta\) chain of C3 and the first 59 residues and the last 103 residues of the \(\alpha\) chain with a molecular weight of 102 kDa. Thus, this truncated C3 is missing many of the C3-ligand binding sites that exist in the \(\alpha\) chain of C3 and it may be the analog of truncated human C3 molecule isolated from human serum allergic dermatitis patients which acts as inhibitor of eosinophil cytotoxicity and neutrophil adherence.\textsuperscript{70}

**OSTEICHTHYES**

The bony fishes possess a well developed complement system with both alternative and classical pathways, similar to those found in mammals.\textsuperscript{71} However, the alternative pathway complement titers in fish appear to be very high when compared to those of mammals.\textsuperscript{72,73} The classical pathway complement titers are only two or three times higher than those of the alternative pathway. In fish, both the optimal temperature for complement activity and the temperature in which complement can be inactivated are considerably lower than those of mammalian complement.\textsuperscript{71} This may reflect the ambient environmental temperature of the water in which fish live. Because the antibody response is not as well developed in fish as in mammals and since antibody production is diminished at lower temperatures,\textsuperscript{74} it is postulated that the alternative pathway plays a leading role in the complement system of bony fishes.

The biochemical characterization of fish complement components is incomplete since D, C1, C3, C4, C4bp and C5 are the only components that have been purified.\textsuperscript{75-79} DNA sequences are only available for C3,\textsuperscript{80} C4bp,\textsuperscript{79a} and C9.\textsuperscript{81} C3s have been isolated from carp,\textsuperscript{82} the gilthead-sea bream (unpublished observations) and trout.\textsuperscript{42,78}
Trout C3s exist in two isoforms, with (C3-1) and without (C3-2) hemolytic activity. They differ in their antigenicity and tryptic peptide map. Both forms have the characteristic two-chain structure (α,β) and contain a thioester site in the α-chain. The deduced amino acid sequence of trout C3-1, when compared with that of human C3, mouse C3, human C4, mouse C5, and human α2-macroglobulin, is 44%, 45%, 28%, 24%, and 15% identical, respectively. In our laboratory we have recently purified and partially sequenced two C3-related proteins termed C3-3 and C3-4. C3-3 is comprised of an α and β chain with molecular weights similar of those from C3-1. C3-3 shows 50% amino acid sequence identity to C3-1 (500 residues analyzed). It fails to be fixed to zymosan upon complement activation (unpublished observations). C3-4 is also comprised of an α and β chain. It shows several amino acid differences when compared with trout C3-1 and trout C3-3 (50 residues analyzed). Both C3-3 and C3-4 contain ConA binding carbohydrates in both the α and β chains in contrast to C3-1 which has such moieties only in the β chain. An antibody directed to C3-3 failed to recognize C3-1 and C3-4, which when combined with the sequence data suggests important structural differences among all these C3s. Whether these proteins are functionally related to the complement system or represent a new family of proteins with different functions remains unclear. The presence of these proteins in other vertebrates with novel functions, different from those of complement, is of great interest to us and is under investigation in our laboratory.

Sea bream C3 presents some notable differences when compared to trout C3. Sea bream C3 contains ConA binding carbohydrates in the α chain, whereas trout has these carbohydrates in the β-chain. The trypsin peptide map and reactivity of trout and sea bream C3 with anti-CVF, and anti-lamprey, anti-trout, anti-Xenopus, anti-axolotl, anti-chicken, anti-rat, and anti-human C3 also show significant differences between these two proteins. For example, anti-CVF antibody reacts strongly with trout C3-1 although it does not recognize sea bream C3. Anti-Xenopus C3 antibody is very reactive with sea bream C3, but it is unreactive with trout C3-1. A partial amino acid sequence of sea bream C3 (90 amino acids) revealed 49% identity with mouse C3 and 67% identity with trout C3-1; trout C3-1 showed 41.3% identity when the same amino acids were compared to mouse C3. These findings, which agree with previous data, suggest that trout is a more primitive teleost than sea bream. In sea bream, complement titers and C3 levels are reduced significantly when fish are chronically stressed for a period of two weeks (Sunyer et al, submitted).

CHONDRICHTHYES

The cartilaginous fishes, represented by the nurse shark Ginglymostoma cirratum, are protected by a rudimentary immune system including complement proteins and immunoglobulins imparting "natural immunity." Antibodies in the nurse shark are unlike those in mammals in that recognition sites to numerous antigens exist prior to immunization. The complement system consists of six complement components, interacting sequentially and providing classical and alternative activation pathways, as well as a terminal, lytic pathway.84 The simpler nature of the shark complement system prohibits direct comparison of these components to mammalian subunits, yet their identity as complement components is ascertained by their ability to interact with mammalian functional counterparts.

CYCLOSTOMATA

The jawless fishes of the class Agnatha are the most primitive of all living vertebrates. The only surviving members of this class, the cyclostomes (round-mouthed eels), are lampreys and the more primitive hagfishes. An alternative pathway in lamprey similar to that in mammals has been identified based on the ability of zymosan to activate this pathway. Lamprey complement was found to lack the macromolecular attack complex (MAC), responsible for
cell lysis upon complement activation.\textsuperscript{85} A protein with opsonic activity homologous to mammalian C3/C4 has been isolated from lamprey (\textit{Lamprea japonica}) and has recently been cloned.\textsuperscript{86} This protein is similar to C4 in its three-chain structure (\(\alpha\), \(\beta\), and \(\gamma\))\textsuperscript{87} although it is more homologous to C3 in amino acid sequence (Tables 2.1 and 2.2). It has been observed, however, that: (a) the C3 convertase cleavage site in lamprey C3 is similar to that in C4 and not to that in C3 (Fig. 2.3) and that (b) Asn-Asn and Glu-Glu sequences are found in the first and second factor I cleavage sites in lamprey C3.

Hagfish C3, which was mistakenly thought to be hagfish Ig,\textsuperscript{88} is a two-chain structure. It contains a thioester site in the \(\alpha\) chain\textsuperscript{89} and acts as an opsonin.\textsuperscript{89} The deduced primary structure from the cDNA of hagfish C3, however, shows one cluster of four basic residues (RRKR) between the putative \(\alpha\) and \(\beta\) chains and another cluster of three basic residues (RRR) between the putative \(\beta\) and \(\gamma\) chains,\textsuperscript{90} which seems to suggest a three-chain structure similar to that of lamprey C3. These findings lead to the hypothesis that hagfish complement may have C3 in two different forms. If the obtained cDNA sequence corresponds to the isolated two-chain C3, then this would indicate that hagfish C3 cannot be cleaved at the RRR linker between the putative \(\beta\) and \(\gamma\) chains. Data show that the N-terminus sequence of \(\alpha\)-chain and \(\gamma\)-chains of the purified three-chain hagfish C3 corresponds to the \(\alpha\)-chain and to \(\gamma\)-chains minus the first residue. Thus, these data support this hypothesis and suggest that this three-chain molecule is the product of enzymatic degradation.

From all data so far reported it appears that cyclostome complement is representative of an intermediate stage in complement phylogeny. In support of this hypothesis is also the amino acid sequence identity of lamprey factor B to that of mouse factor B and C2 which was found to be 33\% and 29\%, thus suggesting that lamprey complement represents a stage before the Bf/C2 gene duplication.\textsuperscript{91}

\section*{Invertebrate Species}

Immune capabilities, albeit primitive, have been detected in a wide variety of invertebrates. Most of the invertebrate immune responses appear to be non-specific and non-inducible. Nevertheless, adaptive immune recognition has been reported in some invertebrates.\textsuperscript{92} Some functional studies show complement-like activity in some of the invertebrate phyla. The presence of an alternative complement-like pathway in lobster, horseshoe crab, and also in starfish has been implied by experiments showing that hemolymph from these species can be activated by CVF.\textsuperscript{93,94} The nature of components involved in these processes has not yet been characterized. In mammals, such activation requires factor B, which suggests that an analogous molecule exists in these species. Similar circumstantial evidence suggests the presence of an alternative pathway in other invertebrate species. CVF has been found to induce C3 convertase activity in a substance present in the hemolymph of the wax moth \textit{Galleria mellonella}.\textsuperscript{95} Nevertheless, such activity may be mediated by any other kind of protease, since in mammals, for instance, trypsin and elastase have the same capabilities. Components of the coelomic fluid of the annelid \textit{Lumbricus terrestris}\textsuperscript{96} and the echinoid \textit{Strongylocentrotus droebachiensis}\textsuperscript{97} reportedly have receptors for C3b. These components are capable of cleaving mammalian C3, although no mechanism has yet been defined. The only molecule with structural similarities to C3 so far purified from invertebrates is the protease inhibitor \(\alpha_2\)M. It has been isolated from the horseshoe crab \textit{Limulus polyphemus},\textsuperscript{98} the crayfish \textit{Pacifastacus leniusculus}\textsuperscript{99} and the lobster \textit{Homarus americanus}.\textsuperscript{100} Other invertebrate proteins showing similarities to complement proteins are a hemolysin from the coelomocytes of the equinoderm \textit{Holothuria polii},\textsuperscript{101} and the factor C of the horseshoe crab \textit{Limulus polyphemus}.\textsuperscript{102} The former appears to be similar to perforin which has considerable amino acid sequence similarity to the proposed binding domains of the proteins of the complement mem-
brane attack complex. $^{103}$ Limulus factor C is part of the blood clotting system of the horseshoe crab, which plays an important role in defense against bacteria. It contains five short consensus repeats (SCRs), like those found in many components of the complement system. $^{102}$ In many arthropods the major defense system is the prophenoloxidase activating system (proPO). It resembles the complement system in that it comprises a complex cascade of serine proteases which is activated by β-1,3 glucans (laminarin, zymosan) and also by LPS. $^{104,105}$ The final products have been reported to have opsonic, coagulating, fungicidal and bactericidal activities. However, no structural homologies with the complement system have been found yet.

**COMPARATIVE ANALYSIS OF C3**

The phylogenetic analysis of C3 from different species is instrumental in delineating the structural elements involved in its different functions. Below we will compare the C3 sequences and briefly review some of the areas that phylogenetic analysis assisted us in understanding C3 structural/functional relationships.

Several C3 regions have been found to be highly conserved (Fig. 2.4, regions A-K). Regions F, and J have been functionally analyzed and include the thioester and the properdin binding site, respectively. The thioester site, GCGEE, is 100% conserved in all species and the amino acid sequences of the surrounding regions are highly similar. The conservation of the thioester bond and its surrounding hydrophobic amino acids emphasizes the functional importance of this region in maintaining, the capacity of C3 to attach to surfaces throughout the evolution. The hydrophobic nature of the region surrounding the thioester bond in C3, C4, and $\alpha_2$-M may shield the thioester from the aqueous environment and nucleophilic attack.

![Fig. 2.4. Average similarity profile among different C3. The average similarity of C3 from different species was calculated using the PLOTSIMILARITY program of the GCG sequence analysis software (Genetics Computer Group, 1991) with a sliding window of 25 residues. The sequences were aligned using the PILEUP program. Segments A-K and I-V are areas of high and low similarity respectively. Regions corresponding to the functional sites of C3 are indicated. The amino acid sequence of different members is based on the indicated references: human C3, $^{2}$ mouse C3, $^{47,48}$ rat C3, $^{50}$ rabbit C3, $^{18,31}$ guinea pig C3, $^{49}$ lamprey C3, $^{46}$ Xenopus C3, $^{62}$ to be published elsewhere), $^{63}$ chicken C3, $^{55}$ cobra C3, $^{58}$ trout C3, $^{42}$](image-url)
The areas J and K of C3 (Fig. 2.4) contain 13 out of 27 Cys of the C3 molecule, 12 of which form disulfide bonds within the 40 kDa fragment and one with the 27 kDa α chain fragment. The area K (about 170 residues) found in all C3, C4, and C5 molecules is absent from α2M molecule and contains the Cys that links the 40 kDa and 27 kDa fragments (Fig. 2.1); the disulfide bridge pattern of other Cys in C3 is also different from that in α2M. These findings, suggest that, although the complement proteins C3/C4/C5 contain many domains with similar gross conformation to α2M, there are significant differences. The properdin (J) binding site is located in an area of high conservation while the CR1, CR2 (IV) and C3aR binding sites are located in areas of low conservation. The I-V segments are areas of low similarity and represent an insertion/deletion in some of the C3s.

The phylogenetic analysis of C3 has greatly assisted studies dealing with the localization of the properdin binding site in human C3\(^{106}\) and in proving that the RGD sequence in C3 is not involved in CR3 binding.\(^{107}\) In addition, the phylogenetic analysis proved to be essential in defining the structural elements involved in the interaction of CR2 with C3d and gp350 protein of Epstein-Barr Virus (EBV). The findings show that: (a) mouse C3 binds to human CR2 while mouse CR2 can bind human C3d but not EBV\(^{108}\) and (b) the binding site in gp350 of EBV for human CR2 has sequence similarity to the CR2 binding site in C3d\(^{109,110}\) suggest that CR2 possesses additional interaction sites for EBV. This is further supported by Scatchard analysis of gp350 binding to CR2, which has both high and low affinity receptor binding sites. Because the human CR2 binding site in mouse C3 is conserved, it was speculated that for efficient binding of EBV to CR2 both sites would be necessary. Recent data by Molina et al.\(^{111}\) using chimeric CR2 molecules constructed by exchanging Mo SCR1-4 as well as synthetic peptides, confirmed that CR2 has two different EBV binding sites. Using chimeric human/mouse CR2 molecules, these same authors as well as Martin et al.\(^{112}\) were able to dissect the EBV, C3d, and OKB7 antibody binding sites and to demonstrate that these sites are different. Because these ligands inhibited each other, it was thought that they were binding to the same site.

The phylogenetic C3 data have also enhanced our understanding of C3-CR1, CR2, H, and B interactions. The analysis of human H, B, CR1 and CR2 binding to the C3s from different species showed that H binds to human, rabbit, porcine, trout and sea bream C3;\(^{42,69}\) CR1 binds to human, rabbit, porcine and Xenopus C3b; CR2 binds to human, rabbit, mouse and Xenopus C3; and B binds to human and porcine C3 (see Table 2.1 and ref. 42). These data suggest that the conservation of sequences comprising the binding sites for H, P, B, CR1, and CR2 are important in maintaining binding to these molecules.\(^{42}\) Human C3 was recently found by our laboratory to have multiple interaction sites for human C3b/C4b binding proteins CR2, H, and B.\(^{113}\) The differential binding of the human factor H, factor B, CR1, and CR2 to different C3s made possible the conclusion that although these four molecules bind to the same domain in human C3, the exact binding sites are different.\(^{42}\)

Moreover, the identification of factor I- and H-like proteins in different species and the determination of factor I specificity has been assisted by phylogenetic studies. The cleavage of human C3b to iC3b by reptile, amphibian, and bony fish plasma proteins has suggested the presence of factor I- and H-like proteins in these species.\(^{114,115}\) Although some data on the structural and functional conservation of C3 interactions have appeared in the literature, they are limited and sometimes speculative; the specific C3 fragments mediating these interactions, as well as the interactions of C3 with autologous C3 ligands, are unknown. Studies from this laboratory have addressed the interactions of C3 from different species with autologous C3-binding proteins. Electrophoretic analysis of the C3
fragments fixed on zymosan after activation of trout, chicken, axolotl, and *Xenopus* complement showed a similar degradation pattern to that observed for human C3. N-terminal amino acid sequence of the 68 kDa and 43 kDa fragments showed that these fragments are analogs of human C3 fragments generated by C5 convertase, and factors I and H. While similar fragments were found in the presence or absence of EGTA (which inhibits the classical complement pathway), no fragments were detected when the activation of the different sera was performed in the presence of EDTA (which inhibits both complement pathways). These data suggest that the species so far tested possess proteins with functions similar to those of human factors B, D, I, and H.

In addition, these data and those of others show that the cleavage site for the C3 convertase (human C3 residues 726-727) is conserved within the C3s from the species tested (Fig. 2.3) and that it is an Arg-Ser. An exception, however, is lamprey C3. When its sequence was aligned to that of other C3, the site that aligns to Arg-Ser was found to be Arg-Asn (no Arg-Ser were found in the surrounding). At the present, it has not yet been confirmed that lamprey C3 is indeed cleaved at this site, or on a site adjacent to it. In relation to factor I-mediated cleavages, the sequence Arg-Ala/Glu is found (based on cDNA sequencing) at the first factor I cleavage site of *Xenopus*, trout, chicken and cobra C3, instead of the Arg-Ser found in mammalian C3 (human C3 residues 1281-1282). At the second factor I cleavage site of trout and cobra C3, an Arg-Thr was found instead of Arg-Ser (human C3 residue 1298-1299). The cleavage of trout C3b by factor I at this Arg-Thr bond was confirmed by sequencing the zymosan-bound fragments of C3 upon complement activation. Alignment of lamprey and hagfish C3 sequences to those of C3 from other species showed that lamprey and hagfish C3 are lacking an Arg-Ser on the sequence surrounding the sites that correspond to first and second factor I cleavage sites of mammalian, chicken, trout, and cobra C3. At the present is unknown whether factor I mediates the cleavage of Arg-Ala/Glu and where it cleaves lamprey and hagfish C3. Concerning the site that iC3b is cleaved to generate C3c and C3dg are the observations that neither Arg-Ser or Arg-Thr sites are found at the putative factor I cleavage sites of C3. In human C3, this site may be an Arg-Glu. Although, no Arg-Ser factor I cleavage sites were found near the Arg-Glu bond, potential cleavage sites were found, which include Arg-Thr and Lys-Thr. If indeed iC3b from species other than human are cleaved to C3c and C3dg, then either this cleavage is mediated by an enzyme different than factor I, or factor I from other species has specificity for Arg/Lys-Thr bonds. In either case, the cleavage of C3b could start in any of the above sites. The findings of Nilsson-Ekdahl et al support the view that the cleavage of C3b to C3c and C3dg could start in sites other than Arg-Glu (human C3). They have identified three C3dg-like fragments 15 with their N-terminus starting at residues 933 (cleavage between Arg-Glu), 939 (cleavage between Lys-Glu) and 919, 924 or 930 (cleavage between Lys-Thr, Arg-Thr or Arg-Leu). The nature of the enzyme mediating some of these cleavages is still a debatable issue.

**PHYLOGENETIC TREE**

The elucidation of the amino acid sequences of several C3- and C3-related proteins (eight C3 proteins, two C4, two C5 and six αM-related proteins) allowed the construction of a phylogenetic tree for these proteins. These proteins were aligned and a phylogenetic tree was constructed using the program PHYLIP. The relative high pairwise percentage identities between sequence pairs of C3 and other C3-related proteins, listed in Table 2.2, are consistent with the hypothesis that all these proteins have diverged from a common ancestor. There is 77%-80% sequence identity among mammalian C3, 50%-53%, 52%-54%, 43%-45% and 51%-54% of mammalian C3 as a group to that of cobra,
chicken, trout and Xenopus C3, respectively. Substantially lower identities (28-33%) of mammalian C3 were seen with lamprey and hagfish C3 which is similar to that observed for C4 (27%-30%) and C5 (26%-29%). The phylogenetic tree shown in Figure 2.5, confirms that C3, C4, C5, and α2M proteins segregate in distinct clades. This is consistent with the duplication event which led to the generation of a complement protein which is the ancestor of C3, C4 and C5, possibly from an ancestor α2M-like protein.

Concerning lampreys and hagfishes, there is no consensus whether they belong to the same phyla or not. Based on morphological analysis and fossil data, lampreys are considered to be more closely related to the jawed vertebrates (gnathostomes), than to hagfishes, suggesting a paraphyletic cyclostomata. Recent molecular analysis of the small subunit (18S) rRNA sequences from hagfishes, lampreys, chondrichthyan fish, tunicates and cephalochordates supports the monophyly of the cyclostomes. Our data are in agreement with this analysis and support the monophyly of cyclostomes.

ACKNOWLEDGMENTS

This work was supported by grants from NSF (DCB90-18751) and NIH AI 30040. Portions of this material appear in Complement Profiles, Volume 1, 1993.

Fig. 2.5. Phylogenetic relationships of C3 and other homologous protein. Sequences were aligned using the program PILEUP and gaps were excluded from phylogenetic analysis. The tree was constructed using the PHYLIP software. The protein sequences for C3, C4, C5, Pregnancy zone protein (PZP), murniglobulin 1 (MUG-1), alpha 1 inhibitor III (A113) and α2M were taken from Swiss Protein and GenBank data bases. The chicken and Xenopus C3 sequences will be published elsewhere. For abbreviations on species see Table 2.1.
REFERENCES


26. Sundstrom SA, Komm BS, Ponce-de-Leon H, et al. Estrogen regulation of tissue-spe-
48. Lundwall A, Wetsel RA, Domdey H, et al. Structure of murine complement component C3 1. Nucleotide sequence of cloned comple-
mentary and genomic DNA coding for the
β chain. J Biol Chem 1984; 259:13851-
13862.
Molecular basis of complement C3 defi-
86:96-106.
50. Misumi Y, Sohda M, Ikehara Y. Nucleotide
and deduced amino acid sequence of rat
18:2178.
Nucleotide sequence of cDNA and derived
amino acid sequence of rabbit complement
component C3 alpha-chain. Immunol In-
52. Thiel S, Bastrup G, Friis-Christiansen P, et
al. Characterization of a lectin in human
plasma analogous to bovine conglutinin.
53. Laursen I, Koch C. Purification of Chicken
C3 and structural and functional character-
ization. Scand J Immunol 1989; 30:
529-538.
Isolation and identification of the third
component of complement of Japanese
55. Mavroidis M, Lambris DJ. Isolation, pri-
mary structure and evolution of chicken
third complement component. J Immunol
1994; (submitted)
56. Kjalke M, Welinder KG, Koch C. Struc-
tural analysis of chicken factor-B-like pro-
tease and comparison with mammalian
complement proteins factor-B and C2. J
57. Koppenheffer TL. Calcium-dependent
complement activity in chicken serum. Dev
Comp Immunol 1991; 15(S1):S104. (ab-
stract)
58. Fritzinger DC, Petrella EC, Connelly MB,
et al. Primary structure of cobra comple-
149:3554-3562.
59. Channe Gowda D, Petrella EC, Tilak Raj
T, et al. Immunoreactivity and function of
oligosaccharides in cobra venom factor. J
60. Cochrane CG, Müller-Eberhard HJ, Aikin
BS. Depletion of plasma complement in vivo
by a protein of cobra venom: its effect on
various immunologic reactions. J Immunol
61. Fritzinger DC, Bredehorst R, Vogel CW.
Complete structure of cobra venom factor
cDNAs. FASEB J 1992; (abstract)
62. Sekizawa A, Fujii T, Katagiri C. Isolation
and characterization of the third component
of complement in the serum of the clawed
133:1436-1443.
63. Avila D, Lambris JD. Isolation and charac-
terization of the third complement com-
ponent of axolotl (Ambystoma mexicanum).
Comp Biochem Physiol 1990; 95B:839-845.
64. Kunnathuglia LM, Chang GH, Sim RB,
et al. Characterization of Xenopus-laevi
complement factor-I structure-conservation
of modular structure except for an unusual
insert not present in human factor-I. Mol
65. Fujii T, Sekizawa A, Katagiri C. Character-
ization of the fourth component of comple-
ment in the serum of the clawed frog Xenopus
laevis. Immunology 1985; 56:743-750.
66. Alexander RJ, Steiner LA. The first com-
ponent of complement from the bullfrog, Rana
catesbeiana: functional properties of C1 and
isolation of subcomponent C1q. J Immunol
67. Sekizawa A, Fujii T, Tochinai S. Membrane
receptors on Xenopus macrophages for two
classes of immunoglobulins (IgM and IgY)
and the third component complement (C3).
68. Grossberger D, Marcuz A, Du Pasquier L,
et al. Conservation of structural and func-
tional domains in complement component
C3 of Xenopus and mammals. Proc Natl Acad
69. Lambris JD, Mavroidis M, Lao Z, Pappas J,
Wang Y. Suner O. Cloning of trout,
chicken, and Xenopus gilli C3: conservation
of functional sites in the C3 from different
species. Mol Immunol 1993; 30:27. (ab-
stract)
70. Minkoff MS, Wong WW, Silberstein DS.
Identification of C3 beta-chain as the hu-
man serum eosinophil cytotoxicity inhibi-
71. Sakai DK. Repertoire of complement in

72. Sunyer JO, Tort LL. Natural haemolytic and bactericidal activities of sea bream (Sparus aurata) are effected by the alternative complement pathway. Vet Immunol Immunopathol 1994; (in press)


76. Nakao M, Yano T, Matsuyama H. Partial purification and characterization of the fourth component (C4) of carp complement. Fish Pathol 1988; 23:243-250.


