Heavy Chain V Region Diversity in the Duck-Billed Platypus (Ornithorhynchus anatinus): Long and Highly Variable Complementarity-Determining Region 3 Compensates for Limited Germline Diversity

Jeannette Johansson, Maria Aveskogh, Barry Munday and Lars Hellman

*J Immunol* 2002; 168:5155-5162; doi: 10.4049/jimmunol.168.10.5155

http://www.jimmunol.org/content/168/10/5155

References

This article cites 50 articles, 20 of which you can access for free at: http://www.jimmunol.org/content/168/10/5155.full#ref-list-1

Subscription

Information about subscribing to *The Journal of Immunology* is online at: http://jimmunol.org/subscription

Permissions

Submit copyright permission requests at: http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts

Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/alerts
Heavy Chain V Region Diversity in the Duck-Billed Platypus 
(Ornithorhynchus anatinus): Long and Highly Variable 
Complementarity-Determining Region 3 Compensates for 
Limited Germline Diversity

Jeannette Johansson,* Maria Aveskogh,* Barry Munday,† and Lars Hellman2*

In this work, to study the emergence of the H chain V region repertoire during mammalian evolution, we present an analysis of 25 independent H chain V regions from a monotreme, the Australian duck-billed platypus, Ornithorhynchus anatinus. All the sequences analyzed were found to form a single branch within the clan III of mammalian V region sequences in a distance tree. However, compared with a classical V gene family this branch was more diversified in sequence. Sequence analysis indicates that the apparent lack of diversity in germline V segments is well compensated for by relatively long and highly diversified D and N nucleotides. In addition, extensive sequence variation was observed in the framework region 3. Furthermore, at least five and possibly seven different J segments seem to be actively used in recombination. Interestingly, internal cysteine bridges in the complementarity-determining region (CDR)3 loop, or between the CDR2 and CDR3 loops, are found in ~36% of the platypus VH sequences. Such cysteine bridges have also been observed in cow, camel, and shark. Internal cysteine bridges may play a role in stabilizing long and diversified CDR3 and thereby have a role in increasing the affinity of the Ab-Ag interaction. The Journal of Immunology, 2002, 168: 5155–5162.

Antibody diversity is generated through a complex series of events in which the number of germline-gene segments and their sequence variation is one important contributing factor. Random combination of V, D, and J segments, the addition of N and P nucleotides during the recombination process, and the addition of single base pair mutations through somatic hypermutations are other important mechanisms.

A relatively large number of germline V gene segments have been found in many species. Based on the degree of sequence identity, the germline VH segments have been shown to form different VH gene families. Seven different VH families have been identified in humans (1) and 15 have been identified in mice (2). The criterion for the presence of a separate family has been defined as a sequence identity of >75% between the most distantly related members of the family (3). The various VH gene families in mammals have been shown to form three major clans (clans I–III) (4). The three mammalian clans have probably coexisted for >400 million years of vertebrate evolution, which indicates that the clans formed before the divergence of early ancestors of mammals and reptiles (5), an evolutionary separation, which probably occurred at least 310 million years ago (6). Fish VH sequences have their closest homologs in mammalian clan III sequences. However, fish also have two additional clans, the teleost and the archaic clan, which seem to lack counterparts in mammals (7).

Based on the information from mouse and human, the presence of a highly diversified germline VH repertoire was long thought to be the general rule for the generation of Ab diversity. However, following the analysis of V gene diversity in chicken (8, 9), rabbit (10), and pig (11), a relatively limited germline variability was found to be common as these species were shown to express only one V gene family. Furthermore, only one functional V gene is used in the rabbit, whereas chicken express only one functional V gene.

The V gene or the V gene family of chicken, rabbit, and pig were all found to belong to clan III. This finding, together with the observation that fish VH sequences have their closest homologs in mammalian clan III sequences, led to the assumption that this is the most ancient of the three mammalian clans (5, 12–15). However, more recent data have shown that both cow (Bos taurus) and sheep express only one VH gene family, which belongs to clan II (16–19). The exclusive expression of clan II sequences could be explained by an inactivation or loss of clan III members in these species. However, the question of which clan appeared first during evolution probably needs further attention. In addition, the frequency of inactivations of various clan members and the subsequent expansion of remaining members in the evolution of the H chain and their importance for the evolution of the H chain repertoire in vertebrates is still an open question.

Until recently most information available on mammalian H chain V region diversity has come from various placental mammals. To obtain a more detailed picture of mammalian V region evolution, we and other groups have recently turned our attention to nonplacental mammals. This has resulted in the characterization of the VH gene repertoire of the American short-tailed opossum (Monodelphis domestica) (20–22). The opossum was shown to express a relatively limited germline V gene repertoire, consisting of only two V gene families, both belonging to clan III. However,
this limitation in variability is to some extent compensated for by a large variation in complementarity-determining region (CDR)\textsuperscript{3} and also by the variability in L chain V regions (23, 24). To increase our understanding of the processes that have shaped the H chain V gene repertoire during mammalian evolution we now turn our attention to the remaining mammalian lineage, the monotremes. Only three living species of monotremes exist today: the duck-billed platypus and two species of echidnas. The monotremes are egg-laying mammals and were therefore regarded as reptile-like or primitive mammals; however, they possess almost all major mammalian features, including a well-developed fur coat, a single bone in the lower jaw, three bones in the middle ear, and mammary glands. Histological studies show that the spleen-, thymus-, and gut-associated lymphoid tissues in the platypus are well developed and comparable in histological structure to those of therian mammals (25). However, in sites where lymph nodes would be expected in marsupials and placental mammals, monotremes were found to have lymphoid nodules that resemble the jugular bodies of the amphibians, which indicate that monotremes have a somewhat more primitive immune system (25).

A biochemical analysis of various Igs in a monotreme was presented in 1973 by Atwell et al. (26), who isolated Igs of two different m.w. from the short-beaked echidna (Tachyglossus aculeatus). The high-m.w. protein resembled human IgM and the low-m.w. Ig resembled human IgG in electrophoretic mobility. Both the high- and low-m.w. Igs consisted of what appear to be equimolar amounts of L and H chains. An N-terminal amino acid sequence of the H chain of the low-m.w. Ig resulted in a sequence that resembles group III of mammalian V gene sequences (27). However, except for this N-terminal sequence, no more detailed molecular analysis of a monotreme Ig protein has been published.

Monmarsupials are thought to have separated from the eutherian (placental) mammals 130 million years ago (28, 29), while the major radiation of the placental mammals probably occurred 70–130 million years ago (28,29), while the appearance of modern postswitch immunoglobulin isotypes in mammalian evolution (H1) shows that the amino acid at that position is highly conserved. When the value is between 1 and 2 (1 < H ≤ 2), the position contains mostly amino acids with similar properties, which correspond to conservative changes. When the H value is >2, the position comprises one of a number of residues. If a position can be occupied by a very large number of alternative amino acids, the H value may approach the theoretical maximum of 4.32 (38). The amino acid sequences used in the Shannon program were first aligned in the CLUSTAL W program. Based on visual inspection, some minor modifications were made.

**Results**

Cloning and sequence analysis

A portion of a spleen from a free-living Tasmanian platypus was obtained by partial splenectomy. Following the operation, the platypus was released back to its normal habitat and was subsequently recaptured in a healthy state. The spleen was used as an mRNA source to construct a total platypus spleen cDNA library. Platypus $\epsilon$- and $\gamma$-chain clones, initially isolated by PCR using degenerate PCR primers, were used to isolate full-length cDNA clones for platypus IgG1 and IgE. Subsequent screening of the cDNA library led to the isolation of a second IgG isotype (IgG2) and two IgA isotypes, IgA1 and IgA2. However, no $\mu$-chain clones have been identified in the platypus library so far. The overall structure of the isolated isotypes resembles the structures of their eutherian counterparts. The complete sequences of platypus IgG1, IgG2, IgE, IgA1, and IgA2 will be presented in two coming communications (Ref. 5 and M. Vernersson, M. Aveskogh, B. Munday, and L. Hellman, manuscripts in preparation).

To avoid bias for any particular V gene family a cDNA fragment from the constant region of the most prevalent IgG isotype in the platypus, the IgG1, was used as probe to screen ~50,000 clones from the nonamplified platypus spleen cDNA library. This resulted in ~350 positive signals. Sixty-six positive clones were chosen for further analysis. Nineteen of the signals were found to represent full-length IgG1 clones. Eleven additional IgG1 clones were studied. Nine contained (in addition to the constant region) the CDR2 and CDR3 regions.

**Materials and Methods**

Construction of a platypus spleen cDNA library

Total RNA was isolated from a platypus spleen by the guanidine thiocyanate method as described previously (35), and poly(A)\textsuperscript{3} RNA was purified using the Poly(A)Tribluel system (Promega, Madison, WI). Subsequently, double-stranded cDNA was synthesized using the TimeSaver cDNA Synthesis kit (Amersham Pharmacia Biotech, Uppsala, Sweden).

Abbreviations used in this paper: CDR, complementarity-determining region; FR, framework region.

K. Belov, L. Hellman, and D. W. Cooper. Characterization of echidna IgM provides insights into the time of divergence of extant mammals. Submitted for publication.
only the D-J segments, and two contained only the J segment. The nucleotide sequences of these 30 clones were determined (Fig. 1). The remainders of the 66 originally positive clones were partial clones that contained only constant region sequences. The entire analysis is based on cDNA clones, and no PCR amplification of the sequenced clones has been performed, which reduces the risk of introducing sequence errors.

In addition, six variable regions derived from clones encoding the H chains of IgA1, IgA2, IgG2, or IgE were included in the sequence comparison shown in Fig. 1. Hence, the majority of the V region sequences presented in this communication originate from the \( \gamma_1 \) chain clones and all sequences originate from postswitch isotypes.

**Sequence alignment**

Sequence alignment using the CLUSTAL W program indicated that all the platypus V segments belong to the same V gene family cysteine residue; □, sequences with two additional cysteine residues within the CDR3; ■, sequences with one cysteine residue in CDR2 and one in CDR3.
They all form a separate branch on the distance tree and they all belong to clan III of mammalian V gene sequences (Fig. 2). However, at the nucleotide level, some of the clones share only 65% sequence identity and not the 75% criterion that defines a V gene family. This latter finding indicates that the platypus V_{H} sequences belong to several closely related V gene families. In contrast, no sequences originating from other species were found within the platypus branch in the phylogenetic tree (Fig. 2).

The finding that all the 25 independent V regions belong to clan III indicates that only clan III members are expressed in the platypus. However, the possibility that V regions belonging to other clans are present in the platypus genome, but have escaped detection due to a low level of expression, cannot be excluded.

As expected, most of the sequence variation in the V region was found in the CDRs (Fig. 1). However, extensive variation was also observed in FR3 (Fig. 1).

The D, N, and P nucleotides, which together make up almost the entire CDR3, range from 25 to 58 bp in length, corresponding to 9–19 amino acid residues (Fig. 3). Therefore, the platypus CDR3s are relatively long, although not exceptionally long, compared with those of many other species (40). They are also very diverse in sequence, which indicates that many D segments are actively used during recombination (Fig. 3A). In addition, N nucleotides seem to contribute significantly to the diversity. However, not many palindromic sequences were found, indicating a relatively small contribution by P nucleotides (Fig. 3A).

A pairwise comparison of all the D segments shown in Fig. 3A indicates that a few D segments have been used more than once. Clones P_{Y_{1}}/4 and P_{Y_{1}}/3 display identity over 12 nt (TGGTAAC TATGG) and clones P_{Y_{1}}/57 and P_{Y_{1}}/71 display only one difference over a 15-nt long region (TAc/tATGTATGGTAGT). Both of these sequences are found in the middle of the CDR3. Furthermore, clones P_{O_{1}}/15 and P_{O_{1}}/3 display one difference over a region of 12 nt (CGGTAc/gTAGTAA), clones P_{O_{1}}/15 and Pe3 display two differences over 16 nt (GTCc/gTCCTACTATTa/gC), and P_{Y_{1}}/2 and Pe3 display one difference over 20 nt (GATTATAGTACTTGt/cAGTAG).

At the 3' end of the CDR3, clones P_{O_{1}}/11 and P_{Y_{1}}/32 share 13 nt (ACTGTGCTTTCGA), clones P_{O_{1}}/11 and P_{Y_{1}}/65 differ at one position over a stretch of 14 nt (GACt/cGTGCTTTCGA), and clones P_{Y_{1}}/35 and P_{O_{2}}/4 are identical over a 10-nt long (GATATGGAGT) region. Interestingly, two of the clones, Pe3 and P_{O_{1}}/15, share highly homologous regions with two other, nonhomologous clones (Fig. 3A). Although highly speculative, this may indicate the use of two consecutive D segments similar to what has been observed in TCR \( \beta \) - and \( \delta \)-chains. The sequences reported in this work have been deposited in the GenBank database and assigned accession numbers AF381289–AF381324.

An analysis of 36 different J region sequences led to the identification of at least five distinct families of J segments (Fig. 4). These families probably represent five individual germline J segments. Based on additional minor differences in sequence, we found indications for as many as seven different germline J segments (Fig. 4). However, it is equally likely that these additional
minor differences originate from somatic hypermutations or allelic variants.

Eleven of the 25 platypus H chain V regions were found to contain additional cysteine residues in their CDR2 or CDR3 (Fig. 5). Two of the clones have one additional cysteine residue, whereas in nine clones there are two additional cysteine residues. In six of the nine clones both cysteines are located in the CDR3, indicating an internal cysteine bridge in the CDR3. The remaining three clones have one cysteine in the CDR2 and one in the CDR3, suggesting a cysteine bridge formation between the CDR2 and CDR3. The role of a single additional cysteine residue is unknown; however, we favor the explanation that these clones are nonfunctional byproducts of the variability-generating mechanism, as free unpaired cysteine residues often lead to uncontrolled complex formations by binding to free cysteines in other molecules (41).

**FIGURE 5.** Additional cysteine residues. Shown are the platypus sequences that contain extra cysteine residues. Black dots indicate extra cysteine residues. Putative cysteine bridges are indicated. , sequences with one extra cysteine residue; , sequences with two additional cysteine residues within the CDR3; , sequences with one cysteine residue in CDR2 and one in CDR3.

**FIGURE 4.** J gene segment alignment: the nucleotide sequence of 36 different J gene segments. They have been grouped into possible subfamilies named J(I) to J(V). Dots indicate identity.

Shannon entropy analysis

To obtain a more accurate estimate of the sequence variation within the V region of platypus Ig H chains and to compare them with the corresponding variation of other species, a Shannon entropy analysis was performed. The Shannon plot was based on the 25 complete platypus V region sequences. This analysis identifies three regions with high variability ($H > 2$) that corresponds to the three CDRs. In addition, extensive variability was observed in FR3 (Fig. 6A). Corresponding analyses were performed on 21 $V_H$ sequences from cow, a species with limited germline variability (Fig. 6B), 48 sequences from the human V$_i$5 family (Fig. 6C), 42 sequences representing seven human V$_i$ families (six from each) (Fig. 6D), and 134 mouse V$_i$ sequences (distributed relatively evenly among the 15 families) (Fig. 6E). When comparing the plots generated from the platypus and cow sequences they were found to be highly similar (Fig. 6, A and B). There is a higher degree of variability at some positions in the CDR3 of the platypus, but the cow may compensate for that by having longer CDR3. The platypus has higher entropy values in the CDR1, CDR2, and FR3 compared with the plot of one human V$_i$ family (Fig. 6, A and C). Nevertheless, the human V$_i$ repertoire is still more variable than the platypus due to the seven families (Fig. 6, A and D). Comparison between the total human and mouse repertoires reveals that the variability between these two species is similar despite the fact that the mouse has twice as many V$_i$ gene families (Fig. 6, D and E).

Discussion

Previous investigations of mammalian Ig evolution have focused on placental mammals and marsupials. In the present report, we demonstrate for the first time an analysis of the H chain V region repertoire in a monotreme, the duck-billed platypus. Our results indicate that monotremes, like chicken, rabbit, pig, cow, sheep, and the American opossum have a relatively limited germline V$_H$ repertoire (8–11, 16–20, 22). The 25 platypus V region sequences were found to form a separate branch on the distance tree, indicating that they all belong to a single V gene family. However, they are more diverse in sequence than a classical V gene family. A possible explanation is that this gene family has expanded extensively in the platypus over a long period of time and has thereby diversified more than V gene families in most other species. This}

The Journal of Immunology

5159

Downloaded from http://www.jimmunol.org/ by guest on April 13, 2017

A scenario is supported by the fact that no V_H sequences from other species appeared within the platypus branch. All of these platypus V region sequences have probably evolved from a single common ancestor or a small family of closely related genes after the separation of monoterms from other mammalian lineages. This together with information from several other species, indicates that deletions of entire V_H clans, or parts of clans, followed by successive rounds of gene duplications may be a relatively common phenomenon during vertebrate evolution.

Although the germline diversity is somewhat limited, the total H chain repertoire does not seem to be limited. We found clear indications for the active use of at least five different J segments. Furthermore, fairly long and highly diversified CDR3 seem to adequately compensate for the limited germline diversity. Relatively few regions of homology between the different clones were found in the CDR3, indicating the active use of many D segments or, alternatively, that the platypus uses a very potent TdT-dependent mechanism to generate N nucleotide diversity. The presence of multiple V_H families entails a lesser need for extensive variation in the CDR1 and CDR2 within the individual gene families. In cow, another species with only one V gene family, a similar degree of variability in the CDR1, CDR2, and CDR3 has been observed. Compensatory modifications to increase the V_H repertoire may arise after massive deletions of entire clans or parts of clans. During evolution, an increase in the germline variation in CDR1 and CDR2 is probably favored. In addition, an increase in CDR3 length and diversity may be even more important for generating variability to compensate for a limited germline V_H repertoire (42). However, problems may arise when the variability and length of CDR3 increases. An increase in the size of the CDR3 probably decreases the stability of this region of the Ab. Usually, no structural motifs like β-pleated sheets or α helices are found in CDR3. The lack of structural motifs may reduce the structural stability of the Ag-binding site due to the generation of a large number of mutually exclusive, and thereby competing, interacting structures. This, in turn, may reduce the affinity of the Ab-Ag interaction. A reduction in the number of potential conformations by the introduction of additional cysteine bridges may be a compensatory mechanism of selective advantage in this case. In the platypus, we found evidence for the existence of additional cysteine bridges in 36% of the V_H sequences. Such bridges seem to occur between CDR2 and CDR3 or as internal cysteine bridges within the CDR3. The internal CDR3 cysteine bridges in platypus are twice as abundant as those between CDR2 and CDR3. The presence of additional cysteine residues has also been observed in cow (16), shark (43), and members of Camelidae (44–46). V_H sequences in cow sometimes have exceptionally long CDR3, with reports of sequences of up to 61 amino acid residues (16). These long CDR3 may contain six or eight cysteine residues, indicating the presence of three or four cysteine bridges. However, the average size of CDR3 in cow is considerably shorter, ~21 amino acids (47, 48). Almost 60% of all
rearranged VH regions in cow seem to have an additional cysteine bridge and, in contrast to the pl atypus, the majority of these are between the CDR2 and CDR3.

Additional cysteine residues and thereby potential cysteine bridges have also been observed in the camel (44). In camels, Iggs both with and without L chains are found. In Iggs that lack L chains, the CH1 domain of the H chain has been deleted (49). The CDR3 in the H chain Abs are long; regions of up to 24 amino acids in length have been observed (50). Additional cysteine residues are present in 75% of these VH regions; however, only ~60% have an even number of additional cysteines indicating functional cysteine bridges. In contrast to platypus and cow, the cysteine bridges in camel are found almost exclusively between CDR1 and CDR3 or between FR2 and CDR3. The role of these cysteine residues in stabilizing the structure, and thereby potentially increasing the affinity of the Ab, is interesting and needs to be studied in more detail. Future analysis of the Ag-Ab interactions of monoclonal Iggs with long and diversified CDR3 and internal cysteine bridges may shed light on this matter. In camels, llamas, and sharks (43), these additional cysteine residues are found primarily in H chain Abs. However, for several reasons H chain Abs are not likely to be present in platypus: 1) the first constant domain, CH1, which is required for the interaction with the L chain, is present in all five platypus H chains isolated (data not shown); 2) to enhance solubility of the nonpaired variable regions, the amino acids that interact with the variable domain of L chains in classical Abs are substituted in H chain Abs of camels; these residues are not substituted in the platypus V regions; 3) L chains of A type have been isolated and cloned from the platypus, and the L chain shows a relatively high degree of diversity and complexity (J. Johansson, J. Salazar, M. Aveskogh, B. Munday, R. Miller, and L. Hellman, manuscript in preparation); and 4) sera from the echidna have been shown to contain at least two classical isotypes in relatively high concentrations, which both display the classical structure with apparent equimolar amounts of L and H chains (26).

A more diverse and hopefully a more complete picture of the evolution of the VH repertoire in vertebrates appears with every new species being investigated. One important conclusion from these studies is that there is an amazing variation in the way in which diversity is being generated. Somatic hypermutation or gene conversion seems to be an important crossroad in evolution. Another example is the large variation in the number and sequence complexity of the germine VH gene pool. A third and unexpected finding was the presence of functional Iggs lacking a L chain in camels. Following the analysis of VH genes in mouse and human it was easy to conclude that somatic hypermutation and a large germine VH gene pool was the general rule for the generation of Ab diversity. However, additional insight into the complexity of evolution has been, and will probably continue to be, obtained with every new species analyzed. The analysis of the V gene repertoire in the platypus gives additional evidence that a high variability can be created with relatively limited germline diversity. Furthermore, cysteine bridges within the CDR3 or between CDR2 and CDR3 may be a more common phenomenon than was earlier expected.

Acknowledgments
We thank Prof. Lars Pilstrom and Molly Vernesson for valuable discussions and for critically reading the manuscript.

References
16. Saini, S. S., B. Allore, R. M. Jacobs, and A. Kaushik. 1999. Exceptionally long bridges between the CDR2 and CDR3. The role of these cysteine residues in stabilizing the structure, and thereby potentially increasing the affinity of the Ab, is interesting and needs to be studied in more detail. Future analysis of the Ag-Ab interactions of monoclonal Iggs with long and diversified CDR3 and internal cysteine bridges may shed light on this matter.


