The Multiple Shark Ig H Chain Genes Rearrange and Hypermutate Autonomously

Catherine Zhu, Wendy Feng, Jeremy Weedon, Pauline Hua, Dimitre Stefanov, Yuko Ohta, Martin F. Flajnik and Ellen Hsu

*J Immunol* published online 29 July 2011
http://www.jimmunol.org/content/early/2011/07/29/jimmunol.1101671

---

**Supplementary Material**
http://www.jimmunol.org/content/suppl/2011/07/29/jimmunol.1101671.1.DC1

**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
The Multiple Shark Ig H Chain Genes Rearrange and Hypermutate Autonomously

Catherine Zhu,* Wendy Feng,* Jeremy Weedon,† Pauline Hua,* Dimitre Stefanov,† Yuko Ohta,‡ Martin F. Flajnik,‡ and Ellen Hsu*†

Sharks and skates are representatives of the earliest vertebrates with an immune system based on V(D)J rearrangement. They possess a unique Ig gene organization consisting of 15 to >50 individual IgM loci, each with one VH, two DH, one JH, and one set of constant region exons. The present study attempts to understand how multiple Ig genes are regulated with respect to rearrangement initiation and to targeting during somatic hypermutation. The linkage of three single-copy IgH genes was determined, and single-cell genomic PCR studies in a neonatal animal were used to examine any relationship between relative gene position and likelihood of rearrangement. Our results show that one to three IgH genes are activated independently of linkage or allelic position and the data best fit with a probability model based on the hypothesis that V(D)J rearrangement occurs as a sequence of trials within the B cell. In the neonatal cell set, two closely related IgH, G2A, and G2B, rearranged at similar frequencies, and their membrane forms were expressed at similar levels, like in other young animals. However, older animals displayed a bias in favor of the G2A isotype, which suggests that although rearrangement at G2A and G2B was randomly initiated during primary repertoire generation, the two very similar IgM sequences appear to be differentially expressed with age and exposure to Ag. We performed genomic single-cell PCR on B cells from an immunized individual to study activation-induced cytidine deaminase targeting and found that hypermutation, like V(D)J rearrangement, occurred independently among the many shark IgH. The Journal of Immunology, 2011, 187: 000–000.

In the mouse, V(D)J recombination at the Ig H chain locus proceeds in two distinct and ordered steps, D to JH followed by VH to DJ (1). This is in part controlled by enhancer-blocking insulator sequences in the 96-kb V–D interval that prevent spreading of open chromatin and simultaneous activation of all elements (2). The differential chromatin activation restricts recombination to DJ formation on one or both alleles before the second step, during which >200 upstream VH occupying >2 Mb organize into multiple loops that gather in close spatial proximity to the DJ, to enable DJ rearrangement to one VH (3). The two phenomena, sequential chromatin domain activation (4) and locus contraction (5, 6), are evolutionary adaptations for a highly complex and spatially extended IgH locus. All components, the VH, D, and JH gene segments and constant (C) region exons, are brought under one set of cis-acting regulators that coordinate gene activation and transcription and accessibility to the recombinase during early B cell differentiation (for a recent review, see Ref. 7). Both alleles are poised for rearrangement, but it is the asynchrony at the VH to DJ step that initiates allelic exclusion and events leading to a single type of Ag receptor expressed per lymphocyte. The underlying causes for one gene acting ahead of the other are the subject of some controversy (8–10).

Ig and V(D)J rearrangement exist in all jawed vertebrates; however, only in bony and cartilaginous fishes are the Ig genes present as multiple loci, or clusters (for reviews, see Refs. 11, 12). The existing cartilaginous fish subclasses, Elasmobranchii (sharks, rays, and skates) and Holocephali (chimeras, ratfish), diverged >350 million years ago, and the IgH of all representative animals examined are arrayed as multiple miniloci. There can be 100–200 individual IgH and IgL genes, each with a few gene segments (V-D1-D2-JH and VL-JL, respectively) and one set of C region exons. This unit is considered ancestral to the classical Ig locus in other vertebrates. In this H chain gene system, the close proximity of the rearranging gene segments (within 2 kb) do not require locus compaction, and they all are simultaneously available to RAG for the three recombination events generating the shark VDJ (13). From the first report (14) of this alternative organization in the earliest of jawed vertebrates, the regulation of gene expression and allelic exclusion have been central questions of interest in an Ig system that is both ancient and conserved.

Our model is the nurse shark, in which we have isolated and classified all of its IgM H chain genes, each of which must rearrange to be expressed (15). There are 9–12 functional IgH genes for which its VH are all members of one family (16), in which the gene number in one subfamily (G4) differs between individuals (15). Thus, in any single differentiating shark B cell, there are multiple potential targets for RAG recombinase as there are for activation-induced cytidine deaminase (AID) after the B cell has

*Department of Physiology and Pharmacology, State University of New York-Downstate, 450 Clarkson Avenue, Box 31, Brooklyn, NY 11203. E-mail address: hsue@hscbklyn.edu

†Scientific Computing Center, State University of New York Health Science Center at Brooklyn, Brooklyn, NY 11203; ‡Department of Microbiology and Immunology, University of Maryland, Baltimore, MD 21201

Received for publication June 7, 2011. Accepted for publication June 30, 2011.

This work was supported by grants from the National Institutes of Health (GM068095 to E.H. and RR006603 to M.F.F.).

The sequences presented in this article have been submitted to GenBank (http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide) under accession numbers JN087503–JN087510.

Address correspondence and reprint requests to Dr. Ellen Hou, Department of Physiology and Pharmacology, State University of New York-Downstate, 450 Clarkson Avenue, Box 31, Brooklyn, NY 11203. E-mail address: hsue@hscbklyn.edu

The online version of this article contains supplemental material.

Abbreviations used in this article: AID, activation-induced cytidine deaminase; BAC, bacterial artificial chromosome; C, constant; CL, conditional likelihood; CMLE, conditional maximum likelihood estimate; GL, germline; PBL, peripheral blood leukocytes; 1R, one recombination event; 2R, two recombination events; 3R, three recombination events; RSS, recombination signal sequences; SHM, somatic hypermutation; sIg, surface Ig-positive; sIgM, surface IgM-positive.

Copyright © 2011 by The American Association of Immunologists, Inc. 0022-1767/11/$16.00
been activated by Ag. AID is found in all vertebrates (17), and mutated shark Ig sequences are scarce in young animals (pups) but abundant in hyperimmunized adults (16, 18, 19).

The present study attempts to understand the regulation and activity of multiple copies of Ig genes within a cell. Pioneering work by Bengtén and coworkers (20) in the bony fish Atlantic cod demonstrated that the many L chain clusters did not possess individual enhancers. Because one set of regulators probably exerts influence over several gene clusters separated by distances of 2.1–4.8 kb, this casts some perplexity as to how cod L chain isotype exclusion is effected but raises interesting prospects for receptor editing (21). We had previously demonstrated H chain exclusion in shark (13); now our first task was to establish whether the shark IgH genes share regulatory elements and so discover if recombination accessibility unfurled on the basis of linkage or allelism. Information on the spatial relationship of the IgM H chain genes is limited in nurse shark and unknown in other species. We linked and mapped three closely related IgH that are single-copy genes. This was followed by single-cell genomic PCR studies in a very young animal to examine the pattern of rearranged IgH relative to their gene position. The paucity of somatic mutations in neonates also affords the best opportunity for scoring VD1, and, based on the results, we constructed probability models to assess the nature of IgH activation in shark precursor B lymphocytes.

We went on to study AID targeting patterns among IgH in single B cells of an older, immunized shark and to ascertain whether genetic exchanges took place between the genes, rearranged or not. In the course of investigating whether an IgH gene in germline (GL) configuration was in fact quiescent in B cells, we found that targeting by somatic hypermutation (SHM) was largely conditional to the rearrangement status of the gene.

Materials and Methods

Animals

Nurse sharks (Ginglymostoma cirratum) were captured off the coast of the Florida Keys. Blood was obtained from the caudal sinus. Peripheral blood leukocytes (PBL) were separated from erythrocytes (RBC) after centrifugation through Ficoll. Genomic DNA and total RNA were isolated using conventional techniques. Genomic DNA obtained from shark-GR PBL, after sorting with anti-IgM mAbs (CB5, CB11, CB16), was described previously in which an aliquot was used for genomic Southern blotting (13). Some animals were sacrificed after capture (shark-33), and others had been immunized with various Ags (shark-GR, shark-JS, shark-PI) (13, 15). At the time of sample collection, shark-GR was 7 y of age, shark-JS 5 to 6 y, shark-PI 3 to 4 y, and shark-33 2.5 y. Shark pups included shark-LA (2.1 y) and sharks-EC and -TH 1.2 y. Some animals were sacrificed after capture (shark-33), and others had been immunized with various Ags (shark-GR, shark-JS, shark-PI) (13, 15).

Single-cell studies were performed with B cells from sharks-GR and -LA, previously described (15). First-round PCR was performed with universal primers in the leader intron (primers GR1, GR2–5) and in JH (primer JH5) (unless specified, oligonucleotide sequences are in Ref. 13). These PCR products were treated with ExoSAP-IT (USB) to remove remaining primers and used separately in four different second-round PCR reactions, as depicted. The first nested reaction (Fig. 1, line 1) was primarily used to see whether any VDJ (result of three rearrangement events, 3R) could be amplified. The universal primers (VG1, VG2–5, and JH6) generate only nonrearranged GL sequences of 1.6 kb bp in the RBC, but in lymphocytes, both the GL as well as rearranged VDI of 350 bp. This reaction allowed us to select positive lymphocyte samples suitable for further analysis; that is, those VDJ-containing samples that were flanked by GL-only signals in the adjacent RBC reactions. This criterion was used to avoid carryover contamination between lymphocytes during picking. For isolation and identification of the VDJ, another set of PCRs (Fig. 1, line 3) was performed on the selected lymphocytes: five reactions each using a forward primer specific for the VH subfamily (Fam1 for G1, Int for G2, GR3N2 for G3, Fam4 for G4, or Fam5 for G5 each in combination with JH6). Examples of such reactions for shark-LA B cells are shown in Supplemental Fig. 1, top panel. The VDI of 400–420 bp was directly sequenced (Geneviz) and cloned into pGEM (Invitrogen). To determine which genes remained nonrecombined, those in GL configuration were amplified in another reaction (Fig. 1, line 4) in which the primers targeted the intersegmental sequence in V-D1 and in D2-JH. These primers were subfamily specific (G1DF/G1JR for G1, FD2-RD2 for G2, G3DF/G3JR for G3, G4DF/G4JR for G4, DR3/4 for G4, and G5DF/G5JR for G5). Examples of such reactions are shown in Supplemental Fig. 1, bottom panel. The identity of the 1.1- to 1.2-kb PCR products was ascertained by restriction enzyme analyses (Supplemental Fig. 2).

To characterize the individual nonrearranged VH gene segment (Fig. 1, line 5), specific primers in the leader intron and D2-JH sequence were used. First-round samples from G1, G3, and G5 are amplified by using subfamily-specific primers in the leader intron and the D2-JH region. Because there are two members of G2 and four of G4, additional gene-specific primers were generated for these genes. These included reverse primers that distinguished G2A from G2B (G2A-JJ, 5′-CGTCACTAACATGTCATGTTCCAAAC-3′; G2B-JJ, 5′-TTAATGAGCCACGACTTTGCTGCG-3′) in combination with forward G2 primer Int.

The individual G4 state is defined by the G4R (G4, G4C, G4D, G4E), primers specific for sequence differences in V and the D-J region were used to amplify the GL genes for direct sequencing (G4A-5′, 5′-TTTTCATCAGTAACACTGG-3′; G4A-3′, 5′-GTCCTGTTCCTCCAGATTAT-3′; G4D-5′, 5′-CATCTTCAGTACAC-3′; G4D-3′, 5′-TTTCTGCTCCAGTTAGCA-3′; G4E-V, 5′-GAAACTAGCGGTTCACTTTTTTGC-3′; G4E-JR, 5′-CTTATTTAAGTGTTC-3′; G4G-5′, 5′-CATCTTCAGTAACATC-3′; and G4G-3′, 5′-TTTCTGCTGCACACTGAG-3′). The CD3R of the all VDJ isolated are provided in the figures. Mutant VDJ sequences were submitted to GenBank (http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide) as JN087503–JN087509.

RT-PCR, PCR, and probes

First-round cDNA was primed in total RNA using an oligonucleotide in the G2 membrane sequence (G2mem4, 5′-GAGGCGAGGGCTGTGTTAA-3′) according to the manufacturer’s instruction for use of the SuperScript III RNase H reverse transcriptase kit (Invitrogen). Subsequent PCRs were performed with primers specific for leader G2 (G2L, 5′-ACGAAATAGCAAGC-ATG-3′) and G2 CHI (G2CH1R, 5′-CGCATTGCGCTGTGTTAG-3′); the optimal annealing temperature was determined by a T-Gradient thermocycler (Biometra), and the elongation time was usually set at 30 s for <500 bp products and 1 min for 800–1200 bp products. After 30 cycles, the ~660 bp PCR product was digested with Bse II, a restriction enzyme site that cleaves it from G2A1 and G2A2 CHI.

RT-PCR was performed on shark-GR spleen RNA to detect pseudogene G8 transcripts. Primers targeting FR1 (V2-5: 5′-TGAYTCAACAGCAGGGCA-3′) and the G8 CH2 (G8CH2R, 5′-GAATGACTTCTTCATGAG-
amplified DNA fragments of ~955 bp from oligo(dT)-primed cDNA. Sequences were cloned into pGEM-T-Easy and their identity ascertained by the unique 9-bp gap in FR1 and the G C region (EU106187).

Long-template PCR (Roche) was performed according to the manufacturer’s instructions for a total of 35 cycles and optimized for the amount of input genomic DNA. Primers for G2B in FR1 (S1, 5′-AGAGGAGGT-TACCTTGGATC-3′) and in J-C intron (J-C3, 5′-ATGGTTAATCCGCA-TAC-3′) were used to raise rearranged products of ~6.2 kb (GL) to 5 kb (fully rearranged VDJ) from B cell genomic DNA. Amplification of G4 sequences from shark-LA RBC DNA was performed with G4-specific primers in leader (G4L, 5′-TCTCTGACTTTCTTATCCC-3′) and in the D2-JH interval [DR3/4 (13)], generating 1772-bp fragments that were cloned into pGEM (Promega). Bacterial colonies were picked with radioactive probe under routine stringency conditions. Probes to H

nylon filters (HyBond-N; GE Healthcare) and the membranes hybridized to 25–50 observable (i.e., excluded from the probability model). This time there are VDJ+ event has occurred. If the VDJ+ event occurs in one of the four trials, Prob(VDJ+) = 1/(1 + 2 + 3 + 4). For each (j) pseudoset, a goodness-of-fit statistic \( G_{k} \) was calculated by CMLE (Equation 3) and the fit of data to an optimal model was conducted in Supplemental Table I. Clone 075N10 (15) carried the C exons of gene G2A and the VH gene segments of G5 (Fig. 2), ~120 kb apart. Because G2B is very closely related to G2A, four BAC clones carrying the G2B C region alone (Supplemental Table I, 399C2, 583K20, 552J13, and 639A12) were analyzed for linkage to any downstream IgH. Their insert ends were sequenced. Each of the four contained one insert end in the G2B J-C intron, at the HindIII site 1476 bp downstream from JH; probes were generated for the opposite ends in two (Fig. 2, boxes E and D) for which a nonrepetitive sequence was available. Probes E and D were hybridized to the dot blots that carried DNA from all IgM BAC clones, and signals were observed only from some G2A and G2B clones. Because neither hybridized to the single BAC clone that carried all of the sequence downstream of the IgH genes operated autonomously, we went on to investigate whether the different IgM isotypes were of the same value to the animal. Two closely related IgH (G2A, G2B) that rearranged at similar frequency in the neonate were examined for expression. The membrane forms were found at similar levels in young animals, but a distinct bias for one (G2A) was observed in adult animals.

Mapping of three IgM genes
The shark-Y BAC library was screened with Cμ3-4 sequence, as previously described (15), and 216 candidate clones were analyzed. Hybridization of dot blots showed that 156 contained V and C components, and 21 carried only C; the first 78 clones were characterized and reported, whereas the remaining 99 were completed at a later time using the same techniques (V. Lee, K. Malecek, and E. Hsu, unpublished observations). The classification of all clones according to IgM subfamily (G1, G2A, G2B, G3, G4, and G5 and pseudogenes G2C, G7, and G8) is available in Supplemental Table I. In an initial approach, we conducted an exact \( \chi^{2} \) goodness-of-fit test (25) separately for each Y strain, using as expected probabilities the CP values under the optimal model. As a second analysis, we calculated a single global \( \chi^{2} \) goodness-of-fit statistic G, conditioned on the marginal distribution of Y, using as expected probabilities the CP values under the optimal model multiplied by the ratio (1 + 2 + 3 + 4). For each Ch strand, we determined the number of insert ends needed to calculate an exact p value for this test is extremely large, we conducted a Monte Carlo simulation in which a dataset constituting the 13C observable cell outcomes. It follows from this model that a maximum of 24 VDJ events can occur in any cell, with exactly one of them being VDJ+. Let \( X_{ij} \) and \( X_{ij}^{\prime} \) represent the number of VDJ+ and VDJ− events, respectively, in a cell having \( Y_{i} \) ascertainmentable genes. From the probability model, we calculated the joint conditional probabilities: \( CP = Prob(X_{ij}^{\prime} = x_{j}, X_{ij} = x_{j} | \theta, Y_{i}) \), for \( x_{j} = 0 \) or 1, \( x_{j} = 0 \ldots \theta - 1 \).

Estimation of \( \theta \) and fit of data to an optimal model was conducted in a manner similar to that used for the fixed-genes model; the Monte Carlo simulation of the probability model data took into account the fact that under this model, the observable outcomes are not all equally probable. We conducted exact \( \chi^{2} \) goodness-of-fit tests of our data to assumptions 4 and 5. Statistical software used were SAS (SAS Institute, Cary, NC) Release 9.2 and StatXact PROCs for SAS, version 9 (Cytel, Cambridge, MA).

Results
Experimental overview
The linkage of three IgM H chain genes was established, and experiments focused on whether there existed any regulation shared between the adjacent or allelic IgH genes. V(D)J recombination and SHM patterns among the functional IgH were studied in single B cells from two individuals, pup shark-LA and adult shark-GR, using the protocol described in Fig. 1. After discovering that pseudogenes also rearranged, we tested probability models that included all IgH to obtain a global picture of IgH rearrangement initiation in A Lymphocyte. Because our results suggested that the IgH genes operated autonomously, we went on to investigate whether the different IgM isotypes were of the same value to the animal. Two closely related IgH (G2A, G2B) that rearranged at similar frequency in the neonate were examined for expression. The membrane forms were found at similar levels in young animals, but a distinct bias for one (G2A) was observed in adult animals.

Statistical methods
We considered two models for which we asked, is the total number of VDJ per cell fixed or constrained, and if so, what value might it be fixed at or have as a maximum? Each probability model was based following assumptions: 1) in each cell, there are 13 genes of interest; 2) 4 of these 13 genes are nonfunctional and never evaluated for VDJ status; 3) in each cell, the number of genes for which VDJ status can be evaluated is a random "nonsense" statistic; 4) VDJ events are equiprobable across the 13 genes; 5) ascertainmentability of a VDJ event is equiprobable across the 9 functional genes; 6) the presence or absence of a VDJ event, when ascertainable, is established without error; and 7) partially rearranged genes such as one recombination event (1R) or two recombination events (2R) should be treated as nonproductive VDJ events.

For the fixed-genes model, there is the assumption (a) that the total number of VDJ events in each cell is fixed across cells and has value θ. For the ith shark cell (i = 1 ... n), let \( Y_{i} \) denote the number of genes evaluated and \( X_{ij} \) the number of observed VDJ events. There are 13Cp ways in which the \( \theta \) VDJ events can be arranged in the 13 genes. Because the sample includes an observation with data (\( X_{i} = 3, Y_{i} = 9 \)), \( \theta \) is constrained to be an integer within the interval (3, 7); \( Y_{i} \) is similarly constrained within the interval (1, 9). For each (\( \theta, Y_{i} \)), we calculated conditional probabilities under the model CP (\( \theta, Y_{i} \)) = Prob(\( X_{ij} = x_{j} | \theta, Y_{i} \), \( x_{j} = 0 \ldots \theta \)). From these probabilities, we calculated the conditional likelihood (CL) of the sample data as CL (\( \theta, Y_{i} \)) = \( \Pi_{x_{j}=1}^{\theta} \cdot Prob(X_{ij} = x_{j} | \theta, Y_{i}) \); we used these numbers to construct a conditional maximum likelihood estimate (CMLE) of \( \theta \).

We asked how well do the sample data fit an optimal (i.e., with \( \theta \) estimated by CMLE) \( \theta \) − VDJ model? In an initial approach, we conducted an exact \( \chi^{2} \) goodness-of-fit test (25) separately for each Y strain, using as expected probabilities the CP values under the optimal model. As a second analysis, we calculated a single global \( \chi^{2} \) goodness-of-fit statistic G, conditioned on the marginal distribution of Y, using as expected probabilities the CP values under the optimal model multiplied by the ratio (1 + 2 + 3 + 4). For each Ch strand, we determined the number of insert ends needed to calculate an exact p value for this test is extremely large, we conducted a Monte Carlo simulation in which a dataset constituting the 13Cp equiprobable outcomes under the probability model (i.e., not the sample dataset) was resampled with replacement k = 1 ... 20,000 times, each pseudoset having both the same n as the sample dataset and the same marginal distribution of Y. For the 4th pseudoset, a goodness-of-fit statistic G4 was calculated; a p value for G is the sample set was estimated by determining its ordinality relative to the distribution of the simulated G4 values.

For the sequential activation model, assumptions 1–7 above remain, but this time, VDJ events occur in each cell as a sequence of trials. For the jth trial, Prob(VDJ+) = 1/(\( \theta - j + 1 \)); the process stops as soon as the first VDJ+ event has occurred. If the VDJ+ event occurs in one of the four nonfunctional genes, then the cell is considered dead and therefore unobservable (i.e., \( \alpha, \omega \) are zero from the probability model). This time there are (9/13) \( \sum_{j=1}^{\theta} P \) observable cell outcomes. It follows from this model that a maximum of 9 VDJ events can occur in any cell, with exactly one of them being VDJ+. Let \( X_{ij} \) and \( X_{ij}^{\prime} \) represent the number of VDJ+ and VDJ− events, respectively, in a cell having \( Y_{i} \) ascertainmentable genes. From the probability model, we calculated the joint conditional probabilities: \( \mathcal{P} = \mathbb{P}(X_{ij}^{\prime} = x_{j}, X_{ij} = x_{j} | \theta, Y_{i}) \), for \( x_{j} = 0 \) or 1, \( x_{j} = 0 \ldots \theta - 1 \).

Estimation of \( \theta \) and fit of data to an optimal model was conducted in a manner similar to that used for the fixed-genes model; the Monte Carlo simulation of the probability model data took into account the fact that under this model, the observable outcomes are not all equally probable. We conducted exact \( \chi^{2} \) goodness-of-fit tests of our data to assumptions 4 and 5. Statistical software used were SAS (SAS Institute, Cary, NC) Release 9.2 and StatXact PROCs for SAS, version 9 (Cytel, Cambridge, MA).
The GL G4 genes in shark-LA RBC DNA were amplified by G4-specific primers and cloned into pGEM (see Materials and Methods). Restriction enzyme analyses of 156 bacterial colony PCR fragments, supported by sequencing, showed there were four different G4 genes. The novel G4D2 (accession number JN087510) closely resembles G4D, differing only by eight substitutions; however, one difference in CDR1 in the VH and another the loss of a HincII site in the V-D interval permitted G4D2 to be distinguished in the single-cell studies described below. Shark-LA carried G4A, G4CG, G4D, and G4D2; the other H chain genes (G1, G2A, G2B, G3, and G5) are single-copy genes, and the shark-LA sequences obtained from RBC matched those previously submitted to GenBank. The restriction enzyme analyses for non-rearranged shark-LA IgH is shown in Supplemental Fig. 2.

Single-cell genomic PCR in neonatal shark

Surface IgM-positive (sIgM+) splenocytes from nurse shark pup LA were isolated, and 54 single cells were picked by hand, alternating with 54 erythrocytes. In the course of the first nested reactions (Fig. 1, line 2), 28 lymphocytes with VDJ were selected after GL-only signals were found in the flanking RBC samples. In Fig. 3 (top panel), 16 B cells are shown in which at least 6 out of 9 IgH genes could be detected. The rearranged IgH were identified as VDJ (Fig. 1, line 3) and their CDR3 shown in Fig. 3 (bottom panel). In each cell, one to three rearrangements are present, and in one case (LA46), there are two in-frame VDJ with CDR3 of viable sizes. In the same cell, a partially rearranged gene (V-DDJ, called 2R) was found; the reason for its noncompletion is that D1 and its 5’ RSS were deleted. The paucity of junctional sequence (Fig. 3, bottom panel, 46G1) does not permit analysis as to how this was generated.

In the established single-copy genes G1, G2A, G2B, G3, and G5, the nonrearranged allele can be detected in GL configuration in 14 out of 17 instances in 12 B cells. For example, in LA16, the G2A VDJ was out-of-frame and presumed to be the first IgH to recombine; the second rearrangement event did not take place at its allele but at G4CG, which recombined to an in-frame VDJ. Such instances suggest that activation of IgH is allele independent. Moreover, because G2A is rearranged as often as and independently of the adjacent G2B, it appears that there is no connection

![Flow diagram for single-cell PCR in shark B cells](https://www.jimmunol.org/)

**FIGURE 1.** Flow diagram for single-cell PCR in shark B cells. The rearranging gene segments (VH, D1, D2, JH) are separated by ~400 bp; the specific distances are from the gene G2A; L is leader. The flanking RSS with 23-bp spacers are represented as open triangles and RSS with 12-bp spacers as filled triangles. Arrows indicate position and orientation of PCR primers. Line 1, First-round PCR. A 40-cycle PCR was performed on B cells and on RBC using universal primers detecting all functional IgH. Thereafter, the PCR products were depleted of oligonucleotides and used in the following four kinds of nested second-round reactions (25–30 cycles). Line 2, Second round nested for VDJ. Universal primers targeting FR1 and JH can amplify VDJ and/or GL fragments in B cells but only GL in the RBC samples. Line 3, Second round nested for VDJ. Specific forward primers reveal which subfamily (G1, G2, G3, G4, G5) the VDJ belongs to. VDJ is isolated and sequenced. Line 4, Second round nested for GL. Subfamily specific primers in the V-D1 and D2-J intersegmental regions amplify the nonrearranged genes. The five subfamily reactions are analyzed by restriction enzyme analyses detailed in Supplemental Fig. 2. Line 5, Second round nested for GL. The unrearranged VH is amplified for sequencing by using subfamily or gene-specific primers. Primer information is supplied in Materials and Methods.

![Mapping of IgH genes G2B, G2A, and G5](https://www.jimmunol.org/)

**FIGURE 2.** Mapping of IgH genes G2B, G2A, and G5. The orientation and order of the IgM H chain genes G2B, G2A, and G5 are shown at top, with the VH gene segments and C exons labeled at left. The distance between G2B leader and its transmembrane exons is 17.9 kb; dotted line indicates map is not drawn to scale. BAC clones, drawn as thick gray lines, include two clones carrying only the C regions of G2B (583K20, 552J13) for which insert-end sequences (boxes E and D) hybridized to BAC clone 312E1. The F sequence from 312E1 (box F) was located 4.4 kb upstream of the G2A leader and 16 kb downstream of the D sequence. BAC clone 075N10 bridged the C exons of G2A and the VH gene segments of G5 (15). BAC clone 099K21 carried only the G5 C exons and extended 35 kb beyond CH1. The distances are shown at the bottom, as determined between the genes. The entire area includes: distance from G2B to G2A (219 kb), from G2A to G5 (140 kb), and the distance G5 to the end of 099K21 (43 kb), a total of 402 kb. Except for 312E1, the BAC clone insert sizes were determined by pulsed-field electrophoresis.
between their accessibility to RAG and their chromosomal positions.

**Rearrangement occurs at pseudogenes**

The single-cell experiments monitored only the functional IgH. Transcripts from four pseudogenes in nonrearranged, nonmutated form have been described (15 and references therein), but it was not clear whether RAG or AID would act upon them. RT-PCR was performed using shark-GR spleen RNA to amplify pseudogene G8 transcripts. Among 23 clones were three partially rearranged V-DJ as well as six VDJ with the J spliced to C region (not shown). Most of these carried substitutions, as did half of the 14 unrearranged VH (G8-15, -28, -31, Fig. 4); it is not known whether the latter sequences were transcribed from a GL or V-DJ configuration. In other experiments, we have found that the pseudogene G2C can also rearrange and hypermutate (C. Zhu and E. Hsu, unpublished observations). We assume that the four pseudogenes (G2C, G6, G7, and G8) are potentially accessible to RAG as they are transcribed and carry canonical RSS (15).

**Sequential activation of IgH**

Because the four pseudogenes are potential targets of RAG, there are 13 IgH loci in shark-LA prepro B cells that can become accessible for rearrangement. To obtain the full picture of all IgH in a shark B lymphocyte, we used data of varying completeness from 36 samples (Supplemental Fig. 3), thereby taking into account missed information. We asked what is the potential number of VDJ events in each cell, and is this number fixed across cells?

Table I shows the results obtained from two kinds of probability models. In the first, the total number of VDJ events in each cell is fixed and has a value of \( u \) to be estimated. In the second, VDJ events occur as a sequence of trials that stop when the first in-frame VDJ occurs. If this happened at a pseudogene, it is considered unobservable. A maximum of \( u \) VDJ events can occur, and...
only one is in-frame. Results of the statistical calculations are detailed in the footnotes to Table I. In the first model, the best estimate was three (Table I, Fixed column), and although most categories showed good fit with the model, the global goodness of fit ($G^2 = 38.0$ with $p = 0.076$) did not.

The sequential model ($u = 4$) fares better than the $u = 3$ model; in the former, no $p$ value is below the customary 0.05 threshold (Table I, Sequential, $u = 4$ column). The $u = 3$ model overall goodness-of-fit statistic $G^2 = 72.5$ with $p = 0.041$, suggesting poor model fit, but for the $u = 4$ model $G^2 = 56.5$ with $p = 0.209$. Thus, of the two types of model examined, the one that fits the data more satisfactorily states that the events (gene activation to rearrangement) occur sequentially and features a maximum of four VDJ events per cell.

**FIGURE 4.** Mutated pseudogene G8 sequences. The G8 cDNA sequence, consisting of leader, leader intron, GL VH gene segment directly spliced to C region, is aligned to the most closely related functional gene, G1 (top). The genomic GL G1 sequence shows leader, leader intron with splice sites (indicated in lowercase), and the VH gene segment demarcated into framework and CDR (underlined). The RSS (heptamer, spacer, nonamer) are labeled. G8 is nonfunctional due to a mutation disabling the intron splice acceptor site. G8 mutants 15, 28, and 31 are compared with the G8 cDNA, with dots representing identity and dashes gaps. The PCR primer is shown in italics.

### Table I. Joint frequency of number of evaluated genes ($Y$) with number of observed VDJ events ($X$)

<table>
<thead>
<tr>
<th>Total IgH Per Cell (Y)</th>
<th>No. of VDJ Events ($X$)</th>
<th>$p$ Value$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$\theta = 3$</td>
<td>$\theta = 3$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>

$Y$ is the total number of IgH, GL or rearranged, detected per cell; $X$ includes VDJ (3R), 2R, and 1R.

$^a$For an exact test of the null hypothesis that data fit the $\chi^2$ VDJ model. Total designates global goodness of fit. $^{\mathrm{CMLE}}$ of discrete parameter $\theta$ is 3.

$^b$Logarithms of the CLs were computed as $\text{CL}_3,Y^2 = 41.3$; $\text{CL}_4,Y^2 = 45.0$; $\text{CL}_5,Y^2 = 56.6$; $\text{CL}_6,Y^2 = 75.5$; $\text{CL}_7,Y^2 = 95.7$. Our CMLE of $\theta$ is 3.

$^c$Logarithms of the CLs were computed as $\text{CL}_3,Y^2 = 53.6$; $\text{CL}_4,Y^2 = 53.0$; $\text{CL}_5,Y^2 = 54.0$; $\text{CL}_6,Y^2 = 55.0$; $\text{CL}_7,Y^2 = 55.6$. Our CMLE of $\theta$ is 4.

G2A is different from G2B

Rearrangement, in-frame as well as nonproductive, is evenly distributed between G2A and G2B among the shark-LA cells (Fig. 3, top panel). This result suggests that the two closely related H chain isotypes are activated for recombination at the same frequency and in this animal are expressed on the cell surface at similar levels. We investigated G2A/G2B expression in pups like shark-LA as well as in adults. First-strand cDNA was primed with the G2 transmembrane sequence, so that the PCR products have a greater likelihood of reflecting the resting cell population (26). The RT-PCR products were blotted, hybridized with vh probe, and the signals analyzed by phosphorimaging. Comparable amounts of DNA were digested with BstE II, a site present only in...
the JH of G2A (Fig. 5A). The intensity of the ∼661 bp BstE II-negative band (G2B) is calculated as a fraction of the total G2 DNA (Fig. 5B). The pups express similar amounts of G2A and G2B (0.47 ± 0.04 SD), whereas the adults overall express significantly less G2B (0.34 ± 0.06 SD). The Mann–Whitney U test gave a value of \( p = 0.021 \), disproving the null hypothesis that the distribution of values is the same across the categories. The adults showed varied skewing in favor of G2A, whether the shark was injected with hen egg lysozyme (shark-JS), Ebola virus (shark-GR), DNP-LPS (shark-PI), or was unimmunized (shark-33).

**Hypermutation at individual IgH genes in a B cell**

The criterion for scoring mutant Ig is the presence of adjacent substitutions that are the hallmark of SHM in sharks (18, 27).

Neonatal Ig sequences are hardly ever mutated, and only one shark-LA rearrangement (162GA) carried substitutions; these consisted of three point mutations and one doublet. The in-frame VDJ in the same cell (16G4CG) contained one point mutation in CDR1. To study patterns of hypermutation among the IgH genes in one cell, sIgM+ cells from the immunized adult shark-GR were analyzed. The VDJ had been sequenced from 13 single cells (13), of which contained mutants (62%). We reanalyzed these eight sequences from 13 single cells (13), of which contained mutants (62%). We reanalyzed these eight cells, as well as eight new ones of this series selected according to the criteria described in Materials and Methods. The five cells in Fig. 6 were chosen for presentation because most of their IgH genes (≥7 out of 10) could be characterized. The VDJ were isolated as described for Fig. 1, reaction 3, as were the GL VH per Fig. 1, reaction 5.

**Screening for SHM at nonrearranged genes**

In the previous experiment, the GL sample size may have been too small to detect SHM occurring at low frequency. GL VH sequences were therefore amplified from genomic DNA from shark-GR sIgM+ cells (13). G2B was chosen because G2B genes in GL configuration are expected to be abundant in shark-GR (see “G2A is different from G2B”). Sequences in GL and various stages of recombination were analyzed. Of 88 GL sequences, 55 carried 1–4 substitutions, a frequency of 0.04% (33 out of 76,912 bp), and 13 independent 1R (VD-D-J) sequences 0.05% (6 out of 11,362 bp). In no case were any tandem substitutions observed. These frequencies were indistinguishable from that obtained from 12 GL sequences from shark-GR erythrocyte DNA, in which the polymerase misincorporation frequency was 0.04% (3 out of 8474 bp). Of the 11 2R (VD-D-J) sequences, 8 of them contained one substitution (1 out of 3768 or 0.03%), but 3 other sequences were mutated throughout VH and the intersegmental region (8–48 changes).

**Discussion**

A limited window for rearrangement

Using mAbs against nurse shark L chain and nurse shark IgM H chain, B cells were respectively isolated from PBL of an immunized shark and spleen of a nonimmunized neonate. In each case, not more than three VDJ were isolated per cell (Fig. 3 (13), although 9 to 10 functional \( \mu \) genes and their alleles were available. As few cells carried more than one in-frame VDJ, H chain exclusion holds throughout ontogeny. Assuming that it is primarily rearranged genes that are transcribed, these results are comparable to what was found by single-cell RT-PCR in the clearnose skate, in which only a few \( \mu \) transcripts were isolated per cell (28). Because the number of IgM H chain genes can vary considerably among cartilaginous fishes, ∼50–100 in horned shark (29), it is the restriction of gene activation that the two systems display in common.

In nurse shark, every \( \mu \) gene must rearrange to be expressed, and RAG acts upon only a few IgH per B lymphocyte (15). The chromatin accessibility appears to affect genes individually and at random, because the pattern of rearranged genes has no correlation with their linkage relationship. Nor was the allele more likely to be activated, because it was usually in GL configuration. These observations suggest that shark IgH gene clusters operate autonomously. Although it would have seemed that multiple loci offer an advantage for multiple rounds of rearrangement, viz. until a productive VDJ is generated, this in fact does not occur because we have so far not detected more than three VDJ in any sIgM+ cell. Moreover, it appears that the pseudogenes could compete for NFs that bring about chromatin activation.

To obtain some idea of the manner in which the IgH are activated, two scenarios were tested, the first being an instructive model in which a set number of genes/chromosome locations become activated, in this way directing suppression of other locations. The second model was one in which gene activation occurred in sequence and stopped with in-frame VDJ in the functional IgH. Our
data showed the most satisfactory overall goodness of fit with the latter model, with up to a maximum of four activated IgH per B cell. We have not cloned four VDJ in a B cell so far, but the sample size was small, and we have not included detecting the pseudogenes (31% of gene pool) in the single-cell assay.

In the mouse, Hewitt and coworkers (30) suggested that a direct interaction between the two IgH alleles leads to the relocation of one to suppressive nuclear compartment after cleavage by RAG on the other. However, interallelic cross talk cannot fully explain H chain exclusion in the shark because linked IgH and their alleles rearrange independently of each other. Perhaps more relevant to the shark observations is a model recently generated for the TCR b locus, which proposes that if the two alleles are autonomous and activated stochastically, they are unlikely to initiate recombination at the same time (10). A sufficient time lag between the two steps, D to J and V to DJ, would permit a productive VDJ at the first allele to effect inhibition on the second, rearranging allele. In shark, a few 2R/1R, like DJ, were found in 36 B cells, suggesting feedback was possible. The feedback signal could come from a surface receptor consisting of H and L chains, if IgL rearranged at the same time, as we have proposed (31), or with an as-yet-unidentified shark surrogate L chain that might be a prejoined (VJ) L chain type like NS3 (18).

Because shark VDJ (VDDJ) is formed in one step, feedback inhibition could be expected to occur during a time lag between the activation of individual H chain genes. If rearrangement initiation is stochastic, this reduces the likelihood that more than one IgH is poised for rearrangement in a cell. Coupled to a time window, the number of opportunities for attempting rearrangement would be limited; our data fit a model with a maximum of four activated IgH. Taken together, IgH autonomy, stochastic activation of IgH, and a restricted time window could explain much of the basis for H chain exclusion in a system with multiple Ig loci. These conclusions might be applied to the amphibian *Xenopus* with tetraploid IgH (32) and the Atlantic salmon with its two independently expressed IgM loci (33, 34).

RAG accessibility is regulated

Some insight into what is involved in regulated activation at a shark IgH gene may be gained by re-examining IgH recombination events in thymocytes (13). Two cell populations are compared in Table II, one in which the sIg+ cells were depleted (thymocytes) and the other positively selected (this study). In single thymocytes, up to seven rearranged genes can be observed, although 82% (46 out of 56) were not fully recombined (1R and 2R). In contrast, only one to three IgH genes were recombined in B cells, but 29 out of 30 were VDJ (3R). There were more recombination events in thymocytes: 109 events in 12 cells in Table II (13 in 1R plus 66 in 2R plus 30 in 3R) or an average of 9 per cell, compared with 5.6 per B cell (89 events in 16 cells, 2 in 2R plus 87 in 3R). It is not that there is less recombinase activity in thymocytes but that RAG might be described as not acting processively at most genes. With such striking differences between B and T cells, the recombination events occurring in thymocytes could not have taken place in a common lymphoid progenitor. Because RAG can be recruited to perform VH to DJ rearrangement in G2 genes of any...
developing lymphocyte, the difference arises from its ability to continue VD to D2 in B cells. Recent findings indicate that RAG activity is regulated by the nature of chromatin structure or the epigenetic environment of the RSS (35, 36).

Chromatin immunoprecipitation assays have revealed that whereas RAG1 binds RSS at active J genes segments, RAG2 is separately brought to the site by binding of modified histone (trimethylated lysine 4 of histone 3), an activating histone mark (36, 37). The RAG proteins recruited at this recombination center function to capture DNA with RSS (e.g., D gene segments) and bring about rearrangement. In thymocytes, we speculate that some kind of recombination centers can form but they are not stable because the epigenetic environment is not optimal for recombination catalytic activity. The fact that we could not detect H chain transcription in thymocytes furthermore suggests that the nature of the chromatin activation is different from B cells (13, 38).

**SHM is not uniform**

We investigated SHM among the IgH genes in single cells to determine which genes were targeted for change and whether gene conversion/sequence exchange could be detected. Because we found occasional GL Ig transcripts in cDNA libraries (E. Hsu, unpublished observation), there was the possibility that transcribing, nonrearranged genes could be bystanders in a hypermutating B cell. Although we have comprehensive information for only five B cells, the results are clear. SHM targeted rearranged genes, and its effect was not necessarily uniform within a cell. The extreme instance is in KS15, in which only one of the two nonproductive VDJ was targeted and carried 34 substitutions and 2 insertions. If tandem conversion/sequence exchange could be detected. Because we found occasional GL Ig transcripts in cDNA libraries (E. Hsu, unpublished observation), there was the possibility that transcribing, nonrearranged genes could be bystanders in a hypermutating B cell. Although we have comprehensive information for only five B cells, the results are clear. SHM targeted rearranged genes, and its effect was not necessarily uniform within a cell. The extreme instance is in KS15, in which only one of the two nonproductive VDJ was targeted and carried 34 substitutions and 2 insertions. If tandem

### Table II. Comparison of IgH gene rearrangement configurations in shark T and B cells

<table>
<thead>
<tr>
<th></th>
<th>Thymocyte</th>
<th>sIgM⁺ Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single cell¹</td>
<td>Up to 8 IgH recombined</td>
<td>Up to 3 IgH recombined</td>
</tr>
<tr>
<td></td>
<td>Average 5</td>
<td>Average 2</td>
</tr>
<tr>
<td>Genomic DNA²</td>
<td>IgH rearrangements</td>
<td>IgH rearrangements</td>
</tr>
<tr>
<td></td>
<td>No IgL rearrangements</td>
<td>IgL rearrangements</td>
</tr>
<tr>
<td>RNA studies³</td>
<td>No H chain mRNA detected</td>
<td>H chain mRNA</td>
</tr>
<tr>
<td></td>
<td>No L chain mRNA detected</td>
<td>L chain mRNA</td>
</tr>
<tr>
<td></td>
<td>No VDJ-J detected</td>
<td>VDJ-J mutants cloned</td>
</tr>
</tbody>
</table>

¹Single thymocyte data are from Ref. 13; the sIgM⁺ column is from present study. 1R is one rearrangement event: VD-D-J or V/D-D-J or V/D-D-J. 2R is two events: VDJ-J or V-DDJ or VD-DJ. 3R is three events or VDJ.

²Rearranged IgH in lymphocyte DNA, as detected by Southern blotting and PCR, were obtained in Refs. 13 and 41. Rearranged IgL in sIgM⁺ genomic DNA were obtained in Ref. 41; and the absence of rearranged IgL in thymus DNA was reported in Ref. 13.

³Northern blotting and RT-PCR (13).

that this phenomenon is at low frequency in the general sIgM⁺ cell population in shark-GR. The difference between results from genomic and cDNA sources suggests that SHM at IgH in GL configuration occurs in a minority of cells and under circumstances that are at present not clear.

**IgM isotypes may not be all the same**

Whether all of the IgMs in a cartilaginous fish serve the same function has not been investigated. G2A and G2B rearranged at the same rate in pup cells (Fig. 3), and the transmembrane receptors they encode were expressed at similar levels in shark-LA as well as three other pup samples (Fig. 5B). However, in all of the older animals, G2A predominated. It is difficult to understand why such a skewing came about based on the V regions, because the few differences are Met/Leu in FR1, Ala/Thr in CDR1, Gly/Ala in CDR2, and Ile/Lys in FR3; the Asp/Gly in JH is often deleted during rearrangement. G2A and G2B differ by 24 aa in the 415 residue C region, and of 17 nonconservative differences, 10 are in the 104-residue CH2 (3 out of 99 in CH1, 3 out of 102 in CH3, and 1 out of 110 in CH4). These differences reflect our previous phylogenetic analyses (15) that the domains in all functional μ genes evolved at comparable rates except for VH and CH2, which are under strong positive selection for sequence diversity. There is little information on shark Ab effector function and none on whether the various IgM isotypes are specialized. The IgM are secreted as pentamers and switch to monomers during an Ab response (39, 40), but there are no data on whether all of the H chains are equally capable of doing so. Because there appears to be G2A prominence in every adult, we propose that not all IgM isotypes are equal. But the basis for expression differences, in the absence of information on effector function or RNA t½, requires further investigation.

In conclusion, we propose that the stochastic formation of a few stable recombination centers allows rearrangement in B cells to proceed by the RAG efficiently capturing nearby RSS that flank the four gene segments. The neighboring IgH genes are too distant for frequent intergenic rearrangement events, even if all are simultaneously accessible to RAG; this partly explains the recombinational autonomy observed. The nurse shark IgH are regulated independently of allele or chromosomal position. The scoring of VDJ in single B cells best fits a probability model based on sequential IgH activation, which implies possibility for feedback inhibition once a productive VDJ has been generated.
Acknowledgments
We thank Ali Zarrin for help in mapping the IgH genes. We also thank Victor Lee, Karolina Malecek, and Anita Lui for work characterizing the nurse shark BAC library.

Disclosures
The authors have no financial conflicts of interest.

References