

Available online at www.sciencedirect.com

Gene xx (2004) xxx–xxx

www.elsevier.com/locate/gene

Evolution of human IgH3'EC duplicated structures: both enhancers HS1,2 are polymorphic with variation of transcription factor's consensus sites

Vincenzo Giambra^a, Alberto Fruscalzo^b, Maria Giufre^{a,1}, Cristina Martinez-Labarga^a,
Marco Favaro^a, Mariano Rocchi^c, Domenico Frezza^{a,*}

^aDipartimento di Biologia Enrico Caffè, Università degli Studi TorVergata, Viale della ricerca scientifica, 00133 Roma, Italy

^bIstituto di Biologia e Patologia Molecolari, CNR, Roma, Italy

^cDAPEG, Sezione di Genetica, Università di Bari, Italy

Received 16 July 2004; received in revised form 20 September 2004; accepted 5 October 2004

Received by M. D'Urso

Abstract

The enhancer complex regulatory region at the 3' of the immunoglobulin heavy cluster (IgH3'EC) is duplicated in apes along with four constant genes and the region is highly conserved throughout humans. Both human IgH3'ECs consist of three loci high sensitive (HS) to DNase I with enhancer activity. It is thus possible that the presence of structural divergences between the two IgH3'ECs and of relative polymorphisms correspond to functional regulatory changes. To analyse the polymorphisms of these almost identical regions, it resulted mandatory to identify the presence of divergent sequences, in order to select distinctive primers for specific PCR genomic amplifications. To this aim, we first compared the two entire IgH3'ECs in silicio, utilising the updated GenBank (GB) contigs, then we analysed the two IgH3'ECs by cloning and sequencing amplicons from independent genomes. In silicio analysis showed that several inversions, deletions and short insertions had occurred after the duplication. We analysed in detail, by sequencing specific regions, the polymorphisms occurring in enhancer HS1,2-A (which lies in IgH3'EC-1, 3' to the C α -1 gene) and in enhancer HS1,2-B (which lies in IgH3'EC-2, 3' to C α -2). Polymorphisms are due to the repetition (occurring one to four times) of a 38-bp sequence present at the 3' of the core of enhancers HS1,2. The structure of both human HS1,2 enhancers has revealed not yet described polymorphic features due to the presence of variable spacer elements separating the 38-bp repetitions and to variable external elements bordering the repetition cluster. We found that one of the external elements gave rise to a divergent allele 3 in the two clusters. The frequency of the different alleles of the two loci varies in the Italian population and allele 3 of both loci are very rare. The analysis of the *Callicebus moloch*, *Gorilla gorilla* and *Pan troglodytes* HS1,2 enhancers showed the transformation from the ancestral structure with the 31- to the 17-bp external element in hominids.

The relevance of the polymorphisms in the HS1,2 enhancers is due to the variable number of binding sites for the transcription factors: NF- κ B, CMYB, BSAP1/2, AP1/4, E47, MyoD and μ E5 and thus to the possible influence of these variations on switch, production of Ig and on maturation of B cells.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Enhancer complex; GenBank contigs; Ancestral gene; Phylogenesis; Italian population

Abbreviations: IgH3'EC, immunoglobulin heavy chain three prime enhancer complex; HS1,2, DNase I hypersensitive sites 1, 2; C α -1 or C α -2, constant heavy chain gene α 1 or 2; LCR, locus control region; $\psi\gamma$, pseudo gamma constant heavy chain gene; $\psi\epsilon$, pseudo epsilon constant heavy chain gene.

* Corresponding author. Tel.: +39 6 72594814; fax: +39 6 2023500.

E-mail address: frezza@uniroma2.it (D. Frezza).

¹ Present address: Dipartimento di Malattie Infettive parassitarie ed immunomediate, Istituto Superiore di Sanità, Roma, Italy.

0378-1119/\$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.10.009

1. Introduction

The structure, and possibly the function, of the enhancers situated in the immunoglobulin regulative region, at the 3' end of the constant heavy chain cluster, is largely conserved from mouse to human (Mills et al., 1997), although in rabbit only one 3' enhancer is present after 13 copies of the constant heavy α gene (Volgina et al., 2000).

The role of the complex regulative region in Ig transcription and in the path of maturation of B lymphocytes is not definitively known (Hu et al., 2000; Saleque et al., 1999). Several transcription factors have been shown to be necessary for B cell maturation (Conley, 2003). The interaction of some of these factors with germ line (GL) promoters and enhancers are evidenced in the maturation process and Ig production of B cells (Laurencikiene et al., 2001). The four enhancers in the mouse and the three in humans at the 3' of C alpha genes behave as an LCR and have a concerted activity in class switch recombination (Manis et al., 1998; Chauveau et al., 1998a). Moreover, HS1,2 enhancer has an important role in germ line transcription where the polymorphic *cis*-acting factor sites can influence the enhancer activity on reporter genes (Pan et al., 2000; Denizot et al., 2001). In mouse knock out of P50/NF- κ B or C-rel subunit cause several abnormalities in Ig transcription (Michaelson et al., 1996) showing differential interactions of the NF- κ B subunits with HS1,2 after B cell stimulation with LPS or CD40 (Zelazowski et al., 2000). The majority of these studies examine the mouse model where the enhancer complex is present in one copy. It is relevant to investigate the activity of the enhancer HS1,2 also in the human enhancer complex where it is duplicated, has polymorphic features and is missing the enhancer HS3-B.

In vitro experiments show interactions of the enhancer specific binding sites with several transcription factors (Besette et al., 2003; Andersson et al., 2000).

The partial duplication encompassing the four constant genes γ 3, γ 1, $\Psi\epsilon$, α 1 in apes did generate the cluster γ 2, γ 4, ϵ , α 2 at the 3' of the $\Psi\gamma$ gene around 60 million of years ago, as suggested by Harindranath et al. (1998); an interesting model for this duplication of the locus was proposed by Sadhu et al. (1997). The duplication of the cluster of the constant heavy genes in the evolution of primates created new opportunities for the immunoresponse not only with a double repertoire of heavy genes but also with the new 3' enhancer complex (Kawamura et al., 1992). In fact, the IgH3'EC at the 3' of C α -1 gene, corresponding to a DNA stretch of \approx 35 kb, was duplicated altogether with the cluster. The genomic structure of the two IgH3'ECs at the 3' of both C α -1 and C α -2 (called IgH3'EC-1 and IgH3'EC-2, respectively) is extensively conserved: in fact, an overall identity of the major features, with little degenerations, is observed, and the differences are very limited (Chen and Birshstein, 1997). Therefore it is technically difficult to attribute each sequence to one of the two loci (Pinaud et al., 1997). Moreover, due to the instability of this specific DNA structure, it is mandatory to adopt caution and appropriate controls for the analyses of DNA polymorphisms of the IgH3'EC regions. Frequent rearrangements have been observed as RFLPs (Bech-Hansen et al., 1983; Frezza et al., 1998) or as deletions or duplications involving large DNA fragments (Bottaro et al., 1991; Rabbani et al., 1996) which could involve also the regulatory regions with a possible influence on Ig expression (Rabbani et al., 1995).

These complex regulatory structures probably play an architectural role in the IgH3' enhancer complex function: it may favour recombination events, matrix interactions and the disposition of the DNA helices for transcription (Pinaud et al., 2001). Several studies demonstrate the influence of the two IgH3'ECs on B-lymphocytes maturation and IgH transcription; moreover, a primary relevance was ascribed to the central enhancer HS1,2 (Chauveau et al., 1998a; Pan et al., 2000). More recently, new evidences indicate the role of the enhancer HS4 within the IgH3'EC (Sepulveda et al., 2004).

We were interested in a detailed analysis of the structure of the duplicated IgH3'ECs and of the boundary regions and in determining the polymorphisms of the enhancers HS1,2-A (3' to C α -1) and HS1,2-B (3' to C α -2). Therefore, we devised a strategy which allowed to selectively amplify the two different loci, in order to check if the fragments obtained were colinear and absent of discrepancies with the sequences of the GB. The detailed analysis allowed to recognise the presence of the new polymorphisms of the HS1,2-A and -B loci with the variation of relative transcription factor sites. Interestingly, the frequencies of the homologous alleles in the two loci resulted highly different in an Italian population sample. The comparison by alignment of the sequences of human, rodent, monkey and apes showed the details of the possible ancestral structure of the conserved enhancer.

2. Material and methods

2.1. In silicio analysis

Contigs of the constant heavy chain region were rescued from the GB with blast programs of the NCBI and from the maps of the chromosome 14 q32 telomeric region (www.ncbi.nlm.nih.gov/mapview/maps.cgi).

Comparison of the sequences of the IgH3'EC-1 and IgH3'EC-2 was performed with Clustal W programs online or with Gene Jockey program. The new contig CHR77 working sequence derived from the junction of five genomic contigs was assembled by the overlapping regions found by blast search. The new sequencing data of the clones of different genomes either from cell lines polymorphic for *Hind*III site or from human peripheral blood cells of healthy donors were compared with the same Clustal W and Blast programs. The WebGene program (in Cabibbo et al., 2002) at the web site <http://125.itba.mi.cnr.it/~webgene/genebuilder.html> was used to find the consensus sites for transcription factors on the polymorphic HS1,2-A and HS1,2-B region.

2.2. The selective PCRs

The 5.4-kb fragment of IgH3'EC-1 was selectively amplified by the primer SA2.5 (5'-GGA TCC CTG TTC CTG ATC ACT G-3') and A2R (5'-GCC CTT CCT

GCC AAC CTG-3') of the sequences U2 and U8, the last being inverted in IgH3'EC-2 together with the HS1,2-B sequence (see Fig. 1A and B). Conversely, to get the IgH3'EC-2 specific fragment of 4.4 kb, the primer A2R was selected respectively within the U8r sequence and it was paired to primer A2F (5'-GCA CTG TCG GCT TAC AGA GG-3') within UB2 which is unique in the IgH3'EC-2.

The reaction conditions were for the SA2.5 and A2R, or A2F and A2R primers: 1.5 units of Taq Polymerase Platinum High Fidelity (Invitrogen, Carlsbad, CA, USA) buffer 1× Platinum High Fidelity (200 mM Tris-HCl, pH 8.4, 500 mM KCl); dNTPs (0.2 mM), MgCl₂ (1.5 mM), primers 15 pmol, water for final volume of 50 μl. The reaction was done at 94 °C for 2 min followed by 10 cycles at 94 °C for

30 s, 59 °C per 30 s, 68 °C per 5 min, followed by 20 more cycles at 94 °C for 30 s, 57 °C for 30 s, 68 °C for 5 min and one final extension at 72 °C for 10 min. A nested PCR was performed for either loci with the same two primers, P3Fw (5'-GAC TCA TTC TGG GCA GAC TTG-3') and D3Rev (5'-GTC CTG GTC CCA AAG ATG G-3'), in order to obtain the polymorphic region of the enhancer HS1,2. The amplification conditions were performed with 1/10 of the volume of the selective PCR, to minimise DNA genomic carryover, with 1 unit of Platinum Taq Polymerase (Invitrogen) and 1× buffer platinum, dNTPs (0.2 mM), MgCl₂ (1.5 mM), primers (15 pmol) adding water to a final volume of 50 μl. The temperature steps of the reaction were as follows: 94 °C for 2 min followed by 30 cycles at 94 °C 30 s, 56 °C for 30 s, 72 °C for 1 min followed by a final

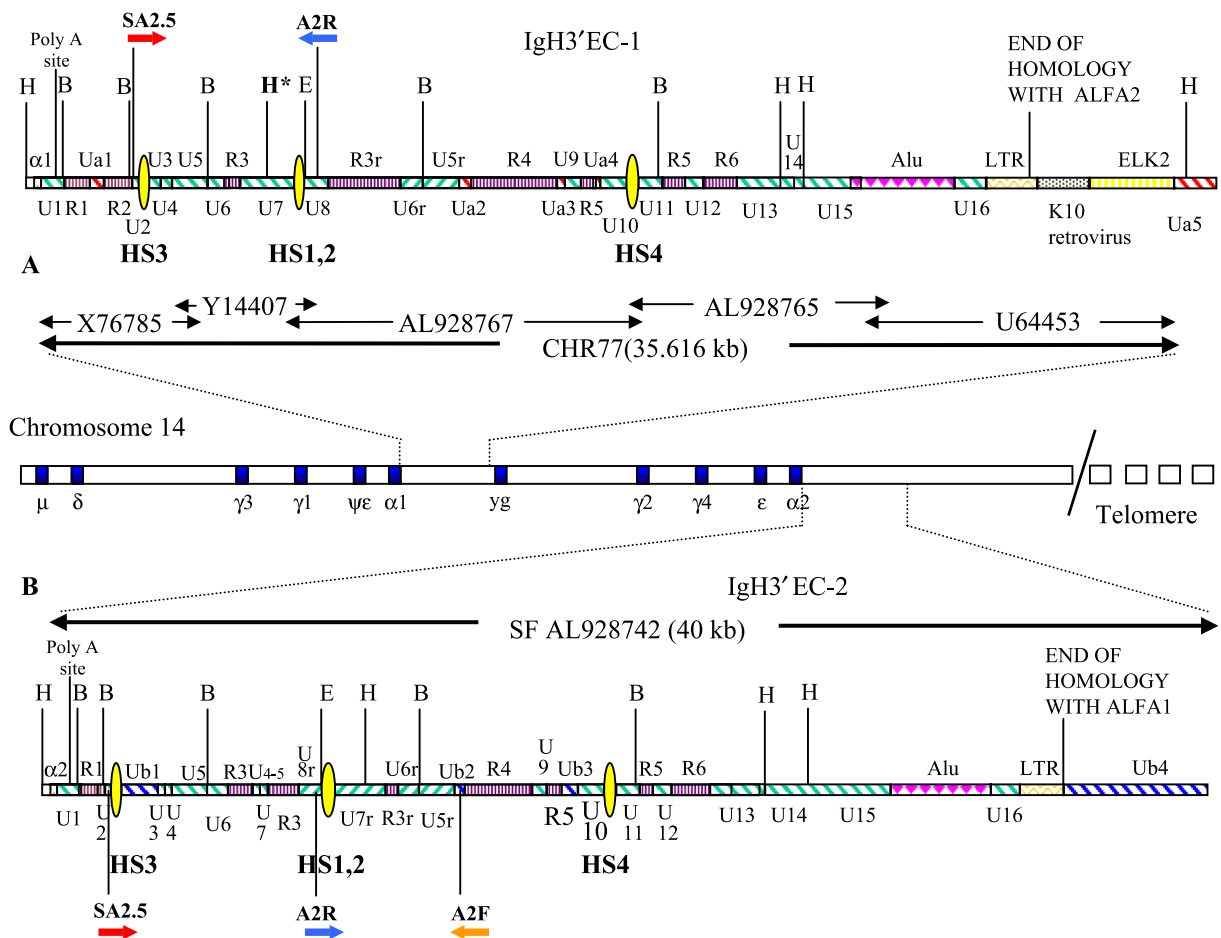


Fig. 1. Genomic organization of the Ig Constant heavy chain genes on chromosome 14q32 region: (A) enlargement of the 3' Cα-1 IgH3'EC-1; (B) enlargement of 3' Cα-2 IgH3'EC-2. The conserved sequence units (U), labelled from left to right with increasing numbers starting from the poly-A site of the membrane Cα exons, are distinguished by the sequences with tandem repeats (R) which are also numbered from left to right. The slashes, green if common to IgH3'EC-1 and IgH3'EC-2, blue with the letter U followed by a or b if unique for either one or the other IgH3'ECs, right and left oriented if present in a duplicated reverse form. The r that follows the Arabic number indicates that the sequence is duplicated in a reverse form. The border of the duplication is inferred by the homology among the two IgH3'ECs and the interruption is signed on the right of both A and B enlargements. The yellow elliptic inserts represent the three enhancers of the IgH3'ECs. The arrows indicate the allocation and direction of the primer sequences for selective amplification of the IgH3'EC-1 or IgH3'EC-2 loci. The contig CHR77 spanning the IgH3'EC-1 region is derived by the assembling of the five contigs reported in the map and in the GenBank with accession nos. X76785, Y14407, AL928767, AL928765 and U64453. The sequence spanning the IgH3'EC-2 is a sub-fragment (SF AL928742) of 40 kb of the contig accession no. AL928742. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

extension at 72 °C for 5 min. Controls with direct amplifications of the same quantities of genomic DNAs as for nested PCR did not result positive in these experimental conditions.

2.3. Monkeys and apes DNA amplifications

The amplifications of the enhancer HS1,2 from monkeys were performed adapting the same human primers used for the direct nonselective PCR of the short fragment. The apes enhancers were amplified using the same selective nested PCRs used for humans. The conditions were the same but for the annealing temperature were applied 2 °C less. The Southern blot analysis was performed on genomic DNA from five samples of *Macacus fascicularis* digested by *Xba*I restriction enzyme and hybridised with the human enhancer probe with the fluorescence system (Roche Diagnostic, Germany). DNAs were extracted from blood using the standard protocol.

2.4. The Italian samples and DNA preparation

The DNA of healthy donors was prepared with standard methods from a representative group of the middle-south Italy either by total blood or peripheral blood lymphocytes. The DNA of a group of donors was prepared by extracting the genomic DNA from the mouth epithelial cells embedded in a sterile cotton plug (Becton Dickinson, Sparks, MD, USA) in that case a lower quantity was harvested but sufficient for PCR analyses. The cotton plugs were removed from the support with a sterile lancet and introduced in a sterile vial with 1.0 ml of TE with SDS 1%, RNase (100 µg/ml), 20 µg/ml of proteinase K, then incubated o.n. at 37 °C gently shaking. The DNA was harvested as suggested by the standard protocol of Microcon (Millipore, Bedford MA, USA).

2.5. Cloning and polymorphisms analysis

The products of selective or not (mice and monkeys) amplifications were cloned with *TOPO XL PCR Cloning kit* (Invitrogen) in the plasmid pCR-XL-TOPO and electropored in *E. coli* TOP10 (Invitrogen) competent cells. The cells were plated on ampicillin selective agar medium, X-gal and incubated o.n. until the white/blue colonies were visible. Minipreps were performed with SIGMA kit (Illinois, USA) and DNA from positive clones extracted, sequenced with M13 forward and reverse universal primers and analysed by automatic AppliedBiosystem capillary 3700 apparatus at the BMR Sequencing Service CRIBI (Padova, Italy). DNA fragments particularly rich in GC repeats were resolved by use of DMSO in the sequencing reactions. The product of direct or nested PCR for the screening of polymorphism frequency in the Italian sample was electrophoresed on agarose gel 3% stained with ethidium

bromide and the alleles of HS1,2-A and HS1,2-B were analysed as evidenced in Fig. 4.

3. Results

3.1. In silico analysis of contigs from GenBank

Search of the GB contigs containing DNA sequences of the Ig 3' IgH3'EC-1 locus identified two novel informative contigs, of 10.8 and 7.3 kb, respectively, corresponding to the GenBank accession nos. AL928767, AL928765. The first of these novel sequences closed the gap between the HS1,2 and HS4 regions and the second the gap between the HS4 region and the pseudo gamma gene ($\psi\gamma$). The analysis of the overlaps of the extremities of the sequences rescued in the GB allowed the reconstitution of the CHR77 working contig (51 kb long) reported in Fig. 1A. The assembled sequence corresponds to the complete IgH3'EC-1 locus.

A partial IgH3'EC-2 sequence allowed to identify the GenBank accession no. AL928742 contig, which is 176 kb long and spans the four constant genes $\gamma 2$, $\gamma 4$, ϵ and $\alpha 2$ together with the complete IgH3'EC-2 (Fig. 1B). The IgH3'EC-2 region is represented in the GB by few independent sequences as compared to IgH3'EC-1, but their colinearity with GenBank accession no. AL928742 contig confirms the identity of the IgH3'EC-2 locus. The three reported contigs represent the most update information about this region of the human genome, except for the polymorphisms.

The reconstitution of IgH3'EC-1 as a continuous sequence allowed the direct comparison of the two constant heavy chain clusters. The structures of both regulatory regions are described in the next paragraph and in Fig. 1A and B, where the conserved or rearranged DNA sequences, as well as the *de novo* insertions, are shown. The analysis of the contigs confirmed the previous results obtained by genomic Southern analysis, which had shown the presence in both loci of internal duplications and inversions (Mills et al., 1997). The homology between the two IgH3'ECs is extended up for 34,550 bp in IgH3'EC-1 (nucleotide 5672 on the GenBank accession no. U64453) and 37,600 bp in IgH3'EC-2 (nucleotide 102.713 on GenBank accession no. AL928742 contig) and excludes the pseudo Elk gene. The 3' termini of the duplication of the constant heavy chain cluster include the LTR but not the unique sequence of the ERV-K10 retrovirus present at the 3' extreme of the IgH3'EC-1 region.

In summary, the duplicated IgH3'EC-1 and 2 regions (see Fig. 1) present 16 conserved sequence units, six repeated sequences and three inversions, which lie between the HS1,2 and HS4 enhancers. The unique sequences specific for each IgH3'EC are four in IgH3'EC-1 and three in IgH3'EC-2. The comparative analysis of the two contigs

was essential to define the unique regions where to allocate the primers for distinctive PCR amplifications, thus allowing the identification of polymorphisms of these regulatory regions.

3.2. Cloning the *IgH3'EC-1* and *IgH3'EC-2* subregions

The correct assembly of the regions described above was verified by new independent clonings from different genomes. To this aim, we developed a double strategy to amplify the two independent *IgH3'EC* regions, one in exclusion of the other (see Section 2.2). The cloned fragments were compared by Blast program to the CHR77 contig (*IgH3'EC-1*) and to the GenBank accession no. AL928742 contig (*IgH3'EC-2*). Each of the clones, which was derived from independent amplifications of five different genomes was identical to the sequence of only one of the two contigs, thus confirming the correctness of the PCR strategy and the accuracy in the construction of the two contigs.

The two loci described in Fig. 1 present conserved sequence units (U) and repeated sequences (R), which are numbered starting from the poly-A site of the membrane exon of the two alpha constant genes (Fig. 1A and B). Several inversions and duplications are present and identify the two *IgH3'ECs*: the “Ua” or “Ub” sequences (see Fig. 1), which are present only in one of the two regions, were utilised to select the primers for PCR distinctive analyses (see Sections 2.2 and 3.3). A confirmation that the selected sequences are not polymorphic came from the population analyses: several hundred genomes amplified with these primers always give the same products, as judged by gel electrophoresis and confirmed by sequencing.

3.3. The polymorphisms of the *HS1,2-A* and *HS1,2-B* enhancers

To study the polymorphisms of the *HS1,2* enhancers of the two *IgH3'ECs*, we applied the selective strategy of PCR amplification of either one of the two *IgH3'ECs*. To verify

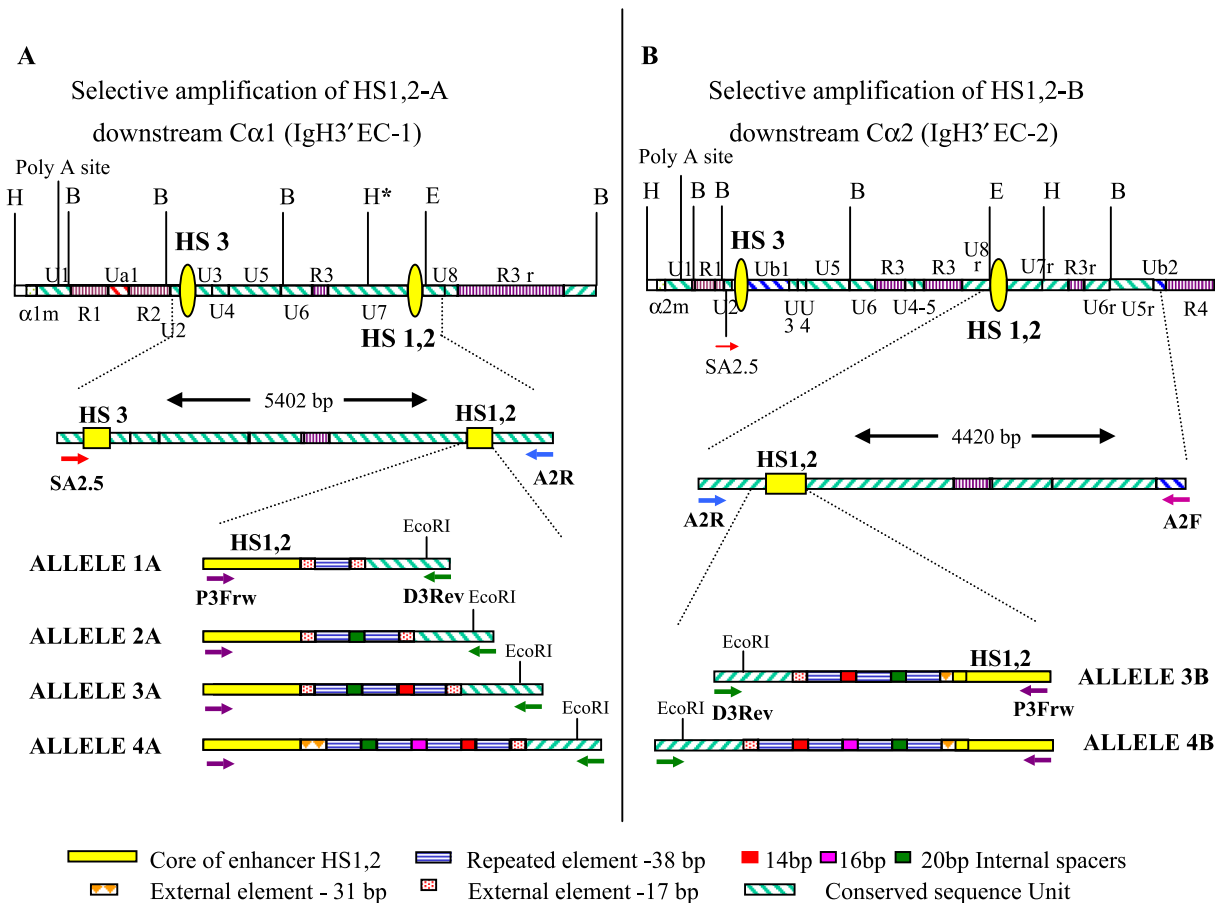


Fig. 2. Strategy of PCR selective amplifications for the *HS1,2-A* and *HS1,2-B* loci and structure of the polymorphisms. The primer positions are indicated by the arrows. The *HindIII* polymorphic site of the *IgH3'EC-1* region is marked by the asterisk (*). The *HS1,2-B* is inverted compared to the *HS1,2-A*, therefore the primer *A2R* is inverted in the two sequences and coupled with primer *SA2,5* amplifies only the *HS1,2-A*. The primer *A2F* corresponds to the *Ub2* unique sequence specific for *IgH3'EC-2* and coupled to *A2R* primer amplifies only the *HS1,2-B* enhancer. In the lower part of the figure are represented schematically in part A the four alleles of *HS1,2-A* and in the part B the two alleles inverted of *HS1,2-B*. The details of the repeated elements of the different alleles are schematically depicted in the legend and show the different elements separating the 38-bp repeated sequence.

the effectiveness of the method, we took advantage of a *Hind*III polymorphic site within the U7 fragment of the IgH3'EC-1 region (Fig. 1): in the genomes, where the restriction site is present between the HS3 and HS1,2 enhancers, the enzyme cuts a *Hind*III fragment of about 23 kb (Frezza et al., 1998). Primers SA2.5 and A2R, selected for the amplification of the HS1,2-A region, are situated in the two contiguous digestion fragments (Figs. 1A and 2A). In fact, the primers did not amplify any visible fragment, when the template genomic DNA containing the homozygous polymorphic *Hind*III site was previously digested. The absence of any other amplified fragment, including the IgH3'EC-2 region, confirmed the specificity of the test. The control with undigested DNA or with DNA from genomes lacking the *Hind*III site gave rise to the proper amplification fragment (5.4 kb, see Fig. 2).

The use of primer A2R paired to primer A2F amplified a unique fragment of the expected length (4.4 kb) corresponding to the IgH3'EC-region, which includes the HS1,2-B enhancer (Figs. 1B and 2B). The selective amplification for the IgH3'EC-2 locus was also controlled by the fact that, on

the contrary of IgH3'EC-1, the DNA digested or undigested with *Hind*III enzyme gave the same product.

The confirmation that, using the two specific primer pairs, only the IgH3'EC-1 or the IgH3'EC-2 clusters were amplified came from cloning and sequencing of the amplicons. In fact, the clones obtained from five independent genomes corresponded to either the working contig CHR77 or to the GenBank accession no. AL928742 contig. In agreement with previous reports (Chen and Birshtein, 1997; Mills et al., 1997; Chauveau et al., 1998b), we could also observe that enhancer HS1,2-B in IgH3'EC-2 is inverted in comparison with enhancer HS1,2-A of IgH3'EC-1 (GenBank accession nos. U84575, U84576, AFO13722, AF013724 and AJ298018).

The selective amplification and cloning of the HS1,2-A and -B enhancers revealed the occurrence of the alleles reported in Figs. 2 and 3B (GenBank accession nos. AJ544218, AJ544219, AJ544220, AJ544221, AY530200 and AY530201). The polymorphic sequences are present immediately at the 3' to the core of the enhancer. The four alleles of IgH3'EC-1 and the two alleles of IgH3'EC-2 vary

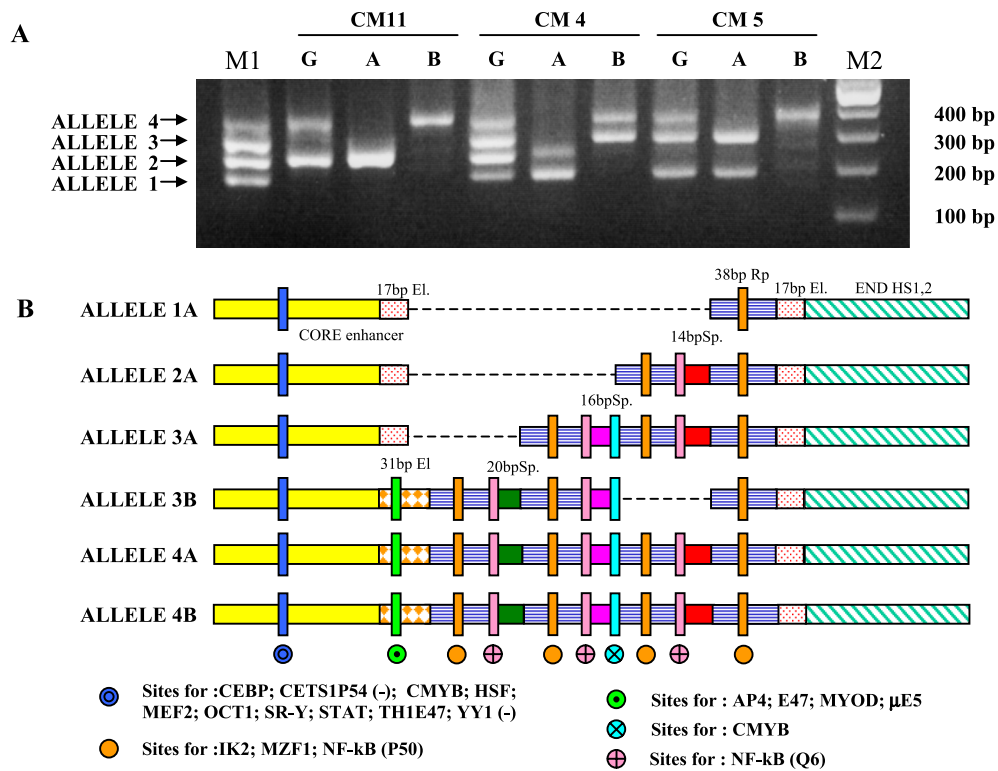


Fig. 3. (A) Agarose 3% gel electrophoresis of the different alleles of HS1,2-A and -B from three different genomic samples. Lane 1: M1, the standard size of the four alleles of a double heterozygous sample; lanes 2-9: genomic samples CM11/CM4/CM5, G nonselective amplification, A selective amplification of HS1,2-A locus and B HS1,2-B locus. In the G lanes, the alleles from both loci are visible. In A or B lane, the alleles of either selective amplification are visible. Lane 11: M2 is the size marker for the ladder of 100 bp. (B) Schematic representation of the six alleles piled up for the conserved and duplicated elements. The structure of the polymorphisms consists of the elements of 38-bp repeated and separated by spacers containing stretches of C of 17 bp on both borders in alleles 1-3 of HS1,2-A and only on one side of alleles 4 of both loci and on allele 3 of HS1,2-B, where on the other side is present the 31-bp spacer with one transcription site element. The internal spacers of 14 and 20 bp have one transcription site or two in the 16-bp element. The 38-bp repeated element presents one transcription site as well as the core of the enhancer. The names of the transcription factors are reported in the legend with the corresponding symbols mapped on the alleles. The sequence of the consensus of the transcription factors reported varies from 9 to 20 bp and the corresponding sequences are fetched with WebGene program cited in Section 2.1.

38-bp element of these alleles (3B, 4A and 4B) presents a transition from T to C at the first position and a transition from C to T at the 15th nucleotide. The differences do not alter the consensus for the transcription factors considered in this work. Each allele was sequenced twice from at least four clones from five independent genomes, either homozygous or heterozygous. The total sequences of the 40 clones analysed did not show the occurrence of other polymorphic sites.

3.4. Variations of consensus sites in the different HS1,2 alleles

The presence of different consensus sites for transcription factors in enhancers HS1,2 is related to the different organisations of the four alleles. The repeated element of 38 bp is present from one to four unvarying copies in the four alleles, and the copies are separated by different spacers in both A and B loci, as shown in Fig. 3B. The 38-bp element presents one copy of the NF- κ B (P50) site which consequently is present from one to four times in the four alleles. A NF- κ B (Q6) site is present at one border of all the internal variable spacers and is thus absent in allele 1 but present from one copy up to three copies in the other alleles. The spacer of 16 bp contains a consensus for the CMYB factor which consequently is present in one copy in allele 2A, and in alleles 3 and 4 of both loci. The 31-bp external element specific of alleles 3B, 4A and 4B contains sites for AP4, E47, MyoD and μ E5. Since the alternative 17-bp external element does not present known *cis* elements, the replacement of this element with that of 31 bp in allele 3B increases the number of binding sites for transcription factors of this allele in comparison with allele 3A. This is the main difference among alleles 3A and 3B. For more details, see Fig. 3B and the GB accession numbers of the six alleles reported in Section 3.3.

3.5. Analyses of HS1,2 in monkeys and apes

The nucleotide sequences of the enhancer HS1,2 of *C. moloch*, *G. gorilla* and *Pan troglodytes* which we isolated and sequenced (GenBank accession nos. AY649406, AY655739 and AY669117) were compared to rodents (GenBank accession nos. X96607, AF314408 of mouse and rabbit, respectively) and humans (GenBank accession no. AJ544218). The alignment of Fig. 4 shows the conservation of the core of the enhancer and the variability of the polymorphic region starting with either the 31- or the 17-bp external spacers. Among apes and man, only one base is substituted in *Gorilla* inside the core region of the enhancer. In the totality of the enhancer, *G. gorilla* and *P. troglodytes* (chimpanzee) have only 4 and 5 nucleotides of difference, respectively, where for *C. moloch* the differences rise to 29 bp. The apes so far analysed have a conserved copy of the allele 1A and the HS1,2-B locus has not yet been cloned but it is supposed to be present with the

Table 1

Allelic frequencies of the enhancers HS1,2-A and -B in the Italian population

Alleles of the enhancer HS1,2	Allelic frequencies of HS1,2-A	Allelic frequencies of HS1,2-B
1	0.439 \pm 0.022	–
2	0.395 \pm 0.021	–
3	0.061 \pm 0.011	0.070 \pm 0.018
4	0.105 \pm 0.014	0.930 \pm 0.018
Total sample	248	100
χ^2	1.47	0.02
Degrees of freedom	5	1
<i>P</i>	0.75< <i>P</i> <0.50	0.900

IgH3'EC-2 duplicate region. The new alleles originated with duplication of the 38-bp element separated by the short spacers of 14, 16, 17 and 20 bp with different *cis* elements for annealing of transcription factors have not been found yet in monkeys or apes. In the human, IgH3'EC-2 presents only the alleles with the 31-bp initial spacer which has been found only in the monkey but not in the apes which were analysed. A Southern analysis of 5 genomic DNA *Xba*I digested of *M. fascicularis* showed the polymorphism of the HS1,2 region (data not shown).

3.6. The frequency of the HS1,2 alleles

The frequency of the different alleles of enhancers HS1,2-A and -B has been determined in a sample of 248 Italian independent healthy donors. In the analysed group, alleles 1A and 2A are respectively present with a frequency of 0.44 and 0.39. Allele 3A is the most rare, with a frequency of 0.06, whereas allele 4A is present with a frequency of 0.10. The χ^2 analysis is highly significant for the equilibrium test (see Table 1).

Alleles 4B and 3B are represented with the respective frequencies of 0.93 and 0.07. The HS1,2-B enhancer has already been described in a sample of French population and shown to be present only in the form of allele 4 (Pinaud et al., 1997). Alleles 1 and 2 were not detected for enhancer HS1,2-B, but this might depend on the size of the sample analysed of 100 diploid genomes.

4. Discussion

The novel genomic contigs spanning the 3' duplicated regions of the IgH genes cluster and present in the GB since the end of 2002 have allowed to assemble a continuous nucleotide sequence of the entire region, from the C α -1 gene up to the IgH3'EC-2, thus providing new information about the structure of the two related IgH3'ECs. This has consented to develop a strategy for specific amplifications of each of the duplicated regions utilising the sequences that diverge in the two IgH3'ECs. The regions corresponding to enhancers HS1,2 of both IgH3'ECs were amplified from several independent genomes and their sequencing, besides

confirming the differences observed in the Human GB, allowed to distinguish the allelic variants of each locus.

Theoretically upon the frequency of duplications of the Ig heavy constant genes present in the human populations (Rabbani et al., 1996), we cannot exclude an extra copy of the enhancer HS1,2, although in the genomes analysed by the selective method, we never found discrepancies or more than two alleles.

With the actual techniques at present it is impossible to obtain information about the structure of the haplotypes with different combinations of the alleles, due to the difficulty of amplifying in a single reaction a fragment spanning about 180 kb.

The analysis of the HS1,2 enhancers has provided two main information: (1) also in the HS1,2-B locus, different alleles are present; (2) the polymorphisms in the two loci seem to have a different origin.

In a screening of 248 Italian healthy donors, we have identified the presence of four alleles of the HS1,2-A locus and two alleles of the HS1,2-B locus from a sample of 100 Italian healthy donors (see Fig. 2). Allele 4-A of HS1,2-A and allele 3-B of HS1,2-B, which have very low frequencies in the Italian population (Table 1), had not been described before.

The different alleles are characterised by the number of repetitions and by the nature of the two external spacers, here distinguished on the basis of their length: 17 and 31 bp (see Figs. 3 and 4). The observation of the structure of the alleles suggests that, as allele 4 is identical in both IgH3'ECs, the duplication of the cluster took place in a genome harbouring that allele. On this presumption, allele 3-B of IgH3'EC-2 could have originated from allele 4 of the same locus, possibly indicating a tendency of the enhancer to curtail the number of repetitions. In the case of the alleles of probable independent origin 1-A, 2-A and 3-A of IgH3'EC-1 there are no indications to favour the hypothesis that deletions from allele 3 produced the two shortest alleles or vice-versa that alleles 2 and 3 derived from allele 1 by duplications; it has however to be taken into account that the ancestral structure must have contained only one element, which eventually would have been amplified and then reduced. It is remarkable that alleles 1 and 2 are highly present in European populations, indicating a positive selection for these two forms compared to the other two alleles. It would be interesting to understand if and why the presence of the shorter forms of the enhancer HS1,2 represent a possible advantage.

In this respect, the cloning and sequencing of the HS1,2 from *Callicebus*, *Gorilla* and chimpanzee evidenced in all three cases the allele 1, although with the 31-bp spacer for the monkey. It is possible that allele 1 is the more commune allele as it is for human European populations (V. Giambra, in preparation). The comparison of the monkey and apes DNA sequences confirmed that the core enhancer is the more conserved part of the enhancer, but among the most recent species it resulted

that the 38mer is also very conserved. It has to be understood why this 38-bp structure is so important to be not changed in the duplications that gave rise to the different alleles.

Our observation (data not shown) that the enhancer of *M. fascicularis* of the only IgH3'EC present is polymorphic supports the hypothesis that different allelic forms were already present before the duplication, and the study of the alleles present in monkeys and apes should help to recognise the polymorphic structure of the ancestral gene. It is worth mentioning that in mouse, rat and rabbit the HS1,2 element shows a 3' conserved part resembling the 31-bp element present in the human alleles 4 of both loci and in allele 3-B of HS1,2-B. The alleles 4 and 3-B which hamper the 31mer could be supposed the oldest human alleles. It can be inferred that the possible transformations toward the actual shape of human alleles started from an ancestral gene with the 31-bp spacer and with one 38-bp element as suggested by the structure of *C. moloch*. It could be also hypothesized that in the hominids, since the duplication of the IgH3'EC happened, the time resulting was not sufficient to introduce in the IgH3'EC-2 locus as many polymorphisms as observed for the IgH3'EC-1 locus. As well, it could be possible that the relative function of the HS1,2-A has more influence on the selection of new alleles which resulted to enrich the immunosystem.

An obvious consequence of the presence of different HS1,2 alleles resides in the variation of the number of copies of the consensus sequences for several transcription factors. In particular, the number of sites for NF- κ B/Q6 and NF- κ B/P50 varies from one to three and from one to four, respectively, in the six alleles. This can be of primary importance, as enhancers of the two IgH3'ECs loci influence the regulation of immunoglobulin production (Denizot et al., 2001) and B lymphocytes maturation (Andersson et al., 2000), as observed also by targeted deletions of enhancer pairs in mouse (Shi and Eckhardt, 2001).

Studies of human genetics should help to understand the role of the polymorphism of the two enhancers, since at least two pathologies have been associated with the increased frequency of allele 2 of HS1,2-A (Aupetit et al., 2000; Frezza et al., 2004).

Acknowledgments

We thank Prof. Olga Rickards and Gabriella Girelli (La Sapienza University—Rome) for providing a part of the blood and DNA control samples, and Giovanna Carignani for suggestions and revision. We gratefully acknowledge for financial support MIUR (Ministero dell'Università e Ricerca Scientifica Cofin), Università di Tor Vergata, Rome, Italy, MIUR (Ministero Italiano della Università e della Ricerca; Cluster C03, Prog.

L.488/92), European Commission (INPRIMAT, QLRI-CT-2002-01325).

References

- Andersson, T., Samuelsson, A., Matthias, P., Pettersson, S., 2000. The lymphoid-specific cofactor OBF-1 is essential for the expression of a V(H) promoter/HS1,2 enhancer-linked transgene in late B cell development. *Mol. Immunol.* 37, 889–899.
- Aupetit, C., Drouet, M., Pinaud, E., Denizot, Y., Aldigier, J.C., Bridoux, F., Cogne, M., 2000. Alleles of the alpha1 immunoglobulin gene 3' enhancer control evolution of IgA nephropathy toward renal failure. *Kidney Int.* 58, 966–971.
- Bech-Hansen, N.T., Linsley, P.S., Cox, D.W., 1983. Restriction fragment length polymorphism associated with immunoglobulin C γ genes reveal linkage disequilibrium and genomic organization. *Proc. Natl. Acad. Sci. U. S. A.* 80, 6952–6956.
- Bessette, B., Guglielmi, L., Cogne, M., Denizot, Y., 2003. Nuclear factors, hs1,2 enhancer and IgA nephropathy. *Kidney Int.* 63, 767.
- Bottaro, A., Cariota, U., De Marchi, M., Carbonara, A.O., 1991. Pulsed field electrophoresis screening for immunoglobulin heavy-chain constant region (IGHC) multigene deletions and duplications. *Am. J. Hum. Genet.* 48, 745–756.
- Cabibbo, A., Grant, P.R., Helmer-Citterich, M. (Eds.), 2002. *The Internet for Cell and Molecular Biologists, Current Applications and Future Potentials.* Horizon Scientific Press, Norfolk, UK.
- Chauveau, C., Decourt, C., Cogne, M., 1998a. Insertion of the IgH locus 3' regulatory palindrome in expression vectors warrants sure and efficient expression in stable B cell transfectants. *Gene* 222, 279–285.
- Chauveau, C., Pinaud, E., Cogne, M., 1998b. Synergies between regulatory elements of the immunoglobulin heavy chain locus and its palindromic 3' locus control region. *Eur. J. Immunol.* 28, 3048–3056.
- Chen, C., Birshtein, B.K., 1997. Virtually identical enhancers containing a segment of homology to murine 3'IgH-E(hs1,2) lie downstream of human Ig C alpha 1 and C alpha 2 genes. *J. Immunol.* 159, 1310–1318.
- Conley, M.E., 2003. Genes required for B cell development. *J. Clin. Invest.* 112, 1636–1638.
- Denizot, Y., Pinaud, E., Aupetit, C., Le Morvan, C., Magnoux, E., Aldigier, J.C., Cogne, M., 2001. Polymorphism of the human alpha1 immunoglobulin gene 3' enhancer hs1,2 and its relation to gene expression. *Immunology* 103, 35–40.
- Frezza, D., Camacho-Vanegas, O., Fruscalzo, A., Favaro, M., Giorgi, S., Scotto d'Abusco, A.S., Gualandi, G., 1998. The region 3' to C α 1 gene of human IG heavy chain displays a polymorphic duplicated sequence and encodes an RNA associated with polysomes. *Gene* 219, 19–24.
- Frezza, D., Giambra, V., Cianci, R., Fruscalzo, A., Giufrè, M., Cammarota, G., Martinez-Labarga, C., Rickards, O., Scibilia, G., Sferlazzas, C., Bartolozzi, F., Magazzù, G.B., Gasbarrini, G., Pandolfi, F., 2004. Increased frequency of the immunoglobulin enhancer HS1,2 allele 2 in coeliac disease. *Scand. J. Gastroenterol.* 39, 1083–1087.
- Harindranath, N., Mills, F.C., Mitchell, M., Meindl, A., Max, E.E., 1998. The human elk-1 gene family: the functional gene and two processed pseudogenes embedded in the IgH locus. *Gene* 221, 215–224.
- Hu, Y., Pan, Q., Pardali, E., Mills, F.C., Bernstein, R.M., Max, E.E., Sideras, P., Hammarstrom, L., 2000. Regulation of germline promoters by the two human Ig heavy chain 3' alpha enhancers. *J. Immunol.* 164, 6380–6386.
- Kawamura, S., Saitou, N., Ueda, S., 1992. Concerted evolution of the primate immunoglobulin alpha-gene through gene conversion. *J. Biol. Chem.* 267, 7359–7367.
- Laurencikiene, J., Deveikaite, V., Severinson, E., 2001. HS1,2 enhancer regulation of germline epsilon and gamma2b promoters in murine B lymphocytes: evidence for specific promoter–enhancer interactions. *J. Immunol.* 167, 3257–3265.
- Manis, J.P., van der Stoep, N., Tian, M., Ferrini, R., Davidson, L., Bottaro, A., Alt, F.W., 1998. Class switching in B cells lacking 3' immunoglobulin heavy chain enhancers. *J. Exp. Med.* 188, 1421–1431.
- Michaelson, J.S., Singh, M., Snapper, C.M., Sha, W.C., Baltimore, D., Birshtein, B.K., 1996. Regulation of 3' IgH enhancers by a common set of factors, including kappa B-binding proteins. *J. Immunol.* 15, 2828–2839.
- Mills, F.C., Harindranath, N., Mitchell, M., Max, E.E., 1997. Enhancer complexes located downstream of both human immunoglobulin C α genes. *J. Exp. Med.* 186, 845–858.
- Pan, Q., Petit-Frere, C., Stavnezer, J., Hammarström, L., 2000. Regulation of the promoter for human immunoglobulin gamma3 germ-line transcription and its interaction with the 3'alpha enhancer. *Eur. J. Immunol.* 30, 1019–1029.
- Pinaud, E., Aupetit, C., Chauveau, C., Cogne, M., 1997. Identification of a homolog of the C alpha 3'/hs3 enhancer and of an allelic variant of the 3'IgH/hs1,2 enhancer downstream of the human immunoglobulin alpha 1 gene. *Eur. J. Immunol.* 27, 2981–2985.
- Pinaud, E., Khamlichi, A.A., Le Morvan, C., Drouet, M., Nalesso, V., Le Bert, M., Cogne, M., 2001. Localization of the 3' IgH locus elements that effect long-distance regulation of class switch recombination. *Immunity* 15, 187–199.
- Rabbani, H., Kondo, N., Smith, C.I.E., Hammarström, L., 1995. The influence of gene deletions and duplications within the IGHC locus on serum immunoglobulin subclass levels. *Clin. Immunol. Immunopathol.* 76, 214–218.
- Rabbani, H., Pan, Q., Kondo, N., Smith, C.I.E., Hammarström, L., 1996. Duplications and deletions of the human IGHC locus: evolutionary implications. *Immunogenetics* 45, 136–141.
- Sadhu, A., Shen, M.L., Hackbarth, M., Hume, E., McKeithan, T.W., 1997. CpG-rich sequences close to the site of duplication within the human IGH constant region. *Immunogenetics* 45, 365–370.
- Saleque, S., Singh, M., Birshtein, B.K., 1999. Ig heavy chain expression and class switching in vitro from an allele lacking the 3' enhancers Dnase I-hypersensitive hs3A and hs1,2. *J. Immunol.* 162, 2791–2803.
- Sepulveda, M.A., Emelyanov, A.V., Birshtein, B.K., 2004. NF-kappaB and Oct-2 synergize to activate the human 3' Igh hs4 enhancer in B cells. *J. Immunol.* 172, 1054–1064.
- Shi, X., Eckhardt, L.A., 2001. Deletional analyses reveal an essential role for the hs3b/hs4 IgH 3' enhancer pair in an Ig-secreting but not an earlier-stage B cell line. *Int. Immunol.* 13, 1003–1012.
- Volgina, V.V., Kingzette, M., Zhai, S.K., Knight, K.L., 2000. A single 3' alpha hs1,2 enhancer in the rabbit IgH locus. *J. Immunol.* 165, 6400–6405.
- Zelazowski, P., Shen, Y., Snapper, C.M., 2000. NF-kappaB/p50 and NF-kappaB/c-Rel differentially regulate the activity of the 3'alphaE-hs1,2 enhancer in normal murine B cells in an activation-dependent manner. *Int. Immunol.* 12, 1167–1172.