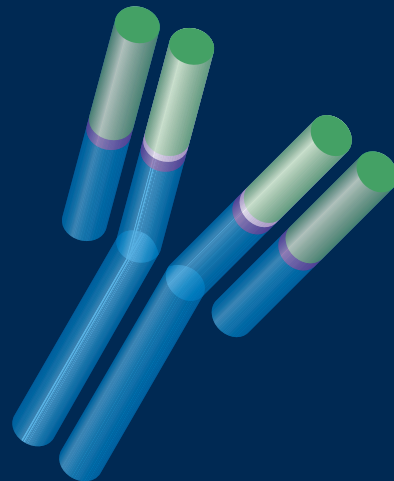
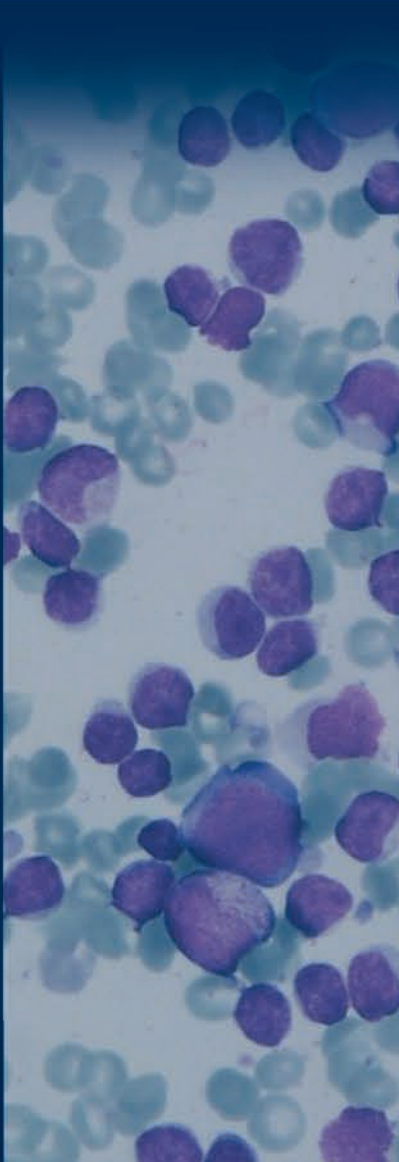


P. Ghia  
R. Rosenquist  
F. Davì

# Immunoglobulin Gene Analysis in Chronic Lymphocytic Leukemia



This publication was supported by



european research initiative on CLL

<http://www.ericll.org>

© Copyright 2009 Wolters Kluwer Health Italy Ltd

ISBN 978-88-7556-377-6

All rights reserved.

Wolters Kluwer Health Italy Ltd  
Via Lanino 5 – 20144 Milan – Italy  
Tel. +39.02.423.4562  
Fax +39.02.422.1200  
[www.wkhealth.com](http://www.wkhealth.com)

*The authors wish to thank Prof. Christer Sundström, Dept. of Pathology, Uppsala University Hospital, Sweden, for supplying the CLL image used on the cover.*

Editorial Management: Siobhan Ward  
Project Management: Rosa Parrella  
Publishing Services: Marita Bassano

This book is protected by copyright. No part of this book may be reproduced in any form by any means, including photocopying, or utilized by any information storage and retrieval system without written permission from the copyright owner, except for brief quotations embodied in critical articles and reviews.

Care has been taken to confirm the accuracy of the information presented and to describe generally accepted practices. However, the author and publisher are not responsible for errors or omissions and for any consequences from application of the information in this book and make no warranty, expressed or implied, with respect to the currency, completeness, or accuracy of the contents of the publication. Application of this information in a particular situation remains the professional responsibility of the practitioner. The author and publisher have exerted every effort to ensure that drug selection and dosage set forth in this text are in accordance with current recommendations and practice at the time of publication.

However, it is the responsibility of the health care provider to ascertain the prescribing info of each drug or device planned for use in their clinical practice.

Printed in Italy by GECA SpA – Cesano Boscone - Milan

This publication was partially funded by



Bayer HealthCare  
Bayer Schering Pharma

# Contributors

## **Chrysoula Belessi**

Hematology Department, Nikea General Hospital, Piraeus, Greece

## **Nicholas Chiorazzi**

The Feinstein Institute for Medical Research, North Shore – LIJ Health System, Manhasset, NY; Department of Medicine, North Shore University Hospital, Manhasset, NY; and Departments of Cell Biology and of Medicine, Albert Einstein School of Medicine, Bronx, NY, USA

## **Frédéric Davi**

Department of Hematology, Hôpital Pitié-Salpêtrière and University Pierre et Marie Curie, Paris, France

## **Paolo Ghia**

Laboratory and Unit of Lymphoid Malignancies, Department of Oncology, Università Vita-Salute San Raffaele and Istituto Scientifico San Raffaele, Milan, Italy

## **Véronique Giudicelli**

IMGT<sup>®</sup>, the international ImMunoGeneTics information system<sup>®</sup>, Université Montpellier 2, Institut de Génétique Humaine, UPR CNRS 1142, Montpellier, France

## **Michael Hallek**

Klinik I für Innere Medizin  
Universität zu Köln, Germany

## **Anton W. Langerak**

Department of Immunology, Erasmus MC, University Medical Center, Rotterdam, The Netherlands

## **Marie-Paule Lefranc**

IMGT<sup>®</sup>, the international ImMunoGeneTics information system<sup>®</sup>, Université Montpellier 2, Institut de Génétique Humaine, UPR CNRS 1142, Montpellier, France

**Carol Moreno**

Institute of Hematology and Oncology, Department of Hematology,  
Hospital Clinic, IDIBAPS, University of Barcelona, Spain

**Kathleen N. Potter**

Molecular Immunology Group, Cancer Sciences Division,  
Southampton University Hospitals Trust, Southampton, UK

**Richard Rosenquist**

Department of Genetics and Pathology, Uppsala University,  
Uppsala, Sweden

**Kostas Stamatopoulos**

Hematology Department and HCT Unit, G. Papanicolaou Hospital,  
Thessaloniki, Greece

**Freda K. Stevenson**

Molecular Immunology Group, Cancer Sciences Division,  
Southampton University Hospitals Trust, Southampton, UK

# Foreword

Chronic lymphocytic leukemia (CLL) represents the most frequent leukemia in the Western world. During the last decade, there has been tremendous progress in elucidating the pathogenesis of this disease. One of the most interesting features discovered during this search is the fact that the leukemia cells express immunoglobulin (IG) that may or may not have incurred somatic hypermutations of the IG heavy variable (IGHV) genes. The outcome of CLL patients with leukemia cells using an unmutated IGHV gene is inferior to those patients with leukemia cells that carry a mutated one.<sup>[1,2]</sup>

In addition to the important observations relating the IG mutation status to clinical behavior, the fact that the IG repertoire in CLL is restricted and also uniquely characterized by the existence of closely similar, stereotyped B cell receptors<sup>[3-5]</sup> implies a role for antigen(s) in leukemogenesis. Recently, it has been found that CLL IG specifically bind certain antigens,<sup>[4]</sup> including cytoskeletal proteins (vimentin, filamin B, and cofilin-1), but also phosphorylcholine-containing antigens (e.g. *Streptococcus pneumoniae* polysaccharides and oxidized low-density lipoprotein).<sup>[6]</sup> Remarkably, some of these antigens represent molecular motifs exposed on apoptotic cells/blebs and bacteria. Taken together, these data suggest that CD5<sup>+</sup> CLL B cells may derive from a cell compartment that produces 'natural antibodies', which may be instrumental in the elimination and scavenging of apoptotic cells and pathogenic bacteria. Recently, the presence of stereotyped B cell receptors has been linked to clinical features for certain subsets of patients, which implies that particular antigen-binding sites may be critical for determining clinical outcome in CLL.<sup>[3,5]</sup>

Given this dynamic and exciting field of research, this book of Drs *Ghia*, *Rosenquist* and *Davi* is a very timely and important contribution that summarizes the current state of this important research area and its clinical implications, especially in the context of clinical trials. The book results from a very fruitful cooperation of an international group of researchers that has worked together during recent years within the European Research Initiative on CLL (ERIC), also with the aim of standardizing the methodology for a reliable and reproducible assessment of the mutational status of IGHV genes in CLL. I wish this book the wide distribution that it highly deserves.

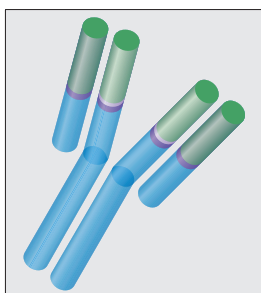
---

Dr. Michael Hallek  
Direktor  
Klinik I für Innere Medizin  
Universität zu Köln

*Michael Hallek*  
26 July, 2008

## References

1. Damle RN, Wasil T, Fais F, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* 1999; 94: 1840-7
2. Hamblin TJ, Davis Z, Gardiner A, et al. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* 1999; 94: 1848-54
3. Tobin G, Thunberg U, Johnson A, et al. Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. *Blood* 2003; 101(12): 4952-7
4. Messmer BT, Albesiano E, Efremov DG, et al. Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. *J Exp Med* 2004; 200(4): 519-25
5. Stamatopoulos K, Belessi C, Moreno C, et al. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: pathogenetic implications and clinical correlations. *Blood* 2007; 109(1): 259-70
6. Lanemo Myhrinder A, Hellqvist E, Sidorova E, et al. A new perspective: molecular motifs on oxidized LDL, apoptotic cells, and bacteria are targets for chronic lymphocytic leukemia antibodies. *Blood* 2008; 111: 3838-48



# Contents

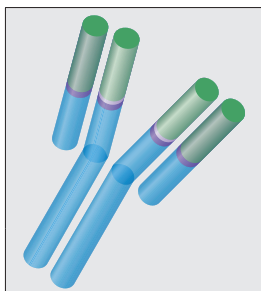
	<b>Preface</b> .....	XIII
	IG V gene analyses: clues to the development, prognosis and treatment of chronic lymphocytic leukemia <i>Nicholas Chiorazzi</i>	
	<b>Introduction</b> .....	1
	Immunoglobulin genes in B cell development <i>Frédéric Davi</i>	
<b>SECTION I</b>	<b>Immunoglobulin gene analysis: technical aspects</b>	
<b>Chapter 1</b>	From the patient to the sequence: selection of material, primers and PCR protocol, clonality analysis, and sequencing protocols <i>Anton W. Langerak</i>	23
<b>Chapter 2</b>	IMGT® standardized analysis of immunoglobulin rearranged sequences <i>Véronique Giudicelli and Marie-Paule Lefranc</i>	33
<b>Chapter 3</b>	Interpretation of ‘trouble sequences’ <i>Chrysoula Belessi</i>	53
<b>SECTION II</b>	<b>Immunoglobulin genes in chronic lymphocytic leukemia: clinical and biological significance</b>	
<b>Chapter 4</b>	Immunoglobulin genes in chronic lymphocytic leukemia: is the B cell receptor the key to understanding pathogenesis and progression? <i>Freda K. Stevenson and Kathleen N. Potter</i>	67
<b>Chapter 5</b>	IGHV gene mutational status and prognosis in chronic lymphocytic leukemia: its relationship to other prognostic markers <i>Paolo Ghia</i>	81
<b>Chapter 6</b>	Stereotyped B cell receptors in chronic lymphocytic leukemia <i>Kostas Stamatopoulos</i>	95





<b>Chapter 7</b>	How to report IG sequence data in clinical routine: ..... 113 cases difficult to categorize <i>Richard Rosenquist</i>
<b>Chapter 8</b>	Immunoglobulin genes as targets for minimal residual ..... 125 disease detection in chronic lymphocytic leukemia <i>Carol Moreno</i>
	<b>Glossary</b> ..... 135





## Preface. IG V gene analyses: clues to the development, prognosis and treatment of chronic lymphocytic leukemia

*Nicholas Chiorazzi*

Chronic lymphocytic leukemia (CLL) is a relatively common and incurable leukemia. Although its etiology remains unknown, over the past decade considerable progress has been made toward understanding the immunobiology of the leukemic cells that overgrow in the disease, and much of this progress has come from analyses of the immunoglobulin (IG) genes expressed by the cells. For instance, it is now clear that CLL patients can be divided into subgroups based on the presence or absence of IGHV somatic mutations.<sup>[1]</sup> Furthermore, IGHV mutations can distinguish between patients that follow an indolent clinical course, living at times for decades often without therapy, and others who succumb relatively rapidly, despite aggressive treatment.<sup>[2,3]</sup>

The disease manifests in clonal expansion of a B lymphocyte characteristically displaying unique phenotypic and molecular features. As with a normal B cell, the most distinctive feature of a CLL cell is the IG antigen receptor it produces, which carries a unique molecular signature determined initially by specific IGH V-D-J and IGK/IGL V-J gene rearrangements primarily occurring during development and secondarily by additional changes (e.g. somatic hypermutation and class-switch recombination) during germinal center reactions. The structure generated by these primary and secondary changes serves as an important identifier of a B cell clone, whether it is leukemic or not. In addition, it can provide clues to the development, maturation and differentiation of the B lymphocyte that gave rise to the malignant clone, as well as practical information relating to prognosis, evaluation of therapeutic efficacy, and as a potential therapeutic target.

In the Introduction to this book, *Frédéric Davi* details the IGH and IGL variable (V) gene rearrangements that create a complete IG protein. The successful accomplishment of this process is essential for the development of a B lymphocyte. Without the tonic and/or antigen-mediated signaling that B cell surface membrane immunoglobulin (mIG) provides, B cell maturation and survival ceases.<sup>[4]</sup>

The characteristics of the IG and its molecular signatures differ between CLL cells and most normal B cells as well as between patients with the indolent versus aggressive forms of the disease.<sup>[5-7]</sup> In Chapter 4, *Freda Stevenson* and *Kathleen Potter* describe these sets of differences, providing cogent arguments to suggest they are reflective of selection for binding sites that react with foreign as well as autologous epitopes, in superantigenic and classical antigen recognition manners. Such mIG-antigen binding interactions and their subsequent distinct physiologic outcomes, such as activation and

Nicholas Chiorazzi  
The Feinstein Institute for  
Medical Research  
North Shore – LIJ Health System  
350 Community Drive  
Manhasset, NY 11030  
USA  
E-mail: nchizzi@nshs.edu



clonal expansion versus anergy and death, for the leukemic clones of CLL patients with unmutated and mutated IGHV may relate to the diversity in clinical course and outcome. These features are also compared and contrasted with other 'indolent' B cell lymphomas.

In theory, the diversity of antigen-binding sites that can be created by the rearrangement process and its accompanying gene junctional variability is enormous ( $>1 \times 10^{12-14}$ ), although its full potential may not be realized due to structural constraints imposed by individual IG V genes and by antigenic selection and stimulation. For example, in pre-B cells, the heavy chain resulting from IGHV-D-J rearrangement pairs with a monomorphic structure, known as surrogate light chain, that permits the further maturation of the B cell to a stage at which a polymorphic light chain resulting from V-J rearrangement becomes associated with the heavy chain. Pairing of the surrogate light chain and a mature light chain with the heavy chain is not a random process, but rather is restricted by amino acid structure. Thus, in the normal B cell repertoire, a skewed association exists for certain IGHV with a subset of IGKV or IGLV. Additional changes in the display of IGHV and IGKV or IGLV occur among developing B lymphocytes based on the antigenic reactivities of mIG, such that IG with unacceptable levels of affinity for autoantigens are either altered by expressing and/or exchanging IGHV, IGKV or IGLV genes ('receptor editing')<sup>[8,9]</sup> or are deleted by apoptosis.<sup>[10]</sup> Finally, the repertoire is further shaped and probably narrowed somewhat by antigen exposure that selects B cells with certain antigen-binding properties for further expansion and introduces somatic changes in V gene structure.<sup>[11]</sup> Thus, the normal, mature B cell antigen-binding repertoire is genetically and environmentally biased. Nevertheless, despite these biases, the repertoire remains remarkably comprehensive ( $\sim 1 \times 10^{12}$ ), reflective of the evolutionary, adaptive mechanisms that support human health and survival.

Taking this into account, *Kostas Stamatopoulos* (Chapter 6) describes the remarkable and apparently CLL-associated similarities that exist in the structure of the antigen-binding domains of mIG in groups of patients with the disease,<sup>[12]</sup> indicating that such strikingly high frequencies cannot have arisen by chance. Rather, he posits that selection by autoantigens and foreign antigens has likely influenced the dramatic similarities in amino acid structure of the receptors between different patient groups. Finally, he provides evidence of 'evolutionary selection' for some of the stereotypic sets, suggesting there is a unique cellular origin for the disease in specific groups.

Although somatic mutations in IGHV genes are valuable indicators of patient survival in CLL, patients with certain stereotypic antigen receptors can share clinical, phenotypic features and follow either indolent or aggressive courses, regardless of their mutation status.<sup>[13]</sup> This suggests that a particular antigen-binding site can be critical in determining clinical presentation and possibly even prognosis. In Chapter 5, *Paolo Ghia* reviews these findings and relates them to other prognostic markers currently in use. Though several additional molecules have been described and repeatedly shown to carry a strong prognostic power, they are all independent of each other and especially of IGHV gene somatic hypermutations. As elaborated by *Ghia*, the challenge now is to prospectively assess a clear hierarchy in terms of

prognostic capability when using several of them in the clinical setting and with individual patients.

In a related vein, *Richard Rosenquist* (Chapter 7) describes approaches to interpreting IG gene sequence data in the clinical setting and gives practical examples on how to report these data in a consistent way. Furthermore, ‘problematic’ cases, in which categorization into ‘mutated’ and ‘unmutated’ is not straightforward, are discussed. A newly established online support and trouble-shooting forum for interpretation of IGHV gene sequence results in CLL is also provided.

The junctional regions of clonal IG and T cell receptor (TR) gene rearrangements are fingerprint-like sequences for each lymphoid malignancy and may serve as markers for clonality. Molecular clonality studies based on the detection of rearranged IG/TR genes can be highly informative for determining the extent of residual disease and the level of clonal escape that can occur after treatment. However, the very nature of this marker is the reason for several limitations and pitfalls that might hamper interpretation of the results. In Chapter 8, *Carol Moreno* discusses several alternative IG-based PCR protocols to follow the level of clonal B cells after therapy. She also offers a brief overview on minimal residual disease (MRD) assessment by multicolor flow cytometry and emphasizing that the benefits of achieving MRD-negative status in patients with CLL require further investigation in large well-controlled trials.

None of the clues provided by IG V genes in CLL would be possible without robust methods to identify germline genes (especially challenging when highly similar), how they join, and how they may have been altered by somatic mutation. Based on the current and most updated literature as well as his laboratory experience on this issue, *Anton Langerak* (Chapter 1) discusses the pros and cons of several technical aspects for reliable and reproducible immunoglobulin gene analysis in CLL, from identification of the clonotypic IGHV-IGHD-IGHJ rearrangement to determining its nucleotide sequence.

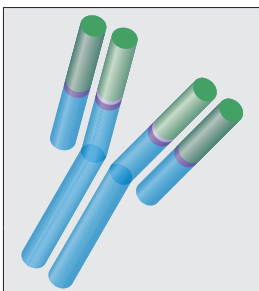
These studies require large, well-characterized and annotated databases of germline and rearranged IG V sequences and standardized tools. The creation of IMGT<sup>®</sup>, the international ImMunoGeneTics information system<sup>®</sup> for IG and TR [<http://imgt.cines.fr>], that is the global reference in immunoinformatics,<sup>[14]</sup> has been the lifelong work of *Marie-Paule Lefranc*, to whom we all are grateful and indebted. In Chapter 2, *Véronique Giudicelli* and *Marie-Paule Lefranc* describe the latest version of the IMGT<sup>®</sup> sequence analytic tools available through this effort. Finally, *Chryssoula Belessi* (Chapter 3) provides suggestions on how to interpret IG sequence data that differ from those routinely encountered – for instance, rearrangements with absent CDR3 anchor ‘landmarks’ or rearrangements rendered unproductive for several reasons such as use of IGHV pseudogenes, out-of-frame junctions, and aberrant somatic hypermutation.

In conclusion, it is essential to note that several of the key advances in the role of IG genes in CLL have come about through the collaboration of a group of European investigators, sponsored in part by the European Research Initiative on CLL (ERIC). I would personally like to applaud the efforts of *Chryssoula Belessi*, *Frédéric Davi*, *Paolo Ghia*, *Carol Moreno*, *Richard*

*Rosenquist*, and *Kostas Stamatopoulos* (listed alphabetically) and *Marie-Paule Lefranc* and her IMGT team for their collegial and unselfish efforts in moving this field forward.

## References

1. Fais F, Ghiotto F, Hashimoto S, et al. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J Clin Invest* 1998; 102: 1515-25
2. Damle RN, Wasil T, Fais F, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* 1999; 94: 1840-7
3. Hamblin TJ, Davis Z, Gardiner A, et al. Unmutated Ig VH genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* 1999; 94: 1848-54
4. Tze LE, Schram BR, Lam KP, et al. Basal immunoglobulin signaling actively maintains developmental stage in immature B cells. *PLoS Biol* 2005; 3: e82
5. Chiorazzi N, Ferrarini M. B cell chronic lymphocytic leukemia: lessons learned from studies of the B cell antigen receptor. *Ann Rev Immunol* 2003; 21: 841-94
6. Stevenson FK, Caligaris-Cappio F. Chronic lymphocytic leukemia: revelations from the B-cell receptor. *Blood* 2004; 103: 4389-95
7. Chiorazzi N, Rai KR, Ferrarini M. Chronic lymphocytic leukemia. *N Engl J Med* 2005; 352: 804-15
8. Tiegs SL, Russell DM, Nemazee D. Receptor editing in self-reactive bone marrow B cells. *J Exp Med* 1993; 177: 1009-20
9. Gay D, Saunders T, Camper S, et al. Receptor editing: an approach by autoreactive B cells to escape tolerance. *J Exp Med* 1993; 177: 999-1008
10. Hartley SB, Crosbie J, Brink R, et al. Elimination from peripheral lymphoid tissues of self-reactive B lymphocytes recognizing membrane-bound antigens. *Nature* 1991; 353: 765-9
11. Klein U, Dalla-Favera R. Germinal centres: role in B-cell physiology and malignancy. *Nat Rev Immunol* 2008; 8: 22-33
12. Murray F, Darzentas N, Hadzidimitriou A, et al. Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia: implications for the role of antigen selection in leukemogenesis. *Blood* 2008; 111: 1524-33
13. Tobin G, Thunberg U, Johnson A, et al. Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. *Blood* 2003; 101: 4952-7
14. Lefranc M-P. WHO-IUIS Nomenclature Subcommittee for Immunoglobulins and T cell receptors report August 2007, 13th International Congress of Immunology, Rio de Janeiro, Brazil. *Dev Comp Immunol* 2008; 32: 461-3



## Introduction. Immunoglobulin genes in B cell development

*Frédéric Davi*

The discovery of the mechanisms responsible for the generation of diversity in antigen receptors has been a major scientific achievement, acknowledged by a Nobel prize in 1987. Since then, it has been a very active field of research for immunologists. The hematological community has also benefited from this knowledge, since, for instance, it has greatly helped to classify the origin of lymphoid malignancies along the B cell differentiation pathway. From a more clinical perspective, immunoglobulin (IG) gene analysis has provided useful tools to determine the clonality of lymphoid proliferations. Recently, the analysis of IG gene sequences has proved to provide essential information regarding the prognosis of chronic lymphocytic leukemia (CLL). However, its wide use as a laboratory routine test requires a basic knowledge of this complex biological system, which constitutes the aim of this article. A selection of reviews that addresses these issues in greater detail is provided at the end.

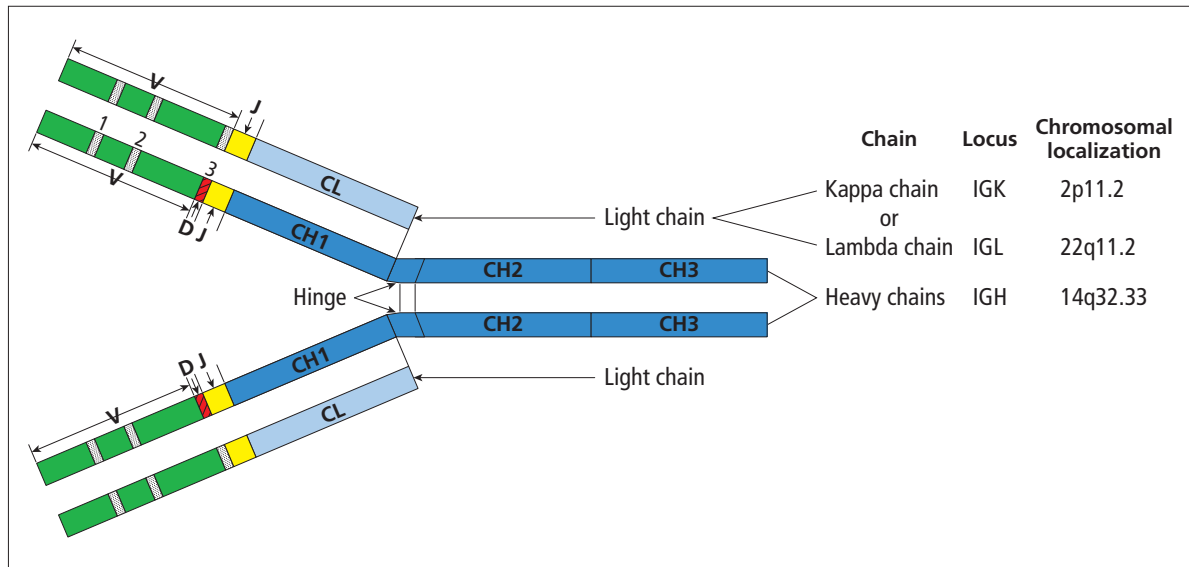
### General overview

The recognition of a vast array of foreign antigens is a cardinal feature of the immune system.<sup>[1]</sup> This is mediated by specialized polymorphic membrane receptors on the surface of B and T lymphocytes, the B cell receptor (BcR) and the T cell receptor (TcR), respectively. The BcR comprises 2 types of components: the recognition unit constituted by a membrane IG, and the signal transmission unit comprising the coreceptors CD79A (Ig-alpha, mb-1) and CD79B (Ig-beta, B29).

An IG is a heterodimer composed of 2 heavy chains and 2 light chains bound by disulfide bridges (Figure 1). Each chain has 2 distinct parts: the N-terminal one, responsible for antigen recognition, is the variable (V) domain. The C-terminal one is the constant (C) region, which for heavy chains, specifies the isotype of the IG and exerts various effector functions. The V domain of heavy chains, or VH, is encoded by 3 types of genes called variable (V), diversity (D), and joining (J), while the V domain of light chains, or VL, is encoded by only V and J genes. The V domain contains 4 relatively conserved framework regions (FR) that maintain the structure of the domain. In between are 3 highly variable stretches which form loops that interact directly with antigens and thus are called complementarity determining regions (CDR). The CDR3 is at the junction of the V, (D) and J genes and has the highest variability.

Frédéric Davi, MD, PhD  
Department of Hematology  
Hôpital Pitié-Salpêtrière  
47-87 Bd de l'Hôpital  
75013 Paris  
France  
E-mail: fred.davi@psl.aphp.fr





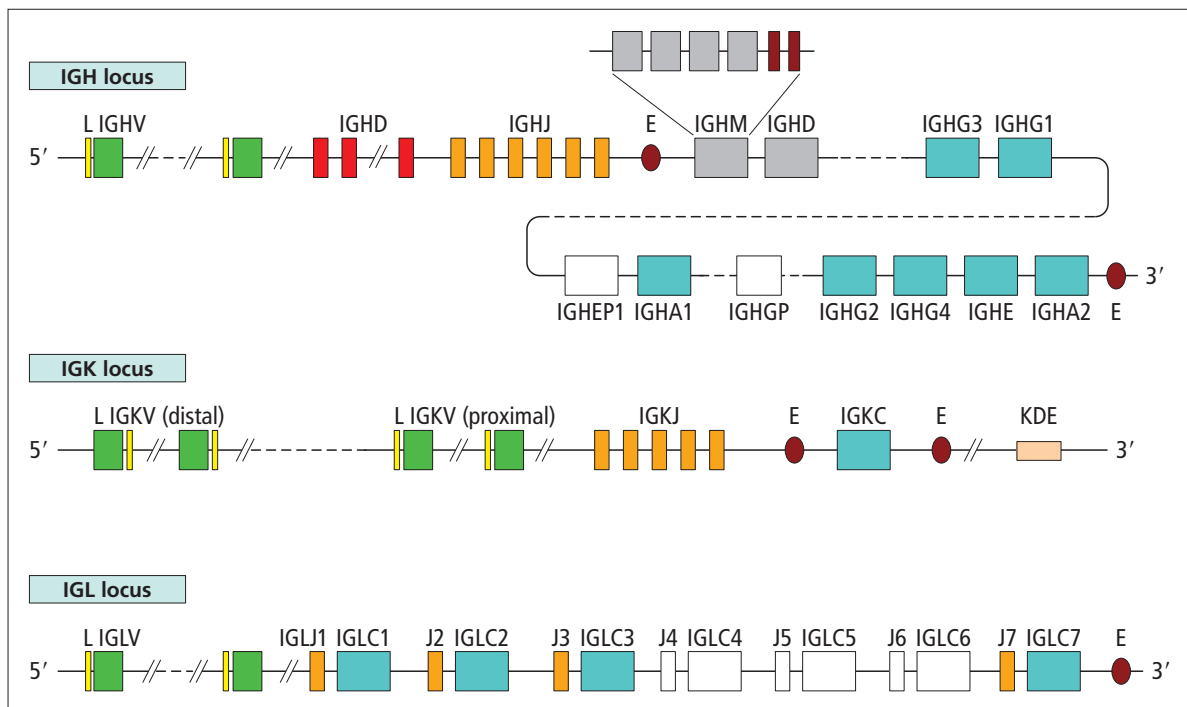
**Figure 1** Schematic representation of an IG molecule. Reproduced with the kind authorization of Marie-Paule Lefranc (IMGT®, the international ImMunoGeneTics information system®; <http://imgt.cines.fr>).

B cell differentiation is a complex process that occurs in two parts. The first takes place in the bone marrow (and the fetal liver), and is antigen-independent. It consists in the assembly, by DNA rearrangement, of the V, D and J genes encoding the variable domain, and leads to the production of B cells expressing a surface membrane-bound IG. The second part occurs in the periphery, within the secondary lymphoid organs, where naive B cells are activated upon antigen encounter, are further diversified by introduction of somatic hypermutations (SHM) in their IG V domains, and undergo class-switch recombination (CSR), in which the C region is replaced by another isotype.<sup>[2]</sup> This results ultimately in the production of high-affinity, switched-memory B cells or differentiation into plasma cells which secrete their IG.

## Genomic organization

### *IGH locus*

Most of the genes coding for IGH are located on chromosome 14q32.33.<sup>[3]</sup> They span about 125 Mb and are organized in 4 clusters V, D, J and C in a 5'-3' orientation (Figure 2). The IGHV genes are close to the telomeric end of the chromosome while the IGHC genes are centromeric. In addition, 35 genes, including IGHV and IGHD, have been found on chromosomes 15 and 16. They have been termed orphans and do not contribute to the IGH repertoire. On chromosome 14, there are 123-129 IGHV genes, depending on the haplotype, of which 38-46 are functional (Tables I, II). Based on nucleotide sequence homology, these functional genes have been divided into 6 to 7 subgroups depending on the haplotype. These subgroups have been classified in clans, with clan I comprising the IGHV1, IGHV5, and IGHV7 subgroups, clan II comprising the IGHV2, IGHV4 and IGHV6 subgroups, while the



**Figure 2** Germline organization of the IG loci. Non-functional genes (pseudogenes or ORF) are represented in white boxes [L = leader; E = enhancer; KDE = kappa deleting element].

IGHV3 subgroup constitutes clan III. There are polymorphic variants of the IGH genes, both in terms of copy number and structure. For instance, an insertion of a 50kb DNA fragment containing 5 IGHV genes is observed in 45% of caucasians. Allelic variants differing by nucleotide substitutions have been described for many IGHV genes, with some genes (e.g. IGHV1-69 and IGHV4-34) having up to 13 alleles (IMGT/GENE-DB; <http://imgt.cines.fr>). Besides these functional genes, the IGH locus also contains open reading frames (ORF) for 4 potential genes which, however, have not yet been found to be transcribed, and 79 pseudogenes due to deleterious point mutations or truncations. Overall, the IGHV genes span over 900 kb of DNA. Furthermore, there are 27 IGHD genes, of which 23 are functional.

They are organized in 4 tandem copies of a 9 kb DNA fragment, and one unique IGHD gene (IGHD7-27) located at about 100 bp 5' of the IGHJ genes. The 9 IGHJ genes are localized on an 8 kb DNA stretch, and include 6 functional genes. As for IGHV genes, there are allelic variants of IGHD and IGHJ genes. Finally, the IGHC locus comprises 11 genes on a 300 kb DNA fragment, of which 2 are pseudogenes.

### IGK locus

The IGK locus is located on the short arm of chromosome 2 (2p11.2) where it spans 1.8 Mb. The IGKV genes have been duplicated and are distributed in 2 clusters separated by 800 kb: an IGKC-proximal cluster, which comprises 40 genes on 600 kb, and an IGKC-distal cluster, the most 5' and centromeric,

**Table I** Number of IG genes

<i>Locus</i>	<i>Chromosome</i>	<i>V</i>	<i>D</i>	<i>J</i>	<i>C</i>	<i>Orphans</i>
IGH	14q32.33	123-129 <sup>a</sup>	27	9	11	35
IGK	2p11.2	76 <sup>b</sup>	0	5	1	25
IGL	22q11.2	73-74 <sup>c</sup>	0	7-711 <sup>d</sup>	7-711 <sup>d</sup>	7

<sup>a</sup> Variable number due to allelic polymorphism by insertion/deletion of 6 genes (among which 5 are on the same 50 kb DNA fragment).

<sup>b</sup> Including 40 genes on the IGKC-proximal cluster and 36 genes on the IGKC-distal cluster.

<sup>c</sup> Allelic polymorphism by insertion/deletion of 1 gene.

<sup>d</sup> Variable number due to allelic polymorphism by amplification of the IGLJ/C2-IGLJ/C3 region resulting in 1 to 4 IGLJ/IGLC genes. Numbers are from the Immunoglobulin FactsBook.<sup>[3]</sup> Numbers of the IGL orphans (one additional gene) and functional IGKV genes (previous ORF now considered as functional) have been updated with the IMGT Repertoire (<http://imgt.cines.fr>).

**Table II** Number of functional IG genes

<i>Locus</i>	<i>V</i>	<i>V subgroups</i>	<i>D</i>	<i>J</i>	<i>C</i>
IGH	38-46 <sup>a</sup>	6-7 <sup>b</sup>	23	6	9
IGK	34-38 <sup>c</sup>	5	0	5	1
IGL	29-33 <sup>d</sup>	10	0	4-5 <sup>e</sup>	4-5 <sup>e</sup>

<sup>a</sup> Variable number due to: (a) allelic polymorphism by insertion/deletion of 6 genes; (b) 2 genes being functional or ORF or pseudogenes.

<sup>b</sup> Allelic polymorphism by insertion/deletion of 1 gene.

<sup>c</sup> Variable number due to 4 genes being functional or pseudogenes.

<sup>d</sup> Variable number due to: (a) allelic polymorphism by insertion/deletion of 1 gene; (b) 3 genes being functional or pseudogenes.

<sup>e</sup> Variable number due to 1 gene being functional or pseudogene. In addition, 1 to 4 additional IGLJ/IGLC genes can be present due to allelic polymorphism.

Numbers are from the Immunoglobulin FactsBook.<sup>[3]</sup>

composed of 36 genes on 400 kb. The 2 clusters are oriented in opposite direction, which determines their mode of rearrangement on IGKJ genes (see below). Individuals with a haplotype lacking the IGKC-distal cluster have been occasionally described. Of the 76 IGKV genes, 34-38 are functional: 4 genes may be functional or pseudogenes depending on the alleles. They belong to 5 subgroups, and polymorphic variants exist, although to a lesser extent than for IGHV genes. There are also 25 orphans located on other regions of chromosome 2 and on other chromosomes (1, 15, 22). All 5 IGKJ genes are functional and localized 5' of a single IGKC gene (Figure 2; Tables I, II).

### ***IGL locus***

The IGL locus spans 1 Mb on chromosome 22q11.2, with the IGLV genes being closer to the centromere. There are 73-74 IGLV genes over 900 kb, among which 29-33 are functional and belong to 10 subgroups (Tables I, II). The IGLJ and IGLC genes are organized in tandem with an IGLJ gene preceding an IGLC gene (Figure 2). Typically, there are 7 IGLJ/C genes, of which 4-5 are functional and encode the four IGL isotypes. There is, however, a polymorphic variation in the number of IGLJ/C genes, since some individuals may carry up to 11 of them, due to an amplification of the IGLJ/C2 - IGLJ/C3 region. There are also 7 orphans on chromosomes 8 and 22 (outside the main IGL locus).

## Generation of the primary B cell repertoire

### Assembly of IG genes

The IG synthesis requires the assembly of the different genes encoding for the V domains of the heavy and light chains (Figure 3). This is accomplished by unique somatic DNA rearrangement events: D-J, then V-D-J for VH, and V-J for VL. These V-(D)-J rearrangements operate in developing lymphocytes within the bone marrow and are tightly regulated.<sup>[1,2,4,5]</sup>

### Mechanisms of V-(D)-J rearrangement

1. **General features:** Rearrangement of the V, D and J genes which constitute the V domain is made possible by the presence of specific DNA sequences, called recombination signal (RS) sequences. They are located at the very 3' end of the V genes, at the 5' end of J genes and flank both sides of D genes

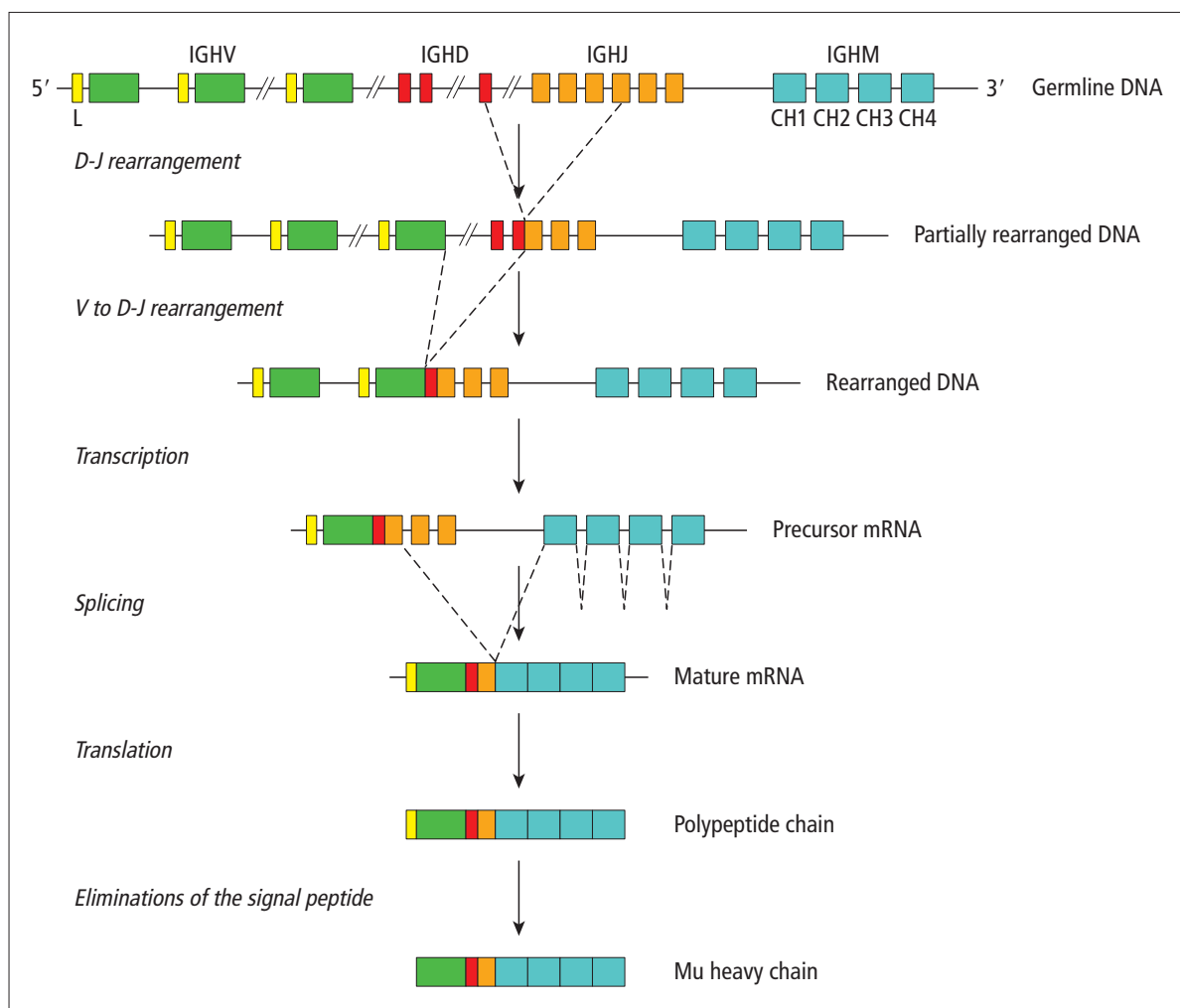
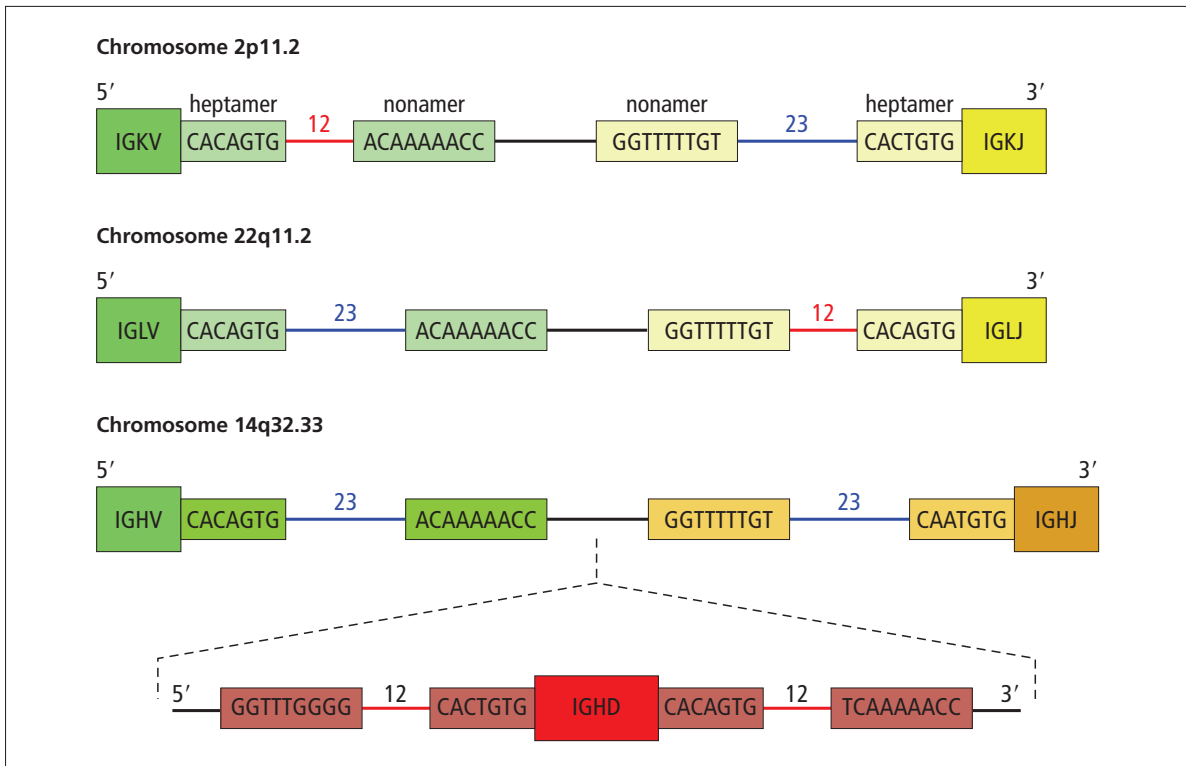


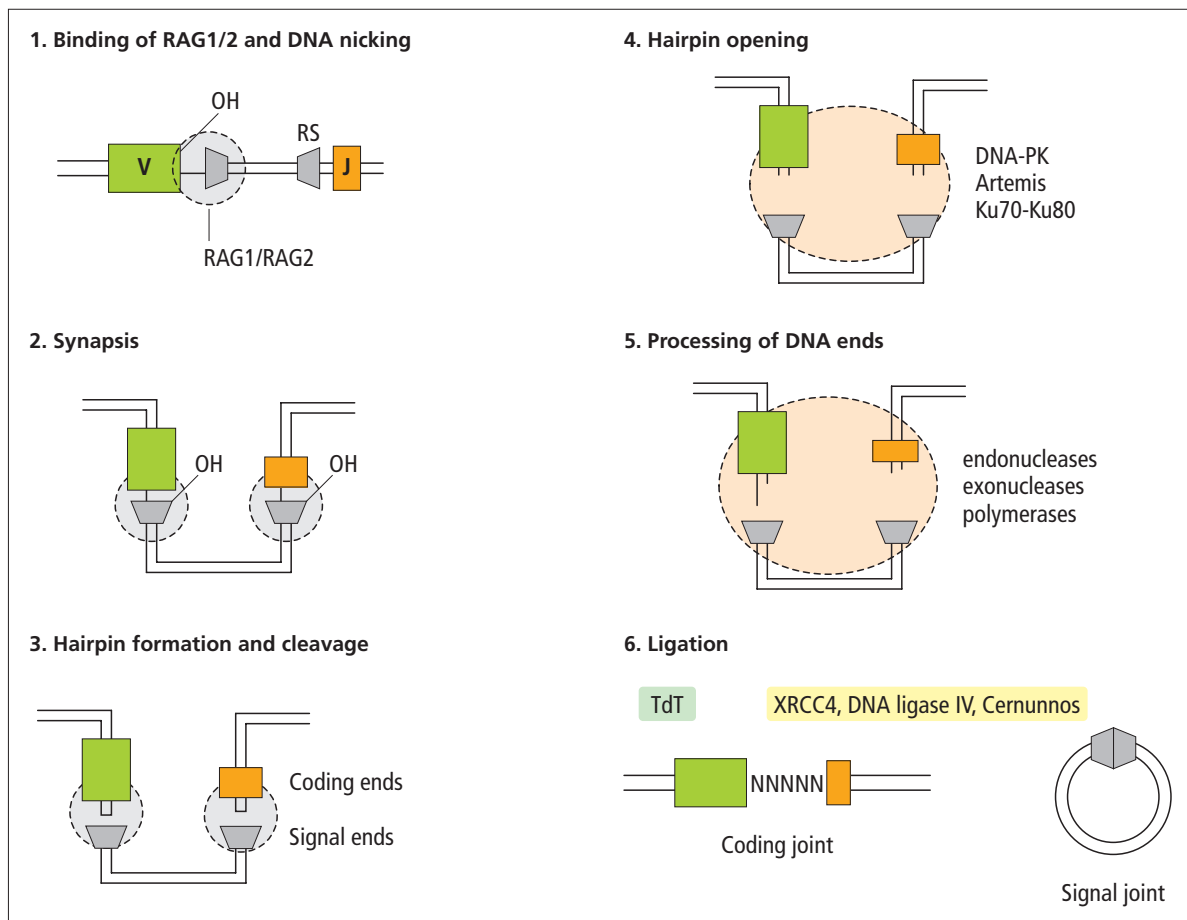
Figure 3 Synthesis of a Mu heavy chain.



**Figure 4** Structure of the recombination signals. Reproduced with the kind authorization of Marie-Paule Lefranc (IMGT®, the international ImMunoGeneTics information system®, <http://imgt.cines.fr>).

(Figure 4). They consist of 2 highly conserved motifs, a heptamer and a nonamer, separated by a non-conserved stretch of 12 or 23 nucleotides. The consensus sequence for the heptamer is 'cacagtg' sequence, while the nonamer is an 'a-t' rich motif (typically 'acaaaacc') [see section on IMGT repertoire (<http://imgt.cines.fr>) for RS sequence logos]. The sequence of spacers is variable; however, their size, which corresponds to 1 or 2 turns of the DNA helix, is fixed. In addition, their localization dictates the so-called 12/23 rule. That is, a recombination in principal only occurs between a gene flanked by a 12 nucleotide-long spacer and its partner flanked by a 23 nucleotide-long spacer. This ensures that the recombination is restricted to the appropriate genes. Thus, IGKV and IGLV genes can rearrange only with J genes, while direct rearrangement to IGHJ genes is not possible for IGHV genes. In fact, at the IGH locus, a D gene has to be rearranged to a J gene (which occurs first), and then a V gene to the partially rearranged D-J gene.

The V-(D)-J rearrangement is mediated by several proteins. Some are lymphoid-specific such as the recombination activating gene 1 and 2 (RAG1 and RAG2), which form a complex that plays a major role in this process. They are expressed only in developing lymphocytes during specific stages. The terminal deoxynucleotidyl transferase (TdT) is also a lymphoid-specific enzyme, which operates after the RAG proteins, and edits the coding junctions. The other enzymes acting during V-(D)-J rearrangement belong

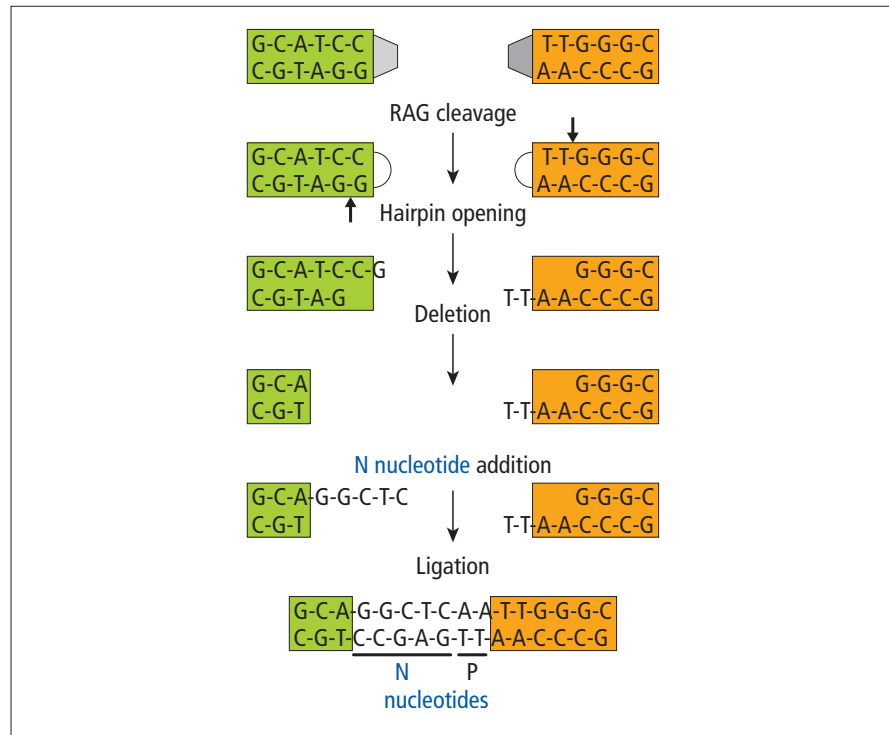


**Figure 5** Sequence of molecular events during V-(D)-J rearrangement [RAG = recombination activating gene; RS = recombination signal; DNA-PK = DNA-dependent protein kinase; TdT = terminal deoxynucleotidyl transferase; XRCC4 = X-ray repair cross-complementing protein 4].

to the general non-homologous-end-joining (NHEJ) machinery, and are recruited to repair the DNA double-strand breaks generated by RAG.

2. **Sequence of events:** The V-(D)-J rearrangement (Figure 5) is a finely orchestrated process which can be divided into 4 steps as follows:
  - Formation of a precleavage complex (synapsis): The RAG1-RAG2 complex recognizes and binds to RS sequences on pairs of genes which have become accessible (see below). These are brought together, thereby creating a DNA loop, and held within the complex.
  - Creation of double-strand breaks (cleavage): RAG1 introduces a single strand nick at the border between the coding end and the heptamer of the RS. This generates a 3' OH group which attacks the phosphodiester bond on the complementary DNA strand, resulting in the formation of a covalently sealed hairpin on the coding ends, and a 5' phosphorylated blunt end on the RS.
  - Processing of the coding ends (Figure 6): The signal ends of both RS can be ligated directly, without any further modification. By contrast,

**Figure 6** Junctional diversity created by the processing of coding ends.



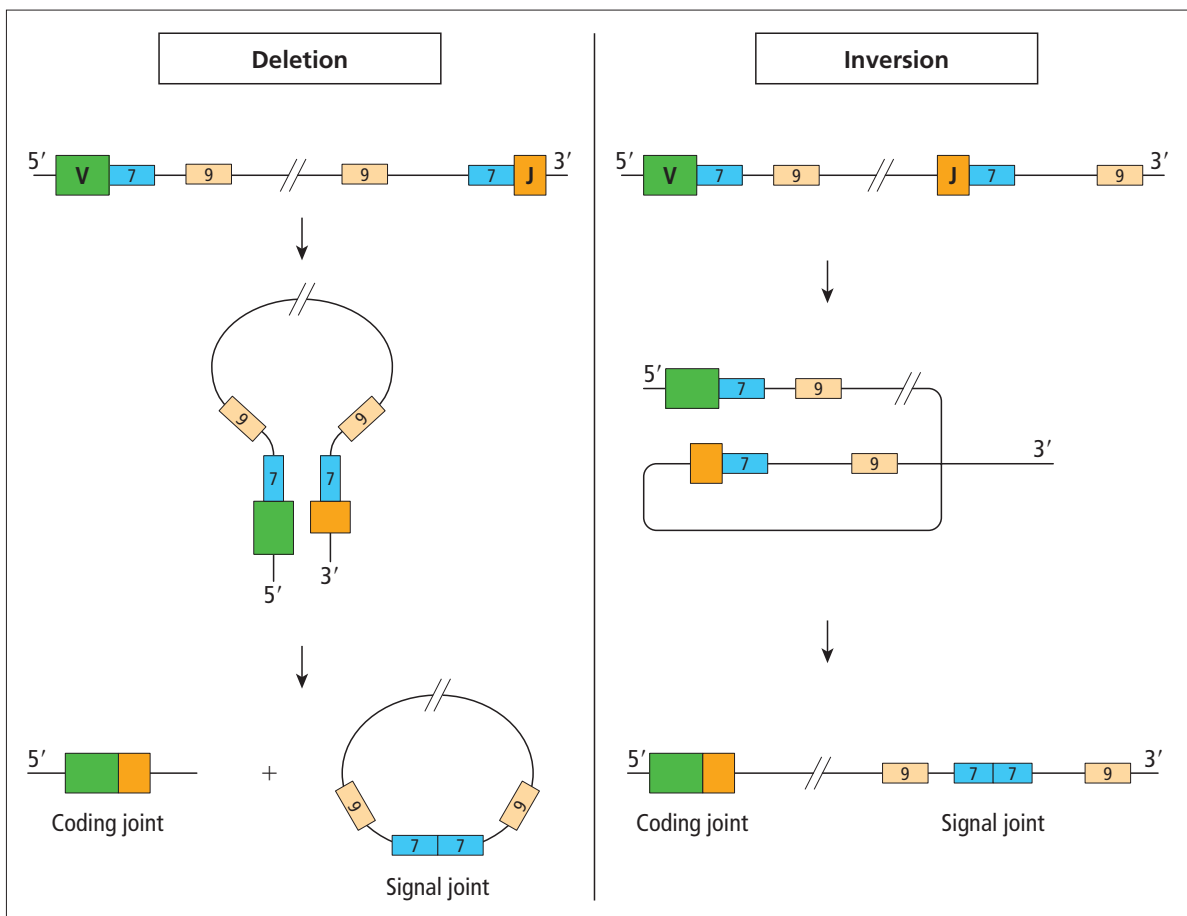
coding ends are submitted to an extensive modification which largely contributes to repertoire diversification. These include: (a) nucleotide trimming ('nibbling') which can be more or less extensive; (b) addition of palindromic (P) nucleotide complementary to an overhanging tail produced when hairpins are cleaved away from the apex; and (c) addition of non-templated nucleotides (N) by the TdT enzyme.

- Finally, the strands are religated, generating coding joints and signal joints.

The RAG1-RAG2 complex is responsible for recognition of the RS and initial cleavage of the DNA. Subsequent steps are controlled by the ubiquitous DNA-repair proteins of the NHEJ pathway, and are also facilitated by the RAG1-RAG2 proteins which remain associated with DNA ends and help target the DNA double-strand breaks. These are recognized and bound by the Ku70 and Ku80 proteins, which also recruit the DNA-dependent protein kinase (DNA-PK) catalytic subunit. The latter activates the Artemis protein for opening the hairpin ends. Other enzymatic activities during processing of coding ends includes unidentified exonucleases, DNA polymerases, and probably other proteins. The insertion of N nucleotides is dependent on the lymphocyte-specific TdT. Its preferential expression at early stages of B cell development explains why it is more active for H rather than L chains. TdT is also inactive during fetal stages, and thus the fetal repertoire lacks N Junctions. Regarding template-dependent DNA synthesis, it seems that the polymerases Pol $\mu$  and Pol $\lambda$  can perform gap-filling synthesis of DNA ends. Finally, the broken DNA strands are brought together and are religated, independently

on coding joints and signal joints, by an enzymatic complex including the X-ray repair cross-complementing protein 4 (XRCC4), the DNA ligase IV, and the newly discovered Cernunnos protein (Figure 5).

- Modes of rearrangement:** Depending on the orientation of the coding genes and their RS, the V-(D)-J rearrangement can occur through 2 main types (Figure 7). The most common is when RS are 'face to face' which results in the deletion of the intervening DNA sequence between the rearranging genes. The excision circle may remain in the cell, but is then lost by dilution after cell division. Alternatively, if RS are in the same orientation, the V-(D)-J rearrangement proceeds by inversion, and the intervening sequence and signal joints remain on the chromosome. This is the case for the IGKV genes of the distal cluster. A third option is when RS are in different orientations but not facing each other; though infrequent, such rearrangements have been described in the IGK locus. Other modes of recombination, such as gene conversion or trans-rearrangement between 2 chromosomes, have been described but play little, if any, role in generation of the human B cell repertoire.



**Figure 7** Types of V-(D)-J rearrangement. Depending on the orientation of the V and J genes and their recombination sequences, rearrangement occurs by deletion (most frequently) or by inversion.

*Final steps of IG synthesis*

DNA rearrangement allows the V-(D)-J genes to be transcribed at high levels in developing precursor B cells. This is due to the fact that the V gene promoter, immediately upstream of the V gene, is brought into the proximity of active regulators, the IG enhancers, located within the J-C introns as well as 3' of the C genes. Of note, the same phenomenon is involved when non-Ig genes, such as BCL2, are aberrantly relocated to the IG locus due to chromosomal translocations. This results in the deregulated and inappropriately high levels of expression of these oncogenes, contributing to the development of lymphoid neoplasia.

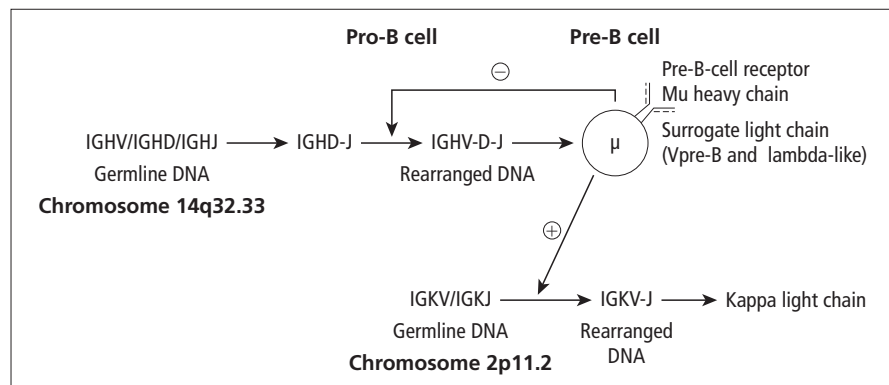
Transcription of the rearranged IG loci will result in the production of a primary transcript which also includes the C region (Figure 3). Intronic sequences are removed from the mature transcript. The leader intron is spliced by joining the leader first exon (L-PART1) to L-PART2 in 5' of the rearranged V gene. The splicing also involves joining the different exons of the IGHC gene that encodes the C region of the heavy chains (there is only a single exon for the IGKC and IGLC genes that encode the C region of the light chains). For IgM<sup>+</sup>IgD<sup>+</sup> B cells, both IGHM (C $\mu$ ) and IGHD (C $\delta$ ) are parts of the primary transcript. However, alternative splicing produces mature mRNA containing the same 5' V-D-J sequence but joined to either an IGHM or IGHD region. After translation of the mRNA, the signal peptide is cleaved off the polypeptide chain in the endoplasmic reticulum and mature IG chains are produced.

*Temporal regulation of V-(D)-J rearrangements*

Rearrangements of the IG genes proceed in the bone marrow along a characteristic developmental pathway.<sup>[1,2]</sup> In the vast majority of cases, although this is not absolute, IGH genes rearrange before light chain genes, with D-J rearrangements preceding V-D-J, and the IGK locus rearranges before the IGL locus.<sup>[6,7]</sup> Thus, it is possible to distinguish different steps of B cell differentiation upon the expression of cell surface markers and the pattern of IG gene rearrangements (and expression). V-(D)-J rearrangements begin in a committed B cell progenitor, called the pro-B cell, upon expression of RAG proteins. At a locus which is initially in a 'germline' configuration, this brings together an IGHD gene and an IGHJ gene, and it often occurs on both alleles. Thereafter, an IGHV gene is joined to the partial D-J rearrangement. Although IGHD genes generally have 2 reading frames, due to the extensive modification of the V-D-J junction, there is roughly 1 chance in 3 that the reading frames of V, D and J are properly conserved. In the case of unsuccessful rearrangement, another attempt takes place on the second allele. Failure to produce a functional V-D-J rearrangement leads to cell death by apoptosis, and these cells are cleared by resident macrophages. Thus, there is a considerable loss of precursor B cells in the bone marrow, although 'salvage' mechanisms by secondary rearrangements exist.

Alternatively, a productive IGHV-D-J rearrangement will lead to the synthesis of an intracellular  $\mu$  chain. At this stage, the  $\mu$  protein associates

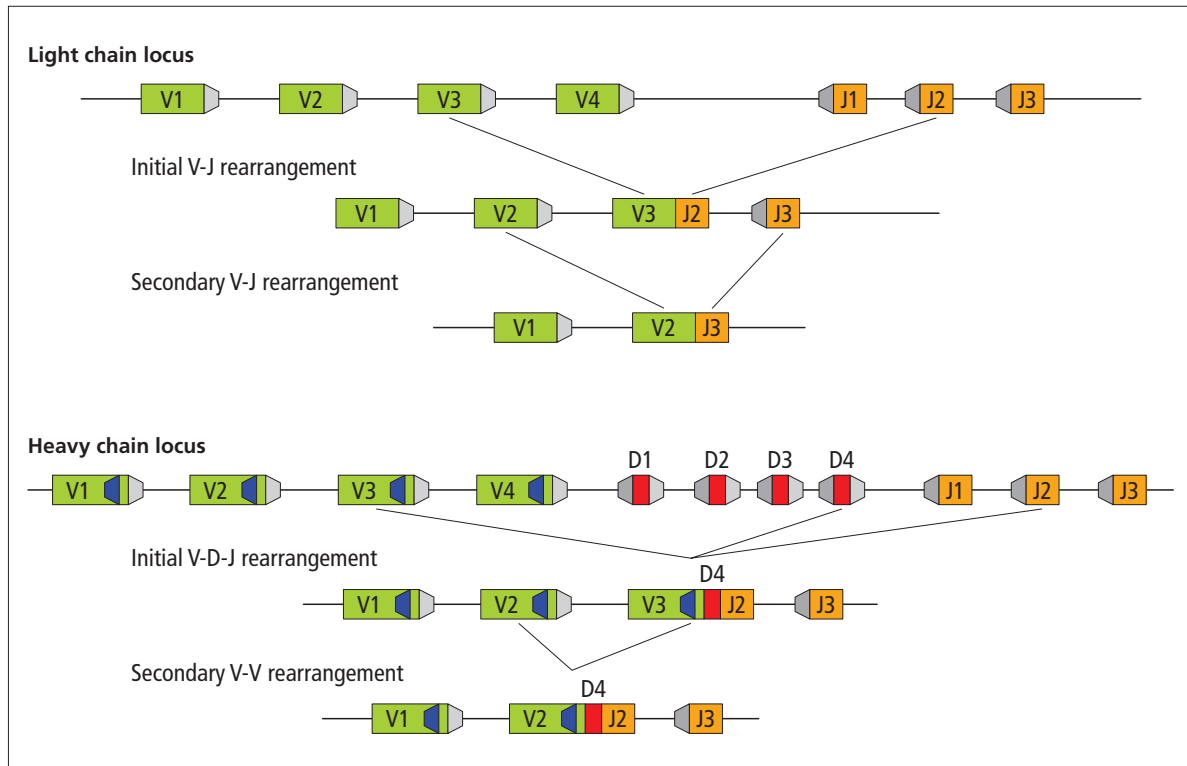
**Figure 8** Regulation of IG gene rearrangements during pre-B cell development. Reproduced with the kind authorization of Marie-Paule Lefranc (IMGT®, the international ImMunoGeneTics information system®; <http://imgt.cines.fr>).



with a surrogate light chain, composed of the products of the invariant and non-rearranging Vpre-B and lambda-like genes. These encode for non-covalently-associated Ig superfamily proteins with a V-like and a C-like domain respectively, which are attached to the  $\mu$  chain by a disulfide bridge. The capacity of the newly synthesized H chain to pair with the surrogate light chain represents a first level of receptor selection. Complexes formed by the  $\mu$  chain and surrogate light chain, associated with the signal transducing CD79A/CD79B heterodimer, are expressed at low levels at the cell surface, and constitute the pre-BcR. This pre-BcR has a crucial role in early B cell development (Figure 8). First, as indicated by its name, it promotes transition from the pro-B to the pre-B stage and allows cell survival. It favors a considerable proliferation of the pre-B cells and thus an expansion of this compartment. It also shuts off RAG expression, thereby inhibiting further V to D-J rearrangement at the second allele, a phenomenon called allelic exclusion. This ensures that a B cell will produce a single H chain, and thus a single antigen specificity, in concordance with the clonal selection theory which stipulates that an individual lymphocyte expresses a unique antigen receptor. However, cells expressing two distinct H chains (failure of allelic exclusion or 'allelic inclusion') can be observed, albeit at a very low frequency. The pre-BcR also down-regulates surrogate light chain transcription, and stimulates rearrangement at the IGK locus. Therefore, after the proliferative phase, the cell cycle stops, RAG proteins are re-expressed and rearrangement at the light chain loci occurs.

A productive IGK rearrangement will eventually be translated to a kappa light chain that will associate with the preformed H chain to constitute an IgM. In addition, as for the H chain, production of a kappa chain will inhibit further rearrangement at the second IGK allele (allelic exclusion) and at the IGL locus (isotypic exclusion). There again, this ensures that only a single light chain is expressed. Conversely, if the first IGK rearrangement is unproductive, an attempt is made on the second allele, and if unsuccessful, the IGL locus will undergo V-J rearrangement. If no productive light chain rearrangement is made, the cell will die unless it is saved by secondary rearrangements.

The IgM associates with the CD79A/CD79B coreceptors, and is expressed at the cell surface as the BcR. In the bone marrow, these immature B cells are submitted to a negative selection process against self-reactivity which forms



**Figure 9** Light chain editing and IGHV replacement.

the basis of central immune tolerance.<sup>[8]</sup> Thus, B cells expressing a BcR which recognize self-antigens with high affinity are deleted. Alternatively, they can escape this fate by undergoing a change in their antigen specificity, a process termed receptor editing (Figure 9). This occurs through reactivation of RAG and secondary rearrangements at the light chain loci, and to a lesser extent the IGH locus. The genomic organization of the light chain genes, especially the IGK locus, allows an upstream V gene recombining to a downstream J gene and deleting the initial V-J joint. Another possibility for light chain editing is switching from kappa to lambda expression. This may be associated with partial deletion of the IGK locus by rearrangement of a DNA sequence (located about 20 kb downstream of the IGKC gene and called kappa deleting element) to either an isolated RS in the IGKJ-IGKC intronic region, or to a 5' IGKV gene. Organization of the IGH locus does not permit such a change of the whole V domain, but a cryptic heptamer embedded in the 3' part of most IGHV genes allows the rearrangement of an upstream IGHV gene, a process called IGHV replacement. The upcoming new IGHV gene then becomes joined to the previously formed CDR3 and IGHJ gene. In addition to the 3' cryptic heptamer, more 5' sites exist which can be used in rare occasions. Multiple rounds of light chain editing and IGHV replacement may occur until non self-reactive BcR are produced.

Of note, receptor editing might also be used by B cells to avoid cell death due to unproductive rearrangements or to the inability of H chain to pair with surrogate or conventional light chain.

## B cell differentiation in the periphery

### *Overview of the sequence of events*

Immature B cells leave the bone marrow and migrate to the spleen where they become transitional B cells and complete their maturation into mature naive B cells.<sup>[1,2]</sup> Depending on the type (and strength) of signals they receive from the microenvironment through the BcR and other receptors, they will become either marginal zone or follicular B cells.

Upon antigen stimulation, naive B cells become activated: they proliferate and differentiate into antibody secreting cells.<sup>[9,10]</sup> Two types of response can be distinguished depending on the nature of antigens and the need of helper T cells. Marginal zone B cells are stimulated by polysaccharides (or lipids) often present on certain pathogens such as encapsulated bacteria, and produce IgM-secreting cells and eventually memory cells. This response is T cell-independent, rapid and represents a first line of 'innate-like' defence against blood-borne antigens. In contrast, follicular B cells which constitute most of the mature naive B cells recognize protein antigens and need the help of T cells in order to respond. After an initial proliferation phase in the primary follicles, some follicular B cells will differentiate into IgM-secreting plasma cells, providing a primary antibody response. Others will enter a 'germinal center reaction' in specialized structures of the secondary follicles. There, activated B cells first differentiate into proliferating centroblasts within the dark zone area. During this phase, somatic hypermutations are introduced into their V domain at an extremely high rate (hence the term hypermutation). Centroblasts then move to the light zone where they become centrocytes and stop dividing. In the presence of antigens presented on follicular dendritic cells and with the help of T cells, centrocytes are selected upon the affinity of their BcR. Those with low affinity are eliminated, while those with high affinity can further expand and differentiate. In parallel with this affinity maturation process, centrocytes undergo isotype class-switch leading to replacement of the original IGHM (C $\mu$ ) region by a different one, IGHG (C $\gamma$ ), IGHA (C $\alpha$ ), or IGHE (C $\epsilon$ ). Finally, this results in the production of memory B cells and plasma cells, respectively, expressing and secreting high affinity IG of different classes.

At the molecular level, the IG genes are submitted to 3 kinds of modifications: (1) somatic hypermutations (SHM) of the V domain; (2) class-switch recombination (CSR) of the C region; and (3) switch from membrane-bound to secreted IG. SHM and CSR are initiated by the same enzyme, activating-induced cytidine deaminase (AID), which plays a pivotal role for both events. However, each can occur without the other. In addition, SHM (and CSR) can occur outside germinal centers, as can be observed for some marginal zone B cells.

### *Somatic hypermutations*

#### *General features*

SHM consists essentially of single base substitutions.<sup>[11,13]</sup> They occur at a frequency of  $\sim 10^{-3}$  per base per cell division, which is about  $10^6$  times higher

than the spontaneous rate of mutation in all other cells. Transcription is necessary and targets the SHM to the V domain. Mutations start downstream of the promoter and extend on 1.5 to 2 kb, with a maximal frequency on the rearranged V(D)J genes, but it spares the C region. Of note, non-Ig genes including oncogenes (such as BCL6) can also be targeted, although at a lower frequency. SHM preferentially targets some hot spots which typically consist of the patterns (a/t)a and g(c/t)(a/t) and their complementary reverse motifs t(a/t) and (a/t)(a/g)c (see IMGT/V-QUEST; in the literature wa, gyw, and complementary tw, wrc). Transitions (e.g. c>t, g>a) are more frequent than transversions (e.g. c>a or g, g>c or t). Both strands can be affected. Mutations can be silent (S) or result in the replacement (R) of an amino acid at a given position, and can localize both in FR and CDR. Typically, since R mutations could improve affinity, they tend to accumulate in the CDRs which are the critical antigen binding sites. In contrast, S mutations are favored and R mutations counter-selected in the FRs in order to preserve the overall structure of the domain. Thus, calculation of R/S ratios in CDR and FR is often used to infer whether an antigen selection has taken place. However, this should be done with caution, especially in the case of a low number of mutations, since one R at a particular place may have major effects on affinity.

#### *Molecular mechanisms of SHM*

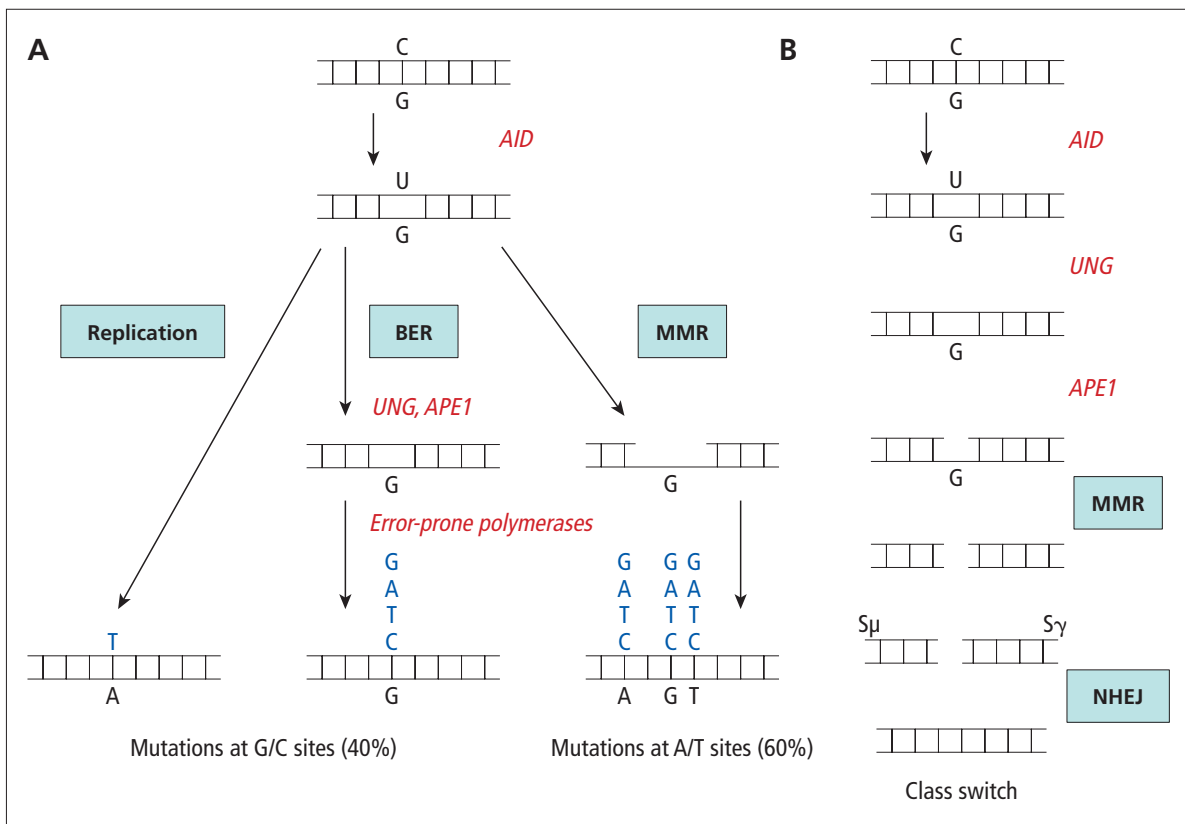
The initial step is mediated by AID which deaminates a 'c' nucleotide into a uridine ('u'), creating a 'u/g' mismatch (Figure 10). As AID acts only on single-strand DNA, transcription is necessary for SHM, since it allows separation of both strands. The 'u/g' mismatch can thereafter be resolved in 3 ways:

1. The 'u' can be considered as a thymidine ('t') during replication, thus creating a c>t mutation.
2. It might be removed by the base excision repair (BER) system, whereby the uracil N-glycosylase creates an abasic site. This is followed by a single-stranded break mediated by the apurinic/apyrimidic endonuclease 1.
3. It may also involve the mismatch repair (MMR) machinery whereby the 'u/g' mismatch is recognized by the MSH2/MSH6 complex and then followed by excision of bases surrounding the initially targeted 'c' nucleotide by the exonuclease 1. This mechanism is thought to be responsible for mutations at 'a/t' basepairs (close to a 'c/g' basepair) which represent about 60% of all mutations.

Finally, in both BER and MMR pathways, breaks and gaps in DNA are repaired by low-fidelity error-prone polymerases (such as Rev1 and Pol $\eta$ ) generating all types of substitution at each site. The 'a/t' versus 'g/c' mutation type might be dependent on the cell cycle phase ('g/c' during G1 phase; 'a/t' during S phase).

#### *Deletions and insertions*

A rare mechanism of SHM leads to the introduction of nucleotide insertions or deletions within IG rearranged sequences; insertions are in



**Figure 10** Model of somatic hypermutation (SHM) and class-switch recombination (CSR).<sup>[11,14,15]</sup> Both share some common mechanisms as they are initiated by activation-induced cytosine deaminase (AID) which necessitates transcription to target single-strand DNA. Cytidines are deaminated into uridines which are then removed by uracil-N-glycosylase (UNG). The replication system transforms the uridine in thymidine leading to a c>t transition (and g>a transition on the opposite strand). In SHM, these abasic sites are further processed by either the base excision repair (BER) or the mismatch repair (MMR) systems. The DNA lesions are repaired by error-prone DNA polymerases, leading to nucleotide transitions and transversions. In CSR, an initial single-strand break is created at the abasic site by the apurinic/aprimidic endonuclease 1 (APE1), followed by a double-strand break induced by MMR enzymes. The fusion of two switch (S) regions breaks originating from different IG classes involves the non-homologous end-joining (NHEJ) pathway in collaboration with several other proteins.

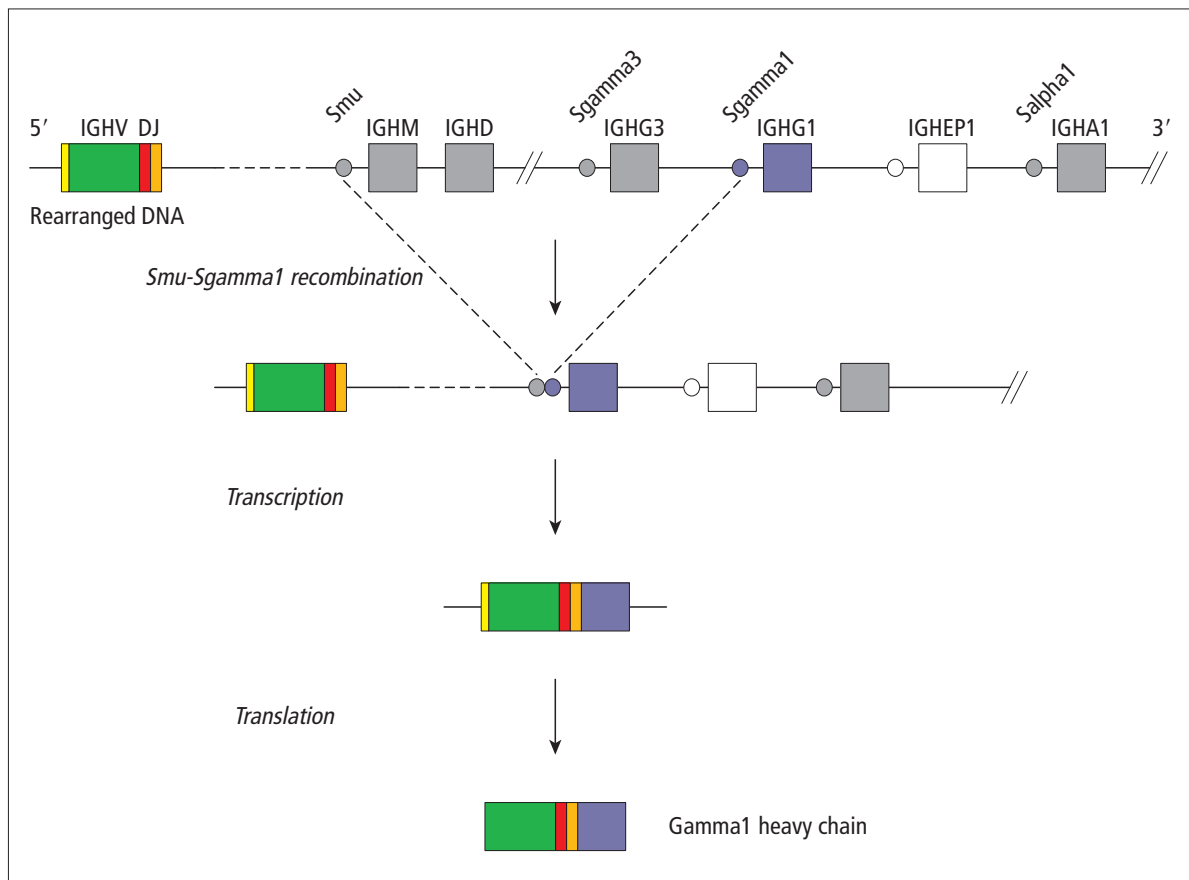
fact duplications of a neighboring sequence. Rearrangements carrying such modifications are productive provided the inserted/duplicated or deleted nucleotides occurred as multiples of 3, thus maintaining the reading frame. The mechanism generating these events is unknown but could involve slippage of the polymerase on particular DNA motifs such as repetitive DNA tracts or stretches capable of forming loop intermediates.

### ***Class-switch recombination***

Switch of IG isotype allows the production of antibodies that perform distinct effector functions.<sup>[14,15]</sup> It requires T cell help through the engagement of CD40 receptors on B cells and secretion of cytokines. The isotype nature depends on the type of antigen and the site of stimulation. This is, in large part, related to the type of helper T cells involved and cytokines they produce.

CSR is mediated by the intrachromosomal recombination of specific DNA sequences located upstream of each C region (except IGHD). These so-called switch regions are 2-3 kb long and contain 'gc' rich repetitive sequences. CSR involves the recombination of the IGHM switch region to that of another C region, with subsequent deletion of the intervening DNA by loop excision. This results in the replacement of the IGHM by a new incoming C region, which is then spliced at the RNA level on the initial V-D-J rearrangement, thus preserving the same V domain (Figure 11).

At the molecular level, CSR shares a number a features with SHM (Figure 10). It necessitates transcription of the C region which starts from a small exon (called I) located upstream of each switch region, resulting in a sterile I-C transcript that does not encode any protein. As mentioned above, CSR is also initiated by AID. Transcription allows the separation of both DNA strands which are then targeted by deamination of multiple 'c' nucleotides by AID. This is followed by generation of single-strand breaks by the BER system, which may be converted into double-strand breaks by the MMR proteins. After deletion of the intervening sequence, fusion of the switch regions is thought to be mediated by the NHEJ systems in association with other proteins.

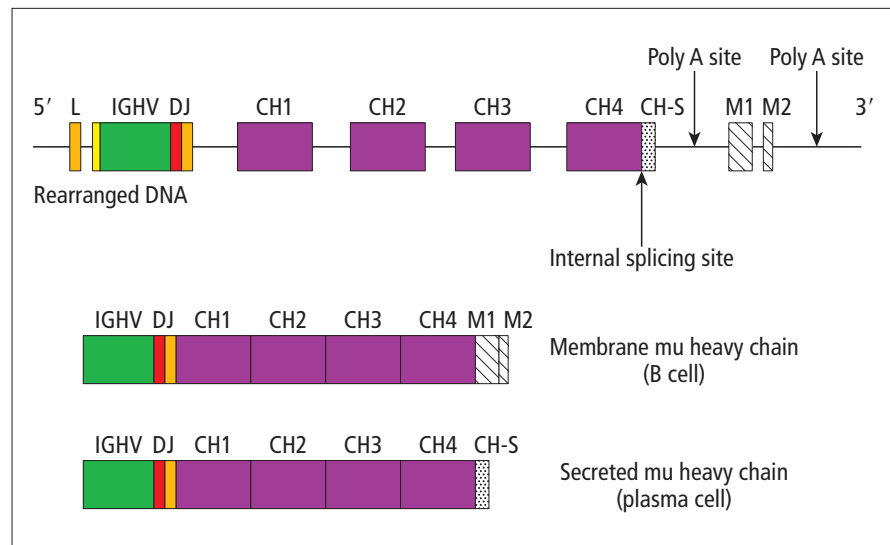


**Figure 11** Schematic representation of heavy chain class-switch: recombination between the switch (S) mu and switch gamma1 regions. Exons of the IGHC genes are not shown and the constant region is not at scale.

### Secretion of IG

Membrane-bound and secreted IG differ in their C terminal portion. This results from the alternative splicing of a primary transcript (Figure 12). The membrane form includes, in addition to the C domains, a transmembrane region and a cytoplasmic region, which are excluded from the secreted form.

**Figure 12** Production of membrane-bound versus secreted IG. Reproduced with the kind authorization of Marie-Paule Lefranc (IMGT®, the international ImMunoGeneTics information system®; <http://imgt.cines.fr>).



### The repertoire of human B cells and the origin of B cell tumors

As discussed, the diversity of the B cell repertoire stems from several sources:

1. *Combinatorial diversity*: The combinatorial process of V domain formation and the vast number of germline genes allows a large number of potential combinations. Taking into consideration only functional genes, there are ~6000 combinations for heavy chains and ~300 for both kappa and lambda chains. Thus, the total number of heavy and light chain combinations is  $\sim 2 \cdot 10^6$ .
2. *Junctional diversity*: Because of the mechanisms described above at the level of the junctions, CDR3 bear the greatest diversity of the molecule. This is particularly true for H chains due to the presence of the IGHD gene and its 2 rearrangements. Overall, it has been estimated that with this additional source of diversity, the potential repertoire may consist of up to  $10^{11}$  different IG.
3. *SHM*: Additional diversification is made by SHM in the mature B cell repertoire.<sup>[13]</sup>

Although the potential repertoire is considerable, the actual expressed repertoire within an individual might be more limited. For instance, as there are about 50 functional IGHV genes, in theory, one might expect to find

in the periphery each IGHV gene expressed at a 2% frequency. However, multiple studies have shown that the normal blood repertoire is biased, with certain IGHV genes being over-represented (such as IGHV4-34, IGHV3-23) while others are under-represented (IGHV3-73). Several factors including gene rearrangement efficiency, the capacity of heavy and light chain pairing, the cellular selection steps described above, type of B cell subset, age and genetic factors may account for these biases. This has to be kept in mind when studying the repertoire of B cell tumors.

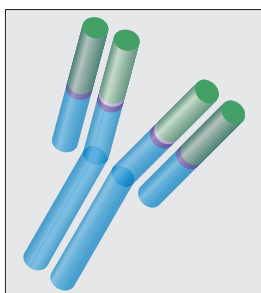
That notwithstanding, the complex process of IG gene maturation from their germline configuration leaves a unique DNA sequence signature in the genome of B cells. This can be used to recognize the monoclonal nature of a B cell population (see Chapter 1), and to track residual tumor cells (see Chapter 8). Overall, IG gene sequences are a testimony of B cells history, and thus their thorough analysis may help to decipher the cell origin of CLL (see further Chapters 4 and 6).<sup>[16-18]</sup>

## References and further reading

1. Rajewsky K. Clonal selection and learning in the antibody system. *Nature* 1996; 381: 751-8
2. Meffre E, Casellas R, Nussenzweig MC. Antibody regulation of B cell development. *Nat Immunol* 2000; 1: 379-85
3. Lefranc M-P, Lefranc G. *The immunoglobulin FactsBook*. London, UK: Academic Press, 2001
4. Schatz DG, Spanopoulou E. Biochemistry of V(D)J recombination. *Curr Top Microbiol Immunol* 2005; 290: 49-85
5. Schlissel MS. Regulating antigen-receptor gene assembly. *Nat Rev Immunol* 2003; 3: 890-9
6. Jung D, Giallourakis C, Mostoslavsky R, et al. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu Rev Immunol* 2006; 24: 541-70
7. Van Zelm MC, Van der Burg M, De Ridder D, et al. Ig gene rearrangement steps are initiated in early human precursor B cell subsets and correlate with specific transcription factor expression. *J Immunol* 2005; 175: 5912-22
8. Nemazee D. Receptor editing in lymphocyte development and central tolerance. *Nat Rev Immunol* 2006; 6: 728-40
9. Carsetti R, Rosado MM, Wardmann H. Peripheral development of B cells in mouse and man. *Immunol Rev* 2004; 197: 179-91
10. Sagaert X, Sprangers B, De Wolf-Peeters C. The dynamics of the B follicle: understanding the normal counterpart of B-cell-derived malignancies. *Leukemia* 2007; 21: 1378-86
11. Peled JU, Kuang FL, Iglesias-Ussel MD, et al. The biochemistry of somatic hypermutation. *Annu Rev Immunol* 2008; 26: 481-511
12. Dörner T, Foster SJ, Brezinschek HP, et al. Analysis of the targeting of the hypermutational machinery and the impact of subsequent selection on the distribution of nucleotide changes in human VHDJH rearrangements. *Immunol Rev* 1998; 162: 161-71
13. Klein U, Goossens T, Fischer M, et al. Somatic hypermutation in normal and transformed human B cells. *Immunol Rev* 1998; 162: 261-80
14. Chaudhuri J, Alt FW. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol* 2004; 4: 541-52. [Erratum in: *Nat Rev Immunol* 2004; 4: 655]

15. Stavnezer J, Guikema JE, Schrader CE. Mechanism and regulation of class switch recombination. *Annu Rev Immunol* 2008; 26: 261-92
16. Stevenson F, Sahota S, Zhu D, et al. Insight into the origin and clonal history of B-cell tumors as revealed by analysis of immunoglobulin variable region genes. *Immunol Rev* 1998; 162: 247-59
17. Küppers R, Klein U, Hansmann ML, Rajewsky K. Cellular origin of human B-cell lymphomas. *N Engl J Med* 1999; 341: 1520-9
18. Küppers R. Mechanisms of B-cell lymphoma pathogenesis. *Nat Rev Cancer* 2005; 5: 251-62





## Section I. Immunoglobulin gene analysis: technical aspects

<b>CHAPTER 1</b>	<b>From the patient to the sequence: selection of material, ... primers and PCR protocol, clonality analysis, and sequencing protocol</b>	<b>23</b>
	<i>Anton W. Langerak</i>	
<b>CHAPTER 2</b>	<b>IMGT® standardized analysis of immunoglobulin rearranged sequences</b>	<b>33</b>
	<i>Véronique Giudicelli and Marie-Paule Lefranc</i>	
<b>CHAPTER 3</b>	<b>Interpretation of 'trouble sequences'</b>	<b>53</b>
	<i>Chrysoula Belessi</i>	





# 1

## From the patient to the sequence: selection of material, primers and PCR protocol, clonality analysis, and sequencing protocol

*Anton W. Langerak*

Chronic lymphocytic leukemia (CLL) is a clinically heterogeneous disorder, mostly showing an indolent disease course, but behaving more aggressively in a subset of patients. Over the years, several clinical and biological parameters have been found to be associated with survival differences in CLL patients. One parameter now stands out as being more informative in terms of prognosis definition: the somatic hypermutation (SHM) status of the immunoglobulin heavy V gene (IGHV). Following the original description in 1999,<sup>[1,2]</sup> it has since been confirmed in many different multi-centre studies that CLL cases showing a rearranged IGHV gene with few or no somatic mutations ('unmutated' IGHV) generally have a less favorable prognosis than cases that exhibit IGHV mutations ('mutated' IGHV). Although differences in the SHM status of CLL cases may reflect a different biological origin (see further Chapter 4), more importantly, they are associated with significant differences in outcome and hence are currently utilized for inclusion or stratification in many clinical protocols.

This chapter discusses several technical aspects of IGHV SHM analysis down the line from patient to sequence, based on the recently published guide to IGHV SHM analysis by the ERIC [European Research Initiative on CLL ([www.ericll.org](http://www.ericll.org))] consortium.<sup>[3]</sup> These recommendations include selection of material, choice of PCR primers and protocols, clonality testing strategies, and sequencing protocols (Table I). Consideration of all of these aspects in any routine clinical diagnostic setting is highly relevant in order to generate a reliable IGHV sequence, which is at the basis of proper IGHV SHM status assignment in individual CLL patients.

### Material selection

Given the high leukocyte count in most CLL patients, peripheral blood (PB), which is easy to obtain, is generally the preferred source of leukemic cells for IGHV SHM analysis (Table I). However, in cases with a low peripheral leukemic burden, other compartments such as bone marrow (BM), lymph node, or any other tissue/fluid in which CLL cells may reside, may be appropriate for obtaining leukemic cells for analysis. As the IGHV SHM status is a stable feature of the leukemic clone, it can be evaluated irrespective of the source of cells. Moreover, SHM status is also considered stable during the disease course, implying that evaluation can be performed at any desired time point. Nevertheless, evaluation of samples with too few leukemic cells

Anton W. Langerak, PhD  
Erasmus MC, University  
Medical Center  
Department of Immunology  
Dr. Molewaterplein 50  
3015 GE Rotterdam  
The Netherlands  
E-mail: a.langerak@erasmusmc.nl

**Table I** Parameters in the IGHV SHM procedure from patient to sequence – recommendations and alternatives

<i>Parameter</i>	<i>Preferred choice</i>	<i>Other options</i>
Source of cells	Peripheral blood (PB)	Bone marrow (BM), lymph node, other
Anticoagulant (PB/BM)	EDTA or CPT tubes	Heparinized tubes
Work-up of cells	Ficoll gradient (PB/BM)	Tissue biopsy or cell suspension of biopsy
Type of nucleic acid	gDNA or RNA/cDNA <sup>a</sup>	–
PCR primers	IGHV leader or IGHV FR1 <sup>a</sup>	IGHV FR2
PCR protocol	None (use existing)	–
<i>Taq</i> polymerase	None	–
Clonality analysis	GS or HD analysis <sup>a</sup>	PAGE (without prior HD)
Sequencing	Direct or via eluted PCR product	Cloning (only with great caution)

<sup>a</sup> See Table II for specific advantages or disadvantages of each option.

*Abbreviations:* BM = bone marrow; CPT = citrate/pyridoxal 5'-phosphate/Tris; EDTA = ethylenediamine tetra-acetic acid; FR = framework region; GS = GeneScan fragment; HD = heteroduplex; IGHV = immunoglobulin heavy V gene; PAGE = polyacrylamide gel electrophoresis; PB = peripheral blood; PCR = polymerase chain reaction.

(e.g. during therapy) is not recommended, as the low tumor load could pose a technical challenge and complicate proper interpretation of the SHM status data concerning the CLL IGHV sequence. When PB or BM aspirates are obtained from the patient, these are best collected in tubes containing EDTA as the anticoagulant or in CPT (citrate/pyridoxal 5'-phosphate/Tris) tubes (Table I). Theoretically, these would best prevent problems with inhibition of PCR amplification. However, in our experience, the use of heparin tubes is a realistic option, implying that heparinized PB and BM samples that are collected for flow cytometric or cytogenetic evaluation can readily be used for IGHV SHM evaluation as well. If PB or BM is the source, Ficoll gradient separation is performed and the mononuclear cell fraction is used for further analysis. Use of whole lysed cell material is not recommended for at least two reasons: (1) the percentage of CLL is lower, which might occasionally create unnecessary complications; and (2) inhibitors of PCR amplification such as hemoglobin are still present. When using lymph node material, either cell suspensions can be made (also appropriate for parallel-flow cytometric analysis) or total biopsies (fresh, archival) can be used.

Starting from either cell fraction or from a biopsy, DNA or RNA can be isolated, as both genomic DNA (gDNA) and complementary DNA (cDNA) can be used as the starting material in IGHV SHM status evaluation (Table I). With respect to the choice of gDNA or cDNA, there is no clear preference, both having their pros and cons (Table II). gDNA, which is most commonly used, has the advantage over RNA/cDNA that no reverse transcription step is required. This is especially useful when archival (paraffin-embedded or stored frozen) cell material is being considered. Also, transportation of material over long distances is less problematic, which is convenient for centralized testing or for sample exchanges between laboratories. On the other hand, when using gDNA, the risk of finding biallelic rearrangements,

<b>Table II</b> Typical advantages and disadvantages of several parameters in the IGHV SHM procedure from patient to sequence		
<i>Parameter</i>	<i>Advantage</i>	<i>Disadvantage</i>
<b>Type of nucleic acid:</b>		
gDNA	More optimal for archival material and transportation No reverse transcription step	Unproductive rearrangement also amplified
RNA/cDNA	Identifies mostly only productive rearrangement Allows isotype identification	Reverse transcription step
<b>PCR primers:</b>		
IGHV leader	Accurate SHM level based on whole IGHV gene	Lower detection rate
IGHV FR1	Widely used in IGHV analysis Higher detection rate	Estimation of SHM level due to lack of information on 5' part IGHV gene
<b>Clonality analysis:</b>		
HD	Unlabeled PCR products allow direct sequencing	Slightly lower detection rate than GS
GS	Slightly better detection rate than HD Optimal visualization of clonality patterns	Labeled PCR products less optimal for direct sequencing

*Abbreviations:* cDNA = complementary DNA; FR = framework region; gDNA = genomic DNA; GS = GeneScan; HD = heteroduplex; IGHV = immunoglobulin heavy V gene; PCR = polymerase chain reaction; SHM = somatic hypermutation.

including an unproductive rearrangement on the second allele, is higher. As this might complicate sequencing interpretation (see further Chapter 7) and/or prolong the procedure, some centres prefer to use cDNA for this reason. However, it should be noted that even though the productive rearrangement at the transcript level is the preferential target, the presence of unproductive transcripts ought not be excluded, as evidenced from some reports (though at a definitely lower frequency as compared with gDNA). A potential drawback is that cDNA-based analysis might mask discrepant cases, for instance when an unmutated, productive rearrangement is co-amplified together with a mutated, unproductive rearrangement at the gDNA level (see also Chapter 7). In addition, using a downstream primer located in the constant region may be advantageous in cases of mutated sequences, since it is not targeted by SHM. Finally, cDNA does allow definition of the isotype, as amplification includes the use of isotype-specific primers.

Taken together, there is no clear scientific rationale for choosing either of the two starting materials, although it would be advisable, in view of more comparable data sets, to standardize the nucleic acid type when starting IGHV SHM evaluation in a multi-centre setting and to explicitly report the source of material when preparing a diagnostic report (see further Chapter 7). Isolation of gDNA or RNA is possible via any method, including several commercially available isolation devices.

## PCR protocols

The concepts and technological approaches of clonality assessment of B cell proliferations (which is, in fact, the first and obligatory step of IGHV sequence

determination), have been extensively reviewed in several reports from the BIOMED-2 consortium.<sup>[4,5]</sup> In order to amplify IGHV rearrangements, several different PCR strategies at the gDNA or cDNA level have been published. Some of these strategies employ consensus IGHV primers; others use IGHV subgroup primers, a set of IGHV subgroup primers multiplexed in one reaction, or primers in the IGHV leader region.<sup>[4,6-8]</sup> On the downstream side, consensus IGHJ or isotype-specific IGHC (mostly IGHM) primers are employed, depending on the choice of gDNA or cDNA as nucleic acid material. Many of these primer sets are being applied routinely in a clinical diagnostic setting, mostly for IGH clonality analysis. In principle, all of these strategies can be utilized for IGHV SHM analysis as well, although as with the type of nucleic acid chosen, there are pros and cons to be considered (Table II).

Strategies employing IGHV leader primers have the intrinsic advantage that the whole IGHV gene is amplified, thus enabling an accurate calculation of the percentage of mutated nucleotides. On the contrary, when using primers within the IGHV gene, such as the frequently used FR1 primers,<sup>[9]</sup> at best, an accurate estimation of the percentage identity can be obtained. In daily practice, this estimation is virtually always accurate enough to classify the individual CLL case in a proper way, although more caution is warranted if the percentage identity is around the 98% cut-off level. In such 'borderline' cases re-evaluation of the IGHV SHM status classification by applying a leader primer-based PCR strategy is recommended. Generally speaking, FR1-based strategies have the advantage that they are already frequently employed in IGH clonality testing, and show a rather high detection rate, especially the widely used multiplex FR1 PCR of the BIOMED-2 consortium.<sup>[4,5]</sup>

Where negative results with the FR1 and consensus PCR strategy are obtained, an FR2 strategy might be considered as alternative (Table I), although it should be noted that the resulting IGHV gene sequences are even shorter to accurately determine the percentage identity. In this case, if a reliable IG rearrangement and a sequence is obtained, an attempt with gene-specific primers should be made in order to get a longer sequence. For all these reasons, an FR3-based evaluation, which does not provide any significant information of the IGHV sequence, is definitely not accurate enough and thus should never be chosen. Rather, a comprehensive approach would include a leader-based strategy next to an FR1-based method. Even though there is no clear rationale for choosing either one as the starting strategy, multi-centre testing for clinical studies is probably only optimal if the PCR protocols and the order in which they are used are standardized in the involved laboratories.

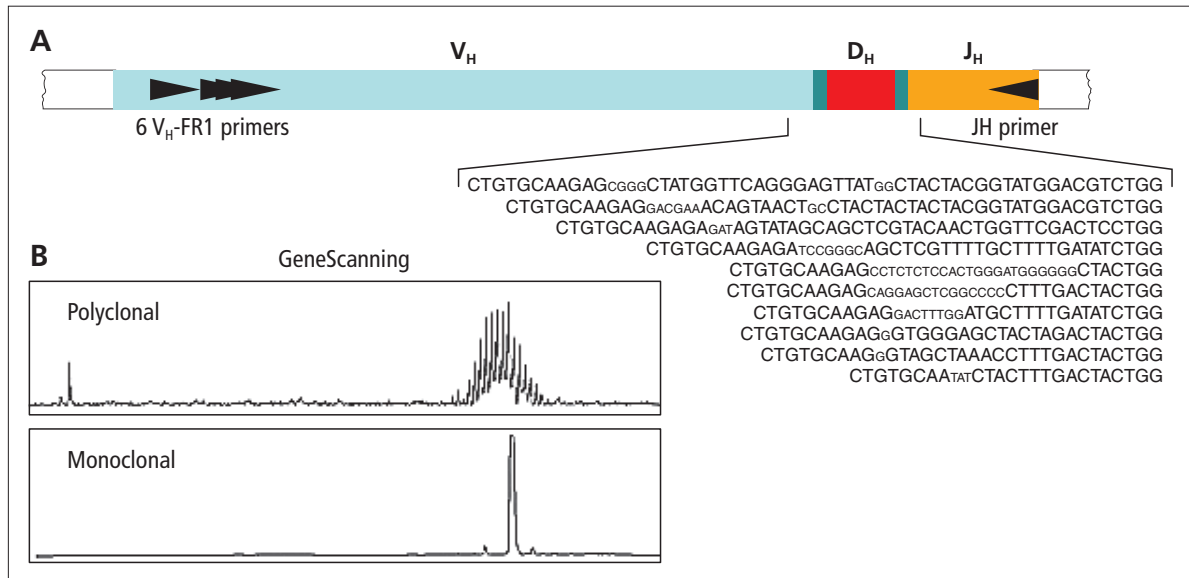
With respect to the detailed PCR protocol (reagents, times, temperatures), there is no preferred approach or critical parameter to be considered and/or optimized for IGHV SHM analysis (Table I). Following the conditions of the original description of the particular PCR protocol as closely as possible is probably the best guarantee for proper amplification. One component of the PCR mixture that needs to be discussed separately is the *Taq* polymerase enzyme. It is important to note that there is no need to use *Taq* polymerase with proofreading activity (Table I). The rate of failures during DNA amplification is so low (in the order  $10^{-4}/10^{-5}$ ), that this does not influence

SHM level determination on a sequence of only a few hundred nucleotides as in any IGHV gene. Evaluation of the SHM status is routinely done via analysis of IGHV genes. Nevertheless, IGKV or IGLV light chain genes could, in principle, also be tested for their SHM level. However, IG light-chain SHM analysis has not found its way into routine practice of CLL testing, most probably because the SHM levels are generally lower and clinical implications have not been validated. Rather, if IGKV/IGLV genes are studied for SHM, this is mostly done for specific research questions, using existing primer sets.<sup>[10,11]</sup>

## Clonality analysis

An important aspect of the whole IGHV SHM analysis procedure is the evaluation of the results of PCR amplification. Firstly, it is very informative to investigate whether PCR amplicons of the correct size are apparent in gel electrophoresis. To this end, agarose gels or preferably polyacrylamide gels should be used, as the latter have a more optimal resolution. Secondly, and perhaps more importantly, further analysis is also required to evaluate the presence of single, monoclonal PCR products as opposed to patterns showing double monoclonal products (biallelic), or even oligoclonality or polyclonality. Oligoclonality and polyclonality might reflect the lack of a true CLL clone in the sample (which would require re-evaluation of the diagnosis) or, alternatively, the presence of a limited amount of leukemic cells [e.g. as occurs in small lymphocytic lymphoma (SLL)]. More importantly, it might be that in those cases, no reliable sequence can be generated because the utilized primer is not annealing to the CLL rearrangement. Hence, in cases of unexpected oligoclonality/polyclonality, before questioning the actual diagnosis, other primer sets should be applied to try to amplify the true CLL clonal IGH rearrangement for IGHV SHM evaluation.

For evaluation of monoclonal vs oligoclonal vs polyclonal patterns, two standardized methods are available. These are GeneScan (GS) fragment analysis<sup>[12,13]</sup> and heteroduplex (HD) analysis,<sup>[14-16]</sup> which have proven to be robust and reliable approaches in a large European consortium on clonality analysis.<sup>[4,5]</sup> These approaches exploit the V-D-J CDR3 (complementarity determining region 3) size heterogeneity, and in HD the differences in the actual CDR3 nucleotide sequences (CDR3 composition) are also taken into account. In GS analysis, fluorescently labeled PCR products are separated by high-resolution gel or capillary electrophoresis, resulting in clear visualization of the level of heterogeneity of the IGH V-D-J PCR products, based on their CDR3 lengths (Figure 1). In HD analysis, PCR products are denatured and enforced to reanneal at low temperatures, resulting in smears or clear bands that reflect polyclonal or monoclonal cases, respectively (Figure 2). In principle, both GS and HD analysis serve to study clonality patterns (Table I), although there are considerations to take into account when choosing either of the two approaches (Table II). GS analysis is slightly more sensitive than HD analysis, and it enables a more optimal visualization of the clonality pattern, which can be especially useful in cases of two rearrangements of fairly similar size, which might otherwise appear as a single band on HD gels.



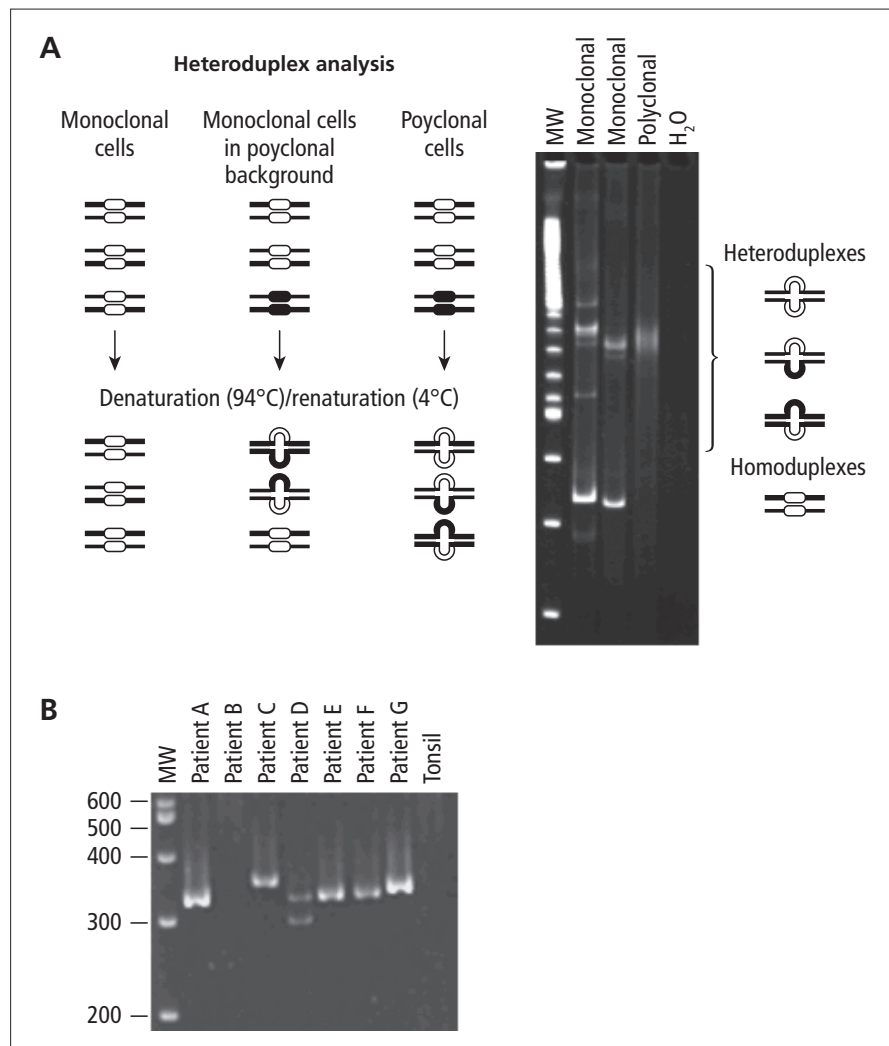
**Figure 1** GeneScan (GS) fragment analysis in clonality testing. **A.** Schematic representation of an IGHV-IGHD-IGHJ rearrangement. Positions of the forward IGHV FR1 (V<sub>H</sub>-FR1) primers and reverse IGHJ primer (J<sub>H</sub>) used in a multiplex PCR reaction are shown. IGH V-D-J gene rearrangements display heterogeneous junctional regions (also known as CDR3 motifs) that differ in size and nucleotide composition. V, D, and J germline nucleotides are shown in large capitals, and randomly inserted nucleotides in small capitals. Upon amplification in an IGHV-FR1 multiplex, PCR reaction products are obtained for further analysis. The junctional heterogeneity in size can be exploited to discriminate between polyclonal and monoclonal IGH V-D-J PCR products using GS fragment analysis. **B.** Example of GS analysis in which fluorochrome-labeled PCR products are denatured for high-resolution fragment analysis of the single-stranded fragments. Monoclonal IGH V-D-J PCR products of identical size thereby give rise to one dominant peak, whereas polyclonal IGH V-D-J PCR products show a Gaussian size distribution.

Such a finding will thereby dictate the sequencing strategy, most probably implying a cloning procedure in order to separate and sequence the two PCR products. On the other hand, with HD analysis, no fluorochrome-labeled primers are required, thus enabling direct sequencing of the PCR products once a monoclonal pattern is observed (see below). When fluorescently labeled PCR products are produced, as in the GS analysis approach, direct sequencing is technically more cumbersome.

### Sequencing protocols

If a monoallelic (and thus monoclonal) rearrangement is observed in clonality analysis (see above), direct sequencing of the PCR product can be performed. The sequencing process is mostly started from the downstream side using either IGHJ consensus or IGHC primers as the reverse primer, depending on the type of nucleic acid being used. From the resulting sequence that is obtained, which covers a large part of the IGHV gene, the IGHV subgroup can be identified, and in a subsequent sequencing experiment, a forward IGHV subgroup or leader primer can be used to generate a complete sequence. Alternatively, parallel to reverse sequencing (with IGHJ or IGHC primers), forward sequencing can be done using an IGHV leader or a consensus IGHV FR1 primer rather than using a mixture of IGHV subgroup primers. Some

**Figure 2** Heteroduplex (HD) analysis in clonality testing. **A.** Schematic overview of HD analysis in which PCR products are denatured (5', 94°C) and reannealed (rapidly cooled at 4°C for 60 min) to form duplexes. In cases of rearranged IGH V-D-J genes, HD analysis exploits the junctional heterogeneity (both in size and composition) to discriminate between polyclonal and monoclonal PCR products. Monoclonal IGH V-D-J PCR products give rise to homoduplexes, as identified by distinct bands, whereas polyclonal IGH V-D-J PCR products mainly form heteroduplexes, resulting in a smear of slow-migrating fragments. **B.** Example of HD analysis in B cell chronic lymphocytic leukemia (CLL) patients showing cases with monoallelic IGH V-D-J rearrangements (patients A, C, E, F, G), biallelic IGH V-D-J rearrangements (patient D), or polyclonality (patient B).



laboratories prefer to have forward as well as reverse sequences from two independent PCR products (total of four sequences), but in our experience there are few, if any, discrepancies that could be attributed to *Taq* polymerase errors. All resulting sequences are then aligned and a consensus sequence can be created for further analysis of the SHM level and IGHV, IGHD, IGHJ gene usage (see further Chapter 2).

In the case of biallelic rearrangements, the two PCR products should be excised from a polyacrylamide gel and eluted, and then sequenced in separate reactions. Alternatively, when multiplex strategies are used, it might be worthwhile to repeat PCR amplification in simplex/single reactions to generate single products that can subsequently be directly sequenced. In cases where both rearrangements utilize an IGHV gene of the same subgroup, the latter approach will fail and gel excision is the only option.

When all other attempts to generate a reliable CLL sequence have failed, subcloning might be considered as final option. Subcloning strategies and subsequent sequencing of individual colonies are cumbersome approaches

that are hardly ever needed. The one situation where they may be required is in the case of biallelic rearrangements where the difference in size of the two PCR products does not allow their physical separation by gel excision. Subcloning might also be useful for looking at intraclonal diversity. Currently this is only done for research purposes, as its prognostic value has not yet been explored.

## Concluding remarks

In the IGHV SHM analysis procedure, many of the technical parameters down the line from patient to sequence have now been standardized in consensus guidelines. Interpretation of sequencing results is discussed in Chapter 2. For several issues, the consensus guidelines show a preference for particular options, but at other points (gDNA vs cDNA; IGHV leader primers vs IGHV FR1 primers; method of clonality analysis) flexibility is larger, partly for historical reasons and because of established preferences in certain laboratories. However, it should be noted that in some cases, the alternative options might provide subtle differences with respect to the end result, which should be taken into account when interpreting the data obtained. Given the lack of a firm scientific rationale to prefer one option over the other, it is therefore recommended to standardize all aspects of the entire IGHV SHM procedure when multi-centre clinical trials are planned and initiated and the inclusion or stratification of patients is based on their IGHV SHM status. This will improve the total data set and greatly facilitate comparability of the results, even though these are being collected in different laboratories.

## References

1. Damle RN, Wasil T, Fais F, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* 1999; 94: 1840-7
2. Hamblin TJ, Davis Z, Gardiner A, et al. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* 1999; 94: 1848-54
3. Ghia P, Stamatopoulos K, Belessi C, et al. ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. *Leukemia* 2007; 21: 1-3
4. Van Dongen JJM, Langerak AW, Brüggemann M, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene rearrangements in suspect lymphoproliferations. Report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 2003; 17: 2257-317
5. Evans PA, Pott Ch, Groenen PJ, et al. Significantly improved PCR-based clonality testing in B-cell malignancies by use of multiple immunoglobulin gene targets: report of the BIOMED-2 Concerted Action BHM4-CT98-3936. *Leukemia* 2007; 21: 207-14
6. Campbell MJ, Zelenetz AD, Levy S, et al. Use of family-specific leader region primer for PCR amplification of the human heavy chain variable region gene repertoire. *Mol Immunol* 1992; 29: 193-203

7. Küppers R, Zhao M, Rajewsky K, et al. Detection of clonal B cell populations in paraffin-embedded tissues by polymerase chain reaction. *Am J Pathol* 1993; 143: 230-9
8. Aubin J, Davi E, Nguyen-Salomon E, et al. Description of a novel FR1 IgH PCR strategy and its comparison with three other strategies for the detection of clonality in B cell malignancies. *Leukemia* 1995; 9: 471-9
9. Matthews C, Catherwood M, Morris TC, et al. Routine analysis of IgVH mutational status in CLL patients using BIOMED-2 standardized primers and protocols. *Leukemia Lymphoma* 2004; 45: 1899-904
10. Stamatopoulos K, Belessi C, Hadzidimitriou A, et al. Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. *Blood* 2005; 106: 3575-83
11. Belessi C, Stamatopoulos K, Hadzidimitriou A, et al. Analysis of expressed and non-expressed IGK locus rearrangements in chronic lymphocytic leukemia. *Mol Med* 2005; 11: 52-8
12. Kneba M, Bolz I, Linke B, et al. Analysis of rearranged T-cell receptor beta-chain genes by polymerase chain reaction (PCR) DNA sequencing and automated high resolution PCR fragment analysis. *Blood* 1995; 86: 3930-7
13. Linke B, Bolz I, Fayyazi A, et al. Automated high resolution PCR fragment analysis for identification of clonally rearranged immunoglobulin heavy chain genes. *Leukemia* 1997; 11: 1055-62
14. Bottaro M, Berti E, Biondi A, et al. Heteroduplex analysis of T-cell receptor gamma gene rearrangements for diagnosis and monitoring of cutaneous T-cell lymphomas. *Blood* 1994; 83: 3271-8
15. Langerak AW, Szczepanski T, Van der Burg M, et al. Heteroduplex PCR analysis of rearranged T cell receptor genes for clonality assessment in suspect T cell proliferations. *Leukemia* 1997; 11: 2192-9
16. Gonzalez M, Gonzalez D, Lopez-Perez R, et al. Heteroduplex analysis of VDJ amplified segments from rearranged IgH genes for clonality assessments in B-cell non-Hodgkin's lymphoma. A comparison between different strategies. *Haematologica* 1999; 84: 779-84

