Available online at www.sciencedirect.com
ScienceDirect
Molecular Immunology 44 (2007) 1057–1064
Molecular Immunology
www.elsevier.com/locate/molimm

Review

Statistical analysis of CDR3 length distributions for the assessment of T and B cell repertoire biases

Patrick Miqueu [a,b], Marina Guillet [b], Nicolas Degauque [a], Jean-Christophe Doré [c], Jean-Paul Soulillou [a], Sophie Brouard [a,*]

[a] *Institut National de la Santé Et de la Recherche Médicale (I.N.S.E.R.M.), Unité 643, "Immunointervention dans les Allo et Xénotransplantations" and Institut de Transplantation Et de Recherche en Transplantation (I.T.E.R.T.), CHU Hôtel-Dieu, 30 Bd Jean Monnet, 44093 Nantes Cedex 01, France*
[b] *TcLand S.A., CHU Hôtel-Dieu, 30 Bd Jean Monnet, 44093 Nantes Cedex 01, France*
[c] *Museum National d'Histoire Naturelle, UMR 5154 CNRS/MNHN – USM 0502 MNHN, 63 rue de Buffon, 75005 Paris, France*

Received 24 May 2006; accepted 15 June 2006
Available online 22 August 2006

Abstract

Complementarity-determining region 3 (CDR3) length distribution analysis explores the diversity of the T cell receptor (TCR) and immunoglobulin (Ig) repertoire at the transcriptome level. Studies of the CDR3, the most hypervariable part of these molecules, have been frequently used to identify recruitment of T and B cell clones involved in immunological responses. CDR3 length distribution analysis gives a clear perception of repertoire variations between individuals and over time. However, the complexity of CDR3 length distribution patterns and the high number of possible repertoire alterations per individual called for the development of robust data analysis methods. The goal of these methods is to identify, quantify and statistically assess differences between repertoires so as to offer a better diagnostic or predictive tool for pathologies involving the immune system. In this review we will explain the benefit of analyzing CDR3 length distribution for the study of immune cell diversity. We will start by describing this technology and its associated data processing, and will subsequently focus on the statistical methods used to compare CDR3 length distribution patterns. Finally, we will address the various methods for assessing CDR3 length distribution gene signatures in pathological states.

Keywords: Spectratyping; Immunoscope; Immunoglobulin; T cell receptor; CDR3; Statistical analysis

1. Studying diversity

T and B cells are central players of adaptive immunity which is defined by its exquisite specificity for distinct molecules and its ability to remember and respond more vigorously to repeated exposures to the same molecules. To ensure an efficient protection of the organism and to distinguish among different, even closely related, pathogenic agents, a huge number of different T and B cell clones are generated (Arstila et al., 1999). This defense mechanism is based on the recruitment, selection and expansion of specific T and B cell clones, dedicated to the destruction of pathogenic agents. The specificity of antigen recognition is ensured by the T cell receptor (TCR) present at the surface of T cells (Tonegawa, 1983) and immunoglobulins (Ig) produced by B cells (Kurosawa and Tonegawa, 1982). Each T cell clone is characterized by its own TCR and each B cell clone by its own Ig. T and B cell diversity is thus dependent on the TCR and Ig. The study of these molecules allows for the identification of biases in the repertoire usage and clones mobilized during immunological responses.

The complementarity-determining region 3 (CDR3) is the most polymorphic region of both the TCR and Ig and interacts with antigen in a specific manner. T and B cell diversity thus correlates with CDR3 variability. Several mechanisms have been described to generate the broad diversity of CDR3. Considering the TCR αβ, which is made up of an α chain and a β chain, the CDR3 is composed of sequences encoded by V and J gene segments in the α chain and sequences encoded by V, D and J gene segments in the β chain. Random selection and

Abbreviations: CDR3, complementarity-determining region 3; TCR, T cell receptor; Ig, immunoglobulin; PCR, polymerase chain reaction

* Corresponding author. Tel.: +33 2 40 08 74 17; fax: +33 2 40 08 74 11.
E-mail address: Sophie.Brouard@univ-nantes.fr (S. Brouard).

0161-5890/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.molimm.2006.06.026

association of respectively the V–J and V–D–J segments during somatic gene rearrangements create a combinatory diversity (Davis and Bjorkman, 1988; Tonegawa, 1983). This mechanism is complemented by a junction diversity created by nucleotide deletions/insertions at the V–J junctions for both chains and D–J junction for the β chain. This complex rearrangement mechanism should in theory allow for around 10^6 different CDR3 sequences for the Vβ chain of the TCR in human blood (Arstila et al., 1999). Sequencing of all mRNA encoding the CDR3 can be performed to assess this variability. This is an exhaustive and precise method, albeit expensive and time-consuming and thus rendering it difficult to apply in the context of immune monitoring. This approach can be used to identify the statistically more frequent CDR3 sequences in the transcriptome. The high frequency of a specific CDR3 sequence illustrates the expansion of its corresponding cell clone. The alternative method of CDR3 length distribution analysis (Cochet et al., 1992; Gorski et al., 1994) is an easier approach to characterize the presence of clonal expansions. CDR3 length distribution analysis has been performed in rodents (Cochet et al., 1992; Douillard et al., 1998), swine (Baron et al., 2001) and humans (Gorski et al., 1994) to assess mobilization of Ig (Baranzini et al., 1999), TCR γδ (Dechanet et al., 1999; Hviid et al., 2000) and TCR αβ (Cochet et al., 1992) repertoires under physiologic conditions as well as in various pathological situations. For example, various approaches to CDR3 length distribution analysis have been employed successfully in numerous studies including in multiple sclerosis (Laplaud et al., 2004; Muraro et al., 2005), hairy cell leukemia (Arons et al., 2005) and graft-versus-host disease following allogeneic stem cell transplantation (Tsutsumi et al., 2004). A highly focused immune repertoire has been shown to be associated with successful containment of HIV (Gorochov et al., 1998), *Mycobacterium tuberculosis* infection (Tully et al., 2005) and clinically effective T cell responses in patients with cancer (Coulie et al., 2001). Finally, precise assessment of T cell repertoire diversity is a powerful tool to identify antigen specific T cells, and gives new insight into how a focused lymphocyte population can govern the immune status.

The CDR3 length distribution analysis method subdivides the CDR3-coding mRNA populations into groups, based on V or V–J gene composition and transcript length. This method analyses the effect of the junction diversity for each V or V–J gene composition tested. The CDR3 length distribution, obtained for each V or V–J gene combination, can identify clonal alterations and expansions. Indeed, following the presentation of a foreign antigen or altered self-antigen by self-major histocompatibility complex molecules, T cell clones are specifically selected and expanded. These selections and expansions lead to modification of the presumed Gaussian-like CDR3 length distribution encountered in physiological conditions (Pannetier et al., 1995). Due to the large number of different V and J genes that could potentially constitute the TCR or Ig, the number of CDR3 length distributions for an individual is substantial, as confirmed by the ImMunoGeneTics database (Lefranc, 2003), a reference resource for studies of Ig and TCR.

2. Technology and data processing

The CDR3 length distribution analysis method is a molecular biology tool based on the analysis of V or V–J gene composition and transcript lengths. Following reverse-transcription into cDNA, CDR3-encoding mRNA is amplified by PCR using specific V and C primers. PCR products are then labeled with a fluorescent C or J primer or radiographic isotope. Finally, amplified fragments are separated by electrophoresis, and visualized by fluorescent sequencing or autoradiography equipment (Fig. 1A). A band intensity pattern is thus obtained. This set of bands represents the total population of mRNA sharing the same V gene or V–J gene combination but varying in length. Bands are usually separated by a space corresponding to a transcript length of three nucleotides; suggesting that the population is composed of in-frame transcripts of different length (Pannetier et al., 1995). Signal processing programs (Liu et al., 1995; Pannetier

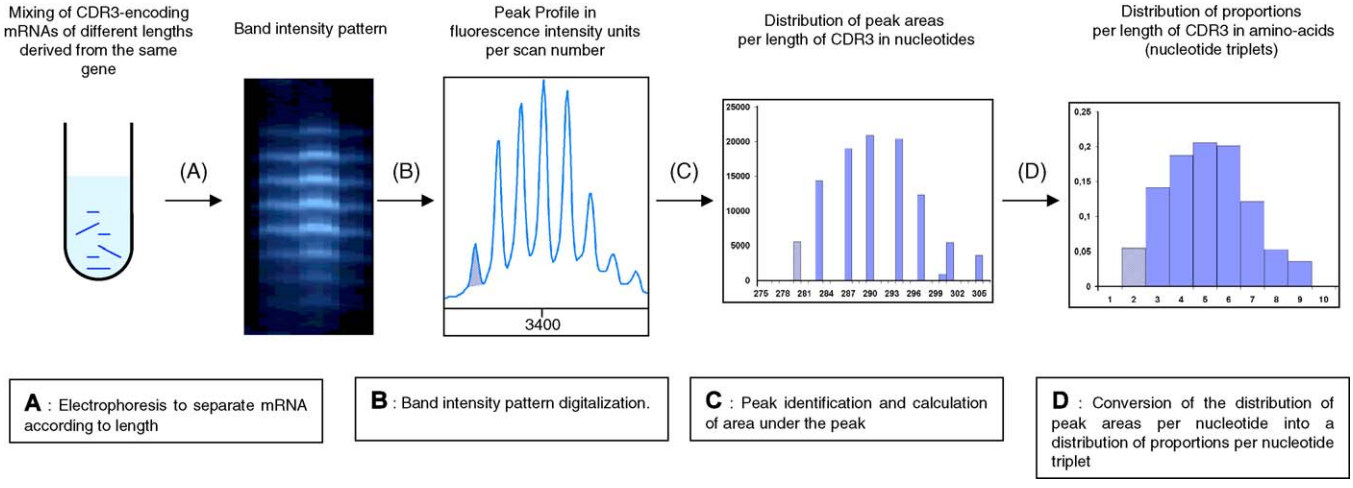

Fig. 1. Signal processing of CDR3 length distribution. The four different steps necessary to turn the electrophoresis signal into quantitative data are outlined. There are three different ways of representing a CDR3 length distribution: as a band intensity pattern, a peak profile or a histogram. Colored in gray, the peak in the profile correspond to the length class in the histogram.

et al., 1995) can be used to digitize the gel image (Fig. 1B) thus enabling quantification of the relative amount of fluorescence/radioactivity in each band. A frequency distribution of the CDR3 lengths (expressed in fluorescence/radioactivity units) is then obtained (Fig. 1C). Once the relative proportions of each length class of mRNA are calculated, the CDR3 length distribution can be displayed as a probability distribution (Fig. 1D). In humans, the TCR β chain has been described as having 65 Vβ genes grouped into 30 families on the basis of the sequence similarity (Folch and Lefranc, 2000). Designing specific labeling primers for each TCR Vβ chain gene, or for a whole Vβ chain family, therefore remains problematic. Among these different Vβ genes families, 26 have been shown to lead to the expression of a functional TCR. This observation tallies with the fact that, in the majority of studies, around 25 (Pannetier et al., 1995) or 26 (Gorski et al., 1994) Vβ genes/family of genes are tested. Nevertheless, the heterogeneity of primers used between studies, together with ambiguity over the Vβ gene names arising from the coexistence of several gene nomenclatures (Arden et al., 1995; Rowen et al., 1996; Wei et al., 1994; Wilson et al., 1988), do not facilitate comparisons between published results.

3. Statistical analysis of CDR3 length distribution

While the method used to calculate CDR3 length distribution is now a validated procedure, the assessment of statistically relevant differences between repertoires of individuals has remained elusive. The challenges encountered when analyzing CDR3 data are two-fold: (1) extraction of information from each CDR3 length distribution and (2) integration of information derived from every CDR3 length distributions. A key observation has been made since the initial publications in this field: recurrent Gaussian-like distribution patterns, usually composed of 8–10 different CDR3 length classes (Pannetier et al., 1993) are observed in healthy individuals (Gorski et al., 1994), as well as naive rats (Douillard et al., 1998), swine (Baron et al., 2001) and mice (Cochet et al., 1992), but also in superantigen-stimulated cells (Musette et al., 1996). In nature and in many biological situations, Gaussian distributions occur when a large number of small effects act additively and independently. Such Gaussian CDR3 length distribution is assumed to be the result of the random mechanism involved in the junction diversity of the TCR. The length of the CDR3 loop greatly influences its shape and ability to fold both on itself and in the company of others loops such as the CDR1 and CDR2 (Davis et al., 1998; Wu et al., 2002). In regard to all these observations, appropriate rearrangements of V (D) J segments resulting in medium-size CDR3 are more likely to occur than rearrangements resulting in short or long CDR3. The recurrence of the Gaussian-like distribution, in healthy or resting situations, inspires numerous statistical methods. From the most trivial to the most sophisticated, these methods have been developed to assign distributions as Gaussian or not. The length of the CDR3 expressed in amino acids can be considered as a discrete random variable. Because a Gaussian CDR3 length distribution is usually composed of 8–10 length classes, counting the number of length classes present in CDR3 length distributions has been proposed as an initial method to evaluate how close a particular CDR3 length distribution is to a Gaussian distribution (Wu et al., 2000). However, despite being easy to implement, the latter method does not take the shape of the distributions into account.

To take this information into consideration, more formal descriptive parameters defining the Gaussian-like distribution can be used. Kurtosis (Fig. 2A), that corresponds to the degree of peakdness in a distribution, and skewness (Fig. 2B), which measures the asymmetry of the distribution (Peggs et al., 2003), give a good indication of the shape of a distribution. However, when the distribution is multimodal, i.e. with more than one major length class, skewness and kurtosis are limited indicators.

The use of statistical methods based on regression analysis or Goodness-of-Fit tests enable the rational assessment of whether a CDR3 length distribution is normally distributed or

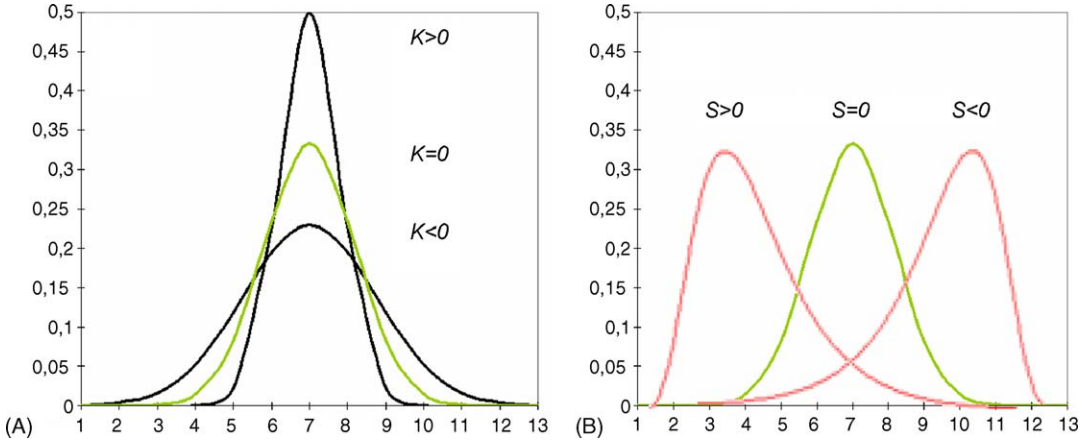

Fig. 2. Skewness and kurtosis are two typical parameters to describe a Gaussian-like distribution. Kurtosis (*K*), as described in (A), is a measure of the peakdness of a distribution. When $K<0$, the distribution is "flatter" than the theoretical Gaussian distribution, colored in green and which $K=0$. Skewness (*S*) is an index of the distribution asymmetry. As shown in (B), the skewness of a theoretical Gaussian distribution is equal to 0, while skewness of a left skewed distribution is positive and skewness of a right skewed distribution is negative. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

not. Regression analysis is a statistical method that chooses a curve that best fits given data points. The performance of the curve fitting is related to the value of the coefficient of determination, R^2. The coefficient value varies from 0 to 1 according to the correlation between the CDR3 length distribution and a statistical model. Linear regression, used to compare two CDR3 length distributions (Naumov et al., 1996), or Quadratic regression (Karadimitris et al., 2000), used to determine whether the log of the intensities of the CDR3 length distribution fits a parabola, have also been tested on CDR3 data (Fig. 3, I). Regression methods have the advantage of giving a score of curve fitting. In comparison, statistical tests can be used to decide whether a proposition is true or not, with a certain level of confidence. With CDR3 length distribution data, Goodness-of-Fit tests make it possible to reject, with a controlled risk, the hypothesis that a CDR3 length distribution fits a spe-

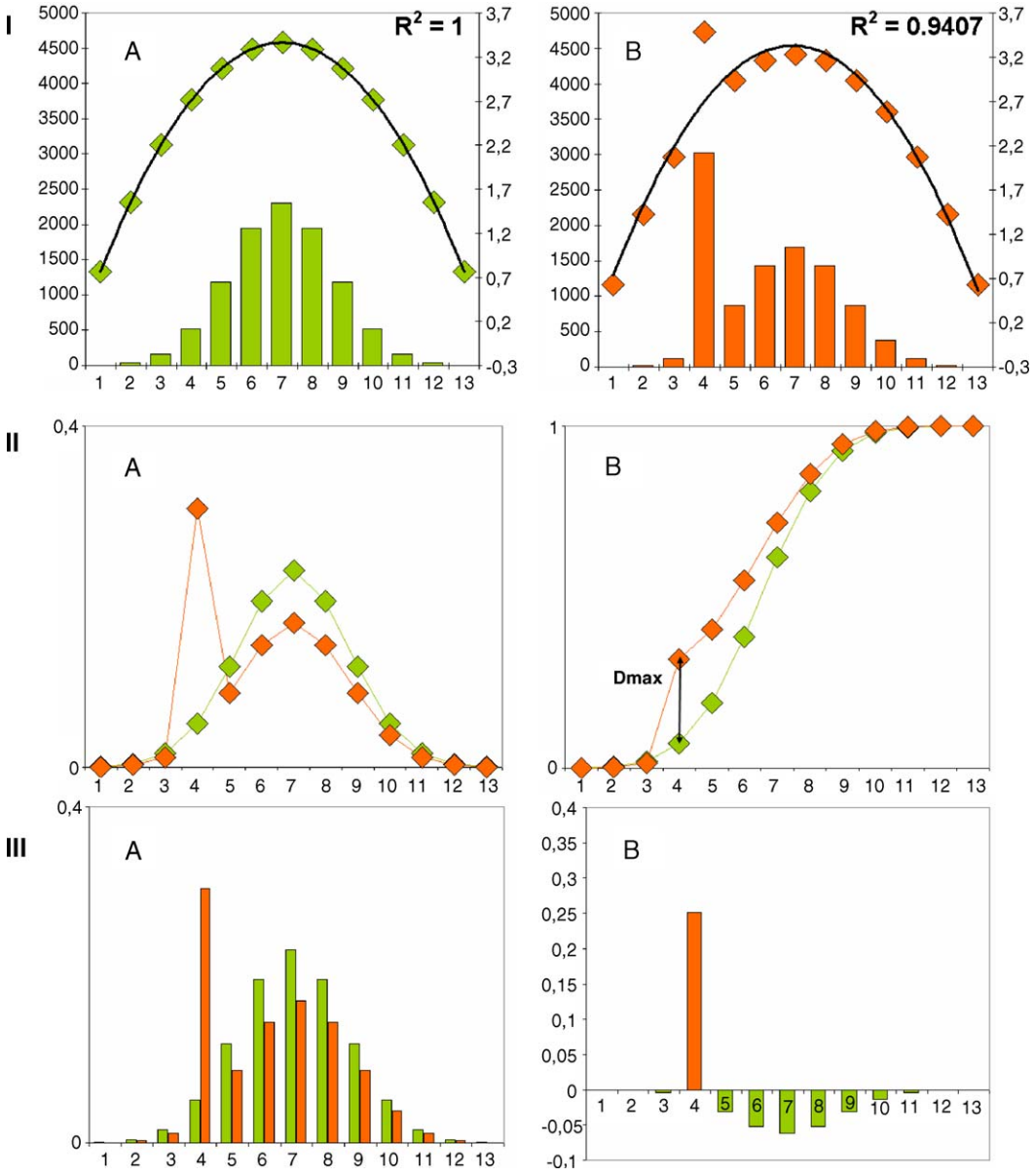

Fig. 3. Three different types of statistical methods to compare a perturbed and a Gaussian CDR3 length distribution. A theoretical Gaussian distribution, colored in green, is compared to a perturbed distribution colored in orange. (I) *Quadratic Regression* is a process by which the equation of a parabola of "best fit" is found for a set of data. When the regression equation truly represents the data set, the *coefficient of determination*, R^2, equals 1. I(A) For a Gaussian distribution, as depicted by the histogram of fluorescence intensities, the logarithm of these intensities (rhombus) are perfectly described by a parabola ($R^2 = 1$). I(B) For a perturbed distribution, quadratic regression does not provide a perfect fit and R^2 is below 1. (II) Statistical tests such as the *Kolmogorov–Smirnov Goodness-of-Fit test* can demonstrate that two distributions are identical. II(A) To compute this test, the discrete distributions, as depicted by the histograms in I(A) and I(B), are transformed into continuous data. II(B) Once the cumulative density functions of the continuous distributions have been calculated, a D-statistics is computed. The value of the D-statistics enables rejection of the null hypothesis of no difference between the distributions. III(A) The Gaussian and perturbed distributions can be depicted as histograms and presented on the same graph. III(B) By calculating the differences in proportions of each length class, it is possible to identify where the clonal expansions occur. The sum of the absolute local differences is the order 1 *Minkowski distance*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

cific distribution such as a Gaussian distribution. For example, Kolmogorov–Smirnov tests were firstly used to show that the CDR3 length distribution of reference samples has a very Gaussian distribution (Gorochov et al., 1998) (Fig. 3, II). Briefly, this test computes a D-statistics (Goodness-of-Fit statistics) between the cumulative density functions of the CDR3 length distribution analyzed and the Gaussian model, and allows for rejection of the null hypothesis of no difference between the distributions. This test has also been successfully used to compare the CDR3 length distribution of healthy individuals (King et al., 2001). Because the Kolmogorov–Smirnov test is restricted to continuous distributions, while CDR3 length distributions are discrete distributions, it could be replaced by the χ^2 Goodness-of-Fit test that has the advantage of using frequencies (Mosley et al., 1998).

The regression and Goodness-of-Fit tests can be used to decide whether or not the distribution is Gaussian, which is very convenient. However, identifying precisely which length classes differ from one distribution to another is essential. Gorochov et al. (1998) proposed to "subtract" an average Gaussian CDR3 length distribution from each CDR3 length distribution obtained from studied samples. The average Gaussian CDR3 length distribution is composed of Gaussian-like CDR3 length distributions of individuals from a reference group of healthy individuals. At a precise length class, a positive difference means that CDR3 transcripts are over-represented in the studied sample CDR3 length distribution in comparison with the average Gaussian CDR3 length distribution (Fig. 3, III). These differences and the distribution shapes can also be represented as a landscape, encompassing all Vβ genes for an individual (Gorochov et al., 1998; Sebille et al., 2001). By adding together all the absolute differences, a percentage of perturbation can be calculated for each gene. This score indicates the number of differences between the tested CDR3 length distribution and a computed reference distribution (Gorochov et al., 1998; Benveniste et al., 2001) or a Gaussian theoretical one (Han et al., 1999). This method is widely used because it pinpoints the length classes that differ from a Gaussian distribution.

Different categories of typical distribution shapes can be defined by a visual analysis of CDR3 length distribution. This method consists in categorizing each CDR3 length distribution according to its shape. Various kinds of categorizations have been used in the literature, from the more basic kind that classifies which is or is not a Gaussian-like distribution (Risitano et al., 2002), to the more sophisticated kind (Demoulins et al., 2003; Guilherme et al., 2000; Sarzotti et al., 2003; Yawalkar et al., 2003). Gaussian-like patterns characterize "*polyclonal*" responses (Pannetier et al., 1995). These patterns illustrate that a large number of clones derived from the random rearrangement process are available (Fig. 4A). Repertoire complexity, depicted by a large number of different-sized PCR products, afford immunological advantages as it as been shown regarding the avoidance of opportunistic infections (Friedman et al., 2001). The term "*oligoclonal*" is used to characterize CDR3 length distribution with a few major CDR3 length classes (Fig. 4B and C). There are two possible interpretations for these skewed patterns: they could either result from oligoclonal expansions over a polyclonal background (Cibotti et al., 1994; Even et al., 1995; Laouini et al., 2000; Pannetier et al., 1995), or, when arising following severe T cell depletion, they could be an indication of a lack of complexity (Gorski et al., 1994; Friedman et al., 2001). In addition, it is important to note that a CDR3 length distribution with a major length (Fig. 4C) does not necessarily result from a monoclonal expansion. Each peak, corresponding to a given CDR3 length, usually contains multiple distinct sequences. The development of a cloning-sequencing approach, coupled with the CDR3 length distribution analysis method, can provide information on the number of distinct types of sequence included in a CDR3 length class (Arstila et al., 1999; Casrouge et al., 2000; Gapin et al., 1998; Hsieh et al., 2004).

As suggested above, the differences that exist between two CDR3 length distributions are due to the existence of clonal expansions. Various methods attempt to explain these differences by identifying where the clonal expansions are located in a skewed CDR3 length distribution. For example, an "expansion" can be statistically pinpointed by a length class proportion

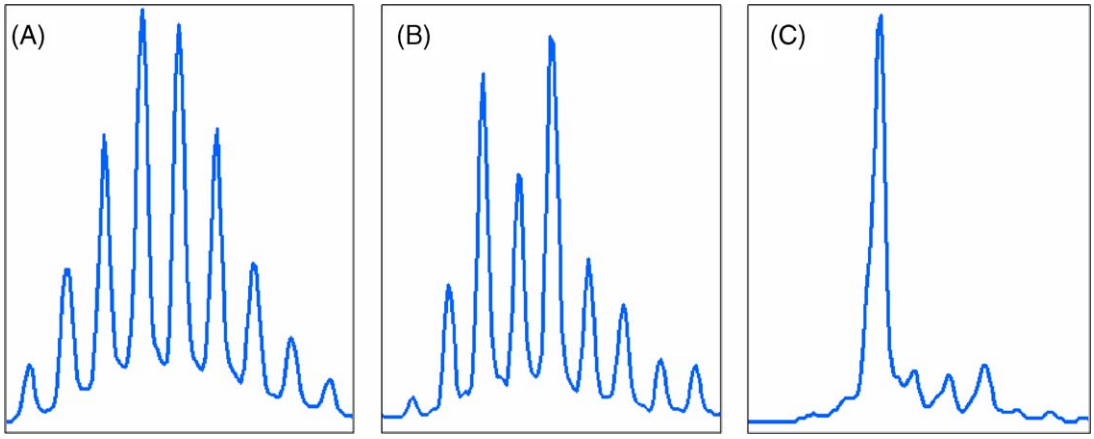

Fig. 4. Examples of CDR3 length distribution patterns (peak profile representation). The (A) profile is a typical Gaussian-like profile which characterize healthy volunteers or superantigen-stimulated cells. It is often classified as belonging to the so-called "*polyclonal*" category. The (B) and (C) profile, which are "perturbed" or "skewed" compared to a Gaussian distribution, reveals the appearance of potential clonal expansions. This kind of pattern is referred to as "*oligoclonal*". The C profile could reveal a "*monoclonal*" pattern with one predominant CDR3 length class.

value that exceeds the value of the length class proportion of the reference CDR3 length distribution by at least 3 (Killian et al., 2002) or 2.5 (Manfras et al., 2004) standard deviations. To identify recurrent expanded CDR3 length classes, a scoring system that assigns a score value to each CDR3 length class of each CDR3 length distribution identifies the same expansions in a group of patients (Collette and Six, 2002).

To conclude, several methods have been developed to assess differences between CDR3 length distributions. Many take into account the underlying pseudo-normality of CDR3 length distribution. Although a visual analysis is easier for biologists, a negative aspect of this method can arise depending on the relative expertise and consistency from day to day of the scientist responsible for analyzing the data. Visually identifying differences in CDR3 length distribution is difficult and can introduce subjective variation and numerous biases. Methods based on quantitative data, using computers, are better at extracting information from images than human observers (Glasbey et al., 1998). These methods quantify and/or statistically assess differences between CDR3 length distributions, or identify clonal expansions.

4. Global analysis

Once information has been extracted from each CDR3 length distribution, understanding the degree of repertoire bias necessitates integration of the particularities of all CDR3 length distributions at the individual and group level. Different methods have been developed to document changes that could occur in a whole repertoire at different time points (King et al., 2001; Hirokawa et al., 2000) and to assess differences between groups (Gorochov et al., 1998). These methods consist in computing one single score value reflecting the overall complexity or clonality of an individual. A comparison of individuals can then be based on differences in this score value. These scores may be derived from the percentage of polyclonal (Talvensaari et al., 2002) or monoclonal CDR3 length distribution (Orsini et al., 2000), the number of the total length classes present in all the CDR3 length distributions (Soiffer et al., 2002; Wu et al., 2000) or the mean percentages of perturbation (Gorochov et al., 1998). Specific statistics (Kepler et al., 2005) are also good indices of the overall complexity of the repertoire. Certain methods combine different overall scores into a complex formula to take into account the advantages of each score (Peggs et al., 2003; Lu et al., 2004). However, one single score cannot summarize all the information contained in 26 CDR3 length distributions, for a single TCR Vβ chain.

To tackle this problem, data reduction methods such as principal component analysis have been proposed (Collette et al., 2003, 2004; Collette and Six, 2002). This mathematical procedure consists in reducing the number of variables, in this case the number of genes, into a smaller number of uncorrelated variables. Once the "dimensionality" of the dataset has been reduced, each individual is characterized by a small number of variables; and statistical tests can be performed more easily. Correspondence factor analysis can also be used for discrimination purposes, as was used to discriminate the TCR Vβ repertoire of rats with allo- or xeno-transplantation and to identify the TCR Vβ genes associated (Guillet et al., 2004). Correspondence factor analysis is an exploratory technique that can be used to locate the variables and the individuals on a correspondence map. However, correspondence factor analysis cannot statistically assess the involvement of a particular gene in a signature in a rigorous manner. Statistical methods such as multiple analysis of variance have been successfully used on perturbation percentages of TCR Vβ genes in order to find the specific genes possibly associated with a state of tolerance (Brouard et al., 2005). This statistical procedure enables multiple dependent variables (genes) to be analyzed simultaneously.

Challenges still remain in determining the correct analytical method to use for CDR3 length distribution. However, in many cases, the CDR3 length distribution analysis has proven to be a predictive method (Ria et al., 2001) on its own or coupled with other kind of data. Indeed, successful efforts have been made to combine quantitative (Bour et al., 1999; Drescher et al., 2000; Even et al., 1995; Gagne et al., 2000; Lang et al., 1997; Lim et al., 2002) or statistical (Hori et al., 2002) techniques to CDR3 length distribution analysis. The purpose is to give relevance of a CDR3 length distribution pattern by correlating it with the quantitative usage of the corresponding gene. In general, initiatives consisting in combining heterogeneous data types with CDR3 length distribution analyses (qualitative transcriptome analyses), such as cell surface protein quantification by flow cytometry (quantitative proteome analysis) (Cossarizza et al., 2004; Dechanet et al., 1999; Pilch et al., 2002; Yawalkar et al., 2003), functional data (cytokines, mixed lymphocyte reaction), clinical data (Schito et al., 2001) or gene expression profiling are promising to better describe T and B repertoires. Development of statistical tools to analyze various types of data mixed together, will make it possible in the near future to model various immunological situations and to extend CDR3 length distribution analysis to the development of prognostic tests.

Acknowledgements

The authors are grateful to Joanna Ashton-Chess who helped in editing the manuscript and to Prof. El Mostafa Qannari and Catherine Ruiz for their helpful remarks concerning this review.

References

Arden, B., Clark, S.P., Kabelitz, D., Mak, T.W., 1995. Human T-cell receptor variable gene segment families. Immunogenetics 42, 455–500.

Arons, E., Sorbara, L., Raffeld, M., Stetler-Stevenson, M., Steinberg, S.M., Liewehr, D.J., Pastan, I., Kreitman, R.J., 2005. Characterization of T-cell repertoire in hairy cell leukemia patients before and after recombinant immunotoxin BL22 therapy. Cancer Immunol. Immunother., 1–11.

Arstila, T.P., Casrouge, A., Baron, V., Even, J., Kanellopoulos, J., Kourilsky, P., 1999. A direct estimate of the human alphabeta T cell receptor diversity. Science 286, 958–961.

Baranzini, S.E., Jeong, M.C., Butunoi, C., Murray, R.S., Bernard, C.C., Oksenberg, J.R., 1999. B cell repertoire diversity and clonal expansion in multiple sclerosis brain lesions. J. Immunol. 163, 5133–5144.

Baron, C., McMorrow, I., Sachs, D.H., LeGuern, C., 2001. Persistence of dominant T cell clones in accepted solid organ transplants. J. Immunol. 167, 4154–4160.

Benveniste, O., Cherin, P., Maisonobe, T., Merat, R., Chosidow, O., Mouthon, L., Guillevin, L., Flahault, A., Burland, M.C., Klatzmann, D., Herson, S., Boyer, O., 2001. Severe perturbations of the blood T cell repertoire in polymyositis, but not dermatomyositis patients. J. Immunol. 167, 3521–3529.

Bour, H., Puisieux, I., Even, J., Kourilsky, P., Favrot, M., Musette, P., Nicolas, J.F., 1999. T-cell repertoire analysis in chronic plaque psoriasis suggests an antigen-specific immune response. Hum. Immunol. 60, 665–676.

Brouard, S., Dupont, A., Giral, M., Louis, S., Lair, D., Braudeau, C., Degauque, N., Moizant, F., Pallier, A., Ruiz, C., Guillet, M., Laplaud, D., Soulillou, J.P., 2005. Operationally tolerant and minimally immunosuppressed kidney recipients display strongly altered blood T-cell clonal regulation. Am. J. Transplant. 5, 330–340.

Casrouge, A., Beaudoing, E., Dalle, S., Pannetier, C., Kanellopoulos, J., Kourilsky, P., 2000. Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes. J. Immunol. 164, 5782–5787.

Cibotti, R., Cabaniols, J.P., Pannetier, C., Delarbre, C., Vergnon, I., Kanellopoulos, J.M., Kourilsky, P., 1994. Public and private V beta T cell receptor repertoires against hen egg white lysozyme (HEL) in non-transgenic versus HEL transgenic mice. J. Exp. Med. 180, 861–872.

Cochet, M., Pannetier, C., Regnault, A., Darche, S., Leclerc, C., Kourilsky, P., 1992. Molecular detection and in vivo analysis of the specific T cell response to a protein antigen. Eur. J. Immunol. 22, 2639–2647.

Collette, A., Bagot, S., Ferrandiz, M.E., Cazenave, P.A., Six, A., Pied, S., 2004. A profound alteration of blood TCRB repertoire allows prediction of cerebral malaria. J. Immunol. 173, 4568–4575.

Collette, A., Cazenave, P.A., Pied, S., Six, A., 2003. New methods and software tools for high throughput CDR3 spectratyping. Application to T lymphocyte repertoire modifications during experimental malaria. J. Immunol. Methods 278, 105–116.

Collette, A., Six, A., 2002. ISEApeaks: an Excel platform for GeneScan and Immunoscope data retrieval, management and analysis. Bioinformatics 18, 329–330.

Cossarizza, A., Poccia, F., Agrati, C., D'Offizi, G., Bugarini, R., Pinti, M., Borghi, V., Mussini, C., Esposito, R., Ippolito, G., Narciso, P., 2004. Highly active antiretroviral therapy restores CD4+ Vbeta T-cell repertoire in patients with primary acute HIV infection but not in treatment-naive HIV+ patients with severe chronic infection. J. Acquir. Immune Defic. Syndr. 35, 213–222.

Coulie, P.G., Karanikas, V., Colau, D., Lurquin, C., Landry, C., Marchand, M., Dorval, T., Brichard, V., Boon, T., 2001. A monoclonal cytolytic T-lymphocyte response observed in a melanoma patient vaccinated with a tumor-specific antigenic peptide encoded by gene MAGE-3. Proc. Natl. Acad. Sci. U.S.A. 98, 10290–10295.

Davis, M.M., Bjorkman, P.J., 1988. T-cell antigen receptor genes and T-cell recognition. Nature 334, 395–402.

Davis, M.M., Boniface, J.J., Reich, Z., Lyons, D., Hampl, J., Arden, B., Chien, Y., 1998. Ligand recognition by alpha beta T cell receptors. Annu. Rev. Immunol. 16, 523–544.

Dechanet, J., Merville, P., Lim, A., Retiere, C., Pitard, V., Lafarge, X., Michelson, S., Meric, C., Hallet, M.M., Kourilsky, P., Potaux, L., Bonneville, M., Moreau, J.F., 1999. Implication of gammadelta T cells in the human immune response to cytomegalovirus. J. Clin. Invest. 103, 1437–1449.

Demoulins, T., Mouthon, F., Clayette, P., Bequet, D., Gachelin, G., Dormont, D., 2003. The same TCR (N)Dbeta(N)Jbeta junctional region is associated with several different vbeta13 subtypes in a multiple sclerosis patient at the onset of the disease. Neurobiol. Dis. 14, 470–482.

Douillard, P., Josien, R., Pannetier, C., Bonneville, M., Soulillou, J.P., Cuturi, M.C., 1998. Selection of T cell clones with restricted TCR–CDR3 lengths during in vitro and in vivo alloresponses. Int. Immunol. 10, 71–83.

Drescher, K.M., Johnston, S.L., Hogancamp, W., Nabozny, G.H., David, C.S., Rimm, I.J., Wettstein, P.J., Rodriguez, M., 2000. V(beta)8(+) T cells protect from demyelinating disease in a viral model of multiple sclerosis. Int. Immunol. 12, 271–280.

Even, J., Lim, A., Puisieux, L., Ferradini, L., Dietrich, P.Y., Toubert, A., Hercend, T., Triebel, F., Pannetier, C., Kourilsky, P., 1995. T-cell repertoires in healthy and diseased human tissues analysed by T-cell receptor beta-chain CDR3 size determination: evidence for oligoclonal expansions in tumours and inflammatory diseases. Res. Immunol. 146, 65–80.

Folch, G., Lefranc, M.P., 2000. The human T cell receptor beta variable (TRBV) genes. Exp. Clin. Immunogenet. 17, 42–54.

Friedman, T.M., Varadi, G., Hopely, D.D., Filicko, J., Wagner, J., Ferber, A., Martinez, J., Brunner, J., Grosso, D., McGuire, L., Korngold, R., Flomenberg, N., 2001. Non-myeloablative conditioning allows for more rapid T-cell repertoire reconstitution following allogeneic matched unrelated bone marrow transplantation compared to myeloablative approaches. Biol. Blood Marrow Transplant. 7, 656–664.

Gagne, K., Brouard, S., Giral, M., Sebille, F., Moreau, A., Guillet, M., Bignon, J.D., Imbert, B.M., Cuturi, M.C., Soulillou, J.P., 2000. Highly altered V beta repertoire of T cells infiltrating long-term rejected kidney allografts. J. Immunol. 164, 1553–1563.

Gapin, L., Fukui, Y., Kanellopoulos, J., Sano, T., Casrouge, A., Malier, V., Beaudoing, E., Gautheret, D., Claverie, J.M., Sasazuki, T., Kourilsky, P., 1998. Quantitative analysis of the T cell repertoire selected by a single peptide-major histocompatibility complex. J. Exp. Med. 187, 1871–1883.

Glasbey, C.A., Berman, M., Hudson, H.M., 1998. Image analysis and tomography. In: Armitage, P., Colton e., T. (Eds.), Encyclopaedia of Biostatistics. Wiley, Chichester, pp. 1980–1998.

Gorochov, G., Neumann, A.U., Kereveur, A., Parizot, C., Li, T., Katlama, C., Karmochkine, M., Raguin, G., Autran, B., Debre, P., 1998. Perturbation of CD4+ and CD8+ T-cell repertoires during progression to AIDS and regulation of the CD4+ repertoire during antiviral therapy. Nat. Med. 4, 215–221.

Gorski, L., Yassai, M., Zhu, X., Kissela, B., Kissella, B., Keever, C., Flomenberg, N., 1994. Circulating T cell repertoire complexity in normal individuals and bone marrow recipients analyzed by CDR3 size spectratyping. Correlation with immune status. J. Immunol. 152, 5109–5119.

Guilherme, L., Dulphy, N., Douay, C., Coelho, V., Cunha-Neto, E., Oshiro, S.E., Assis, R.V., Tanaka, A.C., Pomerantzeff, P.M., Charron, D., Toubert, A., Kalil, J., 2000. Molecular evidence for antigen-driven immune responses in cardiac lesions of rheumatic heart disease patients. Int. Immunol. 12, 1063–1074.

Guillet, M., Gagne, K., Lair, D., Heslan, J.M., Dore, J.C., Soulillou, J.P., Brouard, S., 2004. Different patterns of TCR beta chain regulation following allo- and xeno-transplantation. Xenotransplantation 11, 315–322.

Han, M., Harrison, L., Kehn, P., Stevenson, K., Currier, J., Robinson, M.A., 1999. Invariant or highly conserved TCR alpha are expressed on double-negative (CD3+CD4−CD8−) and CD8+ T cells. J. Immunol. 163, 301–311.

Hirokawa, M., Horiuchi, T., Kitabayashi, A., Kawabata, Y., Matsutani, T., Suzuki, R., Chihara, J., Miura, A.B., 2000. Delayed recovery of CDR3 complexity of the T-cell receptor-beta chain in recipients of allogeneic bone marrow transplants who had virus-associated interstitial pneumonia: monitor of T-cell function by CDR3 spectratyping. J. Allergy Clin. Immunol. 106, S32–S39.

Hori, S., Collette, A., Demengeot, J., Stewart, J., 2002. A new statistical method for quantitative analyses: application to the precise quantification of T cell receptor repertoires. J. Immunol. Methods 268, 159–170.

Hsieh, C.S., Liang, Y., Tyznik, A.J., Self, S.G., Liggitt, D., Rudensky, A.Y., 2004. Recognition of the peripheral self by naturally arising CD25+ CD4+ T cell receptors. Immunity 21, 267–277.

Hviid, L., Akanmori, B.D., Loizon, S., Kurtzhals, J.A., Ricke, C.H., Lim, A., Koram, K.A., Nkrumah, F.K., Mercereau-Puijalon, O., Behr, C., 2000. High frequency of circulating gamma delta T cells with dominance of the V(delta)l subset in a healthy population. Int. Immunol. 12, 797–805.

Karadimitris, A., Manavalan, J.S., Thaler, H.T., Notaro, R., Araten, D.J., Nafa, K., Roberts, I.A., Weksler, M.E., Luzzatto, L., 2000. Abnormal T-cell repertoire is consistent with immune process underlying the pathogenesis of paroxysmal nocturnal hemoglobinuria. Blood 96, 2613–2620.

Kepler, T.B., He, M., Tomfohr, J.K., Devlin, B.H., Sarzotti, M., Markert, M.L., 2005. Statistical analysis of antigen receptor spectratype data. Bioinformatics 21, 3394–3400.

Killian, M.S., Matud, J., Detels, R., Giorgi, J.V., Jamieson, B.D., 2002. MaGiK method of T-cell receptor repertoire analysis. Clin. Diagn. Lab. Immunol. 9, 858–863.

King, D.J., Gotch, F.M., Larsson-Sciard, E.L., 2001. T-cell re-population in HIV-infected children on highly active anti-retroviral therapy (HAART). Clin. Exp. Immunol. 125, 447–454.

Kurosawa, Y., Tonegawa, S., 1982. Organization, structure, and assembly of immunoglobulin heavy chain diversity DNA segments. J. Exp. Med. 155, 201–218.

Lang, R., Pfeffer, K., Wagner, H., Heeg, K., 1997. A rapid method for semi-quantitative analysis of the human V beta-repertoire using TaqManR PCR. J. Immunol. Methods 203, 181–192.

Laouini, D., Casrouge, A., Dalle, S., Lemonnier, F., Kourilsky, P., Kanellopoulos, J., 2000. V beta T cell repertoire of CD8+ splenocytes selected on non-polymorphic MHC class I molecules. J. Immunol. 165, 6381–6386.

Laplaud, D.A., Ruiz, C., Wiertlewski, S., Brouard, S., Berthelot, L., Guillet, M., Melchior, B., Degauque, N., Edan, G., Brachet, P., Damier, P., Soulillou, J.P., 2004. Blood T-cell receptor beta chain transcriptome in multiple sclerosis. Characterization of the T cells with altered CDR3 length distribution. Brain 127, 981–995.

Lefranc, M.P., 2003. IMGT, the international ImMunoGeneTics database. Nucleic Acids Res. 31, 307–310.

Lim, A., Baron, V., Ferradini, L., Bonneville, M., Kourilsky, P., Pannetier, C., 2002. Combination of MHC-peptide multimer-based T cell sorting with the immunoscope permits sensitive ex vivo quantitation and follow-up of human CD8+ T cell immune responses. J. Immunol. Methods 261, 177–194.

Liu, D., Callahan, J.P., Dau, P.C., 1995. Intrafamily fragment analysis of the T cell receptor beta chain CDR3 region. J. Immunol. Methods 187, 139–150.

Lu, J., Basu, A., Melenhorst, J.J., Young, N.S., Brown, K.E., 2004. Analysis of T-cell repertoire in hepatitis-associated aplastic anemia. Blood 103, 4588–4593.

Manfras, B.J., Weidenbach, H., Beckh, K.H., Kern, P., Moller, P., Adler, G., Mertens, T., Boehm, B.O., 2004. Oligoclonal CD8+ T-cell expansion in patients with chronic hepatitis C is associated with liver pathology and poor response to interferon-alpha therapy. J. Clin. Immunol. 24, 258–271.

Mosley, R.L., Koker, M.M., Miller, R.A., 1998. Idiosyncratic alterations of TCR size distributions affecting both CD4 and CD8 T cell subsets in aging mice. Cell. Immunol. 189, 10–18.

Muraro, P.A., Douek, D.C., Packer, A., Chung, K., Guenaga, F.J., Cassiani-Ingoni, R., Campbell, C., Memon, S., Nagle, J.W., Hakim, F.T., Gress, R.E., McFarland, H.F., Burt, R.K., Martin, R., 2005. Thymic output generates a new and diverse TCR repertoire after autologous stem cell transplantation in multiple sclerosis patients. J. Exp. Med. 201, 805–816.

Musette, P., Galelli, A., Truffa-Bachi, P., Peumans, W., Kourilsky, P., Gachelin, G., 1996. The J beta segment of the T cell receptor contributes to the V beta-specific T cell expansion caused by staphylococcal enterotoxin B and *Urtica dioica* superantigens. Eur. J. Immunol. 26, 618–622.

Naumov, Y.N., Naumova, E.N., Gorski, J., 1996. CD4+ and CD8+ circulating alpha/beta T-cell repertoires are equally complex and are characterized by different levels of steady-state TCR expression. Hum. Immunol. 48, 52–62.

Orsini, E., Alyea, E.P., Schlossman, R., Canning, C., Soiffer, R.J., Chillemi, A., Neuberg, D., Anderson, K.C., Ritz, J., 2000. Changes in T cell receptor repertoire associated with graft-versus-tumor effect and graft-versus-host disease in patients with relapsed multiple myeloma after donor lymphocyte infusion. Bone Marrow Transplant. 25, 623–632.

Pannetier, C., Cochet, M., Darche, S., Casrouge, A., Zoller, M., Kourilsky, P., 1993. The sizes of the CDR3 hypervariable regions of the murine T-cell receptor beta chains vary as a function of the recombined germ-line segments. Proc. Natl. Acad. Sci. U.S.A. 90, 4319–4323.

Pannetier, C., Even, J., Kourilsky, P., 1995. T-cell repertoire diversity and clonal expansions in normal and clinical samples. Immunol. Today 16, 176–181.

Peggs, K.S., Verfuerth, S., D'Sa, S., Yong, K., Mackinnon, S., 2003. Assessing diversity: immune reconstitution and T-cell receptor BV spectratype analysis following stem cell transplantation. Br. J. Haematol. 120, 154–165.

Pilch, H., Hohn, H., Freitag, K., Neukirch, C., Necker, A., Haddad, P., Tanner, B., Knapstein, P.G., Maeurer, M.J., 2002. Improved assessment of T-cell receptor (TCR) VB repertoire in clinical specimens: combination of TCR–CDR3 spectratyping with flow cytometry-based TCR VB frequency analysis. Clin. Diagn. Lab. Immunol. 9, 257–266.

Ria, F., van den Elzen, P., Madakamutil, L.T., Miller, J.E., Maverakis, E., Sercarz, E.E., 2001. Molecular characterization of the T cell repertoire using immunoscope analysis and its possible implementation in clinical practice. Curr. Mol. Med. 1, 297–304.

Risitano, A.M., Kook, H., Zeng, W., Chen, G., Young, N.S., Maciejewski, J.P., 2002. Oligoclonal and polyclonal CD4 and CD8 lymphocytes in aplastic anemia and paroxysmal nocturnal hemoglobinuria measured by V beta CDR3 spectratyping and flow cytometry. Blood 100, 178–183.

Rowen, L., Koop, B.F., Hood, L., 1996. The complete 685-kilobase DNA sequence of the human beta T cell receptor locus. Science 272, 1755–1762.

Sarzotti, M., Patel, D.D., Li, X., Ozaki, D.A., Cao, S., Langdon, S., Parrott, R.E., Coyne, K., Buckley, R.H., 2003. T cell repertoire development in humans with SCID after non-ablative allogeneic marrow transplantation. J. Immunol. 170, 2711–2718.

Schito, A.M., Vittinghoff, E., Hecht, F.M., Elkins, M.K., Kahn, J.O., Levy, J.A., Oksenberg, J.R., 2001. Longitudinal analysis of T-cell receptor gene use by CD8(+) T cells in early human immunodeficiency virus infection in patients receiving highly active antiretroviral therapy. Blood 97, 214–220.

Sebille, F., Gagne, K., Guillet, M., Degauque, N., Pallier, A., Brouard, S., Vanhove, B., Delsuc, M.A., Soulillou, J.P., 2001. Direct recognition of foreign MHC determinants by naive T cells mobilizes specific Vbeta families without skewing of the complementarity-determining region 3 length distribution. J. Immunol. 167, 3082–3088.

Soiffer, R.J., Alyea, E.P., Hochberg, E., Wu, C., Canning, C., Parikh, B., Zahrieh, D., Webb, I., Antin, J., Ritz, J., 2002. Randomized trial of CD8+ T-cell depletion in the prevention of graft-versus-host disease associated with donor lymphocyte infusion. Biol. Blood Marrow Transplant. 8, 625–632.

Talvensaari, K., Clave, E., Douay, C., Rabian, C., Garderet, L., Busson, M., Gamier, F., Douek, D., Gluckman, E., Charron, D., Toubert, A., 2002. A broad T-cell repertoire diversity and an efficient thymic function indicate a favorable long-term immune reconstitution after cord blood stem cell transplantation. Blood 99, 1458–1464.

Tonegawa, S., 1983. Somatic generation of antibody diversity. Nature 302, 575–581.

Tsutsumi, Y., Tanaka, J., Miura, Y., Toubai, T., Kato, N., Fujisaw, F., Toyoshima, N., Ota, S., Mori, A., Yonezumi, M., Chiba, K., Kondo, T., Hasino, S., Kobayasi, R., Masauji, N., Kasai, M., Asaka, M., Imamura, M., 2004. Molecular analysis of T-cell repertoire in patients with graft-versus-host disease after allogeneic stem cell transplantation. Leukemia Lymphoma 45, 481–488.

Tully, G., Kortsik, C., Hohn, H., Zehbe, I., Hitzler, W.E., Neukirch, C., Freitag, K., Kayser, K., Maeurer, M.J., 2005. Highly focused T cell responses in latent human pulmonary *Mycobacterium tuberculosis* infection. J. Immunol. 174, 2174–2184.

Wei, S., Charmley, P., Robinson, M.A., Concannon, P., 1994. The extent of the human germline T-cell receptor V beta gene segment repertoire. Immunogenetics 40, 27–36.

Wilson, R.K., Lai, E., Concannon, P., Barth, R.K., Hood, L.E., 1988. Structure, organization and polymorphism of murine and human T-cell receptor alpha and beta chain gene families. Immunol. Rev. 101, 149–172.

Wu, C.J., Chillemi, A., Alyea, E.P., Orsini, E., Neuberg, D., Soiffer, R.J., Ritz, J., 2000. Reconstitution of T-cell receptor repertoire diversity following T-cell depleted allogeneic bone marrow transplantation is related to hematopoietic chimerism. Blood 95, 352–359.

Wu, L.C., Tuot, D.S., Lyons, D.S., Garcia, K.C., Davis, M.M., 2002. Two-step binding mechanism for T-cell receptor recognition of peptide MHC. Nature 418, 552–556.

Yawalkar, N., Ferenczi, K., Jones, D.A., Yamanaka, K., Suh, K.Y., Sadat, S., Kupper, T.S., 2003. Profound loss of T-cell receptor repertoire complexity in cutaneous T-cell lymphoma. Blood 102, 4059–4066.