Human Antibodies 12 (2003) 113–121
IOS Press

Immunoglobulin repertoire of B lymphocytes infiltrating breast medullary carcinoma

B. Kotlan[a,b,c,*], P. Simsa[a], J. Foldi[a], W.H. Fridman[b], M. Glassy[c,d], M. Mc Knight[c,d] and JL Teillaud[b]
[a]*National Medical Center, Institute of Haematology and Immunology, Budapest, Hungary*
[b]*INSERM U255, Centre de Recherches Biomedicales des Cordeliers, Paris, France*
[c]*Rajko Medenica Research Foundation, San Diego, USA*
[d]*Shantha West Inc, San Diego, USA,*

Abstract. Tumor specific peptides recognized by T lymphocytes infiltrating solid tumors, as well as the corresponding T cell receptor (TcR) repertoire usage, have been extensively investigated. By contrast, tumor infiltrating B cells and their immunoglobulin (Ig) repertoire have been studied only in a limited number of tumors. The objective of the present study was to determine, whether DNA sequence analysis of the expressed immunoglobulin variable regions of B cells that infiltrate breast cancer, could be used to reveal a potential specific tumor binding capacity of the antibodies. To answer this question, about 200 expressed Ig heavy (VH) and light chain variable gene (VL) regions were cloned, sequenced and comparatively analysed from a typical medullary beast carcinoma (MBC), where the massive B and plasma cell infiltration correlates with favourable prognosis despite of its high grade. The tumor infiltrating B cell Ig heavy and light chain sequences could be classified into clusters, families and subgroups, based on the identity level to germline, showing a pattern of oligoclonality. Some overrepresented clusters could be determined. In the course of a detailed analysis and search in Blastn database, a number of VH and VL sequences showed more than 99% homology to DNA sequences of Ig VH region, with proved tumor antigen binding capacity. Our data suggest, that potential tumor binder Ig VH and VL sequences might be selected using a detailed immunoglobulin variable region analysis. This new approach might have a benefit for further antibody engineering, as difficulties in search for tumor binders by phage library selection might be reduced and the time for selection shortened.

Keywords: Immunoglobulin repertoire, breast medullary carcinoma, tumor infiltrating B lymphocytes

1. Introduction

T cell receptors (TCR) of tumor infiltrating T cells have been extensively characterized in different types of solid tumors [1,2]. The TCR gene usage has been characterised in a great variety of tumors [3,4], and specific tumor related antigens could be defined by T cell cloning. However, much less attention was paid up to now for the tumor infiltrating B cells, although some interesting findings were achieved in melanoma [5] and some other tumor types [small lung carcinoma [6], colon cancer [7]]. The expressed human B-cell immunoglobulin repertoire in normal fetal [8], newborn [9] and adults [10,11] aswell as that in autoimmune diseases [12] and B cell malignancies [13] were mainly investigated. Guigou V and collegues, 1990 [14] made the first analysis of the expression pattern of Ig VH and V_{κ} families in human adult normal peripheral blood B lymphocytes, thus providing a reference basis for the follow-up of the aquisition of the Ig repertoire in physiological [15,16] and pathological [17] situations. Fetal B lymphocyte repertoire analysis showed a preferential expression of different immunoglobulin genes [8]. V-region repertoire in B chronic lymphocytic leukemia (B-CLL) exhibited considerable diversity, but an antigen-driven stimulation

*Corresponding author: Dr Beatrix Kotlan, National Medical Center, Institute of Haematology and Immunology, Budapest 1113 Dioszegi Street 64, P.O.B. H- 1519 Hungary. Tel.: +36 1 372 43 61; Fax: +36 1 398 06 39, 36 1 466 70 20; E-mail: KotlanB@netscape.net.

ISSN 1093-2607/03/$8.00

and expansion was suggested [18]. Results suggested an antigen-driven response in the generation of IgG4 rheumatoid factor in rheumatoid arthritis disease processes [19].

Only a few data have been accumulated about the immunoglobulin variable gene usage by tumor infiltrating B cells [20–23]. Despite that these data suggested an oligoclonality in two tumor types (breast cancer and ovarian cancer), no selectively tumor-specific binder immunoglobulin variable regions have been reported.

A rare breast tumor type was choosen for our investigations, because of it's strong B lymphocytic and plasmocytic infiltration [24,25]. Human Ig variable regions were amplified specifically with the help of degenerated oligonucleotide PCR primers [26,27] by RT- polymerase chain reaction (RT-PCR). After cloning and sequencing, a detailed comparative DNA sequence analysis was performed using available germline and expressed gene databases. The main characteristics of the immunoglobulin repertoire of B cells infiltrating medullary breast carcinoma was defined and helps determining VH regions that are likely regions with a potential tumor binding capacity.

2. Materials and methods

2.1. Tumor sampling, amplification and sequencing of TIL-B Ig VH and VL

A small piece of tumor tissue (170 μg) exhibiting a strong B cell infiltration by histology was obtained from a tumor removed surgically from a patient with a medullary breast carcinoma. After RNA extraction and cDNA synthesis, VH-JH, $V_\kappa - J_\kappa$ and V_λ-J_λ regions were amplified by PCR using mixtures of primers according to our method described earlier [20]. The immunoglobulin heavy and light chain expressed regions were cloned into pUC18 (SmaI/BAP) / E. coli TG1, and checked for insert content by standard PCR screen method. Altogether, 200 Ig VH and VL inserts were sequenced (Dye Terminator Sequence Reaction Kit and DyeEx Spin kit (Qiagen, Hilden, Germany, ABI PRISM Software and Perkin Elmer Automatic sequencer (Perkin Elmer, New Jersey, USA).

2.2. Process of comparative DNA sequence analysis

First, a DNAplot analysis was carried out with the determined VH-JH, $V_\kappa - J_\kappa$ and V_λ-J_λ region sequences using the IMGT database [28] in order to define the closest germline homology at the variable (V), diversity (D) and join (J) regions. The sequences were than ranked in families and clusters. The family classification was based on the IMGT determination, while the clusterisation was made with the help of V-D-J conformity. Inside the clusters, immunoglobulins sharing the same mutations compose clones.

The sequences were edited and processed with BIOEDIT 5.0.9. sequence editor [29]. For whole lenght comparison, a compiled germline sequence was constructed putting the found V, D and J genes together into one sequence. Multiple sequence alignments were made with ClustalX 1.8 software [30] by clusters to the compiled germline and also among the family members. A phylogenetic tree was made after alignment, and visualised thereafter by TREEVIEW 1.5.2. [31]. An identity matrix was composed to each cluster, which reveals the homology rates. The basic statistical counts were made by EXCEL.

To each well-defined sequence, a homology search was accomplished to find the closest VH and VL sequence homologies with known antigen specificity. Our sequences were compared to KABAT NIH (http://immuno.bme.nwu.edu), to GenBank, and Embnet via NCBI Blastn Engine (http://www.ncbi.nlm.nih.gov/BLAST, http://www.srs.hgmp.mrc.ac.uk). This querry result is termed as "Blastn result".

3. Results

3.1. Defining representative families and clusters of TIL-B expressed immunoglobulin variable regions

Hundered eleven expressed immunglobulin variable regions originating from tumor infiltrating B cells accumulated in one medullary breast carcinoma have been cloned and sequenced. The comparative immunoglobulin repertoire analysis was performed using DNA sequence data analysis softwares (BIOEDIT, Clustal x) and databases accessible through the internet (Kabat NIH, IMGT, NCBI Blastn, Genbank). Our data analysis process is depicted in Fig. 1.

After data proccessing, representative sequences of four VH families (VH3, VH5, VH4, VH1) could be defined. Both type of light chain were present in the tumor infiltrating B cell repertoire [four VK families (V_κ1, V_κ4, V_κ3, V_κ2) and three VL families (V_λ1, V_λ2, V_λ3)] in a ratio of 3 : 2 to κ versus λ. Reffering to $V_\kappa$chains, V_κ1 family outnumbered families V_κ4,

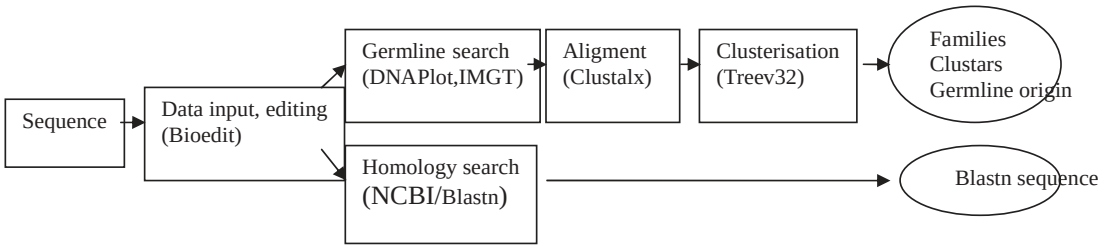

Fig. 1. *Flow chart of comparative sequence data analysis*. The different steps of the analysis are shown (from the processing of row DNA sequence data to the search for already known specificity).

$V_\kappa 3$ and $V_\kappa 2$, as suspected. In the case of V_λ chains, $V_\lambda 1$, $V_\lambda 2$, $V_\lambda 3$ families showed nearly equal numbers of representative members (Table 1). According to the germline sequence comparison through the IMGT database, some distinct clusters could be set-up, based on distinguisable and identical sequences belonging to a given family. Some clusters were overrepresented, while the others had just a few members or were found to be single. As expected in the VH group, the VH3 family was the most abundant, and family 4, 5 and 1 had a lower number of representatives. We found over-represented clusters in the VH3 and the VH5 families, with 11 and 10 sequences respectively, (Figs 2 and 3). The comparison blocks of the over-represented clusters from $V_\kappa 1$ and $V_\lambda 3$ families are shown in Figs 4 and 5 respectively, and demonstrate a high homology to the germline sequences.

3.2. Mutational pattern of TIL-B Ig VH and VL sequences

The identity level among the cluster members (called internal homology) was very high ($< 95\%$). Similarly, the homology level to the germline sequences was in general also high ($< 90\%$). Mismatches could be determined randomly but usually the different gene joining regions were concerned with deletion and / or some mutations. No mutation accumulation in the CDR regions could be revealed in most investigated cases. However, in few cases, the deviation to germline was found to be very high. Since the clusterisation went by V-(D)-J conformity, it sometimes happened that one V gene was being used in more then one cluster.

The VH3/ 1 cluster seems to be highly interesting, as the homology to the germline was only approximately 80%, while the internal identity level was higher than 98%. The mismatches were situated steadily along the VH sequence framework (FR) and complementarity determining regions (CDR) and do not gather only at the CDR regions (Fig. 2). Only one cluster could be defined when analyzing the VH5 family, the identity value of which was 99%. The scatter of mutations is similar to the general distribution found within the other clusters (Fig. 3).

3.3. Blastn query search for finding identical or highly homologous expressed VH and VL sequences with defined specificity

In order to reveal any potential VH specificity, we looked for homologous sequences with defined specificity in the known immunoglobulin databases. As a result for this query, several matches could be obtained. The best fit homologous sequence with verified specificity was the blastn result. Figure 2 represent the resulted query sequence with closest homology level. Several different specificities were found, that derived mostly from autoimmune diseases or B lymphocytes from healthy donors. However in the case of the VH3 cluster (Fig. 2), a sequence that was more than 99% identical to the VH sequence of an antibody directed against a defined type of ganglioside was found. This blastn result sequence appeared to be very close to our cluster, and it's homology to the nearest germline was low.

4. Discussion

Our study represents a detailed immunoglobulin repertoire analysis at the DNA level of B lymphocytes infiltrating a medullary breast carcinoma. As suspected, data in this field might provide important responses for yet unanswered questions. Our findings show a restricted immunoglobulin variable gene region usage, and characteristics of oligoclonality, as suggested in our early study [20]. The results presented herein are in accordance to the findings of Coronella et al. [21] and Hansen et al. [22] who investigated TIL-B immunglobulin repertoire in other medullary breast

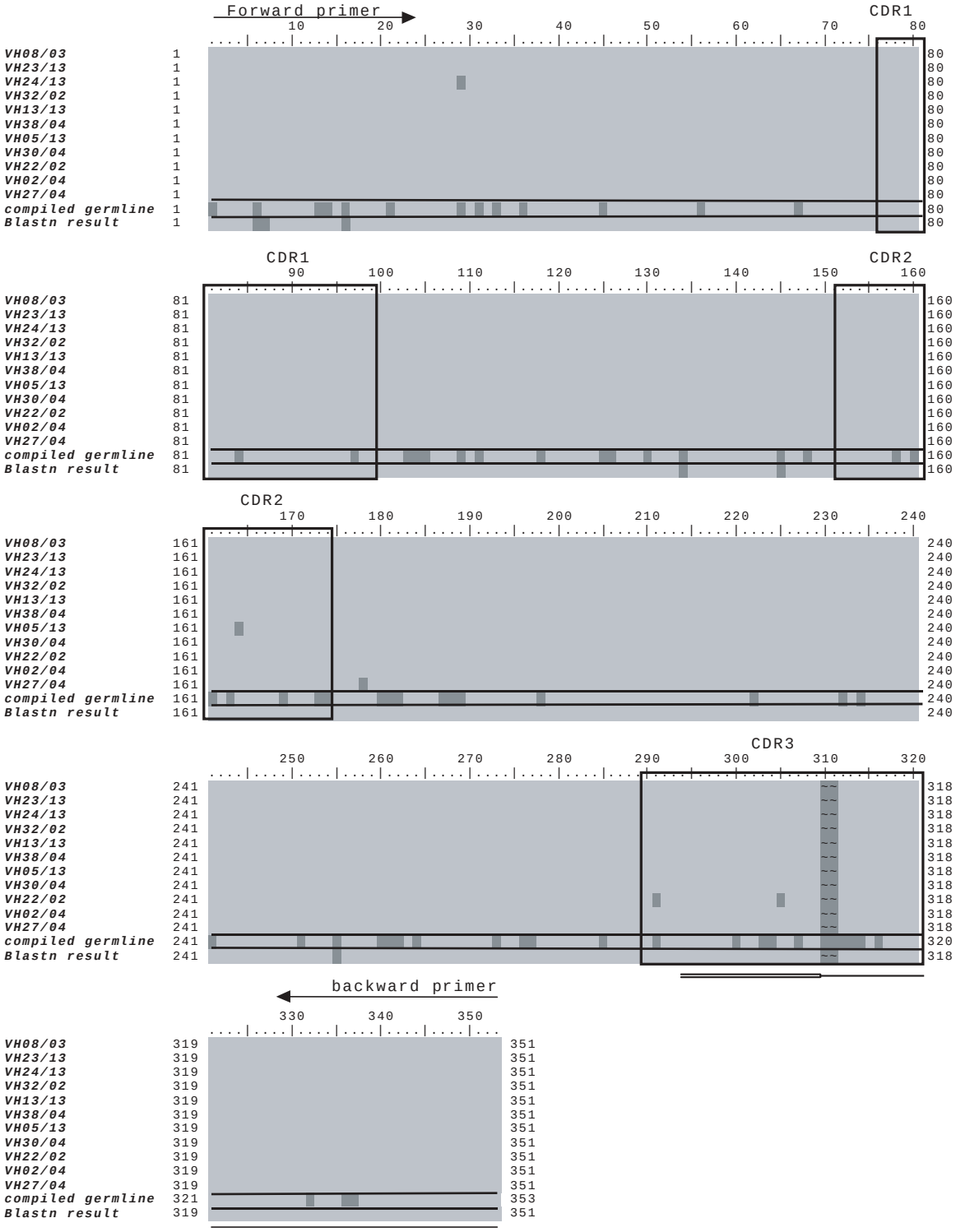

Fig. 2. *Comparison block of the overrepresented clusters from VH3 family.* The largest VH3 cluster is composed from VH3-D4-J4 segments. The cloned sequences are compared to germline sequence data from IMGT and a Blastn query result sequence with known specificities (see in the text). The mismatches are shown in dark grey. The CDR regions are marked and boxed. The arrows show the primers used for cloning. The D segments are boxed with datched lines.

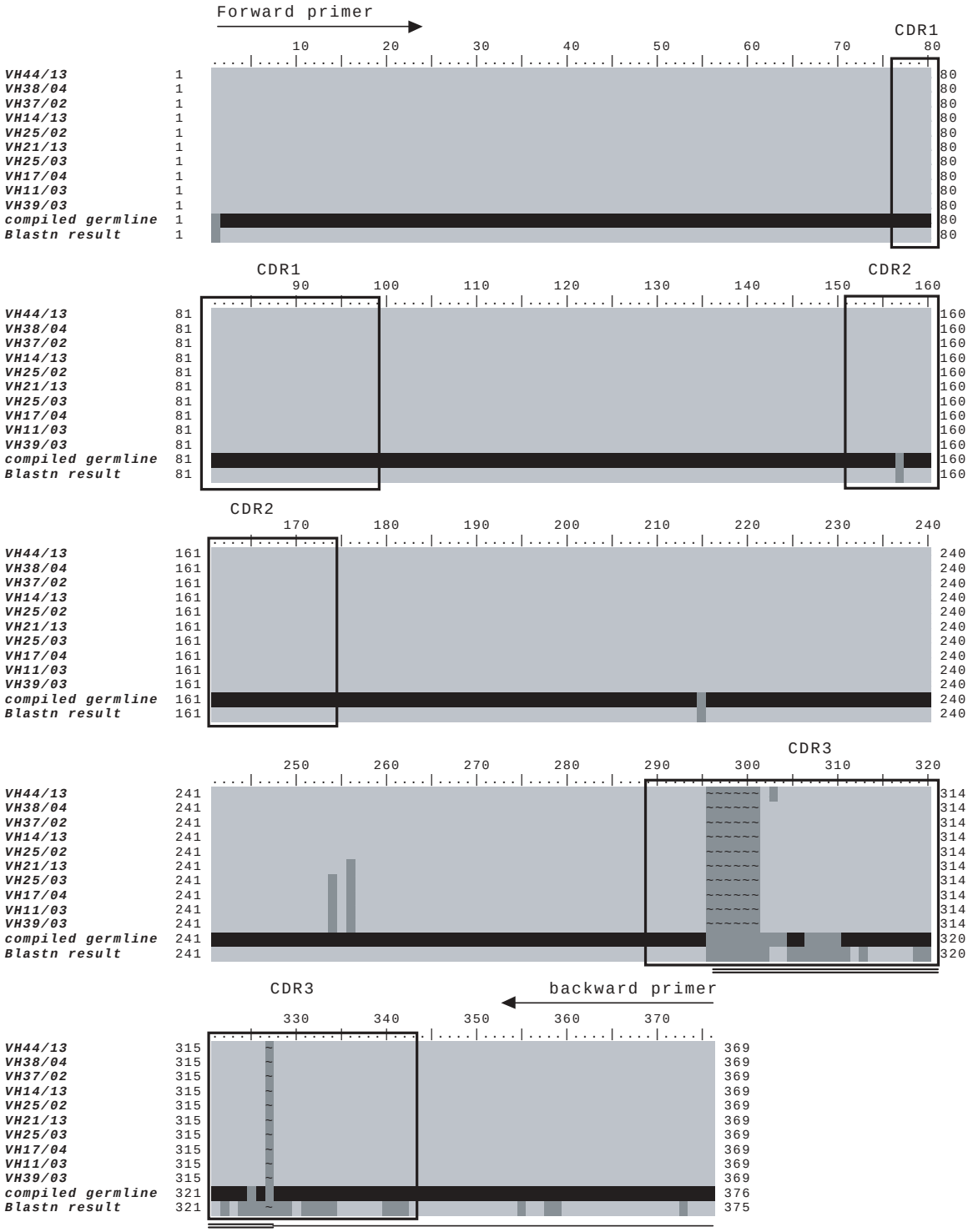

Fig. 3. *Comparison block of another cluster of interest, the VH5.* This VH5 cluster was found as the only cluster of VH5 family with a remarkable number of members. It is composed from VH5-D3-J3 segments. The cloned sequences are compared to germline from IMGT and a Blastn query result sequence with a known specificity (see in the text). The mismatches are shown in dark grey. The CDR regions are marked and boxed. The arrows show the primers used for cloning. The D segments are boxed with datched lines. Some nucleotide deletions occured at both end-joining regions. The D segment of the Blastn result sequence is different, which makes a highly distinct CDR3 region.

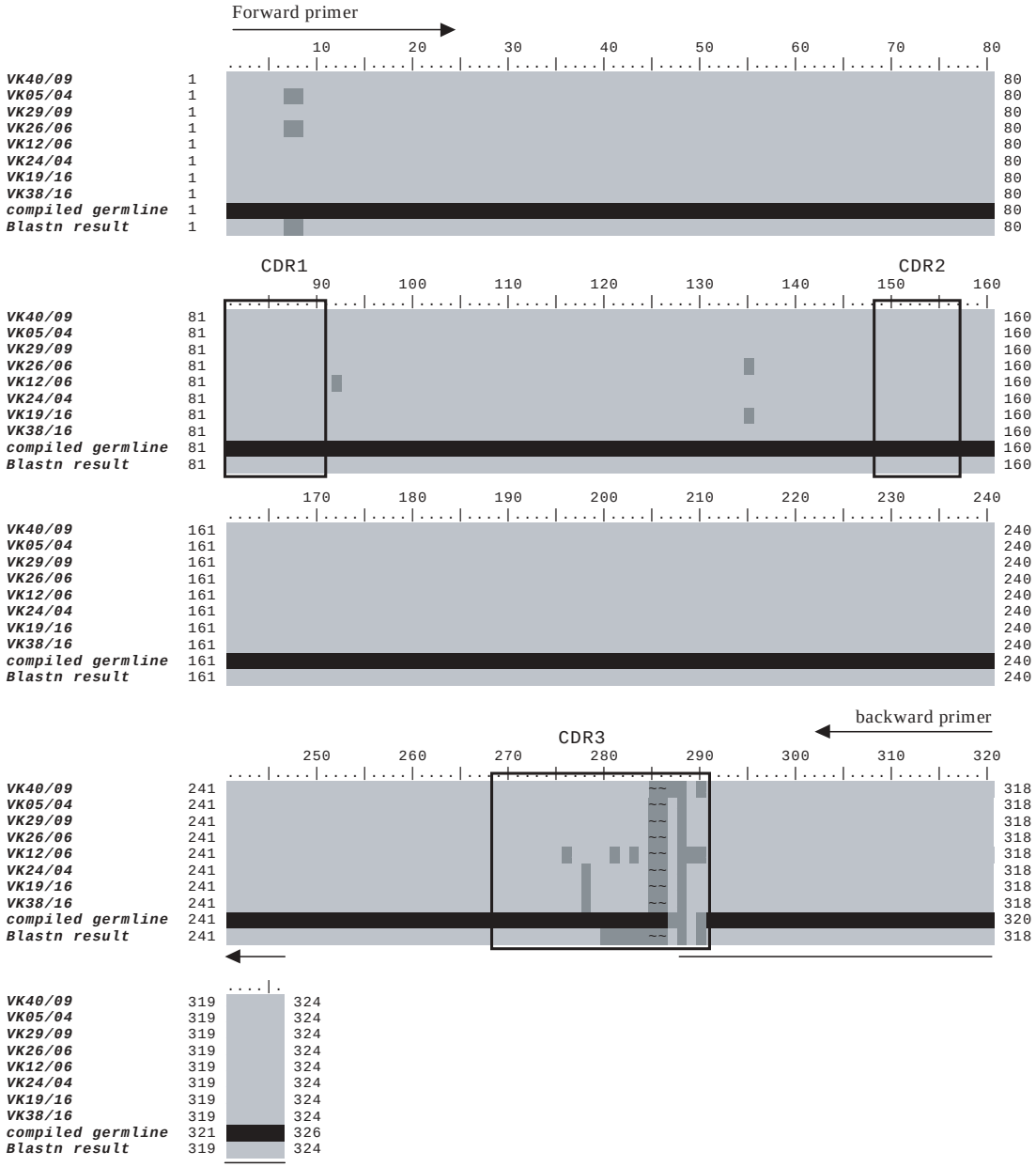

Fig. 4. Comparison block of the over-represented cluster among $V_\kappa 1$ family. The largest cluster among V_κ family is composed of V_κ regions derived from $V_\kappa 1$ and $J_\kappa 1$ genes. The cloned sequences are compared to germline from IMGT and a Blastn query result sequence with a known specificity. The mismatches are shown in dark grey. The CDR regions are marked and boxed. The arrows show the primers used for cloning. The J segment is underlined with a datched line. The homology shows very close relation to the germline sequence. Some mutations and deletions occur at the V-J joint.

carcinoma following our first report. Despite they were also able to define an oligoclonal presence of the immunoglobulins [22], and postulated an antigen – driven selection [21,32], the findings up to know lacked of selective tumor specific targets.

In our present work, out of the 111 DNA sequences of expressed immunglobulin variable regions originating from TIL-B cells in a typical medullary breast carinoma, four VH families, four V_κ families and three V_λ families were found. Among the families, several distinct clusters could be distinguished, amongst which some showed overexpression. The internal homologies

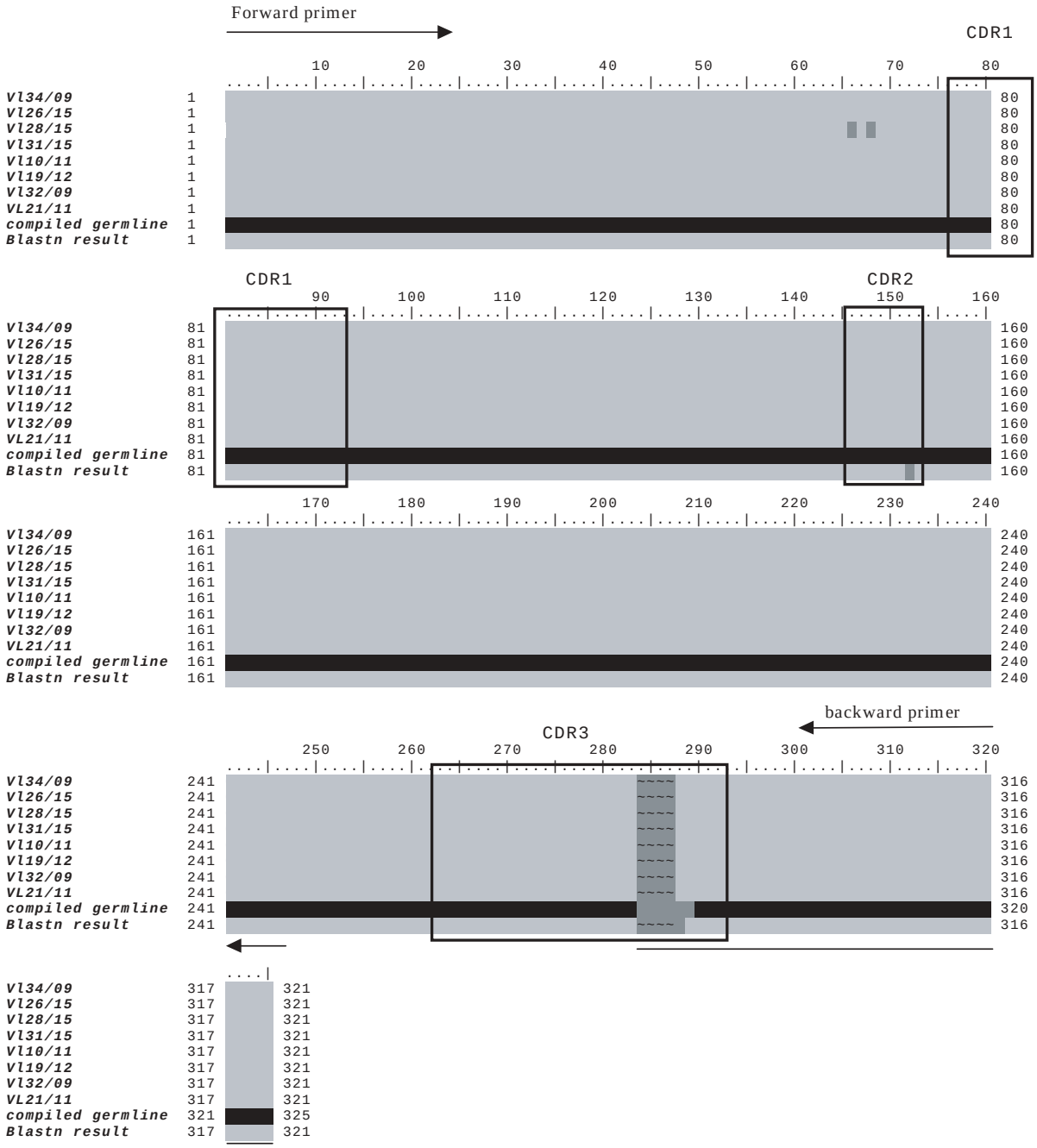

Fig. 5. *Comparison block of a Vλ3 subgroup.* This $V_\lambda 3$ subgroup is composed of $V_\lambda$regions derived from $V_\lambda 3$ and $J_\lambda 1$ genes. The cloned sequences are compared to germline from IMGT and a Blastn query result sequence with a known specificity. The mismatches are shown in dark grey. The CDR regions are marked and boxed. The arrows show the primers used for cloning. The J segment is underlined with a datched line. The homology shows very close relation to the germline sequence. Some mutations and deletions occur at the V-J joint.

within the clusters were very high, and the homology of the expressed VH and VL sequences to the germline was generally also high. The high internal homologies suggest that the mutations observed occur by chance and there was a clonal maturation among them. The high homology to germline sequences propose that a large number of the cloned immunoglobulins might be derived mostly from early stage B1 lymphocytes or still

Table 1
Evaluation of cloned sequences. The families usually consist of several clusters. Summarised data of the TIL-B VH and VL clone sequences' evaluation. The numbers of evaluated sequences and defined clusters in each family are shown. By comparative data analysis the everage identity level and number of germline segments are given

Family	N°-111	N° of clusters	Average identity		N° of used germline V/D/J segment
			Internal	germline	
VH1	2	1	100	93	1/1/1
VH3	24	8(2 large/6 single)	96	94	6/6/6
VH4	11	4	99	97	4/3/3
VH5	10	1	99	95	1/1/1
$V_\kappa 1$	19	7 (1 large/4 single)	97	95	5/4
$V_\kappa 2$	7	3 (1 single)	98	97	2/2
$V_\kappa 3$	1	1	–	89	1/1
$V_\kappa 4$	11	3 (1 large/1 single)	97	96	2/2
$V_\lambda 1$	7	2 (1 single)	93	95	2/1
$V_\lambda 2$	10	1	96	95	1/1
$V_\lambda 3$	9	2 (1 single)	97	96	2/2

haven't entered the hypermutation process. The most over-represented clusters found both among the heavy and light chain families seem to be of high interest, as they might exhibit potential anti-tumor specificities. Although in most of the sequences, no mutation accumulation was found in the complementary determining region genes, in few cases, the deviation to germline was found to be very high. The observation that the mismatches were situated not only along the complementarity determining regions but on the framework regions aswell opens further explanations for mutational pattern. Our results foster the gene selection theory, as in the course of clusterisation by V-(D)-J conformity, one gene was being used sometimes in more than one cluster.

As a result of our Blastn Immunglobulin database query, we got several matches to antibody VH sequences with defined specificity. Most of the different specificities found, corresponded to antibodies isolated in autoimmune diseases or produced by normal B lymphocytes from healthy donors. However in the case of the over-represented VH3/1 cluster, the most homologous VH was originating from a tumor related GD2/GD3 specific antibody, suggesting that anti-ganglioside binding may be one of the major specificities of antibodies produced by TIL B cells in medullary breast carcinoma. The generation of a single chain Fc (scFv) that comprises this VH sequence led to a recombinant antibody fragment with a binding capacity against tumor cells expressing GD3 (Kotlan et al., submitted). Overall, our results show that the immunoglobulin repertoire analysis of tumor infiltrating B cells may have a predictive value for determining a potential binding capacity.

Acknowledgements

This work was supported partly by the Institut Curie, the Association pour la Recherche sur le Cancer (ARC, ARECA program, contract C01-010) and the INSERM. Dr. Kotlan was supported by scholarships and grants from NATO (CLG 978639), Eotvos (XLI/34/2000), OTKA (T030380), the Association pour la Recherche sur le Cancer (ARC), the Société Française du Cancer (SFC), the French embassy in Hungary aswell as the French Ministry of Foreign Affairs, and the Rajko Medenica Research Foundation.

The authors wish to thank the expert technical assistance of Mrs N. Gruel and Mrs Gabriella Furedi Sonkoly. We are also thankful to Dr Andras Varadi (DSci) for helpful discussions.

References

[1] G.M. Halliday, A. Patel and M.J. Hunt et al., Spontaneous regression of human melanoma/nonmelanoma skin cancer: association with infiltrating CD4+ T cells, *World Journal of Surgery* **19** (1995), 352–358.

[2] G. Giegerich, M. Pette and E. Meinl et al., Diversity of T cell receptor α and β chain genes expressed by human T-cells specific for similar myelin basic protein peptide/major histocompatibility complexes, *Eur J Immunol* **22** (1992), 753–759.

[3] A. Van Pel and P. Van Der Bruggen et al., Genes coding for tumor antigens recognized by cytolytic T lymphocytes, in G. Möller, ed., *Immunological Reviews* **145** (1995), 229–250.

[4] R.A. Willemsen, R. Debets, P. Chames and R.L.H. Bolhuis, Genetic engineering of T Cell specificity for immunotherapy of cancer, *Human Immunology* **64** (2003), 56–68.

[5] H. Zhang, D.F. Lake and J.A. Barbuto et al., A human monoclonal antimelanoma single-chain Fv antibody derived from tumor-infiltrating lymphocytes, *Cancer Research* **55** (1995), 3584–3591.

[6] S. Imahayashi, Y. Ichizoshi, I. Yoshino, R. Eifuku, M. Takenoyama and K. Yasumoto, Tumor-infiltrating B-cell-derived IgG recognizes tumor components in human lung cancer, *Cancer. Invest.* **18** (2000), 530–536.
[7] H. Wu, J.D. Pancook, G. Beurlein, T. Chilton, G. Pecht, W.D. Huse and J.D. Watkins, Cloning, isolation and characterization of human tumor in situ monoclonal antibodies, *Cancer. Immunol. Immunother.* **51** (2002), 79–90.
[8] A.M. Cuisinier, V. Guigou and Boubli et al., Preferential expression of VH5 and VH6 immunoglobulin genes in early human B-cell ontogeny, *Scand J Immunol* **30** (1989), 493–497.
[9] Y. Choi, M.H. Rickert, M. Ballow and S.J. Greenberg, Human IgH-V Gene Repertoire in neonatal cord blood, adult peripheral blood, and EBV-transformed cells, *New York Academy of Sciences* **764**, 261–264.
[10] B.D. Stollar, The expressed heavy chain V gene repertoire of circulating B cells in normal adults. In: Immunoglobulin gene expression in development and disease, eds, *Casali P and Silberstein LE Annals of the New York Academy of Sciences* **764** (1995), 264–274.
[11] A.K. Stewart, C. Huang, B.D. Stollar and R.S. Schwartz, High-frequency representation of a single VH gene in the expressed human B cell repertoire, *J. Exp. Med.* **177** (1993), 409–418.
[12] E.K. Shin, T. Akamizu, F. Matsuda, H. Sugawa, J. Fujikura, T. Mori and T. Honjo, Variable regions of Ig heavy chain genes encoding antithyrotropin receptor antibodies of patients with Graves' disease, *J Immunol* **152** (1994), 1485–1492.
[13] H. Schroeder Jr. and G. ighiero, The pathogenesis of chronic lymphocytic leukemia: analysis of the antibody repertoire, *Immunology Today* **15** (1994), 288–294.
[14] V. Guigou, A.M. Cuisinier and C. Tonnelle et al., Human immunoglobulin VH and VK repertoire revealed by in situ hybridization, *Molecular Immunology* **27** (1990), 935–940.
[15] G.P. Cook and I.M. Tomlinson, The human immunoglobulin VH repertoire, *Immunology Today* **16** (1995), 237–242.
[16] J.S. Andris, S.R. Abraham and V. Pascual et al., The human antibody repertoire: heavy and light chain variable region gene usage in six alloantibodies specific for human HLA class I and class II alloantigens, *Molecular Immunology* **32** (1995), 1105–1122.
[17] J.S. Andris, P.H. Ehrlich and L. Ostberg et al., Probing the human antibody repertoire to exogenous antigens: Characterization of the H and L chain V region gene segments from anti-Hepatitis B virus antibodies, *J Immunol* **149** (1992), 4053–4059.
[18] H. Schroeder Jr. and G. Dighiero, The pathogenesis of chronic lymphocytic leukemia: analysis of the antibody repertoire [see comments]. [Review], *Immunology Today* **15** (1994), 288–294.
[19] M. Hakoda, N. Kamatani and A. Taniguchi et al., Generation and molecular characterisation of monoclonal IgG4 rheumatoid factor from a patient with rheumatoid arthritis, *Ann Rheum Dis* **56** (1997), 74–77.
[20] B. Kotlan, N. Gruel, B. Zafrani, G. Furedi, J. Földi, G. Petranyi and W.H. Fridman, Teillaud J.L Immunoglobulin variable regions usage by B-lymphocytes infiltrating a human breast medullary carcinoma, *Immunology Letters* **65** (1999), 143–151.
[21] J.A. Coronella, P. Telleman, G.A. Kingsbury, T.D. Truong, S. Hays and R.P. Junghans, Evidence for an antigen-driven humoral response in medullary ductal breast cancer, *Cancer Research* **61** (2001), 7889–7899.
[22] M.H. Hansen, H. Nielsen and H.J. Ditzel, The tumor-infiltrating B cell response in medullary breast cancer is oligoclonal and directed against the autoantigen actin exposed on the surface of apoptotic cancer cells, *Proc. Natl. Acad. Sci. (USA)* **98** (2001), 12659–12664.
[23] P.M. O' Brien, E. Tsirimonaki, D.W. Coomber, D.W. Millan, J.A. Davis and M.S. Campo, Immunoglobulin genes expressed by B-lymphocytes infiltrating cervical carcinomas show evidence of antigen-driven selection, *Cancer Immunol. Immunother* **5** (2001), 523–532.
[24] O.S. Moore and F.W. Foote, The relatively favourable prognosis of medullary carcinoma of the breast, *Cancer* **2** (1949), 635–640.
[25] R.L. Ridolfi, P.P. Rosen and A. Port et al., Medullary carcinoma of the breast: a clinicopathologic study with 10 year follow-up, *Cancer* **40** (1977), 1365–1385.
[26] J.D. Marks, M. Tristem and A. Karpas et al., Oligonucleotide primers for polymerase chain reaction amplification of human immunoglobulin variable genes and design of family-specific oligonucleotide probes, *European Journal of Immunology* **21** (1991), 985–991.
[27] J.D. Marks, H.R. Hoogenboom and T.P. Bonnert et al., By-passing immunization. Human antibodies from V-gene libraries displayed on phage, *Journal of Molecular Biology* **222** (1991), 581–597.
[28] M.P. Lefranc, IMGT the international ImMunoGeneTics database, *Nucleic Acids Research* **29** (2001), 207–209.
[29] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucl. Acids. Symp. Ser.* **41** (1999), 95–98.
[30] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin and D.G. Higgins, The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Research* **24** (1997), 4876–4882.
[31] R.D.M. Page, TREEVIEW: An application to display phylogenetic trees on personal computers, *Computer Applications in the Biosciences* **12** (1996), 357–358.
[32] M.H. Hansen, H. Nielsen and H.J. Ditzel, Translocation of an antigen to the surface of Medullary Breast Cancer Cells Early in Apoptosis Allows for an antigen-driven antibody response elicited by tumor-infiltrating B cells, *J. Immunol.* **169** (2002), 2701–2711.
[33] B. Kotlan, P. Simsa, J.L. Teillaud, J. Toth, M. Mc Knight and M. Glassy, Immunoglobulin repertoire analysis reveals a potential tumor-antigen binding capacity of B lymphocytes infiltrating human breast carcinomas, manuscript under submission, 2003.