TEMPORAL DYNAMICS OF CLONAL EVOLUTION IN CHRONIC LYMPHOCYTIC LEUKEMIA WITH
STEREOTYPED IGHV4-34/IGKV2-30 ANTIGEN RECEPTORS: LONGITUDINAL IMMUNOGENETIC
EVIDENCE

Running title: Over time intraclonal dynamics of CLL subset #4

Lesley-Ann Sutton¹, Efterpi Kostareli², Evangelia Stalika²,³, Athanasios Tsaftaris³, Achilles
Anagnostopoulos², Nikos Darzentas⁴,³, Richard Rosenquist¹, Kostas Stamatopoulos²,³

¹Department of Immunology, Genetics and Pathology, Rudbeck Laboratory, Uppsala University,
Uppsala, Sweden; ²Hematology Department and HCT Unit, Greece; ³Institute of Applied Biosciences,
CERTH, Thessaloniki, Greece; ⁴Molecular Medicine Program, Central European Institute of
Technology, Masaryk University, Brno, Czech Republic

Corresponding author:
Lesley Ann Sutton, PhD
Department of Immunology, Genetics and Pathology
Rudbeck Laboratory, Uppsala University
SE-751-85 Uppsala, Sweden
E-mail: lesley.sutton@igp.uu.se

Keywords: intraclonal diversification; antigen; over time; stereotyped receptor; somatic
hypermutation
Chronic lymphocytic leukemia (CLL) patients assigned to stereotyped subset #4 possess distinctive patterns of intraclonal diversification (ID) within their immunoglobulin (IG) genes. Though highly indicative of an ongoing response to antigen(s), the critical question concerning the precise timing of antigen involvement is unresolved. Hence, we conducted a large-scale longitudinal study of 8 subset #4 cases totaling 511 and 398 subcloned IG heavy and kappa sequences. Importantly, we could establish a hierarchical pattern of subclonal evolution, thus revealing which somatic hypermutations were negatively or positively selected. In addition, distinct clusters of subcloned sequences with cluster-specific mutational profiles were observed initially, however at later time-points the minor cluster had often disappeared and hence been selected against. Despite the high intensity of ID, it was remarkable that certain residues remained essentially unaltered. These novel findings strongly support a role for persistent antigen stimulation in the clonal evolution of CLL subset #4.
INTRODUCTION

From an immunogenetic perspective, the critical role of the B-cell receptor (BcR) in chronic lymphocytic leukemia (CLL) is underscored by the biased immunoglobulin heavy variable (IGHV) gene repertoire and the allocation of patients into subgroups with markedly different prognosis based on the IGHV gene mutational status of the clonotypic BcRs (1-8). Even more compelling however, is the fact that almost 30% of CLL patients share BcRs with restricted, quasi-identical ‘stereotyped’ IG sequences with highly homologous IG variable heavy chain complementarity determining region 3 (VH CDR3), the key determinant of antigen specificity (9-19). Increasing evidence also indicates that cases expressing such stereotyped BcRs, and consequently assigned to distinct subsets, share biological and clinical features (10, 16, 17, 19-22). These findings, along with other structurally unique features of CLL BcRs, strongly suggest that antigens, superantigens or both may play an active role in the disease.

Elucidation of the antigenic specificity of the clonogenic BcRs in CLL has previously been hindered by technical difficulties; however, more recent procedures utilizing cell lines derived from the neoplastic CLL clone established that these cells can produce BcRs/monoclonal antibodies (mAbs) that bind autoantigens and molecular structures present on apoptotic cells and bacteria such as IgG, vimentin, filamin B, cardiolipin and DNA (23-25). In addition, by using recombinant DNA technologies, the role of antigenic reactivity and the impact of somatic hypermutation (SHM) on CLL mAb specificity have been investigated and revealed that CLL cells likely derive from B cells producing polyreactive, natural antibodies encoded by germline IG genes which either retain or lose polyreactivity due to SHM (26-28). Although the accumulated data on antigen reactivity does not provide the definitive agent driving CLL, they do provide a framework from which comparisons between other entities, most notably autoimmune diseases, can be drawn.

Whilst the results from studies aimed at defining the antigenic reactivity profile of CLL cells convincingly asserts their position in an antigen-driven model of CLL pathogenesis, the timing and
duration of antigenic exposures remain largely unknown. We recently investigated intraclonal diversification (ID) within the IG genes of patients with CLL and found that most cases showed no or low levels of ID. In sharp contrast, we reported intense ID within both the heavy and light chain IG genes of patients assigned to subset #4 (29, 30).

CLL patients assigned to stereotyped subset #4 are characterized clinically by an early age at diagnosis and an indolent disease course and molecularly by BcR IGs that exhibit a series of distinctive immunogenetic features (16). More specifically, they are IgG-switched and composed of heavy chains encoded by the IGHV4-34 gene and light chains encoded by the IGKV2-30 gene. Their VH CDR3 is long and enriched in positively charged residues, being particularly defined by a (K/R)RYY motif at IMGT positions 112.4-112.1 (17, 19). Furthermore, the VH and VK domains demonstrate a high impact of SHM, and are remarkable for carrying shared (‘stereotyped’) amino acid changes induced by SHM: noteworthy amongst these are changes leading to the introduction of negatively charged residues in both the heavy and light chains (17, 31).

The intense ID in subset #4 BcR IGs convincingly implicates antigen selection in the development and evolution of subset #4. However, our previous studies were limited to depicting what was occurring at a single time-point and could not provide insight into the temporal dynamics of the CLL clones. Hence, we gathered a novel and unique data set from serial sampling of 8 subset #4 cases and through this approach we could trace clonal evolution over time and investigate the impact of functionally relevant mutations for subclone selection. Overall, the results reported herein provide conclusive evidence that subset #4 patients continue to acquire SHMs within their IG genes over time and that this observation is best explained by a clear role for antigen selection in the clonal evolution of CLL subset #4.
MATERIALS AND METHODS

Patient group

Peripheral blood samples were collected at multiple time-points from 8 CLL patients who, based on IG gene sequence features and following previously established criteria, were assigned to subset #4 (IGHV4-34/IGKV2-30, 20 amino acid VH CDR3 and IgG-switched) (16, 17, 19). All patients displayed the typical CLL immunophenotype and met the recently revised diagnostic criteria of the National Cancer Institute Working Group (32). Patients’ demographics and clinical and molecular data are summarized in Supplemental Tables 1-2. Cases were analyzed over a 6 year period (range 7 – 72 months, median 20 months) and no patient received treatment during sampling (Supplemental Tables 1 & 3). The diagnostic sample was available, and referred to as time-point 1, for 6 of the 8 patients analyzed. No diagnostic samples were available for the remaining two patients (P0103 and P2451) and therefore the initial sample (time-point 1) analyzed for these patients were 81 and 63 months post diagnosis respectively. Written informed consent was obtained in accordance with the Declaration of Helsinki and the study was approved by the local ethics review committee.

PCR amplification of IG rearrangements

PCR amplification was performed on either genomic DNA (gDNA) or complementary DNA (cDNA) using the IGHV4 leader primer and the antisense IGHC primer for analysis of IGHV-IGHD-IGHJ rearrangements and the VK2 FR1 primer together with the antisense IGKC primer for analysis of the IGKV-IGKJ rearrangements. All amplification reactions were run using the high-fidelity Accuprime Pfx polymerase (Invitrogen) and purified PCR amplicons were subjected to direct sequencing on both strands (16, 17, 29-31).
Subcloning

PCR amplification products were gel-purified with the Qiagen DNA purification columns, ligated into the pCR2.1 vector (Invitrogen), and subsequently transformed into *Escherichia coli* TOP10F’ competent bacteria (Invitrogen). A range of 8 to 39 colonies per case (median, 22 colonies) were chosen randomly and sequenced using the 20 universal primer or M13 primers.

Sequence data analysis and definitions

Sequences obtained from subcloning were analyzed using the Antigen Receptors Research Tool/ARResT at http://bat.infspire.org/arrest, IMGT® databases and the IMGT/V-QUEST tool (http://www.imgt.org) (33, 34). Codons and amino acid positions are according to the IMGT unique numbering system for the V domain. For the IGHV-IGHD-IGHJ rearrangements the entire V region was evaluated (from VH FR1 codon 1 down to the VH CDR3). To avoid misidentification of mutations when the VK2 FR1 consensus primer was used, nucleotide changes in the obtained sequences were evaluated from codon 12 in VK FR1 down to VK CDR3.

Intraclonal diversification analysis

Intraclonal diversification in sets of subcloned sequences obtained from the same sample was assessed by examination of sequence variation in the V domain following definitions previously proposed by our group (29, 30). In brief, all ‘non-ubiquitous’ sequence changes from the germline were evaluated in the counts and characterized as follows: (1) unconfirmed mutation (UCM) - a mutation observed in only one subcloned sequence from the same specimen (unique); (2) confirmed mutation (CM) - a mutation observed more than once among subcloned sequences from the same specimen (partially shared). Amino acid changes resulting from UCMs or CMs are designated by the abbreviations UAA or CAA, respectively.
RESULTS AND DISCUSSION

Incidence, topology and molecular characteristics of intraclonal diversification over time in subset #4 immunoglobulin heavy and light chains

Eight CLL subset #4 cases were investigated for ID by analyzing 511 and 398 subcloned IG heavy and kappa sequences obtained from serial sampling (Supplemental Tables 1-3). Overall, all cases carried intraclonally diversified IG genes while the level of ID observed over time could be described as increasing, decreasing or complex i.e. a mutation appears or disappears and then re-emerges at a subsequent time-point (Figures 1A & B; Supplemental Figures 1-2).

At the nucleotide level, 87 unique confirmed mutations (CMs) were found throughout the IGHV4-34 gene in 374 clones while 137 unconfirmed mutations (UCMs) were also detected. Of the 137 UCMs, only 73 could be considered as ‘truly’ unconfirmed since 42 UCMs were ubiquitous mutations and/or CMs in either another sample or at a different time-point for the same sample. Similarly, the remaining 22 UCMs were ‘confirmed’ by the presence of the same mutation occurring in a single sequence from either a different case or an alternative time-point for the same sample. Analogous to the IGHV4-34 heavy chains of subset #4, their partner IGKV2-30 light chains were also affected by an active ID process, accumulating 53 CMs and 55 UCMs.

All observed CMs and UCMs concerned single-base changes resulting in both silent and replacement mutations while nucleotide substitution analysis revealed that transitions predominated over transversions, in keeping with a canonical SHM process. Detailed information on the ‘type’ (replacement versus silent) and distribution of CMs/UCMs throughout the V region is provided in Supplemental Tables 4-5. At the amino acid (AA) level, 48 confirmed AA changes (CAA) were identified among sets of subcloned IGHV4-34 rearrangements resulting in either a conservative (n=19) or non-conservative (n=29) AA substitution. Such AA changes were also evidenced throughout the IGKV2-30 rearrangements analyzed, and of the 30 CAAs detected, 40% (n=12) resulted in conservative changes (Supplemental Tables 6-8; Supplemental Figure 3).
Detailed analysis of the mutations revealed precise targeting of ID, inferring strong selective pressure for preservation of critical physicochemical properties. In particular, we noted: i) restricted ID patterns, in the form of identical mutations at certain VH/VK positions amongst subclones of different cases e.g. despite glycine at codon 28 (VH CDR1) being mutated in 410/511 (80%) sequences, the only permissible option for substitution was an acidic residue; ii) over time ID hotspots i.e. mutations at certain codons were repeatedly observed during clonal evolution e.g. codon 40 (VH FR2) and codon 66 (VK FR3); and iii) patterns of coincident mutations: as an example, the substitution of isoleucine (I) with valine (V) at codon 21 (VK CDR1) together with the asparagine (N) to aspartic acid (D) change at codon 66 (VK FR3) constitutes a pattern of SHM that could perhaps be interpreted within the context of ‘concerted mutagenesis’ (Figures 1A & B; Supplemental Tables 7-8).

Although, the extent of ID varied between cases with, for example, P3020 and P3916 undergoing limited diversification in comparison to several other cases analyzed, most notably P1422 and P2920, overall the incidence and topology of ID in subset #4 over time are suggestive of selection events governed by structural constraints for optimal antigen recognition. This idea is further underscored by the observation that many UCMs were present as ubiquitous mutations, CMs or UCMs in other samples or at another time-point of the same sample, and thus may be considered as imprints of a highly targeted and selected ID process.

**Patterns of over time clonal branching in CLL subset #4**

The analysis of consecutive samples enabled us to trace diversification within the subset #4 CLL antibody over time, thus revealing which mutations were negatively or positively selected (Figure 2A; Supplemental Figures 1, 2 & 4). Consequently, a ‘stepwise’ accumulation of mutations could be observed with several mutations present in only a fraction of subclones at an early time-point becoming ubiquitous mutations i.e. present in all subcloned sequences of the same case at subsequent time-points e.g. P1939 & P2451 – codon 25 (VK FR1), codon 66 and 86 (VK FR3). The
presence of distinct clusters of subcloned sequences with ‘cluster-specific’ mutational profiles shed light on both their evolutionary relationships, with, for instance, minor clusters appearing or, more often, disappearing over time e.g. in cases P0103 and, especially, P1422 (Figures 2A & 2B; Supplemental Figures 1, 2 & 4). For instance, although a major and a minor clone were observed initially within P1422, a dramatic shift in subclonal composition occurred at later time-points and the minor clone became undetectable. The majority of subcloned sequences at later time-points retained the mutational pattern of the predominant clone while individual clones acquired additional mutations (Supplemental Figures 1-2). A further interesting finding was provided by case P2920 (Figure 2A), which exhibited a pattern of ‘clonal equilibrium’ i.e. two major subclones were present at diagnosis and the relative size of these subclones was maintained over the time-points analyzed. Admittedly, the second time-point available for this case was only 7 months post-diagnosis, however it is plausible that in time one of these clones may predominate due to positive selection.

**Ontogenetic implications**

Examination of the IG sequences of CLL subset #4 antibodies highlighted their resemblance to pathogenic anti-DNA antibodies since both carry IgG-switched BcRs with long positively charged VH CDR3s enriched with arginine residues (35-37). It is therefore conceivable that auto-reactive precursors to the subset #4 CLL clones could arise early in B cell development and subsequently undergo positive selection by DNA, nucleosomes or surface structures of apoptotic cells (38, 39). Thereafter, in order to avoid intense self-reactivity to apoptotic stimulatory material, modifications introduced by SHM could render these clones in a state of diminished responsiveness, perhaps through the introduction of acidic residues in both the heavy and light chain genes (Figure 1; Supplemental Tables 6-8) (40-42). Reinforcing the above theory, studies looking at antigen reactivity profiles in which mutated IGHV4-34 antibodies are reverted to their germline configuration demonstrated increased HEp-2 reactivity and/or acquired polyreactivity (26).
In the periphery these CLL precursors may exist in a quasi-anergic state demonstrating attenuating responses to selecting antigenic elements whilst retaining the ability to engage in superantigen-like interactions (43, 44). Along these lines, it is remarkable that despite the high level of ID, certain critical positions of the VH domain were essentially left unaltered by the ID process. For instance, the W residue at codon 7 in the VH FR1 of IGHV4-34, which plays a critical role in recognition of the N-acetyllactosamine antigenic determinant present on both self-antigens and microbial pathogens (Mycoplasma pneumoniae, CMV, EBV) (45, 46), carried alterations in only 6/511 sequences. It is plausible that selective pressure resulted in maintenance of this auto-antigenic binding motif and consequently the IGHV4-34 clones of subset #4 retained the ability to engage with (auto/super)antigen; as also implied by our previous findings linking subset #4 with persistent, subclinical infection by CMV and EBV (47).

In attempting to reconcile the evidence for the continuous stimulation and selection of fully transformed CLL cells it is worth mentioning another fascinating CLL subset, subset #8, since emerging evidence suggests that the distinct clinical profiles of certain subsets may be linked to their precise functional and genetic make-up (10, 16, 17, 19). Whilst both subset #4 and subset #8 express IgG, differences exist between these subsets with subset #8 (IGHV4-39/IGKV1-39) patients harboring unmutated IG genes, following an aggressive disease course and also having a significantly increased risk for Richter’s transformation (48). In addition, subset #8 cases exhibit a high frequency of trisomy 12, t(14;19)(q32;13) and NOTCH1 aberrations while subset #4 is conspicuous for a remarkably stable genomic profile acquiring few if any aberrations, namely the more favorable prognostic marker del(13q) (49-53). These subsets are further distinguishable due to their distinct immune signaling profiles; subset #4 clones have a particular functional response to stimulation through toll-like receptors (TLRs), exhibiting a TLR7-tolerant profile, whereas subset #8 clones display an unrestricted and intense response to multiple TLR ligands (54, 55).

If one considers the broad reactivity profile of subset #8, it is conceivable that the particularly poly/autoreactive clones do not engage in SHM because their unmutated clonotypic IG
rearrangements confer these cells with ample opportunities to engage in interactions with numerous antigens. In this regard, data generated by Colombo et al, indicating a process of antigenic stimulation/selection leading to expansion of the subset #8 IgG-bearing subclone (56), perhaps accounts for the progressive nature of cases within this subset. However, subset #4, the focus of this study, is extremely enigmatic with the lack of clinical aggressiveness being juxtaposed with extensive intraclonal diversification. Thus, this lack of clinical progression observed in most subset #4 cases despite the presence of clonal evolution may be linked to the proposed anergic state of the malignant cells.

Conclusions

Studying the intricacies of the IG sequence enabled us to trace diversification within the subset #4 CLL antibody. From this a conceptually satisfying model of clonal evolution within subset #4 could be proposed, whereby, periodically, upon contact with microbial or viral pathogens or during bouts of excessive apoptosis induced by various triggers, low affinity autoreactive B cells may be re-activated by engagement of the BcRs and/or other immune receptors (47, 55, 57, 58). Hence clonal evolution may (at least in part) result from the functional interplay between CLL cells, the microenvironment and extrinsic pathogens.

Taken together, this extended immunogenetic analysis advances our understanding of the evolution of CLL subset #4 clones. Admittedly, the advent of high throughput sequencing promises even more in this direction, however, limitations of the existing protocols and platforms currently preclude its usage for reliable assessment of SHM patterns. In addition, biological questions remain, concerning: (i) the precise anatomical location where ongoing SHM occurs and whether T-cell help is required; and (ii) what the relevance of the observed molecular changes (if any) is to the clinical course and outcome? Nevertheless, the intense ID within CLL subset #4 supports the argument that self/neoa- antigens acting in synergy with extrinsic pathogens may drive CLL progenitors or even the malignant cells themselves by continuously triggering BcRs with distinctive structural features.
ACKNOWLEDGMENTS

This work was supported in part by the ENosAI project (code 09SYN-13-880) co-funded by the EU and the Hellenic General Secretariat for Research and Technology to ND and KS; Cariplo Foundation (Milan, Italy) to KS; the Swedish Cancer Society, the Swedish Research Council, and the Lion’s Cancer Research Foundation, Uppsala. ND is currently supported by CEITEC MU (CZ.1.05/1.1.00/02.0068) and SuPReMMe (CZ.1.07/2.3.00/20.0045);

AUTHORSHIP

Lesley-Ann Sutton performed research, analyzed data and wrote the paper. Efterpi Kostareli performed research and analyzed data. Evangelia Stalika performed research. Athanasios Tsafaritis and Achilles Anagnostopoulos supervised research. Nikos Darzentas supervised the immunoinformatics analysis. Richard Rosenquist designed the study and supervised research. Kostas Stamatopoulos designed the study, supervised research and wrote the paper.

DISCLOSURE

The authors declare that they have no competing interests as defined by Molecular Medicine, or other interests that might be perceived to influence the results and discussion reported in this paper.
REFERENCES

33. Lefranc MP, et al. (2009) IMGT, the international ImMunoGeneTics information system. Nucleic Acids Res 37: D1006-1012.


49. Martin-Subero Ji, et al. (2007) A comprehensive genetic and histopathologic analysis identifies two subgroups of B-cell malignancies carrying a t(14;19)(q32;q13) or variant BCL3-translocation. *Leukemia* **21**: 1532-1544.


FIGURE LEGENDS

Figure 1: Restricted patterns and hotspots of intraclonal diversification over time. Figure 1A highlights certain mutations observed within the *IGHV4-34* gene of subset #4 cases whilst Figure 1B illustrates mutations within the *IGKV2-30* gene. Precise targeting of intraclonal diversification (ID) infers strong functional constraints for preservation of critical physicochemical properties. This idea is further underscored by the observation that many unconfirmed mutations (UCMs i.e. mutations present in a single subcloned sequence; single-case and time-point level) were present as ubiquitous mutations, confirmed mutations (CMs i.e. mutations present in more than one but less than all subcloned sequences) or UCMs in other samples or at another time-point of the same sample, and thus may be considered as imprints of the ID process. The substitution of isoleucine (I) with valine (V) at codon 21 (VK CDR1) together with the asparagine (N) to aspartic acid (D) change at codon 66 (VK FR3) provides an example of patterns of SHM that could perhaps be interpreted within the context of ‘concerted mutagenesis’.

Amino acids are represented by a single-letter code. The germline amino acid is indicated in the header with the number referring to the codon position according to the IMGT unique numbering for V-DOMAIN (V-QUEST Version 2.1.2). Within the figures, the number of subcloned sequences containing a certain amino acid is indicated after the single-letter code while amino acids highlighted in white exemplify the phenomenon of UCMs being ‘confirmed by either a different case or another time-point of the same case’. A comprehensive list of amino acid changes observed in all cases analyzed is provided in Supplemental Tables 6-8. *N*: number of subcloned sequences analyzed; *TP*: time-point

Figure 2: Temporal intraclonal dynamics of CLL subset #4. The level of ID observed over time could be described as increasing, decreasing or complex, i.e. a mutation appears or disappears and then re-emerges at a subsequent time-point. (A) Graphs illustrate the patterns of clonal branching over time
in both the *IGHV4-34* and *IGKV2-30* genes of subset #4 CLL cases. The relative size of each subclone is plotted against all time-points analyzed and within each case a cluster of identical subcloned sequences or a unique subcloned sequence is represented by a different colored segment. (B) Schematic representation of the dynamic patterns of intraclonal diversification observed within the *IGHV4-34* gene of case P1422. Although a major and minor clone were observed initially within P1422, a dramatic shift in subclonal composition occurs at later time-points and the minor clone becomes undetectable. The majority of subcloned sequences at later time-points retain the mutational pattern of the predominant clone while individual clones that acquire additional mutations are indicated by the black circles. Additional graphs and sequence information are available in Supplemental Figures 1-4 and Supplemental Tables 6-8. *TP: time-point*
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P0103</td>
<td>1</td>
<td>33</td>
<td>W30</td>
<td>R3</td>
<td>G2</td>
<td>E31</td>
<td>S2</td>
<td>T33</td>
<td>T33</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>28</td>
<td>W27</td>
<td>R3</td>
<td>G8</td>
<td>E20</td>
<td>T20</td>
<td>T20</td>
<td>T28</td>
</tr>
<tr>
<td>P0907</td>
<td>1</td>
<td>33</td>
<td>W33</td>
<td>W32</td>
<td>D33</td>
<td>D22</td>
<td>F21</td>
<td>F22</td>
<td>T33</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>22</td>
<td>W22</td>
<td>D23</td>
<td>D22</td>
<td>Y1</td>
<td>T21</td>
<td>T23</td>
<td>T22</td>
</tr>
<tr>
<td>P1422</td>
<td>1</td>
<td>27</td>
<td>W27</td>
<td>D1</td>
<td>E21</td>
<td>E32</td>
<td>F32</td>
<td>Y1</td>
<td>T27</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>33</td>
<td>W33</td>
<td>D1</td>
<td>E21</td>
<td>E32</td>
<td>F32</td>
<td>Y1</td>
<td>T27</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>25</td>
<td>W24</td>
<td>R1</td>
<td>E21</td>
<td>E25</td>
<td>F27</td>
<td>F25</td>
<td>S20</td>
</tr>
<tr>
<td>P1939</td>
<td>1</td>
<td>35</td>
<td>W35</td>
<td>G1</td>
<td>E34</td>
<td>E1</td>
<td>T35</td>
<td>A35</td>
<td>A1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>35</td>
<td>W35</td>
<td>G1</td>
<td>E1</td>
<td>E34</td>
<td>T35</td>
<td>A35</td>
<td>A1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>21</td>
<td>W21</td>
<td>G1</td>
<td>E34</td>
<td>E1</td>
<td>T35</td>
<td>A35</td>
<td>A1</td>
</tr>
<tr>
<td>P2451</td>
<td>1</td>
<td>35</td>
<td>C1</td>
<td>G1</td>
<td>D33</td>
<td>E1</td>
<td>T33</td>
<td>S33</td>
<td>A3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>33</td>
<td>W25</td>
<td>G1</td>
<td>D33</td>
<td>E33</td>
<td>T33</td>
<td>S33</td>
<td>A3</td>
</tr>
<tr>
<td>P2920</td>
<td>1</td>
<td>26</td>
<td>W25</td>
<td>C1</td>
<td>G26</td>
<td>G26</td>
<td>F26</td>
<td>F26</td>
<td>P15</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>26</td>
<td>W26</td>
<td>G26</td>
<td>G26</td>
<td>F26</td>
<td>F26</td>
<td>P15</td>
<td>S26</td>
</tr>
<tr>
<td>P3020</td>
<td>1</td>
<td>16</td>
<td>W16</td>
<td>G16</td>
<td>G21</td>
<td>F16</td>
<td>S16</td>
<td>S16</td>
<td>T16</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>21</td>
<td>W21</td>
<td>G16</td>
<td>G21</td>
<td>F16</td>
<td>S16</td>
<td>S16</td>
<td>T16</td>
</tr>
<tr>
<td>P3916</td>
<td>1</td>
<td>17</td>
<td>W17</td>
<td>D17</td>
<td>D22</td>
<td>F17</td>
<td>S17</td>
<td>S17</td>
<td>A17</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>22</td>
<td>W22</td>
<td>D17</td>
<td>D22</td>
<td>F22</td>
<td>S22</td>
<td>S22</td>
<td>A17</td>
</tr>
</tbody>
</table>

**Legend:**
- **Acidic (E, D)**
- **Basic (K, R, H)**
- **Aromatic (F, W, Y)**
- **Non-polar, aliphatic (G, A, I, V, L, P, M)**
- **Polar, non-charged (S, T, C, N, Q)**
### Table

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P0103</td>
<td>1</td>
<td>38</td>
<td>I_{38}</td>
<td>S_{33}</td>
<td>S_{31}</td>
<td>G_{6}</td>
<td>N_{38}</td>
<td>D_{32}</td>
<td>Y_{6}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>33</td>
<td>I_{33}</td>
<td>S_{32}</td>
<td>S_{31}</td>
<td>N_{8}</td>
<td>D_{31}</td>
<td>Y_{2}</td>
<td>T_{33}</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8</td>
<td>I_{8}</td>
<td>S_{8}</td>
<td>S_{8}</td>
<td>N_{8}</td>
<td>D_{8}</td>
<td>T_{8}</td>
<td>V_{8}</td>
</tr>
<tr>
<td>P0907</td>
<td>1</td>
<td>16</td>
<td>V_{16}</td>
<td>S_{16}</td>
<td>S_{15}</td>
<td>D_{15}</td>
<td>D_{16}</td>
<td>D_{17}</td>
<td>T_{15}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>17</td>
<td>V_{17}</td>
<td>S_{16}</td>
<td>S_{17}</td>
<td>D_{17}</td>
<td>D_{16}</td>
<td>T_{17}</td>
<td>V_{16}</td>
</tr>
<tr>
<td>P1422</td>
<td>1</td>
<td>14</td>
<td>V_{14}</td>
<td>S_{14}</td>
<td>N_{14}</td>
<td>A_{14}</td>
<td>T_{14}</td>
<td>V_{14}</td>
<td>V_{14}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>26</td>
<td>V_{26}</td>
<td>S_{26}</td>
<td>N_{26}</td>
<td>A_{26}</td>
<td>T_{26}</td>
<td>V_{26}</td>
<td>V_{26}</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>11</td>
<td>V_{11}</td>
<td>S_{11}</td>
<td>N_{11}</td>
<td>A_{11}</td>
<td>T_{11}</td>
<td>V_{11}</td>
<td>V_{11}</td>
</tr>
<tr>
<td>P1939</td>
<td>1</td>
<td>19</td>
<td>I_{17}</td>
<td>V_{2}</td>
<td>S_{2}</td>
<td>C_{17}</td>
<td>N_{17}</td>
<td>D_{17}</td>
<td>G_{2}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>19</td>
<td>I_{19}</td>
<td>S_{1}</td>
<td>C_{18}</td>
<td>N_{19}</td>
<td>D_{19}</td>
<td>T_{19}</td>
<td>V_{19}</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>13</td>
<td>I_{13}</td>
<td>S_{1}</td>
<td>C_{13}</td>
<td>N_{13}</td>
<td>D_{13}</td>
<td>T_{13}</td>
<td>V_{13}</td>
</tr>
<tr>
<td>P2451</td>
<td>1</td>
<td>39</td>
<td>I_{36}</td>
<td>V_{3}</td>
<td>S_{39}</td>
<td>S_{24}</td>
<td>N_{37}</td>
<td>D_{37}</td>
<td>A_{2}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>24</td>
<td>I_{24}</td>
<td>S_{24}</td>
<td>S_{24}</td>
<td>N_{24}</td>
<td>D_{24}</td>
<td>T_{24}</td>
<td>V_{24}</td>
</tr>
<tr>
<td>P2920</td>
<td>1</td>
<td>18</td>
<td>V_{13}</td>
<td>S_{10}</td>
<td>S_{10}</td>
<td>D_{10}</td>
<td>D_{18}</td>
<td>T_{10}</td>
<td>V_{6}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>11</td>
<td>V_{11}</td>
<td>S_{11}</td>
<td>S_{11}</td>
<td>D_{11}</td>
<td>T_{11}</td>
<td>V_{1}</td>
<td>I_{11}</td>
</tr>
<tr>
<td>P3020</td>
<td>1</td>
<td>21</td>
<td>V_{2}</td>
<td>S_{21}</td>
<td>S_{21}</td>
<td>D_{21}</td>
<td>T_{21}</td>
<td>T_{21}</td>
<td>T_{21}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>21</td>
<td>V_{21}</td>
<td>S_{21}</td>
<td>S_{21}</td>
<td>D_{21}</td>
<td>T_{21}</td>
<td>T_{21}</td>
<td>T_{21}</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>22</td>
<td>V_{22}</td>
<td>S_{22}</td>
<td>S_{22}</td>
<td>D_{22}</td>
<td>T_{22}</td>
<td>T_{22}</td>
<td>T_{22}</td>
</tr>
<tr>
<td>P3916</td>
<td>1</td>
<td>15</td>
<td>V_{15}</td>
<td>S_{15}</td>
<td>T_{15}</td>
<td>N_{1}</td>
<td>T_{15}</td>
<td>D_{15}</td>
<td>V_{15}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>13</td>
<td>V_{13}</td>
<td>S_{13}</td>
<td>T_{13}</td>
<td>N_{1}</td>
<td>T_{13}</td>
<td>D_{13}</td>
<td>V_{13}</td>
</tr>
</tbody>
</table>