

# A comparison of human and macaque (*Macaca mulatta*) immunoglobulin germline V regions and its implications for antibody engineering

Philippe Thullier,\* Siham Chahboun and Thibaut Pelat

Groupe de Biotechnologie des Anticorps; Département de Biologie des Agents Transmissibles; Centre de Recherche du Service de Santé des Armées; La Tronche, France

**Key words:** antibody, therapeutic, non-human primate, human, V region, identity, humanization, patent

Seventy-five V regions encoded by the sequenced genome of one *Macaca mulatta* specimen have been identified by homology and paired with similar human counterparts. When the human V region of each pair presented no allelic polymorphism, it was directly compared with its homolog. This was the case for 37 pairs and percents of identity ranged between 84–97%. When the human V region presented allelic polymorphism, this polymorphism was found to be significantly smaller ( $p < 0.0001$ ,  $p < 0.0001$ ,  $p = 0.03$  for IGHV, IGLV, IGKV regions respectively), 4.2-fold on average, than the differences observed between human and macaque V regions. Similar results were obtained when analyzing framework regions (FRs) only. These results, in agreement with others, demonstrate the existence of differences between human and macaque V regions, confirm the need for the humanization of macaque V regions intended for therapeutic use and call into question the validity of patents relying on the “undistinguishable” character of human and macaque V regions or FRs.

## Introduction

The various nano- to picomolar antibody fragments that we have previously isolated<sup>1-5</sup> exemplify how immune phage-displayed libraries from non-human primates (*Macaca fascicularis*) may be efficiently utilized to isolate neutralizing antibodies, which then may be humanized for therapeutic use.<sup>6</sup> This general strategy has been previously detailed<sup>7</sup> and recommendations have been given on how best it may be implemented.<sup>8</sup> In these examples, a high percent of identity between the framework regions (FRs) of the macaque (*Macaca fascicularis*) IgG variable (V) domains and its closest human V region encoded by germline V, diversity (D) and joining J genes have been observed (from 84.2–92%). These high percents of identity are the core of our strategy and facilitate the humanization step. More broadly, all V regions of macaque IgGs available in databanks have been directly compared with all V regions of human IgGs available in the Kabat databank, utilizing H- and G- scores.<sup>9,10</sup> Both scores were mathematically defined as Z-scores, also called standard score,  $Z = ((x - \mu)/\sigma)$ , and were utilized for the comparisons of each macaque V region with different subsets of human V regions. H-score comparisons were performed against whole  $\gamma$ ,  $\lambda$  and  $\kappa$  subsets. G-score comparisons were performed against V regions of the same family to avoid the over-estimation of human

similarity for families over-represented in the Kabat database, such as IGHV3.

H- and G-scores show that V regions from macaque light chains may sometimes be undistinguishable from their human counterparts, but this was not the case for V regions from macaque heavy chains, and so a humanization step is required. For this step, “germline humanization,”<sup>6</sup> which utilizes human germline sequences as templates instead of the more frequently utilized sequences of human expressed IgGs, was shown to be efficient by both scores.

The question of the humanness of V regions from macaque immunoglobulins is also at the core of European patent (EP) 1,266,965. This patent limits in Europe the therapeutic use of macaque antibodies if they are directed against human antigens or brings restrictions regardless of the nature of the antigen in the US (parallel US patent 5,693,780). In both patents, the suitability of macaque V regions for therapeutic use is presented as directly related to the absence of immunogenicity of these regions in humans. This absence of immunogenicity, in turn, is only indirectly demonstrated as a consequence of (1) the impossibility of distinguishing between human and macaque V regions, of which it is stated that “it is impossible therefore to distinguish between variable region immunoglobulin sequences originating from Old World monkeys and those originating from humans

\*Correspondence to: Philippe Thullier; Email: pthullier@yahoo.com

Submitted: 05/03/10; Accepted: 05/30/10

Previously published online: [www.landesbioscience.com/journals/mabs/article/12545](http://www.landesbioscience.com/journals/mabs/article/12545)

DOI: 10.4161/mabs.2.5.12545

based on sequence comparisons” in EP 1,266,965; and (2) the concept of “immunological self,” which states that proteins of a given species do not raise an immune response in that species. Following these two points of reasoning, the “undistinguishable” character of human and macaque V regions would warrant the tolerance of macaque V regions for therapeutic use so that no humanization of these regions would be useful, which contrasts our views. With the words “impossible to distinguish,” the patent more precisely means that “the level of homology between human and monkey (variable region) sequences for a given family is as high as between two human sequences within that “family” (excerpt from EP 1,266,965). The “level of homology” and the “family” concept are not part of the “immunological self” concept, nor part of the concept of immunological tolerance. The “immunological self” concept relates to the various proteins constitutive of one organism, as encoded by its DNA, which are individually recognized and well-tolerated by this organism, so that tolerance has to be considered for each individual protein.

When it comes to the particular case of the immunological tolerance of human recombinant antibodies utilized for therapy, the first source of differences between the recombinant immunoglobulin and the immunoglobulins encoded by the patient’s DNA is the variability observed between individuals or allelic polymorphism. Indeed, a human recombinant antibody may show allelic differences with the “immunological self” of a given patient treated by this antibody if this patient is not the donor of the genetic material encoding the antibody, as is very generally the case. Thus, human recombinant antibodies are not always completely part of every patient’s self and one may recognize that, if macaque V regions were no more different than human V regions are between themselves, there would be no additional risk when utilizing such macaque V regions as compared to utilizing their human counterparts for therapy. One might further argue that, in the case of immunoglobulins, the “immunological self” concept is not strictly applicable because IgGs are mutated during the affinity maturation process. These hypermutations are the second source of differences,<sup>11</sup> so that IgGs are not fully part of the “immunological self” of the person where this process has taken place. These somatic hypermutations and allelic V region polymorphism, as well as constant region polymorphism,<sup>12</sup> are the reasons why IgGs are not perfectly tolerated, as shown by the existence of idiotypic networks<sup>13</sup> and the imperfect tolerance of adalimumab (Humira®).<sup>14</sup>

The “immunological self” concept best applies to IgMs, expressed from rearranged but usually unmutated sequences. Consequently, the question of the suitability of macaque V regions (and therefore variable domains) for therapeutic use may be further expressed as the question of whether or not V regions of human and macaque IgMs, i.e., encoded by unmutated germline sequences, are as similar as V regions of IgMs from different human individuals. This question is also very directly relevant to the patents because they mention immunoglobulins with no more precision, and thus also refer to IgMs. Indeed, if differences were to be observed between human and macaque IgM V regions, they would constitute systematic differences still existing after affinity maturation has taken place on those IgM, i.e.,

between human and macaque IgG V regions. Taking advantage of the availability of *Macaca mulatta* germline sequences, we have addressed this still-unanswered question.

*M. mulatta* (Rhesus monkey) is closely related to *M. fascicularis* (cynomolgus), on which we have worked, as both species are Old World monkeys and both are referred under this name in the patents (Old World monkeys also encompass the more distant baboons). The goal of the study was to identify macaque (*M. mulatta*) V regions encoded by germline V genes, as encountered in IgM, and then compare them to human germline V regions. More precisely, the study was structured in three main steps: (i) identification of putative germline macaque V regions by comparison with human V genes utilizing BLASTP,<sup>15</sup> (ii) identification of the human V regions most similar to each of those macaque V regions and (iii) evaluation of the percent of identity between those pairs of most similar human and macaque V regions. The present study is thus very different from the previous study comparing human and macaque V regions,<sup>10</sup> which focused on V regions from expressed IgGs, thus rearranged and mutated sequences in contrast with germline sequences examined here.

Evaluating the percents of identity between most similar human and macaque V regions by utilizing sequences from one macaque specimen is a two-sided question. If the human V region, and the corresponding V gene, presents no allelic polymorphism, this evaluation is strictly the calculation of the percent of identity between the two V regions. If the human V region presents allelic polymorphism, then the question addressed here becomes the comparison of the percent of identity between most similar macaque and human V regions with the percents of identity existing between this human V region and its alleles. If the percents of identity existing between human alleles are higher than the percent of identity between the human and macaque V regions, the V region of another human is likely to be less foreign, and thus less immunogenic, to a human individual than the V region of a macaque.

## Results

**V regions of  $\gamma$  isotype (IGHV).** The 51 human germline IGHV regions, represented by their allele \*01,<sup>16-18</sup> were aligned on the macaque genome using the BLASTP search program. Twenty-six putative macaque IGHV regions were identified as homologous to these human IGHV regions and further analyzed. The human IGHV regions most similar to these 26 macaque IGHV regions were identified with IMGT/IMGT/DomainGapAlign.<sup>19</sup> Twenty-four different human IGHV regions were found (Table 1). The IGHV region encoded by human IGHV3-23\*04 was identified as the allele most similar to three different macaque IGHV regions (encoded by scaffolds 1099214148171:1052-1950, 1099214739418:108-1401, 1099548049732:81017-82310). Similarly, two human IGHV regions encoded by two alleles of the same gene were identified as most similar to two different macaque IGHV regions (the human IGHV regions encoded by alleles IGHV4-59\*01 and IGHV4-59\*04 are most similar to the macaque IGHV regions encoded by scaffolds 1099548049584:520128-521424 and 1099548049584:994-2281).

**Table 1.** Analysis of the differences between human and macaque IGHV regions

Human V <sub>γ</sub> gene	Macaque V gene ("scaffold") encoding the V region most similar to the region encoded by the human V gene (allele *01)	Human V gene allele encoding the V region most similar to the macaque V region	Nb of alleles of the human V gene encoding the V region most similar to the macaque V region	Human/macaque degree of identity (V region)	Human/macaque degree of identity (FRs only)	Mean interallelic degree of identity (V region)	Mean interallelic degree of identity (FRs only)
IGHV1-f	1099214726682:11453-12476	IGHV1-f*01	1	91.0	90.0	/	/
IGHV3-72	1099548049584:700053-701352	IGHV3-72*01	1	95.0	95.0	/	/
IGHV1-24	1099214764051:1-796	IGHV1-24*01	1	92.0	96.0	/	/
IGHV7-4-1 IGHV7-81	1099548049732:226371-227664	IGHV7-81*01	1	91.0	90.0	/	/
<b>5</b>	<b>4</b>	<b>4</b>		<b>92.3</b>	<b>92.8</b>		
IGHV3-23 IGHV3-53 IGHV3-d IGHV3-66	1099214148171:1052-1950	IGHV3-23*04	5	90.0	92.0	95.8	97.5
IGHV3-35 IGHV3-9 IGHV3-11 IGHV3-20	1099214739418:108-1401	IGHV3-23*04	5	88.0	92.0	95.8	97.5
IGHV3-15 IGHV5-51	1099548049732:81017-82310	IGHV3-23*04	5	89.0	91.0	95.8	97.5
IGHV6-1 IGHV3-15	1099548049110	IGHV6-1*02	2	93.0	92.0	100.0	100.0
IGHV3-15 IGHV5-51	1099548049110:4864-6163	IGHV3-15*05	8	91.0	95.0	96.6	97.7
IGHV4-30-2 IGHV4-30-4 IGHV4-31 IGHV4-39 IGHV4-59 IGHV4-61	1099548049110:62010-63303	IGHV5-51*03	4	92.0	93.0	98.7	98.7
IGHV4-30-2 IGHV4-30-4 IGHV4-31 IGHV4-39 IGHV4-59 IGHV4-61	1099548049584:146129-147419	IGHV4-61*01	6	94.0	96.0	96.8	98.0
IGHV3-74 IGHV3-73	1099548049584:167049-168342	IGHV3-74*01	3	91.0	91.0	99.0	99.0
IGHV3-38 IGHV3-43 IGHV3-48 IGHV3-64 IGHV3-21	1099548049584:175367-176663	IGHV3-73*02	2	94.0	97.0	100.0	100.0
IGHV3-38 IGHV3-43 IGHV3-48 IGHV3-64 IGHV3-21	1099548049584:289868-291155	IGHV3-48*03	3	91.0	92.0	96.5	99.0
IGHV1-58 IGHV1-18 IGHV1-2 IGHV1-3 IGHV1-45 IGHV1-46 IGHV1-8	1099548049584:379299-380592	IGHV1-3*01	2	84.0	83.0	96.0	97.0
IGHV1-18 IGHV1-2 IGHV1-3 IGHV1-45 IGHV1-46 IGHV1-8	1099548049584:38101-39394	IGHV1-2*02	4	90.0	92.0	97.0	97.0

All human IGHV genes are shown on the first (left) column. Macaque IGHV regions, retrieved from the *M. mulatta* sequenced genome by analogy with the regions encoded by human IGHV genes and identified by their encoding DNA (scaffold), are presented in the second column. Alleles of human IGHV regions, most homologous with each macaque IGHV region, are indicated by the name of their gene in the third column. Percents of identity existing between the regions encoded by each pair of most homologous macaque and human IGHV regions are shown in columns four and five (whole V region and framework regions only, respectively). If applicable, the percents of identity existing between alleles of each human IGHV regions are shown in the last columns (whole V region and framework regions only, columns six and seven). (The upper part presents IGHV genes which have no allelic polymorphism, and conversely for the lower part; both parts are separated by a bold black line). In each column, the number of sequences is presented on the last line of each part and on the last line of the table for the whole isotype.

**Table 1.** Analysis of the differences between human and macaque IGHV regions

IGHV1-69	1099548049584:438835-440122	IGHV1-69*09	12	91.0	95.0	96.3	97.9
IGHV4-b	1099548049584:491229-492525	IGHV4-b*01	2	90.0	95.0	99.0	99.0
IGHV2-5							
IGHV2-70	1099548049584:55193-56483	IGHV2-5*10	8	93.0	92.0	97.0	97.9
IGHV2-26							
IGHV4-28	1099548049584:565117-566413	IGHV4-4*02	6	91.0	92.0	95.2	95.4
IGHV3-30							
IGHV3-30-3	1099548049584:612516-613809	IGHV3-33*01	5	91.0	90.0	98.5	99.0
IGHV3-33							
IGHV3-16	1099548049584:654199-655492	IGHV3-7*01	2	89.0	91.0	100.0	100.0
IGHV3-7							
IGHV3-13	1099548049584:97252-98536	IGHV3-13*01	3	88.0	92.0	96.0	97.0
IGHV4-4	1099548049584:520128-521424	IGHV4-59*01	8	92.0	96.0	96.4	96.6
IGHV4-34	1099548049584:994-2281	IGHV4-59*04	8	92.0	95.0	95.4	95.4
IGHV3-49	1099548049732:160496-161795	IGHV3-49*04	5	93.0	95.0	96.8	97.3
46	22	20		90.8	92.7	97.3	98.0
51	26	24		91.0	92.7	97.3	98.0

All human IGHV genes are shown on the first (left) column. Macaque IGHV regions, retrieved from the *M. mulatta* sequenced genome by analogy with the regions encoded by human IGHV genes and identified by their encoding DNA (scaffold), are presented in the second column. Alleles of human IGHV regions, most homologous with each macaque IGHV region, are indicated by the name of their gene in the third column. Percents of identity existing between the regions encoded by each pair of most homologous macaque and human IGHV regions are shown in columns four and five (whole V region and framework regions only, respectively). If applicable, the percents of identity existing between alleles of each human IGHV regions are shown in the last columns (whole V region and framework regions only, columns six and seven). (The upper part presents IGHV genes which have no allelic polymorphism, and conversely for the lower part; both parts are separated by a bold black line). In each column, the number of sequences is presented on the last line of each part and on the last line of the table for the whole isotype.

When a gene encoding one of the identified twenty-four human IGHV regions showed no allelic polymorphism, comparison between human and macaque IGHV regions was performed as an alignment and the percent of identity was calculated. This was the case for four human IGHV regions encoded by genes IGHV1-f\*01, IGHV3-72\*01, IGHV1-24\*01 and IGHV7-81\*01 (upper part of Table 1); with percents of identity of 91, 95, 92 and 91% respectively, no strict identity between these homologous human and macaque IGHV regions was observed. When the comparison of these four homologous human and macaque IGHV regions was restricted to FRs, these percents of identity were equal to 90, 95, 96 and 90%, respectively, so that, again, no strict identity between FRs of human and macaque origins was observed.

Regarding the 20 IGHV human regions showing allelic polymorphism (between 2 and 12 alleles each) and identified as the most similar to macaque IGHV regions (lower part of Table 1), no strict identity between human and macaque homologs was observed, but, due to this allelic polymorphism, such a direct comparison of sequences was not sufficient and allelic polymorphism and human/macaque differences had to be compared by a statistical test. Only four genes (IGHV3-15, IGHV1-69, IGHV2-5, IGHV4-59) have more than 6 alleles and only these four genes would have allowed individual statistical tests to be performed. An ANOVA test was preferred and performed, and the 20 percents of identity between pairs of homologous macaque

and human IGHV regions (91% in average) were found to be very significantly smaller ( $p < 0.0001$ ) than the percents of identity between alleles of human IGHV regions (97.3% in average). The result was the same ( $p < 0.0001$ ) when the comparison was restricted to FRs (92.7% identity between human and macaque FRs of IGHV regions, 98% identity between FRs of human IGHV regions). Regardless the existence of allelic polymorphism of human IGHV regions, the average percent of identity existing between homologous human and macaque IGHV regions was evaluated as 91% (92.7% for FRs) when the average percent of identity between alleles (if any) of human IGHV regions was higher, at 97.3% (98% for FRs).

**V regions of  $\lambda$  isotype (IGLV).** Utilizing BLASTP and alleles \*01 of the 38 human germline IGLV regions,<sup>16-18</sup> 29 macaque putative IGLV regions were discovered. The 27 human IGLV regions most similar to these 29 macaque IGLV regions were identified with IMGT/DomainGapAlign. The IGLV region encoded by IGLV3-21\*01 was identified as the most similar to two macaque IGLV regions (encoded by scaffolds 1099553000069:1588271-1589531 and 1099553000069:1524648-1525908) and the IGLV region encoded by IGLV7-46\*01 was identified as the most similar to two macaque IGLV regions (encoded by scaffold 1099553000069:2046219-2047488 and scaffold 1099553000069:1994667-1995936). When human IGLV regions showed no allelic polymorphism, comparison between human and macaque IGLV regions was straightforward. This was the

case for 13 human IGLV regions (upper part of Table 2). There was no strict identity between these homologous human and macaque IGLV regions, which have percents of identity ranging between 84 and 95% (mean value = 89.2%). When the comparison of these 24 human and macaque IGLV regions was restricted to FRs, the percents of identity ranged between 83.3 and 96.2% (mean value = 91.2%) and, again, no strict identity was observed.

An ANOVA test was performed for the study of the 14 other IGLV human regions, identified as the most similar to macaque IGLV regions, but showing allelic polymorphism of between 2 and 4 alleles each (lower part of Table 2). The percents of identity existing between paired macaque and human IGLV regions (92.9% in average) were found to be very significantly smaller ( $p < 0.0001$ ) than the percents of identity between alleles of human IGLV regions (98.7% in average). The result was the same ( $p < 0.0001$ ) when the comparison was restricted to FRs (94% identity between human and macaque FRs of IGLV regions, 99% identity between FRs of human IGLV regions). Overall, the percent of identity existing between human and macaque IGLV regions was evaluated as 91.2% (92.7% for FRs) when the percent of identity between alleles (if any) of human IGLV regions was 98.7% (99% for FRs).

**V regions of  $\kappa$  isotype (IGKV).** The analysis for the IGKV regions was conducted similarly to the IGHV and IGLV analysis shown previously, but starting from the 52 human IGKV regions.<sup>16-18</sup> Twenty putative macaque IGKV regions were identified, which were most similar to only 15 different human IGKV regions. Indeed, the IGKV region encoded by IGKV1-12\*02 was identified as the most similar to four macaque IGKV regions (encoded by scaffolds 1099548049194:368236-369499, 1099548049194:514207-515470, 1099548049195:12101-13364, 1099548049606:44384-45647), the IGKV region encoded by IGKV1-39\*01 was identified as the most similar to two macaque IGKV regions (encoded by scaffolds 1099548049194:699631-700894 and 1099548049194:720411-721674), and the IGKV region encoded by IGKV2D-29\*01 was identified as the most similar to two macaque IGKV regions (encoded by scaffolds 1099548049195:57878-59156 and 1099214727827:15754-17032). When a gene encoding one of the identified 15 human IGKV regions showed no allelic polymorphism, comparison between human and macaque IGKV regions was straightforward, as was the case for the nine human IGKV regions (upper part of Table 3). There was no strict identity between homologous human and macaque IGKV regions, which had percents of identity ranging between 84 and 97% (mean value = 92.2%). When the comparison of these nine human and macaque IGKV regions was restricted to FRs, these percents of identity ranged between 82.3 and 97.5% (mean value = 92.5%) so that, again, no strict identity was ever observed.

An ANOVA test was performed for the comparison of the six IGKV human regions showing allelic polymorphism (two alleles each) and identified as the most similar to macaque IGKV regions (lower part of Table 3). The six percents of identity between pairs of macaque and human IGKV regions (93.6% in average) were found to be significantly smaller ( $p = 0.03$ ) than the percents of identity between alleles of human IGKV regions (98.8%

in average). The result was the same ( $p = 0.04$ ) when the comparison was restricted to FRs regions (94.3% identity between FRs of human and macaque IGKV regions, 99% identity between FRs of human IGKV regions). Indeed, the small number of IGKV regions tested here may be the explanation of these results at the limit of significance, when ANOVA tests for IGHV and IGKV regions gave highly significant results. Overall, the percent of identity existing between human and macaque IGKV regions was evaluated as 92.8% (93.2% for FRs) when the percent of identity between alleles (if any) of human IGKV regions was 98.8% (99% for FRs).

## Discussion

In the present study, we have utilized the first (\*01) allele of all human V germline genes to identify putative macaque V regions, which were analyzed by comparison with their most similar human V regions. When several putative macaque V regions were identified starting from one human V gene allele, as was the case for 8 human alleles, at least one of these putative macaque V regions is probably functional while the others perhaps are not. Given the high consistency of the observations made in the course of the study, this limited number of occurrences and the associated risk of analysing 10 non-functional genes do not restrict its conclusions. Starting from 141 human V genes, 75 putative macaque V regions only were found and this might indicate a smaller V region repertoire in macaques (*Macaca mulatta*) than in humans; however, not all macaque V genes may have been identified by our procedure and a complete study of the macaque immunoglobulin loci would have been required to identify all these genes, even those perhaps too distant from human genes to be retrieved here. This was not performed and it is acknowledged that the present study could, in effect, be biased toward the identification of macaque V regions that are most human-like. Subsequently, the differences consistently observed here between human and macaque V genes may represent only a part of the total number of differences. A more comprehensive study, encompassing a complete search of the macaque genome for V regions and possibly including a study of the localization and physicochemical properties of the differences would be desirable. In spite or because, of those limitations, the first result to be emphasized is that no strict identity between pairs of homologous (most similar) human and macaque V regions was observed in the course of the present study.

BLASTP was utilized to identify amino acid sequences of macaque V regions similar to human V region, starting from macaque genome sequences as a translation of the macaque genome underlies the BLASTP process. The study was performed at the amino acid level, and not the nucleotide level, due to the relevance for the prediction of immune tolerance. After the identification of macaque V regions, the IMGT/DomainGapAlign tool, recently available on IMGT®, was essential to identify their most identical human counterparts. This search allowed formation of “pairs” of V regions, homologous but encoded by germline V genes of human and macaque origins, which were further studied.



**Table 2.** Analysis of the differences between human and macaque IGLV regions

Human V $\lambda$ gene	Macaque V gene ("scaffold") encoding the V region most similar to the region encoded by the human V gene (allele *01)	Human V gene allele encoding the V region most similar to the macaque V region	Nb of alleles of the human V gene encoding the V region most similar to the macaque V region	Human/macaque degree of identity (V region)	Human/macaque degree of identity (FRs only)	Mean interallelic degree of identity (V region)	Mean interallelic degree of identity (FRs only)
IGLV3-27	1099214733386:978-2238	IGLV3-27*01	1	93.0	93.6	/	/
IGLV3-1	1099553000069:1282904-1284164	IGLV3-1*01	1	85.0	89.7	/	/
IGLV3-19	1099553000069:1552136-1553396	IGLV3-19*01	1	95.0	96.2	/	/
IGLV3-32							
IGLV3-22	1099553000069:1568051-1569332	IGLV3-22*01	1	84.0	83.3	/	/
IGLV5-37	1099553000069:1968429-1969716	IGLV5-37*01	1	87.0	89.7	/	/
IGLV1-44	1099553000069:1985144-1986410	IGLV1-44*01	1	89.0	92.3	/	/
IGLV5-52	1099553000069:2229671-2230958	IGLV5-52*01	1	92.0	91.0	/	/
IGLV5-39							
IGLV5-45	1099553000069:2008529-2009816	IGLV5-48*01	1	91.0	94.9	/	/
IGLV5-48							
IGLV1-50	1099553000069:2031556-2032822	IGLV1-50*01	1	89.0	92.3	/	/
IGLV1-36	1099553000069:1890248-1891514	IGLV1-36*01	1	85.0	87.2	/	/
IGLV6-57	1099553000069:2328740-2330006	IGLV6-57*01	1	89.0	91.0	/	/
IGLV4-3	1099553000069:1414675-1415950	IGLV4-3*01	1	87.0	91.0	/	/
IGLV11-55	1099553000069:2361853-2363140	IGLV11-55*01	1	93.0	93.6	/	/
16	13	13		89.2	91.2		
<hr/>							
IGLV2-33	14164696	IGLV2-18*02	4	92.0	96.2	98.7	98.7
IGLV3-16	1099214728962:318-1578	IGLV3-25*03	3	95.0	96.2	99.0	99.0
IGLV3-25							
IGLV3-10	1099553000069:1464937-1466197	IGLV3-10*01	2	94.0	93.6	98.0	99.0
IGLV2-18							
IGLV2-23	1099553000069:1512475-1513744	IGLV2-8*01	2	94.0	94.9	99.0	99.0
IGLV2-8							
IGLV2-11	1099553000069:1577292-1578561	IGLV2-11*01	3	94.0	96.2	100.0	100.0
IGLV2-14							
IGLV3-21	1099553000069:1588271-1589531	IGLV3-21*01	3	90.0	91.0	96.5	97.5
IGLV3-12	1099553000069:1524648-1525908	IGLV3-21*01	3	90.0	91.0	96.5	97.5
IGLV3-9							
IGLV1-51	1099553000069:1923387-1924653	IGLV1-51*01	2	85.0	85.9	99.0	100.0
IGLV1-40	1099553000069:1928395-1929661	IGLV1-40*01	3	89.0	91.0	97.5	97.5
IGLV1-47	1099553000069:2003910-2005176	IGLV1-47*01	2	94.0	96.2	99.0	100.0
IGLV9-49	1099553000069:2015284-2016562	IGLV9-49*03	3	94.0	93.6	100.0	100.0
IGLV7-46	1099553000069:2046219-2047488	IGLV7-46*01	2	96.0	98.7	99.0	99.0
IGLV7-43	1099553000069:1994667-1995936	IGLV7-46*01	2	94.0	93.0	99.0	99.0
IGLV4-60	1099553000069:2249750-2251025	IGLV4-69*02	2	93.0	92.3	100.0	100.0
IGLV4-69							

Analysis of the differences between human and macaque IGHV regions. All human IGLV genes are shown on the left column. Macaque IGLV regions, retrieved from the *M. mulatta* sequenced genome by analogy with the regions encoded by human IGLV genes, and identified by their encoding DNA (scaffold), are presented in the second column. Alleles of human IGLV regions, most homologous with each macaque IGLV region, are indicated by the name of their gene in the third column. Percents of identity existing between the regions encoded by each pair of most homologous macaque and human IGHV regions are shown in columns four and five (whole V region and framework regions only, respectively). If applicable, the percents of identity existing between alleles of each human IGLV regions are shown in the last columns (whole V region and framework regions only, columns six and seven). The upper part presents IGLV genes which have no allelic polymorphism, and conversely for the lower part; both parts are separated by a bold black line. In each column, the number of sequences is presented on the last line of each part, and on the last line of the table for the whole isotype.

**Table 2.** Analysis of the differences between human and macaque IGLV regions

IGLV10-54	1099553000069:2338581-2339847	IGLV10-54*01	3	99.0	98.7	97.0	97.5
IGLV8-61	1099553000069:2464913-2466182	IGLV8-61*01	3	94.0	94.9	99.0	99.0
22	16	14		92.9	94.0	98.7	99.0
38	29	27		91.2	92.7	98.7	99.0

Analysis of the differences between human and macaque IGHV regions. All human IGLV genes are shown on the left column. Macaque IGLV regions, retrieved from the *M. mulatta* sequenced genome by analogy with the regions encoded by human IGLV genes, and identified by their encoding DNA (scaffold), are presented in the second column. Alleles of human IGLV regions, most homologous with each macaque IGLV region, are indicated by the name of their gene in the third column. Percents of identity existing between the regions encoded by each pair of most homologous macaque and human IGHV regions are shown in columns four and five (whole V region and framework regions only, respectively). If applicable, the percents of identity existing between alleles of each human IGLV regions are shown in the last columns (whole V region and framework regions only, columns six and seven). The upper part presents IGLV genes which have no allelic polymorphism, and conversely for the lower part; both parts are separated by a bold black line. In each column, the number of sequences is presented on the last line of each part, and on the last line of the table for the whole isotype.

This study depended on one exclusive available genome of a macaque specimen, and the macaque allelic polymorphism was thus not taken into account. However, as inferred from the existence of human V regions that do not present such polymorphism, certain macaque genes may also have one allele only. Consequently, differences consistently observed between homologous macaque and human V regions apply, at least in part, to V regions of other macaque specimens. In any case, the observation of one macaque V region, not identical to any of the known human V regions, is sufficient to constitute a counterexample of the “undistinguishable” character of macaque and human V regions. Following the “immunological self” concept, this observation is sufficient to invalid the conclusion that macaque V regions would be as well tolerated as human V regions for therapeutic use. Subsequently, such a counterexample is sufficient to refute the reasoning on which EP 1,266,965 or parallel patent are based and all observations performed in the course of the present study constitute such counterexamples.

For the present analysis, human and macaque V regions were grouped according to isotypes as a former study showed the degrees of similarity between human and macaque V regions to be influenced by their heavy or light chains origin.<sup>10</sup> For the three isotypes, only taking into account the genes for which no human allelic polymorphism is known, 37 direct pair-wise comparisons were performed. Percents of identity between V regions from humans and macaques ranged from 84–97%. When comparisons were restricted to FRs, the percents of identity ranged between 82.3–97.5% so that differences were always observed between human and macaque V regions, indeed reaching as much as 16% for V regions or 17.3% for FRs.

When a human V region is encoded by several alleles, direct comparisons utilizing the human V region most similar to the macaque V region are still possible. As for previous observations, no human/macaque identity was observed, but this conclusion may be regarded as weak. In effect, when allelic polymorphism exists, it might only be a matter of time before a human V region identical to a certain macaque V region may be found, due to the ever increasing availability of human genome sequencing. The conclusion to be reached would be regarded as a much stronger argument if the study evaluated whether or

not the inter-human allelic polymorphism was large enough to compensate for the human/macaque differences.

Allelic and human/macaque differences were compared by an ANOVA test for each isotype, and it was observed that human/macaque differences are significantly (IGKV regions) or very highly significantly (IGHV and IGLV regions) larger than human/human differences ( $p < 0.0001$ ,  $p < 0.0001$ ,  $p = 0.03$  for IGHV, IGLV, IGKV regions, respectively;  $p < 0.0001$ ,  $p < 0.0001$ ,  $p = 0.04$  for the corresponding FRs). The average percent of identity between all human and macaque most similar V regions is 91.6%, to be compared with 98% between all the alleles of human V regions, and 92.9% versus 98.5% for the corresponding FRs. These significantly larger human/macaque differences thus cannot be compensated by the smaller allelic polymorphism; in other words, it is not expected that a human V region would be found identical to a macaque V region.

Revealingly, when the average degrees of difference are considered, as 100 minus the percent of identity, there are in average 4.2-fold increased difference between most similar human and macaque V regions than between human V region alleles (8.4% vs. 2%) and 4.73-fold difference when FRs are considered (7.1% vs. 1.5%). Regardless of the existence or not of allelic polymorphism, germline macaque V regions and their FRs are different than their human counterparts. They may be distinguished based on their sequences only, contrary to what is stated in the above mentioned patents. These differences are likely to be immunogenic, all the more when they are surrounded by mutations brought during affinity maturation, and may form epitopes recognized by humans. In all cases, the human/macaque differences observed here induce risks of immunogenicity that are not encountered with the use of recombinant antibodies encoded by human V regions.

It can be noted that, for the IGKV isotype, there is no trend showing a higher similarity of these V regions with their human counterparts than for the other two isotypes, thus the limit of significance for results regarding IGKV is likely to be due to the smaller number of IGKV regions studied here (20 IGKV regions vs. 26 IGHV and 38 IGLV regions). This might also be the reason why macaque light chains were sometimes found undistinguishable from their human counterparts in a previous study.<sup>10</sup>

**Table 3.** Analysis of the differences between human and macaque IGKV regions

Human V <sub>κ</sub> gene	Macaque V gene ("scaffold") encoding the V region most similar to the region encoded by the human V gene (allele *01)	Human V gene allele encoding the V region most similar to the macaque V region	Nb of alleles of the human V gene encoding the V region most similar to the macaque V region	Human/ macaque degree of identity (V region)	Human/ macaque degree of identity (FRs only)	Mean interallelic degree of identity (V region)	Mean interallelic degree of identity (FRs only)
IGKV5-2	1099548049194:210359-211619	IGKV5-2*01	1	84.0	83.5	/	/
IGKV1D-42	1099548049194:387873-389136	IGKV1D-42*01	1	82.0	82.3	/	/
IGKV1-6	1099548049194:664236-665496	IGKV1-6*01	1	95.0	96.2	/	/
IGKV1-17	1099548049194:699631-700894	IGKV1-39*01	1	95.0	94.9	/	/
IGKV1-5	1099548049194:720411-721674	IGKV1-39*01	1	95.0	94.9	/	/
IGKV6D-41	1099548049194:752119-753382	IGKV6D-41*01	1	91.0	91.1	/	/
IGKV6-21	1099548049194:851038-852301	IGKV6-21*01	1	94.0	93.7	/	/
IGKV6D-21							
IGKV4-1	1099548049194:93296-94577	IGKV4-1*01	1	94.0	93.7	/	/
IGKV2-28	1099548049195:225733-227011	IGKV2-28*01	1	97.0	97.5	/	/
IGKV2D-28							
IGKV2-40	1099548049195:525065-526346	IGKV2-40*01	1	95.0	97.5	/	/
IGKV2D-40							
13	10	9		92.2	92.5		
IGKV1-13							
IGKV1-8							
IGKV1D-13							
IGKV1D-43							
IGKV1D-8							
IGKV1-12							
IGKV1-16							
IGKV1-27	1099548049194:368236-369499	IGKV1-12*02	2	98.0	98.7	100.0	100.0
IGKV1-39							
IGKV1-9							
IGKV1D-12							
IGKV1D-17							
IGKV1D-39							
IGKV1-37							
IGKV1D-37							
IGKV1-NL1	1099548049194:514207-515470	IGKV1-12*02	2	93.0	94.0	100.0	100.0
IGKV1-33	1099548049195:12101-13364	IGKV1-12*02	2	92.0	93.0	100.0	100.0
IGKV1D-33							
IGKV1D-16	1099548049606:44384-45647	IGKV1-12*02	2	96.0	96.0	100.0	100.0
IGKV3-NL4	1099548049194:572921-574184	IGKV3-11*01	2	94.0	94.9	98.0	98.0

Analysis of the differences between human and macaque IGKV regions. All human IGKV genes are shown on the first (left) column. Macaque IGKV regions, retrieved from the *M. mulatta* sequenced genome by analogy with the regions encoded by human IGKV genes, and identified by their encoding DNA (scaffold), are presented in the second column. Alleles of human IGKV regions, most homologous with each macaque IGKV region, are indicated by the name of their gene in the third column. Percents of identity existing between the regions encoded by each pair of most homologous macaque and human IGKV regions are shown in columns four and five (whole V region and framework regions only, respectively). If applicable, the percents of identity existing between alleles of each human IGKV regions are shown in the last columns (whole V region and framework regions only, columns six and seven). The upper part presents IGKV genes which have no allelic polymorphism, and conversely for the lower part; both parts are separated by a bold black line. In each column, the number of sequences is presented on the last line of each part, and on the last line of the table for the whole isotype.



**Table 3.** Analysis of the differences between human and macaque IGKV regions

IGKV3-15							
IGKV3-20							
IGKV3-7							
IGKV3-NL1	1099548049194:684688-685951	IGKV3-20*01	2	94.0	94.9	98.0	98.0
IGKV3-NL3							
IGKV3D-15							
IGKV3D-7							
IGKV2-30							
IGKV2D-24	1099548049194:945813-947091	IGKV2-30*02	2	90.0	89.9	99.0	100.0
IGKV2D-30							
IGKV2-24							
IGKV2D-26	1099548049195:57878-59156	IGKV2D-29*01	2	90.0	92.5	99.0	99.0
IGKV2-29	1099214727827:15754-17032	IGKV2D-29*01	2	92.0	93.7	99.0	99.0
IGKV2D-29							
IGKV3-11							
IGKV3-NL2							
IGKV3-NL5	1099548049606:39933-41196	IGKV3-11*01	2	93.0	93.7	99.0	99.0
IGKV3D-11							
IGKV3D-20							
39	10	6		93.6	94.3	98.8	99.0
52	20	15		92.8	93.4	98.8	99.0

Analysis of the differences between human and macaque IGKV regions. All human IGKV genes are shown on the first (left) column. Macaque IGKV regions, retrieved from the *M. mulatta* sequenced genome by analogy with the regions encoded by human IGKV genes, and identified by their encoding DNA (scaffold), are presented in the second column. Alleles of human IGKV regions, most homologous with each macaque IGKV region, are indicated by the name of their gene in the third column. Percents of identity existing between the regions encoded by each pair of most homologous macaque and human IGKV regions are shown in columns four and five (whole V region and framework regions only, respectively). If applicable, the percents of identity existing between alleles of each human IGKV regions are shown in the last columns (whole V region and framework regions only, columns six and seven). The upper part presents IGKV genes which have no allelic polymorphism, and conversely for the lower part; both parts are separated by a bold black line. In each column, the number of sequences is presented on the last line of each part, and on the last line of the table for the whole isotype.

In conclusion, V regions encoded by germline genes of a macaque (*Macaca mulatta*) specimen were located in its genome by homology with human V regions, and pairs of most similar, germline, human and macaque V regions were formed. No identity between such pairs was ever observed. Moreover, when allelic polymorphisms occur in human V regions, they are significantly smaller than the human/macaque differences and it is unlikely that a new human allele will ever be discovered as being identical to a macaque V gene. These human/macaque V region differences are directly encountered in unmutated IgMs, and also constitute differences upon which mutations introduced during each IgG affinity maturation process will be added. Following the “immunological self” concept, these human/macaque differences could be more immunogenic than the differences due to human allelic polymorphism, when it exists. To minimize such risks of immunogenicity when therapeutic use of macaque V regions is considered, it is certainly useful that these regions are humanized. We have shown such an example of humanization (germline humanization) where the final allele was more germline-like, thus probably less immunogenic, than a functionally equivalent, fully human V

region. The results above may facilitate the access to efficient and well-tolerated recombinant antibodies without undue legal constraints.

## Materials and Methods

**Numbering and nomenclature.** The IMGT unique numbering for V regions and for V domains ([www.imgt.org/textes/IMGTScientificChart/Numbering/IMGTIGVLSuperfamily.html](http://www.imgt.org/textes/IMGTScientificChart/Numbering/IMGTIGVLSuperfamily.html))<sup>20,21</sup> was utilized for this study. V regions are defined as regions encoded by V genes, while variable domains are encoded by V, (D), J genes. IGHV are V regions of  $\gamma$  isotype, and correspondingly for IGLV and  $\lambda$  isotype, and IGKV and  $\kappa$  isotype; this nomenclature allows to distinguish V regions with variable regions, named  $V\gamma$ ,  $V\lambda$  and  $V\kappa$  respectively. Anchor residues, CDRs and FRs were defined according to IMGT.<sup>18</sup> IMGT gene and allele names were utilised in this study, in particular alleles are identified by a number following a \*, itself following the name of the gene (example: IGHV3-23\*04 is an allele of gene IGHV3-23). Otherwise IMGT nomenclature was not always followed and older, but more common wording

was sometimes preferred, e.g., isotype instead of group, or Ig instead of IG.

**Human V regions.** The sequences of human IGHV, IGLV and IGKV germline genes encoding antibodies were retrieved from the IMGT/GENE-DB database (<http://www.imgt.org>).<sup>17</sup> Only functional (F) and open reading frames (ORF) genes were retrieved (partial alleles and pseudogenes (P) were not used). Due to the fact that only a portion of CDR3s is part of the V region, these CDR3 portions were excluded from the study. Consequently, the V region after residue 104 was not taken into account and only the portions comprised between the first residue of FR1 and the last residue of FR3 (2<sup>nd</sup>-CYS 104) were utilised for the study.

**Macaque V regions.** The germline encoded (IgM) sequences of macaque IGHV, IGLV and IGKV regions were retrieved from the sequenced genome of a macaque (*Macaca mulatta*) specimen, available on-line from the Human genome sequencing centre site web (<http://www.hgsc.bcm.tmc.edu>);<sup>22</sup> more precisely the “*M. mulatta* whole genome assembly scaffolds as of 2005-Dec-12” was utilized. The V regions encoded by allele \*01 of each human V gene, from position 1 to 2<sup>nd</sup>-CYS 104, were aligned by BLASTP<sup>15</sup> against the macaque genome in order to identify amino acid sequences regarded as putative macaque V regions in the rest of the study. These putative macaque V regions were identified by the positions on the genomic scaffolds (example: scaffolds 1099214148171:1052-1950), regarded as putative V genes.

**Retrieval of the human V regions, most similar to the macaque V regions.** The alleles of human V region most similar to each macaque putative V regions were retrieved in a second step with IMGT/DomainGapAlign ([www.imgt.org](http://www.imgt.org)),<sup>19</sup> which allows comparisons at the amino acid level. Pairs of most homologous human and macaque V regions were formed. Similarly to macaque V regions, these human V regions were identified by the names of their coding genes.

**Analysis of V regions encoded by macaque and human germline genes.** The percents of identity between paired human and macaque V regions were calculated using ClustalW,<sup>23</sup> for entire

V regions and for FRs only. When the human V region presented no allelic polymorphism, i.e., when there was only one allele for that V region, this calculation made up the entire analysis of the percent of identity between pairs of human and macaque V regions. In addition to that calculation, when the human V region presented allelic polymorphism, the other alleles of the corresponding V gene were translated and defined the allele(s) of this V region. The degree(s) of identity existing between each of these human V regions and their alleles were calculated using ClustalW-multialign tool (<http://mobylipe.pasteur.fr>), for V regions and for FRs only. The percents of identity existing between pairs of human and macaque V region were compared with the percents of identity existing between human V alleles by a statistical test.

**Statistical tests.** For each human V region paired with a macaque V region and presenting allelic polymorphism, the percents of identity corresponding to these human alleles had to be compared with the percents of identity existing between the pair of human and macaque V regions. This comparison had to rely on a statistical test, to assess whether or not the difference was significant. The number of human V gene alleles is generally less than seven, so an analysis for each gene was generally not possible. For this reason, an ANOVA test was performed to simultaneously compare, within each isotype ( $\gamma$ ,  $\kappa$  or  $\lambda$ ), all percents of identity corresponding to the allelic polymorphism with the percents of identity between all human and macaque V regions. The analysis was performed for each isotype because isotypes previously appeared to influence the similarities between human and macaque V regions.<sup>10</sup> These three two-way ANOVA tests were performed with GraphPad Prism version 5.00 for Windows (GraphPad Software, San Diego, CA, USA) ([www.graphpad.com](http://www.graphpad.com)).

#### Acknowledgments

This study was funded by Plan d'étude amont 09co302-1, from Direction générale de l'armement. We thank Jeffrey W. Froude II for manuscript improvements.

#### References

1. Schutte M, Thullier P, Pelat T, Wexler X, Rosenstock P, Hinz D, et al. Identification of a putative Crf splice variant and generation of recombinant antibodies for the specific detection of *Aspergillus Fumigatus*. PLoS One 2009;4:6625.
2. Pelat T, Hust M, Hale M, Lefranc MP, Dubel S, Thullier P. Isolation of a human-like antibody fragment (Scfv) that neutralizes ricin biological activity. BMC Biotechnol 2009; 9:60.
3. Pelat T, Hust M, Laffly E, Condemine F, Bottex C, Vidal D, et al. High-affinity, human antibody-like antibody fragment (single-chain variable fragment) neutralizing the lethal factor (LF) of *Bacillus anthracis* by inhibiting protective antigen-LF complex formation. Antimicrob Agents Chemother 2007; 51:2758-64.
4. Laffly E, Danjou L, Condemine F, Vidal D, Drouet E, Lefranc MP, et al. Selection of a macaque Fab with framework regions like those in humans, high affinity and ability to neutralize the protective antigen (PA) of *Bacillus anthracis* by binding to the segment of PA between residues 686 and 694. Antimicrob Agents Chemother 2005; 49:3414-20.
5. Chassagne S, Laffly E, Drouet E, Herodin F, Lefranc MP, Thullier P. A high-affinity macaque antibody Fab with human-like framework regions obtained from a small phage display immune library. Mol Immunol 2004; 41:539-46.
6. Pelat T, Bedouelle H, Rees AR, Crennell SJ, Lefranc MP, Thullier P. Germline humanization of a non-human primate antibody that neutralizes the anthrax toxin, by in vitro and in silico engineering. J Mol Biol 2008; 384:1400-7.
7. Pelat T, Thullier P. Non-human primate immune libraries combined with germline humanization: an (almost) new, and powerful approach for the isolation of therapeutic antibodies. Mabs 2009; 1:377-81.
8. Pelat T, Hust M, Thullier P. Obtention and engineering of non-human primate (NHP) antibodies for therapeutics. Mini Rev Med Chem 2009; 9:1633-8.
9. Abhinandan KR, Martin AC. Analyzing The “degree of humanness” of antibody sequences. J Mol Biol 2007; 369:852-62.
10. Thullier P, Huish O, Pelat T, Martin AC. The humanness of macaque antibody sequences. J Mol Biol 2010; 396:1439-50.
11. Harding FA, Stickler MM, Razo J, Dubridge RB. The immunogenicity of humanized and fully human antibodies: residual immunogenicity resides in the CDR regions. Mabs 2010; 2:256-65.
12. Jefferis R, Lefranc MP. Human immunoglobulin allotypes: possible implications for immunogenicity. Mabs 2009; 1:332-8.
13. Behn U. Idiotypic networks: toward a renaissance? Immunol Rev 2007; 216:142-52.
14. Sfikakis PP. The first decade of biologic TNF antagonists in clinical practice: lessons learned, unresolved issues and future directions. Curr Dir Autoimmun 2010; 11:180-210.
15. Altschul SF, Gish W. Local alignment statistics. Methods Enzymol 1996; 266:460-80.
16. Lefranc M-P, Lefranc G. The Immunoglobulin Factsbook. San Diego: Academic Press 2001.
17. Giudicelli V, Chaume D, Lefranc MP. IMGT/GENE-DB: A comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Res 2005; 33:256-61.
18. Lefranc MP. Nomenclature of the human immunoglobulin genes. Curr Protoc Immunol 2001; 1:1.

- 
19. Ehrenmann F, Kaas Q, Lefranc MP. IMGT/3D structure-DB and IMGT/DomainGapAlign: A database and a tool for immunoglobulins or antibodies, T cell receptors, Mhc, Igsf and Mhcsf. *Nucleic Acids Res* 2010; 38:301-7.
  20. Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, Truong L, et al. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* 2003; 27:55-77.
  21. Lefranc MP. IMGT unique numbering for immunoglobulins, T-cell receptors and Ig-like domains. *Immunologist* 1999; 7:132-6.
  22. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, et al. Evolutionary and biomedical insights from the Rhesus macaque genome. *Science* 2007; 316:222-34.
  23. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using Clustalw and Clustalx. *Curr Protoc Bioinformatics* 2002; 2:23.