DATE: 10/10/2022

WEBLEM 2

Introduction to Antibody sequence and structure along with Antibody numbering methods such as Kabat, Chothia and other equivalent methods and its importance

Monoclonal antibodies are playing an increasing role in both human and animal health. Different strategies of protein and chemical engineering, including humanization techniques of non-human antibodies were applied successfully to optimize clinical performances of antibodies. Despite the emergence of techniques allowing the development of fully human antibodies such as transgenic Xeno-mice, antibody humanization remains a standard procedure for therapeutic antibodies. An important prerequisite for antibody humanization requires standardized numbering methods to define precisely complementary determining regions (CDR), frameworks and residues from the light and heavy chains that affect the binding affinity and/or specificity of the antibody-antigen interaction. The recently generated deep-sequencing data and the increasing number of solved three-dimensional structures of antibodies from human and non-human origins have led to the emergence of numerous databases. However, these different databases use different numbering conventions and CDR definitions. In addition, the large fluctuation of the variable chain lengths, especially in CDR3 of heavy chains (CDRH3), hardly complicates the comparison and analysis of antibody sequences and the identification of the antigen binding residues.

Numbering Schemes of Antibody Variable Domains

Antibody engineering methods require precise identification of the residues that have an impact on the interaction and/or affinity of the antibody for its target antigen. For example, as mentioned above, CDR-grafting aims to decrease the immunogenicity of non-human antibodies by engineering the variable regions directed against the target antigen. This method requires an accurate identification of the CDRs and therefore an adequate alignment of antibody sequences from human and non-human species. It has been shown that residues from the framework regions might also exert a strong impact on the antibody affinity. Thus, the precise identification of corresponding positions in human and animal immunoglobulin chains is essential. However, the use of different amino acid numbering schemes currently available in the literature is confusing and might lead to aberrant identification of framework and CDR residues. Therefore, it is of crucial importance to understand the different numbering schemes and, consequently, being able to compare them.

Kabat numbering scheme

Over the past decades, sequencing and crystallization of antibodies resulted in significant increase of various sequence and structure databases, which made the comparison of the variable regions from human and animal immunoglobulins possible. In 1970, Kabat and Wu aligned 77 Bence-Jones protein and immunoglobulin light chain sequences in order to study the statistical variability in amino acid composition at the sequential positions of the variable antibody regions. They defined the “variability parameter” as the number of different amino acids at a given position divided by the frequency of the most occurring amino acid at that position. This analysis revealed three hypervariable regions in the variable region of the light chains. The presence of highly conserved residues was also demonstrated, such as the two cysteines that form a disulphide bridge at the inner core of the immunoglobulin domain and a tryptophan residue located immediately after CDRL1. Likewise, three corresponding hypervariable regions were also identified in the variable heavy chain domain. Kabat and Wu postulated that these hypervariable regions would cluster at one side of the folded domain to form a surface responsible for specific antigen recognition and referred to these hypervariable regions as “Complementarity Determining Regions” “CDR”-1,-2, and-3. This hypothesis was later confirmed and further investigated to distinguish antigen-contacting or conformational important residues within these CDRs.

In 1979, Kabat et al. were the first to propose a standardized numbering scheme for the variable regions of immunoglobulins. In their compilation of “Sequences of Proteins of Immunological Interest”, the amino acid sequences of the variable region of the light (λ, κ) and heavy chain of antibodies, as well as the variable region

of T cell receptors (α, β, γ, δ) were aligned and numbered. They observed that the analyzed sequences exhibited variable lengths and that gaps and insertions could only be included at precise positions. Interestingly, the points of insertion were located inside the CDRs, except for CDRL2, but also at some positions inside the framework regions. In the numbering schemes, these insertions are identified and annotated with letters (e.g., 27a, 27b…). It is also noticeable that residue L10 is absent in all the λ light chains, while λ and κ chains are being coded by two different genes, located on different chromosomes. Over the last decades, the accumulation of sequences resulted in the creation of the KABATMAN database.

Although the Kabat numbering scheme is often considered as the standard that is widely adopted for numbering antibody residues, it has some important limitations. Firstly, this scheme was built on the alignments of a limited number of sequences from antibodies with the most common sequence lengths. Consequently, sequences with unconventional insertions or deletions in the CDRs or in the framework regions were not included. Therefore, the original Kabat scheme ignores antibody chains of unconventional lengths, with unique insertions or deletions. However, a useful numbering tool named *ABnum* that numbers the amino acid sequences of variable domains according to a much larger and regularly updated database (Abysis), takes into account insertions of variable lengths, particularly in CDR2 by adding an insertion point at position L54. The second main limitation of the Kabat scheme is that it doesn't match very well with the 3D structure of antibodies. Indeed, the hypervariable regions defined by Kabat do not exactly match with the structural antigen-binding loops. The defined insertion points in CDR-L1 (L27) and CDR-H1 (H35) do not fit with their corresponding positions in the structures. In other words, the corresponding residues (topologically aligned) in crystal structures in CDR-L1 and CDR-H1 don't share the same number in the Kabat numbering scheme.

Chothia numbering scheme

In 1987, Chothia and Lesk introduced a structure-based numbering scheme for antibody variable regions. They aligned crystal structures of antibody variable regions, defined the loop structures that form the CDRs and corrected the position numbers of the insertion points inside CDRL1 and CDRH1 so that they better fit their topological positions. Furthermore, they classified the CDR loops of heavy and light chains in a small number of conserved structures, called “canonical” classes.

Based on the alignment of antibody structures, the Chothia numbering scheme shifts the point of amino acid insertion from position L27 to L30 and from position H35 to H31. It is worth mentioning that the Chothia CDR definition ensures a better correspondence to the structural loops. The loop structure of CDRH3 identified by Chothia matches well the Kabat hypervariable region. In contrast, the other loops are shorter than the hypervariable sequences defined by Kabat, except for CDRH1 which extends from H26 to H32. In any case, the CDRs defined on the hypervariable amino acids according to Kabat and based on loop topology in Chothia's nomenclature have for some CDR's a shifted location and/or comprise deviating loop lengths.

The Chothia numbering scheme possesses the main advantage that topologically aligned residues from different antibodies are localized at the same position number and that the Chothia CDR definition corresponds in most antibody sequences to the structural antigen-binding loops. However, confusion can also arise given the limited use of this numbering scheme compared to the Kabat or the IMGT numbering schemes (see below). Furthermore, a later study published by Chothia et al. changed the insertion point in CDR L1 from residue L30 to L31. However, while investigating the conformation of the antigen-binding loops, of antibodies present in larger databases, they returned to the initial L30 position in 1997. In a similar way, they initially defined an insertion point at position L93 in λ light chains that was shifted to position L95 in their subsequent study. Finally, an important limitation of this numbering scheme is due to the use of the most common CDR sequence lengths, like the Kabat numbering scheme, and therefore the Chothia scheme ignores sequences with unconventional length. However, similarly to the Kabat numbering scheme, this system could be optimized by defining new insertion points.

For numbering antibody KabatMan database can be used.

Kabatman Database:

Kabatman database provides various tools for antibody informatics:

- abYsis - Integrated database and analysis workbench
- abYmod - Antibody modelling software
- abYbank - Antibody sequence and structure data.

    This includes:

    - AbDb - pre-numbered structures from the PDB
    - SACS - Summary of Antibody Crystal Structures
    - EMBLIg - antibody sequences from EMBL-ENA
    - Kabat - FASTA formatted sequences from Kabat (2000)
    - AbPDBSeq - FASTA formatted antibody sequences from the PDB

- Humanness (G)- Assess humanness against expressed sequences in Kabat divided into germline families
- PAPS - Predict VH/VL packing angle
- abYdraw - A downloadable software package for drawing antibody cartoons

The following are included within abYsis, but also available as standalone tools:

- KabatMan: Query the Kabat sequence data
- AbCheck: Test a sequence against the Kabat data for unusual residues
- Chothia canonicals: Identify canonical classes for CDRs from your sequence
- Human subgroups: Assign the human subgroup for your sequence
- Humanness (H): Assess humanness against expressed sequences in Kabat
- Abnum: Apply standard numbering to sequences or structures

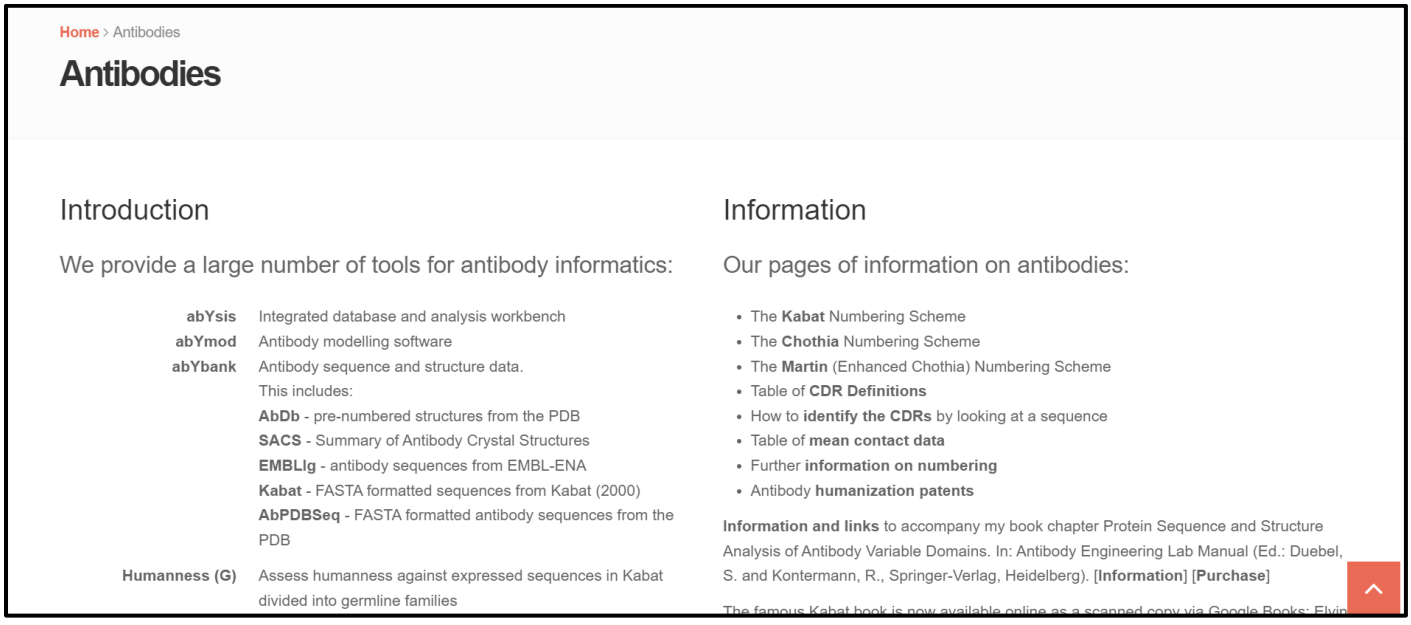

Fig1. Different tools available for Antibodies studies

KabatMan:

KabatMan interface lets you create simple queries for more complex cases, you must write queries directly using the in the KabatMan SQL-like query language. The KabatMan interface is shown below:

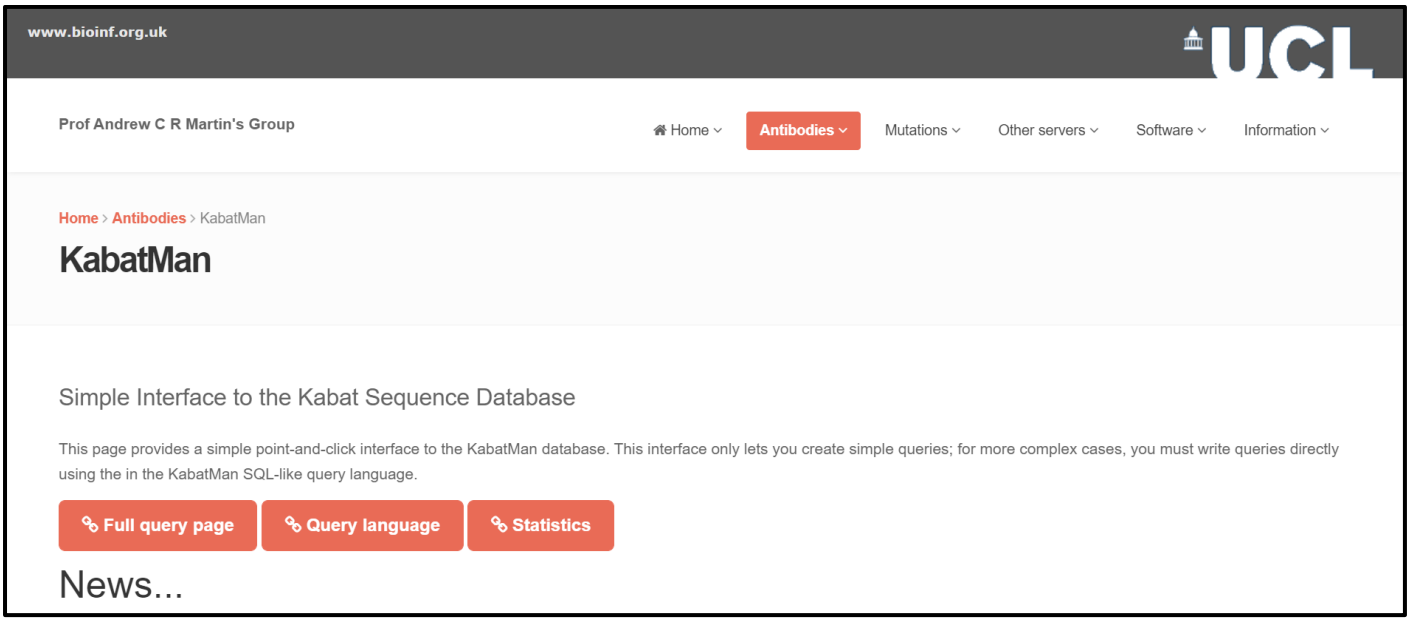

Fig2. Simple Interface to the Kabat Sequence Database

KabatMan - Full queries

This page allows you to query the Kabat antibody sequence database using an SQL-like query language. The main deviation between the language used here and SQL is that clauses within the WHERE statement are combined in reverse polish notation. Also, since there is only one table in the database, there is no FROM statement.

The query language uses three statements: SET, SELECT and WHERE. The SELECT and WHERE commands take you into a mode where clauses are specified with no introductionary SELECT or WHERE keyword. Conversely, the SET command needs to be specified on each line where variables are to be set and must be given *before* any SELECT or WHERE statement. The QUIT or EXIT command is used to leave the program**.**

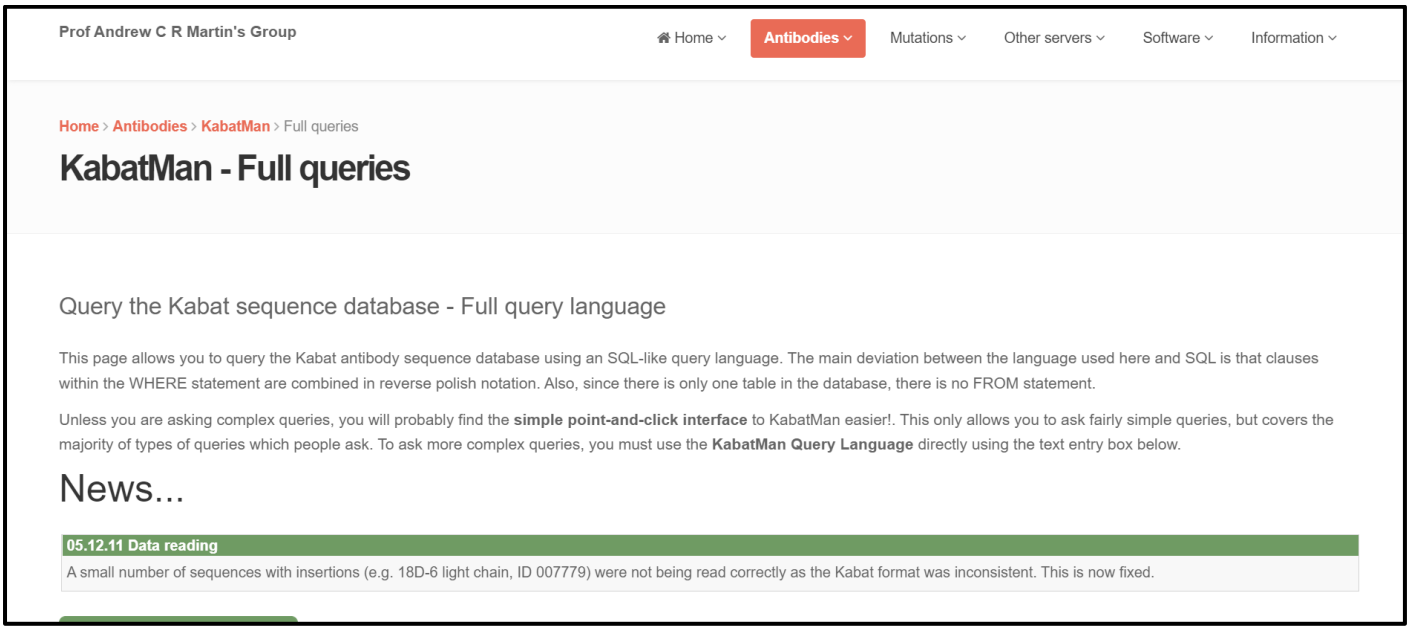

Fig3. KabatMan – Full Query language interface

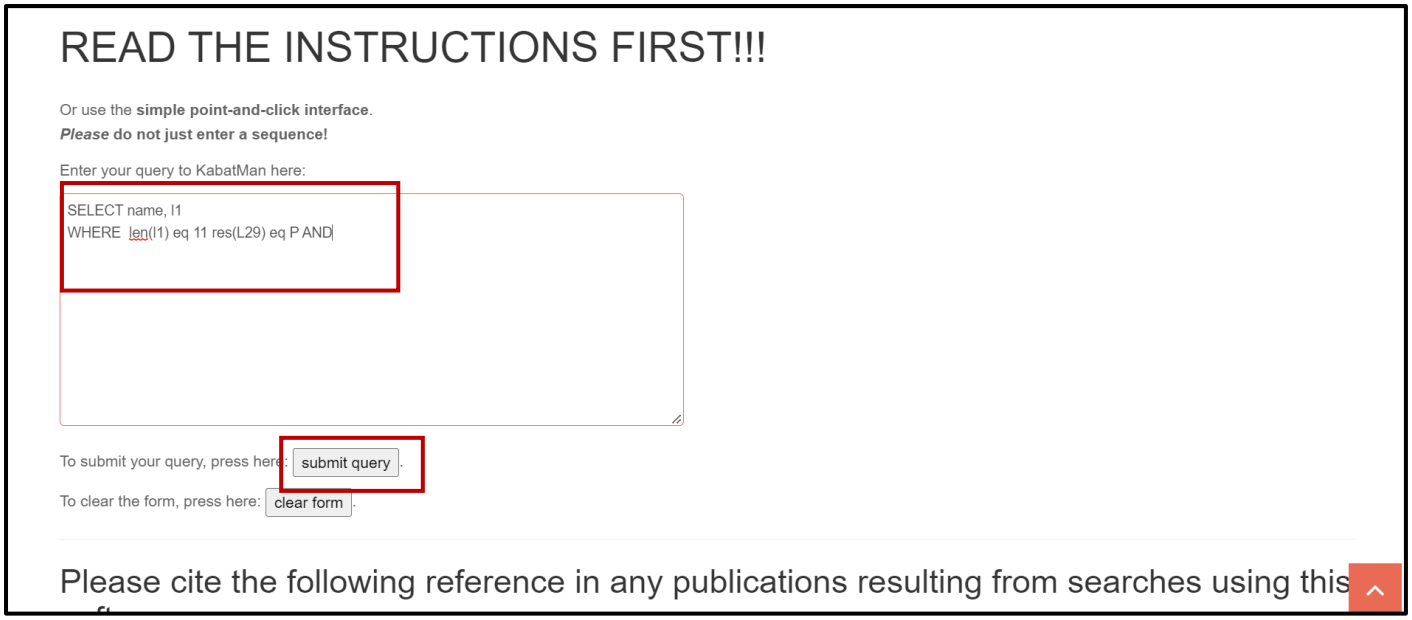

Fig4. Example of query to find all antibodies with 11 residue CDR-L1s and a proline at the sixth position

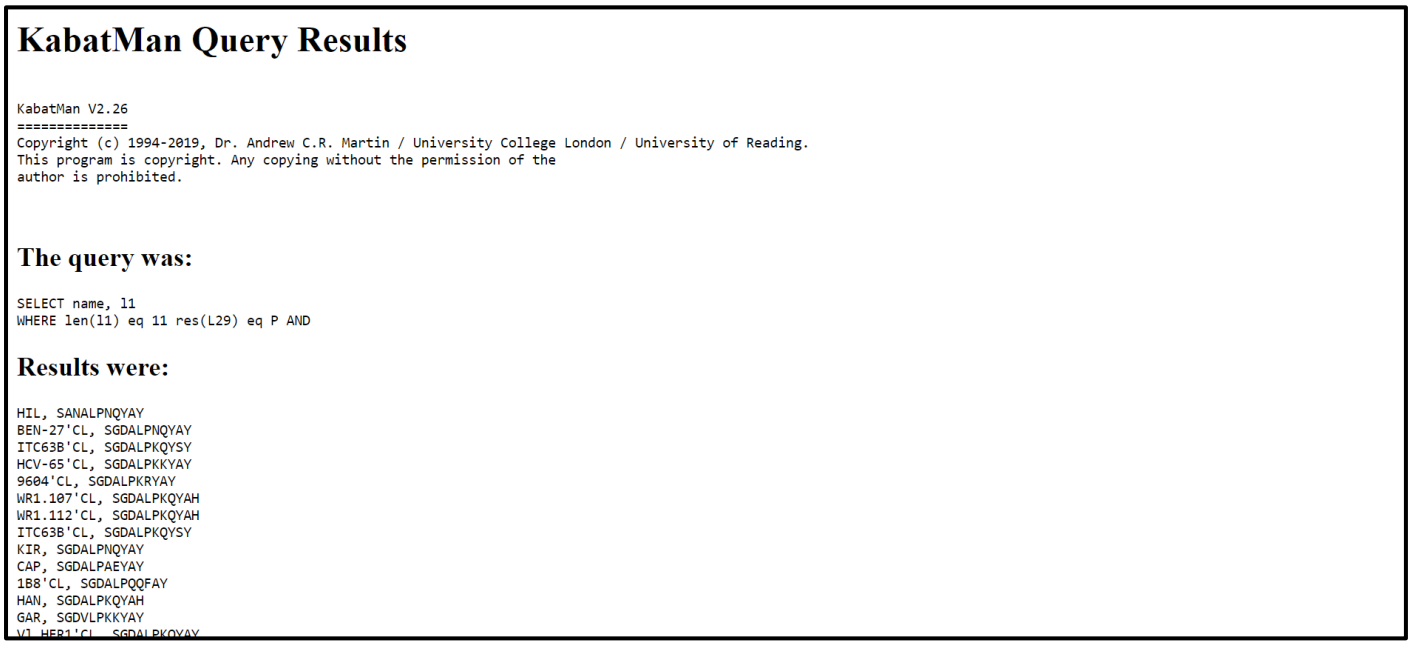

Fig5. Results for the query

The above results show all antibodies with 11 residue CDR-L1s and a proline at the sixth position.

Thus, KabatMan database is using for querying Kabat sequence data. This information is useful for antibody engineering which requires precise identification of the residues that have an impact on the interaction and/or affinity of the antibody for its target antigen. For example, CDR-grafting aims to decrease the immunogenicity of non-human antibodies by engineering the variable regions directed against the target antigen. This method requires an accurate identification of the CDRs and therefore an adequate alignment of antibody sequences from human and non-human species. Moreover, it has been shown that residues from the framework regions might also exert a strong impact on the antibody affinity. Thus, the precise identification of corresponding positions in human and animal immunoglobulin chains is essential.

REFERENCES:

1. *bioinf.org.uk - Prof. Andrew C.R. Martin's group at UCL*. (n.d.). Www.bioinf.org.uk. Retrieved October 10, 2022, from http://www.bioinf.org.uk/abs/
2. Dondelinger, M., Filée, P., Sauvage, E., Quinting, B., Muyldermans, S., Galleni, M., & Vandevenne, M. S. (2018). Understanding the Significance and Implications of Antibody Numbering and Antigen-Binding Surface/Residue Definition. *Frontiers in Immunology*, *9*. https://doi.org/10.3389/fimmu.2018.02278