Protecting amino acid and nucleic acid sequence variants

By Carlos Maíllo

The great success of biologics —drugs produced from living organisms or that contain components of living organisms— over the last decades, including blockbuster monoclonal antibodies and COVID-19 RNA vaccines, has accelerated the shift of pharmaceutical companies towards the development of these complex drugs for addressing many unmet medical needs. This has translated into increasing numbers of patent applications seeking the protection of biological sequences, in particular amino acid and nucleotide sequences.

Patent claims directed to specific biosequences are commonly drafted to also encompass closely related sequence variants in order to block competitors from marketing similar sequences bearing immaterial modifications —which would otherwise had to be addressed under the doctrine of equivalents and its inherent uncertainty.

One of the most common ways to claim sequence variants is by setting a minimum degree of resemblance that the variant and the sequence in question must bear. This degree of resemblance is commonly expressed with terms inherited from the field of comparative genomics, such as *sequence identity*, *sequence similarity*, and *sequence homology*.

The use of these terms in patent documents does not come without its challenges as they were conceived to deal with biological questions, not legal ones. Thus, it has been very much welcomed by practitioners in the biomedical field the presence of a new section in the last version of the EPO *Guidelines for Examination* that entered into force on 1 March 2021, where the use of these terms is discussed (GL 2021, *Part F –Chapter IV, 4.24 - Interpretation of terms such as identity and similarity in relation to amino or nucleic acid sequences*).

According to the new Guidelines, the *percentage of identity* relates to the number of identical residues over a defined length in a given alignment. The new Guidelines also clarify that if no algorithm or calculation method for determining the percentage of identity is defined in the patent, the broadest interpretation will be applied using any reasonable algorithm or calculation method known at the relevant filing date.

Likewise, the new Guidelines point out that if a *percentage of similarity* is used —which is a broader term than *percentage of identity* as it also encompasses conservative substitutions of amino acid residues—, a similarity-scoring matrix should be defined in the description, otherwise the any reasonable matrix known at the filing date will be considered for its determination.

Finally, the new Guidelines indicate that if a *percentage of homology* is used, the term will be considered unclear according to Art. 84 EPC unless the calculation of the percentage of homology is clearly defined in the application as filed.

While the newly introduced section offers some valuable guidance, it still leaves several issues unaddressed that makes it sometimes difficult to interpret the scope of protection of patents claiming sequence variants, in particular those that do not include a clear description of the calculation method to be applied.

For example, even the seemingly clearest term, *sequence identity,* can have multiple meanings and, in many cases, patents do not indicate which one is to be used. As stated above, this term refers to the number of identical residues over a defined length in a given sequence alignment. Therefore, one can obtain completely different identity values when comparing two given sequences depending on the type of algorithm and parameters used for the sequence alignment (e.g., local or global alignment algorithms, type of weight matrix, gap cost value, etc.), and the length defined for calculating the percentage of identity (i.e., the alignment length, query length, or subject length).

Thus, the indication in the new Guidelines that "the broadest interpretation will be applied using any reasonable algorithm or calculation method" can be used by patent proprietors to try to extend the sequence protection as much as possible, but also by third parties to try to invalidate the patent. In order to reduce this uncertainty, it may be advisable to at least identify in the patent document the alignment algorithm to be used.

Perhaps the most common algorithm used to make sequence alignments is BLAST (Basic Local Alignment Search Tool). BLAST standard tool performs local alignments and provides percentages of sequence identity based on alignment length. Thus, BLAST will deliver a 100% sequence identity when aligning a given sequence with any of its fragments, or when aligning a sequence with a much longer sequence that comprises it. Likewise, BLAST will often provide sequence identity values around 100% when aligning different splicing variants of a gene because it will focus on the common regions and will ignore the non-common (spliced-out) parts.

Although BLAST can be a valuable starting point when performing sequence searches and sequence comparisons, the results obtained should be carefully analysed as the type of alignment it performs and the identity values it provides many times do not reflect claim language and can be misleading.

Global alignment algorithms —that try to align every residue within the sequences to be compared— or best-fit alignment algorithms —that seek how to best fit the query sequence into the subject sequence— are generally better suited to compare closely related sequences or sequences of similar size. Therefore, in many cases, these types of algorithms will provide more meaningful results in the context of patents —although they will probably not be considered the broadest interpretation in the cases where no algorithm is specified.

All in all, nowadays it seems to be paramount that patent practitioners in the field of biomedicine fully understand the ins and outs of sequence search and sequence alignment when biological sequences are at play. Knowing just one sequence comparison method may be sufficient for patent drafting, but in no case for assessing claim infringement, as the scope of protection of claims will greatly depend on the comparison method indicated in the patent or, if no method is specified, on the broadest interpretation that can be applied —although it remains to be seen whether the national courts will agree with the EPO's interpretation.