III. Homology as an Indication of Confidence

A. Supporting Data for Homology Claims is Not Necessary for the Patent Law

A biological invention usually includes proteins or nucleic acids as its integral components. The building blocks of proteins and nucleic acids are residues of the small composite molecules - amino acids and nucleosides. Therefore, the combinational order of these residues, *i.e.* the sequence, is the precise description of a relevant protein or nucleic acid. Unlike mechanical inventions in which a structural element can be described as "a handle" or "a pad", this kind of language will always be taken as functional rather than structural in a biological invention, given the existence of a more basal description at the sub-molecular level.

The functionality of a biological sequence is subject to its combinational orders. For nucleic acids, these orders are recognised by transfer RNAs to determine the corresponding amino acids, or form hybrids, hairpins and loops to initiate or terminate certain biological processes.[72] For proteins and polypeptides, these orders are the very basis of their activity sites and three-dimensional structures. Any residue can be potentially critical to the function in question, though usually only a few are truly decisive.

Take two examples with relatively simple settings. To investigate the protein exporting mechanism of a bacterium through its Type Four Secretion System, Annette C. Vergunst *et al.* conducted serial mutations to 17 of the last 30 amino acid residues on the C-terminus of a bacterial protein VirF.[73] For their single mutations, four sites were found showing reduced exporting activity by 50%, the exact value of which was also subject to substituting residues. Double mutations based on these four further suppressed the activity to as close as 0%. It can be seen from this example that the function in question (exporting a fusion protein through a secre-

---

72 *e.g.* amiRNA; antisense RNA; guide RNA of the CRISPR-Cas9 system.

73 Annette C Vergunst and others, 'Positive Charge Is an Important Feature of the C-Terminal Transport Signal of the VirB/D4-Translocated Proteins of *Agrobacterium*' (2005) 102 Proceedings of the National Academy of Sciences 832 <http://www.pnas.org/cgi/doi/10.1073/pnas.0406241102> accessed 10 September 2017.

tion apparatus) remains relatively unchanged during most modification efforts. In some cases, a switch from one function to another needs only a tiny modification. For instance, Armin Djamei *et al.* demonstrated a single mutation to mimic constitutively phosphorylated status of the plant protein VIP1.[74] The phosphorylation status is decisive in this protein's subcellular localisation, which in turn affects its subsequent biological events. In a sense, this change could direct to very different technical effects in terms of patent law. A single mutation to mimic phosphorylation constitutively turns on a tunable function to A;[75] if this site is substituted by other residues, the phosphorylation may never occur, thus the function is directed to B permanently. These two examples serve as an appetiser of how sequence-function is correlated. It is this correlation that gives rise to the argument that a homology claim needs back-up by experimental data to show how sequence-function is precisely correlated for the patented invention.

In the author's view, the demand for supporting data is not well grounded. The purpose of the patent law is, as stated in Article 1 of the Patent Law, "to encourage invention-creation and promote the application of invention-creation".[76] From the Paris Convention[77] to the TRIPS Agreement[78], the patent law is always in the commercial context. Meanwhile, an invention is only recognised by the patent law from the technological perspective. An invention was positively defined in the Patent Law as "new technical solutions proposed for a product, a process or the improvement thereof".[79] Again, as stated in Article 7 of the TRIPS Agreement: "The protection and enforcement of intellectual property rights should contribute to the promotion of technological innovation and to the transfer and dissemination of technology". Therefore, as long as an invention carries out its objective technologically and is useful in an industrial sense, it

74 A Djamei and others, 'Trojan Horse Strategy in *Agrobacterium* Transformation: Abusing MAPK Defense Signaling' (2007) 318 Science 453 <http://www.sciencemag.org/cgi/doi/10.1126/science.1148110> accessed 10 September 2017.

75 "Constitutively" means that the said function of that protein becomes constant.

76 The Patent Law (n 27) Article 1, note: "invention-creation" is coined to include inventions, utility models and designs.

77 Paris Convention for the Protection of Industrial Property (1883, as amended on September 28, 1979).

78 Agreement on Trade-Related Aspects of Intellectual Property Rights (1994) (TRIPS Agreement)

79 The Patent Law (n 27) Article 2.

should be acknowledged by the patent law. Accordingly, the *quid pro quo* of the patent law should confer reasonable protection on this invention. Detailed understanding of why and how the invention works like that remains in the scientific realm, to which the patent law could contribute but not as the primary goal.

Scientific advance and technological progress are twins both favoured by a state's policy. They nevertheless bear distinct meanings, and should not be mistaken for each other. Take the famous quote as an analogy: "Humans lit fires for thousands of years before realising that oxygen is necessary to create and maintain a flame."[80] The control of fire has been considered to be a turning point in the cultural aspect of human evolution.[81] It resulted in significant expansion of human activity, and remarkably enhanced the survivability of humanity. None of these great aspects was compromised for not knowing the later-discovered "high-temperature exothermic redox chemical reaction".[82] It is precisely because of the importance of fire that scientists started to investigate the nature of fire. As a tool to promote the application of innovation, the patent law's purpose of encouraging the dissemination of technology will eventually create an eagerness for scientific knowledge underlying certain important inventions. At that stage, more efforts and resources will flood into the scientific investigation; and most probably an answer is revealed by quite a different person from a very distinct field. Imposing such a duty on the patentee or applicant would possibly impede the dissemination of the invention or deter future incentives to make an invention, which in turn harms the development of science. Thus, the patent law should not excuse itself from dealing with tough questions like the homology claims, by diverting the applicants to explore scientific discovery.

The patent in dispute has successfully identified one specific enzyme that is thermostable and has glucoamylase activity. The technical teaching of the invention is complete in terms of the patent law. Although more experiments can be performed to investigate the conserved motifs, domains and even tertiary structures of the enzyme, it is up to the patentee's choice

---

80 *EMI Group North America v. Cypress Semiconductor*, 268 F.3d 1342 (Fed. Cir. 2001).

81 David Price, 'Energy and Human Evolution' (1995) 16 Population and Environment 301.

82 Stephanie R. Dillon, 'The Chemistry of Combustion' <https://www.chem.fsu.edu/chemlab/chm1020c/Lecture 7/01.php> accessed 10 September 2017.

and should not be an obligation. The obligation belongs to the PRB and the courts to think twice about their position on homology claims. The demand for supporting data is nothing but a need to fulfil the courts' scientific curiosity.

B. Supporting Data for Homology Claims is an Overwhelming Burden

In the PRB's and the courts' decisions, it is frequently argued that in light of so many possible variants, those skilled in the art cannot reasonably predict which one works the invention. Admittedly, had *Novozymes* provided enough data in the patent application, they would not have gone through a hard battle over the validity of the claims. From the author's perspective, the provision of substantial data for a homology claim is only a choice of the patentee in theory, but not a doable job in reality.

To understand why the patentee would abandon such a chance to describe the invention more thoroughly, let us take a little test of mathematics. The patented enzyme has 591 residues in its sequence. At least 99% homology means that ≤5.91 residues can be substituted, which indicates at most five residues, at any position. The Beijing First Intermediate Court, however, stated that "six residues" is also an acceptable meaning of this homology claim, shown in their reasoning - "up to 5-6 residues can be changed".[83] This opinion was conceded by the Beijing High Court, without doubts. Under this setting, the calculation of combinations are as follows:

$$C(591,1) = \frac{591!}{1!(591-1)!} = 591 \quad (1)$$

$$C(591,2) = \frac{591!}{2!(591-2)!} = 174345 \quad (2)$$

$$C(591,3) = \frac{591!}{3!(591-3)!} = 34229735 \quad (3)$$

$$C(591,4) = \frac{591!}{4!(591-4)!} = 5031771045 \quad (4)$$

$$C(591,5) = \frac{591!}{5!(591-5)!} = 590729920683 \quad (5)$$

83 *Boli v PRB* (n 52); *Longda v PRB* (n 52).

$$C(591,6) = \frac{591!}{6!(591-6)!} = 57694622253373 \quad (6)$$

If the bottom line is set at five substitutes at most, the number of combination is the sum of equations (1) to (5); for six substitutes, the sum of (1) to (6). So, the number of combinations are **5.96×10**^{11} and **5.83×10**^{13}. Moreover, a substitute in any of the five or six sites can be any other amino acid except for the original one.[84] It makes an additional multiplication factor of 19, giving the final numbers as **1.12 × 10**^{13} for five and **1.10 × 10**^{15} for six. This calculation is based on "at least 99% homology" to an enzyme with "591 residues", which is a very narrow range and a normal-sized protein. The figure would be remarkably larger for a wider homology range and a larger protein.

We can see that a mere discretion of including an additional residue (5 → 6) into the interpretation of "at least 99% homology" enlarges the number of combinations by 100 times. Strictly speaking, ≤5.91 is not supposed to include the digit six, either in a legal context or in a technical context. However, the courts seemed not having been well informed before exercising their discretion. This demonstrated an insufficient understanding of the relevant technology when they made those reasonings, which renders their other seemingly sound reasonings questionable.

The SIPO Guidelines only require the patentee to exemplify the derived proteins or polypeptides in accordance wih the claimed homology.[85] But this relaxed standard has little chance to survive through the courts' current argument. In essence, the doubt about a homology claim lies in the large population of variants and the lack of knowledge of which one works. It is not something that can be easily fixed with several examples. It will be sarcastic if the PRB and the courts change their attitudes towards the same claim, just because several examples were provided. This is because in front of the astronomical number of variants, any quantity of examples effectively equals to zero in the proportion. Examples do not make any remedy to the argued problem. Therefore, provision of examples according to the SIPO Guidelines does not logically correspond to the PRB's

84 In practice, concerns about properties of the substituted residue will help to reduce the number of candidates.

85 SIPO, *Guidelines for Patent Examination* (2010 Ed.) (the SIPO Guidelines) 355-357. Note, page numbers accord with the English version, available at <http://www.sipo.gov.cn/zhfwpt/zlsqzn/sczn2010eng.pdf> accessed 10 September 2017.

and the courts' argument. Against such argument, exhaustive experimentation is still needed.

There are of course other approaches to facilitate reducing the workload. To identify motifs and domains, alignment of sequences of interest can help to preliminarily predict the conserved regions of the query sequence. However, during the validity proceedings, the sequence alignment data submitted by Novozymes was not well acknowledged. Subsequently, substitutions still need to be carried out to verify the conserved regions. It does not mean that the workload will be automatically lower if possible conserved regions are located. Exactly opposite to this, since the purpose is to support a homology claim, identification of conserved regions may render the patentee in an even awkward situation because when a region is believed to be conserved, the substitutability is thought to be largely limited by those skilled in the art. It means that by identifying one region as conserved, the patentee gives up the possibility that it is in reality not so. Persons skilled in the art will exclude those substitutions[86] made on the alleged "conserved region" from the teaching of the patent. To avoid losing the scope of protection, the patentee would still have to test the substitutions in an exhaustive manner.

The patentee may also resort to structural biology to solve the three-dimensional (3D) structure of the protein, which provides the closest linkage of the sequence-function correlation: sequence - 3D structure - function. That gives the patentee the perfect manner to describe the patented enzyme. However, the correlation of sequence-3D structure becomes another possible question. The courts were reluctant to recognise sequence alignment data as an assumption of conservation during this validity proceeding. There is unlikely to be any difference for a software-based prediction of the 3D structure. Therefore, in an infringement litigation, the patentee need to solve the 3D structure of the alleged infringing goods for the purpose of producing evidence of infringement. Is it a favourable solution? Possibly yes, if the structural biology service is fast and affordable, and if the courts recognise the evidence from computer-facilitated modelling. But the author has doubts over the status quo of its convenience for a litigation, in terms of cost, time, and accountability.

86 Here refers to non-conservative substitution, for conservative substitution see French and Robson (n 15).

In a nutshell, the requirement of supporting data to a homology claim is an overwhelmingly high burden to the patentee. Such requirement will greatly retard the disclosure of new inventions. Without fulfilling it, and most probably not being able to discharge it, a patent can only get a very narrow scope of protection based on its homology claim. This consequence is disproportionate to the technical contributions made by the patentee.[87] The relentlessness in emphasising the importance of experimental data in support of homology claims shows a lack of in-depth analysis of the relevant technology and its relation to the patent law. In light of the above, it is necessary for the PRB and the courts to rethink the meaning of homology language and review their positions on homology claims.

C. Rethinking the Role of Homology Language

1. The Homology Language

Broadly speaking, homology claims are not only limited to those with the words of "identity/similarity/homology". Molecule "hybridisation" and "substitution, deletion or addition" bear the same concept. Hybridisation usually relates to nucleic acids. It refers to the thermodynamic phenomena of two nucleic acid strands annealing together by hydrogen-bond formation between bases from each other strand. Hybridisation is a qualitative description, it can be predicted even without experimentation. Because of the thermodynamic nature and quantitative understanding of chemistry, the strength of hybrid (usually expressed as the Tm; melting temperature) is easily calculated under a given ionic strength environment. Yet, this description is more qualitative than quantitative, as it usually gives an answer of "yes" or "no" under low/moderate/high stringent conditions. Moreover, different numbers of hydrogen bonds formed between A/T and C/G[88] means that the impact of various substitutions may also differ ac-

87 Although a protein-related invention is disclosed with its sequence, the actual teaching is that one "category" of proteins perform the mentioned function. The principle of the function by such "category" of proteins remains the same, as understood by the persons skilled in the art. See William R Pearson, 'An Introduction to Sequence Similarity ("homology") Searching' [2013] Current Protocols in Bioinformatics.

88 Two hydrogen bonds formed between A/T, A=T; three between C/G, C≡G. See J Berg, J Tymoczko and L Stryer, *Biochemistry* (2007) p112.

cordingly. Simply put, changing A or T to others results in a minor alteration in the Tm compared with changing C or G. It makes a prejudice over C and G substitutions in the sequence, which may not have grounds in terms of its biological meanings.[89] "Substitution, deletion or addition" is an operational description, but can be combined with quantitative elements like "substitution, deletion or addition with one or several residues". This type of language has an advantage of delivering the homology concept of the relevant claim to a layperson, because of its description from an operational aspect. Comparing the different types of homology claims, "substitution, deletion or addition" and homology are better choices. They do not distinguish which residue is substituted, and they define the precise relative proximity to the reference sequence in a quantitative manner.

2. The Technical Meaning of Homology

A molecule is the smallest physical entity for a chemical compound that has the chemical properties of that compound. Insofar according to such understanding, a molecule might not be accepted by only partial description, since the partial description of a molecule is not conclusive of the final properties of this molecule. This could be the underlying principle that the PRB did not accept "comprising" or "contain" as a method to describe a protein, though the PRB argued this point based on numerous variants. The requirement of full sequence description brings about one problem that the patentee or applicant will be responsible for any of those "unimportant parts" in the claimed molecules. As argued by the PRB, the rest parts of a molecule may sabotage the claimed function in several ways.[90] At this point, the PRB started to question the enzymatic activity in the claims – the functional limitation. This idea followed in when the analysis continued to the homology claims. Thus, the sequences that are homologous to the disclosed one suffer from the same problem of unpredictability, with regards to their functionality.[91]

---

89 When the function of the nucleic acid in question is to form hybrids or hairpins, this prejudice is justified, as the function *per se* is hybridisation; but when the nucleic acid molecule carries further information, like mRNA or DNA, this prejudice has no justification.

90 PRB Decision No. 17956 (n 12) 15. See Section II.D.1.

91 PRB Decision No. 17956 (n 12) 16. See Section II.D.1. See also Pearson (n 87).

However, the challenge on non-functional sequences is not appropriate in the patent law practice. There are two possible ways to understand the mentioned function in the claims[92] of the '338 patent: 1) the function is part of the title of the subject matter; or 2) the function is used as a functional limitation. Whichever understanding is chosen for the interpretation, it must serve to limit the claims. According to O*n Several Issues concerning the Application of Law in the Trial of Patent Infringement Dispute Cases II*, both preamble portion and characterising portion have the limitation effect. [93] If this rule is followed, those homologous sequences that do not perform the mentioned function will fall outside the claims. For this reason, it is not necessary to make emphasis on the side of non-functional sequences when assessing the validity. Instead, the assessment should be focused on whether those functional ones are claimable.

When a particular sequence with any of its functions is disclosed, the persons skilled in the art do not merely believe that only this specific sequence performs the mentioned function. Rather, those skilled persons will understand that such function can be carried out by sequences similar to the disclosed one.[94]The only ambiguous thing in this understanding is how similar they should be. At this point, the homology range indicated in the claims serves as the basis of their confidence. It does not define absolutely whether any single homologous sequence within this homology range should work, but provides a cut-off value based on which a judgement is less likely to be wrong. The understanding of homology language as a matter of confidence can be demonstrated in the following examples.

In a research article, the authors state that:

> ***In view of the strong conservation*** *of the Skp1 proteins,* ***we predict*** *that* ***Skp1 proteins of other plant species*** *such as N. glauca* ***will interact with VirF*** *similarly to the way in which A. thaliana Skp1 homologs do.*[95]

92 See Section II.B: "an isolated enzyme with glucoamylase activity".

93 See note (n 68) Fa Shi [2016] 1: Article 5.

94 See Vineet Sangar and others, 'Quantitative Sequence-Function Relationships in Proteins Based on Gene Ontology' (2007) 8 BMC Bioinformatics 294 <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-8-294> accessed 10 September 2017.

95 Barbara Schrammeijer and others, 'Interaction of the Virulence Protein VirF of *Agrobacterium tumefaciens* with Plant Homologs of the Yeast Skp1 Protein' (2001) 11 Current Biology 258.

The "strong conservation" refers to a high homology of the said protein "Skp1" across "plant species" other than "*A. thaliana*". We can see from this statement that 1) the authors had no hesitation in nominating similar proteins from other species as Skp1; and 2) the author made clear and unambiguous statement that untested Skp1 proteins from other species can be predicted to function in the same way. Similar views not only exist in scientific publications, but also appear in patent law cases. In the European Patent Office (EPO) case **T0111/00**[96], the claimed cytokine[97] is of human origin. The closest prior art disclosed the sequence of a cytokine of mouse origin. The claim was held to be obvious as 1) there is an incentive to find the claimed cytokine based on the prior art cytokines; and 2) using the mouse cytokine cDNA as a probe to isolate human cytokine is straightforward.[98] A decision for lacking an inventive step demonstrates that the Technical Board of Appeal (TBA) believed that finding the claimed cytokine required no inventive efforts in light of the known cytokines. This belief of the TBA is based on homology between the claimed and the known cytokines. And this belief concurs with the author's opinion of homology as a confidence indicator. These two examples show that the concept of homology indicates a confidence level, which in turn projects the prospect of conducting further research on a particular sequence having such homology.

D. Species of Origin is Not an Effective Limitation

It is worth noting that according to the above two examples, the confidence level acquired from the knowledge of homology did not stop at the boundary of species classification, nor did it stop at the border of genus.[99]

---

96 T 0111/00 (n 66).

97 A category of small proteins that function in cell signalling. See Charles A. Dinarello, 'Historical Review of Cytokines' (2007) 37 European Journal of Immunology S34.

98 The straightforwardness is understood as such: in light of high homology, DNA hybridisation can be reasonably predicted. Therefore, the cDNA of the known protein, or part of it, can be used to probe unknown but suspected homologous DNA from other sources.

99 "Plant" is a kingdom-level description, six levels higher in taxonomic ranking than species; mouse and human converge in the class level – mammalia, four levels higher in taxonomic ranking than species.

Using species of origin as a limitation is thus not reflecting the essence of the technical facts. The Court supported such limitation for the reason that it effectively limited the number of naturally existing variants.[100] This can only be understood as an *ad hoc* solution in front of the patent in dispute, and should not mean that species of origin is a good and the only acceptable limitation.

Species by definition is the largest group of organisms in which two individuals are capable of producing fertile offsprings.[101] They do share some fundamental features, features that are functional to ensure reproduction. No underlying logic can be found to generally ensure proteins from the same species function in the same way, especially an industrially applicable one whose function is determined by human objectives. What if the claimed function is to cause blood precipitation when mixed with anti-A serum? The isoagglutinogen found in human body can work very differently: some people's isoagglutinogen can cause blood precipitation in anti-A serum, some cannot. This is the well-known case of blood cross-matching test for ABO blood group system.[102] Therefore, The Court was wrong in arguing, in its *Novozymes* decision, that within a particular species, there are only several sequences for a protein, and they share the same functions.[103] The only benefit in a species of origin limitation is the limited number of naturally existing variants. Therefore, the Court's acceptance of the species of origin is not well-grounded.

Moreover, when species of origin is used as a limitation, highly conserved sequences may encounter problems. When two sequences from different species are identical,[104] technical aspects of the two sequences will most probably be the same. As a consequence, this type of limitation may embarrass the courts in deciding whether it constitutes an infringement. If no infringement is found, it will technically make such patent non-en-

100 Though the large number of variants should not be a proper perspective in the validity discussion, the Court seemed not able to escape from this argument. This point will be further discussed in the following section.

101 See *Species* at Wikipedia <https://en.wikipedia.org/wiki/Species>.

102 See Fumiichiro Yamamoto, 'Review:ABO Blood Group system—ABH Oligosaccharide Antigens, Anti-A and Anti-B,A and B Glycosyltransferases, and ABO Genes' (2004) 20 Immunohematology 3.

103 *Novozymes* (n 4) 41. See Section II.D.3.

104 See Annex I (n 67). See also University of Missouri-Columbia, 'Identical DNA Codes Discovered in Different Plant Species' <https://www.sciencedaily.com/releases/2012/04/120409164426.htm> accessed 11 September 2017.

forceable; if infringement is found, probably through the doctrine of equivalents,[105] the limitation will prove itself only useful for patent validity not for infringement analysis.

In an infringement proceeding, attribution of an alleged infringing product to a certain species could be tricky. From the scientific perspective, molecular identification of a species usually relies on some specific elements in the organism, exemplified by 16S ribosomal RNA, due to the slow rates of evolution of this region of the gene.[106] It is not scientifically sound to identify a particular species based on an arbitrary sequence in a patent dispute. Even if this has to be done, the only ground to determine the origin of a given sequence is to make comparisons with the reference sequence of a particular species, and by assessing the confidence based on the homology value. From the above discussion, it becomes clear that homology is an inevitable consideration.

E. Concluding Remarks

In view of the discussions in this section, the support requirement for homology claims is neither necessary in terms of patent law nor doable in terms of technological and legal practice. A proper understanding should be given to the homology claims, which is not primarily meant to encompass a population of variants but to indicate a level of confidence on the sequence-function correlation. The use of species of origin as an alternative limitation appears to be a good way to discharge the support requirement, but is in fact only effective to stabilise the disputed patent given the improper arguments submitted from the lower courts. So far, this decision has no further value as a guide for future patent practice. Homology claims, thus, should be acknowledged by the patent law as an important method to describe a biological invention, and the support requirement should be reconsidered with a holistic analysis.

In the following sections, the author will analyse homology claims in the larger context of patentability requirements. The self-consistency of

---

105 For doctrine of equivalents, see note (n 68).

106 CR Woese and GE Fox, 'Phylogenetic Structure of the Prokaryotic Domains: The Primary Kingdoms' (1977) 74 Proceedings of the National Academy of Sciences, USA 5088 <http://www.pnas.org/cgi/doi/10.1073/pnas.74.11.5088> accessed 11 September 2017.

these requirements will be assessed against homology claims, and the predicament in applying Article 26.4 of the Patent Law[107] will be reviewed.

107 The Patent Law (n 27).