Intellectual Property Office

For Creativity and Innovation

BL O/121/08

28 April 2008

PATENTS ACT 1977

APPLICANT Council of Scientific & Industrial Research

ISSUE Whether application number GB 0521163.6 complies with section 1(1)(b)

HEARING OFFICER Dr L Cullen

Introduction

1 Patent application No. GB 0521163.6 ("the application") entitled "A method for storing information in DNA" was lodged on 18 October 2005. It is derived from the corresponding PCT application filed by the Council of Scientific and Industrial Research ("the applicant") on 31 March 2004 and published as WO2004/088585, which claimed priority from an earlier application filed in the USA on 31 March 2003. The application was republished on 18 January 2006 as GB2416242A.

2 In the first examination report under section 18(3), issued on 2 February 2006, the Examiner objected that the original claims lacked novelty and inventive step, and the Agent responded by filing amendments to the original claims. Further reports under section 18(3) were issued to which the agent responded and the novelty objection was successfully overcome. Some clarity objections were also successfully dealt with. However, the Examiner was not satisfied that the amendments filed overcame the inventive step objection raised in his reports and subsequently offered a hearing. This was the only outstanding issue to be resolved.

3 As the Applicant and Examiner failed to reach agreement, a hearing was arranged to resolve the matter, and the Examiner issued a letter on 5 December 2007 summarizing the issues to be addressed at the hearing, based on the claims filed with the Agent's letter of 16 April 2007. The Agent subsequently informed the Examiner in a letter on 7 December 2007 that the Applicant was content for the matter to be decided on the basis of the papers on file.

The Invention

4 The invention relates to a method for storing information in DNA, wherein a

unique sequence of four nucleotides is used so that each character of an extended ASCII character set (i.e. 256 characters) can be represented.

The Claims

5 As amended, claim 1 reads:

> "A method for storing information in DNA using a unique sequence of four DNA bases for representing each character of an extended ASCII character set comprising:
> (a) producing a synthetic DNA molecule comprising encrypted digital information that can be decoded with the use of an encryption key, flanked on each side by a primer sequence; and
> (b) storing the DNA molecule in a storage DNA, which consists of homogenous DNA, heterogeneous DNA or a mixture thereof.

6 A further independent claim, claim 16, is provided which reads:

> "A synthetic DNA molecule comprising encrypted digital information that can be decoded with the use of an encryption key, flanked on each side by a primer sequence, wherein the information is stored using a unique sequence of four DNA bases for representing each character of an extended ACSII character set"

The Law

7 The law regarding inventive step is found in sections 1 and 3 of the Patents Act. The relevant parts of those sections read as follows:

> 1.-(1) A patent may be granted only for an invention in respect of which the following conditions are satisfied, that is to say-
> (a) ….
> (b) it involves an inventive step;
> (c) …
>
> and references in this Act to a patentable invention shall be construed accordingly

> 3. An invention shall be taken to involve an inventive step if it is not obvious to a person skilled in the art, having regard to any matter which forms part of the state of the art by virtue only of section 2(2) above

8 As pointed out by the Examiner, the test for inventive step is found in *Windsurfing*[1], with the elaboration made by the Court of Appeal in *Pozzoli*[2]*.* The four-step test is outlined as follows:

---

[1] Windsurfing International Inc. v Tabur Marine (Great Britain) Ltd [1985] RPC 59

[2] Pozzoli SPA v BDMO SA [2007] EWCA Civ 588, [2007] FSR 37

Step 1(a) Identify the attributes of the person skilled in the art in this case

Step 1(b) Identify the common general knowledge of the person skilled in the art

Step 2 Identify what is the inventive concept of the patent

Step 3 Determine the differences, if any, between the state of the art and the inventive concept

Step 4 Are the differences identified between the invention as claimed and the state of the art obvious to the person skilled in the art or do they require a degree of invention?

The Arguments

9 As the Applicant chose not to attend the oral hearing, the sole arguments regarding inventive step are those present in the correspondence on file.

10 The arguments between the examiner and the agent in relation to the lack of inventive step focused on the disclosure in US 6,312,911 (Bancroft *et al*) and a further two documents, EP 0795862 (IBM) and Bharadwaj *et al*, “*DNA-based high-density memory devices and biomolecular electronics at CSIO*”, Proceedings of SPIE, SPIE, Bellingham, Vol. 4937, 2002, pp 319-325 (referred to as Bharadwaj).

11 The Agent argued that none of these prior art documents disclosed the representation of an extended ASCII character set using sequences of four bases. He also countered that, although Bharadwaj disclosed the use of four DNA bases to encode text, it did not suggest the encoding of the extended ASCII character set. He also argued that the disclosure of EP 0795862 related only to the field of segmentation of data when stored on tape devices.

12 The Examiner acknowledged that US 6,312,911 disclosed the encoding of alphanumeric characters using three DNA bases, and that Bharadwaj disclosed the encoding of numbers, texts or pictures. However, as Bharadwaj disclosed the concept of using more than three bases to represent data, the Examiner considered that it was not inventive to use four bases to represent the extended ASCII set. He submitted that the type and quantity of information being represented was merely an arbitrary choice and dependent upon the overall application of the system and how many encoding bases are available; the more bases available the larger the information set that can be represented. As the use of more than three bases was known, the Examiner considered that the use of four bases to represent the extended ASCII set would be obvious to a skilled worker.

13 The Examiner also considered the Agent’s arguments that the disclosure of EP 0795962 was in a different field, and so would not be considered by the skilled

worker. However, the examiner considered that the skilled worker would be a team consisting of persons who had knowledge of conventional DNA synthesis techniques and of persons who had knowledge of conventional techniques for the encoding, storage and decoding of data. Therefore he considered that the claims indicated did not involve an inventive step.

14 In response the Agent argued that it would not be obvious to the skilled worker to use four DNA bases to represent the extended ASCII character set simply because the use of more than three DNA bases, including the use of four bases, was known. He referred to an additional document (published after the priority date of the application in suit), Smith *et al*, "*Some possible codes for encrypting data DNA",* Biotech Lett, Vol 25, 2003, pp 1125-1130 (referred to as Smith), which proposed three DNA encryption methods and used up to 6 DNA bases per character but only addressed letters and some additional characters, a maximum of 80 characters in total. Consequently, he argued, using six bases could not represent all the 256 extended ASCII codes.

15 The arguments that US 6,312,911 was limited only to alphanumeric characters, and that EP 0795962 relates to a data splitting method meant for a different purpose and is not connected to DNA storage, were also reiterated and the Agent also added that Bharadwaj, which is the applicant's own prior art, does not "completely teach or guide towards either an encryption/coding or the complete encryption method/key".

16 The Examiner pursued his objections under inventive step in light of US 6,312,911 and EP 0795962. He also cited an additional document, McKenney *et al*, "*DNA molecules as standard reference materials I: Development of DNA identification sequences and human mitochondrial DNA reference sequences",* J Res Nat Inst Stand Technol, Vol 102, 1997, pp 53-62, (referred to as McKenney) found during an internet search carried out to confirm the common general knowledge. The McKenney article suggests a tetranucleotide code using all four DNA bases to provide 256 unique codons that can represent the entire ISO Latin-1 character set. He asserted that a person skilled in the art, having knowledge of McKenney and US 6,312,911 would find it obvious to produce a synthetic DNA sequence, such as that of US 6,312,911 to encode the extended ASCII character set using four bases instead of three.

17 In response, the Agent reiterated his arguments regarding the inventive step of the claimed invention over documents US 6,312,911 and EP 0795962 and also argued that McKenney merely teaches the use of the DNA sequence to form molecular tags in conjunction with DNA sequence determination, whereas the claimed invention uses DNA sequences to represent digital data using DNA sequences.

18 A further Examination Report was issued by the Examiner, wherein the four step test of *Windsurfing* was applied in order to assess the inventive step of the invention in light of US 6,312,911 and McKenney. The Examiner upheld his

inventive step objection. In response, the Agent repeated his argument that the teachings of US 6,312,911 are limited to three DNA bases representing alphanumeric characters, and cannot be used to represent most digital documents. He also argued that McKenney does not teach how four DNA bases may be used to encrypt the ISO Latin-1 character set, images, video etc and other kinds of digital information. The document merely asserted that all four DNA bases could be used to provide 256 unique codons, and the Agent argued that these assertions did not have a corresponding enabling disclosure, as the only teaching in the document was the use of triplet codons corresponding to the twenty letters of the amino acids specified by the genetic code.

19 The Examiner subsequently issued an Official Letter, updating the four step *Windsurfing* test to take account of *Pozzoli*, which considered the claimed invention in light of US 6,312,911, EP 0795962, McKenney, and Bharadwaj. The invention was still considered to lack inventive step and a hearing was set for the 12 December 2007.

Analysis

20 I will consider each step of the *Windsurfing/ Pozzoli* approach[1,2] in turn when assessing the obviousness of the present invention in light of the documents cited by the Examiner and taking due account of the case law and the facts of the case.

<u>Step 1a- The person skilled in the art</u>

21 The Examiner considered that the skilled person would be a team of skilled workers: workers skilled in the art of making synthetic DNA and workers skilled in the art of storing digital information. The Agent has disputed the relevance of the art of digital data storage. Given that the technological areas of molecular biology and digital information storage are so distinct, I agree with the Examiner and consider that the skilled person in this case would in fact be a team of workers.

22 In *Halliburton Energy Services Inc v. Smith International (North Sea) Ltd* [2006] R.P.C, 2, Pumfrey J (as he then was) made it clear (in paragraph 39) that the notional person skilled in the art may be in fact comprise a team of people bringing together different skills known in the art:

> “Before I consider the preferred embodiment, it is helpful to consider in rather more detail the qualities of the person through whose eyes the specification is to be read. I have already identified the relevant skills in general terms. The "notional skilled person" who is the addressee of the specification has been described in various ways for various purposes. The skilled person is essentially a legal construct, and not a mere lowest common denominator of all the persons engaged in the art at a particular time. In some cases, of which this is an example, it is clear that the specification is addressed to sets of skills that in the real world would be

possessed by more than one person, and such a specification can be said to be addressed to a team."

23 The Court of Appeal in *Genentech Incs. Patent* [1989] R.P.C. 147, also provides clarification on this point where Mustil LJ stated (at page 278):

> "The successful pursuit of Genentech's research required the deployment of techniques in more than one field: for example, protein sequencing, handling mRNA, building a library, making a probe. I am satisfied on the evidence that there was nobody who united in himself (or herself) all the knowledge and practical skills in each field to a sufficient extent to carry out any kindred project, even if assumed to be non-inventive, on his own. This fact has two corollaries, neither of which I understood to be in dispute. First, that the hypothetical person is a team of persons. Second, that since the search embraced a series of arts, the obviousness of any particular contribution to the ultimate success must be adjudged by reference individually to the hypothetical members of the team, attributing to each the appropriate degree of skill."

24 The team of persons skilled in the art in this case comprises a person skilled in the techniques and methods of molecular biology necessary to provide and manipulate synthetic DNA and a person skilled in the methods of digital information storage including encryption.

<u>Step 1b- The common general knowledge</u>

25 It is clear that at the priority date of the application, the use of DNA sequences to encrypt and/or store information was known. US 6,312,911 cited in the specification, discloses a stenographic method to encode concealed messages within DNA. An encryption key is disclosed using three nucleotides for each alphanumeric symbol. The synthetic DNA is flanked by known primer sequences allowing amplification from e.g. genomic or a heterogeneous mixture of DNA. Bharadwaj discusses the use of DNA for coding and decoding digital information, and discloses the ability to convert text, images or mathematical information into a sequence of nucleotides. The document also discloses the encoding of the flag of India into a DNA sequence using four nucleotides to represent each byte of the 1981 byte JPEG image (see page 322, 2nd paragraph). In addition, the document discloses the use of a four nucleotide code for encoding text (see the Table on page 322).

26 The segmentation of large amounts of data into smaller data blocks was also known, as is demonstrated by EP 0795962. Each data block is flanked by a header and a trailer, with each header containing a sequential identifying number. In addition, it was also well known in the art that the amplification of long pieces of DNA is less efficient and more error-prone than the amplification of shorter fragments. Therefore it was common to segment large pieces of DNA into smaller fragments for amplification purposes. In order to perform this

segmentation of DNA, primers flanking the desired fragments are synthesized, wherein the first primer effectively acts as a starter primer and the second primer effectively acts as an end primer. This is analogous to the flanking of the data block by a header and trailer in EP 0796962. The overlapping DNA sequences and/or primer sequences then allow for the sequential identification of the DNA fragments and assembly into the original large piece of DNA.

Step 2- The inventive concept of the patent

27 As summarized by the Agent in his letters, the inventive concept of the patent application is the use of four base pairs to encrypt an extended ASCII character set within synthetic DNA.

Step 3- Difference(s) over the prior art

28 The encryption method of US 6,312,911 uses three bases to represent alphanumeric characters. As the method is merely for the encryption of text, three nucleotides allows sufficient permutations for this purpose. This is similar to the disclosure of McKenney, where for simplicity they utilized the natural genetic code to encode marker sequences into DNA. This also suited the purpose of the method, which was to tag gene sequences to allow their identification, and therefore did not need to encode more letters than those of the naturally encoded amino acids. However, McKenney does state that if they wanted to encode more characters they could use different permutations of nucleotides, including a permutation involving all four nucleotides to produce 255 codons representing all 255 characters of the ISO-Latin 1 character set.

29 I now turn to Bharadwaj, which discusses the use of DNA for encoding digital information. In particular, I refer you to page 322 of the article. In the 2nd paragraph, it is stated that "*At CSIO we have defined the complete character set in terms of single stranded DNA bases of length four per character*…..*We have the unique capability to convert any text picture or mathematical operation into sequence of DNA bases*". The article does not define what the "character set" is, but any skilled worker would know that it could relate to any character set that can be represented by 4^4, i.e. 256, codons, such as the well known extended ASCII character set or the ISO-Latin 1 character set. This document discloses the encoding of the words "SPIE", "AUSTRALIA", and "MELBOURNE" using a four nucleotide sequence, the same four nucleotide sequence disclosed in the application. It also discloses the encoding of the flag of India, having a JPEG file size of 1981 byte using 7924 nucleotides (i.e. four nucleotides per byte). This again is using the same encryption method disclosed in the specification.

30 The Agent referred to Smith, which was published after the priority date of the application, and so cannot be considered as part of the state of the art. Nevertheless, I will address the arguments associated with this paper. Smith discusses three different methods of encoding information into DNA. The Agent argued that in this document, even by using a codon of six nucleotides it was

only possible to achieve a maximum of 80 codons. This, he argued, would mean that it would be incorrect to assume that it would be possible to use four nucleotides to represent the 256 character extended ASCII set. I disagree with this argument, as there are limitations placed on the structures of the codons described by Smith. The encryption method described whereby a maximum of 80 codons are produced, the "comma" method, does utilize a codon of 6 nucleotides. However, this is not representative of the number of codons available when using six nucleotides, as the first base of the codon is always a G, and the remaining five nucleotides are only made up of A, C and T. In addition, there are further restrictions on the sequence of the codons, as the five nucleotides must necessarily comprise three nucleotides selected from A or T and two C nucleotides. With these restrictions it is only possible to produce 80 codons. I therefore do not think that this disclosure (if it had been part of the state of the art at the time) would direct a person skilled in the art away from the possibility of using a four nucleotide codon to produce up to 256 characters.

Step 4- Was it obvious over the differences?

31 In my mind there are two aspects to the invention that need to be considered, (i) is it obvious to use four nucleotides in an encryption method? and (ii) is it obvious to use DNA to encode the extended ASCII character set?

32 US 6,312,911 and McKenney both disclose the use of a three nucleotide codon, the former method generating a possible 4^3 (64) codons (although only 40 were used), and the latter generating 20 as it only utilized the naturally occurring codons of the genetic code. Bharadwaj demonstrates that codons of up to seven nucleotides can be used, and so length is clearly not an issue. More significant is the disclosure on page 322 of Bharadwaj, where a codon of four nucleotides in length is described. Therefore, I have no doubt in concluding that at the priority date of the application a skilled man would have considered using codons of between three and seven nucleotides in encryption methods. It is also obvious that the longer the codon length, the longer the resulting DNA segments, and therefore I can assume that a skilled person would look at the shortest codon length that addresses his encryption needs.

33 When there are no restrictions placed upon how the four nucleotides can be combined to form a codon it is possible to have a total of 4^4, i.e. 256 codons. If the skilled person wanted to encrypt between 64 and 256 characters using DNA codons, he would appreciate that the use of four nucleotides would provide up to 256 codons. As this would be sufficient for his encryption needs he would not need to consider using a codon length of more than 4 nucleotides.

34 Turning to the second question, would it have been obvious to encode the extended ASCII character set using DNA? The work of McKenney was intended to provide a molecular tag sequence for labeling DNA sequences. However, this article discusses the possibilities of increasing the codon length to a tetranucleotide sequence to provide sufficient codons to specify the complete

ISO Latin-1 character set, which is an extension of the standard ASCII character set and comprises 255 characters. This research group, interested only in tagging DNA standards, clearly recognized the potential in using a codon comprising four nucleotides. If such a use is obvious to a group of molecular biologists not working in the area of data storage, then in my opinion, it would also clearly be obvious to a team of scientists skilled in the art of making synthetic DNA and skilled in the art of storing digital information

35 In addition, it is evident from Bharadwaj that the use of DNA to store digital information was known at the priority date of the application. This paper discusses the capability of DNA to store information, and on page 325 states that the software available has the ability to convert any digital information to a DNA sequence. Moreover, Bharadwaj also discloses the exact same four nucleotide codon encryption code disclosed in the application, and states that "*the complete character set in terms of single strand DNA of length four bases*" has been defined. It is clear from this document that the use of a codon comprising four DNA bases for storing information in DNA was known, and although this article does not explicitly state that the codons were used to represent the extended ASCII character set, given that the author of this paper is the same as the applicant for the present application, I can reasonably assume that this was the case. Even if I am incorrect in this assumption, the Examiner of this application, who is more familiar with the art, has informed me that the extended ASCII character set is a standard form of encoding for digital media known to anyone skilled in this art. This is reflected by the lack of description of the process of encoding in both the application and the Bharadwaj paper; I note that the application only discloses the end result, i.e., the encoding of the Indian flag, and not the method of converting it from its JPEG format to the nucleotide sequence of Example 3. In light of this, even without disclosure of what the "complete character set" of this document might be, I consider that a team skilled in the art of synthesising DNA and storing digital data would realise that this means any character set represented by 4^4 codons, i.e., the extended ASCII character set.

36 Consequently, I consider that US 6,312,911 demonstrates that alphanumeric characters can be encrypted into DNA using codon sequences of three nucleotides. McKenney and Bharadwaj further demonstrate that more characters can be encoded if the size of the codons is increased. In my opinion, a skilled team working in the area of DNA and data storage would appreciate that all 256 character of the extended ASCII set could be stored in DNA using a four nucleotide codon. Claims 1 and 16 therefore lack inventive step.

37 Dependent claims 6-8 and 10-12 are disclosed in US 6,312,911 and therefore these claims are also lacking an inventive step.

<u>EP 0795962</u>

38 The Examiner has cited EP 0795962 to demonstrate that a method of splitting up data using headers and trailers is well known in the art. The document is

concerned with digital information, and particularly with tape pre-formatting. I agree that this demonstrates that methods of storing digital data by splitting it into blocks is well known. I also consider that it is well known that this approach can be used to store data on other electronic media such as CDs and DVDs.

39 However, I think that the issue here also comes down to the nature of DNA itself. A skilled person in the field of molecular biology would be aware of the limitations of the amplification of large DNA sequences, mainly arising from the errors that are incorporated by the enzymes when assembling these large sequences. For this reason, it is standard practice when generating large DNA sequences to segment them, flanked by appropriate starter and terminal primer sequences, and finally stitching the segments together to produce the entire length of sequence required. The fact that this is also done with digital information adds more weight to the argument that the methods of segmenting the data claimed in dependent claims 2-5, 13-15 and 17-20 are obvious to a person skilled in the art.

Conclusion

40 From the above, I conclude that the invention described and claimed in current claims 1, 6-8, 10-12 and 16 of this application does not involve an inventive step in light of the teachings of US 6,312,911 when combined with either one of McKenney or Bharadwaj. I also consider that current dependent claims 2-5, 13-15 and 17-20 are obvious in light of common general knowledge in the field of molecular biology, and the teachings of EP 0795962.

41 As this was the situation when the period for putting the application in order expired on 30 September 2007, the application is refused under Section 18(3) as failing to meet the requirements of inventive step under Section 1(1)(b) of the Act.

Appeal

42 Under Practice Direction to Part 52 of the Civil Procedure Rules, any appeal must be lodged within 28 days.

L Cullen
Deputy Director acting for the Comptroller