Science and Justice 56 (2016) 278–281

Contents lists available at ScienceDirect

Science and Justice

journal homepage: www.elsevier.com/locate/scijus

Providing scientific guidance on DNA to the judiciary

Ian Evett [a], Susan Pope [a,*], Roberto Puch-Solis [b]

[a] Principal Forensic Services, Melbury House, 34 Southborough Road, Bickley, Bromley, Kent BR1 2EB, UK
[b] LGC, 3 Drayton Manor Business Park, Tamworth B78 3GL, UK

ARTICLE INFO

Article history:
Received 6 January 2016
Received in revised form 27 April 2016
Accepted 1 May 2016

ABSTRACT

A series of short documents have been written in response to a request from the UK Judiciary for explanations of research that was commissioned in response to questions they had raised. These related principally to the potential impact of primer binding site mutation (PBSM) but it became clear at an early stage that it was necessary to explain related issues. The three scientific guidance papers (SGPs) that have been prepared thus far are presented in their entirety so that UK scientists may be aware of what has been presented to judges.
Suggestions for further work, including possible communication to jurors are discussed.

1. Introduction

In July 2014, the UK National DNA Database (NDNAD) introduced a range of DNA multiplex chemistries for loading and searching DNA profiles. As a consequence of a briefing on the new systems, the Judiciary requested that an investigation be undertaken into the effect which primer binding site mutations (PBSMs) may be expected to have on the potential incidence of adventitious matches in those cases where samples had been analysed using different PCR chemistries. When such a mutation occurs at one of the loci, it has the effect that the associated allele is not visible in the profile. In the case where a person has two different alleles at a given locus (heterozygous) the effect of a PBSM would be that the profile would appear to be that of an individual with only one allele at that locus (homozygous). This means that the potential exists for an adventitious match as a result of a PBSM when, for example, a crime profile and person profile that have originated from two different individuals are found to be the same as a result of a PBSM in one of the profiles. The Judiciary were concerned that introducing a wider range of PCR chemistries would increase the chance of adventitious matches caused by PBSMs.

This type of adventitious match is different from the incident in 1999 in the UK when the SGM profile from an individual matched that of a crime stain, leading to his arrest. He was released when an upgrade to SGM Plus showed differences at the extra loci tested.

The system for addressing missed matches because of PBSMs is through searches of the NDNAD records for those that differ from the crime profile by a single allele. These are then investigated to see whether any mistyping has occurred, e.g. typographical error or misdesignation of alleles. Such a system does not flag up possible adventitious matches where the two profiles are the same because of a PBSM in one result.

The Judiciary also requested that short documents explaining the scientific background to the issues raised be provided for them. The Lord Chief Justice raised the concerns to the NDNAD management who commissioned an investigation into the potential for such adventitious matches. The work carried out by the authors of this paper who demonstrated, both by theory and using simulations, that the effect of PBSMs is slightly to decrease the adventitious match probability from what it would have been had the same DNA system been used. In order to meet the requirement for short documents explaining the scientific background three documents were produced and agreed with the Lord Chief Justice and his colleagues.

The background to the request is that DNA profiles are now produced and loaded to the NDNAD using several DNA17 PCR chemistries in addition to the existing profiles produced using SGM and SGM Plus chemistries. The primers used in each system vary. It is possible in all PCR chemistries that a PBSM can occur at any locus. The affected allele will not be visible in the profile. So for a locus where a person is heterozygous their profile will appear to be homozygous. If this profile is compared to another profile from the same person that has been analysed using the same PCR chemistry then the two profiles will both appear to be homozygous at the affected locus. However, if the two profiles have been produced using different PCR chemistries where the primers differ at the affected locus, then the two profiles will appear to be different, leading to a false discrimination.

However, it is also possible that the effect of the PBSM could be that two profiles that do in fact differ at a single allele at one locus, but have been analysed using differing PCR chemistries, may appear to match because of the effect of the PBSM. If, for example, a suspect is genotype (a, b) then a PBSM at that locus will lead to a homozygous profile – either (a, a) or (b, b); then, if the suspect is compared with a sample from a

* Corresponding author.
E-mail address: Suepope@principalforensicservices.com (S. Pope).

http://dx.doi.org/10.1016/j.scijus.2016.05.001
1355-0306/

different person who is the same homozygous genotype then there will be an adventitious match which wouldn't otherwise have occurred.

The potential for such matches was investigated [1] and it was shown that the effect of PBSMs was to decrease slightly the adventitious match probability.

It is considered important that scientists who prepare statements and attend court should be fully aware of the guidance that has been presented to the Judiciary on this and related issues. The three papers below are the versions discussed with and agreed by the Judiciary as meeting their needs for Court.

2. Scientific guidance paper 1: DNA match probabilities

The following table shows a DNA SGMPlus profile for a hypothetical single person, as it would be recorded on the National DNA Database (NDNAD).[1] The first row contains the names of 10 locations, or loci, on various chromosomes. At each locus, each person has two *alleles*, one inherited from each parent and these are shown in the second row of the table. The two alleles, taken together, are called a *genotype*. So we see, for example, that this particular donor is genotype (13, 14) at the D8S1179 locus: at this locus we say that the donor is *heterozygous* because the two alleles are different from each other. At locus D16S539, on the other hand, the donor has inherited two alleles of the same type – (9, 9) – and we say that the donor is *homozygous* at this locus.

vWA	TH01	D8S1179	FGA	D21S11	D18S51	D2S1338	D16S539	D19S433	D3S1358
14, 17	6, 9	13, 14	19, 22	29, 30	13, 18	18, 20	9, 9	15, 16	16, 16

At each locus there is a range of alleles to be found in the general population and this enables discrimination between the profiles from different individuals. If we take two unrelated[2] people then it is extremely unlikely that they will have the same ten-locus profile.

Let us imagine that the profile above is from a sample that was recovered from a crime scene, so the identity of the donor is not known. Further, let us imagine that a known person, Mr. X, is profiled, also by the SGMPlus system, and has the same profile as that in the table. So Mr. X will match the crime sample when his profile is searched against the database. If, as a result of this match and in combination with other findings, Mr. X is charged with the particular crime from which the crime sample had been recovered then a court of law will need to address propositions, representing prosecution and defence positions respectively:

2.1. The DNA in the crime sample was left by Mr. X

2.1.1. The DNA in the crime sample was left by some unknown person

It is widely understood that the problem of weighing these two propositions against each other is highly influenced by the probability that a match would have happened if the crime sample had been left by some unknown person. It is widespread practice to consider this unknown person as someone unrelated to Mr. X – leading to the question "what is the probability that some unknown person, unrelated to Mr X, would have the same profile as the crime sample?" This is known as a *match probability*.[3]

The calculation of a match probability is performed one locus at a time. Consider the D8S1179 locus where we find the two alleles 13 and 14 – we are interested in knowing how rare, or how common, each of these is in the population to which the "unknown person" belongs, which we will take to be Caucasian for the sake of discussion. The relative proportions of alleles are estimated from a database created from the profiling of samples from known individuals – typically, a few hundred individuals provide sufficient precision for this purpose.[4] At the D8S1179 locus, for example, we find that approximately 33% of alleles are type 13 and 20% are type 14.[5] These two allele proportions need to be combined in some sort of way to arrive at the probability of the genotype (13, 14) that is of interest in the case that we are considering. The method for doing this accords with a widely accepted approach that takes into account the consideration that human populations are not homogeneous but tend to be structured into subpopulation groups. The formula then embodies a factor, called theta (or Fst) that takes account of substructure – experiments have shown that typical theta values for Caucasian populations are of the order 0.001 but it is usual in casework to use substantially larger values for the sake of conservatism: current UK casework uses theta of 0.03 for Caucasians. If we combine the two allele proportions and theta in the appropriate formula we arrive at a match probability for the D8S1179 locus alone in this particular case of 0.143, or 1 in 7. Across the profile we find match probabilities of 1 in 15, 1 in 10, 1 in 7, 1 in 29 etc. for loci vWA, TH01, D8S1179, FGA etc. respectively. What we seek, of course, is a match probability for the entire profile and there is almost universal agreement, throughout the world, that this may be done by multiplying together the ten single locus match probabilities. In this particular example, this leads to a match probability of approximately one in three trillion.[6]

The very important question that arises is how should this evaluation be presented to a court of law? A question that is sometimes asked of scientists in court is "how can you give such a large figure, based on such a small database?" There are two points to make in reply. The first is that one in three trillion is not a "large figure" – it is an extremely small figure! The second is that the size of the database that was used for estimating allele proportions is of relatively minor importance – the critical factor is the multiplying together of all of the single locus match probabilities. This issue was considered very carefully when the SGMPlus system was introduced into casework by the Forensic Science Service (FSS) in the late 1990s. The statisticians who carried out the validation work at that time (2) considered that the extent of statistical studies into the robustness of the independence assumptions underlying that practice did not justify reporting match probabilities smaller than one in a billion[7] (3). Accordingly, a policy was adopted that in any case where there was a full ten-locus match between a person and crime sample, no case specific match probability would be calculated and a match probability of one in a billion would be reported, whatever the ethnicity of either defendant or offender [2].

This was intended to be an interim policy as it was expected that research databases would be produced of a size and cleanliness that would enable meaningful investigation of the robustness of probabilities such as one in a trillion. For various reasons this has not happened and the FSS "one in a billion" policy has remained in place. However, it should be pointed out that in other countries (notably the USA) such caution is not exercised and it is customary to see match probabilities of one in a quadrillion and even smaller quoted in casework, although no additional research has been carried out to support such figures.

[1] There is also a column that indicates the sex of the donor of the profile, but that is not discussed here because this information is not used in calculating match probabilities.

[2] People who are closely related are much more likely to share alleles – at the extreme, two identical twins would be expected to have the same ten-locus profile.

[3] It is necessary to emphasise that, whereas the match probability provides an adequate measure of evidential weight in simple cases, such as this, it is not adequate in more complicated situations, such as where the crime sample is a mixture. Indeed, there is a real danger that it may lead to an overstatement of evidential weight in such cases.

[4] Note that the NDNAD is not used for estimating allele proportions.

[5] Home Office data at www.gov.uk/government/statistics/dna-population-data-to-support-the-implementation-of-national-dna-database-dna-17-profiling

[6] A billion is a thousand million (nine zeroes) and a trillion is a million million (twelve zeroes). It is common in DNA reporting in the USA to continue this terminology into quadrillion, quintillion and so on.

[7] For example, to investigate independence assumptions associated with a probability of one in a trillion it is desirable to have a database of profiles from at least one million different and unrelated individuals of known provenance.