G89 Use of Proteins to Obtain Measure of Identity in the Absence of Usable DNA

Glendon Parker, PhD, Utah Valley University, Mailstop 179, 800 West University Parkway, Orem, UT 84103*

After attending this presentation, attendees will have a greater appreciation of the potential use of proteins in developing random match probabilities. Attendees will also have greater insight into additional uses of proteomics in forensics.

This presentation will impact the forensic science community by exploring how DNA can be wonderful if it is present in the sample. In the absence of usable DNA, protein can be a potential source of genetic information. This presentation focuses on extracting genetic information from a protein sample after DNA has been degraded or is contaminated. This information can be used to develop random match probabilities.

The use of nuclear DNA to provide matches between a forensic sample and an individual has revolutionized forensic science. DNA matches may have less than a 1 in 10^{13} probability of occurring randomly in the population. Unfortunately DNA is chemically and physiologically labile. In certain contexts, such as in hair shafts, degraded forensic samples, or in contaminated samples such as rape kits, amplification of DNA and subsequent DNA-typing is difficult or not possible due absence of usable DNA. Protein however can theoretically substitute for DNA: the sequence of amino acids is a record of the template DNA and incorporates in the primary amino acid sequence non-synonymous single-nucleotide polymorphisms (SNPs). There are 185,000 non-synonymous SNPs in the human genome (an average of nine amino acid changes per protein). 65,000 of these SNPS have a calculated allelic frequency based on genomic sequencing of bio-geographic populations. Many of these single amino acid polymorphisms (SAPs) can also be detected through tandem mass spectrometry. For this study it is hypothsized that SAPs will be detected in hair protein samples and that enough will be detected to provide a basis for calculating random match probabilities. To test this hypothesis three hairs from one individual were ground and digested with 20µg trypsin overnight. Three aliquots of 0.5% of the sample were then applied to a LC/MS/MS (UHR-qTOF) mass spectrometry instrument. A total of 18,154 peptides were detected, 1,694 of which were unique and not redundant. This corresponded to 285 proteins, 120 of which contained a total of 291 unique and non-redundant polymorphic peptides. In addition, data from the initial application was submitted to the Robust Accurate Identification (RAId) portal operated by the Yu group of the Computational Biology Branch at the National Institutes of Health (www.ncbi.nlm.nih.gov/CBBresearch/Yu). These algorithms are able to search for SAPs. A total of 361 polymorphic peptides were identified using this method. When the results from both searches were combined, polymorphic peptides with corresponding allelic dbSNP frequencies from the Utahn and Northern European populations were identified in 80 proteins. Using the product of all phenotypic frequencies the probability that one person would have that combination of markers was calculated at 5.5×10^{-5}, or one in 18,800. Using loci from 14 of these proteins the chance that one person would have that profile was calculated at 5.8×10^{-5} (or one in 17000). Both major and minor alleles were detected at 4 loci, two of which were identified using the RAId algorithm. These 14 loci provide a basis for developing random match probabilities using peptide information alone. Naturally use of these loci requires additional levels of scrutiny, such as comparison with synthetic peptide standards, confirmation of inferred SNPs using a SNP-chip, confirmation of the robustness of the analysis by repetition, and analysis with hair samples from additional individuals. Higher levels of discrimination are probable with the use of custom reference protein databases, which contain all known SAPs. The precise relationship between amount of hair and level of information, or sensitivity, also needs to be established. While not as genetically powerful as DNA-typing, protein-typing has the potential to significantly contribute to the probative information gained from protein forensic samples in the absence of usable DNA template. This methodology, developed for hair shaft protein, has the potential to be applied to all protein samples that are found in a forensic context.

Single Amino-Acid Polymorphism, Mass Spectrometry, Random Match Probability

Copyright 2012 by the AAFS. Unless stated otherwise, noncommercial *photocopying* of editorial published in this periodical is permitted by AAFS. Permission to reprint, publish, or otherwise reproduce such material in any form other than photocopying must be obtained by AAFS. ** Presenting Author*