

DEVELOPMENT OF AN EXPERT SYSTEM TO INTERPRET STR DNA ANALYSIS RESULTS*

D. Stansberry, J. D. Birdwell, T. W. Wang, J. Carter, R. Tipton, K. Gilbert, K. Budik, J. Pendelton
Laboratory for Information Technologies, The University of Tennessee, Knoxville, TN 37996-2100
L. Kienker, T. Moretti, B. Shea

FBI, Laboratory Division, FBI Academy, Building 12, Quantico, VA 22135

An expert system under development at the University of Tennessee will provide independent interpretation of forensic short tandem repeat (STR) DNA analytical results according to the guidelines of individual laboratories. Capillary run data, from supported capillary electrophoresis systems, will be processed by the expert system to provide fully independent analysis results. Further, the expert system will be entirely automatic. Without human intervention, project files from the data collection software will be batch-imported, analyzed, the STR profiles determined, and reports generated. In addition to designating alleles at the various loci, the expert system will provide identification of, and the rationale for, anomalous peaks, and evaluate the data relative to appropriate controls. A comprehensive event log will be generated by the expert system to support audits and legal disclosures. This event log will contain details of the expert system's analysis and decision-making processes, including the detection, classification, and/or rejection of peaks identified from the STR data.

The design of the expert system is a blend of object-oriented data analysis classes (C++) interfaced to a proven expert-system engine (CLIPS). Computationally expensive analysis algorithms are implemented in C++ for efficiency. The more subtle human-like reasoning algorithms are handled by the expert system engine. The primary algorithms developed and implemented in C++ include fluorescence signal processing, peak detection and classification (including the primer peak), internal size standard peak identification and development of size calibration equations, allelic ladder identification, anomalous peak detection and identification, and mixture identification and analysis. Expert rules have been developed for detecting signal anomalies (spikes and noise), pull-up, broad peaks, bleed-through, stutter, nontemplate nucleotide addition, and off-ladder alleles. Reference data for PCR amplification kits, control samples, and internal sizing standards are stored in a database and can be upgraded as new kits become available.

*The work at the University of Tennessee was supported by the U.S. Department of Justice, Federal Bureau of Investigation under contract J-FBI-03-196. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed, or implied, of the U.S. Government.