TECHNICAL NOTE

Margaret C. Kline,[1] M.S.; David L. Duewer,[2] Ph.D.; Pamela Newall,[3] M.A.; Janette W. Redman[1]; Dennis J. Reeder,[1] Ph.D.; and Melanie Richard,[3] M.Sc.

Interlaboratory Evaluation of Short Tandem Repeat Triplex CTT*

REFERENCE: Kline MC, Duewer DL, Newall P, Redman JW, Reeder DJ, Richard M. Interlaboratory evaluation of short tandem repeat triplex CTT. J Forensic Sci 1997;42(5):897–906.

ABSTRACT: An interlaboratory comparison of typing results for Short Tandem Repeats (STRs) at the GenBank loci HUMCSF1PO, HUMTPOX, HUMTH01, and HUMVWFA31 using the "CTT triplex" and "CTTv quadruplex" has been evaluated. These STRs all have a nominal four basepair (bp) repeat. Seven different samples were distributed to 41 laboratories. The 34 laboratories that returned results used a wide variety of analytical systems. Comparable results were obtained for all samples at all loci when results were reported as an allelic name. Raw sizing results obtained from internal-lane sizing standards differed by nearly five bp at some loci. Many different factors contribute to this observed sizing variability, including choice of sizing standards and matrix composition. Although sizing results can be made more comparable by locus-specific offsets or calibration to a comprehensive set of alleles at each locus, samples typed to the allelic name can now be validly compared regardless of analytical method. Interlaboratory comparison of raw allelic size remains problematic.

KEYWORDS: forensic science, allele sizing, CTTv CSF1PO, TPOX, TH01, vWA, DNA typing, gel electrophoresis, interlaboratory comparison, polymerase chain reaction

We here report the results of a designed interlaboratory comparison of DNA typing using multiplexed Short Tandem Repeat (STR) loci. The study was initiated at the Technical Working Group for DNA Analysis Methods (TWGDAM) 1995 Summer meeting.

[1]Biotechnology Division, Chemical Science and Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD.
[2]Analytical Chemistry Division, Chemical Science and Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD.
[3]Biology Section—DNA, The Center of Forensic Sciences, 25 Grosvenor Street, Toronto, Ontario M7A 2G8.
*Certain commercial equipment, instruments, or materials are identified in this report to specify adequately the results of the interlaboratory study and the experimental procedures used. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.
Preliminary reports have been provided to all participants of the interlaboratory study and have been presented at the 1996 Spring and Summer TWGDAM meetings and at the Seventh International Symposium on Human Identification, Sept. 1996.
This work was supported in part by the National Institute of Justice-U.S. Department of Justice.
Received 19 April 1996; and in revised form 27 Dec. 1997; accepted 3 Jan. 1997.

Seven different samples were prepared at the National Institute of Standards and Technology (NIST) and distributed to the 41 laboratories that had confirmed interest by Oct. 31, 1995. Most participants evaluated the "CTT" triplex, composed of STRs CSF1PO (GenBank locus HUMCSF1PO), TPOX (HUMTPOX), and TH01 (HUMTH01); the remaining participants evaluated the quadruplex "CTTv", which adds vWA (HUMVWFA31) to the CTT STRs. All four STRs are nominally four base pair (bp) repeats, although there is a common TH01 allele ("9.3") that is one bp smaller than the four bp repeat "10" allele.

Thirty-four laboratories reported a total of 46 datasets with some laboratories reporting results for more than one analytical method. Analyst experience ranged from novice to very experienced. Analytical methods included traditional static gel electrophoresis (post-electrophoresis analysis, with fragment size related to migration *distance* after a fixed time) and the newer dynamic methods (real-time electrophoretic analysis, with fragment size related to migration *time* at a fixed distance). A variety of amplification reagents and protocols, gel matrices, and electrophoretic conditions were used by the participating laboratories.

All results that were reported as nominal alleles ("calling to the allele", i.e., assigned an allelic name through reference to a known set of alleles) were in agreement. Results reported as fragment size differed by nearly five bp. Although these sizings were each successfully adjusted to provide correct allelic calls, they could not be directly compared without such adjustment. We attribute most of these differences to sizing protocol and gel matrix composition effects.

In the following sections of this report, we detail: (1) the design and implementation of the interlaboratory comparison, (2) range of methods used by the participants, (3) results of typing by calling the allele name, and (4) results of allele sizing and studies of the various factors that influence interlaboratory sizing variability.

Methods and Materials

In the summer of 1995, TWGDAM's subcommittee on Polymerase Chain Reaction (PCR) methods determined a critical need for evaluation of the "state of the (practical) art" for multiplex STR typing systems. At that time, CTT triplexes of CSF1PO, TPOX, and TH01 were well known, commercially available, and suitable for all STR detection methods. It was decided that a survey study of CTT using a limited number of samples and quantity of materials would provide the needed information if a suitably large number of laboratories would participate. Both manufacturers of commercial

Copyright © 1997 by ASTM International

CTT primer sets, PE-ABD (Perkin-Elmer Applied Biosystems Division, Foster City, CA) and Promega (Promega Corp., Madison, WI), agreed to supply gratis suitable materials and protocol recommendations to all interested parties. NIST coordinated all logistics.

An open invitation to participate in this survey study was made in October 1995 at the TWGDAM Open Meeting preceding the Sixth International Symposium on Human Identification. Because the CTTv quadruplex (CTT plus vWA) was then available, potential participants were requested to specify which multiplexes they wished to evaluate. In early November 1995, the TWGDAM PCR subcommittee supplied NIST with a list of 41 laboratories and the materials they wished to analyze. Table 1 lists the 34 laboratories that returned results by mid-May 1996 and the typing systems that they used.

Because it was necessary to minimize the resource demands on each participant and on the commercial suppliers (and on the NIST coordinators), it was decided to limit the DNA samples to the number that would conveniently fit on one analytical gel. Two cell line samples (supplied gratis by Life Technologies, Inc., Gaithersburg, MD) were included to provide traceability to NIST SRM® 2391 PCR-based DNA Profiling Standard (1). Five unrelated human samples were provided to include some diversity. Fourteen participants received an eighth sample on a novel PCR DNA isolation device (supplied gratis by Schleicher & Schuell, Inc., Keene NH). All samples are described in Table 2.

Study materials were shipped to all laboratories by mid-December 1995. In addition to primer kits and DNA samples, these shipments included: general instructions, information packages for all kits provided, and a results form. This standard form provided for uniform reporting of allelic name and sizing results and for various methodological details, including: primer type, thermocycler model, gel composition, electrophoretic conditions and equipment, detection method and instrumentation, sample denaturation procedure, and amplification method. Participants were also requested to provide high-quality copies of all gel images or electropherograms. Table 3 summarizes the conditions and equipment reported for static electrophoretic methods; Table 4 summarizes the dynamic methods. The "short hand" names for specific equipment and supplies are defined in these tables. All results were collected and evaluated at NIST.

Results and Discussion

Calling to the Allele

All laboratories that "called" allele names, regardless of whether they used static or dynamic analysis methods, reported the same

TABLE 1—*Participants in CTT interlaboratory comparison.*

Laboratory	Analysis*	
	Static	Dynamic
Alabama Department of Forensic Sciences, Birmingham Regional Laboratory	P	
Applied Biosystems Division, Perkin-Elmer Co., Foster City, CA		A
Armed Forces DNA Identity Laboratory, Rockville, MD		A, Pv
California Department of Justice, DNA Berkeley Laboratory	Pv	A
CBR Laboratory Inc., Boston, MA	P	
Cellmark Diagnostics, Germantown, MD	P	
Centre of Forensic Sciences, Biology Section, Toronto, Ontario		Pv
Connecticut State Police Forensic Science Laboratory, Meriden	P	
DNA Laboratory, Bern, Switzerland	P	
Forensic Science Çentre, Adelaide, South Australia		A
Georgia Bureau of Investigation, Department of Forensic Science, Decatur	P	
Illinois State Police, Division of Forensic Services and Identification, Springfield	P	
Kentucky State Police, Central Forensic Laboratory, Frankfort	P	
Lab Corp Biomedical, Research Triangle Park, NC	DNA	
LI-COR Inc., Biotechnoloy Division, Lincoln, NE		P
Michigan State Police, DNA Laboratory, East Lansing	P	
Minnesota Forensic Science Laboratory, St. Paul	P	
National Institute of Standards and Technology, Gaithersburg, MD	P, Pv	A, Pv
New Mexico Department of Public Safety, Santa Fe	P	
North Carolina State Bureau of Investigation, Raleigh	P, Pv	
North Louisiana Criminalistics Laboratory, Shreveport	P	
Office of Chief Medical Examiner, New York City		DNA
Orange County Sheriff-Coroner Department, Santa Ana, CA	P	
Orlando Regional Crime Laboratory, FL		A, Pv
Palm Beach County Sheriff's Office Crime Laboratory, FL	P	
Promega Corporation, Madison, WI	Pv	A, Pv
Royal Canadian Mounted Police, Central Forensic Laboratory, Ottawa, Ontario		A, Pv
Servicio de Huellas Digitales Genéticas, Buenos Aires, Argentina	P	
Suffolk County Crime Laboratory, Hauppauge, NY		A, Pv
University of Nebraska Medical Center, Omaha	P	
University of North Texas Health Science Center at Fort Worth	Pv	
Victoria Forensic Crime Laboratory, Melbourne, Australia	P	
Virginia Division of Forensic Science, Central Laboratory, Richmond	P	
Washoe County Sheriff's Office, Reno, NV	P	

*Laboratories reported results for the following kits:
A–CTT (PE-ABD, Foster City, CA).
P–CTT (Promega Corp., Madison, WI).
Pv–CTTv (Promega Corp., Madison, WI).
DNA–samples only.

TABLE 2—*Samples distributed in CTT interlaboratory comparison exercise.*

Samples	Type	Source	Quantity, Sample Medium*
1	Genomic DNA	Cell line GM09947A	10 ng, 10 μL TE buffer
2	Genomic DNA	Cell line GM09948	10 ng, 10 μL TE buffer
3	Blood stain	A	20 μL whole blood, SS903
4	Blood stain	B	20 μL whole blood, SS903
5	Blood stain	C	20 μL whole blood, SS903
6	Genomic DNA	D	10 ng, 10 μL TE buffer
7	Genomic DNA	E	10 ng, 10 μL TE buffer
8	Blood stain	A	20 μL whole blood, IsoCode

*TE: 10 mmol/L Tris Cl, 0.1 mmol/L EDTA, pH 8.0.
SS903: Schleicher & Schuell 903 Paper (Schleicher & Schuell, Inc., Keene, NH).
IsoCode: PCR DNA sample isolation device (Schleicher & Schuell, Inc.).

allele names for all samples. Laboratories using analytical protocols unable to resolve all neighboring alleles reported two-allele bins that always included the consensus allele. Figure 1 presents a schematic representation of a static gel of the CTT system, showing loading patterns, allele sizes, and allelic ladders. Table 5 lists the reported alleles and binnings for all samples.

Sample 2 was genomic DNA from a cell line having three distinct alleles at locus CSF1PO: allele names 10, 11, and 12. Image intensities for these alleles are not the same, no matter what the imaging system, with allele 12 providing the weakest signal. Twenty seven of the 36 datasets reported all three alleles; the remaining results either noted the presence of a light band without calling the allele or were from laboratories that binned some neighboring alleles.

Samples 3 and 8 were blood stains from the same source having a type of 14,18 at locus vWA. Several participants reported the presence of a minor (much less intense) 15 allele. This apparently real phenomenon is under further investigation.

Although there were no errors in allelic calls attributable to measurement processes, there were some reporting anomalies: (1) One laboratory reported the presence of a minor contaminant in one cell line sample. (2) One laboratory mislabeled datasets for samples 3 and 5. (3) One laboratory reported sizing consistent with allele 6 of locus TH01 but miss-recorded the call as allele

TABLE 3—*Conditions and equipment used with static analysis.*

Detection*	Therm†	Gel‡	Buffer§	Urea‖	N¶
MD	9600	M: 6% Long Ranger	0.6 × TBE	7.0	1
Hitachi & MD	480	P: 4.0/5.0	0.5 × TBE	8.0	1
Hitachi & MD	480	P: 6.0/2.6	0.5 × TBE	7.0	1
MD	9600	P: 6.0/5.0	1.0 × TBE	7.0	1
MD	9600	P: 7.0/3.0	TS − TB	7.0	1
MD	9600	P: 8.0/2.0	TF − TB	7.0	1
MD	9600	P: 9.0/3.3	TS − TBE	0	1
^{32}P	9600	P: 19/1.0	1.0 × TBE	7.0	1
Ag	480	M: 5% Long Ranger	0.5 × TBE	7.0	1
Ag	9600	M: 1X GDG	0.5 × TBE	8.0	1
Ag	MJR	P: 19/1.0	1.0 × TBE	7.0	1
Ag	480	P: 4.0/5.0	0.5 × TBE	7.0	4
Ag	9600	P: 4.0/5.0	0.5 × TBE	7.0	4
Ag	2400	P: 6.0/5.0	0.5 × TBE	7.0	1
Ag	9600	P: 6.0/5.0	0.5 × TBE	7.0	1
Ag	480	P: 6.0/5.0	1.0 × TBE	7.0	3
Ag	9600	P: 6.0/5.0	1.0 × TBE	8.0	1

*Detection: Hitachi–FMBIO™ Fluorescent imaging device, (Hitachi Sortware Engineering America, Ltd., San Bruno, CA).
MD–Molecular Dynamics FluorImager™ SI, (Molecular Dynamics Corporation, Sunnyvale, CA).
^{32}P–radioisotopic label.
Ag–silver stain.
†Thermocycler: 480–DNA Thermocycler 480 (Perkin-Elmer, Norwalk, CT).
2400–GeneAmp® PCR System 2400 (Perkin-Elmer).
9600–GeneAmp PCR System 9600 (Perkin-Elmer).
MJR–MJR MiniCycler™ (MJ Research, Inc., Watertown, MA).
‡"M" indicates a given composition of a proprietary monomeric acrylamide gel.
Long Ranger™ (FMC BioProducts, Rockland, ME).
GDG–GeneAmp Detection Gel (Perkin-Elmer, Norwalk, CT).
"P: x/y" indicates a polyacrylamide gel of composition: x% total acrylamide, y% of the total as bis-acrylamide.
§1.0 × TBE - 89 mmol/L Tris, 89 mmol/L borate, 2 mmol/L EDTA.
TS-TB − (375 mmol/L Tris, 39 mmol/L sulfate) − (104 mmol/L Tris, 28 mmol/L borate).
TF-TB − (375 mmol/L Tris, 30 mmol/L formate) − (104 mmol/L Tris, 28 mmol/L borate).
TS-TBE − (proprietary Tris sulfate) − (1.0 × TBE).
‖Urea concentration in mol/L.
¶Number of submitted datasets collected under stated conditions.

TABLE 4—*Conditions and equipment used with dynamic analysis.*

Ladder*	Detection†	Therm‡	Gel§	Buffer‖	Urea¶	N**
GS500	ABD373	480	P: 6.0/5.0	1.0 × TBE	8.3	2
GS500	ABD373	9600	P: 6.0/5.0	1.0 × TBE	8.3	1
GS2500	ABD373	480	P: 6.0/5.0	1.0 × TBE	8.3	2
GS2500	ABD373	9600	P: 6.0/5.0	1.0 × TBE	8.3	2
GS500	ABD373	9600	P: 6.0/5.0	1.0 × TBE	7.0	1
GS350	ABD377	9600	P: 4.0/5.0	1.0 × TBE	6.0	2
GS500	ABD377	480	P: 4.0/5.0	1.0 × TBE	6.0	4
GS500	ABD377	9600	P: 4.0/5.0	1.0 × TBE	6.0	7
Allelic	LI-COR	MJR	M: 6% Long Ranger	1.0 × TBE	7.0	1

*Sizing ladder: Allelic–external-lane allelic (Promega Corp.).
GS350–internal-lane GENESCAN 350 (PE-ABD).
GS500–internal-lane GENESCAN 500 (PE-ABD).
GS2500–internal-lane GENESCAN 2500 (PE-ABD).
†Detection: ABD373–ABD373 DNA Sequencer (PE-ABD).
ABD377–ABD377 DNA Sequencer (PE-ABD).
LI-COR–LI-COR 4000 infrared scanner (LI-COR, Inc.).
‡Thermocycler: 480–DNA Thermocycler 480 (Perkin-Elmer).
9600–GeneAmp PCR System 9600 (Perkin-Elmer).
MJR–MJR MiniCycler™ (MJ Research, Inc.).
§"M" indicates the given proprietary monomeric acrylamide gel.
"P: x/y" indicates a polyacrylamide gel of composition: x% total acrylamide, y% of the total as bis-acrylamide.
‖1.0 × TBE (89 mmol/L Tris, 89 mmol/L borate, 2 mmol/L EDTA).
¶Urea concentration in mol/L.
**Number of submitted datasets collected under stated conditions.

9. (4) All but one of the samples distributed in this study had a locus TH01 9.3 allele; none had a TH01 10 allele. Nearly all laboratories called the 9.3 alleles without binning. Six laboratories binned some or all of the 9.3 alleles; however, three of these laboratories termed the bin "9.3/10" whereas the other three termed it "10" with a noted definition that all TH01 9.3 and 10 alleles are termed "10". (5) One laboratory chose to report only sizes for all alleles; however, these results were consistent with the average size from other laboratories who successfully called the nominal alleles. Thus, there were incompatible results traceable to sample contamination, sample and data handling errors, and different reporting protocols.

Allele Sizes

Although all laboratories that called nominal alleles based on sizing data did call them successfully, the allele sizings are not directly comparable as reported. Figure 2 presents all reported sizings for all samples for the four STR systems. Ideally, all data for each sample would cluster tightly and symmetrically about the nominal size of the alleles.[4] This is approximately true for locus TH01; however, the reported CSF1PO interlaboratory data vary by more than four bp.

The data at all loci are spread along lines of constant slope. Such strong correlation between sets of measurements on a given sample is characteristic of protocol-specific bias (2). Figure 3 shows the effect of removing much of this bias by adjusting the raw values with a locus- and dataset-specific offset. We approximate these offsets as the average of the average differences between the reported and nominal size of all sizings of a given allele:

$$O_{lk} = \sum_{j=1}^{n_{lk}} \left(A_{kj} - \sum_{i=1}^{n_{lkj}} x_{lkji} \Big/ n_{lkj} \right) \Big/ n_{lk}$$

[4]In this report, the "nominal allele size" is the bp size provided in Ref 6. These were the values used by all laboratories that sized using an external-lane allelic ladder. See below for further discussion.

where A_{kj} is the nominal size of allele j at the given locus, n_{lkj} is the number of sizings of the given allele in the given dataset, n_{lk} is the number of different alleles reported for the given locus, and x_{lkji} is the bp size for sample i having the given allele reported in the given dataset.

This offset correction dramatically decreases the systematic scatter in the data, but does not completely remove it. Table 6 lists the pooled sizing standard deviations (SD) for the data as reported and after offset correction, where the pooling is over all alleles of each locus:

$$SD_k = \sqrt{\sum_{j=1}^{n_{kj}} (n_{kj} - 1) SD_{kj}^2 \Big/ \sum_{j=1}^{n_{kj}} (n_{kj} - 1)}$$

where n_{kj} is the number of samples with allele j at locus k and SD_{kj} is the simple SD calculated over all reported sizings of allele j at locus k. For locus CSF1PO, the offset correction reduces the approximate 99% confidence "match window" of ±3 SD from more than five bp to less than two bp. There is at least one anomalous datum at each of the CTT loci. These values each originated in different laboratories and each differ by at least one bp from other sizings reported for the same allele by that laboratory. The 0.15 bp to 0.30 bp SD for the offset-corrected data is about that previously reported for intergel precision (3,4).

Table 7 lists all the calculated offsets, along with the known protocol factors, for all datasets. Because these offsets are calculated from the same single-gel data to which they are applied, they may not be characteristic of multiple gels run under the same nominal conditions. The virtual identity of the duplicate datasets submitted by laboratory F suggests that the offsets may be characteristic of a particular protocol, at least in the short-term. Long-term (weeks to years) studies of control and evaluation materials are required to access fully the constancy of such offset corrections for a given laboratory.

Cursory examination of Table 7 reveals that: (1) all datasets

TABLE 5—*Nominal alleles.*

Sample	CSF1PO		TPOX		TH01		vWA
	Alleles	Bins*	Alleles	Bins*	Alleles	Bins*	Alleles
1	10,12	9/10, 11/12	8,8	7/8	8,9.3	9.3/10	17,18
2	10,11,12	9/10, 10/11, 11/12	8,9		6,9.3	9.3/10	17,17
3	13,13	12/13, 13/14	8,9		7,9.3	9.3/10	14,18†
4	12,12	11/12	8,8		9,9.3	9.3/10	15,18
5	12,14	11/12, 13/14,	8,11		6,9.3	9.3/10	16,17
6	11,13	10/11, 11/12, 12/13, 13/14	9,10		7,9.3	9.3/10	17,19
7	11,12	10/11, 11/12	8,8		6,7		17,18
8	13,13		8,9		7,9.3	9.3/10	14,18

*Bins as specified by individual participants.
†Several participants reported a faint allele 15 band for this sample.

sized using allelic ladders have uniformly very small offsets, (2) there are one to two bp differences between offsets traceable to different sizing ladders, and (3) different factors have different influence at each of the four loci. Analysis of variance applied to these offsets (results not shown) does identify "sizing ladder" as the dominant factor. However, the data are not sufficiently balanced to evaluate adequately the relative importance and behavior of the other factors and their interactions.

Directly Evaluated Sizing Variables

It is possible to characterize many of the potential sizing factors through a combination of exploratory analysis and directed experiments. Four laboratories evaluated both the PE-ABD and the Promega primer sets using otherwise identical analytical methods.

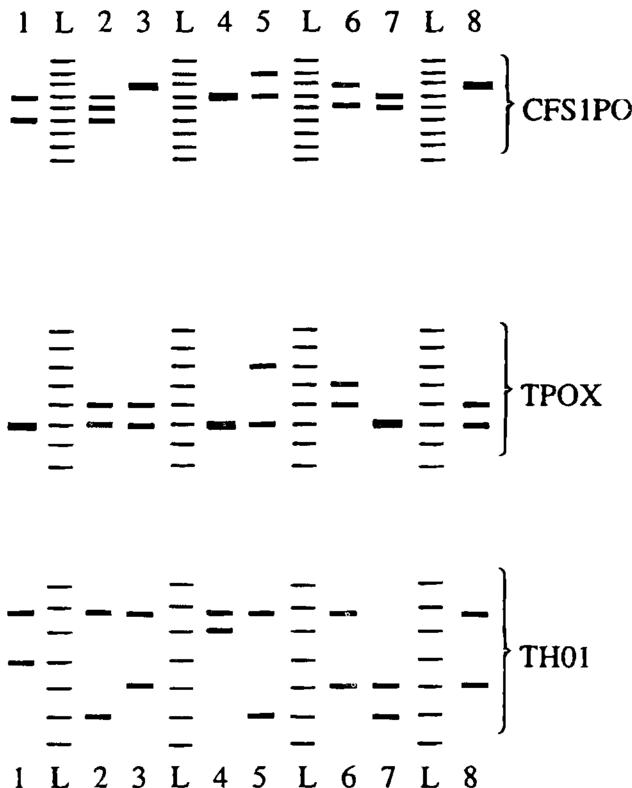

FIG. 1—*Schematic representation of a CTT static gel. Allelic ladders are shown in lanes denoted "L"; samples are in lanes marked with the sample numbers. Band position, width, and intensity mimic those observed.*

Two laboratories similarly evaluated the Hitachi and Molecular Dynamics fluorescence-based image analysis instruments. Analysis of the differences between paired datasets enables direct multi-laboratory evaluation of these factors:

$$\Delta_{kj} = \sum_{l=1}^{n_{lkj}} \left(\sum_{i=1}^{n_{kj}} (x_{lkji} - y_{lkji})/n_{kj} \right) \Big/ n_{lkj}$$

where n_{lkj} is the number of participants that reported sizings for allele j of locus k using the two given sets of paired conditions and x_{lkji} and y_{lkji} are the paired sizings for sample i having this allele.

Primer—Figure 4 shows the lack of appreciable difference between sizings of PCR products from the PE-ABD or Promega primer sets. Because the published allele product sizes of the two primer sets differ by one to two bp depending on locus, this is an unanticipated result (5,6). The contrasts shown in Fig. 4 are calculated from data supplied by four laboratories that evaluated both primer sets by the same analytical method. Contrasts between the primer sets based on all datasets matched by analytical method regardless of laboratory origin also cluster about zero, with only a modest increase in variability.

Image Analysis—Figure 5 shows that there is little difference between sizings provided by the two fluorescence-based static-gel image analysis instruments.

Sizing Variables Confirmed by Additional Experimentation—The choice of sizing ladder and matrix composition are known to affect sizing (7–9). These factors can be characterized in the current study by combining data from laboratories that use similar analytical methods. It is convenient to express the combined results as the difference from the nominal allele size

$$\Delta_{kj} = A_{kj} - \sum_{l=1}^{n_{lkj}} \left(\sum_{i=1}^{n_{kj}} x_{lkji}/n_{kj} \right) \Big/ n_{lkj}$$

where n_{lkj} is the number of participants that reported sizings for this allele using a given set of conditions. The influences of the following factors have been confirmed by directed post-study experimentation within our laboratories.

Sizing Standards—Figure 6a presents average results for datasets grouped by sizing standards, including external-lane allelic ladders

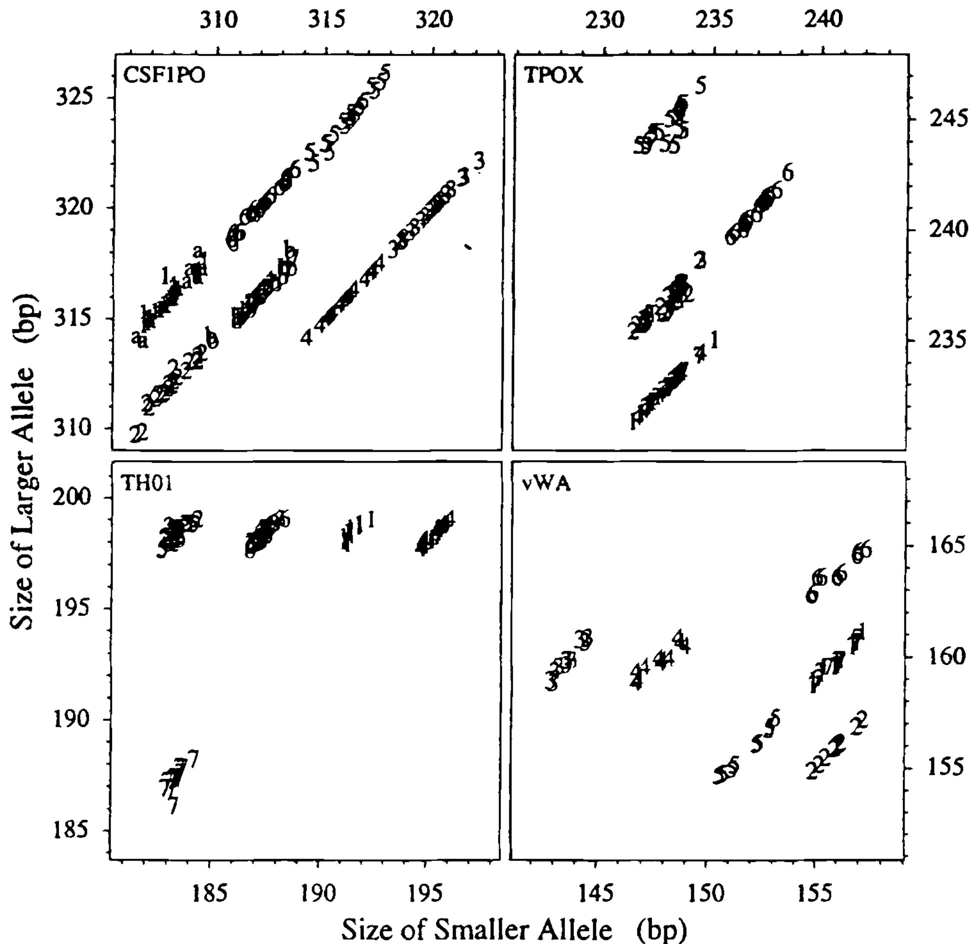

FIG. 2—*Allele sizes of reported data. All data for each loci are presented as scatterplots, with the size of the smaller allele of a given locus plotted against that of the larger allele. The data are labeled with the sample number. For the three-banded pattern of sample 2 at CSF1PO, "2" denotes the pairing of the (smallest, middle) alleles, "a" denotes (smallest, largest), and "b" denotes (middle, largest). All axes of all four scattergrams have the same range of 18 bp.*

and internal-lane sizing standards. Because the components of the GS350 sizing standard are known to be a subset of the GS500 sizing standard, results from the one laboratory that used GS350 have been combined with those using GS500 ("GS350/500").

All results from systems that used allelic ladders for sizing are, as expected from the offsets, very close to the nominal allele size. There are large differences between the combined GS350/500 and the GS2500 sizing standards for at least some of the alleles of all loci.

Figure 6b presents results from sizing an allelic ladder with three different internal-lane sizing standards within one gel: GS350, GS2500, and the experimental BRL25 (Life Technologies, Gaithersburg, MD). Although the absolute magnitude of the differences between the GS350 and GS2500 ladders are somewhat different from those displayed in Fig. 6a, the relative magnitudes among the different loci are quite similar.

Note that the differences between the measured and nominal size of the alleles at a given locus are not constant. There are apparently non-linear and sizing-standard-dependent changes in the offsets within each locus related to nominal allele size. At locus CSF1PO, the offsets for the largest and the smallest allele can differ by more than two bp. Calibration of internal-lane sizing standards to a comprehensive set of alleles at each locus could more fully compensate for interlaboratory sizing differences than does adjustment by a locus-specific constant such as that displayed in Fig. 3.

Gel Matrix/Sequencer—All but one of the datasets sized using the GS350 or GS500 sizing standards used either and ABD 373 DNA Sequencer (PE Applied Biosystems) with 8.3 mol/L urea in the polyacrylamide gel or an ABD 377 DNA Sequencer (PE Applied Biosystems) with 6.0 mol/L urea. One dataset (laboratory G) used an ABD 373 with 7.0 mol/L urea. Figure 7a presents average results for these datasets grouped by urea concentration after subtracting the GS350/500 sizing standard effect discussed above.

The correlation between sequencer and urea concentration limits the utility of these data for isolating instrument-specific and gel matrix-specific sizing effects. However, the unique behavior of the one dataset using 7.0 mol/L urea (laboratory G) suggests that matrix effects could be surprisingly locus-specific.

On request, laboratory G reevaluated all samples using an 8.3 mol/L urea gel matrix but otherwise using the same analytical protocol. Figure 7b contrasts the results of laboratory G using two urea concentrations, thus confirming the urea effect is extremely locus-specific and may be allele-specific.

The two automated sequencers differ in their temperature control capabilities; gel matrices used differ in thickness, length, and total

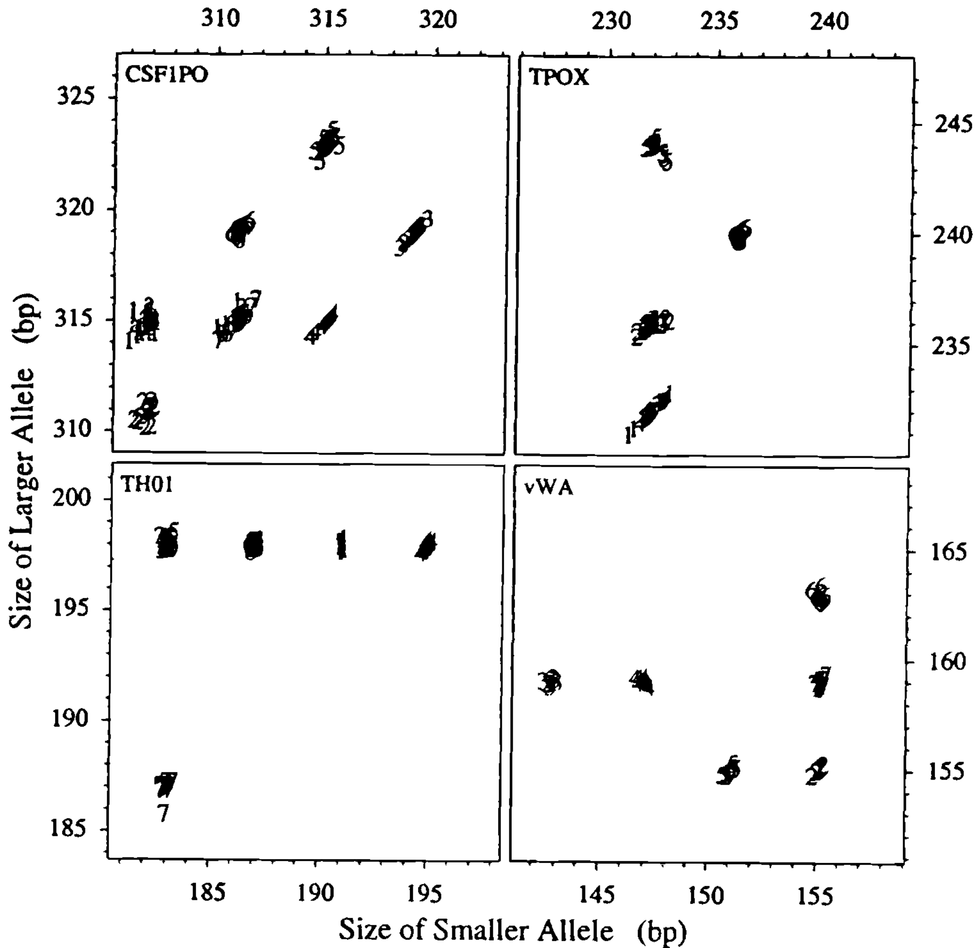

FIG. 3—*Allele sizes after offset correction. Legend as in Fig. 2.*

acrylamide concentration as well as urea concentration. For these reasons, differences intrinsic to the sequencers are difficult to isolate from matrix effects.

Other Sizing Variables

Thermocycler and Amplification Conditions—No laboratory explicitly examined the influence of thermocycler model and/or amplification conditions on sizing. However, there are sufficient data for indirect comparison. After control of sizing standard and urea concentration effects, there is no evidence for thermocycler- or amplification condition-specific PCR product size variability.

TABLE 6—*Pooled standard deviation for sizing CTTv sample alleles.*

Locus	Number of			SD§	
	Sets*	Alleles†	Bands‡	Data‖	Off¶
CSF1PO	22	5	272	0.93	0.29
TPOX	22	4	234	0.73	0.26
TH01	22	5	288	0.35	0.16
vWA	11	6	139	0.78	0.17

*Number of datasets reporting sizings at each locus; many laboratories submitted multiple datasets.
†Number of different alleles represented in the samples at each locus.
‡Total number of sizing results submitted for all alleles at each locus.
§Pooled standard deviations, in bp units, combining individual SDs for each allele of the locus.
‖SD for data as submitted.
¶SD for data after offset correction.

Sample Denaturation—All but one participant denatured samples prior to loading the analytical gel using indistinguishable methods. Only allele type results were reported from this one "native" (without urea) gel. Although all results for this gel at locus CFS1PO were binned, many other participants similarly binned results.

Sizing Algorithm—The software used in our laboratories to convert migration time to bp supports a variety of user-selectable sizing algorithms (10). By oversight, participants in this study were not requested to supply this information nor is there any indication of sizing algorithm on any of the reports provided. A preliminary investigation in the authors' laboratories indicates that there may be systematic differences among the methods. It is known that the choice of image analysis algorithm used to convert migration distance to bp can influence STR sizing.

Sample 8—As previously described, a subset of participants received an eighth sample. This was a bloodstain from the same source as that of sample 3, but was distributed on a novel isolation substrate. All typing results for this sample agreed with those for reported sample 3.

TABLE 7—*Calculated offsets.*

Lab	Conditions					Offsets (in bp)			
	Ladder*	Instr†	Therm‡	Urea§	Primer‖	CSF1PO	TPOX	TH01	vWA
A	Allelic	Hitachi	480	7.0	P	−0.4	0.1	0.1	
A	Allelic	MD	480	7.0	P	−0.4	0.1	0.0	
B	Allelic	LI-COR	MJR	7.0	P	0.1	0.1	0.1	
C	Allelic	Hitachi	480	8.0	Pv	−0.2	−0.1	0.0	−0.1
C	Allelic	MD	480	8.0	Pv	−0.2	−0.1	−0.1	−0.1
D	350	ABD377	9600	6.0	A	1.4	1.4	0.5	
D	350	ABD377	9600	6.0	Pv	1.0	1.1	0.6	1.1
A	500	ABD373	480	8.3	A	2.0	1.7	0.5	
A	500	ABD377	480	6.0	A	1.3	1.7	0.8	
C	500	ABD377	9600	6.0	A	0.5	1.3	0.4	
C	500	ABD377	480	6.0	Pv	0.5	1.4	0.5	0.9
E	500	ABD377	9600	6.0	A	0.7	1.5	0.4	
F	500	ABD377	9600	6.0	A	1.0	1.6	0.5	
F	500	ABD377	9600	6.0	Pv	0.7	1.5	0.4	0.8
F	500	ABD377	9600	6.0	Pv	0.7	1.5	0.4	0.8
G	500	ABD373	9600	7.0	Pv	1.5	2.6	1.2	1.9
G	500	ABD373	9600	8.3	Pv	1.1	1.2	0.9	1.7
H	500	ABD373	480	8.3	Pv	1.4	0.5	0.9	1.7
I	2500	ABD373	9600	8.3	A	2.5	1.1	0.1	
I	2500	ABD373	9600	8.3	Pv	2.3	0.5	0.3	0.2
J	2500	ABD373	480	8.3	A	2.5	0.9	0.2	
J	2500	ABD373	480	8.3	Pv	2.3	0.3	0.5	0.4

*Sizing ladder: Allelic–external-lane allelic (Promega).
GS350–internal-lane GENESCAN 350 (PE-ABD).
GS500–internal-lane GENESCAN 500 (PE-ABD).
GS2500–internal-lane GENESCAN 2500 (PE-ABD).
†Instrument: ABD373–ABD373 DNA Sequencer (PE-ABD).
ABD377–ABD377 DNA Sequencer (PE-ABD).
Hitachi–FMBIO™ Fluorescent Imaging Device (Hitachi America).
LI-COR–LI-COR 4000 infrared scanner (LI-COR).
MD–Molecular Dynamics FluorImager SI (Molecular Dynamics).
‡Thermocycler: 480–DNA Thermocycler 480 (Perkin-Elmer).
9600–GeneAmp PCR System 9600 (Perkin-Elmer).
MJR–MJR MiniCycler (MJ Research).
§Concentration of urea in polyacrylamide gel, mol/L.
‖Primer set: A–PE-ABD-CTT (PE-ABD).
P–Promega-CTT (Promega).
Pv–Promega-CTTv (Promega).

Recommendations

Validly comparable results (either the allele name or a two-allele bin including the consensus allele) were obtained for all samples at all CTTv loci when laboratories reported results as allelic name. Without protocol-specific adjustment, the interlaboratory variability of allele size exceeded the four bp repeat of these STRs. Thus, the most reliable current mechanism for interlaboratory exchange of STR results is the qualitative allelic name.

The agreement in allelic name assignments among laboratories using a variety of measurement systems demonstrates that a single "standard protocol" is not required to achieve standardized results. Such uniform results can be achieved by calibration to a common standard, such as calling sample allele names with reference to sets of well-defined (sequenced or consensus-assigned) human alleles.

Adjustment of raw sizing results with offsets calculated from control samples and/or calibration to allelic ladders may provide effective mechanisms for the interlaboratory exchange of quantitative STR sizing results. Further studies are required to determine the most reliable ways of adjusting the raw sizing data and for

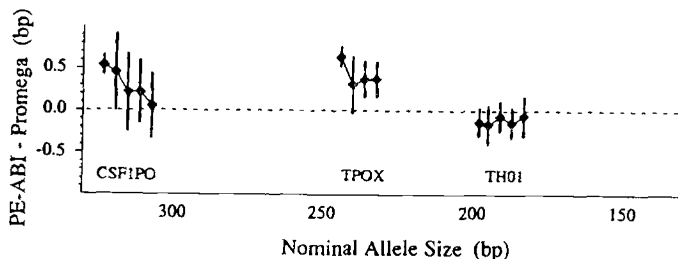

FIG. 4—*Differences between primer sets. The average sizing differences between the PE-ABD and Promega primer sets are presented for all sample CTT alleles. These values are calculated from data supplied by the four laboratories (C, D, I, and J) that evaluated both primer sets under otherwise identical conditions. Vertical bars represent ±1 SD about the average difference.*

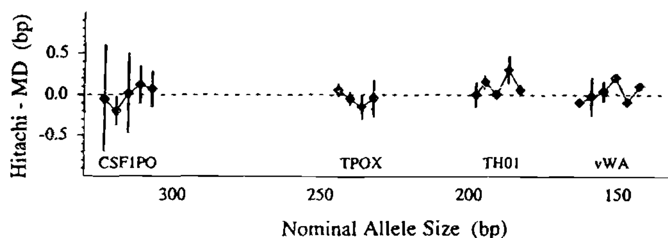

FIG. 5—*Differences between fluorescence imaging devices. The average sizing differences between the Hitachi and Molecular Dynamics imaging systems are presented for all sample CTTv alleles, using data supplied by the two laboratories that evaluated both instruments. Vertical bars represent ±1 SD about the average difference.*

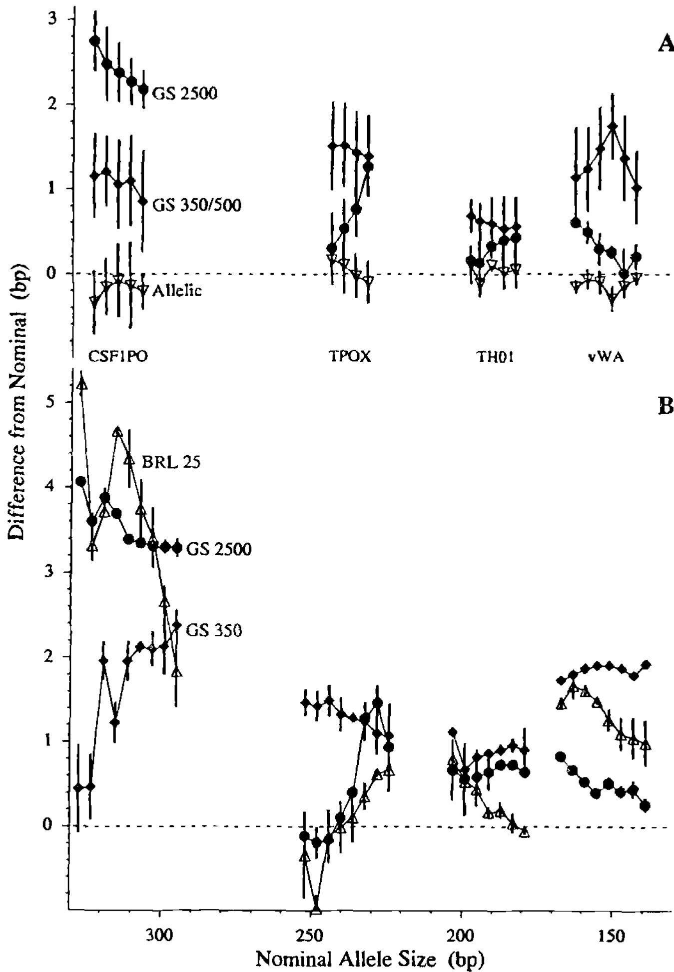

FIG. 6—*Differences among sizing ladders. Figure 6a presents average differences between the reported and the nominal allele size of all sample CTTv alleles grouped by sizing ladder. These values are calculated using all interlaboratory datasets sized with allelic (▽), GS350 or GS500 (♦), and GS2500 (●) sizing ladders. Figure 6b presents the average differences between the observed and nominal size of all components of an allelic ladder. These values are calculated from multiple allelic ladder lanes in one gel run at NIST and sized with three different internal-lane sizing ladders: GS350 (♦), GS2500 (●), and BRL25 (△). Vertical bars represent ±1 SD about the average differences.*

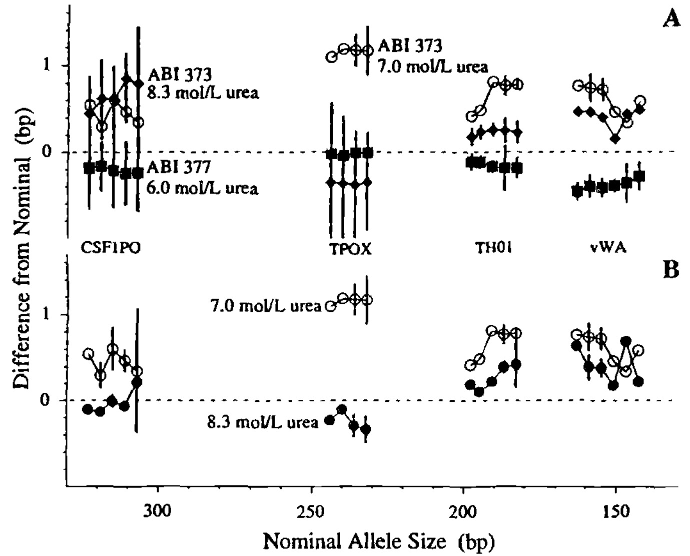

FIG. 7—*Differences among sequencer and/or gel matrices. Figure 7a presents adjusted average differences between the reported and the nominal allele size of all sample CTTv alleles grouped by the combination of sequencer and urea concentration: ABD377 with 6.0 mol/L urea (■), ABD373 with 7.0 mol/L urea (○), ABD373 with 8.3 mol/L urea (♦). These values are calculated using only interlaboratory datasets sized with GS350 or GS500 sizing standards and have been adjusted by subtracting the average GS350/500 ladder difference displayed in Fig. 6. Figure 7b presents the average differences observed in the one laboratory that reported results for 7.0 mol/L urea (○) and 8.3 mol/L urea (●) using otherwise identical analytical procedures. Vertical bars represent ±1 SD about the average differences.*

establishing the intrinsic variability of the adjusted data: (1) the long-term (months to years) electrophoretic stability of current dynamic methods must be established in a number of representative forensic laboratories, and (2) the utility of the various plausible raw data adjustment algorithms must be evaluated. The electrophoretic stability of some common STR sizing systems may be demonstrable using historical data already collected by several forensic laboratories. We believe that the algorithm evaluation can best be accomplished through a designed multilaboratory study of an STR multiplex involving selected samples, controls, sizing standards, and allelic ladders.

Acknowledgments

The enthusiastic cooperation of many analysts and laboratory supervisors made this study possible. We thank them for their largely unsung efforts. We thank Life Technologies, Perkin-Elmer, Promega, and Schleicher & Schuell for donating the materials used in this study.

References

1. National Institute of Standards and Technology. Certificate of Analysis, Standard Reference Material® 2391, PCR-based DNA profiling standard. Gaithersburg (MD): NIST Standard Reference Materials Program; 1995.
2. Youden WJ. Graphical diagnosis of interlaboratory test results. Indust Qual Control 1959;15:1–5.
3. Fourney RM, Fregeau CJ, Bowen JH, Bowen KL, Shutlerr GG, Elliott JC et al. STR binning: practical considerations for statistical evaluation of PCR amplified loci. In: Proceedings of the Sixth International Symposium on Human Identification; October 1995; Scottsdale (AZ), Madison (WI): Promega Corp, 1996;177.
4. Gill P, Urquhart A, Millican E, Oldroyal N, Watson S, Sparkes R et al. A new method of STR interpretation using inferential logic-development of a criminal intelligence database. Int J Legal Med 1996;109:14–22.
5. Perkin-Elmer Corporation. ABD PRISM™ STR Primer Set, PCR Amplification and Typing of TH01, TPOX, and CSF1PO Protocol. Foster City (CA): Perkin-Elmer Applied Biosystems Division; November 1995 P/N 903842 Rev 0;14.
6. Promega Corporation. GenePrint™ STR Systems Technical Manual. Madison (WI): Promega Corp; 1996 Feb Part # TMD006;5.
7. Smith RN. Accurate size comparison of short tandem repeat alleles amplified by PCR. Biotechniques 1995;18:122–8.
8. Worley JM, Mansfield ES, Rubin RB. STR typing accuracy using different molecular weight markers. In: Proceedings of the Sixth International Symposium on Human Identification; October 1995; Scottsdale (AZ). Madison (WI): Promega Corp, 1996;180.
9. Kline MC, Redman JW, Reeder DJ, Duewer DL. Intercomparison

of DNA sizing ladders in electrophoretic separation matrices and their potential for accurate typing of the D1S80 locus. Appl Theor Electrophor 1996;6:33–41.

10. Perkin-Elmer Corporation. GENESCAN 672 Software User's Manual, Rev A. Foster City (CA): Perkin-Elmer Applied Biosystems Division; September 1993 P/N 902842;2–12.

Additional information and reprint requests:
Margaret C. Kline
Biotechnology Division
Chemical Science and Technology Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899