NDLERF

The bioprofiling of illicit drugs

Monograph Series No. 30

Funded by the National Drug Law Enforcement Research Fund
An Initiative of the National Drug Strategy

The bioprofiling of illicit drugs

Leigh Alexander Burgoyne
School of Biological Sciences, Flinders University

David Edward Arnold Catcheside
School of Biological Sciences, Flinders University

Kenneth Paul Kirkbride
Forensic Services, Australian Federal Police

Funded by the National Drug Law Enforcement Research Fund, an initiative of the National Drug Strategy

Produced by the National Drug Law Enforcement Research Fund (NDLERF)
GPO Box 308, Hobart, Tasmania 7001

ISBN: 978-0-9804654-5-7

ISSN: 1449-7476

The research on which this report is based was funded by the National Drug Law Enforcement Research Fund, an initiative of the National Drug Strategy.

Table of Contents

List of Figures

Abstract

It has been found that DNA sequences can be extracted and amplified from typical drug seizures. Non-human DNA in seizures was readily compared for similarities, pair-wise, seizure to seizure and this should be applicable to police intelligence almost immediately and court usage after considerable experience and validation. The technology's limits are explored and future developments are suggested. Drug seizures usually have less DNA than soils but seizures have a potentially useful human content. Even in the relatively small quantities of drug subjected to testing, the human DNA content was sufficient for conventional forensic "trace DNA" techniques to be quite promising. It is suggested that this human content should be treated as a special case of trace DNA. The limited data currently available suggest that in principle the human profiling described in this paper could be conducted by any forensic laboratory around Australia and across most of the world using familiar equipment and techniques. The profiles generated would be compatible with DNA databases such as National Criminal Investigation DNA Database (NCIDD). An application has been made to NDLERF to validate this approach.

The DNA sequences database produced as part of this project has not been included in the report, but is available by contacting the NDLERF Secretariat.

Executive Summary

The hypothesis under examination in the present research is that, as a consequence of illicit drugs being manufactured, "cut", and distributed with no control over contamination, they are likely to contain traces of biological material such as microorganisms, plant remains, and other cellular material, all of which potentially contain DNA. If this hypothesis is correct, and the DNA can be extracted in quantities to allow profiling, then DNA could be used as a means to link apparently unrelated seizures or ascribe their provenance.

This project has proven this hypothesis is quite reasonable, and it is ground-breaking in that it represents the first instance of isolation and profiling of DNA from drug preparations. An additional unexpected finding was that a reasonable proportion of seizures yielded partial human DNA profiles upon extraction. Development of methodology to realise this opportunity to identify individuals involved in handling drugs is the objective of a continuation of this project currently under consideration by NDLERF. The detailed objectives of the proposed continuation are attached below.

Simple techniques for the extraction of human and non-human DNA from drug seizures have been developed as a result of our research. Once extracted from drugs the next step is to profile the DNA. In relation to human DNA, our research has indicated that the standard forensic DNA techniques that are in use in most forensic laboratories around the world are effective for the profiling of DNA originating from drug seizures. Although this finding was unexpected it is arguably the most valuable to emerge from our research. In addition to one of the stated goals of the project (i.e. using DNA profiling as a means of establishing links between apparently unrelated seizures) the human DNA work has yielded extra benefits. Profiles obtained from drugs could be searched against the Australian national DNA database with the possibility of identifying unsuspected conspirators to drug trafficking. Current forensic human DNA profiling is mature, well understood, and well validated; therefore the research described in this report will benefit from significant leverage on that foundation.

In relation to non-human DNA profiling and comparison, the standard forensic techniques are not applicable and a new approach had to be devised. The approach uses a large number of spots of DNA in a two dimensional array on a microscope slide which is flooded with DNA amplified from a drug seizure. Where the amplified DNA shares similarity with a DNA spot on the array it binds to that spot. The DNA is tagged with a coloured fluorescent dye, allowing the degree of binding to be measured for each of the hundreds of spots on the array, forming a profile that can be compared and stored in a database. Pairs of profiles can be further checked for a "match" by tagging the DNA from each drug seizure with a different coloured fluorescent dye, usually green and red respectively. When both seizures bind to the same spot equally, the green plus red leads to yellow fluorescence, allowing the degree of similarity to be determined.

The construction of arrays is a technically-demanding procedure that forensic laboratories in Australia are not equipped to conduct. In order to implement the technique it would be advisable to contract a commercial source of arrays. Costs are likely to be a few hundred dollars per array, potentially much lower if demand is sufficient. The remainder of our technique should pose no serious problem to forensic laboratories as it is no more demanding technically than the techniques currently in use. A device to read the arrays and hybridisation equipment would be required, at a cost of about $200,000.

The array technology employed in this project could be applied now for linking drug seizures based on their non-human DNA content for intelligence purposes, although there is need for further validation to determine the extent of false positive results (i.e. seizures that look the same, but aren't) and false negatives (seizures that look different, but aren't), especially for prosecution work. A survey of a significant number of seizures would be required to address these issues. A survey would also allow refinement of the array with a view to reducing the false positive and false negative rate.

More sophisticated profiling techniques based on direct sequencing of the amplified DNA could replace the array technique as the cost of sequencing declines. However, such changes will not avoid the requirement to conduct a significant survey in order to ascertain the incidence of false positive and false negative results. New sequencing techniques may offer greater throughput at lower unit cost and greater precision as a result of profiling the entire DNA present in a seizure or sequence subset, with the potential to ascribe provenance of a seizure through the identification of geographically characteristic species.

In summary:

- The research conducted has indicated that human DNA can be found in drug seizures and that it can be profiled using commonly accepted methods. Validation of this, as is proposed in a continuation of the project, would allow law enforcement officers to use these DNA profiles in drug investigations to identify a new person of interest through searching against the DNA database, to indicate which seizures have been handled by a known person of interest, or to identify seizures that are linked by the DNA of an unknown person.
- The research conducted has indicated that non-human DNA is present in drug seizures. Within the confines of a limited study it has been identified that DNA can be used to make links between seizures that could be used to inform tactical intelligence. An array-based technique has proved the concept and more sophisticated techniques could improve the process.
- Although DNA profiling of drugs offers the same sort of support to law enforcement that chemical profiling provides, it is an advancement because it can identify persons that are involved (which cannot be done by chemical means) and it can be used on drugs that are difficult to profile using chemical means, such as "ice." Forensic *drug chemistry* laboratories do not currently have the capability to conduct DNA profiling of drugs. Forensic *biology* laboratories do have the capability for human profiling and non-human profiling if arrays become commercially available, but capacity to take on new and additional work is likely to be an issue.

The most immediate recommendation: using the DNA extraction protocols developed in this project we propose to find the proportion of available drug seizures that are likely to have a useful human DNA content and, in general, answer the following questions.

1. Assess the merits of the following profiling techniques and make recommendations as to the best in terms of information reliability:
    a. by standard amplification with Profiler plus ™
    b. by standard amplification with Minifiler™
    c. by the addition of preamplifications (whole genome amplification that the forensic community currently finds acceptable)
    d. by the addition of post-amplification clean-ups.

2. Identify means by which powders might be screened for the presence of DNA prior to profiling.
3. To gain insight into the importance of the time since seizure for success in obtaining profiles.
4. To be able to measure the properties of the seizure that influence DNA- survival-time in deciding whether amplification is likely to be useful (e.g. seizures of near neutral pH and/or dry with no vapours would be expected to preserve DNA much better than those with an "acid head").
5. To determine the interactions between time since seizure and factors that affect DNA stability on the ability to obtain DNA profiles e.g. at low pH there is a more serious concern.
6. To determine the procedures necessary in order to avoid human contamination of drugs by law-enforcement personnel along the chain of custody.

With respect to the project, as it stands, there have been very encouraging pilot studies on Goal 1a. Goal 2 has also been considerably advanced but all the others are virgin fields.

Abbreviations

BSA	bovine serum albumin
Bug	abbreviation for any unidentified micro-organism cultured from seizures
cy3- green	(false colour- excitation at 532nm) fluorophore purchased attached to deoxycytidine triphosphate
cy5- red	(false colour excitation at 635nm) fluorophore purchased attached to deoxycytidine triphosphate
GHB	gamma hydroxyl butyric acid
LSD	lysergic acid diethyl amide
MDMA	3,4-methylenedioxymethamphetamine, ecstasy
PCR	polymerase chain reaction
SNP	single nucleotide polymorphism
STR	short tandem repeat

Summary of Results

The objectives and plan as per the original application and as developed by the steering committee during the progress of the project had a number of aims and objectives. These were

- identify highly discriminating DNA sequences present in four different drug types (heroin, methylamphetamine, ecstasy, and LSD).
- detect similarities between the drugs to determine whether they have a common source and, if so, what is it and what is the body of methods or approach most likely to be fruitful for these highly related purposes.
- Develop a profiling technique based on DNA arrays and ATNA-SMIPS^{tm} technology.
- Develop an experimental database in order to validate the concept and explore the generality of the technique.
- develop new technologies.
- bear in mind any developments that might help determine who has handled the drugs or the materials used to cut it at some stage. (This wasn't implicit in the original objectives but was flagged to the steering committee and extensively discussed).

Various protocols for the preparation and labelling of DNA from various types of seizures including tablets were established early in the project but they were simplified during the course of the project. The earliest techniques developed should still have special application with some seizures but the method of choice was a simple extraction with dilute sodium carbonate because the mild alkalinity of the carbonate solubilises the DNA and the carbonate ion protects the DNA content in at least two ways: first by neutralising any patches of residual acid, from the seizure's manufacture, and second by inactivating divalent metals that catalyse DNA damage. The insoluble residues were centrifuged out and the soluble DNA was then collected on microcon^{TM} filters with any soluble components of the seizures passing through the membrane.

On comparing the carbonate-microcon^{TM} method to other standard procedures, such as for example the "chelex-procedure" that is routine laboratory procedure for textiles, the following points can be made. First and foremost, the carbonate-microcon^{TM} procedure is designed to remove free DNA from surfaces like those of fine crystals or the surface layers of tablets and it would not be expected to efficiently extract high-molecular weight DNA embedded, say, as blood, in a coherent matrix such as a textile as does the "chelex-procedure". Against that, carbonate-microcon^{TM} method will give native DNA and should not break the DNA whereas the "chelex-procedure" gives denatured DNA that has undergone some breakage.

The carbonate procedure also involves less handling than the chelex-procedure and does not use materials difficult to get free of incidental DNA such as chelex resins and, although this last point is not very important when profiling human DNA, we found it to be significant for non-human DNA.

Sequences were cloned from seizures of illicit substances and classified as far as possible, through their relationship to sequences previously lodged in GENBANK and a data-base of sequences was begun from seizures. The value of this data-base to practical forensics has clarified over the last few years and is discussed further in the discussion section of this report.

A summary of the core outcomes from the arrays made from, first, raw seizures and the single-sequence data base are:

- It was relatively easy to compare seizures for similarity by a simple visual observation of pairwise scatterplots made from any two seizures of interest.
- Human DNA was commonly observed in seizures. Its occurrence was studied it was found that, while often at prohibitively low levels, it can be at levels indicating conventional human genome profiling will be practical in a significant number of seizures.
- Non-human DNA, mainly microbial DNA, is very commonly present in seizures of illicit substances and these sequences appear to have the most utility in distinguishing seizures from each other. (As opposed to tracking *who* has handled them). It is considered that the non human sequences are most likely to lead to useful inferences about provenance.

The practical problems of obtaining DNA and amplifying DNA from illicit substances were relatively easily solved, save the problem of LSD paper, but were also the subject of continual incremental development involving optimisation of primers, cycle characteristics, anti-contamination measures, methods of dealing with persistent impurities, challenge-labelling, probing stringencies and artefact-blocking.

The generation of artefacts during amplifications was closely studied and techniques to monitor them and block their effects described. The problem of artefact generation increases steadily as the content of DNA in a seizure falls so that methods of handling it currently set the lower sensitivity limit of this type of analysis. Some of these artefacts were sequenced and some were found to be generated from the primer by template-switching processes and others were anonymous genomic nucleic acids from the reagents. Methods of knowing their influence on the results have been found and, to a point, blocking them is feasible. In the absence of any DNA in the seizure (completely negative assays), the amplifications almost always produce artefacts as product but there are characteristic indications when this is the case.

In pairwise comparisons of seizures, for evidentiary purposes, exclusions are probably more meaningful than inclusions but this is an area of opinion until there are worthwhile statistics on random match frequencies. Generation of sufficient data for measuring random match frequency is a major task for the future. However, the use of comparisons for intelligence purposes appears to be of a much more immediate value as “matches” are very simple to visualise, probably as much or more so than GC-MS results.

Theoretical foundation for the project

DNA is found in the nuclei of cells from living things, such as microbes, fungal spores, plant debris and pollen, as well as human sources. Through analysis of DNA it is possible to deduce the higher taxons from which cellular material originated, sometimes down to the level of species and, in certain circumstances such as the analysis of human DNA, it is possible to distinguish between the DNA from different individuals.

Given that they are manufactured under improvised conditions, drugs are likely to be contaminated with biological material such as skin flakes, saliva, dandruff etc. – from humans, human commensals such as skin and gut bacteria, air-borne dusts with pollens, soils, fly specks, mould spores and similar. These "drop into" drugs whenever they are exposed to the atmosphere during their manufacture, packaging, and distribution. Although they are manufactured under controlled conditions, cutting agents such as sucrose and glucose will unavoidably carry trace biological material that will be incorporated into drug preparations during cutting.

The first aim of the project was to establish whether enough trace biological material is present in typical drug seizures to be collected and analysed. If there is enough material present the second aim was to devise ways of profiling drug samples based on their incidental DNA content for law enforcement purposes. This project is the first of its type in the world.

The non-human DNA content is expected to be both complex and at low concentrations and from almost any conceivable genome type. Thus a method was needed that was not genome-dependent (ruling out specific primers), and which was also compatible with extremely high-gain amplifications.

Profiles were thus generated using single, arbitrary primers at such high amplifications that the sequences selected for amplification were selected by their *efficiency of amplification, not by conventional primer-template specificity*. In effect, amplification is a competitive process, and only the "fittest" sequences thrive and survive to be detected. Profiles derived in this way should be representative of the menu of sequences present in the seizure, not the relative frequencies of the sequences in the seizure, and should be resistant to amplification distortions because these have been allowed to exhaust themselves.

The profiles were to be compared by arrays that are responsive to sequences and not compared by PCR-product lengths. This philosophy is thought to be both robust and that most likely to be compatible with future generations of DNA technology.

Overview of the project's course

Over its three years, the project proceeded approximately as expected. DNA profiles, both non-human and human, were obtained and analysed. When there is a reasonable DNA content there is no serious difficulty in assessing the similarity or difference of one sample with respect to another. The array-profiling of DNA from seizures is novel and required the development of techniques for useful interpretation which are described in this report.

Human DNA was intermittently, even frequently, observed (given the relatively small number of samples examined) in seizures, and observations of these instances led to a few seizures being conventionally profiled for human STR loci. This conventional profiling for the human genome had enough success to stimulate interest and this led to an application for further work on this aspect of seizure DNA. Curently, we don't know whence the human DNA came. For example, it could be forensic chemists as they handle and examine the samples, probably not from the police as they should keep samples intact. It could come from the cutting agent, or country of origin However, this doesn't invalidate the work, but it does mean that contamination in the lab and police station must be minimized. This is now second nature for forensic biologists, but not chemists. Contact leaves a trace, it doesn't matter who makes the contact or when.

With respect to the seizures we had, the heroin seizures seemed to act as if they had quite high DNA levels but this inference from the results was not pursued at this time as the DNA content could only be estimated by examination of amplification products from its DNA and seizures with enough DNA to profile did not have enough DNA in them for a conventional DNA estimation.

We have assembled a small experimental data-base including human but mainly non-human DNA sequences that can be used as a reliable basis upon which to compare seizures. We now have enough experience to suggest that future progress with the profiling of seizure DNA will be successful but exactly how this might best be accomplished is open to conjecture. It is suggested that the best option would appear to proceed via synthetic oligonucleotide arrays constructed from the knowledge gained from our amplified polynucleotide-arrays. This is a recognised and well-developed technology in medical science but is relatively expensive. In the far future, arrays might ultimately be replaced by absolute sequencing after emulsion amplification but, at this time, this prospect is far too distant to plan for in an orderly way although the more immediate prospect of oligonucleotide arrays that exploit our data-base, look quite realistic.

There were some disappointments: we never had enough LSD or GHB seizures to allow us to beat the problem of getting amplified DNA from them. We have processed significant amounts of tablet-form seizures but only MDMA (ecstasy) as we only had a large amount of ecstacy in tablet form. There was clearly observable DNA in about 50% of the tablets, some of which may have levels high enough to give regular STR profiles.

There were other difficulties, but one difficulty – the limited range of drug seizures we had – was positive in some respects as it forced the project to concentrate on some quite large amphetamine seizures and to face-up to their characteristically very low DNA content and this, in turn, forced the project to face up to the molecular issues arising from extremely low DNA levels.

Definitions with brief descriptions of the core technology

A profile: an amplification mix or product mix or "signature-mix"

In this report, these terms refer to the products of a high-gain single-primer amplification derived from the DNA sequences derived from a drug seizure. Because of the highly selective properties of the primers in combination with the amplification procedure, a drug seizure may give a suite of sequences commonly containing approximately four to 12 major components selected by the amplification process itself, from the seizure, and this suite is a seizure's "profile". Such suites of sequences are non-random selections from the seizure DNA and it is because of this non-randomness that they can be regarded as profiles or "signatures" of the much more diverse mixture of sequences that they are derived from. The two-dimensional scatter-plots most commonly presented in this report are visual representations of one seizure's profile being compared to another's against a data-base. This makes them a three-way comparison.

An array-challenge

The term used when using profiles. Like many other phrases in English this can be used for an item (noun), or its application (verb).

- The item called an "array-challenge": the mixture of two DNA extracts each labelled with different fluorescent tags that can be applied to an array.
- The act of doing an "array-challenge" is the application of such solutions to an array.

A challenge might be a mixture of profiles from a heroin seizure and an amphetamine seizure. The colour actually refers to false-colour defined as the colour of the excitation wavelength used to make the fluorophores in the DNA fluoresce rather than the colour of their fluorescence. In this report only two such false-colours are used: red fluorophore cy5 and green fluorophore cy3.

An array

In this report, an array refers to a rectilinear set-out of hundreds or thousands of DNA-containing spots in an area of one or a few square centimetres. The spots are bound to a glass surface by covalent bonds so that they can be challenged by another set of sequences, typically a profile derived from seizures as described above. Very commonly two sets of profiles are compared to each other against such an array using the array as the medium of comparison for assessing similarity/dissimilarity of profiles.

An "array-element", "spot" or "probe" (synonyms)

This refers to a single small spot, a few hundredths of a millimetre in diameter, originally applied to a glass surface as a liquid solution of DNA. As the spot dries, its DNA-content is covalently bound to the glass surface. When such an array of elements is challenged each element/spot will bind any sequences that are, to some degree, homologous to it. As the binding sequences are fluorescently coloured, each spot or element will become appropriately coloured.

The different types of arrays

Mixed sample arrays

This is made from a set-out of unsorted profiles applied directly as spots (array elements) to a suitably receptive glass surface. This type of array is thus hundreds of spots each composed of a "profile" comprising a small *suite of sequences* derived from a seizure. This type of array, with its highly complex spots was important during the early stages of the project. They were challenged by profiles derived from seizures and labelled with fluorescent nucleotides

Single sequence arrays from clones

Single sequences were cloned and isolated from profiles and these single sequences were applied to the glass surface to give spots (array elements). The spots were then challenged by full profiles derived from seizures and labelled with fluorescent nucleotides.

Synthetic oligo-nucleotide arrays

These are probably the future, they have not been used in this project.
The technology of making arrays and challenging them with mixtures of sequences derived from illicit seizures and presenting the results is described in the Methods section.

Usage of the terms "provenance", "match" and "mismatch"

These terms refer to the "end-use" of the data and thus have slightly different meanings when addressing intelligence gathering or prosecution/court end usage The usage of the terms match and mismatch is conventional but as match-probabilities for prosecution/court end usage are further in the future than intelligence-gathering end-usage the terms "match" and "mismatch" are used as simplifications, for their convenience and speed. The writers are aware of the limitations of these terms due to the current lack of sufficient data to determine the frequencies with which these "matches" or "mismatches" would occur by chance. However, the value of the usage of these terms for intelligence-gathering purposes comes from the observation that visually obvious patterns, seen in plots of experimental results, are often quite readily and intuitively appreciated as signatures of relationship. (See the examples described below).

Sources of seizures

Seizures were supplied by South Australian Police and Forensic Science South Australia.

The identification details and any information that was supplied with the seizure are noted where they are used to generate results.

Chapter one: Interpreting the arrays: The signatures of "match" or "mismatch" and "grey zones" between them

Precis: Chapter one uses examples extracted from all the results to illustrate the general rules for the interpretation of the two-way comparisons by scattergrams.

In the discussion below, an array of hundreds of sequences or sequence mixtures on a glass slide is challenged with two sequence mixtures from, for example, two different seizures, one labelled red and one green, and the red and green florescence data used as the X and Y axis values for each spot on the array. This generates a "scatter-plot" with a pattern that summarises the comparison.

A discussion of the meaning of these two dimensional scatter-plot patterns is illustrated below with a discussion of examples of the extreme types of the two dimensional scatter-plot patterns and then the "grey-zones" between.

Examples of the extremes

1. **"Null"***:* no DNA in either sample. In this case there should be nothing on the scatter-plot. This is not observed because of the synthesis of identifiable non-informative sequences. This is discussed extensively later in this report.
2. **"Mismatch"**: when there is DNA in both samples but the sequences are grossly dissimilar. Figure 1.1 is an example of a pattern indicating a strong mismatch between two profiles, in this case *E. coli* DNA, versus a real seizure. The strong mismatch is reflected by the pattern of array spots falling exclusively on or near one or the other of the axes. This implies that both the profiles have many sequences homologous to sequences on the array but have no similarity to each other.
3. **"Match"***:* When the two samples appear to have the same or almost the same spectrum of sequences within them. Figure 1.2 is an example of a pattern indicating a near-perfect match between two seizures, in this case a contrived example comprising a mixture of the the profile from a single seizure part labelled with one dye and part labelled with the other dye, allowing the degree to which identity can be detected by the technology to be assessed. Note how the array spots fall in a single line in the body of the graph showing, as expected, a high degree of match between all the sequences, whatever they may be, within the DNA sample and, moreover, those sequences are present in very similar amounts. Note, however, that it doesn't say that the DNA actually came from the seizure as if the sample (split into two) has little seizure-DNA in it whatsoever they *still* will be identical because they both are the signature of the same reagent impurities. This issue is dealt with later.

Figure 1.1 is a control merely to demonstrate the consequences for the scatter-plot appearance of complete lack-of-identity (i.e. a mismatch) of the profiles from two unrelated samples. One challenge of the array is from a lab strain of *E. coli* amplified with the primer Cpali20mer2G and the other challenge is from heroin seizure, sample S1, amplified with the primer antiseqseq05. (See methods.) The labelled "spots" or elements are all on or very close to the X and Y axes of the scatter-plot. This is as expected for challenges that share no significant similarity. Notice how the

cognate profile is either at the top of its axis (*E. coli* from FTA paper for the Y axis) or near the top of the axis (Heroin seizure S1 on the X axis). Why the heroin challenge cognate array-element was not quite at the top of the X axis was not explored but it probably means that much of the signal in the heroin challenge S1 is coming from human DNA present in the profile and the seizures' spots being more responsive to the human profile spot than to its cognate spot due to the high human content.

Figure 1.1 Data showing the pattern characteristic of a clear mis-match. The array probed is a "mixed-spot array" in which every spot (array element) represents one whole profile, rather than a single-sequence in a spot used later in the project. The profiles that make up the spots on the array have diverse sources but the major source is drug seizures, including the one being used in the challenge. The X axis shows the amount of the heroin seizure's profile in arbitrary fluorescence units that bound to that spot. The Y axis shows the amount of an *E. coli* genome's profile in arbitrary fluorescence units that bound to that spot. The two fluorescences can be resolved because the excitation wavelength of one is quite different to the other's.

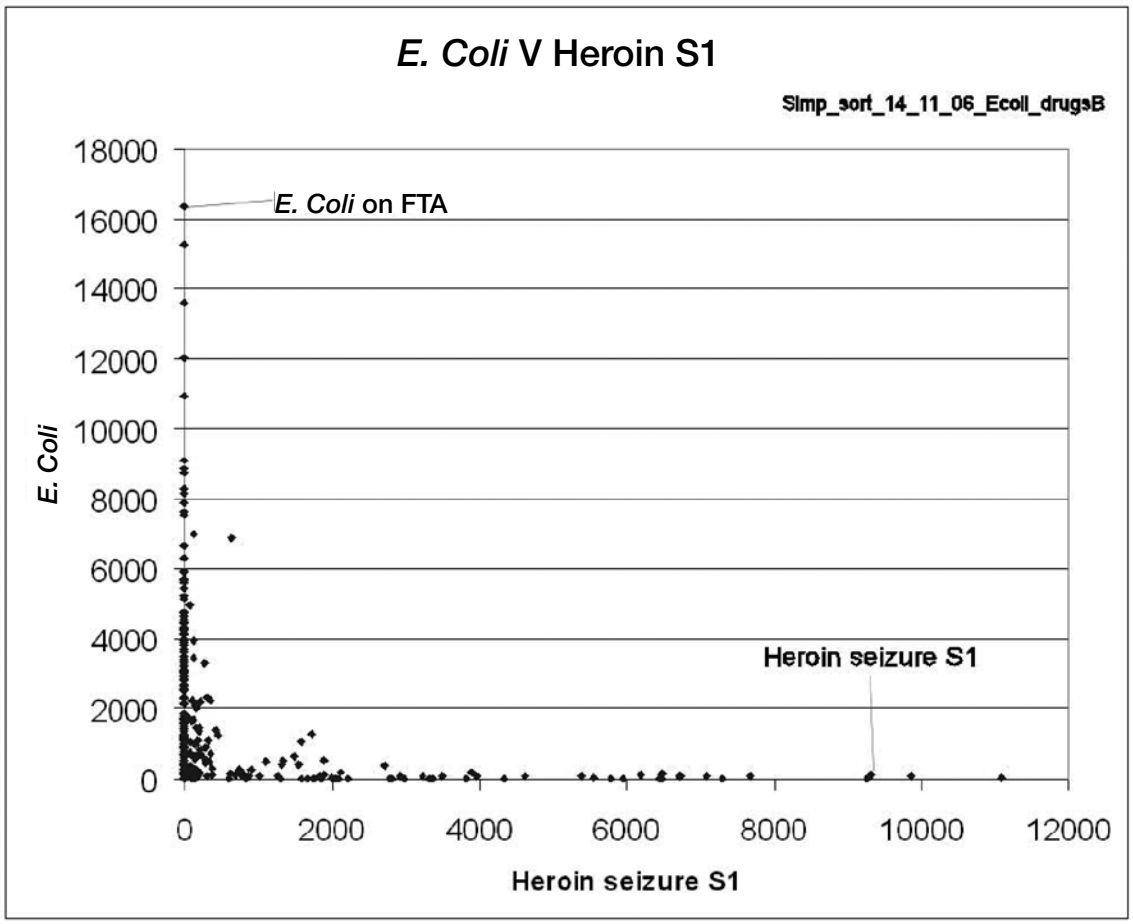


Figure 1.1

Figure 1.2 is a positive-control made simply to demonstrate the consequences of complete identity (match) of the sequences within two samples. In the real world one would never expect such extreme identity from two separate DNA samples, even from the same seizure. Both challenges are exactly the same preparation from exactly the same seizure designated: antiseq05-microcon™ (pellet) methamphetamine REC: 04/B77758-8 bag 0401925-8.27.

As *the two profiles being compared are really exactly the same profile*, albeit labelled differently, the two challenges should be truly identical and the spots should all fall on a straight line. Except for minor variation due to technical issues, this is what is observed; the spots (array elements) that are homologous to the challenges do fall on a line with the different scales for the axis due to the different effectiveness of labelling with the two colours and the very slight divergences from this line are the random noise generated by the technique.

Figure 1.2 A contrived perfect match. The array probed is the same mixed spot array used in Figure 1.1 in which every spot (array element) represents one whole profile, rather than a single-sequence in a spot used later in the project. The X axis shows the amount of the seizure GR142's profile in arbitrary fluorescence units that bound to each spot (cy5, red, label). As this is a positive control, the Y axis shows the amount of *exactly the same DNA* in arbitrary fluorescence units that bound to that same spot but using a different fluorescent label (cy3, green, label) to the X axis. The two fluorescences can be resolved because the excitation wavelengths are quite different.

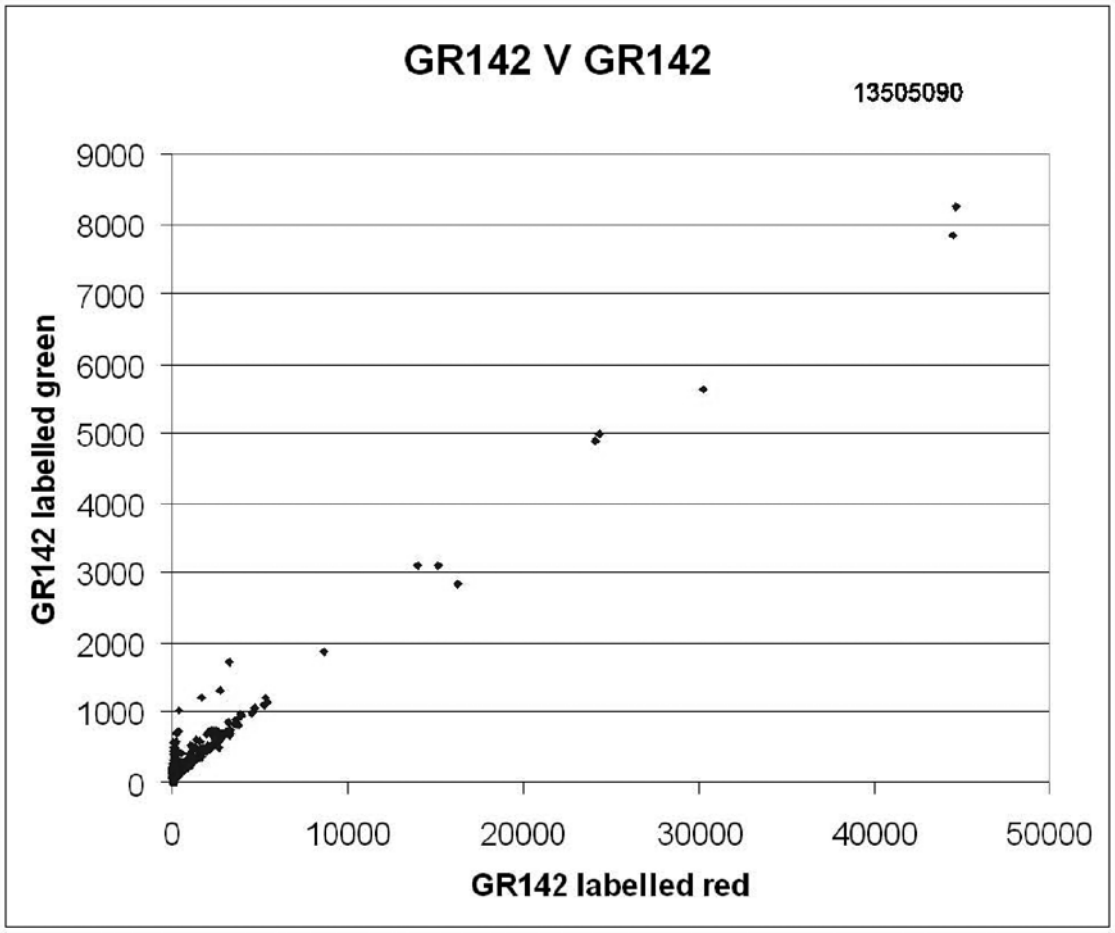


Figure 1.2

The array elements are mixed sequences containing considerable non-informative sequences that do not bind to the challenge DNA. However, this is irrelevant with respect to this particular illustration and is an issue dealt with later. The core matter is that both challenges contain exactly the same sequences and thus give a pattern of a single straight line as expected.

The "grey-zones"

1. **Strong similarity:** an example of a pattern from different but presumptively, very similar profiles (Figure 1.3). Such a pattern is a less perfect case of the pattern in Figure 1.2; it is a scatter around a central line.
2. **Shared features:** an example of when profiles share many sequences but the sequences are not present in closely similar amounts Figure 1.4. Then there is an apparently random scatter of reactive array spots across the scatter-plot that on closer examination appears to possibly be systems of lines radiating from the origin. This is in fact the case as examples with less spots in them show much more clearly.
3. **Near null with only hints of shared features:** *a*n example of profiles that have only a very few sequence and a very few of these are in common and what are in common are present in quite different amounts. In Chapter two, Figure 2.2 shows an example. It is similar to the previous case but in this case the spots seem to fall *in a few sets of lines* that have different slopes. The lines are generated by all those array elements that are homologous to one of the sequences shared by the seizure amplifications, and the slopes of the lines are created by the relative amounts of the sequences in each seizure's profile.

The Figure 1.3 scatterplot is an example of real close similarity, as opposed to the contrived identity in Figure 1.2. The challenging profiles are derived from very closely related seizures. In this case the two profiles are from the water insoluble, alkali soluble extracts from separately-made batches of DNA from closely related seizures so the two challenges are expected to be closely similar and, according to the degree of similarity, fall in a narrow scatter-zone surrounding a notional straight line – to illustrate this, such a notional line has been "dubbed" in.

(Tech note: this slide differs from some later comparisons of NM3 and NM4 in that the known non-informative sequences are "blocked" out with GR170 clones. These are sequences isolated from an amplification, designated GR170, that pilot studies had shown to be rich in non-informative sequences. The presence of high levels of non-informative sequences in GR170 was inferred by the observation that the mixture, designated GR170, showed strong, collective homology to many blank amplifications. Sequences isolated from GR170 can be examined in the data-base (Appendix 2) and all have the prefix GR170).

Figure 1.3 Data from the comparison of two separate samples from the same seizure. This used a "single-sequence array" in which every spot represents one single sequence, not a whole profile as for Figures 1.1 and 1.2. The Challenging profiles were NM3 - REC 06/A40153-7 bags 0503564-71,72,73,74,75 methampheta mine (microcon™ alkali soluble pellet) antiseq05 (cy3) and NM4 - REC 06/A40153-8 bags 0503564-8.1,8.2,8.3,8.4,8.5 methamphetamine (microcon™ alkali soluble pellet) antiseq05 (cy5). The X axis shows the amount of seizure NM4's profile in arbitrary fluorescence units that bound to that spot. The Y axis shows the amount of seizure NM3's profile in arbitrary fluorescence units that bound to that spot. The two fluorescences can be resolved because they have different excitation wavelengths.

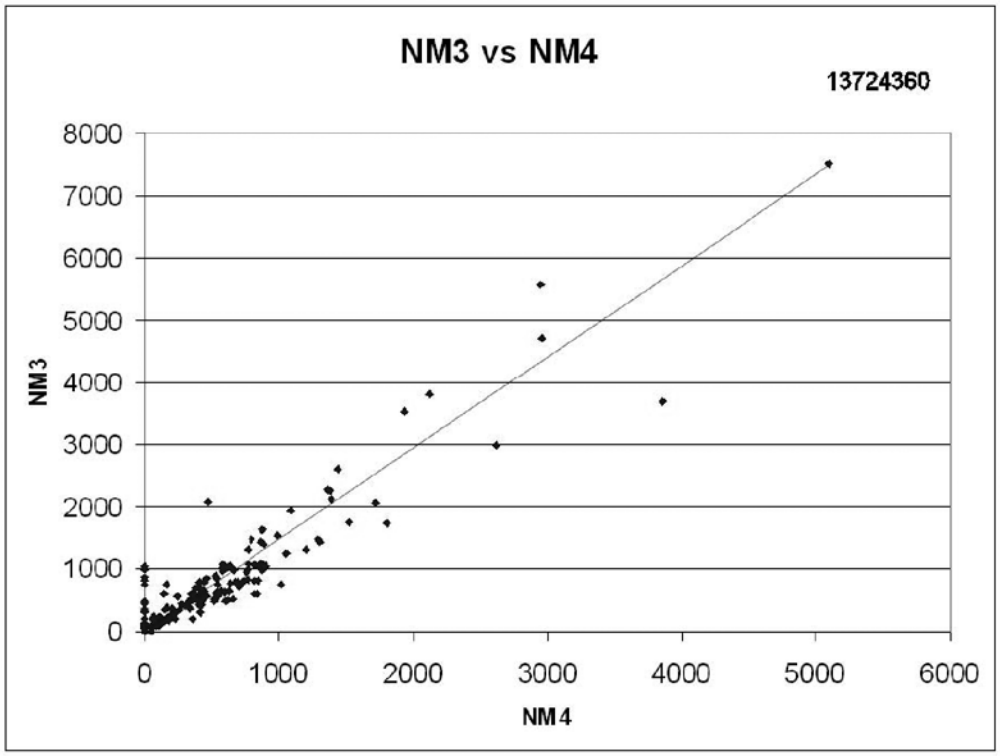


Figure 1.3

When profiles from two closely related seizures are compared (Figure 1.4), the superficially near-random scatter indicates profiles with a lower order of similarity than in Figure 1.3. In this case the two profiles are from the water soluble extracts from separately-made batches of DNA from not-obviously related seizures of methamphetamine. Particularly note that this only looks like a random scatter at first glance but on longer examination, there is the *impression* that much of it is composed of systems of lines radiating from the origin. Systems of lines can be much more clearly seen when there is less similarity or fewer sequences in the profiles being tested against each other (e.g. Figure 2.2).

Figure 1.4 Data from the comparison of samples from two "poorly-related" seizures which have profiles with some elements of similarity. This is a "single-sequence array" in which every spot represents one single sequence, not a whole profile. The challenging profiles were made from ME -(microcon™ sup) methamphetamine seizure. REC 06/A40153-4 bag 0503564-4-amplified with antiseq05 (cy3) and MG -(microcon™ sup) methamphetamine seizure. REC 07/B01458-3 bag 060483-3F-amplified with antiseq05 (cy5). The X axis shows the amount of seizure MG's profile in arbitrary fluorescence units that bound to that spot. The Y axis shows the amount of seizure ME's profile in arbitrary fluorescence units that bound to that spot. The two fluorescences can be resolved because the excitation wavelengths are different.

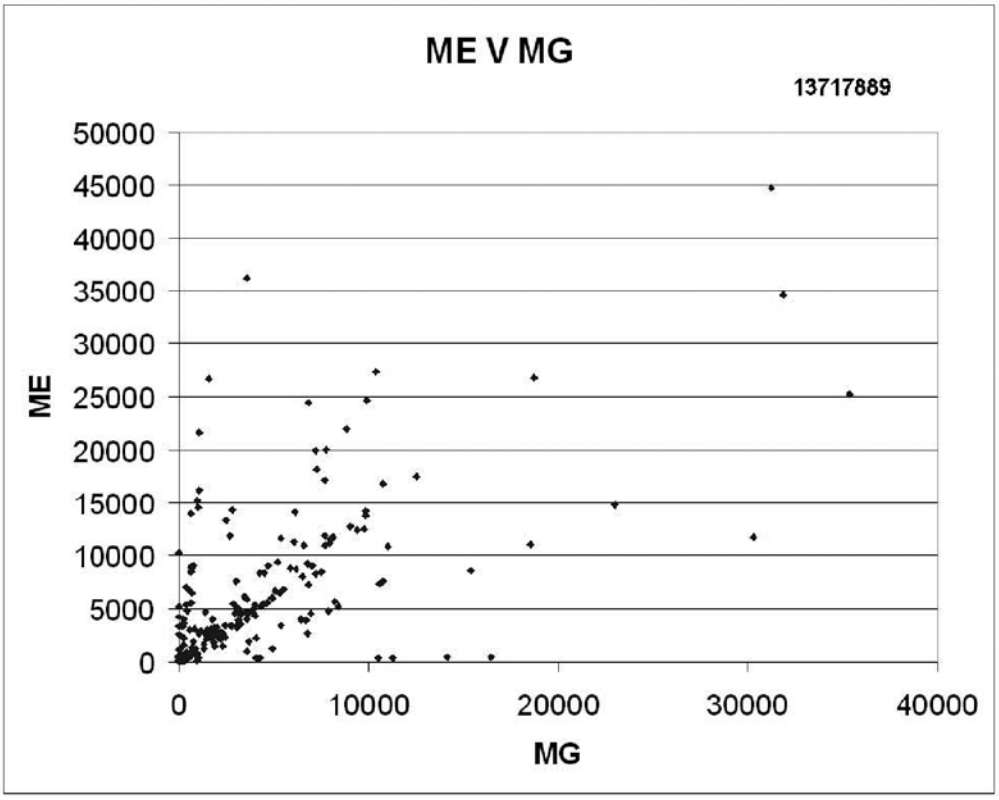


Figure 1.4

Chapter two: Studying the sequences within seizures

Precis: Chapter two first dissects the spectrum of sequence-types found in seizures and considers these types with respect to their probable origin while, simultaneously, developing extraction procedures suitable for soluble drugs. It then goes into the process of interpretation in much more detail than the basics explained in Part 1. A major topic of focus was the unavoidable presence of non-informative sequences; what they are and how to block them and handle them. The final topic in this section is concerned with the extraction of DNA from tablets and the resultant comparisons of tablets to each other.

The common human content

Sequences from the human genome were commonly observed when cloning and sequencing from seizures. This led to a brief pilot-profiling by Profiler plus™ of eight seizure-extracts (2 from each of four seizures) and some trace-DNA mixed profiles were generated. These profiles suggested a new project was justified to follow this further and an application has been made to do so.

- A copy of the project concept for this application is available on request.
- For a sample of the data see Appendix 1. We thank Dr Katrin Both and Ms Karen Lee, Forensic Science SA for the data.

One minor issue concerning human DNA which has not been tested but is an obvious candidate for testing is the prospect of human mitochondrial sequences being useful. However, this has to be assessed against the likelihood of many samples having DNA from more than one individual.

Seizure examinations, general considerations

Gels

DNA samples extracted and amplified from seizures were routinely examined on gels. Figure 2.1 is a representative example.

Figure 2.1 Successful observation of DNA extracted from each of four 20 mg replicates from each of four heroin seizures (S1, S2, S3, and 886) by the solvent-water-Microcon™ procedure. PCR amplification used antiseq05 as primer with two 30ºC annealing cycles, then 62ºC as annealing temperature for two rounds of high-stringency amplifications of 26 cycles then 40 cycles. Amplified products were electrophoresed on an agarose gel.

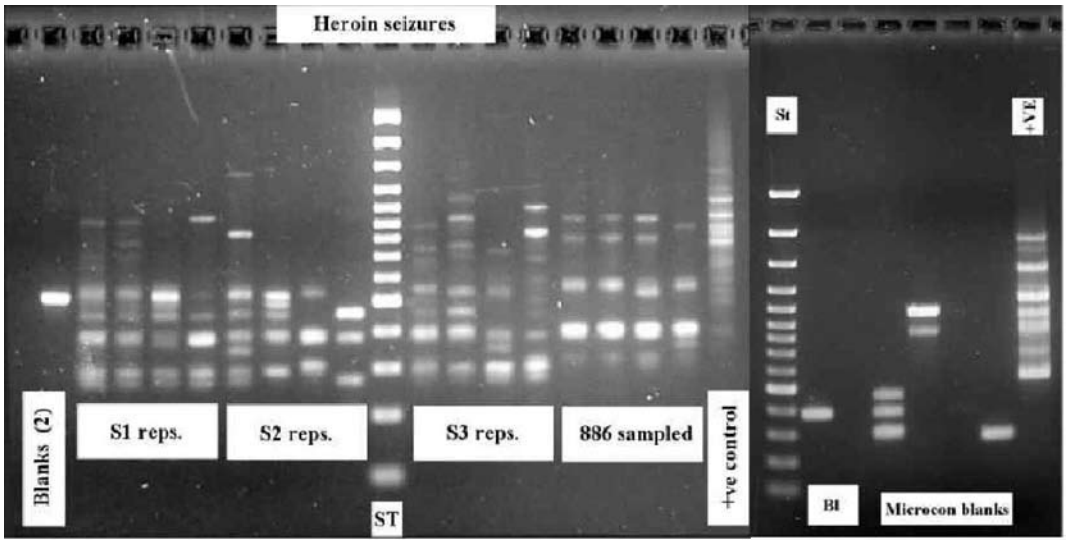


Figure 2.1

Standards:

+ve	*Neurospora crassa* DNA
Bl	reagent alone blank
Microcon™ blanks	amplifications using Microcon™ filters but no drug extract
ST	100 bp ladder

Seizures: 010797 S1, S2, S3 control date: 3/8/05. From a 2001 seizure.

S1 400 mg is 42% heroin, in paracetamol with a trace of ephedrine.

S2 400 mg (34 gm S1 with 18 gm caffeine, 27% heroin).

S3 400 mg (32 gm of S1 treated with 16 gm glucose and 2 gm caffeine, approx 27% heroin).

886 Control date: 1/11/2005. Serial number 000886, divided into three bottles of 350 mg of each. 77.6% heroin as HCl salt. Diluent not specified. Original markings for this set were S1, S2 and S3 but were relabelled S886A, B, C to avoid confusion with the above samples that had used these labels.

Conclusions from this basic type of analysis

At least two procedures provided enough DNA from 20 mg of the above heroin-containing-seizures to allow characterisation of the DNA content, solvent-water-Microcon™ (Figure 2.1) and also solvent-water-DEAE (data not shown).

Subjectively, from band-yields, the DNA yield from solvent-water-Microcon™ is considered most satisfactory and there is also no theoretical reason to believe large losses might be taking place. This method also has the advantage of using familiar reagents that are easy to purify and from this point in the project, sample processing moved over to the water-microcon™ method which is described in detail in the methods section.

It should be noted that this whole study operates at DNA levels far too low for any conventional estimation procedure to be valid. Contamination from reagents and equipment is clearly observable as bands in control blanks (Figure 2.1) but, in our experience with soil-DNA on arrays, this level is manageable. This matter is discussed later in this report.

Some organisms of origin of the sequences subcloned from drug seizures

The sequences identified from seizures were most commonly derived from human sources, either from the human genome itself or from human associated bacteria. By far the commonest identified were skin bacteria such as *Propionibacterium acnes* or human parasites such as *Staphylococcus aureus*. Other sources of identified sequences included *Neurospora crassa* that probably came from the air of our laboratory and caused us to increase our anti-contamination precautions and to include that amongst the sentinel spots placed on arrays. Our later DNA isolations from seizures no longer contained significant Neurospora. However, most sequences from seizures were, as expected, unidentified and thus unclassifiable as had been previously found in another project isolating sequences from soil itself.

Sequences derived from seizures identified from blast searches of GENBANK:

Bacillus licheniformis. Isolated from a heroin seizure. This belongs to *Bacillus subtilis* group commonly found in soil.

Bacillus subtilise. Isolated from a methamphetamine seizure, commonly found in soil.

Bacteriophage SPP1 0.0 Isolated from a heron seizure but commonly from soil bacteria such as *B. subtilis*.

Erwinia. Isolated from a methamphetamine seizure, Erwinia are commonly soil dust-borne organisms, some of which cause bacterial rot of plant material.

Fusarium aywerte. Isolated from a heroin seizure, it's a plant pathogenic fungus

Human sequences were commonly observed in a variety of seizures.

Magnetospirillum. Isolated from a methamphetamine seizure,is a bacterium described as being characteristic of very shallow water; puddles.

"Mus musculus". Isolated from a methamphetamine seizure. Nominally from mouse but the fit is not completely convincing, suggesting it may be from a non-human mammal other than mouse.

Neurospora crassa. From a few seizures. It is a Fungus – probably from strains in use in our lab but it is also common in soils.

Propionibacterium acnes. Isolated from both methamphetamine and MDMA seizures. This is a well-known skin commensal of humans where it can cause acne.

Ralstonia eutropha. Isolated from a methamphetamine seizure. A proteobacterium commonly from soil.

Rhizobium sp. Isolated from both heroin and methamphetamine seizures. It is common in soil where it forms nitrogen fixing symbiotic associations with leguminous plant root nodules.

Salmonella. An isolation from one methamphetamine seizure. Human and small mammal pathogen and commensual.

Shigella. Isolated from a methamphetamine seizure. Human and small mammal pathogen and commensual.

Xanthobacter autotrophicus. Isolated from soil. A bacterium that breaks down xenobiotic contaminants in soil. Commonly found in soil.

Interpreting arrays 1 – examples

This is straightforward and the basics are described with examples in Chapter one.

Two comparisons were chosen as models and examined in detail. The raw data were discussed in the 30 months report. The first comparison was of a methamphetamine with a heroin seizure, two seizures expected to be clearly different. Indeed, a clear difference is observable from mere inspection of the patterns Figure 2.2 as the scatter of spots show some lines, which is not compatible with the seizures having any major similarity. However, there are superficially similar features as many array spots react to both seizures and the sentinel spot (spot GR170, see explanation above Figure 1.3). for the presence of non-informative sequences gave a strong signal so even this similarity is suspicious.

This comparison was repeated in the presence of random hexamer primed reaction products as blocking agents to greatly reduce the contribution of any non-informative sequences. The principles of random hexamer blocking is described in the section on "Blocking" in Chapter four of this report.

Figure 2.3 shows that, in the presence of random hexamer blocking, the remaining similarity fell to negligible levels, exposing seizure ME as having little informative DNA whilst seizure HE still retained its main characteristics. The main sentinel spot, GR170, known from pilot studies to be rich in non-informative sequences, had dropped to coordinates 2175, 2301. This was acceptable in that interesting spots like SD1 and SE5 were well above 8,000 arbitrary fluorescence units on the X axis but the GR170 sentinel spot was well above 5% of 8,000 so it was not negligible. These data illustrate a key component of the project – the development of methodology to clearly distinguish sequence content into informative sequences and non-informative sequences. These categories are discussed and further examined in the section headed *Assessment of blocking non-informative sequences*.

The largest values along the X axis in Figure 2.3 are over spots SE5, SD1, SC5, SE3 – all sequences originally derived from soils rather than drug seizures. None of the sequences were able to be identified by blast searches of the GENBANK databases but, since we have their sequences, primers could easily be designed to specifically extract these sequences if they are present in other seizures. Curiously it seems likely that these sequences are all present in the drug seizure as it seems improbable that four anonymous soil spots should otherwise respond to a seizure when less than a hundred spots on the array are derived from soils. This statistic is so improbable that there is likely to be an underlying reason such as the soil sequences and the detritus sequences in the seizure are similar because they all contain a very common transposon. This is speculation but not wild speculation and if true it has implications for provenance comparisons because transposons are candidates for geographically dependent polymorphisms. This is discussed further in Chapter 4.

Figure 2.2 Data from the comparison of samples from not-obviously-related seizures made from the water soluble extracts from separately-made batches of DNA from ME-(microcon™ sup) REC 06/A40153-4 bag 0503564-4-amplified with antiseq05 (cy3) and HE-(microcon alkali soluble pellet) heroin REC 0E/353051-2 bag 992801, amplified with antiseq05 (cy5) and the single-sequence array in which every spot represents one single sequence was challenged with both sets of amplicons. The X axis shows the intensity of fluorescence emission in arbitrary fluorescence units for spots originating from seizure HE. The Y axis shows the intensity of fluorescence emission for spots originating from seizure ME. The two fluorescences can be resolved because the excitation wavelengths are different.

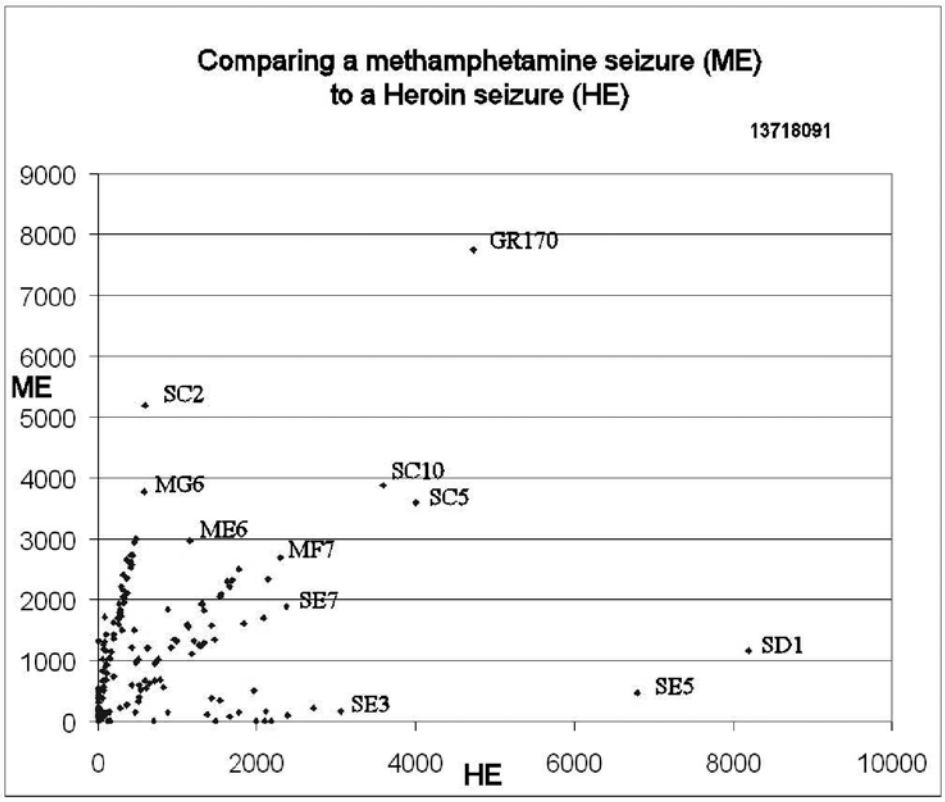


Figure 2.2

The scatterplot exhibits some elements of similarity, reflected by the impression that there are systems of lines radiating from the origin. This is a common observation. This superficially near-random scatter indicates profiles with an even lower order of similarity than those compared in Figure 1.4.

Figure 2.3 This is directly comparable to Figure 2.2, differing only by the presence of blocking DNA that masks uninformative sequences. Note the collapse of spots with respect to the Y axis in comparison to their X axis coordinate which are little changed. For example, array elements (spots) SE5, SD1 and SE3 are still bound strongly by HE sequences but only weakly by ME sequences. In comparison to Figure 2.2 where the two seizures exhibited some similarity, the signal from ME represented by the Y axis has almost collapsed, revealing that the comparison is between a batch with very little informative DNA (ME) where blocking has collapsed its signal, to a batch with larger amounts of informative DNA where the blocker did not collapse its signal much. This shows that the two batches bear weak or no similarity to each other. The principles of blocking are discussed in a later section.

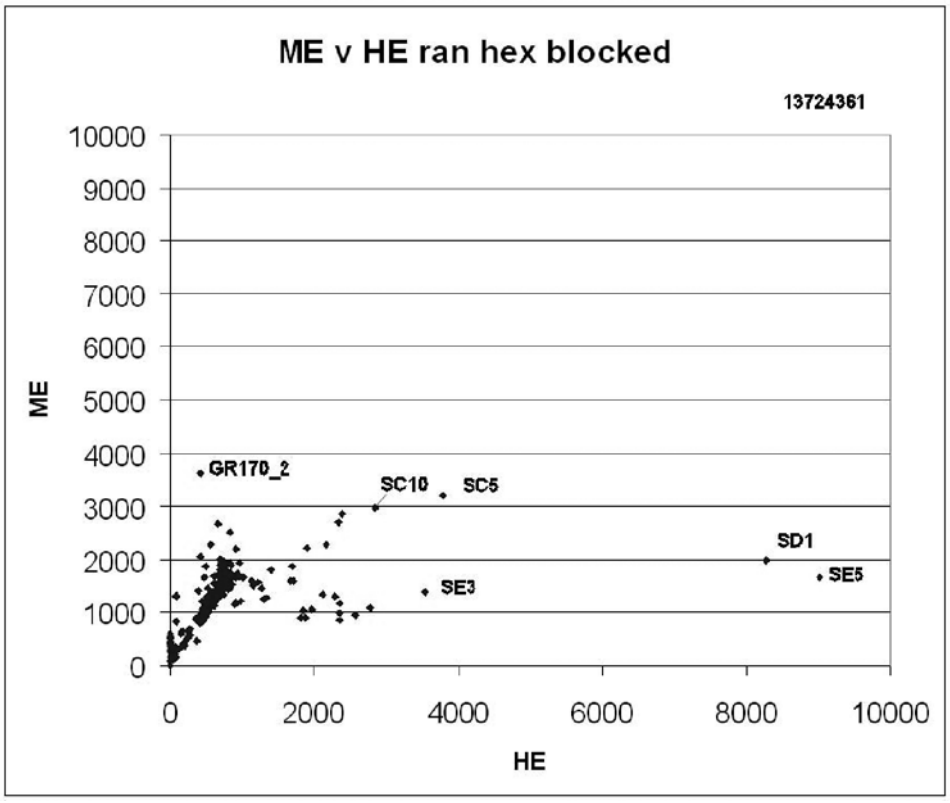


Figure 2.3

Interpreting arrays 2 – distinguishing informative sequences useful for forensic purposes from non-informative sequences that are not valuable for forensic purposes.

Informative sequences include:

- human sequences (only informative when human-specific profiling is used)
- microbial sequences from human sources
- microbial sequences from the environments the drug has been exposed to.

The human contribution is a special issue as it can be either informative or non-informative according to circumstances, as discussed below. However, the non-human but informative sequences are those sequences in seizures, such as bacterial sequences from skin or dust, fungal spores and pollen, that either allow one seizure to be discriminated from another and/or usefully reflect their provenance of the seizure in some way.

Non-informative sequences

The non-informative sequences are those that provide no useful information about the seizure; the most notable examples are sequences such as those that contaminate the enzymes used for the analysis as a result of the practices of the manufacturer supplying the reagents.

Classes of non-informative sequences include:

- The human genome can be non-informative or at least minimally informative. These sequences are, of course, quite informative if human profiling primers are used but when this is not the case the human sequences tend to overwhelm the atnasmips profile. (See Methods section, "*Amplifying DNA from seizures: Making profiles*" and, in this case, their presence or absence is the *only* information they provide).
- There is a contribution from sequences generated by chaotic template-switching during extension of the primers and potentially also from short nucleic acid trash formed from breakdown of DNA.
- There are also the sequences derived from the reagents used for profiling such as DNA contamination from the organisms used to manufacture the reagent enzymes.
- Those sequences from post-seizure handling that have slipped past precautions against contamination.

The course of work during the development of array challenge analysis cannot be separated from careful consideration of the full range of non-informative sequences in-order to eliminate or, at least, account for their contribution to the final assessment. During the first part of this project, they had to be understood and, in the middle months of the project, this appreciation was then used to devise sentinel sequences to be added as spots to the array to monitor them and finally, near the project's end, the accumulated experience was beginning to be used to block them, as a class, out of the analysis.

In principle, non-informative sequences are always present but in practice they are only a serious issue in two circumstances: when the bulk of the DNA is human and/or there is very little DNA at all in the seizure at the time of seizure. Whatever the case, it is important to be able to objectively assess their contribution.

A first step in resolving these issues was to start with the residue after simple water extraction of seizures collected on microcon™ membranes as these appeared to preferentially extract the human contribution. This fractionation may reflect that the human DNA is in cell fragments whereas microbial DNA is inside a prokaryotic cell envelope.

The water-insoluble pellet, the residue from the seizure, was then extracted with alkali to extract sequences from micro-organisms, alive, dead or from their spores. Typically, the DNA levels in such extracts were very low and blocking was probably routinely required because, when samples have very low levels of DNA in the seizure, the non-informative sequences begin to dominate the profiles. (Although this is hindsight).

Assessment of blocking non-informative sequences

The process of blocking consists of adding a mixture of unlabelled sequences, selected to be similar to the non-informational DNA present in profiles, to the mixture of fluorescent labelled profiles used to challenge the array. The unlabelled sequences that are similar to non-informational DNA will reduce binding of the labelled non-informational sequences present in the profile.

Three blocking regimes were considered and compared. The first regime (regime 1) was a defined cocktail of cloned sequences known to be generated by non-informative processes. The clones were taken from the mixture GR170, discussed in the explanation of Figure 1.3 . The second (regime 2)

was the same cocktail at higher concentration. The third (regime 3) was a reagent "blank" that had only random hexamer present as template so the amplification could freely generate artefacts from them and pick up any sequences from the reagent enzymes. The arbitrary but objective method of assessing the three blocking regimes against each other was to use one pair of profiles derived from a seizure, NM3 and NM4, both of which appeared to have a mix of real sequences and non-informative and/or synthetic sequences within them. If any blocking-regime quenched all spots then there would be no informative sequences in a profile, which was not the case with NM3 and NM4. If a blocker was not having any effect, then the array spots would be unchanged and the rank-order of the intensity of array spots before and after blocking would be the same – a plot of rank order before and after blocking would give a straight line with a slope and a trend-line with an R^2 value of 1.0. Conversely if a blocker was working so well that it was completely reorganising the order of intensity of spots, then there would be no real trend-line and the R^2 value would be near zero and, objectively, this would be the best blocking regime of that group of blocking regimes as this is the best objectively measurable effect of blocking. Three experiments were performed with three blocking regimes that gave the following trend-line R^2 values – 0.3836, 0.1726 and 0.0207. The value nearest zero was 0.0207, given by random hexamer-based blocking. It should be noted that future use of arrays composed of oligonucleotides rather than long sequences isolated by PCR and cloning makes possible completely different and probably better ways of addressing the whole issue of non-informative sequences. This is discussed further below.

The first implication of these results is that while blocking can never reduce uninformative values to zero it can reduce them to a level at which they can be cautiously ignored and in critical cases a comparison of blocked and unblocked arrays can readily detect which spots have been insufficiently discounted by the blocking as the effects of blocking are exponential, following simple dilution arithmetic. A spot that shows a very big reduction in signal on blocking will be reduced further, albeit less, on further blocking – heading down to a theoretical "fully-blocked-limit" that can never actually be reached but can be calculated after observing two levels of blocking. Looking at the data below, it is restated that an R^2 value of zero does not *prove* blocking has been perfect but it does indicate that it has been quite substantial and that it is definitely more effective than blockings that give higher R^2 values.

For the blocking experiments using arrays containing spots of clones derived from blank amplifications such as GR170, the "type" example (see Figure 1.3 discussed above), it seems that the template switching products *derived from primer acting as template* are the main source of non-informative sequences. This conclusion is founded on the observation that the template-switching process is definitely *not* random. A study of sequences produced by it has shown that it is a relatively repeatable and predictable process at low template levels because then the template switching processes that cause most spurious product occur on a single, defined, sequence; the primer; and the process itself, in accord with this, has a few relatively well-defined outcomes. Any switching that occurs on other "real" templates merely concatenates template sequences and a moderate amount of template concatenation doesn't seriously interfere with analysis on arrays. Particularly note that we are never concerned with very rare unrepeatable template switching processes as they, by definition, only ever produce an insignificant amount of product. Hypothetically, what if template switching products were highly random? If template switching was very near random it would produce random sequences that would interact weakly with almost everything or nothing at all and in both cases be of no consequence on the array analysis. For our experiments using primer antiseq05, two of the template switching derived clones, sequence GR170.4 and GR170.8, make suitable "control" spots on arrays to detect this class of non-informative sequence in a profile.

Study of blocking out of non-informative sequences in two closely similar samples from the model-seizures (Figures 2.4 to 2.7).

Figures 2.4 to 2.7 compare the seizures NM3 - REC 06/A40153-7 bags 050564-71,72,73,74,75 methamphetamine (microcon™ alkali soluble pellet) antiseq05 (cy3) and NM4 - REC 06/A40153-8 bags 0503564-8.1,8.2,8.3,8.4,8.5 methamphetamine (microcon alkali soluble pellet) antiseq05 (cy5).

To quickly grasp the point of this series of images compare Figure 2.4 to Figure 2.7 with respect to two specific features: the position of spot GR170 (the smaller the signal from it, the better the data because, as discussed, GR170 is mainly non-informative sequences); and the scatter of spots around a notional straight line (the less the scatter, the better the data if, as appears to be the case, seizures NM3 and NM4 are essentially replicates).

The materials chosen for comparison are expected to give scatter-plots in which the position of spots deviate from a straight line primarily due to the presence of ***non***-informative sequences. This is because they have been chosen such that they are equivalent to DNA samples derived from a single seizure but are different preparations from the seizure. The resultant comparisons, are in accord with this, indicating that they were, indeed, essentially, one seizure. They were also very low in total DNA as evidenced from the prominence of non-informative DNA.

The various blocking regimes trialled are detailed in the figure captions.

The first point is that all the primary scatter-plots are more-or-less single lines. This is, of course, the result expected when both seizures are the same (i.e. a "match"). Note: one of these data sets has been used as an example previously (Figure 1.3) when discussing matches.

What was less obvious but very interesting was that, on blocking out non-informative sequences in three different ways, the results became quite different in detail but nevertheless all patterns *remained straight lines*. So both the non-informative material as well as the informative material behaved similarly, a reassuring result as it points to the repeatability of the behaviour both of blockers and the non-informative sequences. This is important because it indicates that blocking uninformative sequences is a viable strategy.

Figure 2.4. Unblocked comparison of NM3 and NM4. Scatter-plot comparing the DNA contents of methamphetamine seizure NM3 to that of methamphetamine seizure NM4. cy3 is the colour (green) that was used to label the sequences derived from the former and cy5 is the colour (red) used to label the sequences from the latter. The X axis scores red fluorescence in arbitrary units set by the voltage on the scanner's photomultiplier, and the Y axis scores green fluorescence in arbitrary units. Each spot on the graph represents one spot (a unique sequence or mixture of sequences) on the array that is responding to either or both of the challenges. The spot's X coordinate indicates the degree to which that sequence has found similar sequences in the red-labelled seizure and the spot's Y coordinate indicates the degree to which that sequence has found similar sequences in the green-labelled seizure. Scanning details and list of sources of spots are available from the raw-data file with serial number corresponding to this slide's serial number. The array, mainly composed of spots of single sequences. However, GR170 (see explanation above Figure 1.3 for GR170) is a whole profile. The X axis shows the amount of seizure NM4's profile in arbitrary fluorescence units that bound to that spot. The Y axis shows the amount of seizure NM3's profile in arbitrary fluorescence units also that bound to that spot. The two fluorescences can be resolved because the excitation wavelengths differ.

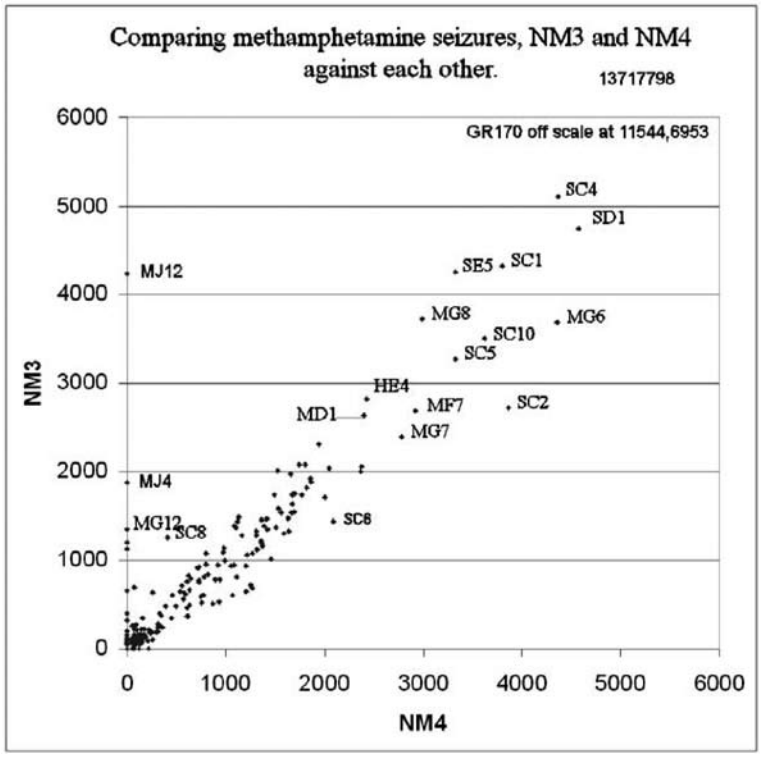


Figure 2.4

Figure 2.5 Comparison of NM3 and NM4 blocked at level 1(level 1=10 pmol GR170.4 (clone)-M13 amplicons, 10pmol GR170.5 (clone)-M13 amplicons, 16 pmol PCR 2.1(T-vector)-M13 amplicons plus 1 nmol antiseq05 primer. Details are otherwise the same as for Figure 2.4.

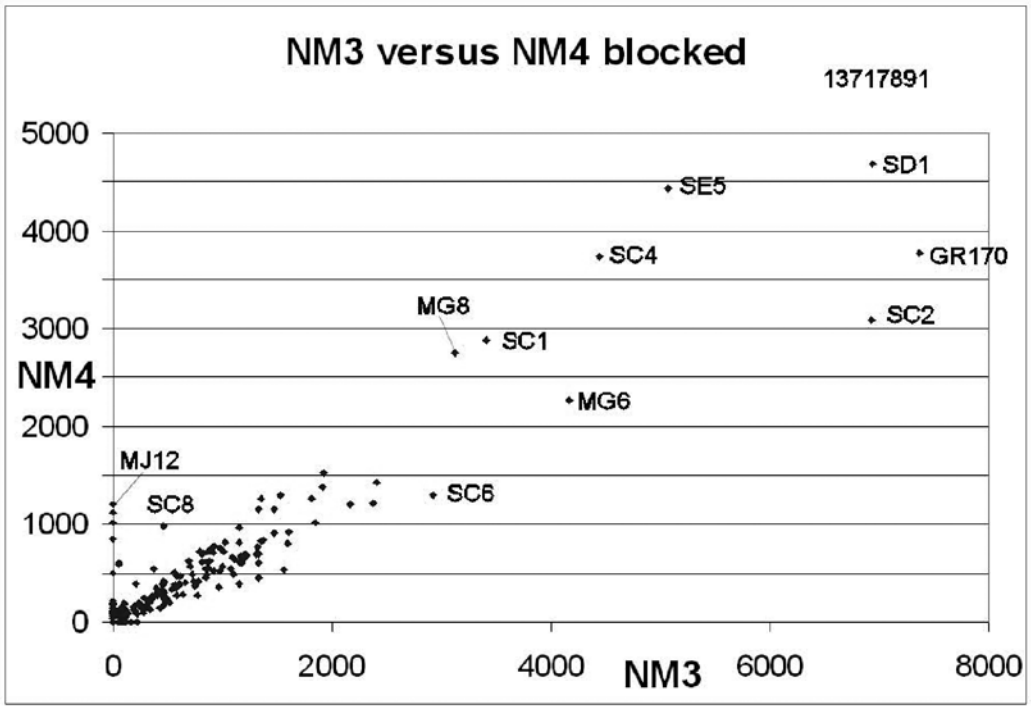


Figure 2.5

Figure 2.6 Comparison of NM3 and NM4 blocked at level 2 (level 2=10 pmol GR170 antiseq05 amplicons,10 pmol GR170.4 (clone)-M13 amplicons, 10 pmol GR170.5 (clone)-M13 amplicons, 16 pmol PCR 2.1(T-vector)-M13 amplicons plus 1 nmol antiseq05 primer). Details are otherwise the same as for Figure 2.4. Note, this is a differently annotated version of the data shown in Figure 1.3.

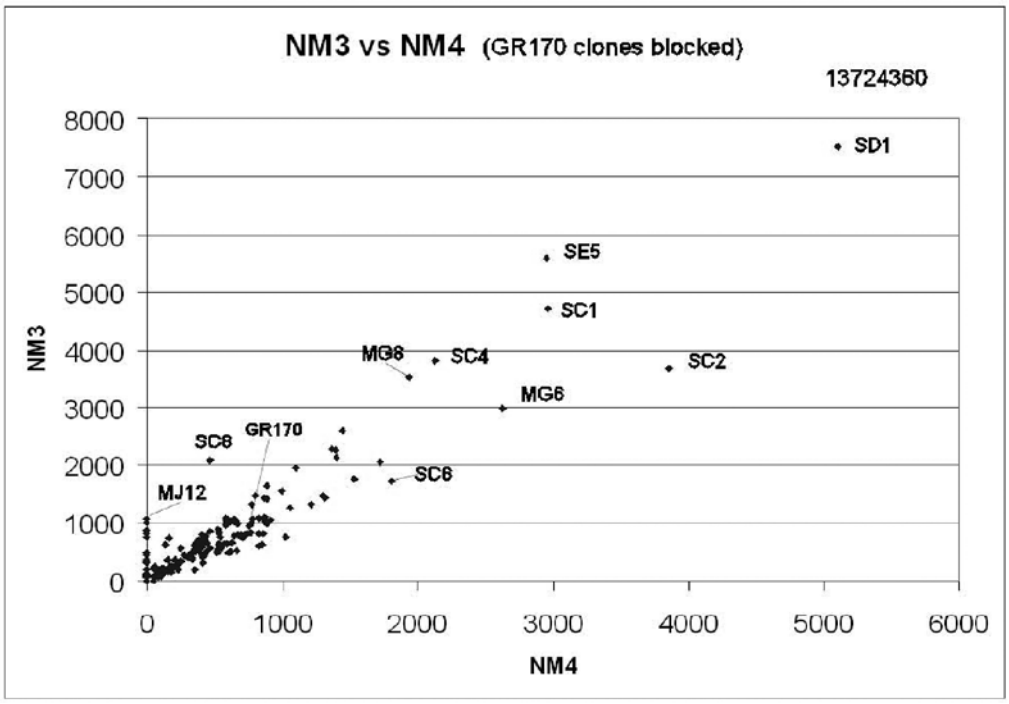


Figure 2.6

Figure 2.7 Comparison of NM3 and NM4 blocked with the products of a standard amplification using random hexamers as template in an otherwise "blank" amplification, using the usual, antiseq05 as primer. Details are otherwise the same as for Figure 2.4.

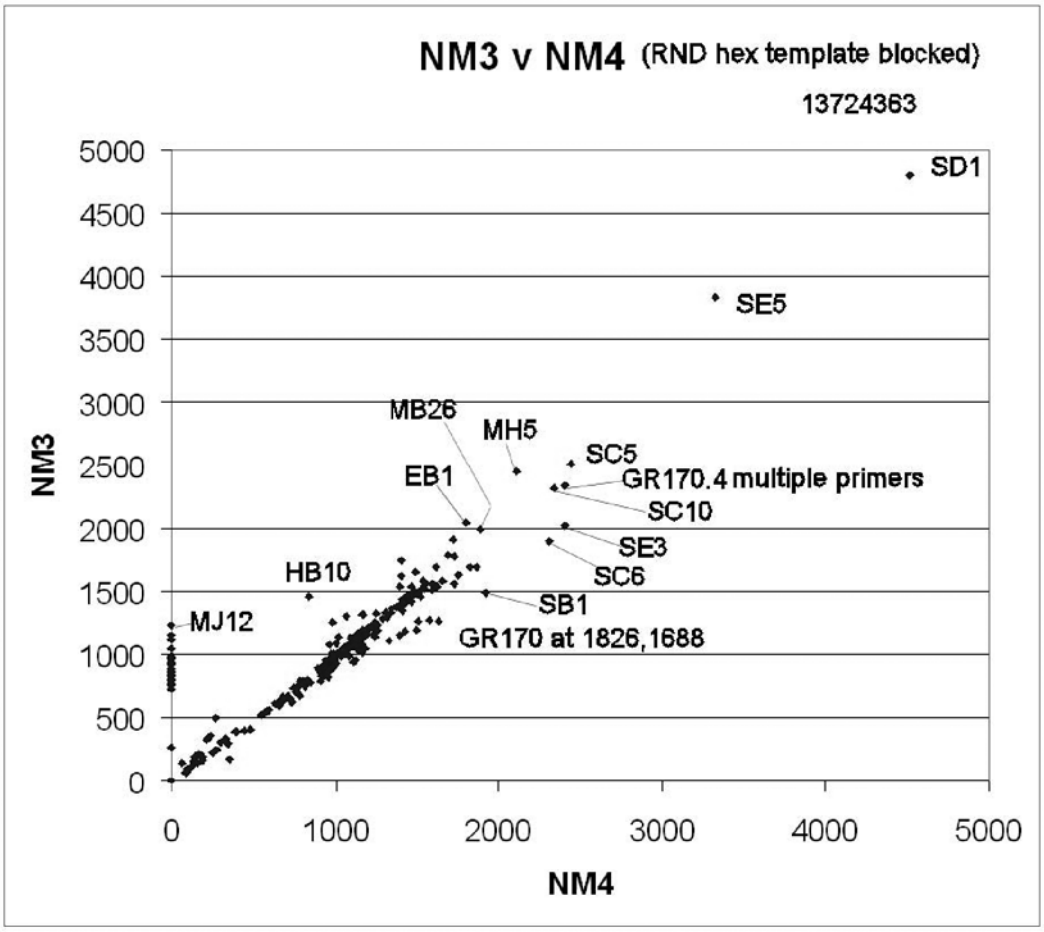


Figure 2.7

The effects of blocking on the rank-order of fluorescence of array spots

To further investigate the effects of blocking on the success of comparisons, the spots in each of Figures 2.4 to 2.7 were ordered with respect to their brightness. The colour red was used for this ranking. The near-linearity of the scatter-plots indicated that either colour would be suitable for ranking and yield similar results. Now, using the rank-orders as dimensions for the axis, each set of data from the un-blocked scatter-plots was plotted against the rank-orders in the three blocked scatter-plots giving rise to plots Figure 2.8 to Figure 2.10.

In summary, the use of random hexamer amplicons as blocking agents (Figure 2.10) was the most effective of this group of blocking treatments since it gave a rank order of spot intensity with the least similarity to the rank order when blocking was not used. This can be assessed by the R^2 value which if 1.0 would suggest no blocking whatsoever and, if zero, that blocking had the broadest range measurable but not that it was necessarily complete. Note that, as discussed above, theoretical blocking can never be physically complete but the use of blockers can allow calculation of what theoretically complete blocking would be.

Figure 2.8 (left) Rank order plot of the fluorescence of each array spot: first blocking regime (low amounts of GR170.4, GR170.5, PCR2.1 and antiseq05 primer) compared with fluorescence from the unblocked array for comparisons of NM3 and NM4. The rank-order was measured from the cy5 signal from Figure 2.5 versus spots on the unblocked array Figure 2.4. Both axes are rank-orders so there are no units.

Figure 2.9 (right) Rank order plot of the fluorescence of each array spot: second blocking regime (high amounts of GR170.4, GR170.5, PCR2.1 and antiseq05 primer) compared with fluorescence from the unblocked array for comparisons of NM3 and NM4. The rank-order was measured from the cy5 signal from Figure 2.6 versus spots on the unblocked array Figure 2.4. Both axes are rank-orders so there are no units.

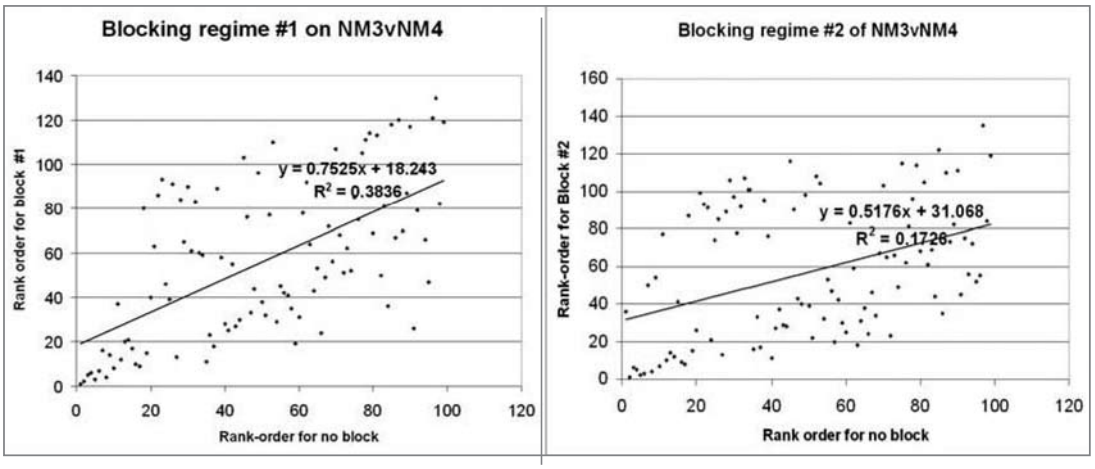


Figure 2.8 **Figure 2.9**

Figure 2.10 Rank order plot of the fluorescence of each array spot: third blocking regime (the products of a standard amplification using random hexamers as template for primer antiseq05.) compared with fluorescence from the unblocked array for comparisons of NM3 and NM4. The rank-order was measured from the cy5 signal from Figure 2.7 versus spots on the unblocked array Figure 2.4. Both axes are rank-orders so there are no units. This is the currently preferred blocking regime.

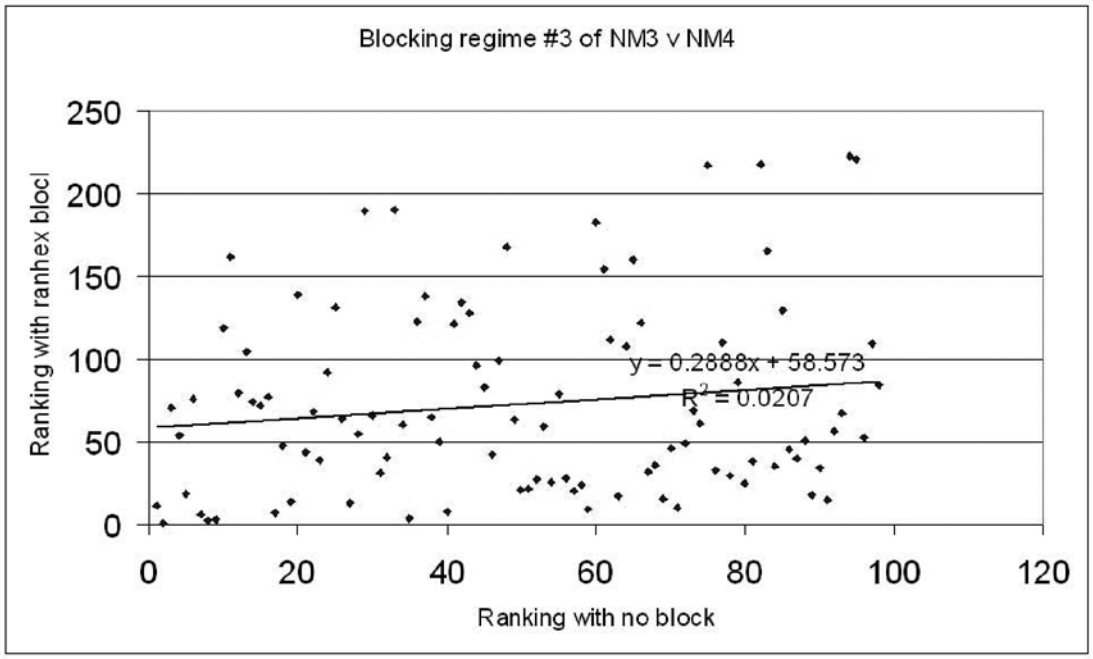


Figure 2.10

Conclusions on blocking

The first point is merely a reminder that none of the above applies to profiling human sequences from seizures using human primers and profiled conventionally. Blocking of non-informative sequences is only an issue when non-specific primers are used at low template levels and the products analysed on arrays, although this is a common reality for the mixed-origin DNA from seizures.

Although it is possible to objectively assess the effectiveness of blocking regimes by comparing blocked and unblocked arrays, no blocking regime can ever be declared perfect or complete. The most practical strategy is to use an obviously effective blocking regime such as the random hexamer template "blank" amplicons used for Figure 2.7 and also including suitable "sentinel" spots on the array to allow continual monitoring of the non-informative sequences. So, useful blocking regimes are available but whilst blocking is effective in making major reduction in the presence of uninformative sequences, it cannot, in principle, entirely eliminate them but can allow calculation of what idealised complete blocking would be like – although to do this calculation requires at least one extra probing to allow the calculations. In practice the use of sentinel spots and a single probing will always be a wise first precaution indicating the presence and approximate degree of the problem, and in the future, the use of oligonucleotide arrays should be designed so as to remove the whole issue or at least to minimise it to such a degree that it is utterly negligible. Beside this, we think the oligonucleotide option has much to recommend it.

Analysis of tablets – results from comparison of some seizures tablet by tablet.

Note: at this point in the project sample-processing moved over to the sodium carbonate-microcon™ method for tablets as described in the methods section.

All the examples discussed above were performed with powdered seizures rather than tablets. Water-soluble powders are the simplest to extract and, using water only, it is possible to get a significant degree of separation of human and non-human DNA. Additionally, seizures with an obviously high content of human DNA were put to one side.

This section is concerned with seizures in tablet form or in capsules and extracted by the sodium carbonate procedure with no attempt to separate human and non-human DNA content.

Three seizures of MDMA (ecstasy) were examined. Each seizure was composed of many tablets and/or capsules and each DNA preparation was made from a single tablet or capsule, except where otherwise indicated. The seizure numbers are, as provided from the law enforcement agency:

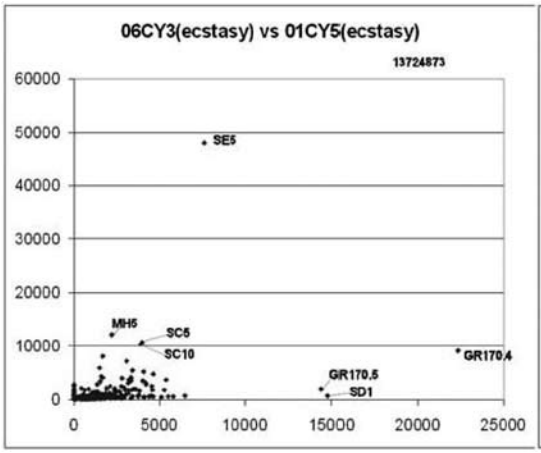
1. REC:06/AQ48078-2 smooth white ecstasy tablets
2. REC:014131 rough yellow ecstasy tablets, maple leaf logo
3. REC:A215200 clear ecstasy capsules containing powder.

DNA was prepared from these seizures by the dilute sodium carbonate method, profiles made and fluorescently labelled using antiseq05 as a primer as specified in the methods section. Pairs of profiles were compared to each other on arrays as specified in the example given in Appendix 5.

Twelve tablets or capsules from the above seizures were subjected to a pre-screen on an agarose gel to see if they had any obvious DNA content. Eight were selected and compared pairwise Figures 2.11 to 2.18.

Figure 2.11 (left) Tablet versus tablet comparison, MDMA seizures. Array 13724873. 06cy3 vs 01cy5, a tablet from seizure 1 versus a tablet from seizure 2.

Figure 2.12 (right) Tablet versus capsule comparison, MDMA seizures. Array 13724375. A2cy3 vs 06cy5, a capsule from seizure 3 versus a tablet from seizure 1. cy3=GREEN label. cy5=RED label. Scattergrams have the X axis red and the Y axis green.



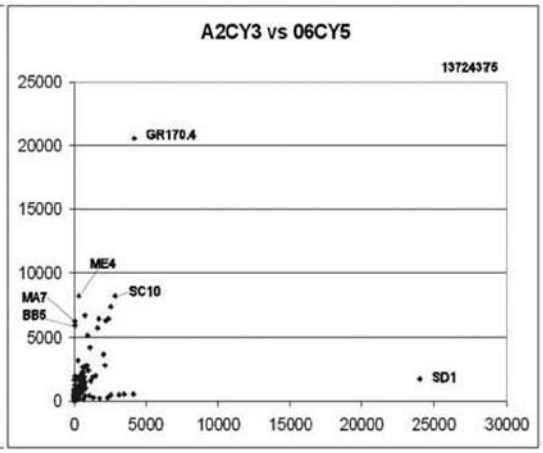
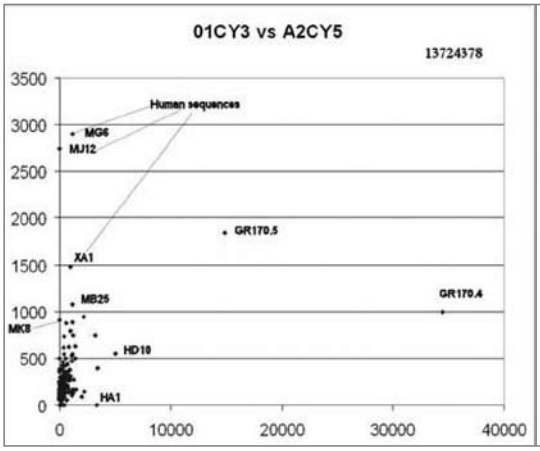


Figure 2.11 **Figure 2.12**

Figure 2.13 (left) Tablet versus capsule comparison, MDMA seizures. Array 13724378. 01cy3 vs A2cy5, a tablet from seizure 2 versus a capsule from seizure 3.

Figure 2.14 (right) Tablet versus tablet comparison, MDMA seizures. Array 13724379. 014cy3 vs 06Acy5 a tablet from seizure 2 versus a tablet from seizure 1. cy3=GREEN label. cy5=RED label. Scattergrams have the X axis red and the Y axis green.



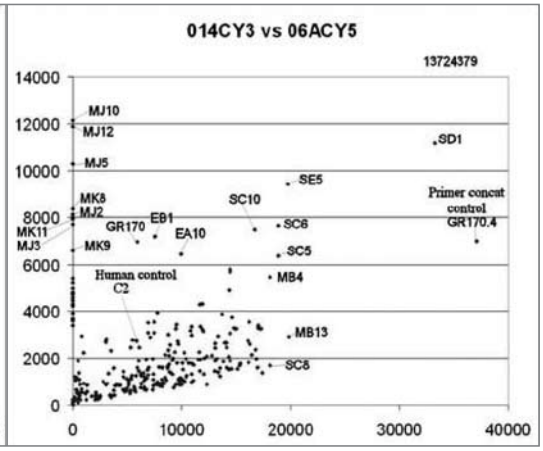


Figure 2.13 **Figure 2.14**

Summary of the array analysis of these seizures:

1. Array 13724873 (06cy3 vs 01cy5, Figure 2.11) shows negligible similarity between the tablets. From the coordinates of GR170.4, the tablet 01cy5 has less DNA in it than does 06cy3 and from the control spot C2 (not shown), they have barely detectable human DNA. Interestingly, they share strong reactivity to one sequence SE6 and some reactivity to SD1.
2. Array 13724375 (A2cy3 vs 06cy5, Figure 2.12) shows negligible similarity between the tablet and capsule. From the coordinates of GR170.4, A2cy3 has less DNA in it than does 06cy5. From the control spot C2 (not shown) there is barely detectable human DNA in either although possibly high enough in 06cy5 to consider an attempt to generate an STR profile. Tablet 06cy5 has a strong reactivity to one sequence: SD1.

3. Array 13724378 (01cy3 vs A2cy5, Figure 2.13) shows negligible similarity between the tablet and capsule. From the coordinates of GR170.4, A2cy5 has much less DNA in it than does 01cy3. From the control spot C2 (not shown) and other loci marked on the figure, there is barely detectable amounts of human DNA in tablet A2cy5 but there is a trace of human DNA in 01cy3, probably just worth considering attempting profiling for STR loci.
4. Array 13724379 (014cy3 vs 06Acy5, Figure 2.14). These tablets are similar in that they both contain human DNA. From the coordinates of GR170.4, 014cy3 has more DNA in it than does 06Acy5. Once again SD1 is common to both seizures.

Scattergram number 4 (Array 13724379) is probably the most DNA-rich comparison and thus the most information rich and will be discussed here in more detail than the others. The scattergram from Array 13724379 is re-presented here as Figure 2.15 along with three subsets of the data points. Figure 2.15 contains the whole, unsorted results for the comparison of two tablets 014cy3 and 06Acy5. Figure 2.16 is the same data set but with the human reactive spots and the spots that react to human commensals *Staphylococcus aureus* and *Propeonobacter acnes* removed from it. Superficially, removal of these spots makes little difference to the appearance of the scattergram, meaning that non-human sources are dominating the appearance scattergram in Figure 2.15. Although these non-human sources are in part related, they are not-identical and the tablets do not have the same DNA content.

Figure 2.15 (left) and **Figure 2.16** (right) **Tablet by tablet comparison of MDMA seizures 014cy3 vs 06Acy5 (Array 13724379).**

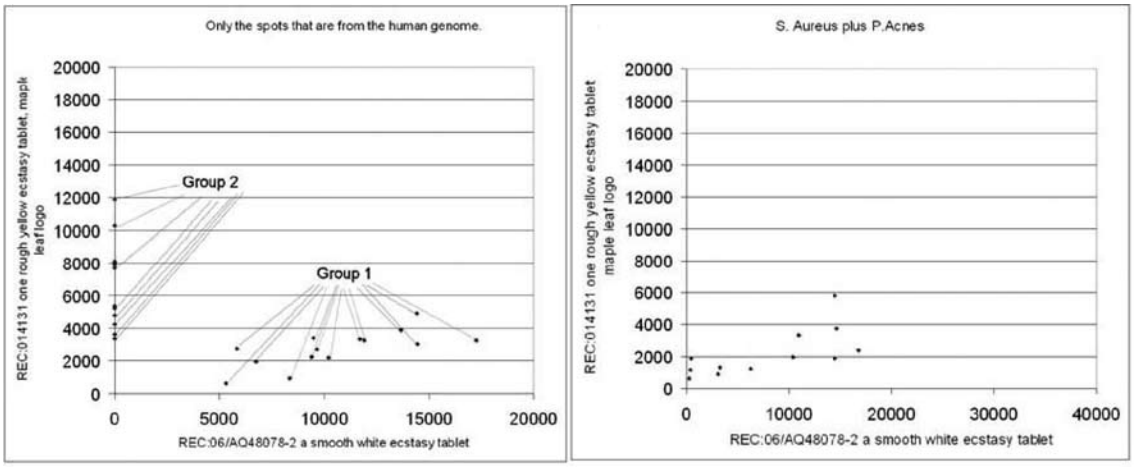


Figure 2.15 **Figure 2.16**

Figure 2.15 has the whole profile, Figure 2.16 has human and known human-related sequences removed.

The array spots with human sequences in them (Figure 2.17) comprise two groups. The group marked 2 only binds sequences from the tablet taken from seizure 2 but the other group binds human sequences from both seizures, although more noticeably from seizure 1 than seizure 2. The spot fluorescences have not been closely normalised so definitive statement about the relative binding of seizure 1 and 2 to the group 2 spots is not appropriate. However, spot patterns in which some human components are not represented in the profile is commonly observed where there are human signals in seizures. Circumstantial evidence indicates that this may be a useful in quantifying ultra-trace human DNA but there are also technical explanations for this concerning inappropriate behaviour of one of the dyes and this possibility is one extra motive for shifting to oligonucleotide arrays that should be much less subject to these possibilities.

Array spots with sequences from the human commensals S. aureus and P. acnes are modestly labelled by both seizures (Figure 2.18). These species may, ultimately, be targets of choice for provenance studies (see discussion).

Figure 2.17 (left) and **Figure 2.18** (right). **Tablet by tablet comparison of MDMA seizures 014cy3 vs 06Acy5 (Array 13724379).**

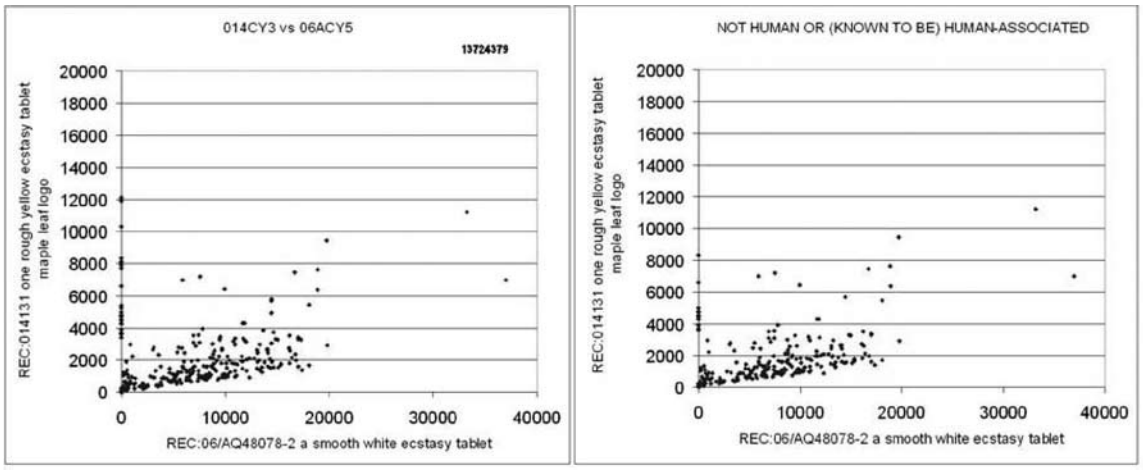


Figure 2.17 **Figure 2.18**

Figure 2.17 shows the human sequences only and Figure 2.18 shows only those sequences from *P. acnes* and *S. aureus*.

Sentinel spots, their values and meanings

These are labelled in Figure 2.14 which compares a tablet from seizure 2 with a tablet from seizure 1.

- C2 (a human control) has coordinates of 5845, 2747 and indicates a significant human DNA signal from both samples.
- GR170.4, coordinates 37044, 6987, is the concatenated primers signal. As usual for drug seizures, it is very strong. It comes from multiple priming sites and is roughly, inversely proportional to the amount of template at the beginning of the amplification.
- GR170.5 coordinates, 11955, 6968, is an unidentified sequence that is, nevertheless, common from seizures or the enzymes used to amplify them.
- SD1 is an unidentified sequence that has been extracted from drug seizures and soils and is common in our seizures. We do not know why. It is noteworthy that many sequences that are found in seizures were found in soil, these denoted by the letter "S", indicating the origin of the type-example.

Overall conclusions from the comparison Array 13724379

Although the two tablets 06Acy5 (tablet from seizure 1) and 014cy3 (tablet from seizure 2) share many sequences in common the DNA profiles cannot be regarded as even similar.

Some of the sequences clearly come from human contact (group 1 on the Figure 2.17) as they are human sequences and both seizures have a clear contribution to the human control (C2) and other human sequences are also labelled (see below for detail). However the group 2 spots on the Y axis are a notable feature of this particular array and, while some will be due to stochastic effects of low DNA causing some human sequence to be present in one sample and not another an examiniation of a number of arrays indicates there is probably also a technical reason due to the behaviour of one fluorophore cy3 discussed in the methods section. This technical reason should simply be avoided in the future by using oligonucleotide arrays (see discussion).

Chapter three: Discussion

Methodology for the extraction of DNA from drug seizures

The methods trialled had subtle advantages and disadvantages. For example, a pre-wash with organic solvents before water or other aqueous solvent extraction reduced the bulk of the seizures to convenient volumes but it took more time and, if the seizure was acidic, gave the acid longer to damage the DNA. The water-only method was persuasively simple but again gave no protection from acid. However, water extraction seemed to give quite a good separation of human and non-human DNA due to the probable but unproven surmise that the human DNA was not protected by encapsulation inside a tough cell wall as is the case for bacterial DNA.

The carbonate procedure was also simple and is probably the method of preference but basic carbonate, or tris, in aqueous solution or mixes with organic solvents can make some of the alkaloid components of many seizures insoluble so they generate a copious precipitate. Amphetamine, for example, is simply insoluble in our routine organic solvent, isopropanol. This is of no importance if one is trying to get the soluble human DNA but can cause considerable difficulty when seeking encapsulated bacterial DNA. Trade-offs are unavoidable.

Theoretical foundation for profiling DNA extracted from drug seizures

Precis: for detecting similarities and differences in DNA content that would allow gathering of intelligence regarding the provenance of and relatedness of drug seizures, the project chose to use a technique designed to meet forensic criteria, not biological goals and, most importantly, to give stable results at very high gain amplifications. The comparison method selected (2-D scattergrams) was chosen to be as simple as possible so it would be most readily interpretable by lay persons lacking specialist knowledge.

The issue of human DNA in seizures is a quite separate issue to the non-human sequences. Human DNA in seizures is a relatively familiar subject as it is really only a special case of the familiar issue of "trace human DNA profiles" which has been extensively reviewed and is currently the subject of considerable research elsewhere. However, this project has developed methods for extracting DNA from drug seizures which are expected to be of value in deriving human profiles.

Thus, without forgetting the human component that we have proposed to investigate in a spin-off project, the main theoretical base of the project is concerned with non-human DNA present in drugs.

The expectation that non-human DNA would be useful for profiling drug seizures arose from a PhD project in our laboratory conducted by James Waters whose thesis work is being prepared for publication. The project developed methods for profiling soil DNA using specially selected high gain amplifications with arrays used to compare the products.

The method was developed specifically to meet forensic criteria rather than the instantaneous state of the population of sequences present in the sample at the time of collection, as might be required by a biologist interested in the total range of species of microorganism present, their taxonomic relationships and their relative frequencies.

In contrast, from the forensic viewpoint, the profile obtained from a sample needs to not only be strongly characteristic of that sample but also minimally affected by small changes in the sample-handling and, above all, insensitive to long amplifications so that it is well adapted to small samples. The relative frequencies of sequences within a sample and their taxonomy are interesting but irrelevant for forensic purposes.

All profiling methods currently give a small suite of sequences that are *characteristic of* the full range of sequences present which in their totality are too complex to be simply compared. Most methods also take into account the relative frequencies of the sequences present *at least to some degree*.

However, a forensically-specialised method should:

- respond to the full menu of sequences present much more than the frequency of any specific sequence within the menu, as the latter is often highly dynamic, particularly during collection-handling and over time
- be useful with very low amounts of template
- use rather than be invalidated by the Darwinistic effects that are always present in amplifications that lead to some sequences being more readily amplified than others. These selective amplification effects can become completely dominating with trace evidence subject to long PCR amplifications
- have a very robust outcome.

On theoretical grounds, very long amplifications with single primers meet the first three criteria reasonably well and, from the experience gained during James Waters' PhD on soil DNA, the theoretical expectations were in accord with observations

Long amplifications with single primers resulted in profiles that reflected the menu of sequences present much more than the original relative frequency of sequences because, at the end of a long amplification, the replication efficiency of any molecule completely swamps out any differences in its starting frequencies. So, whatever were the original frequencies of the molecules amplified, the final frequencies do not reflect them. Put crudely, the most fit molecules predominate whether there was one or a thousand copies at the start of the selection process.

Very long amplifications become progressively less sensitive to the Darwinistic consequences of amplification between replicates because Darwinistic effects tend to level-out as many cycles of selection drives the differences in fitness (delta-fitness) between competing molecules towards zero.

A potential caveat to this expectation is that perhaps subtle variations in amplification conditions could so influence selection that the outcomes of the Darwinistic selection could be highly dependent on such random variations. This caveat was subjected to experimental testing by James Waters (PhD thesis, Flinders University 2004, Investigation of DNA profiling methods for forensic examination of soil evidence) and while slight variations in amplification conditions must, in theory, slightly affect the outcomes of the Darwinistic selection during amplification, this was found to be a minor source of experimental variation *so long as the number of cycles of amplification is large*. It was found that replications of amplifications tend to the same terminal state and this fulfils the fourth criterion: the outcomes are robust.

Thus for profiling DNA extracted from drug seizures, single primers were used with a total of 70 cycles of PCR in two stages.

The resultant products were analysed in three main ways:

1. by cloning fragments and sequencing them
2. by putting unsorted fragment mixtures on arrays and comparing them to other such mixtures (mixed sequences arrays)
3. by making arrays out of sequences cloned from such mixtures and then comparing other mixtures to these clones (single sequence arrays).

The main and immediately obvious method for using such arrays is to compare the profiles from two seizures by mixing them together but with each profile labelled with a different coloured fluorescent tag and determining the relative binding of the two profiles to the array. The results can be presented as a two dimensional scattergram with the intensity of each colour at any array element (spot) providing one of the two coordinates on the scattergram for that spot. One colour provides the spot's Y axis (usually green) and the other colour the spot's coordinate on the X axis (usually red). The shape of the scattergram then makes the degree of similarity of the two seizures visually obvious.

Comparing drug seizures using DNA arrays

Precis: expressing comparisons as two dimensional scattergrams is intuitively understandable with the caveat that, as DNA levels fall, artefacts and non-informative sequences must be increasingly taken into account by sentinel spots and, if indicated, by blocking them from binding to the array.

Excepting the human DNA issue, the crude-meaning of an array pattern derived from comparing two seizures was and is very easy to see visually; even for a lay person. Little training or technical knowledge is needed to assess patterns as matches or mismatches or something in-between.

However, the major technical caution arises from the non-informative sequences that have been described and studied. Once again, excepting the human component, these are from reagent contamination but can be totally synthetic, derived by template-switching and similar corrupt activities by enzymes with insufficient real DNA template to fully occupy their active centres. We spent much time studying the non-informative sequences and believe we have come to an adequate understanding of them and procedures for discounting their influence on a profile. This was not achieved until the last few months of the project. These understandings and procedures are described in the body of the report; however, they should be simply ignorable if this technique should shift over to synthetic oligonucleotide arrays as such arrays can be constructed so as to have no spots reactive with the non-informatives and, of course, a few that are specifically reactive to only them.

It is clear that arrays made from real DNA should always include sentinel spots made from blank amplifications in order to assess the contribution from these non-informative sequences and, if necessary, blank amplifications using random hexamer as template may be used to block the array if they are a serious problem. If the sentinel spots give a signal, non-informative sequences are present and blocking can then be used to annul them from the profile.

Both the problem of the template-switching component of the non-informative sequences and the solution to this problem lies in the interesting fact that template switching operating on excess primer molecules and a few random fragments of real DNA are in fact non-random. The template switching component is remarkably consistent in its overall sequence content, at least over a great many molecules. This means that when long runs of product are formed by template switching, these long runs are substantially homologous to other long runs produced by the same process, even months and years later. If the process wasn't very repeatable and was near-random, the product produced at one time would not significantly hybridise to the random product produced

later. As the former case is the observed situation, the interfering product of template switching can be blocked by just adding large amounts of the same material, albeit unlabelled, to the mix used to challenge an array. As noted, oligonucleotide arrays of the future will simply be designed to ignore the template-switching product and thus, simply, ignore this whole issue and this would be a highly desirable resolution of this issue.

Template switching, chaotic or not, is not a serious problem in the presence of ample real sequences as real, long sequences usually amplify better than template-switching products as template-switching products usually have many more ways they can form hairpin foldbacks that interfere with primer loading. This deduction is in accord with observation as the presence of "real" DNA markedly depresses the formation of template-switching products.

Practical issues concerning the human component versus the non-human component

Human DNA tends to be very easily extracted with water, probably because it is fragmented and skin fragments, unlike bacteria, have no tough coat. In contrast, bacterial DNA from live or dead cells tends to be found in the insoluble pellet left after extraction. This "non-human" pellet has interested us as it often seems almost free of human DNA but it poses another problem as it commonly contains tenacious PCR inhibitors reminiscent of those we found in LSD. We have dealt with this in drug seizures by adding bovine serum albumen (BSA) to amplifications but it would be desirable to avoid using BSA if possible since this is a potential source of further non-informative DNA. It should be noted that an apparently related problem occurs with soil DNA that is often attributed to humic acids but, as we see it in illicit seizures, we suspect that humic acid is not entirely the explanation.

The issue of provenance

Precis: provenance can be inferred by comparing samples of unknown provenance to those of known provenance but the method with greatest generality is likely to be the SNPs in the most common sequences.

Provenance information can be gained from some degree of a "match" between two seizures which if present implies relatedness of provenance. Thus what is known about one seizure can be implied to be also true about the other. This linking of sets of data in a pair-wise fashion is a common logic-strategy in investigative science. However, samples can share much similarity while being quite different in some detail. For example, an array spot may bind polymorphic forms of a sequence. In consequence, the most critical and difficult questions of provenance will probably come down to a sequence by sequence examination of a seizure where the role of arrays will be to point to the sequences present that are in high copy number and are worth examining further.

The extended examination will probably start with primer-pairs specially designed for those particular sequences and analyse the amplification products for single nucleotide polymorphisms (SNPs) in the same way that the variable region of the mitochondrion is currently analysed. The sequences that are likely to be of most practical utility will probably not be, as might be expected, those sequences peculiar to a district or place but will rather be those that are the most common in seizures around the world and we suggest that these may be the ones we have observed from human commensals found in seizures – *Propionibacterium acnes* and *Staphylococcus aureus*. Rare, locality-specific sequences will require massive data-bases to be useful. However, SNPs in common sequences like those of human commensals or perhaps soil transposons should allow

much more compact data-bases and allow much more rational analysis. There are multiple reasons to expect this. In the case of transposons, they are commonly inserted into genomes as both active and inactive fragments that then diverge or decay by apparently random mutation. In the case of human commensals, particularly skin bacteria, this decay and fixation might arise from selective pressures such as, for example, UV exposure and antibiotic exposure that would be different in different societies. Features such as neutral mutations at the third base of codons and lengths between ribosomal gene repeats would also diverge near-randomly. Broadly, these sorts of changes occur and locally fix quite readily and have been used to follow bacterial strain differences since the mid 1990s. (Originally Kostman et al. (1995) *J. Infect. Dis.* 171, 204-208). However, in all these and similar examples, there is always the reality of how the processes creating divergent sequences are counter-balanced by the opposing processes creating uniformity such as selection pressures and diffusion across geographic barriers. Although we have studied this in transposons in eukaryotic genomes (L.M. Smith and L.A. Burgoyne (2001), *Gene* 271, 273-283) it is recognised that the realities would almost certainly be different again in prokaryotic genomes so any project that planned to utilise geographic or ethnic polymorphisms of human commensals would first have to survey polymorphisms in search of those types of loci for which the position of balance of the opposing considerations gave polymorphisms with appropriate utility.

Perspectives and prospects for the future

- *With trace DNA, amplification will be necessary for the foreseeable future and we see no sign of any way non-amplification DNA technology can deliver useful outcomes.*
- *Analyses will probably be sequence-based; not length polymorphism-based.*
- *There will be a mixture of human DNA profiling together with objective analysis of non-human*
- *sequence populations with only minor concern for the species of origin.*
- *The order of development may next be through*
    - *synthetic oligonucleotide arrays (constructed after careful analysis of our or other data-bases)*
    - *then emulsion-amplification will have been assessed for its practical utility; probably indirectly by the groups studying metagenomes*
    - *then, when costs drop low enough, sequence mixtures will be given a preliminary assessment with arrays and subjected to high-throughput-sequencing and sophisticated software linked to large data-bases to allow objectively based inferences to be made.*
- *In the meantime, we believe the human content of drug seizures detected in this project can be developed for objective assessment by conventional analyses of trace DNA residues and the non-human DNA profiles can readily be analysed by the array technology, yielding more subjective information that should be useful in police intelligence for linking the source of drug seizures. This approach is entirely compatible with the likely future developments and represents a stage in their development.*

Chapter four: Methods: Recommended recipes and protocols section

Chemicals and reagents

Chemicals and reagents including enzymes were obtained from commercial sources and not listed unless some special reason or condition exists.

Water and plasticware cleanliness

All water was subjected to ozonizing-UV and ozone such that the water no longer contributes to the background amplification after the usual amplification for trace DNA. Water and plasticware was irradiated in such a way that it was well exposed to the extremely ozone-rich atmosphere produced during irradiation. (Simple but effective irradiation devices were constructed at Flinders University for this purpose. See Appendix 5).

Preparing DNA from seizures

For soluble DNA, the basic principles are simply to get rid of the drug and its diluents such as sucrose etc. with either or both organic solvent and water. The DNA is captured from aqueous solution on microcon™ membranes.

For DNA bound in particles of residues like dead bacteria, water extraction is preceded by an alkaline digest.

The early experiences in this operation were described in the six months report. Seizures with a high content of DNA can have useful preparations made from samples as small as 20 mg but much larger samples, 100 mg and over, are suggested as being most likely to be routinely useful.

Solvent extraction followed by water extraction procedure: in summary, for simply preparing soluble DNA from large amounts of seizures the procedure has satisfactorily and best met the main criteria as specified in the six months report and is useful as it stands. It is, however, only applicable to those seizures with components that are predominantly isopropanol soluble. An example:

1. Using the heroin seizures (S1, S2, S3, 010797) (20 mg per sample, each twice).
2. Add to sample an excess of 70% isopropanol water (1 ml for 20 mg) containing 15 mM tris free base, 0.5 mM EDTA acid. (Explanation: the use of isopropanol is to keep DNA insoluble during early washes, the EDTA is to remove multivalent metals that "mordant" DNA to partuclates, the tris is to raise the pH of the drug to levels that will not damage DNA as heroin is commonly acid).
3. Vortex at RT, 5 mins room temp, then centrifuge at 14K, 5′ room temp. Discard supernatant.
4. Wash pellet if visible, or wash apparently empty tube if pellet is not visible, with 70% isopropanol water (1.0 ml) to get rid of the remnants and the bulk of EDTA which is no longer wanted. Spin 14K, 5′ room temp. Discard supernatant.
5. Comment: all DNA and cells should be in the pellet. So far, none of the isopropanol supernatants are worth studying as they should have no DNA.
6. Now extract soluble DNA. Add water (500 ul) to pellet (500 ul is the volume that Microcon™ can comfortably take and reduce). The few drops of isopropanol residue can be ignored.
7. Vortex, shake or rotate for 30 minutes at RT to dissolve DNA, spin hard (14K, 5′ RT) to get supernatant with DNA. Any DNA remaining in the pellet should be predominantly within microbial or other cells with tough walls – store this pellet or the apparently empty tube for later study.

8. Continuing with the soluble DNA. Microcon™ the supernatants down to almost zero, (about 5 µl) by centrifuging at 2K to 2.5K for about 15'.
9. Wash the near-dry Microcon™ by adding another 100 µl water and spin again (centrifuge at 2.5K total time about 5') down to a small enough volume to just comfortably enable amplification of a pair of replicates from each of about 5 µl final vol.

Simpler water extraction procedure only (Water/Microcon™ Method). Although possibly not as stringent as the first procedure, a simpler procedure that appears to be adequate is as follows and has been the most common procedure used in the latter part of the project.

1. Dissolve drug sample in 0.55 to 1 mL forensic grade H_2O. Spin at 14K for 5 mins at room temp.
2. Transfer supernatant to a clean 1.5 mL tube leaving the pellet behind.
3. Dry the pellet at 60°C for 60 mins (will use pellet in step 1 below).
4. Transfer dissolved drug sample to Microcon™ filter/controls.
5. Spin 500 x g for 35 mins at room temperature.
6. Empty collection tube.
7. Repeat steps 4 to 6 until all of the supernatant has passed through filter.
8. Add 0.2 ml of forensic H_2O to filter.
9. Spin at 500 x g for 20 mins at room temp. and discard supernatant.
10. Repeat steps 8 and 9.
11. Add 10 µl of forensic H_2O to filter.
12. Flick a few times.
13. Invert filter in a clean collection tube.
14. Spin at 13 K for 1 min at room temperature.
15. Sample is in collection tube (approx 10 µl).

Alkali extraction of final pellet

1. Add 200 µl of lysing solution (0.3N NaOH, 0.5mM EDTA, 0.1% SDS) to the dried pellet from earlier.
2. Vortex/rack vigorously to resuspend pellet.
3. Incubate at 37°C for 30 minutes.
4. Centrifuge 14K for 10 minutes at room temp.
5. Transfer supernatant to a Microcon™ filter.
6. Dry the remaining pellet in oven 60°C for 60 mins (will use dry as template in PCR reaction).
7. Centrifuge Microcon™ filters 500xg for 35 mins at room temperature.
8. Discard flow through and wash with 200 µl of forensic H_2O.
9. Centrifuge 500 x g for 20 mins at room temperature.
10. Repeat wash steps 8 and 9 until flow through is clear and no foaming is evident if water is agitated by pumping with the pipette.
11. Invert filter in clean collection tube.
12. Add 10 µl forensic H_2O to filter and flick mix.
13. Centrifuge at 14 K for 1 min at room temperature.
14. DNA is in collection tube.

Technical note: these alkali extracts sometimes contained PCR inhibitors that failed to be removed by Microcon™ similarly to soil extracts.

Extraction of DNA from tablets

Step 1 – dissolution of tablets (ecstasy in the example) in UV-ozonised carbonate solution.

1. Add 1mL of 150 mM sodium carbonate to each tablet (~300 mg) in a 1.5mL microfuge tube.
2. Vortex vigorously at 10 minute intervals for a total of 30 minutes.
3. Centrifuge 14,000rpm/15 minutes to pellet any undissolved material.
4. Transfer supernatant by pipette to a clean 1.5 mL microfuge tube, avoiding the pellet at the bottom of tube and any floating material.
5. Centrifuge again 14,000rpm/5 minutes to pellet any material that was brought over in the previous step.
6. Transfer supernatant to a clean 1.5 mL microfuge tube.
7. At this point the supernatant from 2 tablets from matching batches are pooled into a single 1.5mL tube before processing the supernatant using our standard Microcon™ DNA extraction technique (see below).

Microcon™ processing of tablets extract.

1. Transfer dissolved component of drug sample (the supernatant) to Microcon™ filter/controls.
2. Spin 500 x g for 35 mins at room temperature.
3. Empty collection tube.
4. Repeat steps 1 to 3 until all of the supernatant has passed through filter.
5. Add 0.2 ml of forensic H_2O to filter.
6. Spin at 500 x g for 20 mins at room temp. and discard supernatant.
7. Repeat steps 5 and 6 twice.
8. Add 10 µl of forensic H_2O to filter.
9. Flick a few times.
10. Invert filter in a clean collection tube.
11. Spin at 14,000rpm/1 min. at room temperature.
12. Sample is in collection tube (approx 10 µl).

Amplifying DNA from seizures: Making profiles

The primers – Single primers at high amplification factors, as used in previous soil studies, and refered to as atna-smips primers, were used. The primers sequences are listed below. The selectivity and peculiarities of these primers depends on many factors but an extremely strong factor in their behaviour appears to be simple Darwinian competition during the actual amplification rather than the primer-template homology. Selection of this type means they tend to have the advantage of pseudo-randomly sampling sequences from the whole range of template sequences present rather than just the ones present in highest frequency. An, arguable, disadvantage of this is that genomes with high complexity such as the human genome may overwhelm amplifications and almost eliminate the contribution of less diverse microbial genomes.

Primer list

Sequence Name: Cpali20mer2G	Sequence: GGAGGTGGGCCGGGTGGAGG
Sequence Name: seq 05	Sequence CCCTCGAACACCACCTCC
Sequence Name: antiseq 05	Sequence GGAGGTGGTGTTCGAGGG

PCR mix

For 1–5 µl of DNA solution.

Buffer (10x) 5 µl
$MgCl_2$ (25 mM) 4 µl
dNTPs (50x) 1 µl
Primer (10 M) 3 µl
Red hot polymerase 0.25 µl
Forensic H_2O 35.75 µl (or volume to make 50 µl)
1–5 µl of template solution
TOTAL 50 µl

Cycle conditions. *The stage marked with an asterisk uses the high stringency conditions appropriate for the particular primer.*

This is the "usual" cyling conditions to be taken as the case unless otherwise stated.

1st round PCR

94ºC – 5 min

- 94 ºC – 30 sec
- 30 ºC – 3 min 2 cycles
- 72 ºC – 3 min

- 94 ºC – 30 sec
- 62 ºC – 30 sec* 35 cycles
- 72 ºC – 3 min

72 ºC – 7 min

4 ºC – ∞

2nd round PCR

94ºC – 5min

- 94 ºC – 30 sec
- 62 ºC – 30 sec * 35 cycles
- 72 ºC – 3 min

72 ºC – 7 min

4 ºC – ∞

Making arrays

The DNA samples are laid out, each in their own well in a 384-welled plate in a defined pattern.

The arrays are on glass slides with surfaces prepared to bind DNA applied to them. A programmed spotter dips a pen into each well and applies the pen to a defined position on the glass surface. The spots are then "fixed" on the surface by drying and UV crosslinking.

Preparation PCR reaction products for spotting arrays

1. To a 100 µl PCR reaction, add 10 µl of 3M NaAc pH5.2 and 220 µl of ETOH (99%).
2. Mix briefly and ice for 1 hour.
3. Spin at 14K for 15 mins.
4. A spirate supernatant.
5. Re-suspend in 200 µl of forensic H_2O.
6. Add 20 µl of 3M NaAc pH5.2 and 440 µl of ETOH (99%) to each sample.
7. Mix briefly and ice for 1 hour.
8. Spin at 14K for 15 mins.
9. Aspirate supernatant.
10. Add 200 µl of 70% ETOH and flick mix well or vortex briefly.
11. Spin for 10 mins at 14K.
12. Aspirate supernatant.
13. Re-suspend pellet in 20 µl of Pronto spotting solution.
14. Transfer 10 µl of DNA/spotting solution to 384 well plate.
15. Store plate in -20ºC freezer when not in use.

Spotting

- Spotting was with a quill-type pin. Spots are approximately 140 microns diameter, and spacing, spot centre to spot centre, is 400 microns.
- Spotter is from Biorad. 'BioOdyssey Calligrapher miniarrayer'.
- The array printing software is 'BioOdyssey Calligrapher'.
- The slides are Corning's 'UltraGAPS coated slides' (amino-silane coated).
- Spotting is in Corning's 'Pronto Universal Spotting Solution'.
- Fix spots by 4 hours drying in a vacuum at room temperature with ten seconds exposure to UV in the biorad "genelinker" machine. Approx 20mJoules.

Scanning

- The scanner is 'Genepix 4000A' with 'Genepix Pro version 1.4'.
- Photomultiplier voltages used ranged from 500 to 900 volts, chosen such that no spots are driven to their maximum intensity.
- The colours red and green are "false" colours and actually represent the colours of the excitation wavelengths, 645 mu and 532 mu.

Array challenging

Challenge preparation – fluorescence labelling profiles with cy3 or cy5 deoxyribonucleotide triphosphate.

The labelled nucleotides are from GE healthcare Lifesciences and supplied by the manufacturer as 1 mM solutions (1nmol/µl).

cy3-dCTP product number PA53021

cy5-dCTP product number PA55021

Caution: cy3 can oxidise and this can cause inappropriate, non-specific, binding to some spots so it is stored with antioxidants (e.g. a trace of hydroxyl disulphide).

PCR set-up	*Per 50 µl reaction mix*
Buffer *(10x)*	5 µl
$MgCl_2$ *(25mM)*	4 µl
BSA *(10mg/mL)*	0.5 µl
dGTP *(10mM)*	1 µl
dATP *(10mM)*	1 µl
dTTP *(10mM)*	1 µl
dCTP *(10mM)*	0.4 µl
Primer *(10mM)*	3 µl
Red hot pol.	0.3 µl
Forensic H_2O	30.80 µl

Template: 1 µl template DNA (usually from standard two stage amplification as above)

Label: 2 µl cy3 or cy5.

PCR Cycle conditions

94ºC – 5min

- 94 ºC – 30 sec
- 62 ºC – 30 sec 35 cycles
- 72 ºC – 30 min

72 ºC – 7 min

4 ºC – ∞

Column purify probe reactions using 'Ad-biotec' PCR clean-up kit.

- elute DNA in 50 µl fH_2O

Pre-hybridization of arrays

Prepare hybridization solutions while pre-hybridizing.

1. Warm pre-hybridization solution (5 x SSC, 0.1% SDS, 0.1 mg/ml BSA) to 42°C.
2. Immerse slides in pre-hyb solution (in petri dishes) and incubate at 42°C with gentle rocking for 45 mins.
3. Transfer to 0.1 x SSC and incubate at 25°C for 5 mins (repeat).
4. Transfer to H_2O for 30 secs.
5. Dry by centrifuging at 1,500 rpm for 5 mins (place slide in a falcon tube to centrifuge).

Hybridization

1. Wash cover glass with nuclease free water, followed by ethanol. Dry in Laminar.
2. Making challenge solution: to 50 µl (column purified) of fluorescence labeling reaction mix add:
   a. 250µl formamide
   b. 62.5µl of 20 x SSC
   c. 50µl of 1% SDS
   d. 87.5µl fH_2O.
3. Transfer 50 µl of each of each challenge solution to a clean 1.5 ml screw-top tube (this is enough for one slide). *Add any blocking DNA (see just below) at this point*.
4. Heat the solution (hyb buffer/challenge-profile) to 100°C for 10 mins then place on ice for 2 mins.
5. Pipette challenge solution along edge of slide and bring the cover glass down to the slide allowing their edges to touch. Drop cover glass onto the slide in a manner which traps and spreads the solution between the slide and cover slip.
6. Add forensic H_2O to hybridization chamber to keep humidity up, seal the array in the hybridization chamber.
7. Place the hybridization chamber in a sealed plastic container with some wet paper towels in the bottom.
8. Keep the hybridization chamber upright and place in a hyb incubator at 58°C for 12 – 16 hours.

Technical warning and note: the green (cy3) fluorophore is seriously prone to oxidation or some similar change during the hybridisation and this can generate a variety of artefacts; mainly high backgrounds for green but also some particular spots seem to bind the oxidation products selectively (egour spot MJ12). The spots showing this anomaly very strongly have been noted and they are probably overloaded with very long sequences. Care to use fresh fluorophore and exclude air and prevent drying during probing is essential. In the future the use of oligonucleotide-arrays rather than the long sequences we have sometimes used is expected to remove or solidly reduce this problem.

Blocking

To make blocking sequences, a standard two stage PCR is carried out (see above "Amplifying DNA from seizures: Making profiles"– using the desired source(s) (see just below) of non-informative sequences as template.

1. A cloned sequence containing non-informative sequences such as one cloned from profile GR170, or
2. random hexamer as template (20 ng per 50µl amplification mix) or
3. no template, except those in the reagents.

After the second stage of amplification or, almost equivalent, after a third stage, the total product of non-informative sequences captured is precipitated by standard sodium acetate/ethanol technique and the dry pellet or pellets is added to the challenge mix at item 3 in the section above marked "hybridisation".

Post-hybridization

** Minimize exposure to light as much as is practical at this stage. e.g. turn off overhead laboratory lights and cover petri dishes with black plastic when incubating.

1. Dismantle the chamber and remove the array.
2. Immerse array in 2 x SSC, 0.1% SDS at 58°C and remove cover glass from the slide immediately but gently.
3. Transfer array to 2 x SSC, 0.1% SDS at 58°C for 5 mins.
4. Transfer to 0.1 x SSC, 0.1% SDS at 25°C for 5 mins (repeat step).
5. Transfer to 0.1 x SSC at 25°C for 1 min (repeat step 4 x).
6. Rinse in 0.01 x SSC for 10 secs.
7. Dry by centrifugation at 1,500 x g for 2 mins.
8. Store in dark container until ready to scan.

Processing the results

Scanning raw data. The scanner had its own software that scanned the spots and the areas immediately around them and presented these raw figures as an Excel document. This proprietary software was 'Genepix Pro version 1.4', as noted above, and is not available for display by us.

Converting the raw data into scatterplots: Software was written by us "in house" to extract the core information. See Appendix 2 for explanation and print-out of this software.

Chapter five: Appendices

Appendix 1: Bioprofiling of drugs project – tabulated DNA results

We thank Ms Karen Lee and Katrin Both for processing these profiles.

Sample	Seizure details	D3 S1358	vWA	FGA	D8 S1179	D21 S11	D18 S51	D5 S818	D13 S317	D7 S820	Amelogenin
1A	Meth 0401925 31.10.06	15,NR,17, 18	10,15, NR,17,NR	NR,NR	8,10,13, 15,NR	-	-	11,12	-	-	X,Y
1B	Meth 0401925 31.10.06	15,16,17, 18	15,16,17, 18,19	20,21, 23,24	8,10,12, 13,15,16	30	16,17	9,11, 12	NR,8,11, 12,13,14	NR,NR, NR	X,Y
2A	MDMA 014131 31.10.06	-	-	-	-	-	-	-	-	-	X
2B	MDMA 014131 31.10.06	-	18,19	-	-	-	-	-	-	-	-
3A	PMA 4215200 31.10.06	15,16,NR	15,16,17, 18	21,22	-	29,30,32.2, 33.2	-	10,11, 12,13	-	NR	X
3B	PMA 4215200 31.10.06	14	NR,19	-	7,12,13	-	-	-	11	8	X
4A	Meth 0502425-13 31.10.06	-	-	24	-	-	-	-	-	-	X
4B	Meth 0502425-13 31.10.06	-	-	-	-	-	-	-	-	-	-

KEY

- No result at this locus

NR: Weak result detected below the laboratory's reportable threshold of 50RFU

Appendix 2

Data-processing software

Program1 = simplecondenser9 – converts raw data in comma-delimited form to get a scatterplot. It takes in a text file, comma delimited generated from the raw Excel doc that the scanner generates.

Edit all the text data and heading data, leaving ONLY the columns of figures. Then save this stripped file as comma delimited text and give it a name like R.txt.

It intakes, say R.txt which has many lines, that are each a text version of an Xcel line. (Typically they are too long to display and example here.) This then outputs a working file, giving it the temporary name REDANDGREEN.TXT which looks like this below example of a few lines.

Example of an actual file in text format:

NAME	Colour code	RED	sd	N	GREEN	sd	N	Comment
SA10	undefined	0	0	6	0	0	6	Insig fig in BOTH
HA1	undefined	0	0	6	195	5	6	Insig fig in RED
HA2	undefined	51	16	6	165	6	6	
HA3	undefined	411	7	6	0	0	6	Insig fig in GREEN
HA4	undefined	0	0	6	135	8	6	Insig fig in RED
HA5	undefined	0	0	6	0	0	6	Insig fig in BOTH
HA6	undefined	145	20	6	0	0	6	Insig fig in GREEN
HA7	undefined	0	0	6	161	21	6	Insig fig in RED

These files can be plotted for a “first plot” RED versus GREEN using the Excel plotting facility as the data is tab-delimited and Excel will take it up readily. Then use the two-way scatter plot option of Excel.

The annotation of some spots names can be aided with software that produces a clickable map. It can also be readily done by hand.

The code (example below) is complied in Powerbasic™

```
'SIMPLECONDENSER9.BAS
GLOBAL MONOCOLOURARRAY$()
GLOBAL BOTHCOLOURSARRAY$()
GLOBAL INARRAY$
GLOBAL LINEOFTEXT$
GLOBAL COLUMNCOUNTER&
GLOBAL LINENAMES$()
GLOBAL MEANANDSDDATA#()
GLOBAL LINEDATA#()
GLOBAL ALLLINES$()
GLOBAL OUTITEM$()
GLOBAL FILENNAME$
GLOBAL COLUMNMARKER&
GLOBAL KILLALL$
GLOBAL N&:' THE NAME COUNTER
GLOBAL CURRENTNAME$
GLOBAL NMAX&
GLOBAL LINESMAX&
GLOBAL OUTTEXT$
GLOBAL COUNTALLINES&
GLOBAL TEXTCURRENTCOLUMNCONTENTS$
GLOBAL COLUMNDATA$
GLOBAL COLUMNVALUE&
GLOBAL INSIGNIFICANCE#
GLOBAL ROOTNAMEOCCURENCES#
GLOBAL LOWERLIM$
GLOBAL CGC&:'only values are 1 and 2 (red and green) and also points to the first cood of
MEANANDSDDATA#(2,1000,3)
GLOBAL KKK&
GLOBAL SPILLOVERTEXT$
GLOBAL IDCOLOUR$()


FUNCTION PBMAIN
   DIM LINENAMES$(4000)
   DIM MEANANDSDDATA#(2,3000,3):'coord place 1 is red or green, coord place 2 is for the
   names, coord place 3 is mean,SD,N
   DIM LINEDATA#(4000,3)
   DIM ALLLINES$(4000)
   DIM BOTHCOLOURSARRAY$(4000,2)
   DIM MONOCOLOURARRAY$(4000,2)
   DIM IDCOLOUR$(4000)
```

```
    NMAX&=3999:LINESMAX&=3999
    DIM OUTITEM$(6)

RESET KILLALL$:RESET OUTTEXT$
' INSIGNIFICANCE# IS THE MINIMUM AVERAGE FLUORESECE VALUE OF A SET THAT GETS ALL THE SET KILLED.
OUTTEXT$="USE COMMA DELIMTED TEXT FROM XL- Any set of spots with a mean value of the replicates below a settable value has the whole set just nulled..OK to escape."
MSGBOX OUTTEXT$:RESET OUTTEXT$
'OUTTEXT$="What is the red to green spillover, as a fraction, for the voltage you are using- for 700V it is as below"
'SPILLOVERTEXT$=INPUTBOX$(OUTTEXT$,, "0.001")
RESET N&
OUTTEXT$="enter the plain text, comma delimited, input array's name or accept default name."
INARRAY$ = INPUTBOX$(OUTTEXT$,, "R.TXT")
OPEN INARRAY$ FOR INPUT AS #1
OUTTEXT$="Now enter the lower cut-off level of fluorescence - 0 will give some bleeding strange results - I suggest an absolute minimum of 50."
LOWERLIM$= INPUTBOX$(OUTTEXT$,, "50")
INSIGNIFICANCE#=VAL(LOWERLIM$)
CALL INTAKEALLDATAANDGENERATENAMELIST:' intakes crude comma delimited text list from excel and abstracts sample names.
' info --columns 41 and 42 (column markers 40 and 41) are just before the RED and GREEN means.
CALL PULLOUTREDANDGREENMEANSASTEXT
' CALL CORRECTFORSPILLOVER
CGC&=1:' (CGC&=1 indicates RED data being abstracted.)
CALL FILLMONOCOLOURARRAY: 'puts names and the associated and pointed column's data into rows into MONOCOLOURARRAY$ dim'd (3000,2)
CALL PROCESSCOLUMNFORMEANS: ' for the pointed columns, this gets the means for a particular name
CALL PROCESSCOLUMNFORSDVALUES: 'mean is in MEANANDSDDATA#(CGC&,K&,1), N is put into MEANANDSDDATA#(CGC&,K&,3), sd into MEANANDSDDATA#(CGC&,K&,2)=SD#
FILENNAME$="REDANDGREEN.TXT"
RESET MONOCOLOURARRAY$()
CGC&=2:' (CGC&=2 indicates GREEN data being abstracted
CALL FILLMONOCOLOURARRAY
CALL PROCESSCOLUMNFORMEANS
CALL PROCESSCOLUMNFORSDVALUES
CALL MONITOR
EXITALL:
MSGBOX "Ended - hit OK to escape."
END FUNCTION
```

```
SUB CORRECTFORSPILLOVER
   LOCAL JJ&
   LOCAL SPILL##
   LOCAL REDD##
   LOCAL GREENN##,DEDUCT##
FOR JJ&=1 TO COUNTALLINES&:' FOR ALL LINES WHATSOEVER IN ALLLINES$()
SPILL##=VAL(SPILLOVERTEXT$)
REDD##=VAL(BOTHCOLOURSARRAY$(JJ&,1)):'RED MEANS
DEDUCT##=SPILL##*REDD##
GREENN##=VAL(BOTHCOLOURSARRAY$(JJ&,2)):'get GREEN MEANS
'CORRECT GREEN FOR RED SPILLOVER
GREENN##=GREENN##-DEDUCT##
BOTHCOLOURSARRAY$(JJ&,2)=STR$(GREENN##)
NEXT JJ&
   END SUB


SUB INTAKEALLDATAANDGENERATENAMELIST
   ' Sores a text line, abstracts the name out of the line and creates a list of names. each name is
allowed into the list once only.
   LINEINPUTTING:
   CALL LINEOFINPUT:IF KILLALL$<>"" THEN GOTO BYPASSNAMELISTGENERATOR
   CALL SAVEWHOLELINEINARRAY:IF KILLALL$<>"" THEN GOTO
BYPASSNAMELISTGENERATOR
   CALL GETANAME
   CALL GETIDCOLOUR:' the IDCOLOUR for that name. (The IDCOLOUR and name should be
locked together )
   CALL ADDTONAMELIST
   GOTO LINEINPUTTING:' Do it all again
   BYPASSNAMELISTGENERATOR:
END SUB


SUB GETIDCOLOUR
LOCAL JJ&
COLUMNMARKER&=4:' POINTER TO THE LEFT-MARKER OF THE DESIRED COLUMN (IN THIS
CASE COLUMN 5 OR "E")  TO BE INTAKEN.)
FOR JJ&=1 TO COUNTALLINES&:' FOR ALL LINES WHATSOEVER
LINEOFTEXT$=ALLLINES$(JJ&)
CALL GETACOLUMNSTEXT
IDCOLOUR$(JJ&)=COLUMNDATA$
NEXT JJ&
END SUB


SUB PULLOUTREDANDGREENMEANSASTEXT
```

```
LOCAL JJ&
FOR JJ&=1 TO COUNTALLINES&:' FOR ALL LINES WHATSOEVER IN ALLLINES$()
LINEOFTEXT$=ALLLINES$(JJ&)
COLUMNMARKER&=40
CALL GETACOLUMNSTEXT
BOTHCOLOURSARRAY$(JJ&,1)=COLUMNDATA$:'get RED MEANS
COLUMNMARKER&=41
CALL GETACOLUMNSTEXT
BOTHCOLOURSARRAY$(JJ&,2)=COLUMNDATA$:'get GREEN MEANS
NEXT JJ&
END SUB

SUB FILLMONOCOLOURARRAY:' puts names and the associated pointed columns data into rows in the array MONOCOLOURARRAY$(
LOCAL JJ&
FOR JJ&=1 TO COUNTALLINES&:' FOR ALL LINES WHATSOEVER
LINEOFTEXT$=ALLLINES$(JJ&)
CALL GETANAME
MONOCOLOURARRAY$(JJ&,1)=CURRENTNAME$
CALL GETACOLUMNSTEXT
MONOCOLOURARRAY$(JJ&,2)=BOTHCOLOURSARRAY$(JJ&,CGC&):'CGC& is 1 for the red mean and 2 for the green
NEXT JJ&
END SUB

SUB LINEOFINPUT
IF EOF(1)=0 THEN LINE INPUT #1,LINEOFTEXT$:GOTO CORRECTENDLINEOFINPUT
KILLALL$="KILL":CLOSE#1
CORRECTENDLINEOFINPUT:
END SUB

SUB SAVEWHOLELINEINARRAY
   INCR COUNTALLINES&:IF COUNTALLINES&=>LINESMAX& THEN KILLALL$="KILL":GOTO EXITSAVEWHOLELINEINARRAY:'array overrun protection
   ALLLINES$(COUNTALLINES&)=LINEOFTEXT$
   EXITSAVEWHOLELINEINARRAY:
END SUB
```

```
SUB GETANAME:'Its after the third COMMA.This is REALLY just a special case of
GETACOLUMNSTEXT
LOCAL LENGTHLINE&,J&,COMMACOUNTER&
LOCAL C$
LENGTHLINE&=LEN(LINEOFTEXT$)
RESET CURRENTNAME$
FOR J& = 1 TO LENGTHLINE&
C$=MID$(LINEOFTEXT$,J&,1)
IF COMMACOUNTER&=3 AND ASC(C$)<>44 THEN CURRENTNAME$=CURRENTNAME$+C$
IF ASC(C$)= 44 THEN INCR COMMACOUNTER&
IF COMMACOUNTER&>3 THEN EXIT
NEXT J&
END SUB

SUB ADDTONAMELIST
   LOCAL FLAGG$
   LOCAL J&
   RESET FLAGG$
   IF N&=0 THEN N&=1:LINENAMES$(N&)=CURRENTNAME$:GOTO
EXITENTERCURRENTNAMEINTOLIST
   IF N&=>NMAX& THEN KILLALL$="KILL":GOTO EXITENTERCURRENTNAMEINTOLIST:' AN
OVERUN SAFETY LINE TO BE REMOVED
   FOR J&=1 TO N&
   IF LINENAMES$(J&)=CURRENTNAME$ THEN FLAGG$="ITS AN OLD NAME"
   NEXT J&
   IF FLAGG$ ="" AND N&<NMAX& THEN INCR N&:LINENAMES$(N&)=CURRENTNAME$:
   EXITENTERCURRENTNAMEINTOLIST:
END SUB


SUB GETACOLUMNSTEXT:'processing lineoftext$ GET WHATEEVER MARKEDCOLUMN VALUE
AS TEXT
        LOCAL LENGTHLINE&,J&,COMMACOUNTER&
        LOCAL C$
        LENGTHLINE&=LEN(LINEOFTEXT$)
        RESET COLUMNDATA$:RESET COMMACOUNTER&
        FOR J& = 1 TO LENGTHLINE&
        C$=MID$(LINEOFTEXT$,J&,1)
        IF COMMACOUNTER&=COLUMNMARKER& AND ASC(C$)<>44 THEN
COLUMNDATA$=COLUMNDATA$+C$
        IF ASC(C$)= 44 THEN INCR COMMACOUNTER&:'ASC(C$)= 44 IS A COMMA
        IF COMMACOUNTER&>COLUMNMARKER& THEN EXIT
        NEXT J&
END SUB

SUB PROCESSCOLUMNFORMEANS: ' for the pointed columns, this gets the means for a
```

```
partciular name
LOCAL J&,NAMEOCCURENCES&,K&
LOCAL ACCUMULATOR&
LOCAL MEAN#,TEMPREAL1#,TEMPREAL2#
FOR K&=1 TO N&:'FOR ALL NAMES
RESET NAMEOCCURENCES&
RESET ACCUMULATOR&:RESET MEAN#:RESET TEMPREAL1#:RESET TEMPREAL2#
FOR J&=1 TO COUNTALLINES&:'GETS MEANS AND N
IF MONOCOLOURARRAY$(J&,1)=LINENAMES$(K&)THEN ACCUMULATOR&=ACCUMULATOR
&+VAL(MONOCOLOURARRAY$(J&,2)):INCR NAMEOCCURENCES&
NEXT J&
' GET MEAN AND NUMBER OF CASES
TEMPREAL1#=ACCUMULATOR&:TEMPREAL2#=NAMEOCCURENCES&
MEAN#=TEMPREAL1#/TEMPREAL2#:' ie ACCUMULATOR&/NAMEOCCURENCES&
MEANANDSDDATA#(CGC&,K&,1)=MEAN#:MEANANDSDDATA#(CGC&,K&,3)=NAMEOCCURE
NCES&:'mean is MEANANDSDDATA#(CGC&,K&,1), N is MEANANDSDDATA#(CGC&,K&,3), sd
is MEANANDSDDATA#(CGC&,K&,2)=SD#
IF MEAN#<=INSIGNIFICANCE# THEN MEANANDSDDATA#(CGC&,K&,1)=(-100000):' Killing a
whole set by putting minus 100000 as its mean
NEXT K&
END SUB

SUB PROCESSCOLUMNFORSDVALUES
    ' gets the standard deviation of the mean
  LOCAL LOCALNAME$
  LOCAL J&,NAMEOCCURENCES&,ITEMSCHECK&
  LOCAL ACCUMULATORSQUARES##
  LOCAL MEAN##,TEMPREAL1##,TEMPREAL2##,CURRVAL##,DIFSQ##,VAR##
  LOCAL SD#
  FOR K&=1 TO N&:'FOR ALL NAMES
  MEAN##=MEANANDSDDATA#(CGC&,K&,1)
  RESET MEANANDSDDATA#(CGC&,K&,2):IF MEAN##< -1000 THEN GOTO
IGNORINGTHISNAME
  LOCALNAME$=LINENAMES$(K&)
  RESET ACCUMULATORSQUARES##
  RESET ITEMSCHECK&
  FOR J&=1 TO COUNTALLINES&
  CURRVAL##=VAL(MONOCOLOURARRAY$(J&,2)):DIFSQ##=((CURRVAL##-MEAN##)^2.0)
  IF MONOCOLOURARRAY$(J&,1)=LOCALNAME$ THEN ACCUMULATORSQUARES##=AC
CUMULATORSQUARES##+DIFSQ##
  NEXT J&
  ' GET MEAN AND NUMBER OF CASES
  NAMEOCCURENCES&=INT(MEANANDSDDATA#(CGC&,K&,3))
```

```
    TEMPREAL1##=ACCUMULATORSQUARES##:TEMPREAL2##=NAMEOCCURENCES&
    VAR##=TEMPREAL1##/TEMPREAL2##:' ie ACCUMULATOR&/NAMEOCCURENCES&
=VARIANCE
    SD#=(VAR##^0.5):' now convert VARIANCE to SD of the mean
ROOTNAMEOCCURENCES#=NAMEOCCURENCES&:ROOTNAMEOCCURENCES#=
(ROOTNAMEOCCURENCES#^0.5)
    SD#=SD#/ROOTNAMEOCCURENCES#:'the SD of the mean= SD/(Root of the number
    of items)
    SD#=SD#/MEANANDSDDATA#(CGC&,K&,1):'Expressed as a fraction of mean
    SD#=INT(SD#*100.0):'Expressed as a percentage of the mean
    RESET MEANANDSDDATA#(CGC&,K&,2):IF MEAN##> -1000 THEN MEANANDSDDATA#
(CGC&,K&,2)=SD#:' Set SD to zero if that mean has been killed.
    IGNORINGTHISNAME:
    NEXT K&
    END SUB


SUB MONITOR
LOCAL PLACEINROW&,DIGIVERS&
LOCAL LINOFOUTPUT$,COMMENTER$
OPEN FILENNAME$ FOR OUTPUT AS #3
PRINT#3,"NAME";CHR$(9);"Colour code";CHR$(9);"RED";CHR$(9);"sd%mean";CHR$(9);"N";
CHR$(9);"GREEN";CHR$(9);"sd%mean";CHR$(9);"N";CHR$(9);"Comment"

FOR J&=1 TO N&
LINOFOUTPUT$=LINENAMES$(J&)+CHR$(9)
IDCOLOUR$(J&)="undefined"
LINOFOUTPUT$=LINOFOUTPUT$+IDCOLOUR$(J&)+CHR$(9)
FOR PLACEINROW&=1 TO 3
CGC&=1
DIGIVERS&=INT(MEANANDSDDATA#(CGC&,J&,PLACEINROW&))
IF DIGIVERS&<0 THEN DIGIVERS&=0:COMMENTER$="Insig fig in RED"
OUTITEM$(PLACEINROW&)=STR$(DIGIVERS&)
LINOFOUTPUT$=LINOFOUTPUT$+OUTITEM$(PLACEINROW&)+CHR$(9)
NEXT PLACEINROW&
'
```

```
FOR PLACEINROW&=1 TO 3
CGC&=2
DIGIVERS&=INT (MEANANDSDDATA#(CGC&,J&,PLACEINROW&))
IF DIGIVERS&<0 AND COMMENTER$="" THEN DIGIVERS&=0:COMMENTER$="Insig fig in
GREEN"
IF DIGIVERS&<0 AND COMMENTER$="Insig fig in RED" THEN DIGIVERS&=0:COMMENTER
$="Insig fig in BOTH"
OUTITEM$(PLACEINROW&)=STR$(DIGIVERS&)
LINOFOUTPUT$=LINOFOUTPUT$+OUTITEM$(PLACEINROW&)+CHR$(9)
NEXT PLACEINROW&
LINOFOUTPUT$=LINOFOUTPUT$+COMMENTER$
PRINT#3,LINOFOUTPUT$:RESET LINOFOUTPUT$:RESET COMMENTER$
NEXT J&
CLOSE#3
END SUB
```

Appendix 3

Uv ozone steriliser for destroying DNA contamination

These devices were built to our specifications in our workshop and designed by a member of the workshop staff, Mr Douglas Lloyd Butler. As the radiation is dangerous to the eyes the device had to be equipped with secure safety switches that detect the devices opening and break the circuit to the tubes.

The devices are, essentially, a stainless steel bowl with a cover containing two overhead tubes that emit ozonising UV. Tubes: Type number G4T15H (single socket, 4W 185 nm enhanced (ozonizing) envelope).

Tubes made to size by – Australian Ultraviolet Supplies
23 Northgate Drive
Thomastown 3074

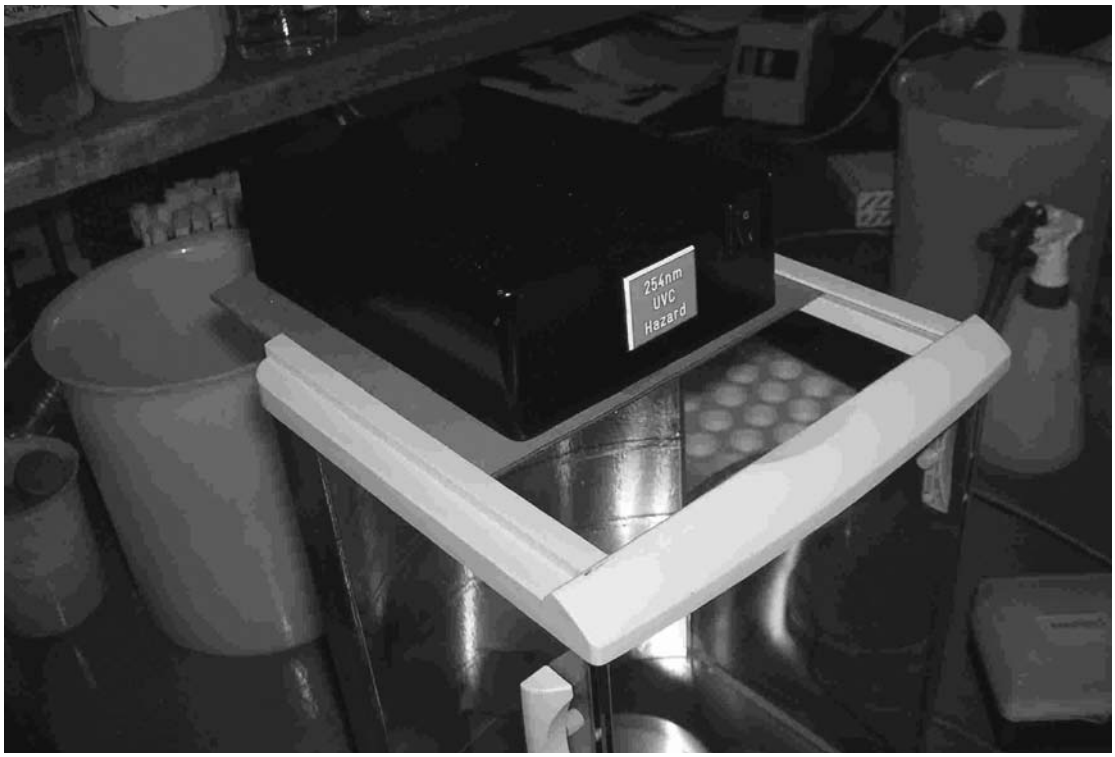
Items of plasticware were exposed to the UV light and the ozone generated by them. Distilled water was exposed directly to both the radiation and the gas in open containers, usually open petri dishes, and then stored in sealed and UV-ozone treated vessels.

The *camera angle distorts the appearance of the relative size of the sections*. The upper lamp housing with on/off switches is smaller than the lower section, which is a 5 litre stainless steel bowl and is in reality larger than the upper section.

Device 1: The steriliser – exterior

Device 1: The steriliser – interior

Device 2: The balance for weighing drugs – exterior



Device 2: The balance for weighing drugs – its UV mounting inverted and exposed