Background documents for presentation at:
IBRO NEUROSCIENCE SCHOOL: Elba Workshop and Summer School
May 4 - 10 Marina di Campo, Elba, Italy
Neurobiology of stress in health and disease; Org: E. Ronald de Kloet (Leiden University)

ELBA SUMMER SCHOOL

The Eurosterone DNA Array

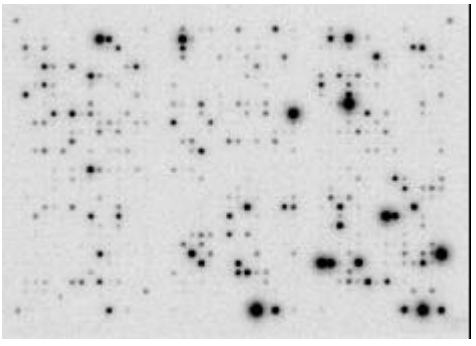
Richard Lathe, University of Edinburgh

Background documents

1. Mirnics, K. (2001) Microarrays in brain research; the good, the bad and the ugly. Nature Reiews Neuroscience 2, 445-447

2. Lathe, R. (1999) Mathematical analysis of ligation reactions and hybridisation parameters. Hons. Neuroscience and Pharmacology, University of Edinburgh.

3. Rose, K., Mason, J.O. and Lathe, R. (2002) Hybridization parameters revisited: solutions containing SDS. BioTechniques, in press.



OPINION

Microarrays in brain research: the good, the bad and the ugly

Károly Mirnics

Making sense of microarray data is a complex process, in which the interpretation of findings will depend on the overall experimental design and judgement of the investigator performing the analysis. As a result, differences in tissue harvesting, microarray types, sample labelling and data analysis procedures make *post hoc* sharing of microarray data a great challenge. To ensure rapid and meaningful data exchange, we need to create some order out of the existing chaos. In these groundbreaking microarray standardization and data sharing efforts, NIH agencies should take a leading role

Microarray data obtained in neuroscience experiments is very different from that obtained in cancer expression profiling owing to its extensive overlap with experimental noise (FIG. 1)[1–3]. At the tissue level, expression changes often do not exceed a two-fold difference between the experimental and control samples[4–8], resulting in a substantial overlap between real expression differences and assay noise. Because of this noise and the cellular heterogeneity of brain tissue, few successful microarray expression profiling studies have been made so far. However, despite the technical difficulties, careful experimental design, combined with extensive data verification and conservative interpretation of data, have yielded a number of groundbreaking brain microarray expression studies[4,8–19].

Data-analysis considerations

Noise of the assay. It is essential to ¶assess assay noise and to establish reliable cutoffs using sound statistical methods. This approach has a proven track record in science; there is no reason to treat microarray data differently. When using complementary DNA (cDNA) arrays, the hybridization of two aliquots of the same sample onto one microarray allows the magnitude and distribution of assay noise to be determined. Using a new sample isolation (and new reverse transcription), this hybridization should be repeated as many times as possible[3,20]. These controls will reveal the combined noise of the microarray assay due to isolation, labelling, hybridization printing, scanning and image-analysis variation. Identification of the exact source of noise in each category, unless excessive, is not particularly critical. However, it is essential to identify probes corresponding to individual genes that tend to be 'noisy' or show preferential labelling with one of the fluorochromes.

Experiment replication. In repeated experiments, always reverse the fluorescent labels between the experimental and control sample[9]. This will help obviate labelling bias. Spot replicates on the array increase the reliability of the fluorescent intensity measurement within an experiment, but they do not eliminate the need of replicated experiments[21], as they do not increase measurement reliability between the experiments. Is it necessary to always repeat the same sample hybridization several times? If interested in changes that are robust, this may not be necessary[4,10]. If assessment of *de novo* induction of gene expression with a drug treatment is required, resources may be better applied by comparing ten treated mice to ten matched controls than comparing three mouse pairs and repeating each hybridization four times. The critical issue is how the data will be analysed. While it is likely that few observations will report false-positive or -negative expression differences across the ten individual comparisons, it is almost certain that a consistently changed gene expression pattern across the majority of the comparisons will be due to biological differences, and not experimental artefact. A relatively simple calculation of probability, taking into account noise cutoffs, number of expressed genes, number of experiments and frequency of changed observations will reliably separate expression differences due to a biological phenomenon from assay noise.

Reliability cutoffs. Once the noise estimate of the system has been established, one can calculate reliability cutoffs for the individual data point observations. These cutoffs will be different for single and repeated experimental comparisons. The greater the number of times the hybridization is repeated (each time starting with a new isolation), the more reliable the gene expression measurement becomes[20,21]. In our system, a single microarray measurement is likely to represent a biological difference between samples at the 99% confidence level if it shows a >1.9-fold difference between the fluorescent intensities. However, 1.3–1.4-fold changes in gene expression might be reliably detected across repeated experiments that involve the same samples[3,20]. There are arguments against reporting the 'most changed' 1% or 5% of genes. Such a measurement neither accounts for the assay noise, nor does it take into account that the biological variability between the samples depends on the nature of the comparison. For example, a comparison between the same cortical region of two adult mice might reveal only a few true expression differences (<0.5%), whereas a brain–liver microarray comparison is likely to report >20% of differentially expressed genes. In both of these experiments, it would be a mistake to use a measurement based on the ordering of a fixed percentage of the most changed genes. For a graphical representation of a different example, see FIG. 1.

Assessment of complex expression differences. This is the most complex part of data-mining. Many excellent expression data analysis methods have evolved or have been adopted for use in microarray analysis, including hierarchical[22] and K-mean clustering, expression-tree harvesting[23], gene shaving[24], self-organizing maps[25,26], gene group analysis[4], factor analysis, support vector machines[27] and principal-component analysis (PCA)[28]. However, the biological question that is being asked and prior knowledge about the biological changes should serve as guides in choosing the data-analysis method. Understanding the strengths and weaknesses of each method is essential — inappropriate data analysis methods will generate 'ordered noise'. For example, hierarchical clustering will find clusters even in randomly generated data points or in data where the samples are unrelated. Furthermore, it might cluster together transcripts that are expressed in different cell types. Similarly, classifying the genes by response into 20 preset groups using K-mean clustering is questionable, unless the investigator has prior reason to expect a preset number of clusters. PCA assessments are relatively easy to perform, they explain the mathematical variability in the data set very well, but relating this variability to biological phenomena represents a complex challenge.

Data verification. If the project is focused on a small group of changes, microarray gene expression data should be verified, preferably using a method that will be informative about the anatomical localization of the transcript (for example, *in situ* hybridization)[29]. This eliminates potential harvesting artefacts and discerns a uniform transcript decrease from a selective messenger RNA (mRNA) loss in a subpopulation of the cells.

For verification of complex gene expression patterns, I recommend a three-step procedure[29]. First, in the initial microarray data-mining, a gene-expression pattern is defined. This is followed by the generation of a second microarray data set using a new group of subjects with the same condition. So, the gene expression pattern discovered in the initial data set is verified on the second data set. Finally, selected single-gene observations, critical for the interpretation of data, are verified by an independent method (for example, *in situ* hybridization).

Data-sharing considerations

As we are breaking new ground in sharing complex data sets, we have to make sure that the guidelines of sharing are based on sound scientific and moral foundations. Sharing inadequately controlled and unreliable data will generate poor results and false conclusions. Mining, sharing and comparing microarray data requires that the investigators who will use the shared data possess an in-depth understanding of the technology and its limitations. Without it, even excellent primary data sets will be compromised and will generate poor secondary analysis and false conclusions.

Microarray data is not readily comparable. In the context of investigating human brain diseases, microarray data is not readily comparable. The list of potentially confounding factors is long, even when using the same microarray probe set. In the sample-preparation phase, many factors will contribute to the final transcriptome variability. They include *post mortem* interval, time of tissue harvest, method of tissue dissection, and nucleic acid isolation procedures. In the labelling phase, choice of the fluorescent dyes, batch of dyes, method of labelling (direct versus indirect) and choice of the reverse transcription primers all might influence the final microarray results. Arraying surface (poly-L-lysine versus aminosylane versus aldehyde), different array preprocessing methods, or batches of fluorescent dyes will also affect fluorescent probe intensities. Similarly, differences in sample hybridization buffers, and temperature or stringency of washes will undoubtedly affect

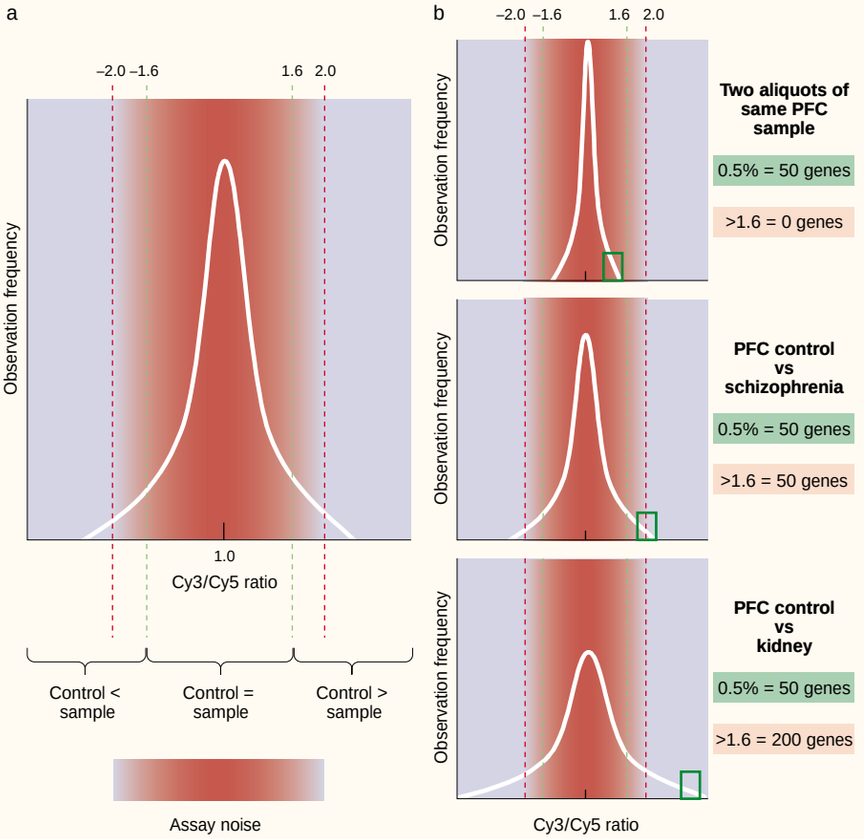


Figure 1 | **Distribution of microarray data in a typical dual-fluorescence cDNA array experiment.** **a** | X axis represents expression difference (Cy3/Cy5 ratio), where the '1.0' bin represents similar standardized fluorescent intensity between the schizophrenia and control samples. Y axis shows the frequency of the probes for each Cy3/Cy5 ratio in the experiment. Note that a significant portion of the obtained data (white line) overlaps with the combined microarray assay noise (red shaded area). In a well-controlled experiment in our system, a 1.6-fold Cy3/Cy5 or Cy5/Cy3 ratio probably represents a true expression difference (green dashed line), whereas the 2.0-fold difference in a single probe's fluorochrome intensity represents a true expression difference in >99% of the observations (red dashed line). **b** | Distribution of data from the same type of the cDNA microarray (n=10,000 genes) is plotted from three different microarray experiments. In the top panel, the same sample was hybridized against itself, with no gene reporting a differential expression over 1.6-fold. In the middle panel, a control sample from prefrontal cortex (PFC) was hybridized against a PFC from a subject with schizophrenia. At the 1.6-fold Cy3/Cy5 signal difference, about 50 genes reported an increased expression (and approximately the same number of genes reported a decreased expression). In the bottom panel, a kidney sample was hybridized against a PFC sample. Here we would expect that >400 genes would be reported differentially expressed over (>200) or below (>200) the 1.6-fold expression cutoff. However, if we would accept the top 0.5% of the 'most changed' genes (green box area) as 'true expression differences', we would report the same number of the most changed genes (50 increased + 50 decreased) in both cases. This would add noise (but not real data) to the top panel, and discard important biological differences from the bottom panel.

microarray signal. The choice of arrayer and microarray scanner, the analysis software and the normalization method will influence the readout of gene expression. Last, results are often extremely difficult to compare across different experimental designs[3] (for example, paired versus pooled comparison), in which different data-mining strategies must be used.

Of course, cross-platform data sharing is even more prone to so-called 'apples and oranges' comparisons. Regardless of what we share and how we might implement such sharing, the essential question is always the same: if one laboratory reports a 3.2-fold decrease in gene expression as a result of a given drug treatment, and a different group reports a 2.3-fold decrease as a result of a second treatment, can we assume that the first treatment had a greater effect? Noting the confounding factors listed above, the most appropriate comparison might be a binary one as shown by Lewohl *et al.*[8], assessing whether the obtained changes were statistically significant within both data sets (for example, change present versus change absent). In current data-sharing efforts, speculation

about the magnitude of the changes might be unreasonable at the present time.

Sharing optical images or analysed expression data. With the difficulties of data comparison in mind, what should be a meaningful format for sharing the data sets that were used to produce a publication? Greyscale images, obtained by a microarray scanner, come closest to raw data, although even their information content depends on nucleic-acid isolation procedures, dye quality, reverse transcription, arraying surface, hybridization, washing steps, and scanner type and setting. The size of a single microarray image may exceed 100 megabytes and image analysis is a complex procedure. However, this procedure allows the investigator to gain the experience that is required for correct interpretation of data. Indeed, analyzing the data from 'unusable' experiments owing to less-than-perfect labelling reactions, poorly performing microarrays, nonlinear signal adjustments and other technical pitfalls have been important for facilitating our own data interpretation.

Unfortunately, microarray optical images are not practical to share, and many investigators are hesitant to go through the full procedure of image and data analysis. Consequently, most neuroscientists will probably prefer to share data already analysed, relying on the judgment of the researchers who generated the initial data set. For cDNA arrays, the most informative measurement is an expression difference between the two samples that were compared on the same microarray. This standardized value, often called balanced differential expression (BDE) is usually expressed as percent of change. BDE, as any processed data, will partially reflect the interpretation of the investigator who generated the data. The mathematical measurements of fluorescent intensities leading to the BDE value will, in addition to the above-mentioned variables, depend on the software used, grid adjustments, type of background measurement, type of signal measurement, method of balancing Cy3/Cy5 intensity, controls used and other user-defined parameters. So, BDE values reflect heavily processed data and often make comparisons between different data sets very difficult.

Sharing fluorescent probe measurements might be practical, but it makes little sense. This is partially processed data, with specific characteristics that are not shared across data sets, and setting the data reliability cutoffs by an investigator who did not mine the microarray images might lead to very different interpretations of the same data set. At the present time, to avoid generation of flawed data, the same investigator should perform the whole image-analysis procedure.

End-user verification of the critical data. The verification of all microarray data points with an independent method is a monumental task and is clearly not possible. Any shared microarray data will probably have false-positive observations. The fraction of shared data that proves to be inaccurate will vary across different data sets, especially when BDE values are shared. Knowing this, the authors of any new manuscript based on shared data must be responsible for verification of the critical observations relevant for the publication regardless of who generated the initial microarray data (see above).

Keeping everybody's interest in mind. Sharing of analysed data is the foundation of science. However, full disclosure of raw data is not a common practice. For example, the histologist does not make publicly available all her/his *in situ* hybridization micrographs, the electrophysiologist does not provide all recording traces from each neuron and the biochemist does not post all putative binding partners found in complex protein-interaction screenings. Science has a long tradition of sharing reagents and already-analysed data sets, but sharing of raw data is not routinely required. Microarray expression values are neither reagents (like knockout mice or antibodies) nor analysed data sets, but can be understood as the laboratory notebook full of uninterpreted measurements.

Tradition would support the idea that the microarray data relevant for the publication would be made available on request once data are published. However, should one have to make publicly available the entire microarray data set, even if the publication focused on only a subset of genes? Again, according to tradition of scientific publications, the answer is no. As an analogy, if during an intracellular recording, responsiveness of a neuron is tested to seven different substances, is the electrophysiologist required to make publicly available all of these data, even if the first publication focused on the responsiveness of the cells to three of the seven substances? Is the anatomist who injects a tracer into a brain region required to share all projections related to the injected structure, even if the focus of the published study is on a particular set of brain areas? Finally, if an investigator makes a custom microarray with two separate subarrays on it — one for genes involved in synaptic transmission, the other to assess G-protein subunits — will the publication detailing synaptic changes have to include the G-protein hybridization data because the measurements were obtained on the same physical surface?

A requirement for full sharing of microarray data sets immediately after the initial publication will place scientists in a very complex position. A series of high-quality microarray experiments can take several years to complete, and it often results in a wealth of data that is analysed, verified and placed in biological context over a prolonged period of time. By making mandatory the sharing of full data sets, regardless of the focus of the publication, the investigator might be penalized academically and in terms of translational impact. Small laboratories or junior investigators failing to publish at a sufficiently rapid pace could become 'scooped' by results generated from their own data sets. In addition, armed with the researcher's own data, much higher throughput entities, such as big academic laboratories or pharmaceutical companies, will probably discern patentable discoveries sooner. Furthermore, disclosure of potentially patentable data might be in direct violation of research policy of many academic institutions.

To satisfy both the demand of the scientific community for sharing and the investigator's right to pursue novel discoveries generated by his data set, a flexible disclosure time frame should be established. For example, we might require a meaningful microarray data set (see above) from federally funded investigators to post into a public data bank[30,31] at some reasonable time after the expiration of funding.

Generating public databases and standards. As a discipline, we want unlimited, standardized and timely public disclosure of the microarray data. In this quest, we should not rely on the sporadic efforts of individual laboratories. The federal granting agencies should consider contracting out brain microarray expression experiments, similar to what was done for the Human Genome Project. Without such a centralized and coordinated project, standardization of microarray experiments will take a long time. This 'brain-expression-profiling project' sponsored by the National Institutes of Health would use standardized harvesting, storage, nucleic acid isolation, hybridization, printing, scanning and analysis procedures. Furthermore, similar array printing and scanning equipment should be used, and microarray probe sets and arraying surfaces should also be standardized.

Similarly, at the level of NIH agencies sponsoring brain research, we should establish 'standard control samples', which would be freely available for all investigators. These

controls, consisting of pooled mRNA would be generated in a standardized fashion by a central laboratory on a large scale. Each federally-funded investigator should be required to perform several hybridizations of this pooled control sample and publicly disclose it with any shared experimental data. This would provide an internal standard, and facilitate more meaningful data sharing between different laboratories and platforms. This approach already has an excellent track record and is routinely used in cancer research[32].

Concluding remarks

I would very much like to participate in a rapid and open exchange of brain microarray data that will result in meaningful comparisons across many groups of investigators. Despite a potential to misinterpret shared data, the potential benefits of data exchange outweigh my concerns. Microarray sharing could enormously advance our understanding of the most complex and debilitating brain diseases, and ultimately we have to keep in mind the millions of people who suffer from brain disorders and are waiting for novel, better approaches to therapy. However, these sharing efforts will have to be coordinated across the whole neuroscience community to establish a solid foundation.

Károly Mirnics is at the Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA.
e-mail: karoly+@pitt.edu

Links

FURTHER INFORMATION Pittarray

1 Luo, Z. & Geschwind, D. H. Microarray applications in neuroscience. *Neurobiol. Dis.* **8**, 183–193 (2001).
2 Geschwind, D. H. Mice, microarrays, and the genetic diversity of the brain. *Proc. Natl Acad. Sci. USA* **97**, 10676–10678 (2000).
3 Mirnics, K. *et al.* in *Methods in Neurogenetics* (eds Chin, H. & Moldin, S.) (CRC Press LLC, in the press).
4 Mirnics, K. *et al.* Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. *Neuron* **28**, 53–67 (2000).
5 Woo, T. U. *et al.* A subclass of prefrontal γ-aminobutyric acid axon terminals are selectively altered in schizophrenia. *Proc. Natl Acad. Sci. USA* **95**, 5341–5346 (1998).
6 Volk, D. W. *et al.* Decreased glutamic acid decarboxylase 67 messenger RNA expression in a subset of prefrontal cortical γ-aminobutyric acid neurons in subjects with schizophrenia. *Arch. Gen. Psychiatry* **57**, 237–245 (2000).
7 Collinge, J. & Curtis, D. Decreased hippocampal expression of a glutamate receptor gene in schizophrenia. *Br. J. Psychiatry* **159**, 857–859 (1991).
8 Lewohl, J. M. *et al.* Gene expression in human alcoholism: microarray analysis of frontal cortex. *Alcohol Clin. Exp. Res.* **24**, 1873–1882 (2000).
9 Geschwind, D. H. *et al.* A genetic analysis of neural progenitor differentiation. *Neuron* **29**, 325–339 (2001).
10 Mirnics, K. *et al.* Disease-specific changes in regulator of G-protein signaling 4 (RGS4) expression in schizophrenia. *Mol. Psychiatry* **6**, 293–301 (2001).
11 Livesey, F. J. *et al.* Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene Crx. *Curr. Biol.* **10**, 301–310 (2000).
12 Luo, L. *et al.* Gene-expression profiles of laser-captured adjacent neuronal sub-types. *Nature Med.* **5**, 117–122 (1999).
13 Thibault, C. *et al.* Expression profiling of neural cells reveals specific patterns of ethanol-responsive gene expression. *Mol. Pharmacol.* **58**, 1593–1600 (2000).
14 Sandberg, R. *et al.* Regional and strain-specific gene expression mapping in the adult mouse brain. *Proc. Natl Acad. Sci. USA* **97**, 11038–11043 (2000).
15 Ginsberg, S. D. *et al.* Expression profile of transcripts in Alzheimer's disease tangle-bearing CA1 neurons. *Ann. Neurol.* **48**, 77–87 (2000).
16 Ginsberg, S. D. *et al.* Predominance of neuronal mRNAs in individual Alzheimer's disease senile plaques. *Ann. Neurol.* **45**, 174–181 (1999).
17 Whitney, L. W. *et al.* Analysis of gene expression in mutiple sclerosis lesions using cDNA microarrays. *Ann. Neurol.* **46**, 425–428 (1999).
18 Vawter, M. *et al.* Application of cDNA microarrays to examine gene expression differences in schizophrenia. *Brain Res. Bull.* (in the press).
19 Hakak, Y. *et al.* Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. *Proc. Natl Acad. Sci. USA* **98**, 4746–4751 (2001).
20 Barlow, C. & Lockhart, D. J. in *Methods in Neurogenetics* (eds Chin, H. & Moldin, S.) (CRC Press LLC, in the press).
21 Lee, M. L. *et al.* Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc. Natl Acad. Sci. USA* **97**, 9834–9839 (2000).
22 Eisen, M. B. *et al.* Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
23 Hastie, T. *et al.* Supervised harvesting of expression trees. *Genome Biol.* **2**, 0003. 0001–0003. 0012 (2001).
24 Hastie, T. *et al.* 'Gene shaving' as a method for identifying distinct sets of genes with similar expression patterns. *Genome Biol.* **1**, 0003. 0001–0003. 0021 (2000).
25 Toronen, P. *et al.* Analysis of gene expression data using self-organizing maps. *FEBS Lett.* **451**, 142–146 (1999).
26 Tamayo, P. *et al.* Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl Acad. Sci. USA* **96**, 2907–2912 (1999).
27 Brown, M. P. *et al.* Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl Acad. Sci. USA* **97**, 262–267 (2000).
28 Raychaudhuri, S. *et al.* Basic microassay analysis: grouping and feature reduction. *Trends Biotechnol.* **19**, 189–193 (2001).
29 Mirnics, K. *et al.* Analysis of complex brain disorders with microarrays: schizophrenia as a disease of the synapse. *Trends Neurosci.* (in the press).
30 Wheeler, D. L. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **29**, 11–16 (2001).
31 Sherlock, G. *et al.* The Stanford microarray database. *Nucleic Acids Res.* **29**, 152–155 (2001).
32 Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).

Acknowledgements

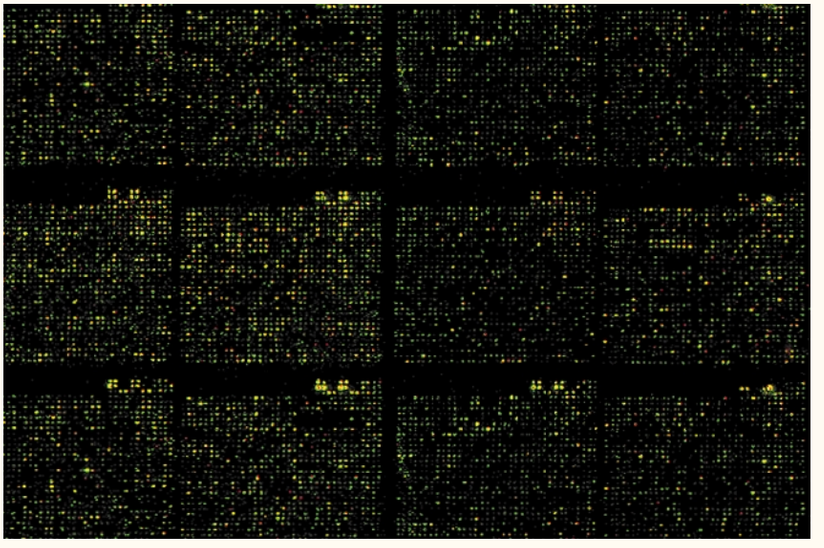
I am grateful to F. A. Middleton, P. Levitt and D. A. Lewis for their critical comments and valuable suggestions regarding the manuscript.

Share your views on microarray data sharing

Has something about this discussion caught your attention? If so, please write to us expressing your point of view. We will continue this debate online and are interested in hearing your opinion.

Send an e-mail to naturereviews@nature.com and flag it for the attention of the *Nature Reviews Neuroscience* editors.

Correspondence will be selected by the editors for publication on the *Nature Reviews Neuroscience* website at http://www.nature.com/reviews/neuro/ where it will be linked to the relevant article.



Hons Neuroscience and Pharmacology 1999-2000
Molecular Biology Module

Mathematical analysis of ligation reactions and hybridisation parameters

Prof. Richard Lathe
Centre for Genome Research and Centre for Neuroscience, University of Edinburgh
Email: rlathe@ed.ac.uk

Some techniques, such as ligation to generate novel constructs, and hybridization (used in almost all molecular biology, from PCR to DNA chips) are so central that the fullest understanding is imperative. In this lecture the student is introduced to the mathematical analysis of the four key paramers (1) ligation dependence on fragment length and concentration (2) DNA/DNA hybridization as a function of temperature, salt and hybrid length (3) DNA probe selectivity as a function of length, and finally (4) the number of clones that one needs to screen using a given hybridization probe.

LIGATION

In the presence of DNA ligase a given DNA molecule with compatible ends can form either (a) a circle, or (b) a dimer (multimer). It is evident that at infinitely low concentration only (a) can occur, while at very high concentration (b) will predominate. To quantitate this phenomenon two parameters are defined:

i = absolute concentration of DNA termini (e.g. in ends / ml or in moles / litre)
j = apparent concentration of one end of a DNA molecule WITH RESPECT TO THE OTHER END OF THE SAME MOLECULE

It follows that when j > i then only circles will be formed, while when i > j then miltimers will be formed.

Calculation of *l*

Consider a molecule 1 kb in length. The average molecular weight of a base-pair is 660 (depending on the base and the counter-balancing ion) and 1kb = 660,000. Thus a typical DNA concentration for recombinant DNA, eg. 5 micrograms (ug) in 100 ul (50 ug/ml) i = 90 nanoM. However, the molecule has two ends, so i = around 160 nanoM.

Larger molecules have less ends per unit weight and the calculation generalises to

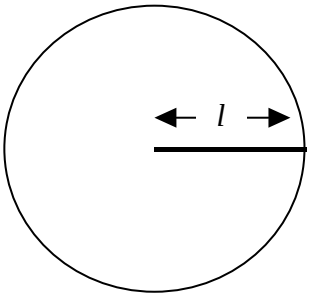
i = 3.2C / L, where L is the length in kb and C is the DNA concentration in ug / ml.

Similar calculations will reveal that a standard 1 kb molecule at 50 ug/ml the number of ends per ml is:

i = 10e14 ends per ml (5 ug 1 kb in 100 ul).

Calculation of j(minimum)

One end of a DNA molecule is constrained, about the other, in a sphere of radius *l* where *l* is the length of the molecule.



If we consider a molecule of 1 kb in length, each bp is 3.4 Angstroms, and 1 kb is 3 x 10e-5 cm. In a sphere of radius *l* the minimum value of j calculates at

j (min) = 6 x 10e12 ends / ml

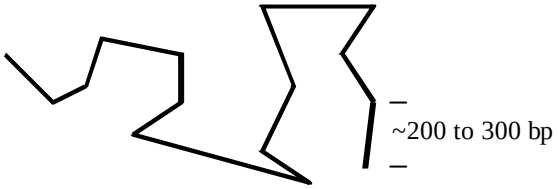
that is roughly 10 nanoM.

Calculation of j(actual)

The actual value of j is much higher, for three reasons.
1. The mathematical model is not rigorous. One could equally consider a model in which the two ends of a molecule are constrained together within a sphere of DIAMERTER = *l*; j values are higher.

2. Empirical factors must be introduced; for instance the DNA itself occupies a finite volume and the contour length is rarely that for the crystallised B form helix.

3. DNA does not behave as a loose string but rather behaves as a flexible polymer of short rigid segments where the random segment length = approx 300 bp, depending on the salt concentration (at higher salt the DNA is more rigid). The minimum segment length was determined from the centrifugation behaviour of lambda DNA and from experimental determination of j values.



The derivation of the formula allowing values of j(actual) to be calculated is complex and need not concern us.

$$j_{act} = \left(\frac{3}{2\pi l b} \right)^{3/2} \text{ ends / ml.}$$

where *l* is the length of the molecule (in cm) and *b* is the random segment length (in cm).

Using experimentally determined values this generalises (approximately) to

Jact = 3.8 x 10e13 / L e3/2 ends /ml

where L = molecule length in kb.

Thus, for a 1 kb molecule,

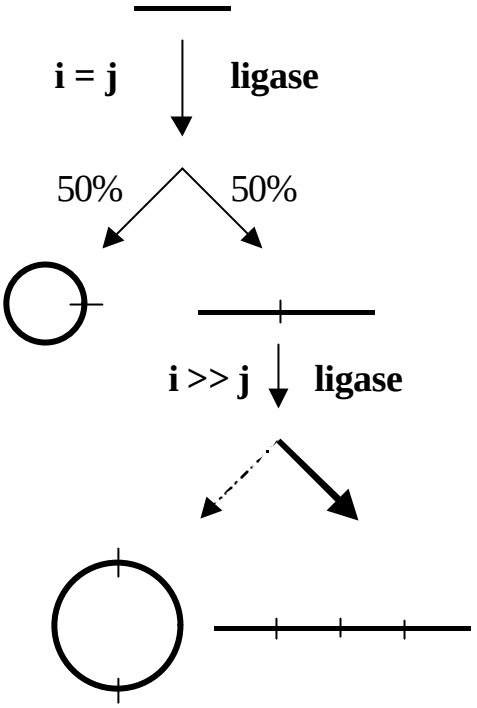
j(actual) = 4 x 10e13 ends /ml

or = 63 nanoM

Weighing I against j

From the above we can immediately see, for a 1 kb molecule (5ug in 100 ul) we are not far away from i=j [I, absolute concentration, was 160 nanoM while j, concentration of one end around the other, was 63 nanoM].

Somewhat longer molecules will increase I and reduce j. But consider what happens if we set *I* = j for a ligation reaction. In the first ligation, circularisation is 50% and dimerisation (or recombinant formation if a mixture of molecules is employed) also represents 50% of the immediate products.



But examine what happens at the next stage. The dimer/recombinant molecule has doubled its length, j has dropped very considerably, and it is now unlikely that the molecule can recircularise to give a viable recombinant! Instead, unproductive multimers are formed.

Much analysis has been performed to determine what are the optimum concentrations when we wish to insert differently sized inserts into a series of standard vectors. The rules emerging are summarised below:

Rules and observations regarding ligation reactions

1. Ensure that there are **an equal number of vector and insert ends** in the reaction.

2. The proportion of correct recombinants increases with increasing vector size, but never exceeds 10% of clones obtained.

3. Optimum concentrations: these are given in the table below.

PLASMID CLONING		
Insert size	Plasmid	Insert

(kb)	vector concen-tration (ug/ml)	concen-tration (ug/ml)
1	250	30
3	60	25
6	25	20
9	12	15
15	4	10
20	2	5

BACTERIOPHAGE CLONING

Because lambda packages naturally as a linear polymer of full-length genomes the concentrations are very different. Optimum concentrations are

Insert = 6kb (large cDNA)
Vector - 1300 ug/ml (yes, >1 mg/ml!)*
Insert - 70 ug/ml

Insert = 18kb (genomic fragment)
Vector - 200 ug/ml
Insert - 30 ug/ml

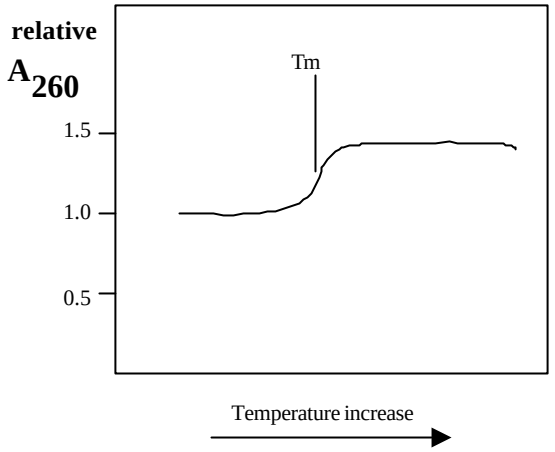
*In practice these concentrations cannot be respected. Typically, for cloning cDNA into lambda the 1 ug of vector is ligated with insert in a final volume of 5 ul

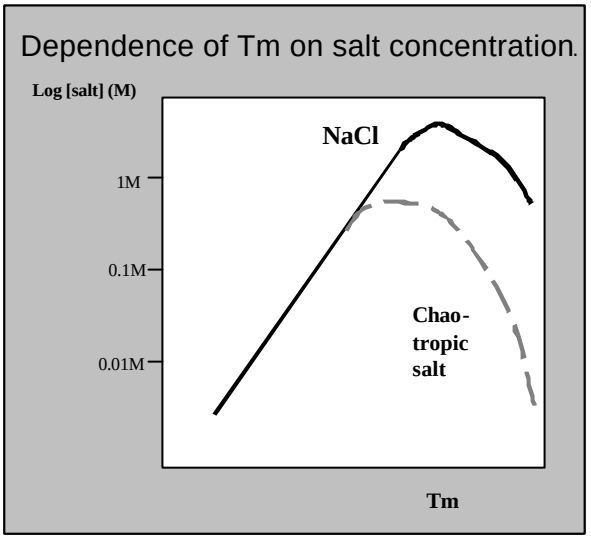
References: Jacobson and Stockmayer (1950) J. Chem. Phys. 18, 1600-1606.
Dugaiczyk et al. (1975) J. Mol. Biol. 96, 171-184.
Legerski and Robberson (1985) J. Mol. Biol. 181, 297-312.

HYBRIDIZATION

Tm measurement. The melting temperature of a DNA duplex is measured by gently heating the DNA and recording at what point the DNA undergoes the transition from double-strand to single-stranded forms that absorb more in the UV range.

Salt [M] stabilises hybrids. The Tm of DNA is critically dependent on the salt concentration of the medium, with increasing salt very strongly stabilising the DNA against denaturation. There is a log relationship between the salt concentration and the Tm.





Note, some salts termed 'chaotropic' destabilise at much lower concentration and can be used to denature DNA - such salts are characterised by very large ion sizes - guanidinium isothiocyanate is a good example (also urea; formamide).

Base composition (%G+C). The stability of a duplex is also governed by the number of G-C pairs (with 3 hydrogen bonds) and the number of A-T pairs (with 2) - more A+T destabilises.

Tm = 16.6 log M + 0.41(%G+C) +81.5

At 300 mM salt, the Tm for long random DNA (G+C = ~50%) is calculated to be around 94oC.

Two further parameters need to be considered.

Length (l). The length of the hybrid is important, and short helices are very much less stable than long ones.

Interstrand homology (h%). Mismatching between the strands will decrease stability.

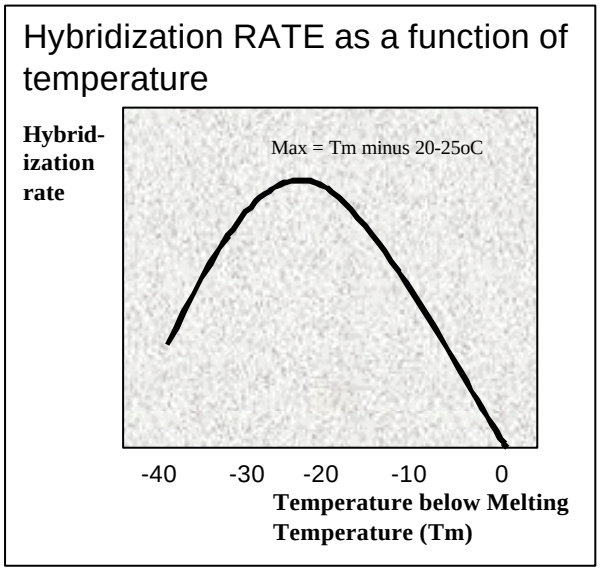
Making some assumptions, we can determine that the Tm in 300 mM NaCl is given by the equation:

Tm = 94 - 820/l - 1.2 (100-h)

For short primers the Tm reduction can be very significant, for instance, an 80-mer will melt 10oC below, while a 20-mer will melt 50oC below the calculated temperature.

Hybridisation temperature

The optimum RATE of hybridisation takes place at about 20 to 25oC below the melting temperature.



Hybridization RATE as a function of temperature

For these reasons hybridisation experiments are usually performed at high temperature - typically 65oC (which is often around 25oC below the melting temperature).

But PCR still works!

Using short primers, PCR amplification often works even when the elongation step is carried out well above the calculated Tm. The reason for this is that the PCR primers are present at enormous concentrations and this alone drives up the Tm very significantly, and perhaps by as much as 20 to 30oC.

PROBE SELECTIVITY

What length of hybridization probe is required to detect a unique gene? A short probe will hybridise to multiple sites in the mammalian genome.

The number of times a given seqence of length l is likely to occur by chance in a genome of length L is given by:

$f = (0.25)^{l} \times 2L$

Basically, to be wholly unique in a mammalian cDNA library a probe of 36 nucleotides is required; in a genomic library this rises to 62 nucleotides, though some compromises can be made.

Reference: *Lathe (1985) J. Molec. Biol. 183, 1-12.*

NUMBER OF CLONES REQUIRED

The number of clones that need to be screened depends on the average insert size, the complexity of the starting material or genome length, and the probability factor required. Small genomes require only a small number of clones, large genomes require large numbers, but this can be reduced by using much larger inserts.

Number of clones required: calculate from formula

$$\text{Number of clones} = \frac{\log(1 - \text{probability})}{\log(1 - \text{Clone length/genome length})}$$

For mammalian genome = ~3 X 109; and selecting a P factor of 99%, if we use a lambda vector we will require 8 x 10e5 clones. As the insert size increases so does the number of clones required decrease:

17kb	lambda	800,000
50 kb	cosmid	300,000
100 kb	BAC	70,000
1000 kb	YAC	14,000

endofdoc

BioTechniques, in press (March 2002)

SHORT TECHNICAL REPORT

Hybridization parameters revisited: solutions containing SDS

Ken Rose, John O. Mason and Richard Lathe*

Department of Biomedical Sciences and Centre for Genome Research, University of Edinburgh, UK

**Address correspondence to R.Lathe, Dept. Biomedical Sciences, Faculty of Medicine, University of Edinburgh, Hugh Robson Building, George Square, Edinburgh EH8 9XD, UK (rlathe@ed.ac.uk).*

Key words: DNA, Hybridization, melting temperature, RNA, SDS

ABSTRACT

Salt concentration governs nucleic acid hybridization according to the Schildkraut-Lifson (1965) equation. High concentrations of sodium dodecyl sulfate (SDS) are used in some common protocols (Church and Gilbert, 1984; Virca et al., 1990) but the effects of SDS on hybridization stringency have not been reported. We investigated hybridization parameters in solutions containing SDS. With targets immobilized on nylon membranes and PCR- or transcription-generated probes we report that the 50% dissociation temperature (Tm) in the absence of SDS was 15-17^{o}C lower than the calculated melting temperature (Tm). SDS had only modest effects on Tm* (1% w/v equating to 8 mM NaCl). RNA/DNA hybrids were ~11^{o}C more stable than DNA/DNA hybrids. Incomplete homology (69%) significantly reduced the Tm* for DNA/DNA hybrids (~14^{o}C; ~0.45^{o}C per % non-homology) but far less so for RNA/DNA hybrids (~2.3^{o}C; ~0.07^{o}C per % non-homology); incomplete homology also markedly reduced the extent of hybridization. On these nylon filters, SDS had a major effect on non-specific binding. Buffers lacking SDS, or with low salt concentration, gave high hybridization backgrounds; buffers containing SDS, or high salt buffers, gave reproducibly low backgrounds.*

INTRODUCTION

Hybridization to sequences immobilized on a solid support is commonly employed to detect sequences related to the hybridization probe while excluding more distantly related nucleic acids. Hybridization experiments performed some decades ago generated the following relation between the concentration of monovalent cation and the melting temperature, Tm, of long (1 kb or more) DNA:

$$Tm = 16.6 \log M + 0.41(\%G+C) + 81.5$$

(10), where M is the monovalent cation concentration (molarity) and % G+C is the percent G+C content. For shorter hybrids (< 1 kb) the expression $\Delta Tm = -820/l$ has been employed (12) where l is the hybrid length in nucleotides and ΔTm the estimated change in Tm (^oC). Somewhat lower values have been advocated elsewhere, including 500/l (5). The change in Tm due to reduced inter-strand homology may be represented by $\Delta Tm = -(100-h)t$, where t is the change in Tm per % non-homology and h is the % homology between the two sequences. Various values have been ascribed to t, varying from 0.04 to 4.8; a midway value of around 1.2^oC per % non-homology has been proposed (7).

With the introduction of new methods of immobilizing (eg. nylon membranes), labelling (polymerase chain reaction, PCR, in vitro transcription) and hybridizing nucleic acids (eg. buffers containing high concentrations of SDS, refs. 6,13) we wished to validate established hybridization parameters. The effects of SDS were not predictable because increasing Na+ concentrations might be expected to increase the stability of the hybrid; a large and potentially chaotropic dodecylsulfate anion could destabilize hybrids, particularly at high concentrations (7% w/v in Church and Gilbert, ref. 6). We standardized a protocol to simulate a common hybridization experiment and investigated, in this model, the effect of various parameters on the extent of hybridization and the intensity of background non-specific hybridization.

MATERIALS AND METHODS

Nucleic acids

Probes and targets were based on ~169 nt segments of the mouse Wnt10A, Wnt10B and Wnt8BB cDNAs (accession numbers U61969 and U61970; ref. 14; AF130349, ref. 8). Equivalent segments of Wnt10A, Wnt10B and Wnt8A (1444-1606, 166 nt, 69% G+C; 805-970, 171 nt, 61% G+C; and 669-824, 160 nt, 66% G+C respectively) were cloned into pBluescript; homology between Wnt10A and Wnt10B segments is 69%. DNA was prepared by a standard CsCl protocol. Probes were radiolabelled by the incorporation of [^{32}P]-nucleotides, either by PCR (DNA probes; standard T3 and T7 forward and reverse primers) or by riboprobe transcription from segments cloned into pBluescript; probes were deproteinized by phenol extraction. DNA probes were denatured (100^oC, 10 min) before use.

Hybridization

Conditions were modified from Church and Gilbert (ref. 6) [1% bovine serum albumin {BSA}, 1 mM Na_2EDTA, 0.5 M $NaPO_4$ pH 7.2, 7% SDS]. DNA targets were denatured (0.5 M NaOH, 1.5 M NaCl, 10 min), applied to filters (Hybond-N, Amersham), renatured (1 M Tris.HCl pH 7.4, 1.5 M NaCl, 10 min), washed (2 x SSC), dried and UV-irradiated (254 nm) before equilibrating in hybridization buffer (0.025 M $NaPO_4$, pH 7.2, 1 mM EDTA, 1% BSA) containing the appropriate concentration of SDS. The total sodium concentrations of were: $NaPO_4$ buffer (25 mM) pH 7.2, [Na^+] = 35 mM; SDS ($CH_3(CH_2)_{10}CH_2OSO_3Na$, MW = 288), 6% w/v, [$Na^+$] = 208 mM; without allowance for partial dissociation. After prehybridization (68^oC, 30 min), the probe was added and incubated overnight. Washing employed hybridization buffer (short washes at room temperature followed by sequential 5 min periods at the designated temperatures as specified). Filter segments were dried before liquid scintillation counting.

RESULTS

We explored the effects of NaCl and SDS on melting temperature and extent of hybridization. This was performed, using 164 nt probes and (plasmid-borne) targets based on the mouse Wnt10A and 10B genes, both for DNA/DNA and RNA/DNA hybrids, and for perfect (100% identity) and imperfect (69% identity) hybrids. Target DNA was immobilized on nylon membranes and hybridized with radiolabelled RNA or DNA probes (prepared by PCR or in vitro transcription) under different salt, temperature and wash conditions. Standardized hybridization was based on the Church and Gilbert (6) protocol; experiments were performed in the presence of BSA, EDTA, a low concentration of $NaPO_4$ (25 mM; a competitor for contaminant radiolabelled inorganic phosphate), and different concentrations of either NaCl or SDS. After washing, the extent of hybridization was determined by direct scintillation counting.

Effect of SDS and NaCl on melting temperature: measured Tm values are less than predicted

To determine the effect of SDS on hybrid stability, hybridization was first performed overnight (68°C). The effective melting temperature (Tm*) was then determined by following hybrid dissociation during serial washing at increasing temperatures. Duplicates were washed (hybridization buffer) at 5°C temperature intervals (each wash, 5 min) over the range 50-90°C. Bound radioactivity was determined after each wash. A typical dissociation curve is plotted in Figure 1. The Tm* values in each case were calculated graphically from the 50% dissociation point (Table 1). For the limitations of this technique see discussion. Both NaCl and SDS stabilized the hybrids; the change of Tm* with increase in NaCl was as predicted by the Schildkraut-Lifson (10) relation, however the measured Tm* values were lower than calculated (15 to 17°C).

Equivalence of SDS and NaCl

Graphically plotted NaCl equivalents were as follows: buffer containing 0.6% SDS was equivalent, with respect to effects on Tm*, to 16.4 mM NaCl; 6% SDS was equivalent to 63 mM NaCl. After subtracting the buffer background this calculated at 7.3 mM NaCl per % SDS (0.6% conditions) and 8.5 mM per % SDS (6% conditions). For further work a round figure of 1% SDS = 8 mM NaCl was adopted, this was substantially less than the calculated figure if SDS was fully ionized (35 mM).

Effect of non-homology on Tm*; relative stability of DNA/DNA and RNA/DNA hybrids

Perfect hybrids, and RNA/DNA hybrids, were most stable. As shown in Table 2A, the Tm* for 69% homologous DNA/DNA hybrids was reduced by 14°C compared to their 100% identical counterparts (0.45°C per % non-homology). For RNA/DNA hybrids, the decline in Tm* was strikingly less, 2.3°C (0.07°C per % non-homology).

RNA/DNA hybrids were considerably more stable than their corresponding DNA/DNA hybrids (Table 2B). The mean differentials were +5.8°C and +17.5°C at 100 and 69% homology respectively. The mean figure, 11.7°C, is comparable with the commonly-adopted figure of 10°C.

Effects on the extent of hybridization

Hybridization temperature was set to 20°C below calculated Tm (perfect hybrids) to achieve near-maximum hybridization rate (4,2,1). Hybridization was performed overnight and, after extensive washing (room temperature) in the same buffers, the absolute quantities of filter-bound radioactivity were determined. Under all conditions tested, non-homology had a major impact on the extent of hybridization (determined after subtraction of the hybridization background obtained from control filters treated in parallel). With reduced inter-strand homology (69%), filter-bound radioactivity fell to 23% for DNA/DNA hybrids, and to 38% for RNA/DNA hybrids. In parallel experiments with less homologous hybrids (57%), the extent of hybridization was further reduced to 10% for DNA/DNA

and 15% for RNA/DNA (not presented). Reduced hybridization rate, in addition to reduced hybrid stability, contributes to the differential hybridization signals generated by perfect and imperfect hybrids.

Hybridization background

In the presence of low salt (10 mM NaCl), and in the absence of SDS, significant but variable levels of radioactivity were retained by the filter despite extensive and stringent washing. The inclusion of SDS in the buffer (0.6 or 6%) or the use of higher salt concentrations (>100 mM) reduced this background substantially (not presented).

DISCUSSION

We have employed a snapshot approach, simulating hybridization experiments as most commonly performed, to estimate hybridization parameters. Our experiments address the temperature at which 50% dissociation takes place, defined here as Tm*, rather than the traditional melting temperature, Tm, defined as the temperature at which 50% association is observed under equilibrium conditions. Tm* may be more appropriate for common hybridization techniques, including Southern and Northern blotting, and DNA microarray hybridization.

Our analysis reconfirms the utility of earlier mathematical analyses of hybridization by Schildkraut and Lifson (10) and by Thomas and Dancis (12) although the apparent melting temperature (Tm*) determined was significantly lower (-17^{o}C) than that calculated from these equations. This may be due to the method employed for Tm* determination, involving serial washing for short periods at increasing stringency. First, the graphically-obtained value of Tm* is an estimate, true values may deviate by several degrees. Second, Tm* reflects dissociation rate rather than Tm as traditionally defined. Third, short wash times should overestimate hybrid stability because the extent of dissociation will be less than during more prolonged incubation (eg. 1 h). However, we saw the reverse effect, with the calculated melting temperature (Tm; ref. 10) substantially exceeding (15 to 17^{o}C) the values (Tm*) measured here. While we have no explanation for this result, the relatively high G+C content of the probes could point to a potential error in the calculation of Tm values according to the 0.41(%G+C) term in the Schildkraut and Lifson (10) equation. Careful reevaluation of this point is warranted. Tm calculations for short oligonucleotides that take sequence into account (9) offer somewhat different values from Schildkraut and Lifson. We also note that polyvalent ions, particularly Mg^{2+}, can make a disproportionate contribution to hybrid stability (15) while under some ionic conditions (eg. tetramethylammonium chloride) A-T pairs can become as stable as G-C pairs (16). Previous hybridization experiments commonly used citrate (to 100 mM) as a chelating agent: EDTA (employed here at 1 mM) is now more prevalent, but at this concentration its ability to sequester specific polyvalent ions may be limited.

However, for comparative assessments (eg. NaCl vs SDS) changes in Tm* values (dissociation) may be predictive of changes in Tm (equilibrium). With the provisos given above, our results suggest that SDS, a common component of hybridization buffers, has no chaotropic (destabilizing) effect on nucleic acid hybrids at the maximum concentration tested (6% w/v). Like NaCl, SDS stabilizes hybrids, with 1% (w/v) SDS being equivalent to approximately 8 mM NaCl in its effects on hybrid melting temperature (rather than to 35 mM if fully ionized).

Under our conditions, the effective melting temperature Tm* for related but mismatched DNA/DNA hybrids was reduced by only 0.6^{o}C per % non-homology, less than the figure of 1.2^{o}C adopted previously 74).

Nevertheless, mismatching significantly reduced the extent of hybridization: at 69% homology the extent of hybridization was reduced by a factor of 5 compared with perfect hybrids, and fell by a factor of 10 at 57% homology (not presented). This accords with earlier reports that Tm reduction due to sequence divergence significantly reduces the rate of association (11,3). Thus probe selectivity for perfect matches versus distantly-related sequences (eg. Southern blotting) is most probably due to a reduction in the rate/extent of hybridization (even after 16 h incubation) rather than to reduced hybrid stability *per se*.

A significant effect of SDS on hybridization background on nylon filters was observed. Low (0.6%) or high (6%) SDS concentrations equally reduced the background, as did NaCl (>100 mM). The mechanism is not known. We also noted that the sharpness of the temperature versus dissociation profile was influenced by salt concentration. At high salt, a temperature range of only 10°C could span fully hybridized to complete dissociation; at the lowest salt concentrations temperature-dependent dissociation was gradual, spanning 25°C or more (data not presented).

We suggest that it is advantageous to perform hybridization in the presence of SDS, as proposed previously (6,13), calculating the melting temperature using the equivalence figure above (1% SDS = 8 mM NaCl), and selecting the hybridization temperature to be approximately 20°C below the Tm. The increased stability of RNA/DNA hybrids should allow stringency conditions (temperature, or equivalent salt) to be increased by 10°C without loss of signal, although caution is advocated if the effective Tm (Tm*) is confirmed to be markedly (15-17°C) below the calculated (10) melting temperature, Tm (this work).

ACKNOWLEDGEMENTS

We would like to thank Melville Richardson for advice and assistance. This work was supported by the BBSRC.

REFERENCES

1. **Anderson, M.L.M. and B.D. Young.** 1985. Quantitative filter hybridization, p 73-111. *In* Hames B.D. and S.J. Higgins (Eds), Nucleic Acid Hybridisation: A Practical Approach. IRL Press, Oxford.
2. **Beltz, G.A., Jacobs, K.A., Eickbush, T.H., Cherbas, P.T. and F. Kafatos.** 1983. Isolation of multigene families and determination of homologies by filter hybridization methods. Meth. Enzymol. *100B:* 266-285.
3. **Bonner, T.I., Brenner, D.J., Neufeld, B.R. and R.J. Britten.** 1973. Reduction in the rate of DNA reassociation by sequence divergence. J. Mol. Biol. *81:* 123-135.
4. **Britten, R.J., Graham, D.E. and B.R. Neufeld.** 1974. Analysis of repeating DNA sequences by reasssociation. Meth. Enzymol. ***29E:*** 363-418.
5. **Britten R.J. and E.H. Davidson.** 1985. Hybridization strategy. p 3-15. *In* Hames B.D. and S.J. Higgins (Eds), Nucleic Acid Hybridisation: A Practical Approach. IRL Press, Oxford.
6. **Church, G.M. and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. U.S.A. *81:* 1991-1995.
7. **Lathe, R.** 1985. Synthetic oligonucleotide probes deduced from amino acid sequence data. Theoretical and practical considerations. J. Mol. Biol. *183:* 1-12.
8. **Richardson, M., Redmond, D., Watson, C.J., and J.O. Mason.** 1999. Mouse Wnt8B is expressed in the developing forebrain and maps to chromosome 19. Mamm. Genome *10:* 923-925.
9. **SantaLucia, J., Jr, Allawi, H.T. and P.A. Seneviratne.** 1996. Improved nearest-neighbor parameters for predicting DNA duplex stability. Biochemistry *35:* 3555-3562.
10. **Schildkraut, C. and S. Lifson.** 1965. Dependence of the melting temperature of DNA on the salt concentration. Biopolymers *3:* 195-208.
11. **Sutton, W.D. and M. McCallum.** 1971. Mismatching and the reassociation rate of mouse satellite DNA. Nature New Biol. *232:* 83-85.
12. **Thomas, C.A.J. and B.M. Dancis.** 1973. Ring stability. J. Molec. Biol. *77:* 43-55.
13. **Virca, G.D., Northemann, W., Shiels, B.R., Widera, G. and S. Broome.** 1990. Simplified northern blot hybridization using 5% sodium dodecyl sulfate. Biotechniques *8:* 370-371.
14. **Wang, J. and G.M. Shackleford.** 1996. Murine Wnt10a and Wnt10b: cloning and expression in developing limbs, face and skin of embryos and in adults. Oncogene *13:* 1537-1544.

15. Willams, A.P., Longfellow, C.E., Freier, S.M./, Kierzek, R. and D.H. Turner. 1989. Laser temperature-jump, spectroscopic, and thermodynamic study of salt effects on duplex formation by dGCATGC. Biochemistry *28:* 4283-4291.

16. Wood, W.I., Gitschier, J., Lasky, L.A. and R.M. Lawn. 1985. Base composition-independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. Proc. Natl. Acad. Sci. U.S.A. *82:* 1585-1588.

Figure 1. **Example of graphic estimation of the effective melting temperature (Tm*) by loss of bound radioactivity during serial washing (5 min) at increasing temperature.** Tm* in this example is estimated to be 63.5°C.

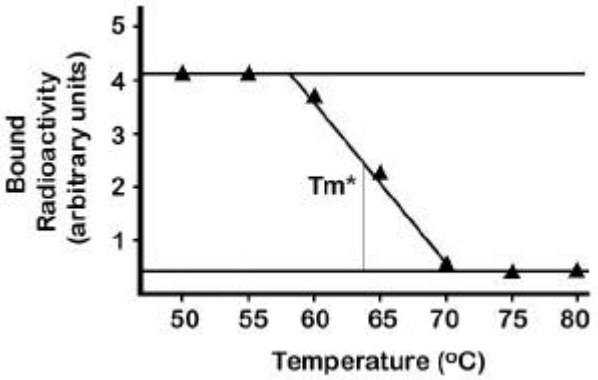


Table 1. Melting temperature (Tm*) as a function of SDS and NaCl concentration for perfect and 69% homologous hybrids

		Melting temperature (Tm*), °C	
Conditions	Hybrid	100% homology	69% homology
10 mM NaCl	DNA	61.5[a]	42
	RNA/DNA	67	65.5
0.6% SDS	DNA	57.5	48.5
	RNA/DNA	63.5	60
100 mM NaCl	DNA	72.5[b]	60
	RNA/DNA	77	76
6% SDS	DNA	68	53
	RNA/DNA	75	72

Values calculated according to refs. (10,12) were [a]79°C [76°C], [b]88.5°C [87°C]; mean differential Tm* = Tm - 17°C [15°C]; [bracketed values allow for incomplete phosphate buffer dissociation].

Table 2. Tm* decline with reduced interstrand homology; increased stability of RNA/DNA versus DNA/DNA hybrids

A: Reduced homology

		Tm* change (°C) due non-homology (69% identity vs 100% identity)
DNA/DNA		
10 mM NaCl		-19.5
100 mM NaCl		-12.5
0.6% SDS		-9.0
6% SDS		-15
	Mean	-14°C (0.45°C per % non-homology)
RNA/DNA		
10 mM NaCl		-1.5
100 mM NaCl		-1
0.6% SDS		-3.5
6% SDS		-3
	Mean	-2.3°C (0.07°C per % non-homology)

B. Elevation of Tm (RNA/DNA vs DNA/DNA), °C

		100% homology	69% homology
10mM NaCl		+5.5	+23.5
100 mM NaCl		+4.5	+16.0
0.6% SDS		+6.0	+11.5
6% SDS		+7.0	+19.0
	Mean	+5.8	+17.5