A calibrated diversity assay for nucleic acid libraries using DiStRO—a Diversity Standard of Random Oligonucleotides

Tatjana Schütze[1,2], Peter F. Arndt[3], Marcus Menger[4], Aniela Wochner[5], Martin Vingron[3], Volker A. Erdmann[1], Hans Lehrach[2], Christian Kaps[6] and Jörn Glökler[2,*]

[1]Institute for Chemistry/Biochemistry, Free University Berlin, Thielallee 63, [2]Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, [3]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, [4]RiNA Netzwerk RNA Technologien GmbH, Takustraße 3, 14195 Berlin, Germany, [5]MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 0QH, UK and [6]TransTissue Technologies GmbH, Tucholskystraße 2, 10117 Berlin, Germany

ABSTRACT

We have determined diversities exceeding 10^{12} different sequences in an annealing and melting assay using synthetic randomized oligonucleotides as a standard. For such high diversities, the annealing kinetics differ from those observed for low diversities, favouring the remelting curve after annealing as the best indicator of complexity. Direct comparisons of nucleic acid pools obtained from an aptamer selection demonstrate that even highly complex populations can be evaluated by using DiStRO, without the need of complicated calculations.

INTRODUCTION

Recent technological developments have employed nucleic acid libraries of staggering diversity. In contrast to arrayed libraries, random combinatorial libraries allow simple and inexpensive handling of greater than a billion different sequences in less than a millilitre. Nowadays the most diverse libraries are used in *Systematic Evolution of Ligands by EXponential enrichment* (SELEX) experiments (1,2) where highly variable initial nucleic acid pools of up to 10^{15} individual sequences are employed (3).

Despite the progress made towards the sequencing of up to 10^7 clones at a time by second generation techniques (4), to date, no method allows the reliable assessment of diversity of larger nucleic acid libraries.

A simple assay to estimate the complexity of DNA has been described earlier (5). The so-called $C_0 t$ analysis is usually performed to measure the size of genomes by observing the absorbance at 260 nm during annealing of complementary DNA fragments of variable lengths.

This assay has been adjusted to be applicable to PCR fragments generated from SELEX experiments using the same principles while reducing ambiguities due to fragment length and primer sequences (6).

A further adjustment to modern real-time PCR equipment, termed AmpliCot, and its application to the measurement of T-cell repertoire diversity has been introduced more recently (7). AmpliCot makes use of the double stranded DNA (dsDNA) binding dye SYBR green to measure the annealing kinetics of even small quantities of DNA. In addition to the $C_0 t_{1/2}$ value (5) commonly used as an indicator of DNA complexity, a linear correlation of the inverse slopes of the kinetic plots and template diversity was described. AmpliCot was evaluated by mixtures of 96 individual clones and the results were used to assign diversities of up to 10^6 by extrapolation. However, apart from the use of such small pools of individual sequences no attempt has been made to validate calculated diversities by calibration of the diversity assays.

To monitor the course of nucleic acid library selections and predict a positive outcome, it is important to reliably determine the quality of libraries and the dynamics of DNA populations during the selection experiments. We sought to develop a general and reliable DNA standard that can be used for the calibration of diversities covering the broad range from single clones up to 10^{12} unique molecules.

MATERIALS AND METHODS

Preparation of dsDNA standard

All randomized synthetic oligonucleotides were synthesized by Purimex RNA/DNA Oligonucleotides (Grebenstein, Germany), using standard phosphoamidite

---

*To whom correspondence should be addressed. Tel: +49 30 8413 1122; Fax: +49 30 8413 1718; Email: gloekler@molgen.mpg.de

chemistry with equally pre-mixed bases at random positions, and were purified via HPLC. Diversity standard oligonucleotide templates were converted to dsDNA by PCR in PTC200-cycler (Biozym Scientific, Hess. Oldendorf, Germany). Briefly, 100 μl reactions with 2 μM primers Ri5 and PO4Ri3 (obtained from Operon, Cologne, Germany with 20 pmol standard oligonucleotide template, 1.5 mM MgCl$_2$, (Promega, USA), 5 U GoTaq®-polymerase (Promega) using four cycles and an annealing temperature of 55°C, yielding an 80 bp PCR product. Residual single-stranded DNA and primers were removed by S1-nuclease (Roboklon, Berlin, Germany) treatment for 15 min at 55°C in S1-buffer (Roboklon). Resulting dsDNA PCR-products were purified by binding to sbeadex® magnetic particles (AGOWA, Berlin, Germany) according to manufacturers' protocol. DNA yield was determined spectrophotometrically (Nanodrop).

Fluorescent diversity assay

The assay was performed using the SYBR Green PCR Core Kit (Applied Biosystems, USA) in 96 well plates with a Biorad Lightcycler device. Each well contained 1.5 μg of dsDNA sample diluted in 50 μl buffer. Reannealing measurements were initiated by denaturation for 2 min at 95°C, and subsequent annealing for 180 min at the respective constant annealing temperature taking one measurement per minute. Remelting was performed starting at 20°C with incremental steps of 0.5°C for 7 s until the final temperature of 98°C was reached.

RESULTS

Design of the diversity standard

Chemical synthesis of oligonucleotides allows the incorporation of random nucleotides at predefined positions. One randomized nucleotide N yields four different sequences and any further randomized nucleotide adds to the diversity exponentially with $D = 4^N$, where N is the number of randomized nucleotides. An even distribution of randomized bases was chosen over a cluster, in order to mimic naturally occurring situations such as somatic

hypermutation in the hypervariable regions of T-cell receptors. We thought that the stringency for small sequence differences is stronger in such an arrangement than in consecutively randomized areas. By this method we designed several oligonucleotide libraries with varying numbers of random nucleotides denoted 2–20 N, based on a single template sequence flanked by conserved primers (Figure 1 and Table 1). The synthetic oligonucleotides were converted to dsDNA by four cycles of PCR and treated by S1 nuclease to remove any unpaired ssDNA. We confirmed the introduction of all four nucleotides at the randomized positions by batch sequencing the oligonucleotides with the conventional Sanger method. Other randomized oligonucleotides obtained from the same supplier, synthesized by the same method, have been subjected to next generation sequencing. Those displayed less than 3.2% deviation from ideal 25% distribution for any base in more than 700 000 sequenced bases in randomized regions (data not shown).

Re-annealing kinetics

As expected we observed that the annealing kinetics varied with the diversity of the used oligonucleotide library. However, despite prolonged annealing times of up to 6 h, the SYBR green fluorescence of 10N and higher stayed well below 50% of its initial value, which made the determination of a $C_0t_{1/2}$ value impossible for libraries with diversities higher than 10^6 (Figure 2a). Triplicate experiments were variable at low diversities (Supplementary Figure 1) and the observed kinetics did not fit the calculations proposed by Baum and McCune (7), especially when diversities exceeded 4×10^3.

Remelting profiles and annealing temperature dependence

We observed that the remelting curves following an annealing step were much more reproducible in terms of melting temperature (Supplementary Figure 2) and characteristic for each of the assayed diversities (Figure 2a and b). Differences are more obvious in the remelting assay when compared with the profiles obtained in the preceding reannealing assay. On an increase in diversity, a decrease in melting temperature from 86°C towards 75°C was
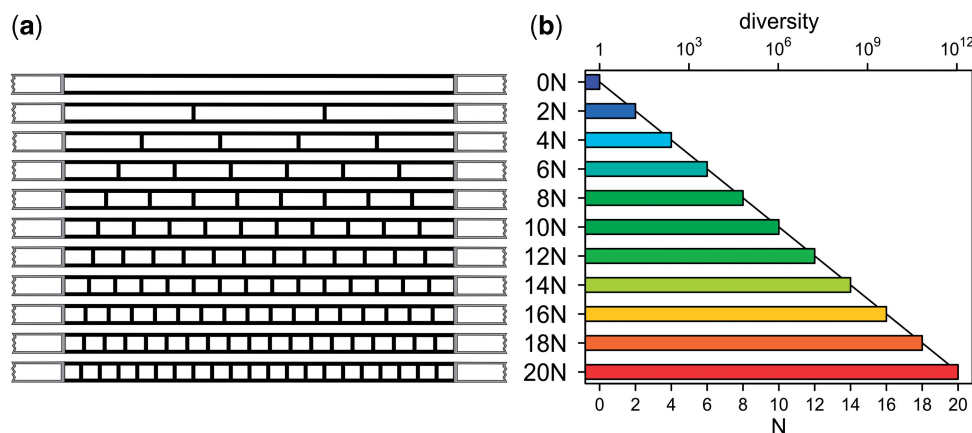
Figure 1. (a) Schematic diagram of diversity standard sequences. Variable positions marked as vertical black bars, flanking primer regions as grey open areas. (b) Variability and resulting diversity of standard DNA fragments.

Table 1. List of all oligonucleotide sequences used in this study

Name	Application	Sequence
Ri5	Primer	5′-GGGAATTCGAGCTCGGTACC
PO4Ri3	Primer	PO_4-5′-CCAAGCTTGCATGCCTGCAG
Div0N	Diversity-Standard	5′-TTCGAGCTCGGTACCGGTAATGT GTAAATCTACTTCCTTCTCAGCGCCCCCGTG GCTGCAGGCATGCAAG
Div2N	Diversity-Standard	5′-TTCGAGCTCGGTACCGGTAATGT GTAAANCTACTTCCTTCTCAGNGCCCCCGTG GCTGCAGGCATGCAAG
Div4N	Diversity-Standard	5′-TTCGAGCTCGGTACCGGTAATGN GTAAATCTNCTTCCTTCNCAGCGCCCNCGTG GCTGCAGGCATGCAAG
Div6N	Diversity-Standard	5′-TTCGAGCTCGGTACCGGNAATGT GNAAATCTNCTTCCTNCTCAGCNCCCCCGNG GCTGCAGGCATGCAAG
Div8N	Diversity-Standard	5′-TTCGAGCTCGGTACCGGNAATGN GTAANTCTANTTCCNTCTCNGCGCNCCCGNG GCTGCAGGCATGCAAG
Div10N	Diversity-Standard	5′-TTCGAGCTCGGTACCGNTAANGT GNAAANCTANTTCNTTCNCAGNGCCNCCGNG GCTGCAGGCATGCAAG
Div12N	Diversity-Standard	5′-TTCGAGCTCGGTACCNGTNATNT GNAAATCNACNTCNTTNTCAGNGCNCCNGTN GCTGCAGGCATGCAAG
Div14N	Diversity-Standard	5′-TTCGAGCTCGGTACCNGNAATNT NTAANTNTACNTNCTTNTNAGCNCNCCCNTN GCTGCAGGCATGCAAG
Div16N	Diversity-Standard	5′-TTCGAGCTCGGTACCNGNANTNT GTNANTNTACNTNCNTCTNANCNCCCNCNTN GCTGCAGGCATGCAAG
Div18N	Diversity-Standard	5′-TTCGAGCTCGGTACCNGNANTNT NTNAATNTNCNTNCNTNTCANCNCNCNCNTN GCTGCAGGCATGCAAG
Div20N	Diversity-Standard	5′-TTCGAGCTCGGTACCNGNANTNTN TNANTNTNCNTNCNTNTNANCNCNCNCNTN GCTGCAGGCATGCAAG

detected, denoting the amount of imperfectly formed heteroduplex. We noticed that the threshold of this shift was partly dependent on the applied annealing temperature. An annealing temperature close to the initial melting point of 80°C leads to a shift towards the heteroduplex population for comparatively low diversities (8 N, 10^5); higher annealing temperatures yielded curves that followed the same pattern as the unmelted, perfectly hybridized samples as described earlier (7,8). Using the lower annealing temperatures of 78°C and 76°C, the shift occurs gradually in higher diversities of 10 N (10^6) and 12 N (10^7), respectively (Supplementary Figure 3). The choice of the annealing temperature will determine the resolution in a given diversity range. Generally, this data demonstrates that there is no continuous linear correlation that determines the annealing kinetics, especially for higher diversities.

Uneven distribution of diversities

As most naturally occurring pools of nucleic acids consist of unevenly distributed sequences, we attempted to simulate this situation by admixtures of varying diversities. The retrieved remelting profiles indicate that by greatly increasing diversity, the pool assumes the kinetics of the lower diversity. This differs from the data

provided by Baum and McCune (7), where one clone was mixed to a final concentration of 50% with a pool of 96 different sequences. Thus strong disparities in clonal distribution may yield misleading values by all melting and annealing kinetics methods.

Evaluation of a SELEX experiment

We applied our new diversity standard to analyze the population dynamics of an aptamer selection against daunomycin (9). We compared PCR aliquots from each selection round (Figure 2c) with our diversity standard (Figure 2b) from the relative distribution of peaks and amplitudes in the remelting assay. The initial population quickly collapses from the initial 10^{15} (25 N) pool to 10^7 (12 N) after just two rounds of selection. The relative plateau of gradual decrease from 10^6 (10 N) to 4×10^4 (6 N) correlates with the emergence of moderate binders in the fifth selection round (10). The final diversity of the tenth round was 1 in 16 (2 N) matching the results obtained by cloning and sequencing (9).

DISCUSSION

The diversity standard was easily synthesized and was shown to perform as anticipated in melting studies.
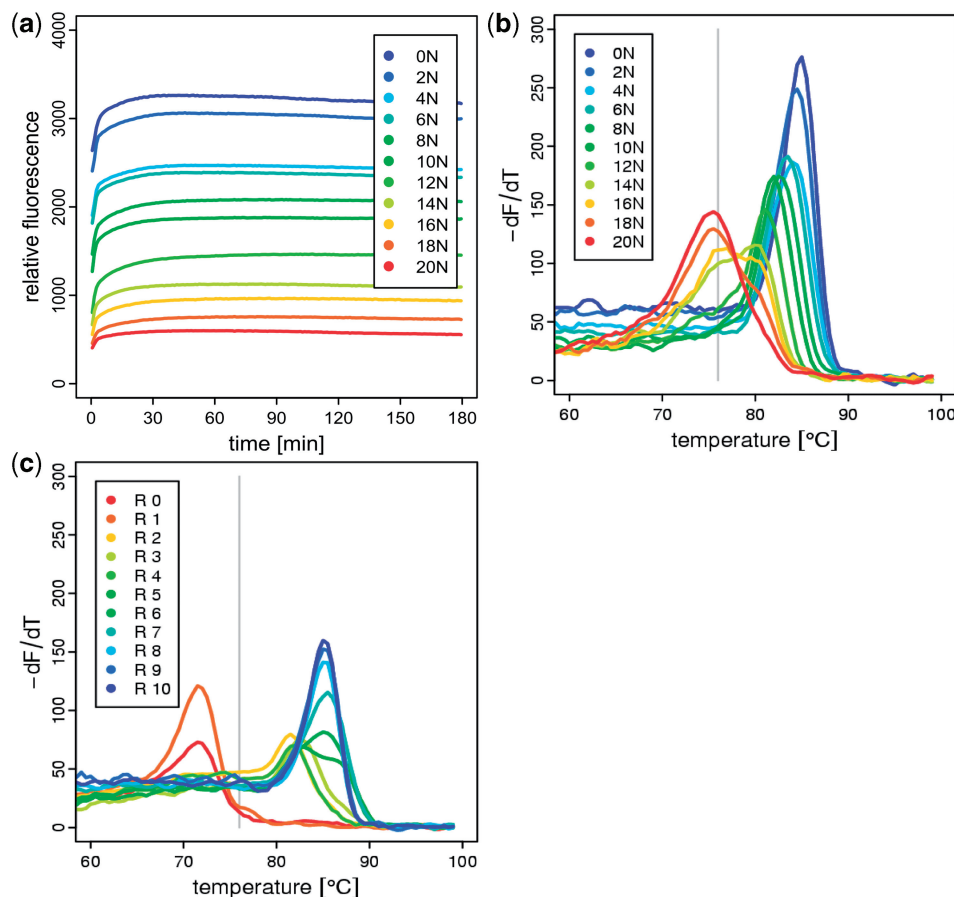
Figure 2. (a) Reannealing kinetics observed for the diversity standard at 76°C, (b) remelting curves for diversity standard after reannealing at 76°C, (c) remelting curves for daunomycin aptamer selection rounds (R 0–R 10) after reanneling at 76°C. Grey lines indicate initial annealing temperature.

We have been able to distinguish DNA populations ranging from single clones to 10^{12} different sequences by using our diversity standard. Previous attempts to describe population diversities by calculation of melting curves lacked necessary internal calibrations and therefore could provide misleading data in some cases. We decided to use the remelting curve as a more characteristic indicator for the analyzed diversity over the usual reannealing profiles. The newly discovered discontinuity may be attributed to the distribution and length of fully annealed regions and corresponding changes in kinetics. Another explanation could be that the exceedingly small concentration of exactly complementary strands diluted into a huge excess of highly similar sequences could result in the formation of kinetically trapped heteroduplexes. Such heteroduplexes may become prevalent beyond $10 N$ (10^6), thus the previously assumed co-linearity of diversities and the annealing process is disrupted.

Other well-established sensitive biochemical assays like ELISA and RT–PCR have already made use of an internal standard for calibration as even small changes may dramatically affect the results. We suggest that a synthetic standard adapted to the sequences of interest can be used to inexpensively and quickly determine diversities of unknown pools. Alternatively, as most nucleic acid libraries are of variable length and consist of unevenly distributed sequences, our DiStRO may serve as a gold standard. Independently performed assays referenced by DiStRO should allow direct comparison of libraries. As naturally derived libraries would display more pronounced differences in clonal copy numbers, our described method using synthetic oligonucleotides will remain partially empirical. However, the defined nature of such oligonucleotides permits reproducible generation and application of the standard. The monitoring of selection experiments of highly complex libraries like in SELEX or other *in vitro* selection techniques (11,12) including library generation (13) is one of the most direct applications. Additionally, calibrated diversity assays could also be useful for the evaluation of DNA pools and provide an economical setup in the rapidly evolving technology of next generation sequencing (14,15); especially paired end sequencing techniques which generate fragments of defined lengths that should facilitate the design and application of a dedicated DiStRO (16).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
2. Ellington,A.D. and Szostak,J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
3. Stoltenburg,R., Reinemann,C. and Strehlitz,B. (2007) SELEX—a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol. Eng.*, **24**, 381–403.
4. Mardis,E.R. (2008) Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.*, **9**, 387–402.
5. Britten,R.J. and Kohne,D.E. (1968) Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science*, **161**, 529–540.
6. Charlton,J. and Smith,D. (1999) Estimation of SELEX pool size by measurement of DNA renaturation rates. *RNA*, **5**, 1326–1332.
7. Baum,P.D. and McCune,J.M. (2006) Direct measurement of T-cell receptor repertoire diversity with AmpliCot. *Nat. Methods*, **3**, 895–901.
8. Ririe,K.M., Rasmussen,R.P. and Wittwer,C.T. (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal. Biochem.*, **245**, 154–160.
9. Wochner,A., Cech,B., Menger,M., Erdmann,V.A. and Glökler,J. (2007) Semi-automated selection of DNA aptamers using magnetic particle handling. *BioTechniques*, **43**, 344, 346, 348.
10. Wochner,A. and Glökler,J. (2007) Nonradioactive fluorescence microtiter plate assay monitoring aptamer selections. *BioTechniques*, **42**, 578, 580, 582.
11. Gupta,R.D. and Tawfik,D.S. (2008) Directed enzyme evolution via small and effective neutral drift libraries. *Nat. Methods*, **5**, 939–942.
12. Lynch,S.A. and Gallivan,J.P. (2009) A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res.*, **37**, 184–192.
13. Stein,V. and Hollfelder,F. (2009) An efficient method to assemble linear DNA templates for in vitro screening and selection systems. *Nucleic Acids Res.*, **37**, e122.
14. Lister,R., Gregory,B.D. and Ecker,J.R. (2009) Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr. Opin. Plant Biol.*, **12**, 107–118.
15. Shendure,J. and Ji,H. (2008) Next-generation DNA sequencing. *Nat. Biotechnol.*, **26**, 1135–1145.
16. Fullwood,M.J., Wei,C., Liu,E.T. and Ruan,Y. (2009) Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res.*, **19**, 521–532.