Droplets as Microreactors for High-Throughput Biology

Valerie Taly, [a, c] Bernard T. Kelly, [b] and Andrew D. Griffiths*[c]

Inspired by the principles of biological evolution, biologists—and others—have in recent decades harnessed the power of “natural” selection to sift through huge libraries of genes and find those with desirable properties. At the same time, the demand for high-throughput biochemical and genetic assays and screens has driven the development of increasingly miniaturised assay systems. An exciting synergy is now emerging between these two fields, whereby the tools of ultrahigh-throughput screening promise to open up new directions in molecular engineering.

"What limit can be put to this power, acting during long ages and rigidly scrutinising the whole constitution, structure and habits of each creature—favouring the good and rejecting the bad? I can see no limit to this power, in slowly and beautifully adapting each form to the most complex relations of life.”

Charles Darwin, The Origin of Species

Introduction

The key to the success of molecular engineering and the burgeoning fields of proteomics and its related disciplines is the ability to discriminate between molecules on the basis of their properties. In the case of molecular engineering, which traditionally encompasses the fields of protein and nucleic acid engineering, one seeks to isolate molecules with desirable properties, such as catalysis of a particular reaction, binding to a particular molecule or surface, or a regulatory activity, whereas in proteomics one tries to decipher the complex web of the cell’s molecular interactions by identifying binding and regulatory partners of each and every protein. Broadly speaking, there are two approaches to the problem of sifting through large collections of genes to identify those that encode “desirable” molecules, whether from an engineering or proteomics point of view. One may inspect the properties of the RNA and protein molecule encoded by each candidate gene directly—essentially a high-throughput screening approach—or one may mimic natural selection more closely, for example, by linking the properties of the encoded products with the survival of the cells that contain them.

High-throughput screening—usually performed in microtitre plates—of collections of typically 10^3–10^5 genes, has been used successfully to engineer proteins with desirable properties, but since there is good theoretical and experimental evidence that larger libraries yield “better” molecules (tighter binders, more active catalysts), the much higher throughputs attainable by evolutionary techniques are highly desirable. In addition, the huge number of potential protein–protein and protein–nucleic acid interactions in the expressed genome limit the usefulness of microtiter-plate-based high-throughput screens; for example, with ~30000 genes in the human genome, a total of ~10^9 potential pair-wise protein–protein interactions must be screened, and realistic microtitre plate screens are limited by handling times and reagent costs to a maximum of 10^6–10^7.

Protein engineers often try to couple the survival of a host cell to the activity of a mutant protein encoded by a gene. Such evolutionary schemes allow a very high-throughput (limited only by the transformation efficiency to ~10^11)[2] but suffer from two main drawbacks. First, they are often complicated by the complex intracellular environment, and secondly they do not normally allow the direct observation and measurement of the properties of individual library members. This, added to the fact that it is often difficult to vary the threshold for selection, robs the experimenter of some degree of control over the search.

All evolutionary systems require a link between genotype (a nucleic acid that can be replicated) and phenotype (a functional trait such as binding or catalytic activity). In nature, this linkage is performed by cellular compartmentalisation. In the laboratory, a very effective strategy is to provide a physical genotype–phenotype linkage. Various techniques have been developed for this purpose, including phage display, plasmid display, ribosomal display, CIS display, mRNA–peptide fusion as well as display on the surface of cells. However, while highly successful in the generation of binding proteins, such methods have met with much less success for the selection of enzymes. Selection can be based on binding to transition-state ana-
logues or enzyme inhibitors but the selection is not directly for catalysis.\[4\]

The model of cellular compartmentalisation has inspired the development of a fully in vitro system, termed in vitro compartmentalisation (IVC).\[5\] The basic idea is to perform billions of experiments by partitioning each experiment into a separate microscopic compartment (working as an individual microreactor). Each compartment is a microdroplet of water that contains all the ingredients for an experiment, dispersed in a water-in-oil (w/o) emulsion. Emulsions are created at ratios of genes to droplets such that most of the droplets contain no more than a single gene. Each aqueous microcompartment serves as an artificial cell where transcription and translation, and the activity of the resulting RNA or protein all take place within the compartment. The oil isolates each microdroplet and therefore prevents exchanges of materials between them. These microdroplets can have a diameter as small as 1 μm (as small as a bacterium). Approximately 50 μL of reaction mixture can thus be dispersed into >10\(^10\) individual microreactors, which makes IVC an ideal tool for high-throughput screening of biological or chemical reactions. It selects for all enzymatic features simultaneously: substrate recognition, product formation, rate acceleration and turnover.

**Directed Evolution in Emulsions**

Initially, IVC was developed for directed evolution purposes. Directed evolution is a powerful approach that harnesses the Darwinian evolutionary principle of iterative rounds of mutation/recombination and selection to evolve new nucleic acids or proteins in the laboratory. Many applications of emulsions in molecular evolution have been described and the isolation of proteins, and most recently RNAs, with binding, enzymatic and regulatory functions has been demonstrated.\[6\] Several formats of IVC have been developed that are of general utility and detailed protocols have recently been published.\[7\] Whilst the first stage of all these laboratory evolution processes is generic (genes are transcribed and translated in discrete compartments), the second stage is not: different phenotypes dictate different means of selection.
Selection for Catalysis by Using Physical Gene–Product Linkage

In the original format of IVC, the substrate was physically linked to the gene. If the gene encoded for an active enzyme, the substrate was converted into a detectable product linked to the gene, whereas genes that encoded inactive enzymes were attached to unmodified substrate molecules. Genes that had attached product could then be separated, for example, by affinity purification by using a product-specific ligand. The simplest applications of this strategy lie in the selection of DNA-modifying enzymes, for which the gene and substrate comprise the same molecule. This system was used to investigate the role of DNA-(cytosine C5)-methyltransferase HhaI residues for the recognition of its target sequence as well as to select variants of DNA methyltransferases HaeIII with new sequence specificities. As described in Figure 1, in this selection format the library genes were linked to the target recognition sequence. Selection was performed by extracting the genes from the emulsion and incubating them with a cognate restriction enzyme that cleaves nonmethylated DNA. If the encoded methyltransferases recognize and methylate this sequence, then the corresponding DNA molecule is protected and can be amplified by PCR. A variant of HaeIII methyltransferase that exhibited a 670-fold enrichment in catalytic efficiency ($k_{cat}/K_m$) for a nonpalindromic target sequence (AGGC) was selected. A comparable improvement (560 times) was obtained for the recognition sequence CGCC. Interestingly, a ninefold improvement was also obtained for the original recognition site (GGCC). This is a rare example of a laboratory-evolved enzyme that catalyzes multiple turnovers by starting from libraries of polymerase genes, and dNTPs, and compartmentalized in the aqueous compartments of water-in-oil emulsions. The emulsion is thermocycled; this results in the release of polymerases and their corresponding genes, which in turn allows DNA replication to proceed. Each polymerase only replicates its own gene to the exclusion of those in other compartments. Only genes that encode active polymerases (green or yellow circles) are replicated, while inactive variants (magenta circles) fail to amplify their own gene and disappear from the gene pool. In addition, genes that encode more active variants will have higher representation in this gene pool (green variants compared to yellow ones). After breaking the emulsion the aqueous compartments are combined. The surviving population is then amplified by PCR and either 4) characterized or 5) submitted to a further round of CSR. The figure was adapted from ref. [11] with permission; copyright National Academy of Sciences, USA, 2001.

IVC has also been used to select DNA polymerases by using so called "compartmentalised self-replication" or CSR (Figure 2). In this method an active polymerase catalyses the replication of the gene that encodes it, within an emulsion droplet, whereas inactive variants fail to amplify their own genes and disappear from the library pool. Polymerases exhibiting higher thermostability and resistance to a potent inhibitor or with modified base specificities were thus selected from gene libraries.

The strategy of selecting for gene–product conjugates can also be used to select reactions with non-DNA substrates. Recently Diels–Alderase ribozymes were selected for intermolecular catalysis and multiple turnovers by starting from libraries that contained 10$^3$ diversified ribozyme genes. Microbeads can be used to couple enzyme-coding genes to multiple (10$^3$) product molecules subsequent to enzymatic conversion thereby selecting for high turnover. Product-coated beads that carry enzyme-coding genes can be detected with fluorescently-labelled anti-product antibodies and enriched by using fluorescence activated cell sorting (FACS). This strategy has been used to select both enzymes and ribozymes.

Figure 3 illustrates how such a procedure was used to directly select for trans-acting ligase ribozymes in multiple-turnover reactions. Genes encoding active restriction endonucleases were then selected by using affinity purification with streptavidin-coated beads.

Figure 2. Selection of polymerase variants by compartmentalised self replication (CSR). 1) A library of genes that encode polymerase variants (grey spheres) is cloned and expressed in E. coli. 2) Bacterial cells carrying the polymerases and corresponding genes are dispersed in a buffer that contains primers, which bind flanking sequences of the polymerase genes, and dNTPs, and compartmentalised within the aqueous compartments of water-in-oil emulsions. 3) The emulsion is thermocycled; this results in the release of polymerases and their corresponding genes, which in turn allows DNA replication to proceed. Each polymerase only replicates its own gene to the exclusion of those in other compartments. Only genes that encode active polymerases (green or yellow circles) are replicated, while inactive variants (magenta circles) fail to amplify their own gene and disappear from the gene pool. In addition, genes that encode more active variants will have higher representation in this gene pool (green variants compared to yellow ones). 4) After breaking the emulsion the aqueous compartments are combined. The surviving population is then amplified by PCR and either 4) characterized or 5) submitted to a further round of CSR. The figure was adapted from ref. [11] with permission; copyright National Academy of Sciences, USA, 2001.
conditions. As for the selection of Diels–Alderase ribozymes mentioned above, such properties of ribozymes were previously not selectable with the widely used technique of systematic evolution of ligands by exponential enrichment (SELEX),[15] which can only select for intramolecular, single-turnover events.

Uncoupling Transcription/Translation and Reaction Steps

Since IVC is a purely in vitro procedure, it allows selection in a wider range of reaction conditions than in vivo systems: one can use substrates, products or reaction conditions that are incompatible with in vivo systems. However, cell-free translation must be performed under defined pH, buffer, ionic strength and metal ion composition.

However, translation can be completely separated from selection by using two sequential emulsification steps, and uncoupling the expression and selection steps allows the use of conditions that are incompatible with transcription/translation. Figure 4 illustrates how such a strategy was used to evolve one of the most efficient enzymes ever described (a phosphotriesterase with a $k_{\text{cat}}$ value of $>10^7 \text{s}^{-1}$).[18] In the first step microbeads that displayed a single gene and multiple copies of the encoded protein variants were formed after translation of the immobilised gene, and the resulting protein was captured with an affinity tag. The microbeads were isolated and re-emulsified in a buffer that contained zinc and carbonate ions to allow the captured inactive apoenzyme to assemble into the catalytically active metalloenzyme. Subsequently, a phosphotriester substrate attached to caged biotin was added. After irradiation of the emulsion to uncage the biotin, the product coated beads that displayed active enzymes and the genes that encoded them were detected by using fluorescently labelled anti-product antibodies and selected by using FACS.

By using a single emulsification procedure it is also possible to select under conditions that are incompatible with the transcription/translation step. This can be done by modifying the content of droplets, without affecting their integrity, once the protein is translated. For example, hydrophobic substrates can be added through the oil phase, the droplet pH can be reduced by delivery of an acid or a substrate contained in the aqueous droplet can be photoactivated.[6] A nanodroplet delivery system was developed for the selection of DNA-nuclease inhibitors,[16] which allowed the transport of various water-soluble solutes including metal ions into the emulsion droplets. Inactive DNA-nuclease were co-compartmentalised with a gene library. After expression of the different variants encoded by the library, the DNA nuclease were activated by delivery of nickel or cobalt ions. Genes encoding nuclease inhibitors survived the digestion and were subsequently amplified and isolated. This method allowed direct selection for regulation (inhibition in this case) rather than simply for binding of the nuclease (as with other in vitro selection approaches such as phage or ribosome display). These experiments led to an increase of well over $10^4$ in affinity as well as selectivity.[26]

Selection for Catalysis by Using FACS to Sort Emulsion Droplets

Although efficient, the selection procedures described above have to be adapted for each enzyme, substrate and reaction, and hence for each evolution experiment. An interesting alternative is to directly select droplets that contain the product of interest and consequently the gene encoding the molecule (RNA or protein) that can catalyse the desired reaction. However, although compartmentalisation in w/o emulsions provides a very efficient way of maintaining the link between the gene, the protein or RNA it encodes and the reaction product, it does not allow direct selection of droplets.

FACS is a powerful technique that holds great promise for high-throughput screening. Modern FACS machines can routinely analyse and sort more than $10^7$ events per hour based on their fluorescence signal. It has already proven very successful for the selection of antibodies with high binding affinity.[27] However, when selecting for catalysis, the product has to

Figure 3. Selection of trans-acting RNA ligase ribozymes by IVC. 1) Streptavidin-coated beads that display one substrate (blue lines; Half 1) as well as one gene from a library of genes that encode ribozyme ligase variants are emulsified with an in vitro transcription/translation mixture, together with a second substrate molecule that bears a fluorophore (yellow lines; Half 2). 2) The genes are transcribed into the corresponding RNA within the compartments. 3) Functional RNA ligase ribozymes ligate the two substrate molecules together; this results in fluorescent beads. 4) The emulsion is broken, all reactions are stopped and the aqueous compartments are combined. 5) Labelled beads are isolated by FACS, and the attached DNA encoding for functional ligase ribozyme is amplified by PCR and either characterised or recompartmentalised for further rounds of selection. The figure was adapted from ref. [14] with permission; copyright the RNA society, 2005.
remained confined in the cell or be captured on the surface of cells or microbeads.\textsuperscript{[13,14,18]}

Recently, we have demonstrated IVC selection based on the direct sorting of intact emulsion droplets by using FACS.\textsuperscript{[19]}

This format of IVC uses fluorogenic substrates that are available for a wide range of enzymes. Moreover, detection of fluorescence is a sensitive method. The initial steps are the same as for the IVC procedures described above. Single genes are compartmentalised with all the components for in vitro expression and compartmentalisation in a water-in-oil emulsion, which contains a soluble substrate that is attached to caged biotin; the substrate is converted to product only in compartments that contain beads and therefore display active enzyme. Consequently, in a compartment that contains a gene encoding an enzyme, the product becomes attached to the gene via the bead. In other compartments, in which the genes do not encode an enzyme for the selected reaction, the intact substrate becomes attached to the gene. The emulsion is broken, and the beads are incubated with anti-product antibodies. Product-coated beads can then be enriched by using flow cytometry together with the genes attached to them, after treatment with a fluorescently labelled antibody; this allows sorting at up to 100,000 events per second. This figure was adapted from ref.\textsuperscript{[13]} with permission; copyright Macmillan Publishers Ltd., 2003.

![Diagram](http://example.com/diagram.png)

Figure 4. Creation of microbead-display libraries, and selection for binding and catalysis by using in vitro compartmentalisation. A) The creation of microbead-display libraries: 1) a repertoire of genes that encode protein variants, each with a common N- or C-terminal epitope tag, are linked to streptavidin-coated beads, which carry antibodies that bind the epitope tag at, on average, less than one gene per bead; 2) the beads are compartmentalised in water-in-oil emulsion to give, on average, less than one bead per compartment, and transcribed and translated in the compartments; 3) consequently, in each compartment the translated protein (10–100 copies) becomes attached to the gene that encodes it via the bead; 4) the emulsion is broken and 5) the microbeads that carry the display library are isolated. To select for binding the microbeads can be affinity purified directly either by using immobilised ligand or labelled with a fluorescent marker and sorted by using flow cytometry. B) Selection for catalysis: 1) microbead-display libraries are resuspended in a reaction buffer of choice and compartmentalised in a second water-in-oil emulsion, which contains a soluble substrate that is attached to caged biotin; 2) the substrate is converted to product only in compartments that contain beads and therefore display active enzyme; 3) the emulsion is then irradiated to uncage the biotin. Consequently, in a compartment that contains a gene encoding an enzyme, 4) the product becomes attached to the gene via the bead. In other compartments, in which the genes do not encode an enzyme for the selected reaction, the intact substrate becomes attached to the gene. 5) The emulsion is broken, and 6) the beads are incubated with anti-product antibodies. Product-coated beads can then be enriched by using flow cytometry (together with the genes attached to them) after treatment with a fluorescently labelled antibody; this allows sorting at up to 100,000 events per second. This figure was adapted from ref.\textsuperscript{[13]} with permission; copyright Macmillan Publishers Ltd., 2003.

![Diagram](http://example.com/diagram.png)

not accessible with in vivo studies. Moreover, since the procedure is completely in vitro, it is not constrained with the necessity to maintain cell viability. In addition, the broad dynamic range of such procedure was demonstrated: clones presenting different catalytic efficiencies are easily distinguishable by using FACS.

In the same way, single bacterial cells expressing a library of variants in the cytoplasm, the periplasm or on their surface can...
also be emulsified and sorted by using FACS. \cite{22} Serum peroxi-
ondase (PON1) is a mammalian enzyme that catalyses the hy-
drolysis and inactivation of a broad range of substrates, includ-
ing plasma homocysteine thiolactone (HcyT). However, even
though PON1 allows a certain detoxification of this toxic metab-
olite, its catalytic efficiency is quite poor (\(k_{\text{cat}}/K_{\text{m}} \approx 75 \text{ m}^{-1} \text{s}^{-1}\)).
Thus, developing a procedure to select for HcyT and thiobuty-
rolactones (TBLs) in general presented a challenge for detec-
tion. The encapsulation of single cells in droplets not only
allows the crucial linkage between genotype and phenotype,
but also provides an unusually high enzyme concentration
(\(>10^4\) enzyme molecules in <10 femtoliter) that enables de-
tection and selection at extremely low signal-to-noise ratios.
A wide range of PON1 variants with a more than 100-fold
improvement in thiolactonase activity were identified. These
results demonstrated the applicability of double emulsions for
single-cell phenotyping.

The ability to perform multiparameter flow sorting also
opens up the possibility of selecting enzymes with two or
more substrates, each of which is hydrolysed to release a
different coloured fluorophore, thereby allowing selection for
substrate specificity.

Selection for Binding by Using IVC

IVC was also used to select proteins and peptides for binding.
Indeed, the compartmentalisation of genes and their encoded
proteins or peptides in emulsions constitutes a way of estab-
lishing a physical link between them. As described below this
link can be covalent \cite{23} or noncovalent \cite{24–27} and direct or
through microbeads. \cite{13, 28}

STABLE \cite{24} (streptavidin–biotin linkage in emulsion) is an in
vitro protein–DNA noncovalent fusion system for the screening
of combinatorial libraries. Biotinylated gene encoding strep-
tavidin (STA)-fused polypeptides are compartmentalised and
transcribed/translated within a w/o emulsion. The encoded
molecules can thus be attached to the encoding DNA through its
biotin label. This DNA display system was used to screen for
peptide ligands \cite{24, 25} and in model selection of glutathione-
transferase for glutathione binding. \cite{26} Noncovalent DNA-en-
coded protein coupling can also be mediated by microbeads
(as described above for catalysis) and this methodology has
also been used to select for binding. \cite{22} DNA-binding proteins
can also be selected with IVC by linking genes and the DNA-
binding protein they encode. \cite{27} Zinc finger proteins (ZFP) were
selected by linking the cognate recognition sequence to the
gene fragments that encoded the ZFP libraries. A peptide-epi-
tope tag was appended to the ZFP; this allowed the affinity
purification of the genes of interest (i.e., by encoding a ZFP
with the ability to bind to the expected DNA sequence).

Bertschinger et al. \cite{22} have described the combination of IVC
and covalent-DNA display. In this case, the library of genes of
interest were linked to a DNA fragment that encoded a DNA
methyltransferase (M.HaeIII) that could form a covalent bond
with a 5'-fluorodeoxycytidine base at the extremity of the DNA
fragment. This resulted in a library in which DNA and protein
were covalently linked. Model selection experiments were suc-
cessfully realised by using specific ligands for the capture of
peptides and globular proteins bound to DNA.

Emulsion PCR

Water-in-oil emulsions that are stable at temperatures higher
than 90 °C can be formulated. PCR in such thermostable emul-
sions was first exploited to evolve DNA polymerases by CSR
(see above). Emulsion PCR (ePCR) constitutes a very efficient
and simple way to perform parallel amplification of single DNA
or RNA molecules. Individual molecules are segregated in
aqueous droplets of emulsion and amplified by PCR in iso-
lation; the genes are either free in solution or linked to beads.
Emulsion PCR has quickly been adapted for a range of other
applications, including sequencing \cite{29, 30} and assessing genetic vari-
ations \cite{31–34} or high-throughput screening of transcription factor
targets. \cite{35} Some of these are described in more detail below.

ePCR to Circumvent PCR Bias

The efficient amplification of genomic libraries—cDNA libraries
and other complex mixtures of genes—by using PCR is imped-
ed by two phenomena. First, short fragments tend to be am-
plicated in preference to larger ones; and, secondly, artefactual
fragments are generated by recombination between homolo-
gous regions of DNA. \cite{36} Recombination in this case occurs
when a primer is partially extended on one template during
one cycle of PCR, and extended further on another template
during a later cycle. Thus, chimeric molecules are generated
and the shorter fragments are then preferentially amplified. A
variety of PCR protocols have been proposed to minimize
these problems, most of which rely on high template concen-
trations and low numbers of PCR cycles. \cite{37} Clearly, however,
such an approach is not viable if little template DNA is avail-
able. In contrast ePCR enables the use of small amounts of
template DNA and high numbers of PCR cycles.

Emulsion PCR was compared to conventional PCR for the
amplification of genomic DNA from Haemophilus parahaemoly-
ticus. \cite{38} For nonemulsified PCR the average size of the ampli-
fied DNA decreased with the number of PCR cycles whereas
for ePCR the size distributions of the amplified DNA remained
essentially constant, irrespective of the number of PCR cycles.

The absence of artefactual PCR products when using emulsion
PCR can be explained by the presence of a single, or at most a
few, template DNA molecules in each droplet. This segregation
of template DNA molecules prevents recombination between
homologous or partially homologous gene fragments during
PCR, thus eliminating the synthesis of short, chimeric products
and other artefacts. It also reduces competition between frag-
ments of different lengths, thus diminishing the bias for ampli-
ifying smaller fragments.

BEAMing: A Method to Assess Genetic Variations

Emulsion PCR has been used to assess genetic variations in a
large population of DNA molecules. \cite{34} A method called BEAM-
(on the basis of its four principal components: beads, emulsion, amplification and magnetics) allows the detection and quantification of sequence variants by transforming each DNA molecule into a fluorescent magnetic particle.

After preamplification of DNA samples by PCR, single DNA molecules are amplified within w/o emulsions and bound to magnetic beads. Each bead then contains thousands of copies of DNA identical in sequence to the original. After breaking the emulsion, hybridisation of fluorescently-labelled oligonucleotides is used for sequence differentiation. The resulting population of beads can then be analysed by FACS. In this way, small numbers of genetic variants can be identified in a population of millions of DNA molecules. This allows the study of variations in gene sequences or transcripts in specific populations or tissues as well as the identification of rare mutations (Figure 6).

A similar procedure was used for the high-throughput screening of transcription factor targets. In this approach, a GLOBE (genetic library on beads) was constructed by solid-phase single-molecule PCR in w/o emulsions. After breaking the emulsion and recovering the beads, in vitro expressed tagged DNA-binding proteins were added. After addition of fluorescently-labelled anti-tag antibodies, the fluorescent beads were sorted by FACS and corresponding target DNA molecules were recovered.

High-Throughput Sequencing with Emulsions

The ability to use ePCR to amplify individual DNA molecules isolated within aqueous compartments allows clonal amplification and provides a completely in vitro system for preparing template DNA for sequencing. Emulsion PCR allows amplification of single template DNA molecules on beads, which leads to microbeads that contain thousands of copies of the original DNA. Two new parallel sequencing methods that use this technology to produce the template for sequencing were recently described. However, the sequencing systems themselves are quite different.

Margulies et al. described an impressive new sequencing technology that combines emulsion compartmentalisation with a readout on microstructured arrays of wells. This technology has the ability to sequence entire bacterial genomes in a matter of hours in a single run. This parallel system is capable of sequencing 25 million bases in a four hour period, which is about 100-times faster than the current state-of-the-art Sanger sequencing method on capillary based platform. The process starts with fragmentation of genomic DNA and subsequent ligation of adaptor sequences to the ends of the fragments. Single-stranded DNA is then attached to a microbead (~28 m in diameter) so that the majority of beads contain no more than a single DNA fragment. The beads are then compartmentalised in a thermostable w/o emulsion (each aqueous compartment contains no more than one bead) and ePCR is performed. At the end of this procedure each DNA sequence has been clonally amplified on a single microbead and

![Figure 6. Accessing the genetic diversity of a DNA population by BEAMing. 1) Streptavidin coated magnetic beads that display biotinylated oligonucleotides (primer 1) are emulsified (w/o) with a template molecule and all components (polymerase, dNTPs and primer 2—a second flanking primer) required for DNA amplification. Since priming by oligonucleotides linked to beads has been found to be much less efficient than priming by the same oligonucleotides free in solution, a small amount of nonbiotinylated forward primer (identical in sequence to the oligonucleotide coupled to the bead) is also compartmentalised. The aqueous compartments contain an average of less than one template molecule and less than one bead. In this scheme, red and green templates represent two DNA molecules with sequences that differ by one or more nucleotides. 2) Thermocycling results in beads coated with thousands of identical copies of the single DNA molecule originally present. 3) The emulsion is broken and the beads are purified with a magnet. After denaturation of the DNA, the beads are incubated with fluorescently labelled oligonucleotide probes (represented in red and green) that can distinguish between the sequences of the different kinds of template. 4) Beads coated with the green or red template are now distinguishable by using flow cytometry (red and green regions) and can be counted within the population. The region that contains beads that failed to hybridise to any probe is grey. Beads that hybridised to both probes would constitute a fourth population, not represented here; FL1: fluorescent channel 1; FL2: fluorescent channel 2. This figure was adapted from ref. [34] with permission; copyright National Academy of Sciences, USA, 2003.]

each bead ends up with roughly 10 million copies of the initial DNA fragment. To perform the sequencing reaction, the DNA template-carrying beads are subsequently loaded into picolitre volume wells etched into the surface of a fibre optic slide with just enough space for one bead. Individual microbeads are then located in wells above an optical fibre that can detect light emitted from the well during the course of a pyrosequencing reaction. This procedure was used to sequence the genome of Mycoplasma genitalium with 96% coverage at 99.96% accuracy in one run. Moreover, ePCR and pyrosequencing were also used, among other things, for the sequencing of mammoth DNA.

In the “multiplex polony sequencing” approach, fragments of DNA that contain short “mate-pairs” of genomic sequence (i.e., short stretches of sequence separated by about 1 kb in the genome) along with universal linker sequences, are amplified on microbeads (~1 μm in diameter) by ePCR to generate microbeads with many single-stranded copies of that fragment. Template-DNA carrying beads are then immobilised in a polyacrylamide gel (this results in a ~ 1.5 cm² monolayer array of disordered, immobilised beads) and mounted in a flow cell to allow automated sequencing-by-ligation and four-colour imaging. This technique was demonstrated by resequencing the genome of an evolved E. coli strain at a fraction of the cost of traditional sequencing techniques.

Current Limitations of Compartmentalisation in Microdroplets

The current limitations of using w/o emulsions to compartmentalize reactions are mostly related to the relatively high polydispersity of the microdroplets. For example, in a directed evolution experiment with IVC, the fact that the droplets are not all the same size means that an identical gene in a different sized droplet can have a different phenotype. This makes it difficult to efficiently select for small differences in activity. Polydispersity also makes it more difficult to achieve conditions in which only a single gene is compartmentalised per droplet: in order to reduce the frequency of droplets that contain more than one gene, the gene concentration must be dropped much lower than would be necessary if the emulsions were monodisperse (in which case genes would be distributed in the droplets as a function of their concentration according to a Poisson distribution). This results in many empty droplets and decreases the selectable repertoire size.

Another limitation is the ability to manipulate the contents of microdroplets after their formation. It is possible to deliver hydrophobic substrates, or ligands, through the oil phase into the water in the droplets and water soluble components can be delivered through nanodroplets, or swollen micelles and the pH can be altered, for example, by the delivery of acetic acid. Substrates or ligands, can be photo-caged, introduced into the droplets during emulsification and remain inert until irradiated. Droplets from two w/o emulsions can also be fused; thus reagents contained in different emulsions can be combined. However, none of these techniques is as universal and controllable as simply being able to pipette reagents together in the well of a microtitre plate.

Finally, it is not possible to measure reaction kinetics in individual droplets in bulk emulsions systems: the screens or selections are based on an end-point assay.

However, as discussed below, we believe that in the future, these problems will be overcome by making and manipulating droplets in microfluidic systems.

Conclusion and Outlook

Compartmentalisation of reactions in microdroplets has already proven very useful for the directed evolution of both enzymes and binding proteins and for high-throughput gene analysis and sequencing. However, the ability to reduce the volume of reaction by a billion fold in comparison to microtitre-plate based approaches (a micrometer diameter droplet has a volume of less than a femtolitre) is likely to find application in a wide range of other domains in which high-throughput and low-reaction volume are needed, for example, drug discovery.

These developments will be helped by recent advances in microfluidics that allow microdroplets to be generated and manipulated in a very controlled and sophisticated manner. Highly disperse single emulsions (< 3% polydispersity) can be created at up to 10000 s⁻¹. Moreover, the resulting aqueous microdroplets can be fused, subdivided, incubated in delay lines and sorted and their contents can be mixed rapidly. Enzymatic reactions can be triggered at defined times in these systems, and the reaction kinetics can be determined by measuring fluorescence at multiple points, which correspond to different times, in the microfluidic system.

High-fidelity manipulation of microdroplets in microfluidic channels combined with the ability to perform chemical and biological experiments in microdroplets offers the possibility of developing powerful automated instruments. These new instruments should provide an unprecedented level of control and ultrahigh-throughput screening that can be applied across many branches of experimental science.

Acknowledgements

This research was supported by the Medical Research Council and by funding from the EC Framework 5 Marie Curie program (the group belongs to the “ENDIRPRO” Training Network). Funding from the CNRS and the Ministère de la Recherche et de l’Education is also gratefully acknowledged.

Keywords: directed evolution · high-throughput screening · in vitro compartmentalization · polymerase chain reaction