TURUN YLIOPISTON JULKAISUJA
ANNALES UNIVERSITATIS TURKUENSIS

SARJA - SER. A I OSA - TOM. 480

ASTRONOMICA - CHEMICA - PHYSICA - MATHEMATICA

Methods of genetic diversity creation and functional display for directed evolution experiments

by

Tuomas Huovinen

TURUN YLIOPISTO
UNIVERSITY OF TURKU
Turku 2013

From

Department of Biochemistry/ Biotechnology
University of Turku
Turku, Finland

Supervised by

Urpo Lamminmäki, Ph.D.
Department of Biochemistry/ Biotechnology
University of Turku
Turku, Finland

and

Petri Saviranta, Ph.D.
Medical Biotechnology
VTT Technical Research Centre of Finland
Turku, Finland

Reviewed by

Professor Kalle Saksela, M.D., Ph.D.
Department of Virology
Haartman Institute
University of Helsinki
Helsinki, Finland

and

Professor Mauno Vihinen, Ph.D.
Department of Experimental Medical Science
Lund University
Lund, Sweden

Opponent

Associate Professor Peter Kristensen, Ph.D.
Department of Engineering
Aarhus University
Aarhus, Denmark

ISBN 978-951-29-5625-8 (PRINT)
ISBN 978-951-29-5626-5 (PDF)
ISSN 0082-7002
Painosalama Oy – Turku, Finland 2013

That's all it takes really,

pressure, and time.

- Ellis Boyd Redding

in *Shawshank Redemption*

CONTENTS

LIST OF PUBLICATIONS

This thesis is based on the following publications, referred to in the text by their Roman numerals (I-V):

I Huovinen T., Brockmann E.C., Akter S., Perez-Gamarra S., Ylä-Pelto J., Liu Y., Lamminmäki U. (2012) Primer extension mutagenesis powered by selective rolling circle amplification. PLoS One 7: e31817.

II Huovinen T, Julin M, Sanmark H, Lamminmäki U. (2011) Enhanced error-prone RCA mutagenesis by concatemer resolution. Plasmid 66(1): 47-51.

III Huovinen T., Sanmark H., Ylä-Pelto J., Vehniäinen M., Lamminmäki U. (2010) Oligovalent Fab display on M13 phage improved by directed evolution. Molecular Biotechnology 44(3): 221-31.

IV Huovinen T., Syrjänpää M., Sanmark H., Brockmann E.C., Azhayev A., Wang Q., Vehniäinen M. and Lamminmäki U. (2013) Two scFv antibody libraries derived from identical VL-VH framework with different binding site designs display distinct binding profiles. Protein Engineering Design and Selection 26(10) 683-93.

V Huovinen T., Syrjänpää M., Sanmark H., Seppä T., Akter S., Khan L. and Lamminmäki U. (2013) Selection performance of an antibody library displayed on filamentous phage coat proteins 9, 3 and truncated 3. Manuscript.

ABBREVIATIONS

A	adenosine (IUPAC), also common abbr. for adenine
ampR	resistant to ampicillin
bp	base pair
BSA	bovine serum albumin
C	cytidine (IUPAC), also common abbr. for cytosine
cDNA	complementary DNA
CDR	complementary determining region
ctzR	resistant to ceftazidime
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dGTPαS	deoxyguanosine 1-thiotriphosphate
DIG	digoxigenin
dITP	deoxyinosine triphosphate
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide
dPTP	5'-triphosphate of 6-(2-deoxy-b-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one
dsDNA	double-stranded DNA
DTPA	diethylenetriamene penta-acetate
dTTP	deoxythymidine triphosphate
dUMP	deoxyuridine monophosphate
dUTP	deoxyuridine triphosphate
ELISA	enzyme-linked immunosorbent assay
EMS	ethyl methanesulfonate

epPCR	error-prone PCR
epRCA	error-prone RCA
F	5-fluorodeoxycytidine
Fab	antigen binding fragment
FACS	fluorescence-activated cell sorting
G	guanosine (IUPAC), also common abbr. for guanine
GC	gas chromatography
IgG	immunoglobulin G
IVC	*in vitro* compartmentalization
IPTG	isopropyl-β-D-thiogalactopyranoside
IUPAC	International union of pure and applied chemistry
MS	mass spectrometry
MMS	methyl methanesulfonate
MNNG	N-methyl-N'-nitro-N-nitrosoguanidine
mRNA	messenger RNA
mut/ntR	mutations per nucleotide per reaction
p3	phage coat protein three; minor coat protein of filamentous phage also known as g3p, pIII
p3Δ	C-terminal domain of p3; also known as $p3^{CT}$
p6	phage coat protein six, also known as pVI
p7	phage coat protein seven, also known as pVII
p8	phage coat protein eight; major coat protein of filamentous phage also known as pVIII
p9	phage coat protein nine, also known as g9p, pIX
PCR	polymerase chain reaction
RAM	rabbit anti-mouse IgG

RCA	rolling circle amplification
RNA	ribonucleic acid
ScFv	single-chain fragment of antibody variable domains
SD	standard deviation
SEC	post-translational secretion pathway
SGS	second generation sequencing, also known as NGS
SOE-PCR	splicing by overlap extension PCR
sRCA	selective rolling circle amplification
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
STR	streptavidin
T	thymidine (IUPAC), also common abbr. for thymine
TAT	twin-arginine translocation
Ti/Tv	transitions-to-transversions-ratio
TRIM	trinucleotide mutagenesis
U	uridine (IUPAC), also common abbr. for uracil
UDG	uracil-DNA-glycosylase
UV	ultraviolet radiation
VH	variable heavy domain of antibody
VL	variable light domain of antibody
Vλ	variable lambda domain of the light chain of antibody
Vκ	variable kappa domain of the light chain of antibody

ABSTRACT

Protein engineering aims to improve the properties of enzymes and affinity reagents by genetic changes. Typical engineered properties are affinity, specificity, stability, expression, and solubility. Because proteins are complex biomolecules, the effects of specific genetic changes are seldom predictable. Consequently, a popular strategy in protein engineering is to create a library of genetic variants of the target molecule, and render the population in a selection process to sort the variants by the desired property. This technique, called directed evolution, is a central tool for trimming protein-based products used in a wide range of applications from laundry detergents to anti-cancer drugs. New methods are continuously needed to generate larger gene repertoires and compatible selection platforms to shorten the development timeline for new biochemicals.

In the first study of this thesis, primer extension mutagenesis was revisited to establish higher quality gene variant libraries in *Escherichia coli* cells. In the second study, recombination was explored as a method to expand the number of screenable enzyme variants. A selection platform was developed to improve antigen binding fragment (Fab) display on filamentous phages in the third article and, in the fourth study, novel design concepts were tested by two differentially randomized recombinant antibody libraries. Finally, in the last study, the performance of the same antibody repertoire was compared in phage display selections as a genetic fusion to different phage capsid proteins and in different antibody formats, Fab vs. single chain variable fragment (ScFv), in order to find out the most suitable display platform for the library at hand.

As a result of the studies, a novel gene library construction method, termed selective rolling circle amplification (sRCA), was developed. The method increases mutagenesis frequency close to 100% in the final library and the number of transformants over 100-fold compared to traditional primer extension mutagenesis. In the second study, Cre/loxP recombination was found to be an appropriate tool to resolve the DNA concatemer resulting from error-prone RCA (epRCA) mutagenesis into monomeric circular DNA units for higher efficiency transformation into *E. coli*. Library selections against antigens of various size in the fourth study demonstrated that diversity placed closer to the antigen binding site of antibodies supports generation of antibodies against haptens and peptides, whereas diversity at more peripheral locations is better suited for targeting proteins. The conclusion from a comparison of the display formats was that truncated capsid protein three (p3Δ) of filamentous phage was superior to the full-length p3 and protein nine (p9) in obtaining a high number of uniquely specific clones. Especially for digoxigenin, a difficult hapten target, the antibody repertoire as ScFv-p3Δ provided the clones with the highest affinity for binding.

This thesis on the construction, design, and selection of gene variant libraries contributes to the practical know-how in directed evolution and contains useful information for scientists in the field to support their undertakings.

TIIVISTELMÄ

Proteiinien muokkauksella tähdätään entsyymien ja sitojareagenssien ominaisuuksien parantamiseen geneettisten muutosten avulla. Tyypillisiä parannettavia ominaisuuksia ovat sitomisvoimakkuus, spesifisyys, kestävyys, tuotto-ominaisuudet ja liukoisuus. Koska proteiinit ovat monimutkaisia biomolekyylejä, geneettisten muutosten vaikutukset ovat vain harvoin tarkkaan ennustettavissa. Siksi suosittu proteiinien muokkausstrategia on luoda kohdemolekyylistä lukuisia geenivariantteja ja asettaa luotu joukko valintakokeeseen, jonka perusteella variantit erottuvat toisistaan tavoitellun ominaisuuden perusteella. Tämä suunnattuna evoluutiona tunnettu tekniikka on keskeinen työkalu kehitettäessä proteiinituotteita, joita käytetään monenlaisissa sovelluksissa vaatteiden pesuaineista syöpälääkkeisiin. Proteiinituotteiden kehitystyön nopeuttaminen edellyttää jatkuvasti uusia menetelmiä, joilla voidaan rakentaa aiempaa laajempia geenikirjastoja ja niille yhteensopivia valintatyökaluja.

Tässä työssä tutkittiin alukepidennysmutageneesitekniikan mahdollisuuksia korkealaatuisempien geenivarianttikirjastojen rakentamiseksi *Escherichia coli*-bakteerin soluihin. Toisessa osajulkaisussa tutkittiin rekombinaatiota menetelmänä, jolla voitaisiin lisätä seulottavissa olevien entsyymivarianttien lukumäärää. Kolmannessa osajulkaisussa kehitettiin valintaprosessi, jonka tarkoituksena oli parantaa vasta-ainefragmenttien ilmentymistä filamenttifaagin pinnalla näyttötekniikkaa varten. Neljännessä osajulkaisussa testattiin geenikirjaston suunnittelustrategioita käytännössä kahdella eri periaatteiden mukaan monimuotoistetulla vasta-ainekirjastolla. Viimeisessä osajulkaisussa vertailtiin faagin eri pintaproteiineihin fuusioidun vasta-ainekirjaston toimintaa valintakokeiden avulla. Samalla tutkittiin vasta-ainefragmenttien Fab (*engl.* antigen binding fragment) ja ScFv (*engl.* single-chain fragment of antibody variable domains) vaikutusta valintakokeen onnistumiseen, jotta selviäisi, mikä on käyttökelpoisin näyttötekniikka käytössä olevan kirjaston hyödyntämiseksi.

Tutkimuksen tuloksena kehitettiin uusi geenivarianttikirjaston rakennusmenetelmä nimeltään sRCA (*engl.* selective rolling circle amplification), joka lisää merkittävästi mutageneesitehokkuutta ja jopa yli satakertaistaa kirjaston muodostavien transformanttien määrän verrattuna tavanomaiseen alukepidennysmutageneesiin. Toisessa osajulkaisussa havaittiin Cre-loxP-rekombinaation olevan sovelias työkalu RCA-satunnaismutageneesin tuloksena syntyvän DNA-vyyhdin pilkkomiseen kehämäisiksi plasmidiyksiköiksi. Uusi menetelmä lisäsikin DNA:n transformaatiotehokkuutta *E. coli*-bakteeriin. Kirjastoseulontojen avulla osoitettiin, että kirjastosta, jossa aminohappojen vaihtelua esiintyi lähempänä vasta-aineen sitomiskohdan keskustaa, löytyi enemmän pienmolekyylejä ja peptidejä tunnistavia vasta-aineita. Ulompana keskustasta sijaitsevien aminohappojen vaihtelu puolestaan tuki proteiineja tunnistavien vasta-aineiden kehitystyötä. Vasta-ainekirjaston faaginäyttötekniikkavertailun johtopäätökset olivat, että käyttämällä filamenttifaagin vasta-aineiden fuusiokumppanina lyhennettyä proteiini kolmea (p3Δ) pystyttiin eristämään runsaammin erilaisia kohdetta tunnistavia vasta-ainemolekyylejä kuin käyttämällä kokopitkää proteiini kolmea tai proteiini yhdeksää (p9). Erityisesti ScFv-p3Δ-formaatissa oleva kirjasto tarjosi sitomisvoimakkuudeltaan parhaita vasta-aineita molekyylikooltaan pienen digoksigeniinin tunnistukseen.

Tutkimus geenivarianttikirjastojen suunnittelusta, rakentamisesta ja seulontaan soveltuvista valintatyökaluista lisää käytännön tietoa suunnattujen evoluutiokokeiden toteuttamiseksi ja on arvokasta tietoa alan tutkijoille heidän tulevissa hankkeissaan.

1 INTRODUCTION

Cultivation of crop species has been practiced by humans for over 10 000 years and the history of selective breeding is equally long [1]. Domestication has gradually resulted in the variety of plant and animal species that we use at present to acquire food and clothes. Selective breeding is directed evolution, in which the desired genetical traits are favoured and the undesired disfavoured by allowing the progeny of the individuals carrying the desired phenotype to reproduce. Some traits were intentionally evolved, such as the size of a grain, whereas some other traits, e.g., the loss of seed shattering, seem to have evolved without intentional human steering [1; 2]. Similar selective practices have been applied to evolve micro-organisms for food processing and medicine manufacturing, of which for example antibiotics, are well-known and of major importance to the welfare of humans and their domesticated companions.

In the early days, microbial production strains were improved without any knowledge of the underlying genetic features. This is a slow process due to the low frequency of new beneficial mutations introduced by nature. Towards modern times, the pace of evolution has been increased by creating genetical diversity with chemicals and irradiation followed by the selection for improved variants. However, the two most important innovations contributing to directed evolution were the establishment of modern recombinant DNA technology and the in-depth knowledge of genes and proteins. Nowadays, instead of whole organisms, individual genes responsible for certain traits are manipulated providing a knowledge-based approach for strain improvement.

From the mid-1980s, proteins of mammalian origin have been expressed with high yields in micro-organisms or production cell lines by recombinant DNA technology [3]. Therapeutics produced by heterologous protein expression have provided a cost-effective or, in some cases, even the only cure for treatment of various human diseases and the share of recombinantly produced protein medicines is predicted to expand further in the future [3]. In addition to pure rational design of gene sequences, collections of synthetic gene variants are routinely generated in laboratories, and the characteristics of the variants are systematically explored. The collections of gene variants, also termed gene libraries, have become a central tool in protein engineering.

Gene libraries are mined either by expressing each variant individually in host cells and analyzing the phenotypic features with an appropriate assay or by applying a display method (**Figure 1**). In display methods, every expressed protein variant is physically linked to the corresponding gene and the expansion of the desired individuals is promoted by exerting an appropriate selection pressure on the pool of displayed geno-phenotype units. When the share of the desired geno-phenotype units is enriched to a satisfying degree in the population, the properties of the individuals are analysed in detail. With display methods, the repertoires of billions of members can be explored, whereas conventional screening efforts are limited to a few thousand members. In this thesis, the

methods for gene library construction and use are reviewed. The results of the present work are novel tools for gene library construction and contain detailed descriptions of the potential and limitations of display methods in search for desired gene products. Especially issues related to the phage display of antibody fragments are discussed.

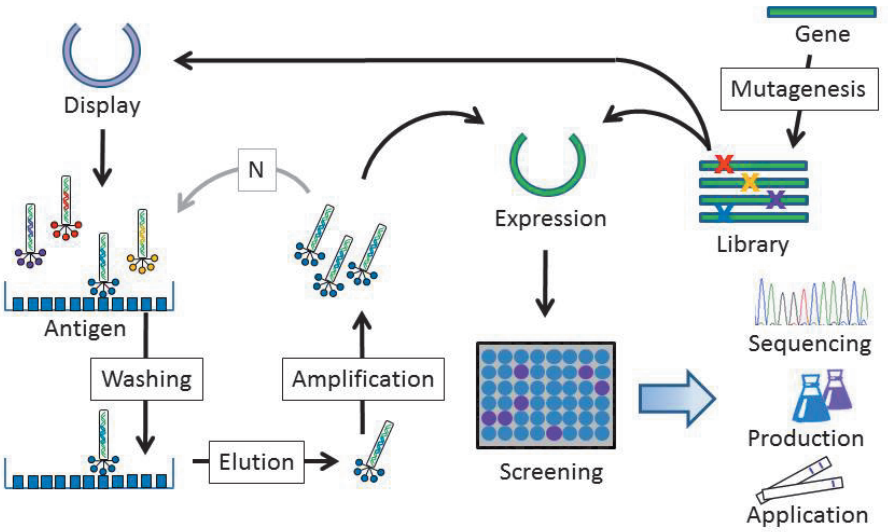

Figure 1. Typical workflows in directed evolution experiments. Gene variants are created with mutagenesis methods and the desired variants in the population are enriched by a selection process, such as phage panning. Alternatively, library clones are directly separated into individual cultures, expressed, and screened for the desired property.

2 REVIEW OF THE LITERATURE

2.1 Methods for gene variant library construction

2.1.1 Random mutagenesis

2.1.1.1 Time

DNA sequences continuously change through spontaneous mutations over time. Single base mutations arise naturally by chemical conversion and replication defects, whereas insertions and deletions are most often related to transposable elements [4]. In one long-term study a laboratory strain of *E. coli* REL606 was cultured in minimal growth medium over 6000 days spanning 40 000 generations [5]. During that time the strain had gained 627 single-nucleotide mutations and 26 deletions, insertions or inversions.

With this mutation rate a single mutation will appear somewhere in the genome in 9 days of continuous culturing. Assuming uniform distribution of mutations and taking into account the genome size of the strain, 4.57×10^6 bp [5], over 115 years is needed to obtain a single mutation per 1000 bp. Because the gene density of *E. coli* is 0.911 ± 0.04 genes per 1000 bp [6], even longer time is required to have a mutation in every gene. It is obvious that the spontaneous mutation rate in *E. coli* is too slow for the needs of protein engineering. Notably, the rate is species dependent as over 100-fold higher mutation rates per genome per replication have been observed in lytic ssRNA viruses than in *E. coli*, which is mainly due to a lack of proof-reading activity of the RNA polymerase employed by the virus [7].

2.1.1.2 Mutator strains

In a closer analysis of the point-mutation rates in the above-mentioned follow-up study on *E. coli* strain REL606 it was noticed that the point-mutation rate increased 70-fold from a natural rate of 1.6×10^{-10} (estimated upper bound) to 1.1×10^{-8} mutations per bp per generation. This change was connected to the emergence of a mutator phenotype around the generation 26 500, which became dominant in the population before the strain achieved 29 000 generations [5]. The increased mutation rate was due to a defect in the *mutT* gene, which is coding for a nucleoside triphosphate pyrophosphohydrolase [8]. This enzyme selectively hydrolyzes 8-oxo-dGTP preventing the misincorporation of 8-oxo-dGTP into dsDNA [9]. 8-oxo-dGTP pairs readily with adenine causing T-A to G-C

transversions and without *mutT* function the transversion rate is increased 100 – 10 000-fold depending on the assay set-up [10; 11]. 8-oxo-dGTP is a by-product of normal aerobic metabolism [12] and, therefore, all organisms have evolved a multitude of damage prevention and repair mechanisms to protect themselves from the damaging influence of this and several other natural products on the integrity of DNA [4].

Knock-outs of the error-reducing mechanisms have been used to speed up the rate of spontaneous mutations. For example, the *E. coli* strain XL1-Red provided commercially by Agilent Technologies has disabling mutations in the above-mentioned gene *mutT*, but also in *mutS* involved in error-prone mismatch repair [13] and in *mutD* making the DNA polymerase III deficient in 3´-5´ exonuclease activity [14]. The mutation rate of this strain was measured to be 5000-fold higher than that of the wild type, corresponding to 1 x 10^{-5} mutations per base pair per generation, i.e. 46 mutations per genome per generation [15; 16]. In a gene mutagenesis experiment, where a target gene of 1000 bp would be inserted into a pUC-series of high copy plasmids with 500 - 700 copies per cell [17], a mutation in five gene copies would be expected in every generation, when propagated in XL1-Red.

In another example, a phage display-compatible *mutD* knock-out strain was developed and used to affinity maturate a ScFv in a high copy phagemid pHEN1 [18]. The mutation rate in this single knock-out was estimated to be 1.7 x 10^{-5} mutations per bp per generation which is on the same scale with the mutation rate of XL-1 Red triple mutant. With multiple rounds of growth in the *mutD* strain and subsequent affinity selections, an anti-hapten ScFv with a 100-fold increase in affinity for the antigen 2-phenyl-5-oxazolone was obtained. Mutator plasmids are another way to induce random mutations. For example, the expression of a nonfunctional *mutD* allele from a plasmid in the host to be mutated resulted in the replacement of the host-expressed wild-type MutD protein in the DNA polymerase III complex and, consequently, to a higher error rate in replication [19].

The disadvantage of the mutator strains is their slow growth rate, as they have a doubling time of 90 – 120 minutes [15; 16]. In addition, the instability of the mutator genome renders long-term maintenance and storage troublesome. From protein engineering point of view, as the mutations are not limited only to the target gene, additional unwanted mutations may generate false positive signal to appear in the following screening step or mask the beneficial effect of a favourable mutation.

2.1.1.3 Mutagens

Before the discovery of PCR techniques random mutations were induced by exposing cells to mutagenic conditions. Random mutagenesis by mutagens is still used in strain development when adapting strains to cultivation conditions [20] or

increasing yields of valuable substances [21]. The list of used mutagens is long and diverse including nitrous acid [22], sodium bisulfite [23; 24], methoxylamine [25] and hydroxylamine [26], to mention some, but the most popular methods in biotechnology are irradiating a strain with ultraviolet light [20], and exposing cells to alkylating agents such as N-methyl-N'-nitro-N-nitrosoguanidine (MNNG or NTG) [27] or ethyl methanesulfonate (EMS) [21; 28].

A practical point of view is to classify the mutagens according to the mechanism of action. Intercalating agents such as proflavin and acridine compounds cause predominantly frameshifts [29], which are seldom beneficial in gene engineering. More useful are different kinds of base modifications resulting in base pair substitutions. Alkylating agents MNNG and methyl methanesulfonate (MMS) are directly mutagenic by transferring methyl-groups on the bases found in dsDNA. For example, O6-methylguanine favourably base pairs with thymine and O4-methylthymine with guanine, causing G-C to A-T and T-A to C-G transitions, respectively [30; 31; 32]. Also ethylating agents, such as EMS, cause the same mispairing, and the mutagenic potency of the three substances increases in the order MMS>MNNG>EMS, correlating with the ability to alkylate the O6 atom of the guanine base [33; 34]. Secondary mutations are also possible with alkylating agents through the launch of the error-prone SOS response [35].

The non-alkylating chemical mutagens hydroxylamine and methoxylamine induce base substitutions through altered base paring in the same way as EMS and MMS, but without bulky side chains. Hydroxylamine and methoxylamine modify cytosine and adenine to N-hydroxy- and N-methoxy-derivatives [36] of which, for example, N6-methoxyadenine pairs with uracil or cytosine depending on the tautomeric form [37; 38].

The loss of the primary amine group of cytosine by deamination is a natural phenomenon with a half-life of 200 years on single-stranded DNA (ssDNA) and 20 000 – 30 000 years on double-stranded DNA (dsDNA) [39; 40]. Deamination results in the formation of uracil, which preferably forms a base pair with adenine causing a C-G to T-A transition [4]. The deamination rate is enhanced by nitrous acid [22; 41], bisulfite [42], and UV irradiation [40]. The methylation of cytosine residues accelerates deamination tenfold and as the deaminated form of 5-methyl-cytosine is the standard base thymine, the change remains undetected by the deamination control surveillance centered around uracil detection [43]. Bisulfite converts cytosine, but not 5-methyl-cytosine to uracil [44] and, therefore, instead of mutagenesis, it has found major use as a tool for mapping methylation sites in genome research [45]. Nitrous acid also deaminates adenine into hypoxanthine, which forms a base pair preferably with cytosine causing A-T to G-C transition [46].

High-energy ionizing radiation such as X-rays, γ rays and α particles are notorious for their ability to damage DNA, but as they may cause danger also to the operator, they are rarely used in biotechnological applications. Instead UV light is a very popular tool to induce random mutations. UV induces the formation of a pyrimidine dimer, which can lead to transitions, transversions,

frameshifts and deletions due to mispairing or due to errors in subsequent repair response [47]. Although a TT dimer is the most abundant aberration induced by UV [48], it only rarely causes mutations as DNA polymerase preferably incorporates A opposite a damaged strand [49]. Because of the inherent A-rule of the DNA polymerase, a CC dimer is frequently changed to a TT dimer. Single C-to-T transitions are encountered as well due to the accelerated deamination rate of cytosines in cyclobutane, i.e. CC dimer, structures [40]. Another mechanism of mutagenesis is that UV radiation generates free radicals, which oxidate guanine to 8-oxo-guanine, which in turn is able to pair with adenine inducing G-C to T-A transversion [50]. However, G-to-T transversions are not very common compared to C-to-T transitions [51].

As already highlighted, different mutagens cause different kinds of changes in DNA (**Table 1**). Higher efficiency mutagenesis and more diverse changes in DNA sequence are achieved by method combinations. For example, a base excision repair deficient *E. coli* strain is significantly more sensitive to nitrous acid than the wild-type strain [52]. Both simultaneous and sequential mutagenesis methods have been applied to develop production strains in the desired direction [53; 54]. Even the implementation of four sequential mutagenesis treatments was reported consisting of UV, low-energy ion beam, atmospheric pressure non-equilibrium discharge plasma, and exposure to MNNG in a study to increase the cellulase production in a *Trichoderma viride* strain [55]. The power of chemical mutagenesis can be further augmented by genome shuffling, in which beneficial mutations from different parental genotypes are brought together via protoplast fusion [56].

The benefit of chemical and physical mutagens is that no prior knowledge of the genetics of the trait to be improved is needed for mutagenesis. Some traits may need simultaneous modification of several genes, which may have very complex interaction networks with each other. In random genome mutagenesis all genes are equally exposed to the mutagens. These agents are simple to use and economically feasible to any laboratory. There are however many limitations in these techniques. Some mutagens, e.g., MNNG, MMS, EMS, and UV, are effective directly *in vivo* but others, e.g., hydroxylamine [26], bisulfite [42], and nitrous acid [22], can be used only with extracted DNA. Different species may also respond differently to the same mutagen. For example, a UV dose at 80 J/m^2 destroys over 99.9% of treated *E. coli* cells, but less than 20% of cyanobacteria cells of the *Anabaena* genus [57].

Table 1. Examples of chemical mutagens.

Mechanism	Representative agent	Typical mutation	Cell penetration
Intercalation	Acridine orange	Frameshift	Yes
Alkylation	Ethyl methanesulfonate	G-C to A-T	Yes
Methoxylation (tautomeric shifting)	Methoxylamine	A-T to G-C	No
Deamination	Nitrous acid	C-G to T-A & A-T to G-C	No
Dimerization	UV irradiation	C-to-T & CC-to-TT	Yes

2.1.1.4 Error-prone reactions with thermophilic enzymes

Chemical mutagenesis is exerted on the whole genome in intact cells or on the extracted transformable DNA element *in vitro*. Most often variants of only a single gene are desired in protein engineering, which is difficult to achieve with chemical mutagens. There are some reported attempts to restrict the effect of the mutagen to a preselected stretch of DNA. In one technique published as "Dual Approach to Random Chemical Mutagenesis" EMS mutagenesis of a plasmid was followed by a restriction enzyme digestion to cleave the target gene, after which it was inserted into a fresh vector for transformation [58]. In another study, sodium bisulfite treatment of a template was combined with a PCR to amplify the target gene containing novel mutations for cloning into a new vector [59]. These are interesting approaches, but still rare examples in the modern times dominated by error-prone PCR.

In error-prone PCR (epPCR), as the name implies, the fidelity of DNA synthesis is intentionally decreased. The error rate of the naturally 3′-5′ exonuclease-deficient, i.e. non-proofreading, Taq DNA polymerase is 2.0×10^{-5} per nucleotide per cycle, which is tenfold higher than the error rate of Pfu DNA polymerase (1.6×10^{-6} per nucleotide per cycle), latter of which is often preferred in molecular cloning because of its higher fidelity [60]. Due to these reasons the early epPCR development was centered on Taq DNA polymerase. The easiest way to decrease fidelity is to add Mn^{2+} which reduces base pairing specificity [61]. A 25-fold increase in error rate was documented by adding 0.7 mM $MnCl_2$ in a

polymerization reaction catalyzed by *E. coli* DNA polymerase I [61], and a similar phenomenon has been observed and widely applied in PCR reactions.

Changing the stoichiometric proportions of dNTPs is another simple way to increase misincorporation. Biased dNTP pools are often used in combination with $MnCl_2$ [62; 63] and, therefore, no exclusive report exists on the true effect of $MnCl_2$ on Taq DNA polymerase fidelity. Instead, the combination of $MnCl_2$ with different stoichiometric proportions of dNTPs has been a subject of vivid discussion in the field of protein engineering [63; 64; 65], the debated issues being the frequency and quality of the mutations.

There is a distinct bias in the type of errors naturally produced by Taq DNA polymerase. Pyrimidine-purine mismatches are more readily incorporated than pyrimidine-pyrimidine or purine-purine mismatches, which is a universal bias in both family A (e.g., Taq DNA polymerase) and family B (e.g., phi29 DNA polymerase) DNA polymerases, leading to a dominant share of transition substitutions [66; 67]. In a nonerror-prone PCR with Taq polymerase and 1 mM dNTPs, the transitions-to-transversions-ratio (Ti/Tv) was 2.6 [66]. In pro-mutagenic conditions containing 20 µM of each dNTP and 0.25 mM $MnCl_2$ the mutagenesis frequency was doubled from the initial 0.0008 to 0.0017 mutations per nucleotide per reaction (mut/ntR), although the Ti/Tv-ratio was essentially unaltered (2.9) [68]. A closer scrutiny of the individual substitutions revealed that two of the four possible transition mutations, namely A-T to G-C and T-A to C-G, covered 63.2% of all mutations and the lowest frequency transversions, G-C to C-G and C-G to G-C, were present only with a 1.4% share [68].

A study on the effect of nucleotide imbalances on the fidelity of T4 DNA polymerase replication had a major influence on the use of high dGTP/dATP-ratio to promote mutagenesis [64]. In this early and excellent study the occurrence of a substitution was deciphered from the reversion of a non-infective phenotype of Φ-X174 bacteriophage back to infective in a non-supressor *E. coli* strain by abolishment of an amber stop codon. Unfortunately, the assay was able to respond to T-A to C-T and A-T to G-C substitutions, but not to the occurrence of C-G to T-A and G-C to A-T transitions. The latter transitions could not be recorded, because there is no C in TAG codon and because changing G for A generates the ochre stop codon TAA.

Against this background it is understandable that epPCR was reported to have been used with dATP/dTTP/dCTP/dGTP concentrations 0.2/1/1/1 mM with a 0.5 mM $MnCl_2$ supplement resulting in 0.0137 ± 0.0029 mut/ntR [63; 65], although a fairly similar mutation rate of 0.0085 ± 0.0043 mut/ntR was achieved with dATP/dTTP/dCTP/dGTP concentrations 1/1/1/0.2 mM [63], a ratio which is exactly the opposite of the earlier recommendation of high dGTP/dATP [64]. Naturally, precaution is needed when comparing mutation frequencies per reaction across studies as the frequency is dependent on the number of cycles implemented. The major difference between the high dGTP/dATP and high dATP/dGTP is the Ti/Tv-ratio, which was the usual 2.7 in the former and only 0.4 in the latter case. The authors of these findings argumented, however, that the

high dATP/dGTP-ratio resulted in a strong mystical bias for substituting A in the template for something else (not shown in detail) and hence, for balanced mutagenesis dATP/dTTP/dCTP/dGTP concentrations of 0.2/1/1/0.2 mM were recommended with a Ti/Tv-ratio of 0.8 and a mutagenesis rate of 0.007 mut/ntR [63]. After all, avoiding an excess of dGTP is crucial as it seems that the predominant transitions are directly linked to the G:T mismatch.

In a reported epPCR method titled "hypermutagenic PCR", a 100-fold increase in mutation frequency was obtained by pushing the dNTP bias a bit further in a Taq DNA polymerase driven reaction with 0.5 mM $MnCl_2$ with 30 µM dATP, 1 mM dTTP, 30 µM dCTP and 1 mM dGTP [62]. The enhancement of the mutation rate from 0.001 to 0.1 mut/nt was dependent on the addition of $MnCl_2$, but there were still more transitions than transversions with a Ti/Tv-ratio of 2.2 with two dominating mutations, A-T to G-C and T-A to C-G, accounting for 61% of all the changes. What was achieved with the high dGTP/dCTP-bias, not experimented in the earlier studies, was an increased bypass of the G:G mismatch from 0.7% [68] to 2.5% [62], which is claimed to be one of the most substantial blocks for the Taq DNA polymerase driven polymerization [69].

As a summary, with imbalanced dNTP pools it is possible to modify the mutational bias. A DNA polymerase with no bias would generate errors with a Ti/Tv-ratio of 0.5 due to four possible transitions and eight possible transversions. This ratio is hardly ever achieved with Taq DNA polymerase in any conditions, and biasing the dNTP pools to the extreme leads to low PCR product yields [63; 70]. Other methods that can be used to increase error rate are to add high doses of $MgCl_2$, to change the pH, to overload DNA polymerase [71], and, naturally, to decrease the template amount and to increase the number of cycles in PCR.

Due to the error preferences of Taq DNA polymerase for A-to-T and T-to-A transversions (and transition mutations in general) adenosine and thymidine nucleotides are four times more likely to be mutated compared to guanosine and cytidine [72]. This skew was targeted with a modified Pfu DNA polymerase lacking the 3´-5´ exonuclease domain [73]. The preferences of this enzyme called mutazyme DNA polymerase are opposite to Taq DNA polymerase as it mutates guanosine and cytidine nucleotides three times more frequently than adenosine and thymidine [73]. Logically, to achieve a more balanced mutagenesis the two error-prone enzymes were combined in the newest mutazyme-containing commercial formulation [74].

Mutazyme (exonuclease-deficient Pfu DNA polymerase) exhibits a 40-fold higher error rate than the native enzyme with transitions as the main substitution type [73; 75]. The processivity of the Mutazyme is increased in the commercial formulations by adding a polymerization enhancing factor in the reaction [76]. The major polymerization enhancing factor of Pfu is a protein called P45, which is a dUTPase that catalyses the degradation of dUTP to dUMP and pyrophosphate [77]. The degradation of dUTP enhances productivity, because

most thermophilic DNA polymerases are inhibited by uridine containing template DNA [78; 79].

Another route to decrease the fidelity of an error-prone DNA polymerase is to genetically modify the dNTP binding site of the polymerase. In one study, the proof-reading ability of Pfu DNA polymerase was first abolished with a point mutation D215A in the exonuclease domain, after which another mutation, D473G in the dNTP binding site, increased the error rate to 7 x 10^{-4} per nucleotide per cycle, which is 500 times higher than in the wild type [80]. By comparison to the exonuclease-deficient enzyme and Taq DNA polymerase reactions in modified reaction conditions, the double mutant was claimed to produce the most unbiased set of random mutations [80]. In addition to Pfu DNA polymerase, the fidelity of Taq DNA polymerase has been genetically engineered, and a single substitution in the dNTP binding site resulted in over 20-fold higher error rate in PCR compared to the wild type [81].

Nucleotide analogs may also be used to produce random mutations in a DNA polymerase catalyzed reaction. Especially the 5'-triphosphate of 6-(2-deoxy-b-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one (dPTP) and 8-oxo-dGTP are efficiently incorporated by Taq DNA polymerase with distinct mutation profiles [82]. P is a bicyclic base analogue, which pairs with adenine or guanine creating transition mutations [83]. 82% of the dPTP-generated mutations in an epPCR reaction with Taq DNA polymerase were A-to-G and T-to-C changes that arise from the preferential incorporation of dPTP opposite adenosine and subsequent pairing with an incoming dGTP [82]. While only 17% of mutations were G-to-A and C-to-T mutations, altogether transitions covered over 99% of all the observed mutations showing the high selectivity of the P base pairing. 8-oxo-dGTP induced only transversions of which T-to-G and A-to-C changes covered 98% of all the mutations arising from the misincorporation of 8-oxo-dGTP opposite adenine which upon replication is substituted with cytosine [82].

A combination of base analogs dPTP and 8-oxo-dGTP would balance the mutational outcome. However, when this strategy was implemented, the mutation profile mostly resembled pPTP fingerprint due to the lower incorporation efficiency of 8-oxo-dGTP [82]. There are also other base analogs that have been used in mutagenesis, e.g., dITP [84], N4-amino-dCTP [85], 2-OH-dATP [86] and 2-OH-dGTP [87], but the reported mutagenesis frequencies in Taq DNA polymerase driven PCR (mut/ntR) with these analogs are less than 1/10 of the mutagenesis frequencies obtained with dPTP and 8-oxo-dGTP [84; 86; 87].

Technically more complicated variations of mutagenesis by thermophilic DNA polymerase have also been developed. One of them, called SeSaM, begins by shearing template DNA into ssDNA fragments [88]. Shearing is achieved by synthesizing template DNA in the presence of thiol-modified dGTP (dGTPαS), followed by iodine treatment cleaving the sugar-phosphate backbone at the sites of the sulfur [88]. Terminal deoxynucleotide transferase catalyses the extension of the ssDNA fragments with one or several universal nucleotides, e.g., dPTP,

that are subsequently used as megaprimers and extended to full-length genes [88; 89]. At the end of the process, the universal bases are replaced by natural bases in a final PCR reaction. The advantages of the SeSaM-method are the increased rates of transversion and semi-controlled mutagenesis of consecutive bases, the latter being a rare event in conventional epPCR. The general methods for increasing error rate in PCR are summarized in **Table 2**.

Table 2. Common procedures for increasing error rate in PCR.

Method	Reference
Imbalanced dNTP pools	[62], [63], [68], [65]
Mn^{2+} supplementation	[62], [63], [68], [65]
Modified DNA polymerase	[73], [80], [81]
Nucleotide analogs	[82], [84], [86], [87]
Less template, more cycles	common knowledge

2.1.1.5 Error-prone reactions with mesophilic enzymes

There are DNA polymerases in higher eukaryotes, which are error-prone by nature. Human DNA polymerase β belongs to the family X of DNA polymerases and is involved in the base excision repair pathway in man [90]. As it produces 10^{-3} to 10^{-4} mutations per nucleotide [91], in one study, random mutations were introduced into antibody genes with a single pass of replication with DNA polymerase β [92]. In a later study, another naturally error-prone human DNA polymerase was used to randomize an amylosucrase gene [93]. The employed enzyme, DNA polymerase η (eta), belongs to the DNA polymerase family Y that exhibit the highest error rates ever recorded of naturally occurring DNA polymerases, namely 10^{-1} to 10^{-3} mutations per nucleotide [94]. Y-family polymerases are involved in translesion synthesis in man and do not possess any proof-reading activity [95].

Comparison of the mutagenesis frequency obtained by the human DNA polymerases β and η correlate well with their informed fidelity as 2×10^{-3} and 1×10^{-2} mutations per nucleotide were obtained by a single pass of replication, respectively [93]. Following this line of reasoning, the most bizarre human DNA polymerase candidate for mutagenesis is the DNA polymerase ι (iota), a Y-family member, which misincorporates guanosine and thymidine opposite a template thymidine with much higher efficiency than adenosine resulting in almost total

degradation of genetic information [96]. Human DNA polymerases produce a similar bias for transition mutations as observed with Taq DNA polymerase. In addition, DNA polymerase η generated deletions at a rate of 6.7×10^{-4} per nucleotide which, on average, corresponds to more than one frameshift per 2 kb gene [93]. Furthermore, an *E. coli* strain with a decreased fidelity of replication was developed, containing targeted mutations in DNA polymerase I, to be used as a mutator strain [97].

The naturally high error rate of the RNA replicase of Qβ bacteriophage has also been exploited in gene diversity generation [98]. Mutation rates of 7×10^{-2} per nucleotide were recorded in exploring the capability of Qβ RNA replicase [98], which were close to the error rates of the most low fidelity human DNA polymerases. The spectrum of substitutions was also more even with Qβ RNA replicase than with conventional epPCR, with an equal presentation of all transversions contrasted by epPCR, which favoured changing A and T over G and C. However, also Qβ RNA replicase showed a strong transition bias for A-T to G-C [98]. Another group of polymerases with high error rate are the reverse transcriptases that have also been exploited to incorporate base analogs [99].

Error-prone RCA is an isothermal DNA amplification method to generate random mutations in circular DNA by lowering the fidelity of phi29 DNA polymerase [100]. When RCA was carried out in the presence of 1.5 mM $MnCl_2$ 0.0035 mut/ntR were obtained [100], which is a lower frequency than reported for epPCR. The benefit of epRCA is the simplicity as a plasmid of interest can be directly amplified in error-prone conditions and transformed to a new host as a DNA concatemer [100]. The efficiency of direct concatemer transformation is low, but it can be easily increased 10- to 50-fold by cutting the concatemer into plasmid-sized units and re-circularizing the product with a ligase, or by implementing Cre/loxP recombination [article II]. The drawbacks of epRCA are that, by default, the whole plasmid is mutated in addition to the gene of interest. $MnCl_2$ decreases the product yields, and there is a strong bias for C-G to T-A mutations covering 66% of all mutations with a Ti/Tv-ratio of 2.7 similar to Taq DNA polymerase [100].

2.1.1.6 Transposons

Some transposons, e.g., mini-Mu, are randomly integrated into a target DNA [101]. A method called trinucleotide exchange mutagenesis (TriNEx) is based on a modified mini-Mu transposon carrying type IIS restriction enzyme sites, which are used to cleave off the integrated transposon from the template sequence with three extra nucleotides leaving blunt ends [102]. The gap is repaired with a DNA cassette carrying a kanamycin resistance gene, the new randomized codon to be inserted and a pair of type IIS restriction enzyme sites. After antibiotic selection and DNA purification, the kanamycin resistance gene is cleaved off and the backbone, with the new extra codon, re-circularized. Other very similar mutagenesis methods based on the random insertion of a transposon carrying

type IIS restriction enzyme sites for subsequent manipulation to delete or insert DNA have also been reported [103; 104]. The main benefit of random mutagenesis by transposons is the ability to change nucleotides in consecutive positions, which rarely occurs with other random mutagenesis methods. Trimer mutagenesis by TriNex results in an exchange of a single codon and double codon in 1/3 and 2/3 insertions, respectively, which may be even more than desired in some applications.

2.1.1.7 Choosing the method

There is plenty of information available on the characteristics and potency of especially *in vivo* mutagens [105], and the full picture of the effects is becoming more and more exact with whole-genome sequence analysis [106; 107]. However, there are only few comparative studies on the efficiency of different random mutagenesis methods in producing mutations for directed evolution experiments. According to one report comparing XL-1 Red strain, hydroxylamine and epPCR methods in *lacZα* mutagenesis, the sequence of increasing mutation frequency was hydroxylamine > XL-1 Red > error prone PCR [108]. In particular, epPCR with nucleotide analogs 8-oxo-dGTP and dPTP provided the highest frequency of mutations along with the best control over the mutation load by altering analog concentrations. Like already discussed above, most mutagenesis methods have a bias for transitions over transversions, which is especially valid with dPTP, 8-oxo-dGTP and hydroxylamine [108]. For a more balanced mutagenesis, a combination of methods is advisable. There are also computer programmes available to model the impact of the chosen method(s) on the anticipated diversity at the protein level [109; 110]. The main issues restricting the method selection are whether *in vivo* or *in vitro* mutagenesis is preferred and whether the target of mutagenesis is known. Other aspects are the available time frame for mutagenesis and the aspired complexity of mutations, e.g. changes at neighbouring codons or consecutive bases. Some characteristics of the random mutagenesis methods are summarized in **Figure 2**.

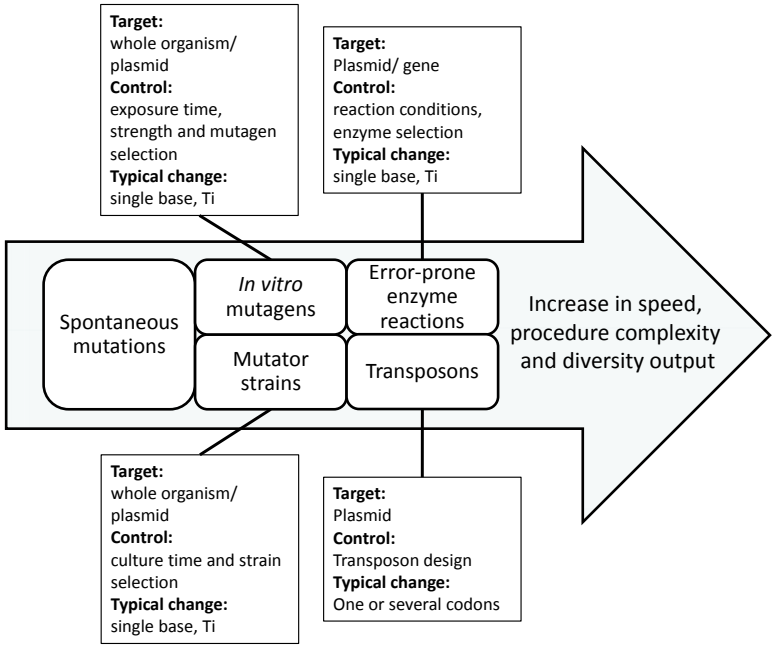

Figure 2. Characteristics of random mutagenesis methods. The frequency and the complexity of mutations introduced into DNA molecules can be increased with various methods. Error-prone enzyme reactions and transposons result in the highest clonal diversities in a short time, whereas the advantage of *in vitro* mutagens and mutator strains is the fairly simple implementation. Ti: transition.

2.1.2 Oligonucleotide-directed mutagenesis

2.1.2.1 Focused random mutagenesis and doped oligonucleotides

Above, epPCR was discussed for random mutagenesis of full-length genes, but naturally, epPCR can be directed to a specified shorter region by using primers as delimiters. A practical application is to randomly mutate a single domain of interest instead of the whole gene. The challenge in this approach is how to incorporate the randomized domain segment to be part of the gene replacing the invariant sequence. This is most often accomplished by inserting complementary sequence parts or restriction enzyme sites at the ends of the segment to be inserted [article I; 111]. There are numerous techniques available to incorporate such DNA elements, and these are discussed in more detail in the chapters

"Construction of highly diverse libraries" and "Construction of small diversity libraries".

The first site-directed mutation was documented in 1978, when Dr. Smith et al. altered the sequence of a DNA oligomer at the synthesis phase and incorporated the oligomer into a gene by primer extension mutagenesis [112]. The obvious consequence from this invention was that by providing mixtures of nucleotides to the growing polymer at the solid-phase synthesis, several variants could be created simultaneously. Nowadays, sequence-specific randomization of one or several positions in a DNA oligomer is a routine operation. If the nucleotides are provided at unequal proportions to favour the wild-type sequence, the diversified oligonucleotides are termed "doped", whereas equal mixture of all four bases would be referred to simply as a random oligonucleotide.

The first doped oligonucleotide dates from the year 1983 and it contained nine consecutive positions that were synthesized in the presence of 75% wild-type nucleotide and 8% of each of the other three nucleotides [113]. After the synthesis of another complementary doped oligonucleotide, the two oligonucleotides were hybridized with each other and ligated to a restriction enzyme site. In comparison to a totally random strategy, a limited number of variants are created by the doped oligonucleotide strategy with most of the members still retaining some of the parental character.

The major benefit of using doped oligonucleotides is the ability to introduce transitions and transversions in a site-specific manner even at sequential positions. Doped oligonucleotides are an effective alternative, when a short stretch of DNA is to be extensively mutated at various positions [114] and in discovery research related to noncoding sequences [115]. Oligonucleotides with equal stoichiometry of mixed bases per site can be ordered from most suppliers with the same price as a single base, but extra costs are added when ordering oligonucleotides with unequal base mixtures. The incorporation of a diversified oligonucleotide pool into the gene of interest results in a library containing mutations only at the site of the oligonucleotide. Full-length gene variants with evenly distributed mutations can also be assembled *in vitro* with modern assembly techniques by designing synthetic doped oligonucleotides with overlapping sequences [116; 117].

2.1.2.2 Oligonucleotides with randomized codons (NNN/NNK/NDT/KMT)

When the aim is to study protein variants, it is far more useful to design the library randomization codon-by-codon than base-by-base. The site-directed mutagenesis methods differ in this respect (**Figure 3**), and there are several solutions for codon-wise mutagenesis, which will be discussed in the next section. The degenerate codon NNN covers all 20 amino acids, but also three stop codons. With a single NNN codon, truncated protein variants are created at 4.7% probability if no skewing occurs due to technical aspects. This can be

avoided by using degenerate codons NNS or NNK (S = C or G and K = G or T), which also code for all 20 amino acids, but only for one stop codon. The reduction of the all-in 64 codons to the chosen set of 32 codons has significant implications, when several residues are diversified in the same gene. The likelihood of obtaining a stop codon in a gene exceeds 50% after ten sequential NNN codons, but with NNS (or NNK) only after sixteen codons.

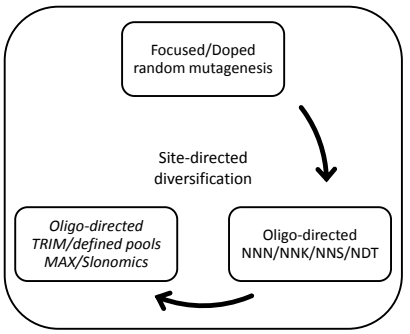

Figure 3. Site-directed diversification techniques. Arrows indicate a shift from base- to codon-wise randomization and, therefore, towards more controlled changes in the polypeptide sequence.

An even more central question for equal presentation of amino acid residues at a certain position in a library is how to cope with the codon redundancy. Tryptophan and methionine are coded by a single codon whereas arginine, leucine and serine are represented in the genetic code by six codons. The NNS/NNK-randomization is beneficial also in this respect as it narrows the gap between the most redundant and least redundant codons from 6:1 to 3:1 (**Figure 4**). For achieving a more balanced set of amino acid residues by incompletely defined bases, some amino acids must be excluded from the library design. A very attractive alternative to NNK is NDT, which is encoding twelve amino acids by twelve codons. This set includes a representative residue of each chemical subgroup characterized by a small (glycine, serine), hydrophobic (leucine, isoleucine, valine), aromatic (tyrosine, phenylalanine), acidic (aspartate), and basic (histidine, arginine) side chain including cysteine and asparagine having a thiol- and amide-group, respectively.

The advantage of NDT over NNK degeneracy is that it enables the exploration of a more diverse structural and functional sequence space with the same number of library variants [118]. For example, a 95% coverage of the structural diversity encompassed by eight consecutive NNK codons (theoretical size: 1.0×10^{12} variants) requires sampling of 3.3×10^{12} clones, but only 1.3×10^{9} clones need to be screened to explore the structural variants generated by eight consecutive NDT codons (theoretical size: 4.2×10^{8} variants) [118]. The practical difference between these two scenarios is that for display systems requiring bacterial transformation, a billion transformants is manageable [article I; 119], but a trillion is beyond the scope.

Smaller degenerate codon repertoires have also been exploited for library build-up especially in the field of synthetic antibodies. The diversification of the heavy chain CDR-loops at a total of 18 positions with a degenerate codon KMT coding for tyrosine, alanine, aspartate and serine, and subsequent selections with phage display resulted in antibodies with nanomolar binding affinities [120]. In a continuation study, the reduction of the tertiary code KMT to a binary code TMT pointed out that sufficient structural diversity is harnessed with a minimal choice between tyrosine and serine for selecting antigen-specific antibodies [121]. Especially tyrosine was pointed out to be a key residue for mediating molecular contacts [122].

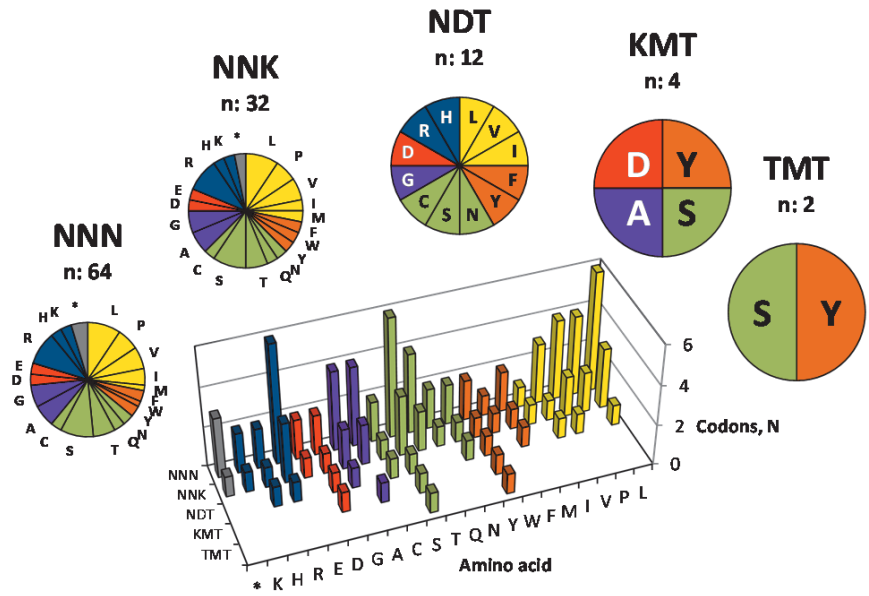

Figure 4. Reduction of genetic code with incompletely defined bases for generation of binding interfaces while maximizing the retention of structural and functional diversity. By a careful selection of the allowed codons in the target position, a more balanced presentation of functionally different amino acids is achieved.

2.1.2.3 Defined primer pools, TRIM and beyond

Often, it is not possible to cover the optimal set of amino acid diversity with a single degenerate codon consisting of incompletely defined bases. One solution is to synthesize each oligonucleotide separately with a defined sequence and pool them afterwards for implementation. Technical proceedings in massively parallel DNA synthesis are making this approach more and more attractive. For example, in one study exploring the limits of a programmable parallel oligo synthesizer, 53 478 unique sequences of ~ 100 nucleotides in length were successfully

synthesized on 23 glass slide arrays [123]. The synthesized oligonucleotides can be detached from the solid support and rendered for further manipulation *in vitro* or cloned forward to biological systems [123; 124]. Naturally, defined diversity pools still suffer from a low success rate of perfect oligonucleotides, which varies from 21 to 58% per slide [123]. The majority of the rest of the oligonucleotides contain one or more base substitutions [123], which is necessarily not a disadvantage from the directed evolution point of view.

Full control of the combinatorial amino acid diversity is obtained by performing oligonucleotide synthesis with trinucleotide phosphoramidites instead of the conventional nucleoside phoshoramidites [125; 126]. In trinucleotide mutagenesis (TRIM) a single codon is used to code for each amino acid and for each residue position, the preferred codons are mixed in desired proportions for integration into the growing oligonucleotide chain (**Figure 5**). This is the method of choice for very complex libraries, such as universal antibody libraries, in which diversity exceeds the sampling capacity emphasizing the importance of library quality [119; 127].

There are also other techniques for trimer mutagenesis taking advantage of enzymatic reactions. One mature technology is called slonomics, which is based on the use of IIS restriction enzymes and self-complementary hairpin-looped DNA oligonucleotides (**Figure 5**) [128]. For the synthesis of all possible sequences, an array of 4096 solid-phase-tethered anchor oligos and 64 dispensable splinker oligos are needed. The synthesis of a particular sequence starts with an anchor oligonucleotide containing the last six bases of the target sequence, of which the very last three nucleotides are present as a 5´-protruding trimer. A splinker oligonucleotide containing a compatible 5´-protruding trimer is annealed and ligated to the anchor, forming a double-stranded double-hairpin fragment, which is subsequently digested with an anchor-specific IIS type restriction enzyme releasing the splinker. At cleavage the splinker oligonucleotide obtains an extra trimer at both the 5´- and the 3´-end of the hairpin loop. The released elongated splinker is ready to be annealed to a second anchor containing the next six nucleotides to be incorporated corresponding to the ninth last to the fourth last nucleotide of the target sequence. Blocks of 18 bp are produced in a single synthesis, and these are then assembled into longer fragments. So far, slonomics has been applied, for example, to the construction of a gene library of G-protein-coupled receptors for finding variants with increased stability [129].

There is yet another older technique, termed MAX randomization or selective hybridization, for integration of desired trimer pools (**Figure 5**) [130]. A single-stranded template oligonucleotide is first synthesized with degenerate codons NNN at the sites of diversification. Then selection oligonucleotides containing six template-complementary bases at the 5´-end and three bases coding for the desired amino acid at the 3´-end are let to hybridize to the NNN-template. In a following ligation step, only the perfectly annealed selection oligonucleotides are ligated to each other. The formed ligation product is selectively amplified in a

PCR to copy the restricted diversity forward. For the total coverage of 20 amino acids, 20 separately synthesized selection oligonucleotides are needed for each NNN-position. Each NNN position also requires a separate set of homing oligonucleotides. Other technical limitations compared to the TRIM and slonomics are that several sequential diversified codon positions cannot be managed and that the reagents are not universally applicable to other targets.

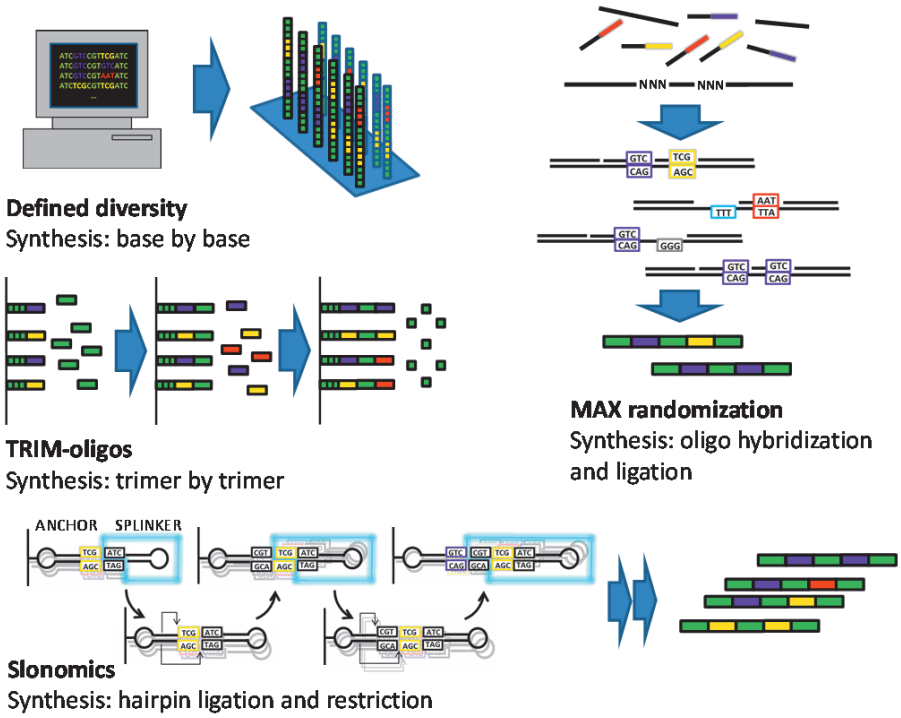

Figure 5. Techniques for codon-wise diversification. The most straight-forward approach is to synthesize each desired variant individually in parallel (defined diversity), whereas the most common approach for codon-wise diversification is to directly add trinucleotides to the growing chain in the synthesis phase (TRIM-oligos). One variation of the trimer approach is to use programmable robotic platform that operates on 4096 arrayed anchor oligos, each of which, passes three nucleotides both from its 5´- and the 3´-end on a growing splinker hairpin (slonomics). Slonomics takes advantage of the catalytical mechanism of a IIS restriction enzyme that cuts DNA outside of its recognition sequence. Upon digestion, the splinker is released and transferred to other wells in the array for the uptake of the next nucleotides. The least applied method for codon-wise diversification is to use a ssDNA molecule as template containing regions of fully randomized sequence (NNN) at the target positions (MAX randomization). Selection oligos with defined sequences (and desired trinucleotides at the 3´-end) are let to hybridize to the template. The correctly ligated dsDNA stretches contain combinations of only the desired codons.

2.1.2.4 General considerations of DNA incorporation

There is a wide range of methods for integrating diversified DNA elements into full-length genes for testing the library function at the protein level. Instead of listing all published methods, of which several are duds due to a lack of users, attention is given here to the methods with most impact. In addition, as the requirements for the size of a gene variant library vary according to the application, techniques for creating a large or a small library are discussed in separate sections. From a practical point of view, the proposed division into small- and large-scale methods is justified. The division is clear by a careful scrutiny of the reported size and quality of the established libraries, irrespective of the theoretical calculations.

From a theoretical point of view, the division of methods, as suitable for large- or small-scale library creation according to the extent of randomization and the number of obtainable library members *in vivo,* is subjective as several parameters affect the judgment. An exactly equivalent experimentation has rarely been applied for an objective comparison of libraries constructed with different methods. Another factor to be considered is DNA transformation, which is more efficient into *E. coli* than to *Bacillus subtilis* or yeast *Saccharomyces cerevisiae*. There are also major differences in the transformation efficiency of *E. coli* depending on the strain and method, e.g. heat shock vs. electroporation [131; 132]. Smaller circular DNA constructs transform with higher efficiency than larger ones [131]. Cell-dependent display systems require transformation of self-replicating or genome-integrating DNA, whereas cell-free display systems function directly with short linear DNA fragments [133]. For these reasons, only library construction methods for cell-dependent display systems with an emphasis on libraries hosted by *E. coli* are discussed in the following paragraphs.

The library construction methods can be evaluated according to several technical parameters. Wild-type gene copies are generally unwanted in the final library, because their presence reduces the diversity of the library. Wild-type copies may also cause a trouble in the selection and, therefore, template background is an important parameter to consider, when pondering between different construction methods. A second parameter is the frequency of unintended display-impairing mutations that inevitably lower the quality of the library. Typically, these are insertions or deletions leading to frameshifts or substitutions in the conserved regions causing misfolding and hence, the loss of function. The total number of obtained library members is a third useful parameter. Not only is the number of unique clones important, but also the relative frequencies of the members in the library. If the unique members are present at very disparate proportions, it may be due to the unequal incorporation of the provided building blocks at the DNA synthesis phase, the defects in library assembly, or the bias arising post transformation. Naturally, the total lack of certain variation of the designed diversity is a more severe defect than biases in the diversity presentation.

There are also other than technical issues to consider when choosing the appropriate library construction method. The commercial availability, the intellectual property landscape and the price may be important factors in decision making. Some methods require rare enzymes or custom synthesis services, which may not be easily accessible. A short hands-on time, simple protocols, reproducibility and scalability are also preferred characteristics of a library construction method. A very detailed analysis of these parameters in relation to each method is out of the scope of this review, but features of the utmost importance are discussed when relevant.

2.1.2.5 Construction of highly diverse libraries

Libraries exceeding one billion members are typically built to develop *de novo* binding functions that do not exist in the parent protein by randomizing surface-exposed residues. The development of the "one fold, many functions" -concept originates from studies in adaptive immunology. Antibodies recognize various kinds of antigens, although all of them are based on the same β-sandwich fold [134]. The majority of the molecular contacts of an antibody responsible for antigen binding are provided by the surface exposed residues in the loops of the β-sandwich fold of the variable domains termed complementary determining regions (CDR) [134]. Later, other analogous biological systems have been found [135; 136], but not surprisingly, the majority of the large library undertakings so far are based on the immunoglobulin or other related scaffolds.

It is logical to begin the scrutiny of the library build-up methods with examples from natural repertoires as similar methods have been applied at a later point in history to harness diversity from synthetic origin. The first universal, i.e. naïve, antibody library was constructed by amplifying rearranged variable light and heavy chain domain genes from B cells by PCR and joining them with a flexible linker for display [137]. Ever since, natural repertoires have been a popular source of antibody library diversity with a continuous flow of construction up-dates of larger and larger repertoires [138; 139; 140; 141; 142; 143]. For example, in the latest Pfizer initiative mRNA of peripheral blood leukocytes originating from 637 human donors and 17 human spleen samples were reverse-transcribed to cDNA and amplified with degenerate family primers to obtain 9 VH, 7 Vκ and 9 Vλ chains [143]. The VL and VH genes were joined with an intervening glycine-serine-linker by PCR assembly creating V_H-$(Gly_4Ser_1)_3$-V_L-constructs, i.e. single chain variable fragments (ScFvs). SfiI restriction endonuclease sites were added at the 5´- and 3´-ends of the amplified ScFvs for cloning the ScFv gene repertoire into a phagemid for display [143].

In the latest naïve library repertoire by Medimmune, cDNA from spleen lymphocytes and fetal liver was used as the source of diversity, and the amplified VH and VL repertoires were directly cloned into a display vector with unique restriction enzyme sites, instead of PCR assembly [144]. The achieved number of transformants of the Pfizer library was 3.1×10^{10} cfu and that of Medimmune 1.3

x 10^{11} cfu by 310 and 200 parallel electroporations, respectively. In both cases the ligation mixture was electroporated into electrocompetent *E. coli* TG1 cells. Other groups have also reported construction of libraries with a comparable size of over 10 billion members mainly with PCR assembly and bimolecular ligation [119; 138; 139]. PCR assembly, known also as splicing by overlap extension (SOE-PCR) [145], is based on identical sequences at one end of the PCR products to be joined. During the annealing phase of thermal cycling, some of the single-stranded PCR fragments will hybridize with a neighbouring fragment, priming each other. At the extension phase, the two DNA strands are extended by DNA polymerase welding the two fragments into one piece (**Figure 6**). The most notable information of the PCR-assembled Pfizer library was the diversity assessment. The size of the primary library was estimated to be 3.5×10^{10} unique members by deep sequencing a sample covering 96 303 heavy and 98 946 light chain reads, which is a perfect match with the total number of obtained colonies in transformation [143], indicating that each colony harbours a unique clone. This finding is more or less valid also for other libraries assembled with the same method.

PCR assembly has also been a popular method for construction of libraries with diversity integrated partly or fully from synthetic oligonucleotides [119; 146; 147; 148; 149]. Although most synthetic repertoires have captured framework regions from template DNA, the construction of VL and VH domain libraries with over a hundred million members by synthetic ssDNA oligonucleotide assembly have been reported [150]. The ssDNA oligonucleotide assembly was accomplished by T4 DNA polymerase without PCR, which is a noteworthy distinction between this method and the ones described above. The final step in most PCR assembly projects is the insertion of the assembled cassette into the final display vector with restriction enzymes and ligase for transformation. Different methods may be employed with other hosts. For example in yeast library construction, a PCR-assembled cassette was successfully integrated into a yeast display vector by recombination *in vivo* [146].

Another construction method with a well-established record of libraries with over a billion members is the Kunkel mutagenesis [122; 151; 152; 153; 154]. This is a primer extension mutagenesis technique in which mutations are incorporated into the final transformable circular DNA without bimolecular ligation [155], which is the most significant difference to the methods described earlier. In Kunkel mutagenesis the temple gene is propagated on a phagemid in an *E. coli* strain lacking the uracil-DNA-glycosylase (UDG) and dUTP pyrophosphatase activities. Due to the two deficiencies, there is a low frequency of uracil bases in place of thymine in the template DNA. The presence of uracil can be exploited as a selection feature as uridylated DNA is biologically inactivated in a wild-type ung^+ dut^+ host [67].

For Kunkel mutagenesis, the template is obtained in single-stranded form from filamentous phage particles that have been passed through the host cells. A mutagenic primer, containing a template complementary sequence on both sides

of the mutated region, is hybridized to the template and extended with a DNA polymerase around the phagemid. The remaining nick is sealed with a DNA ligase present in the enzyme blend. The end product is a covalently closed circular DNA, which is ready for transformation (**Figure 6**). As the product is a heteroduplex of a nascently synthesized mutated strand and the uridylated template strand, the nascently synthesized strand has a selective advantage over the template when the heteroduplex is propagated. Kunkel mutagenesis typically results in 50% mutagenesis efficiency [article I; 155]. In most library initiatives relying on Kunkel mutagenesis, the efficiency of the method is augmented by using templates containing stop-codons at the sites of diversification, which inhibits the display of the template protein [122; 152].

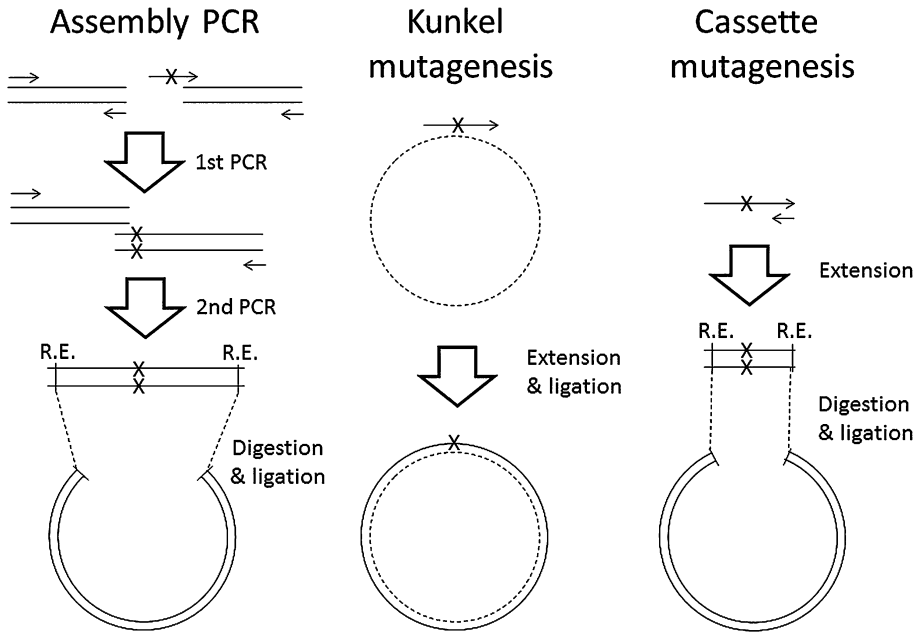

Figure 6. Methods for construction of highly diverse gene libraries.

A more efficient way to improve Kunkel mutagenesis is to decrease the template background prior to transformation. In one solution, the single-stranded template was modified further at the site of diversification to contain a DNA hairpin loop forming a functional restriction enzyme site [154]. The digestion of the parental strand eradicated the replicative potential resulting in 95% mutagenesis efficiencies upon transformation. Another approach, which does not require any template sequence manipulation for template background reduction is called selective RCA [article I]. In this method, the heteroduplex is treated with UDG followed by rolling circle amplification. The resulting concatemer is resolved back to single plasmid units for transformation. The rationale behind this technique is that the abasic sites created by UDG hinder extension by phi29 DNA

polymerase used in the subsequent RCA step. As a result, the nascently synthesized intact circular mutated DNA is selectively amplified. Another benefit of this technique is the ample supply of transformable DNA, which is a particularly lucrative feature for library build-up in yeast and mammalian cell display systems.

A fairly infrequently encountered construction method in the field of large primary libraries containing defined diversity at chosen positions is cassette mutagenesis. In this method, site-directed diversity is created by conventional solid-support synthesis of oligonucleotides into one strand, which is then primed and extended into dsDNA by DNA polymerase [111]. The cassette is digested with a pair of unique restriction enzymes and the diversity, e.g., covering a single CDR loop, is inserted into a template gene replacing a stuffer. This method was practiced by the company Morphosys in the early library versions called HuCAL [111], HuCAL Gold [156] and HuCAL Platinum [127]. However, the CDR diversity in the latest version of the library called Ylanthia, being the largest repertoire of synthetic human antibodies in Fab-format with 1.3×10^{11} claimed members, was constructed with slonomics technology [157].

As described above, slonomics is based on IIS type restriction and ligation for the sequential addition of nucleotide triplets to the growing dsDNA chain [128]. The synthesis of longer regions requires the assembly of short 18 bp blocks generated by the core technology. There is also another antibody library initiative by Pfizer containing slonomics-assembled CDR regions [158]. In both initiatives the final randomized CDR cassettes were inserted into the Fab framework with conventional restriction and ligation. After a full scale-up of the synthesis process, only 1 µg of heavy and light chain DNA was obtained by combination of slonomics and cassette assembly. For the final insertion of the Fab repertoire into the display vector, the heavy and light chains were joined by ligation and extensively amplified in 1000 parallel reactions by PCR to yield 100 µg of Fab cassette. As a result of 800 electroporations into TG1 cells, 4×10^{10} library members were obtained [158]. Diversity was well retained, despite the amplification, as only 1.5% of the CDR-H3 members were observed more than once in the transformed library by deep sequencing 740 000 domains [158].

Cassette mutagenesis requires extensive and careful design of template sequences to obtain a collection of unique restriction enzyme sites for entering diversity at several sites of a single gene. In addition, the efficiency of digestion and ligation of the formed cohesive ends varies from enzyme to enzyme hindering the workflow. In contrast, PCR assembly and Kunkel mutagenesis rely on the high and specific affinity of long stretches of complementary DNA sequences. The uracil excision-based ligation method, also known as USER cloning, shares common features with both the PCR assembly and Kunkel mutagenesis. In this method, PCR is carried out with a primer containing a single uracil at a defined distance from the end of the fragment [159]. The product is treated with a mix of UDG and endonuclease VIII creating a strand break at the site of uracil potentiating the dissociation of the 5´-oligonucleotide piece. The long single-

stranded 3′-extension can be subsequently used to assemble DNA fragments through complementary overlapping ends. Although linear DNA fragment populations comprising 10^{11} members have been created with USER technique for the needs of cell-free display systems [159], in cell-dependent systems the library sizes have so far been modest [160].

2.1.2.6 Construction of small diversity libraries

While large omnipotent primary libraries are utilized to generate binding affinity against any target, small diversity libraries are applied to further improve an existing function, i.e. binding affinity [161], specificity [162] or stability [163]. Another characteristic difference is that as large repertoires require display techniques for selection, small repertoires are most often screened without a physical geno-phenotype linkage. Naturally in screening efforts, the library diversity must be tightly controlled to enable oversampling of the designed diversity [161]. The library size in a small scale experiment is adjusted according to the screening capacity from a thousand enzyme variants manageable with labour-intensive screening methods like GC/MS [162], to over a hundred thousand soluble Fab variants requiring high-throughput ELISA screening [161].

Naturally, the methods for large library creation are applicable to small scale experiments as well, but commercially available methods with minimal preparative procedures are generally preferred for small libraries. PCR assembly is often encountered as the construction method in small diversity experiments [161; 164], but the most common method is probably circular site-directed mutagenesis, which was commercialized by Stratagene as Quikchange kits [165] [74]. In Quikchange, two complementary oligonucleotides are synthesized with mutations in the middle of the primer. The oligonucleotides are hybridized to the target gene on a circular dam-methylated DNA. The desired end product is circular dsDNA of nascently synthesized strands carrying mutations on both strands at the mutagenesis site after PCR cycling. The remaining template and template-mutant chimers are eliminated by DpnI digestion, which cuts both homo- and heteromethylated DNA leaving the mutant product intact [166]. Subsequently, the product is transformed into bacteria for nick sealing and mutant analysis. This is a genuine and fast technique, but only very modest libraries from a few hundred to a few thousand members can be constructed by one reaction [162; 167].

Quikchange has some less cited commercial competitors using different mutagenesis techniques. For example, the GeneEditor site-directed mutagenesis system by Promega is based on the simultaneous annealing of mutagenic oligonucleotide with an oligonucleotide changing the substrate selectivity of an antibiotic resistance gene [168; 169]. The mutants are directly screened from plates containing the selection antibiotic mix, because a plasmid in which the resistance phenotype has emerged also contains, with high probability, the aimed mutations originating from the mutagenesis oligonucleotide, which was provided

in molar excess to the selection oligonucleotide. There are several variants of the double oligonucleotide strategy based on reversion of a frameshift in a resistance gene [170], restoration of the origin of replication [171], and deletion of a restriction site [172].

Another Quikchange challenger is based on inverted PCR, such as the Phusion site-directed mutagenesis kit [173]. In this approach, a whole plasmid is copied by extending phosphorylated primers in opposite directions. One of the primers is carrying the mutations, and the mutagenesis is completed by blunt-end ligation and transformation [173; 174]. In our laboratory, a high efficiency version of the inverted PCR mutagenesis was applied by adding type IIS restriction sites in the primers. After PCR, the product is digested with the type IIS enzyme creating cohesive ends that are ligated to restore the template sequence with the mutations [article IV]. Inverted PCR coupled to type IIS digestion and ligation resulted in libraries of 10^5 to 10^7 transformants by electroporation into XL-1 cells. However, no reports exist of large billion member library creations with any of these methods.

Site saturation mutagenesis is a special application of small libraries constructed by site-directed mutagenesis. In this technique, individual residue positions are randomized to allow all amino acids in the chosen position, while keeping the sequence otherwise invariant [161]. One subtype of this kind of approach is the alanine scanning, in which each position is, one by one, changed to alanine to probe the structural and functional aspects of molecular recognition [175]. In affinity reagent and enzyme development, site saturation mutagenesis is a widely applied method in fine-tuning affinity or substrate selectivity when changes in a limited set of residues at well-known positions are required [161; 176]. Even some exhaustive experiments have been described by the one-by-one randomization of each codon in a gene [164; 177].

2.1.3 Recombination

2.1.3.1 Aims and applications

Recombination is used to shuffle existing gene sequences to obtain additive beneficial mutations into single genes or to remove deleterious mutations. Molecular breeding, i.e. DNA shuffling, is a well-recognized tool for trimming existing protein functions but, strictly speaking, it does not produce *de novo* diversity per se, if that is understood as novel changes at the single base level. Genes can be shuffled either by homologous recombination requiring sequence similarity or by non-homologous recombination potentiating fusions of totally unrelated sequences. Site-specific recombination has been used to augment combinatorial diversity in primary antibody libraries by establishing novel pairs of variable domain genes *in vivo*. In some instances site-specific recombination is

also a powerful technique for library cloning, an alternative to the conventional restriction digestion and ligation reactions.

2.1.3.2 In vivo

Some naïve antibody library builders have included additional genetic elements to increase the library size by *in vivo* recombination [140; 178]. Especially Cre/loxP-recombination has been explored to expand diversity by integrating loxP sites into DNA constructs followed by Cre recombinase induced recombination [179]. At least two large-scale studies exist of using Cre/loxP-recombination to expand combinatorial diversity, in the first of which, VH-CH1 heavy chain repertoire was transformed to *E. coli* and complemented via infection with a phage-packed VL-CL light chain repertoire accompanied with a dummy heavy chain [178]. Both the diversified heavy chains and the invariant dummy sequence were flanked with loxP sites enabling place switching by recombination. A bacteriophage P1 co-infection providing the Cre recombinase resulted in 6.5 x 10^{10} colonies resistant to all three markers originating from the plasmid carrying the heavy chain repertoire (ampicillin), phagemid carrying the light chain repertoire (tetracycline) and P1 bacteriophage (chloramphenicol). In theory, a perfect crossing of 10^8 unique heavy chains with 8 x10^5 unique light chains generates 8 x 10^{13} unique Fab fragments. In practice, the library size in this case was limited by the number of obtained colonies post P1 infection. The reversible nature of loxP recombination is a clear disadvantage of the method [180] as the phage progeny consists of both the perfectly recombined pairs of light and heavy chain and light chain members paired with the dummy companion.

In another initiative, the Cre/loxP-recombination was revisited by infecting cells expressing Cre recombinase constantly with phage carrying loxP-VH-loxP-VL-constructs [140]. A ratio of one cell per 200 phage resulted in the entry of several phagemids per cell, after which Cre recombinase catalysed the formation of all possible combinations of the present VH and VL domains. In this method, the primary phage progeny must be re-infected into fresh cells to restore geno-phenotype linkage prior to selections. Modest libraries consisting of 7 x 10^7 unique VL and VH domains were the foundation for the potential theoretical diversity of 5 x 10^{15} unique ScFvs. However, the authors assessed the final library size to be 3 x 10^{11}, which was mainly limited by the number of cells cultured for re-infection in order to restore the geno-phenotype linkage [140]. Despite the claimed five times larger library size in the latter study, higher affinity antibodies were obtained from the previously described Fab-library of 6.5 x 10^{10} members [178].

The mating of yeast *S. cerevisiae* can be utilized in an analogous manner to the Cre/loxP system in *E. coli*. In mating, two haploid populations with different mating types are let to form diploids, which are then selected for using compatible mating type-linked markers [181]. In this way, two immunoglobulin

sublibraries of light and heavy chains have been brought together for display [181]. Homologous recombination *in vivo* is another common tool for library construction in yeast cells [146]. In fact, compared with transformation of circular DNA, larger repertoires were obtained by co-transformation of linear vector DNA with a library of genes provided as PCR products [182]. It is even possible to introduce site-directed mutations by transforming mere single-stranded oligonucleotides [183]. This is, however, fairly inefficient compared to *E. coli* systems. Simultaneous mutagenesis of three regions of phosphoribosyl-anthranilate isomerase gene in yeast with DNA oligonucleotides succeeded with 0.01% efficiency, and for single sites, mutagenesis efficiency was 5% [184]. *In vivo* recombination and the recent proceedings in yeast transformation protocols have paved the way for yeast display libraries to become a vital competitor of phage display systems as the establishment of a yeast library comprising ten billion members has been reported [185].

2.1.3.3 In vitro

The generation of new variants via homologous DNA recombination *in vitro* most typically relies on PCR assembly. The oldest technology in this niche is sexual PCR dating back to the year 1994 [186]. In sexual PCR, parental genes are fragmented with DNase I digestion and reassembled in a PCR reaction allowing new chimeras to be established. As the name implies, the end result of DNA shuffling superficially resembles the outcome of a crossing over at the meiosis of sexually reproducing organisms. In another method, termed StEP for "staggered extension process", primers are extended for a very short period of time, denatured by heat and let to hybridize again, re-annealing to another template [187]. Typically, homologous genes shuffled by sexual PCR or StEP contained on average only 4 crossovers per gene per round of shuffling with a high share of intact parental sequences [188]. Therefore, another method termed RACHITT, based on more complete fragmentation of the parental sequences, was developed for higher efficiency shuffling [188].

RACHITT stands for "random chimeragenesis on transient templates" [188] and in this method, the source gene variants are first fragmented to single-stranded oligonucleotides. This is achieved by treating the source DNA with a lambda exonuclease destroying the anti-parallel DNA strand followed by a DNase I treatment cutting the remaining ssDNA into shorter pieces. Then, a transient template is prepared of the same source diversity as uridylated ssDNA, which corresponds to the destroyed anti-parallel strand, and the ssDNA fragments are let to hybridize. The non-hybrizing 5´-flaps are cleaved by the endogenous activity of Taq DNA polymerase, whereas the 3´-flaps are removed, and the final gap-filling performed, by Pfu DNA polymerase. The hybridized fragments are ligated by Taq DNA ligase, the template is inactivated by UDG and the chimeras are selectively amplified in a PCR with flanking primers. Although this method yielded on average 14 crossovers per gene [188], it is much more cumbersome to perform than the methods relying directly on PCR for assembly. A fully

fragmentation-free method for recombination, known as synthetic shuffling, has also been devised [189]. If the sequence of the parental genes is known and they have adequate sequence identity, defined and degenerate oligonucleotides can be designed *in silico* to cover the genes. After synthesis, the oligonucleotides are assembled by PCR for selections [189].

The first technique for non-homologous recombination of distantly related genes was published in 1999 and termed ITCHY for "incremental truncation for the creation of hybrid enzymes" [190]. In ITCHY, the two genes to be joined are truncated at opposite ends with exonuclease III. The ends are polished with Klenow fragment and the library of fragments truncated at the 5´-end is ligated to the library of fragments digested at the 3´-end. Unfortunately, this process mainly creates out-of-frame-chimeras. Another issue is the time-controlled exonuclease treatment, which is hardly reproducible. Therefore, a variant of ITCHY, termed thio-ITCHY, was developed. In this method, the genes to be fused were synthesized in the presence of dNTPs doped with a low frequency of α-phosphothioate analogs, which are resistant to exonuclease III digestion [191]. This innovation improves reproducibility as exonuclease digestion can be incubated into completion producing fragments with 3´-thiophosphate linkages. Surely, a combination of ITCHY and DNA shuffling is even more effective in creating chimeras and this combination is better known as SCRATCHY [192].

There are several strategies for recombination at defined sequence sites. Genes have been designed to contain IIS restriction sites that upon digestion and mixing induce directional crossovers at defined places [193]. In single framework libraries, the parental gene templates can be designed to contain unique restriction sites at appropriate crossover positions [article III], which can later be used to shuffle the retrieved specific clones and to segregate mutations. The same task could be performed without integrated designed features in the template sequence with primer sets containing type IIS restriction sites. Even nine fragments have been recombined by the latter method, termed "Golden gate shuffling", including fragment PCR, digestion, mixing and ligation steps [194]. A summary of the recombination methods is given in **Figure 7**.

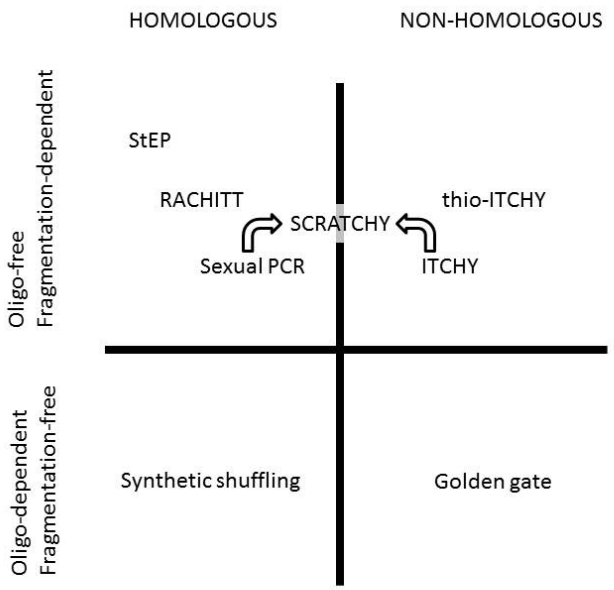

Figure 7. Recombination methods for creating new combinatorial diversity of existing sequences. In homologous recombination the recombinants are constructed by the assembly of similar sequences. Synthetic shuffling is a homologous recombination method that uses fully synthetic oligonucleotides as the building blocks, whereas the other methods are carried out by shuffling fragments of parental sequences. The non-homologous recombination does not require any sequence similarity. ITCHY is based on a blunt-end ligation of two truncated parental sequences and golden gate shuffling takes advantage of type IIS restriction enzyme cut sites introduced by PCR.

2.2 Methods for display and selection

2.2.1 Geno-phenotype linkage and selection

Genotype is the entire set of genes of an individual that together with the environment determines the observable properties of the individual, i.e. the phenotype. In directed evolution experiments a collection of genotypes are created, of which the individuals with the desired phenotype are selected for closer examination. Ranking the individuals according to the property of interest requires carefully planned experimentation. Typical qualities of proteins that are improved by directed evolution are stability [195], specificity [196], expression [129], and affinity [161].

For biotechnological exploitation, access to pure cultures of the desired variants is required and this is managed by spatial separation of the individuals. Most commonly this is achieved by picking colonies of bacteria or yeast cells from agar plates for culturing in separate vials, whereas sorting out a mammalian cell population to single monoclones requires dilutions to single-cell-per-well density. The hunt for the desired phenotype is commonly performed by screening small scale expression cultures of individual clones [161] or, in rare cases, by colony screening on plates, if the product of interest has a host protective phenotype or visually distinct appearance [article II; 197]. The capacity for direct screening of individual clones is limited and, therefore, the largest repertoires are sorted with display techniques. "Display" means a physical link between the expressed protein and the gene coding for it. The conceptual advantage the display techniques enable is the possibility to pool all member variants for selection. The majority of the diversity is lost in a single round of selection but still, the desired individuals qualifying for the selection are retained. After a display-dependent selection, the population contains a high share of individuals expressing the desired trait and, therefore, there is no need for massive screening campaigns in order to find the desired variant. The practical implications of display techniques are that selections are faster and larger repertoires are accessed than by direct screening efforts. A summary of the different display methods is given in **Figure 8**.

The selections of phage displayed libraries may be conducted on a collection of 10^{13} particles in a single 1.5 ml eppendorf tube [119], whereas in cell-free display systems libraries consisting of even 10^{15} members have been managed [198]. In yeast display, 10^7 cells may be routinely sampled with fluorescence-activated cell sorting (FACS) in one run and a pre-selection with magnet activated cell sorting prior to FACS facilitates the mining of even 100-fold larger primary repertoires [199]. The results obtained from a library initiative are dependent on the quality of the displayed library and the mastery of setting effective selection conditions for purposeful enrichment. Selections may, for example, rely on washing the unbound geno-phenotype units from a solid surface coated with the protein of interest or on monitoring the binding of labelled target proteins on the library members in a flow cell. The review of selection conditions is a subject of its own and exceeds the limits of this work. Selection techniques are, however, mentioned if they are an integral part of the display system.

2.2.2 High-throughput screening

Probing between a hundred and a thousand clones manually is easily managed as 96-well microtiter plate expressions, but sampling libraries of several thousand members, or more, requires high level of automation. There are some approaches to alleviate the labor intense nature of screening efforts, which will be briefly summarized here. A colony filter assay has been described in which a large antibody library was expressed directly from cells as a mat of colonies growing on a master filter [200]. The antibodies secreted by the colonies were in contact

with another antigen-coated filter. The antigen-bound antibodies on the second filter were detected and the hit regions at the corresponding site on the master filter collected for another round of colony separation and screening. Specific monoclonal antibodies were obtained after two rounds of filter screening [200]. Another group reported the simultaneous screening of 18 342 clones with antibody arrays [201]. In this technique, samples of expression cultures were spotted at high density on antigen-coated filters, washed and the bound antibodies revealed with a protein-L-horseradish peroxidase conjugate. Nonetheless, the latter technique first required the growth of the individual clones on microtiter well plates before gridding [201]. The antibody arrays have been miniaturized further from the membrane scale to microarrays spotted on microscopic slides [202].

If the screening capacity cannot be augmented to encompass all interesting variants at once, screening may be started with limited diversity libraries followed by the genetic recombination of the found beneficial mutations and sampling of their characteristics. An elegant and massive experiment of this type was carried out by the medical company Bayer in search of a higher affinity variant of adalimumab, the leading antibody drug on the market to treat rheumatoid arthritis [161]. The quest was started by screening site-saturation libraries with diversity set at chosen CDR-loop positions. The libraries consisted of single-site mutants originating from NNK-randomization and double-site mutants originating from NNK-NNK-randomization at consecutive positions. From this screen the positions that made major contributions to the affinity improvement were identified and taken as anchor points. First, all combinations of the mutations at the anchor points, 3840 genotypes in total, were generated and analysed, after which 12 new, less significant, mutations found in the primary screening were sampled using single anchor-point-optimized scaffold as the template. For this purpose, a library was constructed in which each of the new positions was allowed to variate between the affinity-improving and the wild-type residue. The best clone from this screen contained ten changes to the original adalimumab. On top of this, the affinity maturation of adalimumab was finalized by randomizing two further residues. In the three-step screening effort the original adalimumab affinity was first improved tenfold to 10 pM by the anchor point co-optimization, followed by a tenfold improvement to 1 pM by the additional optimization of the second sphere mutations. Eventually, from the final screen of clones originating from the library of two novel randomized positions, a clone with 400 – 500 fM binding affinity was found [161].

2.2.3 Phage display

Phage display is the most widely applied display technology due to its robustness, low price and well-established track record [203; 204]. Several coat proteins of filamentous phage can be used for the display of the protein of interest as direct genetic fusions. The most commonly employed coat proteins for display are p3, p9, and p8, of which p3 and p9 are able to display peptides and

proteins alike [137; 205], whereas p8 is mainly good for peptide display [206]. In addition, p7 and p6 display has been experimented [205; 207], but they remain without widespread applications. There are also numerous genetically engineered variations of display, especially with p3. The fusion of the protein of interest to a truncated p3 lacking the N-terminal infection domains leads to higher display efficiency compared to the full-length p3 display [208] and, in Cysdisplay, p3 is linked to the protein of interest via a disulfide bridge [156]. Selectively infective phage display systems have also been developed in which a non-infective phage displaying a protein of interest regains infectivity through interaction with a binding partner fused to the infectivity-restoring full-length p3 protein or merely to the infection domains [209; 210]. Work has been carried out to facilitate multi-copy display of proteins along the phage rod via p8 by genetically engineering p8 itself [211] or by optimizing the linker properties between p8 and the protein of interest [212].

The numerous display modes developed for the filamentous phage can be classified to display systems operating on phage or phagemid vectors. In phage vector display, the DNA sequence to be displayed is added to the wild-type phage genome. In 1985, in the first report on phage display of this type, peptide-coding sequences were fused to be part of the p3 gene [213]. In this mode, every p3 displays the peptide on the coat. As the fusion peptides may interfere with phage assembly, an extra copy of p3-peptide fusion gene was inserted into the phage genome in the next generation of display vectors allowing also wild-type p3 to be simultaneously expressed [214]. An alternative to the latter is the phagemid system in which the coat fusion is provided from an independently replicating plasmid containing also genetic elements for phage packing [215]. Currently, most large phage libraries are built on phagemids due to the ease of manipulation, higher transformation efficiencies and improved genetic stability. Phagemid systems require an additional infection with a wild-type phage to provide all the other proteins needed to make single-stranded DNA and new virion particles [215]. These phages, termed helper phages, have a modified packing signal to favour the packing of the phagemid instead of the virus genome into the virions being assembled [216]. There are also a variety of helper phage modifications to allow, for example, multivalent display [217] and to inhibit wild-type phage infection with trypsin-cleavable p3 [218].

Several types of protein libraries have been displayed on filamentous phage, of which, antibodies have received most publicity as the pioneers in the field with multiple examples of developed valuable products [219]. The proteins displayed on the N-terminus of p3 must be secreted into the host periplasm for proper display, which is well compatible with antibody fragments that preferably fold in the oxidative environment of the periplasmic space [220; 221]. Libraries of other proteins have also been displayed on filamentous phage including lipocalins [222], fibronectin type III domains [223], Z-domains of protein A [224], and ubiquitins [225], to mention a few. Display may be dramatically improved by changing the signal sequence from post-translational, which is the most widely applied, to co-translational or twin-arginine translocation [221; 226].

Filamentous phage is not very well suited for the display of proteins folding in the cytoplasm and therefore, the successful display of a library may be achieved only by turning to the second popular phage display system, which is the T7 bacteriophage [227]. T7 bacteriophage is a lytic phage, and consequently, display does not require translocation [228; 229]. A more balanced peptide library presentation was also obtained by T7 phage than by filamentous phage indicating translocation-related problems [230]. In relation to the capsid anchor protein, T7 phage supports C-terminal display, whereas N-terminal display is far more used with filamentous phages [213; 231]. The drawback of T7 display is the large genome size of the virion, difficulties in packing the modified DNA for infection and the 10 - 100-fold lower titers compared to filamentous phage samples [231]. Also some other phages have been tested in display, including λ-phage [232], T4 phage [233], baculovirus [234], and adenovirus [235], but they have not yet found a wider audience or are still at a developmental stage. Especially gene therapy by viral delivery is under intensive research and many intriguing undertakings are on-going to improve viral targeting and to evade the immune system with the aid of displayed polypeptides [235; 236; 237].

2.2.4 Cell surface display

Only one year and four months after the emergence of the concept of phage display a foreign antigen was displayed on the outer membrane of *E. coli* as a fusion to the trimeric integral outer membrane protein LamB retaining its antigenic properties as confirmed with antigen-specific immunoprecipitation and immunoelectron microscopy [238]. Since then, peptides [239], enzymes [240], and antibody fragments [241] have been displayed on the surface of *E. coli* for directed evolution experiments. As *E. coli* is a well-studied organism, several pathways have been established for cell surface display.

In the first application using flow cytometry to sort cell-displayed libraries, a C-terminal antibody library was fused to the outer membrane protein OmpA of *E. coli* and sorted according to the ability to bind fluorescently labelled antigen [241]. FACS is still the most common method for cell display selections. In a variant of the OmpA-display, a library is inserted into a loop of OmpA for constrained peptide display [242]. Other proteins used for membrane anchoring are the inactivated form of esterase EstA from *Pseudomonas aeruginosa* and the ice nucleation protein from *Pseudomonas syringe*, both enabling N-terminal display [243; 244]. The outer membrane protein OmpX was engineered by circular permutation to enable simultaneous N- and C-terminal display [245], which was applied to peptide affinity maturation by displaying the library in the N-terminus and a tag in the C-terminus for expression normalization in FACS [246].

Anchoring points other than the outer membrane for *E. coli*-hosted surface display are the flagellum and inner membrane. In one endeavor, which is now commercially available [247], a peptide library was inserted into the active site

loop of the thioredoxin gene, which was then inserted into the flagellin gene enabling display on the flagellum [239]. It was shown by anchored periplasmic expression that at least antibody fragments could be tethered to the inner membrane either by a small N-terminal leader peptide of NlpA or as N-terminal fusions to the filamentous phage p3 [248]. The permeabilization of the outer membrane is required for the fluorescently labelled antigens to access the inner membrane for subsequent sorting. In addition to antibody fragments, inner membrane display has been exploited in engineering G-protein coupled receptors for increased expression [249].

Gram positive bacteria can also be used to present large combinatorial diversities as demonstrated with *Staphylococcus carnosus*, which was harnessed to display affibodies and nanobodies [250; 251]. Yet another platform is to display proteins on the spores of *Bacillus subtilis*, a well-established system for display of defined sequences, although library applications have only recently been reported [252].

Yeast display has been actively exploited ever since the first report dating back to 1997 [253]. Proteins of eukaryotic origin are displayed in a more natural context on yeast than on bacterial host cells. Both post-translational modifications and the ability to express complex proteins, such as full-length antibodies, are central advantages of yeast display [254]. In addition, bacterial surface display is speculated to be sterically hindered by the lipopolysaccharide layer [253]. The most frequently used display technique with *Saccharomyces cerevisiae* is the fusion of protein library to α-agglutinin [255]. The agglutinins are involved in the adhesion of yeast cells during mating [256] and, therefore, the protein-protein interaction surface, tested by nature, is an attractive target for library display. α-agglutinin consists of two subunits of which Aga1p is covalently tethered to the yeast cell wall, and Aga2p is linked to Aga1p via two disulfide bridges [256]. Proteins are displayed as C-terminal fusions to the Aga2p, which is also the natural site for adhesion interaction [253; 257]. Various protein libraries have been displayed via agglutinins on the yeast surface, including antibody fragments [258], single chain TCRs [259; 260], fibronectin III domains [146], transpeptidases [261], and fluorescent proteins [262]. In addition to directed evolution, the agglutinin-site has been widely utilized as a general attachment point in whole-cell biocatalysis for the linking of enzymes operating in the extracellular environment [255].

At least eight other yeast proteins with varying expression levels are suitable anchors for surface display [263]. There are also adapter-mediated surface display systems for yeast, including the protein A and Fc pair, the cohesin and dockerin pair from *Clostridium cellulovorans* [264] and the coiled-coil pair of GR1 and GR2 [265]. In another adapter display system surface-expressed avidins were used to capture library members fused to *in vivo*-biotinylation domains [254]. Avidin capture was reported to allow the display of even full-length antibodies [254]. Besides *S. cerevisiae*, cell display systems have been established for some other yeast species such as *Pichia pastoris* and *Yarrowia lipolytica* [266; 267], but without notable library studies so far. There are also

innovative examples of using yeast display for challenging selections, such as changing protease selectivity [268] or engineering bond-forming enzymes, both requiring custom made multicomponent selection platforms [261].

In addition to bacterial and yeast display, there are reports on mammalian surface display especially for antibody engineering [269; 270]. In one initiative, full-length IgG antibodies were displayed on HEK239 cells and in another, on CHO cells, as a fusion to a C-terminal transmembrane domain [269; 270; 271]. Preliminary studies on insect cell display have been conducted with Sf9 for the display of major histocompatibility complex class II proteins [272] but, to date, the largest libraries on insect cells have been in the range of 10^5 members [273].

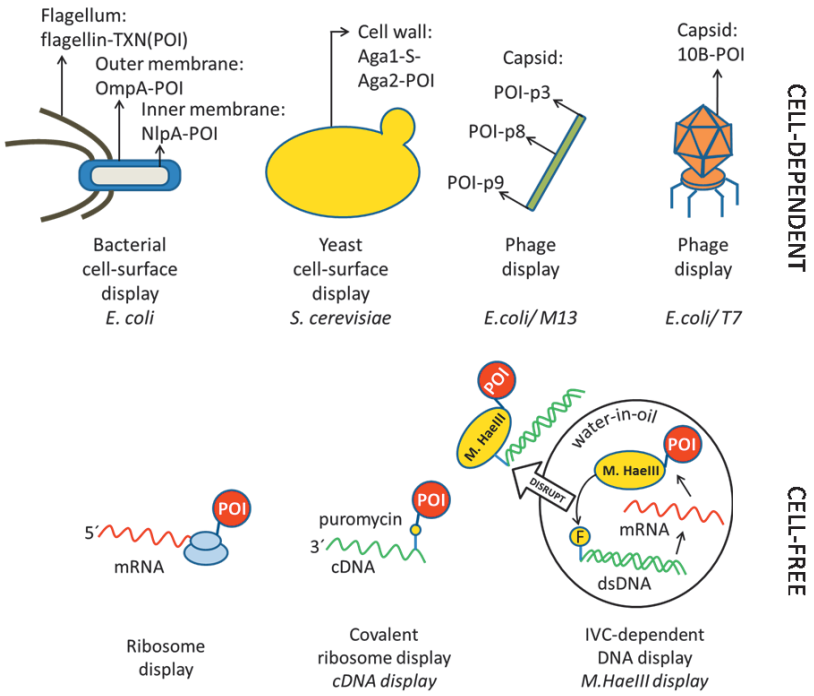

Figure 8. Overview of cell-dependent and cell-free display methods. POI: protein of interest. Arrows indicate different localities used for display.

2.2.5 Ribosome display and related instant links

The reconstitution of translation and transcription *in vitro* without living cells in 1961 [274] was a prerequisite for the development of cell-free display systems. The cell-free transcription and translation extract is obtained by gently lysing *E. coli* cells and separating the soluble fraction by centrifugation [275]. Components, such as amino acids, template DNA and energy source, are supplied in the obtained S30 extract to start translation. Cell-free translation technology

has been further refined by reconstituting translation from purified components minimizing uncontrolled side reactions [276]. In ribosome display, a ribosome remains non-covalently attached at the 3´-end of the translated mRNA while still linked to the nascently synthesized polypeptide due to the absence of stop codons in mRNA [133]. The rescue of the target mRNA is simply performed by reverse transcription and PCR post selection. In a proof-of-concept study in 1994, a target peptide was enriched from a library of 10^{12} members with a monoclonal antibody [133] and the display of full-length proteins was demonstrated three years later [277]. The stability of the system has been enhanced by using low reaction temperature and high concentration of magnesium ions [277]. Measures to silence native mRNA quality control system may also play a role in retaining ribosome-mRNA-complexes intact [277].

Despite the protective measures, ribosome display is fairly labile without covalent linking. Therefore, a technique called mRNA display was developed, in which the mRNA was covalently linked to the translated protein by adding a puromycin-DNA adapter to mRNA prior to translation [278]. The ribosome stalls at the peptidyl acceptor antibiotic puromycin located at the 3´-end, after which the mRNA is reverse transcribed to form a stable mRNA-cDNA hybrid and used for selections. Target genes are rescued via the cDNA by PCR [278]. A variant of this technique is cDNA display (**Figure 8**), which also relies on the puromycin linking and reverse transcription with the exception that the puromycin attachment site is in the centre of a DNA adapter allowing the use of the free 3´-end of the adapter hairpin to prime cDNA synthesis linking the polypeptide covalently to the synthesized puromycin-tagged cDNA, and not to mRNA as in the first version [279]. Instant linking of the formed protein to the coding gene has also been accomplished with a RepA fusion protein [280]. In this method, termed CIS display, a DNA sequence of the origin of replication is added to the gene library DNA and as the RepA-protein fusion is translated, RepA binds to the dsDNA ori-element. Only a single library-scale peptide experiment has been reported using CIS display [281].

One major advantage of the cell-free systems is the possibility to reprogram the genetic code to insert unnatural amino acids into a polypeptide via modified tRNAs. Recent developments in charging tRNAs with ribozymes has widened the spectrum of incorporable unnatural amino acids [282]. For example, large libraries of *in vitro*-translated peptides have been recently connected to form constrained circles with the help of incorporated unnatural amino acid residues followed by selection with mRNA display [283]. The open operating system also allows the addition of other heterologous components, such as chaperones, to aid in protein folding [284].

Another benefit of the open operating system is that the library size is not limited by transformation efficiency as genes are added directly as linear DNA to the transcription and translation mixture. Furthermore, in all selections requiring reverse transcription, the high error rate may bring beneficial mutations during library rescue [285]. In general, mRNA display is a well-established technique

that has been widely applied to engineer fibronectin domains and ankyrin repeat proteins, some of which have already entered clinical trials for medical use [286; 287; 288]. Aside from developing novel pharmaceuticals, ribosome display has also found use as a platform to improve the stability and expression of established protein drugs [289].

2.2.6 In vitro compartmentalization

Double-stranded DNA is more stable than single-stranded DNA and far more stable than mRNA. Consequently, dsDNA would be the preferred genotype partner for establishing geno-phenotype units. Because DNA is not physically in contact with the encoded protein, the problem is that how to cross-link the gene present as dsDNA with the corresponding nascently synthesized polypeptide. Several solutions have been tested to answer this question by isolating the gene and the encoded protein to the same compartment for linking, before the disruption of the isolation and selection with the freshly coupled geno-phenotypes.

One of the most common formulations for *in vitro* compartmentalization (IVC) is the use of water-in-oil emulsions, which typically yield droplets of 2 µm in diameter with 5 fl volumes [290]. A 50 µl reaction may consist of 10^{10} artificial cells. Optimally, each contains a single DNA template that can be amplified, transcribed and translated for geno-phenotype coupling [291]. DNA may be linked to the encoded protein non-covalently as in STABLE-display in which biotinylated DNA binds to the in-droplet expressed streptavidin-protein fusion [292], or via a zinc finger domain fusion binding the recognition element attached to the template DNA [293]. The covalent linking of the translated protein to the corresponding template DNA has been achieved, for example, by fusing the protein of interest with DNA methyltransferase M. *Hae*III (**Figure 8**). M.HaeIII covalently binds to a 5-fluorodeoxycytidine (F) in a methylation target sequence 5´-GGFC-3´ placed at the end of the encoding DNA molecule, establishing a stable link [294].

Geno-phenotype linking may also be arranged via microbeads. In one example a monoclonal antibody attached to streptavidin coated microbeads was used to capture an in-droplet synthesized enzyme [295]. As the template DNA and enzyme substrate were biotinylated, the beads displaying the active enzyme and desired product could be sorted with FACS using labelled anti-product and anti-enzyme antibodies. A streptag variant of the microbead display has also been described [296].

Thermophilic DNA polymerase variants have been successfully evolved by self-replication in intact emulsion droplets [297], which is a fairly straight-forward application for IVC-techniques. However, the most interesting applications of intact droplets are in the field of enzyme engineering, for example, by following fluorescent product accumulation into droplets with FACS. To be amenable to

sorting, the encapsulation of the components in either water-in-oil-in-water double emulsions [298] or liposomes [299] is required. Liposomes have the extra benefit of allowing membrane protein display as they consist of a phospholipid bilayer [299]. Cell-free methods have also been applied to the selection of heterodimeric proteins by allowing gene linking in IVC-droplets followed by mRNA display [300].

2.3 Aspects of functional display

2.3.1 Evolution

During the course of a directed evolution experiment the frequency of alleles in a population are changed. Firstly, the diversity of the allele pool is reduced typically to a few hundred variants and, secondly, the frequency of alleles with experimentally favoured traits is tremendously increased. However, one of the lessons to be learnt from the modern synthesis of evolution theory, emphasizing the reproductive success instead of mere survival [301], is that selection experiments may be substantially skewed by secondary parameters that contribute to the overall fitness of the clones.

Especially in situations in which the fitness of an allele changes over time, the dominating allele at the end is the one with the highest geometric mean fitness [302]. This finding is common in phage display as deleted forms of library members are often encountered after selection on a target antigen [303; 304]. Although the clones that bind to the provided antigen with the required affinity or avidity are enriched in the selection phase, the clones that propagate the best take over the population in the following infection and growth phase. The displayed heterologous proteins often have toxic effects on the *E. coli* host promoting the expansion of the aberrant clones [article III].

Particularly, if the selection pressure is not persuasive enough, or when it is not yet in operation, the alleles are enriched according to the secondary traits. This phenomenon has been strikingly shown with landscape phage displaying peptides as p8 fusions on the filamentous phage capsid [305]. Extensive biases were observed already in the primary phage population prior to selection [305]. Ultimately, peptides that interfere with the phage assembly are entirely excluded from selections, although they would otherwise have excellent binding properties.

2.3.2 Fitness landscape

Proteins do not possess equal evolvability. The natural history of proteins has shown that diverse functions have arisen from some folds, but not from others

[306]. For example antibodies are tunable to recognize a wide variety of structures via mutations and genetic recombination [307] and, therefore, the immunoglobulin scaffold is considered highly evolvable. However, the acquired surface architecture of the protein sets limits to the obtainable fitness increments that additional genetic fine-tuning can offer. Although a local optimum for maximum affinity would be reached by single substitutions, the absolute optimum may remain unachievable, as it would require a fundamentally different surface architecture. This idea is coined in the concepts of rugged and smooth fitness landscape [308; 309]. Directed evolution in a rugged landscape leads to several local adaptive peaks, whereas in a smooth landscape the protein is evolvable to the global absolute maximum fitness via several routes [309].

Protein stability has dual effect on the fitness landscape. On one hand, stability promotes protein evolvability as a stable scaffold tolerates more mutations than an unstable one, which has been demonstrated with cytochrome P450 peroxygenase variants [310]. A library built on a more stable enzyme produced three times as many different solutions for hydrolyzing novel substrates than a library based on an unstable enzyme [310]. On the other hand, the acquisition of novel functions requires structural plasticity, which may be hampered by excessive stability [311; 312]. A recently proposed hypothesis for predicting broad evolvability joins these two ideas together as it states that folds with a physically separated stable core and active site provide high conformational plasticity and stability in the same package implying a smooth fitness landscape [313]. As only a tiny portion of the total available sequence space can be scanned with any combinatorial library initiative, the scan inevitably resembles a sparse shotgun firing pattern. Keeping this in mind, the prediction is that in a rugged fitness landscape of rigid proteins, further variant development would lead to a local, compromised fitness trait, whereas in a smooth landscape of polar (separate core and active site) structures, further development would gradually lead to the ultimate fitness peak.

2.3.3 Library size

The high transformation efficiency of *E. coli* enables the construction of libraries of 10^{11} members, which are displayed on phage or on bacterial surface [314]. Starting from ready-to-transform library DNA, 10^9 - 10^{10} transformants are routinely obtained in our laboratory in one working day [article I; 119]. Yeast transformation is less efficient than bacterial transformation and, therefore, typical yeast display libraries consist of 10^6 -10^7 members although, in rare cases, 10^9- to 10^{10}-membered libraries have been reported [185; 315]. Libraries accessible by *in vitro* compartmentalization are larger or close to the *E. coli* library sizes as the physical dimensions of the compartments resemble the size of bacterial cells [290]. Template dilution to a concentration at which only single template exists in the artificial cell is a prerequisite for IVC-based display techniques setting also limits for the maximum number of members. The instant linking of geno-phenotypes by ribosome display and related techniques provide

significantly larger library sizes up to 10^{13} - 10^{14} members and are ultimately limited only by the number of ribosomes [316]. There is also a difference in the way library DNA is brought to life, as already discussed above. In its most simple form, mere linear DNA with a promoter and the gene to be translated is needed for ribosome display, whereas the DNA to be transformed into *E. coli* must be provided as circular DNA with required replication and selection features.

2.3.4 Randomizing for maximum functionality

Even the highest achievable library size of 10^{14} members is a tiny sample of the full sequence space of a small polypeptide of 100 residues, i.e. 20^{100} variants. From another point of view, screening capacity is most often the main limiting factor especially in enzyme development projects, for which display methods cannot be established. When the physical dimensions of the screening capacity have been fully explored, the only way forward is to improve upon quality, which in directed evolution experiments translates into maximizing the number of functional clones in the population.

Firstly, to be functional the clone must be in an open reading frame. A generally known source of base deletions and insertions are the commercial synthetic oligonucleotides, which may have error rates of 1% per base in microscale synthesis [317]. Libraries requiring the incorporation of multiple oligonucleotides at various positions for diversity generation are especially vulnerable to oligo-derived mutations as a frameshift at any of the randomized sites will destroy the display. There are a few ways to tackle this problem. The mutated genes may be preselected for functionality, for example, as β-lactamase fusions before the final randomization [156]. The quality of the oligonucleotides may also be augmented by choosing the oligonucleotide provider with care or by consensus-based error correction, if oligonucleotides with defined sequences have been used for diversification [318].

Secondly, the integrity of the secondary sequence elements plays a major role in conserving functionality [319; 320]. A β-strand has been claimed to tolerate less mutations than an α-helix [320], and the surface exposed loops allow more diversity than the hydrophobic core [321]. Another general rule states that hydrophobic residues may be exchanged for other hydrophobic residues, as well as hydrophilic residues for other hydrophilic residues, but not for those of the opposite chemical character [321]. The rational design of mutational diversity by the above-mentioned principles aims to retain as many variants as possible in a correctly folded state increasing the overall quality of the library. However, if the aim is to change specificity or evolve affinity, the sites of mutagenesis providing the desired diversity with the highest probability are the residues or regions involved directly in catalysis or in the binding interaction. The implementation of the latter strategy requires more elaborate structural information, whereas a less accurate structural prediction is adequate for the general restriction of

mutagenesis, which is achievable directly from primary sequence with homology modeling [322].

There are several conceptual strategies and computer programmes for guiding the design of focused diversity libraries [323], of which the combinatorial active-site saturation test, i.e. CASTing, is one of the most frequently applied [324]. The idea of CASTing is to simultaneously create site saturation mutagenesis libraries of two residues at the active site of an enzyme. All the combinations of the two varying residues still form a repertoire screenable by any method and are the population with the maximum likelihood of finding altered variants. The library quality is further increased by focusing on the mutational content in the selected positions. Which substitutions are allowed in each position to serve the purpose, and how to capture the diversity in synthetic DNA oligonucleotides in the most non-redundant way? A software programme has also been developed to assist in this dilemma, and to assess the number of clones to be screened for the full coverage of the designed diversity [325]. Especially in antibody library designs, diversity focusing for maximum functionality is fairly straightforward due to the available databases of known antibody sequences [134; 143]. The knowledge of the structure-function relationships of antibody-antigen interaction has inspired library initiatives that, instead of mirroring the full native repertoire, try to evolve affinity reagents for specific subgroups, such as haptens or peptides, by limiting the diversification to carefully selected positions and content to tune the binding profile of the library [326; 327].

2.3.5 Folding, translocation and leader peptides

All proteins are not amenable to filamentous phage display [221; 227]. The history of phage display is linked to peptide and antibody display, which was a fortuitous coincidence leading to successful development projects. Peptides are fairly unstructured if not constrained and, therefore, suitable for display with any platform. Proteins, however, require specific conditions for folding [328] and, consequently, the display of a library of proteins in a new host, disconnected from its natural environment, may not work. To be displayed on the p3 or p8 of a filamentous phage particle, the protein must be secreted into the periplasmic space [221]. Traditionally in filamentous phage display, this is achieved by directing the fusion protein to a post-translational secretion (SEC) pathway with a signal peptide [208]. Antibody fragments are guided to the inner membrane and translocated in an unfolded state by the post-translational signal sequence to the periplasmic space, where folding subsequently takes place as it requires disulfide bond formation [329]. Although the display of antibody fragments is well compatible with the SEC-pathway, the display of cytoplasmic proteins may be severely impaired through SEC, which has been attributed to rapid folding kinetics [330].

Fortunately, there are two alternative pathways for protein secretion in *E. coli*. Co-translational translocation allows the display of more stable scaffolds by

synthesizing the polypeptides of interest directly through the membrane to the periplasmic space [221]. This was dramatically shown with a fibronectin type III domain, which is based on a stable β-sandwich fold without a disulfide bond, by a 1000-fold increase in display when changing from post- to co-translational pathway [331]. The third translocation pathway called twin-arginine translocation (TAT) secretes folded proteins across the plasma membrane enabling the display of cytoplasmic proteins on the filamentous phage [332]. TAT-pathway has been shown to be compatible with truncated p3 and p9 display, whereas display with full-length p3 was impaired [226; 330]. Utmost caution should be taken in generalizing these findings as they most likely are fold-dependent, and the experimental verification of proper display is of central importance prior to any library work.

Naturally, cytoplasmic proteins can be directly displayed on the lytic T7 phage as display does not require export to the periplasm [229]. The cytoplasmic proteins are also compatible with ribosome display techniques, which is the standard platform for working with, e.g., stable ankyrin repeat proteins [288]. Small proteins, such as ScFvs, are amenable to several platforms, whereas the display of large proteins, such as full-length antibodies, is still limited to yeast and mammalian hosts [254; 269]. All in all, when committing to novel library projects, it is clear that understanding the chemical, biological and theoretical limitations of the particular system are necessary for successful implementation. A checklist for new library initiatives is provided in **Figure 9**.

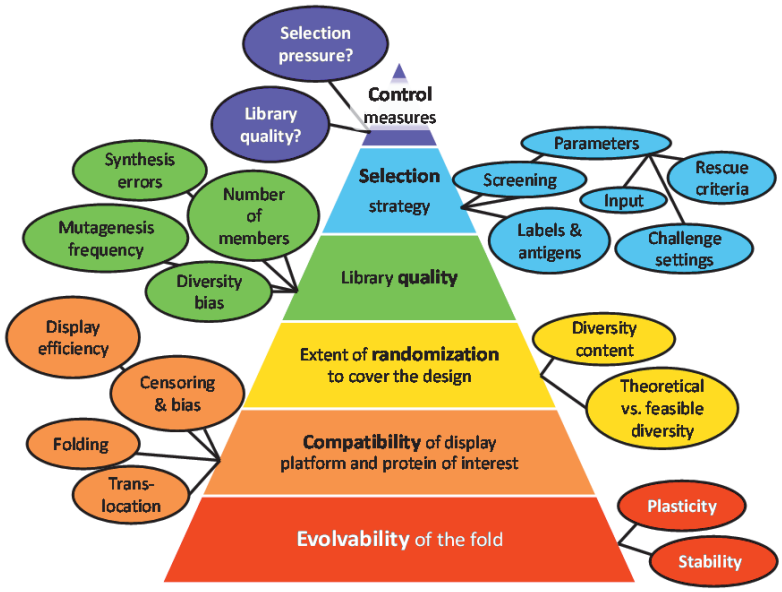

Figure 9. Checklist for new library initiatives.

2.3.6 Evolving enzymes

The establishment of geno-phenotype linkage is the major challenge in engineering proteins capable of catalysis by directed evolution as the substrate and product are dissolved in the surrounding solution. The traditional enzyme engineering relied on a high-throughput screening in well format [333] or on colony screening approaches, if the emerging trait was selectable, e.g., an enzyme conferring a novel antibiotic resistance [article II]. Since then, fluorescence-activated flow cytometry has become a major tool in enzyme engineering as it potentiates the screening of libraries that are several orders of magnitude larger than those manageable with single well assays. However, FACS-selections require individual study designs as there is no generic format suitable for all enzyme selections, in contrast with affinity generation, in which solid phase separation is always an option.

For some FACS-selections, the substrate may be directly turned visible by fluorescence tagging. For example, in selecting sialyltransferases for better catalytic efficiency, the successful transfer of a sialic acid to a fluorescently tagged lactose derivative inhibited the transport of the fluorescent product out of the cells and, as a result of this, the improved catalysis was eminent as a higher cell fluorescence [334]. The activity of generated Cre recombinase variants for recombining at a novel sequence site has also been monitored by FACS [335]. In this study, the exchange of a reporter gene *GFPuv* for *EYFP* by recombination resulted in altered fluorescence, which was discernible in FACS.

There are several novel *in vivo* screening platforms linking an enzymatic reaction with reporter enzyme expression. In one study, cephalosporinases were developed with a yeast-three-hybrid assay, in which the basal transcription of *LacZ* gene was disrupted by the cleavage of a β-lactam substrate linking the DNA-binding protein LexA and transcription activation domain B42 [336]. In another approach, the fusion molecule of a ligand-responsive RNA aptamer and a poly-A tail-cleaving ribozyme was used to modulate the expression of GFP [337]. In the ribozyme-aptamer fusion, either the ligand binding or the cleaving domain was properly folded. In the native state, the ribozyme was actively cleaving the poly-A tails of the GFP transcripts sustaining the low expression level of the reporter, whereas upon ligand binding the ribozyme was misfolded leading to more stable transcripts and higher GFP expression [337].

There are also variations of the cell surface display enabling the selection for catalysis. A general approach for protease trimming was claimed by a system in which the cleavage of an intracellular peptide substrate leads to the display of the cleaved part on the cell surface, which can subsequently be detected with a specific labelled affinity reagent in FACS [268]. Serine proteases have been engineered with a different strategy, in which a substrate was attached to the surface via electrostatic interactions. A functional surface-displayed protease cleaved the substrate, releasing a quencher arm, after which, the fluorescence of the surface-tethered part was restored [240].

Particularly in the early days, phage display was used for enzyme development by using stabilized transition analogs to capture the enzyme-displaying phage for washing [338], or by trapping the products catalyzed by the displayed enzyme to the phage coat, for example, with maleimide activated linkers, followed by a product-specific affinity selection [339]. However, the recent onrush of IVC-methods has dominated the enzyme evolution landscape and will continue to do so in the future. The main benefits of IVC are that the substrate and product are linked to the gene product as all are confined to the same droplet. Especially the FACS-compatible double emulsions have already been applied to the engineering of a β-galactosidase [340] and a paraoxonase [341]. There are also novel ways of droplet manufacture with microfluidics further miniaturizing the screening effort to picoliter scale [342]. With the newest droplet-FACS platforms, 10^7 droplets may be screened in 3 hours, which is a major improvement to the classical robotic screening in microtiter plates, which can reach 10^5 assays per day at best [342].

2.4 The future of experimental directed evolution

The proceedings in *de novo* gene synthesis have made genes available for everybody at diminishing prices. It is trivial to order genes directly as codon-optimized for expression in the final host, although the guidelines regarding the maximum expression of heterologous proteins are still under debate [343]. The progress in massive parallel oligonucleotide synthesis creates new opportunities for the construction of libraries with well-defined diversity and sets new standards for error correction and quality control [344]. Furthermore, the TRIM technology is becoming an ordinary method for library construction as it has recently become commercially available leading to the better control of randomized diversities. In the future, the synthetic genes may even be based on *de novo* protein designs that to date do not exist. One triumph in this field was the computational design of an enzyme that catalyzes the Kemp elimination reaction, for which no natural enzyme has been found [345]. However, in this work extensive directed evolution was also required to further increase the catalytic activity. *De novo* protein design is at an early stage, but major breakthroughs will be seen with the improving computational design algorithms. The results obtained from directed evolution experiments also provide vital information to support the software development.

Second generation sequencing (SGS) techniques have been eagerly implemented in directed evolution research from early on and their significance continues to increase. SGS has a dual role in supporting directed evolution studies. On one hand, massively parallel sequencing is a potentiating tool for analyzing the quality of created libraries in detail [143], for choosing hits from complex experimental set-ups [346] and for gaining knowledge of the properties of proteins via the inspection of mutagenesis profiles [143; 346; 347]. On the other hand, the metagenomics of uncultivable microbes, which is dependent on the massively parallel sequencing for species identification and genome mining, is a

source of thousands of novel enzymes [348], which can be further manipulated by directed evolution, both for the efficient expression in a heterologous host and for the catalysis of unforeseen reactions.

Synthetic biology has progressed from introducing single genes to introducing whole metabolic pathways into the desired host organisms for the production of natural and unnatural products. Metabolic engineering is especially inspired by the visions of biofuel production in micro-organisms and, in this context, directed evolution is a practical tool to adjust the multi-component systems for maximum production [349]. Heterologous pathways have been fine-tuned with promoter libraries [350], transcription factor libraries [351] and the libraries of tunable intergenic regions [352]. Pathway trimming has also been experimented with gene shuffling by integrons [353].

The arsenal of new methods for genome-wide manipulation potentiates new approaches for strain improvement by directed evolution on the organism level. These methods include global transcription machinery engineering (gTME), which is based on mutating transcription regulatory proteins and selecting for the desired phenotypes [354]. In another method, termed multiplex automated genome engineering (MAGE), cells are transformed with ssDNA oligonucleotides targeting specific sequences for replacement by recombination [355]. Using this method, the expression of 24 genes were simultaneously optimized by modulating the ribosome binding site sequences with degenerate oligonucleotides (and by knocking out alternate pathways), followed by a selection phase resulting in a fivefold increase in lycopene production in *E. coli* [355]. Moreover, there are aspirations to develop controllable regulatory elements in the host genome to modulate host metabolism on demand. For example, the specificity of estrogen receptor was changed to respond exclusively to two new synthetic ligands for the establishment of logical gates to shunt gene expression [356]. Another example of genetic circuits is the ligand-responsive riboswitches [337].

The rise of cell-free display platforms predicts accelerated selections on combinatorial repertoires in the future. The open operating system of the cell-free platforms, as well as the progress in engineering whole organisms to express orthogonal tRNA/aminoacyl-tRNA synthetase pairs, have enabled the incorporation of unnatural amino acids into proteins [357]. The unnatural amino acids widen the spectrum of physico-chemical properties of proteins beyond the natural counterparts establishing totally new application areas, e.g., in imaging and immunotherapy [358; 359]. Systems for the incorporation of unnatural amino acids have been further optimized by directed evolution [357]. Another dimension to be explored is the manipulation of binding polypeptides with chemical modifications to obtain novel functions. For example, the bicyclic peptides have been obtained by bridging three cysteines via a tris-(bromoethyl)benzene moiety [360].

If we consider that the minimum tools for gene variant generation are PCR and oligonucleotide synthesis, the directed evolution of single genes has been

practiced from the mid 1980s onward [113; 361]. Our understanding of the functional diversity creation is mostly based on the experiments conducted thereafter, and the rules of thumb for directed evolution continue to be clarified, experiment by experiment, allowing the future scientists to gain results of higher impact with less effort. The representative examples of the recent findings that the scaffolds with a polar active site-core structure correlate with evolvability [313] and that the tyrosine is a key residue for establishing molecular interactions at binding surfaces [122], will most likely turn into new guiding principles.

3 AIMS OF THE STUDY

Directed evolution is a central method for developing affinity reagents and enzymes for biotechnological industry, medicine and diagnostics. The aim of this study was to advance directed evolution research with novel methods for high quality gene library construction and for the improved display of proteins of interest on phage. As antibody phage display is an increasingly applied method, the intention was also to generate information on the differences between antibody diversification schemes and filamentous phage display platforms for maximum performance.

More specifically the aims were:

1. To improve an existing primer extension mutagenesis method with additional enzymatic reactions to produce larger libraries with a higher share of mutated members.

2. To enhance the efficiency of error-prone RCA mutagenesis by tapping more transformants via genetic rearrangement.

3. To develop tools for improved Fab display on M13 phage by directed evolution.

4. To explore the properties of libraries derived from a single antibody scaffold via independent diversification schemes.

5. To compare the performance of antibody fragment libraries displayed on filamentous phage as p3, truncated p3, and p9 fusions in biopanning experiments.

4 SUMMARY OF MATERIALS AND METHODS

A detailed description of the materials and methods can be found in the original publications (I-V). A summary of the used methods and most central materials are presented in **Tables 2**, **3**, **4** and **5** below. Only the novel methods developed in this study are described in detail in section 4.2.

4.1 Bacterial strains, vectors, reagents and methods

Table 3. Bacterial strains.

Bacterial strain	Genotype	Source	Used in
E. coli XL1-Blue	*hsdR17(rK- mK+) endA1 gyrA96(nal^R) thi-1 recA1 relA1 supE44 lac* *F'[::Tn10 proAB$^+$ lacI^q Δ(lacZ)M15]*	Agilent Technologies, USA	I-V
E. coli SS320 (MC1061 F´)	*hsdR mcrB araD139 Δ(araABC-leu)7679 lacX74 galU galK rpsL thi-1* *F'[::Tn10 proAB$^+$lacI^q Δ(lacZ)M15]*	[314]	I-V
E. coli K12 CJ236	*FΔ(HindIII)::cat (Tra+ Pil+ Cam^R) ung-1 relA1 dut-1 thi-1 spoT1 mcrA*	New England Biolabs, USA	I

Table 4. Methods summary.

Method	Used in
PCR, molecular cloning and sequencing	I-V
Kunkel mutagenesis	I
Selective RCA	**I**
Error-prone RCA	II
Concatemer reduction by Cre/LoxP recombination	**II**
Error-prone PCR	III
Phage stock preparation, phage titer determination and biopanning	I,III,IV,V
Protein expression and cell lysis	III,IV,V
MIC determination	II
Western blotting	III
TRFIA and ELISA screening	III,IV,V
Phage immunoassay	III,IV,V
Display efficiency determination	**III,V**
IC50 assays	IV,V
Biolayer interferometry	IV,V
Protein purification	IV

Methods written in bold were developed in this study. See detailed description below.

Table 5. Vectors used for display and screening.

Vector	Description	Resistance	Source	Used in
pEB91	Phage display as fusion to p9. Amber stop between POI and p9.	cm	[119]	I,II
pEB30	Phage display as fusion to full-length p3.	cm	III	III,V
pEB32x	Phage display as fusion to truncated p3.	cm	IV	IV,V
pEB92	Phage display as fusion to p9.	cm	V	V
pEB07	Soluble periplasmic expression as a fusion to β-lactamase.	cm	I	I,IV,V
pAK400	Enhanced periplasmic expression. His tag purification.	cm	[362]	III,V
pAK400ampR	Enhanced periplasmic expression. His tag purification.	amp	II	II
pLK06	Soluble periplasmic expression as fusion to alkaline phosphatase.	amp	IV	IV,V
pLK06H	Soluble periplasmic expression as fusion to alkaline phosphatase. His tag purification.	amp	IV	IV
pLK01	Soluble periplasmic expression as fusion to mouse CL. Binds to RAM plate. His tag purification.	amp	IV	IV

cm: chloramphenicol and amp: ampicillin.

Table 6. Reagents used in the study.

Reagent class	Description	Source	Used in
Helper phage	VCS-M13	Agilent Technologies, USA	I,III,IV,V
Helper phage	Hyperphage M13 K07ΔpIII	Progen Biotechnik, Germany	III,V
Antibody, polyclonal	Rabbit anti-fd (anti-phage)	Sigma-Aldrich, USA	III
Antibody, monoclonal	Mouse anti-M13 9E7	University of Turku, Finland	III
Antibody, monoclonal	Mouse anti-M13 pIII	New England Biolabs, USA	III
Antibody, polyclonal	Sheep anti-mouse IgG HRP conjugated	Amersham Int., UK	III
Antibody, polyclonal	Goat anti-mouse IgG (H+L) biotin conjugated	Zymed, USA	III
Antibody, monoclonal	Anti-PentaHis HRP conjugate	Qiagen, Germany	III
Antibody, monoclonals	Mouse anti-PSA 5A10, 5E4 and 2E9	University of Turku, Finland	IV
Antibody, monoclonals	Mouse anti-PSA H50	Abbott Laboratories, USA	IV
Antibody, polyclonal	Goat anti-rabbit IgG (Forton Bioscience, Morrisville, NC, USA)	Forton Bioscience, USA	IV

Affinity reagent	Streptavidin HRP conjugated	Zymed, USA	III
Affinity reagent	Avidin	Sigma-Aldrich, USA	IV
Antigen	Streptavidin	BioSpa, Italy	IV,V
Antigen	Digoxigenin, digoxigenin-NHS-ester	Sigma-Aldrich, USA	IV,V
Antigen	Microcystin-LR	Åbo Akademi, Finland	IV
Antigen	PSA	University of Turku, Finland	IV
Beads	Dynabeads MyOne Streptavidin T1	Invitrogen, Norway	IV,V
Beads	Dynabeads M-280 Streptavidin	Invitrogen, Norway	IV,V
Beads	Sphero Avidin Magnetic Particles	Spherotech, USA	IV
Plate	96-well rabbit anti-mouse IgG	Kaivogen, Finland	III,IV,V
Plate	96-well streptavidin	Kaivogen, Finland	III,IV,V
Enzyme	Pfu DNA polymerase	Fermentas, Lithuania	I,IV
Enzyme	Phire Hot Start II DNA polymerase	New England Biolabs, USA	I
Enzyme	Phusion DNA polymerase	New England Biolabs, USA	III,IV,V

Enzyme	Phi29 DNA polymerase	Fermentas, Lithuania	I, II
Enzyme	Inorganic Pyrophosphatase	Fermentas, Lithuania	I,II
Enzyme	T4 phosphonucleotide kinase	Fermentas, Lithuania	I
Enzyme	T7 DNA polymerase	Fermentas, Lit./ NEB, USA	I
Enzyme	T4 DNA ligase	Fermentas, Lithuania	I-V
Enzyme	Calf intestine alkaline phosphatase	Fermentas, Lithuania	I, II
Enzyme	Restriction enzymes	Fermentas, Lit./ NEB, USA	I-V
Enzyme	Uracil DNA-glycosylase	New England Biolabs, USA	I
Enzyme	Cre recombinase	New England Biolabs, USA	II
General reagent	Random hexamer primer	Fermentas, Lithuania	I
General reagent	Exoresistant random primer	Fermentas, Lithuania	I

4.2 Methods in detail

4.2.1 Selective RCA

Selective RCA (sRCA) is an enhanced version of the primer extension mutagenesis method described by T.A. Kunkel (schematic illustration of Kunkel mutagenesis in **Figure 6**) [155]. In Kunkel mutagenesis a uracil-containing single-stranded DNA is used as a template for the hybridization of a mutagenic primer. The extension of the primer and the ligation of the remaining nick yield a covalently closed circular dsDNA molecule. In sRCA, this product is treated with UDG and amplified by RCA instead of direct transformation. The removal of uracil bases by UDG from the template strand leads to the selective amplification of the nascently synthesized intact circular DNA strand carrying the desired mutations (**Figure 10**).

This method and variants thereof are also described in article I. For a large-scale library construction, Kunkel-type primer extension mutagenesis was carried out with 5 µg template ss(U)DNA as described in [155; 314; 363]. The formed heteroduplex was treated with 10 U UDG at 37 °C for 1 h and purified with Qiagen PCR purification kit (Qiagen, Hamburg, Germany) to 50 µl 10 mM Tris-HCl (pH 8.5). A 2 µl sample (240 ng i.e. 4×10^{10} molecules of vector pEB91-ScFv) was amplified in 200 µl phi29 DNA polymerase reaction buffer containing 1 mM dNTPs, 50 µM random primers, 4 mM DTT, 0.25 U inorganic pyrophosphatase, and 100 U phi29 DNA polymerase. The reactions were incubated overnight at 30°C and heat-inactivated at 70°C for 10 min. The DNA concatemer produced in RCA by phi29 DNA polymerase was digested with 200 U HindIII to single-plasmid sized units in a 1 ml reaction volume overnight at 37°C, and purified with miniprep kit columns (Qiagen, Hamburg, Germany) in the following way. The 1 ml digest was mixed with 5 ml binding buffer, applied in a 10 ml TERUMO-syringe (Terumo Corp., Tokyo, Japan), which was fastened tightly in a miniprep column. 3 ml of the digest per column was pressed through and, after loading, the purification was continued according to manufacturer's instructions with a final 100 µl elution volume in 10 mM Tris-HCl (pH 8.5).

The HindIII-digested DNA was self-ligated overnight at 16°C by using 5 ng/µl DNA with 0.05 U/µl T4 DNA Ligase in the manufacturer's (Fermentas, St. Leon-Rot, Germany) recommended buffer containing 40 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT and 0.5 mM ATP. The concentration and purification of the ligation was carried out to 50 µl volume containing 230 ng/µl DNA with the syringe-method described above. Ligation was electroporated to freshly prepared SS320 cells by loading 230 ng DNA with 70 µl cells per 0.1 cm gap-with cuvette (Bio-Rad, USA) at settings 200 Ω, 1.25 kV, 25 µF (Gene Pulser, Bio-Rad, USA). Each cuvette was washed twice with 1 ml SOC, recovered at 37 °C with 100 rpm shaking for 1 h and plated on 24x24 cm Bioassay Dishes containing LA with 25

µg/ml cm, 10 µg/ml tet and 1% glucose, 20 ml recovery per plate. Plates were left to dry 1 -2 h and grown o/n at 30 °C.

4.2.2 Concatemer reduction by Cre/loxP recombination

Error-prone RCA is carried out in a phi29 DNA polymerase catalyzed reaction in the presence of $MnCl_2$ with a similar method as described by Fujii et al. [100]. For the concatemer resolution via recombination, a loxP site is inserted into the plasmid prior to amplification. In this study, loxP was ordered as two complementary oligonucleotides that, upon hybridization, readily created HindIII-compatible cohesive ends. The other HindIII recognition sequence in the loxP-cassette was ordered as non-palindromic with a designed point mutation to destroy the site in order to retain HindIII as a unique restriction site in the final loxP construct.

This protocol can also be found in article II. Instead of using Templiphi kit in the error-prone RCA as used by Fujii et al., the reaction was self-formulated and contained the following components: the pAK400ampR vector with inserted loxP site was heat-denatured at 95 °C for 3 min. For error-prone RCA, 25 pg plasmid was amplified in 20 µl 1 x phi29 DNA polymerase buffer with 1.5 mM $MnCl_2$, 1 mM dNTPs, 50 µM random hexamers or exo-resistant random primers, 0.05 U inorganic pyrophosphatase, and 20 U phi29 DNA polymerase at 30 °C o/n. For recombination, the RCA reaction was heat-inactivated at 70 °C for 10 min and 10 µl of the reaction was diluted in 50 µl Cre recombinase buffer supplemented with 1 U Cre recombinase and incubated at 37 °C for 2 h. Reaction was heat inactivated at 70 °C for 10 min and EtOH-precipitated with Pellet Paint Co-precipitant (Merck, Darmstadt, Germany) to 10 µl volume. The original template DNA was removed by digesting 100 ng of the precipitated products with 2 U DpnI in 10 µl volume at 37 °C for 2 h and transformed to *E. coli* SS320 cells.

4.2.3 Display efficiency determination

Display efficiency was determined as Fab (displayed molecule) signal per phage. The number of virion particles in the samples was determined by immunoassay. Phage quantitation by immunoassay requires a standard prepared in the same way as the samples, i.e. 2 x PEG/NaCl-precipitation, with a very well-known infectivity. The study Huovinen et al. 2010 involved multivalent Fab phage stocks and, therefore, a hyperphage stock of the Fab clone 4D5 was chosen as the relative phage particle number standard (rcfu, relative colony forming units) [article III]. The infective titer of the Fab4D5 stock was determined in three independent experiments by counting cmR-colonies. A dilution series of the Fab 4D5 phage stock and sample phage stocks were applied on streptavidin plates coated with biotinylated mouse anti-M13 antibody. Bound phages were detected with a europium-labelled rabbit anti-fd (anti-phage) Ab. The titers of the

unknown samples were determined from the linear-fit of the standard curve (only points at the linear response range included).

Display efficiency determination is also described in article III, but a more detailed protocol is provided here. All washing steps in the phage titer immunoassay were done with 1296-026 Delfia Platewash using Innotrac Wash Buffer and all dilutions were done in Innotrac Assay Buffer containing 50 mM Tris–HCl, pH 7.75, 150 mM NaCl, 0.5 g/l NaN_3, 0.1 g/l Tween 40, 0.5 g/l bovine-γ-globulin, 20 μM DTPA, 5 g/l bovine serum albumin (BSA), and 20 mg/l cherry red. Biotinylated anti-M13 Ab was added on STR wells, 100 ng per well, and incubated at room temperature with slow shaking for 30 min. The plate was washed four times and sample phages (tenfold serial dilutions, $1{:}10^2$ – $1{:}10^5$) and standard dilutions (10^5 cfu/ml - 10^8 cfu/ml) added 100 μl /well, incubated at room temperature with slow shaking for 1 h and washed four times. Europium-labelled anti-phage Ab was added, 13 ng per well, and incubated for 1 h at room temperature with slow shaking. The label was washed four times, the wells enhanced with Delfia Enhancement solution for further 15 min, and the time-resolved fluorescence signal measured with Victor 1420 Multilabel Counter (Wallac Oy, Turku, Finland).

For the determination of the Fab signal, another phage immunoassay was performed on rabbit anti-mouse IgG plates that capture the phage via displayed Fab-moiety. The same amount of phage particles (10^6 rcfu) were applied per well and the assay performed as described above. The display efficiency of the phage clones was calculated in relation to the RAM signal of the original FabD phage stock with the lowest Fab display level (value: 1). In article V, the display efficiency was calculated in relation to the highest RAM/phage-signal in the assay, i.e. hyperphage-superinfected pEB30-Fab phagemid.

5 SUMMARY OF RESULTS AND DISCUSSION

5.1 Selective RCA as a novel method to improve primer extension mutagenesis

5.1.1 Kunkel vs. sRCA - summary of experiments

The motivation for developing Kunkel mutagenesis further arose directly from the lab work as it was noticed that Kunkel mutagenesis resulted in an equal mixture of mutants and template, or sometimes even in a dominant share of the template background [article I]. This is an inherent property of Kunkel mutagenesis, which has been earlier solved either by using stop codons or restriction enzyme recognition elements at the site of mutagenesis [119; 314]. However, in the stop-codon-strategy the template is still present as a major contaminant in the library, albeit unable to be displayed, and in the latter template digestion-strategy, an additional purification step of the transformed library as dsDNA is required for efficient background removal followed by re-transformation. In addition to losses in time and library quality, both of these strategies require prior manipulation of the template sequence and, therefore, they are not universally applicable to all sequences.

As Kunkel product is circular DNA, it is an ideal substrate for rolling circle amplification. Initially, it was reasoned that nicking the template strand would destroy the amplification potential of the template, as only the nascently synthesized strand would be intact circular DNA. This was also proven to be an effective method with a USER enzyme mixture, containing both UDG for uracil excision and a DNA lyase for phosphate backbone nicking at the abasic sites. By further examination of the process, it was deciphered that only UDG is needed to confer the selectivity. This is most probably due to the fact that an abasic site in the template DNA is a strong stalling signal for the family B DNA polymerases [364], including phi29 DNA polymerase. Therefore, phi29 DNA polymerase-driven RCA leads to selective amplification of only the nascently synthesized mutant strand (**Figure 10**).

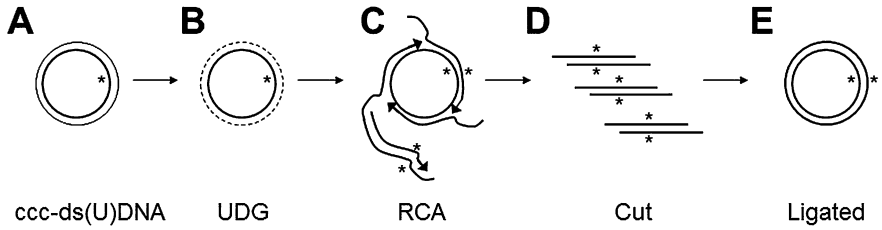

Figure 10. Selective rolling circle amplification (sRCA) flowchart. The heteroduplex, ccc-ds(U)DNA formed in Kunkel mutagenesis (A), is treated with UDG (B) and subsequently amplified with RCA (C) using random hexamers as primers. The resulting DNA concatemer is cut to plasmid-sized units (D) and re-circularized by self-ligation (E) for host transformation. From article I.

These findings were confirmed in three separate studies [article I]. The first proof-of-principle experiment was carried out with a CDR-H1 oligonucleotide coding for 30 genetic variants. An AatII restriction site was present in the template ScFv gene at the site of CDR-H1 for genetic analysis of the outcome. Only 3/10 clones originating from transformed Kunkel product were mutated by analyzing single DNA clone preparations by AatII digestion, whereas after selective RCA, all the picked ten clones were mutants (**Table 7**). In the next experiment, the achievable library size and compatibility to incorporate PCR products was explored. Fourteen VH genes were incorporated into a uridylated VL-library present as ssDNA and consisting of 1×10^8 members. Transformation of 240 ng UDG-treated Kunkel product directly and after RCA, resulted in a library of 3×10^7 and 1×10^{10} cfu, respectively. In this experiment the mutagenesis efficiency was determined for the Kunkel reaction treated with UDG, which was surprisingly high (85%) compared to earlier findings. Naturally, the UDG treatment itself increases the mutagenesis efficiency, but only by 10% in our experiments [article I]. Selective amplification raised the efficiency to 100%.

In the final experiment a ScFv-β-lactamase fusion gene was used as the template containing three stop codons at the CDR-H3 site [article I]. The mutagenesis outcome was easily followed in this system by plating transformed cells on selective agar plates. The mutagenic primer contained seven consecutive NNN codons with a calculated complexity of 10^9 variants. The implementation of sRCA produced similar improvement in the amount of DNA and the number of transformants in both the CDR-H3 mutagenesis experiment as in the preceding VH gene incorporation experiment. The mutagenesis efficiency was increased from 45% obtained with Kunkel mutagenesis to 100% with sRCA. In this experiment, unmodified random hexamers were used, whereas with thio-resistant hexamers only 90% mutagenesis efficiency was obtained. Surprisingly, mutagenesis efficiency was decreased to 0.1% by RCA without UDG treatment indicating the strict dependency of selectivity on the removal of uracils from the template strand.

Table 7. Summary of the sRCA experiments.

Experiment	CDR-H1 mutagenesis	CDR-H3 mutagenesis	VH gene incorporation
Incorporated diversity[a]	30	1×10^9	14
Combinatorial diversity[b]	30	1×10^9	$1x\ 10^{10}$
Kunkel, DNA	N.D.	10 ng	240 ng[c]
after sRCA, DNA	N.D.	525 ng	12 µg
Kunkel, cfu	N.D.	$4x\ 10^5$	3×10^{7c}
after sRCA, cfu	N.D.	8×10^7	1×10^{10}
Kunkel Mut-%	30%	45%	85%[c]
after sRCA Mut-%	100%	90 – 100%[d]	100%

[a]Theoretical number of unique variants in the incoming variable DNA element.

[b]The pre-existing diversity in the template taken into account.

[c]After UDG treatment.

[d]90% efficiency with exoresistant and 100% with unmodified random primers.

5.1.2 Elimination of template-mutant coexistence by sRCA

Mutagenesis efficiency in the CDR-H3 experiment was re-analysed using colony PCR [article I]. In addition to stop codons, a SacII site was present in the template at the CDR-H3 site. Therefore, the amplicon is resistant to cleavage by SacII, if mutagenesis has been successful. In a sample group of 20 colonies originating from the Kunkel sample 8/20 amplicons were perfectly cleaved by SacII, 8/20 were cleavage-resistant and the remaining 4/20 were only partially digested (**Table 8**). A control amplicon derived from the wild-type template was

totally digested by SacII in the same conditions indicating that both mutated and unmutated phagemids had propagated in the 4/20 clones. No partial digestion was observed among the tested sRCA clones in the SacII test. Naturally, in sRCA with exoresistant primers 2/20 amplicons were fully sensitive to SacII as they were template clones.

To verify the presence of double templates in the Kunkel sample, we decided to test a new set of colonies from the selective ampicillin-containing plate, which only allows growth of the mutants with the open reading frame. A pair of primers was used: the forward primer hybridized only to the wild-type template sequence at the CDR-H3 site and the reverse primer hybridized outside the mutated region. In this way, smaller amounts of template DNA could be detected than by digesting the amplicon. In the case of Kunkel heteroduplex, 11/34 clones yielded PCR product indicating that the unmutated template DNA was still present in 32% of the clones considered to be mutants [article I]. Template specific PCR of ampR colonies originating from the sRCA transformation yielded no PCR products.

Table 8. Colony PCR screen of clones created with primer extension mutagenesis of CDR-H3 loop.

Treatment			K	K+U	K+U+R	K+U+R
Random hexamers			none	none	exoresistant	normal
Source plate	Primers	Analysis				
cm (all grow)	WO375 & B1 (All amplified)	SacII-resistant	8/20	9/20	18/20	20/20
		SacII-sensitive	8/20	8/20	2/20	0/20
		Partial digestion with SacII	4/20	3/20	0/20	0/20
cm & amp (only mutants grow)	A1 & pAK400rev (wild type template amplified)	Presence of template DNA	11/34	16/34	0/31	0/34

K: Kunkel, U: UDG, R: RCA, cm: chloramphenicol, and amp: ampicillin.

5.1.3 Extensions of sRCA and future prospects

The template in Kunkel heteroduplex may be rendered unfavourable for rolling circle amplification also by other strategies than uracil incorporation. Nicking endonucleases cut one DNA strand in their recognition element depending on the orientation of the sequence element [365]. This phenomenon was utilized by treating a primer extended covalently closed dsDNA heteroduplex with a nicking endonuclease that cuts only the template strand, after which the intact strand carrying the mutations was favourably amplified in the following RCA reaction (data not shown). However, the development of this method was not continued further as uracil-based sRCA is at least equally potent. Moreover, the nicking endonuclease-dependent method requires the presence of a nicking site that cuts only the template strand, whereas the uracil strategy is universal. Although sRCA has been demonstrated to build *E. coli* hosted libraries, the ability to produce unlimited amount of mutated DNA may find even more use in library construction in other host organisms such as yeasts, e.g. *Kluyveromyces lactis*, in which the heterologous gene is directly targeted for chromosomal integration requiring a large amount of linear DNA in transformation [366].

5.2 Enhancement of error-prone RCA mutagenesis by concatemer reduction to circular monomeric units

Error-prone RCA is an easy-access tool for whole-plasmid mutagenesis. It is claimed to be the most simple mutagenesis method as there is no need for a thermal cycler or specific primers [367]. RCA is turned error-prone by $MnCl_2$ and by reducing template DNA concentration. In the earlier reports on error-prone RCA, the DNA concatemer formed in the multiply-primed RCA reaction is directly transformed to host and segregation to single plasmid units is speculated to occur by homologous recombination in the host cells [100]. In a test reaction to modulate the catalytic specificity of β-lactamase to hydrolyze also ceftazidime, none or only few ceftazidime resistant (ctzR) colonies were obtained due to very low transformation efficiency [article II]. Therefore, methods were sought to improve the transformability of the DNA concatemer in *E. coli*, which can be achieved, for example, by modifying the DNA into plasmid-sized circular units.

The recircularization may be accomplished by digestion with a unique restriction enzyme and subsequent ligation in conditions favouring self-ligation [368]. This method, however, requires two enzymatic reactions and a DNA purification step. In this study, this procedure was replaced with Cre/loxP recombination containing a single recombinase addition and incubation step to yield highly transformable circular DNA [article II]. When loxP site is present in the target plasmid the DNA concatemer produced in a RCA reaction is an ideal substrate for the Cre recombinase catalyzing the excision of the intervening DNA between two adjacent loxP sites in a circular form (**Figure 11**) [369].

This concept was tested on an ampicillin-resistant plasmid carrying the TEM-1 β-lactamase gene by first inserting a loxP element in the plasmid and then performing RCA in error-prone and nonerror-prone conditions. It was found that digestion-ligation and recombination performed with equal efficiency and were superior to direct concatemer transformation in terms of the number of obtained transformants. Recombination produced 20 - 50-fold and 13-fold more transformants per reaction than the direct transformation of the concatemer, in nonerror-prone and error-prone conditions, respectively [article II]. The less significant increase in the number of transformants observed in the error-prone condition is most probably explained by the inhibitory effect of $MnCl_2$ on the amplification. Due to the reduced amplification, the concatemer unit at the end of the reaction is smaller and, therefore, more easily transformed than in nonerror-prone conditions.

Based on seven experiments, the average numbers of ampR colonies with and without recombination were 13 036 ± 6700 cfu and 1183 ± 596 cfu per error-prone RCA reaction, and the numbers of ceftazidime resistant colonies were 115 ± 57 and 9 ± 11 cfu, respectively [article II]. The practical benefit of concatemer resolution is that a full 96-well plate of ctzR colonies can be picked from a single transformed reaction as was also demonstrated in this study. To obtain the same number of colonies by the transformation of the intact concatemer, the collection of all transformants from ten reactions is required. Naturally, inserting loxP into the template plasmid requires some effort, but especially in repeatedly used vector systems, loxP option is a time-saving solution. A major limitation of epRCA is the high fidelity of phi29 DNA polymerase [370] limiting the acquired mutagenesis efficiency. A modified version of phi29 DNA polymerase retaining strand displacement activity and processivity, but having an inherently decreased fidelity, would be a more ideal enzyme for epRCA.

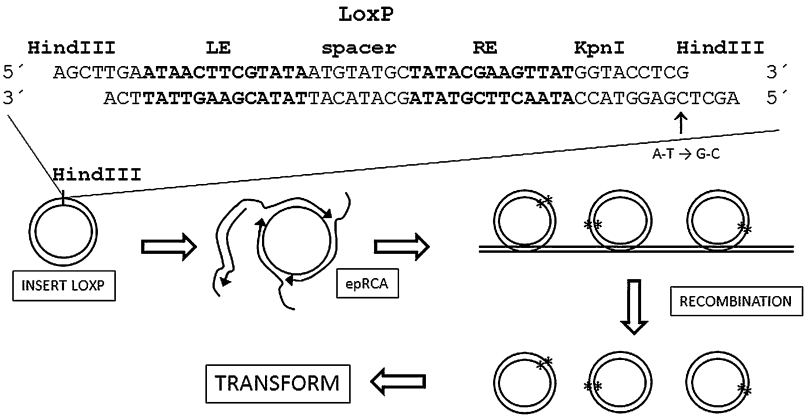

Figure 11. Error-prone RCA with concatemer resolution by Cre/loxP recombination. A LoxP cassette is hybridized from two complementary DNA strands creating directly cohesive ends. The cassette is inserted into a circular plasmid. After error-prone RCA the resulting concatemer is resolved into monomeric circular units with Cre recombinase.

5.3 Directed evolution as a powerful tool to improve oligovalent Fab display on M13 phage

5.3.1 Improved chimeric Fab display on phage by the use of iterative cycles of phage production and selection with rabbit anti-mouse IgG antibodies

Small antigens have only few epitopes available for antibody binding. Consequently, multivalent phage display systems would be beneficial to rescue also those clones that have at least a weak interaction with the antigen. Bearing this in mind, a chimeric Fab template was constructed consisting of codon-optimized human variable domains and mouse constant domains and displayed on filamentous phage with hyperphage superinfection. Hyperphage is a filamentous phage without p3 gene and, therefore, all p3 in hyperphage-packed particles originates from the expressed phagemid [217]. In combination with full-length p3-Fab expression from a phagemid, every p3 should carry the displayed fusion moiety.

The initial multivalent display efficiency of the chimeric Fab on the phage was poor. In order to improve the display, the Fab cassette was randomly mutated by epPCR, displayed on phage and selected on microtiter plates coated with rabbit anti-mouse IgG antibodies (RAM). Selections were carried out both in multi- and monovalent formats using lower (B) and higher (A) mutagenesis frequency. Enrichment for better display was observed in all stocks, but it was superior in libraries with higher mutation frequency (**Figure 12**) [article III].

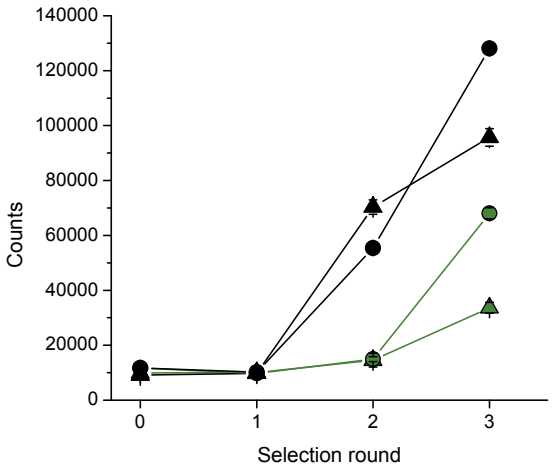

Figure 12. Phage immunoassay of the enrichment of Fab-displaying phages. 1 x 10^7 phages of each panning round were applied to Maxisorp wells coated with rabbit anti-mouse IgG. Bound phages were detected with a europium-labelled anti-phage antibody. Circles: oligovalent libraries; triangles: monovalent libraries; black: more mutated library A; and green: less mutated library B. Adapted from article III.

5.3.2 Two mutations found to affect phage propagation and display efficiency

Clones were picked from the third round of panning for primary and secondary screening and nine confirmed positive clones were chosen for closer genetic analysis. Major deletions in the Fab molecule, or a frameshift, was found in all chosen 5/5 library A members and thus, only the less mutated library B clones turned out to be a valuable source of candidates for further studies [article III]. Two different genotypes were observed among the four sequenced library B members, one of which contained an amber stop codon in VH, which abolishes Fab expression in a non-supressor host. Logically, further investigations were concentrated on the remaining clone with seven mutations that were segregated from each other to study their individual effects in more detail.

Two out of seven mutations were found to have a profound effect on the phage biology [article III]. One of them, $Y102_LF$ (FabF), was located in the core of variable light domain close in sequence to the CDR-L3 loop, while the other mutation caused a change in the start codon of the PelB signal sequence of the heavy chain to code valine instead of methionine (FabM$_{GUG}$). Hyperphage-produced FabF and FabM$_{GUG}$ variants had titers ten- and twofold higher, and Fab display efficiencies 12- and 19-fold higher than the wild type FabD, respectively (**Table 9**). The phage titer of the double mutant was additively increased but the display efficiency was only slightly, if at all, enhanced.

Table 9. Genetic determinants and phage titers of a selected set of clones assayed for display.

Clone	Source	Mutations	Phage titers (rcfu/ml)[a]	Display efficiency
FabD	Original	none	$1.6 \pm 0.1 \times 10^8$	1.0 ± 0.34
FabA1	Phage display	Y102F (VL), M1^{SS}M$_{GUG}$ (PelB-Fd), L19M (VH), L46Q (VH), A55T (VH), T90N (CH1), P123Δ (CH1)	$4.3 \pm 0.2 \times 10^9$	18.5 ± 1.45
FabF	Chimera of FabD & FabA1	Y102F (VL)	$1.2 \pm 0.1 \times 10^9$	11.6 ± 1.33
FabM$_{GUG}$	Chimera of FabD & FabA1	M1^{SS}M$_{GUG}$ (PelB-Fd)	$3.0 \pm 0.2 \times 10^8$	19.0 ± 0.37
FabFM$_{GUG}$	Chimera of FabD & FabA1	Y102F (VL), M1^{SS}M$_{GUG}$ (PelB-Fd)	$3.7 \pm 0.5 \times 10^9$	21.4 ± 3.21

[a] Error bars represent the SD of three independent experiments.

5.3.3 Forces of selection and forced display

There are clearly two selection pressures in the described panning experiment aimed at improving the oligovalent display of Fab molecules. One is the selection for the ability to bind to the provided coating (RAM) on the microtiter plate and the other is the selection for better propagation in the phage amplification phase. As shown with the cell culture assay, the Fab expression in general has toxic effects on the cells, which are further enhanced by isopropyl-β-D-thiogalactopyranoside (IPTG) induction (**Figure 13**) [article III]. Especially, the growth of the original FabD-p3 fusion was the most severely retarded. Against this background, it is easy to conceive that in order to gain a healthier living the Fab gene must be dropped or the expression burden alleviated. This was also observed in most of the studied variants as deletions, frameshifts, or amber stop codons.

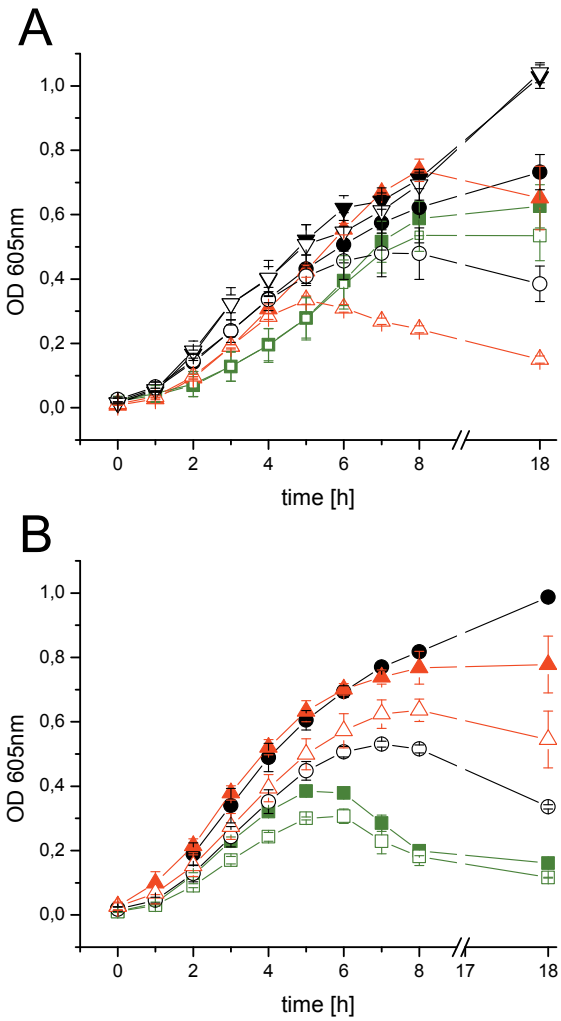

Figure 13. Growth curves of selected clones. (A) Growth curves of clones expressing soluble Fab with 100 µM IPTG (open symbols) and without (filled symbols). (B) Growth curves of clones expressing Fab–p3 fusion with 100 µM IPTG (open symbols) and without (filled symbols). FabD: squares (green); FabF: circles (blue); FabM$_{GUG}$: triangles up (red); XL1-Blue strain without plasmid: triangles down (black). Data are expressed as the mean ± SD of three independently grown cultures. Adapted from article III.

The most striking example of a fit clone was the most wide-spread library A variant lacking the heavy chain altogether and containing a direct light chain fusion to the p3 [article III]. This construction is half the size of a Fab, can easily produce viable phage particles in the forced display system, and is able to bind on the RAM plate. According to the results of this study, caution must be taken in applying the hyperphage system. In the forced display concept, an equal valency display of all variants is anticipated, but this is clearly variating and it seems that even some bald phages are also present. This is most probably explained by

differences in the proteolytic degradation propensity of the fusion moieties, which has also been observed by others [371]. Furthermore, survival of out-of-frame constructs was encountered in this and also in another study on applying hyperphage to enrich open reading frames from a cDNA fragment library [372].

5.4 Augmentation of the overall functional performance of an antibody library by complementary diversity created by independent randomization schemes

5.4.1 Key differences in library design of ScFvP and ScFvM repertoires

The principle design features of the ScFvP library (anti-protein repertoire) were to set the majority of the diversity in the heavy chain, especially at the apex of CDR-H1, CDR-H2, and CDR-H3 loops. Moreover, there was length variation from 7 to 18 residues in CDR-H3 closely following the natural human CDR-H3 repertoire with a mean at 13 residues [373]. In ScFvM (anti-multipurpose) repertoire, the light chain was more diversified than in the ScFvP repertoire by adding variation into the CDR-L1 and CDR-L3 loops. The heavy chain diversity in the ScFvM design, regarding the CDR-H1 and CDR-H2 loops, was shifted "down the β-strand", when viewed from the top of the paratope towards the centre of the antigen binding site, to support contacts with peptide and hapten targets (**Figure 14**).

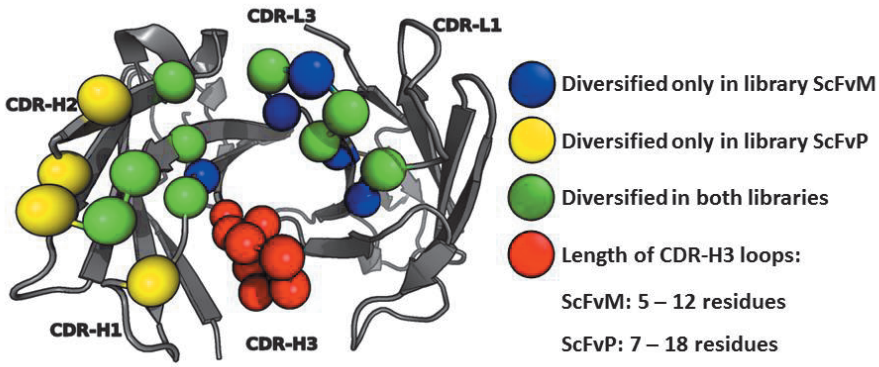

Figure 14. Cartoon presentation of the randomized positions on VH (left) and VL (right) looking at the antigen binding site from above. The positions having variation only in the ScFvP and ScFvM repertoire are marked with yellow and blue, respectively. The shared randomized positions are coloured green and the positions randomized in the CDR-H3 loop are marked in red. Adapted from article IV.

The loop length variation in the CDR-H3 of ScFvM repertoire was limited to short loops of 5 to 12 residues. Although the mean CDR-H3 length of the

observed anti-hapten antibodies in natural repertoires, at ten residues, is on the upper limit of the input lengths [374], it is necessarily not the optimal length distribution for hapten recognition, but an innate constraint from germ-line repertoire and the editing process. Therefore, a shorter CDR-H3 repertoire than that provided by nature was explored as the source of antigen-binding motifs. Moreover, we observed that 86.3% of those antibodies with very short loop lengths, 4 - 7 aa, comprising only 1.7% of all human antibodies are lacking the salt bridge Arg(Lys)106H-Asp116H (R106H-D116H) at the CDR-H3 stem region (**Figure 15**). This observation was taken into account in the ScFvM repertoire by severely restricting the occurrence of the R106H-D116H salt bridge.

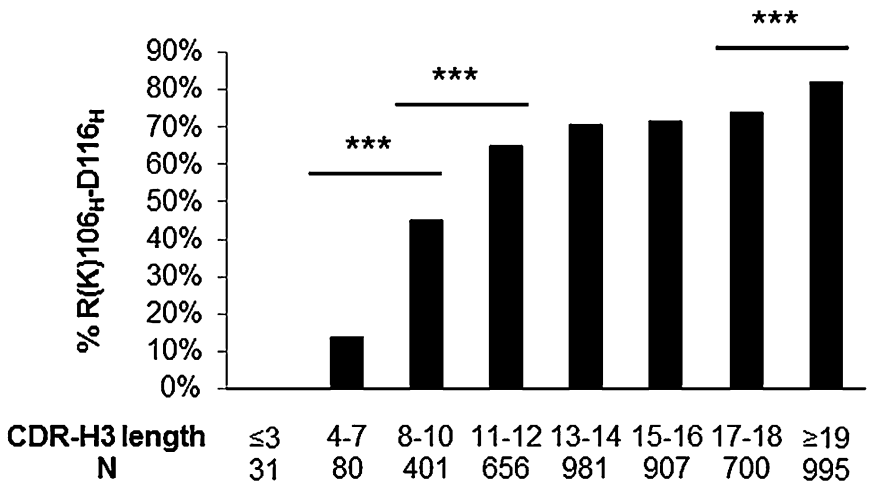

Figure 15. Salt bridge R(K)106H-D116H frequency in human CDR-H3 (N = 4751) length groups (IMGT definition). The analysis is based on the supplementary data set provided by Zemlin et al. [373]. The overall salt bridge frequency differed very significantly from the length-independent distribution ($\chi 2$ value 318, 6 degrees of freedom, $p < 0.0000001$, loop lengths ≤ 3 excluded as they cannot form a salt bridge). Neighbouring length groups with significantly different salt bridge frequencies are highlighted ($\chi 2$ test of independence, *** = $p < 0.0000001$). From article IV.

5.4.2 ScFvM is a superior source of peptide and hapten antibodies

ScFvM and ScFvP repertoires were panned in parallel and as a mixture against different antigens. In parallel selections against streptavidin, the same number of unique antibodies was obtained from both repertoires with the highest measured apparent affinity found from the ScFvP repertoire (**Table 10**) [article IV]. In parallel anti-digoxigenin (anti-DIG) selections, the ScFvM library provided

higher affinity variants than the ScFvP repertoire. The predominant share of retrieved antibodies from the mixed panning experiments that recognized PSA originated from the ScFvP library, whereas all the retrieved anti-McLR and anti-DIG antibodies originated from the ScFvM repertoire based on the origin identification by sequencing. A closer scrutiny of the sequence details of the ScFvM members revealed that the highest affinity anti-DIG and anti-McLR antibodies contained short CDR-H3 loops of four and five residues, respectively, and were without the canonical salt bridge, whereas the CDR-H3 of the highest affinity anti-PSA antibody contained eleven residues with a constraining salt bridge [article IV]. A summary of selections is presented in **Table 10**.

Table 10. Panning and screening summary.

Input library	Antigen	Size g/mol	Screen I	Screen II	Unique	Origin	Highest Observed Affinity
M	STR	60 000	88/96	8/8	6/8	M: 6	Kd[a]: 6.3 ± 1.6 nM (M-STR1)[b]
P	STR	60 000	72/90	8/8	6/8	P: 6	Kd[a]: 2.2 ± 1.0 nM (P-STR3)[b]
M	DIG	391	47/87	8/8	6/8	M: 6	IC50: 158 ± 20 nM (M-DIG1)[b]
P	DIG	391	59/87	8/8	4/7	P: 4	IC50: 1544 ± 42 nM (P-DIG3)
M + P	PSA	30 000	73/186	51/73	19/49	M: 6 P: 11	Kd: 5.3 ± 1.0 nM (M-PSA1)
M + P	MCLR	995	19/186	18/19	6/17	M: 6 P: 0	IC50: 22.8 ± 2.1 nM (M-MCLR1)
M + P	DIG	391	44/186	41/44	3/39	M: 3 P: 0	IC50: 361 ± 74 nM (M-DIG8)[b]

Antigens: STR: streptavidin; DIG: digoxigenin; PSA: prostate specific antigen; and MCLR: microcystin-LR. Screen I: primary screening (pos./all) and Screen II: secondary screening (pos./ selected). All affinity data is based on at least three independent measurements. [a]Apparent affinity of a dimer binding to streptavidin. [b]Measured with purified proteins.

5.4.3 Selection outcome with ScFvM and ScFvP repertoires indicate complementary functional diversity

The final selection of clones against PSA from the mixed library sample was performed with two PSA-capturing antibodies that recognized different epitopes on the opposite sides of a PSA molecule. Five unique clones were found from the 5A10-PSA-panning [article IV]. Among these, the antibodies able to bind to the target in the presence of both 5E4 and 5A10 were exclusively from the ScFvP-repertoire (**Figure 16**). In contrast, the only clones from the 5E4-PSA-panning with any binding ability towards PSA complexed with 5A10 were from the ScFvM-repertoire. Consequently, selected anti-PSA antibodies from ScFvM- and ScFvP-library were able to recognize different epitopes on PSA. Another indication of complementary diversity was observed in the IC50 assays of anti-DIG clones as two ScFvP anti-DIG clones were able to bind the biotinylated digoxigenin, but did not respond to the free digoxigenin [article IV]. Similar clones were not captured from the ScFvM repertoire.

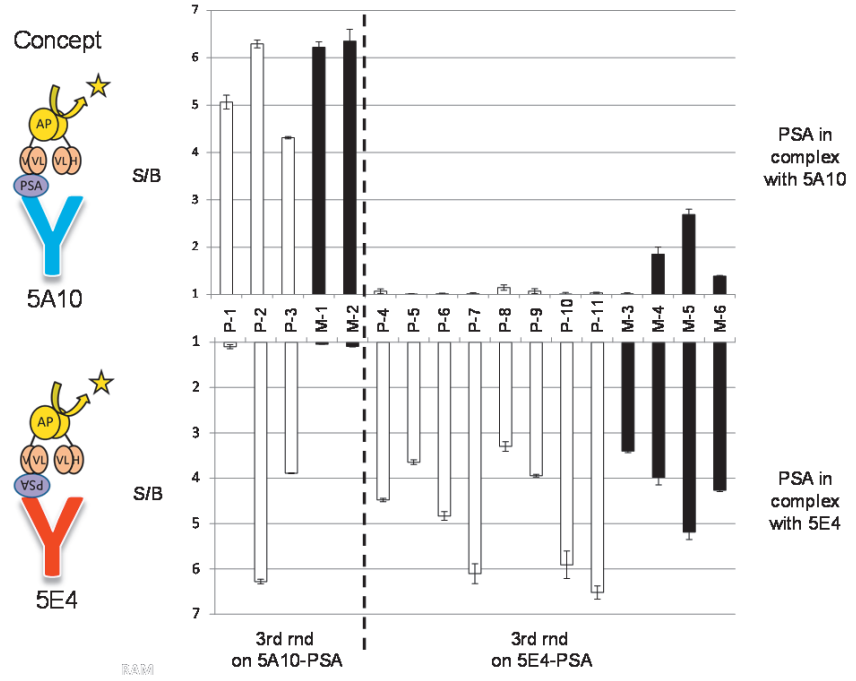

Figure 16. Binding profiles of the extracted anti-PSA antibodies. The ability of the ScFvM (black bars) and ScFvP (white bars) clones to bind PSA as captured on solid-phase either with antibody 5A10 (bars up) or 5E4 (bars down). The origin of the clones, i.e. the selection sample from which they were derived, is indicated below the graph. The error bars represent the SD of the intra-assay variation with three replicas. Adapted from article IV.

In this study, it was demonstrated that by diversifying different residues even the same framework can be modified to provide two libraries with distinct binding profiles. In some studies reported earlier, the likelihood of finding affinity reagents for peptides and haptens was increased by constructing libraries on an antibody scaffold known to bind a member of the target class [326; 327]. In the present strategy, the fundamental principles governing the molecular recognition of different sized antigens was explored in a more disclosed fashion as the starting scaffold was a germ-line gene pair and not an antigen-specific scaffold. Therefore, the principles of molecular recognition of different-sized antigens can be deciphered from the randomization scheme. Unfortunately, the single representative antigens of different classes does not form a panel versatile enough to make conclusions, for example, on the correlations between CDR-H3 loop length and the presence of salt bridge in relation to target size. Experiments with more antigens of various size and chemical nature are required to further clarify the underlying principles with relevant statistical rigor.

5.5 Selection performance of an antibody library and the chosen phage display format

5.5.1 Genetic fusion to capsid proteins p3, p3Δ and p9 support Fab display but with different efficiencies

Despite the popularity of phage display, no library-scale studies have been conducted to compare the selection performance of a library fused to different capsid proteins for display. In this study, an identical antibody repertoire was displayed as Fab-p9, -p3, -p3Δ and ScFv-p3Δ fusions on the filamentous phage (**Figure 17**) and, thereafter, the selection performance was followed in repeated parallel panning reactions against streptavidin and digoxigenin. The effect of valency modulation on the enrichment and diversity of binding phages was studied by superinfecting the Fab-p3 library with hyperphage in parallel with the conventional VCS-M13 superinfection.

We were able to display the Fab library as a fusion to all tested capsid proteins. There was, however, a decrease in the display efficiency of the Fab molecules monitored as Fab/phage-ratio with a sequence: hyperphage-packed Fab-p3 > VCS-M13-packed Fab-p3Δ > VCS-M13-packed Fab-p3 > VCS-M13-packed Fab-p9 (**Figure 18**). The display efficiency of ScFv could not be determined in the same assay as it does not bind to RAM-plates, but in an earlier study carried out with protein L, the ScFv display efficiency was higher as p3Δ than as p9 fusion [119], which is in accordance with the current study. The poor display of Fab-p9 (3.2% of the Fab/phage-ratio obtained for hyperphage-packed Fab-p3) was reflected as markedly slower enrichment or even as a failed experiment in the applied selection conditions [article V]. More than three rounds are, therefore, recommended with Fab-p9 or similar constructs to obtain a satisfying frequency of binding clones. Nevertheless, display via p9 can be considered as a true monovalent display technique, which may be a beneficial feature in the most stringent affinity selections. However, the high number of bald phages must be taken into account, when deciding the number of input phage particles.

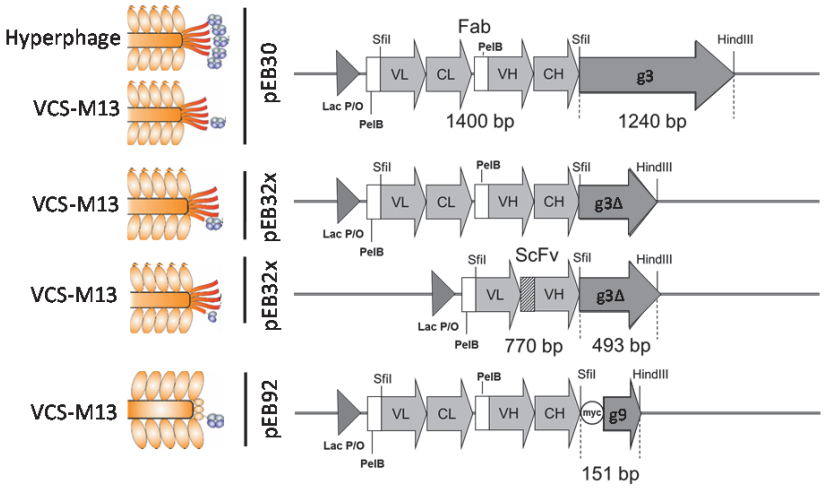

Figure 17. Phage display formats and genetic constructions. Antibody fragments were displayed as N-terminal fusions to the C-terminal phage capsid protein. The capsid genes used for display were the full-length gene-3 (g3), the truncated gene-3 (g3Δ) lacking the infection domains and the gene-9 (g9). Lac operon (Lac P/O) was used to express the constructs and PelB leader peptides were used for periplasmic translocation via the SEC-pathway.

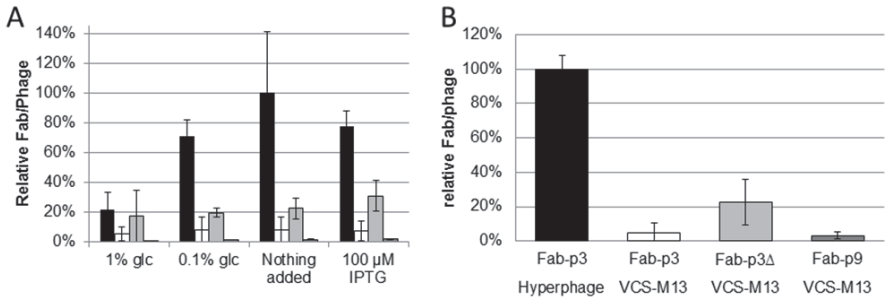

Figure 18. The influence of the phagemid and helper phage on the display level as normalized A) by total phage mass and B) by infectivity. Black: hyperphage Fab-p3; white: VCS-M13 Fab-p3; light grey: VCS-M13 Fab-p3Δ; and dark grey: VCS-M13 Fab-p9. The error bars represent the SD of three independent experiments in figure A and the SD of the intra-assay variation of three replicas in figure B.

5.5.2 Display efficiency and proliferation bias influence the clonal diversity of the retrieved clones

Significant loss of functional diversity was observed both after hyperphage and VCS-M13 packing in the Fab-p3 repertoire (**Table 11**). An increase in the number of frameshift mutants was also evident in ScFv-p3Δ repertoire, but to a lesser extent. Accordingly, a lower number of unique specific Fab clones were obtained from Fab-p3 than from Fab-p3Δ libraries, especially as regards digoxigenin selections (**Table 12**). The diversity of the Fab-p3 repertoire was better retained by oligovalent than monovalent display. However in this study, the use of hyperphage did not result in the emergence of a clone that would have superior characteristics compared to the clones obtained by other display formats. The advantage of oligovalent display is even more questionable, as the highest number of unique specific clones was obtained from p3Δ display libraries. In particular, the highest diversity and affinity antibodies against digoxigenin were retrieved from ScFv-p3Δ repertoire with IC50 values of 112 ± 24 nM and 158 ± 20 nM (**Figure 19**).

A fast enrichment is not always a guarantee for quality. Especially in the monovalent Fab-p3 library, a single anti-DIG monoclone overtook the output. The same clone was observed to be the highest affinity anti-digoxigenin antibody in all screened Fab-p3 outputs and also found in the first set of Fab-p3Δ and Fab-p9 selections. There are alternative explanations for this phenomenon: a low initial diversity, a true challenge (no better clones available), the strong secondary fitness characteristics of the clone, or cross-contamination of the panning reactions. For the most part, the last alternative was excluded by a restriction analysis of the phagemid DNA after selection. Naturally, the phagemid leakage between Fab-p3 libraries (VCS-M13- and hyperphage-infected) is indiscernible and a minor contamination in other libraries is not necessarily visible on stained agarose gels. Consequently, contamination cannot be completely ruled out.

As DIG is a small hapten, the carrier protein streptavidin is exposed to the phages and available target in selection reactions. Therefore, a DTT elution strategy was devised, which in principle should selectively rescue only the digoxigenin-bound phages by cleaving the linker between DIG and STR. Despite the DTT elution, an enrichment of anti-STR clones was observed in DIG selections. This was most probably due to the presence of clones with various affinities against STR and the law of mass action. During the elution (30 min) also STR-binders were dissociated from the carrier and included to continuation. This could be avoided by alternating the carrier between STR and avidin, after which no enrichment of anti-STR phage was observed [articles IV and V].

Although ScFv seems to be superior to Fab in display, the conversion of ScFv into Fab or directly to full-length IgG may bring unpleasant surprises. Some groups have observed a decrease in affinity upon conversion [375], although attention must be paid to possible change in valency as ScFv may form dimers or

higher multimer structures [376], whereas Fab is generally considered to be a monomeric molecule. Possibly due to these issues or easier format-conversion of Fab to full-length IgG, many recent large antibody repertoires have been constructed in Fab format [157; 158]. It is of crucial importance to consider the display related aspects as the chosen format may bias the outcome. Furthermore, as pointed out by the carrier dilemma, selection conditions should be considered with prudence to avoid wrong conclusions on the library potential based on a weak experimental execution.

Table 11. The size, quality, and phage titer of the constructed libraries.

Phagemid	Fusion protein	Trans-formed	Super-infection	in frame	Phage yield[a] I, cfu	Phage yield[a] II, cfu
pAK400-Fab	LC-BLA	2x 10^7	None	8/8	-	-
pEB07-Fab	Fd-BLA	3x 10^8	None	8/8	-	-
pEB30-Fab	Fd-p3	3x 10^9	None	22/24	-	-
pEB30-Fab	Fd-p3	-"-	Hyper-phage	7/23	1.8 ± 0.3 x10^{12}	2.9 ± 1.0 x10^{11}
pEB30-Fab	Fd-p3	-"-	VCS-M13	8/22	7.1 ± 0.5 x10^{13}	3.6 ± 0.8 x10^{12}
pEB32x-Fab	Fd-p3Δ	1x 10^9	VCS-M13	22/24	1.5 ± 0.6 x10^{14}	1.1 ± 0.1 x10^{14}
pEB92-Fab	Fd-p9	5x 10^9	VCS-M13	22/23	3.3 ± 0.2 x10^{14}	2.9 ± 0.8 x10^{14}
pEB32x-scFv	Fd-p3Δ	6x 10^9	VCS-M13	21/34	3.2 ± 0.3 x10^{14}	3.1 ± 7.4 x10^{14}

Table 12. Sequenced anti-DIG clone diversity and highest observed affinities.

		Digoxigenin STR>STR>AVI			Digoxigenin STR>AVI>STR		
Library	Helper phage	N	D	Highest aff. (IC50)	N	D	Highest aff. (IC50)
Fab-p3	Hyperphage	8	1	442 ± 37 nM	8	4	442 ± 37 nM
Fab-p3	VCS-M13	8	1	442 ± 37 nM	8	2	442 ± 37 nM
Fab-p3	Hyper>VCS	8	1	442 ± 37 nM	8	1	442 ± 37 nM
Fab-p3Δ	VCS-M13	7	4	442 ± 37 nM	7	5	1.5 ± 0.1 µM
ScFv-p3Δ	VCS-M13	8	4	158 ± 20 nM	8	6	112 ± 24 nM
Fab-p9	VCS-M13	3	2	442 ± 37 nM	0	0	N.D.

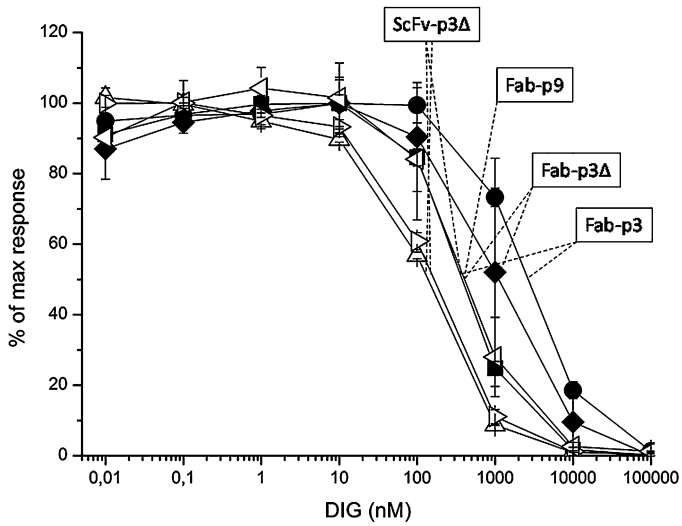

Figure 19. IC50 assay of the three highest affinity anti-DIG antibodies from Fab and ScFv repertoires. The clone marked with a filled square found in Fab-p3, Fab-p3Δ and Fab-p9. The clone marked with a filled diamond found in Fab-p3Δ. The clone marked with a filled circle found in Fab-p3. Open triangles: ScFv-p3Δ.

6 CONCLUSIONS

The nature of directed evolution practiced by humankind has expanded from the traditional breeding concept to a detailed molecular level fine-tuning of single genes and metabolic pathways by designed gene libraries and elaborate selection techniques. Despite the continuously accumulating knowledge of biomolecules, rational design is still totally dependent on directed evolution to finalize design solutions [357]. Therefore, novel library concepts that focus on the most relevant diversity, potentiating tools that encompass larger clonal repertoires, and an understanding of the factors affecting selections are vital information in bringing modern biotechnology forward.

In this thesis, new solutions and insights have been provided in multiple focus areas of laboratory-driven directed evolution. The first two research articles (I and II) described new techniques to improve gene variant library construction. In the first study, a method, termed selective RCA, was introduced. This tool increases the number of transformants and reduces the template background in the final library. In the second study, the efficiency of the simplest random mutagenesis method, termed error-prone RCA, was further enhanced with Cre/loxP-recombination. Especially sRCA has obtained attention in the biotechnology community as a welcome update to Kunkel mutagenesis published in 1985 [155]. Surprisingly, the sRCA study was cited for the first time on account of the finding that phi29 DNA polymerase is capable of using uridylated DNA as template [377]. This is a demonstration among others that applied science may also contribute to basic science.

In the third study (III), a universal method for improving the display of immunoglobulins on phage was described and the role of the found mutations on the display properties discussed. Antibody fragments are foreign molecules to *E. coli* that may severely disturb the host metabolism as demonstrated with the growth experiments of Fab expressing cells in article III. Against this background, it is quite evident that knowledge on the functional aspects of the chosen platform will ultimately lead to a more cost-effective product development. Unfortunately, there are only few research articles on comparing the properties of the well-established display platforms to support researchers. This lack of information was the incentive for the article V. In this study, the selection performance of an identical antibody repertoire as a fusion to different filamentous phage capsid proteins was followed. It was observed that Fab display level on phage was lower as a fusion to p9 than to p3, which in turn, lead to the less efficient enrichment of the binding clones in the p9 repertoire. As a conclusion, in order to maximize the diversity of the output clones in primary library selections, truncated p3 display is a vital alternative, whereas p9 display may be an option for affinity maturation schemes, in which lower valency of displayed molecules is an advantage.

In the fourth research article (IV) a synthetic biology approach was undertaken to design recombinant antibody library repertoires in ScFv format. It was

demonstrated that the reactivity profile of a library could be shifted by targeting different positions in the framework and that both libraries (ScFvP and ScFvM) had their own strengths as the source of antibodies depending on the nature of the target. As exemplified by this study, limiting the searched structural sequence space by the rational design of library diversity, the probability of obtaining more specific antibodies against the desired group of antigens is increased. Conversely, knowledge obtained from the rationally designed focused libraries may also provide new insights into the mechanisms of molecular recognition in general, which is the basis of more efficient synthetic repertoires in the future.

The results presented here contribute to the sciences of experimental directed evolution and synthetic biology that are under continuous development by academic groups and private companies. The grand visions in directed evolution are to harness novel enzymes to produce sustainable energy for all and to deliver drugs to cure diseases that were once considered incurable. While we are not yet there, these kind of initiatives are already on the way [378], progressing with an accelerated speed of evolution.

7 ACKNOWLEDGEMENTS

This study was carried out at the Department of Biochemistry in the Division of Biotechnology at the University of Turku during the years 2007 – 2013. Financial support from the National Technology Agency of Finland (Tekes) and the graduate school of Advanced Diagnostic Technologies and Applications (DIA-NET) is gratefully acknowledged. I wish to thank Professors Timo Lövgren, Kim Pettersson and Tero Soukka for providing excellent research facilities. I am a very fortunate person as I had the opportunity to work under your consistent and just leadership that has led to the relaxed and inspiring atmosphere that we enjoy at the Division of Biotechnology.

I owe my warmest thanks of this thesis to my first supervisor Dr Urpo Lamminmäki. You have given me freedom and support in the optimal formulation during the years, which has resulted in a firm feeling of my researcher identity and even in excessive confidence in my abilities, someone could say. I want to also thank you, Urpo, for your friendship. I have rarely met any as kind person as you. You are endowed with exquisite social skills and grand innovative visions that I have tried to bring to life by giving my best. Thanks to our friendship, we have had active and open knowledge sharing, which I have found very rewarding for the joined efforts.

I wish to thank my second supervisor Dr Petri Saviranta for helpful discussions, analytical thinking and support in my post-graduate studies. Without your earlier contributions to the knowledge base of the antibody engineering group, we would probably not be as far as we are now in our endeavours. I would also like to thank Dr. Jan Weegar for your support and encouragement in the thesis work. I have had very joyful time in sharing laboratory activities and speaking Swedish with you. Tack!

I am grateful to the reviewers, Professors Kalle Saksela and Mauno Vihinen, for your detailed comments and suggestions that have been very valuable in revising the thesis.

Special thanks to the antibody engineering group members: Dr Eeva-Christine Brockmann, Hanna Sanmark, Maria Lahti, Janne Leivo, Sultana Akter, Susan Peréz Gamarra and Markus Vehniäinen as well as to the heart library group members Dr Janita Lövgren, Priyanka Negi, Päivi Malmi and Katja Marttunen. You are great colleagues and I truly have enjoyed both scientific and nonscientific discussions with you. I want to especially acknowledge Eeva for sharing the materials and knowledge in recombinant antibody work during the

whole extent of my studies and Hanna for walking the plank with me from the very beginning to the final plunge into the recombinant Fab and ScFv Ocean.

I wish to thank all co-authors of the articles, of which the ones, not mentioned earlier are: Jani Ylä-Pelto, Yuan Liu, Maria Julin, Markku Syrjänpää, Professor Alex Azhayev, Dr Qi Wang, Titta Seppä and Md. Liton Ferdhos Khan. Of these co-authors, Alex and Qi have given vital contributions to the research materials, whereas the others have performed laboratory work under my supervision. Thank you, Qi, for giving your time in antigen conjugate synthesis and teaching me the basics of conjugate analytics. Thank you, Markku, for your efforts in elucidating the synthetic antibody library performance during your Master thesis and giving me the privilege to guide you. I am also grateful to Katja Marttunen and Liton Ferdhos Khan for trusting your guidance in my hands during the practical work of your Master thesis studies.

I want to acknowledge other DIA-NET graduate students, of which, especially discussions with the Biotechnology-linked peers from the early recruitment (Ari Lehmusvuori, Tanja Savukoski and Minna Ylihärsilä) have been refreshing. Thank you, Ari, for the relaxing discussions in between thesis writing and sharing experience in scholarly matters and coffee flavours. I would like to thank also all other staff at the Division of Biotechnology for providing an enjoyable working environment. Especially Mirja Jaala, Martti Sointusalo, Marja Maula and Sanna Koivuniemi are thanked for excellent care of their duties in the office and laboratory maintenance, not to forget Pirjo Laaksonen, Pirjo Pietilä and Dr Mari Peltola for their helpful advises in europium-labeling, protein production and statistics, respectively. I want to acknowledge Dr Mari Peltola for providing PSA reagents, too, and Dr Jussi Meriluoto for materials and knowledge support for the microcystin research.

I sometimes recollect the active times of the Dengue group at Biotechnology, namely the focused pursuit of immunogenic viral epitopes by Dr Gaurav Batra, Erica Kuusela and Julius Manninen. I especially miss those moments with Dr Gaurav Batra, when we were keeping tabletop centrifuges busy and pipettes in motion in the lab at late hours and debating on science and philosophy. Thank you, Gaurav, for the great time together! I have a lot to learn from your politeness, openness, networking abilities, ambition and commitment.

I am very grateful to my dear friends TTTK for giving delightful insights into life and sharing your lives with me and my family. T´s stand for Timo, Teemu and Teijo, and K for Kalle. Thank you, Timo, for your friendship that has lasted intensive through the years. Thank you, Teemu, for sharing the nature

observations and mosquitos at Kyyjärvi cottage. Thank you, Teijo, for the sailing trips and joined sampling efforts to find the most delicious beer in the world. Thank you, Kalle, for guidance in scientific writing and professional fish cooking. And thanks go also to the spouses of TTTKs, namely Saara, Katja L., Anna and Katja R. for lending your husbands sometimes and for the enjoyable shared moments with whole families.

Thank you, mother and father, Eeva and Risto, for raising me up with the support of my grandmothers, especially Anna-Liisa. You have taught me resilience, creativity, and execution by your own examples, all of which are qualities that have been indeed needed in this task and in life in general. I greatly appreciate your help, dear mother, in nursing the children and baking the best rye bread in the world. I am also in debt to my mother-in-law, Marja-Leena, for her willingness and flexibility to take care of the children, when needed. I wish to thank also my brother's family, Janne, Anne, Aino and Ella Puustelli, my sister's family, Katri, Mikko, Okko, Eevi and Vappu, and my brother-in-law´s family, Martti, Maarit, Jere, Luka and Ellen, for the endless number of enjoyful moments with you. Dr Anne Puustelli is also acknowledged for the professional help in statistics.

I am extremely happy for my two beautiful little diamonds, Taimi and Maikki. Your sparkling eyes cheer me up every day. Dad loves you! Special thanks go to Taimi also for the cover page artwork. Finally, but not least, my beloved wife Katri, there are no words to describe my gratitude to your support during the years. I know that it is not easy to be married to a scientist. Although he would be physically present, he is still absent. And when he should be physically there, he is on the way, as the time schedule he said, was too optimistic and did not hold due to a thousand little excuses. Thank you, sweetheart, for bringing love to my life. I could not imagine life without you. Te amo ab imo pectore.

Turku, November 2013.

Tuomas Huovinen

8 REFERENCES

[1] M.D. Purugganan, and D.Q. Fuller, The nature of selection during plant domestication. Nature 457 (2009) 843-8.

[2] D.Q. Fuller, Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World. Ann Bot 100 (2007) 903-24.

[3] B. Leader, Q.J. Baca, and D.E. Golan, Protein therapeutics: a summary and pharmacological classification. Nat Rev Drug Discov 7 (2008) 21-39.

[4] B.M. Lewin, Genes VIII. Pearson Prentice Hall, Upper Saddle River, N.J., 2004.

[5] J.E. Barrick, D.S. Yu, S.H. Yoon, H. Jeong, T.K. Oh, D. Schneider, R.E. Lenski, and J.F. Kim, Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature 461 (2009) 1243-7.

[6] O. Lukjancenko, T.M. Wassenaar, and D.W. Ussery, Comparison of 61 sequenced Escherichia coli genomes. Microb Ecol 60 (2010) 708-20.

[7] S. Duffy, L.A. Shackelton, and E.C. Holmes, Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet 9 (2008) 267-76.

[8] S.K. Bhatnagar, and M.J. Bessman, Studies on the mutator gene, mutT of Escherichia coli. Molecular cloning of the gene, purification of the gene product, and identification of a novel nucleoside triphosphatase. J Biol Chem 263 (1988) 8953-7.

[9] H. Maki, and M. Sekiguchi, MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. Nature 355 (1992) 273-5.

[10] E.C. Cox, and C. Yanofsky, Altered base ratios in the DNA of an Escherichia coli mutator strain. Proc Natl Acad Sci U S A 58 (1967) 1895-902.

[11] C. Yanofsky, E.C. Cox, and V. Horn, The unusual mutagenic specificity of an E. Coli mutator gene. Proc Natl Acad Sci U S A 55 (1966) 274-81.

[12] A.P. Grollman, and M. Moriya, Mutagenesis by 8-oxoguanine: an enemy within. Trends Genet 9 (1993) 246-9.

[13] B.W. Glickman, and M. Radman, Escherichia coli mutator mutants deficient in methylation-instructed DNA mismatch correction. Proc Natl Acad Sci U S A 77 (1980) 1063-7.

[14] R. Scheuermann, S. Tam, P.M. Burgers, C. Lu, and H. Echols, Identification of the epsilon-subunit of Escherichia coli DNA polymerase III holoenzyme as the dnaQ gene product: a fidelity subunit for DNA replication. Proc Natl Acad Sci U S A 80 (1983) 7085-9.

[15] Agilent Technologies, XL1-Red Competent Cells - Instruction Manual. Agilent Technologies, 2010.

[16] A. Greener, M. Callahan, and B. Jerpseth, An efficient random mutagenesis technique using an E. coli mutator strain. Mol Biotechnol 7 (1997) 189-95.

[17] S. Lin-Chao, W.T. Chen, and T.T. Wong, High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II. Mol Microbiol 6 (1992) 3385-93.

[18] N.M. Low, P.H. Holliger, and G. Winter, Mimicking somatic hypermutation: affinity maturation of antibodies displayed on bacteriophage using a bacterial mutator strain. J Mol Biol 260 (1996) 359-68.
[19] O. Selifonova, F. Valle, and V. Schellenberger, Rapid evolution of novel traits in microorganisms. Appl Environ Microbiol 67 (2001) 3645-9.
[20] U.M. Tillich, S. Lehmann, K. Schulze, U. Dühring, and M. Frohme, The optimal mutagen dosage to induce point-mutations in Synechocystis sp. PCC6803 and its application to promote temperature tolerance. PLoS One 7 (2012) e49467.
[21] M. Wu, C. Tang, J. Li, H. Zhang, and J. Guo, Bimutation breeding of Aspergillus niger strain for enhancing β-mannanase production by solid-state fermentation. Carbohydr Res 346 (2011) 2149-55.
[22] H. Schuster, The method of reaction of desoxyribonucleic acid with nitrous acid. Z Naturforsch B 15B (1960) 298-304.
[23] R. De Giovanni-Donnelly, The mutagenicity of sodium bisulfite on base-substitution strains of Salmonella typhimurium. Teratog Carcinog Mutagen 5 (1985) 195-203.
[24] S. Ermakova-Gerdes, Z. Yu, and W. Vermaas, Targeted random mutagenesis to identify functionally important residues in the D2 protein of photosystem II in Synechocystis sp. strain PCC 6803. J Bacteriol 183 (2001) 145-54.
[25] J.T. Kadonaga, and J.R. Knowles, A simple and efficient method for chemical mutagenesis of DNA. Nucleic Acids Res 13 (1985) 1733-45.
[26] A.E. Stellwagen, and N.L. Craig, Gain-of-function mutations in TnsC, an ATP-dependent transposition protein that activates the bacterial transposon Tn7. Genetics 145 (1997) 573-85.
[27] B.F. Cordero, I. Obraztsova, I. Couso, R. Leon, M.A. Vargas, and H. Rodriguez, Enhancement of lutein production in Chlorella sorokiniana (Chorophyta) by improvement of culture conditions and random mutagenesis. Mar Drugs 9 (2011) 1607-24.
[28] J. Stonesifer, and R.H. Baltz, Mutagenic DNA repair in Streptomyces. Proc Natl Acad Sci U S A 82 (1985) 1180-3.
[29] C.W. Shearman, and L.A. Loeb, On the fidelity of DNA replication. Specificity of nucleotide substitution by intercalating agents. J Biol Chem 258 (1983) 4477-84.
[30] R.F. Newbold, W. Warren, A.S. Medcalf, and J. Amos, Mutagenicity of carcinogenic methylating agents is associated with a specific DNA modification. Nature 283 (1980) 596-9.
[31] B. Singer, S.J. Spengler, H. Fraenkel-Conrat, and J.T. Kuśmierek, O4-Methyl, -ethyl, or -isopropyl substituents on thymidine in poly(dA-dT) all lead to transitions upon replication. Proc Natl Acad Sci U S A 83 (1986) 28-32.
[32] B.D. Preston, B. Singer, and L.A. Loeb, Mutagenic potential of O4-methylthymine in vivo determined by an enzymatic approach to site-specific mutagenesis. Proc Natl Acad Sci U S A 83 (1986) 8501-5.

[33] A.R. Peterson, and H. Peterson, Differences in temporal aspects of mutagenesis and cytotoxicity in Chinese hamster cells treated with methylating agents and thymidine. Proc Natl Acad Sci U S A 79 (1982) 1643-7.
[34] B. Singer, Mutagenesis from a chemical perspective: nucleic acid reactions, repair, translation, and transcription. Basic Life Sci 20 (1982) 1-42.
[35] P.L. Foster, In vivo mutagenesis. Methods Enzymol 204 (1991) 114-25.
[36] E.I. Budowsky, and T.I. Postnova, Mechanism of the mutagenic action of hydroxylamine. X. Certain specificities in the mutagenesis of N-hydroxy and N-methoxy analogs of cytosine and adenine derivatives. Mutat Res 37 (1976) 11-18.
[37] R. Stolarski, B. Kierdaszuk, C.E. Hagberg, and D. Shugar, Hydroxylamine and methoxyamine mutagenesis: displacement of the tautomeric equilibrium of the promutagen N6-methoxyadenosine by complementary base pairing. Biochemistry 23 (1984) 2906-13.
[38] R. Stolarski, B. Kierdaszuk, C.E. Hagberg, and D. Shugar, Mechanism of hydroxylamine mutagenesis: tautomeric shifts and proton exchange between the promutagen N6-methoxyadenosine and cytidine. Biochemistry 26 (1987) 4332-7.
[39] W. Peng, and B.R. Shaw, Accelerated deamination of cytosine residues in UV-induced cyclobutane pyrimidine dimers leads to CC-->TT transitions. Biochemistry 35 (1996) 10172-81.
[40] Y. Barak, O. Cohen-Fix, and Z. Livneh, Deamination of cytosine-containing pyrimidine photodimers in UV-irradiated DNA. Significance for UV light mutagenesis. J Biol Chem 270 (1995) 24174-9.
[41] F. Kaudewitz, Production of bacterial mutants with nitrous acid. Nature 183 (1959) 1829-30.
[42] R. Shapiro, V. DiFate, and M. Welcher, Deamination of cytosine derivatives by bisulfite. Mechanism of the reaction. J Am Chem Soc 96 (1974) 906-12.
[43] W.M. Rideout, G.A. Coetzee, A.F. Olumi, and P.A. Jones, 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. Science 249 (1990) 1288-90.
[44] R.Y. Wang, C.W. Gehrke, and M. Ehrlich, Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. Nucleic Acids Res 8 (1980) 4777-90.
[45] S.J. Clark, J. Harrison, C.L. Paul, and M. Frommer, High sensitivity mapping of methylated cytosines. Nucleic Acids Res 22 (1994) 2990-7.
[46] W. Vielmetter, and H. Schuster, Base specificity in the induction of mutations by nitrous acid in phage T-2. Z Naturforsch B 15B (1960) 304-11.
[47] J.H. Miller, Mutagenic specificity of ultraviolet light. J Mol Biol 182 (1985) 45-65.
[48] R.B. Setlow, and W.L. Carrier, Pyrimidine dimers in ultraviolet-irradiated DNA's. J Mol Biol 17 (1966) 237-54.

[49] I. Tessman, S.K. Liu, and M.A. Kennedy, Mechanism of SOS mutagenesis of UV-irradiated DNA: mostly error-free processing of deaminated cytosine. Proc Natl Acad Sci U S A 89 (1992) 1159-63.

[50] S. Shibutani, M. Takeshita, and A.P. Grollman, Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. Nature 349 (1991) 431-4.

[51] T. Douki, A. Reynaud-Angelin, J. Cadet, and E. Sage, Bipyrimidine photoproducts rather than oxidative lesions are the main type of DNA damage involved in the genotoxic effect of solar UVA radiation. Biochemistry 42 (2003) 9221-6.

[52] O. Sidorkina, M. Saparbaev, and J. Laval, Effects of nitrous acid treatment on the survival and mutagenesis of Escherichia coli cells lacking base excision repair (hypoxanthine-DNA glycosylase-ALK A protein) and/or nucleotide excision repair. Mutagenesis 12 (1997) 23-8.

[53] S. Mourya, and K.S. Jauhri, Production of citric acid from starch-hydrolysate by Aspergillus niger. Microbiol Res 155 (2000) 37-44.

[54] I. Mariam, S.A. Nagra, I. Haq, and S. Ali, Application of 2-factorial design on the enhanced production of calcium gluconate by a mutant strain of Aspergillus niger. Bioresour Technol 101 (2010) 4075-80.

[55] F. Xu, J. Wang, S. Chen, W. Qin, Z. Yu, H. Zhao, X. Xing, and H. Li, Strain improvement for enhanced production of cellulase in Trichoderma viride. Prikl Biokhim Mikrobiol 47 (2011) 61-5.

[56] Y.X. Zhang, K. Perry, V.A. Vinci, K. Powell, W.P. Stemmer, and S.B. del Cardayré, Genome shuffling leads to rapid phenotypic improvement in bacteria. Nature 415 (2002) 644-6.

[57] E. Levine, and T. Thiel, UV-inducible DNA repair in the cyanobacteria Anabaena spp. J Bacteriol 169 (1987) 3988-93.

[58] U. Mohan, and U.C. Banerjee, Molecular evolution of a defined DNA sequence with accumulation of mutations in a single round by a dual approach to random chemical mutagenesis (DuARCheM). Chembiochem 9 (2008) 2238-43.

[59] P. Liu, Y. Hong, Y. Lin, X. Fu, L. Lin, C. Li, X. Meng, and Z. Liu, A frequency-controlled random mutagenesis method for GC-rich genes. Anal Biochem 388 (2009) 356-8.

[60] K.S. Lundberg, D.D. Shoemaker, M.W. Adams, J.M. Short, J.A. Sorge, and E.J. Mathur, High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. Gene 108 (1991) 1-6.

[61] R.A. Beckman, A.S. Mildvan, and L.A. Loeb, On the fidelity of DNA replication: manganese mutagenesis in vitro. Biochemistry 24 (1985) 5810-7.

[62] J.P. Vartanian, M. Henry, and S. Wain-Hobson, Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. Nucleic Acids Res 24 (1996) 2627-31.

[63] R.C. Cadwell, and G.F. Joyce, Randomization of genes by PCR mutagenesis. PCR Methods Appl 2 (1992) 28-33.

[64] N.K. Sinha, and M.D. Haimes, Molecular mechanisms of substitution mutagenesis. An experimental test of the Watson-Crick and topal-fresco models of base mispairings. J Biol Chem 256 (1981) 10671-83.
[65] D. Leung, E. Chen, and G. DV, A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. Technique 1 (1989) 11-15.
[66] K.R. Tindall, and T.A. Kunkel, Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. Biochemistry 27 (1988) 6008-13.
[67] T.A. Kunkel, The mutational specificity of DNA polymerase-beta during in vitro DNA synthesis. Production of frameshift, base substitution, and deletion mutations. J Biol Chem 260 (1985) 5787-96.
[68] J.L. Lin-Goerke, D.J. Robbins, and J.D. Burczak, PCR-based random mutagenesis using manganese and reduced dNTP concentration. Biotechniques 23 (1997) 409-12.
[69] M.M. Huang, N. Arnheim, and M.F. Goodman, Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. Nucleic Acids Res 20 (1992) 4567-73.
[70] L. Wan, M.B. Twitchett, L.D. Eltis, A.G. Mauk, and M. Smith, In vitro evolution of horse heart myoglobin to increase peroxidase activity. Proc Natl Acad Sci U S A 95 (1998) 12825-31.
[71] K.A. Eckert, and T.A. Kunkel, DNA polymerase fidelity and the polymerase chain reaction. PCR Methods Appl 1 (1991) 17-24.
[72] S. Shafikhani, R.A. Siegel, E. Ferrari, and V. Schellenberger, Generation of large libraries of random mutants in Bacillus subtilis by PCR-based plasmid multimerization. Biotechniques 23 (1997) 304-10.
[73] J. Cline, and H. Hogrefe, Randomize Gene Sequences with New PCR Mutagenesis Kit. Stratagene, 2002.
[74] H. Hogrefe, J. Cline, G.L. Youngblood, and R.M. Allen, Creating randomized amino acid libraries with the QuikChange Multi Site-Directed Mutagenesis Kit. Biotechniques 33 (2002) 1158-60, 1162, 1164-5.
[75] J. Cline, J.C. Braman, and H.H. Hogrefe, PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Res 24 (1996) 3546-51.
[76] H. Hogrefe, and J. Cline, Compositions and methods for random nucleic acid molecules. in: T.U.S.P.a.T. Office, (Ed.), US20030152944, Stratagene, United States of America, 2003, pp. 25.
[77] H. Hogrefe, Polymerase enhancing factor (PEF) extracts PEF protein complexes isolated PEF proteins and methods for purifying and identifying same. in: T.U.S.P.a.T. Office, (Ed.), US6183997, Stratagene, United States of America, 2001, pp. 75.
[78] A.Y. Sakaguchi, M. Sedlak, J.M. Harris, and M.F. Sarosdy, Cautionary note on the use of dUMP-containing PCR primers with Pfu and VentR DNA polymerases. Biotechniques 21 (1996) 368-70.
[79] R.S. Lasken, D.M. Schuster, and A. Rashtchian, Archaebacterial DNA polymerases tightly bind uracil-containing DNA. J Biol Chem 271 (1996) 17692-6.

[80] B.D. Biles, and B.A. Connolly, Low-fidelity Pyrococcus furiosus DNA polymerase mutants useful in error-prone PCR. Nucleic Acids Res 32 (2004) e176.

[81] P.H. Patel, H. Kawate, E. Adman, M. Ashbach, and L.A. Loeb, A single highly mutable catalytic site amino acid is critical for DNA polymerase fidelity. J Biol Chem 276 (2001) 5044-51.

[82] M. Zaccolo, D.M. Williams, D.M. Brown, and E. Gherardi, An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. J Mol Biol 255 (1996) 589-603.

[83] P.K. Lin, and D.M. Brown, Synthesis and duplex stability of oligonucleotides containing cytosine-thymine analogues. Nucleic Acids Res 17 (1989) 10373-83.

[84] J.H. Spee, W.M. de Vos, and O.P. Kuipers, Efficient random mutagenesis method with adjustable mutation frequency by use of PCR and dITP. Nucleic Acids Res 21 (1993) 777-8.

[85] K. Negishi, M. Takahashi, Y. Yamashita, M. Nishizawa, and H. Hayatsu, Mutagenesis by N4-aminocytidine: induction of AT to GC transition and its molecular mechanism. Biochemistry 24 (1985) 7273-8.

[86] H. Kamiya, M. Ito, and H. Harashima, Induction of transition and transversion mutations during random mutagenesis PCR by the addition of 2-hydroxy-dATP. Biol Pharm Bull 27 (2004) 621-3.

[87] H. Kamiya, M. Ito, and H. Harashima, Induction of various mutations during PCRs with manganese and 8-hydroxy-dGTP. Biol Pharm Bull 30 (2007) 842-4.

[88] T.S. Wong, K.L. Tee, B. Hauer, and U. Schwaneberg, Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. Nucleic Acids Res 32 (2004) e26.

[89] H. Mundhada, J. Marienhagen, A. Scacioc, A. Schenk, D. Roccatano, and U. Schwaneberg, SeSaM-Tv-II generates a protein sequence space that is unobtainable by epPCR. Chembiochem 12 (2011) 1595-601.

[90] Y. Matsumoto, and K. Kim, Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. Science 269 (1995) 699-702.

[91] W.P. Osheroff, H.K. Jung, W.A. Beard, S.H. Wilson, and T.A. Kunkel, The fidelity of DNA polymerase beta during distributive and processive DNA synthesis. J Biol Chem 274 (1999) 3642-50.

[92] P. Mondon, N. Souyris, L. Douchy, F. Crozet, K. Bouayadi, and H. Kharrat, Method for generation of human hyperdiversified antibody fragment library. Biotechnol J 2 (2007) 76-82.

[93] S. Emond, P. Mondon, S. Pizzut-Serin, L. Douchy, F. Crozet, K. Bouayadi, H. Kharrat, G. Potocki-Véronèse, P. Monsan, and M. Remaud-Simeon, A novel random mutagenesis approach using human mutagenic DNA polymerases to generate enzyme variant libraries. Protein Eng Des Sel 21 (2008) 267-74.

[94] M.T. Washington, R.E. Johnson, S. Prakash, and L. Prakash, Fidelity and processivity of Saccharomyces cerevisiae DNA polymerase eta. J Biol Chem 274 (1999) 36835-8.

[95] T.A. Kunkel, DNA replication fidelity. J Biol Chem 279 (2004) 16895-8.

[96] R.E. Johnson, M.T. Washington, L. Haracska, S. Prakash, and L. Prakash, Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions. Nature 406 (2000) 1015-9.
[97] M. Camps, J. Naukkarinen, B.P. Johnson, and L.A. Loeb, Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. Proc Natl Acad Sci U S A 100 (2003) 9727-32.
[98] G. Kopsidas, R.K. Carman, E.L. Stutt, A. Raicevic, A.S. Roberts, M.A. Siomos, N. Dobric, L. Pontes-Braz, and G. Coia, RNA mutagenesis yields highly diverse mRNA libraries for in vitro protein evolution. BMC Biotechnol 7 (2007) 18.
[99] K.L. Petrie, and G.F. Joyce, Deep sequencing analysis of mutations resulting from the incorporation of dNTP analogs. Nucleic Acids Res 38 (2010) 8095-104.
[100] R. Fujii, M. Kitaoka, and K. Hayashi, One-step random mutagenesis by error-prone rolling circle amplification. Nucleic Acids Res 32 (2004) e145.
[101] S. Haapa, S. Taira, E. Heikkinen, and H. Savilahti, An efficient and accurate integration of mini-Mu transposons in vitro: a general methodology for functional genetic analysis and molecular biology applications. Nucleic Acids Res 27 (1999) 2777-84.
[102] A. Baldwin, K. Busse, A. Simm, and D. Jones, Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx). Nucleic Acids Res 36 (2008) e77.
[103] B.M. Hoeller, B. Reiter, S. Abad, I. Graze, and A. Glieder, Random tag insertions by Transposon Integration mediated Mutagenesis (TIM). J Microbiol Methods 75 (2008) 251-7.
[104] H. Murakami, T. Hohsaka, and M. Sisido, Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs. Nat Biotechnol 20 (2002) 76-81.
[105] P. Anderson, Mutagenesis. Methods Cell Biol 48 (1995) 31-58.
[106] M. Harper, and C.J. Lee, Genome-wide analysis of mutagenesis bias and context sensitivity of N-methyl-N'-nitro-N-nitrosoguanidine (NTG). Mutat Res 731 (2012) 64-7.
[107] S. Flibotte, M.L. Edgley, I. Chaudhry, J. Taylor, S.E. Neil, A. Rogula, R. Zapf, M. Hirst, Y. Butterfield, S.J. Jones, M.A. Marra, R.J. Barstead, and D.G. Moerman, Whole-genome profiling of mutagenesis in Caenorhabditis elegans. Genetics 185 (2010) 431-41.
[108] T.S. Rasila, M.I. Pajunen, and H. Savilahti, Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment. Anal Biochem 388 (2009) 71-80.
[109] T.S. Wong, D. Roccatano, M. Zacharias, and U. Schwaneberg, A statistical analysis of random mutagenesis methods used for directed protein evolution. J Mol Biol 355 (2006) 858-71.
[110] M.J. Volles, and P.T. Lansbury, A computer program for the estimation of protein and nucleic acid sequence diversity in random point mutagenesis libraries. Nucleic Acids Res 33 (2005) 3667-77.

[111] A. Knappik, L. Ge, A. Honegger, P. Pack, M. Fischer, G. Wellnhofer, A. Hoess, J. Wölle, A. Plückthun, and B. Virnekäs, Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. J Mol Biol 296 (2000) 57-86.

[112] C.A. Hutchison, S. Phillips, M.H. Edgell, S. Gillam, P. Jahnke, and M. Smith, Mutagenesis at a specific position in a DNA sequence. J Biol Chem 253 (1978) 6551-60.

[113] M.D. Matteucci, and H.L. Heyneker, Targeted random mutagenesis: the use of ambiguously synthesized oligonucleotides to mutagenize sequences immediately 5' of an ATG initiation codon. Nucleic Acids Res 11 (1983) 3113-21.

[114] K.A. McCormick, and B.D. Cain, Targeted mutagenesis of the b subunit of F1F0 ATP synthase in Escherichia coli: Glu-77 through Gln-85. J Bacteriol 173 (1991) 7240-8.

[115] G. Singh, and T.A. Cooper, Minigene reporter for identification and analysis of cis elements and trans factors affecting pre-mRNA splicing. Biotechniques 41 (2006) 177-81.

[116] A.S. Xiong, Q.H. Yao, R.H. Peng, X. Li, H.Q. Fan, Z.M. Cheng, and Y. Li, A simple, rapid, high-fidelity and cost-effective PCR-based two-step DNA synthesis method for long gene sequences. Nucleic Acids Res 32 (2004) e98.

[117] S. Kosuri, N. Eroshenko, E.M. Leproust, M. Super, J. Way, J.B. Li, and G.M. Church, Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. Nat Biotechnol 28 (2010) 1295-9.

[118] M.T. Reetz, D. Kahakeaw, and R. Lohmer, Addressing the numbers problem in directed evolution. Chembiochem 9 (2008) 1797-804.

[119] E.C. Brockmann, S. Akter, T. Savukoski, T. Huovinen, A. Lehmusvuori, J. Leivo, O. Saavalainen, A. Azhayev, T. Lövgren, J. Hellman, and U. Lamminmäki, Synthetic single-framework antibody library integrated with rapid affinity maturation by VL shuffling. Protein Eng Des Sel 24 (2011) 691-700.

[120] F.A. Fellouse, C. Wiesmann, and S.S. Sidhu, Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. Proc Natl Acad Sci U S A 101 (2004) 12467-72.

[121] F.A. Fellouse, B. Li, D.M. Compaan, A.A. Peden, S.G. Hymowitz, and S.S. Sidhu, Molecular recognition by a binary code. J Mol Biol 348 (2005) 1153-62.

[122] S. Birtalan, R.D. Fisher, and S.S. Sidhu, The functional capacity of the natural amino acids for molecular recognition. Mol Biosyst 6 (2010) 1186-94.

[123] M.A. Cleary, K. Kilian, Y. Wang, J. Bradshaw, G. Cavet, W. Ge, A. Kulkarni, P.J. Paddison, K. Chang, N. Sheth, E. Leproust, E.M. Coffey, J. Burchard, W.R. McCombie, P. Linsley, and G.J. Hannon, Production of complex nucleic acid libraries using highly parallel in situ oligonucleotide synthesis. Nat Methods 1 (2004) 241-8.

[124] R.P. Patwardhan, C. Lee, O. Litvin, D.L. Young, D. Pe'er, and J. Shendure, High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat Biotechnol 27 (2009) 1173-5.

[125] A. Yagodkin, A. Azhayev, J. Roivainen, M. Antopolsky, A. Kayushin, M. Korosteleva, A. Miroshnikov, J. Randolph, and H. Mackie, Improved synthesis of trinucleotide phosphoramidites and generation of randomized oligonucleotide libraries. Nucleosides Nucleotides Nucleic Acids 26 (2007) 473-97.

[126] B. Virnekäs, L. Ge, A. Plückthun, K.C. Schneider, G. Wellnhofer, and S.E. Moroney, Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. Nucleic Acids Res 22 (1994) 5600-7.

[127] J. Prassler, S. Thiel, C. Pracht, A. Polzer, S. Peters, M. Bauer, S. Nörenberg, Y. Stark, J. Kölln, A. Popp, S. Urlinger, and M. Enzelberger, HuCAL PLATINUM, a synthetic Fab library optimized for sequence diversity and superior performance in mammalian expression systems. J Mol Biol 413 (2011) 261-78.

[128] J. Van den Brulle, M. Fischer, T. Langmann, G. Horn, T. Waldmann, S. Arnold, M. Fuhrmann, O. Schatz, T. O'Connell, D. O'Connell, A. Auckenthaler, and H. Schwer, A novel solid phase technology for high-throughput gene synthesis. Biotechniques 45 (2008) 340-3.

[129] K.M. Schlinkmann, M. Hillenbrand, A. Rittner, M. Künz, R. Strohner, and A. Plückthun, Maximizing detergent stability and functional expression of a GPCR by exhaustive recombination and evolution. J Mol Biol 422 (2012) 414-28.

[130] M.D. Hughes, D.A. Nagel, A.F. Santos, A.J. Sutherland, and A.V. Hine, Removing the redundancy from randomised gene libraries. J Mol Biol 331 (2003) 973-9.

[131] J. Sambrook, E.F. Fritsch, and T. Maniatis, Molecular cloning a laboratory manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.

[132] C.T. Chung, S.L. Niemela, and R.H. Miller, One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. Proc Natl Acad Sci U S A 86 (1989) 2172-5.

[133] L.C. Mattheakis, R.R. Bhatt, and W.J. Dower, An in vitro polysome display system for identifying ligands from very large peptide libraries. Proc Natl Acad Sci U S A 91 (1994) 9022-6.

[134] T.T. Wu, G. Johnson, and E.A. Kabat, Length distribution of CDRH3 in antibodies. Proteins 16 (1993) 1-7.

[135] Z. Pancer, C.T. Amemiya, G.R. Ehrhardt, J. Ceitlin, G.L. Gartland, and M.D. Cooper, Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature 430 (2004) 174-80.

[136] N. Nagano, C.A. Orengo, and J.M. Thornton, One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J Mol Biol 321 (2002) 741-65.

[137] J.D. Marks, H.R. Hoogenboom, T.P. Bonnert, J. McCafferty, A.D. Griffiths, and G. Winter, By-passing immunization. Human antibodies from V-gene libraries displayed on phage. J Mol Biol 222 (1991) 581-97.
[138] T.J. Vaughan, A.J. Williams, K. Pritchard, J.K. Osbourn, A.R. Pope, J.C. Earnshaw, J. McCafferty, R.A. Hodits, J. Wilton, and K.S. Johnson, Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. Nat Biotechnol 14 (1996) 309-14.
[139] H.J. de Haard, N. van Neer, A. Reurs, S.E. Hufton, R.C. Roovers, P. Henderikx, A.P. de Bruïne, J.W. Arends, and H.R. Hoogenboom, A large non-immunized human Fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies. J Biol Chem 274 (1999) 18218-30.
[140] D. Sblattero, and A. Bradbury, Exploiting recombination in single bacteria to make large phage antibody libraries. Nat Biotechnol 18 (2000) 75-80.
[141] D.J. Schofield, A.R. Pope, V. Clementel, J. Buckell, S.D.j. Chapple, K.F. Clarke, J.S. Conquer, A.M. Crofts, S.R. Crowther, M.R. Dyson, G. Flack, G.J. Griffin, Y. Hooks, W.J. Howat, A. Kolb-Kokocinski, S. Kunze, C.D. Martin, G.L. Maslen, J.N. Mitchell, M. O'Sullivan, R.L. Perera, W. Roake, S.P. Shadbolt, K.J. Vincent, A. Warford, W.E. Wilson, J. Xie, J.L. Young, and J. McCafferty, Application of phage display to high throughput antibody generation and characterization. Genome Biol 8 (2007) R254.
[142] P. Pansri, N. Jaruseranee, K. Rangnoi, P. Kristensen, and M. Yamabhai, A compact phage display human scFv library for selection of antibodies to a wide variety of antigens. BMC Biotechnol 9 (2009) 6.
[143] J. Glanville, W. Zhai, J. Berka, D. Telman, G. Huerta, G.R. Mehta, I. Ni, L. Mei, P.D. Sundar, G.M. Day, D. Cox, A. Rajpal, and J. Pons, Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. Proc Natl Acad Sci U S A 106 (2009) 20216-21.
[144] C. Lloyd, D. Lowe, B. Edwards, F. Welsh, T. Dilks, C. Hardman, and T. Vaughan, Modelling the human immune response: performance of a 1011 human antibody repertoire against a broad panel of therapeutically relevant antigens. Protein Eng Des Sel 22 (2009) 159-68.
[145] R.M. Horton, Z.L. Cai, S.N. Ho, and L.R. Pease, Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. Biotechniques 8 (1990) 528-35.
[146] D. Lipovsek, S.M. Lippow, B.J. Hackel, M.W. Gregson, P. Cheng, A. Kapila, and K.D. Wittrup, Evolution of an interloop disulfide bond in high-affinity antibody mimics based on fibronectin type III domain and selected by yeast surface display: molecular convergence with single-domain camelid and shark antibodies. J Mol Biol 368 (2007) 1024-41.
[147] C.M. Mahon, M.A. Lambert, J. Glanville, J.M. Wade, B.J. Fennell, M.R. Krebs, D. Armellino, S. Yang, X. Liu, C.M. O'Sullivan, B. Autin, K. Oficjalska, L. Bloom, J. Paulsen, D. Gill, M. Damelin, O. Cunningham, and W.J. Finlay, Comprehensive Interrogation of a Minimalist Synthetic

CDR-H3 Library and Its Ability to Generate Antibodies with Therapeutic Potential. J Mol Biol (2013) 1712-30.

[148] A. Pini, F. Viti, A. Santucci, B. Carnemolla, L. Zardi, P. Neri, and D. Neri, Design and use of a phage display library. Human antibodies with subnanomolar affinity against a marker of angiogenesis eluted from a two-dimensional gel. J Biol Chem 273 (1998) 21769-76.

[149] C.F. Barbas, J.D. Bain, D.M. Hoekstra, and R.A. Lerner, Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem. Proc Natl Acad Sci U S A 89 (1992) 4457-61.

[150] X. Ge, Y. Mazor, S.P. Hunicke-Smith, A.D. Ellington, and G. Georgiou, Rapid construction and characterization of synthetic antibody libraries without DNA amplification. Biotechnol Bioeng 106 (2010) 347-57.

[151] A. Koide, J. Wojcik, R.N. Gilbreth, R.J. Hoey, and S. Koide, Teaching an old scaffold new tricks: monobodies constructed using alternative surfaces of the FN3 scaffold. J Mol Biol 415 (2012) 393-405.

[152] C.V. Lee, W.C. Liang, M.S. Dennis, C. Eigenbrot, S.S. Sidhu, and G. Fuh, High-affinity human antibodies from phage-displayed synthetic Fab libraries with a single framework scaffold. J Mol Biol 340 (2004) 1073-93.

[153] S.S. Sidhu, B. Li, Y. Chen, F.A. Fellouse, C. Eigenbrot, and G. Fuh, Phage-displayed antibody libraries of synthetic heavy chain complementarity determining regions. J Mol Biol 338 (2004) 299-310.

[154] L. Shi, J.C. Wheeler, R.W. Sweet, J. Lu, J. Luo, M. Tornetta, B. Whitaker, R. Reddy, R. Brittingham, L. Borozdina, Q. Chen, B. Amegadzie, D.M. Knight, J.C. Almagro, and P. Tsui, De novo selection of high-affinity antibodies from synthetic fab libraries displayed on phage as pIX fusion proteins. J Mol Biol 397 (2010) 385-96.

[155] T. Kunkel, Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci U S A 82 (1985) 488-92.

[156] C. Rothe, S. Urlinger, C. Löhning, J. Prassler, Y. Stark, U. Jäger, B. Hubner, M. Bardroff, I. Pradel, M. Boss, R. Bittlingmaier, T. Bataa, C. Frisch, B. Brocks, A. Honegger, and M. Urban, The human combinatorial antibody library HuCAL GOLD combines diversification of all six CDRs according to the natural immune system with a novel display method for efficient selection of high-affinity antibodies. J Mol Biol 376 (2008) 1182-200.

[157] T. Tiller, I. Schuster, D. Deppe, K. Siegers, R. Strohner, T. Herrmann, M. Berenguer, D. Poujol, J. Stehle, Y. Stark, M. Heßling, D. Daubert, K. Felderer, S. Kaden, J. Kölln, M. Enzelberger, and S. Urlinger, A fully synthetic human Fab antibody library based on fixed VH/VL framework pairings with favorable biophysical properties. MAbs 5 (2013) 445 - 470.

[158] W. Zhai, J. Glanville, M. Fuhrmann, L. Mei, I. Ni, P.D. Sundar, T. Van Blarcom, Y. Abdiche, K. Lindquist, R. Strohner, D. Telman, G. Cappuccilli, W.J. Finlay, J. Van den Brulle, D.R. Cox, J. Pons, and A. Rajpal, Synthetic antibodies designed on natural sequence landscapes. J Mol Biol 412 (2011) 55-71.

[159] V. Stein, and F. Hollfelder, An efficient method to assemble linear DNA templates for in vitro screening and selection systems. Nucleic Acids Res 37 (2009) e122.
[160] B.R. Villiers, V. Stein, and F. Hollfelder, USER friendly DNA recombination (USERec): a simple and flexible near homology-independent method for gene library construction. Protein Eng Des Sel 23 (2010) 1-8.
[161] C. Votsmeier, H. Plittersdorf, O. Hesse, A. Scheidig, M. Strerath, U. Gritzan, K. Pellengahr, P. Scholz, A. Eicker, D. Myszka, W.M. Coco, and U. Haupts, Femtomolar Fab binding affinities to a protein target by alternative CDR residue co-optimization strategies without phage or cell surface display. MAbs 4 (2012) 341-8.
[162] A. Nobili, M.G. Gall, I.V. Pavlidis, M.L. Thompson, M. Schmidt, and U.T. Bornscheuer, Use of 'small but smart' libraries to enhance the enantioselectivity of an esterase from Bacillus stearothermophilus towards tetrahydrofuran-3-yl acetate. FEBS J 280 (2013) 3084-93.
[163] H. Jochens, D. Aerts, and U.T. Bornscheuer, Thermostabilization of an esterase by alignment-guided focussed directed evolution. Protein Eng Des Sel 23 (2010) 903-9.
[164] K.M. Schlinkmann, A. Honegger, E. Türeci, K.E. Robison, D. Lipovšek, and A. Plückthun, Critical features for biosynthesis, stability, and functionality of a G protein-coupled receptor uncovered by all-versus-all mutations. Proc Natl Acad Sci U S A 109 (2012) 9810-5.
[165] J.C. Bauer, D.A. Wright, J.C. Braman, and R.S. Geha, Circular site-directed mutagenesis. in: U.S.P.a.T. Office, (Ed.), United States Patents, Stratagene and Childrens Medical Center, USA, 1998, pp. 10.
[166] M. McClelland, and M. Nelson, Effect of site-specific methylation on DNA modification methyltransferases and restriction endonucleases. Nucleic Acids Res 20 Suppl (1992) 2145-57.
[167] Agilent Technologies, QuikChange II XL Site-Directed Mutagenesis Kit - Instruction Manual, Agilent Technologies, 2012, pp. 21.
[168] M. Zoller, and M. Smith, Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template. DNA 3 (1984) 479-88.
[169] Promega, GeneEditor in vitro Site-Directed Mutagenesis System - Technical Manual, Promega Corporation, USA, 2006, pp. 28.
[170] M.K. Lewis, and D.V. Thompson, Efficient site directed in vitro mutagenesis using ampicillin selection. Nucleic Acids Res 18 (1990) 3439-43.
[171] H. Ohmori, A new method for strand discrimination in sequence-directed mutagenesis. Nucleic Acids Res 22 (1994) 884-5.
[172] W.P. Deng, and J.A. Nickoloff, Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. Anal Biochem 200 (1992) 81-8.
[173] Thermo Fisher Scientific, Phusion Site-Directed Mutagenesis Kit - Product Information, Thermo Fisher Scientific, Inc., 2013, pp. 7.
[174] A. Hemsley, N. Arnheim, M.D. Toney, G. Cortopassi, and D.J. Galas, A simple method for site-directed mutagenesis using the polymerase chain reaction. Nucleic Acids Res 17 (1989) 6545-51.

[175] B.C. Cunningham, and J.A. Wells, High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science 244 (1989) 1081-5.
[176] M.E. Hulley, H.S. Toogood, A. Fryszkowska, D. Mansell, G.M. Stephens, J.M. Gardiner, and N.S. Scrutton, Focused directed evolution of pentaerythritol tetranitrate reductase by using automated anaerobic kinetic screening of site-saturated libraries. Chembiochem 11 (2010) 2433-47.
[177] W. Huang, J. Petrosino, M. Hirsch, P.S. Shenkin, and T. Palzkill, Amino acid sequence determinants of beta-lactamase structure and activity. J Mol Biol 258 (1996) 688-703.
[178] A.D.W. Griffiths, S C Hartley, O Tomlinson, I M Waterhouse, P Crosby, W L Kontermann, R E Jones, P T Low, N M Allison, T J, Isolation of high affinity human antibodies directly from large synthetic repertoires The EMBO Journal 13 (1994) 3245-60.
[179] K. Abremski, and R. Hoess, Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. J Biol Chem 259 (1984) 1509-14.
[180] K. Abremski, R. Hoess, and N. Sternberg, Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. Cell 32 (1983) 1301-11.
[181] L. Blaise, A. Wehnert, M.P. Steukers, T. van den Beucken, H.R. Hoogenboom, and S.E. Hufton, Construction and diversification of yeast cell surface displayed libraries by yeast mating: application to the affinity maturation of Fab antibody fragments. Gene 342 (2004) 211-8.
[182] A. Zhao, S. Nunez-Cruz, C. Li, G. Coukos, D.L. Siegel, and N. Scholler, Rapid isolation of high-affinity human antibodies against the tumor vascular marker Endosialin/TEM1, using a paired yeast-display/secretory scFv library platform. J Immunol Methods 363 (2011) 221-32.
[183] F. Storici, L.K. Lewis, and M.A. Resnick, In vivo site-directed mutagenesis using oligonucleotides. Nat Biotechnol 19 (2001) 773-6.
[184] N. Pirakitikulr, N. Ostrov, P. Peralta-Yahya, and V.W. Cornish, PCRless library mutagenesis via oligonucleotide recombination in yeast. Protein Sci 19 (2010) 2336-46.
[185] L. Benatuil, J.M. Perez, J. Belk, and C.M. Hsieh, An improved yeast transformation method for the generation of very large human antibody libraries. Protein Eng Des Sel 23 (2010) 155-9.
[186] W.P. Stemmer, DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc Natl Acad Sci U S A 91 (1994) 10747-51.
[187] H. Zhao, L. Giver, Z. Shao, J.A. Affholter, and F.H. Arnold, Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat Biotechnol 16 (1998) 258-61.
[188] W.M. Coco, W.E. Levinson, M.J. Crist, H.J. Hektor, A. Darzins, P.T. Pienkos, C.H. Squires, and D.J. Monticello, DNA shuffling method for

generating highly recombined genes and evolved enzymes. Nat Biotechnol 19 (2001) 354-9.
[189] J.E. Ness, S. Kim, A. Gottman, R. Pak, A. Krebber, T.V. Borchert, S. Govindarajan, E.C. Mundorff, and J. Minshull, Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. Nat Biotechnol 20 (2002) 1251-5.
[190] M. Ostermeier, J.H. Shim, and S.J. Benkovic, A combinatorial approach to hybrid enzymes independent of DNA homology. Nat Biotechnol 17 (1999) 1205-9.
[191] S. Lutz, M. Ostermeier, and S.J. Benkovic, Rapid generation of incremental truncation libraries for protein engineering using alpha-phosphothioate nucleotides. Nucleic Acids Res 29 (2001) E16.
[192] S. Lutz, M. Ostermeier, G.L. Moore, C.D. Maranas, and S.J. Benkovic, Creating multiple-crossover DNA libraries independent of sequence identity. Proc Natl Acad Sci U S A 98 (2001) 11248-53.
[193] K. Hiraga, and F.H. Arnold, General method for sequence-independent site-directed chimeragenesis. J Mol Biol 330 (2003) 287-96.
[194] C. Engler, R. Gruetzner, R. Kandzia, and S. Marillonnet, Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS One 4 (2009) e5553.
[195] E.C. Brockmann, M. Cooper, N. Strömsten, M. Vehniäinen, and P. Saviranta, Selecting for antibody scFv fragments with improved stability using phage display with denaturation under reducing conditions. J Immunol Methods 296 (2005) 159-70.
[196] T. Korpimäki, E. Brockmann, O. Kuronen, M. Saraste, U. Lamminmäki, and M. Tuomola, Engineering of a broad specificity antibody for simultaneous detection of 13 sulfonamides at the maximum residue level. J Agric Food Chem 52 (2004) 40-7.
[197] N.C. Shaner, M.Z. Lin, M.R. McKeown, P.A. Steinbach, K.L. Hazelwood, M.W. Davidson, and R.Y. Tsien, Improving the photostability of bright monomeric orange and red fluorescent proteins. Nat Methods 5 (2008) 545-51.
[198] X. Yan, and Z. Xu, Ribosome-display technology: applications for directed evolution of functional proteins. Drug Discov Today 11 (2006) 911-6.
[199] N. Gera, M. Hussain, and B.M. Rao, Protein selection using yeast surface display. Methods 60 (2012) 15-26.
[200] L. Giovannoni, F. Viti, L. Zardi, and D. Neri, Isolation of anti-angiogenesis antibodies from a large combinatorial repertoire by colony filter screening. Nucleic Acids Res 29 (2001) E27.
[201] R.M. de Wildt, C.R. Mundy, B.D. Gorick, and I.M. Tomlinson, Antibody arrays for high-throughput screening of antibody-antigen interactions. Nat Biotechnol 18 (2000) 989-94.
[202] P. Angenendt, J. Wilde, G. Kijanka, S. Baars, D.J. Cahill, J. Kreutzberger, H. Lehrach, Z. Konthur, and J. Glökler, Seeing better through a MIST: evaluation of monoclonal recombinant antibody fragments on microarrays. Anal Chem 76 (2004) 2916-21.

[203] A.R. Bradbury, S. Sidhu, S. Dübel, and J. McCafferty, Beyond natural antibodies: the power of in vitro display technologies. Nat Biotechnol 29 (2011) 245-54.
[204] G. Løset, and I. Sandlie, Next generation phage display by use of pVII and pIX as display scaffolds. Methods 58 (2012) 40-6.
[205] C. Gao, S. Mao, C.H. Lo, P. Wirsching, R.A. Lerner, and K.D. Janda, Making artificial antibodies: a format for phage display of combinatorial heterodimeric arrays. Proc Natl Acad Sci U S A 96 (1999) 6025-30.
[206] P. Malik, T.D. Terry, F. Bellintani, and R.N. Perham, Factors limiting display of foreign peptides on the major coat protein of filamentous bacteriophage capsids and a potential role for leader peptidase. FEBS Lett 436 (1998) 263-6.
[207] S.E. Hufton, P.T. Moerkerk, E.V. Meulemans, A. de Bruïne, J.W. Arends, and H.R. Hoogenboom, Phage display of cDNA repertoires: the pVI display system and its applications for the selection of immunogenic ligands. J Immunol Methods 231 (1999) 39-51.
[208] C.F. Barbas, A.S. Kang, R.A. Lerner, and S.J. Benkovic, Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. Proc Natl Acad Sci U S A 88 (1991) 7978-82.
[209] M. Dueñas, and C.A. Borrebaeck, Clonal selection and amplification of phage displayed antibodies by linking antigen recognition and phage replication. Biotechnology (N Y) 12 (1994) 999-1002.
[210] K. Gramatikoff, O. Georgiev, and W. Schaffner, Direct interaction rescue, a novel filamentous phage technique to study protein-protein interactions. Nucleic Acids Res 22 (1994) 5761-2.
[211] S.S. Sidhu, G.A. Weiss, and J.A. Wells, High copy display of large proteins on phage for functional selections. J Mol Biol 296 (2000) 487-95.
[212] G.R. Nakayama, G. Valkirs, D. McGrath, and W.D. Huse, Improving the copy numbers of antibody fragments expressed on the major coat protein of bacteriophage M13. Immunotechnology 2 (1996) 197-207.
[213] G.P. Smith, Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science 228 (1985) 1315-7.
[214] G.P. Smith, D.A. Schultz, and J.E. Ladbury, A ribonuclease S-peptide antagonist discovered with a bacteriophage display library. Gene 128 (1993) 37-42.
[215] H.B. Lowman, S.H. Bass, N. Simpson, and J.A. Wells, Selecting high-affinity binding proteins by monovalent phage display. Biochemistry 30 (1991) 10832-8.
[216] M. Russel, S. Kidd, and M.R. Kelley, An improved filamentous helper phage for generating single-stranded plasmid DNA. Gene 45 (1986) 333-8.
[217] S. Rondot, J. Koch, F. Breitling, and S. Dübel, A helper phage to improve single-chain antibody presentation in phage display. Nat Biotechnol 19 (2001) 75-8.
[218] P. Kristensen, and G. Winter, Proteolytic selection for protein folding using filamentous bacteriophages. Fold Des 3 (1998) 321-8.
[219] S.S. Sidhu, and F.A. Fellouse, Synthetic therapeutic antibodies. Nat Chem Biol 2 (2006) 682-8.

[220] J. Rakonjac, J. Feng, and P. Model, Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII. J Mol Biol 289 (1999) 1253-65.
[221] D. Steiner, P. Forrer, M.T. Stumpp, and A. Plückthun, Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. Nat Biotechnol 24 (2006) 823-31.
[222] S. Schlehuber, G. Beste, and A. Skerra, A novel type of receptor protein, based on the lipocalin scaffold, with specificity for digoxigenin. J Mol Biol 297 (2000) 1105-20.
[223] A. Koide, and S. Koide, Monobodies: antibody mimics based on the scaffold of the fibronectin type III domain. Methods Mol Biol 352 (2007) 95-109.
[224] K. Nord, E. Gunneriusson, J. Ringdahl, S. Ståhl, M. Uhlén, and P.A. Nygren, Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. Nat Biotechnol 15 (1997) 772-7.
[225] A. Ernst, G. Avvakumov, J. Tong, Y. Fan, Y. Zhao, P. Alberts, A. Persaud, J.R. Walker, A.M. Neculai, D. Neculai, A. Vorobyov, P. Garg, L. Beatty, P.K. Chan, Y.C. Juang, M.C. Landry, C. Yeh, E. Zeqiraj, K. Karamboulas, A. Allali-Hassani, M. Vedadi, M. Tyers, J. Moffat, F. Sicheri, L. Pelletier, D. Durocher, B. Raught, D. Rotin, J. Yang, M.F. Moran, S. Dhe-Paganon, and S.S. Sidhu, A strategy for modulation of enzymes in the ubiquitin system. Science 339 (2013) 590-5.
[226] J. Speck, K.M. Arndt, and K.M. Müller, Efficient phage display of intracellularly folded proteins mediated by the TAT pathway. Protein Eng Des Sel 24 (2011) 473-84.
[227] M. Dai, J. Temirov, E. Pesavento, C. Kiss, N. Velappan, P. Pavlik, J.H. Werner, and A.R. Bradbury, Using T7 phage display to select GFP-based binders. Protein Eng Des Sel 21 (2008) 413-24.
[228] R.M. Stroud, P. Serwer, and M.J. Ross, Assembly of bacteriophage T7. Dimensions of the bacteriophage and its capsids. Biophys J 36 (1981) 743-57.
[229] A. Rosenberg, K. Griffin, F.W. Studier, M. McCormick, J. Berg, R. Novy, and R. Mierendorf, T7Select® Phage Display System: A powerful new protein display system based on bacteriophage T7, Innovations - Newsletter of Novagen Inc. Novagen, 1996, pp. 1-6.
[230] L.R. Krumpe, A.J. Atkinson, G.W. Smythers, A. Kandel, K.M. Schumacher, J.B. McMahon, L. Makowski, and T. Mori, T7 lytic phage-displayed peptide libraries exhibit less sequence bias than M13 filamentous phage-displayed peptide libraries. Proteomics 6 (2006) 4210-22.
[231] Novagen, T7Select System Manual. Novagen, 2002, pp. 23.
[232] C. Santini, D. Brennan, C. Mennuni, R.H. Hoess, A. Nicosia, R. Cortese, and A. Luzzago, Efficient display of an HCV cDNA expression library as C-terminal fusion to the capsid protein D of bacteriophage lambda. J Mol Biol 282 (1998) 125-35.
[233] Z. Ren, and L.W. Black, Phage T4 SOC and HOC display of biologically active, full-length proteins on the viral capsid. Gene 215 (1998) 439-44.

[234] Y.H. Lin, L.H. Lee, W.L. Shih, Y.C. Hu, and H.J. Liu, Baculovirus surface display of sigmaC and sigmaB proteins of avian reovirus and immunogenicity of the displayed proteins in a mouse model. Vaccine 26 (2008) 6361-7.

[235] S.S. Seregin, Z.C. Hartman, D.M. Appledorn, S. Godbehere, H. Jiang, M.M. Frank, and A. Amalfitano, Novel adenovirus vectors 'capsid-displaying' a human complement inhibitor. J Innate Immun 2 (2010) 353-9.

[236] S. Michelfelder, and M. Trepel, Adeno-associated viral vectors and their redirection to cell-type specific receptors. Adv Genet 67 (2009) 29-60.

[237] R.C. Münch, M.D. Mühlebach, T. Schaser, S. Kneissl, C. Jost, A. Plückthun, K. Cichutek, and C.J. Buchholz, DARPins: an efficient targeting domain for lentiviral vectors. Mol Ther 19 (2011) 686-93.

[238] A. Charbit, J.C. Boulain, A. Ryter, and M. Hofnung, Probing the topology of a bacterial membrane protein by genetic insertion of a foreign epitope; expression at the cell surface. EMBO J 5 (1986) 3029-37.

[239] Z. Lu, K.S. Murray, V. Van Cleave, E.R. LaVallie, M.L. Stahl, and J.M. McCoy, Expression of thioredoxin random peptide libraries on the Escherichia coli cell surface as functional fusions to flagellin: a system designed for exploring protein-protein interactions. Biotechnology (N Y) 13 (1995) 366-72.

[240] M.J. Olsen, D. Stephens, D. Griffiths, P. Daugherty, G. Georgiou, and B.L. Iverson, Function-based isolation of novel enzymes from a large library. Nat Biotechnol 18 (2000) 1071-4.

[241] P.S. Daugherty, G. Chen, M.J. Olsen, B.L. Iverson, and G. Georgiou, Antibody affinity maturation using bacterial surface display. Protein Eng 11 (1998) 825-32.

[242] P.H. Bessette, J.J. Rice, and P.S. Daugherty, Rapid isolation of high-affinity protein binding peptides using bacterial display. Protein Eng Des Sel 17 (2004) 731-9.

[243] S. Becker, S. Theile, N. Heppeler, A. Michalczyk, A. Wentzel, S. Wilhelm, K.E. Jaeger, and H. Kolmar, A generic system for the Escherichia coli cell-surface display of lipolytic enzymes. FEBS Lett 579 (2005) 1177-82.

[244] H.C. Jung, J.M. Lebeault, and J.G. Pan, Surface display of Zymomonas mobilis levansucrase by using the ice-nucleation protein of Pseudomonas syringae. Nat Biotechnol 16 (1998) 576-80.

[245] J.J. Rice, and P.S. Daugherty, Directed evolution of a biterminal bacterial display scaffold enhances the display of diverse peptides. Protein Eng Des Sel 21 (2008) 435-42.

[246] S.A. Kenrick, and P.S. Daugherty, Bacterial display enables efficient and quantitative peptide affinity maturation. Protein Eng Des Sel 23 (2010) 9-17.

[247] Invitrogen, FliTrx Random Peptide Display Library - For the Specific Identification of Protein-Protein Interactions. Invitrogen Corporation, Carlsbad, CA, USA, 2002, pp. 18.

[248] B.R. Harvey, G. Georgiou, A. Hayhurst, K.J. Jeong, B.L. Iverson, and G.K. Rogers, Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from Escherichia coli-expressed libraries. Proc Natl Acad Sci U S A 101 (2004) 9193-8.

[249] C.A. Sarkar, I. Dodevski, M. Kenig, S. Dudli, A. Mohr, E. Hermans, and A. Plückthun, Directed evolution of a G protein-coupled receptor for expression, stability, and binding selectivity. Proc Natl Acad Sci U S A 105 (2008) 14808-13.

[250] N. Kronqvist, J. Löfblom, A. Jonsson, H. Wernérus, and S. Ståhl, A novel affinity protein selection system based on staphylococcal cell surface display and flow cytometry. Protein Eng Des Sel 21 (2008) 247-55.

[251] F. Fleetwood, N. Devoogdt, M. Pellis, U. Wernery, S. Muyldermans, S. Ståhl, and J. Löfblom, Surface display of a single-domain antibody library on Gram-positive bacteria. Cell Mol Life Sci 70 (2013) 1081-93.

[252] N. Gupta, and E.T. Farinas, Directed evolution of CotA laccase for increased substrate specificity using Bacillus subtilis spores. Protein Eng Des Sel 23 (2010) 679-82.

[253] E.T. Boder, and K.D. Wittrup, Yeast surface display for screening combinatorial polypeptide libraries. Nat Biotechnol 15 (1997) 553-7.

[254] J.A. Rakestraw, D. Aird, P.M. Aha, B.M. Baynes, and D. Lipovsek, Secretion-and-capture cell-surface display for selection of target-binding proteins. Protein Eng Des Sel 24 (2011) 525-30.

[255] T. Tanaka, R. Yamada, C. Ogino, and A. Kondo, Recent developments in yeast cell surface display toward extended applications in biotechnology. Appl Microbiol Biotechnol 95 (2012) 577-91.

[256] C.F. Lu, R.C. Montijn, J.L. Brown, F. Klis, J. Kurjan, H. Bussey, and P.N. Lipke, Glycosyl phosphatidylinositol-dependent cross-linking of alpha-agglutinin and beta 1,6-glucan in the Saccharomyces cerevisiae cell wall. J Cell Biol 128 (1995) 333-40.

[257] C. Cappellaro, C. Baldermann, R. Rachel, and W. Tanner, Mating type-specific cell-cell recognition of Saccharomyces cerevisiae: cell wall attachment and active sites of a- and alpha-agglutinin. EMBO J 13 (1994) 4737-44.

[258] E.T. Boder, K.S. Midelfort, and K.D. Wittrup, Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. Proc Natl Acad Sci U S A 97 (2000) 10701-5.

[259] D.H. Aggen, A.S. Chervin, T.M. Schmitt, B. Engels, J.D. Stone, S.A. Richman, K.H. Piepenbrink, B.M. Baker, P.D. Greenberg, H. Schreiber, and D.M. Kranz, Single-chain VαVβ T-cell receptors function without mispairing with endogenous TCR chains. Gene Ther 19 (2012) 365-74.

[260] A.S. Chervin, D.H. Aggen, J.M. Raseman, and D.M. Kranz, Engineering higher affinity T cell receptors using a T cell display system. J Immunol Methods 339 (2008) 175-84.

[261] I. Chen, B.M. Dorr, and D.R. Liu, A general strategy for the evolution of bond-forming enzymes using yeast display. Proc Natl Acad Sci U S A 108 (2011) 11399-404.

[262] T.V. Pavoor, Y.K. Cho, and E.V. Shusta, Development of GFP-based biosensors possessing the binding properties of antibodies. Proc Natl Acad Sci U S A 106 (2009) 11895-900.

[263] J.M. Van der Vaart, R. te Biesebeke, J.W. Chapman, H.Y. Toschka, F.M. Klis, and C.T. Verrips, Comparison of cell wall proteins of

Saccharomyces cerevisiae as anchors for cell surface expression of heterologous proteins. Appl Environ Microbiol 63 (1997) 615-20.
[264] J. Ito, A. Kosugi, T. Tanaka, K. Kuroda, S. Shibasaki, C. Ogino, M. Ueda, H. Fukuda, R.H. Doi, and A. Kondo, Regulation of the display ratio of enzymes on the Saccharomyces cerevisiae cell surface by the immunoglobulin G and cellulosomal enzyme binding domains. Appl Environ Microbiol 75 (2009) 4149-54.
[265] K.C. Wang, C.A. Patel, J. Wang, X. Wang, P.P. Luo, and P. Zhong, Yeast surface display of antibodies via the heterodimeric interaction of two coiled-coil adapters. J Immunol Methods 354 (2010) 11-9.
[266] M. Mergler, K. Wolf, and M. Zimmermann, Development of a bisphenol A-adsorbing yeast by surface display of the Kluyveromyces yellow enzyme on Pichia pastoris. Appl Microbiol Biotechnol 63 (2004) 418-21.
[267] L. Yue, Z. Chi, L. Wang, J. Liu, C. Madzak, J. Li, and X. Wang, Construction of a new plasmid for surface display on cells of Yarrowia lipolytica. J Microbiol Methods 72 (2008) 116-23.
[268] L. Yi, M.C. Gebhard, Q. Li, J.M. Taft, G. Georgiou, and B.L. Iverson, Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. Proc Natl Acad Sci U S A 110 (2013) 7229-34.
[269] P.M. Bowers, R.A. Horlick, T.Y. Neben, R.M. Toobian, G.L. Tomlinson, J.L. Dalton, H.A. Jones, A. Chen, L. Altobell, X. Zhang, J.L. Macomber, I.P. Krapf, B.F. Wu, A. McConnell, B. Chau, T. Holland, A.D. Berkebile, S.S. Neben, W.J. Boyle, and D.J. King, Coupling mammalian cell surface display with somatic hypermutation for the discovery and maturation of human antibodies. Proc Natl Acad Sci U S A 108 (2011) 20455-60.
[270] C. Zhou, F.W. Jacobsen, L. Cai, Q. Chen, and W.D. Shen, Development of a novel mammalian cell surface antibody display platform. MAbs 2 (2010) 508-18.
[271] J. Rogers, E. Choi, L. Souza, C. Carter, C. Word, M. Kuehl, D. Eisenberg, and R. Wall, Gene segments encoding transmembrane carboxyl termini of immunoglobulin gamma chains. Cell 26 (1981) 19-27.
[272] F. Wen, D.K. Sethi, K.W. Wucherpfennig, and H. Zhao, Cell surface display of functional human MHC class II proteins: yeast display versus insect cell display. Protein Eng Des Sel 24 (2011) 701-9.
[273] F. Crawford, E. Huseby, J. White, P. Marrack, and J.W. Kappler, Mimotopes for alloreactive and conventional T cells in a peptide-MHC display library. PLoS Biol 2 (2004) E90.
[274] M.W. Nirenberg, and J.H. Matthaei, The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci U S A 47 (1961) 1588-602.
[275] G. Zubay, In vitro synthesis of protein in microbial systems. Annu Rev Genet 7 (1973) 267-87.
[276] Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, and T. Ueda, Cell-free translation reconstituted with purified components. Nat Biotechnol 19 (2001) 751-5.

[277] J. Hanes, and A. Plückthun, In vitro selection and evolution of functional proteins by using ribosome display. Proc Natl Acad Sci U S A 94 (1997) 4937-42.

[278] R.W. Roberts, and J.W. Szostak, RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc Natl Acad Sci U S A 94 (1997) 12297-302.

[279] J. Yamaguchi, M. Naimuddin, M. Biyani, T. Sasaki, M. Machida, T. Kubo, T. Funatsu, Y. Husimi, and N. Nemoto, cDNA display: a novel screening method for functional disulfide-rich peptides by solid-phase synthesis and stabilization of mRNA-protein fusions. Nucleic Acids Res 37 (2009) e108.

[280] R. Odegrip, D. Coomber, B. Eldridge, R. Hederer, P.A. Kuhlman, C. Ullman, K. FitzGerald, and D. McGregor, CIS display: In vitro selection of peptides from libraries of protein-DNA complexes. Proc Natl Acad Sci U S A 101 (2004) 2806-10.

[281] B. Eldridge, R.N. Cooley, R. Odegrip, D.P. McGregor, K.J. Fitzgerald, and C.G. Ullman, An in vitro selection strategy for conferring protease resistance to ligand binding peptides. Protein Eng Des Sel 22 (2009) 691-8.

[282] H. Murakami, A. Ohta, H. Ashigai, and H. Suga, A highly flexible tRNA acylation method for non-natural polypeptide synthesis. Nat Methods 3 (2006) 357-9.

[283] Y. Yamagishi, I. Shoji, S. Miyagawa, T. Kawakami, T. Katoh, Y. Goto, and H. Suga, Natural product-like macrocyclic N-methyl-peptide inhibitors against a ubiquitin ligase uncovered from a ribosome-expressed de novo library. Chem Biol 18 (2011) 1562-70.

[284] G. Yin, E.D. Garces, J. Yang, J. Zhang, C. Tran, A.R. Steiner, C. Roos, S. Bajad, S. Hudak, K. Penta, J. Zawada, S. Pollitt, and C.J. Murray, Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system. MAbs 4 (2012) 217 - 225.

[285] M.H. Parker, Y. Chen, F. Danehy, K. Dufu, J. Ekstrom, E. Getmanova, J. Gokemeijer, L. Xu, and D. Lipovsek, Antibody mimics based on human fibronectin type three domain engineered for thermostability and high-affinity binding to vascular endothelial growth factor receptor two. Protein Eng Des Sel 18 (2005) 435-44.

[286] R. Mamluk, I.M. Carvajal, B.A. Morse, H. Wong, J. Abramowitz, S. Aslanian, A.C. Lim, J. Gokemeijer, M.J. Storek, J. Lee, M. Gosselin, M.C. Wright, R.T. Camphausen, J. Wang, Y. Chen, K. Miller, K. Sanders, S. Short, J. Sperinde, G. Prasad, S. Williams, R. Kerbel, J. Ebos, A. Mutsaers, J.D. Mendlein, A.S. Harris, and E.S. Furfine, Anti-tumor effect of CT-322 as an adnectin inhibitor of vascular endothelial growth factor receptor-2. MAbs 2 (2010) 199-208.

[287] A. Stahl, M.T. Stumpp, A. Schlegel, S. Ekawardhani, C. Lehrling, G. Martin, M. Gulotti-Georgieva, D. Villemagne, P. Forrer, H.T. Agostini, and H.K. Binz, Highly potent VEGF-A-antagonistic DARPins as anti-angiogenic agents for topical and intravitreal applications. Angiogenesis 16 (2013) 101-11.

[288] R. Tamaskovic, M. Simon, N. Stefan, M. Schwill, and A. Plückthun, Designed ankyrin repeat proteins (DARPins) from research to therapy. Methods Enzymol 503 (2012) 101-34.
[289] A. Buchanan, F. Ferraro, S. Rust, S. Sridharan, R. Franks, G. Dean, M. McCourt, L. Jermutus, and R. Minter, Improved drug-like properties of therapeutic proteins by directed evolution. Protein Eng Des Sel 25 (2012) 631-8.
[290] D.S. Tawfik, and A.D. Griffiths, Man-made cell-like compartments for molecular evolution. Nat Biotechnol 16 (1998) 652-6.
[291] H. Leemhuis, V. Stein, A.D. Griffiths, and F. Hollfelder, New genotype-phenotype linkages for directed evolution of functional proteins. Curr Opin Struct Biol 15 (2005) 472-8.
[292] N. Doi, and H. Yanagawa, STABLE: protein-DNA fusion system for screening of combinatorial protein libraries in vitro. FEBS Lett 457 (1999) 227-30.
[293] Y. Chen, J. Mandic, and G. Varani, Cell-free selection of RNA-binding proteins using in vitro compartmentalization. Nucleic Acids Res 36 (2008) e128.
[294] J. Bertschinger, and D. Neri, Covalent DNA display as a novel tool for directed evolution of proteins in vitro. Protein Eng Des Sel 17 (2004) 699-707.
[295] A.D. Griffiths, and D.S. Tawfik, Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. EMBO J 22 (2003) 24-35.
[296] R. Gan, S. Furuzawa, T. Kojima, K. Kanie, R. Kato, M. Okochi, H. Honda, and H. Nakano, Directed evolution of angiotensin II-inhibiting peptides using a microbead display. J Biosci Bioeng 109 (2010) 411-7.
[297] F.J. Ghadessy, and P. Holliger, Compartmentalized self-replication: a novel method for the directed evolution of polymerases and other enzymes. Methods Mol Biol 352 (2007) 237-48.
[298] K. Bernath, M. Hai, E. Mastrobattista, A.D. Griffiths, S. Magdassi, and D.S. Tawfik, In vitro compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. Anal Biochem 325 (2004) 151-7.
[299] T. Nishikawa, T. Sunami, T. Matsuura, N. Ichihashi, and T. Yomo, Construction of a gene screening system using giant unilamellar liposomes and a fluorescence-activated cell sorter. Anal Chem 84 (2012) 5017-24.
[300] T. Sumida, H. Yanagawa, and N. Doi, In vitro selection of fab fragments by mRNA display and gene-linking emulsion PCR. J Nucleic Acids 2012 (2012) 371379.
[301] S. Wright, Evolution in Mendelian Populations. Genetics 16 (1931) 97-159.
[302] H.A. Orr, Fitness and its role in evolutionary genetics. Nat Rev Genet 10 (2009) 531-9.
[303] R. de Bruin, K. Spelt, J. Mol, R. Koes, and F. Quattrocchio, Selection of high-affinity phage antibodies from phage display libraries. Nat Biotechnol 17 (1999) 397-9.

[304] J. Beekwilder, J. Rakonjac, M. Jongsma, and D. Bosch, A phagemid vector using the E. coli phage shock promoter facilitates phage display of toxic proteins. Gene 228 (1999) 23-31.
[305] G.A. Kuzmicheva, P.K. Jayanna, I.B. Sorokulova, and V.A. Petrenko, Diversity and censoring of landscape phage libraries. Protein Eng Des Sel 22 (2009) 9-18.
[306] T.L. O'Loughlin, W.M. Patrick, and I. Matsumura, Natural history as a predictor of protein evolvability. Protein Eng Des Sel 19 (2006) 439-42.
[307] G. Raghunathan, J. Smart, J. Williams, and J.C. Almagro, Antigen-binding site anatomy and somatic mutations in antibodies that recognize different types of antigens. J Mol Recognit 25 (2012) 103-13.
[308] S. Kauffman, and S. Levin, Towards a general theory of adaptive walks on rugged landscapes. J Theor Biol 128 (1987) 11-45.
[309] P.A. Romero, and F.H. Arnold, Exploring protein fitness landscapes by directed evolution. Nat Rev Mol Cell Biol 10 (2009) 866-76.
[310] J.D. Bloom, S.T. Labthavikul, C.R. Otey, and F.H. Arnold, Protein stability promotes evolvability. Proc Natl Acad Sci U S A 103 (2006) 5869-74.
[311] N. Tokuriki, and D.S. Tawfik, Protein dynamism and evolvability. Science 324 (2009) 203-7.
[312] P.E. Tomatis, S.M. Fabiane, F. Simona, P. Carloni, B.J. Sutton, and A.J. Vila, Adaptive protein evolution grants organismal fitness by improving catalysis and flexibility. Proc Natl Acad Sci U S A 105 (2008) 20605-10.
[313] E. Dellus-Gur, A. Toth-Petroczy, M. Elias, and D.S. Tawfik, What Makes a Protein Fold Amenable to Functional Innovation? Fold Polarity and Stability Trade-offs. J Mol Biol 425 (2013) 2609-21.
[314] S. Sidhu, H. Lowman, B. Cunningham, and J. Wells, Phage display for selection of novel binding peptides. Methods Enzymol 328 (2000) 333-63.
[315] M.J. Feldhaus, R.W. Siegel, L.K. Opresko, J.R. Coleman, J.M. Feldhaus, Y.A. Yeung, J.R. Cochran, P. Heinzelman, D. Colby, J. Swers, C. Graff, H.S. Wiley, and K.D. Wittrup, Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library. Nat Biotechnol 21 (2003) 163-70.
[316] D. Lipovsek, and A. Plückthun, In-vitro protein evolution by ribosome display and mRNA display. J Immunol Methods 290 (2004) 51-67.
[317] S.J. Park, and J.R. Cochran, Protein engineering and design. CRC Press, Boca Raton, 2010.
[318] P.A. Carr, J.S. Park, Y.J. Lee, T. Yu, S. Zhang, and J.M. Jacobson, Protein-mediated error correction for de novo DNA synthesis. Nucleic Acids Res 32 (2004) e162.
[319] N. Tokuriki, F. Stricher, J. Schymkowitz, L. Serrano, and D.S. Tawfik, The stability effects of protein mutations appear to be universally distributed. J Mol Biol 369 (2007) 1318-32.
[320] H.H. Guo, J. Choe, and L.A. Loeb, Protein tolerance to random amino acid change. Proc Natl Acad Sci U S A 101 (2004) 9205-10.
[321] J.U. Bowie, J.F. Reidhaar-Olson, W.A. Lim, and R.T. Sauer, Deciphering the message in protein sequences: tolerance to amino acid substitutions. Science 247 (1990) 1306-10.

[322] X. Qu, R. Swanson, R. Day, and J. Tsai, A guide to template based structure prediction. Curr Protein Pept Sci 10 (2009) 270-85.
[323] T.S. Wong, D. Roccatano, and U. Schwaneberg, Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries. Environ Microbiol 9 (2007) 2645-59.
[324] M.T. Reetz, M. Bocola, J.D. Carballeira, D. Zha, and A. Vogel, Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. Angew Chem Int Ed Engl 44 (2005) 4192-6.
[325] A.E. Firth, and W.M. Patrick, GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. Nucleic Acids Res 36 (2008) W281-5.
[326] H. Persson, J. Lantto, and M. Ohlin, A focused antibody library for improved hapten recognition. J Mol Biol 357 (2006) 607-20.
[327] C.W. Cobaugh, J.C. Almagro, M. Pogson, B. Iverson, and G. Georgiou, Synthetic antibody libraries focused towards peptide ligands. J Mol Biol 378 (2008) 622-33.
[328] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, and P.G. Wolynes, Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 21 (1995) 167-95.
[329] H. Thie, T. Schirrmann, M. Paschke, S. Dübel, and M. Hust, SRP and Sec pathway leader peptides for antibody phage display and antibody fragment production in E. coli. N Biotechnol 25 (2008) 49-54.
[330] N. Velappan, H.E. Fisher, E. Pesavento, L. Chasteen, S. D'Angelo, C. Kiss, M. Longmire, P. Pavlik, and A.R. Bradbury, A comprehensive analysis of filamentous phage display vectors for cytoplasmic proteins: an analysis with different fluorescent proteins. Nucleic Acids Res 38 (2010) e22.
[331] J. Wojcik, O. Hantschel, F. Grebien, I. Kaupe, K.L. Bennett, J. Barkinge, R.B. Jones, A. Koide, G. Superti-Furga, and S. Koide, A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. Nat Struct Mol Biol 17 (2010) 519-27.
[332] B.C. Berks, F. Sargent, and T. Palmer, The Tat protein export pathway. Mol Microbiol 35 (2000) 260-74.
[333] Y.L. Boersma, M.J. Dröge, and W.J. Quax, Selection strategies for improved biocatalysts. FEBS J 274 (2007) 2181-95.
[334] A. Aharoni, K. Thieme, C.P. Chiu, S. Buchini, L.L. Lairson, H. Chen, N.C. Strynadka, W.W. Wakarchuk, and S.G. Withers, High-throughput screening methodology for the directed evolution of glycosyltransferases. Nat Methods 3 (2006) 609-14.
[335] S.W. Santoro, and P.G. Schultz, Directed evolution of the site specificity of Cre recombinase. Proc Natl Acad Sci U S A 99 (2002) 4185-90.
[336] K. Baker, C. Bleczinski, H. Lin, G. Salazar-Jimenez, D. Sengupta, S. Krane, and V.W. Cornish, Chemical complementation: a reaction-independent genetic assay for enzyme catalysis. Proc Natl Acad Sci U S A 99 (2002) 16537-42.
[337] J.K. Michener, and C.D. Smolke, High-throughput enzyme evolution in Saccharomyces cerevisiae using a synthetic RNA switch. Metab Eng 14 (2012) 306-16.

[338] I. Fujii, S. Fukuyama, Y. Iwabuchi, and R. Tanimura, Evolving catalytic antibodies in a phage-displayed combinatorial library. Nat Biotechnol 16 (1998) 463-7.
[339] H. Strobel, D. Ladant, and J.L. Jestin, In vitro selection for enzymatic activity: a model study using adenylate cyclase. J Mol Biol 332 (2003) 1-7.
[340] E. Mastrobattista, V. Taly, E. Chanudet, P. Treacy, B.T. Kelly, and A.D. Griffiths, High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. Chem Biol 12 (2005) 1291-300.
[341] A. Aharoni, G. Amitai, K. Bernath, S. Magdassi, and D.S. Tawfik, High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. Chem Biol 12 (2005) 1281-9.
[342] B. Kintses, C. Hein, M.F. Mohamed, M. Fischlechner, F. Courtois, C. Lainé, and F. Hollfelder, Picoliter cell lysate assays in microfluidic droplet compartments for directed enzyme evolution. Chem Biol 19 (2012) 1001-9.
[343] E. Angov, P.M. Legler, and R.M. Mease, Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding. Methods Mol Biol 705 (2011) 1-13.
[344] S. Ma, I. Saaem, and J. Tian, Error correction in gene synthesis technology. Trends Biotechnol 30 (2012) 147-54.
[345] D. Röthlisberger, O. Khersonsky, A.M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J.L. Gallaher, E.A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K.N. Houk, D.S. Tawfik, and D. Baker, Kemp elimination catalysts by computational enzyme design. Nature 453 (2008) 190-5.
[346] H.B. Larman, G. Jing Xu, N.N. Pavlova, and S.J. Elledge, Construction of a rationally designed antibody platform for sequencing-assisted selection. Proc Natl Acad Sci U S A 109 (2012) 18523-8.
[347] M.W. Traxlmayr, C. Hasenhindl, M. Hackl, G. Stadlmayr, J.D. Rybka, N. Borth, J. Grillari, F. Rüker, and C. Obinger, Construction of a stability landscape of the CH3 domain of human IgG1 by combining directed evolution with high throughput sequencing. J Mol Biol 423 (2012) 397-412.
[348] M.C. Wilson, and J. Piel, Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. Chem Biol 20 (2013) 636-47.
[349] P.P. Peralta-Yahya, F. Zhang, S.B. del Cardayre, and J.D. Keasling, Microbial engineering for the production of advanced biofuels. Nature 488 (2012) 320-8.
[350] H. Alper, C. Fischer, E. Nevoigt, and G. Stephanopoulos, Tuning genetic control through promoter engineering. Proc Natl Acad Sci U S A 102 (2005) 12678-83.
[351] H. Yu, K. Tyo, H. Alper, D. Klein-Marcuschamer, and G. Stephanopoulos, A high-throughput screen for hyaluronic acid accumulation in recombinant Escherichia coli transformed by libraries of engineered sigma factors. Biotechnol Bioeng 101 (2008) 788-96.

[352] B.F. Pfleger, D.J. Pitera, C.D. Smolke, and J.D. Keasling, Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol 24 (2006) 1027-32.
[353] D. Bikard, S. Julié-Galau, G. Cambray, and D. Mazel, The synthetic integron: an in vivo genetic shuffling device. Nucleic Acids Res 38 (2010) e153.
[354] H. Alper, J. Moxley, E. Nevoigt, G.R. Fink, and G. Stephanopoulos, Engineering yeast transcription machinery for improved ethanol tolerance and production. Science 314 (2006) 1565-8.
[355] H.H. Wang, F.J. Isaacs, P.A. Carr, Z.Z. Sun, G. Xu, C.R. Forest, and G.M. Church, Programming cells by multiplex genome engineering and accelerated evolution. Nature 460 (2009) 894-8.
[356] M.J. McLachlan, K. Chockalingam, K.C. Lai, and H. Zhao, Directed evolution of orthogonal ligand specificity in a single scaffold. Angew Chem Int Ed Engl 48 (2009) 7783-6.
[357] E.M. Brustad, and F.H. Arnold, Optimizing non-natural protein function with directed evolution. Curr Opin Chem Biol 15 (2011) 201-10.
[358] J. Grünewald, M.L. Tsao, R. Perera, L. Dong, F. Niessen, B.G. Wen, D.M. Kubitz, V.V. Smider, W. Ruf, M. Nasoff, R.A. Lerner, and P.G. Schultz, Immunochemical termination of self-tolerance. Proc Natl Acad Sci U S A 105 (2008) 11276-80.
[359] J.T. Ngo, J.A. Champion, A. Mahdavi, I.C. Tanrikulu, K.E. Beatty, R.E. Connor, T.H. Yoo, D.C. Dieterich, E.M. Schuman, and D.A. Tirrell, Cell-selective metabolic labeling of proteins. Nat Chem Biol 5 (2009) 715-7.
[360] C. Heinis, T. Rutherford, S. Freund, and G. Winter, Phage-encoded combinatorial chemical libraries based on bicyclic peptides. Nat Chem Biol 5 (2009) 502-7.
[361] R.K. Saiki, S. Scharf, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich, and N. Arnheim, Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230 (1985) 1350-4.
[362] A. Krebber, S. Bornhauser, J. Burmester, A. Honegger, J. Willuda, H.R. Bosshard, and A. Plückthun, Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. J Immunol Methods 201 (1997) 35-55.
[363] K. Bebenek, and T. Kunkel, The use of native T7 DNA polymerase for site-directed mutagenesis. Nucleic Acids Res 17 (1989) 5408.
[364] D. Sagher, and B. Strauss, Abasic sites from cytosine as termination signals for DNA synthesis. Nucleic Acids Res 13 (1985) 4285-98.
[365] S.H. Chan, B.L. Stoddard, and S.Y. Xu, Natural and engineered nicking endonucleases--from cleavage mechanism to engineering of strand-specificity. Nucleic Acids Res 39 (2011) 1-18.
[366] J.J. Krijger, J. Baumann, M. Wagner, K. Schulze, C. Reinsch, T. Klose, O.F. Onuma, C. Simon, S.E. Behrens, and K.D. Breunig, A novel, lactase-based selection and strain improvement strategy for recombinant protein expression in Kluyveromyces lactis. Microb Cell Fact 11 (2012) 112.

[367] R. Fujii, M. Kitaoka, and K. Hayashi, Error-prone rolling circle amplification: the simplest random mutagenesis protocol. Nat Protoc 1 (2006) 2493-7.
[368] D. Christ, K. Famm, and G. Winter, Tapping diversity lost in transformations--in vitro amplification of ligation reactions. Nucleic Acids Res 34 (2006) e108.
[369] D. Metzger, and R. Feil, Engineering the mouse genome by site-specific recombination. Curr Opin Biotechnol 10 (1999) 470-6.
[370] J.A. Esteban, M. Salas, and L. Blanco, Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. J Biol Chem 268 (1993) 2719-26.
[371] B.K. Kay, Phage display of peptides and proteins a laboratory manual. Academic Press, San Diego, Calif. <etc.>, 1996.
[372] M. Hust, M. Meysing, T. Schirrmann, M. Selke, J. Meens, G.F. Gerlach, and S. Dübel, Enrichment of open reading frames presented on bacteriophage M13 using hyperphage. Biotechniques 41 (2006) 335-42.
[373] M. Zemlin, M. Klinger, J. Link, C. Zemlin, K. Bauer, J.A. Engler, H.W. Schroeder, and P.M. Kirkham, Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures. J Mol Biol 334 (2003) 733-49.
[374] A.V. Collis, A.P. Brouwer, and A.C. Martin, Analysis of the antigen combining site: correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. J Mol Biol 325 (2003) 337-54.
[375] C.E. Chan, A.H. Chan, A.P. Lim, and B.J. Hanson, Comparison of the efficiency of antibody selection from semi-synthetic scFv and non-immune Fab phage display libraries against protein targets for rapid development of diagnostic immunoassays. J Immunol Methods 373 (2011) 79-88.
[376] D. Desplancq, D.J. King, A.D. Lawson, and A. Mountain, Multimerization behaviour of single chain Fv variants for the tumour-binding antibody B72.3. Protein Eng 7 (1994) 1027-33.
[377] K. Kitamura, Z. Wang, S. Chowdhury, M. Simadu, M. Koura, and M. Muramatsu, Uracil DNA glycosylase counteracts APOBEC3G-induced hypermutation of hepatitis B viral genomes: excision repair of covalently closed circular DNA. PLoS Pathog 9 (2013) e1003361.
[378] R.E. Cobb, N. Sun, and H. Zhao, Directed evolution as a powerful synthetic biology tool. Methods 60 (2013) 81-90.