In Vitro Evolution of Antibody Affinity using Libraries with Insertions and Deletions

Kalliopi Skamaki

Wolfson College

University of Cambridge

Supervisors:

Dr. Florian Hollfelder

Dr. Ralph Minter

A dissertation submitted for the degree of Doctor of Philosophy

December 2017

Preface

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated.

This dissertation has not been submitted, and is not currently being submitted, for any other degree or qualification at any academic institution and does not exceed 60,000 words.

Kalliopi Skamaki

Acknowledgements

Firstly, I would like to extend my sincerest gratitude to both my supervisors Dr. Florian Hollfelder and Dr. Ralph Minter, who gave me great support and scientific guidance throughout the course of my PhD. I am really grateful for all the help and guidance I received from you both which led to a thoroughly enjoyable PhD experience.

Special thanks to Dr. Stephane Emond who developed the method that allowed me to start my PhD study and all his help. I thank all past and present Hollfelder group members who I had the pleasure of meeting the last four years.

From Medimmune, where I implemented large part of my research, a huge thank you to Dr. Andrew Buchanan and his team for hosting me, especially Dr. Matthieu Chodorge for all his advice and help with of my project. It was great help for me to be a member of your friendly and knowledgeable team.

Many people in Medimmune who took time out of their busy work schedules to impart their experimental knowledge. Thanks to John Andrews for his help with my screening. Ben Kemp and the protein expression team for producing my antibodies, especially Jenifer Spooner. Gareth Rees for his substantial help with my affinity measurements. Bojana Popovic, Jason Breed, Daniel Cannon and Mariana Rangel Pereira for their efforts on crystallography. The DNA chemistry team for their help with sequencing. Other people from Medimmune, Sridharan Sudharsan, Andrew Barnard, Alan Sandercock and from the Biochemistry Department, Kathrine Stott.

Furthermore, I thank Wolfson College for creating such a friendly environment that made my life in Cambridge enjoyable.

I never forget the people who gave me great opportunities to learn before I started my PhD. Dr. Andreas Roussis from the National and Kapodistrian University of Athens. Dr. Hans de Haard and Dr. Peter Ulrich from the company argenx.

Last but not least, I would like to thank my Mother, my Father, all my beloved family and close friends who supported me in all my pursuits.

Finally, I would like to dedicate this thesis to all the people who believed in me and encouraged me throughout the last years.

Abbreviations

BCR	B-cell antigen receptor
bnAbs	Broadly neutralizing monoclonal antibodies
CSR	Class switch recombination
CDRs	Complementarity determining regions
DSF	Differential scanning fluorimetry
Epitopes	Antigenic determinants
ep-PCR	Error-prone PCR
Fab	Fragment antigen binding
Fc	Fragment crystallizable
FWR	Framework regions
FACS	Fluorescence activated cell sorting
GCV	Gene conversion
HCAbs	Heavy chain antibodies
HVLs	Hypervariable loops
Ig	Immunoglobulin
Indels	Insertions and deletions
K_D	Equilibrium dissociation constant
mAbs	Monoclonal antibodies
Paratope	Antigen combining site
scFv	Single-chain variable fragment
V_H	Variable heavy chain domain
V_L	Variable light chain domain

Summary

In Nature, antibodies are capable of recognizing a huge variety of different molecular structures on the surface of antigens. The primary factor that defines the structural diversity of the antibody antigen combining site is the length variation of the complementarity determining region (CDR) loops. Following antigen stimulation, further diversification through the process called somatic hypermutation (SHM) leads to antibodies with improved affinity and specificity. Sequence diversification by SHM is mainly achieved by introduction of point substitutions and a small percentage of insertions/deletions (indels). Although the percentage of indels in affinity matured antibodies is low, probably due to the low rate incorporation of in-frame indels throughout the course of the SHM diversification process, it is likely that the antibody fold can accommodate higher diversity of affinity-enhancing indels. By *in vitro* evolution, other researchers have sampled either only restricted diversity of indels or extended diversity of insertions only in specific positions chosen based on structural information and natural length variation. The aim of this thesis was to study the impact of random and high diversity indels on antibody affinity by *in vitro* evolution.

New approaches for construction of libraries with in-frame amino acid indels were applied to enable sampling of indels of different lengths across the entire antibody variable domains. I followed two different approaches for construction of indel libraries. Firstly, a recently developed random approach allowed the construction of libraries with random insertions and deletions. Secondly, a semi-random approach was developed to build libraries with different lengths of insertions that could be widely applied in future *in vitro* antibody affinity maturation campaigns. Libraries constructed by either of these approaches yielded variants with insertions with improved affinity.

Overall, this thesis demonstrates that insertions besides offering alternative routes to affinity maturation can also be combined with point substitutions to take advantage of additive effects on function.

Table of Contents

1. Introduction

The objective of this thesis is on the development of new approaches for construction of libraries with amino acid insertions and deletions (indels) as a tool for antibody affinity maturation by in vitro evolution and also in helping improve our understanding of protein evolution. This chapter offers an overview of relevant topics for the experimental results presented in later chapters.

1.1. Diversity of the antibody repertoire in Nature

Antibodies, often referred to as immonoglobulins (Ig), are proteins generated throughout the course of the immune response to recognize a huge variety of different molecular structures (antigenic determinants) on the surface of antigens (foreign material) *via* their antigen combining sites. The antigen is eliminated either by inactivation through antibody binding, eg. "neutralizing" a virus by preventing viral entry into host cells, or by employment of other molecules (effector molecules) which trigger antigen destruction, eg. antibody-dependent cellular cytotoxicity (ADCC) by NK cells and phagocytosis by host immune cells.

The structure of the antibody molecule consists of four polypeptide chains, two identical heavy (H) chains (μ, α, γ, δ, ε) and two identical light (L) chains (κ, λ), which form two identical Fab (fragment antigen binding) units and one Fc (fragment crystallizable) unit (**Figure 1**), which are linked by a flexible hinge region. The Fab arms are involved in binding to antigen while the Fc binds to effector molecules. The Fab arm is composed of a variable fragment (Fv) composed of the variable heavy (V_H) and light (V_L) domains and a constant fragment composed of the heavy (C_H1) and light (C_L) domains. The specificity of the antibody is determined by the variability of the Fv, which is conferred by short sequences in the variable domains that sow extreme variation, the hypervariable regions (HV) (three in each variable domain), often referred to as complementarity determining regions (CDRs). The constant Fc region is involved in binding to effector molecules, inducing functions such as complement-dependent cytotoxicity (CDC), complement dependent cell-mediated cytotoxicity (CDCC) and antibody-dependent cellular cytotoxicity (ADCC). These occur by engagement of the Fc region with Fcγ receptors which are expressed on different immune cells (e.g. natural killer cells, mast cells, macrophages and neutrophils) resulting into lysis of

target cells. The Fc also binds to the neonatal Fc receptor (FcRn) resulting IgG serum recycling (Cohen-Solal, 2004). Fc regions differ between different antibody classes (IgG, IgM, IgA, IgD and IgE), leading to triggering of different effector functions.

Structure of the human IgG molecule

Each domain of an IgG molecule folds into a common pattern, the "immunoglobulin fold" (**Figure 1**), which consists of two stacked β-sheets which are further stabilized by a disulphide bond. These highly conserved β-strands are joined by more diverse loops with little secondary structure. The V_H/V_L and C_H1/C_L domain pairing is driven by hydrophobic interactions, while C_H1 and C_L are linked by a disulphide bond. The variable domains are composed of four conserved framework regions (FWR) and hypervariable regions (HV) (T. T. Wu & Kabat, 1970; Kabat & Wu, 1971), the six CDR regions encode structurally variable loops, the hypervariable loops V_L CDR1, CDR2 and CDR3 and the V_H CDR1, CDR2 and CDR3 (**Figure 1**). The CDR loops form the antigen combining site, a single continuous binding surface at the tip of the Fab fragment (Padlan, 1994). The main role of the FWR regions is structural support to the CDR loops but can also affect the conformations of the CDR loops either directly or indirectly by affecting the V_H–V_L interface (Foote & Winter, 1992; Abhinandan & Martin, 2010). The Fc of the IgG is composed of the CH3 domains that are similarly paired, while the CH2 domains do not interact with each other.

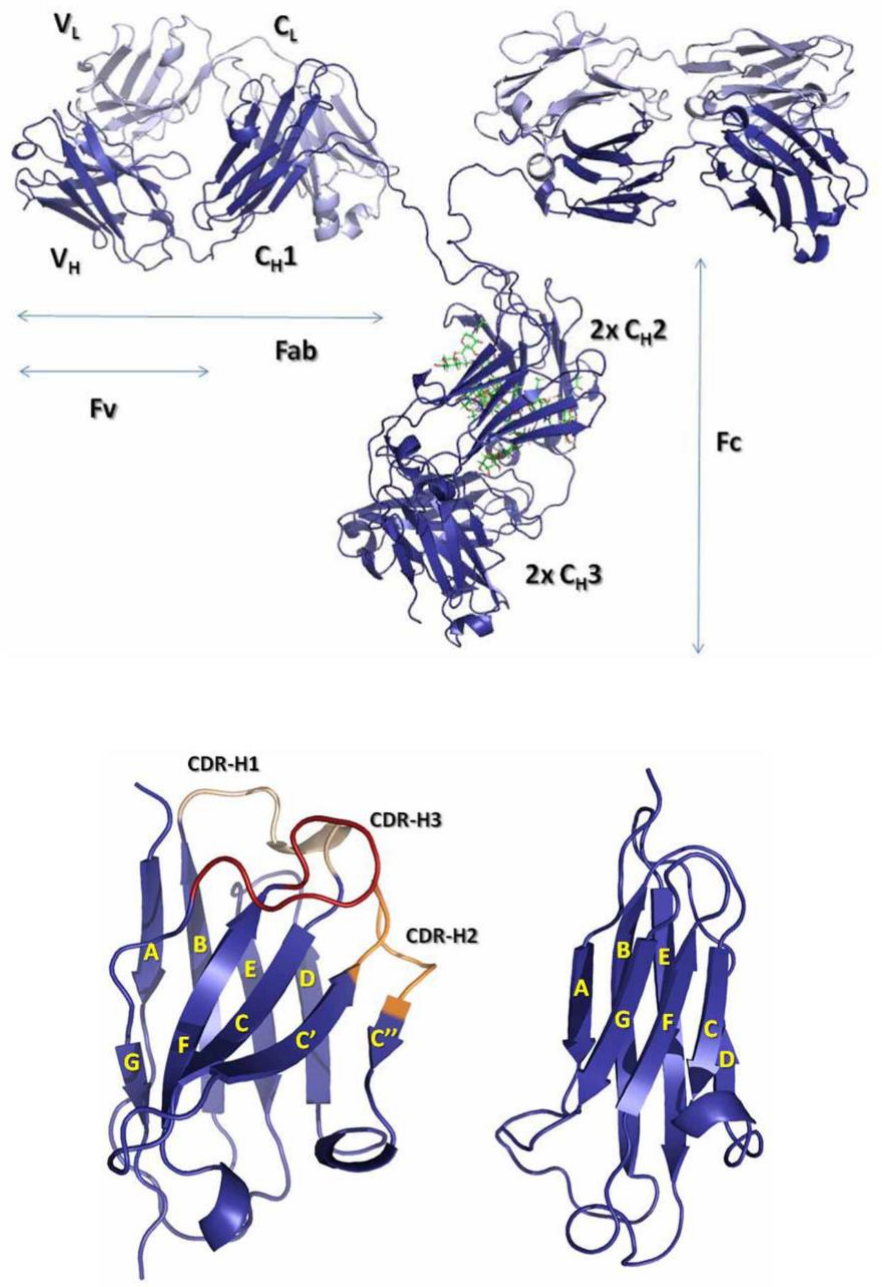


Figure 1. (Above) the IgG molecule and (below) the "immunoglobulin fold", the V_H domain (left) and the CH1 domain (right). The CDRs, CDR1 and CDR2 and CDR3 are shown in yellow, orange and red, respectively (adapted from Finlay & Almagro, 2012).

1.1.1. Primary antibody repertoire

Soluble high affinity antigen-specific antibodies which are produced by plasma cells are the final products of B-cell differentiation (**Figure 2**). The repertoire of B-cell antigen receptors (BCRs), which is the membrane-bound form of the antibody, is initially expressed on the surface of B-cells that have never been exposed to antigen (naïve B-cells). Stimulation of a B-cell by antigen is followed by B-cell proliferation, a process called clonal expansion, that leads to further maturation of antibody affinity and specificity by somatic hypermutation (SHM) and finally generation of specific memory B-cells and plasma cells.

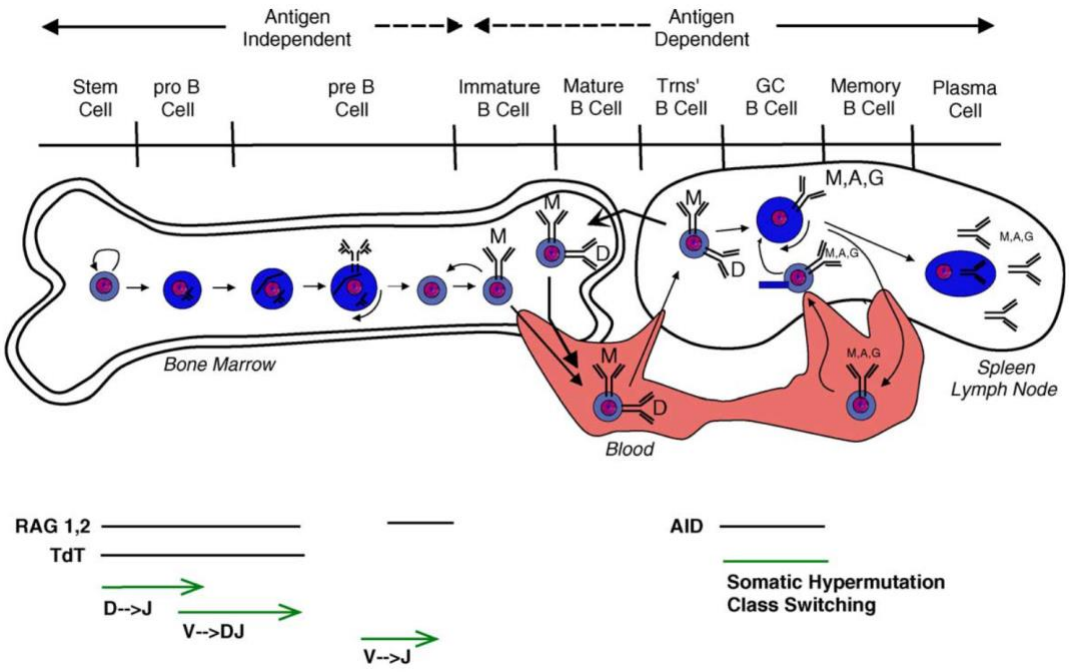


Figure 2. B-cell differentiation and Ig diversification. B cell development starts in the bone marrow where the BCR is expressed as IgM. Mature B cells expressing IgM and IgD migrate from the bone marrow into the blood and the secondary lymphoid organs. Stimulation by antigen causes further maturation of BCR affinity and specificity through the process of somatic hypermutation in the germinal centre (GC) in secondary lymphoid organs and production of memory B-cells and plasma cells. Adapted from (H. W. Schroeder, 2006).

1.1.1.1. Genetic diversity

Humans and mice

The BCR variable (V) heavy and light chain domain repertoires are encoded by rearranged genes that are produced by somatic recombination of multiple gene segments (V(D)J gene segments) which compose the heavy, κ and λ light chain loci. The constant (C) domains are encoded by individual exons which can be combined with different rearranged V-region exons to produce different Ig isotypes. Each V-gene segment contains the coding sequence of the FWRs 1-3 and the N-terminus sequence of the CDR3, while the J gene segments encode the C-terminus sequence of the CDR3 and the FWR4. The V-domains of the light chains are encoded by VJ rearrangement while those of the heavy chains additionally contain D_H segments between the V_H and J_H segments (VDJ) which encode the middle part of the V_H CDR3 (**Figure 3**).

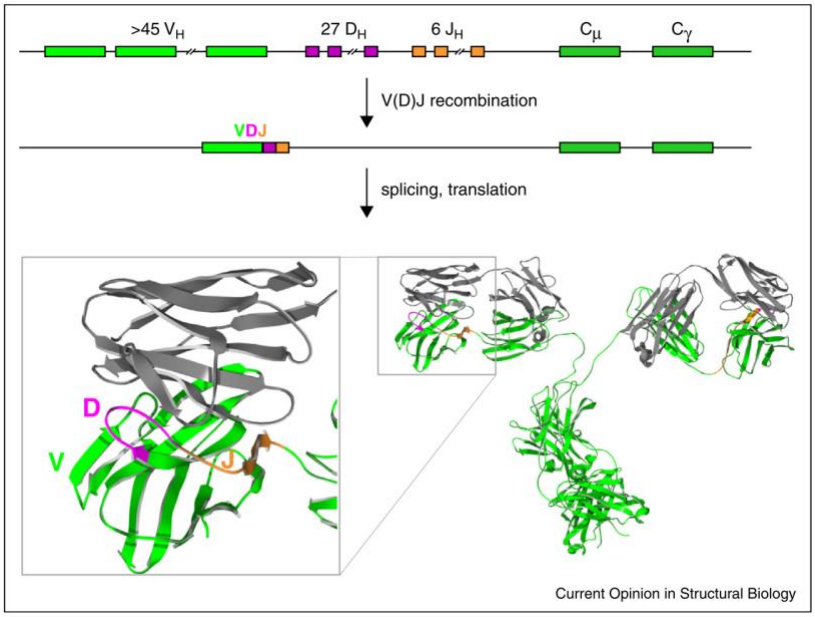


Figure 3. Combinatorial diversity of the V_H-gene derives from VDJ somatic recombination of diverse V_H, D_H and J_H gene segments, which encode the regions FWRs 1-3 to N-terminus CDR3, middle CDR3 and C-terminus CDR3 to FWR4, respectively.

The diversity generated by the process of V(D)J recombination is referred to as combinatorial diversity (**Figures 3 and 4**). The H-chain locus is composed of 80 V_H segments grouped into 7 families of which 40 are functional, 27 D_H and 6 J_H segments. Combination of these segments can produce around 10^4 different heavy chain combinations (H. W. Schroeder, 2006). The light chain loci can produce much less, only around 300 combinations. The κ locus contains 75Vκ and 5Jκ, the Vκ segments belong to 6 different gene families but only 30 of them are known to be functional. The λ locus contains up to 36 functional Vλ segments divided into 3 families and 4 Jλ genes. Strong biases in the usage of specific gene segments and higher frequency of recombination events between certain gene segment have been shown by sanger (Brezinschek, Foster, Dörner, Brezinschek, & Lipsky, 1998; de Wildt, Hoet, van Venrooij, Tomlinson, & Winter, 1999) and next generation (Ig-seq) (Arnaout et al., 2011) sequencing of human naive and memory B cells. For example, 50% of the rearranged genes are composed of only five of the V_H segments (Glanville et al., 2009). Variation of the Ig loci between individuals leads to unique repertoires between individuals (Boyd et al., 2010; Kidd et al., 2012), with differences mostly found in the V_H CDR3, while light chain variants are shared.

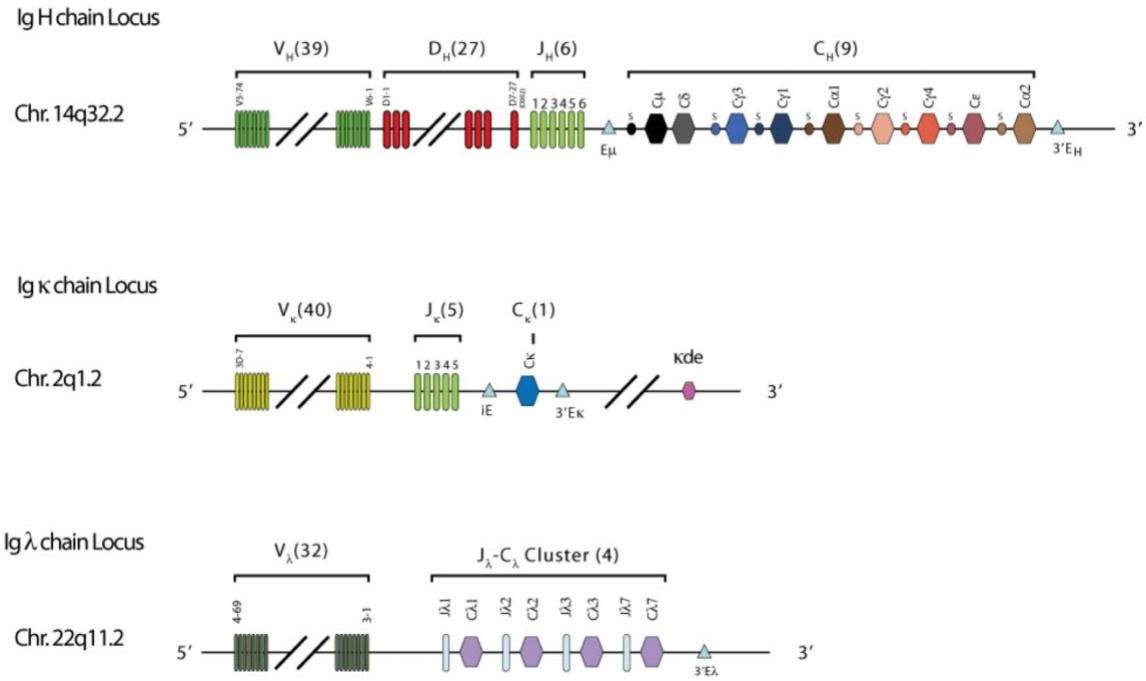


Figure 4. Organization of the V-gene loci. The heavy (Ig H) and light (Ig κ and Ig λ) chain loci are comprised of V, (D) and J gene segments (H. W. J. Schroeder & Cavacini, 2010).

The process of somatic recombination also creates diversity by introduction of random lengths of nucleotides deletions or insertions (N and P) at the junctions of the joined gene segments, the so called junctional diversity. The recombination is performed by RAG recombinases guided by recombination signal sequences (RRS) on the borders of gene segments (Schatz & Ji, 2011). A double-stranded break induced by RAG recombinase cleavage causes the formation of hairpins ends. During the repairing process of the DNA breaks there is loss or gain of nucleotides at the coding ends of the gene segments (junctions). The nucleotide insertions are caused either while filling the gaps of palindromic sequences (P-nucleotides) or by non-template addition of nucleotides (N-nucleotides) by the enzyme terminal deoxynucleotidyl transferase (TdT) (Alt & Baltimore, 1982).

The V_H CDR3 is the most diverse region in length and sequence determining the antibody specificity (Xu & Davis, 2000). Although there are around 10^4 different VDJ combinations, the large diversity of the V_H CDR3 is generated by the junctional diversity (**Figure 5**), that

could theoretically generate more than 1 x 10^7 heavy chain variants (H. W. Schroeder, 2006). Whereas the theoretical V_H CDR3 diversity that can be generated by the random process of somatic recombination is huge, the actual V_H CDR3 diversity is restricted by factors like the amino acid preferences, charge, length and hydrophobicity of the V_H CDR3 loop (Georgiou et al., 2014).

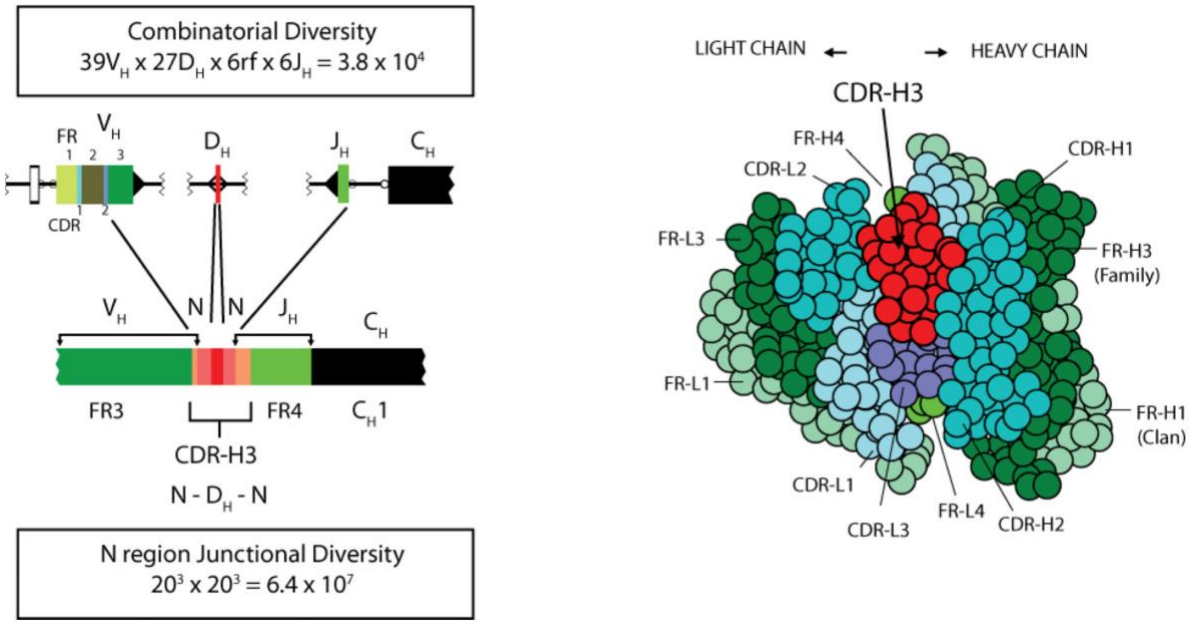


Figure 5. Generation of diversity in the V_H CDR3. Starting from diverse VDJ rearranged genes the diversity is further amplified by introduction of length variation at the junctions of the gene segments (left panel). The antigen-binding site that is formed mainly by the V_H CDR3 (red) (right panel) forms the most diverse part of the antigen combining site enabling huge number of antigen specificities Adapted from (H. W. Schroeder, 2006).

Overall, the primary antibody repertoire is defined by the diversity of antibody gene segments, combinatorial diversity by somatic recombination and junctional diversity. Adding the random combination of L and H chains can result a repertoire of more than 10^{13} different antibody variants in humans (Schroeder, 2006). Accurate estimation of the actual antibody repertoire diversity by high throughput sequencing (Ig-seq) is technically challenging because of limitations of the sequencing technologies (Khan et al., 2016), eg. sequencing

errors, the huge number of B-cells (more than 1 x 10^9 in human peripheral blood) and the fact that only 2% of the total B-cells reside in the peripheral blood (Georgiou et al., 2014).

Other species

Different mechanisms of diversification of the primary antibody repertoire exist in other species. Chickens, unlike humans and mice, do not contain multiple V_H, D_H and J_H gene segments but a single V-gene. Diversity is generated by recombination using multiple V_H-region pseudogenes (~ 50) by a process called gene conversion (**Figure 6**) and further insertions in the D_H-region (Parvari et al., 1988). Interestingly, this process is performed by the enzyme activation-induced deaminase (AID) which in other species executes the process of somatic hypermutation (SHM) (Arakawa, 2002). Huge genetic diversity of the CDRs is generated by multiple gene conversion events (L. Wu et al., 2012).

Additionally to a small classical repertoire, cows possess an unusually long V_H CDR3 repertoire that is encoded by a single V_H (V_H BUL), D_H2 and J_H combination and can exceed the length of sixty amino acids (F. Wang et al., 2013). The diversity of these antibodies is concentrated in the "knob" mini-domain, by diversification of the disulphide patterns using the process of SHM (**Figure 6**).

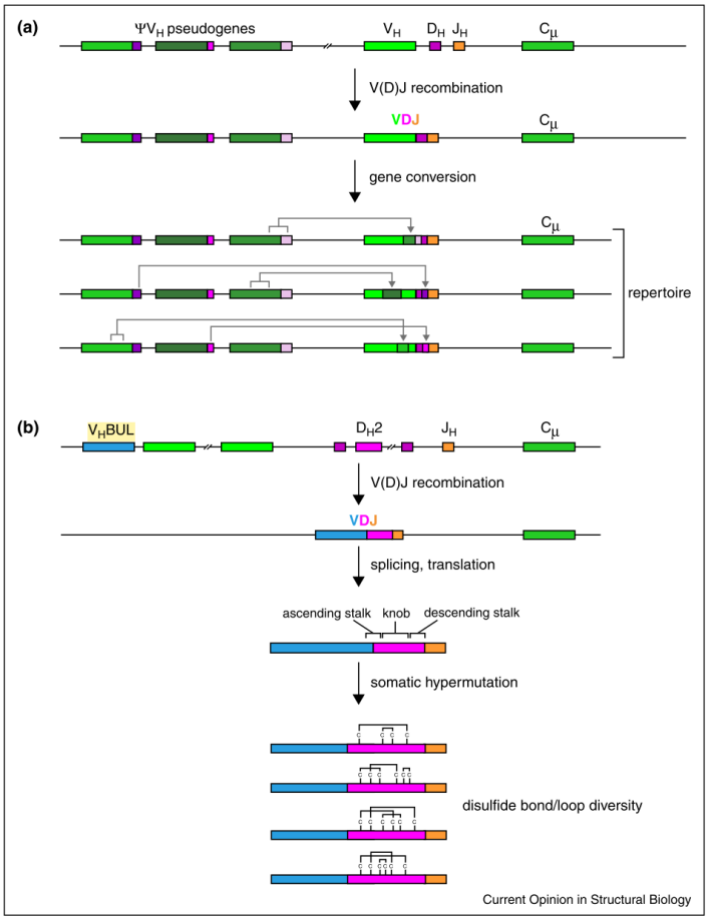


Figure 6. Different mechanisms of genetic diversity generation in chickens and cows. (a) Chickens use gene conversion to produce diverse V-genes from a single VDJ combination, while (b) cows diversify the region encoding a mini-domain ("knob") by diversification of the cysteine pattern of a single rearranged gene through SHM. Adapted from (de los Rios, Criscitiello, & Smider, 2015).

1.1.1.2. Structural diversity

Humans and mice

The structural diversity of the antigen combining site (paratope), which is formed by the CDR loops (**Figure 7**), is accountable for binding to a large diversity of antigenic determinants (epitopes). The topography of the antigen-binding site differs between antibodies binding different sizes of antigens. Binding surfaces that recognize proteins are flat, those that bind peptides form a groove, while hapten binding surfaces use a deep binding site with contact residues buried in the V_H:V_L interface (Almagro, 2004, Raghunathan et al., 2012).

Despite the great diversity of the hypervariable regions, all the CDR loops except from the V_H CDR3 fold into a limited number of main chain conformations, the so called "canonical structures" (Chothia & Lesk, 1987; Al-Lazikani, Lesk, & Chothia, 1997). Recently, comprehensive analysis of a large number of available antigen-antibody complexes redefined the length variation and the diversity of conformations of the hypervariable loops (North, Lehmann, & Dunbrack, 2011). 83% of the non-V_H CDR3 loop conformations of known genetic source (mouse, human, camel or Llama) can be predicted based on the length of the loop and some conserved residues. Analysis of CDR conformations with 28 different lengths (non-V_H CDR3) from mouse, human, camel and Llama identified 72 different loop conformations (clusters) (North, Lehmann, & Dunbrack, 2011), as some CDR lengths can have multiple conformations.

In human and mouse, the large variety of V_L CDR1 loop lengths (10-17 residues) forming a variety of different conformations seems to be correlated with the type of antigen. Short (6–8 residues) loops recognize proteins, while long (11–13 residues) loops peptides and haptens (Collis, Brouwer, & Martin, 2003). The lengths of the other loops are not related to the type of antigen. V_L CDR2 is the most conserved, with either 8 or 12 residues adopting always the same conformation. The V_L CDR3 is quite diverse with 7-13 residues, but in 85% of antibodies it is comprised of 9 amino acids. 70% of the κ-chain V_L CDR3 has a single canonical structure, while the λ-chain V_L CDR3 adopts a wide range of lengths and conformations, of which only a low percentage can be classified as canonical (Chailyan, Marcatili, Cirillo, & Tramontano, 2011). V_H CDR1 is composed of 12-16 residues, with strong bias toward a 13-residue loop (92%), dominated by a single conformation. The V_H CDR2 adopts diverse conformations, composed of 8-12 residues with more common 9- and 10-residue loops.

The V_H CDR3 is the most diverse loop, with high length variation (5-26 amino acids), but most commonly (86%) composed of 7 and 16 residues. The basis of the loop usually adopts specific conformations (Shirai, Kidera, & Nakamura, 1996; Morea, Tramontano, Rustici, Chothia, & Lesk, 1998), but the huge diversity in length and amino acid composition in this loop does not allow the prediction of its overall conformation.

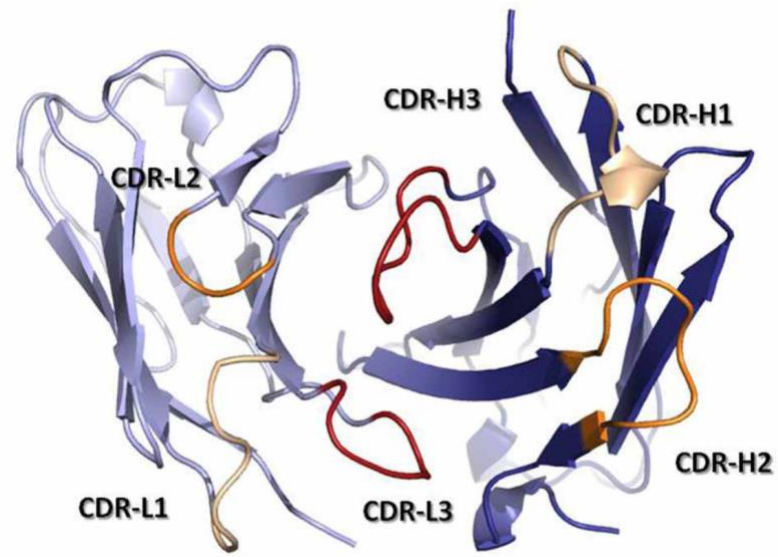


Figure 7. The CDR loops conformations define the antigen-binding site topography (adapted from Finlay & Almagro, 2012).

Taken together, all CDR loops (except from the V_H CDR2) can have considerably different lengths, with some of them varying more than others. The loop length is recognised as the primary factor that defines the topography of the combining site (Al-Lazikani, Lesk, & Chothia, 1997; North, Lehmann, & Dunbrack, 2011), as different CDR lengths can form pockets, grooves or fill in the space to form a flat antibody surface (Vargas-Madrazo, 1995). The combinations of canonical loop structures composing an antigen-binding site are limited (Lara-Ochoa, Almagro, Vargas-Madrazo, & Conrad, 1996). The V_H CDR3 loop mainly defines the topography of the antigen-binding site (Vargas-Madrazo, 1995). A short V_H CDR3 loop can form a cavity to bind peptides, while a long V_H CDR3 loop can from a

finger-like structure that protrude in cavities, like anti-HIV antibodies that penetrate the recessed CD4-binding site of gp120 (Saphire, 2001).

Other species

While the immunoglobulin structure is highly conserved in vertebrates (with exception the jawless fish), some species have evolved immunoglobulins with structurally diverse features (**Figure 8**) (de los Rios, Criscitiello, and Smider 2015). In human and mice, the diversity occurs in all six CDRs, with most variability focused on the CDR3s (Xu & Davis, 2000). Additionally to the classical IgG molecule, the antibody repertoires of sharks (R. L. Stanfield, 2004), camels (Desmyter, Spinelli, Roussel, & Cambillau, 2015) and cows (F. Wang et al., 2013) produce antibodies with unique structural characteristics.

Camels generate antibodies that lack the heavy chain CH1 domain and light chains, referred to as heavy chain antibodies (HCAbs) (Hamers-Casterman et al. 1993). On the course of adaptation from conventional IgG molecules, HCAbs have evolved some unusual characteristics to compensate for the loss of 3 V_L CDRs, like a long V_H CDR3 loop that enable them to interact with epitopes that cannot be accessed by conventional antibodies, eg. enzyme active sites (Flajnik, Deschacht, & Muyldermans, 2011).

Cow antibodies possess unusually long CDR H3 regions with lengths that could exceed the sixty amino acids (F. Wang et al., 2013). The binding site of these antibodies is formed by a "knob" domain placed at the tip of a b-ribbon "stalk" that protrudes far from the traditional antibody paratope site. The diversity of these antibodies is generated by diversification of the disulphide patterns of the "knob" domain. The functional advantage of these antibodies remain unknown, as the types of antigen that are recognised by these antibodies have not been studied yet (Robyn L Stanfield, Wilson, & Smider, 2016). These examples demonstrate

novel structural characteristics that evolved from conventional antibodies to expand the typical repertoire diversity.

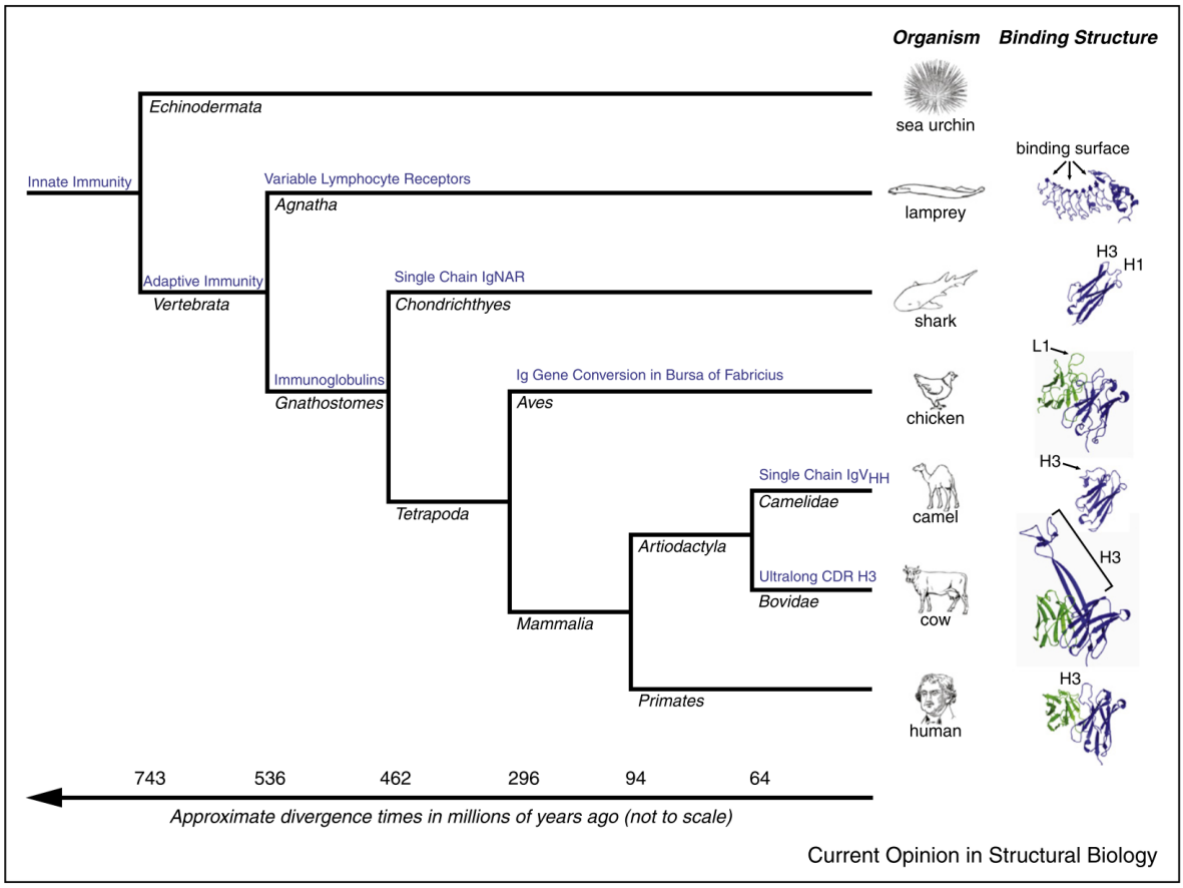


Figure 8. Evolution of immunoglobulins (Ig) with structurally unusual characteristics in different species. Sharks and camelids convergently evolved heavy chain antibodies without light chains. Camelids often have long V_H CDR3 with disulphide bonds. Cows have evolved some antibodies with ultralong V_H CDR3. The binding site of the antibodies is formed by a disulphide-rich mini-domain ("knob") which protrudes far from the typical antibody binding site. Diversity of the knob is generated by diversification of the disulphide patterns. (Right) The diverse structures are shown. Adapted from (de los Rios et al., 2015).

1.1.2. Secondary antibody repertoire

Somatic hypermutation (SHM)

The primary BCR repertoire expressed on the surface of B-cells is capable of recognizing an infinite number of antigenic determinants with low affinity. Encounter of B-cells to their cognate antigen induces a series of events that lead to BCR affinity maturation and finally B-cell differentiation into memory B-cells and plasma cells that produce soluble antibodies. Low affinity B-cells undergo clonal expansion, somatic hypermutation (SHM) and selection of clones that express BCR variants with higher affinity in the germinal centres (GC) (**Figure 9**), which are specialized structures in secondary lymphoid organs (e.g., spleen, lymph nodes), (McHeyzer-Williams, Okitsu, Wang, & McHeyzer-Williams, 2011). SHM involves iterative rounds of diversification of the rearranged V-gene and affinity-based selection of B-cell clones by competition for binding to the antigen that leads to slow elimination of low affinity variants and survival of the higher affinity variants (Tas et al., 2016). Class switch recombination (CSR) that allows the switch from IgM to IgG, IgA, or IgE also occurs in the GC.

There is a biological limit in antibody affinity maturation by SHM, as antibodies with K_D values below a certain threshold do not have further selective advantage (T. R. Poulsen, Meijer, Jensen, Nielsen, & Andersen, 2007; Tine Rugh Poulsen, Jensen, Haurum, & Andersen, 2011). Theoretical analysis of the limits of antibody affinity for any antigen has predicted that the lowest possible K_D value is around $\approx 10^{-10}$ M ($\approx 10^5 - 10^6$ $M^{-1}s^{-1}$/ $\approx 10^{-4}$ s^{-1}) (Foote & Eisent, 1995). The maximum on-rate constant ($10^5 - 10^6$ $M^{-1}s^{-1}$) is restricted by the diffusion coefficients of the antibody and antigen molecules, while the factor limiting the off-rate ($\approx 10^{-4}$ s^{-1}) is the BCR internalization after binding to an antigen. Studies of the on- and off-rates of vaccinated individuals indicate that the limit is K_D of $\approx 10^{-9}$ M, as boost vaccination induced only a small degree of further V-gene sequence diversification without

further increase in affinity (Tine Rugh Poulsen, Jensen, Haurum, & Andersen, 2011). Experimental data on the correlation between the B-cell antigen presentation and the affinity of the BCR-antigen interaction is in agreement with B-cells responding to the antigen only when the affinity of the interaction is between K_D values of 10^{-6} M to 10^{-10} M (Batista & Neuberger, 1998).

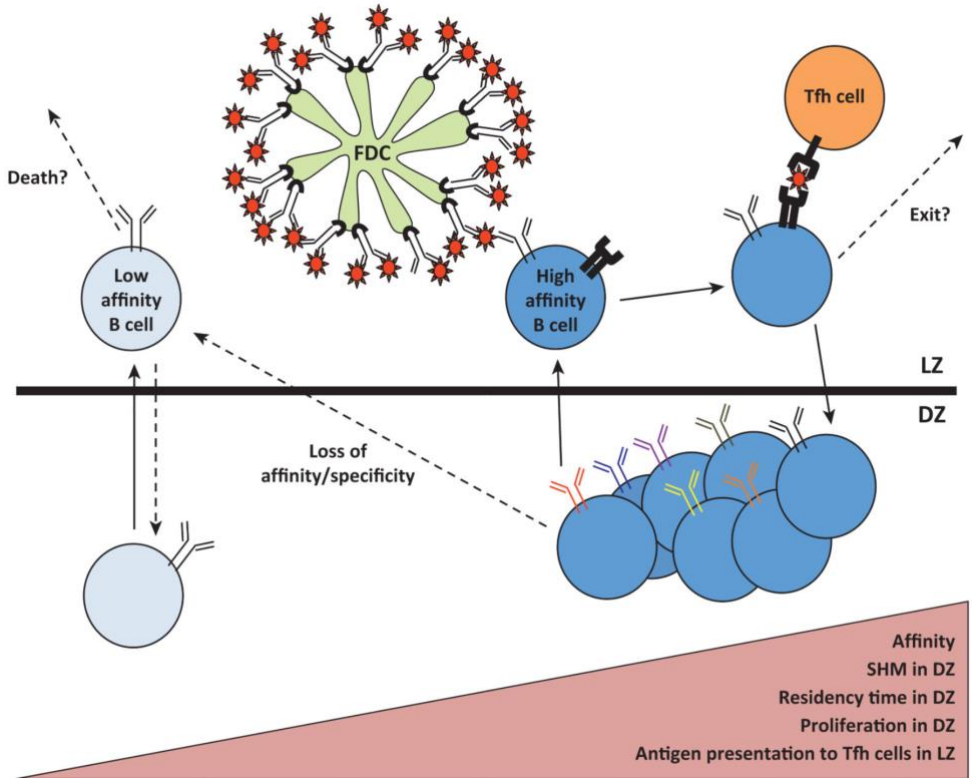


Figure 9. Affinity maturation of B-cells by SHM occurs in the GCs. GCs are divided in two regions known as dark zone (DZ) and light zone (LZ) (Victora & Nussenzweig, 2012). Clonal expansion and SHM occurs in the DZ, while the highest affinity B-cells are selected in the LZ, where the antigen is presented by follicular dendritic cells (FDCs). The clones that bind to the antigen then present it to T follicular helper (Tfh) cells which direct them to the DZ, while B-cells that cannot compete for binding to the antigen do not enter the DZ. The selected B-cells undergo more rounds of proliferation and SHM. Adapted from (Oropallo & Cerutti, 2014).

1.1.2.1. Genetic diversity

Point substitutions

Sequence diversification by SHM is achieved mainly by introduction of point mutations, but also insertions and deletions (indels). High mutation rate in the V-genes is induced by high transcription levels of the corresponding genes (Bachl, Carlson, Gray-Schopfer, Dessing, & Olsson, 2001). Hypermutation targets the rearranged V-region and a part of the 3'-intron but not the constant region (1-2kb downstream the transcription start site). Somatic point mutations are not introduced in a stochastic manner, but high frequency of mutations accumulate in hotspot positions across the V-genes. The hotspot motif of SHM is the sequence DGYW (D=A/G/T; Y=C/T, W=A/T) (Rogozin & Diaz, 2004).

Activation-induced cytidine deaminase (AID) is the enzyme responsible for SHM (Revy et al., 2000) and class switch recombination (CSR). The same enzyme is implicated in homologous recombination of V-genes during gene conversion (Arakawa, 2002). While AID is expressed mainly in B-cells while subjected to SHM, overexpression of this enzyme in other cell lines generates high frequency of mutations (Yoshikawa, 2002; Bowers et al., 2011). The process starts by cytidine deamination on single-stranded DNA (ssDNA) (Pham, Bransteitter, Petruska, & Goodman, 2003; Shen, Ratnam, & Storb, 2005) that results in dU-dG mismatch. The mismatch can lead to different mutational outcomes generating mutations with a bias to transitions over transversions (**Figure 10**) (Neuberger, Harris, Di Noia, & Petersen-Mahrt, 2003). These mutations are generated by pathways that normally repair such mismatches, like the mismatch repair pathway (MMR), but act in an error-prone mode during SHM.

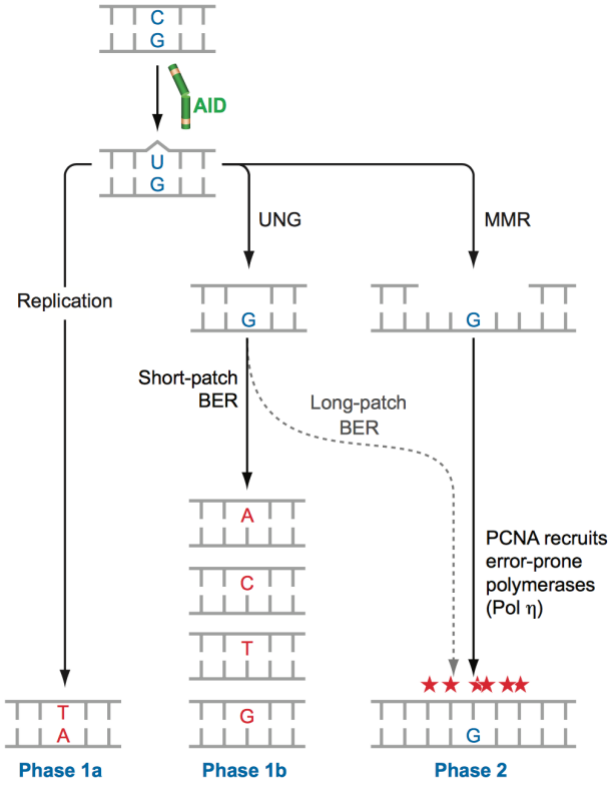


Figure 10. Possible mutational outcomes of cytidine deamination by AID. Phase 1a: If DNA replication occurs, the dU-dG mismatch will become either dC-dG or dT-dA (because A will be added opposite to U), resulting C>T and G>A transition mutations. Phase 1b: if an uracil-DNA glycosylase (UNG) removes dU, all mutations (dA, dT, dC or dG) can result by DNA synthesis opposite to the empty site, resulting in both transition and transversion mutations. Phase 2: dA-dT mutations also occur during SHM, by the mismatch repair (MMR) which generates dA-dT mutations near the dC-dU lesion. Gene conversion in other organisms (chicken or rabbit) starts when the dU-dG mismatch results in recombination with one of the Ig V pseudogenes.

SHM targets preferentially the CDRs than the FWRs of the V-genes according to study of V-genes with frameshifts, consequently rearranged genes that were subjected to somatic hypermutation but were not selected based on affinity (Dörner et al., 1997), but the reason why the same DGYW motifs are targeted more frequently in CDRs is a subject of investigation (Dörner et al., 1997; Wei et al., 2015; Yeap et al., 2015). SHM targets predominantly Ig V-genes, while the DGYW motifs are only partially more enriched in V-genes than in the rest of the genome, indicating that the presence of hotspot motifs is not the only factor determining AID targeting sequences (Di Noia & Neuberger, 2007). Additionally, non-Ig genes controlled by Ig promoters can undergo as high SHM frequencies as the CDRs

of V-genes at positions containing AID motifs (Yélamos et al., 1995). These observations have led to the conclusion that local sequences in CDRs have evolved to direct SHMs preferentially in CDRs than FWRs (Yeap et al., 2015).

Indels

Additionally to point substitutions, the mechanism of SHM introduces a low percentage of in-frame indels. Sequencing analysis of rearranged V_H-genes amplified from DNA of naïve and GC single sorted B-cells by flow cytometry (Klein et al., 1998, Goossens et al., 1998) has shown that V_H-genes of GC B-cells, which have been subjected to SHM but not necessarily selected for improved affinity, but not naive B-cells contain around 4% in-frame and 43% out-of-frame indels, indicating that the highest percentage of indels introduced by SHM are out-of-frame. In 28 out-of-frame V_H-genes they found 20 indels and 319 point mutations, thus only a low percentage (6%) of all mutations were indels, suggesting that the percentage of indels introduced through SHM is very low compared to that of point substitutions. Frequent occurrence of deletions due to SHM was observed in positions with high SHM frequency (Yeap et al., 2015).

Analysis of in-frame indels in V-genes of B-cells that have been subjected to selection by *in vivo* (de Wildt, van Venrooij, Winter, Hoet, & Tomlinson, 1999) or *in vitro* (Bowers et al., 2014) SHM indicate that indel sequences are related to adjacent sequences. This is because insertions occur by duplication of adjacent sequences while deletions by removal of repeated sequences. This observation suggests that indels could be created according to the DNA misalignment model (**Figure 11**) (de Wildt, van Venrooij, Winter, Hoet, & Tomlinson; 1999). The higher percentage of repeats in CDRs compared to FWRs in germline V-gene segments is likely to indicate hotspots for indels (de Wildt, van Venrooij, Winter, Hoet, & Tomlinson, 1999).

In the same study, it has been proposed that most indels share the same hotspots with point mutations as indel duplications or deletions involve AGY motifs (de Wildt et al., 1999). A high percentage of point substitutions that accumulate following the occurence of the insertion that further diversify the inserted sequence has been proposed to be linked either with a shared mechanism of generating indels and point substitutions (de Wildt et al., 1999) or with optimization of the binding paratope during affinity maturation (Bowers et al., 2014).

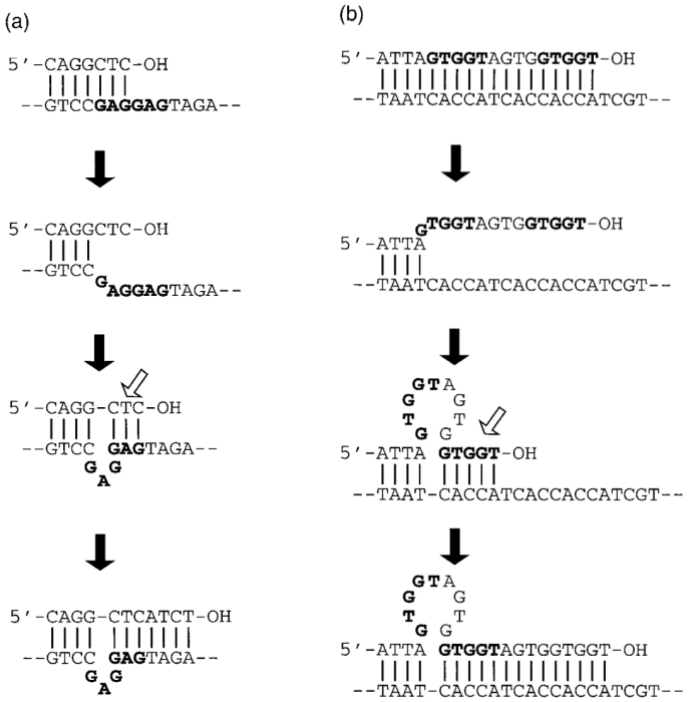


Figure 11. (a) Deletions and (b) insertions could be generated based on the DNA misalignment model. A repeated sequence (in bold) creates a loop because of misalignment. Subsequent extension of the DNA as if it was normally aligned leads to a deletion or an insertion depending on which DNA strand forms the loop. Adapted from (de Wildt et al., 1999).

Affinity matured antibodies

Affinity matured V-genes contain on average 15-20 mutations across the entire variable sequences with an average 18.0 ± 8.1, 9.7 ± 5.4, and 7.7 ± 4.0 mutations for V_H, κ and λ V_L genes, respectively (Tomlinson et al., 1996; Tiller et al., 2007). The percentage of point mutations and indels in V-genes of antibodies that have been selected by SHM has been

determined by sequencing the V-genes from a total of 365 IgG+ B-cells derived from peripheral blood (de Wildt, van Venrooij, et al., 1999; de Wildt, Hoet, van Venrooij, Tomlinson, & Winter, 1999).

In addition to high percentage of point substitutions (10.9% and 7.6% in the V_H and V_L, respectively), affinity matured antibodies also contained indels in 6.5% of the antibodies analyzed (24/365) (de Wildt, van Venrooij, et al., 1999). Similar percentage of in-frame indels (≈1.9% insertions and ≈2% deletions) has been estimated by high throughput sequencing of IgM and IgG memory B-cells isolated by flow cytometry (Briney, Willis, & Crowe, 2012) and in B-cells from patients with rheumatoid arthritis (≈1.4% indels) (Reason & Zhou, 2006). Overall, it has been estimated that only a low percentage, 1.4 – 6.5% of affinity matured antibodies contain indels. The percentage of indels in affinity matured antibodies could be underestimated, because indels in the V_H CDR3 and at the end of the V_L CDR3 cannot be identified due to length variation that is introduced at these positions (junctions) during the initial V(D)J rearrangement. The CDR3 loops can tolerate high length variation, thus it is possible that further length diversification through SHM could contribute to antibody affinity maturation. All these studies agree that the location of indels is mainly in or in close proximity to the CDRs.

Extensively affinity matured antibodies

Most antibodies that have undergone SHM contain 15-20 mutations (Tiller et al., 2007), except from potent broadly neutralizing monoclonal antibodies (bnAbs) to type 1 immunodeficiency virus (HIV-1) that accumulate 40-110 mutations (F. Klein et al., 2013). bnAbs develop only in some patients 2–3 years after infection (Mikell et al., 2011) and contain higher percentage of somatic mutations in comparison to other antibodies isolated

from the same patients (Scheid et al., 2009; Xiao et al., 2009). These mutations are essential for neutralization potency and breadth (Scheid et al., 2009; Xiao et al., 2009).

The reason why such a large number of mutations is required for broad neutralization activity has been studied using the VRC01 class bnAbs, which neutralize HIV-1 by binding to the CD4 site of the gp120 envelop (Kwong & Mascola, 2012). All these antibodies use the same V_H segment (V_H1-2*02) and contain mutations in up to 32% rearranged V_H-gene. Only a fraction of all these mutations in mature bnAbs is required for broad neutralization activity (Zhou et al., 2010; Burton & Mascola, 2015). Recent study has revealed that some mutations critical for neutralization activity have low intrinsic mutation potential, supporting the idea that extensive mutagenesis and selection by SHM needs to occur to achieve targeting of these residues (Hwang et al., 2017).

Analysis of HIV-1 bnAbs estimated that high percentage of the reported bnAbs contain indels (≈40%) acquired during SHM with lengths ranging from 1 to 11 amino acids (Kepler et al., 2014). 27 unique in-frame indels were identified in 56 HIV-1 bnAbs (108 V-genes), thus 25% of the total bnAb V-genes contain indels. The unusually high frequency of indels in bnAbs is due to high levels of SHM which leads not only to accumulation of high percentage of point mutations but also to other more substantial sequence modifications, like indels.

An essential role of insertions in neutralization activity has been demonstrated in some cases. Measurements of the neutralization activity of members of the anti-CD4 bnAb clonal lineage CH31 before and after the occurrence of a 9-amino acid insertion in the CDR1 loop has revealed that broad neutralization activity was acquired only after the occurrence of the insertion (Kepler et al., 2014). Additionally, removal of the FWR3 insertion of another anti-CD4 bnAb (3BNC60) abolished the neutralization activity (Klein et al., 2013). Taken together, high percentage of insertions in bnAbs that occur by extensive SHM seem to have a functional role in broad neutralization activity.

1.1.2.2. Structural basis of affinity maturation

Point substitutions

After selection of antibodies with an antigen-binding topography that confers relatively low affinity, further structural diversification is introduced by SHM. Somatic mutations can be found across the entire V-genes, with the highest percentage of mutations at the antigen combining site, but also on the surface of the variable domains and the V_H: V_L interface (Clark, Ganesan, Papp, & van Vlijmen, 2006). SHM does not change residues that form direct contacts with the antigen, but residues adjacent to those residues. While the diversity in the primary repertoire is concentrated in the centre of the combining site, mutations introduced by SHM are located in the periphery of the binding site (Tomlinson et al., 1996), indicating that the diversity generated by SHM is complementary to that of the primary repertoire diversity.

Analysis of crystal structures of antibody-antigen complexes has shown that SHM improves the shape and molecular complementarity (additional inter-molecular hydrogen bonds, van der Waals contacts) of the antigen combining site to the epitope (Cauerhff, Goldbaum, & Braden, 2004). Additionally to improving the shape complementarity, improved binding has been shown to be achieved by increasing the buried interface with antigen, which is the result of subtle structural rearrangements in the periphery of the interface with antigen (Sundberg et al., 2000; Li, Li, Yang, Smith-Gill, & Mariuzza, 2003).

In addition, increased binding can be attributed to changes in the entropic cost of complex formation. Starting from an antibody that can adopt several combining site conformations (flexible), affinity maturation stabilises the conformation that is beneficial for binding, which is a conformation with smaller conformational change upon binding (paratope preconfiguration) (Wedemayer, 1997; Manivel, Sahoo, Salunke, & Rao, 2000; Schmidt et al., 2013).

Insertions

Analysis of the sites of indels on structures of bnAb-antigen complexes has shown that these are preferentially located close to the antigen-binding site, indicating that indels probably improve interactions with the antigen (Kepler et al., 2014). Co-crystal structure of NIH45-46 with gp120, a clonal variant of the anti-CD4 bnAb VRC01 with a four-residue insertion in the V_H CDR3, has shown that the insertion extends the V_H CDR3 loop introducing new contacts with the antigen, which result an extended buried interface area with the gp120 and enhanced neutralization (X. Wu et al., 2011). Superimposition of the crystal structure of 3BNC60, a clonal variant of NIH45-46 with an insertion in the FWR3, with the co-crystal structure of NIH45-46 with gp120 suggested that the insertion might enhance binding by interacting with a loop region in the gp120 that was truncated from the initial construct (F. Klein et al., 2013). The role of an insertion on broad neutralization was well illustrated on the complex structure of 8ANC195 with gp120 (Scharf et al., 2014). The insertion in the FWR3 was shown to extend the epitope surface allowing interaction with a pocket on gp120 surface that conferred broad neutralization activity. Taken together, the mechanism by which insertions seem to confer broad neutralization activity is by extending the antigen-binding site surface allowing targeting of new surface epitopes on gp120.

1.2. Technologies for discovery of therapeutic monoclonal antibodies

Monoclonal antibodies (mAbs), known for recognizing a single epitope with high affinity and exquisite specificity, are widely used as therapeutics for cancer, autoimmune and infectious diseases, transplantation and other medical conditions (Chan & Carter, 2010; Weiner, 2015). The first great innovation that allowed rapid development of the therapeutic antibody technology was the invention of hybridoma technology, which is still widely used to produce

mouse monoclonal antibodies with defined specificity (Köhler & Milstein, 1975). Before clinical use, the immunogenicity of antibodies derived from animal sources is reduced by the process of "humanization", which is usually performed by grafting the CDRs of a mouse antibody into human V-genes (P. T. Jones, Dear, Foote, Neuberger, & Winter, 1986). Development of technologies for generation of fully human antibodies allowed the wide use of mAbs as therapeutics of human disease (Lonberg, 2008).

Therapeutic antibodies exert their effect by different mechanisms (Chan & Carter, 2010) including binding to a ligand or their cognate receptor to either block their interaction and consequently prevent receptor activation and function (e.g., infliximab, adalimumab) or downregulate the expression of the receptor (e.g., efalizumab, omalizumab). Another important mechanism of action of therapeutic antibodies is the target cell lysis by induction of effector functions (like ADCC, ADCP and CDC) through the interaction of the Fc region with cellular receptors (Melis et al., 2015). More recently, two immune checkpoint inhibitors, the cytotoxic T-lymphocyte antigen 4 (CTLA-4) and the programmed death 1 (PD-1) were targeted to render tumor immunogenic in order to fight cancer through the patient's immune system (Couzin-Frankel, 2013).

Engineered antibody fragments

Genetic engineering allows the production of diverse antibody fragments, like fragments lacking the Fc region, with reduced size or with two different binding sites (bi-specifics). Most therapeutic applications require the Fc fragment of the IgG to increase the FcRn-mediated recycling for extended antibody half-life and to recruit effector functions, but there are some applications where the Fc fragment is not required. For example, presence of the Fc can induce massive cytokine release and associated toxic effects (Holliger & Hudson, 2005). Reduced size could have several benefits including improved pharmacokinetics for tissue

penetration (Holliger & Hudson, 2005) and bi-specific formats (more than 60) offer the possibility to either interact with two different disease targets or redirect T-cells to kill tumor cells (Spiess, Zhai, & Carter, 2015).

The most commonly used antibody fragments are the fragment antigen-binding (Fab) and the single-chain fragment variable (scFv). The Fab fragment (50kDa) has one-third of the size of an IgG molecule (150kDa) and is one of the formats that are being explored for therapeutic applications, in fact these are the most successful fragments accounting for 49% of the fragments that have entered clinical development with three FDA approved fragments (Nelson, 2010). The single-chain variable fragment (scFv) is composed of the variable domains of the heavy and light chain linked by a flexible linker (Skerra & Pluckthun, 1988). The scFvs represent a large percentage (40%) of candidates in clinical development (Nelson, 2010).

1.2.1. Isolation of antibodies with defined specificity

The discovery of therapeutic antibodies with defined specificity against a wide variety of target proteins is generally possible due to the development of appropriate technologies. The classical hybridoma technology has some restrictions, like inability to produce antibodies against antigens that are toxic to the animal and conserved across species (non-immunogenic) (Dufner, Jermutus, & Minter, 2006). Additionally, although K_D values lower than 100 pM have been reported (Rathanaswami et al., 2005), there is an affinity ceiling in the K_D values that can be achieved by the BCR affinity maturation process *in vivo*.

A key technology that allows isolation of specific antibodies from large libraries of functional antibodies is phage display (Smith, 1985), which was adapted to display "fully human" antibody repertoires for selection *in vitro* (McCafferty, Griffiths, Winter, & Chiswell, 1990).

The past decades these technologies were exploited to generate human antibodies with high affinity and specificity (Hoogenboom, 2005). The first mAb that derived from phage displayed antibody libraries was Adalimumab (Humira), an anti-TNF antibody for the treatment of rheumatoid arthritis discovered by Cambridge Antibody Technology (Brekke & Sandlie, 2003).

1.2.1.1. Construction of antibody libraries

Libraries displaying fully human antibody fragments (Fab or scFv) on phage have been constructed by cloning the V-genes of the total B-cell repertoire by RT-PCR (Marks et al., 1991), the so called naïve libraries, which have been used to isolate antibodies with affinities down to the sub-nanomolar range (Winter, Griffiths, Hawkins, & Hoogenboom, 1994; Vaughan et al., 1996; P Holliger & Hoogenboom, 1998). Synthetic libraries are usually made starting by a restricted number of antibody germline genes, eg. one V_H-gene (Griffiths et al., 1994; Viti, Nilsson, Demartis, Huber, & Neri, 2000) or 49 V_H-genes (Hoogenboom & Winter, 1992), by diversification of small number of V_H or V_L CDR3 residues. The fully synthetic human combinatorial antibody libraries (HuCAL) were developed in an effort to mimic the diversity of the human repertoire. The library design was based on 7 V_H and 7 V_L-germline genes that represent each gene family, so starting from 49 V_H:V_L gene combinations diversity was introduced in the CDR3s by mimicking the natural diversity, including length diversity (Knappik et al., 2000). All these synthetic libraries have been used successfully for isolation of antibodies to haptens, proteins and peptides.

1.2.1.2. Selection of specific antibodies

From antibody libraries that typically contain around 10^9-10^{10} members, antibodies that bind specifically to the desired antigen are isolated by *in vitro* selection. Selection is possible

because each phage particle displaying a specific antibody carries the gene encoding it, which can be recovered after selection of the antibody for binding to the desired antigen.

Phage display

This link between the phenotype and the genotype is achieved by displaying an antibody fragment (Fab or scFv) on the surface of filamentous bacteriophage by fusion to the phage coat protein pIII. The filamentous bacteriophage M13, a single-stranded DNA virus, is the most commonly used phage for display, because it can infect *E. coli* containing the F-pilus without causing cell lysis. The pIII phusion protein (pIII-scFv) is encoded by the phagemid, a plasmid with a phage origin of replication and an antibiotic resistance gene (Winter, Griffiths, Hawkins, & Hoogenboom, 1994). A phagemid can be replicated as a plasmid, and also be packaged as single stranded DNA in viral particles when it is "rescued" by a helper phage which provides the proteins necessary for particle assembly.

Biopanning

The selection process called "biopanning" involves iterative rounds of production of phage particles displaying the scFv library in *E. coli*, incubation of the library with the target antigen, washing to remove the variants that do not bind, elution of the particles that bind the antigen and infection of *E. coli* cells with eluted phage (**Figure 12**).

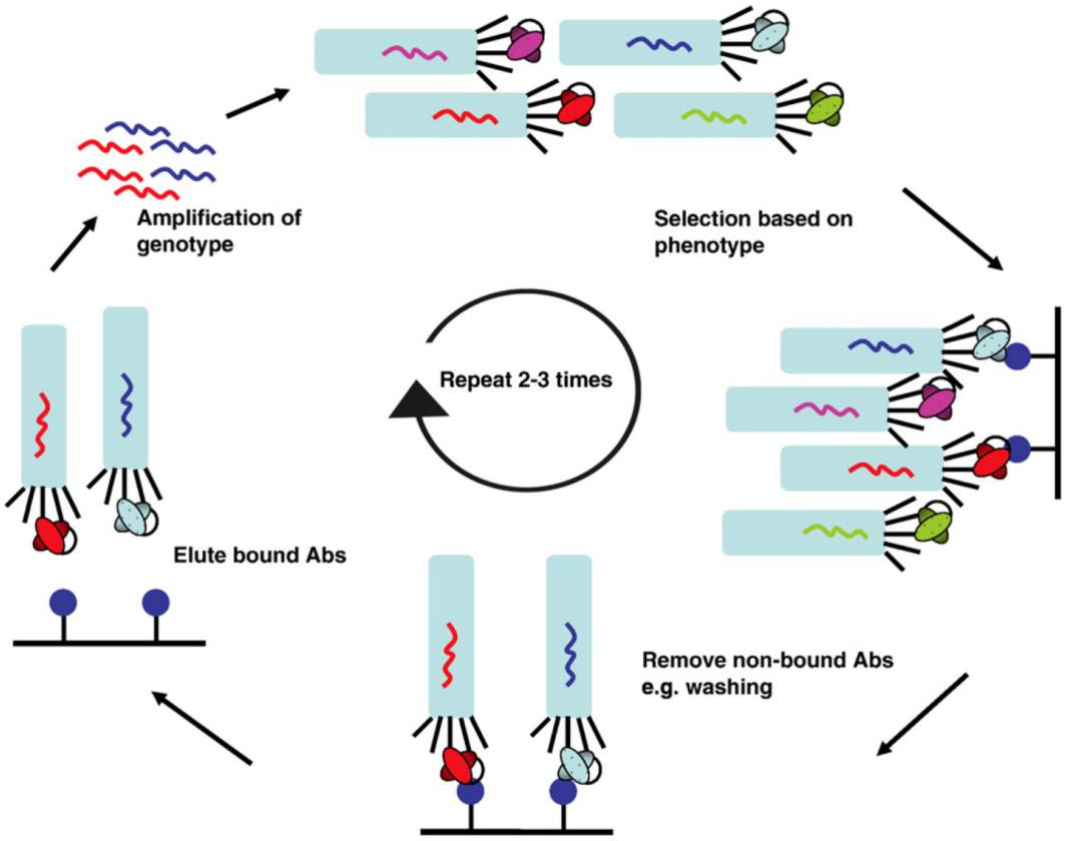


Figure 12. Schematic representation of the biopanning process demonstrated with the phage display method. In phage display, the link between the genotype and phenotype is created by producing phage particles that contain the gene encoding the scFv variants and the scFv fused to the pIII minor coat protein. The phage library is used for selection for binding to the target antigen and particles expressing scFv variants that bind the target remain bound to the antigen after washing. The genotype of the selected variants is recovered by phage elution and *E.coli* infection. This process is repeated a number of times to enrich a pool of antigen-binding scFv variants. Adapted from (Weisser & Hall, 2009). (Weisser & Hall, 2009)

1.2.2. *In vitro* antibody affinity maturation

Discovery of a therapeutic antibody candidate for a disease target is usually followed by improving antibody affinity *in vitro*. The highest possible affinity for the antigen is desirable to achieve high efficacy even at low concentrations, reduced dosing and cost of the therapy (Dufner et al., 2006). In some cases the best affinity is not clear, for example when targeting

a solid tumor a high affinity mAb may be retained at the periphery of the tumour which could result in reduced tumor penetration (Adams et al., 2001; Weiner & Carter, 2005).

The process of directed evolution of antibody affinity may include iterative rounds of genetic diversification and selection or/and screening for improved variants. Construction of libraries with variants of the selected parent antibody is followed by identification of improved variants by selection strategies similar to those described above. Very high affinities, in the low picomolar (Schier et al., 1996; Razai et al., 2005) or even in the fentomolar (E. T. Boder, Midelfort, & Wittrup, 2000) range can be achieved.

1.2.2.1. Strategies for library construction

Point substitutions

Different strategies have been applied to build libraries with variants of the parent antibody with diversity at random or targeted positions. Random approaches, like error-prone polymerase chain reaction (ep-PCR) (Neylon, 2004) and DNA recombination (Stemmer, 1994), introduce diversity at random positions of the entire V-gene, while focused approaches, like CDR walking mutagenesis (W. P. Yang et al., 1995) and rational design, introduce diversity at restricted number of specific positions.

In Nature, diversity is introduced mainly in specific regions, the CDRs, indicating that sufficient beneficial mutations can be found in these regions (Tomlinson et al., 1996). Focused *in vitro* affinity maturation strategies, mimicking the process in Nature, introduce variation only in the CDRs. By saturation mutagenesis (with all possible amino acids at the position) it is impossible to diversify all six CDRs in one library, as each of them is on average nine residues long and each library (with 10^9-10^{10} variants) can only contain all possible combinations of six to seven residues. Thus, the challenge of focused approaches is to identify which CDR positions to target.

One approach that is commonly used to overcome this limitation is called CDR walking, which involves targeted mutagenesis of CDR regions and selection that can be performed in parallel or sequential (Barbas et al., 1994; W. P. Yang et al., 1995). The parallel approach involves randomization of different CDRs in different libraries, parallel affinity-based selections and subsequent combination of the most beneficial mutations from different CDRs in one library. In the same study, the outcome of the parallel approach was compared to that of a sequential approach, which was conducted by randomization of one CDR and selection followed by picking the best variant to be the template for randomization of another CDR. The results of the parallel approach were unpredictable, as beneficial mutations in different CDRs did not always have an additive effect on affinity, but a high improvement (420-fold) was achieved. Whereas using the sequential approach improved variants with small improvements (4-fold) were obtained consistently in each walk with the highest improvements (8-fold) obtained by the V_H CDR3 walk, that seems to be most rapid route to obtain affinity gains.

Since both the V_H and V_L CDR3 regions are the most diverse CDRs in Nature, with the V_H CDR3 focusing by far the highest diversity (Xu & Davis, 2000), a strategy of targeting only these regions was followed to increase the affinity (Schier et al., 1996). By sequential targeting of the V_L CDR3 and then the V_H CDR3, an improvement of more than three orders of magnitude was achieved. Due to the fact that the CDR walking approaches are time consuming, a large number projects have focused on randomization of the V_H CDR3, which is more likely to yield affinity improvements.

Random mutations across the entire V-gene can be introduced by gene amplification by ep-PCR using low-fidelity DNA polymerases. Comparison of libraries containing different number of mutations per scFv gene (2, 4 or 22 mutations) for improved affinity has shown that all these libraries yielded variants with improved affinity, but the greatest affinities

derived from the libraries with the moderate and high mutation rate (Daugherty, Chen, Iverson, & Georgiou, 2000). Interestingly, a high percentage of variants were still functional in the libraries with the high mutation frequency and the majority of beneficial mutations were located in residues distant from the binding site.

The effect of DNA recombination on antibody affinity have been assessed by comparing the outcome of selection for libraries with error-prone PCR with and without a recombination step. The outcome of selection of the recombined library was a population with higher percentage of improved variants including the variant with the highest affinity, indicating that combinatorial library strategies allow sampling of larger sequence space than ep-PCR.

A semi-rational approach, using ribosome display combined with ep-PCR mutagenesis between rounds of selection has been used to identify position (hotspots) that could improve affinity across the entire scFv of an antibody specific for IL-13 (BAK1) (Thom et al., 2006). Identification of mutational hotspots was followed by construction of libraries targeting these positions with the full diversity of amino acid substitutions. These libraries also contained combinations of beneficial mutations in different CDRs to take advantage of additive effects on affinity. The majority of beneficial point substitutions was not found in the CDR3s but in the V_L CDR1, underlining the importance of targeting all CDRs to find mutations that confer affinity improvements. Also, the majority of mutation hotspots (84%) were located in the CDRs, which is in agreement with preferential selection of CDR mutations in human antibodies in Nature.

A disadvantage of the ep-PCR approaches is that these are biased towards conservative substitutions because more than one mutations per codon are rarely introduced, thus sampling of all amino acids in each position is impossible. More recently, a more systematic approach was implemented to study the impact of each mutation across the entire V-gene in affinity, by construction of a single-site saturated mutagenesis library (Koenig et al., 2017). Positions that

enhanced affinity were identified by determining the frequency of each mutation in the library before and after selection by deep sequencing. A variant with improved affinity and stability was identified, with mutation out of the CDRs in a framework position far from the antigen combining site. The fact that the same mutation constitutes one of SHM hotspots in Nature underlines the possibility of finding beneficial mutations throughout the V-region.

Indels

The design of antibody libraries aiming at *in vitro* affinity maturation rarely includes indels. The only platform used for development of therapeutic antibodies that incorporates low percentage of indels during the affinity maturation process is mammalian surface display (Bowers et al., 2011), by using the enzyme AID *in vitro* to generate diversity in the V-gene in mammalian cells, mimicking the process of SHM *in vivo*. The rate of in-frame indels occurring by *in vitro* SHM (≈ 0.05-0.1%) is 10-fold lower than the *in vivo* SHM rate (Bowers et al., 2014). Such a low rate of indels introduced by *in vitro* SHM, arising only by local sequence duplication, shows that this approach samples an even lower diversity of indels than *in vivo* SHM.

In some cases antibody libraries generated by classic methods of site-directed mutagenesis yielded antibodies with improvements attributed to in-frame indels (Finch et al., 2011). These modifications were not intentionally incorporated but probably occurred by mispriming during mutagenesis. Variants of a human anti-IL-6 antibody with enhanced affinity contained either deletion in the V_L CDR3 or insertion in the V_H CDR3. The authors proposed that this may be an additional important strategy to engineer antibody affinity (Finch et al., 2011).
Libraries with length variation in the V_H CDR2 of an anti-hapten antibody were constructed using a method called insertional mutagenesis. The design of libraries with randomized

insertions of different lengths was based on a model structure of the parent antibody 57-2 complexed with 17β-estradiol combined with the observation of natural length variation at the specific position. Libraries with length of insertions up to 4 amino acids on the tip of a b-turn like V_H CDR2 loop were designed with the aim of creating novel contacts with the hapten. Several improved variants were identified, with the highest affinity mutant carrying 12-fold higher affinity than the parent (Lamminmäki et al., 1999). This method was also successfully used to design libraries with insertions in the V_L CDR3 based on structural information with aim changing the specificity of an anti-hapten antibody (Krykbaev, Tsantili, Jeffrey, & Margolies, 2002).

1.2.2.2. Affinity selections

The advent of phage display (McCafferty et al., 1990) was followed by development of more display systems (Weisser & Hall, 2009) that provide linkage of the genetic information (genotype) with the encoded protein (phenotype). While phage display is the most popular method for the isolation of specific antibodies, for maturation of antibody affinity other display technologies are widely used as well (Lipovsek & Plückthun, 2004; Hoogenboom, 2005; Fukuda et al., 2006; Zahnd, Amstutz, & Plückthun, 2007; Bradbury, Sidhu, Dübel, & Mccafferty, 2011).

Other display technologies

Other display technologies that have been used for selection include cell-free formats, like ribosome (Jozef Hanes, Jermutus, Weber-Bornhauser, Bosshard, & Plckthun, 1998), mRNA (Josephson, Ricardo, & Szostak, 2014), CIS (Odegrip et al., 2004) display and methods that involve *in vitro* compartmentalisation in water-in-oil emulsion droplets, like bead surface display (BeSD) (Mankowska et al., 2016) and SNAP display (Houlihan, Gatti-Lafranconi,

Kaltenbach, Lowe, & Hollfelder, 2014). Moreover, bacterial (Georgiou et al., 1997), yeast (Eric T. Boder & Wittrup, 1997) or mammalian (Bowers et al., 2011) cells expressing antibodies on their surface are not commonly used for selection, but for screening by FACS (**Figure 13**).

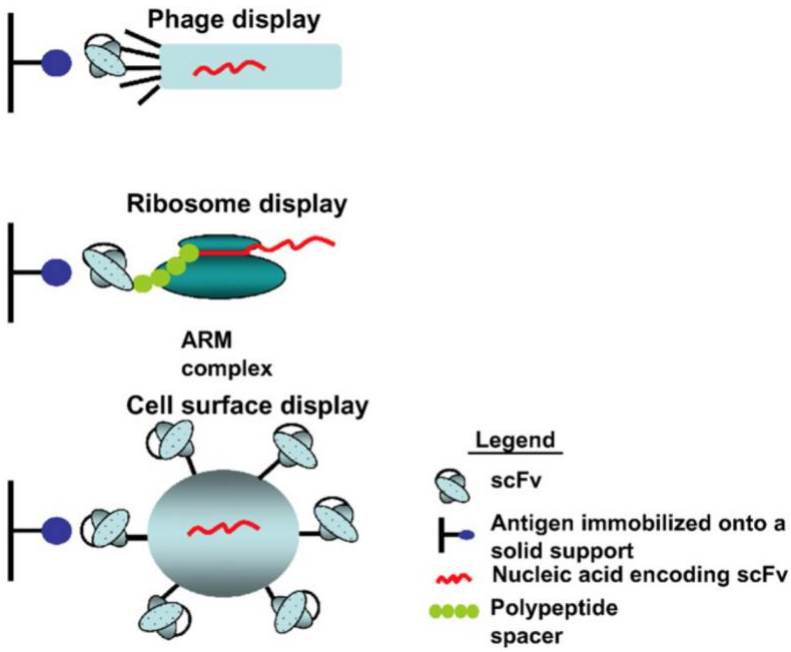


Figure 13. Different display formats for scFv affinity maturation by *in vitro* evolution. (A) Phage display, (B) Ribosome display, (C) Cell surface display. Adapted from (Weisser & Hall, 2009).

One important restriction of the selection process by phage display is the limitation of the library size, that is restricted by the transformation efficiency of cells which is typically $\approx 10^9$ cfu/µg DNA (Dufner et al., 2006). Even lower number of variants can be sampled by cell surface display systems, as these are limited by the number of cells that can be sorted by FACS (practically $\approx 10^6$-10^7) (Feldhaus et al., 2003). *In vitro* expression systems remove this restriction and give access to greater sequence space, as the input of selection is a DNA library that is amplified by PCR, consequently, the library size is not limited by the transformation efficiency of cells, thus allowing libraries theoretically greater than 10^{13} to be explored (Barrick, Takahashi, Ren, Xia, & Roberts, 2001). Another complication of phage

display is that expression takes place into bacterial cells, thus variants are enriched not only by selection, but also for rapid growth in the bacterial host or even toxicity (Plückthun, Schaffitzel, Hanes, & Jermutus, 2000). Furthermore, *in vitro* methods allow selection conditions to be controlled by including detergents and competitors, in addition to extreme pH, temperature and ionic strength since cell viability does not need to be maintained.

Ribosome display

Ribosome display was first developed in 1994 (Mattheakis, Bhattt, & Dower, 1994) with the display of synthetic peptides on ribosomes and subsequently was further adapted for display of scFvs (J Hanes & Plückthun, 1997). The link between genotype and phenotype is realized during *in vitro* translation, by formation of a complex of the translated scFv linked to the ribosome and the mRNA encoding it. The complex remains stable in cold environment (4 °C) and high concentrations of magnesium ions, which can be used for affinity selections via panning to isolate library members with improved affinity for the target protein. After selection, the mRNA pool encoding selected variants is recovered by RT-PCR.

Ribosome display has been proven a powerful technology for *in vitro* antibody affinity maturation. A DNA amplification step by PCR is performed after each round of *in vitro* selection, thus using error-prone DNA polymerases between the rounds of selection have enabled accumulation of multiple beneficial mutations hat conferred affinities up to ≈ 81 pM (J Hanes, Schaffitzel, Knappik, & Plückthun, 2000; Thom et al., 2006). Furthermore, ep-PCR has been combined with recombination between rounds of *in vitro* selection to generate the highest affinity peptide binding mAb to date (K_D=1 pM) (Luginbühl et al., 2006).

Stringency during affinity-based selections

The conditions of affinity-based selections must be designed in order to enrich for the highest affinity variants instead of variants with affinity similar to the parent or improved expression in *E.coli* (Zahnd, Sarkar, & Plückthun, 2010). One strategy to achieve enrichment of improved variants is performing equilibrium selections, by panning with an antigen concentration below the K_D and reducing the concentration by 10-fold after each selection round (Hawkins, Russell, & Winter, 1992; Groves & Nickson, 2012). Another methodology is carrying out off-rate selections, by panning using an excess of competing unlabelled antigen to isolate variants with slower off-rates, which will remain bound to the biotinylated antigen captured on streptavidin beads (Hawkins, Russell, & Winter, 1992; Jermutus, Honegger, Schwesinger, Hanes, & Plückthun, 2001).

1.2.3. *In vitro* evolution of other proteins

Strategies for construction of libraries with indels

Despite the technical difficulties in construction of libraries containing length variation of diverse sequence composition, some protein engineering studies have given insight into the structural and functional effects of indels in proteins other than antibodies. These libraries contained in-frame indels which were constructed either by random (Arpino, Reddington, Halliwell, Rizkallah, & Jones, 2014; Liu, Wei, Dong, Xu, & Liu, 2015) or by targeted mutagenesis approaches (Afriat-jurnou, Jackson, & Taw, 2012). Methods that have been developed for construction of libraries with indels are further analysed in the introduction of **Chapter III** of this thesis.

The essential role of loop length variation in the emergence of new functions was demonstrated by reconstruction of the pathway of evolution of PTE phosphotriesterase

activity, which catalyses the hydrolysis of synthetic substrates, from PTE-like lactonases (PLLs), that exhibit promiscuous phosphotriesterase activity (Afriat-jurnou et al., 2012). Based on structural and phylogenetic analyses a 7-amino acid deletion combined with an epistatic substitution enabled the emergence of HSLase activity while retaining PTE's paraoxonase activity.

More recently developed methods of transposon-mediated mutagenesis (D. Jones, 2005) have been used for construction of libraries with in-frame amino acid deletions across the entire GFP gene (Arpino et al., 2014; Liu et al., 2015). One of these studies showed that the enhanced green fluorescent protein (eGFP) can tolerate amino acid deletions mainly in loops (Arpino et al., 2014). Furthermore, this study revealed that deletion mutations more often occur in helices than in strands, except from positions towards the strand termni. Furthermore, a deletion in the N-terminal 3_{10} helix conferred improved eGFP fluorescence, which enhanced the fluorescence not by affecting the brightness but by optimization of the folding efficiency, therefore increasing the production of functional eGFP in the cells. Another similar study on the UV-optimized GFP variant (GFPUV) examined the idea that decreased fluorescence of GFPUV mutants with deletions could be the result of compromised efficiency of the protein folding and therefore they rescued these variants by introducing mutations that enhanced folding (Liu et al., 2015).

These studies are in agreement with the idea that indels could be a key step in the evolution of new functions (Tóth-Petróczy & Tawfik, 2013), with substitutions in close proximity to the indels compensating or enabling the functional effect of indels (Leushkin, Bazykin, & Kondrashov, 2012).

1.3. Interaction between mutations during affinity maturation

1.3.1. Fitness landscapes and evolutionary trajectories

In order to understand the enormous breadth of protein sequence and function that could be sampled by natural and directed evolution, fitness landscapes in protein sequence space have been used to visualise evolutionary trajectories as a movement through related sequences (Maynard Smith, 1970; Wright, 1930). Protein sequence space shows all the possible combinations of amino acids of a protein of certain length so that neighbouring points represent genotypes which differ by one amino acid mutation. The large number of theoretical possibilities makes the protein sequence space multi-dimensional. For a protein with 100 amino acids sequence space would be represented with 100 dimensions. For simplicity, sequence space is condensed into two axes (**Figure 14**).

Fitness can represent specific properties, such as protein binding affinity or catalytic activity. Flat areas indicate proteins with low fitness while peaks represent desired proteins with high fitness. Thus adaptive evolution can be conceptualized as a walk from a lower to an optimal phenotype across the landscape (Maynard Smith, 1970). In directed evolution, selection or screening methods can explore only a small proportion of a fitness landscape with the aim of isolating phenotypes with higher fitness peaks for a biophysical property such as affinity or stability (Romero & Arnold, 2009).

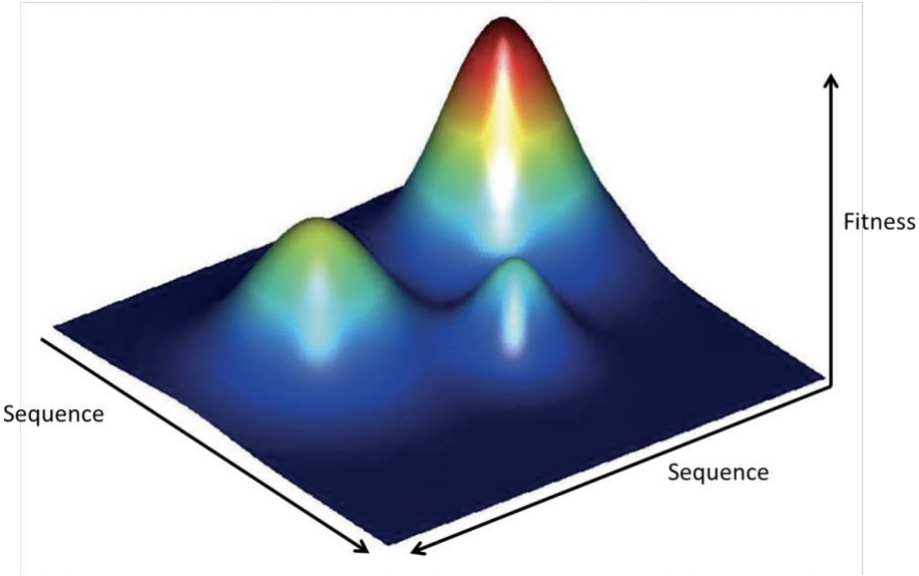


Figure 14. Graphical representation of a fitness landscape. Sequence space is represented by two axes. Height represents protein fitness (vertical axis). The transition from blue to green to yellow to red indicates an increase in fitness. Most of the sequences in the landscape are non-functional (blue) while rare fitness optima are represented in the peaks. If all mutations are additive, the fitness peaks appear smooth while if epistasis occurs between mutations, a rugged landscape with multiple peaks separated by valleys of genotypes with low fitness is observed. Adapted from (de Visser & Krug, 2014).

Experimental studies have generated sequence function maps by using libraries of protein binders coupled to yeast display and high throughput sequencing. Sequence enrichment in the selected versus initial library has been used to assess the fitness (binding) of adaptive mutations on the structure of a protein (Whitehead, Baker, & Fleishman, 2013). Deep mutational scanning represents a powerful method for identifying proteins with high fitness and also enabling experimental investigation of fitness landscape topologies. Another example of mapping a fitness landscape has explored the space of all possible sequences (4^{24} sequences) for short RNAs (24 nt length) by in vitro selection for binding to GTP (Jiménez, Xulvi-brunet, Campbell, Turk-macleod, & Chen, 2013).

The main constraints that are known to shape the fitness landscape of a protein include 1) epistatic interactions observed between mutations and 2) the protein stability threshold (Bershtein, Segal, Bekerman, Tokuriki, & Tawfik, 2006).

1.3.2. Epistatic interactions

Epistasis is the non-additive phenotypic effect of the combination of individual mutations, meaning that the effect of each mutation on function depends on the genetic background on which it occurs (Phillips, 2008). For instance, a beneficial mutation may exert its effect only in the presence of another permissive mutation (Gong, Suchard, & Bloom, 2013). Conversely, a mutation may be beneficial only in the absence of another restrictive mutation (B. C. Dickinson, Leconte, Allen, Esvelt, & Liu, 2013). Epistatic effects can be synergistic, diminishing, antagonistic, aggravating, ameliorating, buffering, compensatory, and reinforcing. Synergistic is when the effect of the combination is beyond that expected from the individual effects, which could mean extra good or extra bad. Positive epistasis means that the phenotype is higher than expected, while negative means lower than expected (Phillips, 2008). Epistatic mutations have been shown in both natural (Ortlund, Bridgham, Redinbo, & Thornton, 2007) and experimental evolution (Bershtein et al., 2006), indicating that the effect of combining mutations is unpredictable.

Similarly, not consistent additivity has been observed in *in vitro* evolution experiments, as combining individual beneficial mutations in different CDRs in a parallel CDR walking approach resulted in unpredictable outcomes (Barbas et al., 1994; W. P. Yang et al., 1995). Measurements of the affinities of intermediate mutants of an *in vivo* affinity maturation pathway has shown that the effect of individual mutations on binding was either positive or neutral, with a high degree of synergy between mutations (P. L. Yang & Schultz, 1999). In other cases, the effect of combining individual somatic mutations was either additive (Ulrich

et al., 1997; England, Nageotte, Renard, Page, & Bedouelle, 1999) or non-interactive (Romesberg, 1998).

1.3.3. Trade-offs between affinity and stability

Natural evolution not only results in proteins with improved activity, but it also maintains soluble, thermodynamically stable proteins (Tokuriki & Tawfik, 2009; Soskine & Tawfik, 2010). In parallel, antibodies need to maintain high thermodynamic stability and reduced aggregation tendency to be considered as therapeutic candidates because these properties could affect manufacturing, storage and serum half-life (Carter & Merchant, 1997). Often mutations that enhance a protein function have a negative effect on stability, indicating a trade-off between function and stability (X. Wang, Minasov, & Shoichet, 2002; Gong et al., 2013). To overcome this effect compensatory mutations have been shown to restore the compromised stability (Tokuriki, Stricher, Serrano, & Tawfik, 2008). Additionally to mutations that improve affinity, compensatory mutations that preserve thermodynamic stability have been shown to accumulate during *in vivo* (Sun et al., 2013; F. Wang, Sen, et al., 2013, retracted recently; Sen et al., 2017) and *in vitro* (Julian, Li, Garde, Wilen, & Tessier, 2017a) antibody affinity maturation.

1.4. Objectives of this thesis

1.4.1. Affinity maturation of a scFv by *in vitro* continuous evolution

The **Chapter II** of this thesis describes efforts to develop a continuous evolution for in vitro evolution of protein binding. Using a continuous evolution set-up sequence space can be explored in a different way, raising the possibility of fundamentally altering the strategy of combinatorial methods. Normally we conduct very stringent selections that yield very few hits that are then re-randomised. However, this is very different from the way Nature evolves biomolecules. The near-neutral theory of natural evolution (Ohta, 1992) suggests that only a fraction of differences between homologous genes lead to improvements in fitness. The remaining majority of observed polymorphisms are the result of genetic drift exploring neutral networks of functionally equivalent genes. Such genetic drift that leads to cryptic variation (neutral to current function) has been suggested to be beneficial to subsequent adaptive evolution (Davies EK, 1999; Masel, 2006; Wagner, 2011). The conditions of the continuous evolution "lagoon' in which competition occurs can be manipulated, so that either regimes of neutral or more stringent adaptive evolution are favoured. For example, will the phylogeny of selected clones be more diverse (in contrast to the severe bottlenecking that a stepwise process implies) and will the broader sampling of sequence space accelerate evolution leading to ultimately better binders?

In this way scFvs relationship between the whole protein sequence and high affinity could be explored for the first time. This objective could be achieved by using continuous *in vitro* evolution. Continuous *in vitro* evolution of binding proteins had not been described in the beginning of this PhD study (January 2014), but the development of a continuous *in vitro* evolution system (PACE) by Esvelt and Carlson (2011 and 2014) suggested a practical way of conducting such experiments. Based on the PACE system we tried to a develop a system of continuous evolution of protein binding properties. As it was shown by Esvelt such a

system would allow large protein populations to evolve over hundreds of rounds of evolution in short time. As the number of rounds performed and the size of population are determinants of *in vitro* evolution experiments, the mutants that could be sampled by PACE are significantly higher than for any other conventional *in vitro* evolution approach, so the improvements in the desired property that could be achieved by PACE are expected to be unprecedented.

1.4.2. Affinity maturation of an entire scFv using libraries with insertions and deletions

In Nature, many steps of the diversification of the primary antibody repertoire include variation of the length of the CDR loops, which can have considerably different lengths with some of them varying more than others, indicating that these can tolerate significant length modification. The loop length is recognised as the primary factor that defines the structural diversity of the antigen combining site.

Following antigen stimulation, further diversification through SHM includes mainly point substitutions and a small percentage of indels. SHM has been shown to introduce subtle changes in the already defined topology of the combining site, thus the percentage of indels could be low because SHM does not introduce major structural changes that would result from an indel. Alternatively, it is likely that the antibody fold can accommodate higher diversity of indels that improve affinity than that observed in Nature, as SHM does not sample random but only biased diversity with a biological limit in the level of affinity that can be achieved, which is sufficient for a biologically competent affinity maturation process. This notion is supported by the fact that the mechanism of SHM has been shown to introduce only a very low percentage of in-frame indels, thus not permitting sampling of high diversity of indel variants. Only anti-HIV bnAbs, which have been subjected to extensive affinity maturation, acquire a higher percentage of indels with a functional role in neutralization.

Consequently, it would be interesting to study the impact of random high diversity indels on antibody affinity by *in vitro* evolution, as it is likely that more indels with beneficial effect on affinity could be found and the structural basis of their function could be elucidated. Other researchers have identified indels that improved antibody affinity, but their methods did not allow them sampling of high diversity indels at random positions. Mammalian surface display introduces only low diversity of indels due to the small library size and the low rate of incorporation of in-frame indels by this method (Bowers et al., 2014). Other researchers have studied long full diversity insertions by insertional mutagenesis, but the insertions were not random, as the design was based on structural information and natural variation (Lamminmäki et al., 1999).

This thesis focuses on the development of new approaches for construction of libraries with in-frame indels that would enable sampling of different lengths of high diversity indels at random positions across the entire antibody variable regions. **Chapter III** describes the application of a random approach that allowed the construction of libraries with random insertions and deletions. **Chapter IV** describes the development of a semi-random approach to build libraries with different lengths of insertions that could be widely applied in future *in vitro* antibody affinity maturation campaigns. Libraries constructed by either of these approaches were subjected to affinity-based selections using appropriate display methods and yielded variants with insertions conferring improved affinity.

2. Continuous *In Vitro* Evolution of Protein Binding

2.1. Introduction

2.1.1. *In vitro* Continuous Evolution

The greatest limitation of *in vitro* evolution is each cycle of diversification and selection is cumbersome and require significant time and effort by a skilled researcher. A process of continuous *in vitro* protein evolution, which would allow protein populations to evolve autonomously by continuous rounds of mutation and selection without the interference of the researcher would allow evolution of large protein populations over hundreds of cycles of evolution, circumventing the limitations in the number of mutants that can be screened by traditional *in vitro* evolution experiments.

In theory, a process of continuous evolution could be a cell culture growing under conditions of selection pressure, where variants of a gene which encode the protein of interest are selected because they confer survival of the cell. For example, a pool of β-lactamase gene variants, that allows growth of bacterial ells growing in medium containing ampicillin. Such a process though would not allow the selection of functional genes, because genomic mutations could permit growth of non-functional genes. Furthermore, the rate of error by cellular replication is not sufficient for generation of large number of mutants.

2.1.2. Filamentous phage

The filamentous phage M13 is a bacteriophage which infects gram-negative bacteria by attaching to receptors on a projection found on the bacterial surface, the F-pili (Mar, & Herbert, 2014). Filamentous phage M13 possesses some characteristics that render it suitable for continuous evolution. Because of its rapid phage cycle compared to *E. coli* cells, it is possible to sustain a phage culture while providing fresh *E. coli* cells. M13 is non-lytic and it does not kill the host cell, but phage particles are continually released and the sole impact on the cells is that it slows down the cell growth rate (Hoffmann-Berling & Mazé, 1964).

Furthermore, release of the first progeny phage particles starts about 10-15 minutes after infection.

The most known phage coat protein because of its common use in phage display, is pIII, which constitutes the only phage coat protein essential for infectivity (Russel & Model, 1989). Phage protein pIII is expressed on the cell inner membrane before incorporation into phage particles. Mature pIII is comprised of three domains, the N1, N2 and CT, which are separated by two glycine-rich linkers. The N-terminal N1 and N2 domains consist of 68 and 132 amino acids, respectively, while the C-terminal (CT) membrane-anchoring domain 149 amino acids. The G1 linker which connect the N1 and N2 domain is shorter than (18 amino-acids) compared to the G2 linker (39 amino-acids). The CT domain adhere to the phage particle, while the N1 and N2 domains are displayed on the surface of the phage (Holliger, Riechmann, & Williams, 1999a).

2.1.3. Filamentous phage infection mechanism

Three to five copies of pIII are attached to the phage surface, but a single copy is sufficient for infectivity (Russel & Model, 1989). Filamentous phages infect *E. coli* cells in a process involving two steps (**Figure 15**). The initial step involves binding of the N2 domain to the F-pilus. Once the phage particle is attached to the cell, the pilus retraction allows the N1 domain to approach and interact with the TolA-C receptor (Riechmann & Holliger, 1997; (Holliger, Riechmann, & Williams, 1999). It has been shown that the N1 domain forms a complex with the N2 domain (Lubkowski, Hennecke, Plückthun, & Wlodawer, 1999), which hides the N1 binding site from the TolA-C until the N2 domain binding to the F-pilus and disruption of the intra-molecular association between two domains (Deng & Perham, 2002).

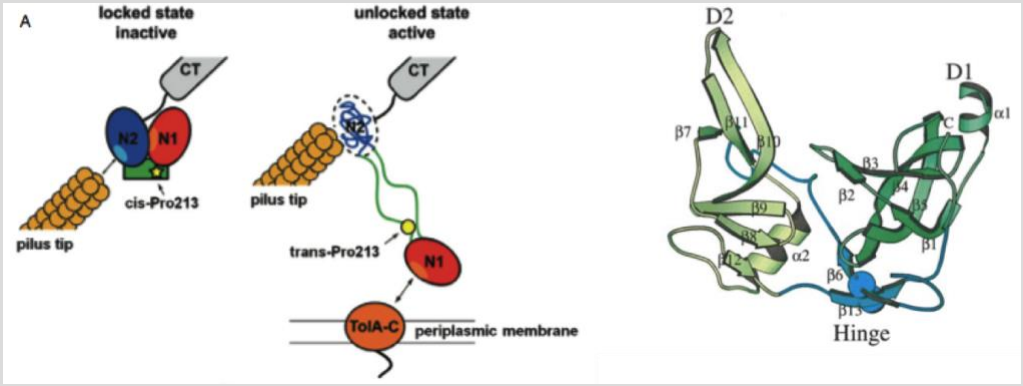


Figure 15. A) Schematic representation of the M13 phage infection process. Starting by interaction of the N2 domain of pIII (blue) with the F-pilus, the intramolecular association between two N1 and N2 domains is disrupted and the N1 domain binding site is accessible for interaction with the TolA-C receptor. Figure adapted from (Marvin D.A. 2014) **B)** N1-N2 domain structure. Figure adapted from (Holliger P, 1999).

2.1.4. Continuous culture of filamentous phage

Researchers have performed continuous culture of filamentous phage (Husimi, Nishigaki, Kinoshita, & Tanaka, 1982) by adding new host cells to a culture of filamentous phage growing in a vessel called cellstat". The dilution rate was adjusted in way that prevented phage wash out from the "cellstat", but did not allow cell division. The purpose of his studies was to compare the fitness of different phage strains, by measuring how quickly the fittest strain took over the "cellstat". Filamentous phage infect only *E. coli* cells that contain the F-pilus on their surface, which is expressed only when the cells are in the log-phase, which depends on the cell density of the cell culture. In order to create a source of host cells growing at constantly at the log-phase, he constructed a device called the turbidostat where the growth phase was kept constant by observing the cell density and regulating the media flow in the turbidostat.

2.1.5. A system for the continuous directed evolution of biomolecules (PACE)

Using a continuous phage culture based on the "cellstat" (Husimi et al., 1982), Esvelt *et al.* (Nature, 2011) constructed the "lagoon" (**Figure 16**). While increasing the cellular mutagenesis rate, the mutagenesis in the lagoon is restricted to the gene of interest. This was achieved by inserting the gene of interest in the phage genome. The phage population is propagating in the lagoon, while fresh host cells are constantly transferred to the lagoon, thus the mutations to the host cells would have minimal impact on the selection of gene variants with mutations. As mentioned above, filamentous phage infectivity depends entirely on the presence of one phage coat protein, the pIII. To selectively enrich the gene mutants with the desired activity they linked the function of interest with the pIII expression, which allows filamentous phage selective propagation. Additionally, the phage titer rises with increasing pIII expression levels, so the pIII infective phage titer could be linked with the activity levels of the evolving protein.

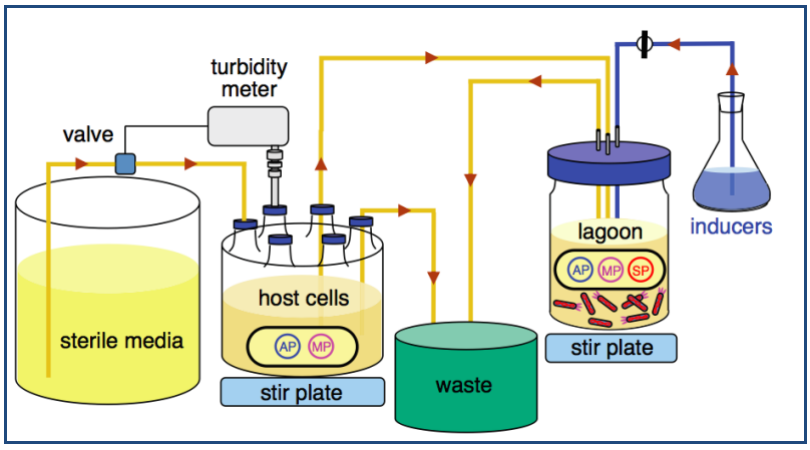


Figure 16. In PACE the phage population is propagating in the lagoon. The host cells are kept in a constant growth phase (mid-log phase) in a separate vessel, where the cell density is measured by a turbidity meter. From this vessel fresh host cells are transferred to the lagoon. Adapted from (Esvelt, Carlson, & Liu, 2011)

The proof of concept study of the phage assisted continuous evolution system (PACE) described the evolution of T7 RNA polymerase variants capable of transcribing DNA downstream the T3 promoter (Esvelt et al., 2011). T7 RNA polymerase is highly specific for the T7 promoter and T7 RNA polymerase variants recognizing the T3 promoter has not been reported before. Starting from less than 3% activity levels on the T3 promoter, T7 RNAP evolution by PACE resulted in variants exhibiting more than 200-fold improvement in 8 days of continuous evolution after around 200 rounds of evolution.

Linkage of the function of interest, the T7 RNA polymerase binding to the T3 promoter, was achieved by deleting gene III from the phage genome and replacing it with the T7 RNA polymerase gene (**Figure 17**). The gene III was cloned into an accessory plasmid (AP) under the control of the T7 promoter, so that pIII expression and therefore the phage infectivity depends on the activity of the T7 RNA polymerase variant on the phage genome. When the evolving gene mutant encodes an active T7 RNA polymerase variant, pIII is expressed and infective phage particles are released. The progeny phage produced from a cell infected with a phage containing an inactive T7 RNA polymerase variant do not induce pIII expression and are not rendered infective. Additionally, the infective phage titer production varies according to the activity levels of the T7 RNA polymerase mutant, which defines the pIII expression levels. Thus in the lagoon only recombinant phage encoding active enzyme variants are infective and propagate. Furthermore, phages containing variants with the highest activity levels will infect more host cells. The evolving population for the T7 RNAP evolution experiments was around 10^{10} phages (Bryan C Dickinson, Leconte, Allen, Esvelt, & Liu, 2013).

In the lagoon, T7 RNAP gene variation was introduced by using host cells containing a mutagenesis plasmid (MP), which encodes a DNA polymerase dominant negative proofreading subunit (**Figure 17**). Expression of this subunit leads to decreased fidelity of the

DNA polymerase. For a 1000 base pairs evolving gene, the mutagenesis rate was estimated to be sufficient to generate all the single and double mutants per round of evolution.

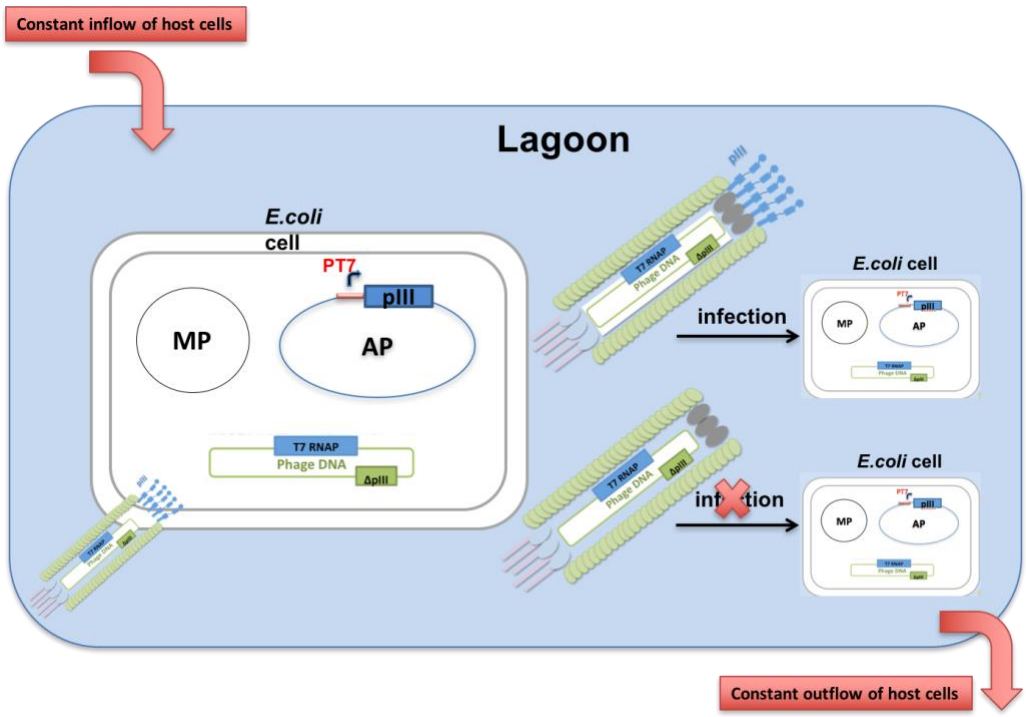


Figure 17. Schematic representation of the PACE system (Esvelt et al., 2011). In the lagoon, there is a linkage of selective replication of T7 RNAP gene variants recognizing the T3 promoter with selective phage propagation. This is achieved by replacing the gene III on the phage genome with the T7 RNAP coding sequence. The gene III is cloned into an accessory plasmid (AP) under the control of the T3 promoter (PT3). T7 RNAP gene variation in the lagoon is introduced by the mutagenesis plasmid (MP), which encodes a DNA polymerase dominant negative proofreading subunit. The host cell strain contains both the AP and the MP. When a phage encoding a functional T7 RNAP variant infects a host cell, it recognizes the T3 promoter and induces the expression of pIII. Consequently, the phages particles that are released from this cell are infective and the functional T7 RNAP variants keep replicating in the lagoon. In contrast, a phage encoding an inactive T7 RNAP does not recognize the T3 promoter and non-infective phage particles are produced from this cell. Continuous inflow and outflow of host cells in the lagoon does not allow the division of host cells in the lagoon.

2.1.6. Selectively Infective Phage (SIP) technology

Selectively infective phage (SIP) technology was developed in 1994 (Dueñas & Borrebaeck, 1994) with aim the selections of binding proteins from libraries of variants of a gene by linking binding to phage infectivity. SIP was further explored and evaluated by other researchers (Krebber, Spada, Desplancq, & Plückthun, 1995; Krebber et al., 1997; Jung, Arndt, Müller, & Plückthun, 1999), but never reached routine use in antibody selections.

SIP is based on the phage coat protein pIII which is essential for phage infectivity. The three domains of pIII (N1, N2 and CT) are joined by two linkers (G1 and G2). The attribute that makes pIII suitable for SIP is the fact that phage infectivity is dependent on the presence of the N-terminal domain (N1). Phages expressing only the CT-N2 domain or CT domain are not infective, but the infectivity is restored when the N1 or N1-N2 is associated with the CT-N2 or CT domain, respectively, by a non-covalent interaction. Thus, an interaction of a scFv with its cognate ligand could trigger phage infectivity. In the most common SIP format, the N1-N2 domains are deleted from the phage genome and fused to the peptide ligand of the scFv, while the scFv coding sequence is cloned into the phage genome so that it can be expressed as a CT- scFv fusion protein on the phage surface. The scFv variants exhibiting high affinity and specificity for the peptide ligand, interact with the N1-N2-ligand fusion protein, which leads to reconstitution of the pIII protein and consequently to a selectively infective phage (**Figure 18**). The coding sequences of scFv mutants that survive the selection are recovered because the phage genome confers resistance to an antibiotic marker and permits growth only of the infected cells.

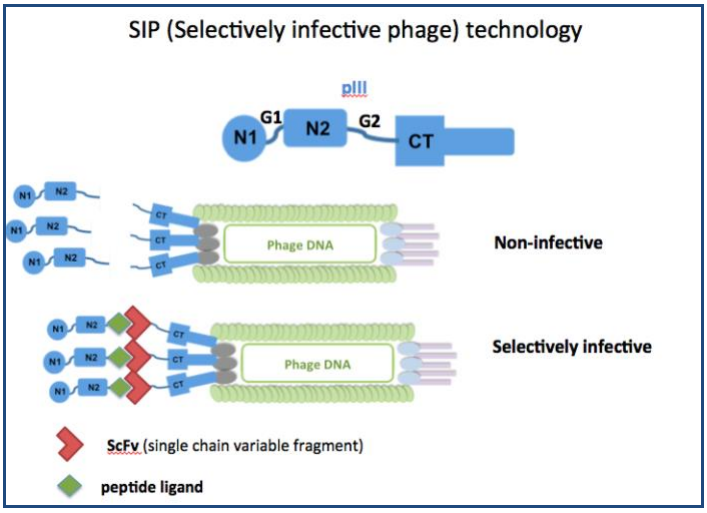


Figure 18. pIII is comprised of the domains N1, N2 and CT joined by flexible linkers (G1 and G2). The N1 domain confers phage infectivity. Phages containing only the CT domain on their surface are non-infective, even when the N1-N2 domain is present. A phage displaying an scFv-CT fusion is rendered infective only if the scFv (red shape) binds to its cognate peptide ligand (green shape). The phages that contain N1-N2-ligand fusion bound on the scFv are called "selectively infective phages". Adapted from (Krebber et al., 1995)

2.1.7. *In vitro* and *in vivo* SIP technologies

There are two different types of SIP: *in vivo* and *in vitro* SIP. For *in vitro* SIP, the phage displaying the binding protein fusion (CT-ScFv) and the N1-N2-ligand fusion are separately produced and purified and then mixed to yield infective phages (**Figure 19**) (Krebber et al., 1997; Pedrazzi, Schwesinger, Honegger, Krebber, & Plückthun, 1997; Cèbe & Geiser, 2000). In in vivo SIP, both the ligand and the scFv fusion proteins are encoded on the phage genome. Appropriate leader sequence drive N1-N2-ligand expression to the bacterial periplasm, while the CT-scFv remains anchored to the bacterial inner membrane. When an interaction occurs in the bacterial periplasm the phage incorporate reconstituted CT-ScFv:N1-N2-ligand protein, which renders it infective (**Figure 19**) (Krebber et al., 1995; Spada, Honegger, & Plückthun, 1998). *In vivo* and *in vitro* SIP has been studied extensively as they could represent fast selection methods for identification of high affinity binding

proteins. But there are some limitations in the application of both *in vitro* SIP and *in vivo* SIP technologies (Jung, Arndt, Müller, & Plückthun, 1999).

There are two published studies referring to affinity selections by *in vivo* SIP (Krebber et al., 1995; Spada, Honegger, & Plückthun, 1998). Krebber in 1995 cloned the CT-scFv coding sequence of an anti-hemagglutinin (HA) peptide scFv (K_D 2×10^{-8}M) on the phage genome. Both the scFv-CT and the N1-N2-HA fusion proteins were encoded on filamentous fd phage genome. After overnight phage production, around 10^{10} selectively infective phages were used to infect *E. coli* cells. The infectivity was quantified by counting the number of colonies after infection: 228 colonies were observed, suggesting that 228 phages out of 10^{10} were infective. Thus the efficiency of the method in this study was very low (2.2×10^{8}-fold lower infectivity than that of the wild-type phage), although it was successfully used for selection of binding proteins *in vivo* (Spada, Honegger, & Plückthun, 1998).

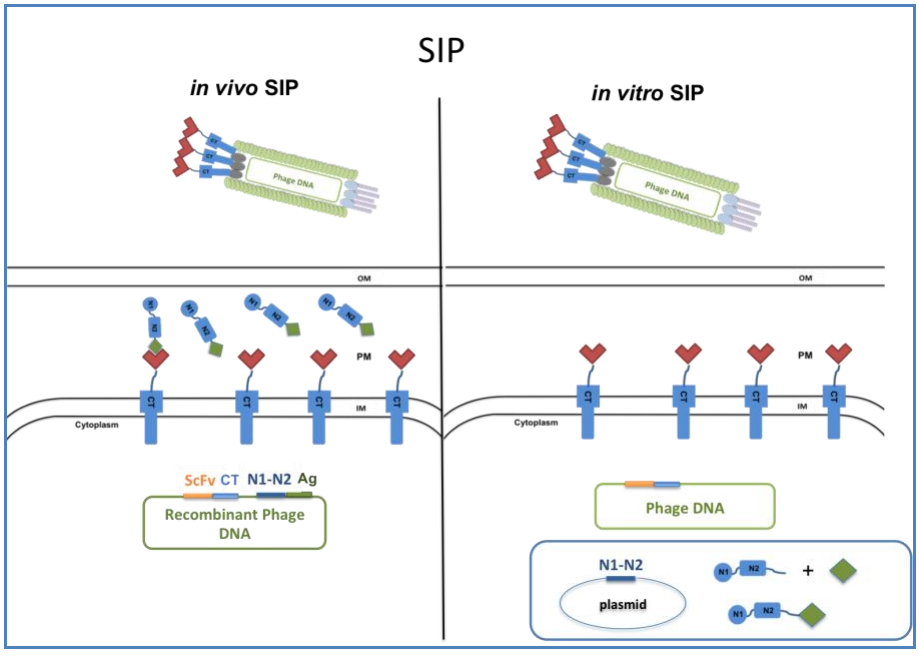


Figure 19. Schematic representation of *in vitro* and *in vivo* SIP technology (Krebber et al., 1997). In *in vivo* SIP both the ScFv-CT and the N1-N2-peptide ligand (Ag) fusion protein are cloned in the phage genome. In *in vitro* SIP, only the ScFv-CT fusion protein is cloned in the phage genome. The N1-N2-Ag fusion protein and phages particles displaying the scFv are produced separately and then mixed *in vitro*. PM, bacterial periplasm

The efficiency of *in vitro* SIP was higher than that of *in vivo* SIP (Krebber et al., 1997). When beta-lactamase (31 kDa) was inserted between the N1 and N2 domains (phage-CT-N2-Bla-N1) the infectivity was 10^4-fold lower than that of the wild-type phage, showing that the Bla insertion caused steric hindrance, probably because of the disruption of the inter-domain interaction between the N1 and N2 domains. The same result (10^4-fold lower infectivity) was obtained when an anti-fluorescein scFv with K_D of 3×10^{-10} M was fused to the CT-N2 domain while the antigen (fluorescein) was conjugated to the N1 domain, to form a selectively infective phage (phage-CT-N2-ScFv: fluorescein-N1). The fact that both a covalent and a non-covalent interaction resulted in the same infectivity suggests that selectively infective phages can be infective. Consequently, this is not a factor that restricts the efficiency of *in vitro* or *in vivo* SIP technology. The efficiency of other factors, as steric hindrance of the scFv insertion and production of selectively infective phages could be the factors resulting in so low infectivity.

2.1.8. Affinity maturation of a scFv by *in vitro* continuous evolution

Using a continuous evolution set-up sequence space can be explored in a different way, raising the possibility of fundamentally altering the strategy of combinatorial methods. Normally we conduct very stringent selections that yield very few hits that are then re-randomised. However, this is very different from the way Nature evolves biomolecules. The near-neutral theory of natural evolution (Ohta, 1992) suggests that only a fraction of differences between homologous genes lead to improvements in fitness. The remaining majority of observed polymorphisms are the result of genetic drift exploring neutral networks of functionally equivalent genes. Such genetic drift that leads to cryptic variation (neutral to current function) has been suggested to be beneficial to subsequent adaptive evolution (Davies EK, 1999; Masel, 2006; Wagner, 2011). The conditions of the continuous evolution

“lagoon’ in which competition occurs can be manipulated, so that either regimes of neutral or more stringent adaptive evolution are favoured. For example, will the phylogeny of selected clones be more diverse (in contrast to the severe bottlenecking that a stepwise process implies) and will the broader sampling of sequence space accelerate evolution leading to ultimately better binders?

In this way scFvs relationship between the whole protein sequence and high affinity could be explored for the first time. This objective could be achieved by using continuous *in vitro* evolution. Continuous *in vitro* evolution of binding proteins has not been described so far, but the development of a continuous *in vitro* evolution system (PACE) by Esvelt and Carlson (2011 and 2014) suggested a practical way of conducting such experiments. Based on the PACE system we would like to use this method for continuous evolution of protein binding properties. As it was shown by Esvelt such a system would allow large protein populations to evolve over hundreds of rounds of evolution in short time. As the number of rounds performed and the size of population are determinants of *in vitro* evolution experiments, the mutants that could be sampled by PACE are significantly higher than for any other conventional *in vitro* evolution approach, so the improvements in the desired property that could be achieved by PACE are expected to be unprecedented.

2.2. Results and Discussion

2.2.1. Using SIP technology to link protein affinity to infective phage production in PACE

To develop a PACE process for continuous evolution of the affinity of protein binders to their cognate ligands, SIP technology was used to link the affinity to the production of infectious progeny. The chosen protein binder was a scFv fused to the CT domain of the pIII protein, by cloning the coding sequence of the CT-protein binder fusion into the genome of a phage M13 strain lacking gene III. To couple pIII protein reconstitution to the desired function, the coding sequence of the N1-N2-peptide ligand fusion could be cloned into an accessory plasmid (AP) which would be present in the host cells. Only phages with variants of the protein binder with high affinity to their cognate peptide ligand would be able to induce sufficient pIII reconstitution, propagate and persist in the lagoon (**Figure 20**).

A possible implication is that high pIII expression before phage infection leads to inhibition of phage infectivity (Carlson, Badran, Guggiana-Nilo, & Liu, 2014a). Particularly, when the half N-terminal domain (residues 1-198) of pIII is expressed the infection is inhibited. Expression of the N1-N1-ligand fusion protein under the control of the lac promoter would probably lead to high background expression and decreased infection due to inhibition of infection by the N1-N2-ligand fusion protein. To deal with this problem in PACE evolution experiments for pIII expression, Carlson and the others used a promoter, the Ppsp promoter, which is induced only after infection of the host cell by the phage. This promoter could be used for N1-N2-ligand peptide expression in SIP and successful propagation of selectively infective phage in the lagoon (**Figure 20**).

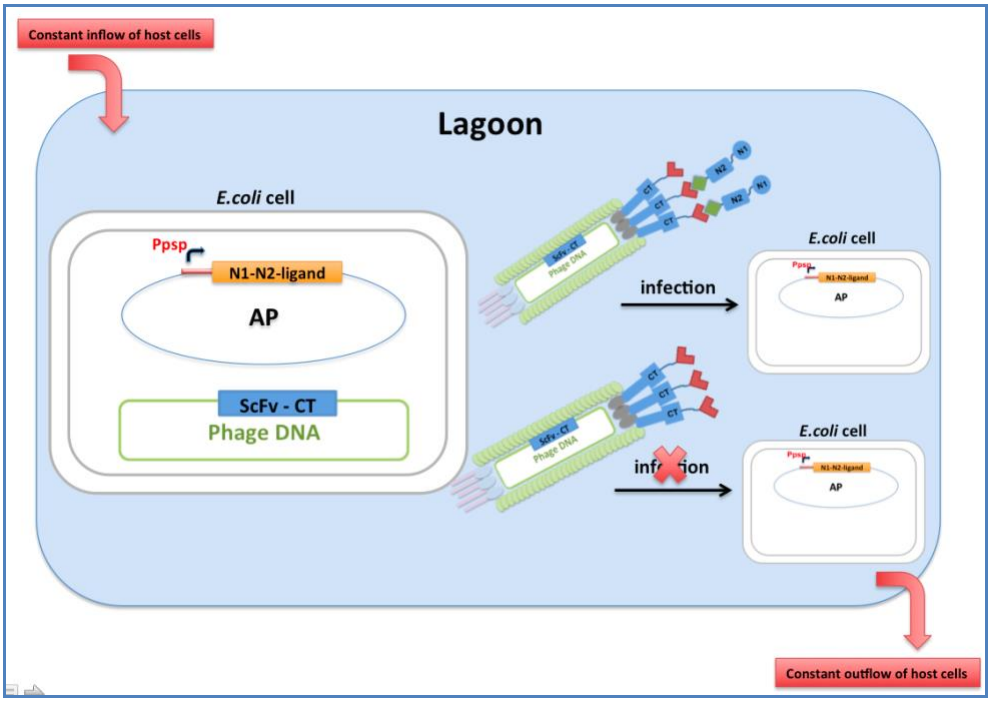


Figure 20. PACE format for affinity maturation by continuous evolution based on SIP technology. In the lagoon, the linkage of selective replication of scFv variants binding the desired ligand peptide is achieved by replacing the gene III on the phage genome with the scFv-CT coding sequence. The N1-N2-ligand coding sequence is cloned into an accessory plasmid (AP) under the control of the Ppsp promoter (phage shock promoter), which is induced only after phage infection.

2.2.2. Factors that determine the efficiency of *in vivo* SIP

As mentioned above, the efficiency of *in vivo SIP* for selections of scFv variants with high affinity was shown to be very low (Krebber C, 1995; Spada S, 1998). For the scFv 17/9 with K_D of 2×10^{-8} M, 228 selectively infective phages were produced out of 10^{10} input phages (1 infective phage out of 10^8 input phages). To apply this method for PACE, all the input phages of a protein binder with K_D as high as 2×10^{-8} M should be selectively infective. Otherwise, the selectively infective phage population would start to drop in the lagoon and the phage encoding the parent protein would not propagate. There are multiple factors that determined the efficiency of *in vivo* SIP in previous experiments that could be improved and proposed solutions are analyzed in the paragraphs below.

Different scFv variants of a phage library may present different levels of expression and degradation of the scFv-pIII protein fusion. scFv expression and degradation levels of the scFv-CT fusion when expressed in *E. coli* periplasm could be factors that impact the levels of scFv display on the phage particles and the percentage of phage that contain intact scFv fusion protein. Therefore, if the parent scFv 17/9 presents high levels of degradation (O'connell, Becerril, Roy-Burman, Daws, & Marks, 2002), an alternative binder with better expression and degradation profile could improve the efficiency of *in vivo* SIP. DARPins could be an alternative to scFvs as these are expected to have higher expressed levels in the *E. coli* periplasm and probably lower degradation levels of the DARPin-CT fusion protein. Consequently, DARPins are expected to have better phage surface display levels and could be the protein scaffold chosen for display in our continuous evolution experiments.

O'Connell (2002) observed that the scFvs display levels on phage particles was higher when a phage instead of a phagemid library was used (O'Connell D, 2002). The authors propose that this could be partially because of the use of the natural gene III phage promoter (PgIII), which would probably lead to lower but well-regulated scFv fusion expression. Increased expression of the scFv expected from the Plac promoter may lead to lower than higher phage surface expression (due to toxicity to *E. coli*). The lac promoter was used to drive the expression of the scFv-CT fusion protein for *in vivo* SIP (Krebber C, 1995; Spada S, 1998). Alternatively, the natural gene III phage promoter (Pg III) couls be tested for better scFv-CT phage surface display levels.

During selectively infective phage production the phages were produced overnight at 25°C. Is the N1-N2-HA fusion protein expected be bound to the phage-scFv after overnight incubation? This is unlikely, as even for an scFv of K_{off} in the range of 10^{-4} s^{-1}, the residence time to the ligand binding site is around 2 hours. In order to reduce the incubation time, a

protocol for phage production with shorter incubation times at higher temperatures could be examined.

When an antibiotic resistance marker gene is inserted into the phage genome, the levels of phage replication are lower than that of the wild-type phage. Different antibiotic resistance markers have different impact on phage propagation (Krebber et al., 1995). For *in vivo* SIP, chloramphenicol was used as it was found to have the lower impact on the phage titer, but there was still 100-fold decrease in the phage titer. When kanamycin was used the phage titer presented a 10^4-fold decrease, even higher than that of chloramphenicol. In PACE experiments, the improved phage vector that was used for evolution of the T7 RNAP had no antibiotic resistance marker (Esvelt et al., 2011). Use of antibiotic resistance markers may have had an impact on the selectively infective phage production, but it is necessary for the infectivity assay by which the selectively infective phage titer determination is performed. As an alternative a plaque assay instead of an infectivity assay for infective phage titer determination could be explored.

When beta-lactamase (31kDa) was inserted between the N2 and CT domains there was a 100-fold decrease in infectivity of the N1-N2-Bla-CT phage compare to the wild type phage (N1-N2-CT). The efficiency of *in vitro* SIP was shown to decrease when larger size protein binders were used (Cèbe & Geiser, 2000). This is a problem that could be addressed by applying an *in vitro* evolution approach to adapt the function of N1-N2 domains of the N1-N2-Bla-CT fusion for SIP.

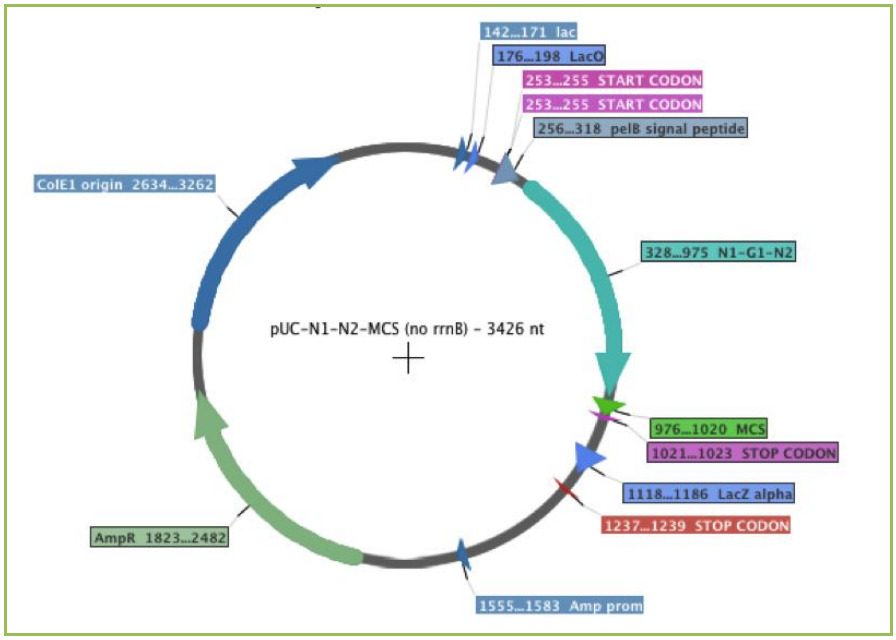
2.2.3. Experimental efforts to improve *in vivo* SIP for PACE

The first experimental step was to reproduce the *in vivo* SIP experiment using the scFv-ligand pair that has been used before by Krebber (1995) while introducing some of the changes according to the observations that were analyzed above (**section 2.2.1**).

2.2.3.1. Construction of vectors for SIP

For the N1-N2-ligand fusion protein expression in the *E. coli* periplasm the vector pUC-N1-N2-MCS (**Figure 21**) was constructed. The N1-N2 coding sequence is followed by an MCS (multi-cloning site) for cloning of the peptide ligand coding sequence. The N1-N2 amino acid sequence that was chosen to be fused to the ligand peptide is shown in the **Figure 22** (Holliger, Riechmann, & Williams, 1999). On the pUC-N1-N2-MCS, the transcription of the N1-N2-ligand fusion is under the control of the lac promoter and the lac operator and the localization of the fusion protein is directed by the signal peptide (pelB) that directs the protein to the bacterial periplasm. After the pUC-N1-N2-MCS vector construction, the HA-peptide coding sequence encoding for a 30 amino-acid long peptide, was cloned into the MCS of the vector to result in the N1-N2-HA expression vector, the pUC-N1-N2-HA (**Figure 23**). Additionally, the vector pUC-g3p was constructed. This vector contains the wild type gene III under the same expression control (pelB, lacP/O) and it could be used as a positive control in some SIP experiments or for infective phage production from phages that encode fully or partially deleted pIII.

A



B

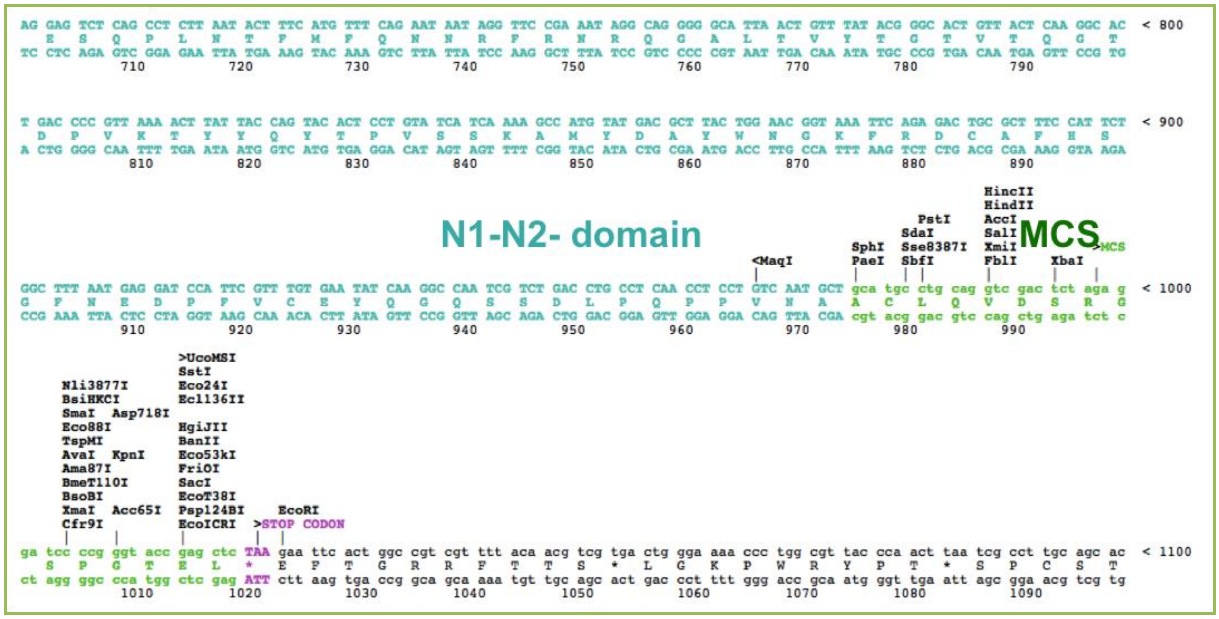


Figure 21. A) Map of pUC-N1-N2-MCS vector B) In the vector pUC-N1-N2-MCS, the N1-N2 coding sequence is followed by an MCS (multi-cloning site) for cloning of the peptide ligand coding sequence.

A

B

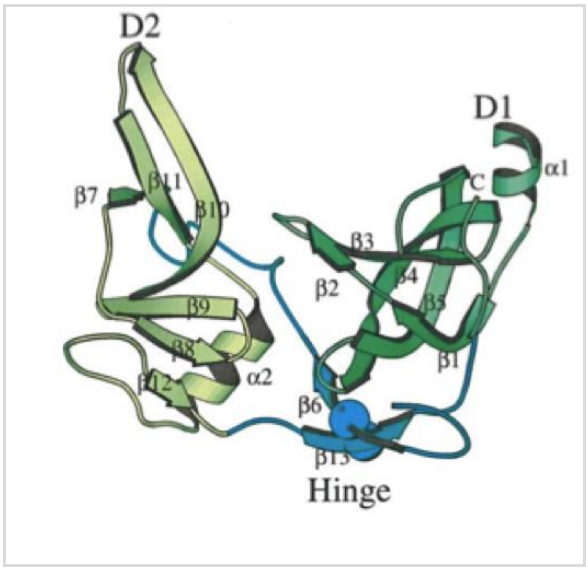


Figure 22. A) The N1-N2 amino-acid sequence that was chosen to be fused to the ligand peptide is shown between the red angles. **B)** The choice of the N1-N2 domain sequence was based on the pIII structure (Holliger, Riechmann, & Williams, 1999).

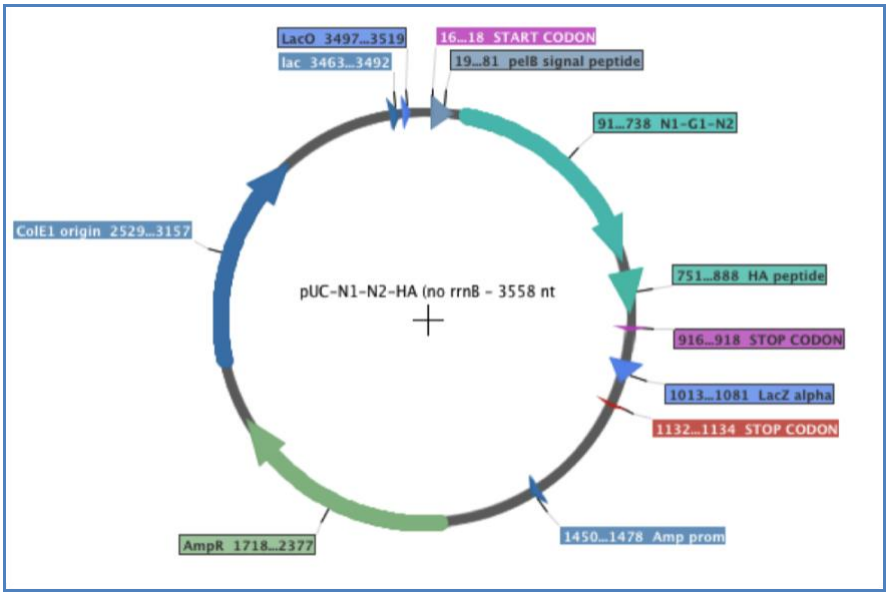


Figure 23. Map of pUC-N1-N2-HA vector. This vector was used for N1-N2-HA fusion protein expression in the *E. coli* periplasm.

2.2.3.2. Construction of recombinant phages

In contrast to the previous SIP experimental efforts, where the filamentous phage fd was used, the phage that was chosen for SIP in this study was VCSM13. VCSM13 is a helper phage designed to be packaged in phage particles with lower efficiency than a phagemid in the presence of a phagemid (mutation Met40Ile in gII). However, VCSM13 is able to replicate in the absence of phagemid DNA. In addition, VCSM13 contains the origin of replication from P15A and the kanamycin resistance gene. VCSM13 was chosen because it was the phage vector used for T7RNAP evolution by PACE before (Esvelt et al., 2011), so its ability to propagate efficiently and to keep artificially inserted large fragment of DNA intact has been proven.

The VCSM13-ScFv-CT phage vector was constructed for the scFv-CT fusion protein expression on the VCSM13 phage surface. The scFv-CT transcription on the phage genome is under the control of the wild-type gene III promoter. The scFv coding sequence was cloned downstream of the gene III signal peptide according to O'Connell (2002), who cloned a scFv library into a phage vector using the gene III signal peptide while adding a signal sequence

cleavage site that is commonly used when scFvs are expressed for phage display (QPAMA). This cloning step is usually achieved by using the restriction enzyme SfiI which recognizes GGCCNNNN^NGGCC sites and allows the insertion of the QPAMA sequence.

The scFv was cloned in a VL-linker-VH orientation as it was done before (Krebber et al., 1995). Part of the sequence of the final VCSM13-scFv-CT phage vector is shown in the **Figure 24**. Using a similar cloning strategy a recombinant phage with deleted N1-N2 domain sequence of the gene III was constructed (VCSM13-CT) for use as a negative control in SIP experiments.

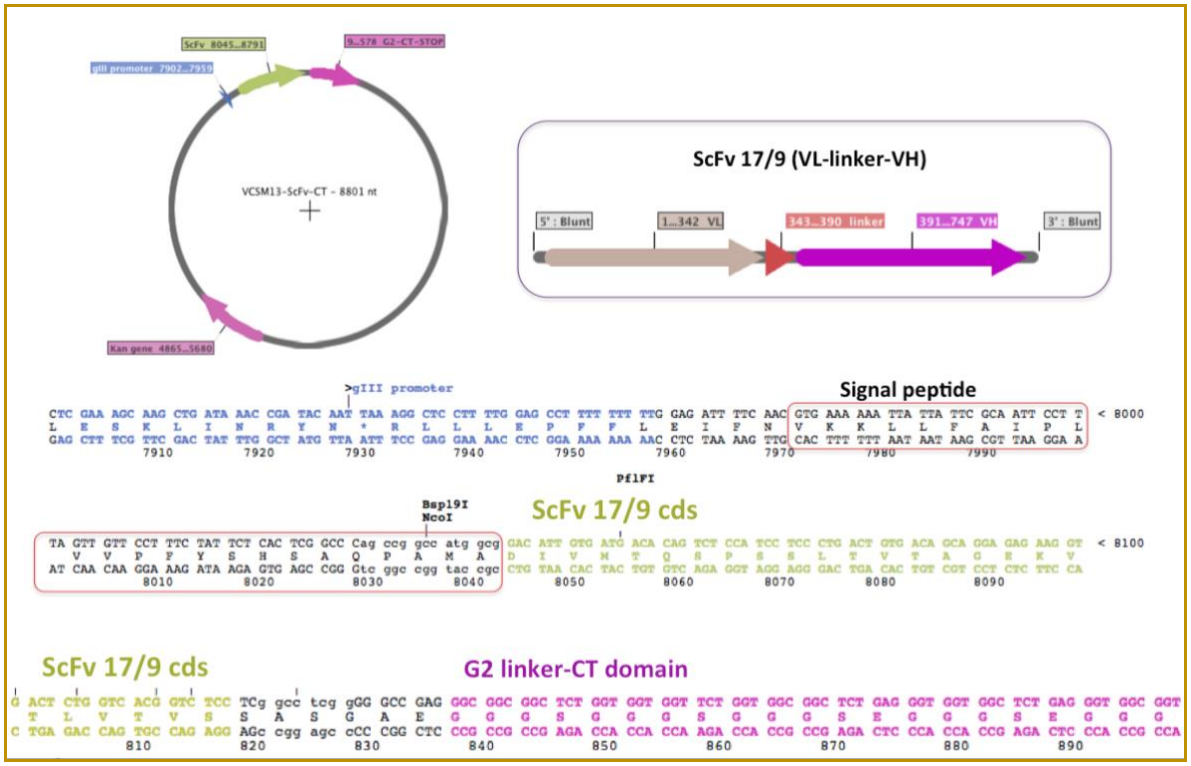


Figure 24. The design of the recombinant phage VCSM13-scFv 17/9-CT. The scFv was cloned in a VL-linker-VH orientation. The vector was designed based on the previously described scFv phage library design (O'connell et al., 2002). The scFv-CT fusion protein was directed to the *E. coli* periplasm using the gene III signal peptide and the signal sequence cleavage site QPAMA. The scFv was fused to the N-terminus of the G2 linker-CT domain.

The cloning strategy for the VCSM13-ScFv-CT phage construction is described in the Methods section. The cloning efficiency using this cloning strategy was low ($5x10^4$ cfu/ ug for *E. coli* cells, i.e. with transformation efficiency $1x10^{10}$ cfu/µg for pUC19 DNA (2×10^5-fold lower than for pUC19). Such efficiency is very low for library construction, so in the future another cloning strategy should be cnsidered for construction of a scFvs or DARPin library.

SIP experiments were performed using the *E. coli* strain S1030 (see Methods for the genotype), as this was the strain used in PACE evolution experiments (obtained from the Liu group from Harvard University).

2.2.3.3. N1-N2-HA expression in *E. coli* periplasm

The *E. coli* periplasmic fraction was obtained by freeze/thaw treatment of S1030/ pUC-N1-N2-HA strain cells. The expression of the N1-N2-HA fusion protein was induced by the removal of glucose and addition of IPTG in the culture medium. The protein contained in the periplasmic fraction were separated by SDS-PAGE and functional expression of the N1-N2-HA protein was confirmed by Western Blot using an anti-HA primary antibody (**Figure 25**), which binds to the epitope recognized by the scFv 17/9. As a negative control, the periplasmic fraction of S1030 cells produced in parallel with the S1030/ pUC-N1-N2-HA periplamic fraction was used.

The size of the protein pIII is around 60 k Da, while that of the pIII CT domain is around 23 k Da (O'connell et al., 2002). It can be concluded that the size of the N1-N2-HA fusion protein was higher than 37 k Da. Since the Western Blot using an anti-HA antibody resulted in only one band which has a size similar to the predicted size of the N1-N2-HA fusion protein, it can be inferred that the N1-N2-HA protein was functionally expressed in the *E. coli* periplasm.

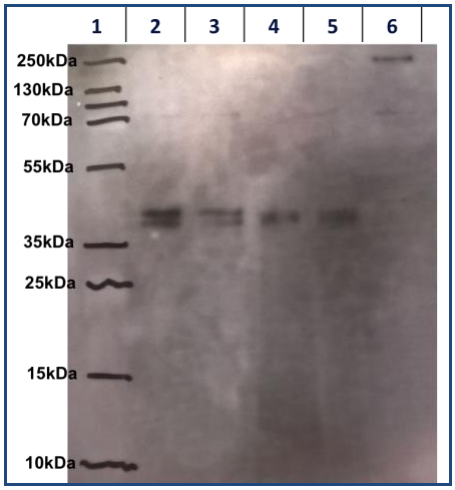


Figure 25. Western Blot of *E. coli* periplasmic fractions separated by SDS-PAGE. An anti-HA antibody was used for detection of the N1-N2-HA fusion protein. Lanes:
1. Protein ladder
2-3. S1030/ pUC-N1-N2-HA *E. coli* periplasmic fraction
4-5. S1030/ pUC-N1-N2-HA *E. coli* periplasmic fraction, with 100 mM DTT
6. Negative Control: S1030 *E. coli* periplasmic fraction

2.2.3.4. VCSM13 titration by plaque assay, infectivity assay and ELISA

VCSM13 titration by ELISA

In *in vivo* SIP, phage particles that contain N1-N2-ligand bound to CT-scFv fusion protein are called selectively infective phages. To estimate the percentage of selectively infective phage particles, a known number of input phages is used to infect *E. coli* cells and the number of colonies that grow in the presence of the appropriate antibiotic equals the number of selectively infective phages. ELISA is a method that could be used for titration of the input phage particles. Phage particles can be detected using an HRP-conjugated antibody against the major coat protein pVIII, which is present at approximately 2700 copies per phage particle and by comparing the signal of the input phages to that of serial dilutions of VCSM13 phage particles with known titer. In the **Figure 26**, VCSM13 was amplified using S1030 host cells and serial dilutions of VCSM13 phage particles in the supernatant of

unkown titer were detected by ELISA. It seems that it is possible to estimate the concentration of phages particles contained in the samples by using this method.

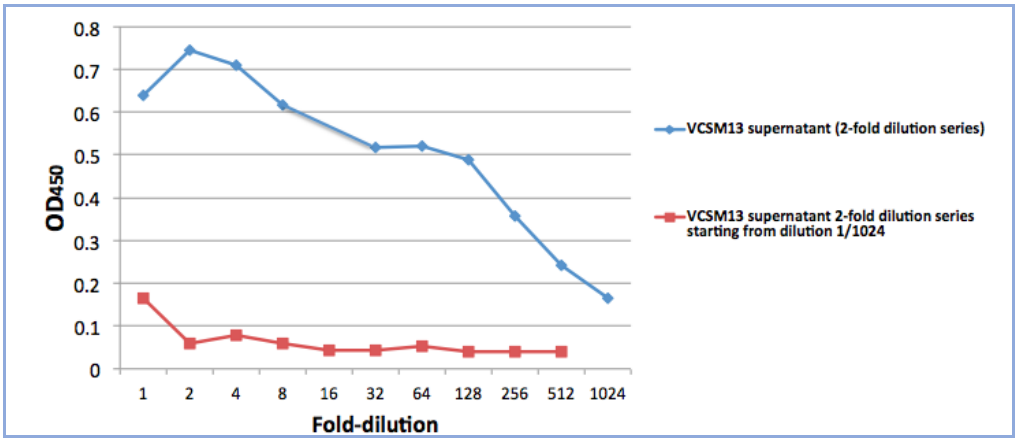


Figure 26. VCSM13 phage supernatant titration by ELISA. Starting from unknown phage titer, the blue line depicts 2-fold serial dilutions of the phage supernatant. The red line shows 2-fold serial dilutions of the same supernatant starting from already 1024-fold diluted sample.

VCSM13 titration by plaque assay

A plaque assay was performed in order to confirm that the titer estimated by plaque assay, ELISA and infectivity is the same. In the **Figure 27**, an ELISA of phage VCSM13 supernatant with titer determined by plaque assay is shown (starting from 6×10^{10} pfu/ ml, 10-fold serial dilutions). The titer of the same sample was also determined by infectivity assay. The estimation of the phage concentration obtained by the plaque assay and the infectivity assay was different, 6×10^{10} pfu/ ml and 2×10^{8} cfu/ ml, respectively.

The reason why the titer may be underestimated by the infectivity assay is not known. In phage display a similar infectivity assay is commonly used for estimation of the phage titers. The difference is that in phage display the phagemid (which is a high copy plasmid of around 4.5 kb) and not the phage genome contains the antibiotic resistance marker gene, which is also a different resistance marker (ampicillin instead of kanamycin). This is why as an alternative to infectivity assay, a plaque assay for SIP selectively infective phage determination was developed as an alternative to infectivity assay.

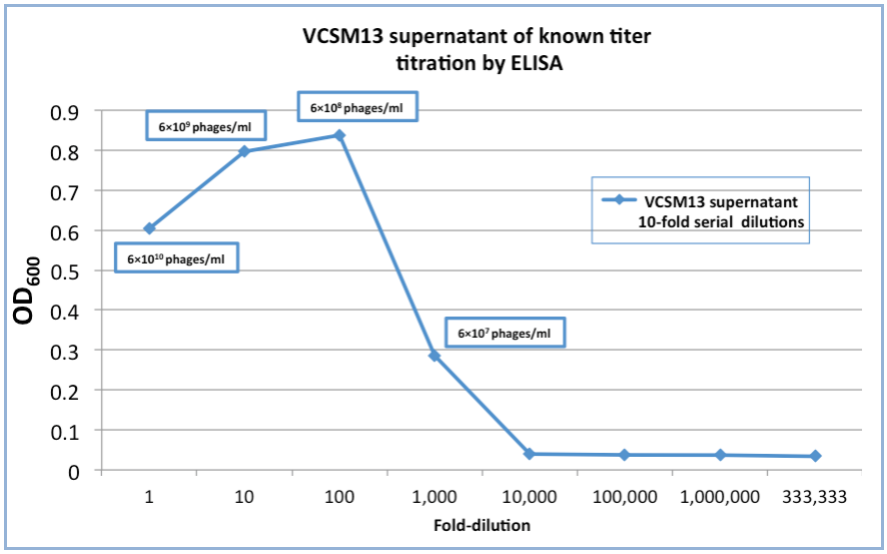


Figure 27. VCSM13 phage supernatant ELISA. Starting from 6×10^{10} phages/ ml, the blue line shows 10-fold dilutions of the sample. The titer of the same supernatant was determined by plaque assay.

2.2.3.5. Determination of scFv phage surface display levels

To assess the levels of scFv-CT fusion expression and degradation, a Western Blot of an SDS-PAGE gel of the phage supernatant could be performed. However, it was not known how many phages are required for successful detection of the scFv-CT fusion and assessment of its expression and degradation levels. Additionally, if the detection was sensitive enough, a Western Blot of the SIP input phages could be performed for detection and quantification of N1-N2-HA protein bound or not bound on the phage-scFv-CT phage.

After 8 hours of phage production (as described in the **Methods section**), the supernatant containing the phage particles produced by S1030 cells infected with VCSM13 and that of cells that had not been infected with VCSM13 were loaded into a SDS-PAGE gel. After protein separation by SDS-PAGE gel electrophoresis, the phages were detected using an anti-CT (pIII) domain antibody (**Figure 28**).

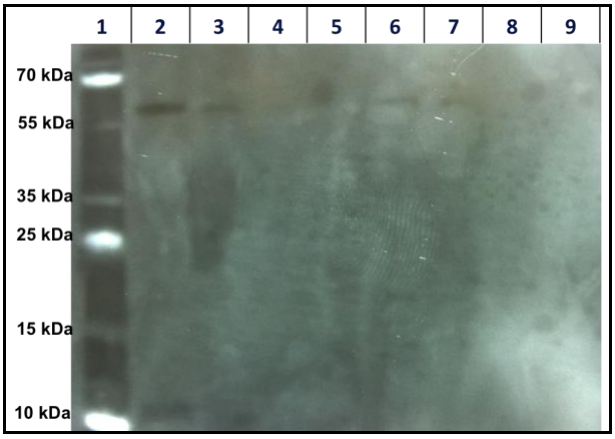


Figure 28. Western Blot of phage supernatant separated by SDS-PAGE. An anti-CT (pIII) antibody was used for detection of the pIII protein. Lanes:

1. Protein ladder
2. VCSM13 phage supernatant (no dilution, 3×10^9 phage particles)
3. VCSM13 phage supernatant (5-fold dilution, 6×10^8 phage particles)
4. VCSM13 phage supernatant (25-fold dilution, 1.2×10^8 phage particles)
5. Negative Control: Supernatant of S1030 culture, non infected with VCSM13
6. VCSM13 phage supernatant (no dilution), with 100mM DTT
7. VCSM13 phage supernatant (1/5 dilution), with 100mM DTT
8. VCSM13 phage supernatant (1/25 dilution), with 100mM DTT
9. Negative Control: Supernatant of S1030 culture, non infected with VCSM13, with 100 mM

The pIII protein size is around 60 k Da, so it was successfully detected. The surface expression and degradation levels of an scFv cannot be assessed by loading directly the phage supernatant directly onto the SDS-PAGE gel. Instead the phages were concentrated by polyethylene glycol (PEG) precipitation.

2.2.3.6. Phage production assay for SIP

As mentioned above, it is unlikely that the N1-N2-HA fusion protein is still bound to the phage- scFv during overnight production of selectively infective phages. In order to reduce the selectively infective phage incubation time before execution of the SIP infectivity assay, an alternative phage production with shorter incubation time at higher temperatures was examined, using the wild-type phage VCSM13 as a positive control.

The difference of the protocol that is outlined in the **Figure 29** compared to the protocols of previous SIP experiments is that starting from a higher OD_{600} of infected cells, so from a higher cell concentration, and growing in higher incubation temperature (37 °C instead of 25 °C), higher phage concentration in the supernatant is expected in shorter incubation time. In a shorter time of incubation more N1-N2-HA fusion protein is expected to be still bound on the phage-scFv phage particles and the selectively infective phages can be detected by an infectivity assay or plaque assay.

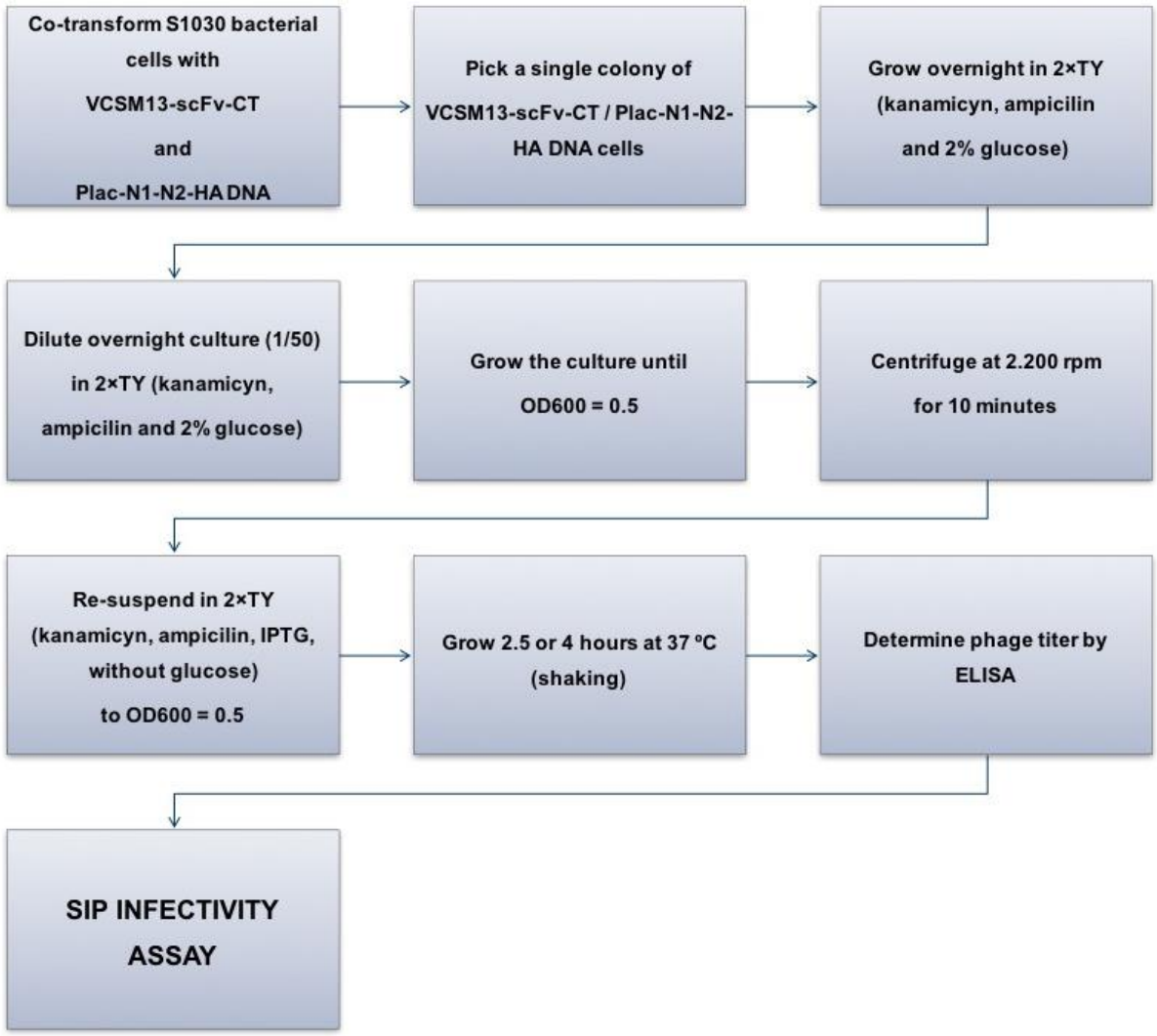


Figure 29. Schematic representation of the proposed phage production protocol for SIP.

The VCSM13 phage titer produced using this protocol was determined by ELISA (**Figure 30**). Indeed, high phage titer was achieved using this protocol (either around 5×10^8 phages/ ml after 4 hours or 5×10^7 phages/ ml after 2.5 hours of incubation). Thus, this protocol could be used for selectively infective phage production.

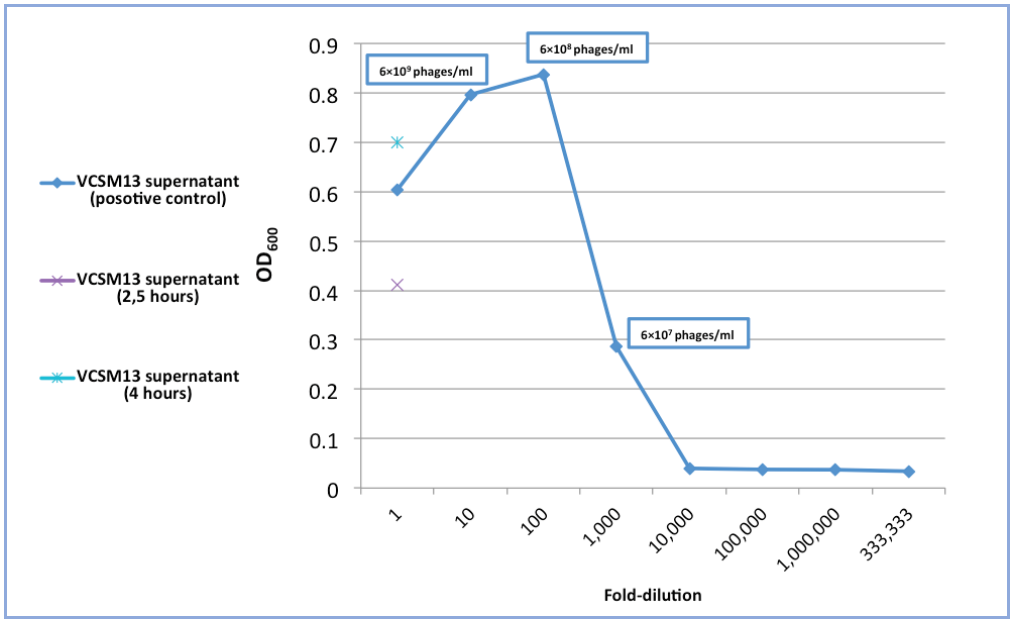


Figure 30. VCSM13 titer determination of VCSM13 phage supernatants produced following the SIP protocol after 2,5 (purple dot) or 4 hours (light blue dot) of phage production. The titer was determined by comparing with known phage titer (dark blue line) (10-fold dilution, starting from 6×10^{10} phages/ml).

2.3. Conclusions

This chapter described efforts to use SIP technology for continuous evolution by PACE (Esvelt et al., 2011). For application of this technology for PACE high efficiency is required. Otherwise, in the lagoon, the selectively infective phage population would start to drop and the phage encoding the parent protein would not propagate. The work on improving SIP technology based on the ideas described in this report was not continued because of the difficulties in improving the efficiency of the SIP technology.

Indeed a system for continuous evolution of protein binding was developed and published in Nature (Badran et al., 2016). The basis of this system was not the SIP technology, but a bacterial two hybrid system (Dove & Hochschild, 1998). But the two-hybrid system was not efficient enough for use in PACE. Thus, before using this system, they optimized the transcriptional activation >200-fold by changing several parameters including the DNA-binding domains in order to increase the protein expression level. Additionally, enhanced transcriptional activation was achieved by modifying the distance between the operator and the promoter, the DNA-binding domain multivalency state, the reporter gene ribosome-binding site, the affinity of the interaction between the promoter and the RNA polymerase and the linker between the DNA-binding domain and the bait.

3. Antibody affinity maturation using libraries with random indels

The chapters III and IV describe my efforts to develop new approaches for construction of libraries with in-frame amino acid indels that would enable sampling of random indels of different lengths across the entire antibody variable domains and develop methods that could be widely applied for antibody affinity maturation by in vitro evolution.

3.1 Introduction

3.1.1. Transposon-based methods for construction of libraries with random indels

In Nature, the percentage of indels in affinity matured antibodies seems to be low, probably because, as discussed in the Introduction chapter, SHM does not sample random but only biased diversity, with insertions occurring only by duplication. Thus, there is the possibility that the antibody fold can tolerate higher percentage of indels that could be beneficial in terms of affinity, which could be identified by sampling higher diversity of random indels by *in vitro* evolution. Other researchers have not been able to sample random and high diversity indels, either because of restrictions in the diversity of indels that can be introduced by their method (Bowers et al., 2014) or because they only used rational instead of random approaches (Lamminmäki et al., 1999).

Several attempts to develop methods for construction of libraries with random indels for *in vitro* evolution of proteins other than antibodies have been described in the literature. The polymerase pol η has been shown to introduce not only point substitutions but also deletions of two or more consecutive bases, while other polymerases, like pol b, generate only single-base indels (Emond et al., 2008). This method could be used to generate variants combining high percentage of point mutations and deletions. Another replication-based approach is RAISE (use of TdT) which is based on gene shuffling to introduce not only point mutations but also indels in the target gene (Fujii, Kitaoka, & Hayashi, 2006). Using this method to mutate TEM β-lactamase some deletions that improved the activity more than point mutations were identified. Overall, these approaches can be used to sample short deletions combined with point mutations, but do not offer the capability to incorporate high diversity of random insertions, as most of these events were single-base additions.

An approach that allows the generation of variants with insertions of variable lengths in random positions by sequence duplication, the so called tandem repeat insertion (TRINS), is

based on assembly PCR and rolling circle polymerization (Kipnis, Dellus-Gur, & Tawfik, 2012). This process is restricted by generation of tandem repeats, instead of high diversity insertions. Random insertion/deletion (RID) mutagenesis has been developed to produce random insertions of any desired length (Murakami, Hohsaka, & Sisido, 2002). This method requires several inter-molecular DNA manipulations and can be difficult to use in generating large libraries. Overall, none of these methods allow the generation of libraries with high diversity in-frame insertions.

Another approach to introduce random modifications into a target gene is by using engineered transposons. Mutagenesis using *in vitro* transposition has been used to insert penta-peptides in random positions of a target gene (Hallet, Sherratt, & Hayes, 1997; Hayes, Hallet, & Cao, 1997). This method is based on insertion of Mu transposons into a target sequence at random positions (Haapa, Taira, Heikkinen, & Savilahti, 1999), which is followed by removal of the transposon by restriction digestion. Modification of the transposon flanking sequences (transposase recognition sites; R sites), which can contain any sequence, result is a specific sequence modification at the position of the removed transposon.

The most important innovation has been developed by Jones (2005), who made an engineered Mu tranposon with a type IIS restriction endonuclease (MlyI) restriction site at the flanking sequences. The design of this transposon leads to removal of a triplet nucleotide upon MlyI digestion while the transposon is taken away from the target gene. Subsequently, the triplet can also be replaced by an alternative triplet nucleotide (Baldwin, Busse, Simm, & Jones, 2008) or the sequence of a whole domain (Edwards, Busse, Allemann, & Jones, 2008), which is achieved by using a secondary cassette that contains the desired modification.

Very recently, Stephane Emond a post-doctoral researcher in Florian Hollfelder's group, based on Jones's method, created novel cassettes that allow the deletions of double or triple

triplets (Emond et al, in preparation; **Figure 31**). Emond's major innovation is the development of a method which is based on the same principle, but with completely novel use of restriction sites and secondary cassettes, allowing the construction of libraries with insertion of triplet nucleotides of different lengths (one, two or three triplets) (Emond et al, in preparation; **Figure 32**). These methods have already been applied by Emond to study enzyme specificity by *in vitro* evolution. This chapter describes the use of these methods to build antibody libraries with high diversity of random indels which were used for *in vitro* affinity maturation.

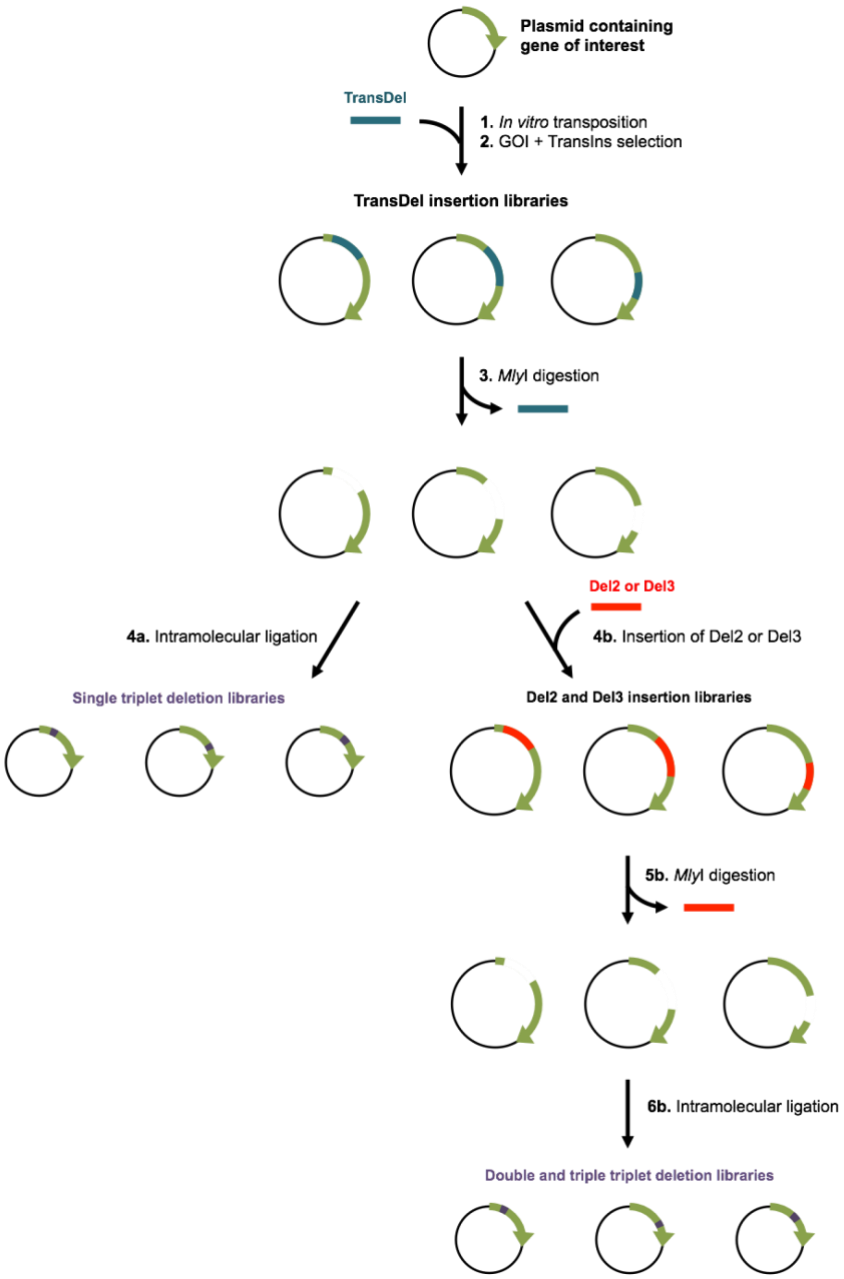


Figure 31. Schematic representation of the transposon-based approach for construction of libraries with random 3-nt (D. D. Jones, 2005), 6-nt and 9nt deletions (Emond et al, in preparation). Insertion of the transposon TransDel at random positions of the gene of interest is followed by removal of the transposon by digestion using the restriction site MlyI at the flanking sequences of the engineered transposon. Intramolecular ligation of the library plasmids results in the deletion of a single triplet nucleotide, while insertion of a secondary cassette (Del2 or Del3) followed by MlyI digestion leads to removal of either double or triple triplet nucleotides. Adapted from (Emond et al, in preparation).

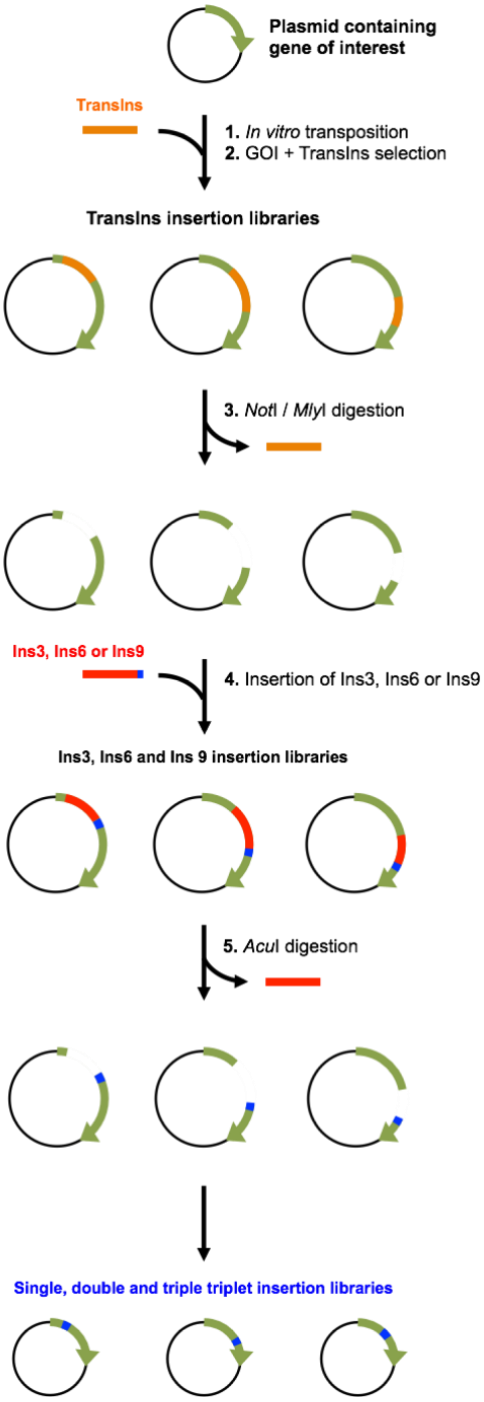


Figure 32. Schematic representation of the transposon-based approach for construction of libraries with random 3-nt, 6-nt and 9nt insertions (Emond et al, in preparation). Insertion of the transposon TransIns at random positions of the gene of interest is followed by removal of the transposon by digestion using the restriction sites MlyI/NotI at the flanking sequences of the engineered transposon. Insertion of secondary cassettes (Ins3, Ins6 or Ins9), which contain NNN triplet insertions followed by AcuI digestion leads to addition of single, double or triple triplet nucleotide insertions. Adapted from (Emond et al, in preparation).

3.1.2. BAK1 as a model antibody for *in vitro* affinity maturation

To study the impact of indels on antibody affinity by *in vitro* evolution, I used as a model the anti-IL-13 antibody BAK1. By choosing BAK1 as a model for indel-based affinity maturation I hoped to compare and contrast the outcomes of *in vitro* evolution using either indels or point substitutions and to test whether the two strategies might synergise.

Previously, BAK1 has been affinity-matured by Thom et al. (2006) using substitution-based approaches. To achieve higher affinity both random and targeted point mutagenesis approaches were used for library construction. Except from the classic approach of construction of targeted V_H CDR3 libraries, the design of libraries targeting other CDRs was guided by scanning the entire BAK1 scFv sequence for mutations that may be related with affinity gains (mutational hotspots). The scanning was performed by ribosome display combined with ep-PCR mutagenesis between the rounds of selection. Identification of mutational hotspots was followed by construction of libraries targeting these hotspots with the full diversity of amino acid substitutions. These libraries also contained combinations of beneficial mutations in different CDRs to take advantage of additive effects on affinity.

The majority of beneficial substitutions were not found in the CDR3s but in the V_L CDR1, underlining the importance of targeting all CDRs to find beneficial point substitutions. Furthermore, the majority of the mutation hotspots were located in the CDRs (84%), which is in agreement with preferential selection of CDR mutations during SHM of human antibodies in Nature. This evolution campaign gave rise to a variant with two point substitutions that conferred large improvement in affinity and improved potency, BAK1.1 (Thom et al., 2006), which has subsequently entered the clinical trials as a therapeutic candidate for the treatment of asthma (May et al., 2012). BAK1.1 has two point substitutions in the loops V_H CDR3 (N99S) and V_L CDR1 (N27I) (**Figure 33**). Hotspots were also identified in other CDRs, the V_H CDR1 and the V_H CDR2. Other variants, with small improvements compared to BAK1.1

(≈2-fold), were generated by a CDR walking mutagenesis approach by combining the BAK1.1 mutations with mutations in the V_H CDR1 or V_H CDR2. Although additionally to the BAK1.1 mutations, the highest affinity variant (BAK1.45) has several mutations in the V_H CDR2, the major improvements in affinity were attributed to the BAK1.1 mutations, which is the variant that was chosen to proceed in clinical tests.

Recently, a co-crystal structure of BAK1.1 with human IL-13 was determined (Popovic et al., 2016), which shows that it is mostly V_L regions that are driving paratope interactions; the V_L CDR2 provides residues that form a network of hydrogen bonds and a salt bridge in the core of binding. The paratope is composed of residues of the V_L CDR1, V_L CDR2, V_L CDR3, V_L FW3 and V_H CDR3.

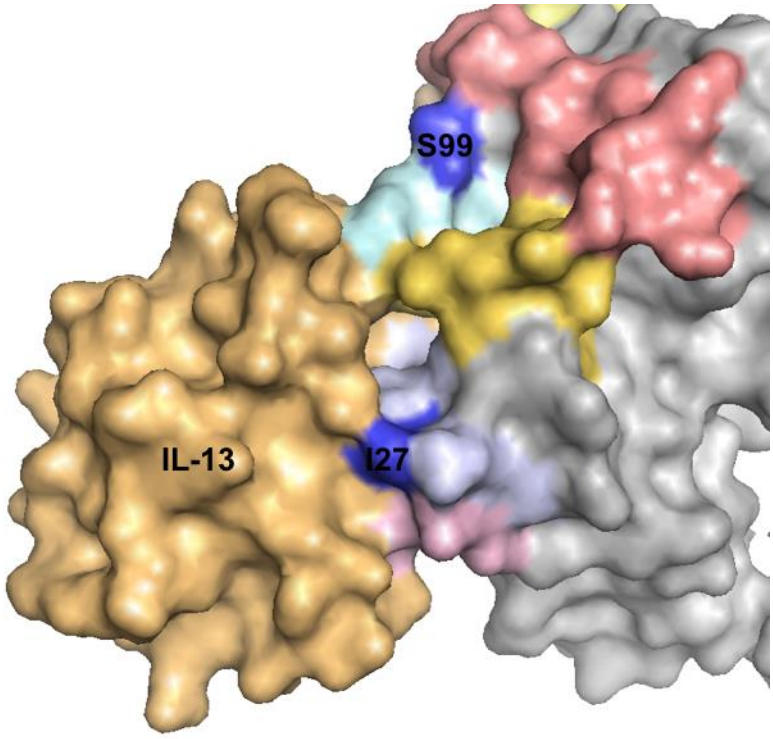


Figure 33. BAK1.1 co-crystal structure with IL-13. The BAK1.1 variant has two substitutions (in blue) that improve BAK1 affinity. The V_H CDR1, V_H CDR2 and V_H CDR3 loops are coloured in wheat, salmon and pale cyan, while the V_L CDR1, V_L CDR2, V_L CDR3 in blue white, pale green and yellow orange, respectively. IL-13 is shown in light orange.

3.2. Results

3.2.1. Library construction

BAK1 scFv libraries with random indels throughout the entire antibody V-gene were constructed using the transposon-mediated approaches developed by Emond (Emond et al., in preparation) (**Fig. 34 and 35**). The BAK1 libraries 3nt-Del, 6nt-Del, 9nt-Del and 3nt-Ins contain 3, 6 and 9 nucleotide deletions and 3 nucleotide insertions (triplet NNN), respectively. Although it is possible to produce libraries with insertions of different lengths (3, 6 and 9nt), only the library with small insertions (3nt-Ins) was constructed, because it was considered important to build libraries that cover the diversity of all the possible variants. For a scFv coding sequence that is comprised of 750 bp, the number of all possible variants with single, double or triple amino acid insertions is 4.8×10^4, 3×10^6 and 2×10^8, respectively. Since the library size that can be achieved with this method is $\approx2\times10^6$ members, it would be difficult to cover the full diversity of a library with insertions longer than a single amino acid. It is technically possible to reach a 10-fold larger library size, suggesting that someone could build a high diversity double amino acid insertion library.

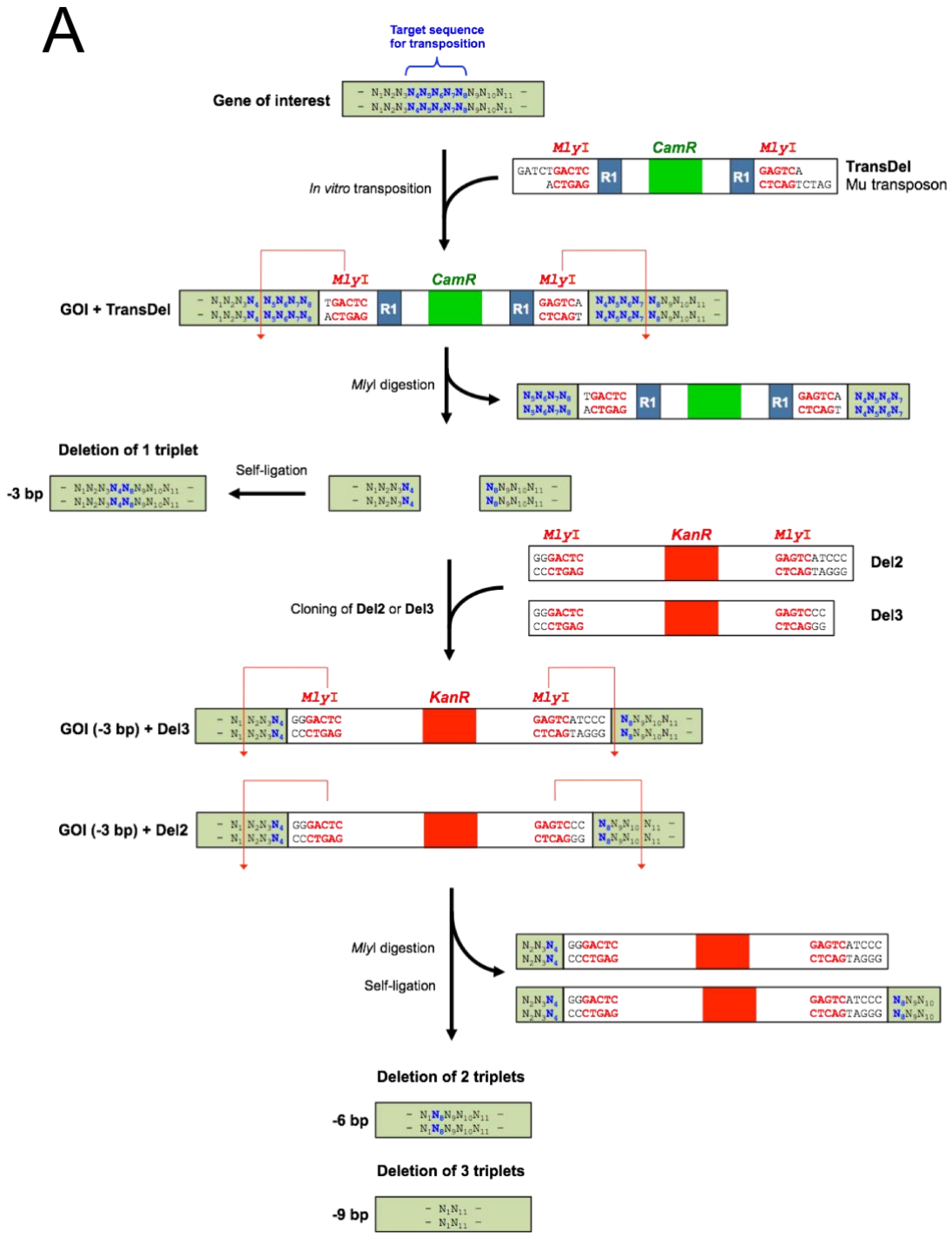


Figure 34. Schematic representation of the process of construction of libraries with single, double or triple triplet deletions (3nt-Del, 6nt-Del and 9nt-Del) of a target gene of interest. The engineered TransDel transposon, which contains type IIS restriction endonuclease (MlyI) flanking sequences, is inserted at random positions of the gene of interest by *in vitro* transposition. The TransDel is designed in a way that removal of the transposon by Mly digestion and self-ligation of the library plasmids results in deletion of a triplet nucleotide. For construction of the 6nt-Del or 9nt-Del libraries, cloning of a secondary cassette (Del2 or Del3) followed by digestion with MlyI is used to remove double or triple triplet nucleotides. Adapted from (Emond et al, in preparation).

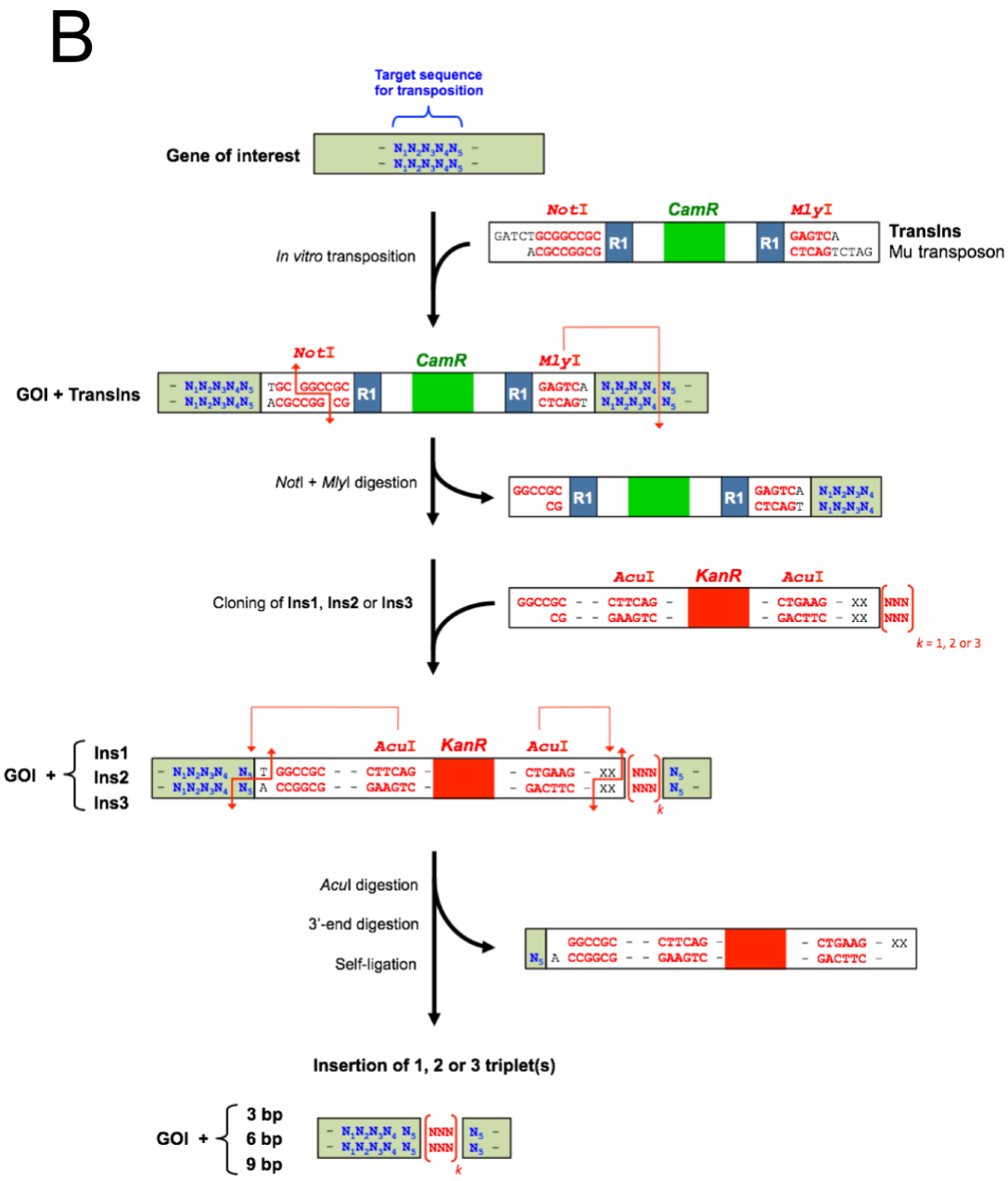


Figure 35. Schematic representation of the process of construction of libraries with single, double or triple triplet insertions (3nt-Ins, 6nt-Ins and 9nt-Ins) of a target gene of interest. The engineered TransIns transposon, which contains MlyI/NotI flanking sequences, is inserted at random positions of the gene of interest by *in vitro* transposition. Removal of the transposon by MlyI/NotI digestion is followed by cloning of an MlyI/NotI cassette (Ins1, Ins2 or Ins3) which contains single, double or triple NNN triplet insertions, respectively. Following digestion with the type IIS restriction endonuclease AcuI, the 3' overhangs are removed by using the Large (Klenow) Fragment of the DNA Polymerase I (3' -> 5' exonuclease activity). Self-ligation of the plasmids leads to insertion of single, double or triple triplet nucleotides. Adapted from (Emond et al, in preparation).

The quality of the libraries was assessed by sequencing of 88 variants from each library and is shown in **Table 1**: the 3nt-Del, 6nt-Del, 9nt-Del and 3nt-Ins libraries contained 86.4, 17, 29 and 35.6 % variants with indels of the desired length, respectively. For each library, the functional library size was calculated by multiplying the total percentage of variants with in-frame indels with the actual library size. The fact that the functional library sizes (1.7×10^6, 3.4×10^5, 5.8×10^5 and 7.1×10^5, respectively) were at least 10-fold higher than the theoretical library sizes (7.5×10^2, 7.5×10^2, 7.5×10^2 and 4.8×10^4, respectively) showed that the percentages of variants with in-frame indels was high enough to cover the theoretical diversity of each library, with actual size $\approx2\times10^6$. The percentage of in-frame indel variants with extra point mutations across the scFv sequence was negligible (0-1.5%). As a consequence, epistatic interactions between indels and random point substitutions should not affect the outcome of the selection, as deleterious or beneficial point substitutions would not mask the effect of beneficial indels.

Furthermore, the sequence analysis (**Table 1**) detected only very few repeats of the same sequence, with very high (> 90%) percentage of variants with indels at unique positions, demonstrating that indels were randomly distributed across the entire scFv sequence and the gene is fully covered. Furthermore, observation of the amino acid residues of the 3nt-Ins library did not reveal any amino acid biases (**Figure 36**): we observed that 100% variants possessed unique insertions, with 14/28 insertions of unique amino acid residue.

Overall, the quality control suggests that libraries with random deletions and high diversity insertions across the entire scFv gene without extra point substitutions were successfully generated and that this custom-made diversity in sequence space can be explored in library selections.

Table 1. Sequence analysis[a] of BAK1 libraries with random indels.

BAK1 indel library	Variants with in-frame indel of desired length %		Variants with frameshifts %	Parent (or linker) %	Variants with indel at unique position %	Variants with unique amino acid insertion %	Theoretical library size	Functional library size
Number of nucleotide (nt) Insertion (Ins) / Deletion (Del)	Total	With extra point mutations						
3 nt-Del	86.4	1.5	13.6	0	90	NA	7.5×10^2	1.7×10^6
6 nt-Del	17	0	67.1	15.9	93	NA	7.5×10^2	3.4×10^5
9 nt-Del	29	0	52	19	83	NA	7.5×10^2	5.8×10^5
3 nt-Ins	35.6	0	47.4	17	100	100	4.8×10^4	7.1×10^5

[a] The analysis was based the sequences of 88 variants from each library. For each library, the functional library size was calculated by multiplication of the total percentage of variants with in-frame by the actual library size, measured by colony number after cloning.

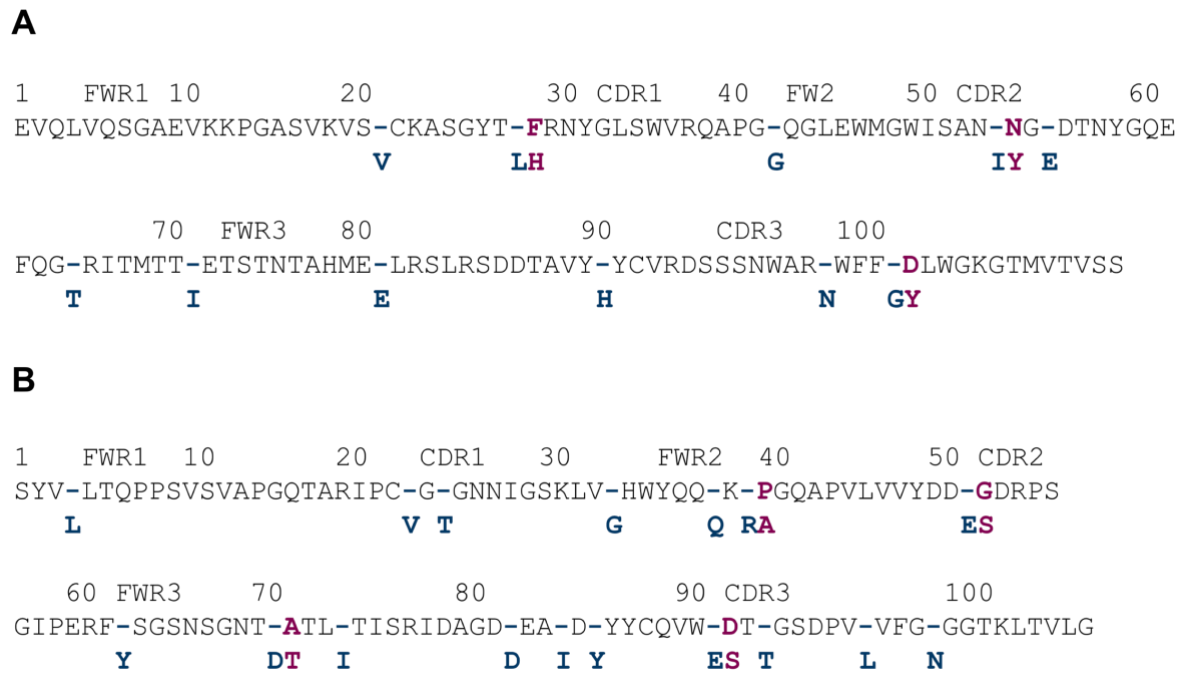


Figure 36. BAK1 library with random single amino acid insertions. Insertions (blue dashes) are randomly distributed across the entire **(A)** V_H and **(B)** V_L regions. The amino acid residues of the insertion or substitution are shown in blue or dark red, respectively.

3.2.2. Affinity-based selections of recombined BAK1 indel libraries

To identify indel variants with improved affinity, panning selections of the BAK1 indel libraries were performed by ribosome display (J Hanes & Plückthun, 1997). **Figure 37** shows an overview of the process that was followed to identify variants with improved affinity from the BAK1 libraries with random indels. Ribosome display was the method of choice because the PCR step before each selection round enables recombination and sampling of very large library sizes (more than 5×10^{11} variants). By contrast, phage display would have only afforded 10^{9} - 10^{10} variants and compartmentalised droplet selections (e.g. SNAP display) would have allowed selections from libraries on the order of 10^{8}.

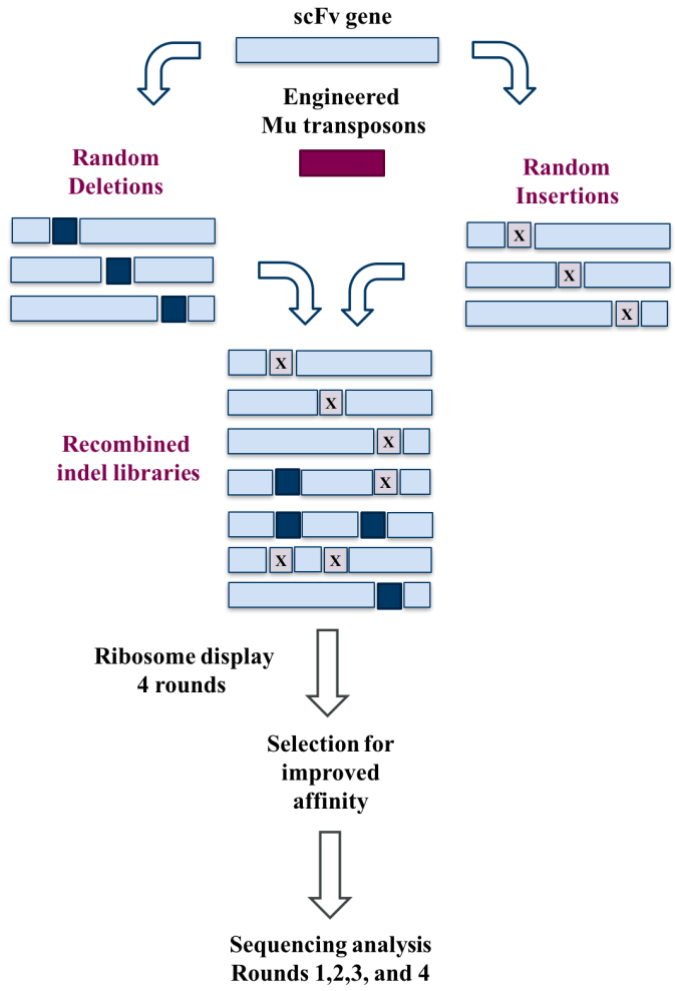


Figure 37. Overview of the process that was followed for identification of improved variants, by recombination of BAK1 libraries with random indels and selection by ribosome display.

To take advantage of potential synergistic effects between indels, variants with more than one indel per gene were created before the first round of selection, by recombination of a mix of the BAK1 indel libraries (3nt-Del, 6nt-Del, 9-nt Del and 3nt-Ins) using the staggered extension process (StEP) *in vitro* recombination (Zhao, 1998). Based on sequencing analysis of 176 variants of the recombined library (**Table 2 and Figure 38**), the input of the selection round 1 was composed of 31.3% parent, 26.9% in-frame single indel variants, 9.2% in-frame double indel variants and 32.6% out-of-frame indels variants. Thus, the functional library sizes of single (1.4×10^{11}) and double (4.6×10^{10}) indel variants were at least 10-fold higher than the theoretical library sizes (5×10^{4} and 1.26×10^{9}), which suggests that not only all possible single variants, but also high number of all possible double indel variants were present in the recombined library. The large diversity of double indel variants in the recombined library shows that the recombination step allowed a search for synergistic interactions between indels.

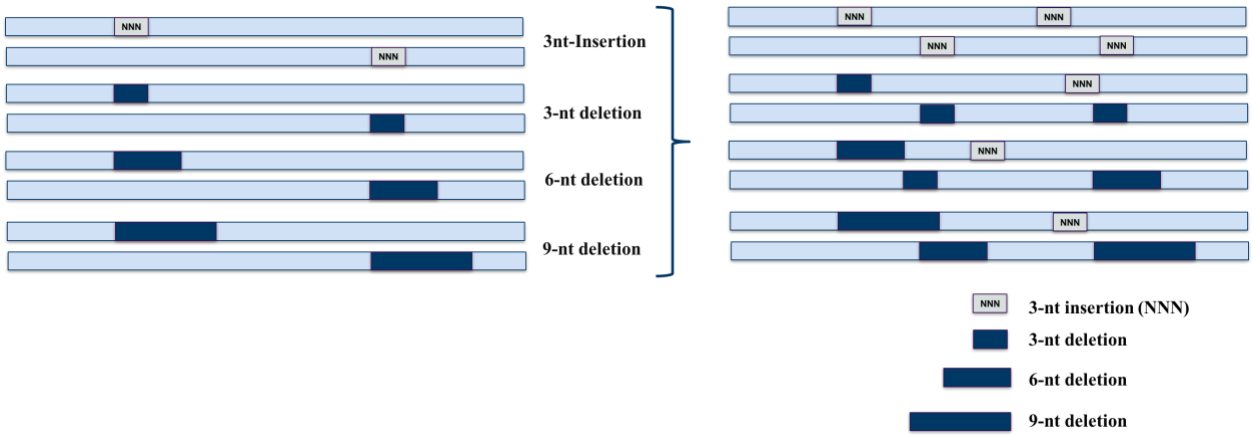
Four rounds of high stringency selections were performed by panning with recombinant biotinylated IL-13 concentrations below the K_D, using the lowest possible IL-13 concentration in each round. The concentrations that were used were 100, 20, 20 and 20 pM for rounds 1, 2, 3 and 4, respectively.

Table 2. Sequence analysis[a] of recombined BAK1 indel library mix.

	Percentage	Number of possible indel variants	Functional library size	Actual library size
Parent	31.3		1.6×10^{11}	
Single indel variants (in-frame)	26.9	5×10^4	1.4×10^{11}	5×10^{11}
Double indel variants (in-frame)	9.2	1.26×10^9	4.6×10^{10}	
Out-of-frame indel variants	32.6		16.3×10^{10}	

[a] The analysis was based on the sequences of 176 variants from each library. The number of possible single indels variants was the sum of possible variants of each indel library (see Table 1). The functional library size was calculated by multiplication of the percentage of in-frame variants with the actual library size.

A



B

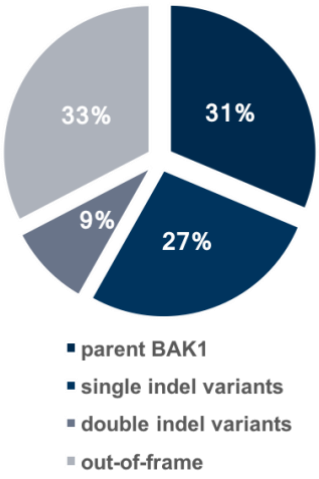


Figure 38. A) Recombination of a mix of the BAK1 indel libraries (3nt-Del, 6nt-Del, 9-nt Del and 3nt-Ins) created variants with more than one indel per gene. **B)** The recombined BAK1 indel library contained high percentage of single and double indels variants.

3.2.3. Sequencing analysis of selection outputs

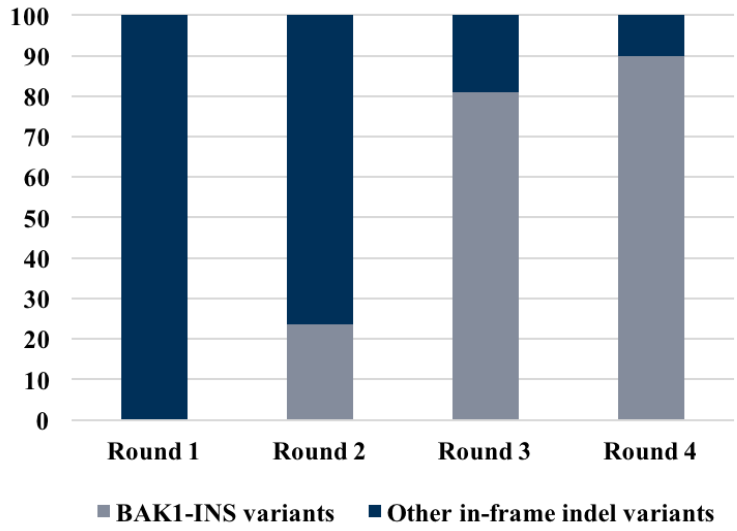
The highest affinity variants were identified by assessing the enrichment of individual variants from the analysis of the sequences of the selection outputs (176 variants from each output) from rounds 1 to 4. Between the selection rounds, an alternative PCR protocol was followed to keep the mutagenesis rate low and avoid contamination of the selection with high affinity variants with point substitutions by using a very high fidelity DNA polymerase instead of the low fidelity polymerase which is usually used for ribosome display selections. Although the percentage of BAK1 variants with point substitutions in round 4 was low, four variants with single point substitutions were enriched (**Table 3**). Two of these point substitutions were found before to improve BAK1 affinity, BAK1.4 (V_L CDR1 N27Y) and BAK1.0 (V_H CDR3 N99S) and the combination of these two point substitutions make up the variant BAK1.1, suggesting that the results are in agreement with the previous outcome of ribosome display selections of BAK1 (Thom et al., 2006). In addition, two more variants with single point substitutions in FWR positions were identified, the V_H FWR2 E46K and V_H FWR3 M80R, which could potentially have a beneficial effect on BAK1 affinity.

Among the in-frame indel variants, a group of variants with an insertion and a substitution at the same position within the V_L FWR3 (BAK1-INS variants, **Figure 39**) were enriched over the rounds, reaching 98% of the total indel variants in round 4; a single amino acid insertion after residue S67 and a substitution of G68. The other in-frame indel variant that was found in round 4 contained a G52a insertion in the V_H CDR2, after the residue S52. The amino acid types that were enriched were insertion of 67aE, 67aV, 67aD or 67aA combined with substitution G68W or G68R. One variant, BAK1-INS1, dominated the selection outcome in round 4 (41%) (**Figure 40**): it has an 67aE insertion and a G68W substitution. Assuming that the most enriched variant had the highest affinity, the variant BAK1-INS1 was converted to an IgG1 format and was characterized in terms of kinetics.

Table 3. Variants with single point substitutions in selection round 4.

Loop	VH FWR2	VH FWR3	VH CDR3	VL CDR1
Residue	**46**	**80**	**99**	**27**
BAK1	E	M	N	N
BAK1.4	E	M	N	**Y**
BAK1.0	E	M	**S**	N
BAK1.01	E	**R**	N	N
BAK1.02	**K**	M	N	N

A



B

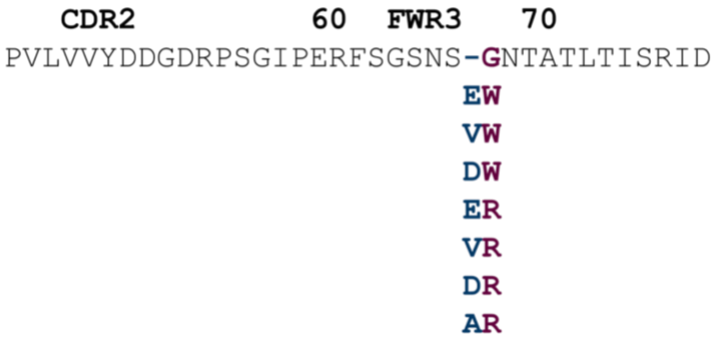


Figure 39. Selections of the recombined BAK1 indel libraries. A group of variants with different amino acid insertion and a substitution at the same position in the V_L FWR3 (BAK1-INS variants) were enriched, reaching 90% of the total indel clones in round 4.

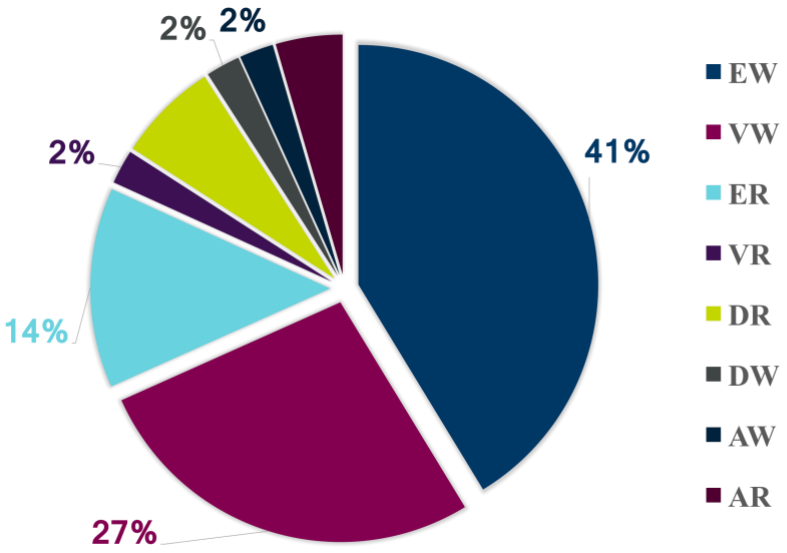


Figure 40. The pie chart shows the predominance of the variant BAK1-INS1 in round 4 (42%), with an insertion V_L 67aE and a substitution G68W. Sequence numbering according to Kabat (T. T. Wu & Kabat, 1970).

3.2.4. Beneficial effect of insertion in the FWR3

To find out if the insertion of the BAK1-INS1 variant conferred an improvement in affinity, the K_D values of the BAK1 IgG variants, which were constructed by site-directed mutagenesis, were measured by surface plasmon resonance using the Biacore (**Table 4**). 14.3-fold higher affinity (K_D=140 pM) was observed for the insertion variant compared to the parent (K_D=2,300 pM). No gain in affinity was observed in variants bearing either substitution G68E or G68W (BAK1-E and BAK1-W), suggesting that the G68W was beneficial only in the presence of the 67a insertion. This observation supports the conclusion that the insertion (67a) was the affinity improving feature of BAK1-INS1. The magnitude of improvement was significant, only 2.8-fold lower than that of the variant with two beneficial substitutions (BAK1.1; K_D = 50 pM).

Although previously the entire BAK1 scFv was scanned for point mutations that improve affinity, no change in affinity was attributed to point substitutions in the V_L FWR3. The observation of affinity-improving indels in this region suggests that an insertion in FWR3 but

no point substitutions enabled large effect on BAK1 affinity. This is the first example of a deliberate library strategy in which insertions complement the effect of substitutions on antibody affinity.

Table 4. Binding kinetics* and thermodynamics** of BAK1 variants

Loop	VH CDR3	VL CDR1	VL FWR3				K_D (pM)	Fold improvement	k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})	ΔΔG (kcal/mol)	ΔΔG1 (kcal/mol)
Residue	99	27	67	67a	68	69						
BAK1	N	N	S	-	G	N	2,300	1	5.6×10^6	1.3×10^{-2}	na	na
BAK1-E	N	N	S	-	E	N	17,100	0.13	3×10^6	4.4×10^{-2}	nd	na
BAK1-W	N	N	S	-	W	N	1,400	1.6	6×10^6	8.7×10^{-3}	nd	na
BAK1.1	S	I	S	-	G	N	50	46	6.2×10^6	3.8×10^{-4}	-3	na
BAK1-INS1	N	N	S	E	W	N	140	14.3	7.1×10^6	1×10^{-3}	-1	na
BAK1-INS1.1	S	I	S	E	W	N	9	256	2.9×10^6	2.5×10^{-5}	-4.9	-0.9

* K_D values of the human IgG1 variants were determined by surface plasmon resonance using the Biacore. The fold improvement in K_D is the ratio of $k_{off}/k_{on.}$ **ΔΔG was calculated from the equations $\Delta G = - RT \ln (K_D^{Parent} / K_D^{Mutant})$ and $\Delta GG_{AB} = \Delta GG_A + \Delta GG_B + \Delta GG1$

3.2.5. Specific amino acid substitution enabled the function of the insertion

The fact that the variant with the highest frequency in round 4 (BAK1-INS1) had a large improvement in affinity confirmed that under the panning conditions of the ribosome display selections, variants with higher affinity than the parent were enriched. This in turn suggests that enrichment of the other BAK1-INS variants was related to affinity improvements. While the presence of the variant 67aE/G68 was in the library was confirmed (data not shown), only the mutant 67aE/G68W (41%) was enriched, suggesting that the insertion was beneficial only in the presence of the 68W substitution. This observation leads to the conclusion that the specific amino acid substitution was essential for the improvement in affinity. This idea is supported by the fact that other variants with G68W were also enriched, 67aV/G68W and 67aA/G68W with 27% and 2.3% in round 4, respectively. Our data do not allow the comparison of these mutants with the 67aE/G68W that would allow us to determine if the

67aE substitution conferred even higher improvement of the variant with 68W substitution, as the K_D values of all these variants have not been measured individually.

Taken together, these results demonstrate that the insertion 67a is beneficial only in the presence of specific amino acid residue (W) at position 68, indicating that specific point substitution of a residue at position flanking the insertion enabled the insertion to exert its function (permissive mutation). We speculate that to identify this variant we had to sample different variants with each of 20 amino acids at this position (68), which underline the importance of sampling a large diversity of insertions and point substitutions around the position of the insertion.

3.2.6. Combined effect of beneficial insertions and substitutions on antibody affinity

To study epistatic interactions of insertions and point substitutions on antibody affinity, we determined the combined effect of the beneficial insertion (BAK1-INS1) and two beneficial substitutions (BAK1.1) on BAK1 affinity (**Figure 41 and Table 4**). Combination of both the insertion and substitutions resulted in a variant (BAK1-INS1.1; K_D = 9 pM) with higher affinity than that with either only substitutions (BAK1.1; K_D = 50 pM) or only the insertion (BAK1-INS1; K_D = 142 pM). BAK1-INS1.1 possessed an overall 256-fold improvement in affinity compared to the parent, which is 6.3-fold higher than that of the substitution variant BAK1.1, with a 6.5-fold lower dissociation constant k_{off} (2.45×10^{-5} s^{-1}). The effect of combining insertion and substitutions was synergistic, as the free energy change that resulted from combining insertion and substitutions (BAK1-INS1.1; -4.9 kcal/mol) was higher than the sum (-4 kcal/mol) of the free energies of the variants with insertion (BAK1-INS1; -1 kcal/mol) and substitutions (BAK1.1; -3 kcal/mol), while for additive mutations the free energy change would be equal to the sum. These results demonstrate that beneficial insertions and point substitutions can have additive effects on affinity, thus for engineering purposes

insertions could be included in CDR walking mutagenesis approaches, as selections of libraries with insertions and point substitutions could be performed in parallel and then combined to exploit additive effects on affinity.

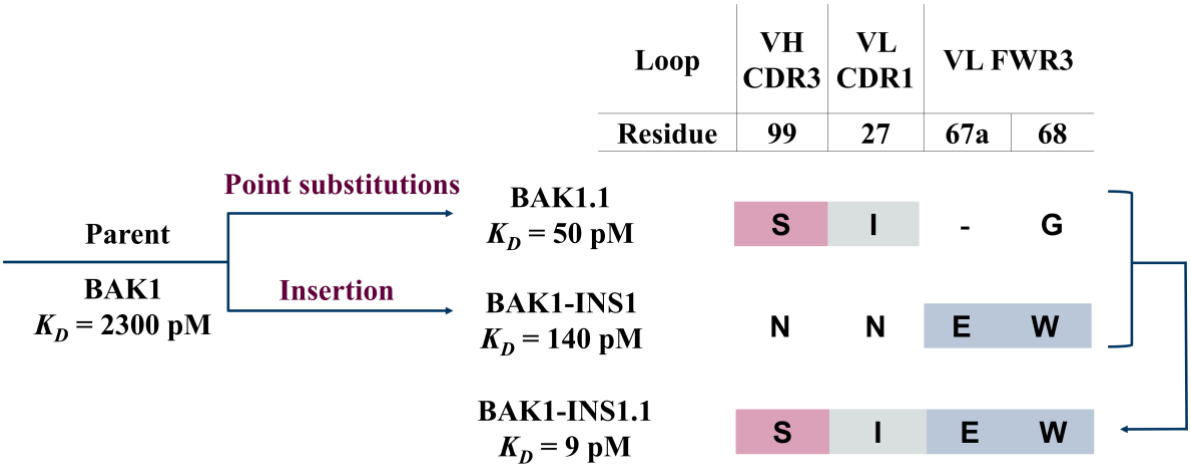


Figure 41. Comparison of the binding kinetics of a BAK1 variant with insertion in the V_L FWR3 (BAK1-INS1), a variant with two beneficial substitutions (BAK1.1) and a variant combining both the INS1 insertion and BAK1.1 substitutions.

3.2.7. Beneficial insertion in the FWR3 loop

The insertion is lodged within the fourth loop into the third framework region (FWR3) (**Figure 42**), sometimes referred to as CDR4 (Fanning & Horn, 2011). In contrast to the other hypervariable loops (CDR1-3), antibodies in Nature show no variation in this loop (Chothia & Lesk, 1987). However, it has been shown that this loop is tolerant to variation (Simon & Rajewsky, 1992) and some anti-HIV bnAbs have insertions in this loop, including PGT121, which originated from the same germline gene (L-Vλ3–21*02) as the BAK1 V_L and has a three amino acid-insertion in the same position as BAK1-INS1 in the FWR3 (Mouquet et al., 2012).

The crystal structure of the variant BAK1.1 in complex with human IL-13 was determined by Popovic et al. (2016). BAK1.1 has two point substitutions in the loops V_H CDR3 (N99S) and V_L CDR1 (N27I), improving BAK1 affinity in combination by 46-fold (**Table 4, Figure 41**), located on the periphery of the antigen-binding site (**Figure 42b**). The position of the single amino acid insertion in the V_L FWR3 loop (67aE/G68W) is close to the antigen-binding site (**Figure 42**), with residues of this loop (S67 and G68) constituting part of the structural paratope (Popovic et al., 2016). Despite the close proximity of these residues to the antigen surface, the only residue that possibly forms a hydrogen bond with antigen is the residue before the positions of the insertion S67 (side chain oxygen), which is 3.5 Å far from the T20 (side chain oxygen), while the residue after the positions of the insertion (G68) does not seem to form contacts with the antigen, as the side chains of these positions (main chain nitrogen from T20 side chain oxygen) are at least 5.4Å far from atoms of the antigen. These observations show that the site of the affinity-enhancing insertion is in close proximity to the antigen-binding site on a partially surface exposed location.

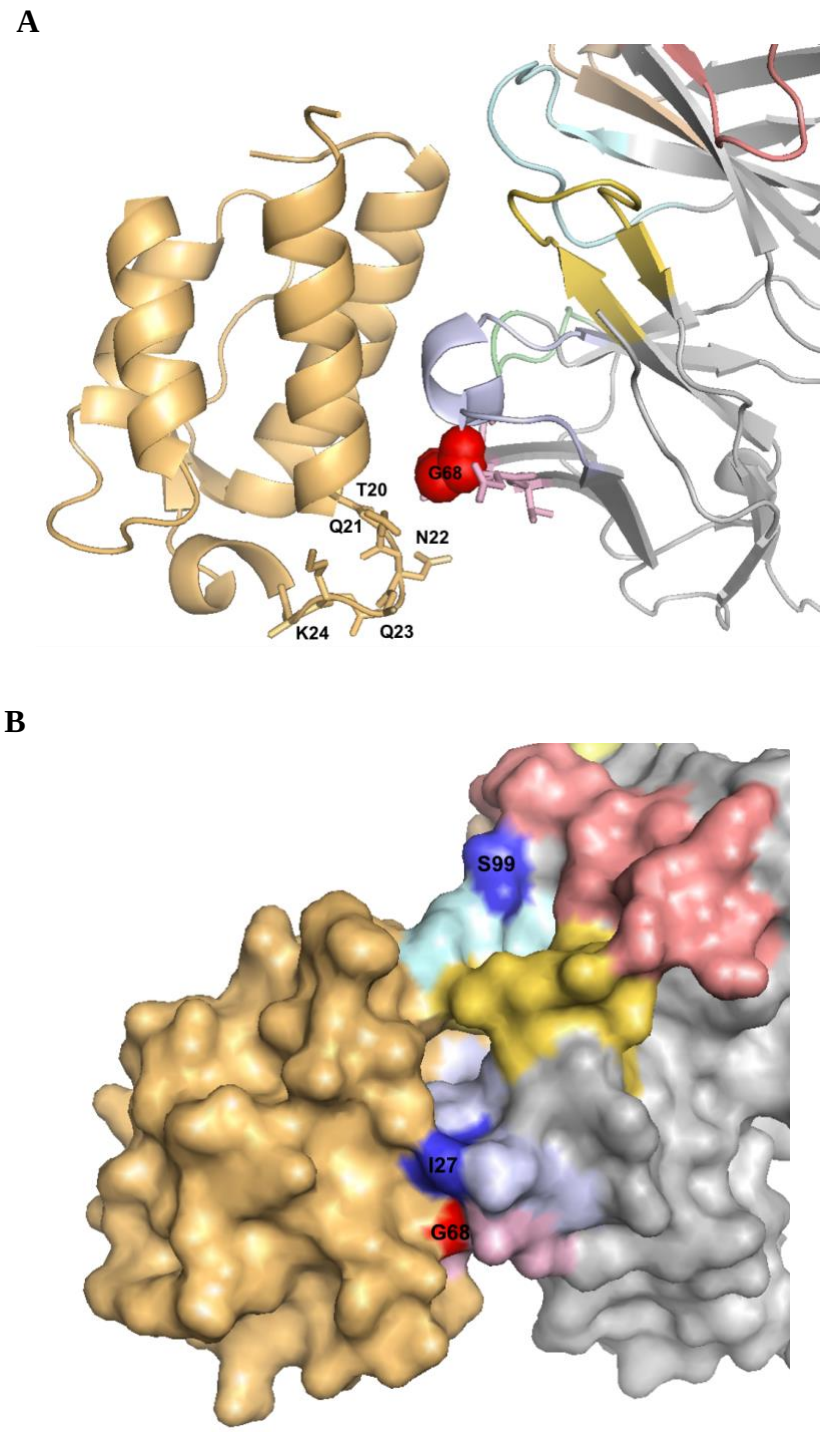


Figure 42. Sites of the beneficial insertions and point substitutions on BAK1.1:IL-13 complex crystal structure. A) The site of the beneficial insertion in the V_L FWR3 loop is in close proximity to the antigen-binding site, after the position G68 (shown as red main chain spheres on the cartoon representation of the structure). **B)** Positions of insertions (in red) and point substitutions (in blue) that not only conferred a large improvement on BAK1 affinity individually, but also synergized to generate a higher affinity antibody are shown on surface representation of the structure. All these modifications are located on surface exposed positions on the periphery of the antigen-binding site. The V_H CDR1, V_H CDR2 and V_H CDR3 loops are coloured in wheat, salmon and pale cyan, respectively, while the V_L CDR1, V_L CDR2, V_L CDR3 in blue white, pale green and yellow orange. IL-13 is shown in light orange.

3.3. Discussion

The work described in this chapter reports on the use of new transposon-based approaches to build antibody libraries with random indels. Quality control of the BAK1 scFv libraries suggested that indels were randomly distributed across the entire scFv and that the single amino acid insertion library was comprised of high diversity insertions. Indeed, this type of search allowed identification of an insertion that improves affinity away from the CDRs, which are the regions usually randomized by targeted approaches, in a non-hypervariable region, the FWR3 loop (Chothia & Lesk, 1987). This is the first time that engineering this loop using insertion libraries resulted in affinity gain, as previous effort to engineer antibody affinity by increasing the length of the FWR3 loop has failed (Young, Watson, Cunningham, MacKenzie, & Huston, 2014).

In Nature, the FWR3 region is not hypervariable and the percentage of insertions in this loop in affinity matured antibodies has not been shown to be higher compared to other FWR regions (de Wildt, van Venrooij, Winter, Hoet, & Tomlinson, 1999; look Briney, Willis, & Crowe, 2012, figure 3A; Bowers et al., 2014). Further evidence that the FWR3 loop is amenable to insertions is provided by the presence of insertions in this loop in several anti-HIV bnAbs which have been subjected to extensive affinity maturation (Klein et al., 2013; Kepler et al., 2015); for example 5 of 17 antibodies with broad neutralization activity (29.4%) have a FWR3 loop insertion based on Figure 1 by Klein et al. (Cell, 2013). Furthermore, one study suggests that the FWR3 loop of an anti-hapten single domain antibody (which lack the light chain) confers high improvement in affinity, as grafting the CDR loops on a different V_HH framework showed that the FWR3 loop (referred to as CDR4) confers a 1000-fold improvement in affinity. The data presented in this chapter combined with the presence of FWR3 loop insertions in extensively matured antibodies (bnAbs) in Nature paints a

consistent picture that FWR3 loop length variation is well tolerated and a potential source of affinity enhancement.

Insertion in the FWR3 (67aE) was beneficial only in the presence of specific amino acid residue at position 68 (G68W). These results indicate epistatic interaction between those modifications, with specific point substitution of a residue at position flanking the insertion enabling the insertion to exert its function (permissive mutation). These results underline the importance of sampling not only high amino acid diversity insertions, but also high diversity point substitutions of residues flanking the position of the insertion in order to identify an insertion variant with improved affinity.

Epistatic interactions between the FWR3 insertion and two beneficial substitutions (BAK1.1) were evaluated by making a BAK1 variant with both FWR3 insertion and BAK1.1 point substitutions. Increased affinity of this variant compared to the variant with only point substitutions, indicates positive epistatic interaction between affinity-enhancing insertion and point substitutions. Insertions have been found to contribute to affinity gains (20 and 40-fold) during *in vitro* affinity maturation of antibodies raised against different antigens by mammalian surface display (Bowers et al., 2011). Although collectively insertions and point substitutions contributed to the overall affinity improvement, it is not clear if these insertions would have had the same effect on affinity if they had been acquired before the occurrence of other substitutions. Overall, this is the first time that was shown that insertions and point substitutions can have additive effects on affinity. Thus, for engineering purposes insertions could be included in CDR walking mutagenesis approaches, as selections of libraries with insertions and substitutions could be performed in parallel and then combined to exploit additive effects on affinity.

G68W substitution was beneficial on affinity only in the presence of the insertion (67aE; permissive mutation), as the mutant with G68W substitution (BAK1-W) but without insertion

at the position 67a did not possess any improvement in affinity. In addition, no point substitutions with large effect on BAK1 affinity were identified in this loop either in previous or in the current study, indicating that point mutations in the FWR3 loop conferred gain in affinity only in the presence of the insertion. These data support the conclusion that insertions and substitutions can have complementary mechanisms of function, leading to the localization of beneficial insertions in different positions than beneficial point substitutions on the same antibody, meaning that insertions offer alternative routes to affinity maturation. For engineering purposes, insertions could be used as a type of sequence modification that can offer additional pathways to affinity maturation, especially in cases where difficulties are encountered with conventional methods using only point substitutions.

Analysis of the site of the FWR3 loop insertion on the BAK1.1:IL-13 complex structure has shown that the position of the insertion constitute part of the structural paratope. The close proximity of the insertion site to the antigen-binding site combined with the large effect on affinity conferred by the insertion, suggests that it is possible that the insertion induces direct interactions of the FWR3 loop with the antigen. Furthermore, the presence of a residue that could form intra-molecular hydrogen bonds and even a salt bridge (67aE; glutamic acid insertion) combined with a residue with aromatic rings that have multiple interaction capabilities with antigen (G68W; tryptophan substitution) (Vindahl, Nielsen, Berg, & Lund, 2013) in the FWR3 loop supports the hypothesis that increased affinity is caused by the insertion due to formation of direct contacts with the antigen. Specifically, the G68W substitution could bury additional hydrophobic surface and possibly make some contacts with the antigen, but probably not pi stacking interactions because of the absence of aromatic residues in close proximity to this residue. Extended hydrophobic surface could also result in lowering the dielectric constant that would benefit electrostatic interactions between polar residues, like aspartic acid (E67a) and serine (S67), with residues of the antigen, like

aspargine (N22) and glutamine (Q21 and Q23) residues that could participate in hydrogen bonds.

The idea that the FWR3 insertion could form new inter-molecular interactions by extending the loop is supported by analysis of other FWR3 insertions in natural antibodies from studies of crystal structures of antibody fragment complexes with their antigens. For example, a FWR3 loop insertion of the anti-HIV bnAb 8ANC195 (Scharf et al., 2014) has been shown to exert its effect by extending the loop and forming intra-molecular contacts, which in turn enable new contacts of FWR3 loop residues around the position of the insertion with residues of the gp120 loops D and V5. The functional effect of the insertion was an extended epitope surface that reached a pocket on gp120 surface, resulting in broadened neutralization activity of the antibody against various viral strains. Another example is the FWR3 loop of the anti-MTX V_HH which was shown to interacts directly with the hapten, resulting in 1000-fold higher affinity than in the absence of the residues that form the inter-molecular contacts (Fanning & Horn, 2011). Overall, the FWR3 insertion could form new inter-molecular contacts either directly or by enabling other loop residues flanking the position of the insertion to interact with the antigen.

Taken together, although the effect of random indels across the entire scFv on antibody affinity was assessed by using libraries with random and high diversity indels, the data presented in this chapter indicate that beneficial insertions are located in loops in close proximity to the antigen-binding site.

4. BAK1 affinity maturation by insertional-scanning mutagenesis

4.1. Introduction

As discussed in introductory sections of previous chapters, Nature appears to sample only low diversity insertions, as this type of modification is introduced by SHM in low frequency and only by sequence duplication. Consequently, by sampling higher diversity of insertions across the variable regions by *in vitro* evolution, one is likely to identify more insertions that confer beneficial effect on antibody affinity. This idea is reinforced by the results presented in the previous chapter, underlining the importance of sampling high amino acid diversity insertions combined with point substitutions in positions flanking the insertion, which may permit the insertion to exert its beneficial effect on affinity.

In the previous chapter sampling random and high diversity insertions across the entire scFv was achieved by using a transposon-based approach to build libraries with single amino acid insertions. Although the construction of libraries with even longer insertions is possible using this method, this type of library for insertions longer than 2 amino acids would be impossible to cover the diversity of all the possible variants (as discussed in chapter II). To overcome this limitation, I developed a semi-random approach that allowed sampling of high diversity insertions longer than a single amino acid. The capability of this method to enable the identification of affinity-enhancing insertions was assessed by searching for beneficial insertions in different loops of the same model antibody, BAK1.

Identification of a beneficial insertion in the V_L FWR3 loop combined with the frequent presence of insertions in that loop of bnAbs to HIV-1, suggest that FWR3 loop insertions can be tolerated and confer improvements in affinity. Random sampling of indels across the entire scFv sequence led to the identification of a beneficial insertion in a loop close to the antigen-binding site, suggesting that beneficial insertions are more likely to be found in loops close to the antigen-binding site. This conclusion is supported by data of affinity matured antibodies in Nature showing beneficial indels mainly occurring in the CDRs 1 and 2 and in

close proximity to the CDRs (de Wildt, van Venrooij, Winter, Hoet, & Tomlinson, 1999; Briney, Willis, & Crowe, 2012; Bowers et al., 2014). It is unclear whether or not the CDR3 loops can accommodate beneficial indels as the analysis of these loops is impossible, because indels in the V_H CDR3 and at the end of the V_L CDR3 cannot be identified due to length variation that is introduced at these positions (junctions) during the initial V(D)J rearrangement. In addition, extensively somatically mutated bnAbs accumulate insertions both in the FWR3 and in the CDR loops (Diskin et al., 2011; F. Klein et al., 2013). Taken together, these data suggest that insertions not only in the FWR3 loop but also in other CDRs could potentially confer significant improvements in affinity.

Assuming that insertions not only in the FWR3 loop but also in other CDRs can confer significant improvements in affinity, I decided to sample different lengths of high diversity insertions in the CDRs of the same antibody (BAK1), in order to find more beneficial insertions on the same antibody. Additionally, according to the findings presented in the previous chapter, it would be desirable to include point substitutions in positions around the insertion, because these could enable the functional effect of the insertion. To find out which CDR positions could be randomized, I followed a semi-random approach, dubbed insertional-scanning mutagenesis (**Figure 43).** This method is based on the notion that positions that can tolerate single amino acid insertions are more likely to tolerate longer insertions as well.

Identification of the positions of the BAK1 gene that can tolerate single amino acid insertions without losing binding to IL-13 was used to inform the construction of libraries with length variation at targeted positions. CDR positions that were found to tolerate insertions were randomized with different lengths of insertions (up to five amino acids), with the full diversity of amino acids, by construction of separate libraries using classic methods of targeted mutagenesis. In contrast to previous methods of insertional mutagenesis (Lamminmäki et al., 1999), knowledge of the structure is not required as the design of

libraries with insertions is not based on natural length variation or structural information, but on screening for binding to the antigen of a library with single amino acid insertions either random or only in the CDRs loops, including the FWR3 loops.

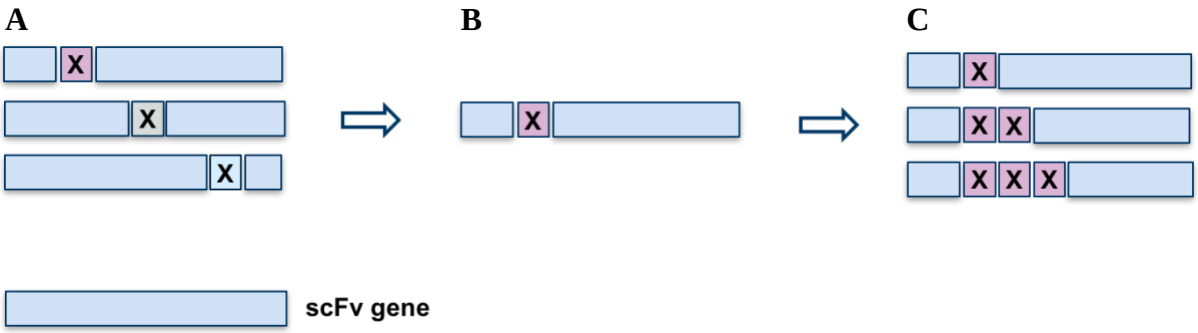


Figure 43. Insertional-scanning mutagenesis. A) The BAK1 library with random single amino acid insertions (see chapter II) (different X colours indicate different amino acid residues), B) was used to identify the BAK1 variants that are still functional and consequently, can tolerate single amino acid insertions (**Figure 45**). Assuming that the variants that tolerate single amino acid insertions might also tolerate longer insertions, these positions were chosen for randomization with full diversity insertions of different lengths. C) The libraries with insertions of different lengths at chosen positions were built using classic methods of targeted mutagenesis.

4.2. Results

4.2.1. Positions that tolerate insertions across the BAK1 scFv

Positions that tolerate insertions were identified by using the library containing random single amino acid insertions across the BAK1 gene (**see chapter III**). By screening a subset (5,632 variants) of this library for binding to human IL-13 using an HTRF assay, BAK1 insertion variants that were still functional were picked and their sequences were determined. The positions that can tolerate insertion across the scFv gene and the frequency of each variant with insertion in specific position (even of different amino acid residue) in the screen are shown in **Figure 44**.

11 positions were confirmed to tolerate single amino acid insertions in the CDR loops, including the FWR3 loop, specifically insertions were found in four different loops, the V_H FWR3, V_H CDR2, V_L CDR3 and V_L FWR3. In the V_L FWR3 loop, the position before the G68 can tolerate single amino acid insertion (5 hits), which is in agreement with the findings presented in the previous chapter, that the BAK1-INS1 variant has a beneficial insertion at this position. A cluster of four positions with high frequency was found in the V_L CDR3 before the D92, T93, G94 and S95 with 4, 16, 18 and 1 hits, respectively. Only one position with high frequency (before A53) was observed in the V_H CDR2. Except from the N-terminus of the V_H, which may tolerate insertions because it is a free terminus with little involvement in binding but with structural plasticity, high frequency of hits was not observed in other regions out of the CDRs. These results indicate that BAK1 can tolerate single amino acid insertions in multiple loops without loss of the binding activity.

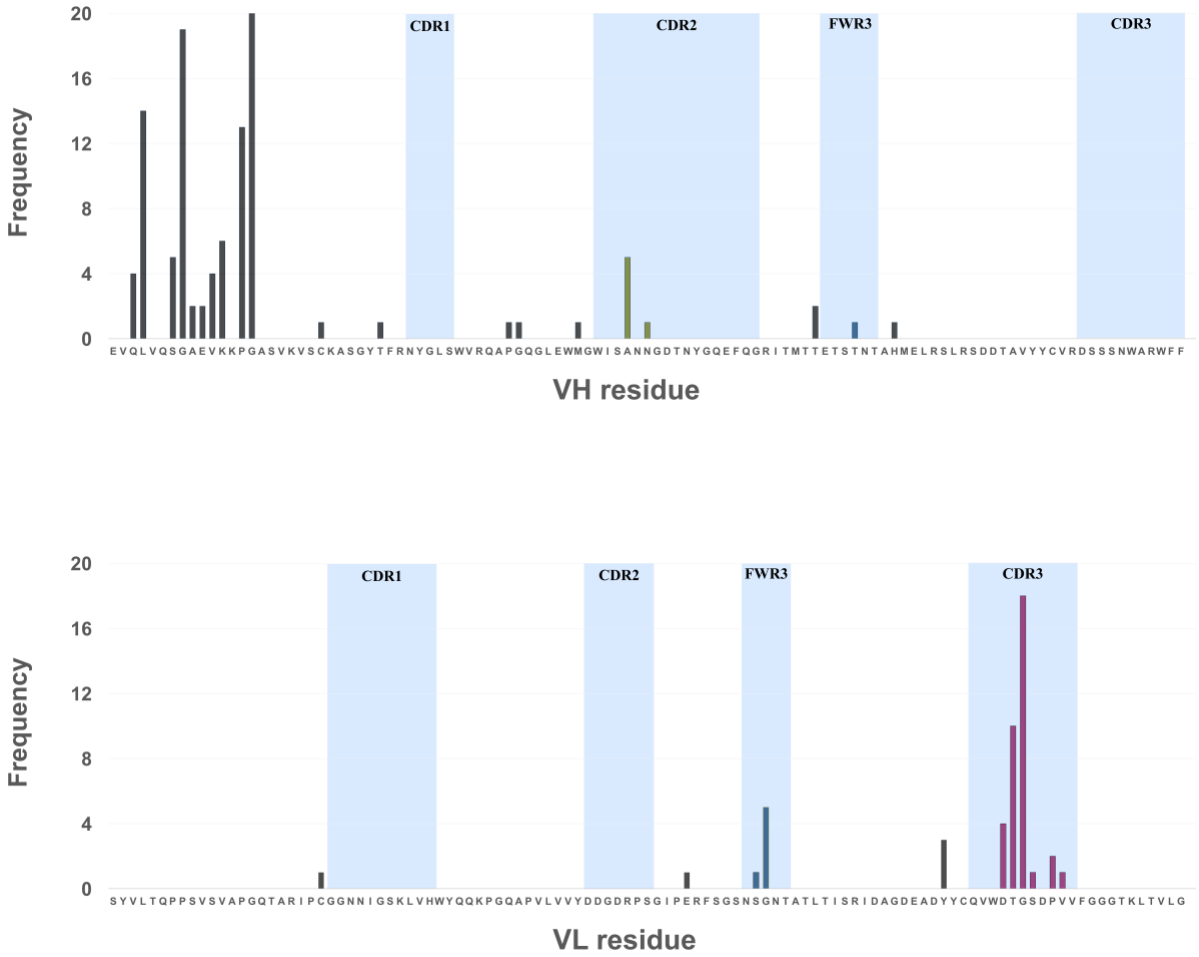


Figure 44. Identification of positions that tolerate single amino acid insertions across the BAK1 scFv were found by screening the BAK1 library for retention of IL-13 binding following single amino acid insertion at different positions. The columns represent a position that tolerate insertion before the indicated residue. The frequency shows the number of hits at a given position (even if the inserted amino acid differs). The loop with the highest frequency, the VL CDR3, was chosen as the target for insertional mutagenesis.

4.2.2. Design of libraries with different lengths of insertions in the V_L CDR3

Assuming that the variants that tolerate single amino acid insertions might also tolerate longer insertions that could also be beneficial on affinity, some of the CDR positions that tolerated insertions were chosen for randomization with full diversity insertions of different lengths. Length variation was considered important to investigate if different insertion lengths would have different effect on affinity and increase the chances of identifying a

beneficial insertion by sampling large diversity of insertions in terms of both length and amino acid composition in each position.

The loop with the highest frequency of insertions in our binding assay, the V_L CDR3, was chosen as the target for insertional mutagenesis. Five libraries with different lengths of insertions, from 0 to 5 amino acids, after the residue W91 were constructed by classic methods of phage library construction by site-directed mutagenesis (**Table 5**). Each library contained 10^9-10^{10} variants with variation in blocks of 5-6 residues (NNS randomization), targeting not only the insertion residues, but also the residues around the insertions with the full amino acid diversity.

Table 5. Design of libraries with different lengths of insertions in the V_L CDR3.

Library		VL CDR3											Theoretical Library size	Actual library size
	Residue	91	92	93	93a	93b	93c	93d	93e	94	95	96		
	BAK1	W	D	T	-	-	-	-	-	G	S	D		
Insertion length	0	W	NNS	NNS	-	-	-	-	-	NNS	NNS	NNS	3.35×10^7	4.6×10^9
	1	W	NNS	NNS	NNS	-	-	-	-	NNS	NNS	D		
	2	W	NNS	NNS	NNS	NNS	-	-	-	NNS	NNS	D	1×10^9	1.4×10^{10}
	3	W	NNS	NNS	NNS	NNS	NNS	-	-	NNS	S	D	1×10^9	2.5×10^{10}
	4	W	NNS	NNS	NNS	NNS	NNS	NNS	-	G	S	D	1×10^9	3.1×10^{10}
	5	W	NNS	NNS	NNS	NNS	NNS	NNS	T	G	S	D	1×10^9	1.4×10^9

4.2.3. Affinity-based selections and screening of the V_L CDR3 insertion libraries

An overview of the design of libraries with different lengths of insertions in the V_L CDR3 by Insertional-scanning mutagenesis and the phage display selections and screening of these libraries to identify variants with improved affinity is shown in **Figure 45**.

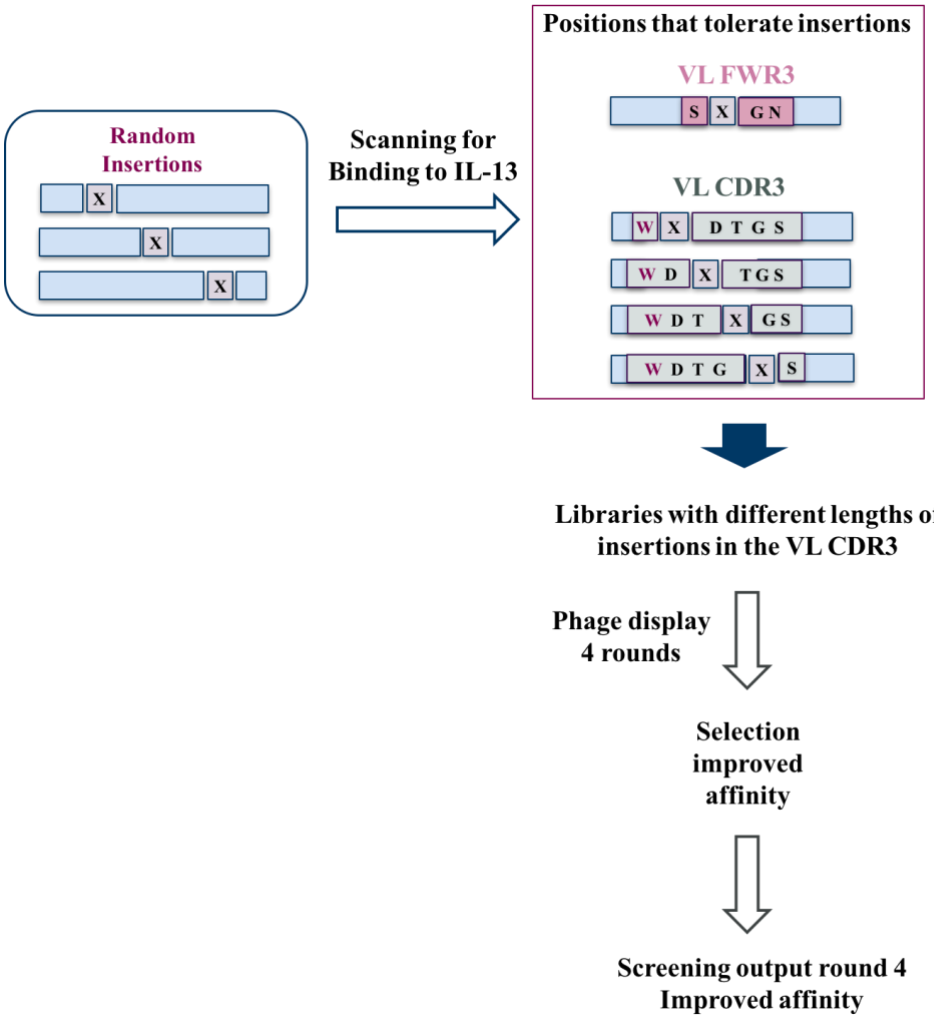


Figure 45. Insertional-scanning mutagenesis (**Figure 43**) of the BAK1 antibody. Scanning of the BAK1 library with random insertions was used to guide the design of libraries with different lengths of insertions in the V_L CDR3. Subsequently, selections of these libraries by phage display were followed by screening for improved variants.

Four rounds of phage display K_D-based selections were performed by panning with a starting recombinant biotinylated IL-13 concentration in the K_D range in the first round, followed by 10-fold lower IL-13 concentration in each round (from 10nM to 0.01 nM). For all the libraries enrichment of specific variants over the rounds was confirmed by sequencing of 44 variants from the outputs of selection rounds 2, 3 and 4 (**Figure 46**). Enrichment of unique variants (40 - 76.5% unique variants in round 4), suggesting that specific variants were enriched under these panning conditions.

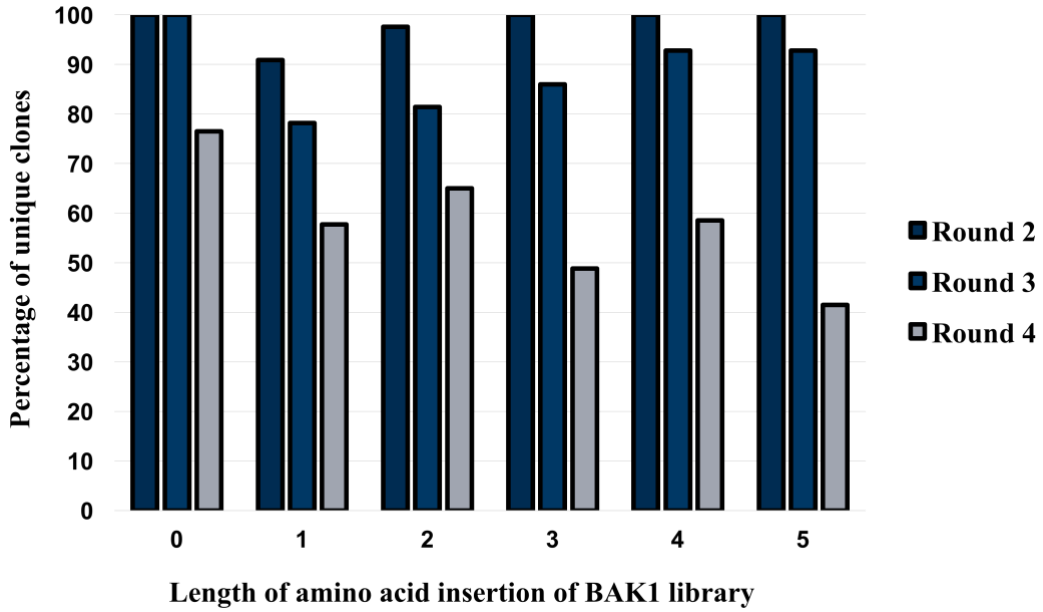


Figure 46. Enrichment of specific variants during phage display selections of V_L CDR3 libraries. Enrichment of unique variants was observed for all libraries, especially the libraries with 1-5 amino acid insertions.

From the round 4 selection output of each library, 384 His_6-tagged scFv periplasmic fractions were screened for binding to human IL-13 using an HTRF competition assay. The sequences of all the screening hits were determined (**Figure 47**) and a few variants from each library were selected for further affinity-based screening. Around half of the variants tested from the libraries with 2-5 amino acids were still binding to IL-13, while from the library with 0/1 amino acid insertions only ¼ of the variants were hits, suggesting that it is more likely to identify variants with improved either expression or affinity from the libraries with 2-5 amino acids. Additionally, the library with 0/1 amino acid insertions presented higher percentage of unique variants (43%), showing higher diversity compared to the other libraries (4.8 - 27.3% unique variants), suggesting that the hits of this library probably contained lower number of improved variants. The library with the lowest diversity of unique hits (5%) was the library with the longest insertions of 5 amino acids, suggesting either a lower number of variants with improved expression for longer insertions or that some variants predominated the selection due to much higher affinity than the other variants.

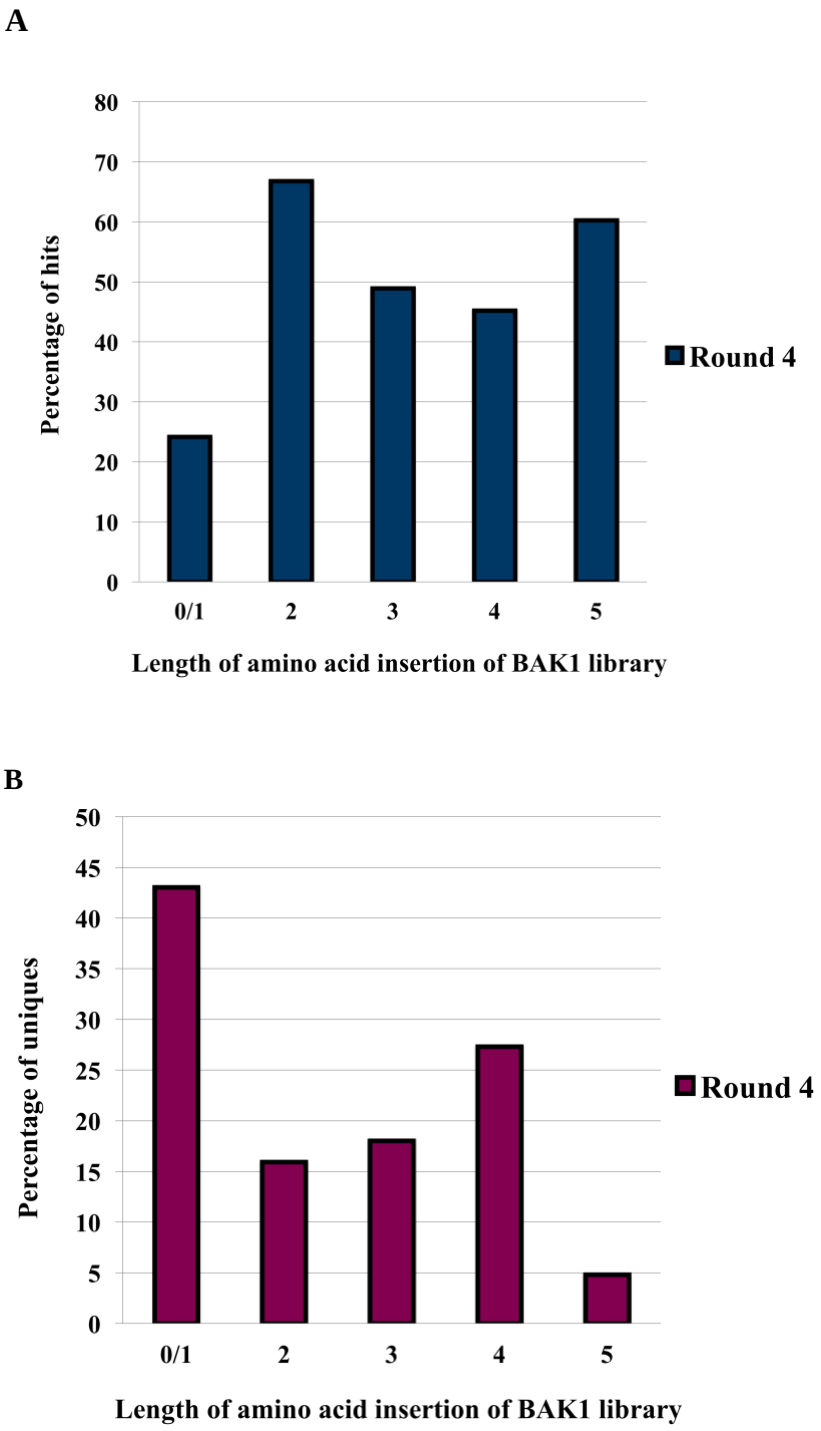


Figure 47. Screening round 4 output of phage display selections of the libraries with insertions in the V_L CDR3 for binding to IL-13. **A)** The percentage of hits in a screen of 384 scFv variants per library. **B)** Sequencing of all the hits was used to analyse the number of unique hits for each library.

The criteria for selecting the variants from each library for further affinity-based screening were the frequency of each unique hit and their signal value in the competition assay. The results of the affinity-based screening are presented in **Table 6.** To identify the variants with improved affinity, an initial HTRF competition affinity screen of purified His_6-tagged scFvs, was followed by picking the variants which seemed to have improvement in affinity rather than in expression. To confirm that these variants were indeed improved, 29 variants were converted to human IgG1 format and tested using a competition ELISA assay. IgG variants with the lowest IC50 values were further characterised using surface plasmon resonance.

Table 6. Screening of BAK1 variants with insertions in the V_L CDR3 for improved affinity.

Loop		VL CDR3											Affinity			
Residue		91	92	93	93a	93b	93c	93d	93e	94	95	96	K_D (pM) Biacore	IC50 HTRF assay	IC50 DELFIA (ug/ml)	HIT Frequency
Parent	BAK1	W	D	T	-	-	-	-	-	G	S	D	2,300	13.3	0.6	
	BAK1.1	W	D	T	-	-	-	-	-	G	S	D	50	0.3	0.06	
0	Ins0.2	W	G	T	-	-	-	-	-	P	D	P	nd	nd	nd	3
	Ins0.3	W	S	W	-	-	-	-	-	S	G	Q	nd	nd	nd	1
	Ins0.4	W	S	E	-	-	-	-	-	S	Q	P	nd	nd	nd	1
	Ins0.5	W	S	V	-	-	-	-	-	E	G	S	nd	nd	nd	1
	Ins0.6	W	S	D	-	-	-	-	-	E	L	G	nd	nd	nd	1
1	Ins1.2	W	S	D	E	-	-	-	-	L	G	D	nd	-	0.3	12
	Ins1.3	W	N	D	E	-	-	-	-	R	G	D	nd	4	0.6	1
	Ins1.4	W	D	P	Y	-	-	-	-	R	G	D	nd	-	0.6	2
	Ins1.5	W	A	G	D	-	-	-	-	K	G	D	nd	-	3.9	15
	Ins1.6	W	S	W	H	-	-	-	-	G	A	D	800	1	0.3	1
	Ins1.7	W	S	E	P	-	-	-	-	L	G	D	nd	3.4	nd	10
	Ins1.8	W	N	W	A	-	-	-	-	D	G	D	nd	8	nd	2
2	Ins2.1	W	S	G	W	D	-	-	-	P	G	D	170	0.8	0.09	103
	Ins2.2	W	S	G	P	G	-	-	-	A	G	D	nd	13.3	nd	1
	Ins2.3	W	P	G	P	M	-	-	-	R	G	D	nd	4.2	nd	4
	Ins2.4	W	N	D	T	D	-	-	-	P	R	D	nd	40	nd	12
	Ins2.5	W	R	T	W	D	-	-	-	S	G	D	nd	0.7	nd	5
	Ins2.6	W	P	G	A	G	-	-	-	P	G	D	nd	36.5	nd	7
	Ins2.7	W	A	D	G	Q	-	-	-	T	G	D	nd	5.9	nd	1
3	Ins3.1	W	T	L	H	D	E	-	-	A	S	D	nd	2.3	nd	28
	Ins3.2	W	T	A	D	F	P	-	-	P	S	D	nd	-	5.2	30
	Ins3.3	W	P	G	P	D	P	-	-	V	S	D	nd	20.9	nd	19
	Ins3.4	W	A	L	W	D	E	-	-	S	S	D	300	0.4	0.03	15
	Ins3.5	W	P	N	W	E	V	-	-	G	S	D	nd	0.9	nd	7
	Ins3.6	W	T	A	R	L	P	-	-	A	S	D	nd	0.35	nd	1
	Ins3.7	W	V	P	W	G	S	-	-	H	S	D	nd	1.2	nd	1
	Ins3.8	W	S	D	R	P	V	-	-	T	S	D	nd	4.3	nd	6
	Ins3.13	W	A	G	P	L	Q	-	-	K	S	D	nd	nd	1.7	1
	Ins3.15	W	P	G	P	D	A	-	-	L	S	D	nd	nd	0.5	1
	Ins3.17	W	N	D	A	N	I	-	-	K	S	D	nd	nd	0.9	1
	Ins3.18	W	N	P	F	D	T	-	-	R	S	D	-	-	4	1
4	Ins4.2	W	V	P	P	H	I	R	-	G	S	D	530	1	0.14	2
	Ins4.3	W	P	S	W	D	D	Q	-	G	S	D	90	1.6	0.06	63
	Ins4.4	W	P	A	W	D	T	V	-	G	S	D	34	1.3	0.07	15
	Ins4.5	W	V	P	G	P	Q	G	-	G	S	D	nd	8.9	nd	14
	Ins4.6	W	V	G	P	E	H	L	-	G	S	D	nd	5.2	nd	9
	Ins4.7	W	S	G	W	E	P	P	-	G	S	D	nd	6.9	1.5	6
	Ins4.9	W	T	T	G	E	P	V	-	G	S	D	nd	nd	1.5	1
	Ins4.10	W	T	T	F	D	Y	P	-	G	S	D	360	nd	0.2	1
	Ins4.11	W	T	E	Q	T	A	K	-	G	S	D	nd	nd	2.1	1
	Ins4.12	W	V	H	E	N	Q	N	-	G	S	D	-	nd	0.1	1
	Ins4.14	W	V	S	K	V	E	E	-	G	S	D	nd	nd	0.7	1
	Ins4.15	W	T	G	G	S	D	S	-	G	S	D	nd	nd	0.9	1
	Ins4.16	W	V	D	P	S	N	P	-	G	S	D	nd	nd	0.6	1
	Ins4.18	W	N	P	S	E	R	R	-	G	S	D	nd	nd	2.7	1
5	Ins5.1	W	V	P	A	T	D	L	T	G	S	D	nd	-	0.2	5
	Ins5.3	W	C	T	F	D	C	Q	T	G	S	D	48	1.9	0.07	88
	Ins5.4	W	V	P	P	S	W	Q	T	G	S	D	nd	0.9	nd	63
	Ins5.5	W	Y	D	R	A	E	I	T	G	S	D	nd	3.8	0.3	4
	Ins5.6	W	V	A	E	D	P	P	T	G	S	D	nd	2.7	0.2	1

4.2.4. Beneficial insertions in the V_L CDR3

K_D measurements of 10 IgG variants (**Table 7**) showed that 8 of those had improvement in affinity, but only variants with two-amino acid insertions or longer had large improvements (> 5-fold), indicating that only insertions of more than one amino acid exert large effects on affinity. All the variants with large improvements, INS2.1, INS3.4, INS4.3, INS4.4 and INS5.3 (13.5, 7.7, 25.5, 67.6 and 48-fold improvement in K_D, respectively), had different lengths of insertions from 2 to 5 amino acids and a motif, 93aW or 93aF combined with 93bD, suggesting that insertion longer than two amino acids combined with this specific motif were required for the insertion to confer a large beneficial effect on affinity. From the library with zero and single amino acid insertions only one variant with single amino acid insertion seemed to hold a small improvement in affinity (INS1.6; 2.9-fold), indicating that variants without insertion in the V_L CDR3 were not improved, which is in agreement with previous data (Thom et al., 2006). These results demonstrate that insertions in the V_L CDR3 have complementary effects with point substitutions on affinity.

Although the variants with long insertions in the V_L CDR3 had even larger improvements in affinity (INS4.4; 67.6-fold) than the variant with a single amino acid insertion in the V_L FWR3 (BAK1-INS1; 14-fold) (**Chapter III**) and compared to variants with shorter insertions in the V_L CDR3, there is not enough data to support the idea that longer insertions result in larger improvements in affinity. Compared to the variant with two point substitutions (BAK1.1, 46-fold), the variants with 4 and 5 amino acid insertions (INS4.3, INS4.4 and INS5.3) had similar improvement in affinity, thus one long insertion had the same effect as two substitutions on BAK1 affinity, demonstrating the large effect of insertions on antibody affinity.

Table 7. Sequences and binding kinetics* of improved BAK1 variants with insertions in the V_L CDR3.

Loop		VH CDR3	VL CDR1	VL CDR3										
Residue		99	27	91	92	93	93a	93b	93c	93d	93e	94	95	96
Parent	BAK1	N	N	W	D	T	-	-	-	-	-	G	S	D
0	BAK1.1	S	I	W	D	T	-	-	-	-	-	G	S	D
1	Ins1.6	N	N	W	S	W	H	-	-	-	-	G	A	D
2	Ins2.1	N	N	W	S	G	W	D	-	-	-	P	G	D
3	Ins3.4	N	N	W	A	L	W	D	E	-	-	S	S	D
4	Ins4.2	N	N	W	V	P	P	H	I	R	-	G	S	D
	Ins4.3	N	N	W	P	S	W	D	D	Q	-	G	S	D
	Ins4.4	N	N	W	P	A	W	D	T	V	-	G	S	D
	Ins4.10	N	N	W	T	T	F	D	Y	P	-	G	S	D
5	Ins5.3	N	N	W	C	T	F	D	C	Q	T	G	S	D

Loop		Affinity					Thermal stability
Residue		K_D (pM)	Fold improvement in K_D	k_{on} (10^6 M^{-1} s^{-1})	k_{off} (10^{-4} s^{-1})	IC50 DELFIA (ug/ml)	Tm (°C)
Parent	BAK1	2,300	1	4.9 ± 0.017	100 ± 0.34	0.6	62
0	BAK1.1	50	46	8.5 ± 0.095	4.3 ± 0.0034	0.06	57
1	Ins1.6	800	2.9	6.5 ± 0.028	5 ± 0.019	0.3	60
2	Ins2.1	170	13.5	8 ± 0.079	10 ± 0.01	0.09	62.5
3	Ins3.4	300	7.7	5.7 ± 0.053	10 ± 0.014	0.03	62
4	Ins4.2	530	4.4	3.1 ± 0.04	10 ± 0.019	0.14	65
	Ins4.3	90	25.5	6.9 ± 0.032	6.3 ± 0.0022	0.06	57.5
	Ins4.4	34	67.6	5.3 ± 0.034	1.8 ± 0.001	0.07	60
	Ins4.10	360	6.4	4.1 ± 0.032	10 ± 0.0096	0.2	62
5	Ins5.3	48	48	2.5 ± 0.018	1.2 ± 0.001	0.07	64.5

* K_D values of the human IgG1 variants were determined by surface plasmon resonance using the Biacore. The fold improvement in K_D is the ratio K_D ^{Variant} / K_D ^{Parent}.

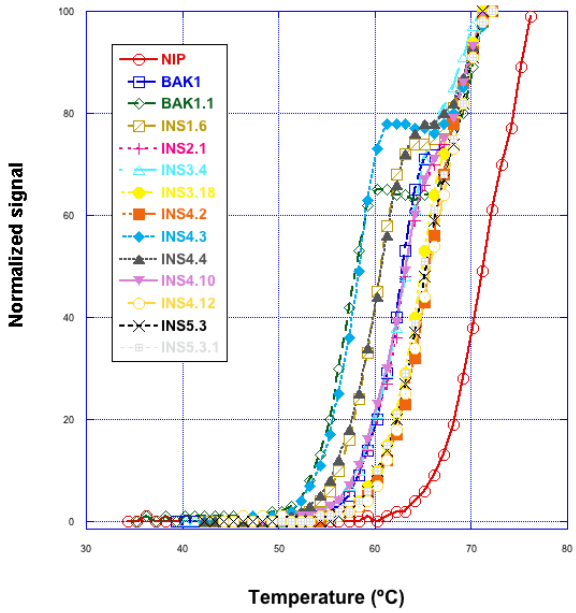
4.2.5. Effect of beneficial insertions on thermal stability

Mutations that improve protein function often have negative effect on stability, indicating a trade-off between function and stability of enzymes (Wang, Minasov, & Shoichet, 2002; Gong, Suchard, & Bloom, 2013) and antibodies (Julian et al., 2015; Julian, Li, Garde, Wilen, & Tessier, 2017). To study the effect of beneficial insertions on stability and compare them with that of beneficial point substitutions, I performed thermal stability measurements of the improved IgG variants by differential scanning fluorimetry (DSF) (Niesen, Berglund, & Vedadi, 2007) (**Figure 48 and Table 7**). IgGs have two unfolding transitions, the CH2

domain unfolding is followed by that of the Fab and the CH3 domain (Tavakoli-Keshe, Phillips, Turner, & Bracewell, 2014; McConnell et al., 2014). DSF does not resolve both peaks, but for the IgG1 variants with the lowest Tm values two peaks were observed. The first peak shows how the Tm of the IgG variants is affected by changes in the variable regions of the Fab.

Most variants with small improvements in affinity (INS2.1, INS3.4, INS4.2) were as stable as the parent IgG (BAK1; Tm=62), thus the insertions did not have a negative impact on antibody stability. Most variants with large improvements in affinity had a loss in stability (INS4.3; Tm=57.5, INS4.4; Tm=60), and the same loss was observed for the variant with two point substitutions (BAK1.1; Tm=57) (**Figure 49**), indicating that the trade-off between affinity and stability was observed for both the insertion and the substation variants with large improvements in affinity. The only exception was the variant with a five-amino acid insertion (INS5.3) which besides conferring large gain in affinity, it also possessed higher stability than the parent (Tm=64.5). This improvement in stability could be attributed to the insertion of two cysteines residues that likely form a disulphide bond that could stabilize the loop. Intra-loop disulphide bond in the V_L CDR3 has not been observed in natural antibodies, suggesting that despite the large beneficial effect on both affinity and stability, this type of modification does not occur in Nature. These data support the conclusion that by engineering this loop with long insertions and high diversity of amino acids we generated an antibody that would not be found in Nature, probably due to sampling of low diversity insertions.

A



B

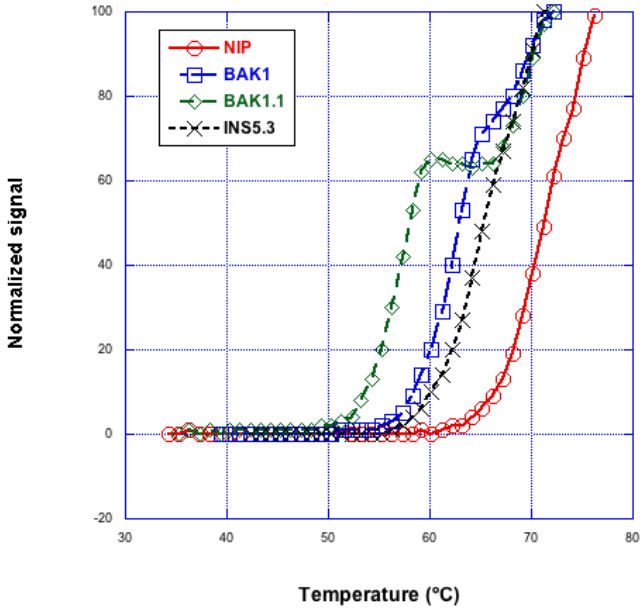


Figure 48. Thermal stability measurements of the IgG variants. The normalized melting curves are shown as measured by DSF. Measurements were repeated twice for all mutants.

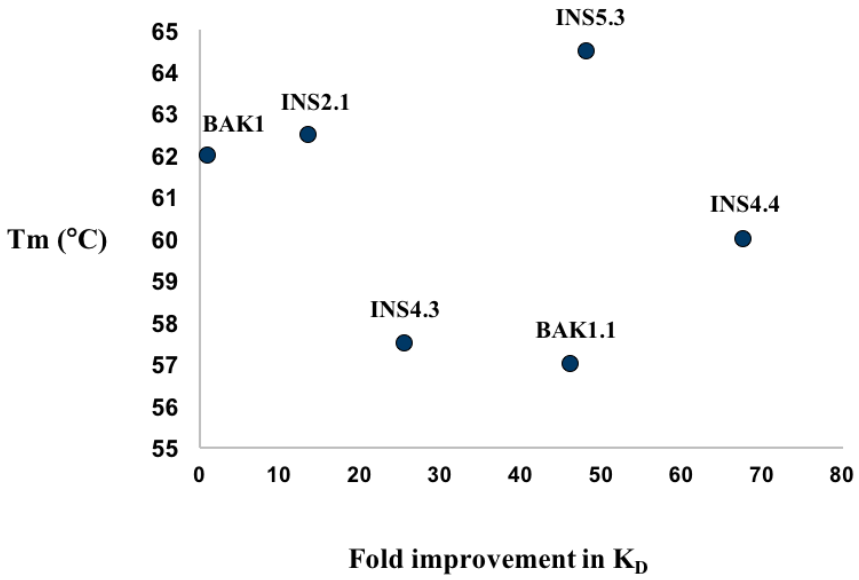


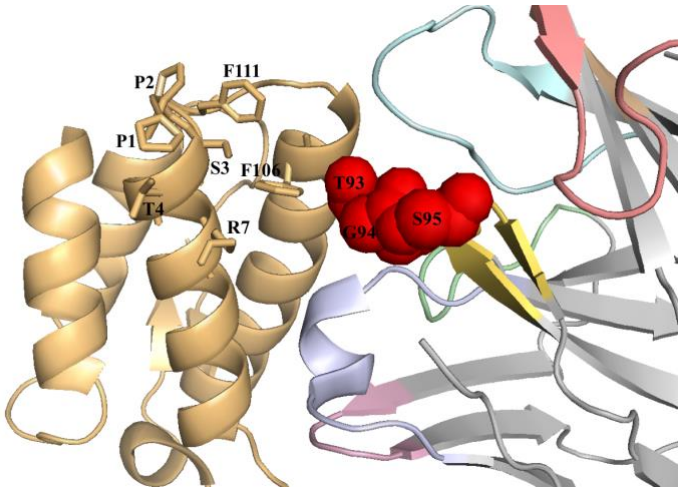
Figure 49. Correlation between melting temperatures and K_D values of IgG variants with large improvements in affinity.

Overall, these results show that insertions that result in gains in affinity can have the same effect as point substitutions on antibody stability, can be neutral, deleterious or beneficial (Julian, Li, Garde, Wilen, & Tessier, 2017b). Insertions induce the same affinity/stability trade-off as has been observed for point substitutions. All these different effects were observed for insertions of different amino acids in the same position in the V_L CDR3, suggesting that specific amino acid substitutions and lengths of insertion in the same position can have very diverse effect on function.

4.2.6. Structural interpretation of the beneficial V_L CDR3 insertions

The crystal structure of the variant BAK1.1 in complex with human IL-13 was determined by Popovic et al. (Popovic et al., 2016). BAK1.1 has two point substitutions in the loops V_H CDR3 (N99S) and V_L CDR1 (N27I), improving BAK1 affinity in combination by 46-fold (**Table 4**), located on the periphery of the binding site (**Figure 50b**). Affinity-enhancing insertions with different lengths, from two to five amino acids were identified in the V_L CDR3 loop (following W91). Some of the residues that were chosen for randomization by insertional mutagenesis, W91 and T93, constitute part of the structural paratope (Popovic et al., 2016) (**Figure 50a**), suggesting close proximity of these residues to the antigen surface. The side chains of these positions have a distance of at least 6.6 Å (T93 oxygen from R7 nitrogen) from atoms of the antigen which means that none of these residues form direct contacts with the antigen. All the positions of the insertions are completely surface exposed (D92 – S95), indicating that the site of beneficial insertions is surface exposed on the periphery of the antigen-binding site.

A



B

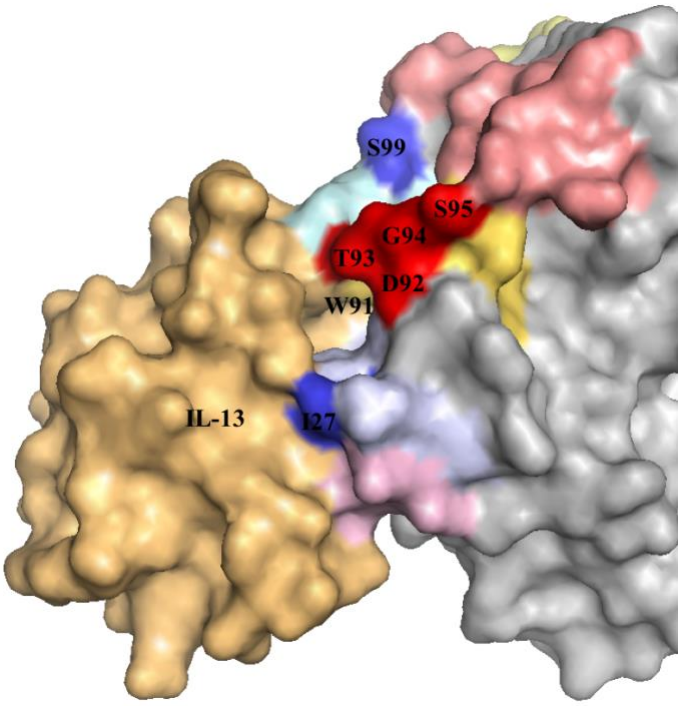


Figure 50. Sites of the beneficial insertions and point substitutions on BAK1.1:IL-13 complex crystal structure. **A)** The site of the beneficial insertions in the V_L CDR3 loop is in close proximity to the antigen-binding site, after the position W91, the positions D92 - S95 (shown as red main chain spheres on the cartoon representation of the structure). **B)** Positions of insertions (in red) and point substitutions (in blue) that conferred large improvements on BAK1 affinity. All these modifications are located on surface exposed positions on the periphery of the antigen-binding site. The V_H CDR1, V_H CDR2 and V_H CDR3 loops are coloured in wheat, salmon and pale cyan, respectively, while the V_L CDR1, V_L CDR2, V_L CDR3 in blue white, pale green and yellow orange. IL-13 is shown in light orange.

4.3. Discussion

The work described in this chapter reports on the use of a new semi-random approach to build antibody libraries with different lengths of insertions, up to six amino acids. Sampling high diversity insertions longer than a single amino acid was considered important because, as discussed in previous sections of this thesis, in Nature only low diversity of insertions is sampled by SHM, therefore, it is likely that sampling high diversity insertions could lead to the discovery of more insertions that confer beneficial effect on antibody affinity. In the previous chapter, sampling insertions no longer than a single amino acid resulted in the identification of an affinity-enhancing insertion in the FWR3 loop of the BAK1 antibody. In this chapter, the method that was developed allowed the discovery of more beneficial insertions in a different loop (V_L CDR3) of the same antibody, suggesting that sampling high diversity insertions of different lengths is important in order to identify beneficial insertions. This idea is also supported by the fact that out of all these high diversity sequences of insertions in the V_L CDR3 that were sampled, only a few variants with a specific motif (positions 93a and 93b; WD or FD) were selected for significantly improved affinity.

The utility of this method was also demonstrated by the fact that the position of the beneficial insertion in the V_L FWR3 (67a) was also identified as one of the positions that can tolerate insertions thus could be randomized by insertional mutagenesis, indicating that this method could have predicted that it is possible to generate variants with improved affinity by randomizing this position. Additionally, the identification of the position that could tolerate insertions in the V_L CDR3 would be difficult to predict without scanning, that allows prediction of the positions that tolerate insertions even without any structural information. However, in another study the same position of an anti-hapten antibody was chosen for randomization by insertional mutagenesis based on the crystal structure to identify variants with modified specificity (Krykbaev et al., 2002).

Furthermore, scanning allows prediction of positions that tolerate insertions even if these do not have length variation in Nature. The first example is the FWR3 loop, which is not hypervariable in Nature, as described in previous chapters. The same conclusion is supported by the identification of insertions longer than those normally found in Nature in the V_L CDR3. BAK1 has a λ-type light chain (IGLV3 family, germline gene; L-Vλ3–21*02). In Nature, the λ-chain V_L CDR3 adopts a wide range of lengths and conformations, composed of 7-13 residues, thus it is possible that this loop can tolerate length variation. Only insertions from two to five amino acids in the 11-amino-acid V_L CDR3 of the parent antibody conferred large improvements in affinity, resulting in 13-15 amino-acid loop, which is longer than normally found in Nature. Thus, although the V_L CDR3 loop can tolerate length variation in Nature, the fact that only longer than single amino-acid insertions would result in improvements in affinity would be difficult to predict.

To compare the outcome of affinity maturation by insertional mutagenesis with that of point mutagenesis in the same loop, V_L CDR3 libraries both with and without insertions in the same positions were constructed. The fact that mutations in the V_L CDR3 loop conferred large gain in affinity only in the presence of two amino acid insertion or longer, suggests complementary mechanisms of function of insertions and point substitutions on antibody affinity. Complementarity is important for engineering purposes, as it means that insertions could be used as a complementary type of sequence modification that could offer alternative routes to affinity maturation.

Taken together, using insertional-scanning mutagenesis to guide the design of libraries with insertions, we were able to sample high diversity insertions of different lengths, which resulted in the discovery of more beneficial insertions on the same antibody, with even higher improvements in affinity, demonstrating that this method can guide the design of libraries with long insertions and could be widely applied in affinity maturation campaigns.

The position of V_L CDR3 loop insertion on the BAK1.1:IL-13 complex structure was shown to be surface exposed on the periphery of the antigen-binding site. The close proximity of the position of the insertion to the antigen combined with the large effect of the insertion on affinity, suggests that it is possible that the insertion makes new contacts with the antigen. Additionally, the presence of a motif (positions 93a and 93b; WD or FD) in all the improved variants with insertions longer than two amino acids suggests that important interactions of these residues can occur only in the presence of an insertion longer than two amino acids, which could be explained by a longer insertion enabling these or other residues to reach and interact with residues on the antigen surface. Specifically, aromatic rings (F106) on the surface of the antigen could achieve pi stacking interactions with 93aW or 93bF and 93bD could form hydrogen bonds or even a salt bridge with residues on the antigen surface (S3, T4, R7).

Furthermore, the idea that the CDR3 insertion could form new contacts with the antigen is supported by studies of the mechanism of function of insertions in affinity matured bnAbs in Nature. Crystal structures of antibody complexes with their antigens have revealed that insertions in both the FWR3 loop (as discussed in section 2.3) and the V_H CDR3 enabled direct interactions with the antigen by extending these loops. The four-residue insertion in the V_H CDR3 of the bnAb NIH45-46, which belongs to the VRC01 clonal lineage (Bjorkman, 2012), introduced not only intra-molecular interactions but also new hydrogen bonds and electrostatic interactions of the insertion residues with the antigen by extending the CDR3 loop and consequently, increasing the buried interface area (Diskin et al., 2011).

Overall, the V_L CDR3 insertion could improve BAK1 affinity by forming new inter-molecular contacts either directly or indirectly by enabling other residues around the position of the insertion to reach the antigen surface and interact with the antigen.

5. Conclusions

This thesis aimed at studying the effect of random and high diversity indels on antibody affinity by in vitro evolution. Insertions that have a beneficial effect on function were found in different locations of the same antibody and structural analysis of these insertions gave insight into the mechanism by which insertions exert their beneficial effect, by extending loops enabling new contacts with the antigen. This thesis demonstrates that insertions besides offering alternative routes to affinity maturation can also be combined with point substitutions to take advantage of additive effects on function. This knowledge improves our understanding on how insertions could be used to engineer antibody affinity.

5.1. Insertions in the CDR and FWR3 loops can confer significant improvements in affinity

Sampling of random indels across the entire variable region of the BAK1 antibody revealed that insertions that confer large affinity improvements can be found not only in the CDR loops, as was expected based on studies of affinity matured antibodies in Nature, but also in a non-hypervariable loop, the FWR3 loop. The fact that the FWR3 loop can accommodate insertions is further supported by the frequent presence of insertions in this loop in extensively affinity matured antibodies (anti-HIV-1 bnAbs), which have been shown to confer broad neutralization activity of the 8ANC195 and 3BNC60 bnAbs (Scharf et al. 2014; Klein et al. 2013). It was expected that affinity-enhancing insertions would be found mainly in the CDRs 1 and 2, as in Nature a low percentage of affinity matured antibodies acquire indels in these loops. The CDR3 loops can tolerate high length variation, thus it is likely that further length diversification of these loops through SHM could contribute to antibody affinity maturation, but the occurrence of indels in these loops cannot be estimated. In this study, insertions in the V_L CDR3 that improve the affinity of the BAK1 antibody were identified, suggesting that length diversification of the V_L CDR3 might also represent a mechanism of affinity maturation in Nature.

5.2. Sampling high diversity insertions enables the discovery of beneficial insertions

In Nature, the low rate of in-frame indels occurring by SHM combined with introduction of only low diversity insertions, as these occur only by sequence duplication, suggests that the diversity generated by SHM is probably not enough to sample high diversity insertions across the entire variable region. Thus, it is likely that the antibody fold can accommodate higher diversity of beneficial insertions. To overcome this limitation, this thesis focused on the development of methods that allowed sampling of high diversity insertions, by using either

an entirely random approach to test short (single) amino acid insertions or a semi-random approach, called insertional-scanning mutagenesis, that uses scanning of random insertions to guide the design of libraries with high diversity longer insertions, up to six amino acids. Indeed, these methods were proven capable of delivering beneficial insertions in more than one loop of the same antibody. In contrast to previous methods of insertional mutagenesis (Lamminmäki et al., 1999), the design of libraries with insertions is not based on natural length variation or structural information, thus knowledge of the structure, which is often not available, is not a prerequisite. Wide application of the methods that were used in this thesis only require scanning for binding to the antigen of a library with single amino acid insertions either at random positions or only in the CDR loops, including the FWR3 loop, as these are the loops that is more like to identify beneficial insertions.

5.3. Enabling point substitutions in positions flanking the insertion

Epistatic interactions between point substitutions is a common factor affecting the functional effect of individual mutations. For example, one mutation may be context dependent, which means that it would be beneficial only in the presence of other, enabling mutations, which could be located far from the mutation. Whether the acquisition of beneficial insertions would depend on the presence of other point substitutions across the antibody variable region was not known before designing the insertion libraries. The discovery of insertions that improved BAK1 affinity in two different loops, independently of the presence of other point substitutions across the scFv gene, points out that the effect of insertions does not necessarily depend on the presence of other mutations. This fact has important engineering implications, as makes possible the identification of beneficial insertions without sampling the combinations of insertions with point substitutions, which would make the search for beneficial insertions impossible because of the magnitude of the libraries that would have to

be constructed or additional epPCR steps. Although the effect of insertions did not depend on point substitutions far from the positions of the insertions, positive epistasis was observed with positions around the insertions, indicating the importance of sampling high diversity of point substitutions in positions flanking the insertion.

5.4. Beneficial insertions and point substitutions can synergize to generate even higher affinity antibodies

Epistatic interactions between indels and point substitutions in Nature are still unclear. Combination of a beneficial insertion (in FWR3 loop) with beneficial point substitutions (BAK1.1 point substitutions) increased the affinity of the variant containing both modifications compared to the variant with only point substitutions, suggesting that insertions and point substitutions can have positive epistatic interactions. These results demonstrate that for engineering purposes, beneficial insertions could be selected separately and then combined with substitutions to exploit additive effects on affinity, thus these could be included in CDR walking mutagenesis approaches. Overall, this is the first time that it was shown that insertions could be included in a parallel CDR walking approach, where affinity maturation using either point substitutions or insertions could be performed separately and then combined to take advantage of additive effects on function.

5.5. Insertions possibly improve affinity by enabling new inter-molecular interactions

Both beneficial insertions in different loops (V_L FWR3 and V_L CDR3) of the BAK1 antibody were located on surface exposed positions on the periphery of the antigen-binding site, suggesting that beneficial insertions are more likely to be found in loops close to the antigen-binding site. The same has been shown for beneficial point substitutions; while the diversity in the primary repertoire is concentrated in the centre of the combining site, point mutations

introduced by SHM are located on the periphery of the binding site (Tomlinson et al., 1996), indicating that these modifications improve affinity while keeping stable the topology that defines specificity. These observations suggest that insertions improve binding by acting in a similar mechanism as point substitutions, without affecting the residues that form direct contacts with the antigen, but residues adjacent to those residues.

Insertions could be involved in making new contacts, like new hydrogen bonds, salt bridges, hydrophobic interactions, intra-molecular interactions with other loops, additional hydrophobic surface and potential pi stacking interactions. Aromatic residues combined with short-chain hydrophilic residues are often found on antibody functional paratopes (Peng, Lee, Jian, & Yang, 2014). This is because aromatic residues (Tyr, Trp, Phe) have multiple interaction capabilities with antigen (Vindahl et al., 2013). The close proximity of the positions of the insertions to the antigen combined with the large effect of the insertions on affinity and the enrichment of specific residues that are frequently found on antibody paratopes (WD or FD motifs) support the idea that the large effect of insertions on BAK1 affinity in both loops (V_L CDR3 and V_L FWR3) could be the result of new contacts with the antigen that was enabled by extended loops.

Furthermore, the idea that insertions may exert their effects by making new contacts with the antigen is supported by studies of the mechanism of function of insertions in affinity matured anti-HIV bnAbs in Nature, which have shown that insertions extend loops and make new contacts with the antigen, either directly or indirectly by enabling other residues around the position of the insertion to reach the antigen surface and interact with the antigen. Another mechanism by which insertions can change antibody affinity is by changing the conformations of adjacent loops, like the 2D1 anti-HA antibody insertion (Krause et al., 2011) at the junction of CDR2 and FWR3, which altered the conformation of the CDR1 loop resulting 35-fold higher affinity. Overall, the mechanisms by which insertions could exert

their function are either by extending loops and making new contacts with the antigen or by altering the conformation of adjacent loops.

Increased affinity by SHM has been shown to involve improvement of the shape and molecular complementarity (additional inter-molecular hydrogen bonds, van der Waals contacts) of the antigen combining site to the epitope (Cauerhff et al., 2004). Additionally, improved binding has been shown to be achieved by increasing the area buried at the interface, which is the result of subtle structural rearrangements on the periphery of the interface with the antigen (Sundberg et al., 2000; Li, Li, Yang, Smith-Gill, & Mariuzza, 2003). The fact that BAK1 beneficial insertions are likely to have gained improved affinity by making new contacts with the antigen on the periphery of the interface, suggests that the improvements in affinity could be attributed to increased buried surface area.

5.6. Complementary effect of insertions to point substitutions on antibody affinity

We found insertions that improve BAK1 affinity in the V_L FWR3 and the V_L CDR3, while beneficial substitutions were found in different loops, the V_H CDR3 and V_L CDR1. The observation that the location of insertions and substitutions that improve BAK1 affinity are in different loops combined with the identification of beneficial insertions in positions where point substitutions did not confer any improvement in affinity, suggests complementary mechanisms of function of insertions and point substitutions on affinity. Thus, insertions have a different mechanism of function than point substitutions, which is in agreement with the observation that insertions exert their effect by extending loops which would not be able to reach the antigen in the absence of the insertion, allowing new inter-molecular interactions. This observation is in contrast to the suggestion that most indels share the same hotspots with point mutations because indel duplications or deletions involve AGY motifs (de Wildt et al., 1999), suggesting same localization of beneficial insertion and point substitutions.

The engineering implications of complementarity are significant. The fact that insertions lead to alternative pathways of evolution of antibody affinity suggests that additional beneficial modifications can be identified by using insertions for affinity maturation. Additionally, the location of insertions and substitutions in different loops point out absence of interaction between those modification, which is in agreement with our finding that their combination can have additive effect on affinity. Overall, for affinity maturation purposes libraries with insertions could be constructed in parallel with libraries with point substitutions and then combine the modifications that improved affinity to take advantage of possible additive effects on function or offer alternatives in cases that it is difficult to identify beneficial point substitutions.

5.7. Why is the percentage of affinity-enhancing insertions in natural antibodies low?

The work reported here provides a greater understanding of the principles governing affinity maturation in Nature. Beneficial insertions were found only on surface exposed positions on the periphery of the binding site area, which is in agreement with the conclusion that affinity-improving insertions introduce only subtle structural rearrangements on the periphery of the interface without affecting the already defined topology of the combining site. Thus, insertions and point substitutions impact affinity by similar mechanisms, by introducing only subtle structural rearrangements on the periphery of the interface area.

Based on studies of the percentage of indels in affinity matured antibodies in Nature, affinity-improving indels were expected to be rare. Thus, the identification of beneficial insertions in two different positions of the same antibody was an unexpected result, indicating that the antibody fold may be able to accommodate higher percentage of indels compared to that observed in natural antibodies. Low occurrence of beneficial insertions in natural antibodies could be attributed to the low rate of incorporation of in-frame insertions during the

diversification process, sampling of only biased diversity of insertions which occur by sequence duplication and the restriction of including the CDR3 regions in the search.

5.8. Outlook

This is the first time that random and high diversity indels across the entire antibody variable region of an antibody were sampled for affinity maturation by *in vitro* evolution. Other researchers have not been able to sample high diversity indels at random positions, either because of restrictions in the diversity of indels that can be introduced by their method (Bowers et al., 2014) or because they only used rational instead of random approaches (Lamminmäki et al., 1999). The methods that I used allowed sampling of random deletions and insertions of different lengths. Although sampling high diversity insertions longer than a single amino acid was not possible by using the entirely random approach (chapter III), I developed an approach that allows the construction of libraries with different lengths of insertions, which could be widely applied in future *in vitro* affinity maturation projects.

As discussed in the introduction, there are two main approaches that have been widely applied for affinity maturation. Parallel or sequential CDR walking mutagenesis (Barbas et al., 1994; W. P. Yang et al., 1995) and scanning by epPCR to guide the design of libraries targeting the CDRs (Thom et al., 2006). The parallel approach involves randomization of different CDRs in different libraries, parallel affinity-based selections and subsequent combination of the most beneficial mutations from each CDR in one library. As described in this thesis, beneficial insertions and point substitutions can have synergistic effects on antibody affinity, suggesting that insertions could be included in parallel CDR walking, as affinity maturation using either point substitutions or insertions could be performed separately and then combined to take advantage of additive effects on function.

Furthermore, an important information gained by this thesis is that insertions and point substitutions can have complementary mechanisms of function, leading to localization of beneficial insertions in different positions than beneficial point substitutions on the same antibody, meaning that insertions offer alternative routes to affinity maturation. For engineering purposes, the fact that insertions lead to alternative pathways of evolution of antibody affinity suggests that insertions can offer an alternative in cases where the identification of beneficial point substitutions has proven difficult. Taken together, including insertions in parallel CDR walking approaches could not only lead to identification of even higher affinity antibodies by taking advantage of additive effects of insertions and point substitutions on function, but can also increase the chances of identifying a beneficial modification.

By insertional-scanning mutagenesis, it is possible to find the positions that could be randomized in length in targeted libraries. This approach requires construction of libraries with single amino acid insertions either by the transposon-based approach (random insertions) or by targeting only the CDRs and the FWR3; for example by construction of a library with single amino acid insertions using the method for construction of single-site saturated mutagenesis library (Koenig et al., 2017). To take advantage of both point substitutions and insertions starting by scanning libraries with either single amino acid insertions or single-site saturated point substitutions in parallel to identify the loops that can possibly improve affinity could be used as a guide for the construction of insertions and point substitution libraries by targeted approaches. The search could even be restricted in the CDR3 loops.

Overall, the approaches for construction of libraries with insertions and deletions that were described in this thesis could be applied in future antibody affinity maturation campaigns to complement and synergize with the classic point substitutions approaches.

6. Materials and Methods

6.1. Continuous *In Vitro* Evolution of Protein Binding

6.1.1. Media

NZY Broth (per Liter): 5g of NaCl, 5 g yeast extract, 10 g NZ amine (casein hydrolysate), Add deionized H_2O to a final volume of 1 Liter. Adjust the pH to 7.5 with NaOH, Autoclave

NZY Top Agar (per Liter): Prepare 1 liter of NZY broth, Add 0.7% (w/v) agarose, Autoclave

2xTY: 10 g/L bacto-yeast extract, 16 g/L bacto-tryptone, 5 g/L NaCl, pH adjusted to 7.0.

Antibiotic concentrations: 100 μg/ml of Ampicillin and 50 μg/ml of Kanamycin, 12.5 μg/ml of Tetracycline

6.1.2. Primers

Table 8. Primers

Primers	
PCR SfiI-ScFv-SfiI amplification	
Fw-SfiI-ScFv	5´-CACTCGGCCCAGCCGGCCATGGCGGACATTGTGATGACACAGTCTC -3´
Rev-SfiI-ScFv	5'- CCCTCGGCCCCCGAGGCC GAGGAGACCGTGACCAGAGTCCCTT -3´
PCR SfiI-VCSM13-SfiI amplification	
Fw-SfiI-VCSM13	5´-TAGGCCTCGGGGGCCGAGGGCGGCGGCTCTGGTGGTGGTTC -3´
Rev-SfiI-VCSM13	5´-GCCATGGCCGGCTGGGCCGAGTGAGAATAGAAAGGAACAACTAAAG -3´
Sequencing	
pUC-g3p	
Fw-pUC-seq-a-lacO	5'– GACCATGATTACGCCAAGCTT –3'
Rev CT-seq	5'– TTAAGACTCCTTATTACGCAGTA –3'
pUC-N1-N2-MCS	
Rev-AP-gIII	5'– GTCTCATGAGCGGATACATA –3'
Rev-N2-seq	5'–AGCATTGACAGGAGGTTGAGG–3'
pUC-N1-N2-HA	
Fw-N1-seq	5'–GAAACTGTTGAAAGTTGTTTAG–3'
VCSM13, VCSM13-ScFv sequencing	
Fw-VCSM13-seq	5'– ACCGATACAATTAAAGGCTCCTT –3'
Fw-VCSM13- seq2	5'– CGCTGCTGAGGGTGACGATCC –3'
Rev-VCSM13-seq	5'– CATTCAACCGATTGAGGGAG –3'

6.1.3. SIP vector construction

For pUC-N1-N2-MCS, pUC-g3p and pUC-N1-N2-HA vector construction, synthetic gene fragments of the pelB-N1-N2-MCS, pelB-MCS-gene III and HA peptide coding sequences flanked by appropriate restriction sites for cloning into the pUC19 vector were ordered from GenScript. Electrocompetent DH10B cells were used for cloning. Colony PCRs were performed using BIOTAQ DNA Polymerase (BIOLINE). The inserted sequences of positive

clones were sequenced with appropriate primers. Strains S1030/pUC-g3p and S1030/pUC-N1-N2-HA were made by transformation of electrocompetent S1030 cells.

6.1.4. Recombinant phage construction

VCSM13 phage particles were obtained from Stratagene and were used to infect S1030 cells, which were then used for VCSM13 RF DNA (replicative form, ds DNA) purification using the protocol for low copy plasmid DNA purification of the Thermo Scientific GeneJET Plasmid Miniprep Kit.

For VCSM13-ScFv-CT phage construction, the phage genome sequence lacking the N1-N2 domains sequence (8 kb fragment) was amplified by PCR with the primers Fw-SfiI-VCSM13 and Rev-SfiI-VCSM13. These primers allowed the insertion of the restriction enzyme SfiI sites for cloning of the scFv fragment using the same sites. The scFv synthetic gene fragment was ordered from GenScript and the ScFv coding sequence was amplified by PCR with primers that introduced the SfiI sites. A similar approach was followed for cloning of the VCSM13-CT phage genome, which lacks the N1-N2 domains sequence. PfuUltra II Fusion HS DNA Polymerase was used for amplification of the SfiI-VCSM13-SfiI (8053 bp) and SfiI-ScFv-Sfi (791 bp) fragments by PCR. The conditions of the reactions are described above:

PCR reaction conditions for SfiI-VCSM13-SfiI phage genome amplification	
VCSM13 DNA (≈100ng)	-
10× PCR buffer	5
10uM Primer 1 (final 0,4 uM)	1
10uM Primer 2 (final 0,4 uM)	1
10mM dNTP's (250uM each)	1,25
PfuUltra II Fusion HS DNA Polymerase	1
ddH2O	-
TOTAL	**50**

PCR reaction conditions for SfiI-ScFv-SfiI fragment amplification	
Plasmid DNA (≈100ng)	-
10× PCR buffer	5
10uM Primer 1 (final 0,4 uM)	1
10uM Primer 2 (final 0,4 uM)	1
10mM dNTP's (250uM each)	1,25
PfuUltra II Fusion HS DNA Polymerase	1
ddH2O	-
TOTAL	**50**

PCR reaction program

1. Denaturation 98°C, 2 min
2. Denaturation 98°C, 20 sec
3. Annealing 68 °C, 20sec
4. Extension 68 °C, 8min for SfiI-VCSM13-SfiI or 30 sec for SfiI-ScFv-SfiI
5. Repeat steps 2-4, 30 cycle
6. 68 °C, 20min
7. Pause, 4 °C

After PCR product cleanup from the 0.8% agarose gel, the PCR product was digested with DpnI to degrade template DNA followed by SfiI overnight (≈ 14 hours) digestion reaction using the Fast Digest enzymes (Thermo Scientific). Then, an overnight 20 µL volume ligation reaction was performed using 100 ng vector and 1: 3 molar ratio vector: insert. 25 ng of the ligation reaction was used for transformation of NEB Turbo Electrocompetent *E. coli* cells with transformation efficiency $1x10^{10}$ cfu/µg for pUC19 DNA. The transformation efficiency of the ligated VCSM13-ScFv-CT DNA was low ($5x10^4$ cfu/ug, i.e. 2×10^5-fold lower than pUC19). The scFv-CT sequence of VCSM13-scFv DNA isolated from positive NEB turbo/VCSM13-scFv clones was obtained by sequencing with primers Fw-VCSM13-seq and Fw-VCSM13-seq2.

6.1.5. pUC-N1-N2-HA fusion protein expression

A single colony of S1030/pUC-N1-N2-HA strain was picked and grown overnight at 37 °C in 5 ml 2TY containing 2 % glucose with AMP. The overnight culture was diluted in 300 mL TG1 in 2TY with AMP (1: 900) with 0.09% glucose (1.35 mL 20 % glucose, final glucose concentration is ~ 0.12 % glucose) and grown at 37 °C until an OD_{600} between 1.0 and 1.5 was reached. 100 mL of 2TY containing AMP (no glucose) and 400 µL IPTG (1 mM final) was then added to the culture (glucose concentration drops below the 0.1% glucose) and grown overnight at 26 °C with very slow shaking.

The periplasmic fraction was prepared by spinning the culture (20 min at 4500 rpm). The supernatant was discarded and pellet was frozen for 90 min at - 80 °C. The pellet was then defrosted (30 min RT) and re-suspended in PBS (25 mL per 400 mL culture) by incubating 1-2 hours on head-over-head at 4 °C. The periplasmic fraction (supernatant) was obtained by centrifugation for 20 min at 4700 rpm (max speed).

6.1.6. Western Blots

The samples were boiled in denaturing SDS buffer, subjected to SDS-PAGE (NuPAGE 10% gel) and electroblotted onto Invitrolon™ PVDF membranes (0.45 µm pore size). The PVDF membrane and SDS-PAGE gel was dipped in 100 % methanol before transfer (50 mL 10× Transfer Buffer containing 250 mM Tris-HCl, 1917 mM glycine and 100 mL methanol in 500 mL total volume. The transfer was performed for 1 hour at 10 V. The membrane was blocked for 45 min in 5% milk/PBS/0.1% Tween 20 and washed three times with PBS.

The membrane was incubated with mouse anti-HA antibody, clone HA-7 SIGMA (1: 1160 dilution in PBS-0.1 % Tween 20) for N1-N2-HA fusion protein detection for 1 hour and washed three times with PBS/0.1 % Tween 20. For phage detection, membranes were incubated with a 1: 3000 dilution of pIII antibody (Mo Bi Tec) in 3% milk/PBS/0.1 % Tween

20. The blot was then incubated with secondary antibody (anti-mouse Antibody-HRP Conjugated, ABCAM) (1: 2000 dilution in PBS-3 % milk, 0.01 % Tween 20) and washed three times with PBS/0.1 % Tween 20. HRP conjugate was detected using the Amersham ECL Prime Western Blotting Detection Reagent.

6.1.7. Phage amplification

An overnight culture of S1030 was diluted in 2TY containing tetracycline and grown until an OD_{600} of 0.5 was reached (mid-log phase). 2 ml of this culture was infected with VCSM13 particles (Stratagene) at a multiplicity of infection of 20:1 (phage-to-cell-ratio), by incubating at 37 °C for 30 minutes (slow shaking). The culture was centrifuged at 2.200 rpm and re-suspended in 20 mL 2TY with kanamycin (final concentration 50 μg/mL). Phages were amplified by culturing the infected cells at 37 °C for 8 hours. The phage supernatant was obtained by centrifuging the culture at maximal speed. Residual cells were killed by heating to 65 °C for 15 minutes. The phage concentration in the supernatant was measured by plaque assay, infectivity assay and ELISA and stored at - 80°C (1 mL phage supernatant containing 0.5 ml aqueous glycerol, 60 % w/w).

6.1.8. Plaque assay

An overnight culture of S1030 was diluted in 2TY containing tetracycline and grown until OD_{600} of 0.5 (mid-log phase). While cells were growing, before reaching the mid-log phase, Top Agar was melted in microwave and 3 mL were dispensed into sterile culture tubes (one per expected phage dilution) and maintained at 47 °C. LB plates (15 mL medium per plate) were pre-warmed at 37 °C. 10^{7}–10^{11}-fold dilutions of phage supernatant were prepared. Aerosol-resistant pipette tips were used to prevent cross-contamination. When the S1030 culture reached the mid-log phase (OD = 0.5), 200 μL into microfuge tubes were dispensed,

one for each phage dilution. To carry out infection, 10 µL of each phage dilution was added to each tube, vortexed quickly, and incubated at room temperature for 1– 5 minutes. Infected cells were transferred to culture tubes containing 47 °C Top Agar. After brief vortexing the culture was immediately poured onto a pre-warmed LB plate and the plate was gently tilted and rotated to spread top agar evenly. The plates were allowed to cool for 5 minutes, and incubated overnight at 37 °C. Plaques were counted on plates that have approximately 100 plaques. Each number was multiplied by the dilution factor for that plate to get phage titer in plaque forming units (pfu) per 10 µL.

6.1.9. Infectivity assay

An overnight culture of S1030 was diluted in 2TY containing tetracycline and grown at 37 °C with shaking (300 rpm) until OD_{600} of 0.5 (mid-log phase). Diluted phage supernatants were mixed with appropriate volumes of S1030 (OD=0.5) and were slowly shaken (150 rpm) at 37 °C for 1 hour. The phage supernatant was serially diluted as described in the **Table 9**.

Table 9. Phage dilutions

Dilution of phage supernatant	Infect 990ul of S1030 (of OD=0,5) with:	Plate on agar plate (Dilution factor)
No dilution	10 µL phage supernatant (10^{-2})	100ul (10^{-3})
10^{-1} dilution	10ul phages supernatant (10^{-3})	100ul (10^{-4})
10^{-2} dilution	10ul phages supernatant (10^{-4})	100ul (10^{-5})
10^{-3} dilution	10ul phages supernatant (10^{-5})	100ul (10^{-6})
10^{-4} dilution	10ul phages supernatant (10^{-6})	100ul (10^{-7})
10^{-5} dilution	10ul phages supernatant (10^{-7})	100ul (10^{-8})

After infection the cultures were plate on LB agar plates containing 50 μg/mL kanamycin and colonies were grown overnight at 37 °C. The phage titer was calculated based on the dilution factor. Colonies were counted on each plate and the titer was calculate using the following formula: [No. of colonies] multiplied by [Dilution factor in plate] = cfu/mL

6.1.10. Phage titration ELISA

Microtiter plate was coated with serial dilutions of phage supernatant in PBS (50 μL per well) and incubate overnight at 4 °C. The next day the plate was washed with PBS 0,05 % Tween 20 and wells were blocked with 300μl PBS with 3% milk for 1 hour at RT (no shacking). Anti-M13 [B62-FE2] HRP conjugated antibody (ab50370, Abcam) was added to the wells (50 μL) after three times washing with PBS-0.05 %Tween 20. The plate was Incubated 1 hour at RT (shaking) and washed 5 times with PBS-0.05 % Tween 20. The signal was detected by adding 50 μL TMB per well and measure OD at 620 nm.

6.1.11. Phage production assay

An overnight culture of S1030/VCSM13-CT-ScFv/pUC-N1-N1-HA or S1030/VCSM13 strain was diluted (1/50) in 2TY containing ampicillin and kanamycin with 2% glucose or 2TY containing kanamycin with 2% glucose, respectively, and grown at 37 °C with shaking (300 rpm) until OD_{600} of 0.5 (mid-log phase). The culture was then centrifuged at 2.200 rpm and re-suspended in 2TY containing ampicillin and kanamycin without glucose, so that the OD_{600} was 0.5. The phage production was done by incubating at 37 °C with shaking (300 rpm) the culture for 2.5 or 4 hours. The phage supernatant was obtained by centrifuged the culture at max speed. Residual cells were killed by heating to 65 °C for 15 minutes.

6.1.12. *E. coli* strain genotypes

NEB turbo: F' $proA^{+}B^{+}$ $lacI^{q}$ ΔlacZM15 / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)Tet^{S} endA1 thi-1 Δ(hsdS-mcrB)5.

DH10B strain: F^{-} endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ^{-}

S1030 strain (Carlson, Badran, Guggiana-Nilo, & Liu, 2014): F'proA+B+ Δ(lacIZY) zzf<Tn10(TetR) lacIQ1PN25-tetR luxCDE/endA1 recA1 galE15 galK16 nupG rpsL(StrR) ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) proBA<pir116 araE201 ΔrpoZ Δflu ΔcsgABCDEFG ΔpgaC λ−

6.2. Molecular Biology

6.2.1. Primers

Table 10: List of primers used in this thesis.

Name	Use	Sequence
		Ribosome display selections
T6te	Ribosome display	CCGCACACCAGTAAGGTGTGCGGTATCACCAGTAG CACCATTACCATTAGCAAG
T7B	Ribosome display	ATACGAAATTAATACGACTCACTATAGGGAGACC ACAACGG
SDCAT-BAK1	Ribosome display	AGACCACAACGGTTTCCCTCTAGAAATAATTTTGT TTAACTTTAAGAAGGAGATATATCCATGGCCGAAG TGCAGC
		Site-directed mutagenesis
Fw-VH99S:	VH CDR3 (N99S)	GAGATTCCAGCAGCAGCTGGGCCCGCTGG
Rev-VH99S	VH CDR3 (N99S)	CCAGCGGGCCCAGCTGCTGCTGGAATCTC
Fw-VL27I	VL CDR1 (N27I)	GGATTCCCTGTGGGGGAAACATCATTGGAAGTAA ACTTGTAC
Rev-VL27I	VL CDR1 (N27I)	GTTTACTTCCAATGATGTTTCCCCCACAGG
Fw-VL68W	VL FWR3 (G68W)	CTGGCTCCAACTCTTGGAACACGGCCACCC
Rev-VL68W	VL FWR3 (G68W)	GGGTGGCCGTGTTCCAAGAGTTGGAGCCAG
Fw-VL68E	VL FWR3 (G68E)	CTGGCTCCAACTCTGAGAACACGGCCACCC
Rev-VL68E	VL FWR3 (G68E)	GGGTGGCCGTGTTCTCAGAGTTGGAGCCAG

Kunkel mutagenesis		
L3-no INS	NNS in VL CDR3	CCTCCGCCGAATACCACGGGSNNSNNSNNSNNSNN CCACACCTGACAATAATAGTCGGCC
L3-1INS	1xNNS insertion in VL CDR3	CCGCCGAATACCACGGGATCSNNSNNSNNSNNSNN CCACACCTGACAATAATAGTCGGCC
L3-2INS	2xNNS insertion in VL CDR3	CCGCCGAATACCACGGGATCSNNSNNSNNSNNSNN SNNCCACACCTGACAATAATAGTCGGCC
L3-3INS	3xNNS insertion in VL CDR3	CCGCCGAATACCACGGGATCACTSNNSNNSNNSNN SNNSNNCCACACCTGACAATAATAGTCGGCC
L3-4INSI	4xNNS insertion in VL CDR3	CCGCCGAATACCACGGGATCACTACCSNNSNNSNN SNNSNNSNNCCACACCTGACAATAATAGTCGGCC
L3-5INSI	5xNNS insertion in VL CDR3	CCGCCGAATACCACGGGATCACTACCAGTSNNSNN SNNSNNSNNSNNCCACACCTGACAATAATAGTCGG CC
FwL3stop	Stop template	CTATTATTGTCAGGTGTGGGATTAATGA AGTGATCCCGTGGTATTCGG
RevL3stop	Stop template	CCGAATACCACGGGATCACTTCATTAATCCCACAC CTGACAATAATAG

6.2.2. Construction of BAK1 indel libraries using engineered transposons

To build BAK1 libraries with random indels using engineered transposons (D. Jones, 2005, Emond et al, in preparation), the BAK1 coding sequence was cloned into the vector pIDR9 while removing the restriction sites that were used in the following digestion steps (MlyI, NotI and AcuI). For the series of digestion and ligation steps that followed (section 2.2.1), plasmids were treated with Fast Digest restriction enzymes (Thermo Scientific) according to the manufacturer's instructions, digestion products were analysed on agarose gels and purified using the Zymo research, DNA Clean & Concentrator TM-5. The transposition reactions of pIDR9-BAK1 were performed using the TransDel or TransIns transposon and the enzyme MuA Transposase (Finnzymes). The conditions of the reaction were as follows:

A 20 µL reaction containing 300 ng pIDR9-BAK1, 50 ng TransDel or TransIns, MuA Transposase (Finnzymes), reaction buffer (5x) and nuclease-free water. The transposition reaction was incubated for 1 hour at 30 °C followed by heat inactivation at 75 °C for 10 min. The cassettes Ins1, Del1, Del2 and Del3 were extracted from pUC57 plasmids. To remove the 3' overhangs 1U DNA Polymerase I Large (Klenow) Fragment (NEB) was used per 1 µg DNA. A 20 µL reaction containing 500 ng DNA, 2 µL ligase buffer (10x), 3.5 µL dNTP's (200 uM) and Klenow fragment (0.5 U/µL, diluted in NEB buffer 2) was incubated for 15 min at 25 °C, stopped by adding 10 mM EDTA (0.4 µL of EDTA, 0.5 M, pH8), incubated 20 min at 75 °C and purified using a Zymo research column. For 20 µL ligation reactions the T4 ligase (Fermentas) was used in a 1:3 molar ratio of vector: insert. Ligations reactions were concentrated using the Zymo research, DNA Clean & Concentrator TM-5 and then transformed into ≈30 µL of E. cloni® 10G Electrocompetent Cells (Lucigen) (> 5 x 10^9 cfu/µg). After electroporation, 300 µL of recovery medium were added and all the cells were plated in LB medium containing appropriate antibiotics after 45 min incubation at 37 °C (slow shaking). After overnight incubation at 37 °C, cells were scraped and collected by centrifugation and plasmid DNA was extracted using a Qiagen plasmid purification kit. The transformation efficiency was higher than 2 x 10^6 colonies in each step.

6.2.3. Subcloning BAK1 indels libraries and quality control

To subclone the BAK1 indel libraries from pIDR9 to pUC_RD (ribosome display vector; **Figure 51**), NEB restriction enzymes were used according to the manufacturer's instructions. Digestion products were analysed on agarose gels and purified using a Qiagen purification kit. 30 µL ligation reactions including vector to insert DNA ratio of 1:3, T4 DNA ligase (NEB) and reaction buffer were incubated at RT for 2 hours and concentrated using the Zymo Research, DNA Clean & Concentrator TM-5 before electroporation of E. cloni® 10G

Electrocompetent Cells (Lucigen) as described above. The same procedure was followed to subclone the BAK1 3nt-Ins library from pUC_RD to pCANTAB6 (phage display vector) before screening of the library.

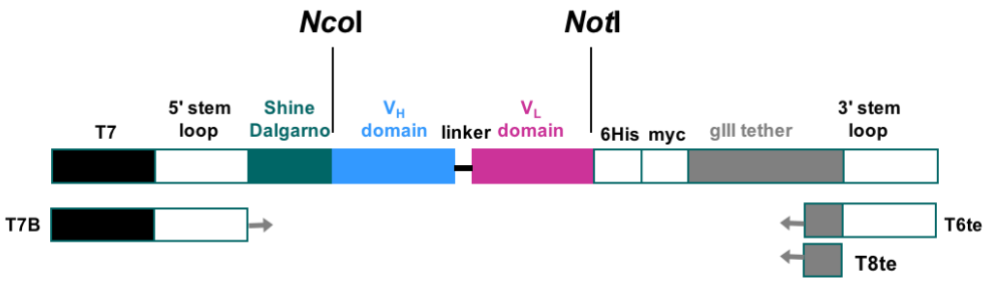


Figure 51. pUC_RD vector

6.2.4. Recombination of BAK1 indels libraries by Step PCR

A protocol for DNA recombination of the BAK1 indel libraries was developed based on the SteP method (Staggered Extension Process In-Vitro DNA Recombination) (Zhao, 1998). The Phusion DNA polymerase (Thermo Scientific, F-531) was chosen instead of the Taq DNA poymerase in order to keep the mutagenesis rate as low as possible. Although different extension times were tested, PCR extension time of 1 sec was chosen because it combined high recombination rate and good quality PCR product. The 50 µL SteP PCR reaction contained a plasmid DNA mix of all the BAK1 indel libraries in pUC_RD, 25 µL of the 2x Phusion master mix, 10uM pUC-Fw and 10uM pUC-Rev. The PCR conditions were as follows: initial denaturation at 95 °C for 30 sec followed by 29 more cycles of denaturation at 95 °C for 10 seconds, annealing at 55 °C for 10 seconds and extension at 72 °C for 1 second and 5 min extra extension at 72 °C in the end. SteP PCR products were purified and concentrated using the DNA Clean & Concentrator TM-5, Zymo Research. This concentrated PCR product was used for *in vitro* transcription for the first round of RD selections. The products of the SteP PCR were also cloned into the pUC_RD vector, individual colonies were

picked in 2 x 96-well plates and sequenced using the same primers (pUC-For/pUC-Rev) to assess the quality of the ribosome display selection input.

6.2.5. Subcloning RD selection outputs into pCANTAB6 for sequencing

To sequence outputs of the ribosome display selections, each output was cloned into pCANTAB6 (phage display vector). Each output was amplified by PCR (with primers T7B/ T6te) and the products were purified from agarose gels using a Qiagen purification kit. 20 μL ligation reactions including vector to insert ratio of 1:3, T4 DNA ligase (NEB) and reaction buffer were incubated at room temperature (RT) for 2 hours, purified and used to transform competent TG1 cells. Individual colonies from each output of rounds 1, 2, 3 and 4 were picked in 2 x 96-well plates per output containing 2 x TYAG medium (2 x TY with 100 μg ml^{-1} ampicillin and 2% glucose) for sequencing, which was performed by the DNA chemistry team at Medimmune.

6.2.6. Site-directed mutagenesis of BAK1 variants

Site-directed mutagenesis of BAK1 scFv gene into the pCANTAB6 vector was performed using non-overlapping oligonucleotides that contained the desired mutation during whole plasmid PCR. Using the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies) following the according to the manufacturer's instructions. The primers that were used are shown in **Table 10** and the BAK1 sequences that were generated are shown in **Figure 52.**

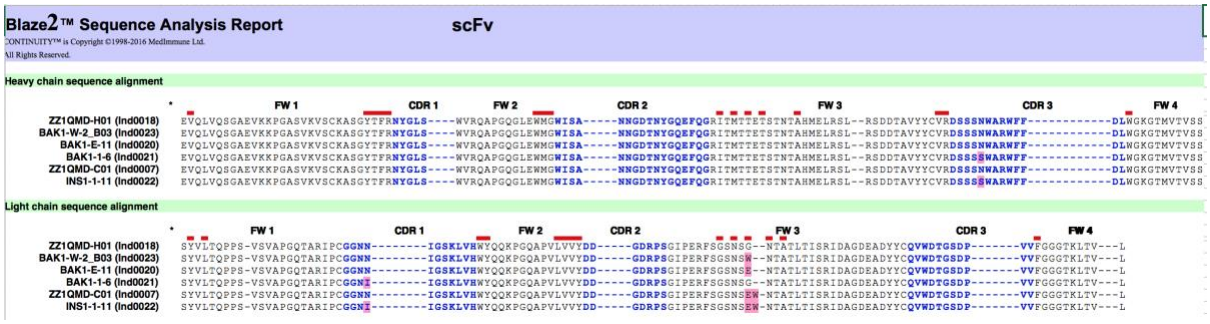


Figure 52. The alignment presents the amino acid sequences of BAK1 IgG variants. The residues are numbered by Kabat numbering system. The CDR residues are highlighted in blue, and introduced point substitutions are marked in pink.

6.2.7. Construction of libraries with insertions in the V_L CDR3 by Kunkel mutagenesis

Kunkel mutagenesis method widely used for the introduction of mutations into phagemids, like pCANTAB6. Kunkel mutagenesis (Sidhu, 1990) was employed to fully randomise blocks of 5 or 6 consecutive amino acid residues per library in the V_L CDR3. To generate the template for saturation mutagenesis a stop codon was introduced into the BAK1 scFv by site-directed mutagenesis (as in section 5.1.6) using the primers FwL3stop and RevL3stop. The BAK1-L3 stop variant was then subcloned into pCantab6 and transformed an *E. coli dut-/ung-* strain called CJ236. Uracil-containing ssDNA (dU-ssDNA) encoding the BAK1-L3 stop variant was purified from M13 phage. Production of dU-ssDNA template was performed by growing CJ236 transformed with stop template DNA into 2 x 5 ml of 2 x TYGAC (2 % glucose, 100 µg/ml ampicillin, 10 µg/ml chloramphenicol) at 37 ˚C (300 rpm) until OD_{600} = 0.8 to 1 (assuming that OD_{600} of 1= 5 x 10^8 cells/ml). Wild-type helper phage was added in multiplicity of infection (MOI) = 10 and allowed the infection by incubating at 37 °C (no shaking) for 10 min. The 1 ml culture was transferred to 30 ml 2 x TYA supplemented with 50 µg/ml kanamycin (to select M13KO7) and 0.25 µg/ml uridine (to enable synthesis of uracil-containing template) which was incubate overnight at 37 °C and 300 rpm. The supernatant was collected after centrifugation and the phage particles were precipitated by adding 1/5 volume of 20% PEG8000/2.5 M NaCl followed by 10 min incubation at room

temperature and centrifugation until the white phage pellet was visible. The pellet was resuspended in PBS and the dU-ssDNA template was purified using a Qiagen Qiaprep Spin M13 kit. Mutagenic oligonucleotides (**Table 10**) encoding the desired mutations (NNS codon insertions) were annealed to the dU-ssDNA template, extended and ligated to form covalently closed circular DNA (ccc-DNA). Annealing was carried out by incubating at 90 °C for 2 min, 50 °C for 3 min and 20 °C for 5 min. The mutagenesis reaction contained 10 mM ATP, 25 mM dNTPs, 100 mM DTT, 30 Weiss units of T4 DNA ligase (6 U/µl) and 30 units T7 DNA polymerase (10 U/µl), which was incubated at 20 ˚C for 3 h. The DNA was purified using a Roche DNA purification kit following the manufacturer's recommendation. Libraries with large number of variants (10^9-10^{10}) were constructed by transforming fresh electrocompetent *E. coli* TG1 cells with the DNA of each reaction.

6.2.8. Preparation of fresh electrocompetent *E. coli* TG1 cells

60 mL of TG1 culture was grown overnight in 2 x TY media at 30 ˚C and 300 rpm. The next day, flasks containing 400 mL medium were inoculated with overnight culture such that the OD_{600} was ~0.1 and were grown at 30 ˚C and 300 rpm until OD_{600} ~0.5-0.6. Once optimum OD_{600} was reached, the cells were chilled for 30 min on ice and 300 mL volumes of culture were divided in centrifugation pots. Following centrifugation for 15 min at 2 ˚C, the supernatant was discarded and the cells were resuspended in 300 mL of ice-cold autoclaved Milli-Q water per pot. The same process was repeated once more and finally the pellet was resuspended in the remaining liquid and transferred into 50 mL Falcon tubes which were filled with cold Milli-Q water. After 10 min of slow speed centrifugation the supernatant was discarded and the cells were resuspended in the remaining liquid. 4 x 400 µl of these cells per library were used for electroporation. Immediately after electroporation 1 mL of 2 x TYG was added to each cuvette and incubated at 37 ˚C and 150 rpm for 1 hour to recover before

plating in large 2 x TYAG bioassay plates (one plate per library), which were incubated overnight at 30 ˚C. In addition, dilution series of each library were plated in 2 x TYAG plates to determine the number of transformed cells, which was used to calculate the size of each library by counting the number of colonies after overnight incubation.

6.2.9. Conversion of BAK1 scFvs to IgGs

To measure the affinities of the variants with insertions in V_L FWR3 or V_L CDR3 by competition ELISA and Biacore, scFv variants were converted to IgG1 format. The BAK1 V_H and V_L sequences were subcloned into the heavy chain vector (pEU1.3) and the light chain vector (pEU4.4), respectively.

6.3. Selections

6.3.1. Ribosome Display selections of recombined BAK1 indel libraries

Isolation of affinity matured antibody fragments (scFv) from the recombined BAK1 indels libraries was done by ribosome display (Jozef Hanes et al., 1998). Selections were performed in solution using biotinylated human IL-13 to allow capture using streptavidin-coated magnetic beads. To isolate fragments with improved affinity over parent a variety of biotinylated IL-13 concentrations below the K_D were used in each round, but only the selection output of the lowest IL-13 concentration which gave a positive outcome was chosen to proceed to the next round. Four rounds of high stringency selections were performed and the concentrations that were finally used were 100, 20, 20 and 20 pM for rounds 1, 2, 3 and 4, respectively.

In vitro transcription of either concentrated PCR product (section 6.1.4) that was the input of the first round of RD selection or amplified selection output between the rounds of selection

was done using the Ribomax Large Scale RNA production system (T7) (Promega), by setting up 50 µL transcription reaction with the following composition: 20 µL linear DNA template (with concentration higher than 50 ng/ µL), 10 µL transcription Buffer (5x), 100 mM rNTP mix and 5 µL T7 enzyme mix. Following incubation of the reaction at 37 ˚C for 2 h, 5 µl of RQ1 RNAse-free DNAse was added and incubated at 37˚C for a further 15 minutes. The mRNA was purified using a ProbeQuant G50 micro column (Amersham Biosciences) according to manufacturer's instructions. The mRNA was quantified by measuring the OD A_{260} of a diluted sample in a spectrophotometer. The mRNA concentrations were determined to be 2.5-3 µg/ mL.

Cell-free translation and stabilisation of scFv-ribosome-mRNA complexes was performed using an S30 extract prepared in-house (Medimmune). The 330 µl translation reactions contained 30 µl mRNA (approximately 6 x 10^{13} molecules), 38.5 µL potassium glutamate (2 M), 26.6. µL magnesium acetate (0.1 M), 5 mg/ mL anti-Protein Disulfide Isomerase (PDI, Sigma), 140 µL S30 translation system and 77 µL Premix X (prepared in-house). The Premix X was contained 250 mM Tris-acetate pH 7.5, 1.75 mM of each amino acid, 10 mM ATP, 2.5 mM GTP, 5 mM cAMP, 150 mM acetylphosphate, 2.5 mg/ mL E. coli tRNA, 0.1 mg/ mL folinic acid and 7.5 % PEG 8000. The translation reaction was stopped by adding 1250 µl chilled HB buffer on ice to stabilise scFv-ribosome-mRNA complexes. 50 mL heparin-block (HB) buffer was prepared by mixing 5 mL 10x *E. coli* wash buffer, 625 µL heparin (200 mg/ mL) and 45 mL sterile-filtered milliQ cold water.

Selection and capture of specifically bound scFv-ribosome-mRNA complexes was performed by mixing 500 µl of stabilised scFv-ribosome-mRNA complexes with 50 µl of de-biotinylated sterile milk to each tube and 30 µl biotinylated antigen at the required concentration. The selections were incubated at 4 ˚C for 2 hours with gentle end-over-end rotation. The complexes were captured by addition 50 µl HB-washed M280 streptavidin-

coated magnetic beads for each selection. KingFisher mL was used to wash the selections in order to remove non-specifically bound complexes. The mRNA was eluted in 220 µL EB20 containing 50 mM Tris-acetate pH 7.5, 150 mM NaCl, 20 mM EDTA, which was then transferred into 400 µl of lysis buffer from the High Pure RNA Isolation kit (Roche) that was used to purify the enriched mRNA following the manufacturer's instructions.

The mRNA was amplified by RT-PCR. The RT reaction contained 4 µL 5x first strand buffer (Invitrogen), 2 µL DTT (100 mM), Reverse (RT) primer, 0.25 µL T8te (100 µM), 0.5 µL dNTP mix (25 mM each), 0.5 µL RNasin (40 U/µl) (Promega) and 0.5 µL Superscript II Reverse Transcriptase (200 U/µl) (Invitrogen). 7.75 µl of this RT Mastermix was mixed with 12.25 µl eluted mRNA. The reactions were incubated in a PCR block at 50 ˚C for 30 min. A PCR reaction was then carried out to amplify the cDNA by mixing 10 µl cDNA with 90 µL of master mix containing 50 µL PCR master mix (2x) (Thermo), 2.5 µL primer SDCAT-BAK1 (10 µM) and 2.5 µL primer T7te (10 µM) and 5 µL DMSO (Sigma). The thermocycling conditions were: denaturation at 95 ℃ for 30 sec followed by 29 more cycles of denaturation at 95 ℃ for 10 sec, annealing at 62 ℃ for 30 sec and extension at 72 ℃ for 1.45 minutes.

6.3.2. Phage Display selections of BAK1 V_L CDR3 insertion libraries

Phage display selections were used to select the variants with improved affinity from the BAK1 V_L CDR3 insertion libraries. Enrichment of improved variants was achieved by equilibrium selections by reducing the concentration of the antigen by 10-fold after each selection round (Hawkins, Russell, & Winter, 1992; Schier et al., 1996; Groves & Nickson, 2012). These selections were performed in solution using a biotinylated IL-13, which was allowed to reach a binding equilibrium with the library of different phage antibodies. The incubation was followed by capturing on streptavidin-coated magnetic beads, washing and eluting the phages which were bound to the soluble biotinylated IL-13.

In order to carry out phage display selections the bacteriophage particles expressing the scFv were produced. Phage particles from the large libraries containing more than 10^9 variants were generated by large-scale (400 mL) growth and helper phage rescue of the scFv library followed by concentration and crude purification of the bacteriophage particles using PEG. 400 mL volume of 2TY media containing 100 µg/ mL ampicillin and 2% glucose were inoculated with frozen library glycerol stock, so that at least a 10-fold excess of the transformed library size was inoculated and grown at 37 ˚C and 280 rpm to OD600 = 0.5-1.0 (mid-log phase). Superinfection with helper phage in order to rescue infective phagemid particles was done by adding 80 µl of $3x10^{13}$ / mL $M13KO7^{trp}$ (multiplicity of infection (MOI) = 10). After infection, 1 hour incubation at 37 ˚C (slow shaking) and centrifugation, the cell pellet was resuspended in 400 mL of 2 x TY medium with 100 µg/ mL ampicillin and 50 µg/ mL kanamycin (no glucose). The removal of glucose allows for expression of scFv-g3p fusion from the phagemid throughout the course of overnight incubation with rapid shaking (280 rpm) at 25 ˚C.

The next day, phage particles were precipitated using polyethylene glycol (PEG). Briefly, this was achieved by centrifugation of the cultures and collection of the supernatant. 3/10 volume of chilled 20% PEG8000/2.5 M NaCl was added and incubated on ice for 1 hour for phage precipitation. The phage pellet was concentrated by spinning, that was repeated to collect all the phages, which were resuspended in 10 mL TE pH 8.0. After addition of 3/10 volume of chilled 20% PEG8000/ 2.5 M NaCl and 1 hour precipitation on ice each phage pellet was centrifugated and resuspended in a small volume 2-5 mL TE pH 8.0. The phage titre was at least 10^{13} cfu/ mL and was used for the first selection round.

For selection, the phage particles were pre-blocked by adding 50 µl phage aliquot to 450 µl of 3% (w/v) skimmed milk powder in PBS (MPBS), which was incubated at room temperature for 1 hour. Magnetic streptavidin-coated beads (Dynabeads) were also pre-

washed with PBS and resuspended in 1 mL of 3% MPBS. The phage library was de-selected against streptavidin beads by adding 50 µl of the blocked beads to the blocked library followed by 1 hour incubation at RT and removal of the beads from the blocked library with a magnetic separator. 500 µl aliquots of each phage library were then incubated with the required concentration of biotinylated antigen, which was rotated end-over-end on a rotor for 1 hour. Bound phage particles were captured by adding 50 µl blocked streptavidin magnetic beads to each selection and the beads were washed (with PBS-Tween) to remove unbound phage by using the Kingfisher 96. Elution of the bound phage was achieved by adding 200 µl freshly made trypsin solution (10 µg/ mL trypsin in 0.1 M sodium phosphate buffer) and incubation for 30 min at 37 ˚C while shaking at 600 rpm. The beads were pelleted with a magnetic separator and the supernatant was used to infect 1 mL of mid-log *E. coli* TG1 cells, which were incubated for 1 hour at 37 ℃ with slow shaking. The cells of each selection output were plated onto 2 x TYAG agar bioassay plates.

6.4. Protein Expression, Purification and Biotinylation

The V_H and V_L coding sequences were cloned into appropriate vectors which were given to the Medimmune expression team who implemented the scFv and IgG expression and purification. The protocols that they followed are described below.

6.4.1 Periplasmic Extraction of scFv from *E. coli*

The production of periplasmic extracts containing his-tagged scFv from *E. coli* (TG1 strain) was implemented as follows: Initially 96 deep well plates filled with 1000 µL medium/well grow-up media (2xTYA, 0.1% glucose and 100 mg/mL ampicillin) were inoculated with culture from a freshly grown plate or frozen glycerol plate. These cultures were incubated at

37 °C for approximately 5 hours shaking at 250 rpm and the scFv expression was induced by adding 1 mM IPTG. After overnight (16 hours) incubation at 30 °C shaking at 250 rpm the cells were pelleted by high speed centrifugation at 4 °C for 15 min. The pellet was resuspended in 400 µl of osmotic shock buffer (MES) containing MOPS (pH 7.4), EDTA and sorbitol (final pH 7.4) and incubated at 4 °C for 30 minutes to lyse the cells. The periplasmic extracts were collected by centrifugation at high speed in chilled centrifuge for 15 min.

6.4.2. His-tagged scFv affinity purification

His-tagged scFvs cloned into the pCantab6 vector were expressed in an *E.coli* (TG1 strain). Overnight cultures (10 mL) were used to inoculate 400 mL of 2 xTY medium (100 µg/mL ampicillin and 0.1% glucose). When the bacteria reached OD_{600} of 0.6 (30 °C, 300 rpm, ~2.5 hours) the scFv expression was induced by adding 1 mM IPTG and incubation for 3 hours at 30 °C. After centrifugation of the cultures (high speed, 10 min at 4 °C) the cell pellet was resuspended in 10 mL of osmotic buffer (TES) containing 200 mM Tris-HCl, 0.5 mM EDTA, 0.5 M sucrose, pH 8.0. Cold 1:5 TES was added (15 mL) and incubated on ice for 30 min. Finally, the cell debris were pelleted by centrifugation in a benchtop centrifuge (high speed, 30 min at 4 °C) and the periplasmic extract was purified on nickel sepharose (GE Healthcare) packed columns followed by buffer exchanged to PBS (NAP-10 columns; GE Healthcare).

6.4.3. Transient transfection of mammalian cells for IgG production

Co-transfection of the heavy chain vector pEU1.3 and lambda light chain vector pEU4.4 into HEK EBNA-293 allowed whole IgG to be expressed and purified by protein A affinity chromatography (GE Healthcare, Little Chalfont, UK). HEK-EBNA cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) containing 10% foetal bovine serum, 250 µg mL^{-1} geneticin (to select for the EBNA marker gene) and 1% non-essential amino acids. Prior

to transfection the medium (20 mL) was replaced with fresh DMEM containing 2% foetal bovine serum and 1x penicillin/streptomycin.

The calcium phosphate transfection method was used to introduce the plasmid DNA into the mammalian cells. 100 μL of $CaCl_2$ (2.5 M) were added in a mixture containing 50 μg of each plasmid DNA in 100 μL pre-warmed filter sterilised distilled water. The pre-warmed DNA mixture was added to 2 x HBS buffer (50 mM Hepes, 280 mM NaCl and 1.5 mM Na_2HPO_4) dropwise whilst mixing on a vortex mixer and incubated for 2-5 min at 37 ℃. The HBS/ DNA solution was added to HEK-EBNA cells, which were incubated overnight at 37 ℃ in 5% CO_2 concentration with humidification.

After overnight incubation the medium was removed and replaced with an appropriate volume of fresh CD-CHO medium containing 8 mM glutamine, 1 x penicillin/streptomycin supplement, 1 x hypoxanthine/thymidine supplement and 1 x OptiMab supplements A and B (25 mL for T150, 60 mL for triple T150 flasks). This was carried out on the first day after transfection has taken place while the harvest was taken 8 days post-transfection, filter sterilised using 0.22 μm steriflips and stored at 4 ℃ until purification.

6.5. Binding assays and measurements

6.5.1. IL-13 biotinylation

Biotinylation of recombinant human IL-13 (PeproTech) was performed by biotinylating surface lysine residues using EZ link NHS-LC-Biotin according to the manufacturer's instructions (Pierce protein research products, Thermo Fisher Scientific) and confirmed by MALDIN TOF mass spectrometry (Medimmune), which was carried out by Alan Sandercock. Biotinylated IL-13 was stored in -80°C in PBS/0.1% BSA after showing that biotinylation did not affect its activity by ELISA, which was done by coating with the same

concentration of IL-13 either biotinylated or not, titration of BAK1-IgG1 (human) and detection using anti-human IgG-HRP antibody.

6.5.2. Homogenous Time Resolved Fluorescence Assays

The 3nt-Ins BAK1 library with random single amino acid insertions was tested by an HTRF binding assay (**Figure 53**), which was performed in 384-well shallow well plates (Corning/Costar). The assay buffer was composed of 0.1% bovine serum albumen, 0.4 M potassium fluoride, 1 x PBS. Each well contained the following reagents diluted in assay buffer (final concentration): streptavidin cryptate (CisBio) (1 nM), biotinylated IL-13 (5 nM), 9-fold dilution of periplasmic fraction (his-tagged scFv), anti-his-XL665 (CISBIO) (20 nM) to give a total assay volume of 20 µL/well.

Screening of the phage display selection outputs of the BAK1 V_L CDR3 insertion libraries for improved affinity was done using an HTRF competition assay (**Figure 53**), which was performed in shallow 384-well plates (Corning/Costar). Each well contained the following reagents diluted in assay buffer: (Eu-labelled) streptavidin cryptate (CisBio) (1 nM), biotinylated IL-13 (1.25 nM), 3-fold dilution of periplasmic fraction (his-tagged scFv), Dylight 650-labelled BAK1-IgG1 (1 nM) to give a total assay volume of 20 µL/well. For further screening for improved affinity, the same assay was used to determine the IC50s of purified his-tagged scFv variants by titration of the scFv with 2-fold dilutions, starting from 100 nM as the highest concentration.

For both assays the reaction mixture was incubated at RT for 4 hours to reach equilibrium binding and then read on a Perkin Elmer Envision plate reader using the following settings: 100 flashes, delay 70 µs, cycle 2000, Excitation UV2 (TRF) 320 nm, Emission APC 665 (Bandwidth 7.5nm), Emission Rhodamine 590 (Bandwidth 20 nm), mirror D400/630. The HTRF ratio was calculated by (665 nm emission/620 nm emission) x 10,000. Non-specific

binding (NSB) was determined by calculation of the HTRF ratio in control wells. Specific binding was calculated by (% delta Fluorescence-NSB)/(Total-NSB) x 100. All data was analysed in Graphpad Prism.

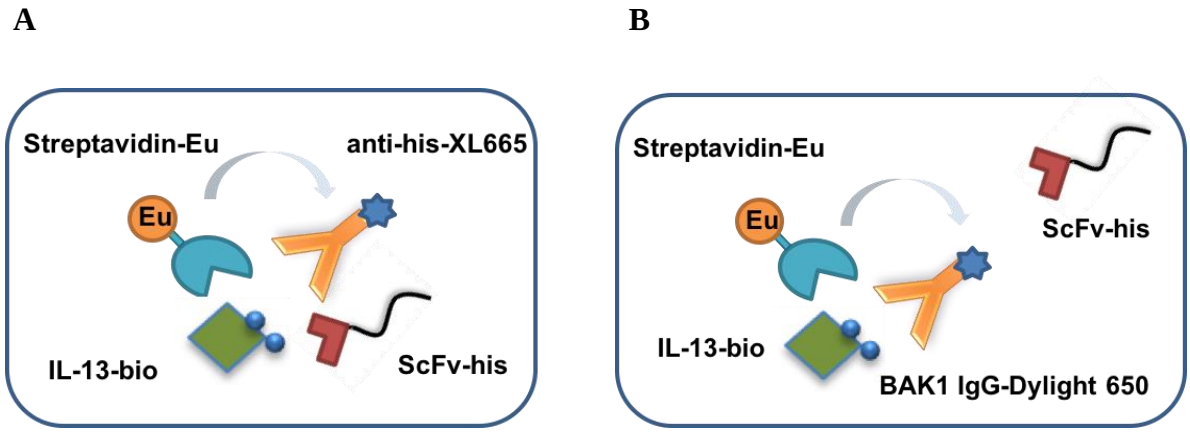


Figure 53. A) HTRF binding assay FRET signal was generated upon binding of biotinylated IL-13 to his-tagged scFv using Eu-labelled streptavidin (donor) and XL665-labbeled anti-his-tag antibody (acceptor). B) Using an HTRF competition assay, his-tagged scFv binding to biotinylated IL-13 competed binding of the Dylight 650-labelled BAK1-IgG (acceptor) resulting in reduced FRET signal, which was detected using Eu-labelled streptavidin (donor).

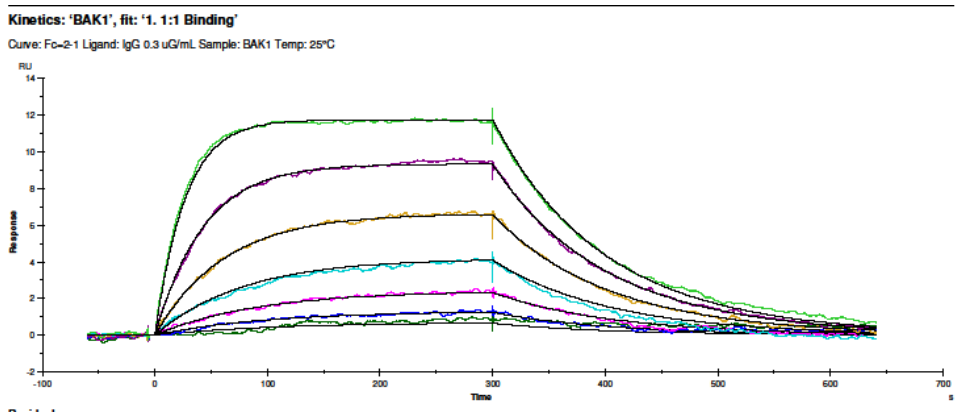
6.5.3. Competition ELISA

A competition ELISA was carried out to identify improved BAK1 IgG1 variants from the V_L CDR3 insertion libraries. Nunc MaxiSorp plates (Thermo Fisher Scientific) were coated with BAK1-IgG in PBS (10 µg/mL) and incubated overnight at 4 ˚C. The wells were blocked with PBS/3% milk for 1 hour at RT and washed with PBS. 50 µL of the mix of a 2-fold serial dilution of the BAK1-IgG (30 µL, starting from 10 µg/mL, final 5 µg/mL) with IL-13-bio (30 µL, 0.2 µg/mL, final 0.1 µg/mL) in PBS-3% milk was incubated at RT (shaking) for 1 hour. After washing with PBS/0.01% Tween binding was detected using a streptavidin conjugated to HRP antibody. Colorimetric detection was done by the addition of 50 µL of the substrate, 3,3′,5,5′-Tetramethylbenzidine (Sigma) and the reaction was stopped with 50 µL 0.5 M H_2SO_4.

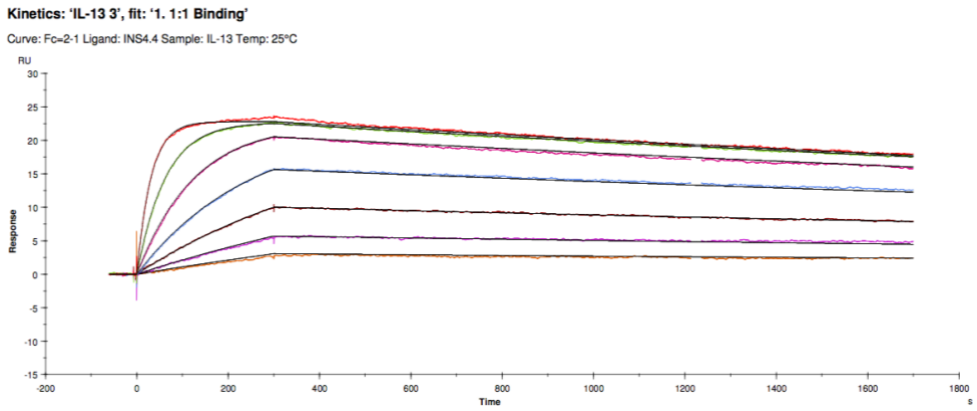
6.5.4. Surface Plasmon Resonance affinity measurements

K_D measurements of the IgG variants were performed using the Biacore T200 (Cambridge University, Biochemistry Department). Biosensor affinity measurements were performed using protein G′-mediated capture immobilisation of the human antibodies by flowing the recombinant human IL-13 (PeproTech) (analyte). A sensor surface pre-immobilized with a recombinant Protein G variant (Sensor Chip Protein G, GE Helthcare) was used for binding of the human IgG1 variants. The buffer (HBS-EP+) contained 0.01 M HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.05% v/v Surfactant P20. IgG analysis was performed by first capturing ~100 resonance units (RUs) of the antibody on the protein G sensor chip (60 s capture at 5 μl/min) and running a 2-fold serial dilution, starting from 6.25 nM IL-13 as analyte at 50 μL/min flow rate with an association time of 5 min. Depending on the strength of the interaction, dissociation was followed for 300 -1800 sec. Regeneration of protein G′ surfaces was performed with two consecutive 20 sec injections of 6 M guanidinium hydrochloride in Dulbecco's PBS.

A



B



C

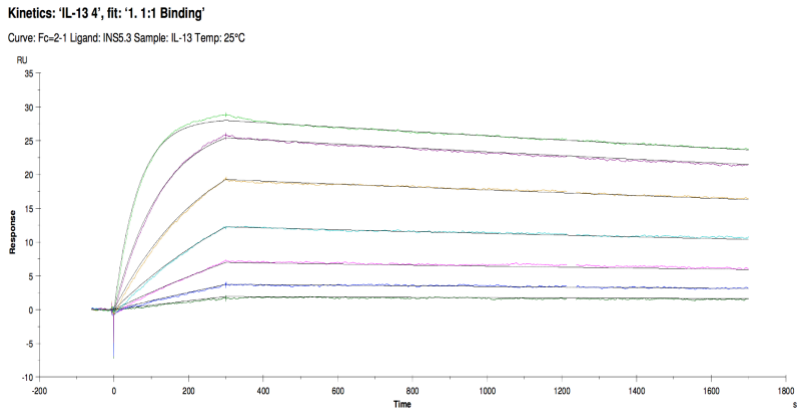


Figure 54. Sensorgram of the A) BAK1 B) INS4.4 and C) INS5.3 IgG1 variants. The data were fitted to 1:1 Langmuir binding model using the Biacore Evaluation software 2.0.3.

6.5.5. Thermal stability measurements

Thermal stability measurements of the BAK1 IgG variants were performed to measure unfolding by monitoring unfolding using SYPRO Orange fluorescence. Purified IgGs were diluted to 5 μM in a Tris buffer (TrisNHCl 50 nM and NaCl 150 nM). SYPRO Orange dye (Invitrogen) was used at a final concentration of 20x. Melting curves were recorded over a temperature range of 25 ℃ to 99 ℃ in duplicate 20 μL reaction volumes using a real-time PCR machine (Rotor-Gene 6000; Corbett Research) with filters to excite at 460 nm and measure emission at 510 nm as the temperature continually increased (ramp rate of 0.5 ℃/min). Data were fitted using GraphPad Prism software.

7. References

Adams, G. P., Schier, R., McCall, A. M., Simmons, H. H., Horak, E. M., Alpaugh, R. K., … Weiner, L. M. (2001). High affinity restricts the localization and tumor penetration of single-chain Fv antibody molecules. *Cancer Research*, *61*(12), 4750–4755. https://doi.org/10.1234/12345678

Afriat-jurnou, L., Jackson, C. J., & Taw, D. S. (2012). Reconstructing a Missing Link in the Evolution of a Recently Diverged Phosphotriesterase by Active-Site Loop Remodeling.

Al-Lazikani, B., Lesk, A. M., & Chothia, C. (1997). Standard conformations for the canonical structures of immunoglobulins 1 1Edited by I. A. Wilson. *Journal of Molecular Biology*, *273*(4), 927–948. https://doi.org/10.1006/jmbi.1997.1354

Almagro, J. C. (2004). Identification of differences in the specificity-determining residues of antibodies that recognize antigens of different size: Implications for the rational design of antibody repertoires. *Journal of Molecular Recognition*, *17*(2), 132–143. https://doi.org/10.1002/jmr.659

Alt, F. W., & Baltimore, D. (1982). Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(13), 4118–4122. https://doi.org/10.1073/pnas.79.13.4118

Arakawa, H. (2002). Requirement of the Activation-Induced Deaminase (AID) Gene for Immunoglobulin Gene Conversion. *Science*, *295*(5558), 1301–1306. https://doi.org/10.1126/science.1067308

Arnaout, R., Lee, W., Cahill, P., Honan, T., Sparrow, T., Weiand, M., … Koralov, S. B. (2011). High-resolution description of antibody heavy-chain repertoires in humans. *PLoS ONE*, *6*(8). https://doi.org/10.1371/journal.pone.0022365

Arpino, J., Reddington, S. C., Halliwell, L. M., Rizkallah, P. J., & Jones, D. D. (2014). Article Random Single Amino Acid Deletion Sampling Unveils Structural Tolerance

and the Benefits of Helical Registry Shift on GFP Folding and Structure. *Structure/Folding and Design*, *22*(6), 889–898. https://doi.org/10.1016/j.str.2014.03.014

Bachl, J., Carlson, C., Gray-Schopfer, V., Dessing, M., & Olsson, C. (2001). Increased Transcription Levels Induce Higher Mutation Rates in a Hypermutating Cell Line. *The Journal of Immunology*, *166*(8), 5051–5057. https://doi.org/10.4049/jimmunol.166.8.5051

Badran, A. H., Guzov, V. M., Huai, Q., Kemp, M. M., Vishwanath, P., Kain, W., … Liu, D. R. (2016). Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. *Nature*, *533*(7601), 58–63. https://doi.org/10.1038/nature17938

Baldwin, A. J., Busse, K., Simm, A. M., & Jones, D. D. (2008). Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx). *Nucleic Acids Research*, *36*(13). https://doi.org/10.1093/nar/gkn358

Barbas, C. F., Hu, D., Dunlop, N., Sawyer, L., Cababa, D. , Hendry, R. M., … Burton, D. R. (1994). In vitro evolution of a neutralizing human antibody to human immunodeficiency virus type 1 to enhance affinity and broaden strain cross-reactivity. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(9), 3809–3813. https://doi.org/10.1073/pnas.91.9.3809

Barrick, J. E., Takahashi, T., Ren, J., Xia, T., & Roberts, R. W. (2001). Large libraries reveal diverse solutions to an RNA recognition problem. *Proc. Natl. Acad. Sci. USA*, *98*(22), 12374–12378. https://doi.org/10.1073/pnas.221467798

Batista, F. D., & Neuberger, M. S. (1998). Affinity dependence of the B cell response to antigen: A threshold, a ceiling, and the importance of off-rate. *Immunity*, *8*(6), 751–759. https://doi.org/10.1016/S1074-7613(00)80580-4

Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., & Tawfik, D. S. (2006). Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature*, *444*(7121), 929–932. https://doi.org/10.1038/nature05385

Bjorkman, P. (2012). Structures of Broadly Neutralizing Anti-HIV Antibodies that Target the CD4 Binding Site on the HIV Envelope. *Science Highlight*, *3*(March), 4–6.

Boder, E. T., Midelfort, K. S., & Wittrup, K. D. (2000). Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proceedings of the National Academy of Sciences*, *97*(20), 10701–10705. https://doi.org/10.1073/pnas.170297297

Boder, E. T., & Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnology*, *15*(6), 553–557. https://doi.org/10.1038/nbt0697-553

Bowers, P. M., Horlick, R. A., Neben, T. Y., Toobian, R. M., Tomlinson, G. L., Dalton, J. L., … King, D. J. (2011). Coupling mammalian cell surface display with somatic hypermutation for the discovery and maturation of human antibodies. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(51), 20455–60. https://doi.org/10.1073/pnas.1114010108

Bowers, P. M., Verdino, P., Wang, Z., Da Silva Correia, J., Chhoa, M., Macondray, G., … King, D. J. (2014). Nucleotide insertions and deletions complement point mutations to massively expand the diversity created by somatic hypermutation of antibodies. *Journal of Biological Chemistry*, *289*(48), 33557–33567. https://doi.org/10.1074/jbc.M114.607176

Boyd, S. D., Gaeta, B. A., Jackson, K. J., Fire, A. Z., Marshall, E. L., Merker, J. D., … Collins, A. M. (2010). Individual Variation in the Germline Ig Gene Repertoire Inferred from Variable Region Gene Rearrangements. *The Journal of Immunology*, *184*(12), 6986–6992. https://doi.org/10.4049/jimmunol.1000445

Bradbury, A. R. M., Sidhu, S., Dübel, S., & Mccafferty, J. (2011). NIH Public Access, *29*(3), 245–254. https://doi.org/10.1038/nbt.1791.Beyond

Brekke, O. H., & Sandlie, I. (2003). Therapeutic antibodies for human diseases at the dawn of the twenty-first century. *Nature Reviews Drug Discovery*, *2*(1), 52–62. https://doi.org/10.1038/nrd984

Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A., & Liu, D. R. (2014). Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature Chemical Biology*, *10*(3), 216–222. https://doi.org/10.1038/nchembio.1453

Carter, P., & Merchant, A. M. (1997). Engineering antibodies for imaging and therapy. *Current Opinion in Biotechnology*, *8*(4), 449–454. https://doi.org/10.1016/S0958-1669(97)80067-5

Cauerhff, A., Goldbaum, F. A., & Braden, B. C. (2004). Structural mechanism for affinity maturation of an anti-lysozyme antibody. *Proceedings of the National Academy of Sciences*, *101*(10), 3539–3544. https://doi.org/10.1073/pnas.0400060101

Cèbe, R., & Geiser, M. (2000). Size of the ligand complex between the N-terminal domain of the gene III coat protein and the non-infectious phage strongly influences the usefulness of in vitro selective infective phage technology. *The Biochemical Journal*, *352 Pt 3*, 841–9.

Chailyan, A., Marcatili, P., Cirillo, D., & Tramontano, A. (2011). Structural repertoire of immunoglobulin λ light chains. *Proteins: Structure, Function and Bioinformatics*, *79*(5), 1513–1524. https://doi.org/10.1002/prot.22979

Chan, A. C., & Carter, P. J. (2010). Therapeutic antibodies for autoimmunity and inflammation. *Nature Reviews Immunology*, *10*(5), 301–316. https://doi.org/10.1038/nri2761

Chothia, C., & Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *Journal of Molecular Biology*, *196*(4), 901–917. https://doi.org/10.1016/0022-2836(87)90412-8

Clark, L., Ganesan, S., Papp, S., & van Vlijmen, H. W. T. (2006). Trends in antibody sequence changes during the somatic hypermutation process. *Journal of Immunology (Baltimore, Md. : 1950)*, *177*(1), 333–340. https://doi.org/10.4049/jimmunol.177.1.333

Cohen-Solal, J. (2004). Fc γ receptors. *Immunology Letters*, *92*(3), 199–205. https://doi.org/10.1016/j.imlet.2004.01.012

Collis, A., Brouwer, A. P., & Martin, A. C. R. (2003). Analysis of the antigen combining site: Correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. *Journal of Molecular Biology*, *325*(2), 337–354. https://doi.org/10.1016/S0022-2836(02)01222-6

Couzin-Frankel, J. (2013). Cancer Immunotherapy. *Science*, *342*(6165), 1432–1433. https://doi.org/10.1126/science.342.6165.1432

Daugherty, P. S., Chen, G., Iverson, B. L., & Georgiou, G. (2000). Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(5), 2029–2034. https://doi.org/10.1073/pnas.030527597

de los Rios, M., Criscitiello, M. F., & Smider, V. V. (2015). Structural and genetic diversity in antibody repertoires from diverse species. *Current Opinion in Structural Biology*, *33*, 27–41. https://doi.org/10.1016/j.sbi.2015.06.002

de Visser, J. A. G. M., & Krug, J. (2014). Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics*, *15*(7), 480–490. https://doi.org/10.1038/nrg3744

de Wildt, R. M., van Venrooij, W. J., Winter, G., Hoet, R. M., & Tomlinson, I. M. (1999). Somatic insertions and deletions shape the human antibody repertoire. *Journal of Molecular Biology*, *294*(3), 701–710. https://doi.org/10.1006/jmbi.1999.3289

Deng, L.W., & Perham, R. N. (2002). Delineating the site of interaction on the pIII protein of filamentous bacteriophage fd with the F-pilus of Escherichia coli. *Journal of Molecular Biology*, *319*(3), 603–14. https://doi.org/10.1016/S0022-2836(02)00260-7

Desmyter, A., Spinelli, S., Roussel, A., & Cambillau, C. (2015). Camelid nanobodies: Killing two birds with one stone. *Current Opinion in Structural Biology*, *32*, 1–8. https://doi.org/10.1016/j.sbi.2015.01.001

Di Noia, J. M., & Neuberger, M. S. (2007). Molecular Mechanisms of Antibody Somatic Hypermutation. *Annual Review of Biochemistry*, *76*, 1–22. https://doi.org/10.1146/annurev.biochem.76.061705.090740

Dickinson, B. C., Leconte, A. M., Allen, B., Esvelt, K. M., & Liu, D. R. (2013). Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(22), 9007–12. https://doi.org/10.1073/pnas.1220670110

Diskin, R., Scheid, J. F., Marcovecchio, P. M., West, A. P., Klein, F., Gao, H., … Bjorkman, P. J. (2011). Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science (New York, N.Y.)*, *334*, 1289–93. https://doi.org/10.1126/science.1213782

Dörner, T., Brezinschek, H. P., Brezinschek, R. I., Foster, S. J., Domiati-Saad, R., & Lipsky, P. E. (1997). Analysis of the frequency and pattern of somatic mutations within nonproductively rearranged human variable heavy chain genes. *Journal of Immunology (Baltimore, Md. : 1950)*, *158*(6), 2779–89. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9058813

Dove, S. L., & Hochschild, A. (1998). Conversion of the ω subunit of Escherichia coli RNA polymerase into a transcriptional activator or an activation target. *Genes and Development*, *12*(5), 745–754. https://doi.org/10.1101/gad.12.5.745

Dueñas, M., & Borrebaeck, C. A. K. (1994). Clonal Selection and Amplification of Phage Displayed Antibodies by Linking Antigen Recognition and Phage Replication. *Bio/Technology*, *12*, 999. Retrieved from https://doi.org/10.1038/nbt1094-999

Dufner, P., Jermutus, L., & Minter, R. (2006). Harnessing phage and ribosome display for antibody optimisation. *Trends in Biotechnology*, *24*(11), 523–529. https://doi.org/10.1016/j.tibtech.2006.09.004

Edwards, W. R., Busse, K., Allemann, R. K., & Jones, D. D. (2008). Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. *Nucleic Acids Research*, *36*(13). https://doi.org/10.1093/nar/gkn363

Emond, S., Mondon, P., Pizzut-Serin, S., Douchy, L., Crozet, F., Bouayadi, K., … Remaud-Simeon, M. (2008). A novel random mutagenesis approach using human mutagenic DNA polymerases to generate enzyme variant libraries. *Protein Engineering, Design and Selection*, *21*(4), 267–274. https://doi.org/10.1093/protein/gzn004

Esvelt, K. M., Carlson, J. C., & Liu, D. R. (2011). A system for the continuous directed evolution of biomolecules. *Nature*, *472*(7344), 499–503. https://doi.org/10.1038/nature09929

Fanning, S. W., & Horn, J. R. (2011). An anti-hapten camelid antibody reveals a cryptic binding site with significant energetic contributions from a nonhypervariable loop. *Protein Science*, *20*(7), 1196–1207. https://doi.org/10.1002/pro.648

Feldhaus, M. J., Siegel, R. W., Opresko, L. K., Coleman, J. R., Feldhaus, J. M. W., Yeung, Y. A., … Wittrup, K. D. (2003). Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library. *Nature Biotechnology*, *21*(2), 163–170. https://doi.org/10.1038/nbt785

Finch, D. K., Sleeman, M. A., Moisan, J., Ferraro, F., Botterell, S., Campbell, J., … Lowe, D. C. (2011). Whole-molecule antibody engineering: Generation of a high-affinity anti-IL-6 antibody with extended pharmacokinetics. *Journal of Molecular Biology*, *411*(4), 791–807. https://doi.org/10.1016/j.jmb.2011.06.031

Finlay, W. J., & Almagro, J. C. (2012). Natural and man-made V-gene repertoires for antibody discovery. *Frontiers in Immunology*, *3*(NOV), 1–18. https://doi.org/10.3389/fimmu.2012.00342

Flajnik, M. F., Deschacht, N., & Muyldermans, S. (2011). A case of convergence: Why did a simple alternative to canonical antibodies arise in Sharks and Camels? *PLoS Biology*, *9*(8). https://doi.org/10.1371/journal.pbio.1001120

Foote, J., & Eisent, H. N. (1995). Commentary Kinetic and affinity limits on antibodies produced during immune responses, *92*(February), 1254–1256.

Foote, J., & Winter, G. (1992). Antibody framework residues affecting the conformation of the hypervariable loops. *Journal of Molecular Biology*, *224*(2), 487–499. https://doi.org/10.1016/0022-2836(92)91010-M

Fujii, R., Kitaoka, M., & Hayashi, K. (2006). RAISE: A simple and novel method of generating random insertion and deletion mutations. *Nucleic Acids Research*, *34*(4). https://doi.org/10.1093/nar/gnj032

Georgiou, G., Ippolito, G. C., Beausang, J., Busse, C. E., Wardemann, H., & Quake, S. R. (2014). The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nature Biotechnology*, *32*(2), 158–168. https://doi.org/10.1038/nbt.2782

Glanville, J., Zhai, W., Berka, J., Telman, D., Huerta, G., Mehta, G. R., … Pons, J. (2009). Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proceedings of the National Academy of Sciences*, *106*(48), 20216–20221. https://doi.org/10.1073/pnas.0909775106

Gong, L. I., Suchard, M. A., & Bloom, J. D. (2013). Stability-mediated epistasis constrains the evolution of an influenza protein. *ELife*, *2013*(2), 1–19. https://doi.org/10.7554/eLife.00631

Griffiths, D., Williams, S. C., Hartley, O., Tomlinson, I. M., Waterhouse, P., Crosby, W. L., … Allison, T. J. (1994). Isolation of high affinity human antibodies directly from large synthetic repertoires. *The EMBO Journal*, *13*(14), 3245–60. https://doi.org/10.1016/0168-9525(94)90126-0

Haapa, S., Taira, S., Heikkinen, E., & Savilahti, H. (1999). An efficient and accurate integration of mini-Mu transposons in vitro: A general methodology for functional genetic analysis and molecular biology applications. *Nucleic Acids Research*, *27*(13), 2777–2784. https://doi.org/10.1093/nar/27.13.2777

Hallet, B., Sherratt, D. J., & Hayes, F. (1997). Pentapeptide scanning mutagenesis: Random insertion of a variable five amino acid cassette in a target protein. *Nucleic Acids Research*, *25*(9), 1866–1867. https://doi.org/10.1093/nar/25.9.1866

Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R., & Plckthun, A. (1998). Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries. *Biochemistry*, *95*(November), 14130–14135. https://doi.org/10.1073/pnas.95.24.14130

Hanes, J., & Plückthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(10), 4937–42. https://doi.org/10.1073/pnas.94.10.4937

Hanes, J., Schaffitzel, C., Knappik, a, & Plückthun, a. (2000). Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nature Biotechnology*, *18*(12), 1287–92. https://doi.org/10.1038/82407

Hawkins, R. E., Russell, S. J., & Winter, G. (1992). Selection of phage antibodies by binding

affinity. Mimicking affinity maturation. *Journal of Molecular Biology*, *226*(3), 889–896. https://doi.org/10.1016/0022-2836(92)90639-2

Hoffmann-Berling, H., & Mazé, R. (1964). Release of male-specific bacteriophages from surviving host bacteria. *Virology*, *22*(3), 305–313. https://doi.org/10.1016/0042-6822(64)90021-2

Holliger, P., & Hudson, P. J. (2005). Engineered antibody fragments and the rise of single domains. *Nature Biotechnology*, *23*(9), 1126–1136. https://doi.org/10.1038/nbt1142

Holliger, P., Riechmann, L., & Williams, R. L. (1999). Crystal Structure of the Two N-terminal Domains of Ê : Evidence for g3p from Filamentous Phage fd at 1 . 9 A Conformational Lability, 649–657.

Hoogenboom, H. R. (2005). Selecting and screening recombinant antibody libraries. *Nature Biotechnology*, *23*(9), 1105–1116. https://doi.org/10.1038/nbt1126

Hoogenboom, H. R., & Winter, G. (1992). By-passing immunisation. Human antibodies from synthetic repertoires of germline VH gene segments rearranged in vitro. *Journal of Molecular Biology*, *227*(2), 381–388. https://doi.org/10.1016/0022-2836(92)90894-P

Houlihan, G., Gatti-Lafranconi, P., Kaltenbach, M., Lowe, D., & Hollfelder, F. (2014). An experimental framework for improved selection of binding proteins using SNAP display. *Journal of Immunological Methods*, *405*, 47–56. https://doi.org/10.1016/j.jim.2014.01.006

Husimi, Y., Nishigaki, K., Kinoshita, Y., & Tanaka, T. (1982). Cellstat - A continuous culture system of a bacteriophage for the study of the mutation rate and the selection process at the DNA level. *Review of Scientific Instruments*, *53*(4), 517–522. https://doi.org/10.1063/1.1137002

Hwang, J. K., Wang, C., Du, Z., Meyers, R. M., Kepler, T. B., Neuberg, D., … Alt, F. W. (2017). Sequence intrinsic somatic mutation mechanisms contribute to affinity maturation of VRC01-class HIV-1 broadly neutralizing antibodies. *Proceedings of the National Academy of Sciences*, 201709203. https://doi.org/10.1073/pnas.1709203114

Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J., & Plückthun, A. (2001). Tailoring in

vitro evolution for protein affinity or stability. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(1), 75–80. https://doi.org/10.1073/pnas.011311398

Jiménez, J. I., Xulvi-brunet, R., Campbell, G. W., Turk-macleod, R., & Chen, I. A. (2013). Comprehensive experimental fi tness landscape and evolutionary network for small RNA. https://doi.org/10.1073/pnas.1307604110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1307604110

Jones, D. (2005). Triplet nucleotide removal at random positions in a target gene: The tolerance of TEM-1 ??-lactamase to an amino acid deletion. *Nucleic Acids Research*, *33*(9), 1–8. https://doi.org/10.1093/nar/gni077

Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S., & Winter, G. (1986). Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature*, *321*(6069), 522–525. https://doi.org/10.1038/321522a0

Josephson, K., Ricardo, A., & Szostak, J. W. (2014). MRNA display: From basic principles to macrocycle drug discovery. *Drug Discovery Today*, *19*(4), 388–399. https://doi.org/10.1016/j.drudis.2013.10.011

Julian, M. C., Lee, C. C., Tiller, K. E., Rabia, L. A., Day, E. K., Schick, A. J., & Tessier, P. M. (2015). Co-evolution of affinity and stability of grafted amyloid-motif domain antibodies. *Protein Engineering, Design and Selection*, *28*(10), 339–350. https://doi.org/10.1093/protein/gzv050

Julian, M. C., Li, L., Garde, S., Wilen, R., & Tessier, P. M. (2017). Efficient affinity maturation of antibody variable domains requires co-selection of compensatory mutations to maintain thermodynamic stability. *Scientific Reports*, *7*(August 2016), 45259. https://doi.org/10.1038/srep45259

Jung, S., Arndt, K. M., Müller, K. M., & Plückthun, A. (1999). Selectively infective phage (SIP) technology: scope and limitations. *Journal of Immunological Methods*, *231*(1–2), 93–104. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10648930

Kabat, E. A., & Wu, T. (1971). Attempts To Locate Complementarity-Determining Residues in the Variable Positions of Light and Heavy Chains. *Annals of the New York Academy*

of Sciences, *190*(1), 382–393. https://doi.org/10.1111/j.1749-6632.1971.tb13550.x

Kepler, T. B., Liao, H. X., Alam, S. M., Bhaskarabhatla, R., Zhang, R., Yandava, C., … Haynes, B. F. (2014). Immunoglobulin gene insertions and deletions in the affinity maturation of HIV-1 broadly reactive neutralizing antibodies. *Cell Host and Microbe*, *16*(3), 304–313. https://doi.org/10.1016/j.chom.2014.08.006

Khan, T. A., Friedensohn, S., de Vries, A. R. G., Straszewski, J., Ruscheweyh, H.-J., & Reddy, S. T. (2016). Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting. *Science Advances*, *2*(3), e1501371–e1501371. https://doi.org/10.1126/sciadv.1501371

Kipnis, Y., Dellus-Gur, E., & Tawfik, D. S. (2012). TRINS: A method for gene modification by randomized tandem repeat insertions. *Protein Engineering, Design and Selection*, *25*(9), 437–444. https://doi.org/10.1093/protein/gzs023

Klein, F., Diskin, R., Scheid, J. F., Gaebler, C., Mouquet, H., Georgiev, I. S., … Bjorkman, P. J. (2013). Somatic Mutations of the Immunoglobulin Framework Are Generally Required for Broad and Potent HIV-1 Neutralization. *Cell*, *153*(1), 126–138. https://doi.org/10.1016/j.cell.2013.03.018

Klein, U., Goossens, T., Fischer, M., Kanzler, H., Braeuninger, A., Rajewsky, K., & Kuppers, R. (1998). Somatic hypermutation in normal and transformed human B cells. *Immunol Rev*, *162*, 261–280. https://doi.org/10.1111/j.1600-065X.1998.tb01447.x

Knappik, Ge, L., Honegger, a, Pack, P., Fischer, M., Wellnhofer, G., … Virnekäs, B. (2000). Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *Journal of Molecular Biology*, *296*(1), 57–86. https://doi.org/10.1006/jmbi.1999.3444

Koenig, P., Lee, C. V., Walters, B. T., Janakiraman, V., Stinson, J., Patapoff, T. W., & Fuh, G. (2017). Mutational landscape of antibody variable domains reveals a switch modulating the interdomain conformational dynamics and antigen binding. *Proceedings of the National Academy of Sciences*, *114*(4), E486–E495. https://doi.org/10.1073/pnas.1613231114

Köhler, G., & Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of

predefined specificity. *Nature*, *256*(5517), 495–497. https://doi.org/10.1038/256495a0

Krause, J. C., Ekiert, D. C., Tumpey, T. M., Smith, P. B., Wilson, I. A., & Crowe, J. E. (2011). An Insertion Mutation That Distorts Antibody Binding Site Architecture Enhances Function of a Human Antibody, *2*(1), 1–8. https://doi.org/10.1128/mBio.00345-10.Editor

Krebber, C., Spada, S., Desplancq, D., Krebber, C, Ge, L., & Pluckthun, a. (1997). Selectively-infective phage (SIP): a mechanistic dissection of a novel in vivo selection for protein-ligand interactions. *Journal of Molecular Biology*, *268*(3), 607–18. https://doi.org/10.1006/jmbi.1997.0981

Krebber, C., Spada, S., Desplancq, D., & Plückthun, a. (1995). Co-selection of cognate antibody-antigen pairs by selectively-infective phages. *FEBS Letters*, *377*(2), 227–31. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8543056

Krykbaev, R., Tsantili, P., Jeffrey, P. D., & Margolies, M. N. (2002). Modifying specificity of antidigoxin antibodies using insertional mutagenesis. *Protein Science : A Publication of the Protein Society*, *11*(12), 2899–908. https://doi.org/10.1110/ps.0223402

Kwong, P. D., & Mascola, J. R. (2012). Human Antibodies that Neutralize HIV-1: Identification, Structures, and B Cell Ontogenies. *Immunity*, *37*(3), 412–425. https://doi.org/10.1016/j.immuni.2012.08.012

Lamminmäki, U., Paupério, S., Westerlund-Karlsson, a, Karvinen, J., Virtanen, P. L., Lövgren, T., & Saviranta, P. (1999). Expanding the conformational diversity by random insertions to CDRH2 results in improved anti-estradiol antibodies. *Journal of Molecular Biology*, *291*(3), 589–602. https://doi.org/10.1006/jmbi.1999.2981

Lara-Ochoa, F., Almagro, J. C., Vargas-Madrazo, E., & Conrad, M. (1996). Antibody-antigen recognition: A canonical structure paradigm. *Journal of Molecular Evolution*, *43*(6), 678–684. https://doi.org/10.1007/BF02202116

Leushkin, E. V, Bazykin, G. A., & Kondrashov, A. S. (2012). Insertions and deletions trigger adaptive walks in Drosophila proteins. *Proceedings. Biological Sciences / The Royal Society*, *279*(1740), 3075–82. https://doi.org/10.1098/rspb.2011.2571

Li, Y., Yang, F., Smith-Gill, S. J., & Mariuzza, R. A. (2003). X-ray snapshots of the maturation of an antibody response to a protein antigen. *Nature Structural & Molecular Biology*, *10*(6), 482–488. https://doi.org/10.1038/nsb930

Liu, S., Wei, X., Dong, X., Xu, L., & Liu, J. (2015). Structural plasticity of green fluorescent protein to amino acid deletions and fluorescence rescue by folding-enhancing mutations. *BMC Biochemistry*, 1–11. https://doi.org/10.1186/s12858-015-0046-5

Lonberg, N. (2008). Fully human antibodies from transgenic mouse and phage display platforms. *Current Opinion in Immunology*, *20*(4), 450–459. https://doi.org/10.1016/j.coi.2008.06.004

Lubkowski, J., Hennecke, F., Plückthun, a, & Wlodawer, a. (1999). Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. *Structure (London, England : 1993)*, *7*(6), 711–22. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10404600

Luginbühl, B., Kanyo, Z., Jones, R. M., Fletterick, R. J., Prusiner, S. B., Cohen, F. E., … Plückthun, A. (2006). Directed Evolution of an Anti-prion Protein scFv Fragment to an Affinity of 1 pM and its Structural Interpretation. *Journal of Molecular Biology*, *363*(1), 75–97. https://doi.org/10.1016/j.jmb.2006.07.027

Mankowska, S. A., Gatti-Lafranconi, P., Chodorge, M., Sridharan, S., Minter, R. R., & Hollfelder, F. (2016). A Shorter Route to Antibody Binders via Quantitative in vitro Bead-Display Screening and Consensus Analysis. *Scientific Reports*, *6*(1), 36391. https://doi.org/10.1038/srep36391

Marks, J. D., Hoogenboom, H. R., Bonnert, T. P., McCafferty, J., Griffiths, A. D., & Winter, G. (1991). By-passing immunization. *Journal of Molecular Biology*, *222*(3), 581–597. https://doi.org/10.1016/0022-2836(91)90498-U

Mattheakis, L. C., Bhattt, R., & Dower, W. J. (1994). An in vitro polysome display system for identifying ligands from very large peptide libraries. *Biochemistry*, *91*(September), 9022–9026. https://doi.org/10.1073/pnas.91.19.9022

May, R. D., Monk, P. D., Cohen, E. S., Manuel, D., Dempsey, F., Davis, N. H. E., … Anderson, I. K. (2012). Preclinical development of CAT-354, an IL-13 neutralizing

antibody, for the treatment of severe uncontrolled asthma. *British Journal of Pharmacology*, *166*(1), 177–193. https://doi.org/10.1111/j.1476-5381.2011.01659.x

Maynard Smith, J. (1970). Natural selection and the concept of a protein space. *Nature*, *225*(5232), 563–564. https://doi.org/10.1038/225563a0

McCafferty, J., Griffiths, A. D., Winter, G., & Chiswell, D. J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. *Nature*. https://doi.org/10.1038/348552a0

McHeyzer-Williams, M., Okitsu, S., Wang, N., & McHeyzer-Williams, L. (2011). Molecular programming of B cell memory. *Nature Reviews Immunology*, *12*(1), 24–34. https://doi.org/10.1038/nri3128

Melis, J. P. M., Strumane, K., Ruuls, S. R., Beurskens, F. J., Schuurman, J., & Parren, P. W. H. I. (2015). Complement in therapy and disease. Regulating the complement system with antibody-based therapeutics. *Molecular Immunology*, *67*(2), 117–130. https://doi.org/10.1016/j.molimm.2015.01.028

Mikell, I., Sather, D. N., Kalams, S. A., Altfeld, M., Alter, G., & Stamatatos, L. (2011). Characteristics of the earliest cross-neutralizing antibody response to HIV-1. *PLoS Pathogens*, *7*(1). https://doi.org/10.1371/journal.ppat.1001251

Mouquet, H., Scharf, L., Euler, Z., Liu, Y., Eden, C., & Scheid, J. F. (2012). Complex-type N -glycan recognition by potent broadly neutralizing HIV antibodies, *109*(47). https://doi.org/10.1073/pnas.1217207109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1217207109

Murakami, H., Hohsaka, T., & Sisido, M. (2002). Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs. *Nature Biotechnology*, *20*(January), 76–81. https://doi.org/10.1038/nbt0102-76

Nelson, A. L. (2010). Hope and hype, (February), 77–83.

Neuberger, M. S., Harris, R. S., Di Noia, J., & Petersen-Mahrt, S. K. (2003). Immunity through DNA deamination. *Trends in Biochemical Sciences*, *28*(6), 305–312. https://doi.org/10.1016/S0968-0004(03)00111-7

Neylon, C. (2004). Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: Library construction methods for directed evolution. *Nucleic Acids Research*, *32*(4), 1448–1459. https://doi.org/10.1093/nar/gkh315

Niesen, F. H., Berglund, H., & Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*, *2*(9), 2212–2221. https://doi.org/10.1038/nprot.2007.321

North, B., Lehmann, A., & Dunbrack, R. L. (2011). A new clustering of antibody CDR loop conformations. *Journal of Molecular Biology*, *406*(2), 228–256. https://doi.org/10.1016/j.jmb.2010.10.030

O'connell, D., Becerril, B., Roy-Burman, A., Daws, M., & Marks, J. D. (2002). Phage versus phagemid libraries for generation of human monoclonal antibodies. *Journal of Molecular Biology*, *321*(1), 49–56. https://doi.org/10.1016/S0022-2836(02)00561-2

Odegrip, R., Coomber, D., Eldridge, B., Hederer, R., Kuhlman, P. A., Ullman, C., … McGregor, D. (2004). CIS display: In vitro selection of peptides from libraries of protein-DNA complexes. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(9), 2806–10. https://doi.org/10.1073/pnas.0400219101

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, *23*(1992), 263–286.

Oropallo, M. A., & Cerutti, A. (2014). Germinal center reaction: Antigen affinity and presentation explain it all. *Trends in Immunology*, *35*(7), 287–289. https://doi.org/10.1016/j.it.2014.06.001

Ortlund, E. A., Bridgham, J. T., Redinbo, M. R., & Thornton, J. W. (2007). Crystal Structure of an Ancient Protein: Evolution by Conformational Epistasis. *Science*, *317*(5844), 1544–1548. https://doi.org/10.1126/science.1142819

Padlan, E. A. (1994). Anatomy of the antibody molecule. *Molecular Immunology*, *31*(3), 169–217. https://doi.org/10.1016/0161-5890(94)90001-9

Parvari, R., Avivi, A., Lentner, F., Ziv, E., Tel-Or, S., Burstein, Y., & Schechter, I. (1988). Chicken immunoglobulin gamma-heavy chains: limited VH gene repertoire,

combinatorial diversification by D gene segments and evolution of the heavy chain locus. *The EMBO Journal*, *7*(3), 739–44. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=454383&tool=pmcentrez&rendertype=abstract

Pedrazzi, G., Schwesinger, F., Honegger, A., Krebber, C., & Plückthun, A. (1997). Affinity and folding properties both influence the selection of antibodies with the selectively infective phage (SIP) methodology. *FEBS Letters*, *415*(3), 289–293. https://doi.org/10.1016/S0014-5793(97)01143-5

Peng, H.P., Lee, K. H., Jian, J.-W., & Yang, A.-S. (2014). Origins of specificity and affinity in antibody-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(26), E2656-65. https://doi.org/10.1073/pnas.1401131111

Pham, P., Bransteitter, R., Petruska, J., & Goodman, M. F. (2003). Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature*, *424*(6944), 103–107. https://doi.org/10.1038/nature01760

Phillips, P. C. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, *9*(11), 855–867. https://doi.org/10.1038/nrg2452

Plückthun, A., Schaffitzel, C., Hanes, J., & Jermutus, L. (2000). In vitro selection and evolution of proteins. *Advances in Protein Chemistry*, *55*, 367–403. https://doi.org/10.1016/j.ymeth.2016.07.022

Popovic, B., Breed, J., Rees, D. G., Gardener, M. J., Vinall, L. M. K., Kemp, B., … May, R. D. (2016). Structural Characterisation Reveals Mechanism of IL-13 Neutralising Monoclonal Antibody Tralokinumab as Inhibition of Binding to IL-13Rα1 and IL-13Rα2. *Journal of Molecular Biology*, *429*(2), 208–219. https://doi.org/10.1016/j.jmb.2016.12.005

Poulsen, T. R., Jensen, A., Haurum, J. S., & Andersen, P. S. (2011). Limits for Antibody Affinity Maturation and Repertoire Diversification in Hypervaccinated Humans. *The Journal of Immunology*, *187*(8), 4229–4235. https://doi.org/10.4049/jimmunol.1000928

Poulsen, T. R., Meijer, P.-J., Jensen, A., Nielsen, L. S., & Andersen, P. S. (2007). Kinetic,

Affinity, and Diversity Limits of Human Polyclonal Antibody Responses against Tetanus Toxoid. *The Journal of Immunology*, *179*(6), 3841–3850. https://doi.org/10.4049/jimmunol.179.6.3841

Rathanaswami, P., Roalstad, S., Roskos, L., Su, Q. J., Lackie, S., & Babcook, J. (2005). Demonstration of an in vivo generated sub-picomolar affinity fully human monoclonal antibody to interleukin-8. *Biochemical and Biophysical Research Communications*, *334*(4), 1004–1013. https://doi.org/10.1016/j.bbrc.2005.07.002

Razai, A., Garcia-Rodriguez, C., Lou, J., Geren, I. N., Forsyth, C. M., Robles, Y., … Marks, J. D. (2005). Molecular evolution of antibody affinity for sensitive detection of botulinum neurotoxin type A. *Journal of Molecular Biology*, *351*(1), 158–169. https://doi.org/10.1016/j.jmb.2005.06.003

Reason, D. C., & Zhou, J. (2006). Codon insertion and deletion functions as a somatic diversification mechanism in human antibody repertoires. *Biology Direct*, *1*, 24. https://doi.org/10.1186/1745-6150-1-24

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., … Durandy, A. (2000). Activation-Induced Cytidine Deaminase (AID) Deficiency Causes the Autosomal Recessive Form of the Hyper-IgM Syndrome (HIGM2). *Cell*, *102*(5), 565–575. https://doi.org/10.1016/S0092-8674(00)00079-9

Riechmann, L., & Holliger, P. (1997). The C-terminal domain of TolA is the coreceptor for filamentous phage infection of E. coli. *Cell*, *90*(2), 351–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9244308

Rogozin, I. B., & Diaz, M. (2004). Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *Journal of Immunology (Baltimore, Md. : 1950)*, *172*(6), 3382–4. https://doi.org/10.4049/JIMMUNOL.172.6.3382

Romero, P. A., & Arnold, F. H. (2009). Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology*, *10*, 866. Retrieved from https://doi.org/10.1038/nrm2805

Romesberg, F. E. (1998). Immunological Origins of Binding and Catalysis in a Diels-Alderase Antibody. *Science*, *279*(5358), 1929–1933. https://doi.org/10.1126/science.279.5358.1929

Russel, M., & Model, P. (1989). Genetic analysis of the filamentous bacteriophage packaging signal and of the proteins that interact with it. *Journal of Virology*, *63*(8), 3284–3295. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2746731

Saphire, E. O. (2001). Crystal Structure of a Neutralizing Human IgG Against HIV-1: A Template for Vaccine Design. *Science*, *293*(5532), 1155–1159. https://doi.org/10.1126/science.1061692

Scharf, L., Scheid, J. F., Lee, J. H., West, A. P., Chen, C., Gao, H., … Bjorkman, P. J. (2014). Article Antibody 8ANC195 Reveals a Site of Broad Vulnerability on the HIV-1 Envelope Spike, 785–795. https://doi.org/10.1016/j.celrep.2014.04.001

Schatz, D. G., & Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. *Nature Reviews Immunology*, *11*(4), 251–263. https://doi.org/10.1038/nri2941

Scheid, J. F., Mouquet, H., Feldhahn, N., Seaman, M. S., Velinzon, K., Pietzsch, J., … Nussenzweig, M. C. (2009). Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature*, *458*(7238), 636–640. https://doi.org/10.1038/nature07930

Schier, R., McCall, A., Adams, G. P., Marshall, K. W., Merritt, H., Yim, M., … Marks, J. D. (1996). Isolation of Picomolar Affinity Anti-c-erbB-2 Single-chain Fv by Molecular Evolution of the Complementarity Determining Regions in the Center of the Antibody Binding Site. *Journal of Molecular Biology*, *263*(4), 551–567. https://doi.org/10.1006/jmbi.1996.0598

Schroeder, H. W. (2006). Similarity and divergence in the development and expression of the mouse and human antibody repertoires. *Developmental and Comparative Immunology*, *30*(1–2), 119–135. https://doi.org/10.1016/j.dci.2005.06.006

Schroeder, H. W. J., & Cavacini, L. (2010). Structure and Function of Immunoglobulins (author manuscript). *Journal of Allergy and Clinical Immunology*, *125*, S41–S52.

https://doi.org/10.1016/j.jaci.2009.09.046.Structure

Mar, N., & Herbert, I. (2014). Isolation of a Bacteriophage Specific for the F + and Hfr Mating Types of Escherichia coli K-12 Author (s): Tim Loeb Stable URL : http://www.jstor.org/stable/1705041 . in the native ceruloplasmins . Such a Types of Escherichia coli K-12, *131*(3404), 932–933.

Shirai, H., Kidera, A., & Nakamura, H. (1996). Structural classification of CDR-H3 in antibodies. *FEBS Letters*, *399*(1–2), 1–8. https://doi.org/10.1016/S0014-5793(96)01252-5

Simon, T., & Rajewsky, K. (1992). A functional antibody mutant with an insertion in the framework region 3 loop of the VHdomain: Implications for antibody engineering. *Protein Engineering, Design and Selection*, *5*(3), 229–234. https://doi.org/10.1093/protein/5.3.229

Skerra, A., & Pluckthun, A. (1988). Assembly of a functional immunoglobulin Fv fragment in Escherichia coli. *Science*, *240*(4855), 1038–1041. https://doi.org/10.1126/science.3285470

Smith, G. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, *228*(4705), 1315–1317. https://doi.org/10.1126/science.4001944

Spiess, C., Zhai, Q., & Carter, P. J. (2015). Alternative molecular formats and therapeutic applications for bispecific antibodies. *Molecular Immunology*, *67*(2), 95–106. https://doi.org/10.1016/j.molimm.2015.01.003

Stanfield, R. L. (2004). Crystal Structure of a Shark Single-Domain Antibody V Region in Complex with Lysozyme. *Science*, *305*(5691), 1770–1773. https://doi.org/10.1126/science.1101148

Stanfield, R. L., Wilson, I. A., & Smider, V. V. (2016). Conservation and diversity in the ultralong third heavy-chain complementarity-determining region of bovine antibodies. *Science Immunology*, *1*(1), 1–21. https://doi.org/10.1126/sciimmunol.aaf7962

Stemmer, W. P. (1994). DNA shuffling by random fragmentation and reassembly: in vitro

recombination for molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(22), 10747–10751. https://doi.org/10.1073/pnas.91.22.10747

Tas, J., Mesin, L., Pasqual, G., Targ, S., Jacobsen, J., Mano, Y., … Victora, G. (2016). Visualizing antibody affinity maturation in germinal centers. *Science*, *351*(6277), 1048–1054. https://doi.org/10.1126/science.aad3439

Tavakoli-Keshe, R., Phillips, J. J., Turner, R., & Bracewell, D. G. (2014). Understanding the relationship between biotherapeutic protein stability and solid-liquid interfacial shear in constant region mutants of IgG1 and IgG4. *Journal of Pharmaceutical Sciences*, *103*(2), 437–444. https://doi.org/10.1002/jps.23822

Thom, G., Cockroft, A. C., Buchanan, A. G., Candotti, C. J., Cohen, E. S., Lowne, D., … Minter, R. (2006). Probing a protein – protein interaction by in vitro evolution. *PNAS*, *103*(20), 7619–7624. Retrieved from http://www.pnas.org/content/103/20/7619.full

Tiller, T., Tsuiji, M., Yurasov, S., Velinzon, K., Nussenzweig, M. C., & Wardemann, H. (2007). Autoreactivity in Human IgG+ Memory B Cells. *Immunity*, *26*(2), 205–213. https://doi.org/10.1016/j.immuni.2007.01.009

Tokuriki, N., Stricher, F., Serrano, L., & Tawfik, D. S. (2008). How protein stability and new functions trade off. *PLoS Computational Biology*, *4*(2), 35–37. https://doi.org/10.1371/journal.pcbi.1000002

Tokuriki, N., & Tawfik, D. S. (2009). Stability effects of mutations and protein evolvability. *Current Opinion in Structural Biology*, *19*(5), 596–604. https://doi.org/10.1016/j.sbi.2009.08.003

Tomlinson, I. M., Walter, G., Jones, P. T., Dear, P. H., Sonnhammer, E. L., & Winter, G. (1996). The imprint of somatic hypermutation on the repertoire of human germline V genes. *Journal of Molecular Biology*, *256*(5), 813–817. https://doi.org/10.1006/jmbi.1996.0127

Tóth-Petróczy, Á., & Tawfik, D. S. (2013). Protein insertions and deletions enabled by neutral roaming in sequence space. *Molecular Biology and Evolution*, *30*(4), 761–771. https://doi.org/10.1093/molbev/mst003

Ulrich, H. D., Mundorff, E., Santarsiero, B. D., Driggers, E. M., Stevens, R. C., & Schultz, P. G. (1997). The interplay between binding energy and catalysis in the evolution of a catalytic antibody. *Nature*, *389*(6648), 271–275. https://doi.org/10.1038/38470

Vargas-Madrazo, E. F. (1995). Canonical Structure Repertoire of the Antigen-binding Site of Immunoglobulins Suggests Strong Geometrical Restrictions Associated to the Mechanism of Immune Recognition. *Journal of Molecular Biology*, *254*, 497–504. https://doi.org/10.1016/j.jmb.2010.08.009

Victora, G. D., & Nussenzweig, M. C. (2012). Germinal Centers. *Annual Review of Immunology*, *30*(1), 429–457. https://doi.org/10.1146/annurev-immunol-020711-075032

Vindahl, J., Nielsen, M., Berg, S., & Lund, O. (2013). Structural analysis of B-cell epitopes in antibody : protein complexes. *Molecular Immunology*, *53*(1–2), 24–34. https://doi.org/10.1016/j.molimm.2012.06.001

Viti, F., Nilsson, F., Demartis, S., Huber, a, & Neri, D. (2000). Design and use of phage display libraries for the selection of antibodies and enzymes. *Methods in Enzymology*, *326*(1991), 480–505. https://doi.org/http://dx.doi.org/10.1016/S0076-6879(00)26071-0

Wagner, A. (2011). The molecular origins of evolutionary innovations. *Trends in Genetics*, *27*(10), 397–410. https://doi.org/10.1016/j.tig.2011.06.002

Wang, F., Ekiert, D. C., Ahmad, I., Yu, W., Zhang, Y., Bazirgan, O., … Smider, V. V. (2013). XReshaping antibody diversity. *Cell*, *153*(6), 1379–1393. https://doi.org/10.1016/j.cell.2013.04.049

Wang, X., Minasov, G., & Shoichet, B. K. (2002). Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *Journal of Molecular Biology*, *320*(1), 85–95. https://doi.org/10.1016/S0022-2836(02)00400-X

Wedemayer, G. J. (1997). Structural Insights into the Evolution of an Antibody Combining Site. *Science*, *276*(5319), 1665–1669. https://doi.org/10.1126/science.276.5319.1665

Weisser, N. E., & Hall, J. C. (2009). Applications of single-chain variable fragment antibodies in therapeutics and diagnostics. *Biotechnology Advances*, *27*(4), 502–520. https://doi.org/10.1016/j.biotechadv.2009.04.004

Whitehead, T. A., Baker, D., & Fleishman, S. J. (2013). *Computational design of novel protein binders and experimental affinity maturation*. *Methods in Enzymology* (1st ed., Vol. 523). Elsevier Inc. https://doi.org/10.1016/B978-0-12-394292-0.00001-1

Winter, G., Griffiths, A. D., Hawkins, R. E., & Hoogenboom, H. R. (1994). Making antibodies by phage display technology. *Annual Review of Immunology*, *12*, 433–55. https://doi.org/10.1146/annurev.iy.12.040194.002245

Wright, S. (1930). Evolution in Medelian Populations, 1–10. https://doi.org/10.1007/BF02459575

Wu, L., Oficjalska, K., Lambert, M., Fennell, B. J., Darmanin-Sheehan, A., Ni Shuilleabhain, D., … Finlay, W. J. J. (2012). Fundamental Characteristics of the Immunoglobulin VH Repertoire of Chickens in Comparison with Those of Humans, Mice, and Camelids. *The Journal of Immunology*, *188*(1), 322–333. https://doi.org/10.4049/jimmunol.1102466

Wu, T., & Kabat, E. A. (1970). An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *The Journal of Experimental Medicine*, *132*(2), 211–50. https://doi.org/10.1084/jem.132.2.211

Wu, X., Yang, Z., Li, Y., Hogerkorp, C., William, R., Seaman, M. S., … Mascola, J. R. (2011). NIH Public Access, *329*(5993), 856–861. https://doi.org/10.1126/science.1187659.Rational

Xu, J. L., & Davis, M. M. (2000). Diversity in the CDR3 Region of VH Is Sufficient for Most Antibody Specificities. *Immunity*, *13*(1), 37–45. https://doi.org/10.1016/S1074-7613(00)00006-6

Yang, P. L., & Schultz, P. G. (1999). Mutational analysis of the affinity maturation of antibody 48G7. *J Mol Biol*, *294*(5), 1191–1201. https://doi.org/10.1006/jmbi.1999.3197

Yang, W. P., Green, K., Pinz-Sweeney, S., Briones, A. T., Burton, D. R., & Barbas, C. F. (1995). CDR walking mutagenesis for the affinity maturation of a potent human anti-HIV-1 antibody into the picomolar range. *Journal of Molecular Biology*, *254*(3), 392–403. https://doi.org/S0022283685706262 [pii]

Yeap, L. S., Hwang, J. K., Du, Z., Meyers, R. M., Meng, F. L., Jakubauskaite, A., … Alt, F. W. (2015). Sequence-Intrinsic Mechanisms that Target AID Mutational Outcomes on Antibody Genes. *Cell*, *163*(5), 1124–1137. https://doi.org/10.1016/j.cell.2015.10.042

Yélamos, J., Klix, N., Goyenechea, B., Lozano, F., Chui, Y. L., González Fernández, A., … Milstein, C. (1995). Targeting of non-Ig sequences in place of the V segment by somatic hypermutation. *Nature*. https://doi.org/10.1038/376225a0

Yoshikawa, K. (2002). AID Enzyme-Induced Hypermutation in an Actively Transcribed Gene in Fibroblasts. *Science*, *296*(5575), 2033–2036. https://doi.org/10.1126/science.1071556

Young, N. M., Watson, D. C., Cunningham, A. M., MacKenzie, C. R., & Huston, J. (2014). The intrinsic cysteine and histidine residues of the anti-Salmonella antibody Se155-4: A model for the introduction of new functions into antibody-binding sites. *Protein Engineering, Design and Selection*, *27*(10), 383–390. https://doi.org/10.1093/protein/gzu018

Zahnd, C., Sarkar, C. A., & Plückthun, A. (2010). Computational analysis of off-rate selection experiments to optimize affinity maturation by directed evolution. *Protein Engineering, Design and Selection*, *23*(4), 175–184. https://doi.org/10.1093/protein/gzp087

Zhou, T., Georgiev, I., Wu, X., Yang, Z.-Y., Dai, K., Finzi, A., … Kwong, P. D. (2010). Structural Basis for Broad and Potent Neutralization of HIV-1 by Antibody VRC01. *Science*, *329*(5993), 811–817. https://doi.org/10.1126/science.1192819