Universität für Bodenkultur Wien
Department für Biotechnologie

Engineering of critical antibody sequences for humanization, deimmunization and germinalization with detailed analysis of their influence on production cells

Dissertation

zur Erlangung des Doktorgrades
an der Universität für Bodenkultur Wien

Eingereicht von

Dipl.-Ing. Patrick Mayrhofer

Wien, Mai 2016

Departmentleiter:

Ao.Univ.Prof. Dipl.-Ing. Dr.nat.techn. Florian Rüker

Betreuer:

Univ.Prof. Dipl.-Ing. Dr.nat.techn. Renate Kunert

Univ.Prof. Dr. Chris Oostenbrink

“Somewhere, something incredible is waiting to be known.”

Carl Sagan

ACKNOWLEDGEMENTS

I would like to deeply thank Renate Kunert for giving me the opportunity to perform this PhD thesis at the Vienna Institute of Biotechnology in a very inspiring environment including many interesting, open and helpful discussions.

Furthermore, I would also like to thank Chris Oostenbrink and his working group at the Institute of Molecular Modeling and Simulation (MMS) for showing me the great potential of molecular dynamics simulations. In this context I would like to especially thank Christian Margreitter and Maria Pechlarner.

I further acknowledge the Austrian Science Fond (FWF) for funding and I really enjoyed being part of the BioToP PhD program in which I found many new colleagues and friends with a diversity of personal and scientific backgrounds. I would like to thank Christian Obinger, Christa Jakopitsch and all other current and former faculty staff and colleagues for providing great advice and input during this program.

I really appreciated getting the opportunity to work in different research labs abroad including the Instituto de Biologia Experimental e Tecnológica (iBET), Portugal, and the La Jolla Institute for Allergy and Immunology, USA. I would like to thank Paula Alves and Alessandro Sette for giving me the resources to learn new concepts in the field of metabolomics and immunology. For scientific and technical assistance I also want to thank Alexey Koshkin, Ana Teixeira, Carla Oseroff, Erin Moore, John Pham and Mikaela Lindvall.

I also would like to thank Jürgen Bode (Hannover Medical School) for having great discussions in the field of targeted gene integration using recombinase-mediated cassette exchange.

I always enjoyed working with my current and former colleagues Alexander Mader, Anna Erian, Clemens Tuppinger, David Reinhart, Julia Hennicke, Linda Schwaigerlehner, Lukas Damjanovic, Philipp Mundsperger, Stefan Bauernfried, Veronika Chromikova, Willibald Steinfellner and Wolfgang Sommeregger. Thank you for your support and fun I have with you during work, lunch and social activities.

I want to thank my family and friends who always provided strong support and fun to build the fundamental basis for me to achieve my set goals. Thank you Andrea, Carina, Clemens, Dominik, Gregor, Maria, Michael, Simon, Theresa and Thomas.

In this regard, I would like to say thank you to my grandparents, Erna and Heinrich, as well as Maria and Herbert. Thank you for always supporting me and my parents in different aspects over many years.

Finally, I want to express my deepest gratitude to my parents, Andrea and Herbert.
In small steps, you showed me how the world might work and helped me to develop the curiosity necessary to further explore the unknown.

Thank you for your endless love and support!

ABSTRACT

Antibody engineering allows the rational optimization of beneficial protein properties defined by the primary amino acid sequence. Non-human monoclonal antibodies (mAb) require modification in the variable region to remove immunogenic sequences, but simultaneously need to retain specificity and affinity. Our model protein, the mouse anti-idiotypic mAb Ab2/3H6, was used for humanization strategies by introducing rationally designed framework mutations. Novel 3H6 variants were expressed and evaluated for binding affinity. The gained information was supplied to develop an upgraded humanization workflow using a prospective design cycle supported by molecular dynamics (MD) simulations. Remaining non-human residues were additionally investigated in detail using T cell epitope prediction tools. Expression of some 3H6 variants in cell culture yielded only moderate production levels and in one case the mAb could not be secreted into the cell culture supernatant. This emphasized the common assumption that the primary amino acid sequence influences not only protein properties but also the cell biology and expression machinery. To investigate these impacts we engineered two different mammalian host cell lines (CHO) for targeted gene integration using recombinase-mediated cassette exchange (RMCE). Expression in RMCE-competent cells enabled us to test different mAb variants under isogenic conditions, eliminating positioning effects resulting from disparate integration loci. We designed and expressed a set of naturally occurring antibodies and germline-derived antibodies by comparing mature antibody sequences to their closest human germline genes. Isogenic RMCE subclones were subjected to proteomics and metabolomics analyses to study the influence of the protein sequence on cellular properties. The presented results demonstrate how different in-silico and in-vitro methods from different scientific fields can be merged to develop novel concepts for antibody engineering.

Keywords: Chinese Hamster Ovary, CHO-K1, DSC, DUKX-B11, HEK293, HIV-1, IgG1, scFv-Fc

ZUSAMMENFASSUNG

Antikörper-Engineering ermöglicht eine gezielte Optimierung vorteilhafter Proteineigenschaften, die durch die primäre Aminosäuresequenz bestimmt werden. Nicht-humane monoklonale Antikörper (mAk) müssen in ihrer variablen Region modifiziert werden, um immunogene Sequenzen zu entfernen, gleichzeitig aber Spezifität und Affinität konservieren. In unseren Humanisierungsstrategien wurde der anti-idiotypische Maus-mAk Ab2/3H6 verwendet und rational entworfene Frameworkmutationen eingefügt. Dann wurden die neuen 3H6 Varianten exprimiert und auf ihre Bindungseigenschaften getestet. Die dabei erhaltenen Informationen wurden in den Prozess der Humanisierung implementiert, um einen prospektiven Designzyklus von in-vitro und in-silico Methoden (molekulardynamischer Simulationen) zu entwickeln. Nicht-humane Positionen wurden zusätzlich in-silico auf T-Zell Epitope untersucht. Bei den Versuchen im Labor konnten bei einigen 3H6 Varianten nur moderate Expressionshöhen erreicht werden und in einem Fall wurde der mAk nicht in den Zellkulturüberstand segregiert. Das bestätigt die Annahme, dass die primäre Aminosäuresequenz nicht nur die Produkteigenschaften beeinflusst, sondern auch die Zellbiologie und Expressionsmaschinerie. Um diese Kausalität zu untersuchen, entwickelten wir zwei verschiedene Säugetier-Wirtszelllinien (CHO) zur zielgerichteten Genintegration unter Verwendung von Rekombinase-mediiertem Kassettenaustausch (RMCE). Verschiedene mAks wurden in RMCE-kompetenten Zellen unter isogenen Bedingungen exprimiert, sodass Positionseffekte durch unterschiedliche Integrationsloki eliminiert werden. Diese RMCE Klone wurden für Proteomik- und Metabolomik-Analysen herangezogen und der Einfluss der Proteinsequenz auf zelluläre Eigenschaften untersucht. Die präsentierten Ergebnisse zeigen wie geeignete in-silico und in-vitro Methoden aus verschiedenen wissenschaftlichen Bereichen verzahnt werden können, um neue Konzepte zur Antikörperentwicklung voranzutreiben.

ABBREVIATIONS

Ab2/3H6	Equal to "3H6", wild-type anti-idiotypic antibody 3H6
BM	Backmutated su3H6 variants
CD	Cluster of differentiation
CDR	Antibody complementarity determining region
CDR H	Heavy chain complementarity determining region
CHO	Chinese hamster ovary
D	Antibody diversity gene region
dhfr	Dihydrofolate reductase
ELISpot	Enzyme-Linked ImmunoSpot
EMA	European Medicines Agency
ER	Endoplasmic reticulum
FDA	US Food- and Drug Administration
FR	Antibody framework region
FR H	Heavy chain framework region
GA3H6	Aggressive graft version of 3H6
GC3H6	Conservative graft version of 3H6
HACA	Human anti-chimeric antibody
HAMA	Human anti-mouse antibody response
HEK	Human embryonic kidney
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
I3	Equal to „CHO-K1 RMCE I3“, a CHO-K1 RMCE cell line
IBET	Instituto de Biologia Experimental e Tecnológica
IEDB	Immune Epitope Database
IgG	Immunoglobulin G
IGHV	Immunoglobulin heavy chain variable gene region
IL	Interleukin
IMGT	Immunogenetics Information System
INF	Interferon
INST	Isotopically non-stationary
J	Antibody joining gene region
JH	Immunoglobulin heavy chain joining gene region
LIAI	La Jolla Institute for Allergy and Immunology
mAb	Monoclonal antibody
MD	Molecular dynamics
MFA	Metabolic flux analysis
MHCII	Major histocompatibility complex class two
PCR	Polymerase chain reaction
PhD	Doctor of Philosophy
RB	Russell bodies
rEF1a	Rat elongation factor 1 alpha promoter
RMCE	Recombinase-mediated cassette exchange
scFv-Fc	Single-chain fragment variable – fragment crystallizable
sec.	Seconds
su3H6	Superhumanized 3H6 antibody
TCR	T-cell receptor
Th	T-helper cell
TR	Antibody training set
V	Antibody variable gene region
vH	Antibody variable heavy chain region
vL	Antibody variable light chain region
wt3H6	Wild-type Ab2/3H6 antibody

OUTLINE

1 INTRODUCTION

Monoclonal antibodies (mAbs) are considered as the fastest growing class of biotherapeutic proteins. Reported US sales for this class of recombinant proteins reached $24.6 billion in 2012 (Aggarwal, 2014). Different criteria of a monoclonal antibody, intended for human application or research, have to be met for therapeutic and economic success.

Binding affinity, specificity, effector functions and immunological side reactions derived from non-human epitopes are the most pronounced features defined by the primary amino acid sequence of mAbs, which can be readily modulated by introducing mutations in various regions of an existing mAb. Such a "template mAb" might be screened from naïve and primed libraries or pools of (immortalized) lymphocytes, which can originate from humans but also other animal species. Although the traditional criterion for antibody clone selection is the biological function, more and more techniques also try to implement secondary antibody features, such as expressability and stability, in the selection process. However, the identification of required mutations for beneficial antibody properties is neither a straightforward nor a generalizable task. Structural frameworks define the evolutionary conserved immunoglobulin fold of antibodies. Localized structural features, such as protein loops or smaller epitopes, are responsible for specific functions and can be rationally or empirically modified to change the physicochemical properties by mutating the respective amino acid positions. Individual amino acids are often preferable for supporting one certain antibody feature (e.g. antibody affinity) but, while doing this exclusively, other features (e.g. stability) might be impeded (Wang et al., 2013). For therapeutic use highest safety standards of the drug compound are highly claimed by the FDA and EMA already in the pre-clinical stage (Lynch et al., 2009). These requirements have to be fulfilled for the biological functions of the drug substance, defined by the protein itself and for the purity of the final formulation, defined by the production process. For economic feasibility and to successfully outcompete other lead candidates a high expression level of the engineered antibody is pursued in large-scale mammalian cell culture technology. Standard and high-end techniques from different disciplines are necessary to assemble the iteration loop of research and development starting with the definition of the primary protein sequence of new mAbs leading over to and executing the clone- and process development. Additionally, a wide array of in-silico methods supports this development already today but more and more in the future will foster the biopharmaceutical technology.

2 AIMS OF THE THESIS

The program for this PhD thesis was based on two topics:

1. Converging in-silico and wet lab methods to develop new strategies for the definition of new mAbs
2. Can product-related and cell-physiological parameters be used to predict and improve the technological production process?

This work contributes to extend our knowledge on proper antibody engineering concepts and to define common criteria of next-generation antibody molecules for optimal biological functions combined with beneficial protein properties for maximum expression under optimal cellular background. Understanding the relation between the individual protein sequence and therapeutic- as well as technological functions guides the search for protein sequence-related expression bottlenecks and the development of novel assays applied for early screening. This leads to novel mAb candidates showing enhanced production levels.

Defined sequence features can be identified for engineering to mediate optimal biological functions but hamper their production process. The fundamental and persisting question is how sequence-related features, mediating biological functions, also govern expressability and interference with cellular physiology on a proteomic or metabolomic level. For example, a pronounced phenotype of cell clones producing difficult-to-express human IgG molecules was described to show distinct antibody aggregates in the endoplasmic reticulum (ER), called “russell bodies” (RB) (Stoops et al., 2012), or intracellular ER antibody crystals (Hasegawa et al., 2011). The authors could trace features leading to this cellular phenotype back to individual antibody chains and currently strong efforts are put into developing novel assays to define the “manufacturability” (Yang et al., 2013) and “developability” (Jarasch et al., 2015; Seeliger et al., 2015) of antibodies at a very early developmental stage. It would be highly advantageous to define optimal therapeutic features of antibody variable regions determined by the primary amino acid sequence, which additionally support large-scale production. Such a general strategy would optimize any monoclonal antibody’s “fitness for large-scale manufacturing” (Stoops et al., 2012).

In this thesis, we developed different substantial demands to rationally enhance antibody protein attributes and demonstrate the power and capabilities of combined in-vitro and in-silico methods in two complementing projects. Antibody humanization aims to reduce the expected immunogenicity of protein sequences of non-human origin, but the humanized antibody variants often exhibit reduced or completely lost binding affinities. In chapter two of this thesis, the experimental work towards a novel prospective humanization design cycle is presented, using molecular dynamics (MD) simulations as a new tool to guide selection of critical backmutations. The computational work was done by Christian Margreitter, a colleague at the Institute of Molecular Modeling and Simulation (MMS) in Chris Oostenbrink’s group. The results were published in a first author publication (Margreitter et al., 2016). Residual immunogenicity might be further reduced by prediction and

deimmunization approaches to remove MHCII restricted epitopes. In-silico prediction tools, provided by the Immune Epitope Database (IEDB), were applied to our Ab2/3H6 model antibody system (Kunert et al., 2002) during a research stay at the La Jolla Institute for Allergy and Immunology (LIAI, USA).

In chapter three, the basic essentials for accurately investigating the influence of antibody germline identity on expression levels were established. Construction of a gene-targeting CHO DUKX-B11 host enabling recombinase-mediated cassette exchange (RMCE) was published in a first author publication (Mayrhofer et al., 2014) and antibody producing subclones were developed to highlight proteomic differences provoked by two similar anti-HIV-1 antibody sequences (3D6 and 2F5). The results were published under co-authorship (Sommeregger et al., 2016). Next, and to improve cellular performance, an engineered CHO-K1 (CHO-K1 RMCE I3) host cell line was established. In this project a set of anti-HIV-1 antibodies (2G12, 2F5, 4B3) and one therapeutic antibody (Ustekinumab) were compared to the closest human antibody germline genes. Afterwards we assembled fully germline antibody variants to match the expression potential, cellular behavior of recombinant clones and proteochemical properties of purified mAbs. Stable cell lines of this antibody set in IgG1 format were established in the novel CHO-K1 RMCE I3 host. Metabolic changes of different fed-batch culture time points were investigated during a research stay at Instituto de Biologia Experimental e Tecnológica (IBET, Portugal) using isotopically non-stationary metabolic flux analysis (INST-MFA).

Additionally, a protocol for conversion of these IgG1 molecules into single-chain fragment variable – fragment crystallizable (scFv-Fc) was submitted for publication in *Methods in Molecular Biology* under first authorship (Mayrhofer and Kunert, under review 2016).

3 HUMANIZATION – REDUCTION OF EXPECTED IMMUNOGENICITY BY SIMULTANEOUSLY MAINTAINING PRIMARY FUNCTION OF ANTIBODY AFFINITY

3.1 Background

Hybridoma technology, developed by Georges Köhler and César Milstein, enables the fast and robust development of monoclonal antibodies directed against the antigen of interest (Köhler and Milstein, 1975). Decades before fully human antibodies could be developed by different surface display systems (Winter et al., 1994) or transgenic mice (Brüggemann et al., 1991), it was possible to immunize animal species such as mice or rats with a protein containing the linear or conformational epitope of interest. After assessment of a proper immune response, antigen-specific B-lymphocytes are harvested from the isolated spleen and fused to immortal myeloma cells by somatic cell fusion. The hybridoma technology still represents a robust and fast method for developing monoclonal antibodies for basic research using well-established protocols (Zhang, 2012). Many different hybridoma cell lines were and are still being developed to significantly advance the understanding about antibodies in general and the antibody-target interaction in particular. A considerable number of such hybridoma-derived antibodies are worth to be tested in clinical trials and to translate into human application. In the course of evolution the human immune system was fine-tuned to recognize protein sequences and structures derived from non-human organisms, such as mice. This feature is often marked by a response called human anti-mouse antibody response (HAMA) in which a high titer level of antibodies directed against the therapeutic antibody is developed (Schroff et al., 1985). This reaction is also accompanied by symptoms such as the cytokine release syndrome, also known as “cytokine storm” with different degrees of severity (Bugelski et al., 2009). In the past notorious examples for adverse side effects were reported for ATG, OKT3, rituximab, as well as the infamous TGN 1412 incident (Attarwala, 2010). High titers of anti-drug antibodies in the human serum reduce efficacy and alter pharmacokinetic properties leading to faster clearance of the therapeutic protein (Holgate and Baker, 2009). The prevention of these adverse side effects is crucial for the safety of the patient and the success of a biotherapeutic candidate. It is of utmost importance to assess expected immunogenicity in the earliest developmental stages of an antibody drug candidate and to eliminate potential adverse effects. Different methods were developed to

assess and reduce the expected immunogenicity and rigorous pre-clinical risk assessment is required by the FDA.

Multiple factors influence the expected immunogenicity of a therapeutic protein in general. These factors cover aspects of all levels of development of a biopharmaceutical process including the design of the specific antibody primary sequence defined in the humanization procedures, culture- and fermentation (upstream) processes or downstream processing and formulation. Therefore, a holistic approach considering all factors is inevitable to prevent immunogenicity of the antibody therapeutic. A comprehensive overview is given by Jawa et al. (2013) in **Table I**. The primary antibody sequence is one major determinant of immunogenicity that might be readily modified by proper antibody engineering practices. Besides process- and formulation-related microheterogeneity, antibody charge variants, glycosylation pattern and aggregates (Joubert et al., 2012; Ratanji et al., 2014; Rombach-Riegraf et al., 2014) also contribute to safety issues and the patient's convenience. Finally, patient-specific factors such as pre-existing antibodies or the immune status and the applied treatment regimen determines the expected anti-drug immunogenicity.

Table I. Overview of potential factors determining protein immunogenicity.
Reprinted from Jawa et al. (2013).

FACTORS	IMPLICATIONS
Intrinsic immunogenic dominant epitopes	Higher risk of immunogenicity based on reactivity to HLA (DR, DP, DQ) alleles
Sequence differences between therapeutic protein and endogenous protein	Cross-reactive neutralizing antibodies to endogenous proteins
Differences in glycosylation patterns. Structural alterations induced by storage condition, production/purification and formulation	Breaking of tolerance due to - Aggregation - Oxidation - Deamidation and degradation - Conformational changes
Route of delivery - Route/frequency of administration - Syringe environment and associated reconstitution - Non-physiologic concentrations	- Differences in antigen presentation following intravenous and subcutaneous delivery - Presence of leachates from syringe (i.e. tungsten or silicone) can facilitate immune responses - Exposure to high dose proteins in formulations can lead to an immune response
Patient-related factors • Immune status - Immunosuppression due to supportive therapies - Autoimmune-diseased state - Genetic features like HLA haplotypes • Intensity of treatment at the first dose • Age	- Immune status can enhance or reduce the likelihood of an immune response - Presence of HLA alleles that have high potential to bind to immune dominant epitopes
Host Cell Proteins (HCP) Recombinant protein therapeutics produced in mammalian cell lines may contain process-related contaminants.	The homology of HCP with human proteins raises concern over the inherent risk of anti-self immune responses.

First attempts to convert non-human antibody sequences into human-like derivatives were developed by the antibody chimerization approach (Morrison et al., 1984). In this basic technique non-human antibody variable regions, comprising complementarity determining regions (CDRs) and framework (FR) sequences, are combined with human constant regions to form animal/human chimeric proteins. However, the human body still recognizes non-human sequences present in the variable regions triggering a human anti-chimeric antibody (HACA) response (Afif et al., 2010). Therefore, more sophisticated humanization techniques were developed. The basic concept is to maintain the primary function of antibodies, i.e. their binding properties that are primarily mediated by their CDRs by equipping them with a human-like framework. The chosen framework must support the correct conformations and dynamics of the CDR-loops required for antigen binding. As a first step to decrease the non-human portions in chimeric antibodies, the CDR-grafting technology was developed by transferring the non-human CDR regions onto properly selected human framework regions (Jones et al., 1986).

One major challenge of CDR-grafting protocols is the conservation of the mAb binding affinity, which might be reduced or even completely lost. Therefore, introduction of murine backmutations into the human framework is required to restore critical FR-CDR interactions leading to a high binding affinity. This process is often elusive and time-consuming since no generally applicable rules exist for selection of proper backmutations. Multiple rounds of introducing mutations within the antibody variable sequence, expression, purification and evaluation of binding affinities are required in the classical design cycle (**Figure 1**). Multiple backmutations may be introduced simultaneously to enhance the chance of getting a high-affinity binding variant. Further cycle rounds are then required to consecutively reduce the number of backmutations to an absolute minimum in a stepwise fashion, to ultimately reduce immunogenicity but maintain affinity. This empirical process is complicated by the fact that each antibody relies on different FR-CDR interaction points for stabilization and no universal procedure is available to identify these requirements a priori.

Information about currently approved therapeutic antibodies may be retrieved from the International Immunogenetics Information System (IMGT) using the IMGT/mAb-DB search tool (Lefranc et al., 2015). Currently, the majority of approved therapeutic antibodies (total 49) is represented by humanized variants (20) compared to antibodies derived from murine (5), human (17) or chimeric (7) origin (**Table II**). This highlights the contemporary importance of this technique to reduce immunogenicity.

Table II. A total of 49 FDA- or EMA approved therapeutic antibodies (+ conjugates) listed in IMGT/mAb on March 13, 2016. Dates indicate first approval. Antibody name substems indicating the origin are underlined.

human	murine	chimeric	humanized
Adalimumab (2002)	Muromonab CD3 (1992)	Abciximab (1993)	Palivizumab (1998)
Panitumumab (2006)	Edrecolomab (1995)	Rituximab (1997)	Trastuzumab (1998)
Canakinumab (2009)	Ibritumomab tiuxetan (2002)	Basiliximab (1998)	Alemtuzumab (2001)
Golimumab (2009)	Catumaxomab (2009)	Infliximab (1998)	Omalizumab (2003)
Ofatumumab (2009)	Blinatumomab (2014)	Cetuximab (2004)	Bevacizumab (2004)
Ustekinumab (2009)		Brentuximab vedotin (2011)	Natalizumab (2004)
Denosumab (2010)		Siltuximab (2014)	Ranibizumab (2006)
Belimumab (2011)			Eculizumab (2007)
Ipilimumab (2011)			Certolizumab pegol (2008)
Raxibacumab (2012)			Nimotuzumab (2008)
Nivolumab (2014)			Tocilizumab (2010)
Ramucirumab (2014)			Pertuzumab (2012)
Alirocumab (2015)			Obinutuzumab (2013)
Daratumumab (2015)			Trastuzumab emtansine (2013)
Evolocumab (2015)			Pembrolizumab (2014)
Necitumumab (2015)			Vedolizumab (2014)
Secukinumab (2015)			Elotuzumab (2015)
			Idarucizumab (2015)
			Mepolizumab (2015)
			Pembrolizumab (2015)
17	5	7	20

3.2 Towards a prospective design cycle using Molecular Dynamics simulations

Almagro and Fransson classified humanization approaches into empirical methods and rational methods (Almagro and Fransson, 2008). Empirical humanization processes rely on selection rather than rationality for chosen mutations to restore or even enhance the affinity. Laborious and expensive tools are required to generate large conformational antibody libraries and to select antibody lead candidates by enrichment technologies (phage-, ribosome- or yeast display) or high-throughput screening (HTS) methods. In contrast, rational methods rely on a design cycle starting by construction of few antibody variants based on selection of rational mutation sites followed by expression, purification and assessment of binding affinities. This highly laborious work can be supported by new computational tools to improve the rational design cycle by providing new knowledge and drive decision-making in the selection process for proper mutations.

MD simulations allow to simulate the movement of atoms of a molecular system over time in a spatial and temporal resolution that cannot be achieved by experimental means (van Gunsteren et al., 2006). By using statistical mechanics, including sophisticated force fields and algorithms for integrating Newton's law of motion, it is possible to generate thermodynamic ensembles of a molecular system that can be time-averaged to deduce macroscopic properties.

We hypothesize that MD simulations will provide crucial information about the dynamical behavior of antibody CDR-loops to evaluate the expected influence of backmutations on the actual binding affinity. Therefore, it should be possible to implement MD simulations as a new tool in the humanization repertoire towards a prospective design cycle for a priori assessment of required backmutations (**Figure 1**). Instead of tedious expression and binding experiments in the wet lab, multiple homology models of mutated antibody variants might be simultaneously subjected to MD simulations in-silico in a high-throughput fashion. Importantly, this procedure only focuses on the dynamics of the CDR regions supported by the underlying antibody framework. The simulation does not include any conformation of the antibody interaction partner (antigen epitope) which is often unknown for a new therapeutic target and it reduces the complexity of the computationally demanding simulation setup.

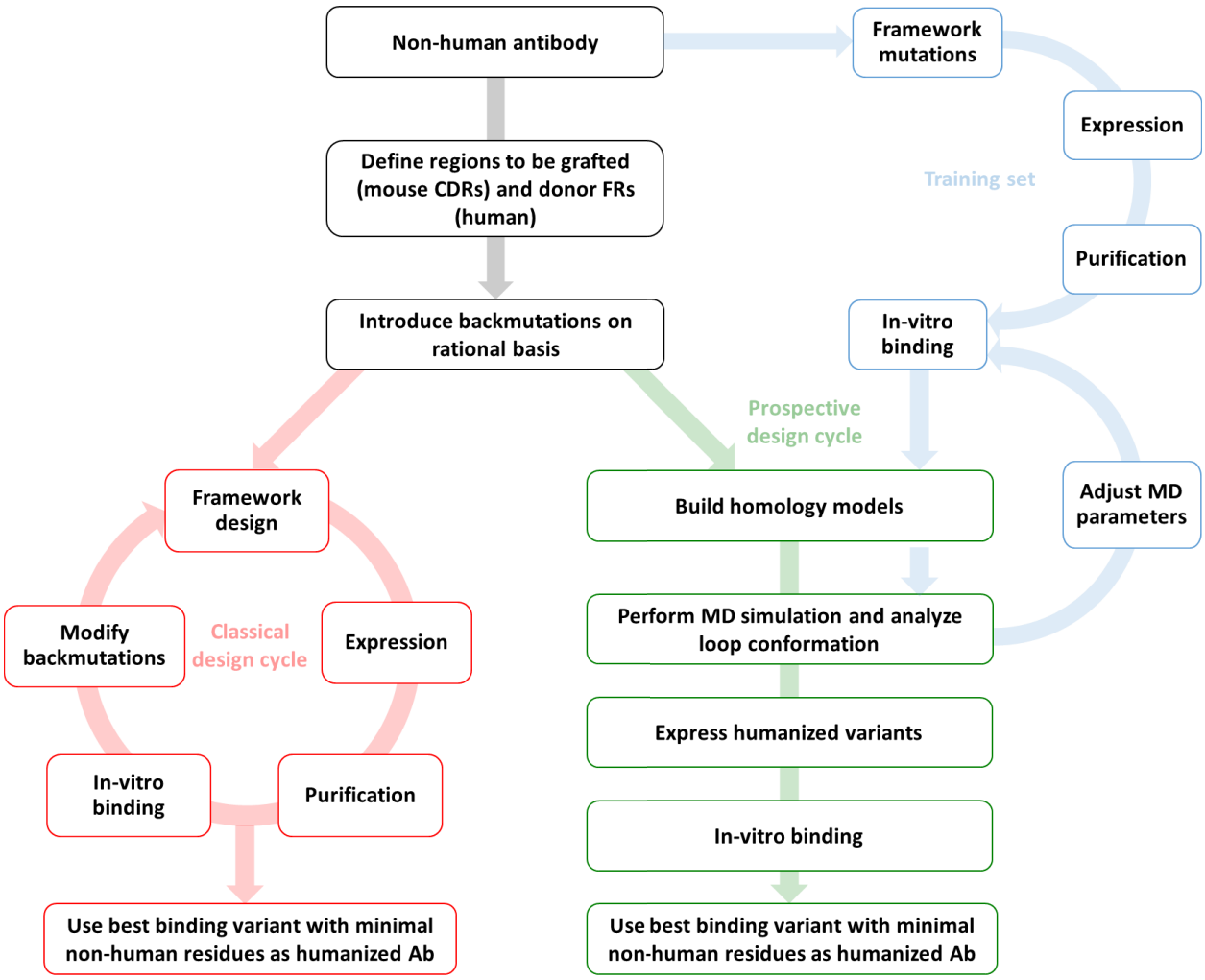
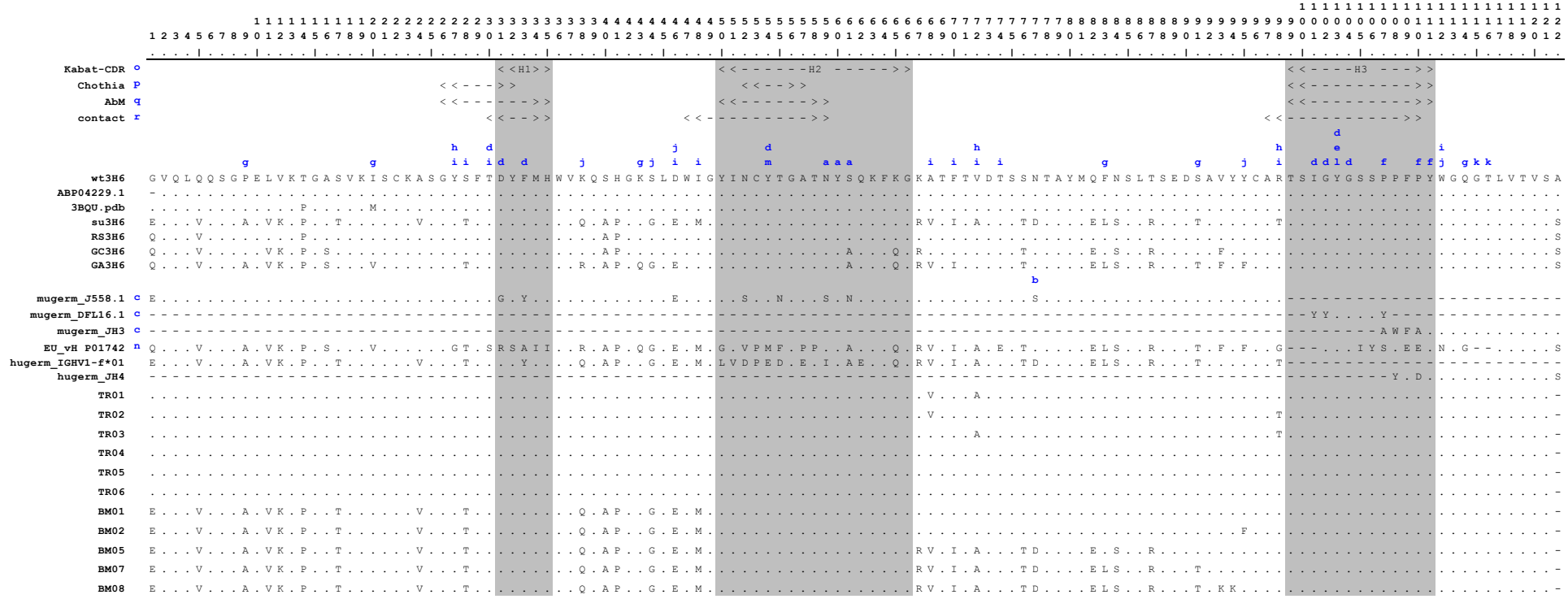


Figure 1. Classical and prospective design cycle for antibody humanization. Classical rational humanization strategies include iterative antibody expression cycles (red) in contrast to a prospective process supported by in-silico simulations (green). For adjustment of MD simulation parameters a training set (blue) was established.

In our lab the anti-idiotypic antibody Ab2/3H6 (Kunert et al., 2002) was developed and subjected to classical humanization techniques such as chimerization or more sophisticated approaches like CDR-grafting, resurfacing and superhumanization (Mader and Kunert, 2010). This antibody is directed against the idiotype of the broadly neutralizing anti-HIV-1 antibody 2F5 and already has served as a suitable model in many different projects. A broad range of knowledge was gathered and reported around this antibody model, including identification of critical amino acids required for binding and supplemented by in-vivo studies (Gach et al., 2007a, 2008a, 2008b). Furthermore, the structure of 3H6 in complex with 2F5 was determined by x-ray crystallography (Bryson et al., 2008) followed by molecular dynamics simulations as a first attempt to rationalize the binding behavior of different humanized variants (de Ruiter et al., 2011). During these studies a special interest was put into the heavy chain of 3H6 that supposedly promotes the majority of the binding to 2F5. Detailed positional sequence information for 3H6vH can be found in **Figure 2**.

Different strategies might be used to find critical (back)mutations for enhanced functions of antibodies. Within the genome, antibodies are generated by recombination of the variable (V), diversity (D) and joining (J) gene segment. The assembled antibody gene is then further matured by somatic mutations which primarily influence specificity, affinity, protein stability and expression levels. We can identify these somatic mutations by aligning the mature antibody sequence to all germline gene segments. For the murine antibody 3H6 (wt3H6) two somatic mutations were found in the vH framework regions (**Figure 2**) when compared to the murine germline J558.1 ("mugerm_J558.1" or "IGHV1S34*01"). Additionally, critical positions for backmutations might be found by using knowledge from literature. Certain positions in the antibody sequence were shown to form the Vernier zone underlying the CDR regions for supporting a correct conformation for high affinity binding (Foote and Winter, 1992). Other positions define the canonical structure class of the hypervariable structural loops (Chothia and Lesk, 1987). It is obvious, that these residues should be preserved during the humanization process. These concepts provide help for the selection process for suitable backmutations, yet are not universally applicable and often not sufficient to restore the full binding affinity in the humanized variant. Additionally, peripheral mutations should be preserved that might have an impact on the overall protein structure such as vH/vL interface residues (Vargas-Madrazo and Paz-García, 2003). Additional information from available experimental data might be useful for the selection process. One source might be alanine scanning experiments to identify critical positions required for correct folding and binding. For the antibody 3H6 also two CDR-grafted versions (GC3H6 and GA3H6) were developed by grafting the murine CDR regions onto the framework regions of the mature human antibody EU (Mader and Kunert, 2010). It was shown that the aggressive graft version (GA3H6) has lost some of its binding affinity compared to the conservative graft variant (GC3H6) which differ in 12 framework positions from each other.



a	N-glycosylation site (Asn-Xaa-Ser/Thr) no influence on binding verified by N59A and tunicamycin (Gach et al., 2007a, 2008a)
b	Position of murine germline sequence with highly dissimilar somatic mutation
c	Closest murine germline sequences of 3H6 (Gach et al., 2007b)
d	2F5 contact residues (Bryson et al., 2008)
e	Critical for binding in alanine scanning experiments (Gach et al., 2008b)
f	Maintained binding in alanine scanning experiments (Gach et al., 2008b)
g	Position considered as critical for having influence on protein structure (Mader and Kunert, 2010)
h	Residues in close proximity or contributing to canonical structure class (Mader and Kunert, 2010)
i	Residues in close contact or contributing to vernier zone (Mader and Kunert, 2010)
j	Residues buried at the vH/vL interface (Mader and Kunert, 2010)
k	Residue missing in EU antibody used for GC3H6 and GA3H6 (Mader and Kunert, 2010)
l	Less flexible in wt3H6 and RS3H6, but more flexible in su3H6, GA3H6 and GC3H6 (de Ruiter et al., 2011)
m	Available for 2F5-hydrogen bonding in wt3H6 and RS3H6, more restricted in su3H6, N52A mutation suggested for su3H6 (de Ruiter et al., 2011)
n	Closest human mature antibody used for GC3H6 and GA3H6 design (Mader and Kunert, 2010)
o	CDR defined by sequence variability (Wu and Kabat, 1970). Integrated into the Abysis web interface (http://www.bioinf.org.uk/abysis/)
p	Structural loops defined by Chothia and Lesk (1987). Integrated into the Abysis web interface (http://www.bioinf.org.uk/abysis/)
q	AbM loops defined by (Martin et al., 1989) as a compromise between Kabat sequence variability and Chothia structural loop (Dübel and Reichert, 2014). Integrated into the Abysis web interface (http://www.bioinf.org.uk/abysis/)
r	Contact regions defined by MacCallum et al. (1996). Integrated into the Abysis web interface (http://www.bioinf.org.uk/abysis/)

Figure 2 Important vH sequence information collected for the model antibody 3H6 and related variants.

These 12 positions also constitute possible candidates for required backmutations, although in the superhumanized variant (su3H6) a different framework was chosen (human germline IGHV1-f*01 and JH4) than for the CDR-grafted variants (mature EU antibody).

For development of the MD simulation-guided humanization approach a first step is to develop the simulation setup and execute the simulation run and analysis. Therefore we generated a training set ("TR") consisting of wild-type 3H6 (wt3H6), a superhumanized 3H6 variant (su3H6) that has lost its binding affinity completely and six double-mutants TR01-TR06 in which critical framework 3 positions of the heavy chain (FR H3) (TR01-TR03), or framework 1-3 of the light chain (TR04-06) were mutated based on the superhumanized counterpart (**Figure 2**).

The training set was subjected to the MD simulation and experimentally expressed in stable CHO pools followed by binding evaluation using biolayer-interferometry. Simulated CDR dynamics could be matched to the experimental binding affinities by using a similarity score calculated from cluster analysis which describes how closely the CDR dynamics of mutant sequence variants resemble the wild-type behavior. This procedure allowed us to formulate a cutoff value of 2 for the similarity score that was considered as threshold to qualitatively distinguish binders from non-binders in the prospective design cycle (**Figure 3**).

The Abysis tool (http://www.bioinf.org.uk/abysis/) allows searching for unusual residues within the antibody sequence, defined by a 1% threshold in the stored antibody sequences giving hints for critical functions of certain positions, therefore supporting optimization of new synthetic antibody variants. Using this knowledge we defined the 3H6 mutants representing the training set, comprising variants with drastically reduced binding affinities and wild-type binding affinities **Figure 3**.

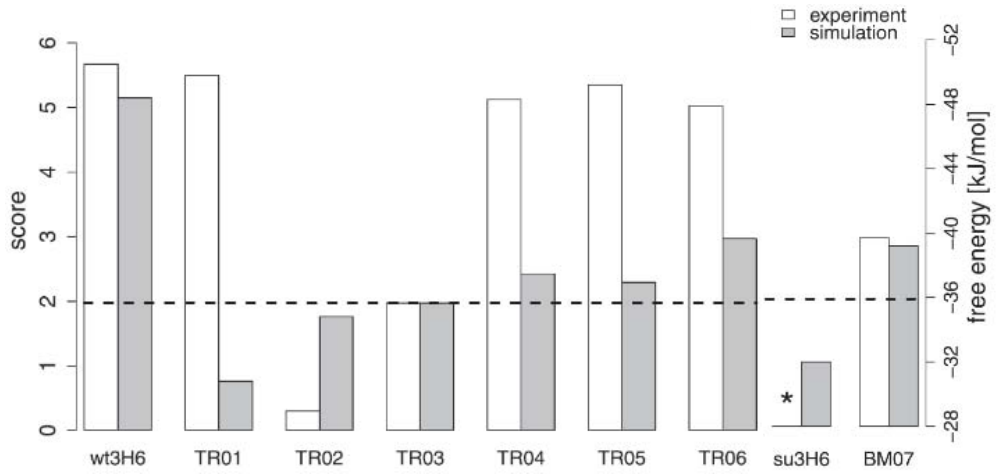


Figure 3. Experimental binding free energies and similarity scores from MD simulations of selected 3H6 variants. Threshold used to discriminate binders from non-binders is indicated as dashed line. * su3H6 showed no binding. Binding free energies were calculated using **Equation 1**. Modified from Margreitter et al. (2016).

Binding affinities determined from k_{off} and k_{on} measurements by biolayer-interferometry assays were reported as binding free energies calculated from **Equation 1**. A change of binding free energies ($\Delta\Delta G_{binding}$) by 5 kJ/mol approximately calculates to one order of magnitude change for the dissociation constant (K_D).

Equation (1) $\Delta G_{binding} = RT \ln(K_D)$

$\Delta G_{binding}$	(Gibbs) binding free energy [kJ/mol]
R	Gas constant [8.3144×10^{-3} kJ/mol/K]
T	abolute temperature [K]
K_D	dissociation constant [M]

The loss in binding affinities of TR02 and TR03 can be attributed to a single R98T^{HC} mutation. Threonine at this very position is only rarely found in human germline V-regions or in antibody sequences compiled in the Abysis database **Table III**.

Table III: Conserved amino acid residues at R98^{HC} equivalent position in the IMGT human germline genes (IMGT/Gene-DB) (Giudicelli et al., 2005) or in 67,586 sequences stored in the Abysis database. Search parameters for IMGT/Gene-DB query: species: Homo sapiens, gene type: variable, functionality: functional, locus: IGH, molecular component: IG.

Residue	IMGT-germlines	Abysis
Arginine (R)	80%	68%
Threonine (T)	5%	4%
Lysine (K)	11%	15%
Histidine (H)	3%	1%
Alanine (A)	1%	2%

For the prospective design cycle we generated backmutated (BM) su3H6 variants ("BM01-BM08") containing murine backmutations at critical FR H3 positions. Selected mutants with a high similarity score (BM01, BM02, BM05, BM07) as well as a low similarity score as a negative control (BM08) were subjected to transient gene expression. Unfortunately, only BM07 gave a high enough titer for subsequent affinity assays (**Table IV**). Interestingly, for BM08 a high intracellular product content could be measured by flow cytometry that was not assembled to functional IgG molecules and secreted into the culture supernatant by HEK293-6E cells. Lysine (K) at positions 93^{HC} and 94^{HC} indeed showed only minimal presence (<1%) in 67,586 screened sequences in the Abysis database highlighting critical functions of these positions. Position 93^{HC} is occupied by Valine (V), Threonine (T), Methionine (M), or Isoleucine (I) in 67%, 11%, 7% and 7% of the sequences, respectively, and showed Lysine (K) only in less than 1%. Position 94^{HC} was even more conserved by the presence of tyrosine in 99% of all the sequences.

Table IV. Backmutated su3H6 variants subjected to transient gene expression in HEK293-6E cells.

Variant ID	Similarity score	Internal ID	Max. titer	comment
su3H6	1.2	RA01	14 µg/mL	Enough material
BM01	2.6	RA08	2.6 µg/mL	
BM02	3.3	RA09	1.3 µg/mL	
BM05	1.8	RA05	5.5 µg/mL	
BM07	2.8	RA02	10 µg/mL	Enough material (>250 µg)
BM08	1.3	RA07	0 µg/mL	

The introduction of a single $T98R^{HC}$ backmutation in the superhumanized variant su3H6 resulted in a partially re-established binding affinity of BM07 and a simulated similarity score above the threshold.

As has become obvious during this project, a fast protocol for generating enough material of multiple antibody IgG variants is required. For this, we used stable CHO-K1 cell pools or transient HEK293-6E cultures to express the 3H6 variants. Typical characteristics and benefit of each process are outlined in **Table V**.

Table V. Comparison of mAb product generation by stable cell pools versus transient cultures.

	Stable cell pools	Transient expression
Advantages	Robust Scalable High titers Material for further studies, cell banking	Very fast (7 days) Flexible for multiple constructs Lower titers
Possible limitations	Few selected variants suitable	Scalability Large amounts of DNA Higher material costs
Time	Medium (1-2 months)	Low (7 days)
Volumetric yield	Medium (25 µg/mL)	Low (10 µg/mL)
Scalability	Fully scalable after clone selection	Limited scale
Plasmid loss	No (stable integration)	High (extrachromosomal)
Material resources	Medium	High

During this project, we could also establish a novel protocol for affinity determination directly from crude culture supernatants using biolayer-interferometry (**Figure 4**). This protocol provides the advantage that even low amounts of antibody protein present in the supernatant can be measured experimentally without the need of prior protein A purification. The results obtained from this assay were confirmed from affinity measurements with protein A purified material.

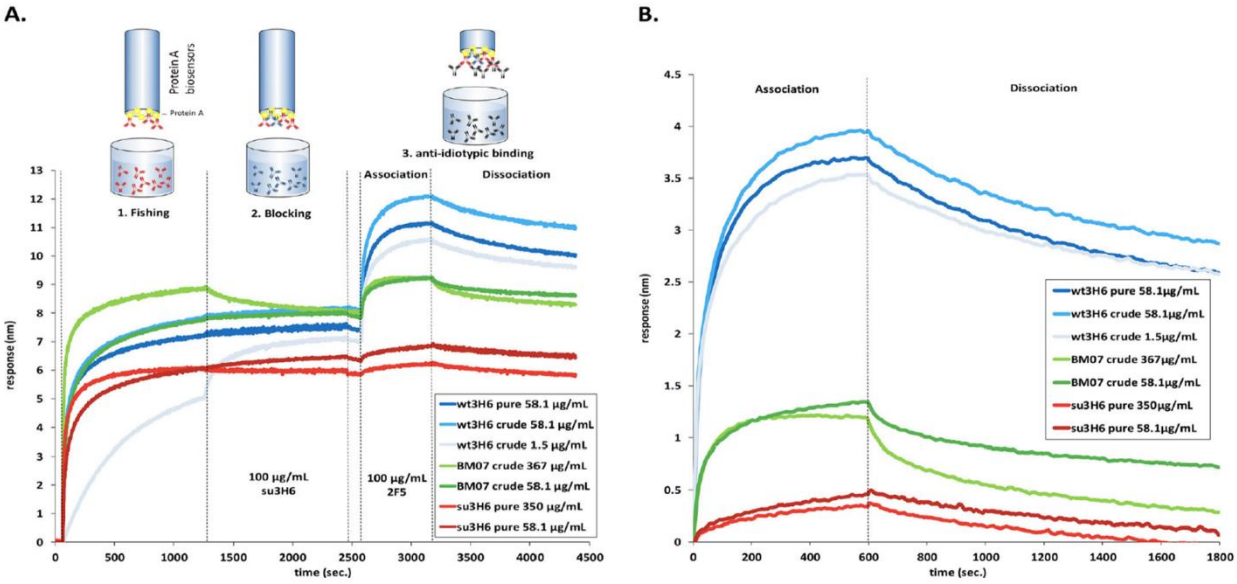
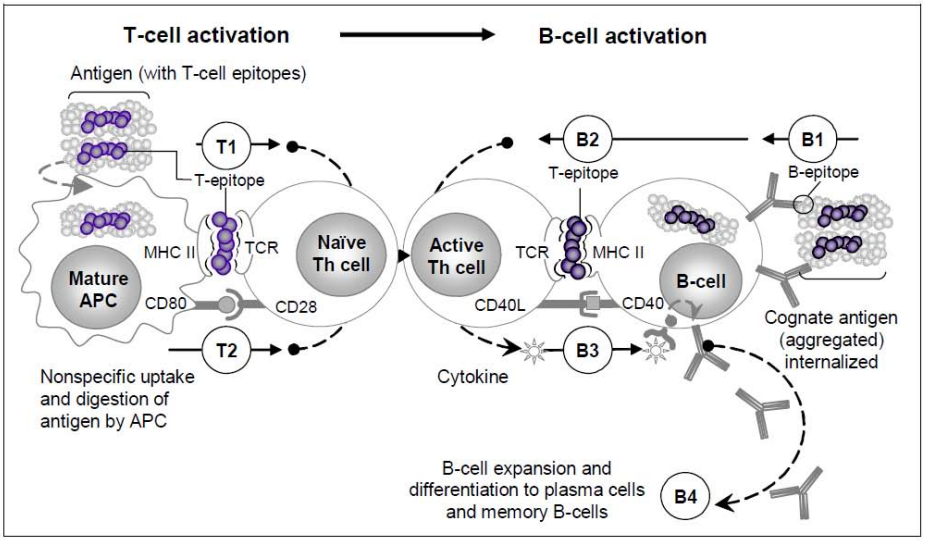


Figure 4. Affinity screening by biolayer-interferometry ("protein A fishing"). A protocol was established to screen antibody affinities directly from concentrated, crude culture supernatants. The ForteBio Octet system was equipped with protein A biosensors to immobilize transiently expressed humanized Ab2/3H6 mutants from concentrated and crude culture supernatants. (**A**) Real-time sensorgram of wt3H6, su3H6 and BM07 at different concentrations. Assay-step times were as follows: 60 sec. baseline in kinetics buffer, 1. Fishing: 1200 sec. immobilization of antibodies from crude culture supernatants, 2. Blocking with 100 µg/mL purified mAb su3H6 for 1200 sec., followed by 120 sec. baseline/washing in kinetics buffer, 3. Binding measurements: 600 sec. mAb 2F5 association (100 µg/mL) with immobilized Ab2/3H6 variants, followed by 1200 sec. dissociation in kinetics buffer only. (B) Association and dissociation curves extracted from raw data and aligned to baseline by the ForteBio software. Reprinted from Margreitter et al. (2016).

3.3 Deimmunization to rationally reduce expected immunogenicity

As was outlined before, the primary driver for protein immunogenicity is the primary protein sequence. In the body, extracellular fluids are scanned by antigen-presenting cells such as macrophages or dendritic cells in a process defined as "extracellular antigen sampling". These professional antigen-presenting cells internalize the non-autologous and potentially immunogenic biotherapeutic protein via phagocytosis or macropinocytosis followed by intracellular processing of the incorporated protein (Roche and Furuta, 2015). Inside the endosomal compartments the internalized proteins are proteolytically cleaved into smaller peptides and loaded onto major

histocompatibility complex class two (MHCII) molecules. During this process the peptide replaces the class II-associated invariant peptide (CLIP), which protects the binding site during transport of the MHCII molecule through the Golgi (Villadangos and Schnorrer, 2007). These MHCII molecules transport their cargo onto the cell surface and present the MHCII-peptide structure to T helper cells. By specific interactions of the ternary complex comprising MHCII, peptide and T-cell receptor (TCR), the T helper cell gets activated for performing Th1 or Th2 effector functions characterized by typical cytokine profiles and surface markers (Mosmann et al., 1986). Th1 cells are involved in activating CD8+ cytotoxic T cells. Th2 helper cells engage with antigen-specific B cells in the lymphoid tissue triggering the activation and differentiation of B-cells into highly productive antibody-secreting plasma cells (**Figure 5**).



APC antigen-presenting cell, **B1** B-cell signal 1, **B2** B-cell signal 2, **B3** B-cell signal 3, **B4** B-cell signal 4, **CD** cluster of differentiation, **MHC** major histocompatibility complex, **T1** T-cell signal 1, **T2** T-cell signal 2, **TCR** T-cell receptor, **Th** T-helper cell

Figure 5. Activation of T-helper cells followed by B-cell activation. Reprinted from De Groot and Moise (2007).

In all of these cell-cell interactions and activation pathways, a complex system of cytokine release and uptake is involved. The signature cytokines for Th2 cells include IL-4, IL-5 and IL-13 and only little secretion of IL-2 or absence of INF-γ, the latter two of which are typical for a Th1 response (Zhu et al., 2010).

The first step in this activation cascade including a protein-specific determinant is the MHCII-peptide interaction. Indeed, much effort has been put into elucidating the characteristics of different MHC alleles to bind a certain set of peptides with common features (Rammensee et al., 1995). Practical applications of this knowledge try to rationalize and predict the interaction of a certain peptide sequence with different MHC alleles (Wang et al., 2008). Difficulties in predicting these interactions arise from the complexity of MHC-peptide interactions. Single promiscuous peptides can be bound by multiple MHC-alleles and one MHC-allele might bind multiple peptides. Furthermore, MHCII

molecules show a high degree of polymorphism, which is determined by three genomic human leukocyte antigen (HLA) loci: HLA-DR, DQ and DP and was developed in course of evolution to account for multiple pathogens, peptide sequences and their temporal mutation rates.

Two criteria are important for a peptide to successfully trigger T cell effector functions: a high binding affinity to the MHC-molecule in order to bind long enough to induce interaction with TCR and a high degree of promiscuity, serving as a weapon against intruding pathogens and mutant variants thereof. These features were highlighted by x-ray crystal structures of MHCII-peptide complexes (McFarland and Beeson, 2002). The peptide binds in an open conformation rather than any secondary structure. In contrast to MHCI molecules, the binding groove of MHCII molecules is open at both ends, therefore longer peptides can be accommodated compared to the MHCI molecule. Binding pockets present in the MHCII binding groove may anchor amino-acid side chains of the bound peptide. Additionally, conserved hydrogen-bonds support a high-affinity binding. In contrast to MHCI, a network of additional sequence independent hydrogen-bonds can be found along the entire peptide length. Multiple peptides can bind the same MHCII molecule. Both the MHCII molecule and its peptides are promiscuous. However, the peptides that bind a certain MHCII molecule share common binding motifs. This feature is used for developing novel prediction tools for MHC-peptide interactions.

The Immune Epitope Database (IEDB) developed at the La Jolla Institute for Allergy and Immunology (LIAI), USA, is a comprehensive resource for B- and T-cell epitopes (Vita et al., 2015). It provides tools for searching experimental data of linear or discontinuous epitopes from B-, T- or MHC ligand assays. Additionally, tools for prediction of different B- and T-cell epitopes were integrated. The input for a MHCII binding prediction is a set of protein- or peptide sequences. The user has to define the prediction method and for which alleles the prediction should be made. Out of seven prediction methods the “IEDB recommended” is set as default. By using this default setting the program choses the best methods suited for a given MHC molecule based on previously observed performance. Because of ongoing research and benchmark studies to evaluate different predictors of certain MHC alleles, the IEDB recommended method might change over time to improve the prediction quality. A HLA allele reference set is provided currently consisting of 27 alleles that should provide a population coverage of >99% (**Table VI**). The output indicates the predicted allele and the peptide to which it binds. For normalizing the relative binding affinities for different MHC-peptide interactions and prediction methods a percentile rank is generated. It is expressed as the percentile compared to five million random samples of 15mers selected from the SWISSPROT database. It is accepted that a percentile rank under 20% represents strong MHC-binders.

Table VI. MHCII allele reference set provided by IEDB used for the MHCII binding prediction.
Reprinted from Greenbaum et al. (2011).

Locus	Allele	Percent of haplotypes	Phenotype frequency
DRB1	DRB1*0101	2.8	5.4
	DRB1*0301	7.1	13.7
	DRB1*0401	2.3	4.6
	DRB1*0405	3.1	6.2
	DRB1*0701	7.0	13.5
	DRB1*0802	2.5	4.9
	DRB1*0901	3.1	6.2
	DRB1*1101	6.1	11.8
	DRB1*1201	2.0	3.9
	DRB1*1302	3.9	7.7
	DRB1*1501	6.3	12.2
	Total	46.2	71.1
DRB3/4/5	DRB3*0101	14.0	26.1
	DRB3*0202	18.9	34.3
	DRB4*0101	23.7	41.8
	DRB5*0101	8.3	16.0
	Total	77.3	87.7
DQA1/DQB1	DQA1*0501/DQB1*0201	5.8	11.3
	DQA1*0501/DQB1*0301	19.5	35.1
	DQA1*0301/DQB1*0302	10.0	19.0
	DQA1*0401/DQB1*0402	6.6	12.8
	DQA1*0101/DQB1*0501	7.6	14.6
	DQA1*0102/DQB1*0602	7.6	14.6
	Total	57.1	81.6
DPB1	DPB1*0101	8.4	16.0
	DPB1*0201	9.2	17.5
	DPB1*0401	20.1	36.2
	DPB1*0402	23.6	41.6
	DPB1*0501	11.5	21.7
	DPB1*1401	3.8	7.4
	Total	76.5	94.5

In this PhD-project basic immunologic in-silico and in-vitro tests (e.g. Enzyme-Linked ImmunoSpot ELISpot) for typical immunogens such as grass pollen allergens were performed in Alessandro Sette's lab at the La Jolla Institute for Allergy and Immunology (LIAI), USA, as well as in-silico analysis of 3H6 sequences. To include information from all possible immune-epitopes but to avoid redundant identification of these epitopes the IEDB user can preprocess the protein sequence by generating 15mers overlapping by 10 amino acids spanning the whole protein sequence (Paul et al., 2013). Twenty-four 15mer peptides were generated from 3H6 sequences to span the vH antibody region. This procedure allows immunogenicity prediction of all possible 9mer core regions in a protein sequence. 15mers generated from the antibody sequences showed different propensity to bind several of the 27 selected reference MHCII alleles (**Figure 6**).

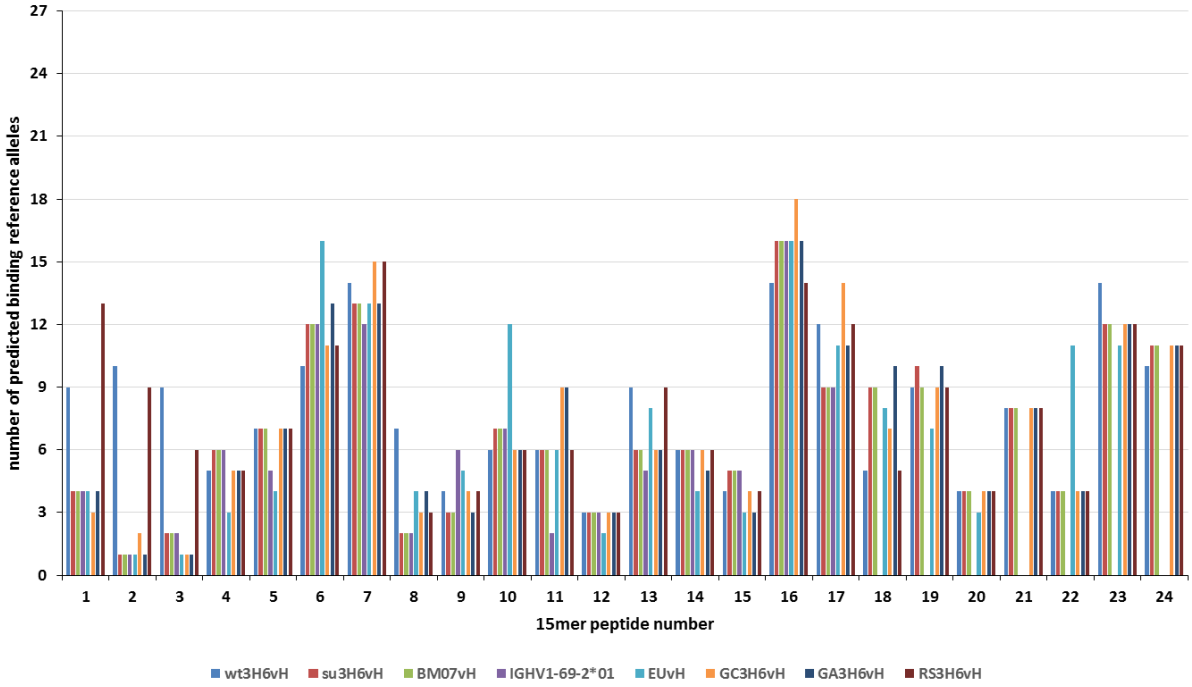


Figure 6. Total number of binding MHCII reference alleles binding to overlapping 15mers peptides generated from selected 3H6vH sequences.

Especially peptide 16 gave very high numbers in all sequences. We cannot conclude that this peptide will be more immunogenic in-vivo since also the human germline IGHV1-69-2*01 (equals to "IGHV1-f*01"), the closest human IMGT germline to the BM07vH sequence, showed a high number of binding MHCII alleles. However, it is generally accepted that for new biotherapeutics under development a lower number of binding alleles is beneficial for reducing immunogenicity and that development of anti-drug antibodies correlates with the presence of CD4+ epitopes (De Groot and Scott, 2007). Importantly, it was shown that often few very promiscuous binders account for the majority of immune reactions (Paul et al., 2015). Therefore, this method should be suitable to evaluate the expected immunogenicity on a relative level. By introducing rationally selected mutations the relative number of binding alleles might be reduced. The single backmutation T98R^{HC} present in BM07vH, that partially restored the binding affinity of the superhumanized variant, also reduced the number of binding alleles by one compared to su3H6vH (**Table VII**). Whether this minimal reduction in the number of MHCII binding alleles also translates into a reduced immunogenicity remains questionable, however it demonstrates that carefully selected mutations might improve affinity and simultaneously reduce immunogenicity. For generating a CDR-grafted antibody with minimal immunogenicity, this deimmunization process should be focused on the murine CDR and the human FR-murine CDR interface, since we can presume that human framework regions (especially germline frameworks in superhumanized variants) are generally well accepted in the human body. Indeed, the highest number of binding alleles was predicted for the murine wild-type 3H6vH sequence followed by CDR-grafted variants (**Table VII**). The lowest number of totally

predicted MHCII alleles could be observed for the superhumanized variants and for the human antibody EU. Statistical analysis might be used in combination with proper in-vitro experiments to define a proper threshold on an acceptable immunogenicity of an antibody sequence. Furthermore, in a humanization project with different humanized variants in hand, the antibody engineer can use in-silico analysis to pre-select those variants with lowest number of binding alleles for further in-vitro confirmation tests.

Table VII. Cumulative number of peptide-MHCII allele interactions of the reference set binding to selected vH 15mers.

variant	number of peptide-MHCII allele interactions
wt3H6vH	189
GC3H6vH	172
GA3H6vH	169
su3H6vH	166
BM07vH	165
EUvH	153

We can assess the changes in expected immunogenicity when murine CDR residues are grafted onto the human germline gene IGHV1-69-2*01 during superhumanization. In **Figure 7** this is demonstrated for sequence BM07vH. The first column in **Figure 7A** indicates the number of sequential 15mers overlapping by ten amino acids. Column two and three indicate the protein sequence of the respective 15mer. Differences are highlighted as red boxes and red letters. Peptides 1 to 4 and 14 to 17 did not differ from the germline gene sequence in BM07vH. Peptides 5 to 13 contain CDR regions representing murine residues in BM07vH required for binding. These murine residues in the CDR regions change the pattern of binding MHCII alleles indicated in column 4 to 30 as red boxes (**Figure 7A**). The number of differences in the type of binding alleles for each peptide is shown in column 31 whereas the total number of binding alleles for each peptide is indicated in column 32 and 33 for BM07vH or germline IGHV1-69-2*01, respectively. The introduction of murine CDR residues in BM07vH increased the number of binding MHCII alleles. The total number of binding alleles increased from 103 to 108 when murine residues were introduced. Only for peptide number nine a decrease in the number of binding alleles could be observed. Importantly, not only the number of binding alleles changed but also the type of MHCII allele (**Figure 7A**: red boxes in column 4 to 30) which might have consequences on certain human individuals carrying these alleles. For example, peptide number six showed a total of twelve predicted MHCII alleles for both the BM07 and the germline sequence, but the type of MHCII alleles were different.

To find positions within the sequence that might have the most pronounced effect on the number of MHCII alleles a ratio describing the increase in number of alleles to introduced mutations was calculated (**Figure 7B**). From that, we can see that one single $Y33F^{HC}$ mutation increases the number of MHCII alleles by two in epitope 5 and by one in epitope 7. However, it was shown that this residue forms contact with 2F5 in the crystal structure 3BQU (Bryson et al., 2008) and therefore cannot be selected as mutation for deimmunization. It might be possible to reduce the number of binding alleles by carefully selecting other residues around this position. Nevertheless, the consequences on protein structure and binding affinity should always be evaluated by proper means. A quite large increase in the total number of alleles is seen in epitope 11, which results from the introduction of a total of ten mutations in a single 15mer. This peptide also contains the CDR H2 glycosylation site of antibody 3H6. Introducing $N59I^{HC}$ mutation in BM07vH theoretically reduces the risk of inducing immunogenic reaction from this glycosylation site, but it does not reduce the number of binding MHCII alleles (not shown).

A.

# 15 mer	BM07vH	IGHV1-69-2*01	HLA-DRB1*01:01	HLA-DRB1*03:01	HLA-DRB1*04:01	HLA-DRB1*04:05	HLA-DRB1*07:01	HLA-DRB1*08:02	HLA-DRB1*09:01	HLA-DRB1*11:01	HLA-DRB1*12:01	HLA-DRB1*13:02	HLA-DRB1*15:01	HLA-DRB3*01:01	HLA-DRB3*02:02	HLA-DRB4*01:01	HLA-DRB5*01:01	HLA-DQA1*05:01/DQB1*02:01	HLA-DQA1*05:01/DQB1*03:01	HLA-DQA1*03:01/DQB1*03:02	HLA-DQA1*04:01/DQB1*04:02	HLA-DQA1*01:01/DQB1*05:01	HLA-DQA1*01:02/DQB1*06:02	HLA-DPA1*02:01/DPB1*01:01	HLA-DPA1*01:03/DPB1*02:01	HLA-DPA1*01/DPB1*04:01	HLA-DPA1*03:01/DPB1*04:02	HLA-DPA1*02:01/DPB1*05:01	HLA-DPA1*02:01/DPB1*14:01	no. of different alleles	binding alleles in BM07vH	binding alleles in IGHV1-69-2*01
1	EVQLVQSGAEVKKPG	EVQLVQSGAEVKKPG																													0	4	4
2	QSGAEVKKPGATVKI	QSGAEVKKPGATVKI																													0	1	1
3	VKKPGATVKISCKVS	VKKPGATVKISCKVS																													0	2	2
4	ATVKISCKVSGYTFT	ATVKISCKVSGYTFT																													0	6	6
5	SCKVSGYTFTDYFMH	SCKVSGYTFTDYYMH												■																■	2	7	5
6	GYTFTDYFMHWVQQA	GYTFTDYYMHWVQQA		■					■																						2	12	12
7	DYFMHWVQQAPGKGL	DYYMHWVQQAPGKGL																								■					1	13	12
8	WVQQAPGKGLEWMGY	WVQQAPGKGLEWMGL																													0	2	2
9	PGKGLEWMGYINCYT	PGKGLEWMGLVDPED				■		■					■	■				■			■	■									7	3	6
10	EWMGYINCYTGATNY	EWMGLVDPEDGETIY		■	■			■	■				■	■				■			■	■									10	7	7
11	INCYTGATNYSQKFK	VDPEDGETIYAEKFQ			■	■		■	■							■		■			■	■									8	6	2
12	GATNYSQKFKGRVTI	GETIYAEKFQGRVTI							■				■													■	■				4	3	3
13	SQKFKGRVTITADTS	AEKFQGRVTITADTS							■																						1	6	5
14	GRVTITADTSTDTAY	GRVTITADTSTDTAY																													0	6	6
15	TADTSTDTAYMELSS	TADTSTDTAYMELSS																													0	5	5
16	TDTAYMELSSLRSED	TDTAYMELSSLRSED																													0	16	16
17	MELSSLRSEDTAVYY	MELSSLRSEDTAVYY																													0	9	9
																														total:	**108**	**103**

B.

peptide	BM07 alleles	IGHV1-69-2*01 alleles	delta no. of alleles	murine mutations	increase of alleles/mutation	alignment (Kabat CDRs underlined)	
5	7	5	2	1	2.0	BM07_vH_5_(21-35)	SCKVSGYTFTDYFMH
						SUGL_vH_5_(21-35)	..........Y..
7	13	12	1	1	1.0	BM07_vH_7_(31-45)	DYFMHWVQQAPGKGL
						SUGL_vH_7_(31-45)	..Y...........
11	6	2	4	10	0.4	BM07_vH_11 (51-65)	INCYTGATNYSQKFK
						SUGL_vH_11 (51-65)	VDPED.E.I.AE..Q
13	6	5	1	3	0.3	BM07_vH_13 (61-75)	SQKFKGRVTITADTS
						SUGL_vH_13 (61-75)	AE..Q..........

Figure 7. Selected 15mer peptides that showed an increase in the number of binding MHCII alleles when murine CDR regions were introduced into the human IGHV1-69-2*01 allele during superhumanization. Murine mutations affect both the number and types of alleles (**A**). Differences in sequence or binding alleles are highlighted red. Inspection of 15mer sequences may provide valuable hints for selection of deimmunizing mutations (**B**).

We assume that epitopes present in human germline IGHV1-69-2*01 are generally accepted within the human body. We cannot exclude that possible new epitopes introduced by murine CDR residues are also present in other human germline sequences. A more rigorous screening of all human germline genes in future projects might identify “novel” epitopes introduced by the murine CDR residues that are not present in any human germline gene and therefore possibly immunogenic.

Similarly to how MD simulations might support the humanization approach for a priori selection of relevant backmutations, also in-silico T-cell epitope prediction tools might reduce the experimental workload in a prospective design cycle (Jawa et al., 2013). However, at the current developmental stage of the in-silico predictive tools, confirmation should always be provided by proper in-vitro/in-vivo tools to continuously improve the prediction power. **Table VIII** gives an overview of methods

available for prediction and confirmation of protein immunogenicity. It is important to emphasize that MHCII-peptide interactions are required for an immune response but not sufficient. For example, not all peptide-MHCII complexes will eventually form ternary complexes with the T-cell receptor. Also, the intracellular pathway and proteolytic process plays a significant role if a potential immune epitope is loaded onto MHCII molecules, presented on the cell surface and eventually leads to an immune response in-vitro/in-vivo.

Table VIII. Overview of available prediction tools and in-vitro/in-vivo confirmation techniques for protein immunogenicity assessment. Further information can be obtained from De Groot and Moise (2007).

Platform	Methods	Description	References
In-silico	EpiMatrix (EvpiVax)	Proprietary T-cell epitope mapping algorithm	Schafer et al., 1998
	IEDB consensus	Free online epitope database including tools for B-, MHCI, MHCII epitope predictions	Vita et al., 2015 Paul et al., 2013
In-vitro	HLA binding assay	a. Measurement of peptides ability to compete for the binding site of a labeled high-affinity ligand to purified MHCII molecules b. direct binding assays c. real-time kinetics assays	Sidney et al., 2008
	T-cell activation assays	Incubation of bulk cultures of PBMCs followed by measurement of a. cytokine secretion (e.g. IFNg, IL-2, IL-4, IL-5) by ELISpot b. expression of intracellular cytokine and transcription factors, extracellular surface marker or direct labeling using tetramers (MHCII:peptide complexes) by flow cytometry c. cell proliferation by 3H-thymidine incorporation	Finco et al., 2014 Findlay et al., 2011
In-vivo	HLA transgenic mice models	Measurement of T-cell responses or development of antibody responses	Brinks et al., 2011

3.4 Discussion

In this project, a workflow towards a prospective design cycle for antibody humanization was established. MD simulations were implemented as a new tool to guide the selection of critical backmutations. The methods and procedures presented in this work lead to improved humanization processes. Especially the use of MD simulations as a new humanization tool provides novel opportunities to pre-select required backmutations based on in-silico experiments. We were able to establish a simulation workflow that correlates the similarity of simulated CDR dynamics to the behavior of the wild-type mAb by calculating experimental binding free energies. In Margreitter et al. (2016) we published the design and construction of a training antibody set by selecting six wt3H6 framework positions, mutated them to their respective su3H6 residue ($A68V^{HC}$, $V72A^{HC}$, $R98T^{HC}$, $V4L^{LC}$, $R45I^{LC}$, $D85Y^{LC}$) and combined them to yield wt3H6 heavy- (TR01-03) or light chain (TR04-05) double-mutants. These six positions were rationally chosen based on knowledge from literature and available crystal structure (PDB 3BQU) determining the Vernier zone, the CDR canonical structure class and the overall protein structure or to be spatially positioned near CDR loop structures or in-between the vH/vL interface. These positions were also selected as backmutation sites in a previous CDR-grafting procedure of 3H6 (Mader and Kunert, 2010). The training set consisting of binders (wt3H6, TR01, TR04, TR05 and TR06) and non-binders (su3H6, TR02 and TR03) allowed us to optimize the MD simulation workflow. The significant loss in binding affinity of wt3H6-mutant TR02 ($A68V^{HC}$/$R98T^{HC}$) and TR03 ($V72A^{HC}$/$R98T^{HC}$) could be attributed to the single framework position $R98T^{HC}$, directly flanking the CDR H3. Sequence analysis of germline or mature antibody genes extracted from the IMGT or Abysis database indiciated a conservation of up to 80% for arginine at this very position, suggesting a fundamental structural function. Recently, with the help of crystal/homology models we identified a complex cation-pi interaction system for this arginine with several tyrosine residues in an arrangement that can be described as a “tyrosine-cage” (unpublished observation). In future work the function of and the residues involved in this tyrosine-cage will be investigated in more detail.

The TR variants were expresssed in CHO cell pools after stable selection with G418. Although this production process took two months, it is a rather robust and easy method to produce sufficient amounts of antibody proteins. More than 1 mg of purified antibody material could be readily prepared and used for accurate determination of binding affinities to the biotinylated idiotype antibody (2F5). The dissociation constant (K_D) was measured with label-free biolayer-interferometry using a streptavidin-biotin capture step which allows the analysis of antibody-antigen interactions in real-time and determination of association (k_{on}) and dissociation (k_{off}) rates by fitting a binding model on the curves. The K_D can be expressed as binding free energies to link the experimental observations of in-vitro data to thermodynamics from MD simulations. Binding free energies are routinely used in the field of MD simulations to describe molecular interactions driven by enthalpic and entropic contributions. Using this relationship the wt3H6 binding free energy ($\Delta G_{binding}$) of -50 kJ/mol translates into a dissociation constant (K_D) of approximately 2 nM.

Based on the optimzed MD simulation from the training setup we simulated different su3H6 backmutation variants (BM01-07) and a double lysine variant BM08 ($V93K^{HC}$/$Y94K^{HC}$). The MD simulation identified only BM08 to give a low similarity score, while BM01-BM07 showed high similarity scores. BM antibody variants were expressed transiently in HEK293-6E cultures using a optimized protocol which combines episomal plasmid replication from the EBNA-1/oriP system, hypothermic cultivation and addition of valproic acid (VPA), an inhibitor of histone deacetylases (iHDAC). However, the total protein yield was much lower than for stable pools. Therefore, we established a protocol to determine the binding behavior of 3H6 variants directly from crude culture supernatants with protein A biosensor capture step and compared the results to the already established streptavidin-biotin setup using biolayer-interferometry with purified protein samples. The bioassay from crude culture supernatants allowed us to qualitatively assess the binding behavior of novel 3H6 variants expressed in transient cultures in 30 mL scale. However, for BM08 no antibody secretion was observed despite intracellular heavy- and light chain expression was confirmed by flow cytometry. Since BM08 represents an artificial antibody sequence it is possible that the chosen double lysine mutations are not compatible with functional assembly of the antibody molecule. Analysis of the BM08 sequence indicate that the double lysine residues show up as highly under-represented feature (<1%) compared to a large set of antibody sequences stored in the Abysis database. In contrast, the BM07 variant was defined by introducing a single $T98R^{HC}$ backmutation, showed an enhanced simulated similarity score, could be expressed in cell cell culture and received partially re-established binding affinity. This variant will be the starting point for further affinity optimization by MD simulations. With this example we were able to prove the great potential of in-silico methods to pre-select certain variants of mAb humanization approaches and to reduce the experimental workload by providing additional knowledge about the influence of distinct amino acids.

An additional aspect for therapeutic drug application are highest safety standards provided by low immunogenicity of the protein. In case of mAbs from rodent origin the humanization process should be combined with a deimmunization process. Reducing the total number of possible peptide-MHCII allele interactions should result in a decrease of expected immunogenicity in-vivo. However, in case of anti-idiotypic antibodies reduction in the number of binding MHCII alleles might be detrimental since the therapeutic effect of this class of antibodies relies on the development of anti-drug antibodies directed against the idiotype as it should be applied in an active vaccination strategy. In contrast to passive administration of mAbs, which requires several and high dose applications over long time-periods for effective therapeutic effects, vaccine proteins are administered only few times to induce an active immune response. Therefore, for antibody 3H6, the ultimate goal might be to maintain certain MHCII epitopes to focus the immune response towards 3H6 sequences responsible for inducing a 2F5-like antibody response but to remove MHCII epitopes at other sites. Presence of certain MHCII epitopes might be tolerated without inducing a severe immune reaction. For an effective vaccination effect, certain MHCII epitopes are even absolutely required to induce an effective immune and memory response and the prediction and identification of such epitopes still is

in the focus of ongoing research projects. By combining in-silico and in-vitro data, the antibody engineer might direct the anti-drug antibody response to the conformation of interest, maintain beneficial T-cell epitopes and remove other non-essential T-cell epitopes.

4 GERMINALIZATION – A CONCEPT TO MODIFY ANTIBODY- AND CELLULAR PROPERTIES

4.1 Background

The limit of specific productivity of recombinant animal cells is often discussed but valuable standards are often not available. Monoclonal antibodies are most often mentioned in combination with extraordinary production rates but in our experience and affirmed by literature, not all antibodies reach identical production titers in bioprocesses. The comparison with the human body as expression system indicated that terminally differentiated plasma cells can express antibodies up to a rate of 10,000 antibody molecules per second (Hibi and Dosch, 1986), which translates into a specific productivity of 215 pg/c/d for an IgG1 antibody assuming an average mass of 150 kDa. Mammalian cell culture processes currently reach a specific productivity of 70-100 pg/c/d (Li et al., 2010; Rita Costa et al., 2010) in industrial scales after four decades of continuous vector-, cell culture- and process optimization. Beside the host cell line and the process parameters, the productivity strongly depends on the primary sequence of the protein of interest. This clearly highlights room for improvements towards the biological limit by integrating knowledge from protein properties and cellular behavior.

Within the human body, functional antibody sequences are generated by recombination of different sets of variable (V), diversity (D) and joining (J) genes (**Figure 8A**). These sets contain up to 46 V-, 27 D- and 6 J functional gene segment in the human heavy chain locus, 35 Vκ and 5 Jκ functional gene segments for the kappa light chain locus or 35 Vλ and 4 Jλ segments for lambda light chain locus (Schroeder, 2006). They are combined by a process called V(D)J-recombination (**Figure 8B**) to generate a theoretical antibody diversity in the order of 10^{13} individual variants when combined with introduction of insertions or deletions at the gene segment junctions and a process called somatic (hyper)mutation (Kim et al., 1981). The latter process is catalyzed by the enzyme activation-induced cytidine deaminase (AID) and is one primary factor for enhancing antibody affinity towards its antigen. Introduction of somatic mutations is primarily focused on the CDR region of an antibody sequence. This property was used to define the Kabat CDR-regions based on sequence variability (Wu and Kabat, 1970). In reality, the human immune system forms distinct combinations of

preferable regions and therefore it needs to be investigated if the primary sequence of antibodies influences the quantity of expression.

Somatic mutations, representing the maturation stage of an antibody sequence, not only influence affinity and specificity but also physicochemical properties of the antibody variable regions such as thermostability (Dimitrov et al., 2014). The interplay of somatic mutations for determining antibody affinity, promiscuity and thermostability is still subject of intense research. For example, there is growing evidence that antibodies are first selected for affinity by reducing promiscuity by somatic mutations, but are later in the maturation process somatically mutated to regain some of its promiscuity (Dimitrov et al., 2014). This phenomenon can also be observed in many broadly neutralizing anti-HIV antibodies showing promiscuous behavior despite their unusual high degree of somatic mutations (Mouquet et al., 2010; Scheid et al., 2009). Another study showed the separation of certain sets of somatic mutations to independently increase affinity or thermostability (Wang et al., 2013). In the same study, it was concluded that additional somatic mutations are required for compensating the decreased thermostability during affinity maturation. In this project we hypothesized an additional feature to the somatic mutations. Does the degree of somatic maturation of an antibody sequence correlate to the in-vitro expression levels and does the primary amino acid sequence of mAbs influence and guide cellular adaptation processes in-vitro? In a further perspective, we want to know whether it is possible to rationally enhance antibody attributes, such as protein stability and expression levels, by simultaneously maintaining affinity.

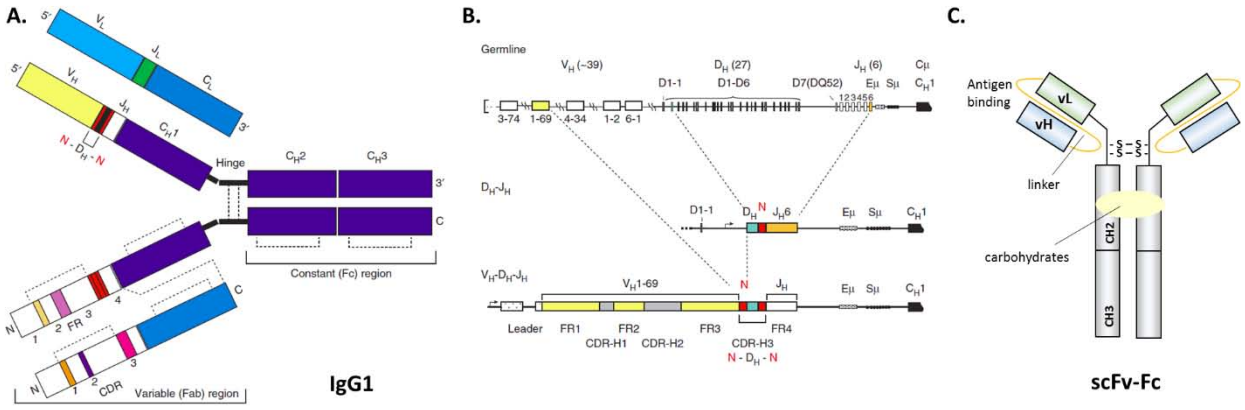


Figure 8. (Genomic) antibody structure and assembly. Individual subdomains of a human IgG1 antibody (**A**) are assembled from a polymorphic set of variable (V), diversity (D) and joining (J) genes by a process called V(D)J-recombination (**B**) (modified from Georgiou et al., 2014). Synthetic constructs such as antibodies in scFv-Fc format can be designed for investigation of the expression system and antigen-antibody interaction (**C**).

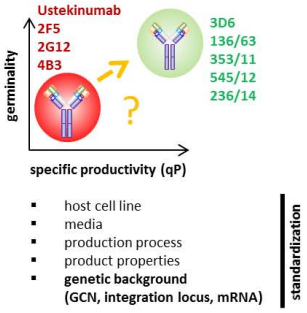
A typical feature of neutralizing anti-HIV antibodies, and especially broadly-neutralizing antibodies such as 2F5, is a high degree of somatic mutations and unusual long CDR-regions to gain access to hidden or flexible parts of the HIV envelope glycoprotein. For 2F5 the Kabat CDR region reaches 22 amino acids in length, indicating a high degree of flexibility of this antibody surface loop, whereas for

other human antibodies most often the CDR H3 loop has a length of 10-14 amino acids (Wu et al., 1993). Broadly neutralizing anti-HIV antibodies are found in relatively low abundance in patients infected by HIV for a long time-period. This persistent infectious viral load and exposure to HIV antigens drives the development of rare antibody variants comprising highly somatically mutated variable regions required for binding to hidden but conserved surface epitopes of different strains and mutants of the virus. Interestingly, these somatic mutations are not only concentrated to the CDR regions but also to the usually conserved framework regions at an unusual high frequency (Klein et al., 2013). In technological aspects, these unique properties of anti-HIV antibodies imply special challenges for production of these monoclonal antibodies and the understanding how protein sequences influence product properties as well as cellular behavior. Since expressability is of great technological importance for this class of antibodies and for other antibodies in general, knowledge about interaction of the protein and the cell biology is significantly valuable. The high somatic mutation rate required for antigen binding might have pronounced effects on the stability and observed expression level in cell cultures. It would be beneficial to determine general sequence-related parameters that influence affinity separated from the influence on (enhanced) expressability.

In the past, we could observe a good correlation between high expression levels and high sequence identity to the closest human germline sequence in different projects. This effect seemed to be vector-, host cell line-, isotype- and operator independent (**Figure 9A**) and could be demonstrated in a systematical approach for antibody 2F5 in comparison to 3D6 (Mader et al., 2012; Sommeregger et al., 2016). In another project, 2G12 expression levels and protein stability could be enhanced by simultaneously maintaining binding properties by introducing 17 germline residues (**Figure 10D**) within the mature antibody framework in IgM antibody molecules (Chromikova et al., 2015).

However, since many factors in mammalian cell cultures influence the observed productivity a standardized cell line development and cultivation platform is highly demanded to compare different antibody expression levels accurately (**Figure 9A**). In a bioprocess technological view, the aspect of standardized media and culture conditions can be guaranteed quite easily by controlled fermentation parameters. However, controlling cellular and genetic parameters such as gene copy number and providing equal transcription levels as a first determinant of transgene expression, is challenging. This, in part, is due to the chromosomal position effect (Wilson et al., 1990) leading to the observation that gene expression level is not always directly proportional to gene copy number (Lattenmayer et al., 2007). Favorably, different antibody sequences should be expressed under isogenic conditions with equal gene copy numbers and under the same cell-physiological background using expression hosts with equal metabolic and proteomic background.

A.



B.

Low expressing antibody ?	IGHV (% aa)	IGHD	IGHJ	IGK/LV (% aa)	IGK/LJ	High expressing antibody?
	IGHV3-9*03 (96%)	IGHD3-22*01	IGHJ3*02	IGKV1-5*03 (98%)	IGKJ3*01	3D6
2F5	IGHV2-5*02 (86%)	IGHD3-3*02	IGHJ6*02	IGKV1-13*02 (85%)	IGKJ4*01	236/14
4B3	IGHV1-69*10 (83%)	IGHD3-22*01	IGHJ6*02	IGLV6-57*01 (90%)	IGLJ3*02	136/63
2G12	IGHV3-21*01 (68%)	IGHD5-12*01	IGHJ3*01	IGKV1-5*03 (84%)	IGKJ1*01	353/11
Ustekinumab	IGHV5-51*03 (88%)	IGHD5-5*01	IGHJ4*02	IGKV1D-16*01 (99%)	IGKJ2*01	554/12

Germline variants

Figure 9. Does germline identity determine expression levels under standardized conditions? The assumption that specific productivity increases with sequence identity to the closest human germline gene is additionally influenced by multifactorial parameters (**A**). In this project a panel of antibody variants consisting of mature antibodies with low germline homology (red) is compared to designed germline variants or high-germinality mAb 3D6 (green) (**B**).

4.2 Construction and definition of mature and germline antibody variants

In order to investigate how the antibody's amino acid sequence influences cellular behavior and protein properties, with special focus on anti-HIV antibodies, a set of germline antibodies was constructed in IgG1 format based on the mature antibody sequence of 2F5, 2G12 and 4B3 (**Figure 9B**). Additionally, a human therapeutic antibody (Ustekinumab) with low sequence homology to its closest germline sequence was included (**Table IX**) in this project.

Table IX. Sequence identity to the closest germline V-sequence of selected fully-human therapeutic antibodies. Protein sequence information was retrieved from patents, PDB database or online resources. Sequence identities below 90% are highlighted in red.

INN name	tradename	origin	target	clinical domain	company	clinical status	IGHV	IGK/LV	Ref.	% aa identity (HV)	% aa identity (L/KV)
Ustekinumab	Stelara	human	IL12/23	immunology	Janssen-Cilag	approved 2009	IGHV5-51*03	IGKV1D-16*01	[1, 2]	88%	99%
panitumumab	Vectibix	human	EGFR	Oncology	Amgen	approved 2006	IGHV4-61*01	IGKV1-33*01	[3]	90%	94%
canakinumab	Ilaris	human	IL-1b	Immunology	Novartis	approved 2009	IGHV3-33*01	IGKV6-21*01	[4]	94%	99%
adalimumab	Humira	human	TNFa	Immunology	AbbVie	approved 2002	IGHV3-9*01	IGKV1-27*01	[5]	94%	96%
golimumab	Simponi	human	TNFa	Immunology	Centocor	approved 2009	IGHV3-30*16	IGKV3-11*01	[6]	95%	99%
Ipilimumab	Yervoy	human	CTLA-4	Oncology	Bristol-Myers Squibb	approved 2011	IGHV3-30*16	IGKV3-20*01	[7]	95%	98%
Belimumab	Benlysta	human	BLyS	immunology	GlaxoSmithKline Human Genome Sciences	approved 2011	n.a.	n.a.	[8]	n.a.	n.a.
Denosumab	Prolia/XGEVA	human	RANK-L	Osteology	Amgen	approved 2010	IGHV3-23*01	IGKV3-20*01	[9]	98%	96%
Ofatumumab	Arzerra	human	CD20	Oncology	Genmab	approved 2009	IGHV3-9*01	IGKV3-11*01	[10]	97%	100%
Ramucirumab	Cyramza	human	VEGFR2	Oncology	Eli Lilly	approved 2014	IGHV3-21*04	IGKV1-12*01	[11]	99%	85%
Nivolumab	Opdivo	human	PD-1	Oncology	Bristol-Myers Squibb	approved 2014	IGHV3-33*03	IGKV3-11*01	[12]	92%	99%

[1] WO2013087911
[2] PDB 3HMX
[3] WO 2012138997 A1
[4] CA2626214A1
[5] WO2013087911
[6] WO2013087911
[7] EP 1503794 B9
[8] http://www.antibodysociety.org/news/approved_mabs.php
[9] DB06643
[10] PDB 3GIZ
[11] PDB 3S37
[12] WO2013173223

All therapeutic human antibodies showed a high degree of similarity to its closest human germline gene. This property reduces the risk for inducing immunogenic reactions within the human body contributing to the success of the drug approval process by EMA or FDA. Another striking feature was the high abundance of antibodies derived from the immunoglobulin variable heavy gene segment IGVH3. Indeed several reports were published that highlight a favorable framework, especially in combination with IGHK1. For these classes of variable segments enhanced thermostabilities and expressability leading to higher lead compound identifications in phage library projects are proposed (Chen et al., 2015; Ewert et al., 2003; Tiller et al., 2013).

To evaluate the influence of somatic mutations we designed germline variants of four distinct antibodies (2F5, 4B3, Ustekinumab, 2G12) by identifying the respective closest human germline V-,

D- and J-gene segments and proper combinations based on Blastp search to identify reported V-D-J combinations (**Figure 10**).

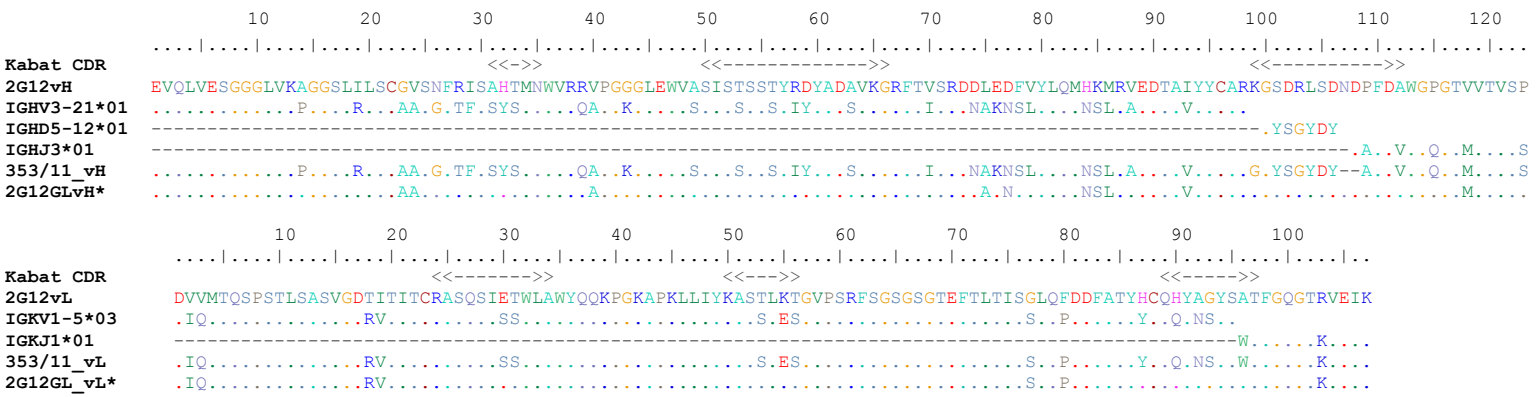
The rationale behind the construction of this mature/germline antibody set is based on a previous proof of principle on how similar antibody sequences influence cellular behavior. This was demonstrated for anti-HIV antibody 2F5 and 3D6 in scFv-Fc format (Mader et al., 2012).

Tailor-made antibody constructs (e.g. scFv-Fc) can be designed by assembly of functional antibody regions. In the past, this was successfully demonstrated for rather small antibody constructs crossing the blood-brain barrier or T-cell engager (BiTe) which are chimeric antibody molecules binding two different antigens simultaneously (bispecific). Construction of artificial molecules allows us to simplify rather complex IgG1 heterotetrameric structures into homodimeric scFv-Fc molecules for single-chain transfections and elevated expression levels but full functionality for antigen binding affinity, variable domain-mediated antibody properties (e.g. stability) and Fc-mediated antibody effector functions. A workflow to design and construct scFv-Fc variants from plasmid templates comprising the antibody sequence in IgG1 format is reported in a submitted manuscript as first-author:

Mayrhofer, Patrick and Renate Kunert. Cloning of single-chain antibody variants by overlap-extension PCR for evaluation of antibody expression in transient gene expression. *Methods in Molecular Biology*, under review 2016


A.
Kabat CDR
3D6_vH
IGHV3-9*03
IGHD3-22*01
IGHJ3*02
3D6_vL
IGKV1-5*03
IGKJ3*01
B.
Kabat CDR
2F5_vH
IGHV2-5*01
IGHD3-3*02
IGHJ6*02
236/14_vH
Kabat CDR
2F5_vL
IGKV1-13*02
IGKJ4*01
236/14_vL
C.
Kabat CDR
4B3_vH
IGHV1-69*10
IGHD3-22*01
IGHJ6*01
136/63_vH
Kabat CDR
4B3_vL
IGLV6-57*01
IGLJ3*02
136/63_vL

D.



*Ref. Chromikova et al., 2015

E.

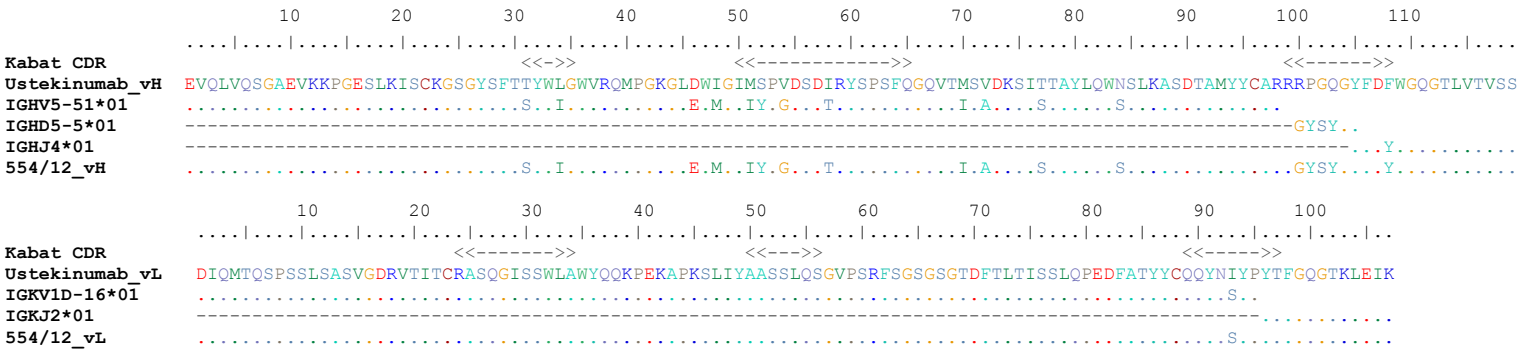


Figure 10. Sequence alignment of anti-HIV antibodies 3D6 (A), 2F5 (B), 4B3 (C) and 2G12 (D) and the therapeutic human antibody Ustekinumab (E) to the closest human germline V-, D-, J-gene segments. Segments were joined to form a germline antibody variant. 2G12GL sequence is indicated as was reported in Chromikova et al. (2015).

4.3 RMCE – A necessary tool for accurate comparison of antibody expression levels

As outlined before, many different factors determine the observed antibody expression level, impeding accurate comparison of different antibody variants. One major factor is the gene copy number of the protein of interest. Traditional methods for stable integration of the antibody-encoding open reading frame rely on random integration resulting in stable cell clones with unpredictable gene copy-numbers at undefined genomic loci. Besides, it is generally accepted that different chromosomal integration sites exhibit various transcription efficiencies based on their chromosomal architectures. Because of this position effect the observed expression levels do not linearly correlate with gene copy number.

To overcome these limitations different gene-targeting techniques were established to insert different genes of interest into the same pre-defined chromosomal loci with consistent gene copy numbers. A big breakthrough in targeted gene integration techniques was achieved with the development of optimized CRISPR/Cas9 systems using so called guide RNAs to specifically define the chromosomal loci to be targeted. Using CRISPR/Cas9 the process of introducing site-specific nucleotide exchanges or insertion/deletions (indels) became feasible. However, at the current developmental stage of this technique the major challenges in using this CRISPR/Cas9 are unpredictable off-target effects. Furthermore, our knowledge about what makes a chromosomal location a transcriptionally highly active locus is still limited. This means that we have the technique in hand to insert any gene of interest but we do not know which locus we should target for generating a stable cell line with high expression capacity. Alternatively, recombinase-mediated cassette exchange (RMCE) is defined as a “tag-and-exchange” technique (Seibler et al., 1998). A recombination-competent cassette containing a reporter/selection marker is stably integrated into the host cell line randomly and expression of the marker is evaluated for production level and stability. The newly established RMCE host can then be used to site-specifically introduce any gene of interest into the pre-selected and characterized genomic locus by exchanging the marker gene under the influence of a recombinase enzyme.

In this project the flipase (FLP)/flipase recognition-target (FRT) system was used (Qiao et al., 2009). A proof of concept using this system was published for using the CHO DUKX-B11 cell line (Mayrhofer et al., 2014). In this publication we demonstrate the construction of a RMCE host cell line, called “DUKX-B11 F3/F” containing a reporter/selection marker flanked by the heterospecific FLP-FRT variants (FRT3 and wild-type FRT) 3’ downstream of a CMV promoter (**Figure 11**).

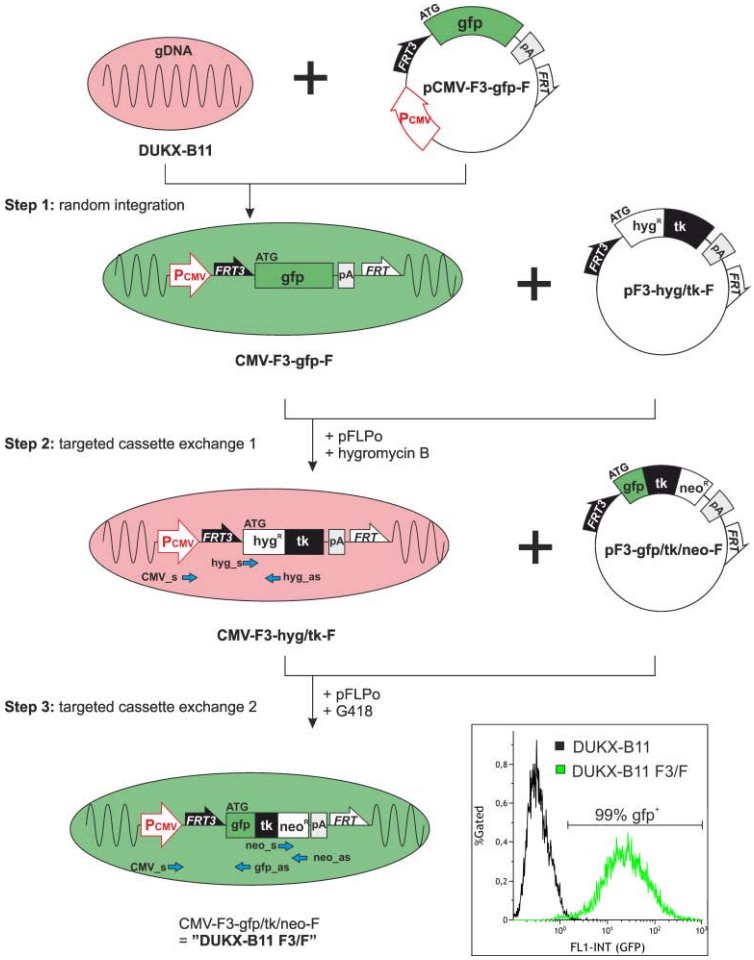


Figure 11. Establishment and homogenous intracellular gfp fluorescence of the RMCE-competent host cell line DUKX-B11 F3/F. DUKX-B11 F3/F was established by random integration of a RMCE cassette on plasmid pCMV-F3-gfp-F (step 1) followed by two consecutive rounds of RMCE (steps 2 and 3) mediated by co-transfection of the targeting vector (pF3-hyg/tk-F or pF3-gfp/tk/neo-F) and the FLP expression plasmid (pFLPo). Stable cell clones were selected by hygromycin B (step 2) or G418 (step 3). Primers used for genomic PCR characterization are indicated as blue arrows. The final DUKX-B11 F3/F cell line was analyzed for gfp fluorescence by flow cytometry shown as a single-parameter (FL1-INT) histogram. Non-fluorescent DUKX-B11 cells were used as negative control. Plasmids are indicated as circles, DUKX-B11 cells are indicated as pink or green ovals, indicating gfp-negative or gfp-positive fluorescence, respectively, and the cassettes integrated into the host genomic DNA (gDNA) are shown. Genetic elements: pCMV CMV promoter, FRT3/FRT spacer-mutant/wild-type FLP recognition target sites, gfp green fluorescent protein, pA poly A signal, ATG start ATG codon, hygR hygromycin phosphotransferase, tk thymidine kinase, neoR neomycin phosphotransferase. Figure reprinted from Mayrhofer et al. (2014).

The special arrangement of the individual elements allowed us to establish a promoter trap to prevent transcription of randomly introduced antibody sequences. Thereby, we could introduce two highly-similar antibody variants 3D6 and 2F5 in scFv-Fc format (**Figure 8C**) into the same genomic locus with equal gene-copy numbers (data not shown). The isogenic cell lines developed for each antibody variant showed similar mRNA transcript levels (**Figure 12A**) and allowed us an unambiguous comparison of 2F5 and 3D6 antibody sequences. We could observe similar growth patterns (**Figure 12B**), but a two-fold higher expression level for 3D6scFv-Fc (**Figure 12C**), which

confirmed previous results using random plasmid integration or bacterial artificial chromosomes (Mader et al., 2012).

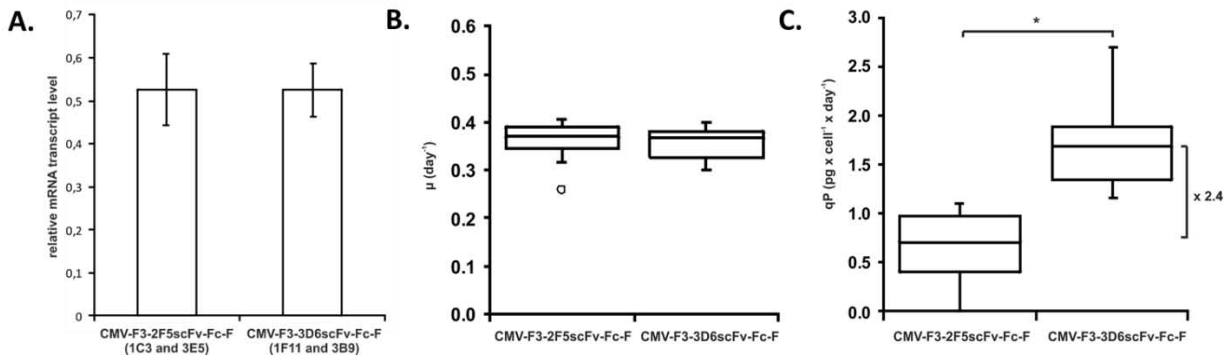


Figure 12. Isogenic antibody producing RMCE subclones show similar mRNA transcript levels but antibody sequence-dependent differences in specific productivities. qPCR analysis of mRNA transcript level was performed for two selected 3D6scFv-Fc- or 2F5scFv-Fc-producing RMCE clones cultivated in spinner flasks (**A**). Samples were taken at three different days and measured in two technical replicates. Total mRNA was reverse transcribed into cDNA and analyzed by qPCR using probes specific for the Fc sequence or β-actin used as an internal standard. Mean $2^{-\Delta Cp}$ values were calculated based on differences of Cp values between β-actin and the Fc sequence. Error bars represent standard deviation. Specific growth rates μ (**B**) and specific productivities qP (**C**) are shown as box plot diagrams of 12 scFv-Fc producing RMCE clones of each antibody variant cultivated for ten consecutive passages in T25 flasks. Boxes represent median, first, and third quartiles of 12 clones. Outliers were defined as values ±1.5 × interquartile range (IQR) and are represented as open circles. Error bars represent sample maxima and minima within 1.5 × IQR. Factor 2.4 represents difference in median-specific productivities between 3D6scFv-Fc- and 2F5scFv-Fc-producing clones. *$p < 0.001$—independent two-sample Student's t test. Figures reprinted from Mayrhofer et al. (2014).

The construction of the novel RMCE host cell line DUKX-B11 F3/F and characterization of isogenic antibody producing daughter cells was published in a first-author manuscript:

> **Mayrhofer, P.**, Kratzer, B., Sommeregger, W., Steinfellner, W., Reinhart, D., Mader, A., Turan, S., Qiao, J., Bode, J., and Kunert, R. (2014). Accurate comparison of antibody expression levels by reproducible transgene targeting in engineered recombination-competent CHO cells. Appl. Microbiol. Biotechnol. *98*, 9723–9733.

A second RMCE host cell line was established with improved production and growth capabilities by using CHO-K1 as parental cell line and a tagging vector containing the CAGGS promoter. We generated this second RMCE competent engineered host cell line because of improved growth properties and higher maximum cell densities compared to the dihydrofolate reductase (dhfr) negative cell line CHO DUKX-B11. A comparison of key attributes between the two RMCE host cell lines is shown in **Table X**.

Table X. Key differences between two RMCE host cell lines DUKX-B11 F3/F and CHO-K1 RMCE I3.

	DUKX-B11 F3/F	CHO-K1 RMCE I3
Original Host	DUKX-B11 (dhfr$^-$) (ATCC CRL-9096)	CHO-K1 (ATCC CCL-61)
Promoter	CMV	CAGGS
Exchange cassette	FRT3-GTN-FRT	FRT3-GTN-PGK-kana/neo-FRT
Reference	Mayrhofer et al., 2014	unpublished

A total of 8 different antibody sequences for 4B3, 2G12, 2F5, Ustekinumab and their germline variants were site-specifically introduced into the I3 RMCE host. The parental cassette to establish the CHO-K1 RMCE I3 host cell line contains the synthetic CAGGS promoter comprising sequences from the CMV immediate early enhancer and chicken β-actin promoter (**Figure 13A**). The CAGGS promoter is placed upstream of the first FRT site (FRT3) to establish a promoter trap and drives expression of the reporter/selection fusion protein (GTN) consisting of the open reading frames of gfp, thymidine-kinase and neomycin-phosphotransferase. The RMCE cassette in-between the two heterospecific FRT sites (FRT3 and FRT) is then replaced by a vector encoding the kappa or lambda light chain, the rat elongation factor 1 alpha promoter (rEF1a) and the IgG1 heavy chain (**Figure 13B**). For a promoter trap, the light chain does not contain any promoter sequence on this plasmid vector and can only be transcribed after proper RMCE cassette exchange to place the light chain coding sequence downstream of the CAGGS promoter.

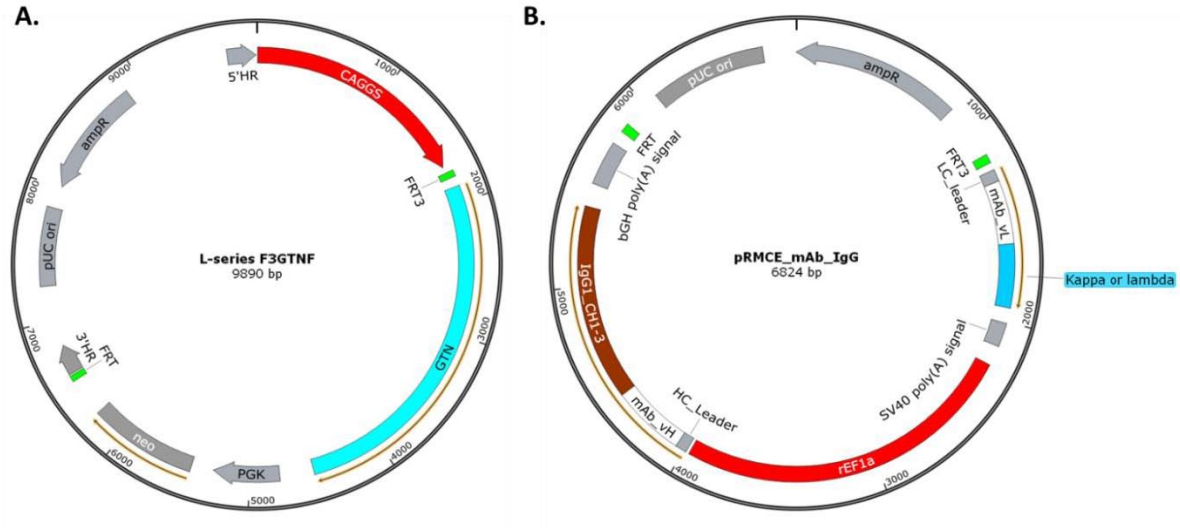


Figure 13. Plasmids to establish the parental CHO-K1 RMCE I3 host (A) or to exchange the RMCE cassette to yield IgG1 antibody producing daughter cells (B). Plasmid L-series F3GTNF contains a CAGGS promoter upstream of the first flipase-recognition target site (FRT) site to drive expression of the GTN fusion protein (GFP/thymidine-kinase/neomycin-phosphotransferase). Furthermore, this plasmid contains sequences for homology recombination (5' and 3' HR) and a neomycin/kanamycin resistance gene for potential use as recombination sites into bacterial artificial chromosome and screening of positive clones. Plasmid pRMCE_mAb_IgG was used to site-specifically replace the RMCE cassette with different antibody sequences in

full-length IgG1 format. In this plasmid the light chain does not contain any promoter sequence. Only after targeted gene integration by RMCE the light and heavy chain are transcribed by the CAGGS or rat elongation factor 1 alpha promoter (rEF1a), respectively.

Culture parameters such as specific productivity and growth rates were determined in exponential growth phase of batch cultures (**Table XI**). Additionally, we monitored intracellular product formation by flow cytometry, antibody specificity by ELISA or biolayer-interferometry and targeted gene integration by PCR. We could establish RMCE daughter clones for all antibody variants. No recombinant clones could be established for antibody 2F5 in a first experiment, therefore the data is not included yet. For Ustekinumab and 2G12 no impaired expression levels could be detected when compared to their germline configuration. In contrast, 4B3 gave a much lower specific productivity than its germline variant 136/63 (**Table XI**).

Table XI. Cell culture parameters of established antibody producing CHO-K1 RMCE I3 daughter clones.

antibody	μ_{max}	qP_{max}	cultivation
2G12 IgG 1B3	0,76	2,95	batch
2G12 IgG 1B11	0,68	3,43	batch
2G12 IgG 1E4	0,71	2,51	batch
353/11 IgG 2A12	0,73	1,97	batch
353/11 IgG 2C7	0,62	1,76	batch
353/11 IgG 2D7	0,73	2,28	batch
Ustek IgG H4	0,66	2,69	batch
Ustek IgG H12	0,69	2,24	batch
Ustek IgG H4	0,79	1,74	routine
Ustek IgG H12	0,78	2,16	routine
554/12 IgG 5F7	0,83	2,03	routine
554/12 IgG 3A11	0,68	2,43	routine
554/12 IgG 5H1	0,85	2,06	routine
4B3 IgG 4A9	0,71	0,6	batch
136/63 IgG E10	0,66	1,75	batch
136/63 IgG 3B6	0,76	1,8	batch
136/63 IgG 4C2	0,73	1,9	batch
236/14 IgG D7	0,98	0,27	batch
236/14 IgG 1B2	0,76	0,29	batch
236/14 IgG 4D9	0,78	0,34	batch
I3	0,72	0	batch
K1	0,78	0	batch

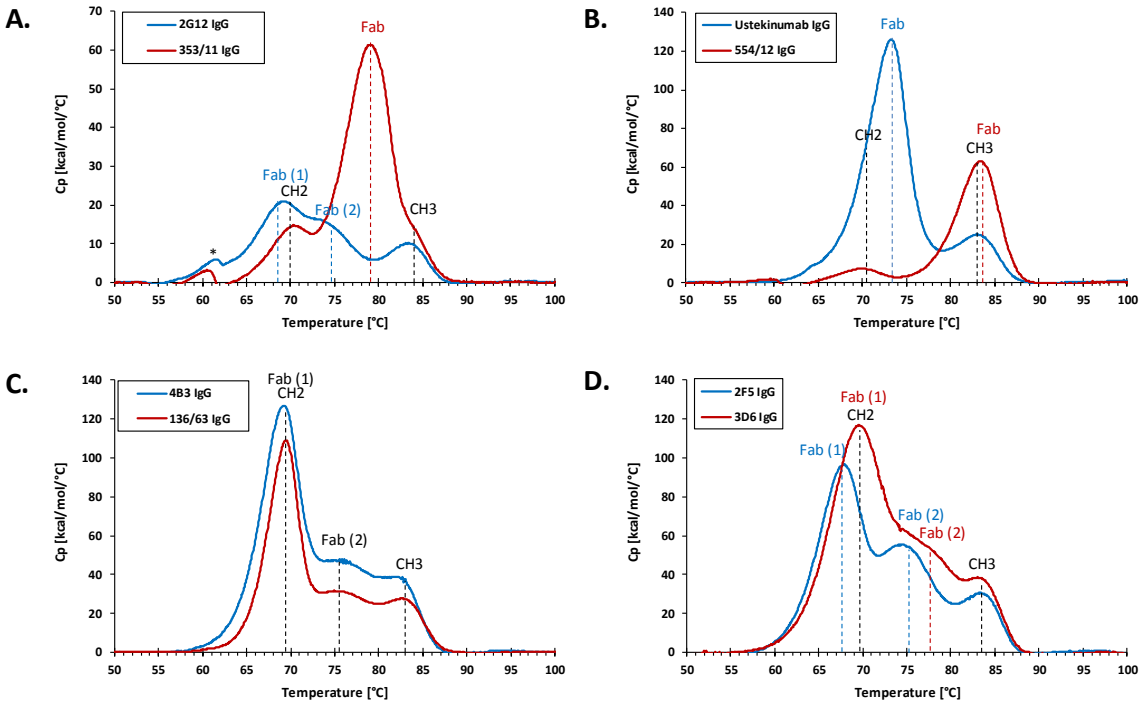
4.4 Antibody product characteristics

Differential scanning calorimetry (DSC) measures the excess heat capacity during an endothermic unfolding process of antibodies. Multiple peak patterns were reported for IgG antibodies. One key report was published by Garber and Demarest (Garber and Demarest, 2007) analyzing 17 different

antibodies derived from human or humanized origin. Consistent with other publications (Ionescu et al., 2008; Mueller et al., 2013), the Fc part of human IgG1 shows two endothermic thermal transitions derived from the CH2, with a lower thermal stability of around 71°C, and from the highly stable CH3 domain, showing a Tm midpoint of 83°C. Although minor variabilities for the transition temperature of the CH2 domain of the Fc part might be observed, caused by differences in CH2-glycosylation pattern (Ionescu et al., 2008), usually CH2 and CH3 transitions of IgG1 remain constant and superimpose well on the thermogram of different antibodies of the same isotype. The Fab domain most often shows a single (cooperative) transition between 57-82°C (Garber and Demarest, 2007), but also double transitions are possible, indicating non-cooperative unfolding of the Fab domain. The Tm of the cooperative Fab unfolding might coincide with the CH2 or CH3 unfolding temperature resulting in a total of only two endothermic unfolding transitions for the full-length IgG molecule. Atlernatively, the Fab domain might also show a distinct cooperative unfolding temperature different from CH2 and CH3 resulting in a three peak thermogram profile for a full length IgG1 antibody. Garber and Demarest (2007) also observed that the cooperative unfolding peak maximum ($C_{P,max}$) for Fab is usually three times higher than for the CH2 or CH3 domain. Such a pattern could be observed for antibody 353/11 IgG (**Figure 14A, red**).

Using this knowledge from literature the individual domains constituting a full-length human IgG1 molecule were assigned to the peaks obtained from DSC measurements from different CHO expressed mAbs (**Figure 14**). Peaks corresponding to the CH2 domain (Tm 67.5-71.6°C) and CH3 domain (Tm 82.5-84.9) were identified. Antibodies 2G12 and its germline variant 353/11 (**Figure 14A**) both showed three endothermic unfolding transitions. For 353/11 the Fab peak is approximately three times higher than the $C_{P,max}$ of CH2 and CH3, indicating a single cooperative unfolding transition for the 353/11 Fab fragment. In contrast, the peak assigned to CH2 in 2G12 showed the highest $C_{P,max}$ and only a minor height for the middle, Fab-assigned peak. Therefore, it can be concluded that the 2G12 Fab fragment unfolds in two transitions in a non-cooperative manner with the first “Fab (1)” transition overlapping with the CH2 signal. More strikingly, the single 353/11 Fab transition showed a higher thermostability (79°C) than both of the 2G12 Fab transitions (68.7°C and 74.6°C), highlighting an overall higher stability for the isotype-matched germline variant 353/11 compared to the mature antibody 2G12.
Ustekinumab and its germline variant 554/12 **(Figure 14B**) both showed only two transition peaks. However, asymmetric peaks indicate that a single Fab transition overlaps with the CH2 or CH3 domain in Ustekinumab or 554/12, respectively. This represents a total increase of 10.7°C for the Fab thermostability for conversion of Ustekinumab (73.4°C) to its germline variant 554/12 (84.1°C). For 2G12 and Ustekinumab an improvement in thermostability of the Fab domain could be demonstrated by removing the somatic mutations resulting in variants 353/11 and 554/12, respectively. Interestingly, a different behavior was observed for 4B3 and its germline variant 136/63 (**Figure 14C**). Both antibody variants showed highly similar DSC thermogram profiles indicating similar stabilities under the given conditions, with two unfolding transitions for the Fab domain indicating non-cooperative unfolding behavior.

Low germinality mAb 2F5 and high germinality mAb 3D6 expressed as scFv-Fc variants in DUKX-B11 RMCE host cells also showed a correlation of higher Fab thermostability with higher germline identity (3D6) during DSC measurements in IgG1 format (**Figure 14D**). The difference in Fab thermostabilities between 2F5 and 3D6 were not as pronounced as for 2G12 or Ustekinumab to their respective germline variant, but a difference of 1.9°C for Fab(1) and 2.3°C for Fab(2) demonstrates higher 3D6 Fab stabilities (**Figure 14E**).



E.

	V^{HC}/V^{LC} germline identity	CH2	CH3	Fab	number of Fab transitions
2G12 IgG	68%/84%	68.7°C	83.3°C	68.7°C & 74.6°C	2
353/11 IgG	100%/100%	70.1°C	84.9°C	79°C	1
Ustek. IgG	88%/99%	71.6°C	82.8°C	73.4°C	1
554/12 IgG	100%/100%	70.6°C	84.1°C	84.1°C	1
4B3 IgG	83%/90%	68.9°C	82.5°C	68.9°C & 76.9°C	2
136/63 IgG	100%/100%	69.3°C	82.8°C	69.3°C & 76.6°C	2
2F5 IgG	85%/85%	67.5°C	83.4°C	67.5°C & 75.4°C	2
236/14 IgG	100%/100%	n.d.	n.d.	n.d.	n.d.
3D6 IgG	96%/98%	69.4°C	83.8°C	69.4°C & 77.7°C	2

low germinality variants
high germinality variants

Figure 14. DSC thermograms of selected mature antibodies and their germline variants. Peaks were assigned to individual IgG1 domains. Measurements were performed in 30 mM phosphate buffer (150 mM NaCl) at pH 6.0 with a scanning rate of 1°C / min. The thermogram was deconvoluted into separate peaks by using sequential two-state transitions. Mature and germline antibody variants were compared for 2G12 and 353/11 (**A**), Ustekinumab and 554/12 (**B**) and 4B3 and 136/63 (**C**). The low germinality antibody 2F5 was compared to high germinality antibody 3D6 (**D**). V-gene identity in heavy chain (V^{HC}) or light chain (V^{LC}) and DSC peaks assigned to CH2, CH3 or Fab domain are summarized (**E**).

Figure 14E summarizes the effects on thermostability by mutating selected positions of the mature antibody variable sequence to the respective germline residue. For 2G12 and Ustekinumab the

thermostability of the Fab domain is clearly increased. Additionally, the unfolding behavior changed from two thermal transition peaks to a single cooperative transition peak of the Fab domain of 2G12 by introducing germline residues. This observation might be attributed to the special feature of antibody 2G12 developing a I-shaped conformation by a unique V_H domain-swap, instead of the conventional Y-shaped conformation of other IgG1 antibodies (Calarese et al., 2003).

To sum up, with DSC experiments we could observe an increase in thermostability for the germline variants of antibody 2G12 and Ustekinumab, whereas no significant improvement was observed for the germline variant of 4B3. Also by introducing germline mutations the behavior in cooperative Fab unfolding changed. The Fab region unfolds in a non-cooperative manner for IgG1 antibody 2G12, 4B3, 136/63 and 2F5. Therefore two Fab transition peaks could be observed for these antibodies. In contrast, a cooperative unfolding behavior (single transition peak) for IgG1 antibody 353/11, Ustekinumab and 554/12 showing a single Fab transition peak can be proposed based on the current DSC results.

4.5 Proteomics and metabolomics for investigation of cell biology

The first example of how the antibody product itself influences cell culture parameters and cellular properties was demonstrated with the antibody pair 2F5 and 3D6 in scFv-Fc format using different vector strategies (Mader et al., 2012). Subclones generated by RMCE in DUKX-B11 F3/F, showing a two-fold higher specific expression for 3D6, were compared to traditional clone development strategies using plasmids or novel strategies using bacterial artificial chromosomes (Sommeregger et al., 2016) in label-free LC-MS proteomic studies **Figure 15**.

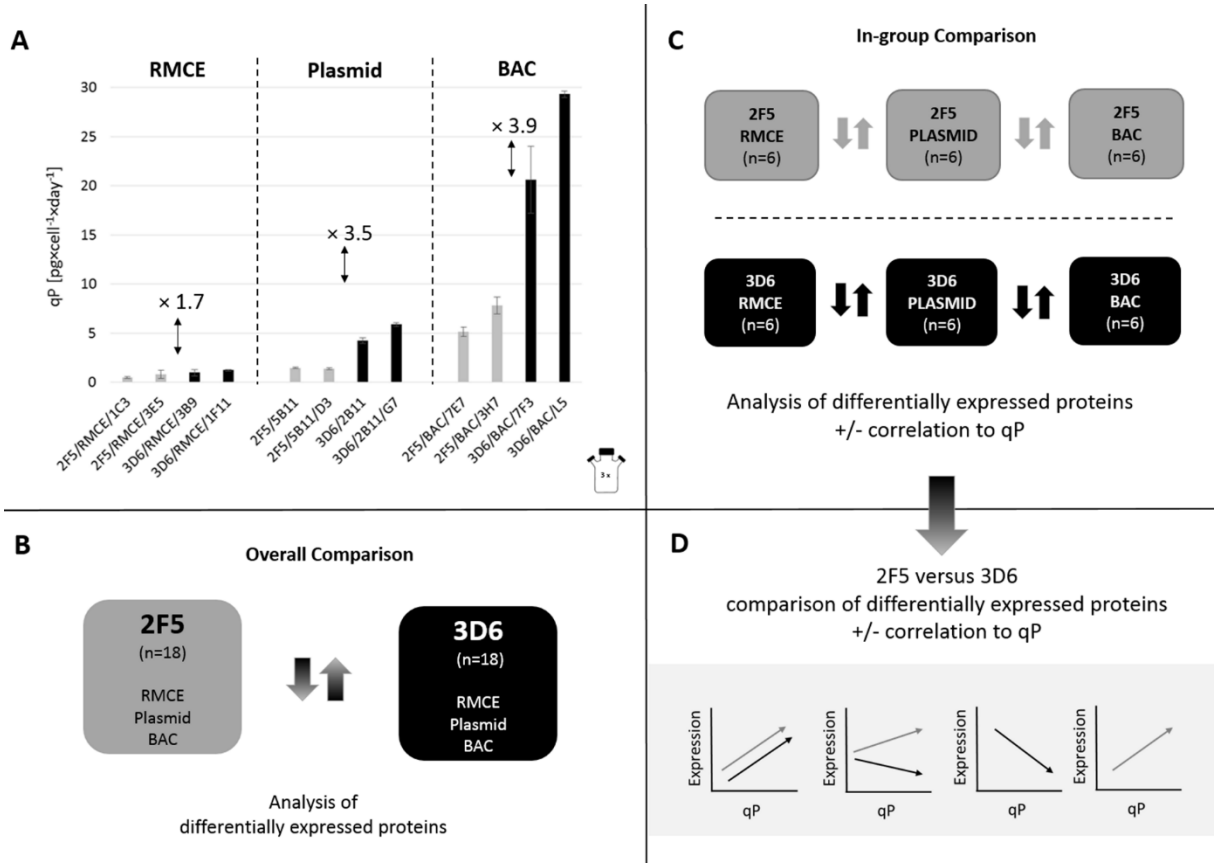


Figure 15. Comparison of proteomic differences of CHO cell lines expressing antibody 2F5 or 3D6 in scFv-Fc format. Expression levels of 3D6 were always higher compared to 2F5 using three different vector strategies (**A**). Data generated from label-free LC-MS-MS analyses were used to compare all 2F5 samples to all 3D6 samples (**B**) or to compare differences for each antibody with different vector strategies (**C**). The data was integrated to find differences in endogenous proteins induced by the antibody sequence or proteomic difference correlated to specific productivity qP (**D**). Figure reprinted from Sommeregger et al. (2016).

Bottlenecks for 2F5 expression were linked to enhanced cellular stress connected with lower levels of Sec231p. The observed effect caused by the 2F5 antibody variable sequence might partially be attributed to individual charge distribution on the CDR H3 loop structure of 2F5. We could also demonstrate membrane association for 2F5 and interaction with liposomes using biolayer-interferometry (unpublished).

The 2F5scFv-Fc and 3D6scFv-Fc DUKX-B11 RMCE clones were analyzed for proteomic changes by label-free LC-MS-MS and published under co-authorship:

> Sommeregger, W., **Mayrhofer, P.**, Steinfellner, W., Reinhart, D., Henry, M., Clynes, M., Meleady, P., and Kunert, R. (2016). Proteomic differences in recombinant CHO cells producing two similar antibody fragments. Biotechnol. Bioeng. doi:10.1002/bit.25957.

In a second example we investigated the ultimate physiological state of the cell (Quek et al., 2010) with a metabolomics approach. Cellular physiology is the result of the combined effects of the

genome, the transcriptome and the proteome and metabolomics analysis allows us to have a deeper look into how the metabolism of a culture changes with different growth phases, media, and expression levels and how to optimally engineer processes and cells for our needs (**Table XII**).

Table XII. Applications of metabolomics studies for CHO cells reported in literature (examples).

Application	Reference
Growth characteristics in different media	Dietmair et al., 2012
Metabolic engineering of TCA cycle	Chong et al., 2010
Characterizing high mAb producing CHO cell lines	Chong et al., 2012
Production-to-consumption lactate shift	Luo et al., 2012
Designing feeding strategies	Ma et al., 2009; Sellick et al., 2011

To investigate whether changes on the metabolome level occur by expression of antibody 136/63 IgG in the CHO-K1 RMCE I3 host we analyzed the metabolic rate of carbohydrates and amino acids at different cultivation phases. Isotopically non-stationary metabolic flux analysis (INST-MFA) was applied using parallel labeling experiments (Antoniewicz, 2013) with [1,2-^{13}C]glucose and [U-^{13}C]glutamine as metabolic tracers suitable for determining fluxes in glycolysis/pentose phosphate pathway (PPP) or tricarboxylic acid cycle (TCA), respectively (Metallo et al., 2009). INST-MFA does not require isotopic steady state with the advantage of reducing required labeling time and the amount of expensive isotopically labeled substances. Compared to bacterial systems label incorporation is slow in mammalian cell cultures and the metabolic steady state might change at longer culture times. Therefore INST-MFA based on metabolic steady state and isotopic non-steady state is the only suitable and accepted method for many cell culture processes.

The two cell lines (IgG 136/63 producing clone 4C2 and the CHO-K1 host cell line) were cultivated in a fed-batch experiment with glucose and glutamine fed on day 2 and 3 (1 g/L glucose, 2 mM L-glutamine) (**Figure 16**). The exponential phase of 4C2 (μ>0.6; day 2 after inoculation) was compared with the stationary phase (μ<0.1; day 4 after inoculation) and the stationary phase (day 4 after inoculation) of CHO-K1. To include replicate cultures, individual shaker tubes were incubated until the first labeling experiment started at day 2 (exponential phase) and are depicted as blue dots in **Figure 16**. Before initiation of ^{13}C-labeling, replicate cultures were pooled (red dots) and split into replicate cultures for stationary phase (green dots), no-label control (black dots), glucose labeling (dark red dots) and glutamine labeling (yellow dots). Before labeling at day 4, replicate stationary cultures were pooled comparable to the exponential phase labeling (red dots) and split for no-label, glucose label and glutamine label cultures. For the labeling experiments the glucose and glutamine was adjusted to 75% with [1,2-^{13}C]Glucose or [U-^{13}C]Glutamine, respectively. The glucose label cultures were fed with unlabeled glutamine and [1,2-^{13}C]Glucose to reach 75% tracer enrichment in

the media and vice versa for the glutamine label cultures. No-label controls were included without addition of any tracer but addition of glucose and glutamine to the same concentration as the label cultures. 136/63 IgG titers in the culture supernatant were measured by biolayer-interferometry (Octet) to calculate the specific productivity (qP) of RMCE clone 136/63 IgG 4C2 (**Figure 16C**).

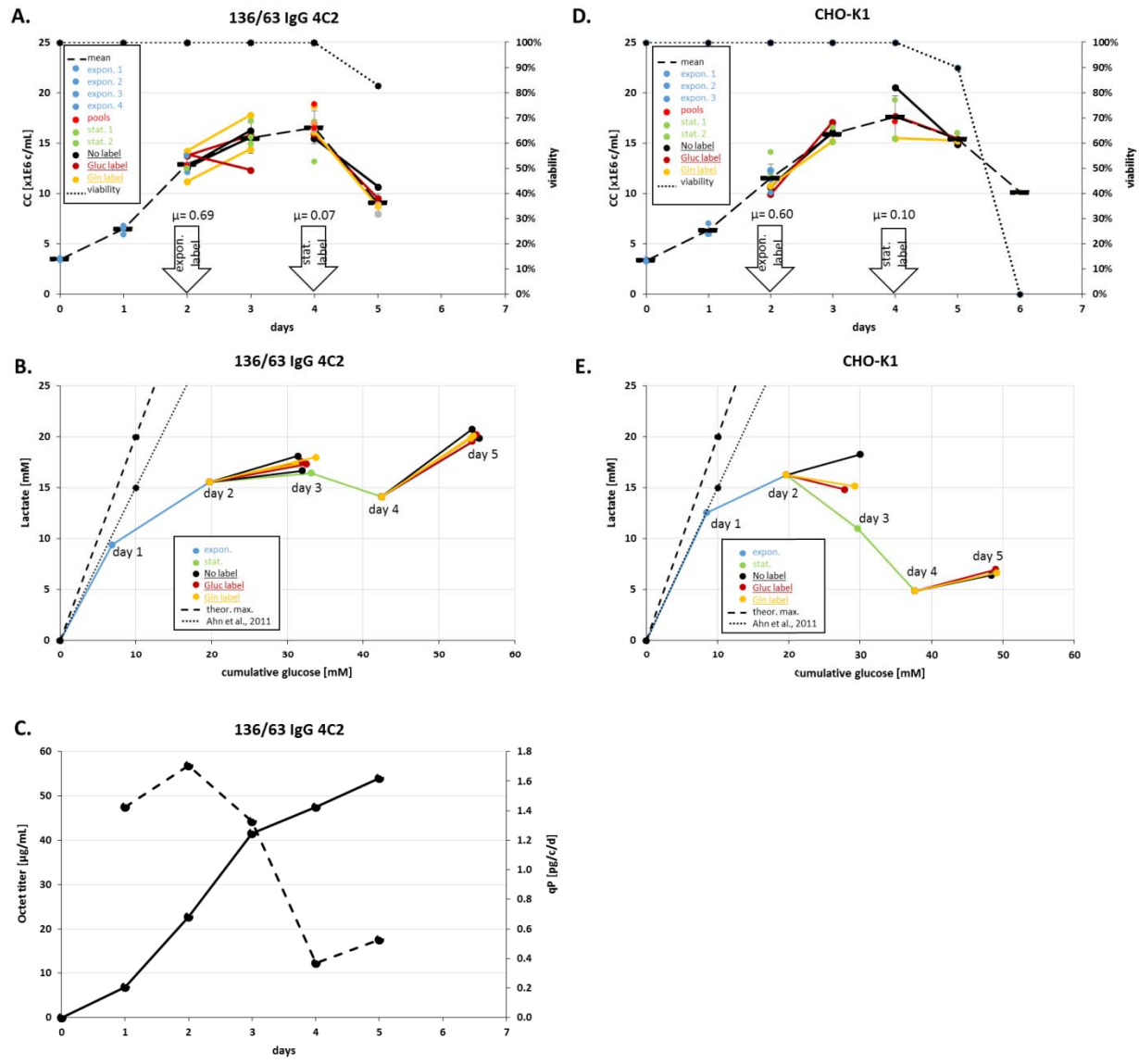


Figure 16. Labeling experiment of antibody producing RMCE cell clones 136/63 IgG 4C2 (A-C) in comparison to CHO-K1 host cell line (D-E). Growth (**A**, **D**) was monitored for a culture operated in fed-batch mode by feeding 1 g/L glucose and 2 mM L-glutamine on day 2 and 3. Duplicate exponential cultures are indicated as blue dots. At day 2 and 4 replicate cultures were pooled (red dots) to initiate the labeling cultures. Replicate stationary cultures are depicted as green dots. Duplicates of labeling cultures are indicated as black, brown and yellow dots/lines at day 2 and 4.

Growth behavior of CHO-K1 and 4C2 were similar reaching maximal peak cell densities of 18 × 10^6 c/mL on day 4 (**Figure 16A, D**) and a maximum specific productivity (qp) of 1.7 pg/c/d for cell line 4C2 (**Figure 16C**). Amino acid consumption/production rates were highly similar (data not shown).

Labeling experiments started at day 2 or 4 by splitting the fed-batch culture and adding [1,2-^{13}C]Glucose or [U-^{13}C]Glutamine label to reach at least 75% enrichment in the medium after pooling the replicate cultures (red dots). Black lines indicate no label, brown lines indicate glucose label and yellow lines indicate glutamine label. One obvious difference between the two cell lines was the lactate production profile. In **Figure 16B** and **E** lactate concentration is plotted against totally consumed cumulative glucose and compared to the theoretically possible lactate formation per molecule glucose (dashed line) or values reported in Ahn and Antoniewicz (2011) (dotted line) for CHO cultures. Both cell lines reached a maximal lactate concentration of 17 mM at day 2 and showed a produced lactate per consumed glucose ratio of 1.5 at day 1 in agreement with reports published by Ahn and Antoniewicz (2011). The maximal possible yield would be two molecules of lactate generated from one glucose molecule. A lower ratio indicates that in the exponential phase a significant portion of the glucose flux is not metabolized to lactate but directed into the TCA cycle for effective ATP production. At day 2 of the experiment a switch from lactate production to –consumption was observed for the CHO-K1 host cell line (**Figure 16E)**. This switch was delayed by one day for the antibody producing cell line 4C2 (**Figure 16B)**. At day 2 lactate production was already significantly reduced in agreement with literature reporting a switch from lactate production to consumption at the transition from peak growth to peak antibody production (Templeton et al., 2013). Interestingly, a re-activation of lactate production was observed (**Figure 16B** and **E**) when glucose and L-glutamine were added at high concentrations (**Table XIII**). The glucose addition was necessary to reach about 75% labeled glucose in the medium following the experimental design of INST-MFA. For both cell lines this effect was more pronounced at day 4 (**Figure 16B and C)**.

Table XIII. Final total glucose and L-glutamine concentrations of label cultures to reach at least 75% ^{13}C-Isotopomer enrichment of the tracers in the culture media.

		Final glucose	Final L-glutamine
136/63 IgG 4C2	Labeling day 2:	10.1 g/L = 56.1 mM	4.3 mM
	Labeling day 4:	4.0 g/L = 22.2 mM	3.4 mM
CHO-K1	Labeling day 2:	10.1 g/L = 56.1 mM	6.8 mM
	Labeling day 4:	3.8 g/L = 21.1 mM	4.5 mM

^{13}C-Isotopomer enrichment of intracellular lactate, amino acids and glycolytic-(phosphoenolpyruvate, 3-phosphoglycerate) or tricarboxylic acid cycle (TCA) intermediates (citrate, α-ketoglutarate, malate, fumarate, succinate) were measured by GC-MS and the combined data,

together with extracellular amino acid consumption/production rates, were integrated into a metabolic network model modified from Carinhas et al. (2016) (**Figure 17**) using the Matlab tool "INCA" to calculate intracellular fluxes and their confidence intervals (**Figure 18**).

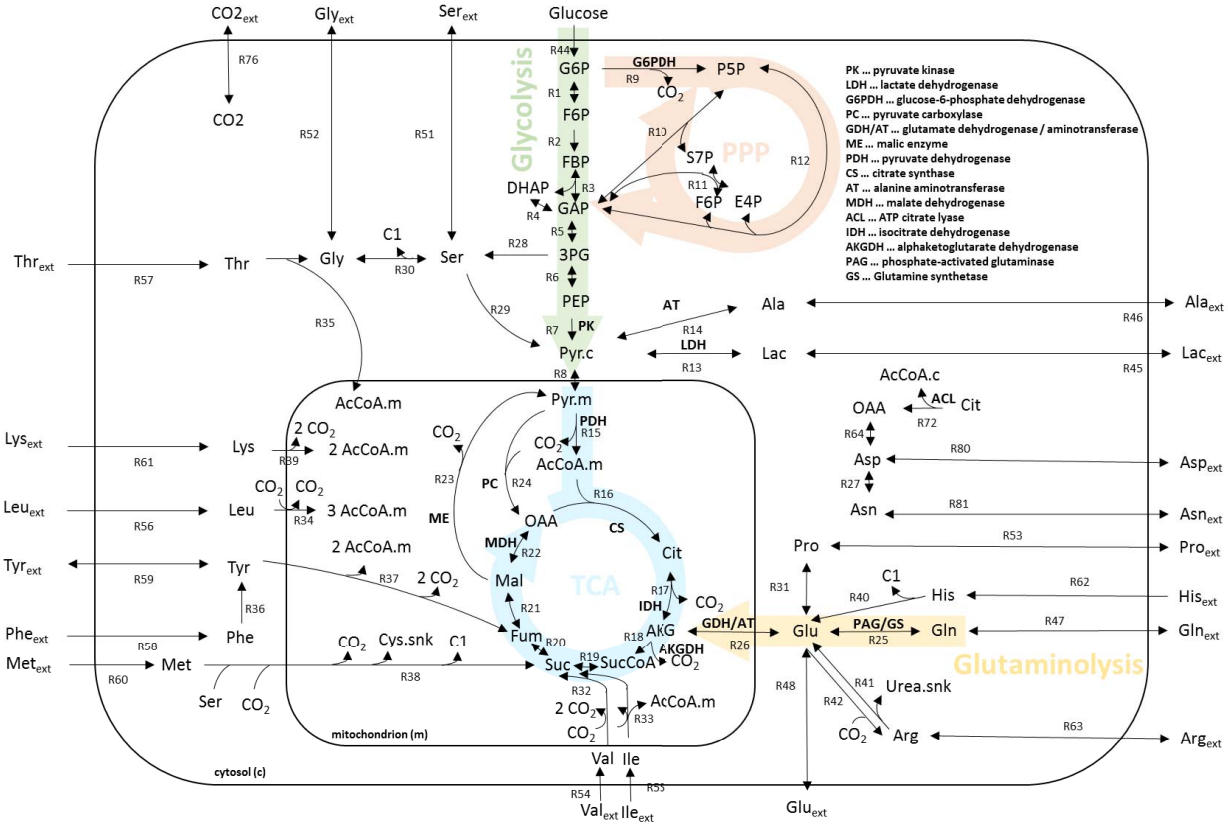


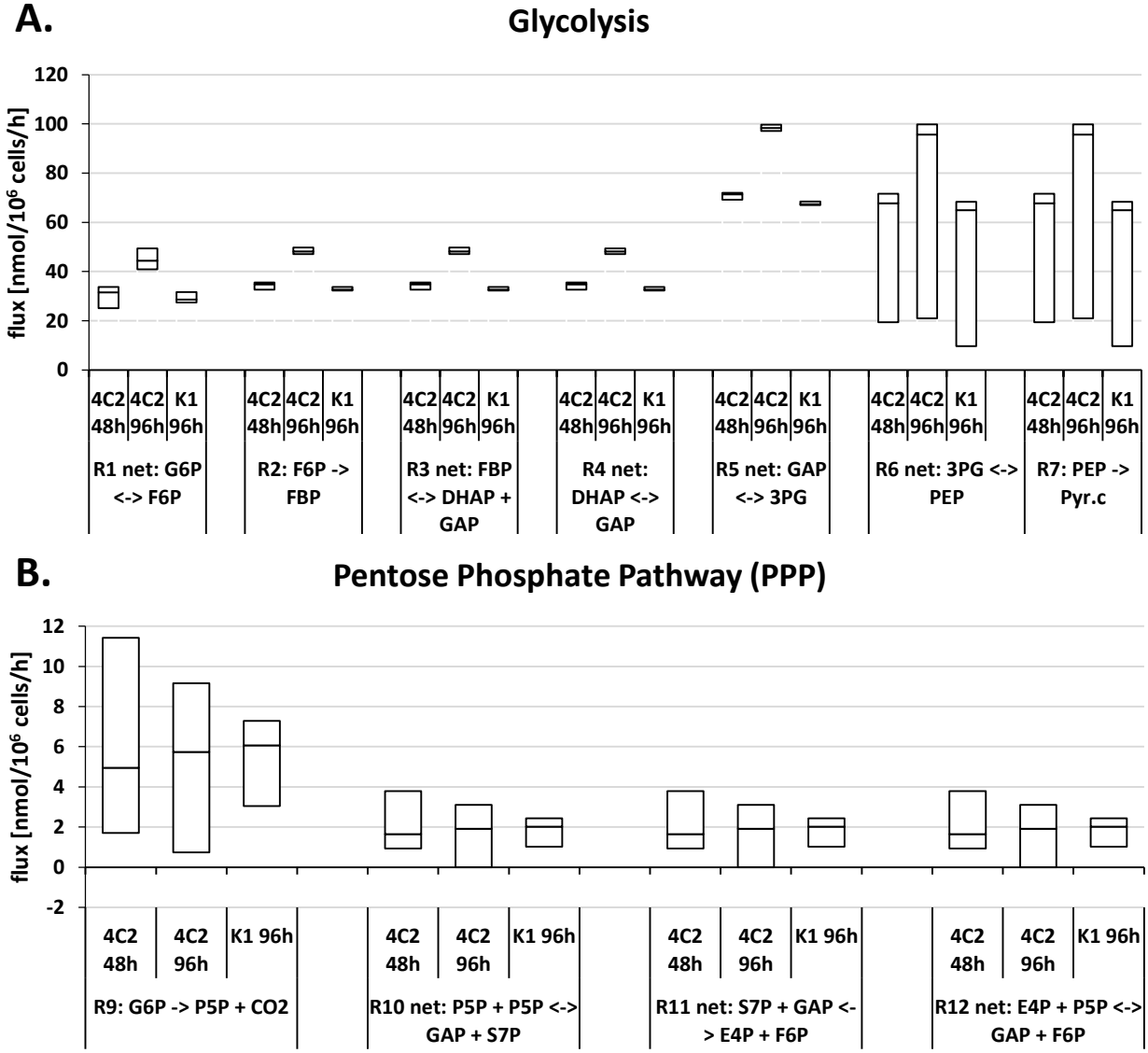
Figure 17. Metabolic network model used for isotopically non-stationary ^{13}C metabolic flux analysis (INST-MFA) integrated into the "INCA"-software.

By comparing the two culture stages (day 2 and 4) of the antibody-producing cell line we could observe an increased glycolytic flux for 4C2 in the labeling experiment at day 4 (**Figure 18A**). Also, an increase in lactate production flux was observed when labeled glucose and glutamine were added at four-day cultures compared to the labeling experiment at two days (**Figure 18C**). No significant difference could be observed for the flux into (reaction R9) or back from (reactions R10, R11 and R12) the oxidative pentose phosphate pathway (PPP) (**Figure 18B**).

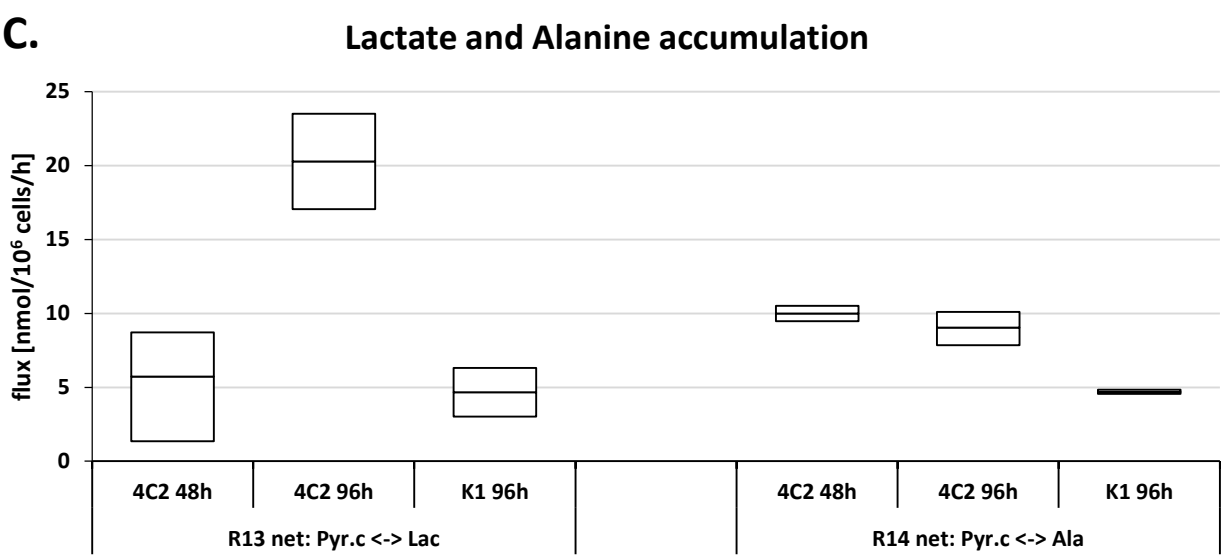
Intracellular fluxes of the TCA-cycle were decreased at later culture stages, but similar for 4C2 and K1 (**Figure 18D-E**). There was a significant reduction in the activity of malic enzyme (reaction R23) during culture time of cell line 4C2.

Amino acid transport rates were generally downregulated at day 4 (**Figure 18N-Q**) but similar for the two cell lines. For aspartate (reaction R80), asparagine (reaction R81), glutamate (R48) and proline (reaction R53) a switch from consumption to secretion was observed.

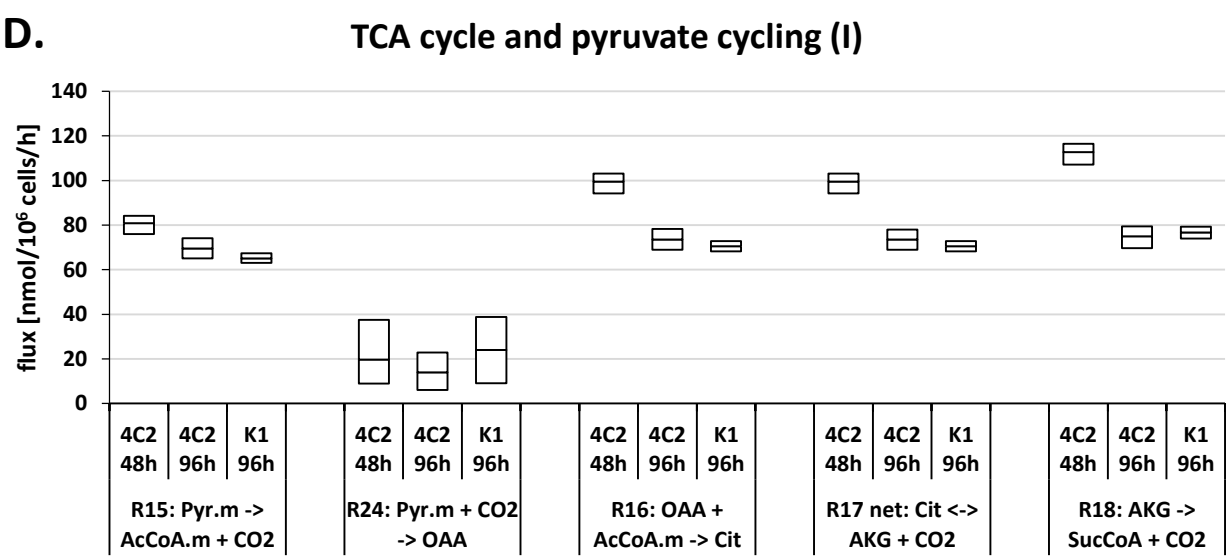
From the metabolomics studies, investigating intracellular fluxes in detail, it can be concluded that under the present expression levels of antibody 136/63 IgG no major metabolic changes were induced in comparison to the CHO-K1 host. Therefore, we can assume that no major differences were induced during establishment and clone selection of parental CHO-K1 RMCE I3 host. 136/63 IgG producing subclones showed good growth characteristics comparable to the wild-type CHO-K1 host cell line. Whether the profile change in lactate and alanine production/consumption is a property induced by the specific antibody sequence has to be determined in more detail by comparison to the CHO-K1 RMCE I3 host cell line and 4B3 IgG producing daughter clones. It is also possible that this property developed from metabolic changes during clone selection of I3 or subsequently during selection of antibody producing daughter clone 4C2. In case the lactate/alanine behavior is induced by the specific antibody sequence, it should be possible to reverse this effect by exchanging the antibody cassette by the original selection/reporter cassette.



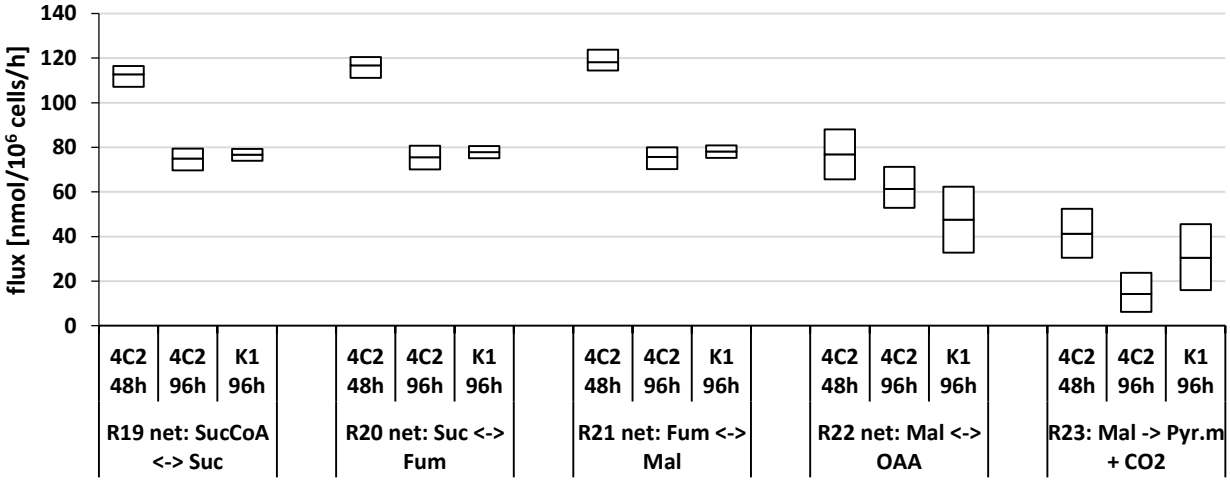
C.


Lactate and Alanine accumulation
flux [nmol/10^6 cells/h]
25
20
15
10
5
0
4C2 48h
4C2 96h
K1 96h
R13 net: Pyr.c <-> Lac
4C2 48h
4C2 96h
K1 96h
R14 net: Pyr.c <-> Ala

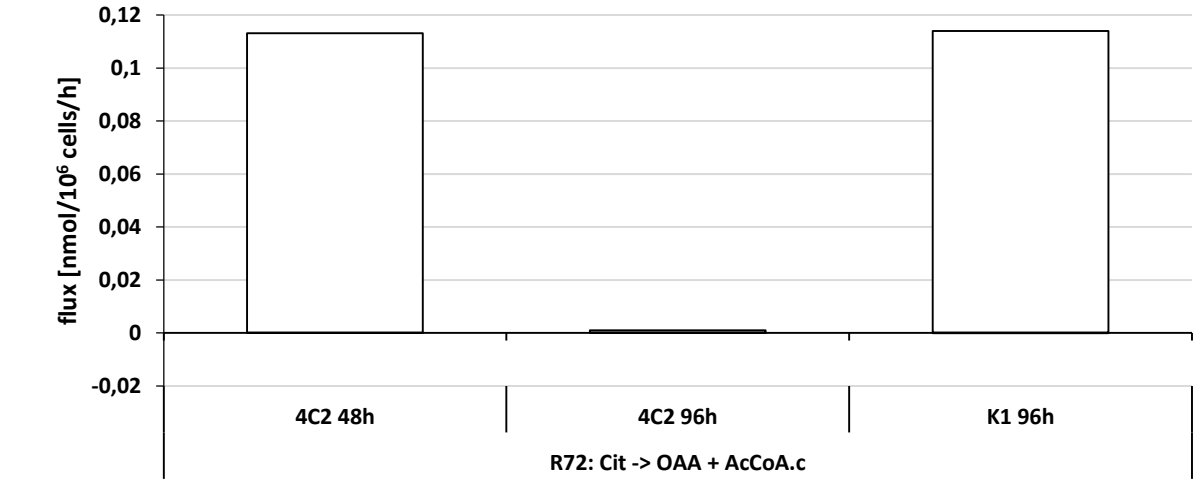
D.


TCA cycle and pyruvate cycling (I)
flux [nmol/10^6 cells/h]
140
120
100
80
60
40
20
0
4C2 48h
4C2 96h
K1 96h
R15: Pyr.m -> AcCoA.m + CO2
R24: Pyr.m + CO2 -> OAA
R16: OAA + AcCoA.m -> Cit
R17 net: Cit <-> AKG + CO2
R18: AKG -> SucCoA + CO2

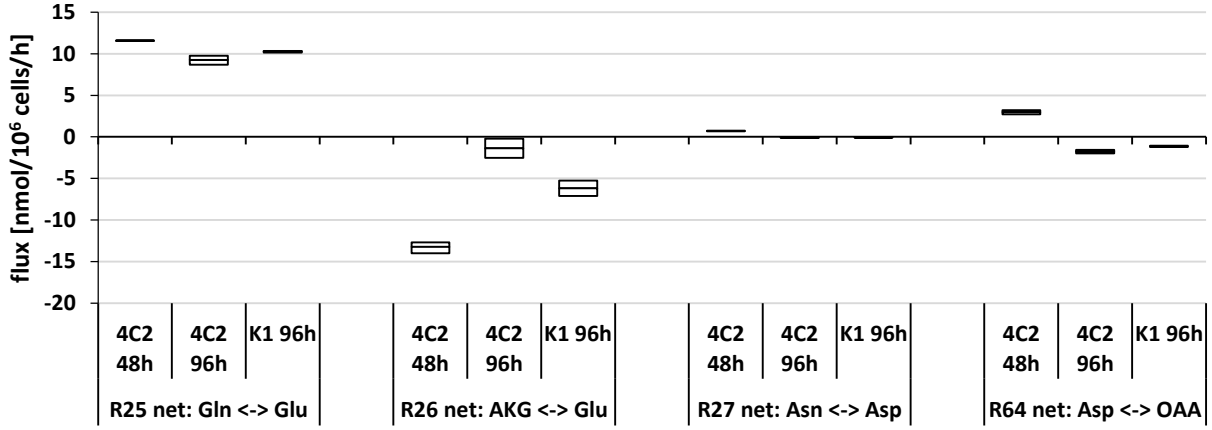
E.


TCA cycle and pyruvate cycling (II)
flux [nmol/10^6 cells/h]
140
120
100
80
60
40
20
0
4C2 48h
4C2 96h
K1 96h
R19 net: SucCoA <-> Suc
R20 net: Suc <-> Fum
R21 net: Fum <-> Mal
R22 net: Mal <-> OAA
R23: Mal -> Pyr.m + CO2

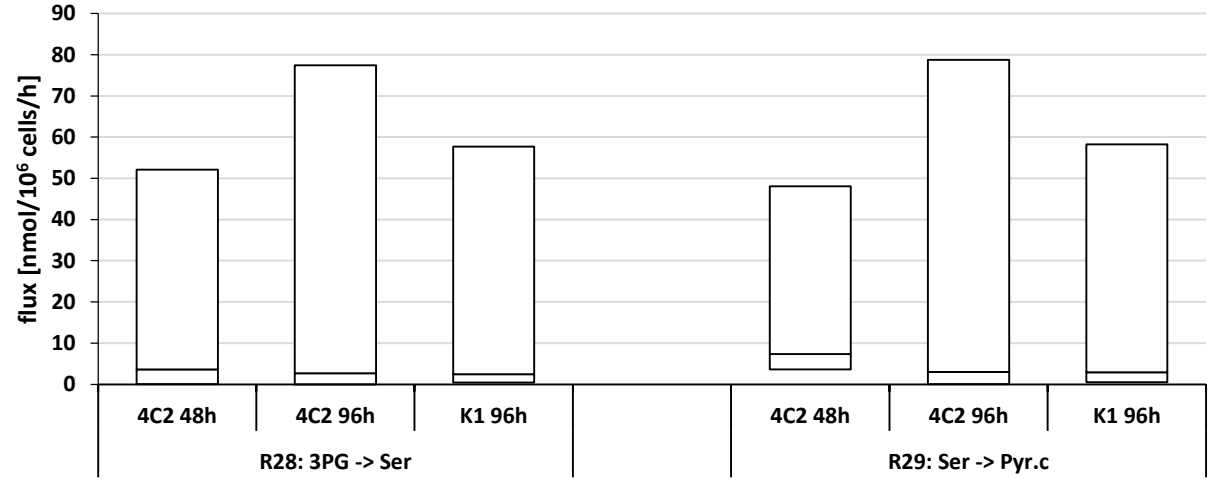
F.

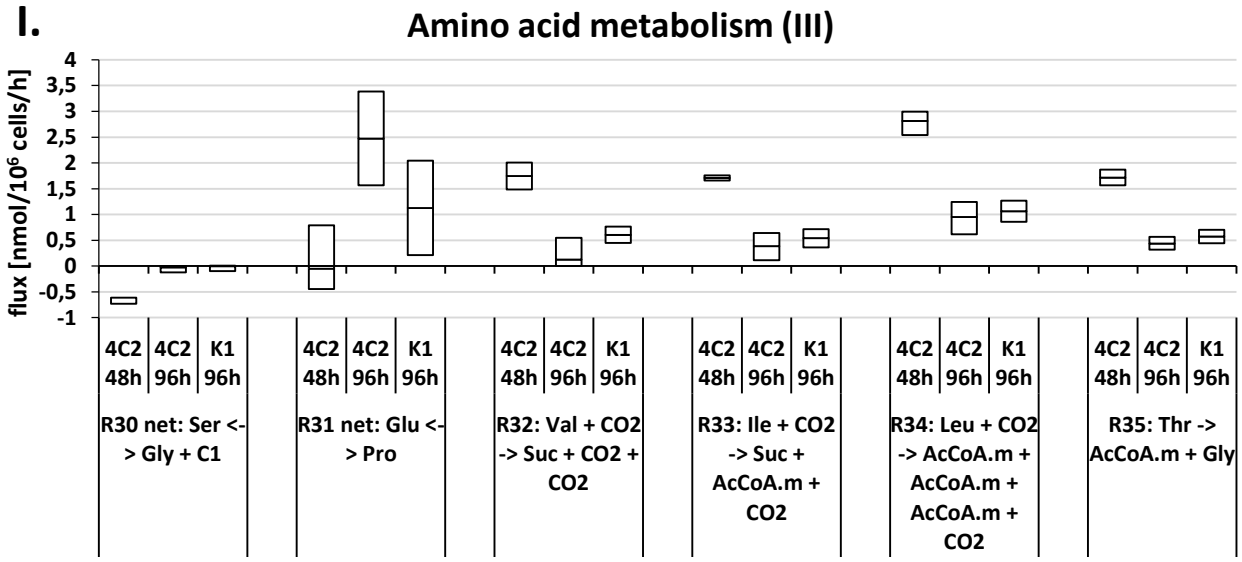

Lipid precursor generation
flux [nmol/10⁶ cells/h]
0,12
0,1
0,08
0,06
0,04
0,02
0
-0,02
4C2 48h
4C2 96h
K1 96h
R72: Cit -> OAA + AcCoA.c

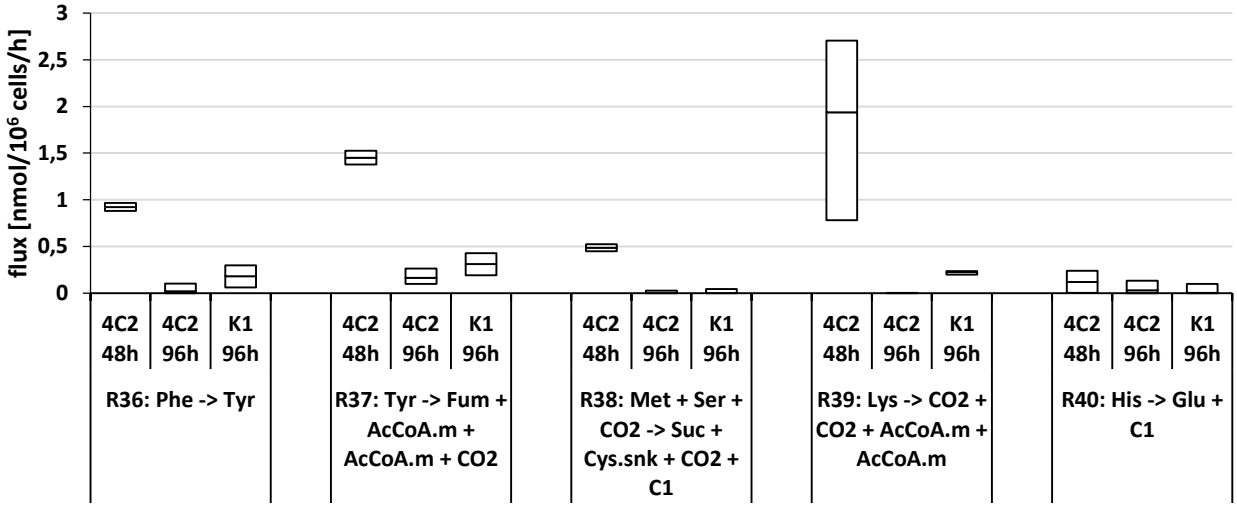
G.

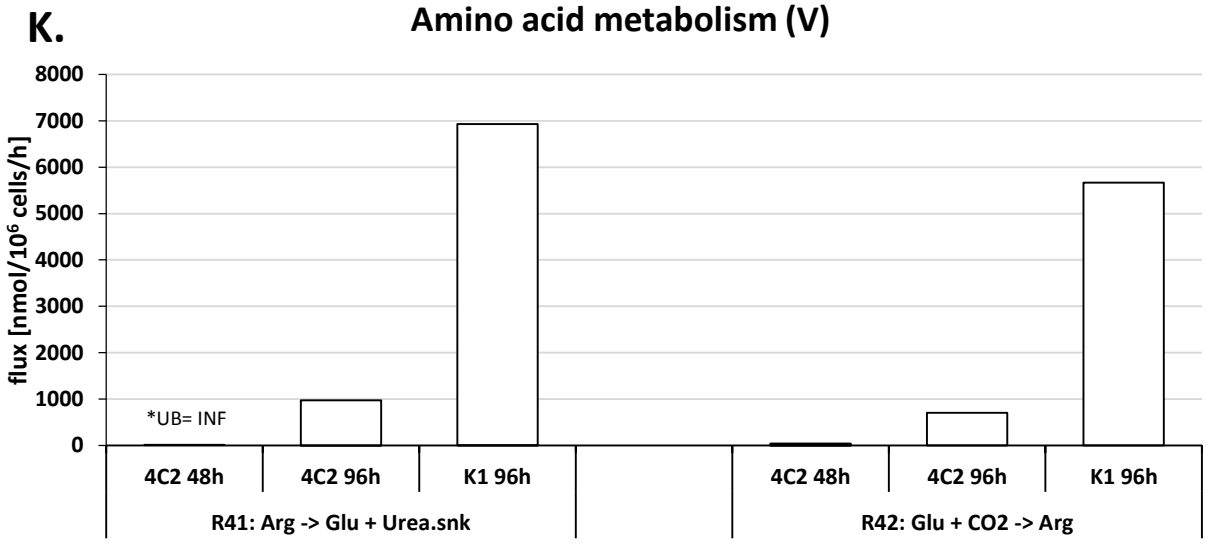

Amino acid metabolism (I)
flux [nmol/10⁶ cells/h]
15
10
5
0
-5
-10
-15
-20
4C2 48h
4C2 96h
K1 96h
R25 net: Gln <-> Glu
R26 net: AKG <-> Glu
R27 net: Asn <-> Asp
R64 net: Asp <-> OAA

H.

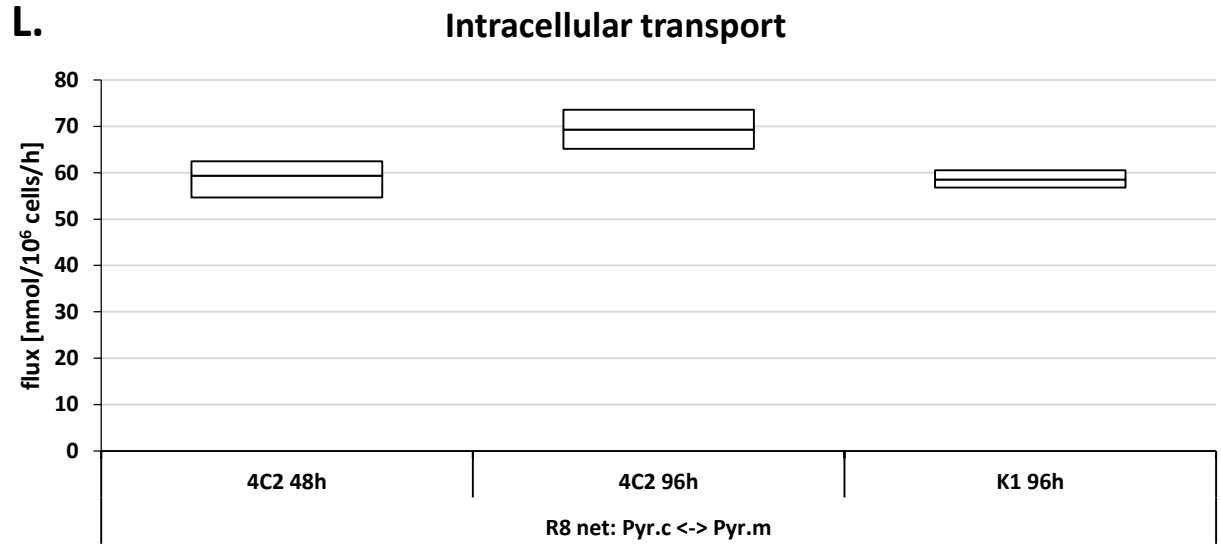

Amino acid metabolism (II)
flux [nmol/10⁶ cells/h]
90
80
70
60
50
40
30
20
10
0
4C2 48h
4C2 96h
K1 96h
R28: 3PG -> Ser
R29: Ser -> Pyr.c


I.
Amino acid metabolism (III)
flux [nmol/10^6 cells/h]
4
3,5
3
2,5
2
1,5
1
0,5
0
-0,5
-1
4C2 48h
4C2 96h
K1 96h
R30 net: Ser <-> Gly + C1
R31 net: Glu <-> Pro
R32: Val + CO2 -> Suc + CO2 + CO2
R33: Ile + CO2 -> Suc + AcCoA.m + CO2
R34: Leu + CO2 -> AcCoA.m + AcCoA.m + AcCoA.m + CO2
R35: Thr -> AcCoA.m + Gly

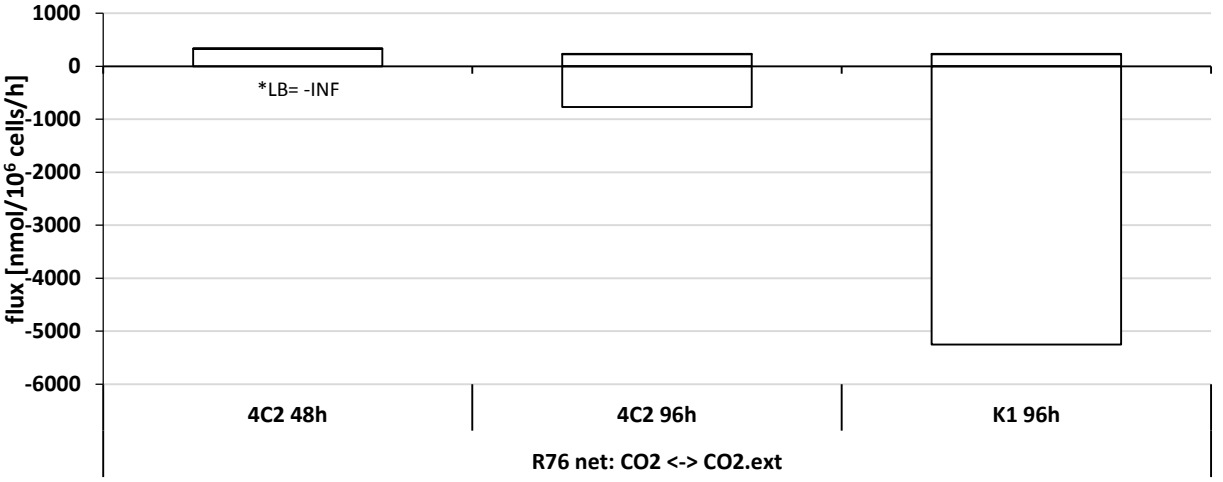

J.
Amino acid metabolism (IV)
flux [nmol/10^6 cells/h]
3
2,5
2
1,5
1
0,5
0
4C2 48h
4C2 96h
K1 96h
R36: Phe -> Tyr
R37: Tyr -> Fum + AcCoA.m + AcCoA.m + CO2
R38: Met + Ser + CO2 -> Suc + Cys.snk + CO2 + C1
R39: Lys -> CO2 + CO2 + AcCoA.m + AcCoA.m
R40: His -> Glu + C1


K.
Amino acid metabolism (V)
flux [nmol/10^6 cells/h]
8000
7000
6000
5000
4000
3000
2000
1000
0
*UB= INF
4C2 48h
4C2 96h
K1 96h
R41: Arg -> Glu + Urea.snk
R42: Glu + CO2 -> Arg

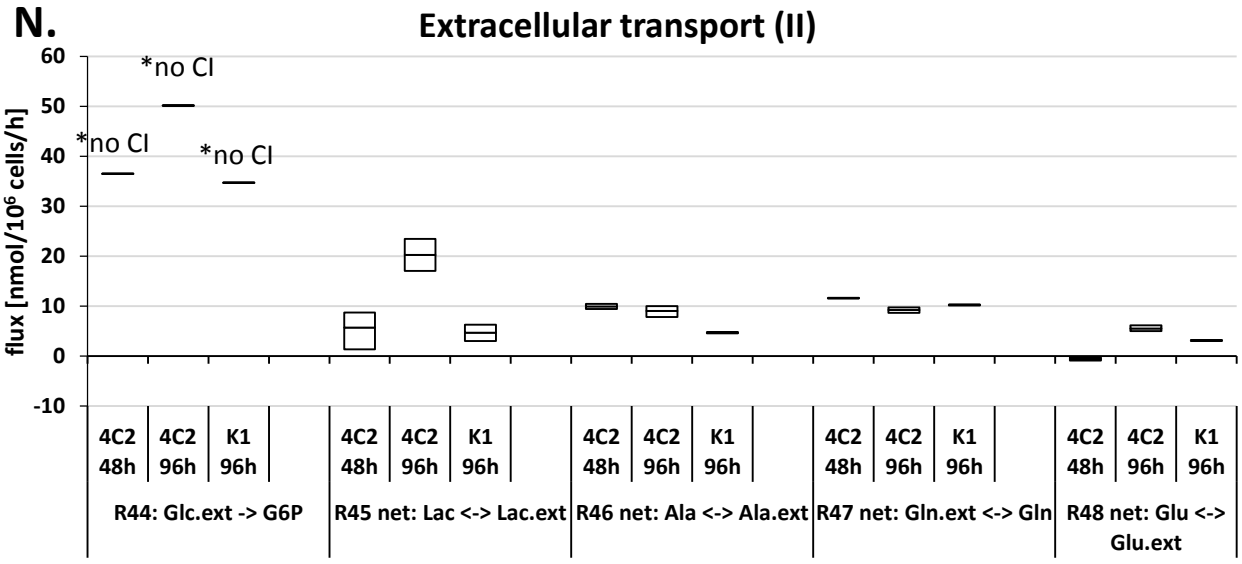
L.


Intracellular transport
flux [nmol/10^6 cells/h]
80
70
60
50
40
30
20
10
0
4C2 48h
4C2 96h
K1 96h
R8 net: Pyr.c <-> Pyr.m

M.


Extracellular transport (I)
flux [nmol/10^6 cells/h]
1000
0
-1000
-2000
-3000
-4000
-5000
-6000
*LB= -INF
4C2 48h
4C2 96h
K1 96h
R76 net: CO2 <-> CO2.ext

N.


Extracellular transport (II)
flux [nmol/10^6 cells/h]
60
50
40
30
20
10
0
-10
*no CI
*no CI
*no CI
4C2 48h
4C2 96h
K1 96h
R44: Glc.ext -> G6P
R45 net: Lac <-> Lac.ext
R46 net: Ala <-> Ala.ext
R47 net: Gln.ext <-> Gln
R48 net: Glu <-> Glu.ext

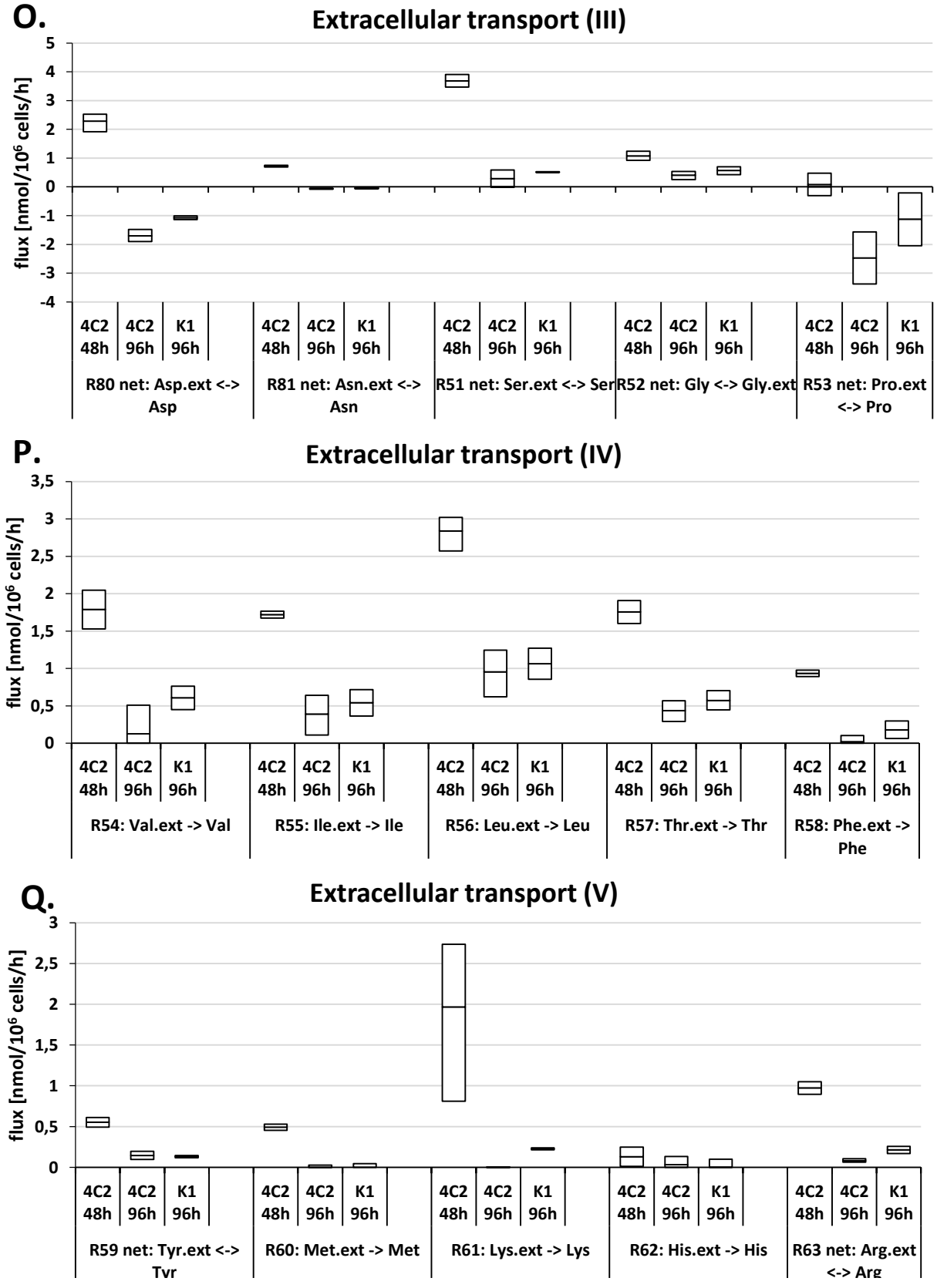


Figure 18. Intracellular metabolic fluxes determined by isotopically non-stationary metabolic flux analysis (INST-MFA). Fluxes were grouped into functional metabolic pathways for glycolysis (**A**), pentose phosphate pathways (**B**), lactate and alanine accumulation (**C**), TCA-cycle and pyruvate cycling (**D-E**), lipid precursor accumulation (**F**), amino acid metabolism (**G-K**), intracellular- (**L**) and extracellular transport (**M-Q**) reactions.

4.6 Discussion

In this project we elaborated and established the technical basis for accurately investigating the influence of somatic mutations versus germline configurations in mAbs. Parameters that have to be evaluated in more detail are protein properties and cellular behavior of recombinant mAb expression. Two engineered CHO host cell lines, based on CHO DUKX-B11 or CHO-K1, were developed (named CHO DUKX-B11 F3/F or CHO-K1 RMCE I3) by integration of a reporter gene cassette containing flanking regions for recombinase mediated cassette exchange. CHO DUKX-B11 F3/F or CHO-K1 RMCE I3 are now used to generate isogenic cell lines expressing different antibodies transcribed from the same genomic locus with identical gene copies. The isogenic transcription leads to highly similar mRNA levels and overcomes limitations arising from chromosomal position effects. The DUKX-B11 F3/F RMCE host cell line proved suitable for investigating the product-induced limited expression of the scFv-Fc antibody fragment 2F5 compared to scFv-Fc 3D6 (Mayrhofer et al., 2014). These two producer clones were used for proteomics analysis of consecutive passages (used as biological replicates) in the exponential growth phase. The analyses elucidated defined proteomic changes attributed to the product of interest (Sommeregger et al., 2016).

The RMCE engineered CHO-K1 host cell line, CHO-K1 RMCE I3, was developed to achieve higher specific growth rates and peak cell densities compared to the dihydrofolate reductase auxotrophic CHO DUKX-B11 F3/F. A diverse set of antibody producing clones was developed showing stable and homogeneous expression of mature or germline IgG1 antibody variants.

The gene-targeting vector used for full-length IgG1 integration into CHO-K1 RMCE I3 only contains a promoter sequence for the antibody heavy chain (**Figure 13B**). A promoter trap strategy was used for the light chain open reading frame. Only upon targeted gene integration the light chain can be transcribed from the CAGGS promoter introduced previously during establishment of the RMCE parental host using the plasmid depicted in **Figure 13A**. Using our vector strategy random integration only promotes transcription of the antibody heavy chain from the rEF1a promoter. Since no light chain molecules are transcribed from randomly integrated plasmids functional antibodies can neither be assembled nor secreted. In contrast to free light chains that can be readily secreted into the culture medium (Knittler and Haas, 1992), heavy chain folding and secretion requires free light chain. It was shown that the CH1 domain does not fold and is therefore tightly bound by a chaperone complex, containing Binding immunoglobulin protein (BiP), which retains the heavy chain within the ER when no light chain is available to support CH1 folding and intradomain disulfide formation (Feige and Buchner, 2014; Lee et al., 1999).

We developed and stabilized by subcloning eight recombinant cell lines expressing four real mAbs (2G12, 4B3, 2F5 and Ustekinumab) and four mAbs constructed by combination of V(D)J germline regions. With these cell lines we elaborated differences in cellular behavior and proteochemical properties assigned to the expressed mAbs.

Targeted gene integration was verified by PCR of the CHO-K1 RMCE I3 subclones' genomic DNA. The exact antibody gene copy numbers and transcript levels remain to be determined by qPCR, droplet digital (dd)PCR or fluorescent in-situ hybridization (FISH). However, because of the negative selection principle and the established promoter trap we can attribute the protein expression observed in RMCE daughter clones to a site-specific gene integration event.

First, the developed cell lines were evaluated for production stability, maximum cell density and product titer. The concept of isogenic sister clones was evaluated by rather identical expression levels. However, we could not identify significant differences between mature and germline antibody variants. It is possible that only at higher expression levels of difficult-to-express variants certain chaperones and other cellular determinants become limiting, leading to observable differences in expression levels. This might explain why in this project no obvious improvement was observed for the germline variant of 2G12, although in previous studies the specific productivity was increased by introducing germline residues in the variable sequence of 2G12 in IgM format (Chromikova et al., 2015). Also, we cannot rule out the possibility that certain synthetically designed V(D)J combinations in the germline variants are not optimal for their expression, although the specific junction between V, D and J genes were constructed based on available combinations in the NCBI database. Another reason for similar expression levels in Ustekinumab and 2G12 compared to their germline variants is the clone selection mechanism. It would be possible that during clone development the cellular capacity determined by e.g. chaperon composition and availability might become suitable for both antibody variants to drive optimal expression. To exclude this phenomenon verification by transient gene expression eliminating clonal selection should be done. Such cultures provide the same proteomic and metabolic background, although the gene copy level cannot be controlled.

As was shown in Sommeregger et al. (2016) a much higher specific productivity can be obtained by multi-copy integration of conventional plasmids or bacterial artificial chromosomes (BACs) in comparison to RMCE. Personal communication with Jürgen Bode (Hannover Medical School) confirmed this observation and highlights the fact that single-copy integration clones cannot compete with multi-copy industrial strains. The expected productivity of antibody producing RMCE subclones finally depends on the effort put into finding stable chromosomal hotspots ("safe harbors") with highest transcription rates of the reporter/selection cassette. Indeed, RMCE cell lines reaching specific productivities of up to 47 pg/c/day were reported for a GS-CHOK1SV RMCE subclone or 11 pg/c/d for subclones with confirmed single-copy integration (Zhang et al., 2015). Interestingly, all of the highly producing clones showed a gene integration at the chromosomal telomeric regions. With our improved CHO-K1 RMCE host cell line we could achieve a maximum specific productivity of 3 pg/c/d for IgG1 molecules. At this point it is important to say that in this project the primary focus was not to establish a highly productive clone but to integrate many different antibody variants into the same chromosomal environment for accurate comparison of their expected expression levels.

One additional potential of targeted gene integration or exchange methods is that a well-characterized antibody producing clone might be "reused" for another, different, antibody sequence. This should lead to subclones with similar product characteristics like glycosylation pattern, similar productivities and stable expression.

With the establishment of single integration clones and the characterization of these stable and highly transcriptionally active "safe harbors" or "hotspots" it should be possible to define targets for the new CRISPR/Cas9 tools. In order to enhance the productivity we might use the flexibility of the RMCE technique based on the availability of different incompatible (heterospecific) FRT sites to integrate multiple copies into the same hotspot by accumulative RMCE (Turan et al., 2013).

DSC measurements of the different mature and germline antibodies showed a trend towards higher stabilities for the germline antibody variants, except for 4B3 which was similar to its germline variant 136/63. Most importantly, no germline variant showed lower thermostabilities than their respective mature antibody sequence highlighting the potential of germinalization for improving protein stability. Reproducible endothermic thermal peak transitions for different antibodies could be identified corresponding to the CH2 and CH3 domain. We found that the Fab region unfolds in two transitions for mAb 2G12, 4B3, 136/63, 3D6 and 2F5 or a single transition for mAb 353/11, Ustekinumab and 554/12. Antibody 236/14 will be measured in future work. Alternatively, to investigate single or multiple transition peaks for the Fab fragment, antibodies might be subjected to papain digestion, followed by purification by protein A affinity- or ion exchange chromatography (Muronetz and Korpela, 2003). The Fc and Fab fragment separation enables unambiguous peak assignment in DSC measurements. Additional experiments might also include the pH or buffer dependency of thermal unfolding. The higher thermostability for germline variants did not translate into clearly higher expression levels in the CHO-K1 RMCE I3 host. Interestingly, although 4B3 and its germline variant 136/63 showed similar thermostabilities the expression levels were significantly different, indicating other expression bottlenecks than protein (thermo)stabilities.

To get more insight in cellular behavior, one of the established IgG1 expressing subclones was successfully subjected to metabolomics studies using ^{13}C-enriched tracer substances. Importantly, a culture process was established to omit the requirement for glucose-free media by properly fine-tuning the addition of glucose or glutamine feeds. This is an advantage given the fact that glucose-free media are often only delivered on special request and sometimes even show different cell culture performance after adjusting glucose levels to equal concentrations when compared to media with glucose (unpublished observation). The metabolic behavior of clone 4C2 expressing IgG1 136/63 was compared to the CHO-K1 host cell line and showed highly similar carbohydrate and amino acid metabolic rates. This highlights the suitability of the established CHO-K1 RMCE I3 host cell line to compare different antibody expression levels on identical metabolic background. On the other hand, we were not able to demonstrate a clear product-induced metabolic change in the 4C2 cell line. One difference in the metabolic behavior between the two cell lines was the earlier shift from lactate production to consumption and overall lower lactate production rates for the CHO-K1

host cell line. However, whether this effect is really a property induced by the product of interest, or simply a result of subclone selection during cell line establishment of the CHO-K1 RMCE I3 host or 136/63 daughter clones, remains to be investigated in more detail. This might be done by subjecting CHO-K1 RMCE I3 to proper metabolomics studies in comparison to CHO-K1 and the 4C2 cell line. Alternatively, if the observed phenotype is really caused by the specific antibody variable sequence, it should be possible to reverse the observed phenotype by exchanging the antibody cassette with the parental selection/reporter marker.

Another reason for the similarities in metabolic behavior might be the level of expression. It is possible that certain proteomic or metabolic changes induced by the product are only pronounced at rather high expression levels.

5 CONCLUSIONS

The conclusions are structured towards the program for this PhD thesis:

1. Converging in-silico and wet lab methods to develop new strategies for the definition of new mAbs

Novel emerging computational tools provide valuable support to existing experimental procedures in the wet lab. In this thesis, the combined approach using in-silico and in-vitro methods to complement each other is demonstrated.

The loss of binding affinity and selection of proper and necessary backmutation sites in the humanized antibody's framework regions, to re-establish antigen binding, constitute major challenges in the humanization process of non-human antibodies. In this PhD program, the experimental workflow towards a prospective humanization design cycle is proposed that uses computational MD simulations to investigate in-silico behavior of structural antibody loops. The analysis of the simulation is based on a similarity score that compares wild-type CDR loop structural dynamics to CDR dynamics of different framework mutants. An arbitrary similarity threshold could be defined that distinguishes binders from non-binders in a qualitative manner. With the established similarity score we can define how similar the behavior of CDR regions of 3H6 mutants is to the wild-type 3H6 CDR behavior, but we cannot rule out that alternative conformational dynamics also contribute significantly to the binding affinity. Therefore, additional features of structural dynamics or antibody-antigen interactions (e.g. hydrogen-bond pattern) will be investigated in-silico and tested by additional antibody variants. Future work will focus on optimizing the simulation analyses for enhanced sensitivity and selectivity to detect even smaller changes in antibody binding affinities.

During this project, it also became evident how important an effective communication and coordinated workflow between research groups of different disciplines is to merge expertise for the development of novel ideas and concepts. The experimental in-vitro work was done at the Department of Biotechnology (DBT) in Renate Kunert's group. Different humanized antibody variants and mutants were designed based on rational mutations, constructed and expressed in mammalian cell cultures using stable or optimized transient transfections. The actual binding affinity was measured with biolayer-interferometry using two different assay setups with purified protein or concentrated crude culture supernatants. The in-silico analyses were performed by the group of Chris Oostenbrink at the Institute of Molecular Modeling and Simulation (MMS). Combined analysis of in-vitro binding affinities and in-silico similarity scores advanced the traditional humanization process by using a novel prospective design cycle leading to superhumanized antibody variants (BM07) with partially re-established in-vitro binding affinities and a high in-silico similarity score. These variants represent the basis for further affinity optimization supported by MD simulations.

To evaluate the remaining rodent sequences of the humanized mAb variants in more detail, in-silico prediction tools were used to determine the number of expected MHCII epitopes possibly triggering anti-drug-antibody responses. Several studies claim that variants with a reduced number of predicted MHCII epitopes are less likely to induce an adverse immune response. Such T-cell responses, would lead to maturation of B-cells into antibody-producing plasma cells characterized by high-titer secretion of antibodies directed against the therapeutic protein. In contrast to supplementation therapy with high-dose treatment for long time-periods, proteins which are used as vaccines require the presence of certain MHCII epitopes to induce a potent helper and memory T-cell response. Therefore, in our approach with the anti-idiotypic antibody not all MHCII epitopes should be considered as bad especially since endogenous human proteins also encode MHCII restricted epitopes. During evolution, the human body has learned to tolerate certain (self) antigens involving specialized regulatory T-cells (Treg). This fact becomes evident when we analyzed human antibody germline sequences, which also showed a relatively high content of predicted MHCII epitopes. Therefore, it cannot be concluded that all MHCII epitopes will ultimately trigger a potent T-helper cell response. An additional statistical analysis will define differences in the number and type of epitope content in germline antibody sequences compared to mature or humanized antibody sequences. Ultimately, the in-silico prediction should be verified by proper in-vitro immunological assays to confirm the induction of an immune response mediated by the ternary peptide/MHCII/T-cell receptor complex.

2. Can product-related and cell-physiological parameters be used to predict and improve the technological production process?

In the humanization project it became evident that artificially constructed antibody variants subjected to computational simulation cannot always be functionally assembled in the present cellular environment and that minor differences in the protein sequence can have a significant impact on the cellular capability to secrete the mAb of interest. In-silico bioinformatics knowledge was used to investigate cellular behavior in relation to physicochemical protein properties. We compared a set of mature primary amino acid sequences of highly-cited anti-HIV-1 antibodies (2G12, 2F5 and 4B3) and one therapeutic antibody (Ustekinumab) to their closest V-, D- and J germline genes found in the IMGT database. The identified germline genes were assembled to establish a set of germline antibody variants. Transcription of the different antibody sequences was standardized at the genomic level by targeted gene integration using RMCE to develop isogenic, homogeneous and stably expressing daughter clones. A proof of concept for RMCE as valuable tool for gene integration could be provided by a DUKX-B11 RMCE host cell line expressing antibody 2F5 and 3D6 in scFv-Fc format. Antibody producing daughter clones were successfully used in label-free proteomics studies by Sommeregger et al. (2016). Additionally, we could establish targeted gene-integration strategies in a CHO-K1 RMCE host not only for single open reading frame expression (e.g. scFv-Fc variants) but also for introducing heavy- and light chains for full-length IgG1 antibody expression. A diverse

antibody set was introduced site-specifically into the CHO-K1 RMCE parental host cell line and cellular properties such as specific growth, productivity, expression stability, clone homogeneity and intracellular product formation was analyzed in different culture processes to evaluate the expressability of the antibodies. Evidence for targeted gene integration was provided by PCR analysis. One RMCE daughter cell line, expressing the germline antibody 136/63 IgG1, could successfully be applied to isotopically non-stationary metabolic flux analysis (INST-MFA) during a research visit at IBET, Oeiras-Portugal, in the group of Paula M. Alves. Enough material was generated and purified to analyze product stability. DSC measurements of purified proteins showed a significant increase in the Fab stability of 2G12 and Ustekinumab by replacing the V-, D- and J mature gene sequences with the germline amino acid residues. Although resistance to temperature-induced unfolding was increased for these two germline variants (353/11 and 554/12) we cannot report significant changes in antibody-specific productivities, intracellular product formation or growth behavior. The observation of equal production capabilities for mature versus germline antibody variants with different thermostabilities can be attributed to protein homeostasis of the RMCE daughter cells that supports expression of the more (thermally) instable antibodies 2G12 and Ustekinumab at the given transcription rates. This homeostasis is assured by different chaperones within the ER important for antibody folding and assembly and it might be possible that the limit for folding-assistance is not reached with the CHO-K1 RMCE I3 host. Only at higher expression rates, using cell clones with randomly introduced multiple copies of plasmids or bacterial artificial chromosomes, differences in expression levels between 2G12 and Ustekinumab compared to their germline variants might become observable. This hypothesis is supported by the fact that a higher fold-change in specific productivity could be observed for the antibody 3D6, a high germline-identity antibody, when compared to 2F5 with low germline sequence identity using CHO clones with multi-copy gene integrations (**Figure 15A**). It is important to note that the primary focus in the development of the RMCE host cell lines was not the generation of high-producing daughter clones but to have a platform technology in hand to control the gene copy number and integration site. Higher productivities by simultaneously controlling the gene copy number might be achieved by RMCE techniques that repeatedly introduce multiple expression cassettes into the same genomic locus by several rounds of RMCE (accumulative RMCE). Alternatively, highly transcriptionally active daughter clones might be selected by rigorous screening of parental RMCE cell clones or by rationally target highly active chromosomal hotspots by emerging gene-targeting techniques such as zinc finger nucleases (ZFN), transcription activator-like effector nucleases (TALEN) or clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated protein 9 (Cas9) techniques. It would also be highly interesting at which expression level a possible bottleneck in expression of the mature antibody becomes obvious and if this threshold is the same for different antibody sets (e.g. 2G12 and 353/11 versus Ustekinumab and 554/12). We further assume that antibodies with lower (thermo)stabilities need more effective folding-assistance mediated by different chaperone activities. Due to lower transition temperatures of 2G12 and 353/11 compared to Ustekinumab and 554/12, we can hypothesize that the difference in expression levels between mature and germline antibody can be detected already at lower expression levels for the 2G12 set,

than for the Ustekinumab set. This hypothesis might be tested by generation of (accumulative) RMCE host cell lines showing wide range of expression levels but defined gene-copy numbers. Alternatively, this array of cell clones might also be established by random integration but requires the exact determination of uncontrolled gene-integration events by tedious qPCR techniques.

In future work the established antibody panel will be converted into scFv-Fc format by proper molecular biological methods (Mayrhofer and Kunert, 2016, under review) and expression differences will be evaluated by transient gene expression. In scFv-Fc antibody formats the main functions of mAbs are maintained such as antigen binding and Fc-related effector functions, especially complement-dependent cytotoxicity (CDC), antibody-dependent cellular cytotoxicity (ADCC) and elongation of in-vivo half-life. Another advantage of these homodimeric molecular structures is that only a single vector has to be transfected, thereby enhancing transfection efficiency and expression levels. Similar to full-length IgG molecules, protein properties such as differences in thermostabilities evoked by the individual antibody variable sequence can be studied in scFv-Fc formats.

6 REFERENCES

Afif, W., Loftus, E.V., Faubion, W.A., Kane, S.V., Bruining, D.H., Hanson, K.A., and Sandborn, W.J. (2010). Clinical Utility of Measuring Infliximab and Human Anti-Chimeric Antibody Concentrations in Patients With Inflammatory Bowel Disease. Am. J. Gastroenterol. *105*, 1133–1139.

Aggarwal, S. (2014). What's fueling the biotech engine-2012 to 2013. Nat. Biotechnol. *32*, 32–39.

Ahn, W.S., and Antoniewicz, M.R. (2011). Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. Metab. Eng. *13*, 598–609.

Almagro, J.C., and Fransson, J. (2008). Humanization of antibodies. Front. Biosci. J. Virtual Libr. *13*, 1619–1633.

Antoniewicz, M.R. (2013). 13C metabolic flux analysis: optimal design of isotopic labeling experiments. Curr. Opin. Biotechnol. *24*, 1116–1121.

Attarwala, H. (2010). TGN1412: From Discovery to Disaster. J. Young Pharm. JYP *2*, 332–336.

Brinks, V., Jiskoot, W., and Schellekens, H. (2011). Immunogenicity of Therapeutic Proteins: The Use of Animal Models. Pharm. Res. *28*, 2379–2385.

Brüggemann, M., Spicer, C., Buluwela, L., Rosewell, I., Barton, S., Surani, M.A., and Rabbitts, T.H. (1991). Human antibody production in transgenic mice: expression from 100 kb of the human IgH locus. Eur. J. Immunol. *21*, 1323–1326.

Bryson, S., Julien, J.-P., Isenman, D.E., Kunert, R., Katinger, H., and Pai, E.F. (2008). Crystal structure of the complex between the F(ab)' fragment of the cross-neutralizing anti-HIV-1 antibody 2F5 and the F(ab) fragment of its anti-i... - PubMed - NCBI.

Bugelski, P.J., Achuthanandam, R., Capocasale, R.J., Treacy, G., and Bouman-Thio, E. (2009). Monoclonal antibody-induced cytokine-release syndrome. Expert Rev. Clin. Immunol. *5*, 499–521.

Calarese, D.A., Scanlan, C.N., Zwick, M.B., Deechongkit, S., Mimura, Y., Kunert, R., Zhu, P., Wormald, M.R., Stanfield, R.L., Roux, K.H., et al. (2003). Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. Science *300*, 2065–2071.

Carinhas, N., Pais, D.A.M., Koshkin, A., Fernandes, P., Coroadinha, A.S., Carrondo, M.J.T., Alves, P.M., and Teixeira, A.P. (2016). Metabolic flux profiling of MDCK cells during growth and canine adenovirus vector production. Sci. Rep. *6*, 23529.

Chen, L., Kutskova, Y.A., Hong, F., Memmott, J.E., Zhong, S., Jenkinson, M.D., and Hsieh, C.-M. (2015). Preferential germline usage and VH/VL pairing observed in human antibodies selected by mRNA display. Protein Eng. Des. Sel. PEDS *28*, 427–435.

Chong, W.P.K., Reddy, S.G., Yusufi, F.N.K., Lee, D.-Y., Wong, N.S.C., Heng, C.K., Yap, M.G.S., and Ho, Y.S. (2010). Metabolomics-driven approach for the improvement of Chinese hamster ovary cell growth: overexpression of malate dehydrogenase II. J. Biotechnol. *147*, 116–121.

Chong, W.P.K., Thng, S.H., Hiu, A.P., Lee, D.-Y., Chan, E.C.Y., and Ho, Y.S. (2012). LC-MS-based metabolic characterization of high monoclonal antibody-producing Chinese hamster ovary cells. Biotechnol. Bioeng. *109*, 3103–3111.

Chothia, C., and Lesk, A.M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. J. Mol. Biol. *196*, 901–917.

Chromikova, V., Mader, A., Hofbauer, S., Göbl, C., Madl, T., Gach, J.S., Bauernfried, S., Furtmüller, P.G., Forthal, D.N., Mach, L., et al. (2015). Introduction of germline residues improves the stability of anti-HIV mAb 2G12-IgM. Biochim. Biophys. Acta BBA - Proteins Proteomics *1854*, 1536–1544.

De Groot, A.S., and Moise, L. (2007). Prediction of immunogenicity for therapeutic proteins: state of the art. Curr. Opin. Drug Discov. Devel. *10*, 332–340.

De Groot, A.S., and Scott, D.W. (2007). Immunogenicity of protein therapeutics. Trends Immunol. *28*, 482–490.

Dietmair, S., Hodson, M.P., Quek, L.-E., Timmins, N.E., Chrysanthopoulos, P., Jacob, S.S., Gray, P., and Nielsen, L.K. (2012). Metabolite profiling of CHO cells with different growth characteristics. Biotechnol. Bioeng. *109*, 1404–1414.

Dimitrov, J.D., Kaveri, S.V., and Lacroix-Desmazes, S. (2014). Thermodynamic stability contributes to immunoglobulin specificity. Trends Biochem. Sci. *39*, 221–226.

Dübel, S., and Reichert, J.M. (2014). Handbook of Therapeutic Antibodies (John Wiley & Sons).

Ewert, S., Huber, T., Honegger, A., and Plückthun, A. (2003). Biophysical properties of human antibody variable domains. J. Mol. Biol. *325*, 531–553.

Feige, M.J., and Buchner, J. (2014). Principles and engineering of antibody folding and assembly. Biochim. Biophys. Acta BBA - Proteins Proteomics *1844*, 2024–2031.

Finco, D., Grimaldi, C., Fort, M., Walker, M., Kiessling, A., Wolf, B., Salcedo, T., Faggioni, R., Schneider, A., Ibraghimov, A., et al. (2014). Cytokine release assays: Current practices and future directions. Cytokine *66*, 143–155.

Findlay, L., Eastwood, D., Ball, C., Robinson, C.J., Bird, C., Wadhwa, M., Thorpe, S.J., Thorpe, R., Stebbings, R., and Poole, S. (2011). Comparison of novel methods for predicting the risk of pro-inflammatory clinical infusion reactions during monoclonal antibody therapy. J. Immunol. Methods *371*, 134–142.

Foote, J., and Winter, G. (1992). Antibody framework residues affecting the conformation of the hypervariable loops. J. Mol. Biol. *224*, 487–499.

Gach, J.S., Maurer, M., Hahn, R., Gasser, B., Mattanovich, D., Katinger, H., and Kunert, R. (2007a). High level expression of a promising anti-idiotypic antibody fragment vaccine against HIV-1 in Pichia pastoris. J. Biotechnol. *128*, 735–746.

Gach, J.S., Quendler, H., Weik, R., Katinger, H., and Kunert, R. (2007b). Partial Humanization and Characterization of an Anti-idiotypic Antibody against Monoclonal Antibody 2F5, a Potential HIV Vaccine? AIDS Res. Hum. Retroviruses *23*, 1405–1415.

Gach, J.S., Quendler, H., Ferko, B., Katinger, H., and Kunert, R. (2008a). Expression, Purification, and In Vivo Administration of a Promising Anti-Idiotypic HIV-1 Vaccine. Mol. Biotechnol. *39*, 119–125.

Gach, J.S., Quendler, H., Strobach, S., Katinger, H., and Kunert, R. (2008b). Structural analysis and in vivo administration of an anti-idiotypic antibody against mAb 2F5. Mol. Immunol. *45*, 1027–1034.

Garber, E., and Demarest, S.J. (2007). A broad range of Fab stabilities within a host of therapeutic IgGs. Biochem. Biophys. Res. Commun. *355*, 751–757.

Georgiou, G., Ippolito, G.C., Beausang, J., Busse, C.E., Wardemann, H., and Quake, S.R. (2014). The promise and challenge of high-throughput sequencing of the antibody repertoire. Nat. Biotechnol. *32*, 158–168.

Greenbaum, J., Sidney, J., Chung, J., Brander, C., Peters, B., and Sette, A. (2011). Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. Immunogenetics *63*, 325–335.

van Gunsteren, W.F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D.P., Glättli, A., Hünenberger, P.H., et al. (2006). Biomolecular modeling: Goals, problems, perspectives. Angew. Chem. Int. Ed Engl. *45*, 4064–4092.

Hasegawa, H., Wendling, J., He, F., Trilisky, E., Stevenson, R., Franey, H., Kinderman, F., Li, G., Piedmonte, D.M., Osslund, T., et al. (2011). In vivo crystallization of human IgG in the endoplasmic reticulum of engineered Chinese hamster ovary (CHO) cells. J. Biol. Chem. *286*, 19917–19931.

Hibi, T., and Dosch, H.-M. (1986). Limiting dilution analysis of the B cell compartment in human bone marrow. Eur. J. Immunol. *16*, 139–145.

Holgate, R.G.E., and Baker, M.P. (2009). Circumventing immunogenicity in the development of therapeutic antibodies. IDrugs Investig. Drugs J. *12*, 233–237.

Ionescu, R.M., Vlasak, J., Price, C., and Kirchmeier, M. (2008). Contribution of variable domains to the stability of humanized IgG1 monoclonal antibodies. J. Pharm. Sci. *97*, 1414–1426.

Jarasch, A., Koll, H., Regula, J.T., Bader, M., Papadimitriou, A., and Kettenberger, H. (2015). Developability assessment during the selection of novel therapeutic antibodies. J. Pharm. Sci. *104*, 1885–1898.

Jawa, V., Cousens, L.P., Awwad, M., Wakshull, E., Kropshofer, H., and De Groot, A.S. (2013). T-cell dependent immunogenicity of protein therapeutics: Preclinical assessment and mitigation. Clin. Immunol. Orlando Fla *149*, 534–555.

Jones, P.T., Dear, P.H., Foote, J., Neuberger, M.S., and Winter, G. (1986). Replacing the complementarity-determining regions in a human antibody with those from a mouse. Nature *321*, 522–525.

Joubert, M.K., Hokom, M., Eakin, C., Zhou, L., Deshpande, M., Baker, M.P., Goletz, T.J., Kerwin, B.A., Chirmule, N., Narhi, L.O., et al. (2012). Highly aggregated antibody therapeutics can enhance the in vitro innate and late-stage T-cell immune responses. J. Biol. Chem. *287*, 25266–25279.

Kim, S., Davis, M., Sinn, E., Patten, P., and Hood, L. (1981). Antibody diversity: somatic hypermutation of rearranged VH genes. Cell *27*, 573–581.

Klein, F., Diskin, R., Scheid, J.F., Gaebler, C., Mouquet, H., Georgiev, I.S., Pancera, M., Zhou, T., Incesu, R.-B., Fu, B.Z., et al. (2013). Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. Cell *153*, 126–138.

Knittler, M.R., and Haas, I.G. (1992). Interaction of BiP with newly synthesized immunoglobulin light chain molecules: cycles of sequential binding and release. EMBO J. *11*, 1573–1581.

Köhler, G., and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. Nature *256*, 495–497.

Kunert, R.E., Weik, R., Ferko, B., Stiegler, G., and Katinger, H. (2002). Anti-idiotypic antibody Ab2/3H6 mimics the epitope of the neutralizing anti-HIV-1 monoclonal antibody 2F5. Aids *16*, 667–668.

Lattenmayer, C., Trummer, E., Schriebl, K., Vorauer-Uhl, K., Mueller, D., Katinger, H., and Kunert, R. (2007). Characterisation of recombinant CHO cell lines by investigation of protein productivities and genetic parameters. J. Biotechnol. *128*, 716–725.

Lee, Y.-K., Brewer, J.W., Hellman, R., and Hendershot, L.M. (1999). BiP and Immunoglobulin Light Chain Cooperate to Control the Folding of Heavy Chain and Ensure the Fidelity of Immunoglobulin Assembly. Mol. Biol. Cell *10*, 2209–2219.

Lefranc, M.-P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., Carillon, E., Duvergey, H., Houles, A., Paysan-Lafosse, T., et al. (2015). IMGT®, the international ImMunoGeneTics information system® 25 years on. Nucleic Acids Res. *43*, D413–D422.

Li, F., Vijayasankaran, N., Shen, A. (Yijuan), Kiss, R., and Amanullah, A. (2010). Cell culture processes for monoclonal antibody production. mAbs *2*, 466–477.

Luo, J., Vijayasankaran, N., Autsen, J., Santuray, R., Hudson, T., Amanullah, A., and Li, F. (2012). Comparative metabolite analysis to understand lactate metabolism shift in Chinese hamster ovary cell culture process. Biotechnol. Bioeng. *109*, 146–156.

Lynch, C.M., Hart, B.W., and Grewal, I.S. (2009). Practical considerations for nonclinical safety evaluation of therapeutic monoclonal antibodies. mAbs *1*, 2–11.

Ma, N., Ellet, J., Okediadi, C., Hermes, P., McCormick, E., and Casnocha, S. (2009). A single nutrient feed supports both chemically defined NS0 and CHO fed-batch processes: Improved productivity and lactate metabolism. Biotechnol. Prog. *25*, 1353–1363.

MacCallum, R.M., Martin, A.C., and Thornton, J.M. (1996). Antibody-antigen interactions: contact analysis and binding site topography. J. Mol. Biol. *262*, 732–745.

Mader, A., and Kunert, R. (2010). Humanization strategies for an anti-idiotypic antibody mimicking HIV-1 gp41. Protein Eng. Des. Sel. PEDS *23*, 947–954.

Mader, A., Prewein, B., Zboray, K., Casanova, E., and Kunert, R. (2012). Exploration of BAC versus plasmid expression vectors in recombinant CHO cells. Appl. Microbiol. Biotechnol. *97*, 4049–4054.

Margreitter, C., Mayrhofer, P., Kunert, R., and Oostenbrink, C. (2016). Antibody humanization by molecular dynamics simulations-in-silico guided selection of critical backmutations. J. Mol. Recognit. JMR *29*, 266–275.

Martin, A.C., Cheetham, J.C., and Rees, A.R. (1989). Modeling antibody hypervariable loops: a combined algorithm. Proc. Natl. Acad. Sci. U. S. A. *86*, 9268–9272.

Mayrhofer, P., and Kunert, R. (2016). Cloning of single-chain antibody variants by overlap-extension PCR for evaluation of antibody expression in transient gene expression. Methods Mol. Biol. (under review).

Mayrhofer, P., Kratzer, B., Sommeregger, W., Steinfellner, W., Reinhart, D., Mader, A., Turan, S., Qiao, J., Bode, J., and Kunert, R. (2014). Accurate comparison of antibody expression levels by reproducible transgene targeting in engineered recombination-competent CHO cells. Appl. Microbiol. Biotechnol. *98*, 9723–9733.

McFarland, B.J., and Beeson, C. (2002). Binding interactions between peptides and proteins of the class II major histocompatibility complex. Med. Res. Rev. *22*, 168–203.

Metallo, C.M., Walther, J.L., and Stephanopoulos, G. (2009). Evaluation of 13C isotopic tracers for metabolic flux analysis in mammalian cells. J. Biotechnol. *144*, 167–174.

Morrison, S.L., Johnson, M.J., Herzenberg, L.A., and Oi, V.T. (1984). Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. Proc Natl Acad Sci USA *21*, 6851–6855.

Mosmann, T.R., Cherwinski, H., Bond, M.W., Giedlin, M.A., and Coffman, R.L. (1986). Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. J. Immunol. Baltim. Md 1950 *136*, 2348–2357.

Mouquet, H., Scheid, J.F., Zoller, M.J., Krogsgaard, M., Ott, R.G., Shukair, S., Artyomov, M.N., Pietzsch, J., Connors, M., Pereyra, F., et al. (2010). Polyreactivity increases the apparent affinity of anti-HIV antibodies by heteroligation. Nature *467*, 591–595.

Mueller, M., Loh, M.Q.T., Tee, D.H.Y., Yang, Y., and Jungbauer, A. (2013). Liquid Formulations for Long-Term Storage of Monoclonal IgGs. Appl. Biochem. Biotechnol. *169*, 1431–1448.

Muronetz, V.I., and Korpela, T. (2003). Isolation of antigens and antibodies by affinity chromatography. J. Chromatogr. B *790*, 53–66.

Paul, S., Kolla, R.V., Sidney, J., Weiskopf, D., Fleri, W., Kim, Y., Peters, B., Sette, A., Paul, S., Kolla, R.V., et al. (2013). Evaluating the Immunogenicity of Protein Drugs by Applying In Vitro MHC Binding Data and the Immune Epitope Database and Analysis Resource. J. Immunol. Res. J. Immunol. Res. \i2013, 2013, e467852.

Paul, S., Lindestam Arlehamn, C.S., Scriba, T.J., Dillon, M.B.C., Oseroff, C., Hinz, D., McKinney, D.M., Carrasco Pro, S., Sidney, J., Peters, B., et al. (2015). Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. J. Immunol. Methods *422*, 28–34.

Quek, L.-E., Dietmair, S., Krömer, J.O., and Nielsen, L.K. (2010). Metabolic flux analysis in mammalian cell culture. Metab. Eng. *12*, 161–171.

Rammensee, H.-G., Friede, T., and Stevanović, S. (1995). MHC ligands and peptide motifs: first listing. Immunogenetics *41*, 178–228.

Ratanji, K.D., Derrick, J.P., Dearman, R.J., and Kimber, I. (2014). Immunogenicity of therapeutic proteins: influence of aggregation. J. Immunotoxicol. *11*, 99–109.

Rita Costa, A., Elisa Rodrigues, M., Henriques, M., Azeredo, J., and Oliveira, R. (2010). Guidelines to cell engineering for monoclonal antibody production. Eur. J. Pharm. Biopharm. Off. J. Arbeitsgemeinschaft Für Pharm. Verfahrenstechnik EV *74*, 127–138.

Roche, P.A., and Furuta, K. (2015). The ins and outs of MHC class II-mediated antigen processing and presentation. Nat. Rev. Immunol. *15*, 203–216.

Rombach-Riegraf, V., Karle, A.C., Wolf, B., Sordé, L., Koepke, S., Gottlieb, S., Krieg, J., Djidja, M.-C., Baban, A., Spindeldreher, S., et al. (2014). Aggregation of Human Recombinant Monoclonal Antibodies Influences the Capacity of Dendritic Cells to Stimulate Adaptive T-Cell Responses In Vitro. PLOS ONE *9*, e86322.

de Ruiter, A., Mader, A., Kunert, R., and Oostenbrink, C. (2011). Molecular Simulations to Rationalize Humanized Ab2/3H6 Activity. Aust. J. Chem. *64*, 900.

Schafer, J.R., Jesdale, B.M., George, J.A., Kouttab, N.M., and De Groot, A.S. (1998). Prediction of well-conserved HIV-1 ligands using a matrix-based algorithm, EpiMatrix. Vaccine *16*, 1880–1884.

Scheid, J.F., Mouquet, H., Feldhahn, N., Seaman, M.S., Velinzon, K., Pietzsch, J., Ott, R.G., Anthony, R.M., Zebroski, H., Hurley, A., et al. (2009). Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. Nature *458*, 636–640.

Schroeder, H.W. (2006). Similarity and divergence in the development and expression of the mouse and human antibody repertoires. Dev. Comp. Immunol. *30*, 119–135.

Schroff, R.W., Foon, K.A., Beatty, S.M., Oldham, R.K., and Morgan, A.C. (1985). Human Anti-Murine Immunoglobulin Responses in Patients Receiving Monoclonal Antibody Therapy. Cancer Res. *45*, 879–885.

Seeliger, D., Schulz, P., Litzenburger, T., Spitz, J., Hoerer, S., Blech, M., Enenkel, B., Studts, J.M., Garidel, P., and Karow, A.R. (2015). Boosting antibody developability through rational sequence optimization. mAbs *7*, 505–515.

Sellick, C.A., Croxford, A.S., Maqsood, A.R., Stephens, G., Westerhoff, H.V., Goodacre, R., and Dickson, A.J. (2011). Metabolite profiling of recombinant CHO cells: designing tailored feeding regimes that enhance recombinant antibody production. Biotechnol. Bioeng. *108*, 3025–3031.

Sidney, J., Assarsson, E., Moore, C., Ngo, S., Pinilla, C., Sette, A., and Peters, B. (2008). Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. Immunome Res. *4*, 2.

Sommeregger, W., Mayrhofer, P., Steinfellner, W., Reinhart, D., Henry, M., Clynes, M., Meleady, P., and Kunert, R. (2016). Proteomic differences in recombinant CHO cells producing two similar antibody fragments. Biotechnol. Bioeng.

Stoops, J., Byrd, S., and Hasegawa, H. (2012). Russell body inducing threshold depends on the variable domain sequences of individual human IgG clones and the cellular protein homeostasis. Biochim. Biophys. Acta BBA - Mol. Cell Res. *1823*, 1643–1657.

Templeton, N., Dean, J., Reddy, P., and Young, J.D. (2013). Peak antibody production is associated with increased oxidative metabolism in an industrially relevant fed-batch CHO cell culture. Biotechnol. Bioeng. *110*, 2013–2024.

Tiller, T., Schuster, I., Deppe, D., Siegers, K., Strohner, R., Herrmann, T., Berenguer, M., Poujol, D., Stehle, J., Stark, Y., et al. (2013). A fully synthetic human Fab antibody library based on fixed VH/VL framework pairings with favorable biophysical properties. mAbs *5*, 445–470.

Turan, S., Zehe, C., Kuehle, J., Qiao, J., and Bode, J. (2013). Recombinase-mediated cassette exchange (RMCE) - a rapidly-expanding toolbox for targeted genomic modifications. Gene *515*, 1–27.

Vargas-Madrazo, E., and Paz-García, E. (2003). An improved model of association for VH-VL immunoglobulin domains: asymmetries between VH and VL in the packing of some interface residues. J. Mol. Recognit. JMR *16*, 113–120.

Villadangos, J.A., and Schnorrer, P. (2007). Intrinsic and cooperative antigen-presenting functions of dendritic-cell subsets in vivo. Nat. Rev. Immunol. *7*, 543–555.

Vita, R., Overton, J.A., Greenbaum, J.A., Ponomarenko, J., Clark, J.D., Cantrell, J.R., Wheeler, D.K., Gabbard, J.L., Hix, D., Sette, A., et al. (2015). The immune epitope database (IEDB) 3.0. Nucleic Acids Res. *43*, D405–D412.

Wang, F., Sen, S., Zhang, Y., Ahmad, I., Zhu, X., Wilson, I.A., Smider, V.V., Magliery, T.J., and Schultz, P.G. (2013). Somatic hypermutation maintains antibody thermodynamic stability during affinity maturation. Proc. Natl. Acad. Sci. U. S. A. *110*, 4261–4266.

Wang, P., Sidney, J., Dow, C., Mothé, B., Sette, A., and Peters, B. (2008). A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach. PLOS Comput Biol *4*, e1000048.

Wilson, C., Bellen, H.J., and Gehring, W.J. (1990). Position Effects on Eukaryotic Gene Expression. Annu. Rev. Cell Biol. *6*, 679–714.

Winter, G., Griffiths, A.D., Hawkins, R.E., and Hoogenboom, H.R. (1994). Making Antibodies by Phage Display Technology. Annu. Rev. Immunol. *12*, 433–455.

Wu, T.T., and 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. J. Exp. Med. *132*, 211–250.

Wu, T.T., Johnson, G., and Kabat, E.A. (1993). Length distribution of CDRH3 in antibodies. Proteins Struct. Funct. Bioinforma. *16*, 1–7.

Yang, X., Xu, W., Dukleska, S., Benchaar, S., Mengisen, S., Antochshuk, V., Cheung, J., Mann, L., Babadjanova, Z., Rowand, J., et al. (2013). Developability studies before initiation of process development: improving manufacturability of monoclonal antibodies. mAbs *5*, 787–794.

Zhang, C. (2012). Hybridoma technology for the generation of monoclonal antibodies. Methods Mol. Biol. Clifton NJ *901*, 117–135.

Zhang, L., Inniss, M.C., Han, S., Moffat, M., Jones, H., Zhang, B., Cox, W.L., Rance, J.R., and Young, R.J. (2015). Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line. Biotechnol. Prog. *31*, 1645–1656.

Zhu, J., Yamane, H., and Paul, W.E. (2010). Differentiation of Effector CD4 T Cell Populations. Annu. Rev. Immunol. *28*, 445–489.

7 SELECTED PUBLICATIONS

1. Margreitter, C.*, **Mayrhofer, P.***, Kunert, R., and Oostenbrink, C. (2016). Antibody humanization by molecular dynamics simulations-in-silico guided selection of critical backmutations. J. Mol. Recognit. JMR *29*, 266–275.

   * authors contributed equally to this work

2. **Mayrhofer, P.**, Kratzer, B., Sommeregger, W., Steinfellner, W., Reinhart, D., Mader, A., Turan, S., Qiao, J., Bode, J., and Kunert, R. (2014). Accurate comparison of antibody expression levels by reproducible transgene targeting in engineered recombination-competent CHO cells. Appl. Microbiol. Biotechnol. *98*, 9723–9733.

3. **Mayrhofer, P.**, and Kunert, R. (2016). Cloning of single-chain antibody variants by overlap-extension PCR for evaluation of antibody expression in transient gene expression. Methods Mol. Biol. (under review).

4. Sommeregger, W., **Mayrhofer, P.**, Steinfellner, W., Reinhart, D., Henry, M., Clynes, M., Meleady, P., and Kunert, R. (2016). Proteomic differences in recombinant CHO cells producing two similar antibody fragments. Biotechnol. Bioeng. doi:10.1002/bit.25957.

Research article

Journal of Molecular Recognition

Received: 21 October 2015, Revised: 19 November 2015, Accepted: 20 November 2015, Published online in Wiley Online Library: 8 January 2016

(wileyonlinelibrary.com) DOI: 10.1002/jmr.2527

Antibody humanization by molecular dynamics simulations—*in-silico* guided selection of critical backmutations

Christian Margreitter[a†], Patrick Mayrhofer[b†], Renate Kunert[b] and Chris Oostenbrink[a*]

Monoclonal antibodies represent the fastest growing class of biotherapeutic proteins. However, as they are often initially derived from rodent organisms, there is a severe risk of immunogenic reactions, hampering their applicability. The humanization of these antibodies remains a challenging task in the context of rational drug design. "Superhumanization" describes the direct transfer of the complementarity determining regions to a human germline framework, but this humanization approach often results in loss of binding affinity. In this study, we present a new approach for predicting promising backmutation sites using molecular dynamics simulations of the model antibody Ab2/3H6. The simulation method was developed in close conjunction with novel specificity experiments. Binding properties of mAb variants were evaluated directly from crude supernatants and confirmed using established binding affinity assays for purified antibodies. Our approach provides access to the dynamical features of the actual binding sites of an antibody, based solely on the antibody sequence. Thus we do not need structural data on the antibody–antigen complex and circumvent cumbersome methods to assess binding affinities. © 2016 The Authors Journal of Molecular Recognition Published by John Wiley & Sons Ltd.

Keywords: molecular dynamics; antibody humanization; conformational clustering; binding affinity; GROMOS

INTRODUCTION

Monoclonal antibodies (mAbs) represent the fastest growing class of biotherapeutic proteins, with US sales reaching $24.6 billion in 2012 (Aggarwal, 2014). As of January 2015, IMGT® (Lefranc *et al.*, 2005), the international ImMunoGeneTics information system (http://www.imgt.org) listed 36 mAbs approved by the FDA or EMEA for human therapeutic use (Poiron *et al.*, 2010). Only 12 of the approved antibodies are of human origin, whereas the majority represents rodent (3), chimeric (7) or humanized antibodies (14), all containing non-human sequences. However, antibodies that are derived from non-human organisms and are applied in human therapies may lead to the human anti-mouse antibody response. Because of their foreign characteristics, they can lead to adverse and potentially harmful side-effects because of altered efficacy and pharmacokinetics (Schroff *et al.*, 1985; Shawler *et al.*, 1985; Sgro, 1995). This indicates the importance for techniques to reduce immunogenicity of antibodies by making them more human-like.

Traditional methods to reduce the risk of severe immunogenic responses to therapeutic antibodies are based on chimerization (Boulianne *et al.*, 1984; Morrison *et al.*, 1984; Neuberger *et al.*, 1985) or complementarity-determining region (CDR)-grafting (Jones *et al.*, 1986; Riechmann *et al.*, 1988; Queen *et al.*, 1989). Further advanced protocols include resurfacing (Roguska *et al.*, 1994), framework shuffling (Dall'Acqua *et al.*, 2005), human content optimization (Lazar *et al.*, 2007), superhumanization (Tan *et al.*, 2002; Hwang *et al.*, 2005), screening of human antibody libraries (Winter *et al.*, 1994; Low *et al.*, 1996; Bradbury *et al.*, 2011) or immunization of transgenic mice (Brüggemann *et al.*, 1991; Taylor *et al.*, 1992; Mendez *et al.*, 1997; Lonberg, 2005). The primary purpose of all of these methods is to keep the risk of adverse side-effects (Hansel *et al.*, 2010) at an absolute minimum. Applied to human subjects, the engineered therapeutic antibody must not trigger any critical human anti-mouse (Schroff *et al.*, 1985; Shawler *et al.*, 1985; Sgro, 1995) or human anti-chimeric (Khazaeli *et al.*, 1994) antibody responses, while the full biological function should be maintained, quantified by a high binding affinity.

However, extensive sequence modifications within the framework regions (FR) during the trial and error based humanization process often result in reduced or even lost binding affinities (Presta *et al.*, 1993; Adams *et al.*, 2006). This effect may be attributed to critical framework positions within the antibody framework sequence, which stabilize the overall protein structure or the V_H/V_L interface (Chothia *et al.*, 1989), contact the antigen directly (Mian *et al.*, 1991) or establish the Vernier zone (Foote and Winter,

* *Correspondence to: Chris Oostenbrink, Institute of Molecular Modeling and Simulation, University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna.*
E-mail: chris.oostenbrink@boku.ac.at

† *Both authors contributed equally to this work.*

a *C. Margreitter, C. Oostenbrink*
Institute of Molecular Modeling and Simulation, University of Natural Resources and Life Sciences, Muthgasse 18, Vienna 1190, Austria

b *P. Mayrhofer, R. Kunert*
Department of Biotechnology, University of Natural Resources and Life Sciences, Muthgasse 18, Vienna 1190, Austria

1992) by providing a suitable physico-chemical environment for a proper conformational ensemble of the CDR loops. In the first step of humanization, the non-human FR are replaced by carefully selected, appropriate human framework sequences. Afterwards, several critical positions within the human framework have to be backmutated to the non-human wild-type. Currently, no universally applicable humanization protocol is available that allows the straightforward, concurrent maintenance of the binding affinity and reduction of the risk for immunogenic responses, i.e. the lowest number of backmutations (BM) necessary.

These choices often have to be made based on empirical knowledge gained from iterative rounds of antibody design, expression and *in-vitro* binding evaluation on a case-by-case basis, making antibody humanization an unpredictable, time-consuming and costly undertaking. It would hence be highly advantageous to predict the effect of potential BM on the binding affinity of mutants, not only because the mere number of potential candidates is tremendous but also because there is an urgent need to understand the underlying physico-chemical mechanisms. Yet, the assessment of binding affinities (i.e. the free energy of binding) by computational tools remains a very demanding task. Docking lacks accuracy (mainly because of the imposed rigidity of bigger molecules), while free energy calculations using molecular dynamics (MD) simulations require structural data on the complex and are far from being readily applied to interactions involving a large molecular interface. Nevertheless, *in-silico* techniques may prove to be a useful addition to the humanization process. In this study, we perform MD simulations to analyse and predict CDR conformations in the humanization process of a mAb. By providing *in-silico* knowledge from MD simulations, proper decisions about critical BM can be made, before testing the designed variants on the bench. In such a prospective design cycle, many different humanized variants, containing different BM, might be assessed *in-silico*. The dynamic behaviour of the CDRs is monitored by simulation and expressed in a score that represents the similarity to the wild-type, the known binder. The most promising mutants are then selected for expression and measurement of binding affinities by experiments. With our technique, we allow for pre-selection of various humanized variants, thus reducing the amount of required experiments during humanization significantly.

In this study the anti-idiotypic antibody Ab2/3H6 directed against the broadly neutralizing anti-HIV-1 antibody 2 F5 (Muster *et al.*, 1993; Kunert *et al.*, 1998) was used as model protein. It was developed from mouse hybridoma (Kunert *et al.*, 2002) and subsequently chimerized (Gach *et al.*, 2007) or humanized by CDR-grafting, resurfacing or superhumanization (Mader and Kunert, 2010). Although not eliciting HIV-1 neutralizing antibodies in first prime/boost studies in BALB/c mice (Gach *et al.*, 2008) or rabbits (Kunert and Mader, 2011), it served as a template for different humanization approaches and MD simulations (de Ruiter *et al.*, 2011) based on the resolved crystal structure (Bryson *et al.*, 2008). The superhumanized variant, su3H6, has lost the binding affinity completely and is therefore suited to be the negative control for the simulation (Mader and Kunert, 2010). An antibody panel consisting of several humanized 3H6 mutants was tested for binding *in-vitro* to refine a similarity score, quantifying the similarity to the original wildtype antibody (wt3H6). The optimized *in-silico* system was then tested to predict the influence of BM on the binding affinity in superhumanized variants.

METHODS

Expression of mAbs

Cell cultures were cultivated in vented 125-ml Erlenmeyer flasks (Corning) on a climo-shaker ISF1-XC (Kuhner) at 140 rpm, 37 °C, 7% CO_2 and 85% humidity. mAb variants used for training of the MD system (TR01-TR06) were expressed using stable transfected cell pools of a serum-free adapted host cell line CHO-K1 (ATCC CCL-61) cultivated in ProCHO5 medium (Lonza, Cat. No. BE12-766Q) supplemented with 4 mM L-glutamine (Biochrom, Cat. No. K0302), 15 mg/l phenol red (Sigma, Cat. No. P0290) and 0.5 mg/ml G418 (Biochrom, Cat. No. A2912).

To study the effect of BM in heavy chain FR of the superhumanized Ab2/3H6 variants, transient expression was performed in HEK293-6E cells (NRC biotechnology Research Institute) (Durocher *et al.*, 2002) cultured in F17 medium (Invitrogen, Life Technologies, Cat. No. A13835-01) supplemented with 4 mM L-glutamine (Biochrom, Cat. No. K0302), 0.1% Kolliphor P188 (Sigma-Aldrich, Cat. No. K4894), 15 mg/l phenol red (Sigma-Aldrich, Cat. No. P0290) and 25 μg/ml G418 (Biochrom, Cat. No. A2912). Transfection of pCEP4 vector (Invitrogen, Cat. No. V044-50) was mediated by polyethylenimine (PEI) transfection using linear 25-kDa PEI (Polysciences, Cat. No. 23966).

Culture supernatants were subjected to concentration by either using Amicon Ultra Centrifugal Filters (0.5 ml, NMWCO 10 kDa, Millipore, Cat. No. UFC501096) or Millipore-Labscale TFF system (Millipore), equipped with a 30-kDa Pellicon cassette (Millipore, Cat. No. PXB030A50) followed by antibody purification by protein A affinity chromatography using the ÄKTA Purifier Station (GE Healthcare) equipped with a HiTrap MabSelect SuRe protein A column (GE Healthcare, Cat. No. 29-0491-04).

Preparation of mAb variants

The mAb variants used for training of the MD system were expressed in CHO-K1 cells with stable cell pools and purified by protein A chromatography (TR01 – TR06).

For studying BM in FR of the heavy chain, transient expression was performed in HEK293-6E cells for different superhumanized Ab2/3H6 variants to allow assessment of re-established binding affinities. After a 7-day production phase, crude culture supernatants were concentrated and diluted in ForteBio kinetics buffer.

Affinity evaluation of mAb variants

All binding studies based on bio-layer interferometry were performed on a ForteBio Octet QK system (Pall ForteBio) equipped with streptavidin (Pall ForteBio, Cat. No. 18-5019) or protein A biosensors (Pall ForteBio, Cat. No. 18-5010).

For immobilization to streptavidin biosensors, purified antibody 2 F5 was biotinylated using the EZ-Link NHS-PEG4-Biotin kit (Thermo Scientific, Cat. No. 21329). Samples and biosensors were equilibrated in kinetics buffer (ForteBio). The streptavidin biosensors were loaded with 20 μg/ml biotinylated 2 F5 antibody, before k_{on} and k_{off} were measured of purified mAb variants by monitoring association and dissociation in kinetics buffer.

Capturing mAbs from crude culture supernatants for quantification already has been proven as an established and robust procedure (Tobias and Kumaraswamy, 2014). In this work we further demonstrate the use of protein A biosensors to determine the binding affinity of the immobilized antibodies to its

antigen (anti-idiotypic antibody). Protein A sensors were equilibrated in kinetics buffer for 60 s before transiently expressed Ab2/3H6 variants were immobilized from the crude and concentrated culture supernatants for 1200 s (Figure 1A). This time period for capturing antibodies from crude culture supernatants was considered as a good trade-off between reaching a suitable sensor saturation level and overall assay time. To block possible free protein A binding sites a blocking step of 1200 s was introduced by submersing the loaded protein A biosensors in high concentration (100 μg/ml) of purified mAb su3H6, showing no interaction with 2 F5. Following a baseline/washing step for 120 s, the association of purified target antibody mAb 2 F5 (100 μg/ml) to protein A-immobilized Ab2/3H6 variants was measured followed by a dissociation step in kinetics buffer only. mAb 2 F5 binding to anti-idiotypic mutants showed a high response for wt3H6 samples of about 3.5 nm in the baseline corrected sensorgram (Figure 1B). Both crude and pure wt3H6 preparations at high concentrations gave similar binding curves. Additionally, also with a very low initial wt3H6 concentration (1.5 μg/ml) a high binding response could be observed reaching a response level of 3.4 nm. These results demonstrate that it is possible to distinguish mAb 2 F5-binders (i.e. wt3H6) from non-binders (i.e. su3H6), the latter of which gave only a very low non-specific response level probably resulting from minimal residual free protein A binding sites. From the experimentally obtained K_d values, the average binding free energies were calculated to be compared to the score obtained from the simulations. To eliminate outliers, data sets have been excluded for which the fit to the theoretical signal curve was calculated to have a correlation coefficient (R^2) below or equal to 0.8. Afterwards, binding free energies were calculated using $\Delta G_{binding} = k_B T \ \ln(K_d)$ for all remaining measurements. Measurements, deviating more than 5.6 kJ/mol from the average (factor 10 in K_d) were iteratively excluded from the calculation. From the remaining values (at least three measurements for each variant), the average binding free energy is reported.

In-silico score calculation

The approach presented here relies solely on the structural and dynamic information retrieved from the monoclonal antibody, as shown in Figure 2. From multiple simulations on the murine antibody, we obtain the most prominent conformations of the CDRs, represented by the central member structures (CMS) of conformational clusters. Subsequently, variants are simulated, and the reproduction of the wild-type reference conformations (CMS) is expressed through a similarity score. It is based on time series of the root-mean-square deviation (RMSD) of the CDR atoms (fitted to the flanking framework backbone; see below) with respect to the wild-type CMS. This means that the score is higher for variants, which are closer to the original rodent antibody in terms of their structural ensembles. The score is calculated according to,

$$\text{score} = \frac{100}{s \cdot m \cdot a} \sum_{i=1}^{a} \sum_{j=1}^{s} \sum_{k=1}^{n} \sum_{x=1}^{m} \begin{cases} 1, & \text{RMSD}_{x,\text{CMS}_i} \leq c_k \\ 0, & \text{else} \end{cases} \quad (1)$$

where c is a vector of thresholds used, $RMSD_{x,CMS}$ is calculated for

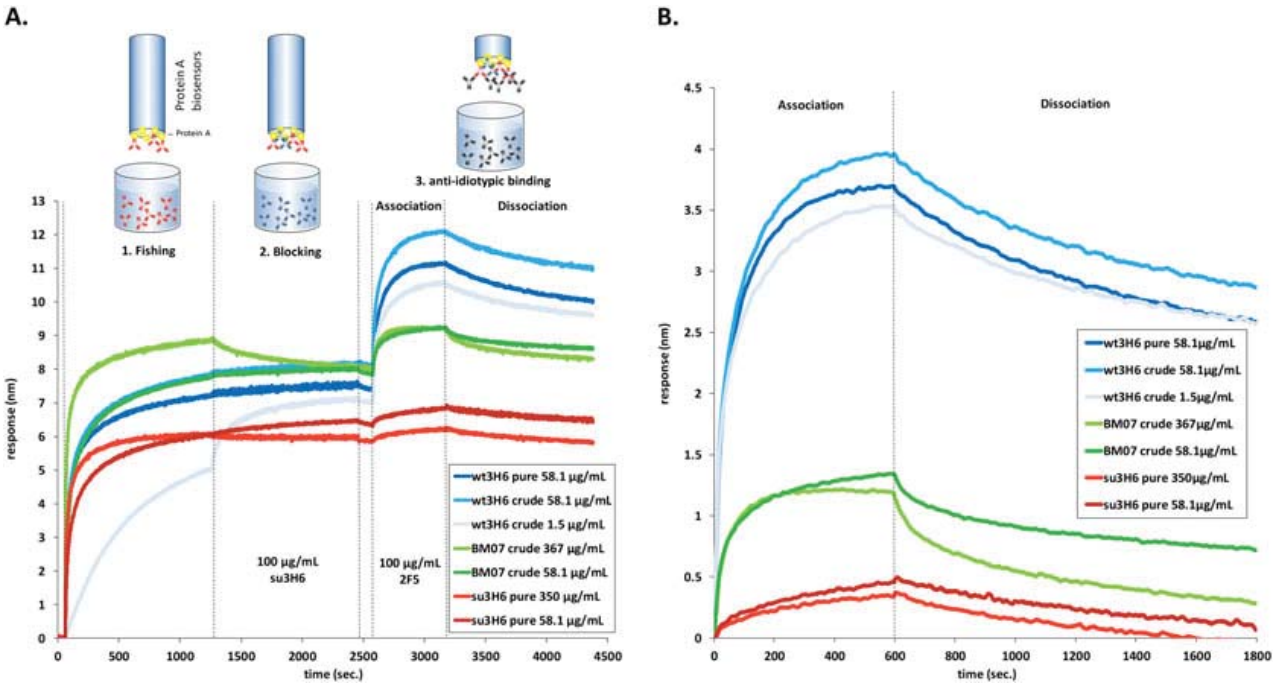


Figure 1. Protein A fishing from crude culture supernatants by bio-layer interferometry. The ForteBio Octet system was equipped with protein A biosensors to immobilize transiently expressed humanized Ab2/3H6 mutants from concentrated and crude culture supernatants. (A) Real-time sensorgram of wt3H6, su3H6 and BM07 at different concentrations. Assay-step times were as follows: 60-s baseline in kinetics buffer, 1. Fishing: 1200-s immobilization of antibodies from crude culture supernatants, 2. Blocking with 100 μg/ml purified mAb su3H6 for 1200 s, followed by 120-s baseline/washing in kinetics buffer, 3. Binding measurements: 600-s mAb 2 F5 association (100 μg/ml) with immobilized Ab2/3H6 variants, followed by 1200-s dissociation in kinetics buffer only. (B) Association and dissociation curves extracted from raw data and aligned to baseline by the fortebio software.

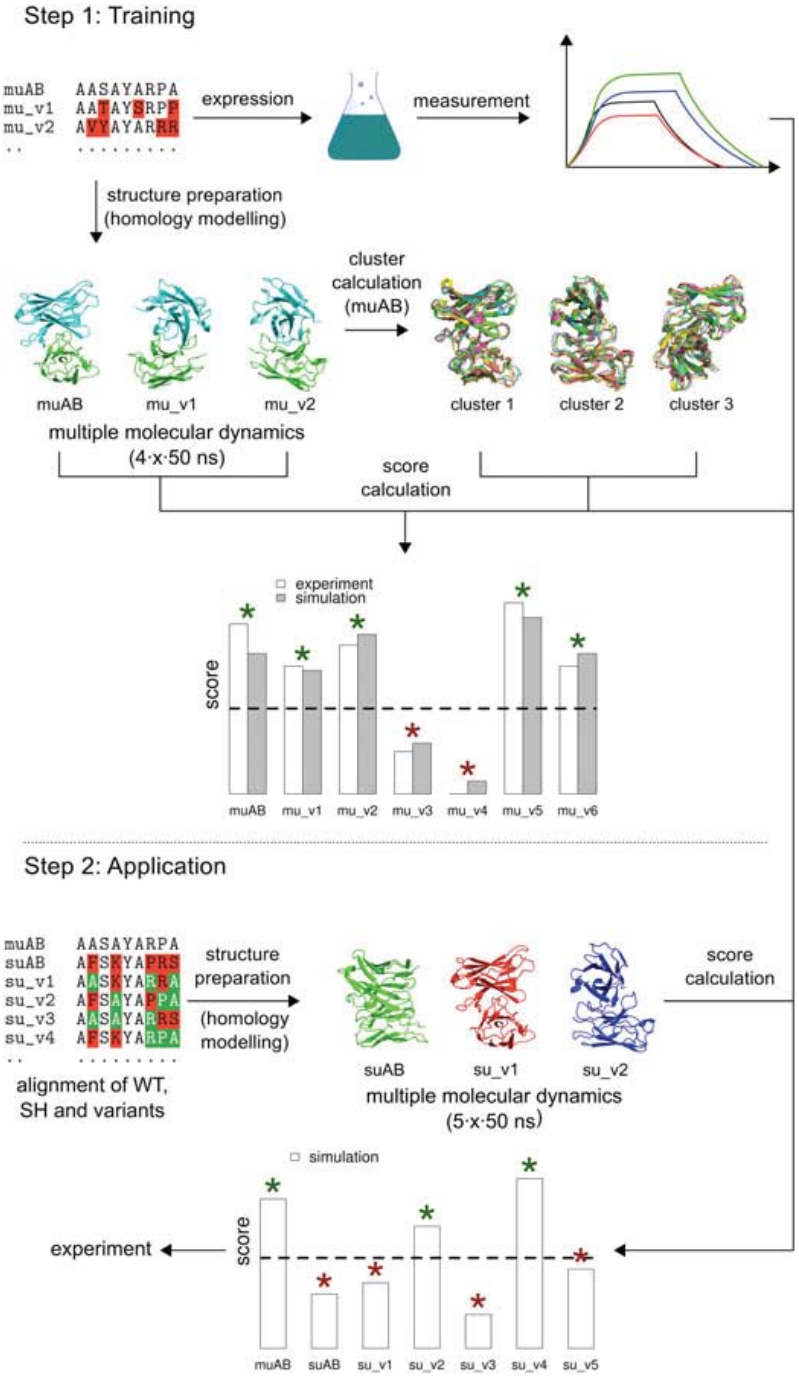


Figure 2. Workflow of the simulation assisted humanization approach. In the training step (above), molecular dynamics simulations of the murine derived wild-type and selected mutant variants are performed and assessed in terms of a score (see equation 1), representing similarity to the wild-type loop conformations. The same set is expressed, and affinities are measured experimentally, allowing for the identification of the qualitative boundary separating binders from non-binders (the latter marked with a red asterisk). The second step (below) starts with the superhumanized antibody variant, and a set with selected backmutations. With the same procedure as before, a computational score can be calculated holding qualitative information on the expected binding affinities. The cutoff determined in the previous step can now be used to classify the results. Note, that this is only an illustrative diagram; the actual values obtained throughout this study are reported in Figures 5 and 6 and Table 1.

a given configuration–CMS pair (in nanometers), *m* is the number of configurations in a trajectory, *n* is the number of thresholds considered, *a* is the number of significant clusters and *s* is the number of replicate simulations for this very variant. In an initial training round, the scores are compared to the binding free energy, experimentally determined by affinity measurements for some variants, to estimate a cutoff of the similarity score. In the second stage, BM of the superhumanized variant are simulated until a candidate with a score above the cutoff is identified, which can then be further optimized. Our approach

is based on the assumption that mutants with comparable structures and dynamics as the original monoclonal antibody also show significant binding. Obviously, the reverse statement is not necessarily true as other conformations/ensembles may bind as well or even better but are disregarded in our approach because they were not present in the murine reference. Furthermore, we assume that induced fit effects upon binding play a minor role and the relevant pre-binding conformations will be sufficiently sampled in the MD simulations, following the conformational selection paradigm (Lee and Craik, 2009; Vogt and Di Cera, 2013).

We have applied this workflow to the murine/wild-type anti-idiotypic Ab2/3H6 antibody, which is directed against the broadly neutralizing anti-HIV-1 antibody 2 F5 and has been studied by our groups earlier (Kunert *et al.*, 1998; Kunert *et al.*, 2002; Mader and Kunert, 2010; de Ruiter *et al.*, 2011). The binding to mAb 2 F5 is mainly facilitated by the third CDR loop of the heavy chain (Ab2/3H6), which simplifies the analysis. In order to cover cases that require multiple loops for proper binding, the above equation can readily be extended. We defined six training variants based on critical framework positions (Supplementary Table S1) of the murine/wild-type Ab2/3H6 (TR01 to TR06; Figure 3), to train the scoring function and 11 backmutation variants based on the non-binding su3H6 antibody (BM01 to BM11; Figure 4), with mutation sites in the FRs selected based on their location in the X-ray structure, vicinity to the Vernier zone, or being in the V_H/V_L interface region.

MD simulations

For wt3H6 (the binding reference), six simulations (of 100 ns each) of the V_H/V_L complex were undertaken. For the compounds in the training set, four simulations (replicates) for each variant were performed with a trajectory length of 50 ns, respectively. The superhumanized version and its derivatives have been simulated five times each (50 ns). Initial coordinates were taken from the crystal structure of the 3H6–2F5 complex (PDB ID: 3BQU) (Bryson *et al.*, 2008); variants were modelled using the programme MOE (MOE, 2014). The simulations were performed without any restraints. All simulations were performed

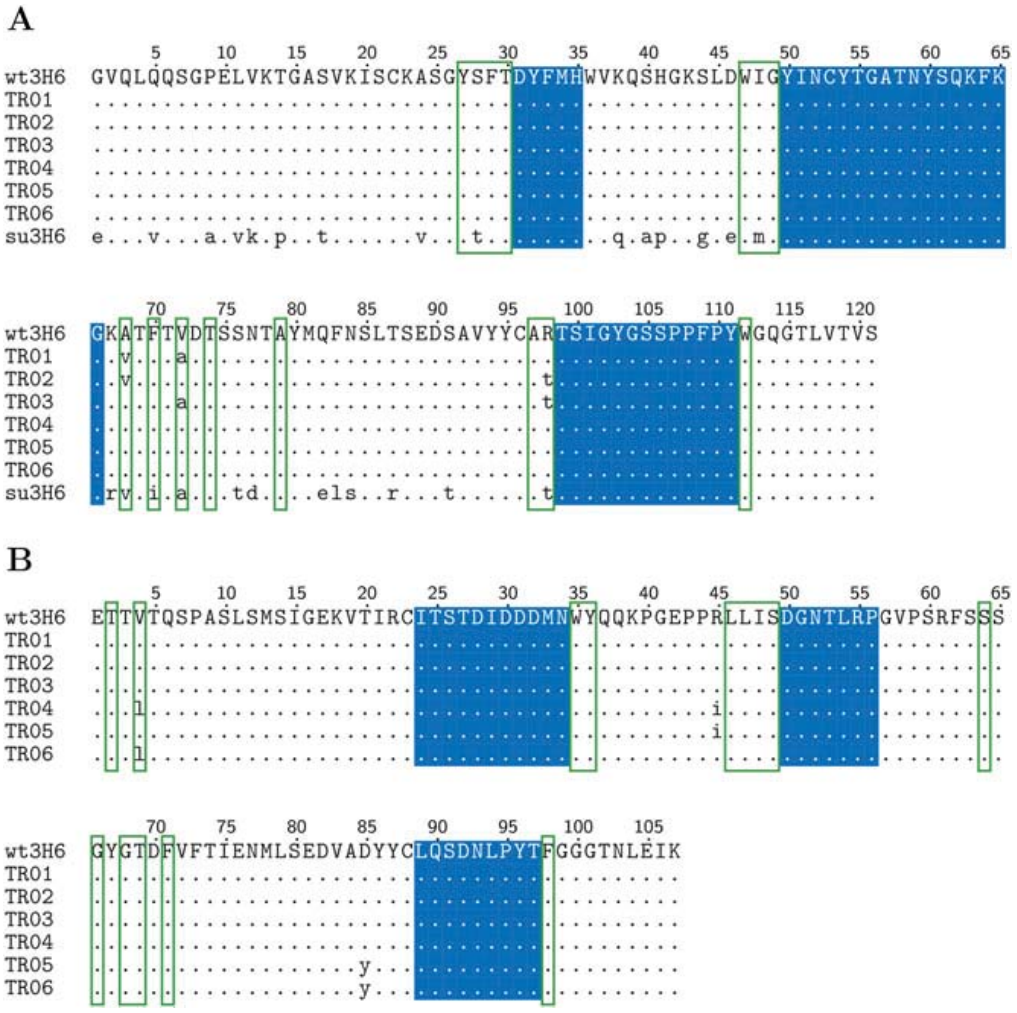


Figure 3. Sequence alignment of the heavy (A) and light (B) chain of the wild-type (wt3H6), the training (TR01–TR06) and the superhumanized (su3H6) variant. The CDR regions as defined by Kabat were conserved during the humanization process and kept constant in all variants throughout this study (blue background). Identical amino acids compared to the murine/wild-type template sequence are indicated as dots. Vernier zone residues of the heavy chain defined by Foote and Winter are marked by green frames.

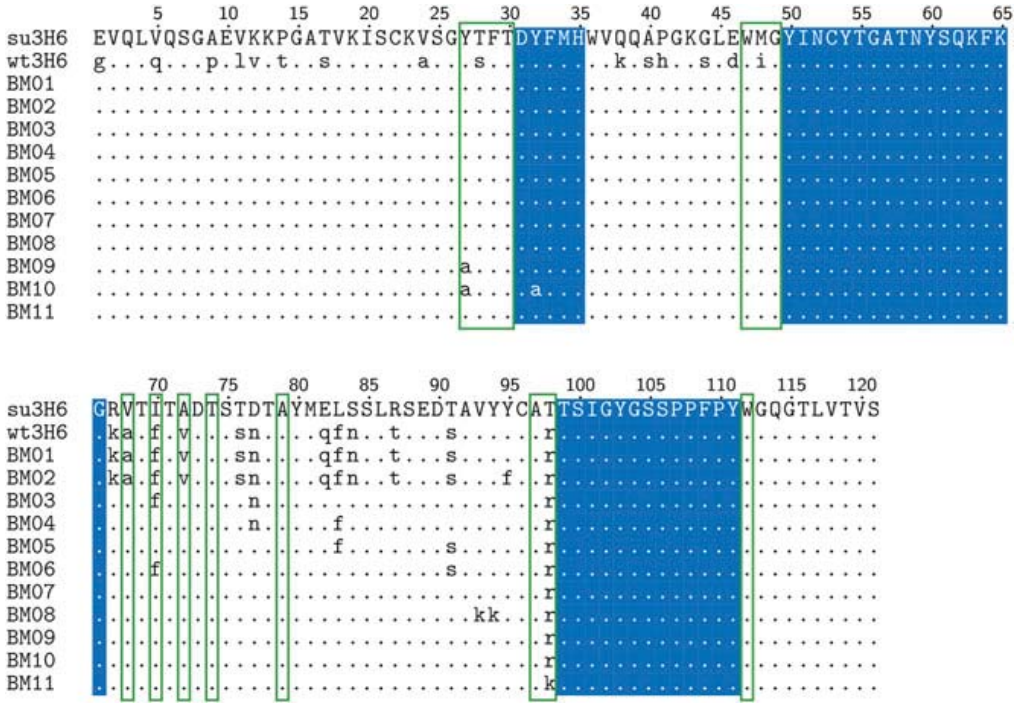


Figure 4. Sequence alignment of the heavy chain of the wild-type (wt3H6), the superhumanized (su3H6) and the simulated (predictive, BM01–BM11) variants. The CDR regions as defined by Kabat were conserved during the humanization process and kept constant in all variants (but BM10) throughout this study (blue background). Identical amino acids compared to the murine template sequence are indicated as dots. Vernier zone residues of the heavy chain defined by Foote and Winter are marked by green frames. No light chain mutations in respect to the superhumanized variant were applied to the backmutation variants.

using the GROMOS (Christen *et al.*, 2005; Schmid *et al.*, 2012) simulation package with the 54A7 parameter set (Oostenbrink *et al.*, 2004; Schmid *et al.*, 2011) in a sufficiently large water-box (0.8-nm minimum solute to box–wall distance). Counter-ions were added (Na^+ and Cl^-) to neutralize the net charge in the box, up to a limit of 15 ions of each type, e.g. to a solute with net charge -9 *e*, 15 Na^+ and 6 Cl^- were added. The rectangular periodic simulation boxes (roughly $5.5 \times 6.5 \times 7.5$ nm) contained approximately 27 000 atoms. Prior to the production runs, the systems were equilibrated from 60 K to 300 K in six discrete steps with a simulation length of 20 ps each and a weak thermostat-coupling with two baths for the solute and solvent (relaxation time of 0.1 ps) and weak barostat-coupling (relaxation time of 0.5 ps and an isothermal compressibility of 4.575×10^{-4} $(kJ\,mol^{-1}\,nm^{-3})^{-1}$). In order to avoid artifacts originating from the same starting structure, only the last 40 ns of each trajectory has been analysed, while for cluster analysis the last 90 ns of the wt3H6 replicates has been used. All simulations in the training set were extended to 100 ns (90 ns analysed) without a significant change in the subsequent analyses (Supplementary Figure S1); therefore, an additional replicate was preferred over longer simulation times in the prediction set. For consistency with the superhumanized and predicted variants, all the reported data is based on the last 40 ns of the first 50 ns of the training set simulations. Weak temperature and pressure coupling (Berendsen *et al.*, 1984) ensured a constant temperature of 300 K and a constant pressure of 1 atm, respectively. SHAKE (Ryckaert *et al.*, 1977) was used to maintain the bond distances at the energy minimum. The integration time step used was 2 fs. Interactions within 0.8 nm were calculated at every time step from a pairlist that was updated every five steps. Intermediate range interactions up to a distance of 1.4 nm were calculated at pairlist updates and kept constant between updates. Long range interactions were approximated with a reaction field contribution (Tironi *et al.*, 1995) to the energies and forces, accounting for a homogeneous medium with a relative dielectric constant (Heinz *et al.*, 2001) of 61 beyond the cut-off of 1.4 nm.

Fitting procedure

In order to compare the conformational ensembles generated by the MD simulations to one another, the respective backbone atoms of FR 3 and 4 of the heavy chain (H:FR3 and H:FR4) were used for a roto-translational least-squares fit. The RMSD calculation afterwards (both for the clustering and the score calculation) was based on all atoms (including side-chain atoms) of complementary determining region 3 of the heavy chain (H:CDR3).

RESULTS AND DISCUSSION

All MD simulations lead to stable trajectories with backbone atom-positional RMSD values to the initial structure of, on average, 0.3 (±0.1) nm for the training set and 0.3 (±0.1) nm for the superhumanized variants (Supplementary Figures S2 and S3). The secondary structure elements of the FR (anti-parallel beta sheets) were maintained throughout. Averaged over simulation time and the FR residues, which are also in β-sheet conformation in the crystal structure, the occurrence of β-sheet conformations amounted to 88% (±2%) for the training set and 83% (±4%) for the superhumanized variants as determined by

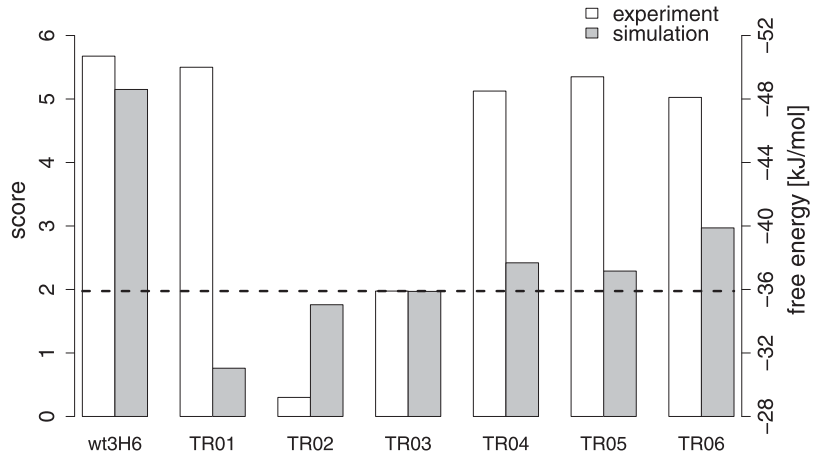


Figure 5. Similarity scores and free energies of binding for the variants. For the calculation of the score, see equation 1. The grey bars indicate a score for the similarity to the murine/wild-type 3H6 antibody (left axis), while the white bars indicate experimentally determined binding free energies (right axis). For TR01, the simulation score and the experimentally determined binding free energy are clearly disagreeing, see main text.

the determine secondary structure of proteins (DSSP) algorithm (Supplementary Figure S4). To calculate the reference structure (s), that are most representative of the conformational ensemble of H:CDR3 in wt3H6, we calculated the cross RMSD matrix of structures collected from all its replicates. A clustering algorithm was applied, using a cut-off of 0.2 nm (Daura *et al.*, 1999). The majority of configurations belonged to the first cluster (52%), while the others were only populated by small amounts (up to 15%), suggesting to use a single representative structure, i.e. a = 1 in equation 1. Without loss of generality, the subsequent analysis could have included additional clusters. Subsequently, the similarity score was computed for the training set, which is shown in Figure 5 together with the experimentally determined binding free energies. From the MD simulations, TR01, TR02 and TR03 were most dissimilar to the original wt3H6. Indeed, for TR02 and TR03 the binding affinity seems to be reduced by 15 – 20 kJ/mol, while the mutations applied to TR04 to TR06 do not seem to affect affinity. Only for TR01, there is no match between the similarity score and the measured binding affinity to 2 F5, possibly because of alternative binding modes (see above).

A superhumanized variant (su3H6) was described earlier, which lost binding affinity completely (Figure 6) (Mader and Kunert, 2010). Our MD simulations confirm that the CDR is significantly different (similarity score of 1.0). Eleven candidates with specific BM (BM01 to BM11) were proposed and are represented in Figure 4. The associated similarity scores for these variants are given in Table 1. Based on the results of the training simulations, the emphasis for the BM was placed on the FR 3 of the heavy chain. In the first variant, BM01, the entire FR was backmutated to the murine/wild-type sequence. The similarity score was

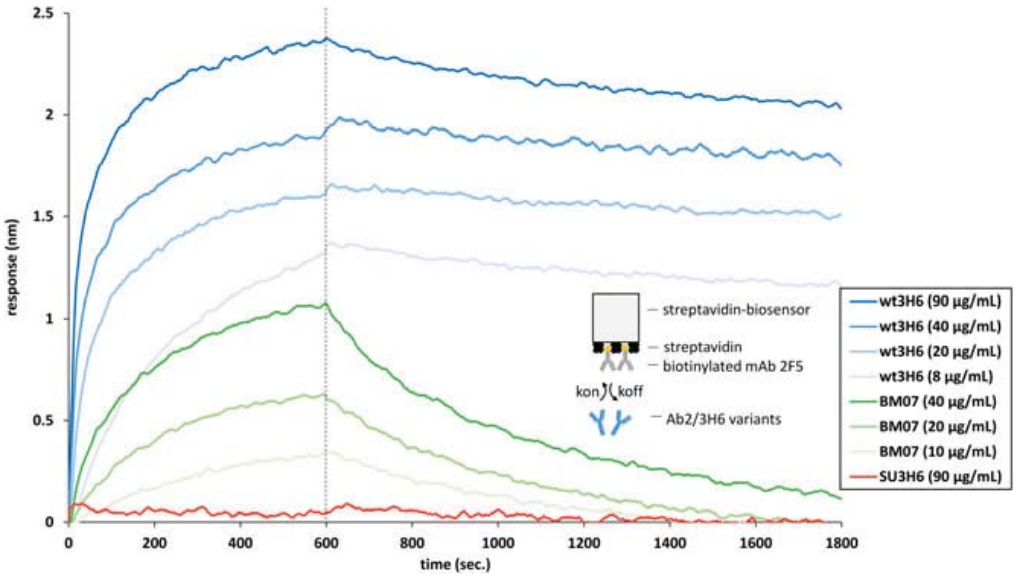


Figure 6. Real-time bio-layer interferometry sensorgrams for determination of anti-idiotypic binding affinities of purified Ab2/3H6 variants to its target antibody mAb 2 F5. Streptavidin biosensors were loaded with biotinylated mAb 2 F5 (20 µg/ml) followed by a washing/baseline step. Association (600 s) and dissociation (1200 s) of different Ab2/3H6 variants were measured at different concentrations or buffer only, respectively.

Table 1. *In-silico* score predictions for all simulated variants based on the (non-binding) su3H6 antibody

Variant	Score
su3H6	1.0
BM01	2.6
BM02	3.3
BM03	3.5
BM04	1.7
BM05	1.8
BM06	2.8
BM07	2.8
BM08	1.3
BM09	2.9
BM10	0.8
BM11	3.2

significantly lower than for wt3H6, indicating that BM01 is not a variant that is to be tested experimentally. Interestingly, variant BM02, which contains the additional mutation Y95F, leads to a significantly higher similarity score. The side-chain of Y95 is pointing into the interface region between the V_L and V_H domains, suggesting that the packing of amino acids at the interface and in the hydrophobic core plays a crucial role. To maintain optimal packing with the human residues from the other FR, variants with fewer BM were proposed in the form of BM03 to BM07. The most interesting variant in this respect is variant BM07, in which only a single backmutation, T98R, was applied and for which a reasonably high similarity score was observed. Next, we proposed an artificial double mutation to lysine at positions 93 and 94 in variant BM08 to have a non-binding variant (negative control) that still contains the T98R mutation. Indeed, the similarity score dropped considerably.

Variant BM07 was transiently expressed using the HEK293–6E system. Because of low transient expression titers the affinity to 2 F5 was measured by a modified protein A fishing setup (see above): Protein A biosensors were immersed in the supernatant to bind the transiently expressed variant. After blocking the remaining binding sites of the sensors with inactive su3H6, the antigen 2 F5 was added, and binding and dissociation could be measured. This superhumanized variant, containing only a single backmutation (BM07), showed a significant improvement in anti-idiotypic binding affinity to mAb 2 F5 resulting in final response levels of 1.4 nm at two different concentrations. Although the binding affinity of BM07 did not reach the full binding capabilities of wt3H6, it showed a significant increase with respect to its precursor molecule, i.e. the non-binding su3H6 antibody (Figure 1B). Based on these qualitative results we expressed BM07 on a larger scale followed by protein A chromatography purification and quantitative assessment of its binding properties using the streptavidin bio-assay setup (Figure 6). The results of the two different methods were qualitatively comparable ($\Delta G_{binding} = -43.5$ kJ/mol for the first and $\Delta G_{binding} = -38.7$ kJ/mol for the latter), indicating that our method for quickly estimating binding affinities from the supernatant is reasonable. Unfortunately, difficulties with expression efficiency have so far prevented the validation of additional superhumanized variants.

Both our simulations and the experiments show, that the backmutation T98R is sufficient to restore binding affinity to a significant extent (Figure 7). This confirms that a single backmutation can be sufficient to (partially) restore binding affinity for superhumanized variants and also indicates a crucial role for R98, which could be successfully predicted by our *in-silico* approach. In fact, 80% of the human heavy chain germline gene sequences retrieved from the IMGT/Gene-DB databank (Lefranc *et al.*, 2005) have an Arg at this very position (Supplementary Table S2). By replacing this residue by threonine, as in su3H6, TR02 and TR03, binding affinity is severely reduced. From the simulations we observe that the charged side chain of R98 interacts with the Y27, Y32, Y111 side-chains (in some sort of "tyrosine-cage") and the T99 backbone (Figure 8). It seems, that through the cation–π interactions in this structural region, certain conformational

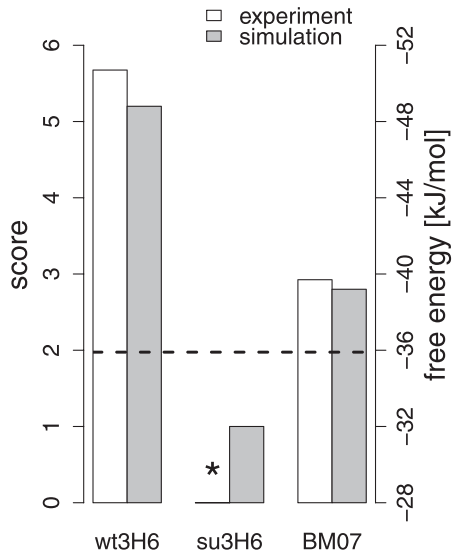


Figure 7. Similarity scores and free energies of binding for the wild-type, the superhumanized antibody and BM07. The experimental binding free energy of su3H6 was below the detection limit.

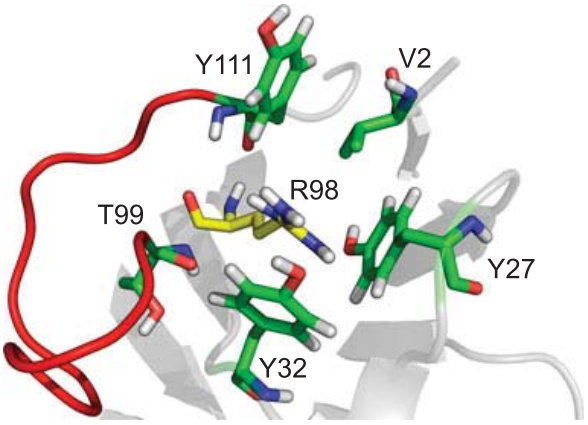


Figure 8. The importance of R98 is likely to be explained by its interaction with the surrounding tyrosines through cation–pi interactions, most notably with one located in CDR loop three (Y111). Therefore, it is of utmost importance that this position remains an arginine in order to retain binding affinity. The "tyrosine cage" at that position could, in conjunction with R98, lead to a more restricted local environment for the CDR loop, which is shown in red.

restrictions are applied to the CDR3 loop (heavy chain), which are crucial for binding. To test this hypothesis, we also investigated other variants computationally (BM09 to BM11; Table 1, Figure 4) which will be the subject of future experimental validation studies. It is remarkable that the double mutation of Y27 and Y32 (BM10) to alanine is predicted to lead to a severe loss of binding, while the single mutation of Y27 (BM09) seems to make no difference. Moreover, a substitution of R98 by lysine (BM11), which also contains a positively charged moiety at a comparable distance to the backbone, seems to be able to function appropriately.

CONCLUSION

We conclude that the presented approach, calibrated with experimental data, allows for useful predictions of the effect that distinct BM have on the binding affinities of antibody variants. We have validated our predictions using well-established experimental techniques and also shown the qualitative agreement of these results with a newly developed efficient method, based on cell culture supernatant. Our computational workflow can be applied as an ab-initio protocol; however, additional information, such as the relative importance of the respective CDR loops in binding, might be included. Moreover, simulations allow for a molecular rationalization of the observed differences, which may help to guide further rounds of compound design where necessary. The approach is readily applicable to different antibodies as only structural information on the original antibody is required and no explicit binding free energy calculations are performed. Presumably, iterative cycles of improvement will be necessary to re-establish a binding affinity, comparable to that of the wild-type, where promising combinations of the previous round are natural candidates.

Acknowledgements

Financial support of the European Research Council (ERC; grant number 260408), the Austrian Science Fund (FWF; grant number P 25056 and project W1224) and the Vienna Science and Technology Fund (WWTF; grant number LS08-QM03) is gratefully acknowledged.

REFERENCES

Adams, CW, Allison, DE, Flagella, K, Presta, L, Clarke, J, Dybdal, N, McKeever, K, Sliwkowski, MX. 2006. Humanization of a recombinant monoclonal antibody to produce a therapeutic HER dimerization inhibitor, pertuzumab. *Cancer Immunol. Immunother. CII* **55**: 717–727. DOI:10.1007/s00262-005-0058-x.

Aggarwal, R. 2014. What's fueling the biotech engine—2012 to 2013. *Nat. Biotechnol.* **32**: 32–39. DOI:10.1038/nbt.2794.

Berendsen, HJC, Postma, JPM, van Gunsteren, WF, DiNola, A, Haak, JR. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**: 3684–3690. DOI:10.1063/1.448118.

Boulianne, GL, Hozumi, N, Shulman, MJ. 1984. Production of functional chimaeric mouse/human antibody. *Nature* **312**: 643–646.

Bradbury, ARM, Sidhu, S, Dübel, S, McCafferty, J. 2011. Beyond natural antibodies: the power of in vitro display technologies. *Nat. Biotechnol.* **29**: 245–254. DOI:10.1038/nbt.1791.

Brüggemann, M, Spicer, C, Buluwela, L, Rosewell, I, Barton, S, Surani, MA, Rabbitts, TH. 1991. Human antibody production in transgenic mice: expression from 100 kb of the human IgH locus. *Eur. J. Immunol.* **21**: 1323–1326. DOI:10.1002/eji.1830210535.

Bryson, S, Julien, J-P, Isenman, DE, Kunert, R, Katinger, H, Pai, EF. 2008. Crystal structure of the complex between the F(ab)′ fragment of the cross-neutralizing anti-HIV-1 antibody 2 F5 and the F(ab) fragment of its anti-idiotypic antibody 3H6. *J. Mol. Biol.* **382**: 910–919. DOI:10.1016/j.jmb.2008.07.057.

Chothia, C, Lesk, AM, Tramontano, A, Levitt, M, Smith-Gill, SJ, Air, G, Sheriff, S, Padlan, EA, Davies, D, Tulip, WR. 1989. Conformations of immunoglobulin hypervariable regions. *Nature* **342**: 877–883. DOI:10.1038/342877a0.

Christen, M, Hunenberger, PH, Bakowies, D, Baron, R, Burgi, R, Geerke, DP, Heinz, TN, Kastenholz, MA, Krautler, V, Oostenbrink, C, Peter, C, Trzesniak, D, van Gunsteren, WF. 2005. The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **26**: 1719–1751. DOI:10.1002/jcc.20303.

Dall'Acqua, WF, Damschroder, MM, Zhang, J, Woods, RM, Widjaja, L, Yu, J, Wu, H. 2005. Antibody humanization by framework shuffling. *Methods San Diego Calif.* **36**: 43–60. DOI:10.1016/j.ymeth.2005.01.005.

Daura, X, van Gunsteren, WF, Mark, AE. 1999. Folding–unfolding thermodynamics of a β-heptapeptide from equilibrium simulations. *Proteins* **34**: 269–280. DOI:10.1002/(SICI)1097-0134(19990215)34:3<269::AID-PROT1>3.0.CO;2-3.

Durocher Y, Perret S, Kamen A. 2002. High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res.* **30**. DOI: 10.1093/nar/30.2.e9. http://nar.oxfordjournals.org/content/30/2/e9

Foote, J, Winter, G. 1992. Antibody framework residues affecting the conformation of the hypervariable loops. *J. Mol. Biol.* **224**: 487–499.

Gach, JS, Quendler, H, Weik, R, Katinger, H, Kunert, R. 2007. Partial humanization and characterization of an anti-idiotypic antibody against monoclonal antibody 2 F5, a potential HIV vaccine? *AIDS Res. Hum. Retroviruses* **23**: 1405–1415. DOI:10.1089/aid.2007.0089.

Gach, JS, Quendler, H, Strobach, S, Katinger, H, Kunert, R. 2008. Structural analysis and in vivo administration of an anti-idiotypic antibody against mAb 2 F5. *Mol. Immunol.* **45**: 1027–1034. DOI:10.1016/j.molimm.2007.07.030.

Hansel, TT, Kropshofer, H, Singer, T, Mitchell, JA, George, AJT. 2010. The safety and side effects of monoclonal antibodies. *Nat. Rev. Drug Discov.* **9**: 325–338. DOI:10.1038/nrd3003.

Heinz, TN, van Gunsteren, WF, Hunenberger, PH. 2001. Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations. *J. Chem. Phys.* **115**: 1125–1136. DOI:10.1063/1.1379764.

Hwang, WYK, Almagro, JC, Buss, TN, Tan, P, Foote, J. 2005. Use of human germline genes in a CDR homology-based approach to antibody humanization. *Methods San Diego Calif.* **36**: 35–42. DOI:10.1016/j.ymeth.2005.01.004.

Jones, PT, Dear, PH, Foote, J, Neuberger, MS, Winter, G. 1986. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* **321**: 522–525. DOI:10.1038/321522a0.

Khazaeli, MB, Conry, RM, LoBuglio, AF. 1994. Human immune response to monoclonal antibodies. *J. Immunother. Emphas. Tumor Immunol. Off. J. Soc. Biol. Ther.* **15**: 42–52.

Kunert, R, Mader, A. 2011. Anti-idiotypic antibody Ab2/3H6 mimicking gp41: a potential HIV-1 vaccine? *BMC Proc.* **5**: P64. DOI:10.1186/1753-6561-5-S8-P64.

Kunert, R, Rüker, F, Katinger, H. 1998. Molecular characterization of five neutralizing anti-HIV type 1 antibodies: identification of nonconventional D segments in the human monoclonal antibodies 2G12 and 2 F5. *AIDS Res. Hum. Retroviruses* **14**: 1115–1128.

Kunert, RE, Weik, R, Ferko, B, Stiegler, G, Katinger, H. 2002. Anti-idiotypic antibody Ab2/3H6 mimics the epitope of the neutralizing anti-HIV-1 monoclonal antibody 2 F5. *AIDS Lond. Engl.* **16**: 667–668.

Lazar, GA, Desjarlais, JR, Jacinto, J, Karki, S, Hammond, PW. 2007. A molecular immunology approach to antibody humanization and functional optimization. *Mol. Immunol.* **44**: 1986–1998. DOI:10.1016/j.molimm.2006.09.029.

Lee, GM, Craik, CS. 2009. Trapping moving targets with small molecules. *Science* **324**: 213–215. DOI:10.1126/science.1169378.

Lefranc, M-P, Giudicelli, V, Kaas, Q, Duprat, E, Jabado-Michaloud, J, Scaviner, D, Ginestoux, C, Clément, O, Chaume, D, Lefranc, G. 2005.

IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res.* **33**: D593–D597. DOI:10.1093/nar/gki065.
Lonberg, N. 2005. Human antibodies from transgenic animals. *Nat. Biotechnol.* **23**: 1117–1125. DOI:10.1038/nbt1135.
Low, NM, Holliger, PH, Winter, G. 1996. Mimicking somatic hypermutation: affinity maturation of antibodies displayed on bacteriophage using a bacterial mutator strain. *J. Mol. Biol.* **260**: 359–368.
Mader, A, Kunert, R. 2010. Humanization strategies for an anti-idiotypic antibody mimicking HIV-1 gp41. *Protein Eng. Des. Sel. PEDS* **23**: 947–954. DOI:10.1093/protein/gzq092.
Mendez, MJ, Green, LL, Corvalan, JR, Jia, XC, Maynard-Currie, CE, Yang, XD, Gallo, ML, Louie, DM, Lee, DV, Erickson, KL, Luna, J, Roy, CM, Abderrahim, H, Kirschenbaum, F, Noguchi, M, Smith, DH, Fukushima, A, Hales, JF, Klapholz, S, Finer, MH, Davis, CG, Zsebo, KM, Jakobovits, A. 1997. Functional transplant of megabase human immunoglobulin loci recapitulates human antibody response in mice. *Nat. Genet.* **15**: 146–156. DOI:10.1038/ng0297-146.
Mian, IS, Bradwell, AR, Olson, AJ. 1991. Structure, function and properties of antibody binding sites. *J. Mol. Biol.* **217**: 133–151.
MOE. 2014. Molecular operating environment (MOE). Chemical Computing Group Inc.: 1010 Sherbrook St. West, Suite #910, Montreal, QC, Canada.
Morrison, SL, Johnson, MJ, Herzenberg, LA, Oi, VT. 1984. Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *Proc. Natl. Acad. Sci. U. S. A.* **81**: 6851–6855.
Muster, T, Steindl, F, Purtscher, M, Trkola, A, Klima, A, Himmler, G, Rüker, F, Katinger, H. 1993. A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1. *J. Virol.* **67**: 6642–6647.
Neuberger, MS, Williams, GT, Mitchell, EB, Jouhal, SS, Flanagan, JG, Rabbitts, TH. 1985. A hapten-specific chimaeric IgE antibody with human physiological effector function. *Nature* **314**: 268–270.
Oostenbrink, C, Villa, A, Mark, AE, van Gunsteren, WF. 2004. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**: 1656–1676. DOI:10.1002/jcc.20090.
Poiron, C, Wu, Y, Ginestoux, C, Ehrenmann, F, Duroux, P, Lefranc, M-P. 2010. IMGT/mAb-DB: the IMGT® database for therapeutic monoclonal antibodies. *Journ. Ouvert Biol Inf. Mathématiques*. Poster 13, Available from: http://www.jobim2010.fr/indexe662.html?q=en/node/56
Presta, LG, Lahr, SJ, Shields, RL, Porter, JP, Gorman, CM, Fendly, BM, Jardieu, PM. 1993. Humanization of an antibody directed against IgE. *J. Immunol. Baltim. Md 1950* (151): 2623–2632.
Queen, C, Schneider, WP, Selick, HE, Payne, PW, Landolfi, NF, Duncan, JF, Avdalovic, NM, Levitt, M, Junghans, RP, Waldmann, TA. 1989. A humanized antibody that binds to the interleukin 2 receptor. *Proc. Natl. Acad. Sci. U. S. A.* **86**: 10029–10033.
Riechmann, L, Clark, M, Waldmann, H, Winter, G. 1988. Reshaping human antibodies for therapy. *Nature* **332**: 323–327. DOI:10.1038/332323a0.
Roguska, MA, Pedersen, JT, Keddy, CA, Henry, AH, Searle, SJ, Lambert, JM, Goldmacher, VS, Blättler, WA, Rees, AR, Guild, BC. 1994. Humanization of murine monoclonal antibodies through variable domain resurfacing. *Proc. Natl. Acad. Sci. U. S. A.* **91**: 969–973.
de Ruiter, A, Mader, A, Kunert, R, Oostenbrink, C. 2011. Molecular simulations to rationalize humanized Ab2/3H6 activity. *Aust. J. Chem.* **64**: 900–909.
Ryckaert, J-P, Ciccotti, G, Berendsen, HJC. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**: 327–341. DOI:10.1016/0021-9991(77)90098-5.
Schmid, N, Eichenberger, AP, Choutko, A, Riniker, S, Winger, M, Mark, AE, van Gunsteren, WF. 2011. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**: 843–856. DOI:10.1007/s00249-011-0700-9.
Schmid, N, Christ, CD, Christen, M, Eichenberger, AP, van Gunsteren, WF. 2012. Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation. *Comput. Phys. Commun.* **183**: 890–903. DOI:10.1016/j.cpc.2011.12.014.
Schroff, RW, Foon, KA, Beatty, SM, Oldham, RK, Morgan, AC. 1985. Human anti-murine immunoglobulin responses in patients receiving monoclonal antibody therapy. *Cancer Res.* **45**: 879–885.
Sgro, C. 1995. Side-effects of a monoclonal antibody, muromonab CD3/orthoclone OKT3: bibliographic review. *Toxicology* **105**: 23–29.
Shawler, DL, Bartholomew, RM, Smith, LM, Dillman, RO. 1985. Human immune response to multiple injections of murine monoclonal IgG. *J. Immunol. Baltim. Md 1950* **135**: 1530–1535.
Tan, P, Mitchell, DA, Buss, TN, Holmes, MA, Anasetti, C, Foote, J. 2002. "Superhumanized" antibodies: reduction of immunogenic potential by complementarity-determining region grafting with human germline sequences: application to an anti-CD28. *J. Immunol. Baltim. Md 1950* **169**: 1119–1125.
Taylor, LD, Carmack, CE, Schramm, SR, Mashayekh, R, Higgins, KM, Kuo, CC, Woodhouse, C, Kay, RM, Lonberg, N. 1992. A transgenic mouse that expresses a diversity of human sequence heavy and light chain immunoglobulins. *Nucleic Acids Res.* **20**: 6287–6295.
Tironi, IG, Sperb, R, Smith, PE, van Gunsteren, WF. 1995. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* **102**: 5451–5459. DOI:10.1063/1.469273.
Tobias, R, Kumaraswamy, S. 2014. Biomolecular binding kinetic assays on the octet platform. *Forte Bio. Appl. Note* **14**: 1–21.
Vogt, AD, Di Cera, E. 2013. Conformational selection is a dominant mechanism of ligand binding. *Biochemistry (Mosc.)* **52**: 5723–5729. DOI:10.1021/bi400929b.
Winter, G, Griffiths, AD, Hawkins, RE, Hoogenboom, HR. 1994. Making antibodies by phage display technology. *Annu. Rev. Immunol.* **12**: 433–455. DOI:10.1146/annurev.iy.12.040194.002245.

SUPPORTING INFORMATION

Additional supporting information can be found in the online version of this article at the publisher's website.

Supplementary material: Antibody humanization by molecular dynamics simulations – in-silico guided selection of critical backmutations

Christian Margreitter,[#a] Patrick Mayrhofer[#b], Renate Kunert[b] and Chris Oostenbrink*[a]

Content:

a C. Margreitter, Dr. C. Oostenbrink
Institute of Molecular Modeling and Simulation
University of Natural Resources and Life Sciences
Muthgasse 18, 1190 Vienna
E-mail: chris.oostenbrink@boku.ac.at

b P. Mayrhofer, Dr. R. Kunert
Department of Biotechnology
University of Natural Resources and Life Sciences
Muthgasse 18, 1190 Vienna
E-mail: renate.kunert@boku.ac.at

# Both authors contributed equally to the work.

* To whom correspondence should be addressed.

wt3H6	properties	su3H6	properties	possible function	position in crystal
$A68^{HC}$	small, aliphatic, hydrophobic	$V68^{HC}$	aliphatic, hydrophobic	backmutated in GC3H6, part of Vernier zone	surface, facing towards CDR-H2/beta-sheet (stabilizing barrel?)
$V72^{HC}$	aliphatic, hydrophobic	$A72^{HC}$	small, aliphatic, hydrophobic	backmutated in GC3H6 and GA3H6, defining canonical CDR structure class, part of Venier zone	surface, underneath CDR-H2
$R98^{HC}$	basic	$T98^{HC}$	polar, uncharged	backmutated in GC3H6 and GA3H6, defining canonical CDR structure class, part of Vernier zone	between CDR-H1 and CDR-H3 (stabilizing Ag binding site)
$V4^{LC}$	aliphatic, hydrophobic	$L4^{LC}$	aliphatic, hydrophobic	backmutated in GC3H6, part of Vernier zone	surface, facing towards CDR-L3/beta-sheet (stabilizing barrel?)
$R45^{LC}$	basic	$I45^{LC}$	aliphatic, hydrophobic	backmutated in GC3H6, buried in V_H/V_L interface, not part of Vernier zone	surface, interaction with CDR-L2?
$D85^{LC}$	acidic	$Y85^{LC}$	aromatic, hydrophobic	backmutated in GC3H6 and GA3H6, affects overall protein stability	interaction with Kappa constant region?

Table S1: Properties of amino acid side-chains selected for establishing wt3H6 double mutants used as a training panel (TR01-06) for MD simulations. The critical functions of these framework positions are described in Mader and Kunert, Protein Eng. Des. Sel. PEDS, 23, 947 – 954 (2010). Spatial positions in the crystal structure 3BQU are summarized.

Search parameters:	Species:	homo sapiens
	Gene type:	variable
	Functionality:	functional
	Locus	IGH
	Molecular component:	IG

Residue	Number of alleles	% of total
Arginine (R)	175	80%
Threonine (T)	10	5%
Lysine (K)	24	11%
Histidine (H)	7	3%
Alanine (A)	2	1%
Total	218	100%

Table S2: Conserved amino acid residues at the position equivalent to R98 in the human IGHV germline genes (IMGT/Gene-DB).

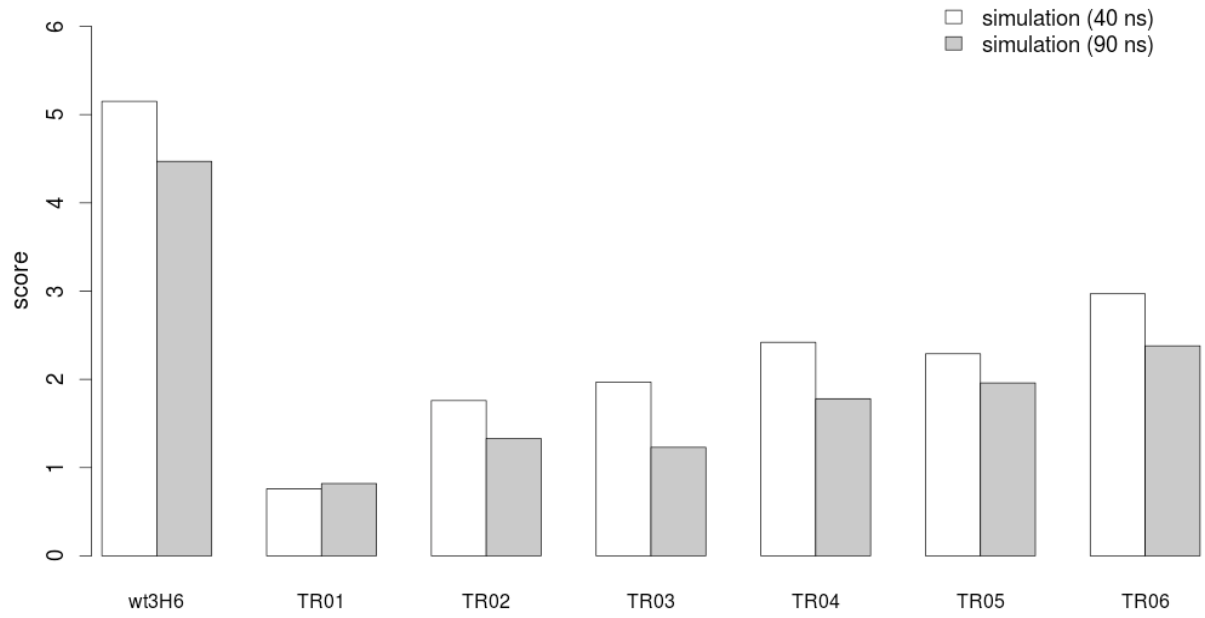


Figure S1: Use of 90 ns trajectories for the score calculation results in a slightly lower absolute score for the variants, while retaining the relative, significant differences. Therefore, it seems reasonable to favor a higher number of replicates per variant over a individual simulation time larger than 50 ns.

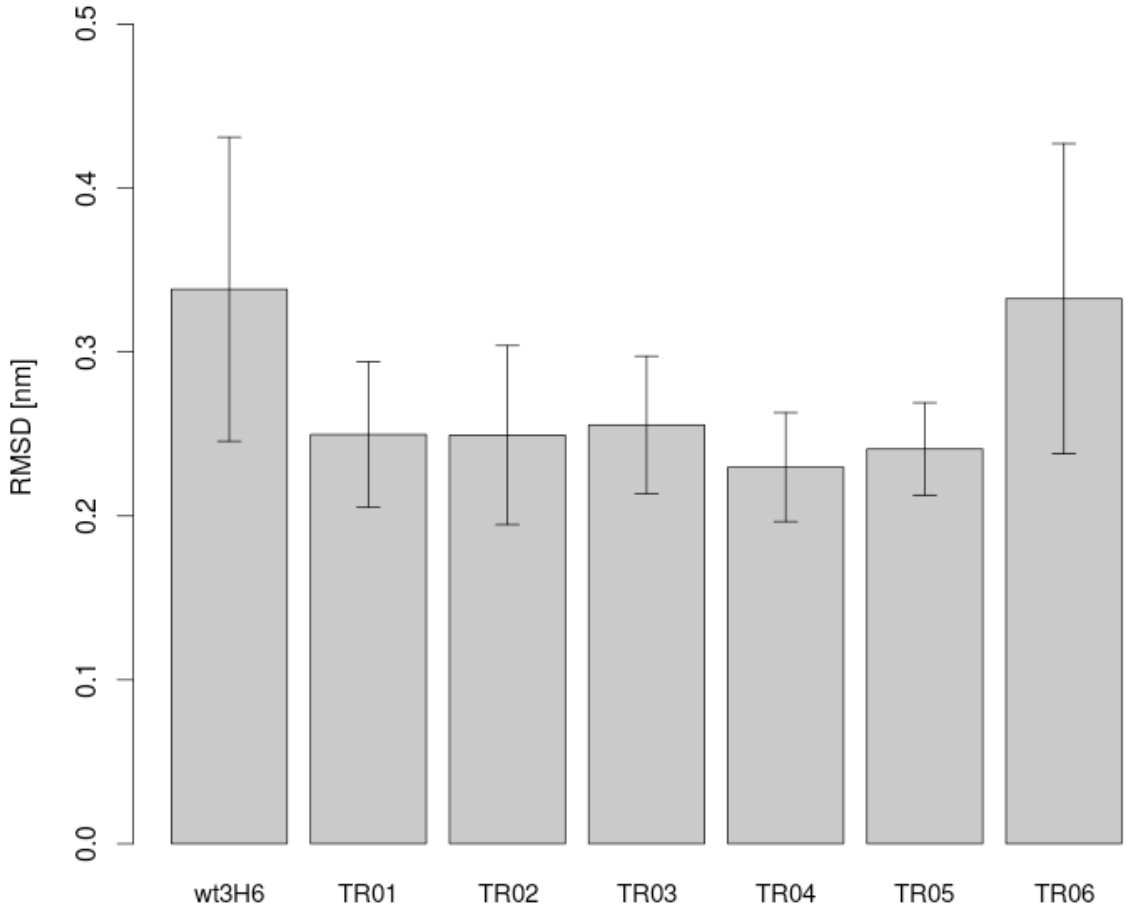


Figure S2: The average backbone atom root-mean-square deviation (RMSD) with respect to the crystal structure over the last 20 ns of all replicates for the training set. The RMSD is calculated for all backbone atoms of the framework regions after a least-squares fit on these atoms. The error bars indicate the standard deviation.

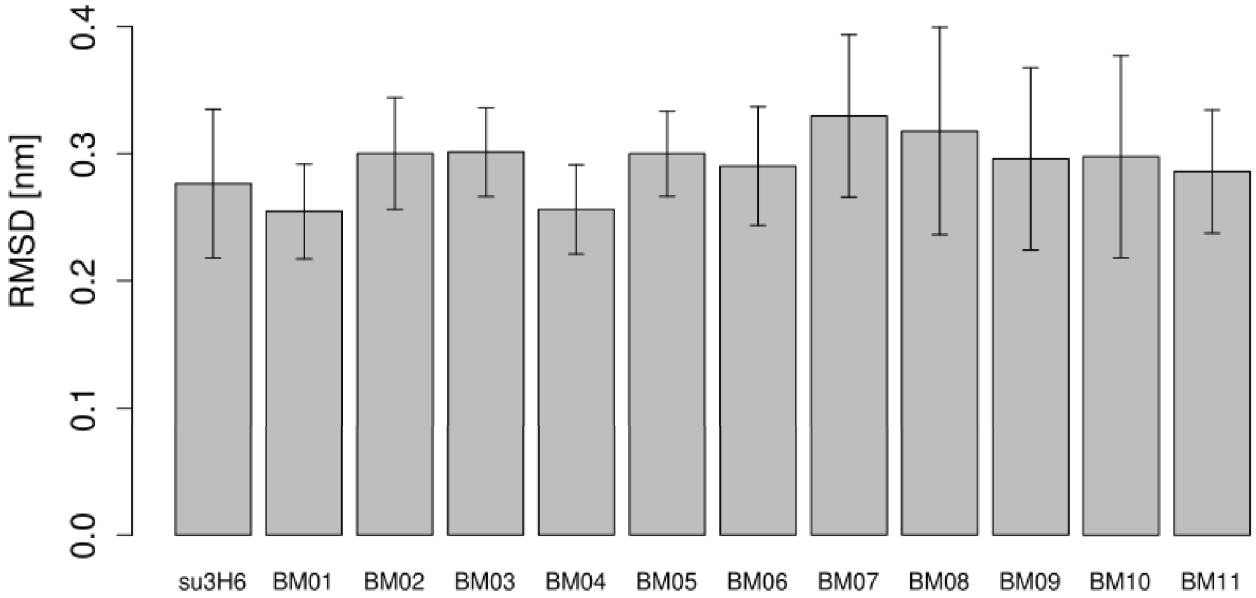


Figure S3: The average backbone atom root-mean-square deviation (RMSD) with respect to the crystal structure over the last 20 ns of all replicates for the superhumanized variants. The RMSD is calculated for all backbone atoms of the framework regions after a least-squares fit on these atoms. The error bars indicate the standard deviation.

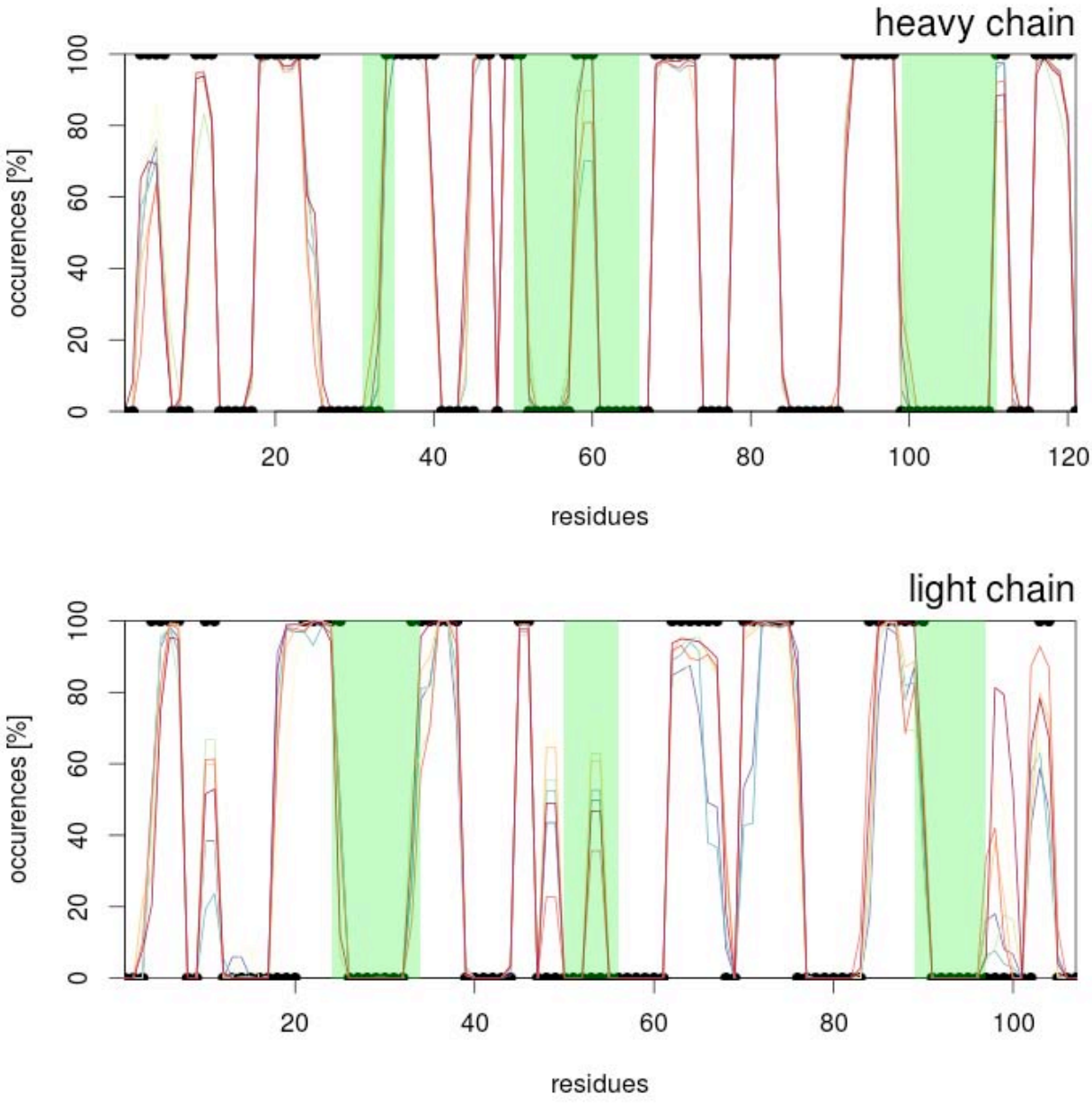


Figure S4: β-strand average occurences of conformations over the last 40 ns over all replicates for the training set as determined by the program DSSP for the framework and complementary-determining regions (CDRs; in green). The classification of the crystal structure is shown in black dots (either 100% or 0%), which agrees well with the mean values of the trajectories. This indicates a stable secondary fold throughout the simulation.

Appl Microbiol Biotechnol (2014) 98:9723–9733
DOI 10.1007/s00253-014-6011-1

Accurate comparison of antibody expression levels by reproducible transgene targeting in engineered recombination-competent CHO cells

Patrick Mayrhofer · Bernhard Kratzer · Wolfgang Sommeregger · Willibald Steinfellner · David Reinhart · Alexander Mader · Soeren Turan · Junhua Qiao · Juergen Bode · Renate Kunert

Received: 18 June 2014 / Revised: 30 July 2014 / Accepted: 1 August 2014 / Published online: 27 August 2014

Abstract Over the years, Chinese hamster ovary (CHO) cells have emerged as the major host for expressing biotherapeutic proteins. Traditional methods to generate high-producer cell lines rely on random integration(s) of the gene of interest but have thereby left the identification of bottlenecks as a challenging task. For comparison of different producer cell lines derived from various transfections, a system that provides control over transgene expression behavior is highly needed. This motivated us to develop a novel "DUKX-B11 F3/F" cell line to target different single-chain antibody fragments into the same chromosomal target site by recombinase-mediated cassette exchange (RMCE) using the flippase (FLP)/FLP recognition target (FRT) system. The RMCE-competent cell line contains a gfp reporter fused to a positive/negative selection system flanked by heterospecific FRT (F) variants under control of an external CMV promoter, constructed as "promoter trap". The expression stability and FLP accessibility of the tagged locus was demonstrated by successive rounds of RMCE. As a proof of concept, we performed RMCE using cassettes encoding two different anti-HIV single-chain Fc fragments, 3D6scFv-Fc and 2F5scFv-Fc. Both targeted integrations yielded homogenous cell populations with comparable intracellular product contents and messenger RNA (mRNA) levels but product related differences in specific productivities. These studies confirm the potential of the newly available "DUKX-B11 F3/F" cell line to guide different transgenes into identical transcriptional control regions by RMCE and thereby generate clones with comparable amounts of transgene mRNA. This new host is a prerequisite for cell biology studies of independent transfections and transgenes.

Keywords Site-directed integration · Cell engineering · Flippase · Specific productivity · RMCE target site

Electronic supplementary material The online version of this article (doi:10.1007/s00253-014-6011-1) contains supplementary material, which is available to authorized users.

P. Mayrhofer · W. Sommeregger · W. Steinfellner · D. Reinhart · A. Mader · R. Kunert (✉)
Department of Biotechnology/Vienna Institute of BioTechnology (BOKU-VIBT), University of Natural Resources and Life Sciences (Vienna), Muthgasse 18, A-1190 Vienna, Austria
e-mail: renate.kunert@boku.ac.at

B. Kratzer
Institute of Immunology, Medical University Vienna, Borschkegasse 8a, A-1090 Vienna, Austria

S. Turan
Department of Genetics, Stanford University School of Medicine, 300 Pasteur Drive, Stanford 94305, CA, USA

J. Qiao
WuXi AppTec Co. Ltd, 288 Fute Zhong Road, Shanghai 200131, China

J. Bode
Institute for Experimental Hematology, Hannover Medical School, Carl-Neuberg-Str.1, D-30625 Hannover, Germany

Introduction

Monoclonal antibodies (mAb) represent the main market fraction of all biotherapeutics with USD 24.6 billion in US sales and a growth of 18.2 % in 2012 (Aggarwal 2014). The main producers for all therapeutic proteins are Chinese hamster ovary (CHO) cells because of their human-like protein expression capabilities, authentic glycosylation patterns, ease of cultivation, and genetic modification (Wurm 2004). Strategies to improve (specific) production titers include genetic engineering of the expression vectors and host cell line combined with optimization of cultivation and feeding strategies (Kim et al. 2012). Cell biological investigations with producer cell

lines under different environmental conditions are especially governed by different -omics techniques to refine production capabilities for biopharmaceutical technology (Dietmair et al. 2012; Datta et al. 2013). Although different reviews have been published in detail, so far the success of complex -omics projects is limited (Kantardjieff et al. 2009; Doolan et al. 2013; Clarke et al. 2011). It can be speculated that the cellular model, methodology, or even the bioinformatic input are not yet appropriate to elucidate complex systemic components of cell biology. One challenge when comparing producer cell lines developed from different transfection reactions is to overcome variations in transcription efficacy caused by the "position effect" (Wilson et al. 1990). Together with residual vector-specific components, this effect is a frequent cause of different epigenetic silencing events, among these are histone deacetylation, distinct histone methylation or phosphorylation steps, and DNA-/promoter-methylation patterns (Mutskov and Felsenfeld 2004; Richards and Elgin 2002). These effects may be triggered by integration of the transgene into heterochromatin, leading to the loss or reduction of expression (Kwaks and Otte 2006). Furthermore, it was reported that the transcriptional strength of a promoter is highly dependent on the chromosomal position and the presence of promoter control elements in the surrounding chromatin environment (Nehlsen et al. 2009). Due to these obvious challenges, gene targeting strategies became more and more important for introducing the gene of interest into predetermined chromosomal loci by using site-specific recombinases, such as Cre/loxP (Fukushige and Sauer 1992; Kito et al. 2002) or flippase (FLP)/FLP recognition target (FRT) (O'Gorman et al. 1991; Huang et al. 2007; Turan et al. 2013) systems. The use of two heterospecific, noncompatible FLP recognition target sites in combination with a screening/selection marker enables the isolation of an engineered host cell line containing a chromosomal recombinant cassette in a predefined, transcriptionally active locus. Subsequently, the cassette downstream of the promoter can be replaced by another coding unit based on the same set of flanking heterospecific FRT sites and free of nonessential vector components ("recombinase-mediated cassette exchange" (RMCE)) exploiting the potential of yeast FLP recombinase (Qiao et al. 2009). Our contribution describes the generation of a new host cell line capable to integrate different transgenes into the same chromosomal locus driven by the same transcriptional control elements. Our new CHO host cell line DUKX-B11 F3/F was developed by random integration of a RMCE-competent cassette followed by target site selections to guarantee a homogenous cell population with an expedient cassette exchange and transcription potential. We verified the potential of the new host cell line by the expression of two model proteins, the antibody fragments 3D6scFv-Fc and 2F5scFv-Fc, at the established RMCE target sites. The clones were grown in T25 and spinner flasks for routine cultivation or batch cultivation experiments to analyze specific transcript- and product-formation capabilities. The established DUKX-B11 F3/F cell line provides a stable, retargetable chromosomal locus, indicated by homogenous intracellular green fluorescent protein (gfp) expression, capable to exchange different transgenes into the same transcription control region by site-specific recombination. After selection, scFv-Fc clones developed by RMCE showed reproducible transcription and secretion levels demonstrating the suitability of DUKX-B11 F3/F for comparison of different recombinant producer cell lines.

Materials and methods

Plasmids

The cytomegalovirus (CMV) promoter sequence-coding vector pCMV-F3-gfp-F was used for random introduction of the initial RMCE-competent cassette as described in Qiao et al. (2009). It contains the CMV promoter sequence driving the expression of the gfp in combination with a HSV thymidine kinase (tk) polyadenylation signal (pA) and two heterospecific FLP recognition target (*FRT*3 and *FRT*) sites downstream of the promoter and 3′ of the pA sequence. The first RMCE-targeting vector pF3-hygromycin phosphotransferase (hyg)/tk-F harbors a fusion protein of hyg/tk and was constructed by PCR amplification of the fusion protein and cloning into *Age*I and *Nhe*I opened pF3-gfp/tk/neomycin phosphotransferase (neo)-F. The second RMCE-targeting vector pF3-gfp/tk/neo-F contains the fusion protein of gfp, tk, and neo ("gtn") as described in Qiao et al. (2009). The FLP recombinase enzyme was provided by co-transfection of the plasmid pFLPo containing the FLP coding region under control of a phosphoglycerate kinase promoter (Raymond and Soriano 2007) (Addgene, plasmid 13793). Plasmids pF3-3D6scFv-Fc-F and pF3-2F5scFv-Fc-F contained the 3D6scFv (GenBank, CAA001551) or the 2F5scFv (sequence obtained from PDB, 2F5A), respectively, combined by a $(GGGGS)_3$ linker and fused to the human IgG1 Fc region (GenBank, CAA49866). 3D6scFv-Fc or 2F5scFv-Fc open reading frames were amplified by PCR and ligated into the *Kpn*I- and *Nhe*I- opened pF3-gfp/tk/neo-F.

Construction of RMCE host DUKX-B11 F3/F

CHO DUKX-B11 cells (ATCC CRL-9096) (Urlaub and Chasin 1980) were routinely cultivated in suspension under protein- and antibiotic-free conditions in a 125-mL spinner flask (Techne) pregased to 5 % CO_2 at 37 °C and 50 rpm. DMEM/HAM's F12 (1:1) medium (Biochrom GmbH) was supplemented with 4 mM L-glutamine (PAA), 0.1 % pluronic-F68 (Sigma-Aldrich), 100 μM hypoxanthine/16 μM thymidine (HT; Invitrogen), 0.25 % soya-peptone/UF (Quest

International GmbH), and a protein-free supplement (Polymun scientific). Plasmid pCMV-F3-gfp-F was introduced by nucleofection (Amaxa cell line Nucleofector kit V, Lonza) of 2×10^7 cells using 20 μg of plasmid DNA in ProCHO5 medium (Lonza) supplemented with 4 mM L-glutamine (PAA) and HT (Invitrogen). Twenty-four hours after the first transfection, 1×10^6 cells were co-transfected with 6 μg pF3-hyg/tk-F and 2 μg pFLPo using 80 μg polyethylenimine (PEI; 25 kDa, linear; Polysciences Inc.) to induce the first RMCE reaction. Stable cell pools were selected by limited dilution in a 96-well plate (Nunc) in ProCHO5 supplemented with 4 mM L-glutamine, HT, and 0.2 mg/mL hygromycin B (Invitrogen). GFP-negative and hygromycin-resistant pools were expanded into T25 flasks and used for a second RMCE reaction by co-transfection of 1×10^6 cells with 6 μg pF3-gfp/tk/neo-F and 2 μg pFLPo using 80 μg PEI. G418-resistant cell pools were selected by limited dilution in ProCHO5 supplemented with 4 mM L-glutamine, HT, and 0.5 mg/mL G418 (PAA) followed by expansion into T25 flasks, a second subcloning step and adaptation into 125-mL spinner flasks (Techne).

Generation and characterization of scFv-Fc producing RMCE subclones CMV-F3-3D6scFv-Fc-F and CMV-F3-2F5scFv-Fc-F

One million cells of RMCE host DUKX-B11 F3/F were co-transfected with 6 μg pF3-3D6scFv-Fc-F or pF3-2F5scFv-Fc-F and 2 μg pFLPo by 80 μg PEI. After 24 h, cells were resuspended in ProCHO5 supplemented with 4 mM L-glutamine, HT, and 1 μM ganciclovir followed by limited dilution in 96-well plates. ScFv-Fc-producing and growing cell pools were selected by enzyme-linked immunosorbent assay (ELISA) and expanded into T25 flasks for analysis of specific growth rates and specific productivities. Two clones of each scFv-Fc variant were chosen for propagation in spinner flasks for routine cultivation by passaging 2×10^5 living cells/mL in a total working volume of 50 mL every 3 to 4 days to keep cells in exponential growth phase. Additional batch experiments in spinner flasks were performed by seeding 2×10^5 living cells/mL in a total working volume of 75 mL in ProCHO5 supplemented with 4 mM L-glutamine and HT without the use of any positive or negative selection. The batch was harvested when the viability dropped below 30 %. Cell number was determined by nuclei preparation with 0.1 M citric acid and 2 % (*w*/*w*) Triton X 100 and particle counting with Coulter counter Z2 (Beckman coulter). Cell viability was determined by 0.5 % trypan blue dye exclusion using a hemocytometer (Neubauer). Product titer in culture supernatant was quantified by sandwich gamma-gamma ELISA. Ninety-six-well Maxisorp plates (Nunc) were precoated with 0.5 μg/mL polyclonal goat anti-human IgG (γ-chain specific) antiserum (Sigma). Standards and samples were quantified in a 2-fold dilution series with goat anti-human IgG-horseradish peroxidase (HRP; γ-chain specific) conjugate (Invitrogen) and stained with orthophenylediamine and H_2O_2 (Merck). The resulting color reaction was measured at 492 nm at a reference wavelength of 620 nm on a microplate reader (Tecan). The specific productivity qP was expressed as picograms per cell per day as described in Lattenmayer et al. (2007).

Measurement of gfp fluorescence and intracellular product accumulation by flow cytometry

Two million cells of DUKX-B11 F3/F were washed and resuspended in PBS for determination of gfp fluorescence. For intracellular scFv-Fc product analysis, 1×10^6 cells were fixed with ice-cold 70 % (*v*/*v*) ethanol and stained with anti-human IgG-R-phycoerythrin (PE) (γ-chain specific) conjugate antibody (Sigma). All flow cytometric analyses were measured on a Gallios flow cytometer (Beckman Coulter). Single cells were gated according to their forward and side scatter properties and analyzed for gfp fluorescence by FL-1 laser channel or for intracellular PE fluorescence by the FL-2 laser channel using Kaluza Analysis Software (Beckman Coulter).

cDNA preparation and determination of mRNA transcript level by qPCR

Two million cells were used for total RNA isolation using Ambion TRI Reagent Solution (Life Technologies) according to the manufacturer's protocol. Residual DNA was removed by DNase (Qiagen) treatment in the presence of RNase inhibitor (Life Technologies). Purified RNA was stored at −20 °C in RNase-free water. Complementary DNA (cDNA) was prepared by reverse transcription with M-MLV reverse transcriptase (Promega) using random primers (Promega) according to the manufacturer's instructions. Concentration and purity was determined with a ND-1000 spectrophotometer (Nano-Drop). Primers for quantitative PCR (qPCR) analysis were designed using the Primer3 Web application (Untergasser et al. 2007) and synthesized by Sigma-Aldrich. A 100-bp amplicon from the Fc part and a 78-bp amplicon of the housekeeping gene β-actin served as an internal control for normalization of the data.

Three-nanogram sample cDNA was denatured for 10 min to reduce broad dispersion of signals as described in Sommeregger et al. (2013) before PCR reaction was started. qPCR analysis including nontemplate and negative controls was performed on the MiniOpticon system (Bio-Rad) using the TaqMan method by denaturation at 95 °C for 5 min followed by 40 cycles at 95 °C for 15 s and 55 °C for 1 min. Fluorescence signal and Cp values were determined by the CFX manager software 2.1 (Bio-Rad) using baseline subtraction and linear regression. The $2^{-\Delta\Delta Cp}$ method (Livak and

Schmittgen 2001) was used for relative quantification of messenger RNA (mRNA) expression levels. For statistical significance, three independent biological samples per cell line were analyzed in two technical runs.

Results

Establishment and characterization of the RMCE host cell line DUKX-B11 F3/F

Suspension-adapted CHO DUKX-B11 cells were transfected with a plasmid containing the CMV promoter 5′ upstream of a cassette containing the gfp reporter and a polyadenylation signal, flanked by two heterospecific FLP recognition target sites (*FRT*3 and *FRT*^{wt}, abbreviated "F3" or "F", respectively; step 1 in Fig. 1). The CHO DUKX-B11 cell line was chosen as an RMCE host because of its broad use in industrial biotechnology. For introduction of a first RMCE-competent "donor cassette", the target unit contained the CMV promoter 5′ upstream of F3 to create a promoter trap in the following transfections. By using an optimized transfection protocol, the genetic expression elements of pCMV-F3-gfp-F were tested in a transient system without positive selection pressure.

The use of a gfp reporter in the first RMCE-targeting construct enables the verification of transfection efficiency. The first RMCE reaction was already initiated 24 h later by co-transfection of the plasmid pF3-hyg/tk-F and the FLP

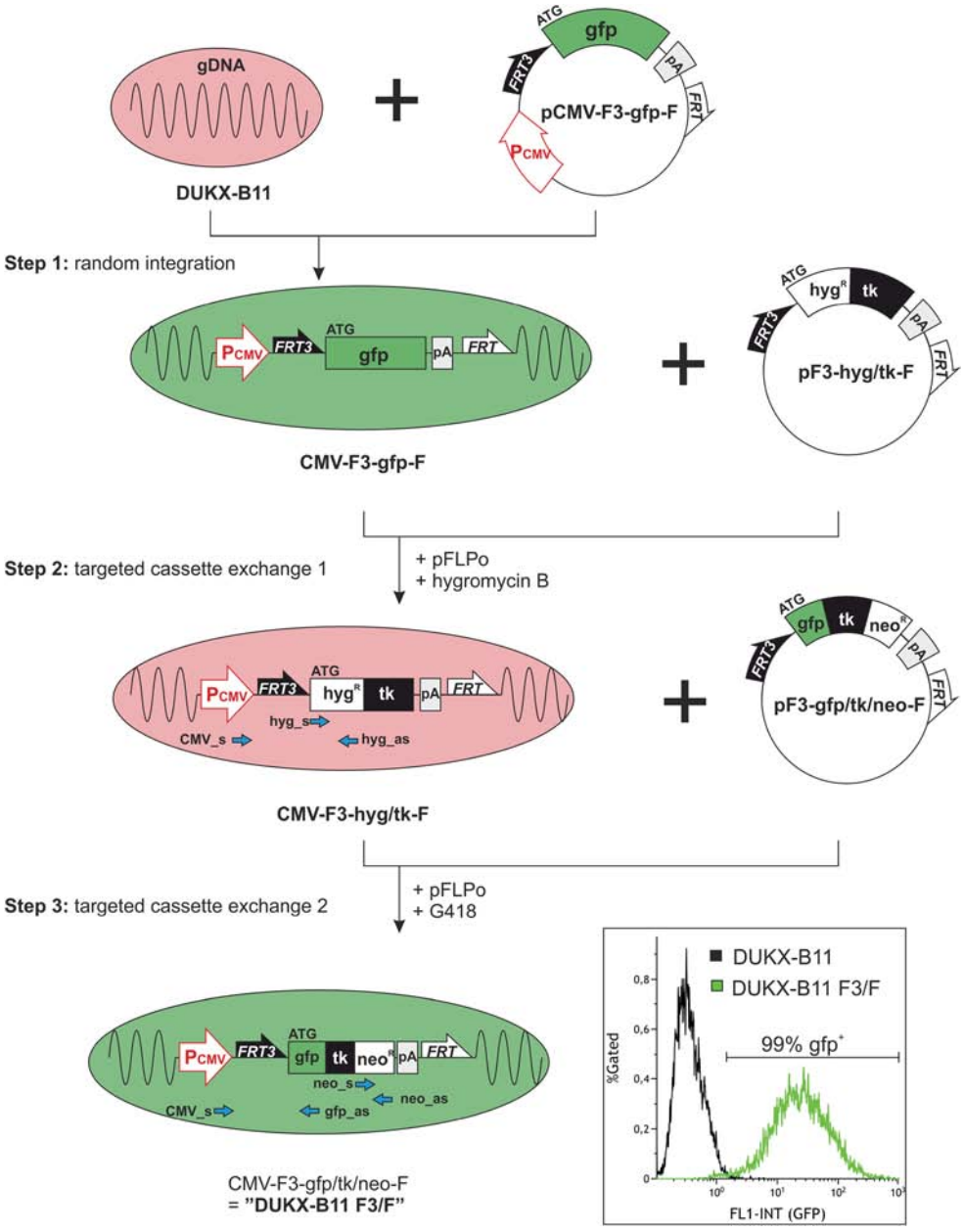


Fig. 1 Establishment and homogenous intracellular gfp fluorescence of the RMCE-competent host cell line DUKX-B11 F3/F. DUKX-B11 F3/F was established by random integration of a RMCE cassette on plasmid pCMV-F3-gfp-F (step 1) followed by two consecutive rounds of RMCE (steps 2 and 3) mediated by co-transfection of the targeting vector (pF3-hyg/tk-F or pF3-gfp/tk/neo-F) and the FLP expression plasmid (*pFLPo*). Stable cell clones were selected by hygromycin B (step 2) or G418 (step 3). Primers used for genomic PCR characterization are indicated as *blue arrows*, and results are listed in Table 1 and Online resource 1. The final DUKX-B11 F3/F cell line was analyzed for gfp fluorescence by flow cytometry shown as a single-parameter (*FL1-INT*) histogram. Nonfluorescent DUKX-B11 cells were used as negative control. Plasmids are indicated as *circles*, DUKX-B11 cells are indicated as *pink* or *green ovals*, indicating gfp-negative or gfp-positive fluorescence, respectively, and the cassettes integrated into the host genomic DNA (*gDNA*) are shown. Genetic elements: *pCMV* CMV promoter, *FRT3*/*FRT* spacer-mutant/wild-type FLP recognition target sites, *gfp* green fluorescent protein, *pA* poly A signal, *ATG* start ATG codon, *hygR* hygromycin phosphotransferase, *tk* thymidine kinase, *neoR* neomycin phosphotransferase

expression plasmid pFLPo to restrict the overgrowth of transfection pools by nonproducers. RMCE donor plasmid pF3-hyg/tk-F comprises a promoterless fusion protein for positive or negative selection by either hygromycin B or ganciclovir, respectively, flanked by the same set of heterospecific *FRT* sites (*FRT3* and *FRT*) (step 2 in Fig. 1). Twenty-four hours after initiation of the first RMCE reaction, the transfection pool was subcloned by limited dilution into 96-well plates in a selection medium containing 200 μg/mL hygromycin B. After 3 weeks, cell growth could be detected in 86 % of the seeded wells by acidification of the phenol red pH indicator. Two out of 576 (0.3 %) totally seeded wells showed a gfp-positive and hygromycin B-resistant cell population that could have been derived from clones containing multiple integrated CMV-F3-gfp-F parental tagging cassettes followed by only partial exchange of some of these targeting cassettes with the F3-hyg/tk-F donor cassettes after the first RMCE reaction. Twelve hyg-resistant clones with gfp-negative phenotype were selected for expansion into T25 flasks. Subclone CMV-F3-hyg/tk-F was selected based on hygromycin B-resistant and gfp fluorescence-negative phenotypes in combination with the highest specific growth rate (data not shown). For evaluation, genomic DNA was extracted and analyzed by PCR (Table 1; Online resource 1) using appropriate primers (blue arrows in Fig. 1). PCR of genomic DNA with primers specific for the CMV promoter and the hyg sequence (CMV_s and hyg_as) generated a specific amplicon, indicating an authentic RMCE reaction with the gfp sequence replaced by the hyg/tk fusion protein. As expected, no amplicon was detected after the first RMCE reaction by PCR using the primer pair specific for the CMV promoter and gfp sequence (CMV_s and gfp_as) confirming the full exchange of parental F3-gfp-F cassettes by RMCE.

Subclone CMV-F3-hyg/tk-F was subjected to a second round of RMCE to replace the hyg/tk fusion protein with a triple fusion protein consisting of gtn (step 3 in Fig. 1). Again, this exchange cassette did not contain any promoter sequence to make use of the promoter trap to prevent transcription of randomly introduced RMCE donor cassettes. At 24 h posttransfection, stable cell clones were selected by limited dilution in G418-containing media to select subclone CMV-F3-gfp/tk/neo-F based on growth characteristics and gfp fluorescence (data not shown). This final RMCE-competent host cell line, designated "DUKX-B11 F3/F", was analyzed for an authentic RMCE cassette exchange by PCR analysis of isolated genomic DNA (Table 1; Online resource 1). A specific PCR amplicon could be detected with primers specifically binding to the CMV promoter and gfp reporter sequence (CMV_s and gfp_as). No amplicons were detected with primers specific for hyg (hyg_s and hyg_as). The neo insert was confirmed by using primers binding only to the neo sequence (neo_s and neo_as) as well as primers binding specifically to the CMV promoter sequence and the neo sequence (CMV_s and neo_as). FISH analysis with gtn-specific probes evidenced a single integration site of the RMCE-competent targeting cassette within the DUKX-B11 F3/F host (Online resource 2). Flow cytometry analysis for gfp fluorescence intensities showed that 99 % of DUKX-B11 F3/F cells showed higher gfp fluorescence intensities than the gfp-negative DUKX-B11 host cells (single-parameter FL-1 INT histogram in Fig. 1) with geometric mean, median, and mode values of 22.4, 24.6, and 27.1, respectively, demonstrating the homogenous gfp expression capability of DUKX-B11 F3/F.

Table 1 PCR characterization on genomic DNA after first RMCE (clone CMV-F3-hyg/tk-F) and second RMCE (clone DUKX-B11 F3/F) reactions using primers depicted as blue arrows in Fig. 1

Subclone	Primer pair				
	CMV_s gfp_as	hyg_s hyg_as	CMV_s hyg_as	neo_s neo_as	CMV_s neo_as
CMV-F3-hyg/tk-F	−	+	+	−	−
DUKX-B11 F3/F	+	−	−	+	+

Specific PCR amplicons are depicted in Online resource 1

3D6scFv-Fc and 2F5scFv-Fc product comparison by RMCE

To test its performance as a genomic target site, two anti-HIV single-chain antibodies 2F5scFv-Fc and 3D6scFv-Fc were cloned into a promoterless RMCE vector flanked by the two heterospecific *FRT* sites (pF3-3D6scFv-Fc-F and pF3-2F5scFv-Fc-F) followed by integration into the DUKX-B11 F3/F genome by RMCE equivalent to step 2 or 3 in Fig. 1.

At 24 h posttransfection, stable antibody-producing subclones were selected by limited dilution and negative selection for absence of tk using the deoxyguanosine analog ganciclovir. Twelve clones for each antibody variant were expanded to T25 flasks, and their growth behavior and productivities were measured for ten consecutive passages.

Similar specific growth rates of all selected subclones were confirmed during the T25 cultivation period, suggesting that the different amino acid sequences of the two scFv-Fc variants have no major influence on the cellular metabolism of the established recombinant cell lines (Fig. 2a). The median specific productivity of 12 3D6scFv-Fc-producing subclones was 2.4-fold higher than that of 12 2F5scFv-Fc-producing subclones (Fig. 2b).

Already in course of clone development for 12 antibody-producing subclones of each variant, a highly reproducible and similar cell behavior in terms of specific productivities (qP) and growth rates (μ) could be observed (Fig. 2a, b). By

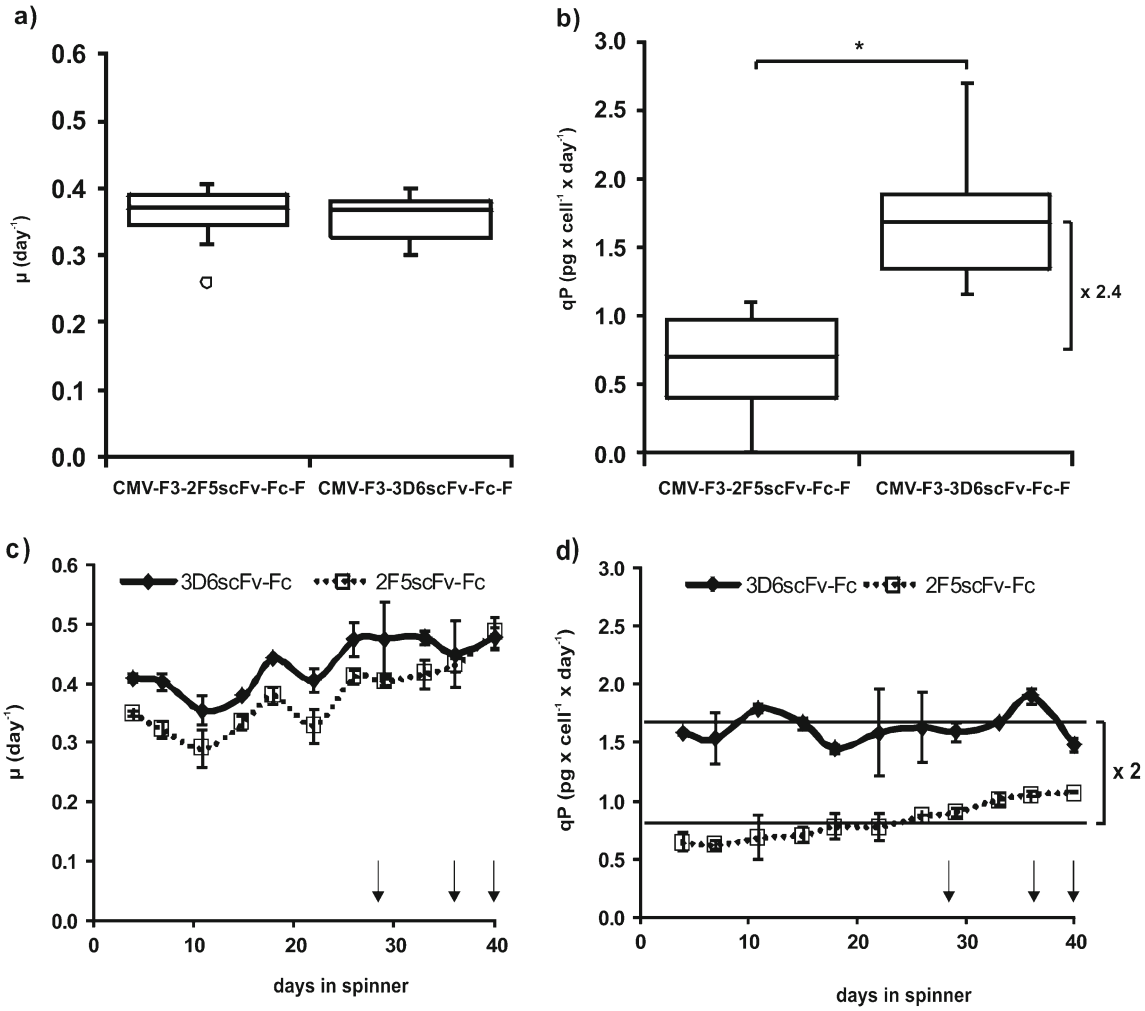


Fig. 2 Analysis of specific growth rates and productivities of scFv-Fc producing RMCE cell clones for ten consecutive passages in T25 and routine cultivation in spinner flasks. **a** Specific growth rates μ and **b** specific productivities qP are shown as box plot diagrams of 12 scFv-Fc producing RMCE clones of each antibody variant cultivated for ten consecutive passages in T25 flasks. *Boxes* represent median, first, and third quartiles of 12 clones. Outliers were defined as values ±1.5×interquartile range (IQR) and are represented as open circles. *Error bars* represent sample maxima and minima within 1.5×IQR. Factor 2.4 represents difference in median-specific productivities between 3D6scFv-Fc- and 2F5scFv-Fc-producing clones. *$p<0.001$—independent two-sample Student's *t* test. **c** Specific growth rates μ and **d** specific productivities qP of two scFv-Fc-producing clones of each antibody variant are shown as mean values in spinner cultivation for 40 days. *Error bars* represent ±SEM. Days of sampling for flow cytometry and qPCR analyses are shown as *arrows*. Factor 2 represents the mean difference in qP of 3D6scFv-Fc- and 2F5scFv-Fc-producing cell clones

contrast, the development of scFv-Fc-producing clones with conventional random integration of plasmids resulted in much higher variability in specific productivities and growth rates between individual subclones producing the same antibody variant, indicated by a higher interquartile range (midspread) in Online resource 3. Therefore, it is assumed that even with a small number of selected RMCE subclones, different cell lines with highly similar transcription efficiencies, indicated by the specific RNA/DNA ratio, can be obtained. By using conventional random transgene integration much higher screening effort is necessary to find two subclones with similar expression behavior. This is because random integration of transgenes into different chromosomal loci results in different transcription efficiencies and additional positional effects contribute to the final mRNA levels together with different RNA polymerase/DNA interactions.

Although RMCE exchange gave highly similar productivities and growth rates (Fig. 2a, b), little variance can be observed due to quantitation variance or local heterogeneity in culture conditions such as temperature or nutrient gradients during T25 flask cultivation. Therefore, the two best performing 3D6scFv-Fc (1F11 and 3B9) and 2F5scFv-Fc (1C3 and 3E5) clones were selected based on qp and μ for consecutive passages in spinner vessels for 40 days without any selection pressure. During this period of routine cultivation, similar specific growth rates were observed for all scFv-

Fc-producing subclones (Fig. 2c). A factor two difference was noted for the specific productivities of 3D6scFv-Fc-producing subclones compared with 2F5scFv-Fc-producing subclones (Fig. 2d). In the batch experiments without selection pressure, the maximum cell density reached 2×10^6 cells/mL after 7 or 9 days for subclone CMV-F3-3D6scFv-Fc-F 1F11 or CMV-F3-2F5scFv-Fc-F 1C3, respectively (Fig. 3a). Whereas maximum cell viability could be maintained for 3 days, it dropped to 30 % after 9 days in batch culture for both cell clones. A final product titer of 13 μg/mL was obtained for subclone CMV-F3-3D6scFv-Fc-F 1F11 exhibiting a constantly decreasing specific productivity starting from 3.3 pg $cell^{-1}$ day^{-1}. For subclone CMV-F3-2F5scFv-Fc-F 1C3, a final titer of only 3.4 μg/mL could be reached with a specific productivity of 0.8 pg $cell^{-1}$ day^{-1} that remained constant for the first 6 days during batch cultivation.

Intracellular product formation and mRNA levels of scFv-Fc-producing subclones

The four scFv-Fc-producing cell clones 1F11, 3B9, 1C3, and 3E5 chosen for routine cultivation in spinner flasks were analyzed for intracellular product formation and mRNA levels at three independent sampling days (indicated by arrows in Fig. 2c, d) to identify possible bottlenecks responsible for the differences in specific productivities of the two antibody variants. A single-parameter histogram at day 40 in spinner cultivation is depicted in Fig. 4a. A single peak of intracellular PE fluorescence (FL-2 signal intensity) was obtained from intracellular scFv-Fc for all individual sampling days (Online resource 4), as was expected for a RMCE guided exchange of transgenes. Comparing the median fluorescence intensity of two 2F5scFv-Fc and 3D6scFv-Fc-producing cell clones only

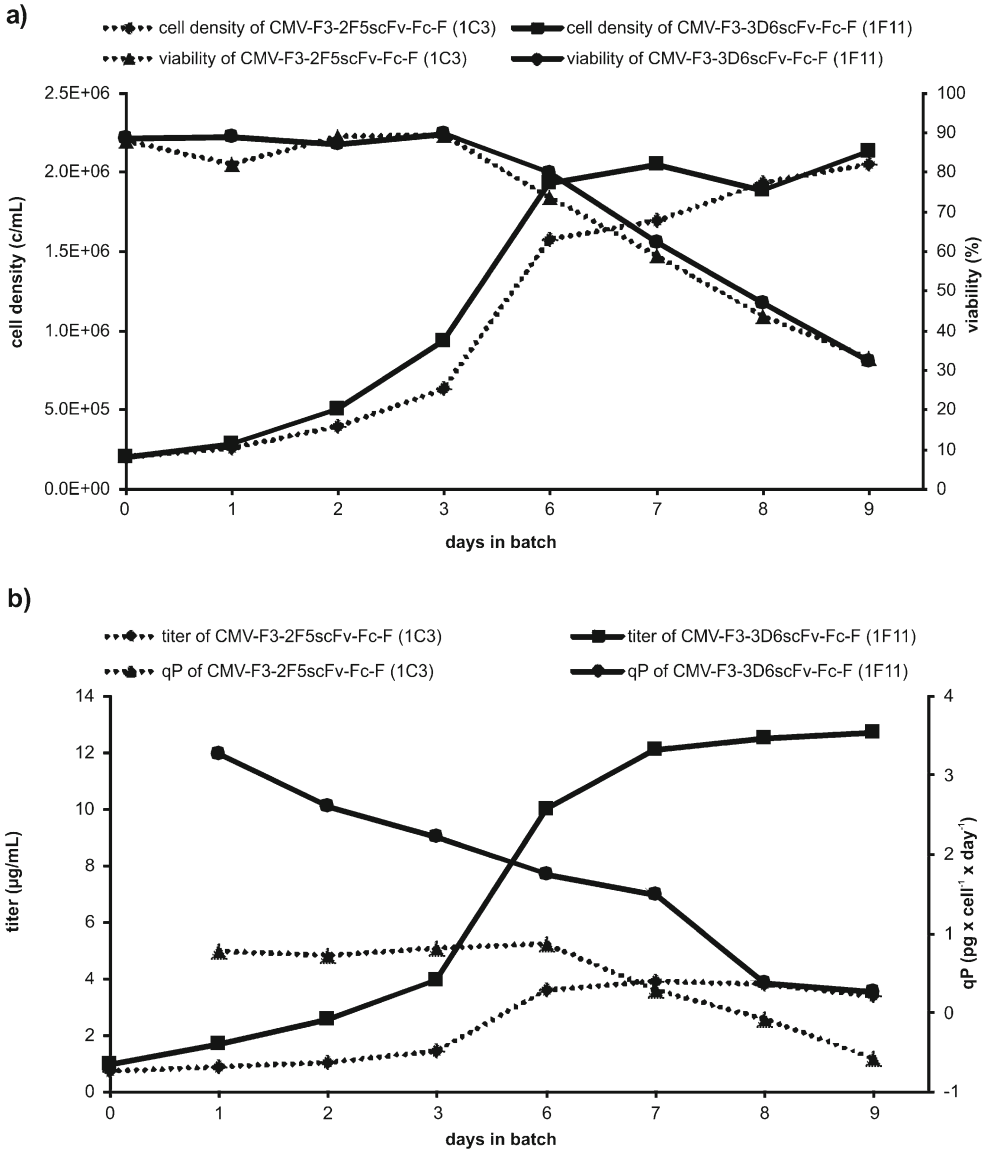


Fig. 3 Batch experiments of subclone CMV-F3-3D6scFv-Fc-F 1F11 and CMV-F3-2F5scFv-Fc-F 1C3 in spinner flasks. 1F11 and 1C3 were selected for spinner batch experiments by seeding 2×10^5 cells/mL in ProCHO5 supplemented with 4 mM L-glutamine and hypoxanthine/thymidine without using any selection pressure. During the 9-day batch experiment, cell density and viability (**a**) as well as specific productivity qp and final product titer (**b**) were measured

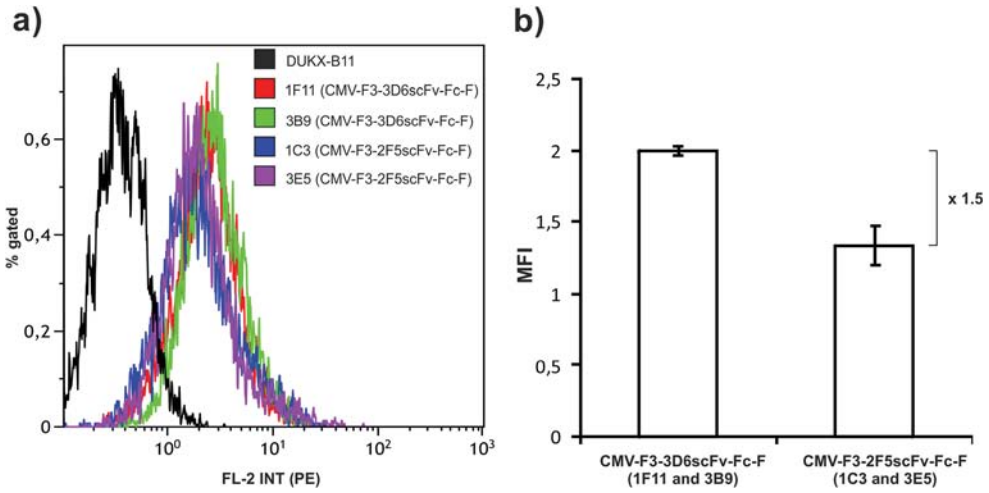


Fig. 4 Flow cytometric analysis of intracellular product formation of four selected scFv-Fc clones cultivated in spinner flasks. Samples were taken at three different days from four scFv-Fc-producing cell clones cultivated in spinner flasks. Cells were fixed by ethanol treatment and labeled with anti-huIgG-γ-chain-R-phycoerythrin (PE) antibody. **a** Single-parameter FL-2 histogram of samples taken at day 40 in spinner cultivation. DUKX-B11 host cell line was used as negative control. **b** Mean values of intracellular product formation of two 3D6scFv-Fc- or 2F5scFv-Fc-producing cell clones for each antibody variant are compared by median fluorescence intensities (MFI) showing a difference of 1.5 in intracellular product content. *Error bars* represent ±SEM

minor variations were measured between the different sampling times indicated by small error bars. The median fluorescence intensity (MFI) of 3D6scFv-Fc-producing clones was a factor 1.5 higher than the 2F5scFv-Fc-producing clones (Fig. 4b). This difference does not necessarily reflect the 2-fold difference in specific productivities. Relative mRNA levels compared with beta-actin reference level were constant in all scFv-Fc-producing RMCE clones cultivated in spinner flasks (Fig. 5; Online resource 5). These data support the capacity of the new host cell line as a tool for the comparison of clones at a post transcriptional level.

Discussion

Genomic targets enabling FLP RMCE can either be established by random integration (electroporation and transfection; Turan et al. 2013) or using the scout functions and integration preferences of retroviral vectors (Kuehle et al. 2014). Once the single-copy status has been documented the respective candidate sites lend themselves to the FLP-mediated integration of compatible expression cassettes. These conversions maintain the single-copy status of the target, they do not introduce unwanted vector sequences nor do they cause inadvertent rearrangements due to the underlying enzymatic principles.

In biotechnology, common exploitations of this strategy concern the definition of unique highly expressed loci for which a comprehensive toolbox has been introduced by Qiao et al. (2009). Based on the available target and donor vector backbones, our present study follows somewhat different goals, as it addresses the expression potential of various transgene architectures. This approach is valid if candidate constructs can be compared at predefined sites with moderate or even low intrinsic expression potential. Access to these requirements can be provided in case the general

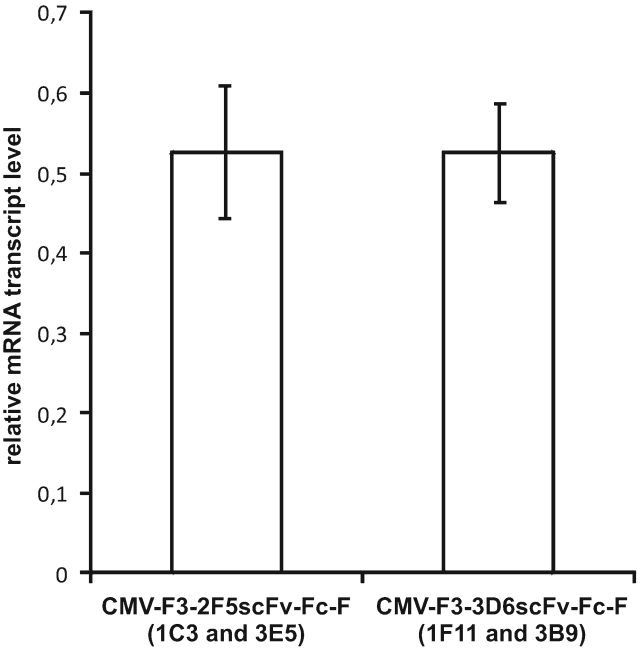


Fig. 5 qPCR analysis of mRNA transcript level of two selected 3D6scFv-Fc- or 2F5scFv-Fc-producing RMCE clones cultivated in spinner flasks. Samples were taken at three different days and measured in two technical replicates. Total mRNA was reverse transcribed into cDNA and analyzed by qPCR using probes specific for the Fc sequence or β-actin used as an internal standard. Mean $2^{-\Delta Cp}$ values were calculated based on differences of Cp values between β-actin and the Fc sequence. *Error bars* represent standard deviation

selection scheme (illustrated in Fig. 1) is streamlined in that only the second round of selections (step 3) is stringently controlled whereas step 2 obeys relaxed conditions that can be verified with just moderate investments of time.

The aim of this study was to establish the DUKX-B11 F3/F host cell line capable of introducing different antibody variants into the same genomic locus by RMCE. The primary goal was to develop this target site for a controlled and predictable integration of different genes of interest into the same chromosomal position irrespective of the achieved final specific productivity. Therefore, in this study no effort was made to identify high-producer clones by stringent and time-consuming screening and selection procedures or to characterize the targeted genomic locus in more detail by southern blot analysis or DNA sequencing methods. The rationale behind the development of the RMCE host cell line DUKX-B11 F3/F is to provide a tool that enables investigation of the expression levels of different proteins, in particular antibody variants, in an industrially widely used cell line under reproducible and identical transcriptional status of the integrated gene cassettes. Therefore, the primary focus was given to select RMCE compatible target sites with long-term stability and accessibility to the transcription machinery for gene expression as well as accessibility to the FLP recombinase enzyme to ensure efficient RMCE reactions.

The initial RMCE-competent target cassette, pCMV-F3-gfp-F, is controlled by a CMV promoter upstream of the *FRT*3 site resulting in a promoter trap as described in Qiao et al. (2009) (Fig. 1, step 1). The promoter trap overcomes expression of randomly introduced transgenes not targeted specifically to the chromosomal RMCE target site by FLP mediated recombination. The procedure for establishing DUKX-B11 F3/F was based on the expression of the CMV-F3-gfp-F cassette without selection for stable gfp-producing cell clones in a way that potent RMCE targets were already available 24 h after the first transfection with pCMV-F3-gfp-F. The first RMCE recombination events between the target cassette CMV-F3-gfp-F and the donor cassette F3-hyg/tk-F might have occurred either at the chromosomal level, where the gfp cassette was integrated randomly into the genome and afterwards the replacement occurred by RMCE. Alternatively and also possible, the cassette exchange reactions might have occurred already at the transient episomal stage where RMCE was initiated between the two cassettes still present as episomal plasmids. However, either way, the gfp cassette needs to be replaced by the hyg/tk fusion cassette by RMCE and integrated into the host genome leading to hyg/tk expression driven by the CMV promoter introduced by the first parental plasmid. The stability and accessibility of the donor cassette CMV-F3-hyg/tk-F could be demonstrated by cultivation in hygromycin B, followed by induction of a second RMCE reaction to replace the F3-hyg/tk-F cassette by another RMCE donor cassette, F3-gfp/tk/neo-F (step 3 in Fig. 1) followed by G418 selection. This second RMCE reaction and positive selection procedure is necessary to establish cell clones with high RMCE potential, at loci with appropriate transcription (mRNA) levels. At this point a negative selection procedure would be possible by using ganciclovir to select clones that have lost the tk fusion protein by site-specific cassette exchange.

The individual RMCE reactions were tracked with specific PCR reactions (Table 1; Online resource 1), indicating the success of both cassette exchange events. Absence of specific amplicons for hyg and presence of amplicons specific for the neo sequence under control of the CMV promoter in the final RMCE host DUKX-B11 F3/F indicated a successful integration of the triple fusion protein gfp/tk/neo into the target site by RMCE. The final RMCE host cell line DUKX-B11 F3/F resists G418 and exhibits a gfp-positive phenotype as indicated by flow cytometry (single-parameter FL-1 INT histogram in Fig. 1). The homogenous gfp signal profile, indicated by similar median, mean and mode values, implies that gfp expression in all cells of the population is driven by the same transcription control elements as was expected for a targeted RMCE integration of the gfp cassette. FISH results indicate a single integration locus of the gfp/tk/neo RMCE-targeting cassette in DUKX-B11 F3/F (Online resource 2).

For proof of concept, 3D6scFv-Fc and 2F5scFv-Fc antibody variants were stably introduced into DUKX-B11 F3/F employing the tagged RMCE locus (Fig. 2; Online resource 3). The analysis of twelve subclones of each variant showed similar growth characteristics but protein specific differences in specific secretion levels (Fig. 2a, b). Importantly, the two selected scFv-Fc producing subclones for each variant proved stable and highly similar specific productivities as indicated by small error bars in Fig. 2d. This demonstrates the success of targeted integration of different transgenes by RMCE to induce predictable and reproducible expression levels in DUKX-B11 F3/F. The capability of DUKX-B11 F3/F for targeted integration of different genes of interest into RMCE target sites for reproducible expression patterns is supported by homogenous and similar intracellular scFv-Fc expression patterns (Fig. 4), as well as reproducible and stable transgene mRNA levels (Fig. 5). Thereby, DUKX-B11 F3/F is suitable to investigate product specific differences in the posttranscriptional expression cascade under a chromatin status-independent setup.

A factor of 2- to 3-fold higher specific productivity could be detected for 3D6scFv-Fc subclones compared with 2F5scFv-Fc subclones using DUKX-B11 F3/F as the expression host for targeted integration of the transgenes by RMCE. The pronounced differences in specific productivity were observable during continuous T25 and spinner flask cultivation as well as in batch culture experiments. Mader et al. (2013) reported a factor of three to 4-fold lower specific productivity of 2F5scFv-Fc-producing subclones compared with 3D6scFv-Fc cell clones established by random

introduction of conventional plasmid expression vectors or alternatively by bacterial artificial chromosomes (BAC) harboring the same expression elements.

The four step procedure to generate antibody producing DUKX-B11 clones presented in this study has certain advantages compared with simple two-step "tag-and-exchange" strategies reported in literature (Askew et al. 1993). The introduction of a first gfp cassette under control of a trapped promoter allows the easy verification of transfection efficiency in the transient stage. At this point also, a selection of gfp high-producing cells could be performed by fluorescence-activated cell sorting as was demonstrated in Qiao et al. (2009). It is known that the efficiency of RMCE is dependent on the chromosomal position of the integrated target cassette as was demonstrated by Vooijs et al. (2001) for the Cre/loxP system. CHO cells are especially prone to spontaneous genetic rearrangements (Cao et al. 2012; Kim et al. 2011) that might lead to a loss of RMCE capability of the targeted chromosomal cassette. To ensure a RMCE-competent locus, the stably introduced hyg/tk cassette was exchanged by two additional RMCE reactions to introduce the gfp/tk/neo fusion and finally the scFv-Fc antibody cassette. The finally generated scFv-Fc producer cell lines demonstrate the long-term potency of the target site for RMCE by two different transgene cassettes with varying nucleotide sequence length between the two heterospecific *FRT* sites.

Although the primary goal of this study was not to generate a RMCE host cell line conferring high specific productivities, if desired, improvements in qP can be reached by systematic use of S/MAR elements (Qiao et al. 2009). An alternative option is enabled by lentiviral transduction, at the single-copy level into transcriptionally competent loci as described by Oberbek et al. (2011). The system could also be improved by using bacterial artificial chromosomes (BAC) as open chromatin and copy number-dependent expression vehicles (Blaas et al. 2009) in combination with serial, accumulative RMCE reactions for introduction of multiple gene copies with various available heterospecific *FRT* sites into the same genomic locus. We wish to emphasize, however, that for present purposes, a less stringent selection procedure for choosing chromosomal loci with moderate expression capabilities was used to investigate transgene-intrinsic limitations in low-producing antibody variants (exemplified by 2F5scFv-Fc) that might otherwise be overlaid by the high-expression activity of a high-expression locus.

Global analysis methods in systems biology to identify bottlenecks common to different antibody producer cell clones attract more and more attention (Gupta and Lee 2007). One limitation in different -omics techniques is the generation of subclones with identical transcription capabilities derived from the same transfection reaction or, even more challenging, derived from different transfection reactions of different antibody variants. The difficulties arise on the one hand from randomly introduced expression cassettes leading to unpredictable gene copy numbers and different transcription efficiencies influenced by different surrounding chromosomal environments (position effect). We here propose that the new available DUKX-B11 F3/F cell line is a RMCE-competent host to facilitate the rapid generation of subclones with comparable and stable gene copy numbers, mRNA, and intracellular product levels guided by integration into the preselected chromosomal locus. Proof of concept was provided during the host cell development with the exchange of the initial cassette for the active heterologous transgene cassette (CMV-F3-gfp/tk/neo-pA-F). Therefore, it is expected that exchange of this cassette with any scFv-Fc coding sequence will lead to clones that are more robust against epigenetic silencing and genetic instabilities compared with subclones generated by transgene integration into "new and unknown" chromosomal sites by random integration events. It can be expected that the DUKX-B11 F3/F host cell line is useful in systems biology in general and in following proteomics studies, as an explicit application, to establish different antibody producing cell lines with moderate but highly similar and robust expression capabilities more rapidly since the adaption process is shortened by targeted integration into an active and stable locus.

Further analysis of the generated scFv-Fc-producing cell clones by qPCR analysis with primers specific for targets other than β-actin might be used for a more in-depth comparison of the generated scFv-Fc-producing subclones. Such targets might include components of the ER compartment, apoptotic pathway, or folding/secretion machinery.

In summary, a stable RMCE host, DUKX-B11 F3/F, could be established with homogenous expression capabilities of a gtn fusion protein. The stably integrated RMCE target sites can be replaced by different genes of interest leading to homogenous cell populations. We were able to develop recombinant clones with stable and reproducible mRNA levels and intracellular product formation, as well as similar growth and expression capabilities controlled by targeted integration of transgene cassettes into the same chromosomal environment by RMCE.

Our study significantly expands existing RMCE tools by a slight modification of already existing strategies to generate genomic target loci to guide different transgenes specifically into chromosomal loci with reproducible transcription characteristics.

Acknowledgments The work of PM and WSo was funded by the PhD program "BioToP-Biomolecular Technology of Proteins" (Austrian Science Funds, FWF Project W1224). Work in the MHH laboratory Work in the authors' laboratory (ST, JQ, JB) was supported by the CliniGene Network of Excellence (European Commission FP6 Research Program, contract LSHBCT-2006-018933 until 2011), the Excellence Initiative "REBIRTH" (REgenerative BIology to Reconstructive Therapy), the SFB 738 (Optimierung konventioneller und innovativer Transplantate), and a ReGene (Regenerative Medicine and Biology) grant by the BMBF.

Conflict of interest The authors declare that they have no conflict of interest.

References

Aggarwal SR (2014) What's fueling the biotech engine—2012 to 2013. Nat Biotechnol 32(1):32–9. doi:10.1038/nbt.2794

Askew GR, Doetschman T, Lingrel JB (1993) Site-directed point mutations in embryonic stem cells: a gene-targeting tag-and-exchange strategy. Mol Cell Biol 13(7):4115–24. doi:10.1128/MCB.13.7.4115

Blaas L, Musteanu M, Eferl R, Bauer A, Casanova E (2009) Bacterial artificial chromosomes improve recombinant protein production in mammalian cells. BMC Biotechnol 9:3. doi:10.1186/1472-6750-9-3

Cao Y, Kimura S, Itoi T, Honda K, Ohtake H, Omasa T (2012) Construction of BAC-based physical map and analysis of chromosome rearrangement in Chinese hamster ovary cell lines. Biotechnol Bioeng 109(6):1357–67. doi:10.1002/bit.24347

Clarke C, Doolan P, Barron N, Meleady P, O'Sullivan F, Gammell P, Melville M, Leonard M, Clynes M (2011) Predicting cell-specific productivity from CHO gene expression. J Biotechnol 151(2):159–65. doi:10.1016/j.jbiotec.2010.11.016

Datta P, Linhardt RJ, Sharfstein ST (2013) An 'omics approach towards CHO cell engineering. Biotechnol Bioeng 110(5):1255–71. doi:10.1002/bit.24841

Dietmair S, Nielsen LK, Timmins NE (2012) Mammalian cells as biopharmaceutical production hosts in the age of omics. Biotechnol J 7(1):75–89. doi:10.1002/biot.201100369

Doolan P, Clarke C, Kinsella P, Breen L, Meleady P, Leonard M, Zhang L, Clynes M, Aherne ST, Barron N (2013) Transcriptomic analysis of clonal growth rate variation during CHO cell line development. J Biotechnol 166(3):105–13. doi:10.1016/j.jbiotec.2013.04.014

Fukushige S, Sauer B (1992) Genomic targeting with a positive-selection lox integration vector allows highly reproducible gene expression in mammalian cells. Proc Natl Acad Sci U S A 89(17):7905–9

Gupta P, Lee KH (2007) Genomics and proteomics in process development: opportunities and challenges. Trends Biotechnol 25(7):324–30. doi:10.1016/j.tibtech.2007.04.005

Huang Y, Li Y, Wang YG, Gu X, Wang Y, Shen BF (2007) An efficient and targeted gene integration system for high-level antibody expression. J Immunol Methods 322(1–2):28–39. doi:10.1016/j.jim.2007.01.022

Kantardjieff A, Nissom PM, Chuah SH, Yusufi F, Jacob NM, Mulukutla BC, Yap M, Hu WS (2009) Developing genomic platforms for Chinese hamster ovary cells. Biotechnol Adv 27(6):1028–35. doi:10.1016/j.biotechadv.2009.05.023

Kim M, O'Callaghan PM, Droms KA, James DC (2011) A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. Biotechnol Bioeng 108:2434–2446. doi:10.1002/bit.23189

Kim JY, Kim YG, Lee GM (2012) CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Appl Microbiol Biotechnol 93(3):917–30. doi:10.1007/s00253-011-3758-5

Kito M, Itami S, Fukano Y, Yamana K, Shibui T (2002) Construction of engineered CHO strains for high-level production of recombinant proteins. Appl Microbiol Biotechnol 60(4):442–8. doi:10.1007/s00253-002-1134-1

Kuehle J, Turan S, Cantz T, Hoffmann D, Suerth JD, Maetzig T, Zychlinsk D, Klein C, Steinemann D, Baum C, Bode J, Schambach A (2014) Modified lentiviral LTRs allow Flp recombinase-mediated cassette exchange and in vivo tracing of „factor-free" induced pluripotent stem cells. Mol Ther 22(5):919–28. doi:10.1038/mt.2014.4

Kwaks TH, Otte AP (2006) Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. Trends Biotechnol 24(3):137–42. doi:10.1016/j.tibtech.2006.01.007

Lattenmayer C, Trummer E, Schriebl K, Vorauer-Uhl K, Mueller D, Katinger H, Kunert R (2007) Characterisation of recombinant CHO cell lines by investigation of protein productivities and genetic parameters. J Biotechnol 128(4):716–25. doi:10.1016/j.jbiotec.2006.12.016

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25(4):402–8. doi:10.1006/meth.2001.1262

Mader A, Prewein B, Zboray K, Casanova E, Kunert R (2013) Exploration of BAC versus plasmid expression vectors in recombinant CHO cells. Appl Microbiol Biotechnol 97(9):4049–54. doi:10.1007/s00253-012-4498-x

Mutskov V, Felsenfeld G (2004) Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. EMBO J 23(1):138–49. doi:10.1038/sj.emboj.7600013

Nehlsen K, Schucht R, da Gama-Norton L, Krömer W, Baer A, Cayli A, Hauser H, Wirth D (2009) Recombinant protein expression by targeting pre-selected chromosomal loci. BMC Biotechnol 9:100. doi:10.1186/1472-6750-9-100

Oberbek A, Matasci M, Hacker DL, Wurm FM (2011) Generation of stable, high-producing CHO cell lines by lentiviral vector-mediated gene transfer in serum-free suspension culture. Biotechnol Bioeng 108(3):600–10. doi:10.1002/bit.22968

O'Gorman S, Fox DT, Wahl GM (1991) Recombinase-mediated gene activation and site-specific integration in mammalian cells. Science 251(4999):1351–5. doi:10.1126/science.1900642

Qiao J, Oumard A, Wegloehner W, Bode J (2009) Novel tag-and-exchange (RMCE) strategies generate master cell clones with predictable and stable transgene expression properties. J Mol Biol 390(4):579–94. doi:10.1016/j.jmb.2009.05.012

Raymond CS, Soriano P (2007) High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. PLoS One 2(1):e162. doi:10.1371/journal.pone.0000162

Richards EJ, Elgin SC (2002) Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell 108(4):489–500. doi:10.1016/S0092-8674(02)00644-X

Sommeregger W, Prewein B, Reinhart D, Mader A, Kunert R (2013) Transgene copy number comparison in recombinant mammalian cell lines: critical reflection of quantitative real-time PCR evaluation. Cytotechnology 65(5):811–8. doi:10.1007/s10616-013-9606-y

Turan S, Zehe C, Kuehle J, Qiao J, Bode J (2013) Recombinase-mediated cassette exchange (RMCE)—a rapidly-expanding toolbox for targeted genomic modifications. Gene 515(1):1–27. doi:10.1016/j.gene.2012.11.016

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35:W71–4. doi:10.1093/nar/gkm306

Urlaub G, Chasin L (1980) Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. Proc Natl Acad Sci U S A 77:4216–4220

Vooijs M, Jonkers J, Berns A (2001) A highly efficient ligand-regulated Cre recombinase mouse line shows that LoxP recombination is position dependent. EMBO Rep 2(4):292–7. doi:10.1093/embo-reports/kve064

Wilson C, Bellen HJ, Gehring WJ (1990) Position effects on eukaryotic gene expression. Annu Rev Cell Biol 6:679–714. doi:10.1146/annurev.cb.06.110190.003335

Wurm FM (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 22(11):1393–8. doi:10.1038/nbt1026

Supplementary Material

Applied Microbiology and Biotechnology

Accurate comparison of antibody expression levels by reproducible transgene targeting in engineered recombination-competent CHO cells

Mayrhofer P.[1], Kratzer B.[2], Sommeregger W.[1], Steinfellner W.[1], Reinhart D.[1], Mader A.[1], Turan S.[3], Qiao J.[4], Bode J.[5] , Kunert R.[1]*

[1] Department of Biotechnology/Vienna Institute of BioTechnology (BOKU – VIBT), University of Natural Resources and Life Sciences (Vienna), Muthgasse 18, A-1190 Vienna, Austria

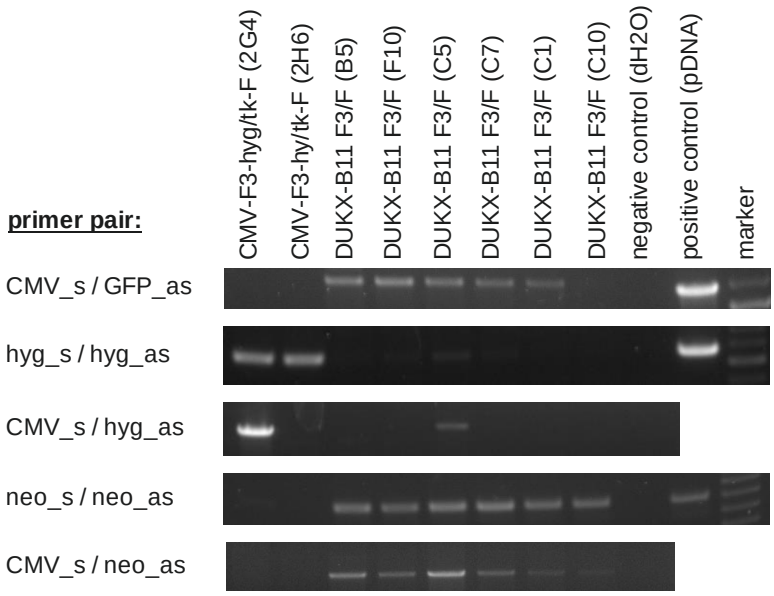
[2] Institute of Immunology, Medical University Vienna, Borschkegasse 8a, A-1090 Vienna, Austria

[3] Department of Genetics, Stanford University School of Medicine, Alway Building, Room M-334, 300 Pasteur Drive, Stanford, CA 94305-5120, USA

[4] WuXi App Tec Co. Ltd, 288 Fute Zhong Road, Shanghai 200131, China

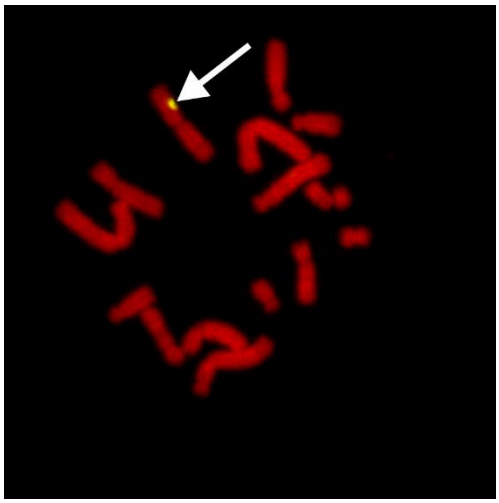
[5] Hannover Medical School/Institute for Experimental Hematology, OE 6960, Carl-Neuberg-Str.1, D-30625 Hannover, Germany

* Corresponding author: email: renate.kunert@boku.ac.at; tel: +43 1 47654 6595; fax: +43 1 47654 6675



Online Resource 1 PCR characterization of genomic DNA isolated from different CMV-F3-hyg/tk-F or DUKX-B11 F3/F subclones as summarized in Table 1.

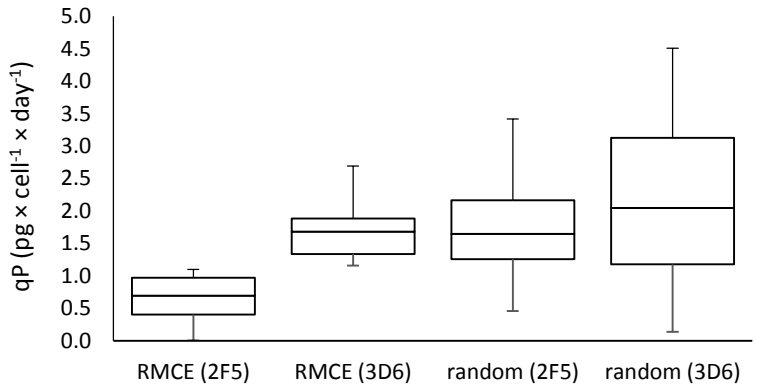
Subclone CMV-F3-hyg/tk-F (2G4) showed an authentic RMCE reaction with the first RMCE donor cassette (step 2 in Fig. 1) and was selected for the second RMCE reaction resulting in the final DUKX-B11 F3/F (B5) clone showing the gfp/tk/neo (GTN) cassette integrated behind the CMV promoter by RMCE. If available, plasmid DNA (pDNA) was used as positive control



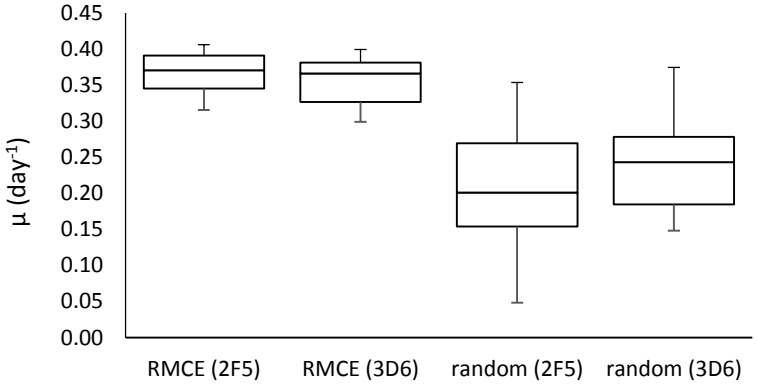
Online Resource 2 Fluorescence in-situ hybridization (FISH) analysis of DUKX-B11 F3/F with probes specifically binding to the CMV-F3-gfp/tk/neo-F RMCE target site.

3 µg of plasmid DNA pF3-gfp/tk/neo-F was labeled using the DIG-Klenow-labeling kit DIG High prime (Roche) according to the manufacturer's protocol. For preparation of metaphase spreads cells in exponential phase were treated with 0.2 µg/mL demecolcine for 4 hours. DIG-labeled and hybridized probes were detected using triple detection antibody setup. Sequential addition of mouse anti-DIG (Roche), goat anti-mouse-FITC and rabbit anti-goat-FITC detection antibodies gave a single fluorescent locus (white arrow) under a SP5 II laser scanning confocal microscope (Leica) at 488 nm using propidium iodide counter-staining at 561 nm

a.)

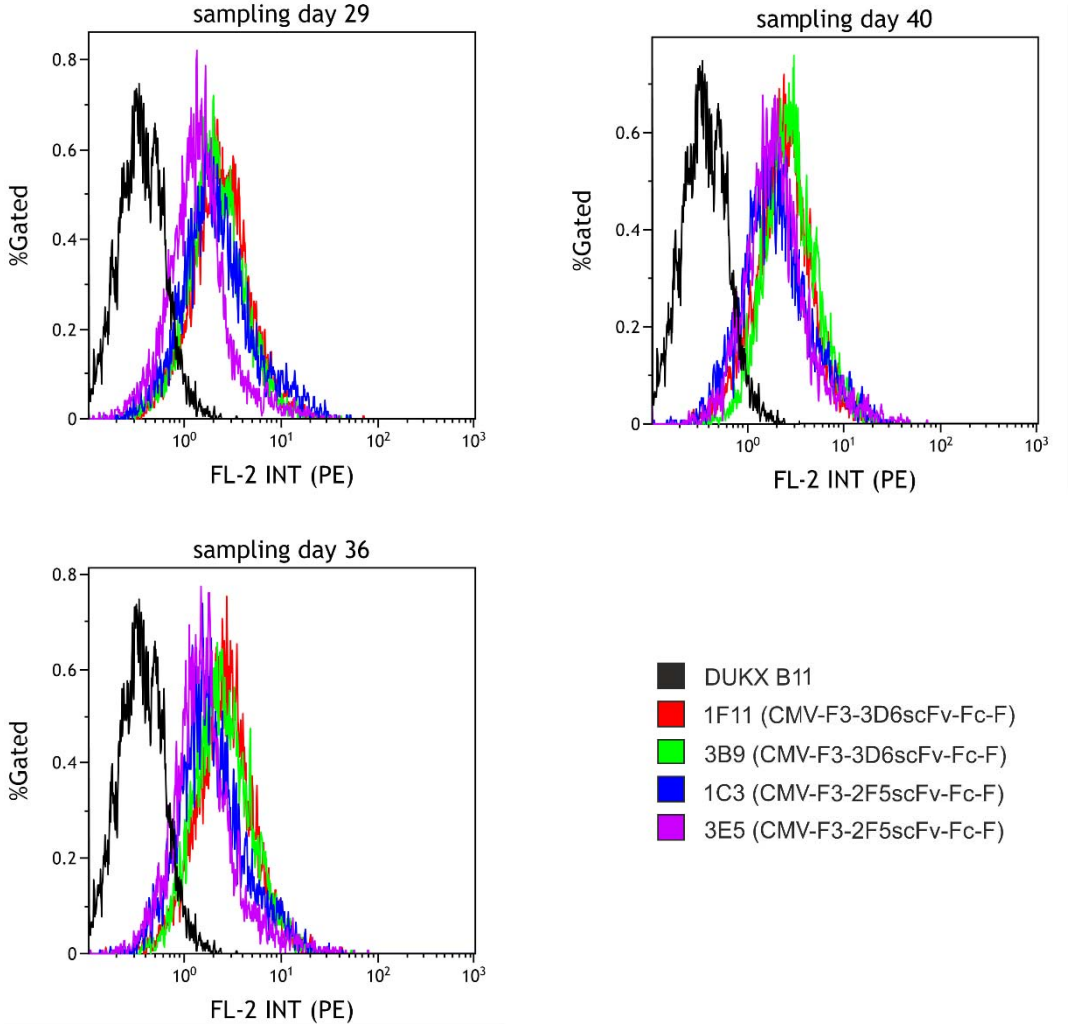


b.)



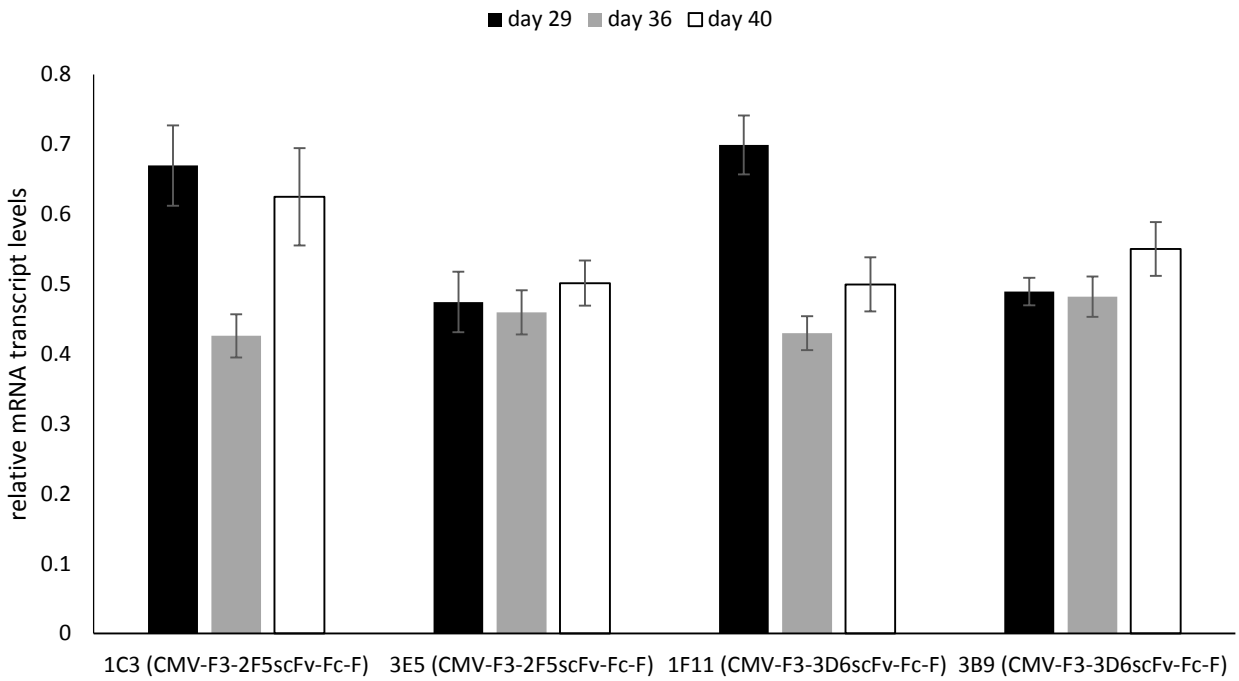
Online Resource 3 Distribution of specific productivities and growth rates of scFv-Fc producing subclones developed by targeted RMCE compared to random integration of the transgene.

RMCE subclones were developed as described in Materials and Methods and analyzed as described in Fig. 2a and Fig. 2b. Subclones generated by random integration of the plasmid DNA were established by using the same transfection and selection procedures as with the RMCE subclones. (**A**) Specific productivities (qP) with an interquartile range (midspread) of 0.56 and 0.54 pg × $cell^{-1}$ × day^{-1} of the 2F5scFv-Fc and 3D6scFv-Fc subclones generated by RMCE or 0.91 and 1.95 pg × $cell^{-1}$ × day^{-1} of 2F5scFv-Fc and 3D6scFv-Fc generated by random integration of the transgenes, respectively. (**B**) Specific growth rates (μ) with an interquartile range (midspread) of 0.05 and 0.05 day^{-1} of the 2F5scFv-Fc and 3D6scFv-Fc subclones generated by RMCE or 0.12 and 0.09 day^{-1} of 2F5scFv-Fc and 3D6scFv-Fc generated by random integration of the transgenes, respectively



Online Resource 4 Homogenous intracellular product formation of scFv-Fc producing subclones at three independent sampling days measured by flow cytometry.

ScFv-Fc producing subclones cultured in spinner flasks were fixed by ice-cold ethanol and labeled with anti-huIgG-γ-chain-R-phycoerythrin (PE) antibody. Single parameter FL-2 histograms were used to determine the mean values of the median fluorescence intensities (MFI) at three independent sampling days, as indicated in Fig. 4b



Online Resource 5 Relative mRNA transcript levels of two 2F5scFv-Fc producing subclones and two 3D6scFv-Fc producing subclones sampled at three independent culture days.

Samples were measured in two technical and three biological replicates. Total mRNA was reverse transcribed into cDNA and analyzed by qPCR using probes specific for the Fc sequence or β-actin used as an internal standard. Mean $2^{-\Delta Cp}$ values were calculated based on differences of Cp values between β-actin and the Fc sequence. Error bars represent standard deviation

Title:

Cloning of single-chain antibody variants by overlap-extension PCR for evaluation of antibody expression in transient gene expression

Running head:

scFv-Fc antibody cloning

Authors:

Patrick Mayrhofer and Renate Kunert*

Vienna Institute of BioTechnology, Department of Biotechnology
University of Natural Resources and Life Sciences, Vienna
Muthgasse 18, 1190 Vienna
E-mail: renate.kunert@boku.ac.at

*To whom correspondence should be addressed

Abstract:

Single-chain fragment variable – fragment crystallizable antibody constructs (scFv-Fc) are homodimeric proteins representing valuable alternatives to heterotetrameric full-length IgG molecules to study protein properties and product dependent cellular behavior. In contrast to naturally occurring antibodies, these artificial molecules are assembled from functional antibody domains to reduce molecule complexity and enhance antibody expression levels. The scFv-Fc format retains critical antibody functions such as antigen binding affinity and antibody effector functions. Here we present a protocol to convert the full-length anti-HIV-1 IgG1 antibody 2F5 into a scFv-Fc construct. Variable and constant regions are amplified by conventional PCR reactions and assembled by a single overlap-extension PCR reaction. The amplified product is then cloned into a mammalian expression vector suitable for high-titer transient gene expression. This workflow can be applied to any antibody sequence by adapting the specific primer sequences to the antibody of choice.

Keywords:

HEK293, monoclonal antibodies, transient gene expression, anti-HIV-1 2F5

1. Introduction

Since their description as magic bullets, monoclonal antibodies (mAbs) have been continuously developed to represent the fastest growing class of biotherapeutic proteins with US sales reaching $24.6 billion in 2012 (**1**).

Without doubt, the primary function of these biomolecules is the affinity to their respective antigen. High affinity antigen binding is mediated by the specific amino acid sequence of the complementarity-determining regions (CDR) present in the variable regions of an antibody molecule. The correct conformation of the CDR loops is supported by flanking framework (FR) regions. The constant regions connected to the C-terminus of variable regions define the isotype of an antibody. Naturally occurring isotypes can be assigned to subclasses IgA, IgD, IgE, IgG and IgM determined by the different and rather invariable ("constant") regions of a full-length antibody molecule. For an IgG1 heavy chain this includes the constant heavy 1 (CH1), constant heavy 2 (CH2) and constant heavy 3 (CH3) region. For the light chain, the constant part is established by either a kappa (κ) or lambda (λ) constant region. The molecular structure of an antibody can further be classified by the functional regions. The fragment crystallizable (Fc) was shown to force crystallization of mAbs in X-ray crystallographic studies and consists of region CH2 and CH3. The antigen binding fragment (Fab) contains the variable fragment (Fv) region, formed by the heavy (vH) and light chain (vL) variable domain together with the CH1 and light chain constant region. Distinct regions in the Fc part are responsible for antibody receptor-mediated effector functions. For example, within the human body the Fc part interacts with molecules of the complement system or Fc-gamma receptors (FcγR) to mediate complement-dependent cytotoxicity (CDC) or antibody-dependent cellular cytotoxicity (ADCC), respectively (**2**). In addition to complete mAbs, corresponding antibody fragments or fusion proteins thereof are applied in clinical applications with increasing attention (**3, 4**). For various reasons, smaller molecules are preferable in human application (e.g. to cross the blood-brain barrier) which led to the definition of scFv molecules representing a single-chain protein of the two variable antibody regions connected by a synthetic linker peptide. The linker enables flexible interaction of the vH and the vL sequence to form the antigen binding pocket. Such scFv fragments can be produced in prokaryotic expression systems but they tend to be unstable with limited in vivo half-life. Alternatively, single-chain fragment variable – fragment crystallizable (scFv-Fc) constructs represent homodimeric biomolecules with reduced complexity compared to heterotetrameric IgG1 antibody molecules but maintained critical antibody functions, such as antigen binding and receptor-mediated effector functions. Another advantage of these molecules in technological aspects is that only one chain has to be expressed within the mammalian host to assemble a fully functional scFv-Fc molecule by disulfide-bridge dimerization of two identical chains containing leader-vH-linker-vL-hinge-CH2-CH3.

In this protocol we describe the construction of scFv-Fc antibodies based on plasmid templates encoding full-length IgG1 sequences (**Figure 1**). Critical domains (leader, vH, vL, hinge-CH2, CH3) are

amplified by conventional PCR reactions using primers containing linker sequences (**Table I**) and assembled by overlap-extension PCR. The PCR constructs are then ligated into a commercially available mammalian expression plasmid and can then be transfected into HEK293 cells for efficient transient expression of the scFv-Fc molecules.

2. Materials:

1. Use sterile pipette tips, tubes and autoclaved ultrapure deionized water (dH_2O).

2. Plasmids:
   Commercially available pCEP4 vector (Invitrogen), a high-copy number plasmid containing an Epstein-Barr virus nuclear antigen 1 (EBNA-1) expression cassette and origin of replication (oriP) required for episomal plasmid replication and propagation to daughter cells after cell division. High-level transcription is mediated by a cytomegalovirus (CMV) immediate early enhancer/promoter. For 2F5 template sequences any plasmid containing heavy- or light chain of the IgG1 anti-HIV-1 antibody 2F5 can be used.

3. LB-amp medium and agar plates:
   10 g/L tryptone, 5 g/L yeast extract, 170 mM NaCl, pH 7.0, 1.5% agar-agar (Merck) for plates, 100 µg/mL ampicillin.

4. Plasmid purification:
   LB-amp medium, small-scale purification using Plasmid Miniprep Kit I peqGOLD (peqlab) and large-scale purification with NucleoBond Xtra Midi EF (Macherey-Nagel), thermoshaker incubator.

5. DNA quantification by Nanodrop 2000 (Thermo Scientific).

6. Gel-electrophoresis:
   Bromophenol Blue/Xylene Cyanol FF (BX) DNA loading dye (6×, Thermo Fisher), Generuler DNA ladder mix (Thermo Fisher), agarose gels consist of 1% (w/v) agarose (peqlab) in Tris (0.5 M)-Acetate-EDTA (0.05 M) (TAE) buffer and 0.3 µg/mL ethidium bromide (EtBr), gel-electrophoresis chamber (Biorad) with TAE-buffer (+ 0.3 µg/mL EtBr), Gel Doc illumination chamber (Biorad).

7. High-fidelity PCR for fragment amplification or overlap-extension PCR:
   KAPA HiFi PCR kit (peqlab) containing KAPA HiFi DNA Polymerase (1 U/µL), dNTP mix (40 mM) and fidelity reaction buffer (5×), C1000 thermal cycler (Biorad).

8. Colony PCR for screening *E. coli* colonies:
   Taq DNA polymerase kit (NEB) containing Taq DNA Polymerase (5 U/µL) and Thermopol reaction buffer (10×), dNTP mix (40 mM), C1000 thermal cycler (Biorad).

9. Oligonucleotide primer sequences (**Table I**) ordered from Sigma-Aldrich (**Note 4**).

[**Table I** near here**]**

10. DNA purification by isopropanol precipitation:
    3 M sodium acetate adjusted to pH 5.2, isopropanol, 70% (v/v) ethanol.

11. Agarose gel extraction and purification of PCR products:
    Wizard Plus SV Gel and PCR clean up system (Promega).

12. Restriction of overlap-extension PCR amplicon and plasmids:
    High-fidelity (HF) variant of KpnI (NEB) and XhoI (NEB), restriction buffer CutSmart (NEB).

13. Ligation: T4 DNA Ligase (400 U/µL, NEB), T4 DNA ligase reaction buffer (10×, NEB).

14. SOC-medium:
    20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl, 3 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose.

15. Transformation and Cryopreservation: electrocompetent *E. coli* Top10 (Invitrogen), SOC medium, LB-amp agar plates, Gene Pulser Xcell electroporator (Biorad), Thermomixer Comfort (Eppendorf), glycerol (Merck).

16. Transfection:
    HEK293-6E cells (National Research Council, Canada) stably expressing the EBNA-1 protein for episomal replication of plasmids containing oriP sequences, Freestyle F17 expression medium (Invitrogen), 4 mM L-glutamine (Biochrom), 0.1% Kolliphor P188 (Sigma-Aldrich), 15 mg/L phenol red (Sigma-Aldrich), 25 µg/mL G418 (Biochrom), linear polyethylenimine (25 kDa, Polysciences), tryptone N1 (TekniScience) , 50 mL conical cell culture tubes (Corning), climo-shaker ISF1-XC (Kuhner).

3. Methods

In a first step three individual fragments are amplified from different plasmids containing heavy- or light chain variable and constant (hinge-CH2-CH3) sequences of the IgG1 antibody 2F5. These fragments comprise i) leader-vH-linker ii) linker-vL-hinge and iii) hinge-CH2-CH3 (**Figure 1A**). The three fragments are then combined in one reaction for overlap-extension PCR (**Figure 1B**). Afterwards the 2F5 scFv-Fc fragment is cloned (**Figure 1C**) into the commercially available mammalian expression vector pCEP4 (**Note 1**). The established mammalian expression vector containing the scFv-Fc construct is then subjected to transient gene expression in mammalian cells to yield the fully functional antibody variant (**Figure 1D**).

3.1 Generation of 2F5 fragment 1, 2 and 3 from IgG template plasmids

1. Generate and purify template plasmid DNA for fragment amplification using a "Miniprep" purification kit according to manufacturer's protocol (**Note 2**):
Inoculate 10 mL LB-amp medium with a cryopreserved *E. coli* stock, containing the heavy- or light chain sequence of the 2F5 IgG1 molecule, using a sterile pipette tip. Incubate for 12-16 h at 37°C and 200 rpm. Pellet the cell suspension by centrifugation for 10 min at 5,000 × g and 4°C. Discard the culture supernatant. Resuspend the pellet in 250 µL solution I (complemented with RNase A), transfer everything into an Eppendorf tube and vortex vigorously to disrupt any cell clumps (critical step to enhance efficiency of cell lysis). Cell lysis is initiated by adding 250 µL alkaline solution II and mixing by inverting the tube six times. Incubate for two minutes at room temperature. Neutralize the clear lysate by adding 350 µL solution III and mix by inverting the tube six times. Centrifuge the tube for 10 min at 10,000 × g and 4°C to remove precipitated proteins and genomic DNA. Transfer the supernatant onto a silica membrane column and centrifuge for 1 min at 10,000 × g and 4°C. Discard the flow-through and wash the bound plasmid DNA with 500 µL "PW" plasmid kit-buffer and 750 µL DNA wash kit-buffer by centrifugation for 1 min at 10,000 × g and 4°C after each washing step. The purified plasmid DNA is then eluted from the column by adding 100 µL autoclaved deionized water (dH_2O). Concentration and quality of the plasmid DNA solution is determined by Nanodrop (**Note 3**). This procedure usually gives yields higher than 10 µg for a 10 mL *E. coli* suspension carrying the pCEP4 vector. Adjust plasmid DNA concentration to 100 ng/µL.

2. Preparation of primer stock solutions:
All lyophilized primer samples (**Table I**) are resuspended in dH_2O according to manufacturer's instructions to give a 100 µM primer master stock solution that is routinely stored at -20°C (**Note 4**). A 10 µM primer working stock solution is prepared and used for setting up the PCR reactions.

3. Preparation of a PCR master mix:
For amplifying the three 2F5 fragments in three separate PCR reactions a master mix (3 ×) containing all substances but primers and plasmid DNA is prepared: At first, mix 50.3 µL dH_2O with 15 µL fidelity reaction buffer, 2.3 µL dNTP mix and 1.5 µL KAPA DNA polymerase (**Note 5**). Mix, vortex and spin down. Prepare three PCR-tubes, each containing 0.75 µL of one forward primer (primer 1, 3 or 5), 0.75 µL of one reverse primer (primer 2, 4 or 6) and 0.5 µL plasmid template DNA (50 ng total per reaction, two tubes containing the 2F5 heavy chain sequence with primer pair 1+2 or 5+6 and one tube containing the 2F5 light chain plasmid with primer pair 3+4). 23 µL of PCR master mix is added to each tube.

4. PCR and gel-electrophoresis:
Use the thermocycler program 1 (**Table II**) to amplify each fragment from the 2F5 template plasmids. After finishing the PCR reaction, 5 µL of BX-buffer is added and the amplicons are loaded individually on a 1% agarose gel (+ 0.3 µg/mL EtBr) for gel-electrophoresis in TAE buffer (+ 0.3 µg/mL EtBr) at 100 V for approx. 1h (**Note 6**).

5. Isolation of the specific PCR amplicons:
Cut the PCR amplicons showing the correct size (fragment 1: 512 bp, fragment 2: 362 bp, and fragment 3: 710 bp) under UV-light (**Note 8**). Extract and purify the DNA from the sliced agarose gel using commercial extraction kits (**Note 9**). Elute purified fragments with dH_2O. Measure DNA concentrations and adjust concentration to 10 ng/µL (**Note 10**).

3.2 Overlap-extension PCR

To generate the full-length insert by overlap-extension PCR a total of 50 ng DNA template (fragment 1, 2 and 3) is used in the PCR reactions. Equimolar amounts for each individual fragment are adjusted based on the fragment length of each fragment. As a first PCR step, 10 cycles without any primers are initiated by using thermocycler program 2 (**Table II**) to generate full-length templates based on overlapping regions of the three fragments. A second PCR is initiated with the same solution by adding outer primer 1 and 6 followed by PCR program 3 (**Table II**).

1. PCR without primers:
Mix 12.5 µL dH_2O, 5 µL fidelity reaction buffer, 0.75 µL dNTP, 1.6 µL fragment 1 (16 ng) , 1 µL fragment 2 (10 ng), 2.2 µL fragment 3 (22 ng) and 0.5 µL KAPA DNA Polymerase. Vortex and spin down. Start the PCR cycle program 2 (**Table II**).

2. Amplification of full-length template using outer primers:
Add 0.75 µL primer 1 and 0.75 µL primer 6 to the PCR reaction solution and start PCR cycle program 3 (**Table II**). After the PCR reaction has finished, 5 µL of BX-buffer is added and the solution loaded onto

a 1% agarose gel (+0.3 µg/mL EtBr) for gel-electrophoresis in TAE buffer (+0.3 µg/mL EtBr) at 100 V for approx. 1h.

3. Isolation of the specific PCR amplicon:
Cut the PCR amplicon showing the correct size (1,543 bp) under UV-light. Purify the DNA from the agarose gel using commercial kits. Elute the purified fragment with 50 µL dH_2O (**Note 10**).

3.3 Restriction and cloning into pCEP4 expression vector

1. Generation of sticky ends by KpnI/XhoI double-digestion of the PCR amplicon:
Add 5 µL CutSmart buffer, 0.5 µL KpnI and 0.5 µL XhoI to the eluted PCR amplicon solutions and incubate for 3h at 37°C. Heat inactivate for 20 min at 65°C (**Note 11**).

2. Purification of restricted PCR amplicon by isopropanol precipitation:
Pre-cool centrifuge to 4°C. Add 5 µL of 3 M sodium acetate (pH 5.2) and 35 µL isopropanol. Pellet the DNA precipitate by centrifugation at 15,000 × g for 30 min. Remove the supernatant by carefully inverting the tube without disturbing the pellet (hardly visible). Wash the pellet with 1 mL 70% (v/v) ethanol. Centrifuge at 15,000 × g for 15 min, decant the supernatant and dry the pellet for 5-20 min. Dissolve in 15 µL dH_2O.
Measure DNA concentrations and adjust concentration to 10 ng/µL.

3. Restriction of pCEP4 vector backbone:
Digest 3 µg pCEP4 plasmid DNA in a total of 50 µL containing 5 µL CutSmart buffer, 0.5 µL KpnI and 0.5 µL XhoI in dH_2O. Incubate for 3h at 37°C. Heat inactivate for 20 min at 65°C (**Note 12**).

4. Purification of digested pCEP4 vector backbone:
Add 10 µL BX-buffer to 50 µL inactivated restriction solution, load onto a 1% agarose gel (+ 0.3 µg/mL EtBr) for gel-electrophoresis in TAE buffer (+0.3 µg/mL EtBr) at 100 V for approx. 1h. Cut the linearized plasmid DNA showing the correct size (10,157 bp) under UV-light. Extract and purify the DNA from the agarose gel using commercial extraction kits. Elute the linearized vector with 50 µL dH_2O. Measure DNA concentrations and adjust to 10 ng/µL.

5. Ligation:
To find the optimal molar insert : vector ratio and to check for ligation efficiency three parallel reactions are set up. Prepare a master mix (3 ×) containing 29.2 µL dH_2O, 15 µL linearized pCEP4 vector (50 ng per reaction), 6 µL T4 DNA ligase reaction buffer, 3 µL T4 DNA ligase. Transfer 17.7 µL of the master mix into each of three tubes. Add to tube 1: 2.3 µL dH_2O, to tube 2: 0.8 µL digested 2F5scFv-Fc insert (8 ng) and 1.5 µL dH_2O and to tube 3: 2.3 µL digested 2F5scFv-Fc insert (23 ng),

resulting in a molar vector : insert ratio of 1 : 0, 1 : 1 and 1 : 3, respectively. Incubate for 10 min at room temperature or overnight at 16°C. Heat inactivate by incubation at 65°C for 10 min. Chill on ice.

6. Transformation into electro-competent TOP10 *E. coli:*
Chill electroporation cuvettes on ice and gently thaw 40 µL TOP10 aliquots on ice. 3 µL ligation mixture is added to 40 µL electrocompetent TOP10 *E. coli* aliquots followed by applying an electric pulse (1.8 kV, 25 µF, 200 Ω). Immediately, 250 µL SOC medium is added to regenerate cells by incubation for one hour at 37°C and 400 rpm on a thermoshaker incubator to induce the ampicillin resistance gene. 50 and 100 µL of the cell suspension is plated onto LB-amp selective agar (100 µg/mL ampicillin) and incubated overnight at 37°C.

7. Colony PCR:
For preparation of a master mix (10 ×) use 256.8 µL dH_2O, 30 µL ThermoPol buffer, 6 µL dNTP mix, 3 µL forward primer 7 (screen_seq_CMV_for), 3 µL reverse primer 8 (screen_2F5vL_rev), and 1.2 µL Taq DNA polymerase (**Note 13**). Aliquote 30 µL in PCR tubes. Prepare Eppendorf tubes containing 50 µL LB-amp medium and one LB-amp agar plate. By using sterile toothpicks or pipette tips transfer a single colony of the overnight LB-amp agar plate into the colony PCR solution and stir. Transfer the same toothpick or pipette tip into the tube with liquid LB-amp medium and then streak onto the LB-amp plate to inoculate the liquid culture and the LB-agar plate as a backup. Repeat this step for 10 individual colonies (**Note 14**). Incubate the LB-amp agar backup plate overnight at 37°C and the liquid LB-amp tubes at 37°C and 400 rpm in the thermomixer. Positive clones after PCR screening are propagated in liquid overnight cultures by adding 50 µL cell suspension to 10 mL fresh LB-amp medium and incubation at 37°C and 200 rpm in the thermoshaker incubator. The next day a cryostock and DNA isolation ("Miniprep") is done. For cryopreservation of positive colonies mix 625 µL of exponential growth cultures with 375 µL 80% glycerol solution to yield a 30% glycerol cryostock that is stored at -80°C. Plasmid DNA is purified by commercial miniprep kits as described above. The integrity of the sequence is checked by restriction enzyme digestion control and DNA Sanger-sequencing . For control digestion use 0.5 µg plasmid DNA in a total of 50 µL containing dH_2O, 5 µL CutSmart buffer, 0.5 µL restriction enzyme 1 and 0.5 µL restriction enzyme 2. Restriction enzyme 1 and 2 should be chosen according to the specific antibody sequence in order to ideally cut the plasmid once in the vector sequence and once within the antibody insert. For the pCEP4_2F5scFv-Fc plasmid presented here, this can be done with AgeI (NEB) and EcoRV (NEB) to yield two fragments of 2.5 kb and 9.2 kb in length. Incubate for 3h at 37°C. Add 10 µL BX-Buffer and load 12 µL (100 ng DNA) onto a 1% agarose gel. Run the gel at 120 V for approx. 1h. For the DNA Sanger-sequencing reaction 1.2 µg plasmid DNA and 3 µL of 10 µM primer stock 7 (screen_seq_CMV_for) in a total volume of 15 µL are pre-mixed and sent for sequencing service (**Note 15**).

8. Plasmid preparation:
Enough material for transient gene expression can be purified using commercial DNA purification kits (**Note 16**)

4. Notes

Note 1: pCEP4 can be obtained from Invitrogen. Other mammalian expression vectors might be used as well.

Note 2: The procedure here is described for Plasmid Miniprep Kit I, peqGOLD (peqlab). Other commercial miniprep kits might be suitable as well.

Note 3: The quality of purified plasmid DNA is usually assessed by absorbance ratios at 260 nm / 280 nm and 260 nm /230 nm. Low protein or phenol contamination of purified plasmid DNA preparations usually give 260 nm / 280 nm ratios of about 1.8. 260 nm /230 nm ratios of 2.0 – 2.2 are generally obtained from pure preparations free of EDTA, phenols and carbohydrates.

Note 4: Primers in our lab are routinely ordered from Sigma-Aldrich using a synthesis scale of 0.025 µmol, desalted as purification method and dried:
https://www.sigmaaldrich.com/configurator/servlet/DesignTool?prod_type=STANDARD
Primers ordered from other suppliers might be suitable as well.

Note 5: Use high-fidelity DNA polymerases such as KAPA HiFi DNA polymerase (error rate: 2.8×10^{-7}). Other high-fidelity enzymes such as Phusion DNA polymerase (NEB, error rate: 4.4×10^{-7}) might be used as well using modified PCR cycle conditions.

Note 6: EtBr is strongly mutagenic. Working with EtBr requires personal protective equipment such as protective goggles, laboratory coat and nitrile gloves. Use separate space for procedures involving EtBr. If possible, use a separate room.

Note 7: Optimal annealing temperature for each primer pair can be determined using a temperature gradient for the annealing step.

Note 8: For working under UV-light use protective equipment to cover skin and eyes. Exposure time of preparative DNA samples to UV-light should be minimized to prevent degradation and introduction of random mutations.

Note 9: In our lab we routinely use Wizard Plus SV Gel and PCR clean up system (Promega) for purifying PCR reaction or agarose gels. Products from other suppliers might be used as well.

Note 10: For concentrating DNA samples isopropanol precipitation or vacuum evaporation might be used. Individual fragments can be sequenced at this point to check for errors.

Note 11: Only XhoI can be heat-inactivated. KpnI is removed by isopropanol precipitation or gel-electrophoresis.

Note 12: To check for restriction efficiency we suggest to include following controls: i) negative control containing no restriction enzyme ii) KpnI only control iii) XhoI only control. To prevent co-purification of single cut plasmids an additional step for dephosphorylation of linearized plasmids with calf intestinal alkaline phosphatase (CIP) might be included following isopropanol purification to remove active KpnI.

Note 13: For this qualitative screen a low fidelity DNA polymerase is sufficient. We routinely use the Taq DNA Polymerase (NEB).

Note 14: Success rate for finding positive colonies containing the gene of interest correctly integrated into the plasmid DNA depends on the ligation and transformation efficiency. More colonies should be screened at lower cloning efficiencies.

Note 15: This protocol is used for a Barcode Economy Run at Microsynth AG. Conditions might change for other sequencing services.

Note 16: Possible suppliers: Macherey-Nagel NucleoBond Xtra Midi EF or Maxi EF or QIAGEN Plasmid Midi Kit, Maxi kit, Mega kit or Giga kit.

5. References

1. Aggarwal, S. (2014). What's fueling the biotech engine-2012 to 2013. Nat. Biotechnol. *32*, 32–39.

2. Hansel, T.T., Kropshofer, H., Singer, T., Mitchell, J.A., and George, A.J.T. (2010). The safety and side effects of monoclonal antibodies. Nat. Rev. Drug Discov. *9*, 325–338.

3. Beck, A., Wurch, T., Bailly, C., and Corvaia, N. (2010). Strategies and challenges for the next generation of therapeutic antibodies. Nat. Rev. Immunol. *10*, 345–352.

4. Carter, P.J. (2006). Potent antibody therapeutics by design. Nat. Rev. Immunol. *6*, 343–357.

Figure Captions:

Figure 1. Assembly of scFv-Fc fragments.
Three individual fragments i) leader-vH-linker, ii) linker-vL-hinge and iii) hinge-CH2-CH3 are amplified from heavy- and light chain 2F5IgG plasmid templates with primer pairs 1+2, 3+4 or 5+6, respectively (**A**). The three fragments are then assembled to a single open reading frame by overlap-extension PCR and amplified with primers 1+6 (**B**). The single amplicon is then cloned into a suitable expression vector using flanking KpnI and XhoI restriction sites (**C**) for transient expression of the 2F5scFv-Fc homodimer (**D**). Primer sequences depicted as small arrows can be found in **Table I**.

Table Captions:

Table I. Primer sequences used for construction of 2F5scFv-Fc.
overlapping regions, ***RESTRICTION SITES***, 2F5-SPECIFIC SEQUENCES

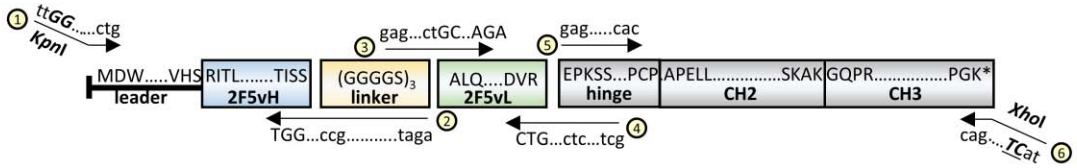
Table II. Thermocycler programs used for PCR amplification.

Tables:

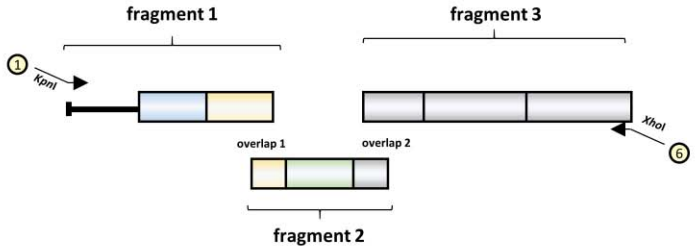
no.	Name	Sequence
①	**KpnI_HC-Leader_for**	tt***GGTACC***gccaccatggactggacctg
②	**2F5vH+linker_rev**	**agatccaccacctccgctaccgcctc**ccccagatcctccgccgccGCTGCTGATGGTCACGGT
③	**2F5vL+linker_for**	**gaggcggtagcggaggtggtggatct**GCTCTGCAGCTGACCCAGA
④	**2F5vL_hinge_rev**	**gctgctcttgggctc**CCTCACGTCCACCCTGGTC
⑤	**mutFc_for**	**gagcccaagagcagc**gacaagacccacac
⑥	**XhoI_CH3_rev**	ta***CTCGAG***ctatcacttgccgggggac
⑦	**screen_seq_CMV_for**	atcaacgggactttccaaaa
⑧	**screen_2F5vL_rev**	GATGGTCAGGGTGAACTCG

	Program 1	Program 2	Program 3	Program 4
step 1 (initial denaturation)	95°C for 180 sec	95°C for 180 sec	95°C for 180 sec	95°C for 300 sec
step 2 (denaturation)	98°C for 20 sec	98°C for 20 sec	98°C for 20 sec	95°C for 30 sec
step 3 (annealing, **Note 7**)	65°C for 25 sec	65°C for 15 sec	65°C for 15 sec	60°C for 30 sec
step 4 (extension)	72°C for 15 sec	72°C for 45 sec	72°C for 45 sec	68°C for 60 sec
repeat (step 2 to 4)	20 cycles	10 cycles	30 cycles	30 cycles
Step 5 (final extension)	72°C for 300 sec	72°C for 300 sec	72°C for 300 sec	68°C for 300 sec
primers	fragment 1: 1 + 2 fragmetn 2: 3 + 4 fragment 3: 4 + 5	none	1 + 6	7 + 8
amplicon:	fragment 1: 512 bp fragment 2: 362 bp fragment 3: 710 bp	1,543 bp	1,543 bp	862 bp

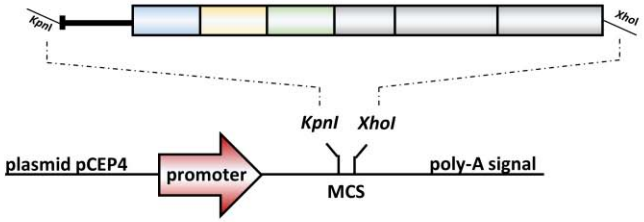
A.


ttGG.....ctg
KpnI
gag...ctGC..AGA
gag.....cac
MDW.....VHS
leader
RITL.....TISS
2F5vH
(GGGGS)3
linker
ALQ....DVR
2F5vL
EPKSS...PCP
hinge
APELL.....SKAK
CH2
GQPR.....PGK*
CH3
XhoI
cag....TCat
TGG...ccg.....taga
CTG...ctc...tcg

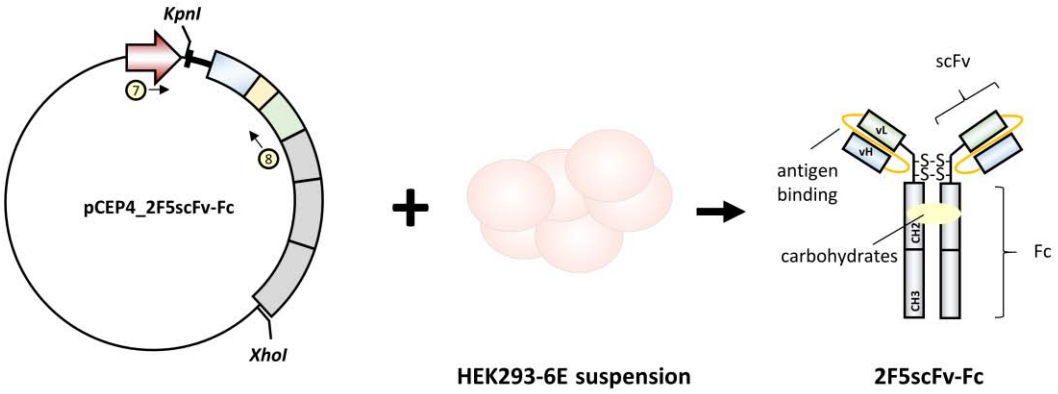
B.


fragment 1
fragment 3
KpnI
XhoI
overlap 1
overlap 2
fragment 2

C.


KpnI
XhoI
plasmid pCEP4
promoter
poly-A signal
MCS

D.


KpnI
pCEP4_2F5scFv-Fc
XhoI
HEK293-6E suspension
scFv
vL
vH
antigen binding
carbohydrates
CH2
CH3
Fc
2F5scFv-Fc

ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Proteomic Differences in Recombinant CHO Cells Producing Two Similar Antibody Fragments

Wolfgang Sommeregger,[1,2,3] Patrick Mayrhofer,[1] Willibald Steinfellner,[1] David Reinhart,[1] Michael Henry,[4] Martin Clynes,[4] Paula Meleady,[4] Renate Kunert[1]

[1]Vienna Institute of BioTechnology (VIBT), University of Natural Resources and Life Sciences, Muthgasse 18, B, 5th Floor, 1190 Vienna, Austria; telephone: +43 1 47654-6595; fax: +43 1 47654-6675; e-mail: renate.kunert@boku.ac.at

[2]Polymun Scientific GmbH, Klosterneuburg, Austria

[3]Bilfinger Industrietechnik Salzburg GmbH, Salzburg, Austria

[4]National Institute for Cellular Biotechnology (NICB), Dublin City University, Dublin 9, Ireland; telephone: +353 700-5910; fax: +353 700-5484; e-mail: paula.meleady@dcu.ie

ABSTRACT: Chinese hamster ovary (CHO) cells are the most commonly used mammalian hosts for the production of biopharmaceuticals. To overcome unfavorable features of CHO cells, a lot of effort is put into cell engineering to improve phenotype. "Omics" studies investigating elevated growth rate and specific productivities as well as extracellular stimulus have already revealed many interesting engineering targets. However, it remains largely unknown how physicochemical properties of the recombinant product itself influence the host cell. In this study, we used quantitative label-free LC-MS proteomic analyses to investigate product-specific proteome differences in CHO cells producing two similar antibody fragments. We established recombinant CHO cells producing the two antibodies, 3D6 and 2F5, both as single-chain Fv-Fc homodimeric antibody fragments (scFv-Fc). We applied three different vector strategies for transgene delivery (i.e., plasmid, bacterial artificial chromosome, recombinase-mediated cassette exchange), selected two best performing clones from transgene variants and transgene delivery methods and investigated three consecutively passaged cell samples by label-free proteomic analysis. LC-MS-MS profiles were compared in several sample combinations to gain insights into different aspects of proteomic changes caused by overexpression of two different heterologous proteins. This study suggests that not only the levels of specific product secretion but the product itself has a large impact on the proteome of the cell.

Biotechnol. Bioeng. 2016;9999: 1–11.

Paula Meleady and Renate Kunert contributed equally to this work.
Correspondence to: R. Kunert and P. Meleady
Contract grant sponsor: Polymun Scientific GmbH, Klosterneuburg, Austria
Contract grant sponsor: University of Natural Resources and Life Sciences, Vienna
Contract grant sponsor: Austrian Science Fund, FWF Project
Contract grant number: W1224
Received 5 September 2015; Revision received 25 November 2015; Accepted 14 February 2016
Accepted manuscript online xx Month 2016;
Article first published online in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/bit.25957

KEYWORDS: Chinese hamster ovary cells; bottleneck; endoplasmic reticulum; secretion; specific productivity

Introduction

Chinese hamster ovary (CHO) cells are the most frequently used mammalian host for the production of biopharmaceuticals (Walsh, 2014). For a better understanding of the biological background and to improve the production system "omics"-based research has given insights into cellular changes triggered by overexpression, higher growth rates or extracellular stimulus like hypothermic cultivation, butyrate treatment, or hyperosmotic conditions (Datta et al., 2013; Kildegaard et al., 2013; Kim et al., 2012). We hypothesize that the proteome of production clones is additionally and considerably influenced by the nascent and mature recombinant protein, characterized by distinct biochemical features (e.g., structure, stability, surface charge distribution) and challenging post-translational modifications as well as secretion. To create genetically more defined and comparable recombinant cell lines, the CHO DUKXB11-RMCE host cell line capable of targeted gene integration by recombinase-mediated cassette exchange (RMCE) was described (Mayrhofer et al., 2014) and used for recombinant protein expression of scFv-Fc versions of the two anti-HIV-1 antibodies 3D6 and 2F5 (Kunert et al., 1998) under isogenic conditions. Although, the same RMCE host cell line was used and transgene copy numbers as well as levels of transcript were identical, a twofold difference in specific productivity between 2F5- and 3D6-scFv-Fc producers was observed. Following this observation, we applied two additional strategies of transgene delivery and established 2F5- and 3D6-scFv-Fc producers, once with a common plasmid strategy and secondly with the *Rosa26* bacterial artificial chromosome (BAC) strategy (Blaas et al., 2009, 2012; Kunert and Casanova, 2013; Mader et al., 2013; Zboray et al., 2015), applying the same CHO DUKXB11 host

cell line that was used for the generation of the RMCE cell line. Independent from the transgene delivery system, we regularly identified clones with higher specific productivities for 3D6-scFv-Fc compared to 2F5-scFv-Fc producing cell lines. These findings support the assumption that the protein's physicochemical properties in particular the unique variable heavy domain (V_H)/variable light domain (V_L) sequences of the antibody fragment, trigger different intracellular responses within the recombinant cells.

In this study, we aimed to identify differences on a proteomic level in recombinant CHO cells producing the scFv-Fc version of either 2F5 or 3D6 rather than comparing the recombinant cell lines to the respective host cell line. In order to minimize bias caused by clonal selection, screening, and the genomic variability of CHO cells, we included the two best producing clones for each scFv-Fc variant of three transgene delivery methods (RMCE, plasmid, and BAC). Understandably, not only the transgene but also the different vector systems generated cell lines with varying specific production rates for both products. Samples from the 12 clones (two best producers, two products, three transgene delivery methods) were collected from three consecutive passages and unfractionated whole-cell-lysates were analyzed using quantitative label-free LC-MS proteomic analyses and bioinformatics tools. Evaluation of LC-MS data was performed by two types of comparisons. The "transgene comparison" was applied to identify proteomic changes caused by the two products while the "transgene delivery comparison" aimed to elucidate proteomic changes resulting from elevated production rates. In summary, we could show that not only the specific product secretion rate but also the recombinantly produced protein itself considerably influences the cell's physiology. These findings suggest that host engineering strategies might only work for a subset of recombinant proteins.

Materials and Methods

Antibody Fragments

3D6-Fv (PDB: 1DFB) and 2F5-Fv (PDB: 2F5B), respectively, were combined via a $(GGGGS)_3$ linker to the human immunoglobulin G (IgG)-1 Fc region (GenBank, CAA49866). 2F5-scFv-Fc and 3D6-scFv-Fc sequence alignment (79.3% identity) is shown in Supplemental Figure S1. Nucleotide sequences were codon optimized and synthesized (Geneart, Regensburg, Germany).

Cell Culture

3D6- and 2F5-scFv-Fc expressing cell lines were used for the generation of samples for proteomic and thermal stability analyses. Both antibody fragments were produced using three different vector systems each transfected in the same protein-free adapted CHO-DUKX-B11 (ATCC CRL-9096) (Urlaub and Chasin, 1980) host cell lines. The applied transgene delivery systems were indicated by (i) common plasmid vectors; (ii) *Rosa26* BACs; and (iii) RMCE, for the generation of the recombinant cell lines. The two best performing clones for each strategy and product were used in this study ($n = 12$). The cell line establishment of the 12 clones has been previously described (Mader et al., 2013; Mayrhofer et al., 2014). All clones were cultivated in suspension in 125-mL spinner flasks (Techne, Thermo Fisher Scientific, Waltham, MA) at 37°C, 7% CO_2, and 50 rpm. ProCHO5 (Lonza, Basel, Switzerland) was used for all clones supplemented with 4 mM L-glutamine and hypoxanthine/thymidine (HT) without any selection pressure. Cells were passaged into fresh media every 3–4 days.

Cell Counting and Viability Determination

The cell concentration was calculated by counting the nuclei of cells lyzed in 0.1 M citric acid and 2% (w/w) Triton X 100 with the particle counter Z2 (Beckman Coulter, Brea, CA). Viability was determined by trypan blue exclusion method using a Neubauer cell counting chamber. Growth rate μ (d^{-1}) was calculated according to Equation (1) where X (cells) represents the total number of viable cells and t (d) the cultivation time in days.

$$\mu = \ln\left(\frac{X_1}{X_0}\right) \times \frac{1}{(t_1 - t_0)} \quad (1)$$

Product Concentration and Specific Productivity

The product concentration in the cell supernatants was determined by sandwich enzyme-linked immunosorbent assay (ELISA) as previously described (Mader et al., 2013; Mayrhofer et al., 2014). Briefly, 96-well Maxisorp plates (Nunc, Thermo Fisher Scientific) were coated with polyclonal goat anti-human IgG (γ-chain specific) (I3382, Sigma–Aldrich, St. Louis, MO). Goat anti-human IgG horseradish peroxidase conjugate (γ-chain specific) (62-8420, Life Technologies, Carlsbad, CA) was used as the detection antibody. Staining was initiated with orthophenylediamine and H_2O_2 and the resulting color reaction was measured at 492 nm and a reference wavelength of 620 nm on a micro-plate reader (Tecan, Männedorf, Switzerland). Specific productivity qP ($pg \times cell^{-1} \times day^{-1}$) was calculated according to Equation (2) where P (pg) represents the product amount.

$$qP = \mu \times \frac{(P_1 - P_0)}{(X_1 - X_0)} \quad (2)$$

Differential Scanning Calorimetry

2F5 and 3D6 antibody fragments were purified by Protein A chromatography from pooled culture supernatants from all clones ($n = 12$) using a 1 mL pre-packed MabSelect SuRe resin on an ÄKTA Purifier (both GE Healthcare, Little Chalfont, UK) according to the manufacturer's instructions. Eluates were desalted and buffer-exchanged to 40 mM phosphate, 150 mM NaCl, pH 6.0 using PD MidiTrap G-25 units (GE Healthcare) according to the manufacturer's instructions. Finally, sample quantification was performed on a NanoDrop spectrophotometer (Thermo Fisher Scientific) by applying the scFv-Fc respective theoretical extinction coefficients. Thermal denaturation of the scFv-Fc samples was monitored using automated differential scanning calorimetry (DSC). All DSC measurements were performed in duplicates on a VP-DSC MicroCal LLC equipment (GE Healthcare). Protein

solutions were sampled from 96-well plates using the robotic attachment. The protein concentration of all samples was 2–3 μM. The temperature profile was recorded between 20 and 100°C with a scan rate of 1°C/min. The results were evaluated and fitted with the Origin 7.0 software (OriginLab, Northampton, MA). The unfolding states of the antibodies were fit using the non-two state unfolding model within the software.

Sample Preparation for Label Free LC-MS Analysis

Sampling of each clone was performed at three consecutive passages (biological replicates). Ten million cells were harvested before each splitting by centrifugation and washed twice with cold PBS. Washed cell pellets were immediately snap frozen in liquid nitrogen and stored at −80°C prior to further preparation. Cells were thawed on ice and lyzed in lysis buffer (7 M urea, 2 M thiourea, 4% CHAPS, 30 mM Tris, pH 8.5) for 1 h at room temperature. Whole cell lysates were cleaned up using ReadyPrep 2D Cleanup Kit (Bio-Rad Laboratories, Hercules, CA) and purified proteins resuspended in label free buffer (6 M urea, 2 M thiourea, 10 mM Tris, pH 8.0). Protein concentrations were determined in triplicate using a Bradford 1× Dye Reagent and a Quick start BSA standard (both Bio-Rad). A total of 14 μg of protein were re-suspended in 46.6 μL of 50 mM ammonium bicarbonate. Reduction was performed by adding 0.5 μL of 0.5 M DTT at 56°C for 20 min. Afterwards, samples were alkylated by adding 1.4 μL of 0.55 M iodoacetamide and then incubated for 35 min at room temperature. Digestion was performed by adding 1 μg MS grade Trypsin Gold (Promega, Fitchburg, WI) and 0.01% ProteaseMax (Promega) overnight at 37°C. Trifluoroacetic acid (TFA) was added to a final concentration of 0.5% to inactivate the trypsin. Samples were frozen at −20°C prior to analysis by LC-MS/MS.

Quantitative Label Free LC-MS Analysis

Nano LC–MS-MS was carried out essentially as previously described (Linge et al., 2014). Briefly, the analysis was performed using an Ultimate 3000 RSLCnano system (Dionex, Thermo Fisher Scientific) coupled to a hybrid linear ion trap/Orbitrap mass spectrometer (LTQ Orbitrap XL; Thermo Fisher Scientific). Samples were thawed and sonicated to ensure an even suspension and 1 μg of digested proteins were loaded onto a C18 trap column (C18 PepMap, 300 μm i.d. × 5 mm, 5 μm particle size, 100 μm pore size; Dionex) and desalted for 5 min. The trap column was then switched online with the analytical column (PepMap C18, 75 μm i.d. × 500 mm, 3 μm particle, and 100 μm pore size; Dionex), and peptides were eluted in a 300 min gradient. Data were acquired with Xcalibur software, version 2.0.7 (Thermo Fisher Scientific). The mass spectrometer was operated in data-dependent mode and externally calibrated.

Data Analysis

Differential proteomic analysis using label-free LC-MS/MS was carried out using Progenesis QI for proteomics version 1.0 (NonLinear Dynamics Limited, Newcastle upon Tyne, UK), essentially as recommended by the manufacturer and as previously described (Clarke et al., 2012). The raw data obtained from each of the LC-MS/MS runs per sample was processed using Progenesis QI for proteomics software. Several criteria were used to filter the data before exporting the Progenesis output files to Proteome Discoverer 1.4 (Thermo Fisher Scientific) for protein identification: peptide features with adjusted ANOVA *P*-value ≤ 0.05 between experimental groups, mass features with charge states from +1 to +3, and the number of isotopes was set to 3 or less. All MS/MS spectra were exported from Progenesis software as an mgf file and searched against CHO-specific protein sequence databases, using a combination of the translated NCBI genomic database (Baycin-Hizal et al., 2012) containing 24,927 entries (fasta file downloaded January 2014) and the expressed cDNA database (BB-CHO) (Meleady et al., 2012) containing 14,627 entries, through Proteome Discoverer 1.4 and the search algorithms Mascot and SequestHT. The search parameters used for all searches on Proteome Discoverer 1.4 were as follows: precursor mass tolerance set to 20 ppm, fragment mass tolerance set to 0.6 Da; up to two missed cleavages were allowed, carbamidomethylation set as a fixed modification, and methionine oxidation set as a variable modification. For re-importation back into Progenesis LC-MS software for further analysis only peptide identifications with MASCOT ion peptide scores above 40 or peptides with XCorr scores >1.9 for singly charged ions, >2.2 for doubly charged ions and >3.75 for triply charged ions from Sequest analysis were accepted. To filter out target peptide spectrum matches (target-PSMs) over the decoy-PSMs, a fixed false discovery rate (FDR) of 1% was set at the peptide level. Proteins were only then considered as being differentially expressed between experimental groups if they had an adjusted ANOVA *P*-value ≤ 0.05 and were identified by ≥ 2 peptides.

Experimental Design

We used two different strategies to explore differences within the proteome of 2F5- and 3D6-scFv-Fc producers.

Transgene Comparison

In a first approach, all 2F5-scFv-Fc samples ($n = 18$) were compared to all 3D6-scFv-Fc samples ($n = 18$) and differentially expressed proteins were evaluated (Fig. 1B). Using this transgene comparison, we could identify proteome changes which are related to the expressed product.

Transgene Delivery Comparison

In the second data evaluation approach, we performed 2F5- and 3D6-scFv-Fc transgene delivery comparisons of RMCE versus plasmid versus BAC (each $n = 6$, resulting from two clones analyzed in three consecutive passages) (Fig. 1C). Proteins that were statistically significant differentially expressed between RMCE, plasmid, and BAC clones were obtained and only expression profiles that positively or negatively correlated to the specific productivity of the distinct products caused by the use of different transgene delivery strategies were taken into account and used for further comparison of 2F5-scFv-Fc and 3D6-scFv-Fc differential expression profiles (Fig. 1D). This strategy allowed us to identify proteins that

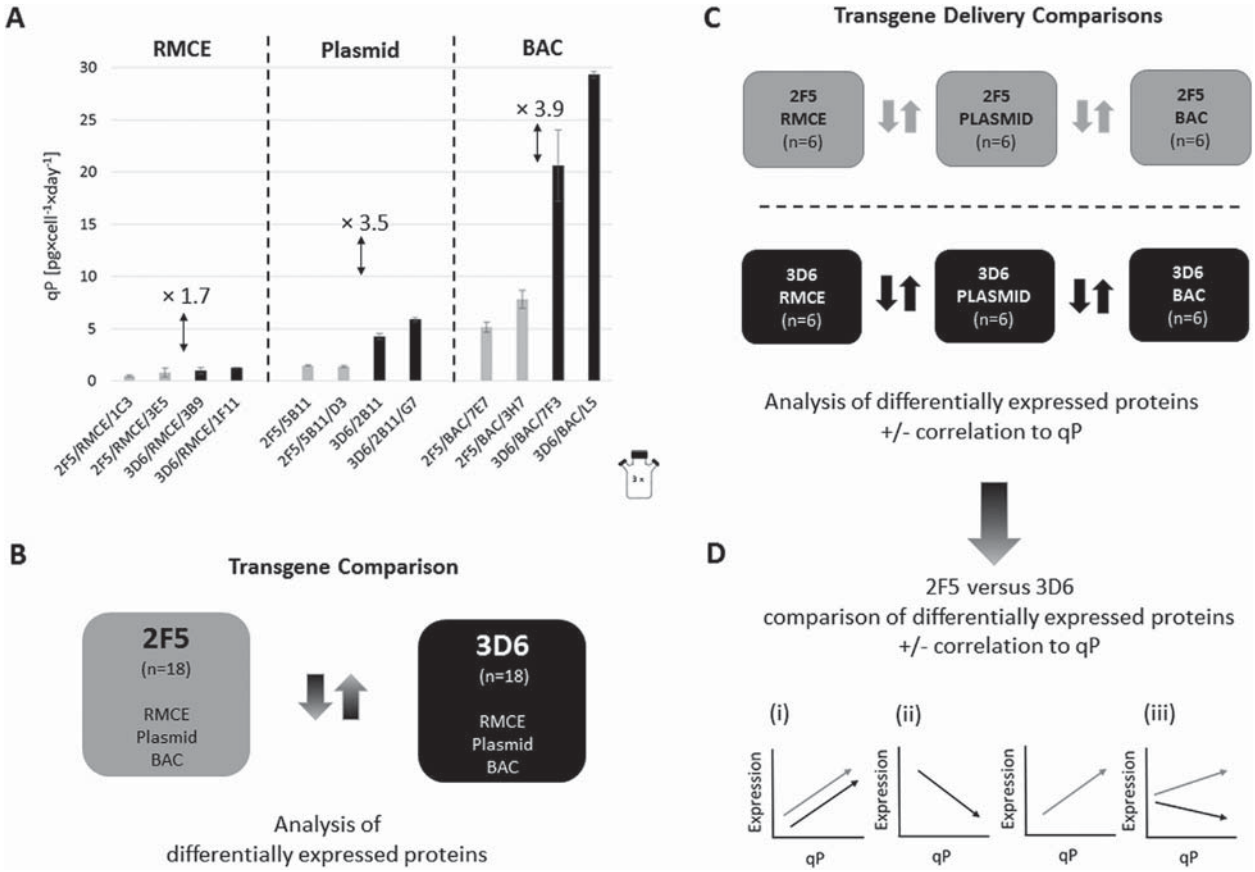


Figure 1. Schematic work-flow representation: (**A**) Cell specific productivities were identified and samples for proteomic analysis were taken. (**B**) LC-MS-MS data of all 2F5-scFv-Fc producers were compared with 3D6-scFv-Fc data to explore global transgene-specific proteome changes. (**C**) 2F5- and 3D6-scFv-Fc transgene delivery comparisons were performed and differentially expressed proteins with positive or negative correlation to qP were evaluated. (**D**) Subsequently, differentially expressed proteins that showed a ± correlation to qP in the in-group comparisons of 2F5- and 3D6-scFv-Fc producers were compared to each other.

are up/down regulated with higher specific productivities in both groups. Furthermore, we could identify proteins that were differentially expressed only in one of the groups as well as protein expression profiles that were opposing between the two groups.

Pathway Analysis

Identified differential proteins were manually assigned to mouse official gene IDs, according to the protein names, for further gene list analysis using GeneCodis3 (http://genecodis.cnb.csic.es/) (Carmona-Saez et al., 2007; Nogales-Cadenas et al., 2009; Tabas-Madrid et al., 2012).

Western Blot

Cellular protein lysates of relevant cell clones were pooled to confirm the experimental data of the proteomics comparisons. A total of 15 μg of total protein was mixed with 2× Laemmli sample buffer, heated to 95°C for 5 min and cooled prior to loading on 4–12% NuPAGE Bis-tris pre-cast gels (Life Technologies). Resolved proteins were electrophoretically blotted to Amersham Hybond P membranes (GE Healthcare, Buckinghamshire, UK). Membranes were blocked in 5% milk powder dissolved in 0.1% Tween in Tris-Buffered Saline (TBST), washed with TBST, and incubated with primary antibody (rabbit polyclonal antibody (pAb) to heat shock protein 60 (Hspd1), ab46798; rabbit pAb to catalase (Cat), ab1877; rabbit pAb to SERPINB1, ab47731, Abcam, Cambridge, UK) diluted in TBST including 5% milk powder. Subsequently, membranes were washed and incubated with secondary HRP-conjugated antibodies (Goat anti-rabbit pAb, P0448 Dako, Agilent Technologies, Santa Clara, CA) diluted in TBST including 5% milk powder. Western blots were developed using WesternBright ECL Spray (Advansta, Menlo Park, CA). Subsequently, the membranes were washed in TBST and re-probed with loading control antibodies (mouse mAb to ß-actin (Actb), ab8226, Abcam or mouse mAb to α-tubulin (Tuba1a), Sigma–Aldrich) and secondary antibody (goat anti-mouse pAb, P0447, Dako) as described above.

Results

Cell Culture

For each product (2F5- and 3D6-scFv-Fc) and each strategy of transgene delivery (RMCE, plasmid, and BAC), the two best

performing cell lines identified by specific productivity and growth rate were cultivated ($n = 12$). The recombinant scFv-Fc producers were routinely cultivated in spinner flasks and samples for proteomic analysis were taken before passaging during three consecutive passages (biological replicates) at high viabilities (>90%). Cell line specific parameters were determined and are summarized in Table I and depicted in Figure 1A. For both products, the clones generated with RMCE had a low specific productivity, clones generated using plasmids displayed a medium qP and BAC clones displayed a high specific production rate. Overall, the specific productivities of the 3D6-scFv-Fc producers were always higher compared to the 2F5-scFv-Fc clones. The mean difference between 2F5- and 3D6-scFv-Fc clones was 1.7× (RMCE), 3.5× (plasmid), and 3.9× (BAC). Growth rate correlated slightly negatively with qP and was highest for the RMCE-generated 2F5- and 3D6-scFv-Fc producers with only minor differences between them.

2F5-and 3D6-scFv-Fc Thermal Stability

A typical human IgG1 DSC curve contains three unfolding transitions of the constant heavy domains C_H2, C_H3 and fragment, antigen binding (Fab) domain. The herein tested single chain antibody fragments revealed three unfolding transitions that superimposed quite well (Fig. 2). However, some differences of the respective melting temperatures could be identified. Unfolding of the C_H3 domain was almost identical for both 3D6- and 2F5-scFv-Fc at ~83°C. The unfolding transition of the C_H2 domain of 2F5-scFv-Fc (65.0 ± 0.2°C) occurred slightly earlier than that of 3D6scFv-Fc (65.7 ± 0.1°C). The Fv's showed the lowest resistance to thermal unfolding and occurred 1.4°C earlier for 2F5-scFv-Fc (52.1 ± 0.1) than for 3D6-scFv-Fc (53.5 ± 0.2).

Proteomics

We performed label-free LC-MS proteomic analyses of recombinant CHO cell lines producing two different types of scFv-Fc fragments to explore changes caused by the recombinant protein (Fig. 1B) as well as differences evoked by elevated specific production rates (Fig. 1C and D) caused by the use of different transgene delivery strategies.

Transgene Comparison

To gain insight into proteomic changes caused by the type of recombinant scFv-Fc being produced, we compared all 2F5- ($n = 18$) with all 3D6-scFv-Fc ($n = 18$) samples. Principal component analysis of all statistically significant differential peptides ($P \leq 0.05$) shows a clear separation between 2F5- and 3D6-scFv-Fc samples (Supplemental Fig. S2). We could identify 60 proteins being differentially expressed between the two groups ($P \leq 0.05$; ≥1.2-fold change; ≥2 peptides used for quantification) (Supplemental Table SI). A selection of the identified differential expressed proteins is listed in Table II. Peptides were also searched against a human database. However, we could not identify the recombinant product itself to be statistically significantly different between the heterogeneous groups of 2F5- and 3D6-scFv-Fc producers, most probably due to the quantitative variability of intracellular product in the heterogeneous sample group.

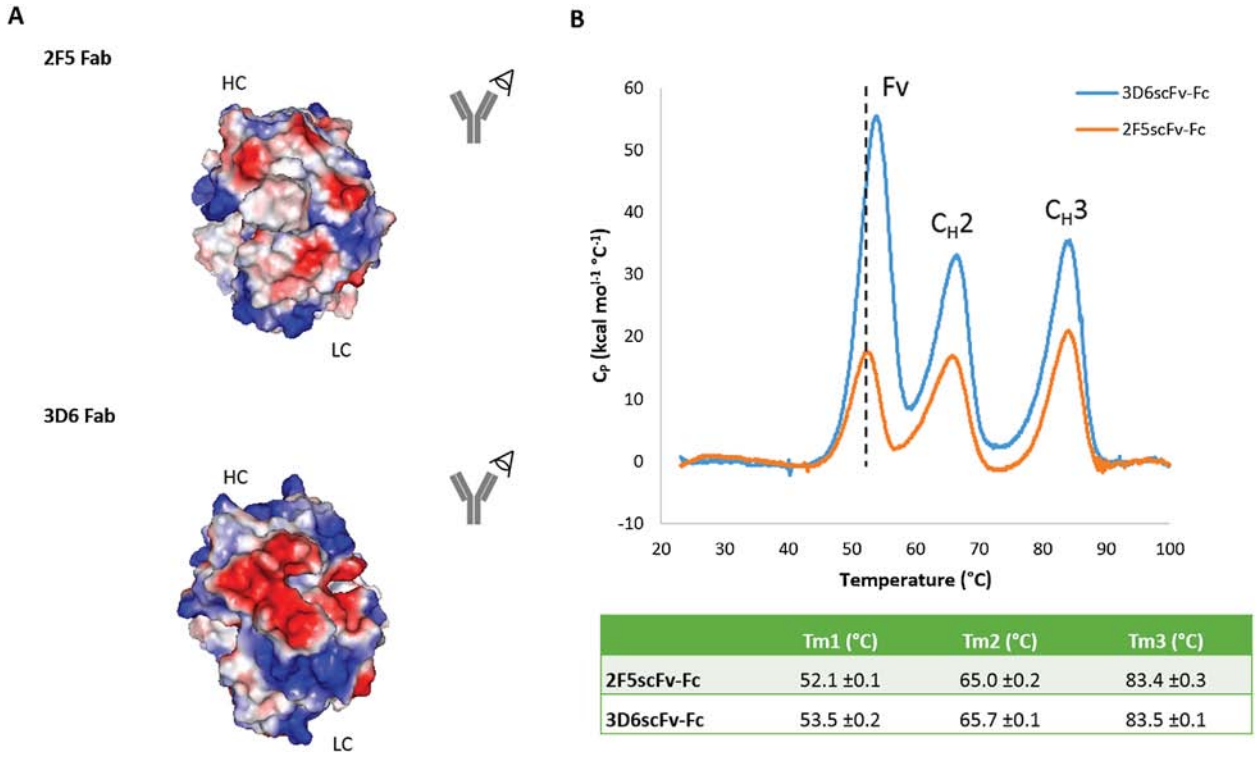
Subsequently, we performed singular enrichment analysis using the official gene IDs (mouse equivalent) of the 60 identified proteins using the web-based tool GeneCodis3. Singular enrichment analysis of GOSlim process revealed three biological processes to be statistically significant enriched (hypergeometric *P*-value ≤0.05). These identified cellular processes were cell proliferation, protein folding, and extracellular matrix organization (Supplemental Table SVIII). Two of the identified differential proteins, the 60 kDa heat shock protein (Hspd1) and Catalase (Cat) were validated by western blots (Fig. 3A). The western blots of pooled lysates from all 2F5- ($n = 18$) and 3D6-scFv-Fc ($n = 18$) samples confirmed the overall higher abundance of Hspd1 in 2F5-scFv-Fc samples as well as higher abundance of Catalase in 3D6-scFv-Fc samples compared to ß-actin (Actb) which was used as loading control.

Transgene Delivery Comparison

To explore proteomic changes caused by increasing product secretion rates resulting from different transgene delivery strategies

Table I. Cell line performances in terms of specific growth rates and specific productivities. Comparison of 3D6 scFv-Fc clones and 2F5scFv-Fc clones generated with different strategies for transgene delivery.

Clone	μ (d^{-1})	Mean μ	qP ($pg \times cell^{-1} \times d^{-1}$)	Mean qP	Fold diff. in qP
2F5/RMCE/1C3	0.47	0.46	0.48	0.66	1.70
2F5/RMCE/3E5	0.46		0.84		
3D6/RMCE/3B9	0.38	0.40	1.00	1.12	
3D6/RMCE/1F11	0.43		1.24		
2F5/PLASMID/5B11	0.40	0.39	1.49	1.45	3.50
2F5/PLASMID/5B11/D3	0.38		1.42		
3D6/PLASMID/2B11	0.35	0.36	4.25	5.08	
3D6/PLASMID/2B11/G7	0.38		5.91		
2F5/BAC/7E7	0.30	0.33	5.17	6.49	3.85
2F5/BAC/3H7	0.36		7.81		
3D6/BAC/7F3	0.29	0.35	20.62	24.97	
3D6/BAC/L5	0.41		29.32		



	Tm1 (°C)	Tm2 (°C)	Tm3 (°C)
2F5scFv-Fc	52.1 ±0.1	65.0 ±0.2	83.4 ±0.3
3D6scFv-Fc	53.5 ±0.2	65.7 ±0.1	83.5 ±0.1

Figure 2. (**A**) Computationally modeled Fab fragment of the original 2F5 (PDB: 2F5B) and 3D6 (PDB: 1DFB) IgG using the PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC. The surface model is colored by the underlying residue charge: red is negative, blue is positive, and white is neutral. (**B**) DSC thermogram of 2F5- and 3D6-scFv-Fc. The Fv, CH2, and CH3 domains and their respective unfolding transitions of are indicated.

for 2F5 and 3D6-scFv-Fc producing cell lines, we performed transgene delivery comparisons of RMCE ($n = 6$) versus plasmid ($n = 6$) versus BAC ($n = 6$) samples and evaluated the statistically significant proteins differentially expressed between RMCE, plasmid, and BAC clones. The outcome was firstly evaluated separately and afterwards combined. We identified 109 (for 2F5-scFv-Fc producers, Supplemental Table SII) and 212 (for 3D6-scFv-Fc producers, Supplemental Table SIII) differentially expressed proteins that displayed either a positive or negative correlation to qP or μ (highest relative abundance in BAC and lowest relative abundance in RMCE clones or vice versa; $P \leq 0.05$; $\geq 1.5\times$ fold change; ≥ 2 peptides used for quantification).

Subsequently, using this approach we were able to cluster the results into three groups of differentially expressed proteins:

(1) Expression pattern correlating to qP in both antibody groups (2F5 and 3D6). This first group of differentially expressed proteins was identified in 2F5- as well as in 3D6-scFv-Fc clones describing positive/negative correlation according to the transfection system—lowest relative abundance in the group of RMCE and highest relative abundance in the group of BAC generated clones or vice versa (reflects a correlation to qP or μ). We identified 61 proteins that show the same expression pattern in both group comparisons. Three proteins were poorly annotated hypothetical proteins and could not be assigned. Singular enrichment analysis of GOSlim process using the gene IDs of the 58 remaining proteins (Supplemental Table SIV, Fig. 4) revealed four processes to be statistically significant enriched (hypergeometric *P*-value ≤ 0.05). These cellular processes were cell cycle, vesicle-mediated transport, cell death, and transport in general (Supplemental Table SIX). Selected differentially expressed proteins showing the same expression pattern in both transgene delivery comparisons are listed in Table III. As internal control when searching the differential peptides against the human database, we detected the constant region of immunoglobulin (Ig) γ-1 chain in both transgene delivery comparisons to be differential (Supplemental Fig. S3). 2F5 Ig γ-1 chain was identified by 11 peptides with a 4.6-fold change from RMCE to BAC. 3D6 Ig γ-1 chain was identified by 7 peptides with a 10.1-fold change from RMCE to BAC. To validate the proteomics results, we confirmed the differential expression of the protein leukocyte elastase inhibitor A (Serpinb1a) by Western blot (Fig. 3B). The increase of Serpinb1a levels from low in RMCE clones to high in BAC clones was confirmed for 2F5-scFv-Fc as well as in 3D6-scFv-Fc producers compared to α-tubulin (Tub1a) which was used as loading control.

(2) Differentially expressed proteins identified in only one antibody group. The second group summarizes proteins that were only identified to be differential for one transgene delivery comparison (either 2F5- or 3D6-scFv-Fc). We identified 51 (2F5-scFv-Fc) and 154 (3D6-scFv-Fc) proteins that were only detected to be positively or negatively correlating to the specific productivities or growth rates in one of the two transgene delivery comparisons (RMCE vs. plasmid vs. BAC). However, many of these identified proteins were

Table II. Transgene comparison: table of selected proteins differential between 2F5 ($n=18$) and 3D6 scFv-Fc ($n=18$) samples ($P \leq 0.05$; $\geq 1.2\times$ fold change; ≥ 2 peptides used for quantification) and sorted by fold change.

Description	Gene ID	Fold change	Highest mean	Function
Galectin-3	Lgals3	1.75	2F5	Anti-apoptosis; mRNA processing; extracellular matrix organization
DNA replication licensing factor MCM5	Mcm5	1.58	2F5	DNA replication initiation
60 kDa Heat shock protein, mitochondrial[a]	Hspd1	1.50	2F5	Protein folding; apoptosis
N-acetyltransferase 10	Nat10	1.49	2F5	Histone acetylation; metabolic process
Nucleolin	Ncl	1.43	2F5	Chromatin decondensation
10 kDa Heat shock protein, mitochondrial	Hspe1	1.43	2F5	Protein folding
Importin-5	Ipo5	1.27	2F5	Protein import into nucleus
Metastasis-associated protein MTA2	Mta2	1.22	2F5	Histone deacetylation, DNA packaging; apoptosis
Glutathione S-transferase P 2	Gstp2	2.74	3D6	Metabolic process
Glutathione S-transferase P 1	Gstp1	2.19	3D6	Apoptosis; metabolic process; regulation of stress-activated MAPK cascade
Peroxiredoxin-1	Prdx1	1.76	3D6	Response to oxidative stress; cell proliferation; regulation of stress-activated MAPK cascade
Golgi-associated plant pathogenesis-related protein 1	Glipr2	1.64	3D6	Regulation of ERK1 and ERK2 cascade
Protein disulfide-isomerase A3	Pdia3	1.58	3D6	Folding; apoptosis
Catalase[a]	Cat	1.52	3D6	Response to oxidative stress
Calreticulin	Calr	1.47	3D6	Protein folding; proliferation
SEC23-interacting protein	Sec23Ip	1.38	3D6	Intracellular protein transport; golgi organization
Protein disulfide-isomerase A4	Pdia4	1.33	3D6	Protein secretion; protein folding
Thioredoxin reductase 1, cytoplasmic	Txnrd1	1.29	3D6	Response to oxidative stress; proliferation
Glutathione S-transferase Mu 6	Gstm6	1.28	3D6	Metabolic process
Eukaryotic translation initiation factor 5A-1	Eif5a	1.28	3D6	Apoptosis; proliferation; translation
Eukaryotic initiation factor 4A-I	Eif4a1	1.26	3D6	Translation
Glutathione S-transferase omega-1 isoform 1	Gsto1	1.26	3D6	Metabolic process
DnaJ homolog subfamily C member 7	Dnajc7	1.23	3D6	Protein folding
Protein disulfide-isomerase	P4Hb	1.23	3D6	Protein folding
Endoplasmin	Hsp90b1	1.23	3D6	Protein folding; apoptosis
GrpE protein homolog 1, mitochondrial	Grpel1	1.21	3D6	Protein folding

[a]Results validated via Western blotting (Fig. 3).

also identified in the respective other antibody transgene delivery comparison, but they were not correlating with the specific productivity or the growth rate (highest/lowest abundance in plasmid clones). These proteins are assumed to be not strictly regulated and therefore not further discussed. After exclusion of such proteins, we finally identified 9 and 32 proteins that were only detected to be differential in the transgene delivery comparison of the 2F5-scFv-Fc or 3D6-scFv-Fc, respectively, and not detected at all for the respective other group (Supplemental Tables SV and SVI).

(3) Differentially expressed proteins showing opposite expression patterns in the two antibody groups. We identified two proteins that showed exactly the opposite expression pattern in the two transgene delivery comparisons, positive correlation to qP in one transgene delivery comparison, and negative correlation in the other

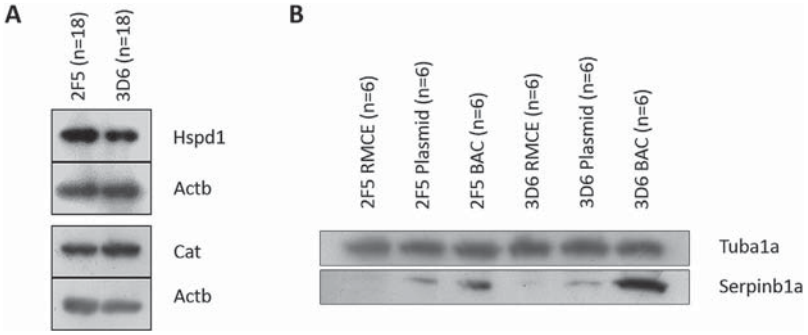


Figure 3. Western blot validations of proteomics results of the two differential (2F5s vs. 3D6s) proteins 60 kDa heat schock protein (Hspd1) and catalase showing increased expression of Hspd1 (fold difference: proteomic 1.50× and densitometric 1.24×) and decreased expression of catalase (fold difference: proteomic 1.52× and densitometric 1.36× in the group of 2F5s ($n=18$) compared to 3D6s ($n=18$) (**A**) and of the protein leukocyte elastase inhibitor A (Serpinb1a) showing increasing expression of Serpinb1a from RMCE to plasmid to BAC for 2F5-scFv-Fc and 3D6-scFv-Fc producers (**B**), where proteomics results showed a 5.8 and 18-fold increase of Serpinb1a from RMCE to BAC for 2F5- and 3D6-scFvFc producers, respectively; 15 μg of total protein from pooled lysates were loaded each and ß-actin (Actb) or α-tubulin (Tuba1a) was used as internal loading control.

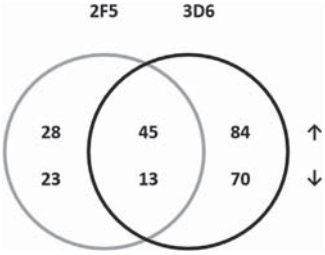


Figure 4. Venn diagram: Comparison of differentially expressed proteins correlating to qP identified in the 2F5- and 3D6-scFv-Fc transgene delivery comparisons; ↑: proteins positively correlating with qP, ↓: proteins negatively correlating with qP.

transgene delivery comparison (≥1.5-fold change in at least one group). The identified proteins were Cullin-associated NEDD8-dissociated protein 1 (Cand1), a regulator of SCF ubiquitin ligases and Protein SON (Son), a mRNA splicing cofactor (Supplemental Table SVII). Cand1 and SON both were positively correlating with qP in the group of 3D6-scFv-Fc producers whereas negatively correlating with qP in the group of 2F5-scFv-Fc producers.

Discussion

Cell Culture

Three different strategies (RMCE, plasmid, and BAC) for transgene delivery were used to generate recombinant 2F5- and 3D6-scFv-Fc producers (Mader et al., 2013; Mayrhofer et al., 2014). The RMCE clones showed the lowest specific productivity and highest growth rate. Clones generated by *Rosa 26* BACs as transgene vehicles had the highest specific production rate and lowest growth rate for both products and clones generated by random integration of plasmid vectors were found in between (Fig. 1A). The RMCE clones have only one transgene copy integrated by recombination in a specific but undefined chromosomal locus. The plasmid clones, have presumably higher transgene copy numbers in unspecified loci. The clones that were generated using the *Rosa26* BAC technology have their transgenes in a well-defined transcriptional highly active environment provided by the BAC but no fixed transgene copy number. This enables a simple explanation for the distinct specific production rates based on the transgene delivery method. Additionally, we identified significant differences between product-specific expression rates of the two antibody fragments, which increased with increasing specific productivity. The mean qP difference between 2F5-scFv-Fc and 3D6-scFv-Fc was 1.7-fold for the low-producing RMCE clones, 3.5-fold for the medium producing plasmid clones, and 3.9-fold for the BAC clones. Transcript levels of all RMCE clones used in this study were investigated (Mayrhofer et al., 2014) and despite equal levels of transcripts, 2F5-scFv-Fc clones had a twofold lower mean specific productivity than 3D6-scFv-Fc clones over a period of 10 passages. This is very similar to the herein described difference of 1.7-fold during a shorter period of time. Mader et al. (2013) described a three to fourfold different specific productivity between 2F5- and 3D6-scFv-Fc plasmid- and BAC-generated producers, which is in accordance with the results presented in this study. Since both techniques are based on random integration, gene copy numbers and transcript levels can vary between 2F5- and 3D6-scFv-Fc plasmid- and BAC-derived producers. However, Mader et al. found no correlation between reduced gene copy numbers or levels of transcript and the reduced production rate of 2F5scFv-Fc compared to the 3D6 antibody fragment. Especially the different expression rates of isogenic RMCE clones lead to the assumption that the established CHO cells are more challenged with the production of the 2F5-scFv-Fc antibody fragment compared to the 3D6-scFv-Fc. We suppose that physicochemical properties of the nascent polypeptide might cause diminished translational activity or post-translational processing.

Thermal Stability

Already previous studies indicated that antibodies which are more stable against thermal unfolding are able to generate clones with improved expression rates (Buchanan et al., 2013; Garber and Demarest, 2007). In our study, DSC measurements revealed that the Fv domains were the least stable and unfolded first at 52.1 ± 0.1 (2F5scFv-Fc) and 53.5 ± 0.2 (3D6scFv-Fc). It remains elusive to which extent the thermal stability difference of 1.4°C contributed to the herein observed 1.7 to 3.9-fold different protein

Table III. Transgene delivery RMCE ($n=6$) versus plasmid ($n=6$) versus BAC ($n=6$); selected differential expressed proteins correlating with qP or μ identified for 2F5 scFv-Fc and/or 3D6 scFv-Fc producers (number of peptides used for quantitation ≥2; Anova *P*-value ≤0.05; fold change ≥1.5×).

		3D6			2F5			
Description	Gene ID	Fold change	Max	Min	Fold change	Max	Min	Function
Leukocyte elastase inhibitor A[a]	Serpinb1a	18.04	BAC	RMCE	5.77	BAC	RMCE	Regulation of proteolysis
Ig γ-1 chain C region	*IGHG1*	10.13	BAC	RMCE	4.62	BAC	RMCE	Recombinant product
Cathepsin B	Ctsb	5.14	BAC	RMCE	3.51	BAC	RMCE	Proteolysis
Galectin-1	Lgals1	4.81	BAC	RMCE	4.91	BAC	RMCE	Proliferation
Gelsolin	Gsn	2.25	BAC	RMCE	2.57	BAC	RMCE	Vesicle-mediated transport
Ras-related protein Rab-1A	Rab1	1.66	BAC	RMCE	2.01	BAC	RMCE	Vesicle-mediated transport
Vesicle-trafficking protein SEC22b	Sec22b	1.65	BAC	RMCE	1.69	BAC	RMCE	Vesicle-mediated transport
Heme oxygenase 1	Hmox1	4.81	RMCE	BAC	3.41	RMCE	BAC	ER-stress, proliferation
Proteasome-associated protein ECM29-like	Ecm29	3.71	RMCE	BAC	2.21	RMCE	BAC	Proteasome adaptor

[a]Result validated via Western blotting (Fig. 3).

expression but the proportion of our results are in good agreement with the published data.

Proteomics

In order to understand cellular proteomic differences of the clones producing these antibody fragments, we applied quantitative label-free LC-MS proteomic analyses and performed two types of comparisons. According to the population of the included clones, we defined different thresholds for statistical analyses but for both analyses we considered only proteins identified by two or more peptides with an adjusted *P*-value ≤ 0.05. For the "2F5 versus 3D6 transgene comparison," a fold-change cut-off of 1.2 was applied since we did not expect severe differences in this comparison of two phenotypically heterogeneous groups. For the "2F5 and 3D6 transgene delivery comparison," we increased the cut-off to 1.5 since we compared relatively small and homogenous groups with comparably diverse phenotypes.

Transgene Comparison

The overall proteomic comparison of the two groups of 2F5-scFv-Fc ($n = 6$) and 3D6-scFv-Fc ($n = 6$) producers in triplicates revealed 60 statistically significant differentially expressed proteins. According to the heterogeneity of the population, the recombinant product itself was not identified to be differentially expressed. This is in accordance with previously published intracellular flow cytometry data (Mader et al., 2013; Mayrhofer et al., 2014) showing comparably no severe differences in intracellular product levels between 2F5- and 3D6-sc-FvFc producers. Therefore the significantly lower specific productivities for 2F5-scFv-Fc producers strengthen the assumption that translation and/or ER throughput is decelerated in 2F5-scFv-Fc clones. Gene enrichment analysis of the 60 differentially expressed proteins highlighted among others the processes of proliferation and protein folding to be significantly enriched. We found very interesting candidates taking part in several, probably phenotype related, processes. A detailed description of their cellular functions can be found in the supplemental material. The higher abundance of proteins involved in proliferation, apoptosis, and cellular stress (f.i. Lgals3, Hspd1, Hspe1, Mcm5) in the group of the apparently difficult to produce 2F5-scFv-Fc clones suggests an adjustment to permanent stressful cellular conditions. Furthermore, the higher abundance of proteins interacting with chromatin and histones in the group of 2F5-scFv-Fc producers (f.i. Nat10, Ncl, Ipo5, Mta2) might indicate abnormal regulation of the chromatin structure. The identification of four differentially expressed Glutathione S-transferases (Gstp2, Gstp1, Gstm6, and Gsto1) higher abundant in the group of 3D6-scFv-Fc producers indicates a high impact of S-glutathionylation in cellular regulation (e.g., stress response and control of cell-signaling pathways). The higher relative abundances of several proteins involved in folding (Pdia3, Calr, Pdia4, Dnajc7, P4Hb, Hsp90b1, Grpel1) and translation (Eif5a, Eif4a1) in the group of 3D6-scFv-Fc producers may be attributed to their overall higher specific production rate. Also the elevated levels of oxidative stress markers (Prdx1, Cat, Txnrd1) may reflect the overall higher ER throughput of 3D6-scFv-Fc producers. Reactive oxygen species (ROS) can be produced as by-products of oxygen-utilizing enzymatic reactions, such as the mitochondrial respiratory chain. Furthermore, there is accumulating evidence that protein folding, endoplasmic reticulum (ER) stress, and the production of ROS are interlinked (Malhotra and Kaufman, 2007). One very interesting finding was that Sec23Ip, a protein involved in the organization of ER exit sites, was less abundant in 2F5-scFv-Fc producers (higher in 3D6-scFv-Fc producers). Sec23Ip depletion or overexpression alters ER exit sites morphology and a reduced level delays export from the ER (Ong et al., 2010; Shimoi et al., 2005). Hasegawa et al. (2011) recently reported a striking recombinant CHO phenotype that showed intracellular crystallization of a recombinantly produced model IgG. The study of this particular cell line provided insights in maximum cellular secretory capacity and identified the ER export as a rate-limiting bottleneck for the particular IgG. These results could be an indication of aggravated ER exit of 2F5-scFv-Fc fragments, and Sec23Ip might play a crucial role for the decreased secretory throughput. Hasegawa et al. suggested that an acidic cluster on the surface of the V_H complementarity determining regions (CDRs) of their model IgG was important for its in vivo crystallization. Taking a closer look at the surface charge of the variable regions of 2F5 and 3D6 (Fig. 2) reveals individual surface charge distributions between the two antibodies. Therefore, we speculate that the surface charge might contribute to the limited secretory throughput.

Transgene Delivery Comparison

In the second data evaluation approach, we investigated proteomic changes caused by elevated specific productivities and compared RMCE versus plasmid versus BAC clones for both antibody groups of 2F5- and 3D6-scFv-Fc producers separately. We were only interested in expression patterns that correlated with the specific productivity or growth rate and therefore neglected identified statistically significant differentially expressed proteins that showed the highest or lowest abundance in the group of plasmid-derived clones. Several of the identified proteins were already somehow linked to expression in recombinant mammalian cells (f.i. Anxa1, Hsp90ab1, Hspd1, Lgals1, Pdia3, Pdia6, and Vim) (Alete et al., 2005; Baik et al., 2006, 2008; Meleady et al., 2008, 2011; Nissom et al., 2006; Seth et al., 2007; Van Dyk et al., 2003; Wingens et al., 2015; Yee et al., 2008) (Supplemental Tables SII and SIII). We identified 58 proteins plus the immunoglobulin γ-chain (recombinant product) correlating with specific productivities in both transgene delivery comparisons of 2F5- and 3D6-scFv-Fc producers. Fourty-five of these proteins were positively and thirteen negatively correlating to specific productivity. Since these proteins were identified for both transgene delivery comparisons, we consider them to be commonly up- or down-regulated with increasing specific productivities. Gene enrichment analysis revealed cell cycle, vesicle-mediated transport as well as transport in general to be significantly enriched, which is in accordance with the phenotypical observations. Cell death was also identified to be statistically enriched and may be related to the counterbalance of cellular stress. The highest fold-change from low producing RMCE clones to high producing BAC clones was identified for the protein leukocyte elastase inhibitor A (Serpinb1a). Serpinb1a negatively

regulates the activity of neutrophil proteases (Cooley et al., 2001) and overexpression of Serpinb1 was shown to increase recombinant IgG productivity in CHO (Lin et al., 2015). Cathepsin B, which was also positively correlating with specific productivity, is a protease predominantly present in lysosomes. It is well described that autophagy and lysosomal degradation is activated during ER stress (Cheng and Yang, 2011; Kaminskyy and Zhivotovsky, 2012). These results suggest that independent of the produced scFv-Fc fragment ER stress increases with increasing specific productivity. Galectin-1 levels were also increasing in cells with a higher specific productivity in both antibody transgene delivery comparisons. There is evidence that Galectin-1 is involved in the regulation of cell growth. Higher Galectin-1 levels in higher producing but slower growing cell lines suggest that Galectin-1 may be an interesting target for proliferation engineering. We also identified three proteins, gelsolin, Ras-related protein Rab-1A, and vesicle-trafficking protein SEC22b, which are all involved in vesicle mediated transport, to be expressed at increased levels in the higher producing cell lines in both groups. Assuming that secretion engineering might improve the achievable productivities in recombinant CHO cells (Peng and Fussenegger, 2009), these three proteins may be interesting targets for overexpression. Remarkably, we identified fewer proteins that were negatively correlating to specific productivity in both groups. The proteasome-associated protein ECM29 was identified to be down-regulated with increasing specific productivity in both groups. It is believed that ECM29 serves as an adaptor for coupling 26 S-proteasomes to specific cellular compartments (Gorbea et al., 2010). Summarizing the observed results, an increase of lysosomal proteins and a decrease of proteasome associated proteins with increasing specific productivities has been shown. We speculate that at low specific productivities and moderate ER stress levels, the degradation of unfolded proteins is sufficient via the ubiquitin proteasome system. With increasing ER throughput and increasing levels of unfolded proteins the recombinant cells might have to switch to macro-autophagy to survive. It has been previously described that the ubiquitin proteasome system and autophagy are interconnected and inhibition of proteasome function leads to activation of autophagy to compensate for the reduced proteasome function (Ding et al., 2007; Korolchuk et al., 2010).

The fact that we identified numerous proteins only in one transgene delivery comparison is another hint that the type of recombinant product itself has an impact on the proteome of recombinant CHO cells.

We identified only two proteins that displayed an opposing trend in the expression pattern comparing the two groups. Both proteins, however, were only slightly changed in one group and only passed the threshold of >1.50-fold change in the other group. The proteins were Cullin-associated NEDD8-dissociated protein 1 (Cand1) as well as protein SON. Cand1 is involved in the regulation of SCF ubiquitin ligases (Chua et al., 2011; Olma and Dikic, 2013). Ubiquitination plays a crucial role in the nuclear factor-κB (NFκB) pathway, endocytic trafficking, DNA repair, and protein degradation (Grabbe et al., 2011). SON acts as an mRNA splicing cofactor and promotes splicing of many cell-cycle and DNA-repair transcripts (Ahn et al., 2011). Since these proteins were the only ones identified that showed an opposing expression pattern comparing the identified proteins positively or negatively correlating to the specific productivity in the transgene delivery comparison of 2F5- and 3D6-scFv-Fc producers, we believe that these proteins might play a crucial role in cellular regulations leading to the observed phenotypes.

Conclusion

We could clearly show that distinct recombinant proteins evoke different proteomic responses within the cells. Furthermore, we found indications that the presumably hard to produce 2F5-scFv-Fc fragment induces permanent stress within the recombinant CHO cells. The higher abundance of histone and chromatin interacting proteins in the group of 2F5-scFv-Fc producers might be a result of abnormal regulation of the chromatin structure. Furthermore, in context with a study published by Hasegawa et al. (2011), the lower levels of Sec23lp may be connected to the decelerated ER-throughput, and probably caused by the surface charge distribution on the CDRs of the 2F5-scFv-Fc .

In the second part of this study, we found evidence that ER stress is increasing with higher specific productivities for both groups of 2F5- and 3D6-scFv-Fc producers. The presented data also suggest that macroautophagy might play a crucial role for survival of ER-stressed high-producers. Interestingly, we could identify two proteins, Cand1 and SON, which showed the opposite correlation to the specific production rate in the two groups. Therefore, these proteins might be some key regulators of specific production rates.

In summary, we could show that the level of specific product secretion largely influences the cell's proteome. More striking is the fact that we could show that the recombinantly produced protein itself considerably influences the cell's physiology. In a consequence, these findings also suggest that host engineering strategies might only work for a subset of recombinant proteins.

This work has been financially supported by Polymun Scientific GmbH, Klosterneuburg, Austria and the University of Natural Resources and Life Sciences, Vienna. WS and PM were also supported by the PhD program "BioToP-Biomolecular Technology of Proteins" (Austrian Science Fund, FWF Project W1224). We thank Alexander Mader, Andreas Gili, Bernhard Kratzer, Bernhard Prewein, and Martina Hofbauer for their help with cell line development.

References

Ahn EY, DeKelver RC, Lo MC, Nguyen TA, Matsuura S, Boyapati A, Pandit S, Fu XD, Zhang DE. 2011. SON controls cell-cycle progression by coordinated regulation of RNA splicing. Mol Cell 42(2):185–198.

Alete DE, Racher AJ, Birch JR, Stansfield SH, James DC, Smales CM. 2005. Proteomic analysis of enriched microsomal fractions from GS-NS0 murine myeloma cells with varying secreted recombinant monoclonal antibody productivities. Proteomics 5(18):4689–4704.

Baik JY, Joo EJ, Kim YH, Lee GM. 2008. Limitations to the comparative proteomic analysis of thrombopoietin producing Chinese hamster ovary cells treated with sodium butyrate. J Biotechnol 133(4):461–468.

Baik JY, Lee MS, An SR, Yoon SK, Joo EJ, Kim YH, Park HW, Lee GM. 2006. Initial transcriptome and proteome analyses of low culture temperature-induced expression in CHO cells producing erythropoietin. Biotechnol Bioeng 93(2):361–371.

Baycin-Hizal D, Tabb DL, Chaerkady R, Chen L, Lewis NE, Nagarajan H, Sarkaria V, Kumar A, Wolozny D, Colao J, Jacobson E, Tian Y, O'Meally RN, Krag SS, Cole RN, Palsson BO, Zhang H, Betenbaugh M. 2012. Proteomic analysis of Chinese hamster ovary cells. J Proteome Res 11(11):5265–76.

Blaas L, Musteanu M, Eferl R, Bauer A, Casanova E. 2009. Bacterial artificial chromosomes improve recombinant protein production in mammalian cells. BMC Biotechnol 9:3.

Blaas L, Musteanu M, Grabner B, Eferl R, Bauer A, Casanova E. 2012. The use of bacterial artificial chromosomes for recombinant protein production in mammalian cell lines. Methods Mol Biol 824:581–593.

Buchanan A, Clementel V, Woods R, Harn N, Bowen MA, Mo W, Popovic B, Bishop SM, Dall'Acqua W, Minter R, Jermutus L, Bedian V. 2013. Engineering a therapeutic IgG molecule to address cysteinylation, aggregation and enhance thermal stability and expression. MAbs 5(2):255–262.

Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM, Pascual-Montano A. 2007. GENECODIS: A web-based tool for finding significant concurrent annotations in gene lists. Genome Biol 8(1):R3.

Cheng Y, Yang JM. 2011. Survival and death of endoplasmic-reticulum-stressed cells: Role of autophagy. World J Biol Chem 2(10):226–231.

Chua YS, Boh BK, Ponyeam W, Hagen T. 2011. Regulation of cullin RING E3 ubiquitin ligases by CAND1 in vivo. PLoS ONE 6(1):e16071.

Clarke C, Henry M, Doolan P, Kelly S, Aherne S, Sanchez N, Kelly P, Kinsella P, Breen L, Madden SF. 2012. Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. BMC Genomics 13(1):656.

Cooley J, Takayama TK, Shapiro SD, Schechter NM, Remold-O'Donnell E. 2001. The serpin MNEI inhibits elastase-like and chymotrypsin-like serine proteases through efficient reactions at two active sites. Biochemistry 40(51):15762–15770.

Datta P, Linhardt RJ, Sharfstein ST. 2013. An 'omics approach towards CHO cell engineering. Biotechnol Bioeng 110(5):1255–1271.

Ding WX, Ni HM, Gao W, Yoshimori T, Stolz DB, Ron D, Yin XM. 2007. Linking of autophagy to ubiquitin-proteasome system is important for the regulation of endoplasmic reticulum stress and cell viability. Am J Pathol 171(2):513–524.

Garber E, Demarest SJ. 2007. A broad range of Fab stabilities within a host of therapeutic IgGs. Biochem Biophys Res Commun 355(3):751–757.

Gorbea C, Pratt G, Ustrell V, Bell R, Sahasrabudhe S, Hughes RE, Rechsteiner M. 2010. A protein interaction network for Ecm29 links the 26 S proteasome to molecular motors and endosomal components. J Biol Chem 285(41):31616–31633.

Grabbe C, Husnjak K, Dikic I. 2011. The spatial and temporal organization of ubiquitin networks. Nat Rev Mol Cell Biol 12(5):295–307.

Hasegawa H, Wendling J, He F, Trilisky E, Stevenson R, Franey H, Kinderman F, Li G, Piedmonte DM, Osslund T, Shen M, Ketchem RR. 2011. In vivo crystallization of human IgG in the endoplasmic reticulum of engineered Chinese hamster ovary (CHO) cells. J Biol Chem 286(22):19917–31.

Kaminskyy V, Zhivotovsky B. 2012. Proteases in autophagy. Biochim Biophys Acta 1824(1):44–50.

Kildegaard HF, Baycin-Hizal D, Lewis NE, Betenbaugh MJ. 2013. The emerging CHO systems biology era: Harnessing the 'omics revolution for biotechnology. Curr Opin Biotechnol 24(6):1102–1107.

Kim JY, Kim YG, Lee GM. 2012. CHO cells in biotechnology for production of recombinant proteins: Current state and further potential. Appl Microbiol Biotechnol 93(3):917–930.

Korolchuk VI, Menzies FM, Rubinsztein DC. 2010. Mechanisms of cross-talk between the ubiquitin-proteasome and autophagy-lysosome systems. FEBS Lett 584(7):1393–1398.

Kunert R, Casanova E. 2013. Recent advances in recombinant protein production: BAC-based expression vectors, the bigger the better. Bioengineered 4(4):258–261.

Kunert R, Rüker F, Katinger H. 1998. Molecular characterization of five neutralizing anti-HIV type 1 antibodies: Identification of nonconventional D segments in the human monoclonal antibodies 2G12 and 2F5. AIDS Res Hum Retroviruses 14(13):1115–1128.

Lin N, Brooks J, Sealover N, George HJ, Kayser KJ. 2015. Overexpression of Serpinb1 in Chinese hamster ovary cells increases recombinant IgG productivity. J Biotechnol 193:91–99.

Linge A, Maurya P, Friedrich K, Baretton GB, Kelly S, Henry M, Clynes M, Larkin A, Meleady P. 2014. Identification and functional validation of RAD23B as a potential protein in human breast cancer progression. J Proteome Res 13(7):3212–3222.

Mader A, Prewein B, Zboray K, Casanova E, Kunert R. 2013. Exploration of BAC versus plasmid expression vectors in recombinant CHO cells. Appl Microbiol Biotechnol 97(9):4049–4054.

Malhotra JD, Kaufman RJ. 2007. The endoplasmic reticulum and the unfolded protein response. Semin Cell Dev Biol 18(6):716–731.

Mayrhofer P, Kratzer B, Sommeregger W, Steinfellner W, Reinhart D, Mader A, Turan S, Qiao J, Bode J, Kunert R. 2014. Accurate comparison of antibody expression levels by reproducible transgene targeting in engineered recombination-competent CHO cells. Appl Microbiol Biotechnol 98(23):9723–9733.

Meleady P, Doolan P, Henry M, Barron N, Keenan J, O'Sullivan F, Clarke C, Gammell P, Melville MW, Leonard M, Clynes M. 2011. Sustained productivity in recombinant Chinese hamster ovary (CHO) cell lines: Proteome analysis of the molecular basis for a process-related phenotype. BMC Biotechnol 11:78.

Meleady P, Henry M, Gammell P, Doolan P, Sinacore M, Melville M, Francullo L, Leonard M, Charlebois T, Clynes M. 2008. Proteomic profiling of CHO cells with enhanced rhBMP-2 productivity following co-expression of PACEsol. Proteomics 8(13):2611–2624.

Meleady P, Hoffrogge R, Henry M, Rupp O, Bort JH, Clarke C, Brinkrolf K, Kelly S, Müller B, Doolan P, Hackl M, Beckmann TF, Noll T, Grillari J, Barron N, Pühler A, Clynes M, Borth N. 2012. Utilization and evaluation of CHO-specific sequence databases for mass spectrometry based proteomics. Biotechnol Bioeng 109(6):1386–1394.

Nissom PM, Sanny A, Kok YJ, Hiang YT, Chuah SH, Shing TK, Lee YY, Wong KT, Hu WS, Sim MY, Philp R. 2006. Transcriptome and proteome profiling to understanding the biology of high productivity CHO cells. Mol Biotechnol 34(2):125–140.

Nogales-Cadenas R, Carmona-Saez P, Vazquez M, Vicente C, Yang X, Tirado F, Carazo JM, Pascual-Montano A. 2009. GeneCodis: Interpreting gene lists through enrichment analysis and integration of diverse biological information. Nucleic Acids Res 37(suppl 2):W317–W322.

Olma MH, Dikic I. 2013. Cullins getting undressed by the protein exchange factor Cand1. Cell 153(1):14–16.

Ong YS, Tang BL, Loo LS, Hong W. 2010. p125A exists as part of the mammalian Sec13/Sec31 COPII subcomplex to facilitate ER-Golgi transport. J Cell Biol 190(3):331–345.

Peng R-W, Fussenegger M. 2009. Engineering the secretory pathway in mammalian Cells. Cell line development, Netherlands: Springer. p 233–248.

Seth G, Philp RJ, Lau A, Jiun KY, Yap M, Hu WS. 2007. Molecular portrait of high productivity in recombinant NS0 cells. Biotechnol Bioeng 97(4):933–951.

Shimoi W, Ezawa I, Nakamoto K, Uesaki S, Gabreski G, Aridor M, Yamamoto A, Nagahama M, Tagaya M, Tani K. 2005. p125 is localized in endoplasmic reticulum exit sites and involved in their organization. J Biol Chem 280(11): 10141–10148.

Tabas-Madrid D, Nogales-Cadenas R, Pascual-Montano A. 2012. GeneCodis3: A non-redundant and modular enrichment analysis tool for functional genomics. Nucleic Acids Res 40(W1):W478–W483.

Urlaub G, Chasin LA. 1980. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. Proc Natl Acad Sci USA 77(7):4216–4220.

Van Dyk DD, Misztal DR, Wilkins MR, Mackintosh JA, Poljak A, Varnai JC, Teber E, Walsh BJ, Gray PP. 2003. Identification of cellular changes associated with increased production of human growth hormone in a recombinant Chinese hamster ovary cell line. Proteomics 3(2):147–156.

Walsh G. 2014. Biopharmaceutical benchmarks 2014. Nat Biotechnol 32(10):992–1000.

Wingens M, Gätgens J, Schmidt A, Albaum SP, Büntemeyer H, Noll T, Hoffrogge R. 2015. 2D-DIGE screening of high-productive CHO cells under glucose limitation-basic changes in the proteome equipment and hints for epigenetic effects. J Biotechnol 201:86–97.

Yee JC, de Leon Gatti M, Philp RJ, Yap M, Hu WS. 2008. Genomic and proteomic exploration of CHO and hybridoma cells under sodium butyrate treatment. Biotechnol Bioeng 99(5):1186–1204.

Zboray K, Sommeregger W, Bogner E, Gili A, Sterovsky T, Fauland K, Grabner B, Stiedl P, Moll HP, Bauer A, Kunert R, Casanova E. 2015. Heterologous protein production using euchromatin-containing expression vectors in mammalian cells. Nucleic Acids Res 43(16):e102.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Proteomic differences in recombinant CHO cells producing two similar antibody fragments

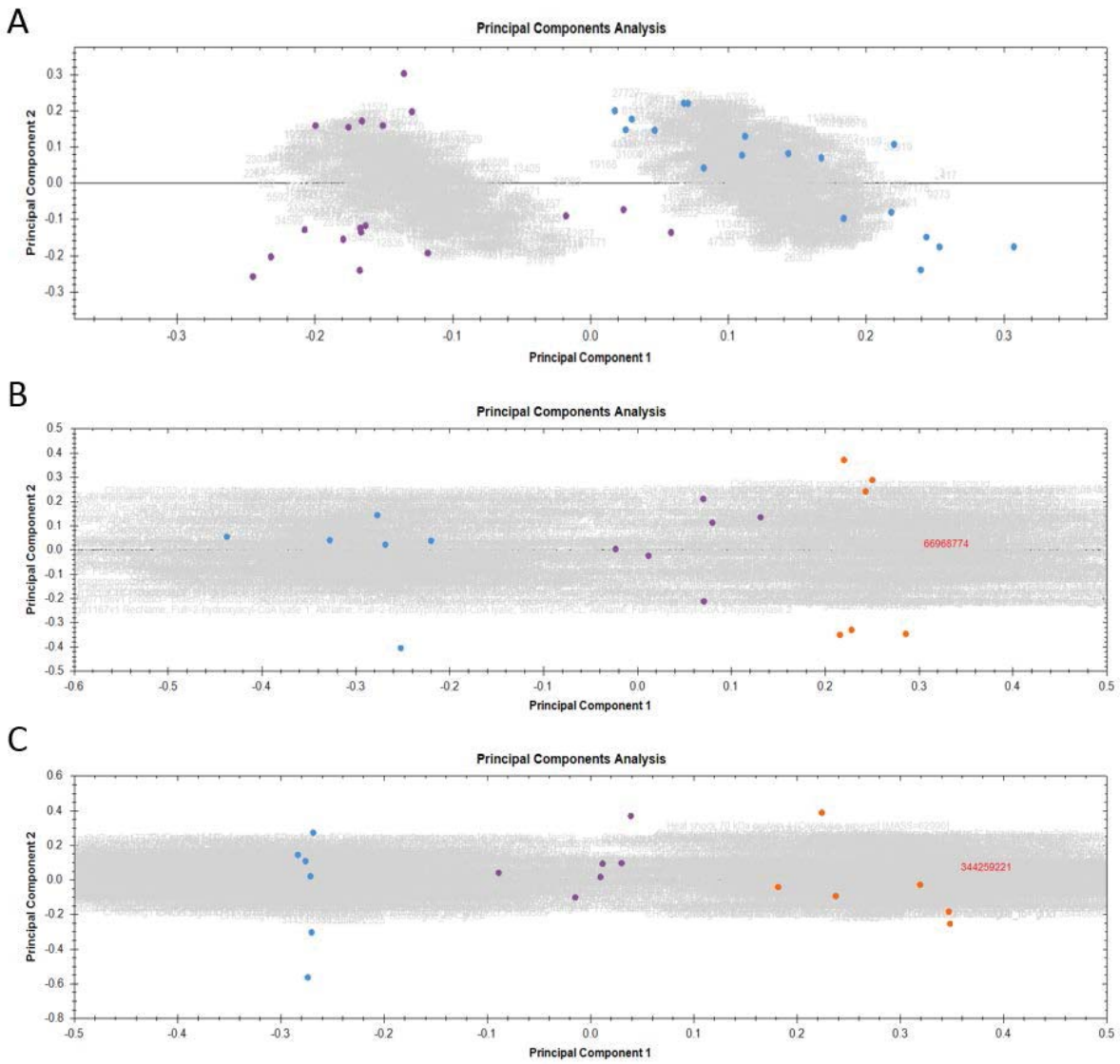
Supplemental material

scFv-Fc fragments

```
2F5scFv-Fc-aa      RITLKESGPPLVKPTQTLTLTCSFSGFSLSDFGVGVGWIRQPPGKALEWLAII-YSDDDK
3D6scFv-Fc-aa      EVQLVESGGGLVQPGRSLRLSCAASGFTFNDYA--MHWVRQAPGKGLEWVSGISWDSSSI
                   .: * ***  **:* ::* *:*: ***::.*:.  : *:** ***.***:: * :....

2F5scFv-Fc-aa      RYSPSLNTRLTITKDTSKNQVVLVMTRVSPVDTATYFCAHRRGPTTLFGVPIARGPVNAM
3D6scFv-Fc-aa      GYADSVKGRFTISRDNAKNSLYLQMNSLRAEDMALYYCVKGRDYYDSGG-----YFTVAF
                    *: *:: *:**::*.:**.: * *. :   * * *:*.: *      *       . *:

2F5scFv-Fc-aa      DVWGQGITVTISSGGGGSGGGGSGGGGSALQLTQSPSSLSASVGDRITITCRASQGVTSA
3D6scFv-Fc-aa      DIWGQGTMVTVSSGGGGSGGGGSGGGGSDIQMTQSPSTLSASVGDRVTITCRASQSISRW
                   *:****  **:***************** :*:*****:********:********.::

2F5scFv-Fc-aa      LAWYRQKPGSPPQLLIYDASSLESGVPSRFSGSGSGTEFTLTISTLRPEDFATYYCQQLH
3D6scFv-Fc-aa      LAWYQQKPGKVPKLLIYKASSLESGVPSRFSGSGSGTEFTLTISSLQPDDFATYYCQQYN
                   ****:****. *:****.**************************:*:*:********* .

2F5scFv-Fc-aa      FYPHTFGGGTRVDVREPKSSDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVT
3D6scFv-Fc-aa      SY--SFGPGTKVDIKEPKSSDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVT
                    *  :** **:**::*********************************************

2F5scFv-Fc-aa      CVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYK
3D6scFv-Fc-aa      CVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYK
                   ************************************************************

2F5scFv-Fc-aa      CKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVE
3D6scFv-Fc-aa      CKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVE
                   ************************************************************

2F5scFv-Fc-aa      WESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKS
3D6scFv-Fc-aa      WESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKS
                   ************************************************************

2F5scFv-Fc-aa      LSLSPGK
3D6scFv-Fc-aa      LSLSPGK
                   *******
```

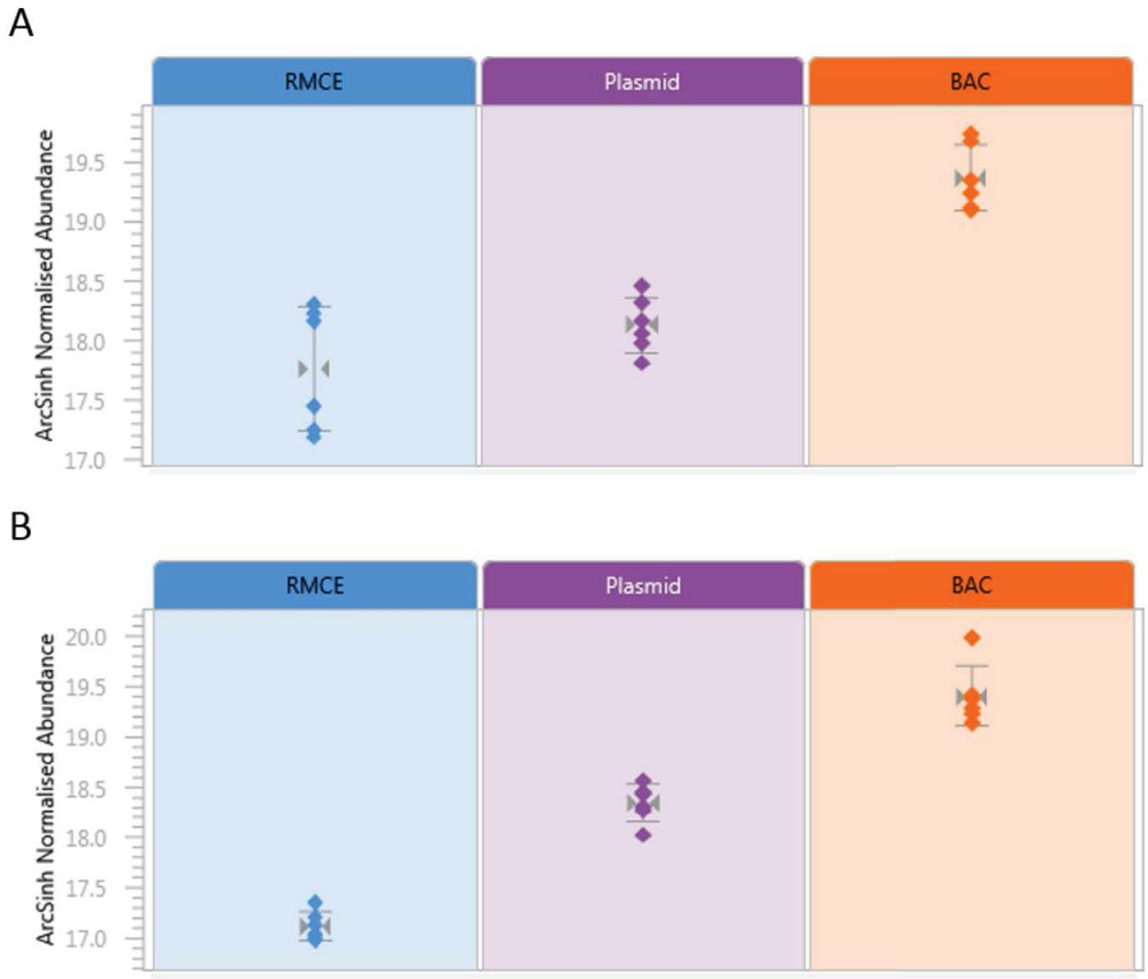
Supplemental Figure 1: Clustal sequence alignment of 2F5- and 3D6-scFv-Fc.

Principal component analysis (PCA)



Supplemental Figure 2: Progenesis output: PCA of all differential peptides (p≤0.05) used for protein identification; A: transgene comparison of 3D6 (n=18, blue) versus 2F5 (n=18, purple) samples; B: 2F5 transgene delivery comparison of RMCE (n=6, blue) versus Plasmid (n=6 purple) versus BAC (n=6, orange) samples; C: 3D6 transgene delivery comparison of RMCE (n=6, blue) versus Plasmid (n=6 purple) versus BAC (n=6, orange) samples.

Recombinant product identified in proteomic analysis



Supplemental Figure 3: Progenesis output: Normalised abundance of human Ig gamma-1 chain C region in 2F5-scFv-Fc samples (A) and 3D6-scFv-Fc samples (B).

Identified DE proteins (transgene comparison)

Supplemental Table 1: Transgene comparison: 2F5-scFv-Fc (n=18) versus 3D6-scFv-Fc (n=18) identified differential proteins (n=60); number of peptides used for quantitation ≥ 2; Anova p Value ≤ 0.05; fold change **≥ 1.2x**. PCA of features used for identifications are shown in supplemental figure 2, A. Table is sorted by fold-change.

Description	Gene ID	Peptides used for quant.	Anova (p)	Fold change	Highest mean
Galectin-3	Lgals3	2	0.00	1.75	2F5
Calcium/calmodulin-dependent protein kinase type II subunit delta	Camk2d	2	0.00	1.59	2F5
DNA replication licensing factor MCM5	Mcm5	2	0.01	1.58	2F5
60 kDa heat shock protein, mitochondrial	Hspd1	10	0.00	1.50	2F5
N-acetyltransferase 10	Nat10	2	0.02	1.49	2F5
Nucleolin	Ncl	7	0.00	1.43	2F5
10 kDa heat shock protein, mitochondrial	Hspe1	5	0.00	1.43	2F5
Heme oxygenase 2	Hmox2	3	0.02	1.42	2F5
Splicing factor 3B subunit 1	Sf3b1	3	0.00	1.36	2F5
NADPH:adrenodoxin oxidoreductase, mitochondrial	Fdxr	3	0.01	1.35	2F5
Tropomyosin alpha-4 chain	Tpm4	2	0.02	1.35	2F5
Intracellular adhesion molecule 1	Icam1	2	0.04	1.35	2F5
Putative ribosomal RNA methyltransferase NOP2	Nop2	2	0.01	1.29	2F5
Lanosterol synthase	Lss	3	0.00	1.28	2F5
BRI3-binding protein	Bri3bp	2	0.00	1.28	2F5
Importin-5	Ipo5	2	0.02	1.27	2F5
Drebrin-like	Dbnl	3	0.03	1.26	2F5
Antigen KI-67	Mki67	2	0.02	1.25	2F5
Nuclear mitotic apparatus protein 1	Numa1	2	0.00	1.23	2F5
Filamin-B	Flnb	2	0.04	1.23	2F5
Metastasis-associated protein MTA2	Mta2	2	0.01	1.22	2F5
Annexin A2	Anxa2	2	0.02	1.22	2F5
Cysteine desulfurase, mitochondrial	Nfs1	2	0.00	1.22	2F5
Serine/arginine repetitive matrix protein 2	Srrm2	3	0.00	1.22	2F5
Aldehyde oxidase	Aox1	2	0.01	1.21	2F5
Succinyl-CoA:3-ketoacid-coenzyme A transferase 1, mitochondrial	Oxct1	3	0.00	1.21	2F5
DNA topoisomerase 1	Top1	2	0.01	1.21	2F5
Glutathione S-transferase P 2	Gstp2	3	0.00	2.74	3D6
Glutathione S-transferase P 1	Gstp1	3	0.00	2.19	3D6
Peroxiredoxin-1	Prdx1	6	0.00	1.76	3D6

Sulfide:quinone oxidoreductase, mitochondrial	Sqrdl	2	0.03	1.75	3D6
Golgi-associated plant pathogenesis-related protein 1	Glipr2	2	0.00	1.64	3D6
[9]Protein disulfide-isomerase A3	Pdia3	2	0.01	1.58	3D6
Catalase	Cat	2	0.00	1.52	3D6
Calreticulin	Calr	4	0.00	1.47	3D6
Basement membrane-specific heparan sulfate proteoglycan core protein	Hspg2	3	0.01	1.45	3D6
Von Willebrand factor A domain-containing protein 5A	Vwa5a	2	0.03	1.42	3D6
Hydroxymethylglutaryl-CoA lyase, mitochondrial	Hmgcl	2	0.00	1.39	3D6
SEC23-interacting protein	Sec23Ip	2	0.00	1.38	3D6
ATP-citrate synthase	Acly	2	0.00	1.38	3D6
Fumarylacetoacetase	Fah	3	0.01	1.36	3D6
Glutamate--cysteine ligase regulatory subunit	Gclm	4	0.02	1.35	3D6
Alpha-enolase	Eno1	11	0.00	1.33	3D6
Protein disulfide-isomerase A4	Pdia4	3	0.00	1.33	3D6
Pyruvate carboxylase, mitochondrial	Pcx	4	0.01	1.31	3D6
Thioredoxin reductase 1, cytoplasmic	Txnrd1	3	0.02	1.29	3D6
Glucose-6-phosphate 1-dehydrogenase	G6Pdx	4	0.00	1.28	3D6
Glutathione S-transferase Mu 6	Gstm6	4	0.01	1.28	3D6
Eukaryotic translation initiation factor 5A-1	Eif5a	2	0.01	1.28	3D6
Nucleobindin-2	Nucb2	2	0.00	1.27	3D6
Eukaryotic initiation factor 4A-I	Eif4a1	3	0.00	1.26	3D6
Glutathione S-transferase omega-1 isoform 1	Gsto1	3	0.00	1.26	3D6
Alcohol dehydrogenase (NADP+)	Adh5	2	0.00	1.26	3D6
DnaJ homolog subfamily C member 7	Dnajc7	2	0.02	1.23	3D6
Importin subunit beta-1	Kpnb1	6	0.00	1.23	3D6
Protein disulfide-isomerase	P4Hb	3	0.02	1.23	3D6
Endoplasmin	Hsp90b1	5	0.01	1.23	3D6
GrpE protein homolog 1, mitochondrial	Grpel1	3	0.00	1.21	3D6
Nucleoside diphosphate kinase A	Nme1	2	0.02	1.21	3D6
Chloride intracellular channel protein 4	Clic4	3	0.01	1.21	3D6

Discussion of identified DE proteins (transgene comparison)

Galectin-3 (Lgals3), which was found to be present in higher amounts in 2F5-scFv-Fc producers, is an interesting protein that is involved in several cellular processes like adhesion, cell cycle progression and apoptosis as well as inflammatory processes (Dumic et al. 2006). In the cytoplasm Galectin-3 interacts with the apoptosis repressor Bcl-2; in the nucleus, Galectin-3 is a required pre-mRNA splicing factor (Haudek et al. 2010). Two other proteins with elevated levels in the group of 2F5-scFv-Fc producers are heat shock protein 60 (Hsp60/Hspd1) and its co-chaperone heat shock protein 10 (Hsp10/Hspe1). Amongst other heat shock proteins, Hspd1 was shown to be important for cellular survival under stressful conditions (Rossi et al. 2002). Another versatile higher abundant protein in the group of 2F5-scFv-Fc producers is DNA replication licensing factor MCM5 (Mcm5). Mcm5 is one of the key proteins in DNA replication initiation but it was also shown to be essential in Stat1-mediated transcriptional activation (Snyder et al. 2005). N-acetyltransferase 10 (Nat10) was as well found to be higher in 2F5-scFv-Fc producers. Nat10 influences histone acetylation and up-regulates telomerase activity through transactivation of hTERT promoter (Lv et al. 2003). Furthermore, it was described that Nat10 is involved in DNA damage response and increases resistance to genotoxicity (Liu et al. 2007). The protein nucleolin was found to be 1.43-fold higher in the 2F5 producers. As a DNA-binding protein, nucleolin triggers chromatin decondensation, facilitates transcription and modulates DNA replication (Angelov et al. 2006). Nucleolin is predominantly present in the nucleolus, where it also interacts with rRNA and assists in its maturation and processing (Ginisty et al. 1998). In the cytoplasm, nucleolin interacts with mature mRNAs. Recently, nucleolin was shown to associate with several mRNAs encoding proteins with roles in cell growth and proliferation, including Bcl-2, p53, cyclin I, and Akt1 (Abdelmohsen et al. 2011). Importin-5 which was also found to be slightly higher in 2F5-scFv-Fc producers mediates the import of ribosomal proteins and core histones into the nucleus (Baake et al. 2001; Jäkel and Görlich 1998). Another very interesting protein up regulated in 2F5-scFv-Fc producers is Mta2, which is part of NuRD, the nucleosome remodelling and deacetylase complex (Xue et al. 1998). NuRD prevents the accumulation of spontaneous DNA damage and regulates apoptotic responses through p53 and p21 (Smeenk et al. 2010). Interestingly we could identify four Glutathione S-transferases (Gstp2, Gstp1, Gstm6 and Gsto1) to be less abundant in 2F5-scFv-Fc producers (higher in 3D6-scFv-Fc producers). S-glutathionylation has an impact on several proteins including signalling proteins, transcription factors as well as heat shock proteins (Tew et al. 2011).

Peroxiredoxin-1 (Prdx1) levels were also higher in 3D6-scFv-Fc producing CHO cells. Peroxiredoxin-1 has an important role in maintaining the redox state *in vivo*. Furthermore, Prdx1 and other members of the Prdx family, are also involved in regulating growth factor signalling pathways (Rhee et al. 2003). Golgi-associated plant pathogenesis-related protein 1 (Glioma pathogenesis-related protein 2 (Glipr2)) was also higher in abundance in 3D6-scFv-Fc samples. Overexpression of Glipr2 in human kidney 2 (HK-2) cells suggested that it has an ERK 1/2 activating function. ERK 1/2 signalling itself plays a central role in cell proliferation and apoptosis (Mebratu and Tesfaigzi 2009). Catalase, which showed elevated expression in 3D6 samples, is an antioxidant enzyme found in nearly all living organisms that are exposed to oxygen and catalyzes the reduction of H_2O_2 to water (Kirkman and Gaetani 2007) and is also a marker of oxidative stress. Thioredoxin reductase 1, which was also slightly increased in 3D6 producers, is another indicator of oxidative stress. Reactive oxygen species (ROS) can be produced as by-products of oxygen-utilizing enzymatic reactions, such as the mitochondrial respiratory chain. Furthermore, there is accumulating evidence that protein folding, endoplasmic reticulum (ER) stress and the production of ROS are interlinked (Malhotra and Kaufman 2007). One very interesting finding was that Sec23Ip, a protein involved in the organization of ER exit sites, was less abundant in 2F5-scFv-Fc producers (higher in 3D6-scFv-Fc producers). Sec23Ip depletion or overexpression alters ER exit sites morphology and a reduced level delays export from the ER (Ong et al. 2010; Shimoi et al. 2005). Furthermore, we detected several proteins involved in folding (Pdia3, Calr, Pdia4, Dnajc7, P4Hb, Hsp90b1, Grpel1) and translation (Eif5a, Eif4a1) to be higher abundant in 3D6-scFv-Fc producers.

Identified DE proteins (transgene delivery comparison)

Supplemental Table 2: Transgene delivery comparison: 2F5-scFv-Fc RMCE (n=6) versus Plasmid (n=6) versus BAC (n=6) samples. Identified differentially expressed proteins (n=109) between RMCE versus Plasmid versus BAC correlating to qP and μ (highest BAC and lowest RMCE or vice versa); number of peptides used for quantitation ≥ 2; Anova p Value ≤ 0.05; fold change **≥ 1.5x**. PCA of features used for identification are shown in supplemental figure 2, B. Table is sorted by fold-change and positive or negative correlation with qP.

Description	Peptides used for quant.	Anova (p)	Fold change	Highest mean	Lowest mean
Leukocyte elastase inhibitor A	7	0,0000	5,77	BAC	RMCE
Cornifin-A	3	0,0115	5,41	BAC	RMCE
[2,4,6]Galectin-1	9	0,0000	4,91	BAC	RMCE
NADH-cytochrome b5 reductase 1	2	0,0002	3,76	BAC	RMCE
Cathepsin B	7	0,0001	3,51	BAC	RMCE
Hypothetical protein LOC100752925	2	0,0000	3,42	BAC	RMCE
Myosin-9	4	0,0021	3,22	BAC	RMCE
Alpha-N-acetylgalactosaminidase	4	0,0000	3,20	BAC	RMCE
Hypothetical protein LOC100752970	3	0,0000	3,12	BAC	RMCE
Cysteine-rich with EGF-like domain protein 2	3	0,0000	3,01	BAC	RMCE
Plectin	2	0,0004	2,77	BAC	RMCE
Actin-related protein 2/3 complex subunit 1B	2	0,0061	2,71	BAC	RMCE
[9]Actin-related protein 2/3 complex subunit 2	6	0,0004	2,67	BAC	RMCE
[10]Gelsolin	3	0,0000	2,57	BAC	RMCE
Heat shock protein beta-1	2	0,0117	2,51	BAC	RMCE
Isoleucyl-tRNA synthetase, mitochondrial	5	0,0000	2,41	BAC	RMCE
[2,4,6,9]Vimentin	24	0,0017	2,31	BAC	RMCE
Myoferlin	2	0,0003	2,28	BAC	RMCE
[4]Macrophage-capping protein	5	0,0000	2,27	BAC	RMCE
Glutathione peroxidase 1	2	0,0140	2,22	BAC	RMCE
Zyxin	3	0,0004	2,17	BAC	RMCE
Glycerol-3-phosphate dehydrogenase 1-like protein	2	0,0021	2,17	BAC	RMCE
Transgelin-2	4	0,0004	2,15	BAC	RMCE
Succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial	5	0,0000	2,13	BAC	RMCE
Pyridoxal-dependent decarboxylase domain-containing protein 1	2	0,0001	2,11	BAC	RMCE
LIM domain and actin-binding protein 1	4	0,0028	2,10	BAC	RMCE
Coronin-1B	8	0,0000	2,08	BAC	RMCE
Myosin-10	3	0,0010	2,05	BAC	RMCE
Rho GTPase-activating protein 1	4	0,0000	2,05	BAC	RMCE

Ras-related protein Rab-1A	2	0,0000	2,01	BAC	RMCE
Translocation protein SEC62	2	0,0001	2,01	BAC	RMCE
HIV Tat-specific factor 1 homolog	2	0,0119	1,97	BAC	RMCE
Actin-related protein 2	4	0,0008	1,97	BAC	RMCE
[9]ATP-citrate synthase	4	0,0001	1,97	BAC	RMCE
Transmembrane protein 43	2	0,0000	1,96	BAC	RMCE
Pyruvate carboxylase, mitochondrial	2	0,0144	1,95	BAC	RMCE
[6]Peptidyl-prolyl cis-trans isomerase B	7	0,0001	1,94	BAC	RMCE
Filamin-B	29	0,0000	1,91	BAC	RMCE
Ras-related protein Ral-B	2	0,0025	1,91	BAC	RMCE
[9]Microtubule-associated protein 1B	2	0,0000	1,89	BAC	RMCE
Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform	3	0,0002	1,89	BAC	RMCE
Nucleobindin-2	3	0,0001	1,85	BAC	RMCE
Estradiol 17-beta-dehydrogenase 12	3	0,0000	1,85	BAC	RMCE
Sulfated glycoprotein 1	10	0,0005	1,84	BAC	RMCE
4-trimethylaminobutyraldehyde dehydrogenase	2	0,0000	1,84	BAC	RMCE
Actin-related protein 3	5	0,0030	1,83	BAC	RMCE
Acid ceramidase	2	0,0002	1,77	BAC	RMCE
Src substrate cortactin	6	0,0016	1,76	BAC	RMCE
Glutathione S-transferase Mu 7	2	0,0004	1,76	BAC	RMCE
Myosin regulatory light chain 12B	2	0,0067	1,75	BAC	RMCE
Destrin	3	0,0008	1,75	BAC	RMCE
Reticulocalbin-3	3	0,0021	1,73	BAC	RMCE
Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12	2	0,0000	1,70	BAC	RMCE
NADH-cytochrome b5 reductase 3	3	0,0000	1,69	BAC	RMCE
Vesicle-trafficking protein SEC22b	3	0,0028	1,69	BAC	RMCE
[1]Fumarate hydratase	3	0,0000	1,67	BAC	RMCE
UPF0556 protein C19orf10-like	2	0,0012	1,67	BAC	RMCE
MHC class I antigen Hm1-C4	3	0,0176	1,66	BAC	RMCE
LIM and SH3 domain protein 1	8	0,0000	1,66	BAC	RMCE
Receptor expression-enhancing protein 5	2	0,0000	1,66	BAC	RMCE
Annexin A11	5	0,0009	1,66	BAC	RMCE
Long-chain specific acyl-CoA dehydrogenase, mitochondrial	8	0,0016	1,64	BAC	RMCE
Calmodulin	3	0,0008	1,62	BAC	RMCE
3-keto-steroid reductase	2	0,0002	1,61	BAC	RMCE
V-type proton ATPase subunit G 1	2	0,0005	1,58	BAC	RMCE
V-type proton ATPase subunit B, brain isoform	2	0,0005	1,58	BAC	RMCE
Septin-11	2	0,0011	1,57	BAC	RMCE
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A	3	0,0000	1,57	BAC	RMCE
Vesicle-associated membrane protein-associated protein A	5	0,0108	1,56	BAC	RMCE
Protein canopy-like 2	3	0,0000	1,56	BAC	RMCE
AP-2 complex subunit mu-like isoform 2 [Monodelphis domestica]	2	0,0001	1,53	BAC	RMCE

Golgi apparatus protein 1	4	0,0018	1,53	BAC	RMCE
Bone marrow stromal antigen 2	3	0,0441	1,52	BAC	RMCE
Prostaglandin E synthase 3	3	0,0053	5,14	RMCE	BAC
Heme oxygenase 1	7	0,0002	3,41	RMCE	BAC
Lipoprotein lipase	2	0,0000	2,99	RMCE	BAC
Nucleosome-binding protein 1	3	0,0060	2,80	RMCE	BAC
Glutamate--cysteine ligase regulatory subunit	5	0,0000	2,48	RMCE	BAC
[7]Macrophage migration inhibitory factor	2	0,0084	2,41	RMCE	BAC
Exportin-5	2	0,0000	2,33	RMCE	BAC
Phosphoribosylformylglycinamidine synthase	2	0,0000	2,32	RMCE	BAC
Anamorsin	3	0,0000	2,31	RMCE	BAC
Intracellular adhesion molecule 1	8	0,0000	2,30	RMCE	BAC
carbonyl reductase	2	0,0007	2,21	RMCE	BAC
Proteasome-associated protein ECM29-like	2	0,0028	2,21	RMCE	BAC
Spliceosome RNA helicase BAT1	4	0,0025	2,14	RMCE	BAC
[7]Thioredoxin reductase 1, cytoplasmic	7	0,0281	1,97	RMCE	BAC
Protein phosphatase 1G	3	0,0004	1,91	RMCE	BAC
Aldehyde dehydrogenase, mitochondrial	6	0,0006	1,91	RMCE	BAC
Lamina-associated polypeptide 2, isoforms alpha/zeta	8	0,0000	1,89	RMCE	BAC
Ribose-phosphate pyrophosphokinase 1	3	0,0000	1,87	RMCE	BAC
Protein SON	2	0,0006	1,79	RMCE	BAC
[9]Heat shock protein 75 kDa, mitochondrial	7	0,0000	1,78	RMCE	BAC
Proliferating cell nuclear antigen	3	0,0002	1,71	RMCE	BAC
Thioredoxin-like protein 1	4	0,0001	1,69	RMCE	BAC
Basic leucine zipper and W2 domain-containing protein 2	5	0,0000	1,68	RMCE	BAC
Flavin reductase	10	0,0183	1,66	RMCE	BAC
Cullin-4B	2	0,0035	1,63	RMCE	BAC
[7]Glutathione S-transferase P 2	2	0,0014	1,63	RMCE	BAC
6-phosphogluconate dehydrogenase, decarboxylating	9	0,0001	1,62	RMCE	BAC
Nuclear autoantigenic sperm protein	6	0,0005	1,61	RMCE	BAC
Proliferation-associated protein 2G4	10	0,0000	1,59	RMCE	BAC
UV excision repair protein RAD23 homolog B	2	0,0002	1,58	RMCE	BAC
BUB3 budding uninhibited by benzimidazoles 3 homolog (yeast), isoform CRA_c	4	0,0001	1,58	RMCE	BAC
Drebrin	6	0,0001	1,57	RMCE	BAC
Eukaryotic translation initiation factor 3 subunit J	3	0,0026	1,57	RMCE	BAC
[2,3,9]Heat shock protein HSP 90-beta-like	34	0,0001	1,56	RMCE	BAC
GMP synthase [glutamine-hydrolyzing]	3	0,0000	1,55	RMCE	BAC
Large proline-rich protein BAT3	2	0,0002	1,53	RMCE	BAC

[1-10] also mentioned in: [1]Alete et al., 2005; [2]Baik et al., 2006; [3]Baik et al., 2008; [4]Meleady et al., 2008; [5]Meleady et al., 2011; [6]Nissom et al., 2006; [7]Seth et al., 2007; [8]Van Dyk et al., 2003; [9]Wingens et al., 2015; [10]Yee et al., 2008

Supplemental Table 3: Transgene delivery comparison: 3D6-scFv-Fc RMCE (n=6) versus Plasmid (n=6) versus BAC (n=6) samples. Identified differentially expressed proteins (n=212) between RMCE versus Plasmid versus BAC correlating to qP and μ (highest BAC and lowest RMCE or vice versa); number of peptides used for quantitation ≥ 2; Anova p value ≤ 0.05; fold change **≥ 1.5x**. PCA of features used for identification are shown in supplemental figure 2, C. Table is sorted by fold-change and positive or negative correlation with qP.

Description	Peptides used for quant.	Anova (p)	**Fold change**	Highest mean	Lowest mean
Leukocyte elastase inhibitor A	8	0,0000	18,04	BAC	RMCE
Pantetheinase	3	0,0000	6,84	BAC	RMCE
Hypothetical protein LOC100752925	2	0,0000	5,54	BAC	RMCE
Cathepsin B	6	0,0000	5,14	BAC	RMCE
[2,4,6]Galectin-1	8	0,0000	4,81	BAC	RMCE
Myoferlin	2	0,0000	4,58	BAC	RMCE
Hypothetical protein LOC100752970	3	0,0000	4,25	BAC	RMCE
Sulfide:quinone oxidoreductase, mitochondrial	2	0,0000	4,21	BAC	RMCE
[4]Macrophage-capping protein	5	0,0000	4,17	BAC	RMCE
Cysteine-rich with EGF-like domain protein 2	3	0,0000	4,16	BAC	RMCE
[10]Trifunctional enzyme subunit beta, mitochondrial	5	0,0000	4,02	BAC	RMCE
Acyl-CoA synthetase family member 2, mitochondrial	2	0,0000	3,76	BAC	RMCE
NADH-cytochrome b5 reductase 1	2	0,0000	3,65	BAC	RMCE
UTP--glucose-1-phosphate uridylyltransferase	3	0,0000	3,59	BAC	RMCE
[10]Trifunctional enzyme subunit alpha, mitochondrial	11	0,0000	3,51	BAC	RMCE
Protein-glutamine gamma-glutamyltransferase 2	3	0,0000	3,50	BAC	RMCE
Delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial	2	0,0000	3,45	BAC	RMCE
Acyl-CoA-binding protein	2	0,0001	3,10	BAC	RMCE
Epididymal secretory protein E1	2	0,0001	2,90	BAC	RMCE
LIM domain and actin-binding protein 1	5	0,0000	2,83	BAC	RMCE
[10]LDLR chaperone MESD	4	0,0007	2,78	BAC	RMCE
Cysteine and glycine-rich protein 1	5	0,0000	2,77	BAC	RMCE
Plectin	3	0,0004	2,74	BAC	RMCE
60S acidic ribosomal protein P2	2	0,0233	2,74	BAC	RMCE
Cathepsin Z	9	0,0000	2,72	BAC	RMCE
Major vault protein	4	0,0000	2,69	BAC	RMCE
182 kDa tankyrase-1-binding protein	6	0,0000	2,60	BAC	RMCE
V-type proton ATPase subunit B, brain isoform	2	0,0000	2,57	BAC	RMCE
Protein canopy-like 3	2	0,0000	2,57	BAC	RMCE
Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform	3	0,0000	2,55	BAC	RMCE

Annexin A11	5	0,0000	2,46	BAC	RMCE
Dipeptidyl peptidase 2	2	0,0026	2,44	BAC	RMCE
Receptor expression-enhancing protein 5	2	0,0000	2,42	BAC	RMCE
Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	2	0,0000	2,38	BAC	RMCE
Mannose-1-phosphate guanyltransferase beta	2	0,0000	2,32	BAC	RMCE
EH domain-containing protein 4	6	0,0010	2,30	BAC	RMCE
Alpha-N-acetylgalactosaminidase	4	0,0007	2,30	BAC	RMCE
Peroxiredoxin-6	5	0,0000	2,30	BAC	RMCE
Epoxide hydrolase 1	7	0,0000	2,30	BAC	RMCE
[5]Annexin A4	3	0,0001	2,27	BAC	RMCE
Pyridoxal-dependent decarboxylase domain-containing protein 1	2	0,0000	2,27	BAC	RMCE
[2,4,6,9]Vimentin	27	0,0006	2,26	BAC	RMCE
[10]Gelsolin	2	0,0014	2,25	BAC	RMCE
Destrin	3	0,0015	2,17	BAC	RMCE
Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12	2	0,0000	2,16	BAC	RMCE
Isoleucyl-tRNA synthetase, mitochondrial	5	0,0000	2,16	BAC	RMCE
Dynactin subunit 1	3	0,0002	2,16	BAC	RMCE
[3]Protein disulfide-isomerase A5	2	0,0016	2,15	BAC	RMCE
Coronin-1B	7	0,0000	2,14	BAC	RMCE
General vesicular transport factor p115	2	0,0000	2,12	BAC	RMCE
Proteasome subunit alpha type-6	2	0,0248	2,10	BAC	RMCE
Coronin-7	2	0,0035	2,10	BAC	RMCE
V-type proton ATPase subunit G 1	2	0,0000	2,10	BAC	RMCE
[4,9]Vinculin isoform 1	5	0,0156	2,09	BAC	RMCE
Legumain	3	0,0002	2,09	BAC	RMCE
Sorbitol dehydrogenase	5	0,0000	2,09	BAC	RMCE
Neutral amino acid transporter A	4	0,0001	2,08	BAC	RMCE
Glycerol-3-phosphate dehydrogenase 1-like protein	2	0,0001	2,07	BAC	RMCE
Cathepsin D	3	0,0004	2,06	BAC	RMCE
Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	4	0,0231	2,03	BAC	RMCE
Proteasome subunit beta type-5	3	0,0000	2,01	BAC	RMCE
Actin-related protein 2/3 complex subunit 1B	3	0,0029	2,01	BAC	RMCE
Thioredoxin domain-containing protein 12	2	0,0000	1,99	BAC	RMCE
Transmembrane protein 43	3	0,0000	1,98	BAC	RMCE
Alpha-2-macroglobulin receptor-associated protein	5	0,0014	1,98	BAC	RMCE
Nucleobindin-2	3	0,0000	1,98	BAC	RMCE
Plastin-3	7	0,0000	1,98	BAC	RMCE
Calcium-binding mitochondrial carrier protein SCaMC-1	3	0,0001	1,95	BAC	RMCE
Ribosome-binding protein 1	11	0,0000	1,94	BAC	RMCE
[2,3]Protein disulfide-isomerase A3	24	0,0000	1,93	BAC	RMCE
3-hydroxyisobutyrate dehydrogenase, mitochondrial	2	0,0000	1,92	BAC	RMCE
[4]Annexin A5	5	0,0002	1,91	BAC	RMCE

[6]Peptidyl-prolyl cis-trans isomerase B	7	0,0000	1,90	BAC	RMCE
Lysosomal alpha-glucosidase	3	0,0003	1,87	BAC	RMCE
Transgelin-2	3	0,0004	1,86	BAC	RMCE
Pyruvate carboxylase, mitochondrial	7	0,0000	1,85	BAC	RMCE
3-keto-steroid reductase	2	0,0006	1,84	BAC	RMCE
Estradiol 17-beta-dehydrogenase 12	3	0,0007	1,84	BAC	RMCE
Rho GTPase-activating protein 1	4	0,0000	1,81	BAC	RMCE
Talin-1 isoform 1	8	0,0000	1,81	BAC	RMCE
S-formylglutathione hydrolase	5	0,0000	1,81	BAC	RMCE
Sulfated glycoprotein 1	12	0,0001	1,80	BAC	RMCE
Alpha-actinin-1	16	0,0000	1,79	BAC	RMCE
Sorcin	2	0,0000	1,78	BAC	RMCE
Septin-11	6	0,0000	1,76	BAC	RMCE
V-type proton ATPase subunit E 1	2	0,0000	1,75	BAC	RMCE
Cation-independent mannose-6-phosphate receptor	4	0,0122	1,75	BAC	RMCE
Calmodulin	4	0,0008	1,75	BAC	RMCE
Dipeptidyl-peptidase 3	3	0,0000	1,74	BAC	RMCE
Vesicle-associated membrane protein-associated protein A	2	0,0006	1,74	BAC	RMCE
HIV Tat-specific factor 1 homolog	4	0,0000	1,73	BAC	RMCE
Mesencephalic astrocyte-derived neurotrophic factor	5	0,0008	1,72	BAC	RMCE
[4]UDP-N-acetylhexosamine pyrophosphorylase-like protein 1	2	0,0009	1,72	BAC	RMCE
Drebrin-like protein	4	0,0000	1,71	BAC	RMCE
Eukaryotic translation initiation factor 4 gamma 1	2	0,0077	1,71	BAC	RMCE
[6,9]Hypoxia up-regulated protein 1	5	0,0018	1,71	BAC	RMCE
[4]Adenylyl cyclase-associated protein 1	8	0,0001	1,70	BAC	RMCE
LIM and SH3 domain protein 1	9	0,0004	1,69	BAC	RMCE
Eukaryotic translation initiation factor 5A-1	2	0,0002	1,68	BAC	RMCE
[9,10]Moesin	10	0,0001	1,68	BAC	RMCE
Methylosome protein 50	2	0,0000	1,67	BAC	RMCE
Ras-related protein Rab-1A	3	0,0000	1,66	BAC	RMCE
[9]Cytosolic non-specific dipeptidase	3	0,0005	1,66	BAC	RMCE
Myosin-9	2	0,0022	1,66	BAC	RMCE
Annexin A6	5	0,0008	1,65	BAC	RMCE
E3 SUMO-protein ligase RanBP2	2	0,0041	1,65	BAC	RMCE
Laminin subunit gamma-1	2	0,0151	1,65	BAC	RMCE
Vesicle-trafficking protein SEC22b	3	0,0006	1,65	BAC	RMCE
Spectrin alpha chain, brain	2	0,0035	1,64	BAC	RMCE
[2,4,7]Protein disulfide-isomerase A6	2	0,0004	1,64	BAC	RMCE
Peptidyl-prolyl cis-trans isomerase FKBP1A	2	0,0340	1,63	BAC	RMCE
[9]Microtubule-associated protein 1B	6	0,0000	1,62	BAC	RMCE
Sialic acid synthase	3	0,0000	1,62	BAC	RMCE
UPF0556 protein C19orf10-like	2	0,0040	1,59	BAC	RMCE
Putative RNA-binding protein Luc7-like 2	2	0,0009	1,59	BAC	RMCE

Golgi reassembly-stacking protein 2	2	0,0017	1,56	BAC	RMCE
Plasminogen activator inhibitor 1 RNA-binding protein	3	0,0095	1,56	BAC	RMCE
Uncharacterized protein C10orf88 homolog	2	0,0433	1,56	BAC	RMCE
Src substrate cortactin	4	0,0000	1,56	BAC	RMCE
COP9 signalosome complex subunit 1	2	0,0011	1,56	BAC	RMCE
Kinesin-1 heavy chain	7	0,0005	1,55	BAC	RMCE
Myosin-10	2	0,0056	1,55	BAC	RMCE
[2,3,5,6]Annexin A1	11	0,0009	1,54	BAC	RMCE
[10]Hydroxymethylglutaryl-CoA synthase, cytoplasmic	5	0,0090	1,54	BAC	RMCE
Ubiquitin-conjugating enzyme E2 L3	2	0,0006	1,53	BAC	RMCE
Protein canopy-like 2	3	0,0000	1,53	BAC	RMCE
Catalase	11	0,0001	1,52	BAC	RMCE
[1]Cofilin-1	5	0,0009	1,52	BAC	RMCE
Cullin-associated NEDD8-dissociated protein 1	2	0,0039	1,52	BAC	RMCE
[4]Aldose reductase-related protein 2	4	0,0000	6,81	RMCE	BAC
Heme oxygenase 1	7	0,0000	4,81	RMCE	BAC
Agrin	2	0,0077	3,82	RMCE	BAC
Proteasome-associated protein ECM29-like	2	0,0000	3,71	RMCE	BAC
Elongation factor 1-delta	2	0,0000	3,23	RMCE	BAC
Heterogeneous nuclear ribonucleoprotein H3	2	0,0000	2,91	RMCE	BAC
Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	3	0,0000	2,87	RMCE	BAC
Basement membrane-specific heparan sulfate proteoglycan core protein	24	0,0000	2,79	RMCE	BAC
Lipoprotein lipase	2	0,0000	2,75	RMCE	BAC
Succinate-semialdehyde dehydrogenase, mitochondrial	2	0,0001	2,42	RMCE	BAC
Transcription intermediary factor 1-beta	3	0,0000	2,33	RMCE	BAC
Lamina-associated polypeptide 2, isoforms alpha/zeta	9	0,0000	2,22	RMCE	BAC
C-4 methylsterol oxidase	2	0,0000	2,16	RMCE	BAC
Prostaglandin F2 receptor negative regulator	2	0,0000	2,16	RMCE	BAC
[10]Inosine-5'-monophosphate dehydrogenase 2	2	0,0004	2,15	RMCE	BAC
[10]Thiosulfate sulfurtransferase	2	0,0000	2,12	RMCE	BAC
Nucleosome-binding protein 1	2	0,0048	2,10	RMCE	BAC
Heterogeneous nuclear ribonucleoprotein A1	3	0,0000	2,04	RMCE	BAC
39S ribosomal protein L12, mitochondrial	3	0,0000	2,03	RMCE	BAC
Exportin-5	2	0,0001	1,97	RMCE	BAC
DnaJ homolog subfamily A member 1	2	0,0006	1,94	RMCE	BAC
Putative ribosomal RNA methyltransferase NOP2	7	0,0009	1,94	RMCE	BAC
Chromobox protein homolog 3	4	0,0000	1,94	RMCE	BAC
Cysteine desulfurase, mitochondrial	3	0,0000	1,94	RMCE	BAC
Putative adenosylhomocysteinase 2	3	0,0000	1,93	RMCE	BAC
Mitochondrial carnitine/acylcarnitine carrier protein	3	0,0004	1,92	RMCE	BAC
Serpin H1	3	0,0006	1,91	RMCE	BAC
Anamorsin	2	0,0000	1,90	RMCE	BAC

Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	13	0,0000	1,83	RMCE	BAC
Proline synthase co-transcribed bacterial homolog protein	2	0,0008	1,81	RMCE	BAC
FK506-binding protein 4	8	0,0000	1,81	RMCE	BAC
Condensin complex subunit 1	2	0,0006	1,81	RMCE	BAC
[7]Myb-binding protein 1A	14	0,0020	1,80	RMCE	BAC
ERO1-like protein alpha	2	0,0014	1,79	RMCE	BAC
Basic leucine zipper and W2 domain-containing protein 2	5	0,0000	1,78	RMCE	BAC
Ubiquitin-like modifier-activating enzyme 1	2	0,0001	1,78	RMCE	BAC
[Protein ADP-ribosylarginine] hydrolase	2	0,0000	1,78	RMCE	BAC
Nucleolar phosphoprotein p130	3	0,0218	1,76	RMCE	BAC
Eukaryotic initiation factor 4A-II	2	0,0004	1,75	RMCE	BAC
[4]Thymidylate synthase	3	0,0133	1,75	RMCE	BAC
Deoxyuridine 5'-triphosphate nucleotidohydrolase	3	0,0124	1,74	RMCE	BAC
[2,5,6]60 kDa heat shock protein, mitochondrial	18	0,0000	1,73	RMCE	BAC
rRNA 2'-O-methyltransferase fibrillarin	4	0,0007	1,73	RMCE	BAC
Inorganic pyrophosphatase 2, mitochondrial	2	0,0001	1,72	RMCE	BAC
[1]Dihydrolipoyl dehydrogenase, mitochondrial	3	0,0000	1,72	RMCE	BAC
ATPase family AAA domain-containing protein 3	3	0,0001	1,71	RMCE	BAC
Intracellular adhesion molecule 1	7	0,0002	1,71	RMCE	BAC
Sjogren syndrome antigen B	2	0,0006	1,70	RMCE	BAC
[4,7]Nucleophosmin	3	0,0004	1,70	RMCE	BAC
[4,5]Calponin-3	5	0,0032	1,69	RMCE	BAC
NHP2-like protein 1	3	0,0002	1,68	RMCE	BAC
High mobility group protein B2	2	0,0082	1,68	RMCE	BAC
Lamin-B1	6	0,0000	1,67	RMCE	BAC
[7]DNA topoisomerase 2-alpha	5	0,0004	1,66	RMCE	BAC
Structural maintenance of chromosomes protein 4	2	0,0048	1,66	RMCE	BAC
Histone H2B type 1-H	2	0,0004	1,65	RMCE	BAC
Ribose-phosphate pyrophosphokinase 1	2	0,0000	1,64	RMCE	BAC
4F2 cell-surface antigen heavy chain	2	0,0002	1,63	RMCE	BAC
Nuclear autoantigenic sperm protein	7	0,0001	1,63	RMCE	BAC
tRNA (cytosine-5-)-methyltransferase NSUN2	4	0,0000	1,61	RMCE	BAC
[6]Annexin A2	8	0,0132	1,61	RMCE	BAC
Heterogeneous nuclear ribonucleoprotein L	4	0,0000	1,60	RMCE	BAC
Heterogeneous nuclear ribonucleoprotein L	5	0,0000	1,60	RMCE	BAC
Protein phosphatase 1G	2	0,0009	1,59	RMCE	BAC
Serine/arginine-rich splicing factor 7	4	0,0018	1,58	RMCE	BAC
Torsin-1B-like	2	0,0010	1,57	RMCE	BAC
Thioredoxin domain-containing protein 5	4	0,0000	1,57	RMCE	BAC
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	2	0,0000	1,56	RMCE	BAC
[6,7]High mobility group protein B1	5	0,0043	1,56	RMCE	BAC
Polyadenylate-binding protein 4	3	0,0004	1,56	RMCE	BAC

U3 small nucleolar RNA-interacting protein 2	2	0,0000	1,55	RMCE	BAC
DnaJ homolog subfamily C member 7	2	0,0019	1,55	RMCE	BAC
Heterogeneous nuclear ribonucleoprotein F	4	0,0000	1,55	RMCE	BAC
Structural maintenance of chromosomes protein 2	3	0,0036	1,55	RMCE	BAC
[10]RuvB-like 2	4	0,0001	1,53	RMCE	BAC
Nucleolin	9	0,0001	1,53	RMCE	BAC
Septin-9	2	0,0004	1,52	RMCE	BAC
[10]ADP-sugar pyrophosphatase	3	0,0000	1,52	RMCE	BAC
[1,7]Voltage-dependent anion-selective channel protein 1	6	0,0000	1,52	RMCE	BAC
GTP:AMP phosphotransferase, mitochondrial	2	0,0002	1,52	RMCE	BAC
Protein RCC2	5	0,0015	1,51	RMCE	BAC
Proliferation-associated protein 2G4	10	0,0007	1,50	RMCE	BAC
Monocarboxylate transporter 1	3	0,0000	1,50	RMCE	BAC

[1-10] also mentioned in: [1]Alete et al., 2005; [2]Baik et al., 2006; [3]Baik et al., 2008; [4]Meleady et al., 2008; [5]Meleady et al., 2011; [6]Nissom et al., 2006; [7]Seth et al., 2007; [8]Van Dyk et al., 2003; [9]Wingens et al., 2015; [10]Yee et al., 2008

Supplemental Table 4: Transgene delivery comparison RMCE (n=6) vs Plasmid (n=6) vs BAC (n=6); same expression pattern for 2F5-scFv-Fc and 3D6-scFv-Fc; Combined identified differential proteins (n=58+1); number of peptides used for quantitation ≥ 2; Anova p Value ≤ 0.05; fold change **≥ 1.5x** in both in-group comparisons. Table is sorted by fold-change and positive or negative correlation with qP.

		3D6			2F5				
Description	**Gene ID**	**Peptides used for quant.**	**Anova (p)**	**Fold change**	**Peptides used for quant.**	**Anova (p)**	**fold change**	**Highest mean**	**Lowest mean**
Leukocyte elastase inhibitor A	Serpinb1a	8	0.000	**18.04**	7	0.000	**5.77**	BAC	RMCE
Ig gamma-1 chain C region	***IGHG1***	7	0.000	***10.13***	11	0.000	***4.62***	BAC	RMCE
Cathepsin B	Ctsb	6	0.000	**5.14**	7	0.000	**3.51**	BAC	RMCE
Galectin-1	Lgals1	8	0.000	**4.81**	9	0.000	**4.91**	BAC	RMCE
Myoferlin	Myof	2	0.000	**4.58**	2	0.000	**2.28**	BAC	RMCE
Macrophage-capping protein	Capg	5	0.000	**4.17**	5	0.000	**2.27**	BAC	RMCE
Cysteine-rich with EGF-like domain protein 2	Creld2	3	0.000	**4.16**	3	0.000	**3.01**	BAC	RMCE
NADH-cytochrome b5 reductase 1	Cyb5r1	2	0.000	**3.65**	2	0.000	**3.76**	BAC	RMCE
LIM domain and actin-binding protein 1	Lima1	5	0.000	**2.83**	4	0.003	**2.10**	BAC	RMCE
Plectin	Plec	3	0.000	**2.74**	2	0.000	**2.77**	BAC	RMCE
V-type proton ATPase subunit B, brain isoform	Atp6v1b2	2	0.000	**2.57**	2	0.001	**1.58**	BAC	RMCE
Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform	Ppp2r2a	3	0.000	**2.55**	3	0.000	**1.89**	BAC	RMCE
Annexin A11	Anxa11	5	0.000	**2.46**	5	0.001	**1.66**	BAC	RMCE
Receptor expression-enhancing protein 5	Reep5	2	0.000	**2.42**	2	0.000	**1.66**	BAC	RMCE
Alpha-N-acetylgalactosaminidase	Naga	4	0.001	**2.30**	4	0.000	**3.20**	BAC	RMCE
Pyridoxal-dependent decarboxylase domain-containing protein 1	Pdxdc1	2	0.000	**2.27**	2	0.000	**2.11**	BAC	RMCE
Vimentin	Vim	27	0.001	**2.26**	24	0.002	**2.31**	BAC	RMCE
Gelsolin	Gsn	2	0.001	**2.25**	3	0.000	**2.57**	BAC	RMCE
Destrin	Dstn	3	0.001	**2.17**	3	0.001	**1.75**	BAC	RMCE
Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12	Gng12	2	0.000	**2.16**	2	0.000	**1.70**	BAC	RMCE
Isoleucyl-tRNA synthetase, mitochondrial	Iars2	5	0.000	**2.16**	5	0.000	**2.41**	BAC	RMCE

Coronin-1B	Coro1b	7	0.000	**2.14**	8	0.000	**2.08**	BAC	RMCE
V-type proton ATPase subunit G 1	Atp6v1g1	2	0.000	**2.10**	2	0.001	**1.58**	BAC	RMCE
Glycerol-3-phosphate dehydrogenase 1-like protein	Gpd1l	2	0.000	**2.07**	2	0.002	**2.17**	BAC	RMCE
Actin-related protein 2/3 complex subunit 1B	Arpc1b	3	0.003	**2.01**	2	0.006	**2.71**	BAC	RMCE
Transmembrane protein 43	Tmem43	3	0.000	**1.98**	2	0.000	**1.96**	BAC	RMCE
Nucleobindin-2	Nucb2	3	0.000	**1.98**	3	0.000	**1.85**	BAC	RMCE
Peptidyl-prolyl cis-trans isomerase B	Ppib	7	0.000	**1.90**	7	0.000	**1.94**	BAC	RMCE
Transgelin-2	Tagln2	3	0.000	**1.86**	4	0.000	**2.15**	BAC	RMCE
Pyruvate carboxylase, mitochondrial	Pcx	7	0.000	**1.85**	2	0.014	**1.95**	BAC	RMCE
3-keto-steroid reductase	Hsd17b7	2	0.001	**1.84**	2	0.000	**1.61**	BAC	RMCE
Estradiol 17-beta-dehydrogenase 12	Hsd17b12	3	0.001	**1.84**	3	0.000	**1.85**	BAC	RMCE
Rho GTPase-activating protein 1	Arhgap1	4	0.000	**1.81**	4	0.000	**2.05**	BAC	RMCE
Sulfated glycoprotein 1	Psap	12	0.000	**1.80**	10	0.000	**1.84**	BAC	RMCE
Septin-11	Sept11	6	0.000	**1.76**	2	0.001	**1.57**	BAC	RMCE
Calmodulin	Calm1	4	0.001	**1.75**	3	0.001	**1.62**	BAC	RMCE
Vesicle-associated membrane protein-associated protein A	Vapa	2	0.001	**1.74**	5	0.011	**1.56**	BAC	RMCE
HIV Tat-specific factor 1 homolog	Htatsf1	4	0.000	**1.73**	2	0.012	**1.97**	BAC	RMCE
LIM and SH3 domain protein 1	Lasp1	9	0.000	**1.69**	8	0.000	**1.66**	BAC	RMCE
Ras-related protein Rab-1A	Rab1	3	0.000	**1.66**	2	0.000	**2.01**	BAC	RMCE
Myosin-9	Myh9	2	0.002	**1.66**	2	0.004	**2.79**	BAC	RMCE
Vesicle-trafficking protein SEC22b	Sec22b	3	0.001	**1.65**	3	0.003	**1.69**	BAC	RMCE
Microtubule-associated protein 1B	Mtap1b	6	0.000	**1.62**	2	0.000	**1.89**	BAC	RMCE
Src substrate cortactin	Cttn	4	0.000	**1.56**	6	0.002	**1.76**	BAC	RMCE
Myosin-10	Myh10	2	0.006	**1.55**	3	0.001	**2.05**	BAC	RMCE
Protein canopy-like 2	Cnpy2	3	0.000	**1.53**	3	0.000	**1.56**	BAC	RMCE
Heme oxygenase 1	Hmox1	7	0.000	**4.81**	7	0.000	**3.41**	RMCE	BAC
Proteasome-associated protein ECM29-like	Ecm29	2	0.000	**3.71**	2	0.003	**2.21**	RMCE	BAC
Lipoprotein lipase	Lpl	2	0.000	**2.75**	2	0.000	**2.99**	RMCE	BAC
Lamina-associated polypeptide 2, isoforms alpha/zeta	Tmpo	9	0.000	**2.22**	8	0.000	**1.89**	RMCE	BAC
Nucleosome-binding protein 1	Hmgn5	2	0.005	**2.10**	3	0.006	**2.80**	RMCE	BAC
Exportin-5	Xpo5	2	0.000	**1.97**	2	0.000	**2.33**	RMCE	BAC
Anamorsin	Ciapin1	2	0.000	**1.90**	3	0.000	**2.31**	RMCE	BAC

Basic leucine zipper and W2 domain-containing protein 2	Bzw2	5	0.000	**1.78**	5	0.000	**1.68**	RMCE	BAC
Intracellular adhesion molecule 1	Icam1	7	0.000	**1.71**	8	0.000	**2.30**	RMCE	BAC
Ribose-phosphate pyrophosphokinase 1	Prps1	2	0.000	**1.64**	3	0.000	**1.87**	RMCE	BAC
Nuclear autoantigenic sperm protein	Nasp	7	0.000	**1.63**	6	0.000	**1.61**	RMCE	BAC
Protein phosphatase 1G	Ppm1g	2	0.001	**1.59**	3	0.000	**1.91**	RMCE	BAC
Proliferation-associated protein 2G4	Pa2g4	10	0.001	**1.50**	10	0.000	**1.59**	RMCE	BAC

Supplemental Table 5: Transgene delivery comparison RMCE (n=6) vs Plasmid (n=6) vs BAC (n=6); 2F5-scFv-Fc identified differential proteins (n=9) correlating to qP or μ that were not identified in the in-group comparison of 3D6-scFv-Fc at all; number of peptides used for quantitation ≥ 2; Anova p Value ≤ 0.05; fold change **≥ 1.5x** in both in-group comparisons. Table is sorted by fold-change and positive or negative correlation with qP.

Description	Gene ID	Fold change	Highest mean	Lowest mean
Heat shock protein beta-1	Hspb1	**2.51**	BAC	RMCE
Translocation protein SEC62	Sec62	**2.01**	BAC	RMCE
MHC class I antigen Hm1-C4	-	**1.66**	BAC	RMCE
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A	Stt3a	**1.57**	BAC	RMCE
Thioredoxin reductase 1, cytoplasmic	Txnrd1	**1.97**	RMCE	BAC
Proliferating cell nuclear antigen	Pcna	**1.71**	RMCE	BAC
Cullin-4B	Cul4b	**1.63**	RMCE	BAC
Glutathione S-transferase P 2	Gstp2	**1.63**	RMCE	BAC
Eukaryotic translation initiation factor 3 subunit J	Eif3j	**1.57**	RMCE	BAC

Supplemental Table 6: Transgene delivery comparison RMCE (n=6) vs Plasmid (n=6) vs BAC (n=6); 3D6 scFv-Fc identified differential proteins (n=32) correlating to qP or μ that were not identified in the in-group comparison of 2F5-scFv-Fc at all; number of peptides used for quantitation ≥ 2; Anova p Value ≤ 0.05; fold change **≥ 1.5x** in both in-group comparisons. Table is sorted by fold-change and positive or negative correlation with qP.

Description	Gene ID	Fold change	Highest mean	Lowest mean
Epididymal secretory protein E1	Npc2	**2.9**	**BAC**	**RMCE**
60S acidic ribosomal protein P2	Rplp2	**2.74**	**BAC**	**RMCE**
Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	Nsdhl	**2.38**	**BAC**	**RMCE**
Mannose-1-phosphate guanyltransferase beta	Gmppb	**2.32**	**BAC**	**RMCE**
Dynactin subunit 1	Dctn1	**2.16**	**BAC**	**RMCE**
Proteasome subunit alpha type-6	Psma6	**2.1**	**BAC**	**RMCE**
Coronin-7	Coro7	**2.1**	**BAC**	**RMCE**

Vinculin isoform 1	Vcl	**2.09**	**BAC**	**RMCE**
Legumain	Lgmn	**2.09**	**BAC**	**RMCE**
Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	Ppp2r1a	**2.03**	**BAC**	**RMCE**
Proteasome subunit beta type-5	Psmb5	**2.01**	**BAC**	**RMCE**
Cation-independent mannose-6-phosphate receptor	Igf2r	**1.75**	**BAC**	**RMCE**
Eukaryotic translation initiation factor 4 gamma 1	Eif4g1	**1.71**	**BAC**	**RMCE**
Peptidyl-prolyl cis-trans isomerase FKBP1A	Fkbp1a	**1.63**	**BAC**	**RMCE**
Putative RNA-binding protein Luc7-like 2	Luc7l2	**1.59**	**BAC**	**RMCE**
Golgi reassembly-stacking protein 2	Gorasp2	**1.56**	**BAC**	**RMCE**
Plasminogen activator inhibitor 1 RNA-binding protein	Serbp1	**1.56**	**BAC**	**RMCE**
Hydroxymethylglutaryl-CoA synthase, cytoplasmic	Hmgcs1	**1.54**	**BAC**	**RMCE**
Ubiquitin-conjugating enzyme E2 L3	Ube2l3	**1.53**	**BAC**	**RMCE**
Agrin	Agrn	**3.82**	**RMCE**	**BAC**
Succinate-semialdehyde dehydrogenase, mitochondrial	Aldh5a1	**2.42**	**RMCE**	**BAC**
C-4 methylsterol oxidase	Sc4mol	**2.16**	**RMCE**	**BAC**
Thiosulfate sulfurtransferase	Tst	**2.12**	**RMCE**	**BAC**
Mitochondrial carnitine/acylcarnitine carrier protein	Slc25a20	**1.92**	**RMCE**	**BAC**
Thymidylate synthase	Tyms	**1.75**	**RMCE**	**BAC**
Structural maintenance of chromosomes protein 4	Smc4	**1.66**	**RMCE**	**BAC**
Histone H2B type 1-H	Hist1h2bh	**1.65**	**RMCE**	**BAC**
tRNA (cytosine-5-)-methyltransferase NSUN2	Nsun2	**1.61**	**RMCE**	**BAC**
Thioredoxin domain-containing protein 5	Txndc5	**1.57**	**RMCE**	**BAC**
High mobility group protein B1	Hmgb1	**1.56**	**RMCE**	**BAC**
Nucleolin	Ncl	**1.53**	**RMC E**	**BAC**
Septin-9	Sept9	**1.52**	**RMCE**	**BAC**

Supplemental Table 7: Transgene delivery comparison RMCE (n=6) vs Plasmid (n=6) vs BAC (n=6); differential expressed proteins correlating with qP or μ showing the opposite expression pattern for 3D6-scFv-Fc and 2F5-scFv-Fc producers (n=2) (number of peptides used for quantitation ≥ 2; Anova p Value ≤ 0.05; fold change **≥ 1.5x** in at least one of the two in-group comparisons).

		3D6					2F5				
Description	**Gene ID**	**Peptides used for quant.**	**Anova (p)**	**Max fold change**	**max**	**min**	**Peptides used for quant.**	**Anova (p)**	**Max fold change**	**max**	**min**
Cullin-associated NEDD8-dissociated protein 1	Cand1	**2**	**0.004**	**1.52**	BAC	RMCE	3	0.019	1.22	RMCE	BAC
Protein SON	Son	**3**	**0.036**	**1.27**	BAC	RMCE	2	0.001	1.79	RMCE	BAC

Gene Enrichment Analysis

Supplemental Table 8: GeneCodis3 output: GOSlim process singular enrichment analysis of gene list (n=60) of proteins identified to be differential between 2F5- and 3D6-scFv-Fc clones (transgene comparison).

Items	Details	Support	List size	Reference support	Reference size	hyperg. pValue	Genes
GO:0008283	cell proliferation (BP)	3	60	159	37681	0.002	Mki67,Prdx1,Txnrd1
GO:0006457	protein folding (BP)	6	60	117	37681	0.000	Grpel1,Hspe1,Hspd1,Hsp90b1, Calr,Dnajc7
GO:0030198	extracellular matrix organization (BP)	2	60	85	37681	0.008	Lgals3,Hspg2

Supplemental Table 9: GeneCodis3 output: GOSlim process singular enrichment of gene list (n=58) of proteins that showed the same expression profile in both transgene delivery comparisons. The protein Ecm29 was not identified by GeneCodis3 and is therefore not part of this enrichment analysis.

Items	Details	Support	List size	Reference support	Reference size	hyperg. pValue	Genes
GO:0007049	cell cycle (BP)	5	57	528	37681	0.001	Ppm1g,Calm1,Nasp, Sept11,Anxa11
GO:0016192	vesicle-mediated transport (BP)	3	57	168	37681	0.002	Gsn,Rab1,Sec22b
GO:0008219	cell death (BP)	2	57	50	37681	0.003	Vapa,Hmox1
GO:0006810	transport (BP)	7	57	1615	37681	0.011	Xpo5,Lasp1,Atp6v1g1,Nasp, Atp6v1b2,Rab1,Sec22b

References

Abdelmohsen K, Tominaga K, Lee EK, Srikantan S, Kang MJ, Kim MM, Selimyan R, Martindale JL, Yang X, Carrier F and others. 2011. Enhanced translation by Nucleolin via G-rich elements in coding and non-coding regions of target mRNAs. Nucleic Acids Res 39(19):8513-30.

Alete DE, Racher AJ, Birch JR, Stansfield SH, James DC, Smales CM. 2005. Proteomic analysis of enriched microsomal fractions from GS-NS0 murine myeloma cells with varying secreted recombinant monoclonal antibody productivities. Proteomics 5(18):4689-704.

Angelov D, Bondarenko VA, Almagro S, Menoni H, Mongelard F, Hans F, Mietton F, Studitsky VM, Hamiche A, Dimitrov S. 2006. Nucleolin is a histone chaperone with FACT-like activity and assists remodeling of nucleosomes. The EMBO Journal 25(8):1669-1679.

Baake M, Bäuerle M, Doenecke D, Albig W. 2001. Core histones and linker histones are imported into the nucleus by different pathways. European Journal of Cell Biology 80(11):669-677.

Baik JY, Joo EJ, Kim YH, Lee GM. 2008. Limitations to the comparative proteomic analysis of thrombopoietin producing Chinese hamster ovary cells treated with sodium butyrate. J Biotechnol 133(4):461-8.

Baik JY, Lee MS, An SR, Yoon SK, Joo EJ, Kim YH, Park HW, Lee GM. 2006. Initial transcriptome and proteome analyses of low culture temperature-induced expression in CHO cells producing erythropoietin. Biotechnol Bioeng 93(2):361-71.

Dumic J, Dabelic S, Flögel M. 2006. Galectin-3: an open-ended story. Biochimica et Biophysica Acta (BBA)-General Subjects 1760(4):616-635.

Ginisty H, Amalric F, Bouvet P. 1998. Nucleolin functions in the first step of ribosomal RNA processing. The EMBO Journal 17(5):1476-1486.

Haudek KC, Spronk KJ, Voss PG, Patterson RJ, Wang JL, Arnoys EJ. 2010. Dynamics of galectin-3 in the nucleus and cytoplasm. Biochimica et Biophysica Acta (BBA)-General Subjects 1800(2):181-189.

Jäkel S, Görlich D. 1998. Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. EMBO J 17(15):4491-502.

Kirkman HN, Gaetani GF. 2007. Mammalian catalase: a venerable enzyme with new mysteries. Trends in Biochemical Sciences 32(1):44-50.

Liu H, Ling Y, Gong Y, Sun Y, Hou L, Zhang B. 2007. DNA damage induces N-acetyltransferase NAT10 gene expression through transcriptional activation. Molecular and Cellular Biochemistry 300(1-2):249-258.

Lv J, Liu H, Wang Q, Tang Z, Hou L, Zhang B. 2003. Molecular cloning of a novel human gene encoding histone acetyltransferase-like protein involved in transcriptional activation of hTERT. Biochemical and Biophysical Research Communications 311(2):506-513.

Malhotra JD, Kaufman RJ. 2007. The endoplasmic reticulum and the unfolded protein response. Seminars in Cell & Developmental Biology 18(6):716-731.

Mebratu Y, Tesfaigzi Y. 2009. How ERK1/2 activation controls cell proliferation and cell death: Is subcellular localization the answer? Cell Cycle 8(8):1168-75.

Meleady P, Doolan P, Henry M, Barron N, Keenan J, O'Sullivan F, Clarke C, Gammell P, Melville MW, Leonard M and others. 2011. Sustained productivity in recombinant Chinese hamster ovary (CHO) cell lines: proteome analysis of the molecular basis for a process-related phenotype. BMC Biotechnol 11:78.

Meleady P, Henry M, Gammell P, Doolan P, Sinacore M, Melville M, Francullo L, Leonard M, Charlebois T, Clynes M. 2008. Proteomic profiling of CHO cells with enhanced rhBMP-2 productivity following co-expression of PACEsol. Proteomics 8(13):2611-24.

Nissom PM, Sanny A, Kok YJ, Hiang YT, Chuah SH, Shing TK, Lee YY, Wong KT, Hu WS, Sim MY and others. 2006. Transcriptome and proteome profiling to understanding the biology of high productivity CHO cells. Mol Biotechnol 34(2):125-40.

Ong YS, Tang BL, Loo LS, Hong W. 2010. p125A exists as part of the mammalian Sec13/Sec31 COPII subcomplex to facilitate ER-Golgi transport. J Cell Biol 190(3):331-45.

Rhee SG, Chang T-S, Bae YS, Lee S-R, Kang SW. 2003. Cellular Regulation by Hydrogen Peroxide. Journal of the American Society of Nephrology 14(suppl 3):S211-S215.

Rossi MR, Somji S, Garrett SH, Sens MA, Nath J, Sens DA. 2002. Expression of hsp 27, hsp 60, hsc 70, and hsp 70 stress response genes in cultured human urothelial cells (UROtsa) exposed to lethal and sublethal concentrations of sodium arsenite. Environmental health perspectives 110(12):1225.

Seth G, Philp RJ, Lau A, Jiun KY, Yap M, Hu WS. 2007. Molecular portrait of high productivity in recombinant NS0 cells. Biotechnol Bioeng 97(4):933-51.

Shimoi W, Ezawa I, Nakamoto K, Uesaki S, Gabreski G, Aridor M, Yamamoto A, Nagahama M, Tagaya M, Tani K. 2005. p125 is localized in endoplasmic reticulum exit sites and involved in their organization. J Biol Chem 280(11):10141-8.

Smeenk G, Wiegant WW, Vrolijk H, Solari AP, Pastink A, van Attikum H. 2010. The NuRD chromatin–remodeling complex regulates signaling and repair of DNA damage. The Journal of Cell Biology 190(5):741-749.

Snyder M, He W, Zhang JJ. 2005. The DNA replication factor MCM5 is essential for Stat1-mediated transcriptional activation. Proceedings of the National Academy of Sciences of the United States of America 102(41):14539-14544.

Tew KD, Manevich Y, Grek C, Xiong Y, Uys J, Townsend DM. 2011. The role of glutathione S-transferase P in signaling pathways and S-glutathionylation in cancer. Free Radical Biology and Medicine 51(2):299-313.

Van Dyk DD, Misztal DR, Wilkins MR, Mackintosh JA, Poljak A, Varnai JC, Teber E, Walsh BJ, Gray PP. 2003. Identification of cellular changes associated with increased production of human growth hormone in a recombinant Chinese hamster ovary cell line. Proteomics 3(2):147-56.

Wingens M, Gätgens J, Schmidt A, Albaum SP, Büntemeyer H, Noll T, Hoffrogge R. 2015. 2D-DIGE screening of high-productive CHO cells under glucose limitation--basic changes in the proteome equipment and hints for epigenetic effects. J Biotechnol 201:86-97.

Xue Y, Wong J, Moreno GT, Young MK, Côté J, Wang W. 1998. NURD, a Novel Complex with Both ATP-Dependent Chromatin-Remodeling and Histone Deacetylase Activities. Molecular Cell 2(6):851-861.

Yee JC, de Leon Gatti M, Philp RJ, Yap M, Hu WS. 2008. Genomic and proteomic exploration of CHO and hybridoma cells under sodium butyrate treatment. Biotechnol Bioeng 99(5):1186-204.

8 CURRICULUM VITAE

Patrick Mayrhofer, M.Sc.
Kaasgrabengasse 22A/10– A-1190 Vienna – phone 0043/664 73916263 – mail:
mayrhofer.patrick@gmail.com

EDUCATION

International PhD Program - Biomolecular Technology of Proteins (BioToP)
(BOKU, University of Natural Resources and Life Sciences Vienna)
03.2013 – present

Master Study – Biotechnology (BOKU, University of Natural Resources and Life Sciences Vienna)
02.2011 – 02.2013
academic degree: Dip.-Ing. (M.Sc)

Study Abroad (University of Wisconsin-Madison)
08.2012 – 12.2012

Bachelor Study – Food- and Biotechnology (BOKU, University of Natural Resources and Life Sciences Vienna)
10.2007 – 02.2011
academic degree: Bakk. Techn. (B.Sc)

RESEARCH EXPERIENCE

Antibody humanization guided by molecular dynamics (MD) simulations
03.2013 – present, Supervisor: Renate Kunert, Co-supervisor: Chris Oostenbrink
- Transient Antibody Expression in CHO and HEK cell lines
- protein purification
- Protein binding studies (ELISA, Biolayer interferometry)
- biophysical characterization (DSC, ECD)

Research visit – Instituto de Biologia Experimental e Tecnológica (iBET), Oeiras, Portugal
02.2016 – 04.2015, Supervisor: Paula M. Alves, Ana Teixeira
- isotopically non-stationary metabolic flux analysis (INST-MFA)
- GC/MS analyses

Research visit – La Jolla Institute for Allergy and Immunology (LIAI), CA, USA
01.2015 – 04.2015, Supervisor: Alessandro Sette
- in-vitro and in-silico immunological studies of proteins
- training within animal care facility (mice)

Influence of amino acid sequence on growth and secretion of recombinant CHO cell lines
01.2012 – present, Supervisor: Renate Kunert, Co-supervisor: Alexander Mader

- stable cell line development
- recombinant protein expression
- molecular cloning
- gene targeting by recombinase mediated cassette exchange (RMCE)

NSF Research Experiences for Undergraduates (REU) – Georgia Institute of Technology, GA, USA
05.2010 – 08.2010, Supervisor: Nicholas Hud, Co-supervisor: Irena Mamajanov
- nucleic acid assembly
- mass spectrometry and electronic circular dichroism (ECD)

Synthesis of Francisella Lipid A substructures
09.2009-11.2009, Supervisor: Paul Kosma, Alla Zamyatina, Co-supervisor: David Baum
- synthesis of carbohydrate moieties

SUPERVISION EXPERIENCE

Bachelor theses and diploma-theses of students

PRESENTATIONS AND POSTERS

Mayrhofer, P; Mader, A; Kratzer, B; Reinhart, D; Steinfellner, W; Sommeregger, W; Kunert, R (2015). Identification of bottlenecks in antibody expression using targeted gene integration. [Poster] [24th European Society for Animal Cell Technology (ESACT) Meeting: C2P2: Cells, Culture, Patients, Products, Barcelona, Spain, MAY 31-JUNE 3, 2015], BMC Proceedings, 9, P7-P7

Mayrhofer, P; Mader, A; Kunert, R (2014).
RMCE reference sites - A valuable tool for comparing antibody expression capabilities in CHO cells. [POSTER] NEW BIOTECHNOL. 2014; 31: S186-S187.

Mayrhofer, P; Mader, A; Kunert, R (2014)
Versatile RMCE Integration of Different Antibody Variants into Predictable Reference Sites in CHO Cells. [Poster]
PEGS - Protein and Antibody Engineering Summit , MAY 5-9, 2014, Boston, USA

Mayrhofer, P; Mader, A; Margreitter, C; Oostenbrink, C; Kunert, R (2014)
Transient gene expression for evaluation of antibody humanization by molecular dynamics simulation. [Talk]
Human Antibodies And Hybridomas 2014, MAR 31- APR 2, 2014, Vienna, AUSTRIA

Scientific Publications

Mayrhofer, P., and Kunert, R. (2016). Cloning of single-chain antibody variants by overlap-extension PCR for evaluation of antibody expression in transient gene expression. Methods Mol. Biol. (under review).

Sommeregger, W., Mayrhofer, P., Steinfellner, W., Reinhart, D., Henry, M., Clynes, M., Meleady, P., and Kunert, R. (2016). Proteomic differences in recombinant CHO cells producing two similar antibody fragments. Biotechnol. Bioeng. doi:10.1002/bit.25957.

Margreitter, C.*, Mayrhofer, P.*, Kunert, R., and Oostenbrink, C. (2016). Antibody humanization by molecular dynamics simulations-in-silico guided selection of critical backmutations. J. Mol. Recognit. JMR 29, 266–275.

* authors contributed equally to this work

Mayrhofer, P; Kratzer, B; Sommeregger, W; Steinfellner, W; Reinhart, D; Mader, A; Turan, S; Qiao, J; Bode, J; Kunert, R (2014).
Accurate comparison of antibody expression levels by reproducible transgene targeting in engineered recombination-competent CHO cells.
Appl Microbiol Biotechnol. 2014; 98(23):9723-9733