UNIVERSITY OF TURKU

UNVEILING THE SECRETS OF DNA:

Improved expression and phage display efficiency of synthetic recombinant binding proteins in *E. coli* through modulation of codon usage

Antti Kulmala

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. AI OSA – TOM. 624 | ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA | TURKU 2020

UNIVERSITY OF TURKU

UNVEILING THE SECRETS OF DNA:

Improved expression and phage display efficiency of synthetic recombinant binding proteins in *E. coli* through modulation of codon usage

Antti Kulmala

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA - SER. AI OSA – TOM. 624 | ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA | TURKU 2020

University of Turku

Faculty of Science and Engineering
Department of Biochemistry
Molecular Biotechnology and Diagnostics
Doctoral Programme in Molecular Life Sciences

Supervised by

Associate Professor, Urpo Lamminmäki, PhD
Department of Biochemistry
Molecular Biotechnology and Diagnostics
University of Turku
Turku, Finland

University Teacher, Tuomas Huovinen, PhD
Department of Biochemistry
Molecular Biotechnology and Diagnostics
University of Turku
Turku, Finland

Reviewed by

Researcher, Johan Nilvebrant, PhD
Division of Protein Engineering
KTH Royal Institute of Technology
Stockholm, Sweden

Senior Researcher, Tommi Kajander, PhD
Institute of Biotechnology
University of Helsinki
Helsinki, Finland

Opponent

Associate Professor, Vesa Hytönen, PhD
Faculty of Medicine and Health Technology
Tampere University
Tampere, Finland

ISBN 978-951-29-8064-2 (PRINT)
ISBN 978-951-29-8065-9 (PDF)
ISSN 0082-7002 (Print)
ISSN 2343-3175 (Online)
Painosalama Oy, Turku, Finland 2020

To all the people who supported me during this project

UNIVERSITY OF TURKU
Faculty of Science and Engineering
Department of Biochemistry
Molecular Biotechnology and Diagnostics
ANTTI KULMALA: Unveiling the secrets of DNA: Improved expression and phage display efficiency of synthetic recombinant binding proteins in *E. coli* through modulation of codon usage
Doctoral Dissertation, 148 pp.
Doctoral Programme in Molecular Life Sciences
April 2020

ABSTRACT

Recombinant binding proteins are becoming increasingly important in various applications, including diagnostics, basic research, and therapeutics. Therefore, the demand for recombinant binding proteins will increase in the future, making it essential to enhance protein production methods. One approach is to improve heterologous expression, in which proteins are produced outside their native hosts. *Escherichia coli* has been a workhorse of heterologous expression for decades due to easy cultivation, cost efficiency, well-known genetics, and compatibility with phage display. However, the production of heterologous proteins in *E. coli* can sometimes be very difficult. Often the problems in heterologous expression are related to codon usage, which works as a control mechanism of protein translation, especially in bacteria. Different organisms do not use codons in the same manner, and in some cases, a heterologously expressed gene can include codons that are rarely or too frequently used in *E. coli*, which can disturb the codon usage derived control of translation. Many studies have reported improved yields of heterologous proteins produced in *E.coli* when the codon usage of the heterologous gene has been recoded to suit better for *E. coli*. Most often, recombinant binding proteins used in the previously mentioned applications are not native proteins of *E. coli* either, and many times their production or phage display in *E. coli* can be very cumbersome. In this thesis, we aimed to improve expression and phage display properties of two essential recombinant binding proteins in *E. coli* through modulation of codon usage. One binding protein was an antibody fragment called fragment antigen binding (Fab) and another artificial recombinant binding protein called Designed Ankyrin Repeat Protein (DARPin). In the first publication, the expression and phage display efficiency of a human anti-digoxigenin Fab fragment was improved by codon harmonizing the selected segments of the Fab fragment gene. In the second publication, expression and secretion properties of Fab fragments were improved by selecting enhanced signal sequence variants from PelB signal sequence libraries, which included only codon changes. In the third publication, the signal sequence libraries were used to enhance Sec dependent phage display of an anti-GFP DARPin and a DARPin library.

KEYWORDS: Codon usage, protein expression, *E. coli*, Fab fragment, DARPin

TURUN YLIOPISTO
Luonnontieteiden ja tekniikan tiedekunta
Biokemian laitos
Molekulaarinen biotekniikka ja diagnostiikka
ANTTI KULMALA: DNA:n salaisuuksien paljastaminen: Synteettisten rekombinanttisitojaproteiinien tuoton ja faaginäyttötehokkuuden parantaminen *E. colissa* kodonikäytön muokkauksen avulla
Väitöskirja, 148 s.
Molekulaaristen biotieteiden tohtoriohjelma
Huhtikuu 2020

TIIVISTELMÄ

Rekombinanttisitojaproteiineista on tulossa yhä tärkeämpiä diagnostiikassa, perustutkimuksessa ja lääkinnällisissä sovelluksissa. Tästä syystä rekombinanttisitojaproteiinien kysyntä tulee kasvamaan tulevaisuudessa, minkä vuoksi erilaisten proteiinituottomenetelmien kehittäminen on olennaisen tärkeää. Yksi lähestymistapa on kehittää heterologista ekspressiota, jossa proteiineja tuotetaan luonnollisen isännän ulkopuolella. *Escherichia coli* on vuosikymmeniä toiminut heterologisen proteiinituoton työhevosena, johtuen helpoista ja halvoista kasvatusmenetelmistä, hyvin tunnetusta genetiikasta sekä yhteensopivuudesta faaginäyttötekniikan kanssa. Proteiinien tuottaminen luonnollisen isännän ulkopuolella voi kuitenkin toisinaan olla hyvin vaikeaa. Heterologisen proteiinituoton ongelmat liittyvät monesti kodonikäyttöön, joka toimii proteiinien translaation kontrollimekanismina erityisesti bakteereissa. Eri organismit eivät käytä kodoneja samalla tavalla, ja joissakin tapauksissa heterologisesti ilmennetty geeni voi sisältää kodoneja, joita käytetään harvoin tai liian usein *E. colissa*, mikä puolestaan voi häiritä kodonikäyttöpohjaista proteiinitranslaation kontrollointia. Monet tutkimukset ovat raportoineet heterologisten proteiinien tuottotasojen kasvaneen, kun heterologisesti ilmennetyn geenin kodonikäyttöä on muokattu *E. colille* sopivaksi. Rekombinanttisitojaproteiinit, joita käytetään yllä mainituissa sovelluksissa, eivät useimmiten ole *E. colin* luonnollisia proteiineja. Tästä syystä niiden tuottaminen ja faaginäyttö *E. colissa* voi olla hyvin vaivalloista. Tämän väitöstutkimuksen tarkoituksena oli parantaa kahden rekombinanttisitojaproteiinin proteiinituottoa ja faaginäyttöä *E. colissa* kodonikäytön muokkauksen avulla. Yksi sitojaproteiini oli vasta-aineen osa nimeltään ”anigeeniä sitova osa” (Fab-osa) ja toinen keinotekoinen rekombinanttisitojaproteiini nimeltään ”suunniteltu ankyriinitoistoproteiini” (DARPin). Ensimmäisessä julkaisussa ihmisen digoxigeniiniä sitovan Fab-osan tuottoa ja faaginäyttöä parannettiin harmonisoimalla Fab-osan geenin valikoitujen alueiden kodonikäyttö. Toisessa ja kolmannessa julkaisussa Fab-osien ja DARPiinien tuotto- sekä eritysominaisuuksia parannettiin valikoimalla parantuneita signaalisekvenssivariantteja PelB signaalisekvenssikirjastoista, jotka sisälsivät vain kodonimuutoksia.

ASIASANAT: Kodonikäyttö, proteiinituotto, *E. coli*, Fab-osa, DARPin

Table of Contents

Abbreviations

AB	Assay buffer
AMEF	Antibody-mediated enzyme formation
ATP	Adenosine triphosphate
BLA	Biologics License Application
CAI	Codon Adaptation Index
CDR	Complementary-determining region
CL, CH	Constant light, Constant heavy
cytFkpA	Cytoplasmic FkpA
Cκ	Constant kappa
dAbs	Domain antibodies
DARPin	Designed Ankyrin Repeat Protein
DARTs	Dual affinity retargeting proteins
DNA	Deoxyribonucleic acid
DnaJ	Cytoplasmic chaperone
DnaK	Cytoplasmic chaperone
Dsb	Disulfide forming protein
EMA	European Medicines Agency
Erv1p	Sulfhydryl oxidase
Fab	Fragment antigen binding
FACS	Fluorescence-activated cell sorting
Fc	Fragment crystallizable
FcγR	Fragment crystallizable gamma receptor
FDA	Food and Drug Administration
FkpA	Periplasmic chaperone
FW2	Framework 2
GFP	Green fluorescent protein
Gor	Glutathione oxidoreductase
GroEL/ES	Cytoplasmic chaperones
GrpE	Cytoplasmic chaperone
HCAbs	Heavy chain antibodies
HlyA	α-haemolysin

HuCAL	Human Combinatorial Antibody Library
HV2 and 4	Hypervariable loops 2 and 4
Ig	Immunoglobulin
IPTG	isopropyl-β-D-thiogalactopyranoside
kcal/mole	kilocalories per mole
kDa	Kilodalton
Lac	Lactose
Leu-5	Leucine 5
MAA	Marketing Authorization Application
MBP	Maltose-binding protein
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
NUPACK	Nucleic Acid Package
OCCA	Oligonucleotide-directed chelate complementation assay
OD	Optical density
ORF	Open reading frame
PCR	Polymerase chain reaction
PDI	Protein disulfide isomerase
PEG/NaCl	Polyethylene glycol/sodium chloride
PelB	Pectate lyase B signal peptide
phoA	Alkaline phosphatase
pI	isoelectric point
PPIase	Peptidyl prolyl *cis-trans* isomerase
PTM	Post-translational modification
RpoD	Global sigma factor
SB	Super broth
ScAb	Single-chain antibody
ScFv	Single-chain variable fragment
SD	Shine-Dalgarno
SecB	Chaperone involved in Sec dependent translocation
Skp	Periplasmic chaperone
SOC	Super optimal broth with catabolite repression
sRCA	Selective rolling circle amplification
SRP	Signal recognition particle
tAI	tRNA Adaptation Index
TAT	Twin-arginine translocation
TNF-α	Tumor necrosis factor alpha
tRNA	Transfer ribonucleic acid
TrxB	Thioredoxin oxidoreductase
TSS	Translation start site

UDG	Uracil DNA glycosylase
VL, VH	Variable light, Variable heavy
VNAR	Variable domain of New Antigen Receptor
V_κ	Variable kappa
V_λ	Variable lambda
WC	Watson-Crick

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Kulmala A, Huovinen T and Lamminmäki U. Effect of DNA sequence of Fab fragment on yield characteristics and cell growth of *E. coli*. *Sci Rep*, 2017; 7: doi:10.1038/s41598-017-03957-6.

II Kulmala A, Huovinen T and Lamminmäki U. Improvement of Fab expression by screening combinatorial synonymous signal sequence libraries. *Microb Cell Fact*, 2019; 18: doi:10.1186/s12934-019-1210-1.

III Kulmala A, Lappalainen M, Lamminmäki U and Huovinen T. Synonymous codon mutations in post-translational signal peptide increase phage display efficiency of DARPins to the same level as achieved with co-translational translocation. *Manuscript*

1 Introduction

Efficient protein expression is essential when producing different proteins for various purposes, and one of the cornerstones of contemporary biotechnology is the ability to produce proteins in organisms that do not normally produce them. This procedure is called heterologous expression [1,2]. *Escherichia coli* is a popular host for heterologous expression due to the cost efficiency, easy cultivation and compatibility with phage display, which is an essential part of many directed evolution experiments [3,4]. Nevertheless, heterologous expression campaigns implemented by using *E. coli* often face hurdles like low expression or issues with solubility *et cetera*. In recent years, researchers have discovered that codon usage works as an underlying control mechanism in different phases of the protein expression by controlling, for example, translation speed and mRNA secondary structure formation [5]. Different species do not use codons in the same manner, but the identities of preferred and avoided codons vary between species, sometimes significantly [6]. This discrepancy in codon usage between species can disturb the control mechanism, and thus cause problems, especially in heterologous expression [7]. The effect of codon usage on translation is particularly pronounced in prokaryotes [8], and in many cases expression of mammalian proteins has proven to be very challenging without recoding the genes to suit better for *E. coli* [2].

One important group of mammalian proteins expressed in *E. coli* is recombinant antibodies, which are widely used in different applications, including diagnostics, basic research, and therapeutics. In *E. coli*, almost all recombinant antibodies are expressed as antibody fragments to facilitate the production and to ensure compatibility with phage display [3,4]. Furthermore, utilization of recombinant antibody fragments can be highly beneficial in some applications like diagnostics and imaging [3,9,10]. Among recombinant antibody fragments, Fab fragments (fragment antigen binding) are especially attractive; they are more stable compared to single-chain fragments and can be more easily converted to full-length antibodies [11,12]. However, due to different reasons, the expression levels of Fab fragments are many times quite low [13–16]. In addition to the recombinant antibodies, binders based on non-immunoglobulin scaffolds have gotten more attention, because they are simpler than recombinant antibodies, and easier to

express in *E. coli*. However, also non-immunoglobulin based scaffolds have putative shortcomings when it comes to clinical applications and overthrowing recombinant antibodies [17,18]. Furthermore, although many binders based on non-immunoglobulin scaffolds are highly expressing, there can be problems regarding phage display, for example, when utilizing commonly used high-capacity post-translational translocation pathway Sec [19]. This thesis aimed to improve the expression of a synthetic human Fab fragment and display efficiency of a class of non-immunoglobulin scaffold based binder called DARPins (Designed Ankyrin Repeat Proteins) in *Escherichia coli* through modulation of codon usage.

2 Review of literature

This literature review is composed of three main chapters, which connect the thesis to a broader context and explain and cover the topics of the thesis. The first chapter briefly introduces different recombinant binding protein formats, while keeping emphasis mainly on clinically relevant or promising binders. The second chapter describes how the modulation of codon usage – the central theme of the thesis – affects protein expression in *E. coli*. Finally, the improvement of recombinant antibody expression is an extensively studied topic, and hence the third chapter reviews different approaches utilized to improve the expression of recombinant antibodies in *E. coli*.

2.1 Recombinant binding protein formats

2.1.1 Binders based on typical immunoglobulin scaffold

Immunoglobulins, which are key players of the natural immune system, can be divided into five different classes according to the type of heavy chain. The immunoglobulin classes are IgM, IgE, IgD, IgA and IgG, of which IgG class is the most abundant in human serum and also almost exclusively used in therapeutic applications as full-length antibodies [10,20,21]. Full-length antibodies based on IgG are large (~150 kDa) multidomain proteins composed of two light and heavy chains, which can be further divided into variable light domain (V_L), constant light domain (C_L), variable heavy domain (V_H) and three constant heavy domains (C_H1, C_H2 and C_H3) (**Figure 1**). Six regions called complementary-determining regions (CDR), which are evenly distributed between the V_H (CDRH1, 2 and 3) and V_L (CDRL1, 2 and 3), are responsible for binding of the antigen [17,21]. The modular nature of the antibodies makes it possible to generate smaller fragments from the antibodies, and these fragments can be modified even further. The fragments can be produced, for example, by the means of genetic engineering or enzymatic treatment. The most common antibody fragments based on typical immunoglobulin scaffold are single-chain variable fragment (ScFv), fragment antigen binding (Fab)

and domain antibodies (dAbs) [22]. The antibody fragment formats and their derivatives are illustrated in **Figure 1**.

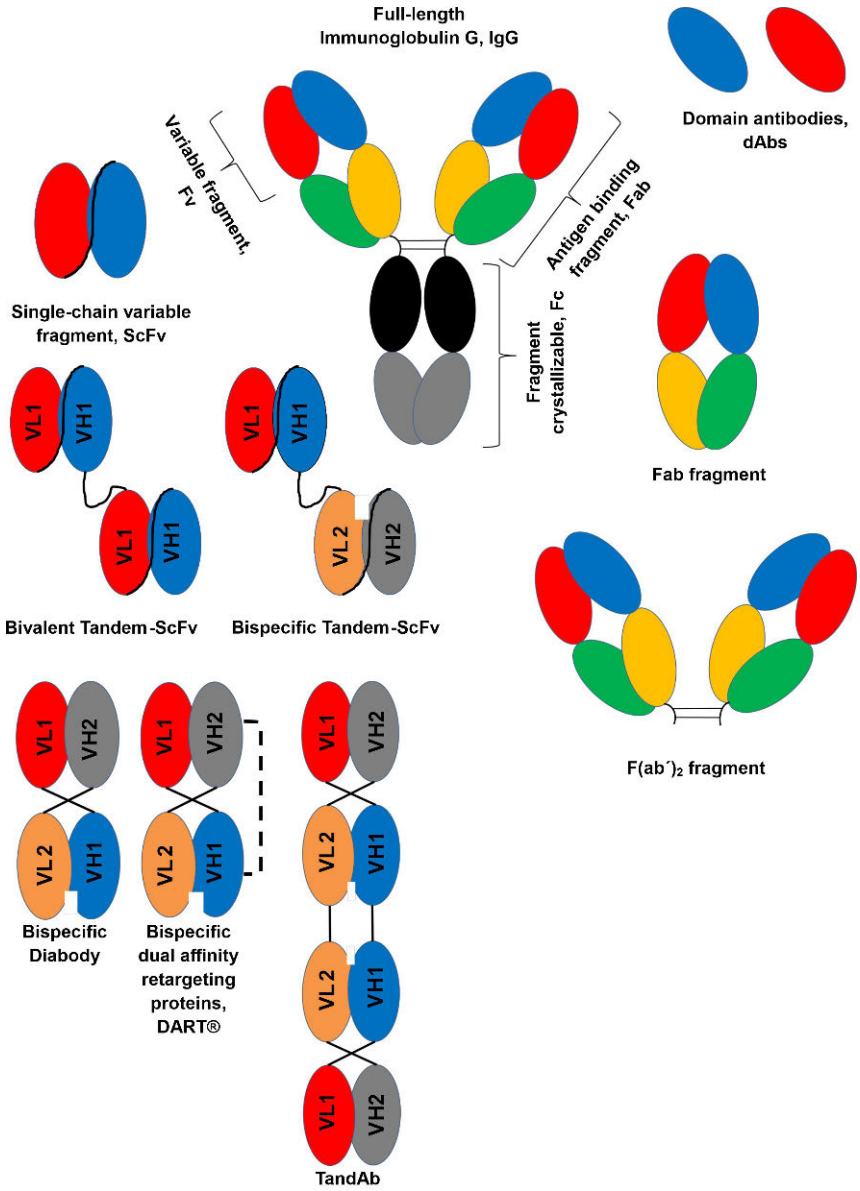

Figure 1. Schematic structures of a full-length antibody (Immunoglobulin G), different antibody fragments, and antibody fragment derivatives. Variable light and heavy domains are colored in red and blue, respectively. Constant light domain and constant heavy 1 domain are colored in green and orange. Constant heavy domains 2 and 3 are colored in black and grey. Black lines between the domains of ScFvs are linker sequences, which connect the domains. Dashed line represents disulfide bond.

ScFvs are 25–27 kDa sized fragments that are composed of V_L and V_H domains connected by the flexible linker, which is usually rich in serine and glycine residues [10,22–24] (**Figure 1**). The linker is usually 15 residues long (the linker should be at least ~3.5 nm long), and is not prone to form ordered secondary structures [25]. The ScFv fragment can be further modified into several different derivatives like Tandem-ScFvs, which connects two different ScFvs via ~25 residues long helical or flexible peptide linker (**Figure 1**). Consequently, Tandem-ScFvs can be bivalent or bispecific [26–28]. Currently there are two ScFv format based therapeutic agents, Brolucizumab (humanized ScFv) and Blinatumomab (Murine bispecific Tandem-ScFv), approved by the European Medicines Agency (EMA) or the Food and Drug Administration (FDA) [29]. Further ScFv derivatives include Diabodies [30], dual affinity retargeting proteins DARTs® [31] and TandAbs [32] (**Figure 1**). In Diabody format, the linker between the two variable domains is so short that intrachain V_H–V_L pair cannot form, but instead interchain V_H–V_L pairs are formed with another construct with the same arrangement [30]. DART® format is basically the same as Diabody format, but it has an interdomain disulfide bridge [31]. TandAb format in turn connects two Diabodies to form a tetravalent binder [32].

Fabs are ~55 kDa sized fragments that contain C_L and C_H1 domains in addition to the V_H and V_L domains [10,33] (**Figure 1**). The light and heavy chains are assembled together via interdomain disulfide bond between C_L and C_H1 [22]. Fabs have many advantages compared to ScFvs, including stability, native structure and easier conversion into the IgG format without the risk of compromising binding affinity. Furthermore, Fabs can be detected by various secondary antibodies [12,22,34]. So far, there are three Fab fragments approved by the EMA and FDA [29]. Another Fab based format $F(ab')_2$ is composed of two antigen binding arms connected by the Ig hinge region (**Figure 1**). The $F(ab')_2$ is approximately twice the size of Fab fragment (~110 kDa). Both formats, Fab and $F(ab')_2$, can be produced enzymatically by digesting the with papain or pepsin, respectively, or by the means of genetic engineering [10,22].

Domain antibodies, dAbs, are 11–15 kDa sized individual V_H or V_L domains, which binding potential was initially demonstrated in phage display selection against lysozyme [35] (**Figure 1**). Domain antibodies are prone to aggregate, but certain changes, including replacement of hydrophobic residues with more hydrophilic residues at positions 44, 45 and 47 of the V_H [36], and extension of CDRH3 loop may improve the biophysical properties of the domain antibodies [37]. Individual domain antibodies might possess some potential in clinical applications, but until this day, domain antibodies have been mostly used as fusion partners in larger multispecific constructs [22,28,38,39].

2.1.2 Binders based on atypical immunoglobulin scaffold

Besides typical IgG scaffold (IgG1), Camels, Llamas and Alpacas have two types of IgGs (IgG2 and IgG3) that are different from other mammals (**Figure 2**). These heavy chain IgGs, or HCAbs, are special because they don't contain the light chain (hence the name) or the C_H1 domain, which are present in the typical IgGs. The difference between IgG2 and IgG3 is the length of the hinge region, which is shorter in IgG2. Since HCAbs don't bear the light chain or C_H1 domain, the variable heavy domains (V_HH) of the HCAbs have naturally evolved to function as a single domain. Therefore, there has been an interest to use the V_HH domains as individual domains (also known as Nanobodies), which are 12–15 kDa in size, soluble and stable [22,40–44]. Due to the aforementioned characteristics, the natural design of the Nanobody has also been used as a template to improve the biophysical properties of human dAbs [36]. Compared to the V_H of the typical IgG, the Nanobody has more hydrophilic framework 2 (FW2) region, longer CDR1 and CDR3 loops, and one extra disulfide bond between CDR1 and CDR3. The hydrophilic residues in the FW2 region of the Nanobody are those, which usually interact with V_L in typical IgGs [45,46]. The renal cut-off value is ~60 kDa, which means that individual Nanobodies are rapidly filtered out from the bloodstream by the kidneys, usually within two hours [47,48]. The rapid clearance can be advantageous in some cases, like certain cancer treatments [42] or imaging [48], but disadvantageous in inflammatory or infectious diseases [47,49]. At the moment, there is one Nanobody based drug, Cablivi™, which is approved by the EMA and FDA for the treatment of thrombotic thrombocytopenic purpura [29].

Yet another rarity, which is found in cartilaginous fishes, is a variable domain of New Antigen Receptor (VNAR) [50], which is similar to the Nanobody, since it does not have the light chain [22] (**Figure 2**). Further similarities include small size (11–12 kDa) and high stability [22,41]. However, one striking difference compared to the variable domains of typical IgGs and Nanonody is the lack of CDR2 region. Furthermore, the VNARs contain regions called hypervariable loops, HV2 and HV4 (HV4 also affects the binding of the antigen), which are frequently mutated, and located between the CDR1 and CDR3 [51,52]. The CDR3 loop of the VNARs can be extraordinarily long, up to 40 residues [53], but the average length is about 14–18 residues [54].

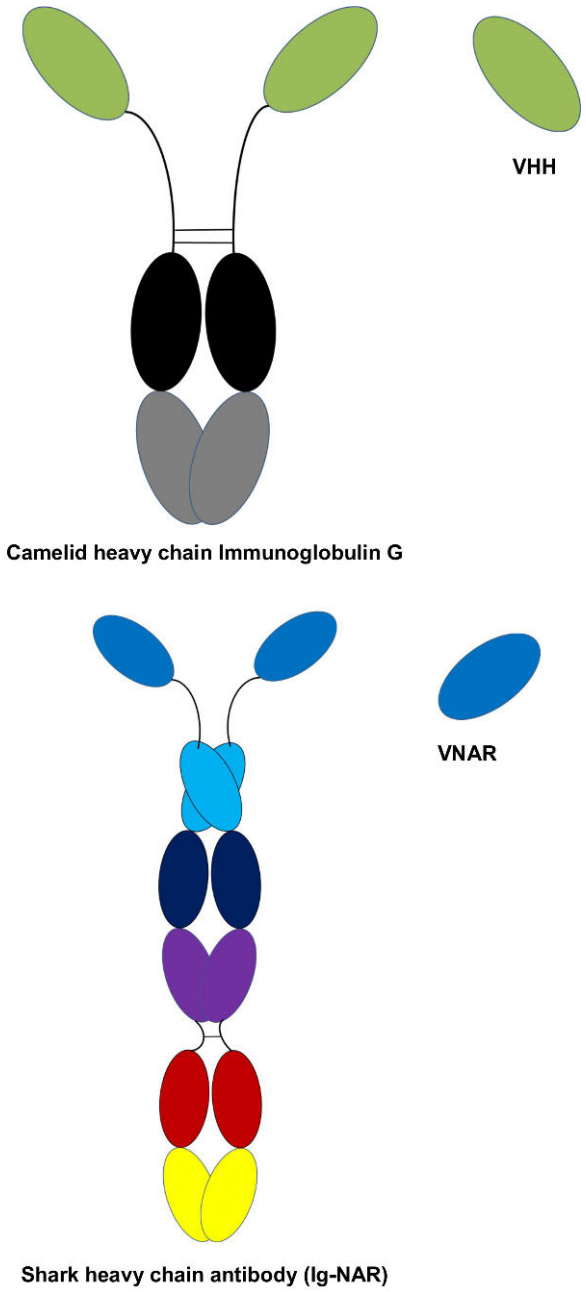

Figure 2. Schematic structures of a camelid heavy chain immunoglobulin G, V_HH domain (or Nanobody), shark heavy chain antibody (Ig-NAR) and new antigen receptor (VNAR). Constant heavy domains 2 and 3 of the camelid heavy chain IgG are colored in black and grey. V_HH domain is olive green. Constant domains 1, 2, 3, 4 and 5 of the shark heavy chain antibody are colored in light blue, dark blue, purple, dark red and yellow, respectively. VNAR fragment is colored in blue.

2.1.3 Binders based on non-immunoglobulin scaffold

Binders based on non-immunoglobulin scaffolds (non-IgG scaffolds) provide a worthy addition to the set of recombinant binders, since many of the non-IgG scaffolds can be easily expressed in high yields in *E. coli*, they are simple (e.g. no need for post-translational modifications, small size and most of them don't contain disulfide bonds) and many of them are very stable [17,18,55]. Almost all binders based on non-IgG scaffolds, which are of clinical interest, are derived from naturally occurring molecules [18], including Adnectins (fibronectin type III) [56–58], Anticalins (lipocalins) [59], Avimers (A-domain of cell surface receptors) [60], Fynomers (FYN tyrosine kinase) [61,62], Kunitz domains (Kunitz-type protease inhibitors) [63–65], Knottins (found from many species e.g. *Conus magnus*) [66,67] and Affibodies (protein A from *Staphylococcus aureus*) [68]. However, there is a one type of binder called DARPin (Designed Ankyrin Repeat Protein), which is of artificial origin. DARPins are composed of varying number of ankyrin repeats, which occur abundantly in nature, between N-capping and C-capping domains. DARPins allow very flexible design and engineering, since ankyrin repeats between the capping domains can be added, shuffled or deleted [69,70]. Structures of the binders based on non-immunoglobulin scaffolds are represented in the **Figure 3**.

To date, the most successful non-IgG scaffold-based binders from therapeutic point-of-view have been Knottin and Kunitz domain, since there are two approved drugs based on Knottin (Prialt and Linzess) and one based on Kunitz domain (Ecallantide) [18]. Recently, the FDA accepted Biologics License Application (BLA) of a DARPin based drug called Abicipar, which is used for the treatment of neovascular age-related macular degeneration. Furthermore, the EMA approved the Marketing Authorization Application (MAA) of Abicipar [71].

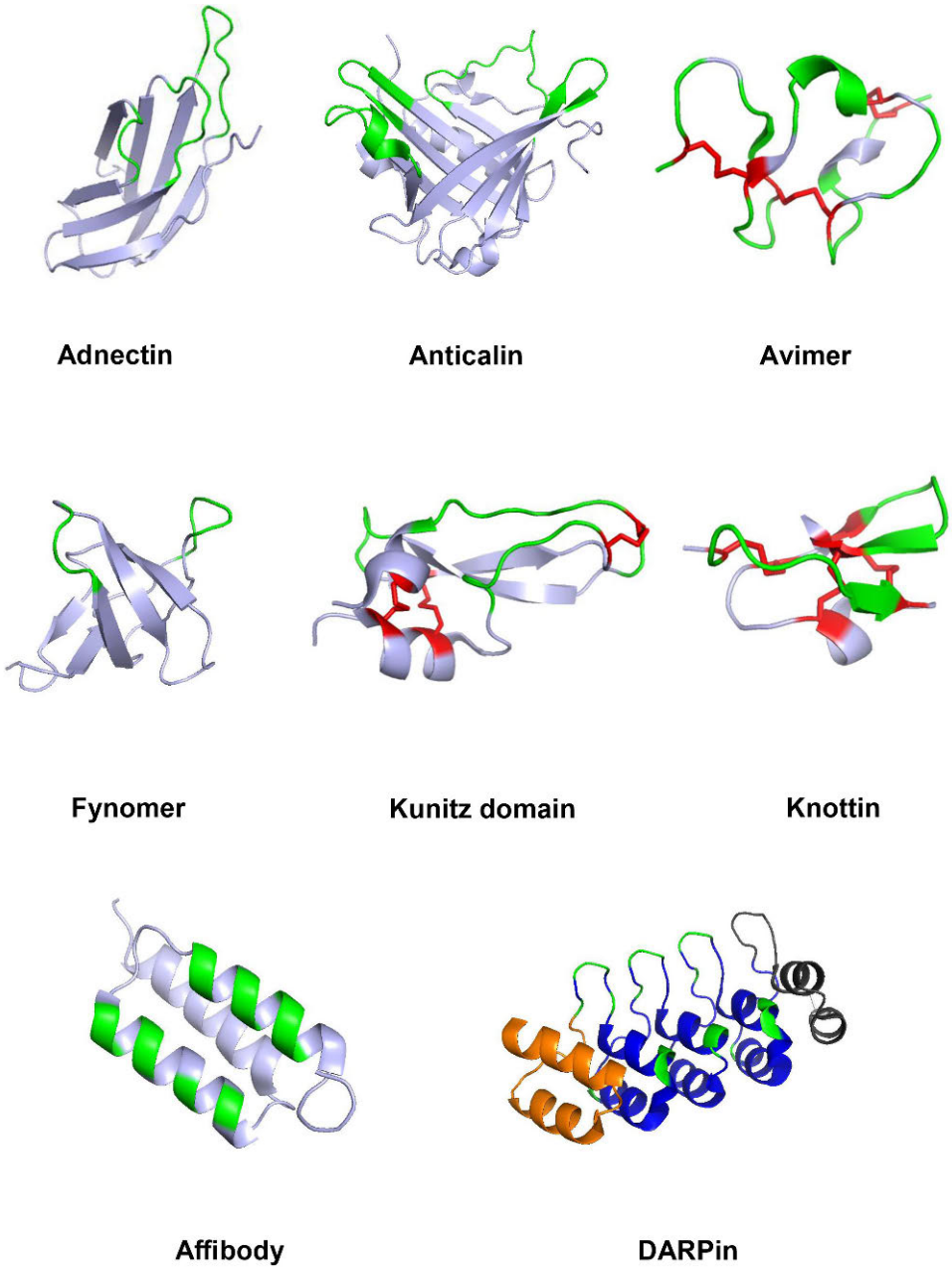

Figure 3. Cartoon structures of the binders based on non-immunoglobulin scaffold. Structures of Adnectin (PDB ID: 1ttg), Anticalin (3bx7), Avimer (1ajj), Fynomer (1m27), Kunitz domain (1kth), Knottin (2it7), Affibody (1q2n) and DARPin (1svx) were retrieved from Protein Data Bank and visualized by PyMOL Viewer. Disulfide bridges are shown in red and antigen contact sites in green. N-capping and C-capping domains of DARPin are shown in orange and gray, respectively.

2.2 Synonymous codon usage and its effects on protein expression in *E. coli*

2.2.1 Synonymous codon usage bias

The genetic code is redundant, which means that there are 64 different codons encoding 20 amino acids and three translation stop signals. Due to the redundancy of the genetic code, each amino acid, excluding methionine and tryptophan, is encoded by more than one codon. Codons that encode the same amino acid are called "synonymous codons" [72]. The synonymous codons are not used equally, but some synonymous codons are preferred (i.e. more frequently used) over others. The identities of preferred codons can vary significantly between species [73] and between genes [74] in the genome of an organism. This phenomenon is called "synonymous codon usage bias" [6,75,76] (**Figure 4**). If synonymous codon is changed to another synonymous codon encoding the same amino acid, it is called "synonymous mutation". There are two main hypotheses to explain the existence of codon usage bias: mutation hypothesis and natural selection hypothesis. The mutation hypothesis claims that some codons are more favorable targets for mutations than others, and therefore, occur in lower frequencies [77,78]. The selection hypothesis in turn suggests that an organism gains fitness benefits (e.g. more efficient and accurate translation of proteins) as a result of some synonymous mutation(s), which is/are then passed forward [79]. It is important to note that these hypotheses do not exclude each other, but they co-exist, and shape the codon usage of organisms together [79,80].

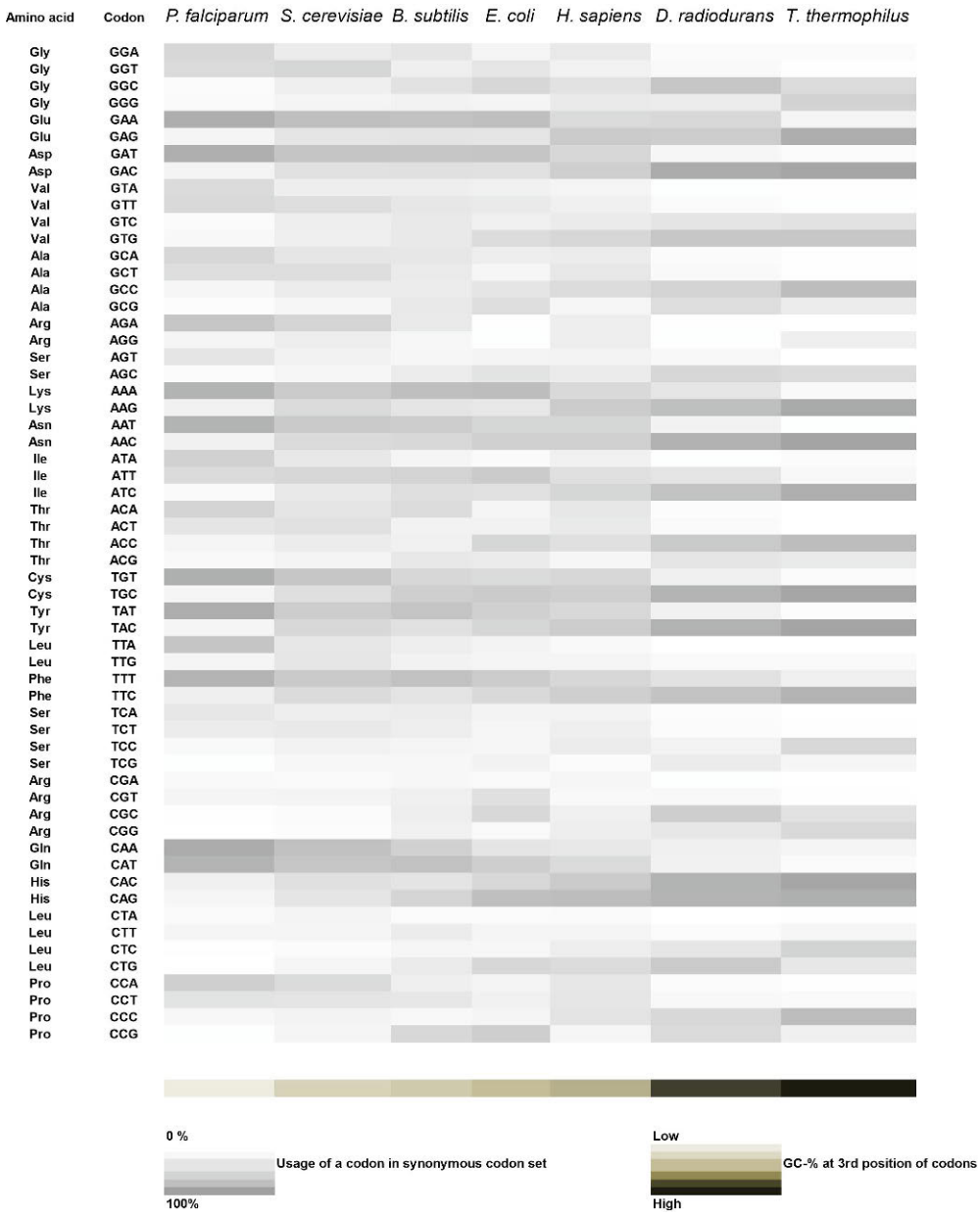

Figure 4. Illustration of codon usage bias within organisms and between organisms. The codon usage frequencies of the selected organisms were retrieved from Kazusa Codon Usage Database [81] and visualized by a greyscale heat map. The heat map represents the usage distribution of codons in the synonymous codon set. Methionine, tryptophan and the stop codons are excluded from the dataset. A rectangular heat map below the codon usage heat map illustrates the occurrence frequency of guanine and cytosine bases at the third position of codons.

2.2.2 Concept of codon optimality

The concept of codon optimality is traditionally based on a hypothesis according to which highly expressed genes are under strong selection pressure in terms of translational efficiency [82,83]. Therefore, codons that frequently occur in the highly expressed genes are considered to be the most optimal codons in terms of translation [83]. In addition to the enrichment in the highly expressed genes, the optimal codons are usually recognized by abundant tRNA species [84,85] and pair with tRNAs via Watson-Crick (WC) interaction instead of wobble interaction [85]. Generally, tRNA abundance correlates with faster translation rate of a codon [75,82,86], but also the mode of codon–anti-codon interaction has an effect (WC interaction is thought to be "faster") [87]. In contrast, rare codons (codon usage frequency < 1 % in the genome of *E. coli*), which are enriched in lowly expressed genes, are traditionally considered as translationally non-optimal, because they are recognized by low abundance tRNA species and therefore translated slowly [7,88–90]. During the years, many different codon usage tables have been established to rank all the 64 codons and to point out the most optimal codons in the codon set [83–85,91]. Consequently, the traditional approach for increasing protein expression – with varying success – has been codon-optimization, which replaces all the codons of a gene with the most optimal codon encoding that same amino acid [1,16,92–96]. However, as for the efficient expression of proteins, judging codons as optimal or non-optimal solely based on previously mentioned characteristics is not entirely valid, because the "optimality" of a codon also depends on the location of the codon in a gene, and other context dependent variables like codon-pair bias. Local and other context-dependent variables, which have an influence on the choice of the most suitable codon in terms of expression, are reviewed in the following sections.

2.2.3 Codon usage at the start of the open reading frame

In many organisms, around the first 30–50 codons (~20 codons in *E. coli*) of an open reading frame (ORF) are generally translated with rare codons [97,98]. Tuller and colleagues showed that rare codons at the start of the ORF increase translational efficiency, which led to a theory called "ramp". According to the "ramp theory", rare codons at the start of the ORF slow down the ribosomes at the initial stage of translation elongation, thus reducing "ribosomal traffic jams" [97]. However, further studies implemented with *E. coli* suggested that increased translational efficiency was not due to the codon rarity itself, but due to reduced mRNA secondary structures near the translation start site (TSS). The studies suggested that reduced mRNA secondary structures near the TSS make the ribosome binding site more accessible to the ribosome, thereby facilitating the

translation initiation and increasing protein expression [99–101]. The finding made sense because rare codons in *E. coli* are usually AT-rich at the third position, and besides, Bentele et al. demonstrated that rare codons are only preferred at the start of the ORF if they are AT-rich [100–102]. Regarding the relationship between mRNA secondary structures near the TSS and protein expression, the most important region relative to the start codon was found to be -4–+37 nucleotides [99], in which the position +10 nucleotides exhibited the strongest correlation between mRNA secondary structures and protein expression [100]. Furthermore, mRNA secondary structures with Gibbs free energy (ΔG) less than -10 kcal/mole near the TSS can seriously hamper or totally inhibit translation [99,103]. It should be noted that mRNA secondary structures with Gibbs free energy less than -10 kcal/mole are rare in the endogenous genes of *E. coli* [104].

2.2.4 Codon usage between domains and smaller structural motifs

Domain boundaries are segments that separate higher-ordered structures from each other [96,105]. Thanaraj and Argos studied 37 multi-domain proteins of *E. coli*, and found out that optimal codons are avoided in translationally slow domain boundary segments, while many rare codons are preferred in those segments [105]. Translationally slow segments have been associated with co-translational folding and solubility *in vivo* [105–107]. However, in a subsequent study, Saunders and Dean analyzed large protein set including proteins from *E. coli* and did not find any evidence of enrichment of rare codons in the domain boundaries. Instead, the authors observed enrichment of optimal codons (i.e. fast translated codons) in the domain boundaries [108]. A very recent and comprehensive study didn't observe any enrichment of rare codons in the domain boundaries either, but detected slowly translated regions between smaller structural elements [109], which was concurred by the aforementioned study [108]. The discrepancy in results regarding domain boundaries might be explained by the idea according to which some domains benefit from the enrichment of rare codons in the domain boundaries, whereas some domains benefit from the enrichment of optimal codons. The former scenario may occur if domain folds cooperatively and the latter in non-cooperative folding [110]. On the other hand, rare codons in the domain boundaries can have a beneficial effect on folding if, for example, the folding of the N-terminal domain might be interfered by the C-terminal domain. In this case, rare codon clusters give more time to the N-terminal domain to fold correctly [111,112]. The correct folding guided by the codon usage can affect even the oxidation states of cysteine [113]. Lastly, codon usage is not, however, the only way to decrease the speed of

translation in translationally slow segments, but stable hairpin structures formed in these segments can potentially have the same effect [105,114,115].

2.2.5 Codon usage in different secondary structure units

A study by Thanaraj and Argos initially suggested that α-helices are encoded by optimal codons, and β-structures and loops by codons that are translated slowly [116]. However, subsequent studies have shown that codon preferences in different secondary structure units are not so unambiguous. For instance, the codon preference can depend on positional location in the secondary structure (e.g. start, center or end). This is especially prevalent in the case of asparagine codon AAC of *E. coli*, which is highly preferred at the C-terminal end of β-structures [108,117]. Furthermore, some codons like valine codon GTT can be preferred at the start of α-helix, but not preferred at the center of the α-helix; in contrast, the situation is the opposite in the case of GTG. Codon optimality can also function as a signal for secondary structure transitions. Optimal codons signal the start of a coil region, whereas translationally less-optimal codons signal the start of a helix or strand. Also, regarding two-domain proteins α,β, α/β and α+β of *E. coli*, the second domain is translated faster than the first one 60 %, 100 %, 75 % and 50 % of the times, respectively. This further affects codon selection in different secondary structure units [108]. In addition, secondary structure composition of a protein can potentially dictate whether a codon is a "former" or "breaker" of a certain secondary structure. In all-alpha proteins CAC codon is "helix former", while in α+β and α/β proteins it is "helix breaker". Similarly, CTA codon is strand former in all beta proteins, but breaker in α+β and α/β proteins [118].

2.2.6 Codon usage, protein aggregation and mistranslation

Mistranslation-induced protein misfolding is a hypothesis according to which optimal codons – due to their translational accuracy – are favored in the regions where missense errors (normal rate 1/1000–1/10000 codons) would be highly deleterious for protein structure, and therefore cause misfolding and aggregation of a protein [7,119–121]. Furthermore, it has been suggested that GroEL chaperone dependency of a protein affects codon selection as well [122]. The sensitive regions comprise especially buried and functionally important regions of a protein, as well as regions where mutations would cause major changes in Gibbs free energy ($\Delta\Delta G > 3.0$) [120,123]. The missense errors in translation are caused by the competition between cognate tRNAs (carrying the right amino acid) and near-cognate tRNAs (carrying the wrong amino acid). Therefore, a given codon is prone to misreading if the codon is decoded by low-abundance cognate tRNA, and there

are high-abundance near-cognate tRNAs that are able to pair with the codon. For example, Kramer and Farabaugh showed that arginine codons AGA (one copy of cognate tRNA) and AGG (one copy of cognate tRNA), and asparagine codon AAT (no cognate tRNA) are highly susceptible to misreading by $tRNA_{UUU}^{Lys}$ (tRNA copy number = 6) [97,124]. The interactions between U•U and U•G (AGG and AGA) are stable at the second base position of a codon [125], while the third base position (wobble position) is not so strictly regulated by the ribosome [126]. Interestingly, in most of the bacterial genomes analyzed by Shah and Gilchrist, focal tRNA abundances exhibited a positive correlation with the abundance of neighboring tRNA [127], suggesting that suppression of missense errors caused by tRNA competition might be embedded into the genetic code, at least in bacteria. However, too fast translation speed, for example by using excessive amounts of rapidly translated optimal codons, can also cause protein aggregation [87,128]. Recoding of the luciferase gene by the most rapidly translated codons caused increased aggregation of the protein [87].

2.2.7 Codon pair bias

Two codons residing next to each other are not randomly selected, but certain codon pairs are preferred. This phenomenon is called codon pair bias, which also affects translation speed [129,130]. For example, leucine–alanine codon pairs CTG-GCA and CTG-GCG are favored, whereas CTG-GCC is underrepresented [129]. Some codon pairs, like ACG-CTG pair, can have an inhibitory effect on translation, although the pair frequently occurs in the genome of *E. coli*. In contrast, some codon pairs have favorable position-dependent effects on translation (e.g. rare codon pairs before stop codon) [131]. Nevertheless, at some positions, rare arginine codon pairs can be highly deleterious [132]. Certain codon pairs are universally avoided in bacteria, archaea, and eukaryotes. Such codon pairs include, for example, nnUAnn, nnGGnn, nnGnnC, nnCGCn, GUCCnn, nnCnnA, and UUCGnn. Codon pairs nnGCnn, nnCAnn, and nnUnCn are favored [133]. The mechanism of translation enhancing effect by codon pair bias is not fully understood, but it is thought to be caused by the interaction between two tRNAs at P-site and A-site of the ribosome [134]. Furthermore, in the case of amino acid pairs that are composed of two identical amino acids, the pair is preferentially coded by the pair of identical synonymous codons (e.g. TTT-TTT and TTC-TTC) [135].

2.2.8 Codon usage and mRNA stability

Codon usage modulated translation initiation and translation elongation have a significant effect on mRNA degradation, although it was previously thought that

mainly GC content affects the stability of the mRNA via higher thermodynamic stability [99,136,137]. Efficient initiation and elongation lead to a higher ribosomal density, which in turn leads to a longer mRNA half-life [138]. Furthermore, the expression of 14,000 sequence variants in *E. coli*, as well as the combination of competition assay and deep sequencing revealed that efficient translation initiation combined with longer mRNA a half-life yields more protein per mRNA molecule, and thereby reduces the "fitness cost" of protein expression [139]. In addition the aforementioned variables affecting mRNA half-life, avoiding RNase E cleavage sites is important as well [99,140].

2.2.9 Codon usage in signal sequences

Whole-genome analysis of *E. coli* showed that rare codons occur more frequently in signal sequences compared to the mature portions of a gene. The enrichment of rare codons in the signal sequences was even more pronounced than at the start of the ORF of non-secretory genes [141]. Later, Zalucki and colleagues studied the effect of synonymous codon usage of signal sequence on the expression of maltose-binding protein (MBP) and β-lactamase in *E. coli*. When rare codons were changed to the most optimal codons, the expression of β-lactamase decreased by fourfold and the expression of MBP decreased by 20-fold [142,143]. In a subsequent study, Zalucki and colleagues discovered that changing rare codons to optimal codons, in fact, affects the folding of the transported protein, which in this case was MBP. Due to the change in synonymous codon usage in the signal sequence, the MBP folded into a conformation that was sensitive to proteases leading to degradation of the MBP [144]. However, a recent study, which examined the effect of synonymous codon usage of light and heavy chain PelB signal sequences on Fab fragment expression, discovered that the position of a rare codon is more important than the sheer number of rare codons in the signal sequence. In this case, the Leu-5 position of the light chain PelB signal sequence was particularly important. Furthermore, the study suggested that decreased mRNA secondary structures at the translation start site and rare codon at the Leu-5 exhibit independent, but additive effects on expression [145]. The finding might be related to the hypothesis according to which reduced mRNA secondary structures in the translation start site (efficient translation initiation) and rare codons in the signal sequence (slower movement of the ribosome) affect protein translocation efficiency and chaperone recycling [146]. It has also been suggested that rare codons in the signal sequences of prokaryotes might work in a similar fashion as the signal recognition particle (SRP) pathway or facilitate the interaction with the SRP [98,141]. Moreover, it has been shown that routing to the SRP pathway in *E. coli* may also occur by means of SD-like (Shine-Dalgarno) sequences, which cause ribosomal arrest [147,148].

2.2.10 Translation of codons during amino acid starvation

Isoacceptor tRNAs are tRNAs, which are charged with the same amino acid, but recognize different synonymous codons (or set of synonymous codons). These isoacceptor tRNAs respond differently to amino acid starvation, which means the deprivation of an amino acid. For example, $tRNA^{Leu}_5$ encoding leucine codon TTG is highly insensitive to amino acid starvation, whereas $tRNA^{Leu}_2$ (CTC, CTT) and $tRNA^{Leu}_3$ (CTA, CTG) are highly sensitive to the same condition. During amino acid starvation, charging of sensitive tRNAs decreases down to zero, whereas tRNAs insensitive to amino acid starvation remain charged. The sensitivity of an isoacceptor tRNA to amino acid starvation is not dependent on tRNA copy number or codon usage frequency. In fact, many abundant tRNA species are sensitive to amino acid starvation [149,150]. The benefit of using codons encoded by the tRNAs insensitive for starvation was shown in a study in which 40 DNA polymerase and 40 ScFv gene variants, which differed only in terms of codon usage, were expressed in *E. coli* [151]. In the genome of *E. coli*, the codons that are encoded by the tRNAs sensitive to amino acid starvation are used as "control points", which regulate the transcription of an operon responsible for the synthesis of a given amino acid. For instance, in the case of *leu* operon, which is responsible for leucine biosynthesis, the leucine codon CTA is used as a control point (i.e. regulatory codon) in a short-lived leader peptide that precedes structural genes. When the leucine levels are low, it takes more time than in a normal situation for the ribosome to translate the leader peptide including the regulatory codon. This event causes a change in mRNA secondary structure further downstream, which allows RNA polymerase to continue transcription. In the opposite condition, faster translation of the regulatory codon causes a change in mRNA secondary structure that prevents the transcription [150,152,153]. In contrast, codons that are insensitive to amino acid starvation are preferentially used in enzymes that synthesize the given amino acid [150].

2.2.11 Preserving codon landscape by codon harmonization

The codon harmonization approach is based on the idea according to which each gene intrinsically contains the blueprint – or non-random occurrence of slowly and fast translated codons at different positions in the gene – for optimal translation kinetics, which thereby leads to optimal protein folding. This non-random occurrence of codons in the gene is called codon landscape. In the codon harmonization approach, a heterologous gene is recoded with synonymous codons of a production organism, which mimic the codon usage frequency of an original host [5,96] (**Figure 5**). Thus, the codon landscape is preserved. In contrast, the codon-optimization approach abolishes the codon landscape, because in this

approach only the most frequently used/optimal synonymous codons of the production organism are used at each position in the heterologous gene. The codon harmonization approach has been successfully applied in the heterologous expression of different proteins [87,96].

Amino acid sequence of human IGHV3	Native codon usage of human IGHV3		Codon usage of human IGHV3 optimized for *E. coli*		Codon usage of human IGHV3 harmonized for E. coli	
E	GAG	0,58	GAA	0,7	GAA	0,7
V	GTG	0,46	GTG	0,4	GTG	0,4
Q	CAG	0,73	CAG	0,7	CAG	0,7
L	CTG	0,4	CTG	0,46	CTG	0,46
V	GTG	0,46	GTG	0,4	GTG	0,4
E	GAG	0,58	GAA	0,7	GAA	0,7
S	TCT	0,19	AGC	0,33	TCG	0,16
G	GGG	0,25	GGC	0,46	GGT	0,29
G	GGA	0,25	GGC	0,46	GGT	0,29
G	GGC	0,34	GGC	0,46	GGT	0,29
L	TTG	0,13	CTG	0,46	TTG	0,12
V	GTA	0,12	GTG	0,4	GTA	0,17
Q	CAG	0,73	CAG	0,7	CAG	0,7
P	CCT	0,29	CCG	0,55	CCT	0,17
G	GGG	0,25	GGC	0,46	GGT	0,29
G	GGG	0,25	GGC	0,46	GGT	0,29
S	TCC	0,22	AGC	0,33	TCG	0,16
L	CTG	0,4	CTG	0,46	CTG	0,46
R	AGA	0,21	CGC	0,44	CGT	0,36
L	CTC	0,2	CTG	0,46	TTA	0,15
S	TCC	0,22	AGC	0,33	TCG	0,16
C	TGT	0,46	TGC	0,58	TGT	0,42
A	GCA	0,23	GCG	0,38	GCA	0,21
A	GCC	0,4	GCG	0,38	GCG	0,38
S	TCT	0,19	AGC	0,33	TCG	0,16

Figure 5. Illustration of the codon harmonization approach. N-terminal part of the human IGHV3 gene is used as an example. The grey segments indicate the positions that are affected by the codon harmonization. The numbers indicate the codon usage frequency of a given codon. The codon usage frequencies were obtained from Kazusa Codon Usage Database [81].

2.3 Approaches for improving recombinant antibody expression in *E. coli*

2.3.1 *E. coli* as an expression platform for recombinant antibodies

E. coli has been an important organism for therapeutic recombinant protein expression since the early 1980s when the FDA approved recombinant insulin, which was produced in *E. coli*, for treatment of diabetes [154]. As for the production of functional antibody fragments in *E. coli*, three significant milestones were achieved in 1988. First, Arne Skerra and Andreas Plückthun reported a successful assembly and secretion of both variable domains into the periplasmic space [155] and subsequently Bird et al. published a single-chain variable fragment (ScFv) that joined the two variable fragments with peptide linker [23]. The third milestone was a successful secretion of an active chimeric fragment antigen binding (Fab) described by Better et al. [33]. One milestone also worth mentioning was achieved in 2002, when Simmons et al. reported successful expression of a full-length aglycosylated IgG in *E.* coli [156]. Still, ScFv and Fab fragments are the most common recombinant antibody formats expressed in *E. coli* [157,158] and the organism itself can be considered as the standard for laboratory scale production of antibody fragments [159]. *E. coli* possesses several characteristics that make it a popular production organism. First and foremost, a vast amount of experience (including protocols, general practices, and genetic information) has been accumulated over the last 30 years [159], which makes *E. coli* easy to use and manipulate genetically [160]. Further advantages include fast growth, inexpensive reagents, easy scale-up, high yields and wide selection of strains and cloning tools [154,160,161]. Despite many favorable characteristics, *E. coli* also has some shortcomings when it comes to the production of complex molecules like recombinant antibodies. The cytoplasm of *E. coli* is reducing, which means that disulfide bonds cannot be formed in the cytoplasm. Correct disulfide bond formation is essential for the correct folding of recombinant antibodies, and traditionally, the recombinant antibodies have been transported to the oxidative environment of the periplasm to ensure the correct formation of the disulfide bonds [159]. However, the periplasmic expression may be hampered by inefficient translocation, which reduces yields [15]. Moreover, the periplasmic compartment is much smaller than the cytoplasmic compartment, and therefore the capacity of the periplasm (e.g. limited chaperone availability and sheer space to contain expressed proteins) for the expression is lower [159]. One major drawback, besides the aforementioned problem, is the lack of glycosylation machinery that prohibits especially the production of glycosylated full-length IgGs in *E. coli*. In addition, *E.*

coli doesn't have sophisticated folding machinery like mammalian cells do, which makes it even more difficult to express complex molecules of heterologous origin, including recombinant antibodies [157,158].

Phage display is an essential part of recombinant antibody development campaigns implemented in *E. coli*. The phage display technology is based on the display of a protein of interest (e.g. recombinant antibody) on the surface of a phage particle and the physical linkage between phenotype (displayed protein) and genotype (gene encoding the protein inside the phage particle) [162–165] (**Figure 6a**). The linkage between phenotype and genotype is a crucial feature when antigen-specific binders are selected and enriched, for example, from antibody libraries during a process called biopanning (**Figure 6b**). The displayed protein of interest is usually fused to the minor coat protein p3 (p8 and p9 fusions are also used) of a filamentous M13 phage, which is a non-lytic phage enabling continuous production of new phage particles displaying the protein of interest [166,167]. The gene of the displayed protein of interest can be inserted in the phage genome and fused to the coat protein of choice, or the fusion protein gene can be expressed from a separate vector called phagemid, in which case the rest of the phage genome is provided by so-called helper phage infection [166]. The latter approach is more popular in the construction of large phage libraries since it enables easier genetic manipulation and larger library sizes [166]. For a successful display of proteins on the surface of the phage particle, the fusion protein has to be transported to the periplasm in order to fold correctly. Therefore, phage display is hindered by the same problems as soluble expression of recombinant antibodies. In phage display applications, the recombinant antibodies are displayed as antibody fragments [12,166].

Different approaches have been studied to overcome the hurdles described above, and to improve yields and biophysical properties of recombinant antibodies expressed in *E. coli*. The following sections cover the achievements related to the improvement of recombinant antibody expression in *E. coli* and antibody phage display.

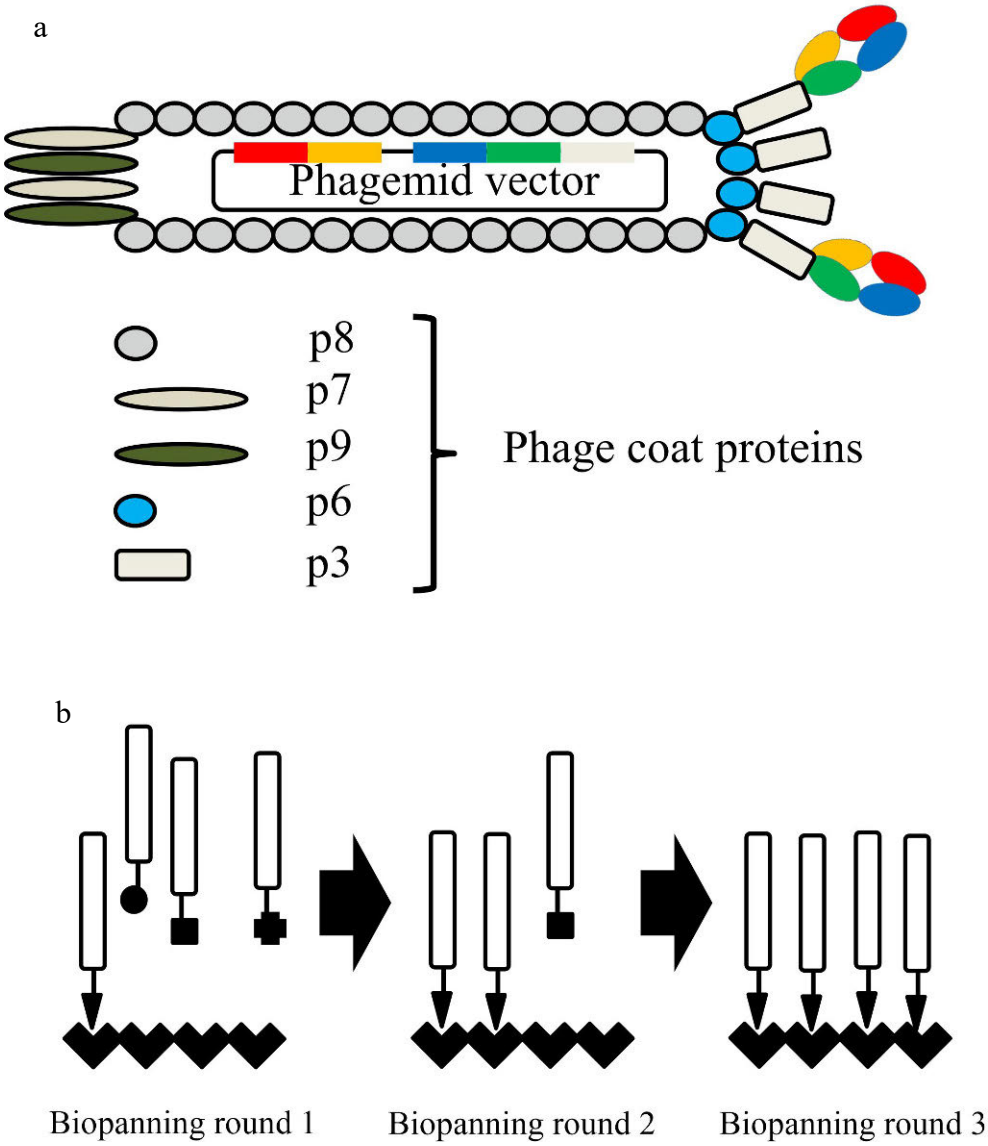

Figure 6. Illustration of phage display. (a) Representation of a filamentous M13 phage displaying a Fab fragment fused to a p3 coat protein. A phagemid vector encoding the fusion protein is shown inside the phage particle. Phage coat proteins are denoted below the phage. (b) Representation of a biopanning process. Phage pool displaying different versions from the same protein (e.g. recombinant antibody) are represented to the antigen. Unbound phages are washed away and bound phages are eluted. *E. coli* cells are infected with the eluted phage pool, and specific binders are enriched. Then the selection procedure is repeated and target specific binders are further enriched.

2.3.2 Co-expression of chaperones

In many cases, heterologous proteins expressed in *E. coli* are prone to aggregate, which is usually caused by the inadequate amount of chaperones and foldases [168]. Therefore, one popular approach for preventing protein aggregation, and thereby increase the amount of active protein, has been the co-expression of different molecular chaperones and foldases [169]. The molecular chaperones, which have been observed to improve the expression of recombinant antibodies, can be broadly divided into two groups: cytoplasmic chaperones (GroEL/ES, DnaKJE and trigger factor) and periplasmic chaperones/foldases (FkpA, Skp, DsbC and DsbA) [158,170].

2.3.2.1 Periplasmic chaperones FkpA and Skp

Periplasmic chaperone peptidyl prolyl *cis-trans* isomerase (PPIase) FkpA is a homodimer that can be divided into two domains: N-terminal domain and C-terminal domain. Both domains have different and independent functions; the C-terminal domain of FkpA exhibits PPIase activity, and the N-terminal domain exhibits chaperone activity [171]. Another periplasmic chaperone Skp is a 17 kDa size trimer that has a central cavity formed by the three subunits [172]. Skp binds its substrate in the cavity and thus prevents the aggregation of the substrate [172]. Skp helps outer membrane protein intermediates to fold properly, and its function is really essential, since the absence of Skp causes protein aggregation in the periplasm [173]. Many studies have shown that co-expression of FkpA and Skp improve the binding activity, solubility, and phage display of different ScFv fragments in periplasmic expression [173–178]. Moreover, co-expression of FkpA or Skp can cause a considerable increase in leakage of ScFv fragment to the extracellular medium (~40 mg/l in shake flask culture) [174]. However, although both chaperones have beneficial effects on the aforementioned characteristics when co-expressed independently, they don't show synergistic effects when co-expressed simultaneously [174]. In contrast to the previously mentioned studies, in which the chaperones were transported to the periplasm, Levy et al. co-expressed FkpA without a signal sequence (cytFkpA) thus retaining the chaperone in the cytoplasm. The authors found out that cytFkpA improves the periplasmic expression of Fab fragments containing kappa light chains and that cytFkpA, in fact, works even better than the native FkpA. A similar effect regarding Fab fragments containing lambda light chains was not observed, because unlike the kappa light chains, the lambda light chains do not contain *cis*-prolines that are the substrates for the PPIase activity of FkpA. The authors speculated that the beneficial effect on periplasmic expression is due to the isomerization of key prolines of the kappa light chains in the cytoplasm, which removes so called "folding bottleneck" subsequent to the

translocation when Sec-dependent secretory pathway is used [179]. Previously, Levy et al. had also demonstrated that cytoplasmic co-expression of Skp resulted in 5–6-fold (0.8 mg/l/OD600) increase in Fab fragment yield when also the Fab fragment was expressed in the cytoplasm of *E. coli* Origami (DE3) [15]. Lastly, some implications are suggesting that the co-expression of Skp can be absolutely crucial for the activity and phage display of some antibody fragments, for example, single-chain antibodies (scAb) [180,181].

2.3.2.2 Periplasmic foldases DsbC and DsbA

DsbC and DsbA are periplasmic foldases (chaperones that are actively involved in the folding process of proteins in ATP-dependent manner), which belong to a disulfide-bond forming protein system that is located in the periplasmic space and inner membrane of *E. coli* [168,182,183]. The disulfide-bond forming protein system also includes proteins DsbB, DsbD and DsbG [175,183]. The role of each Dsb protein is nicely illustrated by Ario de Marco [170]. In short, oxidized DsbA forms disulfide bonds by oxidizing consecutive thiol groups (DsbA serves as electron pair acceptor) in the nascent polypeptide, and during the process, DsbA is reduced. DsbB re-charges, or oxidizes, thiol groups of DsbA. DsbA does not have a proofreading activity, and therefore, the task of DsbC and DsbG is to reshuffle the incorrectly paired cysteine bridges. DsbC and DsbG are kept in reduced form by DsbD, which in turn is reduced by the cytoplasmic thioredoxin [170,175,183] (**Figure 7**). Like the previously discussed chaperones, also co-expression of DsbC has been shown to exhibit highly beneficial effects on periplasmic and cytoplasmic expression of ScFvs and Fabs [15,168,184–187]. However, unlike FkpA and Skp, DsbC and DsbA showed positive synergistic effects on ScFv and Fab expression when they were co-expressed simultaneously [183,185]. Nevertheless, the synergistic effect was not observed in full-length IgG expression [188]. In addition, the choice of optimal foldase can be dependent on the chosen promoter. Recently, it was discovered that DsbC works better than DsbA when they are used in combination with a weaker phoA promoter, but when the promoter is changed to a stronger Lac promoter, the situation is the opposite. It was suggested that when the expression level is high, the formation rate of disulfide-bonds catalyzed by DsbA is the limiting factor rather than the proofreading capacity of DsbC and *vice versa* [185].

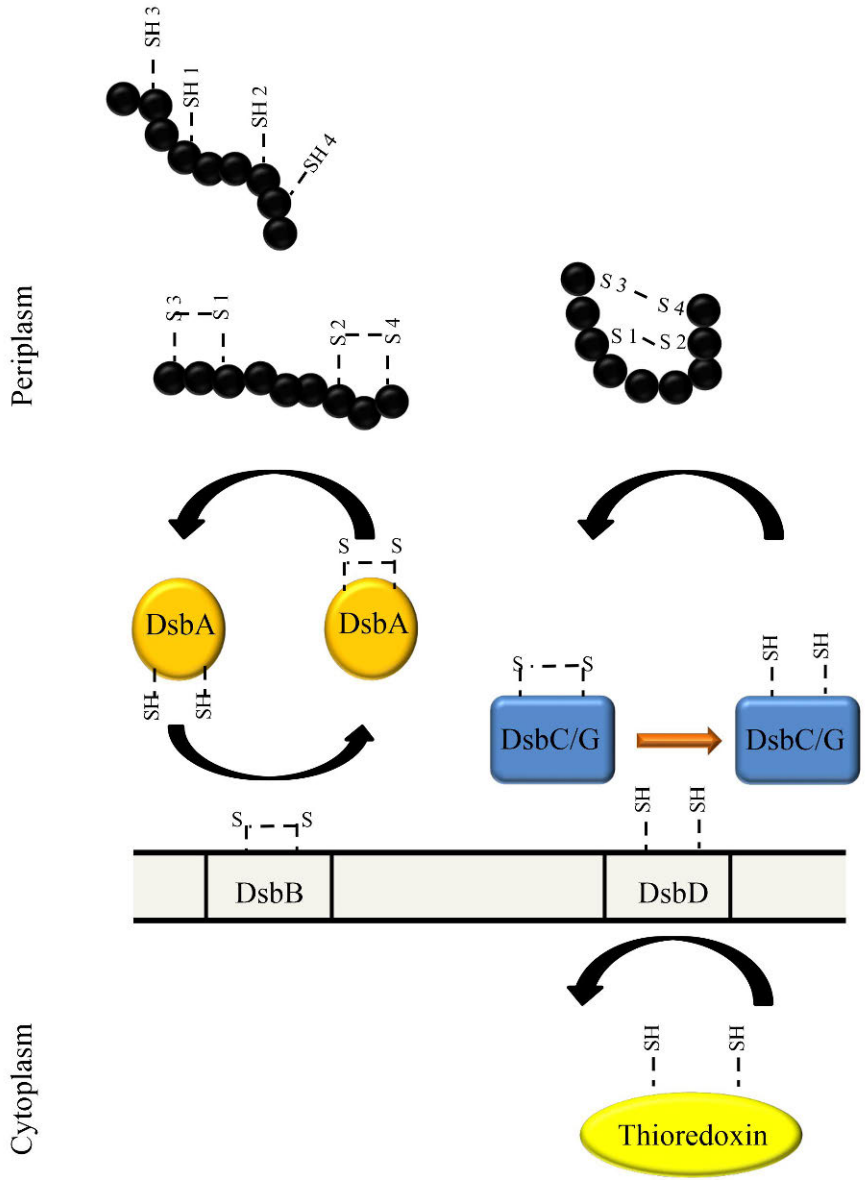

Figure 7. Illustration of the disulfide-bond forming protein system. DsbA forms disulfide bonds by oxidizing consecutive thiol groups and during the process DsbA is reduced. DsbB recharges DsbA. DsbA does not have a proofreading activity, and therefore, the task of DsbC and DsbG is to reshuffle the incorrectly paired cysteine bridges. DsbC and DsbG are kept in reduced form by DsbD, which in turn is reduced by the cytoplasmic thioredoxin. The figure is adapted from [170].

2.3.2.3 Cytoplasmic chaperones

The beneficial effects of co-expression of cytoplasmic chaperones on periplasmic expression have been very limited, and some studies have reported that the co-expression of certain cytoplasmic chaperones in periplasmic expression can be either ineffectual or even detrimental [168,174,184]. Nevertheless, simultaneous co-expression of three cytoplasmic chaperones DnaK, DnaJ and GrpE (DnaKJE system) proved to be useful in periplasmic expression in a study conducted by Hu et al. [184]. DnaK is a highly expressed cytosolic chaperone, which binding to substrates is regulated by co-chaperones DnaJ and GrpE [189]. DnaJ is involved in substrate delivery to DnaK and stimulation of substrate binding to DnaK, while GrpE regulates substrate release [189,190]. The study by Hu et al. revealed that the co-expression of DnaKJE system improved the yield of functional ScFv by 100-fold [184]. The success of the DnaKJE system in the periplasmic expression might be related to DnaK's function as a holdase; it keeps its substrate in a translocation-competent unfolded conformation like SecB that also acts as holdase [191,192]. In the cytoplasmic expression, however, the situation is different. In two studies describing the cytoplasmic expression of Fab fragment and thioredoxin fused ScFv, the results showed that co-expression of DnaKJ (no GrpE) or DnaKJE systems can be deleterious or negligible for the expression [15,193]. Furthermore, in contrast to the periplasmic expression [168,174,184], the co-expression of chaperones GroEL/ES and trigger factor exhibited a positive effect in the cytoplasmic expression [15,187,193]. Trigger factor is a constitutively expressed, ribosome bound chaperone that is the first chaperone to interact with nascent polypeptide. When the polypeptide is released from the trigger factor, approximately 70 % fold spontaneously [194]. GroEL/ES is a complex that forms a "barrel with a lid". GroEL forms the barrel shape and seven GroES co-chaperones work as a lid. The "barrel" contains polypeptides and provides a protected environment for the polypeptides (up to 60 kDa) to fold [189]. The function of trigger factor and GroEL/ES might explain why they work better in cytoplasmic expression than periplasmic expression.

2.3.3 Expression of recombinant antibodies in the cytoplasm

Recombinant antibodies contain disulfide-bonds, which is why they have to be transported into the oxidizing periplasmic compartment instead of the reducing cytoplasmic compartment. However, there are some problems concerning periplasmic expression, including the small volume of the periplasmic space (8–16 % of the total cell volume) and limited capacity of the secretion machinery [195]. This has prompted research groups to find ways to enable the cytoplasmic expression of recombinant antibodies. One way of expressing recombinant

antibodies in the cytoplasm is to use mutant *E. coli* strains, which have a partially oxidizing cytoplasm. These strains, which include for example Origami, Origami derivatives and SHuffle (overexpresses DsbC as well), are deficient in thioredoxin reductase (*trxB*) and glutathione oxidoreductase (*gor*) genes. Both enzymes are involved in a mechanism that maintains the reducing environment in the cytoplasm [196,197], and many studies have shown that the *trxB*/*gor* strains indeed improve the cytoplasmic expression of recombinant antibodies compared to the cytoplasmic expression of recombinant antibodies in the strains with reducing cytoplasm [15,187,193,198–200]. The cytoplasmic expression of recombinant antibodies has also been directly compared to that of periplasmic expression. Venturi et al. compared the cytoplasmic expression and the periplasmic expression of two Fab fragments, which were expressed in Origami and JM83 (wild type with respect to TrxB and Gor genes), respectively. The authors showed that the cytoplasmic expression of the Fabs resulted in remarkable 50–250-fold (10–30 mg/l) increase in expression levels compared to that of the periplasmic expression [201]. Similar results regarding ScFv fragment expression was observed as well [197]. There has been an increasing interest in expressing aglycosylated full-length IgGs in *E. coli*, and different approaches have been explored to achieve efficient expression of full-length IgGs. One way of enhancing the expression of aglycosylated full-length IgGs is to express them to the cytoplasm like Fab and ScFv fragments. Robinson and colleagues proved that the cytoplasmic expression is superior to (~10-fold increase) the periplasmic expression, when it comes to the expression of aglycosylated full-length IgGs [202]. Furthermore, the binding activity and the stability of the IgG expressed in the cytoplasm of *E. coli trxB*/*gor* strain were equal to that of the same IgG that was produced in hybridomas [202].

There are also other means of expressing recombinant antibodies in the cytoplasm, since disruption of the mechanism maintaining the reducing environment in the cytoplasm of *E. coli* is not essential, as shown by the research group of Lloyd Ruddock. First, the group demonstrated that co-expression of sulfhydryl oxidase Erv1p (catalyzes disulfide bond formation) does not only enable the cytoplasmic expression of disulfide-bond containing proteins in BL21 (DE3) pLySRARE strain, but it makes it also more efficiently than *trxB*/*gor* strain Rosetta-gami [203]. In the same study, the authors showed that simultaneous co-expression of disulfide isomerase (in this case DsbC) with Erv1p exhibited a highly positive synergistic effect [203]. Based on these findings, the group developed a system called CyDisCo. The CyDisCo system, which co-expresses Erv1p and protein disulfide isomerase (PDI), can be used to efficiently express natively folded ScFvs (240 mg/l) and Fabs (42 mg/l) in the cytoplasm of *E. coli* [195].

Yet another way of expressing recombinant antibodies in the cytoplasm is to use so called “intrabodies”. Intrabodies are recombinant antibodies for which the

formation of disulfide bonds is not essential, but they can be biologically functional regardless of missing disulfide bonds [158]. Therefore, the intrabodies can be expressed to the cytoplasm. Different mutation and selection strategies have been employed in combination to obtain intrabodies with improved cytoplasmic expression profile [198,204,205], but it is also possible to develop intrabodies simply by fusing maltose-binding protein (MBP) to different ScFvs [206]. By using the combination of random mutagenesis and Antibody-mediated enzyme formation (AMEF) based selection, it was possible to develop an ScFv intrabody mutant that exhibited very high cytoplasmic expression levels in shake flask cultures (0.5 g/l) and fermentation (3.1 g/l), but however, the thermal stability of the mutant was lower than the thermal stability of the parental ScFv [198]. In the previously mentioned approach in which MBP was fused to different ScFvs, it was possible to select intrabodies with improved expression and stability, although the expression levels in shake flask cultures (200 mg/l) [206] were not as high as the expression levels reached by Martineau et al [198].

2.3.4 Optimization of recombinant antibody genes

Sometimes the produced recombinant antibody itself needs to be modified due to various reasons. It may be intrinsically suboptimal in terms of biophysical properties (e.g. isoelectric point, pI, that is difficult for formulation development) [207], it lacks some wanted features (e.g. interaction of aglycosylated Fc-part with Fc-receptor) [202], or it has unwanted features (e.g. post-translational modification sites, PTMs) [207]. Furthermore, in some cases, co-expression of chaperones can cause increased metabolic stress to the cells, thereby leading to lower expression levels [154]. In these situations, the gene of the recombinant antibody itself has to be optimized. Different gene optimization approaches for improving the aforementioned characteristics are described below.

2.3.4.1 Effect of different variable domain combinations

Besides different expression profiles, the antibody variable domains exhibit different biophysical properties like pI, stability, and aggregation propensity [37,207]. In order to find optimal individual domains and domain combinations, Ewert et al. conducted a thorough biophysical characterization of germline consensus domains of seven human VH families (V_H1a, V_H1b, V_H2, V_H3, V_H4, V_H5 and V_H6) and seven human VL families (V_κ1, V_κ2, V_κ3, V_κ4, V_λ1, V_λ2 and V_λ3), which were derived from Human Combinatorial Antibody Library (HuCAL® by MorphoSys) [37,208]. The authors discovered that the best VH and VL domains in terms of expression and stability were the V_H3 and V_κ3 domains. Moreover, the

authors addressed the aggregation problem, which is characteristic for VH domains [209–211], by extending the length of the CDRH3 loop to 17 amino acids. Different VH and VL domain were tested in different combinations in ScFv format, and the most optimal combinations with regard to yield and stability were H3κ3, H1bκ3, H5κ3, and H3κ1. It is also noteworthy that pairing V_H3 with V_λ domains markedly improved the performance V_λ domains, which as individual domains, were poorly expressing and unstable [37].

In another study, different VH/VL combinations were extensively analyzed during the construction of a fully synthetic human Fab antibody library Ylanthia (by MorphoSys) [207]. After in silico analysis of various parameters of individual frequently used VH and VL germline sequences, the authors tested 400 VH/VL pairs (20 VH x 20 VL) of which 95 were subjected to more detailed biophysical analysis. According to the analysis, V_H3 containing Fabs were expressed, on average, more efficiently than Fabs containing other VH family members, albeit the difference was insignificant. However, phage display levels of the Fabs bearing V_H3 were significantly higher than the display levels of Fabs bearing other VH family members [207]. The result is in line with two other studies, in which the V_H3 family was observed to be highly prevalent after selections [212,213]. As for the expression of Fabs containing V_κ or V_λ, the median expression levels of V_λ containing Fabs were significantly higher than the expression levels of V_κ containing Fabs. However, V_κ makes Fabs significantly more stable and resistant to aggregation compared to V_λ [207]. The preference of V_κ over V_λ in therapeutic antibody development was highlighted by a study, in which 137 antibodies in different clinical phases were analyzed [214]. Of the approved therapeutic antibodies analyzed in the study (48 antibodies), only 4 % contained V_λ light chain. Nevertheless, it should be noted that the choice of optimal VL is not necessarily so simple, but it can be affected by the chosen recombinant antibody format [212].

2.3.4.2 Modulation of amino acid and codon sequence

Recombinant antibodies with improved biophysical properties and expression profiles can also be obtained by exploring the effects of various amino acid identities at different positions. For instance, proteins can be engineered more resistant to aggregation by introducing amino acids that increase net charge at specific positions [215]. This finding was concurred in a study, which aimed to increase aggregation resistance of antibody heavy chain variable domains (dAbs) by introducing mutations to the CDR loops. The results indicated that increased aggregation resistance of the dAbs was associated with acidic isoelectric point (5.1±1.1) [216]. Based on the same methodology, Dudgeon et al. introduced mutations to surface-exposed regions of human V_H3 and $V_\kappa3$ domains, and found

out that negatively charged amino acids, especially aspartate, in CDRH1 of V_H3 and in CDRL2 of $V_κ3$ (also position 49 outside CDRL2) contributed to the increased aggregation resistance of the domains. These mutations were also additive, in other words, triple mutants exhibited better biophysical properties and expression than the respective single and double mutants [217].

According to the biophysical properties and the expression, the VH families can be divided into two classes: the good (V_H1, V_H3 and V_H5) and the bad (V_H2, V_H4 and V_H6). By comparing the amino acid sequences of the two classes and by employing structural analysis, Ewert et al. determined the positions of the amino acid residues possibly contributing to the biophysical properties and the expression. Six positions were identified based on the analyses, and single amino acid substitutions were introduced to two ScFvs (2C2 and 6B3) bearing the V_H6 domain, which is suboptimal in many respects [218]. The bad class amino acid was replaced by the good class amino acid at each position. The substitutions at positons (AHo numbering [219]) 5 (Q→V), 16 (S→G), 58 (T→I) and 76 (S→G) enhanced the stability of both ScFvs, whereas substitutions at positions 72 (V→D) and 90 (S→Y) affected the stability of 6B3 ScFv only. Apart from the position 58, the substitutions at other positions improved the expression of both ScFvs. Furthermore, the best result in terms of stability (2C2 = 72.3 kJ/mole; 6B3 = n.d.) and expression (2C2 = 5.1 mg/l; 6B3 = 1.7 mg/l) was obtained when all the mutations were combined [218]. The improved properties were attributed to the effect of residues at positive torsion angle positions on local conformation (16 and 76), removal of unsatisfied hydrogen bonds (58), increased β-sheet forming propensity (5 and 90) and replacing hydrophobic solvent-exposed residues with hydrophilic residues (72) [218]. The deleterious effect of the hydrophobic surface exposed residues has also been described by Forsberg et al. [220]. However, extensive mutagenesis of 45 residue positions (based on positional variability of amino acid residues and entropy) in a Fab fragment gene revealed that at some positions also the "bad class" residues (e.g. from V_H2 and/or V_H6) can be the most optimal ones when it comes to stability [221].

Modulation of codon usage has also been used to increase expression levels of recombinant antibodies [16,95,151,222]. The traditional approach has been codon-optimization of genes. The approach was successfully used to increase the expression level of ScFv fragment by 14-fold [223], but another study that described the codon-optimization of ScFv did not observe any improvement in expression after codon-optimization [151]. In the case of Fab fragments, the codon-optimization was even deleterious, because it rendered Fab fragments toxic to the host cells [16,95]. However, codon harmonization of selected segments of a Fab gene alleviated the toxic effects caused by the codon-optimization, and furthermore, increased the expression levels to 10 mg/l in shake flask culture [95].

2.3.4.3 Enhancing the interaction between aglycosylated Fc-part and FcγR-receptors

Interaction between the Fc-part of an antibody and the FcγR receptors (FcγR) is crucial for the function of the immune system and most of the therapeutic antibodies [224,225]. All the FcγR receptors, except neonatal Fc and C-lectin type receptors (Type II receptors), belong to the group of classical Type I receptors, which include FcγRI, FcγRIIa, FcγRIIb, FcγRIIc, FcγRIIIa, and FcγRIIIb. Of the classical receptors, all but FcγRIIb are activating receptors. The interaction between Type I receptors and Fc-part is markedly affected by the structure of the hinge region, which in turn is affected by antibody subclass and glycosylation of asparagine residue 297 in the CH2 region. However, the neonatal Fc receptor is not sensitive to glycosylation [224].

Since *E. coli* does not have a glycosylation machinery, alternative ways have been explored to enable interaction between the aglycosylated Fc-part and the FcγR receptors. To this end, Jung and colleagues established an Fc library that was screened for binding to FcγRI by using FACS (fluorescence-activated cell sorting). The results indicated that two mutations at positions 382 (E→V) and 428 (M→I) in the CH3 region restored the interaction between aglycosylated Fc part and FcγRI [226], and the importance of the positions regarding the interaction was confirmed by another study as well [202]. Nevertheless, the aglycosylated Fc-part carrying the two mutations was not bound by the other Type I Fc receptors, unlike the glycosylated counterpart, but it retained the pH-dependent interaction with the neonatal Fc receptor. Regardless, the aglycosylated Fc-part mutant was able to induce ~40 % and ~70 % tumor cell killing, when monocyte-derived dendritic cells were mixed with Her-2 expressing breast cancer cells in 25:1 and 100:1 ratios, respectively [226]. In a subsequent study, Jung and colleagues were able to obtain an aglycosylated mutant Fc-part (298 S→G; 299 T→A; 390 N→D; 382 E→V; 428 M→L) that was bound by two FcRIIa alleles with remarkably greater affinity when compared to Herceptin, but unfortunately the affinity towards FcγRI was significantly decreased [227]. Finally, it can be said that recombinant antibodies carrying aglycosylated and engineered Fc-parts have potential also in therapeutic applications.

2.3.5 Secretion of recombinant antibodies in *E. coli*

Gram-negative bacteria have five types of secretion pathways, of which type I and type II pathways are the most popular in recombinant antibody production in *E. coli*. The type I secretion pathway can be used for secretion of proteins into the culture medium in a single-step manner, and signal peptide of α-haemolysin (HlyA) has been the most exploited for that purpose [228]. The type II pathway

includes a post-translational Sec-dependent pathway, a co-translational SRP (signal recognition particle) pathway and a TAT pathway (twin-arginine translocation) (**Figure 8**). Characteristic for the TAT pathway is that it can transport folded proteins across the inner membrane [228,229]. The Sec-dependent pathway has been mostly employed to secrete recombinant antibodies into the periplasmic space, but the SRP pathway can be used as well [230]. The secretion capacity of the SRP pathway is lower than the capacity of the Sec-dependent pathway [188], but in some cases, it can be beneficial to use co-translational transportation, which helps to keep transported recombinant antibodies in transportation competent state.

The SRP pathway outperformed the Sec-dependent pathway in the expression of full-length IgGs [188,231], ScFv [232] and phage displayed single-chain Fab [233]. However, it was observed by Lee et al. that overwhelming the capacity of the SRP pathway in the production of full-length IgG caused reduced fitness of the cells. The fitness was increased when both pathways, SRP and Sec, were used in combination for the secretion of light and heavy chain into the periplasmic space, respectively [188]. Interestingly, similar fitness benefits were not achieved if the heavy chain was secreted via the SRP-pathway and the light chain via the Sec-dependent pathway [234]. Moreover, Zhou et al. discovered that increased hydrophobicity of the hydrophobic region of the heavy chain signal peptide, which uses the Sec-dependent pathway, increased the expression efficiency of full-length IgG [231]. In contrast to the previously mentioned results, a study that compared four Sec-dependent pathway signal peptides and three SRP-pathway signal peptides found out that signal peptides using the SRP-pathway were inferior to the signal peptides using the Sec-dependent pathway in phage display of ScFv. A similar trend was observed in soluble expression as well [229]. Contradictions in the results are not surprising since even the smallest of changes in the signal peptide

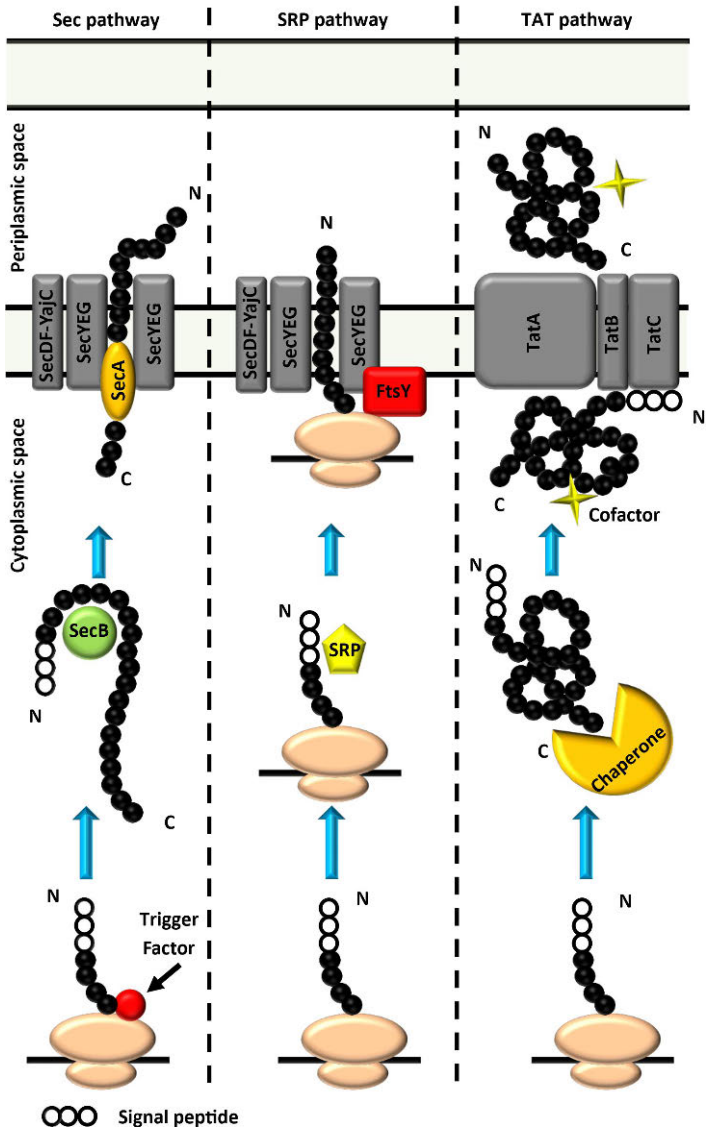

Figure 8. Type II secretory pathways in *E. coli*. Sec pathway is a post-translational pathway where a polypeptide is first fully synthesized, after which a chaperone SecB binds to the polypeptide. SecB keeps the polypeptide in translocation competent state and guides the polypeptide to a membrane-bound protein SecA. The polypeptide is translocated through a channel formed by a SecYEG complex. SecA works as a plunger, pushing the polypeptide through the SecYEG channel. SRP pathway is a co-translational pathway where signal recognition particle (SRP) recognizes the polypeptide when it emerges from the ribosome. The formed complex is then co-translationally transported to the SecYEG complex, where the SRP interacts with a protein called FtsY and the polypeptide is translocated through the channel. TAT pathway transports fully folded proteins. First, the folded polypeptide interacts with a chaperone. Then a co-factor binds to the polypeptide and the polypeptide interacts with TAT machinery. The signal sequence interacts with TatC, and subsequently, the polypeptide is translocated to the periplasmic space.

can have notable effects on expression. For example, only one synonymous codon change in the signal peptide can cause almost 2-fold difference in expression [145]. In general, modulation of codon usage of signal peptides using the Sec-dependent pathway has been an efficient way to increase the expression of Fab fragments [145,235], ScFv fragment [236] and full-length IgG [234].

In addition to the periplasmic secretion, different efforts have been made to secrete recombinant antibodies directly into the culture medium. For example, the addition of 2 % glycine and 1 % Triton X-100 to culture medium increased the secretion of TNF-α fused ScFv to the culture medium by 170-fold [237]. Another study described the successful utilization of HlyA transport system in extracellular secretion of ScFv [238], but also a secretion system based on osmotically induced protein Y has been used to excrete ScFv [239].

2.3.6 Choice of promoter

Among other variables affecting recombinant antibody expression, the choice of promoter is also one of the most important things to consider when optimizing the expression [154,161,240]. A study by Zhong et al. compared the effect of T7, lac, trc, and triple tac promoters on Fab expression, and observed that the triple tac was the most optimal promoter in both *E. coli* strains tested during the study. In contrast, the T7 was the worst promoter in both strains. The lac promoter was the second-best in BL21 (DE3) pRARE2 and the third-best in JM109 [14]. However, the lac promoter was the most optimal in full-length IgG expression among the other tested promoters trc, tet, araBAD and T7 [234], which is partially in line with the study by Zhong et al. In a recently developed system called *RiboTite*, which simultaneously controls transcription and translation, the lac promoter was used together with riboswitch that controls translation [241]. The *RiboTite* system was subsequently used successfully in the expression of three different ScFvs [242]. In addition, by co-expressing mutant RpoD (global sigma factor) and by controlling the transcription with phoA promoter, it was possible to obtain a marked increase in expression of full-length IgGs [243].

2.3.7 N-linked glycosylation in *E. coli*

Over 20 years ago, it was discovered by Szymanski and colleagues that *Campylobacter jejuni* has a general glycosylation system, which is governed by a *pgl* gene cluster [244]. A few years later, Wacker et al. successfully transferred the glycosylation system into *E. coli*, and proved that the system is responsible for N-linked glycosylation of proteins. Furthermore, it was found that oligosaccharyltransferase PglB, which is responsible for glycan transfer, is a key

player in the glycosylation process [245]. PglB is also highly similar to a catalytic subunit of eukaryotic oligosaccharyltransferase STT3, and therefore, the N-glycosylation system of *C. jejuni* is thought to be similar to the respective eukaryotic system [245–248]. Different *E. coli* strains capable of glycosylation and plasmids carrying *pgl* gene clusters have been used to produce glycosylated recombinant antibodies in *E. coli*.

Lizak et al. introduced different versions from the PglB glycosylation consensus sequence D/E-Y-N-X-S/T (Y, X ≠ P) to the linker sequence of ScFv fragment, and discovered that optimized consensus sequence DQNAT [249] in the middle of the linker provided the highest glycosylation efficiency. Subsequent MS analysis showed that the ScFv was homogenously decorated with typical N-glycan of *C. jejuni*. However, when pharmacokinetic properties of the glycosylated ScFv was compared to that of non-glycosylated ScFv, the authors observed that the difference in clearance rate was insignificant [250]. Furthermore, Fisher et al. reported that the introduction of DQNAT sequence at position N297 of the Fc part of IgG enabled efficient glycosylation, but not interaction with $Fc_{\gamma}RI$ [251].

3 Aims of the study

Recombinant binding proteins including antibodies, antibody fragments, and binders based on non-immunoglobulin scaffolds have become increasingly important in various applications. However, it has proven to be challenging to express some recombinant binder formats in sufficient amounts (e.g. Fab fragments) in bacterial hosts, or to display some binders efficiently on the surface of phage particles (e.g. DARPins) when using the Sec-dependent pathway. In many cases, modulation of codon usage has rescued the expression and phage display properties of difficult proteins. Recent studies have shown that codon usage has a significant regulative role in protein translation. Nevertheless, the effects of codon sequence on the regulation of protein expression are still not fully understood, and many features to which codon usage has an effect, are not thoroughly examined. The general aim of the study was to improve expression and phage display properties of synthetic human Fab fragments and DARPins in *E. coli* by means of codon usage modulation and to provide more information on parameters that shape the codon usage in different parts of the genes. More specifically the aims were:

1. To study the effect of synonymous codon usage mutations introduced to different gene segments of the Fab fragment on cell growth, expression and phage display.

2. To explore the effect of different strains and genetic elements (e.g. strain and antibiotic resistance markers) on codon usage.

3. To improve Fab fragment expression and secretion by selecting improved variants from PelB signal sequence libraries containing synonymous codon mutations in the PelB signal sequences of light and heavy chain.

4. To enhance phage display efficiency of DARPins translocated via Sec dependent pathway by selecting improved variants from the synonymous PelB signal sequence libraries.

5. To understand better the mechanisms through which the codon sequence features govern protein expression.

4 Materials and Methods

This chapter provides a summary of the materials and methods used in the study. Detailed descriptions are presented in the relevant original publications **I-III**.

4.1 Cloning of genetic variants

The genetic variants 1, 2, 3, 4 and 5 used in the publication **I**, were produced by replacing codon harmonized segments in a synthetic human anti-digoxigenin Fab fragment gene with respective codon-optimized segment. Cloning of the segments was implemented by using traditional cloning methods. The variants 1, 2, 3, 4 and 5 were produced by using SfiI/SexAI, SexAI/MscI, MscI/PacI, PacI/XhoI and XhoI/SfiI restriction enzyme combinations, respectively. All the restriction enzymes, except SfiI (Thermo Scientific, Waltham, USA), were obtained from New England Biolabs (Ipswich, USA).

4.2 Expression of Fab fragments

Synthetic human anti-digoxigenin (**I**, **II**), anti-microcystin MC-LR (**II**) and anti-SORLA (**II**) Fab fragments were expressed in *E. coli* strains XL1-Blue (*recA1 endA1 gryA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI^{q}ZΔM15 Tn10 (Tet^r)]*) (Stratagene, LaJolla, USA) (**I**, **II**) and BL21 (*B F$^-$ ompT gal dcm lon hsdS$_B$(r$_B^-$ m$_B^-$) [malB$^+$]$_{K-12}$(λ^S)*) (Merck, Darmstadt, USA) (**I**). The heavy chains of the Fab fragments were expressed as C-terminal fusions to 6 x histidine tag from pAK400 [252] (**I**) or pLK04 vectors [253] (**I**, **II**). Overnight pre-cultures were diluted to OD (600 nm) 0.1 in 20 ml (**I**), 5 ml (**II**), 2 x 200 ml (**I**) or 500 ml (**II**) of SB medium (1 % glucose and appropriate antibiotic). The cultures were grown at 37 °C with 300 rpm shaking to OD (600 nm) 0.7–0.9. The appropriate amount of cells were pelleted by centrifugation and re-suspended in fresh SB medium (1 mM IPTG and appropriate antibiotic) to OD (600 nm) 0.5 (**I**, **II**), 0.6 (**II**) or 0.7 (**I**). The cells were cultured at 26 °C with 300 rpm shaking for 3 h (**I**, **II**) or 4 h (**I**). After induction, 1 ml samples were taken from the cultures and the cells were disrupted by sonication (**I**, **II**) or periplasmic extraction (**II**).

The primary screenings (**II)** were started by picking individual colonies to the wells of 96-well plates containing 160 µl of SB medium (1 % glucose and 100 µg/ml amp) and growing overnight at 37 °C with 900 rpm shaking with 70 % humidity. The next day, the wells of 96-well plate containing 200 µl of SB medium (0.05 % glucose and 100 µg/ml amp) were inoculated with 4 µl of overnight cultures. The cells were cultured for 4 h like done above and subsequently 10 µl of 4 mM IPTG was added to the cultures. After the addition of IPTG, the temperature was lowered to 26 °C and cells were grown overnight. The cells were lysed by freeze–thaw method.

4.3 Phage production

The synthetic human anti-digoxigenin Fab fragment (**I**) and an anti-GFP DARPin (**III**) were displayed on the surface of phage particle by expressing the proteins from pEB32x vector [254] as fusions to truncated p3 phage coat protein in XL1-Blue (**I**, **III**) and SS320 (*hsdR mcrB araD139 Δ(araABC-leu)7679 lacX74 galU galK rpsL thi-1 [F' proAB lacI^{q}ZΔM15 Tn10 (Tet^r)]*) [255] (**III**). Phage productions were implemented in 20 ml (**I**, **III**) and 200 µl (**III**) production volumes. In brief, cells were grown to OD (600 nm) 0.5 (20 ml production) or turbidity (200 µl production) in SB medium (1% glucose and appropriate antibiotics) at 37 °C with appropriate shaking and humidity (only in 200 µl production). Subsequently, the cells were infected with VCS-M13 helper phage and incubated 30 (20 ml production) or 45 (200 µl production) min at 37 °C. The cells were pelleted by centrifugation, medium was discarded and the pellets were dissolved in fresh SB medium containing 0.05 % glucose (20 ml production) or no glucose (200 µl production), appropriate antibiotics and $MgCl_2$. The cells were grown overnight at 26 °C (**I**) or 37 °C (**III**) with appropriate shaking and humidity. In 20 ml phage productions, kanamycin was added to the cultures after 1 h from medium exchange, and subsequently the cultivation was continued overnight. In 20 ml phage productions, phages were purified from culture supernatants by PEG/NaCl precipitation (**I**, **III**). The phage titers were determined by OCCA (oligonucleotide-directed chelate complementation assay) [256] (**I, III**) or by total phage immunoassay (**III**).

4.4 Establishment of synonymous signal sequence libraries

The synonymous PelB signal sequence libraries used in the publication **II** were established by introducing synonymous codon mutations to the n-region, hydrophobic region and c-region of PelB signal sequences of light and heavy chain

of the anti-digoxigenin Fab fragment. The respective regions were simultaneously mutated in the light and heavy chain PelB signal sequences, thereby resulting in three libraries. Uridylated single-stranded templates of the n-region, the hydrophobic region and the c-region in pEB32x vector were produced as described by Sidhu et al. [255] with supplemented uridine, and the diversifications of the regions were carried out by using Kunkel mutagenesis method [257] as applied by Huovinen et al. [258]. The remaining template sequences in the n-region, the hydrophobic region and the c-region diversification reactions were digested with BglII (Thermo Scientific), BamHI (Thermo Scientific) and NotI (Thermo Scientific) restriction enzymes, respectively. Subsequently, purified digestion reactions were treated with uracil-DNA glycosylase (UDG) enzyme (Thermo Scientific), which creates abasic sites by removing uracil bases from the templates. Phi29 polymerase used in the sRCA stops at abasic sites, and is thereby unable to amplify the template sequences [258]. The UDG treated Kunkel reactions were amplified by using selective rolling circle amplification (sRCA) [258] in overnight amplification reaction. The sRCA products were digested into linear single-plasmid segments with XhoI restriction enzyme (Thermo Scientific). The XhoI digestions were performed two times. The linear single-plasmid segments were re-circularized in overnight reactions by using T4 DNA ligase (Thermo Scientific). Purified ligation reactions were transformed into SS320 cells by using electroporation and after recovery in SOC medium, the transformation reactions were plated on LA plates (0.5 % glucose and chloramphenicol and tetracycline) and incubated overnight at 37 °C. The next day, cells were collected from the plates.

In order to reduce the amount of non-productive clones in the libraries, they were cloned from pEB32x vector to pEB07 vector [258], which expresses the protein of interest as a C-terminal fusion to the β-lactamase gene. Thus, in the presence of ampicillin, frameshifted or truncated library clones are not viable. The libraries in the pEB32x vector were first amplified by PCR, and subsequently the purified amplicons were digested with ApaI (Thermo Scientific) and XhoI (Thermo Scientific) restriction enzymes. The digested amplicons were cloned into ApaI and XhoI digested pEB07 vector by using T4 DNA ligase. The pEB07 vector initially carried anti-digoxigenin Fab fragment equipped with parental PelB signal sequences, which were removed during the ApaI and XhoI digestion of the pEB07. The libraries in the pEB07 vector were transformed into SS320 cells, recovered in SOC medium (low glucose, 0.05 %) and plated on LA plates (0.05 % glucose and ampicillin). The libraries were then subjected to ampicillin selection by gradually increasing the ampicillin concentration. After the selection procedure, the libraries were amplified by PCR, digested with ApaI and XhoI and ligated to the pLK04 vector by using T4 DNA ligase. The pLK04 vector was used for the screening of the libraries.

The same synonymous PelB signal sequence libraries were also used in the publication **III** to increase the phage display efficiency of DARPins. The n-region and hydrophobic region libraries in pEB32x vector were digested with SfiI (Thermo Scientific) restriction enzyme, and subsequently further digested with XhoI to reduce the Fab fragment background. The SfiI digestion leaves the light chain PelB attached to the pEB32x. A SfiI digested anti-GFP DARPin gene was ligated to the libraries by using T4 DNA ligase, transformed into SS320 cells, recovered in SOC medium and plated on LA plates (0.5 % glucose and chloramphenicol and tetracycline). The next day, the cells were collected from the plates

4.5 Colony PCR for sequencing

Sequencing samples in the publication **II** were produced by using colony PCR method. In brief, cell samples were diluted in MQ water, and Fab fragment genes in the diluted cell samples were amplified by PCR. The PCR reactions were purified by enzymatic purification method in which PCR products are treated with FastAP (Thermo Scientific) and Exonuclease I (Thermo Scientific). The reactions were incubated at 37 °C for 15 min and inactivated at 85 °C for 15 min. Samples were sequenced at Macrogen Inc. (Seoul, South Korea).

4.6 Immunoassays

In this study, immunoassays were used to measure the amount of immunoreactive Fab fragment (**I**, **II**), phage immunoreactivity (**I**, **III**) and total phage amount (**III**). The immunoassays included the following steps: (1) addition of 100 µl of biotinylated target or capture antibody diluted in Assay Buffer (AB) (Kaivogen, Turku, Finland) to the streptavidin coated wells of 96 well plate (Kaivogen), (2) incubation with slow shaking at RT for 30 min (**III**) or 60 min (**I**, **II**), (3) two washes, (4) addition of 100 µl of analyte diluted in AB, (5) incubation with slow shaking at RT 30 min (**III**) or 60 min (**I**, **II**), (6) two washes, (7) addition of 25 ng per 100 µl of europium labeled detection antibody diluted in AB, (8) incubation with slow shaking at RT for 30 min (**III**) or 60 min (**I**, **II**), (9) four washes, (10) addition of 200 µl of Enhancement solution (Wallac, Turku, Finland), (11) incubation with slow shaking at RT for 10 min and (12) measurement with Victor 1420 Multilable Counter (Wallac). The reaction volume in the immunoassays was 100 µl (apart from the addition of Enhancement solution). Biotinylated digoxigenin (100 µl per well of 1 µM bio-dig in AB) was used as a target in the publications **I** and **II**. Biotinylated GFP (30 pmole in 100 µl of AB) was used as a target in the publication **III**. Biotinylated mouse anti-phage antibody (25 ng per 100 µl in

AB) (University of Turku, Turku, Finland) was used as a capture antibody in the publication **III**. Europium labeled anti-human Fab 2A11 (Hytest Ltd, Turku, Finland) was used for the detection of Fab fragments (**I**, **II**) and europium labeled mouse anti-phage antibody was used for the detection of phages (**III**).

4.7 Bioinformatics

Different codon usage metrics and bioinformatics software were used to analyze the relationship between different coding sequence dependent variables and protein expression or phage display efficiency. These metrics and software are introduced in this section

4.7.1 Codon usage metrics

Relative adaptiveness value, w_i, is a method for ranking individual codons (e.g. based on occurrence frequency in highly expressed genes) among the set of synonymous codons encoding the same amino acid [83] (**II**). Codon adaptation index, CAI, is a geometrical mean of the relative adaptiveness values [83] (**II**). CAI tells how much a given gene includes codons that are frequently used in highly expressed genes or in the genome of *E. coli.* Similar metric called tRNA adaptation index, tAI, is also a geometrical mean of the relative adaptiveness values, which in this case however, are based on tRNA abundance and mode of codon-anticodon binding. tAI shows the extent to which gene is adapted to the cellular tRNA pool [85] (**I**).

4.7.2 Bioinformatics software

NUPACK is a web software that is used for the analysis and design of DNA/RNA structures [259]. NUPACK was used in the publication **I** for mRNA secondary structure analysis of genes. Visual Gene Developer is a software that can be used for the analysis of multiple parameters (e.g. codon usage and mRNA secondary structure) [260]. It was used in the publication **II** for codon usage analysis and sliding window analysis of positional mRNA secondary structure. IBM SPSS Statistics 22 was used for the statistical analyses (Armonk, USA) (**I**, **II**, **III**).

5 Results and Discussion

This chapter summarizes the results and discussion presented more in detail in the relevant original publications. The summary of results and discussion follows the chronological publication order of the articles. In the first publication (**I**), the expression and phage display efficiency of the anti-digoxigenin Fab fragment were improved by codon harmonizing the selected segments of the Fab fragment gene. In the second publication (**II**), expression and translocation properties of the anti-digoxigenin Fab fragment were improved by selecting enhanced variants from synonymous PelB signal sequence libraries, which included only synonymous codon mutations. Furthermore, the versatility of the best signal sequence was demonstrated with two other Fab fragments. The synonymous PelB signal sequence libraries were also used to enhance phage display efficiency of an anti-GFP DARPin and a DARPin library (**III**). Previously, phage display of DARPins by using the Sec-dependent translocation pathway had been highly inefficient.

5.1 Codon harmonization of the anti-digoxigenin Fab fragment improves yield characteristics and cell growth of *E. coli*

A codon-optimized synthetic anti-digoxigenin ScFv fragment [261] was initially selected from a synthetic ScFv phage display library [254] and subsequently converted into a Fab fragment by adding codon-optimized human constant light (C_κ) and constant heavy (C_H1) domains to the framework. However, the newly formed codon-optimized synthetic human anti-digoxigenin Fab fragment caused severe toxicity to the host cell, and therefore couldn't be produced (**Figure 9a**). In order to solve the problem, we decided to employ a codon harmonization method, which had been previously applied to enhance the expression and folding of heterologous proteins [87,96,113]. Codon harmonized segments included frame2-CDRL2-frame3 of variable light domain (49–97 by IMGT numbering), constant light domain and constant heavy domain. In addition, two codon pairs, one at the start of the variable light region and another in the heavy chain PelB, were de-optimized. After harmonization of the segments and de-optimization of the codon pairs, the viability of the cells was restored (**Figure 9a**). The codon harmonized

Fab fragment (containing harmonized segments and de-optimized codon pairs) was also relatively highly expressing in shake flask culture, yielding 4 mg of immunoreactive Fab fragment in 2 x 200 ml culture volume. Codon harmonization of the constant domains led to a striking increase in the diversity of codon usage. The diversity of codon usage for the codon-optimized C_H1 and C_κ were 18/61 and 20/61, respectively, whereas the same values for the codon harmonized segments were 40/61 and 44/61 (**Figure 9b**). The difference was not so prominent between the codon-optimized and codon harmonized frame2-CDRL2-frame3 segments.

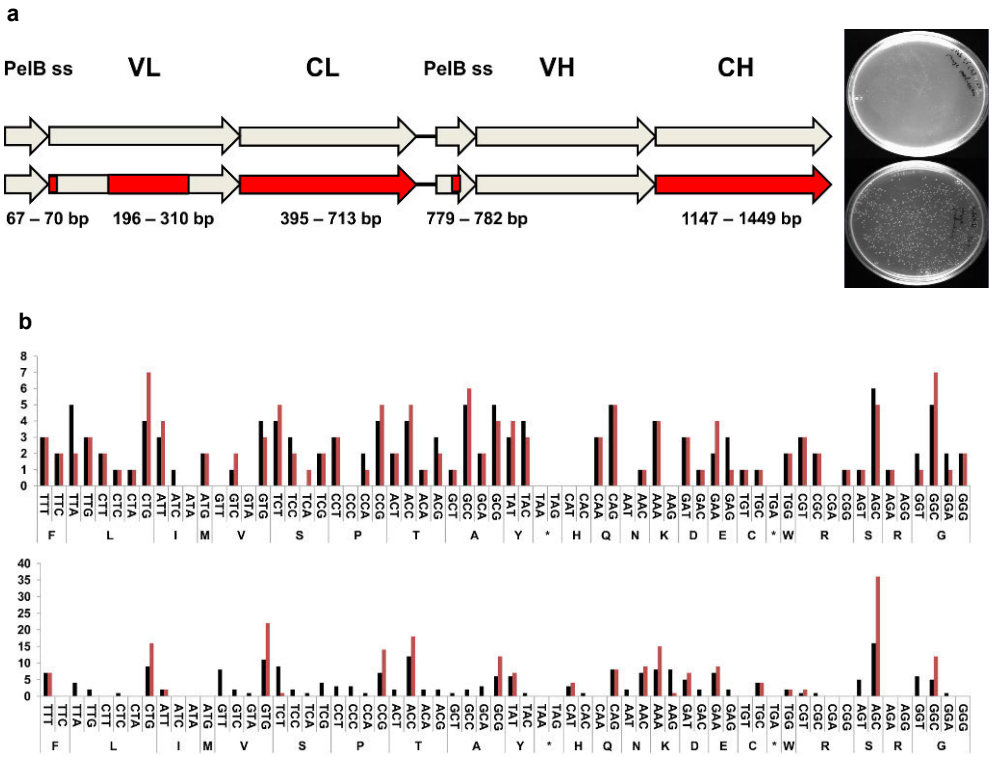

Figure 9. Codon-optimized and codon harmonized synthetic human anti-digoxigenin Fab fragments. (**a**) The upper gene represents codon-optimized Fab gene and the lower codon harmonized Fab gene. Red segments indicate the harmonized/de-optimized segments (positions are indicated below). PelB ss = signal sequence of pectate lyase B, VL = variable light, CL = constant light, VH = variable heavy and CH = constant heavy. The petri dish images show the transformation results of codon-optimized Fab gene (upper) and codon harmonized Fab gene (lower). The Fab fragments were expressed from a bicistronic mRNA. (**b**) The total codon counts in the codon-optimized (red bars) and harmonized variable (black bars) segments (upper diagram) and in the codon-optimized and harmonized constant segments (lower diagram). (Both images are adapted from publication **I**)

After indicating the beneficial effect of codon harmonization on cell growth and Fab fragment yield, we aimed to find out the segment that had the most substantial effect on the aforementioned properties. To this end, codon harmonized and de-optimized segments in the codon harmonized Fab fragment were replaced one-by-one with respective codon-optimized segments by using traditional cloning methods. The cloning procedure resulted in five codon variants, each of which contained four codon harmonized/de-optimized segments and one codon-optimized segment. Hereafter, the codon variants are referred to as variant 1 (optimized codon pair at the start of the Fab gene), variant 2 (codon-optimized frame2-CDRL2-frame3 segment), variant 3 (codon-optimized Cκ), variant 4 (optimized leucine codon pair in the heavy chain PelB signal sequence) and variant 5 (codon-optimized CH1). The codon harmonized Fab containing all the harmonized/de-optimized segments is hereafter referred to as Fab0.

5.1.1 Codon-optimized constant light domain is a highly deleterious segment in many respects

The codon variants and the Fab0 were expressed from pAK400 and pLK04 vectors in XL1-Blue and BL21 strains in four different vector/strain combinations. The vectors pAK400 and pLK04 were identical apart from the antibiotic resistance marker. The vector pAK400 carried chloramphenicol acetyltransferase gene and the vector pLK04 carried TEM-1 beta-lactamase gene. For phage display experiments, the codon variants and the Fab0 were expressed from pEB32x vector in XL1-Blue. The expression and phage display experiments were repeated three times with each vector/strain combination. All the codon variants were compared to the Fab0 in terms of cell growth, the total yield of the immunoreactive Fab fragment and phage display efficiency. During the experiments, it became evident that especially the variant 3 had a highly deleterious effect on all the aforementioned properties. The variant 3 exhibited significantly negative effects on cell growth (i.e. generation times) with almost all the tested vector/strain combinations (except pAK400/BL21) (**Table 1**). The negative effect on cell growth with pLK04/XL1-Blue and pLK04/BL21 combinations was, in fact, so severe that no viable colonies were formed despite six independent transformation attempts. The total yields of the variant 3 were 82 % lower in XL1-Blue and BL21 strain compared to that of the Fab0 (**Table 2**). Furthermore, the variant 3 caused significant (Mann-Whitney U, 2-tailed, $p = 0.002$) decrease in phage display efficiency along with the variant 2 (Mann-Whitney U, 2-tailed, $p = 0.012$) when compared to the Fab0. The results were well in-line with earlier findings made by Lin and colleagues, who described the detrimental effect of codon-optimization of the light chain on the expression of a Fab fragment [16]. The authors found out that

codon-optimization of the light chain increased the amount of insoluble Fab fragment, thereby causing toxic effects on *E. coli*.

Table 1. The average generation times. The average generation times of the harmonized Fab fragment (Fab0), codon-optimized Fab fragment (sFab) and the codon variants from three independent cultures are shown. The standard deviations of the generation times are shown in parentheses.

The average generation times (min)					
Variants	**XL1-Blue**			**BL21**	
	pAK400	**pEB32x**	**pLK04**	**pAK400**	**pLK04**
Fab0	**45 (±0.8)**	**37 (±1.2)**	**56 (±2.2)**	**53 (±4.2)**	**109 (±11.3)**
sFab	**No growth**	**No growth**	**N.D.**	**N.D.**	**N.D.**
Variant 1	**45 (±2.2)**	**40 (±4.5)**	**105 (±73.6)***	**51 (±5.7)**	**107 (±18.4)**
Variant 2	**98 (±6.2)***	**42 (±5.0)**	**80 (±31.4)**	**67 (±2.1)**	**106 (±16.3)**
Variant 3	**77 (±17.1)***	**90 (±44.3)***	**No growth**	**44 (±6.4)**	**No growth**
Variant 4	**45 (±4.5)**	**37 (±1.2)**	**88 (±21.1)***	**64 (±11.3)**	**66 (±17.7)**
Variant 5	**47 (±0.8)**	**44 (±10.7)**	**85 (±39.0)**	**56 (±2.1)**	**108 (±19.1)**

* = Significantly different from Fab0; $p \leq 0.05$
N.D. = not determined
No growth = a given gene could not be cloned or cultivated in liquid culture

5.1.2 The effect of codon usage modulation on expression can depend on used strain and antibiotic resistance marker

During the expression experiments, we observed that some codon variants were strikingly affected by the antibiotic resistance marker of the expression vector. The average total yields of the variant 2 were 79 % lower than the average total yields of the Fab0 with pAK400/XL1-Blue combination, but when the variants and the Fab0 were expressed from pLK04 vector in XL1-Blue strain, the average total yields of the variant 2 increased by 111 % compared to that of the Fab0 (**Table 2**). However, the average total yields of the variant 2 decreased significantly compared to the Fab0, when expressed from the pLK04 vector in BL21 strain (**Table 2**). The change of antibiotic resistance marker had even more dramatic effect on the expression of the variant 5 (**Table 2**). With pAK400/XL1-Blue combination, the variant 5 exhibited similar expression levels compared to Fab0, but with pLK04/XL1-Blue combination, the expression of the variant 5 was severely impaired and couldn't be reliably detected (**Table 2**). Lastly, in terms of expression, the variant 1 and variant 4 showed subtle preference for pAK400/BL21

and pLK04/XL1-Blue combinations, respectively (**Table 2**). The results indicate that modulation of codon usage for optimal heterologous expression is not straightforward, but different factors can potentially affect the optimality of the codon usage, thereby making the prediction of the outcome of the codon usage modulation more difficult.

Table 2. The average total yields of immunoreactive Fab fragment. The average total yields of immunoreactive Fab fragment of the harmonized Fab fragment (Fab0), codon-optimized Fab fragment (sFab) and the codon variants from three independent cultures are shown. The standard deviations of the total yields are shown in parentheses.

The average total yields of immunoreactive Fab fragment (µg/ml)				
Variants	**XL1-Blue**		**BL21**	
	pAK400	**pLK04**	**pAK400**	**pLK04**
	Total yield	**Total yield**	**Total yield**	**Total yield**
Fab0	**2.62 (±1.04)**	**1.68 (±0.21)**	**6.15 (±0.90)**	**7.80 (±2.78)**
sFab	**N.D.**	**N.D.**	**N.D.**	**N.D.**
Variant 1	**2.83 (±0.83)**	**1.94 (±0.84)**	**7.72 (±1.61)***	**7.92 (±2.26)**
Variant 2	**0.58 (±0.12)*****	**3.54 (±1.32)*****	**7.69 (±1.96)**	**2.49 (±3.38)****
Variant 3	**0.47 (±0.04)*****	**N.D.**	**1.08 (±0.26)****	**N.D.**
Variant 4	**2.9 (±1.55)**	**2.40 (±0.39)****	**7.26 (±1.48)**	**9.00 (±4.09)**
Variant 5	**2.18 (±0.53)**	**<0.0015*****	**7.16 (±1.27)**	**7.24 (±3.23)**

* = Significantly different from Fab0
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$
N.D. = not determined

5.2 Improved expression of Fab fragments through screening combinatorial synonymous signal sequence libraries

In the second publication, we studied how synonymous codon mutations introduced into the light chain and the heavy chain PelB signal sequences affect the yields of immunoreactive Fab fragments expressed from a bicistronic mRNA. To this end, we used previously mentioned codon harmonized anti-digoxigenin Fab fragment (Fab0) as a framework in order to establish three combinatorial signal sequence libraries, which contained synonymous codon mutations in the n-region (N–0 library), hydrophobic region (H–0 library), or c-region (C–0 library) of the light and heavy chain PelB signal sequences (**Figure 10**).

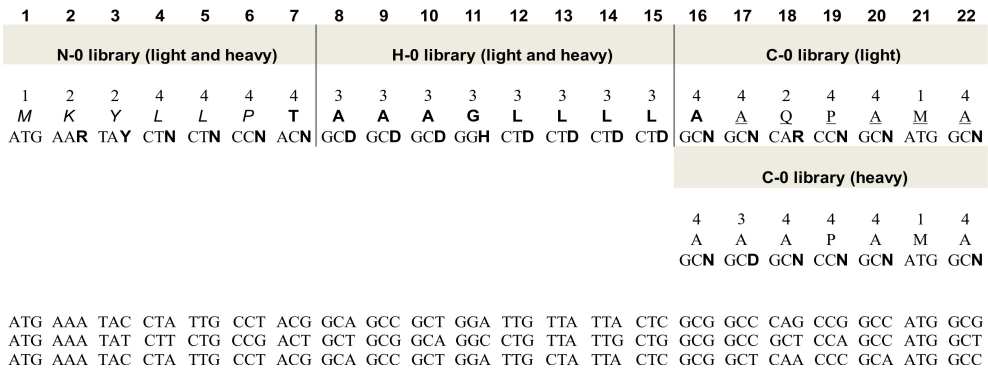

Figure 10. The design of the signal sequence libraries. The boundaries of each library are marked with vertical bars. The amino acid sequence is shown above the codon sequence. Numbers above the amino acid sequence illustrate the positional diversity. Bold numbers above the table show the position number for each amino acid in the sequence. Letters in italics denote the natural n-region, bold letters denote the natural hydrophobic region and underlined letters denote the natural c-region. The codon sequences below the randomization schemes of the libraries are parental light chain PelB (upper), parental heavy chain PelB (middle) and wild-type PelB (lower). **H** = C, T, A; **D** = A, G, T; **Y** = C, T; **V** = A, G, C; **R** = A, G. (Adapted from the publication **II**).

In each library, the given region was simultaneously diversified in both PelB signal sequences. The libraries were fused to beta-lactamase gene (pEB07 vector) and cultured in the presence of ampicillin in order to decrease the number of non-productive signal sequence clones (frameshifted or truncated clones) and thereby enrich productive signal sequence clones in the libraries. After the enrichment of the productive signal sequence clones, the libraries were cloned into the pLK04 expression vector, and 279 clones from each library (837 in total) were cultured and expressed on 96-well plates in 200 µl culture volume. The impact of synonymously mutated signal sequence clones on the expression of immunoreactive Fab fragment was analyzed by using time-resolved fluorometry based immunoassay, and subsequently ranking the signal sequence clones based on fluorescence intensities. Top 10 % cohort included 40 clones from the N–0 library, 21 clones from the H–0 library, and 23 clones from the C–0 library. From the top 10 % cohort, all the signal sequence clones bearing correctly mutated signal sequences (including only planned synonymous mutations) in either one or both chains were selected for more tightly controlled secondary screening implemented in 5 ml culture volume. The secondary screening included 30, 10 and 18 signal sequence clones from the N–0, H–0 and C–0 libraries, respectively (**Figure 11**). Of the secondary screened N–0 library signal sequence clones, 43 % exhibited at least twofold, and 17 % at least threefold increase in Fab fragment expression compared to the Fab0 (carrying parental PelB signal sequences). Furthermore, the best signal

sequence clone (N–0 2), which exhibited over fivefold increase in Fab fragment expression compared to the Fab0, was obtained from the N–0 library. In the H–0 library, 40 % and 0 % of the clones exhibited at least twofold or threefold increase in expression compared to the Fab0, respectively. Respective percentages for the C–0 library were 22 % and 6 %. The fraction of clones showing decreased expression in N–0, H–0 and C–0 libraries were 23 %, 30 % and 22 %, respectively. Moderately high fraction of clones showing decreased expression in each library highlights the difficulty of controlling the variation of the expression levels in the primary screening campaigns implemented in 96-well format. The high variation arises from the challenges to normalize the cell number and the growth phase of the cells in each well before induction.

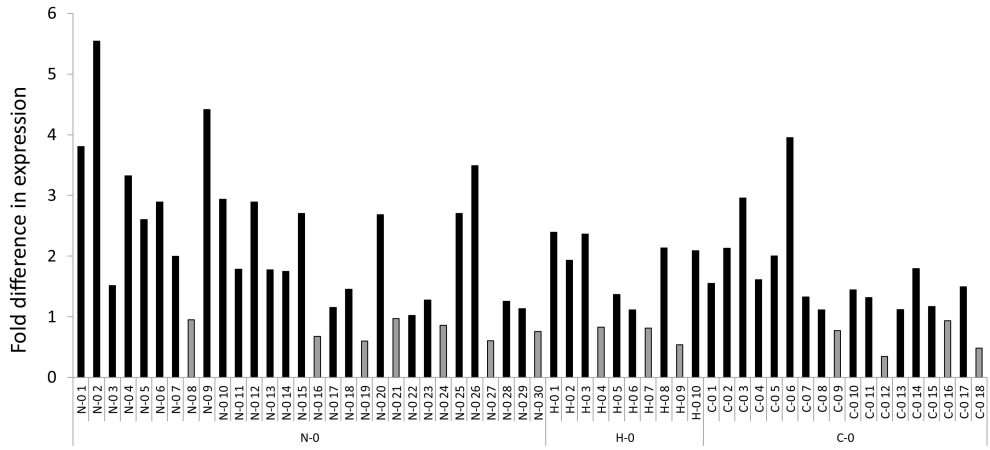

Figure 11. Fold difference in expression compared to the Fab0. The libraries and clones of each library are shown on the x-axis and separated by vertical lines. Gray bars indicate decreased expression compared to the Fab0.

The best signal sequence clone N–0 2 was expressed in triplicate in 500 ml culture volume to demonstrate that the results are also applicable in more relevant production scales. On average, the PelB codon sequences of the N–0 2 increased the total yields of immunoreactive Fab fragment by more than threefold compared to the Fab0. Furthermore, the PelB codon sequences of the N–0 2 significantly improved Fab fragment translocation efficiency as 78 % of the total amount of the immunoreactive Fab fragment was found in the periplasm, whereas in the case of the Fab0, the translocation efficiency was only 56 % (T-test, Equal variances assumed, Sig. 2-tailed $p = 0.005$). Finally, the light chain and the heavy chain PelB codon sequences of the N–0 2 were cloned into two other Fab fragments, anti-microcystin Fab and anti-SORLA (sortilin-related receptor) Fab, in order to see if

the improved PelB codon sequences can be used to increase the expression of other Fab fragments as well. Indeed, the introduction of the PelB codon sequences of the N–0 2 increased the expression in the case of both Fab fragments. The yields of the immunoreactive anti-microcystin Fab fragment equipped with N–0 2 PelBs increased by 3.9-fold compared to the anti-microcystin Fab fragment equipped with parental PelBs. The yields of the immunoreactive anti-SORLA Fab fragment equipped with parental PelBs were below the reliable detection limit of the employed immunoassay, but the N–0 2 PelBs increased the yields clearly above the detection limit. Together the results show that the modulation of the codon usage of the signal sequence is a potential approach for increasing the expression of Fab fragments with different target specificities.

5.2.1 Synonymous codon identity at fifth leucine position is especially important for Fab fragment expression

The relationship between codon usage of the PelB signal sequence and the expression levels was analyzed by using different codon usage metrics and statistical methods. The analyses revealed that synonymous codon identity at fifth leucine position located in the n-region of the PelB signal sequence (N–0 library) had a significant effect on the expression of immunoreactive Fab fragment. The relationship between synonymous codon usage at fifth leucine position and the Fab fragment expression was found by plotting the expression levels of the clones against relative adaptiveness values (w_i) of codons at each codon position in each library and analyzing the significance of the correlations with Spearman rank correlation coefficient (**Figure 12**).

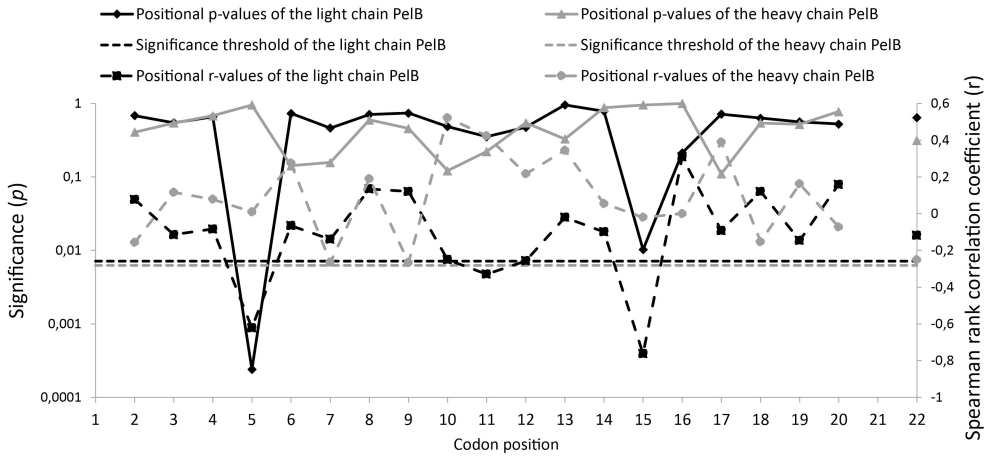

Figure 12. Correlation between the relative adaptiveness values and the expression levels of immunoreactive Fab fragment. Black line (with diamond markers) indicates the *p*-value at given position in the light chain PelB and grey line (with triangle markers) indicates the *p*-value at given position in the heavy chain PelB. Dashed black line (with square markers) indicates the Spearman rank correlation coefficient at given position in the light chain PelB and dashed grey line (with sphere markers) indicates the Spearman rank correlation coefficient at given position in the heavy chain PelB. Horizontal dashed black and grey lines indicate the levels below which the correlations are significant after Bonferroni correction for multiple correlations in the light chain PelB (0.00714) and in the heavy chain PelB (0.00625) respectively. Positions where solid black or grey line goes below the dashed vertical lines are considered as statistically significant. (Adapted from publication **II**)

Further examination of the fifth leucine position was implemented by pooling the N–0 library clones based on the codon identity at the fifth leucine position and comparing the average expression levels between the pools. The further examination showed that the most optimal leucine codon CTG caused a significant decrease in the expression of immunoreactive Fab fragment compared to leucine codons CTA (Kruskal-Wallis Adj. Sig. $p = 0.005$) and CTT (Kruskal-Wallis Adj. Sig. $p = 0.035$). CTA codon is one of the rare codons in *E. coli* [7,90]. The finding was confirmed by using Leu-5 CTA and Leu-5 CTG single-codon variants, which differed from each other only in terms of synonymous codon identity at the fifth leucine position. The Leu-5 CTA single codon variant exhibited on average 1.8-fold higher expression of immunoreactive Fab fragment than the Leu-5 CTG single-codon variant (T-test, Equal variances assumed, Sig. 2-tailed $p = 3.62 \times 10^{-9}$). The average expression levels of the single-codon variants were also compared to that of respective codon pools, but no significant differences were observed. The result indicates that synonymous codon identity at the fifth leucine position is indeed an important factor affecting the expression levels. The Leu-5 CTA single-codon variant was also compared to a variant having a rare [Leu-4]-[Leu-5]-[Pro-6]

codon run, because previous studies have described the beneficial effects of rare/non-optimal codons in the signal sequences on protein expression [142–144]. Surprisingly, the rare codon run variant exhibited 1.3-fold lower expression of immunoreactive Fab fragment than the Leu-5 single codon variant (T-test, Equal variances assumed, Sig. 2-tailed p = 0.000016). Furthermore, no significant correlations were observed between the expression levels and CAI values of the three signal sequence regions. The data shows that in terms of protein expression, the position of the rare/non-optimal codon is more important than the sheer number of rare/non-optimal codons.

5.2.2 Secondary structures of the mRNA in the translation initiation regions of the light chain and the heavy chain affect Fab fragment expression

In addition to the analysis of the relationship between the codon usage and the expression of immunoreactive Fab fragment, we also analyzed the impact of local mRNA secondary structures on the expression of immunoreactive Fab fragment. The mRNA sequences of the signal sequence clones in each library were folded by using the "mRNA profile v2" module of the Visual Gene Developer, which employs sliding window analysis in order to determine local Gibbs free energy of the mRNA at each base position. Then the expression levels of the clones were plotted against the positional Gibbs free energy in each library, and the significance of the correlations was analyzed as described previously. The sliding window analysis indicated that reduced mRNA secondary structures in 30 nt region centered around 155–157 and 159–160 nt correlated significantly with increased expression of immunoreactive Fab fragment in the N–0 library (**Figure 13**). The positions that correlated significantly with the expression of immunoreactive Fab fragment were located in the translation initiation region between the Shine-Dalgarno sequence and the start codon. This result was in line with the previous studies, which have reported the same relationship between decreased mRNA secondary structures and increased protein expression in the translation initiation region [99–101]. Surprisingly, however, reduced mRNA secondary structures in 30 nt region centered around 886 nt, which was located in the proximity of the heavy chain Shine-Dalgarno sequence, correlated significantly (r = –573, p = 0.0009; Bonferroni corrected alpha level = 0.0013) with decreased expression of immunoreactive Fab fragment in the N–0 library.

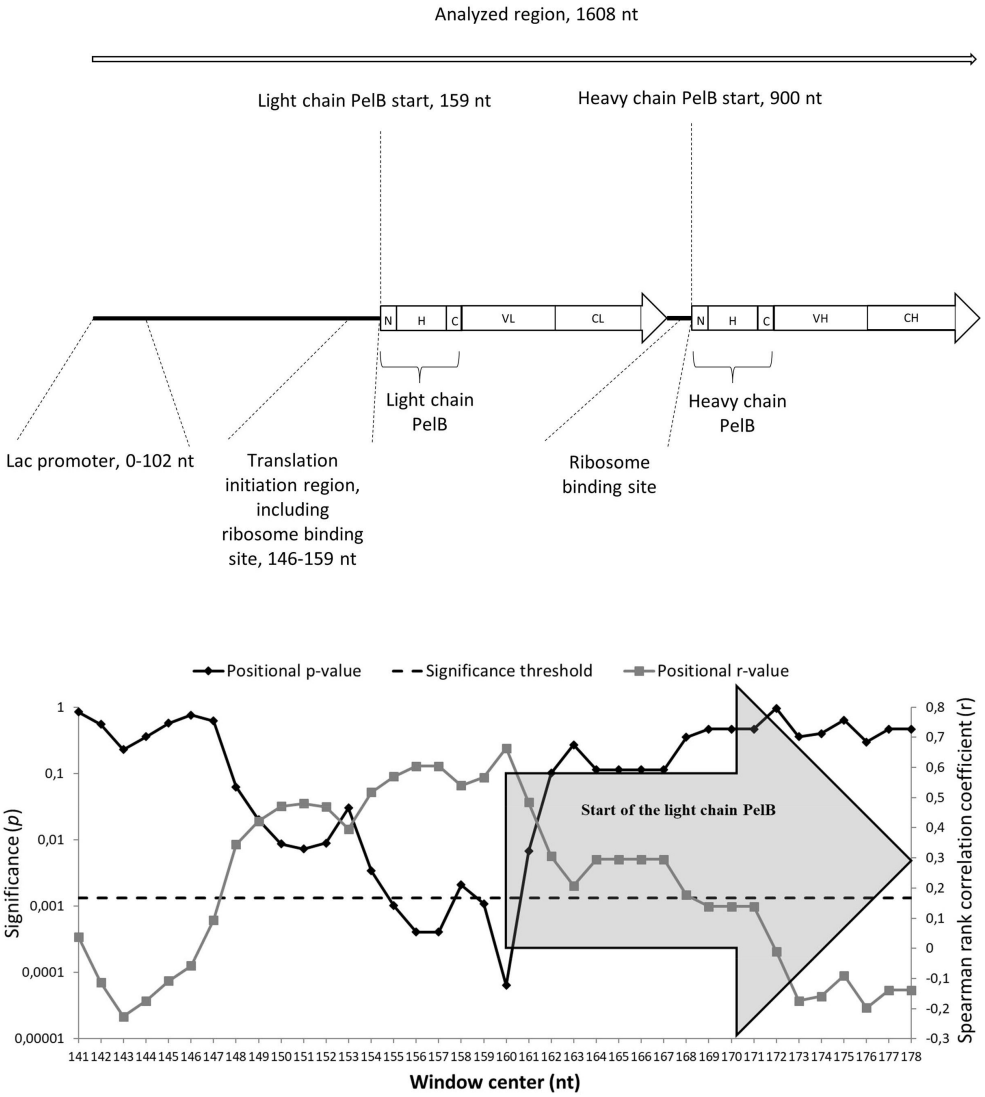

Figure 13. The mRNA secondary structure analysis of the library clones. The upper figure shows the folded region. The arrow above the illustration of the bicistronic Fab fragment gene indicates the length of the analyzed region. Positions of the Lac promoter, translation initiation region, the light chain PelB start and the heavy chain PelB start are shown. The lower figure shows the correlation between positional Gibbs free energy of mRNA secondary structure and Fab expression in the light chain region of the N-0 library. Black line indicates the *p*-value at given position and dashed black line indicates the level below which the correlation is significant. Grey line indicates the Spearman rank correlation coefficient value at given position. The grey arrow denotes the start and the translational direction of the light chain PelB. (Modified from publication **II**)

We also observed that when rare leucine codon CTA resided at the fifth leucine position, reduced mRNA secondary structures in the translation initiation region

correlated significantly with increased expression of immunoreactive Fab fragment (r = 0.845, p = 0.034). A similar trend was not observed with other analyzed leucine codons, indicating that the effect was codon specific. The relationship between mRNA secondary structures in the translation initiation region and codon usage at the fifth leucine position was further explored by using single-codon variant Leu-5 CTA and a new single-codon variant Leu-5 TTA. Both variants had exactly the same Gibbs free energy profile, but different relative adaptiveness values (CTA, w_i = 0.0724; TTA, w_i = 0.2636). The expression levels of immunoreactive Fab fragment produced by the single-codon variant Leu-5 CTA were on average (five independent expressions) 26 % higher than the respective expression levels of the single-codon variant Leu-5 TTA after 3 h of induction. The difference between the variants was significant as well (T-test, Equal variances assumed, Sig. 2-tailed p = 0.000004). Previously it has been suggested that AT-rich rare/non-optimal codons are enriched at the N-terminus of the genes to reduce mRNA secondary structures at the start of the genes [99–101], but our data implied that reduced mRNA secondary structures in the translation initiation region and the occurrence of a rare codon at the fifth leucine position might have additive but independent roles, at least in signal sequences. Previously Zalucki et al. had hypothesized that the combination of rapid translation initiation (reduced mRNA secondary structures in the translation initiation region) and rare/non-optimal codons in the signal sequence could be associated with higher efficiency in protein translocation or recycling of chaperones [146].

5.2.3 Synonymous alanine codon pairs in the hydrophobic region of the heavy chain PelB increase Fab fragment expression

Signal sequence clones in the H–0 library carrying synonymous alanine codon pairs [Ala-8]–[Ala-9] in the hydrophobic region of the heavy chain PelB signal sequence exhibited, on average, almost twofold higher expression of immunoreactive Fab fragment than signal sequence clones in the H–0 library without synonymous alanine codon pairs at the respective position. The difference between the two groups was also statistically significant (T-test, Equal variances assumed, Sig. 2-tailed p = 0.006). In general, signal sequence clones in the H–0 library had a high frequency of synonymous codon pairs. The high frequency of synonymous codon pairs in the hydrophobic region might be related to the structural sensitivity of the region. Cannarrozzi et al. reported that synonymous codon pairs can increase translation speed up to 30 % [262], and Pechmann and Frydman, in turn, reported that hydrophobic regions have a high frequency of codons, which are translated rapidly [263]. Therefore, it is plausible that the purpose of the synonymous codon pairs in the hydrophobic region is to increase the translation speed.

5.3 Enhancement of Sec dependent translocation of DARPins by using synonymous PelB signal sequence libraries

Many previous studies have shown that Sec dependent translocation of stable and fast-folding proteins is highly inefficient due to the post-translational nature of the pathway, but the translocation efficiency of these proteins can be substantially improved when they are directed to the co-translational SRP pathway [19,264–266]. This also applies to the phage display of DARPins, which are highly stable and fast-folding proteins. Steiner et al. tested several different signal sequences (including PelB), which directed the transported DARPins either to the Sec pathway or to the SRP pathway, in order to achieve efficient phage display of DARPins. The authors observed that the co-translational SRP pathway was superior to the post-translational Sec pathway when it came to the translocation of DARPins. Among the signal sequences using the SRP pathway, the DsbA signal sequence was the most efficient [19]. Nevertheless, the Sec dependent translocation pathway has a greater transportation capacity than the SRP pathway [188], which is why improved translocation of the stable and fast-folding proteins via the Sec pathway would be highly attractive. The high transportation capacity is desirable because the amounts of the expressed protein of interest can sometimes be extremely high [2]. These points coupled with the good results obtained from the previous study, in which the expression and Sec dependent translocation of Fab fragments were improved by using synonymous PelB signal sequence libraries, encouraged us to explore the potential of the libraries in enhancing the Sec dependent translocation of an anti-GFP DARPin, and thereby improve Sec dependent phage display efficiency of the DARPin. The anti-GFP DARPin gene was cloned into two previously established synonymous PelB signal sequence libraries, of which one contained synonymous codon mutations in the n-region (DARPin N library) and another in the hydrophobic region (DARPin H library). By primary screening 900 clones from both libraries and secondary screening selected clones in five replicates, we were able to identify clones that exhibited significantly improved phage display of the anti-GFP DARPin compared to the parental PelB. In total, nine clones exhibited significantly improved phage display levels. Seven clones (DN1, -2, -5, -6, -8, -9 and -10) (**Figure 14 a**) were found from the DARPin N library and two (DH4 and DH7) (**Figure 14 b**) from DARPin H library. The best DARPin N library clone DN5 and the best DARPin H library clone DH4 showed on average over 9-fold (Mann-Whitney U, Asymp. 2-tailed sig. $p = 0.009$) and 10-fold improvement (Mann-Whitney U, Asymp. 2-tailed sig. $p = 0.009$) in phage display of the anti-GFP DARPin, respectively, when compared to the parental PelB. Astonishingly, the best DARPin N library clone exhibited significant, 1.8-fold (Mann-Whitney U, Asymp. 2-tailed sig. p =

0.027) improvement in display of the anti-GFP DARPin also when compared to the DsbA signal sequence, which directs proteins of interest to the SRP pathway. However, in addition to the planned synonymous mutations, the best DARPin N library clone DN5 included one unexpected A8V mutation. Also, it should be noted the average display levels of the best DARPin H library clone DH4 were better (~1.6-fold) than the average display levels of the anti-GFP DARPin carrying DsbA signal sequence, although the difference was not significant.

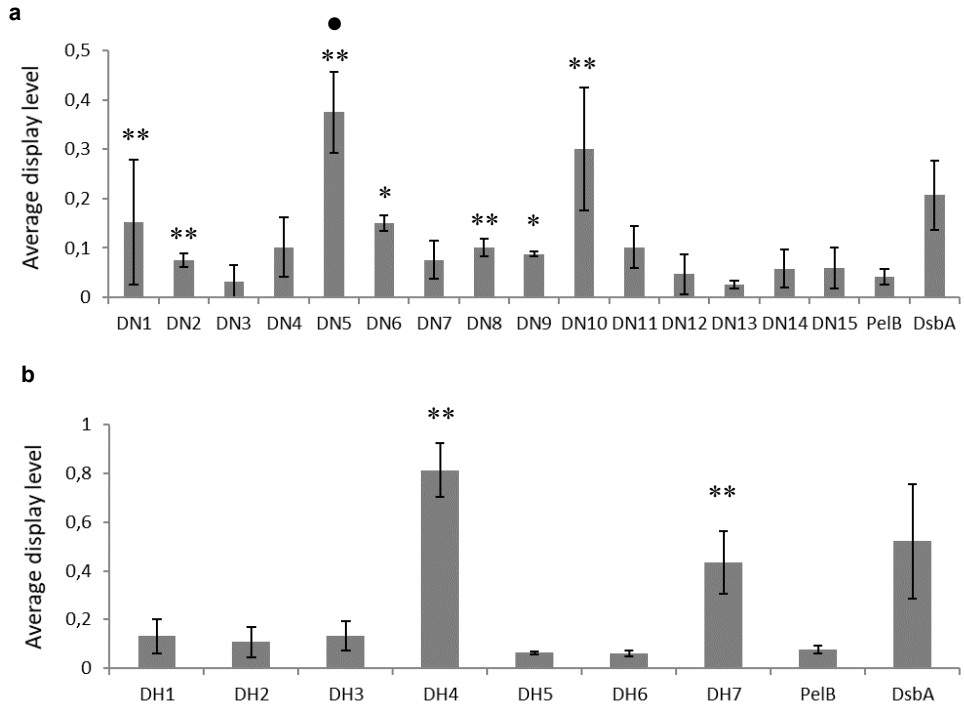

Figure 14. Average display levels of the secondary screened PelB signal sequence clones. (a) DARPin N library and (b) DARPin H library. The display level is the time-resolved europium signal measured from the immunoassay detecting the specific phage binding to the biotinylated GFP divided by the signal measured from total phage immunoassay. Error bars represent standard deviation of five replicate phage productions (exceptions DN4, 6, 9, 11, 14; n = 4) Significant differences to the parental PelB and DsbA are shown by asterisk (*) and sphere (●), respectively. Mann-Whitney U, *(●) $p < 0.05$; **(●●) $p < 0.01$.

We also tested the capability of the best signal sequence clones to improve the display of a DARPin library on the surface of phage particles by fusing DN5, DN10 and DH4 signal sequences to a DARPin library containing serine/tyrosine binary diversity at 23 positions. In order to analyze whether n-region and hydrophobic region mutations would have additive effects on display levels, two combinatory mutants (DN5–DH4 and DN10–DH4) were created. The DN5–DH4

combination was designed not to contain A8V substitution of DN5 to exclude the effect of the substitution on display levels, and to investigate only the impact of synonymous mutations. In general, all tested PelB signal sequence clones exhibited statistically significant (according to Mann-Whitney U test) improvement in display of the binary library compared to the parental PelB. Clones DN5, DN10, DH4, DN5–DH4 and DN10–DH4 showed ~7-fold (p = 0.008), ~2-fold (p = 0.032), ~4-fold (p = 0.008), ~3-fold (p = 0.008) and ~6-fold (p = 0.008) increase in display, respectively, when compared to the parental PelB (**Figure 15**). Furthermore, clones DN5 and DN5–DH4 significant improvement compared to the DsbA signal sequence (**Figure 15**). Both signal sequences improved the display of the binary library by almost 2-fold (p = 0.008) when compared to DsbA. The combinatory clone DN5–DH4 didn't show any significant improvement in display efficiency when compared to the DN5 indicating that the A8V substitution was not the determining factor for higher display and that n-region and hydrophobic region sequences did not have a synergistic effect on the display level. The variant DN10 was the only variant exhibiting significantly lower display level of binary library DARPins compared to DsbA (**Figure 15**), which is surprising since the clone DN10 was slightly better in displaying anti-GFP DARPin than DsbA. Furthermore, the DN10–DH4 showed decreased display levels of the binary library compared to the DH4 alone, which together with the DN5–DH4 results suggest that n-region codon usage is a major factor in dictating the display efficiency.

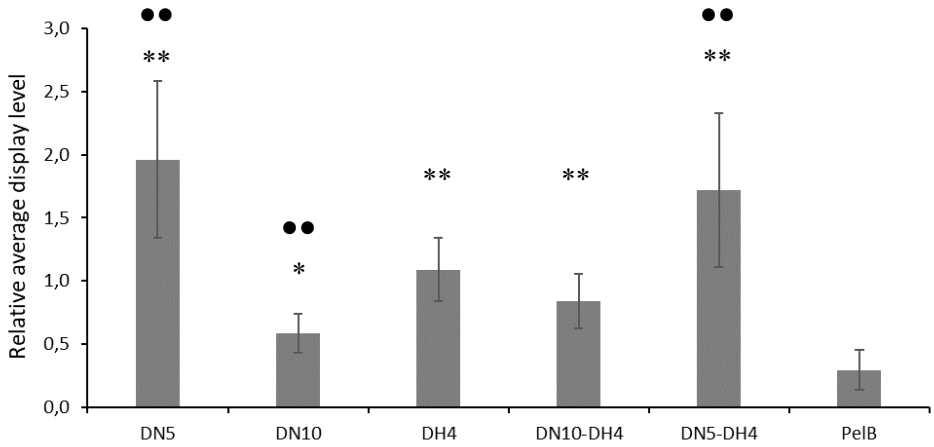

Figure 15. Effect of PelB signal sequence clones on average display levels of the DARPin serine/tyrosine binary library. All the average display levels are relative to the display levels of binary library displayed with DsbA signal sequence. Error bars represent standard deviation of five independent phage productions. Significant differences to the parental PelB and DsbA are shown by asterisk (*) and sphere (●), respectively. Mann-Whitney U, *(●) $p < 0.05$; **(●●) $p < 0.01$.

6 Conclusions

Different binding proteins, such as antibodies, antibody fragments, and many binders based on non-immunoglobulin scaffolds, have already established themselves in therapeutics, diagnostics, and basic research, but they are becoming increasingly important in such applications. For example, it has been predicted that therapeutic antibody markets alone will grow up to 15 % by 2022, reaching a worth of 200 billion [267]. Therefore, efficient engineering and production of the binding proteins is vital. During the recent years, researchers have discovered that codon usage is a major factor dictating the translational efficiency. Nevertheless, many codon usage related factors are still not very well understood, and the effects of codon usage on the expression of some highly relevant proteins are not sufficiently examined or not examined at all. In this thesis, we studied the impacts of codon usage on the expression and phage display efficiency of two relevant binding protein formats, Fab fragments and DARPins, in *E. coli*. We successfully improved the expression and phage display efficiency of Fab fragments as well as phage display efficiency of the anti-GFP DARPin and DARPin library through modulation of codon usage. Based on the results obtained from the original publications, the following conclusions can be drawn:

(I) The codon harmonization approach is an efficient way to drastically increase the expression of Fab fragment, even if it was initially highly toxic to the host cell. Furthermore, the outcome of the modulation of codon usage can depend on the employed strain and antibiotic resistance marker, which suggests, that one gene design is not necessarily the most optimal one in all conditions.

(II) By screening improved synonymous PelB signal sequences from the synonymous PelB signal sequence libraries, the expression levels of different Fab fragments can be efficiently increased, and simultaneously provide new design parameters regarding the PelB signal sequence.

(III) The synonymous PelB signal sequence libraries can also be used to improve the display efficiency of Sec pathway translocated DARPins. Previously, the translocation of the stable and fast-folding DARPins via the post-translational Sec pathway has been highly inefficient. Moreover, the results suggest that synonymous PelB signal sequence libraries can be used to improve secretion of proteins, which are very different from each other.

In conclusion, it can be argued that the modulation of codon usage is a worthy approach among the methods used to improve protein expression and phage display efficiency. The modulation of codon usage is an attractive way to increase protein expression when, for example, amino acid sequence cannot be altered (due to change in physicochemical properties), or one cannot add anything to the system (e.g. chaperone co-expression). This thesis provides new ideas and ways to improve protein expression and phage display efficiency by means of codon usage modulation.

Acknowledgements

This study was carried out at the Department of Biotechnology, University of Turku, during the years 2016–2020. Financial support from the Doctoral Programme in Molecular Life Sciences (DPMLS) and Business Finland is gratefully acknowledged.

I want to thank Professor Kim Pettersson, Professor Tero Soukka and, Associate Professor Urpo Lamminmäki for giving me the chance to pursue a doctoral degree at the Department of Biotechnology as well as for providing nice, relaxed and innovative working atmosphere at the Department.

I am deeply grateful for my supervisors Associate Professor Urpo Lamminmäki and University Teacher Tuomas Huovinen for all the support and guidance you have given me during this project. The balance of freedom and responsibility that you gave me was just right, and it helped me to grow as a confident professional in the field of antibody engineering. Furthermore, I want to say that your comprehensive knowledge on antibody engineering and positive attitude truly inspired me. Thank you! In addition, I want to express my appreciation for my thesis steering committee members Dr. Tony Wahlroos and Dr. Jan Weegar.

A big thank you goes to the pre-examiners Dr. Johan Nilvebrant and Dr. Tommi Kajander for reviewing this thesis, and for all the valuable comments.

At the Department of Biotechnology, I have had a privilege to work with great colleagues. I want to thank Priyanka, Sultana, Hanna, and Eeva for sharing the office with me and for creating a nice atmosphere. Also, I want to thank Niklas, Parvez and Sherif for long Friday lunches at Ilo restaurant, Champions League match day studios at Pikku-Torre, and an unforgettable trip to Barcelona. Lastly, I owe Satu Lahtinen a big thank you for helping me with all the practicalities regarding graduation, dissertation, and printing of the thesis.

I also want to thank Matias Lappalainen for implementing the essential experiments in for publication III.

I am forever thankful for my mom, dad, sister Ilona and brother Tuomas for a never-ending support and unconditional love. I couldn’t have made it this far without you. You have always been there for me whenever I have needed you, and encouraged me in everything whatever I have decided to do. Special thanks go to

my three-year-old sororal nephew, Lenni. Playing with you always reminds me of things that are most important in life.

I have always been surrounded by great friends, of which I want to thank my closest friends Lauri, Heikki, Antti, and Juho. Spending time with you has always helped me to detach myself from work. Thank you for the unforgettable moments, great discussions, and NHL tournaments.

Last but definitely not least I want to thank my girlfriend, Jenna. My words can't express the gratitude I am feeling. During this project, you have always been the first person to encounter my moments of joy and desperation, and thus experienced this journey with me as authentic as it gets. Thank you for always being there for me and understanding me.

20.4.2020

Antti Kulmala

List of References

1. Elena C, Ravasi P, Castelli ME, Peirú S, Menzella HG. Expression of codon optimized genes in microbial systems: Current industrial applications and perspectives. Front Microbiol. 2014;5: 1–8. doi:10.3389/fmicb.2014.00021
2. Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. Trends Biotechnol. 2004;22: 346–353. doi:10.1016/j.tibtech.2004.04.006
3. Nelson AL. Antibody fragments: hope and hype. MAbs. 2010;2: 77–83. doi:10.4161/mabs.2.1.10786
4. Schirrmann T, Meyer T, Schütte M, Frenzel A, Hust M. Phage display for the generation of antibodies for proteome research, diagnostics and therapy. Molecules. 2011;16: 412–426. doi:10.3390/molecules16010412
5. Quax TEF, Claassens NJ, Söll D, van der Oost J. Codon Bias as a Means to Fine-Tune Gene Expression. Mol Cell. 2015;59: 149–161. doi:10.1016/j.molcel.2015.05.035
6. Plotkin JB, Kudla G. Synonymous but not the same: The causes and consequences of codon bias. Nat Rev Genet. 2011;12: 32–42. doi:10.1038/nrg2899
7. Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coil. Curr Opin Biotechnol. 1995;6: 494–500. doi:10.1016/0958-1669(95)80082-4
8. Lithwick G, Margalit H. Hierarchy of sequence-dependent features associated with prokaryotic translation. Genome Res. 2003;13: 2665–2673. doi:10.1101/gr.1485203
9. Tate J, Ward G. Interferences in immunoassay. Clin Biochem Rev. 2004;25: 105–120.
10. Holliger P, Hudson PJ. Engineered antibody fragments and the rise of single domains. Nat Biotechnol. 2005;23: 1126–1136. doi:10.1038/nbt1142
11. Hust M, Meyer T, Voedisch B, Rülker T, Thie H, El-Ghezal A, et al. A human scFv antibody generation pipeline for proteome research. J Biotechnol. 2011;152: 159–170. doi:10.1016/j.jbiotec.2010.09.945
12. Omidfar K, Daneshpour M. Advances in phage display technology for drug discovery. Expert Opin Drug Discov. 2015;10: 651–669. doi:10.1517/17460441.2015.1037738
13. Ukkonen K, Veijola J, Vasala A, Neubauer P. Effect of culture medium, host strain and oxygen transfer on recombinant Fab antibody fragment yield and leakage to medium in shaken E. coli cultures. Microb Cell Fact. 2013;12: doi:10.1186/1475-2859-12-73. doi:10.1186/1475-2859-12-73
14. Zhong N, Loppnau P, Seitova A, Ravichandran M, Fenner M, Jain H, et al. Optimizing production of antigens and Fabs in the context of generating recombinant antibodies to human proteins. Muyldermans S, editor. PLoS One. 2015;10: e0139695. doi:10.1371/journal.pone.0139695. doi:10.1371/journal.pone.0139695
15. Levy R, Weiss R, Chen G, Iverson BL, Georgiou G. Production of correctly folded Fab antibody fragment in the cytoplasm of Escherichia coli trxB gor mutants via the coexpression of molecular chaperones. Protein Expr Purif. 2001;23: 338–347. doi:10.1006/prep.2001.1520
16. Lin B, Renshaw MW, Autote K, Smith LM, Calveley P, Bowdish KS, et al. A step-wise approach significantly enhances protein yield of a rationally-designed agonist antibody fragment in E. coli. Protein Expr Purif. 2008;59: 55–63. doi:10.1016/j.pep.2008.01.002

17. Vazquez-Lombardi R, Phan TG, Zimmermann C, Lowe D, Jermutus L, Christ D. Challenges and opportunities for non-antibody scaffold drugs. Drug Discov Today. 2015;20: 1271–1283. doi:10.1016/j.drudis.2015.09.004
18. Simeon R, Chen Z. In vitro-engineered non-antibody protein therapeutics. Protein Cell. 2018;9: 3–14. doi:10.1007/s13238-017-0386-6
19. Steiner D, Forrer P, Stumpp MT, Plückthun A. Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. Nat Biotechnol. 2006;24: 823–831. doi:10.1038/nbt1218
20. Aldeghaither D, Smaglo BG, Weiner LM. Beyond peptides and mAbs - Current status and future perspectives for biotherapeutics with novel constructs. J Clin Pharmacol. 2015;55: S4–S20. doi:10.1002/jcph.407
21. Chiu ML, Gilliland GL. Engineering antibody therapeutics. Curr Opin Struct Biol. 2016;38: 163–173. doi:10.1016/j.sbi.2016.07.012
22. Bates A, Power CA. David vs. Goliath: The Structure, Function, and Clinical Prospects of Antibody Fragments. Antibodies. 2019;8: 28. doi:10.3390/antib8020028
23. Bird RE, Hardman KD, Jacobson JW, Johnson SYD, Kaufman BM, Lee S, et al. Single-Chain Antigen-Binding Proteins. Science (80-). 1988;242: 423–426.
24. Helma J, Cardoso MC, Muyldermans S, Leonhardt H. Nanobodies and recombinant binders in cell biology. J Cell Biol. 2015;209: 633–644. doi:10.1083/jcb.201409074
25. Huston JS, Levinson D, Mudgett-Hunter M, Tai MS, Novotny J, Margolies MN, et al. Protein engineering of antibody binding sites: Recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in Escherichia coli. Proc Natl Acad Sci U S A. 1988;85: 5879–5883. doi:10.1073/pnas.85.16.5879
26. Gruber M, Schodin BA, Wilson ER, Kranz DM. Efficient tumor cell lysis mediated by a bispecific single chain antibody expressed in Escherichia coli. J Immunol. 1994;152: 5368–74. Available: http://www.ncbi.nlm.nih.gov/pubmed/8189055
27. Mallender WD, Voss EW. Construction, expression, and activity of a bivalent bispecific single-chain antibody. J Biol Chem. 1994;269: 199–206.
28. Brinkmann U, Kontermann RE. The making of bispecific antibodies. MAbs. 2017;9: 182–212. doi:10.1080/19420862.2016.1268307
29. Antibody Society. Antibody therapeutics approved or in regulatory review in the EU or US. 2019 [cited 22 Nov 2019]. Available: https://www.antibodysociety.org/resources/approved-antibodies/
30. Holliger P, Prospero T, Winter G. “Diabodies”: Small bivalent and bispecific antibody fragments. Proc Natl Acad Sci U S A. 1993;90: 6444–6448. doi:10.1073/pnas.90.14.6444
31. Johnson S, Burke S, Huang L, Gorlatov S, Li H, Wang W, et al. Effector cell recruitment with novel Fv-based dual-affinity re-targeting protein leads to potent tumor cytolysis and in vivo B-cell depletion. J Mol Biol. 2010;399: 436–449. doi:10.1016/j.jmb.2010.04.001
32. Affimed. Affimed pipeline. 2019 [cited 22 Nov 2019]. Available: https://www.affimed.com/rock-platform/pipeline/
33. Better M, Chang CP, Robinson RR, Horwitz AH. Escherichia coli secretion of an active chimeric antibody fragment. Science (80-). 1988;240: 1041–1043. doi:10.1126/science.3285471
34. Hust M, Dübel S. Mating antibody phage display with proteomics. Trends Biotechnol. 2004;22: 8–14. doi:10.1016/j.tibtech.2003.10.011
35. Ward ES, Gussow D, Griffiths AD, Jones PT, Winter G. Binding activities of a repertoire of single immunoglobulin variable domains secreted from Escherichia coli. Nature. 1989;341: 544–546. doi:10.1101/sqb.1989.054.01.033
36. Davies J, Riechmann L. “Camelising” human antibody fragments: NMR studies on VH domains. FEBS Lett. 1994;339: 285–290.

37. Ewert S, Huber T, Honegger A, Plückthun A. Biophysical properties of human antibody variable domains. J Mol Biol. 2003;325: 531–553. doi:10.1016/S0022-2836(02)01237-8
38. O'Connor-Semmes RL, Lin J, Hodge RJ, Andrews S, Chism J, Choudhury A, et al. GSK2374697, a Novel Albumin-Binding Domain Antibody (AlbudAb), Extends Systemic Exposure of Exendin-4: First Study in Humans - PK/PD and Safety. Clin Pharmacol Ther. 2014;96: 704–712. doi:10.1038/clpt.2014.187
39. Scott MJ, Lee JA, Wake MS, Batt K V., Wattam TA, Hiles ID, et al. 'In-Format' screening of a novel bispecific antibody format reveals significant potency improvements relative to unformatted molecules. MAbs. 2017;9: 85–93. doi:10.1080/19420862.2016.1249078
40. Maass DR, Sepulveda J, Pernthaner A, Shoemaker CB. Alpaca (Lama pacos) as a convenient source of recombinant camelid heavy chain antibodies (VHHs). J Immunol Methods. 2007;324: 13–25. doi:10.1016/j.jim.2007.04.008
41. Reader RH, Workman RG, Maddison BC, Gough KC. Advances in the Production and Batch Reformatting of Phage Antibody Libraries. Mol Biotechnol. 2019;61: 801–815. doi:10.1007/s12033-019-00207-0
42. Cortez-Retamozo V, Backmann N, Senter PD, Wernery U, De Baetselier P, Muyldermans S, et al. Efficient Cancer Therapy with a Nanobody-Based Conjugate. Cancer Res. 2004;64: 2853–2857. doi:10.1158/0008-5472.CAN-03-3935
43. Van Der Linden RHJ, Frenken LGJ, De Geus B, Harmsen MM, Ruuls RC, Stok W, et al. Comparison of physical chemical properties of llama V(HH) antibody fragments and mouse monoclonal antibodies. Biochim Biophys Acta - Protein Struct Mol Enzymol. 1999;1431: 37–46. doi:10.1016/S0167-4838(99)00030-8
44. Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hammers C, Songa EB, et al. Naturally occurring antibodies devoid of light chains. Nature. 1993;363: 446–448. doi:10.1038/363446a0
45. Sabir JSM, Atef A, El-Domyati FM, Edris S, Hajrah N, Alzohairy AM, et al. Construction of naïve camelids VHH repertoire in phage display-based library. Comptes Rendus - Biol. 2014;337: 244–249. doi:10.1016/j.crvi.2014.02.004
46. Govaert J, Pellis M, Deschacht N, Vincke C, Conrath K, Muyldermans S, et al. Dual beneficial effect of interloop disulfide bond for single domain antibody fragments. J Biol Chem. 2012;287: 1970–1979. doi:10.1074/jbc.M111.242818
47. Cortez-Retamozo V, Lauwereys M, Hassanzadeh Gh. G, Gobert M, Conrath K, Muyldermans S, et al. Efficient tumor targeting by single-domain antibody fragments of camels. Int J Cancer. 2002;98: 456–462. doi:10.1002/ijc.10212
48. Harmsen MM, De Haard HJ. Properties, production, and applications of camelid single-domain antibody fragments. Appl Microbiol Biotechnol. 2007;77: 13–22. doi:10.1007/s00253-007-1142-2
49. Harmsen MM, Van Solt CB, Fijten HPD, Van Setten MC. Prolonged in vivo residence times of llama single-domain antibody fragments in pigs by binding to porcine immunoglobulins. Vaccine. 2005;23: 4926–4934. doi:10.1016/j.vaccine.2005.05.017
50. Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, Flajnik MF. A new antigen receptor gene family that undergoes rearrangement and ... Nature. 1995;374: 168–173. doi:10.1038/nature01633.1.
51. Zielonka S, Empting M, Grzeschik J, Könning D, Barelle CJ, Kolmar H. Structural insights and biomedical potential of IgNAR scaffolds from sharks. MAbs. 2015;7: 15–25. doi:10.4161/19420862.2015.989032
52. Stanfield RL, Dooley H, Verdino P, Flajnik MF, Wilson IA. Maturation of Shark Single-domain (IgNAR) Antibodies: Evidence for Induced-fit Binding. J Mol Biol. 2007;367: 358–372. doi:10.1016/j.jmb.2006.12.045

53. Feng M, Bian H, Wu X, Fu T, Fu Y, Hong J, et al. Construction and next-generation sequencing analysis of a large phage-displayed VNAR single-domain antibody library from six naïve nurse sharks. Antib Ther. 2019;2: 1–11. doi:10.1093/abt/tby011
54. Dooley H, Flajnik MF, Porter AJ. Selection and characterization of naturally occurring single-domain (IgNAR) antibody fragments from immunized sharks by phage display. Mol Immunol. 2003;40: 25–33. doi:10.1016/S0161-5890(03)00084-1
55. Jenkins TP, Fryer T, Dehli RI, Jürgensen JA, Fuglsang-Madsen A, Føns S, et al. Toxin neutralization using alternative binding proteins. Toxins (Basel). 2019;11: 1–28. doi:10.3390/toxins11010053
56. Koide A, Bailey CW, Huang X, Koide S. The fibronectin type III domain as a scaffold for novel binding proteins. J Mol Biol. 1998;284: 1141–1151. doi:10.1006/jmbi.1998.2238
57. Xu L, Aha P, Gu K, Kuimelis RG, Kurz M, Lam T, et al. Directed evolution of high-affinity antibody mimics using mRNA display. Chem Biol. 2002;9: 933–942. doi:10.1016/S1074-5521(02)00187-4
58. Dineen SP, Sullivan LA, Beck AW, Miller AF, Carbon JG, Mamluk R, et al. The Adnectin CT-322 is a novel VEGF receptor 2 inhibitor that decreases tumor burden in an orthotopic mouse model of pancreatic cancer. BMC Cancer. 2008;8: 1–10. doi:10.1186/1471-2407-8-352
59. Beste G, Schmidt FS, Stibora T, Skerra A. Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. Proc Natl Acad Sci U S A. 1999;96: 1898–1903. doi:10.1073/pnas.96.5.1898
60. Silverman J, Lu Q, Bakker A, To W, Duguay A, Alba BM, et al. Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains. Nat Biotechnol. 2005;23: 1556–1561. doi:10.1038/nbt1166
61. Schlatter D, Brack S, Banner DW, Batey S, Benz J, Bertschinger J, et al. Generation, characterization and structural data of chymase binding proteins based on the human Fyn kinase SH3 domain. MAbs. 2012;4: 497–508. doi:10.4161/mabs.20452
62. Grabulovski D, Kaspar M, Neri D. A novel, non-immunogenic Fyn SH3-derived binding protein with tumor vascular targeting properties. J Biol Chem. 2007;282: 3196–3204. doi:10.1074/jbc.M609211200
63. Bode W, Huber R. Natural protein proteinase inhibitors and their interaction with proteinases. Eur J Biochem. 1992;204: 433–451.
64. Roberts BL, Markland W, Siranosian K, Saxena MJ, Guterman SK, Ladner RC. Protease inhibitor display M13 phage: selection of high-affinity neutrophil elastase inhibitors. Gene. 1992;121: 9–15. doi:10.1016/0378-1119(92)90156-J
65. Roberts BL, Markland W, Ley AC, Kent RB, White DW, Guterman SK, et al. Directed evolution of a protein: Selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. Proc Natl Acad Sci U S A. 1992;89: 2429–2433. doi:10.1073/pnas.89.6.2429
66. Colgrave ML, Craik DJ. Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: The importance of the cyclic cystine knot. Biochemistry. 2004;43: 5965–5975. doi:10.1021/bi049711q
67. Smith HS, Deer TR. Safety and efficacy of intrathecal ziconotide in the management of severe chronic pain. Ther Clin Risk Manag. 2009;5: 521–534. doi:10.2147/tcrm.s4438
68. Nord K, Gunneriusson E, Ringdahl J, Stahl S, Uhlen M, Nygren P. Binding proteins selected from combinatorial libraries of an a-helical bacterial receptor domain. Nat Biotechnol. 1997;15: 772–777.
69. Forrer P, Stumpp MT, Binz HK, Plückthun A. A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. FEBS Lett. 2003;539: 2–6. doi:10.1016/S0014-5793(03)00177-7
70. Binz HK, Stumpp MT, Forrer P, Amstutz P, Plückthun A. Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J Mol Biol. 2003;332: 489–503. doi:10.1016/S0022-2836(03)00896-9

71. Molecular Partners. INTERIM MANAGEMENT STATEMENT – Q3 2019: Encouraging Additional Phase 3 Data Presented for Abicipar, and Start of Phase 1 Trial for MP0310, the First Immuno-oncology DARPin® with a Novel Therapeutic Design. In: Press Release [Internet]. 2019 [cited 2 Dec 2019]. Available: https://www.molecularpartners.com/interim-management-statement-q3-2019-encouraging-additional-phase-3-data-presented-for-abicipar-and-start-of-phase-1-trial-for-mp0310-the-first-immuno-oncology-darpin-with-a-novel-t/
72. Chung BK-S, Lee D-Y. Computational codon optimization of synthetic gene for protein expression. BMC Syst Biol. 2012;6: 134. doi:10.1186/1752-0509-6-134
73. Grantham R, Gautier C, Gouy M, Mercier R, Pavé A. Codon catalog usage and the genome hypothesis. Nucleic Acids Res. 1980;8: 197. doi:10.1093/nar/8.1.197-c
74. Ikemura T. Codon Usage and tRNA Content in Unicellular and Multicellular Organisms'. 1985;2: 13–34.
75. Chaney JL, Clark PL. Roles for Synonymous Codon Usage in Protein Biogenesis. Annu Rev Biophys. 2015;44: 143–66. doi:10.1146/annurev-biophys-060414-034333
76. Roth A, Anisimova M, Cannarozzi GM. Measuring codon usage bias. Codon Evol Mech Model. 2012. doi:10.1093/acprof:osobl/9780199601165.003.0013
77. Akashi H. Synonymous Codon Usage in Drosophila melanogaster: Natural Selection and Translational Accuracy. Genet Soc Am. 1994;136: 927–935.
78. Stoletzki N, Eyre-Walker A. Synonymous codon usage in Escherichia coli: Selection for translational accuracy. Mol Biol Evol. 2007;24: 374–381. doi:10.1093/molbev/msl166
79. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. TL - 12. Nat Rev Genet. 2011;12 VN-r: 32–42. doi:10.1038/nrg2899
80. Novoa EM, Ribas de Pouplana L. Speeding with control: Codon usage, tRNAs, and ribosomes. Trends Genet. 2012;28: 574–581. doi:10.1016/j.tig.2012.07.006
81. Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from the international DNA sequence databases; its status 1999. Nucleic Acids Res. 1999;27: 292. doi:10.1093/nar/27.1.292
82. Hanson G, Coller J. Translation and Protein Quality Control: Codon optimality, bias and usage in translation and mRNA decay. Nat Rev Mol Cell Biol. 2018;19: 20–30. doi:10.1038/nrm.2017.91
83. Sharp PM, Li W. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 1987;15: 1281–1295.
84. Ikemura T. Correlation between the abundance of Escherichia coli transfer RNAs and the occurence of the respective codon choice that is optimal for the E. coli translational system. J Mol Biol. 1981;151: 389–409.
85. dos Reis M, Savva R, Wernisch L. Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res. 2004;32: 5036–5044. doi:10.1093/nar/gkh834
86. Gingold H, Pilpel Y. Determinants of translation efficiency and accuracy. Mol Syst Biol. 2011;7: 481. doi:10.1038/msb.2011.14
87. Spencer PS, Siller E, Anderson JF, Barral JM. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. J Mol Biol. 2012;422: 328–335. doi:10.1016/j.jmb.2012.06.010
88. Clarke IV TF, Clark PL. Rare codons cluster. PLoS One. 2008;3. doi:10.1371/journal.pone.0003412
89. Zhang S, Zubay G, Goldman E. Low-usage codons in Escherichia. Gene. 1991;105: 61–72.
90. European Molecular Biology Laboratory. Protein Expression, E. coli, Optimisation of expression levels. In: Protein Expression and Purification Core Facility [Internet]. 2019 [cited 11 Nov 2019]. Available: https://www.embl.de/pepcore/pepcore_services/protein_expression/ecoli/optimisation_expression_levels/
91. Dana A, Tuller T. Mean of the Typical Decoding Rates: A New Translation Efficiency Index Based on the Analysis of Ribosome Profiling Data. G3-Genes Genomes Genet. 2015;5: 73–80. doi:10.1534/g3.114.015099

92. Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O. Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. Protein Expr Purif. 2008;59: 94–102. doi:10.1016/j.pep.2008.01.008
93. Wang X, Li X, Zhang Z, Shen X, Zhong F. Codon optimization enhances secretory expression of Pseudomonas aeruginosa exotoxin A in E. coli. Protein Expr Purif. 2010;72: 101–106. doi:10.1016/j.pep.2010.02.011
94. Marlatt NM, Spratt DE, Shaw GS. Codon optimization for enhanced Escherichia coli expression of human S100A11 and S100A1 proteins. Protein Expr Purif. 2010;73: 58–64. doi:10.1016/j.pep.2010.03.015
95. Kulmala A, Huovinen T, Lamminmäki U. Effect of DNA sequence of Fab fragment on yield characteristics and cell growth of E. coli. Sci Rep. 2017;7: 10.1038/s41598-017-03957–6. doi:10.1038/s41598-017-03957-6
96. Angov E, Hillier CJ, Kincaid RL, Lyon JA. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. PLoS One. 2008;3: e2189. doi:10.1371/journal.pone.0002189. doi:10.1371/journal.pone.0002189
97. Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, et al. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell. 2010;141: 344–354. doi:10.1016/j.cell.2010.03.031
98. Clarke TF, Clark PL. Increased incidence of rare codon clusters at 5' and 3' gene termini: implications for function. BMC Genomics. 2010;11: 118. doi:10.1186/1471-2164-11-118
99. Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene expression in Escherichia coli. Science (80-). 2009;324: 255–258. doi:10.1126/science.1170160
100. Goodman DB, Church GM, Kosuri S. Causes and effects of N-terminal codon bias in bacterial genes. Science (80-). 2013;342: 475–479. doi:10.1126/science.1241934
101. Bentele K, Saffert P, Rauscher R, Ignatova Z, Blüthgen N. Efficient translation initiation dictates codon usage at gene start. Mol Syst Biol. 2013;9: doi:10.1038/msb.2013.32. doi:10.1038/msb.2013.32
102. Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from the international DNA sequence data bank; its status 1999. Nucleic Acids Res. 1999;27: 292.
103. Osterman IA, Evfratov SA, Sergiev P V., Dontsova OA. Comparison of mRNA features affecting translation initiation and reinitiation. Nucleic Acids Res. 2013;41: 474–486. doi:10.1093/nar/gks989
104. Tuller T, Waldman YY, Kupiec M, Ruppin E. Translation efficiency is determined by both codon bias and folding energy. Proc Natl Acad Sci U S A. 2010;107: 3645–50. doi:10.1073/pnas.0909910107
105. Thanaraj T a, Argos P. Ribosome-mediated translational pause and protein domain organization. Protein Sci. 1996;5: 1594–1612. doi:10.1002/pro.5560050814
106. Cortazzo P, Cervenansky C, Marin M, Reiss C, Ehrlich R, Deana a. Silent mutations affect in vivo protein folding in Escherichia coli. Biochem Biophys Res Commun. 2002;293: 537–541. doi:10.1016/S0006-291X(02)00226-7
107. Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat Struct Mol Biol. 2009;16: 274–280. doi:10.1038/nsmb.1554
108. Saunders R, Deane CM. Synonymous codon usage influences the local protein structure observed. Nucleic Acids Res. 2010;38: 6719–6728. doi:10.1093/nar/gkq495
109. Chaney JL, Steele A, Carmichael R, Rodriguez A, Specht AT, Ngo K, et al. Widespread position-specific conservation of synonymous rare codons within coding sequences. PLoS Comput Biol. 2017;13: 1–19. doi:10.1371/journal.pcbi.1005531

110. O'Brien EP, Vendruscolo M, Dobson CM. Kinetic modelling indicates that fast-translating codons can coordinate cotranslational protein folding by avoiding misfolded intermediates. Nat Commun. 2014;5: 2988. doi:10.1038/ncomms3988
111. Purvis IJ, Bettany AJE, Santiago TC, Coggins JR, Duncan K, Eason R, et al. The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. J Mol Biol. 1987;193: 413–417. doi:10.1016/0022-2836(87)90230-0
112. Sander IM, Chaney JL, Clark PL. Expanding Anfinsen ' s Principle : Contributions of Synonymous Codon Selection to Rational Protein Design Expanding Anfinsen ' s Principle : Contributions of Synonymous Codon Selection to Rational Protein Design. J Am Chem Soc. 2014;136: 858–861. doi:10.1021/ja411302m
113. Buhr F, Jha S, Thommen M, Mittelstaet J, Kutz F, Schwalbe H, et al. Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. Mol Cell. 2016;61: 341–351. doi:10.1016/j.molcel.2016.01.008
114. López D, Pazos F. Protein functional features are reflected in the patterns of mRNA translation speed. BMC Genomics. 2015;16: 513. doi:10.1186/s12864-015-1734-7
115. Faure G, Ogurtsov AY, Shabalina SA, Koonin E V. Role of mRNA structure in the control of protein folding. 2016;44: 10898–10911. doi:10.1093/nar/gkw671
116. Thanaraj TA, Argos P. Protein secondary structural types are differentially coded on messenger RNA. Protein Sci. 1996;5: 1973–1983. doi:10.1002/pro.5560051003
117. Orešič M, Shalloway D. Specific correlations between relative synonymous codon usage and protein secondary structure. J Mol Biol. 1998;281: 31–48. doi:10.1006/jmbi.1998.1921
118. Gu W, Zhou T, Ma J, Sun X, Lu Z. Propensities of Synonymous Codons. 2003;2: 150–157.
119. Drummond DA, Wilke CO. Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. Cell. 2008;134: 341–352. doi:10.1016/j.cell.2008.05.042
120. Zhou T, Weems M, Wilke CO. Translationally Optimal Codons Associate with Structurally Sensitive Sites in Proteins. Mol Biol Evol. 2009;26: 1571–1580. doi:10.1093/molbev/msp070
121. Ogle JM, Ramakrishnan V. Structural Insights Into Translational Fidelity. Annu Rev Biochem. 2005;74: 129–177. doi:10.1146/annurev.biochem.74.061903.155440
122. Warnecke T, Hurst LD. GroEL dependency affects codon usage-support for a critical role of misfolding in gene evolution. Mol Syst Biol. 2010;6: 1–11. doi:10.1038/msb.2009.94
123. Lee Y, Zhou T, Tartaglia GG, Vendruscolo M, Wilke CO. Translationally optimal codons associate with aggregation-prone sites in proteins. Proteomics. 2010;10: 4163–4171. doi:10.1002/pmic.201000229
124. Kramer EB, Farabaugh PJ. The frequency of translational misreading errors in E. coli is largely determined by tRNA competition. Rna. 2007;13: 87–96. doi:10.1261/rna.294907
125. Grosjean HJ, M D. On the physical basis for ambiguity. Science (80-). 1978;75: 610–614.
126. Ogle JM, Brodersen DE, Clemons W.M. J, Tarry MJ, Carter AP, Ramakrishnan V. Recognition of cognate transfer RNA by the 30S ribosomal subunit. Science (80-). 2001;292: 897–902. doi:10.1126/science.1060612
127. Shah P, Gilchrist MA. Effect of Correlated tRNA abundances on translation errors and evolution of codon usage bias. PLoS Genet. 2010;6. doi:10.1371/journal.pgen.1001128
128. Siller E, DeZwaan DC, Anderson JF, Freeman BC, Barral JM. Slowing Bacterial Translation Speed Enhances Eukaryotic Protein Folding Efficiency. J Mol Biol. 2010;396: 1310–1318. doi:10.1016/j.jmb.2009.12.042
129. Gutman GA, Hatfield GW. Nonrandom utilization of codon pairs in Escherichia coli. Proc Natl Acad Sci. 1989;86: 3699–3703. doi:10.1073/pnas.86.10.3699
130. Irwin B, Heck JD, Hatfield GW. Codon Pair Utilization Biases Influence Translational Elongation Step Times. Biochemistry. 1995;270: 22801–22806.
131. Boycheva S, Chkodrov G, Ivanov I. Codon pairs in the genome of Escherichia coli. Bioinformatics. 2003;19: 987–998. doi:10.1093/bioinformatics/btg082

132. Hu X, Shi Q, Yang T, Jackowski G. Specific replacement of consecutive AGG codons results in high-level expression of human cardiac troponin T in Escherichia coli. Protein Expr Purif. 1996;7: 289–93. doi:10.1006/prep.1996.0041
133. Tats A, Tenson T, Remm M. Preferred and avoided codon pairs in three domains of life. BMC Genomics. 2008;9: 463. doi:10.1186/1471-2164-9-463
134. Buchan JR, Aucott LS, Stansfield I. tRNA properties help shape codon pair preferences in open reading frames. Nucleic Acids Res. 2006;34: 1015–1027. doi:10.1093/nar/gkj488
135. Shao ZQ, Zhang YM, Feng XY, Wang B, Chen JQ. Synonymous codon ordering: A subtle but prevalent strategy of bacteria to improve translational efficiency. PLoS One. 2012;7. doi:10.1371/journal.pone.0033547
136. Boel G, Letso R, Neely H, Price WN, Wong K-H, Su M, et al. Codon influence on protein expression in E. coli correlates with mRNA levels. Nature. 2016;529: 358–363. doi:10.1038/nature16509
137. Komar AA. The Yin and Yang of codon usage. Hum Mol Genet. 2016;25: R77–R85. doi:10.1093/hmg/ddw207
138. Edri S, Tuller T. Quantifying the effect of ribosomal density on mRNA stability. PLoS One. 2014;9. doi:10.1371/journal.pone.0102308
139. Frumkin I, Schirman D, Rotman A, Li F, Zahavi L, Mordret E, et al. Gene Architectures that Minimize Cost of Gene Expression. Mol Cell. 2017;65: 142–153. doi:10.1016/j.molcel.2016.11.007
140. Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc Natl Acad Sci U S A. 2002;99: 9697–9702. doi:10.1073/pnas.112318199
141. Power PM, Jones RA, Beacham IR, Bucholtz C, Jennings MP. Whole genome analysis reveals a high incidence of non-optimal codons in secretory signal sequences of Escherichia coli. Biochem Biophys Res Commun. 2004;322: 1038–1044. doi:10.1016/j.bbrc.2004.08.022
142. Zalucki YM, Jennings MP. Experimental confirmation of a key role for non-optimal codons in protein export. Biochem Biophys Res Commun. 2007;355: 143–148. doi:10.1016/j.bbrc.2007.01.126
143. Zalucki YM, Gittins KL, Jennings MP. Secretory signal sequence non-optimal codons are required for expression and export of beta-lactamase. Biochem Biophys Res Commun. 2008;366: 135–141. doi:10.1016/j.bbrc.2007.11.093
144. Zalucki YM, Jones CE, Ng PSK, Schulz BL, Jennings MP. Signal sequence non-optimal codons are required for the correct folding of mature maltose binding protein. Biochim Biophys Acta - Biomembr. 2010;1798: 1244–1249. doi:10.1016/j.bbamem.2010.03.010
145. Kulmala A, Huovinen T, Lamminmäki U. Improvement of Fab expression by screening combinatorial synonymous signal sequence libraries. Microb Cell Fact. 2019;18: 1–17. doi:10.1186/s12934-019-1210-1
146. Zalucki YM, Beacham IR, Jennings MP. Biased codon usage in signal peptides: a role in protein export. Trends Microbiol. 2009;17: 146–150. doi:10.1016/j.tim.2009.01.005
147. Fluman N, Navon S, Bibi E, Pilpel Y. mRNA-programmed translation pauses in the targeting of E. coli membrane proteins. Elife. 2014;3: e03440. doi:10.7554/eLife.03440
148. Li G-W, Oh E, Weissman JS. The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature. 2012;484: 538–541. doi:10.1038/nature10965
149. Dittmar KA, Sørensen MA, Elf J, Ehrenberg M, Pan T. Selective charging of tRNA isoacceptors induced by amino-acid starvation. EMBO Rep. 2005;6: 151–157. doi:10.1038/sj.embor.7400341
150. Elf J, Nilsson D, Tenson T, Ehrenberg M. Selective charging of tRNA isoacceptors explains patterns of codon usage. Science (80-). 2003;300: 1718–1722. doi:10.1126/science.1083811
151. Welch M, Govindarajan S, Ness JE, Villalobos A, Gurney A, Minshull J, et al. Design parameters to control synthetic gene expression in Escherichia coli. PLoS One. 2009;4: e7002. doi:10.1371/journal.pone.0007002. doi:10.1371/journal.pone.0007002

152. Elf J, Ehrenberg M. What makes ribosome-mediated transcriptional attenuation sensitive to amino acid limitation? PLoS Comput Biol. 2005;1: 0014–0023. doi:10.1371/journal.pcbi.0010002
153. Wessler SR, Calvo JM. Control of leu operon expression in Escherichia coli by a transcription attenuation mechanism. J Mol Biol. 1981;149: 579–597. doi:10.1016/0022-2836(81)90348-X
154. Huang CJ, Lin H, Yang XM. Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. J Ind Microbiol Biotechnol. 2012;39: 383–399. doi:10.1007/s10295-011-1082-9
155. Skerra A, Plückthun A. Assembly of a functional immunoglobulin Fvfragment in Escherichia coli. Science (80-). 1988;240: 1038–1041. doi:10.1126/science.3285470
156. Simmons LC, Reilly D, Klimowski L, Shantha Raju T, Meng G, Sims P, et al. Expression of full-length immunoglobulins in Escherichia coli: Rapid and efficient production of aglycosylated antibodies. J Immunol Methods. 2002;263: 133–147. doi:10.1016/S0022-1759(02)00036-4
157. Frenzel A, Hust M, Schirrmann T. Expression of recombinant antibodies. Front Immunol. 2013;4: doi: 10.3389/fimmu.2013.0021. doi:10.3389/fimmu.2013.00217
158. Lee YJ, Jeong KJ. Challenges to production of antibodies in bacteria and yeast. J Biosci Bioeng. 2015;120: 483–490. doi:10.1016/j.jbiosc.2015.03.009
159. de Marco A. Recombinant antibody production evolves into multiple options aimed at yielding reagents suitable for application-specific needs. Microb Cell Fact. 2015;14: 1–17. doi:10.1186/s12934-015-0320-7
160. Spadiut O, Capone S, Krainer F, Glieder A, Herwig C. Microbials for the production of monoclonal antibodies and antibody fragments. Trends Biotechnol. 2014;32: 54–60. doi:10.1016/j.tibtech.2013.10.002
161. Ayyar BV, Arora S, Ravi SS. Optimizing antibody expression: The nuts and bolts. Methods. 2017;116: 51–62. doi:10.1016/j.ymeth.2017.01.009
162. Smith GP. Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. Science (80-). 1985;228: 1315–1317. doi:10.1126/science.4001944
163. McCafferty J, Griffiths A, Winter G, DJ C. Phage antibodies filamentous phage dispalying antibody variable domains. Nature. 1990;348: 552. Available: https://www.nature.com/articles/348552a0
164. Barbas CF, Kang AS, Lerner RA, Benkovic SJ. Assembly of combinatorial antibody libraries on phage surfaces: The gene III site. Proc Natl Acad Sci U S A. 1991;88: 7978–7982. doi:10.1073/pnas.88.18.7978
165. Breitling F, Dübel S, Seehaus T, Klewinghaus I, Little M. A surface expression vector for antibody screening. Gene. 1991;104: 147–53. doi:10.1016/0378-1119(91)90244-6
166. Ledsgaard L, Kilstrup M, Karatt-Vellatt A, McCafferty J, Laustsen AH. Basics of antibody phage display technology. Toxins (Basel). 2018;10. doi:10.3390/toxins10060236
167. Huovinen T, Syrjänpää M, Sanmark H, Seppä T, Akter S, Khan LMF, et al. The selection performance of an antibody library displayed on filamentous phage coat proteins p9, p3 and truncated p3. BMC Res Notes. 2014;7: 1–16. doi:10.1186/1756-0500-7-661
168. Zhang Z, Li ZH, Wang F, Fang M, Yin CC, Zhou ZY, et al. Overexpression of dsbc and dsbg markedly improves soluble and functional expression of single-chain fv antibodies in escherichia coli. Protein Expr Purif. 2002;26: 218–228. doi:10.1016/S1046-5928(02)00502-8
169. Makino T, Skretas G, Georgiou G. Strain engineering for improved expression of recombinant proteins in bacteria. Microb Cell Fact. 2011;10: 32. doi:10.1186/1475-2859-10-32
170. de Marco A. Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. Microb Cell Fact. 2009;8. doi:10.1186/1475-2859-8-26
171. Saul FA, Arié JP, Vulliez-le Normand B, Kahn R, Betton JM, Bentley GA. Structural and Functional Studies of FkpA from Escherichia coli, a cis/trans Peptidyl-prolyl Isomerase with Chaperone Activity. J Mol Biol. 2004;335: 595–608. doi:10.1016/j.jmb.2003.10.056

172. Walton TA, Sandoval CM, Fowler CA, Pardi A, Sousa MC. The cavity-chaperone Skp protects its substrate from aggregation but allows independent folding of substrate domains. Proc Natl Acad Sci U S A. 2009;106: 1772–1777. doi:10.1073/pnas.0809275106
173. Ow DS, Lim DY, Morin Nissom P, Camattari A, Wong V V. Co-expression of Skp and FkpA chaperones improves cell viability and alters the global expression of stress response genes during scFvD1.3 production. Microb Cell Fact. 2010;9: 1–14. doi:10.1186/1475-2859-9-22
174. Sonoda H, Kumada Y, Katsuda T, Yamaji H. Effects of cytoplasmic and periplasmic chaperones on secretory production of single-chain Fv antibody in Escherichia coli. J Biosci Bioeng. 2011;111: 465–470. doi:10.1016/j.jbiosc.2010.12.015
175. Wang R, Xiang S, Feng Y, Srinivas S, Zhang Y, Lin M, et al. Engineering production of functional scfv antibody in E. coli by co-expressing the molecule chaperone skp. Front Cell Infect Microbiol. 2013;4: 1–9. doi:10.3389/fcimb.2013.00072
176. Mavrangelos C, Thiel M, Adamson PJ, Millard DJ, Nobbs S, Zola H, et al. Increased yield and activity of soluble single-chain antibody fragments by combining high-level expression and the Skp periplasmic chaperonin. Protein Expr Purif. 2001;23: 289–295. doi:10.1006/prep.2001.1506
177. Bothmann H, Plückthun A. The periplasmic Escherichia coli peptidylprolyl cis,trans-isomerase FkpA: I. Increased functional expression of antibody fragments with and without cis-prolines. J Biol Chem. 2000;275: 17100–17105. doi:10.1074/jbc.M910233199
178. Bothmann H, Plückthun A. Selection for a periplasmic factor improving phage display and functional periplasmic expression. Nat Biotechnol. 1998;16: 376–380. doi:10.1038/nbt0498-376
179. Levy R, Ahluwalia K, Bohmann DJ, Giang HM, Schwimmer LJ, Issafras H, et al. Enhancement of antibody fragment secretion into the Escherichia coli periplasm by co-expression with the peptidyl prolyl isomerase, FkpA, in the cytoplasm. J Immunol Methods. 2013;394: 10–21. doi:10.1016/j.jim.2013.04.010
180. Chauhan S, Hou CY, Jung ST, Kang TJ. Detection and purification of backbone-cyclized proteins using a bacterially expressed anti-myc-tag single chain antibody. Anal Biochem. 2017;532: 38–44. doi:10.1016/j.ab.2017.06.003
181. Hayhurst A, Harris WJ. Escherichia coli Skp chaperone coexpression improves solubility and phage display of single-chain antibody fragments. Protein Expr Purif. 1999;15: 336–343. doi:10.1006/prep.1999.1035
182. Hoffmann JH, Linke K, Graf PCF, Lilie H, Jakob U. Identification of a redox-regulated chaperone network. EMBO J. 2004;23: 160–168. doi:10.1038/sj.emboj.7600016
183. Sun XW, Wang XH, Yao YB. Co-expression of Dsb proteins enables soluble expression of a single-chain variable fragment (scFv) against human type 1 insulin-like growth factor receptor (IGF-1R) in E. coli. World J Microbiol Biotechnol. 2014;30: 3221–3227. doi:10.1007/s11274-014-1749-2
184. Hu X, O'Hara L, White S, Magner E, Kane M, Gerard Wall J. Optimisation of production of a domoic acid-binding scFv antibody fragment in Escherichia coli using molecular chaperones and functional immobilisation on a mesoporous silicate support. Protein Expr Purif. 2007;52: 194–201. doi:10.1016/j.pep.2006.08.009
185. Rodriguez C, Nam DH, Kruchowy E, Ge X. Efficient Antibody Assembly in E. coli Periplasm by Disulfide Bond Folding Factor Co-expression and Culture Optimization. Appl Biochem Biotechnol. 2017;183: 520–529. doi:10.1007/s12010-017-2502-8
186. Jurado P, Ritz D, Beckwith J, De Lorenzo V, Fernández LA. Production of functional single-chain Fv antibodies in the cytoplasm of Escherichia coli. J Mol Biol. 2002;320: 1–10. doi:10.1016/S0022-2836(02)00405-9
187. Heo MA, Kim SH, Kim SY, Kim YJ, Chung J, Oh MK, et al. Functional expression of single-chain variable fragment antibody against c-Met in the cytoplasm of Escherichia coli. Protein Expr Purif. 2006;47: 203–209. doi:10.1016/j.pep.2005.12.003

188. Lee YJ, Kim HS, Ryu AJ, Jeong KJ. Enhanced production of full-length immunoglobulin G via the signal recognition particle (SRP)-dependent pathway in Escherichia coli. J Biotechnol. 2013;165: 102–108. doi:10.1016/j.jbiotec.2013.03.007
189. Sala A, Bordes P, Genevaux P. Multitasking SecB chaperones in bacteria. Front Microbiol. 2014;5: 1–12. doi:10.3389/fmicb.2014.00666
190. Genevaux P, Georgopoulos C, Kelley WL. The Hsp70 chaperone machines of Escherichia coli: A paradigm for the repartition of chaperone functions. Mol Microbiol. 2007;66: 840–857. doi:10.1111/j.1365-2958.2007.05961.x
191. Kurt N, Rajagopalan S, Cavagnero S. Effect of Hsp70 chaperone on the folding and misfolding of polypeptides modeling an elongating protein chain. J Mol Biol. 2006;355: 809–820. doi:10.1016/j.jmb.2005.10.029
192. Bechtluft P, Kedrov A, Slotboom DJ, Nouwen N, Tans SJ, Driessen AJM. Tight hydrophobic contacts with the secB chaperone prevent folding of substrate proteins. Biochemistry. 2010;49: 2380–2388. doi:10.1021/bi902051e
193. Sonoda H, Kumada Y, Katsuda T, Yamaji H. Functional expression of single-chain Fv antibody in the cytoplasm of Escherichia coli by thioredoxin fusion and co-expression of molecular chaperones. Protein Expr Purif. 2010;70: 248–253. doi:10.1016/j.pep.2009.11.003
194. Hoffmann A, Bukau B, Kramer G. Structure and function of the molecular chaperone Trigger Factor. Biochim Biophys Acta - Mol Cell Res. 2010;1803: 650–661. doi:10.1016/j.bbamcr.2010.01.017
195. Gaciarz A, Veijola J, Uchida Y, Saaranen MJ, Wang C, Hörkkö S, et al. Systematic screening of soluble expression of antibody fragments in the cytoplasm of E. coli. Microb Cell Fact. 2016;15: 10.1186/s12934-016-0419–5. doi:10.1186/s12934-016-0419-5
196. de Marco A. Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. Microb Cell Fact. 2009;8: 26. doi:10.1186/1475-2859-8-26
197. Jurado P, De Lorenzo V, Fernández LA. Thioredoxin fusions increase folding of single chain Fv antibodies in the cytoplasm of Escherichia coli: Evidence that chaperone activity is the prime effect of thioredoxin. J Mol Biol. 2006;357: 49–61. doi:10.1016/j.jmb.2005.12.058
198. Martineau P, Jones P, Winter G. Expression of an antibody fragment at high levels in the bacterial cytoplasm. J Mol Biol. 1998;280: 117–127. doi:10.1006/jmbi.1998.1840
199. Santala V, Lamminmäki U. Production of a biotinylated single-chain antibody fragment in the cytoplasm of Escherichia coli. J Immunol Methods. 2004;284: 165–175. doi:10.1016/j.jim.2003.10.008
200. Napathorn SC, Kuroki M, Kuroki M. High expression of fusion proteins consisting of a single-chain variable fragment antibody against a tumor-associated antigen and Interleukin-2 in Escherichia coli. Anticancer Res. 2014;34: 3937–3946.
201. Venturi M, Seifert C, Hunte C. High level production of functional antibody Fab fragments in an oxidizing bacterial cytoplasm. J Mol Biol. 2002;315: 1–8. doi:10.1006/jmbi.2001.5221
202. Robinson MP, Ke N, Lobstein J, Peterson C, Szkodny A, Mansell TJ, et al. Efficient expression of full-length antibodies in the cytoplasm of engineered bacteria. Nat Commun. 2015;6. doi:10.1038/ncomms9072
203. Hatahet F, Nguyen VD, Salo KEH, Ruddock LW. Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of E. coli. Microb Cell Fact. 2010;9: 67.
204. Proba K, Wörn A, Honegger A, Plückthun A. Antibody scFv fragments without disulfide bonds made by molecular evolution. J Mol Biol. 1998;275: 245–253. doi:10.1006/jmbi.1997.1457
205. Seo MJ, Jeong KJ, Leysath CE, Ellington AD, Iverson BL, Georgiou G. Engineering antibody fragments to fold in the absence of disulfide bonds. Protein Sci. 2009;18: 259–267. doi:10.1002/pro.31

206. Bach H, Mazor Y, Shaky S, Shoham-Lev A, Berdichevsky Y, Gutnick DL, et al. Escherichia coli maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies. J Mol Biol. 2001;312: 79–93. doi:10.1006/jmbi.2001.4914
207. Tiller T, Schuster I, Deppe D, Siegers K, Strohner R, Herrmann T, et al. A fully synthetic human Fab antibody library based on fixed VH/VL framework pairings with favorable biophysical properties. MAbs. 2013;5: 445–470. doi:10.4161/mabs.24218
208. Knappik A, Ge L, Honegger A, Pack P, Fischer M, Wellnhofer G, et al. Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. J Mol Biol. 2000;296: 57–86. doi:10.1006/jmbi.1999.3444
209. Jäger M, Gehrig P, Plückthun A. The scFv fragment of the antibody hu4D5-8: Evidence for early premature domain interaction in refolding. J Mol Biol. 2001;305: 1111–1129. doi:10.1006/jmbi.2000.4342
210. Jäger M, Plückthun A. Folding and assembly of an antibody Fv fragment, a heterodimer stabilized by antigen. J Mol Biol. 1999;285: 2005–2019. doi:10.1006/jmbi.1998.2425
211. Wirtz P, Steipe B. Intrabody construction and expression III: Engineering hyperstable VH domains. Protein Sci. 2008;8: 2245–2250. doi:10.1110/ps.8.11.2245
212. Schwimmer LJ, Huang B, Giang H, Cotter RL, Chemla-Vogel DS, Dy F V., et al. Discovery of diverse and functional antibodies from large human repertoire antibody libraries. J Immunol Methods. 2013;391: 60–71. doi:10.1016/j.jim.2013.02.010
213. Kügler J, Wilke S, Meier D, Tomszak F, Frenzel A, Schirrmann T, et al. Generation and analysis of the improved human HAL9/10 antibody phage display libraries. BMC Biotechnol. 2015;15: 1–15. doi:10.1186/s12896-015-0125-0
214. Jain T, Sun T, Durand S, Hall A, Houston NR, Nett JH, et al. Biophysical properties of the clinical-stage antibody landscape. Proc Natl Acad Sci U S A. 2017;114: 944–949. doi:10.1073/pnas.1616408114
215. Chiti F, Calamai M, Taddei N, Stefani M, Ramponi G, Dobson CM. Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. Proc Natl Acad Sci U S A. 2002;99: 16419–16426. doi:10.1073/pnas.212527999
216. Jespers L, Schon O, Famm K, Winter G. Aggregation-resistant domain antibodies selected on phage by heat denaturation. Nat Biotechnol. 2004;22: 1161–1165. doi:10.1038/nbt1000
217. Dudgeon K, Rouet R, Kokmeijer I, Schofield P, Stolp J, Langley D, et al. General strategy for the generation of human antibody variable domains with increased aggregation resistance. Proc Natl Acad Sci U S A. 2012;109: 10879–10884. doi:10.1073/pnas.1202866109
218. Ewert S, Honegger A, Plückthun A. Structure-based improvement of the biophysical properties of immunoglobulin VH domains with a generalizable approach. Biochemistry. 2003;42: 1517–1528. doi:10.1021/bi026448p
219. Honegger A, Plückthun A. Yet another numbering scheme for immunoglobulin variable domains: An automatic modeling and analysis tool. J Mol Biol. 2001;309: 657–670. doi:10.1006/jmbi.2001.4662
220. Forsberg G, Forsgren M, Jaki M, Norin M, Sterky C, Enhörning Å, et al. Identification of framework residues in a secreted recombinant antibody fragment that control production level and localization in Escherichia coli. J Biol Chem. 1997;272: 12430–12436. doi:10.1074/jbc.272.19.12430
221. Demarest SJ, Chen G, Kimmel BE, Gustafson D, Wu J, Salbato J, et al. Engineering stability into Escherichia coli secreted Fabs leads to increased functional expression. Protein Eng Des Sel. 2006;19: 325–336. doi:10.1093/protein/gzl016
222. Hu S, Wang M, Cai G, He M. Genetic code-guided protein synthesis and folding in Escherichia coli. J Biol Chem. 2013;288: 30855–30861. doi:10.1074/jbc.M113.467977
223. Sharma SK, Suresh MR, Wuest FR. Improved soluble expression of a single-chain antibody fragment in E. coli for targeting CA125 in epithelial ovarian cancer. Protein Expr Purif. 2014;102: 27–37. doi:https://doi.org/10.1016/j.pep.2014.07.007

224. Jennewein MF, Alter G. The Immunoregulatory Roles of Antibody Glycosylation. Trends Immunol. 2017;38: 358–372. doi:10.1016/j.it.2017.02.004
225. Sazinsky SL, Ott RG, Silver NW, Tidor B, Ravetch J V., Wittrup KD. Aglycosylated immunoglobulin G1 variants productively engage activating Fc receptors. Proc Natl Acad Sci U S A. 2008;105: 20167–20172. doi:10.1073/pnas.0809257105
226. Jung ST, Reddy ST, Kang TH, Borrok MJ, Sandlie I, Tucker PW, et al. Aglycosylated IgG variants expressed in bacteria that selectively bind FcγRI potentiate tumor cell killing by monocyte-dendritic cells. Proc Natl Acad Sci U S A. 2010;107: 604–609. doi:10.1073/pnas.0908590107
227. Jung ST, Kelton W, Kang TH, Ng DTW, Andersen JT, Sandlie I, et al. Effective phagocytosis of low Her2 tumor cell lines with engineered, aglycosylated igg displaying high FcγRIIa affinity and selectivity. ACS Chem Biol. 2013;8: 368–375. doi:10.1021/cb300455f
228. Mergulhão FJM, Summers DK, Monteiro GA. Recombinant protein secretion in Escherichia coli. Biotechnol Adv. 2005;23: 177–202. doi:10.1016/j.biotechadv.2004.11.003
229. Thie H, Schirrmann T, Paschke M, Dübel S, Hust M. SRP and Sec pathway leader peptides for antibody phage display and antibody fragment production in E. coli. N Biotechnol. 2008;25: 49–54. doi:10.1016/j.nbt.2008.01.001
230. Frenzel A, Hust M, Schirrmann T. Expression of recombinant antibodies. Front Immunol. 2013;4: 1–20. doi:10.3389/fimmu.2013.00217
231. Zhou Y, Liu P, Gan Y, Sandoval W, Katakam AK, Reichelt M, et al. Enhancing full-length antibody production by signal peptide engineering. Microb Cell Fact. 2016;15: 1–11. doi:10.1186/s12934-016-0445-3
232. Lee YJ, Jeong KJ. Enhanced production of antibody fragment via SRP pathway engineering in Escherichia coli. Biotechnol Bioprocess Eng. 2013;18: 751–758. doi:10.1007/s12257-013-0111-0
233. Koerber JT, Hornsby MJ, Wells JA. An improved single-chain fab platform for efficient display and recombinant expression. J Mol Biol. 2015;427: 576–586. doi:10.1016/j.jmb.2014.11.017
234. Makino T, Skretas G, Kang T-H, Georgiou G. Comprehensive engineering of Escherichia coli for enhanced expression of IgG antibodies. Metab Eng. 2011;13: 241–251. doi:https://doi.org/10.1016/j.ymben.2010.11.002
235. Humphreys DP, Carrington B, Bowering LC, Ganesh R, Sehdev M, Smith BJ, et al. A plasmid system for optimization of Fab′ production in Escherichia coli: Importance of balance of heavy chain and light chain synthesis. Protein Expr Purif. 2002;26: 309–320. doi:10.1016/S1046-5928(02)00543-0
236. Stemmer WPC, Morris SK, Kautzer CR, Wilson BS. Increased antibody expression from Escherichia coli through wobble-base library mutagenesis by enzymatic inverse PCR. Gene. 1993;123: 1–7. doi:10.1016/0378-1119(93)90531-7
237. Yang J, Moyana T, Mackenzie S, Xia Q, Xiang J. One hundred seventy-fold increase in excretion of an FV fragment-tumor necrosis factor alpha fusion protein (sFV/TNF-α) from Escherichia coli caused by the synergistic effects of glycine and Triton X-100. Appl Environ Microbiol. 1998;64: 2869–2874.
238. Fernandez LA, Sola I, Enjuanes L, De Lorenzo V. Specific secretion of active single-chain Fv antibodies into the supernatants of Escherichia coli cultures by use of the hemolysin system. Appl Environ Microbiol. 2000;66: 5024–5029. doi:10.1128/AEM.66.11.5024-5029.2000
239. Cheng CM, Tzou SC, Zhuang YH, Huang CC, Kao CH, Liao KW, et al. Functional production of a soluble and secreted single-chain antibody by a bacterial secretion system. PLoS One. 2014;9: 1–7. doi:10.1371/journal.pone.0097367
240. Tegel H, Ottosson J, Hober S. Enhancing the protein production levels in Escherichia coli with a strong promoter. FEBS J. 2011;278: 729–739. doi:10.1111/j.1742-4658.2010.07991.x

241. Morra R, Shankar J, Robinson CJ, Halliwell S, Butler L, Upton M, et al. Dual transcriptional-Translational cascade permits cellular level tuneable expression control. Nucleic Acids Res. 2016;44. doi:10.1093/nar/gkv912
242. Horga LG, Halliwell S, Castiñeiras TS, Wyre C, Matos CFRO, Yovcheva DS, et al. Tuning recombinant protein expression to match secretion capacity. Microb Cell Fact. 2018;17: 1–18. doi:10.1186/s12934-018-1047-z
243. McKenna R, Lombana TN, Yamada M, Mukhyala K, Veeravalli K. Engineered sigma factors increase full-length antibody expression in Escherichia coli. Metab Eng. 2019;52: 315–323. doi:10.1016/j.ymben.2018.12.009
244. Szymanski CM, Ruijin Y, Ewing CP, Trust TJ, Guerry P. Evidence for a system of general protein glycosylation in Campylobacter jejuni. Mol Microbiol. 1999;32: 1022–1030. doi:10.1046/j.1365-2958.1999.01415.x
245. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, et al. N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. Science (80-). 2002;298: 1790–1793. doi:10.1126/science.298.5599.1790
246. Kelleher DJ, Gilmore R. An evolving view of the eukaryotic oligosaccharyltransferase. Glycobiology. 2006;16: 47–62. doi:10.1093/glycob/cwj066
247. Abu-Qarn M, Eichler J. Protein N-glycosylation in Archaea: Defining Haloferax volcanii genes involved in S-layer glycoprotein glycosylation. Mol Microbiol. 2006;61: 511–525. doi:10.1111/j.1365-2958.2006.05252.x
248. Li L, Woodward R, Ding Y, Liu X wei, Yi W, Bhatt VS, et al. Overexpression and topology of bacterial oligosaccharyltransferase PglB. Biochem Biophys Res Commun. 2010;394: 1069–1074. doi:10.1016/j.bbrc.2010.03.126
249. Chen MM, Glover KJ, Imperiali B. From peptide to protein: Comparative analysis of the substrate specificity of N-linked glycosylation in C. jejuni. Biochemistry. 2007;46: 5579–5585. doi:10.1021/bi602633n
250. Lizak C, Fan YY, Weber TC, Aebi M. N-linked glycosylation of antibody fragments in Escherichia coli. Bioconjug Chem. 2011;22: 488–496. doi:10.1021/bc100511k
251. Fisher AC, Haitjema CH, Guarino C, Çelik E, Endicott CE, Reading CA, et al. Production of secretory and extracellular N-linked glycoproteins in Escherichia coli. Appl Environ Microbiol. 2011;77: 871–881. doi:10.1128/AEM.01901-10
252. Krebber A, Bornhauser S, Honegger A, Bosshard HR, Plückthun A. Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. J Immunol Methods. 1997; 35–55.
253. Huovinen T, Julin M, Sanmark H, Lamminmäki U. Plasmid enhanced error-prone RCA mutagenesis by concatemer resolution. Plasmid. 2011;66: 47–51. doi:10.1016/j.plasmid.2011.03.004
254. Huovinen T, Syrjänpää M, Sanmark H, Brockmann E, Azhayev A, Wang Q, et al. Two ScFv antibody libraries derived from identical VL – VH framework with different binding site designs display distinct binding profiles. Protein Eng Des Sel. 2013;26: 683–693. doi:10.1093/protein/gzt037
255. Sidhu SS, Lowman HB, Cunningahm BC, Wells JA. Phage display for selection of novel binding peptides. Methods Enzymol. 2000;328: 333–363.
256. Lehmusvuori A, Manninen J, Huovinen T, Soukka T. Homogenous M13 bacteriophage quantification assay using switchable lanthanide fluorescence probes. Biotechniques. 2012;53: 301–303. doi:10.2144/0000113954
257. Kunkel TA. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci U S A. 1985;82: 488–492. doi:10.1073/pnas.82.2.488
258. Huovinen T, Brockmann E, Akter S, Perez-Gamarra S, Yla J, Liu Y, et al. Primer extension mutagenesis powered by selective rolling circle amplification. PLoS One. 2012;7: e31817. doi:10.1371/journal.pone.0031817. doi:10.1371/journal.pone.0031817

259. Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, et al. Software news and updates NUPACK : analysis and design of nucleic acid systems. J Comput Chem. 2010;32: 170–173. doi:10.1002/jcc
260. Jung SK, McDonald K. Visual gene developer: A fully programmable bioinformatics software for synthetic gene optimization. BMC Bioinformatics. 2011;12: 1–13. doi:10.1186/1471-2105-12-340
261. Brockmann EC, Akter S, Savukoski T, Huovinen T, Lehmusvuori A, Leivo J, et al. Synthetic single-framework antibody library integrated with rapid affinity maturation by VL shuffling. Protein Eng Des Sel. 2011;24: 691–700. doi:10.1093/protein/gzr023
262. Cannarrozzi G, Schraudolph NN, Faty M, von Rohr P, Friberg MT, Roth AC, et al. A role for codon order in translation dynamics. Cell. 2010;141: 355–367. doi:10.1016/j.cell.2010.02.036
263. Pechmann S, Frydman J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nat Struct Mol Biol. 2013;20: 237–43. doi:10.1038/nsmb.2466
264. Debarbieux L, Beckwith J. The reductive enzyme thioredoxin 1 acts as an oxidant when it is exported to the Escherichia coli periplasm. Proc Natl Acad Sci U S A. 1998;95: 10751–10756. doi:10.1073/pnas.95.18.10751
265. Jonda S, Huber-Wunderlich M, Glockshuber R, Mössner E. Complementation of DsbA deficiency with secreted thioredoxin variants reveals the crucial role of an efficient dithiol oxidant for catalyzed protein folding in the bacterial periplasm. EMBO J. 1999;18: 3271–3281. doi:10.1093/emboj/18.12.3271
266. Schierle CF, Berkmen M, Huber D, Kumamoto C, Boyd D, Beckwith J. The DsbA Signal Sequence Directs Efficient, Cotranslational Export of Passenger Proteins to the Escherichia coli Periplasm via the Signal Recognition Particle Pathway. J Bacteriol. 2003;185: 5706–5713. doi:10.1128/JB.185.19.5706
267. Grilo AL, Mantalaris A. The Increasingly Human and Profitable Monoclonal Antibody Market. Trends Biotechnol. 2019;37: 9–16. doi:10.1016/j.tibtech.2018.05.014

UNIVERSITY OF TURKU

Painosalama Oy, Turku, Finland 2020

ISBN 978-951-29-8064-2 (PRINT)
ISBN 978-951-29-8065-9 (PDF)
ISSN 0082-7002 (Print)
ISSN 2343-3175 (Online)