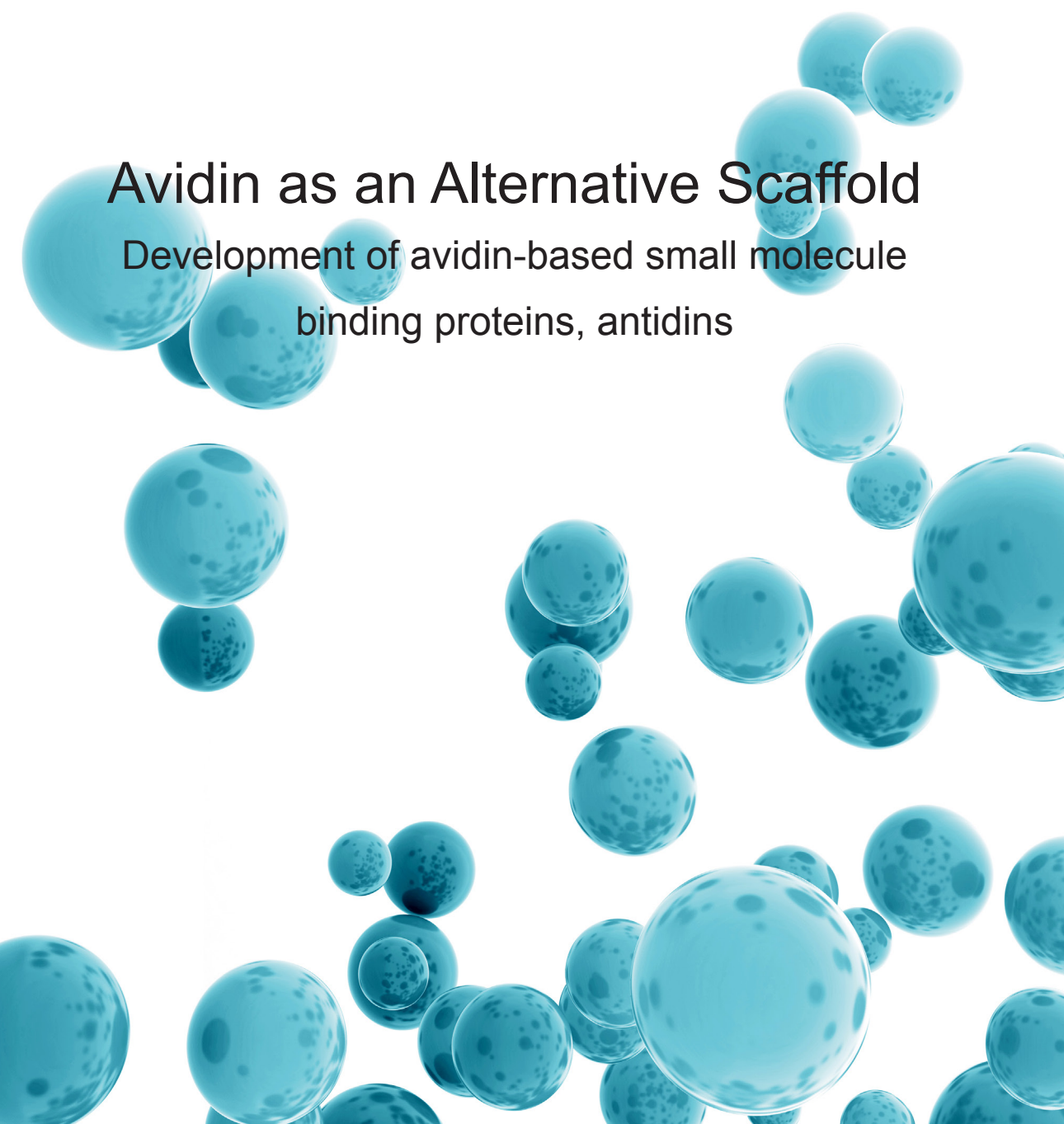


SOILI LEHTONEN

Avidin as an Alternative Scaffold

Development of avidin-based small molecule
binding proteins, antidins





SOILI LEHTONEN

Avidin as an Alternative Scaffold

Development of avidin-based small molecule
binding proteins, antidotes



ACADEMIC DISSERTATION

To be presented, with the permission of
the Faculty Council of the Faculty of Medicine and Life Sciences
of the University of Tampere,
for public discussion in the auditorium F115 of the Arvo building,
Arvo Ylpön katu 34, Tampere,
on 15 September 2017, at 12 o'clock.

UNIVERSITY OF TAMPERE

SOILI LEHTONEN

Avidin as an Alternative Scaffold

Development of avidin-based small molecule
binding proteins, antidotes

Acta Universitatis Tamperensis 2305
Tampere University Press
Tampere 2017

ACADEMIC DISSERTATION

University of Tampere, Faculty of Medicine and Life Sciences

National Doctoral Programme in Informational and Structural Biology (ISB)

Doctoral Programme in Biomedicine and Biotechnology

Finland

Supervised by

Professor Emeritus Markku Kulomaa

University of Tampere

Finland

Associate Professor Vesa Hytönen

University of Tampere

Finland

Reviewed by

Docent Urpo Lamminmäki

Tampere University of Technology

Finland

Professor Stefan Ståhl

KTH Biotechnology Royal Institute of Technology

Sweden

The originality of this thesis has been checked using the Turnitin OriginalityCheck service in accordance with the quality management system of the University of Tampere.

Copyright ©2017 Tampere University Press and the author

Cover design by

Mikko Reinikka

Acta Universitatis Tamperensis 2305

ISBN 978-952-03-0513-0 (print)

ISSN-L 1455-1616

ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1809

ISBN 978-952-03-0514-7 (pdf)

ISSN 1456-954X

<http://tampub.uta.fi>

This PhD thesis is dedicated to our dear children:

Jubo, you are the sunshine of my life. Don't lose the enthusiasm you have for everything!

And our dear unborn baby, take your time to develop despite the defense. We'll meet you soon!

ACKNOWLEDGEMENTS

My deepest gratitude goes to my supervisors Professor Emeritus Markku “Kuku” Kulomaa and Associate Professor Vesa Hytönen. Kuku welcomed me to his group and has provided me many interesting opportunities to broaden my knowledge and grow as a scientist through the years. Thank you, Kuku, for your professional support and for providing a warm, pleasant working atmosphere. I wish to thank my other supervisor Vesa for professional guidance and everlasting optimism. Your enthusiastic attitude towards science and your endless thirst for knowledge has been very inspiring. Additionally, I want to thank you both for the first-class working environment and facilities.

I am thankful for Dr. Tiina Riihimäki for introducing me to the fascinating world of phage display and kindling my interest in protein engineering. Thank you for encouraging and teaching me. I also want to thank Dr. Barbara Taskinen for brainstorming and successful collaboration on the DNA library design and phage display. You two have been excellent coworkers, and without you this work would not have been possible! I would also like to thank Antti Tullila and Dr. Tarja Nevanen from VTT Technical Research Center of Finland for sharing ideas and successful collaboration. I also want to thank Sampo Kukkurainen for great work with the modelling and simulations, as well as for creating the Perl script for analyzing the DNA sequencing results. Further thanks go to Dr. Purvi Jain who later joined the group for half a year and taught me new tricks in phage display.

During these years I have had the pleasure to supervise three Master’s theses and a Bachelor’s thesis, as well as to guide several summer students. Thank you Elina Ojala, Rolle Rahikainen, Niklas Kähkönen, Sandra Posch, and Meri Uusi-Mäkelä, who were eager and talented students and contributed to my different projects. I wish to express my gratitude to Ulla Kiiskinen, Niklas Kähkönen, Outi Väätäinen, Latifeh Azizi, and Laura Kananen for the excellent technical support and endless willingness to help. Special thanks to my dear “lunch buddies” Dr. Tiina Riihimäki, Dr. Juha Määttä, Dr. Tiia Koho, Dr. Jenni Leppiniemi, Rolle Rahikainen, Dr. Jenita Pärssinen, Dr. Sanna Auer, Dr. Minna Hankaniemi, Niila Saarinen, Magdaléna von Essen and Latifeh Azizi for enjoyable conversations (about work and everything else) and good times over these years. It has been a pleasure working with you! I

would like to also thank other MBT and PD group members, the new and former ones: the late Docent Henri Nordlund, Dr. Satu Helppolainen, Dr. Jarkko Valjakka, Dr. Olli Laitinen, Dr. Anssi Mähönen, Anssi Nurminen, Dr. Inger Vikholm-Lundin, Dr. Vasył Mykuliak and various students over the years for creating an excellent working environment.

I am grateful to my thesis committee members Professor Emeritus Pentti Tuohimaa, Professor Kristiina Takkinen, and Dr. h.c. Jouko Haapalahti for interesting discussions and valuable comments during my thesis project. Additionally, I wish to express my sincere gratitude to all my collaborators and co-authors not already mentioned: Dr. Martina Rangl, Dr. Andreas Ebner, Professor Peter Hinterdorfer, Masi Koskinen, Nitin Agrawal, Professor Mark Johnsson, and Dr. Tomi Airenne. I would like to thank my reviewers Professor Stefan Ståhl and Dr. Urpo Lamminmäki for their valuable comments and feedback to improve my thesis, and Dr. Eloise Mikkonen for proofreading the English.

I also would like to thank the Academy of Finland (grants to M.S.K.) and the National Doctoral Programme in Informational and Structural Biology (ISB) for funding, and a personal grant from the Finnish Cultural Foundation for finalizing the thesis. In addition to funding, ISB also enabled me to make connections with other graduate students throughout Finland by organizing excellent annual Spring and Winter meetings. Furthermore, I would like to thank the Scientific Foundation of the City of Tampere for a grant for printing this PhD thesis. I also thank the Pirkanmaa Hospital District for financial support.

The balance in my life has come from my dear friends with whom I have forgotten the stress of work. Thank you all for enjoyable moments during these years: Anu, Jouni, Kati, Vesku, Elli, Jussi, Sami, Kaisa, Hanna, Mikko, Sari, Henna, Laura, Jouni, Jossu, Joonas, Anna, Suvi, and Aino. I am thankful for the numerous good times and experiences we have shared together!

Last but not least, I am grateful to my family, relatives, and family-in-law for their support and care over the years. I wish to communicate a great expression of gratitude to my parents for all their support and encouragement during my studies. I also want to thank my parents-in-law for their hospitality and all their help. Help from both parents has been priceless during the last couple of years. Kiitokset tuesta ja muistamisista rakkaille Mummuillen! Thank you Antti, Johanna, Mikko, and Terhi for your hospitality and enjoyable times.

Finally, my deepest gratitude goes to my dear husband Jukka and our dear son Juho. You bring so much love, joy, and happiness to my life. Thank you, Jukka, for

sharing your life with me, being my best friend, patiently supporting, encouraging,
and loving me. I love you!

Tampere, August 2017

A handwritten signature in cursive script, reading "Suli Lehtinen". The signature is written in dark ink on a white background. The first name "Suli" is written in a simple, rounded cursive, and the last name "Lehtinen" is written in a more fluid, connected cursive style, ending with a long, horizontal flourish.

ABSTRACT

Antibodies, the specific binding proteins produced by our immune system for virtually any foreign target molecule, have been used as valuable laboratory reagents in the life sciences and as therapeutic drugs in medicine for almost a century. However, in recent years these antibody-based affinity reagents have been challenged by novel types of binding proteins developed through directed evolution, offering preferable properties in certain applications. The aim of this thesis was to develop avidin scaffold-based novel binding proteins for small molecule targets, which are challenging to develop high-affinity antibodies for. Avidin is an exceptionally stable protein from chicken egg-white, known for its high affinity towards its native ligand, biotin ($K_d \sim 10^{-15} \text{M}$). This is the strongest non-covalent interaction known to nature between a protein-ligand pair. The structure-function of avidin is well known, and avidin is known to tolerate genetic modifications well, especially when located in the loop regions. Above all, lipocalins belonging to the same “calycin” superfamily as avidin, have already been successfully used as alternative scaffolds.

Avidin libraries were constructed by targeting loop region amino acid residues involved in biotin binding for randomization, and phage display was used for selection. Testosterone was chosen as the first target molecule, resulting in the selection of avidin variants, antidin, with micromolar affinities. With the help of affinity maturation, the remaining biotin-binding affinity was significantly reduced while improving the affinity towards testosterone. Later several avidin libraries were pooled together and used for the selection of binders towards several diagnostically relevant small molecules in parallel. The obtained results showed that an avidin scaffold has potential to be used as an alternative scaffold for several small molecules.

As a result of this PhD project, new methods for library construction were established. The new Gateway-compatible phagemid (pGWphagemid) vector was constructed to enable more efficient library generation utilizing the homologous recombination of the bacteriophage lambda. Furthermore, a short-cut to the traditional DNA-shuffling protocol was introduced preserving the high quality of the DNA library while reducing time required for library construction.

TIIVISTELMÄ

Vasta-aineet ovat immuunipuolustusjärjestelmämme tuottamia erityisiä sitojaproteiineja, jotka kykenevät tunnistamaan lähes minkä tahansa vieraan molekyylin. Niitä onkin hyödynnetty jo melkein vuosisadan ajan sekä arvokkaina laboratorio-reagensseina biotieteissä että lääkkeinä. Viime vuosina vasta-aineisiin perustuvat reagenssit ovat saaneet haastajikseen uudenlaisia sitojaproteiineja, jotka tarjoavat tiettyihin sovelluksiin vasta-aineita parempia ominaisuuksia. Tämän väitöskirjan tavoite oli kehittää avidiinin rakenteeseen perustuvia uusia sitojaproteiineja pienmolekyyleille. Avidiini on poikkeuksellisen stabiili proteiini, joka on peräisin kananmunan valkuaisesta ja tunnetaan sen erittäin korkeasta affiniteetista luontaista ligandiaan, biotiinia, kohtaan ($K_d \sim 10^{-15} \text{M}$). Tämä on voimakkain tunnettu eikovalenttinen vuorovaikutus proteiinin ja sen ligandin välillä. Avidiinin rakenteen ja toiminnan välinen yhteys on hyvin tunnettu, ja proteiinin tiedetään kestävän hyvin geneettistä muokkausta, varsinkin silmukkarakenteidensa osalta. Lisäksi samaan calyxiini-rakennepereheeseen avidiinin kanssa kuuluvien lipokaliinien sitomisominaisuuksia on onnistuneesti jo muokattu.

Avidiinikirjastot koottiin kohdennetun satunnaismutageneesin avulla muokkaamalla avidiinin silmukkarakenteissa sijaitsevia biotiinin sitomiseen osallistuvia aminohappotähteitä. Sitomisominaisuuksiltaan halutunkaltaiset avidiinivariantit valikoitiin sitten faaginäyttömenetelmän avulla. Ensimmäisenä kohdemolekyylinä toimineelle testosteronille saatiin valikoitua mikromolaarisen affiniteetin omaava sitoja. Affiniteettimaturaation avulla sen ristireaktiota biotiinille saatiin vähennettyä. Myöhemmin kolme erilaista avidiinikirjastoa yhdistettiin ja niistä seulottiin onnistuneesti erilaisia diagnostisesti merkittäviä pienmolekyylejä sitovia avidiinivariantteja, ns. antidiineja. Avidiini-rakenteen osoitettiin näin olevan muokattavissa sitomaan useampiakin erilaisia kohdemolekyylejä, joten sitä voitaisiin käyttää uusien sitojaproteiinien kehittämisessä.

Tässä väitöskirjaprojektissa kehitettiin myös kirjaston kokoamisessa käytettäviä menetelmiä. Uusi Gateway-kloonaukseen yhteensopiva phagemid-vektori (pGWphagemid) koottiin hyödyntämään bakteriofagi lambdan käyttämää homologista rekombinaatiota. Samalla suoraviivaistettiin DNA:n sekoitusmenetelmällä koottavien kimeeristen geenikirjastojen kokoamisprotokollaa, jolloin DNA-

kirjaston korkea laatu ja diversiteetti saatiin paremmin säilytettyä samalla lyhentäen kirjaston kokoamisessa tarvittavaa aikaa.

TABLE OF CONTENTS

Acknowledgements	5
Abstract	9
Tiivistelmä.....	11
ABBREVIATIONS	16
ORIGINAL PUBLICATIONS.....	19
1 INTRODUCTION	21
2 REVIEW OF THE LITERATURE	24
2.1 Avidin and other biotin-binding proteins.....	24
2.1.1 Avidin.....	24
2.1.2 Avidin protein family.....	27
2.2 (Strept)avidin-biotin technology.....	31
2.2.1 Engineered (strept)avidins	31
2.3 Directed evolution of proteins	36
2.4 Construction of mutant libraries	38
2.4.1 Random mutagenesis.....	39
2.4.2 Site-directed random mutagenesis.....	39
2.4.3 Recombination techniques	41
2.5 Mutant library selection methods.....	43
2.5.1 Phage Display	44
2.5.2 Alternative selection systems.....	47
2.6 Modified affinity proteins.....	48
2.6.1 Engineering of antibodies and antibody fragments	50
2.6.2 Alternative scaffolds to replace antibodies	53
3 AIMS OF THE STUDY	58
4 MATERIALS AND METHODS.....	59
4.1 Molecular cloning methods.....	59
4.1.1 Phagemid vectors (I-III)	59
4.1.2 Bacterial expression vectors (I, III).....	60

4.1.3	Targeted mutagenesis (III).....	60
4.1.4	Bacterial strains (I-III)	60
4.1.5	DNA sequencing (I-III)	61
4.2	Directed evolution of avidin by phage display.....	62
4.2.1	Construction of phagemid libraries (I-III)	62
4.2.2	Selection by biopanning	65
4.2.3	Validation of pannings.....	66
4.2.4	Screening (I-III)	68
4.3	Expression and purification of the recombinant proteins	69
4.3.1	Protein production of avidin variants (I-III)	69
4.3.2	Affinity purification (I-III)	70
4.3.3	Detection of protein expression via SDS-PAGE (I-III)	70
4.4	Biophysical analyses	71
4.4.1	UV-Vis spectrophotometry (I-III)	71
4.4.2	Differential scanning calorimetry (DSC) (I, III)	71
4.4.3	Size-exclusion chromatography (SEC) (I).....	71
4.4.4	Size Exclusion Chromatography with Static Light Scattering (SEC-SLS) Analysis (III)	72
4.4.5	X-ray Crystallography (III)	72
4.5	Determination of ligand binding interactions	73
4.5.1	Fluorometric assay (III).....	73
4.5.2	Surface plasmon resonance (I)	74
4.5.3	Protein Microplate Assay (I, III).....	74
4.5.4	Molecular Dynamics (MD) Simulations (III)	75
4.5.5	Interaction analysis by Molecular Recognition Force Spectroscopy (I).....	75
5	RESULTS AND DISCUSSION	76
5.1	Functional display of avidin on the M13 phage (I)	76
5.2	Library design of genetic antidin libraries (I, III).....	79
5.3	Capture and characterization of antidins (I, III).....	83
5.3.1	Steroid-binding antidins (I, III).....	85
5.3.2	Other antidins (III).....	90
5.4	Rational design of point mutations to improve the properties of the selected antidins (III).....	95
5.4.1	Antidins with enhanced thermal stability	95
5.4.2	MD simulations showed the importance of N118 for steroid binding	96
5.4.3	The crystal structure of sbAvd-2(I117Y)	96
5.5	Improving the library construction and selection.....	100
5.5.1	Construction of new phagemid vectors (II)	100
5.5.2	Shortcut in DNA shuffling protocol (II)	100
5.6	Future plans	103
6	CONCLUSIONS	105

7	REFERENCES.....	109
8	ORIGINAL PUBLICATIONS	125

ABBREVIATIONS

3D	three-dimensional
Amp ^R	ampicillin resistance
<i>AVD</i> / AVD	avidin gene / protein
<i>AVR</i> / AVR	avidin-related gene / protein
Avd	avidin
BBP	biotin-binding protein
BNP	biotinyl <i>p</i> -nitrophenyl ester
CDRs	complementarity determining regions
Cam ^R	chloramphenicol resistance
dAb	domain antibodies composed of single domain
DARPin	designed ankyrin repeat protein
dcAvd	dual-chain avidin
DNA	deoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i> bacterium
Fab-fragment	antigen-binding fragment of an antibody
Fc-region	constant region of antibody
FACS	fluorescence-activated cell sorting
Fn3	fibronectin type III domain
Fv-region	variable domain of antibody, consisting of VL and VH
GOI	gene-of-interest
HABA	2-(4'-hydroxybenzene)azobenzoic acid
Ig	immunoglobulin
ITCHY	incremental truncation for the creation of hybrid enzymes
k _a	association rate constant
K _d	dissociation constant
k _{diss}	dissociation rate constant
L3,4	loop structure of Avd between β-sheets 3 and 4
LB	Lysogeny broth
Lcn	lipocalin

LR cloning	homologous recombination -based cloning between <i>attL</i> and <i>attR</i> sites
MBP	maltose-binding protein
mRNA	messenger-RNA
Nb	nanobodies, also known as single-domain antibodies (sdAb)
PCR	polymerase chain reaction
pI	isoelectric point
RACHITT	random mutagenesis on transient templates
RNA	ribonucleic acid
POC	point-of-care
RT	room temperature
SA	streptavidin
SB	Super broth
scAvd	single-chain avidin
scFv	single-chain Fv
SCRATCHY	DNA-shuffled mixture of two ITCHY-libraries
SD	standard deviation
sdAb	single-domain antibodies, also known as domain antibodies (dAb) or nanobodies (Nb)
SHIPREC	sequence homology-independent protein recombination
SPA	staphylococcal protein A
StEP	staggered extension process
T _m	transition midpoint temperature
tRNA	transfer-RNA
VH	a heavy chain variable domain
VL	a light chain variable domain
V _{tot}	total volume
wt	wild-type

ORIGINAL PUBLICATIONS

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

- I.** Riihimäki TA, **Hiltunen S**, Rangl M, Nordlund HR, Määttä JA, Ebner A, Hinterdorfer P, Kulomaa MS, Takkinen K, Hytönen VP. Modification of the loops in the ligand-binding site turns avidin into a steroid-binding protein. *BMC Biotechnol.* 2011 Jun 9;11:64.
- II.** **Lehtonen SI***, Taskinen B*, Ojala E, Kukkurainen S, Rahikainen R, Riihimäki TA, Laitinen OH, Kulomaa MS, Hytönen VP. Efficient preparation of shuffled DNA libraries through recombination (Gateway) cloning. *Protein Eng Des Sel.* 2015 Jan;28(1):23-8. (*Equal contribution)
- III.** **Lehtonen SI**, Tullila A, Agrawal N, Kukkurainen S, Kähkönen N, Koskinen M, Nevanen TK, Johnson MS, Airenne T†, Kulomaa MS, Riihimäki TA, Hytönen VP. Artificial Avidin-Based Receptors for a Panel of Small Molecules. *ACS Chem Biol.* 2016 Jan 15;11(1):211-21.

The original publications are reproduced with the permission of the copyright holders.

RESPONSIBILITIES OF SOILI LEHTONEN IN THE ARTICLES COMPRISING THIS THESIS

Article I: Tiina Riihimäki was mainly responsible for this study. I participated in the construction of the sbAvd-1(L3,4) library, panning of both libraries Avd(L1,2) and sbAvd-1(L3,4), characterization of the antidins sbAvd-1 and sbAvd-2, and the writing of the article. The MRFS-analyses were done by Martina Rangl at Johannes Kepler University Linz, Austria.

Article I has also been published earlier as part of the PhD thesis of Tiina Riihimäki.

Article II: I was mainly responsible for planning, practical work, and the writing of the article. Barbara Taskinen contributed significantly to executing some of the experiments and the writing process.

Article III: I was mainly responsible for planning, practical work, and writing of the article. Antti Tullila participated in the selection of the phages, carried out at the laboratory of VTT Biotechnology (Espoo). The crystal structure of sbAvd-2(I117Y) was solved by Nitin Agrawal and Tomi Airenne. Tiina Riihimäki participated in the construction of the phage display libraries, as well as the writing process.

All these studies were executed under the supervision of Professor Emeritus Markku Kulomaa and Associate Professor Vesa Hytönen.

1 INTRODUCTION

Naturally evolved proteins, usually composed of 20 standard amino acids, are responsible of all the cellular processes that make life possible. These processes include molecular transport, catalysis, signaling, and ligand recognition, just to name some. In order to facilitate these processes, a protein must first recognize its target, whether it is DNA for gene regulation, an antigen for triggering an immune response or a substrate for catalysis. Molecular recognition is central to all biological processes. Ligand binding occurs via a combination of weak, non-covalent interactions: ionic, hydrogen-bonding, and hydrophobic interactions. Additionally, shape complementarity is essential (Banta *et al.*, 2013).

In addition to biochemical phenomena, biochemical measurements and diagnostics also depend on specific and high affinity interactions between molecules. The reagents capable for selectively recognizing biomolecules are essential in many areas of the life sciences including bioseparation, diagnostics, imaging, and therapy (Nygren, 2008). Antibodies, products of the humoral immune response, have been utilized for almost a century both in the life sciences and medicine. They enable efficient responses against almost every foreign macromolecular substance because of two distinct mechanisms: 1) The robust immunoglobulin domain architecture of antibodies consists of a rigid scaffold that supports six hypervariable loops, capable of forming highly diverse binding sites. 2) An efficient genetic mechanism creates sequence diversity step-wise at the somatic level, whereby an inherited set of gene segments is randomly recombined and followed by hypermutation events (Skerra, 2003).

Although it is possible to analyze protein structures and show which amino acid residues are important and crucial for certain protein-ligand interactions, we are still far from the situation where one could design a protein scaffold with desired functional properties completely *de novo* (Tracewell and Arnold, 2009). Directed evolution is a method carried out in the laboratory and aims to mimic natural evolution. Increased understanding of the molecular mechanisms behind the immune system has enabled adaptation of these principles *in vitro* using combinatorial protein engineering principles in creation of both antibodies or antibody fragments, and artificial binding proteins (Skerra, 2003). All these

combinatorial strategies require 1) a library (a pool of single gene variants), and 2) a means of screening or selecting from that library (Neylon, 2004). Different mutagenesis methods are available for creating a library from which random, oligonucleotide-directed mutagenesis is the most applied when scaffold-based new binders are engineered. Directed evolution has proven to be an invaluable tool for protein engineering: It has been successfully used to improve various protein properties (e.g. catalytic activity, thermostability, enantioselectivity and binding affinity) in thousands of experiments (Nov, 2014).

During the last 20 years, novel types of binding proteins developed by combinatorial biotechnology have challenged classical antibody-based affinity reagents. Functional selection and screening enables affinity reagents (both protein and nucleic acid-based) to be routinely identified from a range of different libraries based on their binding ability towards the desired target structure (Nygren, 2008). However, this thesis will only deal with protein-based affinity reagents. So-called protein scaffolds, or alternative scaffolds, have been used to generate novel types of binding proteins for various applications both in research and medicine (Skerra, 2000).

The ultimate aim of this PhD thesis was to engineer the binding site of chicken avidin protein with novel target specificities via directed evolution methods using phage display for selection. Avidin, and its bacterial analogue, streptavidin, are exceptional proteins with their extraordinarily high affinity towards their ligand biotin. Together they form (strept)avidin-biotin technology, which has been studied and utilized extensively over the past four decades (Avraham *et al.*, 2015). Although the basis for numerous biotechnological applications rest upon the high affinity between these proteins and their ligand, biotin, the robustness of these proteins further promotes and diversifies their applications. Directed evolution of avidin was inspired by the success of engineering the lipocalin scaffold belonging to the same ‘calycin’ superfamily as avidin. The term ‘antidin’ was adopted following the example of many other alternative scaffolds modified to be used in novel binding purposes like antibodies. Antidins are prepared by reshaping the ligand-binding pocket of avidin via protein engineering in order to recognize a novel ligand.

In this study we used random oligonucleotide-directed mutagenesis to engineer the antidins. The amino acid residues in the loop structures that participate in biotin binding were randomized and cognate variants with affinity towards the used target molecules were selected from the resulting libraries. Antidins were produced in *E. coli* and subsequently studied in detail for their structure and function. Furthermore, a more efficient method for the subcloning of a DNA shuffled library was

established through generation of a new pGWphagemid vector compatible with Gateway® cloning.

2 REVIEW OF THE LITERATURE

2.1 Avidin and other biotin-binding proteins

The members of the avidin protein family are well known for their high affinity towards D-biotin and structural stability. In addition to avidin, found as a minor component in the egg-white of chicken, chickens also contain avidin related proteins closely resembling egg-white avidin (Laitinen *et al.*, 2002). Together they form the chicken avidin gene family (see section 2.1.2). Additionally, similar “avidins” have been found from both prokaryotes and eukaryotes, but not yet in any mammalian species (see section 2.1.2). Moreover, the chicken genome encodes also other biotin-binding proteins: chicken egg yolk has been found to contain biotin-binding proteins I and II (BBP-I/-II) (White and Whitehead, 1987).

Avidin along with these other biotin-binding proteins belong to the ‘calycins’, protein structural superfamily, which also contains lipocalins, fatty-acid binding proteins, a group of metalloproteinase inhibitors, and triabin (Flower *et al.*, 2000). In addition to overall structure similarity (a repeated +1 topology β -barrel) shared by the members of the calycin superfamily, calycin proteins have conserved main chain conformations, amino acid side chains, and thus also the interactions they make. These properties form a structural signature characteristic of the superfamily (Flower, 1993; Flower *et al.*, 2000).

2.1.1 Avidin

Avidin is a positively charged chicken egg-white protein that binds extremely tightly to its natural ligand, the small water-soluble vitamin, D-biotin. The affinity (of femtomolar) between avidin and biotin is the strongest non-covalent interaction known to nature (Green, 1975). A typical antigen:antibody complex is three to six orders of magnitude weaker (Hudson and Souriau, 2003). Avidin is a homotetramer (~60 kDa), where each subunit of 128 amino acids is arranged in an eight-stranded antiparallel (up-and-down) β -barrel with the D-biotin binding site inside the cavity (Rosano *et al.*, 1999) (Figure 1).

The high affinity of the avidin:biotin interaction can be explained by 1) shape complementarity, 2) several hydrogen bonds, and 3) hydrophobic interactions. Out of these the first mentioned, shape complementarity is the most important parameter. The biotin-binding site is a deep, pear-shaped pocket, whose volume as well as three-dimensional structure and orientation of the residues participating in hydrogen bonding with the vitamin are predetermined as complementary to that of the incoming vitamin (Livnah *et al.*, 1993; Rosano *et al.*, 1999). In the absence of biotin, the biotin-binding pocket contains five molecules of water that mimic the structure of biotin in the binding site until biotin is bound (Rosano *et al.*, 1999): In streptavidin the water molecules have been found to vacate the binding site escaping through the water channel near the back of the binding pocket as biotin enters (Hyre *et al.*, 2002). In the apo-protein (without ligand) the vitamin-binding pocket is fairly open due to the flexible L3,4-loop, thus allowing fast access of biotin to the binding site. When biotin is bound, it is buried inside the central pocket of the β -barrel-structured protein with the vitamin's bicyclic ring at the bottom of the cavity. The vitamin is trapped by conformational readjustments of the protein primarily involving the stiffened L3,4-loop, but also L5,6. During the process, three amino acid residues of L3,4 contribute additional interactions with biotin (Rosano *et al.*, 1999). Thus, the high affinity of biotin to avidin stems from an extremely slow dissociation rate (Green, 1963a; Green, 1990).

Biotin binding of avidin involves a highly stabilized network of polar and hydrophobic interactions. In the biotin-bound state, five aromatic residues form a "hydrophobic box" for biotin binding, and five hydrogen bonds are formed with an ureido-ring and with valeryl side chain of biotin, respectively (Livnah *et al.*, 1993). Additionally, the subunits one and two interact functionally (forming a functional dimer), where the major agent, the W110 located in L7,8, participates in biotin binding with the neighboring subunit (and vice versa) (Livnah *et al.*, 1993). X-ray studies (Livnah *et al.*, 1993; Pugliese *et al.*, 1993; Rosano *et al.*, 1999) have shown that for such a high affinity interaction as the avidin-biotin pair forms, shape complementarity is more important than hydrogen bonds and hydrophobic interactions. The nature of this interaction is thus exceptional in the sense of binding energy per atom (Kuntz *et al.*, 1999).

Besides its strong biotin-binding ability, avidin is an exceptionally stable protein based on its high thermostability (Gonzalez *et al.*, 1999) and its ability to resist proteases (Hiller *et al.*, 1991; Ellison *et al.*, 1995), as well as extreme pH and other denaturing conditions (Green, 1975). Stability is already high without its ligand, but as the biotin-bound form it is even stronger (Gonzalez *et al.*, 1999). As avidin is a

tetramer, the protein contains altogether three regions of monomer-monomer interactions: subunit interfaces 1–2 (forming the already mentioned functional dimer), 1–3 (which is important for the oligomeric stability of the protein) and 1–4 (forming the structural dimer). All these interactions contribute to the rigidity of the quaternary structure (Livnah *et al.*, 1993).

Avidin is a positively charged protein (pI 10.5) (Melamed and Green, 1963) due to the eight arginine and nine lysine residues that each monomer possesses (DeLange and Huang, 1971). In chicken, avidin is glycosylated (10% carbohydrates of its composition) at residue Asn17, but glycosylation is not found to be important for the stability of the protein nor the ability to bind biotin (Hiller *et al.*, 1987; Bayer *et al.*, 1995; Wang *et al.*, 1996). The protein has been successfully expressed and purified in various organisms: *E. coli* (Hytönen *et al.*, 2004a), insect cells (Airenne *et al.*, 1997), *Pichia pastoris* (Zocchi *et al.*, 2003), and even in corn (Kusnadi *et al.*, 1998).

Besides biotin, avidin has been shown to be able to bind various biotin derivatives (Green, 1963b), as well as some other small molecules that are chemically different from biotin: an azo dye, commonly known as HABA (2-(4'-hydroxybenzene)azobenzoic acid) (Green, 1965) and other azo molecules, including a phenyl derivative of HABA (Repo *et al.*, 2006).

2.1.2 Avidin protein family

2.1.2.1 Chicken avidin gene family

In a chicken, avidin is expressed in the egg-laying chick's oviduct under the influence of the steroid hormone, progesterone (Hertz *et al.*, 1943; O'Malley and McGuire, 1968; Korenman and O'Malley, 1968). Avidin has also been shown to be produced in a number of other tissues of both male and female chickens after bacterial or viral infections, inflammation, or tissue trauma (Elo *et al.*, 1979a; Elo *et al.*, 1979b; Elo *et al.*, 1980a; Elo *et al.*, 1980b).

In addition to egg-white avidin, the chicken avidin gene family also contains avidin homologues: The seven other members are known as avidin-related genes (*AVRs*) from which *AVR4* and *AVR5* have identical coding sequences. The other *AVRs* are 94–99% identical, whereas the identity between the different *AVRs* and *AVD* genes ranges from 91% to 95%. The number of differing *AVR* genes seems to differ between chickens and even between cells within the same chicken (Kunnas *et al.*, 1993).

Compared to avidin, the AVR proteins (with sequence identity between 69–78% with avidin) have numerous amino acid substitutions located in subunit interface regions while most of the biotin-binding residues are preserved. Despite this they form extraordinarily stable tetramers similar to those of avidin. Differences however, are found in biotin binding and physico-chemical properties: glycosylation and charge properties. Additionally, avidin and AVR2 differ from all other AVRs, which have an uneven number of cysteine residues, thus enabling most of the AVRs to form inter-subunit disulphide-bridges in addition to intra-subunit disulphide bonds (Laitinen *et al.*, 2002; Hytönen *et al.*, 2005a).

AVRs have a rigid conserved L3,4 structure, which explains the high heat stability of AVRs exceeding that of avidin's. AVR4/5 (henceforth AVR4) displays the highest heat stability among all the biotin-binding proteins so far characterized (Eisenberg-Domovich *et al.*, 2005), which can mostly be explained by the I117Y substitution and the different L3,4-loop as compared to that of avidin. Furthermore, the sequence difference between $\beta 3$ and $\beta 5$ of avidin and AVRs, although important for heat stability, can explain the heightened hydrolytic activity towards biotinyl *p*-nitrophenyl ester (BNP) (Hayouka *et al.*, 2008) and the weaker biotin-binding affinity compared to avidin (Laitinen *et al.*, 2002). Biotin-binding affinities, however, vary over a wide range of values: AVR4 having almost as high an affinity as avidin ($K_d \sim 3.6 \times 10^{-14} \text{M}$) (Hytönen *et al.*, 2004b) and AVR2 having the lowest affinity ($k_{\text{diss}}(\text{AVR2}) \sim 215.55 \times 10^{-6} \text{s}^{-1}$, $k_{\text{diss}}(\text{AVR4}) \sim 0.18 \times 10^{-6} \text{s}^{-1}$, $k_{\text{diss}}(\text{AVD}) \sim 0.05 \times 10^{-6} \text{s}^{-1}$) (Hytönen *et al.*, 2005a). The K109I mutation of AVR2 seems to be the critical difference between AVR2 and all other AVRs, partially explaining its weaker biotin-binding affinity (Hytönen *et al.*, 2005a).

The theoretical pIs of AVR3 and AVR4 are basic, resembling avidin (~ 10), while AVR1, AVR6, and AVR7 are neutral ($\text{pI} \sim 7$), and AVR2 is acidic with a $\text{pI} \sim 5$. There is variation also in the glycosylation pattern of these proteins. Polyclonal anti-avidin detects AVRs poorly, and monoclonal antibodies raised against avidin cannot recognize any AVRs. Additional glycosylation patterns may mask antibody epitopes. Moreover, the variability among amino acid residues on the surface of the protein can explain this result (Laitinen *et al.*, 2002).

2.1.2.2 Other members of avidin protein family

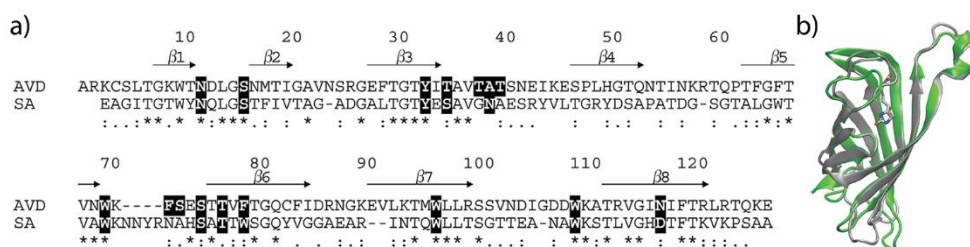


Figure 2. The comparison of avidin (Avidin) and streptavidin (SA). (A) An alignment of Avidin and SA. The numbering is according to the avidin sequence. The residues forming direct interactions with biotin are shown in black. The secondary-structure elements are indicated by arrows on top of the alignment. In the case of avidin, an intrasubunit disulfide bridge is formed between Cys4 and Cys83. (B) The superimposed monomer structure of Avidin (in grey, [PDB:2AVI]) (Livnah *et al.*, 1993) and SA (in green, [PDB:1MK5]) (Hyre *et al.*, 2006) with biotin (shown in sticks). Subfigure A is modified from (Laitinen *et al.*, 2007), and B is drawn by Vesa Hytönen with VMD 1.9.1 software using the coordinates [PDB: 2AVI] (Livnah *et al.*, 1993) and [PDB:1MK5] (Hyre *et al.*, 2006). The coordinates of the biotin residue shown are from avidin-biotin structure. Subfigure A is reprinted from (Laitinen *et al.*, 2007) with permission from Elsevier (copyright 2007).

Other avidins have been found and characterized both from eukaryotic and prokaryotic species. Out of these, streptavidin is the best known, due to its almost as high biotin-binding affinity ($K_d \sim 10^{-14}$ M) (Green, 1990) compared to chicken avidin. Streptavidin is a bacterial analogue of avidin secreted by *Streptomyces avidinii* (Tausig and Wolf, 1964; Chalet and Wolf, 1964). Later, *S. venezuelae* was also found to produce streptavidins (Bayer *et al.*, 1995). Although streptavidin is functionally and structurally highly similar to avidin, it has quite a different primary sequence (Figure 2) (Argarana *et al.*, 1986). However, the amino acid residues participating directly in biotin binding are well conserved, and thus the main difference in primary structure between avidin and streptavidin is found on the loop structures: They seem to have gone through deletions and insertions during evolution. Additionally, streptavidin contains 25 alanines compared to the five of avidin, thus it has neutralized its pI (6.9) through the natural alanine scanning process (Wilchek and Bayer, 1999). As streptavidin is a prokaryotic protein, it is not glycosylated like chicken avidin.

In eukaryotes, similar avidins have been found in all egg-laying species: birds, reptiles, and amphibians (Hertz and Sebrell, 1942; Jones and Briggs, 1962; Korpela *et al.*, 1981). Hytönen *et al.* (2003) have characterized poultry egg-white avidins from

Table 1. Different avidins found from eukaryotic and prokaryotic species.

name	from	biotin-binding (K _d)	oligomeric state	calculated pI	thermal stability DSC (T _m) / SDS-PAGE (T _r)	references
Avidin	chicken, <i>Gallus gallus</i> and <i>Gallus domesticus</i>	1×10 ⁻¹⁵ M	tetrameric	10.4	T _m : 83 °C → 117 °C T _r : 60 °C → 90 °C	(Green, 1975; Laitinen <i>et al.</i> , 1999)
Xenavidin	a frog, <i>Xenopus tropicalis</i>	1× 10 ⁻¹³ M	tetrameric	8.9	T _m : n/d T _r : <22 °C → 75 °C	(Määttä <i>et al.</i> , 2009)
Zebavidin	a zebrafish, <i>Danio rerio</i>	~10 ⁻⁹ M	tetrameric	7.9	T _m : 67.8 °C → 80.0 °C T _r : <22 °C → ~60–70 °C	(Taskinen <i>et al.</i> , 2013)
Bjavidins	amphioxus <i>Branchiostoma japonicum</i>	1: 1.6 × 10 ⁻⁶ M 2: 8.6 × 10 ⁻⁸ M *	tetrameric	1: 6.8 2: 4.7	T _m : n/d T _r : n/d	(Guo <i>et al.</i> , 2017)
Strongavidin	a sea urchin, <i>Strongylocentrotus purpuratus</i>	> 10 ⁻¹² M	tetrameric?	3.9	T _m : n/d T _r : <22 °C → 70 °C	unpublished (Veneskoski, 2009)
Tamavidins	a basidiomycete fungus, <i>Pleurotus cornucopiae</i>	strong affinity 1: for biotin 2: for biotin and 2-iminobiotin resembling (strept)avidins	tetrameric	1: 6.2 2: 7.4	1: n/d 2: T _r : 78 °C → >99.9 °C	(Takakura <i>et al.</i> , 2009)
Lentiavidins	shiitake mushroom, <i>Lentinula edodes</i>	n/d Lentiavidin 1 binds biotin	n/d	1: 3.9 2: 4.4	T _m : n/d T _r : n/d	(Takakura <i>et al.</i> , 2016)
Streptavidin	gram ⁺ soil bacteria <i>Streptomyces avidinii</i> and <i>S. venezuelae</i>	~10 ⁻¹⁴ M	tetrameric	6.1	T _m : 75 °C → 112 °C	(Tausig and Wolf, 1964; Chaiet and Wolf, 1964)
Bradavidin I	symbiotic nitrogen-fixing, gram ⁺ bacteria <i>Bradyrhizobium japonicum</i>	~10 ⁻¹⁰ M	tetrameric	6.3	T _m : n/d T _r : 65 °C → 85 °C	(Nordlund <i>et al.</i> , 2005b; Leppiniemi <i>et al.</i> , 2012)
Bradavidin II		<10 ⁻¹⁰ M	without clearly defined oligomeric state	9.6	T _m : 75 °C → 98 °C T _r : n/d	(Helppolainen <i>et al.</i> , 2008; Leppiniemi <i>et al.</i> , 2013)
Rhizavidin	symbiotic nitrogen-fixing, gram ⁺ bacteria <i>Rhizobium etli</i>	tight, comparable to bradavidin I	dimeric	4.0	T _m : 75 °C → 101 °C T _r : n/d	(Helppolainen <i>et al.</i> , 2007; Meir <i>et al.</i> , 2009)
Shwanavidin	marine, gram ⁻ proteobacterium <i>Shewanella denitrificans</i>	tight, similar to that for rhizavidin and streptavidin	dimeric	4.7	T _m : 74 °C → n/d (>95 °C) T _r : n/d	(Meir <i>et al.</i> , 2012)
Hoefavidin	marine, gram ⁻ α-proteobacterium, <i>Hoeflea phototrophica</i> DFL-43 ^T	tight, comparable to rhizavidin	dimeric	4.0	T _m : 85 °C → 96 °C T _r : n/d	(Avraham <i>et al.</i> , 2015)
Burkavidin	a soil-dwelling bacteria, <i>Burkholderia pseudomallei</i>	K _d < 10 ⁻⁷ M (probably underestimate)	tetrameric	5.3	T _m : n/d T _r : 80 °C → >99 °C	(Sardo <i>et al.</i> , 2011)

*(analyzed by ELISA

gram^{+/−} gram-positive or -negativeeukaryotic
prokaryotic

duck, goose, ostrich, and turkey (Hytönen *et al.*, 2003). Table 1 summarizes the different avidins characterized from various species, from both eukaryotes and prokaryotes. In addition of these, several avidin-like genes have been found, however they do not bind biotin or have very low affinity towards it (e.g. rhodavidin, fibropellin, and burkavidin1).

2.2 (Strept)avidin-biotin technology

The recombinant forms of chicken avidin and its bacterial analogue, streptavidin, collectively (strept)avidin, are proteins that are widely used in a number of diverse applications in the life sciences, e.g. in protein purification, labeling techniques, nanotechnology, diagnostics, and targeted drug delivery. They can be efficiently produced by both prokaryotic and eukaryotic expression systems (Sano and Cantor, 1990; Hytönen *et al.*, 2004a; Airene *et al.*, 1997; Zocchi *et al.*, 2003; Kramer *et al.*, 2000). (Strept)avidin-biotin technology is based on the extremely tight and specific affinity between (strept)avidin and biotin ($K_d \approx 10^{-14}$ – 10^{-16} M) (Green, 1990), as well as their hyper-thermostability: (Strept)avidins can tolerate heat, denaturants, low and high pHs, and are even resistant to the activity of a number of proteolytic enzymes. A flexible protein scaffold of (strept)avidins can be adjusted for different approaches through protein engineering, and biotin can be easily chemically coupled to different molecules. To conclude, this is a rare protein class that is versatile and capable of providing scaffolds for multiple purposes. Furthermore, revealing the molecular details explaining these unexceptional properties has been one driving force explaining the interest for (strept)avidins. For reviews, see Laitinen *et al.*, 2006 and Laitinen *et al.*, 2007.

2.2.1 Engineered (strept)avidins

Over the years, both avidin and streptavidin have been extensively modified. Although avidin was found much earlier than streptavidin, the corresponding gene was not cloned until 1995 (Wallén *et al.*, 1995), whereas streptavidin was cloned already in 1986 (Argarana *et al.*, 1986). The X-ray structure was also determined earlier for streptavidin due to the lack of glycosylation (Hendrickson *et al.*, 1989; Weber *et al.*, 1989). Thus, streptavidin characterization has led the way in elucidating the high biotin-binding affinity. One aim has been to reveal the different aspects

influencing the exceptional properties of (strept)avidins, the high-affinity ligand binding, and structural stability. Another point has been in applications to regulate the physicochemical and biotin-binding properties, and thus broaden the potential application spectrum. Therefore another aim has been to overcome some of the inherent drawbacks of the (strept)avidin-biotin system: 1) irreversibility of the interaction, 2) aggregation tendency caused by tetramerization in some applications, and 3) non-optimal pharmacokinetics in *in vivo* applications (Laitinen *et al.*, 2006). Although this thesis is about avidin, some streptavidin modifications are also included in this chapter since the biotin-binding residues are well conserved: The functional changes due to modifications in streptavidin have been found to take place in avidin as well, and vice versa.

2.2.1.1 Modifications for biotin binding and stability

As avidin and streptavidin are considered extreme examples of tight ligand binding, both rational and site-directed random mutagenesis has been used to study the role of different biotin-binding residues in a vast amount of publications (reviewed in Laitinen *et al.*, 2007). As the wild type (wt) binding site displays a virtually perfect fit with biotin, mutants with reduced affinity can therefore be expected upon the introduction of almost any kind of mutation in the respective area (Wilchek and Bayer, 1999).

First, the hydrophobic lining of the biotin-binding pocket was studied. As the importance of tryptophan residues for avidin had been shown by Green already in 1963 (Green, 1963b), expectedly, substituting the tryptophan residues of streptavidin one by one with alanine or phenylalanine resulted in reduced biotin and 2-iminobiotin affinities (Chilkoti *et al.*, 1995). Then a similar approach was used to study the residues forming hydrogen bonds with biotin ureido oxygen (Klumb *et al.*, 1998; Marttila *et al.*, 2003). Avidin mutant Y33H was shown to bind biotin in a pH-dependent manner (Marttila *et al.*, 2003). The D128A mutation of streptavidin, breaking a hydrogen bond to a ureido NH group, was shown to be an important intermediate in a simulated dissociation pathway of biotin from streptavidin, as it resulted in a concerted structural alteration of bound biotin and binding contact residues (Freitag *et al.*, 1999; Hyre *et al.*, 2002).

Tetramer stability of (strept)avidin has been enhanced through interprotomer disulphide bridges (Reznik *et al.*, 1996; Nordlund *et al.*, 2003b), which yielded higher transition midpoint temperatures (T_m) in the absence of biotin, but did not have a major effect on biotin-binding properties. Furthermore, the naturally existing

intrasubunit disulfide bridges (between Cys4 and Cys83) were shown to be important for the stability of avidin: Mutated cysteine residues caused lower T_m values in the absence of biotin (close to the T_m of apo-streptavidin (Gonzalez *et al.* 1999)), although biotin was able to restore the stability to be comparable with wt avidin (Nordlund *et al.*, 2003b). Streptavidin is naturally devoid of cysteines (Argarana, *et al.* 1986).

Since AVR4 was found to be the most stable biotin-binding protein characterized so far (T_m 106.4 °C) (Hytönen *et al.*, 2004b), molecular modelling (Hytönen *et al.*, 2004b) and detailed structural analysis of AVR4 (Eisenberg-Domovich *et al.*, 2005) inspired the development of more stable avidin mutants (Hytönen *et al.*, 2005b). Using chimeragenesis to combine a 21-amino acid segment from AVR4 with avidin, a significantly more stable (T_m 96.5 °C) chimeric avidin protein, ChiAvd, was developed (compared to native avidin, T_m 83.5 °C) resembling AVR4 with its biotin-binding properties and resistance against proteinase K (Hytönen *et al.*, 2005b). Introducing an AVR4-inspired point mutation into a subunit interface of avidin resulted in Avd(I117Y), which compared to avidin had significantly increased thermostability (T_m 97.5 °C) but preserved its high biotin-binding properties (Hytönen *et al.*, 2005b). Finally, by combining chimeragenesis with point mutation, a hyperthermostable ChiAVD(I117Y) was constructed (T_m 111.1 °C) (Hytönen *et al.*, 2005b). Later, ChiAvd(I117Y) was shown to be resistant to various harsh organic solvents (Määttä *et al.*, 2011), and it could even be printed using an ink-jet printer (Heikkinen *et al.*, 2011).

In attempt to turn subunit association and biotin binding of avidin into a pH-sensitive phenomena, individual amino acid residues have been replaced with histidines. Out of the resulting mutants, Avm(M96H), Avm(M96H, W110H), and Avm(I117H, W110H) showed consistently predictable behavior, and were thus the most promising variants from an application point of view (Nordlund *et al.*, 2003a). Later, the Avm(M96H) mutant was named a switchable avidin mutant, as it was found to disassemble into its monomers by treatment with a combination of mild acid with SDS (Pollheimer *et al.*, 2013). In order to improve the variant, point mutations were added to lower the net charge of the protein and a R114L-mutation improved the affinity towards conjugated biotins, thus creating Switchavidin (Taskinen *et al.*, 2014b).

2.2.1.2 Modifications for oligomericity and valency

Stability of the avidin structure has been studied by introducing mutations into the interface regions of the protein (Laitinen *et al.*, 1999; Laitinen *et al.*, 2001). A fully monomeric avidin form, monoavidin, was produced in a subsequent study by mutating two interface residues: W110K and N54A (Laitinen *et al.*, 2003). The biotin-binding affinity decreased to $\sim 10^{-7}\text{M}$, the monomer was more sensitive to proteinase K digestion, and it was only weakly recognized by a polyclonal antibody (Laitinen *et al.*, 2003). In the case of streptavidin, the double mutant Q95A, W120K (analogous to monoavidin) was oligomeric, but showed a significant decrease in biotin-binding affinity (Wu and Wong, 2005). Instead, the monomeric form of streptavidin was achieved by mutating two biotin-binding residues T90A and D128A (Qureshi and Wong, 2002). These results show that although avidin and streptavidin have quite similar characteristics, it is not always possible to use direct analogy between these two proteins in order to achieve the desired mutants.

From an application point of view, it is beneficial to be able to alter the biotin-binding affinity of some subunits while preserving the high affinity of the rest. Thus dual-chain avidin (dcAvd) with two distinct biotin-binding sites in the form of a polypeptide was constructed (Nordlund *et al.*, 2004). This was accomplished through creating two different circularly permuted forms of avidin (cpAvd5 \rightarrow 4 and cpAvd6 \rightarrow 5), which connect by a short glycine-serine-rich linker. These cpAvds introduced new termini located in loops (L4,5 and L5,6, respectively) (Figure 3) (Nordlund *et al.*, 2004). The dcAvd scaffold formed dimers (pseudotetramers) in solution, resembling wt Avd with its functional and structural characteristics. The scaffold was later further engineered by insertion of point mutations I117C in cpAvd5 \rightarrow 4 and V115H in cpAvd6 \rightarrow 5 in order to obtain a dcAvd derivative, dcAvd(I117C \rightarrow 4I117H \rightarrow 5), assembling only into one of the theoretical two conformations (Hytönen *et al.*, 2006). Introducing mutations into selected subunits yielded dual-affinity avidins (Hytönen *et al.*, 2005c; Hytönen *et al.*, 2006; Leppiniemi *et al.*, 2011; Riihimäki *et al.*, 2011a). Since dcAvd enables combining of only two different binding sites, the scaffold was modified further to yield single-chain avidin (scAvd): It consists of two dcAvds connected by another GS-rich linker and enables modification of each of the subunits independently (Nordlund *et al.*, 2005a).

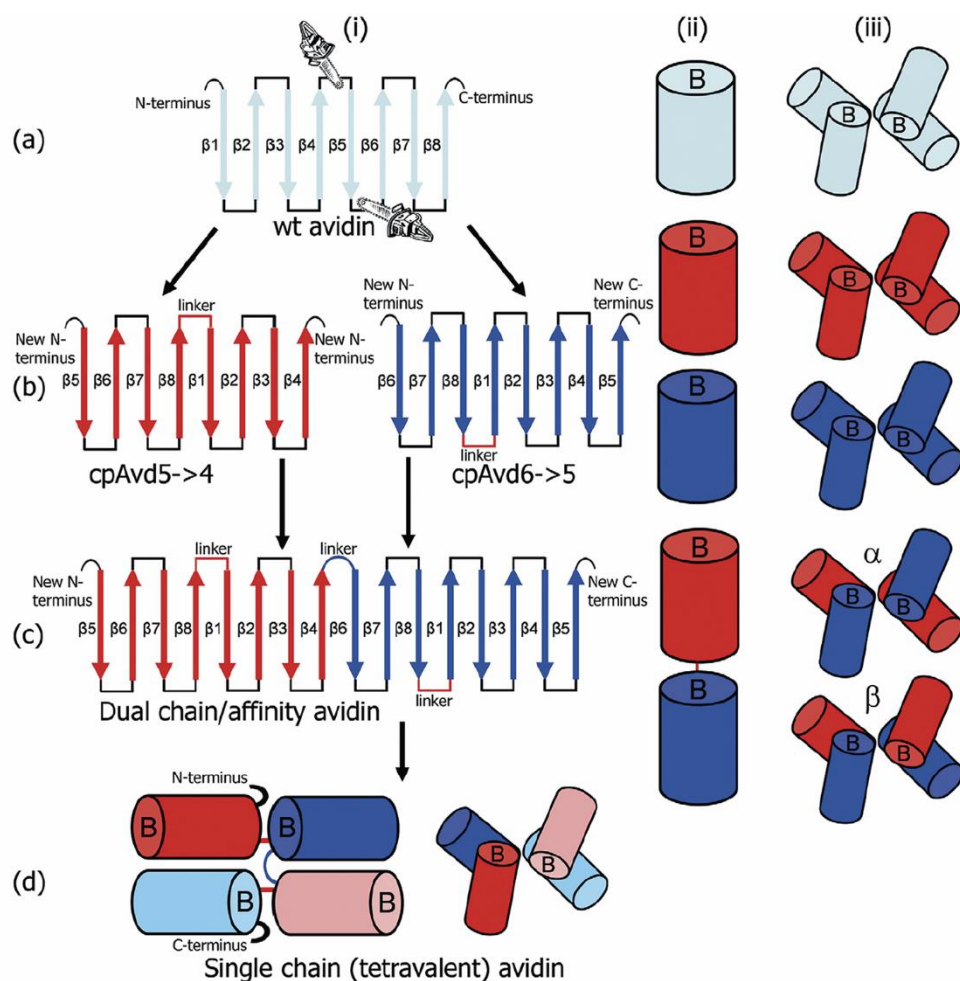


Figure 3. A schematic figure showing the engineering strategy of dual-chain (dcAvd) and single-chain avidins (scAvd). a) Wild-type avidin topology (i) with its tertiary (ii) and quaternary structure (iii). b) At first two different circularly permuted avidin mutants were generated: One with new termini introduced into avidin L4,5 (the upper saw in a), resulting in the circularly permuted avidin (cpAvd)5→4; and another with the new termini in L5,6 (the lower saw in a), resulting in cpAvd6→5. The quaternary structure of these cpAvids are composed of four identical subunits as in the case of wt Avid. c) Dual-chain avidin is generated by joining the C-terminus of cpAvid5→4 to the N-terminus of cpAvid6→5, and it enables combining of two different properties together by modifying only one of its domains (Nordlund *et al.*, 2004). However without further modifications, two different quaternary structure outcomes are possible (Hytönen *et al.*, 2006). d) By fusing two dcAvids together, a single-chain avidin (scAvid) is generated enabling combination of four independently modified subunits (Nordlund *et al.*, 2005a). Figure reprinted from Laitinen *et al.*, 2007, with permission from Elsevier.

2.2.1.3 Other modifications to improve applicability

In addition to the above-mentioned, charge properties (Nardone *et al.*, 1998; Marttila *et al.*, 1998), glycosylation (Marttila *et al.*, 2000), biodistribution (Chinol *et al.*, 1998), and ligand specificity (Reznik *et al.*, 1998) of (strept)avidins have also been modified. The high pI (10.5) of avidin has hindered its use and preferred the usage of the naturally quite neutral streptavidin (pI 6.1) even though chicken egg-white avidin is more readily available and is thus less expensive. A set of charge-reduced mutants of avidin therefore have been developed, without compromising the high biotin-binding affinity (Nardone *et al.*, 1998; Marttila *et al.*, 1998). In addition to the basic charge, glycosylation can also cause non-specific binding, and thus deglycosylated variants of avidin have been developed (Marttila *et al.*, 2000; Hytönen *et al.*, 2004a). Biodistribution is important in the applications involving *in vivo* conditions: Chinol *et al.* (1998) have shown that PEGylated avidins raise plasma half-life and yields lower liver and kidney to blood ratios compared to unPEGylated avidins (Chinol *et al.*, 1998). Reznik *et al.* (1998) showed how ligand-binding specificity could be fine-tuned to direct the binding to favor 2-iminobiotin instead of biotin as a ligand (Reznik *et al.*, 1998).

2.3 Directed evolution of proteins

Directed evolution is an important tool for protein engineering, and enables the generation of protein variants with improved or novel properties. In order to mimic the process of natural evolution, directed evolution, which refers to the artificial selection process conducted in the laboratory, requires first the construction of a protein library from which to select the protein variants with desired properties. There are different options for constructing the library (described in the next chapter), and the choice of the method depends on the scaffold to be engineered as well as the aim. Moreover, the designated strategy to select (chapter 2.5) the desired variants has an effect on the design process as well, since it defines the upper limit of the library size. The design process is one of the most important steps, as the other parts (constructing the library and selection process) cannot compensate for bad design no matter how well they are performed.

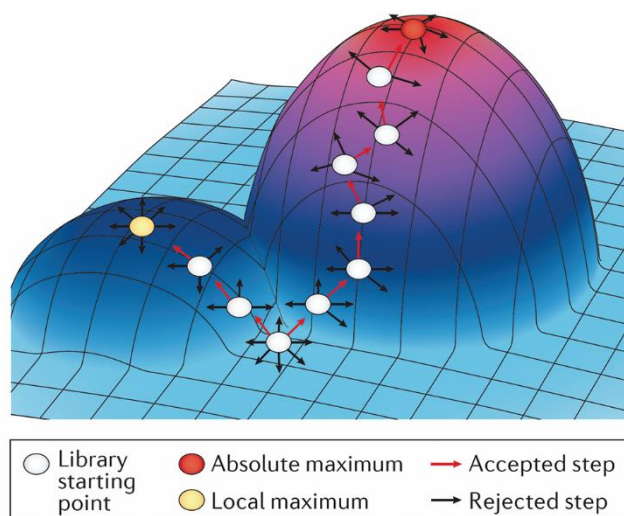


Figure 4. Fitness landscape. Search for the fittest variant in directed evolution is usually visualized as a series of steps within a 3D fitness landscape, where sequence space (all the possible genotypes of the library members) is shown as a square with x- and y-axes, and the z-axis illustrates the fitness of the individual clone towards the target. Thus, the goal of directional evolution is to climb towards peak activity levels. Directed evolution can lead to absolute maximum activity levels but can also become trapped at local fitness maxima in which library diversification is insufficient to cross “fitness valleys” and access neighboring fitness peaks. Figure is modified from Packer and Liu, 2015 and reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Genetics, copyright (2015).

The process of designing new binding proteins starts with scaffold selection (Banta *et al.*, 2013). The protein scaffold describes a polypeptide framework with a well-defined 3D-structure that can be modified through mutations or insertions and deletions. The libraries, often consisting of up to tens of billions of different variants, are generated at the DNA level, followed by expression. By applying selective pressure, variants with a desired phenotype can then be isolated from the library. The term library size refers to the total number of transformants. A library’s effective or functional size, on the other hand, is the number of distinct, functional full-length protein variants in the library. This number can differ significantly from the library size, depending on the strategy used for library construction, and how successfully it was used. The sequence space is defined as the pool of all the protein sequence variants that could possibly be generated and included in the library. The process of directed evolution is often illustrated with a picture, called a fitness landscape (Figure 4) (Nov, 2014).

2.4 Construction of mutant libraries

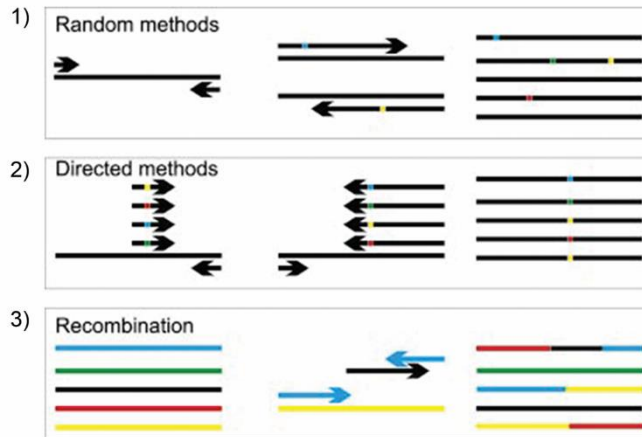


Figure 5. Alternative approaches for creation of protein-encoded DNA-libraries: 1) Random mutagenesis, where changes are created at random along a whole gene. 2) Site-directed random mutagenesis that involve randomization at specific positions within a gene sequence. 3) Recombination techniques, which do not directly create new sequence diversity but instead combine existing diversity in new ways by bringing together portions of existing sequences and mixing them in novel combinations. The bars in the figure represent genes, the short arrows primers in a PCR reaction, and the longer arrows nascent (elongating) DNA-strands. Different colours represent different sequences. Thus, variability is created by introducing mutations in the first two, while the third method combines existing sequences based on their homology. Figure is modified and reprinted by permission from Neylon, 2004.

For library construction, one first needs to select the appropriate protein scaffold, or more precisely, the DNA encoding it. Methods for creating protein-encoding DNA libraries can be divided into three approaches: 1) Random mutagenesis, 2) site-directed random mutagenesis, and 3) recombination techniques, which combine portions of existing sequences and mix them by creating hybrid proteins with novel combinations (Figure 5). Homologous recombination employs mutations already found in natural homologous proteins, which are shown to be functional in nature, thus improving the functionality of the constructed library. Mutagenesis approaches, on the other hand, involving randomization of amino acid residues can have unwanted and unexpected influences on protein function and stability, and therefore especially random mutagenesis libraries contain high amounts of nonfunctional proteins (Neylon, 2004).

2.4.1 Random mutagenesis

Random mutagenesis creates mutations at random positions along the gene, and it can be carried out using either *in vivo* or *in vitro* strategies. It is an especially useful approach when functionally significant positions of the protein are not known. The easiest option is to use bacterial mutator strains (e.g. XL1-Red) that have defects in one or several DNA repair pathways leading to a higher mutation rate (for a review, see Muteeb and Sen, 2010). However, due to relatively low mutation rates and the inability to target just the gene of interest for the mutagenesis, *in vitro* mutagenesis strategies are strongly preferred. Thus, one of the most popular approaches for generating libraries for directed evolution experiments is to use error-prone PCR. It generates point mutations during PCR amplification in the gene-of-interest (GOI) by utilizing low fidelity of DNA polymerase under certain conditions (e.g. with varying $MnCl_2/MgCl_2$ concentrations, biased dNTP concentrations, the usage of mutagenic dNTP analogues, and increased concentrations of Taq DNA polymerase). Additionally, the average number of mutations per clone can be increased by simply increasing the number of PCR cycles, since mutations accumulate with each cycle of PCR amplification (Cadwell and Joyce, 1992).

2.4.2 Site-directed random mutagenesis

Site-directed methods offer a very powerful route to randomize specific chosen residue positions and regions within the protein. This method requires the availability of structural information, either as a solved X-ray structure or a good homology model of the scaffold-of-choice. Once the appropriate sites or contiguous regions have been selected, mutations are introduced using e.g. synthetic oligonucleotides containing degenerate codons (Banta *et al.*, 2013).

The important thing to consider especially with a site-directed random (also known as saturation) mutagenesis library, is the aimed library size. When all 20 amino acids are allowed at each of the randomized positions and n denotes the number of randomized positions, the size of the protein sequence space is theoretically 20^n . However, if degenerate codons are used, there are usually 32–64 codons encoding the 20 amino acids (Table 2 and 3), and thus 32–64ⁿ genes is required to encode 20ⁿ proteins. This is due to codon bias: degenerate codons cause the obligatory use of redundant codons in addition to those required for encoding the 20 amino acids, and thus amino acids are represented unevenly (Hughes *et al.*, 2003). As the sequence space grows very rapidly with n , randomization efficiency is progressively lost, and

Table 2. Degenerate base abbreviations

Degenerate base designation	Actual bases coded	Degenerate base designation	Actual bases coded
N	A, G, C, T	M	A, C
V	A, C, G	R	A, G
H	A, C, T	W	A, T
D	A, G, T	S	C, G
B	C, G, T	Y	C, T
		K	G, T

Table 3. Properties of different degenerate codons

Degenerate codon	N:o of codons	N:o of amino acids	N:o of stop codons
NNN	$4^3=64$	20 (ranging from 1/64 to 6/64)	all stop codons (3/64)
NNB	$4*4*3=48$	20 (ranging from 1/48 to 5/48)	1 stop (1/48)
NNK(S)	$4*4*2=32$	20 (ranging from 1/32 to 3/32)	1 stop (1/32)
NNY	$4*4*2=32$	15 (no Trp, Gln, Glu, Met, Lys)	none
“MAX” (Hughes <i>et al.</i> , 2003), mixture of standard degenerate oligonucleotides (Tang <i>et al.</i> , 2012; Kille <i>et al.</i> , 2013) or usage of synthetic trinucleotides	20	20	none

NNN and NNK(S) are the most often used degenerate codons. NNY was used in (III). MAX codons were introduced in (Hughes *et al.*, 2003) enabling only one codon for each amino acid.

gene library sizes rapidly exceed cloning capability. Thus, libraries often tend to be orders of magnitude smaller than is required for full amino acid coverage (Hughes *et al.*, 2003). In order to cover the sequence space with a manageable sized library, saturation experiments usually involve only $n \leq 6$ randomized positions.

In an ideal situation for randomization, there would be a predefined distribution over the 20 amino acids with zero probability for a stop codon. This can be achieved by using the special MAX oligonucleotides (Hughes *et al.*, 2003), or through a proper mixture of several standard degenerate oligonucleotides (Tang *et al.*, 2012; Kille *et al.*, 2013). Yet another possibility, although more expensive, is to order the gene of interest synthetically. All of these methods enable the possibility to define the used codons in the specific amino acid positions.

Information from several different sources such as mathematical, computational and evolutionary models can be combined to generate (semi)rational libraries: The selection of a smaller, yet chemically balanced subset of amino acids at each of the randomized positions (instead of all 20) significantly reduces the sequence space. These semi-rational libraries are also easier to explore, and the usage of rationally designed alphabets also improves selection efficiency (Nov, 2014). One possibility is to select residues that represent each type of amino acid (polar, charged, hydrophobic, etc.) by using the NDT-codon (12 codons/12 aa), which reduces the complexity of libraries (Reetz *et al.*, 2008). Furthermore, various computational software have been developed in order to ease the library construction design process, for example GLUE-IT (<http://guinevere.otago.ac.nz/cgi-bin/aef/glue-IT.pl>) (Firth and Patrick, 2008), CASTER (http://www.kofo.mpg.de/media/2/D1108347/0987095526/ISM_tools.zip) (Reetz and Carballeira, 2007), and TopLib (<http://stat.haifa.ac.il/~yuval/toplib/>) (Nov, 2012). For a comprehensive review, see Nov, 2014.

Different studies have revealed that there are general rules about the shape and amino acid composition in the antigen-binding site that can be used as a guide to construct antibody phage display libraries through biased random mutagenesis (Collis *et al.*, 2003). Fellouse *et al.* have utilized restricted randomization and constructed a series of Fab-libraries using either binary code, which restricts randomization only to Tyr and Ser found to be enriched in the CDR-regions of antibodies (Fellouse *et al.*, 2005), or using degenerate codons encoding only four amino acid residues (Fellouse *et al.*, 2004; Fellouse *et al.*, 2006). Despite the extreme restriction, high affinity binders were selected in both cases that are comparable with binders selected from naïve libraries.

2.4.3 Recombination techniques

Homologous recombination employs mutations already shown to be functional in nature, found in homologous parent genes. In *in vitro* DNA recombination, novel DNA sequences are formed as fragments from two or more homologous parent genes, and are randomly assembled into chimeric genes. These sequence homology-dependent methods depend on DNA sequence identity for generating diversity, and thus, crossover positions are biased to be created between genes at loci sharing the highest homology and cannot form between regions with low homology.

Furthermore, when sequences have less than 70% sequence identity, there is a severe bias toward parental recombination (Lutz *et al.*, 2001).

DNA shuffling is a widely used directed evolution approach generating diversity through homologous recombination, combining useful mutations from individual related genes. In this method developed by Stemmer in 1994, chimeric gene libraries are generated by random fragmentation from a pool of related genes (by DNase I), followed by reassembly of the fragments with a self-priming polymerase chain reaction. Crossovers are created by template switching in the areas of sequence homology (Stemmer, 1994). As a source of diversity, one can either use naturally occurring homologous genes or mutant genes created previously, e.g. by random mutagenesis, to combine selected point mutations in novel combinations (Cramer *et al.*, 1998; Neylon, 2004). DNA shuffling enables many parent gene sequences to be recombined simultaneously, and thus generates multiple crossovers per reassembled sequence. However, the annealing-based reassembly limits the recombination process by aggregating the crossovers in regions of high sequence identity (Moore *et al.*, 2001). DNA shuffling was the first recombination method described, and it is still one of the most commonly used DNA recombination protocols.

Staggered extension process (StEP) is another *in vitro* DNA recombination approach (Zhao *et al.*, 1998), where the template sequences are primed followed by repeated cycles of denaturation and extremely short annealing/polymerase-catalyzed extension. This enables the growing fragments in each cycle to anneal to different templates based on their sequence complementarity and extend further to create recombination cassettes. StEP is continued until full-length genes are formed, possibly followed by a gene amplification step, if desired (Zhao *et al.*, 1998).

Random mutagenesis on transient templates (RACHITT) is an approach developed by Coco *et al.* (2001). The method differs from others by relying on single-stranded (rather than double-stranded) fragments that are allowed to hybridize onto a full-length single-stranded homologous gene. This template strand is synthesized to incorporate uracil, enabling its subsequent degradation. Un-hybridized 5' and 3' termini of the fragments are trimmed using nucleases, gaps are filled and fragments are ligated. Finally, the template strand is digested and the chimeric strand is made double-stranded through PCR (Coco *et al.*, 2001). In practice, RACHITT is more challenging compared to the other gene shuffling strategies, however it has been shown to seriously improve genetic diversity.

Incremental truncation for the creation of hybrid enzymes (ITCHY) is a method to create combinatorial fusion libraries between genes independent of DNA

homology (Ostermeier *et al.*, 1999). Incremental truncation is based on exonuclease III, which creates a library of all possible single base pair deletions of a given piece of DNA. A few years later, Sieber *et al.* (2001) introduced their method for sequence homology-independent protein recombination (SHIPREC) that can similarly create libraries of single-crossover hybrids of unrelated or distantly related proteins (Sieber *et al.*, 2001). The main drawback of these methods is that the members of these libraries can contain only one crossover site per gene.

By combining different recombination methods, one can create more diverse libraries. Since an ITCHY library has all theoretically possible crossover points, DNA shuffled ITCHY libraries, called SCRATCHY libraries, are more diverse compared to the traditional DNA shuffling libraries where crossover points are limited to precise regions of DNA identity. In practice, a SCRATCHY library is created by constructing first two ITCHY libraries: One with gene A on the N-terminus, and another with gene B on the N-terminus. Then DNA fragments of A-B and B-A fusions that are approximately the same size as the original genes are isolated, amplified and treated as in the normal DNA shuffling reaction (Kawarasaki *et al.*, 2003).

2.5 Mutant library selection methods

Once the DNA library is constructed, the translated protein variants are either screened or selected to isolate mutants with the desired trait. There are many different methods available. However, each new protein function requires an individually designed selection or screening method (Bratkovic, 2010). In library screening individual variants of a library are tested and their function is compared to the wild-type function. This is often a tedious approach, as it involves preferentially all but at least a selected part of the variants to be screened. However, it is usually necessary in order to find for example desired enzymatic functions. In contrast, searching for the fittest variant based on its affinity towards a ligand or another protein is an easier task, at least in theory, since the whole library can be subjected to selection. Selection links the desired phenotype with the clone genotype and its survival in the assay. The protein and the DNA, cDNA or mRNA that encodes it can be linked indirectly in phage and cell-surface display, or directly as in the case of mRNA or ribosome display (Lin and Cornish, 2002). This section will focus on the most commonly used selection method, phage display, with only a brief description of the other selection methods.

2.5.1 Phage Display

(Bacterio)phages are a diverse group of prokaryotic cells (bacteria and archae) capable of infecting viruses. Phage display, established by Smith (1985), is a versatile and robust technique in which (poly)peptide variants are displayed on the phage surface (Smith, 1985). The method involves the gene of interest to be inserted into the phage genome and its expression as a fusion with one of the coat proteins. Thus there is a physical link in the form of the phage particle, between each protein variant and its cognate DNA sequence. The principle of phage display is simple: The desired target-binding phages are selected from a phage library consisting of phages displaying a wide diversity of (poly)peptides on their surface. The technique has been reviewed extensively (for example, see Pande *et al.*, 2010; Bratkovic, 2010).

Most phage display work has been carried out utilizing filamentous Ff-phage strains, of which M13 is the most commonly used. The affinity selection process of the phage particles, called “(bio)panning”, is based on the specific binding properties of the fusion proteins they are displaying. First, the phage-display library is exposed to the target-of-interest to allow target-specific proteins on the surface of the phages to bind their targets. Phages not binding to the surface will be washed away, and then the bound ones can be eluted e.g. with low pH, high salt, or a solution containing the ligand of interest. These selected phages can be amplified simply by infecting bacteria and subjected to further rounds of panning, typically with more stringent selection conditions. Ideally, already one round of selection would be enough to select the best binders from the library, but in practice, nonspecific binding limits the achievable enrichment and thus selection and amplification needs to be repeated several rounds. The obtained selected clones can then be analyzed by sequencing and various *in vitro* binding assays (Hoogenboom, 2005).

Two parameters are manipulated during the panning process: Yield (the fraction of the phage particles with a given fitness surviving selection) and stringency (the degree to which proteins with higher fitness are favored over the others) (Smith and Petrenko, 1997). The condition of the first selection round should be manipulated towards lower stringency and then gradually raise conditions towards a higher stringency with ongoing selection rounds. Otherwise, the most promising candidates are easily lost during the first panning round. In practice, this is accomplished by

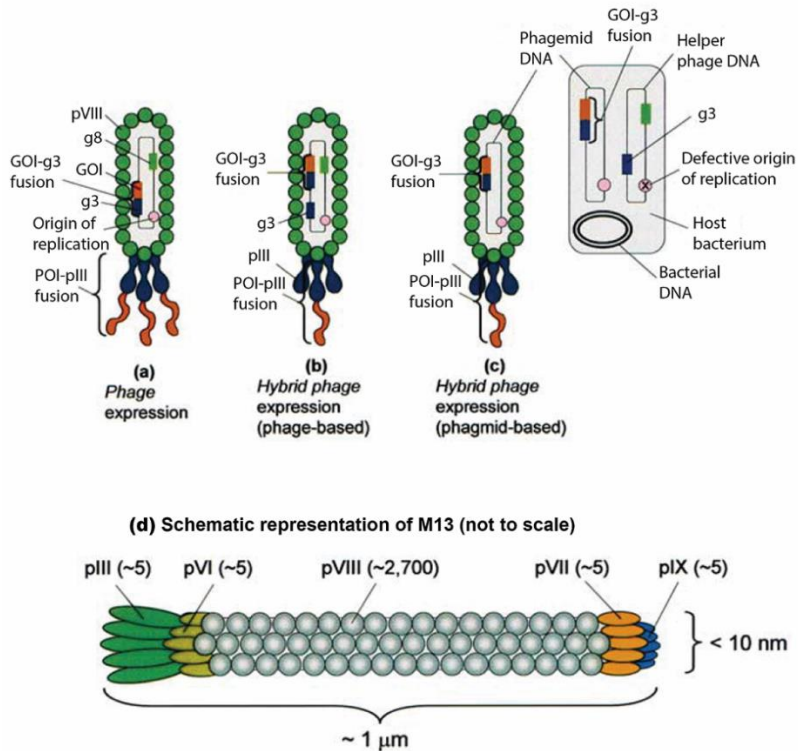


Figure 6. Bacteriophage M13-based phage display. a-c) Phage display constructs according to Smith's classification (Smith and Petrenko, 1997): a) Type 3 display, where the GOI is fused to phage gIII resulting in phage particles in which all the coat surface pIII-proteins are fusion proteins. b) Type 33 display, where gIII-fusion is incorporated as an additional element in the phage genome resulting in hybrid phages encoding both wt and fusion versions of the pIII. c) A type 3+3 system is encoded by two separate genomes: The wild type is carried by a phage called the helper phage, while the recombinant version is encoded by a special kind of plasmid called a phagemid. Like other plasmids, phagemid has a plasmid replication origin for replication in *E. coli* and an antibiotic resistance gene, but also a filamentous phage replication origin. The last mentioned characteristic is inactive until the cell is infected with the helper phage. The phage origin on the helper phage DNA has been modified to favor packaging of phagemid DNA instead of helper phage DNA. Thus the progeny virions are mosaics as their coats are composed of a mixture of recombinant and wild type pIII carrying phagemid DNA inside. d) Filamentous Ff-phages including M13 are naturally flexible rods with dimensions $\sim 1 \mu\text{m} \times 6 \text{ nm}$, mainly composed of a tube of helically arranged molecules of the major coat protein pVIII. The M13 phage particles are the most commonly used in phage display consisting of a single-stranded (+) DNA molecule as the phage genome surrounded by thousands of copies of the phage coat proteins. These coat proteins include ~ 2700 copies of the major coat protein pVIII, about five copies each of the minor coat proteins pIII and pVI at one end of the phage particle, and similarly five copies each of minor coat proteins pVII and pIX at the opposite end of the phage particle. Figure is modified and reprinted from Willats, 2002, copyright 2002, with permission from Springer.

manipulating 1) the target concentration, 2) the duration of the library with target, 3) the washing conditions (e.g. duration, detergent concentration, or pH value), and/or 4) the elution conditions. Ideally, specific elution conditions with high affinity and highly specific ligands should be used in order to guarantee the selection of specific binders. Usually such eluents are not available and thus non-specific elution conditions have to be used (Bratkovic, 2010).

There are monovalent and polyvalent systems available that can efficiently display short peptides or large proteins. Generally, monovalent systems are phagemid vector based (Mead and Kemper, 1988; Cesareni, 1988; Bass *et al.*, 1990) display systems, whereas polyvalent systems are phage vector based (Smith, 1985) (Figure 6). The phage vector system is based on the entire phage genome where GOI is inserted as a fusion to one of the phage surface proteins (in most cases gene III is used), thus essentially every copy of the particular coat protein contains GOI as a fusion component (unless a single phage vector carries two copies of the surface protein, recombinant fusion, and wild type genes) (type 33 display in Figure 6). The phagemid system is based on a phagemid-plasmid carrying the phage surface protein fusion gene. Furthermore, the so-called helper phage, providing all the other proteins, is needed to produce functional phages. Helper phages are modified Ff phages, which contain antibiotic resistance genes, an additional origin of replication, and a severely disabled packaging signal. Phagemid vectors provide a major advantage through their smaller size and ease of cloning, allowing for large library sizes. The display level of GOIs has been the main difference between these systems. However, the valency of display can be controlled by the use of special helper phages enabling polyvalent display also from the phagemid vector (e.g. Hyperphage (Rondot *et al.*, 2001), Ex-phage (Baek *et al.*, 2002), and Phaberge (Soltes *et al.*, 2003)). The use of these helper phages on the other hand requires displaying the GOI with a complete pIII.

The transformation efficiency of the bacterial host forms the limiting factor for phage display library sizes. Most phage display libraries have between 10^7 – 10^9 clones, although libraries of $>10^{10}$ have been achieved (Simeon & Chen 2017). Thus, many repeated electroporations (Dower *et al.*, 1988) are necessary to create large libraries. Typical phage display starts with the construction of a phagemid DNA library *in vitro*, followed by its transformation into competent *E. coli* cells (Bratkovic, 2010). One of the reasons for its popularity is the robustness of the system. The high stability of the phage enables an extensive range of conditions to be used in order to drive the selective pressure in the desired direction. Thus even for example high temperature or denaturant concentrations can be used to select for high stability proteins. The technology is well suited to approaches involving a common protein fold. However,

the method with non-lytic filamentous phages is limited for protein and peptides that can be efficiently expressed in the periplasm of *E. coli*. To circumvent this limitation, lytic phage strains T7 (Rosenberg *et al.*, 1996), T4 (Efimov *et al.*, 1995; Ren and Black, 1998), and lambda-based (Maruyama *et al.*, 1994; Sternberg and Hoess, 1995) systems have been developed.

2.5.2 Alternative selection systems

Cell surface display: The principle of cell surface display (Ståhl and Uhlén, 1997; Boder and Wittrup, 1997; Ernst *et al.*, 1998) resembles phage display: A library of protein variant encoding genes is fused to such a membrane protein, which is able to present the proteins on the cell surface. Cell surface display is compatible with fluorescence-activated cell sorting (FACS) techniques that are normally used for high-throughput screening. As in the case of phage display, the transformation efficiency of the host cell sets the upper limits for the obtainable library sizes. In the case of yeast and other eukaryotic cells, maximal library sizes are smaller than in the case of bacterial cell surface display systems (Lin and Cornish, 2002).

Ribosome display and mRNA display: Ribosome and mRNA display take place completely *in vitro* and thus they are not limited by transformation efficiency, enabling larger libraries ($>10^{12}$) compared to phage or cell surface display methods. This enables library construction to be faster as well. Additionally, the whole selection process composed of several cycles can be performed efficiently in a short time, since amplification by PCR can be coupled with PCR-mediated recombination or random mutagenesis (Jijakli *et al.*, 2016). Phenotype and genotype are linked by either a non-covalent (ribosome display) or covalent (mRNA display) complex which forms between the mRNA and the translated peptide. Both of these display formats have been successfully used to identify low-pM affinity binders, as quickly as after only a single selection round (Schilling *et al.*, 2014). The major limitation of these approaches is an intrinsically unstable mRNA molecule that necessitates the selection to be carried out under strict RNase-free conditions or at low temperature (Simeon and Chen, 2017).

The ribosome display introduced (Mattheakis *et al.*, 1994; Hanes and Plückthun, 1997) is based on an *E. coli* S30 coupled transcription/translation system, in which GOI lacking stop codons leads to a situation where both the translated peptide and mRNA remain associated with the ribosome complex. However, the GOI needs to be followed by a spacer sequence, which provides the (poly)peptide with enough

distance to fold properly. The ribosome complexes can be affinity selected based on the binding properties of the translated (poly)peptides towards the target of interest. The cognate mRNA can be used for PCR amplification once it has been dissociated from the complex and reverse transcribed to cDNA (Hanes and Plückthun, 1997). Even in the case of very high-affinity binding of the ribosomal complex to the desired target, the dissociation of the ribosomal complex can be achieved with mild nonspecific elution by simply adding EDTA.

mRNA display differs from ribosome display by introducing puromycin, an antibiotic mimicking the aminoacyl end of tRNA, thus forming a covalent linkage between the mRNA and translated peptide (Roberts and Szostak, 1997). Puromycin is attached to the 3' end of stop codon-lacking mRNA followed by a DNA spacer. When the mRNA has been translated *in vitro*, the ribosome pauses at the DNA spacer and allows a puromycin molecule at the end of the mRNA to enter the ribosome A site and react with the polypeptide chain. In the mRNA display experiment, the mRNA library is first ligated to a chemically synthesized 30-bp DNA sequence with a puromycin molecule at the 3' end. During *in vitro* translation, covalent complexes are formed between the mRNA and peptides via puromycin. In contrast to the ribosome display, these covalent complexes are then isolated and reverse transcribed before selected against an immobilized target. These selected complexes can then be PCR amplified for sequencing, or further rounds of mutation and selection (Roberts and Szostak, 1997).

2.6 Modified affinity proteins

Natural antibodies with their excellent molecular recognition properties have been long utilized in various areas of biological research, in biotechnological and chemical use, as well as in bioanalytics, medical diagnostics and therapy. The revelation of the molecular mechanism behind the natural evolution of antibodies have tremendously helped the process for directed evolution of antibodies. And later the same principles were used to create other, so-called alternative or protein scaffolds, not based on an immunoglobulin scaffold. The design of alternative scaffold molecules have been directed specifically to be better suited to applications in which the versatile antibodies offer only a limited number of options (see section 2.6.2) (Skerra, 2003).

Figure 7 shows various proteins used as scaffolds for the generation of artificial binding proteins. They include antibody derivatives (see section 2.6.1), proteins imitating antibody structure with rigid scaffold presenting multiple loops for

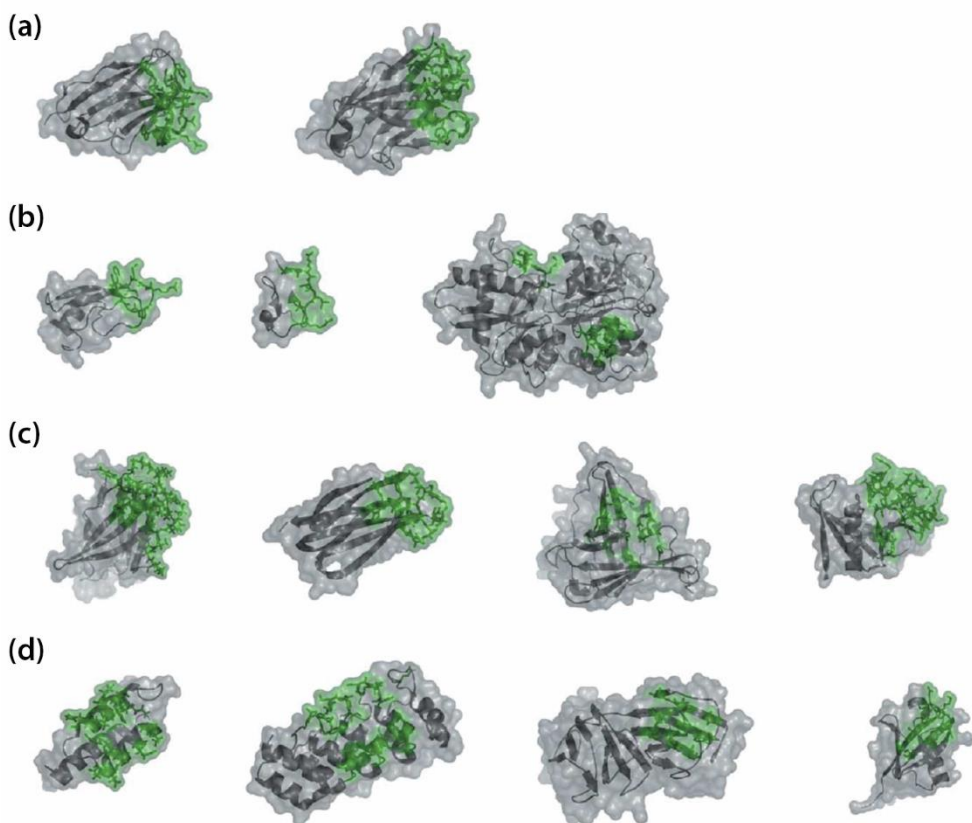


Figure 7. Different protein scaffolds used for the generation of artificial binding proteins. In the structures, randomized surface regions are highlighted in green containing the amino acid side-chain positions that have been modified or where additional residues have been inserted. The secondary structure of the conserved framework and the conserved surface of the protein are displayed in grey. (a) Antibody derivatives: Single-domain antibodies of human origin (chain A from 1OHQ) and of camelidea (chain E from PDB entry 1KXQ). (b) Rigid protein frameworks for the insertion or randomization of single loop peptides: the Kunitz domain (chain A from 1AAP), Squash-type protease inhibitor (chain I from 1H9H, residues 1 to 28), and human serum transferrin (1A8E). (c) Proteins resembling antibodies by presenting multiple loops for randomization on rigid frameworks: Human cytotoxic T lymphocyte-associated antigen 4 (1AH1), tenth fibronectin type III domain (1FNA), lipocalin (chain A from 1BBP), and C-type lectin-like domain (1TN3). (d) Proteins providing secondary structure elements for randomization: the Z-domain of Protein A (2SPZ), ankyrin repeat protein (1MJ0), γ -crystallin (1AMM), and ubiquitin (1UBQ). Figure adapted and reprinted from Hey *et al.*, 2005, copyright 2005, with permission from Elsevier.

Table 4. Properties of alternative scaffolds.

Name	Affibody	Anticalin	Antidin	DARPin (Designed Ankyrin Repeat Protein)	Monobody / Adnectin
Scaffold	three-helix bundle	Lipocalin β -barrel	avidin β -barrel	Ankyrin repeat (AR) protein consisting of varying number of ankyrin repeat motifs	Fibronectin type III domain (Fn3) β -sandwich
Original source	Z-domain of Staphylococcal protein A (SPA)	apolipoprotein D, bilin-binding protein, FABP (fatty acid - binding protein)	avidin from chicken	derived from naturally occurring AR- proteins	human 10 Fn3
Secondary structure	3 helix bundle	β -barrel	β -barrel	Ankyrin repeat motif	β -sandwich
Size	6.5 kDa	~20 kDa	~60 kDa	~14–18 kDa	~10 kDa
Oligomericity	monomer	monomer	tetramer	monomer (dimer and trimer available)	monomer
Example ligands	Taq DNA polymerase, human insulin, human apolipoprotein A1, Her2 and Her3	fluorescein, digoxigenin, CTLA-4, ED-B, the MET extracellular domain, A β - peptide and estradiol	progesterone, testosterone	MBP, JNK2, p38, Her2 and EpCAM	ubiquitin, TNF- α , MBP, lysozyme and IgG
Reference(s)	review Löfblom <i>et al.</i> , 2010	review Richter <i>et al.</i> , 2014	Riihimäki <i>et al.</i> , 2011b; Lehtonen <i>et al.</i> , 2016	review Plückthun <i>et al.</i> , 2015	review Lipovchek <i>et al.</i> , 2011

randomization (Fn3, lipocalins, avidin), as well as proteins providing secondary structure elements for randomization (SPA Z-domain, Ankyrin repeat proteins). Table 4 combines the properties of the best-established forms of alternative scaffolds (Jost and Plückthun, 2014) compared to traditional antibodies and avidin-structure based antidiins, the topic of this PhD thesis.

2.6.1 Engineering of antibodies and antibody fragments

Antibodies, also known as immunoglobulins, are biological recognition elements that form the core of the vertebrate immune system. An immunoglobulin- γ (IgG) is the most abundant antibody circulating in the blood of mammals, produced mainly by plasma cells. It is a large-sized (~150kDa) protein with a Y-shaped structure, which is used by the immune system to identify and neutralize pathogens such as bacteria and viruses. The glycosylated protein is composed of four polypeptide

chains (two heavy and two light chains), and it has multiple disulphide bonds. Each antigen-binding site is composed of a heavy and a light chain variable domain (VH and VL, respectively). More precisely, six hypervariable loop regions also known as complementarity determining regions, CDRs, form the antigen-binding site. These CDR segments vary among immunoglobulins with different specificities both in amino acid sequence and length, and therefore the overall Ig fold needs to be stable enough to support them. Immunoglobulin forms thus a bipartite protein structure, with a conserved scaffold supporting a structurally variable active site. The diversity of CDRs forms the basis for the unique capability of the immune system to efficiently create adaptor proteins against almost every given target (Skerra, 2003).

In many applications where antibodies have been utilized for their binding ability, only the functional antigen-binding region of an antibody (Fab or Fv) is needed, and thus systems based on antibody fragments have been created (Figure 8). Furthermore, these fragments enable the use of phage display for selection, since whole antibodies cannot be expressed in *E. coli*. However, despite the smaller size, these functional fragments are still composed of two weakly associated variable domains containing six hypervariable loops, which are demanding to manipulate simultaneously. In single-chain Fv (scFv), VH and VL are linked through a flexible spacer (Hosse *et al.*, 2006).

The idea of utilizing single VH domains as a scaffold was first put into practice by Ward *et al.* already in 1989 (Ward *et al.*, 1989), and they introduced the term single domain antibodies (abbreviated as dAbs or sdAbs). These sdAbs (referring to either the VH- or VL-chain) are the smallest antigen-binding fragments of IgG. However, this unpaired Ig domain has a tendency to aggregate since it has to expose a significant area of its hydrophobic surface (normally shielded by association with the partner domain) to solvent (Ward *et al.*, 1989). A few years later in 1993, Hamers-Casterman *et al.* published an article describing the discovery of an interesting class of antibodies from camels without light chains and thus their VH domains are soluble (Figure 8) (Hamers-Casterman *et al.*, 1993). Later this natural subclass of IgGs consisting of a pair of heavy Ig chains were found in other camelids like llamas (Muyldermans and Lauwereys, 1999). Additionally, cartilaginous fish (e.g. sharks) contain H-chain only antibodies (Greenberg *et al.*, 1995), named IgNAR (New Antigen Receptor) and its variable domain as V-NAR (Roux *et al.*, 1998). These VH domains have increased surface hydrophilicity and a much longer CDR-H3 compared to human IgGs (Roux *et al.*, 1998).

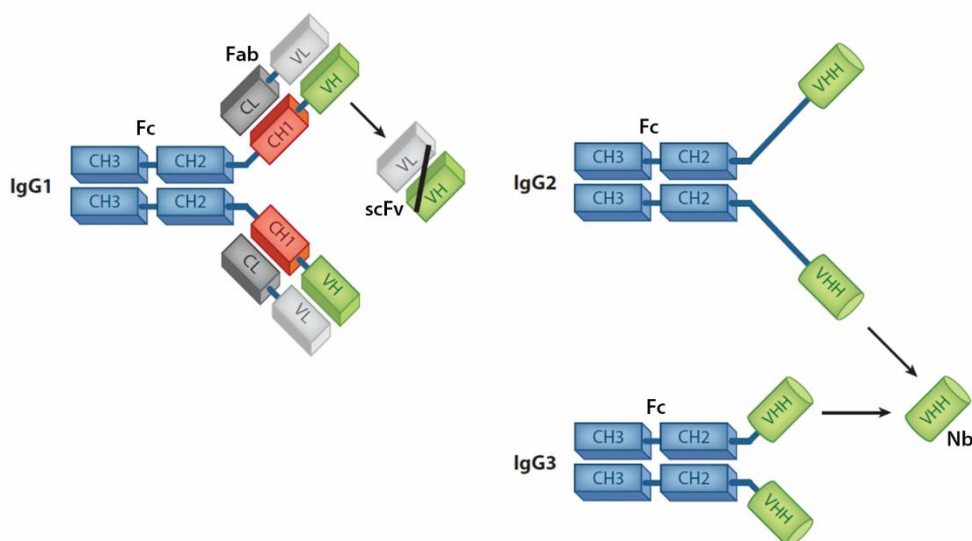


Figure 8. Schematic figure representing naturally occurring antibodies in camelid sera and the different antibody derivatives that can be constructed based on those. IgG1 represents the conventional Y-shaped antibodies containing two light (L) chains (composed of constant and variable domains, CL and VL) and two heavy (H) chains (composed of CH1-3 and VH). Additionally the sera of camelids contain also two types of homodimeric heavy-chain antibodies (HCAbs), IgG2 and IgG3, containing only heavy chains. The figure indicates the difference between various antibody derivatives. Fab-fragment consists of one constant and one variable domain of each of the heavy and light chain joined by a disulphide bond. The smallest possible functional antigen-binding fragment, single-chain Fv (scFv), can be generated from conventional antibodies and consists of a VH-VL pair linked by an oligopeptide. Likewise, the smallest functional antigen-binding fragment of HCAbs is called a single-chain VHH, also known as a Nanobody (Nb) and as a single-domain antibody (sdAb). Figure modified and reprinted with permissions from Muyldermans, 2013 (copyright 2013, Annual Reviews).

However, despite the drawbacks of antibodies in certain applications, the use of modified monoclonal antibodies is a growing field in next-generation antibody therapeutics, especially in the field of oncology and targeted drug delivery, and also in imaging. Different kinds of engineered antibodies have been created in numerous ways to meet the requirements of medicine and bioindustry (Evans and Syed, 2014; Jain *et al.*, 2007). Molecular engineering has had a huge effect on the clinical success of antibodies in recent years (Jain *et al.*, 2007). Antibodies have been modified to alter their specificity, affinity and valency, and alter their molecular size, but also to change their immunogenicity, effector functions, and pharmacokinetic properties (Jain *et al.*, 2007). Currently, antibody-drug conjugates designed for targeted drug delivery represent the most prominent and innovative antibodies under development

(Evans and Syed, 2014). Many engineered antibodies are in the market and clinical trials (Evans and Syed, 2014).

2.6.2 Alternative scaffolds to replace antibodies

Although natural antibodies have been widely utilized both in therapy, medical diagnostics and bioanalytics, as well as for biotechnological and chemical uses for years, they are not without limitations. As novel applications are emerging, the demands for affinity reagents are constantly growing. Conventional antibodies have several less desirable structural features for applications such as requiring harsher conditions or certain therapeutic strategies, like drug targeting (Skerra, 2003).

The most common problem associated with antibodies is their large, complex structure: They are 150-kDa multidomain glycoproteins containing disulphide bridges. Production in mammalian cell cultures is time-consuming and expensive, while only the antibody fragments can be expressed in prokaryotic cells. Furthermore, the necessity for expression in periplasmic space due to disulfide bridges leads to poor yield when compared to cytosolic expression. Even the smallest functional fragment, the Fv fragment, is composed of ~250 amino acids and is composed of two different protein chains. Additional efforts are required at a genetic level to clone and manipulate these genes in parallel. Moreover, in the case of the Fv fragment, the two variable domains are only weakly associated, and thus this also causes practical problems. Even the scFv fragments with a linker often have low stabilities and the extra linker segment may affect the antigen-binding capability. Additionally, it is demanding to manipulate altogether six hypervariable loops of the binding site simultaneously. Therefore, along with antibody fragments (discussed previously), other so-called protein or alternative scaffolds have been used in order to generate artificial receptor proteins with desired ligand specificities (reviewed by Skerra, 2000; Skerra, 2003).

Logically, first only scaffolds resembling immunoglobulins with their exposed loops were modified. Later the demands for ideal protein structure have loosened a bit, and the general guidelines for an alternative scaffold is that it should be able to improve the majority of the limitations associated with Ig domains without compromising affinity or specificity (Binz *et al.*, 2005). Ideally, these protein scaffolds should be small-sized and have a robust, monomeric structure with a low degree of post-translational modifications. Their genetic engineering should be simple, with an option to produce functional fusion proteins. Importantly, these protein scaffolds

should contain a region tolerating extensive modification in order to create novel binding sites. The region should be possible to be reshaped both through amino acid replacements and by insertions and deletions without compromising the protein folding capability (Skerra, 2003).

Protein scaffolds will be solely concentrated on here (therefore e.g. oligonucleotide and peptide molecules, Aptamers, are excluded) and introduce below a few different classes of engineered affinity proteins that have proven potential in therapeutic, diagnostic, and/or biotechnological applications. These can be divided into proteins, where ligand-binding amino acid residues are located in exposed loops (Anticalins and Monobodies/Adnectins), and those where they are located in the secondary structure (Affibodies and DARPins).

2.6.2.1 Affibodies

Affibody molecules are small, single domain proteins of only 6.5 kDa, which have proven to be capable of binding to any given protein target with high affinity and specificity. They consist of 58 amino acid residues devoid of cysteines folding into a three-helix bundle (reviewed by Löfblom *et al.*, 2010). The protein framework of an affibody is based on the staphylococcal protein A (SPA) Z-domain that can be efficiently expressed and secreted at high yield in *E. coli* (reviewed by Nygren, 2008; Skerra, 2000). The Z-domain is highly soluble, a proteolytically and thermally stable domain that naturally binds to the Fc-region of immunoglobulin via interactions with surface-exposed residues on the helices.

Affibody molecule libraries are constructed by combinatorial randomization of the residues involved in the natural binding interface. A total of 13 solvent-accessible surface residues of the Z-protein-domain scaffold were identified as significant for target binding based on X-ray crystallography analysis, and those were chosen for randomization to construct phagemid libraries (Nord *et al.*, 1995; Nord *et al.*, 1997). Other selection systems have been tested as well (reviewed by Nygren, 2008), but hitherto all *de novo* selections have been carried out using phage display. Affinity maturation has been achieved using either helix shuffling or sequence alignment combined with directed combinatorial mutagenesis (Gunneriusson *et al.*, 1999; Orlova *et al.*, 2006). Affibodies have been selected to bind several target proteins (reviewed by Nygren, 2008; Löfblom *et al.*, 2010; Ståhl *et al.*, 2017) including human insulin, apolipoprotein A1, Taq DNA polymerase, IgA, Alzheimer amyloid beta peptides, and the breast cancer markers ErbB2 (Her2) and ErbB3, (Her3) with K_d values ranging from micromolar to picomolar (Nord *et al.*, 1997; Gunneriusson *et al.*,

1999; Rönnmark *et al.*, 2002; Rönnmark *et al.*, 2003; Orlova *et al.*, 2006; Malm *et al.*, 2013). The affibody molecules with one of the highest affinities (~ 20 pM) have been achieved for affinity-matured variants directed to Her2 and Her3 (Orlova *et al.*, 2006; Eigenbrot *et al.*, 2010; Malm *et al.*, 2013). Since their first description, the suitability of affibodies has been investigated in many different and diverse applications, and they have been described as possibly the most successfully engineered protein scaffold for *in vivo* imaging. Recently, the first therapeutic affibody molecules entered clinical development (reviewed by Nygren, 2008; Banta *et al.*, 2013; Ståhl *et al.*, 2017).

2.6.2.2 DARPins (Designed Ankyrin Repeat Domain)

Ankyrin repeat proteins consist of tightly packed 33 residue long repeat motifs. The Ankyrin repeat motif is responsible for mediating a broad range of protein interactions and there are cytosolic, membrane-bound, and secreted forms. It is composed of a β -turn, two antiparallel α -helices and an unstructured loop, which reaches the β -turn of the next repeat. Usually there are four to six repeats, forming a structure with a large solvent-accessible surface and a continuous hydrophobic core with a groove-like binding surface (Li *et al.*, 2006; Bork, 1993), although in nature a single protein with up to 29 consecutive repeats have been found (Walker *et al.*, 2000). Functionally designed AR proteins (DARPins) have been selected from consensus combinatorial libraries, composed of varying numbers of repeat motifs flanked by special terminal repeats of AR domains, so-called capping repeats, which shield the hydrophobic core (Binz *et al.*, 2003; Forrer *et al.*, 2003). The molecular mass of a DARPin depends on the number of repeating units, 14–18 kDa being the average size (corresponding to 4–5 repeating units). These libraries have six diversified residues located in the β -turn and the first α -helix of the AR module forming possible interactions with the target. So far 27 framework residues have been shown to bind to a variety of targets (reviewed by Plückthun, 2015), including *E. coli* maltose-binding protein (MBP), eukaryotic protein kinases JNK2 and p38 (Binz *et al.*, 2004), Her2 (Zahnd *et al.*, 2006; Zahnd *et al.*, 2007), and tumor-associated antigen epithelial cell adhesion molecule (EpCAM) (Stefan *et al.*, 2011) with K_d values ranging from low nanomolar to low picomolar affinities. Most DARPins have been shown to possess high stability against heat or denaturants, and even in high protein concentrations. Proteins can be efficiently expressed in the cytoplasm of *E. coli*. Most display or selection methods are applicable with DARPin libraries, from which ribosome display is the most utilized (reviewed by Plückthun, 2015).

2.6.2.3 Monobodies / Adnectins

Monobodies or Adnectins are based on the fibronectin type III domain (Fn3), which is a small autonomous folding unit involved in specific molecular recognition or ligand binding. It is estimated to occur in up to ~2% of all animal proteins (Bork and Doolittle, 1992). The Fn3 (more precisely the 10th of 15 repeating units in human fibronectin, called ¹⁰Fn3, but referred to henceforth simply as Fn3) belongs to the IgG superfamily and naturally binds integrins. It is a small, monomeric β -sandwich protein of 94 amino acid residues without disulfide bonds. The protein structure resembles a trimmed Ig variable VH domain consisting of seven β -strands connected with three exposed loops (Koide *et al.*, 1998). Additionally, Fn3 is a highly thermostable protein, with a T_m value over 80 °C (Parker *et al.*, 2005).

Fn3 has a β -sandwich fold, which closely resembles immunoglobulin domains. Logically, because of the resemblance with immunoglobulins, the Fn3-based combinatorial libraries have been primarily constructed by randomization of the residues in three of the flexible loop structures on one side of the protein, but also through the side-and-loop surface by randomizing residues in a single loop and the face of the β -sheet (Koide *et al.*, 2012). Since its development as a molecular scaffold (Koide *et al.*, 1998), Fn3 has become one of the most widely used non-antibody scaffolds (Koide *et al.*, 2012). Different ligand binding Monobody/Adnectin-variants have been selected via mRNA, phage, and yeast surface display with K_d values ranging from micromolar to picomolar (reviewed by Lipovsek, 2011; Simeon and Chen, 2017). Targets include ubiquitin (Koide *et al.*, 1998), TNF- α (Xu *et al.*, 2002), maltose-binding protein (MBP) (Gilbreth *et al.*, 2008; Koide *et al.*, 2007), human and yeast small ubiquitin-like modifiers (hSUMO4 and ySUMO) (Koide *et al.*, 2007), lysozyme (Hackel *et al.*, 2008), and IgG (Hackel and Wittrup, 2010).

2.6.2.4 Anticalins

Anticalins are based on a lipocalin scaffold. Lipocalins constitute a structural family of small, functionally diverse, robust extracellular proteins, which are all comprised of a single polypeptide chain of 150–190 residues. The lipocalin (Lcn) family members act as storage or transport proteins for mainly hydrophobic and/or chemically sensitive organic compounds in many organisms. Twelve different secreted human lipocalins have been identified in the blood or tissue fluids. Despite the low sequence homology, lipocalins share a conserved 3D-fold dominated by eight antiparallel β -strands, which form a cup-shaped β -barrel core with a C-terminal

α -helix attached to its side. At the open end of the cup, the four structurally variable loops connecting the β -strands together with adjoining residues within the beta-barrel framework form the natural ligand-binding site (Skerra, 2008). Engineered lipocalins with novel target specificities (anticalins) are not glycosylated and can be even devoid of disulfide bonds, and are typically ~ 20 kDa in size. Anticalins are thermostable with $T_{ms} > 70$ C (Schlehuber and Skerra, 2002; Olwill *et al.*, 2013; Rauth *et al.*, 2016), with the proteins being easily produced in *E. coli* (Breustedt *et al.*, 2006).

Anticalins have been engineered typically by randomizing a range of 16–24 residues in the loop structures of the natural ligand-binding site by applying phage display or bacterial surface display techniques (Gebauer and Skerra, 2012). Four natural lipocalin scaffolds were selected for the construction of combinatorial libraries based on their known 3D-structures: the bilin-binding protein (BBP) from the butterfly *Pieris brassicae* (Beste *et al.*, 1999), the human apolipoprotein D (ApoD) (Vogt and Skerra, 2004), the human lipocalin 1 (Lcn1), also known as tear lipocalin (Tlc) (Breustedt *et al.*, 2005), and the human lipocalin 2 (Lcn2), also known as neutrophil gelatinase-associated lipocalin (NGAL) or siderocalin (Goetz *et al.*, 2002). The first anticalins were selected from the first lipocalin BBP library with high affinities for the dye fluorescein (Beste *et al.*, 1999) and the plant steroid digoxigenin with affinities in the low nanomolar range (Schlehuber *et al.*, 2000; Schlehuber and Skerra, 2002).

During the last years, anticalins have been versatily developed against various different target formats, including proteins, peptides, and small molecules such as T-cell coreceptor CTLA-4 (Schönfeld *et al.*, 2009), extra-domain B (ED-B) positive Fn (Gebauer *et al.*, 2013), MET extracellular domain (Olwill *et al.*, 2013), Alzheimer β -amyloid peptide (Rauth *et al.*, 2016), and estradiol (Liu *et al.*, 2012). The affinities (K_d) range from nano- to picomolar range. Monomeric human Lcn2 has shown to be a particularly promising scaffold for anticalin engineering, both for its excellent binding properties and stability (reviewed by Richter *et al.*, 2014).

3 AIMS OF THE STUDY

The general objective of the study was to broaden the usability of avidin as a reagent and develop new avidin-based artificial binding proteins to be used in applications requiring harsher conditions and in point-of-care (POC) tests, because of avidin's high stability. The specified objectives are as follows.

- 1) To construct an avidin phage display library for selection of avidin-scaffold based binders with new specificities from the library
- 2) To show the usability of avidin as an alternative scaffold to develop binding proteins for several different ligands
- 3) To construct a new phagemid vector for library construction enabling Gateway LR cloning
- 4) To develop a novel DNA shuffling library construction method

4 MATERIALS AND METHODS

All the materials and methods used during this study are described in detail in the original publications I–III with only a brief summary provided here.

4.1 Molecular cloning methods

4.1.1 Phagemid vectors (I–III)

The phagemid vector used to construct the avidin phage display libraries was based on the VTT Fab phagemid vector (pBluescript SK+ derived phagemid, VTT Technical Research Centre of Finland, Biotechnology, Espoo, Finland). It contains a *pelB* signal peptide preceding the Fab fragment followed by a C-terminal fraction of the minor coat protein pIII (aa 198–406), and it enables expression as a monovalent display mode (3+3). The phagemid has *f1*(+) and *pUC* replication origins, and the ampicillin resistance gene (*Amp^R*). The vector was digested with *NheI* and *NotI* (Thermo Fisher Scientific Inc) restriction enzymes to replace the Fab fragment with the avidin cDNA library (cut with the same restriction enzymes and ligated using T4 DNA Ligase (Thermo Fisher Scientific Inc) (I, III). For gel extractions, the GFX DNA purification kit (GE Healthcare, Uppsala, Sweden) was used.

The phagemid vector was modified to improve the cloning efficiency and to enable Gateway LR cloning (II). The phagemid vector was thus digested with *EcoRI* and *NotI* (Thermo Fisher Scientific, Inc.), to replace the *pelB* signal peptide and Fab fragment with the *attR*-flanked Gateway cassette (PCR amplified from a pTriEx-1.1 Gateway destination vector that was kindly provided by Professor Kari Airene, University of Eastern Finland, Kuopio, Finland). The Gateway cassette contains *attR* recombination sites flanking a *wdB* gene and the chloramphenicol resistance gene (*Cam^R*). The constructed pGWphagemid vector (II) enables construction of the avidin phage display libraries by Gateway LR cloning using the *wdB* suicide gene for negative selection (Katzen, 2007).

4.1.2 Bacterial expression vectors (I, III)

For efficient and tightly-regulated protein expression, selected avidin mutants were subcloned into the pET101/D expression vector (Invitrogen) under the T7lac promotor via directional TOPO® cloning, followed by heat shock transformation into *E. coli* TOP10 (Invitrogen) cells according to the manufacturer's instructions. At the same time, the ompA signal peptide needed for periplasmic expression was inserted in front of the avidin coding sequence by SES-PCR using 5' primers, which also contained the 5' CACC needed for the directional TOPO cloning reaction. The 3' primer added the 6xHis-tag needed for protein purification when expressing avidin mutants with a putative new function.

In study II, the recombinant proteins with the ompA signal peptide were also expressed directly from the pGWphagemid vector where the gene is under the lac promoter. The amber stop codon enables expression of free avidins as well (especially if expressed in another strain than the amber suppressor strain). The basal protein expression is repressed with glucose and induced by IPTG.

4.1.3 Targeted mutagenesis (III)

Targeted mutagenesis was used to add the point mutations I117Y and N118D for selected antidiins. QuikChange® mutagenesis (Agilent Technologies) was used to target mutations into antidiins in pET101/D-based expression vectors. The primers were designed and mutagenesis was performed according to the instructions of the manufacturer (Stratagene/Agilent Technologies).

4.1.4 Bacterial strains (I-III)

Several *Escherichia coli* strains were used for the purposes of the study: *E. coli* XL1-Blue (Stratagene) electrocompetent cells were used for phage display, because the strain contains the F' episome necessary for phage infection. Additionally, XL1-Blue contains the supE44 mutation, and thus the amber codon is not recognized as a stop codon in this strain – instead, it is translated into glutamine (Eggertsson & Söll, 1988). This enables the usage of the amber codon (II) between the avidin gene and C-terminal pIII. *E. coli* strains TOP10 (Invitrogen) and DH5α (Invitrogen) were used for plasmid propagations, except in the case of pGWphagemid (II) which was propagated in *E. coli* DB3.1 (Invitrogen) due to the *ccdB* suicide gene (Katzen, 2007).

The *ccdB* is a cytotoxic protein lethal to most *E. coli* strains. For recombinant protein expression, the *E. coli* BL21-AI (Invitrogen) strain was used. This strain enables high-level recombinant protein expression from the T7-based expression vector pET101/D (I, III). Furthermore, BL21-AI enables expression of chimeric proteins without the pIII-fusion, in case the amber codon resides in the junction (II).

4.1.5 DNA sequencing (I-III)

Colony-PCR was used prior to plasmid purification and subsequent DNA sequencing, when appropriate, to ensure the subcloning was successful. DreamTaq® (ThermoFisher Scientific), GoTaq® Green (Promega), or MyTaq® (Bioline) was utilized here according to the recommendations of the manufacturer. Bacterial cultures were prepared by inoculating with single colonies or with 10 µl of cell culture glycerol stocks from master plates, and cultured overnight in LB or SB medium supplemented with appropriate antibiotics and glucose (1%, w/v) to repress the basal protein expression from the vector. Minipreps were prepared according to the instructions of the manufacturer by using the GeneJET™ Plasmid Purification Kit, or by automated plasmid purification using the NucleoSpin Robot-96 Plasmid Kit (Macherey-Nagel) and pipetting robot (Genesis RSP 100, Tecan) with Gemini software (v. 3.40 SP1, Tecan).

Sanger sequencing was performed by PCR-amplifying the insert with appropriate primers and the amplified fragments were directly cycle-sequenced using the BigDye® Terminator v3.1 Cycle Sequencing kit (as described in detail in II) and visualized on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). The sequencing results were analyzed using various software, depending on the number of samples to be analyzed. The programs DNAMAN v. 4.11 (Lynnon Biosoft) or SnapGene (GSL Biotech LLC) were used for analyzing up to 16 samples. When the constructed library or enriched clones from the library were sequenced, the sequencing results were analyzed at DNA and protein levels using Perl Script (as described in detail in II). The resulting protein sequences were aligned using a ClustalW alignment algorithm (EBI web server) and further analyzed using GeneDoc (v. 2.7.0) (Nicholas *et al.*, 1997) and MEGA 5 (Tamura *et al.*, 2011) software.

4.2 Directed evolution of avidin by phage display

4.2.1 Construction of phagemid libraries (I-III)

In order to display avidin variants on the surface of the M13 phages, two expression constructs were used: 1) sole fusion without a stop codon in between the fusion partners (I, III), and 2) amber-version, where the amber codon was inserted between the Avd gene and C-terminal portion of the M13 surface protein pIII (II). The latter enables the assembly of a tetrameric protein on the surface of the phage, as the amber suppressor strain XL1-Blue is used (Sidhu *et al.*, 2000). Since suppression of the amber codon is not absolute (depends highly on the reading context (Eggertsson and Söll, 1988) and the growth phase), efficiency ranges between 65–93% (Singaravelan *et al.*, 2010), and both pIII-fusions and free proteins are expressed. Additionally, two different signal peptides, pelB (I, III) (from *Erwinia carotovora* pectate lyase B, (Lei *et al.*, 1987) and ompA (II) (from *Bordetella avium* outer-membrane protein A, (Gentry-Weeks *et al.*, 1992), were used, both utilizing the Sec-pathway (for a review, see Low *et al.*, 2013).

In order to create avidin variants, either directed random mutagenesis or DNA shuffling (described below) was used, followed by subcloning into a phagemid-vector (described in 4.1.1). The constructed DNA libraries were then electroporated into *E. coli* XL1-Blue (Stratagene) cells, and stored as glycerol stocks and minipreps (see section 4.2.1).

4.2.1.1 Directed random mutagenesis (I, III)

The libraries with randomization targeting loop regions of the avidin β -barrel (I, III) were constructed using a two-step PCR strategy, essentially as previously described in (Barbas 3rd *et al.*, 2001). Avidin cDNA was used as a template. Special primers were designed that contained NNN or NNY degenerate codons at the amino acid residues to be randomized (Table 3). Parallel to these special 3' primers, 5' primers were designed with overlapping DNA sequences enabling gel extracted and purified PCR fragments to be combined in the second amplification step. At the 5' and 3' end of avidin the desired restriction enzyme sites for NheI and NotI were added with another set of primers.

4.2.1.2 DNA shuffling (II)

To construct a DNA shuffled phage display library, the cDNAs of the genes *AVR2* (Hytönen *et al.*, 2005a) and *AVR4* (Hytönen *et al.*, 2004b) were used as starting material to construct the asymmetric templates *attL1-ompA-AVR4* and *AVR2-attL2*. DNA family shuffling was performed essentially as previously described (Stemmer, 1994; Ikeuchi *et al.*, 2003; Niederhauser *et al.*, 2012). Templates were digested with DNaseI, followed by the primerless reassembly reaction and Gateway LR cloning reaction (excisive reaction of bacteriophage lambda: $attL \times attR \rightarrow attB + attP$).

4.2.1.3 Storage and validation of the constructed libraries (I-III)

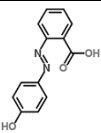
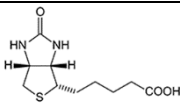
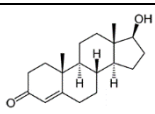
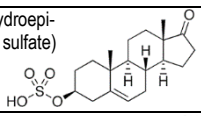
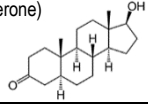
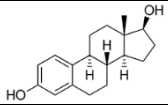
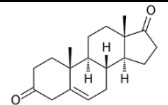
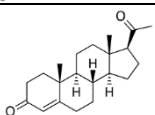
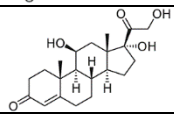
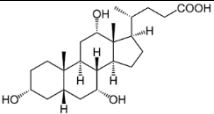
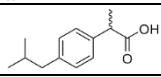
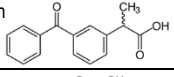
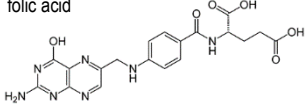
From the several parallel transformation reactions performed to construct a library, a portion was plated on LB_{amp} plates to be able to calculate the size and verify the quality of the constructed library (by DNA sequencing the clones, see 4.1.5). The remaining liquid from the transformation reactions of the same library were pooled together and diluted with SB medium containing appropriate antibiotics, cultured overnight and used to prepare glycerol stocks and minipreps of the constructed library.

In order to analyze the quality of the library, master plates of 96 wells (deep well plates, ABgene, Thermo Scientific, Surrey, UK) were prepared: Single colonies from the library transformants were cultured in SB freezing buffer supplemented with the appropriate antibiotics and glucose (1%, w/v) for 6-7 h. The glycerol concentration was then increased and the master plates were stored at -80 °C. Likewise, output plates from each panning round (II, III) were used to pick clones for master plates to be sequenced later.

4.2.1.4 Production of phages and PEG precipitation (I-III)

Glycerol stocks of the XL1-Blue library transformants were used to prepare 10 ml cultures that were infected with the helper phage (VCS-M13, $\sim 10^{11}$ pfu, Stratagene). The culture was then diluted to 100 ml (V_{tot}) and grown for 2h +37 °C shaker (III: in the presence of wt avidin (0.1 mg/ml)). Finally, kanamycin (70 µg/ml) was added and the temperature lowered to +28 °C overnight. The next morning, a sample of 3 mL was used for plasmid isolation (GeneJET Miniprep Kit) and the rest of the

Table 5. Target molecules used in biobanning.

target molecule	conjugated version (conjugation level)	MW	used for	publication
HABA (2-(4'-hydroxy- benzene)azo- benzoic acid) 	BSA (n/d)	242.23 g/mol	panning	I
D-biotin 	BSA (1:1), casein ^e	244.31 g/mol	panning, binding test, binding analyses	I–III
testosterone 	testosterone 3-(O- carboxymethyl)-oxime : BSA (30:1 and 7:1)/ : AP (5:1)/ : HSA (29:1)	288.42 g/mol	panning, binding test	I, III
DHEAS (dehydroepi- androsterone sulfate) 	-	368.49 g/mol	binding test	I
DHT (dihydrotestosterone) 	-	290.44 g/mol	binding test	I
estradiol 	BSA (n/d)	272,39 g/mol	binding test	I
androstenedione 	-	286.42 g/mol	binding test	I
progesterone 	progesterone 3-(O- carboxymethyl)oxime : AP (3:1) / : HSA (26:1)	314.46 g/mol	panning, binding analyses	I, III
hydrocortisone 	hydrocortisone 3-(O- carboxymethyl)oxime : AP (8:1) / : HSA (32:1)	362.46 g/mol	panning, binding analyses	III
cholic acid 	cholic acid : AP (n/a)/ : HSA (n/a)	408.57 g/mol	panning, binding analyses	III
(S)-(+)-ibuprofen 	(S)-(+)-ibuprofen : AP (3:1)/ : HSA (15:1)	206.29 g/mol	panning, binding analyses	III
(S)-(+)-ketoprofen 	(S)-(+)-ketoprofen : AP (4:1) / : HSA (28:1)	254.29 g/mol	panning, binding analyses	III
folic acid 	folic acid : AP (1:1) / : HSA (2:1)	441,14 g/mol	panning, binding analyses	III

n/a = no ionization in MALDI (III)

culture was used for double PEG precipitation (described in detail in Niederhauser *et al.*, 2012) in order to purify and concentrate the phages.

4.2.2 Selection by biopanning

4.2.2.1 Target molecules

The aim was to select avidin variants based on their binding properties, with the target molecules used in biopanning were: 2-(4'-hydroxy-benzene)azo-benzoic acid (HABA) (I), testosterone (I, III), biotin (II), progesterone, hydrocortisone, cholic acid, ibuprofen, ketoprofen, and folic acid (III) (Table 5). HABA was used in the first panning test (I) and biotin was used in order to select efficient biotin-binders within the DNA-shuffled AVR2/4 library (II). The other small target molecules were chosen based on their diagnostic relevance.

4.2.2.2 Biopanning method with a microplate (I)

Microwell plates (Nunc Immunosorb or Maxisorp™, Nunc, Roskilde, Denmark) were coated with a small ligand of interest conjugated to BSA (testosterone- or HABA-BSA and, as a negative control, BSA). Phages displaying avidin variants in BSA-PBS were allowed to bind to the immobilized target molecule and non-binders were washed away. Bound phages were eluted with 0.1M hydrochloric acid containing D-biotin (10 µM) or testosterone (10 µM or 35 µM), neutralized, and used to infect freshly grown XL1-Blue cells. The phages were amplified overnight with the help of VSC-M13 helper phages (Stratagene) and PEG-precipitated the following day as described above (see section 4.2.1). The biopanning step was repeated and altogether three to four panning rounds were performed, increasing the stringency of washing conditions every panning round in order to decrease nonspecific binding.

4.2.2.3 Biopanning with a magnetic particle processor (II–III)

In studies II and III, the panning was performed on the surface of coated magnetic particles operated by the magnetic particle processor Precipitor® (Abnova) (II) or KingFisher® (ThermoElectron) (III). Magnetic particles (Magnetic Bead Carboxyl

(Abnova, Taipei City, Taiwan) (II) or Dynabeads M-270 (Epoxy, Invitrogen) (III)) were coated with protein-conjugated small ligands according to the recommendations of the manufacturer. The conjugates used were biotinylated BSA (II) or AP-conjugated progesterone 3-(O-carboxymethyl)oxime, hydrocortisone 3-(O-carboxymethyl)oxime, testosterone 3-(O-carboxymethyl)-oxime, cholic acid, S-(+)-ibuprofen, S-(+)-ketoprofen, and folic acid (III) (Table 5).

For the first round of biopanning, the PEG-precipitated phages were combined with a small sample of the blocked, coated magnetic beads (Table 6). Instead of amplifying phages to be PEG precipitated, the phages were amplified in small scale cultures (1 ml) on 96-deep well plates (deep well plates, ABgene, Thermo Scientific, Surrey, UK) from which the supernatant was directly used for the next round of biopanning. The used protocol is modified from Turunen *et al.*, 2009. Altogether four to five rounds of biopanning were performed, increasing the stringency of the panning conditions (the amount of magnetic particles decreases, the incubation time shortens, and the number of washing steps increases) every round. The exact conditions of each of the panning rounds are listed in Table 6.

4.2.3 Validation of pannings

4.2.3.1 Analysing the integrity of the phagemid vector

In order to verify the integrity of the phagemid vector and the presence of a correct-sized insert, either digestion analysis (I, III) or PCR amplification (II) was performed for the created phage-isolated plasmids.

4.2.3.2 Dot blot (II)

Dot blot was used to analyse whether phages displaying biotin-binding variants were enriched during the panning rounds. A sample of phages from each panning round was applied to a nitrocellulose sheet. After blocking and washing, the blot was analyzed with BTN-AP (1:5000, Vector Laboratories, Burlingame, CA, USA) as a probe using BCIP/NBT substrate.

Table 6. Exact conditions in panning rounds performed with magnetic particle processors (II, III). Table modified and combined from II and III. Reprinted with permissions from Lehtonen *et al.*, 2016 and Lehtonen *et al.*, 2015, copyright 2016 American Chemical Society; and from Oxford University Press, copyright 2014.

Round	Phages	Input (cfu/ml)	Depletion	Beads	Incubation	PBST washes	Elution
1st	200 µl of PEG-precipitated phages + 200 µl PBST (0.05%)	2.1×10^{12} (II) / 2.1×10^{12} (III)	20 µg BSA (II) / AP ^a (III)	5 µl	4°C, shaker, O/N	3 x 1 min (II) / 5x 20 s (III)	20 min, 0.1M HCl
2nd	200 µl of supernatant from the 1 st round phage cultures ^b	1.4×10^{10} (II) / 1.8×10^{10} (III) ^c	20 µg BSA (II) / AP ^a (III)	4 µl	RT, magnetic particle processor, 4x 15 min	3 x 1 min (II) / 4x 20 s (III)	20 min, 0.1M HCl
3rd	200 µl of supernatant from the 2 nd round phage cultures ^b	2.3×10^{10} (II) / 1.3×10^{10} (III) ^c	20 µg BSA (II) / AP ^a (III)	3 µl	RT, magnetic particle processor, 4x 15 min	3 x 1 min (II) / 4x 20 s (III)	20 min, 0.1M HCl
4th	200 µl of supernatant from the 3 rd round phage cultures ^b	4.1×10^9 (II) / 1.8×10^{10} (III) ^c	20 µg BSA (II) / AP ^a (III)	2 µl	RT, magnetic particle processor, 2x 15 min	4 x 1 min (II) / 4 x 1 min (III)	20 min, 0.1M HCl
5th (III)	200 µl of supernatant from the 4 th round phage cultures ^b (III)	1.8×10^{10} (III) ^c	20 µg AP ^a (III)	1 µl (III)	RT, magnetic particle processor, 2x 7.5 min	4x 2 min (III)	20 min, 0.1M HCl

^a alkaline phosphatase (AP)

^b incubation was performed in the presence of Tween (0.05%)

^c calculated average titer between several parallel panning reactions

4.2.3.3 Phage-pool microplate assay

Phage-pool microplate assays were used to monitor the enrichment of phages expressing functional ligand-binding proteins on their surface. For that purpose, a sample of culture supernatant from phage amplifications after each round was directly used (as previously described by Turunen *et al.* 2009 (II, III)). For the ELISA, Nunc immunosorp/Maxisorp® plates were coated with protein (HSA or casein) - conjugated target molecule (progesterone, hydrocortisone, testosterone, cholic acid, ibuprofen, ketoprofen, folic acid, or biotin). Different carrier proteins were used than in the panning process. As negative controls, wells were coated similarly with carrier proteins (HSA or casein-containing milk). After blocking wells, phage culture supernatant was added. Finally, after washing off unbound phages, bound phages were detected using anti-M13/HRP (GE healthcare), which binds to the major coat protein pVIII, and can be detected using HRP-substrate. ABTS (2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) solution was used as a substrate, and the plates were read at A415 with a microplate reader (Bio-Rad 680 XR).

4.2.4 Screening (I-III)

Different microplate assays based screenings (described below) were used to analyze the individual clones obtained during the panning process. Based on the results obtained, clones were selected for plasmid purification and DNA sequencing as described in section 4.1.5.

4.2.4.1 Microplate assay for phages (I, III)

Phage ELISAs were used to study the enrichment of individual phage variants expressing functional ligand-binding proteins on their surface. For that purpose, phages were produced from individual clones (according to the protocol described by Kingsbury and Junghans, 1995 (I), or alternatively, essentially as previously described by Turunen *et al.*, 2009 (III)). For these assays, Nunc immunosorp/Maxisorp® plates were coated with protein (BSA or HSA) – conjugated to target molecule (progesterone, hydrocortisone, testosterone, cholic acid, ibuprofen, ketoprofen or folic acid). As negative controls, wells were coated

similarly with carrier proteins (BSA or HSA). In study III, different carrier proteins were used compared to the panning process. The bound phages were detected as described in the section 4.2.3.3 above.

4.2.4.2 Avidin-biotin displacement (ABD) assay (II)

Enrichment of functional biotin-binding clones (II) was studied also in the so-called avidin-biotin displacement (ABD) assay, where cell lysates from individual protein expression clones were studied on a microplate. The ABD-assay was used to get an estimate of the biotin dissociation rate constant. For the analysis of each clone, one well was coated with casein and two wells were coated with casein-SS-BTN. After blocking wells, clarified cell lysate was added to the wells. The unbound phages were washed off, and biotinylated alkaline phosphatase (BTN-AP) was used to detect the bound proteins. After washing, PBS alone or supplemented with biotin (10 µg/ml) was added to the wells and incubated for 30 minutes. The plates were washed again, and the bound BTN-AP was detected using phosphatase substrate (1 mg/ml, Sigma-Aldrich, St. Louis, MO, USA) in DEA buffer (1M diethanol amine, 0.5mM MgCl₂, pH 9.8) and read at A405 with microplate reader 680XR (Bio-Rad Laboratories Inc., Hercules, CA, USA). The signals from the wells treated with free biotin were compared with untreated wells in order to analyze the rate of biotin displacement (the ratio of the signals). This gives an estimate for the biotin dissociation rate, which is independent of the amount of expressed functional protein. ABD-assay was performed essentially as described in Taskinen *et al.*, 2014a.

4.3 Expression and purification of the recombinant proteins

4.3.1 Protein production of avidin variants (I-III)

The recombinant proteins were produced on deep well plates (ABgene, Thermo Scientific, Surrey, UK), in bottle cultures (essentially as described in Hytönen *et al.*, 2004a) or in a 7.5-L fermentor (Labfors 3, Infors, Bottmingen, Switzerland) using the fed-batch culturing method based on pO₂-stat as described previously (Määttä *et al.* 2011). The expression strain was BL21-AI (I-III) or XL1-Blue (II). Protein expression was repressed with glucose, and depending on the expression vector

used, induced by the addition of IPTG (II) or IPTG + L-arabinose (I, III), and carried out at 28 °C.

4.3.2 Affinity purification (I–III)

The recombinant proteins with a 6xHistag were purified using affinity chromatography on a Ni-NTA (QIAGEN) column or with an automated ÄKTA Purifier protein purification system (GE Healthcare Life Sciences) with HisTrap FF Crude columns (GE Healthcare). Bacterial cell pellets were suspended in a binding buffer (phosphate buffer containing a low concentration of imidazole) and lysed by sonication (Sonics & Materials Vibra Cell™) of lysozyme-treated cells (50 µg/ml lysozyme incubated with the cell suspension on ice for 30 min) or by homogenization (EmulsiFlex-C3 homogenizator, Avestin Inc.). For elution, either step-wise (with gravity flow) or gradient elution (with ÄKTA Purifier) was used increasing the imidazole to a maximum of 700mM. Finally, the best elution fractions based on SDS-PAGE-gels and Western blots were pooled together, and the imidazole concentration was reduced to 20mM using step-wise dialysis as the proteins were dialyzed into the measurement buffer (20mM NaH₂PO₄/Na₂HPO₄, 1M NaCl, 20mM imidazole pH 7.4).

In the case of recombinant biotin-binding proteins (II) from small-scale productions, bacterial cell pellets were suspended in PBS containing 1M NaCl, sonicated (Sonics & Materials Vibra Cell™) and bound to biotin resin (D-biotin Sepharose™ 4 Fast Flow, Affiland).

4.3.3 Detection of protein expression via SDS-PAGE (I–III)

Samples from different steps of protein purification were analyzed by staining an SDS-PAGE gel with Coomassie, or alternatively, transferred to a nitrocellulose sheet to be analyzed by immunoblotting. To detect avidin-pIII fusions, monoclonal anti-pIII antibody (1:5,000, MoBiTec GmbH) with AP-conjugated goat anti-mouse antibody (1:30,000, Sigma-Aldrich) were used. To detect avidin variants, polyclonal anti-avidin (University of Oulu) or monoclonal avidin antibody TDA8 (1:5,000) (Kulomaa *et al.*, 1978) with AP-conjugated goat anti-rabbit antibody (1:30,000, Sigma-Aldrich) were used.

4.4 Biophysical analyses

4.4.1 UV-Vis spectrophotometry (I-III)

The concentration and purity ratios (260/280nm and 230/260nm) of the purified plasmids and proteins were analyzed with a NanoDrop 2000 UV-Vis spectrophotometer. Additionally, 20mM progesterone and testosterone stocks were verified with the NanoDrop by measuring wavelength at 245nm, and $\epsilon(245\text{nm})$ 14,000 and 16,596 for progesterone and testosterone, respectively.

4.4.2 Differential scanning calorimetry (DSC) (I, III)

The thermal stability of the avidin variants was studied in the presence and absence of ligands using an automated VP-Capillary DSC System (Microcal Inc.). Protein samples dialyzed into the measurement buffer (see section 4.3.2) were degassed prior to measurement. In study I, protein monomer concentration in the cell was 15 μM (0.225 mg/ml) and the ligand (D-biotin (Sigma-Aldrich) or testosterone (Steraloids Inc.)) was 50 μM . In study III, protein monomer concentration in the cell was 7 μM and ligand concentration was 21 μM . Thermograms were recorded between 20 and 140 °C with a heating rate of 2.0 °C/min. Obtained results were analyzed using the Origin 7.0 DSC software suite (GE Healthcare, Microcal Inc.). In study I, temperature transition midpoints (T_m) were recorded from the highest peaks, whereas in study III, T_m s were recorded with multiple non-two-state equations (and Levenberg-Marquardt iterations) for fitting.

4.4.3 Size-exclusion chromatography (SEC) (I)

The oligomeric state of the avidin variants was analyzed with size-exclusion chromatography (SEC) using a Superdex200 10/300 GL column (GE Healthcare) connected to an ÄKTApurifier™. The measurement buffer (see section 4.3.2) was used as the liquid phase with a flow rate of 0.5 ml/min at room temperature. The column was calibrated using a gel filtration mixture (thyroglobulin, β -globulin, ovalbumin, myoglobulin, and vitamin B12 (Bio-Rad Laboratories)) as a molecular mass standard. Protein samples of 90–193 μg (monomer concentration) in a volume of 500 μl were used in the analysis.

4.4.4 Size Exclusion Chromatography with Static Light Scattering (SEC-SLS) Analysis (III)

The proteins were analyzed for molecular weight (static light scattering) and hydrodynamic size (dynamic light scattering) using a liquid chromatography instrument (CBM-20A, Shimadzu Corporation) equipped with an autosampler (SIL-20A), UV-Vis (SDP-20A), and a fluorescence detector (RF-20Axs), as well as a Zetasizer μ V light scattering detector (Malvern Instruments Ltd.). The instrument was controlled using Lab Solutions Version 5.51 (Shimadzu Corporation) and OmniSEC 4.7 (Malvern Instruments Ltd.). Samples (~ 50 μ g in 10–100 μ l) were injected onto a Superdex200 5/150GL column (GE Healthcare) and equilibrated with the buffer the proteins were dialyzed in (see section 4.3.2). A flow rate of 0.1 ml/min at 20 °C was used. Molecular weight determination was done by calculating a standard curve based on the elution volume of the molecular weight markers (cytochrome C, 12.4 kDa; carbonic anhydrase, 29 kDa; ovalbumin, 44 kDa; BSA, 66 kDa (Sigma-Aldrich), or alternatively, using the light-scattering intensity-based determination protocol involving BSA (monomeric peak) in SLS detector calibration using a Malvern microV detector and the OmniSEC software (Malvern Instruments Ltd.).

4.4.5 X-ray Crystallography (III)

Protein sbAvd-2(I117Y) (1.8 mg/ml in the measurement buffer, see section 4.3.2) was crystallized by our collaborators in the laboratory of Professor Mark Johnson at Åbo Akademi University (Turku, Finland) using the vapor diffusion method. The crystal formed within a few weeks on 96-well triple sitting drop iQ plates (ITP Labtech) used with the mosquito® liquid handling robot (ITP Labtech) and a temperature controlled crystallization incubator (RUMED® model 3201) set-up at 21 °C.

The initial X-ray diffraction properties were analyzed using a PX Scanner (Agilent Technologies). For data collection, cryoprotectant (CryoProtX™, Molecular Dimensions) was added to the crystallization drop just before it was frozen in liquid nitrogen. The data were collected (at the European Synchrotron Radiation Facility, beamline ID23-2, Grenoble, France) and processed using XDS (Kabsch, 1993). Initial phase estimates for the structure factors were obtained using the molecular replacement program Phaser (McCoy *et al.*, 2007) within the CCP4i GUI (Collaborative Computational Project, Number 4, 1994; Potterton *et al.*, 2003).

AVR2 [PDB:1WBI] (Hytönen *et al.*, 2005a) was used as the template structure for molecular replacement. For refinement, Refmac5 (Murshudov *et al.*, 2011) was used, and the structure was rebuilt in several cycles and manually edited using Coot (Emsley and Cowtan, 2004). Solvent atoms were added using ARP/wARP (Langer *et al.*, 2008; Morris *et al.*, 2003; Lamzin and Wilson, 1993; Perrakis *et al.*, 1999) and Coot. To check the final structure and for creating figures, PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.) and Bodil (Lehtonen *et al.*, 2004) were used. For validating the final structure, the inbuilt tools of Coot (Emsley and Cowtan, 2004), and MolProbity (Davis *et al.*, 2007) within the Phenix software suite (Adams *et al.*, 2002) were used. The final coordinates and structure factors of sbAvd-2(I117Y) were deposited in the Protein Data Bank (Berman *et al.*, 2000; Berman *et al.*, 2002) with the accession code 4U46.

4.5 Determination of ligand binding interactions

4.5.1 Fluorometric assay (III)

The binding affinity of the unconjugated small molecules to avidin variants was determined utilizing intrinsic fluorescence originating from the aromatic amino acid residues (mainly tryptophan) of avidin and the fluorescence quenching caused by ligand binding. In brief, protein samples of 100nM were excited at 280nm, and the emission was collected at 350nm using a QuantaMaster Spectrofluorometer (Photon Technology International, Inc.) with 2nm slits. The assay was performed in a quartz cuvette with stirring at 25 °C. The ligand was added to the protein sample in small aliquots, and fluorescence intensity was monitored after a short incubation. First experiments were performed with ligand concentrations of between 66nM–30μM, and repeated for protein-ligand pairs with a K_d value less than 500nM using smaller amounts of the ligand (6nM–50μM) added to the protein sample. The dissociation constant (K_d) was determined from the resulting quenching curve using a GraphPad Prism (GraphPad Software, Inc.). The data were fitted to a quadratic equation for tight binding interactions, which takes ligand depletion and non-specific binding into account. For biotin-binding measurements, up to 350nM protein samples were used in order to achieve a reasonable B_{max} value.

4.5.2 Surface plasmon resonance (I)

The binding kinetics of steroid-binding avidin variants, sbAvd-1 and sbAvd-2, was analyzed with a BIAcore X optical biosensor (Biocore, Uppsala, Sweden). Testosterone-BSA was coupled to the carboxymethylated dextran layer of a sensor chip using EDC/NHS-chemistry (1000 RU, 40 μ l/min flow rate). The binding of sbAvd-1, sbAvd-2, and wt Avd (as a negative control) samples on testosterone-BSA-coated chips was measured using different protein concentrations and the kinetic constants were determined using the BiaEvaluation software according to the manufacturer's instructions. Testosterone-binding was more closely detected by measuring binding in the presence of varying concentrations (0.75–50 μ M) of free testosterone. To evaluate the binding specificity of the steroid-binding avidin variants, the binding of sbAvd-1 and sbAvd-2 to the testosterone-BSA surface was competed against free steroid hormone molecules (testosterone, DHEAS, androstendione, estradiol, and DHT (Steraloids Inc.)) and free biotin (Sigma-Aldrich).

The biotin-binding kinetics were determined by preparing a second sensor chip coupled with biotin (~130RU). However, it is important to note that the determination of the bound mass is not very accurate in the case of small molecules because the immobilization can change the physicochemical properties of the surface.

4.5.3 Protein Microplate Assay (I, III)

The ligand-binding specificity of selected avidin forms was determined with a microplate assay. In the assays, wells of MaxiSorp F96 microplates (Nunc) were coated by applying 100 μ l of PBS containing 500ng of BSA- or HSA-conjugated ligands (listed in Table 5) at +37 °C for 2 hours. As negative controls, carrier proteins HSA (from VTT Technical Research Center of Finland, Espoo) or BSA (Sigma-Aldrich) were used. After blocking and washing the wells, purified proteins (~100 ng/well) in measurement buffer were added. In case the proteins were preincubated beforehand with biotin, 10 μ M D-biotin (Sigma-Aldrich) was used. The avidin variants bound to the wells were detected using polyclonal anti-avidin (University of Oulu) as a probe. After incubation with AP-conjugated coat anti-rabbit IgG (Sigma-Aldrich) and applying phosphatase substrate solution (1mg/ml pNPP (Sigma-Aldrich) in 1M diethanolamine pH 9.8 with 0.5mM MgCl₂), A₄₀₅ was measured with a microplate reader (Bio-Rad 680 XR).

4.5.4 Molecular Dynamics (MD) Simulations (III)

The antidin sbAvd-1 homology model (based on [PDB:1VYO] (Repo *et al.*, 2006)) was prepared, and testosterone was docked into the biotin-binding pocket using GOLD (Hornak *et al.*, 2006). Testosterone was examined only in a binding orientation where it enters the binding pocket with its 17-OH group ahead, since the testosterone-conjugate used in biopanning is coupled to its carrier protein via the 3-keto oxygen. Tetrameric sbAvd-1 with four bound testosterone molecules was solvated in TIP3P water, and the system charge was neutralized with chloride ions, resulting in a total of 52,833 atoms. The Amber ff99SB force field (Jones *et al.*, 1997) was used for the protein, and for testosterone the GAFF force field (Wang *et al.*, 2004), with the simulation run in NAMD 2.9 (Phillips *et al.*, 2005) using a time step of 1fs. A 50-ns MD simulation was produced at 300K and atmospheric pressure using the Langevin temperature and pressure controls. Analyses and figures were prepared using VMD 1.9.1 (Humphrey *et al.*, 1996) and Pymol 1.7 (Shrödinger, LLC).

4.5.5 Interaction analysis by Molecular Recognition Force Spectroscopy (I)

The interaction of sbAvd-1 and sbAvd-2 with testosterone and biotin was studied also by molecular recognition force spectroscopy (MRFS) by our collaborators in the laboratory of Professor Peter Hinterdorfer at Johannes Kepler University Linz (Linz, Austria). The proteins were covalently bound to modified mica sheets via lysines as previously described (Kamruzzahan *et al.*, 2004). The testosterone was covalently attached to an AFM tip via a flexible spacer, heterobifunctional Fmoc-PEG-NHS crosslinker, as previously described (Wildling *et al.*, 2009). All MRFS experiments were performed on a Pico SPM I (Agilent Technologies, Santa Clara, CA). All the used, modified cantilevers had nominal spring constants between 10-30pN/nm (Veeco Instruments, Santa Barbara, CA), and the effective spring constants were determined by the thermal noise method (Hutter and Bechhoefer, 1993; Butt and Jaschke, 1995). Force-distance cycles were completed using a z-range of 200-300nm. Sweep durations were adjusted between 0.25–4s. During one data set of 1000 force-distance curves, the lateral tip position was changed (a few hundred nm) about every 100 curves to ensure that the binding events were statistically reasonable. To prove the specificity of the binding, the ligand-binding sites of proteins were blocked by ~1h incubation with free testosterone (200nM) added to the measuring solution.

5 RESULTS AND DISCUSSION

Based on the original papers (I–III) and unpublished material.

5.1 Functional display of avidin on the M13 phage (I)

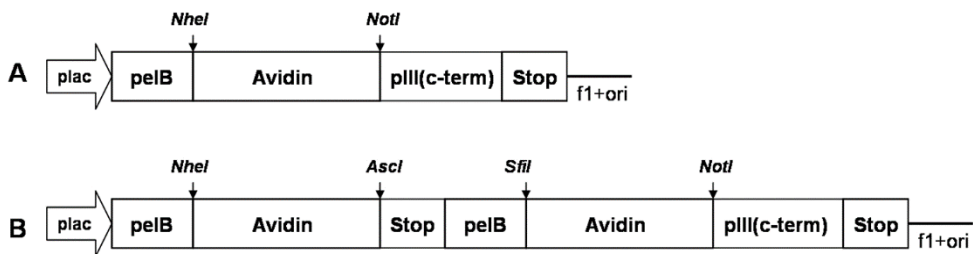


Figure 9. Two different display constructs: a) Avid (or as a positive control for panning test, Avid(N118M)) protein is produced solely as a fusion with C-terminal pIII (referred to as Avid-pIII display); and b) Avid (or Avid(N118M)) displayed with the corresponding free subunits (referred to as Avid/Avid-pIII display). The latter mimic the situation with an amber stop codon. Figure from I, reprinted with permission from Riihimäki *et al.*, 2011b.

To establish the avidin phage display system, avidin was chosen to be displayed on the M13 bacteriophage as a fusion with the C-terminal region of the minor coat protein pIII. Two different display constructs were used to evaluate the strategies for displaying the avidin scaffold on the M13 phage: 1) Avidin was produced either solely as a fusion with pIII (referred to as the Avid-pIII display construct, Figure 9a), or 2) with avidin subunits (referred to as the Avid/Avid-pIII display construct, Figure 9b). The idea in production of free subunits is to mimic the generally used amber-stop codon technique, in which the read-through of the amber stop codon by the tRNA results in the production of free subunits. The free subunits should enhance the functional assembly of the tetrameric avidin scaffold, because membrane anchoring of avidin by the pIII fusion partner might have a negative effect on the oligomerization of avidin subunits and thus to the function, i.e. biotin binding of avidin. Oligomerization of avidin takes place in the periplasmic space of *E. coli* when

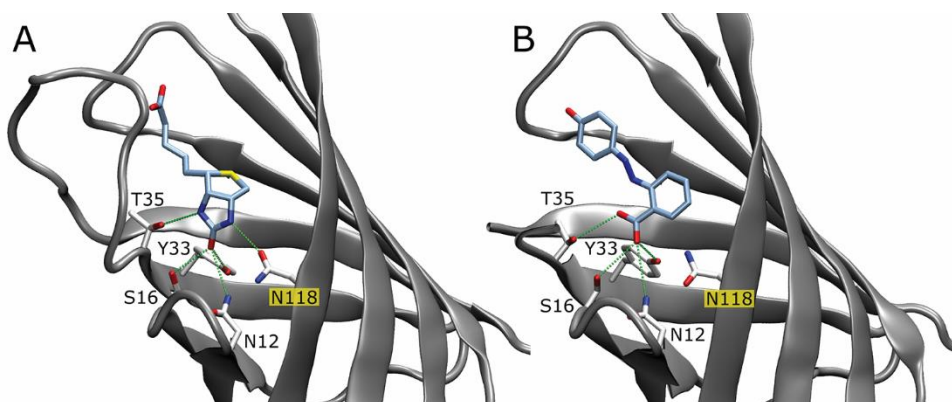


Figure 10. Structural comparison of the Avid-BTN and Avid-HABA complexes. Three-dimensional structure of avidin monomer complexed with biotin (A) and HABA (B). Hydrogen bonds formed between the ligand and Avid are shown as dashed green lines. When comparing the interaction of the ligands with Avid, a clear and significant difference is seen in the interaction with asparagine 118: There is no hydrogen-bonding partner for N118 in HABA, although other key residues interacting with the biotin ureido ring also interact with the carboxyl group of HABA. Figure from I (Additional file 1), reprinted by permission from Riihimäki *et al.*, 2011b.

the Sec pathway utilizing ompA signal peptide from *Bordetella avium* is used (Hytönen *et al.*, 2004a), and thus the usage of the pelB signal, utilizing the same secretion pathway, should work similarly.

In order to validate the constructs (Figure 9) and nascent phages, Western blot was used to analyze the size and oligomeric state of the avidin-pIII fusions, whereas the biotin binding of avidin displaying phages was studied in biopanning. In addition to wt Avid, the Avid(N118M) mutant was also displayed on the phages using two separate constructs in parallel (Figure 9). At first biotin-coated surfaces were used in biopanning to enrich the phages and to show that avidin-displaying phages are functional. Next, the mixture of avidin and Avid(N118M) phages were used in panning against 4-hydroxyazobenzene-2-carboxylic acid (HABA). Previously, Avid(N118M) has been shown to have ~1.5-fold increased HABA-affinity ($K_d = 5.2\mu\text{M}$) and a million-fold reduced biotin-affinity compared to wt Avid (Määttä *et al.*, 2008) (Figure 10), and was thus used as a positive control.

In Western blot (I, Figure 2), polyclonal anti-avidin showed bands that matched the size of the Avid-pIII fusion (~38kDa) and free avidin (~15kDa), indicating that both of these avidin variants were successfully expressed from the corresponding constructs (Figure 9). Monoclonal anti-pIII revealed that a portion of the Avid-pIII fusion was partially proteolytically cleaved (~35kDa). Moreover, since the avidin

Table 7. Results from the control panning to select HABA-binders. The percentage of Avd and Avd(N118M) sequences from different panning rounds. The output phage amount is shown as colony forming units (cfu) per milliliter of phage stock. Table modified from I. Reprinted by permission from Riihimäki *et al.*, 2011b.

Selection round	Output (cfu/ml)	Avd-pIII / Avd(N118M)-pIII (seq %)	Output (cfu/ml)	Avd/Avd-pIII / Avd(N118M)/Avd(N118M)-pIII (seq %)
1 st	3000	25 / 75	18 000	33 / 67
2 nd	400	50 / 50	12 000	25 / 75
3 rd	3000	0 / 100	14 000	0 / 100

constructs were based on a monovalent display mode (3+3), intact pIII (~58 kDa) expressed from the helper phage was also detected. Panning against HABA resulted in a clear enrichment of Avd(N118M) phages as expected. After only three panning rounds of selection, phages displaying the Avd(N118M) mutant outcompeted the wt Avd phage population, which indicated the high selectivity of the produced phages (Table 7). Moreover, phagemid DNA with the avidin display expression unit was confirmed by restriction enzyme digestion analysis to be stable during the panning rounds (data not shown).

To conclude, these tests showed that functional Avd could be displayed on the surface of M13 phages. Retrospectively, however, one could have compared the phagemid containing pelB signal with the previously used successful ompA signal peptide, since it has been shown there are differences in the expression level from protein to protein (Low *et al.*, 2013). Avd/Avd-pIII-constructs were shown to be more efficient, but since these constructs were not optimal for library construction (enabling chimeric Avds to form), they were abandoned. However, it would have been easy to add the amber stop codon in between avidin and c-terminal pIII to replace the Avd/Avd-pIII construct. This was executed later, first with DNA-shuffled libraries (Niederhauser *et al.*, 2012).

5.2 Library design of genetic antidin libraries (I, III)

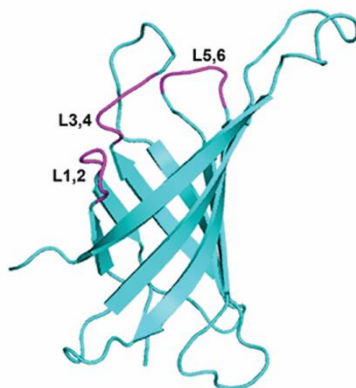


Figure 11. Avidin monomer [PDB: 1AVD] shown with named loop areas targeted for randomization (in magenta) in different antidin libraries. Figure modified from III and reprinted with permission from Lehtonen *et al.*, 2016. Copyright 2016 American Chemical Society.

In order to create novel small molecule recognition proteins, antidins, based on the avidin scaffold, the site-directed random mutagenesis strategy was used for which the loops adjacent to the ligand-binding site of avidin were selected as targets (Figure 11) based on the 3D structure of avidin [PDB:2AVI] (Livnah *et al.*, 1993). Four to nine amino acid residues were randomized per library in one or two loops of avidin using the degenerate codons NNN or NNY. NNN enables all the 20 different amino acids including all stop codons, whereas NNY restricts randomization to 15 of the 20 proteinogenic amino acids eliminating all stop codons. Table 8 summarizes the details of the constructed libraries. The phage display libraries were based on the monovalent display mode (3+3), and the library was expressed as a sole fusion with C-terminal pIII (monovalent pIII display).

First, the Avd L1,2 library was constructed in which a loop between β -strands 1 and 2, L1,2, (consisting of residues N12, D13, L14, G15 and S16) was randomized using the degenerate codon NNN, and the library was used to select testosterone-binding variants (I). In order to further lower the biotin-binding affinity and to decrease the cross-reactivity of the selected steroid-binding avidin, the affinity maturation library, sbAvd-1 L3,4 (I) (later named as AvLib-3 (III)), was constructed containing a randomized loop area between β -strands 3 and 4. The L3,4 loop is highly important for biotin binding as it “locks” biotin inside the binding pocket. In L3,4, there are four amino acid residues that form direct interactions with biotin

(T35, T38, A39, and T40, Figure 2), and thus a region containing residues 35–38 (Figure 11) were randomized using the degenerate codon NNY.

Later, two new libraries were constructed (III): (1) AvLib-1, utilizing the same strategy as the Avd L1,2 library, but this time a part of the loop between β -strands 5 and 6 of avidin (amino acid residues F72, S73, E74, and S75) was selected for randomization with the degenerate codon NNN, and (2) AvLib-2, where the L1,2-loop together with a part of the L5,6-loop (N12, D13, L14, G15, S16, F72, S73, E74, and S75) were randomized with the degenerate codon NNY. The previously constructed AvLib-3 library was pooled together with AvLib-1 and AvLib-2 before biopanning (III).

The first constructed Avd L1,2 library (I) was found to consist of approximately 1×10^5 individual clones based on sequencing results and transformation efficiency (the theoretical size of the library was $20^5 = 3.2 \times 10^6$ based on all the different amino acid residues). Thus, only a fraction equaling 3% of the theoretical sequence space was covered (I). The theoretical sequence space of the created libraries AvLib-1 ($20^4 = 1.6 \times 10^5$) and AvLib-3 ($15^4 = 0.05 \times 10^6$) was fully covered based on the number of the unique transformants (2.1×10^6 and 1.4×10^6 , respectively), whereas from the library AvLib-2 ($15^9 = 3.8 \times 10^{10}$) only a fraction (0.02%) was covered as the library contained only 5.9×10^6 unique transformants (I, III) (Table 8).

The usage of degenerate codons creates bias in a library due to the fact that 1) there are several different codons encoding the same amino acid residues and thus certain amino acids are overrepresented compared to others. Additionally, 2) bias is also caused by tRNAs present in *E. coli*. On the other hand, degenerate codons are easy to use, and library construction using them is quite inexpensive. However, the use of the degenerate codon NNN (Table 3) on reflection should have been completely avoided, since it drastically increases the screening effort: There are 64 codons per amino acid position to be randomized including all the stop codons. Thus the library contains a number of mutants having at least one stop codon (3/64 possibility at each site, resulting in truncated proteins). Instead of NNN, the wiser option would have been to choose NNK (Table 3). The screening effort would have been less (32 codons per amino acid position to be randomized) containing all the amino acids but eliminating 2/3 of the stop codons. Therefore, already four or five NNN-randomized amino acid residues in the libraries AvLib-1 (III) and Avd L1,2 (I), respectively, would require at least ($64^4 =$) 16.8×10^6 and ($64^5 =$) 1.1×10^9 colonies, respectively, to theoretically cover all the codon options. When taking into account stop codons causing premature termination of transcription however, full coverage would require even bigger libraries: 1.2×10^8 and 8.7×10^9 , respectively (Nov, 2014).

The constructed libraries thus represented only negligible fractions of this sequence space (Table 8). Luckily, full coverage is not desirable as such: When the library is big and the number of all possible distinct variants is large, any of the top-performing variants is likely to meet design requirements (Nov, 2014). Thus, TopLib computer software (<http://stat.haifa.ac.il/~yuval/toplib/>, last accessed 24.5.2017) was used to calculate the probabilities for finding these top variants from our constructed libraries (Nov, 2012). The probability for existence of one of the top five candidates from the constructed Avd L1,2 library (I) is 0.11, one of the top 10 candidates is 0.22 and one of the top 100 candidates >0.9 . In contrast, the probability for existence of one of the top 3 candidates from AvLib-1 is 1 (Nov, 2014). In the case of libraries AvLib-1 (III) and Avd L1,2 (I), NNK would have meant the requirement of 1.05×10^6 and 3.36×10^7 clones, respectively. Full coverage of the libraries would have meant 1.35×10^7 and 5.17×10^8 , respectively. If we will still continue to speculate, with the colony numbers obtained (Table 8), these libraries (Avd L1,2 (I) and AvLib-1 (III)) would have included one of the top eight binders with a probability >0.95 .

Later the degenerate codon NNY (Table 3) was selected instead of the previously used NNN to avoid the possibility of any stop codons (AvLib-2 and AvLib-3 (III)). However, it also entirely eliminates five amino acids: Trp (W), Gln (Q), Glu (E), Met (M), and Lys (K). Nevertheless, when considering the enriched amino acid residues found in the antigen recognition sites of the antibodies (Collis *et al.*, 2003), the important amino acid residues for the antigen binding site, Tyr (Y) and Ser (S), are still present in the NNY-library, although unfortunately Trp (W) is omitted. The clear advantage of the NNY codon is the reduction of the required library size (only 32 codons per amino acid position without stop codons). However, even with the NNK-codon, the library size of the constructed AvLib-3 would have been enough for theoretically certainly containing one of the top two binders – and even the best binder with a probability of 0.95 (Nov, 2014).

To sum up, the Avd L1,2 library construction strategy was a bit too ambitious, especially for the first library. In practice, the hit found in the Avd L1,2 library is probably just one of the top 120 variants. Thus, we could have reached a much better candidate for the basis of the affinity maturation library by using a bit wiser library construction scheme. The AvLib-3 was fully covered, but in the case of AvLib-2, the probability of containing even one of the top 100 candidates in the library is only 0.02. Retrospectively, one should have used NNK or NNY for the construction of the Avd L1,2 and AvLib-2-libraries, and possibly also restricted (especially in the case of AvLib-2) the sites to be randomized.

Table 8. Characteristics of the avidin gene libraries. Table modified from III and reprinted with permission from Lehtonen *et al.*, 2016. Copyright 2016 American Chemical Society.

Library name	Gene	Randomized loops	Randomized amino acids	Theoretical size** ($\times 10^6$)	Library size*** ($\times 10^6$)	Effective library size **** ($\times 10^6$)	Degenerate codon	Completeness	N:o possible codon variants***** ($\times 10^6$)
Avd L1,2	wt Avd	L1,2	N12 , D13, L14, G15, S16	3.20	0.1	0.08	NNN	0.02366	1 073.00
Avlib-1	wt Avd	L5,6	F72 , S73 , E74, S75	0.16	2.1	0.15	NNN	0.94470	16.80
Avlib-2	wt Avd	L1,2 & L5,6	N12 , D13, L14, G15, S16 , F72 , S73 , E74, S75	38 443.34	5.9	38 443.34	NNY	0.00015	35 184 372.09
Avlib-3*	sbAvd-1	L3,4	T35 , A36, V37, T38	0.05	1.4	0.05	NNY	1.00000	1.05

* = affinity maturation library sbAvd-1 L3,4 (Riihimäki *et al.*, 2011b)

** based on possible amino acid residues (NNN= 20 / NNY=15)

*** calculated from transformation activity

**** number of distinct amino acids (not including stop codons); calculated mainly using GLUE-IT (up to 6 codons)

***** based on possible codons (NNN = 64 codons, NNY = 32 codons)

amino acids directly involved in biotin-binding are marked in bold

On the other hand, the total number of nine amino acid positions at maximum for randomization is still low compared to for example the anticalin libraries, typically containing 16–24 randomized residues (Gebauer and Skerra, 2012). The number of randomized residues is also similar to other successful protein scaffolds (Lipovsek, 2011; Löfblom *et al.*, 2010; Plückthun, 2015). Many of these libraries have been prepared using defined mixtures of nucleotide triplets ensuring quite equal amino acid residue distribution, at the same time avoiding stop codons and unfavorable residues (e.g. cysteines) completely. Enabling 19 amino acid residues for example at each of 20 positions still leads to a huge number of theoretical combinations ($\sim 10^{25}$). Even state-of-the-art phage display libraries enable complexities of only 10^{10} – 10^{11} , and thus permit limited sampling only for libraries as huge as these. Nevertheless, high affinity binding proteins have been able to select in these circumstances. Broad sampling is intended to achieve coverage of different binding modes and/or epitopes on the target. Of course, the number of different codons may be further reduced, but this also restricts the chemistry of the allowed side chains (Gebauer and Skerra, 2012). However, successful library design and construction, per se, is not a guarantee for a successful outcome, although it enhances the probability for selecting a variant with the desired properties. The selection conditions, especially the selection stringency, have a vital role for the outcome. It is essential that the selection pressure is not too high to enable recovery of the rare sequences.

5.3 Capture and characterization of antidins (I, III)

In order to select avidin variants capable of binding small target molecules other than biotin from the constructed phage display libraries, biopanning was used. The small molecules were chosen as targets for their size (enabling them to fit in the deep ligand-binding pocket of avidin), as well as their clinical relevance. Thus, the development of a specific binding protein for any of these molecules would be valuable, enabling their utilization in clinical applications or in point-of-care tests. Among the target molecules chosen for example, steroid hormones are difficult for high-affinity monoclonal antibody development due to their hydrophobic nature with only a few functional groups. On the other hand, they are interesting target candidates for the avidin scaffold which naturally provides a deep ligand-binding pocket, which would be preferable for a steroid-binding molecule. According to Schlehuber and Skerra (2005), the optimal steroid-binding pocket would possess a

hydrophobic cavity with shape complementarity generated through aromatic side chains and appropriately placed hydrogen bond donors and acceptors (Schlehuber and Skerra, 2005).

At first, the steroid hormone testosterone was used as the sole target (I) for the phage display library Avd L1,2, which was panned against BSA-conjugated testosterone-3-carboxymethyloxime (testosterone-3-CMO, Table 5). A couple of other small molecules (estradiol, amphetamine, morphine, THC) were also later used as targets for this library, but with no success (unpublished data). An enriched avidin variant, the steroid-binding avidin sbAvd-1, was used to construct an affinity maturation library (sbAvd-1 L3,4 (I) – later named AvLib-3 (III)) from which an improved avidin variant, sbAvd-2, with a higher affinity towards testosterone was obtained.

Later, three different avidin phage display libraries (AvLib1–3) were pooled together and used for biopanning against seven different diagnostically relevant ligands in parallel (III). These ligands were progesterone, hydrocortisone, testosterone, cholic acid, ibuprofen, ketoprofen, and folic acid (Table 5). Target molecules included different steroids since the previous work (I) had yielded successfully captured steroid-binding avidins from the AvLib-3 library using testosterone as a target. Furthermore, including different steroids as target molecules would provide important data about the specificity of the enriched variants (through possible cross-reactivity, or lack of it). Overall, the selected set of ligands has variability both in molecular size and chemical structure, and thus enabled us to explore the potential of the generated protein libraries to accommodate a variety of ligands.

In order to select several small molecule target-binders in parallel from the pooled AvLib1–3 libraries, semi-automatic panning with a magnetic particle processor was used. Considering the immobilization of the antigen for the panning, the main advantage of using magnetic beads (in III) over microtiter plates (I) is the increase in surface area, which enables higher antigen concentrations and thus leads to a more efficient panning process. Especially in cases where actual binding phages are extremely rare in the library, it is crucial to recover these phages in the very first round of selection (as well as in subsequent rounds), thus minimizing the repetitive isolation of identical phage sequences (so-called “sibling” phages) derived from replication of a single parent (McConnell *et al.*, 1999). Additionally the washing step is more efficient, as moving the magnetic particles from one well to another reduces the background of unspecific binders on surfaces and only minimal volumes are transferred. Importantly, however, the magnetic plate processor enables selection of

uniform quality for several different target molecules in parallel (McConnell *et al.*, 1999).

After pannings, sequencing and microplate assays were used to screen the enriched variants. The selected set of avidin variants, antidins, were produced in *E. coli* and purified with immobilized nickel affinity chromatography, after which they were characterized using various methods (described in sections 4.4–4.5). Semi-automatic panning yielded avidin variants with binding activity for six out of seven target molecules. Protein productions and purifications of the selected antidins yielded 2–10mg/L of protein with >95% purity based on SDS-PAGE analysis. Antidins were found to have a tetrameric quaternary structure as in the case of wt avidin, but their ligand-binding properties differed significantly from wt avidin (Tables 10 and 11). Wt avidin had generally negligible affinity toward the ligands other than biotin, except in the case of progesterone, for which $K_d \sim 3 \mu\text{M}$ (Table 10 and 11) was achieved.

5.3.1 Steroid-binding antidins (I, III)

The steroid-binding avidin variants sbAvd-1 and sbAvd-2, enriched from Avd L1,2 and the affinity maturation (AvLib-3) library, respectively, were the first antidins characterized (I). The later performed semi-automatic biopanning of the pooled phage libraries AvLib1–3 (III) against different target ligands also yielded the enrichment of sbAvd-2 from the AvLib-3 library in panning against testosterone and progesterone (III, Table 2). Table 9 shows the avidin variants enriched to bind different steroid molecules.

The microplate assay with a set of different small molecule targets used in study III showed that steroid-binding avidin variants (sbAvds) can clearly bind to progesterone and testosterone. Importantly, sbAvds showed no binding to the carrier proteins (bovine serum albumin, BSA, and human serum albumin, HSA) of the used small molecule conjugates (listed in Table 5). All sbAvds showed a strong preference for binding progesterone over testosterone: Affinities (K_d) for progesterone were in the low micromolar range, whereas the affinities for testosterone were 10–40-fold weaker, and in the micromolar range. Additionally sbAvd-5 and sbAvd-6 were also enriched during ketoprofen selection (III, Table 2), but affinities toward the non-steroid ketoprofen were significantly weaker ($\sim 3 \mu\text{M}$) compared to progesterone. DSC measurements revealed that the sbAvds had preserved the high thermal stability of wt avidin ($T_m \sim 80^\circ\text{C}$).

Table 9. The enriched target steroid binding antidin sequences compared to the original sequence of wt avidin. The partial avidin monomer with the modified portions of the loop structures are shown (in purple), as well as the amino acid residues of wt avidin in those positions. X-axis in each sequence logo indicates the numbers of the amino acid residues and possible Y-axis indicates the sequence conservation at each position (measured in bits). The empty space in the table means that no changes were made to the original sequence at that site. Amino acids are coloured according to their chemical properties: polar amino acids (G,S,T,Y,C) are green, neutral (Q, N) purple, basic (K,R,H) blue, acidic (D,E) red, and hydrophobic (A,V,L,I,P,W,F,M) amino acids are black. The avidin monomer with biotin image is modified from III, and sequence logos were drawn with the WebLogo (v. 2.8.2) server (<http://weblogo.berkeley.edu/logo.cgi>, accessed last time 12.7.2017). Reprinted with permission from Lehtonen *et al.*, 2016. Copyright 2016 American Chemical Society.

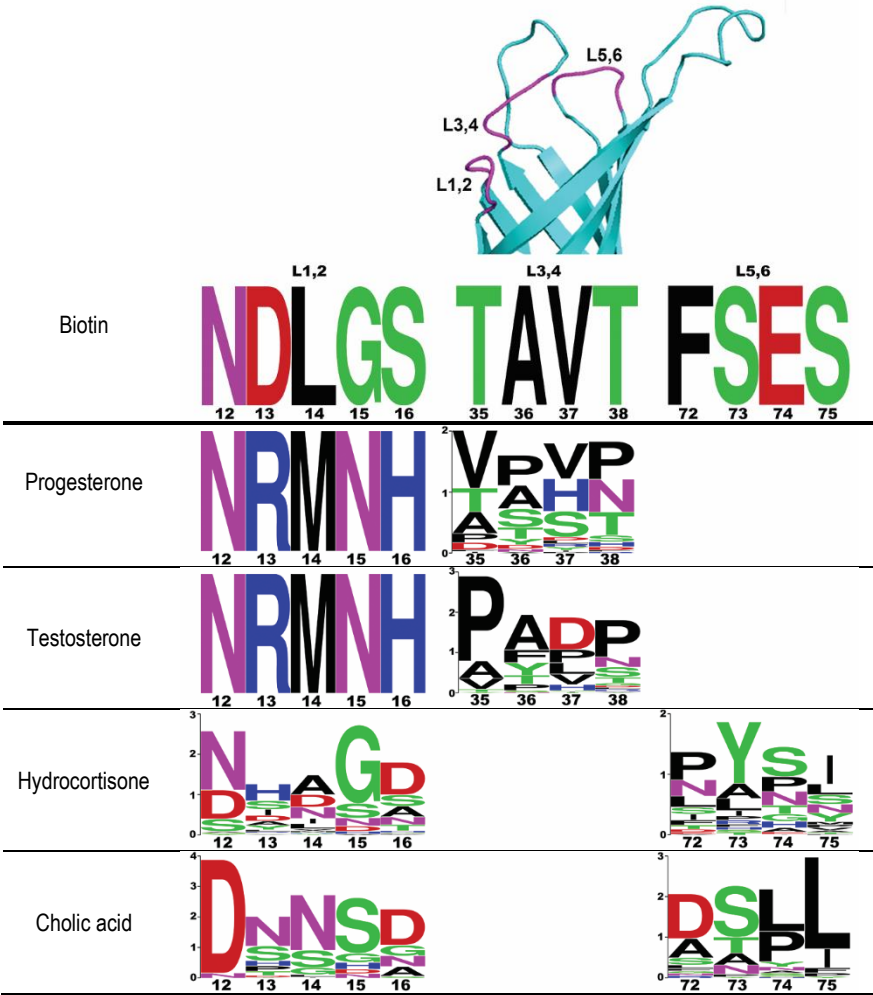


Table 10. Determined T_m and K_d values of steroid-binding antidiins. Table modified from III and reprinted with permission from Lehtonen *et al.*, 2016. Copyright 2016 American Chemical Society.

Protein	Ligand	DSC T_m (°C)	ΔT_m (°C)	Fluorometry K_d (nM) ^a	SPR k_a (1/Ms)	SPR K_d (1/s)	SPR K_D (nM)
sbAvd-1		80.6 ^b	± ^c				
	biotin	83.2 ^b	-	n/a	4.2 x 10 ⁵	5.6 x 10 ⁻⁴	1.4
	progesterone	90.7	-	174 ±3			
	testosterone	83.1	-	6299 -	1.0 x 10 ³	9.5 x 10 ⁻³	9000
sbAvd-2		81.2	±1.1				
	biotin	81.6	±0.7	6649 ±504	1.0 x 10 ³	6.8 x 10 ⁻⁴	660
	progesterone	94.3	±0.5	111 ±35			
	testosterone	86.7	±1.0	2986 ±1438	813	8.5 x 10 ⁻³	11 000
sbAvd-3		80.6	±1.2				
	biotin	80.8	±0.8	8580 ±673			
	progesterone	92.3	±0.4	249 ±90			
	testosterone	84.6	±0.8	1841 -			
sbAvd-4		79.6	±0.8				
	biotin	79.9	±0.6	4731 -			
	progesterone	91.5	±1.5	135 -			
	testosterone	85.1	-	2172 -			
sbAvd-5		83.3	±1.4				
	biotin	83.0	±1.1	7312 -			
	progesterone	96.0	±0.8	108 -			
	testosterone	88.8	-	5.5 1630 -			
	ketoprofen	84.5	-	1.2 2647 -			
sbAvd-6		81.8	-				
	biotin	81.6	±1.7	3303 -			
	progesterone	93.9	±1.0	117 ±26			
	testosterone	86.9	-	5.0 1469 -			
	ketoprofen	83.3	-	1.5 3058 -			
wt Avd		80.8	±0.1				
	biotin	118.7	±1.7	37.9 n/d ^d			
	progesterone	83.0	±1.6	2.2 3355 -			
	hydrocortisone	82.1	±0.6	1.3 n/a ^e			
	testosterone	82.7	±0.3	1.8 n/a			
	cholic acid	81.5	±0.3	0.6 n/a			
	ketoprofen	82.3	±0.1	1.5 n/a			
	folic acid	81.5	±0.2	0.6 n/a			

^a K_d obtained using 100nM protein concentration, and a fit for tight ligand binding, which accounts for the ligand depletion. In the case of biotin-binding measurements, up to a 350nM protein concentration was used to achieve measurable fluorescence quenching. ^bThe result from I. ±SD, standard deviation, is shown, when more than one measurement. ^dn/d, not determined. ^en/a, the data could not be fitted to the model, or the total fluorescence quenching due to the addition of the ligand was less than 2000 units.

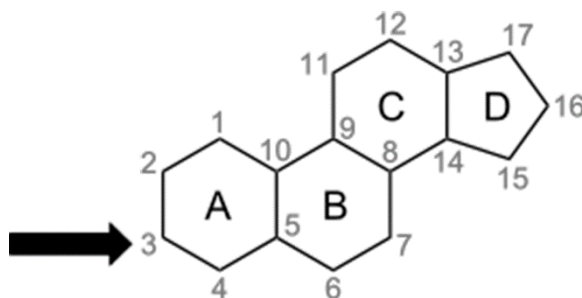


Figure 12. The steroid skeleton forming the basic structure of steroids. The four ring structures are named A–D and the figure shows the numbering of the carbon atoms in the structure. The arrow indicates the site from which most steroid molecules used in panning are conjugated (C-3). Cholic acid is an exception, as it is conjugated through one of its intrinsic carboxyl groups residing in the side group attached to C-17.

Testosterone-binding properties. The first sbAvds (sbAvd-1 and sbAvd-2) were characterized using a biosensor for surface plasmon resonance (SPR) (I). The later determined K_d value for testosterone binding of sbAvd-1 based on the fluorescence-quenching curve ($\sim 6\mu\text{M}$, Table 10) (III) was quite consistent with the inhibition assay (50% inhibition with 5–10 μM testosterone) executed with a testosterone-coated chip in SPR (I). However, in the case of sbAvd-2, the K_d value $\sim 3\mu\text{M}$ (Table 10) (III) differed significantly from the previous result where 50% inhibition on the SPR-chip was already achieved with 750nM free testosterone (I). This could be explained partially by the lower coating level of the chip with sbAvd-2 compared to sbAvd-1, as seen also in Figure 5 (in I). However, it is difficult to predict due to the oligomericity of antidin and possible heterogeneity within the BSA-conjugated ligands. The specificity of these sbAvds to the testosterone-BSA-coated sensor chip was analyzed by binding competition using biotin and a selected set of different steroids (testosterone, dehydroepiandrosterone, androstenedione, estradiol, and dihydrotestosterone). According to the obtained results, (I, Figure 5) sbAvd-1 shows the most cross-reactivity towards testosterone-resembling androgens (dehydroepiandrosterone (DHEAS) and androstenedione) as well as for biotin, whereas the inhibition of sbAvd-2 binding to testosterone-BSA was most efficiently achieved by testosterone and DHEAS. The inhibition effect of biotin was clearly weaker for sbAvd-2 (I). Unfortunately, progesterone was not included in this assay, since the affinity towards it was only revealed later (III). Results obtained from the molecular recognition force spectroscopy (MRFS), utilized in order to study the interaction between sbAvds (more specifically sbAvd-1 and sbAvd-2) and testosterone on a single molecule level, were well in line with the other results, showing sbAvd-2 had a higher affinity towards testosterone (I).

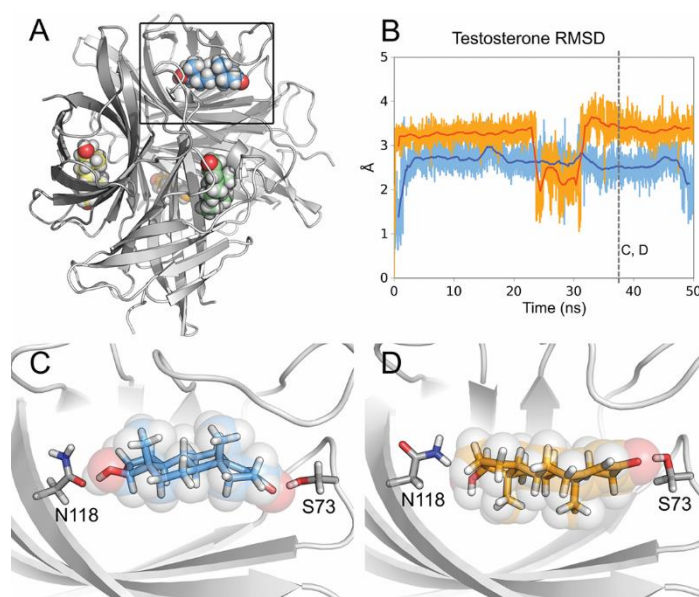


Figure 13. Molecular dynamic simulation of sbAvd-1 with bound testosterone molecules. (A) Testosterone binding to a tetrameric sbAvd-1 model in a 50ns molecular dynamic simulation. (B) The distances of two testosterone molecules (blue, orange) in reference to their docked positions are measured using root-mean-square displacement (RMSD). (C, D) Snapshots showing the two testosterone molecules in the ligand-binding pockets at 37.5ns as indicated in B. Figure from III, reprinted with permission from Lehtonen *et al.*, 2016. Copyright 2016 American Chemical Society.

SbAvd-1, as the first antidin, was used in molecular dynamic (MD) simulations with testosterone (III) in order to model the interaction via ligand docking, thus helping to design further mutations to the protein. A 50ns MD-simulation of tetrameric sbAvd-1 with four testosterone molecules bound to the ligand-binding pockets revealed two main binding modes, where testosterone methyl groups pointed either toward the L3,4-loop (51% of the time) or toward the β 7 strand (41% of the time). Both of these modes involved the formation of two hydrogen bonds between the testosterone molecule and sbAvd-1, where the participating amino acid residues were Asn118 and Ser73. More specifically, the interactions were formed between testosterone 17-OH and the side chain of Avd Asn118, and between the 3-keto oxygen and the side chain of Avd Ser73 (N118 and S73, respectively, in Figure 13) (III). Testosterone was examined only in a binding orientation where it enters the binding pocket with its 17-OH group ahead (Figure 13), since the testosterone-conjugate used in biopanning was coupled to its carrier protein via the 3-keto oxygen. However, we cannot be sure if the free testosterone prefers entering the ligand-binding pocket with the other end first.

Progesterone-binding properties. Since the previously selected antidiins, sbAvd-1 and sbAvd-2, were later isolated from panning against progesterone, binding affinities were determined using a fluorometric assay and found to be $K_d=174\text{nM}$ and $K_d=111\text{nM}$ for sbAvd-1 and sbAvd-2, respectively (Table 10) (III). The affinities of sbAvd-3–6 for progesterone and testosterone are within the same, low micromolar range ($K_d=108\text{--}249\text{nM}$) (Table 10) (III).

Biotin-binding properties. The clear difference between the antidiins sbAvd-1 and sbAvd-2, was also seen when comparing their binding to the biotin-coated sensor chip in SPR. The measurements revealed almost 500-fold decreased biotin-binding affinity for sbAvd-2 due to the modification of the L3,4-loop ($K_d=1.4\text{nM}$ and 660nM for sbAvd-1 and sbAvd-2, respectively) (I). However, the determined K_d for sbAvd-2 with the fluorometric assay was even weaker ($K_d\sim 6.6\mu\text{M}$) (Table 10) (III). This 10-fold difference can be explained by the avidity effect: The situation is completely different when the tetrameric avidin variant is bound to a surface-coated ligand (as in SPR) compared to free ligand (this case in the fluorometric assay) (Jung *et al.*, 2000; Vauquelin and Charlton, 2013). Altogether, the biotin-binding affinities of sbAvd-2–6 were drastically lower ($3\text{--}9\mu\text{M}$) (Table 10), as compared to wt avidin (and significantly lower compared to sbAvd-1), and the preincubation of sbAvds with biotin could not inhibit progesterone binding in the microplate analysis (Figure 15) (III).

To conclude, sbAvds are promising antidiins, showing that even a change of 7–8 amino acid residues can result in a clear difference in ligand specificity without destabilizing the protein scaffold. The T_m values around $80\text{ }^\circ\text{C}$ showed that sbAvds were able to retain the high thermal stability. Although the obtained affinities and the specificities of the so-far developed antidiins are not comparable with antibodies, they show that avidin can be used as a scaffold and ligand-binding properties can be tailored. These antidiins could possibly be used for example as a sample pretreatment phase to concentrate steroids from a biological or environmental sample.

5.3.2 Other antidiins (III)

Hydrocortisone and cholic acid -binding antidiins. Two hydrocortisone- and two cholic acid-binding antidin forms (hbAvd-1, hbAvd-2, cabAvd-1, and cabAvd-2) enriched from the hydrocortisone and cholic acid selections (III) with T_m values around $64\text{ }^\circ\text{C}$ for the hbAvds and $57\text{ }^\circ\text{C}$ for the cabAvds (Table 11) had significantly more destabilized structures compared to the other antidiins. The protein structures

could be stabilized however by adding the ligands used in the selection of these variants (Table 11) (III). The microplate analysis (Figure 14) showed hbAvd-2 and cabAvd-2 to be cross-reactive with several ligands in the study: progesterone, hydrocortisone, testosterone, and cholic acid. Even though the target ligands (hydrocortisone and cholic acid) are slightly larger molecules compared to progesterone and testosterone, they are also steroids. Affinities were determined utilizing the fluorescence quenching assay generally yielding micromolar affinities towards their target molecules. However, measurements revealed that the enriched avidin variants can have even higher affinity (in the low micromolar range) towards the smaller steroids, progesterone and testosterone (Table 11) (III). This can be an indication of several things: 1) The ligand-binding pocket as such is not big enough for the bigger steroids. 2) Especially in the case of hydrocortisone, it might have been better to conjugate it from the other end of the molecule, allowing the hydrophobic steroid rings to go in first.

Nevertheless, compared to wt Avd, a significant reduction in biotin-binding affinities were achieved (2.4–34 μ M) (Table 11). Among the hydrocortisone and cholic acid -binding antidiins, cabAvd-2 was the most promising showing a clearly changed ligand-binding preference: The determined affinity for cholic acid was 64nM, whereas affinity toward biotin was 4.7 μ M (Table 11). Interestingly, cabAvd-2 shows how one can reach a similar affinity compared to sbAvds towards progesterone by combining different amino acid residues in the loop regions important for ligand-binding (Table 9 and 11).

Folic acid -binding antidiin. Selection against folic acid resulted in a captured antidiin, fabAvd-1, with T_m 68 °C. The determined K_d values showed a low micromolar affinity (K_d =188nM) toward folic acid and significantly decreased biotin-binding affinity (K_d ~16 μ M) (Table 11) (III). In the microplate assay, the protein showed a clear binding response to folic acid (Figure 14), which however could be competed with biotin (Figure 15). Also the thermal stabilizing effect of folic acid (three times molar excess) was modest (ΔT_m =3.7 °C) compared to the determined affinity (Table 11). This inconsistency between the obtained K_d values for fabAvd-1 and behavior on the microplate assay, as well as the modest thermal stabilizing effect of folic acid, could be explained by the relatively large size of folic acid ($M_{\text{folic acid}}$ =441.4 g/mol) compared to biotin (M_{biotin} =244.3 g/mol): Due to the size difference, folic acid, being almost twice as big as biotin, most probably cannot completely fit into the binding site of fabAvd-1. Therefore it cannot enhance the structural cooperativity either, in contrast to biotin binding to (strept)avidin (Rosano *et al.*, 1999). Likewise, if the target ligand cannot fit into the binding pocket,

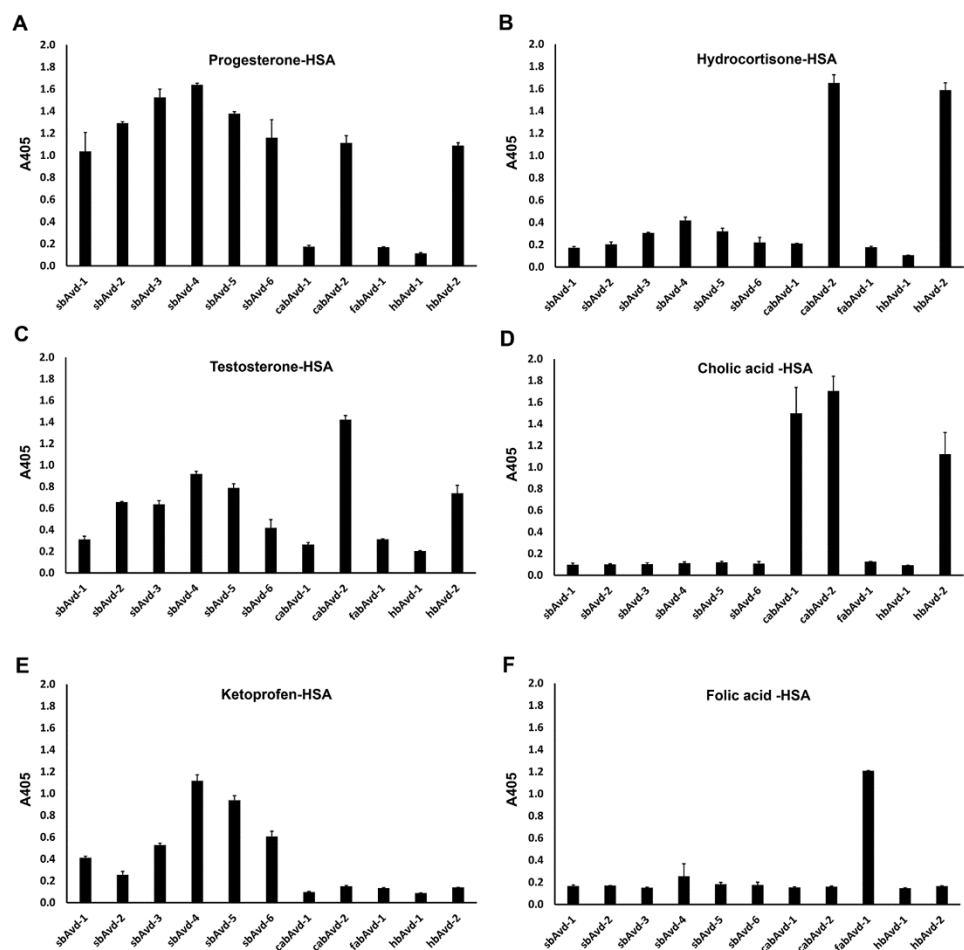


Figure 14. Microplate assay to determine the ligand-binding characteristics of the antidins (using a concentration of $\sim 1 \mu\text{g/ml}$) with human serum albumin (HSA)-conjugated ligands: (A) Progesterone-HSA, (B) hydrocortisone-HSA, (C) testosterone-HSA, (D) cholic acid-HSA, (E) ketoprofen-HSA, and (F) folic acid-HSA. In the graph, bars represent the mean value of two parallel measurements and error bars indicate standard deviation. Figure from III, reprinted with permission from Lehtonen *et al.*, 2016. Copyright 2016 American Chemical Society.

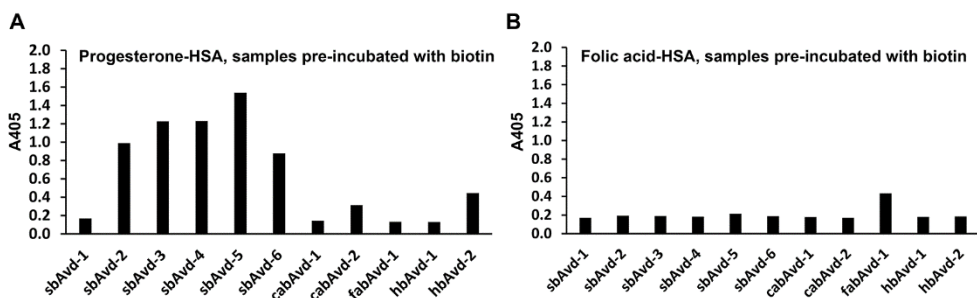


Figure 15. Biotin-inhibition microplate assay. Small molecule ligand binding of antidiins was competed with pre-incubated biotin. In the assay, a protein concentration of $\sim 1 \mu\text{g/ml}$ was used and the protein sample was pre-incubated with $10 \mu\text{M}$ D-biotin (Biochemica, Fluka). The binding of antidiins pre-incubated with biotin was studied on A) progesterone-HSA coated wells and B) folic acid-HSA coated wells. Figure modified from III and reprinted with permission from Lehtonen *et al.*, 2016. Copyright 2016 American Chemical Society.

biotin as the smaller ligand, is able to fit into the ligand-binding pocket and thus inhibit binding to the folic acid-coated surface, as seen in Figure 15. However, among the smaller-sized ligands, hydrophobic progesterone and testosterone ($\Delta T_m = 6.0^\circ\text{C}$ and 4.3°C , respectively) seem to stabilize fabAvid-1 slightly better compared to biotin ($\Delta T_m = 0.7^\circ\text{C}$) (Table 11) (III).

To sum up, despite the total number of mutations in these antidiins being comparable to sbAvids, hbAvids, cabAvids, and fabAvid-1 had significantly lower thermal stabilities (Table 11). Furthermore, the steroids hydrocortisone and cholic acid, conjugated at the opposite ends of their structures, also showed that the conjugation site might have an important role in panning success: An antidin with a stronger affinity was obtained towards the slightly larger cholic acid ($MW_{\text{cholic acid}} 408.57 \text{ g/mol}$ vs. $MW_{\text{hydrocortisone}} 362.46 \text{ g/mol}$, Table 5) compared to those obtained towards hydrocortisone. Therefore, when panning avidin variants against these slightly larger molecules, it might be better to use molecules conjugated at their tail parts, as in the case of cholic acid, thus enabling molecules to hide their hydrophobic steroid structure deep in the binding pocket of avidin.

It is also good to keep in mind that binding on a surface is quite different compared to the situation in a homogeneous solution. As shown already in 2000 by Jung *et al.*, if the ligand-coating density on the surface is high enough to contain two biotin molecules accessible for a single streptavidin tetramer to bind, the dissociation rate becomes much slower: The ratio of the off-rate constant in this situation is approximately the square of the situation where the protein is bound to the surface by a single subunit. This squaring can be explained in that SA is immobilized to the

Table 11. Properties of hydrocortisone-, cholic acid-, and folic acid-binding antidiins. Table modified from III and reprinted with permission from Lehtonen *et al.*, 2016. Copyright 2016 American Chemical Society.

Protein	Ligand	DSC T _m (°C)	Δ T _m (°C)	Fluorometry K _d (nM) ^a
hbAvd-1		64.3 ±6.1 ^b		
	biotin	61.7 ±3.5	-2.6	n/d ^c
	progesterone	81.2 ±4.1	16.9	n/a ^d
	hydrocortisone	70.4 ±5.7	6.2	1513 -
	testosterone	79.9 ±2.7	15.7	n/a
	cholic acid	63.8 ±4.3	-0.5	n/a
hbAvd-2		63.8 ±2.5		
	biotin	67.5 ±1.4	3.7	2401 -
	progesterone	83.6 -	19.8	252 ±30
	hydrocortisone	69.4 ±3.8	5.7	2784 -
	testosterone	77.5 -	13.7	750 -
	cholic acid	70.5 ±2.2	6.8	n/a
cabAvd-1		57.1 ±0.0		
	biotin	57.3 ±0.2	0.2	33761 -
	progesterone	71.2 ±0.3	14.2	1756 -
	hydrocortisone	57.7 ±0.1	0.7	n/a
	testosterone	69.0 -	12.0	2411 -
	cholic acid	65.2 ±0.4	8.1	1700 ±163
cabAvd-2		56.9 ±0.2		
	biotin	58.9 ±0.5	2.0	4722 -
	progesterone	81.3 ±3.4	24.5	114 -
	hydrocortisone	64.9 -	8.0	4400 -
	testosterone	75.2 ±4.8	18.4	105 -
	cholic acid	79.9 ±0.9	23.0	64 -
fabAvd-1		68.1 ±0.6		
	biotin	68.8 ±1.3	0.7	16199 ±461
	progesterone	74.1 ±1.3	6.0	n/a
	hydrocortisone	69.1 ±1.3	1.0	n/d
	testosterone	72.5 -	4.3	n/a
	folic acid	71.8 ±0.6	3.7	188 -
wt Avd		80.8 ±0.1		
	biotin	118.7 ±1.7	37.9	n/d
	progesterone	83.0 ±1.6	2.2	3355 -
	hydrocortisone	82.1 ±0.6	1.3	n/a
	testosterone	82.7 ±0.3	1.8	n/a
	cholic acid	81.5 ±0.3	0.6	n/a
	ketoprofen	82.3 ±0.1	1.5	n/a
	folic acid	81.5 ±0.2	0.6	n/a

^aK_d obtained using 100nM protein concentration, and a fit for tight ligand binding, which accounts for the ligand depletion. In the case of biotin-binding measurements, up to a 350nM protein concentration was used to achieve measurable fluorescence quenching. ^b±SD, standard deviation, is shown, when more than one measurement. ^cn/d, not determined. ^dn/a, the data could not be fitted to the model, or the total fluorescence quenching due to the addition of the ligand was less than 2000 units.

surface by binding two biotin molecules. In case one bond breaks, the other site is still linked to one surface-immobilized biotin (Jung *et al.*, 2000). This phenomenon explains the difference between microplate assay results or SPR measurements, and fluorometric assays. A further difference stems from the need to use a conjugated ligand with the surface-based methods.

5.4 Rational design of point mutations to improve the properties of the selected antidins (III)

In order to improve the properties of the selected antidins as well as to examine the effects of additional mutations, two different point mutations were added to the selected antidins: 1) AVR4-inspired I117Y based on previous results with avidin (Hytönen *et al.*, 2005b) and 2) molecular dynamics simulation based N118D justified more thoroughly below (see section 5.4.2). The point mutations were added to the pET101 expression vectors of antidins using QuikChange-mutagenesis. The properties of these point-mutated antidins were examined using the same methods as with the other antidins.

5.4.1 Antidins with enhanced thermal stability

In order to study the effects of additional mutations to ligand binding of antidins, selected promising variants (sbAvd-2, cabAvd-2, and fabAvd-1) were chosen to be further stabilized by the point mutation I117Y known to strengthen the subunit 1,3-interface, and thus significantly improve the thermal stability of avidin (Hytönen *et al.*, 2004b; Hytönen *et al.*, 2005b). As expected, the I117Y modification resulted in higher T_m values (apo-protein T_m s increased by 15–35 °C) (III, Table 3), and protein yields (2- to 6-fold increase in yields obtained from 0.5L cultures). At the same time, however, the ligand binding properties of these further modified antidins were affected. The point mutation seemed to increase the biotin-binding affinity of those antidins (cabAvd-2 and fabAvd-1) which originally had significantly lower thermal stability compared to wt avidin (K_d 4.7 μ M and 16 μ M \rightarrow 1.5 μ M and 1 μ M, respectively), whereas practically no change was detected in the case of sbAvd-2 (\sim 7 μ M) (III, Table 3). In the case of sbAvd-2 the point mutation mostly affected the steroid-binding affinities: The affinity towards progesterone decreased (111nM \rightarrow 180nM), whereas the affinity towards testosterone increased (\sim 3.0 μ M \rightarrow 1.7 μ M). In the case of cabAvd-2, the affinity decreased for cholic acid (64nM \rightarrow 163nM), and increased for progesterone and testosterone (114nM and 105nM \rightarrow 104nM and 52nM, respectively). In the case of fabAvd-1, surprisingly the affinity towards folic acid increased (188nM \rightarrow 64nM) (III, Table 3).

5.4.2 MD simulations showed the importance of N118 for steroid binding

The 50-ns MD-simulation of sbAvd-1 with bound testosterone molecules suggested two hydrogen bonds formed between testosterone and antidin, where the participating amino acid residues were Asn118 and Ser73 (Figure 13). Since testosterone differs from progesterone only by the functional group at C-17 (Table 5), the point mutation N118D was assumed to have an effect on progesterone binding, as Asp is not capable of forming a hydrogen bond with progesterone 17-O. In order to test this assumption, the point mutation was introduced to steroid-binding avidin variant sbAvd-2.

The N118D resulted in decreased progesterone-binding affinity, as expected, based on the determined K_d for the antidin sbAvd-2(N118D) (913nM) compared to sbAvd-2 (180nM). Additionally, the affinity towards testosterone seemed to increase ($\sim 3.0\mu\text{M} \rightarrow 1.6\mu\text{M}$) (III, Table 3). The N118D mutation showed also to significantly destabilize sbAvd-2: Thermal stability decreased from 81.2 to 65.3 °C (III, Table 3). The Asn118 residue has thus clearly a significant role in steroid binding.

5.4.3 The crystal structure of sbAvd-2(I117Y)

Crystallization trials were performed on sbAvd-2(I117Y), and other sbAvds, both in the absence and presence of steroids aiming to clarify molecular characteristics at the atomic level. Crystals only formed from sbAvd-2(I117Y), from which the ligand-free X-ray structure was determined at 1.95Å resolution (Table 12, Figure 16).

The sbAvd-2(I117Y) [PDB: 4U46] contains two monomers per asymmetric unit, and the biological assembly is clearly tetrameric. The primary structures of sbAvd-2(I117Y) and chicken avidin [PDB:1AVD] (Pugliese *et al.*, 1993) share over 90% similarity in the amino acid sequence, and despite the eight amino acid residue mutations of sbAvd-2(I117Y), the folds are highly similar β -barrel structures. When the C α atoms of sbAvd-2(I117Y) (chain A) and chicken avidin [PDB:1VYO] (chain A) are superimposed, the root-mean-squared deviation (RMSD) is 0.2Å.

The differences in sequence and structure between wt avidin and sbAvd-2(I117Y) are located in the L1,2-loop (mutations D13R, L14M, G15N, S16H), in the L3,4-loop (mutations T35A, A36T, T38N), and at the subunit 1,3-interface of the protein, where the I117Y mutation is located. Unfortunately, the effect of the mutations in the L3,4-loop for ligand binding cannot directly be derived from the determined apo-structure, since the L3,4-loop residues 38–42 were not visible in the X-ray structure.

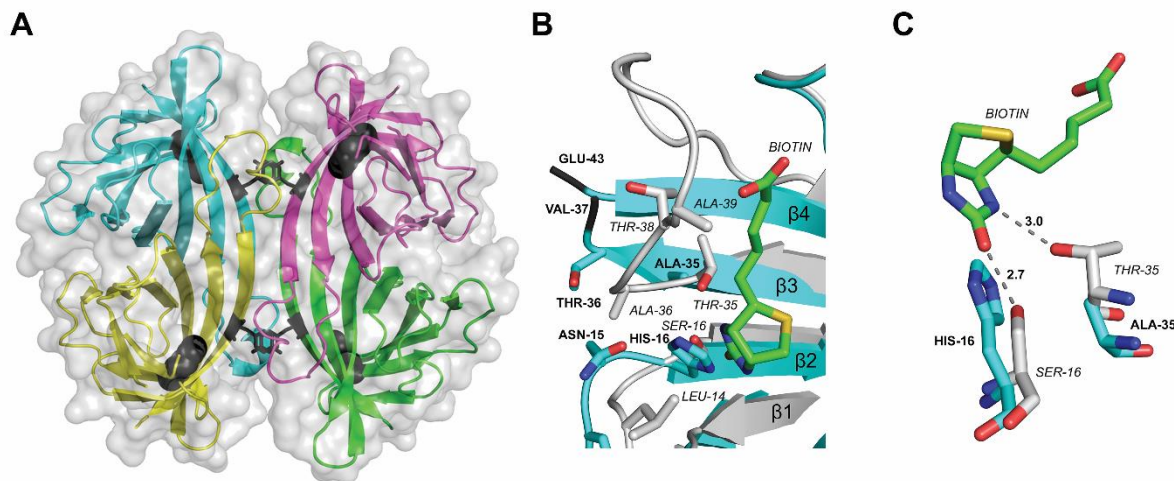


Figure 16. Crystal structure of sbAvd-2(I117Y) [PDB: 4U46]. (A) The quaternary structure of sbAvd-2(I117Y) is shown with a transparent surface. The subunits are shown with different colours: Subunit I in cyan, subunit II in green, subunit III in magenta, and subunit IV in yellow. Tyr117 residues are shown in black sticks at subunit interfaces and His16 residues as black spheres at the bottom of the ligand-binding pockets. (B) Comparison of the selected L1,2- and L3,4-loop residues (of chain A) of sbAvd-2(I117Y) (with cyan carbon atoms and bold labels) and wt Avd [PDB:1AVD] (Pugliese *et al.*, 1993) (with grey carbon atoms and labels in italics) is shown in stick models. Bound biotin of the wt Avd structure is shown also as a stick model (with green carbon atoms). β -sheets 1–4 are numbered. The C α -atoms of Val37 and Glu43 of sbAvd-2(I117Y) are coloured dark grey to indicate the missing region of the L3,4-loop. (C) Comparison of His16 and Ala35 of sbAvd-2(I117Y), and the equivalent residues Ser16 and Thr35 of wt Avd. The hydrogen bonds between biotin and the residues of wt Avd are indicated with dashed lines. Colouring of the residues is as in B. Figure from III, reprinted with permission from Lehtonen *et al.*, 2016. Copyright 2016 American Chemical Society.

This is probably due to high thermal motions of the loop in the absence of ligand, which have been observed repetitively with wt Avd causing L3,4 to be invisible in crystal structures (Livnah *et al.*, 1993; Rosano *et al.*, 1999).

In chicken avidin–biotin complex structures [PDB:2AVI and PDB:1AVD] (Livnah *et al.*, 1993; Pugliese *et al.*, 1993), the L3,4-loop acts as a lid that closes over the biotin-binding site by interacting with the ligand. In these holo-structures, Thr35 forms a hydrogen bond with one of the nitrogen atoms of the bicyclic ring system of biotin (Livnah *et al.*, 1993; Pugliese *et al.*, 1993). Due to the T35A mutation in sbAvd-2, this interaction with biotin is lost. When the sbAvd-2(I117Y) crystal structure is compared to the biotin-complex structure of chicken avidin [PDB:2AVI, 1AVD], the imidazole side chain of His16 is shown to occupy roughly the equivalent space corresponding to the location of the ureido ring oxygen atom of biotin (Livnah *et al.*, 1993; Pugliese *et al.*, 1993). Hence, His16 of sbAvds cannot hydrogen bond to biotin in the same way as the equivalent Ser16 does in the avidin–biotin complex structure.

In the sbAvd-2(I117Y) mutant, the AVR4-inspired mutated residue Tyr117 packs against Tyr117 (π - π stacking) on the 1,3-interface (resembling Tyr115 of AVR4 (Eisenberg-Domovich *et al.*, 2005)) and forms two hydrogen bonds: one with Lys94 of the 1,3-interface and one with a structural water molecule. The π - π stacking makes the Tyr117-residues closely packed together, which causes spatial movement also to the neighboring residues. The Asn118 is shown to be important for progesterone binding, thus the I117Y mutation may explain the decreased affinity of sbAvd-2(I117Y) towards progesterone.

To sum up, the T35A and S16H mutations are probably the most important amino acid changes reshaping the ligand pocket of sbAvd-2(I117Y) and changing its ligand-binding preference compared to wt avidin: Both cause the interactions to be lost with biotin. It has been experimentally shown that the L3,4-loop modification of sbAvd-2 further decreases its biotin-binding affinity compared to sbAvd-1 and improves the affinity for progesterone binding, while it does not have strong effect on testosterone binding. The T35A mutation could serve as an explanation for this phenomenon as Ala35 in sbAvd-2 could favor interactions with the extra methyl group of progesterone not present in the otherwise highly similar testosterone. In this binding mode, the 20-keto group of progesterone (and equivalent group of testosterone) could in turn be involved in an interaction with one of the side chain nitrogen atoms of His16.

Table 12. Structure determination statistics for sbAvd-2(I117Y). Table from III, reprinted with permission from Lehtonen *et al.*, 2016. Copyright 2016 American Chemical Society.

Data processing^a	
Space group	P2 ₁ 2 ₁ 2
Unit cell:	
a, b, c, (Å)	74.47, 79.80, 43.07
α, β, γ (°)	90, 90, 90
Wavelength (Å)	0.87260
Beamline	ID14-1, ESRF
Resolution (Å) ^b	25-1.95 (2.05-1.95)
Observed Reflections ^b	92876 (12995)
Unique Reflections ^b	19264 (2650)
I/sigma ^b	20.32 (5.09)
Rfactor (%) ^b	7.8 (52.4)
Completeness ^b	99.5 (99.9)
Refinement	
Matthews coefficient	2.02
R _{work} (%) ^c	18,10 %
R _{free} (%) ^c	21,10 %
Monomers (asymmetric unit)	2
R.m.s.d:	
Bond lengths (Å)	0,015
Bond angles (°)	1.62

^aThe numbers in parenthesis refer to the highest resolution bin

^bData from XDS (Swillens, 1995)

^cData from Refmac 5 (Tamura *et al.*, 2011)

Recently, the new X-ray structure of progesterone-bound sbAvd-2(I117Y) was determined at 2.80Å resolution [PDB:5LUR] (Agrawal *et al.*, manuscript), still without precise information of the complete L3,4-loop. The L3,4-loop is suspected to stay an open conformation to accommodate the slightly larger progesterone ligand, and cannot thus seal the ligand inside the binding site, as observed in the wt Avd–biotin complex [PDB: 2AVI, 1AVD] (Livnah *et al.*, 1993; Pugliese *et al.*, 1993). However, there is spatial rearrangement of the L1,2 residues to accommodate the incoming progesterone. Met14 forms a hydrophobic interaction with the methyl group of the 20-acetyl group of progesterone. And as speculated, the I117Y-mutation (due to π - π stacking with Tyr117 of another subunit) causes the side chain of Asn118 to move slightly away from the binding site, thus explaining the reduced progesterone-binding affinity.

5.5 Improving the library construction and selection

5.5.1 Construction of new phagemid vectors (II)

In order to construct high-diversity DNA libraries, an efficient and robust subcloning step is essential. Especially in the case of phage display, where the library size is limited by the transformation efficiency, the quality and quantity of the DNA library is critical. Thus, in order to improve the quality of the constructed phage display libraries, we made a Gateway cloning compatible (Katzen, 2007) phagemid vector (pGWphagemid) by replacing its *pelB-Fab* gene of the phagemid vector with the Gateway cassette.

The pGWphagemid vector contains the Gateway cassette, consisting of the *attR*-flanked *ccdB* suicide gene and chloramphenicol resistance gene (Cam^R), upstream of the C-terminus of pIII gene. This construct enables the construction of pIII-phagemid libraries via Gateway cloning. The cognate gene should be *attL*-flanked in order to be cloned in frame with the C-terminal pIII fragment (aa 198–406). The vector was utilized successfully in library construction: Among the sequenced ($n=45$) clones of the DNA library prepared by directly transforming the reassembly reaction into XL1-Blue cells, no empty vectors were found.

Gateway cloning stems on the *ccdB* suicide gene theoretically restricting the use of the methodology for cell lines that do not bear the F' episome. Phage display, on the other hand, requires an *E. coli* strain with the F' episome as this is the requirement for phages to be able to infect cells. However, we showed that it is possible to construct a phage display library using the F' episome bearing *E. coli* XL1-Blue cells. Additionally, the transformation efficiencies of *ccdB*-containing (empty) pGWphagemid vectors were compared to conventional phagemid vector and only $<0.1\%$ transformants were found. Despite the fact that XL1-Blue cells can tolerate low levels of *ccdB*, the negative selection pressure during the following biopanning step ensures the quick removal of *ccdB*-bearing phagemids.

5.5.2 Shortcut in DNA shuffling protocol (II)

The new pGWphagemid vector was tested in the construction of a model library, where two genes encoding for avidin-related proteins AVR2 and AVR4, were DNA shuffled together. The selected AVR2 and AVR4 genes share $\sim 95\%$ DNA

sequence identity and 82% identity in their primary protein structures. Furthermore, their tertiary and quaternary structures resemble each other. However, these proteins differ substantially in their biotin-binding and physicochemical properties, which makes them an interesting pair to be DNA shuffled. AVR2 and AVR4 proteins represent the opposite ends among the AVR family members: AVR2 has the lowest biotin-binding affinity and it has an acidic charge; whereas AVR4 resembles avidin in having basic charge (pI 10) and binds biotin and 2-iminobiotin with almost as high affinity as avidin, as well as being a hyperthermostable protein. Furthermore, the proteins differ with respect to stabilizing interactions of their 1–3 interface: In the case of AVR2, mainly polar interactions are involved, whereas hydrophobic interactions are used in the case of AVR4 (resembling thus avidin) (Hytönen *et al.*, 2005a).

In order to eliminate parental products in the LR cloning reaction ($attL \times attR \rightarrow attB + attP$), the asymmetric templates *attL1-ompA-AVR4* and *AVR2-attL2* (inspired by the article from Ikeuchi *et al.*, 2003) were used as starting materials for the DNA shuffling reaction. At the same time, the conventional DNA shuffling approach was simplified, streamlining the whole approach as gel extractions and amplification steps were omitted and the reassembly product was directly used in the LR cloning reaction. Figure 1 (in II) shows the streamlined approach with the consecutive steps of DNA shuffling. To compare the transformation efficiencies between the streamlined approach versus the conventional approach, transformation tests with highly competent cells were performed. The feasibility of the method was demonstrated in phage display by constructing a model phage library in a two-step process, where the LR cloning reactions were first transformed to *E. coli* DH5 α (in order to achieve zero background) and subsequently transformed into *E. coli* XL1-Blue (for a native phage display library). Table 13 shows the results from these transformation tests where the streamlined Gateway-based DNA shuffling reaction (reassembly reaction with no PCR-amplification) was compared with PCR-amplified reassembly reaction and a gel purified *attL*-flanked positive control. The streamlined Gateway-based DNA shuffling enables a reasonable transformation efficiency of 9.0×10^6 cfu/ μ g, although it is lower than that obtained from controls (1.5×10^8 and 1.8×10^8 cfu/ μ g, respectively).

According to the sequencing results, the constructed model library showed almost only correctly cloned inserts: All of the clones were shuffled at the DNA level, but some of the translated sequences were still identical to the parental proteins, due to the high similarity of the parental genes. The diversity of the library was high with 75% unique clones, and only a low amount of the analyzed clones

Table 13. Quality and characteristics of the DNA-shuffled model libraries based on sequence analysis after transformation tests into *E. coli* DH5α and XL1-Blue cells. Table from II, reprinted by permission from Oxford University Press, copyright 2014.

		Transformation efficiency cfu/μg	Analyzed clones	With shuffled insert	Intact CDS	Unique clones	Clones with point mutations
ElectroMAX [™] DH5α-E [™] Competent cells (Invitrogen)	Nonamplified	9.0 x 10 ⁶	43	43 100%	36 83.7%	33 91.7%	3 8.3%
	Amplified	1.5 x 10 ⁸	44	44 100%	38 86.4%	38 100%	18 47.4%
	attL-flanked AVR4	1.8 x 10 ⁸	2	-	2 100%	-	-
XL1-Blue Electro-Comp Cells (Stratagene)	Nonamplified	5.2 x 10 ⁶	45	45 100%	37 84.4%	33 89.2%	4 10.8%
	Amplified	8.1 x 10 ⁷	46	46 100%	38 82.6%	38 100%	22 57.9%
	attL-flanked AVR4	7.3 x 10 ⁷	2	-	2 100%	-	-

containing additional point mutations. Clones had on average 2.8 crossovers based on translated sequences. Most importantly, the diversity of the library did not seem to be affected by the additional transformation into DH5α. The mutation frequency seemed to be higher in XL1-Blue compared to DH5α, however when the mutation frequencies determined at the protein level per amino acid residue were compared with the previously reported 0.7% mutation frequency per nucleotide (Stemmer, 1994), they were at a similar range.

Due to the small amount of differences between the parental genes, the theoretical library size of the shuffled AVR2/AVR4 library was relatively small (possible unique clones $2^{23} = 8.4 \times 10^6$ based on differing amino acid residues). However, as shown in Table 13, bigger sequence spaces can be easily covered with this approach. Moreover, further diversity through incorporating point mutations can be easily achieved by using a non-proofreading DNA polymerase enzyme, such as Taq, in the amplification step. This was also observed in our study (Table 13). In order to demonstrate the functionality of the modified protocol to construct DNA-shuffled libraries, the phage display method was used to enrich the functional biotin-binding proteins from the model AVR2/AVR4 library. Already after three rounds of panning, a clear enrichment of biotin-binding specific phages was observed and in the microplate assay, clear binding to biotinylated casein-coated wells was detected with only a low background signal.

As a result from the phage display selection, an enrichment of efficiently expressing biotin-binding chimeric AVR mutants was achieved, outcompeting the parental proteins in expression levels. During selection, the frequency of unique clones were reduced as expected from 74.4% in the native library to 18.9%. The demonstration of library panning showed that functional phages can be obtained and enriched from a library based on their properties despite the shortcut in library construction.

To conclude, the described shortcut method is generally compatible with construction of DNA-shuffled libraries. Although the usability of the new Gateway-compatible phagemid vector was demonstrated here with a DNA-shuffled library, it is suitable for the construction of DNA libraries through any approach whether utilizing mutagenesis or recombination. The *attL*-flanked *AVR4* in Table 13 demonstrates the transformation efficiencies make possible for example site-directed random mutagenesis library construction. Moreover, we have been able to reach even larger library sizes later with the same XL1-Blue cells (transformation efficiency 4.8×10^8 cfu/ μ g, data not shown).

5.6 Future plans

Since the publication of the last of the original articles (III), new steps have been taken towards improving antidins: Firstly, in order to circumvent the drawback of the negative selectable marker for theoretically restricting the use of F' episome bearing cells, the new pGWphagemid vector was constructed replacing the original counter selectable marker, the *cadB* gene, with another, *SacBR* (Traore and Zhao, 2011). *SacB* and *SacR* (together noted as *SacBR*) are *Bacillus subtilis*-derived genes that encode the enzyme, levansucrase, involved in sucrose metabolism. In the presense of sucrose, levansucrase synthesizes levan, and as there is no further metabolic pathway for levan in gram-negative bacteria, it accumulates in *E. coli* and becomes toxic (Traore and Zhao, 2011). Furthermore, two new synthetic libraries were designed for progesterone- and hydrocortisone-selections based on the enriched avidin variant sequences (III). The libraries were synthesized utilizing TRIM-technology by GeneArt (ThermoFisher) and subcloned into the new pGWSacBRphagemid. In both libraries, three residues important for biotin binding were selected for total randomization, whereas other residues were only allowed to undergo restricted randomization based on the enriched residues in publication III.

In order to improve the specificities and affinities of the sbAvds, affinity maturation libraries should be constructed (allowing for the sequence of L1,2 to be changed as well) and utilized in the biopanning process, aiming to decrease the crossreactivity by negative selection. One possibility would be to use a strategy successfully used with other protein scaffolds, containing a much higher number of amino acid residues randomized at once. Future libraries could be further improved by the addition of a protease (e.g. trypsin) cleavage site and 6xhistag linking avidin mutant and the C-terminal pIII. The first mentioned would allow for a gentle elution strategy in case high-affinity binders are achieved, whereas histag would enable purifying of the new variants efficiently also when expressed from a phagemid vector.

To combine (strept)avidin technology and novel antidins, scAvd-technology is especially tempting, as it enables the construction of special building blocks with defined affinities in defined subunits of the tetrameric structure.

6 CONCLUSIONS

Changing the binding properties of a high-affinity molecule without altering and thus destabilizing the overall scaffold structure is a challenging task. Avidin however has been studied for decades, and is known to withstand genetic modification quite well, especially when they are targeted in the highly variable loop structures, which are naturally diverse among the avidin gene family (Laitinen *et al.*, 2006). The natural ligand of avidin, biotin, differs significantly from steroids, such as progesterone, for which antidins with the best affinities were achieved: Biotin is a water-soluble vitamin composed of an ureido ring fused with a tetrahydrothiophene ring attached to a valeryl side chain, whereas the steroid hormones used in these studies are cholesterol-derived hydrophobic molecules with only a few functional groups attached to their steroid backbone (Figure 12, Table 5).

Much has happened over a decade of avidin engineering through directed evolution and phage display. As soon as the first phages expressing avidin on their surface were constructed and their functionality verified, construction of the first avidin-based libraries was started (I). The Avd libraries constructed with saturation mutagenesis (Avd L1,2 and AvLib1–3) were based on the sole fusion of Avd with c-terminal pIII (I, III). Even though the oligomeric constructs could have been more efficient in selection, the theoretical monovalency of the sole Avd-pIII fusion might have enabled higher affinity binders to be selected, since there is less of the avidity effect caused by the oligomeric Avd. During these years, the entire methodological system has been improved: Library construction was enhanced through modification of the used phagemid vector to enable more efficient construction of bigger phage display libraries. Additionally different screening methods for avidin variants have been established. We also generated a shortcut method for the construction of DNA-shuffled libraries, which both saves the time required for the library construction, and more importantly, preserves the high quality of the library (II). Furthermore, panning with a magnetic bead processor combined with small-scale phage production (Turunen *et al.*, 2009) enables more efficient selection as several different conditions tested in parallel.

The first constructed avidin phage display library, Avd L1,2, was tested against several small molecule targets, from which testosterone was the only one yielding

promising results. At that time, we did not know wt Avd had an intrinsic affinity to progesterone, which could have been useful to know, and probably explains the panning success with testosterone. The intrinsic affinity of wt Avd towards progesterone probably played a major role in testosterone-panning combined with the small-sized Avd L1,2 library (containing changes only in L1,2). Speculatively, the obtained library size could have reached a much better candidate to form the basis for the affinity maturation process, simply with a bit wiser library construction scheme.

Comparing the complexities of the libraries to other successful protein scaffolds, the whole avidin library construction scheme is still in its infancy: For example in anticalin libraries, 16–24 residues are typically randomized. We have been able to show however, that the ligand-binding specificity of avidin can be changed. From the pooled avidin libraries (AvLib1–3), enriched variants were obtained towards six out of seven small molecule targets showing micro- to low micromolar affinities (III). Maybe some day, one of the antidinins will be utilized in an application for medicine or the bioindustry.

From the developed antidinins, sbAvds are possibly the most promising ones at the moment. They show that as little as a 7–8 amino acid residue change can result in a clear difference in ligand specificity, while retaining the high thermal stability of avidin ($T_m \sim 80^\circ\text{C}$). However, the affinities (and the specificity) towards progesterone are not high enough to compete with antibodies on the market. Interestingly, compared to sbAvds, the cabAvd-2 shows how one can reach a similar affinity to progesterone by combining different amino acid residues in the loop regions important for ligand binding. The change of these nine amino acids however, significantly reduced the thermal stability of avidin, but stability could be retained and actually even improved with one additional point mutation (I117Y). Moreover, this point mutation seemed to improve steroid affinity as well. This antidin could be a promising scaffold to form the basis for affinity maturation, based on its affinities towards several steroids.

Hydrocortisone and cholic acid are steroids as well, although slightly larger molecules compared to progesterone and testosterone. As seen from the results, the enriched avidin variants binding hydrocortisone and/or cholic acid have cross-reactivity towards these smaller-sized steroids. Interestingly, these targets enriched the avidin variants from the AvLib-2 library but not from AvLib-3. This can be an indication of several things: 1) The sequence of sbAvd-1 (D13R, L14M, G15N, S16H) in L1,2 is probably not the best possible to form the basis for the high-affinity steroid-binder. Not even in the case of cholic acid, which resembled more

testosterone as a conjugated ligand than hydrocortisone, since its conjugation was done at the opposite end of the steroid structure compared to most of the other steroid-conjugates (conjugated from C-3, see Figure 12). This probably explains the high affinity of cabAvd-2 also towards testosterone (and progesterone). 2) The unmodified L3,4 might be suitable as such for steroids already intrinsically containing hydrophobic alanine and valine residues. 3) The slightly larger target molecules seem to favor modifications in L5,6-loop towards a more hydrophobic environment.

One parameter to consider when assessing the potential of avidin as an alternative scaffold are the limits in the size of a target that avidin can accommodate. Folic acid is clearly already too big. This could explain the inconsistency between the obtained K_d value and modest thermal stabilizing effect, as well as the result that preincubation of fabAvd-1 with biotin could decrease the binding of the protein on a folic acid-coated surface. Based on its molecular weight ($M_{\text{folic acid}}=441.4$ g/mol) compared to biotin ($M_{\text{biotin}}=244.3$ g/mol), it is unlikely that folic acid can completely fit into the binding site of fabAvd-1. Interestingly, cholic acid and hydrocortisone differ only 46 Da from each other in their molecular mass, cholic acid being a slightly bigger molecule. However, binders were more successfully obtained for cholic acid, probably based on the orientation used in the panning. When the ligand-binding pocket of avidin is considered, it is probably better that the hydrophobic portion of the molecule goes in first, as in the case of biotin. Basically the size of all the small molecules used here as targets (I, III) exceeds the size of the original biotin-binding pocket of avidin (242 \AA^3). When this is compared to the size of the ligand-binding pocket of steroid-hormone receptors, usually around 445 \AA^3 (Li *et al.*, 2005), there is a significant size difference. With such modest modifications that we have performed so far, the ligand-binding pocket size is not able to be changed radically.

Well over 30 different protein scaffolds with different topologies and folds have been investigated as alternatives to antibodies (Binz and Plückthun, 2005). They offer a large set of options to be used in different applications ranging from affinity chromatography to tissue staining, and diagnostic applications to therapeutic applications. Avidin-scaffold based antidins would have potential in applications where highly stable protein structures are required (e.g. in sample concentration and point-of-care diagnostic tests). Rarely are protein scaffolds as stable as avidin, including both thermal and chemical stability. Even though the affinities and specificities of these newly developed antidins are still inadequate to compete with antibodies in for example diagnostic tests based on their high stability, they might still be useful in sample preparation, for example to concentrate steroids from a sample. ScAvd-technology also provides interesting possibilities, since the

development of multivalent molecules would be beneficial in enabling focused orientation of other molecules.

Interestingly, avidin was first discovered when studying the hormone function of chick oviduct and possible gene regulation. Avidin was found to be expressed by progesterone (Korenman and O'Malley, 1968). Later, during the characterization of novel antidins, we found that avidin has intrinsic affinity towards progesterone ($K_d \sim 3\mu\text{M}$; III). This may indicate some kind of feedback loop, where avidin negatively regulates its expression. To our knowledge, no one has studied this yet. Recently, novel avidins from simple animals have also been observed to have progesterone-regulated expression (Guo *et al.*, 2017). It would be interesting to study, whether they share the intrinsic low affinity for progesterone as well.

7 REFERENCES

- Adams, P.D., R.W. Grosse-Kunstleve, L.W. Hung, T.R. Ioerger, A.J. McCoy, N.W. Moriarty, R.J. Read, J.C. Sacchettini, N.K. Sauter, and T.C. Terwilliger. 2002. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr.D Biol.Crystallogr.* 58:1948-1954.
- Agrawal, N., S.I. Lehtonen, M. Uusi-Mäkelä, P. Jain, T.A. Riihimäki, V.P. Hytönen, M.S. Kulomaa, M.S. Johnson, and T.T. Airenne. Manuscript. Docking and X-ray studies on the binding of small molecules to antidiins.
- Airenne, K.J., C. Oker-Blom, V.S. Marjomaki, E.A. Bayer, M. Wilchek, and M.S. Kulomaa. 1997. Production of biologically active recombinant avidin in baculovirus-infected insect cells. *Protein Expr.Purif.* 9:100-108.
- Argarana, C.E., I.D. Kuntz, S. Birken, R. Axel, and C.R. Cantor. 1986. Molecular cloning and nucleotide sequence of the streptavidin gene. *Nucleic Acids Res.* 14:1871-1882.
- Avraham, O., A. Meir, A. Fish, E.A. Bayer, and O. Livnah. 2015. Hoefavidin: A dimeric bacterial avidin with a C-terminal binding tail. *J.Struct.Biol.* 191:139-148.
- Baek, H., K.H. Suk, Y.H. Kim, and S. Cha. 2002. An improved helper phage system for efficient isolation of specific antibody molecules in phage display. *Nucleic Acids Res.* 30:e18.
- Banta, S., K. Dooley, and O. Shur. 2013. Replacing antibodies: engineering new binding proteins. *Annu.Rev.Biomed.Eng.* 15:93-113.
- Barbas 3rd, C.F., D.R. Burton, J.K. Scott, and G.J. Silverman. 2001. Phage display: a laboratory manual.
- Bass, S., R. Greene, and J.A. Wells. 1990. Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins.* 8:309-314.
- Bayer, E.A., T. Kulik, R. Adar, and M. Wilchek. 1995. Close similarity among streptavidin-like, biotin-binding proteins from *Streptomyces*. *Biochim.Biophys.Acta.* 1263:60-66.
- Berman, H.M., T. Battistuz, T.N. Bhat, W.F. Bluhm, P.E. Bourne, K. Burkhardt, Z. Feng, G.L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J.D. Westbrook, and C. Zardecki. 2002. The Protein Data Bank. *Acta Crystallogr.D Biol.Crystallogr.* 58:899-907.
- Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235-242.
- Beste, G., F.S. Schmidt, T. Stibora, and A. Skerra. 1999. Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *Proc.Natl.Acad.Sci.U.S.A.* 96:1898-1903.
- Binz, H.K., P. Amstutz, A. Kohl, M.T. Stumpp, C. Briand, P. Forrer, M.G. Grutter, and A. Plückthun. 2004. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat.Biotechnol.* 22:575-582.
- Binz, H.K., P. Amstutz, and A. Plückthun. 2005. Engineering novel binding proteins from nonimmunoglobulin domains. *Nat.Biotechnol.* 23:1257-1268.

- Binz, H.K., and A. Plückthun. 2005. Engineered proteins as specific binding reagents. *Curr.Opin.Biotechnol.* 16:459-469.
- Binz, H.K., M.T. Stumpp, P. Forrer, P. Amstutz, and A. Plückthun. 2003. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J.Mol.Biol.* 332:489-503.
- Boder, E.T., and K.D. Wittrup. 1997. Yeast surface display for screening combinatorial polypeptide libraries. *Nat.Biotechnol.* 15:553-557.
- Bork, P. 1993. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins.* 17:363-374.
- Bork, P., and R.F. Doolittle. 1992. Proposed acquisition of an animal protein domain by bacteria. *Proc.Natl.Acad.Sci.U.S.A.* 89:8990-8994.
- Bratkovc, T. 2010. Progress in phage display: evolution of the technique and its application. *Cell Mol.Life Sci.* 67:749-767.
- Breustedt, D.A., I.P. Korndörfer, B. Redl, and A. Skerra. 2005. The 1.8-Å crystal structure of human tear lipocalin reveals an extended branched cavity with capacity for multiple ligands. *J.Biol.Chem.* 280:484-493.
- Breustedt, D.A., D.L. Schönfeld, and A. Skerra. 2006. Comparative ligand-binding analysis of ten human lipocalins. *Biochim.Biophys.Acta.* 1764:161-173.
- Butt, H., and M. Jaschke. 1995. Calculation of thermal noise in atomic force microscopy. *Nanotechnology.* 6:1-7.
- Cadwell, R.C., and G.F. Joyce. 1992. Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* 2:28-33.
- Cesareni, G. 1988. Phage-plasmid hybrid vectors. *Biotechnology.* 10:103-111.
- Chalet, L., and F.J. Wolf. 1964. The Properties of Streptavidin, a Biotin-Binding Protein Produced by Streptomyces. *Arch.Biochem.Biophys.* 106:1-5.
- Chilkoti, A., P.H. Tan, and P.S. Stayton. 1995. Site-directed mutagenesis studies of the high-affinity streptavidin-biotin complex: contributions of tryptophan residues 79, 108, and 120. *Proc.Natl.Acad.Sci.U.S.A.* 92:1754-1758.
- Chinol, M., P. Casalini, M. Maggiolo, S. Canevari, E.S. Omodeo, P. Caliceti, F.M. Veronese, M. Cremonesi, F. Chiolerio, E. Nardone, A.G. Siccardi, and G. Paganelli. 1998. Biochemical modifications of avidin improve pharmacokinetics and biodistribution, and reduce immunogenicity. *Br.J.Cancer.* 78:189-197.
- Coco, W.M., W.E. Levinson, M.J. Crist, H.J. Hektor, A. Darzins, P.T. Pienkos, C.H. Squires, and D.J. Monticello. 2001. DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat.Biotechnol.* 19:354-359.
- Collaborative Computational Project, Number 4. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.D Biol.Crystallogr.* 50:760-763.
- Collis, A.V., A.P. Brouwer, and A.C. Martin. 2003. Analysis of the antigen combining site: correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. *J.Mol.Biol.* 325:337-354.
- Cramer, A., S.A. Raillard, E. Bermudez, and W.P. Stemmer. 1998. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature.* 391:288-291.
- Davis, I.W., A. Leaver-Fay, V.B. Chen, J.N. Block, G.J. Kapral, X. Wang, L.W. Murray, W.B. Arendall 3rd, J. Snoeyink, J.S. Richardson, and D.C. Richardson. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35:W375-83.

- DeLange, R.J., and T.S. Huang. 1971. Egg white avidin. 3. Sequence of the 78-residue middle cyanogen bromide peptide. Complete amino acid sequence of the protein subunit. *J.Biol.Chem.* 246:698-709.
- Dower, W.J., J.F. Miller, and C.W. Ragsdale. 1988. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* 16:6127-6145.
- Efimov, V.P., I.V. Nepluev, and V.V. Mesyanzhinov. 1995. Bacteriophage T4 as a surface display vector. *Virus Genes.* 10:173-177.
- Eggertsson, G., and D. Söll. 1988. Transfer ribonucleic acid-mediated suppression of termination codons in *Escherichia coli*. *Microbiol.Rev.* 52:354-374.
- Eigenbrot, C., M. Ultsch, A. Dubnovitsky, L. Abrahmsen, and T. Härd. 2010. Structural basis for high-affinity HER2 receptor binding by an engineered protein. *Proc.Natl.Acad.Sci.U.S.A.* 107:15039-15044.
- Eisenberg-Domovich, Y., V.P. Hytönen, M. Wilchek, E.A. Bayer, M.S. Kulomaa, and O. Livnah. 2005. High-resolution crystal structure of an avidin-related protein: insight into high-affinity biotin binding and protein stability. *Acta Crystallogr.D Biol.Crystallogr.* 61:528-538.
- Ellison, D., J. Hinton, S.J. Hubbard, and R.J. Beynon. 1995. Limited proteolysis of native proteins: the interaction between avidin and proteinase K. *Protein Sci.* 4:1337-1345.
- Elo, H.A., O. Jänne, and P.J. Tuohimaa. 1980a. Avidin induction in the chick oviduct by progesterone and non-hormonal treatments. *J.Steroid Biochem.* 12:279-281.
- Elo, H.A., M.S. Kulomaa, and P.J. Tuohimaa. 1979a. Avidin induction by tissue injury and inflammation in male and female chickens. *Comp.Biochem.Physiol.B.* 62:237-240.
- Elo, H.A., M.S. Kulomaa, and P.J. Tuohimaa. 1979b. Progesterone-independent avidin induction in chick tissues caused by tissue injury and inflammation. *Acta Endocrinol.(Copenh).* 90:743-752.
- Elo, H.A., S. Räisänen, and P.J. Tuohimaa. 1980b. Induction of an antimicrobial biotin-binding egg white protein (avidin) in chick tissues in septic *Escherichia coli* infection. *Experientia.* 36:312-313.
- Emsley, P., and K. Cowtan. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr.D Biol.Crystallogr.* 60:2126-2132.
- Ernst, W., R. Grabherr, D. Wegner, N. Borth, A. Grassauer, and H. Katinger. 1998. Baculovirus surface display: construction and screening of a eukaryotic epitope library. *Nucleic Acids Res.* 26:1718-1723.
- Evans, J.B., and B.A. Syed. 2014. From the analyst's couch: Next-generation antibodies. *Nat.Rev.Drug Discov.* 13:413-414.
- Fellouse, F.A., P.A. Barthelémy, R.F. Kelley, and S.S. Sidhu. 2006. Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code. *J.Mol.Biol.* 357:100-114.
- Fellouse, F.A., B. Li, D.M. Compaan, A.A. Peden, S.G. Hymowitz, and S.S. Sidhu. 2005. Molecular recognition by a binary code. *J.Mol.Biol.* 348:1153-1162.
- Fellouse, F.A., C. Wiesmann, and S.S. Sidhu. 2004. Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc.Natl.Acad.Sci.U.S.A.* 101:12467-12472.
- Firth, A.E., and W.M. Patrick. 2008. GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. *Nucleic Acids Res.* 36:W281-5.
- Flower, D.R. 1993. Structural relationship of streptavidin to the calycin protein superfamily. *FEBS Lett.* 333:99-102.

- Flower, D.R., A.C. North, and C.E. Sansom. 2000. The lipocalin protein family: structural and sequence overview. *Biochim.Biophys.Acta.* 1482:9-24.
- Forrer, P., M.T. Stumpp, H.K. Binz, and A. Plückthun. 2003. A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Lett.* 539:2-6.
- Freitag, S., V. Chu, J.E. Penzotti, L.A. Klumb, R. To, D. Hyre, I. Le Trong, T.P. Lybrand, R.E. Stenkamp, and P.S. Stayton. 1999. A structural snapshot of an intermediate on the streptavidin-biotin dissociation pathway. *Proc.Natl.Acad.Sci.U.S.A.* 96:8384-8389.
- Gebauer, M., A. Schiefner, G. Matschiner, and A. Skerra. 2013. Combinatorial design of an Anticalin directed against the extra-domain b for the specific targeting of oncofetal fibronectin. *J.Mol.Biol.* 425:780-802.
- Gebauer, M., and A. Skerra. 2012. Anticalins small engineered binding proteins based on the lipocalin scaffold. *Methods Enzymol.* 503:157-188.
- Gentry-Weeks, C.R., A.L. Hultsch, S.M. Kelly, J.M. Keith, and R. Curtiss 3rd. 1992. Cloning and sequencing of a gene encoding a 21-kilodalton outer membrane protein from *Bordetella avium* and expression of the gene in *Salmonella typhimurium*. *J.Bacteriol.* 174:7729-7742.
- Gilbreth, R.N., K. Esaki, A. Koide, S.S. Sidhu, and S. Koide. 2008. A dominant conformational role for amino acid diversity in minimalist protein-protein interfaces. *J.Mol.Biol.* 381:407-418.
- Goetz, D.H., M.A. Holmes, N. Borregaard, M.E. Bluhm, K.N. Raymond, and R.K. Strong. 2002. The neutrophil lipocalin NGAL is a bacteriostatic agent that interferes with siderophore-mediated iron acquisition. *Mol.Cell.* 10:1033-1043.
- Gonzalez, M., C.E. Argarana, and G.D. Fidelio. 1999. Extremely high thermal stability of streptavidin and avidin upon biotin binding. *Biomol.Eng.* 16:67-72.
- Green, N.M. 1990. Avidin and streptavidin. *Methods Enzymol.* 184:51-67.
- Green, N.M. 1975. Avidin. *Adv.Protein Chem.* 29:85-133.
- Green, N.M. 1965. A Spectrophotometric Assay for Avidin and Biotin Based on Binding of Dyes by Avidin. *Biochem.J.* 94:23C-24C.
- Green, N.M. 1963a. Avidin. 1. the use of (14-C)biotin for Kinetic Studies and for Assay. *Biochem.J.* 89:585-591.
- Green, N.M. 1963b. Avidin. 3. the Nature of the Biotin-Binding Site. *Biochem.J.* 89:599-609.
- Greenberg, A.S., D. Avila, M. Hughes, A. Hughes, E.C. McKinney, and M.F. Flajnik. 1995. A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature.* 374:168-173.
- Gunneriusson, E., K. Nord, M. Uhlén, and P. Nygren. 1999. Affinity maturation of a Taq DNA polymerase specific affibody by helix shuffling. *Protein Eng.* 12:873-878.
- Guo, X., J. Xin, P. Wang, X. Du, G. Ji, Z. Gao, and S. Zhang. 2017. Functional characterization of avidins in amphioxus *Branchiostoma japonicum*: Evidence for a dual role in biotin-binding and immune response. *Dev.Comp.Immunol.* 70:106-118.
- Hackel, B.J., A. Kapila, and K.D. Wittrup. 2008. Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling. *J.Mol.Biol.* 381:1238-1252.
- Hackel, B.J., and K.D. Wittrup. 2010. The full amino acid repertoire is superior to serine/tyrosine for selection of high affinity immunoglobulin G binders from the fibronectin scaffold. *Protein Eng.Des.Sel.* 23:211-219.

- Hamers-Casterman, C., T. Atarhouch, S. Muyldermans, G. Robinson, C. Hamers, E.B. Songa, N. Bendahman, and R. Hamers. 1993. Naturally occurring antibodies devoid of light chains. *Nature*. 363:446-448.
- Hanes, J., and A. Plückthun. 1997. In vitro selection and evolution of functional proteins by using ribosome display. *Proc.Natl.Acad.Sci.U.S.A.* 94:4937-4942.
- Hayouka, R., Y. Eisenberg-Domovich, V.P. Hytönen, J.A. Määttä, H.R. Nordlund, M.S. Kulomaa, M. Wilchek, E.A. Bayer, and O. Livnah. 2008. Critical importance of loop conformation to avidin-enhanced hydrolysis of an active biotin ester. *Acta Crystallogr.D Biol.Crystallogr.* 64:302-308.
- Heikkinen, J.J., L. Kivimäki, J.A. Määttä, I. Mäkelä, L. Hakalahti, K. Takkinen, M.S. Kulomaa, V.P. Hytönen, and O.E. Hormi. 2011. Versatile bio-ink for covalent immobilization of chimeric avidin on sol-gel substrates. *Colloids Surf.B Biointerfaces*. 87:409-414.
- Helpolainen, S.H., J.A. Määttä, K.K. Halling, J.P. Slotte, V.P. Hytönen, J. Jänis, P. Vainiotalo, M.S. Kulomaa, and H.R. Nordlund. 2008. Bradavidin II from *Bradyrhizobium japonicum*: a new avidin-like biotin-binding protein. *Biochim.Biophys.Acta*. 1784:1002-1010.
- Helpolainen, S.H., K.P. Nurminen, J.A. Määttä, K.K. Halling, J.P. Slotte, T. Huhtala, T. Liimatainen, S. Ylä-Herttuala, K.J. Airenne, A. Närvänen, J. Jänis, P. Vainiotalo, J. Valjakka, M.S. Kulomaa, and H.R. Nordlund. 2007. Rhizavidin from *Rhizobium etli*: the first natural dimer in the avidin protein family. *Biochem.J.* 405:397-405.
- Hendrickson, W.A., A. Pähler, J.L. Smith, Y. Satow, E.A. Merritt, and R.P. Phizackerley. 1989. Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proc.Natl.Acad.Sci.U.S.A.* 86:2190-2194.
- Hertz, R., R.M. Fraps, and W.H. Sebrell. 1943. Induction of Avidin Formation in the Avian Oviduct by Stilbestrol Plus Progesterone. 52.
- Hertz, R., and W.H. Sebrell. 1942. Occurrence of Avidin in the Oviduct and Secretions of the Genital Tract of several Species. *Science*. 96:257.
- Hey, T., E. Fiedler, R. Rudolph, and M. Fiedler. 2005. Artificial, non-antibody binding proteins for pharmaceutical and industrial applications. *Trends Biotechnol.* 23:514-522.
- Hiller, Y., E.A. Bayer, and M. Wilchek. 1991. Studies on the biotin-binding site of avidin. Minimized fragments that bind biotin. *Biochem.J.* 278 (Pt 2):573-585.
- Hiller, Y., J.M. Gershoni, E.A. Bayer, and M. Wilchek. 1987. Biotin binding to avidin. Oligosaccharide side chain not required for ligand association. *Biochem.J.* 248:167-171.
- Hoogenboom, H.R. 2005. Selecting and screening recombinant antibody libraries. *Nat.Biotechnol.* 23:1105-1116.
- Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*. 65:712-725.
- Hosse, R.J., A. Rothe, and B.E. Power. 2006. A new generation of protein display scaffolds for molecular recognition. *Protein Sci.* 15:14-27.
- Hudson, P.J., and C. Souriau. 2003. Engineered antibodies. *Nat.Med.* 9:129-134.
- Hughes, M.D., D.A. Nagel, A.F. Santos, A.J. Sutherland, and A.V. Hine. 2003. Removing the redundancy from randomised gene libraries. *J.Mol.Biol.* 331:973-979.

- Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J.Mol.Graph.* 14:33-8, 27-8.
- Hutter, J.L., and J. Bechhoefer. 1993. Calibration of atomic-force microscope tips *Rev. Sci. Instrum.* 64 (7):1868-1873.
- Hyre, D.E., L.M. Amon, J.E. Penzotti, I. Le Trong, R.E. Stenkamp, T.P. Lybrand, and P.S. Stayton. 2002. Early mechanistic events in biotin dissociation from streptavidin. *Nat.Struct.Biol.* 9:582-585.
- Hyre, D.E., I. Le Trong, E.A. Merritt, J.F. Eccleston, N.M. Green, R.E. Stenkamp, and P.S. Stayton. 2006. Cooperative hydrogen bond interactions in the streptavidin-biotin system. *Protein Sci.* 15:459-467.
- Hytönen, V.P., J. Hörhå, T.T. Airene, E.A. Niskanen, K.J. Helttunen, M.S. Johnson, T.A. Salminen, M.S. Kulomaa, and H.R. Nordlund. 2006. Controlling quaternary structure assembly: subunit interface engineering and crystal structure of dual chain avidin. *J.Mol.Biol.* 359:1352-1363.
- Hytönen, V.P., O.H. Laitinen, T.T. Airene, H. Kidron, N.J. Meltola, E.J. Porkka, J. Hörhå, T. Paldanius, J.A. Määttä, H.R. Nordlund, M.S. Johnson, T.A. Salminen, K.J. Airene, S. Ylä-Herttuala, and M.S. Kulomaa. 2004a. Efficient production of active chicken avidin using a bacterial signal peptide in *Escherichia coli*. *Biochem.J.* 384:385-390.
- Hytönen, V.P., O.H. Laitinen, A. Grapputo, A. Kettunen, J. Savolainen, N. Kalkkinen, A.T. Marttila, H.R. Nordlund, T.K. Nyholm, G. Paganelli, and M.S. Kulomaa. 2003. Characterization of poultry egg-white avidins and their potential as a tool in pretargeting cancer treatment. *Biochem.J.* 372:219-225.
- Hytönen, V.P., J.A. Määttä, H. Kidron, K.K. Halling, J. Hörhå, T. Kulomaa, T.K. Nyholm, M.S. Johnson, T.A. Salminen, M.S. Kulomaa, and T.T. Airene. 2005a. Avidin related protein 2 shows unique structural and functional features among the avidin protein family. *BMC Biotechnol.* 5:28.
- Hytönen, V.P., J.A. Määttä, T.K. Nyholm, O. Livnah, Y. Eisenberg-Domovich, D. Hyre, H.R. Nordlund, J. Hörhå, E.A. Niskanen, T. Paldanius, T. Kulomaa, E.J. Porkka, P.S. Stayton, O.H. Laitinen, and M.S. Kulomaa. 2005b. Design and construction of highly stable, protease-resistant chimeric avidins. *J.Biol.Chem.* 280:10228-10233.
- Hytönen, V.P., H.R. Nordlund, J. Hörhå, T.K. Nyholm, D.E. Hyre, T. Kulomaa, E.J. Porkka, A.T. Marttila, P.S. Stayton, O.H. Laitinen, and M.S. Kulomaa. 2005c. Dual-affinity avidin molecules. *Proteins.* 61:597-607.
- Hytönen, V.P., T.K. Nyholm, O.T. Pentikäinen, J. Vaarno, E.J. Porkka, H.R. Nordlund, M.S. Johnson, J.P. Slotte, O.H. Laitinen, and M.S. Kulomaa. 2004b. Chicken avidin-related protein 4/5 shows superior thermal stability when compared with avidin while retaining high affinity to biotin. *J.Biol.Chem.* 279:9337-9343.
- Ikeuchi, A., Y. Kawarasaki, T. Shinbata, and T. Yamane. 2003. Chimeric gene library construction by a simple and highly versatile method using recombination-dependent exponential amplification. *Biotechnol.Prog.* 19:1460-1467.
- Jain, M., N. Kamal, and S.K. Batra. 2007. Engineering antibodies for clinical applications. *Trends Biotechnol.* 25:307-316.
- Jijakli, K., B. Khraiwesh, W. Fu, L. Luo, A. Alzahmi, J. Koussa, A. Chaiboonchoe, S. Kirmizialtin, L. Yen, and K. Salehi-Ashtiani. 2016. The in vitro selection world. *Methods.* 106:3-13.
- Jones, G., P. Willett, R.C. Glen, A.R. Leach, and R. Taylor. 1997. Development and validation of a genetic algorithm for flexible docking. *J.Mol.Biol.* 267:727-748.
- Jones, P.D., and M.H. Briggs. 1962. The distribution of avidin. *Life.Sci.(1962).* 1:621-623.

- Jost, C., and A. Plückthun. 2014. Engineered proteins with desired specificity: DARPin, other alternative scaffolds and bispecific IgGs. *Curr.Opin.Struct.Biol.* 27:102-112.
- Jung, L.S., K.E. Nelson, P.S. Stayton, and C.T. Campbell. 2000. Binding and Dissociation Kinetics of Wild-Type and Mutant Streptavidins on Mixed Biotin-Containing Alkylthiolate Monolayers. *Langmuir.* 16 (24):9421-9432.
- Kabsch, W. 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography.* 26:795-800.
- Kamruzzahan, A.S., F. Kienberger, C.M. Stroh, J. Berg, R. Huss, A. Ebner, R. Zhu, C. Rankl, H.J. Gruber, and P. Hinterdorfer. 2004. Imaging morphological details and pathological differences of red blood cells using tapping-mode AFM. *Biol.Chem.* 385:955-960.
- Katzen, F. 2007. Gateway(R) recombinational cloning: a biological operating system.
- Kawarasaki, Y., K.E. Griswold, J.D. Stevenson, T. Selzer, S.J. Benkovic, B.L. Iverson, and G. Georgiou. 2003. Enhanced crossover SCRATCHY: construction and high-throughput screening of a combinatorial library containing multiple non-homologous crossovers. *Nucleic Acids Res.* 31:e126.
- Kille, S., C.G. Acevedo-Rocha, L.P. Parra, Z.G. Zhang, D.J. Opperman, M.T. Reetz, and J.P. Acevedo. 2013. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth.Biol.* 2:83-92.
- Kingsbury, G.A., and R.P. Junghans. 1995. Screening of phage display immunoglobulin libraries by anti-M13 ELISA and whole phage PCR. *Nucleic Acids Res.* 23:2563-2564.
- Klumb, L.A., V. Chu, and P.S. Stayton. 1998. Energetic roles of hydrogen bonds at the ureido oxygen binding pocket in the streptavidin-biotin complex. *Biochemistry.* 37:7657-7663.
- Koide, A., C.W. Bailey, X. Huang, and S. Koide. 1998. The fibronectin type III domain as a scaffold for novel binding proteins. *J.Mol.Biol.* 284:1141-1151.
- Koide, A., R.N. Gilbreth, K. Esaki, V. Tereshko, and S. Koide. 2007. High-affinity single-domain binding proteins with a binary-code interface. *Proc.Natl.Acad.Sci.U.S.A.* 104:6632-6637.
- Koide, A., J. Wojcik, R.N. Gilbreth, R.J. Hoey, and S. Koide. 2012. Teaching an old scaffold new tricks: monobodies constructed using alternative surfaces of the FN3 scaffold. *J.Mol.Biol.* 415:393-405.
- Korenman, S.G., and B.W. O'Malley. 1968. Progesterone action: regulation of avidin biosynthesis by hen oviduct in vivo and in vitro. *Endocrinology.* 83:11-17.
- Korpela, J.K., M.S. Kulomaa, H.A. Elo, and P.J. Tuohimaa. 1981. Biotin-binding proteins in eggs of oviparous vertebrates. *Experientia.* 37:1065-1066.
- Kramer, K.J., T.D. Morgan, J.E. Throne, F.E. Dowell, M. Bailey, and J.A. Howard. 2000. Transgenic avidin maize is resistant to storage insect pests. *Nat.Biotechnol.* 18:670-674.
- Kulomaa, M.S., H.A. Elo, and P.J. Tuohimaa. 1978. A radioimmunoassay for chicken avidin. Comparison with a [¹⁴C]biotin-binding method. *Biochem.J.* 175:685-690.
- Kunnas, T.A., M.J. Wallén, and M.S. Kulomaa. 1993. Induction of chicken avidin and related mRNAs after bacterial infection. *Biochim.Biophys.Acta.* 1216:441-445.
- Kuntz, I.D., K. Chen, K.A. Sharp, and P.A. Kollman. 1999. The maximal affinity of ligands. *Proc.Natl.Acad.Sci.U.S.A.* 96:9997-10002.

- Kusnadi, A.R., E.E. Hood, D.R. Witcher, J.A. Howard, and Z.L. Nikolov. 1998. Production and purification of two recombinant proteins from transgenic corn. *Biotechnol.Prog.* 14:149-155.
- Laitinen, O.H., K.J. Airene, A.T. Marttila, T. Kulik, E. Porkka, E.A. Bayer, M. Wilchek, and M.S. Kulomaa. 1999. Mutation of a critical tryptophan to lysine in avidin or streptavidin may explain why sea urchin fibropellin adopts an avidin-like domain. *FEBS Lett.* 461:52-58.
- Laitinen, O.H., V.P. Hytönen, M.K. Ahlroth, O.T. Pentikäinen, C. Gallagher, H.R. Nordlund, V. Ovod, A.T. Marttila, E. Porkka, S. Heino, M.S. Johnson, K.J. Airene, and M.S. Kulomaa. 2002. Chicken avidin-related proteins show altered biotin-binding and physico-chemical properties as compared with avidin. *Biochem.J.* 363:609-617.
- Laitinen, O.H., V.P. Hytönen, H.R. Nordlund, and M.S. Kulomaa. 2006. Genetically engineered avidins and streptavidins. *Cell Mol.Life Sci.* 63:2992-3017.
- Laitinen, O.H., A.T. Marttila, K.J. Airene, T. Kulik, O. Livnah, E.A. Bayer, M. Wilchek, and M.S. Kulomaa. 2001. Biotin induces tetramerization of a recombinant monomeric avidin. A model for protein-protein interactions. *J.Biol.Chem.* 276:8219-8224.
- Laitinen, O.H., H.R. Nordlund, V.P. Hytönen, and M.S. Kulomaa. 2007. Brave new (strept)avidins in biotechnology. *Trends Biotechnol.* 25:269-277.
- Laitinen, O.H., H.R. Nordlund, V.P. Hytönen, S.T. Uotila, A.T. Marttila, J. Savolainen, K.J. Airene, O. Livnah, E.A. Bayer, M. Wilchek, and M.S. Kulomaa. 2003. Rational design of an active avidin monomer. *J.Biol.Chem.* 278:4010-4014.
- Lamzin, V.S., and K.S. Wilson. 1993. Automated refinement of protein models. *Acta Crystallogr.D Biol.Crystallogr.* 49:129-147.
- Langer, G., S.X. Cohen, V.S. Lamzin, and A. Perrakis. 2008. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat.Protoc.* 3:1171-1179.
- Lehtonen, J., D. Still, V. Rantanen, J. Ekholm, D. Björklund, Z. Iftikhar, M. Huhtala, S. Repo, A. Jussila, J. Jaakkola, O. Pentikäinen, T. Nyrönen, T. Salminen, M. Gyllenberg, and M. Johnson. 2004. BODIL: a molecular modeling environment for structure-function analysis and drug design. *J.Comput.Aided Mol.Des.* 18:401-419.
- Lehtonen S.I., B. Taskinen, E. Ojala, S. Kukkurainen, R. Rahikainen, T.A. Riihimäki, O.H. Laitinen, M.S. Kulomaa, and V.P. Hytönen. 2015. Efficient preparation of shuffled DNA libraries through recombination (Gateway) cloning. *Protein Eng Des Sel.* Jan;28(1):23-8.
- Lehtonen S.I., A. Tullila, N. Agrawal, S. Kukkurainen, N. Kähkönen, M. Koskinen, T.K. Nevanen, M.S. Johnson, T.T. Airene, M.S. Kulomaa, T.A. Riihimäki, and V.P. Hytönen. 2016. Artificial Avidin-Based Receptors for a Panel of Small Molecules. *ACS Chem Biol.* Jan 15;11(1):211-21.
- Lei, S.P., H.C. Lin, S.S. Wang, J. Callaway, and G. Wilcox. 1987. Characterization of the *Erwinia carotovora* pelB gene and its product pectate lyase. *J.Bacteriol.* 169:4379-4383.
- Leppiniemi, J., T. Grönroos, J.A. Määttä, M.S. Johnson, M.S. Kulomaa, V.P. Hytönen, and T.T. Airene. 2012. Structure of bradavidin-C-terminal residues act as intrinsic ligands. *PLoS One.* 7:e35962.
- Leppiniemi, J., J.A. Määttä, H. Hammaren, M. Soikkeli, M. Laitaoja, J. Jänis, M.S. Kulomaa, and V.P. Hytönen. 2011. Bifunctional avidin with covalently modifiable ligand binding site. *PLoS One.* 6:e16576.

- Leppiniemi, J., A. Meir, N. Kähkönen, S. Kukkurainen, J.A. Määttä, M. Ojanen, J. Jänis, M.S. Kulomaa, O. Livnah, and V.P. Hytönen. 2013. The highly dynamic oligomeric structure of bradavidin II is unique among avidin proteins. *Protein Sci.* 22:980-994.
- Li, J., A. Mahajan, and M.D. Tsai. 2006. Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry.* 45:15168-15178.
- Li, Y., K. Suino, J. Daugherty, and H.E. Xu. 2005. Structural and biochemical mechanisms for the specificity of hormone binding and coactivator assembly by mineralocorticoid receptor. *Mol.Cell.* 19:367-380.
- Lin, H., and V.W. Cornish. 2002. Screening and selection methods for large-scale analysis of protein function. *Angew.Chem.Int.Ed Engl.* 41:4402-4425.
- Lipovsek, D. 2011. Adnectins: engineered target-binding protein therapeutics. *Protein Eng.Des.Sel.* 24:3-9.
- Liu, J., B. Ning, M. Liu, Y. Sun, Z. Sun, Y. Zhang, X. Fan, Z. Zhou, and Z. Gao. 2012. Construction of ribosome display library based on lipocalin scaffold and screening anticalins with specificity for estradiol. *Analyst.* 137:2470-2479.
- Livnah, O., E.A. Bayer, M. Wilchek, and J.L. Sussman. 1993. Three-dimensional structures of avidin and the avidin-biotin complex. *Proc.Natl.Acad.Sci.U.S.A.* 90:5076-5080.
- Löfblom, J., J. Feldwisch, V. Tolmachev, J. Carlsson, S. Ståhl, and F.Y. Frejd. 2010. Affibody molecules: engineered proteins for therapeutic, diagnostic and biotechnological applications. *FEBS Lett.* 584:2670-2680.
- Low, K.O., N. Muhammad Mahadi, and R. Md Illias. 2013. Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. *Appl.Microbiol.Biotechnol.* 97:3811-3826.
- Lutz, S., M. Ostermeier, G.L. Moore, C.D. Maranas, and S.J. Benkovic. 2001. Creating multiple-crossover DNA libraries independent of sequence identity. *Proc.Natl.Acad.Sci.U.S.A.* 98:11248-11253.
- Määttä, J.A., T.T. Airene, H.R. Nordlund, J. Jänis, T.A. Paldanius, P. Vainiotalo, M.S. Johnson, M.S. Kulomaa, and V.P. Hytönen. 2008. Rational modification of ligand-binding preference of avidin by circular permutation and mutagenesis. *Chembiochem.* 9:1124-1135.
- Määttä, J.A., Y. Eisenberg-Domovich, H.R. Nordlund, R. Hayouka, M.S. Kulomaa, O. Livnah, and V.P. Hytönen. 2011. Chimeric avidin shows stability against harsh chemical conditions--biochemical analysis and 3D structure. *Biotechnol.Bioeng.* 108:481-490.
- Määttä, J.A., S.H. Helppolainen, V.P. Hytönen, M.S. Johnson, M.S. Kulomaa, T.T. Airene, and H.R. Nordlund. 2009. Structural and functional characteristics of xenavidin, the first frog avidin from *Xenopus tropicalis*. *BMC Struct.Biol.* 9:63-6807-9-63.
- Malm, M., N. Kronqvist, H. Lindberg, L. Gudmundsdotter, T. Bass, F.Y. Frejd, I. Höiden-Guthenberg, Z. Varasteh, A. Orlova, V. Tolmachev, S. Ståhl, and J. Löfblom. 2013. Inhibiting HER3-mediated tumor cell growth with affibody molecules engineered to low picomolar affinity by position-directed error-prone PCR-like diversification. *PLoS One.* 8:e62791.
- Marttila, A.T., K.J. Airene, O.H. Laitinen, T. Kulik, E.A. Bayer, M. Wilchek, and M.S. Kulomaa. 1998. Engineering of chicken avidin: a progressive series of reduced charge mutants. *FEBS Lett.* 441:313-317.
- Marttila, A.T., V.P. Hytönen, O.H. Laitinen, E.A. Bayer, M. Wilchek, and M.S. Kulomaa. 2003. Mutation of the important Tyr-33 residue of chicken avidin: functional and structural consequences. *Biochem.J.* 369:249-254.

- Marttila, A.T., O.H. Laitinen, K.J. Airene, T. Kulik, E.A. Bayer, M. Wilchek, and M.S. Kulomaa. 2000. Recombinant Neutralite avidin: a non-glycosylated, acidic mutant of chicken avidin that exhibits high affinity for biotin and low non-specific binding properties. *FEBS Lett.* 467:31-36.
- Maruyama, I.N., H.I. Maruyama, and S. Brenner. 1994. Lambda foo: a lambda phage vector for the expression of foreign proteins. *Proc.Natl.Acad.Sci.U.S.A.* 91:8273-8277.
- Mattheakis, L.C., R.R. Bhatt, and W.J. Dower. 1994. An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proc.Natl.Acad.Sci.U.S.A.* 91:9022-9026.
- McConnell, S.J., T. Dinh, M.H. Le, and D.G. Spinella. 1999. Biopanning phage display libraries using magnetic beads vs. polystyrene plates. *BioTechniques.* 26:208-10, 214.
- McCoy, A.J., R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, and R.J. Read. 2007. Phaser crystallographic software. *J.Appl.Crystallogr.* 40:658-674.
- Mead, D.A., and B. Kemper. 1988. Chimeric single-stranded DNA phage-plasmid cloning vectors. *Biotechnology.* 10:85-102.
- Meir, A., E.A. Bayer, and O. Livnah. 2012. Structural adaptation of a thermostable biotin-binding protein in a psychrophilic environment. *J.Biol.Chem.* 287:17951-17962.
- Meir, A., S.H. Helppolainen, E. Podoly, H.R. Nordlund, V.P. Hytönen, J.A. Määttä, M. Wilchek, E.A. Bayer, M.S. Kulomaa, and O. Livnah. 2009. Crystal structure of rhizavidin: insights into the enigmatic high-affinity interaction of an innate biotin-binding protein dimer. *J.Mol.Biol.* 386:379-390.
- Melamed, M.D., and N.M. Green. 1963. Avidin. 2. Purification and Composition. *Biochem.J.* 89:591-599.
- Moore, G.L., C.D. Maranas, S. Lutz, and S.J. Benkovic. 2001. Predicting crossover generation in DNA shuffling. *Proc.Natl.Acad.Sci.U.S.A.* 98:3226-3231.
- Morris, R.J., A. Perrakis, and V.S. Lamzin. 2003. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* 374:229-244.
- Murshudov, G.N., P. Skubák, A.A. Lebedev, N.S. Pannu, R.A. Steiner, R.A. Nicholls, M.D. Winn, F. Long, and A.A. Vagin. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr.D Biol.Crystallogr.* 67:355-367.
- Muteeb, G., and R. Sen. 2010. Random mutagenesis using a mutator strain. *Methods Mol.Biol.* 634:411-419.
- Muyldermans, S. 2013. Nanobodies: natural single-domain antibodies. *Annu.Rev.Biochem.* 82:775-797.
- Muyldermans, S., and M. Lauwereys. 1999. Unique single-domain antigen binding fragments derived from naturally occurring camel heavy-chain antibodies. *J.Mol.Recognit.* 12:131-140.
- Nardone, E., C. Rosano, P. Santambrogio, F. Curnis, A. Corti, F. Magni, A.G. Siccardi, G. Paganelli, R. Losso, B. Aprea, M. Bolognesi, A. Sidoli, and P. Arosio. 1998. Biochemical characterization and crystal structure of a recombinant hen avidin and its acidic mutant expressed in *Escherichia coli*. *Eur.J.Biochem.* 256:453-460.
- Neylon, C. 2004. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res.* 32:1448-1459.
- Nicholas, K.B., H.B.J. Nicholas, and D.W.I. Deerfield. 1997. GeneDoc: Analysis and Visualization of Genetic Variation. *Embnew.News.* 4.

- Niederhauser, B., J. Siivonen, J.A. Määttä, J. Jänis, M.S. Kulomaa, and V.P. Hytönen. 2012. DNA family shuffling within the chicken avidin protein family - A shortcut to more powerful protein tools. *J.Biotechnol.* 157:38-49.
- Nord, K., E. Gunneriusson, J. Ringdahl, S. Ståhl, M. Uhlen, and P.A. Nygren. 1997. Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nat.Biotechnol.* 15:772-777.
- Nord, K., J. Nilsson, B. Nilsson, M. Uhlén, and P.A. Nygren. 1995. A combinatorial library of an alpha-helical bacterial receptor domain. *Protein Eng.* 8:601-608.
- Nordlund, H.R., V.P. Hytönen, J. Hörhä, J.A. Määttä, D.J. White, K. Halling, E.J. Porkka, J.P. Slotte, O.H. Laitinen, and M.S. Kulomaa. 2005a. Tetravalent single-chain avidin: from subunits to protein domains via circularly permuted avidins. *Biochem.J.* 392:485-491.
- Nordlund, H.R., V.P. Hytönen, O.H. Laitinen, and M.S. Kulomaa. 2005b. Novel avidin-like protein from a root nodule symbiotic bacterium, *Bradyrhizobium japonicum*. *J.Biol.Chem.* 280:13250-13255.
- Nordlund, H.R., V.P. Hytönen, O.H. Laitinen, S.T. Uotila, E.A. Niskanen, J. Savolainen, E. Porkka, and M.S. Kulomaa. 2003a. Introduction of histidine residues into avidin subunit interfaces allows pH-dependent regulation of quaternary structure and biotin binding. *FEBS Lett.* 555:449-454.
- Nordlund, H.R., O.H. Laitinen, V.P. Hytönen, S.T. Uotila, E. Porkka, and M.S. Kulomaa. 2004. Construction of a dual chain pseudotetrameric chicken avidin by combining two circularly permuted avidins. *J.Biol.Chem.* 279:36715-36719.
- Nordlund, H.R., O.H. Laitinen, S.T. Uotila, T. Nyholm, V.P. Hytönen, J.P. Slotte, and M.S. Kulomaa. 2003b. Enhancing the thermal stability of avidin. Introduction of disulfide bridges between subunit interfaces. *J.Biol.Chem.* 278:2479-2483.
- Nov, Y. 2014. Probabilistic methods in directed evolution: library size, mutation rate, and diversity. *Methods Mol.Biol.* 1179:261-278.
- Nov, Y. 2012. When second best is good enough: another probabilistic look at saturation mutagenesis. *Appl.Environ.Microbiol.* 78:258-262.
- Nygren, P.A. 2008. Alternative binding proteins: affibody binding proteins developed from a small three-helix bundle scaffold. *Febs J.* 275:2668-2676.
- Olwill, S.A., C. Joffroy, H. Gille, E. Vigna, G. Matschiner, A. Allersdorfer, B.M. Lunde, J. Jaworski, J.F. Burrows, C. Chiriaco, H.J. Christian, M. Hülsmeier, S. Trentmann, K. Jensen, A.M. Hohlbaum, and L. Audoly. 2013. A highly potent and specific MET therapeutic protein antagonist with both ligand-dependent and ligand-independent activity. *Mol.Cancer.Ther.* 12:2459-2471.
- O'Malley, B.W., and W.L. McGuire. 1968. Studies on the mechanism of action of progesterone in regulation of the synthesis of specific protein. *J.Clin.Invest.* 47:654-664.
- Orlova, A., M. Magnusson, T.L. Eriksson, M. Nilsson, B. Larsson, I. Höidén-Guthenberg, C. Widström, J. Carlsson, V. Tolmachev, S. Ståhl, and F.Y. Nilsson. 2006. Tumor imaging using a picomolar affinity HER2 binding affibody molecule. *Cancer Res.* 66:4339-4348.
- Ostermeier, M., J.H. Shim, and S.J. Benkovic. 1999. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat.Biotechnol.* 17:1205-1209.
- Packer, M.S., and D.R. Liu. 2015. Methods for the directed evolution of proteins. *Nat.Rev.Genet.* 16:379-394.

- Pande, J., M.M. Szewczyk, and A.K. Grover. 2010. Phage display: concept, innovations, applications and future. *Biotechnol.Adv.* 28:849-858.
- Parker, M.H., Y. Chen, F. Danehy, K. Dufu, J. Ekstrom, E. Getmanova, J. Gokemeijer, L. Xu, and D. Lipovsek. 2005. Antibody mimics based on human fibronectin type three domain engineered for thermostability and high-affinity binding to vascular endothelial growth factor receptor two. *Protein Eng.Des.Sel.* 18:435-444.
- Perrakis, A., R. Morris, and V.S. Lamzin. 1999. Automated protein model building combined with iterative structure refinement. *Nat.Struct.Biol.* 6:458-463.
- Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J.Comput.Chem.* 26:1781-1802.
- Plückthun, A. 2015. Designed ankyrin repeat proteins (DARPs): binding proteins for research, diagnostics, and therapy. *Annu.Rev.Pharmacol.Toxicol.* 55:489-511.
- Pollheimer, P., B. Taskinen, A. Scherfler, S. Gusenkov, M. Creus, P. Wiesauer, D. Zauner, W. Schöffberger, C. Schwarzing, A. Ebner, R. Tampé, H. Stutz, V.P. Hytönen, and H.J. Gruber. 2013. Reversible biofunctionalization of surfaces with a switchable mutant of avidin. *Bioconj.Chem.* 24:1656-1668.
- Potterton, E., P. Briggs, M. Turkenburg, and E. Dodson. 2003. A graphical user interface to the \it CCP4 program suite. *Acta Crystallographica Section D.* 59:1131-1137.
- Pugliese, L., A. Coda, M. Malcovati, and M. Bolognesi. 1993. Three-dimensional structure of the tetragonal crystal form of egg-white avidin in its functional complex with biotin at 2.7 Å resolution. *J.Mol.Biol.* 231:698-710.
- Qureshi, M.H., and S.L. Wong. 2002. Design, production, and characterization of a monomeric streptavidin and its application for affinity purification of biotinylated proteins. *Protein Expr.Purif.* 25:409-415.
- Rauth, S., D. Hinz, M. Börger, M. Uhrig, M. Mayhaus, M. Riemenschneider, and A. Skerra. 2016. High-affinity Anticalins with aggregation-blocking activity directed against the Alzheimer beta-amyloid peptide. *Biochem.J.* 473:1563-1578.
- Reetz, M.T., and J.D. Carballera. 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat.Protoc.* 2:891-903.
- Reetz, M.T., D. Kahakeaw, and R. Lohmer. 2008. Addressing the numbers problem in directed evolution. *Chembiochem.* 9:1797-1804.
- Ren, Z., and L.W. Black. 1998. Phage T4 SOC and HOC display of biologically active, full-length proteins on the viral capsid. *Gene.* 215:439-444.
- Repo, S., T.A. Paldanius, V.P. Hytönen, T.K. Nyholm, K.K. Halling, J. Huuskonen, O.T. Pentikäinen, K. Rissanen, J.P. Slotte, T.T. Airenne, T.A. Salminen, M.S. Kulomaa, and M.S. Johnson. 2006. Binding properties of HABA-type azo derivatives to avidin and avidin-related protein 4. *Chem.Biol.* 13:1029-1039.
- Reznik, G.O., S. Vajda, T. Sano, and C.R. Cantor. 1998. A streptavidin mutant with altered ligand-binding specificity. *Proc.Natl.Acad.Sci.U.S.A.* 95:13525-13530.
- Reznik, G.O., S. Vajda, C.L. Smith, C.R. Cantor, and T. Sano. 1996. Streptavidins with intersubunit crosslinks have enhanced stability. *Nat.Biotechnol.* 14:1007-1011.
- Richter, A., E. Eggenstein, and A. Skerra. 2014. Anticalins: exploiting a non-Ig scaffold with hypervariable loops for the engineering of binding proteins. *FEBS Lett.* 588:213-218.
- Riihimäki T.A., S. Hiltunen, M. Rangl, H.R. Nordlund, J.A. Määttä, A. Ebner, P. Hinterdorfer, M.S. Kulomaa, K. Takkinen, and V.P. Hytönen. 2011b. Modification of the loops in the ligand-binding site turns avidin into a steroid-binding protein. *BMC Biotechnol.* 11:64.

- Riihimäki, T.A., S. Kukkurainen, S. Varjonen, J. Hörhå, T.K. Nyholm, M.S. Kulomaa, and V.P. Hytönen. 2011a. Construction of chimeric dual-chain avidin by tandem fusion of the related avidins. *PLoS One*. 6:e20535.
- Roberts, R.W., and J.W. Szostak. 1997. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc.Natl.Acad.Sci.U.S.A.* 94:12297-12302.
- Rondot, S., J. Koch, F. Breitling, and S. Dubel. 2001. A helper phage to improve single-chain antibody presentation in phage display. *Nat.Biotechnol.* 19:75-78.
- Rönmark, J., H. Grönlund, M. Uhlén, and P.A. Nygren. 2002. Human immunoglobulin A (IgA)-specific ligands from combinatorial engineering of protein A. *Eur.J.Biochem.* 269:2647-2655.
- Rönmark, J., C. Kampf, A. Asplund, I. Höiden-Guthenberg, K. Wester, F. Pontén, M. Uhlén, and P.A. Nygren. 2003. Affibody-beta-galactosidase immunoconjugates produced as soluble fusion proteins in the *Escherichia coli* cytosol. *J.Immunol.Methods*. 281:149-160.
- Rosano, C., P. Arosio, and M. Bolognesi. 1999. The X-ray three-dimensional structure of avidin. *Biomol.Eng.* 16:5-12.
- Rosenberg, A., K. Griffin, F.W. Studier, M. McCormick, J. Berg, R. Novy, and R. and Mierendorf. 1996. T7Select® Phage Display System: A powerful new protein display system based on bacteriophage T7. in *Novations - Newsletter of Novagen, Inc.* 6:1-6.
- Roux, K.H., A.S. Greenberg, L. Greene, L. Strelets, D. Avila, E.C. McKinney, and M.F. Flajnik. 1998. Structural analysis of the nurse shark (new) antigen receptor (NAR): molecular convergence of NAR and unusual mammalian immunoglobulins. *Proc.Natl.Acad.Sci.U.S.A.* 95:11804-11809.
- Sano, T., and C.R. Cantor. 1990. Expression of a cloned streptavidin gene in *Escherichia coli*. *Proc.Natl.Acad.Sci.U.S.A.* 87:142-146.
- Sardo, A., T. Wohlschlager, C. Lo, H. Zoller, T.R. Ward, and M. Creus. 2011. Burkavidin: a novel secreted biotin-binding protein from the human pathogen *Burkholderia pseudomallei*. *Protein Expr.Purif.* 77:131-139.
- Schilling, J., J. Schöppe, and A. Plückthun. 2014. From DARPins to LoopDARPins: novel LoopDARPin design allows the selection of low picomolar binders in a single round of ribosome display. *J.Mol.Biol.* 426:691-721.
- Schlehuber, S., G. Beste, and A. Skerra. 2000. A novel type of receptor protein, based on the lipocalin scaffold, with specificity for digoxigenin. *J.Mol.Biol.* 297:1105-1120.
- Schlehuber, S., and A. Skerra. 2005. Lipocalins in drug discovery: from natural ligand-binding proteins to "anticalins". *Drug Discov.Today*. 10:23-33.
- Schlehuber, S., and A. Skerra. 2002. Tuning ligand affinity, specificity, and folding stability of an engineered lipocalin variant -- a so-called 'anticalin' -- using a molecular random approach. *Biophys.Chem.* 96:213-228.
- Schönfeld, D., G. Matschiner, L. Chatwell, S. Trentmann, H. Gille, M. Hülsmeier, N. Brown, P.M. Kaye, S. Schlehuber, A.M. Hohlbaum, and A. Skerra. 2009. An engineered lipocalin specific for CTLA-4 reveals a combining site with structural and conformational features similar to antibodies. *Proc.Natl.Acad.Sci.U.S.A.* 106:8198-8203.
- Sidhu, S.S., G.A. Weiss, and J.A. Wells. 2000. High copy display of large proteins on phage for functional selections. *J.Mol.Biol.* 296:487-495.
- Sieber, V., C.A. Martinez, and F.H. Arnold. 2001. Libraries of hybrid proteins from distantly related sequences. *Nat.Biotechnol.* 19:456-460.

- Simeon, R., and Z. Chen. 2017. In vitro-engineered non-antibody protein therapeutics. *Protein Cell*.
- Singaravelan, B., B.R. Roshini, and M.H. Munavar. 2010. Evidence that the supE44 mutation of *Escherichia coli* is an amber suppressor allele of *glnX* and that it also suppresses ochre and opal nonsense mutations. *J.Bacteriol.* 192:6039-6044.
- Skerra, A. 2008. Alternative binding proteins: anticalins - harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities. *Febs J.* 275:2677-2683.
- Skerra, A. 2003. Imitating the humoral immune response. *Curr.Opin.Chem.Biol.* 7:683-693.
- Skerra, A. 2000. Engineered protein scaffolds for molecular recognition. *J.Mol.Recognit.* 13:167-187.
- Smith, G.P. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science.* 228:1315-1317.
- Smith, G.P., and V.A. Petrenko. 1997. Phage Display. *Chem.Rev.* 97:391-410.
- Soltes, G., H. Barker, K. Marmai, E. Pun, A. Yuen, and E.J. Wiersma. 2003. A new helper phage and phagemid vector system improves viral display of antibody Fab fragments and avoids propagation of insert-less virions. *J.Immunol.Methods.* 274:233-244.
- Stahl, S., T. Gräslund, A. Eriksson Karlström, F.Y. Frejd, P.A. Nygren, and J. Löfblom. 2017. Affibody Molecules in Biotechnological and Medical Applications. *Trends Biotechnol.*
- Stahl, S., and M. Uhlén. 1997. Bacterial surface display: trends and progress. *Trends Biotechnol.* 15:185-192.
- Stefan, N., P. Martin-Killias, S. Wyss-Stoeckle, A. Honegger, U. Zangemeister-Wittke, and A. Plückthun. 2011. DARPin recognizing the tumor-associated antigen EpCAM selected by phage and ribosome display and engineered for multivalency. *J.Mol.Biol.* 413:826-843.
- Stemmer, W.P. 1994. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc.Natl.Acad.Sci.U.S.A.* 91:10747-10751.
- Sternberg, N., and R.H. Hoess. 1995. Display of peptides and proteins on the surface of bacteriophage lambda. *Proc.Natl.Acad.Sci.U.S.A.* 92:1609-1613.
- Swillens, S. 1995. Interpretation of binding curves obtained with high receptor concentrations: practical aid for computer analysis. *Mol.Pharmacol.* 47:1197-1203.
- Takakura, Y., K. Sofuku, M. Tsunashima, and S. Kuwata. 2016. Lentiavidins: Novel avidin-like proteins with low isoelectric points from shiitake mushroom (*Lentinula edodes*). *J.Biosci.Bioeng.* 121:420-423.
- Takakura, Y., M. Tsunashima, J. Suzuki, S. Usami, Y. Kakuta, N. Okino, M. Ito, and T. Yamamoto. 2009. Tamavidins--novel avidin-like biotin-binding proteins from the Tamogitake mushroom. *Febs J.* 276:1383-1397.
- Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol.Biol.Evol.* 28:2731-2739.
- Tang, L., H. Gao, X. Zhu, X. Wang, M. Zhou, and R. Jiang. 2012. Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers. *BioTechniques.* 52:149-158.
- Taskinen, B., T.T. Airenne, J. Jänis, R. Rahikainen, M.S. Johnson, M.S. Kulomaa, and V.P. Hytönen. 2014a. A novel chimeric avidin with increased thermal stability using DNA shuffling. *PLoS One.* 9:e92058.

- Taskinen, B., D. Zauner, S.I. Lehtonen, M. Koskinen, C. Thomson, N. Kähkönen, S. Kukkurainen, J.A. Määttä, T.O. Ihalainen, M.S. Kulomaa, H.J. Gruber, and V.P. Hytönen. 2014b. Switchavidin: reversible biotin-avidin-biotin bridges with high affinity and specificity. *Bioconjug.Chem.* 25:2233-2243.
- Taskinen, B., J. Zmurko, M. Ojanen, S. Kukkurainen, M. Parthiban, J.A. Määttä, J. Leppiniemi, J. Jänis, M. Parikka, H. Turpeinen, M. Rämetsä, M. Pesu, M.S. Johnson, M.S. Kulomaa, T.T. Airenne, and V.P. Hytönen. 2013. Zebavidin--an avidin-like protein from zebrafish. *PLoS One.* 8:e77207.
- Tausig, F., and F.J. Wolf. 1964. Streptavidin--a substance with avidin-like properties produced by microorganisms. *Biochem.Biophys.Res.Commun.* 14:205-209.
- Tracewell, C.A., and F.H. Arnold. 2009. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr.Opin.Chem.Biol.* 13:3-9.
- Traore, S.M., and B. Zhao. 2011. A novel Gateway(R)-compatible binary vector allows direct selection of recombinant clones in *Agrobacterium tumefaciens*. *Plant.Methods.* 7:42-4811-7-42.
- Turunen, L., K. Takkinen, H. Söderlund, and T. Pulli. 2009. Automated panning and screening procedure on microplates for antibody generation from phage display libraries. *J.Biomol.Screen.* 14:282-293.
- Vauquelin, G., and S.J. Charlton. 2013. Exploring avidity: understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands. *Br.J.Pharmacol.* 168:1771-1785.
- Veneskoski, K. 2009. pro gradu thesis: Strongavidini - Biotiinia sitova merisilestä peräisin oleva avidiinin kaltainen proteiini.
- Vogt, M., and A. Skerra. 2004. Construction of an artificial receptor protein ("anticalin") based on the human apolipoprotein D. *Chembiochem.* 5:191-199.
- Wada, A., S.Y. Sawata, and Y. Ito. 2008. Ribosome display selection of a metal-binding motif from an artificial peptide library. *Biotechnol.Bioeng.* 101:1102-1107.
- Walker, R.G., A.T. Willingham, and C.S. Zuker. 2000. A *Drosophila* mechanosensory transduction channel. *Science.* 287:2229-2234.
- Wallén, M.J., M.O. Laukkanen, and M.S. Kulomaa. 1995. Cloning and sequencing of the chicken egg-white avidin-encoding gene and its relationship with the avidin-related genes Avr1-Avr5. *Gene.* 161:205-209.
- Wang, C., M. Eufemi, C. Turano, and A. Giartosio. 1996. Influence of the carbohydrate moiety on the stability of glycoproteins. *Biochemistry.* 35:7299-7307.
- Wang, J., R.M. Wolf, J.W. Caldwell, P.A. Kollman, and D.A. Case. 2004. Development and testing of a general amber force field. *J.Comput.Chem.* 25:1157-1174.
- Ward, E.S., D. Gussow, A.D. Griffiths, P.T. Jones, and G. Winter. 1989. Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. *Nature.* 341:544-546.
- Weber, P.C., D.H. Ohlendorf, J.J. Wendoloski, and F.R. Salemme. 1989. Structural origins of high-affinity biotin binding to streptavidin. *Science.* 243:85-88.
- White, H.B., 3rd, and C.C. Whitehead. 1987. Role of avidin and other biotin-binding proteins in the deposition and distribution of biotin in chicken eggs. Discovery of a new biotin-binding protein. *Biochem.J.* 241:677-684.
- Wilchek, M., and E.A. Bayer. 1999. Foreword and introduction to the book (strept)avidin-biotin system. *Biomol.Eng.* 16:1-4.

- Wildling, L., P. Hinterdorfer, K. Kusche-Vihrog, Y. Treffner, and H. Oberleithner. 2009. Aldosterone receptor sites on plasma membrane of human vascular endothelium detected by a mechanical nanosensor. *Pflügers Arch.* 458:223-230.
- Willats, W.G. 2002. Phage display: practicalities and prospects. *Plant Mol.Biol.* 50:837-854.
- Wu, S.C., and S.L. Wong. 2005. Engineering soluble monomeric streptavidin with reversible biotin binding capability. *J.Biol.Chem.* 280:23225-23231.
- Xu, L., P. Aha, K. Gu, R.G. Kuimelis, M. Kurz, T. Lam, A.C. Lim, H. Liu, P.A. Lohse, L. Sun, S. Weng, R.W. Wagner, and D. Lipovsek. 2002. Directed evolution of high-affinity antibody mimics using mRNA display. *Chem.Biol.* 9:933-942.
- Zahnd, C., F. Pecorari, N. Straumann, E. Wyler, and A. Plückthun. 2006. Selection and characterization of Her2 binding-designed ankyrin repeat proteins. *J.Biol.Chem.* 281:35167-35175.
- Zahnd, C., E. Wyler, J.M. Schwenk, D. Steiner, M.C. Lawrence, N.M. McKern, F. Pecorari, C.W. Ward, T.O. Joos, and A. Plückthun. 2007. A designed ankyrin repeat protein evolved to picomolar affinity to Her2. *J.Mol.Biol.* 369:1015-1028.
- Zhao, H., L. Giver, Z. Shao, J.A. Affholter, and F.H. Arnold. 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat.Biotechnol.* 16:258-261.
- Zocchi, A., A.M. Jobe, J.M. Neuhaus, and T.R. Ward. 2003. Expression and purification of a recombinant avidin with a lowered isoelectric point in *Pichia pastoris*. *Protein Expr.Purif.* 32:167-174.

8 ORIGINAL PUBLICATIONS

RESEARCH ARTICLE

Open Access

Modification of the loops in the ligand-binding site turns avidin into a steroid-binding protein

Tiina A Riihimäki¹, Soili Hiltunen¹, Martina Rangl², Henri R Nordlund¹, Juha AE Määttä¹, Andreas Ebner², Peter Hinterdorfer², Markku S Kulomaa¹, Kristiina Takkinen³ and Vesa P Hytönen^{1*}

Abstract

Background: Engineered proteins, with non-immunoglobulin scaffolds, have become an important alternative to antibodies in many biotechnical and therapeutic applications. When compared to antibodies, tailored proteins may provide advantageous properties such as a smaller size or a more stable structure.

Results: Avidin is a widely used protein in biomedicine and biotechnology. To tailor the binding properties of avidin, we have designed a sequence-randomized avidin library with mutagenesis focused at the loop area of the binding site. Selection from the generated library led to the isolation of a steroid-binding avidin mutant (sbAvd-1) showing micromolar affinity towards testosterone ($K_d \sim 9 \mu\text{M}$). Furthermore, a gene library based on the sbAvd-1 gene was created by randomizing the loop area between β -strands 3 and 4. Phage display selection from this library led to the isolation of a steroid-binding protein with significantly decreased biotin binding affinity compared to sbAvd-1. Importantly, differential scanning calorimetry and analytical gel-filtration revealed that the high stability and the tetrameric structure were preserved in these engineered avidins.

Conclusions: The high stability and structural properties of avidin make it an attractive molecule for the engineering of novel receptors. This methodology may allow the use of avidin as a universal scaffold in the development of novel receptors for small molecules.

Keywords: protein engineering, avidin scaffold, phage display, steroid hormone, testosterone

Background

Antibodies are the most widely used biomolecules for therapeutic, diagnostic and research applications, because they can be generated against virtually any molecule using protein engineering techniques (for a review see [1]). However, antibodies have certain fundamental disadvantages such as the complex architecture of their antigen-binding site, low stability, and a rather large size [2-5]. Moreover, the production of full-size antibodies is relatively expensive [6]. To overcome these limitations and to improve therapeutic antibodies, antibodies have been extensively engineered [7]. For example the size of the antibody molecule has been reduced by producing single-domain antigen-binding derivatives [8]. In addition to extensive antibody engineering, a

versatile repertoire of tailored biomolecules from non-immunoglobulin protein scaffolds have been generated [4,9,10]. Anticalins, derived from the lipocalin fold, are a good example of engineered proteins [11]. The β -barrel structure of lipocalins is thermostable and robust and serves as an excellent scaffold for engineering novel receptors. They have been modified to bind novel ligands, such as fluorescein and digoxigenin, with affinities comparable with antibodies [12].

Chicken avidin (Avd), known for its extremely high affinity towards the water-soluble vitamin H, D-biotin, has been widely used in life science research applications [13]. Aside from biotin, Avd also binds dyes and peptides, which share no significant structural similarity with biotin [14,15]. Avd provides an attractive robust scaffold for the development of novel receptors, and Avd has many advantageous properties such as high chemical and thermal stability, a deep ligand binding site optimized for the binding of small molecules, and

* Correspondence: vesa.hytonen@uta.fi

¹Institute of Biomedical Technology, University of Tampere and Tampere University Hospital, FI-33520 Tampere, Finland

Full list of author information is available at the end of the article

an oligomeric nature enabling signal amplification. Moreover, the structure of Avd is well characterized [16,17] and numerous engineered forms of Avd have been described [18]. Engineered Avd forms, in which the two pairs of the binding sites (dual-chain Avd) [19-21] or all four binding sites (single-chain Avd) [22] can be independently manipulated, have been developed.

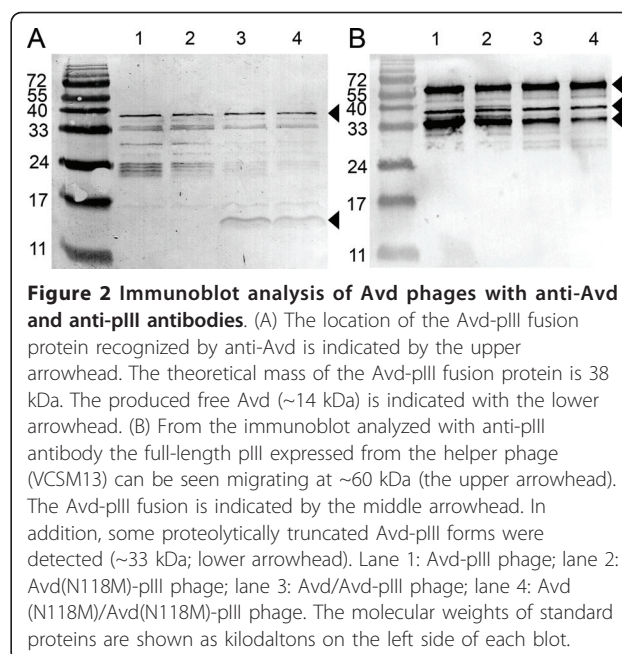
In the present study Avd was modified to bind steroid hormones. Avd mutant sbAvd-1, which was captured using the phage display selection [23], has a micromolar affinity to testosterone. The steroid-binding protein sbAvd-1 was characterized and further engineered to decrease cross-reactivity towards other molecules, especially towards Avd's natural ligand biotin. The resulting sbAvd-2 mutant with a modified loop between β -strands 3 and 4 was found to prefer steroid hormones over biotin in ligand binding.

Results

Functional display of Avd protein on the M13 phage

Avd was displayed on the surface of the M13 phage as a fusion with the C-terminal region of the minor coat protein pIII. Two different strategies for displaying the Avd scaffold in the active form on the M13 phage were evaluated. In the first display construct, Avd was produced solely as a fusion with pIII (Avd-pIII; Figure 1A), whereas in the Avd/Avd-pIII display construct, free Avd subunits were produced in addition to the pIII fusion (Figure 1B).

The size and oligomeric state of the Avd-pIII fusions were analyzed by SDS-PAGE and by western blot. Polyclonal anti-Avd (Figure 2A) indicated expression of both the Avd-pIII fusion (upper arrow, ~38 kDa) and the free Avd (lower arrow, ~15 kDa). The production of free subunits should enhance the functional assembly of the tetrameric Avd scaffold, especially if membrane anchoring of Avd by the pIII fusion partner has a negative effect on the oligomerization of Avd subunits. This strategy mimics the generally used amber-stop codon technique, in which free subunits are produced by the

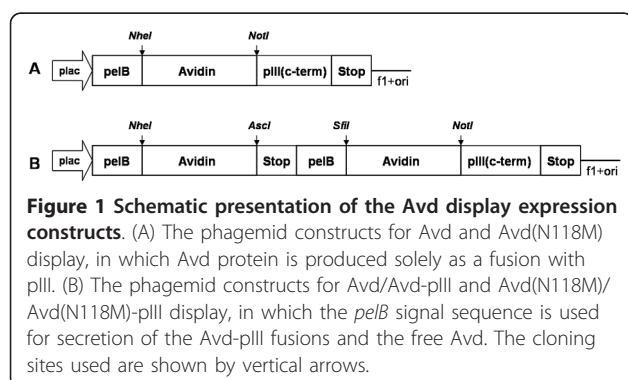


read-through of amber stop codon by the tRNA. We assume that the oligomerization of Avd, secreted by the *pelB* signal sequence, occurs in the periplasmic space of *E. coli*, allowing the display of functional Avd on the phage, as is the case for the display of the antibody Fab fragment, which requires folding of the heavy and light chains for the assembly of a functional antibody molecule [24].

A portion of the Avd-pIII fusion was partially proteolytically cleaved, as can be seen from the blot analyzed with the anti-pIII monoclonal antibody (Figure 2B, ~35 kDa band). Because Avd display constructs designed in this study were based on the monovalent display mode (3 + 3), the intact pIII (migrating at ~58 kDa) expressed from the helper phage was also detected from the blot (Figure 2B, uppermost arrow).

Avd-displaying phages were functional because they bound specifically to the biotin-coated surfaces, and phages were efficiently amplified even after several panning rounds. Moreover, phagemid DNA with the Avd display expression unit, which was confirmed by the restriction enzyme digestion analysis, was stable during the panning rounds (data not shown).

To analyze the functionality of phages, the mixture of amplified Avd and Avd mutant N118M phages were screened by panning against 4-hydroxyazobenzene-2-carboxylic acid (HABA). As determined in our previous study, the biotin binding affinity of the Avd mutant N118M was reduced ~1,000,000-fold ($K_d = 4.2 \times 10^{-9}$ M) and HABA-affinity was increased ~1.5-fold ($K_d = 5.2 \times 10^{-6}$ M) compared to wtAvd [25] (Additional file 1). During the panning procedure a clear enrichment of



Avd(N118M) phages over wt Avd phages was detected, which was an indication of the high selectivity of the produced phages. After only three rounds of selection phages displaying the Avd mutant N118M outcompeted the wtAvd phage population (Additional file 2).

Capture of a steroid-binding avidin

The loops adjacent to the ligand-binding site were selected for randomization [16]. First we created a library (Avd L1,2 library) in which residues N12, D13, L14, G15, and S16 were randomized. These amino acids form a loop between β -strands 1 and 2 (Figure 3). Codon NNN was used for randomization, and therefore, all 20 different amino acids were present, including all stop codons. The Avd L1,2 library was ligated into a phagemid as a sole fusion with the C-terminal portion of pIII. Based on sequencing results and transformation efficiency, the Avd L1,2 library was found to consist of approximately 1×10^5 individual members. When a stretch of five amino acid residues is completely randomized, the theoretical library size is 3.2×10^6 .

In the panning experiments Avd L1,2 loop library phages were introduced onto a testosterone-coated surface. The phage genomes carrying the mutated cDNA were found to be stable during the selection. A clear enrichment of sequences in the randomized loop area was observed, indicating the success of the selection conditions. Interestingly, we detected N12 as being a highly conserved amino acid residue among the

enriched pool of proteins. An Avd variant, named sbAvd-1, that carried the sequence N12, R13, M14, N15, H16 was selected for further analysis.

The specificity of sbAvd-1 can be tuned by additional mutations in the loop between β -strands 3 and 4

To further lower the biotin-binding affinity and to decrease the cross-reactivity of steroid-binding Avd, a library (sbAvd-1 L3,4) was generated in which the loop area between β -strands 3 and 4 was randomized (Figure 3). The loop between β -strands 3 and 4 is highly important for biotin binding of Avd because this loop 'locks' biotin into the binding site. In this process, three amino acid residues in the loop form direct interactions with biotin [16]. In the library four amino acids (T35, A36, V37 and, T38) were randomized using the NNY codon. The use of this codon covers 14 of the 20 naturally occurring amino acids while eliminating all of the stop codons. The library consisted of approximately 1.4×10^6 individual members, when calculated from the sequencing results and transformation efficiency, which exceeds the theoretical size of the library calculated based on the possible combinations of amino acid residues (3.8×10^4).

Binders from the sbAvd-1 L3,4 library were selected by phage display panning against a testosterone surface. In every panning round washes were adjusted according to the number of output colonies. The quality of the phage genomes carrying the mutated cDNA was evaluated by DNA sequencing at various stages during selection. A combination of acid and testosterone was used for elution. The selected sbAvd-1 phage clones were evaluated by microplate analysis using BSA-testosterone as a target ligand and utilizing M13-antibody to determine the amount of bound phages. The sbAvd-1 variant that showed the highest binding activity in the microplate assay (data not shown) had the sequence A35, T36, V37, N38. This mutant, named sbAvd-2 was selected for comparative analysis with sbAvd-1.

Production and purification of Avd mutants

Proteins were produced in soluble form in the *E. coli* strain BL21-AI using N-terminal OmpA bacterial secretion signal from *Bordetella avium* [26]. Proteins were purified by Ni-NTA affinity chromatography that yielded ~2 mg/L pure protein. According to gel-filtration analysis, in solution, both sbAvd-1 (51 kDa) and sbAvd-2 (53 kDa) showed tetrameric state (Additional file 3) similar to that of wtAvd (Avd expressed in *E. coli* 53 kDa, chicken Avd 60 kDa [26]). The slight decrease in molecular weight of Avd expressed in bacteria can be explained by the lack of glycosylation.

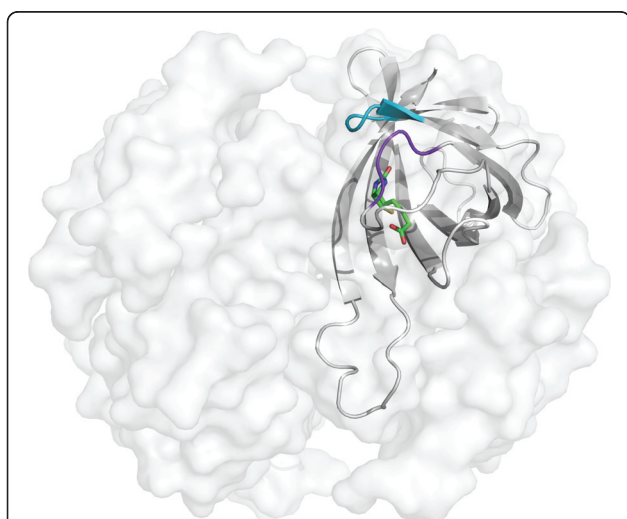
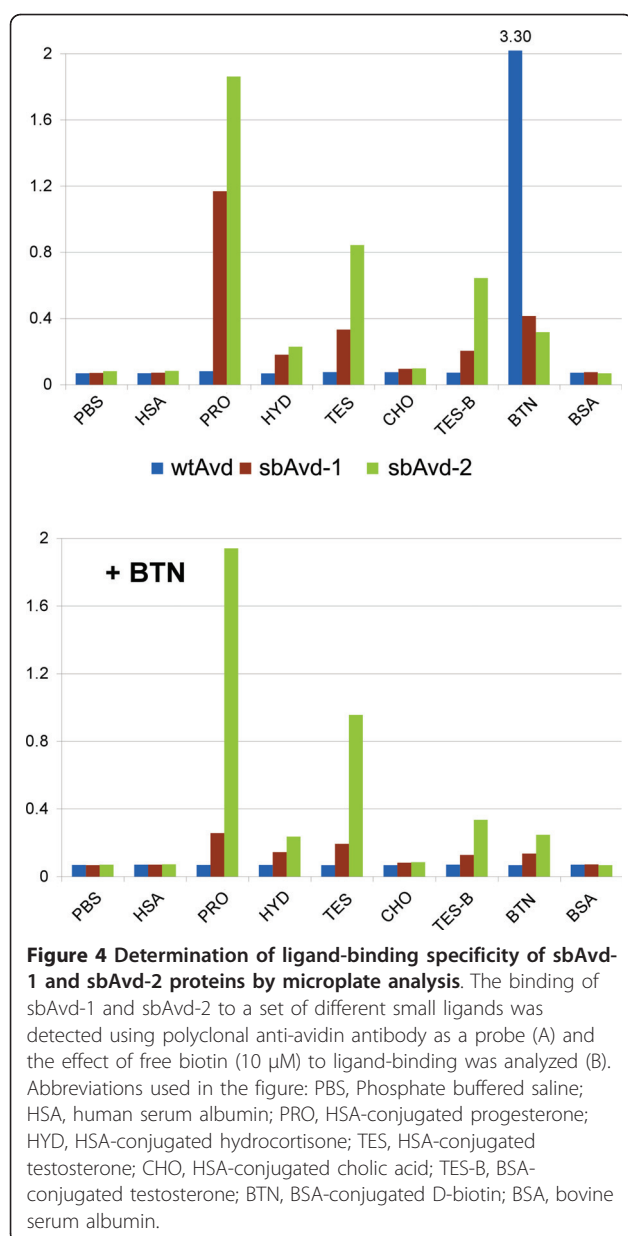


Figure 3 Three-dimensional structure of wtAvd with the loops chosen for random mutagenesis marked. One subunit of tetrameric wtAvd with bound biotin (PDB: 2AVI) [16] is shown in the figure. The randomized amino acid residues are N12, D13, L14, G15, and S16 (shown in blue) in the loop between β -strands 1 and 2, and T35, A36, V37, and T38 (shown in purple) in the loop between β -strands 3 and 4.



Determination of ligand-binding specificity of Avd forms by microplate assay

The binding specificity of proteins was analyzed by microplate assay, in which a set of different small molecules were used as a target molecules (Figure 4A). WtAvd was used as a negative control, and we detected no affinity towards the ligands except biotin. Importantly, sbAvd-1 and sbAvd-2 did not bind the proteins used as carriers for small molecules (bovine serum albumin (BSA) and human serum albumin (HSA)). However, these modified Avds showed clear binding to testosterone and progesterone.

The binding to the surface-immobilized ligands was inhibited by pre-incubating the proteins with 10

μ M D-biotin (Figure 4B). Free biotin significantly inhibited the binding of sbAvd-1 to steroids, indicating notable affinity towards biotin. However, in the case of sbAvd-2 the binding of testosterone and progesterone was not affected by biotin, showing a clear decrease in biotin-binding affinity.

Biosensor analyses of steroid-binding Avds

The kinetic constants of testosterone- and biotin-binding to steroid-binding proteins were determined with surface plasmon resonance (SPR) analysis. The purified sbAvd-1 and sbAvd-2 bound to a testosterone-BSA-coated sensor chip with similar affinities (Table 1), whereas wtAvd showed no binding to the testosterone surface. Furthermore, testosterone-binding was inhibited with varying testosterone concentrations (0.75–50 μ M, data not shown). In the case of sbAvd-1, the 50% inhibition was achieved between 5–10 μ M testosterone, which is consistent with the determined testosterone surface binding affinity. Interestingly, the 50% inhibition was already achieved in 750 nM testosterone concentration in the sbAvd-2 measurements. This result suggests a much higher binding affinity towards free testosterone than that measured towards surface-immobilized testosterone.

The biotin binding of the steroid-binding proteins was analyzed by the biotin-coated sensor chip. The comparison of sbAvd-1 and sbAvd-2 revealed that the biotin-binding affinity decreased almost 500-fold due to the modification of the loop between β -strands 3 and 4 (Table 1). This result appears to be consistent with the previous study in which mutation of T35A alone decreased the biotin-binding affinity of wtAvd approximately 200-fold [20].

The specificity of the steroid-binding was analyzed by binding competition using the following steroids: testosterone, dehydroepiandrosterone sulfate (DHEAS), androstenedione, estradiol, and dihydrotestosterone (DHT). In addition, the binding to the surface was also competed for by biotin. SbAvd-1 was found to be cross-reactive with androgens highly similar to testosterone (DHEAS and androstenedione), and had noticeable affinity towards biotin (Figure 5A). Interestingly, DHT showed less efficient inhibition compared to testosterone, suggesting lower binding affinity towards this steroid form. The binding of sbAvd-2 to testosterone was found to be most efficiently inhibited by testosterone and DHEAS (Figure 5B), whereas the inhibition caused by biotin was clearly weaker than that in the case of sbAvd-1. This result is consistent with the microplate analysis results and with the binding kinetic constants determined with SPR analysis.

Table 1 Kinetic parameters and the determined thermostability of sbAvds

Protein	Ligand	SPR			DSC		
		k_a (1/Ms)	k_d (1/s)	K_D (M)	T_m (°C)	ΔT_m (°C)	$\Delta H \times 10^4$ (cal/mol)
wtAvd	-	-	-	-	85.5	-	5.4
wtAvd	Btn	n.d.	n.d.	n.d.	123.2	37.7	12.8
wtAvd	Tes	-	-	no binding	86.2	0.7	5.6
sbAvd-1	-	-	-	-	80.6	-	5.2
sbAvd-1	Btn	4.2×10^5	5.6×10^{-4}	1.4×10^{-9}	83.2	2.6	6.4
sbAvd-1	Tes	1.0×10^3	9.5×10^{-3}	9.0×10^{-6}	81.5	0.9	5.7
sbAvd-2	-	-	-	-	82.5	-	4.6
sbAvd-2	Btn	1.0×10^3	6.8×10^{-4}	6.6×10^{-7}	83.0	0.5	4.2
sbAvd-2	Tes	813	8.5×10^{-3}	1.1×10^{-5}	83.1	0.6	4.8

The kinetic parameters determined by SPR analysis. The biotin binding affinity of avidin is too high to be determined by SPR. The transition midpoint of thermal unfolding (T_m) and the calorimetric heat of unfolding were determined by DSC. Delta T_m represents the increase of T_m in the presence of 50 μ M ligand.

Interaction analysis of steroid-binding Avds by MRFS

Molecular recognition force spectroscopy (MRFS) [27] was used to study the interaction between sbAvds and testosterone on a single molecule level. An atomic force microscopy (AFM) tip was functionalized with a single testosterone molecule [28] and repeatedly approached and retracted from the sbAvd-1- or sbAvd-2-coated surface (Figure 6A). The binding forces were measured in force-distance cycles, whereby the deflection (force) of the cantilever was recorded as a function of the tip-sample distance. For evaluation, 1,000 force-distance cycles were recorded, and probability density functions (pdf) were generated from the detected interaction forces [29]. The tip-tethered testosterone was found to form a complex with the sbAvds; the retraction led to a

downward bending of the cantilever until a particular force was reached resulting in the rupture of the bond between testosterone and sbAvd (Figure 6B and 6C). The most probable unbinding force [29] was found to be similar for both of the testosterone-sbAvd complexes: 40 pN at a constant pulling velocity of 600 nm/s. From 1,000 recorded force-distance cycles, sbAvd-1 showed 183 detected interactions and sbAvd-2 showed 215 events.

To prove the specificity of the measured interactions, control experiments were performed. For these experiments, free testosterone was injected into the measuring solution to preoccupy the binding sites of the steroid-binding protein immobilized on the surface. The retraction curve identical to the approaching curve was

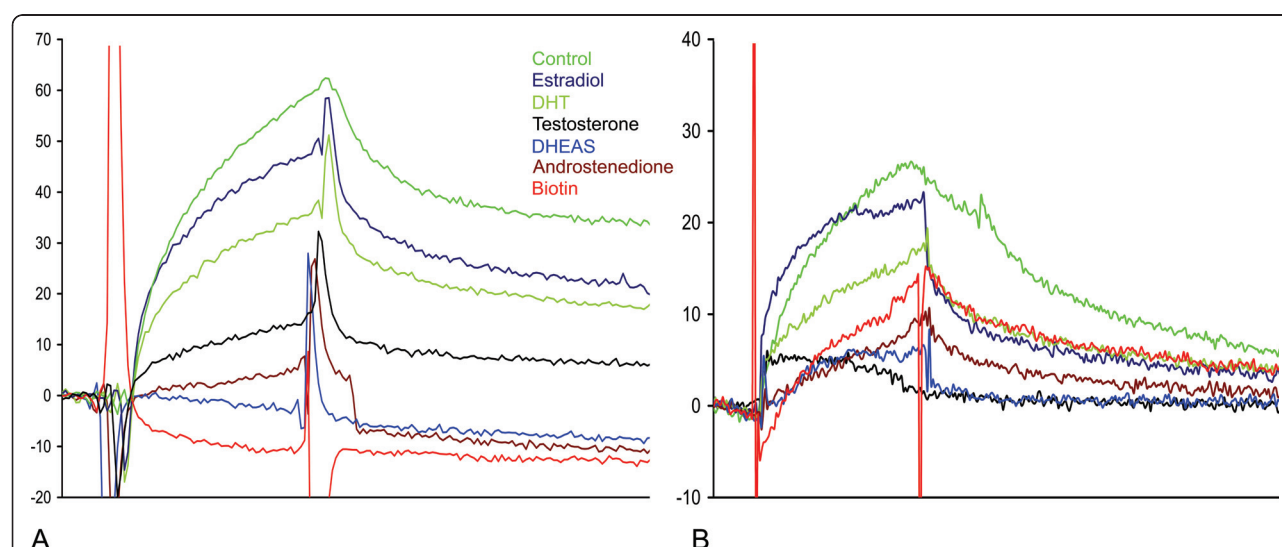
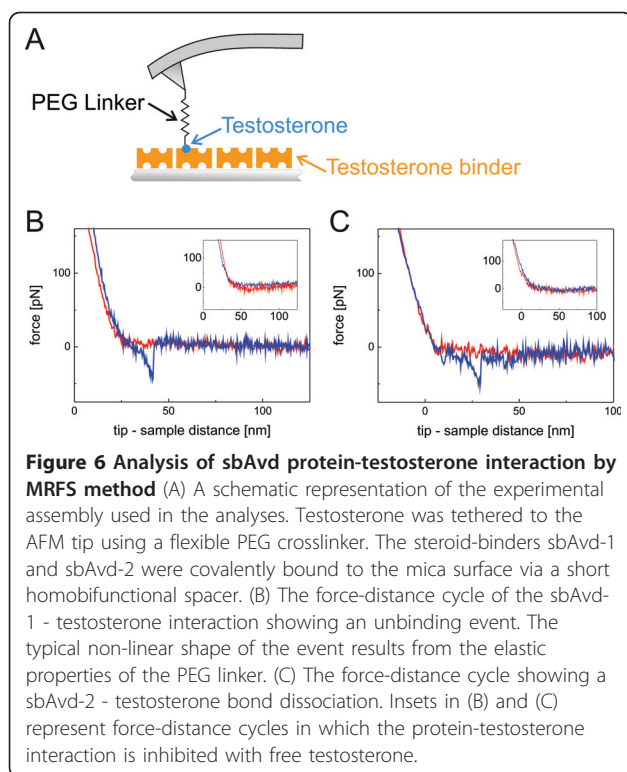


Figure 5 Inhibition analysis of sbAvd-1 and sbAvd-2 proteins by the SPR method. The binding of the sbAvd-1 and the sbAvd-2 to a CM5 sensor chip functionalized with testosterone-BSA was measured in the presence of 50 μ M inhibitors. (A) The binding of the sbAvd-1 protein was totally inhibited by dehydroepiandrosterone, androstenedione, and biotin. (B) The binding of the sbAvd-2 protein was totally inhibited by dehydroepiandrosterone or testosterone, but not biotin. This result is due to the markedly decreased affinity of the protein towards biotin. Samples: Protein sample, green; protein with estradiol, dark blue; protein with DHT, olive; protein with testosterone, black; protein with DHEAS, blue; protein with androstenedione, brown; protein with biotin, red.



detected (insets of Figure 6B and 6C), which indicated the total inhibition of the binding by free testosterone. Additionally, when the binding sites were preoccupied with free testosterone, the number of binding events dropped down to 67 from 183 detected interactions in the case of sbAvd-1. In case of sbAvd-2 the number of interactions dropped from 215 to 21 after addition of free testosterone.

Differential scanning calorimetry reveals the high stability of the engineered proteins

The high thermal stability of wtAvd (temperature transition midpoint (T_m) = 85.5°C) was preserved in the steroid-binding mutants; these proteins showed similar T_m values compared to wtAvd (sbAvd-1 T_m value of 80.6°C and sbAvd-2 T_m value of 82.5°C) (Table 1, Additional file 4). Ligand binding often stabilizes a protein and raises the T_m value. As expected, the addition of biotin stabilized wtAvd effectively, increasing the T_m value almost 40°C. This T_m indicates a very high binding affinity.

The effect of biotin-binding on the protein stability of steroid-binding proteins was much smaller than in the case of wtAvd, thus showing decreased biotin affinity. In fact, sbAvd-2 showed a negligible increase in T_m in the presence of biotin (ΔT_m = 0.5°C). The presence of testosterone only slightly increased the T_m values of sbAvds (ΔT_m = 0.6-0.9°C). Actually, a similarly small

increase was detected in the ΔT_m value of wtAvd when testosterone was added. However, because the testosterone-binding of wtAvd was not detected in other analyses (SPR and microplate analysis) the mechanism of stabilization remains unclear.

Discussion

In the present study, Avd proteins were displayed on the M13 phage in a monovalent form fused with the C-terminal region (aa residues 198-406) of the coat protein pIII. The crucial requirement for the functional display on phage is successful expression of the protein in *E. coli*. As has been earlier reported Avd protein can be efficiently expressed in a soluble form in *E. coli* [26]. The close relative of Avd, streptavidin, has been displayed on bacterial phages as a pVIII fusion [30,31]. However, to our best knowledge, chicken Avd has not been previously displayed on phages for screening purposes.

To evaluate the different modes for Avd display, we generated constructs expressing pIII fused to Avd and constructs that expressed free Avd subunits in addition to the pIII fusions. Based on the expression analysis, binding properties and selection experiments, the construct that expressed both fusion and free Avd showed an enhanced assembly of functional, tetrameric Avd on the phage compared to the Avd-pIII fusion alone (Additional file 2). However, the construct that expressed pIII-fused Avd was chosen for use in the selection system for the generation of gene libraries because the use of only one Avd gene in the protein display would ensure that the selected proteins could be assembled to homotetramers.

The potential of the developed platform for screening novel Avd-based receptors was investigated. Five residues in the loop between β -strands 1 and 2 of Avd were randomized to generate a population of diverse Avd genes. The gene population was screened for novel binding properties using the phage display method [23]. We observed the enrichment of Avd phages binding to testosterone-coupled BSA. The Avd variant sbAvd-1, which carried the sequence N12, R13, M14, N15, H16, was further analyzed and found to bind free testosterone with an affinity similar to BSA-conjugated testosterone. Moreover, we detected that the biotin-binding affinity of sbAvd-1 was not completely diminished; it was still high enough to inhibit sbAvd-1 from binding to testosterone-BSA (Table 1, Figures 4 and 5). This finding confirmed that testosterone occupies the same binding site in Avd as does biotin. Based on the cross-creativity measurements, sbAvd-1 also binds other steroid hormones, such as progesterone and DHEAS (Figures 4 and 5).

SbAvd-1 was further engineered to decrease biotin binding and to improve steroid specificity. From the

phage display selections, the sbAvd-1-derived protein, named sbAvd-2, with a mutated loop between β -strands 3 and 4 was captured. This mutant showed similar or slightly decreased binding affinity towards immobilized testosterone, whereas the biotin-binding affinity was clearly decreased (Table 1, Figures 4 and 5). We also noticed that sbAvd-2 bound more tightly to free testosterone than did sbAvd-1 (Figure 5). Significantly, sbAvd-2 preferred steroids as a ligand over biotin; this finding is important when considering the applications for biofluids, in which biotin is often present in relatively high concentrations.

This study and our preliminary results from experiments with a number of different target ligands (Hiltunen S, Riihimäki TA *et al.*, unpublished data) suggest that Avd-based receptors for various different small molecules can be tailored. These Avd-based receptors may be valuable tools for diagnostic use in the future.

Conclusions

The current study provides a promising platform for the selection of tailored ligand-binders evolved from the Avd scaffold in the monovalent pIII protein display. The Avd scaffold has characteristics that are beneficial in protein engineering, such as high thermal and chemical stability, simple folding and an optimal structure for small ligands. Furthermore, Avd can be modified rather freely without major change in the fold [18]. Novel Avds that can simultaneously bind multiple ligands could become next-generation molecular tools for clinical and diagnostic applications [22,32]. Importantly, novel Avd-based receptors could be used in applications that require harsh conditions.

Methods

Construction of phagemid vectors

All of the basic recombinant DNA methods were performed essentially as previously described [33]. Appropriate restriction sites were added to the cDNA of the Avd core sequence [34] by PCR using the primers Avd_NheI_5' and Avd_NotI_3' (Additional file 5). PCR products were first subcloned into the pCR[®]2.1-TOPO plasmid by TOPO TA-cloning (Invitrogen) and the plasmids were transformed into *E. coli* TOP10 cells. Plasmids were isolated from colonies that contained the inserts based on the blue-white screening. Avd fragments were cut out from the pCR[®]2.1-TOPO-plasmid using the NheI and NotI restriction enzymes and ligated into the phagemid vector (pBluescript SK+ derived phagemid, Research Center of Finland, Biotechnology, Espoo, Finland). The cDNA of the Avd was cloned into the phagemid vector as an N-terminal fusion to the C-terminal domain (amino acids 198-406) of the minor phage coat protein III. In the Avd/Avd-pIII constructs,

the coding sequence of the free Avd was subcloned using the primers Avd_NheI_5' and Avd_AscI_stop_3' (Additional file 5). The Avd-pIII expression cassette was generated by subcloning a fragment amplified from the primers Avd_SfiI_5' and Avd_NotI_3' (Additional file 5). The nucleotide sequences of the Avd constructs were verified by sequencing on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems) according to the protocols recommended by the manufacturer (ABI PRISM BigDye Terminator Cycle Sequencing Kit v.1.1, Applied Biosystems).

Amplification of phage particles

All of the basic phage display methods were performed essentially as previously described [35]. Phagemid vectors with the Avd insert were transformed into chemically competent *E. coli* XL1-Blue cells (Stratagene, La Jolla, CA) with the heat shock method. Phage stocks of the different Avd display constructs were made from individual colonies picked from the transformation plates into super broth (SB) medium supplemented with the appropriate antibiotics and glucose. The bacterial cultures were infected with the helper phage (10^{12} pfu/ml) VCS-M13 (Stratagene, LaJolla, CA) for amplification of Avd phages. Phages were PEG precipitated and analyzed by SDS-PAGE and western blotting following immunostaining with the polyclonal rabbit α -avd IgG (University of Oulu, 1:5000) and the monoclonal mouse anti-pIII IgG (Biosite, Sweden, 1:2000) antibodies.

Functionality test of Avd-displaying phages

The functionality of the Avd phages was tested by panning the phages on surfaces coated with BSA conjugated to HABA (HABA-BSA) [36]. As a positive control for selection, the Avd mutant N118M was also displayed on the phages. Panning was performed essentially as previously described [35]. For elution, vigorous shaking in 100 mM hydrochloric acid containing 10 μ M D-biotin (Biochemica, Fluka, 14400) was performed. In total, three selection rounds were performed. After each round the integrity of the Avd expression units was analyzed by restriction enzyme digestions of the phagemid DNA samples. Importantly, both single and double Avd and Avd(N118M) construct phages were competed against each other. Equivalent amounts of phages displaying Avd or Avd(N118M) mutant were mixed and biopanned. The ratio between constructs was determined with DNA sequencing.

Construction of Avd L1,2 library

To construct an Avd DNA library, amino acids N12, D13, L14, G15, and S16 in the loop between beta strands 1 and 2 (L1,2) were randomized. The libraries were constructed essentially as previously described

[35]. A nucleic acid fragment of 105 base pairs was amplified with the primers Avd_NheI_5' and Loop 1-2_R1_3' (Additional file 5) using wt Avd cDNA as template. Parallel to this fragment, a nucleic acid fragment with 357 base pairs was PCR-amplified with the primers Loop 1-2_R2_5' and Avd_NotI_3' (Additional file 5) using wt Avd as a template. The PCR strategy is presented in Additional file 6. The desired amplification products were separated by agarose gel electrophoresis and isolated from the gel using the Nucleo Spin Extract II kit (Macherey-Nagel) according to the manufacturer's instructions. PCR products were combined in the second amplification step in the presence of the PCR primers Avd_NheI_5' and Avd_NotI_3' (Additional file 5). This amplification resulted in a DNA fragment of 462 base pairs. The fragment was isolated from the gel and cut with the restriction enzymes *NheI* and *NotI* (Fermentas) followed by purification using the Nucleo Spin Extract II kit (Macherey-Nagel). Fragments were ligated into the phagemid vector and the resulting ligation product was transformed into electrocompetent cells of the *E. coli* strain XL1-Blue (Stratagene) by electroporation. Transfected bacteria were infected with the VCS-M13 helper phage (Stratagene) and phages were harvested from the culture and purified with PEG precipitation. The amount of phage particles was determined by titration.

Selecting steroid-binders from the Avd L1,2 library

The Avd L1,2 library was panned against the steroid hormone testosterone. As a negative control, wtAvd-displaying phages were also panned against testosterone. NUNC immunosorp plates were coated with a testosterone-BSA conjugate (Sigma, T-3392, 1 µg). Phages were preincubated in the BSA-coated wells to prevent non-specific binding. Three to four selection cycles were conducted, and the stringency of washing conditions was increased every panning round to decrease nonspecific binding. In the first panning round wells were washed three times with phosphate buffered saline containing 0.05% Tween (PBS-Tween) and five times with phosphate buffered saline (PBS). Phages were eluted by vigorous shaking for 10 minutes in 100 mM hydrochloric acid containing 10 µM D-biotin (Biochemica, Fluka, 14400). Biotin was used for elution because it was probable that after randomization of the loop L1,2 the resulting Avd would still have a rather high affinity towards biotin. In the final panning round 10 µM testosterone (Steraloids Inc., USA) was used in addition to 100 mM hydrochloric acid for elution. The eluted phage solutions were neutralized with 2 M Tris. Eluted phages were used to infect *E. coli* XL1-Blue cells and aliquots of the infected bacteria were plated to quantify the amount of eluted phages. Phages were amplified and purified as

described earlier. Precipitated phages were then used for the next round of selection. After every panning round, results were verified by sequencing (20 sequences), and the number of phage particles was determined by phage titration.

Generation and screening of the SbAvd-1 L3,4 library

An SbAvd-1-derived DNA library was constructed and ligated into the phagemid as described above. In the library amino acids T35, A36, V37, and T38 in the loop between beta strands 3 and 4 (L3,4) were randomized; thus, in the construction of the library the primers Avd_NheI_5' and 3_4R_1_3' and primers 3_4R_2_5' and Avd_NotI_3' (Additional file 5) were used in the PCR. SbAvd-1 cDNA was used as a template. Ligated phagemid was transformed into electrocompetent cells of *E. coli* strain XL1-Blue (Stratagene) by electroporation. Transfected bacteria were infected with the VCS-M13 helper phage (Stratagene) and phages were harvested from the culture and purified with PEG precipitation. The amount of phage particles was determined by titration.

The sbAvd-1 L3,4 library was biopanned against the steroid hormone testosterone similarly as described above, with some exceptions. NUNC immunosorp plates were coated with a testosterone-BSA conjugate (Sigma, T-3392, 200 ng). Phages were preincubated in BSA-coated wells before the panning procedure. Four selection cycles were performed. Phages were eluted by vigorous shaking for 10 minutes in 100 mM hydrochloric acid containing 35 µM testosterone (Steraloid Inc., USA). After each panning round, results were verified by sequencing (20 sequences), and the number of phage particles was determined by phage titration. Additionally, every panning round was screened for binders by anti-M13 as previously described [37]. For the ELISA, (NUNC) immunosorp plates were coated with a testosterone-BSA conjugate (Sigma, T-3392, 1 µg), wells were blocked with 5% milk solution, and detection was performed with Anti-M13/HRP (GE Healthcare) and read with a microplate reader (Bio-Rad 680 XR).

Production and purification of Avd mutants

For biochemical analyses, the proteins were produced in *E. coli* strain BL21-AI (Invitrogen) using expression vector pET101/D (Invitrogen) [26]. After sonication (Sonics & Materials Vibra Cell™) and DNaseI (New England Bio Labs) treatment of *E. coli* cells, the purification of the proteins was conducted using Ni-NTA affinity chromatography according to the instructions of the manufacturer (QIAGEN).

The oligomeric state of the proteins was assayed with fast protein liquid chromatography (FPLC) gel-filtration using an ÄKTApurifier™ HPLC equipped with a

Superdex 200 10/300 GL column (Tricorn, Amersham Biosciences, GE Healthcare). The column was calibrated using the gel-filtration mixture (thyroglobulin, γ -globulin, ovalbumin, myoglobin, and vitamin B12; Bio-Rad Laboratories) as a molecular-mass standard. Sodium phosphate buffer (20 mM, pH 7.4) with 1M NaCl and, 20 mM imidazole was used as the liquid phase. Protein samples of 90-193 μ g in a volume of 500 μ l were used in the analysis.

DSC measurements of wtAvd and steroid-binding Avds

Proteins (0.225 mg/ml) were analyzed in sodium phosphate buffer (20 mM, pH 7.4), containing 20 mM imidazole and 1 M NaCl. D-biotin (Biochemica, Fluka, 14400) and testosterone (Steraloids Inc., USA) were diluted with the measurement buffer to a final concentration of 50 μ M. All solutions were degassed prior to measurements to avoid air bubbles. An automated capillary VP-DSC instrument (GE Healthcare, MicroCal, Northampton, USA) was used to measure the stability of the proteins with or without ligands. During the measurement, protein samples were heated from 20°C to 130°C at a scanning rate of 120°C/h. Feedback mode was set to low and the filler period was 8 s. Temperature transition midpoints (T_m) were recorded from the highest peaks and the calorimetric heat changes (ΔH) were calculated using the MicroCal Origin 7 software (GE Healthcare, MicroCal, Northampton, USA).

Determination of ligand-binding specificity of Avd forms by microplate assay

MaxiSorp F96 microplate wells (NUNC) were coated with 500 ng of conjugated ligand (HSA-conjugated progesterone, hydrocortisone, testosterone, and cholic acid (Technical Research Center of Finland, Espoo, Finland); or BSA-conjugated testosterone (A6958-000, Steraloids Inc., USA), and biotin (Jenni Leppiniemi, University of Tampere, Finland)), and with 500 ng of the carrier proteins HSA and BSA (A7906, Sigma) in 100 μ l of PBS for 2 hours 37°C. Carrier proteins were used as negative controls. Plates were washed three times with PBS-Tween, blocked with 0.5% BSA-PBS for 30 min, and then washed again. Proteins (0.9 μ g/ml) in sodium phosphate buffer (20 mM, pH 7.4) with 1M NaCl and 20 mM imidazole, were added to the wells and incubated for 1 h. A portion of wtAvd (Belovo, Bastogne, Belgium), sbAvd-1 and sbAvd-2 was preincubated with 10 μ M D-biotin (Biochemica, Fluka, 14400). Bound Avd was detected by rabbit α -avd IgG (University of Oulu) and with alkaline phosphatase conjugated goat anti-rabbit IgG (A3937, Sigma). After adding the phosphatase

substrate solution (1 mg/ml pNPP (S0942, Sigma) in 1 M diethanolamine pH 9.8, with 0.5 mM $MgCl_2$), the plates were read after 15 minutes at A405 with a microplate reader (Bio-Rad 680 XR).

Biosensor analyses of steroid-binding Avds

A BiAcCore X optical biosensor (Biacore, Uppsala, Sweden) was used for the analysis of binding kinetics. Testosterone-BSA was coupled to the carboxymethylated dextran layer of a sensor chip using standard amine coupling chemistry (1000 RU, 40 μ l/min flow rate). Samples of sbAvd-1, and sbAvd-2 were diluted in 50 mM sodium phosphate containing 1 M NaCl, and the same buffer was used in the measurements. The binding of the sbAvd-1 and sbAvd-2 samples on testosterone-BSA coated chips was measured and the kinetic constants were determined from the measurements performed with different protein concentrations using the BiaEvaluation software according to the manufacturer's instructions. WtAvd was used as a negative control.

The binding of steroid-binding proteins to the testosterone-BSA surface was competed with free steroid hormone molecules (testosterone, DHEAS, androstenedione, estradiol, and DHT (Steraloids Inc., USA)) and free biotin (Biochemica, Fluka, 14400) to evaluate the specificity of binding. Additionally, steroid-binding was more closely detected by measuring the binding in the presence of varying concentrations (0.75 μ M-50 μ M) of inhibiting testosterone.

For the determination of biotin-binding kinetics, a sensor chip was prepared as follows: diaminoethylene was first attached to the surface using a mixture containing 0.2 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and 0.05 M N-hydroxysuccinimide (NHS) in water. Second, to introduce amino groups to the surface, 1 M ethylenediamine (Fluka 03550) in water was applied. Finally, 5 mM biotin N-succinimidyl ester (Biochemica, Fluka, 14405) in 50% DMSO was injected on the surface (40 μ l/min flow rate, ~130 RU; please note that the determination of the bound mass is not very accurate in case of small molecules because the immobilization can change the physicochemical properties of the surface).

Molecular Recognition Force Spectroscopy experiments

Molecular recognition force spectroscopy (MRFS) was used to study the interaction of the produced proteins with testosterone and biotin. Steroid-binding proteins were covalently bound to modified mica sheets via lysines as previously described [38]. Testosterone was coupled to the AFM tip via the heterobifunctional Fmoc-PEG-NHS crosslinker as previously described

[28], resulting in a covalent attachment of testosterone via a flexible spacer (Figure 6A).

All MRFS experiments were performed on a Pico SPM I (Agilent Technologies, Santa Clara, CA). All modified cantilevers used had nominal spring constants between 10-30 pN/nm (Veeco Instruments, Santa Barbara, CA). The effective spring constants of the cantilevers were determined by the thermal noise method [39,40]. Force-distance cycles were completed using a z-range of 200 and 300 nm. Sweep durations were adjusted between 0.25 and 4 s. During one data set of 1000 force-distance curves, the lateral tip position was changed (a few hundred nm) about every 100 curves to ensure that the binding events were statistically reasonable. The specificity of the binding was proved by adding free testosterone (200 nM) into the measuring solution and incubating for approximately 1 h to block the ligand-binding sites of the proteins.

Additional material

Additional file 1: Structural comparison of Avd-BTN and Avd-HABA complexes. 3D-structures of Avd complexed with biotin (A) and HABA (B). These two ligands have different hydrogen bonding interactions with the amino acids at the binding site. (A) Hydrogen bonds formed between the ureido ring of biotin and Avd are shown here as dashed green lines. (X-ray crystallographic structure (PDB 2AVI)) [41]. (B) A significant difference between biotin and HABA binding to Avd is seen in the interaction of the ligand with asparagine 118. Although other Avd key residues interacting with the BTN ureido ring also interact with HABA, there is no hydrogen bonding partner for N118 in HABA. Again, hydrogen bonds formed between the carboxyl group of HABA and Avd are shown as dashed green lines. The loop between β -strands 3 and 4 was not resolved in the avd-HABA complex (coordinates kindly provided by Prof. Oded Livnah). The figure was made with the VMD program [42].

Additional file 2: Sequencing results and input from the control selections. The percentage of wt Avd and Avd(N118M) mutant sequences after sequencing analysis from the different rounds of HABA selection. The amount of input phages is shown as colony forming units (cfu) per milliliter of culture.

Additional file 3: Gel-filtration chromatograms of sbAvd-1 and sbAvd-2 proteins. The chromatograms of steroid-binding proteins determined at wavelength of 280 nm by gel-filtration. Besides the main peak, there is also a small peak of oligomeric form of the protein in the chromatogram of sbAvd-1 protein (gray curve). It is a typically observation also in the case of wtAvd [26]. The chromatogram of sbAvd-2 protein is shown with black curve. The tetrameric form of the protein is dominant in both samples.

Additional file 4: The effect of ligand-binding to the stability of wtAvd, sbAvd-1 and sbAvd-2 proteins. DSC thermograms were obtained from the protein sample (0.225 mg/ml) by scanning temperature range of 20°C to 130°C with heating rate of 120°C/min. The analysis was conducted in the absence and presence of ligands (50 μ M).

Additional file 5: The cloning primers used in the study. The restriction enzyme cleavage sites are indicated in *italics*.

Additional file 6: A schematic presentation of the construction strategy of the Avd L1,2 library. for 1-2 loop library a nucleic acid fragment of 105 base pairs was amplified (Step 1, A) with the primers Avd_NheI_5' and Loop 1-2_R1_3' (schematically referred in the figure as Avd_mutant_3') using wtAvd cDNA as a template. Parallel to this, a nucleic acid fragment with 357 base pairs, was PCR-amplified (Step 1, B) with the primers Loop 1-2_R2_5' (schematically referred in the figure as

Avd_combine_5') and Avd_NotI_3'; also using wtAvd as a template. Amplified fragments were combined in a second amplification step in the presence of PCR primers Avd_NheI_5' and Avd_NotI_3', wherein a DNA fragment of 462 base pairs was obtained.

Abbreviations

AFM: atomic force microscopy; BSA: bovine serum albumin; BTN: D-biotin; DHEAS: dehydroepiandrosterone sulphate; DHT: dihydrotestosterone; DMSO: dimethyl sulfoxide; EDC: 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride; HABA: 4'-hydroxyazobenzene-2-carboxylic acid; HSA: human serum albumin; MRFS: molecular recognition force spectroscopy; NHS: N-hydroxysuccinimide; PBS: phosphate buffered saline; PEG: polyethylene glycol; sbAvd: steroid-binding avidin; SB: super broth; SDS-PAGE: sodium-dodecyl sulfate polyacrylamide gel electrophoresis; wtAvd: wild type avidin.

Acknowledgements

We thank the laboratory staff of VTT (Technical Research Center of Finland) Biotechnology, Espoo, Finland, for excellent technical assistance. We thank MSc. Antti Tullila for kindly providing us HSA-conjugates. We also thank the staff at the Molecular Biology group, Department of Biological and Environmental Science, University of Jyväskylä for their assistance. We thank MSc. Sampo Kukkuriainen for help in molecular modeling. We would also like to thank MSc. Kaisa Helttunen, Professor Kari Rissanen and Dr. Juhani Huuskonen for their help in the preparation of the HABA-BSA conjugate. We thank Hong Chang for participation in the experimental work. The study was financially supported by the Academy of Finland (115976, 121236); the National Technology Agency of Finland (BioFace 40055/08); European Micro and Nano Technology support program (FFG 421695); Pirkanmaa Hospital District and Tampere Graduate Program in Biomedicine and Biotechnology.

Author details

¹Institute of Biomedical Technology, University of Tampere and Tampere University Hospital, FI-33520 Tampere, Finland. ²Institute of Biophysics, Johannes Kepler University Linz, 4040 Linz, Austria. ³VTT Technical Research Centre of Finland, FI-02044 VTT, Finland.

Authors' contributions

TAR conceived the study, constructed the Avd libraries, performed most of the experiments, and contributed to the writing of the manuscript; SH participated in the construction of the Avd libraries and in the experimental work and contributed to the writing of the manuscript; JAEM participated in the experimental work and contributed to the writing of the manuscript; MR, AE and PH performed MRFM measurements and contributed to the writing of the manuscript; HRN, MSK and KT conceived the study and supervised the work; VPH conceived and designed the study, supervised the work and contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Received: 29 April 2011 Accepted: 9 June 2011 Published: 9 June 2011

References

- Hoogenboom HR: Selecting and screening recombinant antibody libraries. *Nat Biotechnol* 2005, **23**:1105-1116.
- Hoess RH: Protein design and phage display. *Chem Rev* 2001, **101**(10):3205-3218.
- Sarikaya M, Tamerler C, Jen AK, Schulten K, Baneyx F: Molecular biomimetics: nanotechnology through biology. *Nat Mater* 2003, **2**:577-585.
- Binz HK, Amstutz P, Pluckthun A: Engineering novel binding proteins from nonimmunoglobulin domains. *Nat Biotechnol* 2005, **23**:1257-1268.
- Skerra A: Alternative non-antibody scaffolds for molecular recognition. *Curr Opin Biotechnol* 2007, **18**:295-304.
- Werner RG: Economic aspects of commercial manufacture of biopharmaceuticals. *J Biotechnol* 2004, **113**:171-182.
- Hudson PJ, Souriau C: Engineered antibodies. *Nat Med* 2003, **9**:129-134.
- Saerens D, Ghassabeh GH, Muyldermans S: Single-domain antibodies as building blocks for novel therapeutics. *Curr Opin Pharmacol* 2008, **8**:600-608.

9. Nygren PA, Skerra A: **Binding proteins from alternative scaffolds.** *J Immunol Methods* 2004, **290**:3-28.
10. Gebauer M, Skerra A: **Engineered protein scaffolds as next-generation antibody therapeutics.** *Curr Opin Chem Biol* 2009, **13**:245-255.
11. 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 USA* 1999, **96**:1898-1903.
12. Schlehuber S, Skerra A: **Tuning ligand affinity, specificity, and folding stability of an engineered lipocalin variant – a so-called 'anticalin' – using a molecular random approach.** *Biophys Chem* 2002, **96**:213-228.
13. Laitinen OH, Nordlund HR, Hytönen VP, Kulomaa MS: **Brave new (strept) avidins in biotechnology.** *Trends Biotechnol* 2007, **25**:269-277.
14. Green NM: **Avidin.** *Adv Protein Chem* 1975, **29**:85-133.
15. Repo S, Paldanius TA, Hytönen VP, Nyholm TK, Halling KK, Huuskonen J, Pentikäinen OT, Rissanen K, Slotte JP, Airene TT, Salminen TA, Kulomaa MS, Johnson MS: **Binding Properties of HABA-Type Azo Derivatives to Avidin and Avidin-Related Protein 4.** *Chem Biol* 2006, **13**:1029-1039.
16. Livnah O, Bayer EA, Wilchek M, Sussman JL: **Three-dimensional structures of avidin and the avidin-biotin complex.** *Proc Natl Acad Sci USA* 1993, **90**:5076-5080.
17. Rosano C, Arosio P, Bolognesi M: **The X-ray three-dimensional structure of avidin.** *Biomol Eng* 1999, **16**:5-12.
18. Laitinen OH, Hytönen VP, Nordlund HR, Kulomaa MS: **Genetically engineered avidins and streptavidins.** *Cell Mol Life Sci* 2006, **63**:2992-3017.
19. Nordlund HR, Laitinen OH, Hytönen VP, Uotila ST, Porkka E, Kulomaa MS: **Construction of a dual chain pseudotetrameric chicken avidin by combining two circularly permuted avidins.** *J Biol Chem* 2004, **279**:36715-36719.
20. Hytönen VP, Nordlund HR, Hörhå J, Nyholm TK, Hyre DE, Kulomaa T, Porkka EJ, Marttila AT, Stayton PS, Laitinen OH, Kulomaa MS: **Dual-affinity avidin molecules.** *Proteins* 2005, **61**:597-607.
21. Riihimäki TA, Kukkurainen S, Varjonen S, Hörhå J, Nyholm TKM, Kulomaa MS, Hytönen VP: **Construction of chimeric dual-chain avidin by tandem fusion of the related avidins.** *PLoS One* 2011, **6**:e20535.
22. Nordlund HR, Hytönen VP, Hörhå J, Maatta JA, White DJ, Halling K, Porkka E, Slotte JP, Laitinen OH, Kulomaa MS: **Tetravalent single chain avidin: From subunits to protein domains via circularly permuted avidins.** *Biochem J* 2005, **392**:485-491.
23. Smith GP: **Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface.** *Science* 1985, **228**:1315-1317.
24. Hoogenboom HR, Griffiths AD, Johnson KS, Chiswell DJ, Hudson P, Winter G: **Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains.** *Nucleic Acids Res* 1991, **19**:4133-4137.
25. Määttä JA, Airene TT, Nordlund HR, Jänis J, Paldanius TA, Vainiotalo P, Johnson MS, Kulomaa MS, Hytönen VP: **Rational modification of ligand-binding preference of avidin by circular permutation and mutagenesis.** *Chembiochem* 2008, **9**:1124-1135.
26. Hytönen VP, Laitinen OH, Airene TT, Kidron H, Meltola NJ, Porkka E, Hörhå J, Paldanius T, Määttä JA, Nordlund HR, Johnson MS, Salminen TA, Airene KJ, Ylä-Herttua S, Kulomaa MS: **Efficient production of active chicken avidin using a bacterial signal peptide in Escherichia coli.** *Biochem J* 2004, **384**:385-390.
27. Hinterdorfer P, Dufrene YF: **Detection and localization of single molecular recognition events using atomic force microscopy.** *Nat Methods* 2006, **3**:347-355.
28. Wildling L, Hinterdorfer P, Kusche-Vihrog K, Treffner Y, Oberleithner H: **Aldosterone receptor sites on plasma membrane of human vascular endothelium detected by a mechanical nanosensor.** *Pflügers Arch* 2009, **458**:223-230.
29. Baumgartner W, Hinterdorfer P, Schindler H: **Data analysis of interaction forces measured with the atomic force microscope.** *Ultramicroscopy* 2000, **82**:85-95.
30. Avrantinis SK, Stafford RL, Tian X, Weiss GA: **Dissecting the streptavidin-biotin interaction by phage-displayed shotgun scanning.** *Chembiochem* 2002, **3**:1229-1234.
31. Sidhu SS, Weiss GA, Wells JA: **High copy display of large proteins on phage for functional selections.** *J Mol Biol* 2000, **296**:487-495.
32. Leppiniemi J, Määttä JA, Hammaren H, Soikkeli M, Laitaoja M, Jänis J, Kulomaa MS, Hytönen VP: **Bifunctional avidin with covalently modifiable ligand binding site.** *PLoS One* 2011, **6**:e16576.
33. Sambrook J, Fritsch EF, Maniatis T: **Molecular Cloning: A Laboratory Manual.** Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; 1990.
34. Gope ML, Keinänen RA, Kristo PA, Conneely OM, Beattie WG, Zarucki-Schulz T, O'Malley BW, Kulomaa MS: **Molecular cloning of the chicken avidin cDNA.** *Nucleic Acids Res* 1987, **15**:3595-3606.
35. Barbas CF III, Burton DR, Scott JK, Silverman GJ: **Phage Display, A Laboratory manual.** Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; 2001.
36. Hofstetter H, Morpurgo M, Hofstetter O, Bayer EA, Wilchek M: **A labeling, detection, and purification system based on 4-hydroxyazobenzene-2-carboxylic acid: an extension of the avidin-biotin system.** *Anal Biochem* 2000, **284**:354-366.
37. Kingsbury GA, Junghans RP: **Screening of phage display immunoglobulin libraries by anti-M13 ELISA and whole phage PCR.** *Nucleic Acids Res* 1995, **23**:2563-2564.
38. Kamruzzahan AS, Kienberger F, Stroh CM, Berg J, Huss R, Ebner A, Zhu R, Rankl C, Gruber HJ, Hinterdorfer P: **Imaging morphological details and pathological differences of red blood cells using tapping-mode AFM.** *Biol Chem* 2004, **385**:955-960.
39. Hutter JL, Bechhoefer J: **Calibration of atomic-force microscope tips.** *Review of Scientific Instruments* 1993, **64**:1868-1873.
40. Butt HJ, Jaschke M: **Calculation of thermal noise in atomic force microscopy.** *Nanotechnology* 1995, **6**:1-7.
41. Livnah O, Bayer EA, Wilchek M, Sussman JL: **Three-dimensional structures of avidin and the avidin-biotin complex.** *Proc Natl Acad Sci USA* 1993, **90**:5076-5080.
42. Humphrey W, Dalke A, Schulten K: **VMD: visual molecular dynamics.** *J Mol Graph* 1996, **14**:33-8, 27-8.

doi:10.1186/1472-6750-11-64

Cite this article as: Riihimäki et al.: Modification of the loops in the ligand-binding site turns avidin into a steroid-binding protein. *BMC Biotechnology* 2011 **11**:64.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



Short Communication

Efficient preparation of shuffled DNA libraries through recombination (Gateway) cloning

Soili I. Lehtonen^{1,†}, Barbara Taskinen^{1,2,†}, Elina Ojala^{1,3},
Sampo Kukkurainen^{1,2}, Rolle Rahikainen^{1,2}, Tiina A. Riihimäki¹,
Olli H. Laitinen⁴, Markku S. Kulomaa^{1,5}, and Vesa P. Hytönen^{1,2,*}

¹BioMediTech, FI-33014 University of Tampere, Biokatu 6, Finland, ²Fimlab Laboratories, Biokatu 4, FI-33520 Tampere, Finland, ³Next Biomed Technologies NBT Oy, Viikinkaari 6, FI-00790 Helsinki, Finland, ⁴The Center for Infectious Medicine, Department of Medicine HS, Karolinska Institute, Karolinska University Hospital, Huddinge F59, SE-141 86 Stockholm, Sweden, and ⁵Tampere University Hospital, FI-33521 Tampere, Finland

*To whom correspondence should be addressed. E-mail: vesa.hytönen@uta.fi

[†]The authors contributed equally to the article.

Edited by Arne Skerra

Received 27 January 2014; Revised 4 September 2014; Accepted 25 October 2014

Abstract

Efficient and robust subcloning is essential for the construction of high-diversity DNA libraries in the field of directed evolution. We have developed a more efficient method for the subcloning of DNA-shuffled libraries by employing recombination cloning (Gateway). The Gateway cloning procedure was performed directly after the gene reassembly reaction, without additional purification and amplification steps, thus simplifying the conventional DNA shuffling protocols. Recombination-based cloning, directly from the heterologous reassembly reaction, conserved the high quality of the library and reduced the time required for the library construction. The described method is generally compatible for the construction of DNA-shuffled gene libraries.

Key words: DNA library, DNA shuffling, directed evolution, phage display recombination cloning

Introduction

DNA (family) shuffling generates diversity through recombination and is a powerful tool for the directed evolution of proteins. Useful mutations which have been previously shown to be functional in nature are combined from individual genes (Stemmer, 1994a, b; Cramer *et al.*, 1998). Generally, conventional digestion–ligation methods are used in the construction of the shuffled DNA libraries, but these methods involve cumbersome purification and amplification steps which radically reduce the diversity of the shuffled library. Thus, the generation of highly diverse DNA libraries represents a bottleneck in the directed evolution approach (Lutz and Patrick, 2004; Wang *et al.*, 2006).

The robust and highly efficient Gateway cloning method (Hartley *et al.*, 2000) is a unique, *in vitro*, site-specific recombination technology based on the well-characterized recombination system of bacteriophage lambda. This method facilitates efficient cloning and

high-throughput DNA transfer, while maintaining the orientation and reading frame of the fragment(s) of interest (Hartley *et al.*, 2000). Since its introduction, Gateway cloning has been widely used in different applications such as systems biology (Brasch *et al.*, 2004; Marsischky and LaBaer, 2004; Katzen, 2007; Festa *et al.*, 2013) and life science applications, including the cloning of the cDNA libraries. Gateway cloning is also a popular method for the generation of protein expression vectors (Structural Genomics Consortium *et al.*, 2008), and it has also been used to construct a phage display library (Gao *et al.*, 2008). However, there are only few examples of directed evolution studies where Gateway cloning was utilized (van den Berg *et al.*, 2006; Hedde and Mazaleyrat, 2007; Gruet *et al.*, 2012).

In the present paper, we describe a straightforward method for the construction of shuffled DNA libraries using Gateway cloning directly from a reassembly reaction, which streamlines the cloning procedure and increases the diversity of the library. As a proof of the concept, our

model avidin library, constructed of the avidin-related genes AVR2 and AVR4, was turned into a phage display library and its biopanning resulted in avidins expressing more efficiently than parental proteins.

Results and discussion

We constructed a Gateway cloning compatible phagemid vector (pGW phagemid) containing the Gateway cassette upstream of the C-terminus of *pIII* gene. The Gateway cassette consists of the *attR*-flanked *ccdB* gene and the chloramphenicol resistance gene (*Cm^R*) (Supplementary Fig. S1). The construct enables the Gateway cloning of *attL*-flanked gene of interest in frame with the C-terminal pIII fragment (aa 198–406) (Fig. 1C). This allows the target protein to be displayed on the surface of bacteriophage M13 as a pIII fusion protein. The Gateway cassette was subcloned into a phagemid, replacing the original sequence consisting of the *pelB* signal peptide gene and the *Fab* gene (Supplementary Fig. S2).

For the construction of the model library, we used the genes encoding for avidin-related proteins AVR2 and AVR4, which share ~95% DNA sequence identity and 82% identity in their primary protein structures. Their tertiary and quaternary structures are highly similar; however, these genes were selected because of the fundamental differences in their biotin-binding and physicochemical properties (Laitinen *et al.*, 2002; Eisenberg-Domovich *et al.*, 2005; Hytönen *et al.*, 2005). The possible improvements that can be achieved by the DNA shuffling of avidin proteins have previously been successfully explored in Niederhauser *et al.* (2012). To eliminate parental products in the LR cloning reaction (excisive reaction (*attL* × *attR* → *attB* + *attP*)), we used the asymmetric templates *attL1-OmpA-AVR4* and *AVR2-attL2* as the starting materials for the DNA shuffling reaction (Supplementary Fig. S3) (Ikeuchi *et al.*, 2003). Figure 1 shows the consecutive steps of DNA shuffling confirmed by agarose gel electrophoresis (AGE) analysis. As shown in the schematic figure, the conventional approach was simplified as no gel extractions or amplification steps were performed. The reassembly product was directly used in the LR cloning reaction, which streamlines the approach.

We show in Table I that the streamlined Gateway-based DNA shuffling enables transformation efficiency of 9.0×10^6 cfu/μg by using highly competent cells (ElectroMAX™ DH5α-E™ Competent cells, Invitrogen). By comparison, PCR-amplified reassembly reaction and a gel-purified *attL*-flanked positive control resulted in transformation efficiencies of 1.5×10^8 and 1.8×10^8 cfu/μg respectively. The lower transformation activity directly after reassembly might reflect the relatively low amount of *attL*-flanked genes in the reaction. Alternatively, partially reassembled DNA fragments may inhibit the LR reaction. However, reasonable transformation efficiency can be obtained even without amplification. In practice, 1 μg corresponds to seven standard LR cloning reactions, making the protocol feasible in terms of reagent costs (LR-clonase kit is sufficient for 20 reactions; Gateway® LR Clonase® Enzyme mix, Invitrogen™, Thermo Fisher Scientific, Inc.).

Gateway cloning stems on *ccdB* suicide gene, theoretically restricting the use of the methodology for cell lines not bearing F' episome. However, we show that it is possible to construct a phage display library using F' episome bearing *Escherichia coli* XL1-Blue cells: No empty vectors were found from sequenced ($n = 45$) clones of the DNA library prepared by transforming XL1-Blue cells directly after LR cloning of the reassembly reaction (Table I). We also compared the transformation efficiency of *ccdB*-containing (empty) pGW phagemid to conventional phagemid vector and found only <0.1% transformants (data not shown). Therefore, although XL1-Blue cells can tolerate low levels of

ccdB, *ccdB*-bearing phagemids may be quickly removed by the negative selection pressure during the following selection step.

The theoretical library size of the shuffled AVR2/AVR4 library is relatively small (possible unique clones $2^{23} = 8.4 \times 10^6$), due to a small amount of differences between the parental genes. Table I shows that even bigger sequence spaces can be covered using the approach. Moreover, if non-proofreading DNA polymerase such as Taq (Stemmer, 1994a) is used in the amplification step, incorporated point mutations will add further diversity to the library. This was also demonstrated here, since the PCR-amplified library shows an increased amount of point mutations (Table I).

To demonstrate the feasibility of the method in phage display, a model phage library was constructed. The LR cloning reactions were first transformed to *E. coli* DH5α cells to remove possible non-recombinants and to amplify the DNA-shuffled AVR2/AVR4 gene library. Due to the use of relatively inefficient competent cells in this experiment, the size of the library was 2.4×10^5 cfu, and based on transformation efficiency 7.0×10^4 cfu/μg. The constructed model library covered the sequence space of the AVR2/AVR4 library (theoretical size 8.4×10^6) to some extent. The plasmids were subsequently extracted from DH5α cells and used to transform *E. coli* XL1-Blue cells to construct the phage display library, with a transformation efficiency of 9.0×10^5 cfu/μg.

The quality of the AVR2/AVR4-shuffled model library was analyzed through sequencing; both after the first transformation into *E. coli* DH5α cells and after the subsequent transformation into *E. coli* XL1-Blue cells (native phage display library). Supplementary Table SI shows an overview of the sequencing results (for complete sequence alignments of the clones in DH5α and XL1-Blue cells, see Supplementary Figs S4 and S5, respectively). The library showed almost exclusively correctly cloned inserts. All of the clones were shuffled at the DNA level, while some of them (<10%) showed translated sequences identical to the parental protein which, after careful analysis of the DNA sequences, were found to be resulting from DNA shuffling because of high similarity of the parental sequences. The diversity of the library was high, having 75% unique clones. The clones in the phage display library had on average 2.8 crossovers based on translated sequences (Supplementary Fig. S6). Importantly, according to our analysis, the additional transformation into *E. coli* DH5α cells did not affect the diversity of the library. A low amount of the analyzed clones contained random point mutations that were not present in any of the parental genes. The mutation frequency in *E. coli* XL1-Blue clones was higher than the frequency in *E. coli* DH5α clones (Supplementary Table I). Nevertheless, the mutation frequencies per amino acid residue determined at the protein level were 0.1% for XL1-Blue clones and 0.07% for DH5α clones, respectively, which were in similar range with the previously reported 0.7% mutation frequency per nucleotide (Stemmer, 1994a).

The model library was screened for functional biotin-binding proteins by phage display. An enrichment of biotin-binding-specific phages was observed already after three rounds of panning (Supplementary Table SII). Throughout the panning process the integrity of the pGWphagemid vector was assessed by isolation of the plasmids from the output phages and subsequent PCR amplification of the inserts (Fig. 2A). The original phage display library and the pool of phages obtained after each panning round were screened for active, biotin-binding chimeric AVR2/AVR4 proteins using both a microplate assay (Fig. 2B) and a dot blot method (Fig. 2C). After the third and fourth panning rounds, clear binding to biotinylated casein-coated wells was detected, and only a low background signal was observed (Fig. 2B). Similar results were obtained through dot blot analysis (Fig. 2C).

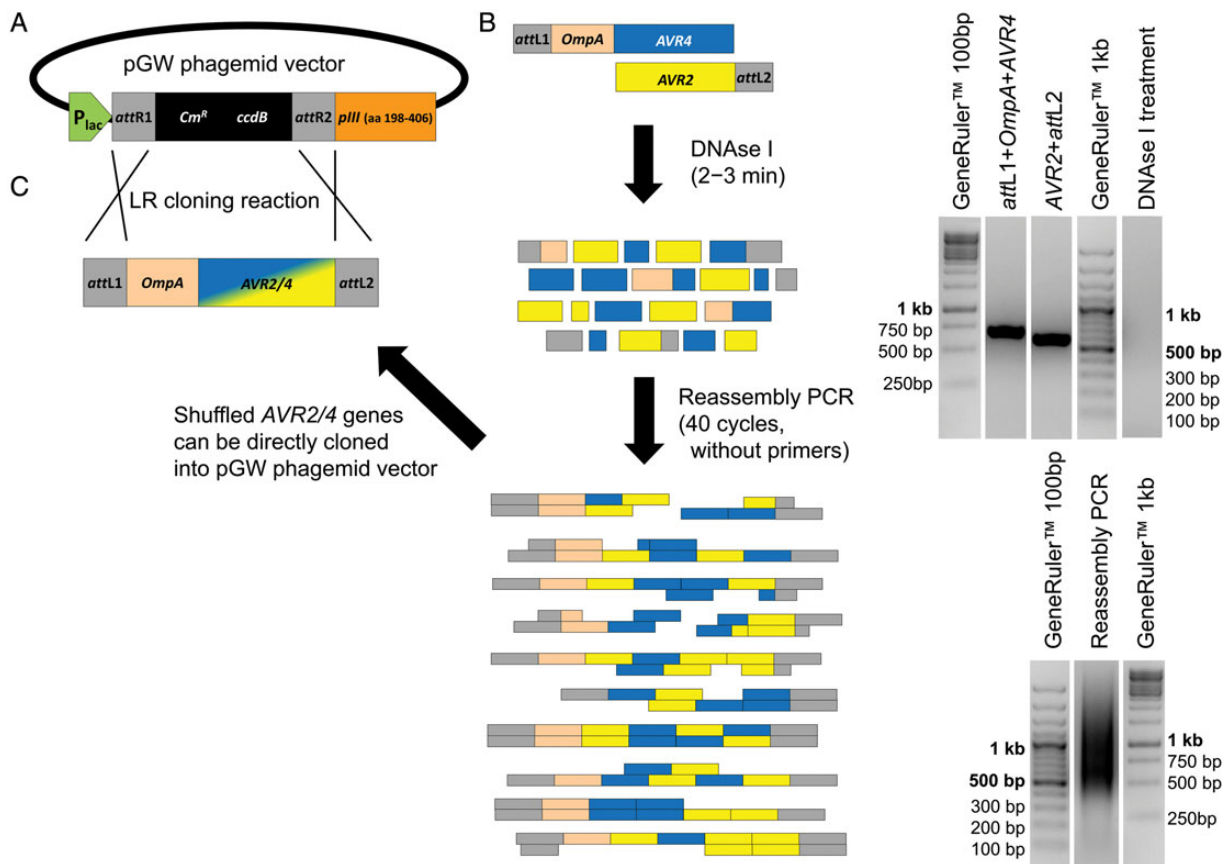


Fig. 1. Overview of the model library construction by DNA family shuffling and Gateway cloning. (A) Constructed pGW phagemid vector for the 3 + 3 phage display containing the *attR*-flanked Gateway cassette with *ccdB* suicide and chloramphenicol resistance (*Cm^R*) genes. (B) The flowchart shows the strategy described in this study to construct a DNA-shuffled library using an improved protocol. Asymmetric templates were used as parental genes to create fragments by DNase I digest. Aliquots of the templates prior and after DNase I digestion, and from the reassembly reaction were analyzed by AGE. After DNase I digestion, the fragments were not clearly visible on the gel because of the heterogeneous size and relatively low concentration of these products. However, the final reassembly resulted in a large amount of DNA fragments of different sizes observed as a smear on the gel. Importantly, neither an additional amplification step of the reassembled genes, nor the purification of the reassembly reaction was needed prior to the LR cloning step (C). Instead, reassembly reaction was directly used in the LR cloning reaction to construct our model library of chimeric *AVR2/AVR4* genes.

Table I. Quality and characteristics of the DNA-shuffled model libraries based on sequence analysis after transformation tests into *E. coli* DH5 α and XL1-Blue cells

		Transformation efficiency ^a cfu/ μ g	Analyzed clones	With shuffled insert ^b	Intact CDS ^c	Unique clones ^d	Clones with point mutations ^e
ElectroMAX TM DH5 α -E TM Competent cells (Invitrogen)	Nonamplified	9.0×10^6	43	43 100.0%	36 83.7%	33 91.7%	3 8.3%
ElectroMAX TM DH5 α -E TM Competent cells (Invitrogen)	Amplified ^f	1.5×10^8	44	44 100.0%	38 86.4%	38 100%	18 47.4%
ElectroMAX TM DH5 α -E TM Competent cells (Invitrogen)	<i>attL</i> -flanked <i>AVR4</i>	1.8×10^8	2	– –	2 100.0%	– –	– –
XL1-Blue Electro-Comp Cells (Stratagene)	Nonamplified	5.2×10^6	45	45 100.0%	37 84.4%	33 89.2%	4 10.8%
XL1-Blue Electro-Comp Cells (Stratagene)	Amplified ^f	8.1×10^7	46	46 100.0%	38 82.6%	38 100%	22 57.9%
XL1-Blue Electro-Comp Cells (Stratagene)	<i>attL</i> -flanked <i>AVR4</i>	7.3×10^7	2	– –	2 100.0%	– –	– –

^aCalculated by using the total amount of DNA (insert + vector) in the electroporation reaction.

^bInsert is constructed of *OmpA* signal peptide preceding shuffled *AVR2/AVR4* gene. No parental DNA sequences were detected. However, due to high similarity of the parental sequences, in some cases DNA shuffling did not result in changes on protein level.

^cIntact coding sequence (CDS) of shuffled *AVR2/AVR4* including *OmpA* signal peptide and in frame with *pIII*.

^dNumber of sequences that was not identical to any other of the analyzed sequences and had intact CDS.

^eSequences that contain mutations that are not found in the parental genes.

^fAmplified with non-proofreading MyTaqTM DNA polymerase (Bioline). This results in a number of point mutations and thus also increases the number of unique clones.

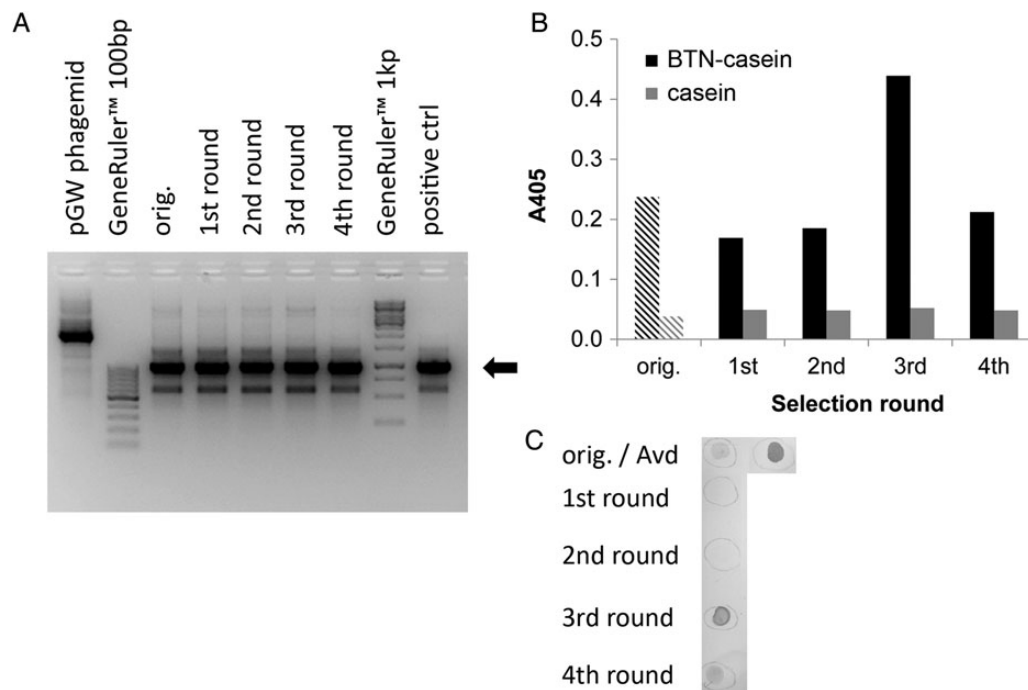


Fig. 2. Analyses to verify the biopanning process. (A) The integrity of the pGW phagemid vector, containing the desired-sized insert, was followed throughout the biopanning process by isolating the plasmids from the native phage display library (orig.) and from the output phages after each panning round. The insert was subsequently amplified by PCR. A sequence-verified plasmid containing chimeric *AVR2/AVR4* gene in pGW phagemid and pGW phagemid with Gateway cassette were used as positive and negative controls, respectively. (B) To verify that the phages, containing the pGW phagemid library, displayed active biotin-binding chimeric *AVR2/AVR4* proteins on their surface, the pool of phages prior to the panning (orig.) and obtained after each panning round were analyzed via microplate assays using biotin-conjugated casein (BTN-casein). As a negative control, the phages were also incubated on casein-coated microplate wells. The native phage display library (orig.) is shown with dashed bars, because the phages were PEG-precipitated in contrast to the pool of phages obtained after each panning round, which were taken directly from the phage culture supernatants. (C) A dot blot analysis was performed to verify the results gained from the microplate assay. A sample of the phage culture supernatants or PEG-precipitated phages of the native library (orig.) was detected using an alkaline phosphatase-conjugated biotin. As a positive control, recombinant avidin produced in *E. coli* (5 µg) was used.

Proteins expressed in 72 individual clones were analyzed after the third (44 clones) and fourth (28 clones) panning rounds by sequence analysis and avidin–biotin displacement (ABD) assay. ABD assay estimates the rate of dissociation of the protein–biotin complex and measures the concentration-independent tightness of the ligand binding. Overall, the phage display selection resulted in an enrichment of biotin-binding chimeric AVR mutants expressing more efficiently than parental AVRs (Fig. 3). Because expression of biotin-binding proteins is potentially harmful for the bacteria, we propose that the developed methodology has potential for other challenging proteins.

Sequencing revealed six enriched sequences (identical sequences found in multiple analyzed clones) and 12 unique sequences (Fig. 3A). Clones with identical sequences were grouped into Groups A–F. The groups with the highest enrichment, Groups A–C, showed positive signals in the ABD assay (Fig. 3C) and expression levels of Groups B and C were increased over the parental proteins (Fig. 3B, Supplementary Fig. S7). The low signal observed for AVR4 in Fig. 3C is due to low expression levels of the AVR4–phagemid construct. The expected reduction in the amount of unique clones after panning was observed from 74.4% in the native library to 18.9%.

Materials

Construction of the pGW phagemid vector

The Fab phagemid vector (pBluescript SK+ derived phagemid; Research Center of Finland, Biotechnology, Espoo, Finland) (Supplementary Fig.

S2A) was used to construct the Gateway compatible pGW phagemid vector. The signal peptide (pelB) and the Fab-encoding region were removed by digestion with the *Eco*RI and *Not*I restriction site enzymes (Thermo Fisher Scientific, Inc., Waltham, MA, USA), followed by shrimp alkaline phosphatase treatment (Thermo Fisher Scientific, Inc.).

The Gateway cassette was PCR amplified from a pTriEx-1.1 Gateway destination vector (provided by Professor Kari Airenne, University of Eastern Finland, Kuopio, Finland), and is described in detail in Supplementary Data. The resulting 1700-bp DNA fragment was digested with *Eco*RI and *Not*I (Thermo Fisher Scientific, Inc.). Because the *ccdB* is a cytotoxic protein lethal to most *E. coli* strains, the constructed vector was propagated in *E. coli* DB3.1 (Invitrogen™, Thermo Fisher Scientific, Inc.). The authenticity of the constructed pGW phagemid was confirmed through restriction analysis (not shown) and DNA sequencing.

Construction of the chimeric model library

Parental genes, *AVR2* (Hytönen *et al.*, 2005) and *AVR4* (Hytönen *et al.*, 2004b), were amplified from pBVboostFG-AVR2 and pGEM®-T Easy-AVR4 vectors respectively, as schematically shown in Supplementary Fig. S3. These DNA fragments, flanked asymmetrically with Gateway *attL* elements were used as templates in the following DNA shuffling experiments. All the PCR (100 µl) contained 5 ng of template plasmid, 0.2 mM dNTP mix (Fermentas, Burlington, Canada), 10 µM 5' and 3' primers, and 2 U Phusion™ High-Fidelity DNA polymerase (Thermo Fisher Scientific, Inc.) in 1X Phusion GC buffer in the presence of 5%

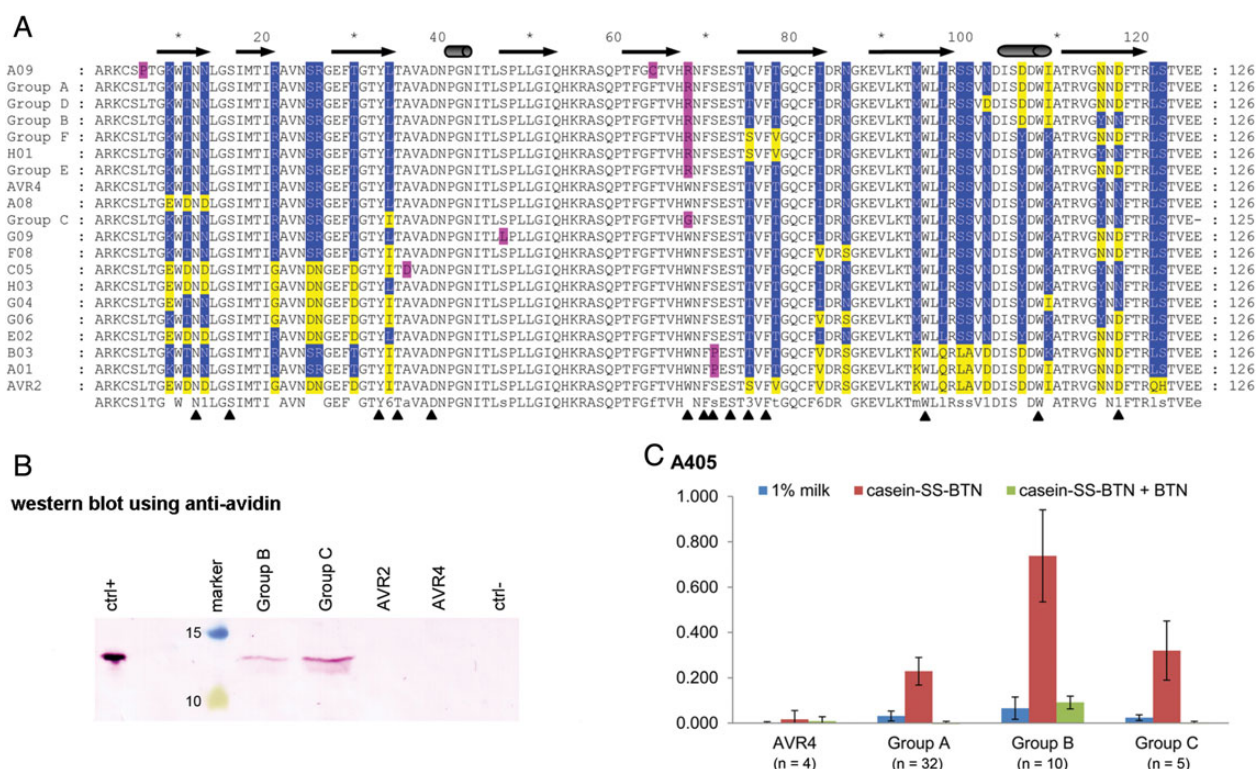


Fig. 3. Enrichment of the biotin-binding clones from the model library. (A) Multiple sequence alignment of the enriched clones after four panning rounds. The secondary structure elements, shown on top of the alignment, correspond to the AVR2 [PDB:1WBI] (Hytönen *et al.*, 2005) and AVR4 [PDB: 1Y52] (Eisenberg-Domovich *et al.*, 2005) structures. The residues originating from AVR2 and AVR4 are shown in yellow and blue, respectively. The point mutations resulting in amino acid changes are shown in purple. The triangles at the bottom of the alignment highlight the residues involved in biotin binding. (B) Expression levels of selected clones in comparison with the parental proteins. The proteins were detected using a polyclonal antibody against chicken avidin (Kulomaa *et al.*, 1978). ctrl+: positive control, bacterial expressed chicken avidin; ctrl-: negative control, BL21-AI cells without pGW phagemid; marker, molecular weight standard. (C) Results from the ABD assay of the proteins of groups A–C. The amount of analyzed clones for each group is given.

dimethyl sulfoxide. The primers used in the PCR are listed in Supplementary Table SIII. Correct-sized PCR fragments were gel purified using an illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, Piscataway, NJ, USA).

DNA family shuffling was performed essentially as previously described (Stemmer, 1994a; Zhao and Arnold, 1997; Niederhauser *et al.*, 2012) and is described in detail in Supplementary Data. The reassembled genes were used as entry clones in the LR cloning reactions, each containing 150 ng of destination vector (pGW phagemid) and 400–500 ng of reassembly reaction product. Table I LR cloning reactions contained 640 ng of reassembly reaction product, or 150 ng of amplified reassembly reaction or *attL*-flanked AVR4 as entry clones. Gateway® LR clonase™ II (Invitrogen™, Thermo Fisher Scientific, Inc.) enzyme mix (2 µl) was added and the reactions were incubated overnight at RT. The reactions were terminated after addition of 1 µl of Proteinase K at 37°C for 10 min. Transformation of the LR cloning reaction into *E. coli* cells and preparation of the DNA library are described in detail in Supplementary Data. Masterplates from the transformations were prepared (see Supplementary Data) and isolated plasmids from the masterplates were sequenced (see Supplementary Data).

Selection and screening of model library

Production of phages was done essentially as in Niederhauser *et al.* (2012) and is described in detail in Supplementary Data. Four rounds of biopanning were performed using the automated magnetic bead platform, Precipitator™ (Abnova, Taipei City, Taiwan) and the parameters are listed in Supplementary Table SIV (for the detailed

biopanning protocol, see Supplementary Data). After every round of biopanning, phages were amplified on microplates essentially as previously described (Turunen *et al.*, 2009) and is described in detail in Supplementary Data.

Stability of the DNA library in pGW phagemid. After each round of amplification from the overnight *E. coli* cultures, the plasmid DNA was isolated and the inserts were PCR amplified using primers 1867 and 886 (Supplementary Table SIII).

Microplate assay. The enrichment of phages expressing functional biotin-binding proteins on their surface was analyzed through phage ELISA, as previously described (Turunen *et al.*, 2009).

Dot blot analysis. The phages were analyzed after each panning round through dot blotting using biotinylated alkaline phosphatase (1 : 5000) as a probe (Vector Laboratories, Burlingame, CA, USA). Five microliter samples of the phages were applied on nitrocellulose and BCIP/NBT was used for detection.

ABD assay

Estimation of the biotin-binding properties of chimeric proteins was done using a microplate-based avidin–biotin displacement assay essentially as in Taskinen *et al.* (2014). The assay is described in detail in Supplementary Data.

Protein expression tests

One clone from sequence Groups B and C was tested for protein production in *E. coli* BL21-AI. Proteins were produced and captured with

D-biotin resin essentially as in Hytönen et al., (2004a) and the procedure is described in detail in Supplementary Data. The captured proteins were analyzed on Western blot.

Conclusion

We have demonstrated that Gateway cloning can be efficiently combined with DNA shuffling in the preparation of DNA mutant libraries. In contrast to a previously described method employing the Gateway cloning system with DNA shuffling (van den Berg et al., 2006), the presented approach is more efficient and requires fewer steps. Gruet et al. (2012) have shown that the (epPCR) library construction can be simplified by eliminating the BP (integrative reaction ($attB \times attP \rightarrow attL + attR$)) step of the Gateway cloning technology, and suggested that their method could possibly be applied to the creation of DNA shuffling libraries as well, but their method does not simplify the DNA shuffling process itself. The model DNA libraries prepared here were of high quality, and we did not detect any significant number of clones with errors in the sequence recombination. This expands the use of Gateway technology for the research utilizing directed evolution methods. We also demonstrated that this platform is suitable in combination with phage display selection, which resulted in functional chimeric proteins, thus validating the proper design of these novel DNA tools. We believe that these findings make DNA shuffling a more attractive and powerful method for directed evolution studies.

Supplementary data

Supplementary data are available at PEDS online.

Acknowledgements

The authors acknowledge Ulla Kiiskinen, Latifeh Azizi, Niklas Kähkönen, Laura Kananen and Meri Uusi-Mäkelä for their excellent technical support. The authors thank Professor Kari Airene for his continuous support. The authors also thank Chloe Thomson and Conor Kelly for checking the language of the manuscript. The authors acknowledge the infrastructure support from Biocenter Finland.

Conflict of interest: none declared.

Funding

This work was supported by grants from the Academy of Finland [project numbers 136288 to V.P.H., 140978 to V.P.H., 121236 to M.S.K., 261285 to M.S.K.]; MNT-ERA.net/Tekes—the Finnish Funding Agency for Technology and

Innovation [40407/09 to M.S.K.]; Pirkanmaa Hospital District (to V.P.H. and to M.S.K.); Sigrid Jusélius foundation (to V.P.H.); National Doctoral Programme in Informational and Structural Biology (to S.I.L. and B.N.); and Tampere Graduate Program in Biomedicine and Biotechnology (to S.K. and T.R.).

References

- Brasch, M.A., Hartley, J.L. and Vidal, M. (2004) *Genome Res.*, **14**, 2001–2009.
- Cramer, A., Raillard, S.A., Bermudez, E. and Stemmer, W.P. (1998) *Nature*, **391**, 288–291.
- Eisenberg-Domovich, Y., Hytönen, V.P., Wilchek, M., Bayer, E.A., Kulomaa, M. S. and Livnah, O. (2005) *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 528–538.
- Festa, F., Steel, J., Bian, X. and Labaer, J. (2013) *Proteomics*, **13**, 1381–1399.
- Gao, H., Pattison, D., Yan, T., et al. (2008) *PLoS One* **3**, e2983.
- Gruet, A., Longhi, S. and Bignon, C. (2012) *Microb. Cell Fact.*, **11**, 14–2859–11–14.
- Hartley, J.L., Temple, G.F. and Brasch, M.A. (2000) *Genome Res.*, **10**, 1788–1795.
- Hedde, C. and Mazaleyrat, S.L. (2007) *Protein Eng. Des. Sel.*, **20**, 327–337.
- Hytönen, V.P., Laitinen, O.H., Airene, T.T., et al. (2004a) *Biochem. J.*, **384**, 385–390.
- Hytönen, V.P., Nyholm, T.K., Pentikäinen, O.T., et al. (2004b) *J. Biol. Chem.*, **279**, 9337–9343.
- Hytönen, V.P., Määttä, J.A., Kidron, H., et al. (2005) *BMC Biotechnol.*, **5**, 28.
- Ikeuchi, A., Kawarasaki, Y., Shinbata, T. and Yamane, T. (2003) *Biotechnol. Prog.*, **19**, 1460–1467.
- Katzen, F. (2007) *Expert Opin. Drug Discov.*, **2**, 571–589.
- Kulomaa, M.S., Elo, H.A. and Tuohimaa, P.J. (1978) *Biochem. J.*, **175**, 685–690.
- Laitinen, O.H., Hytönen, V.P., Ahlroth, M.K., et al. (2002) *Biochem. J.*, **363**, 609–617.
- Lutz, S. and Patrick, W.M. (2004) *Curr. Opin. Biotechnol.*, **15**, 291–297.
- Marsischky, G. and LaBaer, J. (2004) *Genome Res.*, **14**, 2020–2028.
- Niederhauser, B., Siivonen, J., Määttä, J.A., Jänis, J., Kulomaa, M.S. and Hytönen, V.P. (2012) *J. Biotechnol.*, **157**, 38–49.
- Stemmer, W.P. (1994a) *Proc. Natl Acad. Sci., U.S.A.* **91**, 10747–10751.
- Stemmer, W.P. (1994b) *Nature*, **370**, 389–391.
- Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium, Graslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., et al. (2008) *Nat. Methods*, **5**, 135–146.
- Taskinen, B., Airene, T.T., Jänis, J., Rahikainen, R., Johnson, M.S., Kulomaa, M. S. and Hytönen, V.P. (2014) *PLoS One*, **9**, e92058.
- Turunen, L., Takkinen, K., Söderlund, H. and Pulli, T. (2009) *J. Biomol. Screen.*, **14**, 282–293.
- van den Berg, S., Lofdahl, P.A., Hard, T. and Berglund, H. (2006) *J. Biotechnol.*, **121**, 291–298.
- Wang, T.W., Zhu, H., Ma, X.Y., Zhang, T., Ma, Y.S. and Wei, D.Z. (2006) *Mol. Biotechnol.*, **34**, 55–68.
- Zhao, H. and Arnold, F.H. (1997) *Nucleic Acids Res.*, **25**, 1307–1308.

Artificial Avidin-Based Receptors for a Panel of Small Molecules

Soili I. Lehtonen,[†] Antti Tullila,^{‡,⊥} Nitin Agrawal,^{§,⊥} Sampo Kukkurainen,^{†,||} Niklas Kähkönen,[†] Masi Koskinen,[†] Tarja K. Nevanen,[‡] Mark S. Johnson,[§] Tomi T. Airenne,[§] Markku S. Kulomaa,[†] Tiina A. Riihimäki,[†] and Vesa P. Hytönen^{*,†,||}

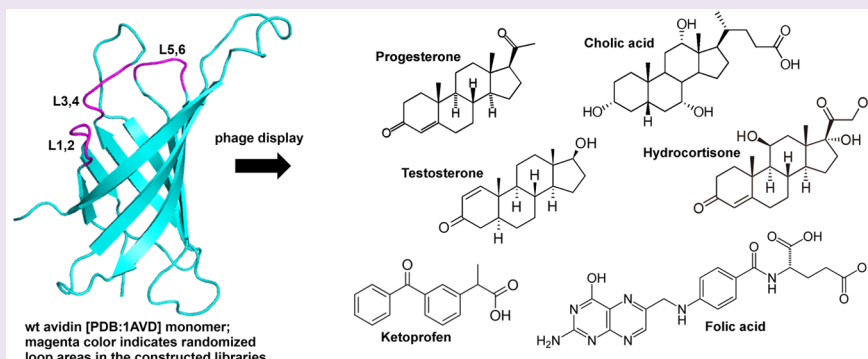
[†]BioMediTech, University of Tampere, Biokatu 6, FI-33014 Tampere, Finland

[‡]VTI Technical Research Centre of Finland, Tietotie 2, FI-02044 Espoo, Finland

[§]Structural Bioinformatics Laboratory, Biochemistry, Faculty of Science and Engineering, Åbo Akademi University, Tykistökatu 6A, FI-20520 Turku, Finland

^{||}Fimlab Laboratories, Biokatu 4, FI-33520 Tampere, Finland

S Supporting Information



ABSTRACT: Proteins with high specificity, affinity, and stability are needed for biomolecular recognition in a plethora of applications. Antibodies are powerful affinity tools, but they may also suffer from limitations such as low stability and high production costs. Avidin and streptavidin provide a promising scaffold for protein engineering, and due to their ultratight binding to D-biotin they are widely used in various biotechnological and biomedical applications. In this study, we demonstrate that the avidin scaffold is suitable for use as a novel receptor for several biologically active small molecules: Artificial, chicken avidin-based proteins, antindins, were generated using a directed evolution method for progesterone, hydrocortisone, testosterone, cholic acid, ketoprofen, and folic acid, all with micromolar to nanomolar affinity and significantly reduced biotin-binding affinity. We also describe the crystal structure of an antidin, sbAvid-2(I117Y), a steroid-binding avidin, which proves that the avidin scaffold can tolerate significant modifications without losing its characteristic tetrameric beta-barrel structure, helping us to further design avidin-based small molecule receptors.

Small molecule recognition has a central role in various technologies, such as diagnostics, separation methods, and drug development.¹ Protein-based applications for biomedical research and biotechnology have increased in number, use, and complexity, and various protein scaffolds have been engineered to generate improved or unique biomolecules.^{2,3} The use of antibodies and their engineered forms provide a fruitful resource for sensitive and highly specific protein tools for biomedical and biotechnological applications.^{4,5} Additionally, a variety of nonimmunoglobulin scaffolds (e.g., affibodies, Kunitz domains, anticalins), having designed characteristics, are showing promise in many applications and may even provide properties superior to antibodies.^{2,6,7} Small molecules are especially challenging for development of high-affinity receptors, and there are a limited number of suitable scaffolds available.⁸

Avidin is a frequently utilized protein tool in a plethora of life science applications, including imaging, purification, labeling, targeting, and detection.^{9,10} This chicken-derived tetrameric beta-barrel protein has many desirable properties: it is rather small in size (60 kDa), it can be efficiently produced in *E. coli*,¹¹ its structure is well-known (e.g., [PDB: 1VYO]),¹² and it is highly resistant to a wide range of denaturing conditions such as high temperatures, extreme pH values, and many organic solvents.^{13,14} Furthermore, its structure–function relationship has been widely explored using rational mutagenesis,¹⁵ and several avidins with tailored and novel functions exist, such as dual-chain avidin,^{16–19} single-chain avidin,²⁰ extremely thermo-

Received: June 9, 2015

Accepted: November 9, 2015

Published: November 9, 2015

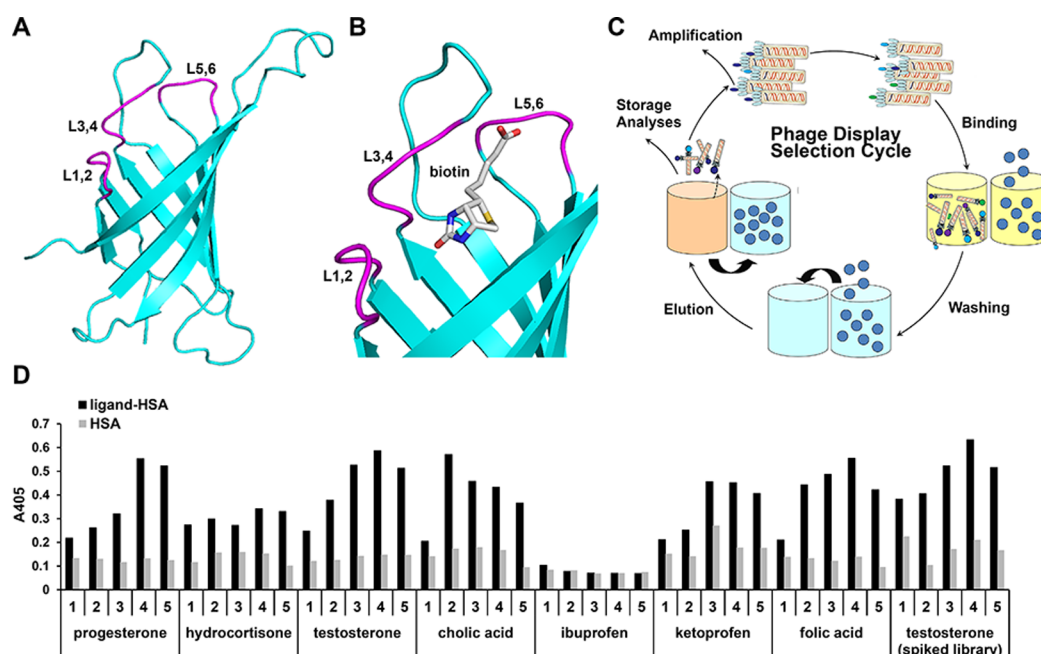


Figure 1. Library design and selection. (A) Side view of the wt avidin [PDB: 1AVD]³⁹ monomer. Carbon atoms are colored in cyan; magenta color indicates the randomized loop areas in the AvLib-1–3 libraries. (B) The binding pocket of wt avidin [PDB: 1AVD]³⁹ shown with biotin. Magenta color indicates the randomized loop areas in the AvLib libraries. (C) Schematic presentation of the semiautomatic phage display cycle performed using magnetic particle processor to capture the new recognition molecules. (D) The enrichment of the phages carrying the desired binding property was analyzed in a phage microplate assay after panning. In the graph, black bars represent the response from phages bound to the wells coated with human serum albumin (HSA) conjugated ligands, and gray bars represent the response from phages nonspecifically bound to HSA.

stable avidin,^{14,21} and steroid-binding avidin,²² thus expanding even further the usability of this scaffold.

The ligand-binding site of avidin has optimal features for small molecule recognition, and the femtomolar avidin–biotin interaction is actually one of the tightest noncovalent interactions known.²³ D-biotin (vitamin B₇) is a water-soluble molecule, and its binding to avidin is coordinated by several polar amino acid residues, which are hydrogen bonded to the oxygen, nitrogen, and sulfur atoms of biotin (e.g., [PDB: 2AVI]).²⁴ On the other hand, several aromatic and hydrophobic residues provide a complementary fit by surrounding the intervening aliphatic tail of the ligand. Besides biotin, avidin also has affinity toward some other small molecules that are chemically different from biotin: an azo dye, commonly known as HABA (2-(4'-hydroxybenzene)azobenzoic acid)—utilized in concentration measurements²⁵—and other azo molecules, including a phenyl derivative of HABA.¹² Therefore, the binding cavity of avidin offers potential as a scaffold to be engineered and mutated in order to recognize small molecules other than the currently known ligands.

One widely used strategy for modulation of the ligand-binding characteristics of a target scaffold is via loop randomization, because loop areas are often directly and critically involved in ligand binding and generally flexible in the absence of a ligand. For example, in addition to engineering the hypervariable loops in the complementary-determining regions (CDR) of antibodies,^{26,27} lipocalins have also been successfully modified via loop randomization resulting in anticalin scaffolds.^{28–30} Anticalins are high-affinity receptors for a variety of targets, mainly peptides and proteins,³⁰ but are also suitable for binding small molecules like digoxigenin.²⁸

Here, we describe the isolation and characterization of antidins, artificial receptors for small molecule recognition with the avidin scaffold obtained using a directed evolution

approach.³¹ Gene libraries (AvLib-1–3) were generated using targeted random mutagenesis and screened for the binding of a selected set of small molecules (progesterone, hydrocortisone, testosterone, cholic acid, ketoprofen, ibuprofen, and folic acid) using a semiautomated phage display procedure. The small molecule targets used here were chosen based on their diagnostic relevance. The selected set of ligands varies in both molecular size and chemical structure, thus enabling us to explore the potential of the generated protein libraries to accommodate a variety of ligands. The interactions between the produced antidin scaffolds and the small ligands were analyzed by using calorimetry, fluorometry, and microplate assays.

We demonstrate that the avidin scaffold is suitable to be used as a novel receptor for several biologically active small molecules: Antidins having micromolar to nanomolar affinities toward targets were captured with significantly reduced affinity toward biotin. Importantly, the high thermal stability and tetrameric structure of avidin were retained. Furthermore, we describe the structure of an antidin, sbAvid-2(I117Y), a steroid-binding avidin, which showed proper folding of the protein and helps us to better understand the structural determinants behind the increased affinity toward steroids generated using this phage display strategy.

RESULTS AND DISCUSSION

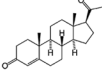
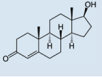
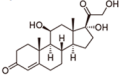
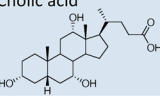
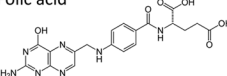
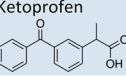
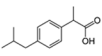
Library Design and Selection. In order to create novel small molecule recognition proteins based on the avidin scaffold, we utilized a loop randomization strategy (Figure 1A,B) for which the randomized amino acids were chosen based on the 3D structure of the avidin–biotin complex [PDB: 2AVI].²⁴ Two new libraries were created: (1) AvLib-1, where the loop between the $\beta 5$ and $\beta 6$ strands (L5,6-loop) of avidin was selected for partial manipulation (Table 1), and (2) AvLib-

Table 1. Characteristics of the AvLib libraries

library name	gene	randomized loops	randomized amino acids	theoretical size ($\times 10^6$)	library size ^a ($\times 10^6$)	degenerate codon
AvLib-1	wt avidin	L5,6	F72, S73, E74, S75	0.16	2.1	NNN
AvLib-2	wt avidin	L1,2 and L5,6	N12, D13, L14, G15, S16, F72, S73, E74, S75	38 443.00	6.7	NNY ^b
AvLib-3 ^c	sbAvd-1 ^d	L3,4	T35, A36, V37, T38	0.05	1.7	NNY ^b

^aCalculated from transformation activity. ^bNNY codon was used to limit the sequence space and to eliminate the introduction of stop codons. The degenerate codon excludes also the codons for five amino acids (Trp, Gln, Glu, Met, and Lys). ^cThis library has been used in a previous study (sbAvd-1 L3,4 library).²² ^dDescribed in ref 22.

Table 2. Structures of Selection Targets and Enriched Sequences from the Selections

Ligand	Code name	N12	D13	L14	G15	S16	T35	A36	V37	T38	F72	S73	E74	S75	Frequency (%)
 Progesterone	sbAvd-5		R	M	N	H	V	P	H	P					23
	sbAvd-1		R	M	N	H									14
	sbAvd-6		R	M	N	H	V	S	S	N					12
	sbAvd-2		R	M	N	H	A	T	V	N					11
	sbAvd-3		R	M	N	H	P	A	D	P					4
	sbAvd-4		R	M	N	H	P	Y	L	S					2
 Testosterone	sbAvd-3		R	M	N	H	P	A	D	P					40
	sbAvd-4		R	M	N	H	P	Y	L	S					15
	sbAvd-2		R	M	N	H	A	T	V	N					11
	sbAvd-5		R	M	N	H	V	P	H	P					8
 Hydrocortisone	hbAvd-1		H	A	G	D					P	Y	S	I	33
	hbAvd-2	D	S	N	S	A					L	A	N	L	13
	hbAvd-3	D	I	D	G	N					N	L	T	N	10
 Cholic acid	cabAvd-1	D	N	N	S	D					D	S	L	L	48
	cabAvd-2	D	S	S	S	N					A	T	P	I	16
	cabAvd-3	D	P	G	H	G					S	A	P	L	9
 Folic acid	fabAvd-1		N	T	D	T					S	P	L	S	97
	fabAvd-2		C	S	D	T						P	I	S	3
 Ketoprofen	sbAvd-5		R	M	N	H	V	P	H	P					86
	sbAvd-6		R	M	N	H	V	S	S	N					8
 Ibuprofen	ibAvd-1		L	R	T	D					D	Y	G	V	75
	ibAvd-2										L	P	P	V	25

2, where the L1,2-loop together with the L5,6-loop were partially randomized (Table 1). For the selection process, a previously constructed avidin DNA library for steroids,²² herein referred to as AvLib-3, was pooled together with AvLib-1 and AvLib-2 before biopanning. The AvLib-3 library is based on the sbAvd-1²² sequence with the L3,4-loop partially randomized (Table 1). In the case of AvLib-1 and AvLib-3, the theoretical sequence space of the created libraries was fully covered as judged by calculating the library size from the unique transformants, whereas in the case of the AvLib-2 library, only a fraction (0.02%) of the enormous sequence space ($15^9 = 3.8 \times 10^{10}$ possible unique clones) was covered (Table 1). The library design, however, was not optimal due to the degenerate codons used: NNN used in AvLib-1 increases the screening effort as it enables also stop codons, whereas NNY used in other libraries eliminates all stop codons but also excludes the codons for five amino acids (Trp, Gln, Glu, Met, and Lys). The avidin libraries were based on the monovalent display mode (3 + 3) and expressed as a sole fusion protein with the C-terminal portion of pIII as described.²²

Five rounds of selection were performed for pooled phage libraries using a magnetic particle processor (Figure 1C) combined with phage amplification on microplates without the

conventional PEG precipitation.³² This method enables performing multiple selections simultaneously toward several different target molecules. The magnetic beads were coated with alkaline phosphatase (AP) conjugated target ligands: progesterone 3-(O-carboxymethyl)oxime, hydrocortisone 3-(O-carboxymethyl)oxime, testosterone 3-(O-carboxymethyl)oxime, cholic acid, S-(+)-ibuprofen, S-(+)-ketoprofen, or folic acid. The aim was to explore the variation potential of the generated libraries, and the target molecules included different steroids as the previous work showed that steroid-binding avidins can be captured from the AvLib-3 library by using testosterone in a simpler panning method.²² The enrichment of specific phages was screened by M13 phage microplate analysis. As shown in Figure 1D, the selection was clearly efficient, and in most cases the responses from the phage screening increased until the fourth panning round. In the case of selection against ibuprofen, no clear enrichment was detected, indicating that the selection procedure used was not successful for this ligand. In the future, other selection strategies could be investigated for this ligand.³³ Up to 96 clones from the fifth panning round were DNA sequenced, and in all cases enriched sequences were obtained (Table 2): The progesterone, testosterone, and ketoprofen binders mostly had alterations only in the L3,4-

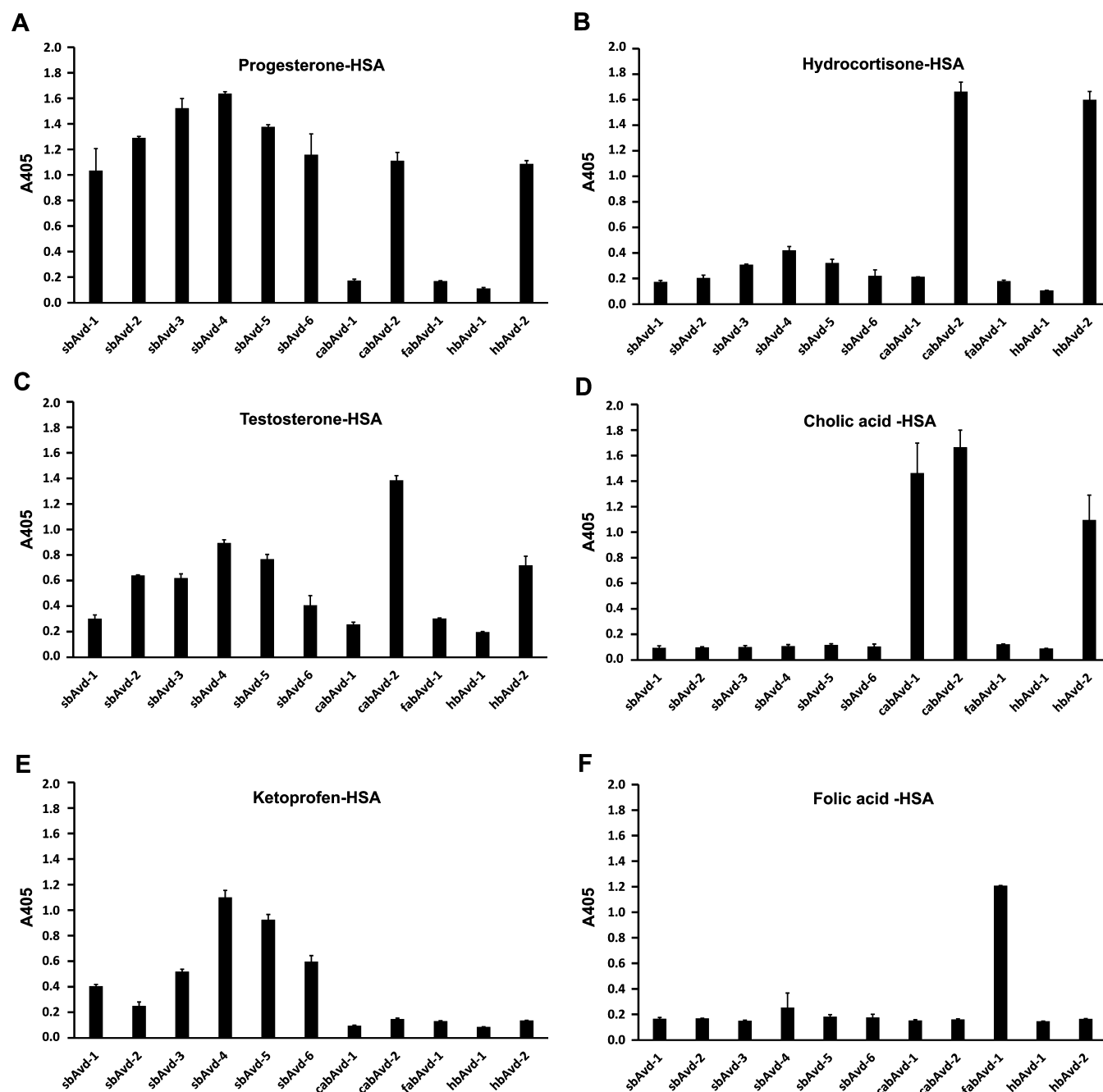


Figure 2. Microplate assay to determine the ligand-binding characteristics of the antidins with human serum albumin (HSA)-conjugated ligands. (A) Progesterone-HSA, (B) hydrocortisone-HSA, (C) testosterone-HSA, (D) cholic acid-HSA, (E) ketoprofen-HSA, and (F) folic acid-HSA. In the graph, bars represent the mean value of two parallel measurements. Error bars indicate the standard deviation. Six sbAvids showed clear affinity toward steroid coated surfaces (A, C). Besides steroid binding, sbAvid-4, sbAvid-5, and sbAvid-6 also showed affinity toward ketoprofen (E). Two mutants, hbAvid-2 and cabAvid-2, showed affinity toward multiple ligands (A, B, C, D). Based on the microplate assay, the most specific interactions among the antidins were detected with cabAvid-1 and fabAvid-1 on cholic acid (D) and folic acid (F) surfaces, respectively.

loop of sbAvid-1, whereas forms with mutations in the L1,2 and L5,6 loops of wt avidin were enriched in the selections with other ligands (Supporting Information).

Antidins Show Micromolar to Nanomolar Affinity toward Target Small Molecules. The panel of enriched avidin mutants were produced in *E. coli* and characterized. The proteins were purified by immobilized nickel affinity chromatography, which typically yielded 2–10 mg L⁻¹ of protein with >95% purity after a single purification step without further optimization (Supporting Information Figure 1). The

oligomeric state of the produced mutants was defined by SEC-SLS analysis (Supporting Information Table 1), which indicated a tetrameric quaternary structure as is seen for wt avidin. The ligand-binding properties of the modified avidin forms, antidins, and wt avidin were analyzed using a microplate assay monitoring the binding to the studied small molecule ligands (Figure 2), by determining the effect of the ligands on the thermal stability of the antidins (Table 3, Supporting Information Figure 2) and by determining the changes in

Table 3. Determined T_m and K_d Values of Antidins

protein	ligand	DSC T_m (°C)	SD ^a	ΔT_m (°C)	fluorometry K_d (nM) ^b	SD
sbAvd-1		80.6 ^c	- ^d			
	biotin	83.2 ^c	-	2.6	n/a	
	progesterone	90.7	-	10.1	174	3.0
	testosterone	83.1	-	2.5	6299	-
sbAvd-2		81.2	1.1			
	biotin	81.6	0.7	0.4	6649	504.2
	progesterone	94.3	0.5	13.2	111	35.1
	testosterone	86.7	1.0	5.6	2986	1437.8
sbAvd-2 (I117Y)		96.6	0.1			
	biotin	96.9	0.1	0.3	7084	470.2
	progesterone	100.1	1.3	3.5	180	-
	testosterone	97.7	0.9	1.1	1719	-
sbAvd-2 (N118D)		65.3 ^e	-			
	biotin	65.9 ^e	-	0.6	n/a	
	progesterone	70.7 ^e	-	5.4	913	-
	testosterone	69.1 ^e	-	3.8	1642	-
sbAvd-3		80.6	1.2			
	biotin	80.8	0.8	0.2	8580	673.2
	progesterone	92.3	0.4	11.7	249	89.6
	testosterone	84.6	0.8	4.0	1841	-
sbAvd-4		79.6	0.8			
	biotin	79.9	0.6	0.3	4731	-
	progesterone	91.5	1.5	11.9	135	-
	testosterone	85.1	-	5.5	2172	-
sbAvd-5		83.3	1.4			
	biotin	83.0	1.1	-0.3	7312	-
	progesterone	96.0	0.8	12.7	108	-
	testosterone	88.8	-	5.5	1630	-
	ketoprofen	84.5	-	1.2	2647	-
sbAvd-6		81.8	-			
	biotin	81.6	1.7	-0.2	3303	-
	progesterone	93.9	1.0	12.0	117	26.2
	testosterone	86.9	-	5.0	1469	-
	ketoprofen	83.3	-	1.5	3058	-
		64.3	6.1			
hbAvd-1		61.7	3.5	-2.6	n/d ^e	
	progesterone	81.2	4.1	16.9	n/a ^f	
	hydrocortisone	70.4	5.7	6.2	1513	-
	testosterone	79.9	2.7	15.7	n/a	
	cholic acid	63.8	4.3	-0.5	n/a	
		63.8	2.5			
hbAvd-2		67.5	1.4	3.7	2401	-
	progesterone	83.6	-	19.8	252	30.3
	hydrocortisone	69.4	3.8	5.7	2784	-
	testosterone	77.5	-	13.7	750	-
	cholic acid	70.5	2.2	6.8	n/a	
		57.1	0.0			
cabAvd-1		57.3	0.2	0.2	33761	-
	progesterone	71.2	0.3	14.2	1756	-
	hydrocortisone	57.7	0.1	0.7	n/a	
	testosterone	69.0	-	12.0	2411	-
	cholic acid	65.2	0.4	8.1	1700	163.3
		56.9	0.2			
cabAvd-2		58.9	0.5	2.0	4722	-
	progesterone	81.3	3.4	24.5	114	-
	hydrocortisone	64.9	-	8.0	4400	-
	testosterone	75.2	4.8	18.4	105	-
	cholic acid	79.9	0.9	23.0	64	-
		91.5	-			
cabAvd-2 (I117Y)		91.5	-	0.0	1450	9.2
	progesterone	97.0	0.2	5.5	104	-

Table 3. continued

protein	ligand	DSC T_m (°C)	SD ^a	ΔT_m (°C)	fluorometry K_d (nM) ^b	SD
fabAvd-1	testosterone	94.9	-	3.4	52	-
	cholic acid	92.8	-	1.3	163	-
		68.1	0.6			
	biotin	68.8	1.3	0.7	16199	461.0
	progesterone	74.1	1.3	6.0	n/a	
	hydrocortisone	69.1	1.3	1.0	n/d	
fabAvd-1 (I117Y)	testosterone	72.5	-	4.3	n/a	
	folic acid	71.8	0.6	3.7	188	-
		94.2 ^g	-			
	biotin	94.2 ^g	-	0.0	927	-
	folic acid	94.7 ^g	-	0.5	64	24.3
		80.8	0.1			
wtAvd	biotin	118.7	1.7	37.9	n/d	
	progesterone	83.0	1.6	2.2	3355	-
	hydrocortisone	82.1	0.6	1.3	n/a	
	testosterone	82.7	0.3	1.8	n/a	
	cholic acid	81.5	0.3	0.6	n/a	
	ketoprofen	82.3	0.1	1.5	n/a	-
	folic acid	81.5	0.2	0.6	n/a	

^aSD, Standard deviation. ^b K_d obtained using 100 nM protein concentration, and a fit for tight ligand binding, which accounts for the ligand depletion. In the case of biotin-binding measurements, up to 350 nM protein concentration was used to achieve measurable fluorescence quenching. ^cThe results from ref 22. ^dOnly one measurement ^en/d, not determined. ^fn/a, the data could not be fitted to model, or the total fluorescence quenching due to addition of the ligand was less than 2000 units. ^g T_m determined using 2-fold protein and ligand concentrations in the cell compared to other samples.

intrinsic fluorescence (Table 3, Supporting Information Figure 3).

sbAvids. Six antidin forms binding progesterone and testosterone (Table 2) were analyzed and named steroid-binding avidins (sbAvids). The scaffolds of sbAvid-1 and sbAvid-2 have been described previously,²² and the testosterone-binding characteristics of sbAvid-2 have been analyzed by molecular atomic force microscopy;³⁴ the current study provides more detailed information on the progesterone binding of these mutants. The affinities of the new sbAvid variants for progesterone and testosterone are within the same range with sbAvid-2. All sbAvids had a strong preference for binding progesterone over testosterone: Affinities (K_d) toward testosterone were in the micromolar range in all cases, whereas the affinities toward progesterone were 108–249 nM (Table 3 and Supporting Information Figure 3). Additionally, sbAvid-5 and sbAvid-6 were enriched during the ketoprofen selection (Table 2); the affinities toward ketoprofen were significantly weaker ($\sim 3 \mu\text{M}$, Table 3) as compared to steroid binding. The biotin-binding affinities of sbAvid-2–6 were drastically lower ($3\text{--}9 \mu\text{M}$, Table 3), as compared to wt avidin (K_d in the fM range³⁵), and in the microplate analysis the preincubation of sbAvids with biotin could not inhibit progesterone binding (Supporting Information Figure 4A). Importantly, the DSC measurements showed that the modifications did not have a destabilizing effect on the protein scaffolds, since all of the sbAvids had retained the high thermal stability of avidin with T_m values around 80 °C (Table 3, Supporting Information Figure 2). The ΔT_m values of the sbAvids in the presence of different ligands correlated well with their K_d values (Supporting Information).

hbAvids and cabAvids. Antidin forms enriched from the hydrocortisone (hbAvid-1 and hbAvid-2) and the cholic acid (cabAvid-1 and cabAvid-2) selections had significantly more destabilized structures than the other antidins studied here. The T_m values were around 64 °C for hbAvids and around 57 °C for

cabAvids (Table 3). However, the addition of ligands used in the selection of these forms significantly stabilized the proteins ($\Delta T_m = 5.7\text{--}23$ °C), especially in the case of cabAvid-2.

The microplate analysis (Figure 2) revealed that hbAvid-2 and cabAvid-2 have affinities toward progesterone ($K_d = 252$ nM and 114 nM, respectively), hydrocortisone (2.8 μM and 4.4 μM , respectively), testosterone (750 nM and 105 nM, respectively), and cholic acid (cabAvid-2, 64 nM) that were determined utilizing the intrinsic fluorescence (Table 3). Progesterone had a strong stabilizing effect on all of the studied hbAvid and cabAvid forms ($\Delta T_m = 14\text{--}25$ °C, Table 3). The biotin-binding affinities of these proteins were weak as compared to wt avidin (2.4–34 μM). The ligand-binding preference was clearly changed for cabAvid-2: cabAvid-2 had a biotin-binding affinity of 4.7 μM , whereas the affinity for cholic acid was 64 nM (Table 3).

fabAvid-1. An antidin captured from the selection against folic acid, fabAvid-1, showed a clear binding response to folic acid in the microplate analysis (Figure 2). However, when the binding competed with biotin, the response dropped, indicating some residual biotin-binding affinity (Supporting Information Figure 4B). When the ligand binding of fabAvid-1 was studied in more detail by fluorescence spectroscopy, a K_d of 188 nM toward folic acid (Table 3) was detected with significantly decreased biotin-binding affinity ($K_d \sim 16 \mu\text{M}$). However, the thermal stabilizing effect of folic acid (three times molar excess) was modest ($\Delta T_m = 3.7$ °C), which could be due to the relatively large size of folic acid ($M_{\text{folic-acid}} = 441.4 \text{ g mol}^{-1}$ vs $M_{\text{biotin}} = 244.3 \text{ g mol}^{-1}$). For this reason, it is unlikely that folic acid can completely fit into the binding site of fabAvid-1, and thereby enhance structural cooperativity, i.e. stronger inter-subunit association and increased structural order upon binding like in the case of binding of biotin to streptavidin and chicken avidin.^{13,24}

Control Measurements. The interaction of wt avidin with the small molecule ligands used in this study was analyzed, and

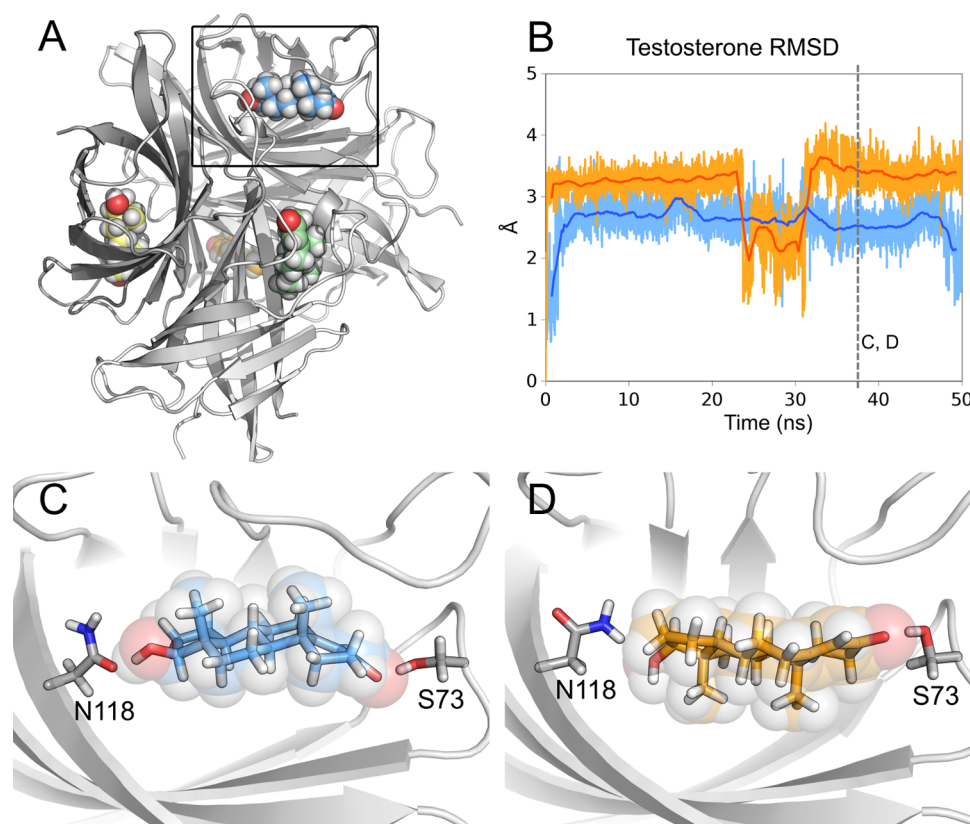


Figure 3. Molecular dynamics simulation of sbAvid-1 with bound testosterone molecules. (A) Testosterone binding to a tetrameric sbAvid-1 model in a 50 ns molecular dynamics simulation. (B) The distances of two testosterone molecules (blue, orange) in reference to their docked positions are measured using root-mean-square displacement (RMSD). (C, D) Snapshots showing the two testosterone molecules in the ligand-binding pockets at 37.5 ns as indicated in B.

in general, only a low-level response was seen, indicating that wt avidin had only negligible affinities toward the ligands (Supporting Information). By measuring the changes in intrinsic fluorescence, negligible changes (Supporting Information Figure 3A) were seen, except in the case of progesterone ($K_d \sim 3 \mu\text{M}$, Table 3). It is good to keep in mind the fact that avidins belong to the same structural superfamily as lipocalins,³⁶ which are known to transport small hydrophobic molecules such as steroids.³⁷

Antidins with Enhanced Thermal Stability. In order to further stabilize selected promising variants of antidins (sbAvid-2, cabAvid-2 and fabAvid-1) and to examine the effects of additional mutations on their ligand binding, the point mutation I117Y was introduced into the proteins. The I117Y mutation is known to significantly improve the thermal stability of avidin by strengthening the subunit 1,3-interface.²¹ The mutation was originally inspired by the equivalent position, residue Tyr115, in the avidin-related protein 4 (AVR4) structure [PDB: 1Y55];³⁸ Tyr115 has been shown to be important for the higher thermal stability of AVR4 in comparison to avidin. As expected, the I117Y modification increased the T_m of the apo-proteins by 15–35 °C (Table 3). Moreover, the yields from protein production increased 2- to 6-fold in 0.5 L cultures. The point mutation I117Y affected the ligand binding of these antidins: sbAvid-2(I117Y) showed some decrease in progesterone binding ($K_d = 180 \text{ nM}$) when compared to sbAvid-2 ($K_d = 111 \text{ nM}$), whereas the mutation had only a minor effect on the biotin- and testosterone-binding affinities (Table 3). As compared to cabAvid-2, antidin cabAvid-

2(I117Y) showed an increase in biotin binding (1.5 vs 4.7 μM), some increase in progesterone and testosterone binding (104 vs 114 nM and 52 vs 105 nM, respectively), but a modest decrease in cholic acid binding (163 vs 64 nM). The affinity of fabAvid-1(I117Y) toward folic acid increased (64 vs 188 nM), along with increased biotin-binding affinity (1 vs 16 μM), as compared to fabAvid-1 (Table 3).

MD Simulation of sbAvid-1 Shows the Importance of Asn118 for Steroid Binding. Two main binding modes were observed during a 50 ns molecular dynamics simulation of sbAvid-1 with four bound testosterone molecules (Figure 3): the testosterone methyl groups either pointed toward the $\beta 7$ strand (41% of the time) or toward the L3,4 loop (51% of the time). In both modes, the testosterone 17-OH group formed a hydrogen bond with the side chain of Asn118 and the 3-keto oxygen with the side chain of Ser73 (Figure 3). Since testosterone and progesterone differ only in regard to the functional group at C-17, the point mutation N118D was introduced into sbAvid-2 and assumed to affect progesterone binding. And, indeed, sbAvid-2(N118D) had decreased affinity for progesterone (913 vs 180 nM), while the affinity for testosterone slightly increased (1.6 vs 3.0 μM , Table 3). However, the N118D mutation destabilized sbAvid-2 (the thermal stability decreased from 81.2 to 65.3 °C, Table 3). These results indicate that Asn118 is an important residue for steroid binding. Moreover, since Ile117 is located next to Asn118, the importance of Asn118 for ligand binding may also explain why sbAvid-2(I117Y) had decreased affinity to progesterone.

The Crystal Structure of sbAvd-2(I117Y). To elucidate the molecular characteristics of sbAvids at the atomic level, crystallization trials of sbAvd-2(I117Y), and other sbAvids, were performed in the absence and presence of steroids; crystals only formed from sbAvd-2(I117Y). The X-ray structure of a ligand-free sbAvd-2(I117Y) was determined at 1.95 Å resolution (Table 4, Figure 4A). Two monomers were found per

Table 4. Structure Determination Statistics for sbAvd-2(I117Y)

data processing ^a	
space group	P2 ₁ 2 ₁ 2
unit cell:	
a, b, c (Å)	74.47, 79.80, 43.07
α, β, γ (deg)	90, 90, 90
wavelength (Å)	0.87260
beamline	ID23-2, ESRF
resolution (Å) ^b	25–1.95 (2.05–1.95)
observed reflections ^b	92876 (12995)
unique reflections ^b	19264 (2650)
I/σ ^b	20.32 (5.09)
R _{factor} (%) ^b	7.8 (52.4)
completeness ^b	99.5 (99.9)
refinement	
Matthews coefficient	2.02
R _{work} (%) ^c	18.10%
R _{free} (%) ^c	21.10%
monomers (asymmetric unit)	2
RMSD	
bond lengths (Å)	0.015
bond angles (deg)	1.62

^aThe numbers in parentheses refer to the highest resolution bin. ^bData from XDS.⁴⁶ ^cData from Refmac 5.⁴⁷

asymmetric unit in the sbAvd-2(I117Y) crystals, and the biological assembly was clearly tetrameric. The primary structure of sbAvd-2(I117Y) is over 90% identical in amino

acid sequence with chicken avidin [PDB: 1AVD],³⁹ and despite the mutations, the fold of sbAvd-2(I117Y) is highly similar to that of the chicken avidin β-barrel structure. The Cα atoms of sbAvd-2(I117Y) (chain A) superimpose with chicken Avid [PDB: 1VYO]¹² (chain A) with a root-mean-squared deviation (RMSD) of 0.2 Å.

The sequence and structural differences between wt avidin and sbAvd-2(I117Y) are located in the L1,2-loop (mutations D13R, L14M, G15N, S16H), in the L3,4-loop (mutations T35A, A36T, T38N), and at the subunit 1,3-interface of the protein, where the I117Y mutation is located (Figure 4). Residues 38–42 of the L3,4-loop were not visible in the X-ray structure probably due to high thermal motion of the loop in the absence of the ligand. Therefore, the effect of the mutations T35A, A36T, and T38N for ligand binding cannot directly be derived from the determined structure. Moreover, there were a few unidentified peaks in the electron density maps (Supporting Information), and two blobs of density within the ligand-binding pocket were not modeled.

In the chicken avidin–biotin complex structures [PDB: 2AVI²⁴ and PDB: 1AVD³⁹], the L3,4-loop adopts a “closed” conformation, with the loop acting as a lid that closes over the biotin-binding site by interacting with the ligand. In these structures, Thr35 is hydrogen bonded to one of the nitrogen atoms of the bicyclic ring system of biotin.^{24,39} This interaction is lost in sbAvd-2 due to the T35A mutation that, together with the S16H mutation, could explain the drop in the biotin-binding affinity of sbAvd-2. In the sbAvd-2(I117Y) crystal structure, the imidazole side chain of His16 occupies roughly the space equivalent to that where the ureido ring oxygen atom of biotin is located in the biotin–complex structure of chicken avidin [PDB: 2AVI, 1AVD]^{24,39} (Figure 4B,C); consequently, His16 cannot form a hydrogen bond to biotin similarly as the side-chain oxygen atom of the equivalent Ser16 does in the avidin–biotin complex structure. Therefore, the T35A and S16H mutations are probably the most important amino acid changes reshaping the ligand pocket of sbAvd-2(I117Y) and

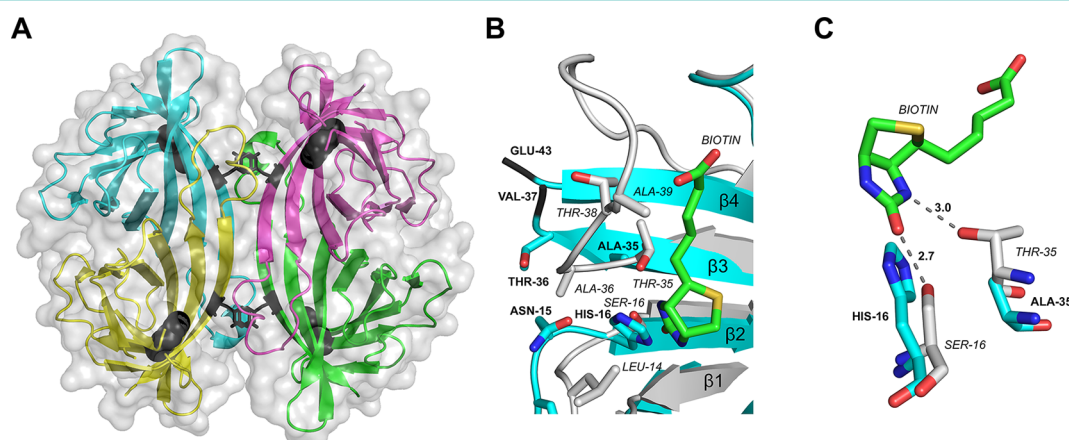


Figure 4. Crystal structure of sbAvd-2(I117Y). (A) The quaternary structure of sbAvd-2(I117Y) is shown with a transparent surface. Subunit I, cyan; subunit II, green; subunit III, magenta; and subunit IV, yellow. Tyr117 residues (black sticks) at the subunit interfaces and His16 residues (black spheres) at the bottom of the ligand-binding pockets are shown. (B) Comparison of selected L1,2-loop and L3,4-loop residues (chain A; stick models) of sbAvd-2(I117Y) (cyan carbon atoms; bold labels) and wt avidin [PDB: 1AVD]³⁹ (gray carbon atoms; labels in italics) are shown. Bound biotin of the wt avidin structure is shown as a stick model (green carbon atoms). β-sheets 1–4 are numbered. The Cα-atoms of Val37 and Glu43 of sbAvd-2(I117Y) are colored dark gray to indicate the missing region of the L3,4-loop. (C) Comparison of His16 and Ala35 of sbAvd-2(I117Y) and the equivalent residues Ser16 and Thr35 of wt avidin. The hydrogen bonds between biotin and Ser16 and Thr35 of wt avidin are indicated with dashed lines. Coloring of the residues is as in B.

altering the ligand-binding preferences of sbAvd-2(I117Y) as compared to wt chicken avidin.

As compared to sbAvd-1, the L3,4-loop modification of sbAvd-2 further decreases its biotin-binding affinity and improves the binding of progesterone, while it does not strongly affect testosterone binding (Table 3, Supporting Information Figure 4A). This may be explained by the T35A mutation: Ala35 in sbAvd-2 could favor interactions with the extra methyl group of progesterone that is not present in the otherwise highly similar testosterone. In this binding mode, the 20-keto group of progesterone (and the equivalent group of testosterone) could in turn be involved in interaction with one of the side chain nitrogen atoms of His16.

In AVR4, Tyr115 packs (π π stacking) against Tyr115 on the 1,3-interface and forms a hydrogen bond with the residues Lys92 and Asn117 on the 1,3-interface, as well as with one structural water molecule. In the sbAvd-2(I117Y) mutant, however, Tyr117 forms only two hydrogen bonds, one with Lys94 of the 1,3-interface and one with a structural water molecule.

CONCLUSIONS

Here, we have demonstrated that the avidin scaffold is suitable to be used as a novel receptor for several biologically active small molecules. The ligand-binding preferences of the wt avidin and antidins are very different: the biotin-binding affinity of wt avidin is the strongest known in the nature, whereas antidins have only a modest biotin-binding affinity. Furthermore, wt avidin has only negligible affinities toward the studied ligands, whereas antidins show micro- to nanomolar affinities toward them. The different ligand preferences between wt avidin and antidins may enable the use of antidins in applications where biotin-containing biological fluids restrict the use of wt avidin. Importantly, the high thermal stability of avidin, and its characteristic tetrameric beta-barrel structure, was retained in antidins despite the genetic modifications in the structure. Moreover, antidins were produced in *E. coli* with yields up to 10 mg L⁻¹ enabling inexpensive and scalable production. All in all, antidins could be used as an alternative to antibody scaffold-based methods and are promising candidates for the development of novel diagnostics and biomedical applications.

METHODS

Phage Display Libraries. Three avidin mutant libraries (AvLib1–3 libraries, see Table 1) with randomization targeted to loop regions of the β -barrel structure were prepared using a two-step PCR strategy and ligated into a phagemid vector (pBluescript SK+, VTT Technical Research Centre of Finland) as previously described.²² The resulting plasmids were electroporated into *E. coli* XL1-Blue (Stratagene; Supporting Information). The cultures were infected with VSC-M13 helper phages (Stratagene, 10¹¹ pfu mL⁻¹) essentially as in ref 40 and cultured overnight in the presence of wt avidin (0.1 mg mL⁻¹) at 28 °C on a shaker. The phages were collected by PEG precipitation,⁴⁰ and the phage titers were determined.⁴⁰ The AvLib libraries were pooled together (1:1:1) based on their titers. As a positive control, the AvLib-3 library²² was spiked with 100× excess of phages carrying a previously selected sbAvd-2 clone.²²

Semiautomated Biopanning Procedure. Magnetic beads (Dynabeads M-270 Epoxy, Invitrogen) were coated with alkaline phosphatase (AP)-conjugated ligands according to instructions from the manufacturer. The coating of the beads was then confirmed by Micro BCA Protein Assay Kit (Thermo Scientific) essentially as described³² (Supporting Information). Semiautomated biopanning on microplates³² was performed with a KingFisher Magnetic Particle

Processor (Thermo Fisher Scientific). A total of five rounds of biopanning were performed (Supporting Information Table 2) as described in detail in the Supporting Information. The effectiveness of the selection procedure was estimated from the output/input ratios of the phage titers (Supporting Information Table 3), and the enrichment of the specific phages was monitored using the M13 phage microplate assay³² with human serum albumin (HSA)-conjugated ligands. The conjugation of the ligands to AP and HSA is described in the Supporting Information. Finally, up to 96 output colonies were picked from each panning round for each ligand, and DNA sequencing was performed as described⁴¹ (Supporting Information).

Protein Production and Purification. The most enriched avidin mutants were subcloned with OmpA signal peptide¹¹ and C-terminal 6x His-tag into pET101/D expression vector (Invitrogen)²² and produced in *E. coli* strain BL21-AI (Invitrogen) essentially as previously described¹¹ (Supporting Information). The point mutations I117Y and N118D were introduced into sbAvd-2 by utilizing the QuikChange method (Invitrogen) (Supporting Information). For some of the protein variants, 4-L fed-batch cultures were also performed in a fermentor (Labfors Infors 3, Infors HT) as described previously.¹⁴

The proteins were purified using affinity chromatography on a Ni-NTA (QIAGEN) column²² or with an automated ÄKTA Purifier protein purification system (GE Healthcare Life Sciences) with HisTrap FF Crude columns (GE Healthcare; Supporting Information). The proteins were dialyzed into the measurement buffer (20 mM NaH₂PO₄/Na₂HPO₄, 1 M NaCl, 20 mM imidazole, pH 7.4).

Protein Microplate Assays. The ligand-binding specificity of selected avidin forms was determined with a microplate assay essentially as described.²² In the assays, well surfaces were coated with HSA-conjugated ligands.

Size Exclusion Chromatography with Static Light Scattering Analysis. The proteins were analyzed for molecular weight (static light scattering) and hydrodynamic size (dynamic light scattering) using a liquid chromatography instrument (CBM-20A, Shimadzu Corporation) equipped with an autosampler (SIL-20A), UV-vis (SPD-20A), and a fluorescence detector (RF-20AX) as well as Zetasizer μ V light scattering detector (Malvern Instruments Ltd.) essentially as described.⁴² Samples (~50 μ g in 10–100 μ L) were injected in a Superdex200 5/150GL column (GE healthcare) and equilibrated with the measurement buffer. A flow rate of 0.1 mL min⁻¹ at 20 °C was used.

Differential Scanning Calorimetry (DSC). The thermal stability of the studied proteins in the presence and absence of ligands was analyzed using an automated VP-Capillary DSC System (Microcal Inc.) essentially as described⁴² (Supporting Information). Protein samples in the measurement buffer were degassed prior to the measurement. The protein concentration in the cell was 7 μ M, and the ligand concentration was 21 μ M.

Fluorometric Assay. The binding affinity of the unconjugated small molecules to avidin mutants was determined utilizing the intrinsic fluorescence originating from the aromatic amino acid residues (mainly tryptophan and tyrosine) of avidin and the fluorescence quenching caused by ligand binding. In brief, 100 nM protein samples in 50 mM NaH₂PO₄/Na₂HPO₄ and 650 mM NaCl at pH 7 were excited at 280 nm, and emission was collected at 350 nm using QuantaMaster Spectrofluorometer (Photon Technology International, Inc.) with 2 nm slits. The assay was performed in a quartz cuvette with stirring at 25 °C. The ligand was added to the protein sample in small aliquots, and the fluorescence intensity was monitored after a short incubation. The dissociation constant (K_d) was determined from the resulting quenching curve using GraphPad Prism (GraphPad Software, Inc.). The data were fitted to a quadratic equation⁴³ for tight binding interactions (as described in detail in the Supporting Information), which takes ligand depletion and nonspecific binding into account.

MD Simulations. The antidin sbAvd-1 model (based on PDB: 1VYO¹²) was prepared, and testosterone was docked into the biotin

binding pocket using GOLD.⁴⁴ MD simulation of the complex was run in NAMD⁴⁵ as described in detail in the [Supporting Information](#).

X-ray Crystallography. Protein sbAvd-2(I117Y) (1.8 mg mL⁻¹; in the measurement buffer) was crystallized using the vapor diffusion method as described in the [Supporting Information](#). The initial X-ray diffraction properties were analyzed using a PX Scanner (Agilent Technologies). The data were collected (at the European Synchrotron Radiation Facility, beamline ID23-2, Grenoble, France) and processed, and the structure was determined as described in the [Supporting Information](#) (see Table 4).

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acschembio.5b00906](https://doi.org/10.1021/acschembio.5b00906).

Supplemental Figures 1–5, Tables 1–4, methods, results and discussion, and references (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +358 40 190 1517. Fax: +358 33 641 471. E-mail: vesa.hytonen@uta.fi.

Author Contributions

[†]Equal contribution

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Grant funding from the Academy of Finland (257814 (M.S.J.), 272283 (M.S.J.), 136288 (V.P.H.), 273192 (V.P.H.), 261285 (M.S.K.), and 272288 (M.S.K.)), VTT Technical Research Centre of Finland, Sigrid Juselius Foundation, Joe Pentti and Tor Borg Memorial Fund, and Åbo Akademi Center of Excellence in Cell Stress and Aging is gratefully acknowledged. We thank the Academy of Finland FIRI program (141398), bioinformatics infrastructure support (J.V. Lehtonen) from Biocenter Finland, and CSC IT Center for Science for laboratory and computational infrastructure support. We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities, and we would like to thank the local contacts for assistance in using beamline ID23-2. The Finnish National Protein Crystallography Consortium (FinnProCC) is acknowledged for data collection. We thank Biocenter Finland for support in protein production and characterization infrastructure. We also thank U. Kiiskinen, J. Määttä, S. Posch, L. Azizi, F. Kangasniemi, E. Ojala, J. M. Lehtonen, and H. Boer for technical support.

■ REFERENCES

- (1) Poma, A., Whitcombe, M., and Piletsky, S. (2013) Plastic Antibodies, in *Designing receptors for the next generation of biosensors* (Piletsky, S. A., and Whitcombe, M. J., Ed.), Vol. 12, pp 105–125, Springer, Heidelberg.
- (2) Banta, S., Dooley, K., and Shur, O. (2013) Replacing antibodies: engineering new binding proteins. *Annu. Rev. Biomed. Eng.* 15, 93–113.
- (3) Binz, H. K., Amstutz, P., and Pluckthun, A. (2005) Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.* 23, 1257–1268.
- (4) Adams, J. J., and Sidhu, S. S. (2014) Synthetic antibody technologies. *Curr. Opin. Struct. Biol.* 24, 1–9.
- (5) Chen, W., Gong, R., Ying, T., Prabakaran, P., Zhu, Z., Feng, Y., and Dimitrov, D. S. (2014) Discovery of novel candidate therapeutics

and diagnostics based on engineered human antibody domains. *Curr. Drug Discovery Technol.* 11, 28–40.

(6) Ruigrok, V. J., Levisson, M., Eppink, M. H., Smidt, H., and van der Oost, J. (2011) Alternative affinity tools: more attractive than antibodies? *Biochem. J.* 436, 1–13.

(7) Gebauer, M., and Skerra, A. (2009) Engineered protein scaffolds as next-generation antibody therapeutics. *Curr. Opin. Chem. Biol.* 13, 245–255.

(8) Hosse, R. J., Rothe, A., and Power, B. E. (2006) A new generation of protein display scaffolds for molecular recognition. *Protein Sci.* 15, 14–27.

(9) Dundas, C. M., Demonte, D., and Park, S. (2013) Streptavidin-biotin technology: improvements and innovations in chemical and biological applications. *Appl. Microbiol. Biotechnol.* 97, 9343–9353.

(10) Laitinen, O. H., Nordlund, H. R., Hytonen, V. P., and Kulomaa, M. S. (2007) Brave new (strept)avidins in biotechnology. *Trends Biotechnol.* 25, 269–277.

(11) Hytonen, V. P., Laitinen, O. H., Airenne, T. T., Kidron, H., Meltola, N. J., Porkka, E. J., Horha, J., Paldanius, T., Maatta, J. A., Nordlund, H. R., Johnson, M. S., Salminen, T. A., Airenne, K. J., Yla-Herttuala, S., and Kulomaa, M. S. (2004) Efficient production of active chicken avidin using a bacterial signal peptide in *Escherichia coli*. *Biochem. J.* 384, 385–390.

(12) Repo, S., Paldanius, T. A., Hytonen, V. P., Nyholm, T. K., Halling, K. K., Huuskonen, J., Pentikainen, O. T., Rissanen, K., Slotte, J. P., Airenne, T. T., Salminen, T. A., Kulomaa, M. S., and Johnson, M. S. (2006) Binding properties of HABA-type azo derivatives to avidin and avidin-related protein 4. *Chem. Biol.* 13, 1029–1039.

(13) Gonzalez, M., Argarana, C. E., and Fidelio, G. D. (1999) Extremely high thermal stability of streptavidin and avidin upon biotin binding. *Biomol. Eng.* 16, 67–72.

(14) Maatta, J. A., Eisenberg-Domovich, Y., Nordlund, H. R., Hayouka, R., Kulomaa, M. S., Livnah, O., and Hytonen, V. P. (2011) Chimeric avidin shows stability against harsh chemical conditions—biochemical analysis and 3D structure. *Biotechnol. Bioeng.* 108, 481–490.

(15) Laitinen, O. H., Hytonen, V. P., Nordlund, H. R., and Kulomaa, M. S. (2006) Genetically engineered avidins and streptavidins. *Cell. Mol. Life Sci.* 63, 2992–3017.

(16) Nordlund, H. R., Laitinen, O. H., Hytonen, V. P., Uotila, S. T., Porkka, E., and Kulomaa, M. S. (2004) Construction of a dual chain pseudotetrameric chicken avidin by combining two circularly permuted avidins. *J. Biol. Chem.* 279, 36715–36719.

(17) Hytonen, V. P., Nordlund, H. R., Horha, J., Nyholm, T. K., Hyre, D. E., Kulomaa, T., Porkka, E. J., Marttila, A. T., Stayton, P. S., Laitinen, O. H., and Kulomaa, M. S. (2005) Dual-affinity avidin molecules. *Proteins: Struct., Funct., Genet.* 61, 597–607.

(18) Leppiniemi, J., Maatta, J. A., Hammaren, H., Soikkeli, M., Laitaoja, M., Janis, J., Kulomaa, M. S., and Hytonen, V. P. (2011) Bifunctional avidin with covalently modifiable ligand binding site. *PLoS One* 6, e16576.

(19) Riihimäki, T. A., Kukkurainen, S., Varjonen, S., Horha, J., Nyholm, T. K., Kulomaa, M. S., and Hytonen, V. P. (2011) Construction of chimeric dual-chain avidin by tandem fusion of the related avidins. *PLoS One* 6, e20535.

(20) Nordlund, H. R., Hytonen, V. P., Horha, J., Maatta, J. A., White, D. J., Halling, K., Porkka, E. J., Slotte, J. P., Laitinen, O. H., and Kulomaa, M. S. (2005) Tetraivalent single-chain avidin: from subunits to protein domains via circularly permuted avidins. *Biochem. J.* 392, 485–491.

(21) Hytonen, V. P., Maatta, J. A., Nyholm, T. K., Livnah, O., Eisenberg-Domovich, Y., Hyre, D., Nordlund, H. R., Horha, J., Niskanen, E. A., Paldanius, T., Kulomaa, T., Porkka, E. J., Stayton, P. S., Laitinen, O. H., and Kulomaa, M. S. (2005) Design and construction of highly stable, protease-resistant chimeric avidins. *J. Biol. Chem.* 280, 10228–10233.

(22) Riihimäki, T. A., Hiltunen, S., Rangl, M., Nordlund, H. R., Maatta, J. A., Ebner, A., Hinterdorfer, P., Kulomaa, M. S., Takkinen, K., and Hytonen, V. P. (2011) Modification of the loops in the ligand-

binding site turns avidin into a steroid-binding protein. *BMC Biotechnol.* 11, 64–6750–11–64.

(23) Green, N. M. (1975) Avidin. *Adv. Protein Chem.* 29, 85–133.

(24) Livnah, O., Bayer, E. A., Wilchek, M., and Sussman, J. L. (1993) Three-dimensional structures of avidin and the avidin-biotin complex. *Proc. Natl. Acad. Sci. U. S. A.* 90, 5076–5080.

(25) Green, N. M. (1965) A Spectrophotometric Assay for Avidin and Biotin Based on Binding of Dyes by Avidin. *Biochem. J.* 94, 23C–24C.

(26) Igawa, T., Tsunoda, H., Kuramochi, T., Sampei, Z., Ishii, S., and Hattori, K. (2011) Engineering the variable region of therapeutic IgG antibodies. *MAbs* 3, 243–252.

(27) Miersch, S., and Sidhu, S. S. (2012) Synthetic antibodies: concepts, potential and practical considerations. *Methods* 57, 486–498.

(28) Schlehuber, S., and Skerra, A. (2002) Tuning ligand affinity, specificity, and folding stability of an engineered lipocalin variant – a so-called ‘anticalin’ – using a molecular random approach. *Biophys. Chem.* 96, 213–228.

(29) Gebauer, M., and Skerra, A. (2012) Anticalins small engineered binding proteins based on the lipocalin scaffold. *Methods Enzymol.* 503, 157–188.

(30) Richter, A., Eggenstein, E., and Skerra, A. (2014) Anticalins: exploiting a non-Ig scaffold with hypervariable loops for the engineering of binding proteins. *FEBS Lett.* 588, 213–218.

(31) Yuan, L., Kurek, I., English, J., and Keenan, R. (2005) Laboratory-directed protein evolution. *Microbiol. Mol. Biol. Rev.* 69, 373–392.

(32) Turunen, L., Takkinen, K., Soderlund, H., and Pulli, T. (2009) Automated panning and screening procedure on microplates for antibody generation from phage display libraries. *J. Biomol. Screening* 14, 282–293.

(33) Lou, J., Marzari, R., Verzillo, V., Ferrero, F., Pak, D., Sheng, M., Yang, C., Sblattero, D., and Bradbury, A. (2001) Antibodies in haystacks: how selection strategy influences the outcome of selection from molecular diversity libraries. *J. Immunol. Methods* 253, 233–242.

(34) Rangl, M., Leitner, M., Riihimäki, T., Lehtonen, S., Hytonen, V. P., Gruber, H. J., Kulomaa, M., Hinterdorfer, P., and Ebner, A. (2014) Investigating the binding behaviour of two avidin-based testosterone binders using molecular recognition force spectroscopy. *J. Mol. Recognit.* 27, 92–97.

(35) Green, N. M. (1963) Avidin. 1. the use of (14-C)biotin for Kinetic Studies and for Assay. *Biochem. J.* 89, 585–591.

(36) Flower, D. R. (1993) Structural relationship of streptavidin to the calycin protein superfamily. *FEBS Lett.* 333, 99–102.

(37) Åkerström, B. (2006) *Lipocalins*, pp 204, Molecular biology intelligence unit, Landes Bioscience, Georgetown, TX.

(38) Eisenberg-Domovich, Y., Hytonen, V. P., Wilchek, M., Bayer, E. A., Kulomaa, M. S., and Livnah, O. (2005) High-resolution crystal structure of an avidin-related protein: insight into high-affinity biotin binding and protein stability. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 61, 528–538.

(39) Pugliese, L., Coda, A., Malcovati, M., and Bolognesi, M. (1993) Three-dimensional structure of the tetragonal crystal form of egg-white avidin in its functional complex with biotin at 2.7 Å resolution. *J. Mol. Biol.* 231, 698–710.

(40) Niederhauser, B., Siivonen, J., Maatta, J. A., Janis, J., Kulomaa, M. S., and Hytonen, V. P. (2012) DNA family shuffling within the chicken avidin protein family - A shortcut to more powerful protein tools. *J. Biotechnol.* 157, 38–49.

(41) Lehtonen, S. I., Taskinen, B., Ojala, E., Kukkurainen, S., Rahikainen, R., Riihimäki, T. A., Laitinen, O. H., Kulomaa, M. S., and Hytonen, V. P. (2015) Efficient preparation of shuffled DNA libraries through recombination (Gateway) cloning. *Protein Eng., Des. Sel.* 28, 23–28.

(42) Taskinen, B., Airenne, T. T., Janis, J., Rahikainen, R., Johnson, M. S., Kulomaa, M. S., and Hytonen, V. P. (2014) A novel chimeric avidin with increased thermal stability using DNA shuffling. *PLoS One* 9, e92058.

(43) Swillens, S. (1995) Interpretation of binding curves obtained with high receptor concentrations: practical aid for computer analysis. *Mol. Pharmacol.* 47, 1197–1203.

(44) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Genet.* 65, 712–725.

(45) Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802.

(46) Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* 26, 795–800.

(47) Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMACS for the refinement of macromolecular crystal structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 67, 355–367.