Novel methods and screens for portein stability engineering

Jose, Asial Ignacio

2014

Jose, A. I. (2014). Novel methods and screens for portein stability engineering. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/58910>

<https://doi.org/10.32657/10356/58910>

Singapore International Graduate Award Scholarship

NOVEL METHODS AND SCREENS FOR PROTEIN STABILITY ENGINEERING

ASIAL IGNACIO JOSE

2014

Supervisors

Prof. Pär Nordlund

Dr. Tobias Cornvik

School of Biological Sciences - Division of Biomedical Structural Biology
61 Biopolis Drive, Proteos – Singapore 138673

NOVEL METHODS AND SCREENS FOR PROTEIN STABILITY ENGINEERING

ASIAL IGNACIO JOSE

School of Biological Sciences

A thesis submitted to the Nanyang Technological University in
partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2014

ACKNOWLEDGEMENTS

I would like to thank Prof. Pär Nordlund and Dr. Tobias Cornvik for accepting me in their laboratory. Thank you for your support, for allowing me to work on the exciting projects presented in this manuscript, as well as for providing a stimulating work environment. I am very thankful for the guidance and fruitful discussions that allowed me to grow as a researcher. I would like to stress the importance of Dr. Tobias Cornvik's close supervision of the projects presented in this report, which allowed me to progress smoothly with my research. I thank also Nanyang Technological University and Singapore International Graduate Award Scholarship for allowing me to join the University, and providing stipend during this period.

Many thanks to the people that have been directly involved in these projects: Max Wu, Yue Xiang Cheng, Henrik Engman, Priya Ramani and Henrik Björkling. Your contribution to this work was very valuable, and I'm deeply thankful for your hard work, for your many suggestions and pertinent questions, as well as for your support in difficult times.

I am extremely grateful to all the people that have supported me, advised me, encouraged me and contributed to my wellbeing and happiness throughout these four years: Martina Nilsson, Josefine Karlsson, Lars-Göran Dahlgren, Fariborz Nasertorabi, Andreas Larsson, Anna Jansson, Yin Hoe Yau, Susana Geifman Shochat, Kelly Hew, Dan Chen, Kit Yeung, Lucy Pan, Yin Hua Chen, Agnes Jong, Dina Darwis, Daniel Sim, Hsiang Ling Teo, Aline Reynaud, Julien Lescar, Albert Gustov, Jaslyn Lee, Wai Chung Ng, Andrienne Tho, Misa Noda, James Fong, Irene Fong, among many others. Special thanks to Yi Jun Lin for her support.

Many thanks to my family in Argentina, for understanding and supporting me throughout these many years. To the Mars family, who is always a pillar in my life. To Gaëlle Pencreac'h, Leon Otten, Pascale de Philip, Goetz Parsiegla, Pedro Coutinho, Eric Bourhis, Lauric Haber and Germaine Fuh, for your key contributions to my career. To my friends, for your friendship. To all of you, my deepest gratitude.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	i
TABLE OF CONTENTS	ii
FIGURES	v
TABLES.....	viii
ABBREVIATIONS	ix
ABSTRACT	xi
INTRODUCTION.....	1
1. Proteins: folding, unfolding and aggregation.....	1
1.1. Protein synthesis and folding.....	1
1.2. Protein unfolding and aggregation	3
2. Applications of aggregation resistant proteins	4
2.1. Protein stability in the biopharmaceutical industry	4
2.2. Protein stability for industrial biocatalysis	5
2.3. Protein stability for research applications	6
3. Approaches to improve the thermal aggregation-resistance of proteins	7
3.1. Buffer optimization and formulation.....	7
3.2. Protein construct optimization.....	8
3.3. Stabilization by mutagenesis.....	8
3.4. Stabilization by interaction with ligand.....	9
4. Protein stabilization by mutagenesis.....	10
4.1. Current approaches to engineer thermostable protein variants.....	10
4.2. Directed evolution strategies	13
4.3. Current screening strategies and their problems.....	15
4.4. Current activity-independent screens: advantages and limitations	16
4.5. Characterization of thermal stability of proteins variants	19
5. Combinatorial approaches for the generation of binders <i>de novo</i>.....	25
5.1. Polyclonal and Hybridoma technologies, and their limitations	25
5.2. Combinatorial approaches.....	26
5.3. Phage display	27

5.4. Library generation strategies.....	29
5.5. Libraries with a biased amino-acid composition.....	31
6. Scaffolds for the generation of binders	32
7. Stabilizing binders as crystallization chaperones	34
8. Methods to validate protein-protein interactions.....	35
8.1. Enzyme-linked immunosorbent assay	35
8.2. Surface plasmon resonance.....	38
8.3. Biolayer interferometry	40
9. Recombinant protein production and purification.....	41
9.1. The expression host.....	41
9.2. DNA template and construct optimization	42
9.3. Cloning of targets.....	43
9.4. Production and purification.....	46
10. Projects, aim and strategy	48
10.1. Protein stabilization by mutagenesis	48
10.2. Protein stabilization by interaction with a protein or peptide partner	50
MATERIALS AND METHODS	52
1. Materials	52
2. Methods.....	53
2.1. Cloning of the target genes	53
2.2. Determination of Temperature of cellular aggregation (T_{cagg})	54
2.3. Error-prone library generation	56
2.4. Library screening.....	57
2.5. Protein production and purification	57
2.6. Biophysical characterization	59
2.7. Crystallization and structural determination.....	60
2.8. Surface Plasmon Resonance.....	61
2.9. Biolayer Interferometry.....	61
2.10. Long-term stability study for IL1RA.....	62
2.11. TEV protease activity	62
2.12. DHAA activity.....	63
2.13. Generation of a phage display library	63
2.14. Phage display library panning.....	66
2.15. Validation of phage display-selected clones by phage-ELISA.....	67
2.16. Protein stabilization by interaction with a partner	68
2.17. Generation of a co-expression library with a binary code	69

2.18. Screening for binders by interaction-Colony Filtration	70
RESULTS.....	71
1. Target selection, production and purification	71
2. Screening for stabilized protein variants by Hot-Colony Filtration	72
2.1. Thermal stress leads to protein aggregation <i>in vivo</i>	72
2.2. A direct read-out for screening improvements in protein stability in directed evolution approaches	75
2.3. Evolution of protein thermostability for structural biology	78
2.4. Stabilizing biotechnology tool-proteins and protein drugs	81
2.5. Stabilization of enzymes	97
3. Stabilization of proteins by interaction with a partner	107
3.1. Selection of binders by phage display	107
3.2. Stabilization of proteins by interaction with a partner	117
DISCUSSION	132
1. Identification of stabilized protein variants by Hot-Colony Filtration	132
2. Stabilization of proteins by interaction with a partner	134
2.1. Generation of binders by phage display	134
2.2. Stabilization of proteins by interaction with a partner	136
CONCLUSION AND FUTURE DIRECTIONS	139
BIBLIOGRAPHY	141
SUPPLEMENTARY INFORMATION	150
AUTHOR'S PUBLICATIONS	155
POSTERS	156

FIGURES

FIGURE 1: ENERGY LANDSCAPE OF PROTEIN FOLDING AND AGGREGATION.	2
FIGURE 2: GENERAL OVERVIEW OF A TYPICAL DIRECTED EVOLUTION APPROACH TO IMPROVE A DESIRED PROTEIN CHARACTERISTIC.	12
FIGURE 3: OVERVIEW OF THE MAIN APPROACHES AVAILABLE FOR THE RANDOMIZATION OF DNA SEQUENCES.	13
FIGURE 4: SCHEMATIC REPRESENTATION OF CURRENT ACTIVITY INDEPENDENT SCREENS.	18
FIGURE 5: TYPICAL CURVES OBTAINED IN BIOPHYSICAL CHARACTERIZATION EXPERIMENTS TO DETERMINE PROTEIN STABILITY.	21
FIGURE 6: SCHEMATIC ILLUSTRATION OF A PROTEIN CRYSTALLIZATION PHASE DIAGRAM.	23
FIGURE 7: DETERMINATION OF PROTEIN STRUCTURE BY X-RAY CRYSTALLOGRAPHY.	24
FIGURE 8: SCHEMATIC REPRESENTATION OF ANTIBODY GENERATION BY POLYCLONAL AND HYBRIDOMA TECHNOLOGY.	26
FIGURE 9: SELECTION OF <i>DE NOVO</i> GENERATED BINDERS AGAINST A DESIRED ANTIGEN BY PHAGE DISPLAY.	29
FIGURE 10: SCHEMATIC REPRESENTATION OF THE KUNKEL MUTAGENESIS METHOD.	31
FIGURE 11: NEW DEVELOPMENTS IN BINDING MOLECULES.	34
FIGURE 12: ENZYME-LINKED IMMUNOSORBENT ASSAY (ELISA)	37
FIGURE 13: PRINCIPLE OF SURFACE PLASMON RESONANCE (SPR).	39
FIGURE 14: PRINCIPLE OF BIOLAYER INTERFEROMETRY (BLI).	40
FIGURE 15: SCHEMATIC REPRESENTATION OF LIGATION INDEPENDENT CLONING (LIC).	45
FIGURE 16: SCHEMATIC REPRESENTATION OF THE COLONY FILTRATION BLOT (CoFi).	48
FIGURE 17: SCHEMATIC REPRESENTATION OF THE HOT COLONY FILTRATION BLOT (HOT-CoFi). ..	50
FIGURE 18: SCHEMATIC REPRESENTATION OF THE INTERACTION COLONY FILTRATION BLOT (iCoFi)	51
FIGURE 19: DETERMINATION OF THE TEMPERATURE OF CELLULAR AGGREGATION (T_{cagg}) OF DIVERSE PROTEINS BY HOT COLONY FILTRATION	74
FIGURE 20: EVOLVING PROTEIN THERMAL STABILITY BY RANDOM MUTAGENESIS AND HOT-CoFi SCREENING.	77
FIGURE 21: STRUCTURE OF MUTANT NXR1-1 AT 1.5 Å (LIGHT BLUE – PDB 4HVY), SUPERIMPOSED TO WILD-TYPE STRUCTURE (GRAY, TRANSPARENT – PDB 3GG6).	79
FIGURE 22: ANALYSIS OF MUTATIONS OBTAINED FOR ANTI-GRB2 SCFV.	82

FIGURE 23: SPR DATA FOR ANTI-GRB2 WT AND MUTANTS.	83
FIGURE 24: DETERMINATION OF THE TEMPERATURE OF CELLULAR AGGREGATION T_{cagg} FOR THE VARIANTS OF 4D5-VH SELECTED ON THE SECOND ROUND OF HOT-CoFl.	85
FIGURE 25: TCAGG ESTIMATION FOR 2 ND GENERATION VARIANTS WITH SER ₁₀ CDR-H3.	86
FIGURE 26: SOLUBILITY AND RESISTANCE TO TEMPERATURE OF THE BEST MUTANTS AND WT, MONITORED FROM 40°C TO 95°C.	86
FIGURE 27: RESIDUES IN THE VH DOMAIN FAVORABLE FOR STABILIZATION.....	88
FIGURE 28: STRUCTURAL ANALYSIS OF VH-36 AND COMPARISON WITH THE VH DOMAIN FROM 4D5- FAB.....	91
FIGURE 29: CONTRIBUTIONS OF EACH MUTATION TO THE STABILITY OF VH-36.	92
FIGURE 30: SPR EXPERIMENTS OF IL1RA (A) WT, (B) IL1RA-2 AND (C) IL1RA-8, IN TRIPLICATES.	94
FIGURE 31: STRUCTURE OF IL1RA (BLUE) COMPLEXED WITH IL1R (ORANGE).	95
FIGURE 32: STABILITY OF IL1RA AND MUTANTS.	96
FIGURE 33: ACTIVITY ASSAY FOR TEV VARIANTS OBTAINED FROM THE FIRST ROUND OF SCREENING BY HOT-CoFl.	99
FIGURE 34: ACTIVITY ASSAY FOR TEV VARIANTS OBTAINED FROM THE SECOND ROUND OF SCREENING BY HOT-CoFl.	100
FIGURE 35: ACTIVITY CHARACTERIZATION OF TEV WT AND MUTANTS AT ROOM-TEMPERATURE WITHOUT (A) AND WITH (B) A 20MIN HEAT-CHALLENGE AT 60°C.	101
FIGURE 36: TEV STRUCTURE SHEDS LIGHT INTO THE STABILIZING EFFECT OF MUTATIONS.	102
FIGURE 37: DHAA WT AND MUTANTS, ACTIVITY AND STABILITY.....	103
FIGURE 38: INFLUENCE OF ENZYME CONCENTRATION IN TCP DEHALOGENATION BY DHAA.....	106
FIGURE 39: FN3 PHAGE DISPLAY LIBRARY CONSTRUCTION AND ANALYSIS.....	109
FIGURE 40: SINGLE-POINT PHAGE ELISA IN 96-WELL FORMAT FOR ALL TARGETS.....	111
FIGURE 41: PRELIMINARY BINDING ANALYSIS OF CLONES SELECTED BY COMPETITIVE PHAGE ELISA.	113
FIGURE 42: BINDING KINETIC ANALYSIS OF SELECTED CLONES BY SURFACE PLASMON RESONANCE.	115
FIGURE 43: BINDING KINETIC ANALYSIS OF SELECTED CLONES BY BIOLAYER INTERFEROMETRY....	116
FIGURE 44: PROTEIN STABILIZATION THROUGH INTERACTION WITH A PARTNER.....	118
FIGURE 45: COMPLEX STRUCTURE OF TWO TARGETS STUDIED BY INTERACTION CoFl.	121
FIGURE 46: PROTEIN STABILIZATION BY INTERACTION WITH PARTNER, MONITORED BY DIFFERENTIAL SCANNING FLUORIMETRY	122

FIGURE 47: ISOTHERMAL AGGREGATION CURVES OF MBP AND NPDZ, ALONE OR WITH DIFFERENT RATIOS OF THEIR CORRESPONDING PARTNER, DETERMINED BY iCoFi.....	124
FIGURE 48: IDENTIFICATION OF BINDERS BY INTERACTION-COLONY FILTRATION.	126
FIGURE 49: VH LIBRARY DESIGN.....	128
FIGURE 50: CONFIRMATION iCoFi FOR ANTI-GRB2 VH CLONES VH.AGRB2.1 AND VH.AGRB2.2..	129
FIGURE 51: BIOLAYER INTERFEROMETRY CHARACTERIZATION OF INTERACTIONS BETWEEN GRB2-SH2 DOMAIN AND DIFFERENT VH DOMAINS SELECTED BY INTERACTION-CoFi.....	131

TABLES

TABLE 1: EXAMPLES OF PROTEINS ENGINEERED FOR STABILIZATION, AND THEIR INDUSTRIAL IMPORTANCE.	9
TABLE 2: MAIN STRATEGIES TO ENGINEER PROTEIN STABILITY.	10
TABLE 3: COMPARISON OF MAIN STABILITY SCREENING OR SELECTION TECHNIQUES.....	15
TABLE 4: SUMMARY OF VECTORS USED IN THIS STUDY	54
TABLE 5: DESCRIPTION OF TARGETS STABILIZED	76
TABLE 6: SUMMARY OF MUTANTS OBTAINED FOR NXR1	78
TABLE 7: NXR1-1 STRUCTURE DATA COLLECTION AND REFINEMENT STATISTICS	80
TABLE 8: SUMMARY OF MUTANTS OBTAINED FOR ANTI-GRB2 ScFv.....	81
TABLE 9: VH-36 STRUCTURE DATA COLLECTION AND REFINEMENT STATISTICS	90
TABLE 10: SUMMARY OF MUTANTS OBTAINED FOR IL1RA.....	93
TABLE 11: SUMMARY OF MUTANTS OBTAINED FOR TEV	98
TABLE 12: SUMMARY OF MUTANTS OBTAINED FOR DHAA	104
TABLE 13: SEQUENCE OF BC AND FG LOOPS FOR SELECTED FN3 PHAGE CLONES	112
TABLE 14: TARGETS TESTED FOR STABILIZATION BY INTERACTION WITH PARTNER.....	117
TABLE 15: SUMMARY OF THE TEMPERATURES OF CELLULAR AGGREGATION (T_{CAGG}) OBTAINED BY INTERACTION CoFi BLOT FOR THREE DIVERSE TARGETS	119
TABLE 16: SUMMARY OF THE HALF-LIFE ($T_{1/2}$) OBTAINED BY ISOTHERMAL AGGREGATION MONITORED BY iCoFi BLOT.....	123
TABLE 17: SEQUENCES OF ANTI-GRB2 VH CLONES SELECTED	129

ABBREVIATIONS

Ag: antigen	Fc: antibody constant (or crystallizable) fragment
APH: aminoglycoside phosphotransferase	FN3: fibronectin type III domain
AU: arbitrary units	Fv: variable fragment
BLI: biolayer interferometry	GFP: green-fluorescent protein
bp: base-pairs	GPCR: G-protein coupled receptor
BSA: bovine serum albumin	GST: glutathione S transferase
CCD: charge-coupled device	h: hours
CD: circular dichroism	HHV1: Herpes Simplex virus-1
CDR: complementarity determining region	Hot-CoFi: CoFi at high temperature
CH2: antibody constant heavy domain 2	HRP: horseradish peroxidase
CoFi: Colony-Filtration	IC₅₀: inhibitory concentration-50
DARPin: designed ankyrin repeat protein	iCoFi: interaction-CoFi
DCP: 1,2-dichloropropane	Ig: immunoglobulin
DCB: 1,2-dichlorobutane	IL1RA: interleukin-1 receptor antagonist
DHAA: haloalkane dehalogenase	IMAC: immobilized metal ion affinity chromatography
DNA: deoxyribonucleic acid	IMP: integral membrane protein
DSC: differential scanning calorimetry	IPTG: isopropyl β -D-1-thiogalactopyranoside
DSF: differential scanning fluorimetry	ITC: isothermal titration calorimetry
DSLS: differential static light scattering	KSHV: Kaposi's Sarcoma-Associated Herpesvirus
EBV: Epstein-Barr virus	LB: Luria-Bertani media
<i>E. coli:</i> <i>Escherichia coli</i>	LIC: ligation independent cloning
ELISA: enzyme-linked immunosorbent assay	LIC-Fw: LIC forward overhang
epPCR: error-prone PCR	LIC-Rv: LIC reverse overhang
Fab: antibody antigen-binding fragment	MBP: maltose-binding protein
	MEGAWHOP: megaprimer PCR of whole plasmid
	min: minutes

mRNA: messenger ribonucleic acid

MW: molecular weight

NAD⁺: Nicotinamide adenine dinucleotide

NMR: nuclear magnetic resonance

NPDZ: N-terminal PDZ domain from protein Harmonin

NXR1: NUDIX hydrolase domain of NUDT18

nt: nucleotides

ORF: open reading frame

PAGE: poly-acrylamide gel electrophoresis

PBM: PDZ binding motif

PCR: polymerase chain reaction

PEG: polyethylene glycol

RIA: radioimmunoassay

RMSD: root-mean-square deviation

rpm: revolutions per minute

RT : room-temperature

SAM: sterile alpha motif from protein Sans

scFv: single-chain variable fragment

SDS: sodium dodecyl sulfate

SEC: size exclusion chromatography

sec: seconds

SPR: surface plasmon resonance

ssDNA: single-stranded DNA

StEP: staggered extension process

T_{1/2}: half-life

T_{agg}: aggregation temperature

TB: Terrific Broth

T_{cagg}: cellular aggregation temperature

TCEP: tris(2-carboxyethyl)phosphine

TCP: 1,2,3-trichloropropane

TEV: Tobacco-Etch virus

T_m: melting temperature

T. thermophilus: *Thermus thermophilus*

USD: United States dollars

V-NAR: shark variable new antigen receptor

VH: antibody variable domain from the heavy chain

VL: antibody variable domain from the light chain

WT: wild-type

ΔASA: buried accessible surface area

ΔC_p: heat capacity

ΔH_m: enthalpy

ΔS: entropy

ABSTRACT

Novel methods and screens for protein stability engineering

Protein stability is often a limiting factor in the development of commercial proteins and biopharmaceuticals, as well as for biochemical and structural studies. In this thesis, I present different methods developed with the aim of improving protein stability, either by mutagenesis or by interaction with a protein partner.

For protein stabilization through mutagenesis, a new high-throughput colony-based stability screen is described, which is a direct and biophysical read-out of intrinsic protein stability. As the method is generic and activity independent, it can easily be applied to a wide range of proteins.

For protein stabilization through interaction with a protein partner, the use of phage display was explored. As an alternative and complement to phage display, I present the development of a new biophysical screen that monitors protein stabilization upon binding in cells and crude extracts, and its use in *de novo* binder generation using an autonomous VH domain as scaffold.

INTRODUCTION

1. Proteins: folding, unfolding and aggregation

This thesis deals with the study and improvement of protein thermal stability inside the cell. Different approaches were explored, including stability improvement by directed evolution, and stabilization by interaction with a partner. In this section, important concepts to understand this work are explained.

1.1. Protein synthesis and folding

Protein synthesis within the cell is a process involving a myriad of elements from the molecular machinery that leads to the recognition of a deoxyribonucleic acid (DNA) molecule, its transcription into a messenger ribonucleic acid (mRNA), to finally be translated into a polypeptide sequence, which folds into a specific three-dimensional native structure capable of performing one or many specific functions (1). After synthesis, protein quality control and the maintenance of proteome homeostasis is achieved by an integrated network of several hundreds of proteins, such as molecular chaperones and their regulators, the ubiquitin-proteasome system, and the autophagy system. The chaperone system consists of a group of proteins that assist in the folding of proteins by interacting with the folding intermediates and preventing them from aggregating. They are also involved in their rescue and re-folding in case of destabilization and misfolding. The ubiquitin-proteasome and the autophagy systems mediate the timely removal of irreversibly misfolded and aggregated proteins (2).

Protein folding is a complex process in which a one-dimensional, unstructured, nascent amino-acid sequence adopts its specific three-dimensional native structure, through a process that can happen in a microsecond. Even though this

is one of the most fundamental processes in biology, it is still poorly understood. Current hypothesis of protein folding agree that the energy landscape of protein folding is funnel shaped: there are many open, unfolded, high-energy structures but few low-energy, native-like conformations (**Figure 1**). The polypeptide adopts random conformations, which strive towards the states of lowest free energy, following the laws of thermodynamics. To do so, protein sequences first fold in local nucleation units of secondary structures, such as helices, sheets and turns, followed by growth into more global structures. Small < 100-residue proteins tend to acquire their native three-dimensional structure rapidly, but for the remaining 90% of the proteins (which contain more than 100 residues) folding intermediates are the rule. The bottom of the energy landscape funnel is rugged; i.e. there is a range of co-existing conformation isomers, all of which are at an energy minimum. The more flexible a protein is, the larger the ensemble of conformers (3-5).

Figure 1: Energy landscape of protein folding and aggregation.

Extracted from (4). The purple surface shows the multitude of conformations funneling to the native state. The pink area shows the many conformations moving towards amorphous aggregates and amyloid fibrils.

The main factors contributing to protein folding are intramolecular interactions: hydrogen bonds, van der Waals interactions, backbone angle preferences, electrostatic interactions, hydrophobic interactions and chain entropy (3). However, the folding intermediates are also exposed to intermolecular interactions, which can lead to protein aggregation.

1.2. Protein unfolding and aggregation

Protein aggregation is defined as the association of two or more non-native protein molecules, which tend to collapse in aqueous solution. Aggregation is often irreversible, and mainly driven by hydrophobic forces. The aggregation process usually leads to the formation of amorphous structures composed of denatured proteins interacting together. Alternatively, aggregation can also lead to ordered, fibrillar structures known as amyloids, composed mainly of beta-strands. Amyloid fibrils are toxic to the cells, and are associated with neurodegenerative diseases such as Alzheimer, Parkinson and Huntington's disease, among many others (4, 6, 7).

The aggregation propensity of proteins is greatly enhanced in the crowded environment of the cell. Protein synthesis and folding occurs in the presence of 300-400 g/L of proteins and other macromolecules, which increases the tendency of protein intermediates to interact with other proteins. For this reason, molecular chaperones are usually required to prevent aggregation and aid folding proteins to reach their native structure (4).

Temperature increases can also lead to protein unfolding and aggregation. During the temperature-induced unfolding, proteins undergo a conformational change from a relatively well-ordered structure, characteristic of the low-energy bottom of the folding landscape, to a random-coil structure characteristic of the high-energy top area of the energy landscape. This leads to the exposure of hydrophobic residues to water molecules, as well as to other molecules in solution. Intermolecular interactions through the exposed hydrophobic areas lead to protein aggregation (8). Interestingly, proteins tend to start aggregation

before they are fully unfolded. Partial unfolding of protein is characterized by a molten globule state, where the tertiary structure is absent, while the secondary structure is still preserved. They have large patches of contiguous hydrophobic residues exposed, which are very prone to aggregation (6).

2. Applications of aggregation resistant proteins

The evolutionary pressure on protein stability is limited to the conservation of function in its natural environment, and is dependent on factors such as a specific ionic composition, the presence of protein or small molecule partners, and a given temperature at which the organism coding for the protein lives (9-11). However, in many research and industrial applications the proteins of interest are deprived from their natural environment. When isolated, they are solubilized in an artificial buffer solution, with a simple composition compared to the natural cell environment. Moreover, industrial applications usually require extreme conditions, such as the presence of organic solvents, extreme pH, high salinity and/or high temperatures (10). Therefore, most biotechnological and industrial applications require or would benefit from the availability of more stable protein variants.

2.1. Protein stability in the biopharmaceutical industry

One area that would benefit from aggregation-resistant proteins is the pharmaceutical industry. It represents a market of roughly 600 billion USD, of which 100 billion USD are attributed to the ~200 biopharmaceuticals approved so far. Many of them, such as monoclonal antibodies Rituxan (to treat non-Hodgkin's lymphoma), Herceptin (breast cancer) and Avastin (various cancers) or recombinant proteins erythropoietin (anemia) and insulin (diabetes), are blockbusters representing sales for ~10 billion USD each a year (12).

More efficient strategies of stabilizing proteins could potentially reduce efforts in protein manufacturing and formulation, as well as improve the safety of protein drugs. Biopharmaceuticals that have reached the clinic are generally highly

thermostable (13-16): e.g. human Growth Hormone denatures at 78°C (14), anti-Her2 antibody trastuzumab unfolds at temperatures above 68°C (16), insulin above 72°C (13), and tissue plasminogen activator at 73°C (15). In contrast many protein drug candidates in development are suffering from poor stability (17, 18), which affects efficacy and can lead to increased immunogenicity of the biopharmaceutical; inhibiting their advancement to the clinic. In several cases, such as interferon- α 2a, these effects have been shown to arise from protein aggregation, often induced by exposure to ambient temperatures during protein handling or storage (19, 20). Finally, an increasing number of reports suggest that most proteins or peptides are able to aggregate into amyloid assemblies under suitable destabilizing conditions, such as increased temperatures (7, 21). The peptide hormone glucagon, for example, fibrillates easily if mishandled by prolonged storage at a concentration of 2.5 mg/mL or higher at 37°C (7).

2.2. Protein stability for industrial biocatalysis

As for the pharmaceutical industry, another field where protein stability is of interest is the biotechnology industry, where enzymes are commonly employed. Biocatalysts are used in a wide range of applications, such as in agriculture, paper, textile, the bio-energy industry, or household care, which are multi-billion dollar markets. Industrial enzymes account for roughly 4 billion USD per year of sales, and display an annual growth of 6% (22). To minimize costs, industrial biocatalysis requires enzymes that are highly stable at the conditions where they are applied. Thermal stability, stability in organic solvents, robustness in acid or alkaline environments, among others, are desired properties. For instance, thermal stability of biocatalysts is of major importance for industrial processes, since performing enzymatic reactions at high temperature has many advantages such as higher solubility of the reactants, higher reaction rates, lower viscosity and less microbial contaminants (23, 24). Unfortunately, in most cases, the original enzyme is not directly suitable for the desired application due to the “hostile” conditions at which the reactions are performed. Therefore, during

process development considerable engineering efforts are carried out to stabilize enzymes, particularly to increased temperatures (25-28).

2.3. Protein stability for research applications

Protein stability is also of great importance in basic research as stable proteins are more amenable to biochemical and structural studies. More stable proteins are easier to handle, are more stable during purification and have increased shelf life. Moreover, thermal stability has been shown to increase the likelihood of crystallization of both soluble and membrane proteins, presumably by reducing protein flexibility and extending protein lifetime in the crystallization trial. This concept is now one of the most promising avenues on the path to obtain crystal structures of challenging proteins, such as GPCRs (29, 30) and protein kinases (31).

In directed evolution studies, enzyme robustness is usually necessary so that more mutations can be tolerated. If an enzyme's active site cavity needs to be expanded, for example, to accommodate a new bulkier substrate, this would lead to the destabilization of the mutated enzyme. If the parental enzyme is only marginally stable, the introduction of such mutations would lead to protein aggregation. However, if the enzyme was previously stabilized by the introduction of mutations in other areas, the mutations in the active site could be tolerated. In other words, protein robustness correlates with higher "evolvability", i.e. the capacity of a protein to withstand mutations that would lead to the development of new functions, such as the modification of substrate specificity or the introduction of new catalytic activities for enzymes. The more a protein is stable, the more amenable it is to be evolved to perform a new function, as it tolerates more easily destabilizing mutations that are necessary for a new activity (11, 32).

3. Approaches to improve the thermal aggregation-resistance of proteins

As we have seen previously, protein stabilization is crucial in diverse areas of biological sciences. There have been considerable efforts to develop ways to stabilize proteins, and nowadays there are several validated strategies. Buffer optimization is a straightforward approach, as ionic composition, pH and the presence of stabilizing additives such as detergents, sugars or polyethylene glycols, can have a significant influence in protein stability (7, 33, 34). The introduction of certain mutations can lead to an increase in stability, and there are a great number of successful examples of protein stabilization by this approach in the literature (25, 26, 28, 35-40). Finally, significant developments in screening technologies have made possible the identification of ligands that induce protein stabilization upon binding (33, 41-43).

3.1. Buffer optimization and formulation

Buffer optimization and formulation is the most common way to stabilize a protein. This approach is widely used by the biopharmaceutical industry, the biotechnology industry, and by many research groups.

Ericsson U. B. *et al.* analyzed the influence of twenty-three different buffers on the thermal stability of twenty-five proteins (33). The temperature at which the proteins unfolded varied significantly according to the buffer used. Screening for additives such as sugars, ions, and cofactors, was also shown to be a feasible route to stabilize the proteins tested.

The addition of salts or sugars favors a compact state of proteins in solution, as they are preferentially excluded from the surface of the protein. Protein stabilizers such as polyols, polyethylene glycol, or other polymers sterically hinder protein-protein interactions and limit diffusion, and are commonly used in biopharmaceutical formulation to prevent protein aggregation. Polar amino acids such as arginine and glutamate are also used, as they prevent protein-protein interactions and cover exposed hydrophobic areas. Non-ionic detergents can also sometimes reduce shear-induced aggregation, as well as prevent heat-

induced aggregation. Finally, cyclodextrins are also employed as they suppress aggregation by interaction with hydrophobic residues (7). For biotechnological applications such as biocatalysis, beside the previously mentioned options, protein stability can be increased by immobilization, such as covalent linkage onto a solid support (44).

3.2. Protein construct optimization

Instead of changing the buffer composition, there are numerous methods described where the protein itself is modified. The generation of N- and C-terminal truncated constructs of a given protein can lead to significant variations in recombinant expression levels and solubility, as described in **section 9**. Malawski G. A. *et al.* demonstrated that the different constructs have substantial variations in thermal stability. Moreover, the constructs with higher stability presented a higher crystallization-likelihood. Construct optimization is therefore a feasible route for thermal stability improvement (31).

3.3. Stabilization by mutagenesis

The most common strategy employed to increase protein thermal stability is mutagenesis. Some successful examples of this approach can be found in **Table 1**. Detailed information on the strategies employed for the generation of the improved variants will be provided in **section 4**.

Table 1: Examples of proteins engineered for stabilization, and their industrial importance.

Protein	Property	Applications	Reference
Xylanase	Stability at high temperature.	Digestion of hemicellulose (paper and animal food industry).	(26)
	Stability in alkaline pH.	Bio-bleaching of pulp (paper industry).	(38)
	Stability in the absence of calcium.	Elimination of calcium and its precipitates from fermentation and downstream processing (food industry and multi-enzyme industrial processes).	(35)
Laccase	Stability in organic solvent.	Bleaching of pulp and textile dyeing, organic synthesis, bioremediation.	(40)
α-amylase	Stability at low pH and high temperature.	Starch hydrolysis in low pH industrial processes.	(37)
Phenylacetone Monooxygenase	Stability at high temperature.	Enantioselective oxidation of ketones (Baeyer-Villiger reaction).	(36)
Fructose biphosphate aldolase	Stability at high temperature, in organic solvent.	Synthesis of sugar analogues.	(28)
Lipase	Stability at high temperature.	Digestion of lipids (detergent industry).	(25)

3.4. Stabilization by interaction with ligand

As described in **section 1**, the energy landscape of a protein is funnel-shaped, and rugged at the bottom. The rugged bottom represents the many different conformation isomers co-existing in solution. The interaction of a protein with a ligand re-shapes the energy landscape, shifting the equilibrium of populations towards the conformation that fits best the ligand. This event leads in many cases to stabilization of the protein (5, 45).

There are numerous examples in the literature of stabilization of proteins by interaction with ligands, such as small molecules (33, 41-43), peptides (46-48) and protein partners (49). While many methods are available to validate the

stabilization obtained from known interaction partners (refer to next section), few options are available to screen for new ligands. DSF and DSLS have successfully been used to screen for new small molecule ligands (33, 41, 43), but no ligand-induced stabilization screens have been reported for peptides or proteins. Current approaches for the selection of binding peptides and proteins do not focus on the stabilization of the target, but rather on affinity and specificity towards the target, the most popular technique being phage display (50-53). Phage display technology will be explained in details in **section 5**.

4. Protein stabilization by mutagenesis

4.1. Current approaches to engineer thermostable protein variants

Three main strategies are currently being applied to engineer stable protein variants by mutagenesis: rational design, consensus design and directed evolution (**Table 2**) (10, 11, 54).

Table 2: Main strategies to engineer protein stability.

Technique	Advantages	Disadvantages
Rational design	Small number of variants to test.	Structural information needed. Advantageous mutations are difficult to predict.
Consensus design	Small number of variants to test. No structural knowledge needed. Semi-rational design.	Large set of homologs needed. Phylogenetic bias towards sequences from a common ancestor. Evolution of homologs in nature does not usually correlate with evolution of a single desired property.
Directed evolution	No structural knowledge needed. Rapid identification of improved clones. Identification of non-predictable mutations.	Large library of mutants needed. Rapid and reliable screening method needed.

In rational design approaches, experimental and structural information are used to predict mutations that could lead to a more stable variant of a protein. A small set of possible stable variants are created by site-directed mutagenesis, and assayed for improved stability (10, 55). However, the method can only be applied to well-characterized proteins, for which extensive structural information is available, as stabilizing mutations are difficult to predict. Furthermore, the success rate is low (11).

Consensus design is a semi-rational approach based on the hypothesis that at a given position of a protein sequence, the consensus amino acid of the alignment of protein homologues, contributes more to the stability of the protein than the non-consensus amino acids (10, 56). No structural or experimental evidence is needed to guide the design of the protein variants, which allows its applicability to a wider range of targets. As only consensus protein sequences are tested, the set of protein variants to test for increased stability is small. However, a large group of protein homologues is required to identify statistically relevant consensus sequences. The phylogenic relationship between the aligned homologues is also of extreme importance, as the consensus sequence would be strongly biased towards homolog sequences from a common ancestor. As a consequence, the method may often fail unless large numbers of functional and structurally similar proteins are aligned to reduce phylogenic bias (54).

Directed evolution consists of a Darwinian selection of mutants of a protein among a library of variants generated *in vitro* (**Figure 2**). The improved variants are selected for their higher fitness over the population under a particular selection pressure or screening condition, which correlates to the desired property. As the library of mutants can be generated at random and screened for the desired trait, no prior experimental or structural information is needed (10, 11, 32). Typically 0.01%-0.5% single amino acid mutations are beneficial, 30-50% are strongly deleterious, and 50-70% are neutral or slightly deleterious (11, 32). Therefore large libraries of variants of the protein have to be generated ($>10^4$ individuals) to identify the ones having improved properties. Means to generate highly diverse libraries are thus needed, as well as stringent screening systems to identify the improved variants.

Figure 2: General overview of a typical directed evolution approach to improve a desired protein characteristic.

Modified from (11). A protein target is chosen for optimization, as well as a desired property to improve. Additional information available of the target allows the selection of the best starting DNA sequences as template (e.g. selected gene fragments or codon-optimized ORFs). The sequence is mutated using random mutagenesis, site-directed mutagenesis or a recombination method, generating a library of mutants. A functional assay allows the selection of improved variants from the library. New rounds of sequence diversification and selection are performed until the identification of variants that meet the desired improvement, allowing their biotechnological or industrial application.

4.2. Directed evolution strategies

Libraries of randomized variants of a protein are first created at the DNA level, as this molecular medium is easy to read, manipulate, and amplify. Provided that a link is maintained between the protein variant and its coding gene, selected mutants can be identified by DNA sequencing (57).

Three main directed evolution strategies are available: random mutagenesis, site-directed mutagenesis, and recombination methods (**Figure 3**) (57-61).

Figure 3: Overview of the main approaches available for the randomization of DNA sequences.

Modified from (57).

A: Random mutagenesis is based on the introduction of errors during the amplification of DNA, at random positions. PCR conditions can be optimized to modulate the rate of mutation.

B: Directed mutagenesis is an oligonucleotides-based mutagenesis method for the introduction of selected mutations at specific positions within the open reading frame. The libraries generated are usually of small size and high quality.

C: Recombination methods are based on random fragment shuffling of a group of ORFs, which lead to the generation of new DNA molecules containing portions of different parental sequences. The most employed recombination techniques are DNA shuffling and Staggered Extension Process.

Random mutagenesis is based on the introduction of errors during the amplification of DNA in a PCR reaction. In most common error-prone PCR (epPCR) protocols, a low fidelity DNA polymerase is used in combination with a suboptimal buffer composition, e.g. containing small amounts of Mn^{2+} (instead of the natural cofactor Mg^{2+}) as well as biased concentrations of dNTPs (57, 61).

Site-directed mutagenesis strategies, also known as saturation mutagenesis, are based on the amplification of DNA with oligonucleotides containing randomized codons at specific positions. The libraries created are mutated only at selected positions in the gene, and the oligonucleotides can be synthesized in a biased manner to avoid the introduction of stop codons, or undesired amino acids. These focused libraries are usually of higher quality, and of smaller size, compared to epPCR. However, this method requires prior structural data to guide the selection of which sites to mutate, and as for rational design, it is difficult to predict which positions to mutate (57, 60).

Recombination methods are based on the re-assembly of gene fragments to create libraries of full-length gene chimeras. The most widely used methods are based on homology-driven recombination, also known as *in vitro* sexual recombination, either by DNA shuffling or Staggered Extension Process (StEP) (58, 59). In DNA shuffling, a pool of homolog genes are fragmented in small 50-200 bp fragments by DNase I; the fragments are then randomly re-annealed, and extended to a full-length library of recombined genes by PCR (58, 59, 61). In the StEP method, a pool of full-length homolog genes are used as templates in a PCR reaction; by performing extremely short denaturation and re-annealing/extension cycles, partially extended fragments re-anneal to different templates, thus generating recombined genes (58). These methods require the availability of many homologues of the protein of interest to perform the recombination, or the prior selection of promising variants from epPCR libraries (57, 59).

4.3. Current screening strategies and their problems

When a suitable library has been obtained, the next step in the process is to screen for increased stability. Most stability screening strategies are based on specific activity assays, usually performed in multi-well microtiter plates or in bacteria/yeast colonies (**Table 3**) (25, 26, 28, 35-38, 40, 62).

Table 3: Comparison of main stability screening or selection techniques.

Method	Read-out	Advantages	Disadvantages
Enzymatic screen	Colorimetry or fluorescence	Well established. Only selects active mutants.	Limited screening capacity. Only applicable to enzymes. Expensive. Indirect read-out.
Phage display	Phage binding or infectivity	High screening capacity: $>10^7$ variants per experiment.	Limited to proteins that can be displayed on phage. Indirect read-out.
THR	Resistance to antibiotic	High screening capacity: $>10^7$ variants per experiment Activity-independent. Inexpensive.	Special host system: <i>T. thermophilus</i> . Limited temperature range to test: 60-75 °C. Protein expression at 60-75 °C. Indirect read-out.
Tripartite Fusion System	Resistance to antibiotic	High screening capacity: $>10^7$ variants per experiment Activity-independent. Inexpensive.	Screens for "foldability" and protease resistance, rather than stability. Engineering of appropriate linker needed. Indirect read-out.

Generally, the screening consists of a residual activity read-out, usually by colorimetry or fluorescence, after a denaturation step (typically heat denaturation) (62, 63). These methods are well established and widely employed in the protein-engineering field. However, they are only applicable to enzymes, and there is a need of a specific assay for each enzymatic activity. If a new

enzyme is to be evolved to perform a reaction that has not been studied previously, a new method has to be developed specifically; the new assay has to provide information on activity in a high-throughput manner. This limits the number of enzymes that can be studied by a single research team, and requires specialization on the enzymatic reaction of interest (25, 26, 28, 35-38, 40, 62, 63). In the case of proteins with binding activity, phage display selection systems (described in **section 5**) can be used to select for stabilized variants (64-67). The assay is usually based on the conservation of binding after a denaturation selection step (65-67). The main drawback of these techniques, which limit their general applicability, is the need to display the protein of interest as a fusion to one of the bacteriophage's coat proteins. The fusion protein has to be correctly exported into *E. coli* periplasm, and correctly displayed on the phage surface (65-67).

4.4. Current activity-independent screens: advantages and limitations

There have been attempts to develop generic (thermo)stability screens that are independent of protein function and have enough throughput to screen libraries of significant size: these include Proside (64, 68), THR (69) and the Tripartite Fusion System (70) (**Figure 4**).

Proside technology (64, 68) is a phage-based method that relies on the fact that bacteriophage M13 coat protein III is essential for phage infectivity towards *E. coli* cells. A target protein is expressed as a fusion protein, inserted between the N1- and N2- N-terminal domains and the C-terminal domain of phage protein III. Phages are produced expressing this fusion construct at their surface. Variants for the target protein are generated and displayed on the phage surface; the phages are subjected to incubation in the presence of proteases; finally, *E. coli* cells are infected with the protease-treated phages. Phages displaying protein variants that are more tightly folded are less susceptible to protease digestion and have higher infectivity. After a few rounds of selection, the phage libraries are enriched in clones displaying correctly folded and more rigid variants of the target. This technology relies on an indirect read-out of protein folding through

protease resistance and phage infectivity, rather than thermostability *per se*. Phages that delete the inserted gene would be favored, and could be rapidly enriched. Moreover, the need to display the protein on the bacteriophage's surface considerably limits the number of targets amenable to the technique.

THR (69) is based on the expression of a thermostable variant of kanamycin-acetyltransferase fused to the target gene in *Thermus thermophilus*. Variants of the target protein are generated, transformed into *T. thermophilus* cells and plated onto kanamycin-containing agar-plates. This organism grows at temperatures ranging between 60°C and 75°C. The stability of the target protein influences the overall stability of the fusion construct; less stable variants promote the aggregation of the fusion protein, leading to a reduced resistance to kanamycin. The only variants expected to grow are those with sufficient stability at 60°C – 75°C. The main limitations of the method are the host used, which is not a validated recombinant protein expression system, and the narrow screening temperature. Moreover, as for Proside, there is a selection pressure for the deletion of the target gene.

The Tripartite Fusion System (70) is based on a split β -lactamase containing the target gene inserted between the two fragments and expressed in *E. coli*. Variants of the target protein are created, libraries transformed into bacteria, and plated onto ampicillin-containing agar. Protein variants that express a correctly folded protein allow for the complementation of the split β -lactamase fragments, conferring resistance to ampicillin, while unfolded variants are susceptible to aggregation, and protease degradation, and do not allow to complementation to occur. This method is not a thermostability screen but rather a “foldability” and protease resistance screen without temperature challenge. Furthermore it requires the correct insertion of the target gene in between the two fragments of the reporter. As for the previous methods mentioned, the deletion of the target gene could be a major risk of false positives.

Figure 4: Schematic representation of current activity independent screens.

A. Proside. A library of phages contains different variants of a protein displayed as a fusion to the phage coat g3p protein, which is required for phage infectivity. The library is exposed to proteases; the most stable clones are able to resist digestion while less stable variants are degraded. The undigested phages are able to infect new *E. coli* cells and are amplified.

B. THR. A library is created coding for variants of a protein of interest (red) as a fusion to a thermostable version of kanamycin-acetyltransferase (blue) and transformed into *Thermus Thermophilus*. Bacteria are grown at a temperature ranging from 60°C to 75°C. The most stable clones remain soluble and confer kanamycin resistance, while the least stable variants unfold and aggregate, being kanamycin-sensitive.

C. Tripartite Fusion System. A library is created coding for variants of a protein of interest (red) as a fusion of two halves of beta-lactamase at the N- and C-termini (blue). Variants that are correctly folded allow for the complementation of the N- and C-terminal fragments of β -lactamase, conferring ampicillin resistance. Unfolded or aggregating variants do not allow complementation to occur, and are ampicillin-sensitive.

4.5. Characterization of thermal stability of proteins variants

4.5.1. Biophysical characterization of protein stability

After potential improved variants have been selected by any screening method, the mutants have to be verified. This is usually done by comparing the temperature at which the wild-type protein and the selected mutants unfold and/or aggregate. Biophysical methods are ideal for this purpose, and the two main techniques that have been traditionally employed are Differential Scanning Calorimetry (DSC) and Circular Dichroism (CD) spectroscopy. In the last decade, however, two high-throughput techniques have also emerged: Differential Scanning Fluorimetry (DSF) and Differential Static Light Scattering (DSLS). All the mentioned methods have been cross-validated and, in most cases, they generate very similar denaturation temperatures (33, 41, 43, 71-77).

In DSC experiments (**Figure 5A**), protein samples are subjected to a linear increase in temperature. A reference solution (buffer without protein) undergoes the same increase in temperature, in parallel. The amount of heat necessary to increase the temperature in both solutions is compared, and the difference between the protein and the reference corresponds to the excess heat capacity ΔC_p . As the protein unfolds, the heat capacity increases, reflecting the increased energy necessary to break the tight interactions within the folded protein. The heat capacity reaches a maximum when half of the proteins in the sample are unfolded, and then decreases as the majority of the protein becomes unfolded. The melting curve obtained is sigmoidal, and the maximum ΔC_p corresponds to the melting temperature of the protein, or T_m (71, 74).

CD spectroscopy (**Figure 5B**) is employed to measure and quantify the chirality of proteins subjected to a gradual increase of temperature (or denaturant concentration). This provides information on the evolution of the secondary and tertiary structure of the protein as a function of temperature. The melting temperature T_m is defined as the temperature at which half of the proteins have lost their tertiary structure (and are thus unfolded) (72, 75).

Differential Scanning Fluorimetry (**Figure 5C**) is a relatively new high-throughput technique to monitor the unfolding of up to 96 or 384 protein samples subjected to a gradual increase in temperature. Protein samples are mixed with a fluorescent dye that is quenched in aqueous solution, but fluoresce when present in a hydrophobic environment. As proteins unfold, the hydrophobic core becomes accessible to the dye, which leads to an increase in fluorescence that can be detected using real-time PCR equipment. The maximum fluorescence is obtained when all proteins are unfolded, and the melting temperature T_m is defined as the temperature corresponding to the midpoint of the signal transition (33, 41, 73, 76).

Differential Static Light Scattering (DSLS – **Figure 5D**) is another high-throughput method to monitor the aggregation of proteins as a function of temperature. Proteins are subjected to gradually increasing temperatures and the scattering of light is monitored through a CCD camera. When proteins start to aggregate, they start to scatter light, attaining a maximum when all proteins in the sample aggregate. The temperature of aggregation, T_{agg} , is defined as the temperature at which half of the maximum light scattering signal transition is obtained (43, 76, 77).

Figure 5: Typical curves obtained in biophysical characterization experiments to determine protein stability.

A: Typical Differential Scanning Calorimetry profile (extracted from (71)). As the protein unfolds, the heat necessary to increase the temperature of the protein solution is higher than the heat necessary to obtain the same increase in a reference solution not containing the protein. This heat difference is known as the excess heat capacity, ΔC_p . The melting temperature is defined as the temperature corresponding to the maximum excess heat capacity. Additional thermodynamic information can be analyzed, e.g. enthalpy of unfolding (ΔH_m), which provides information on the secondary structure of the protein. Using ΔC_p and ΔH_m , the entropy ΔS and the standard free energy of the unfolding reaction ΔG can be calculated.

B: Typical UV-Circular Dichroism profile (extracted from (75)). The fraction of unfolded protein corresponds to the ellipticity change of the solution measured at a given wavelength. Far-UV CD (216nm – closed symbols) provides information of the changes in secondary structure of the protein, while near-UV CD (280nm – open symbols) accounts for changes in the tertiary structure. The melting temperature T_m is defined as the temperature corresponding to half of the maximum ellipticity change. In this example, the profiles of two different proteins are superposed, and their corresponding T_m (T_{m-1} and T_{m-2}) indicated.

C: Typical Differential Scanning Fluorimetry profile (extracted from (76)). As the protein unfolds, the hydrophobic core is exposed and interacts with a fluorescent dye. The dye is quenched in aqueous solution, but fluoresce in a hydrophobic environment. The increase in fluorescence correspond thus to the unfolding of the protein with increase in temperature. The melting temperature T_m is defined as the temperature corresponding to half of the maximum fluorescence transition. In this example, the profiles of two different proteins are superposed, and their corresponding T_m (T_{m-1} and T_{m-2}) indicated.

D: Typical Differential Static Light Scattering profile (extracted from (76)). As proteins aggregate, the light scattering of the protein sample increases. This increase in light scattering is detected by a CCD camera. The aggregation temperature T_{agg} of the protein is defined as the temperature corresponding to half of the maximum light scattering intensity transition. In this example, the profiles of two different proteins are superposed, and their corresponding aggregation temperature T_{agg} (T_{agg-1} and T_{agg-2}) indicated.

4.5.2. Structural characterization of stabilized variants

To understand why and how the mutations obtained are contributing to a more stable protein, it is usually very helpful to determine the protein's three-dimensional structure. Structural biology provides atomic understanding of every biological process, and among them, it helps understand the process governing protein function, folding and stability. Protein structural analysis is regularly used in directed evolution studies, as it is a useful tool to rationalize the mutations obtained and how they help improve a particular property.

Two complementary technologies are commonly used: X-ray crystallography and nuclear magnetic resonance (NMR). NMR is useful to study protein structure in solution, and can provide information on the dynamic aspect of protein structure, including flexibility, protein folding as well as the dynamics of protein-ligand interactions. X-ray crystallography provides structural information from protein crystals, and can provide atomic-level resolution, which is particularly useful when studying small changes in protein structure, such as side-chain rearrangements (1). This section focuses on structure determination by X-ray crystallography, as it is the most common technique employed in directed evolution studies, and the method employed in this work.

The process of protein crystallization consists in a change of state of protein molecules, from a soluble form into a solid, ordered form known as protein crystals. Protein crystals are formed under particular conditions in which the solution is brought into supersaturation. This is achieved by varying the concentration of precipitant, protein, additives, pH, temperature, among other parameters (78). This process can be illustrated by a phase diagram (**Figure 6**). It consists of four different zones representing different degrees of supersaturation: a zone of high supersaturation where the protein will precipitate; a zone of moderate supersaturation where spontaneous nucleation takes place; a metastable zone (just below the nucleation zone) of lower supersaturation, where crystals are stable and may grow but no further nucleation can take place; and the zone of undersaturation, where the proteins are fully dissolved and will not crystallize (78).

Figure 6: Schematic illustration of a protein crystallization phase diagram.

The adjustable parameters can be: precipitant or additive concentration, pH, temperature, among others. Four major crystallization methods are presented: (i) microbatch, (ii) vapour diffusion, (iii) dialysis and (iv) free interface diffusion. Image extracted from (79).

Structure determination is accomplished by a series of steps that can be time consuming and require optimizations (**Figure 7**). First, proteins are produced recombinantly and purified in sufficient quantity (refer to **section 9**). Then, crystal screens are set which allow the exploration of hundreds of different conditions that could lead to the formation of crystals: e.g. different buffers, pH, ionic strength, additives. This is typically done in 96-well format crystallization plates, following different techniques (e.g. vapor diffusion, batch; sitting drop, hanging drop). If any condition produce crystals, they can be further optimized by testing slight variations of those conditions. The crystals obtained are extracted from the crystallization plates, soaked in cryoprotectant to avoid the formation of ice upon freezing, and then rapidly froze in liquid nitrogen.

The crystals are subjected to a beam of X-ray radiation at different angles of incidence. Each angle of incidence generates a diffraction pattern, which is collected by a detector. This is typically performed in a synchrotron facility. The diffraction patterns obtained are missing one important information: the phase information for each diffraction obtained. This can be obtained by different

means, including molecular replacement (using a homologue protein for which the structure has been solved) and from anomalous signals (from heavy atoms and seleno-methionine for example). Specialized software is used to generate a three-dimensional protein structure from the diffraction patterns and phase information (80, 81).

Figure 7: Determination of protein structure by X-ray crystallography.

Extracted from (82). Refer to main text for description of each step.

5. Combinatorial approaches for the generation of binders *de novo*

5.1. Polyclonal and Hybridoma technologies, and their limitations

Traditionally the only binding proteins that researchers could isolate were antibodies. The first approach developed, polyclonal antibody generation (**Figure 8**), is still widely used and relies on the injection of a given antigen into animals such as mice, rabbits or lamas. The animals develop an immune reaction, and the serum that contains a mixture of reactive antibodies can be isolated. There are several limitations to this method: not all antigens lead to the isolation of polyclonal antibodies; the samples have low purity; and the volume of serum obtained limits supply (51).

Hybridoma technology (**Figure 8**) was developed more than thirty years ago to alleviate the problems of purity and supply of polyclonal technology. After immunization of animals, immune cells are isolated and fused to tumor cells. This leads to the immortalization of the fusion cells (i.e. hybridomas). Hybridoma cells can be isolated, and each clone probed for expression and secretion of antibodies and antigen binding. The antibodies obtained are monoclonal, i.e. they are unique and not a mixture, and can be produced in large quantities by clonal expansion of the isolated cells. Despite the advantages over polyclonal technology, there are still limitations to the method: the antibody selection takes place in the animals serum, and therefore requires antigens that are stable in this environment; many antigens such as membrane proteins, nucleic acids and protein complexes are less suitable. In addition, it is difficult to improve the affinities of the antibodies obtained, which requires the isolation of the DNA and affinity optimization using other techniques; for biopharmaceutical applications, the antibody has to be “humanized”, a painstaking process of framework mutation to mimic human antibody sequences (51).

Figure 8: Schematic representation of antibody generation by polyclonal and hybridoma technology.

Modified from (51). In polyclonal technology, animals are injected with an antigen, generating an immune response, after which the reactive serum can be isolated. For Hybridoma technology, immune cells are isolated after immunization, and fused to tumor cells to generate immortal hybridoma cells. Each hybridoma cell generated is isolated, and probed to detect the ones secreting reactive antibodies. Those cells can be amplified to produce monoclonal antibodies in large quantities.

5.2. Combinatorial approaches

Combinatorial libraries were developed to circumvent the limitations of polyclonal and hybridoma technology. The main goal is to have more control of the antibody selection process by performing this step not in animals but *in vitro*. DNA libraries that code for millions to trillions of different antibodies are generated *in vitro* either by isolation of DNA from natural immune repertoires (human donors or animals) or by design and *in vitro* synthesis. Each antibody coded by the library is expressed recombinantly, and linked to the coding DNA. The link between genotype and phenotype is the key element of these

technologies, as it allows retrieving the DNA sequence of an antibody after it has been selected to bind a particular antigen (83).

There are several combinatorial technologies to generate and isolate antibodies. These include phage display, ribosome display, yeast display, mRNA display and *in vitro* compartmentation, among others. They all have the advantage of generating antibodies *in vitro*, which allows the careful control of selection conditions. This expands the potential targets for which antibodies can be generated to virtually any molecule, including proteins, DNA, membrane proteins, protein complexes, as well as small molecules, specific protein domains, or proteins with specific post-translational modifications. The antibodies obtained are unique; they can be produced recombinantly in large quantities and purified to high degree. Fully human antibodies can be generated, eliminating the need of humanization. The isolation of the coding DNA permits easy further manipulation to improve affinity, specificity, and stability. DNA subcloning opens the possibility to introduce new functions such as dimerization, multimerization, and fusion to enzymes and tags (51, 83, 84).

Detailed explanations of each method are available in a review by Hoogenboom H. R. (83). Phage display is explained in details in this section, as it is the most widely used combinatorial technique for antibody generation and most relevant to my work.

5.3. Phage display

Phage display technology was developed more than twenty years ago, and has since then evolved constantly into a robust and versatile technique for generation of binders (antibodies, peptides and novel binders) *in vitro* (51, 84-86).

In phage display technology (**Figure 9**), billions of variants of a DNA sequence (coding, for example for a random peptide, or different variants of an antibody) are fused to one of the bacteriophage's coat protein genes, allowing the display of the corresponding protein variants at the surface of the phage. There are several coat proteins that can be used for display, the most popular being protein III and

protein VIII. Protein III is present in the tip of the phage in three to five copies, and is used for monovalent display; this format is preferred for the selection of binders with high affinity and specificity. Protein VIII is present in roughly 3,000 copies, and is therefore employed for multivalent display, this configuration is ideal for the selection of weak binders, as avidity allows sufficient binding to the targets to withstand the stringency of the selection process (87, 88). As each phage particle encapsulates a vector coding for the protein variant displayed on its surface, there is a link between the phenotype observed, and the corresponding genotype (51).

Extremely diverse libraries of peptides or proteins can be subjected to selection procedures *in vitro*. The basic principle behind the selection process is the following: the phages are put in contact with the antigen, which can be immobilized or in solution – the phages with affinity to the target bind; washing steps allow for the removal of non-binding variants; finally, binding phages are retrieved and propagated in *Escherichia coli* cells. This procedure is repeated on the newly propagated phages, to eventually obtain after a few rounds of selection a population dominated by antigen-binding phages (50-53). As the selection is performed in the test tube, the researcher has great freedom in the selection conditions: differences in antigen presentation (immobilized or in solution), number of washes, buffer conditions, temperature, incubation time, the presence of competitors, and counter-selections allow to tailor the selection process for the isolation of the ideal binder (88).

Figure 9: Selection of *de novo* generated binders against a desired antigen by phage display.

Schematic representation based on (51).

A library of phages displaying billions of variants of a protein of interest is put in contact with a target antigen (Ag). The clones with affinity for the target bind, while non-binding phages are removed by washing. Bound phages are retrieved, and used to infect *E. coli* cells. This leads to the amplification of the binding phages, and the generation of a new library enriched with binders. A few rounds of selection (or “panning”) should lead to a library where the majority of members bind the target.

5.4. Library generation strategies

Combinatorial approaches for binder generation rely on the careful design and creation of a library. A very diverse library is required, containing a majority of functional clones, i.e. producing full-length proteins in their folded state. Antibody libraries are the most commonly available; they are usually in the form

of single-chain variable fragments (scFv) or Fab (the different formats would be explained in detail in **section 6**). Two strategies for library generation are widely employed: cloning of antibody sequences from natural repertoires; and design and synthesis of the antigen-binding loops (i.e. complementary determining regions, or CDRs) which are introduced into a single antibody scaffold (i.e. the “template” antibody) (89).

For libraries generated from natural repertoires, the DNA sequences coding for the antibody repertoire are amplified from serum of donors (human or animal) by PCR. The introduction of restriction sites allow for the cloning of the amplified sequences into a vector suitable for phage display. The same strategy can be applied for the introduction of synthetic CDRs; the loops are synthesized as oligonucleotides, which are cloned into a destination vector containing appropriate restriction sites in the antibody framework. Synthetic CDRs can also be introduced by Kunkel mutagenesis (**Figure 10**) (50, 90). In this technique, a single-stranded DNA (ssDNA) sequence of the template vector is produced in a special *E. coli* strain that introduced uracil residues into the DNA sequence. The oligonucleotides containing the combinatorial CDR sequences to be introduced are annealed to the ssDNA template vector. The complementary DNA sequence is produced using a DNA polymerase (typically T7 DNA polymerase), and ligated using a DNA ligase. Upon transformation, *E. coli* cells degrade the uracil-containing template and produce a new strand complementary to the mutated strand, which does not contain uracils.

Figure 10: Schematic representation of the Kunkel mutagenesis method.

5.5. Libraries with a biased amino-acid composition

To generate phage display libraries from synthetic CDRs, careful design is necessary to generate maximal functionality and antigen binding capacity in the library with minimal library size. The transformations efficiencies into *E. coli* limit library size to about 10^{10} unique members. If completely degenerate codons

are used, the theoretical library size is extremely diverse, which leads to a small fraction of clones being physically represented: if we mutate five positions with completely random NNN codons, the theoretical library size is $5^4 = 5 \times 10^4$ unique members, and about 23% of them would contain a stop codon. By restricting the third codon position to either G/C (NNK) or G/T (NNS) only 32 codons are used; all amino-acids are still coded, but the library size is reduced to $5^3 = 2 \times 10^2$ members, and only ~16% of them would contain stop codons.

By varying the codon composition during oligonucleotide synthesis, tailored amino-acid compositions can be obtained, reducing further library size while increasing functionality and binding potential. Pioneering work by Fuh G., Sidhu S.S. and co-workers showed that this approach, when applied to carefully selected positions of an antibody scaffold, leads to highly functional libraries with members able of binding a wide variety of targets with high affinity and specificity (50, 86). The development of phosphoramidite cassette technology, in which tri-nucleotides coding for specific amino-acids are “pre-made” and available to be used for oligonucleotide synthesis, further expanded the potential of tailored amino-acid compositions. Now the exact proportions of each amino-acid can be designed, favoring residues more likely involved in antigen binding such as Tyrosine, or structural diversity such as Glycine and Serine, and eliminating stop codons and undesirable residues such as Cysteine (91, 92).

6. Scaffolds for the generation of binders

Antibody technology has developed greatly in the last thirty years. Polyclonal, hybridoma and phage display technologies have allowed the widespread availability of antibody molecules as detection reagents, purification reagents and biopharmaceuticals. Antibodies remain the gold standard of binding molecules, however over the last decade a great number of alternative binding proteins have emerged to complement antibodies for specific applications (Figure 11).

Immunoglobulins of the IgG type are the norm, but other types such as camel “heavy chain” immunoglobulins and shark Ig-NARs have emerged as alternatives

(Figure 11A). DNA engineering technologies have allowed the dissection of antibody molecules into smaller building blocks that can be re-assembled in a myriad of new original configurations. Camelid V_HH, shark V-NAR, and human V_H or V_L domains exploit the isolated small immunoglobulin variable domain, which has some favorable properties such as fast tissue penetration and rapid clearance from the bloodstream (particularly interesting for radio-imaging). Similarly constant domains (e.g. CH2 domain) can be engineered for analogous applications. Single-chain Fv (scFv) is a popular format in which V_H and V_L domains are joined with a flexible linker. Antibody Fab is composed of a single binding arm of the antibody molecule, and maintains binding to the target while eliminating effector functions present in the Fc fragment. By re-arranging all these building blocks, multivalent and multispecific molecules can be obtained, which is useful to increase affinity for the former, and tissue- or cell-specificity for the latter (93-96).

Non-antibody scaffolds are also gaining popularity (**Figure 11B**). The concept is to take a protein with favorable properties (e.g. high stability, high expression levels in inexpensive *E. coli*, lack of disulfide bridges, lack of post-translational modifications) and use it as template for the generation of binders, by mutating carefully selected residues on its surface. There are numerous successful examples of this approach, such as the Anticalins, Fibronectin type 3 domain (FN3 monobody), Designed Ankirin Repeat Proteins (DARPin), and affibodies (96).

Figure 11: New developments in binding molecules.

A. Immunoglobulin-based scaffolds and new re-engineered formats. Modified from (93).

B. Some examples of alternative non-immunoglobulin scaffolds. Modified from (96).

7. Stabilizing binders as crystallization chaperones

The increasing availability of affinity reagents with high affinity and specificity towards any given target has led to their growing application in structural biology projects. One of the most promising avenues to generate crystal structures of difficult targets, such as membrane proteins, structured RNA molecules, protein kinases, and enzymes in specific conformations, is the use of binders as co-crystallization partners, or “crystallization chaperones”. Binders can lock a target protein into a specific conformation, which increases the homogeneity of the sample. They can also facilitate crystal contacts, favoring the generation of the crystal lattice. They can be used to stop enzymatic reactions at specific stages, facilitating their structural study. The stability of the target can be increased during purification. Structured-RNA molecules can be rendered less flexible by interacting with a binder. For all these potential applications, chaperone-assisted crystallography is currently a very attractive option to increase crystallization success in structural genomic approaches. There are

successful examples of this approach using Fab molecules, as well as non-immunoglobulin binders such as DARPin and FN3 monobodies (97).

8. Methods to validate protein-protein interactions

When a binder has been obtained the interaction with its target has to be confirmed and the affinity determined. There are numerous validated techniques to confirm protein-protein interactions *in vitro*. These include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), chromatography, isothermal titration calorimetry (ITC), surface plasmon resonance (SPR), and biolayer interferometry (BLI) (98, 99). Each technique has its advantages and disadvantages, but only ELISA, SPR and BLI will be described, as those were used for this study.

8.1. Enzyme-linked immunosorbent assay

ELISA (**Figure 12A**) is a robust and widely used technique for the validation of protein-protein interactions, which was developed more than forty years ago (100). It is a solid-phase technique in which one of the interaction partners (e.g. an antigen) is immobilized onto a solid support (microtiter plate), usually by adsorption. The immobilized antigen is then incubated with the analyte (e.g. a primary antibody). After a period of incubation and washing, the bound antigen is detected through an analyte-specific antibody conjugated with Horseradish peroxidase (HRP) or other reporter enzymes (e.g. HRP-conjugated secondary antibody). The assay is developed through colorimetric or fluorescent signal using the reporter enzyme's activity (101), high signal indicating binding.

Two ELISA variants relevant to this work are phage-ELISA (**Figure 12B**) and phage competitive-ELISA (**Figure 12C**) (102). In phage-ELISA, the antigen is immobilized, and probed with a phage displaying the potential binder in its surface. After incubation and washing, the bound phages are detected using an HRP-conjugated anti-phage antibody, and developed as previously.

In phage competitive-ELISA, a constant concentration of phages is incubated with increasing concentrations of antigens in solution (i.e. competitor antigen). As competitor antigen concentration increases, increasing number of phages are blocked. After a period of incubation, the phage-competitor mix is incubated with immobilized antigens. Only free-phages are able to bind the immobilized antigens. Development is performed as for phage-ELISA. At low concentrations of competitor, most phages are in their free form: they bind the immobilized antigen and give strong signal upon development. As competitor concentration increases, more phages are blocked, leading to a decrease of signal until a minimum plateau. The midpoint of the transition is the Inhibitory Concentration-50 (IC_{50}), which correlates with the binder's affinity.

Figure 12: Enzyme-linked immunosorbent assay (ELISA)

A. Standard ELISA. Antigen is immobilized, blocked, and probed with a specific primary antibody. A Horseradish peroxidase (HRP)- conjugated secondary antibody is then used to detect the primary antibody, and developed by chemiluminescence or colorimetry.

B. Phage-ELISA. Same principle as for the standard ELISA, but the primary antibody is replaced by the phage particle displaying the binding protein or peptide on its surface.

C. Phage competitive-ELISA. Phage particles are incubated in the presence of increasing amounts of competitor antigen in solution, followed by incubation with the immobilized antigen. At low competitor concentrations, the majority of phages are available for binding to the immobilized antigen, but as the concentration increases, more phages are blocked. The samples are developed in a similar manner as for phage-ELISA.

8.2. Surface plasmon resonance

Surface plasmon resonance (**Figure 13A**) is a label-free technique developed thirty years ago into a robust technology to study protein-protein and protein-ligand interactions (103). It is a solid-phase method providing direct read-out of the interaction between an analyte in solution and an antigen attached to the surface of a sensor chip. The SPR sensor chip is composed of a glass surface coated with a thin 50-nm layer of gold, which has been derivatized on one side to allow for simple covalent attachment of molecules using different types of chemistries. On the opposite side of the chip, polarized light is shone onto the gold surface through a prism, its reflection monitored by a detector. At a specific angle, the gold electrons are excited, and resonate creating an electromagnetic wave, which absorbs energy from the photons. This angle is named the refractive angle, at which there is minimum reflected light. The immobilization of antigen generates a specific refractive angle, which is modified upon binding of an analyte. This change of angle is measured, and corresponds to the SPR refractive signal used to characterize the biomolecular interactions (98, 104).

SPR is a micro-fluidic technique in which an analyte is injected into a flow cell and passed through a sensor chip immobilized with an antigen (**Figure 13B**). Before injection, the refractive signal is minimal and correspond to the buffer signal (which is removed by blanking against an empty reference cell). Upon injection, analytes bind the immobilized antigen, increasing the immobilized mass and thus the refractive signal. A maximum is eventually reached, at which analyte association and dissociation are at equilibrium. At the end of the injection, analytes gradually dissociate leading to decreasing refractive signals. The association phase is characterized by rate of association: the association constant k_a (in $M^{-1}.s^{-1}$). The dissociation phase is characterized by a rate of dissociation: the dissociation constant k_d (in s^{-1}). Binding kinetics are defined as $K_D = k_d/k_a$ (in M).

Figure 13: Principle of Surface Plasmon Resonance (SPR).

A. Schematic representation of interaction characterization by SPR. As injection starts, analytes are passed through a blank flow cell and then by a flow cell contained the antigen immobilized on the sensor chip. On the opposite face of the sensor chip a thin layer of gold is excited by polarized light through a prism, and the reflected light is captured on a detector, following a specific refractive angle. Upon binding, the interaction between analyte and antigen creates a transient increase in the immobilized mass, leading to a change of refractive angle. Binding reaches eventually saturation (at sufficiently high concentration of analyte), giving maximum resonance signal. At the end of the injection, analytes dissociate gradually from the antigen, leading to a return of the refractive signal to its original baseline.

B. Typical sensograms obtained for association, saturation and dissociation (after blanking). Upon injection, analytes start to interact with the immobilized antigen, leading to an increase in refractive signal at a constant rate named association constant (k_a). Saturation is obtained during injection when the rate of association reaches equilibrium with the rate of dissociation of molecules (k_d). At the end of the injection, analytes dissociate gradually at a constant dissociation rate k_d .

8.3. Biolayer interferometry

Biolayer interferometry (**Figure 14**) is a relatively new technique compared to ELISA or SPR, which is gaining popularity in recent years. It is an optical analytical method that measures the interference pattern of two surfaces: a layer of immobilized protein sample, and an internal reference layer. White light is shone through glass optical fiber. Part of the light is reflected at the tip of the fiber, travels back through the fiber, and is detected at the other end of the optical path. Changes in the thickness of the glass surface at the tip of the fiber produce changes in the optical path of the reflected light, leading to a change of phase of the reflected light (i.e. a wavelength shift). The antigen is immobilized on a sensor tip and connected to the fiber-optic light source. The tip is dipped into a solution containing the analyte, typically with agitation. The binding of the analyte to the antigen increases the thickness of the surface, which is characterized by a wavelength shift of the reflected light that can be measured through a detector (at the other end of the optical path). Dissociation is measured by dipping the tip in buffer without analyte, and detecting reduction of surface thickness over time (99).

The sensograms obtained by BLI are very similar to the ones obtained by SPR (**Figure 13B**), but binding signal is described by an increase in biolayer thickness (in nm) and not by refractive signal.

Figure 14: Principle of BioLayer Interferometry (BLI).

Image extracted from (99). Refer to previous figure for typical binding kinetic profiles.

9. Recombinant protein production and purification

Proteins are the core of biomedical and biological sciences. Their study and characterization usually requires milligram quantities of relatively pure protein samples, which are typically obtained by recombinant protein expression. For all the characterizations performed on pure proteins in this study, proteins were produced in *E. coli* and purified in our laboratory. In this section, the procedures performed are explained.

The Structural Genomics initiative, a worldwide effort to produce, purify and characterize tens of thousands of different proteins from Eubacteria, Archea and Eukarya, has helped define a common high-throughput strategy for the production of recombinant proteins (105). Our protocols are in accordance with those guidelines.

9.1. The expression host

Escherichia coli is the most common expression host for the production of recombinant proteins. It is inexpensive, robust, easy to manipulate, production time is short and recombinant expression yields can be as high as 20% of total cellular protein content (105, 106).

The molecular toolbox available for this organism is extensive. Among those tools, there is a great selection of *E. coli* strains that have been developed, and can be used for tailored applications. BL21(DE3), for example, allows expression through bacteriophage T7 RNA polymerase and is deficient in *lon* and *ompT* proteases (105). Rosetta-2 strain is a BL21 (DE3) derivative carrying additional rare codons for the expression of eukaryotic proteins (107). Origami strains are particularly suited for the formation of disulfide bonds in the cytoplasm (107). C41 is a BL21-variant well-suited for the expression of toxic proteins and membrane proteins (108). All these strains can be tested to identify the most suitable for the expression of the protein of interest.

9.2. DNA template and construct optimization

Another important parameter to consider, when producing a protein, is the open-reading frame (ORF) used as template, as well as its boundaries. A protein of interest can have different isoforms or splice variants, or the expressed polypeptide could undergo post-translational modifications such as signal peptide removal. The DNA can be obtained from its natural source, or produced by gene synthesis technology. If the latter option is pursued, the codons have to be specified, so one can choose to follow the original DNA sequence, or to modify it to follow *E. coli*'s most common codons. All these parameters have to be taken into account when selecting the best template ORF to be cloned into an expression vector (105).

Small truncations of constructs in the N- and C-termini, which might have minimal influence on activity, can have dramatic effects in expression and solubility. About 50% full-length proteins from prokaryotes, and ~10% full-length proteins from eukaryotes can be produced in soluble form in *E. coli* (109). However, Gräslund S. *et al.* demonstrated that optimizing the protein N- and C-termini could double the success rate. This is due to the fact that the expression levels and degree of solubility of recombinant proteins can vary significantly even by slight modifications in the protein construct boundaries (110).

The production of a specific domain of a protein can be sufficient for many applications, as single domains can usually perform their functions independently. As for the full-length protein, the boundaries of the domain can influence the expression levels and solubility of the construct. Therefore, generating several constructs for each domain of interest can increase the chances of successful expression (105). There are numerous domain-boundary prediction servers based in sequence homology, e.g. Foldpro, Robetta and Maopus (111). However, their accuracy is still limited.

If we take all these different parameters into account, the number of constructs to be tested (to increase the chances of successful expression of a protein) can be very large. Therefore, high-throughput techniques allowing rapid cloning and expression tests in small-scale are essential.

9.3. Cloning of targets

As seen in the previous section, for every protein target, a great number of constructs might be tested to identify the ones that are producing higher expression yields and better purification profiles. A generic, high-throughput cloning method is necessary for this purpose.

The most commonly employed expression vectors, such as the pET vectors, use the T7 RNA polymerase to drive protein production, under the control of a lac operator. These vectors are designed to express proteins in strains such as BL21(DE3), which contain a genomic copy of the T7 RNA polymerase under the control of a lac repressor. Under normal growth conditions, the T7 RNA polymerase is not produced and therefore there are negligible expression levels of the target protein. However, upon induction the polymerase is produced, leading to a strong expression of the recombinant protein (107).

Purification tags allow rapid and specific purification of the target protein. There are many tags available for this purpose, but the poly-histidine tag is the most widely employed. This tag can be fused to the 5' or 3' end the construct, and allows the purification of the target protein by Immobilized Metal Affinity Chromatography (IMAC). A protease cleavage site, such as the tobacco etch virus (TEV) site, can be added between the target protein and the tag, allowing the removal of the tag if needed (105). Other affinity tags, such as Strep-II tag are also available, as well as solubility tags such as Maltose Binding Protein (MBP) or Glutathione S Transferase (GST) (107).

When performing cloning in a high-throughput manner, the classical ligation-dependent cloning is not always convenient. This method relies on restriction enzyme digestion followed by ligation of the insert and the vector. The restriction sites to be used can sometimes be present in the construct DNA, which requires their removal before the cloning can be performed. Ligation-independent methods are a popular alternative to circumvent this problem. Two methods are usually employed: recombination-based technologies such as the Gateway system, or ligation-independent cloning (LIC) (105, 112).

Ligation-independent cloning (**Figure 15**) (113) is a non-commercial method based on the annealing of the vector and insert through ~15- nucleotide long complementary sequences on both ends of the insert and vector. The vector is linearized by enzyme digestion, while the insert is amplified with primers containing ~15-nucleotide long overhang sequences composed of only three nucleotides, sequences that are also present in the vector's cloning site. As an example, we could imagine an overhang sequence only composed of A, G, and T, and the complementary sequence being composed of only T, C and A.

Both vector and insert are treated with T4 DNA polymerase, in the presence of a single deoxynucleotide. In our example, the insert would be treated in the presence of C, while the vector in the presence of G. The polymerase has a 3' to 5' exonuclease activity, and therefore digests the DNA in the 3' to 5' sense until it finds a base complementary to the deoxynucleotide present in the reaction. The base absent in the overhang sequence is present in the vector, flanking the sequences to digest, and the polymerase stops digesting at these sites. This generates the ~15-nucleotide long overhangs in both insert and vector, which are complementary. The T4-treated vector and insert are mixed together, and allowed to anneal, generating a circular vector that is ligated by the cellular machinery upon transformation into bacteria.

Figure 15: Schematic representation of Ligation Independent Cloning (LIC).

9.4. Production and purification

Given the high number of constructs to test for expression, the first step in protein production is to identify the constructs that produce the highest yields of protein and maintain solubility throughout the purification. To do so, high-throughput miniaturized methods have been developed, which allow for the production and purification of hundreds of protein constructs using 1-20 mL of culture for expression, and 96-well plates for small-scale IMAC purifications. A small amount of IMAC agarose beads (e.g. ~25 μ L per well) is distributed in a 96-well filter plate, equilibrated with buffer, and incubated with the small-scale expression soluble fraction. The undesired proteins and contaminants are eliminated by successive washes, in the presence of low concentrations of imidazole (an organic compound with a structure similar to the Histidine side-chain ring), followed by centrifugation. The His-tagged proteins are eluted in the presence of high concentration of imidazole, by centrifugation (105, 110). Several growth conditions can be tested in parallel, such as expression strains, IPTG concentrations, expression temperature, protein production time, and media, and which can have a significant effect in yields and solubility of the constructs tested. SDS-PAGE analysis of the purified samples is performed, and the most promising constructs, strains and culture conditions are selected to produce and purify the protein in large-scale (114).

Large-scale expression is usually performed in 1-3L cultures, using baffled shake flasks and rich media (ideal to obtain high cell densities). Cultures are grown at 37°C until mid-to-late log phase, then protein expression is performed overnight at 15-25°C with IPTG induction. The incubation at 37°C allows rapid cell growth, reaching high cell density. Protein expression at lower temperature is optimal for high yields in soluble protein, presumably because the slower protein production rates allow more time for the proteins to fold (105).

After protein expression, constructs are usually purified using a two-step purification strategy: IMAC purification followed by size-exclusion chromatography (SEC). IMAC combines high protein binding capacity with relatively high specificity, and elution with imidazole is mild. SEC separates proteins according to their size, which allows for the isolation of specific

populations of the target protein. This is particularly useful to eliminate aggregates of the target, which might be detrimental for many applications, as well as impurities present after IMAC. These procedures can be automated using chromatography systems such as the ÄKTA systems (GE Healthcare). TEV cleavage followed by IMAC subtraction (i.e. the binding of impurities onto the IMAC resin while the poly-histidine-free protein of interest is eluted), and ion exchange chromatography can be used as supplementary “polishing” steps for increased purity (105).

10. Projects, aim and strategy

The aim of this project was to develop simple, robust and cost-effective screening methodologies to identify stabilized proteins. Two complementary paths were taken: the stabilization of proteins by mutagenesis, and their stabilization by an interaction partner.

10.1. Protein stabilization by mutagenesis

As described throughout the introduction, one popular strategy to stabilize a protein is through mutagenesis, particularly using directed evolution approaches. We decided to explore the potential development of a generic, high-throughput and robust method to generate more stable variants of proteins, regardless of their function or origin. This method was based on the Colony Filtration blot (CoFi blot) developed by Cornvik T., Nordlund P. and colleagues (**Figure 16**), which was originally developed to identify proteins or protein constructs that express at high level and in a soluble form in *E. coli* (115-117).

Figure 16: Schematic representation of the Colony Filtration blot (CoFi).

A library of plasmids coding for different proteins (red dots), constructs or variants of a protein are transformed into bacteria and plated on agar. Isolated colonies are transferred to a Durapore 0.45µm filter membrane, and protein expression is induced. Colonies are lysed by placing the filter membrane on top of a lysis sandwich. The latter is composed of a nitrocellulose membrane and a Whatman paper soaked in lysis buffer. Upon cell lysis, soluble proteins are able to migrate through the filter membrane, and bind the nitrocellulose membrane. Insoluble proteins aggregate and are not able to diffuse through the filter. Soluble variants can be identified by western-blot, through a specific tag.

In the Colony Filtration method, plasmid libraries coding different proteins, protein variants or protein constructs are transformed into *E. coli* (or other microorganisms) and plated on agar plates. This “Master” plate is duplicated by carefully placing a 0.45 µm filter membrane on top of the colonies, and then peeling it from the colonies. The membrane, containing a copy of the master plate, is placed on top of an agar plate, colonies facing up, and protein expression is induced; the Master plate conserves part of the bacterial colonies, and they can be re-grown.

After induction, the filter membrane is transferred on top of a lysis sandwich. The filter sandwich is composed of a nitrocellulose membrane on top of a Whatman paper, the latter soaked in lysis buffer. Upon cell lysis, the filter membrane allows the diffusion of soluble proteins, while aggregated proteins are retained. The soluble proteins diffuse through the filter and bind to the nitrocellulose membrane, and can be detected through a specific tag. If the proteins are His-tagged, for example, they can be detected using an anti-His tag antibody, conjugated with horseradish peroxidase. The clones in the screened plate expressing higher levels of soluble protein show a strong signal after development. By superimposing the developed blot to the Master plate, the colonies which express a protein variant in high levels can be retrieved.

Thermal stress of a purified protein above a certain critical temperature, usually leads to irreversible protein unfolding, followed by rapid aggregation (77). We hypothesized that a similar scenario might be in effect for recombinant proteins over-expressed in *E. coli*. When cells are subjected to increased temperatures the proteins might unfold and aggregate in a similar way as purified proteins. If that was the case, the CoFi blot could be adapted into a broadly applicable technique for the determination of proteins aggregation temperature in crude samples, such as *E. coli* cells. We named the method Hot-CoFi blot (**Figure 17**).

Another promising application would be the adaptation of the CoFi blot into a simple, generic, fast and reliable activity-independent screening technique for the isolation of thermostable protein variants. For this purpose, methods to easily generate high-quality libraries of randomized proteins were implemented,

and the screening method was optimized to isolate improved variants in a reliable manner.

Figure 17: Schematic representation of the Hot Colony Filtration blot (Hot-CoFi).

The procedure is similar to the CoFi blot, with the following modifications: the induced colonies are incubated at high temperature (arbitrary selected according to the desired stringency), before performing the lysis step at the same elevated temperature. Unstable proteins aggregate and are not able to diffuse through the filter. Stable proteins remain soluble, diffuse through the filter, and can be identified by western-blot, through a specific tag.

10.2. Protein stabilization by interaction with a protein or peptide partner

As a complementary strategy for protein stabilization, we decided to explore the potential of stabilizing proteins through the interaction with *de novo* generated partners. Two strategies were followed: the generation of binders by phage display (**Figure 9**), and the generation of binders directly with a new method derived from the Hot CoFi, named interaction-CoFi (**Figure 18**).

For the phage-display based approach, the 10th fibronectin type III domain (FN3) was chosen as a scaffold. This scaffold has been extensively studied as an alternative scaffold to antibodies and antibody domains, as it shares the antibody-like fold with the advantage of lacking disulfide bridges, it has high production yields in *E. coli*, high stability and toleration to mutations in the loops

(118-120). For the introduction of diversity, two loops were randomized by Kunkel mutagenesis, aiming for a bias in amino-acid composition similar to that encountered in antibodies (50).

For the interaction-CoFi (iCoFi) (**Figure 18**), we reasoned that the interaction of a target protein with a stable protein partner would lead to an increase of the target's apparent temperature of cellular aggregation (T_{cagg}). Such stabilization could be detected by Hot-CoFi blot. As a proof-of-concept, known interaction partners were co-expressed, and the T_{cagg} was measured by Hot-CoFi against the T_{cagg} of the target alone. As a more elaborate proof-of-concept, we aimed at validating iCoFi as an effective alternative to phage display for the screening and selection of stabilizing binders generated *de novo*. For such purposes, a minimalistic library of antibody fragments was generated and screened for stabilization of selected targets.

Figure 18: Schematic representation of the interaction Colony Filtration blot (iCoFi)

A plasmid coding for the protein of interest (red dot) is co-transformed with a library of co-expression plasmids coding for potential partners (blue dot) into bacteria and plated on agar. The selection procedure is identical as for the Hot-CoFi. Colonies containing protein partners that do not interact do not lead to the stabilization of the target, causing its aggregation. If the partners do interact, the complex leads to the stabilization of the target, leading to a partial or full population of soluble proteins. Stabilized targets diffuse through the filter, and can be identified by western-blot, through a specific tag.

MATERIALS AND METHODS

1. Materials

All materials were purchased from the following companies:

- Buffers and salts: Sigma-Aldrich (St. Louis, MO, USA), Merck-Millipore (Billerica, MA, USA) and Affymetix (Sta. Clara, CA, USA).
- Enzymes: New England Biolabs (NEB, Ipswich, MA, USA), except for KOD HotStart and KOD Xtreme polymerases (Merck-Millipore, Billerica, MA, USA).
- Random mutagenesis: Genemorph II Mutagenesis Kit from Agilent Technologies (Santa Clara, CA, USA)
- Membranes: Durapore membranes were from Merck-Millipore (Billerica, MA, USA), Nitrocellulose membranes from Thermo-Fischer Scientific (Waltham, MA, USA), Whatman paper from Whatman (Maidstone, Kent, UK).
- Protease inhibitors: Inhibitor Cocktail Set III, EDTA-Free, from Merck-Millipore (Billerica, MA, USA).
- Anti-His probe: Thermo-Fischer Scientific (Waltham, MA, USA).
- Super Signal West Dura chemiluminescence kit: Thermo-Fischer Scientific (Waltham, MA, USA).
- DNA, ssDNA, PCR and gel purification kits were from Qiagen (Venlo, Netherlands).
- Plates: Petri dishes and 96-well plates were from NUNC, Thermo-Fischer Scientific (Waltham, MA, USA); 3-well Intelli-plate96 from Art Robbins Instruments (Sunnyvale, CA, USA).
- CCD camera: Fujifilm (Fujifilm LAS-4,000, Tokyo, Japan).
- LEX bioreactor system: Harbinger Biotechnology & Engineering (Toronto, ON, Canada).
- DLS detection system: StarGazer, from Harbinger Biotechnology & Engineering (Toronto, ON, Canada).

- AKTA Xpress protein purification system: GE Healthcare (Little Chalfont, United Kingdom).
- Surface Plasmon Resonance T-3,000 system and CM5 chips: GE Healthcare (Little Chalfont, United Kingdom).
- Biolayer Interferometry system BLItz from ForteBio (Menlo Park, CA, USA)
- Real-Time PCR detection system: iCycler iQ, from Bio-Rad (Hercules, CA, USA).
- Sypro-orange: from Bio-Rad (Hercules, CA, USA).
- Crystallization solutions: JCSG+ Suite screen from Qiagen (Venlo, Netherlands); Crystal Screen HT and Crystal Screen Cryo HT from Hampton Research (Aliso Viejo, CA, USA).
- *E. coli* strains: Rosetta2 from Novagen (GE Healthcare, Little Chalfont, United Kingdom), TG1 from Lucigen (Middleton, WI, USA), CJ236 from NEB (Ipswich, MA, USA).
- Biotinylation kits from Thermo Fischer (Waltham, MA, USA).

2. Methods

2.1. Cloning of the target genes

The UL14, UL18, ORF49 and ORF69 genes were kindly provided by Dr. Jurgen Haas (University of Edinburgh), anti-Grb2 scFv gene was kindly provided by Dr. Sachdev S. Sidhu (University of Toronto), and a solubility-improved variant of TEV protease (121) by Dr. Helena Berglund (Karolinska Institutet). The genes coding for IL1RA (isoform 2), 4D5 VH domain, CH2 domain, TEV-SUMO-TEV, APH, anti-APH DARPIn, NPDZ, SAM-PBM, and anti-MBP DARPIn, were synthesized (Genscript), while NXR1 was cloned from the human MGC collection and MBP from vector pNBH7C.

All the genes studied were subcloned by Ligation Independent Cloning (LIC), according to described procedures (122), with the exception of TEV protease, which was already provided in a suitable vector. The destination vector varied

according to the target studied; vectors, antibiotic resistances, primers used and variations in T4 treatment are summarized in **Table 4**.

All targets studied for Hot Colony Filtration of soluble proteins were cloned into pNIC-CH3. All co-expression targets studied were cloned into pNIC-CTHF. Co-expression partners and the VH library were cloned into pCDF-LIC-CS or pCDF-LIC-bla.

Table 4: Summary of vectors used in this study

Vector	Antibiotic	LIC (5' overhang)	LIC (3' overhang)	T4 vector	T4 insert	Tag
pNIC-CH3	Kanamycin	5'AGAGGAGATAATT AATG 3'	5'AATGGTGGTGATG ATGGTGCGC 3'	dCTP	dGTP	C-ter His ₆
pNIC-CTHF	Kanamycin	5'TTAAGAAGGAGAT ATACTATG 3'	5'GATTGGAAGTAGA GGTTCTCTGC 3'	dCTP	dGTP	C-ter Flag- His ₆
pCDF-LIC-CS	Streptomycin	5'AAAGAGGAGATAA TTAATG 3'	5'TTGGAAAGTAGAGG TTCTCTGC 3'	dCTP	dGTP	C-ter StrepII
pCDF-LIC-bla	Ampicillin	5'AAAGAGGAGATAA TTAATG 3'	5'TTGGAAAGTAGAGG TTCTCTGC 3'	dCTP	dGTP	C-ter StrepII
pTH24	Ampicillin	-	-	-	-	C-ter His ₆

Vectors were linearized by digestion with *BfuAI* (NEB), according to the manufacturers protocol. The genes were amplified by PCR using gene-specific primers that contained universal LIC forward and LIC reverse overhangs. Treatment of the insert and the linearized vector with T4 DNA polymerase, in the presence of a single deoxy-nucleotide, generates 15-nucleotide-long overhangs by the polymerase's 3'-exonuclease activity. T4 polymerase-treated insert and vector were annealed, and used to transform a bacterial cloning strain. The presence of a negative selection marker gene, *SacB* (coding for *B. subtilis* levansucrase, conferring sensitivity to sucrose), allows the selection of the vector containing the desired gene on LB-agar plates in the presence 5% sucrose and a specific antibiotic.

2.2. Determination of Temperature of cellular aggregation (T_{cagg})

The temperature of cellular aggregation (T_{cagg}) was determined following a modified protocol based on the Colony Filtration Blot (116, 117). Rosetta2 cells (Novagen) were transformed with the expression plasmids, harboring the gene

of interest fused with a C-terminal His₆-tag, and the cells were plated. Individual colonies were picked and grown overnight at 37°C in Terrific Broth supplemented with antibiotics (50 µg/mL Kanamycin or Ampicillin, according to the vector, and 34 µg/mL Chloramphenicol). 2 µL-aliquots of culture were plated in as many LB-agar plates+antibiotics as temperatures to screen and incubated at 37°C until a lawn was visible. Colonies were transferred to a Durapore 0.45 µm filter membrane (Merck-Millipore), and placed on LB-agar plates (colonies facing up) supplemented with antibiotics and 0.2 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG), except for the single-chain antibody, VH domain and IL1RA, for which 1 µM, 25 µM and 30 µM IPTG were used, respectively. Induction was performed overnight at room temperature (RT) for protein expression.

After protein production, the induction plates were incubated at temperatures ranging from 20°C to 80°C, with temperature intervals of 10°C, for 30 min. In the meantime, a lysis sandwich was prepared, composed of a Whatman paper, soaked in CoFi lysis buffer [20 mM Tris, pH 8.0, 100 mM NaCl, 0.2 mg/ml lysozyme, 11.2 U/mL Benzonase Endonuclease (Merck-Millipore) and 1:1,000 dilution of Protease Inhibitor Cocktail Set III, EDTA-Free (Merck-Millipore)], and a nitrocellulose membrane (Merck-Millipore). The lysis sandwich was incubated 10 min at the screening temperature. The Durapore membrane was then transferred on top of the lysis sandwich, with the colonies facing up, and incubated at the screening temperature for another 30 min.

Cells lysis was further improved by three freeze-thaw cycles at -80°C and RT, respectively (30 min each). The Durapore membrane and the Whatman paper were discarded, and the nitrocellulose membrane was incubated in blocking buffer [TBS-T buffer (20 mM Tris, pH 7.5, 500 mM NaCl, 0.05% (vol/vol) Tween 20) and 1% Bovine Serum Albumin (BSA)] for 1 h at RT.

After blocking, the nitrocellulose membrane was washed three times in TBS-T, 10 min at RT, with shaking at 70 revolutions per minute (rpm). It was then incubated in TBS-T containing a 1:5,000 dilution of HisProbe-HRP (Thermo Scientific), for 1 h at RT, with shaking at 30 rpm. Three washing steps followed, performed as previously described.

The nitrocellulose membrane was developed using Super Signal West Dura chemiluminescence kit (Thermo Scientific), using a CCD camera (Fujifilm LAS-4,000). Signal intensities were measured using Multi-Gauge software (Fujifilm) over the area corresponding the colonies, and normalized by area (unit: Arbitrary Intensity Units per mm²) .

For T_{cagg} determination from liquid cultures, protein expression was performed in liquid cultures, the cells were pelleted, and resuspended in CoFi lysis buffer. Aliquots of 50-100 µL were dispensed into a 96-well PCR plate. Samples were subjected to a temperature gradient, with temperatures varying according to the desired range to study, for 15 minutes. After heating, samples were cooled to room-temperature, and centrifuged at 3,000 g for 20 min. The supernatants were filtered through a 0.45 µm filter plate, and 2 µL of the flow-through (from each temperature tested) was applied onto a nitrocellulose membrane. The development was performed as previously. As the dots are usually of even diameter, the signal intensities were measured without normalization, as average intensities with imaging software ImageJ (123).

2.3. Error-prone library generation

Error-prone PCR was performed using Genemorph II Mutagenesis Kit (Agilent Technologies), using the cloned genes as template, according to the manufacturer's protocol. An average of 50 ng of insert template was used, in a 30 cycles PCR reaction with phosphorylated primers, leading to an average of ~2-3 mutations per gene in the final plasmid library. The error-prone PCR product was gel purified (Qiagen Gel purification Kit).

The plasmid library was created by Megaprimer PCR of Whole Plasmid (MEGAWHOP) reaction (124), using an optimized protocol described here. Typically, 1-10 ng of plasmid containing the wild-type (WT) gene was used as template, and amplified using ~1 ng of insert per bp of insert length of purified epPCR insert as megaprimer (e.g. 500 ng for a 500 bp insert), in a PCR reaction with KOD Xtreme polymerase (Merck -Millipore). The buffer composition was the standard recommended by the manufacturer, with the addition of 1 mM

NAD⁺ and 40 U of Taq DNA ligase, and using 0.5 U of KOD Xtreme polymerase. The PCR conditions were as follows: 94°C for 2 min, (98°C for 10 sec, 65°C 30 sec, 68°C for 6 min) for 30 cycles, 4°C on hold. After PCR, 20 U of *DpnI* (NEB) was added to the PCR reaction, and incubated for 3 h at 37°C.

DH10B cells were electroporated with the *DpnI*-treated MEGAWHOP product yielding libraries of $\sim 10^5$ unique members.

2.4. Library screening

Rosetta2 cells were transformed with the mutagenesis libraries, and plated on 24.5 cm diameter square LB-agar plates supplemented with 50 µg/mL Kanamycin (Ampicillin, for TEV protease) and 34 µg/mL Chloramphenicol. These are considered master plates.

They were screened using the same protocol as described for T_{cagg} determination, but at a single temperature, typically 5-10°C above the purified protein's melting temperature (T_m) or aggregation temperature (T_{agg}) for 30 min. VH domain was an exception, as it was screened at 80°C for round 1 (16°C above T_{agg}) and 90°C for round 2, the induction plate being incubated for 1 h at that temperature.

Positives colonies were selected according to signal intensity, and picked from the master plate. To avoid cross-contaminations or false positives, each picked colony was re-streaked onto agar plates, and four new colonies picked from the re-streak. They were verified by re-growing them as 2 µL bacterial lawns, and repeating the same screening procedure, and including the wild-type clone as a control. Positive mutants, confirmed from the verification step, were produced in large-scale for biophysical characterization.

2.5. Protein production and purification

Selected positive clones were produced in large-scale cultures using 0.75 L or 1.5 L of TB + antibiotics, and incubated at 37°C using a LEX bioreactor system (Harbinger Biotechnology & Engineering). When OD_{600nm} reached 2, the culture

was cooled to 18°C for 30 min, protein production was induced with 0.5 mM IPTG (except for IL1RA, VH domain and scFv, induced with 25 μ M, 30 μ M and 1 μ M IPTG, respectively), and incubated overnight at 18°C.

For long-term stability experiments performed for IL1RA at high concentration, protein production was performed in a larger scale. For WT, 12 L of TB + antibiotics were used, while 6 L were used for each mutant.

Cultures were harvested by centrifugation at 4,000 g, 25 min at 4°C. The pellets were weighed, and solubilized in 1.5 volumes:weight of pellet lysis buffer [100 mM HEPES, pH 8.0, 500 mM NaCl, 10 mM Imidazole, 10% Glycerol, 1:1,000 dilution of Protease Inhibitor Cocktail Set III, EDTA-Free (Merck-Millipore)]. Cells were lysed by performing one cycle of freeze-thawing at -80°C and RT, respectively, followed by sonication. The lysed cells were centrifuged at 46,000 g, for 25 min at 4°C, the supernatant collected and filtered through a 1.2 μ m filter syringe.

The filtrate was then purified by Immobilized Metal Ion Affinity Chromatography (IMAC), followed by Size Exclusion Chromatography (SEC). The IMAC column (Ni-NTA agarose, Qiagen) was washed with 20 column volumes of IMAC wash 1 buffer [20 mM HEPES, pH 7.5, 500 mM NaCl, 10 mM Imidazole, 10% Glycerol], followed by 20 volumes of IMAC wash 2 buffer [20 mM HEPES, pH 7.5, 500 mM NaCl, 25 mM Imidazole, 10% Glycerol]. The desired protein was eluted with 5 column volumes of IMAC elution buffer [20 mM HEPES, pH 7.5, 500 mM NaCl, 500 mM Imidazole, 10% Glycerol]. The IMAC eluate was loaded onto a Superdex-75 or Superdex-200 SEC column (GE Healthcare), according to protein size, for further purification, and eluted in SEC buffer [20 mM HEPES, pH 7.5, 300 mM NaCl, 10% Glycerol]. TCEP was added to all the purification buffers to a final concentration of 0.5 mM, except for proteins containing disulfide bonds.

For the stability improvement of the VH-domain, IMAC purification was performed as described, but further purified using a 1 mL HiTrap rProtein A FF column (GE Healthcare). To do so, the Protein A column was equilibrated with 10 mL of SEC buffer, the IMAC eluate was loaded. The column was washed with 10 mL of SEC buffer. The protein was eluted with 0.1 M glycine, pH3, and neutralized by 1:5 v:v of neutralization buffer [1 M Tris, pH8, 1.5 M NaCl, 10% glycerol].

For the characterization of the VH-domain clones isolated through interaction CoFi, periplasmic extraction was performed, followed by purification through protein A. For periplasmic extraction the pellet was solubilized in 1 volume ice-cold TES buffer [0.2 M Tris pH 8.0, 0.5 mM EDTA, 0.5 M sucrose, 1:1,000 dilution of Protease Inhibitor Cocktail Set III, EDTA-Free (Merck-Millipore)] for 30 min at 4°C, after which 1 volume of ice-cold milliQ water was added and mixed for another 30 min. The solution was centrifuged at 16,000 g for 15 min at 4°C. The supernatant was loaded onto a 1 mL Protein A column and purified as previously.

For long-term stability experiments performed for IL1RA, purification followed the standard IMAC and SEC purification procedures, but for SEC the buffer was changed to CSE buffer [10 mM sodium citrate, 140 mM NaCl, 0.5 mM EDTA, pH 6.5].

2.6. Biophysical characterization

The denaturation temperature of the purified protein variants was determined by Differential Scanning Fluorimetry, except for UL18 and VH-domain for which Differential Static Light-Scattering was used instead, according to published methods (42, 43).

To perform Differential Scanning Fluorimetry, 5 µg of the proteins were diluted into a 25 µL SEC buffer solution containing 5x SYPRO Orange fluorescent dye (Bio-Rad). Proteins were tested in triplicates, the fluorescence was monitored using a 96-well Real-Time PCR detection system (iCycler iQ, from Bio-Rad), from 25°C to 95°C, with a gradual temperature increase of 1°C every 10 sec.

For studies of protein stabilization by interaction with a partner, target protein and partner were mixed in different ratios (optimal ratios varied for each target, and are specified in the results section). Melting temperatures were determined by DSF as described.

For Differential Static Light-Scattering experiments, 5 µg of protein was diluted into 50 µL of SEC buffer (for VH domain, 1 mg/ml of protein was used instead to obtain aggregation curves), and dispensed in triplicates in 384-well optical

bottom plates with black upper structure (Nunc). The samples were sealed with 50 μ L of mineral oil (Sigma-Aldrich), and centrifuged at 15°C, 3 min at 3,200 g. Static Light-Scattering was monitored using a DLS detection system (StarGazer, from Harbinger Biotechnology & Engineering), in a temperature range of 25°C to 80°C, with a gradual increase in temperature of 1°C /min.

The melting temperatures and aggregation temperatures were determined using the Boltzmann equation, by the instrument's software.

2.7. Crystallization and structural determination

NXR1 WT and mutants were used in crystallization trials at a concentration of 10 mg/mL in SEC buffer, by sitting-drop vapor diffusion on a 3-well Intelli-plate96 (Art Robbins Instruments) using the JCSG+ Suite screen (Qiagen), and incubated at room-temperature. Mutants 3 and 5 gave crystals in 0.1 M tri-sodium citrate, pH5 and 3.2 M ammonium sulfate. Mutant 5 crystals were mounted and frozen. Diffraction data was collected at the National Synchrotron Radiation Research Center (Taiwan) and the structure was solved and refined using CCP4 suite (125) and Phenix (126), by molecular replacement using PDB ID 3GG6 as a template.

For VH-36, crystallization trials were performed at a concentration of 20 mg/mL in PBS, by sitting-drop vapor diffusion on a 3-well Intelli-plate96 (Art Robbins Instruments) using the Crystal Screen HT Suite (Hampton Research). Crystals were obtained in wells C4 (condition A: 0.2 M ammonium sulfate, 30% PEG 8,000) and C6 (condition B: 0.2 M sodium acetate trihydrate, 0.1 M sodium cacodylate trihydrate pH 6.5, 30% PEG 8,000), 15% Glycerol was used as cryoprotectant. Diffraction data was collected at the Australian Synchrotron (Australia) and the structure was solved and refined using CCP4 suite (125) and Phenix (126), by molecular replacement using PDB ID 3B9V as a template.

2.8. Surface Plasmon Resonance

Surface Plasmon Resonance was performed using a Biacore T-3,000 system (GE Healthcare) and CM5 amine-coupling chips (GE Healthcare), following the manufacturer's recommendations.

For IL1RA, WT and mutants were immobilized at ~400 response units (RU). IL1 type-I receptor (R&D Systems) was injected in SPR buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3.4 mM EDTA, 0.005% surfactant P-20) at a flow of 30 μ L/min for 6 min, followed by a dissociation period of 1 h. Analyte concentration varied from 10 nM to 312.5 pM in 1:2 dilutions. Remaining bound protein was dissociated with 15 μ L of 20 mM HCl.

For anti-Grb2 ScFv, Grb2-SH2 domain was immobilized at ~80 RU. ScFv WT and mutants were injected in SPR buffer at a flow of 30 μ L/min for 2 min, followed by a dissociation period of 10 min. Analyte concentration varied from 100 nM to 3.125 nM (except for scfv-1, for which concentrations varied from 200 nM to 6.25 nM) in 1:2 dilutions. Remaining bound protein was dissociated with 15 μ L of 20 mM HCl.

For anti-Xylanase FN3 clones, Xylanase was immobilized at ~3,000 RU. Anti-Xylanase FN3 clones FN.Xyl.H2 and FN.Xyl.H4 were injected in SPR buffer at a flow of 30 μ L/min for 4 min, followed by a dissociation period of 7 min. Analyte concentration varied from 100 μ M to 6.25 μ M in 1:2 dilutions.

For anti-MBP FN3 clone, MBP was immobilized at ~8,000 RU. Anti-MBP FN3 clone FN.MBP.F1 were injected in SPR buffer at a flow of 30 μ L/min for 4 min, followed by a dissociation period of 7 min. Analyte concentration varied from 76 μ M to 19 μ M in 1:2 dilutions.

2.9. Biolayer Interferometry

Biolayer Interferometry (BLI) was performed on a BLItz system (ForteBio), using streptavidin coated sensors (ForteBio), and antigens biotinylated *in vitro* using EZ-Link NHS-PEG4-Biotin biotinylation kit (Thermo Scientific), according

to the manufacturers recommendations. All kinetic analysis were performed in kinetics buffer [1X PBS, 0.01% BSA, 0.02% Tween 20, 0.005% NaN₃].

For all antigens, immobilization was performed for 2 min using 4 µL of biotinylated antigen at a concentration of 50 µg/mL. Binding was performed in the 4 µL drop-holder for 2 min, while dissociation was done in 250 µL kinetics buffer (tube holder), with agitation at 2,200 rpm. Concentrations of analyte varied from target to target, and are specified in the results section. Sensors were regenerated by exposure to 10 mM HCl for 30 sec, followed by 1 min incubation in kinetics buffer.

2.10. Long-term stability study for IL1RA

Long-term stability studies were performed for IL1RA WT and mutants in a similar manner as previously described (34). IL1RA was concentrated to 50 mg/mL in CSE buffer. For experiments at 30°C, 10 µL of protein sample was used, and 20 µL for 40°C. The protein sample was dispensed in a 384-well black flat transparent bottom plate (NUNC), and covered with 50 µL of mineral oil. The protein-containing wells were surrounded by buffer blanks dispensed in identical manner. Protein aggregation was measured for seven days at 405 nm (using a Tecan Infinite M200 plate reader). Measurements were performed every minute for the first 5 h, and every 30 min for the rest of the experiment. Solutions were homogenized by performing 3 sec shaking before every measurement.

2.11. TEV protease activity

To assess the activity of the TEV protease WT and mutants, a substrate was designed: a fusion protein containing a N-terminal His₆-tag, and two SUMO subunits linked by a flexible (GGGS)₂-ENLYFQ-(GGGS)₂ linker (SUMO-TEV-SUMO). 35 µg of substrate were mixed with 0.5 µg of TEV protease in a 25 µL reaction. The reaction was performed at RT, and 3 µL aliquots were removed at different time points. Substrate digestion was monitored by SDS-PAGE.

To determine activity after a heat challenge, the TEV protease sample was heated at 60°C for 20 min before performing the activity assay as described.

2.12.DHAA activity

To test the activity of DHAA WT and mutants against 1,2,3-trichloropropane (TCP), a 96-well format assay was established, based on previously described protocols (127). DHAA WT and mutants were added to 200 μ L DHAA assay solution [1 mM BES (N,N -bis(2-hydroxyethyl)-2-aminoethanesulfonic acid), 150 mM NaCl, 0.05 mM BTB (bromothymol blue), 10 mM TCP, pH7.8], at a final concentration of 0.15 mg/mL, and the change of colour was measured at 620nm using a Tecan Infinite M200 plate reader at room-temperature. Selected mutants were further characterized by measuring their activity at 40°C, under the same conditions.

Residual activity measurements after thermal challenge were performed by submitting an enzyme solution at a concentration of 3 mg/mL in DHAA dilution buffer (1 mM BES, 150 mM NaCl, pH 7.8), to a 15 min thermal stress at 35°C, 45°C, 50°C and 55°C, followed by rapid cooling to RT. The enzyme residual activity is measured as previously.

To assess the influence of protein concentration on activity and stability, enzyme concentrations were varied between 0.15 mg/mL, 0.75 mg/mL and 1.5 mg/mL. The DHAA assay solution was modified by increasing BES concentration to 10 mM, and activity was measured at 40°C.

2.13.Generation of a phage display library

A fusion gene was generated by gene synthesis. It was composed, from 5' to 3', of DsbA signal sequence, the 10th Fibronectin type III domain, a hexa-histidine tag, amber stop codon, and the C-terminal domain of gene g3p (coding for residues 249-406) from M13 bacteriophage, in accordance to previously described constructs (118, 120). The fusion gene was cloned into pPCR-Script II SK (+) phagemid vector, in frame with the start codon of the lac promoter originally

regulating β -galactosidase expression. The newly created vector was named pIA003.

CJ236 cells were transformed with pIA003 vector, and uracil-containing single-stranded DNA (ssDNA) was generated by infection with M13K07 helper-phage in the presence of 0.25 μ g/mL uridine and purified according to previously described procedures (128, 129).

This ssDNA template was mutated by Kunkel mutagenesis to remove loops BC and FG and replace them with a stop codon and a PmlI restriction site, according to previously described procedures (128, 129). In brief, primers are phosphorylated using T4 polynucleotide kinase (NEB); phosphorylated primers are annealed to the ssDNA template by PCR (90°C for 3 min, 50°C for 5 min, and 4°C on hold); finally, the complementary DNA strand is extended by T7 DNA polymerase (NEB) and ligated simultaneously by T4 DNA ligase (NEB) for 3h at room-temperature.

Both mutations were performed simultaneously, using the following primers:

BG-stop 5'- CACCGTAGGTGATACGGTAGTATTACACGTGCGGCGCTCCCAAG -3',

FG-stop 5'- CGGTAGTTGATAGAGATCGGTTTAGATTACACGTGGGTAACCGCGTAAACGGTG -3'.

The mutagenesis was used to transform Mach1 chemically-competent cells and plated, plasmids isolated and sequenced to confirm the introduction of the desired mutations. The new vector was named pIA003-stop.

Template pIA003-stop was used to generate the phage display libraries. The BC loop length was varied between six and eight residues, while the FG loop length was varied between seven and nineteen residues. To do so, thirty-nine independent libraries were created (using the same Kunkel mutagenesis procedures) by combining the following primers targeting either the BC or the FG loop:

BC-6 5'- CACCGTAGGTGATACGGTAGTA [FED]_{x6} CGGCGCGTCCCAAGAG -3'

BC-7 5'- CACCGTAGGTGATACGGTAGTA [FED]_{x7} CGGCGCGTCCCAAGAG -3'

BC-8 5'- CACCGTAGGTGATACGGTAGTA [FED]_{x8} CGGCGCGTCCCAAGAG -3'

Where [FED] correspond to a special codon composed of a biased proportion of nucleotides. FED = F(C=0.1,A=0.45,G=0.45), E(C=0.25,T=0.45,A=0.15,G=0.15), D(C=0.25,T=0.15,A=0.45,G=0.15)

FG-7	5'- CGGTAGTTGATAGAGATCGGTTTAGA [ZYX] _{x7} GGTAACCGCGTAAACGGTGATG -3'
FG-8	5'- CGGTAGTTGATAGAGATCGGTTTAGA [ZYX] _{x8} GGTAACCGCGTAAACGGTGATG -3'
FG-9	5'- CGGTAGTTGATAGAGATCGGTTTAGA [ZYX] _{x9} GGTAACCGCGTAAACGGTGATG -3'
FG-10	5'- CGGTAGTTGATAGAGATCGGTTTAGA [ZYX] _{x10} GGTAACCGCGTAAACGGTGATG -3'
FG-11	5'- CGGTAGTTGATAGAGATCGGTTTAGA [ZYX] _{x11} GGTAACCGCGTAAACGGTGATG -3'
FG-12	5'- CGGTAGTTGATAGAGATCGGTTTAGA [ZYX] _{x12} GGTAACCGCGTAAACGGTGATG -3'
FG-13	5'- CGGTAGTTGATAGAGATCGGTTTAGA [ZYX] _{x13} GGTAACCGCGTAAACGGTGATG -3'
FG-14	5'- CGGTAGTTGATAGAGATCGGTTTAGA [ZYX] _{x14} GGTAACCGCGTAAACGGTGATG -3'
FG-15	5'- CGGTAGTTGATAGAGATCGGTTTAGA [ZYX] _{x15} GGTAACCGCGTAAACGGTGATG -3'
FG-16	5'- CGGTAGTTGATAGAGATCGGTTTAGA [ZYX] _{x16} GGTAACCGCGTAAACGGTGATG -3'
FG-17	5'- CGGTAGTTGATAGAGATCGGTTTAGA [ZYX] _{x17} GGTAACCGCGTAAACGGTGATG -3'
FG-18	5'- CGGTAGTTGATAGAGATCGGTTTAGA [ZYX] _{x18} GGTAACCGCGTAAACGGTGATG -3'
FG-19	5'- CGGTAGTTGATAGAGATCGGTTTAGA [ZYX] _{x19} GGTAACCGCGTAAACGGTGATG -3'

Where [ZYX] correspond to a special codon composed of a biased proportion of nucleotides. ZYX = Z(C=0.24,G=0.76), Y(C=0.31,T=0.34,A=0.17,G=0.18), X(C=0.38,T=0.19,A=0.26,G=0.17).

The Kunkel mutagenesis reactions were purified using Qiagen Gel Purification Kit, and eluted in 160 µL. Each purified reaction was used to transform 2x 200 µL of TG1 electro-competent cells. After 1 hour recovery in 5 mL SOC media at 37°C, 10¹⁰ cfu/mL M13K07 helper phage was added to each culture and incubated for one additional hour. The cultures were transferred to 250 mL 2YT media supplemented with 50 µg/mL Carbenicillin and 50 µg/mL Kanamycin, and incubated for 6h at 37°C, followed by 20h at 30°C, with agitation.

Phage libraries were purified individually by 2 rounds of precipitation with 1:5 volumes of PEG/NaCl buffer [20% PEG 8,000, 2.5M NaCl] and resuspension in PBS [10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 137 mM NaCl, 2.7 mM KCl, pH 7.4] following published methods (128, 129). The final yield for each library was approximately 20mL of phage at a concentration of ~5.0 ×10¹³ pfu/mL in a 1:1

PBS:Glycerol solution, which was stored at -20°C until further use. For each library, quality was assessed by sequencing 96 individual clones.

2.14. Phage display library panning

All phage libraries were mixed in equal ratio, diluted in 20 volumes of PBS, precipitated with PEG/NaCl buffer and resuspended in PBS.

Initially, solid-phase phage panning was performed on NUNC Maxisorp plates, following procedures described either by Koide S. and co-workers (128), or by Fuh G. and co-workers (86). In these procedures, the target antigen is immobilized on Maxisorp plates by overnight incubation at 4°C, and blocked with 0.5% BSA, 0.5% casein, or 1% milk and 0.05% - 0.5% Tween-20. The phage library, at concentration of $\sim 10^{13}$ phages/ml is blocked in PBS-Tween (PBS-T, Tween concentration varies from 0.05% to 0.5% according to the stringency desired in the selection) with 0.5% BSA, 0.5% casein, or 1% milk for 1 h at RT. The blocked phages are incubated with the target for 2 h at room-temperature. A series of washes in PBS-T were performed; the number of washes and the concentration of Tween were varied according to the stringency desired. Finally, the bound phages were eluted by incubation with 100 μ L of 0.1M HCl for 10 min, and then neutralized with 1:5 volumes of 1M Tris, pH 9.0. The eluted phages can be used to infect TG1 cells growing in exponential phase, and new phages produced by the addition of M13K07 helper phage, in 25 mL of 2YT media supplemented with 50 μ g/mL Kanamycin and 50 μ g/mL Carbenicillin, incubated for 6 h at 37°C followed by 20h at 30°C, with agitation. The amplified phages are finally purified by PEG/NaCl precipitation, and resuspended in PBS for a new round of selection.

Solid phase panning was performed on 5 different targets: Maltose Binding Protein (MBP), Grb2 SH2 domain, Phosphoinositide 3-Kinase 3C (PIK3C3), Xylanase, and SUMO-TEV-SUMO. However, selections did not lead to the selection of any specific binder, as assessed by single-point phage-ELISA (128).

As a more stringent selection procedure, phage solution-phase sorting was performed according to Bostrom J. *et al.* (130), for the same targets as for solid-phase sorting, after a first round of solid-phase selection. Antigens (at a concentration of 0.5 mg/mL in PBS) were biotinylated *in vitro* using Pierce EZ-Link Sulfo-NHS-LC Biotin kit, in a protein to biotin molar ratio of 1:4, for 1 h at room temperature. The reaction was quenched by adding 1M Tris-HCl, pH7.5 to a final concentration of 0.1 M, and incubated on ice for 30 min. The phage libraries (at a concentration of 5×10^{12} /mL) were blocked by PBS supplemented with 0.5% BSA and 0.1% Tween for 1 h at RT. The blocked phage solution was incubated with 250 nM biotinylated antigen for 1 h. In the meanwhile, a NUNC Maxisorp plate had been coated with 5 µg/mL neutravidin in PBS overnight at 4°C, and blocked with 0.5% BSA and 0.1% Tween 1 h at RT. The library-antigen mixes were transferred to the neutravidin coated plate, and incubated for 10 min at RT. Plates were then washed eight times with PBS-0.1% Tween, and bound phages eluted with 0.1 M HCl as previously.

Exponentially growing TG1 cells were infected with the eluted phages, and plated. 96 individual phages were picked from each library, and phage produced by addition of M13K07 helper phage and incubation overnight at 37°C in TB supplemented with 50 µg/mL Kanamycin and 50 µg/mL Carbenicillin.

2.15. Validation of phage display-selected clones by phage-ELISA

Culture supernatants containing phage were used to check for binding by single-point phage ELISA. To do so, 30 µL of supernatant were mixed with 120 µL PBS supplemented with 0.5% BSA and 0.1% Tween. After 1 h incubation at RT, the blocked phages are applied to a NUNC Maxisorp plate coated with the antigen and blocked as previously. After 15 min incubation, the plate was washed three times with PBS-0.1% Tween. Anti-M13 antibody-HRP conjugated was added at a concentration of 1:5,000 and incubated 1 h at RT. The plate was washed four times with PBS-0.1% Tween and developed by adding 100 µL/well of Pierce TMB-Turbo substrate, incubated 10-30 min, after which the reaction was blocked by adding 50 µL 2M phosphoric acid.

Phage clones that seemed positive in the single-point phage ELISA were amplified in 25 mL TG1 cultures and purified by PEG/NaCl precipitation. These purified phages were used to perform dose-response and competitive ELISAs (130). For dose-response ELISA, serial dilutions of phage are applied to a NUNC Maxisorp plate coated with the antigen and developed in the same way as for phage single-point ELISA. For phage competitive ELISA, a constant, non-saturating concentration of phage is mixed with serial dilutions of the antigen in a PBS, 0.5% BSA, 0.1% Tween solution, and incubated 1 h at RT. The mixture is applied to a NUNC Maxisorp plate coated with the antigen and incubated for 10 min at RT. The plate is then washed and developed in the same manner as previously.

2.16. Protein stabilization by interaction with a partner

To be able to co-express two proteins partners in *E. coli*, a new ligation-independent-cloning (LIC) vector was created by gene-synthesis that would be compatible with pET-based expression vectors. This vector is based on pCDF-1b (GE Healthcare) and contains CloDF13 as bacterial origin of replication, a f1 bacteriophage origin, *aadA* Streptomycin/Spectinomycin resistance gene (a *bla* β -lactamase resistance version was also created), T7 promoter and terminator, and a LIC-compatible sequence containing two flanking LIC-overhang sequences and BfuAI restriction sites, *sacB* as negative selection marker (sucrose-sensitivity) and C-terminal TEV-cleavage site and StrepII-tag.

The protein targets were cloned into pNIC-CTHF vector, while interaction partners were cloned into pCDF-LIC-CS (using the ampicillin-resistant version, except for scFv for which the streptomycin-resistant version was used). The temperature of cellular aggregation T_{cagg} was determined for targets alone, or when co-expressed with the interaction partners, from liquid protein expression culture as previously described. The same experiment was also performed by mixing in different ratios independent cultures of target and partner.

For isothermal aggregation experiments, the same procedure was performed as for T_{cagg} determination, but the samples were incubated at a single temperature, 1°C-5°C below the target's T_{cagg} . Samples were extracted from the PCR machine at regular intervals, aggregated protein removed by centrifugation followed by filtration, and dotted on a nitrocellulose membrane for development with anti-His probe as previously described.

2.17. Generation of a co-expression library with a binary code

The superstable VH-domain generated by Hot-CoFi, VH-36-Ser, was chosen as template for co-expression library construction, and cloned into pCDF-LIC-CS (bla). Single-stranded DNA was produced as previously described, and CDR-H2 and CDR-H3 were randomized by Kunkel mutagenesis using the following primers:

CDR-H2: 5' GAATGGGTTGCAAGGATTTMCCTTMTTMTGGTTMTACTTMTTATGCCGATAGCGTCAAG 3'
CDR-H3 (7) 5' CGTCTATTATTGTGGCCGC [TMT]×7 CCTATGGACTACAGGGGTC 3'
CDR-H3 (10) 5' CGTCTATTATTGTGGCCGC [TMT]×10 CCTATGGACTACAGGGGTC 3'

Two libraries were created: VH-YS7, composed of CDR-H2+CDR-H3 (7 residues in length); and VH-YS10, composed of CDR-H2+CDR-H3 (10 residues in length). Some modifications had to be introduced into the Kunkel procedure to obtain optimal results. Primer phosphorylation was performed as previously, but primer annealing was done at 90°C for 3 min, 50°C for 5 min and 37°C on hold. Complementary strand extension and ligation were performed at 37°C for 3h. T7 DNA polymerase was inactivated at 75°C for 20 min, after which the reaction was treated with 100U Exonuclease III (NEB) and 30U uracil-deglycosylase (NEB) for 1 h at 37°C.

The libraries were purified using Qiagen Gel Extraction Kit, and electroporated into TG1 cells. Each library obtained contained $\sim 1.2 \times 10^5$ unique members, and were purified using Qiagen Spin Miniprep plasmid purification kits.

2.18. Screening for binders by interaction-Colony Filtration

The initial proof-of-concept experiment for the identification of binders among a population of non-binders was performed using control proteins. Two different targets, co-expressed with their corresponding interaction partners, were used as positive controls: Grb2 + anti-Grb2 scFv (WT or M1), MBP + anti-MBP DARPin. The negative control was the target co-expressed with an unrelated protein: Grb2+IL1RA, MBP + anti-ERK DARPin. Cells expressing the positive controls were mixed with cells expressing the negative control partners, with negative control cells in large excess (typically 1:100 positive to negative ratio) and plated. The colonies were transferred onto a Durapore membrane, and screened as previously described for Hot-CoFi blot, at a temperature close to the target's T_{cagg} . Positive controls were expected to give high chemiluminescence signal, while negative controls low signal.

To validate that the VH libraries generated were functional and could be used to isolate binders by iCoFi, Grb2-SH2 domain was used as a proof-of-concept target. As the affinities of the binders were expected to be weak, a TorT signal peptide was introduced at Grb2 N-terminus (TorT-Grb2) so that both target and binder could be directed to the periplasm. Rosetta2 cells were transformed with TorT-Grb2 plasmid and plated. A single colony was used to generate electrocompetent cells. These cells were transformed with a library mix composed of library VH-YS7 and VH-YS10 in a ratio of 1:8, and plated on LB-agar plates containing 17 µg/mL Chloramphenicol, 25 µg/mL Kanamycin and 50 µg/mL Carbenicillin. The screen was performed as previously described for Hot-CoFi, at a screening temperature of 50°C (30 min incubation, followed by 30 min lysis-filtration). Colonies judged as positives had a chemiluminescence signal roughly two times higher than background levels (from random colonies surrounding the positives). These positives were picked, streaked onto agar plates, and four new colonies per positive were submitted to a new round of iCoFi screen for confirmation. Positive colonies on the confirmation iCoFi experiment were picked, the VH clones sequenced, and produced in large-scale for further characterization.

RESULTS

1. Target selection, production and purification

The overall goal of the projects presented in this manuscript, was to develop and validate new technologies for the improvement of protein stability, regardless of their function, origin, or size. Generic methods, that can be applied easily and quickly to any protein that a researcher is studying, and for which there is a requirement for stabilization. Having this in mind, in order to validate the method, we chose a diverse set of proteins which have been found to be important in different biopharmaceutical and commercial applications, and for which we had no previous experience. The choice of targets was an important, time-consuming process: extensive literature research, careful selection of the most representative targets, and search for prior information of successful expression in *E. coli*. For targets with disulfide bonds, it was also necessary to choose appropriate signal peptides to direct expression towards the periplasm.

After choosing the targets, they had to be cloned in our expression vectors, and expression tested in small-scale. For many of the targets, insufficient expression was observed and they had to be discarded. For others, expression conditions had to be optimized, e.g. by performing IPTG concentration gradients in small scale. Hundreds of protein variants were produced recombinantly in *E. coli*, and purified following different strategies: for most of them IMAC purification followed by size-exclusion chromatography (SEC); for antibody domains, by periplasmic extraction followed by protein A purification.

A total of about 150 proteins were purified in milligram quantities for characterization. This number does not include the very many that had to be produced several times, nor those that were tested in the initial phase of the project but had to be discarded for diverse reasons: insufficient yield in large scale production, high stability of wild-type protein, or difficulty to monitor unfolding or aggregation *in vitro* with the conventional equipment available in

our laboratory (Differential Scanning Fluorimetry and Differential Static Light Scattering).

It is out of scope for this report to give further details of the results obtained for expression and purification of the many targets tested, such as small-scale purification results, SEC profiles, or SDS-PAGE of purified proteins. However, it is important to bare in mind that significant efforts were carried “behind the scenes”, to produce the results described in this manuscript.

2. Screening for stabilized protein variants by Hot-Colony Filtration

2.1. Thermal stress leads to protein aggregation *in vivo*

As mentioned earlier, thermal stress of a purified protein above a certain critical temperature, usually leads to irreversible protein unfolding and rapid aggregation. Our working hypothesis is that a similar scenario happens for recombinant proteins over-expressed in *E. coli*. When cells are subjected to increased temperatures the proteins unfold and aggregate in a similar way as purified proteins. If this is the case, we should be able to monitor this phenomenon with a modified version of CoFi blot (116), named Hot-CoFi blot. In this method, colonies expressing a target protein are incubated at increasing temperatures, and protein aggregation is monitored through a filtration step followed by dot blot (**Figure 19A**).

We selected a set of nine proteins with a midpoint of unfolding (T_m) of the purified protein ranging from 40°C to 68°C, which were cloned into expression plasmids, transformed into *E. coli* and plated. The colonies were then transferred to a filter membrane that previously has been shown to let soluble proteins pass through while retaining aggregates (116). After expression, the colonies were incubated at a wide range of temperatures after which they were subjected to concurrent lysis and colony filtration (116) (**Figure 19A,B**). In this step, the filter membrane retains aggregates, while soluble proteins diffuse from the lysed colonies onto a nitrocellulose membrane. A fusion tag then allows for the detection of the target protein on the membrane. A high signal indicates the presence of soluble protein, while signal loss accounts for protein aggregation in

the cell. For all proteins tested, there was a good correlation between signal loss and the T_m determined by differential scanning fluorimetry (DSF) that measures the unfolding of the purified proteins (**Figure 19C**) (42). This correlation strongly indicates that it is the unfolding and subsequent aggregation of the protein inside the cell that we are measuring. We therefore consider our screen to be a biophysical screen (i.e. it measures the unfolding and aggregation of proteins in a direct manner), analogous to differential static light scattering (43). This sets it apart from previous published thermostability screens used for directed evolution of stability (for which the unfolding and aggregation event is not measured, but rather inferred indirectly using a secondary read-out such as loss of activity). Interestingly, the midpoint of the cellular aggregation curve (named here T_{cagg}) was generally a few degrees lower than the T_m obtained by DSF ($\sim 5^\circ\text{C}$ lower on average). This difference might reflect the influence of the macromolecular crowding of the cell on protein *in vivo* thermal stability.

To increase the throughput of the method, the T_{cagg} determination was later performed in crude cell extracts, in a liquid format. To do so, protein expression was performed in liquid cultures, cell pellets were resuspended in lysis buffer, aliquoted in PCR plates, and submitted to the same range of temperatures using a temperature gradient in a PCR machine. After the temperature challenge, plates were centrifuged, and the supernatant filtered using a $0.45\ \mu\text{m}$ filter plate. The filtrate was dotted on a nitrocellulose membrane and developed as previously. Results obtained were comparable to the T_{cagg} obtained at the colony level (data not shown).

The wide temperature range accessible by the method covers the T_m of most mesophilic proteins, making the method widely applicable.

Figure 19: Determination of the temperature of cellular aggregation (T_{cagg}) of diverse proteins by Hot Colony Filtration

A. Schematic representation of the Hot-CoFi method. Below the T_{cagg} proteins are soluble and are able to diffuse through the Durapore membrane and bind the nitrocellulose membrane. As temperature is increased, more proteins aggregate, leading to a reduction of diffusion towards the nitrocellulose.

B. Cellular aggregation curves obtained using Hot-CoFi. Signals obtained represent chemoluminescence signal in arbitrary units (AU) from the whole colony, and standardized in mm^2 to account for colony size. Blue dots represent the signals obtained and used for the fitting of sigmoidal curves. Red dots were excluded from the curve fitting; for UL18 and ORF69 they correspond to very intense signals that burned all the substrate on the membrane, leading to an apparent weaker signal.

C. Comparison of melting temperatures (T_m) obtained by DSF against temperatures of cellular aggregation (T_{cagg}) obtained by Hot-CoFi. UL18 is not included as it was not possible to determine its T_m .

2.2. A direct read-out for screening improvements in protein stability in directed evolution approaches

The efficiency of our method (hereafter referred to as Hot Colony Filtration, or Hot-CoFi) to assess protein unfolding and aggregation in cells encouraged us to implement it as a screen for improvement of thermostability using directed evolution approaches. A diverse set of proteins was chosen for stabilization, representing several main applications where protein stability is crucial. For biotechnology and biopharmaceutical applications, a single-chain Fv antibody (against Grb2 SH2 domain, based on Genentech's 4D5 scaffold), Interleukin-1 receptor antagonist (IL1RA, Anakinra, developed by Amgen for Rheumatoid arthritis treatment), and immunoglobulin VH and CH2 domains (currently developed as new scaffolds for binder generation) (94, 95) were chosen. For structural biology, NXR1 (the NUDIX hydrolase domain of human NUDT18) and numerous viral proteins (commonly characterized by loosely-packed cores and unstructured regions) (131) were selected. Finally for biocatalyst stabilization applications, two enzymes were chosen: TEV protease – a highly specific cysteine protease (132), haloalkane dehalogenase (DHAA) – an enzyme with potential use for bioremediation of industrial wastes (127, 133) (**Table 5**).

For all targets, random mutagenesis libraries were generated using error prone PCR in combination with library cloning using Megaprimer PCR of Whole Plasmid (MEGAWHOP) (124). Each library contained $\sim 10^5$ unique members as determined by serial dilution plating after electroporation. Sequencing of 12 clones randomly picked from each library showed an average of one to three amino-acid mutations per gene and less than 10% wild-type sequences. The libraries were screened using Hot-CoFi at a temperature of $\sim 5^\circ\text{C} - 10^\circ\text{C}$ above T_m for the wild-type (WT) protein (**Figure 20**). Positive clones were further verified by re-streaking the picked colonies and re-screening using the Hot-CoFi blot, with WT as a control. Proteins of the clones that were positive in the second screen were produced in large scale and purified by IMAC and SEC. To determine the T_m of the purified proteins, differential scanning fluorimetry (42) was

performed and to determine the midpoint of aggregation (T_{agg}) differential static light scattering was used (43).

Table 5: Description of targets stabilized

Name	Organism	Function	T_m / T_{agg}^*
NXR1	Human	NUDIX domain of NUDT18. Hydrolysis of nucleoside diphosphate derivatives	40.3°C
ORF69	KSHV	Associates with nuclear matrix	51.0°C
UL18	HHV1	Capsid protein	54.4°C*
ORF49	KSHV	Early lytic phase transcription factor	48.3°C
UL14	HHV1	Enhancement of nuclear localization of UL17 and UL26	43.7°C
ScFv	Mouse	Single-chain Fv binding Grb2-SH2 domain (based on Genentech's 4D5 scaffold)	68.4°C
VH	Mouse	Immunoglobulin VH domain from 4D5 humanized IgG (anti-Her2, Herceptin)	64.9°C*
CH2	Human	IgG1 CH2 domain ("nanobody" scaffold)	56.1°C
IL1RA	Human	Interleukin-1 receptor antagonist. Rheumatoid Arthritis treatment (Anakinra, developed by Amgen)	56.6°C
TEV protease	TEV	Highly-specific cysteine protease	52.7°C
DHAA	<i>Rhodococcus rhodochrous</i>	Haloalkane dehalogenase. Potential use in bioremediation and industrial catalysis	51.5°C

Virtually all of the clones selected after the confirmation screen were also more thermostable after purification (95% of positives that were purified and for which a T_m or T_{agg} was obtained), and the range of improvement in thermal stability after just one round of mutagenesis ranged from 5.6°C – 26.6 °C with a median of 8.5°C for the best mutants (**Figure 20B,C**). See **Supplementary Table 1** and **Supplementary Table 2** for a complete list of positive mutants obtained and their corresponding stability. For many of the targets we also further characterized the biophysical and biochemical properties of the mutants and compared them to wild type.

Figure 20: Evolving protein thermal stability by random mutagenesis and Hot-CoFi screening.

A. Typical result obtained from screening a random mutagenesis library by Hot-CoFi. Dots with high intensity represent colonies expressing stable variants of the target protein.

B. Thermal stability improvement obtained for the best mutant of each target (and corresponding standard deviation). For all targets, DSF values are shown, except for UL18(*) and VH-domain(*) for which DSLS had to be used due to poor behavior in DSF.

C. In vitro thermal shift curves of each target's WT and best mutant. For all targets, DSF values are shown, except for UL18(*) and VH-domain(*) for which DSLS had to be used due to poor behavior in DSF. WT is shown in black, the best 1st and 2nd generation mutants are red and blue, respectively.

2.3. Evolution of protein thermostability for structural biology

Initial thermostabilization studies were performed on NXR1 (PDB ID 3GG6). This human protein domain represents a typical structural biology target, for which the structure was previously solved at a resolution of 2.1Å using the multi-construct approach. After screening ~5,000 clones using Hot-CoFi, three stabilized mutant proteins were obtained (**Table 6**). The largest improvements of T_m was 26.6°C for mutant NXR1-1, arising from two mutations. In this case one mutation actually targets the active site, and is therefore likely to abolish activity.

Table 6: Summary of mutants obtained for NXR1

Target	Mutant name	Selection round	Mutations	T_m (°C)	Standard deviation (°C)
NXR1	WT	-	-	40.27	0.28
	NXR1-1	Round 1	V45M, E91V	66.85	0.03
	NXR1-2	Round 1	E95K, S107F	55.76	0.07
	NXR1-3	Round 1	E91G	49.95	0.28

Variants not characterized are not shown. They are available in the Supplementary Table 1.

NXR1 WT and mutants were subjected to crystallization trials. Minimal crystallization efforts yielded a crystal diffracting to 1.5Å for the most stable mutant (NXR1-1), without requiring ligand-stabilization or optimization of crystallization conditions (**Figure 21, Table 7**). This is in line with previous reports suggesting that thermal stabilization leads to increased likelihood of crystallization and better diffraction data (29, 30, 33).

The great stabilization obtained for NXR1-1 can be rationalized from the structure, and arises from both mutations. Mutation E91V removes the constrained glutamate side-chain, for which the two most favored rotamers would lead to a steric clash with the beta strand it faces, and replaces it with valine, which on the contrary interacts favorably with that beta strands backbone through hydrophobic interactions. Mutation V45M is interesting, as this replacement creates significant stabilizing changes. M45 interacts with both C43 and W115 through rare sulfur to aromatic NH and sulfur to cysteine sulphydryl interactions (134), respectively, requiring significant movements of M45 side-chain. This leads to a 90 degrees tilting of W115 side-chain, which

induces the displacement for R111, creating a new charge-charge interaction with E109.

Figure 21: Structure of mutant NXR1-1 at 1.5 Å (light blue – PDB 4HVY), superimposed to wild-type structure (gray, transparent – PDB 3GG6).

(A). Overall structure superposition. RMSD = 0.405. **(B).** Zoom into the mutations and structural changes. The structure is shown as a cartoon, with important residues highlighted as sticks. WT residues are gray. For the mutant, C is red for M45 and V91, and green for other residues; N is dark blue, sulfur is yellow, oxygen is red.

Table 7: NXR1-1 structure data collection and refinement statistics

PDB ID	4HVY
Wavelength (Å)	1.0
Resolution range (Å)	28.52 - 1.46 (1.512 - 1.46)
Space group	I 41 2 2
Unit cell	90.2 90.2 75.03 90 90 90
Total reflections	160294 (24019)
Unique reflections	26703 (2679)
Multiplicity	6.0 (6.1)
Completeness (%)	98.45 (100.00)
Mean I/sigma(I)	15.91 (1.91)
Wilson B-factor	22.56
R-sym	0.042 (0.633)
R-factor	0.1663 (0.2648)
R-free	0.2105 (0.3064)
Number of atoms	1370
macromolecules	1176
ligands	24
water	170
Protein residues	135
RMS(bonds)	0.007
RMS(angles)	1.15
Ramachandran favored (%)	99
Ramachandran outliers (%)	0
Clashscore	11.07
Average B-factor	34.70
macromolecules	30.70
solvent	58.70

Statistics for the highest-resolution shell are shown in parentheses.

2.4. Stabilizing biotechnology tool-proteins and protein drugs

2.4.1. Stabilization of a single-chain antibody fragment

Single chain Fv (scFv) antibody fragments are important tools in modern biotechnology. They are, for example, used in library-based generation of renewable binders (135), which can be further developed into intact antibodies and protein drugs. ScFv also have direct potentials as protein drugs due to improved tissue penetration as compared to IgGs (93). However, the stability of scFvs can vary widely, which often compromises their usefulness in medical applications and as biotechnology tools (136). Hot-CoFi screening of random mutagenesis libraries of anti-Grb2 scFv, based on 4D5 scaffold (anti-Her2 IgG, Herceptin), produced several stabilized mutants, the best of which has a T_m that is 9.5°C higher than WT (**Table 8**). The analysis of the mutations (**Figure 22**) suggests CDR-H3 as a hot-spot for stabilizing mutations. Indeed, five out of eight stable variants isolated by Hot-CoFi contain mutations in this loop, highlighting its critical role for scFv stability. There also seems to be a hot-spot for stabilizing mutations in the first 10 residues of the sequence, suggesting the N-terminus as a potential target to further improve stability using more site-directed libraries.

Table 8: Summary of mutants obtained for anti-Grb2 ScFv

Target	Mutant name	Selection round	Mutations	T_m (°C)	Standard deviation (°C)
ScFv	WT	-	-	68.42	0.10
	ScFv-1	Round 1	W225G	77.96	0.13
	ScFv-2	Round 1	M7K, A226D	71.53	0.03
	ScFv-3	Round 1	T8I, I109V, H227P	71.39	0.16
	ScFv-4	Round 1	M1T, A226D	71.56	0.18
	ScFv-5	Round 1	A226D	71.72	0.11
	ScFv-6	Round 1	I(-8)T S59P	69.94	0.05
	ScFv-7	Round 1	S3P, D4G	69.60	0.10
	ScFv-8	Round 1	S55T, F86S, Q208H	70.67	0.09

Figure 22: Analysis of mutations obtained for anti-Grb2 scFv.

Signal peptide is shown in grey and italics. CDRs are underlined in red. Mutations are shown below the WT sequence, and colored according to the mutant: scFv-1 is red, scFv-2 is orange, scFv-3 yellow, scFv-4 pink, scFv-5 cyan, scFv-6 blue, scFv-7 purple and scFv-8 green.

To confirm that the stabilized mutants retained the affinity of the WT, the affinity against the Grb2-SH2 domain was characterized by SPR (**Figure 23**). Most mutants retained a binding affinity similar to WT. However, the most stable variant (scFv-1), containing a single mutation (W225G) in CDR-H3, lost considerable binding affinity, from an original ~50pM for WT to ~43nM. The reduction in binding affinity of this mutant might be due to the adoption of a new conformation by CDR-H3 that is not favorable for Grb2 binding. Trp is frequently involved in antigen binding (137). Its replacement by Gly is a dramatic change, which leaves position 225 without a side-chain. Therefore, the significant loss of affinity observed for scFv-1 could also indicate a key role of W225 in antigen binding.

Figure 23: SPR data for anti-Grb2 WT and mutants.

Experimental data is coloured according to the variant: WT is grey, scFv-1 red, scFv-2 orange, scFv-3 yellow, scFv-6 blue, scFv-7 purple and scFv-8 green. Regression curves are shown in black. The regressions follow a 1:1 binding kinetics model, except for scFv-1, for which a “two-state conformational change” kinetics model was used.

2.4.2. Stabilization of antibody VH domain

As described in the introduction, other than scFv and Fabs, there are numerous new scaffolds used to generate recombinant binding proteins by phage display. Most of these would benefit from increased scaffold thermostability to allow the generation of libraries of improved quality (e.g. DARPins (138, 139) and Monobodies (140)). VH-domains are the variable domain of the human immunoglobulin heavy chain. They have potential for the generation of binders and for pharmaceutical applications. The small size of VH-domains allows enhanced tumor penetration (94), and makes them good candidates to target so called cryptic epitopes, i.e. narrow cavities present in viral surface antigens which are poorly accessible to intact antibodies (93). Broader application of this scaffold has, however, been hampered by poor stability, mainly due to the loss of stabilizing interactions with the light-chain in the intact antibody (93, 94). To test the Hot-CoFi strategy to screen for the stabilization of recombinant binder scaffolds, we chose 4D5 VH-domain as our parental scaffold, as it has a relatively high aggregation temperature ($T_{agg} \sim 65^{\circ}\text{C}$ at 1 mg/mL) and its stability has been extensively studied by phage-display (94). Screening of $\sim 40,000$ clones by Hot-CoFi at 80°C lead to the identification of 22 unique positive clones (32 total). These variants were pooled together, and used as templates for a new epPCR library created by MEGAWHOP. The 2nd round of screening was performed at 90°C , leading to the identification of 31 additional unique positive clones (35 total), which were further characterized. Due to the high aggregation temperature of these clones, their biophysical characterization was performed by a modified version of the T_{cagg} estimation, in which the thermal stress was performed on full protein lysate in a PCR machine, using a gradient from 40°C to 95°C , followed by centrifugation and filtration to remove aggregates. The soluble fraction was then dotted on a nitrocellulose membrane and developed (**Figure 24**).

The T_{cagg} estimated for WT is $\sim 50^{\circ}\text{C}$, while the best mutant, VH-36, has a $T_{cagg} \sim 72^{\circ}\text{C}$ with potential partial reversible folding (as the background signal obtained for this mutant above 75°C is superior to the background obtained for most mutants and WT). This correspond to a 22°C improvement in aggregation temperature.

Figure 24: Determination of the temperature of cellular aggregation T_{cagg} for the variants of 4D5-VH selected on the second round of Hot-CoFi.

In a previous study, Sidhu S.S. and co-workers found mutations that stabilize the 4D5-VH domain. However, these mutations were found to require the CDR-H3 for the stabilization, as the mutations lead to a conformational change of this loop to cover the hydrophobic patch exposed by the lack of interactions with a light-chain (94). To study the dependence on CHR-H3 for the stabilization of our mutants, we replaced this loop by a deca-serine loop in nine of the 2nd generation mutants and in 4D5-VH WT. The residues replaced were 95-100a, as those positions are usually mutated in phage display libraries (91, 129). A new T_{cagg} estimation was performed, this time from 25°C to 100°C (**Figure 25, Figure 26**). Surprisingly, we see a stabilization of the VH WT by the loop replacement, leading to a $T_{cagg} \sim 60^\circ\text{C}$ (+10°C), with a temperature of cellular aggregation for the best mutant, VH-36-Ser of $T_{cagg} \sim 78^\circ\text{C}$ (+6°C). This suggests that the best

mutant selected is not dependent on the sequence of CDR-H3 for stability. It is interesting to note that the 2nd best mutant, VH-33, increases its T_{cagg} when CDR-H3 is replaced by the poly-serine loop to $T_{cagg} \sim 77^\circ\text{C}$, but low signals below 50°C seem to indicate a reduced solubility of this variant at lower temperatures. This could be seen as a potential partial dependence of VH-33 for the WT CDR-H3, which is not observed for VH-36.

Figure 25: T_{cagg} estimation for 2nd generation variants with Ser₁₀ CDR-H3.

Figure 26: Solubility and resistance to temperature of the best mutants and WT, monitored from 40°C to 95°C .

By comparing the positions of the mutations obtained in this work, with the positions suggested by Sidhu S.S. and co-workers (141) as favorable for stabilization of the VH domain (**Figure 27A,B**), it is possible to highlight common mutations, but also differences. Tonikian R. *et al.* suggest Kabat positions 35, 37, 39, 44, 45, 47, and 50 (141). In this work, of those only positions H35 and R50 were confirmed (G44 is also mutated in some of our mutants, but in low frequency). However, additional stabilizing positions were identified: A78, S93, Y100a, A100b, and W103. It is interesting to note that, by comparing round 1 with round 2 of Hot-CoFi (**Figure 27B**), we can clearly see enrichment of the favorable mutations mentioned. The area targeted by the mutations suggested by Sidhu S.S. and co-workers is adjacent to the one identified in this work (**Figure 27C,D**).

The best two mutants identified by this work, VH-33 and VH-36, contain the mutations R50S, A78V, S93G, A100bP and H35D, S93G, A100bP, W103R, respectively. Interestingly, mutations in R50 and H35D are mutually exclusive in all the clones selected (**Supplementary Table 2**). We speculated that H35D creates a new salt-bridge interaction with R50, as they are located in close proximity. This charge-charge interaction would be located at the center of the hydrophobic patch created by the removal of the light-chain interactions, therefore lowering the hydrophobicity and “stickiness” of the area and increasing stability.

Figure 27: Residues in the VH domain favorable for stabilization.

A. 4D5 sequence and positions mutated. Residues are numbered according to the Kabat nomenclature (142). CDRs are underlined, positions suggested by Sidhu and co-workers as favorable for stabilization (141) are highlighted by an asterisk; the most common mutations obtained in this work are shown below the corresponding WT residue, in red if frequent, or grey if rare.

B. Mutation frequencies obtained for clones selected by Hot-CoFi round 1 (blue) and round 2 (red). Frequencies are calculated on unique sequences, by dividing the number of clones containing a particular residue mutated by the total number of clones obtained per round.

C-D. Residues favorable for stabilization by mutagenesis as suggested by Sidhu S.S. and co-workers (141) (C) and as suggested by this work (D). Representation based on the 4D5-Fab structure (PDB ID 1FVC). 4D5-VH is shown in cartoon representation in blue, parts of the light-chain are shown in orange for reference. Suggested stabilizing amino-acid positions are shown as sticks, in yellow (C) or red (D). H35 and R50 are indicated with an arrow.

In order to get additional insight into the contribution of each mutation to the high stability of the best mutant obtained, VH-36, we solved its structure. Diffractions from two different crystals were solved in parallel, each from a different crystallization condition (**Table 9**): condition A (20 mg/ml of protein in PBS, 0.2 M ammonium sulfate, 30% PEG 8,000), condition B (20 mg/ml of protein in PBS, 0.2 M sodium acetate trihydrate, 0.1 M sodium cacodylate trihydrate pH 6.5, 30% PEG 8,000). Both conditions produced crystals diffracting at high resolution, 1.5 Å and 1.6 Å, respectively.

Each structure contained two molecules of VH-36 per asymmetric unit. Chain A was in both cases fully solved, while chain B lacked electron densities for residues D98-P100b, located in CDR-H3, suggesting flexibility in this loop. The superposition of chain A and chain B from both structures show a highly conserved structure (RMSD=0.4 Å), with variations only in CDR-H3 (**Figure 28A**). The superposition of VH-36 chain A with 4D5-VH domain from 4D5-Fab structure (PDB ID 1FVC), also shows the conservation of the structure (RMSD=1 Å), with the exception of CDR-H3 (**Figure 28B**). In the case of 4D5-Fab CDR-H3 is extended, while for VH-36 CDR-H3 is folded and covering the hydrophobic area usually interacting with the light-chain.

Table 9: VH-36 structure data collection and refinement statistics

	Condition A	Condition B
Wavelength (Å)	1.0	1.0
Resolution range (Å)	19.46 - 1.5 (1.58 - 1.5)	16 - 1.6 (1.69 - 1.6)
Space group	P 1 21 1	P 1 21 1
Unit cell	35.03 83.24 44.19 90 99.74 90	35.16 83.66 44.23 90 99.82 90
Total reflections	295282	246992
Unique reflections	39323	33175
Multiplicity	7.5 (7.4)	7.4 (7.4)
Completeness (%)	99.09 (97.44)	99.97 (100.00)
Mean I/sigma(I)	16.18 (2.30)	15.62 (2.79)
Wilson B-factor	15.96	15.14
R-merge	0.072 (0.799)	0.082 (0.675)
R-work	0.1850 (0.2824)	0.1845 (0.2336)
R-free	0.2163 (0.3203)	0.2145 (0.3014)
Number of atoms	2048	2017
macromolecules	1775	1770
ligands	5	0
water	268	247
Protein residues	233	233
RMS(bonds)	0.006	0.007
RMS(angles)	1.08	1.10
Ramachandran favored (%)	99	98
Ramachandran outliers (%)	0	0
Clashscore	1.72	2.31
Average B-factor	20.30	20.50
macromolecules	18.40	18.60
ligands	34.70	-
solvent	32.30	33.70

Statistics from two crystals are detailed (condition A and condition B).

Figure 28: Structural analysis of VH-36 and comparison with the VH domain from 4D5-Fab.

A: Superposition of VH-36 chain A and chain B from both structures solved (crystallization conditions A and B). Both chains A are shown in blue, while chains B are shown in red. Ribbon representation.

B. Superposition of VH-36 (chain A), with the VH domain from 4D5-Fab (PDB ID 1FVC). VH-36 is shown in blue, while 4D5-VH is shown in gray. Ribbon representation.

C. Surface representation of 4D5-VH (left) and VH-36. Basic residues are shown in blue, acidic residues are red, polars are green and non polars are shown in white.

The analysis of 4D5-VH domain (from 4D5-Fab structure) and VH-36 surfaces (**Figure 28C**), shows that the WT VH domain has an extensive hydrophobic area usually covered by the light-chain. In VH-36, the conformational change in CDR-H3 leads to a more compact structure, in which this hydrophobic surface is covered by polar and charged residues. The analysis of the mutated residues

shows that each of them is implicated in the stabilization of VH-36 (**Figure 29**). D35 is involved in charge-charge interactions with R50, as previously suspected, increasing stability while lowering the aggregation propensity of the area. G93 and P100b promote the folding of CDR-H3, covering most of the former light-chain interaction surface. The replacement of tryptophan by arginine at position 103, removes a hydrophobic residue that would be exposed, placing a hydrophilic and charged residue instead. R103 is therefore also beneficial and contributes to the higher stability and lower aggregation propensity of VH-36.

Figure 29: Contributions of each mutation to the stability of VH-36.

VH-36 is shown in cartoon representation, in blue. D35, R50, G93, P100b and W103 are shown as sticks, and each atom colored in a specific color: C is yellow, N is blue, O is red.

2.4.3. Stabilization of a protein drug

Protein drug stability is critical for their therapeutic application, as low stability hampers their efficacy and safety. IL1RA was chosen as a test example for the stabilization of a protein drug as it is biopharmaceutical characterized by an

aggregation propensity linked to poor thermal stability (34, 143). Therefore, extensive efforts have been made to improve the stability of IL1RA by re-formulation (144) in conjunction with studies of long-term storage and thermal stability (34, 143, 145).

A first cycle of random mutants of IL1RA were produced and screened for improved stability. Four mutant proteins were isolated, all having a stability improvement of $\sim 6^{\circ}\text{C}$ over WT. All positive mutants were combined for a second cycle with additional random mutagenesis followed by stability screening. Five new mutants were obtained with improved stabilities of up to 9.5°C over WT (**Table 10**). Interestingly, most variants in the library contain mutations combined from the four different 1st cycle stabilized clones (as observed from sequencing random clones from the library – data not shown). The most stable mutant, IL1RA-8 (round 2), combines one mutation from IL1RA-1 (A12T) and two mutations from IL1RA-2 (H54N, F148L).

Table 10: Summary of mutants obtained for IL1RA

Target	Mutant name	Selection round	Mutations	T _m (°C)	Standard deviation (°C)
IL1RA	WT	-	-	56.04	0.51
	IL1RA-1	Round 1	A12T, P50H	62.28	0.13
	IL1RA-2	Round 1	P50H, H54N, F148L	61.93	0.35
	IL1RA-3	Round 1	M-7R, W16R, M65L	61.99	0.33
	IL1RA-4	Round 1	A12V	62.07	0.23
	IL1RA-5	Round 2	A12T, P50H, F148L	64.10	0.04
	IL1RA-6	Round 2	S8N, A12V, P50H, H54N, F148L	64.36	0.07
	IL1RA-7	Round 2	A12V, P50H, H54N, F148L	64.34	0.05
	IL1RA-8	Round 2	A12T, H54N, F148L	65.52	0.02
	IL1RA-9	Round 2	A12T, F148L	62.87	0.04

The mutations that are present in the best clone IL1RA-8 are highlighted in red. Most clones share some of those mutations.

To assess whether the mutations affect binding to type I IL-1 receptor (IL1R), WT, IL1RA-2 (round 1) and IL1RA-8 (round 2) were further characterized by SPR. Both mutants have a similar K_D to IL1R compared to WT (**Figure 30**), with a final K_D of 554 pM for IL1RA-8 (290 pM for WT).

Figure 30: SPR experiments of IL1RA (A) WT, (B) IL1RA-2 and (C) IL1RA-8, in triplicates.

IL1RA WT and mutants were immobilized, and IL1-RI injected at concentrations ranging from 10 nM to 312.5 pM, in 2-fold dilutions.

Subsequently, we mapped the mutated residues of IL1RA-8 in the IL1RA-IL1R complex structure (**Figure 31**). The mutations cluster in the opening of IL1RA beta-barrel, suggesting that this area is favorable for stabilization. As can be seen in the structure, they are located far from the interaction site with IL1R domains one and two, but in proximity of the interaction site with domain three. A12 and F148 do not participate in this interaction, while H54 interacts with IL1R's T297 through main chain to main chain interactions (146). Interestingly, in the IL1RA apo structure, H54 has no side-chain density and is part of a partially disordered loop comprising residues 51 – 54 (147). This particular loop changes

conformation and is stabilized upon receptor binding (146). H54N might thus contribute to the marginal reduction of binding. Other stabilized variants contain mutation P50H, a mutation that confirms the importance of this loop for stability. New mutations in this area might allow the selection of further stabilized variants with retention or improvement of affinity.

Figure 31: Structure of IL1RA (blue) complexed with IL1R (orange).

Structure from PDB ID 1IRA. IL1RA residues A12, H54 and F148 are highlighted in red; IL1R residue T297 is highlighted in orange.

In light of these results and the potential therapeutic application of these variants, accelerated stability studies were performed using buffer conditions and concentrations similar to the formulation of the clinical version of IL1RA, Anakinra, as previously described (34). This consists in the analysis of protein aggregation kinetics at temperatures above room-temperature, to assess long-term stability. The best 1st and 2nd cycle mutants (IL1RA-2 and IL1RA-8, respectively) were compared with WT protein in a thermal challenge in CSE buffer, at a concentration of 50 mg/ml over one week at 30°C (**Figure 32**). In

these conditions the WT protein starts to aggregate after 2 h, IL1RA-2 after ~10 h, while IL1RA-8 shows no sign of aggregation even after seven days.

Equivalent long-term stability experiments were performed at 40°C (**Figure 32**). WT started aggregating after 25 min and had aggregated completely in 1 h. IL1RA-2 started aggregating at a similar time-point (33 min), but aggregated at a slower rate, leading to a plateau only after 2 days. IL1RA-8 started to aggregate after 5 –10 hours, and reached a plateau after four days.

Figure 32: Stability of IL1RA and mutants.

Aggregation is measured at OD_{405nm}, in triplicates, at a protein concentration of 50 mg/mL, and at a temperature of 30°C or 40°C. WT is in blue, IL1RA-2 in orange, and IL1RA-8 in red. After 7 days, the samples were heated at 80°C for 30 min to obtain the maximum aggregation signal (filled circles).

2.5. Stabilization of enzymes

2.5.1. Stabilization of TEV protease

To study the usefulness of the Hot-CoFi method for stabilization of biotechnologically important enzymes, the TEV protease was chosen as a test case. For this target, seventeen positive clones were obtained, with stability improvements of up to 9.8°C (**Table 11**). Activity experiments were carried out at room temperature, to compare the performance of these stabilized variants against WT. Most mutant proteins maintained identical activity to the WT (**Figure 33**). The mutants with increased stability but decreased activity obtained have mutations in close spatial proximity to the TEV protease active site, suggesting plasticity in this area, which could be unfavorable for stability but necessary for activity.

Five active and diverse variants (mutants TEV-1, TEV-6, TEV-8, TEV-14 and TEV-17) were selected for a new cycle of random mutagenesis library creation and screening. Twenty-two variants were obtained with up to 14.9°C improvement in stability over WT (**Table 11**). Eight of these were fully active (**Figure 34**). When heat-challenged at 60°C for 20 min before performing activity assays, the WT protein was inactive, while the most stable 1st and 2nd cycle mutants were partially and fully active, respectively (**Figure 35**).

Table 11: Summary of mutants obtained for TEV

Target	Mutant name	Selection round	Mutations	T _m (°C)	Standard deviation (°C)
TEV	WT	-	-	52.14	0.29
	TEV-1	Round 1	D140E	56.36	0.08
	TEV-2	Round 1	E119G, F185L	61.91	0.04
	TEV-3	Round 1	F53L, E119G	61.21	0.10
	TEV-4	Round 1	E119G, E201V	56.55	0.13
	TEV-5	Round 1	L69M, E119G	60.92	0.03
	TEV-6	Round 1	E119G	61.55	0.13
	TEV-7	Round 1	E119G, I179L	60.63	0.06
	TEV-8	Round 1	N25S, L251S	56.82	0.20
	TEV-10	Round 1	S147T, N189Y, N232D	54.96	0.09
	TEV-11	Round 1	P26T, S147T, N189Y	58.64	0.07
	TEV-12	Round 1	E119G, D161A	59.71	0.09
	TEV-13	Round 1	T43I, S170A	54.69	0.02
	TEV-14	Round 1	H180L, Y261S	57.53	0.17
	TEV-15	Round 1	P108H, S135T, N187K	53.94	0.02
	TEV-16	Round 1	H180Q, F230I, Q239R, K262N	56.08	0.01
	TEV-17	Round 1	Q87L	57.39	0.05
	TEV-18	Round 2	Y8D, N25S, T43I, E119G, V212D, L251S	60.71	0.15
	TEV-19	Round 2	Y24F, E119G, D140E	62.30	0.24
	TEV-20	Round 2	N25S, T84A, R118W, E119G, L247H	63.75	0.24
	TEV-21	Round 2	E119G, S170A, N189I, L251S, K262T	63.21	0.23
	TEV-22**	Round 2	N25S, I179V, L251S	52.60	0.26
	TEV-24	Round 2	T43I, K80N, E119G, S170A, I176V	62.88	0.08
	TEV-26	Round 2	F53L, E119G, L223M, L251S	61.63	0.09
	TEV-27	Round 2	N25S, E119G, Q206H, L247P, L251S	63.90	0.30
	TEV-28	Round 2	E119G, S170A, T186N, L251S	64.70	0.24
	TEV-29	Round 2	E119G, F230S, L251S	63.47	0.10
	TEV-30	Round 2	N25S, M100L, E119G, L251S	62.36	0.14
	TEV-31	Round 2	N25S, E119G, L251S	61.14	0.12
	TEV-32	Round 2	V76L, E119G, L251S	60.99	0.12
	TEV-33	Round 2	E119G, M137W	61.27	0.11
	TEV-34	Round 2	F104Y, E119G, S170A, L251S	60.93	0.05
	TEV-35	Round 2	L7S, N25S, Q87L, E119G, L251S	67.01	0.09
	TEV-36	Round 2	N25S, E119G, F230L, V241D, L251S	63.18	0.02
	TEV-37	Round 2	E119G, G226D, L251S	62.00	0.03
	TEV-38	Round 2	N25S, E119G, Q163L, S170A, A182V, P240S, L251S	62.44	0.12
	TEV-39**	Round 2	E119G, G162E, S170A	53.06	0.19

Variants not characterized are not shown. They are available in the Supplementary Table 1.

** The mutant has a lower T_m than WT or the parental clone.

Figure 33: Activity assay for TEV variants obtained from the first round of screening by Hot-CoFi.

Digestion at room-temperature of a substrate containing a TEV protease site is monitored over time, by SDS-PAGE.

Figure 34: Activity assay for TEV variants obtained from the second round of screening by Hot-CoFi.

Digestion at room-temperature of a substrate containing a TEV protease site is monitored over time, by SDS-PAGE.

Figure 35: Activity characterization of TEV WT and mutants at room-temperature without (A) and with (B) a 20min heat-challenge at 60°C.

The sequences of all the stabilized 2nd cycle mutants were analyzed, and the frequency of each mutation was calculated. The most conserved mutations were highlighted on TEV protease structure in **Figure 36**. E119G is the most conserved mutation (frequency of 94%), followed by L251S (78%), N25S (44%), and S170A (28%). The most stable variant, TEV-35 (round 2), combines the three most frequent mutations: one from TEV-6 (E1119G) and two from TEV-8 (N25S, L251S). In addition, it contains one mutation from TEV-17 (Q87L), plus one new mutation (L7S). Interestingly, the three most stable active variants, TEV-35, TEV-20 and TEV-26, contain mutations Q87L, T84A and F53L respectively, all located in the same small hydrophobic core, suggesting that this area might be amenable for further stabilization without compromising TEV protease activity.

Figure 36: TEV structure sheds light into the stabilizing effect of mutations.

TEV structure (PDB ID 1LVM) is shown in blue (chain A is opaque, chain B transparent). E119G is highlighted in purple, L251S as a red sphere at the C-terminus (as the residue is missing in the structure), N25S is orange and S170A is yellow. Q87L, T84A and F53L are green. TEV digestion product (which indicates the active site) is shown in gray.

2.5.2. Stabilization of Haloalkane Dehalogenase

Haloalkane Dehalogenase (DHAA) from *Rhodococcus rhodochrous* is an enzyme of particular industrial interest, because it is able to selectively convert several industrial waste chemicals such as 1,2-dichloropropane (DCP), 1,2,3-trichloropropane (TCP), and 1,2-dichlorobutane (DCB), into more valuable chlorhydrins (127). An example of such applications is the evolution of DHAA for the enantioselective conversion of TCP into (R) or (S)- DCP, which are further converted into epichlorhydrins, building blocks of many fine chemicals such as atorvastatin (cholesterol lowering drug), (S)-viloxazine (anti-depressive) or (S)-bisopropol (β -blocker) (148).

DHAA activity against TCP can be easily monitored by a colorimetric assay, which links the release of hydrochloride molecules to a change of pH (**Figure 37A**). This type of assays have been used in directed evolution approaches to improve DHAA activity against TCP (133), and the enzyme's thermal stability

(127). We chose to use DHAA a proof-of-concept example for the use of Hot-CoFi to improve enzyme properties for industrial applications.

Figure 37: DHAA WT and mutants, activity and stability

A. Enzymatic dehalogenation of TCP by DHAA. DCP and hydrochloric acid are produced.

B. Initial activity of DHAA WT and mutants at room temperature or 40°C.

C. Residual activity after heat-challenge.

The initial and residual activity are calculated on the linear phase of the curve, at the start of the reaction, and which correspond to the maximum speed of reaction under the conditions studied.

DHAA WT T_m is 53.3°C. A 1st round of selection by Hot-CoFi at 62°C lead to the identification of seven clones which were more stable (as determined by T_{cagg} measurements, data not shown) and active in preliminary activity assays (data not shown). They were used as parental templates to generate a new MEGAWHOP library followed by a Hot-CoFi screen at 72°C. Six new stabilized variants were identified (**Table 12**), the best of which (DHAA-8) has a $T_m=62.5^\circ\text{C}$ (improvement of +9.2°C over WT).

Table 12: Summary of mutants obtained for DHAA

Target	Mutant name	Selection round	Mutations	T _m (°C)	Standard deviation (°C)
DHAA	WT			53.31	0.68
	DHAA-5	Round 1	E99V, A172V	59.44	0.41
	DHAA-6	Round 1	A172V	58.57	2.94
	DHAA-7	Round 1	G171C, A172V	61.84	0.35
	DHAA-8	Round 2	G171C, A172V, L194I	62.55	0.84
	DHAA-10	Round 2	A172V, N227D	60.44	0.81

Variants not characterized are not shown. They are available in the Supplementary Table 1.

The three more stable variants from round 1 and 2 were further characterized by measuring their activity at RT and 40°C (**Figure 37B**). The maximum activity was measured during the linear increase of signal observed at the beginning of the reaction, showing a modest improvement of up ~20% at RT (for DHAA-10) and at 40°C (for DHAA-6). Interestingly the activity of all DHAA variants, including WT, is about two times higher at 40°C than at RT, highlighting the potential benefit of performing the bioremediation reaction at higher temperatures.

To test the robustness of the WT and mutants against heat-stress, they were subjected to a 20 min heat challenge at 35°C, 45°C, 50°C and 55°C followed by activity measurements at RT (**Figure 37C**). WT has a 50% loss of activity when exposed at 45°C, and a complete loss when incubated at 50°C, while round 1 and round 2 produced mutants retaining full activity when challenged at 45°C, more than 90% of the activity after 50°C treatment, and up to 14% residual activity after challenge at 55°C.

Increasing the temperature of the reaction to 40°C doubled the speed of reaction. Another option to increase productivity in an industrial setting would be to increase enzyme concentration. We decided to test the behavior of WT and the three best mutants (DHAA-7, -8 and -10) with higher enzyme concentrations at 40°C (**Figure 38**). Buffer strength was increased ten times to be able to monitor the reaction more accurately, and to slow down the change of pH over time. The increase in concentration led to an increase of the initial activity observed, as expected, but not in a linear manner. Indeed, the initial conversion rates doubled when enzyme concentrations were increased from 0.15 mg/mL to 0.75 mg/mL, but did not increase significantly when the concentration was increased to 1.5

mg/mL. This might be explained by limitations in the speed of diffusion of the substrate.

The final yield, as extrapolated from the final OD_{620nm}, was the highest at 0.75 mg/mL, suggesting that this enzyme concentration is optimal in the reaction setting of this study. At this concentration, the reaction continues throughout the 1.5 h measured, while at 1.5 mg/mL it seems to plateau after 40 min.

All the mutants seem to be active throughout the 1.5 h of measurement, while WT seems to aggregate in a concentration dependent manner, as suggested by the increase in turbidity of the reaction (which leads to an increase in OD_{620nm}). WT does not aggregate at 0.15 mg/mL, but aggregates rapidly after 35 min at 0.75 mg/mL, and after 15 min at 1.5 mg/mL. The aggregation propensity of DHAA WT could be due to a lower stability to the change of pH produced during the reaction, and/or to the lower thermal stability of the protein.

DHAA-7 has the mutations G171C, A172V, DHAA-8 has G171C, A172V, L194I and DHAA-10 A172V, N227D. It seems that G171 and A172 are hot-spots for stabilization. This is in accordance to previous work highlighting the importance of these residues in DHAA stability, and confirm the importance of N227 as well (149, 150).

The mutants obtained are more stable than wild-type, present similar activity towards TCP, while being able to withstand a reaction setting at higher temperature and higher protein concentration. The variants generated are therefore more apt for industrial applications.

Figure 38: Influence of enzyme concentration in TCP dehalogenation by DHAA.

A. Colorimetric assay of DHAA activity at 40°C and 0.15 mg/mL enzyme

B. Colorimetric assay of DHAA activity at 40°C and 0.75 mg/mL enzyme

C. Colorimetric assay of DHAA activity at 40°C and 1.5 mg/mL enzyme

D. Initial activity as a function of DHAA concentration

The initial activity is calculated on the linear phase of the curve, at the start of the reaction, and which correspond to the maximum speed of reaction under the conditions studied.

3. Stabilization of proteins by interaction with a partner

3.1. Selection of binders by phage display

3.1.1. Library creation

Phage display is a widely accepted method for the selection of binders generated *de novo*. An extremely diverse library of phages, displaying billions of variants of a protein at their surface, have the potential to contain clones that bind a diverse set of targets. Selection procedures allow for the isolation of the binders specific for a desired antigen (51).

It is essential to select an appropriate framework protein, into which diversified loops will be introduced, as well as the correct amino-acid composition of those loops, to obtain a library with the potential to contain high affinity binders to a diverse set of potential targets (50, 86, 120, 151).

We chose the 10th fibronectin type III (FN3) as the preferred scaffold, since it is a well established scaffold. It has been successfully employed to create phage display libraries, which have been used to select several high affinity binders. Moreover, it has high expression in *E. coli*, high stability, and do not contain disulfide bridges, which facilitates cytoplasmic expression (118-120, 128). A phage display vector was created based on previously published work by Koide S. and co-workers (118, 128). A fusion construct was created composed of, from 5' to 3', DsbA signal sequence, FN3 gene, His₆ tag, amber stop, and the C-terminal portion of bacteriophage M13 g3p protein; and cloned into pPCR-Script SK(-) vector (**Figure 39A**).

FN3 contains three loops amenable to randomization, BC, DE and FG (**Figure 39B**). A phage display library was created by randomizing loops BC and FG, since these loops have the highest plasticity (119). Amino-acid diversity was designed to mimic a naïve antibody repertoire composition, following Lee C.V. *et al* strategy (50), by skewing the proportion of bases (A, T, G, C) in each codon during oligonucleotide synthesis. The bias introduced at the nucleotide level leads to a biased codon composition, and eventually amino-acid frequencies following a desired design.

In naïve murine and human antibody repertoires, the length of CDR-H3 varies significantly, the majority of members having loops of 7-16 residues in mice, and 10-22 residues in humans (152). To mimic naïve antibody libraries, additional structural diversity was designed by varying the loop length: the BC loop was varied between 6 and 8 residues, while the FG loop was varied between 7 and 19 residues. Due to all possible BC-FG combinations this leads to a total of 39 libraries. After library creation, the library quality and amino-acid composition was verified by sequencing 96 randomly picked clones from each library, and compared with the original design (**Figure 39C,D**).

The template FN3 sequence used for library creation contains a stop codon in the BC and FG loops. Sequencing shows a 47% success in replacing both loops with a randomizing sequence without disturbing the reading frame. As the total library size obtained was $\sim 8 \times 10^{10}$ unique members, this leads to a functional library of $\sim 3.8 \times 10^{10}$ members, which is comparable to library sizes reported by other groups (50, 91, 120). The skewed amino-acid composition obtained follows the desired design, with significant differences only for His, Asn and Cys in the BC loop (+6.5%, +4.6% and -3.2% compared with the design, respectively) and Gly and Pro in FG loop (-3.4% and +3.1%, respectively). These differences might arise from differences in coupling efficiencies for the bases during oligonucleotide synthesis, or preferred oligonucleotide binding during library synthesis.

Figure 39: FN3 phage display library construction and analysis.

A. Schematic representation of the vector used for the phage library.

B. Structure of the 10th Fibronectin type III domain. Cartoon representation based on structure of WT FN3 (PDB ID 1FNA). The WT structure is shown in blue (this includes the DE loop), while the mutated loops BC and FG are highlighted in orange and red, respectively.

C-D: Theoretical and observed amino-acid composition for FN3 library. (C) BC loop, (D) FG loop.

3.1.2. Selection of binders against diverse targets

To test whether the FN3 library contain binders against a wide variety of targets, five proteins were chosen for panning as proof-of-concept antigens: Maltose Binding Protein (MBP), Grb2 SH2 domain, Phosphoinositide 3-Kinase 3C (PIK3C3), Xylanase, and SUMO-TEV-SUMO.

These targets were subjected to solid phase phage display selections, in which the library is applied to plates coated with the antigen, and non-binding phages are removed by washing. The stringency can be controlled by the number of washes, the presence of a competitor such as excess of BSA, casein or milk, as well as by the concentration of detergent Tween-20 (0.05%-0.5%) used in the binding and wash buffer. Bound phages are retrieved and amplified, and new rounds of selections are performed.

A wide variety of solid phase selection strategies were performed, by modifying the stringency of the selection (different number of washes, presence of competitor, Tween concentration, etc.), but none allowed for the selection of specific binders (data not shown). Sequencing showed that the majority of the clones selected contained stop codons in the FN3 gene. This suggests that at each round of selection, these truncated constructs were enriched, probably due to higher growth and phage production rates during phage amplification, compared to full-length FN3 variants.

To increase the stringency of the selection procedures (and eliminate more accurately the truncated FN3 phages), a solution sorting strategy was pursued. A first round of selection was performed in solid phase, in the presence of 1% milk and 0.05% Tween, followed by a second round of solution sorting against 250 nM of biotinylated antigen in the presence of 0.5% BSA and 0.1% Tween. The number of rounds of selection was kept to a minimum to avoid enrichment of truncated clones, two rounds giving optimal results.

After the second round, the eluted binding phages were used to infect TG1 cells and plated. For each target, 96 individual clones were picked randomly, phages produced in 1 mL cultures, and binding tested by phage single-point binding ELISA (**Figure 40**). The phage clones that gave high signal were sequenced (**Table 13**).

Figure 40: Single-point phage ELISA in 96-well format for all targets.
Potential binding clones (highlighted in with red circles) were picked for sequencing

Table 13: Sequence of BC and FG loops for selected FN3 phage clones

Target	Clone Name	BC	FG
MBP	FN.MBP.A1	Stop	Stop
	FN.MBP.D9	KKSLNCF	Frame-shift
	FN.MBP.H12	PRHYYYYS	Stop
	FN.MBP.F1	YRSHIG	GADRRNRSGDVHTA
Xylanase	FN.Xyl.A5	VHSHYD(amber)	Stop
	FN.Xyl.H4	YNINFRN	NHESPRFNL
	FN.Xyl.B9	DMLSGNYH	Frame-shift
	FN.Xyl.H2	SYHHYDL	SVATRDSNAH
PIK3C3	FN.PIK.D6	YEFSKT	Stop
	FN.PIK.G8	IINCDSN	Frame-shift
	FN.PIK.H9	Stop	Frame-shift
	FN.PIK.C12	ICFSFD	YTRDCNTHY
	FN.PIK.G12	RNITYHV	Frame-shift
Grb2	FN.Grb2.A1	Stop	Stop
	FN.Grb2.A9	Frame-shift	Frame-shift
	FN.Grb2.C10	Frame-shift	Frame-shift
	FN.Grb2.B5	Stop	Stop
SUMO- TEV- SUMO	FN.SUMO.F2	HYFSHHYN	Frame-shift
	FN.SUMO.H11	YHVNVD	RYDGNVSDSADCAD
	FN.SUMO.H7	DHHNLYT	TDNHAGDH(amber)
	FN.SUMO.H3	GDSVYYDN	double clone
	FN.SUMO.A1	(amber)SNCFY	(amber)RGPHHGNSH
	FN.SUMO.G6	TYVSCRC	EPYYTHVASH

Remark: amber codons (amber) code for Glu in suppressor strains such as TG1.

The presence of clones that produce high signal in single-point ELISA but have stop codons or frame-shifts seems to support the hypothesis of higher phage production for these clones, which leads to higher background signal as phage concentrations are not normalized. Nevertheless, there were several clones that have a complete FN3 sequence and high single-point ELISA signal.

Phage FN3 clones that were confirmed to be functional at the sequence level (no stops or frame shifts) were amplified and purified, and further characterized by phage competitive ELISA, to determine the apparent inhibitory concentration

IC₅₀ (**Figure 41**). All clones seem to show binding, and the majority has an inhibitory concentration IC₅₀ in the micromolar range, with the exception of anti-Xylanase clones which did not show significant inhibition.

Figure 41: Preliminary binding analysis of clones selected by competitive phage ELISA.

Remark: Xylanase did not produce typical inhibition curves, but the FN3 clones were nevertheless further characterized.

All the clones tested by phage competitive ELISA were cloned into an expression plasmid for expression as non-fusion proteins in the cytoplasm, and purified by IMAC and SEC. Binding kinetics were tested both by Surface Plasmon Resonance

(**Figure 42**) and by Biolayer Interferometry (**Figure 43**). Not all targets yielded kinetic constants in both methods (it is worth noting that limited optimizations were performed): SPR yielded kinetic constants for FN.MBP.F1, FN.Xyl.H2 and FN.Xyl.H4, while BLI produced kinetic constants for FN.SUMO.G6, FN.SUMO.H11, FN.PIK.C12 and FN.MBP.F1. This highlights the complementary of both methods to study protein-protein interactions.

SPR and BLI confirmed binding interactions in the low micromolar range for all the clones. The K_D values are not precise, as the signals obtained in both methods are weak, and therefore prone to inexactitudes. Nevertheless, they confirm that we selected binders for four out of five targets tested. This corroborates that the FN3 library is functional, and that it contains binders against a wide variety of targets.

The affinities obtained are in accordance with previously reported affinities obtained for naïve FN3, Fab and peptide phage libraries (50, 53, 118) but inferior to Fab-dimer libraries (50) or libraries highly skewed for Tyr and Ser (91, 120, 151). The weak affinities obtained probably arise from the randomization of only two loops, as well as a library design that might not be optimal for this scaffold. The use of multiple loops might provide an extended binding surface, as well as higher avidity within the binding molecule (i.e. the binding of one loop would provide a platform for the binding the other loops, and inversely the dissociation of one loop would more difficultly lead to the dissociation of the remaining bound loops). The skewed amino-acid composition was originally designed for an antibody molecule, and might not be adapted for the FN3 scaffold; a more careful design might prove beneficial.

Figure 42: Binding kinetic analysis of selected clones by Surface Plasmon Resonance.

Experimental measurements are shown in black, while kinetic regression curves are shown in blue. Binding constants are indicated.

Figure 43: Binding kinetic analysis of selected clones by Biolayer Interferometry.

Experimental measurements are shown in black, while kinetic regression curves are shown in blue. Binding constants are indicated.

3.2. Stabilization of proteins by interaction with a partner

3.2.1. Interaction of target protein with a more stable partner leads to stabilization

In the previous sections, we have demonstrated that we can monitor the temperature-induced aggregation of proteins inside the cell using Hot-CoFi. We also demonstrated that the method is robust for the identification of variants with improved stability in directed evolution campaigns. As described in the introduction, interactions of a protein with a ligand can lead to thermal stabilization of the protein *in vitro* (33, 41-43, 46-49). We reasoned that Hot-CoFi could be adapted to monitor the thermal stabilization of proteins by a ligand in crude, unpurified samples. We named this modified method interaction-Colony Filtration, or iCoFi.

We determined the T_{cagg} of a diverse set of proteins (**Table 14**) when expressed alone, as well as in the presence of a known interaction partner (**Figure 44**, **Table 15**). The interaction partners were tested both in co-expression cultures, as well as by mixing in different ratios cultures expressing each partner alone. When target and partner were mixed from independent expression cultures, the mixtures were supplemented with an unrelated target (anti-ERK DARPin) so that each ratio would contain the same amount of total crude extract (e.g. 100 μ L of expression culture from the target was supplemented with either 1000 μ L of culture from unrelated partner, or 1000 μ L of known partner, to give a target to partner ratio of 1:0 or 1:10, respectively).

Table 14: Targets tested for stabilization by interaction with partner

Target	Partner	Affinity	Reference
Grb2	anti-Grb2 scFv	50 pM	This study
MBP	anti-MBP DARPin	4.4 nM	(153)
NPDZ	SAM	1.3 nM	(154)

Figure 44: Protein stabilization through interaction with a partner.

A. Schematic representation of the interaction-CoFi blot (iCoFi). A plasmid coding for the protein of interest (red dot) is co-transformed with a co-expression plasmid coding for potential partners (blue dot) into bacteria and protein expression is induced. If the partners do interact, the complex leads to the stabilization of the target, leading to a change in the temperature of cellular aggregation (T_{cagg}) of the target protein. The development procedure is identical as for the Hot-CoFi.

B. Schematic representation of vector pCDF-LIC-CS.

C. Examples of stabilization of diverse proteins by interaction with their corresponding partner. Aggregation curves were obtained by T_{cagg} estimation from liquid cultures expressing either the target alone, or co-expressed with a known interaction partner. For MBP and NPDZ, different target to partner ratios were tested by mixing separate cultures.

Table 15: Summary of the temperatures of cellular aggregation (T_{cagg}) obtained by interaction CoFi blot for three diverse targets

Target	Condition	T_{cagg} (°C)	St. Dev. (°C)
Grb2	Grb2 alone	39.9	0.4
	Grb2 + anti-Grb2 co-expression	51.1	0.4
MBP	MBP alone	46.1	0.4
	MBP + anti-MBP co-expression	48.7	0.5
	MBP 1:1 anti-MBP	50.0	0.4
	MBP 1:5 anti-MBP	47.4	0.3
	MBP 1:10 anti-MBP	49.2	0.3
NPDZ	NPDZ alone	41.3	0.1
	NPDZ + SAM co-expression	41.7	0.1
	NPDZ 1:5 SAM	42.5	0.4
	NPDZ 1:10 SAM	42.1	0.1

All experiments were performed in triplicates, except for Grb2, which was done only once.

A new co-expression vector (**Figure 44B**) was designed for this purpose and produced by gene synthesis. This phagemid is based in vector pCDF-1b but was modified significantly to contain a ligation-independent cloning site, a C-terminal Strep-II tag, as well as a f1 origin of replication. The vector was named pCDF-LIC-CS, with a streptomycin resistance (pCDF-LIC-CS) as well as an ampicillin resistance version (pCDF-LIC-bla).

For some targets, the presence of a partner leads to a significant stabilization (**Figure 44C**). The co-expression of the scFv with Grb2, for example, leads to an increase of Grb2 apparent T_{cagg} of 11°C, from $T_{cagg}=40^{\circ}\text{C}$ to $T_{cagg}=51^{\circ}\text{C}$. For others the stabilization is small, but reproducible. This is the case for example for NPDZ interacting with SAM, which leads to a modest 0.4°C to 1.2°C stabilization, according to the ratio of protein to partner tested. MBP interacting with anti-MBP shows also a modest 1.2-3.1°C increase in T_{cagg} .

The smaller stabilizations obtained for NPDZ and MBP as compared to the Grb2-anti-Grb2 scFv interaction can be analyzed through the structure of the complexes (**Figure 45**). Anti-MBP DARPin interacts with its target through a relatively small surface area that does not seem to be directly implicated in structural stabilization. The interaction between MBP and anti-MBP DARPin buries an accessible surface area (ΔASA) of 611.2 Å², relatively small compared to

other protein-protein interactions, which tend to have ΔASA above 800 Å (153). Moreover, this DARPin binds MBP in its open, ligand-free conformation, which could be a less stable conformation.

In the case of SAM-NPDZ interaction, binding is performed through two distinct domains, a SAM domain and a C-terminal PBM domain, forming a clamp around NPDZ, which buries a large ΔASA of 1,309 Å (**Figure 45**) (154). This mode of interaction would be expected to stabilize considerably both proteins. One possible explanation for the low stabilization observed is the expression construct used in our experiments. The main contributor to the SAM-NPDZ interaction is SAM's C-terminal PBM domain, particularly the last three residues (T459-L461) (154). However, our expression construct contains a C-terminal Strep-II tag downstream of L461. This additional sequence could lead to a significant reduction in affinity through steric hindrance of T459-L461 interactions with NPDZ.

The stabilizations observed in crude extracts was confirmed on purified proteins, by monitoring the melting temperature of targets alone or in the presence of their corresponding partner, by Differential Scanning Fluorimetry (**Figure 46**). All targets show an increased T_m when their partner is added. This confirms that our method is monitoring an increase in stability induced by the binding event, which prevents the unfolding followed by aggregation of the target protein.

Figure 45: Complex structure of two targets studied by interaction CoFi.

A. MBP in complex with anti-MBP DARpin (PDB ID: 1SVX).

B. NPDZ in complex with SAM-PBM (PDB ID 3K1R).

Figure 46: Protein stabilization by interaction with partner, monitored by Differential Scanning Fluorimetry

To further validate that the modest stabilizations obtained for MBP and NPDZ are real, aggregation experiments at static temperature were performed (i.e. isothermal aggregation). The protein target was incubated with its partner in a ratio of 1:1, 1:5 and 1:10 (of independent expression cultures), as well as in co-expression. The samples were incubated at a temperature close to the target T_{cagg} , measuring remaining soluble protein by centrifugation, filtration and dot blot as previously described (**Table 16, Figure 47**).

For both targets tested, isothermal aggregation in the presence of a partner leads to an increased half-life as compared with the target alone. The increase is maximal for target to partner ratios of 1:10, highlighting that protein ratios are important to obtain maximum stabilization. For MBP, a ratio of 1:10 leads to an increase in half-life at 50°C of 12.5 min. For NPDZ the same ratio increases $T_{1/2}$ by 8.4 min at 42°C.

These results confirm that proteins can be stabilized by interaction with a partner, and this stabilization can be detected by the modified version of Hot-CoFi presented here. The protein to partner ratio seems to be important, as excess of partner seems to lead to greater stabilizations.

Table 16: Summary of the half-life ($T_{1/2}$) obtained by isothermal aggregation monitored by iCoFi blot

Target	Temperature (°C)	Condition	$T_{1/2}$ (min)	R^2
MBP	50	MBP alone	7.58	0.984
		MBP + anti-MBP co-expression	12.89	0.950
		MBP 1:1 anti-MBP	8.15	0.986
		MBP 1:5 anti-MBP	11.95	0.982
		MBP 1:10 anti-MBP	20.11	0.977
NPDZ	42	NPDZ alone	6.62	0.966
		NPDZ + SAM co-expression	5.01	0.823
		NPDZ 1:1 SAM	6.61	0.979
		NPDZ 1:5 SAM	9.74	0.926
		NPDZ 1:10 SAM	15.00	0.989

All experiments were performed in triplicates.

Figure 47: Isothermal aggregation curves of MBP and NPDZ, alone or with different ratios of their corresponding partner, determined by iCoFi.

3.2.2. Screening of binders by interaction-CoFi: proof-of-concept

Encouraged by these promising results, we decided to explore the potential use of iCoFi for the identification of binders leading to stabilization of a target among a population of non-binders, a situation encountered in *de novo* generation of binders. To do so, we co-transformed Grb2 with IL1RA: this combination of proteins should not lead to an increase of Grb2 aggregation temperature, and therefore is a negative control. A bacterial culture of cells harboring Grb2+IL1RA plasmids was spiked with cells harboring Grb2+scFv WT and Grb2+ScFv M1 plasmids in ratios of approximately 400:1 and 100:1, respectively. scFv WT has an affinity of 50~pM while scFv M1 of ~50nM. The spiked culture was plated, and iCoFi was performed at 50°C (**Figure 48A**). By measuring the chemiluminescence signal intensities of the colonies, three distinctive populations of colonies could be distinguished: low, medium and high intensity colonies. The low intensity colonies were attributed to the negative control, the ones giving medium intensity to colonies harboring Grb2+scFv M1 plasmids, while high intensity colonies to Grb2+scFv WT. Eight high intensity colonies from iCoFi at 50°C were picked, and seven of them confirmed as scFv by PCR with specific primers (data not shown).

A similar experiment was performed for MBP. Cells co-expressing MBP with anti-ERK DARPin were used as negative control, and spiked in a ratio of ~100:1 with cells co-expressing MBP + anti-MBP DARPin. After plating, iCoFi was performed at 60°C, giving similar results as for the previous target (**Figure 48B**). The positive clones were confirmed to be anti-MBP by sequencing (data not shown).

Figure 48: Identification of binders by interaction-Colony Filtration.

A. Identification of colonies co-expressing scFv WT or M1 with Grb2 among colonies expressing Grb2 with a non-binder (IL1RA) by iCoFi at 50°C.

B. Identification of colonies co-expressing MBP + anti-MBP DARPin among colonies expressing MBP with a non-binder (anti-ERK DARPin) by iCoFi at 60°C.

3.2.3. Screening of *de novo*-generated binders by interaction-CoFi

An important potential application for the iCoFi method is to identify binders generated *de novo*. This would be an alternative to methods such as phage display, ribosome display or yeast surface display, but performed at the colony level. In a similar manner as for the FN3 library, a protein scaffold is chosen; carefully selected loops are randomized at specific positions, with a desired amino-acid composition. The library is subjected to selection by interaction-CoFi against a wide variety of targets, and identifying particular clones that interact specifically with a desired target, and stabilize it.

A significant handicap for iCoFi when performing selections on randomized libraries is the limited screening capacity. Display technologies such as yeast display, phage display or ribosome display can select clones from libraries containing a diversity of typically 10^7 , 10^{10} and 10^{12} unique members, respectively (83). With iCoFi, we could expect to screen comfortably a maximum of ten plates containing 10^5 colonies each, i.e. 10^6 unique members. Diversity must be therefore considerably restricted. Previous works by Fellouse F.A. *et al.*

have demonstrated that antibody libraries randomized with a binary code of random combinations of Serine and Tyrosine residues are sufficient to generate specific binders with affinities in the nanomolar range (91, 151). This strategy seems ideal for our purposes. However, in previous reports at least three loops have been randomized, with library diversities of $\sim 10^{10}$ (91, 151). A potential advantage of our method is the high stringency of the biophysical stability screen, which should limit the number of false positives.

The superstable autonomous VH domain, VH-36 identified by Hot-CoFi was selected as our scaffold for library generation. This scaffold is based in the 4D5-Fab, which has been used extensively in phage display approaches, including libraries with a binary code (91, 151). We hypothesized that the fold and loop exposure was not significantly changed in this construct as compared to the Fab, and therefore we could follow a similar library design as in previous works (91, 151) (**Figure 49A**). However, once we solved the structure of the VH-36 scaffold, some differences became apparent particularly in CDR-H3 (**Figure 49B**). Nevertheless, the residues to be mutated conserved a similar solvent exposure, and therefore seemed a valid choice for randomization. CDR-H3, the main interaction loop, was randomized with either seven- or ten-residues Tyr/Ser random loops, allowing high flexibility. CDR-H2 was only randomized at positions 52, 53, 54, 56, 58, also with Tyr/Ser combinations, and therefore is more rigid. This corresponds to a theoretical library size of $\sim 7.4 \times 10^4$ unique members, which could be comfortably screened by iCoFi. The library was generated in our in-house co-expression vector pCDF-LIC-bla, by Kunkel mutagenesis. $\sim 80\%$ of clones contained at least one loop randomized, and $\sim 50\%$ were randomized in both.

Figure 49: VH library design.

Cartoon representation (blue) based either on Fab structure 1FVC (A) or VH-36 structure (B). Residues to be mutated with Tyr/Ser are highlighted in yellow.

Grb2 SH2 domain was selected as our first, proof-of-concept target. As the VH library is expressed in the *E. coli* periplasm through a stII signal peptide, we directed the expression of Grb2 into the same compartment, but through a different signal peptide, TorT.

The VH library was electroporated into cells containing the TorT-Grb2 plasmid, and plated. Interaction-CoFi was performed following Hot-CoFi procedures, screening ~20,000 clones at 50°C. Two clones were confirmed as positives, VH.aGrb2.1 and VH.aGrb2.2 (**Figure 50**). Sequencing showed that VH.aGrb2.1 contains mutations in both CDR-H2 and CDR-H3, while VH.aGrb2.2 has mutations only in CDR-H3 (**Table 17**).

Figure 50: Confirmation iCoFi for anti-Grb2 VH clones VH.aGrb2.1 and VH.aGrb2.2.

Each colony picked in the iCoFi screen was re-streaked, and four new colonies were picked per clone. They were submitted to a new round of interaction-CoFi at 50°C. Confirmed positive clones produce higher signals as compared to random clones from the library.

Table 17: Sequences of anti-Grb2 VH clones selected

	CDR-H2										CDR-H3													
	51	52	52sa	53	54	55	56	57	58	59	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	101
VH template	I	Y	P	T	N	G	Y	T	R	Y	R	S	S	S	S	S	S	S	S	S	S	P	M	D
VH.aGrb2.1	I	Y	P	S	Y	G	Y	T	Y	Y	R	S	S	Y	S	S	S	Y	Y	S	S	P	M	D
VH.aGrb2.2	I	Y	P	T	N	G	Y	T	R	Y	R	Y	S	S	S	S	Y	Y	Y	S	S	P	M	D

Both VH clones were produced in large scale and purified through protein A columns, confirming the conservation of the VH fold. Binding to Grb2 was studied by biolayer interferometry, against the target immobilized through a biotin-PEG₄-linker (**Figure 51**). VH.aGrb2.1 has an affinity of $K_D \sim 56 \mu M$, while VH.aGrb2.2 has an affinity of $K_D \sim 36 \mu M$. The specificity of these clones was tested by performing binding kinetics against an unrelated target, MBP, which

was immobilized in a similar manner as Grb2. Both clones present some unspecific binding to MBP, with affinities in the mid-micromolar to low-millimolar range (the fits of these curves are unreliable, as it can be appreciated by the standard error in k_a). Control experiments performed with VH-36-Ser (the template used for library construction) showed no binding for both Grb2 and MBP at the highest concentration tested (50 μ M), confirming that the binding obtained for the selected clones is due to the interactions through CDR-H3 and/or CDR-H2. Interestingly, VH.aGrb2.2 has mutations only in CDR-H3, but shows similar affinity as compared to VH.aGrb2.1. This indicates that it is possible to generate affinity and specificity through a single CDR.

Figure 51: Biolayer interferometry characterization of interactions between Grb2-SH2 domain and different VH domains selected by interaction-CoFi.

The target Grb2 was immobilized through a PEG₄-biotin linker onto streptavidin-coated sensors, and binding kinetics analyzed **(A)**. Unspecific binding was analyzed using sensors coated with MBP-PEG₄-biotin **(B)**. VH-36-Ser was also analyzed as a negative control **(C)**.

Experimental measurements are shown in black (except for the negative control against MBP, which is shown in orange), while kinetic regression curves are shown in blue. Binding constants are indicated.

DISCUSSION

1. Identification of stabilized protein variants by Hot-Colony Filtration

In this report, I describe a new method to study protein thermal stability of over-expressed proteins in cells, and its use for the selection of stabilized protein variants using a directed evolution approach. Our method is rapid, robust, and efficient in detecting stabilized proteins. Furthermore, it is effective over a wide range of screening temperatures and is generic as it is potentially applicable to most soluble proteins expressed in a colony-forming microorganism.

Our method has several advantages over existing activity-independent thermostability screens available, e.g. Proside (68), THR (69) and the Tripartite Fusion System (70), thanks to its direct readout and broad applicability. Proside (68) is a phage-based method that relies on an indirect read-out of protein folding through protease resistance and phage infectivity, rather than thermostability. Moreover, the need to display the protein on the bacteriophage's surface considerably limits the number of targets amenable to the technique. THR (69) is based on the expression of a thermostable variant of kanamycin-acetyltransferase fused to the target gene in *T. thermophilus*. However, the host used is not a validated recombinant protein expression system, and the screening temperature window is narrow (60°C – 75°C). The Tripartite Fusion System (70) is based on a split β -lactamase containing the target gene inserted between the two fragments and expressed in *E. coli*. It is not a thermostability screen but rather a “foldability” and protease resistance screen without temperature challenge. Furthermore it requires the correct insertion of the target gene in between the two fragments of the reporter. Moreover none of these methods have become the standard way of generating more stable variants of a target protein.

Our method is a robust biophysical assay allowing an easy and accurate determination of aggregation temperatures of proteins in complex cellular environments, without biases introduced by reporter fusion partners or other indirect readouts. Interestingly, the temperature of cellular aggregation T_{cagg}

obtained at the colony level correlates well with the melting temperature T_m acquired by differential scanning fluorimetry, which requires purified samples. This suggests that Hot-CoFi monitors the unfolding and subsequent aggregation of proteins, and therefore is a biophysical method. This opens new exciting possibilities for *in vivo* biophysical characterization and evolution of proteins, specially relevant for proteins of unknown function and/or recalcitrant to purification, or proteins that require high stability in complex and crowded environments.

We have demonstrated the broad applicability of our method for the identification of protein variants with improved stability by generating more thermostable variants of 11 diverse proteins, as well as exemplified potential applications. The simplicity of the method allows it to be used by any laboratory, requiring only standard laboratory equipment. Many targets can be evolved in parallel, with a throughput of up to ~100,000 clones per experiment.

From previous work in the literature typical stabilizations obtained with other methods, in the first round of selection, is in the range of 2-5°C (25-28) and the stabilization seen in our experiments is two to three times better than these (median stabilization is 8.5°C). The expanded screening capacity, as well as the direct biophysical read-out of our screen, might explain the increased improvement of thermostability per selection round observed. Moreover, the fact that we can confirm the improved stability of 95% of all characterized clones highlights the excellent stringency and robustness of the method.

We believe that our stability screen is the preferred option when no functional assays are available for a potential target protein. All of the structural biology targets studied here have no activity assay available, and were successfully stabilized with Hot-CoFi. Moreover, our screen allowed the identification of several stabilized viral proteins variants, including the viral capsid protein UL18, suggesting Hot-CoFi as a new tool for recombinant vaccine development.

Our characterization of IL1RA, TEV, DHAA, scFv and VH-domain highlights the great potential of Hot-CoFi to screen for improvements in the stability and biophysical properties of biopharmaceuticals and biotechnological proteins. Resistance to aggregation either by heat-stress or long-term storage at room-

temperature would be of great benefit for both Biotech and Biopharma industries, as it expands the potential industrial applications of enzymes for the former, and the safety and formulation options of biopharmaceuticals for the latter. We believe our method has great potential applicability in both industries for such purposes.

Finally, for many of our targets (NXR1, IL1RA, scFv, VH, TEV, DHAA), Hot-CoFi automatically identified hotspots in the protein where stabilizing mutations are clustered. This seems to indicate that the data generated with this method will also shed light on fundamental aspects of protein stability.

2. Stabilization of proteins by interaction with a partner

2.1. Generation of binders by phage display

Phage display is a widely applied technique for the selection of binders generated *de novo*. Successful library generation is dependent on the choice of protein scaffold to use, library design – i.e. which amino-acids to mutate and the diversity allowed, library size obtained, and its quality (50, 86, 118, 119, 128, 151). In this work, we have successfully generated a phage display library based on the 10th Fibronectin type III domain. The library size of $\sim 8.0 \times 10^{10}$ unique members, of which 47% of clones are functional, in agreement with previously published works (50, 86, 120). Two FN3 loops, BC and FG, were successfully mutated using a skewed codon composition, allowing for the tailored introduction of amino-acid diversity biased for residues commonly seen in naïve antibody repertoires (50). The obtained library followed the library design.

Library panning for five diverse targets yielded binding clones for four of them, with affinities in the low-micromolar range. For the target Grb2, for which we did not manage to select specific binders, new optimizations might be required in the panning selections, or it might be sufficient to screen more clones by single-point ELISA (here we only screened 96 clones). The binding affinity between the selected binders and the targets can be further improved by affinity maturation procedures. To do so, two complementary approaches can be

pursued: the introduction of diversity in the DE loop which was not randomized here, and at the FN3 N-terminus (which is in close proximity with the paratope); and the “soft” or “restricted” randomization of the BC and FG loops. For “soft” randomization, the best-selected clones are chosen as new templates, and each position is randomized with a skewed amino-acid composition so that at each position ~50% remains WT, and 50% can be any other residue (130). This allow for the fine-tuning of the amino-acid composition in the loops, to allow a tighter binding to the target. The “restricted” randomization is similar to the “soft” randomization, but each position is randomized with a restricted set of residues instead, e.g. homolog residues.

The generation of the FN3 library and the selection of binders against a diverse set of targets allowed us to establish in-house all the methods necessary for phage display *de novo* library generation and selection of binders. The current library can now be applied to select for binders of interest for our laboratory. These binders can be used, for example, as crystallization chaperones to stabilize proteins and improve crystallization likelihood, a strategy that has a growing number of success examples (97).

The knowledge and experience acquired in library creation and selection allows us to be well positioned to undertake new strategies and generate new improved libraries. Some new directions that could be pursued are the generation of Serine- and Tyrosine-rich FN3 libraries; the randomization of all BC, DE and FG loops, and eventually the FN3 N-terminus; and the use of other protein scaffolds. Tyr/Ser- rich libraries (also known as minimalistic libraries) have been very successful at generating low-nanomolar binders (91, 120). However, the techniques required to generate them are identical to the ones used in this work, the only difference being the synthesis of special oligonucleotides: the synthesis is performed using trinucleotides, which allows for the generation of very precise codon compositions (91, 120).

In this work, only two loops were randomized, but randomization of at least three or four loops seems to be the norm for libraries containing high-affinity binders (50, 86, 91, 120). The randomization of DE loop in addition to BC and FG might be thus beneficial. The FN3 N-terminus could also be extended and

randomized, providing a prolonged binding surface; indeed, this area is in close proximity to the binding surface composed of BC, DE, and FG loops (**Figure 39B**). Each protein scaffold provides a particular topology of the binding surface available for interactions. The shape of that interaction surface, as well as its plasticity, influences the general capacity of a specific scaffold to bind to a diverse set of targets (155). The generation of additional libraries based on other scaffolds would allow us to have a toolkit of different display scaffolds, increasing the success rate for *de novo* generation of binders.

One potential candidate to be used as a new scaffold for library generation is the superstable autonomous human VH-domain identified by Hot-CoFi, VH-36. VH domains have great potential for the development of new biotechnological and therapeutic applications, as hypothesized by extrapolating properties observed in autonomous camelid V_HH domains. V_HH are well suited to be used as building blocks that can be assembled together to produce multivalent or multi-specific constructs. Their small size provides good tissue penetration and fast clearance from the blood stream, making them suitable for delivery of toxic payloads into solid tumors, or for imaging with radioisotopes (94). Our human VH variant is particularly well suited for library generation using a minimalistic diversity strongly biased for Tyr/Ser (91, 151), as the replacement of the original CDR-H3 of the stabilized clone VH-36 by a deca-serine (VH-36-Ser) led to an increased stabilization. This suggests that a library biased for Tyr/Ser would maintain structural integrity, which would make it highly functional.

2.2. Stabilization of proteins by interaction with a partner

During this study we have demonstrated that protein-protein interactions can lead to protein thermostabilization, and that this phenomenon can be monitored in cells using a modified version of Hot-CoFi, which we named interaction-CoFi (iCoFi). We have successfully detected the stabilization of three diverse protein targets by co-expressing them with known interaction partners, as well as by mixing cultures of target and partner expressed separately. These stabilizations can be monitored accurately by iCoFi. The degree of stabilization varies

significantly from target to target, as we report stabilizations ranging from 0.5°C to 10°C. Further studies are required to determine the influence of the partner stability and the affinity towards the target in the stabilization observed.

The identification of *de novo*-generated binders by iCoFi represents a groundbreaking new application for the method. A VH library was carefully designed to minimize library size while maintaining high functionality. The use of a binary code of Tyr/Ser, the focus on only two loops, CDR-H2 and CDR-H3, and only two lengths for the latter, allowed for a minimal library size. Our proof-of-concept experiments confirmed that iCoFi could identify VH clones binding and stabilizing Grb2 SH2 domain. Two clones were obtained with affinities between 30 μ M and 60 μ M. Interestingly, one of these clones was mutated only in CDR-H3, suggesting it might be possible to obtain binding from libraries focusing on a single loop. These clones have partial specificity towards the target, as they also bind the unrelated target MBP, albeit with significantly lower affinity. Affinity and specificity could be improved during affinity maturation campaigns. One straightforward approach to do so is to use the clone mutated only in CDR-H3, VH.aGrb2.2, as template and generate a new library in which a Tyr/Ser binary code would be applied to CDR-H1 and CDR-H2.

After obtaining binding from the three CDRs, further improvements could be obtained from carefully designed “soft” or “restricted” randomization libraries, in a similar approach as suggested for the FN3 library. In this case, in order to reduce the library size, randomization could be focused only on CDR-H3, and on Ser residues only. Indeed, CDR-H3 dominates most antigen binding sites, and Tyr residues are usually involved in binding, while small residues provide plasticity to the loop (156). In previous studies by Fellouse F.A. *et al.* (91, 151), where different minimalistic Fab libraries were studied, CDR-H3 was consistently allowed higher diversity, and there was a correlation between higher diversity in CDR-H3 and higher affinity in the binders selected. Moreover, structural studies on one of the targets selected, anti-DR5 Fab-YSd1, showed that CDR-H3 contributes to ~40% of surface area buried upon antigen binding, and Tyr residues represent 99% solvent-accessible surface buried upon binding in that loop (151). Zemlin *et al.* (152) studied the amino-acid composition of CDR-H3 in naïve human and murine immunoglobulin databases, and showed a dominant

presence of Tyr (~15-30%), Gly (~15%), and Ser (~10%). One simple strategy for minimalistic affinity maturation would be to generate a “glycine-scan” library, in which Ser residues would be allowed to mutate to Gly with a frequency of 50%. The introduction of glycine at key positions might improve affinity significantly.

Overall, the results obtained for *de novo* generation of binders by iCoFi seem very positive, and further efforts should be carried to optimize this method and render it a competitive alternative or complement to current methods such as phage-display. We envisage that our strategy is particularly useful to identify binders that stabilize the target. These stabilizing binders could be ideal as crystallization chaperones.

CONCLUSION AND FUTURE DIRECTIONS

Throughout this report, we have extensively analyzed the potential applications of Colony Filtration blot variants, Hot-CoFi and iCoFi, to study protein stability in cells, as well as to stabilize proteins either through mutagenesis or by interaction with a partner. Phage display was also explored as an alternative to generate binders, which could potentially stabilize protein targets upon binding.

Hot-CoFi was successfully developed as an easy, rapid, and reliable biophysical screen to study and improve protein stability inside the cell. The method is robust, and is ready to be used extensively by our lab and the overall research community.

Interaction-CoFi has been developed as an effective biophysical method to study protein-protein interactions in crude cell extracts, as well as to identify new binders against any target generated *de novo*. The proof-of-concept results presented in this report show the potential of the method to become a valid alternative or complement to more standard methods to generate binders *de novo*, such as phage display. Due to the potential of this method, focus would be put into demonstrating its broad applicability, and optimizing it for higher throughput and higher affinity of the binders selected.

Phage display has been successfully implemented in our laboratory, as presented in this study. We have developed a highly diverse library based on the 10th fibronectin type III domain, and demonstrated that it contains binders against a broad panel of antigens. The affinities obtained are in the low-micromolar range, and could be significantly improved by modifying the library design. Future efforts should be focused on capitalizing on the experience acquired, and generating an improved library, either on the FN3 scaffold or other scaffolds such as the superstable VH domain developed in this study. A biased library towards Tyr/Ser residues seems like the best alternative.

Future directions should also include the combination of phage display and iCoFi into a unified strategy to generate tight binders with optimal biophysical

properties in a high-throughput manner. The high library sizes obtained for phage display, combined with the stringency and accuracy of our biophysical method in identifying and confirming binders that stabilize the target upon interaction, could be developed into a powerful platform for binder generation *de novo*. These binders could be used for a wide range of applications, including molecular diagnostics, biopharmaceuticals, imaging, and as affinity reagents for research.

BIBLIOGRAPHY

1. A. L. Lehninger, D. L. Nelson, M. M. Cox, *Lehninger principles of biochemistry*. (W.H. Freeman, New York, ed. 6th, 2013).
2. F. U. Hartl, A. Bracher, M. Hayer-Hartl, Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324 (Jul 21, 2011).
3. K. A. Dill, J. L. MacCallum, The protein-folding problem, 50 years on. *Science* **338**, 1042 (Nov 23, 2012).
4. F. U. Hartl, M. Hayer-Hartl, Converging concepts of protein folding in vitro and in vivo. *Nature structural & molecular biology* **16**, 574 (Jun, 2009).
5. S. Kumar, B. Ma, C. J. Tsai, N. Sinha, R. Nussinov, Folding and binding cascades: dynamic landscapes and population shifts. *Protein science : a publication of the Protein Society* **9**, 10 (Jan, 2000).
6. E. Y. Chi, S. Krishnan, T. W. Randolph, J. F. Carpenter, Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm Res* **20**, 1325 (Sep, 2003).
7. S. Frokjaer, D. E. Otzen, Protein drug stability: a formulation challenge. *Nat Rev Drug Discov* **4**, 298 (Apr, 2005).
8. C. Scharnagl, M. Reif, J. Friedrich, Stability of proteins: temperature, pressure and the role of the solvent. *Biochim Biophys Acta* **1749**, 187 (Jun 1, 2005).
9. N. E. Labrou, Random mutagenesis methods for in vitro directed enzyme evolution. *Current protein & peptide science* **11**, 91 (Feb, 2010).
10. M. Lehmann, M. Wyss, Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Current opinion in biotechnology* **12**, 371 (Aug, 2001).
11. P. A. Romero, F. H. Arnold, Exploring protein fitness landscapes by directed evolution. *Nature reviews. Molecular cell biology* **10**, 866 (Dec, 2009).
12. G. Walsh, Biopharmaceutical benchmarks 2010. *Nat Biotechnol* **28**, 917 (Sep, 2010).
13. K. Huus, S. Havelund, H. B. Olsen, M. van de Weert, S. Frokjaer, Thermal dissociation and unfolding of insulin. *Biochemistry* **44**, 11171 (Aug 23, 2005).
14. F. Mulinacci, M. A. Capelle, R. Gurny, A. F. Drake, T. Arvinte, Stability of human growth hormone: influence of methionine oxidation on thermal folding. *J Pharm Sci* **100**, 451 (Feb, 2011).
15. J. T. Radek, F. J. Castellino, A differential scanning calorimetric investigation of the domains of recombinant tissue plasminogen activator. *Arch Biochem Biophys* **267**, 776 (Dec, 1988).
16. A. A. Wakankar *et al.*, Physicochemical stability of the antibody-drug conjugate Trastuzumab-DM1: changes due to modification and conjugation processes. *Bioconjug Chem* **21**, 1588 (Sep 15, 2010).
17. J. Willuda *et al.*, High thermal stability is essential for tumor targeting of antibody fragments: engineering of a humanized anti-epithelial

- glycoprotein-2 (epithelial cell adhesion molecule) single-chain Fv fragment. *Cancer Res* **59**, 5758 (Nov 15, 1999).
18. M. A. Arndt *et al.*, Generation of a highly stable, internalizing anti-CD22 single-chain Fv fragment for targeting non-Hodgkin's lymphoma. *Int J Cancer* **107**, 822 (Dec 10, 2003).
 19. H. Schellekens, Bioequivalence and the immunogenicity of biopharmaceuticals. *Nat Rev Drug Discov* **1**, 457 (Jun, 2002).
 20. A. J. Chirino, A. Mire-Sluis, Characterizing biological products and assessing comparability following manufacturing changes. *Nat Biotechnol* **22**, 1383 (Nov, 2004).
 21. M. Stefani, Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. *Biochim Biophys Acta* **1739**, 5 (Dec 24, 2004).
 22. B. Erickson, Nelson, P. Winters, Perspective on opportunities in industrial biotechnology in renewable chemicals. *Biotechnology journal* **7**, 176 (Feb, 2012).
 23. B. van den Burg, V. G. Eijssink, Selection of mutations for increased protein stability. *Current opinion in biotechnology* **13**, 333 (Aug, 2002).
 24. V. G. Eijssink, S. Gaseidnes, T. V. Borchert, B. van den Burg, Directed evolution of enzyme stability. *Biomolecular engineering* **22**, 21 (Jun, 2005).
 25. S. Ahmad, M. Z. Kamal, R. Sankaranarayanan, N. M. Rao, Thermostable *Bacillus subtilis* lipases: in vitro evolution and structural insight. *J Mol Biol* **381**, 324 (Aug 29, 2008).
 26. C. Dumon *et al.*, Engineering hyperthermostability into a GH11 xylanase is mediated by subtle changes to protein structure. *J Biol Chem* **283**, 22557 (Aug 15, 2008).
 27. L. Giver, A. Gershenson, P. O. Freskgard, F. H. Arnold, Directed evolution of a thermostable esterase. *Proc Natl Acad Sci U S A* **95**, 12809 (Oct 27, 1998).
 28. J. Hao, A. Berry, A thermostable variant of fructose biphosphate aldolase constructed by directed evolution also shows increased stability in organic solvents. *Protein engineering, design & selection : PEDS* **17**, 689 (Sep, 2004).
 29. T. Warne *et al.*, Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* **454**, 486 (Jul 24, 2008).
 30. T. Warne, M. J. Serrano-Vega, C. G. Tate, G. F. Schertler, Development and crystallization of a minimal thermostabilised G protein-coupled receptor. *Protein Expr Purif* **65**, 204 (Jun, 2009).
 31. G. A. Malawski *et al.*, Identifying protein construct variants with increased crystallization propensity--a case study. *Protein science : a publication of the Protein Society* **15**, 2718 (Dec, 2006).
 32. J. D. Bloom, S. T. Labthavikul, C. R. Otey, F. H. Arnold, Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* **103**, 5869 (Apr 11, 2006).
 33. U. B. Ericsson, B. M. Hallberg, G. T. Detitta, N. Dekker, P. Nordlund, Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **357**, 289 (Oct 15, 2006).
 34. A. A. Raibekas *et al.*, Anion binding and controlled aggregation of human interleukin-1 receptor antagonist. *Biochemistry* **44**, 9871 (Jul 26, 2005).

35. S. R. Andrews *et al.*, The use of forced protein evolution to investigate and improve stability of family 10 xylanases. The production of Ca²⁺-independent stable xylanases. *J Biol Chem* **279**, 54369 (Dec 24, 2004).
36. M. T. Reetz, S. Wu, Laboratory evolution of robust and enantioselective Baeyer-Villiger monooxygenases for asymmetric catalysis. *J Am Chem Soc* **131**, 15424 (Oct 28, 2009).
37. T. H. Richardson *et al.*, A novel, high performance enzyme for starch liquefaction. Discovery and optimization of a low pH, thermostable alpha-amylase. *J Biol Chem* **277**, 26501 (Jul 19, 2002).
38. Q. Wang, T. Xia, Enhancement of the activity and alkaline pH stability of *Thermobifida fusca* xylanase A by directed evolution. *Biotechnology letters* **30**, 937 (May, 2008).
39. A. Worn *et al.*, Correlation between in vitro stability and in vivo performance of anti-GCN4 intrabodies as cytoplasmic inhibitors. *J Biol Chem* **275**, 2795 (Jan 28, 2000).
40. M. Zumarraga *et al.*, In vitro evolution of a fungal laccase in high concentrations of organic cosolvents. *Chemistry & biology* **14**, 1052 (Sep, 2007).
41. F. H. Niesen, H. Berglund, M. Vedadi, The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature protocols* **2**, 2212 (2007).
42. M. W. Pantoliano *et al.*, High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **6**, 429 (Dec, 2001).
43. G. A. Senisterra *et al.*, Screening for ligands using a generic and high-throughput light-scattering-based assay. *J Biomol Screen* **11**, 940 (Dec, 2006).
44. C. Mateo *et al.*, Immobilization of enzymes on heterofunctional epoxy supports. *Nature protocols* **2**, 1022 (2007).
45. K. Teilum, J. G. Olsen, B. B. Kragelund, Protein stability, flexibility and function. *Biochim Biophys Acta* **1814**, 969 (Aug, 2011).
46. Z. Reich *et al.*, Stability of empty and peptide-loaded class II major histocompatibility complex molecules at neutral and endosomal pH: comparison to class I proteins. *Proc Natl Acad Sci U S A* **94**, 2495 (Mar 18, 1997).
47. K. Saito, M. Oda, A. Sarai, T. Azuma, H. Kozono, Bound peptide-dependent thermal stability of major histocompatibility complex class II molecule I-Ek. *Biochemistry* **43**, 10186 (Aug 10, 2004).
48. G. Graziano, F. Catanzano, C. Giancola, G. Barone, DSC study of the thermal stability of S-protein and S-peptide/S-protein. *Biochemistry* **35**, 13386 (Oct 15, 1996).
49. C. J. Layton, H. W. Hellinga, Quantitation of protein-protein interactions by thermal stability shift analysis. *Protein science : a publication of the Protein Society* **20**, 1439 (Aug, 2011).
50. C. V. Lee *et al.*, High-affinity human antibodies from phage-displayed synthetic Fab libraries with a single framework scaffold. *J Mol Biol* **340**, 1073 (Jul 23, 2004).
51. S. W. Michnick, S. S. Sidhu, Submitting antibodies to binding arbitration. *Nature chemical biology* **4**, 326 (Jun, 2008).

52. R. Tonikian, Y. Zhang, C. Boone, S. S. Sidhu, Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. *Nature protocols* **2**, 1368 (2007).
53. S. S. Sidhu, H. B. Lowman, B. C. Cunningham, J. A. Wells, Phage display for selection of novel binding peptides. *Methods in enzymology* **328**, 333 (2000).
54. C. Jackel, J. D. Bloom, P. Kast, F. H. Arnold, D. Hilvert, Consensus protein design without phylogenetic bias. *J Mol Biol* **399**, 541 (Jun 18, 2010).
55. S. Ewert, A. Honegger, A. Pluckthun, Stability improvement of antibodies for extracellular and intracellular applications: CDR grafting to stable frameworks and structure-based framework engineering. *Methods* **34**, 184 (Oct, 2004).
56. M. J. Serrano-Vega, F. Magnani, Y. Shibata, C. G. Tate, Conformational thermostabilization of the beta1-adrenergic receptor in a detergent-resistant form. *Proc Natl Acad Sci U S A* **105**, 877 (Jan 22, 2008).
57. C. Neylon, Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res* **32**, 1448 (2004).
58. H. Zhao, W. Zha, In vitro 'sexual' evolution through the PCR-based staggered extension process (StEP). *Nature protocols* **1**, 1865 (2006).
59. S. Harayama, Artificial evolution by DNA shuffling. *Trends in biotechnology* **16**, 76 (Feb, 1998).
60. M. T. Reetz, J. D. Carballeira, Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature protocols* **2**, 891 (2007).
61. B. Shen, PCR approaches to DNA mutagenesis and recombination. An overview. *Methods Mol Biol* **192**, 167 (2002).
62. P. C. Cirino, R. Georgescu, Screening for thermostability. *Methods Mol Biol* **230**, 117 (2003).
63. A. S. Bommarius, J. M. Broering, J. F. Chaparro-Riggers, K. M. Polizzi, High-throughput screening for enhanced protein stability. *Current opinion in biotechnology* **17**, 606 (Dec, 2006).
64. A. Martin, F. X. Schmid, V. Sieber, Proside: a phage-based method for selecting thermostable proteins. *Methods Mol Biol* **230**, 57 (2003).
65. K. Famm, L. Hansen, D. Christ, G. Winter, Thermodynamically stable aggregation-resistant antibody domains through directed evolution. *J Mol Biol* **376**, 926 (Feb 29, 2008).
66. L. Jespers, O. Schon, K. Famm, G. Winter, Aggregation-resistant domain antibodies selected on phage by heat denaturation. *Nat Biotechnol* **22**, 1161 (Sep, 2004).
67. A. Drevelle *et al.*, Disulfide bond substitution by directed evolution in an engineered binding protein. *Chembiochem : a European journal of chemical biology* **10**, 1349 (May 25, 2009).
68. V. Sieber, A. Pluckthun, F. X. Schmid, Selecting proteins with improved stability by a phage-based method. *Nat Biotechnol* **16**, 955 (Oct, 1998).
69. H. Chautard *et al.*, An activity-independent selection system of thermostable protein variants. *Nat Methods* **4**, 919 (Nov, 2007).
70. L. Foit *et al.*, Optimizing protein stability in vivo. *Molecular cell* **36**, 861 (Dec 11, 2009).

71. G. Bruylants, J. Wouters, C. Michaux, Differential scanning calorimetry in life science: thermodynamics, stability, molecular recognition and application in drug design. *Current medicinal chemistry* **12**, 2011 (2005).
72. K. Kuwajima, Circular dichroism. *Methods Mol Biol* **40**, 115 (1995).
73. J. J. Lavinder, S. B. Hari, B. J. Sullivan, T. J. Magliery, High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. *J Am Chem Soc* **131**, 3794 (Mar 25, 2009).
74. M. M. Lopez, G. I. Makhatadze, Differential scanning calorimetry. *Methods Mol Biol* **173**, 113 (2002).
75. R. Ruller, L. Deliberto, T. L. Ferreira, R. J. Ward, Thermostable variants of the recombinant xylanase A from *Bacillus subtilis* produced by directed evolution show reduced heat capacity changes. *Proteins* **70**, 1280 (Mar, 2008).
76. G. A. Senisterra, P. J. Finerty, Jr., High throughput methods of assessing protein stability and aggregation. *Mol Biosyst* **5**, 217 (Mar, 2009).
77. M. Vedadi *et al.*, Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A* **103**, 15835 (Oct 24, 2006).
78. N. E. Chayen, Turning protein crystallisation from an art into a science. *Current opinion in structural biology* **14**, 577 (Oct, 2004).
79. N. E. Chayen, E. Saridakis, Protein crystallization: from purified protein to diffraction-quality crystal. *Nat Methods* **5**, 147 (Feb, 2008).
80. A. Joachimiak, High-throughput crystallography for structural genomics. *Current opinion in structural biology* **19**, 573 (Oct, 2009).
81. G. Rhodes, *Crystallography made crystal clear : a guide for users of macromolecular models*. Complementary science series (Elsevier/Academic Press, Amsterdam ; Boston, ed. 3rd, 2006), pp. xxv, 306 p.
82. E. Abola, P. Kuhn, T. Earnest, R. C. Stevens, Automation of X-ray crystallography. *Nature structural biology* **7 Suppl**, 973 (Nov, 2000).
83. H. R. Hoogenboom, Selecting and screening recombinant antibody libraries. *Nat Biotechnol* **23**, 1105 (Sep, 2005).
84. A. R. Bradbury, S. Sidhu, S. Dubel, J. McCafferty, Beyond natural antibodies: the power of in vitro display technologies. *Nat Biotechnol* **29**, 245 (Mar, 2011).
85. G. P. Smith, Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315 (Jun 14, 1985).
86. S. S. Sidhu *et al.*, Phage-displayed antibody libraries of synthetic heavy chain complementarity determining regions. *J Mol Biol* **338**, 299 (Apr 23, 2004).
87. J. W. Kehoe, B. K. Kay, Filamentous phage display in the new millennium. *Chemical reviews* **105**, 4056 (Nov, 2005).
88. H. R. Hoogenboom, Overview of antibody phage-display technology and its applications. *Methods Mol Biol* **178**, 1 (2002).
89. S. Miersch, S. S. Sidhu, Synthetic antibodies: concepts, potential and practical considerations. *Methods* **57**, 486 (Aug, 2012).
90. T. A. Kunkel, Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci U S A* **82**, 488 (Jan, 1985).

91. F. A. Fellouse *et al.*, High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *J Mol Biol* **373**, 924 (Nov 2, 2007).
92. R. N. Gilbreth, K. Esaki, A. Koide, S. S. Sidhu, S. Koide, A dominant conformational role for amino acid diversity in minimalist protein-protein interfaces. *J Mol Biol* **381**, 407 (Aug 29, 2008).
93. P. Holliger, P. J. Hudson, Engineered antibody fragments and the rise of single domains. *Nat Biotechnol* **23**, 1126 (Sep, 2005).
94. P. A. Barthelémy *et al.*, Comprehensive analysis of the factors contributing to the stability and solubility of autonomous human VH domains. *J Biol Chem* **283**, 3639 (Feb 8, 2008).
95. D. S. Dimitrov, Engineered CH2 domains (nanoantibodies). *mAbs* **1**, 26 (Jan-Feb, 2009).
96. S. D. Nuttall, R. B. Walsh, Display scaffolds: protein engineering for novel therapeutics. *Current opinion in pharmacology* **8**, 609 (Oct, 2008).
97. S. Koide, Engineering of recombinant crystallization chaperones. *Current opinion in structural biology* **19**, 449 (Aug, 2009).
98. R. Karlsson, A. Larsson, Affinity measurement using surface plasmon resonance. *Methods Mol Biol* **248**, 389 (2004).
99. J. Concepcion *et al.*, Label-free detection of biomolecular interactions using BioLayer interferometry for kinetic characterization. *Combinatorial chemistry & high throughput screening* **12**, 791 (Sep, 2009).
100. E. Engvall, P. Perlmann, Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry* **8**, 871 (Sep, 1971).
101. D. J. Reen, Enzyme-linked immunosorbent assay (ELISA). *Methods Mol Biol* **32**, 461 (1994).
102. H. B. Lowman, Phage display of peptide libraries on protein scaffolds. *Methods Mol Biol* **87**, 249 (1998).
103. B. Liedberg, C. Nylander, I. Lunström, Surface plasmon resonance for gas detection and biosensing. *Sensors and Actuators* **4**, 299 (//, 1983).
104. B. Liedberg, C. Nylander, I. Lundstrom, Biosensing with surface plasmon resonance--how it all started. *Biosensors & bioelectronics* **10**, i (Fall, 1995).
105. C. Structural Genomics *et al.*, Protein production and purification. *Nat Methods* **5**, 135 (Feb, 2008).
106. F. R. Schmidt, Recombinant expression systems in the pharmaceutical industry. *Applied microbiology and biotechnology* **65**, 363 (Sep, 2004).
107. J. Koehn, I. Hunt, High-Throughput Protein Production (HTPP): a review of enabling technologies to expedite protein production. *Methods Mol Biol* **498**, 1 (2009).
108. B. Miroux, J. E. Walker, Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**, 289 (Jul 19, 1996).
109. P. Braun, J. LaBaer, High throughput protein production for functional proteomics. *Trends in biotechnology* **21**, 383 (Sep, 2003).
110. S. Graslund *et al.*, The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. *Protein Expr Purif* **58**, 210 (Apr, 2008).

111. S. Kirillova, S. Kumar, O. Carugo, Protein domain boundary predictions: a structural biology perspective. *The open biochemistry journal* **3**, 1 (2009).
112. Y. Peleg, T. Unger, Resolving bottlenecks for recombinant protein expression in E. coli. *Methods Mol Biol* **800**, 173 (2012).
113. C. Aslanidis, P. J. de Jong, Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* **18**, 6069 (Oct 25, 1990).
114. R. Vincentelli *et al.*, High-throughput protein expression screening and purification in Escherichia coli. *Methods* **55**, 65 (Sep, 2011).
115. T. Cornvik *et al.*, An efficient and generic strategy for producing soluble human proteins and domains in E. coli by screening construct libraries. *Proteins* **65**, 266 (Nov 1, 2006).
116. T. Cornvik *et al.*, Colony filtration blot: a new screening method for soluble protein expression in Escherichia coli. *Nat Methods* **2**, 507 (Jul, 2005).
117. S. L. Dahlroth, P. Nordlund, T. Cornvik, Colony filtration blotting for screening soluble expression in Escherichia coli. *Nature protocols* **1**, 253 (2006).
118. A. Koide, C. W. Bailey, X. Huang, S. Koide, The fibronectin type III domain as a scaffold for novel binding proteins. *J Mol Biol* **284**, 1141 (Dec 11, 1998).
119. V. Batori, A. Koide, S. Koide, Exploring the potential of the monobody scaffold: effects of loop elongation on the stability of a fibronectin type III domain. *Protein engineering* **15**, 1015 (Dec, 2002).
120. J. Wojcik *et al.*, A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. *Nature structural & molecular biology* **17**, 519 (Apr, 2010).
121. S. van den Berg, P. A. Lofdahl, T. Hard, H. Berglund, Improved solubility of TEV protease by directed evolution. *J Biotechnol* **121**, 291 (Feb 10, 2006).
122. C. Aslanidis, P. J. de Jong, Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* **18**, 6069 (Oct 25, 1990).
123. T. J. Collins, ImageJ for microscopy. *BioTechniques* **43**, 25 (Jul, 2007).
124. K. Miyazaki, Creating random mutagenesis libraries by megaprimer PCR of whole plasmid (MEGAWHOP). *Methods Mol Biol* **231**, 23 (2003).
125. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760 (Sep 1, 1994).
126. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213 (Feb, 2010).
127. H. Zhao, A pH-indicator-based screen for hydrolytic haloalkane dehalogenase. *Methods Mol Biol* **230**, 213 (2003).
128. A. Koide, S. Koide, Monobodies: antibody mimics based on the scaffold of the fibronectin type III domain. *Methods Mol Biol* **352**, 95 (2007).
129. J. Bostrom, G. Fuh, Design and construction of synthetic phage-displayed Fab libraries. *Methods Mol Biol* **562**, 17 (2009).
130. J. Bostrom, C. V. Lee, L. Haber, G. Fuh, Improving antibody binding affinity and specificity for therapeutic development. *Methods Mol Biol* **525**, 353 (2009).

131. N. Tokuriki, C. J. Oldfield, V. N. Uversky, I. N. Berezovsky, D. S. Tawfik, Do viral proteins possess unique biophysical features? *Trends in biochemical sciences* **34**, 53 (Feb, 2009).
132. S. van den Berg, P. A. Lofdahl, T. Hard, H. Berglund, Improved solubility of TEV protease by directed evolution. *J Biotechnol* **121**, 291 (Feb 10, 2006).
133. M. Pavlova *et al.*, Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nature chemical biology* **5**, 727 (Oct, 2009).
134. P. Zhou, F. Tian, F. Lv, Z. Shang, Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. *Proteins* **76**, 151 (Jul, 2009).
135. K. Colwill, S. Graslund, A roadmap to generate renewable protein binders to the human proteome. *Nat Methods* **8**, 551 (Jul, 2011).
136. A. Worn, A. Pluckthun, Stability engineering of antibody single-chain Fv fragments. *J Mol Biol* **305**, 989 (Feb 2, 2001).
137. I. S. Mian, A. R. Bradwell, A. J. Olson, Structure, function and properties of antibody binding sites. *J Mol Biol* **217**, 133 (Jan 5, 1991).
138. H. K. Binz, M. T. Stumpp, P. Forrer, P. Amstutz, A. Pluckthun, Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J Mol Biol* **332**, 489 (Sep 12, 2003).
139. G. Interlandi, S. K. Wetzel, G. Settanni, A. Pluckthun, A. Caflisch, Characterization and further stabilization of designed ankyrin repeat proteins by combining molecular dynamics simulations and experiments. *J Mol Biol* **375**, 837 (Jan 18, 2008).
140. A. Koide, M. R. Jordan, S. R. Horner, V. Batori, S. Koide, Stabilization of a fibronectin type III domain by the removal of unfavorable electrostatic interactions on the protein surface. *Biochemistry* **40**, 10326 (Aug 28, 2001).
141. R. Tonikian, S. S. Sidhu, Selecting and purifying autonomous human variable heavy (VH) domains. *Methods Mol Biol* **911**, 327 (2012).
142. G. Johnson, T. T. Wu, Kabat database and its applications: 30 years after the first variability plot. *Nucleic Acids Res* **28**, 214 (Jan 1, 2000).
143. R. F. Latypov *et al.*, Biophysical characterization of structural properties and folding of interleukin-1 receptor antagonist. *J Mol Biol* **368**, 1187 (May 11, 2007).
144. B. S. Chang, G. Reeder, J. F. Carpenter, Development of a stable freeze-dried formulation of recombinant human interleukin-1 receptor antagonist. *Pharm Res* **13**, 243 (Feb, 1996).
145. S. Krishnan, A. A. Raibekas, Multistep aggregation pathway of human interleukin-1 receptor antagonist: kinetic, structural, and morphological characterization. *Biophys J* **96**, 199 (Jan, 2009).
146. H. Schreuder *et al.*, A new cytokine-receptor binding mode revealed by the crystal structure of the IL-1 receptor with an antagonist. *Nature* **386**, 194 (Mar 13, 1997).
147. H. A. Schreuder *et al.*, Refined crystal structure of the interleukin-1 receptor antagonist. Presence of a disulfide link and a cis-proline. *Eur J Biochem* **227**, 838 (Feb 1, 1995).
148. J. G. van Leeuwen, H. J. Wijma, R. J. Floor, J. M. van der Laan, D. B. Janssen, Directed evolution strategies for enantiocomplementary haloalkane

- dehalogenases: from chemical waste to enantiopure building blocks. *Chembiochem : a European journal of chemical biology* **13**, 137 (Jan 2, 2012).
149. K. A. Gray *et al.*, Rapid evolution of reversible denaturation and elevated melting temperature in a microbial haloalkane dehalogenase. *Adv. Synth. Catal.* **343**, 607 (July 17, 2001).
 150. T. Koudelakova *et al.*, Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel. *Angewandte Chemie* **52**, 1959 (Feb 11, 2013).
 151. F. A. Fellouse *et al.*, Molecular recognition by a binary code. *J Mol Biol* **348**, 1153 (May 20, 2005).
 152. M. Zemlin *et al.*, Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures. *J Mol Biol* **334**, 733 (Dec 5, 2003).
 153. H. K. Binz *et al.*, High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat Biotechnol* **22**, 575 (May, 2004).
 154. J. Yan, L. Pan, X. Chen, L. Wu, M. Zhang, The structure of the harmonin/sans complex reveals an unexpected interaction mode of the two Usher syndrome proteins. *Proc Natl Acad Sci U S A* **107**, 4040 (Mar 2, 2010).
 155. R. J. Hosse, A. Rothe, B. E. Power, A new generation of protein display scaffolds for molecular recognition. *Protein science : a publication of the Protein Society* **15**, 14 (Jan, 2006).
 156. S. Koide, S. S. Sidhu, The importance of being tyrosine: lessons in molecular recognition from minimalist synthetic binding proteins. *ACS chemical biology* **4**, 325 (May 15, 2009).

SUPPLEMENTARY INFORMATION

Supplementary table 1: Summary of all the variants obtained, their mutations and Tm (or Tagg*)

Target	Mutant name	Selection round	Mutations	Tm / Tagg* (°C)	Standard deviation (°C)
NXR1	WT	-	-	40.27	0.28
	NXR1-1	Round 1	V45M, E91V	66.85	0.03
	NXR1-2	Round 1	E95K, S107F	55.76	0.07
	NXR1-3	Round 1	E91G	49.95	0.28
	NXR1-4	Round 1	E91K	Not scaled up	
UL14	WT	-	-	44.23	0.86
	UL14-1	Round 1	V77D, V81L, A85G	54.53	0.21
	UL14-2	Round 1	F43S, F47Y, I79N, A136S, P216H	Insufficient yield	
UL18*	WT	Round 1	-	54.43*	0.48
	UL18-1	Round 1	S212L	60.65*	0.60
	UL18-2	Round 1	S212L, T247M	61.03*	0.36
ORF49	WT	-	-	48.05	0.35
	ORF49-1	Round 1	K8N, A36T, K186R	57.80	0.23
	ORF49-2	Round 1	L90I, E116Q, A155V, G157D, T239I	51.92	0.26
	ORF49-3	Round 1	Q207H	53.08	0.06
	ORF49-4	Round 1	V218I, C294Y	50.66	0.17
	ORF49-5	Round 1	A155V, D278N, Q302E	51.30	0.32
	ORF49-6	Round 1	K119E, L125V, Q207H	51.82	0.09
	ORF49-7	Round 1	K187E, T193S	57.65	0.31
	ORF49-8	Round 1	V218I, R273S	54.43	0.20
	ORF49-9	Round 1	Q207H, K242I	53.45	0.60
ORF69	WT	-	-	51.05	0.09
	ORF69-1	Round 1	R157K, V179D, A201P	52.38	0.16
	ORF69-2	Round 1	S35L, C141Y	57.11	0.10
	ORF69-3	Round 1	C173Y	Not scaled up	
	ORF69-4	Round 1	C141W	54.85	0.07
ScFv	WT	-	-	68.42	0.10
	ScFv-1	Round 1	W225G	77.96	0.13
	ScFv-2	Round 1	M7K, A226D	71.53	0.03
	ScFv-3	Round 1	T8I, I109V, H227P	71.39	0.16
	ScFv-4	Round 1	M1T, A226D	71.56	0.18
	ScFv-5	Round 1	A226D	71.72	0.11
	ScFv-6	Round 1	I(-8)T S59P	69.94	0.05
	ScFv-7	Round 1	S3P, D4G	69.60	0.10
	ScFv-8	Round 1	S55T, F86S, Q208H	70.67	0.09
VH-domain	WT	-	-	64.87*	0.20
	VH-1	Round 1	R50G	73.35*	0.45
	VH-2	Round 1	L115P	68.84*	0.11
	VH-3	Round 1	H35D, V48I	Two overlapping transitions*	
CH2	WT	-	-	56.12	0.39
	CH2-1**	Round 1	K11E, V49D, L74P	46.58	0.33
	CH2-2**	Round 1	K13N, V31G, G106D, A108V	42.73	0.16
	CH2-3	Round 1	V5D, V29A, V31I, F40V, V70A	Insufficient yield	
	CH2-4	Round 1	K13E	61.75	0.18

	CH2-5	Round 1	V5A, K13E, V24D, Q76H, K91N	Insufficient yield	
	CH2-6	Round 1	V5D, N41D	Insufficient yield	
	CH2-7	Round 1	E34D	58.06	0.15
	CH2-8	Round 1	V5D, R57P, K82I	Insufficient yield	
IL1RA	WT	-	-	56.04	0.51
	IL1RA-1	Round 1	A12T, P50H	62.28	0.13
	IL1RA-2	Round 1	P50H, H54N, F148L	61.93	0.35
	IL1RA-3	Round 1	M-7R, W16R, M65L	61.99	0.33
	IL1RA-4	Round 1	A12V	62.07	0.23
	IL1RA-5	Round 2	A12T, P50H, F148L	64.10	0.04
	IL1RA-6	Round 2	S8N, A12V, P50H, H54N, F148L	64.36	0.07
	IL1RA-7	Round 2	A12V, P50H, H54N, F148L	64.34	0.05
	IL1RA-8	Round 2	A12T, H54N, F148L	65.52	0.02
	IL1RA-9	Round 2	A12T, F148L	62.87	0.04
TEV	WT	-	-	52.14	0.29
	TEV-1	Round 1	D140E	56.36	0.08
	TEV-2	Round 1	E119G, F185L	61.91	0.04
	TEV-3	Round 1	F53L, E119G	61.21	0.10
	TEV-4	Round 1	E119G, E201V	56.55	0.13
	TEV-5	Round 1	L69M, E119G	60.92	0.03
	TEV-6	Round 1	E119G	61.55	0.13
	TEV-7	Round 1	E119G, I179L	60.63	0.06
	TEV-8	Round 1	N25S, L251S	56.82	0.20
	TEV-9	Round 1	Q246R	53.94	0.20
	TEV-10	Round 1	S147T, N189Y, N232D	54.96	0.09
	TEV-11	Round 1	P26T, S147T, N189Y	58.64	0.07
	TEV-12	Round 1	E119G, D161A	59.71	0.09
	TEV-13	Round 1	T43I, S170A	54.69	0.02
	TEV-14	Round 1	H180L, Y261S	57.53	0.17
	TEV-15	Round 1	P108H, S135T, N187K	53.94	0.02
	TEV-16	Round 1	H180Q, F230I, Q239R, K262N	56.08	0.01
	TEV-17	Round 1	Q87L	57.39	0.05
	TEV-18	Round 2	Y8D, N25S, T43I, E119G, V212D, L251S	60.71	0.15
	TEV-19	Round 2	Y24F, E119G, D140E	62.30	0.24
	TEV-20	Round 2	N25S, T84A, R118W, E119G, L247H	63.75	0.24
	TEV-21	Round 2	E119G, S170A, N189I, L251S, K262T	63.21	0.23
	TEV-22**	Round 2	N25S, I179V, L251S	52.60	0.26
	TEV-23	Round 2	E119G, R121L, L168V, S170A, K242R, L251S, P279A	Insufficient yield	
	TEV-24	Round 2	T43I, K80N, E119G, S170A, I176V	62.88	0.08
	TEV-25	Round 2	E119G, S170A, L251S	Insufficient yield	
	TEV-26	Round 2	F53L, E119G, L223M, L251S	61.63	0.09
	TEV-27	Round 2	N25S, E119G, Q206H, L247P, L251S	63.90	0.30
	TEV-28	Round 2	E119G, S170A, T186N, L251S	64.70	0.24
	TEV-29	Round 2	E119G, F230S, L251S	63.47	0.10
	TEV-30	Round 2	N25S, M100L, E119G, L251S	62.36	0.14
	TEV-31	Round 2	N25S, E119G, L251S	61.14	0.12
	TEV-32	Round 2	V76L, E119G, L251S	60.99	0.12
	TEV-33	Round 2	E119G, M137W	61.27	0.11
	TEV-34	Round 2	F104Y, E119G, S170A, L251S	60.93	0.05
	TEV-35	Round 2	L7S, N25S, Q87L, E119G, L251S	67.01	0.09
	TEV-36	Round 2	N25S, E119G, F230L, V241D, L251S	63.18	0.02
	TEV-37	Round 2	E119G, G226D, L251S	62.00	0.03

	TEV-38	Round 2	N25S, E119G, Q163L, S170A, A182V, P240S, L251S	62.44	0.12
	TEV-39**	Round 2	E119G, G162E, S170A	53.06	0.19
DHAA	WT			53.31	0.68
	DHAA-1	Round 1	G171V	Not scaled up	
	DHAA-2	Round 1	F113L	Not scaled up	
	DHAA-3	Round 1	D106E	Not scaled up	
	DHAA-4	Round 1	A63V	Not scaled up	
	DHAA-5	Round 1	E99V, A172V	59.44	0.41
	DHAA-6	Round 1	A172V	58.57	2.94
	DHAA-7	Round 1	G171C, A172V	61.84	0.35
	DHAA-8	Round 2	G171C, A172V, L194I	62.55	0.84
	DHAA-9	Round 2	D106E, G171C, A172V, K263M	Not scaled up	
	DHAA-10	Round 2	A172V, N227D	60.44	0.81
	DHAA-11	Round 2	D106E, G171C, A172V	Not scaled up	
	DHAA-12	Round 2	F113L	Not scaled up	
	DHAA-13	Round 2	D106E	Not scaled up	

* T_{agg} is used since we were unable to obtain a T_m for the protein.

** Mutant has a lower T_m than WT or the parental clone.

Supplementary table 2: Summary of clones obtained for Hot-CoFi screen of human VH domain.

Clone	Round	Mutations	T _{cagg} (°C)
WT	Round 1		49.6
VH-1	Round 1	R50S	N/A
VH-2	Round 1	R50M, A79V	N/A
VH-3	Round 1	H35D	N/A
VH-4	Round 1	A79V	N/A
VH-5	Round 1	R50S, W110R	N/A
VH-6	Round 1	A(-4)T, H35D	N/A
VH-7	Round 1	S97G	N/A
VH-8	Round 1	A79V, A88T	N/A
VH-9	Round 1	H35D, S120T	N/A
VH-10	Round 1	H35D, W110L	N/A
VH-11	Round 1	R50T	N/A
VH-12	Round 1	H35Q	N/A
VH-13	Round 1	Y33H, A79V	N/A
VH-14	Round 1	R50M	N/A
VH-15	Round 1	R50G	N/A
VH-16	Round 1	R50M	N/A
VH-17	Round 1	Y105N	N/A
VH-18	Round 1	S97G	N/A
VH-19	Round 1	R50S	N/A
VH-20	Round 1	R50G	N/A
VH-21	Round 1	Y105H, W110R	N/A
VH-22	Round 1	R50S, A88V	N/A
VH-23	Round 1	NO MUTATION	N/A
VH-24	Round 1	R50M	N/A
VH-25	Round 1	R50M	N/A
VH-26	Round 1	V2A	N/A
VH-27	Round 1	NO MUTATION	N/A
VH-28	Round 1	NO MUTATION	N/A
VH-29	Round 1	A106G	N/A
VH-30	Round 1	S97G, A106P	N/A
VH-31	Round 1	L11V, T54A	N/A
VH-32	Round 1	R50M	N/A
VH-33	Round 2	R50S, A79V, S97G, A106P	72.88
VH-34	Round 2	R50S, A79V, A88T	69.39
VH-35	Round 2	L11V, R50S, T54A, S97G	65.66
VH-36	Round 2	H35D, S97G, A106P, W110R	73.45
VH-37	Round 2	H35D, T54A, Y105H, W110R	70.15
VH-38	Round 2	R50S, S97G, A106P	70
VH-39	Round 2	R50T, A79V	70.64
VH-40	Round 2	H35D, A92G, W110L	70.02
VH-41	Round 2	H35D, Y105N	65.4

VH-42	Round 2	H35D, Y105H, W110R	66.74
VH-43	Round 2	H35D, K76E, A79V, Y105H, W110R	69.74
VH-44	Round 2	R50S, S97G, A106P	65.51
VH-45	Round 2	R50S, Y105H, W110R	68.48
VH-46	Round 2	H35D, T91S, W110L	67.03
VH-47	Round 2	H35D, V64A, Y105H, W110R	64.03
VH-48	Round 2	H35D, S97G, L99W, A106P	69.27
VH-49	Round 2	S85N	53.69
VH-50	Round 2	A23V, A79V, A88T	65.67
VH-51	Round 2	R50T, A79V	72.62
VH-52	Round 2	R50S, S97G	65.4
VH-53	Round 2	R50T, A79V	70.49
VH-54	Round 2	R50M, A79V, A106G	67.37
VH-55	Round 2	R50M, A79V, Y105N	67.96
VH-56	Round 2	R50T, A79V, W110R	70.88
VH-57	Round 2	R50T, Y105H, W110R	67.94
VH-58	Round 2	H35D, Y105N	64.99
VH-59	Round 2	R50M, A79V, Y105N	67.32
VH-60	Round 2	H35D, Y105H, W110R	67.03
VH-61	Round 2	R50S, S97G, A106P	67.53
VH-62	Round 2	H35D, S97G, A106P	68.85
VH-63	Round 2	Y33H, G44D, S97G, A106P	53.18
VH-64	Round 2	R50S, Y105N	67.73
VH-65	Round 2	Y33H, R50S, A79V, A106P	71.59
VH-66	Round 2	H35D, S97G	67.48
VH-67	Round 2	P14S, H35D, W110L	64.08

N/A: not acquired. Round 1 mutants were not characterized by T_{cagg} determination.

AUTHOR'S PUBLICATIONS

I. Asial , Y. X. Cheng, H. Engman, M. Dollhopf, B. Wu, P. Nordlund, T. Cornvik. **Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell.** *Nature Communications* **4**, 2901 (Dec 19, 2013).

I. Asial , Y. X. Cheng, H. Engman, P. Ramani , P. Nordlund, T. Cornvik. **Generation of a thermostable autonomous antibody variable domain by Hot Colony Filtration.** *Manuscript under preparation.*

I. Asial , H. Engman, P. Ramani, H. Bjorkling, P. Nordlund, T. Cornvik. **Monitoring protein-protein interactions inside the cell by interaction-induced thermostabilization.** *Manuscript under preparation.*

I. Asial, H. Engman, P. Ramani , P. Nordlund, T. Cornvik. **Screening antibody libraries in cells through binding-induced target stabilization.** *Manuscript under preparation.*

POSTERS

2010. **Improving protein thermostability using in-vitro evolution in combination with a novel screening method.** *IBC Next Generation Protein Therapeutics*. San Francisco, USA.

2011. **Activity-Independent Thermostability Screening.** *IBC Next Generation Protein Therapeutics*. San Francisco, USA.

2012. **Activity-Independent Thermostability Screening.** *Keystone Symposia meeting on High-Throughput Structural Biology*. Keystone, USA.