LiP-Chip a quick and affordable method for assaying protein thermal stability

Judith Villain

Université de Paris

Université de Paris

École doctorale Frontières de l'Innovation en Recherche et Éducation

(ED 474)

Center for Interdisciplinary Research in Biology

LiP-Chip: A quick and affordable method for assaying protein thermal stability

Par Judith Villain

Thèse de doctorat de biologie moléculaire et structurale et biochimie, biophysique moléculaire

Dirigée par Olivier Rivoire
Et par Kimberly Reynolds
Co-encadrée par Clément Nizak

Présentée et soutenue publiquement le 1er Octobre 2020

Devant un jury composé de :

Stéphanie Descroix, Directeur de recherche, Institut Curie, rapporteur

Kendra Frederick, Assistant professor, UT Southwestern Medical Center, examinateur

Philippe Minard, Professeur, Université Paris Saclay, rapporteur

Fabio Sterpone, Directeur de recherche, Université de Paris, examinateur

LiP-Chip: une méthode rapide et abordable pour mesurer la stabilité thermique des protéines

Comprendre quantitativement les protéines est essentiel à leur ingénierie et à leurs applications. Cependant, l'étude de la stabilité des protéines repose sur des méthodes souvent coûteuses et longues. Nous proposons une nouvelle technique pour estimer la stabilité thermique des protéines : LiP-Chip.

LiP-Chip combine l'approche de la protéolyse limitée (LiP) à différentes températures avec la quantification sur des gels d'électrophorèse de protéines miniaturisés sur puce (Chip). L'étape LiP ne nécessite que des équipements de laboratoire standard et peut facilement être multiplexée et automatisée. Comme preuve de principe de l'intérêt de combiner LiP avec de telles « puces protéiques », nous avons utilisé les puces P80 et le Bioanalyseur 2100 d'Agilent. La technique de LiP-Chip pourrait être encore améliorée grâce au développement de dispositifs microfluidiques adaptés.

Nous montrons que LiP-Chip peut être appliquée à des quantités de protéines inférieures à celles utilisées typiquement par la calorimétrie et le dichroïsme circulaire. Plus important encore, la séparation de taille des échantillons sur puce permet l'utilisation d'échantillons de protéines purifiées simplement via une étiquette poly-histidine sans étape de purification supplémentaire sur les colonnes de chromatographie. LiP-Chip peut ainsi fournir une estimation rapide, abordable et facile à utiliser de la stabilité thermique des protéines.

Dans la première partie nous introduirons de manière générale les différents concepts importants pour l'étude de la stabilité des protéines. Pour commencer nous présenterons le lien entre la stabilité des protéines et leur repliement en structure tri-dimensionnelles. Nous mettrons en

évidence la relation entre la structure des protéines et leurs fonctions, puis nous énoncerons les lois thermodynamiques qui régissent le repliement des protéines pour mener aux mesures quantitatives de la stabilité des protéines. Enfin nous énumèrerons les méthodes qui permettent de suivre le dépliement des protéines.

Dans la seconde partie introductive nous définirons la protéolyse limitée, nous aborderons son histoire puis proposerons de l'appliquer à l'étude de la stabilité des protéines et détaillerons les méthodes quantitatives d'analyses des résultats de la protéolyse limitée.

Cette introduction sera suivie de la présentation détaillée du développent de la méthode LiP-Chip. Nous exposerons les protéines modèles utilisées pour valider la méthode LiP-Chip. Nous détaillerons précisément notre analyse des données en expliquant la provenance de celles-ci puis discuterons les fonctions mathématiques utilisées pour modéliser ces données ainsi que les erreurs standards sur les estimateurs. Enfin nous signalerons l'importance du choix des points de mesures expérimentales pour leur analyse.

Nous retracerons ensuite les étapes qui ont permis la création de la méthode LiP-Chip. Nous fournirons les détails expérimentaux qui ont permis la validation de cette méthode, dans un premier temps sans la puce microfluidique (LiP off Chip) puis avec celle-ci. Nous préciserons notamment la reproductibilité de la méthode LiP-Chip, sa robustesse envers différent ratios enzyme/substrat et aux différents temps d'incubations ainsi que le préchauffage des échantillons. Nous examinerons les effets d'agrégation des protéines et étendrons la méthode pour des échantillons de faible concentration en protéines. Nous recommandons un protocole général pour la LiP-Chip et résumerons toutes les caractéristiques de la méthode.

Dans un ultime chapitre nous explorerons trois applications de la méthode LiP-Chip. Nous adresserons deux types de protéines, un anticorps par l'exemple d'un fragment d'anticorps synthétique (scFv) et un enzyme (la

dihydrofolate réductase). Nous nous intéresserons pour finir aux influences de l'environnement sur la stabilité des protéines.

Nous conclurons ce manuscrit avec les questions ouvertes pour lesquelles Lip-Chip est applicable et proposerons des perspectives à la méthode Lip-Chip.

Mots clefs : Stabilité des protéines, Protéolyse limitée, Lab-on-a-chip, Puces protéiques microfluidiques

LiP-Chip: A quick and affordable method for assaying protein thermal stability

A quantitative understanding of proteins is essential for their engineering and their application. The study of protein stability however, relies mostly on methods that tend to be cost- and time-consuming. We propose a new technique to estimate the thermal stability of proteins: LiP-Chip.

LiP-Chip combines the approach of limited proteolysis (LiP) at different temperatures with the quantification on lab-on-a-chip protein gels (Chip). The LiP step requires only standard laboratory equipment and can easily be multiplexed and automated. As a proof of principle for the advantageous combination of LiP with protein chips, we used the P80 chips in combination with the 2100 Bioanalyzer from Agilent. LiP-Chip could be further improved through the development of tailored microfluidic devices.

We show that LiP-Chip can be applied on protein amounts below the typical range used by stability studies with Differential Scanning Calorimetry and Circular Dichroism. Most importantly, the size separation of the samples on the chip allows the use of protein samples from his-tag purification without additional purification on chromatography columns. LiP-Chip can thus provide a quick, affordable and easy-to-use estimation of protein thermal stability.

Keywords : Protein stability, Limited Proteolysis, Lab-on-a-chip, microfluidic protein gel

Acknowledgements

This work would not have been possible without the help of all the wonderful people around me. First, I want to thank my husband Guillaume for always being by my side and helping me through any situation.

I would like to appreciate the help from my supervisors in defining and guiding the scientific projects, and securing funding. Especially the scientific discussions with Kim, her mentorship, and her extraordinary kindness enriched my studies and my work.

In addition to my supervisors, I wish to thank my TAC tutors Dule and Yann, and my colleagues from both labs. In particular Steven and Mégane, who were always available for discussions and whose help I could count on.

My gratitude extends to my doctoral school, especially David and Elodie, for providing me with the means of returning to France and extending my PhD.

I would like to also extend my thanks to the Kortemme lab and Rangathan lab for sharing their plasmids and proteins with me.

I am grateful to the SMBP lab at ESPCI for welcoming me with open arms and teaching me about mass spectrometry. I also thank the technology platform at IPGG and the Joliot & Vriz team at the CIRB for providing me with help and equipment for conducting my experiments.

Furthermore, I would like to express my appreciation to the administrative and technical staff at the CIRB.

Finally, I wish to thank my family and friends for their support.

Preface

To talk about your science is to talk about yourself. Science is never independent of the scientist - your history, your way of thinking influences the path your science takes. And in return, your scientific project with all its wonders and adversities shapes you as a person. Before I talk about the path that my scientific project has taken over the past 1.5 years, I would like to take some time to share my story with you. I think that it highlights quite well what a PhD can teach us beyond the hard facts and scientific skills - the essential transversal skills that help us on any of our paths in the future.

I started my PhD with an intriguing interdisciplinary and international project between the group of Olivier Rivoire in Paris and the lab of Kimberly Reynolds in Dallas. Both labs are united in their goal to investigate the general patterns underlying cellular systems and individual proteins by using statistical analysis of large-scale sequence data in combination with wet-lab experiments. After having done internships in both labs separately, I was excited to work on a joint project. My initial project revolved around the biological meaning of gene coupling. During the first year, the theoretical work in Paris helped me understand the relationship between the experimental measure of gene essentiality and the statistical measure of conservation. Embarking on a plane to Dallas in summer 2017, I was excited to extend my project by experimental work on the coupling between gene pairs and the coupling between genes and the environment.

After only a few months of work filled with high-throughput bacterial growth experiments, my PhD came to a full stop. On a bright December morning, a car hit me on my way to the lab. What followed was a yearlong struggle to get back on my feet, both medically and scientifically. It was not possible for me to stay in the US and continue my work there. I returned to France, but had lost my project and one year of funding. I never thought that I would need to face the fear of crossing the street, the inability to walk on my own - wondering if I would be able to dance at my wedding - while fighting to continue my PhD.

When I had finally recovered enough to work again, it felt as if I was back to zero. Starting my PhD from scratch with a new project. We went away from the studies of genes and environment coupling and focused on the patterns describing protein stability. I have to admit that it took me some time to mourn the loss of my first project and to accept the importance and wonders of the new one. What kept me going was the perspective to learn new lab techniques and to dive into the disciplines of thermodynamics and biochemistry. Despite the differences between my first and second project when it comes to the approach and the lab techniques, it is fascinating to realize that we still come to the same conclusion. Whether it is the experimental method for determining gene essentiality, the genetic background for cellular functions, or the buffer conditions for protein stability - we cannot fully understand a system independent of its environment.

And just as biological systems are influenced by their environment, science is influenced by the scientist. It felt like a start from zero with the new project, but I was not the same person as when I started my PhD. Throughout my first project, and the obstacles I overcame after the accident, I had learned how to organize my time, to be methodical and resilient, and to find motivation and inspiration in difficult situations. It does not need an accident to learn these things, although it is much easier to see the development when faced with such a sudden and complete interruption. I believe that the PhD provides us with many obstacles to give us opportunities to learn and advance. It is part of the scientific process to encounter problems, to analyze different possibilities, rethink what we know, and to learn and adapt. And as much as science is influenced by the scientist, a scientist is shaped by their science. A PhD does not only give us the opportunity to advance scientifically, but helps us to expand our personal boundaries.

List of abbreviations

This list contains abbreviations that are used in several places throughout this thesis. We have excluded commonly used abbreviations (including common material in molecular biology and biochemical laboratories), abbreviations that are only used once, and abbreviations that are not integral to the understanding of the main concepts of this thesis.

α-Lac	α-Lactalbumin
bNAb	Broadly neutralizing antibody
BSA	Bovine Serum Albumin
CD	Circular Dichroism
CDR	Complementarity determining region
Chip	Microfluidic protein gel
C_H1-3	Constant regions 1 to 3 of an antibody's heavy chain
C_L	Constant region of an antibody's light chain
C_m	Denaturant concentration at the midpoint of the unfolding transition
°C	Degree Celsius
DCA	Direct coupling analysis
Δ	Delta indicates a difference/change between two states of a parameter
ΔG^0	Gibbs free energy under standard conditions
$\Delta G_U^{\ddagger}$	Gibbs free energy of activation of unfolding
ΔH_{cal}	Change in enthalpy experimentally determined with DSC
$\Delta H_U^{\ddagger}$	Enthalpy of activation of unfolding
ΔH_{vH}	Change in van't Hoff enthalpy
$\Delta S_U^{\ddagger}$	Entropy of activation of unfolding
DHFR	Dihydrofolate reductase
DSC	Differential Scanning Calorimetry
ecDHFR	DHFR from *Escherichia coli*
e.g.	For example

E/S	Enzyme/Substrate
Fab	Antigen-binding fragment
Fc	Crystallizable fragment
f_N	Frequency of folding
FR	Framework region
f_U	Frequency of unfolding
$G_{N/U}$	Gibbs free energy of the folded state (N) or unfolded ensemble (U)
$H_{N/U}$	Enthalpy of the folded state (N) or unfolded ensemble (U)
His-tag	Poly-histidine tag
H_{vH}	van't Hoff enthalpy
K	Equilibrium constant
k_B	Boltzmann constant
k_D	Rate constant of digestion
k_N	Rate constant of folding
k_U	Rate constant of unfolding
LC-MS	Liquid chromatography-mass spectrometry
LiP	Limited proteolysis
LiP-Chip	Thermal stability method (Limited proteolysis and protein chips)
LiP-MS	Thermal stability method (Limited proteolysis and LC-MS)
MALDI-TOF	Matrix-assisted laser desorption ionization time-of-flight
MS	Mass spectrometry
m/z	Mass-to-charge ratio
N	Natively folded state of a protein
$[N]_{eq}$	Equilibrium concentration of natively folded protein
P80	Agilent protein chip with mass range from 5 to 80 kDa
PBS	Phosphate buffered saline
PDZ	PSD-95/Drosophila Discs Large/Zona Occludens-I
pI	Isoelectric point
PK	Proteinase K

pK_a	Negative logarithm to the base 10 of the acid dissociation constant
p_N	Probability of a protein to be folded
PSD-95^{PDZ3}	Third PDZ domain of the postsynaptic density protein 95
p_U	Probability of a protein to be unfolded
q-PCR	Quantitative/real-time polymerase chain reaction
R	Gas constant
$S_{N/U}$	Entropy of the folded state (N) or unfolded ensemble (U)
SC	Selection coefficient
scFv	Single-chain variable fragment
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SE	Standard error
T	Temperature
T_m	Melting temperature
$T_{m,app}$	Apparent melting temperature
Trp	Tryptophan
TS	Transition state
TST	Transition state theory
U	Unfolded ensemble of a protein
$[U]_{eq}$	Equilibrium concentration of unfolded protein
V_H	Variable region of an antibody's heavy chain
V_L	Variable region of an antibody's light chain

Table of Contents

Introduction

Protein folding and stability

Fold and function

Nature provides a vast variety of proteins with very diverse functions. The Human Proteome Map reports more than 30 000 experimentally identified proteins in the human proteome (1,2). All these diverse proteins are on their most basic level chains of amino acids. It is the folding into specific conformations that enables those chains of amino acids to perform their function (3). A disruption of the native fold can lead to the loss of protein function. The ability of a protein to stay in its native fold and not unfold or misfold is referred to as protein stability. A protein's stability is a highly investigated topic in medicine and pharmacology. Alzheimer's disease for example is partly attributed to the aggregation of misfolded β-amyloid peptides of the amyloid precursor protein (4). Aggregates form when unfolded or misfolded proteins assemble into larger structures. These structures range from disordered (amorphous) to ordered, such as amyloid fibrils, and can be highly stable (5,6). Apart from its amino acid sequence, a protein's stability is strongly influenced by its environment (7). A change in pH, temperature, or co-solvent concentration among others, can significantly change protein stability and increase a protein's aggregation propensity (6,8). One example from pharmaceutical sciences is the filamentous aggregation of insulin at low pH (6). Determining a formulation for therapeutic proteins that provides sufficient protein stability is a major challenge in pharmacology (8).

The interest in protein stability is not limited to the medical and pharmaceutical sectors. In many industrial applications, proteins have to function under harsh conditions for an extended amount of time. An important branch in food industry is the largescale saccharification of starch. This process converts starch into sugars with the use of specific amylolytic enzymes called amylases (9). The process is often employed at elevated temperatures (e.g. 60°C - 90°C) and varying pH levels. Most naturally occurring proteins are only marginally stable and cannot keep their native fold under such demanding conditions (3), resulting in a loss of function. Proteins specifically designed or evolved for high stability are less prone to unfold or misfold under those conditions and can function where natural proteins cannot. In the example of the conversion of starch to

sugars, different design strategies have been employed to produce amylases with higher thermal stability (10).

The above examples from medicine, pharmacology, and food industry show that proteins and their stability and function are important in many different areas. A quantitative understanding of proteins is integral for their engineering and their application. With the vast amount of sequence data available, a fundamental question in protein science is the relationship between a protein's sequence and its properties. Studies attempting to answer this question often require large experimental data sets and statistical inference (11–14). For the analysis of functional properties of proteins, a wide variety of high-throughput methods exist: From phage-display for protein binding to droplet-microfluidics for enzymatic activity. The study of protein stability however, relies mostly on methods that tend to be cost- and time-consuming. The importance of protein stability for a wide range of applications highlights the need for strategies to accurately measure a protein's stability under different environmental conditions. The goal of this thesis was to propose an affordable, quick, and easy-to-use method for the estimation of protein thermal stability. Our LiP-Chip method does not require elaborate purification steps and will enable detailed studies of the natural design principles underlying stability, as well as rapid characterization of designed or evolved proteins under different environmental conditions.

Thermodynamics of protein (un)folding

Gibbs free energy

Folding, unfolding, and misfolding of proteins are a dynamic process that is guided by the energies of the different conformations (3). The change in Gibbs free energy (ΔG) between the folded, misfolded, and unfolded states determines how stable a protein is. The larger the gap between the native fold and the other conformations, the more energy input is required for the protein to unfold or misfold, and the more stable it is. The Gibbs free energy change between two states is directly linked to their difference in enthalpy (ΔH) and entropy (ΔS), as well as the temperature (T) of the system:

$$\Delta G = \Delta H - T * \Delta S \qquad \text{(Eq. 1)}$$

ΔG is an indicator for the spontaneity of a process to occur, given constant pressure and constant temperature. If ΔG is below zero, the change from one state to the other will occur

spontaneously. At ΔG equal to zero no net change is observed. A positive ΔG indicates that energy is required in order for a change from one state to the other to happen. For proteins, the unfolding process at moderate temperatures is not spontaneous, with a ΔG_U above zero. Hence, at moderate temperatures, proteins in a solution will mostly be in their native fold.

The ΔH term in Equation 1 describes the heat transferred when changing from one state to another. A positive ΔH means that heat energy is required for the process. Enthalpy (H) is determined by non-covalent atomic interactions, such as Van der Waals and Coulomb interactions. Typically, a protein in native conformation has lower enthalpy than an unfolded protein due to the various interactions in the native fold. Therefore, ΔH_U of protein unfolding is positive, resulting in a positive ΔG_U at moderate temperatures.

With increasing temperature, the entropy term becomes more and more important for the value of ΔG. Entropy (S) refers to the microstates a system in a certain state can occupy. An unfolded protein can spatially arrange itself in many more ways than a natively folded protein. Therefore, ΔS_U between the folded and the unfolded conformation is positive. This has brought up the question why proteins fold. To answer this question, we have to consider the interaction of the protein with the solvent. This reduces the entropy loss upon protein folding due to the hydrophobic effect (15). The hydrophobic effect is largely attributed to the hydrophobic amino acids. In a folded protein, these amino acids are often buried inside the core of the protein, well protected from the polar solvent. In unfolded proteins on the other hand, hydrophobic amino acids are exposed to the solvent. This leads to the formation of hydration shells of ordered water molecules around the hydrophobic amino acids. Ordering of the solvent reduces the entropy of the solvent, resulting in a smaller difference between the entropy of the folded and unfolded conformation when considering the protein and the solvent together.

Equation 1 shows that the influence of ΔS on ΔG depends on the temperature. At a certain temperature, ΔH and $T * \Delta S$ have the same value, resulting in a ΔG of zero. This temperature is called melting temperature (T_m). When analyzing protein unfolding, T_m indicates the temperature at which the Gibbs free energy of the folded conformation is equal to the Gibbs free energy of the unfolded conformation. At temperatures above the T_m, the entropy term becomes larger than the enthalpy term and ΔG_U becomes negative. Therefore, at temperatures above T_m the unfolding of a protein is spontaneous.

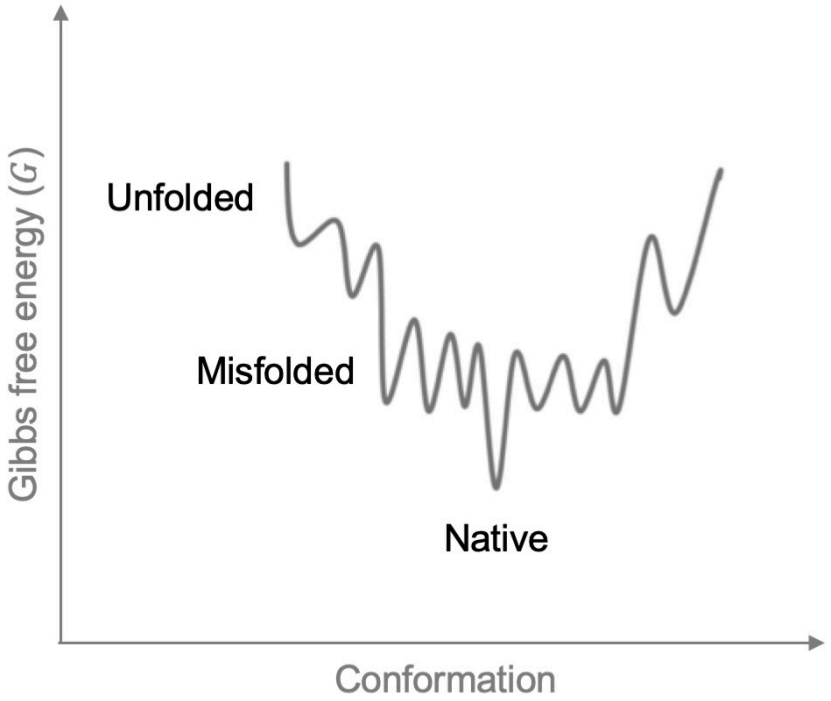


Figure 1. Conformational energy landscape.

Proteins can occupy different conformations (unfolded, misfolded, native) that have different Gibbs free energies (G). The image shows a schematic of an energy landscape of a marginally stable protein at constant temperature and pressure. The energy of the unfolded conformation is higher than the energy of the misfolded or native conformations. Proteins with this energy landscape would therefore most likely be natively folded. [Source: modified from (3)]

Boltzmann distribution

We have introduced the T_m as the temperature at which the Gibbs free energy of the unfolded state is equal to the Gibbs free energy of the folded state. Another interpretation of the T_m is the temperature at which the probability of a protein to be folded is equal to the probability of a protein to be unfolded. The Boltzmann distribution allows us to calculate the probability of a system to be in a specific state based on the energy of that state and the energies of all possible states of the system. This probability is not only applicable at T_m, but can be calculated at any temperature. The probabilities of a protein to be folded (p_N) or unfolded (p_U) are therefore given by the following equations:

$$p_U = \frac{e^{-G_U/k_B*T}}{\sum_i e^{-G_i/k_B*T}} \quad \text{(Eq. 2)}$$

$$p_N = \frac{e^{-G_N/k_B*T}}{\sum_i e^{-G_i/k_B*T}} \quad \text{(Eq. 3)}$$

with the Boltzmann constant k_B and temperature T in Kelvin. G_U and G_N are the Gibbs free energy of the unfolded state and the folded state, respectively. The sum over i indicates the sum over all possible states the protein can occupy (e.g. folded, unfolded, and misfolded states) with their respective Gibbs free energies G_i.

Let us now consider the ratio between p_U and p_N:

$$\frac{p_U}{p_N} = e^{-(G_U-G_N)/k_B*T} = e^{-\Delta G_U/k_B*T} \quad \text{(Eq. 4)}$$

This provides a new formula for the calculation of the change in Gibbs free energy of protein unfolding (ΔG_U):

$$\Delta G_U = -k_B * T \ln \frac{p_U}{p_N} \quad \text{(Eq. 5)}$$

Equation 5 allows us to calculate ΔG_U without the knowledge of ΔH_U and ΔS_U. All we need to know is the ratio between p_U and p_N at temperature T. On the other hand, we can confirm with Equation 5 our claim that the probabilities of being folded and unfolded are equal at T_m. We know that ΔG_U is zero at T_m. Thus, unless T_m itself is zero, p_U/p_N has to be one.

In the next section we will show how we can relate the ratio between the probabilities to the equilibrium constant (K).

Equilibrium constant

For simplicity, we assume that the unfolding of a protein follows a two-state model. That means that we disregard any misfolded or intermediate conformations and consider only the change between the natively folded (N) and the unfolded (U) states. It has been shown that this assumption holds well for many proteins (16–18).

A two-state model of protein unfolding is the following:

$$N \underset{k_N}{\overset{k_U}{\rightleftharpoons}} U \qquad \text{(Model 1)}$$

with N the natively folded protein and U the unfolded protein. The rate constants of unfolding and folding are given by k_U and k_N, respectively. The equilibrium constant (K) is defined as the ratio between k_U and k_N.

One advantage of the two-state model of unfolding is that we can determine the amount of protein in one state by measuring the amount of protein in the other state, given the total amount of protein. Knowing the amount of protein in each state allows us to calculate frequencies for unfolding (f_U) and folding (f_N), which approach the probabilities p_U and p_N if the number of observed proteins is large enough.

At thermodynamic equilibrium, the net change between the amount of folded and unfolded protein is zero and the following relationship holds:

$$\frac{p_U}{p_N} = \frac{f_U}{f_N} = \frac{[U]_{eq}}{[N]_{eq}} = \frac{k_U}{k_N} = K \qquad \text{(Eq. 6)}$$

with $[\,]_{eq}$ referring to the concentration at thermodynamic equilibrium and K the equilibrium constant. Thus, at thermodynamic equilibrium, we can calculate K not only from its definition of the ratio between k_U and k_N, but also from the ratio between the concentration of unfolded ($[U]_{eq}$) and folded ($[N]_{eq}$) protein.

It is important to mention the case of $K = 1$. This is the point at which the energies of the folded and unfolded state are the same and a protein solution will contain approximately half of the proteins folded and half unfolded. This occurs only at the T_m.

Combining Equation 5 and Equation 6 gives us the following equation, that is widely used in the thermodynamic analysis of chemical and biological processes:

$$\Delta G_U = -R * T \ln K \qquad \text{(Eq. 7)}$$

We substituted the Boltzmann constant k_B with the gas constant R to allow us to analyze data in molar concentrations instead of the amount of particles.

With Equation 6 and the assumption of a two-state model, we can simplify Equations 2 and 3:

$$\frac{[U]_{eq}}{[U]_{eq} + [N]_{eq}} = f_U = p_U = \frac{1}{1 + e^{\Delta G_U / R*T}} \qquad \text{(Eq. 8)}$$

$$\frac{[N]_{eq}}{[U]_{eq} + [N]_{eq}} = f_N = p_N = \frac{1}{1 + e^{-\Delta G_U / R*T}} \qquad \text{(Eq. 9)}$$

with $\Delta G_U = G_U - G_N$.

Equations 8 and 9 allow us to link experimentally observable values, such as equilibrium concentrations of unfolded or folded protein, with the thermodynamic theory of protein unfolding. The relationship between ΔG_U and f_N follows a sigmoidal curve with three distinct phases (Figure 2). In the pre-transition phase, the energy difference between the folded and the unfolded conformation is so large that the probability of being unfolded is almost zero. In an imaginary experiment, we would therefore observe mostly folded protein, resulting in an estimated folding frequency f_N close to one. In the transition region, the unfolding probability increases sharply and we observe less and less folded protein. The assumption of the two-state model provides us with the information that the amount of folded protein decreases in the same way as the amount of unfolded protein increases. The same relationship holds for the folding and unfolding probabilities. At ΔG_U of zero, the probabilities of being folded and unfolded are the same, with a value of 0.5. The third phase is termed post-transition phase. Here, most of the proteins in our experiment are unfolded and f_N approaches zero. The same relationship, but inverse, holds for ΔG_U and f_U (Suppl. Figure 1).

In the following section, we will discuss experimental approaches used to quantify protein folding.

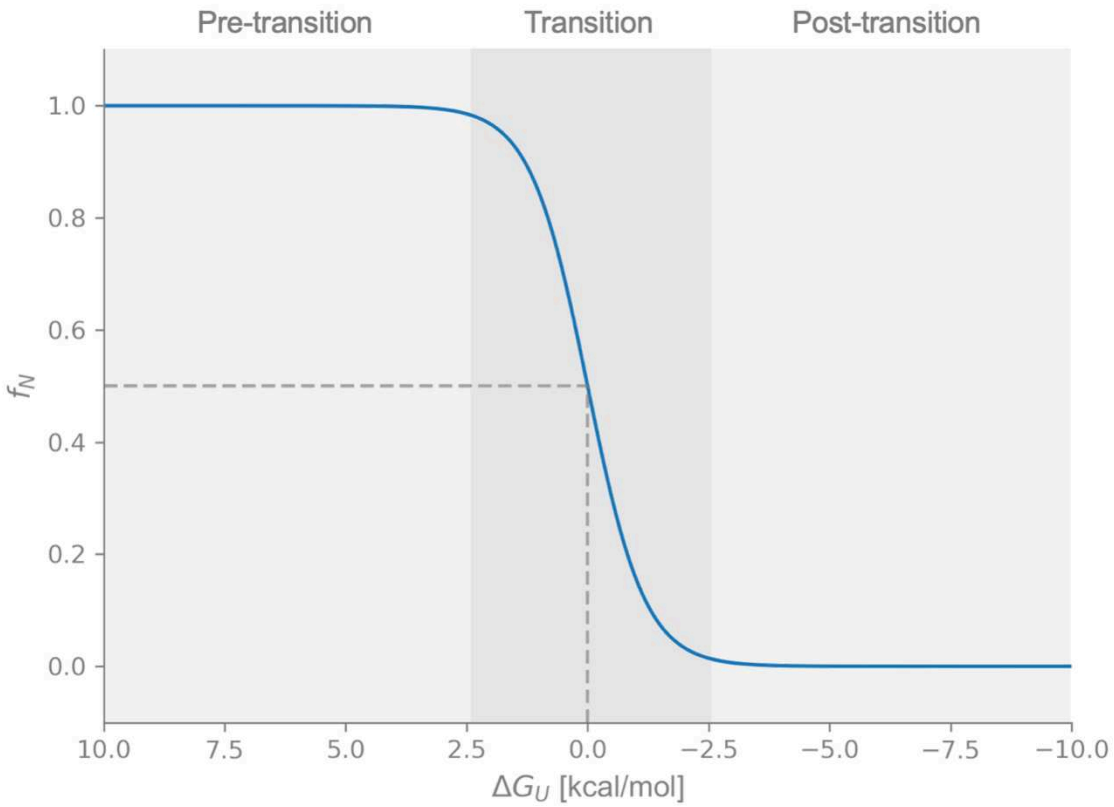


Figure 2. Link between experiment and thermodynamic theory.

The sigmoidal curve was generated with Equation 9 over a ΔG_U range between 10 and -10 kcal/mol. The differently shaded regions correspond to the pre-transition, transition, and post-transition phases (from left to right). The dotted line indicates the point at which the probability of being folded (p_N) or unfolded (p_U) are the same, $f_N = f_U = 0.5$, and $\Delta G_U = 0$.

Quantifying protein folding

As described above, the stability of a protein refers to its ability to keep its native fold. Quantifying protein stability therefore means quantifying protein folding. This can be achieved with different approaches that differ either in the experiments that are performed or in the parameters that are estimated. A quick comparison of the chemical stability and the thermal stability approach is given in Table 1. We will discuss these approaches, as well as the kinetic stability approach, in more detail in the following subsections.

Table 1. Comparison between chemical stability and thermal stability

	Chemical stability	**Thermal stability**
Estimated parameters	Standard ΔG (ΔG^0), C_m	ΔH, ΔS, ΔG, T_m
Denaturation	Chemical denaturant (e.g. urea)	Temperature denaturation
Advantages	Only affects ΔG, Near equilibrium, No aggregation	Fast & automated, Different ΔH estimates, Aggregation quantification
Disadvantages	Time consuming, Reversibility problems, Highly pure protein	Aggregation problems, Reversibility problems, Temperature effect on buffer, Highly pure protein

Chemical stability

The goal of the chemical stability approach is to estimate the standard ΔG (ΔG^0) of a protein, meaning ΔG under standard conditions (e.g. 25°C, 1 atm, pH 7.0). The term chemical stability relates to the unfolding of the protein with chemical denaturants. The protein of interest is mixed with different denaturant concentrations and the amount of unfolded or folded protein is measured at each denaturant concentration. A widely used denaturant for estimating the chemical stability of proteins is urea. Urea leads to the unfolding of protein by binding to non-polar and polar amino acids, replacing water molecules as preferred solvent, and leading to

hydrophobic hydration (19). It has been shown experimentally that the relationship between ΔG and urea concentration is linear in the transition phase (20,21):

$$\Delta G = \Delta G^0 - m * [\text{urea}] \qquad \text{(Eq. 10)}$$

with ΔG^0 the standard ΔG, m indicating the slope of the linear relationship, and [urea] the urea concentration. This simple relationship between the urea concentration and ΔG is one of the advantages of the chemical stability approach. Chemical denaturation affects only one parameter of Equation 4: the ΔG. The temperature is kept constant. Usually, a temperature of 25°C is chosen in accordance with standard conditions. It is also worth mentioning that the use of chemical denaturant prevents protein aggregation.

Another advantage of chemical stability is that a state close to the thermodynamic equilibrium can be reached by incubating the protein with the denaturant for an extended amount of time. In many studies, the equilibration time is longer than ten hours (22–24). Even though the long equilibration time allows the assumption of thermodynamic equilibrium, it also poses a disadvantage. Chemical stability experiments take much longer than experiments estimating thermal stability, as we will discuss further down. Additionally, not all chemical denaturations are reversible. Reversibility however, is a key aspect of the thermodynamic theory. In order to be able to estimate thermodynamic parameters, the reversibility of chemical protein denaturation has to be shown by gradually diluting out the denaturant, measuring the amount of folded or unfolded protein, and comparing it to the initial results. If reversibility of the unfolding process is not given, only apparent parameters (parameter values under the specific experimental conditions) can be estimated.

Further limitations of the chemical stability approach are possible effects of the denaturant concentration on protease activity, as well as the need for highly pure protein for the experiments.

Let us discuss what an ideal experiment for estimating the chemical stability of a protein could look like. We mix the protein of interest with urea at several different concentrations (covering the transition region, as well as parts of the pre- and post-transition regions (Figure 2)). In order to get close to the thermodynamic equilibrium, we let the protein/urea mixes equilibrate for ten hours. After equilibration, we measure the amount of folded protein (or unfolded protein) at different urea concentrations with a method of our choice (we will discuss widely used methods later on). Given the knowledge of the total amount of protein in each sample, we can calculate

the fraction of folded protein (f_N) at each urea concentration. As discussed above, the addition of urea only affects the value of ΔG. Therefore, chemical stability experiments show a sigmoidal relationship between the urea concentration and the amount of folded protein (Figure 3A), as the relationship between ΔG_U and f_N shown in Figure 2. Assuming a two-state unfolding process and converting the raw data to f_N, we can estimate the equilibrium constant K from these results with $K = f_N/(1 - f_N)$. By using the estimated value for K in Equation 7, we can obtain an estimate of ΔG_U for each urea concentration. The chemical stability parameter that is used for comparing different proteins however, is the standard Gibbs free energy change of protein unfolding (ΔG_U^0). We can estimate ΔG_U^0 in accordance with Equation 10 through linear extrapolation of the ΔG_U values at the different urea concentrations within the transition region (Figure 3B). In addition to the change in Gibbs free energy, chemical stability studies provide the denaturant concentration at the midpoint of the unfolding transition (C_m) (Figure 3A). The C_m indicates the concentration at which ΔG_U is equal to zero and the folded and unfolded conformations are equally probable.

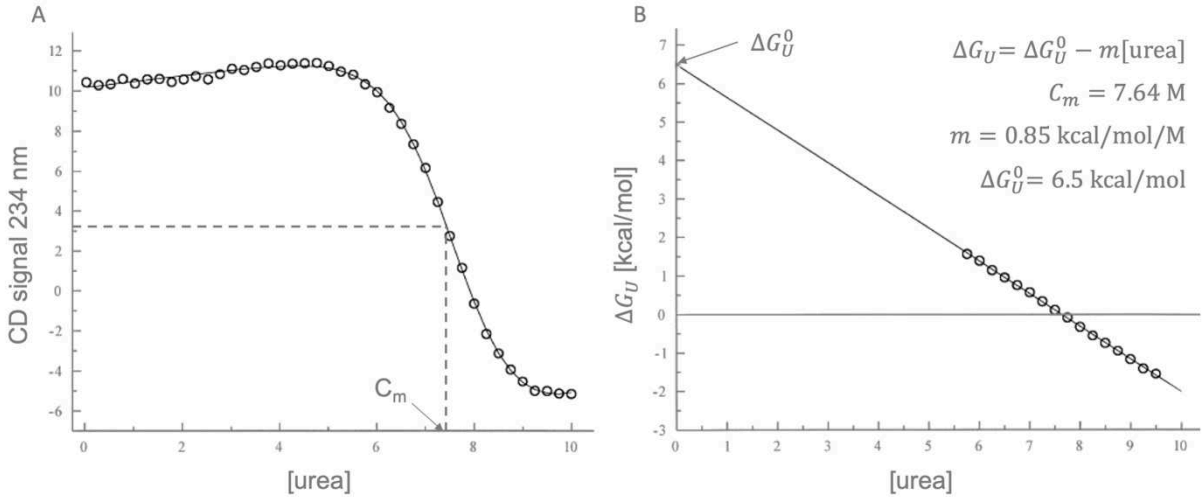


Figure 3. Urea denaturation curve for RNase Sa and linear extrapolation of ΔG_U^0.

The figure is based on a chemical stability study of RNase Sa by Pace *et al.* (21). (A) Circular Dichroism (CD) analysis at 234 nm of RNAse Sa at 25°C, pH 5.0, 30 mM sodium acetate buffer and different urea concentrations. The CD signal at 234 nm is used as a proxy for the amount of folded protein. C_m denotes the urea concentration at which 50% of the proteins are unfolded. (B) Linear extrapolation of the data in the transition phase of A. The standard Gibbs free energy of unfolding (ΔG_U^0) corresponds to the intersection with the y-axis. [Source: modified from (21)]

Thermal stability

Another way of denaturing proteins is by temperature. This is the basis for the thermal stability approach (Figure 4). The thermodynamic parameters that can be estimated through thermal stability experiments are more numerous than with chemical stability experiments, reaching from ΔH, to ΔS, to ΔG. Additionally, the melting temperature T_m can be obtained and offers another way of comparing thermal stabilities of different proteins. As the temperature corresponding to the midpoint of the unfolding transition, T_m indicates the temperature at which ΔG_U is zero and the proteins are as probable to be unfolded as folded.

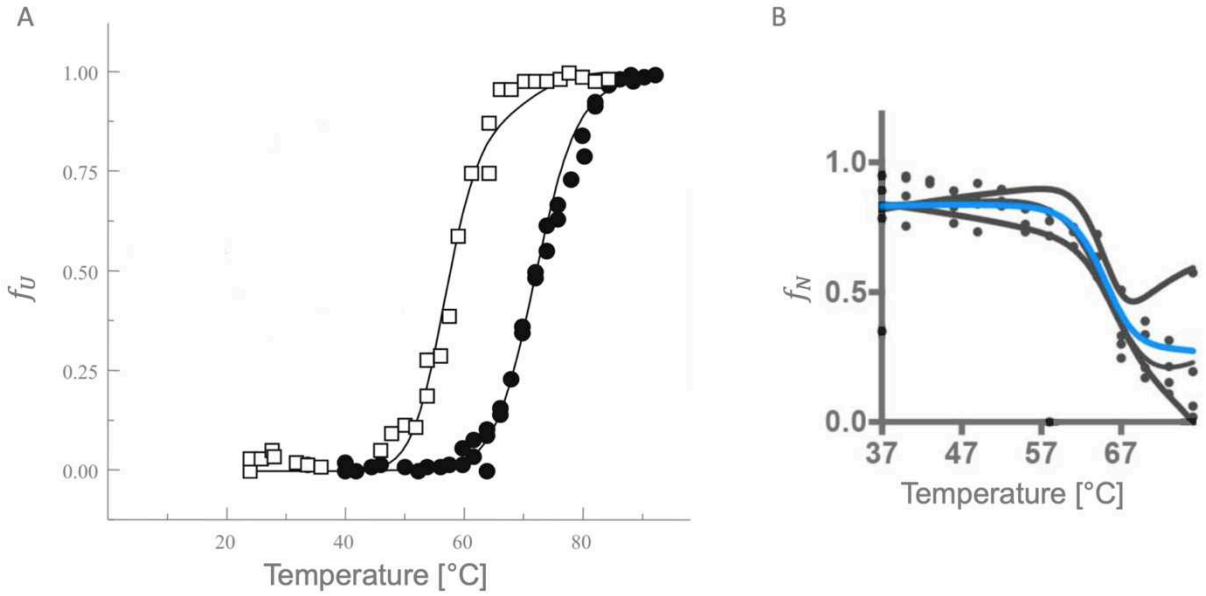


Figure 4. Data from different thermal stability studies.

The above data are chosen examples of thermal stability studies. (A) Change in the fraction of unfolded protein upon thermal denaturation of Ct-Hsp70 measured with CD (black circles) or Tryptophan (Trp) fluorescence (white squares) [Source: modified from (25)]. The difference between the denaturation curves obtained with the two methods was explained with a multi-step unfolding of Ct-Hsp70 and properties of each method for measuring the intermediate unfolding reactions. (B) Thermal unfolding of BSA monitored with LiP-MS [Source: modified from (26)]. The data points represent triplicates corresponding to one peptide. The gray lines are the fits of each replicate. The blue line is the average over all measured peptides and was used to estimate the apparent melting temperature.

One advantage of the thermal stability approach over the chemical stability approach is its fast and often automated experimentation. In machines that can gradually increase the temperature of the sample and readout the amount of unfolded or folded samples, a thermal stability experiment can be accomplished in about one hour. Many studies choose a temperature increment of 0.5 - 1°C/min (27–34). It is assumed that this time is enough for the sample to reach the thermodynamic equilibrium between the folded and unfolded conformations. In order to verify this assumption, the reversibility of the denaturation by temperature has to be shown. For this, the sample is cooled down to its initial temperature (often 25°C) after the first round of denaturation and a second round of denaturation is conducted. If the protein unfolding with temperature is reversible, the results from the second denaturation will correspond to the first one. If that is not the case, the assumption of thermodynamic equilibrium does not hold and only apparent parameters can be estimated.

Another advantage of thermal stability studies lies in the method of Differential Scanning Calorimetry (DSC), which we will discuss in more detail in the next section. DSC provides two different ways of estimating ΔH; one based on experimental results of heat absorption by the sample, the other relying on the thermodynamic relationships as described above. Depending on the nature of the unfolding process, these two estimates of ΔH might differ. A comparison of both can further illuminate cooperative unfolding or the presence of intermediates.

The next topic is at the same time an advantage and a disadvantage in the thermal stability approach: Aggregation. When proteins are denatured at elevated temperatures, they might aggregate. If the aggregation is too dominant, it disturbs the results too much and no thermodynamic parameters can be estimated. On the other hand, aggregation is a widely observed problematic that is worth studying in more detail.

Other disadvantages of the thermal stability approach include possible effects of the temperature on the buffer and the need for highly pure protein, as it is the case in the chemical stability approach.

It is interesting to note that chemical denaturation and thermal denaturation does not necessarily result in the same unfolded conformations (35). This can lead to slightly different estimates of the thermodynamic parameters. Furthermore, some proteins that do not show reversible unfolding in chemical stability studies might do so in thermal stability studies, and vice versa.

An ideal experiment for estimating the thermal stability of a protein might consist of a protein sample that is gradually heated up from a temperature far below the T_m to a temperature above it. The amount of natively folded (or unfolded) protein is measured for each temperature step. This corresponds to thermal stability experiments with several different measurement methods (apart from DSC) that we will introduce in more detail in the section 'Methods for monitoring protein unfolding'. The change in temperature affects the fraction of folded protein (f_N) not only through ΔG_U (see Equation 1 for the link between ΔG and temperature (T)), but also through the temperature term of the Boltzmann distribution (Equation 9). Nevertheless, the relationship between f_N and temperature follows a sigmoidal curve (Figure 5A). The same holds for f_U (Suppl. Figure 2). In the same way as we can estimate ΔG_U with the chemical stability approach, we can estimate it from our imaginary thermal stability experiment by deducing K from f_N and using Equation 7. The midpoint of the transition phase indicates T_m. In addition to ΔG_U and T_m, thermal stability experiments allow us to estimate ΔH and ΔS of protein unfolding. The enthalpy is here referred to as van't Hoff enthalpy (H_{vH}), based on the equation proposed by Jacobus Henricus van 't Hoff, that relates ΔH to K. The van't Hoff equation (Equation 12) can be obtained by combining Equation 1 and Equation 7 and differentiation with respect to the temperature (T):

$$\ln K = \frac{-\Delta H}{R * T} + \frac{\Delta S}{R} \qquad \text{(Eq. 11)}$$

$$\Delta H_{vH} = R * T^2 \frac{d \ln K}{dT} \qquad \text{(Eq. 12)}$$

If ΔH and ΔS are constant, Equation 11 describes a linear relationship between the natural logarithm of K and ΔH and ΔS over the inverse of the temperature ($1/T$). Indeed, if we plot this relationship with the data from the imaginary experiment shown in Figure 5A, we obtain a straight line (Figure 5B). The slope of that line corresponds to $-\Delta H/R$, the intercept with the y-axis to $\Delta S/R$, with R the gas constant. Thus, the thermal stability approach allows us to estimate the change in Gibbs free energy of the protein unfolding (ΔG_U), as well as a dissection of ΔG_U into its enthalpic (ΔH_U) and entropic (ΔS_U) terms.

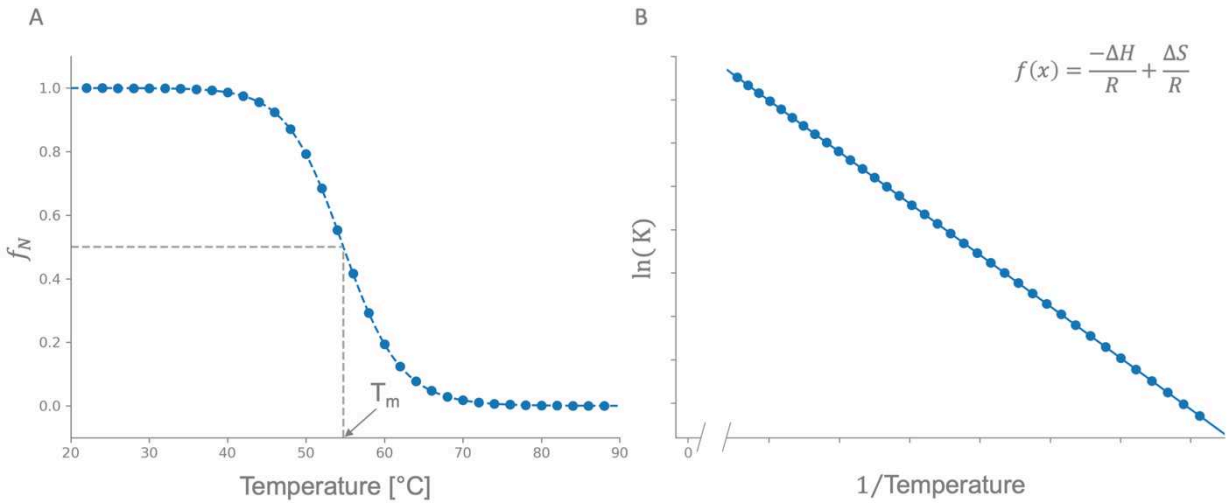


Figure 5. Schematic of data from an ideal thermal stability experiment.

(A) The sigmoidal curve was generated with Equation 9 combined with Equation 1 over increasing temperatures. The T_m is highlighted with the dotted line and represents the temperature at which 50% of the proteins are unfolded and ΔG_U is zero. (B) Shows the relationship between K, ΔH, and ΔS. With the data from (A), the assumption of a two-state model, and Equation 6 we can estimate K. Equation 11 then provides the means for calculating ΔH and ΔS.

Kinetic stability

The chemical stability and thermal stability approaches are focusing on the estimation of the change in Gibbs free energy between the unfolded and the folded conformations (ΔG_U). Important for the unfolding of proteins however, is also the Gibbs free energy of activation ($\Delta G_U^{\ddagger}$). When considering protein unfolding from a purely thermodynamic point of few, only the energy difference between the unfolded (U) and folded (N) conformations are considered. From a kinetics standpoint however, the two conformations are separated by a hypothetical transition state (TS), that imposes an energy barrier (Figure 6). In order for a protein to unfold, it has to overcome the energy barrier from N to U; defined as $\Delta G_U^{\ddagger}$. The estimation of $\Delta G_U^{\ddagger}$ is the goal of the kinetic stability approach. Transition state theory (TST) provides the means to link thermodynamics and kinetics. We will not go into detail about TST, but only provide a broad overview of the general concept applied to protein unfolding. In contrast to the previously explained chemical stability and thermal stability approaches, TST does not require the unfolding reaction to be at thermodynamic equilibrium between N and U. Instead, it assumes that the equilibrium consists between N and TS. With the Eyring equation (Equation 13), we can link the rate constant of unfolding (k_U) to $\Delta G_U^{\ddagger}$:

$$k_U = \frac{\kappa k_B T}{h} e^{-\Delta G_U^{\ddagger}/RT} \qquad \text{(Eq. 13)}$$

with transition coefficient κ (often assumed 1), Boltzmann constant k_B, Planck constant h, gas constant R, and temperature T.

Therefore, we can estimate $\Delta G_U^{\ddagger}$ with experimentally obtained data on k_U. Furthermore, the relationship between Gibbs free energy, entropy, and enthalpy as stated in Equation 1 also holds for the here estimated Gibbs free energy of activation. In a similar way as described for the thermal stability approach with the van't Hoff equation (Equation 12), we can derive a linear relationship providing information on the enthalpy of activation ($\Delta H_U^{\ddagger}$) and the entropy of activation ($\Delta S_U^{\ddagger}$) from k_U (36).

It should be mentioned that TST is not the only theory that can be used to estimate kinetic stability. An empirical law for estimating the activation energy is the Arrhenius equation, that we will not cover here. The activation energy roughly corresponds to the $\Delta H_U^{\ddagger}$ as estimated with TST.

As Figure 6 shows, $\Delta G_U^{\ddagger}$ and ΔG_U are connected. However, $\Delta G_U^{\ddagger}$ provides a different perspective of the unfolding reaction, namely the time of unfolding. The higher the energy of the TS, the longer it will take for proteins to unfold. Thus, ΔG_U is related to the ratio between the conformations at thermodynamic equilibrium, whereas $\Delta G_U^{\ddagger}$ is influenced by the time it will take to reach that equilibrium.

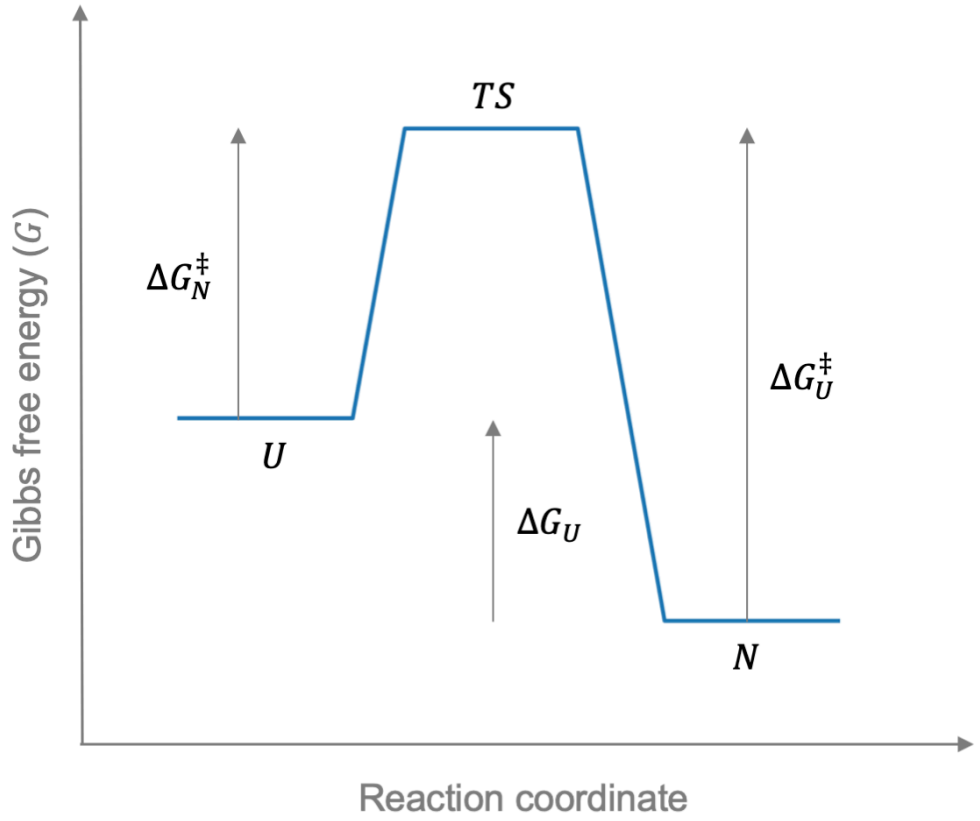


Figure 6. Energy profile of the protein folding/unfolding reaction.

Protein unfolding is a reaction that occurs from the natively folded state (N), over the transition state (TS), to the ensemble of unfolded states (U). From the thermodynamic point of view, we consider the change in Gibbs free energy between N and U (ΔG_U). The kinetic approach considers the change in Gibbs free energy between N and TS ($\Delta G_U^{\ddagger}$) or U and TS ($\Delta G_N^{\ddagger}$).

Methods for monitoring protein unfolding

In the previous section, we have described several different approaches for quantifying protein stability: Chemical stability, thermal stability, and kinetic stability. These approaches refer to general concepts that can be realized with a variety of experimental methods. In this section, we will introduce some of the most widely used of those experimental methods (Table 2).

We can differentiate between three types of methods. The first type are methods that provide a general means for measuring the amount of folded or unfolded protein. These methods can be used for all three approaches of estimating protein stability, described above. The second type of methods are specific for the estimation of protein thermal stability. The third type are high-throughput techniques that rely on display methods.

General methods:

Intrinsic tryptophan fluorescence has been used for a long time to measure the amount of unfolded protein. As the name indicates, this method relies on the natural fluorescence of the amino acid tryptophan (Trp). Trp is hydrophobic and thus, in most natively folded proteins, buried in the core (37). There, it is protected from the polar solvent. Most natural proteins contain only a few Trp amino acids, making it easier to connect a change in Trp fluorescence to protein unfolding. The change in Trp fluorescence upon protein unfolding results from the emission sensitivity of Trp to the local environment (38). The Trp emission displays a red shift in increasingly polar environments. Thus, the change in the maximum Trp emission wavelength under conditions of increasing protein denaturation (chemical or thermal), is a good proxy for the amount of unfolded protein.

In addition to a change in the maximum Trp emission wavelength for unfolded proteins, the fluorescence intensity of Trp is influenced by the protein's conformation. This relationship is less clear than the shift in wavelength. In some proteins, the Trp fluorescence is quenched by neighboring amino acids in the folded protein (37). Upon unfolding, the Trp fluorescence increases. In other proteins, Trp fluorescence is stronger in the natively folded conformation and is quenched by the solvent upon unfolding. In both cases, with increasing or decreasing values, the change in Trp fluorescence intensity between folded and unfolded proteins provides a means for measuring the amount of unfolded protein.

Intrinsic Trp fluorescence can easily be measured with a spectrophotometer and does not require labeling of the samples. Samples can contain protein ranging from pg to mg. However,

these samples need to contain highly pure protein to avoid emission signals from contaminants. In addition to possible signals from contaminants, the background fluorescence of the buffer solution needs to be considered. This background fluorescence might change with denaturant concentration or temperature. Another limitation relevant to this study is that only proteins containing at least one Trp amino acid can be assayed with the intrinsic Trp fluorescence method. Especially for small protein domains, such as PSD-95/Drosophila Discs Large/Zona Occludens-I (PDZ) or antibody fragments, this is not always the case. For larger proteins however, intrinsic Trp fluorescence has been successfully employed to estimate their chemical or thermal stabilities (18,30,34).

Circular Dichroism (CD) is a spectroscopy method based on circularly polarized light. It is widely used to assay structural changes of biomolecules (39,40). For proteins, far-UV CD gives insight into their secondary structure. Different structural elements, such as α-helices, β-sheets, and loops, absorb the circularly polarized light to different extends (41). This results in distinct CD spectra, depending on the presence or absence of such structural elements. Protein unfolding, either with chemical or heat denaturation, undergoes significant structural changes. Therefore, comparing CD spectra of protein solutions under different denaturing conditions provides a measure of unfolded protein in the solution, paving the way for the estimation of protein stability.

CD studies require special CD spectrophotometers that are equipped for the analysis of circularly polarize light in addition to the acquisition of absorbance and fluorescence data of standard spectrophotometers. As the intrinsic Trp fluorescence method, CD samples do not need to be labelled. However, the presence of distinct structural elements, such as α-helices or β-sheets in the natively folded protein, is advantageous. In order to clearly attribute changes in CD spectra to the unfolding of the protein of interest, CD studies require highly pure protein. As mentioned above for intrinsic Trp fluorescence, the CD spectra might also be influenced by the background signal of the buffer. This has to be taken into account during data analysis. CD spectroscopy finds wide applicability in studies estimating chemical or thermal protein stability (18,28,30–32,42). Such studies usually report samples ranging from 20 to 400 µg of protein for the stability estimation.

Limited Proteolysis (LiP) is a method that makes use of the preference of proteases to digest unfolded proteins rather than folded ones. LiP plays an integral role in this project. We will

therefore discuss it in much more detail in the section 'Limited Proteolysis'. Nevertheless, LiP should be mentioned here as one of the methods that can be applied to measure the amount of natively folded protein in chemical stability and thermal stability studies (22,26,43). Samples for LiP do not require labelling and can reach from purified protein to cell lysate. However, LiP alone is only half of the method for measuring the amount of folded protein. It needs to be combined with a quantification technique. Commonly used techniques are based on mass spectrometry (MS) or protein gels (see also 'Quantification of LiP results'). The amount of protein required for an estimation of protein stability with LiP varies with the quantification technique. It can range from pg to mg amounts.

Methods specific to protein thermal stability:

Differential Scanning Calorimetry (DSC) was developed in the 1960s and has since been used extensively to study thermodynamic properties of proteins and other biomolecules (44). DSC monitors the difference in heat capacity between a sample and a buffer reference, while they are gradually heated. The heat capacity of a protein sample changes with the ratio between folded and unfolded proteins. The maximum heat capacity is reached at T_m, where approximately half of the proteins are folded and half are unfolded. As mentioned above, DSC provides two different estimates of ΔH_U. One is based on the total amount of heat needed to unfold all proteins in the sample. This estimate is entirely experimental and is termed ΔH_{cal}. The second estimate uses the concepts from thermodynamics as described in section 'Thermodynamics of protein (un)folding'. It especially links the van't Hoff equation (Equation 12) to the measured quantities and is thus called ΔH_{vH}. Depending on the states involved in the protein unfolding, ΔH_{cal} and ΔH_{vH} might differ. Equality between the two estimates indicates a cooperative two-state unfolding, whereas unfolding follows a multi-state process if ΔH_{cal} exceeds ΔH_{vH}. Apart from the two estimations of enthalpy changes, DSC provides all necessary information to estimate ΔS_U and ΔG_U. For more information on DSC, please refer to the paper by Freire (45).

DSC requires a differential scanning calorimeter, a specialized machine only applicable to DSC experiments. DSC has been used in a large amount of protein thermal stability studies (28,30,32,33). The typical protein amount for DSC studies is 100 to 500 µg of highly pure protein.

Thermal shift assays are a group of high-throughput methods based on fluorescent signals (intrinsic or from dyes) that are used to assay protein stability in a variety of different conditions; from different buffer solutions, to ligand binding or mutations (46). A thermal shift assay that is gaining more and more popularity is the ThermoFluor assay (also referred to as Thermal Denaturation Assay (TDA) or Differential Scanning Fluorimetry (DSF)) (47,48). The ThermoFluor assay employs the fluorescent dye SYPRO Orange which binds to hydrophobic surfaces (48). The fluorescence of unbound SYPRO Orange is quenched by water. When a protein unfolds however, its hydrophobic core is exposed to the solvent. SYPRO Orange can bind and an increase in fluorescence can be detected, allowing the measurement of the amount of unfolded protein. ThermoFluor assays are conducted on q-PCR machines. The 96-well plate setup is especially useful to quickly scan a protein's thermal stability in many different conditions in parallel. Similar to many other methods described above, ThermoFluor assays require the protein samples to be highly pure. In contrast to the more established methods of CD and DSC however, ThermoFluor assays can be conducted with as little as 3 µg of protein, and thus are a good alternative when only small amounts of protein are available. On the other hand, the use of a fluorescent dye that binds to hydrophobic surfaces comes with its limitations. ThermoFluor assays are not applicable to proteins or peptides with exposed hydrophobic surfaces, as might be expected for antibody fragments. Nevertheless, the cost-affective, quick, and high-throughput applicability of ThermoFluor is increasing its popularity for thermal stability studies of proteins.

Display methods

Another type of method for essaying protein stability are based on display methods. These combine the high-throughput ability of phage display or yeast surface display (YSD) with selection of stable proteins based on limited proteolysis, temperature, pH, or the cell's quality control system (49–53). The great advantage of the display methods is their applicability to large libraries without the need for tedious purification. This allows the quick selection and identification of proteins with enhanced stability. This stability however, does not correspond directly to the thermodynamic stability parameters described above. Although a correlation between the stability score from a YSD approach and thermodynamic parameters, such as T_m and ΔG_U has been shown, the display methods do not allow their estimation and one of the above mentioned methods has to be employed (50–52). Furthermore, the source of the selected protein stability is less clear with the display methods. The selected proteins are for example

less prone to aggregation, rather than more thermally stable (51). Thus, many studies refer to the selection of aggregation-resistant or soluble proteins, than thermodynamically stable proteins (51–53). Nevertheless, display methods provide a powerful tool for the selection of general protein stability in large libraries.

All experimental methods mentioned above are successfully used for the estimation of protein stability. They all have their own advantages and disadvantages. Some methods require highly pure protein samples, some are low-throughput, time-consuming, or cost-intensive. Almost all of these methods struggle with protein aggregation. Given the importance of different environmental conditions and aggregation for proteins in all sectors (see 'Fold and function'), there is a need for a quick and cost-effective method with sufficient throughput, that can provide protein stability data under different conditions and identify aggregation. To this means, we are proposing the LiP-Chip method that we will introduce in detail in the next chapter. First however, we will discuss LiP and its existing applications to estimate protein stability in the next section.

Table 2. Methods for monitoring protein stability

Technique	**Stability approach**	**Typical protein amount**	**Sample purity**	**Throughput**
Tryptophan fluorescence	Chemical, thermal, kinetic stability	Highly variable (pg - mg)	High	High
Circular Dichroism (CD)	Chemical, thermal, kinetic stability	20 - 400 µg	High	Low
Limited Proteolysis (LiP)	Chemical, thermal, kinetic stability	Highly variable (pg - mg)	Low	Medium
Differential Scanning Calorimetry (DSC)	Thermal stability	100 - 500 µg	High	Low
ThermoFluor assay	Thermal stability	≥ 3 µg	High	High
LiP-Chip	Thermal stability	10 - 100 µg	Medium	Medium
Display methods	General protein stability (thermodynamic stability, aggregation, folding rates)	Not applicable	No purification	High

Limited Proteolysis

History of LiP

Proteolysis describes the process of peptide bond hydrolysis. It can occur spontaneously on a very long time scale, but is usually catalyzed by proteases. The hydrolysis of only a limited amount of peptide bonds is referred to as limited proteolysis (LiP) or limited hydrolysis (54,55). Originally, LiP described a natural process that governs for example the conversion of inactive precursor proteins to active proteins. In the 1940's and 1950's, several papers discussed that LiP is responsible for the activation of trypsin from trypsinogen and the conversion of fibrinogen to fibrin (55,56).

Nowadays however, LiP refers to a widely used laboratory method. In order to artificially hydrolyze only a limited amount of peptide bonds, the protein of interest is exposed to a low amount of a protease for a short amount of time. The degree of the proteolysis depends strongly on the chosen ratio between protease and protein concentration (Enzyme/Substrate (E/S) ratio) and the time of the exposure (incubation time). The protease will hydrolyze the most flexible digestion sites and not have time to fully digest the protein of interest. In order for hydrolysis to occur, the amino acid region around the digestion site needs to fit into the active site of the protease (57–59). The more flexible a protein region is, the more likely it can assume a conformation that fits into the protease's active site.

The E/S ratio and the incubation time are strongly linked. With a long incubation time even a low E/S ratio can lead to extensive proteolysis, whereas a short incubation time allows the usage of high E/S ratios.

LiP results in distinct digestion patterns for different proteins or different environmental conditions. The applications of LiP are highly diverse and comprise studies of protein folding, protein structure and structural changes, and estimation of thermal and chemical stability (22,26,43,60–66). A study from 1977 by Laemmli and colleagues used LiP for peptide mapping to distinguish different proteins (60). They exposed several proteins for a limited amount of time to different proteases. Each protein/protease combination resulted in a distinct and reproducible peptide pattern. These peptide patterns can be used to differentiate even highly related proteins, such as α- and β-Tubulin.

The same approach of LiP with several proteases can be used not only to distinguish different proteins, but also to analyze a protein's structure (66). As mentioned above, the limited exposure of the protein to the protease leads to the hydrolysis of the most flexible digestion sites. This in turn enables the identification of structural features of the protein. Numerous studies successfully applied LiP to identify protein domains, loops, and disordered regions (58,59,66).

Another important application of LiP is the study of conformational changes (62,64,65). Here, LiP makes use of the fact that structural changes lead to a change in the exposure of the digestion sites. Thus, comparing the LiP digestion patterns of a protein under different conditions gives insight into the protein's conformations under those conditions. This analysis can not only highlight conformational changes due to differing environments, but also help to identify interactions with ligands (43,67).

LiP to estimate protein stability

As we described above, LiP leads to the hydrolysis of peptide bonds only at the most flexible sites in a protein. This mechanism also explains the often observed protease resistance of native proteins (68). Proteins in their native conformation can be rather rigid and thus, are less prone to hydrolysis by proteases. Upon unfolding however, a protein becomes more flexible and more easily hydrolyzed by proteases. Therefore, using LiP under different denaturing conditions enables the estimation of a protein's stability.

The use of LiP for the estimation of a protein's chemical stability is termed Pulse Proteolysis (22). It follows the general structure of chemical stability studies, often using urea as a chemical denaturant and equilibrating the protein-denaturant mix for several hours at 25°C. After equilibration, a broad-specificity protease, such as thermolysin, is added and digests for a short amount of time. For the estimation of the protein's chemical stability, the amount of undigested protein serves as a proxy for the natively folded proteins and is measured over the different denaturant concentrations. The original Pulse Proteolysis paper used sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to quantify the amount of undigested protein (22), but other quantification methods are possible.

LiP can not only be used to estimate the chemical stability of protein, but also its thermal stability. In 2004, Minde *et al.* showed that LiP can be applied to estimate the thermal stability of proteins in purified form and in cell lysate (43). In their Fast-PP method, they used the

thermostable protease thermolysin and exposed the protein of interest to it over different temperatures for approximately 1 min. In order to estimate the apparent melting temperature ($T_{m,app}$ - the melting temperature under the specific experimental conditions) of the protein of interest, they observed the change of undigested protein over increasing temperatures with SDS-PAGE. They successfully showed that a ligand, such as maltose for Maltose Binding Protein or a heme for cytochrome C, can increase the stability of the respective protein. Furthermore, they reported that the estimated $T_{m,app}$ is robust towards different E/S ratios. In contrast to our findings, Minde *et al.* observed a decrease in stability with increasing incubation time.

Leuenberger *et al.* took the application of LiP further than individual proteins and applied a combination of LiP and liquid chromatography-mass spectrometry (LC-MS), termed LiP-MS, to estimate the thermal stability of the proteome of several organisms (26). They were able to estimate the thermal stability of approximately 300 proteins within the cell-lysate and deduced a direct relationship between thermal stability and abundance, as well as an inverse relationship between thermal stability and aggregation.

Quantification of LiP results

The limited proteolysis of the protein of interest in only the first step to characterize protein stability. It needs to be combined with a quantification method. In this section, we will discuss three different quantification methods that have their individual advantages and limitations.

Mass spectrometry

Mass spectrometry (MS) comprises several techniques for identifying the individual components of a sample by their differences in mass-to-charge ratio (m/z) (Figure 7). It is highly sensitive and enables differentiation to an atomic level. MS analyses for proteins are usually achieved by fully digesting the protein with the protease trypsin. The specificity of trypsin (for the amino acids lysine and arginine) leads to predictable peptides. Each peptide will have a different mass and their ions will have specific charges, enabling their separation. The protein peptides after trypsin digestion are ionized in the mass spectrometer (e.g. with a laser) and their abundance is identified as intensity over their mass-to-charge ratio.

Its high resolution in m/z and high sensitivity makes mass spectrometry very powerful. Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry for

example can detect proteins as low as 1 picomole and peptides as low as 100 attomole (69). MS techniques can be used as a high-throughput quantification method for samples containing highly diverse proteins, such as in the cell-wide analysis of protein thermal stability from Leuenberger *et al.* (26). For the estimation of protein thermal stability, the proteome was first exposed to LiP with a broad-specificity protease (proteinase K (PK) or thermolysin) at different temperatures. The proteins unfold at high temperatures depending on their thermal stability. LiP favors the digestion of unfolded protein. The samples were subsequently fully digested with trypsin. For proteins that were mostly natively folded at a certain temperature, the resulting peptides would largely originate from the tryptic digestion. Unfolded proteins however, would produce peptides that show either one or both ends corresponding to a digestion by the broad-specificity protease. Following the depletion of fully tryptic peptides of a protein over increasing temperatures gives insight into the increase of the amount of unfolded protein and thus, allows the estimation of thermal stability. Leuenberger *et al.* used the LiP-MS approach to estimate the thermal stability of more than 8 000 proteins across four species.

LiP-MS is very well applicable to estimate the stability of protein libraries composed of proteins that are highly different. However, if the proteins of interest are very similar to each other (e.g. point mutants), their tryptic peptides are similar and it might not be possible to identify which peptide originated from which protein. Although there are different ways of dealing with this problematic, they all reduce the throughput and increase the preparation time. For this project, we investigated the MALDI-TOF mass spectrometry technique that provides the parallel analysis of many samples through a 96-well format (Figure 7). Due to the incredible sensitivity of mass spectrometry techniques however, the samples need to be specifically prepared and salts removed in a time-consuming dialysis step. We thus set out to identify a method that would allow a quick and cost-effective estimation of protein thermal stability. As a first step, we turned to the widely used protein gel electrophoresis.

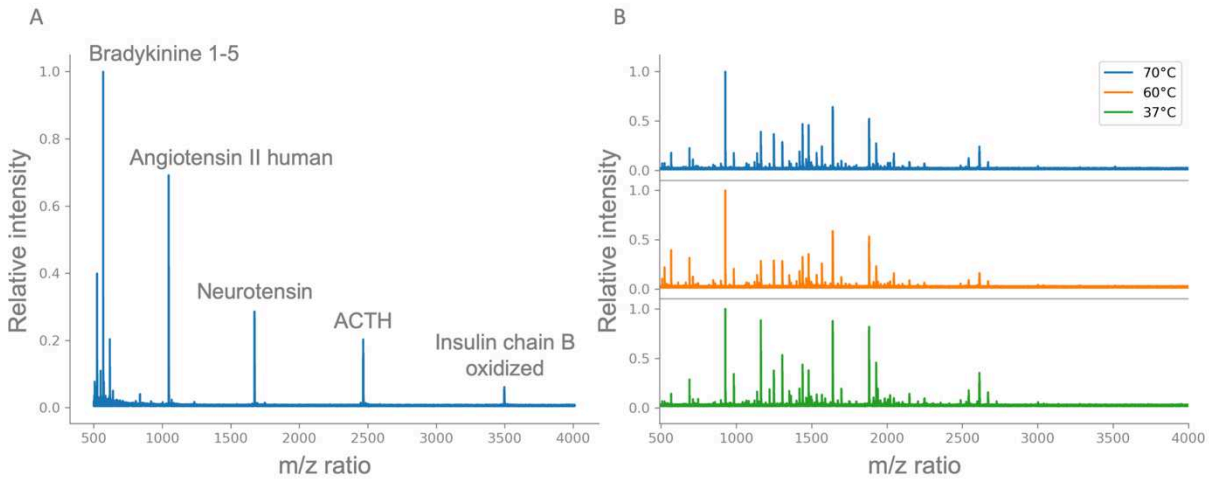


Figure 7. Mass spectra of the calibration standard and LiP of BSA at different temperatures.

The mass spectra were obtained by MALDI-TOF mass spectrometry. The intensities were normalized by the largest intensity of each spectrum. For all experiments, the α-Cyano-4-hydroxycinnamic acid (HCCA) matrix was used. (A) Mass spectrum of the calibration standard used throughout the experiments. This calibration standard contained the following proteins: Bradykinine 1-5 (573.3150 m/z), Angiotensin II human (1046.5420 m/z), Neurotensin (1672.9180 m/z), Adrenocorticotropic homone (ACTH, 2465.1990 m/z), and Insulin chain B oxidized (3494.6510 m/z). (B) Mass spectra of LiP of Bovine Serum Albumin (BSA) at 37°C, 60°C, and 70°C (from bottom to top). LiP was conducted with an E/S ratio of 1/30 and an incubation time of 5 min with preheating. The BSA concentration was 0.1 µg/µL in a reaction volume of 20 µL. For a detailed analysis, individual peptides need to be analyzed.

SDS-PAGE

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) has been used for half a century to analyze proteins (70). The main principle is the separation of proteins and peptides according to their size. In order to analyze the length of the proteins and eliminate the effect of their secondary structure on the migration through the gel, the sample is treated with sodium dodecyl sulfate (SDS). As a detergent, SDS denatures proteins and builds complexes called protein-SDS micelles. These protein-SDS micelles are negatively charged due to the negative charge of SDS. With the application of an electric field, the linearized and charged protein-SDS micelles migrate through the gel. Smaller peptides run through the polyacrylamide gel faster than larger proteins, leading to the size separation. By comparing the sample proteins to a protein ladder containing proteins of known size and concentration, the mass and concentration of the proteins in the sample can be estimated. In this study, we used the

commercially available Novex™ Tris-Glycine Mini Gels (WedgeWell™ Format) from Thermo Fisher Scientific with polyacrylamide concentrations of 16% or gradient gels with 4-12%. For proteins of masses below 15 kDa, we chose the Tricine system from Novex™ to facilitate a sharper resolution in the small size range.

After size separation of the sample, the gel needs to be stained to allow the readout of the separated proteins and peptides. A widely used visual dye for the staining of proteins after electrophoresis is colloidal Coomassie blue (71). It binds non-covalently to proteins through Van der Waals forces and hydrophobic interactions. Pictures of the stained protein gels are subsequently taken and band intensities analyzed with image processing software (see also 'Data source'). We used the PageBlue™ gel staining (Thermo Fisher Scientific), containing the Coomassie blue dye G-250. The sensitivity of this Coomassie blue dye is 5 ng/band with a dynamic range of 5 - 500 ng (Table 3). Throughout our experiments, we estimated the minimal amount of protein for the quantification of LiP results with SDS-PAGE and Coomassie blue staining to be 2.5 µg per sample (based on results from Suppl. Figure 6). Over ten temperatures this amounts to 25 µg for the estimation of thermal stability with LiP. However, the required amount of protein strongly depends on the chosen LiP parameters, as well as the protein gel and staining method, and optimizations are possible.

In combination with the size separation on the SDS-PAGE and the post-electrophoretic gel staining, the analysis of a protein sample takes approximately three hours according to the vendor. In our experience however, the gel staining results in a high background signal unless the protein gel is destained overnight. The SDS-PAGE is parallelizable depending on the size of the gel. For the commercially available protein gels used in this study, the maximum number of sample wells was 15.

SDS-PAGE is a very useful method for the analysis of protein samples. However, it is rather time consuming and the image-based quantification is error-prone. Throughout this project, we used SDS-PAGE mostly for initial assessments of LiP parameters. We will discuss in the following subsection a mostly automated microfluidic approach to SDS-PAGE.

Protein chip

In this section, we will discuss the standard protein chips commercially available from Agilent Technologies (P230 and P80) for the 2100 Bioanalyzer. As a proof of principle, we worked with the P80 chips.

These protein chips are the microfluidic approach to SDS-PAGE (72). They rely on an intertwined set of microfluidic channels pushing the protein sample through a main channel filled with a gel matrix consisting of polydimethylacrylamide (PDMA) and SDS (Figure 8). Before introducing the sample into the protein chip, it is treated with SDS, and if the protein of interest contains disulfide bridges, with Dithiothreitol (DTT). The SDS denatures the protein and forms protein-SDS micelles. The sample is then loaded onto the protein chip, an electric field is applied, and the sample's contents are separated by size in the main channel. The labeling of the proteins and peptides contained in the sample is achieved with a fluorescent dye that binds non-covalently to the protein-SDS micelles and is measured at the end of the main channel (72). One drawback of this labeling technique however, is the binding of the fluorescent dye not only to the protein-SDS complexes, but also to the SDS micelles. Therefore, a destaining step is included before the detection of the fluorescence, by diluting the SDS from the gel matrix. This SDS dilution does not only lead to a reduction of the background fluorescence, but also to a marked increase in the fluorescence corresponding to the sample contents.

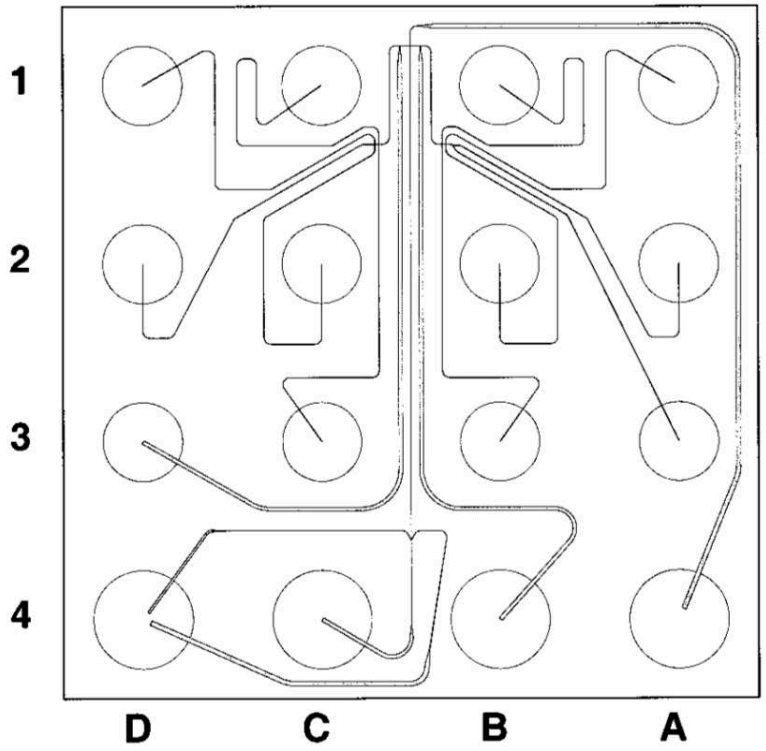


Figure 8. Chip design of the Agilent Technology protein chips.

The wells are shown as circles, the microfluidic channels as lines. Wells A4 and C4 are the separation buffer and waste well, respectively. Wells B4 and D3 are loading wells for the gel matrix. Well D4 is the SDS dilution well. Well D2 is for the protein ladder. The other ten wells are reserved for samples. [Source: (72)]

The protein chips are designed for ten samples (Figure 8). One after the other, the ten samples are separated by size in the main channel of the chip. The P80 chips chosen for this study can assay proteins and peptides in a mass range from 5 – 80 kDa, ideal for experiments with small protein fragments. The sensitivity of the P80 chips is reported as 6 ng/µL for carbonic anhydrase in phosphate buffered saline (PBS) and 15 ng/µL for BSA in PBS (Agilent Technologies). The quantitative range is 60 – 2 000 ng/µL, and the qualitative range 6 - 4 000 ng/µL with a sample volume of only 4 µL. We have further adapted the protocol for low protein concentrations and obtain more than enough sample for the estimation of thermal stability with LiP-Chip from only 10 µg of protein (see 'Low protein concentration'). Table 3 shows a comparison of the classical SDS-PAGE with Coomassie blue staining as used in this study and the P80 protein chip. The sensitivity of the protein chip is more than three times higher than the sensitivity of the SDS-PAGE with Coomassie blue staining and the dynamic range is more than two times larger. However, the most important advantage of the protein chips over the SDS-PAGE with Coomassie blue staining is the vastly reduced analysis time. This is achieved by a decrease of

the staining time from approximately 1 h to 100 ms, as well as a reduction of the dilution time to 0.3 s (72).

Despite its advantages in sensitivity and analysis time, the protein chips are limited to a maximum of ten samples. Furthermore, we will show later that unusual behavior of the background fluorescence can occur and distort the results (see 'Reproducibility of LiP-Chip'). It might be possible to scale up the throughput and reduce the reaction volume by using protein chips with higher sensitivity, microtiter plates, pipetting robots, or more elaborate microfluidic approaches.

Table 3. Comparison of SDS-PAGE (Coomassie blue staining) and P80 protein chip

	SDS-PAGE (Coomassie blue staining)	**P80 protein chip**[1]
Sensitivity	5 ng/band[2]	1.6 ng/peptide[3]
Dynamic range	5 - 500 ng	1.6 - 1 067 ng[4]
Analysis time	90 minutes[5] + 95 minutes[6] to 16 hours[7]	30 minutes
Sample volume	20 µL[8]	4 µL
Number of samples	14[9]	10

1 The specifications from Agilent Technologies were converted to units comparable to the PageBlue™ specifications from Thermo Fisher Scientific

2 Evaluated for β-galactosidase and BSA (Thermo Fisher Scientific)

3 Evaluated for carbonic anhydrase in PBS (Agilent Technologies)

4 Evaluated for carbonic anhydrase and β-lactoglobulin in PBS (Agilent Technologies)

5 Approximate time of sample separation on 4-12% Novex™ Tris-Glycine Mini Gels, WedgeWell™ Format

6 Analysis time for the PageBlue™ gel staining (Thermo Fisher Scientific)

7 Approximate time for sufficient PageBlue™ gel staining observed in this study

8 Recommended sample volume for the Novex™ Tris-Glycine Mini Gels, WedgeWell™ Format

9 Maximal number of samples (without ladder) for the Novex™ Tris-Glycine Mini Gels, WedgeWell™ Format

Developing LiP-Chip

We propose a new technique to estimate the thermal stability of proteins: LiP-Chip.

LiP-Chip combines the approach of limited proteolysis (LiP) at different temperatures with the quantification on lab-on-a-chip protein gels (Chip). The LiP step requires only standard laboratory equipment and can easily be multiplexed and automated. For the sample analysis on the chip we used the 2100 Bioanalyzer from Agilent. This machine is relatively inexpensive and available in many labs and research platforms. In addition to automated electrophoresis of proteins, the 2100 Bioanalyzer is highly versatile and provides chips for DNA and RNA electrophoresis, as well as flow cytometry. We will show in the following sections that LiP-Chip provides an estimate of the apparent melting temperature of proteins. This measure allows us to distinguish between proteins of diverse stability. Furthermore, LiP-Chip can be applied to protein amounts below the typical range used by stability studies with DSC and CD.

Validation proteins

For the development and validation of the LiP-Chip method we chose two well-known and commercially available proteins: Bovine Serum Albumin (BSA) and α-Lactalbumin (α-Lac).

Bovine Serum Albumin:

BSA is a model protein that is widely employed in biological studies and commonly used for the validation of biological techniques and equipment. It consists of α-helices and loops, has a size of approximately 66 kDa, and forms seventeen disulfide bridges in its native conformation (Figure 9). The T_m of BSA is reported between 56°C and 69°C (26,73). This large range of T_m values results not only from the different estimation methods, but also from differences in purity and protein concentration. Nevertheless, we can deduce an appropriate temperature range for our analysis of BSA with LiP-Chip. In order to fit our data to the previously described model of protein unfolding (see 'Thermodynamics of protein (un)folding'), the ideal temperature range for our LiP-Chip analysis should cover several temperatures below the T_m at which the protein is well folded, and several high temperatures at which the protein is unfolded. For BSA, we chose a temperature range between 37°C and 70°C.

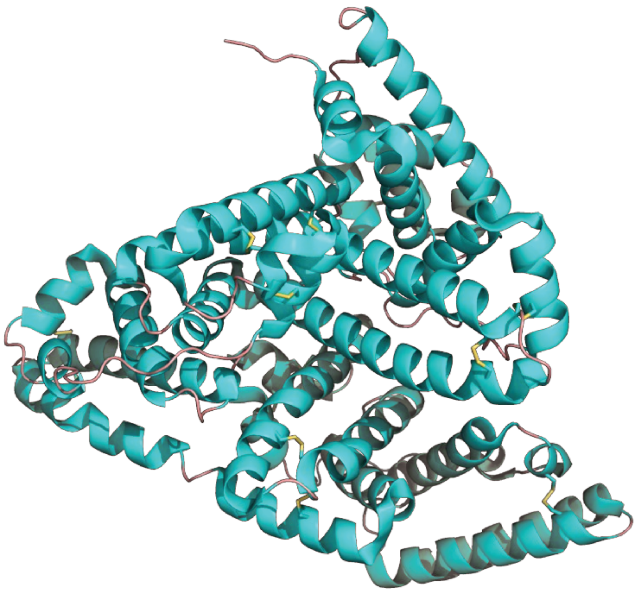


Figure 9. Crystal structure of BSA.

Crystal structure of BSA from PDB entry 4F5S rendered with PyMol. Only Chain A of two chains is displayed. The α-helices are depicted in light blue, β-sheets in pink (not present), and loops are salmon-colored. The seventeen disulfide bridges are shown in yellow.

α-Lactalbumin:

The second validation protein we chose is the whey protein α-Lac (Figure 10). With a size of approximately 14 kDa it is much smaller than BSA and contains only four disulfide bridges. It can assume multiple partially unfolded intermediate states, as well as the molten globule form under several conditions, and is thus widely used as a model for studying protein folding (74). Although α-Lac is highly similar to lysozyme C in sequence and structure, they strongly differ in their biological function (75). The presence of α-Lac is limited to mammalian species in which the protein regulates the synthesis of lactose. The range of stability of α-Lac is even broader than for BSA ranging from 34.5°C to 67.5°C and depends very strongly on environmental conditions and the method of stability estimation (76). This is due to the fact that α-Lac binds metal cations, most efficiently calcium ions (Ca^{2+}). The holo form with bound Ca^{2+} is much more stable than the apo form and exhibits a three-state unfolding (76,77). Other metal cations, such as Mg^{2+}, Na^{+}, and K^{+}, bind α-Lac at much lower rates, leading to only slight increases in the stability (77). Apo- α-Lac on the other hand has been shown to follow a two-state unfolding with an estimated melting temperature around 34.5°C (76). For the validation of the LiP-Chip method we have chosen to work with α-Lac from bovine milk (Sigma-Aldrich) in PBS and assay a temperature range between 4°C and 50°C.

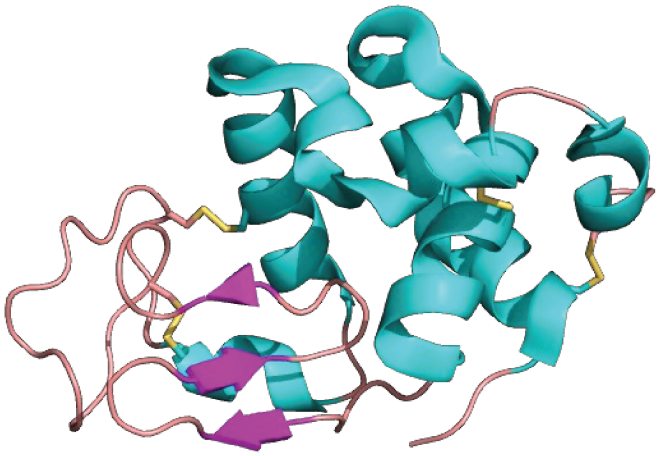


Figure 10. Crystal structure of α-Lac.

Crystal structure of apo α-Lac from bovine (PDB 1F6R) rendered with PyMol. Only chain A of six chains is displayed. The α-helices are depicted in light blue, β-sheets in pink, and loops are salmon-colored. The four disulfide bridges are shown in yellow.

Data analysis

In this section, we will discuss the general procedure of our data analysis, where the data comes from, which functions we fit, and what the error of our parameter estimation is. As an example experiment, we chose a LiP-Chip experiment of BSA with a BSA concentration of 2 µg/µL in a 15 µL reaction volume. The E/S ratio was 1/2 000 and the incubation time 7 min without preheating.

Data source

The thermal stabilities in this project were obtained with two different readout methods: SDS-PAGE and protein chips (see also 'Quantification of LiP results'). The readout with SDS-PAGE takes longer than the protein chip and provides less quantitative results, but offers a slight increase in sample size (ten samples in the protein chip, fourteen on the SDS-PAGE). Thus, we mostly used SDS-PAGE in an initial step to identify the best E/S ratio and incubation time for a subsequent LiP-Chip analysis of a protein. The four additional samples on the SDS-PAGE allowed us to assay a wider range of temperatures with different E/S ratios. In order to obtain the relative amount of undigested protein from an SDS-PAGE picture, we followed a densitometry analysis with ImageJ from the SYBIL (Systems biology for the functional validation of genetic determinants of skeletal diseases) project (78). This analysis consists of marking the area of interest by hand, and identifying and selecting the peak corresponding to the undigested protein (Figure 11 and 'Material and Methods'). The quality of the results strongly depends on the quality of the SDS-PAGE picture. Furthermore, the manual aspects of the densitometry analysis make it rather subjective. Nevertheless, this approach provides sufficient results for a first estimation of appropriate LiP-Chip parameters (see 'Validation of LiP off Chip').

For a more quantitative analysis of a protein's thermal stability we used protein chips in combination with the Agilent 2100 Bioanalyzer. The subsection 'Protein chip' in the section 'Quantification of LiP results' gives a detailed explanation of the protein chips used in this project. In short, protein chips are a microfluidics version of SDS-PAGE. With protein chips we can analyze a total of ten samples in as little as 30 minutes. The content of each sample is separated by size in a main channel. Larger proteins and peptides have a longer migration time than smaller ones. At the end of the channel, fluorescence intensity proportional to the amount

of protein is measured. The analysis software of the Bioanalyzer (2100 Expert Software) reports a spectrum for each sample (Figure 11). These spectra contain the fluorescence intensities over different migration times and thus, different protein sizes. Larger peptides take longer to pass the detector than smaller peptides. Thus, the raw fluorescence intensity of large peptides appears naturally higher than that of small peptides. The 2100 Expert Software provides a correction of the fluorescence signal according to a peptide's migration time, as well as a removal of the background fluorescence.

In addition to the peaks corresponding to the digested and undigested protein, the chip spectrum also contains peaks generated by the machine, as well as signals from a lower and an upper marker. These markers are provided in the chip preparation kit and are mixed with each sample at the same concentration. In contrast to the lower marker, the fluorescence intensity of the upper marker is not influenced by the digested proteins. We therefore used the upper marker to normalize for possible variability between the wells of a chip. For this, we divided the raw fluorescence intensities of a sample by the fluorescence intensity of the corresponding upper marker. We refer to the normalized fluorescence intensity corresponding to the peak of the undigested protein as relative undigested protein.

Before every detailed stability estimation of a new protein, we identified the migration time of the full protein. Our estimation of thermal stability is based on the amount of undigested protein after LiP at different temperatures. A typical LiP-Chip experiment consists of LiP samples at ten different temperatures. These temperatures should range over the pre-transition, transition, and post-transition phase (Figure 2). We will give a more detailed explanation of ideal temperature ranges further down. For each temperature, we measured the fluorescence signal at the migration time corresponding to the full protein. The fraction of undigested protein over different temperatures follows a sigmoidal curve (Figure 5). In the following paragraph we will introduce the function we chose to fit our data.

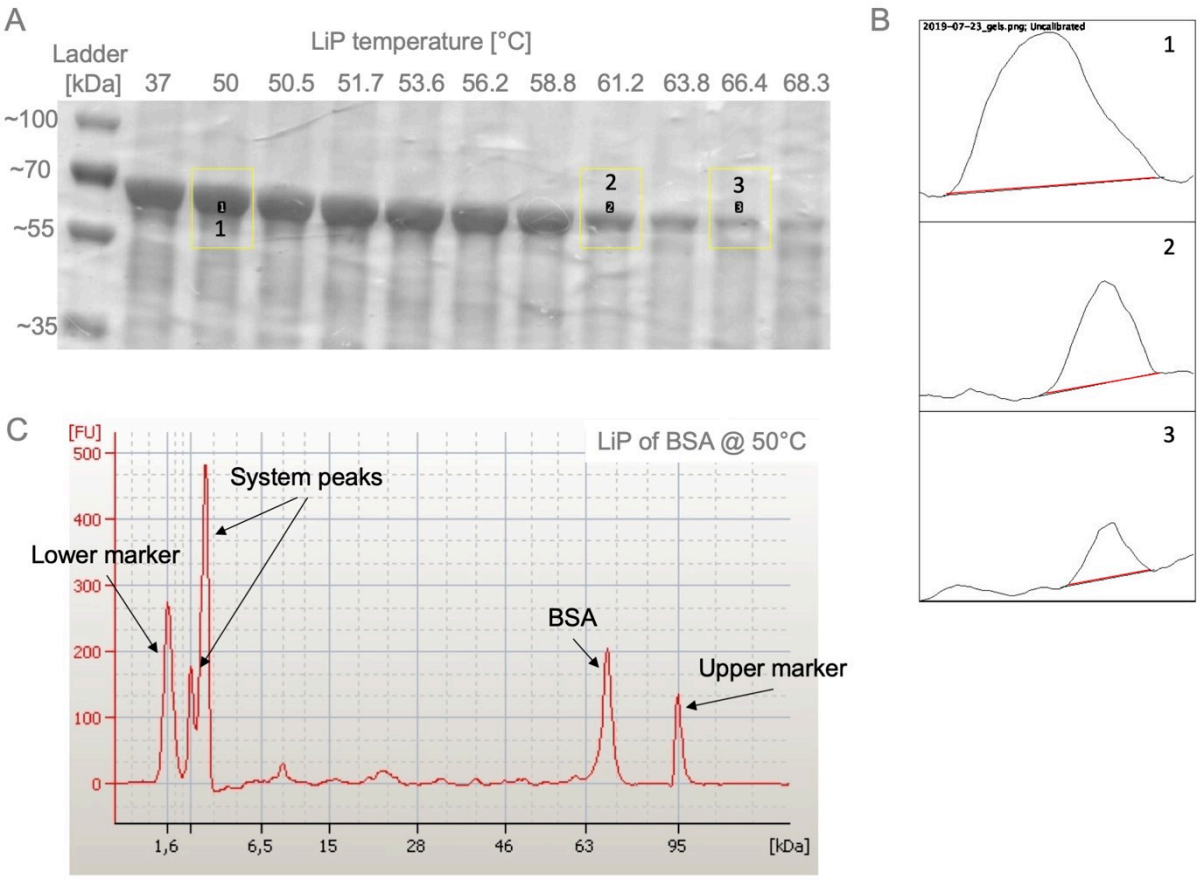


Figure 11. Densitometry analysis and protein chip analysis of LiP of BSA at different temperatures.

(A) Picture of a protein gel after SDS-PAGE of LiP samples of BSA at temperatures from 37°C to 68.3°C. The total amount of protein loaded was 10 µg per well. The band just below the 70 kDa ladder band corresponds to the undigested BSA. The SDS-PAGE was stained with PageBlue™ staining solution. The picture was taken in color with a standard scanner and converted into an 8-bit type with the ImageJ software. For the densitometry analysis, ImageJ was used to select the areas for analysis by hand. These areas are marked with yellow squares and numbered with 1 to 3 from left to right. For display purposes, we only selected three bands. (B) Intensity signals generated by the densitometry analysis of the marked areas in (A). The number in the plots corresponds to the number of the squares in (A). The lines highlighted in red have to be drawn by hand to select the peak of interest. The area under the peaks corresponds to the signal intensity of the undigested BSA and can be calculated with ImageJ. (C) The electropherogram of LiP of BSA at 50°C generated by the 2100 Bioanalyzer. The x-axis shows molecular size in kDa, the y-axis the raw fluorescence intensity in arbitrary units FU. The electropherogram contains a signal for the upper and lower markers of the sample, as well as peaks generated by the machine (system peaks). The peak corresponding to the undigested BSA is visible above 63 kDa. The lower peaks result from partly digested BSA proteins.

Fitting function

Protein stability data follows a sigmoidal curve (Figure 2). With a short incubation time and the assumption that proteolysis is much faster than unfolding, this holds also for the amount of undigested protein (22). At low temperatures (pre-transition phase), the amount of undigested protein is constant. This is followed by a sharp decrease in undigested protein around the melting temperature (transition phase). At high temperatures (post-transition phase), the digestion is maximal. The resulting amount of protein is only limited by the digestion rate of unfolded protein, the protease-susceptibility of the natively folded protein, and the incubation time of LiP.

In order to compare thermal stabilities of different proteins, we are interested in estimating the apparent melting temperature ($T_{m,app}$) of each protein. We specifically make the distinction between T_m and $T_{m,app}$. As described above (see 'Thermodynamics of protein (un)folding'), classical thermodynamics defines the T_m as the temperature at which ΔG is zero; or in other words at which the probabilities of a protein to be folded or unfolded are equal. This definition however, relies on the assumption of equilibrium conditions. This assumption does not hold in our experiments. The data obtained with the LiP-Chip method depend not only on the thermodynamic unfolding of the protein, but also on experimental parameters, such as incubation time and E/S ratio. Nevertheless, we can extract an apparent stability measure by defining $T_{m,app}$ as the temperature at which we reach the midpoint of digestion between the pre-transition and the post-transition plateaus (Figure 2).

For the fitting of our data, we focused on functions that allow an easy readout of $T_{m,app}$. One function that empirically fits well our data and that we used in the data analysis throughout this project is the following:

$$y(x) = \frac{A - D}{1 + \left(\frac{x}{C}\right)^B} + D \qquad \text{(Eq. 14)}$$

This model has been extensively used for the analysis of bioassays with sigmoidal behavior, such as dose-response curves or optical density of growing cultures (79,80). It contains four parameters: The upper plateau (A), the lower plateau (D), the steepness of the slope in the transition phase (B), and the inflection point (C). In our analysis, the variable x corresponds to the temperature of LiP in degree Celsius (°C). The result $y(x)$ is the normalized fluorescence

intensity of undigested protein at a given temperature. The apparent stability measure $T_{m,app}$ can directly be extracted from Equation 14 as the parameter C.

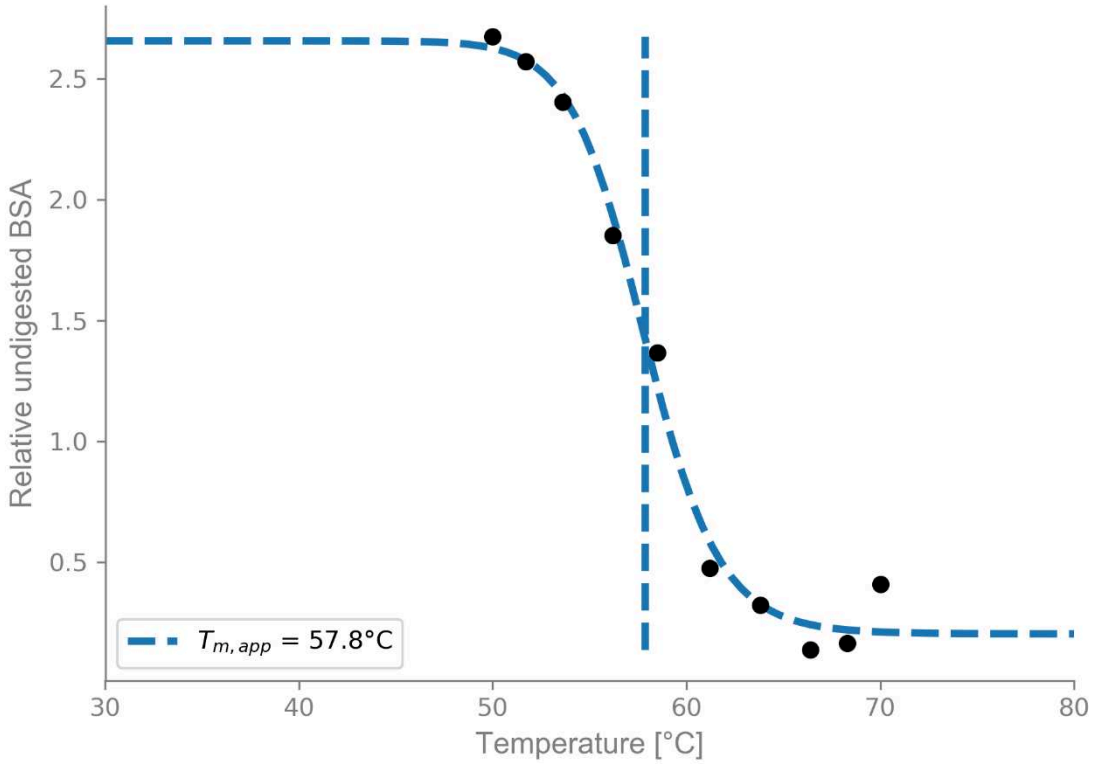


Figure 12. Fitted sigmoidal curve to LiP-Chip data of BSA.

(Black) Relative amount of undigested BSA based on the normalized fluorescence intensity of the peak corresponding to undigested BSA in a LiP-Chip experiment over ten temperatures. (Blue) The curve shows the optimal fit for Equation 14 fitted to the LiP-Chip data of BSA. The vertical line marks the inflection point of the curve at a $T_{m,app}$ of 57.8°C.

We used in-house python scripts for our data analysis. The goal was to find the optimal parameters for Equation 14 to fit the relative amount of undigested protein from LiP at ten different temperatures and to extract the $T_{m,app}$. Finding an optimal fit of a function to data is an iterative process. It starts with a guess for the parameters of the function (either generic guesses, such as 1, or specific guesses). The values predicted with the function and those initial parameters is compared to the data. The objective of the iterations is to minimize the distance between the predicted values and the data. For this purpose, parameters are changed in each iteration and the distance between predicted values and data is recalculated. Depending on the method used to change the parameters, initial guesses, and boundaries on the parameter space, the results of the optimization may vary. It is therefore important to choose these conditions

appropriately. For our data analysis, we used the *curve_fit* function from the scipy optimize package with the non-linear least squares optimization method. We chose biologically meaningful boundaries for the fit parameters. The result $y(x)$ corresponds to the relative amount of undigested protein. This value can either be zero or positive. Thus, parameter D can only be in the range $[0, \infty)$. For parameter A we can set a more precise boundary. A reasonable assumption is that during our LiP-Chip experiment no additional protein can be generated. Therefore, the largest value that the pre-transition plateau can reach is the amount of protein that the experiment was initiated with. Ideally, the initial amount of protein can be measured by replacing one LiP sample with a sample of the protein without PK. However, this data was not available for most of our experiments. We therefore estimated the initial amount of protein from observations based on several LiP-Chip experiments or LiP experiments quantified with SDS-PAGE.

In order to choose the boundaries for the remaining two parameters, we have to consider the process we are analyzing: The unfolding and subsequent digestion of protein with increasing temperature. We know that proteins unfold more and more with increasing temperature (see 'Quantifying protein folding'). Thus, it is reasonable to assume a negative slope in the transition phase. Given Equation 14, we can therefore assume that B is in the range between $[0, \infty)$. The proteins in this study are all mesophilic and show moderate stabilities within the range between freezing and boiling temperatures. Hence, we set the boundaries of parameter C (corresponding to $T_{m,app}$) between 0°C and 100°C.

With the above mentioned boundary conditions and generic initial guesses, Equation 14 fits well to the LiP-Chip data of our validation protein BSA with an estimated $T_{m,app}$ of 57.8°C (Figure 12). A good fit of Equation 14 to LiP-Chip data can also be observed for the validation protein α-Lac (Suppl. Figure 3). Furthermore, our estimation of $T_{m,app}$ is robust towards different fitting functions. We obtain similar results with a logistic function that is often used in the estimation of protein stability at equilibrium conditions (see 'Logistic fitting function' and Suppl. Figure 4). For a more detailed analysis of the fit, we applied the Bootstrap method to estimate the standard error of our $T_{m,app}$ estimation.

Standard error of the estimator

The previous paragraph highlighted that Equation 14 is a reasonable model for fitting LiP-Chip data. We need to consider however, that we are limited by a sample size of ten data points. How confident can we be in the estimated $T_{m,app}$? Can we optimize our data analysis by choosing the LiP temperatures appropriately? In order to answer the first question we analyzed the standard error (SE) of our $T_{m,app}$ estimation using the bootstrap method.

The bootstrap method is widely applied to estimate the SE and confidence interval of an estimator. For LiP-Chip, we are interested in the SE of the estimated $T_{m,app}$. Bootstrap starts with a data set; in our case ten LiP-Chip data points. The goal of the method is to create a large number (e.g. 1 000) of new data sets with the same size of the original one by randomly resampling the original data with replacement (81). That means that the new data sets can contain some of the original data points several times and some not at all. For each new data set, the estimator is calculated anew. Thus, with the new data sets we can generate a distribution of the estimator. This allows us to calculate the mean value of the estimator, as well as the standard deviation (SD). This SD of the resampling distribution of the estimator corresponds to the SE of the estimator.

We applied the bootstrap method with 1 000 resampling steps to the LiP-Chip experiment of BSA. The boundary conditions were chosen as mentioned above: B between 0 and 100, C and D in $[0, \infty)$, and A between 0 and the initial amount of BSA used in the experiment. Based on several experiments of BSA with and without PK, we estimated this initial amount of BSA to be at most eight times the maximal amount observed for the LiP samples.

Equation 14 contains four parameters. From the LiP-Chip method however, we only obtain ten data points. This produces a statistical limitation for the bootstrap method. Due to the small sample size and the random resampling of the data points, some resampled data sets might not contain sufficient unique data points to fit Equation 14. We therefore excluded resampled data sets with less than four unique data points. This does not occur very often and excluding the few cases does not significantly change the distribution of $T_{m,app}$.

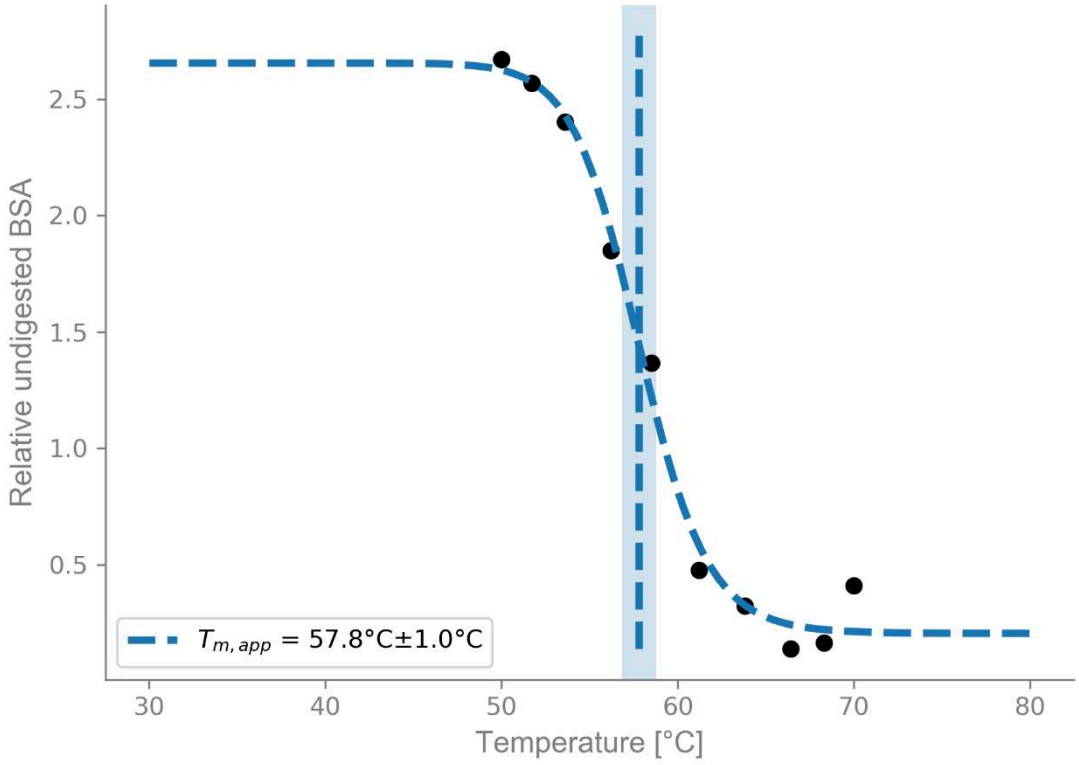


Figure 13. Bootstrap analysis of LiP-Chip of BSA.

The upper boundary of parameter A was chosen as the initial amount of BSA in the experiment. This amount of estimated to be 8 times the maximum value of relative undigested BSA observed for the LiP samples. The boundaries for the other parameters were chosen as follows: B ∈ $[0,\infty)$, C ∈ [0,100], D ∈ $[0,\infty)$. The displayed curve corresponds to the fit of all ten data points. The mean over the $T_{m,app}$ estimates from the bootstrap method is marked with a vertical line. The SE of the $T_{m,app}$ estimation is represented by the shaded area.

Using the bootstrap method on the example LiP-Chip experiment of BSA, the mean value for $T_{m,app}$ was 57.8°C with a SE of 1°C (Figure 13). This error of estimation is reasonably low. It is important to note however, that the SE of the $T_{m,app}$ estimation is influenced by the choice of data points and variability in the data. In the next paragraph, we analyze the importance of the different data points.

Importance of data points

The protein chips allow for the analysis of ten samples at a time. This limit on the sample size requires appropriate choosing of the tested LiP temperatures. In order to analyze the importance of the chosen data points at different LiP temperatures, we removed one of the data points at a time and fitted Equation 14 to the remaining nine data points (Figure 14).

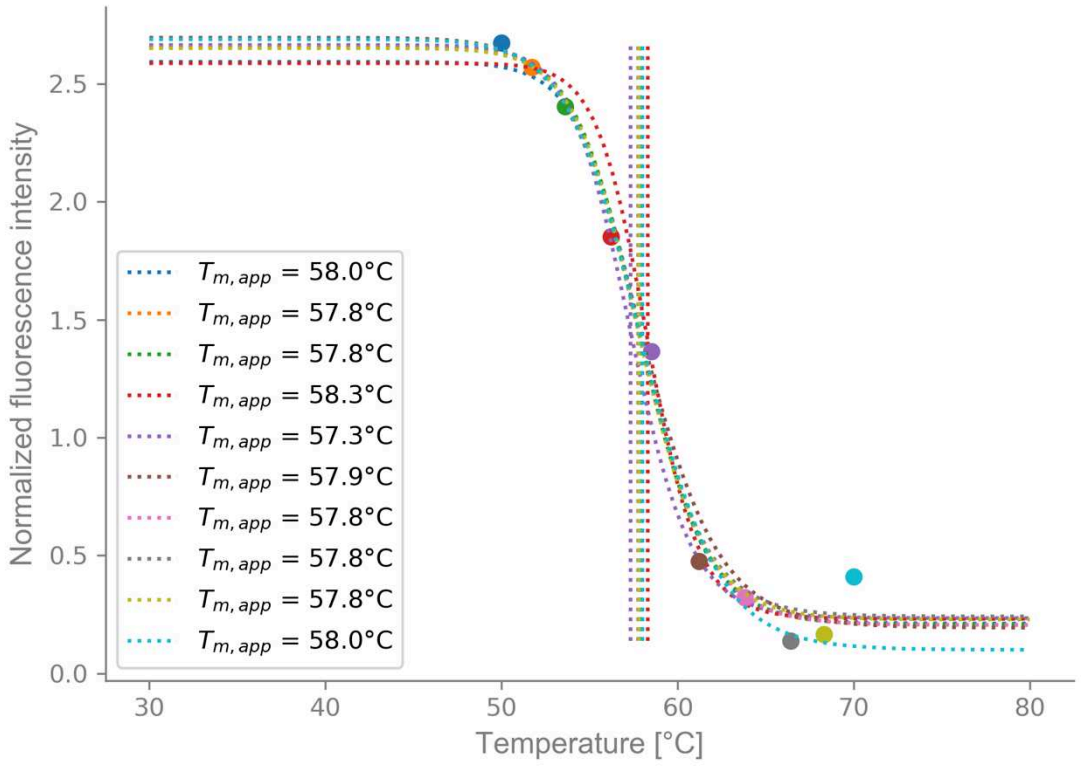


Figure 14. Effect of data point removal on $T_{m,app}$ estimation.

Data points were removed one at a time and Equation 14 was fitted to the remaining nine data points. The color of the data point indicates the removed point for the fitted curve and estimated $T_{m,app}$ (vertical line) of the same color. The estimated $T_{m,app}$ values for each color are reported in the legend.

The value of $T_{m,app}$ estimated on the basis of all ten data points is 57.8°C (Figure 12). For most of the data points, their removal does not change that value. Removing one of the two temperatures close to $T_{m,app}$ (56.2°C and 58.5°C, red and purple in Figure 14) however, leads to the largest deviation in the estimation of $T_{m,app}$. This indicates the importance of these data points close to the midpoint of the transition phase for fitting the data.

In addition to removing single data points, we analyzed the effect of removing pairs of data on the estimation of $T_{m,app}$. We calculated the difference between the estimated $T_{m,app}$ with all ten data points and the estimated $T_{m,app}$ with the removed data (Figure 15). The pairwise analysis strengthens the importance of the two data points around the $T_{m,app}$. No matter which other data point we removed, removing any of those two points from the analysis changed the estimation of $T_{m,app}$.

Apart from the two data points close to the midpoint of transition, removing the first two data points also affects the estimation of $T_{m,app}$. This is due to the limited sample size of the protein chips and the resulting lack of further data points on the pre-transition plateau. A weaker effect is also visible for the temperatures 66.4°C and 68.3°C. We suggest that a good distribution of LiP temperatures does not only include data points closely around the $T_{m,app}$, but also coverage of the pre- and post-transition plateaus. We confirmed this suggestion with an analysis of synthetic LiP-Chip data (Suppl. Figure 5).

It is important to mention the data point at 70°C. When removing this particular point, the post-transition plateau parameter (D) of the fit changes visibly (Figure 14 light blue curve). Compared to the data points at slightly lower temperatures, the value at 70°C seems too high. This is most likely due to aggregation of the unfolded BSA protein (see also ‘Validation of LiP off Chip’).

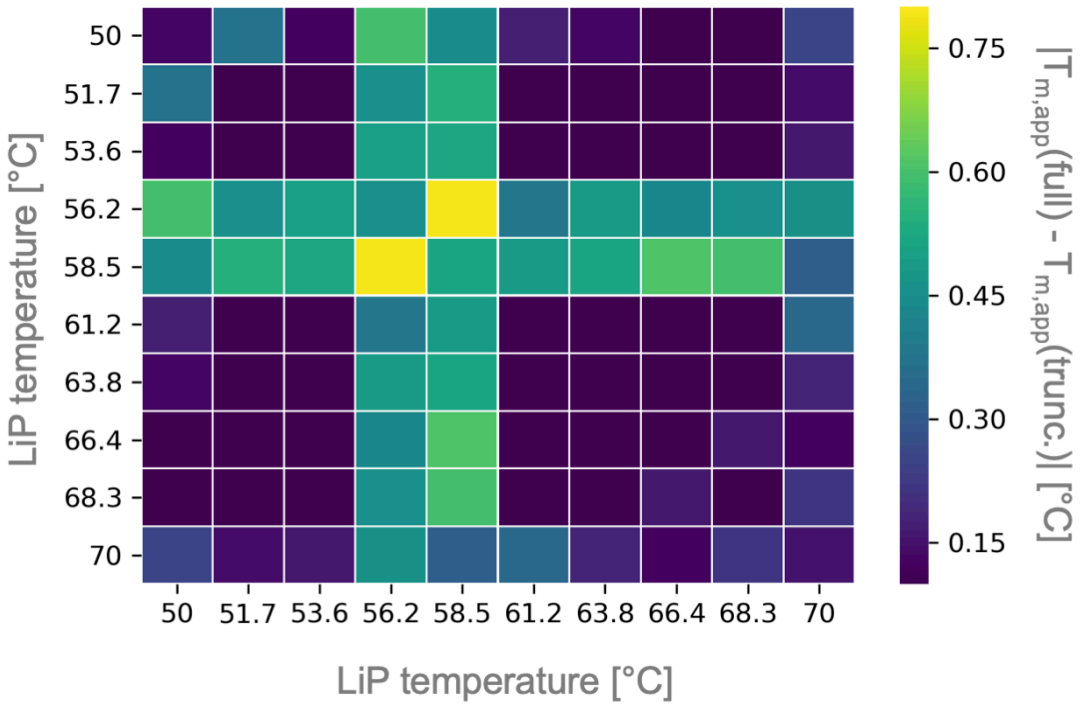


Figure 15. Single and pairwise effect of data point removal on $T_{m,app}$ estimation.

The effect on $T_{m,app}$ estimation is calculated as the absolute value of the difference between the estimation with all ten data points ($T_{m,app}$(full)) and the estimation after the removal of one or two data points ($T_{m,app}$(trunc.)). The single point removals are represented on the diagonal. Effects of pairwise point removals are shown off-diagonal. A dark blue color indicates no effect of the data removal for the estimation of $T_{m,app}$, whereas a yellow color indicates a difference in estimated $T_{m,app}$ after data removal.

Summary of data analysis

We showed in this section that the model we chose (see Equation 14) is well suitable for fitting LiP-Chip data. It allowed us to easily extract values for $T_{m,app}$. Having ten data points is a limitation, but we were nevertheless able to estimate a reasonably low SE with the bootstrap method. Our analysis of the importance of the individual data points highlights that the $T_{m,app}$ estimation can be improved by choosing the LiP-Chip temperatures appropriately. We suggest to use one sample to estimate the total amount of protein without PK in order to find a precise boundary for fitting the pre-transition plateau. Further, two temperatures should be picked in the pre-transition phase, two temperatures in the post-transition phase, and five temperatures evenly spread throughout the transition phase. We will describe an experimental procedure for choosing such LiP temperatures in the section ‘Recommended protocol’.

LiP-Chip

Validation of LiP off Chip

As a first step before using LiP in combination with protein chips, we validated the general idea of combining protein electrophoresis with our LiP approach on traditional protein gels. We conducted these first experiments with the validation protein BSA and established that an E/S ratio of 1/2 000 with an incubation time of 7 min and a final BSA concentration of 2 µg/µL is appropriate for a quantification with SDS-PAGE and Coomassie blue staining. In order to detect the melting temperature within the wide range of T_ms determined through literature research (see 'Validation proteins'), we covered a temperature range from 37°C to 73°C (Figure 16). The band of the undigested BSA is visible just below the 70 kDa marker of the ladder in the far left lane. The intensity of this band strongly decreases between the LiP at 58.8°C and 61.2°C.

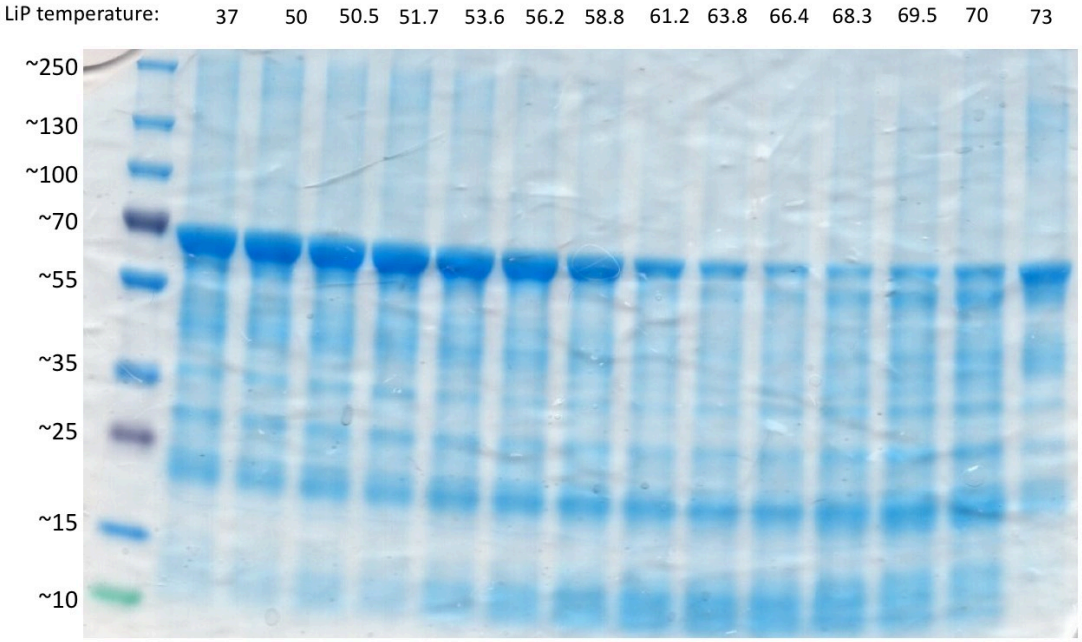


Figure 16. Quantification of LiP of BSA with SDS-PAGE and Coomassie blue staining.

The E/S ratio of the LiP experiment was 1/2 000, the incubation time 7 min without preheating. The BSA concentration was 2 µg/µL in a reaction volume of 15 µL. Each well was loaded with a total protein amount of 10 µg. The gel electrophoresis was conducted with a 4-20% Tris-Glycine gel and the following parameters: 25 mA, 200 V, 130 min. The gel staining was done overnight with PageBlue™ and the destaining in ddH_2O over the course of one day. The band of undigested BSA is visible just below the 70 kDa marker of the ladder.

As described in 'Data analysis', we used ImageJ to quantify the intensities of the bands corresponding to the undigested BSA, fitted a sigmoidal curve to the data, and estimated the $T_{m,app}$ of BSA as the inflection point of the resulting curve (Figure 17). Figure 17 does not only show the analyzed data from the protein gel depicted in Figure 16, but the results from a duplicate LiP experiment of BSA quantified with a separate SDS-PAGE. With estimated $T_{m,app}$ of 59.1°C and 58.2°C, these experimental replicates show a good reproducibility. Furthermore, these $T_{m,app}$ values are well within the range of T_m proposed throughout literature.

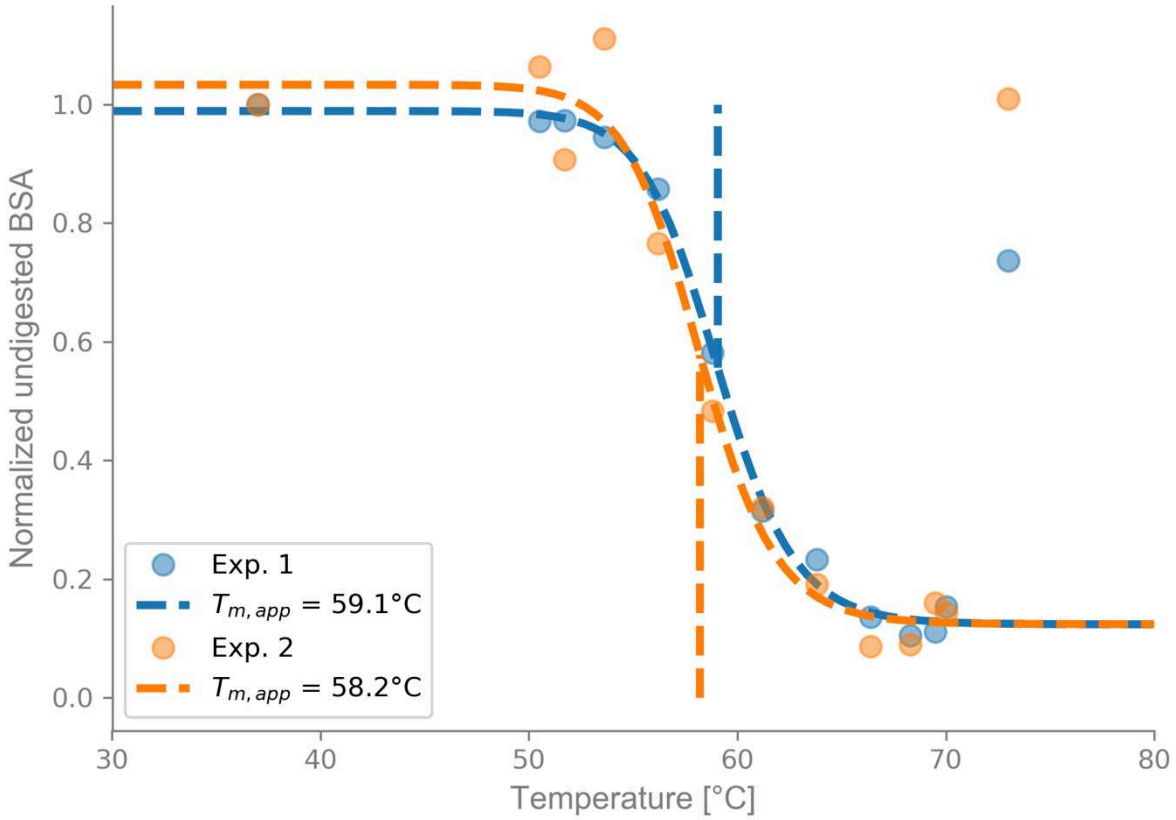


Figure 17. Estimated thermal stability of BSA from LiP and SDS-PAGE.

The data shows the normalized band intensities of the undigested BSA for duplicate experiments over different LiP temperatures. Each data set was normalized by the respective BSA band intensity at 37°C. For the LiP experiments, the E/S ratio was 1/2 000 and the incubation time 7 min without preheating. The BSA concentration was 2 µg/µL in a reaction volume of 15 µL. After heating of the LiP samples with the addition of SDS and DTT, 10 µg of total protein amount was used for the SDS-PAGE analysis. The 4-20% Tris-Glycine protein gels were stained overnight with PageBlue™ and destained in ddH_2O over the course of one day. The data were processed with ImageJ and analyzed as described in 'Data analysis'. For the fit of the sigmoidal curve, the outlier points at 73°C were excluded. The resulting estimated $T_{m,app}$ are 59.1°C for Exp. 1 and 58.2°C for Exp. 2.

The behavior of BSA throughout LiP over temperatures between 37°C and 68.3°C approaches the expected sigmoidal curve. A slight increase in undigested BSA is visible for the samples around 70°C. The amount of undigested BSA from the LiP sample at 73°C however, is quite variable between the duplicates and does not fit the sigmoidal curve. This is most likely due to protein aggregation. The temperature of 73°C highly destabilizes BSA, leading to the misfolding and unfolding of BSA and subsequent formation of protein aggregates. Such aggregates are protected from digestion by PK, but can be resolved by SDS. Thus, a large amount of aggregates might be visible on the SDS-PAGE as undigested protein. This hypothesis is further favored by the observation of a large variability between the duplicates. For further discussion see 'Protein aggregation'.

Another possibility for the unexpectedly high amount of undigested BSA at 73°C is a reduced activity of PK. Although it has been shown that the activity loss of PK between 37°C and 76°C is negligible (26), it is possible that a slightly reduced PK activity might contribute to the increase in undigested BSA. A rigorous estimation of the PK activity over the assayed temperatures could be used to normalize the observed digestion by the protease activity. For the analysis of BSA, sufficient data points on the post-transition plateau are available and the removal of the data point at 73°C does limit the estimation of $T_{m,app}$. For proteins with melting temperatures above 70°C however, we advise the use of thermostable proteases, such as thermolysin. These thermostable proteases need to be considered with caution. In contrast to PK, thermolysin cannot be inactivated by heat, but requires treatment with specific chemicals. Furthermore, thermolysin is more specific than PK, reducing the effect of the limited proteolysis.

Figure 17 shows that our LiP experiments of BSA with PK over a temperature range from 37°C to 70°C and subsequent quantification with SDS-PAGE is well suitable for the estimation of the $T_{m,app}$. Nevertheless, the quantification with SDS-PAGE and Coomassie blue staining remains time-consuming and error-prone (see 'Quantification of LiP results' and 'Data source'). Furthermore, the visualization of protein bands with Coomassie blue staining requires sufficient amount of protein in the LiP reaction. Throughout our experiments, we estimated the minimal amount of protein for LiP over ten temperatures with a quantification on SDS-PAGE with Coomassie blue staining to be 25 µg (based on results from Suppl. Figure 6). This amount is highly dependent on the assayed protein and the LiP conditions, such as E/S ratio and incubation time.

The quantification of LiP with SDS-PAGE and Coomassie blue staining is a valuable first approach to evaluate the LiP parameters and the temperature range to be assayed in more detail with LiP-Chip. For the next section, we will move on to the evaluation of our LiP-Chip method.

Reproducibility of LiP-Chip

In a previous section, we evaluated our data analysis and showed that the standard error of our $T_{m,app}$ estimation is reasonably low to compare thermal stabilities of proteins estimated with LiP-Chip. The data analysis however, is not the only source of error in a method based on biochemical experiments and technical equipment. One source of noise is LiP itself (experimental reproducibility). We will therefore discuss the reproducibility of the LiP-Chip experiment over several experimental replicates. Another source of noise is the variation between protein chips (technical reproducibility). We analyzed the influence of the 2100 Bioanalyzer protein chips on the reproducibility of LiP-Chip by comparing the results from the same LiP experiment quantified on different protein chips (technical replicates). One specific aspect of the technical reproducibility is an unusual increase in the background fluorescence of certain sample wells of the protein chip. We termed this the baseline effect.

Due to the availability of a limited amount of replicates (duplicates in most cases), we do not claim to estimate absolute statistics. Nevertheless, our results indicate the general applicability and limitations of the LiP-Chip method.

Experimental replicates

We define experimental replicates as reruns of the entire LiP-Chip experiment under the same conditions and with protein and protease samples from the same stock. Our rigorously timed and documented protocol allowed us to minimize variability of experimental replicates as much as possible (see ‘Material and Methods’). Nevertheless, biological experiments will always contain variability, with sources as mundane as the temperature of the room, the day of the week, or the quality of the experimenter’s breakfast.

An analysis of experimental duplicates of our validation protein BSA is shown in Figure 18. With an estimated $T_{m,app}$ of 58.9°C and 59.9°C, the difference between the stability measures of the two experimental replicates is only 1°C. This is within the range of experimental error reported by other thermal stability studies (28,31,48,76).

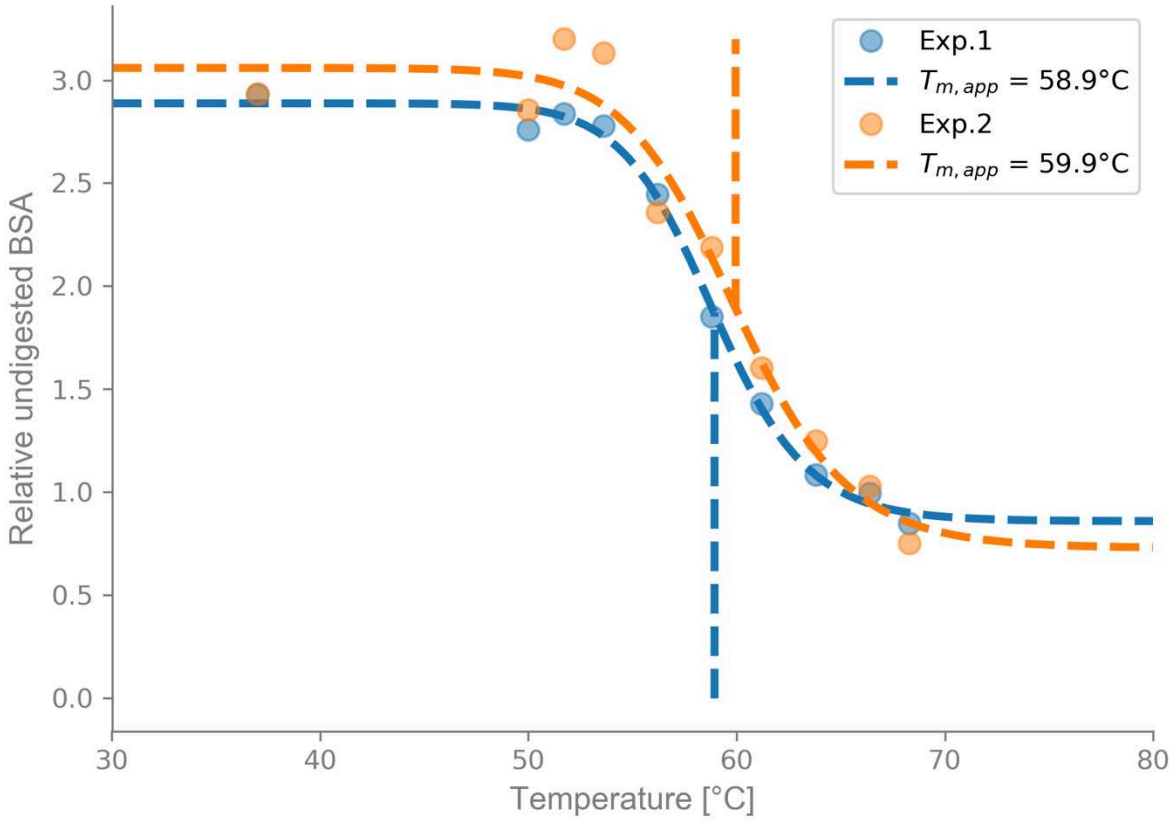


Figure 18. Experimental replicates of LiP-Chip with BSA.

The blue and the orange data represent two separate LiP-Chip experiments of BSA under the same conditions. For both experiments, the E/S ratio was 1/2 000, the incubation time 7 min with preheating of 5 min. The BSA concentration in the reaction was between 1.7 µg/µL and 2 µg/µL in a reaction volume of 100 µL. The data was processed and analyzed as described in 'Data analysis'. The estimated $T_{m,app}$ are 58.9°C and 59.9°C.

The difference between the estimated $T_{m,app}$ values of duplicate LiP-Chip experiments with our validation protein α-Lac is even smaller (Figure 19). The replicates have apparent melting temperatures of 38.2°C and 38.6°C. The estimated $T_{m,app}$ are highly similar despite the presence of several data points deviating from the fitted curve of Exp. 1. Furthermore, the relative amount of undigested α-Lac in Exp. 1 is approximately three times larger than the amount of undigested α-Lac in Exp. 2. This is due to the large difference in the fluorescence intensity of the upper marker between the duplicate experiments (Suppl. Figure 7). This difference in marker intensity can have several sources and it is difficult to identify the exact cause. Nevertheless, our duplicate experiments with experimental replicates of α-Lac suggests that the chip to chip differences in the fluorescence intensities does not have a strong effect on the estimation of $T_{m,app}$.

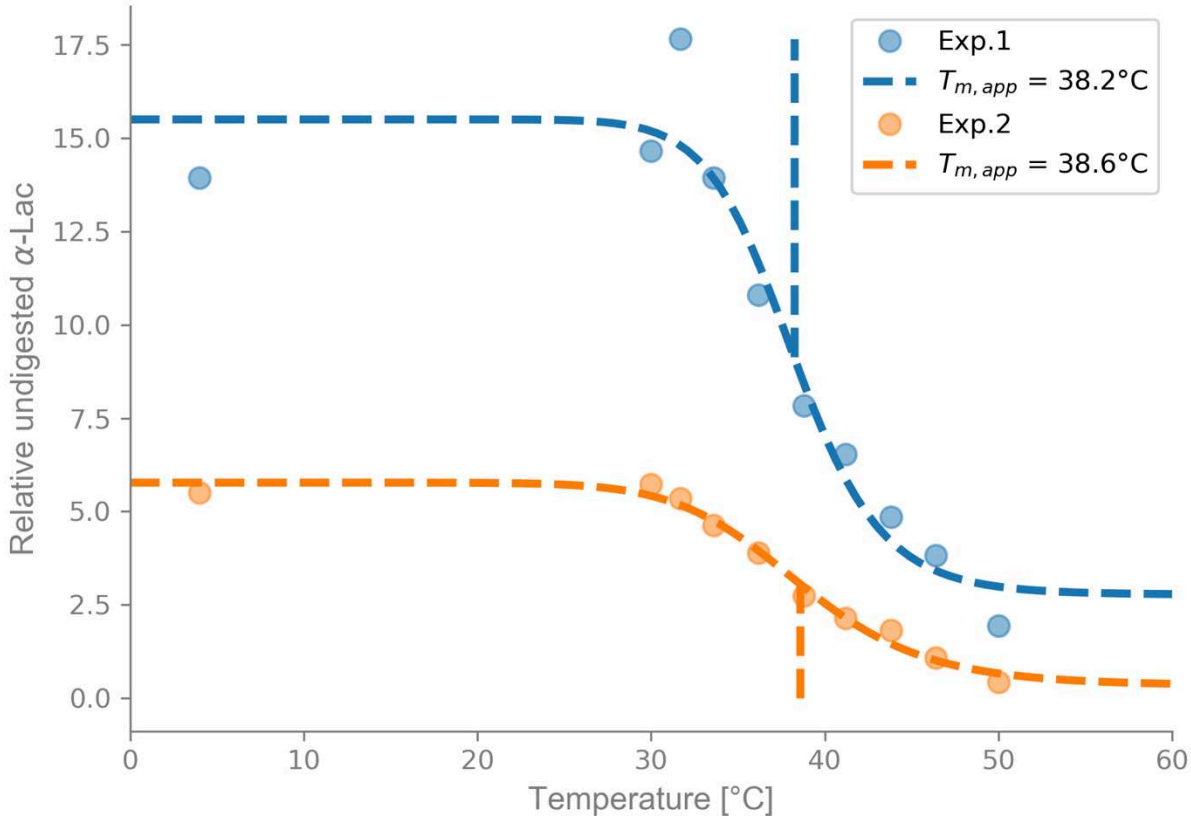


Figure 19. Experimental replicates of LiP-Chip with α-Lac.

The blue and the orange data represent two separate LiP-Chip experiments of α-Lac under the same conditions. For both experiments, the E/S ratio was 1/500, the incubation time 1 min without preheating. The α-Lac concentration in the reaction was 0.5 µg/µL in a reaction volume of 20 µL. The data was processed and analyzed as described in 'Data analysis'. The estimated $T_{m,app}$ are 38.2°C and 38.6°C.

Technical replicates

Technical replicates refer to one LiP experiment that was analyzed several times on separate protein chips. The comparison of technical replicates allows us to evaluate the effect of the 2100 Bioanalyzer protein chip technique on our results. As reported for the experimental replicates of α-Lac (Figure 19), we observe a large chip-to-chip difference for the technical replicates of BSA due to variations in the absolute fluorescence intensities of the upper marker (Figure 20 and Suppl. Figure 8). Despite this variability however, the estimated $T_{m,app}$ of the technical replicates with BSA are highly similar, with 59.9°C and 60.0°C. This indicates that LiP-Chip has a good technical reproducibility that is not strongly influenced by chip-to-chip variability.

Nevertheless, the large variations in upper marker intensity has to be taken into account when designing an experiment involving protein chips. In this project, we have chosen to limit the amount of assayed temperatures to ten in order to use one chip per stability estimation. If more than ten samples are desired, it is possible to split a stability estimation over several chips, as long as a means for normalization between chips is provided.

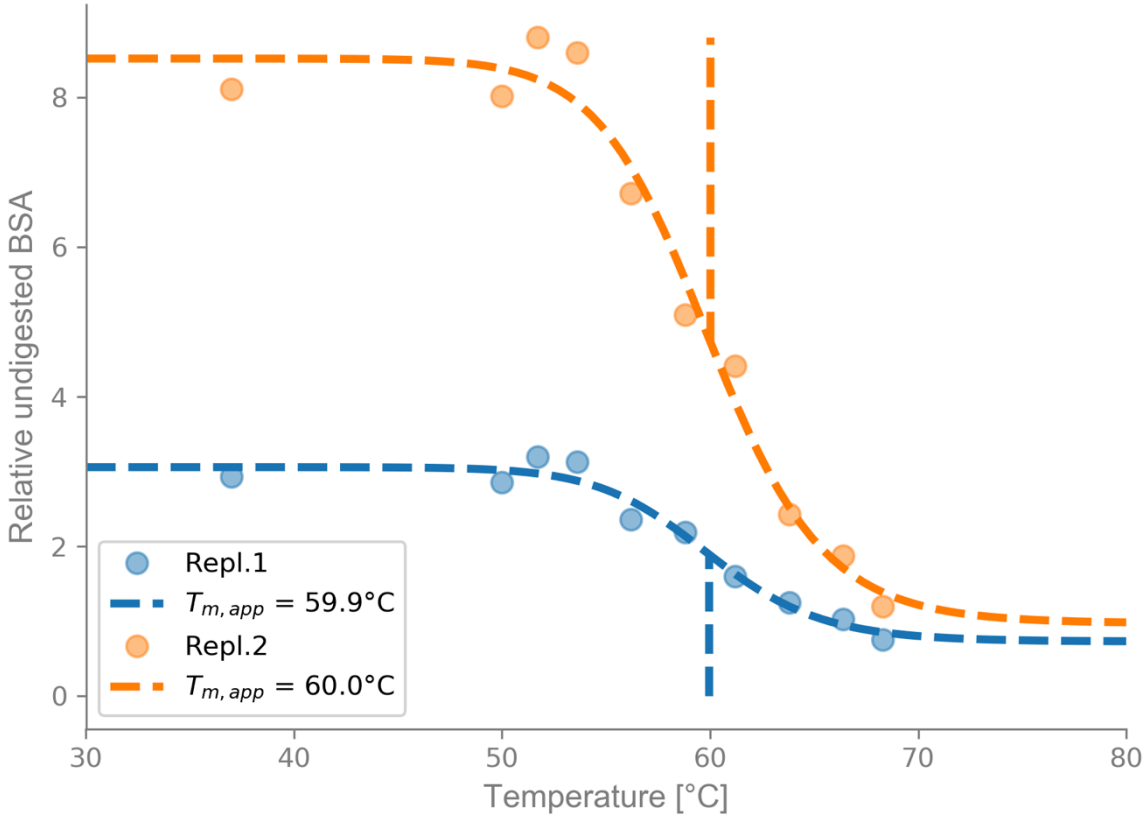


Figure 20. Technical replicates of LiP-Chip with BSA.

The blue and the orange data represent the same LiP-Chip experiment of BSA quantified on different protein chips. Repl. 1 corresponds to Exp. 2 of Figure 18. The E/S ratio was 1/2 000, the incubation time 7 min with 5 min preheating. The BSA concentration in the reaction was 2 µg/µL in a reaction volume of 100 µL. The data was processed and analyzed as described in 'Data analysis'. The estimated $T_{m,app}$ are 59.9°C and 60.0°C.

The comparison of the technical duplicates with BSA shows that the reproducibility with the Bioanalyzer can be quite high. There is however, one technical issue that we have encountered multiple times. We term this problem the baseline effect and will discuss it in more detail in the next subsection.

Baseline effect

We define as the baseline effect an increase of the background fluorescence with well number. The background fluorescence corresponds to the baseline of the fluorescence spectrum (Figure 11C). Ideally, each sample analyzed with the same chip should have the same background fluorescence. In approximately 50% of our experiments however, we observed an increase in the background fluorescence from sample six or seven on (Suppl. Figure 10). The background fluorescence for later samples was not only generally higher, but showed a linear increase with migration time. Although the 2100 Expert Software applies an internal baseline correction to the sample data, the baseline effect influences the measured amount of the upper marker and the protein of interest. This in turn affects the estimated $T_{m,app}$. For the technical duplicates of LiP-Chip with α-Lac, we obtained a difference of the estimated $T_{m,app}$ of 1.6°C (Figure 21). Repl. 1 did not have the baseline effect, whereas Repl. 2 had an increasingly strong baseline effect for the samples between 40°C and 50°C. The baseline effect results in a technical reproducibility that is higher than the experimental reproducibility with chips without the baseline effect (see Figure 19). Nevertheless, when taking into account the error of our $T_{m,app}$ estimations (see 'Standard error of the estimator'), the larger difference in estimated $T_{m,app}$ is just within the estimated SE (Suppl. Figure 9). Thus, as expected for technical replicates, the duplicates would most likely not be identified as significantly different. It is nonetheless important to be aware of the baseline effect and to take it into account when designing a LiP-Chip experiment. A more detailed analysis of the baseline effect can be found in the supplementary information.

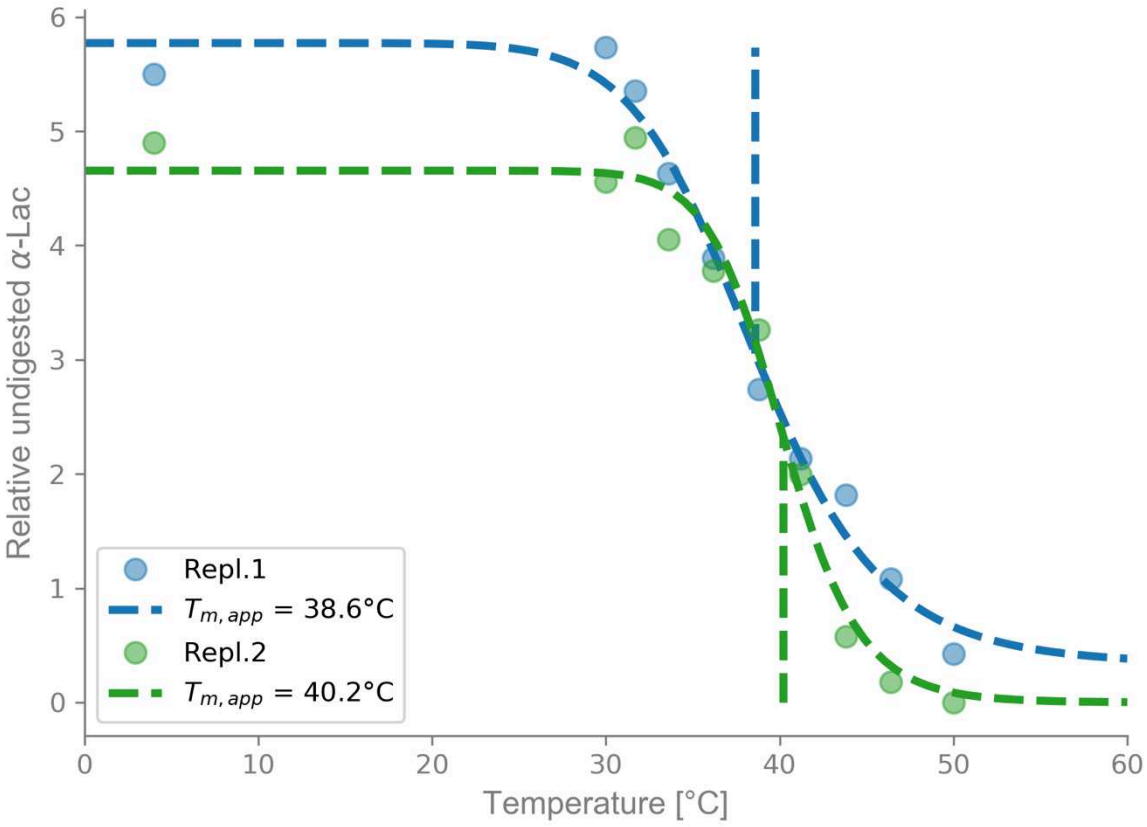


Figure 21. Technical replicates of LiP-Chip with α-Lac.

The blue and the green data represent the same LiP experiment of α-Lac quantified on two different protein chips. Repl. 1 corresponds to Exp. 2 in Figure 19. The E/S ratio was 1/500, the incubation time 1 min without preheating. The α-Lac concentration in the reaction was 0.5 µg/µL in a reaction volume of 20 µL. The data was processed and analyzed as described in 'Data analysis'. The estimated $T_{m,app}$ are 38.6°C and 40.2°C. Repl. 1 did not have the baseline effect, whereas Repl. 2 had the baseline effect from well seven on (corresponding to the LiP temperatures above 40°C). The replicate quantification was conducted on two different 2100 Bioanalyzer experiments.

Interestingly, the influence of the baseline effect on the estimation of $T_{m,app}$ is rather reproducible. With three technical replicates, one without (Repl. 1 in Figure 21) and two with the baseline effect (Repl. 2 in Figure 21 and Figure 22, and Repl. 3 in Figure 22), the estimated $T_{m,app}$ of the experiments with baseline effect are the same and show a large difference to the experiment without baseline effect (Figure 22 and Suppl. Figure 11). Furthermore, the sample order of Repl. 3 was altered compared to the other experiments. Here, the three samples affected by the increase of the background fluorescence are the samples around 30°C. For the other replicate with baseline effect, the samples above 40°C were affected. The results of Repl. 3 are much more variable than the results of Repl. 2. Due to the normalization with the upper marker, samples with larger protein amount are more strongly influenced by the baseline effect, than

samples with low protein amount. Therefore, we advise to use the last wells of a protein chip for samples at high temperatures above the T_m or to reduce the assayed LiP temperatures by three and only use the first seven wells of the protein chip. Otherwise, the addition of an internal marker, such as a large and known amount of BSA in LiP-Chip analyses of proteins smaller than BSA, could help to better correct for the baseline effect (see 'Detailed analysis of the baseline effect' and Suppl. Figure 12B).

Another solution could be to change to a different protein chip. In this study, we used the P80 chip that consists of a gel-matrix containing a fluorescent dye. The high-sensitivity HSP-250 protein chip on the other hand, uses pre-labelling of the samples and does not require the fluorescence dye in the gel-matrix. This might remove the baseline effect, but increases the processing time of the sample. A better removal of the baseline effect without increasing the experimental time might be achieved with a more elaborate baseline correction in the data processing. In addition to the corrected fluorescence data, the 2100 Expert Software provides the raw data without baseline correction or peak alignment. A detailed analysis of the raw data could give insight into the baseline effect and allow for a more adapted processing of the data.

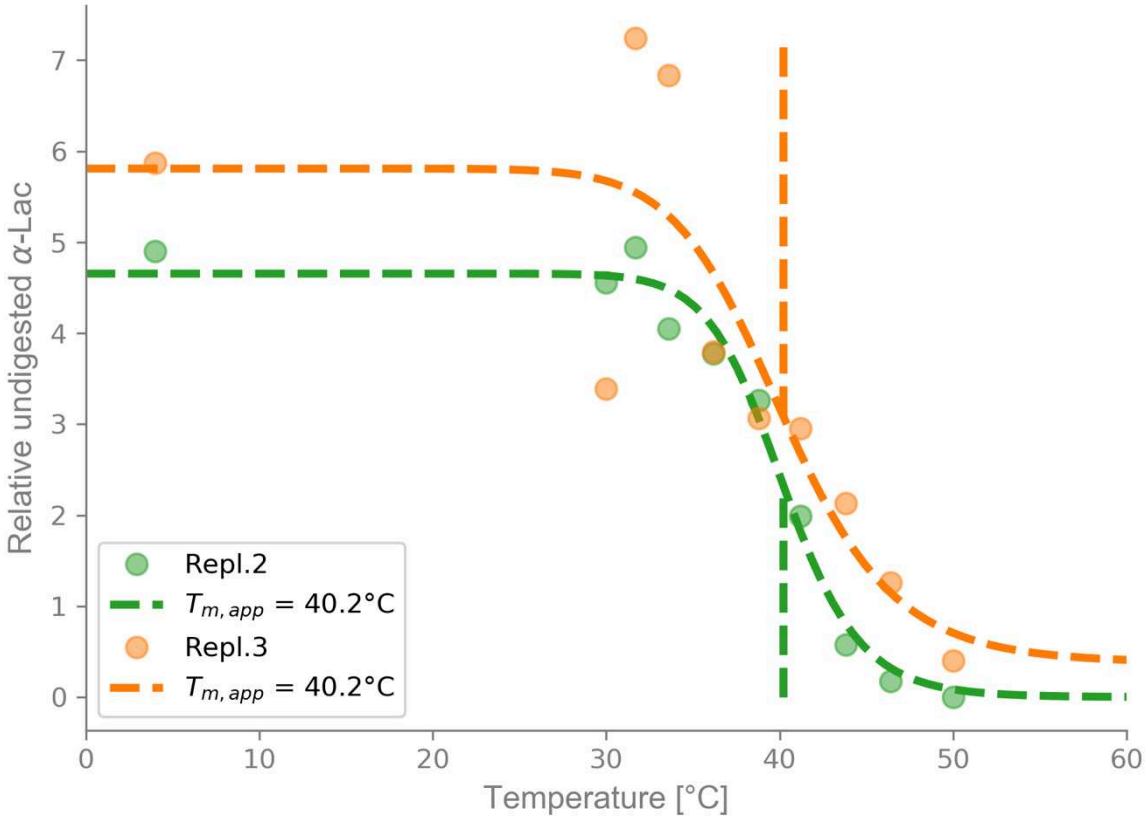


Figure 22. Technical replicates of LiP-Chip of α-Lac with baseline effect.

The green and the orange data represent the same LiP-Chip experiment of α-Lac quantified on two different protein chips. Repl. 2 corresponds to Repl. 2 in Figure 21. The E/S ratio was 1/500, the incubation time 1 min without preheating. The α-Lac concentration in the reaction was 0.5 µg/µL in a reaction volume of 20 µL. The data was processed and analyzed as described in 'Data analysis'. The estimated $T_{m,app}$ are 40.2°C for both replicates. Both replicates had the baseline effect. For Repl. 2 the LiP temperatures above 40°C were affected, for Repl. 3 the LiP temperatures around 30°C. The replicate quantification was conducted on two different 2100 Bioanalyzer experiments.

Robustness to Enzyme/Substrate ratios

The ratio between the protease and the protein of interest (E/S ratio) is one of the parameters of LiP. Literature on LiP often states the need for adjusting the E/S ratio for every protein individually. If the E/S ratio is too high, the proteolysis is mostly limited by the rate of hydrolysis. Thus, the proteins in the sample will be maximally digested at all LiP temperatures, preventing the observation of the change in unfolded proteins around the T_m. With a too low E/S ratio on the other hand, the chance of the protease to encounter a protein is too low to obtain a representative sample for the estimation of the amount of unfolded proteins.

The precise optimization of the E/S ratio for each protein would add a large amount of work, especially in mutation saturation studies or studies of protein libraries. With LiP-Chip, we can show that the E/S ratio is robust within wide boundaries and does not require extensive optimization.

For these experiments, we chose the validation protein BSA and exposed it to different concentrations of PK over a range of temperature for an incubation time of 7 min. The assayed E/S ratios were 1/200, 1/2 000, and 1/10 000. The incubation time and the E/S ratio are strongly linked. With a long incubation time, even a low E/S ratio can lead to extensive digestion of the protein of interest. On the contrary, a short incubation time allows the usage of high E/S ratios while conserving the effect of temperature on the unfolding and subsequent digestion of the protein of interest. Our results show that LiP-Chip of BSA with PK over an incubation time of 7 min is robust for E/S ratios between 1/2 000 and 1/10 000 (Figure 23). The estimated $T_{m,app}$ are 59.9°C and 58.6°C, respectively. This difference of 1.3°C between the estimated $T_{m,app}$ is within the error of the LiP-Chip method as described in 'Reproducibility of LiP-Chip'.

In addition to the robust E/S range, our results show that with an incubation time of 7 min the E/S ratio of 1/200 is too high for the estimation of the $T_{m,app}$ of BSA. In contrast to the data with E/S ratios of 1/2 000 and 1/10 000, the results for a ratio of 1/200 do not follow a sigmoidal behavior corresponding to protein unfolding. In fact, the amount of undigested BSA slightly increases with increasing temperature. This behavior is likely due to aggregation. BSA tends to aggregate at a temperature above 50°C (82). This is further favored by high protein concentrations, such as the BSA concentration of 2 µg/µL in this experiment. Aggregates can be quite stable and will protect the protein from digestion with proteases. During the analysis on the protein chip however, we treat the LiP samples with SDS, leading to potential solvation

of the aggregates. Thus, the observed amount of undigested protein might be higher if aggregates had formed during LiP. We will discuss this in more detail in 'Protein aggregation'.

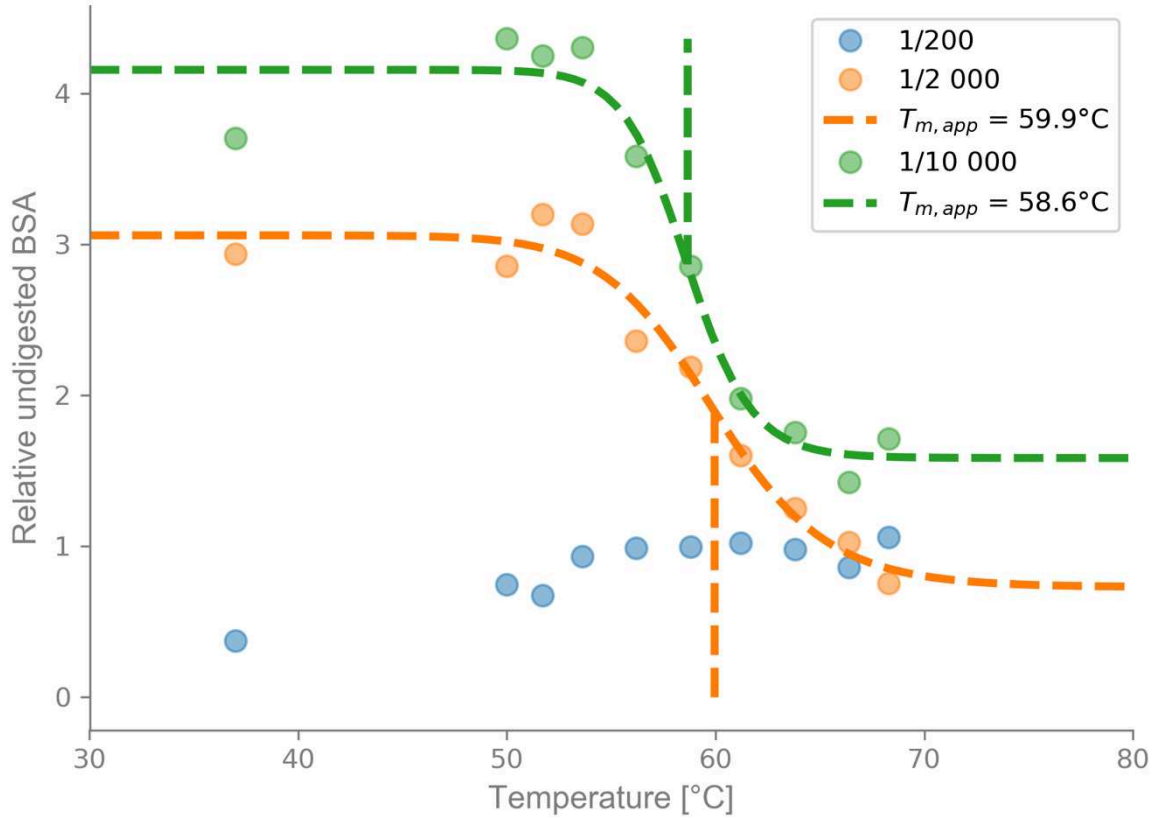


Figure 23. LiP-Chip stability of BSA over different E/S ratios.

The incubation time for all three E/S ratios was 7 min. The final BSA concentration was 2 µg/µL in a reaction volume of 100 µL. The BSA was preheated to the respective LiP temperatures for 5 minutes. (Blue) E/S ratio of 1/200. The data does not follow the expected sigmoidal curve. (Orange) E/S ratio of 1/2 000. The $T_{m,app}$ is estimated at 59.9°C. (Green) E/S ratio of 1/10 000. The estimated $T_{m,app}$ is 58.6°C. The $T_{m,app}$ for the orange and the green data were obtained by fitting Equation 14.

Our experiments represent only a first analysis of the robustness of LiP-Chip to different E/S ratios. They indicate that LiP-Chip is robust over a wide range of E/S ratios for the validation protein BSA. Additionally, our data shows a clear characteristic of an E/S ratio that is not appropriate. Thus, it is not necessary to evaluate many different E/S ratios for each protein of interest, but testing a few reasonably chosen ratios is sufficient for finding appropriate LiP conditions.

For a more detailed analysis of the robustness of LiP-Chip, further experiments could include the search for the higher and lower limit of E/S ratios by testing more ratios between 1/200 and

1/2 000, as well as above 1/10 000. The estimation of the full robustness range can then be used in an analysis of the relationship between the incubation time and the E/S ratio. We expect that the robustness range shifts towards higher E/S ratios with decreasing incubation time. Interesting is also to apply this robustness analysis to other proteins, such as our second validation protein α-Lac. This comparison of robustness ranges over different proteins would give more insight into the need for E/S ratio adjustments for studies of many different proteins.

Different incubation times

The incubation time is another important parameter of LiP. A change in incubation time will lead to a change in the overall amount of undigested protein. In an extreme case of an incubation time of several hours, this can lead to a full digestion of the protein of interest, making an estimation of $T_{m,app}$ impossible. As described above, we expect the incubation time to be inversely linked to the E/S ratio. With a long incubation time requiring a low E/S ratio and a short incubation time a higher E/S ratio for the LiP-Chip stability estimation.

In their 2004 paper, Minde *et al.* claim that the incubation time changes the estimated $T_{m,app}$ (43). They report decreasing $T_{m,app}$ values with increasing incubation times from 66 seconds to 11 minutes for the entire LiP process of Maltose Binding Protein. From a theoretical point of view however, the estimated $T_{m,app}$ should not change with the incubation time. A possible effect of an increased incubation time might rather be the overall lowering of the amount of undigested protein. We would however, still expect a clear decrease around the same temperature as for other incubation times. The reduction in $T_{m,app}$ by Minde *et al.* might be due to the low sensitivity of Coomassie-stained protein gels.

Our preliminary results show that a change in incubation time within a range from 7 to 20 min does not affect the estimated $T_{m,app}$ for BSA. For that, we analyzed the LiP-Chip stability of BSA with PK at an E/S ratio of 1/2 000 over three different incubation times: 1 min, 7 min, and 20 min. The estimated $T_{m,app}$ are 59.9°C for 7 min incubation and 61.0°C for 20 min (Figure 24). This is within the error of LiP-Chip (see 'Reproducibility of LiP-Chip'). The results of the experiment with an incubation time of 1 min shows large variability in the reported amount of undigested BSA. This strongly influences the estimation of $T_{m,app}$ (Suppl. Figure 13). Nevertheless, with an increase of the incubation time from 7 min to 20 min, and considered with caution an incubation time of 1 min, we do not observe a decrease of estimated $T_{m,app}$ for our validation protein BSA with increasing incubation time.

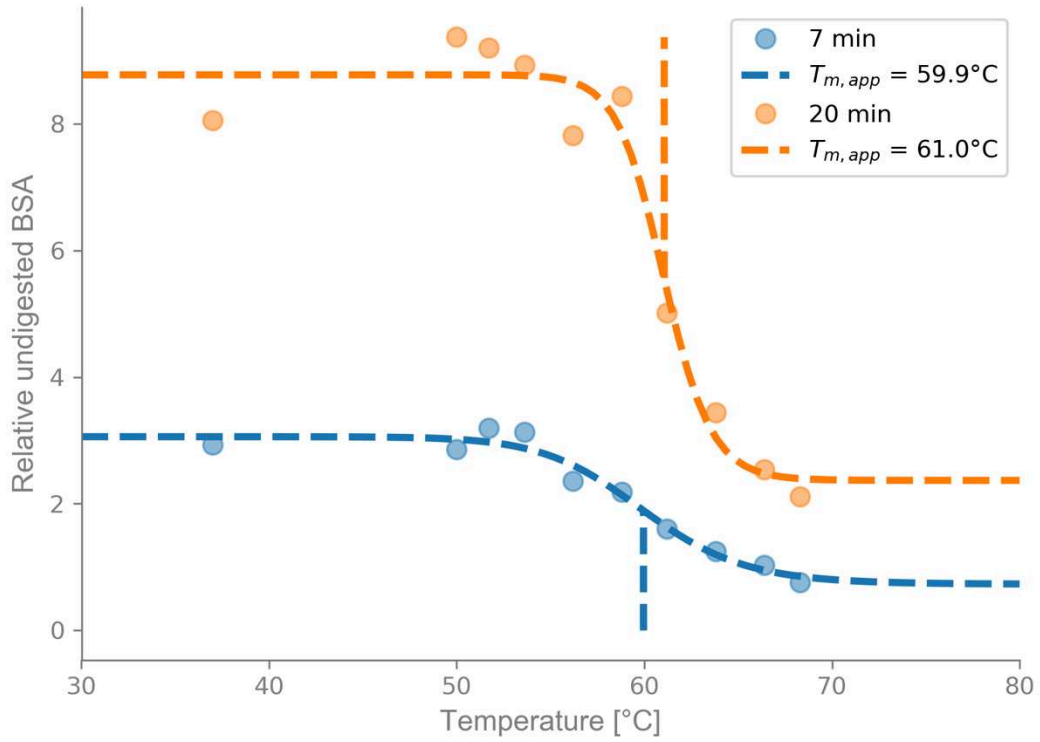


Figure 24. LiP-Chip stability of BSA over different incubation times.

The blue data represents an incubation time of 7 min and corresponds to the experiment with E/S 1/2 000 in Figure 23. The orange data was generated with an incubation time of 20 min. The E/S ratio for all incubation times was 1/2 000 with a BSA concentration of 2 µg/µL in 100 µL reaction volume. The BSA samples were preheated for 5 min to their respective temperatures. The data was fitted with Equation 14. The estimated $T_{m,app}$ are 59.9°C for an incubation time of 7 min and 61.0°C for an incubation time of 20 min.

Our analysis can easily be extended by assaying more incubation times. For a detailed analysis of the relationship between the incubation time and the E/S ratio, it could be interesting to determine the upper and lower boundaries. How long do we need to incubate so that the digestion is too strong to show the unfolding around the $T_{m,app}$? Which minimal incubation time is necessary to sufficiently digest the unfolded proteins in the sample for an estimation of $T_{m,app}$? The link between the incubation time and the E/S ratio implies that a change in E/S ratio would shift the boundaries of the incubation time. At a lower E/S ratio, we expect the full digestion of the protein of interest to take longer than at a higher E/S ratio. Nevertheless, our results suggest that within certain boundaries, the choice of incubation time and E/S ratio is robust towards changes. This would allow the comparison of several different proteins under the same conditions and simplifies their comparison with regard to their thermal stability.

Sample preheating

In this section, we want to discuss the aspect of preheating the protein sample for the estimation of the thermal stability. Some studies preheat their samples, others do not (26,43). From a theoretical point of view preheating would allow the protein sample to reach thermodynamic equilibrium. This is a prerequisite for the estimation of thermodynamic parameters, such as ΔG and absolute T_m. A protein sample would be exposed to its LiP temperature for an extended amount of time prior to the addition of the protease. The protein sample will approach a state of zero net change between the folded and unfolded conformations. At this point, the protease can be added to digest only the unfolded proteins and provide a measure of unfolding at the different LiP temperatures. From a theoretical standpoint, preheating can be compared to the extended incubation of a protein sample with a denaturant in chemical stability studies. Technically however, there are a few issues that we need to address when preheating. With our LiP-Chip protocol using PCR machines, the preheating of small sample volumes and subsequent opening of the machine or removal of the sample for the addition of the protease, can lead to condensation. Furthermore, small volumes of preheated protein cool down quickly during the addition of the protease (kept at 4°C to limit autolysis). This increases the experimental error. The sample volume should therefore be scaled up, increasing the necessary amount of protein for each experiment with preheating. For example, a small preheating volume of 10 µL of BSA were very difficult to handle and resulted in a large variability between the LiP samples (Suppl. Figure 14). By scaling up the volume of the BSA samples to 100 µL for each LiP temperature, we were able to estimate the thermal stability of BSA (Figure 25 orange). It is important to note that for BSA the sample preheating shifts the estimated $T_{m,app}$ to slightly lower temperatures (Figure 25). This highlights the need for a uniform protocol when comparing the thermal stability of different proteins.

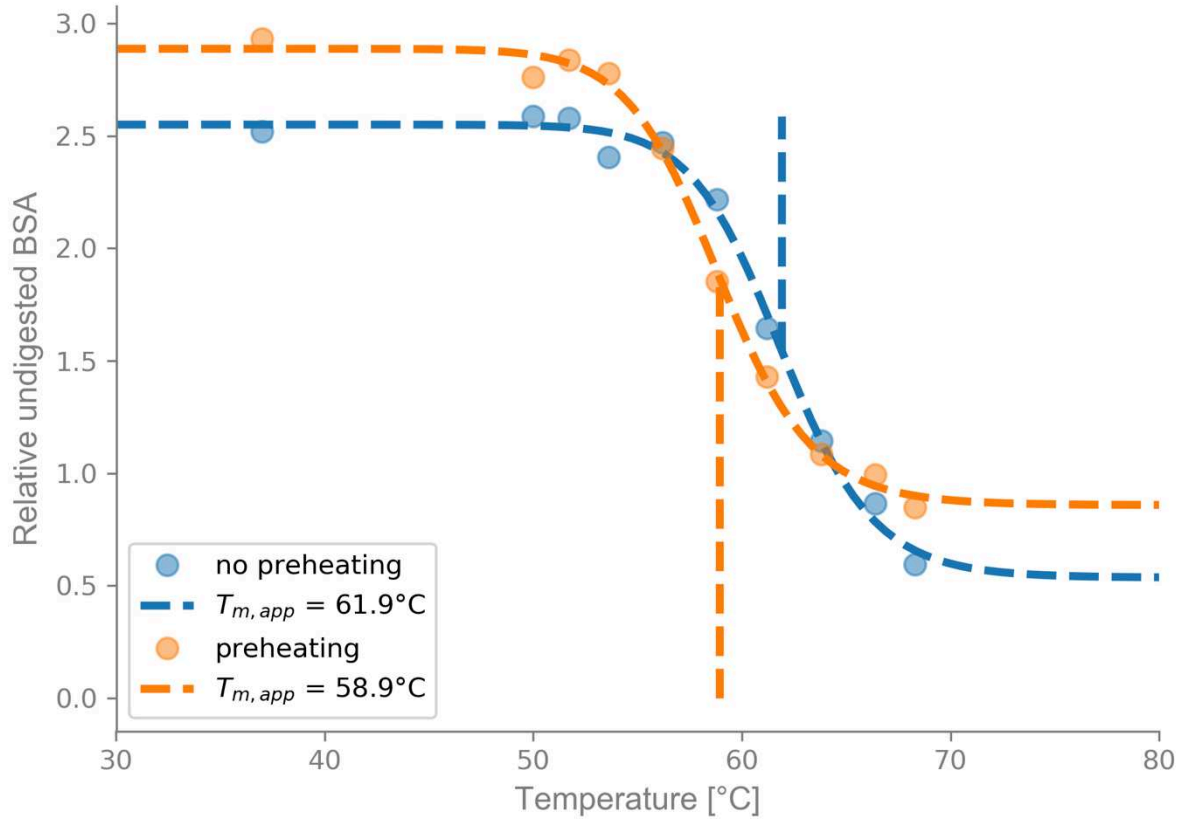


Figure 25. LiP-Chip experiment of BSA with and without preheating.

Both experiments were conducted with an E/S ratio 1/2 000. The blue data was generated with an incubation of 7 min without preheating. The BSA concentration was 2 µg/µL in a reaction volume of 15 µL. The orange data corresponds to an incubation time of 7 min with preheating of 5 min. The BSA concentration was 1.7 µg/µL in a reaction volume of 100 µL. This data is the same as Exp.1 in Figure 18. The data was fitted with Equation 14 and results in an estimated $T_{m,app}$ of 61.9°C for the experiment without preheating and 58.9°C with preheating.

Although theoretically the equilibration in chemical denaturants and preheating achieve the same goal, they can practically be very different. The most striking difference is protein aggregation. It can prevent the sample to approach the thermodynamic equilibrium and influence the estimated thermal stability. Aggregation is the oligomerization of misfolded or unfolded proteins and thus, linked to a protein's stability. In chemical stability experiments, the unfolding of the proteins is achieved through binding to the denaturant. This largely prevents the aggregation of the unfolded proteins. Thermal unfolding on the other hand, relies solely on denaturation by temperature. Therefore, the more proteins in a sample are unfolded, the more likely they aggregate. As assemblies of multiple proteins, aggregates are less accessible to proteases than individual unfolded proteins, distorting the measured amount of folded and

unfolded protein. We will show in the next subsection that aggregation can generally influence LiP-Chip, even without preheating. Since aggregation is temperature- and time-dependent however, it is further favored by preheating.

The effect of preheating on the protein aggregation depends on a protein's sequence and environment. Not all unfolded or misfolded proteins form aggregates under any condition. It is therefore important to consider the protein of interest and the question behind the experiment when deciding on sample preheating. For a protein that is known to be aggregation-prone and that is only available in small amounts, the preheating step should be omitted. On the other hand, preheating can introduce another layer of information when analyzing for example mutations of a protein. With preheating, it might be possible to identify a mutation that decreases aggregation, but does not increase thermal stability. Varying the time of sample preheating, can give insight into the propensity of a protein to aggregate. However, these applications of sample preheating are outside of the scope of this thesis. For the validation of the LiP-Chip method with the protein α-Lac, as well as for the other proteins in this study except for BSA, we did not apply the sample preheating.

Protein aggregation

Protein aggregation is an important property of any protein solution. As previously mentioned, it plays an integral role in neurodegenerative diseases, such as Alzheimer's disease, and influences the use of proteins as therapeutics. Protein aggregates refer to the assembly of unfolded or misfolded proteins into larger structures. Every protein has an intrinsic propensity to aggregate, depending on the environment (5,8). The pH, temperature, and protein concentration among others, can strongly influence the formation of protein aggregates. Due to the temperature denaturation, thermal stability experiments can be affected by protein aggregation. Some studies even used aggregation as a proxy for thermal stability, as for example in the Thermal Proteome Profiling (TPP) method (83). However, protein aggregation is a complex process with diverse outcomes, from disordered amorphous aggregates to ordered fibrils. Some aggregates are irreversible, others dissolve and the proteins refold into their native conformation when the temperature is lowered. The detailed understanding of protein aggregation is still under investigation and requires further research.

As the LiP-Chip method relies on the denaturation of proteins by temperature, it can be affected by protein aggregation. Aggregates that form during LiP of the protein sample are more protected from protease digestion than the unfolded proteins. The subsequent preparation of the LiP samples for the protein chip requires the use of SDS to unfold the proteins and peptides and facilitate their separation in the gel matrix. SDS is a denaturant that can dissolve protein aggregates. Thus, aggregation in the LiP samples can lead to an overestimation of the amount of natively folded protein. Since aggregation is temperature dependent, this overestimation of folded protein is stronger at high LiP temperatures. We observed such an aggregation effect during our LiP-Chip experiments with the single-chain variable fragment (scFv) of the A2 antibody (Figure 26). We will discuss A2 scFv in more detail in the following chapter.

The comparison between an experiment with strong aggregation and an experiment without, highlights the two parameters that we can adjust to avoid protein aggregation: protein concentration and E/S ratio. It is well known that protein concentration influences the formation of aggregates (6). One major challenge of protein purification for example, is the aggregation of overexpressed protein. Depending on the aggregation propensity of the protein of interest, as well as the environmental conditions, high protein concentrations can distort the results so much that an estimation of $T_{m,app}$ is not possible. With LiP-Chip however, we can analyze samples with concentrations as little as 67 ng/µL. We will discuss the specific LiP-Chip protocol for low concentrations in the next subsection.

The protein concentration in the LiP samples can not only be directly reduced to avoid aggregation, but also indirectly through an increase in protease. Increasing the E/S ratio might lead to more digestion of unfolded protein and a decrease in their assembly to aggregates. A decreased E/S ratio in combination with an increase in protein concentration as shown in Figure 26 however, can favor protein aggregation. Our preliminary results from LiP-Chip of A2 scFv shows the need for taking aggregation into account when designing a LiP-Chip experiment. Further experiments with different combinations of protein concentrations and E/S ratios might provide more insight into the interplay of these two parameters, as well as give information on the relationship between aggregation and thermal stability.

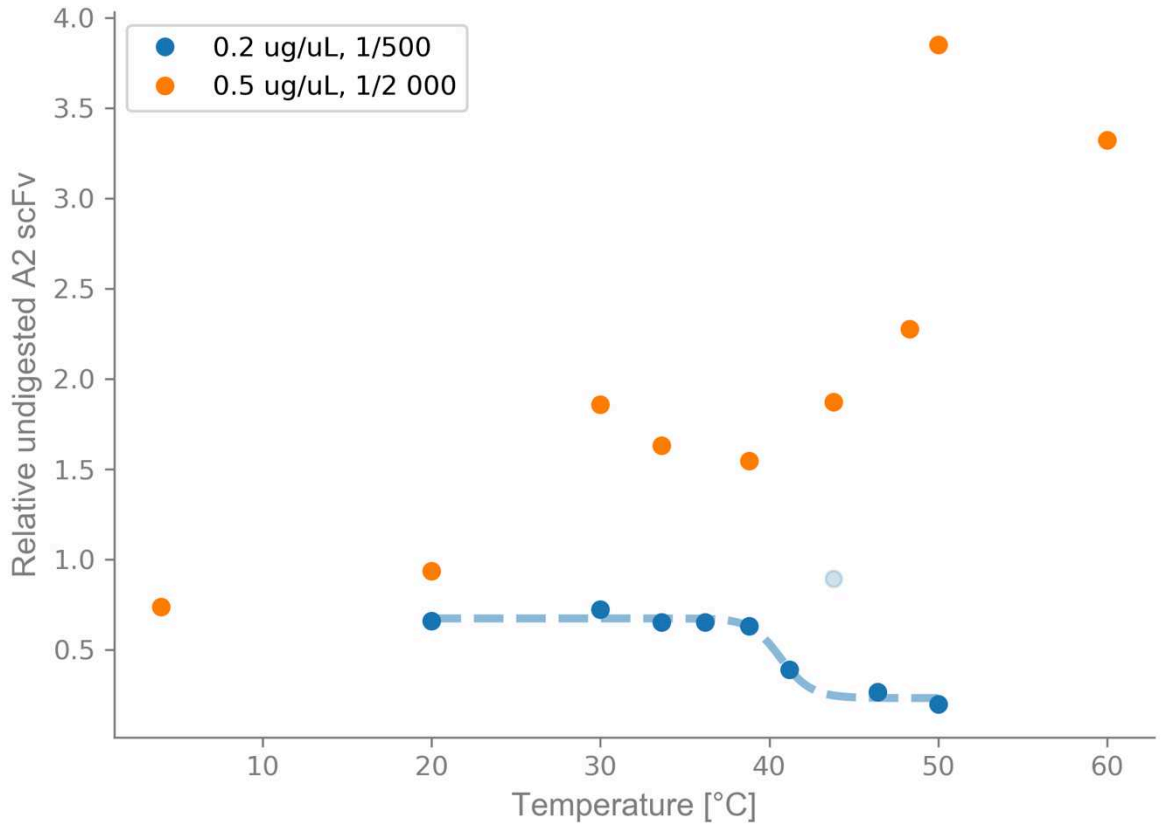


Figure 26. LiP-Chip of A2 scFv at different protein concentrations and E/S ratios.

(Orange data) LiP experiment with E/S ratio of 1/500 and an incubation time of 1 min without preheating. The A2 scFv concentration was 0.2 µg/µL in a reaction volume of 20 µL. (Blue data) LiP experiment with E/S ratio of 1/2 000 and an incubation time of 1 min without preheating. The A2 scFv concentration was 0.5 µg/µL in a reaction volume of 20 µL. The data point in light blue corresponds to an outlier due to the baseline effect (Suppl. Figure 15). It was excluded for the fitting of the sigmoidal curve. The rest of the data was fitted with Equation 14. The estimated $T_{m,app}$ is 40.7°C.

Low protein concentration

Agilent reports for the 2100 Bioanalyzer protein chips with a range from 5 to 80 kDa (P80) a quantitative range between 60 and 2 000 ng/µL and a qualitative range between 6 and 4 000 ng/µL with a sample volume of 4 µL (Agilent Technologies). The sample preparation protocol for the protein chips includes a large dilution with water to reduce the salt concentration. This results in a range of 16 to 533 ng on the chip for a quantitative assessment of a protein or peptide and a range of 1.6 to 1 067 ng for a qualitative assessment. For the LiP-Chip method, we are interested in measuring the amount of undigested protein. Thus, we have to make sure that at least for low and medium temperatures, the amount of undigested protein after the LiP step is within the quantitative range of the protein chip. We can influence the digestion with the chosen E/S ratio and incubation time. A lower E/S ratio or a shorter incubation time can lead to lower digestion and can help to shift the amount of undigested protein into the quantitative range. The appropriate E/S ratio and incubation time will depend on the protein of interest. We advise to rather adjust the E/S ratio than the incubation time. A short incubation time is more favorable with respect to aggregation. It could therefore be helpful to conduct a first LiP experiment with a known amount of the protein of interest at a few temperatures (e.g. one below and one above the expected T_m) with a short incubation time (e.g. 1 min) and several E/S ratios. Including a sample of the protein of interest at a known concentration without the protease can provide insight into the resulting absolute amount of undigested protein after LiP.

But what can we do if the available amount of protein is so low that we cannot adjust the LiP conditions enough to be able to analyze the LiP results on the protein chip? As mentioned above, the sample preparation protocol from Agilent contains a large dilution step. This results in an amount of 16 ng of protein on the chip from a 4 µL sample with 60 ng/µL of that protein. As a proof of principle, we have explored the reduction of the dilution with LiP-Chip experiments of BSA in PBS buffer. Additionally, we have applied the adjusted protocol to the third PDZ domain of the postsynaptic density protein 95 ($PSD\text{-}95^{PDZ3}$) (see 'PDZ domain' in the supplementary information).

With our adjusted protocol, we achieve a total amount of 200 ng on the chip from a 3 µL sample with 67 ng/µL protein concentration. In order to assure the correct concentrations of the protein chip reagents despite the reduced dilution of the sample, we diluted the reagents beforehand in water (see 'Material and Methods'). To avoid problems with the protein chip due to high salt concentrations, we investigated the use of PK in solutions with reduced salt concentrations

(Suppl. Figure 16). The LiP of BSA with PK is not affected by salt conditions as low as 0.01 x PBS. With the adjusted protocol, we can clearly identify differences of undigested BSA between LiP at 37°C and 70°C with an initial BSA concentration of only 67 ng/µL and a LiP sample volume of 15 µL (see Suppl. Figure 17 and 'Material and Methods'). After sample preparation for the protein chip analysis with SDS and DTT, our protocol yields 30 µL. The analysis on the protein chip requires only 6 µL of prepared sample. With the use of high-precision pipets for small volumes or pipetting robots, the amount of protein necessary for a LiP-Chip experiment could be further reduced.

Our preliminary results open doors for the LiP-Chip analysis of proteins at low concentrations. Depending on the protein of interest and its protease susceptibility, the protein concentration could be further reduced by adjusting the E/S ratio or incubation time as mentioned above, or by reducing the reaction volume. For a wide-spread application of the adjusted protocol however, further optimization might be required. The use of a different sample buffer for instance could have a strong influence on the quality of the results. In some cases, a reduction of salts by dialysis might be necessary, increasing the overall preparation time. Furthermore, we recommend the use of an internal marker at a known concentration (e.g. BSA), since the fluorescence intensity of the upper marker provided by the commercial kit is low under the adjusted protocol.

Recommended protocol

In this section, we want to give a general overview of the protocol we recommend for the estimation of thermal stability of a protein with the LiP-Chip method. Detailed protocols can be found in the supplementary information.

Initial evaluation of LiP parameters

As a very first step, the researcher should evaluate the appropriate E/S ratio and range of temperatures for the LiP-Chip stability estimation. Interesting E/S ratios to test are 1/100, 1/500, 1/1 000, and 1/2 000. The tested temperatures could cover a range from 20°C to 70°C in 10°C steps, depending on how much is known about the expected T_m of the protein of interest. Other parameters, such as incubation time and protein concentration might have to be adjusted as well. For a first evaluation however, we recommend an incubation time of 1 min and a protein concentration between 0.2 µg/µL and 0.5 µg/µL. Depending on the number of different conditions, we used SDS-PAGE or protein chips for this first parameter evaluation to measure the amount of undigested protein. It is also useful to include a sample of the protein at known concentration without the protease to be able to assess the absolute amount of digestion during LiP under different conditions.

Choosing appropriate LiP-Chip parameters

With the results from the initial LiP parameter evaluation, the researcher can choose the appropriate parameter values for the estimation of the thermal stability of the protein of interest with LiP-Chip. The researcher should choose the E/S ratio that ensures that the digestion is not too strong for a quantification of the amount of undigested protein on the protein chip. In order to find the best range of LiP temperatures, the researcher must look for the two temperatures in the initial evaluation that show a clear drop in the amount of undigested protein from the lower to the higher temperature. An ideal temperature range to estimate the $T_{m,app}$ of the protein of interest is a range of 30°C around the temperatures with the drop in undigested protein. With 10°C below the lower temperature, the 10°C that cover the drop, and 10°C above the upper temperature. In that way, the assayed temperatures should cover the pre-transition phase, transition phase, and post-transition phase (Figure 2). As mentioned in the section about the 'Data analysis', two temperatures should be picked evenly in the pre-transition phase, two temperatures in the post-transition phase, and five temperatures in the transition phase. In order

to better estimate the boundaries for the data analysis, every LiP-Chip experiment should contain a sample of the protein of interest at know concentration without the protease.

Advice on protein samples

As a general rule for biological experiments, we advise the researcher to make stock solutions at appropriate concentrations and store them as aliquots that correspond to the volumes needed for one experiment (including extra volume for the pipetting error). For protein studies in particular, the solutions containing proteins should be shock frozen in liquid nitrogen and stored at -80°C (for long storage) or -20°C (for shorter storage). The thawed protein solutions should be kept on ice to reduce aggregation or autolysis in the case of PK. Nevertheless, aggregation can occur during the freeze-thaw process. Centrifugation of the protein sample for 5 minutes at 15 000 x g can help to remove such aggregates. Given the autolysis of PK however, we recommend the aggregation removal only for the protein of interest. PK on the other hand should be used immediately after thawing. A well prepared and timed organization of the experiment is helpful.

LiP step

To cover the temperature range of 30°C, we recommend the use of PCR machines with temperature gradient. PCR machines provide additional means of automation through the use of PCR strips or plates and multi-channel pipettes, or even pipetting robots. As a temperature profile for the LiP step of the LiP-Chip method we propose the following:

- 1 min at the specific LiP temperature.
- 5 min at 97°C to inactivate PK (ideally with a uniform ramp time for all LiP temperatures).
- 5 min at 10°C to let the LiP samples cool down and facilitate easy handling.

The cooled LiP samples should be shock frozen in liquid nitrogen and stored at -20°C for further analysis on the protein chip.

Chip step

We will report here only the preparation for the Agilent P80 protein chips for the 2100 Bioanalyzer. The general approach is in accordance with the protocols provided by Agilent. For

low protein concentrations however, we adjusted the sample preparation as reported in a detailed protocol in the supplementary information (‘Material and Methods’).

The reagents from the protein chip kit should be adjusted to room temperature for 30 minutes prior to the preparation of the chip. This ensures the proper functioning of the gel mixes, as well as the resuspension of the SDS in the buffer. The reagents should not be kept at room temperature for more than 60 minutes to avoid degradation. Highly important is also to protect the reagent with the fluorescent dye from light.

As described in ‘Quantification of LiP results’, the detection of proteins and peptides on the protein chip is facilitated by the intercalation of the fluorescent dye with the protein-SDS micelles. For this, the protein sample has to be first treated with SDS for 5 min at 95°C. In the case of proteins containing disulfide bonds, DTT should be present in the SDS-containing buffer.

The preparation of the protein chip with the gel-dye mix and the destaining solution is done on the Agilent chip-priming station and follows the providers protocol. Important to note is that the gel-dye mix contains DMSO and should be handled with care. The Agilent protocol gives specific advice on how to pipet gels and samples into the wells of the protein chip. These advice should be meticulously followed to prevent the formation of air bubbles or the spillage of sample that can disturb the analysis of the chip. Especially important is for example to insert the pipet tip to the bottom of the well and tilt it slightly before dispersing its content. The chip priming station is universal for all Bioanalyzer chips, including different protein chips, DNA chip, RNA chips, and flow cytometry, but has to be specifically adjusted for each chip type. We have observed problems with the air seal between the chip priming station and the P80 protein chips, leading to the baseline effect, despite the adjusted settings for protein chips (see ‘Detailed analysis of the baseline effect’). This highlights the need for improvements in protein chip preparation or protein chip design.

The following table includes several problems that we encountered in the course of the development of the LiP-Chip method and their possible causes and solutions.

Table 4. Troubleshooting advice

Problem	Possible causes and solution
No change in the amount of undigested protein observable for the LiP temperature, but moderate difference between the LiP samples and the protein sample without PK	The temperature range might not include the $T_{m,app}$ -> Change the temperature range
No change in the amount of undigested protein for the LiP temperatures and weak difference between LiP samples and protein sample without PK	The protein digestion during LiP is too low -> Increase E/S ratio or incubation time
No change in the amount of undigested protein for the LiP temperatures and strong difference between LiP samples and protein sample without PK	The protein digestion during LiP is too strong -> Decrease E/S ratio or incubation time
The amount of undigested protein increases with temperature	This might be due to protein aggregation -> Reduce the protein concentration in the reaction
White sediments after LiP	Strong protein aggregation -> Reduce the protein concentration in the reaction
Peak of undigested protein is not detectable for LiP samples and sample without PK	The protein concentration is below the qualitative range of the protein chip -> Increase the protein concentration or decrease the sample dilution (see 'Low protein concentration')
The upper marker intensity increases strongly for the last few samples	Baseline effect -> Reanalyze samples on another chip
Upper marker not detectable and protein peaks at unusually high migration times	The gel-dye mix is too old -> Make sure that you use the gel-dye mix for a maximum of four weeks

Summary of LiP-Chip

In the previous sections, we have shown that the LiP-Chip method is well applicable to estimate the thermal stability of our validation proteins BSA and α-Lac. The different $T_{m,app}$ of BSA and α-Lac of approximately 60°C and 38°C respectively, highlights the applicability of LiP-Chip over a wide range of protein stabilities. The $T_{m,app}$ of BSA and α-Lac estimated with LiP-Chip are in accordance with reported melting temperatures obtained with different thermal stability methods. However, we need to consider that LiP-Chip experiments are not at the thermodynamic equilibrium and are influenced by the E/S ratio and incubation time. Therefore, the $T_{m,app}$ might differ from the T_m as estimated with other methods, such as DSC or CD. Nevertheless, we expect that the $T_{m,app}$ correlates with the T_m. LiP-Chip could thus be used to quantitatively rank proteins with different stabilities, such as mutants or orthologs. For a more detailed analysis, our validation with BSA and α-Lac can be extended to establish a calibration curve relating $T_{m,app}$ estimated with LiP-Chip and T_m from other methods. Such a calibration curve can for example be obtained by comparing the $T_{m,app}$ from LiP-Chip with published T_ms for several point mutants of a protein. Widely studied proteins, such as PDZ domains or Dihydrofolate reductase (DHFR), can provide T_m data from different methods.

In addition to the general applicability of LiP-Chip to estimate the thermal stability of proteins, we observed good robustness of the $T_{m,app}$ estimation with respect to the E/S ratio and incubation time. Nevertheless, these two parameters remain important for the application of LiP-Chip and should not be neglected when analyzing a new protein.

We showed that the LiP-Chip method has a high experimental reproducibility of less than 1°C for the experiments with the validation proteins. The technical reproducibility however, is occasionally hindered by the baseline effect. It is worth to investigate solutions to this baseline effect or to switch to other protein chips or tailored microfluidic devices for further studies.

Another advantage of LiP-Chip is its ability to identify protein aggregation. This is an interesting avenue that further studies could explore.

Application of LiP-Chip

In the previous chapter, we have shown that the LiP-Chip method is well applicable to estimate the thermal stability of commercially available proteins, such as BSA and α-Lac. In this chapter, we want to discuss the use of LiP-Chip on non-commercial proteins, from the widely known enzyme Dihydrofolate reductase (DHFR) to antibody fragments. We will especially discuss current problems and possible solutions of the LiP-Chip application to various proteins.

Antibodies

Antibodies are a fascinating class of proteins. They are part of the adaptive immune system and span an incredible diversity due to their continuous evolution throughout a vertebrate's life time. The general function of antibodies is to bind to antigen and to trigger downstream events. This is facilitated by a highly conserved structure of two symmetric modules that bind antigen on one end (antigen-binding fragment (Fab)) and trigger downstream events on the other end (crystallizable fragment (Fc)) (Figure 27) (84). We can further distinguish two chains (light chain and heavy chain), as well as highly conserved constant regions (C_L and C_H1 to C_H3) and variable regions (V_L and V_H). The binding of antibodies to antigen is mostly mediated by six hypervariable loops, called complementarity determining regions (CDRs), that are evenly spread over V_H and V_L and interspersed by framework regions (FRs) (85,86). Of the six CDRs, the V_H CDR3 seems to be most determining for the specificity of the antigen binding.

Antibodies are extensively studied as therapeutics in pharmaceutical research. For the interests of this lab however, the continuous evolution of antibodies through affinity maturation is most intriguing. Very generally speaking, during affinity maturation antibodies undergo cycles of mutations and selection against a specific antigen until they sufficiently bind that antigen. Understanding the mechanisms and the constraints behind the natural evolution of antibodies can shed light on the general relationship between a protein's sequence and its properties, such as stability, specificity, affinity, and evolvability. For the diverse projects in this lab, we focus on a widely used synthetic version of antibodies, called single-chain variable fragment (scFv) (Figure 27). The scFv antibodies consist only of the V_H and V_L regions connected by a short linker and are able to bind to antigen (87,88). They are easily expressible in standard bacterial models, such as *Escherichia coli*.

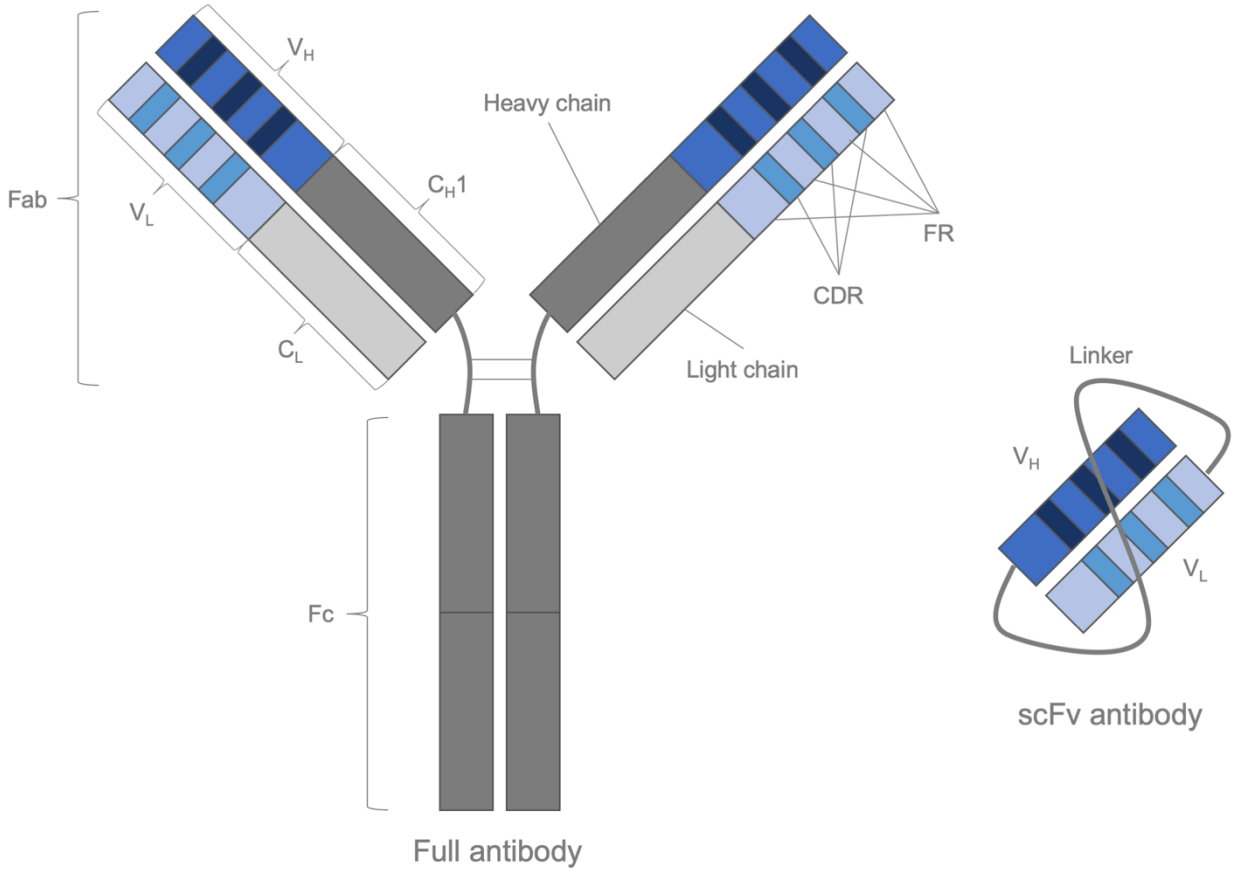


Figure 27. Schematic of the structure of a full antibody and an scFv antibody.

(Left) The image shows a rough schematic of an antibody, consisting of two symmetric modules that can be further separated into heavy chains and light chains. The antigen-binding fragment (Fab) is shown on top and contains the variable regions (V_H and V_L), as well as the constant region of the light chain (C_L) and one of the three constant regions of the heavy chain (C_H1). The other constant regions of the heavy chain constitute the crystallizable fragment (Fc). In the variable regions we can further identify three complementarity determining regions (CDRs) interspersed with framework regions (FRs). (Right) This is a schematic of a single-chain variable fragment (scFv) antibody. It consist of the V_H and the V_L regions fused by a linker.

A2 scFv antibody

We chose A2 scFv to validate the applicability of LiP-Chip on scFv antibodies (89). A2 scFv was isolated from the Tomlinson I+J library during a phage-display screen for binding to the synthetic polymer poly(vinylpyrrolidone). We worked with the A2 scFv antibody due to its good expression and secretion in *E. coli* and its reasonable purification yield. These qualities hint towards a sufficient stability for LiP-Chip experiments. The Tomlinson I+J library is composed of over 100 million scFv antibodies based on one human antibody framework. The scFv antibodies contain V_H and V_L domains linked by a flexible Glycine-Serine linker. Members of the Tomlinson library have been used in many studies, but to our knowledge the thermal stability of A2 scFv has not been reported.

Expression and purification of A2 scFv

From a previous study, we have the A2 scFv as a fusion with a phage gene (pIII) on a plasmid containing a pelB signal sequence for secretion of the protein into the medium when expressed in *E. coli* (Suppl. Figure 18). The A2 scFv and the phage gene are separated by an amber stop codon, leading to an expression of the fusion protein in the TG1 *E. coli* strain and an expression of only A2 scFv in other *E. coli* strains. For the expression of A2 scFv alone (Suppl. Figure 19), we transformed the plasmid into the *E. coli* BL21 (DE3) expression strain. The purification of A2 scFv is facilitated by a poly-histidine tag (his-tag) that we combined with the Econo-column purification system from Bio-Rad and the HisPur Cobalt Resin from Thermo Fisher Scientific. We will give an overview of the purification of A2 scFv here. For a more detailed protocol, please refer to the 'Material and Methods'. In order to limit leakage expression of the antibody due to insufficient suppression of the lac operon, we used 1% glucose in our overnight medium. We then diluted the overnight culture into fresh medium with only 0.1% glucose and incubated the diluted culture at 37°C until an optical density around 0.5 was reached. At this point, we added isopropyl β-D-1-thiogalactopyranoside (IPTG) to a concentration of 1 mM to induce antibody expression. The antibody is equipped with a pelB signal peptide, leading to its transport into the periplasm of the bacterium. There, the signal peptide is cleaved and the antibody subsequently secreted into the medium. We therefore harvested only the supernatant after antibody expression overnight at 30°C. We proceeded with the purification of the antibody from the supernatant. In order to achieve high protein concentrations, we eluted into several fractions, qualitatively evaluated the presence of the antibody in the fractions with a dot blot,

and proceeded with the fractions containing an abundance in antibody (Figure 28). Before using the purified protein in LiP-Chip experiments, we dialyzed it against PBS to reduce the amount of imidazole. For the direct comparison of the applicability of LiP-Chip to the commercially available validation proteins BSA and α-Lac, we used PBS buffer during the purification, dialysis, and LiP-Chip experiments of A2 scFv.

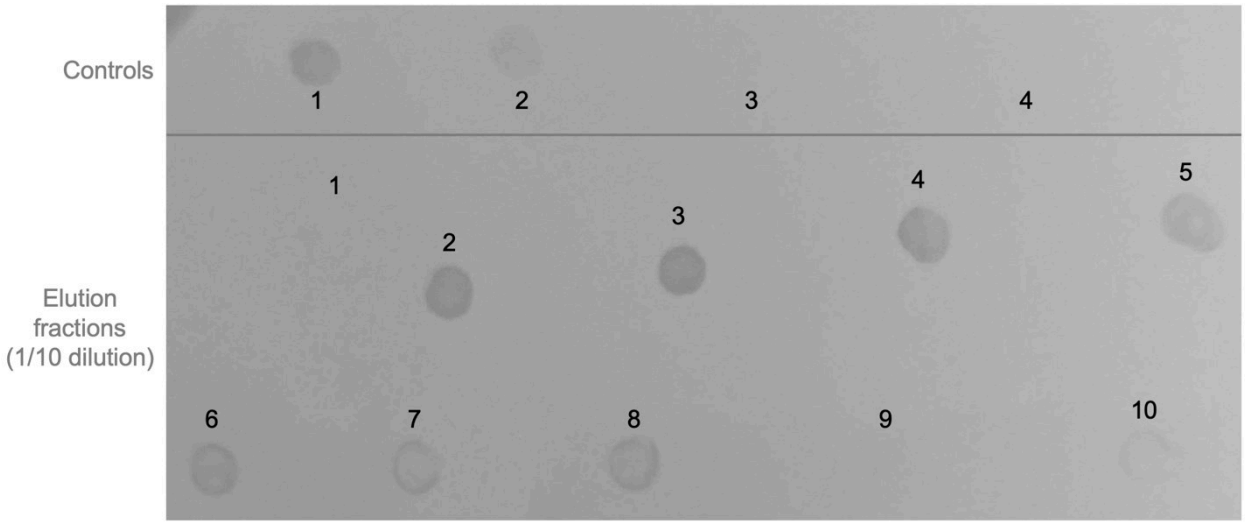


Figure 28. Dot blot of A2 scFv purification.

(Upper row) Four controls are represented: 1 = positive control with A2 scFv supernatant from previous purification. 2 = flowthrough of supernatant after incubation with the HisPur resin for approximately four hours. This shows that we are losing some A2 scFv despite the long incubation time. 3 = flowthrough of first wash. 4 = flowthrough of third and last wash. (Lower two rows) These dots show the ten elution fractions diluted 1/10 in PBS. As expected, elution 1 does not contain a sufficient amount of A2 scFv. Elutions 2, 3, and 4 were chosen for further dialysis and experiments. (All) For each dot, we used 2 µL of sample. The primary antibody was Monoclonal Anti-polyHistidine antibody produced in mouse (Sigma-Aldrich) and the secondary antibody Peroxidase AffiniPure Donkey Anti-Mouse IgG (H+L) (Interchim). For a detailed protocol see 'Material and Methods'.

Thermal stability of A2 scFv

The thermal stability of A2 scFv has not been published before. In a very first LiP experiment, we therefore evaluated the amount of undigested A2 scFv for a few widely spread LiP temperatures between 30°C and 60°C and quantified the results with SDS-PAGE (Figure 29 and Suppl. Figure 6). From this preliminary LiP experiment, we estimated the $T_{m,app}$ to be between 40°C and 50°C and evaluated the temperature range between 30°C and 60°C to be appropriate for further LiP-Chip experiments with the LiP parameters of E/S ratio 1/200 and an incubation time of 1 min.

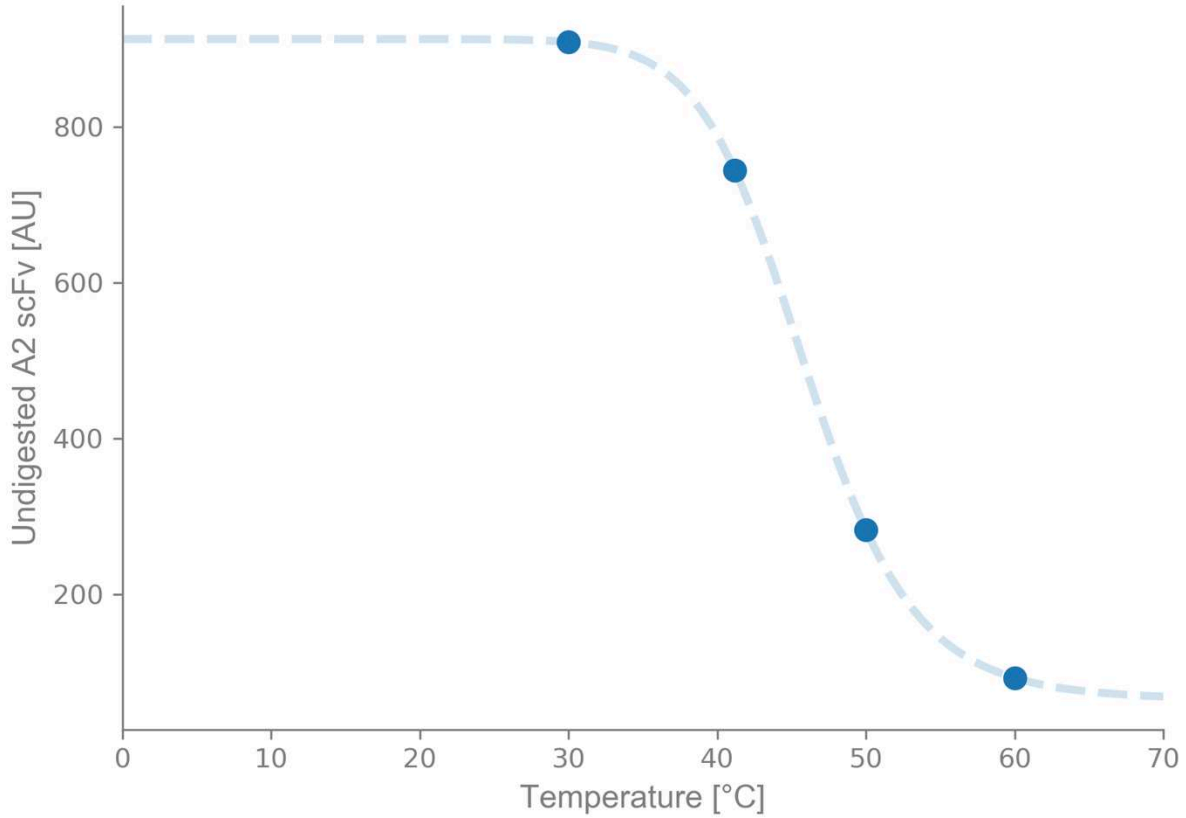


Figure 29. LiP experiment of A2 scFv with quantification on SDS-PAGE.

The LiP experiment was conducted with an E/S ratio of 1/200 and an incubation time of 1 min without preheating. The A2 scFv concentration in the reaction was 250 ng/µL in a volume of 55 µL. We used a 4-20% Tris-Glycine gel and loaded 2.5 µg of total protein amount into each well. The SDS-PAGE was run for 100 min with 25 mA, 200 V. The protein gel was stained for 1 h in PageBlue™ staining solution and destained overnight in ddH_2O. Given the limit of only four data points, the fitted sigmoidal curve serves only as a rough guideline of a possible digestion curve, but was not used to estimate a $T_{m,app}$.

The LiP-Chip experiments however, proofed to be more challenging for the purified A2 scFv than for the commercial validation proteins. We initially conducted the LiP-Chip experiments with a low A2 scFv concentrations of 200-250 ng/µL with the normal protein chip protocol from the vendor. Even after decreasing the E/S ratio to 1/500, the A2 scFv is quite susceptible to the digestion with PK already at low temperatures (Figure 30). This results in a low amount of undigested A2 scFv detectable on the protein chip over the different LiP temperatures and thus, an increased error between samples.

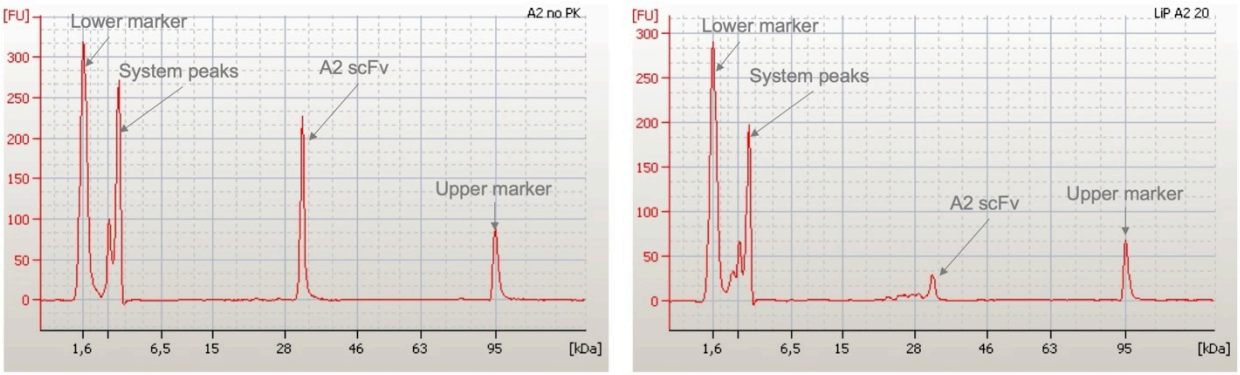


Figure 30. Comparison of A2 scFv peak spectra with and without LiP.

(Left) Bioanalyzer spectrum of a sample containing A2 scFv without PK. (Right) Bioanalyzer spectrum of LiP of A2 scFv with PK at 20°C. The LiP experiment was conducted with an E/S ratio of 1/500 and an incubation time of 1 min without preheating. The sample with PK and the sample without PK were prepared in the same way. Both had an A2 scFv concentration of 200 ng/µL in a 20 µL reaction volume.

Nevertheless, LiP-Chip experiments suggest a lower $T_{m,app}$ (between 39.9°C and 40.7°C) than proposed by the LiP quantification on SDS-PAGE (Figure 31). The difference in estimated $T_{m,app}$ from the preliminary LiP experiment quantified with SDS-PAGE and the LiP-Chip experiments can have several reasons. During the LiP-Chip experiments, we observed peaks of digested A2 scFv very close to the undigested protein, with a difference in migration time of only 0.5 milliseconds, corresponding to approximately 1.5 kDa size difference. The resolution of the chosen Tris-Glycine SDS-PAGE for proteins and peptides in the low mass range of 25 kDa is not sufficient to clearly identify such small size differences (Suppl. Figure 6). A different protein gel (e.g. Tricine gels) or a higher polyacrylamide concentration and longer running time can provide a better resolution. On the one hand, the analysis with SDS-PAGE consisted of only four data points. Thus, the results need to be considered with caution. Furthermore, the results from the LiP-Chip experiments are somewhat messy due to the low amount of undigested A2 scFv.

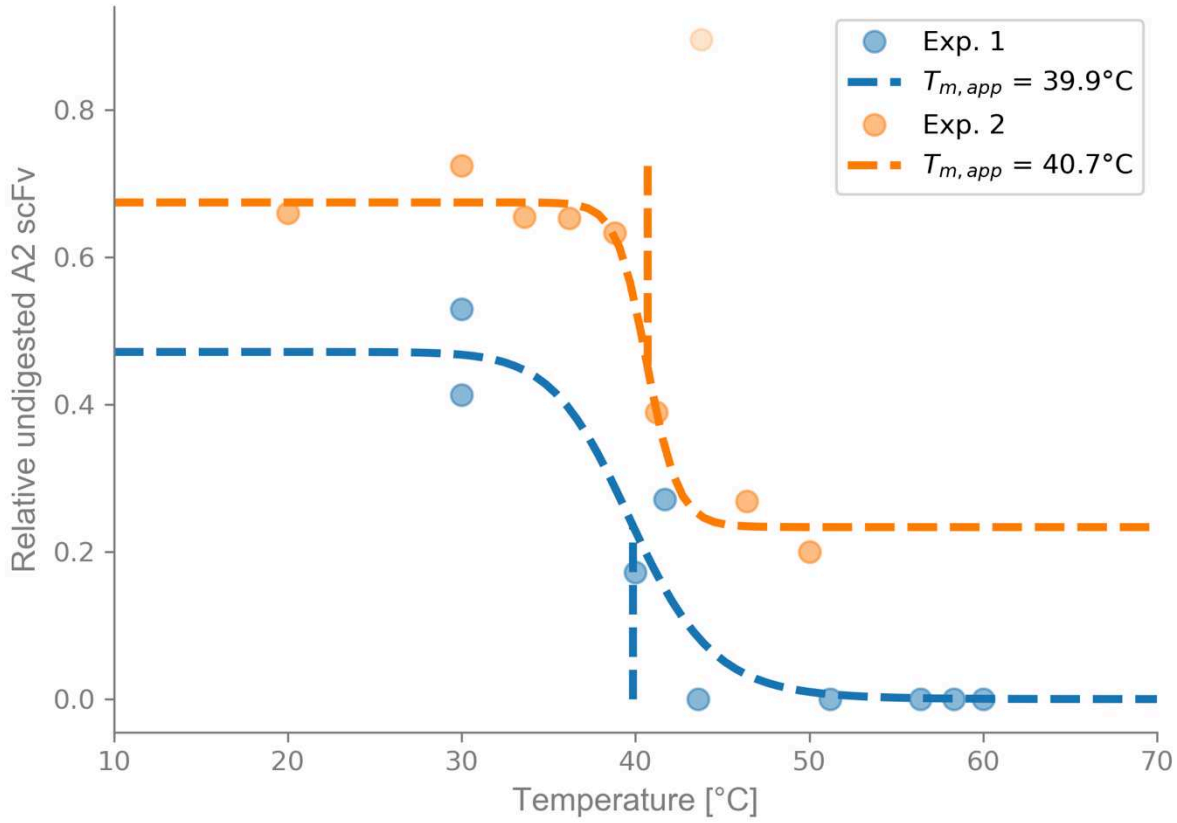


Figure 31. LiP-Chip stability of A2 scFv.

(Blue) Exp. 1 had an E/S ratio of 1/200 and an incubation time of 1 min without preheating. The A2 scFv concentration was 250 ng/µL in 55 µL reaction volume. The two data points at 30°C represent duplicate LiP experiments at 30°C and indicate the error between samples. For the fitting of the sigmoidal curve, the mean between the duplicate data points was used. (Orange) Exp. 2 had an E/S ratio of 1/500 and an incubation time of 1 min without preheating. The A2 scFv concentration was 200 ng/µL in 20 µL reaction volume. This data corresponds to the blue data in Figure 26. The data point in light orange at 43.8°C is most likely an outlier due to the baseline effect (Suppl. Figure 15) and was excluded for the fitting of the sigmoidal curve and the estimation of $T_{m,app}$.

The protein concentration in these first LiP-Chip experiments of A2 scFv were limited by low protein yields from small scale purifications. As a first measure to increase the quality of the LiP-Chip results for A2 scFv, we scaled up the purification and optimized the purification protocol. We acquired concentrations above 2 µg/µL in a 1 mL elution volume from originally 250 mL of bacterial culture. We used this newly purified A2 scFv, at a much higher concentration than the first purifications, for further LiP-Chip experiments. Additionally, we evaluated a reduced E/S ratio of 1/2 000 (Figure 32).

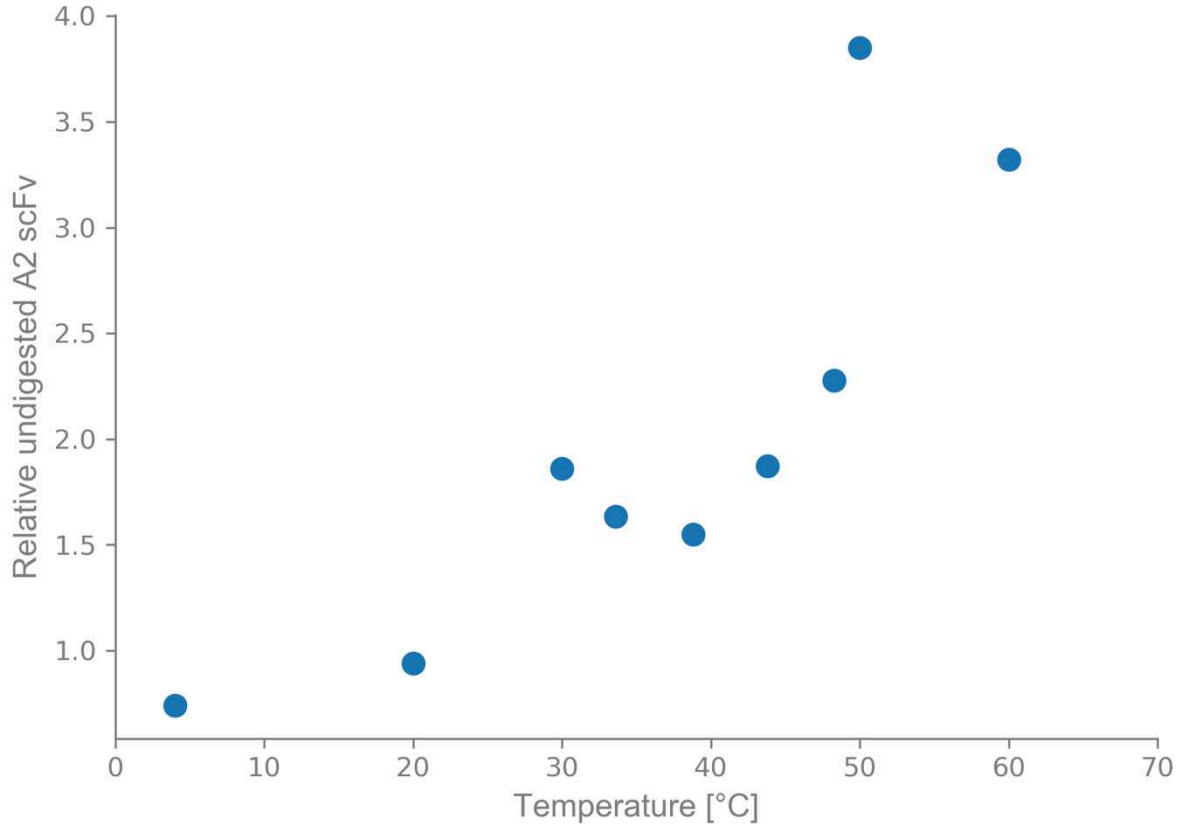


Figure 32. A2 scFv aggregation.

The LiP experiment was conducted with an E/S ratio of 1/2 000 and an incubation time of 1 min without preheating. The A2 scFv concentration was 500 ng/µL in a reaction volume of 20 µL. A strong increase of undigested protein is visible with increasing LiP temperature, indicating progressive aggregation of A2 scFv. This data corresponds to the orange data in Figure 26.

Using an increased A2 scFv concentration and a decreased E/S ratio however, did not result in LiP-Chip results of higher quality. In fact, we observe strong aggregation of the protein with increasing temperature (see also 'Protein aggregation'). We therefore wanted to return to the LiP parameters that had worked previously (E/S ratio = 1/500 and A2 scFv concentration = 250 ng/µL) and reevaluate the LiP-Chip stability of A2 scFv from there. For this, we used the A2 scFv purified at high concentration and diluted it in PBS buffer for the LiP-Chip experiment. To our astonishment, we were not able to reproduce the thermal stability data as shown in Figure 31. Instead, we observed increasingly lower stability of A2 scFv (Figure 33). Let us first discuss the data points from 20°C on. Based on observing large differences of undigested A2 scFv between LiP temperatures, replicate 1 (blue data) suggests a melting temperature of A2 scFv between 30°C and 36.2°C, whereas replicate 2 (orange data) shows a large difference

already between 20°C and 30°C. The data of the replicate experiments in Figure 33 do not allow the reasonable estimation of $T_{m,app}$. Nevertheless, they clearly show problems with the stability of A2 scFv from the purification batch with high concentration. An explanation of this could be found in the high protein concentration. ScFv antibodies represent only a small fragment compared to natural antibodies. They lack all stabilizing properties exhibited by the constant regions and can suffer from aggregation. Such aggregation in the protein stock solution could possibly affect the LiP-Chip samples in diverse ways, influencing the stability estimation with LiP-Chip. Another highly important aspect of protein stability is the buffer solution. We will discuss this in more detail further down. A different buffer could possibly stabilize A2 scFv and reduce its aggregation propensity, leading to more reproducible data even at high protein concentration.

Let us now talk about the additional data point at 4°C. We added this LiP temperature in an attempt to strengthen the pre-transition plateau, as we have done for the validation protein α-Lac (Figure 21). For A2 scFv however, we observe a large decrease in undigested protein between 4°C and 20°C for both experiments. This can have several reasons. On the one hand we have observed that A2 scFv is quite prone to digestion with PK (Figure 30). It is very likely however, that the activity of PK is much lower at 4°C than at 20°C, leading to less proteolysis of A2 scFv at the lower temperature. This would artificially increase the difference of undigested protein between 4°C and 20°C. A normalization of the LiP-Chip results with the activity of PK at the different LiP temperatures might reduce this problem. On the other hand, the decrease of folded protein at low temperatures could also be intrinsic to thermal denaturation of scFv antibodies. A study by Montoliu-Gaya *et al.* observed such a behavior in their thermal denaturation experiments of the scFv-h3D6 antibody and its mutants (30). The sigmoidal curves of the chemical denaturation experiments of the same proteins did not display such a baseline slope. Experiments with more data points at low temperatures could provide insight into this phenomenon. Another reason could be multi-state unfolding. Some scFv antibodies have been shown to exhibit multi-state unfolding (30,90,91). This might be favored by the linker between the V_H and the V_L part of the scFv. Although the linker used for A2 scFv does not contain a digestion site for PK, residues of V_H just before the linker could be hydrolyzed by PK. It would be interesting to investigate the conformation and possible unfolding intermediates further. For such a study, we recommend the use of CD for a detailed analysis of the conformation of A2 scFv and DSC for the evaluation of single- of multi-state unfolding.

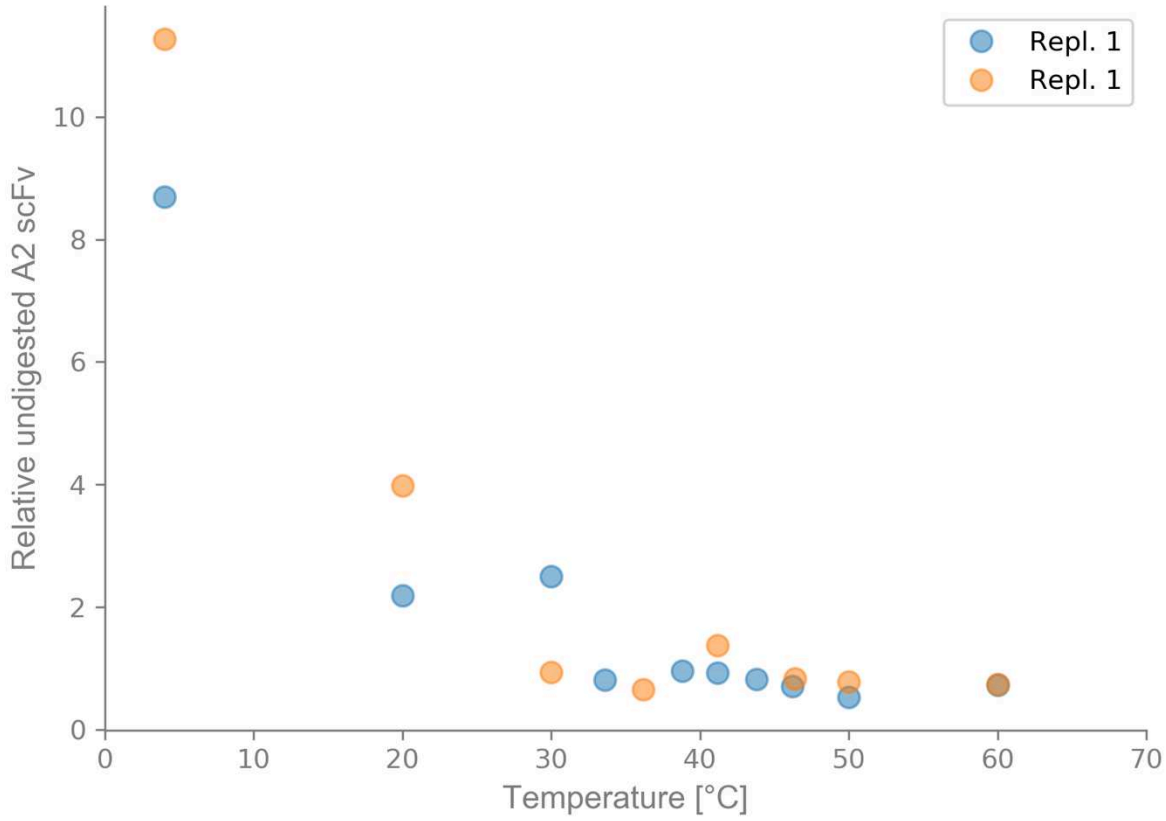


Figure 33. LiP-Chip of A2 scFv after purification at high concentration.

The two LiP-Chip data sets correspond to LiP-Chip replicates under the same conditions: E/S ratio of 1/500, incubation time of 1 min, no preheating, A2 scFv concentration of 250 ng/µL in a reaction volume of 20 µL. In both cases, two different A2 scFv aliquots with a concentration of 2.22 µg/µL from the same purification were used. (Blue) Disregarding the first strong decrease in undigested A2 scFv between 4°C and 20°C, the next strong decrease happens between 30°C and 36.2°C. (Orange) The strong decrease of undigested A2 scFv happens already below 30°C. (Both) The amount of digestion between 4°C and 20°C is very different. We describe possible reasons in the main text.

For further analysis of A2 scFv with the LiP-Chip method, we suggest the purification of the protein in a different buffer. Commonly used buffers other than PBS are for example Tris-HCl or HEPES buffer. The A2 scFv concentration during purification and LiP-Chip experiments should be kept low. To further prevent aggregation, the E/S ratio should not be too low. An E/S ratio of 1/500 with an incubation time of 1 min and an A2 scFv concentration of 200 ng/µL is a good starting point for further experiments. In order to increase the detection of the undigested protein with the protein chip, we recommend the use of our adjusted protocol for low protein concentrations (see 'Low protein concentration' and 'Material and Methods'). For the detection of the undigested A2 scFv on the protein chip, we observed good results with the adjusted protocol (Suppl. Figure 20).

Due to the time constraints of this project and the many unknown parameters of the scFv antibodies, we chose to focus our interest on a scientific question regarding a well-known and highly studied protein: Dihydrofolate reductase.

Dihydrofolate reductase

The enzyme Dihydrofolate reductase (DHFR) plays an important role in the folate metabolism, which is present in every organism. DHFR reduces 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydrofolate (THF) with the help of the cofactor nicotinamide adenine dinucleotide phosphate (NADPH), oxidizing it to $NADP^+$. The action of DHFR lies upstream of many processes important for cell proliferation and cell growth, such as the de novo synthesis of purines and pyrimidines, the production of thymidylate, and the synthesis of several amino acids. The central role of DHFR for cells makes it an important protein in many different aspects: from medical and pharmaceutical studies to a model system in biological research. Although DHFR is present throughout all organisms and exhibiting the same function, bacterial and mammalian DHFR have evolved to be different in sequence and structure. The antibiotic Trimethoprim and the chemotherapy agent Methotrexate make use of that difference, targeting either the bacterial DHFR or the mammalian protein.

DHFR is not only limited to medical and pharmaceutical studies, but is also a well-established model system for studying the relationship between conformational change and catalysis. The enzymatic activity of *E. coli* DHFR (ecDHFR) is guided by conformational changes over a range of five distinct complexes and a transition state (Figure 34) (92–94).

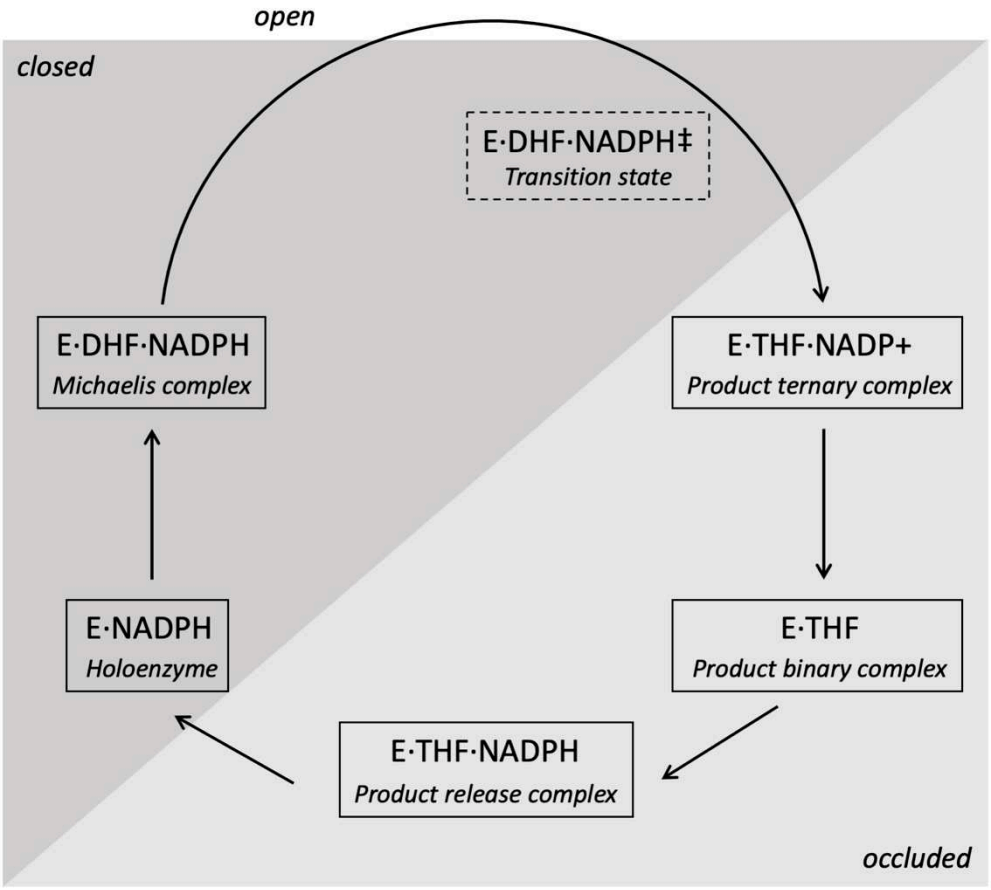


Figure 34. Complexes and conformational changes in the catalytic cycle of ecDHFR.

The catalytic cycle of ecDHFR occurs over five distinct complexes and a transition state. The Met20 loop in the holoenzyme, the Michaelis complex, and the transition state is in the closed conformation. It is postulated that the change from the Michaelis complex to the transition state occurs over an open conformation (93). For the product ternary complex, product binary complex, and product release complex, the Met20 loop is in the occluded conformation.

These conformational changes are especially focused around the Met20 loop of ecDHFR (Figure 35). This loop is located just below the active site and can move between four conformations: closed, occluded, open, and disordered. We will omit the disordered conformation that has been suggested to result from time-averaged change between the closed and the occluded conformations (93). The Met20 loop of ecDHFR assumes the closed conformation in the holoenzyme (E·NADPH), the Michaelis complex (E·DHF·NADPH), and the transition state (E·DHF·NADPH‡). In the closed conformation, the Met20 covers the active site, excluding solvent and bringing the cofactor and the substrate close together. The hydride transfer from NADPH to DHF however, is not sufficient to create THF. It needs to be accompanied by a protonation (94). The source of the protonation is still an open question. One hypothesis suggests that the Met20 loop of the Michaelis complex changes into an open conformation to allow a water molecule to enter into the active site as a proton donor (93), followed by a change to the closed conformation in the transition state. For the three product complexes (product ternary complex (E·THF·NADP+), product binary complex (E·THF), and product release complex (E·THF·NADPH)), the Met20 loop is in the occluded conformation. This conformation excludes part of the cofactor from the active site and facilitates the release of the product THF. The relationship between conformational change and catalysis of ecDHFR suggests that protein stability would influence the protein's activity (34).

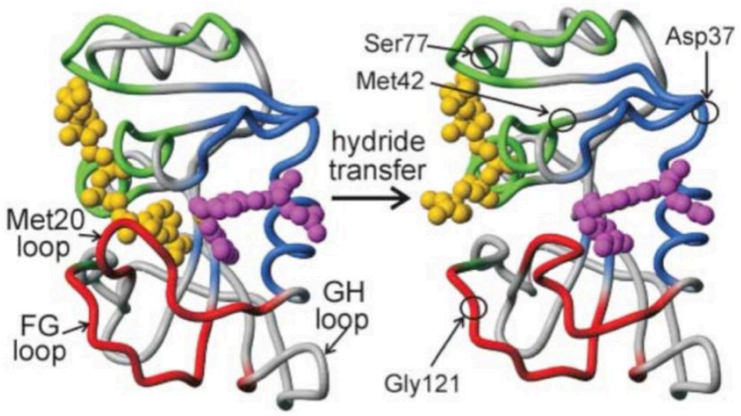


Figure 35. Conformations of ecDHFR Met20 loop.

The left structure represents E·folate·NADP+ (PDB 1RX2) as a model for the Michaelis complex with the Met20 loop in the closed conformation. The right shows the structure of E·5,10-dideazaTHF(ddTHF)·NADPH (PDB 1RX6) as a model for the product ternary complex with the Met20 loop in the occluded conformation. The cofactor is represented in yellow, the folate and ddTHF in purple. (Red) Active site loops, (blue) substrate-/product-binding residues, (green) cofactor-binding markers. [Source : (95)]

Choice of mutants

To investigate the relationship between DHFR activity, stability and in vivo function, Thompson *et al.* measured the growth rate effects of all possible ecDHFR mutations in cellular environments with and without the Lon protease (29). Among other tasks, Lon is responsible for protein quality control by digesting misfolded or mutant proteins (96). Selecting for cell growth as a proxy of ecDHFR activity on a library of all single-point mutants of ecDHFR gives highly different results in the environment with and without Lon. In the absence of Lon, many more mutants are advantageous than in the presence of the protease. This difference in mutational effect was quantified with the change in selection coefficient (SC) between the mutation in the environment with Lon and without Lon.

$$\Delta SC = SC_{+Lon} - SC_{-Lon}$$

Thus, a negative ΔSC indicates a mutant that is more advantageous in the absence of Lon. The role of Lon as a quality control suggests that those mutants might be highly active, but not very stable. Twenty-four single-point mutants were chosen for an analysis of their T_m with the CD method (Figure 36). The results show a correlation between the ΔSC and the T_m, confirming the hypothesis that the mutations advantageous in the absence of Lon and disadvantageous in the presence of Lon are destabilizing mutations.

Table 5. ecDHFR and core mutants chosen for this study.

Mutation	Apparent T_m (CD measurement)	$\Delta Selection\ coefficient$
ecDHFR	57.4°C	0
I41A	39.0°C	-1.24
I41V	53.3°C	-0.07
C85L	42.4°C	-0.78
I91A	55.2°C	-0.6

[Data source: (29)]

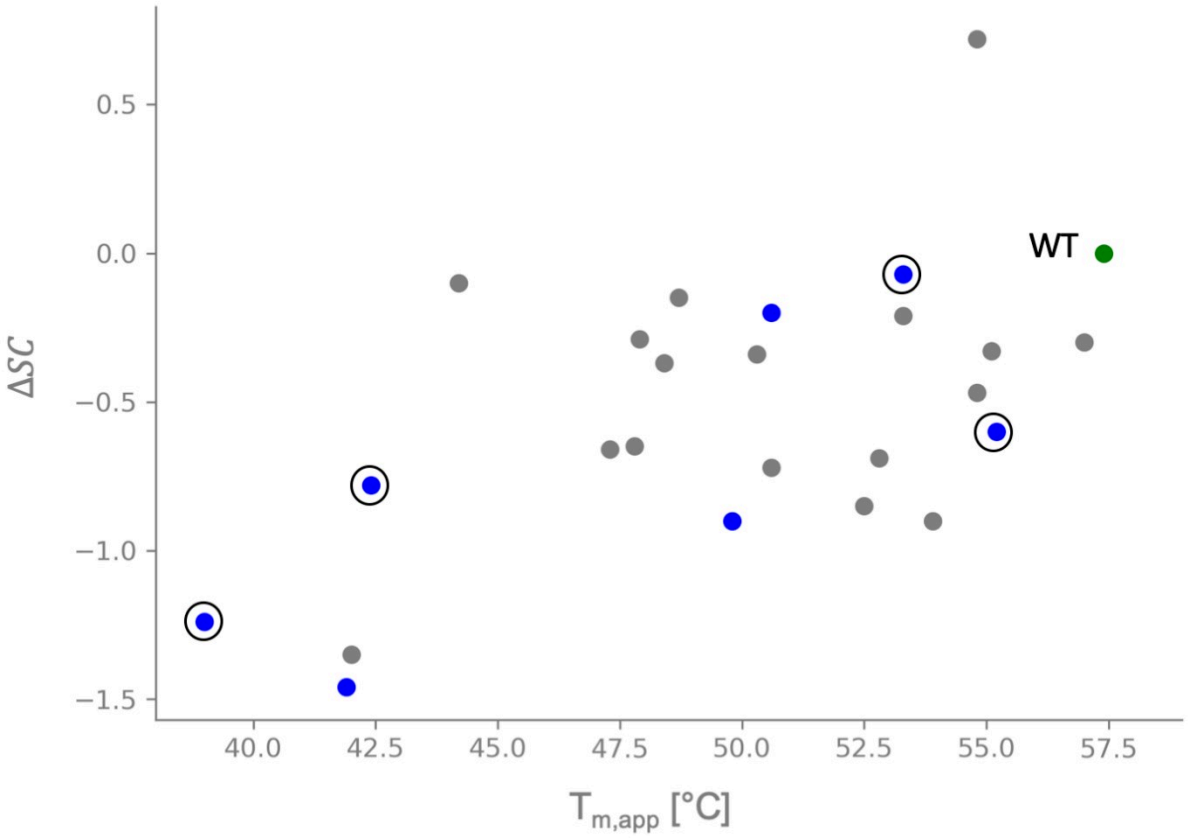


Figure 36. Melting temperature and ΔSC of ecDHFR single-point mutants.

The blue colored data points represent mutants in the core region of ecDHFR. The four mutants further considered in this work are marked with an additional black circle. The green data point shows the wild-type (WT) ecDHFR. [Data source: (29)]

To more deeply test the relationship between thermal stability and proteolytic sensitivity, we planned to first estimate the thermal stability of the wild-type ecDHFR and four of the twenty-four mutants with the LiP-Chip method. The set of twenty-four mutants contains seven single-point mutations in the core region of ecDHFR (Figure 36). These seven mutants span a wide range of T_m and ΔSC values. We chose to focus on four of these core mutants from the lowest to the highest T_m with diverse ΔSC (Table 5). Additionally, we have chosen two mutations at the same position with very different effects on both parameters. We expect that the LiP-Chip measurements will be well correlated to the T_m estimations with CD, as well as to the ΔSC.

Expression and purification of DHFR

We received plasmids containing the wild-type ecDHFR and the four selected mutants from the Reynolds laboratory and the Kortemme laboratory (Suppl. Figure 21 and Suppl. Figure 22). We transformed the plasmids into *E. coli* BL21 (DE3), a strain that is optimized for protein expression. To facilitate purification, each construct contained a his-tag. We used the Econo-column system from Bio-Rad in combination with the HisPur Cobalt Resin from Thermo Fisher Scientific for the in-house purification of ecDHFR and its mutants. In accordance with our experiments with the validation proteins BSA and α-Lac, we used PBS buffer throughout the purification and the subsequent LiP-Chip experiments. We provide a detailed purification protocol in the 'Material and Methods' and will give only a rough overview here. For a high yield in ecDHFR expression, we diluted overnight cultures of the *E. coli* strain BL21 (DE3) containing the plasmid into terrific broth (TB) and let them grow to an optical density of approximately 0.7. We then induced the expression of the wild-type ecDHFR or the mutants with IPTG. In order to allow slow cell growth and slow protein expression that favors proper folding of the expressed proteins over an extended amount of time, we incubated the induced cultures overnight at 18°C. The ecDHFR and its mutants are expressed intracellularly. We therefore harvested the cells after overnight expression of the protein of interest, released the cell contents by sonication, and removed cell debris through ultra-centrifugation. The his-tag allowed us to column purify ecDHFR and its mutants with the HisPur Cobalt Resin. From protein expression in 1 L of culture, we eluted in volumes of approximately 1 mL with protein concentrations between 9 and 14 mg/mL. For the ecDHFR mutant I41V, we validated the successful purification with dot blot and SDS-PAGE (Figure 37 and Suppl. Figure 23).

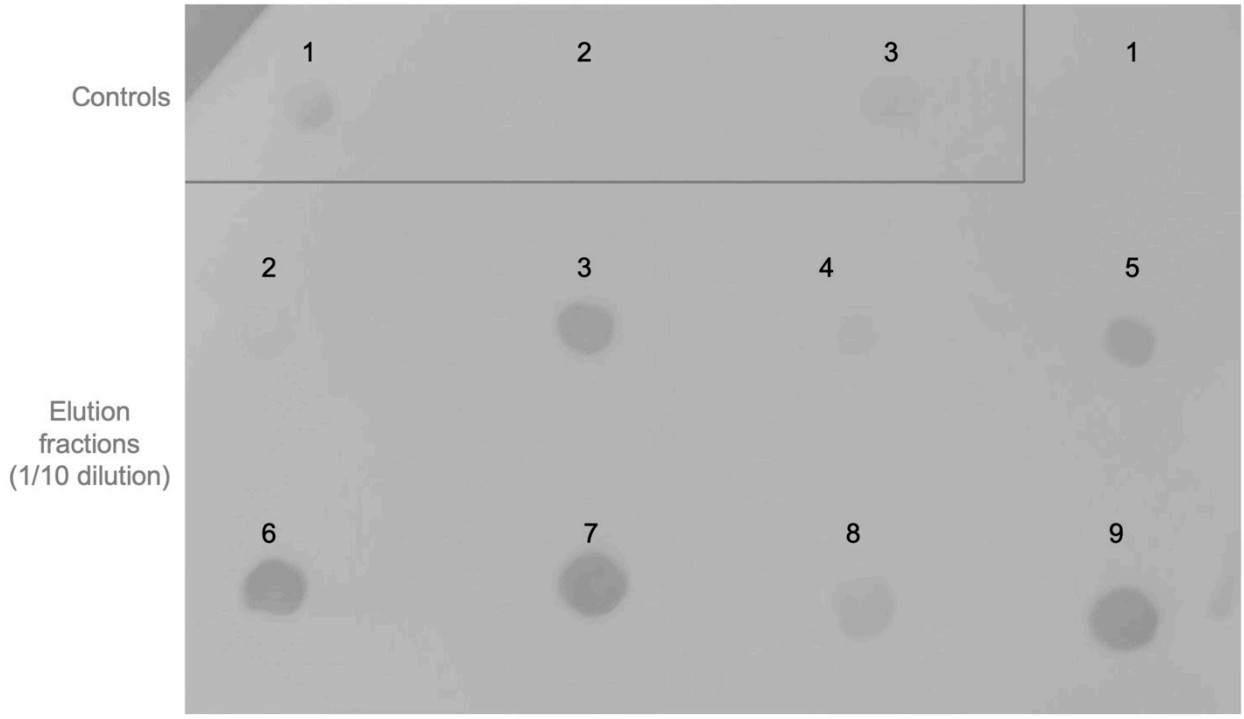


Figure 37. Dot blot of ecDHFR I41V elution fractions.

All dots correspond to 2 µL of sample. (Upper left) The section labeled 'Controls' contains three dots: 1 = positive control with A2 scFv, 2 = negative control with BSA, 3 = flowthrough after several washes. (Rest) The other nine dots represent the fractions from elution on the column. The fractions were diluted 1/10 in PBS. Fractions 1 and 2 have expectedly low amounts of ecDHFR I41V. Although elution fraction 4 does not show positive results on this dot blot, it was confirmed as containing a sufficient amount of ecDHFR I41V with other dot blots of the same fractions.

LiP-Chip of ecDHFR I41V

The first ecDHFR variant we had successfully purified was the I41V mutant. This particular ecDHFR mutant has been shown to have a T_m of 53.3°C (29). In order to establish appropriate LiP parameters, we evaluated the digestion of ecDHFR I41V between 30°C and 70°C with two different E/S ratios (1/200 and 1/1 000) (Suppl. Figure 24). The E/S ratio of 1/200 is too high and leads to a digestion of ecDHFR I41V below the detectable amount. An increase of undigested protein into the detectable range at 70°C however, indicates aggregation of ecDHFR I41V. For the E/S ratio 1/1 000, the band corresponding to the undigested protein is well visible for all temperatures. Due to the suboptimal quality, band intensity, and resolution of the protein

gel, the densitometry analysis is rather messy. Nevertheless, we can use it to confirm the trend suggested by a visual analysis of the protein gel. In disagreement with the assumed T_m of 53.3°C, we observe a strong decrease in undigested ecDHFR I41V already between 30°C and 40°C (Figure 38). As for the E/S ratio of 1/200, the LiP with E/S ratio 1/1 000 hints to aggregation of ecDHFR I41V at 70°C.

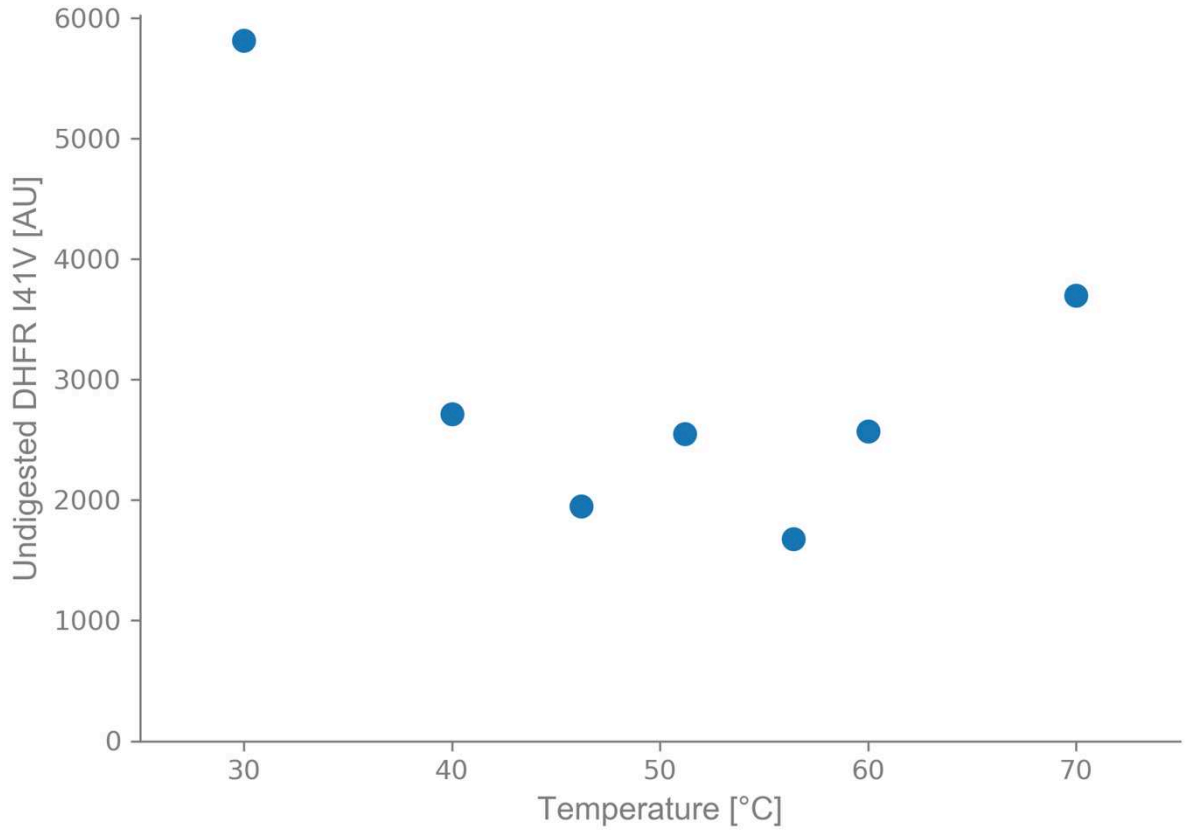


Figure 38. Quantification of LiP of ecDHFR I41V with SDS-PAGE.

The LiP experiment was conducted with an E/S ratio of 1/1 000 and an incubation time of 1 min without preheating. The ecDHFR I41V concentration was 500 ng/µL in a reaction volume of 20 µL. We quantified the LiP results on a 4-20% Tris-Glycine gel, run for 90 min at 25 mA and 200 V. Each well contained 5 µg of total amount of protein. The gel was stained with PageBlue™ staining for 1 h and destained overnight in ddH_2O. The densitometry analysis followed instructions from the SYBIL project (78).

A LiP experiment of ecDHFR I41V with E/S ratio of 1/500 over a temperature range from 20°C to 80°C shows a decrease in undigested protein even earlier: Between 20°C and 30°C (Suppl. Figure 23). This hints at similar stability problems as observed with the A2 scFv antibody.

We further evaluated appropriate E/S ratios in the range from 1/500 to 1/10 000, as well as a different protein concentration (1 µg/µL), for the application of LiP-Chip for the estimation of the thermal stability of ecDHFR I41V (Figure 39). We tested the different E/S ratios for two temperatures, one below (37°C) and one above (65°C) the expected T_m. This LiP-Chip experiment highlighted the extend of the aggregation of ecDHFR I41V. All samples showed clear aggregation after LiP at 65°C in the form of white precipitates. For the highest E/S ratio of 1/500, we observed white precipitation even after LiP at 37°C. We attempted to remove those aggregates by centrifugation before analyzing the samples on the protein chip. At E/S ratios above 1/5 000, the proteolysis of ecDHFR I41V at 37°C is already quite strong, leaving only a small range for the evaluation of the decrease in undigested protein around the melting temperature. The E/S ratios of 1/5 000 and 1/10 000 provide a larger change in undigested protein between LiP at 37°C and 65°C and are thus, more appropriate for further experiments.

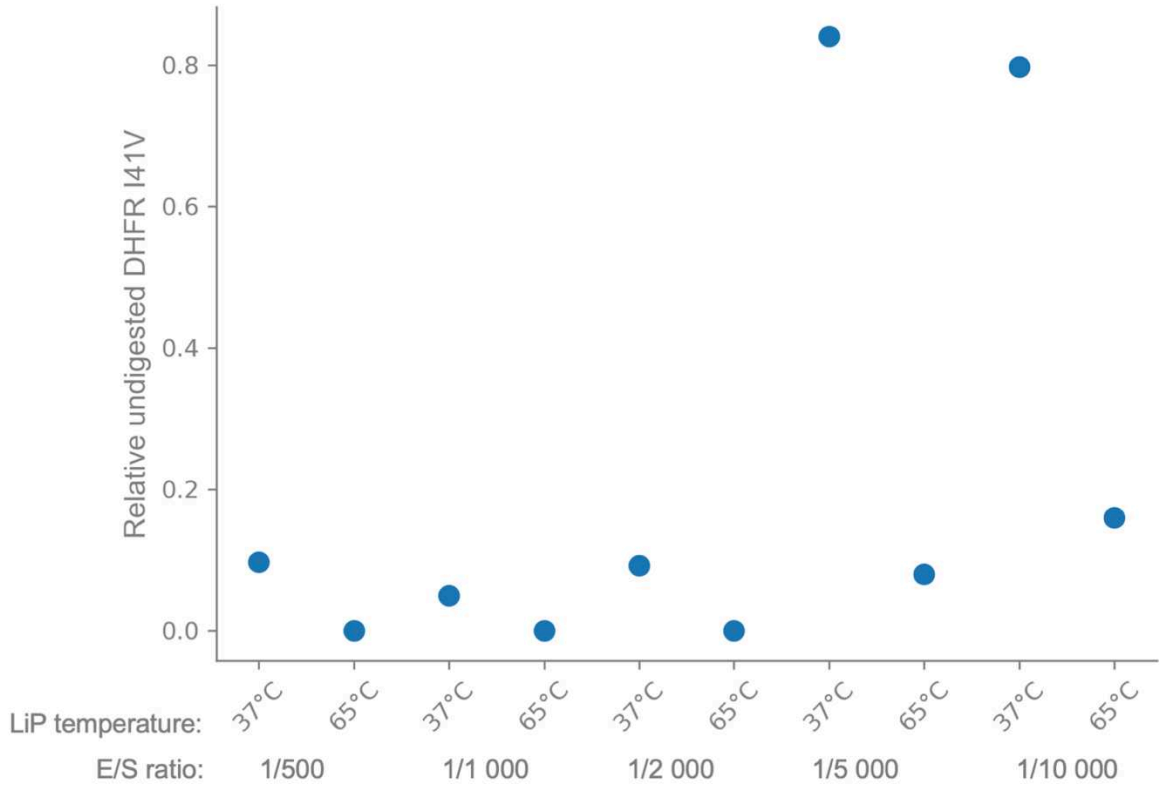


Figure 39. LiP-Chip of ecDHFR I41V at two temperatures with different E/S ratios. The data shows the relative amount of undigested ecDHFR I41V after LiP at 37°C and 65°C with different E/S ratios (1/500, 1/1 000, 1/2 000, 1/5 000, 1/10 000). The incubation time was 1 min without preheating. The concentration of ecDHFR I41V was 1 µg/µL in a reaction volume of 20 µL. The LiP samples were analyzed with the P80 protein chips on the 2100 Bioanalyzer and processed as described in 'Data analysis'.

We chose to proceed with the E/S ratio of 1/10 000 due to its lower amount of visible aggregation compared to the samples with E/S ratio of 1/5 000. We conducted a LiP-Chip experiment of ecDHFR I41A over ten temperatures spread between 37°C and 65°C (Figure 40). In order to remove possible aggregates, we centrifuged the samples before analyzing them on the protein chip. Unfortunately, the results do not show the expected sigmoidal curve. They rather display large deviations in the amount of undigested ecDHFR I41V from LiP temperature to LiP temperature.

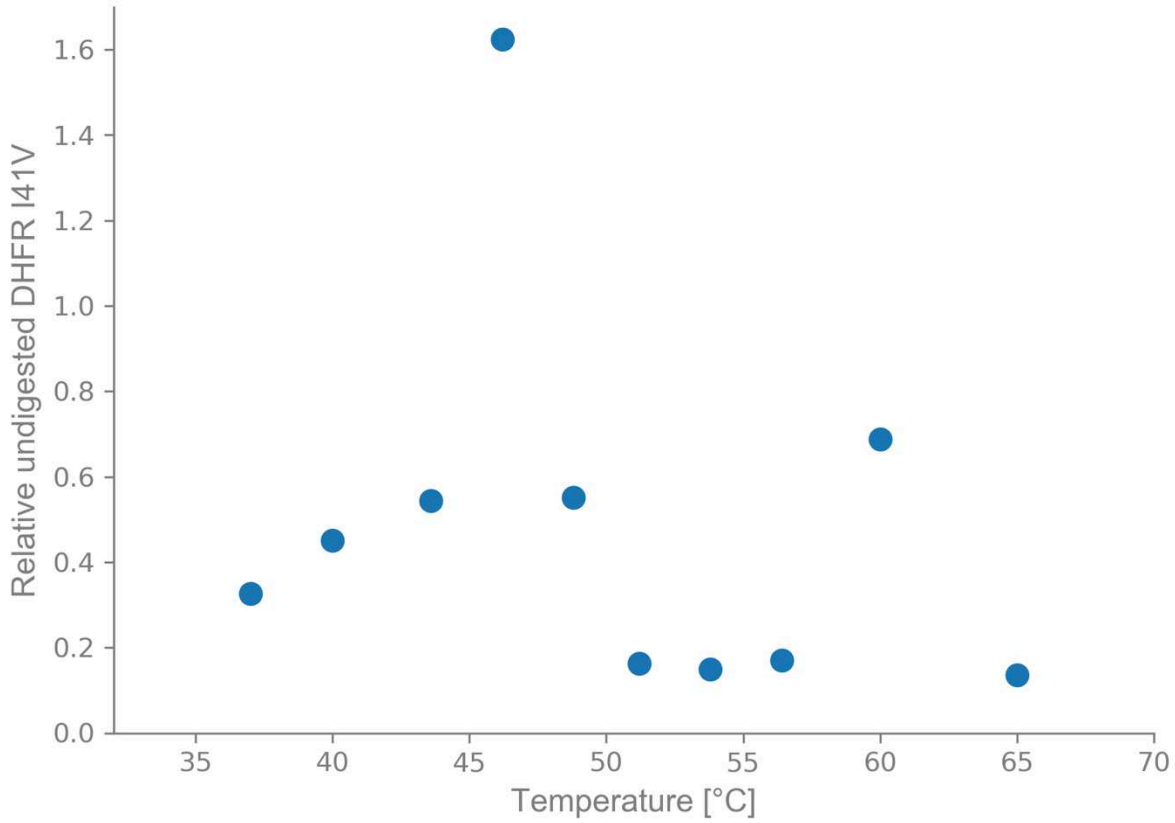


Figure 40. LiP-Chip analysis of ecDHFR I41V.

LiP of ecDHFR I41V was conducted with an E/S ratio of 1/10 000 and an incubation time of 1 min without preheating. The concentration of ecDHFR I41V was 1 µg/µL in a reaction volume of 20 µL. The LiP samples were analyzed with the P80 protein chip on the 2100 Bioanalyzer and processed as described in 'Data analysis'.

The current pandemic prevented us from acquiring further LiP-Chip data on ecDHFR I41V, as well as the wild-type and the other mutants. Nevertheless, we want to share possible explanations for the problems encountered with our preliminary experiments. The large differences of undigested ecDHFR I41V between LiP temperatures shown in Figure 40 might be influenced by aggregation. Aggregation is a stochastic process that can arise spontaneously. It can be sufficient for a few misfolded proteins to aggregate and form an aggregation nucleus, leading to a fast spread of protein aggregation throughout the sample. Although we attempted to remove aggregates by centrifugation, this can only remove large aggregates. Soluble microaggregates remain in the sample and can initiate further aggregation (97). We therefore recommend to reduce the protein concentration in the LiP-Chip experiments between 200 and 500 ng/µL.

A source for the low stability of ecDHFR I41V suggested by our experiments could be the PBS buffer. The CD experiments for the estimation of the thermal stability by Thompson *et al.* was conducted in Tris-HCl buffer with additional salt (29). Furthermore, the purified proteins were stored in a stabilizing solution of Tris-HCl with additional salt and glycerol. We suggest to change the buffer for the purification, storage, and possibly experimentation of the ecDHFR variants for further studies. In the following section, we will discuss properties of different widely used buffers.

Stability and environment

We have mentioned above that the stability of a protein is not only defined by the protein itself, but is highly influenced by its environment. One fundamental example is the small increase in entropy upon protein unfolding due to the formation of hydration shells around hydrophobic regions and the resulting entropy decrease of the solvent.

For laboratory experiments, the environment is quite simple and can be controlled through the choice of buffer, additives, pH, and temperature. It is not always evident however, which choice of environment is optimal or even appropriate for an experiment. This is mainly due to the complexity of mechanisms underlying protein properties and a lack of detailed understanding of their interactions with the environment. Thus, the current strategy for finding an appropriate environment for protein experiments or an optimal formulation for protein therapeutics, is empirical and relies on scanning a wide range of options.

The scope of this project is not to give a detailed introduction to the different environments or to understand their effect on protein stability. We want to give only a small overview of commonly used environments for studies of protein thermal stability. This can provide information for further LiP-Chip experiments with the above mentioned A2 scFv or ecDHFR and its mutants.

One important protein property to consider when choosing an appropriate environment is the isoelectric point (pI) of the protein. This is the pH at which the net charge of the protein is zero, meaning there are as many positively as negatively charged ions (98,99). At a pH above the pI, the protein is negatively charged, below the pI positively. The pI of a protein corresponds to the composition of the pK_as (negative logarithm of the acid dissociation constants) of the residues. The pK_a of a residue depends on its local environment, leading to different pIs for different conformations of the protein, such as natively folded and unfolded. It was originally postulated that proteins are most stable at their pI (addressed in (99)). However, it has since been shown in mutation studies that the conformational stability of proteins does not correlate with their pI (99). The aggregation propensity of a protein on the other hand, is highest around the protein's pI. It is therefore important to choose a pH that is appropriate for the experiment. During protein purification, a pH far away from the pI can lead to less aggregation and a higher yield. Studies of protein aggregation however, might be best conducted at a pH close to the pI of the protein of interest.

After choosing a pH for the experiment, a buffer with a corresponding pH range needs to be chosen. Three widely used buffers in biological and biochemical studies are phosphate buffered saline (PBS), tris(hydroxymethyl)aminomethane (Tris), and 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES). They all have their individual advantages and limitations.

PBS for example is a very specific and well-defined variant of a phosphate buffer that contains additional potassium ions and has a pH of approximately 7.4 (Table 6). The independence of the pH of PBS towards temperature changes makes it a good candidate for thermal stability studies. A different pH range can be achieved by preparing phosphate buffers at different ratios and pH than PBS. Exploring a different pH range is especially import for A2 scFv with an estimated theoretical pI of 7.18 (100).

Some people suggest that Tris or Trizma® buffers are more stabilizing for proteins than phosphate buffers. However, there does not seem to be published scientific evidence for this. Tris buffers are often used for protein purification, refolding, and storage. Despite the dependence of the pH of Tris buffers on the temperature, these buffers have been successfully used not only in chemical stability studies (24,101), but also in thermal stability studies (29,101). For further analyses of A2 scFv and the ecDHFR variants, we recommend to explore the purification in a Tris buffer. For the LiP-Chip methods, it needs to be evaluated if Tris is compatible with LiP at different temperatures. According to the specifications from Agilent, Tris buffers are compatible with the Bioanalyzer protein chips.

Another group of buffers are the Good's buffers. These buffers were specifically selected for the use in biological and biochemical studies. Good's buffers, such as HEPES, buffer in a wide pH range around the physiological pH (Table 6). In contrast to Tris, they are non-toxic to cell lines and are thus often used for cell cultures. Given the origin of the A2 scFv antibody, a Good's buffer might be a good choice if sufficient stability cannot be achieved in the PBS or Tris buffers.

Apart from the buffer, additives can be used to stabilize proteins and prevent aggregation, such as glycerol, salts, and sugars (102). Other additives however, need to be considered with care. His-tag purification of proteins for example, uses imidazole at different concentrations in the purification buffers. Imidazole itself is a buffer and can alter the pH of the solutions. In order to avoid effects of changing pH on the stability of the protein, the purification solutions should be precisely controlled for their pH at the temperature of their use.

Table 6. Commonly used buffers

Buffer	pH range	Notes
PBS	7.3-7.5[1]	Independent of temperature
Tris	7.0-9.0[2]	Temperature-dependent
HEPES	6.8-8.2[2]	Non-toxic to cell lines

1 Source (103)
2 Source (104)

Conclusion and perspective

We developed the LiP-Chip method: An affordable, quick, and easy-to-use method for the estimation of protein thermal stability. It combines the approach of LiP by PK over different temperatures with quantification on protein chips. We showed that LiP-Chip is well applicable for the estimation of the thermal stability of two commercially available proteins (BSA and α-Lac). These proteins are very different in their size, structure, and range of $T_{m,app}$ from around 39°C to around 60°C. We furthermore adapted the LiP-Chip method to be applicable to protein amounts as low as 10 µg for the estimation of thermal stability.

We took the first steps towards the validation of LiP-Chip for proteins from his-tag purification. Our first experiments on the A2 scFv antibody shows promising results. Based on this and the reported use of LiP on cell lysates (43), we are confident that LiP-Chip can be used for the estimation of the thermal stability of proteins from simple purifications or even overexpressed proteins in cell lysates. More importantly however, we suggest that the encountered issues during our project result from the protein purification itself, especially the environment. This highlights further the need for more detailed understanding of the effect of the environment on protein stability, as well as methods that allow for quick, easy, and affordable screening of protein stability in different environments. LiP-Chip is a first step towards such a widely useable method.

In addition to the estimation of $T_{m,app}$ with LiP-Chip, the method can easily be modified to provide an estimate of the Gibbs free energy of activation of the unfolding process ($\Delta G_U^{\ddagger}$). If a sample without PK is included on the protein chip, we can establish the initial amount of protein in the LiP samples. In combination with the knowledge of the incubation and the amount of undigested protein after the LiP experiment, we can estimate the rate constant of digestion (k_D). On the given timescale, we can assume k_D to be similar to the rate constant of unfolding (k_U). With the help of Equation 13, we can then obtain $\Delta G_U^{\ddagger}$ for the different LiP temperatures and thus, a kinetic stability for our protein of interest. Investigating not only the thermodynamic stability of proteins, but also their kinetic stability, can provide more insight into the mechanisms underlying protein stability and help in further approaches to tackle problems of protein engineering and application.

We identified several limitations of the LiP-Chip method and propose solutions to overcome them. The baseline effect influences the quality of the LiP-Chip results and increases the error between replicates. This limitation could be addressed with a custom analysis of the protein chip raw data or the use of other labeling methods, such as provided by the high-sensitivity protein chip from Agilent (HSP-250).

Another area, where we can propose improvement, is the required amount of protein. Although 10 µg for the estimation of thermal stability with LiP-Chip is below the minimal required amount of protein in analyses with CD or DSC, it is three times more than for ThermoFluor assays. However, the current protocol is easily adaptable to smaller volumes and lower protein amounts with the use of pipetting robots, high-precision pipettes for small volumes, or microfluidic approaches.

The largest bottleneck of LiP-Chip is the restriction to ten samples on the protein chip. This leads to only a medium throughput of the method. There are other quantification methods with higher throughput available, such as capillary electrophoresis (67). Nevertheless, we propose to tackle the limitations of LiP-Chip with tailored microfluidic devices. A tailored microfluidic device could not only help to increase the throughput, reduce the sample volume, and evade the baseline effect. It might also provide a means to combine the LiP step and the quantification on one device. Such a device would be highly useful to any protein studies for the evaluation of appropriate environmental conditions.

Finally, we want to give a small overview of projects and open questions that could benefit from the LiP-Chip method:

The first proposed project is the prediction of proteins with different stabilities and their subsequent analysis with LiP-Chip. It has been shown that there is a correlation between protein stability and coevolutionary coupling (13,105). It might therefore be possible to use DCA to predict mutations that alter the stability of a protein and even provide the means to determine the strength and direction of the stability alteration. LiP-Chip could be applied to estimate the thermal stability of designed proteins without the need for elaborate purification or highly specific instrumentation.

A widely discussed open question concerning protein stability is its link with evolvability. It has been postulated that more stable proteins are more robust towards mutations and can thus more easily evolve (106). Antibodies provide an ideal model to study if thermal stability is linked to a protein's potential as an evolutionary starting point. Comparing the stabilities of germline and matured antibodies, as well as their evolvability can shed more light on the question. A study from our lab shows that libraries based on germline antibody frameworks contain more variants that can bind new targets, than libraries based on matured antibody frameworks (107). We have made a first step towards assessing the average thermal stability within those libraries by optimizing their purification protocol and purifying the germline library (see 'Material and Methods').

When it comes to antibodies, there are still many open questions. Interesting to us is to analyze and quantify the link between thermal stability and other properties of the antibody, such as flexibility, affinity, neutralization ability, and polyreactivity. How does the stability change over affinity maturation? Many matured antibodies, and especially broadly neutralizing antibodies (bNAbs), are less stable than the corresponding germline antibody (27,108). In some cases however, matured antibodies without broadly neutralizing ability are reported with higher stabilities as their respective germline antibody. What affects this increase in stability? Are higher stability and broad neutralization mutually exclusive? It has also been reported that affinity maturation can follow different flexibility trajectories (109). The LiP approach assumes a higher digestion propensity with higher protein flexibility. Is there a link between the

overall flexibility and the thermal stability? Such questions can be addressed with scFv antibodies. Methods such as ThermoFluor assays might have problems with exposed hydrophobic regions of scFv antibodies. The LiP-Chip method could provide an alternative that does not require an extensive amount of protein or long analysis time.

Another topic that LiP-Chip could be applied to, is the effect of the environmental conditions on the thermal stability of proteins. If LiP-Chip is useable with cell lysates, this would enable the study of genetic mutations and their protein stability in different cell environments. As Thompson et al. showed, a small change in the cell environment (presence of absence of the Lon protease) can result in the enrichment of DHFR variants with highly different thermal stability (29). A tailored microfluidic adaptation of the current LiP-Chip method could further enable detailed studies of the link between environmental conditions and protein stability, possibly making the choice of environment for protein studies and formulations easier.

We can think of a variety of questions that could benefit from the LiP-Chip method or a more advanced integrated device. Although there is still some work to be done before LiP-Chip can be applied more widely, we have made substantial progress in providing a quick, affordable, and easy-to-use method for the estimation of protein thermal stability.

Supplementary information

Logistic fitting function

The second function we used to fit our LiP-Chip data is the following logistic function:

$$y(x) = \frac{A - D}{1 + e^{(C - x/-B)}} + D \qquad \text{(Eq. 15)}$$

As the function in Equation 14, the logistic function in (Eq. 15) has four parameters. The upper and lower plateaus are represented by the parameters A and D, respectively. The parameter B signifies the slope of the transition. With parameter C as the midpoint of the transition slope, Equation 15 allows an easy readout of $T_{m,app}$. This function is applied in other studies of protein stability, such as thermal stability estimations with CD (29,42). The Equation 15 can be derived from the Boltzmann distribution when assuming a two-state process and using the Gibbs-Helmholtz relationship. It is therefore often used in studies that estimate thermodynamic parameters for processed assumed to be at equilibrium. That is not the case with our LiP-Chip method. Nevertheless, our data shows a sigmoid behavior and can be well fitted with Equation 15 (Suppl. Figure 4). It is important to note though, that we cannot estimate absolute thermodynamic parameters. The parameters obtained from our experiments are apparent parameters that depend on the experimental conditions.

Detailed analysis of the baseline effect

The baseline effect was reproducible with two different Bioanalyzer instruments. According to the Agilent technical support, the increase of the background could be due to an insufficient air seal between the chip priming station and the protein chip. Indeed, an air seal test showed insufficient seal for protein chips. This seems to be a general attribute of the protein chips though. We observed the same sealing problems for protein chips with two different chip priming stations. Furthermore, both chip priming stations showed correct air seal for DNA chips, but not for protein chips. We therefore postulate that the baseline effect is an intrinsic problem of the protein chips that needs to be considered in the experimental design and data analysis.

The analysis software of the Bioanalyzer (2100 Expert Software) tries to remove the background fluorescence in an internal pre-processing step. Such a correction however, is not

sufficient to remove the baseline effect. We can clearly identify chips with the baseline effect from the steep increase in fluorescence of the upper marker for later samples (Suppl. Figure 12A). Unfortunately, this increase in baseline fluorescence affects the upper marker much stronger than the protein sample. This might be due to the much larger amount of protein sample in the analysis than the amount of upper marker. In Suppl. Figure 12B, we compare the results of three protein chips with replicates of BSA. One chip had a strong baseline effect for samples eight to ten. A second chip had a slight baseline effect, whereas a third chip did not show an increase in background fluorescence. For the two chips with baseline effect, the fluorescence intensity of the upper marker is strongly affected in later samples. The intensity of the peak corresponding to the undigested BSA however, is not noticeably affected by the background increase. This leads to a strong underestimation of the relative amount of undigested BSA for samples in wells eight to ten due to the normalization with the upper marker.

It is important to also mention the larger fluorescence intensities for the first sample. This is highly reproducible throughout all our experiments. Contrary to the baseline effect however, the larger fluorescence intensity for sample one is proportional between the upper marker and the protein sample and can be removed by normalization with the upper marker.

PDZ domain

The PSD-95/Drosophila Discs Large/Zona Occludens-I (PDZ) domains represent a widespread protein interaction module of only 80-100 amino acids. It can be found in a vast variety of organisms, from animals, to yeast, plants, and bacteria (110–112). In animals, it is an important part of the formation and function of signal transduction complexes. The 3D-structure of PDZ domains is highly conserved, with most of them consisting of five to six β-sheets and two α-helices - although there are a few deviations, such as a third α-helix in the variant PSD-95^{PDZ3} (Suppl. Figure 25) (113,114). Despite the conserved structure throughout the PDZ family, the functions and especially the binding partners of the variants are highly diverse (115).

The small size, abundance over many organisms, high structural conservation, and functional diversity makes PDZ domains a popular model for the understanding of protein interactions and allostery, as well as the relationships between sequence, structure, and function. Especially the vast amount of evolutionary sequence data of PDZ domains has fueled studies investigating energetic interactions within proteins through statistical analysis approaches (11–13,116,117). At the current moment, the PDZ pfam family consists of 55 769 sequences over 2 879 species (118).

In an initial scope of this project, we wanted to explore the relationship between a protein's thermal stability and its pairwise energetic interactions. For this, we focused on the third PDZ domain of the postsynaptic density protein 95 (PSD-95^{PDZ3}) and the statistical analysis approach of direct coupling analysis (DCA). DCA uses evolutionary sequence data to infer pairwise residue couplings based on coevolution that correspond to native contacts (119–122). Additionally, the technique provides a score based on those statistical couplings of coevolving residues for each member of the analyzed protein family. It has been shown that the coevolutionary couplings identified with DCA provide information on protein folding, protein-protein interactions, and fitness landscapes (14,123–125). Given the strong link between protein folding and protein stability, we proposed the possibility of using DCA for the prediction of a protein's thermal stability. We were planning on validating this hypothesis by calculating the DCA scores and thermal stabilities of single point mutants of PSD-95^{PDZ3} and comparing the link of both parameters of the set of mutants.

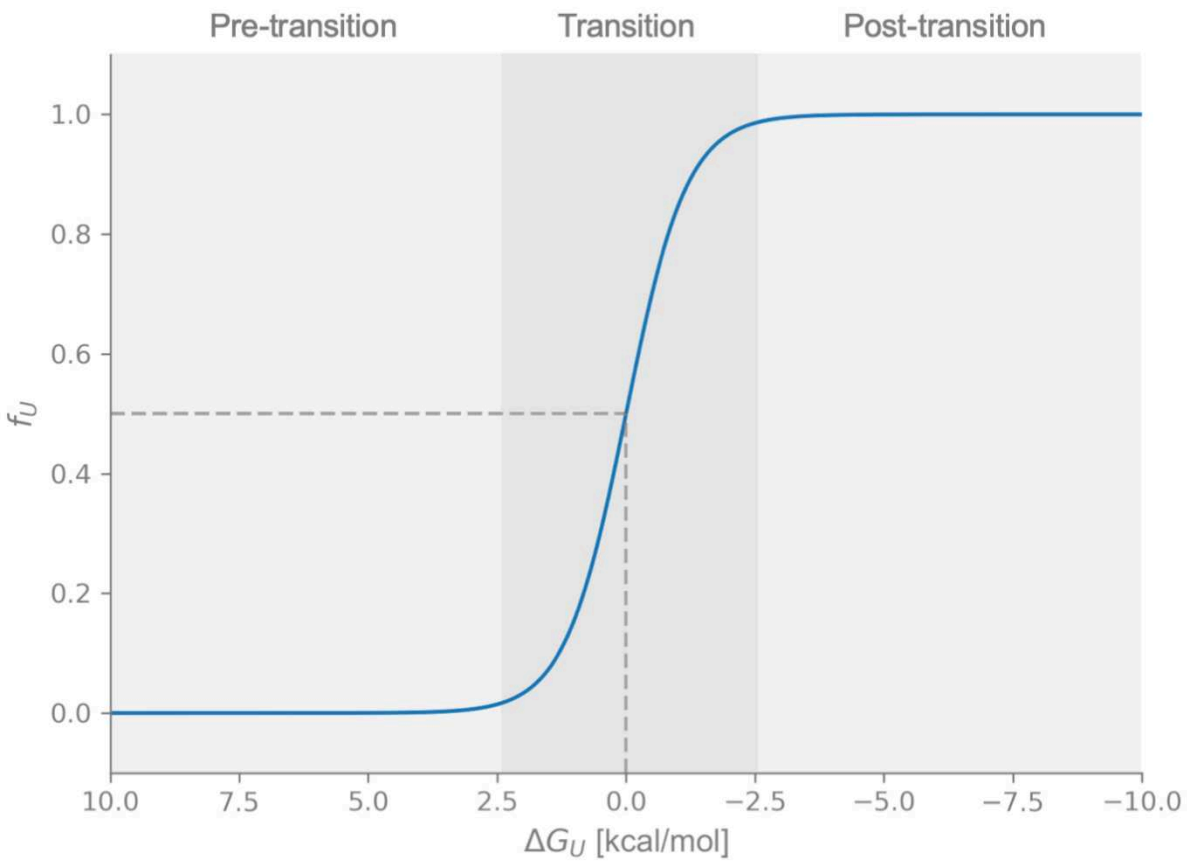
We received purified rat PSD-95PDZ3 (wild-type) and several of its point mutants as a gift from the Ranganathan lab. As a first step, we adjusted the protein chip protocol from Agilent for the usage with low protein concentrations to account for a limited amount of protein for some mutants (see 'Material and Methods'). We validated the adjusted protocol with BSA and successfully applied it to the wild-type PSD-95PDZ3 (Suppl. Figure 17 and Suppl. Figure 27). The well-visible peak in the electropherogram corresponding to PSD-95PDZ3 opens the doors for a LiP-Chip analysis with protein concentrations as low as 67 ng/µL in reaction volumes of only 15 µL. Thus, a whole LiP-Chip stability analysis over ten temperatures would require only 10 µg of protein, which is lower than the usually required amounts for stability studies with DSC or CD (Table 2).

Before an application of the adjusted protocol to the set of PSD-95^{PDZ3} mutants, we evaluated appropriate LiP parameters (E/S ratio and incubation time) for the wild-type protein. The T_m of the wild-type PSD-95^{PDZ3}, as established by the Ranganathan lab, is 68.8°C. We therefore analyzed a temperature range from 37°C to 72°C with E/S ratios from 1/100 to 1/1 000 and incubation times from 30 seconds to 5 minutes. With an E/S ratio of 1/300 and an incubation time of 30 seconds, we can see a clear decrease in the amount of undigested wild-type protein between 37°C and 70.4°C (Suppl. Figure 28).

In order to estimate the thermal stability of PSD-95^{PDZ3} with LiP-Chip, more temperatures in the range from 42°C to 72°C should be analyzed with the above described LiP conditions (E/S

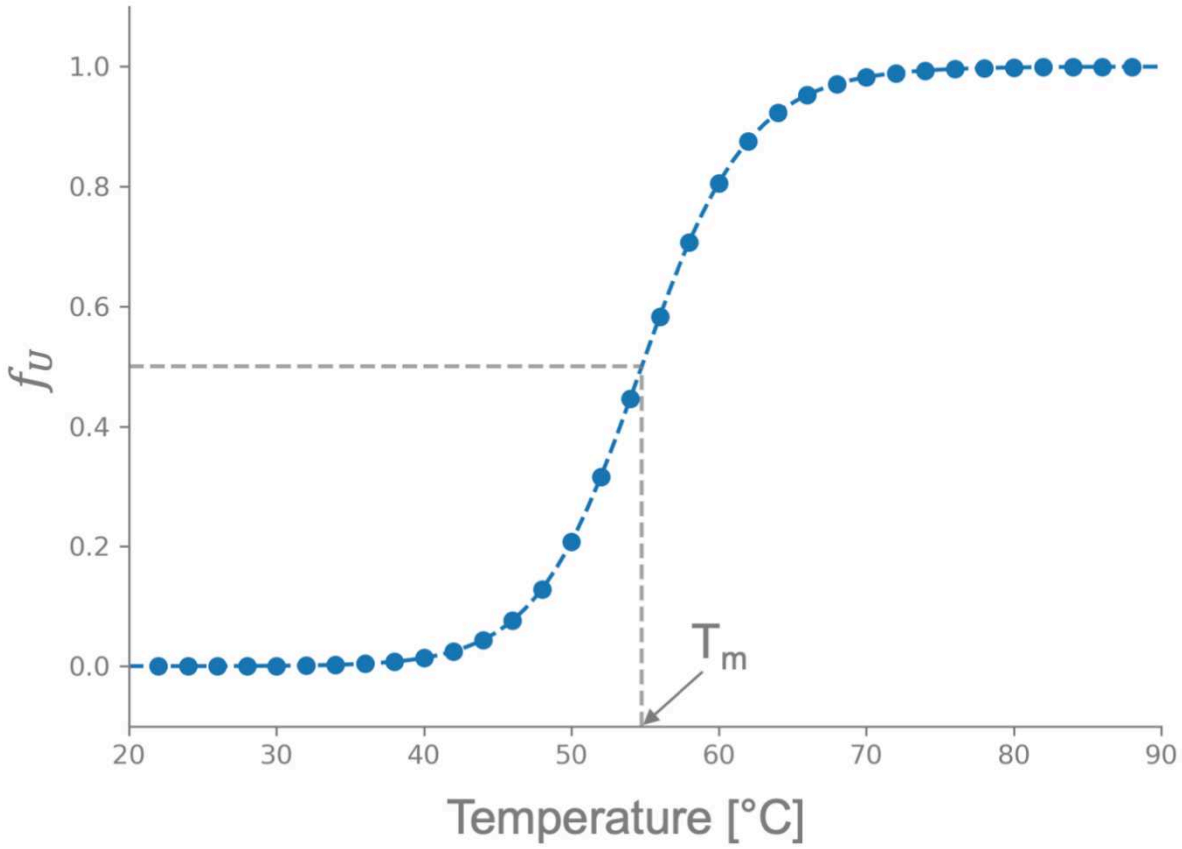
ratio of 1/300 and incubation time of 30 sec). Before we could proceed to conduct extensive LiP-Chip experiments on the wild-type PSD-95^{PDZ3} and the set of single point mutants, two papers showing the correlation between protein stability and coevolutionary coupling for single point mutants of the PDZ domain surfaced (13,105).

Supplementary figures



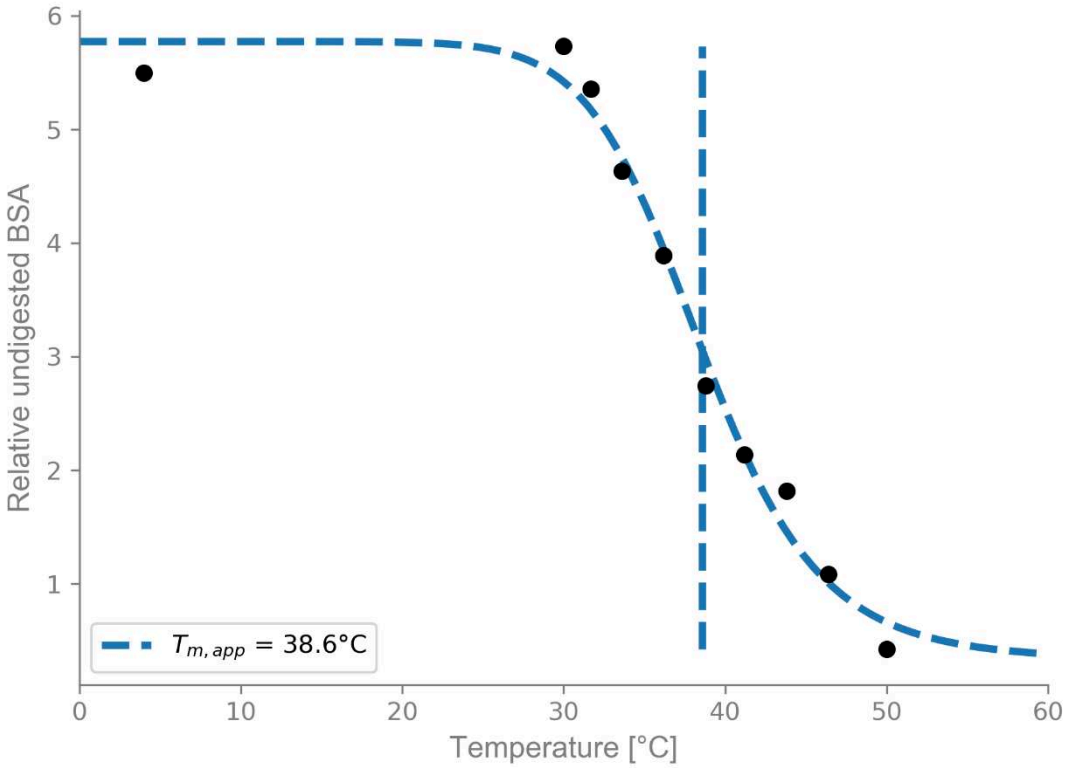
Suppl. Figure 1. Link between experiment and thermodynamic theory.

The sigmoidal curve was generated with Equation 8 over a ΔG_U range between 10 and -10 kcal/mol. The differently shaded regions correspond to the pre-transition, transition, and post-transition phases (from left to right). The dotted line indicates the point at which the probability of being folded or unfolded are the same, $f_U = f_N = 0.5$, and $\Delta G_U = 0$.



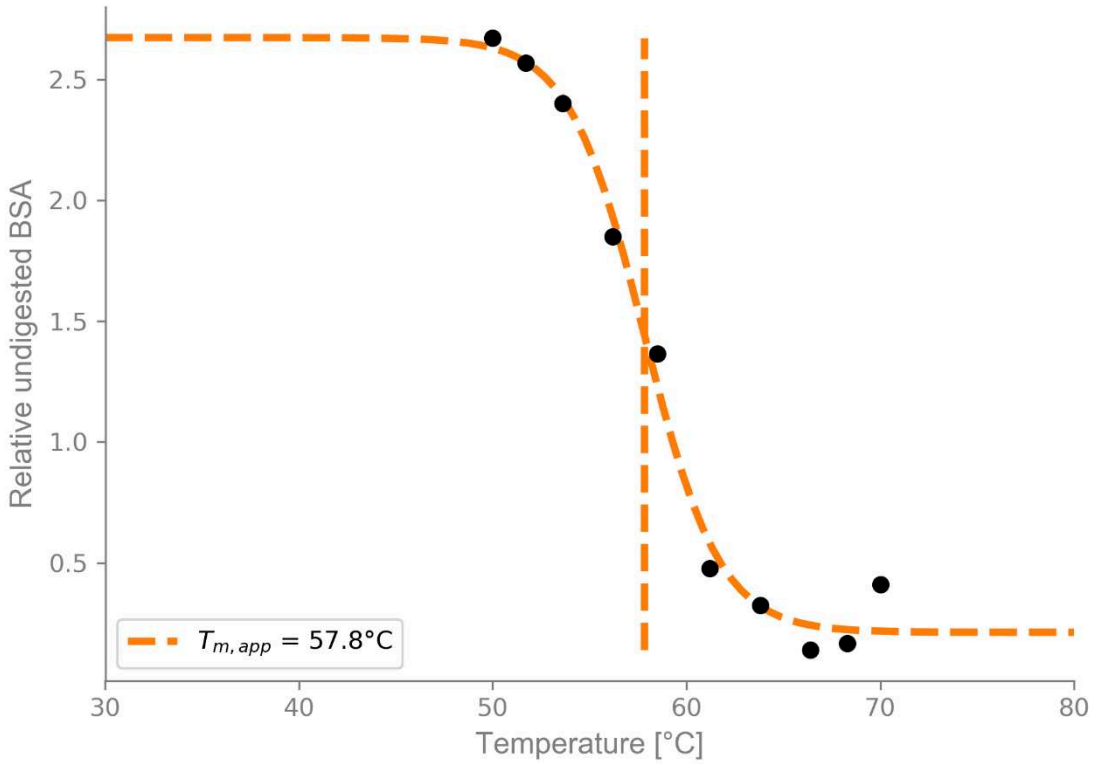
Suppl. Figure 2. Schematic of data from an ideal thermal stability experiment.

The sigmoidal curve was generated with Equation 8 combined with Equation 1 over changing temperature. The T_m is highlighted with the dotted line and represents the temperature at which 50% of the proteins are unfolded and ΔG_U is zero.



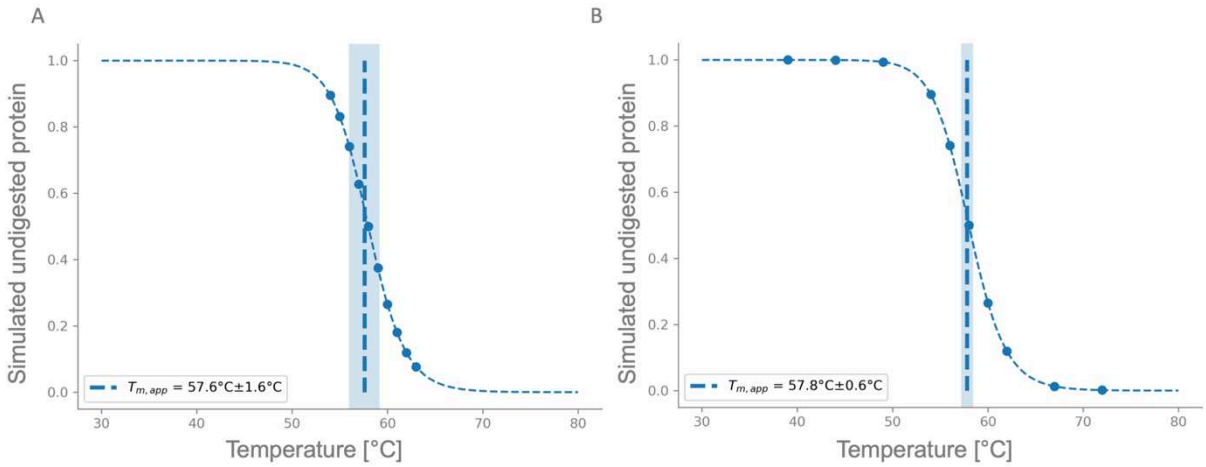
Suppl. Figure 3. Fitted sigmoidal curve to LiP-Chip data of $\boldsymbol{\alpha}$-Lac.

(Black) Relative amount of undigested α-Lac based on the normalized fluorescence intensity of the peak corresponding to undigested α-Lac in a LiP-Chip experiment over ten temperatures. (Blue) The curve shows the optimal fit for Equation 1 fitted to the LiP-Chip data of α-Lac. The vertical line marks the inflection point of the curve at a $T_{m,app}$ of 38.6°C.



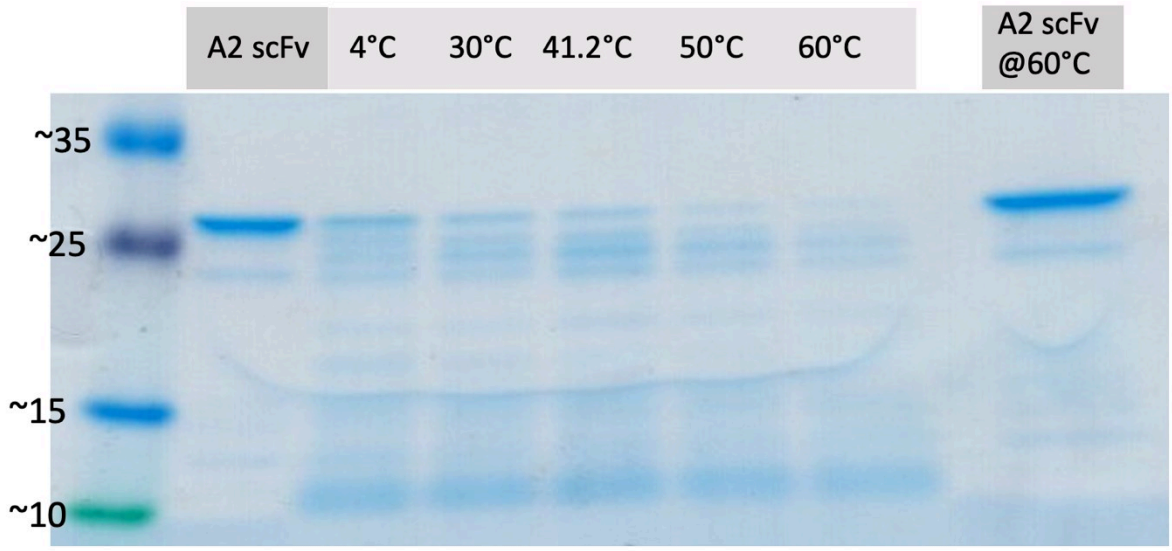
Suppl. Figure 4. Fitted logistic function to LiP-Chip data of BSA.

The fitting function is described in 'Logistic fitting function'. (Black) Relative amount of undigested BSA based on the normalized fluorescence intensity of the peak corresponding to undigested BSA in a LiP-Chip experiment over ten temperatures. (Blue) The curve shows the optimal fit for Equation 15 fitted to the LiP-Chip data of BSA. The vertical line marks the inflection point of the curve at a $T_{m,app}$ of 57.8°C.



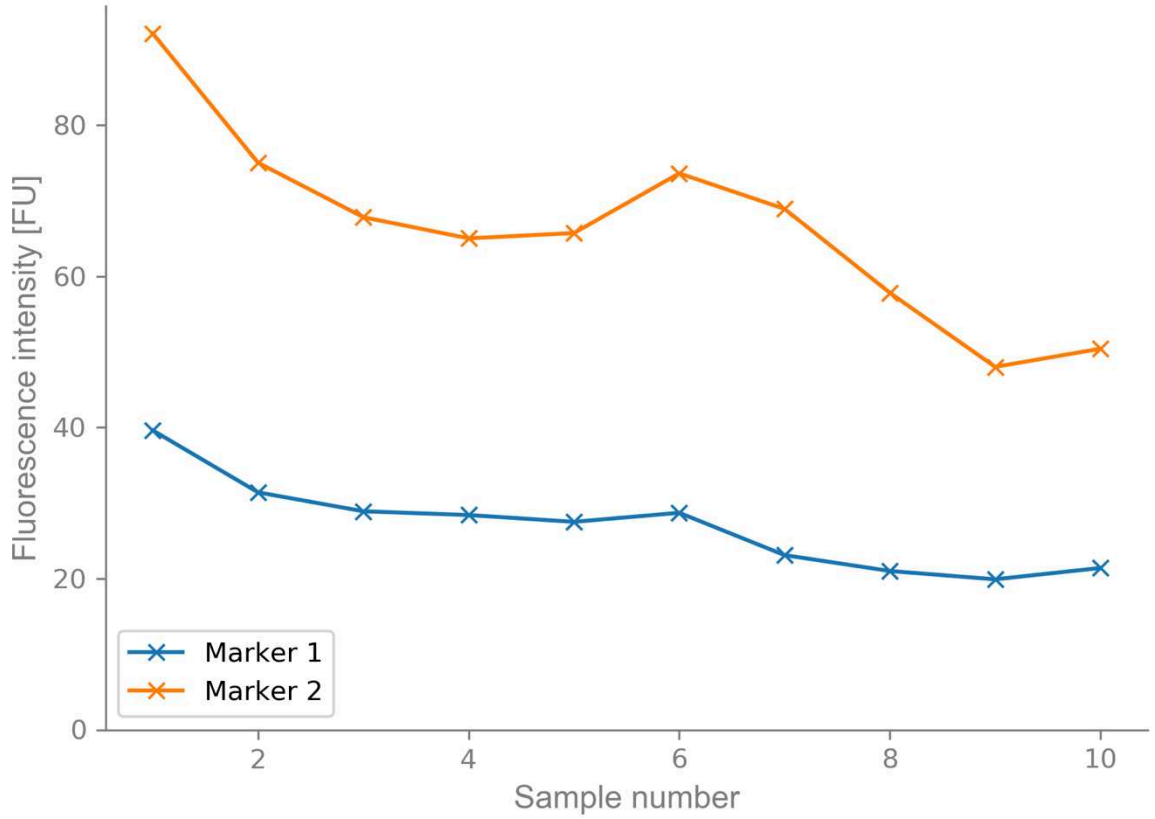
Suppl. Figure 5. Data simulation of phase coverage.

Both data were simulated with the following parameters for Equation 14: A = 1.0, B = 30, C = 58, D = 0. The perfect data points and the optimal fit are shown as circles and sigmoidal curve. For the actual data simulation, a normally distributed error with a standard deviation (SD) of 0.1 was added to each data point. 1 000 simulated data sets were generated and their $T_{m,app}$s estimated. The mean and the SD of the estimated $T_{m,app}$s is displayed in the legend. (A) The simulated LiP temperatures cover the transition phase. The mean $T_{m,app}$ is 57.6°C with a SD of 1.6°C. (B) The simulated LiP temperatures cover all three phases. The mean $T_{m,app}$ is 57.8°C with a SD of 0.6°C. The SD is thus smaller if the pre- and post-transition phases are covered in addition to the transition phase.



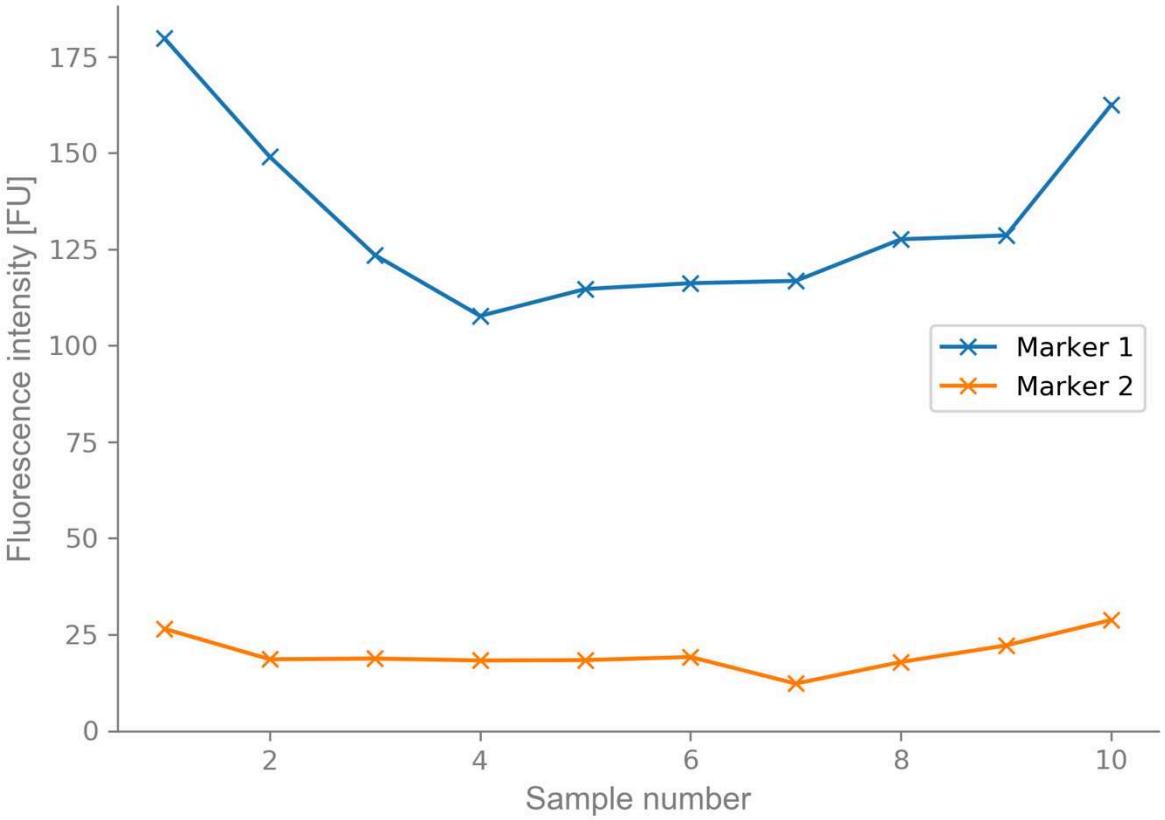
Suppl. Figure 6. Quantification of LiP of A2 scFv with SDS-PAGE and Coomassie blue staining.

The LiP experiment was conducted with an E/S ratio of 1/200 and an incubation time of 1 min without preheating. The A2 scFv concentration was 0.25 µg/µL in a reaction volume of 55 µL. Each lane was loaded with a total protein amount of 2.5 µg. The SDS-PAGE was run for 100 min with 25 mA and 200 V. The protein gel was stained for 1 h with PageBlue™ and destained overnight in ddH_2O. The undigested A2 scFv is visible just above the 25 kDa ladder band.



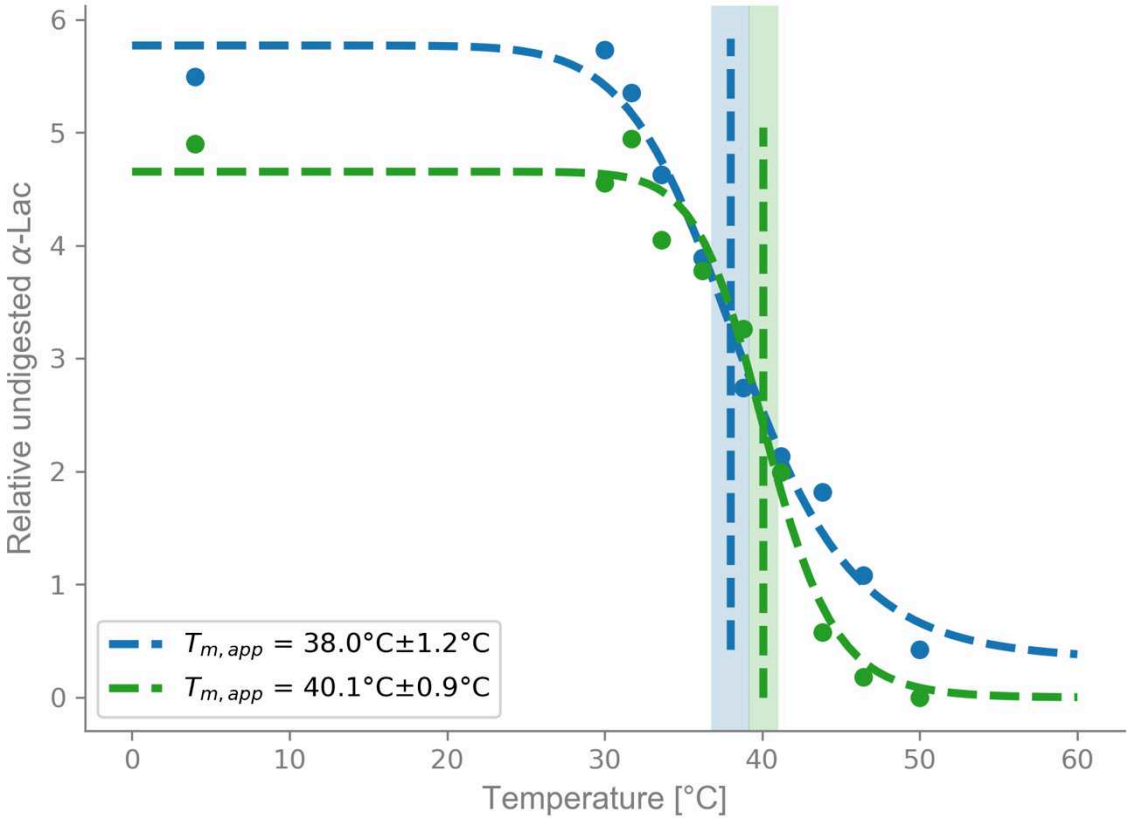
Suppl. Figure 7. Upper marker fluorescence intensities of LiP-Chip experimental replicates with α-Lac.

The absolute fluorescence intensity corresponding to the upper marker from sample well one to ten for two different LiP-Chip experiments of α-Lac. The absolute values are highly different between the two experiments.



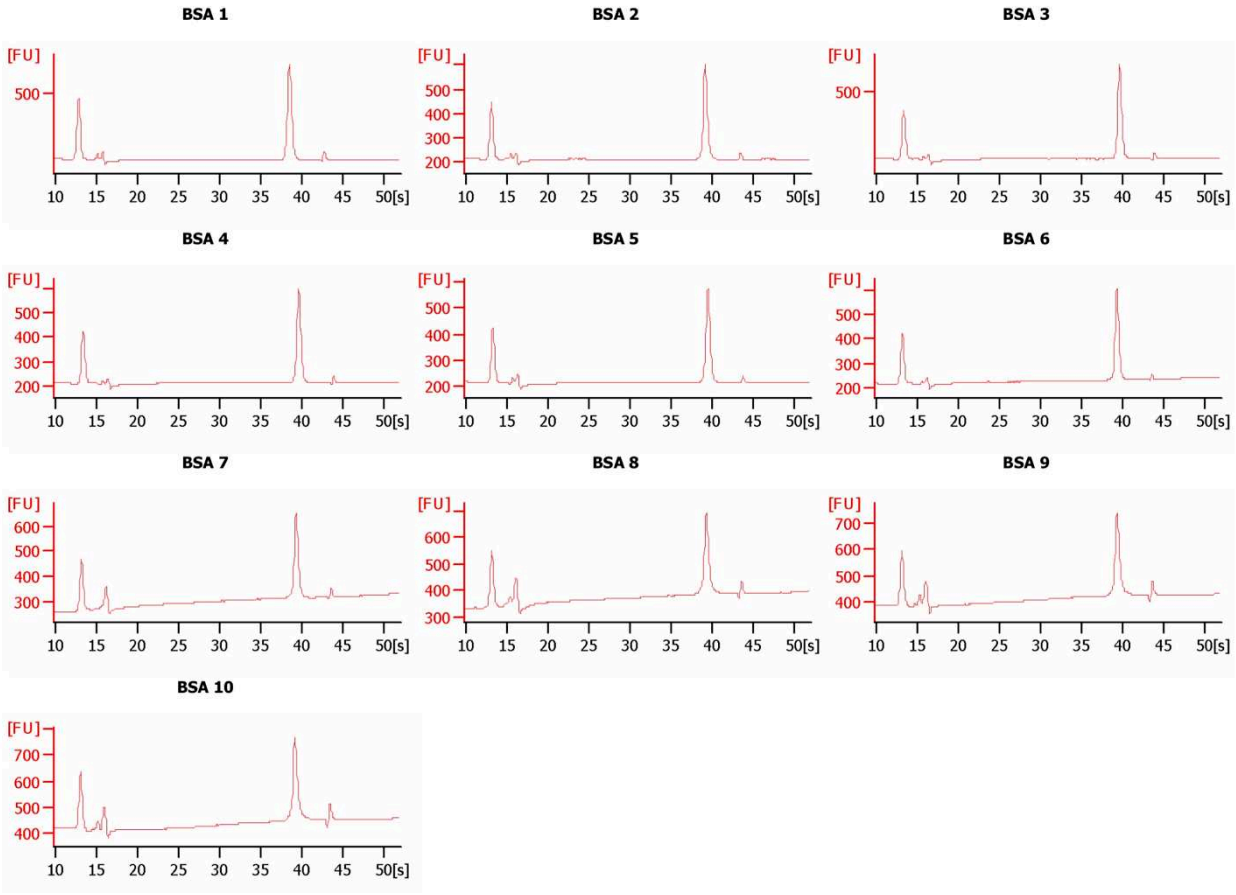
Suppl. Figure 8. Upper marker fluorescence intensities of LiP-Chip technical replicates with BSA.

The absolute fluorescence intensity corresponding to the upper marker from sample well one to ten for two different protein chip quantifications of the same LiP experiment of BSA. The absolute values are highly different between the two experiments.



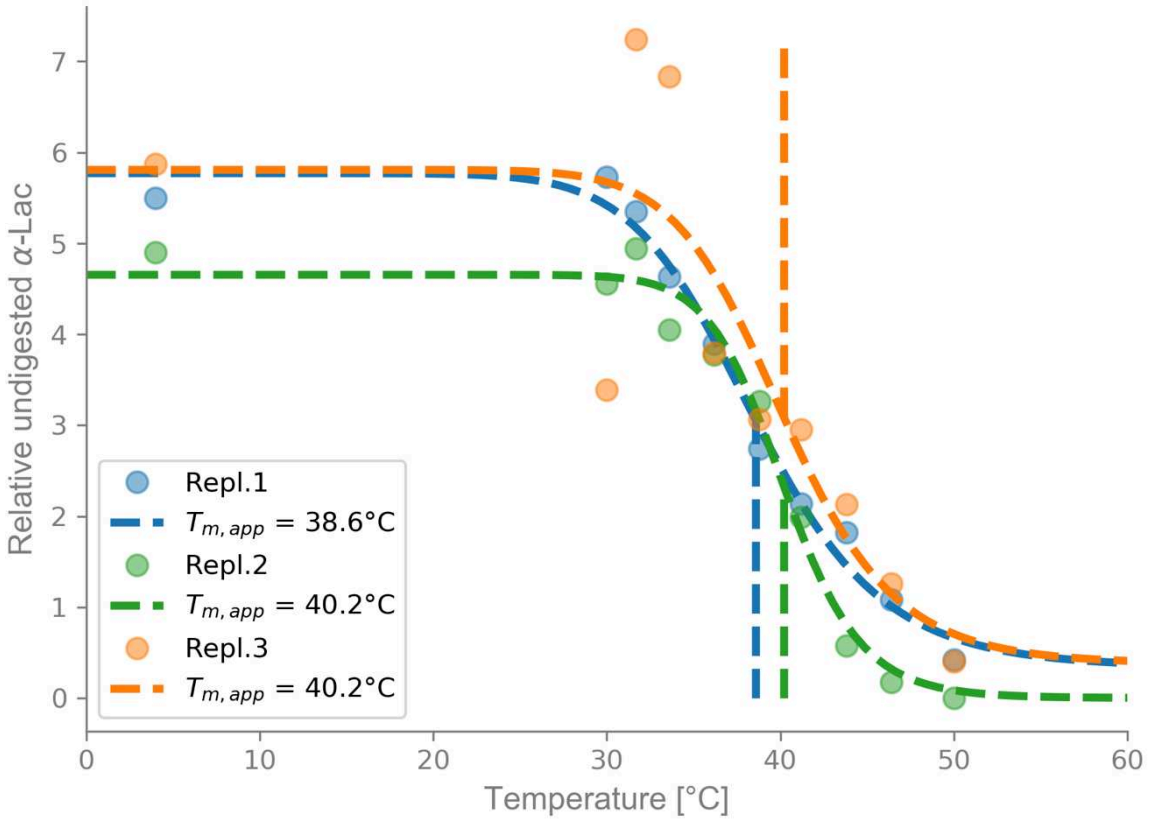
Suppl. Figure 9. Estimated SE of $T_{m,app}$ estimation for technical LiP-Chip replicates of α-Lac.

The blue and the green data represent the same LiP experiment, but a quantification on two different protein chips. The data from the protein chips was normalized by the fluorescence intensity of the upper marker in each sample well. The resulting amount of relative undigested protein was fitted with Equation 14. The SE was estimated with the bootstrap method as described in 'Standard error of the estimator'. The upper boundary of parameter A was chosen as two times the maximum over the measured data, in accordance with experiments containing α-Lac without PK at a known concentration. The E/S ratio of the LiP step was 1/500, the incubation time 1 min without preheating. The α-Lac concentration was 0.5 µg/µL in a reaction volume of 20 µL.



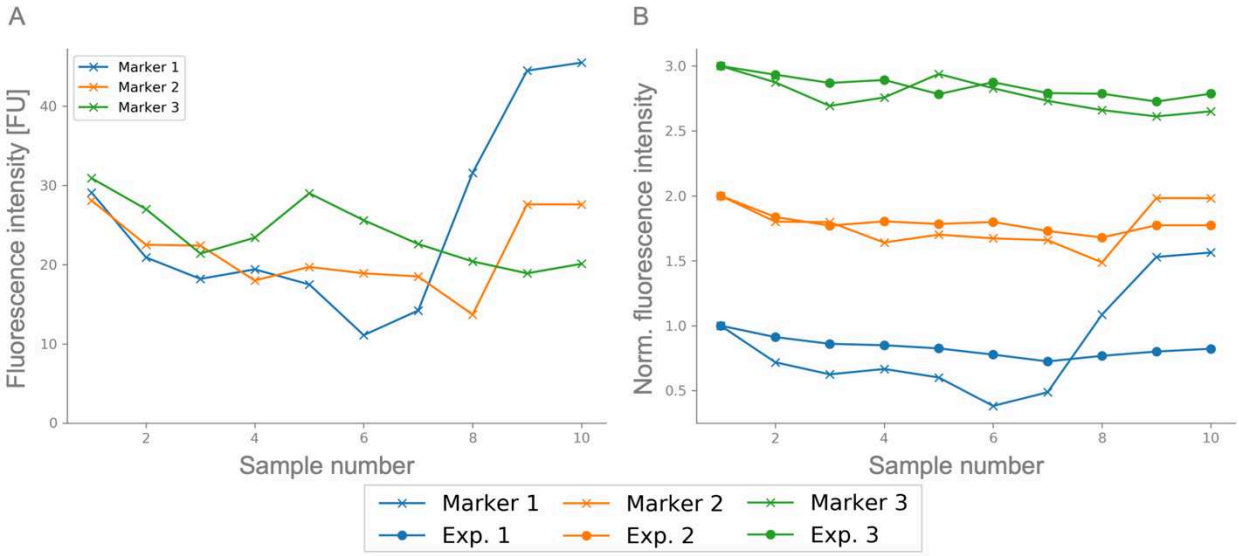
Suppl. Figure 10. Spectra corresponding to raw data from a protein chip with baseline effect.

The experiment consisted of ten BSA samples from the same aliquot. The increase in overall fluorescence intensity from approximately 200 FU to above 400 FU is well visible from sample seven on. Additionally, the baseline for samples seven to ten show a positive slope over increasing migration time. The 2100 Expert Software tries to remove the baseline for each sample in a preprocessing step. This internal correction is not sufficient to eliminate the strong increase in fluorescence intensity for the upper marker in wells with increased baseline.



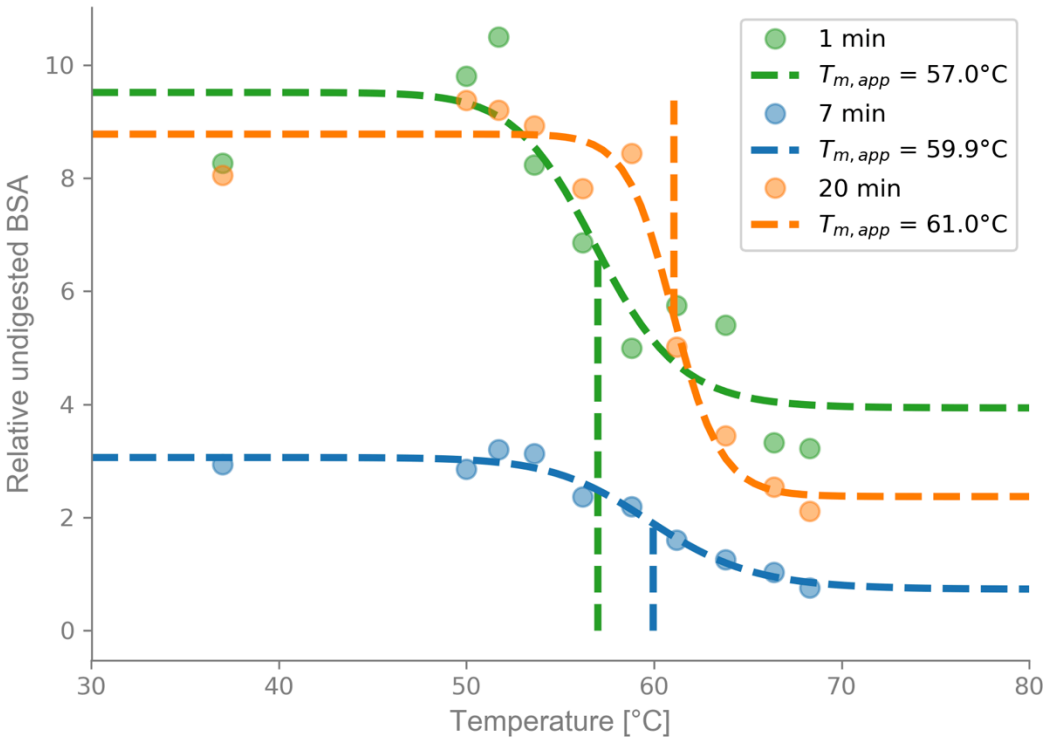
Suppl. Figure 11. Technical replicates of LiP-Chip of α-Lac with and without baseline effect.

All three data sets correspond to the same LiP experiment evaluated on three different protein chips and two different 2100 Bioanalyzer instruments. (Blue) Repl. 1 did not have the baseline effect. (Green) Repl. 2 had a baseline effect and was conducted on a different instrument than the other two replicates. (Orange) Repl. 3 had a baseline effect. The E/S ratio was 1/500, the incubation time 1 min without preheating. The α-Lac concentration was 0.5 µg/µL in a reaction volume of 20 µL. The data were processed and analyzed as described in 'Data analysis'.



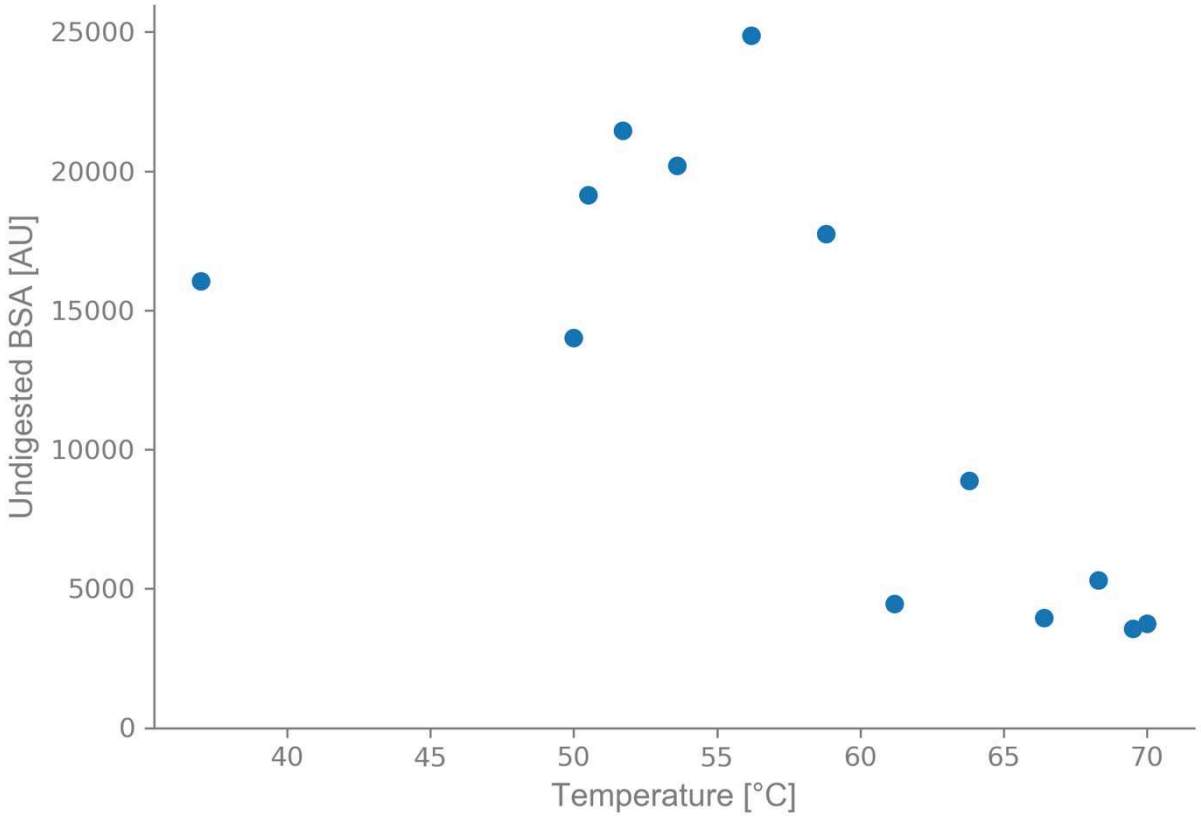
Suppl. Figure 12. Baseline effect for BSA replicates.

The three colors depict three different protein chips with ten BSA samples for each. (A) The absolute fluorescence intensities of the upper marker in each sample well. (Blue) The baseline effect is well visible by the strong increase of the upper marker intensity for wells eight and up. (Orange) A slight baseline effect is depicted by the increased upper marker intensity of wells nine and ten. (Green) No baseline effect for this protein chip. (B) For the easy comparison of the baseline effect on the upper marker and on the BSA sample over the three protein chips, the upper marker intensities and BSA intensities were normalized and the results from the three experiments were shifted by a value of 1 for Exp. 2 and 2 for Exp. 3. The crosses depict the upper marker intensities normalized by the upper marker in the first sample well of the respective protein chip. The dots represent the fluorescence intensity corresponding to BSA, normalized with the value of the first sample of the respective protein chip. For all experiments, the normalized absolute fluorescence intensity of BSA is well reproducible, despite the presence of the baseline effect in protein chips one and two.



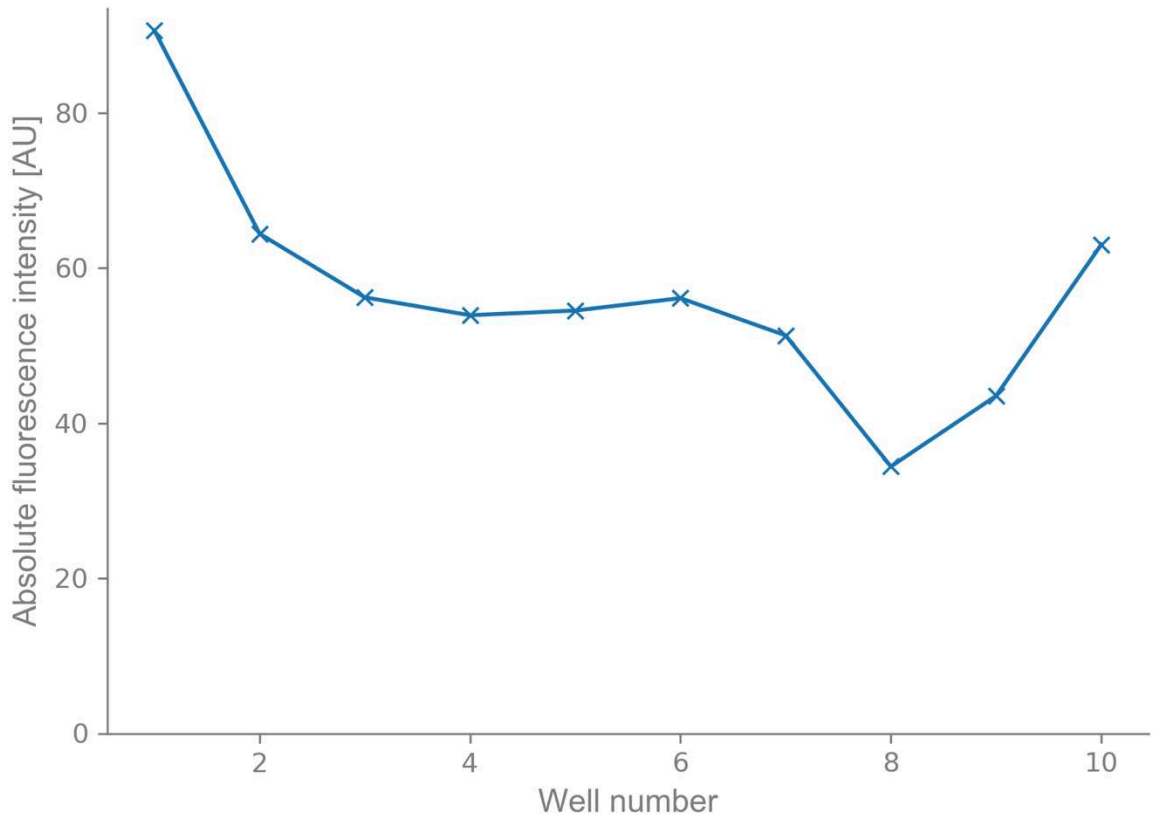
Suppl. Figure 13. LiP-Chip stability of BSA over different incubation times.

The E/S ratio for all incubation times was 1/2 000. The BSA concentration was 2 µg/µL in a reaction volume of 100 µL. The data was fitted with Equation 14. The estimated $T_{m,app}$ are 57°C for an incubation time of 1 min, 59.9°C for 7 min incubation, and 61.0°C for an incubation time of 20 min.



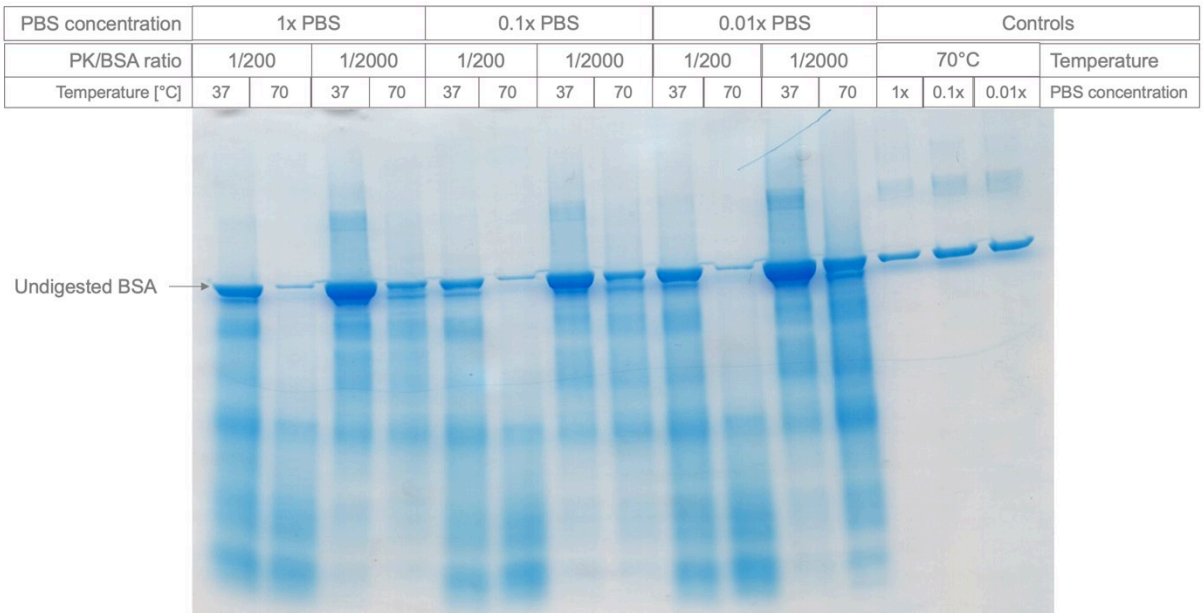
Suppl. Figure 14. LiP of BSA with preheating in small volumes.

The LiP experiment was conducted with an E/S ratio of 1/2 000 and an incubation time of 5 min with preheating of 5 min. The preheating volume was 10 µL. The BSA concentration was 2 µg/µL in a reaction volume of 15 µL. The quantification was done with SDS-PAGE on a 4-20% Tris-Glycine gel run for 135 min with 25 mA and 200 V. Each well contained a total protein amount of 10 µg. The gel was stained with PageBlue™ staining solution overnight and destained in ddH_2O over the course of one day. The densitometry analysis followed instructions from the SYBIL project (78).



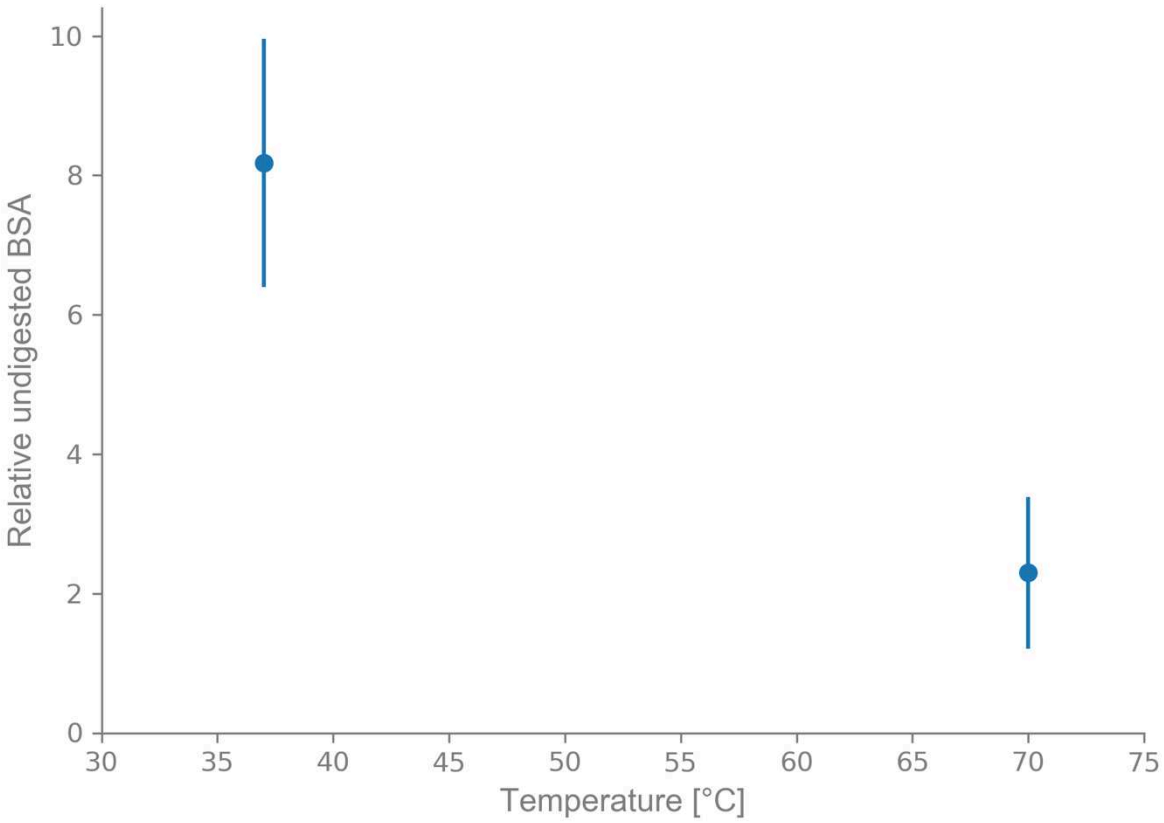
Suppl. Figure 15. Upper marker fluorescence indicating outlier data point.

The data shows the absolute fluorescence intensity of the upper marker in wells 1 to 10 corresponding to the LiP-Chip experiment of A2 scFv depicted in blue in Figure 26 and in orange in Figure 31. As observed in most experiments, the fluorescence intensity in the first well is higher than in the following wells. The dip and increase in fluorescence intensity for the last three wells indicates the baseline effect. The observed outlier in the relative fluorescence intensity of the undigested A2 scFv for sample 8 (43.8°C in Figure 26 and Figure 31) is most likely due to the low fluorescence intensity of the upper marker for that sample.



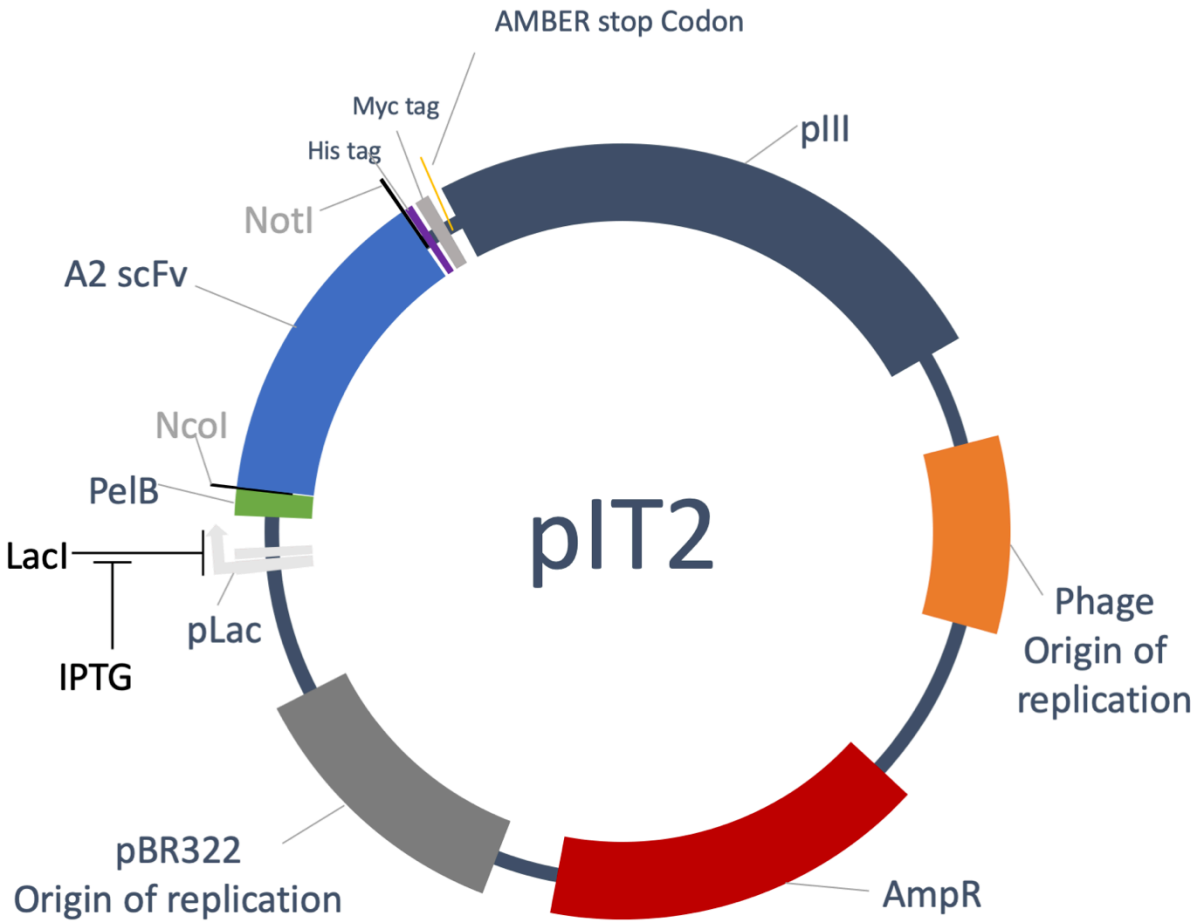
Suppl. Figure 16. LiP of BSA with different PBS dilutions.

We conducted LiP experiments of BSA at 37°C and 70°C with E/S ratios of 1/200 and 1/2 000 in different PBS dilutions. The incubation time was 5 min without preheating. The results were quantified with SDS-PAGE on a 16% Tricine gel run for 90 min with 100 mA and 125 V. The wells with the LiP samples contained a total protein amount of 10 µg, the control samples 1 µg. The gel was stained with PageBlue™ overnight and destaining in ddH_2O over the course of several hours. The difference in undigested BSA between 37°C and 70°C is visible at all PBS dilutions and all E/S ratios.



Suppl. Figure 17. LiP-Chip of BSA with low concentration protocol.

The LiP experiment was conducted with an E/S ratio of 1/200 and an incubation time of 5 min without preheating. The BSA concentration was 67 ng/µL in a reaction volume of 15 µL. The protein chip was used as described in the protocol adjusted for low concentrations (see 'Material and Methods'). The error bars corerspond to a mix of four replicates (two experimental replicates and two technical replicates).



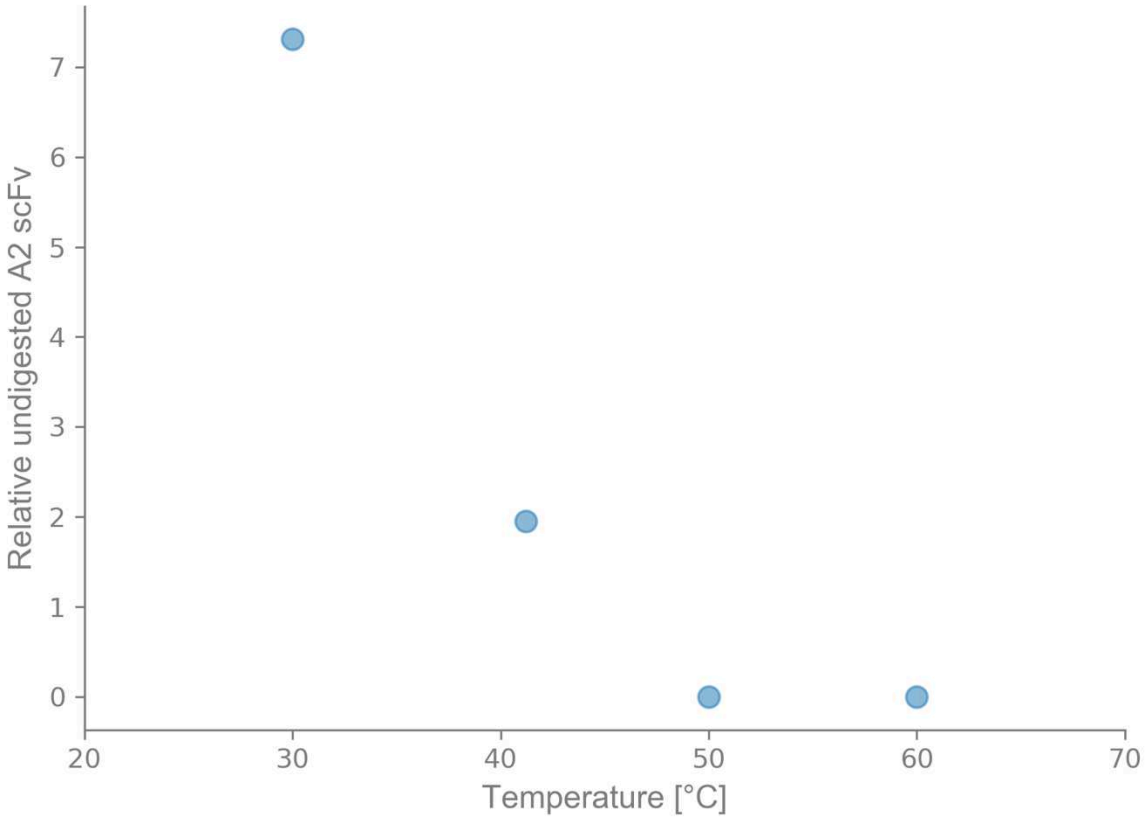
Suppl. Figure 18. Plasmid map of A2 scFv in pIT2.

The pIT2 plasmid contains an ampicillin resistance cassette (red), a pBR322 origin of replication (grey), and a phage origin of replication (orange). The A2 scFv gene (blue) is regulated by a pLac promoter. For secretion it is fused to a PelB signal peptide (green) and for purification and detection it is followed by a his-tag (purple) and a myc-tag (light grey). A2 scFv is separated from the pIII gene (dark blue) through an amber stop codon (yellow). In amber-suppressor strains, such as *E. coli* TG1, the fusion protein of A2 scFv and pIII is produced. In the *E. coli* strain BL21 (DE3), the ribosome stops at the amber stop codon and the pIII is not expressed.

```
10         20         30         40         50         60
AEVQLLESGG GLVQPGGSLR LSCAASGFTF SSYAMSWVRQ APGKGLEWVS YIDTSGSYTS
         70         80         90        100        110        120
YADSVKGRFT ISRDNSKNTL YLQMNSLRAE DTAVYYCAKY GSGFDYWGQG TLVTVSSGGG
        130        140        150        160        170        180
GSGGGGSGGG GSTDIQMTQS PSSLSASVGD RVTITCRASQ SISSYLNWYQ QKPGKAPKLL
        190        200        210        220        230        240
IYSASYLQSG VPSRFSGSGS GTDFTLTISS LQPEDFATYY CQQYANTPAT FGQGTKVEIK
        250        260
RAAAHHHHHH GAAEQKLISE EDLNGAA
```

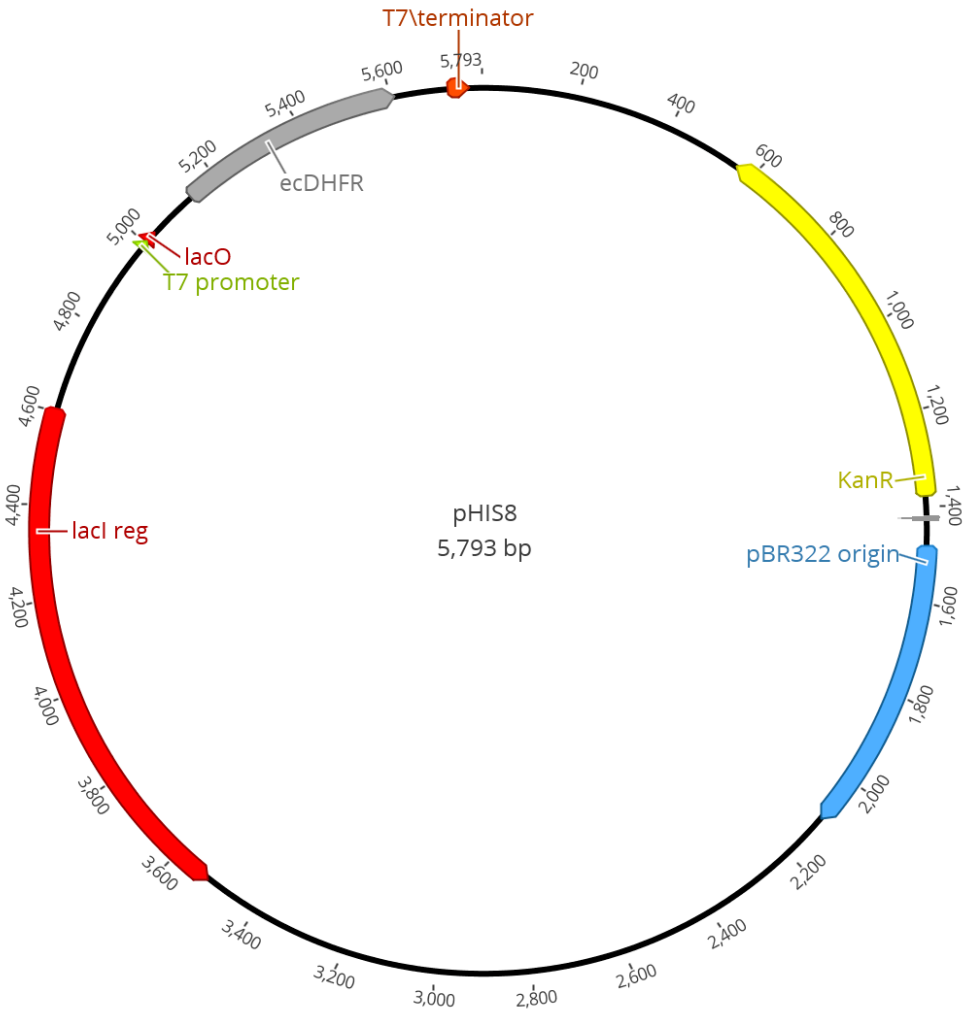
Suppl. Figure 19. Amino acid sequence of A2 scFv.

Amino acid sequence of A2 scFv (89). The molecular weight calculated with the ProtParam tool from ExPASy is 28142.94 Da (100).



Suppl. Figure 20. LiP-Chip experiment of A2 scFv adapted for low concentrations.

The E/S ratio was 1/200 with an incubation time of 1 min without preheating. The A2 scFv concentration was approximately 0.25 µg/µL in a reaction volume of 55 µL. The LiP samples were quantified with the protein chip according to the chip preparation adapted for low protein concentrations (see 'Material and Methods'). These preliminary results show a strong decrease in undigested A2 scFv between 30°C and 40°C. No undigested A2 scFv is detectable in the LiP samples at 50°C and 60°C.



Suppl. Figure 21. Plasmid map of ecDHFR in pHIS8.

The pHIS8 plasmid contains a kanamycin resistance cassette (yellow), a pBR322 origin of replication (blue), and a lacI regulatory gene and a lac operon (red). The ecDHFR gene (grey) is regulated by a T7 promoter (green).

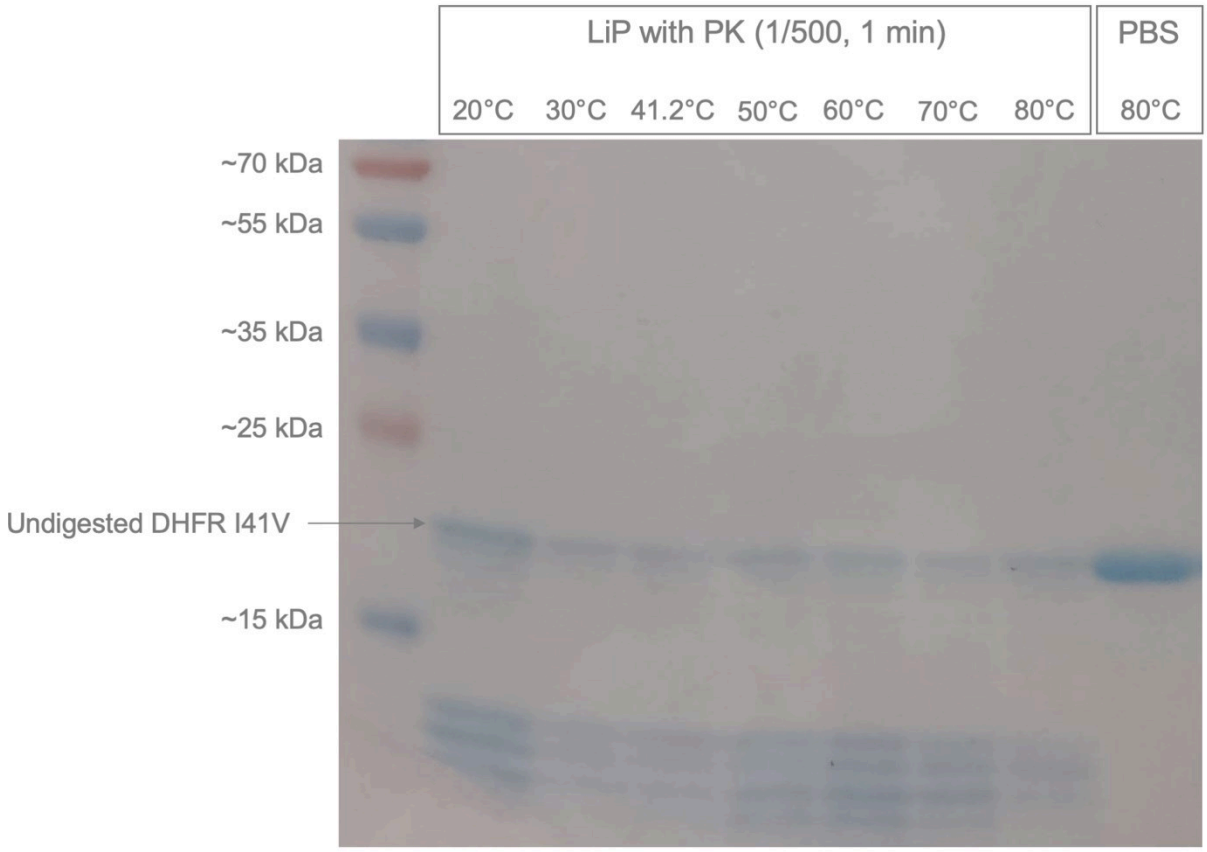
```
        10         20         30         40         50         60
KHHHHHHHHG GLVPRGSHGM MISLIAALAV DRVIGMENAM PWNLPADLAW FKRNTLNKPV

        70         80         90        100        110        120
VMGRHTWESI GRPLPGRKNI ILSSQPGTDD RVTWVKSVDE AIAACGDVPE IMVIGGGRVY

       130        140        150        160        170
EQFLPKAQKL YLTHIDAEVE GDTHFPDYEP DDWESVFSEF HDADAQNSHS YCFEILERR
```

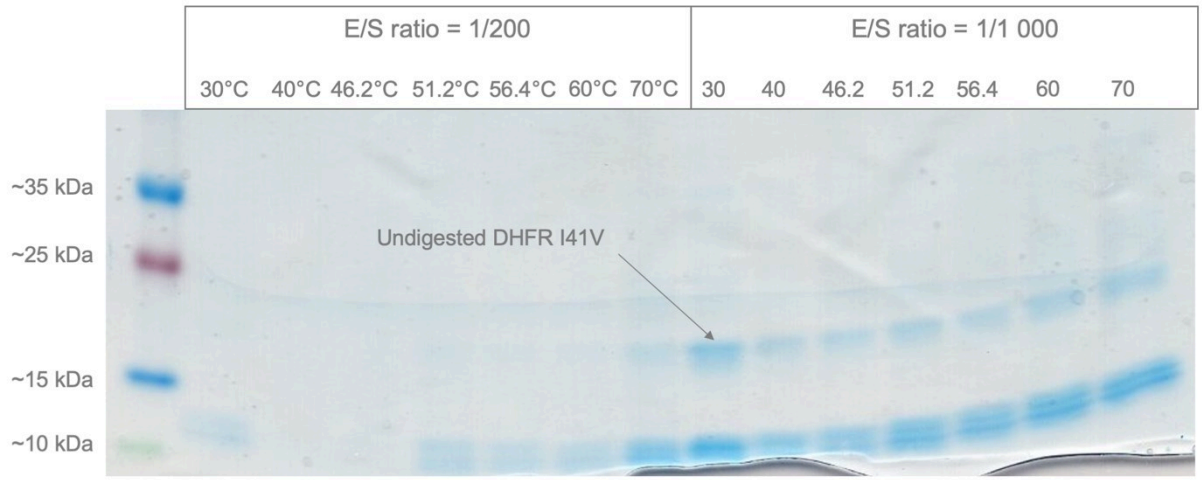
Suppl. Figure 22. Amino acid sequence of ecDHFR I41V.

Amino acid sequence of ecDHFR I41V received as a gift from the Kortemme laboratory. The molecular weight calculated with the ProtParam tool from ExPASy is 20259.87 Da (100). The mutation of the isoleucine at position 41 (without the his-tag and linker) to valine is marked in red.



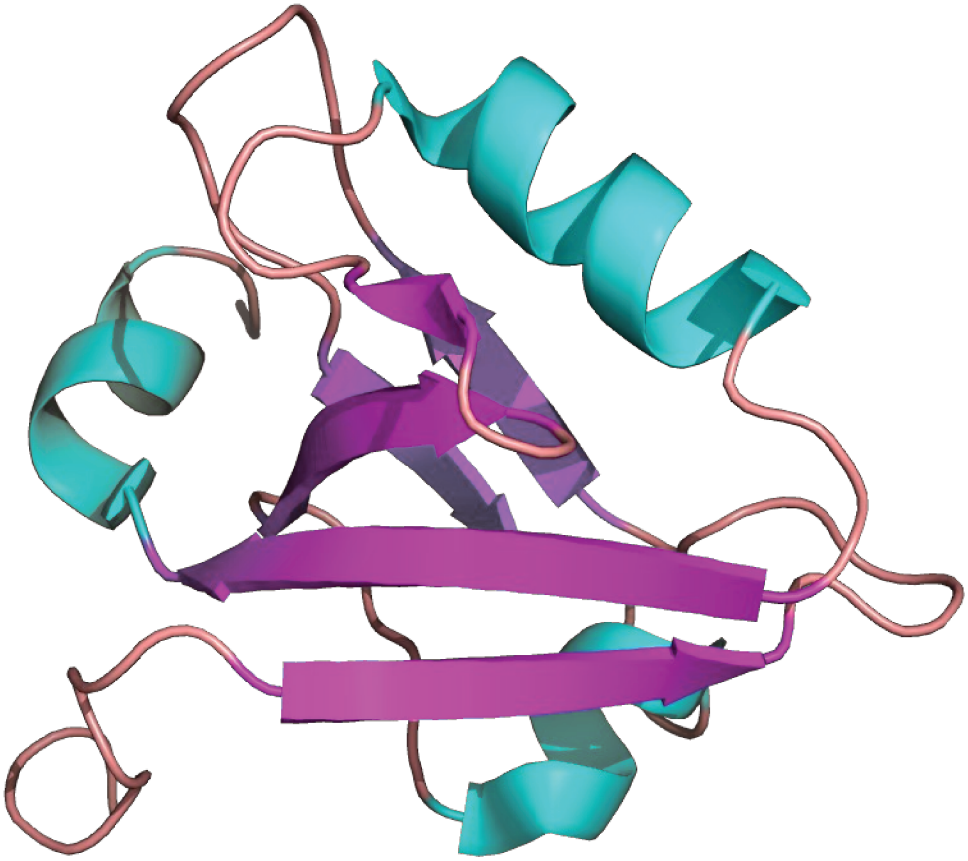
Suppl. Figure 23. Validation of purification and LiP-Chip of ecDHFR I41V.

The far right lane shows the purified and dialyzed ecDHFR I41V after following the LiP protocol at 80°C without PK. Only the band of the undigested ecDHFR I41V is visible, indicating the success of the purification. Furthermore, the heating of ecDHFR I41V does not lead to substantial protein degradation. The other lanes containt the PageRuler™ plus protein ladder and the LiP samples with temperatures from 20°C to 80°C. The E/S ratio was 1/500 and the incubation time 1 min without preheating. The A2 scFv concentration was 0.5 µg/µL in a reaction volume of 20 µL. Each lane on the protein gel contains a total protein amount of 5 µg. The 4-20% Tris-Glycine gel was run for 90 min with 25 mA and 200 V. It was stained for 1 h with PageBlue™ staining solution and destained overnight in ddH_2O.



Suppl. Figure 24. LiP-Chip of ecDHFR I41V with different E/S ratios.

The LiP-Chip experiments were conducted with two different E/S ratios: 1/200 and 1/1 000. The incubation time was 1 min without preheating. The A2 scFv concentration was 0.5 µg/µL in a reaction volume of 20 µL. Each lane on the protein gel contains a total protein amount of 5 µg. The 4-20% Tris-Glycine gel was run for 90 min with 25 mA and 200 V. It was stained for 1 h with PageBlue™ staining solution and destained overnight in ddH_2O.



Suppl. Figure 25. Crystal structure of the third PDZ domain from PSD-95.

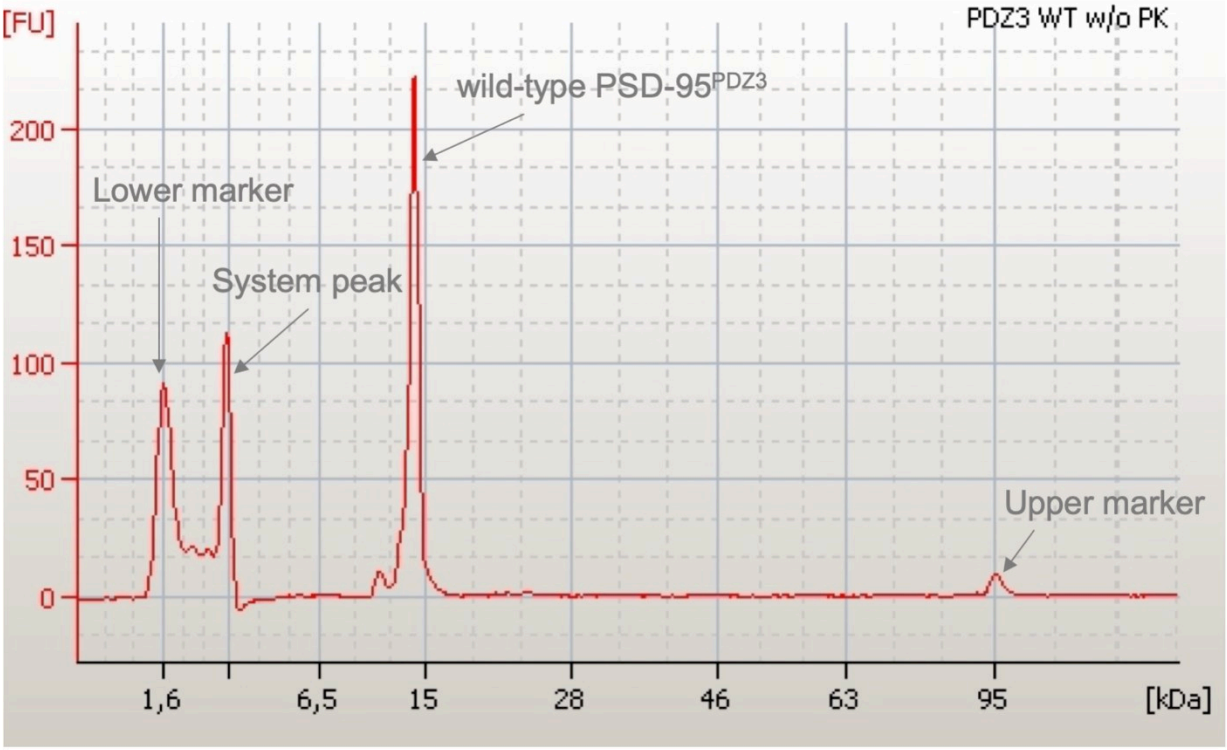
Crystal structure of PSD-95^{PDZ3} (PDB 5MZ7) rendered with PyMol. The ligand Azidohomoalanine was removed for this image. The three α-helices are depicted in light blue, the six β-sheets in pink, and loops are salmon-colored.

```
        10         20         30         40         50         60
GSPEFLGEED IPREPRRIVI HRGSTGLGFN IVGGEDGEGI FISFILAGGP ADLSGELRKG

        70         80         90        100        110
DQILSVNGVD LRNASHEQAA IALKNAGQTV TIIAQYKPEE YSRFEANSRV DSSGRIVTD
```

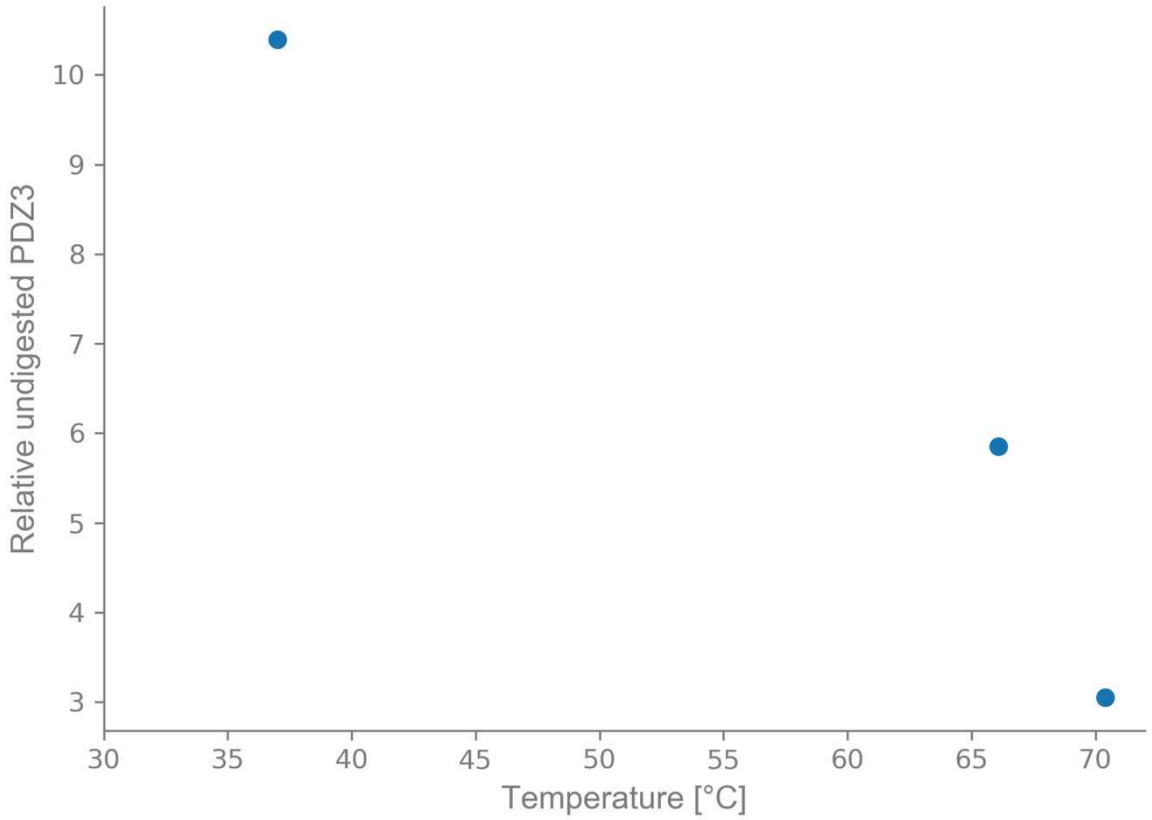
Suppl. Figure 26. Amino acid sequence of wild-type PSD-95^{PDZ3}.

Amino acid sequence of the rat PSD-95^{PDZ3} variant received as a gift from the Ranganathan lab. The molecular weight calculated with the ProtParam tool from ExPASy is 12.724 kDa (100).



Suppl. Figure 27. Electropherogram from the protein chip analysis of wild-type PSD-95^{PDZ3} with the adjusted protocol for low protein concentrations.

Spectrum produced by the 2100 Expert Software of a sample with 67 ng/µL PSD-95PDZ3 in PBS. The total amount of PSD-95PDZ3 loaded on the chip was 200 ng. The peak just below the 15 kDa tick corresponds to the undigested PSD-95PDZ3 with a molecular mass of 12.724 kDa (Suppl. Figure 26). The small peak to the left of the PSD-95PDZ3 peak might originate from slight degradations of the protein. The upper marker and the lower marker are well visible at 95 kDa and 1.6 kDa, respectively.



Suppl. Figure 28. LiP-Chip of PSD-95PDZ3 for selected temperatures.

The E/S ratio was 1/300, the incubation time 30 sec without preheating. The PSD-95^{PDZ3} concentration in the reaction was 67 ng/µL in a reaction volume of 15 µL. The assayed LiP temperatures were 37°C, 66.1°C, and 70.4°C. The amount of relative undigested PDZ3 results from the fluorescence intensity of the peak corresponding to the full PSD-95^{PDZ3} normalized by the intensity of the upper marker in the same well.

Material and Methods

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Monoclonal Anti-polyHistidine antibody produced in mouse	Sigma-Aldrich	Cat#H1029
Peroxidase AffiniPure Donkey Anti-Mouse IgG (H+L)	Interchim	Cat#715-035-150
A2 scFv	This work	N/A
Germline antibody library	This work	N/A
Bacterial and Virus Strains		
Escherichia coli BL21 (DE3)	New England BioLabs	Cat#C2527H
Escherichia coli BL21 (DE3) + pIT2-A2scFv	This work	N/A
Escherichia coli BL21 (DE3) + pHIS8-3 Darpa DHFR	This work	N/A
Escherichia coli BL21 (DE3) + pHIS8 ecDHFR I41V	This work	N/A
Escherichia coli BL21 (DE3) + pHIS8 ecDHFR I41A	This work	N/A
Escherichia coli BL21 (DE3) + pHIS8 ecDHFR C85L	This work	N/A
Escherichia coli BL21 (DE3) + pHIS8 ecDHFR I91A	This work	N/A
Escherichia coli T7 Express	New England BioLabs	Cat#2566H
Escherichia coli T7 Express + germline library	This work	N/A
Escherichia coli T7 Express + limited library	This work	N/A
Escherichia coli T7 Express + bNAb library	This work	N/A
Material, Chemicals, Peptides, and Recombinant Proteins		
HisPur Cobalt Resin	ThermoFisher	Cat#89964
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Sigma-Aldrich	Cat#I5502
Imidazole	Sigma-Aldrich	Cat#I5513
DL-Dithiothreitol solution (DTT)	Sigma-Aldrich	Cat#43816-10ML
Phosphate buffered saline (PBS)	Sigma-Aldrich	Cat#P4417
Skim Milk Powder	Sigma-Aldrich	Cat#70166
TWEEN 20	Sigma-Aldrich	Cat#P9416
TMB Enhanced One Component HRP Membrane Substrate	Sigma-Aldrich	Cat#T9455
Amersham Protran Nitrocellulose Blotting Membrane	Cytiva	Cat#10600003
Proteinase K (PK)	New England BioLabs	Cat#P8107S
Bovine Serum Albumin (BSA)	Sigma-Aldrich	Cat#A2153
α-Lactalbumin (BLA)	Sigma-Aldrich	Cat#L6385
PSD-95^{PDZ3} and point mutants	Ranganathan lab	N/A
ecDHFR	This work	N/A
ecDHFR I41V	This work	N/A

Critical Commercial Assays		
Econo-Column (1.0 x 20 cm)	Bio-Rad	Cat#7374152
Econo-Column Funnel	Bio-Rad	Cat#7310003
Agilent Protein 80 Kit	Agilent Technologies	Cat#5067-1515
2100 Bioanalyzer Instrument	Agilent Technologies	Cat#G2939BA
Float-a-lyzer g2 ce 3,5 - 5 kda 1 mL	Fisher Scientific	Cat#G235029
Novex 4-20% Tris-Glycine Mini Gels, WedgeWell format, 15-well	ThermoFisher	Cat#XP0425BOX
Novex Tris-Glycine SDS Running Buffer (10x)	ThermoFisher	Cat#LC26754
Novex™ 16% Tricine Protein Gels, 1.0 mm, 15-well	ThermoFisher	Cat#EC66955BOX
Novex 16% Tris-Glycine Mini Gels, WedgeWell format, 15-well	ThermoFisher	Cat#XP00165BOX
Novex™ Tricine SDS Buffer Kit	ThermoFisher	Cat#LC1677
PageBlue™ protein staining solution	ThermoFisher	Cat#24620
Recombinant DNA		
pIT2-A2 scFv	Soshee et al. (2014)	N/A
pHIS8-3 Darpa DHFR	Reynolds lab	N/A
pHIS8 DHFR mutants (I41V, I41A, C85L, I91A)	Kortemme lab	N/A
Software and Algorithms		
2100 Expert Software	Agilent Technologies	Cat#G2946CA

LiP-Chip protocol

LiP: (without preheating)

1. Thaw the protein aliquots (stored at -80°C) on ice
2. Remove aggregates:
    - Centrifuge for 5 min at 15 000 x g in a centrifuge at 4°C
    - Transfer the supernatant without aggregation pellet into a cold microfuge tube and continue the experiment with this
3. Measure the concentration of the protein (e.g. Eppendorf photometer D30 with µCuvette)
4. Dilute the protein to the desired concentration with cold buffer (e.g. PBS) and store the sample on ice
5. Thaw a PK aliquot on ice
    - Make sure that all following steps until step 8 are done before the PK is thawed
    - Continue immediately when PK is thawed to reduce the time of autolysis
6. Start the PCR machines and let them heat to the desired temperatures with the following protocol:
    - Lid heating at 100°C
    - *[Temperature/temperature range]* for 1 min
    - Pause
    - *[Temperature/temperature range]* for *[incubation time]*
    - 97°C for 5 min (adjust the ramp time to be approximately the same for all LiP temperatures)
    - 10°C for 5 min
    - 10°C on hold
7. Prepare PCR tubes according to the number of samples with 10 µL protein each and store them on ice
    - We recommend PCR tubes on a strip with individual caps for easy usage
    - Spin down the samples quickly in a microcentrifuge to make sure they are on the bottom of the tube

8. Make a dilution of PK with cold buffer to the desired E/S ratio and store it on ice
    - Fill the PK dilution into PCR tubes according to the protein samples to facilitate quick and easy pipetting with a multi-channel pipette
9. Add 10 µL PK to the prepared PCR tubes with protein
    - Use a multi-channel pipette to reduce the difference in digestion time between samples
    - Pipet to the bottom of the tube, mix by pipetting up and down, and spin quickly in a microcentrifuge
    - Proceed quickly to the next step!
10. Heat the mixed samples in the PCR machines with the above protocols (after 'Pause')
11. Take out the samples as soon as they have cooled down to 10°C for 5 min
12. Spin them down quickly in a microcentrifuge, shock freeze them in liquid nitrogen, and store them at -20°C

LiP: (with preheating)

Steps 1 to 4 are as described in 'LiP: (without preheating)'.

5. Start the PCR machines and let them heat to the desired temperatures with the following protocol:
    - Lid heating at 100°C
    - *[Temperature/temperature range]* for 1 min
    - Pause
    - *[Temperature/temperature range]* for 5 min
    - Pause
    - *[Temperature/temperature range]* for *[incubation time]*
    - 97°C for 5 min (adjust the ramp time to be approximately the same for all LiP temperatures)
    - 10°C for 5 min
    - 10°C on hold
6. Prepare PCR tubes according to the number of samples with 150 µL protein each and store them on ice

7. Thaw a PK aliquot on ice
8. Preheat the PCR tubes with the protein samples to the LiP temperatures
    - Make sure that the preheating is done the moment the PK is thawed and the dilution is made (see next step)
9. Make a dilution of PK with cold buffer to the desired E/S ratio and store it on ice
    - Fill the PK dilution into PCR tubes according to the protein samples to facilitate quick and easy pipetting with a multi-channel pipette
10. Add 5 µL PK to PCR tubes according to the number of LiP samples
    - Spin quickly in a microcentrifuge to ensure that the solution is on the bottom
11. Add 95 µL of preheated protein to the PCR tubes with PK
    - Use a multi-channel pipette to prevent differences in digestion time between samples
12. Heat the mixed samples in the PCR machines with the above protocols (after the second 'Pause')
13. Take out the samples as soon as they have cooled down to 10°C for 5 min
14. Spin them down quickly in a microcentrifuge, shock freeze them in liquid nitrogen, and store them at -20°C

Chip:

To reduce the time of the experiment, the protein chip can be prepared while the LiP step is conducted. This protocol is according to the Agilent protocol and contains additional notes from our experiences.

1. Start the computer and the 2100 Agilent Bioanalyzer, start the 2100 Expert Software and label the samples in the software
2. Clean the electrodes
3. Let the sample buffer (SB) (if required with DTT) and the ladder adjust to room temperature (RT) for 30 min
    - It is good to start with this 15-20 min before step 4
4. Take the gel-dye mix (GD) and the destaining solution (DS) out of the fridge and let it adjust to RT for at least 30 min protected from light
    - Do not let them at RT for longer than 1 h

5. While GD and DS adjust, vortex the SB and the ladder, spin them down in a microcentrifuge, fill ten PCR tubes with 2 µL SB each and store them at RT
6. Thaw the LiP samples on ice
7. Set a PCR machine to the following protocol:
    - TSP lid heating at 100°C
    - 95°C on hold
    - 95°C for 5 min
    - 25°C on hold
8. Add 4 µL of sample to the 2 µL SB prepared in step 5
    - Mix by pipetting up and down an spin quickly in a microcentrifuge
9. Heat the sample-SB mixes and the ladder to 95°C in the PCR machine for 5 min
10. When the PCR machine reaches 25°C, remove the samples, spin them in a microcentrifuge, let them further adjust to RT for 5 min, then spin them for 15 s
11. Add 84 µL Milli-Q water to all samples and the ladder, vortex for 5 s and spin them down quickly in a microcentrifuge
12. Set up the Chip Priming Station according to the Agilent protocol
13. Load 12 µL GD into the designated well and push down the plunger until it is held down by the Chip Priming Station. Wait for exactly 1 min. Release the plunger and let it rise on its own for 5 s. Very slowly bring the plunger back to the 1 mL mark
    - Insert the pipet tip to the bottom of the well and tilt it slightly to avoid air bubbles
    - The mixes are viscous, pay when pipetting to not introduce air bubbles
    - GD contains DMSO and needs to be handled with care
14. Add 12 µL GD into the other designated wells and 12 µL DS into its well
15. Add 6 µL sample into each of the ten sample wells and 6 µL of the ladder into well 11
    - Insert the pipet tip to the bottom and tilt it slightly to avoid air bubbles
16. Quickly place the chip into the Bioanalyzer and start the assay
17. Shock freeze the remaining samples in liquid nitrogen and store them at -20°C
18. When the assay is done, clean the electrodes

LiP-Chip protocol for low protein concentrations

LiP:

1. Thaw the protein aliquots (stored at -80°C) on ice
2. Remove aggregates:
    - Centrifuge for 5 min at 15 000 x g in a centrifuge at 4°C
    - Transfer the supernatant without aggregation pellet into a cold microfuge tube and continue the experiment with this
3. Measure the concentration of the protein (e.g. Eppendorf photometer D30 with µCuvette)
4. Dilute the protein to 100 ng/µL with cold buffer (e.g. PBS) and store the sample on ice
5. Thaw a PK aliquot on ice
    - Make sure that all following steps until step 8 are done before the PK is thawed
    - Continue immediately when PK is thawed to reduce the time of autolysis
6. Start the PCR machines and let them heat to the desired temperatures with the following protocol:
    - Lid heating at 100°C
    - *[Temperature/temperature range]* for 1 min
    - Pause
    - *[Temperature/temperature range]* for *[incubation time]*
    - 97°C for 5 min (adjust the ramp time to be approximately the same for all LiP temperatures)
    - 10°C for 5 min
    - 10°C on hold
7. Prepare PCR tubes according to the number of samples with 10 µL protein each and store them on ice
    - We recommend PCR tubes on a strip with individual caps for easy usage
    - Spin down the samples quickly in a microcentrifuge to make sure they are on the bottom of the tube

8. Make a dilution of PK with cold Milli-Q water to the desired E/S ratio and store it on ice
    - Fill the PK dilution into PCR tubes according to the protein samples to facilitate quick and easy pipetting with a multi-channel pipette
9. Add 5 µL PK to the prepared PCR tubes with protein
    - Use a multi-channel pipette to reduce the difference in digestion time between samples
    - Pipet to the bottom of the tube, mix by pipetting up and down, and spin quickly in a microcentrifuge
    - Proceed quickly to the next step!
10. Heat the mixed samples in the PCR machines with the above protocols (after 'Pause')
11. Take out the samples as soon as they have cooled down to 10°C for 5 min
12. Spin them down quickly in a microcentrifuge, shock freeze them in liquid nitrogen, and store them at -20°C

Chip:

1. Prepare the sample buffer (SB) (if required with DTT) 1:45 diluted with Milli-Q water (referred to as dSB)
    - Let the SB adjust to room temperature (RT) for 30 min
    - Make aliquots of 170 µL and store them at -20°C
2. Start the computer and the 2100 Agilent Bioanalyzer, start the 2100 Expert Software and label the samples in the software
3. Clean the electrodes
4. Let dSB and the ladder adjust to RT for 30 min
    - It is good to start with this 15-20 min before step 4
5. Take the gel-dye mix (GD) and the destaining solution (DS) out of the fridge and let it adjust to RT for at least 30 min protected from light
    - Do not let them at RT for longer than 1 h
6. While GD and DS adjust, vortex the dSB and the ladder, spin them down in a microcentrifuge and store them at RT
7. Thaw the LiP samples on ice

8. Set a PCR machine to the following protocol:
    - TSP lid heating at 100°C
    - 95°C on hold
    - 95°C for 5 min
    - 25°C on hold
9. Add 15 µL of dSB to each sample
    - Mix by pipetting up and down an spin quickly in a microcentrifuge
10. Heat the sample-dSB mixes and the ladder to 95°C in the PCR machine for 5 min
11. When the PCR machine reaches 25°C, remove the samples, spin them in a microcentrifuge, let them further adjust to RT for 5 min, then spin them for 15 s
12. Add 84 µL Milli-Q water to the ladder, vortex for 5 s and spin them down quickly in a microcentrifuge
13. Set up the Chip Priming Station according to the Agilent protocol
14. Load 12 µL GD into the designated well and push down the plunger until it is held down by the Chip Priming Station. Wait for exactly 1 min. Release the plunger and let it rise on its own for 5 s. Very slowly bring the plunger back to the 1 mL mark
    - Insert the pipet tip to the bottom of the well and tilt it slightly to avoid air bubbles
    - The mixes are viscous, pay when pipetting to not introduce air bubbles
    - GD contains DMSO and needs to be handled with care
15. Add 12 µL GD into the other designated wells and 12 µL DS into its well
16. Add 6 µL sample into each of the ten sample wells and 6 µL of the ladder into well 11
    - Insert the pipet tip to the bottom and tilt it slightly to avoid air bubbles
17. Quickly place the chip into the Bioanalyzer and start the assay
18. Shock freeze the remaining samples in liquid nitrogen and store them at -20°C
19. When the assay is done, clean the electrodes

SDS-PAGE quantification protocol

For most of the SDS-PAGE in this work, we used 4-20% Tris-Glycine gels and ran them for 90 to 130 min with 25 mA per gel and 200 V. For a few experiments, we used 16% Tris-Glycine gels and ran them for 90 min with 40 mA and 125 V. Due to the small size of PSD-95^{PDZ3}, we used 16% Tricine gels to facilitate higher resolution in the low kDa-range. For each sample well, we usually used a volume of 20 µL and protein amounts of 1 µg for full proteins and 2.5 µg to 10 µg for LiP samples.

Gel staining:

1. Submerge the gel in PageBlue^{TM} staining solution for 1 h on a rotating plate
2. Rinse the gel with ddH_2O and submerge it in ddH_2O for 10 min on a rotating plate
3. Repeat step 2 twice
4. Destain the gel in ddH_2O overnight on a rotating plate (Replace the water occasionally)

Band quantification:

This protocol is taken from the tutorial 'Densitometry using ImageJ' from the Sybil project (78). It used the freely available image processing software ImageJ.

1. Convert the image to 8-bit (Image → Type → 8-bit)
2. Highlight the band of interest in the first lane with the square selection tool and press Command + 1 (Ctrl + 1 on Windows)
3. Duplicate the square by clicking its center, drag it to the next lane, and press Command + 2. Repeat this for all lanes except for the last one
    - The squares will always be at the same height as the square in the first lane
4. Repeat setting the square for the last lane as in step 3, but press Command + 3. This opens a window with the lane plots
5. In the window with the lane plots, distinguish the peak of interest from the background by drawing a line from the beginning to the end of the peak with the line tool
6. Select the peaks of interest with the magic wand tool to obtain the area of the delimited peak. The report will automatically open in a new window
7. Use the area under the peaks in the 'Data analysis'

Dot blot protocol

1. Prepare 40 mL of a 2% skim milk solution in PBS
2. Cut the nitrocellulose blotting membrane and place it into a petri dish
3. Pipette 2 µL of each sample onto the membrane and let them dry for approximately 5 min
4. Add 20 µL of skim milk solution to the petri dish with the membrane and incubate on a rotating plate for 30 min
5. Add 15 µL Tween20, 7 µL of primary antibody (here: Monoclonal Anti-polyHistidine antibody produced in mouse) and 3 µL of secondary antibody (here: Peroxidase AffiniPure Donkey Anti-Mouse IgG (H+L)) to the rest of the skim milk solution and fill it into a new petri dish
6. Move the membrane into the new petri dish prepared in step 5 and incubate on a rotating plate for 1 h
7. Move the membrane to a new petri dish, cover it with Milli-Q water, and let it incubate for a few minutes
8. Dry the membrane between two sheets of paper and place it into a new petri dish
9. Add approximately 600 µL HRP membrane substrate over the membrane and let it incubate for a few minutes until the dots appear

Transformation protocol

If the plasmid is received on a filter paper, first resuspend it in the following way:

1. Place the filter paper into a microfuge tube using tweezers
2. Add 20 - 30 µL sterile water
3. Incubate at room temperature for 10 min
4. Spin for 1 min in a microcentrifuge
5. Use directly for transformations or store at -20°C

The transformation protocol mostly follows the protocols from New England BioLabs for their BL21 (DE3) and T7 Express strains.

1. Thaw the cells on ice for approximately 8 min
2. Add 2 µL of the plasmid to the cell mixture. Carefully flick 4-5 times
    - The total amount of plasmid should be between 1 pg and 100 ng
3. Incubate on ice for 30 min. Don't mix
4. Warm the outgrowth media to 37°C
5. Heat shock for 10 s in a water bath of 42°C. Don't mix
6. Place on ice for 30 s to 1 min
7. Pipette 950 µL warmed outgrowth media into the mixture. Don't put the cells on ice
8. Place into a shaker at 37°C for 1 h with strong shaking
9. Warm the selection plates to 37°C
10. Plate 100 µL of cells onto one plate
11. Pellet cells by centrifuging for 1 min in a microcentrifuge. Discard 800 µL of the supernatant, resuspend the cell pellet in the rest of the supernatant, plate the resuspended cells
12. Incubate plates overnight at 37°C

Expression and purification protocol - A2 scFv

Expression:

1. Make an overnight culture in 2x YT with 100 µg/mL ampicillin and 1% glucose
2. Incubate for approximately 20 h at 30°C (time can be reduced by incubating at 37°C)
3. Dilute the overnight culture to an optical density of 0.06 in 2x YT with 100 µg/mL ampicillin and 0.1% glucose
4. Incubate at 37°C until an optical density around 0.5 is reached
    - Takes approximately 1.5 h for BL21 (DE3) with pIT2-A2 scFv
5. Add IPTG to a concentration of 1 mM
    - Be aware the IPTG is light sensitive and should also not undergo freeze-thaw cycles
6. Incubate overnight at 30°C

Purification:

Work on ice or in the cold room as much as possible!

7. Harvest the supernatant:
    - Centrifuge 10 min, 8 000 x g, 4°C
    - Mover supernatant to new tubes
    - Centrifuge 10 min, 8 000 x g, 4°C
    - Filter supernatant with 0.2 µm filters
    - Keep the supernatant in the fridge or on ice
8. Clean the HisPur Cobalt Resin in batch:
    - Fill two 2-mL microfuge tubes with 500 µL resin each
    - Centrifuge 2 min, 700 x g, 4°C
    - Discard the supernatant
    - Add 1 mL PBS and mix
    - Centrifuge 2 min, 700 x g, 4°C
    - Discard the supernatant (the resin should not dry out. Mix it right away with the sample for purification or PBS)

9. Mix the resin with the bacterial culture supernatant containing the protein of interest and let them incubate for approximately 1 h at 4°C with stirring
10. Wash the Econo-Column (and -Funnel) with 70% ethanol, followed by 3-4 times with PBS
11. Fill the resin/supernatant mix into the Econo-Column and let it flow through. Collect a small sample of the flow through for further analysis
    - Make sure that the resin does not dry out
12. Wash at least three times with 1 mL PBS
    - Collect the last drops of the third wash and check the absorbance at 280 nm. Continue with the next step if it is close to the baseline, otherwise wash again with PBS
13. Elute the his-tagged protein in fractions with 1-2 mL PBS + 150 mM imidazole
    - Add the elution buffer while the latch is open and collect 3 to 5 drops for each fraction
14. Check the protein concentration in the fractions with dot blot and select up to 1 mL for dialysis
15. Clean the Econo-Column with Milli-Q water, 70% ethanol, and rinse with Milli-Q water
16. Dialyze up to 1 mL of the purified protein in 1 L PBS with three buffer exchanges (3 h, 4h, overnight)
17. Assess the protein amount with a spectrophotometer using the extinction coefficient (46090 $M^{-1}cm^{-1}$ for A2 scFv). For the validation of purity use SDS-PAGE.
18. Aliquot working stocks, shock freeze them in liquid nitrogen, and store them at -80°C

Expression and purification protocol - germline library

Expression:

1. Make an overnight culture at an optical density of approximately 0.1 in 2x YT with 100 µg/mL ampicillin and 1% glucose
2. Incubate for approximately 20 h at 30°C (time can be reduced by incubating at 37°C)
3. Dilute the overnight culture to an optical density of 0.06 in 2x YT with 100 µg/mL ampicillin and 0.1% glucose
4. Incubate at 37°C until an optical density around 0.1 is reached
    - Takes approximately 1 h for BL21 (DE3) with pIT2-germline library
5. Add IPTG to a concentration of 0.5 mM
    - Be aware the IPTG is light sensitive and should also not undergo freeze-thaw cycles
6. Incubate overnight at 30°C

Purification:

Work on ice or in the cold room as much as possible!

7. Harvest the supernatant:
    - Centrifuge 10 min, 8 000 x g, 4°C
    - Mover supernatant to new tubes
    - Centrifuge 10 min, 8 000 x g, 4°C
    - Filter supernatant with 0.2 µm filters
    - Keep the supernatant in the fridge or on ice
8. Clean the HisPur Cobalt Resin in batch:
    - Fill two 2-mL microfuge tubes with 500 µL resin each
    - Centrifuge 2 min, 700 x g, 4°C
    - Discard the supernatant
    - Add 1 mL PBS and mix
    - Centrifuge 2 min, 700 x g, 4°C

- Discard the supernatant (the resin should not dry out. Mix it right away with the sample for purification or PBS)

9. Mix the resin with the bacterial culture supernatant containing the protein of interest and let them incubate for approximately 1 h at 4°C with stirring
10. Wash the Econo-Column (and -Funnel) with 70% ethanol, followed by 3-4 times with PBS
11. Fill the resin/supernatant mix into the Econo-Column and let it flow through. Collect a small sample of the flow through for further analysis
    - Make sure that the resin does not dry out
12. Wash at least three times with 1 mL PBS
    - Collect the last drops of the third wash and check the absorbance at 280 nm. Continue with the next step if it is close to the baseline, otherwise wash again with PBS
13. Elute the his-tagged protein in fractions with 1-2 mL PBS + 150 mM imidazole
    - Add the elution buffer while the latch is open and collect 3 to 5 drops for each fraction
14. Check the protein concentration in the fractions with dot blot and select up to 1 mL for dialysis
15. Clean the Econo-Column with Milli-Q water, 70% ethanol, and rinse with Milli-Q water
16. Dialyze up to 1 mL of the purified protein in 1 L PBS with three buffer exchanges (3 h, 4h, overnight)
17. Assess the protein amount with a spectrophotometer using the extinction coefficient (35535 $M^{-1}cm^{-1}$ for mean original germline sequence). For the validation of purity use SDS-PAGE.
18. Aliquot working stocks, shock freeze them in liquid nitrogen, and store them at -80°C

Expression and purification protocol - ecDHFR

Expression:

1. Make an overnight culture in LB media with 35 µg/mL kanamycin
2. Incubate overnight at 37°C
3. Dilute the overnight culture 1/100 into Terrific broth with 35 µg/mL kanamycin
4. Incubate at 37°C until an optical density around 0.7 is reached
    - Takes approximately 1.5 h for BL21 (DE3) with pHis-8 ecDHFR I41V
    - Take a 1 mL sample for later analysis, spin 1 min in microcentrifuge, remove supernatant, shock freeze pellet in liquid nitrogen, and store at -20°C
5. Add IPTG to a concentration of 0.25 mM
    - Chill the cultures on ice until the incubator reaches 18°C
    - Be aware the IPTG is light sensitive and should also not undergo freeze-thaw cycles
6. Incubate overnight at 18°C

Purification:

Work on ice or in the cold room as much as possible!

7. Measure the optical density of the culture and take a sample that corresponds to the same amount of cells as the 1 mL sample taken in step 4. Spin the sample for 1 min in a microcentrifuge, remove the supernatant, shock freeze it and store it at -20°C
8. Harvest the cell pellet: Centrifuge 10 min, 6 000 x g, 4°C
9. Remove the supernatant and weigh the cell pellet into flacon tubes and resuspend it 1:5 (w:v) in PBS
    - The weight depends on the maximum volume for the sonicator samples. In our case, we used 1-1.5 g in approximately 5 mL PBS
10. Place the sample tubes on ice and lyse the cells by sonication
    - 3 x 15 s with 1 min pause
11. Use an ultracentrifuge to separate cell debris and supernatant: 30 min, 40 000 x g, 4°C
    - Take a 100 µL sample of the supernatant and a sample of the pellet. Shock freeze them in liquid nitrogen and store them at -20°C for quality control

12. Move the supernatant to a new tube
13. Clean the HisPur Cobalt Resin in batch:
    - Fill two 2-mL microfuge tubes with 500 µL resin each
    - Centrifuge 2 min, 700 x g, 4°C
    - Discard the supernatant
    - Add 1 mL PBS and mix
    - Centrifuge 2 min, 700 x g, 4°C
    - Discard the supernatant (the resin should not dry out. Mix it right away with the sample for purification or PBS)
14. Mix the resin with the bacterial culture supernatant containing the protein of interest and let them incubate for approximately 1 h at 4°C with stirring
15. Wash the Column (and -Funnel) with 70% ethanol, followed by 3-4 times with PBS
16. Fill the resin/supernatant mix into the Econo-Column and let it flow through. Collect a small sample of the flow through for further analysis
    - Make sure that the resin does not dry out
17. Wash at least three times with 1 mL PBS
    - Collect the last drops of the third wash and check the absorbance at 280 nm. Continue with the next step if it is close to the baseline, otherwise wash again with PBS
18. Elute the his-tagged protein in fractions with 1-2 mL PBS + 150 mM imidazole
    - Add the elution buffer while the latch is open and collect 3 to 5 drops for each fraction
19. Check the protein concentration in the fractions with dot blot and select up to 1 mL for dialysis
20. Clean the Econo-Column with Milli-Q water, 70% ethanol, and rinse with Milli-Q water
21. Dialyze up to 1 mL of the purified protein in 1 L PBS with three buffer exchanges (3 h, 4h, overnight)
22. Assess the protein amount with a spectrophotometer using the extinction coefficient (33585 $M^{-1}cm^{-1}$ for ecDHFR)
23. Aliquot working stocks, shock freeze them in liquid nitrogen, and store them at -80°C

MALDI-TOF MS protocol

Make sure to use appropriate equipment, such as glassware, specially treated plastic, and ultra-pure LC/MS water.

Reduction:

1. Thaw the samples, sonicate them, and spin them in a microcentrifuge
2. Sonicate the DTT vial for a short while
3. Dilute the samples 1:2 in PBS
4. Add 2 µL 100 mM DTT to each sample and incubate for 30 min at 56°C with agitation
5. Let the samples cool down for 5-10 min at room temperature (RT)

Alkylation:

6. Thaw the Iodoacetamide slowly at RT
    - Protect it from light
7. Add 2.1 µL Iodoacetamide to each sample
8. Incubate for 20 to 60 min at RT (protected from light)

Trypsin digestion:

9. Add trypsin to each sample at a 1/5 E/S ratio
10. Incubate for 2 to 16 h at 37°C with agitation
11. Let the samples cool down to RT
12. Add 5 µL of 5% Trifluoroacetic acid (TFA) to each sample
13. Verify on a pH-paper that the pH is below 5. If not, add more TFA in 1 µL steps
14. Store the samples at -20°C

Desalting:

15. Sonicate the samples to thaw them
16. Prepare two tubes with 50% Acetonitile (ACN) + 0.1% TFA, two tubes with 0.1% TFA, and one tube for waste
17. Use the Zip-Tip C18 tips
    - Maximum volume 10 µL

- Maximum load 1 µg
- Make sure not to pull in air!

18. Humidify the column
    - 3 x 10 µL 50% ACN + 0.1% TFA
    - Discard the first two times
    - Keep the last 10 µL to prevent the column from drying out
19. Wash the column
    - 3 x 10 µL 0.1% TFA
    - Discard into waste
20. Fix peptides
    - Pipet 10 µL of sample up and down 10 times
21. Wash the column
    - 4 x 10 µL 0.1% TFA
    - Discard to waste
22. Elute the peptides
    - Take 5 µL 50% ACN + 0.1% TFA and pipet into a new tube
    - Pipet up and down 5-10 times
    - Repeat the previously two steps twice without pipetting the already pipetted 5 µL
    - Keep the tubes open until the ACN is mostly evaporated

Plate preparation:

23. Clean the 96-well metal plate with 50% methanol and metal polish
24. Sonicate the 30% and 60% α-Cyano-4-hydroxycinnamic acid (HCCA) solvents
25. Add 250 µL 30% HCCA solvent to one tube with HCCA matrix and 250 µL 60% HCCA solvent to another tube of HCCA matrix. Vortex
    - 30% HCCA solvent + HCCA matrix = 30% HCCA solution
    - 60% HCCA solvent + HCCA matrix = 60% HCCA solution
26. Sonicate for 5-10 min

27. Sonicate the calibration standard
28. Add 10 µL 60% HCCA solution to the calibration standard
29. Sonicate for 5-10 min
30. Pipet 1 µL of calibration standard with HCCA solution into each of the 13 calibration spots on the 96-well metal plate
31. Sonicate the samples and bring them all up to 5 µL with 0.1% TFA
32. Pipet 0.3 µL of the samples into the circles on the 96-well metal plate
33. Add 0.6 µL 30% HCCA solution to the sample spots on the plate
    - Make sure the samples don't dry out before adding the HCCA solution
34. Let the plate dry
35. Set up the software 'TOFTOF Series Explorer'
36. Place the 96-well metal plate into a plate holder and clean it with compressed air and precision wipes
37. Load the plate into the instrument and start the laser
    - Laser needs to warm up and stabilize for 20 min, but turns off after 30 min when not used
38. Align the plate
    - The distortion factor (RMS) is computed. The closer it is to zero, the better. A RMS greater than 120 leads to the rejection of the alignment
39. Calibrate the plate
40. Test the plate by analyzing the calibration samples outside of the calibration method
41. Start the automatic data acquisition for the samples

Bibliography

1. Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, et al. A draft map of the human proteome. Nature. 2014 May;509(7502):575–81.

2. Human Proteome Map [Internet]. [cited 2020 May 4]. Available from: http://www.humanproteomemap.org

3. Goldenzweig A, Fleishman SJ. Principles of Protein Stability and Their Application in Computational Design. Annu Rev Biochem. 2018 Jun 20;87(1):105–29.

4. Ross CA, Poirier MA. Protein aggregation and neurodegenerative disease. Nat Med. 2004 Jul;10(S7):S10–7.

5. Stefani M. Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. Biochim Biophys Acta BBA - Mol Basis Dis. 2004 Dec;1739(1):5–25.

6. Zapadka KL, Becher FJ, Gomes dos Santos AL, Jackson SE. Factors affecting the physical stability (aggregation) of peptide therapeutics. Interface Focus. 2017 Dec 6;7(6):20170030.

7. Senske M, Törk L, Born B, Havenith M, Herrmann C, Ebbinghaus S. Protein Stabilization by Macromolecular Crowding through Enthalpy Rather Than Entropy. J Am Chem Soc. 2014 Jun 25;136(25):9036–41.

8. Roberts CJ. Therapeutic protein aggregation: mechanisms, design, and control. Trends Biotechnol. 2014 Jul;32(7):372–80.

9. Satyanarayana T, Noorwez SM, Kumar S, Rao JLUM, Ezhilvannan M, Kaur P. Development of an ideal starch saccharification process using amylolytic enzymes from thermophiles. Biochem Soc Trans. 2004 Apr 1;32(2):276–8.

10. Rubingh DN. Protein engineering from a bioindustrial point of view. Curr Opin Biotechnol. 1997 Aug;8(4):417–22.

11. McLaughlin Jr RN, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R. The spatial architecture of protein function and adaptation. Nature. 2012 Nov;491(7422):138–42.

12. Salinas VH, Ranganathan R. Coevolution-based inference of amino acid interactions underlying protein function. eLife. 2018 Jul 20;7:e34300.

13. Morcos F, Schafer NP, Cheng RR, Onuchic JN, Wolynes PG. Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. Proc Natl Acad Sci. 2014 Aug 26;111(34):12408–13.

14. Figliuzzi M, Jacquier H, Schug A, Tenaillon O, Weigt M. Coevolutionary Landscape Inference and the Context-Dependence of Mutations in Beta-Lactamase TEM-1. Mol Biol Evol. 2016 Jan;33(1):268–80.

15. Harano Y, Kinoshita M. Translational-Entropy Gain of Solvent upon Protein Folding. Biophys J. 2005 Oct;89(4):2701–10.

16. Ramprakash J, Doseeva V, Galkin A, Krajewski W, Muthukumar L, Pullalarevu S, et al. Comparison of the chemical and thermal denaturation of proteins by a two-state transition model. Anal Biochem. 2008 Mar;374(1):221–30.

17. Touchette NA, Perry KM, Matthews CR. Folding of dihydrofolate reductase from Escherichia coli. Biochemistry. 1986;25(19):5445–52.

18. Dumoulin M, Conrath K, Van Meirhaeghe A, Meersman F, Heremans K, Frenken LGJ, et al. Single-domain antibody fragments with high conformational stability. Protein Sci. 2009 Apr 13;11(3):500–15.

19. Rossky PJ. Protein denaturation by urea: Slash and bond. Proc Natl Acad Sci. 2008 Nov 4;105(44):16825–6.

20. Pace CN. [14] Determination and analysis of urea and guanidine hydrochloride denaturation curves. In: Methods in enzymology. Elsevier; 1986. p. 266–80.

21. Pace CN, Shaw KL. Linear extrapolation method of analyzing solvent denaturation curves. Proteins Struct Funct Bioinforma. 2000;41(S4):1–7.

22. Park C, Marqusee S. Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding. Nat Methods. 2005 Mar;2(3):207–12.

23. Ahmad S, Kumar V, Ramanand KB, Rao NM. Probing protein stability and proteolytic resistance by loop scanning: A comprehensive mutational analysis. Protein Sci. 2012 Mar;21(3):433–46.

24. Monsellier E, Bedouelle H. Quantitative measurement of protein stability from unfolding equilibria monitored with the fluorescence maximum wavelength. Protein Eng Des Sel. 2005 Sep 1;18(9):445–56.

25. Fuertes M. Thermodynamic stability of the C-terminal domain of the human inducible heat shock protein 70. Biochim Biophys Acta BBA - Proteins Proteomics. 2004 Jun;1699(1–2):45–56.

26. Leuenberger P, Ganscha S, Kahraman A, Cappelletti V, Boersema PJ, von Mering C, et al. Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. Science. 2017 Feb 24;355(6327):eaai7825.

27. Henderson R, Watts BE, Ergin HN, Anasti K, Parks R, Xia S-M, et al. Selection of immunoglobulin elbow region mutations impacts interdomain conformational flexibility in HIV-1 broadly neutralizing antibodies. Nat Commun. 2019 Dec;10(1):654.

28. Consalvi V, Chiaraluce R, Giangiacomo L, Scandurra R, Christova P, Karshikoff A, et al. Thermal unfolding and conformational stability of the recombinant domain II of glutamate dehydrogenase from the hyperthermophile Thermotoga maritima. Protein Eng Des Sel. 2000 Jul;13(7):501–7.

29. Thompson S, Zhang Y, Ingle C, Reynolds KA, Kortemme T. Modulating the cellular context broadly reshapes the mutational landscape of a model enzyme. BioRxiv. 2019;848010.

30. Montoliu-Gaya L, Martínez JC, Villegas S. Understanding the contribution of disulfide bridges to the folding and misfolding of an anti-Aβ scFv: Disulfide Bridges in scFv Folding and Misfolding. Protein Sci. 2017 Jun;26(6):1138–49.

31. Guo J, Luk LYP, Loveridge EJ, Allemann RK. Thermal Adaptation of Dihydrofolate Reductase from the Moderate Thermophile *Geobacillus stearothermophilus*. Biochemistry. 2014 May 6;53(17):2855–63.

32. Tian J, Woodard JC, Whitney A, Shakhnovich EI. Thermal Stabilization of Dihydrofolate Reductase Using Monte Carlo Unfolding Simulations and Its Functional Consequences. Briggs JM, editor. PLOS Comput Biol. 2015 Apr 23;11(4):e1004207.

33. Bershtein S, Serohijos AWR, Bhattacharyya S, Manhart M, Choi J-M, Mu W, et al. Protein Homeostasis Imposes a Barrier on Functional Integration of Horizontally Transferred Genes in Bacteria. Achtman M, editor. PLOS Genet. 2015 Oct 20;11(10):e1005612.

34. Rodrigues JV, Bershtein S, Li A, Lozovsky ER, Hartl DL, Shakhnovich EI. Biophysical principles predict fitness landscapes of drug resistance. Proc Natl Acad Sci. 2016 Mar 15;113(11):E1470–8.

35. Narayan A, Bhattacharjee K, Naganathan AN. Thermally versus Chemically Denatured Protein States. Biochemistry. 2019 May 28;58(21):2519–23.

36. Ptáček P, Šoukal F, Opravil T. Introduction to the Transition State Theory. In: Ptáček P, Opravil T, Šoukal F, editors. Introducing the Effective Mass of Activated Complex and the Discussion on the Wave Function of this Instanton [Internet]. InTech; 2018 [cited 2020 Jun 30]. Available from: http://www.intechopen.com/books/introducing-the-effective-mass-of-activated-complex-and-the-discussion-on-the-wave-function-of-this-instanton/introduction-to-the-transition-state-theory

37. Royer CA. Probing Protein Folding and Conformational Transitions with Fluorescence. Chem Rev. 2006 May;106(5):1769–84.

38. Ghisaidoobe A, Chung S. Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Förster Resonance Energy Transfer Techniques. Int J Mol Sci. 2014 Dec 5;15(12):22518–38.

39. Manning MC. Protein structure and stability assessment by circular dichroism spectroscopy. In ACS Publications; 1993.

40. Martin SR, Schilstra MJ. Circular dichroism and its application to the study of biomolecules. Methods Cell Biol. 2008;84:263–93.

41. Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc. 2006 Dec;1(6):2876–90.

42. Smith CA, Shi CA, Chroust MK, Bliska TE, Kelly MJS, Jacobson MP, et al. Design of a Phosphorylatable PDZ Domain with Peptide-Specific Affinity Changes. Structure. 2013 Jan;21(1):54–64.

43. Minde DP, Maurice MM, Rüdiger SGD. Determining Biophysical Protein Stability in Lysates by a Fast Proteolysis Assay, FASTpp. Uversky VN, editor. PLoS ONE. 2012 Oct 3;7(10):e46147.

44. Prenner E, Chiu M. Differential scanning calorimetry: An invaluable tool for a detailed thermodynamic characterization of macromolecules and their interactions. J Pharm Bioallied Sci. 2011;3(1):39.

45. Freire E. Differential scanning calorimetry. In: Protein stability and folding. Springer; 1995. p. 191–218.

46. Senisterra G, Chau I, Vedadi M. Thermal denaturation assays in chemical biology. Assay Drug Dev Technol. 2012;10(2):128–36.

47. Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc. 2007 Sep;2(9):2212–21.

48. Ericsson UB, Hallberg BM, DeTitta GT, Dekker N, Nordlund P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal Biochem. 2006 Oct;357(2):289–98.

49. Kristensen P, Winter G. Proteolytic selection for protein folding using filamentous bacteriophages. Fold Des. 1998 Oct;3(5):321–8.

50. Rocklin GJ, Chidyausiku TM, Goreshnik I, Ford A, Houliston S, Lemak A, et al. Global analysis of protein folding using massively parallel design, synthesis, and testing. Science. 2017 Jul 14;357(6347):168–75.

51. Jespers L, Schon O, Famm K, Winter G. Aggregation-resistant domain antibodies selected on phage by heat denaturation. Nat Biotechnol. 2004 Sep;22(9):1161–5.

52. Famm K, Hansen L, Christ D, Winter G. Thermodynamically Stable Aggregation-Resistant Antibody Domains through Directed Evolution. J Mol Biol. 2008 Feb;376(4):926–31.

53. Klesmith JR, Bacik J-P, Wrenbeck EE, Michalczyk R, Whitehead TA. Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. Proc Natl Acad Sci. 2017 Feb 28;114(9):2265–70.

54. Neurath H. Protein Structure and Enzyme Action. Rev Mod Phys. 1959 Jan 1;31(1):185–90.

55. Davie EW, Neurath H. Identification of a peptide released during autocatalytic activation of trypsinogen. J Biol Chem. 1955;212(2):515--530.

56. Scheraga HA, Laskowski Jr M. The fibrinogen-fibrin conversion. Elsevier; 1957. (Advances in Protein Chemistry; vol. 12).

57. Schechter I, Berger A. On the size of the active site in proteases. I. Papain. Biochem Biophys Res Commun. 1967 Apr;27(2):157–62.

58. Fontana A, De Laureto PP, Spolaore B, Frare E, Picotti P, Zambonin M. Probing protein structure by limited proteolysis. Acta Biochim Pol. 2004 Jun 30;51(2):299–321.

59. Arnold U, Köditz J, Markert Y, Ulbrich-Hofmann R. Local fluctuations vs. global unfolding of proteins investigated by limited proteolysis. Biocatal Biotransformation. 2005 Jan;23(3–4):159–67.

60. Cleveland DW, Fischer SG, Kirschner MW, Laemmli UK. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. J Biol Chem. 1977 Feb 10;252(3):1102–6.

61. Dieckmann R, Pavela-Vrancic M, von Döhren H, Kleinkauf H. Probing the domain structure and ligand-induced conformational changes by limited proteolysis of tyrocidine synthetase 1. J Mol Biol. 1999 Apr;288(1):129–40.

62. Feng Y, De Franceschi G, Kahraman A, Soste M, Melnik A, Boersema PJ, et al. Global analysis of protein structural changes in complex proteomes. Nat Biotechnol. 2014 Oct;32(10):1036–44.

63. Fontana A, Polverino de Laureto P, De Filippis V, Scaramella E, Zambonin M. Probing the partly folded states of proteins by limited proteolysis. Fold Des. 1997 Apr;2(2):R17–26.

64. Schopper S, Kahraman A, Leuenberger P, Feng Y, Piazza I, Müller O, et al. Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry. Nat Protoc. 2017 Nov;12(11):2391–410.

65. Polverino de Laureto P, Taddei N, Frare E, Capanni C, Costantini S, Zurdo J, et al. Protein Aggregation and Amyloid Fibril Formation by an SH3 Domain Probed by Limited Proteolysis. J Mol Biol. 2003 Nov;334(1):129–41.

66. Picotti P, Marabotti A, Negro A, Musi V, Spolaore B, Zambonin M, et al. Modulation of the structural integrity of helix F in apomyoglobin by single amino acid replacements. Protein Sci. 2004 Jun;13(6):1572–85.

67. Lyer LK, Phanse R, Xu M, Lan W, Krause ME, Bolgar M, et al. Pulse Proteolysis: An Orthogonal Tool for Protein Formulation Screening. J Pharm Sci. 2019 Feb;108(2):842–50.

68. Hubbard SJ. The structural aspects of limited proteolysis of native proteins. Biochim Biophys Acta BBA - Protein Struct Mol Enzymol. 1998 Feb;1382(2):191–206.

69. Mesbah K, Thai R, Bregant S, Malloggi F. DMF-MALDI: droplet based microfluidic combined to MALDI-TOF for focused peptide detection. Sci Rep. 2017 Dec;7(1):6756.

70. Brunelle JL, Green R. One-dimensional SDS-polyacrylamide gel electrophoresis (1D SDS-PAGE). In: Methods in enzymology. Elsevier; 2014. p. 151–9.

71. Chevalier F. Standard Dyes for Total Protein Staining in Gel-Based Proteomic Analysis. Materials. 2010 Oct 20;3(10):4784–92.

72. Bousse L, Mouradian S, Minalla A, Yee H, Williams K, Dubrow R. Protein Sizing on a Microchip. Anal Chem. 2001 Mar;73(6):1207–12.

73. Michnik A. Thermal stability of bovine serum albumin DSC study. J Therm Anal Calorim. 2003 Feb 1;71(2):509–19.

74. Chaudhuri A, Chattopadhyay A. Lipid binding specificity of bovine α-lactalbumin: A multidimensional approach. Biochim Biophys Acta BBA - Biomembr. 2014 Aug;1838(8):2078–86.

75. Greene LH, Grobler JA, Malinovskii VA, Tian J, Acharya KR, Brew K. Stability, activity and flexibility in α-lactalbumin. Protein Eng Des Sel. 1999 Jul;12(7):581–7.

76. Apenten RKO. A three-state heat-denaturation of bovine α-lactalbumin. Food Chem. 1995;52(2):131–3.

77. Permyakov EA, Berliner LJ. α-Lactalbumin: structure and function. FEBS Lett. 2000;473(3):269–74.

78. Densitometry using ImageJ | SYBIL [Internet]. [cited 2020 Apr 20]. Available from: https://www.sybil-fp7.eu/node/95

79. DeLean A, Munson PJ, Rodbard D. Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves. Am J Physiol-Endocrinol Metab. 1978 Aug 1;235(2):E97.

80. Sebaugh JL, McCray PD. Defining the linear portion of a sigmoid-shaped curve: bend points. Pharm Stat. 2003 Jul;2(3):167–74.

81. B. Efron. Bootstrap Methods: Another Look at the Jackknife. Ann Stat. 1979;7(1):1–26.

82. Sigma Aldrich. Product information - Albumin from bovine serum.

83. Mateus A, Kurzawa N, Becher I, Sridharan S, Helm D, Stein F, et al. Thermal proteome profiling for interrogating protein interactions. Mol Syst Biol. 2020;16(3):e9232.

84. Edelman GM, Cunningham BA, Gall WE, Gottlieb PD, Rutishauser U, Waxdal MJ. The covalent structure of an entire γG immunoglobulin molecule. Proc Natl Acad Sci. 1969;63(1):78–85.

85. Wu TT, Kabat EA. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J Exp Med. 1970;132(2):211–50.

86. Xu JL, Davis MM. Diversity in the CDR3 Region of VH Is Sufficient for Most Antibody Specificities. Immunity. 2000 Jul;13(1):37–45.

87. Huston JS, Levinson D, Mudgett-Hunter M, Tai MS, Novotny J, Margolies MN, et al. Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in Escherichia coli. Proc Natl Acad Sci. 1988 Aug 1;85(16):5879–83.

88. Ahmad ZA, Yeap SK, Ali AM, Ho WY, Alitheen NBM, Hamid M. scFv Antibody: Principles and Clinical Application. Clin Dev Immunol. 2012;2012:1–15.

89. Soshee A, Zürcher S, Spencer ND, Halperin A, Nizak C. General In Vitro Method to Analyze the Interactions of Synthetic Polymers with Human Antibody Repertoires. Biomacromolecules. 2014 Jan 13;15(1):113–21.

90. Wörn A, Plückthun A. Different Equilibrium Stability Behavior of ScFv Fragments: Identification, Classification, and Improvement by Protein Engineering. Biochemistry. 1999 Jul;38(27):8739–50.

91. Pedroso I, Irún MP, Machicado C, Sancho J. Four-State Equilibrium Unfolding of an scFv Antibody Fragment. Biochemistry. 2002 Aug;41(31):9873–84.

92. Fierke CA, Johnson KA, Benkovic SJ. Construction and evaluation of the kinetic scheme associated with dihydrofolate reductase from Escherichia coli. Biochemistry. 1987 Jun 30;26(13):4085–92.

93. Sawaya MR, Kraut J. Loop and Subdomain Movements in the Mechanism of Escherichia coli Dihydrofolate Reductase: Crystallographic Evidence. Biochemistry. 1997 Jan;36(3):586–603.

94. Schnell JR, Dyson HJ, Wright PE. Structure, Dynamics, and Catalytic Function of Dihydrofolate Reductase. Annu Rev Biophys Biomol Struct. 2004 Jun 9;33(1):119–40.

95. Boehr DD, McElheny D, Dyson HJ, Wright PE. The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis. Science. 2006 Sep 15;313(5793):1638–42.

96. Lee I, Suzuki CK. Functional mechanics of the ATP-dependent Lon protease- lessons from endogenous protein and synthetic peptide substrates. Biochim Biophys Acta BBA - Proteins Proteomics. 2008 May;1784(5):727–35.

97. Bondos S. Methods for Measuring Protein Aggregation. Curr Anal Chem. 2006 Apr 1;2(2):157–70.

98. Tanford Charles. Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins. J Am Chem Soc. 1962 Nov;84(22):4240–7.

99. Shaw KL, Grimsley GR, Yakovlev GI, Makarov AA, Pace CN. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. Protein Sci. 2001 Jun;10(6):1206–15.

100. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. In: The proteomics protocols handbook. Springer; 2005. p. 571–607.

101. Eftink MR. The use of fluorescence methods to monitor unfolding transitions in proteins. Biophys J. 1994 Feb;66(2):482–501.

102. Protein Purification - Extraction and Clarification - Choice of lysis buffer and additives - EMBL [Internet]. [cited 2020 Jun 19]. Available from: https://www.embl.de/pepcore/pepcore_services/protein_purification/extraction_clarification/lysis_buffer_additives/

103. Buffer Preparations and Recipes | AAT Bioquest [Internet]. [cited 2020 Jun 19]. Available from: https://www.aatbio.com/resources/buffer-preparations-and-recipes

104. Buffer Reference Center [Internet]. Sigma-Aldrich. [cited 2020 Jun 19]. Available from: https://www.sigmaaldrich.com/life-science/core-bioreagents/biological-buffers/learning-center/buffer-reference-center.html

105. Lui S, Tiana G. The network of stabilizing contacts in proteins studied by coevolutionary data. J Chem Phys. 2013 Oct 21;139(15):155103.

106. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. Proc Natl Acad Sci. 2006 Apr 11;103(15):5869–74.

107. Schulz S, Boyer S, Smerlak M, Cocco S, Monasson R, Nizak C, et al. Parameters and determinants of responses to selection in antibody libraries. bioRxiv. 2019;712539.

108. Shehata L, Maurer DP, Wec AZ, Lilov A, Champney E, Sun T, et al. Affinity Maturation Enhances Antibody Specificity but Compromises Conformational Stability. Cell Rep. 2019 Sep;28(13):3300-3308.e4.

109. Ovchinnikov V, Louveau JE, Barton JP, Karplus M, Chakraborty AK. Role of framework mutations and antibody flexibility in the evolution of broadly neutralizing antibodies. eLife. 2018 Feb 14;7:e33038.

110. Sakarya O, Conaco C, Egecioglu O, Solla SA, Oakley TH, Kosik KS. Evolutionary Expansion and Specialization of the PDZ Domains. Mol Biol Evol. 2010 May 1;27(5):1058–69.

111. Ponting CP. Evidence for PDZ domains in bacteria, yeast, and plants. Protein Sci. 2008 Dec 31;6(2):464–8.

112. Muley VY, Akhter Y, Galande S. PDZ Domains Across the Microbial World: Molecular Link to the Proteases, Stress Response, and Protein Synthesis. Gojobori T, editor. Genome Biol Evol. 2019 Mar 1;11(3):644–59.

113. Lee H-J, Zheng JJ. PDZ domains and their binding partners: structure, specificity, and modification. Cell Commun Signal. 2010 Dec;8(1):8.

114. Gautier C, Visconti L, Jemth P, Gianni S. Addressing the role of the α-helical extension in the folding of the third PDZ domain from PSD-95. Sci Rep. 2017 Dec;7(1):12593.

115. Nourry C, Grant SGN, Borg J-P. PDZ Domain Proteins: Plug and Play! Sci Signal. 2003 Apr 22;2003(179):re7–re7.

116. Lockless SW. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. Science. 1999 Oct 8;286(5438):295–9.

117. Raman AS, White KI, Ranganathan R. Origins of Allostery and Evolvability in Proteins: A Case Study. Cell. 2016 Jul;166(2):468–80.

118. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2018 Oct 24;47(D1):D427–32.

119. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein-protein interaction by message passing. Proc Natl Acad Sci. 2009 Jan 6;106(1):67–72.

120. Lunt B, Szurmant H, Procaccini A, Hoch JA, Hwa T, Weigt M. Inference of direct residue contacts in two-component signaling. In: Methods in enzymology. Elsevier; 2010. p. 17–41.

121. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci. 2011 Dec 6;108(49):E1293–301.

122. Cocco S, Feinauer C, Figliuzzi M, Monasson R, Weigt M. Inverse Statistical Physics of Protein Sequences: A Key Issues Review. Rep Prog Phys. 2018 Mar 1;81(3):032601.

123. Sulkowska JI, Morcos F, Weigt M, Hwa T, Onuchic JN. Genomics-aided structure prediction. Proc Natl Acad Sci. 2012 Jun 26;109(26):10340–5.

124. Feinauer C, Szurmant H, Weigt M, Pagnani A. Inter-Protein Sequence Co-Evolution Predicts Known Physical Interactions in Bacterial Ribosomes and the Trp Operon. Keskin O, editor. PLOS ONE. 2016 Feb 16;11(2):e0149166.

125. dos Santos RN, Morcos F, Jana B, Andricopulo AD, Onuchic JN. Dimeric interactions and complex formation using direct coevolutionary couplings. Sci Rep. 2015 Nov;5(1):13652.

Table of contents

Table des illustrations