The Effect of Point Mutations on the Biophysical Properties of an Antimicrobial Peptide Development of a Screening Protocol for Peptide Stability Screening

Pohl, Christin Alina; Zalar, Matja; Bialy, Inas El; Indrakumar, Sowmya; Peters, Günther H.J.; Friess, Wolfgang; Golovanov, Alexander P.; Streicher, Werner W.; Noergaard, Allan; Harris, Pernille

Published in:
Molecular Pharmaceutics

Link to article, DOI:
[10.1021/acs.molpharmaceut.0c00406](https://doi.org/10.1021/acs.molpharmaceut.0c00406)

Publication date:
2020

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Pohl, C. A., Zalar, M., Bialy, I. E., Indrakumar, S., Peters, G. H. J., Friess, W., Golovanov, A. P., Streicher, W. W., Noergaard, A., & Harris, P. (2020). The Effect of Point Mutations on the Biophysical Properties of an Antimicrobial Peptide: Development of a Screening Protocol for Peptide Stability Screening. *Molecular Pharmaceutics*, 17(9), 3298-3313. <https://doi.org/10.1021/acs.molpharmaceut.0c00406>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Article

The effect of point mutations on the biophysical properties of an anti-microbial peptide: development of a screening protocol for peptide stability screening

Christin Pohl, Matja Zalar, Inas El Bialy, Sowmya Indrakumar, Günther H. J. Peters, Wolfgang Friess, Alexander P. Golovanov, Werner W Streicher, Allan Noergaard, and Pernille Harris

Mol. Pharmaceutics, **Just Accepted Manuscript** • DOI: 10.1021/acs.molpharmaceut.0c00406 • Publication Date (Web): 01 Jul 2020

Downloaded from pubs.acs.org on July 6, 2020

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The effect of point mutations on the biophysical properties of an anti-microbial peptide: development of a screening protocol for peptide stability screening

Christin Pohl^{1,2}, Matja Zalar³, Inas El Bialy⁴, Sowmya Indrakumar², Günther H.J. Peters², Wolfgang Friess⁴, Alexander P. Golovanov³, Werner w. Streicher¹, Allan Noergaard^{1*} and Pernille Harris^{2*}

¹Novozymes A/S, Krogshoejvej 36, 2880, Bagsvaerd, Denmark
²Technical University of Denmark, Department of Chemistry, Kemitorvet 207, 2800 Kongens Lyngby, Denmark
³Manchester Institute of Biotechnology and Department of Chemistry, Faculty of Science and Engineering, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK.
⁴Ludwig-Maximilians-Universitaet Muenchen; Department of Pharmacy; Pharmaceutical Technology and Biopharmaceutics; Butenandtstrasse 5, 81377 Muenchen, Germany
*corresponding authors: ph@kemi.dtu.dk; alnq@novozymes.com

ABSTRACT

Therapeutic peptides and proteins show an enormous potential on the pharmaceutical market, but high costs in discovery and development are limiting factors so far. Single or multiple point mutations are commonly introduced in protein drugs to increase their binding affinity or selectivity. They can also induce adverse properties, which might be overlooked in a functional screen, such as a decreased colloidal or thermal stability, leading to problems in later stages of the development. In this study, we address the effect of point mutations on the stability of the 4.4 kDa anti-microbial peptide plectasin, as a case study. We combined a systematic high throughput biophysical screen of the peptide thermal and colloidal stability using dynamic light scattering and differential scanning calorimetry in combination with structure-based methods including small angle X-ray scattering, analytical ultracentrifugation and nuclear magnetic resonance spectroscopy. Additionally, we applied molecular dynamics simulations to link obtained protein stability parameters to the protein’s molecular structure. Despite their predicted structural similarities, all four plectasin variants showed a substantially different behavior in solution. We observed an increasing propensity of plectasin to aggregate at higher pH, and the introduced mutations influenced the type of aggregation. Our strategy for systematically assessing the stability and aggregation of protein drugs is generally applicable and is of particular relevance given the increasing number of protein drugs in development.

KEYWORDS

protein-protein interaction; protein characterization; protein aggregation; protein engineering; aggregation assessment; pharmaceutical screening; peptide screening

INTRODUCTION

Protein and peptide drugs exhibit enormous clinical potential due to their high specificity and low toxicity. Subsequently, the number of peptide and protein therapeutics increased substantially in the last decade ^{1,2}. However, in contrast to small molecules, they face different challenges with respect to their production, formulation and administration and maintaining the stability of protein drugs in solution remains a big challenge

^{3,4}. In particular, aggregation has been shown to be a major issue in protein drug development since it can result in reduced activity ⁵ and immune response ⁶⁻⁸. The study of protein aggregates, aggregation pathways and factors is widely discussed in academia and industry ⁹⁻¹¹. A comparison of different types of aggregates or self-associating species remains complicated ¹². The pH value has been shown to have a major influence on protein aggregation ^{13,14}. With automation and technological advances, high throughput screening of protein drugs for their thermal and colloidal stability became standard for assessing manufacturability ¹⁵. In addition, the use of computer based methods shows great potential to screen proteins before going into later stages of development ¹⁶. Therefore, the development of systematic approaches as well as the public availability of screening data for a high number of different protein drug candidate molecules is of high importance to accelerate and improve protein drug development.

In this study we developed a strategy to screen and characterize the effect of point mutations on protein stability and aggregation as a function of pH. As a case study, we applied the approach on the anti-microbial peptide plectasin and three triple-point mutation variants. Plectasin is a 40-amino acid anti-microbial peptide with a distinct tertiary structure, which shows antibacterial activity against Gram positive bacteria ¹⁷ (Figure 1). It binds the pyrophosphate unit of Lipid II, preventing its assembly into the bacterial cell wall ¹⁸. Its high-resolution structure was solved using NMR spectroscopy (PDBID: 1ZFU) and subsequently by X-ray crystallography on a racemic crystal (PDBID: 3E7U) ^{17,19}. Plectasin, shown in Figure 1A, consists of an α -helix (M13 - S21), two antiparallel β -strands (G28 - A31; V36 - C39) and three disulfide bridges (C4-C30, C15-C37, C19-C39). The wildtype, N5S D9A K26R (Figure 1B), D9S Q14K V36L (Figure 1C) and D9N M13L Q14R (Figure 1D) triple-mutant variants are referred to as PPI40, PPI41, PPI42 and PPI43, respectively, in this article. The mutations were introduced in an attempt to increase the biological activity of the peptide. In PPI41, the D9A mutation replaces a negatively charged amino acid with a more hydrophobic amino acid, increasing the isoelectric point (pI) of this variant. The other two mutations in PPI41 stay within the same charge/hydrophobicity/polarity group of amino acids as the amino acid found in the wildtype. In PPI42, the D9S mutation exchanging a negatively charged for a polar amino acid and the Q14K mutation exchanging a polar for a positively charged amino acid, which led to a higher pI of the protein. The V36L mutation is an exchange of hydrophobic amino acids residues. In PPI43, the D9N and the Q14R mutation increased the pI of the protein. The M13L mutation is an exchange of hydrophobic amino acids. Generally, the charge of a protein depends on the pH, and within the typical formulation pH range the plectasin variants show a higher net positive charge than the plectasin wildtype (Figure 1F). The theoretical isoelectric point increases by 0.5 -0.9 pH units. The molecular weight of plectasin is not considerably changed by the mutations (Figure 1F). The grand average hydropathy (GRAVY) values ²⁰ indicate that PPI41 is slightly more hydrophobic than the other variants and that overall the plectasins are rather hydrophilic.

selected conditions where more labor-intensive methods were used to study selected formulations. Oligomerization and aggregation of plectasin was structurally characterized using nuclear magnetic resonance spectroscopy (NMR), analytical ultracentrifugation (AUC) and small-angle X-ray scattering (SAXS). Additionally, molecular dynamics (MD) simulations were performed to simulate the protein structure of the variants and identify changes in surface charge, aggregation prone areas and flexibility as a function of pH in relation to the mutated amino acids in order to understand the experimental observations on a molecular level. In a high throughput screening plectasin was screened for wide pH range and ionic strength. We chose pH as the parameter for the detailed screening as it is known to be a key parameter to influence protein aggregation²⁶ and we observed significant differences among the variant. Our results showed a high structural similarity for all variants but a marked difference in aggregation propensity and solution stability. Thus, our findings emphasize the difficulties in protein drug development when it comes to the use of platform technologies. In our study, we demonstrate how a combination of high throughput methods and structural investigation can help to understand protein colloidal stability and aggregation in a better way, which is a step towards a more time- and cost-effective way in developing and screening protein drugs.

Figure 2: Schematic overview of the experimental approach applied in this study. The methods and conditions shown in black represent the actual applied methods in this study. The methods and conditions shown in grey represent alternative methods that have not been used.

MATERIAL AND METHODS

Sample preparation

If not stated otherwise, all plectasin samples were dialyzed into the desired condition using Slide-A-Lyzer™ 2000 MWCO dialysis cassettes (Thermo Fisher) with a buffer exchange after 2 and 4 hours and then continued overnight, ensuring a 200-fold dilution each step. The protein concentration after dialysis was measured using a NanoDrop™ 8000 Spectrophotometer (ThermoFisher). If not stated otherwise, all measurements were

performed at a protein concentration of 2 mg/ml (455 μ M), obtained by dilution into the final condition. Filtration of the sample and buffer was performed using Luer-Lok™ syringes (BD) and 0.22 μ m Millex®-GV filter (Merck) or 0.02 μ m Whatman Anotop™ 10 filter (GE Healthcare Life Science). The pH reported was measured on the dialysis buffer if not otherwise stated.

DSC measurements

All plectasin samples were dialyzed in 10 mM Acetate pH 4.5 as described above and the concentration was measured. A protein stock solution of 20 mg/ml was obtained by dilution with the filtered dialysis buffer (0.2 μ m). The respective formulations were obtained by a 10 times dilution into the final condition, taken pH shift through dilution into account. Measurement was performed with a MicroCal VP-Capillary DSC (Malvern Panalytical) with a scanning rate of 200°C/h from 25 to 125°C. The change in heat capacity ΔC_p [cal/°C] was detected. Analysis was performed with Origin 7 including the MicroCal VP-Capillary add-in. The respective buffer spectrum was deducted and the baseline before and after the measurement was picked. The thermal unfolding point of the protein, $T_{1/2}$, was chosen from the detected peak maximum in change in heat capacity. Final graphs were made with Origin® 2019 (OriginLabs). All measurements were performed in biological triplicates.

DLS measurements

All plectasin samples were dialyzed in 10 mM Acetate pH 4.5 as described above, filtered (0.02 μ m) and the concentration was measured. A protein stock solution of 20 mg/ml was obtained by dilution with the filtered dialysis buffer (0.02 μ m). The respective formulations were obtained by a 10 times dilution. The measurement was performed with a DynaPro® Plate Reader™ II (Wyatt Technology) using Aurora 384 LV/EB plates (Brookes Life Science Systems). Silicone oil (Sigma Aldrich) was used for sealing the wells. All measurements were performed isothermally at 25°C with 5 s acquisition time and 20 acquisitions per well. All formulations were measured in technical triplicates. t_0 measurement was performed at room temperature on three independent samples (biological triplicates), before one was taken for the stress study. The cumulant fit was used for analysis. Analysis was performed with DYNAMICS version 7.8.1.3 and final graphs were made with Origin® 2019 (OriginLabs).

DLS/SLS B_{22} and k_D measurements

All plectasin variants were dialyzed into 10 mM Acetate (pH 3.5, 4.0, 4.5, 5.0, and 5.5) as described above. Upon dialysis the pH was measured and adjusted to the desired pH using hydrochloric acid or sodium hydroxide. For each formulation, a concentration series of 10 points was prepared with concentrations between 3 and 16 mg/ml. All samples were centrifuged at 10,000 rpm for 10 minutes before measurement to eliminate large aggregates. Samples were filled in triplicates of 10 μ l into Aurora 1536 well plates. The measurements were done using DynaPro® Plate Reader™ III (Wyatt Technology) where SLS and DLS were measured simultaneously. The measurements were done isothermally at 25 °C with 60 percent laser power. 15 acquisitions were measured with 30 seconds acquisition time. All plates used were calibrated using Dextran standard at concentrations 0 to 10 mg/ml. Silicone oil (Sigma Aldrich) was used for sealing the wells. Analysis was performed using DYNAMICS version 7.8.2.18 and final graphs were made with Origin® 8 SR 2 (OriginLabs). Only acquisitions with autocorrelation functions displaying monomeric behavior were taken for the B_{22} and k_D calculations.

NMR spectroscopy

All plectasins samples were prepared as described above. Upon dialysis the pH was measured and adjusted to the desired pH using hydrochloric acid or sodium hydroxide. All NMR experiments were acquired at 25 °C on Bruker Avance III 800 MHz spectrometer equipped with 5mm TCI cryoprobe and temperature control unit. Data was acquired and processed using Topspin 3.5 (Bruker), and further analyzed in Dynamics Center 2.5.1 (Bruker) and OriginPro 9.1 (OriginLab).

At each pH value, a combination of four experiments was acquired to characterize effect of pH on plectasins structure and stability. At each pH point, 1D ^1H , 2D natural abundance ^1H - ^{13}C -heteronuclear single quantum correlation (HSQC) experiment with sensitivity enhancement, gradient coherence selection, and multiplicity editing was acquired. Translational self-diffusion coefficients (D_t) were measured by standard pulse program stebppg1s19 with additional water signal presaturation during the relaxation delay. Diffusion time and gradient length were set to 49.9 and 4 ms, respectively. D_t values were derived by fitting the data to the Stejskal-Tanner equation in Dynamics Center 2.5.1 (Bruker), with the errors in D_t estimated using Monte Carlo simulations, as implemented in Dynamics Center. Apparent hydrodynamic radius of protein clusters, R_h^{NMR} , as derived from NMR diffusion measurements, was calculated using Stokes-Einstein equation:

$$R_h^{\text{NMR}} = \frac{kT}{6\pi D_t \eta}$$

Where T is the absolute temperature, k is the Boltzmann constant and η is the viscosity. As all measurements were performed in dilute conditions, the viscosity was set to the viscosity of water at 298 K and was 0.89 mPas.

Changes in proton transverse relaxation rates (R_2) were monitored using a series of modified Carr-Purcell-Meiboom-Gill (CPMG) experiments with temperature compensation. The fixed echo time to allow elimination of diffusion and J-modulation effects was 1.4 ms. The T_2 values were determined by fitting the data to a single exponential function in the Dynamics Center 2.5.1. The T_2 errors were estimated using the Monte Carlo simulation as implemented in Dynamics Center 2.5.1 (Bruker).

SAXS measurements

Sample preparation was performed as described above. All samples and dialysis buffers were filtered (0.02 μm). The absorbance of all samples was measured using a NanoDropTM 8000 Spectrophotometer (ThermoFisher). Upon dialysis the pH was measured and adjusted to the desired pH using hydrochloric acid or sodium hydroxide. A concentration series was obtained by dilution with the dialysis buffer. Measurements were performed at the German Electron Synchrotron DESY at the P12 EMBL BioSAXS beamline and The European Synchrotron ESRF at BM29 BioSAXS beamline. The data analysis was performed using the ATSAS software package version 2.8.4²⁷. Final graphs were made with Origin[®] 2019(OriginLabs). The summary of the SAXS data collection is shown in Table S1.

AUC measurements

All samples were diluted with the filtered (0.02 μm) dialysis buffer with a 10 times dilution before dialysis. Dialysis was performed for 48h with exchanging the buffer 10 times during dialysis, ensuring a dilution of at least 200-fold each step. The absorbance of all samples was measured using a NanoDropTM 8000 Spectrophotometer (ThermoFisher). All samples were diluted to an absorbance of $A_{280} = 1$ (0.45 mg/ml). The

measurements were performed as a sedimentation velocity run with an Optima XL-1 ultracentrifuge (Beckman-Coulter) at 50000 rpm. Data analysis was performed using sedfit version 16.1c [28][29].

Constant pH MD simulations

The crystal structure of plectasin wildtype (PDB: 3E7U) served as a template to introduce the mutations to construct the plectasin variants. Minimization was performed with the PrepWiz tool from Schrödinger release 2016-3 (Schrödinger, LLC, New York, NY, USA)³⁰ and the protonation state of titratable amino acids in dependency of pH was determined using the H++ server (<http://biophysics.cs.vt.edu/H++>)³¹ and fixed during the simulations. The internal dielectric constant was set to 10 and the external dielectric constant to 80. The PDB structure was generated using the output topology and coordinate files (Amber format) using ambpdb tool³². MD simulation were performed in Amber as all-atom classical constant pH molecular dynamics (MD) simulations³³ in explicit solvent. The MD simulation setup is explained in detail in our previous work³⁴. They were carried out in Amber 16 implementing the amber force field ff99SB³⁵ with water represented using the TIP3P water model³⁶. Hydrogen bonds were constrained using the SHAKE algorithm³⁷. Each variant was simulated at five different pH ranging from 3.5 to 5.5 with a 0.5 increment corresponding to the pH range used in the experiments. A truncated octahedron water box with 15Å padding in all directions was used. Each system was simulated for 100 ns in duplicates starting from a random seed number.

The simulations were analyzed with CPPTRAJ³⁸ in Amber16, and VMD 1.9.3³⁹. Clustering was based on grouping similar conformations into distinct clusters based on agglomerative hierarchical cluster method^{40,41}.

RESULTS

Thermal stability of plectasin variants

Plectasin wildtype PPI40 and the three variants PPI41, PPI42 and PPI43 were initially characterized towards their thermal and colloidal stability at a wide pH range from pH 3.5 to pH 9.0 using differential scanning calorimetry (DSC) and dynamic light scattering (DLS). However, PPI40 and PPI42 showed aggregation and precipitation at neutral pH. Therefore, plectasin was further characterized at an acidic pH range from 3.5 to 5.5 with 0.5 units increments, where no visible aggregates or precipitation could be observed. The addition of sodium chloride did not improve plectasins' thermal and colloidal stability in step 1 of our study (data not shown). Therefore, plectasins' were screened only on their response to pH in step 2. Plectasin showed a two-state thermal induced unfolding measured with DSC (Figure S1). Plectasins' thermal unfolding $T_{1/2}$ was determined to 90 to 100°C (Figure 3, Figure S1). However, plectasins' onset of unfolding T_{onset} was much lower at approx. 60°C. A clear difference between the wildtype PPI40 and the three variants PPI41, PPI42 and PPI43 could be observed. While $T_{1/2}$ of the wildtype appeared to be almost independent of pH at the tested conditions, all the mutant peptides showed a gradual increase in thermal stability with increasing pH. At pH 3.5, $T_{1/2}$ of all variants was approx. 5°C lower than $T_{1/2}$ of the wildtype, while at higher pH $T_{1/2}$ was similar for all plectasins.

Figure 3: Heat denaturation temperature $T_{1/2}$ of all plectasin variants determined with DSC (■ PPI40, ● PPI41, ▲ PPI42, ▼ PPI43).

Colloidal stability of plectasin variants

To assess the colloidal stability for the plectasin variants, a time dependent heat stress study was performed by incubating samples at 25, 40 and 50°C. Samples were characterized with DLS before incubation (t_0) and after two (t_2) and four weeks (t_4). The DLS measurements, which detect mutual diffusion of the molecules, showed an increase in the apparent hydrodynamic radius R_h^{DLS} at t_0 with increasing pH for all plectasins (Figure 4A). This increase did not correlate with an increase of polydispersity, which was high with high standard deviation (Figure 4B). The increase in R_h^{DLS} indicated an enhanced formation of oligomers or aggregates at higher pH, but the high polydispersity suggested the presence of oligomers already at low pH. Due to the plectasins' rigid protein structure, we did not expect pronounced differences in the protein structure upon a change of condition (Figure 1A-D). Surprisingly, all plectasins showed slightly different apparent radii, which seemed independent of the pH. From the expected structure of the variants, which were build based on the crystal structure of the wildtype and energy-minimized after introducing the mutations³⁰, no significant change in R_h was expected (Table S5)⁴². The R_h^{DLS} of PPI43 was well below 1 nm, which was even lower than the theoretically expected R_h of approx. 1.07 nm. This deviation can be explained by the effect of intermolecular repulsion in different extends for plectasins on their mutual diffusion rates detected by DLS.

When incubated at 25°C, plectasin did not show a significant change in R_h^{DLS} after 2 weeks except for PPI42, which aggregated at pH 5.0 and 5.5 (Figure 5). After 4 weeks incubation, no significant change could be observed. Minor changes in R_h^{DLS} are most likely due to the high polydispersity in the samples. No difference between incubation at 40°C and 25°C could be observed. Incubation at 50°C affected plectasins colloidal stability significantly leading to aggregation in many conditions for four plectasins. After 4 weeks (data not shown), many formulations showed increased aggregation and precipitation and no high-quality data could be obtained. No clear correlation with pH could be observed, but aggregation of plectasins' appeared to occur shortly after start of the incubation at high temperature.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 4: Influence of pH on the colloidal stability of plectasin measured with DLS. A: Hydrodynamic radius R_h^{DLS} . B: %PD. The cumulant fit was used for analysis of the data (see also Figure S21).

Figure 5: Colloidal stability study of plectasin when stored at different temperatures represented in a heat map (abscissa: pH; ordinate: molecule). The samples were measured with DLS and the cumulant fit was used for analysis. Conditions of $R_h^{DLS} > 2$ nm are shown in dark red, conditions at which no dataset could be obtained are shown in black.

Self-interaction propensity of the plectasin variants

The osmotic virial coefficient B_{22} and the diffusion interaction parameter k_D were determined by static light scattering (SLS) and DLS. The negative values of B_{22} for PPI40 and PPI42 indicated overall net attraction and a higher tendency for association (Figure 6A). An increase of pH had no significant effect on B_{22} for PPI40 and PPI42. At pH 5.0 and 5.5 no reliable B_{22} could be obtained for PPI42. PPI41 and PPI43 showed higher, positive B_{22} values within the studied pH range, indicating overall repulsion of the molecules and subsequently a lower tendency for aggregation. Increasing the pH resulted in a decrease in B_{22} . Apart from pH 3.5, PPI43 showed significantly stronger repulsion compared to PPI41.

Similar to the B_{22} results, k_D measured by DLS was negative for both PPI40 and PPI42 (Figure 6B). For PPI40 k_D increased with increasing pH, however, k_D remains close to zero or below. An increase in pH had no effect on the k_D values of PPI42. PPI41 showed positive k_D values for the whole pH range and k_D increased with pH indicating less self-attraction. PPI43 showed a lower k_D compared to PPI41, which was not the case for B_{22} . With increasing pH, the k_D values of PPI43 decreased which is in agreement with the B_{22} trend.

Figure 6: Protein-protein interaction (PPI) of plectasin measured with SLS and DLS. A: Osmotic virial coefficient (B_{22}) determined with SLS. B: Diffusion interaction parameter (k_D) determined with DLS.

Structural investigation and characterization of the plectasin variants using NMR

We used 1D, 2D ^1H as well as 2D ^{13}C NMR to detect structural changes in the plectasin variants upon pH change. Comparison of 2D ^1H - ^{13}C HSQC spectra (Figure S2) revealed that all three mutants are structurally equivalent to the wildtype plectasin, as expected from the model energy-minimized structures. Upon pH changes only the charged residues showed local chemical shift perturbations, presumably due to changes in their ionic state, while the overall structure of all plectasin variants did not change since the fingerprint signal pattern was preserved (Figure S3-S6).

The relative signal intensity in ^1H spectra tended to decrease with higher pH at the same protein concentration for all plectasins (Figure 7A), indicating the formation of aggregated protein clusters with broader signals and

1
2 slower tumbling rates. This effect was most pronounced in PPI42, while PPI40, PPI41 and PPI43 were less
3 affected. PPI42 lost more than a half of its initial signal intensity, indicating significant formation of large, NMR-
4 invisible soluble aggregated clusters at high pH. To further assess the size of NMR-visible protein oligomers,
5 we measured the translational self-diffusion coefficient (D_t) and the transverse relaxation rate (R_2). With the
6 formation of larger NMR-visible oligomers, translational movement is slowed down, reflecting in lower D_t
7 values, while the R_2 values generally increase with the increased oligomer size. PPI40, PPI41 and PPI43
8 showed a gradual increase of R_2 , while R_2 of PPI42 stayed rather constant (Figure 7B). R_2 was highest for
9 PPI41. The D_t of PPI40 was not affected by pH, whereas the D_t values of PPI41, PPI42 and PPI43 decreased
10 with increasing pH (Figure 7C). The hydrodynamic radius R_h^{NMR} derived from D_t varied slightly between the
11 variants, reflecting different aggregates formed and different aggregation propensities, but stayed fairly
12 constant over the pH range, except for PPI42, which showed an increased R_h^{NMR} at pH 5.0 and 5.5 (Figure
13 7D). The R_h^{NMR} values obtained from measuring self-diffusion coefficients were higher than the R_h^{DLS} obtained
14 from measured mutual diffusion, which may be explained by the higher protein concentration in NMR samples,
15 indicating a concentration dependent association of plectasin as well as the effect of inter-molecular repulsions
16 on mutual diffusion coefficients.
17
18
19
20

Figure 7: Effect of pH on plectasin derived from NMR experiments. A: Relative intensity of ^1H plectasin spectra across different pH values compared with the spectral intensity at pH 3.5 as reference. B: Proton transverse relaxation rates (R_2). C: Translational self-diffusion coefficients (D_t) D: Apparent hydrodynamics radius (R_h^{NMR}) calculated from D_t values

Structural investigation and characterization of the plectasin variants using SAXS

Derived parameters from the initial analysis of the SAXS curves are the apparent molecular weight (MW) and the radius of gyration R_g^{SAXS} , which were determined for plectasin at pH 3.5 and pH 5.0 (Figure 8). The apparent MW , determined from the pair distance distribution $p(r)$ function (Figure S9), was approx. 7-8 kDa at a protein concentration $c = 1$ mg/ml for all plectasins at both pH values, which is higher than the molecular weight of the monomer (4.4 kDa). Plectasins' apparent MW increased at higher concentration at pH 3.5 (Figure 8A) and pH 5.0 (Figure 8B), indicating a concentration dependent oligomerization, but at pH 3.5 the curves plateaued for all variants at $c = 5$ mg/ml. At pH 5.0, plectasins' MW increased gradual for all variants but PPI42, which showed an almost exponential increase of apparent MW . A similar trend could be observed for R_g^{SAXS} (Figure 8C,D). While R_g^{SAXS} hardly changed with increasing concentration at pH 3.5, a clear increase could be observed at pH 5.0. The form of the $p(r)$ functions confirmed a higher propensity to form oligomeric species at pH 5.0 than pH 3.5 for all plectasins, but also the presence of some oligomeric species at low pH (Figure S9), explaining the high polydispersity seen in the DLS measurements. PPI41 seemed to form larger species even at pH 3.5, which agrees with higher R_2 observed in the NMR measurements. However, in our initial screen PPI41 showed a high colloidal stability and no visible precipitation over a wide pH range, indicating that the oligomers formed by PPI41 are quite stable. PPI43 showed a significantly lower R_g^{SAXS} at pH 5.0 than the other plectasin variants, which was already observed in DLS measurements.

In all measurements, repulsion within the sample could be observed in the scattering curve at pH 3.5 as the intensity at low q values decreased with higher protein concentration (Figure 9). The most pronounced repulsion was observed for PPI43 (Figure 9D), followed by PPI42 (Figure 9C), then PPI40 (Figure 9A) and PPI41 was the least repulsive (Figure 9B). This order corresponds to the order of R_h^{DLS} observed, indicating that indeed interparticle repulsion in different extends influenced the mutual diffusion of plectasin. As an effect of plectasins' repulsion the R_g^{SAXS} is most likely underestimated. Repulsion could be also observed at pH 5.0, but oligomerization and aggregation could be observed as well (Figure S8). Especially PPI42 showed aggregation at higher protein concentration.

Figure 8: Investigation of change of pH on plectasin with SAXS. A: Apparent MW at pH 3.5. B: Apparent MW at pH 5.0. C: R_g^{SAXS} at pH 3.5. D: R_g^{SAXS} at pH 5.0.

Figure 9: SAXS curves (concentration normalized) of plectasin at pH 3.5. A: PPI40. B: PPI41. C: PPI42. D: PPI43. The part of the SAXS curve at low $q \text{ (nm}^{-1}\text{)}$ showing interparticle interaction is plotted. Repulsion is indicated by a black arrow.

Separation of components in plectasin samples by AUC

According to analytical ultracentrifugation (AUC) performed at protein concentration of 0.45 mg/ml the major species ($\sim 95\%$) for all plectasins at low and high pH has a sedimentation coefficient of approx. 1S (Figure 10). This corresponds to the sedimentation coefficient of the plectasin monomer (Table S5). The derived MW and R_h^{AUC} was slightly higher at pH 5.0 than at pH 3.5 (Figure 10E). Larger components in the samples were only present at a low amount below 5 % and differed slightly between the plectasin variants. A small percentage of different oligomeric species could be observed (Figure 10A-D, inset), which seemed to be slightly more prominent at pH 3.5. However, very large aggregates would not be visible within this measurement as they sediment to the cell bottom before the measurement starts.

The ratio R_g/R_h is an indicator of the particles shape with a value of ~ 0.775 for globular protein and higher values for elongated molecules. Therefore, the R_g^{SAXS}/R_h^{AUC} ratio was determined for all variants at pH 3.5 and pH 5.5 (Figure 10F,G, Table S3). From theoretical values of ratio of R_g and R_h from the minimized

computational models, we expect a R_g/R_h ratio of approx. 0.85 for all plectasins, if monomeric (Table S5). At pH 3.5, PPI40, PPI42 and PPI43 agreed with the expected R_g/R_h (Figure 10F) and seemed to be independent of the protein concentration, except from PPI40 showing a slightly increasing ratio. PPI41's R_g^{SAXS}/R_h^{AUC} ratio of approx. 0.95 was independent of the protein concentration. At pH 5.0, the R_g^{SAXS}/R_h^{AUC} ratio of the variants differed more than at pH 3.5 and the R_g^{SAXS}/R_h^{AUC} ratio increased significantly with concentration for all plectasins (Figure 10G), indicating the formation of more elongated particles with increasing concentration. The effect was most pronounced for PPI42 (see also Table S3). The difference in repulsion in the system might, however, have an influence on the R_g^{SAXS}/R_h^{AUC} ratio.

E

Protein	pH	sw	sw(20,w)	percentage	MW (Da)	R_h^{AUC} (nm)
PPI40	3.5	1.01	0.90	96.80	4270	1.26
PPI40	5	1.03	0.92	98.48	4465	1.29
PPI41	3.5	0.93	0.83	94.00	3841	1.23
PPI41	5	1.04	0.93	96.61	4492	1.29
PPI42	3.5	0.92	0.82	91.35	3899	1.23
PPI42	5	1.00	0.89	96.34	4216	1.24
PPI43	3.5	0.93	0.83	97.75	3952	1.24
PPI43	5	0.93	0.83	96.43	4398	1.39

Figure 10: Investigation of plectasin in solution with AUC. A-D: Continuous distribution of plectasin at pH 3.5 (solid lines) and pH 5.0 (dashed lines). Inserts show the peaks of oligomeric fractions. The theoretical sedimentation coefficient of the respective monomeric species, calculated with HullRad⁴², is indicated with the dashed grey line. E: Properties of the species corresponding to the main peak are shown. F,G: R_g^{SAXS}/R_h^{AUC} ratio at pH 3.5 and pH 5.0.

MD simulations show differences in flexibility for plectasin variants

MD simulations were performed in duplicates to understand structural dynamics in the pH range of 3.5 to 5.5. During the 100 ns simulation, the secondary structural elements stayed intact. This was expected due to the three disulfide bonds, which contribute distinctly to the structure of plectasin³⁴. Only little variation of the secondary structure of the plectasin mutated peptides, could be observed, which changed slightly upon change of pH (Figure 1, Figure S14, Figure S16). Based on hierarchical clustering of the MD simulation trajectories, the most representative structure from each simulation was taken for further analysis. We analyzed the aggregation prone areas of plectasin using CamSol v2.2²². The overall solubility score of all plectasin variants showed a positive value, indicating good solubility. However, we observe a decreasing trend with increasing pH (Figure S13A). The solubility score of the individual amino acids identified the N- and C-terminus as the least soluble areas of plectasin (Figure S13B-E). Within the amino acid stretch 16-26, most amino acids showed a decrease in solubility upon increase of pH. The structural properties upon change of pH were further analyzed. We observed high differences in amino acid positions among the representative structures for bulky amino acids on the proteins surface, indicating flexibility, for example W8, F35, K23 and K38. The most representative structures of PPI41 across the investigated pH range appeared to be more uniform than PPI40, PPI42 and PPI43 (Figure 11, Figure S16). PPI42 showed large variations across the investigated pH range in the position of the protein's backbone in the N-terminal loop region, especially at P7,

W8 and E10 (Figure 11, Figure S15). The backbone of PPI40 and PPI43 showed small variations among the representative structures, but only around W8 (Figure 11D). In the C-terminal region, Y29 and Y40 are close in space and show different degrees of variation in the variants. In PPI41, this variation appeared to be less (Figure 12), indicating less flexibility. The mutated amino acid R26 (from K26) builds a salt bridge with the C-terminus (Figure 12B). N5S and D9A in PPI41 are located in the N-terminal loop and are close in space in the protein. L36 (changed from V36) in PPI42 takes up more space. M13L in PPI43 was reported to lead to a shorter α -helix⁴³.

Figure 11: Representative structure of the N-terminal loop of all four plectasin variants at pH 3.5 - 5.5 from MD simulations (low pH:light blue; high pH:dark blue); A: PPI40. B: PPI41 C: PPI42D: PPI43. The loop is shown as sticks and backbone structure only for all plectasins, respectively.

Figure 12: C-terminal region of PPI40 and PPI41 at pH 3.5 - 5.5 (low pH:light blue; high pH:dark blue).A: PPI40. B: PPI41. Y29 and Y40 are shown in brown, K26 in PPI40 is shown in pink and R26 in PPI41 is shown in cyan. The salt bridge between Y40 and R26 in PPI41 is indicated by the dashed line and the distance is shown.

DISCUSSION

Rationale behind the choice of methods and screening setup

The approach of combining biophysical and structural characterization in as an approach to study pharmaceutics has been applied in recent studies^{44–46}. However, in most studies the proteins of interest have been characterized using only one structural method. Protein aggregation is still a major issue in prote4in drug development⁴⁷, but the research focus on peptide aggregation has mainly been focused on the disease linked formation amyloid fibrils^{48,49}. The scope of this study was therefore to develop a new approach of peptide drug characterization that can be applied and adapted to complex and aggregating systems. To assess plectasin stability intrinsic fluorescence was initially attempted. However, the protein's only tryptophan W8 is located in a loop and already solvent exposed in the native state. No change in the fluorescence between the folded and unfolded state could be observed by nano differential scanning fluorimetry (nanoDSF) and isothermal chemical denaturation (ICD) (data not shown). As an alternative to high throughput nanoDSF, DSC was chosen to determine the thermal stability of plectasin. Plectasin showed a thermally induced unfolding of approx. 100°C, which implies, that unfolding should not be an issue at relevant temperatures. However, the onset of unfolding detected with DSC was much lower (approx. 60°C) and temperature induced aggregation can therefore not be excluded due to a high $T_{1/2}$, as we have proven with the accelerated stress test. Subsequently, size exclusion chromatography (SEC) was used to study plectasin behavior as a function of pH. However, interaction of plectasin with the column matrix could be observed. The usage of acetonitrile or arginine disrupted this interaction, but both approaches will also interfere the protein-protein interactions to be studied. Furthermore, previous results have shown binding of arginine to plectasin⁵⁰. Plectasin showed a very low solubility (< 1 mg/ml) at neutral and high pH¹⁷. The addition of salt did not lead to significant

improvement of plectasin colloidal stability. Therefore, a pH range of pH 3.5 to 5.5 in acetate buffer was chosen for the in-dept investigation of plectasin (step 2) in this case study. However, the principle of our screening approach is easy to adapt to other proteins and different techniques or different screening conditions in terms of the buffer choice or adding excipients.

Plectasin shows pH dependent aggregation that differ among the variants

DLS gave us a first indication that the formation oligomers and aggregates of plectasin is influenced by pH, as R_h^{DLS} increased at higher pH. However, light scattering is very sensitive to the presence of a few large aggregates, leading to high polydispersity, poor reproducibility and high error, making it difficult to distinguish between different types of oligomers or aggregates formed. The polydispersity of all plectasin samples was rather high due to the presence of oligomers or aggregates. An increase of R_h^{DLS} could be observed upon increasing pH, most likely due to a shift of the equilibrium between monomer and oligomers towards oligomerization and aggregation. This was confirmed by the diffusion interaction parameter (k_D) and the osmotic virial coefficient (B_{22}). All plectasin variants showed high affinity for self-attraction, becoming more pronounced at higher pH. Self-attraction was stronger for PPI40 and PPI42 as compared to PPI41 and PPI43. These results agree with our findings in NMR, which indicated that aggregated protein clusters with broader signals and slower tumbling rates were formed for all variants with pH increase. This was most pronounced in PPI42, while PPI40, PPI41 and PPI43 were less affected. D_t value of PPI42 is much lower than $1.74 \cdot 10^{-10}$ m²/s predicted by HYDROPRO⁵¹, indicating that in NMR conditions PPI42 is oligomerizing in solution, forming NMR-visible oligomers. In addition, significant decrease in PPI42 signal intensity at higher pH indicate the formation of large soluble NMR-invisible clusters. The D_t and R_2 values determined by NMR did not change significantly with pH for PPI42. For PPI40, PPI41 and PPI43 we observed an increase in R_2 , which is in agreement with the gradual increase of R_h^{DLS} and in line with gradual increase of molecular size of soluble oligomers.

SAXS measurements demonstrated an apparent MW larger than that of the monomer at low and high pH, even at low protein concentration, indicating the presence of oligomers and aggregates at both pH. Due to the observed repulsion, the apparent MW is most likely underestimated. The MW increase with protein concentration indicated increasing presence of larger oligomeric species. This increase of MW was much more pronounced at pH 5.0 than at pH 3.5, which was also indicated by DLS and NMR. AUC was therefore performed to see what sizes of oligomers or aggregates were present in solution. It could be shown that the main species has a sedimentation coefficient of approx. 1S, which corresponds to the monomer (Figure 10, Table S4, Table S5). All plectasins showed small fractions of larger species. The peaks of small oligomeric species seemed to be slightly more prominent at pH 3.5 As plectasin is a peptide we needed to centrifuge with high centrifugal force (50.000g). Therefore, large aggregates formed upon increase of pH to pH 5.0, might have already reached to bottom of the cell before the measurement started and could therefore be not detected.

R_h is defined as an equivalent hard sphere diffusing with the same speed as the measured molecule. R_g is defined as the mass weighted average size from the molecules center. For an elongated molecule, R_g becomes larger resulting in a larger R_g/R_h ratio. This effect could be observed with increasing concentration (Figure 10). This increase in the R_g^{SAXS}/R_h^{AUC} ratio (Table S3) is at least twice as high at pH 5.0 compared to 3.5 for all plectasin variants, indicating a higher propensity to form non-spherical particles like oligomers or aggregates. Most likely the effect of concentration and pH on the R_g/R_h ratio is underestimated as we observed repulsive effects between the species in SAXS. This gives us another indication that the formation of oligomers and aggregates is concentration dependent and more pronounced at pH 5.0 than at pH 3.5. In addition to the

differences in aggregation propensity and aggregates formed, we observed an increase of $T_{1/2}$ for PPI41, PPI42 and PPI43, but not for the plectasin wildtype with DSC, which might be due to the higher pI in the plectasin variants compared to the wildtype.

All techniques indicated an increasing propensity of plectasin to aggregate at high pH, which differed from variant to variant. The formation of large protein clusters seemed to be most prominent in PPI42, as seen in NMR and SAXS. In DLS, SAXS and AUC we could detect small fractions of different oligomers also at pH 3.5 for all plectasins, which seemed to be smaller in size than at pH 5.0. However, as separation techniques apart from AUC turned out to be complicated to apply to plectasin, a characterization of the different oligomers formed was not pursued further in this study.

Advantages and limitations of different biophysical methods used in this study

Using a portfolio of different complementary techniques allowed us to draw a holistic picture and understand a protein interaction behavior as complex as the one of plectasin better. It is very important to question experimental results and keep possible artefacts, advantages and limitations of the techniques used for analysis in mind. In our study, we derived biophysical parameters like R_h and R_g from different techniques (Table 1). R_h^{DLS} , derived from mutual diffusion coefficients, differed between the variants and was unrealistically small, especially for PPI43, which may be explained by the presence of molecular repulsion in the system, which makes molecules repel each other and hence appear to diffuse faster relative to each other. This repulsion could be also seen from the shape of the SAXS curves, especially at pH 3.5. On the other hand, R_h^{NMR} was calculated using experimentally measured translational self-diffusion coefficients, which are not affected by molecular repulsion. The differences in R_h^{NMR} between the plectasin variants derive most likely from different quantities of NMR-visible oligomers. AUC measures R_h directly on particles separated based on their size and is also affected by protein repulsion, and hence often is considered a gold standard in protein size determination. Therefore, R_h^{AUC} was used for the calculation of the R_g/R_h value. AUC showed that more than 95% of all plectasins is present as monomer. However, protein concentration used for AUC was lower than that used in all other techniques, which may have shifted equilibrium towards dissociated state; besides, the presence of larger rapidly sedimenting aggregates can be missed, which might explain why we did not see significant differences between the variants in AUC. Additionally it might explain why we could only observe a small quantities of oligomeric species at pH 3.5, but not at pH 5.0.

Protein-protein interaction could be derived from the k_D analysis of the DLS measurements, the B_{22} analysis of the SLS measurements and from the shape of the SAXS curves (Table 1). The mostly repulsive behavior of PPI41 and PPI43 seen in the DLS and SLS measurements, represents the sum of all protein-protein interactions in the sample and agreed with their higher colloidal stability at neutral pH seen in DLS in the initial screen (step 1). However, it has been shown that multicomponent complexes and their protein diffusion can be influenced by factors unrelated to protein-protein interaction⁵². Furthermore, k_D has an offset as compared to B_{22} , meaning a slightly negative k_D can indicate slight net repulsion⁵³. In our SAXS measurements we observed repulsive behavior in different extents for all plectasins. At pH 5.0, in addition to the repulsion, attraction and oligomerization could be observed, which might be explained by the formation of oligomers which still showed repulsive behavior.

For screening, high throughput and little sample consumption are important factors. In our study DLS had the highest throughput and the least sample consumption, considering, that it can be measured in a 1526 well plate with only 10 μ l sample in each well. It is also very sensitive towards the presence of aggregates. This makes it a useful method to detect these aggregates, but in our study, it led to difficulties in reproducibility.

1
2 Nevertheless, in combination with other biophysical characterization, DLS is a valuable method for in-dept
3 screening. The broad applicability of DLS to different conditions and protein sizes makes it a good tool for the
4 first evaluation of a large number of different samples.
5 Due to the above mentioned difficulties using SEC, AUC was the only technique, which involved separation
6 of the particles in solution in this study. AUC can give valuable information about particles present in solution
7 but comes with a low throughput.
8 SAXS and NMR give structural information about the molecules and species present. NMR is able to provide
9 high-resolution insights into the system and is very sensitive towards changes in structure and environment.
10 Additionally, NMR can give valuable information about kinetics of a system. SAXS is a low-resolution
11 technique, but the throughput is most times higher and the sample consumption is lower than with NMR,
12 making it more feasible for screening. In case of monodisperse samples or in-line separation with SEC (SEC-
13 SAXS), modeling can be applied to SAXS data, which makes it a valuable technique for biologics despite of
14 lower resolution.
15 For a broad applicability to different biological systems, the size limitations of the different methods play an
16 important role. NMR applications may be limited for proteins of larger size. SAXS can be a very powerful when
17 applied to multi-domain protein systems and their subunits when they are monodisperse in solution. However,
18 polydispersity complicates the SAXS data analysis and interpretation as seen in this study. In DLS a wide
19 range of different protein sizes, can be measured, but the amount of detailed information of the protein is
20 limited and will most likely not be the only method applied when characterizing a protein system. AUC has
21 generally no limitations in terms of protein size when the sedimentation force applied through spinning is
22 carefully chosen, but there may be upper limits on protein concentration used, which may distort results for
23 systems with concentration-dependent oligomerization. The combination of experimental and computational
24 results is powerful to explain observed experimental results on a molecular level. Computational methods like
25 MD can be in principle applied to all different protein sizes, however, the bigger the protein, the longer the
26 computational time and the more expensive the simulations become.

34 Table 1: Summary of derived radii and observed protein-protein interactions (PPI) in this study.

	pH	R_h^{DLS}	R_h^{NMR}	R_h^{AUC}	R_g^{SAXS}	PPI ^{DLS}		PPI ^{SLS}		PPI ^{SAXS}	
		(nm)	(nm)	(nm)	(nm)	attraction	repulsion	attraction	repulsion	attraction	repulsion
PPI40	3.5	1.07	1.25	1.26	1.08	++		+			+++
	5.0	1.46	1.25	1.29	1.12			+		+	++
PPI41	3.5	1.16	1.50	1.23	1.13		+++		+++		++
	5.0	1.5	1.63	1.29	1.26		+++		++	++	+
PPI42	3.5	1.01	2.25	1.23	1.05	+		+			+++
	5.0	1.37	2.75	1.24	1.21	+		+		++	+
PPI43	3.5	0.83	2.10	1.24	0.98		+++		+++		++++
	5.0	1.22	1.90	1.39	1.01	+			+++	+	++

48
49 **MD simulations point towards flexibility as a cause for plectasins' aggregation**

50
51 Molecular dynamics simulations provide a good complementary tool to explain effects seen experimentally on
52 a molecular level. With MD simulation, we could observe that the point mutations did not affect the overall
53 plectasin structural fold, which could be confirmed by NMR. Previously, studies have shown that flexibility and

conformational changes can contribute to fibril formation and aggregation of peptides^{54,55}. In light of these studies, MD simulations can be used as a screening approach for peptides to access their colloidal stability. The overall solubility score of plectasin showed a slightly decreasing trend with increasing pH, which agrees with our experimental findings, that plectasin have a higher aggregation propensity with increasing pH. The analysis of the individual solubility score revealed that especially the N- and C-terminus are aggregation prone. Many small patches in plectasin become lower in solubility with increasing pH (Figure S13). Especially the amino acid stretch 16-26 included many amino acids with a decrease of solubility upon increasing pH. However, this does not explain the differences in the aggregation behavior between the different plectasin variants. Therefore, we analyzed structural changes of the representative structures from the MD simulations upon changes in pH, indicating flexibility and different possible conformations. These local motions that arises due to flexibility in the loop regions could be captured well within the 100ns time frame, since loop motions occur in the ps-ns range⁵⁶. All variants showed an increased fluctuation within the amino acid stretch 6-13, which represents the N-terminal loop of plectasin (Figure S22). Among the most representative structures derived from the MD simulations, the N-terminal loop region in plectasin showed large variation among the variants. The change in amino acid at position 9 in all variants seemed to have an impact in the loop stability of plectasin. We could observe large deviation in the position of the protein's backbone within the investigated pH range in the area of E10 in PPI40 and PPI42 at high pH, which correlated with a higher propensity of these variants to aggregate. In PPI41, the introduced S5 and A9 mutations were close in space which might influence the N-terminal loop stability positively through interaction, as PPI41 seemed to be generally less flexible in that region. In contrast, PPI42 showed large backbone variations in this region, especially at W8 and E10, which might be an explanation for the high propensity of PPI42 to aggregate upon increasing pH. The mutation sites at position 9 or 14 might disrupt the backbone and lead to alternative conformations obtained by PPI42. In the region of Y29 and Y40, we observed large fluctuations in all variants but PPI41. Y29 and Y40 generally seemed to influence each other through π stacking. In PPI41, the C-terminus at Y40 builds a salt bridge to the mutated amino acid R26, leading to lower flexibility of the Tyrosine's. This might cause PPI41's high colloidal stability over a wide pH range. From comparing the experimental results with our findings in MD, we conclude that an increase in flexibility within the protein is strongly linked to oligomerization and aggregation in plectasin. Generally, equilibria and the types of oligomers forms seems to be influenced by the introduced mutations.

CONCLUSION

Plectasin represents a complicated system with respect to protein-protein interaction, oligomerization and aggregation in solution. Despite similarities in the solution structures of the wildtype and three variants investigated in this study, all four plectasins show markedly different solution characteristics. As such, plectasin turned out to be a challenge in terms of high throughput screening normally used for protein drug development. Employing DLS, we observed a pH dependent oligomerization or aggregation which differed between variants. DSC showed an increase in thermal unfolding temperature with an increase of pH for all plectasin variants. An increase of the temperature of unfolding is a widely used metric for sample stability assessment¹⁵. However, an in-depth characterization is challenging with high throughput methods only. Therefore, we combined the screening of plectasin with lower throughput, more labor-intensive methods like NMR, AUC and SAXS. These methods give much more details, but can only be applied on selected conditions, which were chosen based on the initial screening results. Additionally, based on our experimental results we conducted MD simulations to identify aggregation prone areas and analyze the impact of the mutations on the

1 proteins molecular structure. With this combination of methods, we were able to get a more detailed picture
2 of the behavior of proteins in solution. The introduced mutations do not significantly affect the monomeric
3 protein structure but have a clear impact on the proteins' tendency to oligomerize and aggregate. The
4 plectasin variants showed different pH dependent behavior forming oligomers at higher pH to a different extent.
5 These oligomerizations appeared to be concentration dependent.
6

7
8 In summary, our strategy of combining high throughput screening with structural characterization is generally
9 applicable and adjustable to other protein systems. The combination of biophysical, structural and
10 computational methods represents a new approach of pharmaceutical screening, which is especially relevant
11 for complex systems such as peptides. All methods have different advantages and limitations, which must be
12 taken into account for every system individually. The number of conditions and investigated proteins is flexible
13 and techniques can be replaced or added, depending on the applicability to the system being studied.
14 Furthermore, screening of excipients for stabilization is generally possible, as long as the limitations of certain
15 techniques are taken into account. We thus suggest a wide adoption of the strategies outlined for the
16 characterization of protein therapeutics.
17
18
19
20
21

22 **ASSOCIATED CONTENT**

23
24 Supporting information

25
26 SI₁: Structural parameters derived from SAXS; SI₂: Parameters for SAXS data collection; SI₃: R_g/R_h ratio as a
27 function of protein concentration; SI₄: Sample components measured with AUC; SI₅: Theoretical biophysical
28 parameters of plectasin; SI₆: Temperature induced unfolding curves (DSC); SI₇: ^1H - ^{13}C HSQC of all plectasin
29 variants; SI₈: ^1H - ^{13}C HSQC of PPI40 as a function of pH; SI₉: ^1H - ^{13}C HSQC of PPI41 as a function of pH; SI₁₀:
30 ^1H - ^{13}C HSQC of PPI42 as a function of pH; SI₁₁: ^1H - ^{13}C HSQC of PPI43 as a function of pH; SI₁₂: SAXS curves
31 of plectasin at pH 3.5; SI₁₃: SAXS curves of plectasin at pH 5.0 (low q area); SI₁₄: $p(r)$ functions of plectasin
32 at pH 3.5 and pH 5.0; SI₁₅: D_{max} of plectasin at pH 3.5 and pH 5.0; SI₁₆: Kratky plots of plectasin at pH 3.5 and
33 pH 5.0; SI₁₇: SAXS curves of plectasin at pH 5.0; SI₁₈: Solubility score of plectasin; SI₁₉: Overlay of simulated
34 plectasin structures at pH 3.5 and pH 5.0; SI₂₀: Structure of the N-terminal loop of plectasin as a function of
35 pH; SI₂₁: Structure of the backbone of plectasin as a function of pH; SI₂₂: Solubility score mapped on the
36 surface of PPI40; SI₂₃: Solubility score mapped on the surface of PPI41; SI₂₄: Solubility score mapped on the
37 surface of PPI42; SI₂₅: Solubility score mapped on the surface of PPI43; SI₂₆: Autocorrelation function and
38 cumulant and regularization fit of plectasin at pH 3.5 and pH 5.0. SI₂₇: RMSF plots of plectasin at pH 3.5 and
39 pH 5.5.
40
41
42
43
44
45
46
47
48

49 **AUTHOR INFORMATION**

50
51 Corresponding Authors

52
53 *email: alnq@novozymes.com; ph@kemi.dtu.dk

54
55 ORCID ID:

Christin Pohl: 0000-0001-6193-4016

Matja Zalar: 0000-0002-4807-0177

Inas ElBialy: 0000-0002-5468-9363

Sowmya Indrakumar: 0000-0002-5144-3261

Günther H.J. Peters: 0000-0001-9754-2663

Wolfgang Friess: 0000-0002-8807-309X

Alexander P. Golovanov: 0000-0002-8592-3984

Werner Streicher: 0000-0002-2425-5839

Allan Nørgaard: 0000-0003-1850-9879

Pernille Harris: 0000-0002-6806-4903

Authors Contribution

Christin Pohl, Allan Nørgaard and Pernille Harris wrote the manuscript. Pernille Harris, Allan Nørgaard, Alexander P. Golovanov, Wolfgang Friess and Günther H.J. Peters planned and supervised the study. Christin Pohl conducted, analyzed and evaluated the preliminary DLS and DSC screening studies (step 1). Christin Pohl and Matja Zalar evaluated preliminary data and designed screening conditions of step 2 in this study. Christin Pohl conducted analyzed and evaluated DSC, DLS, AUC and SAXS studies. Matja Zalar conducted, analyzed and evaluated NMR studies. Inas ElBialy conducted, analyzed and evaluated DLS and SLS intermolecular interaction studies. Sowmya Indrakumar conducted, analyzed and evaluated MD simulation studies. All authors corrected and approved the final manuscript. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGEMENTS

The synchrotron SAXS data was collected at beamline P12 operated by EMBL Hamburg at the Petra II storage ring (DESY, Hamburg, Germany), and at the BioSAXS beamline, BM29, at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. For local support we thank Melissa Gräwert, Stefano Da Vela, Andrey Gruzinov, Karen Manalastas (all: EMBL, Hamburg) and Martha Brennich (ESRF, Grenoble)

This work was funded by European Union's Horizon 2020 research and innovation program (grant agreement no. 675074).

We thank the Danish Agency for Science, Technology, and Innovation for funding the instrument center DanScatt.

Simulations were performed at the high-performance computing (HPC) services at DTU and in-house CPU/GPU cluster facilities at DTU chemistry.

We thank Alina Kulakova for her assistance in the SAXS measurements.

The first author would like to thank the whole PIPPI consortium for continuous collaboration and constructive discussions.

ABBREVIATIONS

AUC: Analytical ultracentrifugation

B_{22} : Osmotic virial coefficient

°C: degree celsius

DLS: dynamic light scattering

DSC: differential scanning calorimetry

ICD: isothermal chemical denaturation

k_D : Diffusion interaction parameter

kDa: kilo dalton

MD: Molecular dynamics

MW : Molecular weight

nanoDSF: Nano differential scanning fluorimetry

NMR: Nuclear magnetic resonance

PPI40: Plectasin wildtype

PPI41: 5S 9A 26R variant of plectasin

PPI42: 9S 14K 36L variant of plectasin

PPI43: 9N 13L 14R variant of plectasin

$p(r)$: distance distribution (function)

R_g : Radius of gyration

R_h : Hydrodynamic radius

D_{max} : maximum particle dimension

Rmsd Root mean square deviation

SAXS: Small-angle X-ray scattering

REFERENCES

1. Walsh, G. Biopharmaceutical benchmarks 2018. *Nat. Publ. Gr.* **36**, 1136–1145 (2018).
2. Mullard, A. 2018 FDA drug approvals. *Nat. Rev. Drug Discov.* **18**, 85–89 (2019).
3. Sato, A. K., Viswanathan, M., Kent, R. B. & Wood, C. R. Therapeutic peptides: technological advances driving peptides into development. *Curr. Opin. Biotechnol.* **17**, 638–642 (2006).
4. Henninot, A., Collins, J. C. & Nuss, J. M. The Current State of Peptide Drug Discovery: Back to the Future? *J. Med. Chem.* **61**, 1382–1414 (2018).
5. Wang, W. & Kelner, D. N. Correlation of rFVIII Inactivation with Aggregation in Solution. **20**, (2003).
6. Moussa, E. M. *et al.* Immunogenicity of Therapeutic Protein Aggregates. *J. Pharm. Sci.* **105**, 417–430 (2016).
7. Sethu, S., Govindappa, K., Park, K. & Sathish, J. Immunogenicity to Biologics : Mechanisms , Prediction and Reduction. 331–344 (2012). doi:10.1007/s00005-012-0189-7
8. Ratanji, K. D. *et al.* Immunogenicity of therapeutic proteins : Influence of aggregation Immunogenicity of therapeutic proteins : Influence of aggregation. **6901**, (2014).
9. Carpenter, J. F. *et al.* Overlooking subvisible particles in therapeutic protein products: Gaps that may compromise product quality. *J. Pharm. Sci.* **98**, 1201–1205 (2009).
10. Singh, S. K. *et al.* An industry perspective on the monitoring of subvisible particles as a quality attribute for protein therapeutics. *J. Pharm. Sci.* **99**, 3302–3321 (2010).
11. Mahler, H. C., Friess, W., Grauschopf, U. & Kiese, S. Protein aggregation: Pathways, induction factors and analysis. *J. Pharm. Sci.* **98**, 2909–2934 (2009).
12. Nahri, L. O., Schmit, J., Bechtold-Peters, K. & Sharma, D. Classification of Protein Aggregates. *J. Pharm. Sci.* **101**, 493–498 (2011).
13. Xiong, Y. L. Influence of pH and Ionic Environment on Thermal Aggregation of Whey Proteins. *J. Agric. Food Chem.* **40**, 380–384 (1992).
14. Pfefferkorn, C. M., McGlinchey, R. P. & Lee, J. C. Effects of pH on aggregation kinetics of the repeat domain of a functional amyloid, Pmel17. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21447–21452 (2010).
15. Jarasch, A. *et al.* Developability assessment during the selection of novel therapeutic antibodies. *J. Pharm. Sci.* **104**, 1885–1898 (2015).
16. Bommarius, A. S., Broering, J. M., Chaparro-Riggers, J. F. & Polizzi, K. M. High-throughput screening for enhanced protein stability. *Curr. Opin. Biotechnol.* **17**, 606–610 (2006).
17. Mygind, P. H. *et al.* Plectasin is a peptide antibiotic with therapeutic potential from a saprophytic fungus. *Nature* **437**, 975–980 (2005).

18. Schneider, T. *et al.* Plectasin, a Fungal Defensin, Targets the Bacterial Cell Wall Precursor Lipid II. *Science* (80-.). **328**, 1168–1172 (2010).
19. Mandal, K. *et al.* Racemic crystallography of synthetic protein enantiomers used to determine the X-ray structure of plectasin by direct methods. *Protein Sci.* **18**, 1146–1154 (2009).
20. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
21. Mandal, K. *et al.* Racemic crystallography of synthetic protein enantiomers used to determine the X-ray structure of plectasin by direct methods. *Protein Sci.* **18**, 1146–1154 (2009).
22. Sormanni, P., Aprile, F. A. & Vendruscolo, M. The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility. *J. Mol. Biol.* **427**, 478–490 (2015).
23. Schrödinger, LLC. *The {PyMOL} Molecular Graphics System, Version~1.8.* (2015).
24. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. *John M. Walk. Proteomics Protoc. Handb.* 571–607 (2005).
25. Peri, S., Steen, H. & Pandey, A. GPMAW--a software tool for analyzing proteins and peptides. *Trends Biochem. Sci.* **26**, 687–689 (2001).
26. Wang, W., Nema, S. & Teagarden, D. Protein aggregation-Pathways and influencing factors. *Int. J. Pharm.* **390**, 89–99 (2010).
27. Franke, D. *et al.* ATSAS 2.8 : a comprehensive data analysis suite for small-angle scattering from macromolecular solutions . *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
28. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
29. Schuck, P., Perugini, M. A., Gonzales, N. R., Hewlett, G. J. & Schubert, D. Size-distribution analysis of proteins by analytical ultracentrifugation: Strategies and application to model systems. *Biophys. J.* **82**, 1096–1111 (2002).
30. Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided. Mol. Des.* (2013). doi:10.1007/s10822-013-9644-8
31. Gordon, J. C. *et al.* H++: A server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* (2005). doi:10.1093/nar/gki464
32. Case, D. A. *et al.* The Amber biomolecular simulation programs. *Journal of Computational Chemistry* (2005). doi:10.1002/jcc.20290
33. Mongan, J., Case, D. A. & McCammon, J. A. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* (2004). doi:10.1002/jcc.20139
34. Indrakumar, S. *et al.* Conformational Stability Study of a Therapeutic Peptide Plectasin Using Molecular Dynamics Simulations in Combination with NMR. *J. Phys. Chem. B* acs.jpcc.9b02370 (2019). doi:10.1021/acs.jpcc.9b02370
35. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma.* (2010). doi:10.1002/prot.22711

- 1
2 36. Mahoney, M. W. & Jorgensen, W. L. A five-site model for liquid water and the reproduction of the
3 density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* (2000).
4 doi:10.1063/1.481505
5
- 6 37. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of
7 motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* (1977).
8 doi:10.1016/0021-9991(77)90098-5
9
- 10 38. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of
11 molecular dynamics trajectory data. *J. Chem. Theory Comput.* (2013). doi:10.1021/ct400341p
12
- 13 39. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* (1996).
14 doi:10.1016/0263-7855(96)00018-5
15
- 16 40. Tan, P.-N., Steinbach, M. & Kumar, V. Chapter 8: Cluster Analysis: Basic Concepts and Algorithms.
17 in *Introduction to Data Mining* 487–568 (2005). doi:10.1016/0022-4405(81)90007-8
18
- 19 41. Wolf, A. & Kirschner, K. N. Principal component and clustering analysis on molecular dynamics data
20 of the ribosomal L11·23S subdomain. *J. Mol. Model.* (2013). doi:10.1007/s00894-012-1563-4
21
- 22 42. Fleming, P. J. & Fleming, K. G. HullRad: Fast Calculations of Folded and Disordered Protein and
23 Nucleic Acid Hydrodynamic Properties. *Biophys. J.* **114**, 856–869 (2018).
24
- 25 43. Zhang, Y. *et al.* In vitro and in vivo characterization of a new recombinant antimicrobial peptide,
26 MP1102, against methicillin-resistant *Staphylococcus aureus*. *Appl. Microbiol. Biotechnol.* **99**, 6255–
27 6266 (2015).
28
- 29 44. Fukuda, M., Moriyama, C., Yamazaki, T., Imaeda, Y. & Koga, A. Quantitative Correlation between
30 Viscosity of Concentrated MAb Solutions and Particle Size Parameters Obtained from Small-Angle X-
31 ray Scattering. *Pharm. Res.* **32**, 3803–3812 (2015).
32
- 33 45. Liu, J., Yadav, S., Andya, J., Demeule, B. & Shire, S. J. Analytical Ultracentrifugation and Its Role in
34 Development and Research of Therapeutic Proteins. *Methods Enzymol.* **562**, 441–476 (2015).
35
- 36 46. Wishart, D. S. Characterization of biopharmaceuticals by NMR spectroscopy. *TrAC - Trends Anal.*
37 *Chem.* **48**, 96–111 (2013).
38
- 39 47. Wang, W. Advanced protein formulations. *Protein Sci.* **24**, 1031–1039 (2015).
40
- 41 48. Knowles, T. P. J., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein
42 misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15**, 384–396 (2014).
43
- 44 49. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188–1203
45 (2012).
46
- 47 50. Indrakumar, S. Characterization of protein-exciipient interactions and protein-protein interactions on
48 the molecular level using in-silico simulation methods. (Technical University of Denmark, 2019).
49
- 50 51. Ortega, A., Amorós, D. & García De La Torre, J. Prediction of hydrodynamic and other solution
51 properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* **101**, 892–898 (2011).
52
- 53 52. Sorret, L. L., DeWinter, M. A., Schwartz, D. K. & Randolph, T. W. Challenges in Predicting Protein-
54 Protein Interactions from Measurements of Molecular Diffusivity. *Biophys. J.* **111**, 1831–1842 (2016).
55
- 56 53. Menzen, T. I. M. & Friess, W. Temperature-Ramped Studies on the Aggregation , Unfolding , and
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Interaction of a Therapeutic Monoclonal Antibody. *Pharm. Biotechnol.* 445–455 (2013).
doi:10.1002/jps.23827

54. Valerio, M. *et al.* Early events in protein aggregation: Molecular flexibility and hydrophobicity/charge interaction in amyloid peptides as studied by molecular dynamics simulations. *Proteins Struct. Funct. Bioinforma.* **58**, 110–118 (2004).

55. Zapadka, K. L., Becher, F. J., Gomes dos Santos, A. L. & Jackson, S. E. Factors affecting the physical stability (aggregation) of peptide therapeutics. *Interface Focus* **7**, 20170030 (2017).

56. Kmiecik, S. *et al.* Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **116**, 7898–7936 (2016).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23 point mutation dependent oligomerization and aggregation