New Disulphide Bond in Cystatin-Based Protein Scaffold Prevents Domain-Swap-Mediated Oligomerization and Stabilizes the Functionally Active Form

Matja Zalar* and Alexander P. Golovanov*

Manchester Institute of Biotechnology and Department of Chemistry, School of Natural Sciences, Faculty of Science and Engineering, University of Manchester, 131 Princess Street, Manchester M1 7DN, U.K.

ABSTRACT: Peptide aptamers built using engineered scaffolds are a valuable alternative to monoclonal antibodies in many research applications because of their smaller size, versatility, specificity for chosen targets, and ease of production. However, inserting peptides needed for target binding may affect the aptamer structure, in turn compromising its activity. We have shown previously that a stefin A-based protein scaffold with AU1 and Myc peptide insertions (SQT-1C) spontaneously forms dimers and tetramers and that inserted loops mediate this process. In the present study, we show that SQT-1C forms tetramers by self-association of dimers and determine the kinetics of monomer–dimer and dimer–tetramer transitions. Using site-directed mutagenesis, we show that while slow domain swapping defines the rate of dimerization, conserved proline P80 is involved in the tetramerization process. We also demonstrate that the addition of a disulphide bond at the base of the engineered loop prevents domain swapping and dimer formation, also preventing subsequent tetramerization. Formation of SQT-1C oligomers compromises the presentation of inserted peptides for target molecule binding, diminishing aptamer activity; however, the introduction of the disulphide bond locking the monomeric state enables maximum specific aptamer activity, while also increasing its thermal and colloidal stability. We conclude that stabilizing scaffold proteins by adding disulphide bonds at peptide insertion sites might be a useful approach in preventing binding-epitope-driven oligomerization, enabling creation of very stable aptamers with maximum binding activity.

INTRODUCTION

Peptide aptamers are proteins that consist of short target-binding polypeptide loops embedded within a stable protein scaffold, designed to bind specifically to a defined target. Engineered protein scaffolds are typically based on small native globular proteins, modified to remove original function and include new subcloning sites for adding the interchangeable loops. To achieve desired specificity and affinity, the sequences containing the desired binding epitope(s) (typically up to 10–15 residues) are usually inserted instead of the original loops. In principle, peptide aptamers mimic the antibody-based molecular recognition but typically have a much smaller frame (often ~15 kDa) and less complex structure and do not require post-translational modifications and therefore can be often produced in simpler recombinant expression systems.1 Peptide aptamers are applied in various research tasks, including the development of combinatorial protein libraries for protein recognition,2,3 studies of protein function and their interactions,4 diagnostic tools,5 biosensors,6 imaging agents,7 and as biotherapeutics.8 As such, peptide aptamers are an emerging valuable alternative to monoclonal antibodies which until now have prevailed as the “gold standard” for affinity binding studies.

More than 50 structurally diverse nonimmunoglobulin scaffolds have been reported to date.1 While protein scaffolds are designed to be as stable as possible, insertion of modified loops may however unintentionally destabilize them, leading to aggregation and reduction in thermal stability9 or cause larger structural rearrangements such as domain-swap oligomerization.10 Changes to protein tertiary and quaternary structures may influence conformation or presentation of the binding loops themselves, thus compromising target binding.

To explore in detail the structural and functional consequences of loop insertions, we are using a model engineered protein scaffold derived from stefin A, named SQT.11 Stefin A belongs to the cystatin superfamily of cysteine protease inhibitors, which also includes stefin B and cystatin C.12 SQT has three possible insertion sites for peptides, namely, the N-terminus, loop 1 and loop 2. While it has been...
shown in the original publication\textsuperscript{11} that SQT retains the secondary structure upon various peptide insertions, we have demonstrated in our previous study that an SQT variant, named SQT-1C, with AU1 and Myc peptides inserted into loop 1 and loop 2, respectively, has decreased thermal stability and poor solution behavior.\textsuperscript{10} Insertion of these epitopes led to spontaneous formation of interconverting monomeric, dimeric, and tetrameric species in solution, with such oligomerization directly mediated by the inserts in the engineered loops.\textsuperscript{10} Although the problem with domain-swap oligomerization and destabilization has been identified, it was not clear what the functional consequences of this oligomerization were, and how this structural instability could be prevented. 

In this present study, we have further explored the kinetics and mechanism of SQT-1C oligomerization. We determined that tetramerization occurs through self-association of domain-swapped dimers, with the formation of these dimers being the rate-limiting step. We have designed two SQT-1C variants. In the first variant, a P80G point mutation was introduced to explore the role of conserved proline 80 in tetramerization kinetics. For the second variant, a double mutant was designed, creating a disulphide bond which locked the configuration of the inserted loop 1. This drastically stabilized the monomeric species and prevented formation of domain-swapped dimers. Additionally, we show that oligomerization of SQT-1C reduces its target-binding capacity, whereas the disulphide bond-stabilized monomer had the highest specific activity. We conclude that stabilizing protein scaffolds by adding disulphide bonds at peptide insertion sites to stabilize the engineered loops might be a useful approach for preventing binding-epitope-driven oligomerization, while simultaneously also improving their thermal and colloidal stability.

\section*{RESULTS}

\textbf{SQT-1C Oligomerizes through Monomer–Dimer–Tetramer Pathway.} As previously shown\textsuperscript{10} monomeric SQT-1C is in equilibrium with dimeric and tetrameric species in solution; however, the exact oligomerization pathway has not been established. To determine the kinetic model of SQT-1C oligomerization, we have isolated monomeric, dimeric, and tetrameric protein fractions (Table 1) and followed the re-equilibration kinetics of each fraction using size exclusion chromatography (SEC). As shown in Figure 1A, monomeric SQT-1C first forms dimers, with tetramerization occurring only after a substantial amount of dimers accumulate in solution. This indicates that dimers act as an intermediate state on the oligomerization pathway to formation of tetramers. In the isolated dimer fraction (Figure 1B), the dimer population quickly converts to monomers and tetramers. Over time, the fraction of monomers remains constant, whereas the association of dimers into tetramers becomes predominant, with dimer concentration decreasing and tetramer population increasing. This further supports the observation that tetramerization occurs by association of dimers and that dimers are only an intermediate state in the oligomerization pathway. Finally in the tetrameric fraction, partial dissociation of tetramers into dimers and monomers occurs already during sample preparation, with the final equilibrium of predominantly tetramer population followed by monomer and then dimer reached after 2 h of incubation (Figure 1C). These data clearly demonstrate that SQT-1C is in dynamic equilibrium between monomeric, dimeric, and tetrameric species. Moreover, less than 1% of species with molecular weight higher than 60 kDa were present throughout the SEC experiments, indicating that tetramers are the preferred final state with no further higher-order aggregation occurring in the time frame of the experiments. SQT-1C oligomerization can be described by a sequential monomer–dimer–tetramer self-association model\textsuperscript{13} shown in Figure 2. We were able to fit successfully all the experimental SEC data on SQT-1C oligomerization kinetics at various protein concentrations to this model using DynaFit 4 software,\textsuperscript{14} obtaining a single set of global kinetic parameters. As shown in Figure 3, the monomer–dimer–tetramer model describes the data completely, further supporting the choice of the model.

\begin{table}[h]
\centering
\caption{Molecular Weights of Protein Oligomers (in kDa) as Determined by SEC–MALS}
\begin{tabular}{|l|c|c|c|}
\hline
 & SQT-1C & SQT-1C\textsuperscript{Nmyc} & SQT-1C\textsuperscript{Q46C,N59C} \\
\hline
monomer & 15.3 ± 0.5 & 15.3 ± 0.7 & 14.3 ± 0.5 \\
\hline
dimer & 32 ± 1 & 32 ± 1 & 30 ± 1 \\
\hline
tetramer & 62 ± 2 & 59 ± 2 & 61 ± 2 \\
\hline
\end{tabular}
\end{table}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure1.png}
\caption{SEC analysis of SQT-1C species interconversion. Time evolution of monomers, dimers, and tetramers after incubation of 5 mg/mL SQT-1C samples at 25 °C starting from (a) monomeric, (b) dimeric, and (c) tetrameric preisolated species.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure2.png}
\caption{Model of SQT-1C oligomerization. \( k_1 \) and \( k_{−1} \) are the on and off rates of dimer formation while \( k_2 \) and \( k_{−2} \) are the on and off rates for dimer–tetramer transition, respectively.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure3.png}
\caption{Structural Rearrangement of SQT-1C Monomers is the Rate-Limiting Step of Protein Dimerization. In order
Table 2. Summary of Estimated Rate Constants for SQT-1C Oligomerization at Different Temperatures

| T, °C | $k_1$ (M$^{-1}$ s$^{-1}$) | $k_2$ (s$^{-1}$) | $k_3$ (M$^{-1}$ s$^{-1}$) | $k_4$ (s$^{-1}$) |
|-------|--------------------------|---------------|--------------------------|---------------|
| 20    | 0.018 ± 0.003            | 0.1 ± 5       | 0.52 ± 0.2               | 9 ± 5         |
| 22    | 0.02 ± 0.005             | 0.1 ± 5       | 8.7 ± 7.1                | 0.1 ± 5       |
| 25    | 0.058 ± 0.008            | 0.1 ± 0.1     | 1.7 ± 0.6                | 2 ± 2         |
| 27    | 0.096 ± 0.01             | 0.1 ± 7       | 1.8 ± 0.5                | 3 ± 2         |
| 30    | 0.47 ± 0.12              | 210 ± 40      | 6.4 ± 3.7                | 2 ± 0.1       |
| 33    | 0.53 ± 0.34              | 500 ± 2000    | 7.7 ± 3.9                | 6 ± 2         |
| 35    | 1.25 ± 0.27              | 10 ± 100      | 110 ± 990                | 5 ± 8         |

*Reported errors are standard deviations obtained during data fitting in DynaFit 4 using the default settings.

to further explore the kinetics of SQT-1C oligomerization, we measured the temperature dependence of rate constants for SQT-1C dimerization and tetramerization, which were then used to estimate the apparent activation energy of these processes. Fits to experimental data are shown in Figure S1, whereas the obtained estimates of rate constants are shown in Table 2. The on-rates for both dimerization $k_1$ and tetramerization $k_2$ are significantly slower than expected for a simple diffusion-limited self-association process where the rate constants typically range from $10^5$ to $10^6$ M$^{-1}$ s$^{-1}$. This is consistent with structural rearrangements occurring on a slow timescale, responsible for both association steps. Furthermore, the rate for the monomer to dimer ($k_1$) reaction is an order of magnitude slower than that of dimer to tetramer transition ($k_2$) indicating that the two processes are governed by different structural rearrangement mechanisms and that dimer formation is the limiting step in the SQT-1C oligomerization pathway. Our previous molecular modeling data suggested that dimerization proceeds via the domain-swapping mechanism. Figure 4 shows Arrhenius plots of the temperature dependence data of SQT-1C dimerization and tetramerization rates $k_1$ and $k_2$. Fitting the data to the linearized Arrhenius equation allowed estimation of the apparent activation energies ($E_a$) for both processes. $E_a$ for SQT-1C monomer to dimer transition, $S_3 ± 5$ kcal/mol, was greater than that for dimer–tetramer transition, $36 ± 6$ kcal/mol. Activation energies for SQT-1C dimerization are consistent with previously reported values of $55 ± 4$ kcal/mol for domain-swap oligomerization of stefin B. Similarly, apparent activation energy for SQT-1C tetramerization is similar to that of stefin B tetramerization, $28 ± 3$ kcal/mol, and is consistent with slow reactions accompanied by minor, local conformational changes.

**Rationale for SQT-1C Mutant Design.** For the cystatin family, two distinct steps of protein association have been reported previously, both involving structural rearrangement and hence relatively slow timescale. The first one is domain-swap dimerization, where the domain swap occurs through extension of the conserved hydrophobic five-residue “cystatin motif” (QVVAG) in loop 1. In SQT, this motif has been mutated (to QVLAS) and split to accommodate peptide loop insertion into the scaffold;[9] therefore, this motif itself can no longer be responsible for the domain swapping (Figure S). For SQT-1C, our previous experiments and modeling suggest that it is the engineered loops themselves that drive domain-swap-mediated dimerization and further tetramerization. The second known step of cystatin association is a so-called hand-shaking mechanism, where the trans to cis isomerization of conserved P74 in loop 2 is required for the association of two domain-swapped dimers into a stable tetramer.[17]

To confirm the mechanism of SQT-1C oligomerization and to find way of preventing it, we have created two mutants. In the first mutant, named SQT-1C$^{P80}$, we have mutated the residue P80 (corresponding to P74 conserved in other cystatins) to glycine (Figure S), removing the possibility of slow trans–cis isomerization while allowing more conformational flexibility. Such a change could potentially either eliminate tetramerization, as shown previously for stefin B[17] and Na+–K+–ATPase,[20] or accelerate it due to increased flexibility of the loop 2. In the second mutant (SQT-1C$^{Q46C,N59C}$), we have introduced a disulfide bond across the base of loop 1, between $\beta1$ and $\beta2$ strands, by a double Q46C and N59C mutation (Figure Sb), in an attempt to stabilize a specific topology and prevent structural rearrangement. A similar approach has been used previously on cystatin C, where prevention of domain swapping using disulfide bridges inhibited dimerization and fibril formation.[21] The position of cysteines in SQT-1C$^{Q46C,N59C}$ was chosen for two reasons. First, their position at the end of the $\beta2$ and start of $\beta3$
should prevent opening of loop 1 in the monomer or covalently trap the domain-swapped dimer, preventing the interconversion of monomers and dimers and allowing their purification. Second, these mutations are positioned outside the restriction site NheI in loop 1 and hence should not affect insertion of the target-binding peptides into the SQT scaffold between L48 and A56 residues situated at the base of the loop (Figure 5B,D).

**Proline 80 is Involved in SQT-1C Tetramerization.** After separation of refolded SQT-1C P80G variant on SEC coupled with multiangle light scattering (SEC–MALS), three elution peaks were identified corresponding to monomer, dimer, and tetramer fractions (Table 1 and Figure S2). Far-UV circular dichroism (CD) spectra of freshly isolated monomeric SQT-1CP80G species showed little difference to that of isolated SQT-1C, indicating that the secondary structure of SQT-1C is retained in the SQT-1CP80G mutant (Figure S2). Additionally, only minor chemical shift perturbations of residues located next to the mutated residues were identified in 2D 1H−15N heteronuclear single quantum coherence (HSQC) spectra, further indicating that the 3D structure of SQT-1C monomer is retained in the P80G mutant (Figure S2). The melting temperature of SQT-1CP80G was $54 \pm 1$ °C, compared to $56 \pm 1$ °C for SQT-1C, as measured by the intrinsic fluorescence peak shift upon heating, showing that the P80G mutation does not significantly affect thermal stability. The colloidal stabilities of SQT-1C and SQT-1C$^{P80G}$, measured by static light scattering at 266 nm as onset temperature of aggregation ($T_{agg}$), were also very similar and coincided with their melting temperatures, suggesting major aggregation happening once the protein becomes thermally unfolded (Figure 7).

**Figure 5.** SQT-1C mutation scheme. (a) Model of SQT-1C based on PDB ID 6QB2 (b) scheme of loop 1 and mutations present in SQT-1C$^{Q46C,N59C}$ mutant. (c) Scheme of loop 2 and mutation introduced in SQT-1CP80G variant. (d) Sequence alignment of stefin A, SQT, SQT-1C, SQT-1C$^{P80G}$, and SQT-1C$^{Q46C,N59C}$. Secondary structure of SQT-1C, as determined from its structure (PDB ID 6QB2) is shown below sequences. In all panels, positions and names of restriction sites used to add functional loops to SQT scaffold are shown in pink, inserted loops are shown in blue, and unmodified regions in gray. Mutation site P80G is depicted in orange while Q46C and N59C are shown in green.

**Figure 6.** NMR chemical shift perturbation analysis of mutant variants. Per residue weighted backbone amide chemical shift perturbations for (a) SQT-1CP80G and (b) SQT-1C$^{Q46C,N59C}$ compared to SQT-1C show that chemical shift perturbations occur only around the mutation sites and at residues in spatial proximity of the mutated sites. Asterisks (*) denote mutation sites.
Dimerization rates $k_1$ and their temperature dependence were similar to those of SQT-1C, as evident from Arrhenius plots (cf. Figures 8A and 4A), with similar apparent activation energy, indicating that the P80 is not involved in the dimerization process. However, the estimated activation energy for dimer to tetramer transition was significantly lower for SQT-1C compared to SQT-1C (cf. Figures 8B and 4B).

These experiments overall reveal that although P80 in SQT-1C is not involved in the dimerization process, it is involved in tetramerization, and the trans-cis isomerization of this residue is likely a contributing factor.

**Monomeric State of SQT-1C can be Stabilized by Addition of a Disulphide Bond.** To stabilize the monomeric form of SQT-1C and prevent structural rearrangement leading to dimerization, a double Q46C and N59C mutant was produced so that a disulphide bond can spontaneously form at the base of loop 1 between $\beta2$ and $\beta3$ strands. Additionally, any domain-swapped dimers formed during refolding and oxidation will be also covalently stabilized, preventing their dissociation into monomers. The SQT-1CQ46C,N59C mutant was expressed, refolded, oxidized, and purified as described in the Materials and Methods. After separation on the size exclusion column, individual monomeric, dimeric, and tetrameric species of SQT-1C were isolated for further analysis (Table 1 and Figure S4). Notably, large populations of higher molecular weight oligomers were visible on SEC–MALS trace (Figure S4), likely formed by misfolding and cross-linking via disulphide bonds during the refolding/oxidation step. Far-UV CD spectra of freshly isolated SQT-1CQ46C,N59C monomeric species showed little difference to the SQT-1C monomer, indicating that the secondary structure of SQT-1C is retained (Figure S4). Additionally, only minor chemical shift perturbations of residues located next to the mutated residues were identified in 2D $^1$H–$^13$C HSQC spectra, further indicating that the 3D structure of the SQT-1C monomer is retained in SQT-1CQ46C,N59C (Figures 6B and S4). Moreover, the CD spectra show that isolated covalently cross-linked SQT-1CQ46C,N59C dimers and tetramers are structurally similar to dimers and tetramers formed by SQT-1C (Figure S5). In addition to structure retention, introduction of disulphide bond drastically increased the melting temperature of the monomeric species above 95 °C, with only minor changes in intrinsic fluorescence signal observed across temperature ramps (Figure 7). Additionally, only a slight increase in static light scattering (SLS) at 266 nm over increasing temperature was observed, much less than for SQT-1C or SQT-1C, further indicating that this bridged mutant is colloidaly stable up to very high temperatures (Figure 7). Hydrogen–deuterium (H–D) exchange rates for SQT-1CQ46C,N59C were lower than those for SQT-1C measured previously, but exchange still occurs within minutes (Figure S6), suggesting that even after addition of disulphide bond the monomeric structure somewhat lacks long-lived hydrogen bond networks.

To test whether the introduction of the disulphide bond between $\beta2$ and $\beta3$ sheets stabilized the monomeric species against transition into dimers and tetramers, we tested SQT-1CQ46C,N59C oligomerization kinetics across a range of temperatures and protein concentrations. While a small fraction (<5%) of dimers was present in initially isolated monomeric species due to lack of column resolution and slight overlap between the elution peaks of monomers and dimers in SEC, there was no further significant interconversion of monomeric species into dimers or higher oligomers observed over time in any of the tested conditions, even at higher concentrations (Figure S7). We can therefore conclude that engineered disulphide bond prevents the opening of the monomeric species through loop 1, hence preventing the domain swapping and dimerization and subsequent tetramerization. Consequently, SQT-1CQ46C,N59C stays in solution as a stable monomer, with greatly enhanced colloidal and thermal stability.

### Table 3. Summary of Estimated Rate Constants for SQT-1CQ46C,N59C Oligomerization at Different Temperatures

| $T\, ^\circ C$ | $k_1 \, (M^{-1} s^{-1})$ | $k_2 \, (M^{-1} s^{-1})$ | $k_f \, (s^{-1})$ | $k_2 \, (s^{-1})$ |
|--------------|-----------------|-----------------|---------------|---------------|
| 22           | 0.02 ± 0.01     | 0.1 ± 0.01      | 3 ± 0.1       | 6 ± 0.4       |
| 25           | 0.09 ± 0.01     | 3 ± 0.3         | 2.1 ± 0.3     | 7 ± 1         |
| 27           | 0.14 ± 0.01     | 5 ± 3           | 2.8 ± 0.6     | 10 ± 20       |
| 30           | 0.29 ± 0.02     | 0.1 ± 0.2       | 3 ± 2         | 10 ± 3        |
| 33           | 0.88 ± 0.03     | 0.2 ± 3         | 7 × 10$^8$ ± 3 × 10$^8$ | 14 ± 3        |
| 35           | 2.3 ± 2         | 15 ± 13         | 1.6 ± 1.1     | 2 ± 3         |

*Reported errors are standard deviations obtained during data fitting in DynaFit 4 using the default settings.*
implicit, assumptions in the engineered scaffold design is that the inserted target-binding loops are held by the scaffold in a correct conformation optimal for their binding and that the scaffold is stable enough to maintain this conformation throughout its preparation, storage, and usage lifecycle. In the case of SQT-1C, dimerization and tetramerization clearly change the conformation of binding epitope within loop 1, from hairpin to extended conformation, and also change its solvent exposure. Therefore the question arises: does this oligomerization alter the functional competency of SQT-1C? While isolated monomeric and oligomeric fractions of SQT-1C could not be tested previously for functionality due to fast interconversion between the species, SQT-1C<sup>Q46C,N59C</sup> mutant yielded stable monomers, dimers, and tetrators which now can be separated. Hence, we tested the binding efficiency of monomeric, dimeric, and tetrameric species of SQT-1C<sup>Q46C,N59C</sup> using an enzyme linked immunosorbent assay (ELISA) experiment with commercially available polyclonal antibodies against AU1 and Myc peptides located in loops 1 and 2, respectively.

The most efficient AU1-mediated binding (on total protein quantity basis) was observed for the monomeric protein, followed by dimer and then tetramer, for a range of total protein concentrations (Figure 9A). Both dimers and tetraters had a smaller binding capacity than monomers, presumably because of composition of several factors, namely, steric clashes, partial burial, and non-optimal extended conformation of loop 1 in domain-swapped oligomers. Binding efficiency of tetramer with AU1 antibody was not dissimilar to binding of this antibody to standalone Myc peptide, used as a negative control (Figure 9A). Interestingly, efficiency of SQT-1C<sup>Q46C,N59C</sup> binding to Myc peptide present in loop 2 is only slightly reduced by dimerization, whereas tetramerization significantly decreases the SQT-1C<sup>Q46C,N59C</sup> ability to present the Myc peptide to the respective antibody (Figure 9B), presumably because of burial of loop 2 within the tetramerization interface. Control reactions performed with 20 μg/mL standalone Myc peptide (positive control) showed similar efficiency of binding to Myc antibodies as denatured SQT-1C<sub>1C</sub>; however, SQT-1C<sup>Q46C,N59C</sup> monomers exhibited even higher binding efficiency (Figure 9B). Binding of SQT-1C<sup>Q46C,N59C</sup> monomers to both AU1 and Myc antibodies was consistently more efficient than that of denatured SQT-1C which was added to ELISA reactions as a control (Figure 9). It can be envisaged that upon addition and dilution, this control WT SQT-1C partially refolds and partitions into a usual mixture of monomers, dimers, and tetraters; thus, it is expected to have an appreciable binding affinity for both antibodies. These results clearly indicate that SQT-1C monomers possess the highest specific activity toward target binding, whereas for domain-swapped dimers and particularly tetraters, the specific binding activity is significantly reduced. This finding provides a rationale for stabilizing a specific form of an engineered protein scaffold, in the case of SQT-1C it is monomeric form, to achieve maximum specific activity, as well as to improve its thermal and colloidal stability and prevent domain swapping. These several beneficial effects can be achieved simultaneously by the introduction of a single disulphide bond at the base of target-binding loop which otherwise drives domain swapping and oligomerization.

**DISCUSSION**

The ability of engineered protein scaffolds to retain their structural, thermal, and colloidal stability upon insertion of various peptide loops needed for their target-binding function is crucial for their research and industrial applications. As such, the small frame of scaffolds needs to absorb additional steric strains introduced by the inserted loops and to maintain the correct conformation. We have shown that protein scaffold SQT-1C forms domain-swap dimers that further associate into stable tetraters, in a similar way to oligomerization pathway of other proteins in the cystatin family. Slow kinetics and high apparent activation energy of SQT-1C dimerization are consistent with large structural rearrangement needed for dimer formation. Interestingly, the apparent activation energy for SQT-1C dimerization is roughly twice smaller than that reported for stefin A and is similar to the activation energy in the nucleation phase of fibrillation reaction for stefin B, the less stable of the two stefins. While monomer—dimer and dimer—tetramer transitions occur on a similar timescale to domain-swap dimerization and subsequent tetramerization and fibrillation of other members of the cystatin family, SQT-1C oligomerizes at room temperature, whereas oligomerization of other members of the cystatin family, including stefin A, stefin B, and cystatin C, occurs only at elevated temperature, in the presence of organic solvents or in the presence of denaturants. This suggests that the insertion of the specific peptides in SQT-1C significantly lowers the stability of the protein, making it more likely to form domain-swapped dimers.

Domain-swap dimerization in the cystatin family has been previously reported to occur through extension of the conserved hydrophobic five-residue “cystatin motif” (QVQAG) in Loop 1 as a consequence of frustration of this hairpin hinge region. It has been shown recently that this motif, when engineered into the hinge of a β-hairpin, causes domain swapping of otherwise nondomain-swapped proteins. On the other hand, it has been established that mutations in this hinge region can slow down or even completely eliminate domain-swap oligomerization of cystatins. Even though V48L mutation has been introduced into this particular motif in SQT to introduce NheI restriction site for insertion of peptides between L48 and the following alanine into the scaffold and to prevent domain-swap oligomerization, SQT-1C still forms domain-swap dimers at room temperature. As we have shown previously by measuring the temperature of unfolding of T<sub>μ</sub> and studies of H−D exchange rates, the SQT-1C structure somewhat lacks a
stable hydrogen bonding network. It is likely that structural frustration introduced by AU1 peptide in the loop 1 destabilizes the construct, with the frustration relieved by fully extending AU1 peptide conformation. This is achieved in the domain-swapped dimers, which associate further into tetramers. Introducing the disulphide bond stabilizing loop 1 in a hairpin configuration greatly increases thermal and colloidal stability of SQT-1C<sup>Q46C,N59C</sup> and essentially prevents monomer–dimer transition, allowing isolation of a stable monomeric form, which also exhibits the highest specific activity toward binding antibodies for both loops 1 and 2.

In proteins from the cystatin family, tetramerization occurs via a so-called hand-shaking mechanism, where the trans to cis isomerization of conserved P74 in loop 2 of these proteins drives association of two domain-swapped dimers into a stable tetramer. In SQT-1C, the correspondent residue is P80, which was mutated here to a glycine to remove contribution from proline isomerization, producing a SQT-1C<sup>P80G</sup> mutant. We show that P80G mutation significantly reduces the activation energy needed for tetramer formation. This suggests that trans–cis isomerization of P80 in SQT-1C may be one of the transitions needed for loop 2 to adopt conformation favorable for tetramerization and engage in interaction with the neighboring chain. Overall the tetramers of the SQT-1C<sup>P80G</sup> mutant are structurally similar to those formed by SQT-1C.

This is in contrast to previous observations in stefin B, where mutation of the conserved proline disrupted the typical pathway of oligomerization, leading to fibril formation. Instead, amorphous aggregates were formed without clear monomer–dimer–tetramer transition. This highlights the subtle differences between the engineered protein scaffold, and its native ancestors.

Engineering disulphide bonds into the protein core is generally a well-established method to increase protein stability. A disulphide bond was successfully introduced previously in cystatin C to prevent its dimerization and eliminate fibril formation. Here, we have introduced a disulphide bond between β2 and β3 sheets of SQT by a double Q46C and N59C mutation in an attempt to stabilize the monomeric form and prevent interconversion of monomeric species into dimers and tetramers. However, the effect of this disulphide bond on SQT protein scaffold was quite dramatic, not only locking the structure in monomeric form and preventing oligomerization but also raising the melting temperature and onset temperature of aggregation above 95 °C. This increase surpasses the 79.9 °C melting temperature of the original “empty” SQT scaffold itself. From our functional binding experiments, we found that this monomeric form had the highest specific binding activity, compared with dimers and tetramers, suggesting that monomers ensure the best presentation of the target-binding epitopes. This further implies that in the case of SQT scaffold, its major degradation pathway, formation of soluble dimers and tetramers, is detrimental to its functional activity. As it can be anticipated that addition of target-binding loops in other small engineered protein scaffolds may introduce similar strains on the core structure, leading to domain swapping and subsequent oligomerization, we propose that adding disulphide bonds at the base of ligand binding loop(s) may increase scaffold stability and maximize its specific target-binding activity.

### MATERIALS AND METHODS

#### Plasmids
Synthesized codon-optimized gene constructs of SQT-1C and two mutants, SQT-1C<sup>P80G</sup> and SQT-1C<sup>Q46C,N59C</sup>, were obtained from GeneArt (Thermo Fisher Life Technologies) and subcloned into pET21a+ vector with a cleavable hexa-histidine tag as previously described.

#### Protein Expression and Purification
All three SQT-1C variants were expressed as previously described. Although SQT-1C and SQT-1C<sup>P80G</sup> were purified as previously reported, cell pellets of SQT-1C<sup>Q46C,N59C</sup> were resuspended in denaturing buffer (20 mM NaPi, 500 mM NaCl, 6M GdnHCl, 5 mM tris(2-carboxyethyl)phosphine, pH 8.0) with 0.5% v/v Triton X-100 (Sigma-Aldrich). Resuspended pellets were then lysed by sonication with Sonopuls HD 3200 ultrasonic homogenizer equipped with TT13/F2 probe (Bandelin) and clarified by centrifugation at 30 000g for 30 min at 4 °C. Supernatants were transferred onto Ni-NTA resin (Qiagen) in a gravity flow column and incubated for 90 min at 25 °C. After incubation, columns were washed with respective denaturing buffers supplemented with 10 mM imidazole. The bound material was eluted with 500 mM imidazole in denaturing buffer. Refolding and oxidation of SQT-1C<sup>Q46C,N59C</sup>, enabling disulphide bond reshuffling, was achieved by 1:10 v/v rapid dilution where 1 mM reduced GSH and 0.25 mM oxidized GSH were added to refolding buffer (20 mM NaPi, 150 mM NaCl, 5 mM ethylenediaminetetraacetic acid, pH 7.2), followed by overnight dialysis into refolding buffer without GSH. Finally, the proteins were purified on a Superdex 200 26/600 HiLoad column (GE Life Sciences) and pre-equilibrated with refolding buffer. All SQT-1C variants eluted as a set of well-defined oligomers, allowing isolation of monomeric, dimeric, and tetrameric fractions. Isolated species were then re-concentrated to the desired concentration using Vivaspin 20 centrifugal devices with a 5 kDa molecular weight cutoff (Sartorius Stedim Biotech GmbH). Protein concentrations were estimated by absorbance at 280 nm (ε = 14900 M<sup>−1</sup> cm<sup>−1</sup>). Molecular weights of protein species were determined using SEC coupled with multangle light scattering (SEC–MALS) run at 25 °C. Protein samples (200 μg) were injected on Superdex 200 10/300GL column (GE Life Sciences) and passed through a Wyatt DAWN Heleos II EOS 18-angle laser photometer (Wyatt Technology) coupled to a Wyatt Optilab rEX (Wyatt Technology) refractive index detector. Data analysis was performed in ASTRA 6.1 software (Wyatt Technology).

#### CD Spectroscopy
Far-UV CD spectra of individual oligomeric species were acquired on an Applied Photophysics Chirascan using a 0.01 cm path length quartz cell, immediately after separation in SEC column at protein concentration of 1 mg/mL. The wavelength was varied from 190 to 280 nm with 0.5 nm step and acquisition time of 3 s per point. For each CD spectrum, three scans were averaged and smoothed.

#### Static Light Scattering and Intrinsic Fluorescence
SLS and intrinsic fluorescence measurements were conducted simultaneously using an UNcLe (Unchained Labs) across a temperature ramp from 20 to 90 °C with a heating rate 1 °C min<sup>−1</sup>. Data were processed using the UNcLe analysis software, as per manufacturer’s recommendations. Melting temperatures (T<sub>m</sub>) of all three SQT-1C variants were determined using temperature dependence of the first derivative of the barycentric mean (BCM) of fluorescence intensity, following standard instrument procedure. SLS at 266 nm was used as an
indicator of protein colloidal stability, where the onset of aggregation temperature ($T_{\alpha_1}$) was defined as the temperature at which the measured scattering signal reaches 10% of its maximum value.

**Monitoring SQT-1C Oligomer Transitions by SEC.**

Oligomerization kinetics of all SQT-1C variants was measured using SEC, with a Superdex 200 10/300GL column (GE Life Sciences) attached to an Agilent 1100 Series HPLC system (Agilent Technologies) at 25 °C, and pre-equilibrated in refolding buffer. Elution of samples was detected at 280 nm. Isolated monomer fractions of SQT-1C fractions at 5 mg/mL were incubated for 24 h at 20, 22, 25, 27, 30, 33, 35, and 40 °C with 10 µL sample aliquots injected onto the SEC column every 90 min. Kinetic data were obtained as a series of single independent runs and kinetics at 25 °C was measured twice to check for data reproducibility. For concentration-dependent oligomerization kinetics at various concentrations and temperatures, data were processed and analyzed using MARS (BMG LABTECH) and OriginPro 9.5.1 (OriginLab). The experimental SEC data on monomer−dimer−tetramer transitions, isolated monomers were incubated at 25 °C at 1, 5, and 10 mg/mL, and the kinetics of oligomerization was analyzed as described above. All peaks in SEC traces were integrated and the kinetics of oligomerization was analyzed as described above. The ratio between dimer−tetramer transition, isolated monomers was incubated at 25 °C at 1, 5, and 10 mg/mL, and the kinetics of oligomerization was analyzed as described above. All peaks in SEC traces were integrated and finally, the ratio between different oligomers was calculated. Data analysis was performed in ChemStation (Agilent Technologies) and OriginPro 9.1 (OriginLab). The experimental SEC data on oligomerization kinetics at various concentrations and temperatures were then fit to the monomer−dimer−tetramer model using DynaFit 4 software.13 Apparent activation energies for monomer−dimer−dimer−tetramer transitions were estimated by fitting the data to the linearized Arrhenius equation:

$$\ln(k) = \ln(A) - \frac{E_a}{RT},$$

where $k$ is rate constant for each oligomerization step, $A$ the pre-exponential factor, $R$ a gas constant, $T$ is absolute temperature, and $E_a$ the apparent activation energy.

**NMR Experiments.**

NMR samples were prepared by adding 5% v/v 2H$_2$O to 1 mM 15N-labeled protein solutions in 20 mM sodium phosphate buffer, 150 mM NaCl, pH 7.2. All NMR spectra were acquired at 25 °C on 800 MHz Bruker AVANCE III spectrometer equipped with 5 mm triple resonance TCI CryoProbe and temperature control unit. The spectra were acquired and processed using Bruker TopSpin 3.5 and analyzed using NMRFAM-SPARKY20 and Dynamics Center 2.2.4 (Bruker). Backbone assignment of SQT-1C has been previously described10 (BMRB ID 27757). The backbone assignment of SQT-1C was then transferred to SQT-1C$^{P80G}$ and SQT-1C$^{Q46C,N59C}$ 2D 1H−15N HSQC spectra by matching peak positions. Assignment of shifted cross-peaks of SQT-1C$^{Q46C,N59C}$ was additionally verified using 3D TROSY-based HNCA and HNCO experiments from standard Bruker pulse program library. Proton−deuterium (H−D) exchange rates of SQT-1C$^{Q46C,N59C}$ were measured as previously described.10 The weighted chemical shift changes of backbone amide groups ($\delta_{NH}$) due to point mutations in SQT-1C$^{P80G}$ and SQT-1C$^{Q46C,N59C}$ were calculated as

$$\Delta\delta_{NH} = \sqrt{0.5(\Delta\delta_H)^2 + (\Delta\delta_N/10)^2},$$

where $\Delta\delta_H$ and $\Delta\delta_N$ were chemical shift changes in proton and nitrogen dimensions, respectively.

**Enzyme Linked Immunosorbsent Assay.** To examine the ability of SQT-1C$^{Q46C,N59C}$ oligomers to present inserted peptides to target antibodies, plastic MaxiSorb plates (Nunc) were coated with 1, 2, 5, 10, and 20 µg/mL of SQT-1C$^{Q46C,N59C}$ monomers, dimers, and tetramers in phosphate-buffered saline pH 7.4 (PBS) overnight at 4 °C with shaking. c-Myc protein at 20 µg/mL was used as a positive control for anti-Myc antibody binding and as negative control for anti-AU1 antibody binding, whereas 20 µg/mL denatured SQT-1C was used as another positive control for both antibodies. Additionally, bovine serum albumin (BSA) was used as a negative control. Protein concentrations were measured by UV absorbance at 280 nm wavelength. c-Myc protein sample was kindly provided by Dr Matthew Cliff and Prof Jon Waltho (University of Manchester). All samples were measured in triplicates. Plates were blocked with 2% (w/v) BSA (Sigma-Aldrich) in PBS at 25 °C for 2 h. Plates were then incubated with either goat anti-AU1 primary antibody (ab3400, Abcam) diluted 1:2000 or goat anti-Myc tag primary antibody (ab9132, Abcam) diluted 1:25 000 for 2 h at 25 °C, followed by incubation at 25 °C for 1 h with rabbit anti-gold secondary antibody labeled with horseradish peroxidase (ab6741, Abcam) diluted 1:50 000 with PBS. Between incubation steps, plates were washed using 0.05% (v/v) Tween 20 in PBS. After incubation with secondary antibody, plates were incubated with the 3,3′,5,5′-tetramethy benzidine (TMB) substrate (Abcam) for 15 min at 25 °C. The reaction was stopped by addition of 450 nm Stop Solution for TMB Substrate (Abcam), and the absorbance was read at 450 nm using a multilaw plate reader CLARIOstar (BMG LABTECH). Data were processed and analyzed using MARS (BMG LABTECH) and OriginPro 9.5.1 (OriginLab).

---

**ASSOCIATED CONTENT**

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.9b02269.

---

**AUTHOR INFORMATION**

**Corresponding Author**

*E-mail: a.golovanov@manchester.ac.uk. Phone: +44 (0) 161 306 5813. Fax: +44 (0) 161 306 5201.*

**ORCID**

Matja Zalar: 0000-0002-4807-0177

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

**Author Contributions**

M.Z. and A.P.G conceived and designed the study. M.Z. conducted the experiments and analyzed the data. The manuscript was written through contributions of both authors.

**Funding**

This work was supported by EU Horizon 2020 Research and Innovation program under the Marie Skłodowska-Curie grant agreement No 675074.

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

We would like to thank Dr Robin Curtis and Prof Alan Dickson, and their groups for letting us use their research facilities. We are especially grateful to Prof Jon Waltho for...
useful discussions and his comments on the work. We would also like to thank Dr. Matthew Cliff from the MIB NMR Facility for technical support with NMR spectrometers, Dr. Tom Jowitt from Biomolecular Analysis Facility for technical support with SEC-MALS, Dr. Derren Heyes from the MIB Biophysics Facility for technical support with the CD spectrometer, and Dr. Tim Eyes from SynBioChem for advice on ELISA experiments.

**ABBREVIATIONS**

SEC, size-exclusion chromatography; SEC−MALS, SEC coupled with multangle light scattering; SLS, static light scattering; BCM, barycentric mean; HSQC, heteronuclear single quantum coherence; ELISA, enzyme linked immunosorbent assay

**REFERENCES**

(1) Reverdatto, S.; Bura, D.; Shekhtman, A. Peptide aptamers: development and applications. *Curr. Top. Med. Chem.* 2015, 15, 1082−1101.

(2) Colas, P.; Cohen, B.; Jessen, T.; Grishina, I.; McCoy, J.; Brent, R. Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature 1996*, 380, 548−550.

(3) Nord, K.; Gunneriusson, E.; Ringdahl, J.; Ståhl, S.; Uhle, M.; Nygren, P.-Å. Binding proteins selected from combinatorial libraries of an α-helical bacterial receptor domain. *Nat. Biotechnol.* 1997, 15, 772−777.

(4) Skerra, A. Alternative non-antibody scaffolds for molecular recognition. *Curr. Opin. Biotechnol.* 2007, 18, 295−304.

(5) Retout, M.; Valkenier, H.; Triffaux, E.; Doneux, T.; Bartik, K.; Braylants, G. Rapid and Selective Detection of Proteins by Dual Trapping Using Gold Nanoparticles Functionalized with Peptide Aptamers. *ACS Sens.* 2016, 1, 929−933.

(6) Justino, C. I. L.; Freitas, A. C.; Pereira, R.; Duarte, A. C.; Santos, T. A. P. R. Recent developments in recognition elements for chemical sensors and biosensors. *TrAC, Trends Anal. Chem.* 2015, 68, 2−17.

(7) Skerra, A. Anticalins as alternative binding proteins for therapeutic use. *Curr. Opin. Mol. Ther.* 2007, 9, 336−344.

(8) Friedman, M.; Orlova, A.; Johansson, E.; Eriksson, T. L. J.; Hőlden-Guthenberg, I.; Tolmachev, V.; Nilsson, F. Y.; Stålh, S. Directed evolution to low nanomolar affinity of a tumor-targeting epidermal growth factor receptor-binding affibody molecule. *J. Mol. Biol.* 2008, 376, 1388−1402.

(9) Hoffmann, T.; stadler, L. K. J.; Busby, M.; Song, Q.; Buxton, A. T.; Wagner, S. D.; Davis, J. J.; Ko Feronig, P. Structure-function studies of an engineered scaffold protein derived from stefin A: Development of the SQM variant. *Protein Eng., Des. Sel.* 2010, 23, 403−413.

(10) Zalar, M.; Indrakumar, S.; Levy, C. W.; Tunnicliffe, R. B.; Peters, G. H. J.; Golovanov, A. P. Studies of the oligomerization mechanism of a cystatin-based engineered protein scaffold. *Sci. Rep.* 2019, 9, 9067.

(11) Stadler, L. K. J.; Hoffmann, T.; Tomlinson, D. C.; Song, Q.; Lee, T.; Busby, M.; Nyathi, Y.; Gendra, E.; Tiede, C.; Flanagan, K.; Cockell, S. J.; Wipat, A.; Harwood, C.; Wagner, S. D.; Knowles, M. A.; Davis, J. J.; Keegan, N.; Ko Feronig, P. Structure-function studies of an engineered scaffold protein derived from stefin A: II. Development and applications of the SQF variant. *Protein Eng., Des. Sel.* 2011, 24, 751−763.

(12) Turk, V.; Stoka, V.; Turk, D. Cystatins: biochemical and structural properties, and medical relevance. *Front. Biosci.* 2008, 13, 5406−5420.

(13) Garai, K.; Frieden, C. The Association−Dissociation Behavior of the ApoE Proteins: Kinetic and Equilibrium Studies. *Biochemistry* 2010, 49, 9533−9541.

(14) Kuzmič, P. Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV protease. *Anal. Biochem.* 1996, 237, 260−273.

(15) Schreiber, G.; Haran, G.; Zhou, H.-X. Fundamental Aspects of Protein−Protein Association Kinetics. *Chem. Rev.* 2009, 109, 839−860.

(16) Škerget, K.; Višňov, A.; Pomep-Novak, M.; Turk, V.; Waltho, J. P.; Turk, D.; Žerovnik, E. The mechanism of amyloid-fibril formation by stefin B: Temperature and protein concentration dependence of the rates. *Proteins 2009*, 74, 425−436.

(17) Jenko Kokalj, S.; Gunčar, G.; Štrem, I.; Morgan, G.; Rabzelj, S.; Kenig, M.; Staniforth, R. A.; Waltho, J. P.; Žerovnik, E.; Turk, D. Essential Role of Proline Isomerization in Stefin B Tetramer Formation. *J. Mol. Biol.* 2007, 366, 1569−1579.

(18) Jerala, R.; Žerovnik, E. Accessing the global minimum conformation of stefin A dimer by annealing under partially denaturing conditions. *J. Mol. Biol.* 1999, 291, 1079−1089.

(19) Sanders, A.; Craven, C. J.; Higgins, L. D.; Giannini, S.; Conroy, M. J.; Houndslow, A. M.; Waltho, J. P.; Staniforth, R. A. Cystatin forms a tetramer through structural rearrangement of domain-swapped dimers prior to amyloidogenesis. *J. Mol. Biol.* 2004, 336, 165−178.

(20) Geering, K.; Jaunin, P.; Jaisser, F.; Merillard, A. M.; Horisberger, J. D.; Mathews, P. M.; Lemas, V.; Fambrough, D. M.; Rossier, B. C. Mutation of a conserved proline residue in the beta-subunit ectodomain prevents Na(+)-K(+)-ATPase oligomerization. *Am. J. Physiol.* 1993, 265, C1169−C1174.

(21) Nilsson, M.; Wang, X.; Rodziwicz-Motowidlo, S.; Janowxsi, R.; Lindström, V.; Önerfford, P.; Westemark, G.; Frzonka, Z.; Jaksolks, M.; Grubb, A. Prevention of Domain Swapping Inhibits Dimerization and Amyloid Fibril Formation of Cystatin C. *J. Biol. Chem.* 2004, 279, 24236−24245.

(22) Žerovnik, E.; Virden, R.; Jerala, R.; Turk, V.; Waltho, J. P. On the mechanism of human stefin B folding: I. Comparison to homologous stefin A. Influence of pH and trifluoroethanol on the fast and slow folding phases. *Proteins: Struct., Funct., Genet.* 1998, 32, 296−303.

(23) Staniforth, R. A.; Giannini, S.; Higgs, L. D.; Conroy, M. J.; Houndslow, A. M.; Jerala, R.; Craven, C. J.; Waltho, J. P. Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily. *EMBO J.* 2001, 20, 4774−4781.

(24) Rodziwicz-Motowidlo, S.; Iwaszkiewicz, J.; Osowska, R.; Czaplewska, P.; Sobolewski, E.; Szymańska, A.; Stachowiak, K.; Liwo, A. The role of the Val57 amino-acid residue in the hinge loop of the human cystatin C. Conformational studies of the beta2-L1-beta3 segments of wild-type human cystatin C and its mutants. *Biopolymers* 2009, 91, 373−383.

(25) Ferreira, D. U.; Komives, E. A.; Wyolnes, P. G. Frustration in biomolecules. *Q. Rev. Biophys.* 2014, 47, 285−363.

(26) Macarenaunas, N. M.; Gosavi, S. Protein Domain-Swapping Can Be a Consequence of Functional Residues. *J. Phys. Chem. B* 2016, 120, 6929−6939.

(27) Nandwani, N.; Surana, P.; Negi, H.; Mascarhenas, N. M.; Uedaonkar, J. B.; Das, R.; Gosavi, S. A five-residue motif for the design of domain swapping in proteins. *Nat. Commun.* 2019, 10, 452.

(28) Olrikowska, M.; Jankowska, E.; Kołodziejczyk, R.; Jaksolks, M.; Szymańska, A. Hinge-loop mutation can be used to control 3D domain swapping and amyloidogenesis of human cystatin C. *J. Struct. Biol.* 2011, 173, 406−413.

(29) Dombkowski, A. A.; Sultana, K. Z.; Craig, D. B. Protein disulfide engineering. *FEBS Lett.* 2014, 588, 206−212.

(30) Lee, W.; Tonelli, M.; Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* 2014, 31, 1325−1327.