Short flexible peptides linking non-interacting protein domains appear to resist proteolysis by facilitating domain motions that sterically inhibit protease approach

Snehal Waghmare  
Indian Institute of Science Education and Research Mohali

Pumananda Guptasarma (✉ guptasarma@iisermohali.ac.in)  
Indian Institute of Science Education and Research Mohali

Research Article

Keywords: protein fusion, linker peptide, proteases, proteolytic susceptibility, thermostable domains

Posted Date: June 26th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3091794/v1

License: ©  This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Our objective in this paper was to identify any general characteristics of linker peptides that could potentially facilitate their escape from proteolytic degradation in fusion proteins, following their placement between two evolutionarily-unrelated, non-interacting, proteolytically-resistant protein domains. We selected Coh2 [an all-beta cohesin domain from *C. thermocellum* CipA], and BSX [a beta/alpha barrel xylanase domain from *Bacillus sp. NG-27*], and joined these domains through five different linker peptide sequences: (i) Rigid [3 repeats of N-EAAAK-C], (ii) Flexible [two repeats of N-SGGGG-C], (iii) Nat-Full [42 residues of a Coh2-adjacent linker in CipA], (iv) Nat-Half [a 21 residues-long derivative of Nat-Full] and (v) Nat-Quarter [a 9 residues-long derivative of Nat-Full]. The fusion proteins were produced, and size-exclusion chromatography and denaturing as well as non-denaturing electrophoresis were performed to assess the proteolytic susceptibilities of linker peptide sequences during (a) storage of the fusion proteins at 4°C, through proteolysis effected by trace amounts of proteases naturally present in solution, and (b) exposure at room temperature to Subtilisin A, a non-sequence-specific protease. In both cases, proteolytic degradation of the linker was observed in Rigid, Nat-Full, Nat-Half and Nat-Quarter, but not in Flexible. Our further data and analyses suggest that, unlike with the other four other linkers, Flexible is able to escape proteolysis by using its conformational flexibility to facilitate motions in its flanking (non-interacting) domains to deflect approaching proteases, thus allowing such freely-moving domains to sterically protect the linker's backbone, in ways that would be unlikely if the flanking domains were to interact.

Introduction

Protein engineering experiments can involve the fusion of protein domains, or proteins constituted of a single subunit. Quite often, the domains/proteins that are fused together do not naturally adjoin each other in the sequence(s) of any naturally-occurring protein(s). In such cases, since such domains/proteins have not co-evolved to exist upon the same polypeptide chain or possibly even function within the same biochemical pathway(s), very little scope exists for their engaging in any mutual interactions with each other. Thus, fusion proteins do not necessarily constitute rigid systems of mutually-interacting protein domains [1–3]. They can also involve freely moving domains that do not interact with each other and just happen to be located upon the same polypeptide chain. In principle, such domains can potentially be linked by peptide sequences that are naturally predisposed to be either rigid, or flexible. Choices regarding whether a linker is to be rigid, or flexible, or short, or long, must depend upon the degrees of freedom that one wishes to incorporate, in allowing each domain in the fusion protein to perform its own function with minimal interference from any other domain, using backbone bond rotations in linker peptide sequences. The success, or failure, of any protein fusion experiment is thus significantly determined by the choice of linker peptide sequences that one makes, since linkers can both (a) affect the functions, conformations, and/or conformational stabilities of flanking domains, and (b) happen to be naturally susceptible to undergoing proteolytic degradation by proteases present in the fusion protein's environment, or caused to be susceptible to such proteolysis (through the influence of flanking domains). As a consequence,
biochemists and biotechnologists tend to exercise as much care as they possibly can, in choosing linkers that do not themselves undergo proteolysis to release free domains that have no scope for remaining together through any noncovalent interactions. Unfortunately, such a choice of linkers is made difficult by the fact that linker sequences can potentially vary greatly in length, sequence, sequence-complexity, and conformational characteristics. Little is yet known about which general factors are likely to influence the proteolytic susceptibilities of linker peptides joining two non-interacting protein domains, and in what manner.

In the work presented in this paper, our objective was to focus solely upon the differential proteolytic susceptibilities of a small number of linker peptide sequences that could be held to be representative of different types of linkers (rigid, flexible, long, short, natively-present et cetera). Since we wished to study linker proteolytic susceptibilities without distractions involving either the effects of linkers upon the conformations of flanking domains, or the effects of flanking domains upon the conformations of linkers, we determined that we would use only linker-flanking domains that are themselves autonomously-folding, highly-thermostable, resistant to proteolysis, and incapable of engaging in mutual interactions. The rationale for choosing highly thermostable flanking domains could be summarized as follows. High conformational stability in proteins is typically associated with high resistance to proteolysis. This is mainly because high thermal stability tends to be correlated with better chain folding and, therefore, with better shielding of the protein's polypeptide backbone from the action of proteases present in the surrounding solvent [4–6]. In our experiments for this paper, we reduced the likelihood of proteolytic degradation occurring in flanking domains by using domains that are already known to be thermostable; initially suspected to be proteolytically-resistant and thereafter also confirmed to be proteolytically-resistant. The result was an effective elimination of scope for occurrence of proteolysis within flanking domains. This allowed us to focus solely upon the occurrence of proteolysis within linker peptide sequences.

Two thermostable domains were chosen for being fused: (i) Coh2, a Type I cohesin from Clostridium thermocellum CipA, and (ii) BSX, a Bacillus sp. Xylanase. The three-dimensional structure of BSX consists of a single (β/α)8-fold barrel, as is evidenced by the protein's atomic coordinates sourced from the Protein Data Bank, using the PDB ID, 2F8Q [7]. The three-dimensional structure of Coh2 consists of a nine-stranded barrel, with a jelly-roll fold topology consisting of two flattened β-sheets, as is evidenced by the protein's coordinates sourced from the Protein Data Bank, using the PDB ID, 1OHZ, Chain A [8]. We fused Coh2 to BSX at the genetic level, using five different linkers: a rigid linker, a flexible linker, and three additional linkers that were derived from the sequence of a 42 residues-long (native) peptide sequence joining the C-terminus of the Coh2 domain to the N-terminus of an adjacent carbohydrate binding module (CBM) in Clostridium thermocellum CipA. Further details about these linkers are available in the materials and methods section and the Supplementary Information file.

Briefly, we show that out of the five linkers with different characteristics that were used by us to join the same two protein domains (Coh2 and BSX), the linker that we already knew to be the most unstructured
(Flexible) proved, somewhat counter-intuitively, to also be the linker that consistently displayed maximal escape from proteolysis. Through a systematic elimination of possibilities, we concluded that it is this linker’s ability to facilitate motions in its flanking domains that allows those domains to sterically inhibit the approach of proteases, leading to the linker’s remarkable ability to escape proteolysis. Notably, in the only other discussion of differential linker proteolytic susceptibilities that we have been able to find in the published literature, a linker equivalent to Rigid was postulated to be better than a linker equivalent to Flexible, based on protein expression data and assumptions regarding possible proteolytic susceptibilities [9], i.e., the exact opposite conclusion, based on indirect inferences rather than upon direct experimental evidence concerning proteolytic susceptibilities. We think that the difference in perspective between the said report, and the work that is reported in this paper, owes to a fundamental difference in the nature(s) of the flanking domains, and their potential for engaging in mutual interaction(s), since our proposal is that a flexible linker is advantageous only when its flanking domains do not interact, and not when they do interact.

We hasten to clarify, therefore, that the work presented in this paper by no means suggests that Flexible is an ideal linker for use in every conceivable situation. We only propose that our work shows that a short and conformationally-flexible linker like Flexible is anticipated to be less proteolytically-sensitive than other (longer and/or more rigid) linkers, if it is placed between domains that do not interact, such that the flexibility of the linker imparts freedom of movement to its flanking domains. An apt analogy here would be the flexible metal chain of a ‘nunchuck’ which is used in the martial arts. Such a chain joins two rigid flanking rods that have no mutual interactions with each other. Thus, the flexibility of the nunchuck’s central chain allows its flanking rigid rods to move around with great freedom and, therefore, to sterically prevent the approach of other objects and bodies, including swords that can cut the central chain. Similarly, we aver that the flexibility of a linker placed between non-interacting flanking domains allows it to harness the natural (and independent) motions of its flanking domains to protect itself sterically from approaching proteases. Our results thus present a novel perspective that speaks to the importance of the nature of flanking domains, in deciding upon which peptide linkers to use for a particular protein fusion.

Materials and Methods

Choice of linkers. Nat-Quarter. N- NATPTKGAT-C (9 residues); Nat-Half: N-NATPTKGATPTNTATPTKSAT-C (21 residues); Nat-Full: N-NATPTKGATPTNTATPTKSATPTPRPSVPTNTPTNPANTP-C (42 residues); Rigid: N- EAAAKEAAAKEAAAK-C (15 residues); Flexible: N- SGGGGSGGGG-C (10 residues). Nat-Quarter and Nat-Half describe the first 9 and 21 residues, respectively, of Nat-Full, a linker lying on the C-terminal side of Coh2 in C. thermocellum CipA. Rigid is an archetypal rigid linker, consisting of 3 repeats of N-EAAAK-C. Flexible is an archetypal flexible linker, consisting of 2 repeats of N-SGGGG-C.

Construction of fusion proteins. To link Coh2 and BSX with the above linkers, (i) DNA encoding Coh2 (PDB ID: 1OHZ) was PCR-amplified from C. thermocellum genomic DNA (strain ATCC 27405); and (ii) DNA encoding BSX (PDB ID: 2F8Q) was gotten commercially synthesised. These were fused genetically through splicing by overlap extension (SOE) PCR, to produce five different Coh2-linker-BSX-encoding
fusion DNA sequences, using appropriately-designed primers incorporating the said linkers. The resultant genes were PCR amplified, digested by NdeI, and XhoI, and ligated into the T7-based expression vector, pET23a, for production in fusion with a C-terminal 6xHis affinity tag in *Escherichia coli* strain BL21 Star (DE3) pLysS, as described below. The details of oligonucleotide primers used (Supplementary Information file; pages 2–3), and the amino acid sequences of the fusion proteins constructed (Supplementary Information file; pages 4–5) are provided separately.

**Protein expression and purification.** Plasmid vectors (pET23a) bearing genes encoding Coh2-BSX fusions were transformed into XL-1 Blue *E. coli* to produce plasmids for DNA sequencing. These were then transformed into *E. coli* BL21 Star (DE3) pLysS for protein expression and purification. Overexpression was induced during growth of transformed cells in LB media by 1 mM IPTG in the mid-exponential phase of culture growth, at an optical density (O.D.\textsubscript{600}) of 0.6. Following induction, cells were incubated for 8 hours at 37°C, sedimented and then lysed. Expressed 6xHis-tagged protein fusions were chromatographically purified from clarified lysates through Ni-NTA affinity chromatography, using columns from GE Healthcare. Protein yields were typically 4 to 5 mg per litre of culture. All five Coh2-BSX fusion proteins, and the individual Coh2 and BSX proteins, were purified under non-denaturing conditions before transfer into 20 mM Na-HEPES buffer (pH 7.5, containing 100 mM NaCl, and 2 mM CaCl\textsubscript{2}). The transfer into this buffer was carried out through buffer-exchange upon a Superdex-75 size exclusion chromatography (SEC) column from GE Healthcare, using the GE Akta Purifier 10 chromatographic workstation. The same column was also set up and used in analytical mode to examine the elution behaviour of Coh2-linker-BSX fusion proteins incorporating different linkers, following 40 days of storage. The correctness of each construct was verified through DNA sequencing. The correctness of the identities of Coh2 and BSX were verified through peptide mass-fingerprinting mass spectrometry for Coh2 (using MALTI-Q-TOF) and BSX (using ESI-Q-TOF).

**SDS–PAGE analysis.** SDS-PAGE gels (12%) were used to examine the covalent (chain) integrity of Coh2-linker-BSX protein fusions. Protein markers with molecular weights ranging from 14.4 kDa to 116 kDa were run alongside. To detect proteolytic degradation, SDS-PAGE was performed both immediately following purification (0 days), and also following 15 days of storage at 4°C.

**Proteolysis-susceptibility assay.** Subtilisin A (a ~ 27 kDa serine protease) from *B. licheniformis* was sourced from Merck (Sigma), USA. The protease was added to the fusion constructs, and to Coh2 and BSX, individually, to examine the linker’s degree of resistance to proteolysis by exogenously-added Subtilisin A; a protease known to hydrolyse peptide bonds without any sequence specificity \cite{10}. Proteins (flanking domains, or fusion constructs) were incubated with Subtilisin for 2 h or 8 h at 50°C, using Subtilisin A: Protein molar ratios that varied from 1:1 to 1:1000, using stock solutions of Subtilisin A prepared in 20 mM Na-HEPES buffer (pH 7.5, containing 100 mM NaCl, and 2 mM CaCl\textsubscript{2}) and stock solutions of flanking domains or fusion constructs. Samples were analysed on 12% SDS–PAGE to assess proteolytic degradation.
Xylanase activity assay. Recombinant purified proteins were assayed for enzymatic activity upon Birchwood xylan (Sigma/Merck) by mixing the protein with xylan in HEPES buffer of pH 7.5, containing 100 mM NaCl, and 2 mM CaCl₂, using a final protein concentration of 0.5 µM and a final xylan concentration of 0.5% during the reaction. The reaction was incubated for 2 h at 50 °C. Dinitrosalicylic acid (DNSA) reagent (75 µl) was added to the reaction (50 µl) and heated at 80 °C for 30 min, for coupling of the reagent to the reducing ends of sugars released through glycosidic bond hydrolysis, assayed by the measurement of the colour developed upon heating through use of absorption measurements at 540 nm, following appropriate dilution (e.g., five-fold) to make measurements within the absorption range of the spectrophotometer used. Since only comparative (and not absolute) assays are presented, in order to examine any effects of the presence of the Flexible linker, and the Coh2 domain (in isolation and in fusion) upon the folding and enzymatic activity of the BSX xylanase, only relative data is presented using absorption values at 540 nm.

Results

Chromatographic examination of construct integrity during storage reveals that Flexible escapes proteolysis the most, amongst all linkers.

Figures 1A and 1B show mutually-normalized analytical chromatographic gel filtration (size exclusion) elution profiles of the five Coh2-linker-BSX fusion protein constructs. The profiles shown in Fig. 1A were collected immediately following purification. The profiles shown in Fig. 1B were collected following storage of purified protein for a period of 40 days at 4°C (arbitrarily taking this period to be representative of a typical period of storage in protein science laboratories, before the production and purification of a new batch of protein). The objective was to compare the chromatographic profiles of freshly prepared protein with protein subjected to storage, to examine whether storage causes proteolysis by trace amounts of proteases present in solution.

In Fig. 1A, each of the five constructs is observed to elute as a single species. Further, in each of the five elution profiles, the elution of this single species is observed to correspond to the elution volume anticipated to be seen for a monomeric form of each construct. Thus, elution volumes vary between 9.93 and 10.24 ml, which is the range of elution volumes in which these constructs are anticipated to elute from any ~ 24 ml Superdex-75 Increase (GE) gel filtration column. Given that the sizes of the flanking domains (Coh2 and BSX) are identical in all five constructs, the minor differences in elution volumes amongst constructs seem to be clearly attributable to differences in the size(s) of their constituent linkers. Therefore, in keeping with the reduction in hydrodynamic volumes of constructs, and increase in elution volumes, that is anticipated to be observed with reducing linker lengths, the five fusion proteins could be seen to elute from the column in order of reducing linker length: Nat-full (9.93 ml; 42 residues), Nat-half (10.18 ml; 21 residues), Rigid (10.20 ml; 15 residues), and Flexible (10.24 ml; 10 residues). Only with Nat-Quarter (10.15 ml; 9 residues) did we see a somewhat anomalous behavior.
In Fig. 1B, presenting the behaviour of the same five fusion constructs after 40 days of storage at 4°C, it is observed that constructs no longer necessarily elute as single-species, as earlier observed, between 9.93 and 10.24 ml. Instead, constructs are now observed to elute as two new peaks of significance, at ~11.5 ml, and ~13.5 ml, respectively, for four of the five constructs (i.e., in constructs incorporating Nat-Quarter, Nat-half, Nat-full and Rigid linkers). Further, constructs are also observed to elute as two additional minor peaks at ~19 ml and ~20 ml, for two of four constructs (i.e. in constructs incorporating Nat-full and Nat-half) displaying peaks at ~11.5 ml, and ~13.5 ml. With the fifth construct incorporating the Flexible linker, however, only a single-species identical to that seen for the elution of freshly purified protein in Fig. 1A was predominantly observed to elute at exactly the same volume as before, in sharp contrast to what was observed with the other four constructs. With the construct incorporating Flexible, negligible minor peaks were observed at ~11.5 ml and ~13.5 ml, and no significant peaks were observed at either ~19 ml or ~20 ml. These results clearly suggest the occurrence of substantial degradation in the linker peptides in four out of the five constructs during storage, but no degradation in the linker peptide in the fifth construct.

Next, we confirmed that the species eluting at ~11.5 ml, and ~13.5 ml, respectively, correspond to free BSX, and free Coh2. The control experiment demonstrating this is shown in Fig. 1C. In this figure, the elution of recombinant BSX and recombinant Coh2 are seen at ~11.5 ml, and ~13.5 ml, i.e., at exactly the elution volumes seen in the constructs that had undergone proteolytic scission. Notably, we did not find it particular surprising that BSX and Coh2 themselves appeared to have survived proteolysis upon storage, despite the evident proteolysis that had taken place in the linker peptides present within the same constructs. This is because BSX and Coh2 are thermostable domains, and (as emphasized in the introduction section), expected to resist proteolysis.

Consequently, one could say that at this point in the study (pending electrophoretic confirmation of conclusions; presented in the very next section) that (i) it stands established that trace amounts of proteases, present in solution, attack linker peptide backbones in four out of the five Coh2-linker-BSX constructs, during storage at 4°C. One could also conclude (ii) that this leads to physical separation of BSX and Coh2, leading to their elution from the column as independent, non-interacting entities. Further, one could also propose (iii) that the covalent integrity of BSX and Coh2 are maintained, owing to their own resistance to proteolysis (although this is only fully established using electrophoretic examination, in the next section). Mysteriously, it also stands established (iv) that the trace amounts of proteases that attack the linker peptide sequences in the other four fusion Coh2-linker-BSX constructs, somehow fail to do so in the Coh2-linker-BSX construct incorporating the Flexible linker.

**Denaturing electrophoretic examination of construct integrity during storage also reveals that Flexible escapes proteolysis the most, amongst all linkers.**

Gel filtration (size-exclusion) chromatographic experiments cannot really address the question of whether any intra-domain proteolytic cuts might have also occurred, within the polypeptide backbones of Coh2 or BSX, during storage at 4°C, i.e., chromatography cannot confirm whether cuts might have occurred that
did not, however, lead to physical separation of the polypeptides generated by such cuts. In this regard, it may be noted that it is theoretically possible for a folded protein to undergo proteolysis at one or two sites and still remain in a natively-folded state, through the continuance of most noncovalent interactions stabilizing the native structure. Therefore, the only way of examining whether any such cuts had indeed occurred would be to perform SDS-PAGE experiments (under denaturing conditions). Such experiments could then allow us to use any covalent separation of such polypeptides forced by the occurrence of whole-chain denaturation in the presence of SDS, to conclude that some cuts had indeed occurred. On the other hand, if no cuts whatsoever had occurred in the polypeptide backbones of BSX and Coh2, one would find only bands corresponding to intact polypeptide chains of BSX and Coh2, in the SDS-PAGE.

In Fig. 2A, we present SDS-PAGE profiles of the five fusion constructs, collected immediately after their purification. As with the chromatographic experiments, these SDS-PAGE experiments reinforce conclusions about the absence of degradation in any of the five fusion constructs at this stage. On the 12% SDS-PAGE used for this experiment, each fusion construct displays a mobility that is clearly attributable to the size of the extended (i.e., SDS-bound, and SDS- and heat-unfolded) polypeptide chain of construct. Therefore, the construct incorporating Nat-full shows the poorest mobility, owing to its possession of the longest linker sequence. Constructs incorporating Nat-half and Rigid display mobilities that are intermediate to those of constructs incorporating Nat-full, on the one hand, and Flexible and Nat-Quarter, on the other hand. Constructs incorporating Flexible and Nat-Quarter display the highest mobility, as anticipated, owing to their possession of the smallest linker peptide sequences joining the Coh2 and BSX domains. None of the constructs are observed to display any signs of anomalous mobility, or indeed any signs of having undergone significant degradation, in Fig. 2A.

In Fig. 2B, we present the SDS-PAGE profiles of the same protein constructs, this time after ~ 15 days of storage at 4°C, rather than after ~ 40 days of storage. Remarkably, even by this time point, there are signs of significant levels of degradation having occurred within the linker peptide sequences, in at least some of the constructs. Satisfyingly, (i) such degradation is observed in the same four constructs that had displayed a physical separation of Coh2 and BSX domains resulting from degradation of linker peptide sequences, in the chromatographic profiles in Fig. 2A. Equally satisfyingly, (ii) no degradation of either the Coh2 domain, or the BSX domain, could be seen in any single instance, fully establishing that only the linker peptide sequence had undergone proteolysis during storage. Notably, Coh2, and BSX, being well-folded domains with folding equilibria that keep the bulk of their populations in a folded state at all times, remain intact during storage, and undergo no degradation either (a) in a state of genetic fusion with each other, or (b) following separation through proteolytic scission of linker peptides. This much is clearly evident from the observation that there is no lowering in the intensity of the bands that correspond to intact Coh2-linker-BSX which cannot be attributed to (and compensated for by) the heightening of intensity of bands corresponding to the release of free BSX, and Coh2. In all cases, no visible band suggests the occurrence of degradation within BSX or Coh2. Thus, evidently only linker sequences undergo cleavage during storage at 4°C. The fact that this occurs at low rates, over a period of days, could owe to either the lower activity anticipated for proteolytic enzymes at low temperatures, or the low amounts of such proteases present in solution in the protein's environment.
In each case, in Fig. 2B, degradation during storage is observed to produce three distinct polypeptide species resolvable through SDS-PAGE: (i) an un-degraded (residual) Coh2-linker-BSX population [size > 60 kDa]; (ii) a population corresponding to BSX [size ~ 45 kDa], and (iii) a population corresponding to Coh2, with the linker peptide still presumably present at its C-terminus [size > 20 kDa). Figure 2B thus establishes that linkers undergo proteolysis at their BSX-adjoining ends, in all four constructs that undergo linker degradation, namely those incorporating Nat-Quarter, Nat-half, Nat-full and Rigid. What is most especially notable about Fig. 2B, however, is not the observation that BSX and Coh2 are not vulnerable to degradation by proteases present in the environment. Nor is it the observation that the linker peptide sequences in four out of the five fusion constructs undergo degradation. What is most notable instead is the fact that the Coh2-linker-BSX construct incorporating the Flexible linker (marked Flexible in Figs. 1 and 2) is clearly seen to have undergone almost no degradation, exactly as was concluded from the chromatography data, whereas all other constructs have clearly undergone degradation. This result confirmed by electrophoresis, continues to remain counter-intuitive, because the Flexible linker may be anticipated to be quite incapable of adopting any particular structure, as it is constituted of two serine residues and eight glycine residues (and because a glycine residue possesses only a hydrogen atom for a side chain, and typically imparts high conformational freedom to the chain wherever it is present). The linker peptide sequence, Flexible, is a popular linker that is used for the creation of fusion proteins precisely because the flexibility of Flexible ensures the least restriction of functioning of the flanking domains [11–12].

Our results thus suggest that Flexible undergoes no proteolysis when placed between well-folded domains that themselves resist proteolysis, even though it remains unstructured itself. We observe this even though a priori considerations would suggest that Flexible is also the most likely of the five linkers to undergo proteolysis, owing to its unstructured nature and its presumed inability to shield its peptide bonds from the surrounding aqueous solvent (and any proteases present in such a solvent). Further, it must be remarked here that such an observation has previously not been made because experiments involving poorly-folded flanking domains do also involve proteolysis occurring at multiple sites during storage, which smear-out distinctions between the ‘proteolytic vulnerabilities of linker sequences’ and the ‘proteolytic vulnerabilities of the sequences of the flanking domains’. In these experiments, since we ensured that no proteolysis occurs in the two flanking domains, and carried out experiments with different linkers that facilitated comparisons, we were able to make this distinction and clearly show that Flexible is the linker that maximally escapes proteolysis. Such a result could only have been obtained after ensuring that every other parameter remains equal in every other experiment conducted, i.e., with all linkers flanked by identical domains (with identical conformational stabilities), and with fusions subjected to identical physical and chemical conditions for testing, as well as identical methods of testing, and using only identical protein concentrations, and identical durations of incubation.

Chromatographic and native-PAGE experiments demonstrate Coh2 and BSX engage in no mutual interactions.
The Supplementary Information file presents the amino acid sequences of Coh2, BSX and all five fusion constructs, and also the mass spectrometry data confirming the covalent chemical identities of Coh2 and BSX. In Fig. 2C, we demonstrate by way of confirmation, in two different ways, that Coh2 and BSX display no interactions with each other. The main figure panel in Fig. 2C presents non-normalized gel filtration chromatographic data for both (i) elution of Coh2, or BSX, loaded independently upon a ~ 24 ml Superdex-200 Increase (GE) column (note: this is a different column from the one that was used for experiments reported in Figs. 1A, 1B and 1C), and also (ii) elution of Coh2 and BSX mixed, incubation overnight in each other's presence at 50°C, and then loaded together upon the same column.

Figure 2C shows that if Coh2 (which gives rise to the profile shown in blue, when it is loaded by itself) and BSX (which gives rise to the profile shown in red, when it is loaded by itself) are added to each other and then loaded onto the Superdex-200 Increase column after an overnight incubation, the resultant elution (which gives rise to the profile shown in black) appears to be the equivalent of a numerical summation of the curves obtained through independent loadings of Coh2 (the blue profile) and BSX (the red profile). This is very rigorous evidence that Coh2 and BSX do not interact, since if they did interact to produce a Coh2-BSX complex of any size, this would lead to the creation of a new entity that would display a smaller elution volume than BSX, due to its being bigger in size. Notably, however, no such peak is seen.

A second piece of evidence demonstrating the lack of interactions between Coh2 and BSX is shown in the inset within the Fig. 2C panel, which presents the results of a native PAGE in which no separate gel band is seen when Coh2 and BSX are added to each other and incubated overnight, i.e., there is no additional band that is not already clearly attributable to being derived from the loading of either Coh2, or BSX, upon the native gel, as shown in adjacent lanes on the gel shown in the inset to Fig. 2C. Therefore, considering the native PAGE and the chromatography data together, it may be said that we have clearly established that Coh2 and BSX do not interact with each other.

As an aside, it may be noted that tryptophan residues are absent in Coh2, whereas BSX contains 11 tryptophan residues. This causes Coh2 to display a much poorer absorption at 280 nm than BSX during gel filtration. As a consequence, when no normalization is done, based on peak absorption values of Coh2 and BSX (note: such a normalization was performed for Figs. 1A, 1B and 1C), the peak absorption of Coh2 can be seen to be very much lower than the peak absorption of BSX, even though the concentration of the loaded Coh2 happens to be twice the concentration of BSX, as was the case for the experiment done for the data shown in Fig. 2C.

Amongst all fusion constructs, the one incorporating the Flexible linker escapes proteolysis by Subtilisin A the most.

To further reinforce our counter-intuitive result showing that Flexible, which is anticipated to be an unstructured linker peptide sequence, displays the least susceptibility to proteolysis amongst five linker peptides tested, during storage at 4°C, we decided to further examine the resistance of all of these constructs by additionally exposing them to low concentrations of a non-specifically-acting serine protease, Subtilisin A, using either 2 h, or 8 h, incubations, and using varying molar ratios of Subtilisin.
Protein. It may be noted that Subtilisin A is a protease which is commonly used to test the relative proteolytic vulnerabilities of different regions of folded proteins, as well as to probe for regions that are not folded as tightly as other regions, through methods involving limited proteolytic digestion, which is typically performed using varying Subtilisin A:protein ratios, and/or varying durations of incubation of subtilisin with the substrate protein, or construct [13]. In such experiments which involve limited protein digestion, Subtilisin A helps to identify regions that are poorly-structured in a significant fraction of the population at any given time, since enough time is not given for the protease to degrade the entire protein population into small peptides. The assumption that is implicitly made in interpreting such experiments is that the nature of the folding equilibrium determines both (a) the initial species formed through *limited* proteolytic digestion prior to complete digestion, and (b) the time required for complete degradation, or *limit* digestion.

Ordinarily, Subtilisin A:protein ratios of 1:1000 are capable of degrading unstructured regions of proteins within minutes (or tens of minutes) of exposure to substrate proteins. In Fig. 3A, we show the behaviour of Coh2 and BSX in isolation, and in each other's presence, on a 12% SDS-PAGE, as controls displaying the expected mobilities of free Coh2 and BSX. In Fig. 3B, we show that virtually the entire population of Coh2-*Flexible*-BSX remains un-degraded even after 2 h of incubation with Subtilisin A, when the ratio of Subtilisin A:Coh2-*Flexible*-BSX happens to be 1:1000, with only hints of free BSX and Coh2 released through proteolysis of the linker being visible in the form of bands. At ratios that are more favourable to proteolysis, substantial fractions of the population can still be seen to have remained un-degraded. Significant degradation is seen only once relative concentrations of Subtilisin A and Coh2-*Flexible*-BSX approach similar or neighbouring orders of magnitude. Furthermore, only when a ratio of Subtilisin A: Coh2-*Flexible*-BSX of 1:1 is used can Coh2-*Flexible*-BSX be observed to have undergone complete proteolytic digestion over the duration of a 2 h incubation, and become invisible as a consequence. As the relative amounts of Subtilisin A are raised to higher and higher levels with respect to the protein substrate, with ratios changing from 1:000 to 1:1, and before complete degradation is observed, some degradation can be seen to have occurred even in the BSX and Coh2 domains, generating some additional gel bands (left unmarked in Fig. 3).

By way of comparison, in Figs. 3C and 3D, we show similar experiments performed with similar concentrations of Coh2, or BSX, and similar molar ratios of Subtilisin A:Protein. These figure panels show that between Subtilisin A:Protein molar ratios of 1:20 and 1:1000, all three protein species, i.e., Coh2, BSX and Coh2-*Flexible*-BSX undergo no significant degradation, with the possible exception of Coh2-*Flexible*-BSX which shows a very minute but detectable amount of proteolytic release of free BSX through the degradation of the *Flexible* linker even in the range of molar ratios between 1:20 and 1:1000. This constitutes a very remarkable resistance to degradation of all species concerned, and suggests that the *Flexible* linker does indeed escape proteolysis almost to the same extent as Coh2 and BSX are themselves resistant to proteolysis in the Coh2-*Flexible*-BSX construct, with this now being verified not just through the linker’s remarkable ability to escape proteolysis during storage but also upon exposure to Subtilisin A.
What remains to be examined is how the other four constructs behave upon exposure to Subtilisin A, in comparison with the Coh2-Flexible-BSX construct (marked as Flexible in all figures), to see whether the results are similar to what is seen with storage at 4 °C. In Figs. 4A-D, we present experiments comparing the degradation of all five constructs by Subtilisin A in the following manner. Figure 4A presents results of a 2 h incubation in which the ratio of Subtilisin A:Protein was 1:1000, with a total protein (fusion protein construct) loading in each lane of 3.07 µg (or 16 µl of a 3.18 µM protein reaction solution in which Subtilisin was separately present). Maintaining the same total protein loading, Fig. 4B presents results of a reaction in which the severity of the proteolysis was raised 2-fold, i.e., a using 2 h incubation, but with a Subtilisin A:Protein ratio of 1:500, in which twice the amount of subtilisin is available to cut the protein construct. In Fig. 4C, the severity is raised by a further 2-fold with respect to Fig. 4B, i.e., by increasing the duration from 2 h to 8 h. Finally, in Fig. 4D, the severity is raised by yet another 2-fold with respect to Fig. 4C, i.e., with the ratio being increased to 1:500. In other words, going from Fig. 4A to Fig. 4D, through Figs. 4B and Fig. 4C, there is a progressive 2-fold increase of severity of proteolysis from panel to panel, amounting to an 8-fold increase in severity overall.

What is observed in Fig. 4 is once again that the construct incorporating Flexible displays the maximal escape from proteolysis, surviving the 8-fold increase in severity of proteolysis from Fig. 4A to Fig. 4D, in keeping with anticipations raised by the data in Fig. 3B. In comparison, all other constructs can be seen to undergo progressively increasing amounts of limited proteolysis as the severity of the reaction is increased 8-fold from the results presented in Fig. 4A to Fig. 4D. Even the construct incorporating Nat-Quarter, which appears to be the second-best in escaping proteolysis, after the construct incorporating Flexible, undergoes significant proteolytic degradation in the 8 h reaction using a ratio of 1:500, as shown in Fig. 4D. However, in comparison, the construct incorporating Flexible remains largely un-degraded demonstrating that it maximally escapes proteolysis by Subtilisin A amongst all constructs, just as it most escapes proteolysis by trace amounts of proteases present in solution during storage at 4 °C.

**Coh2 and BSX remain fully and independently functional in the fusion construct Coh2- Flexible- BSX.**

This section presents control experiments to demonstrate that the non-interacting BSX and Coh2 domains in the Coh2-Flexible-BSX fusion construct remain not just folded but also functional, as a consequence of remaining folded. In Fig. 5A, we present data on the relative activities of BSX as (i) a lone enzyme in solution, and compare this the activity that is observed (ii) when BSX is incubated with Coh2, and (iii) when BSX exists in genetic fusion with Coh2 in the construct that we call Coh2-Flexible-BSX, which is sometimes also simply referred to as Flexible. As a control, an enzyme activity assay was also performed with Coh2 alone, and confirmed to show no activity. The data in Fig. 5A establishes that BSX is as active by itself as it is active when it is in the presence of Coh2 in solution, or in genetic fusion with Coh2. This establishes not so much that Coh2 and BSX do not interact with each other (which is something that we have already shown in a previous section) but that they do not negatively influence each other when exist upon the same polypeptide chain.
In Fig. 5B, we establish that just as the Coh2-\textit{Flexible}-BSX fusion construct supports xylanase activity as demonstrated in Fig. 5A, the fusion construct also supports the function of a folded form of Coh2. The function of Coh2, which is a cohesin domain, is to bind to proteins/enzymes that host a dockerin domain in their sequence. One such dockerin-hosting multi-domain protein is the cellulose, CelF, from \textit{Clostridium thermocellum}. Since CelF was made in our lab for some other purpose, and readily available, we examined whether CelF is able to bind to Coh2-\textit{Flexible}-BSX through cohesin-dockerin interactions. The results are shown in Fig. 5B. We used gel filtration chromatography on a Superdex-200 Increase (GE) column, to examine the behaviour of CelF, Coh2-\textit{Flexible}-BSX, and a mixture of CelF and Coh2-\textit{Flexible}-BSX and see whether the Coh2 domain of the Coh2-\textit{Flexible}-BSX fusion construct is capable of interacting with CelF which contains a dockerin domain, to give rise to a larger complex that elutes at a smaller elution volume. As Fig. 5B demonstrates, CelF was found to elute at 13.89 ml, while the Coh2-\textit{Flexible}-BSX fusion construct was found to elute at 14.301 ml, being smaller than CelF (which has a size of 80 kDa). Upon mixing, Coh2-\textit{Flexible}-BSX and CelF bind to each other to yield a complex that elutes at the much smaller elution volume of 12.275 ml, in addition to an unbound population of CelF, thus proving that Coh2 in Coh2-\textit{Flexible}-BSX remains competent to bind to the dockerin domain of CelF. Details of the dockerin domain and its sequence can be seen in the Supplementary Information file.

\textbf{Discussion and Conclusions}

An unstructured linker like \textit{Flexible} could not possibly resist proteolysis without assistance from its flanking domains. Three kinds of assistance could be conceived to owe to steric hindrance offered by Coh2 and/or BSX to the approach of proteases, and owe to any one or more of the following: (a) the sizes of Coh2 and BSX (i.e., their hydrodynamic volumes); (b) the distance separating Coh2 and BSX along the chain (i.e., the linker length); and (c) the motions of Coh2 and BSX, with respect to each other, (i.e., the freedom of movement facilitated by a combination of a flexible linker, and a lack of interaction between Coh2 and BSX). Interestingly, controls addressing the first of the two above possibilities have already been described above, in the results section, in that (i) the identities of the flanking domains and, therefore, the hydrodynamic volumes of the flanking domains, are identical in all five constructs, and (ii) the lengths of the linkers used and, therefore, the physical proximity of Coh2 to BSX, are also nearly identical in at least three of the five constructs [i.e., \textit{Rigid} (15 residues), \textit{Flexible} (10 residues), and \textit{Nat-Quarter} (9 residues)]. Since differential proteolytic susceptibility is still observed between constructs incorporating \textit{Rigid} and \textit{Nat-Quarter}, on the one hand, and the construct incorporating \textit{Flexible}, on the other hand and further, despite their possessing flanking domains with identical sizes, and linkers with similarities lengths, it is clearly not either the hydrodynamic volumes of Coh2 and BSX, or the proximity of Coh2 and BSX, that can be held responsible for \textit{Flexible}’s remarkable escape from proteolysis to the highest degree amongst all used linkers. This suggests that differential susceptibilities owe to differences in the relative motions of Coh2 and BSX. Presumably, \textit{Flexible} facilitates the greatest relative motions of its flanking domains, by dint of its being flexible. It is reasonable to argue that such motions result in larger hydrodynamic volumes, than the actual volumes, in the individual flanking domains. In turn, this could create greater steric hindrance to the approach of proteases. In other words, we would like to
propose that the flanking domains function like ‘fans’ or ‘whisks’ that sterically ‘bat away’ approaching proteases, in the manner in which the flexibility of the chain of a ‘nunchuck’ allows its flanking rods to bat away approaching objects.

Our work thus shows that an unstructured and flexible linker essentially indirectly arranges for its own protection from the action of approaching proteases by imparting greater independence (of motion) to its flanking domains and reciprocally receiving greater steric protection from proteolysis. We have demonstrated both (A) that an unstructured and flexible linker of 10 residues (Flexible, consisting of the glycine- and serine-rich sequence, N-SGGGGSGGGG-C) resists degradation by proteases present in trace amounts in solution, and also (B) to a significant extent by exogenously added Subtilisin A, if the linker is flanked by two well-folded domains that are themselves resistant to proteolysis. We have shown that such a level of resistance is not seen in linkers of comparable size that differ in flexibility (and are more rigid) and presented evidence and arguments indicating that Flexible's escape from proteolysis owes to its facilitation of flanking domains motions (assisting steric inhibition of protease access).

Such evidence and arguments do, of course, satisfactorily explain why Rigid undergoes proteolytic degradation. However, they do not satisfactorily explain why the naturally-occurring linker, Nat-full, and its derivatives, Nat-Half, and Nat-Quarter, display high proteolytic susceptibility, considering that nature uses Nat-Full in the protein, CipA. We suggest that these linkers derived from CipA (in which Nat-full acts as a natural linker separating Coh2 and a CBM domain) are rich in proline and threonine residues that tend to be heavily glycosylated [14], since linkers joining domains in bacterial cellulases are known to be protected from proteolysis through such glycosylation in certain organisms [15]. This is something that does not naturally occur during the production of proteins in E. coli.

In summary, the work presented in this paper demonstrates that the flexibility of the Flexible linker rather counter-intuitively causes the linker to escape proteolysis in a certain structural context, i.e., when this flexible linker is flanked by two protein domains that have no scope whatsoever for interacting with each other. This is because it is this lack of scope for mutual interactions between the flanking domains which ensures they move with respect to each other, and sterically inhibit the approach of proteases. In a different structural context, e.g., in a context in which Flexible acts as a loop between domains that interact with each other, we have no doubt that the outcome could be profoundly different.

Declarations

Acknowledgements

We thank the Department of Biotechnology, Government of India, for a project grant (Grant No: BT/PR/31706/PBD/26/705/2019) funding a Hyperthermophile Enzyme Hydrolase Research Centre (HEHRC) in PG’s lab. SW thanks the University Grants Commission for a doctoral research fellowship.

Conflict of Interest Statement
The authors declare that there are no competing financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

Author Contributions Statement

SW participated in the conception of the work, performed the experiments, analyzed the results, and participated in the writing of the manuscript. PG participated in the conception of the work, interpretation of results and writing of the manuscript.

References


**Figures**
Figure 1

Chromatographic examination of protein construct integrity as a function of storage reveals that *Flexible escapes proteolysis, unlike other linkers*. **A.** Data collected immediately following purification. **B.** Data collected after storage of purified proteins for a period of 40 days at 4 °C. **C.** Profiles of independently chromatographed BSX and Coh2.
Electrophoretic examination of protein construct integrity as a function of storage reveals that Flexible escapes proteolysis, unlike other linkers. A. SDS-PAGE profiles of the five fusion constructs collected immediately after purification. B. SDS-PAGE profiles of the five fusion constructs collected after 15 days of storage at 4 °C. C. Gel filtration chromatographic data for both (i) elution of Coh2, or BSX, loaded
independently upon a ~24 ml Superdex-200 Increase (GE) column and also (ii) elution of Coh2 and BSX mixed together. The inset to Figure 2C represents the results of a native PAGE.

Figure 3

**Subtilisin A-based assay for the proteolytic susceptibility of the Coh2-**Flexible-**BSX construct.** A. Behaviour of Coh2 and BSX in isolation, and in each other’s presence, on a 12 % SDS-PAGE. B. SDS-PAGE
showing species generated from Coh2-**Flexible**-BSX after 2 h of incubation with Subtilisin A, using different subtilisin:Coh2-**Flexible**-BSX ratios. **C.** SDS-PAGE showing species generated from Coh2 after 2 h of incubation with Subtilisin A, using different subtilisin:Coh2 ratios. **D.** SDS-PAGE showing species generated from BSX after 2 h of incubation with Subtilisin A, using different subtilisin:BSX ratios. Note: species of comparable mobility are annotated with the same number, in panels B, C and D, and the legend explaining these numbers is included in a box placed to the right of panel A, in which the symbol ‘+’ placed to the right of any number denotes degradation species derived from the band corresponding to the said number.
Figure 4

**Subtilisin A-based assay for the proteolytic susceptibilities of all Coh2-linker-BSX constructs.** The panels show species generated upon incubation of Subtilisin A:construct, as follows. **A.** 2 h incubation using a Subtilisin A:construct ratio of 1:1000. **B.** 2 h incubation using a Subtilisin A:construct ratio of 1:500. **C.** 8 h incubation using a Subtilisin A:construct ratio of 1:1000. **D.** 8 h incubation using a Subtilisin A:construct ratio of 1:500. Note: There is a progressive 2-fold increase in the severity of conditions used for
proteolysis from Figure 4A to Figure 4B, and also from Figure 4B to Figure 4C, and Figure 4C to Figure 4D, amounting to an 8-fold increase in severity between Figure 4A and Figure 4D.

Figure 5

Evidence of full and independent functionality of Coh2 and BSX in the fusion construct, Coh2-**Flexible-BSX**. A. Relative xylan-degrading activity in BSX alone (marked ‘BSX’), BSX incubated with Coh2 (marked...
‘BSX+Coh2’), BSX in fusion with Coh2 in the construct, Coh2-Flexible-BSX (marked ‘Flexible’), and Coh2 alone (marked ‘Coh2’) of identical molarities (0.5 µM each). B. Gel filtration chromatographic profile examining the elution of CelF (including a Coh2 binding-competent dockerin, marked ‘CelF’), the Flexible fusion construct (marked ‘Coh2-Flexible-BSX’), and a mixture of CelF and the Flexible fusion construct (marked ‘CelF + Coh2-Flexible-BSX’) of identical molarities (5.0 µM each).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- WaghmareandGuptasarmaProteinJSupplementaryFile20623.pdf