Flexible peptide linkers display higher resistance to proteolysis than rigid linkers by facilitating movements that inhibit the approach of proteases

Purnananda Guptasarma[1] and Snehal Waghmare[1]

[1]Indian Institute of Science Education and Research Mohali

January 30, 2023

Abstract

The differential proteolytic susceptibilities of different types of linkers joining the same two protein domains were investigated, to identify linkers displaying the highest resistance to degradation by proteases in the environment. Five linkers, namely *Rigid* [3 repeats of EAAAK], *Flexible* [two repeats of SGGGG], *Nat-Full*, *Nat-Half* and *Nat-Quarter* [3 linkers of 42, 21, and 9 residues, respectively, derived from a naturally-occurring Coh2-adjacent sequence] were used to fuse two thermostable, and proteolytically-resistant, domains, namely Coh2 (from *C. thermocellum* CipA), and BSX (from *Bacillus sp.* NG-27). The produced fusions were examined for linker degradation, using electrophoretic (SDS-PAGE) and chromatographic (size exclusion) experiments, following different periods of storage at 4 0C. *Rigid, Nat-Full, Nat-Half* and *Nat-Quarter* were degraded, and released free Coh2 and BSX. However, *Flexible* displayed a remarkable resistance to both environmental proteases and Subtilisin A (a non-specific protease). Our data suggests that *Flexible* imparts freedom of movement to Coh2 and BSX, and then reciprocally receives protection against proteolysis from Coh2 and BSX, through their dynamic steric inhibition (or 'batting away') of protease approach.

Introduction

Protein engineering experiments involving fusion of different proteins/domains that do not naturally adjoin each other forces biotechnologists to have to choose suitable linker peptides, which can vary in length, complexity, and conformational characteristics. The success or failure of the fusion experiment is determined by the choice of linker peptide, for two primary reasons: (i) the linker can affect the function, conformation, and/or conformational stability of flanking proteins/domains; (ii) the linker can itself be naturally susceptible to undergoing proteolytic degradation by proteases present in the environment, or become susceptible to proteolysis (as a result of the influence of its flanking proteins/domains). In the work presented here, our objective was to focus upon differential proteolytic susceptibilities of different types of linkers, without distractions involving either (a) the effects of linkers upon conformations of flanking domains, or (b) effects of flanking domains upon conformations of the linkers, to the extent possible. Therefore, the approach we adopted was to use thermostable flanking proteins/domains that typically display autonomy of folding, as well as autonomy of conformational stabilization, in addition to high (natural) resistance to proteolysis [which is a natural corollary, because high thermal stability translates into a better-folded state, leading to lower scope for proteases to gain access to peptide bonds (Mukherjee and Guptasarma, 2005)]. In experiments presented below, through the use of thermostable flanking domains, we reduced the likelihood of proteolytic degradation in flanking domains. This allowed us to examine only the linker peptide for proteolytic degradation.

Two thermostable domains, (i) Coh2, a Type I cohesin from *Clostridium thermocellum* CipA, and (ii) BSX, a *Bacillus sp.* Xylanase, were joined. The structure of BSX consists of a single (β/α)8-fold barrel (Manikandan *et al* ., 2006). The structure of Coh2 consists of a nine-stranded barrel with jelly-roll fold topology, consisting

of two flattened β-sheets (Carvalho *et al* ., 2003). Coh2 and BSX were fused with the use of five different linkers: a rigid linker, a flexible linker, and three linkers derived from the sequence of a 42 residues-long (native) linker joining the C-terminal end of Coh2, in *Clostridium thermocellum* CipA, to the N-terminal end of an adjoining carbohydrate binding module (CBM). Below, we show that the linker that is the most structurally-flexible and unstructured is also the linker that maximally escapes proteolysis. A systematic elimination of possibilities indicates that this counter-intuitive result owes to the linker's facilitation of motions in its flanking domains. The motions appear to negatively affect the ability of proteases to approach the linker's (otherwise scissile) peptide bonds.

Materials and Methods

Choice of linkers. *Nat-Quarter* : N- NATPTKGAT-C (9 residues); *Nat-Half* : N-NATPTKGATPTNTATPTKSAT-C (21 residues);*Nat-Full* : N-NATPTKGATPTNTATPTKSATATPTRPSVPTNTPTNTPANTP-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. thermocellulm* 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 was PCR-amplified from *C. thermocellum* genomic DNA (strain ATCC 27405); and (ii) DNA encoding BSX was gotten commercially synthesised. Coh2 and BSX were fused through splicing by overlap extension (SOE) PCR, using appropriately-designed primers incorporating the said linkers. Coh2-linker-BSX-encoding fusions were PCR amplified, digested by NdeI, and XhoI, and ligated into the T7-based expression vector pET23a, to be produced in fusion with a C-terminal 6xHis affinity tag.

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 that were later transformed into *E coli* BL21 Star (DE3) pLysS, towards protein over-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 ($O.D._{600}$) 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 chromatography, using columns from GE Healthcare. Protein yields were typically 4 to 5 mg per litre of culture. All five Coh2-BSX fusion proteins were purified under non-denaturing conditions and transferred into 20 mM Na-HEPES buffer (pH 7.5, containing 100 mM NaCl, and 2 mM $CaCl_2$) 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 set up was also used in analytical mode, to examine elution behaviour of Coh2-linker-BSX fusion proteins incorporating different linkers, following 40 days of storage.

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 Coh2-linker-BSX fusion made using the *Flexible* linker (i.e., Coh2-*Flexible* -BSX), to examine the linker's degree of resistance to proteolysis by exogenously-added Subtilisin A, which is known to hydrolyse peptide bonds without any sequence specificity (Markland, 1971). Coh2-*Flexible*- BSX was incubated with Subtilisin for 2 h at 50 °C, using Subtilisin A:Coh2-*Flexible*-BSX molar ratios varying from 1:1 to 1:1000 (or 0.001:1), using stock solutions of Subtilisin A prepared in 20 mM Na-HEPES buffer (pH 7.5, containing 100 mM NaCl, and 2 mM $CaCl_2$) and stock solutions of Coh2-*Flexible*- BSX as mentioned above. Samples were analysed on 12 % SDS–PAGE to assess proteolytic degradation.

Results and Discussion

Chromatographic examination of protein construct integrity as a function of storage. Figures 1A, and 1B, respectively, show the mutually-normalized analytical chromatographic gel filtration elution profiles of the five Coh2-linker-BSX fusion protein constructs. Profiles shown in Figure 1A were collected immediately after purification, while profiles shown in Figure 1B were collected after storage at 4 °C for 40 days. In Figure 1A, all constructs are seen to elute as monomers (between 9.93 and 10.24 ml), with minor differences in elution volume attributable to known differences in the size(s) of constituent linkers. Briefly, the constructs roughly eluted in the order of reducing (known) linker length, and reducing (anticipated) hydrodynamic volume [*Nat-full* (9.93 ml; 42 residues), *Nat-Quarter* (10.15 ml; 9 residues), *Nat-half* (10.18 ml; 21 residues), *Rigid*(10.20 ml; 15 residues), and *Flexible* (10.24 ml; 10 residues)], with only *Nat-quarter* behaving anomalously. Figure 1B shows that, after 40 days of storage at 4 °C, the single elution peaks of the constructs (between 9.93 and 10.24 ml), degenerated into two new peaks of significance, at ˜11.5 ml, and ˜13.5 ml, with four of the five constructs (incorporating *Nat-Quarter, Nat-half, Nat-full* and *Rigid* linkers); two additional minor peaks at ˜19 and ˜20 ml with two of the five constructs (incorporating *Nat-full* and *Nat-half*); but no peaks whatsoever at ˜13.5, ˜19 or 20 ml, and only an insignificant peak at ˜11.5 ml, with the fifth construct (incorporating *Flexible*). This demonstrates that four constructs underwent significant degradation, whereas the construct incorporating *Flexible* did not undergo degradation. The species eluting at ˜11.5 ml, and ˜13.5 ml, respectively, were confirmed to be BSX, and Coh2, through control experiments. Since BSX and Coh2 are thermostable, they resist proteolysis. Therefore, trace amounts of proteases in the proteins' environment attack only the linker backbones in four of the five Coh2-linker-BSX constructs, to physically separate BSX and Coh2 and cause them to elute independently.

Hosted file

`image1.emf` available at https://authorea.com/users/494044/articles/621762-flexible-peptide-linkers-display-higher-resistance-to-proteolysis-than-rigid-linkers-by-facilitating-movements-that-inhibit-the-approach-of-proteases

Figure 1. Chromatographic examination of protein construct integrity as a function of storage, through-monitoring of gel filtration chromatographic elution profiles. *Panel A.* Gel filtration chromatography profiles of all constructs on the day of purification.*Panel B.* Gel filtration chromatography profiles of all constructs after 40 days of storage in a refrigerator.

Gel filtration chromatographic results do not, however, address the question of whether there have been any additional intra-domain proteolytic cuts within the backbones of BSX and Coh2. To examine whether there are any such cuts, we performed SDS-PAGE experiments (under denaturing conditions), to investigate whether any sections of Coh2, or BSX, had become covalently separated from other sections, to remain remain non-covalently associated during chromatography.

Electrophoretic examination of protein construct integrity as a function of storage. Figure 2A presents SDS-PAGE profiles of constructs collected immediately after purification, demonstrating the absence of degradation. On a 12 % SDS-PAGE, each construct displayed a mobility attributable to the size of the extended (i.e., SDS-bound, and SDS- and heat-unfolded) polypeptide, with the construct incorporating*Nat-full* showing the least mobility due to its possession of the longest linker sequence, and constructs incorporating *Nat-half*and *Rigid* displaying mobilities 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 because they have the smallest linkers. No construct displayed signs of anomalous mobility, or significant degradation. Figure 2B presents the SDS-PAGE profiles collected after 15 days of storage at 4 °C, by which time (despite the anticipated lower activity of proteolytic enzymes naturally present in the environment, at low temperatures; unless derived from psychrophile organisms in refrigerators), sufficient degradation was observed to have already occurred in four constructs. Figure 2B shows that Coh2, and BSX, being well-folded domains with folding equilibria keeping the bulk of their populations folded at any time, largely remain intact and undergo no visible degradation, either while they exist

in fusion, or after their separation through proteolytic scission of the linker, as is evident from the fact that there is no lowering in intensity of bands corresponding to intact Coh2-linker-BSX that are not compensated for by the intensities of bands corresponding to free BSX, or Coh2. No visible bands suggest degradation of BSX and Coh2 and, in each case, in Figure 2B, degradation produces three distinct species: (i) the residual (un-degraded) Coh2-linker-BSX [size > 60 kDa] population; (ii) a population corresponding to BSX [size ˜45 kDa], and (iii) a population corresponding to Coh2, with the linker peptide still present at the C-terminus [size > 20 kDa). Figure 2B thus establishes that linkers undergo proteolysis at their BSX-adjoining ends, in constructs incorporating *Nat-Quarter, Nat-half, Nat-full*and *Rigid* .

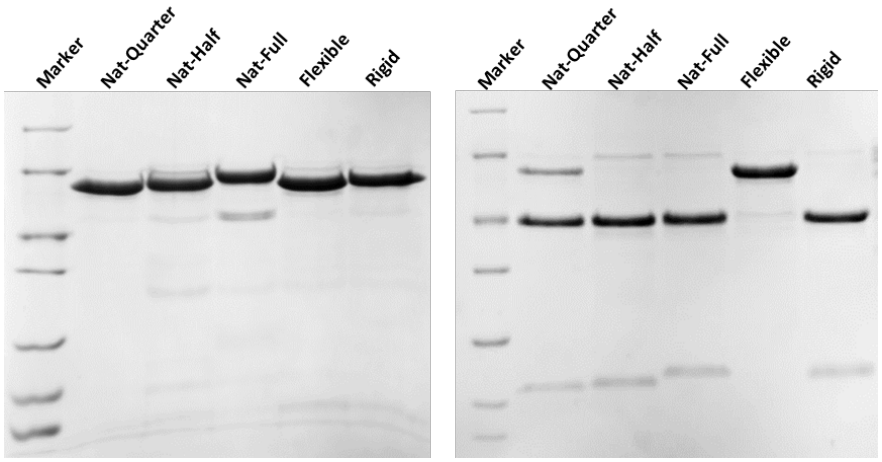

Figure 2. Electrophoretic examination of protein construct integrity as a function of storage, through monitoring by SDS-PAGE.*Panel A.* Behaviour of constructs on the day of purification.*Panel B.* Behaviour of constructs following 15 days of storage in a refrigerator.

Counter-intuitive proteolytic susceptibilities of linker peptides. What is most especially notable about Figure 2B is not the observation that BSX and Coh2 are not vulnerable to degradation by proteases present in the environment. This is expected, since they are thermostable domains. Furthermore, prior work in our lab (data not shown) had caused us to choose these domains for these experiments. What is notable instead is the fact that the Coh2-*Flexible*- BSX construct undergoes almost no degradation. This result is somewhat counter-intuitive. The linker called *Flexible* cannot adopt a particular structure, as it is constituted of two serine residues and eight glycine residues (since glycine possesses only a hydrogen atom as its side chain, and this typically imparts great conformational freedom to regions of proteins chains that contain the residue, glycine).*Flexible* is, in fact, a popular linker used in protein fusions (Argos, 1990; Chen *et al* ., 2013) precisely because its own lack of intrinsic structure prevents it from restricting the motions (and activities) of flanking domains. Our results suggest that*Flexible* undergoes no proteolysis despite remaining unstructured, although it is the least likely of five different linkers to undergo proteolysis, when placed between well-folded domains that themselves resist proteolysis. It must be remarked that, in experiments involving poorly-folded flanking domains, proteolysis can (and does) occur at multiple sites upon storage, with this ordinarily smearing-out any distinctions between 'proteolytic vulnerabilities of linkers' and 'proteolytic vulnerabilities of flanking domains'. Here, having ensured that no proteolysis occurs in flanking domains, we observe that the linker least likely to resist proteolysis is actually the one that maximally resists proteolysis. Furthermore, this result is obtained after ensuring that every parameter remains equal in all conducted experiments, i.e., with linkers possessing identical flanking domains (with identical conformational stabilities), and with fusions subjected to identical physical and chemical conditions for testing, and identical methods of testing, using identical protein concentrations, and identical durations of incubation.

Probing of the proteolytic susceptibility of the 'Flexible' linker through use of Subtilisin A. Before proceeding

to rationalize the counter-intuitive result that an unstructured linker peptide such as *Flexible* undergoes the least proteolysis, we decided to further examine the resistance of *Flexible* to proteolysis in the context of its presence in Coh2-*Flexible*- BSX, by exposing Coh2-*Flexible*- BSX to the non-specifically-acting serine protease, Subtilisin A, using 2 h of incubation and different molar ratios of Subtilisin A:Coh2-*Flexible* -BSX. Subtilisin A is commonly used to test the relative proteolytic vulnerabilities of different regions of a folded protein, as well as to probe for regions that are not as well-folded as other regions, through limited proteolytic digestion performed using varying Subtilisin A:protein ratios, and/or varying durations of incubation (Sharma and Guptasarma, 2008). In such experiments, 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 population into small peptides, the assumption being 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 (or 0.001:1) degrade unstructured regions of proteins within minutes (or tens of minutes) of exposure. Figure 3 shows that virtually the entire population of Coh2-*Flexible* -BSX remains un-degraded even after 2 h of incubation with Subtilisin A, using a Subtilisin A:Coh2-*Flexible* -BSX ratio of 0.001:1, with only hints of separated BSX and Coh2 bands being visible. At ratios that are more favourable to proteolysis, substantial fractions of the population can still be seen to remain un-degraded, and significant degradation is seen only when relative concentrations of Subtilisin A and Coh2-*Flexible* -BSX approach the same order of magnitude. Only when a ratio of 1:1 is used is Coh2-*Flexible* -BSX observed to have been completely destroyed. As the relative amounts of Subtilisin A rise to high levels, and before complete degradation is observed, some degradation can be seen to have occurred even in the BSX and Coh2 domains, with this generating some additional gel bands (left unmarked in Figure 3).

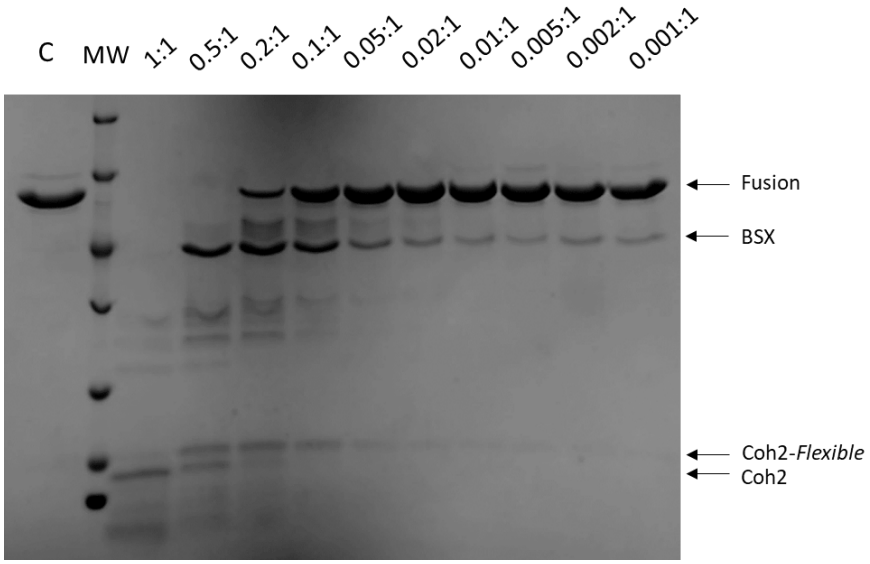

Figure 3. Electrophoretic examination of protein construct integrity as a function of exposure of Coh2-*Flexible*- BSX (3.4 μM) to Subtilisin A (varying concentrations annotated above the figure), using SDS-PAGE. The lane marked 'C' shows the Coh2-Flexible-BSX construct that was not exposed to Subtilisin A. The lane marked 'MW' shows molecular weight markers of 116, 66.2, 45, 35, 25, 18.4 and 14.4 kDa, respectively (top to bottom).

Understanding the counter-intuitive resistance to proteolysis of the 'Flexible' linker. An unstructured linker like *Flexible*could not possibly resist proteolysis without assistance from its flanking domains. Three kinds of assistance may be conceived to arise from steric hindrance offered by Coh2 and/or BSX to the approach

of proteases, deriving from: (a) the sizes (or hydrodynamic volumes) of Coh2 and BSX; (b) the distance separating Coh2 and BSX; and (c) the motions of Coh2 and BSX, with respect to each other. Control experiments addressing the first two possibilities have already been described above, in that the sizes of the flanking domains are identical in all five constructs, and in that the physical proximity of Coh2 to BSX is also nearly identical in at least three constructs [*Rigid* (15 residues), *Flexible* (10 residues), and *Nat-Quarter* (9 residues)]; however, with proteolysis still seen to occur differentially in other constructs and the construct incorporating*Flexible* . This establishes that it is neither the sizes (volumes) of Coh2 and BSX, nor the proximity of Coh2 and BSX, which are alone responsible for *Flexible* 's remarkable resistance of proteolysis. This leaves only the third possibility, namely that differential susceptibilities arise from differences in the motions of Coh2 and BSX.*Flexible* presumably facilitates the greatest relative motions of its flanking domains, because it is flexible. It is not unreasonable to argue that such motions could serve to create large apparent hydrodynamic volumes of flanking domains (i.e., much bigger hydrodynamic volumes than actual volumes). This, in turn, could effectively create greater steric hindrance to the approach of proteases. In other words, the flanking domains probably function like 'fans' or 'whisks' that 'bat-away' any approaching proteases. Our work shows that an unstructured and flexible linker can arrange for its protection from proteases by imparting greater independence of motion to its flanking domains, and by receiving, in return, protection from proteolysis by approaching proteases.

Conclusion

We have shown 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 to a significant extent by exogenously added Subtilisin A, when it is flanked by two well-folded domains that are themselves also resistant to proteolysis. In addition, we have presented evidence indicating that the resistance of *Flexible* to proteolysis owes to its facilitation of motions by its flanking domains which presumably sterically inhibit the access of proteases to the linker. This argument suitably explains why *Rigid* undergoes proteolytic degradation, but not why the naturally-occurring linker,*Nat-full* , displays high proteolytic susceptibility. We suggest that linkers in CipA (from which *Nat-full* is derived) are rich in proline and threonine residues that tend to be heavily glycosylated (Gerwig *et al* ., 1993), since linkers joining domains in bacterial cellulases are known to be protected from proteolysis by glycosylation (Langsford et el., 1987), which does not occur during production in*E. coli* .

Declaration

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.

References

Argos, P. (1990). An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *J. Mol. Biol.* **211** , 943-958.

Carvalho, A. L., Dias, F. M., Prates, J. A., Nagy, T., Gilbert, H. J., Davies, G. J., ... & Fontes, C. M. (2003). Cellulosome assembly revealed by the crystal structure of the cohesin–dockerin complex. *Proc. Natl. Acad. Sci.* **100** , 13809-13814.

Chen, X., Zaro, J. L., & Shen, W. C. (2013). Fusion protein linkers: property, design and functionality. *Adv. Drug Deliv. Rev.* **65** , 1357-1369.

Gerwig, G. J., Kamerling, J. P., Vliegenthart, J. F., Morag, E., Lamed, R., & Bayer, E. A. (1993). The nature of the carbohydrate-peptide linkage region in glycoproteins from the cellulosomes of*Clostridium thermocellum* and *Bacteroides cellulosolvens* .*J. Biol. Chem.* **268** , 26956-26960.

Langsford, M. L., Gilkes, N. R., Singh, B., Moser, B., Miller, R. C., Warren, R. A. J., & Kilburn, D. G. (1987). Glycosylation of bacterial cellulases prevents proteolytic cleavage between functional domains. *FEBS Lett.* **225** , 163-167.

Manikandan, K., Bhardwaj, A., Gupta, N., Lokanath, N.K., Ghosh, A., Reddy, V.S., Ramakumar, S (2006). Crystal structures of native and xylosaccharide-bound alkali thermostable xylanase from an alkalophilic Bacillus sp. NG-27: structural insights into alkalophilicity and implications for adaptation to polyextreme conditions, *Protein Sci.* **15** , 1951–1960.

Mukherjee, S., & Guptasarma, P. (2005). Direct proteolysis-based purification of an overexpressed hyperthermophile protein from Escherichia coli lysate: a novel exploitation of the link between structural stability and proteolytic resistance. *Protein Exp. Purif.* **40** , 71-76.

Sharma, S. & Guptasarma, P. (2008). Evidence of native-like substructure (s) in polypeptide chains of carbonic anhydrase deposited into insoluble aggregates during thermal unfolding. *Protein J.***27** , 50-58.