Molecular BioSystems

Backbone circularization of Bacillus subtilis family 11 xylanase increases its thermostability and its resistance against aggregation

Journal:	*Molecular BioSystems*
Manuscript ID	MB-ART-05-2015-000341.R2
Article Type:	Paper
Date Submitted by the Author:	21-Sep-2015

Complete List of Authors:	Di Ventura, Barbara; University of Heidelberg, Theoretical Bioinformatics Waldhauer, Max; University of Heidelberg, Schmitz, Silvan; University of Heidelberg, Ahlmann-Eltze, Constantin; University of Heidelberg, Gleixner, Jan; University of Heidelberg, Schmelas, Carolin; University of Heidelberg, Huhn, Anna; University of Heidelberg, Bunne, Charlotte; University of Heidelberg, Buescher, Magdalena; University of Heidelberg, Horn, Max; University of Heidelberg, Klughammer, Nils; University of Heidelberg, Kreft, Jakob; University of Heidelberg, Schaefer, Elisabeth; University of Heidelberg, Bayer, Philipp; University of Heidelberg, Kraemer, Stephen; German Cancer Research Center (DKFZ), Division of Theoretical Bioinformatics Neugebauer, Julia; German Cancer Research Center (DKFZ), Division of Theoretical Bioinformatics Wehler, Pierre; University of Heidelberg, Mayer, Matthias; Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH), DKFZ-ZMBH-Alliance, Eils, Roland; University of Heidelberg, ; German Cancer Research Center (DKFZ), Division of Theoretical Bioinformatics

Backbone circularization of *Bacillus subtilis* family 11 xylanase increases its thermostability and its resistance against aggregation

Max C. Waldhauer[a*], Silvan N. Schmitz[a*], Constantin Ahlmann-Eltze[a], Jan G. Gleixner[a], Carolin C. Schmelas[a], Anna G. Huhn[a], Charlotte Bunne[a], Magdalena Büscher[a], Max Horn[a], Nils Klughammer[a], Jakob Kreft[a], Elisabeth Schäfer[a], Philipp A. Bayer[a], Stephen G. Krämer[b], Julia Neugebauer[b], Pierre Wehler[a], Matthias P. Mayer[c], Roland Eils[abd] and Barbara Di Ventura[a#]

[a] Center for Quantitative Analysis of Molecular and Cellular Biosystems (BioQuant), University of Heidelberg, Germany

[b] German Cancer Research Center (DKFZ), Division of Theoretical Bioinformatics, Heidelberg, Germany

[c] Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH), DKFZ-ZMBH-Alliance, Heidelberg, Germany

[d] Institute for Pharmacy and Molecular Biotechnology (IPMB), Division of Bioinformatics and Functional Genomics, Heidelberg, Germany

[*] These authors contributed equally to this work

[†] The authors are the iGEM team Heidelberg 2014

[#] to whom correspondence should be addressed: barbara.diventura@bioquant.uni-heidelberg.de

Abstract

The activity of proteins is dictated by their three-dimensional structure, the native state, and is influenced by their ability to remain in or return to the folded native state under physiological conditions. Backbone circularization is thought to increase

protein stability by decreasing the conformational entropy in the unfolded state. A positive effect of circularization on stability has been shown for several proteins. Here, we report the development of a cloning standard that facilitates implementing the SICLOPPS technology to circularize proteins of interest using split inteins. To exemplify the usage of the cloning standard we constructed two circularization vectors based on the *Npu* DnaE and gp41-1 split inteins, respectively. We use these vectors to overexpress in *Escherichia coli* circular forms of the *Bacillus subtilis* enzyme family 11 xylanase that differ in the identity and number of additional amino acids used for circularization (exteins). We found that the variant circularized with only one additional serine has increased thermostability of 7 °C compared to native xylanase. The variant circularized with six additional amino acids has only a mild increase in thermostability compared to the corresponding exteins-bearing linear xylanase, but is less stable than native xylanase. However, this circular xylanase retains more than 50 % of its activity after heat shock at elevated temperatures, while native xylanase and the corresponding exteins-bearing linear xylanase are largely inactivated. We correlate this residual activity to the fewer protein aggregates found in the test tubes of circular xylanase after heat shock, suggesting that circularization protects the protein from aggregation under these conditions. Taken together, these data indicate that backbone circularization has a positive effect on xylanase and can lead to increased thermostability provided the appropriate exteins are selected. We believe that our cloning standard and circularization vectors will facilitate testing the effects of circularization on other proteins.

Introduction

In order to be used in biotechnological and medical applications enzymes must often withstand high temperatures, extreme pHs, transport and conservation and be resistant to detergents and organic solvents. Thermophiles represent a natural source of enzymes with many of these important features.[1] Yet, finding the appropriate thermophilic counterpart of a given enzyme might be tedious and obtaining large amounts of it might be difficult, because genes encoded by thermophiles cannot always be overexpressed in mesophilic hosts.[2-4] Moreover, since hyperthermophilic enzymes have evolved to function at high temperatures, they often show reduced activity at lower temperatures.[5, 6] These inconveniences can be avoided by directly

engineering mesophilic enzymes to become highly thermostable. To this aim, several approaches are available such as directed evolution, rational design principles like the introduction of disulfide bonds and the consensus concept based on sequence homology.[7] The rational engineering approaches are challenging because of the lack of a single mechanism responsible for empowering proteins with high thermal stability. Nonetheless, a rational design principle that seems to be generally applicable is the docking of N- and C-termini and the anchoring of loose ends. Indeed, loops and termini are the parts of a protein that unfold first during thermal denaturation, since these are the regions with the highest temperature factors in a protein structure.[8] Structural studies of thermophilic proteins showed that loops are stabilized by either being shortened via formation or extension of secondary structures or better anchored to the rest of the protein via ion pairing, hydrogen bonds or hydrophobic interactions.[9, 10] N- and C-termini are also seen to interact for mutual stabilization,[10] or to form bonds with neighboring residues that anchor them better to the protein core.[11] Natural proteins for which N- and C-termini are connected via a peptide bond, the so-called circular proteins, are a beautiful manifestation of this principle because they are particularly thermostable even when of mesophilic origin.[12-14] Moreover, due to the absence of free amino and carboxy-terminal regions, circular proteins are protected from the action of exopeptidases.[15]

This has led to the idea of circularizing the backbone of proteins of interest as a mean to enhance their stability either due to a decreased conformational entropy of the unfolded state, or to a more rapid folding because of a reduced number of folding pathways or due to resistance to exopeptidases.[16] Circularization is likely to not interfere with protein structure and function,[15, 17, 18] unless protein activity is dependent on the termini themselves or conformational strain is introduced in the circularization process. Surely, circularization of proteins whose termini are close in space is easier to achieve using short linkers. Actually, proteins often have their N- and C-termini close together and surface exposed,[19] therefore circularization should be feasible in many cases.

The advent of the split intein technology[20] has allowed simplifying the circularization protocol, going from the more laborious *in vitro* chemical ligation[21] procedure to the *in vivo* circularization via simple expression of a modified gene into a bacterial host.[18] Inteins are autocatalytic protein domains that excise themselves out of proteins and in

doing so connect the flanking regions with a peptide bond which is indistinguishable from a bond made by the ribosome (Fig. S1a).[22-25] The splicing reaction requires the presence of specific amino acids that are conserved in all inteins, such as a cysteine or serine at the N-terminus and an asparagine at the C-terminus.[23] Split inteins are constituted by two separate domains that need to associate to reconstitute a functional intein.[23, 26-28] Therefore, during the process of splicing, split inteins make fusions between two previously independent proteins (Fig. S1b). Split inteins can be used to modify or label proteins post-translationally, allowing the regulation of processes in cells[29, 30] or whole organisms.[31] Importantly, some inteins can be controlled by external signals, such as temperature,[31, 32] small ligands[29, 30, 33, 34] and light.[35-37]

Taking advantage of the mechanism of split intein splicing, Scott *et al.* devised SICLOPPS, a strategy for generating circular proteins *in vivo*.[18] In this design, the gene encoding the C-terminal split intein is fused to the beginning of the gene encoding the protein of interest, while the gene encoding the N-terminal split intein is fused to its end (Fig. 1). This leads to the expression in the cells of a fusion protein that contains the target protein flanked by the split inteins. After binding to each other, the split inteins excise themselves out and connect the N- and C-termini of the same protein, generating a circular protein. It is important to note that for most inteins the splicing efficiency is highly dependent on the identity of the extein residues found at the splice junctions,[23] to which we refer here as "exteins" for simplicity. Moreover, in order for the splicing to take place, the first residue of the C-extein ought to be a cysteine, serine or threonine.[23] When circularizing a protein of interest, the exteins must be included in the design and remain in the final product as a scar (Fig. 1). Therefore, "native" and "linear" proteins do not coincide: while clearly the native protein is also linear, the term linear protein most often refers to either a linearized version of the circular protein obtained by digestion or to the protein that contains the exteins but that was never circularized.

The SICLOPPS technology has been successfully used to generate circular peptide libraries[38-41] and several circular proteins.[16, 18, 42-44] In these latter cases, the corresponding circularization constructs have been created using classical type II restriction enzymes that leave scars in the circularized proteins. In the majority of the cases, extra amino acids ought to be added for achieving fast and efficient intein

splicing, therefore it is possible to select appropriate restriction enzymes for which the scars coincide with the desired exteins. Yet, this strategy leads to the addition of fixed extra amino acids. However, it would be advantageous to have the possibility to select the identity and number of the exteins for any gene of interest, without laborious and tedious search for the right restriction enzymes for each case.

Here we aimed at simplifying and generalizing the *in vivo* SICLOPPS circularization method by creating a scarless cloning standard that could be compatible with any gene of interest and almost all split inteins. To exemplify the use of the cloning standard we constructed two general-purpose circularization vectors, based on the very efficient split inteins *Npu* (*Nostoc punctiforme* PCC73102) DnaE[45-47] and gp41-1.[48] These vectors are general-purpose vectors because cloning of any gene of interest is possible thanks to the cloning standard we devised. We use these vectors to circularize family 11 xylanase from *B. subtilis* in *E. coli* cells. Xylanases are enzymes used in the pulp, paper and food industries because they degrade xylan, the main component of plant hemicellulose. Because the industrial processes in which they are used require elevated temperatures, there is a high need for thermostable xylanases. Several improved engineered or natural (thermophilic) xylanases have been reported.[49-51] Backbone circularization of xylanase as a means to increase its thermostability has not been attempted so far. Therefore, we wanted to study the effects that circularization might have on this protein. We could circularize three variants of xylanase differing in the exteins added to achieve circularization. We found that circularization effectively increases the thermostability of xylanase, provided only a serine is added to the protein. Moreover, this circular variant shows improved resistance against aggregation. This holds true also for the circular variant bearing six additional amino acids, which is not improved compared to native xylanase in terms of thermostability.

Results

Development of a scarless cloning standard for the *in vivo* circularization of a protein of interest using the SICLOPPS technology

Often the best working split intein/exteins combination has to be experimentally determined for each protein of interest, leading to a potentially large number of

intein/exteins combination constructs to clone. We aimed at developing an easy cloning strategy to carry out SICLOPPS-based protein or peptide circularization that would not require any tailor-made adjustment for each target to circularize. To this aim we aligned the protein sequences of all inteins that we found in the InBase database[52] to extrapolate standard overhangs that could be compatible with most inteins. This is possible due to the highly conserved nature of the intein termini: 79 % of the unique sequences of N- and continuous inteins in InBase begin with a cysteine, and the second position almost always features a hydrophobic amino acid. Sequence conservation is even more pronounced at the C-terminus of inteins, where 94 % of the unique C- and continuous inteins end on asparagine, and the penultimate amino acid is a histidine in most cases. Importantly, we based our cloning standard on type IIS restriction sites that allow scarless Golden Gate assembly.[53] This is an essential feature that gives flexibility in defining the exteins to be used for circularization. We chose to use enzymes that create 4-base pair overhangs, which are commonly used for Golden Gate assembly, and positioned their cleavage sites such that the overhangs coincide with the first and last four amino acids of the intein sequence. This fixes the terminal amino acids to cysteine and asparagine, respectively, but also restricts the possibilities for the second and second-to-last positions. Nevertheless, due to the aforementioned sequence conservation, we were able to identify 4-base pair motifs which match 54 % and 92 % of all intein N- and C-termini, respectively, and 49 % of all whole intein sequences (Fig. 2a). If one allows the amino acids in the first and last two positions of the intein to be substituted by chemically similar ones, this number is raised to 82 %. It should then, however, be determined whether these mutations have adverse effects on the splicing reaction.

Among the naturally split inteins listed in InBase, which are particularly interesting for a SICLOPPS application, 86 % strictly conform to the standard overhangs, while 91 % satisfy the requirements when allowing mutations. This property makes the cloning standard especially suited for general-purpose protein circularization vectors, although it should be noted that also continuous inteins may be artificially split.[54]

An extended version of this standard, which also considers other applications of inteins, is described in BBF RFC 105 (http://hdl.handle.net/1721.1/96071).

Construction of two circularization vectors based on the *Npu* DnaE and gp41-1 split inteins

To facilitate the future cloning of circularization constructs, we created two general-purpose circularization vectors based, on the *Npu* DnaE (pCIRC$_{Npu}$) and gp41-1 (pCIRC$_{gp41\text{-}1}$) split inteins, respectively (Fig. 2b). We chose these two split inteins because they are among the most efficient ones known to date.[45, 46, 48] gp41-1 is, moreover, known to tolerate a wider range of exteins than *Npu* DnaE.[48]

pCIRC$_{Npu}$ and pCIRC$_{gp41\text{-}1}$ follow our cloning standard and the SICLOPPS technology for circularization. They both contain a T7 promoter, a strong ribosome binding site (RBS), a kanamycin resistance marker, and the monomeric red fluorescent protein (mRFP) expression cassette as insert to be replaced by the gene of interest, which allows for easy screening of positive clones (Fig. 2b).

Using the circularization vectors to produce circular xylanases in *E. coli*

We analyzed the published three-dimensional structure of xylanase (PDB 1XXN) to estimate the distance between its termini. We hypothesized that one additional amino acid is sufficient to connect the termini without necessarily introducing strain. gp41-1 has been shown to splice efficiently in the presence of a serine or cysteine as C-terminal extein.[48] *Npu* DnaE, instead, is most efficient in combination with the non-native RGKCWE extein sequence.[55] In order to increase the chances of obtaining the three circular xylanase variants, we opted for trying the three exteins (S, C or RGKCWE) with gp41-1 and two exteins (RGKCWE and C) with *Npu* DnaE.

Given our cloning standard, for the construction of the different circular variants we had to create only three PCR products consisting of the amplified gene encoding xylanase (gene name: xynA BSU18840; uniprot ID P18429 – XYNA_BACSU) from the *B. subtilis* genome encoding the appropriate exteins in the primers. These products are then compatible with both pCIRC$_{Npu}$ and pCIRC$_{gp41\text{-}1}$. Note that we eliminated the N-terminal secretion signal peptide of xylanase. Indeed, while having the protein secreted into the medium can be advantageous and might make a purification step dispensable, we reasoned that secretion of the circular form might be problematic if not impossible. In order to treat all proteins in the same way, we decided to express all forms in the cytoplasm of the cells and proceed with the purification step. The PCR products were cloned into the circularization vectors pCIRC$_{Npu}$ and pCIRC$_{gp41\text{-}1}$ giving rise to the expression plasmids pCIRC$_{Npu}$-^{C}xynA, pCIRC$_{Npu}$-^{CWE}xynA^{RGK}, pCIRC$_{gp41\text{-}1}$-^{C}xynA, pCIRC$_{gp41\text{-}1}$-^{S}xynA, and pCIRC$_{gp41\text{-}1}$-^{CWE}xynA^{RGK} (Fig. 3). A

PCR product of the native xynA gene was cloned into unmodified pSB1K30 giving rise to the expression plasmid for native xylanase pNAT-xynA (Fig. 3). This latter construct was further modified by adding the extein coding sequences to either the 5' or the 3' end of the xylanase coding sequence giving rise to the expression plasmids for linear xylanases pLIN-^{C}xynA, pLIN-xynA^{C}, pLIN-^{S}xynA, pLIN-xynA^{S}, pLIN-xynA^{RGKCWE} (Fig. 3). Moreover, to investigate how important the termini of native xylanase are for the overall stability of the protein, the expression plasmids pLIN-$^{\Delta}$xynA and pLIN-xynA$^{\Delta}$ for linear xylanases with a deletion of either the C-terminal or N-terminal amino acid were created.

Since it has previously been shown that circular proteins have a higher electrophoretic mobility than linear ones on SDS-PAGE gels,[15] we analyzed the crude lysates to see whether circularization occurred. We found that all combinations worked but *Npu* DnaE/C (Fig. S2). We then proceeded with the production of three circular xylanases, plus native and linear controls. *E. coli* BL21(DE3) cells transformed with the corresponding constructs were grown in an auto-induction medium[56] to overexpress the proteins, which were then purified as described in the Experimental section.

The difference in mobility between circular and linear xylanases was again assessed by SDS-PAGE, this time using purified proteins (Fig. 4a). To confirm that the species with higher mobility represent indeed circular xylanases, we performed MS/MS sequencing of fragments generated by digestion of the circular RGKCWE xylanase variant with the endoproteinase GluC. We identified one such fragment ending with the sequence RGKCWE (Fig. 4b), a sequence not genetically encoded in that order in pCIRC-^{CWE}xynA^{RGK}, but consistent with the backbone-circularized xylanase. The low abundance linear by-product (Fig. 4a, faint band at the same size as linear xylanase) that results from the spontaneous cleavage reaction of both the C- and N- inteins, where no new peptide bond is formed, was used as a control. While traces of the fragment indicating circularization could be detected in this sample, another fragment ending with RGK only was present in far greater amounts (Fig. 4c).

Circularization increases the thermostability of xylanase

We then used differential scanning fluorimetry (DSF) to estimate the melting temperatures (T_m) of all the purified xylanase variants (Table 1). Interestingly, we found that the exteins strongly modulate the effects that backbone circularization has on the thermostability of xylanase: only the circular S variant was – as expected for a

circular protein – more thermostable than native xylanase (T_m = 63.8 °C versus T_m = 56.9 °C, respectively). Notably, circularization either increased (S and RGKCWE variants) or left unchanged (C variant) the T_m compared to the respective linear forms (Table 1). Yet, addition of the exteins to the linear forms decreased the T_m compared to native xylanase in all cases. Especially addition of RGKCWE to the C-terminus led to a considerable destabilization of the protein. This was also seen when deleting the C-terminal tryptophan. In contrast, deletion of the N-terminal alanine did not affect the stability.

We further measured the activities at different temperatures of two forms (one linear and circular) for the S and RGKCWE variants plus native xylanase and derived the temperature optimum for catalysis (Fig. 5 and Fig. S3. These values were in good accordance with the T_m (Table 1) and were consistent with previously published results for native *B. subtilis* xylanase.[57]

Effect of circularization on the aggregation behavior of xylanase

Since the circular RGKCWE variant was not more thermostable than native xylanase we asked whether circularization could equip this protein with tolerance against heat shock. Therefore, we exposed this circular variant and the corresponding native and linear controls to various temperatures for ten minutes followed by rapid cool down to 37 ºC and measured their activity at 37 ºC (Fig. 6a and b). We found that, under these experimental conditions, circular xylanase bearing the RGKCWE sequence retained significant activity after being heated to temperatures from 70 ºC to 99 ºC compared to native and linear proteins. Beyond this important effect of circularization on the residual activity of xylanase we noted that all proteins had minimal residual activity after heat shock at temperatures slightly above the T_m value, followed by an increase in residual activity for heat shock at higher temperatures (Fig. 6a and b).

To understand if the residual activity was dictated by aggregation caused by heat-induced protein denaturation, we conducted a classical supernatant/pellet assay[58] in which soluble material is separated from insoluble, aggregated material via centrifugation at $18{,}000 \times g$. We found a good qualitative accordance between the results obtained with the residual activity and aggregation assays: for all proteins there was a maximum of aggregates for a temperature around the T_m, and appearance of some soluble material above this critical temperature (Fig. 6c and d). The amount

of soluble material was particularly high for circular xylanase, explaining why for this variant we measured the highest residual activity in this temperature interval (Fig. 6a and b).

In order to find an explanation for the low amount of aggregation after heat shock at 70 °C to 99 °C, we tested additional heat shock conditions for circular xylanase (Fig. 6e). As observed in the previous assay, incubation at 60 °C for ten minutes followed by rapid cool down in icy water caused complete aggregation, while after incubation at 90 °C about a quarter of the protein remained soluble. Interestingly, incubation at 60 °C for ten minutes followed by incubation at 90 °C for ten minutes and rapid cool down led to results similar to 90 °C incubation only. Moreover, doubling the incubation time from ten to twenty minutes did not cause much more aggregation at 90 °C.

We performed the aggregation assay also for the circular and linear S variants and found that also the circular S variant is more soluble than its linear counterpart after heat shock at high temperatures (Fig. 6f), albeit this effect is less pronounced than for circular versus linear RGKCWE variants.

Discussion

In this study we circularized *B. subtilis* family 11 xylanase using only a cysteine, only a serine, or the RGKCWE peptide as linkers. Interestingly, thermostability was increased only for the protein circularized with a serine. These results show that the selection of the linker has a strong influence on the overall stability of the protein. According to basic polymer theory, cross-links in a polymer chain lead to a decrease in the entropy of the unfolded state, thus stabilizing the folded state.[59] Therefore, circularization should always lead to an increase in thermostability, provided no conformational strain is introduced. This is exemplified by the circular C and S variants: albeit differing from each other by a single atom, only the S variant is more thermostable than native xylanase, suggesting that adding a cysteine results in conformational strain in the circular protein. Moreover, changes in the amino acid sequence of a protein, which are necessary for cross-link formation, can also affect its stability. The linear RGKCWE variant was indeed significantly less thermostable than native xylanase, suggesting that addition of the RGKCWE peptide to the C-terminus had itself a destabilizing effect. This is likely due to a disturbance of the local

environment of the C-terminus. In line with these results, we found that eliminating the C-terminal tryptophan led to a dramatic decrease in thermostability of xylanase (Table 1). Interestingly, the circular RGKCWE variant was more thermostable than its linear counterpart, consistent with the predicted decrease in conformational entropy of the unfolded state resulting from circularization. However, circularization cannot compensate for the destabilization of the protein caused by the addition of these six amino acids.

The importance of the termini for the stability of proteins is well known. Backbone circularization involves modification of the termini. Often additional amino acids need to be introduced to have efficient intein splicing; moreover, sometimes a linker covering the distance between the termini should be introduced in order to avoid conformational strain. Since interactions of such linkers with the rest of the protein are difficult to assess and usually not considered, they may introduce more flexibility and could be an initiation site for unfolding and therefore counterbalance the increase in stability gained by circularization. As it is not easy to predict the impact on the protein of this modification in combination with the circularization itself, it is highly advisable to test several constructs, as done here for xylanase. With the two vectors presented in this study, obtaining an expression construct for a particular protein requires only a single cloning step. Indeed, so far the SICLOPPS technology has been more offering a guideline on how to achieve *in vivo* circularization of a protein or peptide of interest starting from any expression plasmid.[60] This is reflected by the fact that the reports on proteins circularized using the SICLOPPS technology so far used each a personalized plasmid.[16, 18, 42-44] While being able to create a circularization vector out of any vector available in the laboratory is surely an advantage, the need to get the DNA source for the inteins remains, as well as the need to search for the appropriate restriction enzymes compatible with the exteins. Using our pCIRC$_{Npu}$ and pCIRC$_{gp41\text{-}1}$ it will be possible to quickly test several exteins combinations with two of the most efficient split inteins known to date thanks to the usage of overhangs created specifically for this application. A similar approach has already been used in the commercially available IMPACT system, which employs the type IIS restriction enzyme SapI.[61] While this enzyme creates 3-base pair overhangs and thus only requires the very terminal amino acids be fixed, enzymes which create the 4-base pair

overhangs used in our cloning standard are more commonly used and may already be available in many laboratories.

Moreover, since our cloning standard is designed to be compatible with most inteins, the construction of further vectors, which can be used with the same inserts, is also easy. Facilitating the cloning is surely an important aspect to encourage future experiments on circular proteins. However, it should be kept in mind that the bottleneck is in the downstream analyses that require protein purification and *in vitro* biochemical assays in most cases.

Beyond thermostability, a desirable feature of proteins is the ability to withstand thermal inactivation caused by misfolding and aggregation upon heat-induced unfolding. For instance, in the polymerase chain reaction, DNA polymerases need to remain active after exposure to the high temperature required for DNA denaturation. We performed an assay to determine the residual activity after exposure to high temperatures of the circular RGKCWE variant, its linear counterpart and native xylanase. We made the interesting observation that these proteins show a minimum of residual activity at a temperature around the T_m, above which they regain some activity (Fig. 6a and b). The aggregation assay establishes a connection between the residual activity and the amount of protein aggregates formed under these heat shock conditions (Fig. 6c and d). These data demonstrate that there is a temperature around the T_m at which the amount of aggregates is highest, but for temperatures above this threshold there is a significant increase in enzymatically active, soluble material. Since this holds true for all variants studied, we conclude that it is a characteristic of xylanase independent of circularization.

It has been shown that, for proteins consisting solely of β-sheets like xylanase, the intermediates in aggregate formation are not fully unfolded, but still contain native structure elements.[62] This is in agreement with the observation that, besides hydrophobic effects, intermolecular β-sheet formation is a driving force in the formation of stable aggregates.[63] Consequently, at high temperatures where xylanase is fully disordered, the formation of β-sheet-based aggregates may be largely reduced. Furthermore, temperatures close to or slightly above the T_m are considered to be ideal for aggregate formation, while at high temperatures, aggregates are unstable.[63, 64]

Interestingly, in a previous study on the same xylanase, a very similar thermal inactivation assay was performed.[50] Native xylanase at the same pH and concentration

was shown to be completely inactivated after incubation at high temperatures by irreversible thermal denaturation. However, it is likely that the heating and cooling speeds prior to and after the incubation were different. In order to avoid aggregation during temperature changes, we performed heating up or cooling down of the samples as fast as possible. We speculate that the circular RGKCWE variant does not actually aggregate at very high temperatures, but only during the heating and cooling phases. The results of the heat shock experiments we performed by exposing the protein first to 60 ºC and subsequently to 90 ºC in comparison to incubation at 60 °C only (Fig. 6e) strongly suggest that the aggregates even dissolve at high temperatures.

Protein concentration is also known to have a critical effect on aggregation behavior.[63] At very low concentrations, aggregation does not occur, because the proteins likely do not interact with each other. Increasing the concentration usually leads to a strong increase in aggregation. For instance, native xylanase and the linear RGKCWE variant at a concentration of 0.02 mg/ml remain partially soluble after incubation at 70 °C (Fig 6a), while they almost completely aggregate at a concentration of 0.1 mg/ml (Fig. 6b). Surprisingly, at least under the conditions we used, this does not apply to the circular RGKCWE variant. Independent of the concentration, approximately half of the activity is retained after incubation. We can only speculate about the reason for the different aggregation behavior of the circular RGKCWE variant. It might be possible that circularization is incompatible with certain stable aggregates. Another explanation might be that circularization leads to a more compact unfolded state, which decreases the rate of collision of molecules and thus results in slower aggregation. Interestingly, a similar explanation was proposed to explain the effects of disulfide bonds in T4 lysozyme.[65]

Taken together, our results show that circularization has positive effects on xylanase, leading to either an increase in thermostability (circular S variant) or a decrease in aggregation propensity (circular RGKCWE variant and, to some extent, circular S variant). While we were happy to see that circularization could lead to increased thermostability in at least one case, this circular variant would likely not directly bring a benefit to the industrial processes in which xylanase is used. However, backbone circularization can be implemented on other xylanases sharing homology to the one used here that have a higher thermostability to begin with. Moreover, further mutations can be introduced, for instance by directed evolution, on circular xylanase

to increase its thermostability even more. We thus emphasize the use of backbone circularization as strategy to create proteins suitable for biomedical and biotechnological applications.

Experimental

Unless otherwise specified, chemical reagents were purchased from Sigma-Aldrich, St. Louis, MO, USA, and used without further purification.

Cloning

Enzymes used for standard cloning procedures were purchased from New England Biolabs, Ipswich, MA, USA. PCR amplification, overlap extension PCR (OE-PCR)[66] and circular polymerase extension cloning (CPEC)[67] were performed using Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher Scientific, Waltham, MA, USA). Oligonucleotides were ordered from Sigma-Aldrich. BioBricks from the Registry of Standard Biological Parts (http://partsregistry.org/) are indicated by their part name (BBa_number).

Circularization vectors: pCIRC$_{Npu}$ and pCIRC$_{gp41-1}$

Two DNA oligos forming an insert consisting of a NotI and an XbaI restriction site, a strong RBS derived from the T7 phage gene 10a[68] (BBa_K1362090) followed by an arbitrary sequence of homology needed for circular polymerase extension cloning (CPEC) with appropriate overhangs for cloning with EcoRI and SpeI were annealed. The insert was cloned into the expression vector pSB1K30 (BBa_K1362093) using EcoRI and SpeI. The resulting cloning intermediate was called pSB1K30-T7RBS-IGEMHD, because the arbitrary sequence codes for the amino acid sequence IGEMHD.

In order to create pCIRC$_{Npu}$, four PCR products with overlapping homologous regions were created. pSB1K30-T7RBS-iGEMHD was linearized between the T7 RBS and the IGEMHD sequence by PCR, forming the vector fragment. For the first insert, the *Npu* DnaE C-intein coding sequence was amplified from pVS07,[45] kindly provided by H. D. Mootz. The second insert was a PCR product of an mRFP expression cassette (BBa_J04450) with BsaI restriction sites pointing outwards added on both sides. For

the third insert, the *Npu* DnaE N-Intein coding sequence was amplified from pVS41,[69] also provided by H. D. Mootz. Two stop codons were added by the reverse primer. The three inserts were assembled to one large insert by OE-PCR. The large insert was inserted into the vector fragment via CPEC, resulting in pCIRC$_{Npu}$ (BBa_K1362000). The same vector fragment was also used for the assembly of pCIRC$_{gp41\text{-}1}$. In addition, one large insert with overlapping homologous regions to the vector fragment was created, containing the gp41-1 C-intein coding sequence, and mRFP expression cassette with BsaI restriction sites pointing outwards on both sides and the gp41-1 N-intein coding sequence. The insert was assembled from PCR products of BBa_K1362160 and BBa_K1362161. A detailed protocol on how to clone genes of interest into the pCIRC$_{intein}$ plasmids can be found in the Electronic Supplementary Information. pCIRC$_{Npu}$ and pCIRC$_{gp41\text{-}1}$ are available through the Addgene repository (www.addgene.org).

Circular xylanase expression constructs: pCIRC$_{Npu}$-^{C}xynA, pCIRC$_{Npu}$-^{CWE}xynA^{RGK}, pCIRC$_{gp41\text{-}1}$-^{C}xynA, pCIRC$_{gp41\text{-}1}$-^{S}xynA, and pCIRC$_{gp41\text{-}1}$-^{CWE}xynA^{RGK}

The xylanase insert for pCIRC$_{Npu}$-^{CWE}xynA^{RGK} was created by PCR amplification from a *B. subtilis* 168 colony with primers binding to the xynA coding sequence. A BsaI restriction site, the intein-originating overhang sequence CAAC and the sequence coding for the non-native C-extein CWE[55] were added by the forward primer. The (5'-3') sequence coding for the N-extein RGK,[55] the intein-originating overhang (5'-3') sequence TGCT and a BsaI restriction site were added by the reverse primer. The insert was cloned into pCIRC$_{Npu}$ using Golden Gate assembly[53] with subsequent re-ligation due to a BsaI restriction site being present in the xylanase coding sequence.

pCIRC$_{Npu}$-^{C}xynA, pCIRC$_{gp41\text{-}1}$-^{C}xynA, pCIRC$_{gp41\text{-}1}$-^{S}xynA, and pCIRC$_{gp41\text{-}1}$-^{CWE}xynA^{RGK} were created analogously with the respective extein coding sequences added by the primers. The corresponding inserts were amplified from pNAT-xynA and cloned in pCIRC$_{Npu}$ or pCIRC$_{gp41\text{-}1}$.

Native xylanase expression construct: pNAT-xynA

The xylanase coding sequence was amplified from pCIRC$_{Npu}$-^{CWE}xynA^{RGK}. An XbaI restriction site and an RBS (BBa_B0034) were added by the forward primer. A

BioBrick suffix containing a PstI restriction site was added by the reverse primer. This insert was cloned into pSB1K30 using XbaI and PstI.

Linear xylanase expression constructs pLIN-^{C}xynA, pLIN-xynA^{C}, pLIN-^{S}xynA, pLIN-xynA^{S}, pLIN-xynA^{RGKCWE}, pLIN-$^{\Delta}$xynA and pLIN-xynA$^{\Delta}$

In order to create pLIN-xynA^{RGKCWE}, two DNA oligos forming an insert coding for the amino acids RGKCWE with overhangs designed for scarless Golden Gate cloning into pNAT-xynA were annealed. pNAT-xynA was linearized between the xylanase coding sequence and the N-extein coding sequence and Esp3I (BsmBI isoschizomer) restriction sites were added on both sides. The linearized pNAT-xynA and the annealed oligo insert were assembled to form pLIN-xynA^{RGKCWE} by Golden Gate cloning with Esp3I (Thermo Fisher Scientific, Waltham, MA, USA).

For pLIN-xynA^{C}, overlapping mutagenesis primers adding the coding sequence for a C were designed. The PCR product was directly used for transformation, yielding pLIN-xynA^{C}.

pLIN-^{C}xynA, pLIN-^{S}xynA and pLIN-$^{\Delta}$xynA were created analogously to pNAT-xynA. The forward primers were adjusted in order to either add the coding sequence for the extein or omit the N-terminal A of the xylanase coding sequence.

pLIN-xynA^{S} and pLIN-xynA$^{\Delta}$ were created similarly, but without a PstI site at the end of the insert. The inserts were cloned into pSBX1K30 using XbaI and SpeI. Here, the reverse primers were adjusted in order to either add the coding sequence for S or omit the C-terminal W of the xylanase coding sequence.

Expression and Purification

E. coli BL21(DE3) transformed with any of the described plasmids were grown in ZYM-5052 auto-induction medium[56] at 37 °C. When the optical density at 600 nm (1 cm path length) reached 0.6, the temperature was shifted to 20 °C. The cells were harvested by centrifugation the next morning, resuspended in 50 mM sodium phosphate buffer (pH 6.0), and lysed by sonication. The crude lysate was cleared by centrifugation at 20,000 × *g*. The supernatant was applied to a column packed with CM Sephadex C-50 equilibrated with the same buffer at a linear flow rate of 0.5 cm/min. The column was then washed with 5 column volumes of the buffer. A linear gradient from 0 M to 1 M NaCl was applied to the column and 5 mL fractions were

collected. The fractions were analyzed by SDS-PAGE. Fractions containing xylanase, as determined by the presence of a band at the expected molecular weight in the gels, were pooled and the buffer was exchanged with 20 mM sodium phosphate (pH 5.7) using an Amicon Ultra-15 centrifugal filter unit (10 kDa nominal molecular weight cutoff, EMD Millipore, Bedford, MA, USA). The circular S variant was subsequently incubated at 60 °C for 15 min in order to cause the linear xylanase by-products and other contaminants to denature and aggregate, while the circular product would remain soluble (compare Fig. 4a and Fig S4). The aggregates were removed by centrifugation at 20,000 × *g* and the supernatant was used for all further experiments. For long-term storage, the enzymes were mixed with an equal volume of 100 % glycerol, filter sterilized, and kept at -20 °C.

Mass Spectrometry

Mass spectrometric analysis was performed at the Core Facility for Mass Spectrometry and Proteomics of the ZMBH (Zentrum für Molekulare Biologie der Universität Heidelberg). Samples were reduced with DTT, alkylated with iodoacetamide and digested with Endoproteinase GluC (New England Biolabs, Ipswich, MA, USA) using a Digest pro MS liquid handling system (Intavis, Cologne, Germany) following a protocol as described by Catrein *et al.*[70] Digested peptides were then extracted from the gel pieces with 50 % acetonitrile/0.1 % trifluoroacetic acid (TFA), concentrated nearly to dryness in a SpeedVac vacuum centrifuge and diluted to a total volume of 30 µL with 0.1 % TFA. 10 µL of the sample were analyzed by a nanoHPLC system (Waters, Milford, MA, USA) coupled to an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) as follows: the samples were loaded onto a 20 cm C18 column with a flow rate of 30 µL/min. MeCN and H_2O, both containing 0.1 % formic acid, were used as mobile phases. Peptides were separated and eluted with a linear gradient of 3 % to 50 % MeCN over the course of 30 min at a flow rate of 300 nL/min. MS data were acquired with an automatic switch between a full scan and up to ten data-dependent MS/MS scans.

Tandem mass spectra were extracted by Mascot Daemon without grouping or smoothing and analyzed using Mascot (version 2.4.1, Matrix Science, Boston, MA, USA). Mascot was set up to search against a manually compiled database containing the protein of interest and a second entry of two concatenated sequences of the same

protein. The protease was set to GluC, the fragment ion mass tolerance was 0.50 Da and the parent ion tolerance was 20 ppm. Carbamidomethylation of cysteine was specified as a fixed modification, and deamidation of asparagine and glutamine and oxidation of methionine as variable modifications. Scaffold (version 4.0.3, Proteome Software, Portland, OR, USA) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0 % probability by the Peptide Prophet algorithm[71] with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at a probability greater than 99.0 % and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm.[72]

Xylanase Activity Assays

Xylan degradation by xylanase is assayed by interrupting the reaction and measuring the amount of released xylose. To quantify the xylose, we chose a standard assay based on 3,5-dinitrosalicylic acid (DNS),[73] and modified a published protocol[74] to use a thermocycler (Mastercycler Nexus GSX1, Eppendorf, Hamburg, Germany). Substrate solution was prepared by dissolving 1 % (w/v) beechwood xylan in 0.1 M MES-NaOH (pH 5.7) and heating in boiling water for 3 hours. The solution was cleared by sterile filtration. Stop solution containing 1 % (w/v) DNS, 30 % (w/v) potassium sodium tartrate tetrahydrate and 0.4 M NaOH was prepared by mixing the components and stirring at room temperature for an hour. In general, 10 µL of enzyme dilution and 90 µL of substrate solution were mixed and left for 10 min at a defined temperature for the xylan degradation reaction. Subsequently, 100 µL of stop solution were added and the temperature was changed to 98 °C for 10 min for the xylose-dependent reduction of DNS to 3-amino-5-nitrosalicylic acid. After cooling down, the absorption at 540 nm was measured using 96-well microplates (Greiner Bio-One) in a plate reader (Tecan Infinite® M200, Tecan, Männedorf, Switzerland). The absorption of a xylanase-lacking control was subtracted to obtain the difference in absorption, which is proportional to the xylanase activity.

Differential Scanning Fluorimetry

To determine the melting point of the xylanase variants, the Protein Thermal Shift Dye Kit (Life Technologies, Thermo Fisher Scientific, Waltham, MA, USA) was

employed. The assay was prepared in triplicates according to the manufacturer's recommendations using a protein concentration of 100 µg/mL in the absence of glycerol. The assay was run in a StepOnePlus Real Time PCR System (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA) with a step and hold program (0.1 K per step for 15 s) from 25 °C to 99 °C. A sigmoid curve given by

$$I = \frac{\alpha_{\mathrm{N}} + \beta_{\mathrm{N}} \cdot T + (\alpha_{\mathrm{D}} + \beta_{\mathrm{D}} \cdot T) \cdot \exp\left(\frac{-\Delta G_{\mathrm{m}}}{RT}\right)}{1 + \exp\left(\frac{-\Delta G_{\mathrm{m}}}{RT}\right)}$$

(equation 9 in Clarke and Fersht[75], adapted for thermal denaturation), where

$$\Delta G_{\mathrm{m}} = \Delta H_{\mathrm{m}} \cdot \left(1 - \frac{T}{T_{\mathrm{m}}}\right) + \Delta C_{\mathrm{m},p} \cdot \left(T - T_{\mathrm{m}} - T \cdot ln\left(\frac{T}{T_{\mathrm{m}}}\right)\right)$$

(equation 4 in Nicholson and Scholtz[76]), was fitted to the measured fluorescence intensity data using non-linear regression. Here, I is the fluorescence intensity, α_{N} and α_{D} are the baseline fluorescence of the native and denatured states, respectively, β_{N} and β_{D} are the temperature-dependent change in fluorescence for the native and denatured state, respectively, T is the current temperature, T_{m} is the temperature at which half of the protein is denatured, ΔH_{m} is the molar unfolding enthalpy at $T = T_{\mathrm{m}}$, $\Delta C_{\mathrm{m},p}$ is the molar change in heat capacity of the unfolding process, and R is the universal gas constant. The equation was fitted separately to each of three replicate per protein and the obtained parameters were averaged. While a remaining trend was discernible in the residuals, they were generally in the range of 2.5 % of the measured values. The model was thus considered valid for the determination of the melting temperature.

Determination of temperature optimum for catalysis

Activity assays were performed at seven temperatures between 45 °C and 65 °C, using a modified version of a published protocol[77]. The purified enzymes were diluted in either 0.1 M MES-NaOH (pH 5.7) buffer or 20 mM sodium phosphate buffer (pH 5.7) to a concentration of 20 µg/mL. For the activity assay 90 µL of substrate solution were pre-warmed at reaction temperature for 2 min in the thermocycler. Subsequently, 10 µL of enzyme dilution were added and allowed to react for 10 min, until stop solution was added. The xylanase activity was determined as described above. For every replicate, a third-order polynomial was fitted to the measured activity with respect to the temperature, and its local maximum was calculated. These values were then averaged over three replicates.

Residual activity after heat shock assay

Enzyme dilutions in 0.1 M MES-NaOH (pH 5.7) with concentrations of 100 µg/mL (containing 6.3 % glycerol and 1.26 mM DTT) and 20 µg/mL (containing 1.26 % glycerol and 0.25 mM DTT) were prepared. 10 µL of enzyme dilution were heat shocked at temperatures between 37 °C and 99 °C. After 10 min at the heat shock temperature, the enzyme dilutions were cooled down to 37 °C. After one additional minute at 37 °C, 90 µL of substrate solution were added in order to measure residual activity at 37 °C for 10 min, until stop solution was added. The temperature changes between the individual steps were carried out at the maximum ramp rates of the thermocycler (approx. 5 K/s for heating up and 3.5 K/s for cooling down). The xylanase activity was determined as described above. The data were normalized to activity after a 37 °C heat shock.

Aggregation after heat shock assay

Purified enzymes in 20 mM sodium phosphate buffer (pH 5.7) were diluted to a concentration of 100 µg/mL. Aliquots of 50 µL were heat shocked at different temperatures between 37 °C and 99 °C. The aliquots were placed in the pre-warmed thermocycler and cooled down in ice-cold water after 10 min. Subsequently the aliquots were centrifuged for 30 min at 18,000 × *g*. The supernatant was removed and mixed with 16.7 µL 4X Laemmli sample buffer (Bio-Rad Laboratories, Hercules, CA, USA) containing 100 mM DTT (4X concentration). 50 µL of 20 mM sodium phosphate buffer (pH 5.7) and 16.7 µL of 4X Laemmli sample buffer containing DTT were added to the pellet. The samples were incubated at 95 °C for 10 min and run on Tris-Glycine-SDS gels. The gels were then submerged in a 1:1 methanol/water mixture saturated with chloroform for 10 min, washed twice in a methanol/water mixture and 5 times in water over the course of 30 min. The fluorescent signal was visualized using a UV transilluminator (312 nm) for 5 min. The band intensities were quantified using a two-dimensional curve fitting approach and the percentage of protein in the supernatant was calculated.

Acknowledgments

We thank Henning D. Mootz for providing plasmids pVS07 and pVS41 containing the *Npu* DnaE split intein, Oliver Wichmann and Mathias Utz for help with protein purification, Stefan Huber for support with differential scanning fluorimetry measurements, and Joel Beaudouin for discussions and critical reading of the manuscript. Mass spectrometry was performed by Bernd Heßling at the core facility for mass spectrometry and proteomics of the Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH). This work was supported by the Klaus Tschira Foundation, the Dietmar Hopp Foundation, CellNetworks, the Helmholtz Association, the BioQuant, the DKFZ and the University of Heidelberg.

References

1. M. S. Urbieta, E. R. Donati, K. G. Chan, S. Shahar, L. L. Sin and K. M. Goh, *Biotechnology advances*, 2015.
2. S. H. Bjornsdottir, O. H. Fridjonsson, J. K. Kristjansson and G. Eggertsson, *Extremophiles : life under extreme conditions*, 2007, **11**, 283-293.
3. F. E. Jenney, Jr. and M. W. Adams, *Annals of the New York Academy of Sciences*, 2008, **1125**, 252-266.
4. A. Hidalgo, L. Betancor, R. Moreno, O. Zafra, F. Cava, R. Fernandez-Lafuente, J. M. Guisan and J. Berenguer, *Applied and environmental microbiology*, 2004, **70**, 3839-3844.
5. K. Ishikawa, H. Ishida, Y. Koyama, Y. Kawarabayasi, J. Kawahara, E. Matsui and I. Matsui, *Journal of Biological Chemistry*, 1998, **273**, 17726-17731.
6. C. Vieille and G. J. Zeikus, *Microbiol Mol Biol Rev*, 2001, **65**, 1-43.
7. M. Lehmann and M. Wyss, *Current opinion in biotechnology*, 2001, **12**, 371-375.
8. P. A. Karplus and G. E. Schulz, *Naturwissenschaften*, 1985, **72**, 212-213.
9. G. Auerbach, R. Ostendorp, L. Prade, I. Korndorfer, T. Dams, R. Huber and R. Jaenicke, *Struct Fold Des*, 1998, **6**, 769-781.
10. G. Auerbach, R. Huber, M. Grattinger, K. Zaiss, H. Schurig, R. Jaenicke and U. Jacob, *Structure*, 1997, **5**, 1475-1483.
11. S. Macedo-Ribeiro, B. Darimont, R. Sterner and R. Huber, *Structure*, 1996, **4**, 1291-1301.
12. D. J. Craik and N. M. Allewell, *J Biol Chem*, 2012, **287**, 26999-27000.
13. L. Cascales and D. J. Craik, *Organic & biomolecular chemistry*, 2010, **8**, 5035-5047.
14. M. Trabi and D. J. Craik, *Trends in biochemical sciences*, 2002, **27**, 132-138.
15. H. Iwai and A. Pluckthun, *Febs Lett*, 1999, **459**, 166-172.
16. C. Siebold and B. Erni, *Biophys Chem*, 2002, **96**, 163-171.
17. H. Iwai, A. Lingel and A. Pluckthun, *J Biol Chem*, 2001, **276**, 16548-16554.
18. C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnon and S. J. Benkovic, *Proceedings of the National Academy of Sciences of the United States of America*, 1999, **96**, 13638-13643.

19. J. M. Thornton and B. L. Sibanda, *Journal of molecular biology*, 1983, **167**, 443-460.
20. H. Wu, Z. Hu and X. Q. Liu, *Proceedings of the National Academy of Sciences of the United States of America*, 1998, **95**, 9226-9231.
21. D. P. Goldenberg and T. E. Creighton, *Journal of molecular biology*, 1983, **165**, 407-413.
22. M. Q. Xu and T. C. Evans, Jr., *Current opinion in biotechnology*, 2005, **16**, 440-446.
23. N. H. Shah and T. W. Muir, *Chem Sci*, 2014, **5**, 446-461.
24. N. I. Topilina and K. V. Mills, *Mobile DNA*, 2014, **5**, 5.
25. M. Q. Xu and T. C. Evans, Jr., *Methods*, 2001, **24**, 257-277.
26. A. S. Aranko, A. Wlodawer and H. Iwai, *Protein engineering, design & selection : PEDS*, 2014, **27**, 263-271.
27. D. W. Wood and J. A. Camarero, *J Biol Chem*, 2014, **289**, 14512-14519.
28. H. D. Mootz, *Chembiochem : a European journal of chemical biology*, 2009, **10**, 2579-2589.
29. G. Skretas and D. W. Wood, *Protein science : a publication of the Protein Society*, 2005, **14**, 523-532.
30. A. R. Buskirk, Y. C. Ong, Z. J. Gartner and D. R. Liu, *Proc Natl Acad Sci U S A*, 2004, **101**, 10505-10510.
31. M. P. Zeidler, C. Tan, Y. Bellaiche, S. Cherry, S. Hader, U. Gayko and N. Perrimon, *Nature biotechnology*, 2004, **22**, 871-876.
32. E. Adam and F. B. Perler, *J Mol Microbiol Biotechnol*, 2002, **4**, 479-487.
33. H. D. Mootz and T. W. Muir, *Journal of the American Chemical Society*, 2002, **124**, 9044-9045.
34. H. D. Mootz, E. S. Blum, A. B. Tyszkiewicz and T. W. Muir, *Journal of the American Chemical Society*, 2003, **125**, 10561-10569.
35. W. Ren, A. Ji and H. W. Ai, *J Am Chem Soc*, 2015, **137**, 2155-2158.
36. J. K. Bocker, K. Friedel, J. C. Matern, A. L. Bachmann and H. D. Mootz, *Angew Chem Int Ed Engl*, 2015, **54**, 2116-2120.
37. J. Binschik, J. Zettler and H. D. Mootz, *Angew Chem Int Ed Engl*, 2011, **50**, 3249-3252.
38. F. H. Gohard, D. J. St-Cyr, M. Tyers and W. C. Earnshaw, *Open biology*, 2014, **4**, 140163.
39. K. R. Lennard and A. Tavassoli, *Chemistry*, 2014, **20**, 10608-10614.
40. S. A. El-Mowafi, J. N. Alumasa, S. E. Ades and K. C. Keiler, *Antimicrobial agents and chemotherapy*, 2014, **58**, 5500-5509.
41. A. Tavassoli and S. J. Benkovic, *Nature protocols*, 2007, **2**, 1126-1133.
42. Z. Zhao, X. Ma, L. Li, W. Zhang, S. Ping, M.-Q. Xu and M. Lin, *J. Microbiol. Biotechnol*, 2010, **20**, 460-466.
43. N. K. Williams, P. Prosselkov, E. Liepinsh, I. Line, A. Sharipo, D. R. Littler, P. M. Curmi, G. Otting and N. E. Dixon, *J Biol Chem*, 2002, **277**, 7790-7798.
44. G. Volkmann, P. W. Murphy, E. E. Rowland, J. E. Cronan, X.-Q. Liu, C. Blouin and D. M. Byers, *J Biol Chem*, 2010, **285**, 8605-8614.
45. J. Zettler, V. Schutz and H. D. Mootz, *Febs Lett*, 2009, **583**, 909-914.
46. H. Iwai, S. Zuger, J. Jin and P. H. Tam, *Febs Lett*, 2006, **580**, 1853-1858.
47. K. Jagadish, R. Borra, V. Lacey, S. Majumder, A. Shekhtman, L. Wang and J. A. Camarero, *Angewandte Chemie-International Edition*, 2013, **52**, 3126-3131.

48. P. Carvajal-Vallejos, R. Pallisse, H. D. Mootz and S. R. Schmidt, *Journal of Biological Chemistry*, 2012, **287**, 28686-28696.
49. L. Song, A. Tsang and M. Sylvestre, *Biotechnology and Bioengineering*, 2015, **112**, 1081-1091.
50. K. Miyazaki, M. Takenouchi, H. Kondo, N. Noro, M. Suzuki and S. Tsuda, *J Biol Chem*, 2006, **281**, 10236-10242.
51. S. P. George, A. Ahmad and M. B. Rao, *Bioresour Technol*, 2001, **78**, 221-224.
52. F. B. Perler, *Nucleic acids research*, 2002, **30**, 383-384.
53. C. Engler, R. Kandzia and S. Marillonnet, *Plos One*, 2008, **3**, e3647.
54. S. Brenzel, T. Kurpiers and H. D. Mootz, *Biochemistry*, 2006, **45**, 1571-1578.
55. M. Cheriyan, C. S. Pedamallu, K. Tori and F. Perler, *J Biol Chem*, 2013, **288**, 6202-6211.
56. F. W. Studier, *Protein expression and purification*, 2005, **41**, 207-234.
57. R. Ruller, J. C. Rosa, V. M. Faça, L. J. Greene and R. J. Ward, *Biotechnology and applied biochemistry*, 2006, **43**, 9-15.
58. A. Mogk, T. Tomoyasu, P. Goloubinoff, S. Rudiger, D. Roder, H. Langen and B. Bukau, *The EMBO journal*, 1999, **18**, 6934-6949.
59. P. J. Flory, *J Am Chem Soc*, 1956, **78**, 5222-5235.
60. A. Tavassoli and S. J. Benkovic, *Nature protocols*, 2007, **2**, 1126-1133.
61. S. Chong, F. B. Mersha, D. G. Comb, M. E. Scott, D. Landry, L. M. Vence, F. B. Perler, J. Benner, R. B. Kucera and C. A. Hirvonen, *Gene*, 1997, **192**, 271-281.
62. D. Bratko and H. Blanch, *The Journal of chemical physics*, 2003, **118**, 5185-5194.
63. J. Gsponer and M. Vendruscolo, *Protein and peptide letters*, 2006, **13**, 287-293.
64. B. Vekhter and R. S. Berry, *The Journal of chemical physics*, 1999, **110**, 2195-2201.
65. R. Wetzel, L. J. Perry, W. A. Baase and W. J. Becktel, *Proceedings of the National Academy of Sciences of the United States of America*, 1988, **85**, 401-405.
66. R. Higuchi, B. Krummel and R. Saiki, *Nucleic acids research*, 1988, **16**, 7351-7367.
67. J. Quan and J. Tian, *Plos One*, 2009, **4**, e6441.
68. P. O. Olins and S. H. Rangwala, *J Biol Chem*, 1989, **264**, 16973-16976.
69. V. Schütz and H. D. Mootz, *Angewandte Chemie*, 2014, **126**, 4197-4201.
70. I. Catrein, R. Herrmann, A. Bosserhoff and T. Ruppert, *The FEBS journal*, 2005, **272**, 2892-2900.
71. A. Keller, A. I. Nesvizhskii, E. Kolker and R. Aebersold, *Analytical chemistry*, 2002, **74**, 5383-5392.
72. A. I. Nesvizhskii, A. Keller, E. Kolker and R. Aebersold, *Analytical chemistry*, 2003, **75**, 4646-4658.
73. G. L. Miller, *Anal. Chem.* , 1959, **31**, 426-428.
74. K. Miyazaki, M. Takenouchi, H. Kondo, N. Noro, M. Suzuki and S. Tsuda, *J Biol Chem*, 2006, **281**, 10236-10242.
75. J. Clarke and A. R. Fersht, *Biochemistry*, 1993, **32**, 4322-4329.
76. E. M. Nicholson and J. M. Scholtz, *Biochemistry*, 1996, **35**, 11369-11378.

77. R. Ruller, L. Deliberto, T. L. Ferreira and R. J. Ward, *Proteins*, 2008, **70**, 1280-1293.

Fig. 1 Schematic overview of intein-mediated backbone circularization. The protein of interest is depicted in cyan.

Fig. 2 Design of the cloning standard. (a) Sequence logo generated from all 511 (N-terminus)/507 (C-terminus) distinct intein sequences listed in InBase.[51] The sequence alignment of the two split inteins used in this work is shown above. The two 4-base pair overhangs to be used in scarless Golden Gate assembly together with all their possible translations are shown at the bottom. (b) Features and usage of the circularization vector $pCIRC_{intein}$, where 'intein' stands for the selected split intein to be used. The insert is created by PCR amplification of the sequence coding for the protein of interest (POI). The sequences coding for the exteins, the overhangs derived in (a) and BsaI restriction sites have to be added by the primers. The insert is cloned in a one-pot reaction into pCIRC, replacing the mRFP expression cassette using Golden Gate assembly. The vector confers resistance to kanamycin. RBS, ribosome binding site.

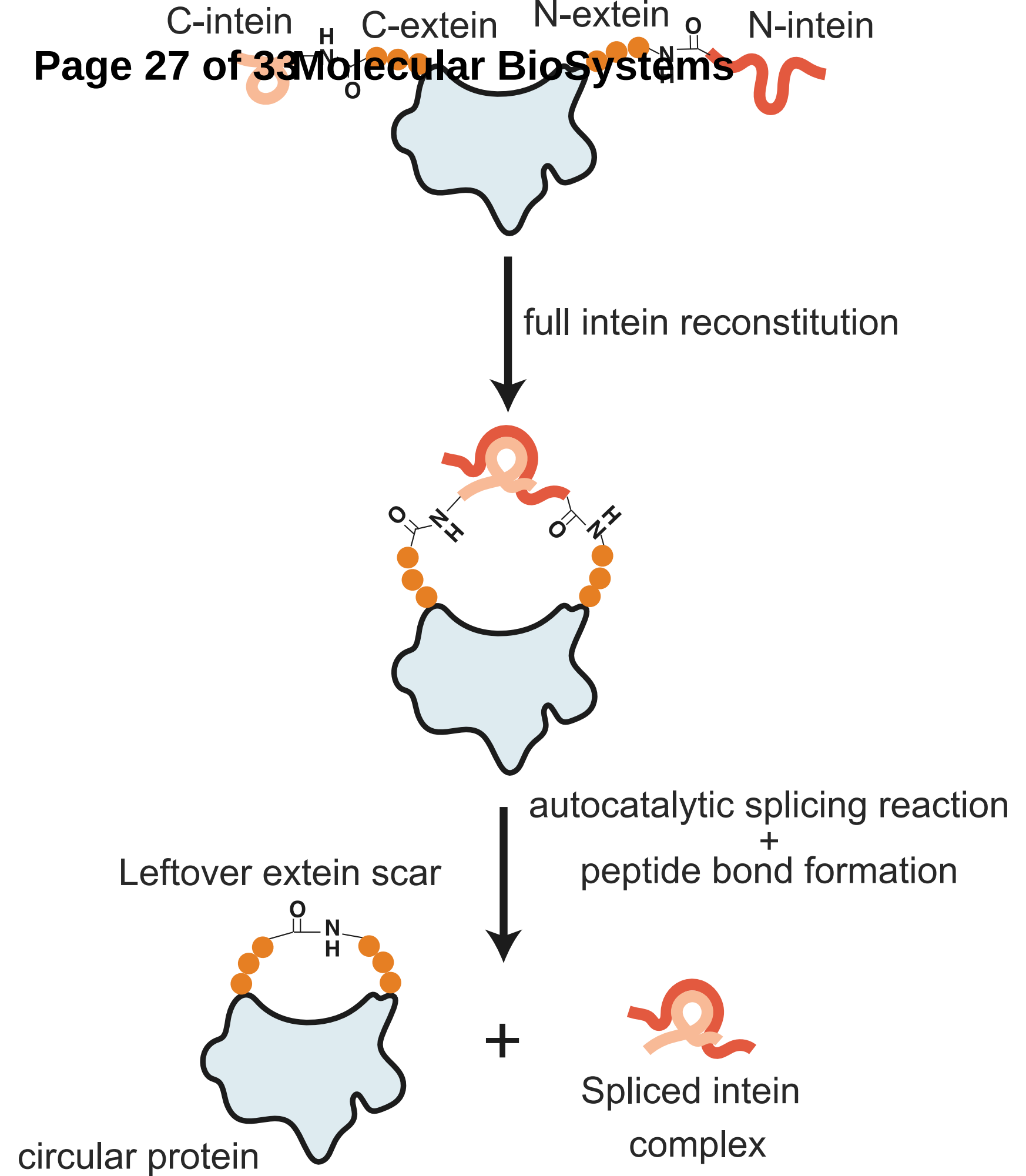
Fig. 3 Schematic representation of the xylanase constructs used in this study. The first and last amino acids of xylanase, the exteins, and the terminal catalytic residues of the split inteins are indicated. Prom, promoter. RBS, ribosome binding site. ATG, initiator methionine.

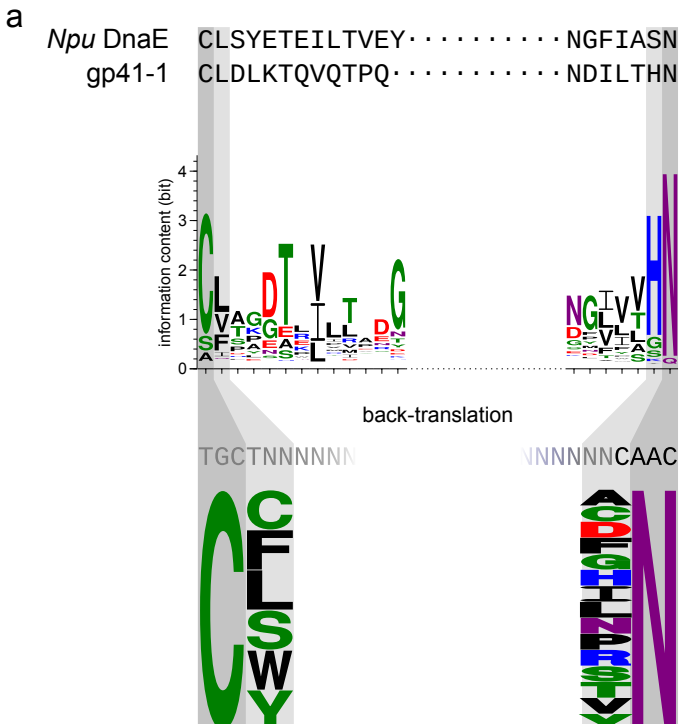
Fig. 4 Proof of circularization**.** (a) SDS-PAGE of circular xylanase variants (circ) and of their linear counterparts (lin). The different exteins used for circularization are indicated in bold. (b) MS/MS spectrum of the predicted C-terminal GluC fragment of native xylanase with both N- and C-exteins added at the C-terminus as expected for circular xylanase. b and y ions that were identified in the MS/MS spectrum are shown below and above the peptide sequence and highlighted in orange and blue, respectively. Ions in the spectrum were labeled independently of their charge state. The parent ion, designated M, is shown in green. Only the most prominent signals have been labeled. The symbols ° and * denote loss of H_2O and NH_3, respectively. (c) ESI-MS extracted ion chromatogram of masses corresponding to the fragment from (b), and the C-terminal xylanase fragment with only the N-intein added to the C-terminus (as expected for the linear splicing by-product).

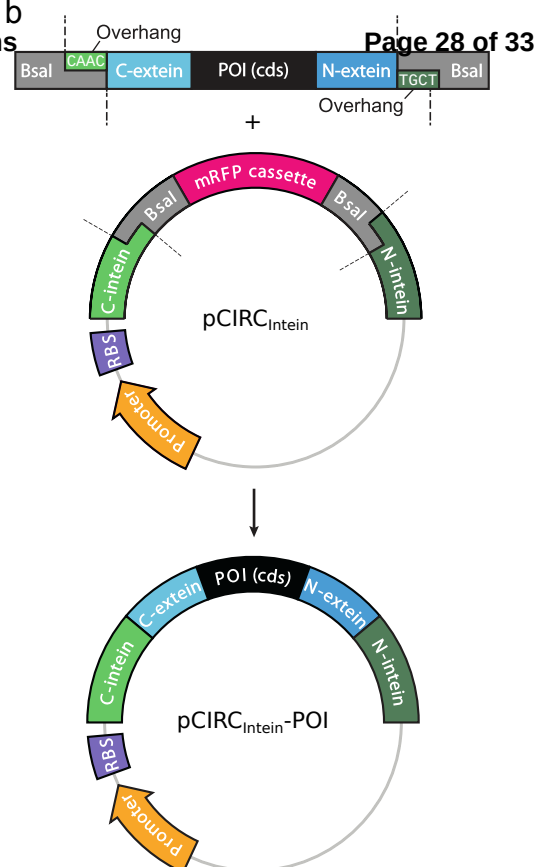
Fig. 5 Temperature dependence of catalytic activities for the circular and linear S variants and native xylanase. Enzymes were used at a final concentration of 2 μg/mL in the absence of glycerol and reducing agents. The data points are normalized to the local maximum of a fitted 3rd degree polynomial for each variant. Error bars represent the standard error of the mean for three independent replicates.

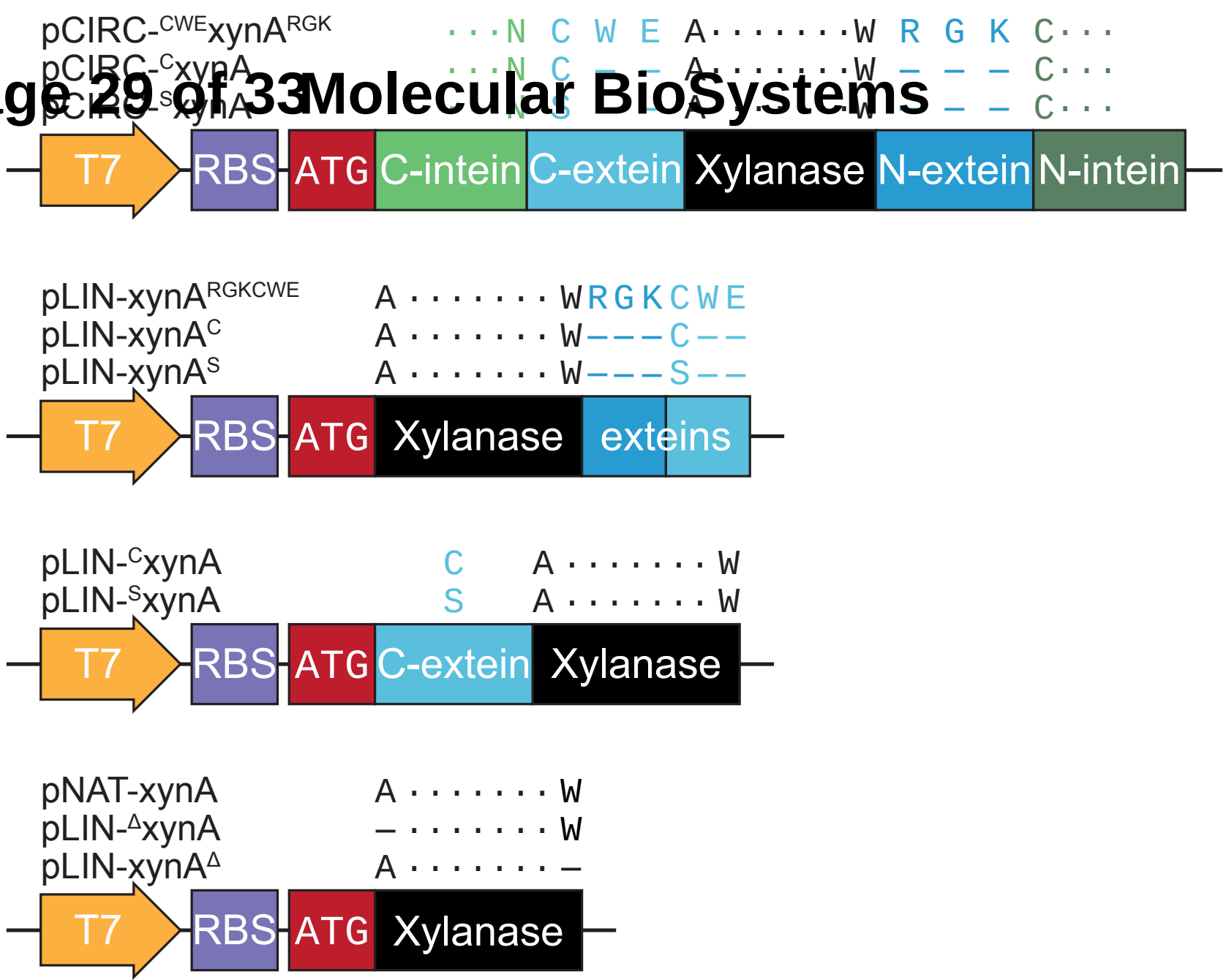
Fig. 6 Residual activity and aggregation of xylanase variants after heat shock. (a, b) Residual activity at 37 °C after heat shock at the indicated temperatures. The data were normalized to the activity after heat shock at 37 °C. Enzymes were used at a concentration of 20 μg/mL (a) or 100 μg/mL (b) during heat shock and diluted down to 2 μg/ml (a) or 10 μg/mL (b) in the activity assay by the addition of the substrate. Error bars represent the standard error of the mean for three independent replicates. (c) Representative SDS-PAGE of circular, linear and native xylanase after heat shock at the indicated temperatures. Circular and linear indicate the RGKCWE variants. The gel was stained with chloroform and imaged using ultraviolet illumination. S, supernatant. P, pellet. (d) Quantification of the aggregation after heat shock experiment shown in (c). Error bars represent the standard error of the mean for three independent replicates. (e) Aggregation of the circular RGKCWE variant after the indicated heat shock regimes. Two conditions used in (c) – "60 °C for 10 min" and "90 °C for 10 min" – were repeated for direct comparison. (f) As in (d) but for the circular and linear S variants. The linear S variant here indicates xylanase with the additional serine at the N-terminus.

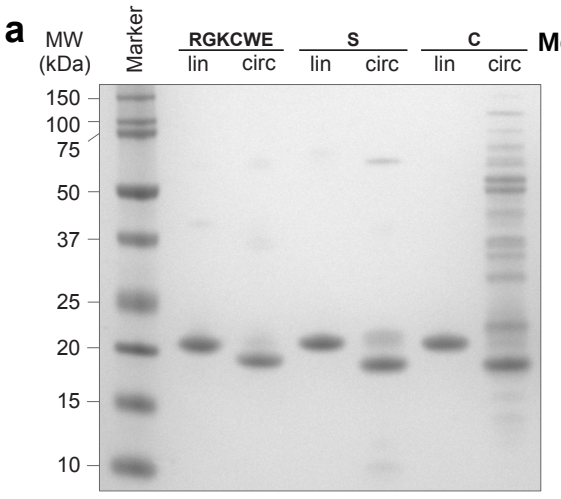
Table 1 Melting temperatures and temperature optima for catalysis for xylanase variants. T_m, melting temperature. Tolerances indicate the standard error of the mean of three independent replicates.

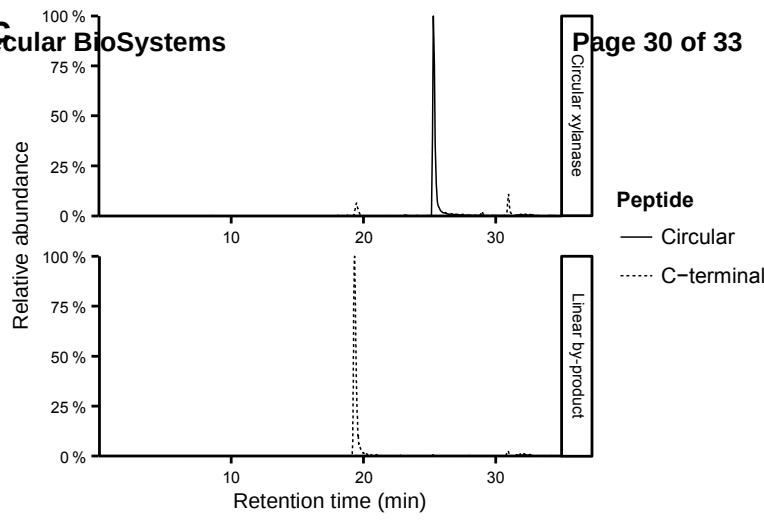

C-intein
C-extein
N-extein
N-intein
full intein reconstitution
autocatalytic splicing reaction
+
peptide bond formation
Leftover extein scar
+
Spliced intein
complex
circular protein

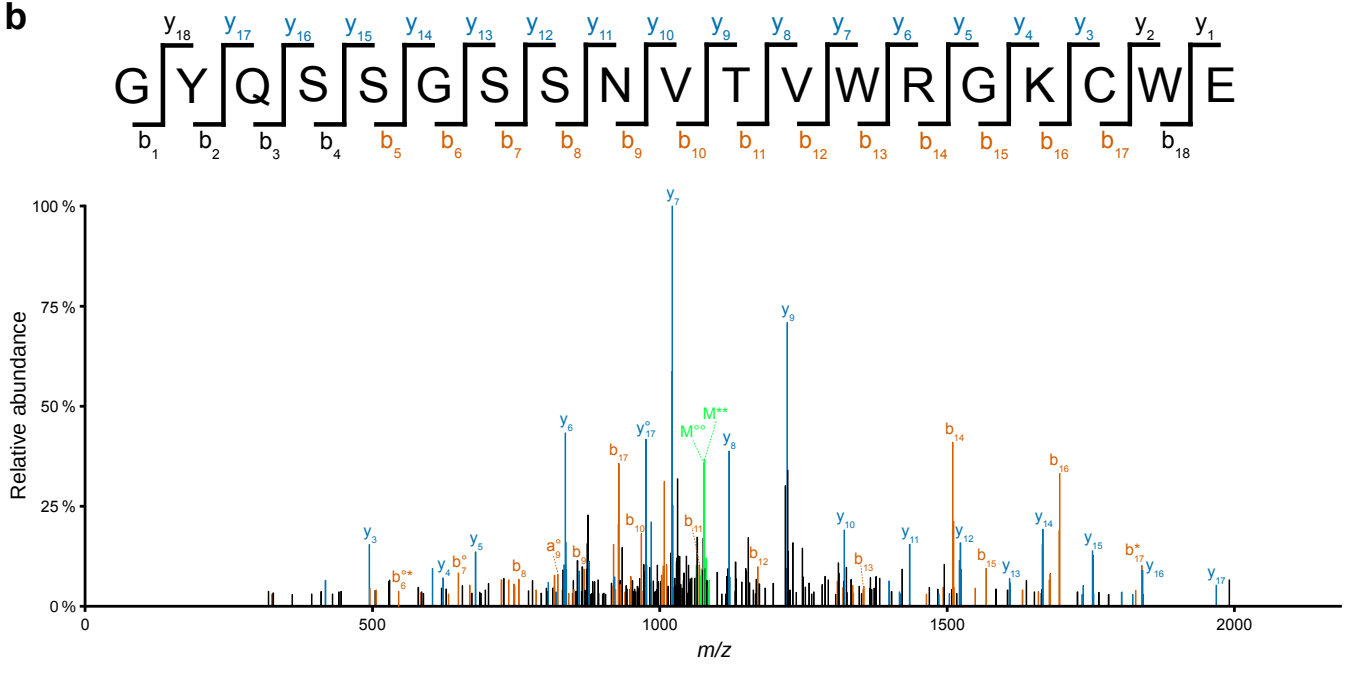

a
Npu DnaE CLSYETEILTVEY·····NGFIASN
gp41-1 CLDLKTQVQTPQ·····NDILTHN
information content (bit)
back-translation
TGCTNNNNNN
NNNNNNCAAC

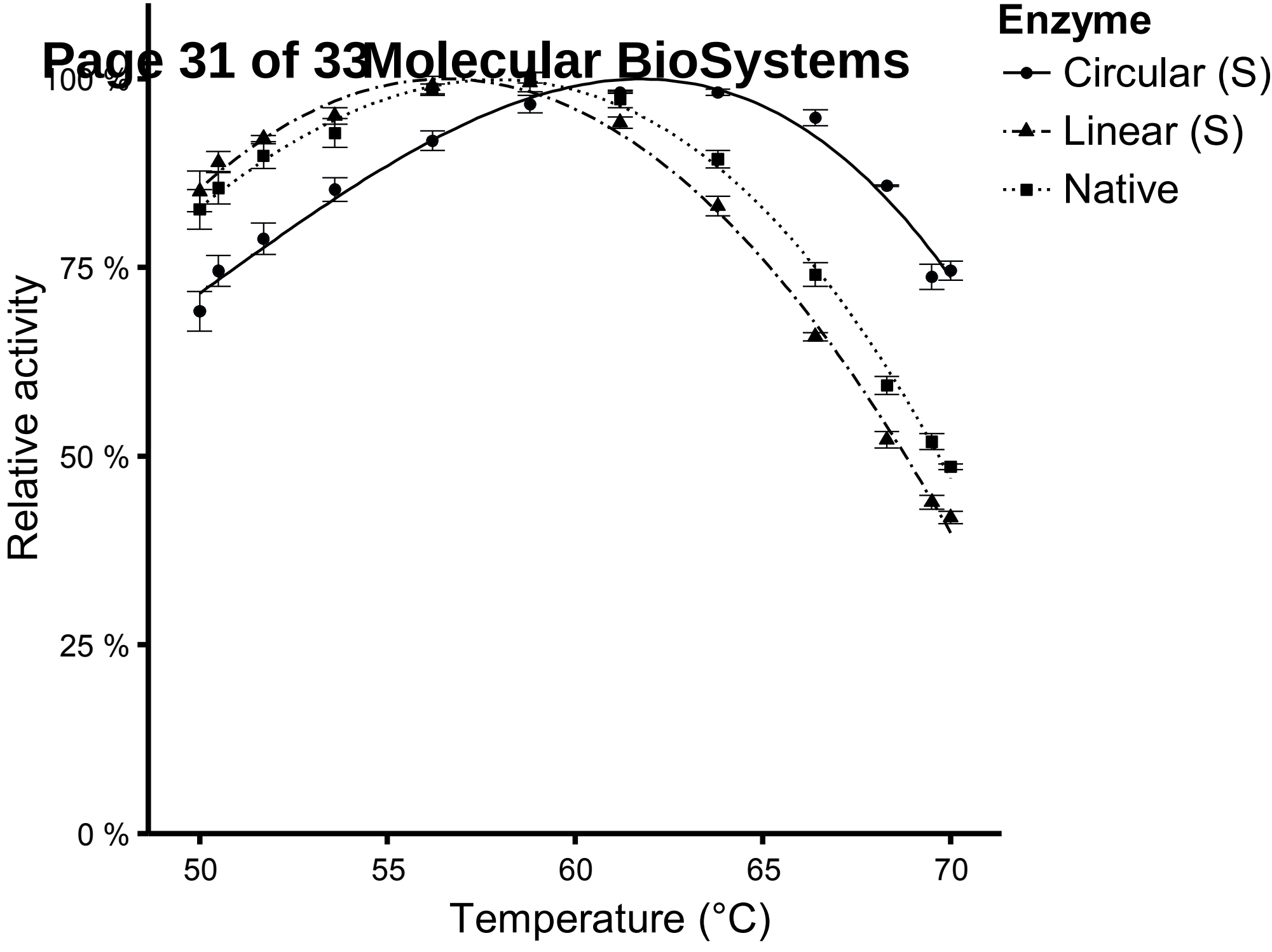

b
Overhang
BsaI
CAAC
C-extein
POI (cds)
N-extein
TGCT
BsaI
Overhang
+
mRFP cassette
BsaI
BsaI
C-intein
N-intein
pCIRCIntein
RBS
Promoter
POI (cds)
C-extein
N-extein
C-intein
N-intein
pCIRCIntein-POI
RBS
Promoter

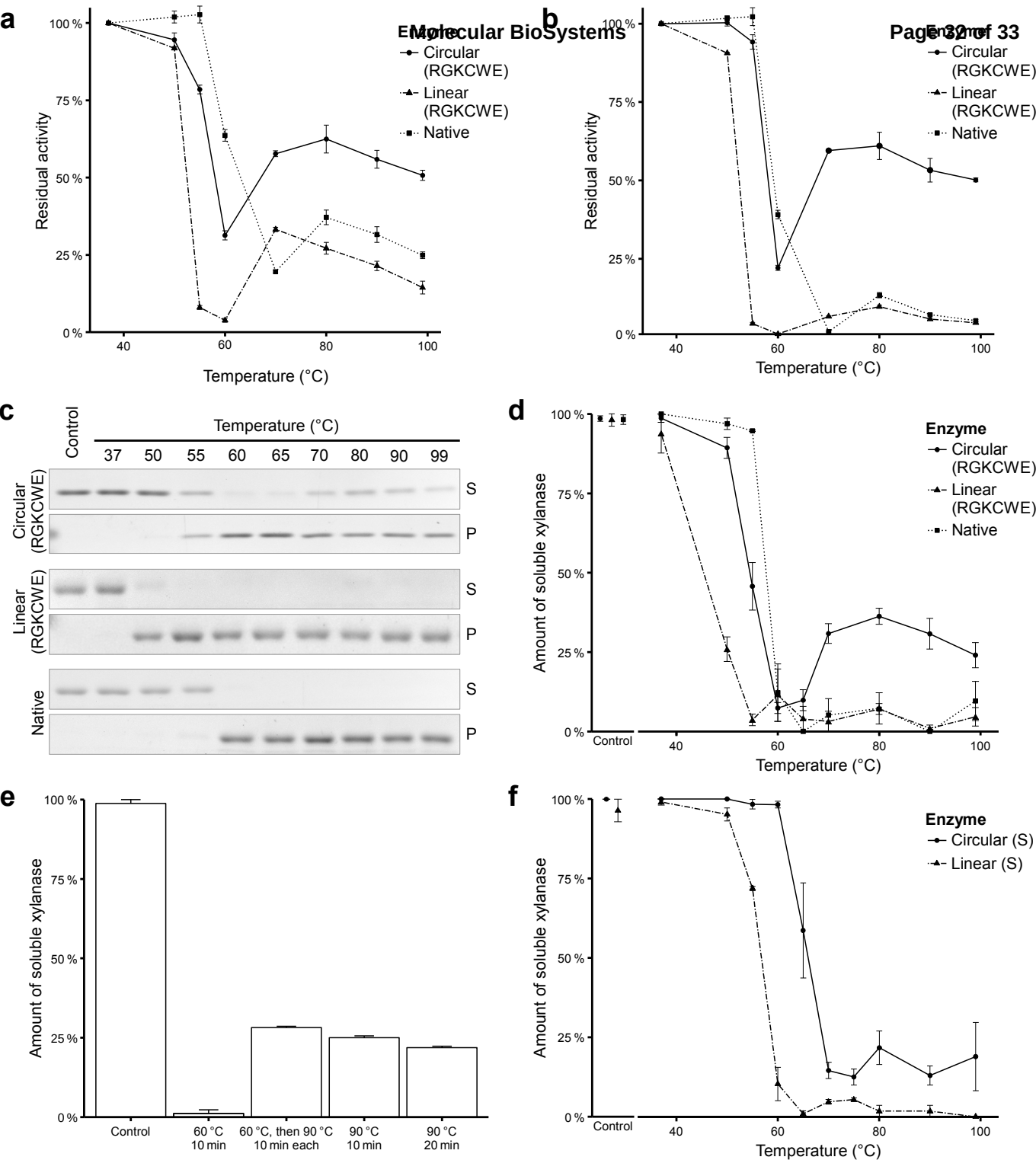

pCIRC-CWExynARGK ···N C W E A·····W R G K C···
pCIRC-CxynA ···N C – – A·····W – – – C···
pCIRC-SxynA ···N S – – A·····W – – – C···
T7
RBS
ATG
C-intein
C-extein
Xylanase
N-extein
N-intein
pLIN-xynARGKCWE A·····W R G K C W E
pLIN-xynAC A·····W – – – C – –
pLIN-xynAS A·····W – – – S – –
T7
RBS
ATG
Xylanase
exteins
pLIN-CxynA C A·····W
pLIN-SxynA S A·····W
T7
RBS
ATG
C-extein
Xylanase
pNAT-xynA A·····W
pLIN-ΔxynA –·····W
pLIN-xynAΔ A·····–
T7
RBS
ATG
Xylanase


a
MW (kDa)
Marker
RGKCWE
S
C
lin
circ
lin
circ
lin
circ
150
100
75
50
37
25
20
15
10


c
100 %
75 %
50 %
25 %
0 %
Relative abundance
Circular xylanase
Linear by-product
10
20
30
Retention time (min)
Peptide
Circular
C−terminal


b
G Y Q S S G S S N V T V W R G K C W E
Relative abundance
100 %
75 %
50 %
25 %
0 %
0
500
1000
1500
2000
m/z


Enzyme
Circular (S)
Linear (S)
Native
Relative activity
100 %
75 %
50 %
25 %
0 %
50
55
60
65
70
Temperature (°C)


a
b
Residual activity
Temperature (°C)
Enzyme
Circular (RGKCWE)
Linear (RGKCWE)
Native
c
Control
Temperature (°C)
37 50 55 60 65 70 80 90 99
Circular (RGKCWE)
Linear (RGKCWE)
Native
S
P
d
Amount of soluble xylanase
Control
e
Amount of soluble xylanase
Control
60 °C 10 min
60 °C, then 90 °C 10 min each
90 °C 10 min
90 °C 20 min
f
Amount of soluble xylanase
Enzyme
Circular (S)
Linear (S)

Exteins added	Variant/Position	T_m (℃)	Temperature Optimum for Catalysis (℃)
(none)	native	56.9 ± 0.2	58.0 ± 0.2
	deletion of N-terminal A	57.1 ± 0.1	n. d.
	deletion of C-terminal W	42.58 ± 0.09	n. d.
S	linear; N-terminus	55.5 ± 0.2	56.5 ± 0.2
	linear; C-terminus	56.4 ± 0.2	n. d.
	circular	63.78 ± 0.09	61.7 ± 0.3
C	linear; N-terminus	55.0 ± 0.3	n. d.
	linear; C-terminus	56.9 ± 0.2	n. d.
	circular	54 ± 1[†]	n. d.
RGKCWE	linear; C-terminus	49.1 ± 0.6	51.4 ± 0.2
	circular	53.4 ± 0.1	54.2 ± 0.4

[†]As a note of caution, since we could not truly separate the circular C variant from other proteins (linear form, non-spliced precursors, other contaminants), and the T_m of the linear form is similar to the one measured, we could not determine the exact T_m of the circular form.

n. d., not determined.