The FEBS Journal

FEBS PRESS science publishing by scientists

Creation of active TIM barrel enzymes through genetic fusion of half-barrel domain constructs derived from two distantly related glycosyl hydrolases

Prerna Sharma, Pallavi Kaila and Purnananda Guptasarma

Department of Biological Sciences, Centre for Protein Science, Design and Engineering, Indian Institute of Science Education and Research (IISER), Mohali, SAS Nagar, India

Keywords
beta/alpha barrels; chimera; enzyme engineering; enzymes; half-barrel domains; TIM barrels

Correspondence
P. Guptasarma, Centre for Protein Science, Design and Engineering, Department of Biological Sciences, Indian Institute of Science Education and Research (IISER), Mohali, Knowledge City, Sector-81, SAS Nagar, Manauli P.O., Punjab 140306, India
Fax: +91 172 2240266
Tel: + 91 172 2293151 and +91 9815417265
E-mails: guptasarma@iisermohali.ac.in and guptasarma@yahoo.com

(Received 20 January 2016, revised 10 September 2016, accepted 14 October 2016)

doi:10.1111/febs.13927

Diverse unrelated enzymes that adopt the beta/alpha (or TIM) barrel topology display similar arrangements of beta/alpha units placed in a radial eight-fold symmetry around the barrel's axis. The TIM barrel was originally thought to be a single structural domain; however, it is now thought that TIM barrels arose from duplication and fusion of smaller half-barrels consisting of four beta/alpha units. We describe here the design, expression and purification, as well as characterization of folding, activity and stability, of chimeras of two TIM barrel glycosyl hydrolases, made by fusing different half-barrel domains derived from an endoglucanase from *Clostridium cellulolyticum*, CelCCA and a beta-glucosidase from *Pyrococcus furiosus,* CelB. We show that after refolding following purification from inclusion bodies, the two half-barrel fusion chimeras (CelCCACelB and CelBCelCCA) display catalytic activity although they assemble into large soluble oligomeric aggregated species containing chains of mixed beta and alpha structure. CelBCelCCA displays hyperthermophile-like structural stability as well as significant stability to chemical denaturation (C_m of 2.6 M guanidinium hydrochloride), whereas CelCCA-CelB displays mesophile-like stability (T_m of ~ 71 °C). The endoglucanase activities of both chimeras are an order of magnitude lower than those of CelB or CelCCA, whereas the beta-glucosidase activity of CelBCelCCA is about two orders of magnitude lower than that of CelB. The chimera CelCCACelB shows no beta-glucosidase activity. Our results demonstrate that half-barrel domains from unrelated sources can fold, assemble and function, with scope for improvement.

Enzyme
Pyrococcus furiosus beta-glucosidase (CelB, EC: 3.2.1.21). *Clostridium cellulolyticum* endoglucanase A (CelCCA, EC: 3.2.1.4).

Introduction

Triosephosphate isomerase barrel proteins represent the most common of enzyme folds. Their structures exhibit a radial symmetry in the arrangement of eight constituent 'strand-loop-helix' motifs, also called beta/alpha units. The beta/alpha units form two coaxial cylinders which together constitute the TIM barrel.

Abbreviations
CDD, conserved domain database; DNSA, 3,5-dinitrosalicylic acid; Gdm.HCl, guanidinium hydrochloride; IMAC, imidazole metal affinity chromatography; IPTG, isopropyl β-D-1-thiogalactopyranoside; NTA, nitrilotriacetic acid; PDB, Protein Data Bank; RCSB, research collaborator for structural bioinformatics; rms, root mean square; SOE, splicing by overlap extension; TIM, triosephosphate isomerase.

Eight beta strands come together to form the inner (hydrophobic) cylinder/barrel which forms the core of the protein. Eight alpha helices come together to form the outer (somewhat more hydrophilic) cylinder/barrel which faces the solvent. The polypeptide chains alternates between the two cylinders as it winds around the barrel's axis. The eight loops that connect beta strands in the inner barrel to alpha helices in the outer barrel are known as beta/alpha loops; these come together to form the mouth of the barrel, and the catalytic activity of the TIM barrel is also carried by these loops [1,2]. The eight smaller loops connecting alpha helices in the inner barrel to the beta strands in the outer barrel are known as alpha/beta loops. Although there is very little sequence conservation among $(\beta/\alpha)_8$ barrel proteins, the different sequences of all TIM barrel proteins somehow manage to adopt the same overall structural fold, ostensibly owing to the overall distribution of similar residues (polar, nonpolar and charged residues) at topologically similar positions in the sequences and structures of different proteins representing this class; a contention which is supported by structure-based sequence alignments [3].

Given the modularity of their structure, therefore, TIM barrel proteins are well suited for examining the question of whether folding mechanism is dictated by native topology or whether conserved amino acids are essential for the formation of some crucial folding intermediates on the pathway to folded native structure [4,5]. Interestingly, not only do TIM barrel proteins display a regular order and high symmetry but they also display some autonomy in the folding of groups of beta/alpha units distributed along the polypeptide chain [6], although up until the beginning of the current century, TIM barrels were widely considered to be single domain proteins. Thus, $(\beta/\alpha)_8$ barrels are favourite subjects of study of protein chemists/engineers who occasionally feel encouraged to design new $(\beta/\alpha)_8$ barrels by fusing, mixing and matching of chain sections. The first fraction of a $(\beta/\alpha)_8$ barrel to be identified and demonstrated to be semiautonomous was the 'half-barrel' consisting of two sets of four beta/alpha units each, first noticed in two highly evolutionarily related proteins of the histidine biosynthesis pathway, HisA and HisF. These two proteins show great sequence and structure conservation, with each retaining trace evidences of an internal gene duplication event, involving two sets of four beta/alpha units each, not evident in the sequences of other TIM barrels [7–9]. The half-barrel sections of HisA as well as HisF show symmetry in the position of certain conserved residues, in that the same set of such residues are present at similar positions in both half-barrels. These findings support hypotheses concerning the possible divergent evolution of TIM barrels, first validated using HisF [1], by constructing two half-barrel domains, HisF-N and HisF-C, and demonstrating that they are structured and assembled into homodimers and higher order oligomers. These half-barrels are inactive when expressed and purified separately, but structured and assembled into heterodimers with their partner half-barrels, and also active, when coexpressed and allowed to fold conjointly *in vivo*, or refolded conjointly *in vitro*.

The homodimers and higher order oligomers formed by such half-barrel domains, when expressed and folded by themselves, ostensibly owe to the tendency to bury hydrophobic surfaces away from the solvent, thus escaping aggregation through hydrophobic collapse [1]. Interestingly, a full $(\beta/\alpha)_8$ barrel generated by fusing the N-terminal half-barrel domain of HisF (HisF-N) to the C-terminal half-barrel domain of HisA (HisA-C) mimicked the evolution of the double half-barrel fold [10]. On similar lines, Sterner and coworkers showed that it is possible to generate a stable and compact, native-like or wild-type-like $(\beta/\alpha)_8$ barrel by fusing two halves from two different TIM barrels (HisA-N and HisF-C, and vice versa) or even by fusing two identical half-barrels (duplicated HisF-C, or HisF-C-C*) derived from the same protein. Such experiments were supported by further fine tuning of residue–residue interactions involved in folding and assembly of chain sections and/or subunits, selected through random mutagenesis [11].

Several *in vitro* experiments have shown that novel stable proteins can be formed through the combinatorial shuffling of polypeptide segments [12–15]. Attempts have also been made to form a folded $(\beta/\alpha)_8$ barrel called Octarellin, by the eightfold repetition of a single designed β/α unit derived from three $(\beta/\alpha)_8$ barrel proteins [16,17]. The latest of these *de novo* strategy-based proteins is Octarellin VI, made through the application of rational design and computational approaches using the Rosetta server [18]. Similarly, an ideal $(\beta/\alpha)_8$ barrel protein has been made by following a two-step *de novo* approach, in which an idealized artificial backbone of eight β-strands and eight α-helices, joined to each other by α/β turns, was first created. This was subsequently optimized by fitting in suitable amino acids with optimal rotameric conformations, in a sequence generated by an automated selection algorithm [19–21]. Recent work by Huang and coworkers explains the design of fourfold symmetrical $(\beta/\alpha)_8$ barrels guided by geometrical and chemical principles, opening further new avenues for making

custom-made enzymes based on the basic $(\beta/\alpha)_8$ barrel structural scaffold [22].

The extensive background information provided above supports the likely autonomous nature of folding of groups of beta/alpha units in most $(\beta/\alpha)_8$ barrels, and indicates that a high degree of plasticity could be associated with the assembly of such autonomous units into folded protein structures. Notwithstanding the successes of the experiments described above, it must be noted that they were all performed using either the same protein (reshuffling individual units) or by recombining proteins with very high structure and sequence similarity. In contrast, in the work presented in this paper, attempts were made to fuse half-barrel domains of two very distantly related glycosyl hydrolases; one a mesophile-derived endoglucanase/cellulase from *Clostridium cellulolyticum* (CelCCA) and the other a hyperthermophile-derived β-glucosidase from *Pyrococcus furiosus* (CelB) which have very little conservation of sequence, and which have also evolved to fold and function in temperature environments which are very different. Their similarity lies mainly in the fact that the catalytic dyad of residues in both enzymes consists of two glutamate residues which are present at nearly equivalent positions on loops 4 and 7, in the fourth and seventh beta/alpha units of both proteins. Our aim was to explore the hetero-assembly, folding, stability and presence of activity (if any) in chimeras formed by recombining half-barrel domains of CelCCA and CelB, with a particular interest in examining the presence of activity, as the two catalytic residues that are required to come together in the correct geometry to give rise to activity are provided by two different half-barrel domains sourced from organisms with no known common evolutionary ancestry.

Results

Protein design

As can be seen in Table 1, CelCCA (RCSB PDB ID 1EDG) is a monomeric mesophile endoglucanase from *C. cellulolyticum* with a molecular weight of 43 kDa, whereas CelB (RCSB PDB ID 3APG) is a homotetrameric hyperthermophile beta-glucosidase from *P. furiosus* with a molecular weight of 55 kDa. The catalytic sites of CelCCA and CelB each consist of two glutamate residues located at topologically equivalent positions in the two proteins, as shown in Fig. 1. The design of the chimeric enzyme constructs (hybrids) is shown in Fig. 2A,B in which the two halves of each designed protein are contributed by half-barrel domains sourced from CelB and CelCCA. The constructs are arrived at by identifying and demarcating the boundary separating the two half-barrel domains of each protein by identifying the last residue of the fourth β/α-unit of each parent, then nominally cleaving the polypeptide chain after this residue, creating an N-terminal half-barrel domain and a C-terminal half-barrel domain for each of the two proteins, and fusing the N-terminal half-barrel of CelB with the C-terminal half-barrel of CelCCA to create CelBCelCCA and, conversely, the N-terminal half-barrel of CelCCA with the C-terminal half-barrel of CelB to create CelCCA-CelB. The sequences of these chimeric proteins were submitted to the site hosting the software, RAPTORX, for model building [23]. The predicted structures are shown in Fig. 2C,D, respectively, for CelCCACelB and CelBCelCCA. The predicted structure of CelB-CelCCA was found to show maximum similarity with the sequence of chain A of 3APG (CelB), with 35% helical content, 13% beta strand content and 50% random coil content. The predicted structure of CelCCA-CelB showed maximum similarity with 3WDP and 1EDG, representing the PDB coordinates of a mutant of CelB and CelCCA respectively. The model built with 3WDP as template was selected for further analysis and found to have 42% helical content, 9% beta strand content and 48% random coil content. The coordinates of the models of the chimeric proteins thus obtained were analysed using the software, PDBEFOLD, for structure-based sequence alignments and comparisons involving the chimeric proteins and the native forms of CelB and CelCCA. A result of this exercise was that CelCCA was found to have 54% sequence similarity with CelB and the structures of the two proteins could be superimposed with a root-mean-square (rms) deviation of ~ 2.8 Å. Although the sequence

Table 1. Comparison of the characteristics of the CelCCA and CelB enzymes.

S. no.	Parameters	CelCCA	CelB
1.	Length	380 amino acids	472 amino acids
2.	Molecular weight (KDa)	43.10	54.6
3.	Isoelectric point (pI)	6.00	5.42
4.	Size of first half-barrel	217 amino acids	254 amino acids
5.	Size of second half-barrel	163 amino acids	218 amino acids
6.	Quaternary Structure	Monomer	Tetramer
7.	Residues constituting catalytic dyad	E170 (Loop 4) and E307 (Loop 7)	E207 (Loop 4) and E372 (Loop 7)

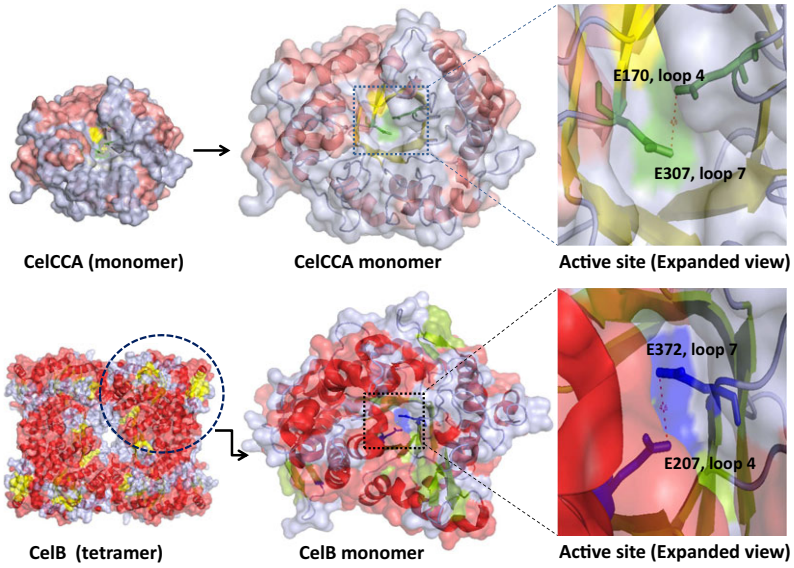

Fig. 1. An overview of the structural scaffolds of CelCCA and CelB showing the positions of the catalytic glutamate, E170 and E307 in CelCCA, and E207 and E372 in CelB.

similarity is reasonably high, giving rise to a correspondingly high structural similarity, it may be noted that the sequence identity between the two proteins is only 13%.

The estimates of the overall secondary structural contents of the proteins, as a fraction of their total length, were found to be 59% and 39%, respectively, when CelCCA and CelB were submitted as queries, as can be seen in Table 2. The secondary structure estimates suggest that the predicted structures of the chimeras share a greater degree of similarity with CelB (Table 2). The helices, strands and loops of the two proteins were compared in terms of sequence identity and conservation, and the results are shown in Table 3. The sequences of CelCCA and CelB were also subjected to conserved domain search through use of the conserved domain database (CDD) tool of NCBI [24]. In the CDD search results, CelCCA was matched with a cellulase belonging to glycosyl hydrolase family 5 (with residues 48–346, i.e. 267 residues in total, matched by the database) with accession number pfam00150. CelB, a beta-glucosidase, was matched with a protein with accession number COG2723 belonging to the beta-glucosidase/6-phospho beta-glucosidase/beta-galactosidase (carbohydrate transport and metabolism) class where residues 1–460 were analysed. The results of the CDD analyses can be seen in Fig. S1. The analyses and comparisons of the sequences (and particularly of the loop sequences) reveal that residues forming the catalytic dyad, i.e. the two glutamate residues, E170 (loop 4) and E307 (loop 7) in CelCCA, and E207 (loop 4) and E372 (loop7) in CelB, are present at topologically equivalent positions in both proteins, as can be seen in Table 3. The distance between the two glutamate residues is also ~ 3.8–4.1 Å in both proteins, as can be seen in Fig. 1.

An important point here is that there are extra loop sequences in the parent enzymes, some of which form secondary structural elements, and these were retained in the chimeras as can be seen in Fig. 2D, as these could be important for the folding, stability and oligomerization of the half-barrel domain structures derived from the parent enzymes. Figure 3 shows the sequences of CelB and CelCCA aligned with the sequences of CelB-CelCCA and CelCCACelB, along with the locations of the catalytic glutamates. Based on the literature, prediction analyses, and analyses of these structures and sequences, it is reasonable to conclude that the chimeras could potentially show enzymatic activity if (a) there were proper formation, and closure of the inner beta barrel core, through hydrophobic and hydrogen bonding interactions among the beta strands of the two half-barrels, and (b) if the catalytic residues were positioned, as a consequence, at almost equivalent positions mimicking the conformations of the native parent enzymes. Of course, given the facts that the native form of CelCCA is a monomer, and the native form of CelB is a homo-tetramer, it is also reasonable to expect that a largely folded chimeric enzyme construct could still engage in intermolecular interactions and thus end up forming large oligomers in unexpected ways, either owing to the tendency of the CelB parent (a natural homo-tetramer) to oligomerize or owing to the need to bury away some hydrophobic surfaces exposed through incomplete folding, or partial misfolding, in any section of the chimera's polypeptide chain.

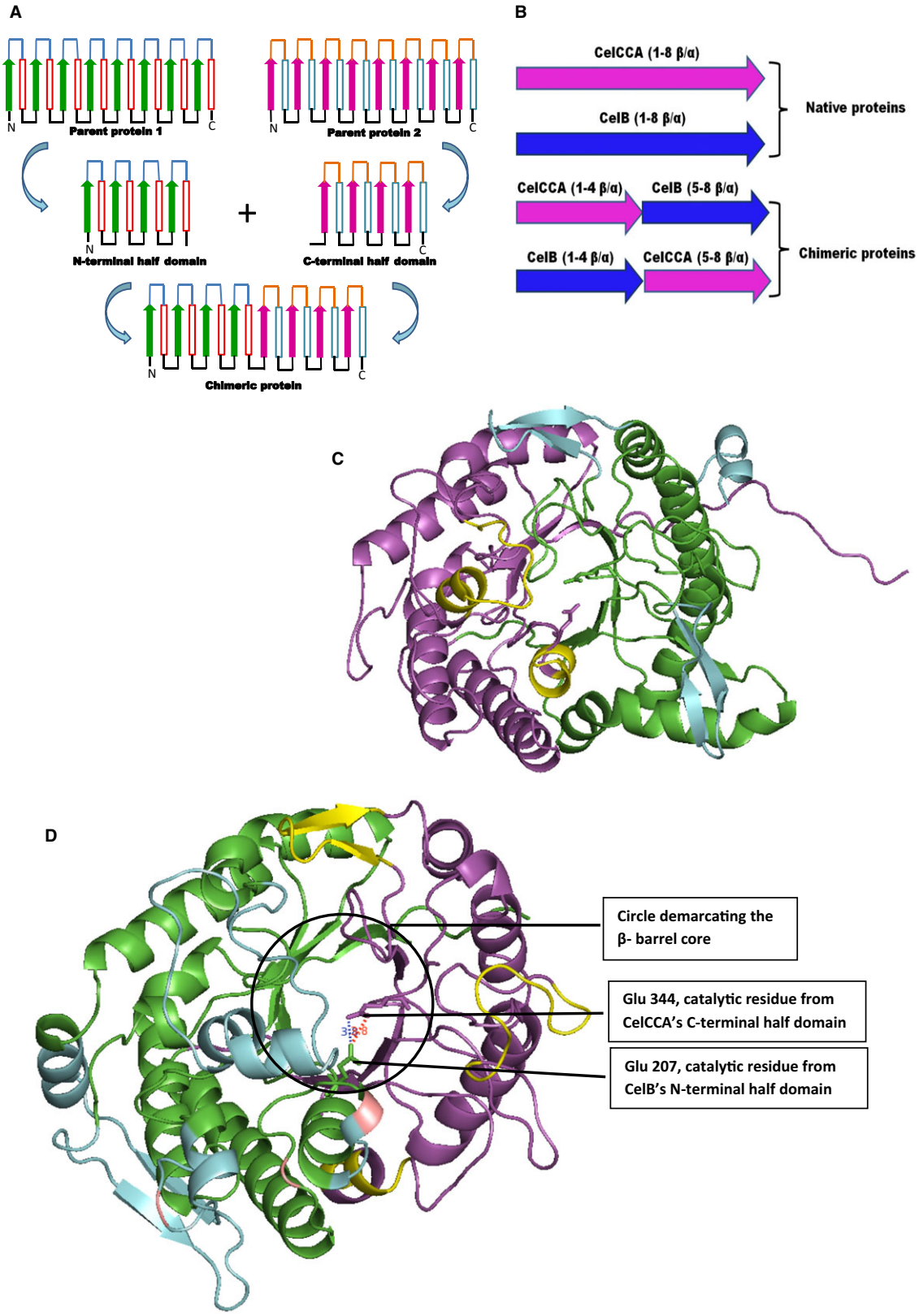
A
N
Parent protein 1
C
N
Parent protein 2
C
+
N
N-terminal half domain
C
C-terminal half domain
N
Chimeric protein
C
B
CelCCA (1-8 β/α)
CelB (1-8 β/α)
Native proteins
CelCCA (1-4 β/α)
CelB (5-8 β/α)
CelB (1-4 β/α)
CelCCA (5-8 β/α)
Chimeric proteins
C
D
Circle demarcating the β- barrel core
Glu 344, catalytic residue from CelCCA's C-terminal half domain
Glu 207, catalytic residue from CelB's N-terminal half domain

Fig. 2. (A) A general schematic figure depicting the design and construction of chimeric proteins consisting of half-barrels segments sourced from two different parent $(\beta/\alpha)_8$ barrel or TIM barrel proteins. (B) A figure depicting the design of construction of the two chimeras, CelCCACelB and CelBCelCCA, from two different parent proteins, CelCCA and CelB. (C) Structure of CelCCACelB predicted by the RaptorX server. The part shown in green is the half-barrel (N-terminal) domain contributed by CelB and the part shown in purple is the half-barrel (C-terminal) domain contributed by CelCCA, with catalytic glutamates shown in stick format. The extra sequences for CelB and CelCCA, respectively, are shown in sky-blue and yellow. (D) Structure of CelBCelCCA predicted by the RaptorX server. The part shown in green is the half-barrel (N-terminal) domain contributed by CelB and the part shown in purple is the half-barrel (C-terminal) contributed by CelCCA, with catalytic glutamates shown in stick format. The extra sequences for CelB and CelCCA, respectively, are shown in light blue and yellow. The residues sourced from CelCCA which are engaged in subunit–subunit interactions in native CelCCA are shown in pink, in this modelled chimera. The hydrophobic β-barrel core which is expected to be formed by the eight β-strands, with four strands contributed by each half domain, is shown encircled.

Table 2. Comparison of the sequence and structural similarities between the CelCCA and CelB proteins. The information provided was generated by submitting the structural coordinates of the modelled structures of the chimeras (obtained using i-TASSER) to the program, PDBEFOLD, together with the structural coordinates of the native proteins for which crystal structures are known (RCSB).

	Target					
	RMSD (Å)	Seq. Id. %	Sec. Str. %	RMSD (Å)	Seq. Id. %	Sec. Str.%
Query	CelCCA			CelB		
CelCCA	–	–	–	2.8–2.89	13	59
CelB	2.79–2.84	13	39	–	–	–
CelCCACelB	3.12	31	44	0.9	58	96
CelBCelCCA	2.77	40	53	0.69	68	95

Protein expression and yields

The chimeric proteins were purified under denaturing conditions and subjected to refolding, following confirmation of identity. The SDS/PAGE in Fig. 4A shows the parent proteins, CelB (lane 1) and CelCCA (lane 2), as well as the chimeric proteins, CelBCelCCA (lane 4) and CelCCACelB (lane 5), with apparent sizes similar to their estimated sizes of ~ 43, ~ 55, ~ 49 and ~ 51.5 kDa respectively. The yields of the native proteins, CelCCA and CelB, were ~ 30 mg per litre and 4–5 mg per litre of overnight grown culture, respectively, whereas for the chimeras, CelBCelCCA and CelCCACelB, the yields of the refolded proteins were 2–3 mg per litre and 1–1.5 mg per litre respectively.

Hydrodynamic size analyses of parent proteins and chimeras

To assess the oligomeric status of each protein, size exclusion (gel filtration) chromatography and dynamic light scattering studies were performed, following purification and refolding. The results are shown in Fig. 4B. Although the parent proteins, CelCCA and CelB, were found to elute, respectively, as monomer (16.2 mL) and tetramer (10.8 mL), from a Superdex-200 column with a bed volume of 24 mL, the two chimeric proteins, CelCCACelB, and CelBCelCCA, were found to elute mainly as large soluble oligomers (or soluble aggregates) in the void volume (~ 8 mL) of the same gel filtration column. As mentioned above, this could be expected to occur either because of the multimer-forming interactions of the half-barrel sourced from CelB, or because of the need to bury away hydrophobic surfaces which are incompletely buried through chain refolding. The chimeric proteins were also found to be subject to some degradation, with degraded peptides seen during chromatography (~ 16 mL, 20 mL), as seen in Fig. 4B. For further studies, using dynamic light scattering, only eluting oligomers were used. The results are shown in Fig. 4C. The diameters of CelB, CelCCA, and CelBCelCCA, were estimated to be 11.43, 5.79 and 22.12 nm, respectively, which correspond roughly to that expected of a tetramer, a monomer and an oligomer respectively. The CelCCACelB protein sample, however, showed two peaks, corresponding to 8.96 nm and 22.15 nm. It is probably thus a mixed population ranging from dimer to oligomer.

Refolding and structure formation in chimeras

As the chimeric proteins, CelCCACelB and CelBCelCCA, were derived through protein engineering, and not designed by nature to fold into native

Table 3. Comparison of the secondary structural elements in CelCCA and CelB. The sequences of structurally (and topologically) analogous beta-alpha loops, alpha helices and beta strands in the two proteins are shown, with information on sequence residue length (*L*), percentage homology (*H*%) and sequence identity (*I*%). Loops shown with underlined contribute catalytically important glutamate residues (shown in bold).

		CelCCA				CelB			
		Sequence	*L*	*H*%	*I*%	Sequence	*L*	*H*%	*I*%
Beta-alpha loops	1	NLGNTFDAFNGTNITNELDYE TSWSGIKTTK	31	51.61	22.58	SWSGFQFEMGLPGSEVESDWWVWVHD KENIASGLVSGDLPENGPA	45	35.55	15.55
	2	IPVSWHPHVSGSDYKISD	18	50	16.66	GIEWARIFPKPTFDVKVDVEKDEEGNIISVDVP	33	27.27	9.09
	3	NTHHDVDKVKGYFPSS	16	37.5	25	NLYHWPLPLWIHDPIAVRKLGPDRAPA	27	22.22	14.8
	4	**E**GMN**E**PRLVGHANEWWPELTNS	22	40.9	9	**E**PNVVYNQGYINLRSGFPPGYLS	23	39	8.6
	5	PGYVASP	7	71	14.2	AFAWHDPLA	9	55	11.11
	6	HAYCPWNFAGLAMADGGTNA WNINDS	26	50	19.25	NYYSRLVYGAKDGHLVPLPGYGFMSERGG FAKSGRPASDFGWEMY	45	28.88	11.11
	7	**E**CGAVDKNNL	10	50	30	**E**NGMADAADRY	11	45.45	27.27
	8	WDNNNFSGTGELFGFFDRRSC QFKFP	26	73.07	19.23	HWSLTDNYEWAQGFRMRFGLVYVDFETK KRYLR	33	57.57	15.15
Alpha helices	1	QMIDAIKQK	9	66.66	22.22	YWHLYKQDHDIAEKL	15	40	13.33
	2	VWMNRVQEVVNYCIDN	16	81.25	6.25	ESTIKELEKIANMEALEHYRKIYSDWK	27	48.14	3.7
	3	QYMASSKKYITSVWAQIAAR	20	65	–	GWLDEKTVVEFVKFAAFVAYHLDDL	25	52	–
	4	DVVDSINCINQLNQDFVNTVRAT	23	60.86	43.47	FEAAEKAKFNLIQAHIGAYDAIKE	24	58.33	4.16
	5	DGAT	4	50	25	EEYKDEVEEIRKKDYEFVTILHSK	24	8.33	4.16
	6	KDQSEVTWFMDNIYNK	16	68.75	12.5	PEGLENLLKYLNNAYE	16	68.75	12.5
	7	KTRVEYMSYYVAQAKAR	17	41.17	11.76	YRPHYLVSHLKAVYNAMK	18	38.88	11.11
	8	EIIDGMVKY	9	66.66	22.22	PSALVFREIATQKEIPEELAHLAD	24	25	8.33
Beta strands	1	GW	2	100	50	MFGY	4	50	25
	2	TVR	3	66.67	33.33	CIRG	4	50	25
	3	YVIL	4	75	–	KTFIL	5	60	–
	4	HLIF	4	25	–	MWSTMN	6	16.66	–
	5	LMC	3	66.67	–	SVGVIY	6	33.33	–
	6	IIVSV	5	60	25	IGV	3	100	33.33
	7	VIIG	4	75	50	MIIT	4	75	50
	8	LCIL	4	50	25	VRGYL	5	40	20

conformation, it is conceivable that some portions would have hydrophobic patches exposed, leading to a tendency to aggregate or to form oligomers. Similarly, interactions involving folded chain sections from CelB (which is a homo-tetramer) could also result in oligomer formation. Upon expression, the chimeras were both initially found located within inclusion bodies. They were extracted under denaturing conditions, purified through imidazole metal affinity chromatography (IMAC) and subjected to refolding in 20 mM Tris buffer, pH 8.0, supplemented with 5% glycerol. Glycerol stabilizes protein structure, and aids in masking exposed hydrophobic patches and preventing aggregation [25]. The refolded chimeras were examined for the formation of secondary structure through circular dichroism (CD) and intrinsic tryptophan fluorescence studies. The far-UV CD spectra of the chimeras are shown in Fig. 4D, along with the CD spectra of the parent CelB and CelCCA enzymes. The spectra show that the chimeras are well-folded, exhibiting a mixture of alpha helical and beta sheet structures like the parent enzymes, i.e. they display the two negative bands characteristic of alpha helices, with minima at 208 and 222 nm, and also clearly exhibit contributions from beta sheet structures which generally display a negative maximum at 216 nm, but are subsumed by the negative maxima owing to the alpha helices which show greater mean residue ellipticity than beta structures. The fluorescence spectra of the chimeras and parent enzymes are shown in Fig. 4E. The fluorescence emission spectra of the two chimeras indicate that they are well-folded, and that they bury away their aromatic residues. The wavelengths of maximal fluorescence emission (λ_{max}) of CelCCA and CelB are observed at 343 nm. The λ_{max} of CelBCelCCA and CelCCACelB, respectively, are observed at 333 and 338 nm. The parent enzymes as well as the chimeras thus display wavelengths of maximal emission at much shorter λ_{max} wavelengths than ~ 352 nm (which is the λ_{max} shown by tryptophan

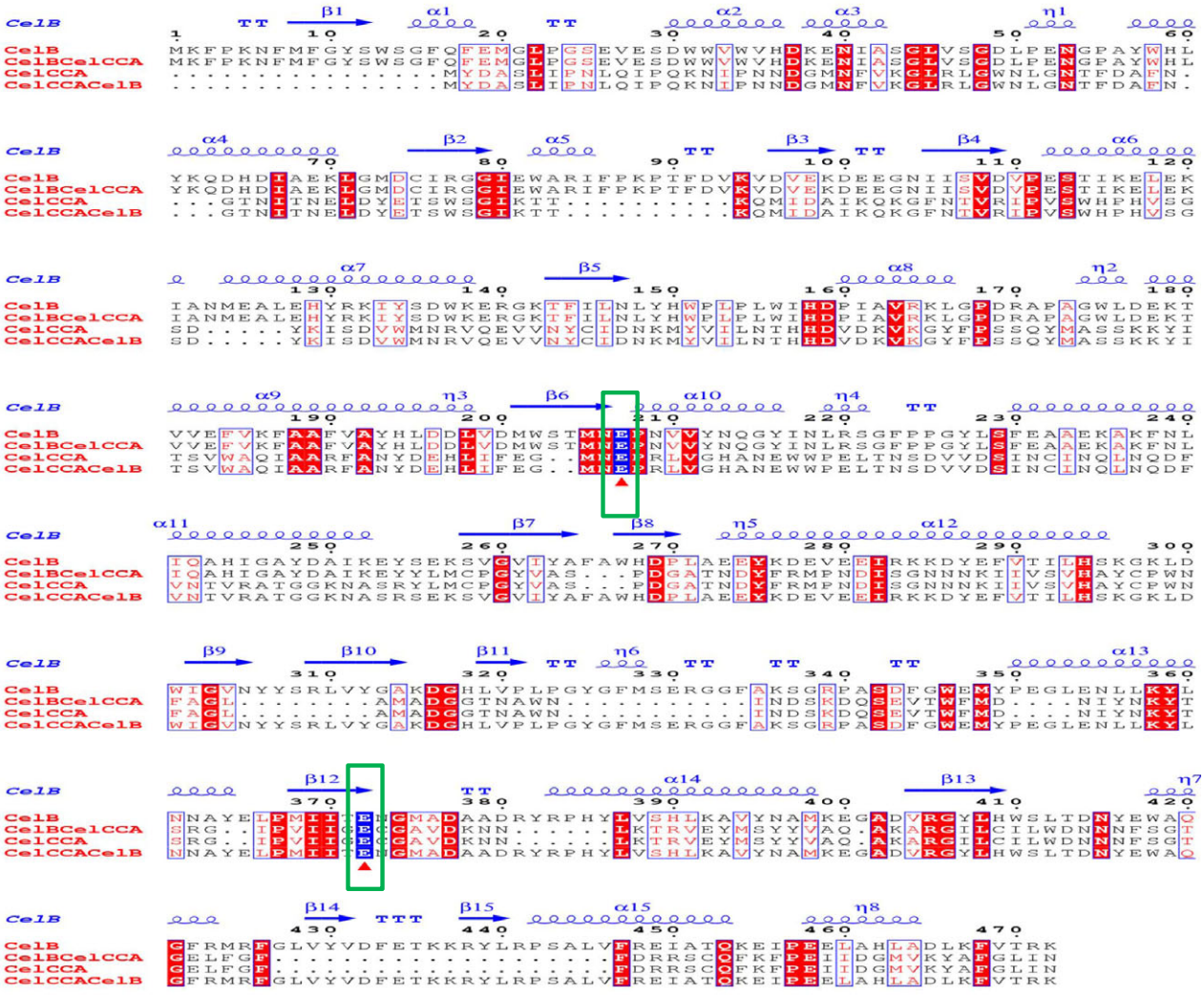

Fig. 3. Sequence alignment of the two parent proteins, CelCCA and CelB, and the two chimeric proteins, CelCCACelB and CelBCelCCA. The active site glutamate residues in CelCCA and CelB are highlighted in blue with a blue arrow underlining the residues at topologically equivalent positions; the likely active site glutamate residues in the chimeras are also similarly highlighted and underlined. Residues displaying identity are shown in red, in the alignment, while residues displaying similarity are shown in boxes. The figure was made using the software ESPRIPT3 [29].

residues exposed to aqueous solvent in a protein that fails to undergo folding).

Stability to thermal and chemical denaturation

The chimeras were examined for their thermal as well as chemical stability. For determining thermal stability, the proteins were heated from 25 to 90 °C at a scan rate of 1 °C·min^{-1}. The mesophilic parent enzyme, CelCCA, was found to have a melting temperature (T_m) of ~ 70 °C, whereas the hyperthermophile parent enzyme, CelB, did not show any denaturation over this range of temperatures, consistent with its evolutionary origin in a hyperthermophile archaeon, *P. furiosus*. The results of these analyses are shown in Fig. 5A, together with the analyses of the behaviour of the chimeras. Interestingly, the chimera, CelCCACelB, showed thermal stability similar to its mesophile parent, CelCCA, with a melting temperature of ~ 71 °C, although most of this chimera's sequence was derived from the thermophile parent enzyme, CelB. Equally interestingly, the other chimera, CelBCelCCA behaved similarly to its hyperthermophile parent, CelB, showing resistance to thermal denaturation, i.e. only insignificant spectral changes were seen to accompany heating, and these

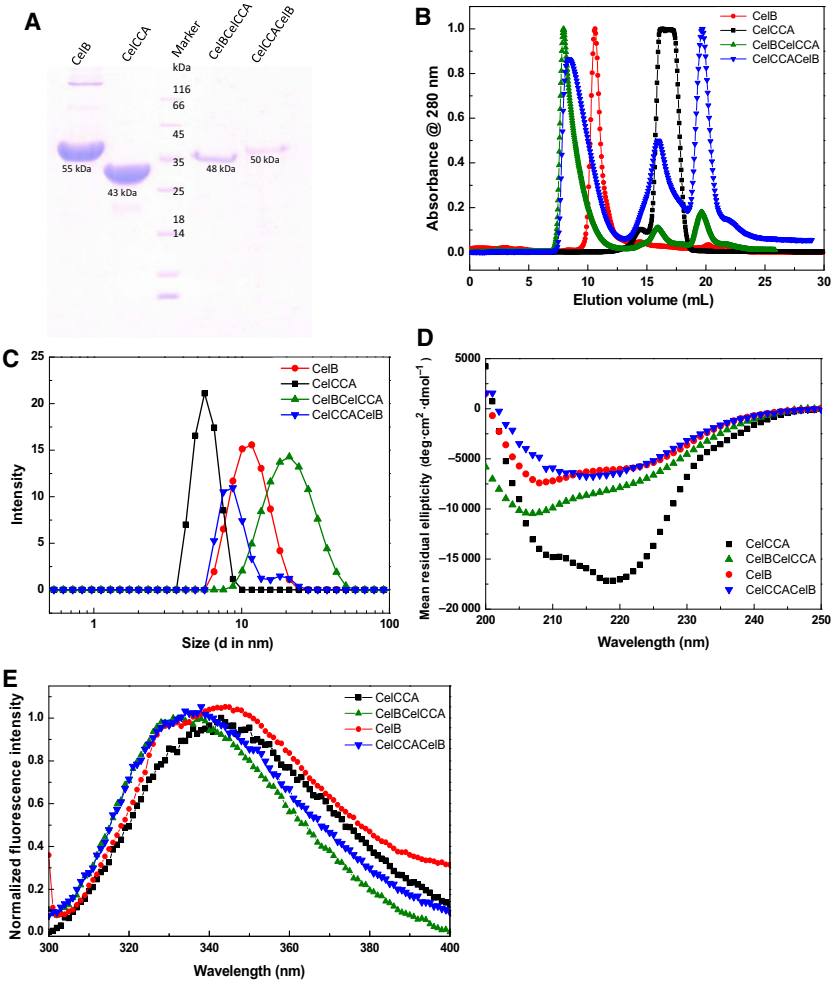

Fig. 4. (A) SDS/PAGE showing the purified proteins, CelB (~ 55 kDa, Lane 1), CelCCA (~ 43 kDa, Lane 2), protein molecular weight markers (Lane 3), CelBCelCCA (~ 49 kDa, Lane 4) and CelCCACelB (~ 52 kDa, Lane 5). (B) Gel filtration chromatographic profiles of the native parent proteins and the two chimeras. (C) Intensity versus size plots of the dynamic light scattering (DLS) data for the native parent proteins and the two chimeras. (D) Far-UV CD spectra of the native parent proteins and the two chimeras. (E) Fluorescence emission spectra of the native parent proteins and the two chimeras.

were fully reversed during subsequent cooling, as the temperature was varied from 25 °C to 90 °C and back to 25 °C. The data presented in Fig. 5A thus suggest that the N-terminal half-barrel determines the thermal stability of the entire chimeric enzyme, consistent with other reports in which the stability of TIM barrels appears to be determined by the N-terminal half-barrel [1]. Upon being subjected to chemical denaturation by guanidinium hydrochloride, the chimeras unfolded as given below: the highest operational stability was seen in CelB (C_m of 3.67 M) followed by CelBCelCCA (C_m of 2.6 M) and CelCCA (C_m of 2.0 M). The data from which these values are derived are shown in Fig. 5B. We could not assess the chemical denaturation profile of CelCCACelB because of the low yield and instability of the protein, which made it difficult to create multiple separate solutions of the protein of high enough concentration to perform reliable chemical denaturation experiments. The increased operational chemical stability of the CelBCelCCA (chimera) in comparison to CelCCA appears once again to owe to the N-terminal half-barrel domain being derived from the hyperthermophile parent enzyme, CelB. It may be noted that Park and Marqusee have shown that in order to analyse the stability of a multimeric protein, which is a challenging task (because of the intrinsic difficulty in handling the mathematical model for the folded multimer-unfolded monomer equilibrium), one can adopt the concept of effective stability, ΔGeff (_−RTlnKeff), where Keff is the equilibrium constant expressed in monomer units [26]. In the analysis of the effect of the denaturant upon ΔGeff, the dependence of effective stability on denaturant concentration (effective m-value) is simply the m-value of its monomeric unit when the protein is in folded state. However, in unfolded state, its stability depends on denaturant concentration with the m-value of its multimeric form, with the effective *m*-value at the C_m being a good approximation of the apparent m-value determined by fitting the equilibrium unfolding data from

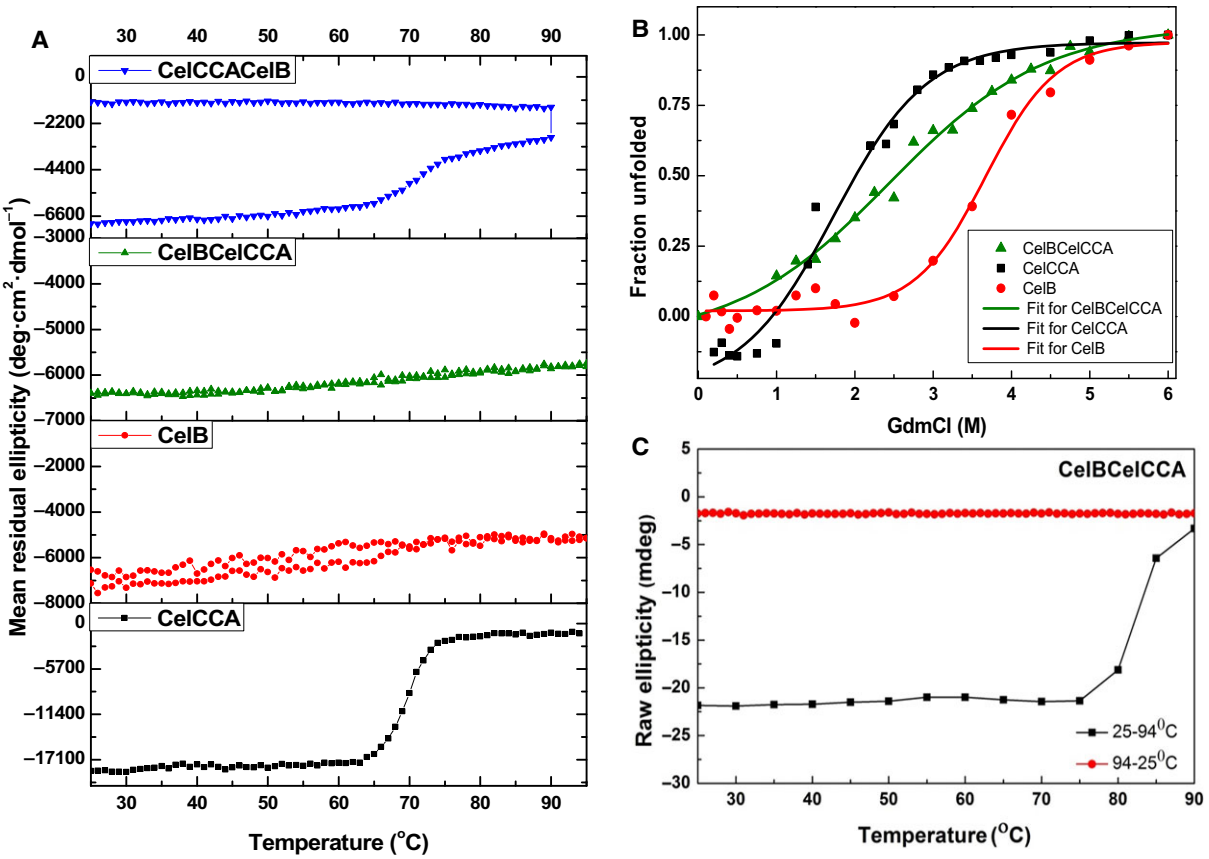

Fig. 5. (A) The thermal denaturation profiles of the parent proteins and the two chimeras, monitored in terms of the decrease in mean residue ellipticity (structure) at 222 nm with increase in temperature from 25 °C to 90/95 °C. For CelB and the two parent proteins, the ellipticity data are also plotted for decrease in temperature from 90/95 °C to 25 °C. (B) The chemical denaturation profiles for guanidinium hydrochloride induced equilibrium denaturation, monitoring ellipticity at 222 nm as a function of denaturant concentrations. (C) Thermal denaturation of CelBCelCCA, refolded by a different cosolute mediated refolding protocol. The change in raw ellipticity at 222 nm is plotted with increase (black) and decrease (red) in temperature, between 25 °C and 94 °C.

multimeric proteins with a two-state monomer model. Therefore, sometimes the fitting based on a two-state model can be used and the effective free energy, ΔGeff value (calculated by assuming a two-state model and monomers, during fitting), can be used as a parameter of oligomer stability [26]. However, although we performed such an analysis, the data are not presented here as many in the structural-biochemical community have reservations regarding the admissibility of this treatment. We choose to present only the data relating to the operational stability, in terms of the apparent values of the T_m and C_m.

In passing, with regard to the thermal unfolding data, the bulk of which was presented previously, in Fig. 5A, we also point out that in a separate set of experiments in which CelBCelCCA was refolded under somewhat different conditions (using arginine as a cosolute along with 7 M glycerol), the CelBCelCCA chimera which was otherwise observed to be extremely thermostable (undergoing no significant unfolding) did undergo substantial unfolding, displaying a melting temperature of ~ 85 °C, as shown in Fig. 5C, with such unfolding being irreversible. In other words, while the thermostability was still present, it was somewhat reduced when the protein was subjected to somewhat different conditions of refolding.

Detection of endoglucanase activity and its temperature and pH dependence

The two parent proteins, CelB and CelCCA, and the two chimeras, CelBCelCCA and CelCCACelB, were examined for the presence of endoglucanase activity. Both chimeras showed endoglucanase activity. All four proteins were examined for the presence of activity in the temperature range of 20 to 90 °C, and

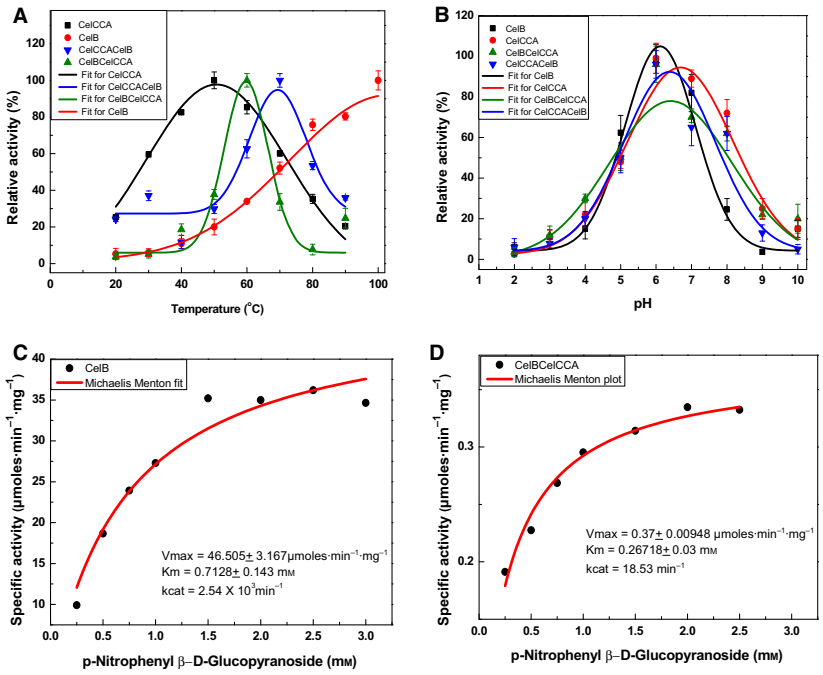

Fig. 6. (A) Changes in activity as a function of temperature, shown in relation to peak activity observed, in the two parent proteins and the two chimeras. (B) Changes in activity as a function of pH, shown in relation to peak activity observed, in the two parent proteins and the two chimeras. (C) Michaelis–Menten plot of β-glucosidase activity observed in the native parent protein, CelB. (D) Michaelis–Menten plot of β-glucosidase activity observed in the chimera, CelBCelCCA.

characterized in terms of activity units, wherever possible. Although CelCCA and CelB showed maximum activity at 50 °C and 95 °C, CelCCACelB and CelBCelCCA showed maximum activity at 60 °C and 70 °C respectively, as seen in Fig. 6A. The optimum pH for endoglucanase activity in all the four proteins was determined to lie in the pH range of 6.1–6.7, as seen in Fig. 6B. The endoglucanase activity units for the enzymes were also calculated and found to be 59.3, 12.03 and 15.36 units·mg^{-1} for CelCCA, CelB and CelBCelCCA, respectively, as shown in Table 4. Interestingly, all three enzymes showed activity in assays involving a total incubation of 1 h, but with CelCCACelB, no substantial activity could be detected with a 1-h incubation. Therefore, in characterizing the activity of the four proteins under different pH and temperature conditions, we performed overnight incubations with all four enzymes, and the results are reported in Table 5.

Table 4. Comparison of the endoglucanase activities of native CelCCA and CelB, and the two chimeric proteins, CelBCelCCA and CelCCACelB, determined using the DNSA assay and measurement of absorbance at 540 nm (conditions of the assay are described in the methods section). ND, 'not determined'.

Sample	OD_{540}	μM of glucose released	Specific activity (U·mg^{-1})
Blank	0.0368	0.197006	0
CelBCelCCA	0.0551	0.368682	15.36
CelB	0.0494	0.315209	12.03
CelCCA	0.1675	1.42313	59.30
CelCCACelB	ND	ND	ND

Table 5. Comparison of the presence or absence of endoglucanase and β-glucosidase activities in the native and chimeric proteins produced and studied.

Activity	CelCCA	CelB	CelBCelCCA	CelCCACelB
Endoglucanase	+	–/+ (up to 5/6)	+	+
Glucosidase	–	+	+	+

Detection of β-glucosidase activity

All four proteins were also examined for the presence of β-glucosidase activity (inherently present in CelB). Both chimeras were found to be active as β-glucosidases and, as already mentioned above, they were also found to be active as endoglucanases, as detailed in Table 6. However, as with the endoglucanase activity of CelCCACelB which we were not able to characterize in detail (i.e. through the determination of activity parameters), we were similarly not able to characterize β-glucosidase activity in CelCCACelB in detail, due to the fact that it is unstable and degrades rapidly and also because it is a slow enzyme and only shows its activity in overnight incubation reactions. Of the two parent enzymes, CelB and CelCCA, the parent enzyme, CelCCA, being an endoglucanase, did not

Table 6. Comparison of kinetic parameters determined for β-glucosidase activity in the native and chimeric proteins.

Protein	Specific activity (μmoles·min^{-1}·mg^{-1})	K_m (mM)	k_{cat} (min^{-1})	k_{cat}/K_m (M^{-1}·min^{-1})
CelB	46.505 ± 3.167	0.7128	2.54×10^3	3.56×10^6
CelCCA	nd	nd	Nd	nd
CelBCelCCA	0.37 ± 0.0095	0.267	18.53	6.9×10^4
CelCCACelB	ND	ND	ND	ND

show any detectable β-glucosidase activity, as was to be expected, even when reactions were overloaded with very high concentrations of the enzyme and overnight incubations were performed. Therefore, only CelB, the parent enzyme which is a β-glucosidase, and the chimera, CelBCelCCA, were subjected to experiments aimed at determining kinetic parameters, and characterizing β-glucosidase activity. The specific activity (V_{max}) of CelB (~ 46.5 μmoles·min^{-1}·mg^{-1}) was found to be 100 times more that of the chimera, CelBCelCCA (0.37 μmoles·min^{-1}·mg^{-1}). This is understandable given the nature and extent of the engineering performed. Also, the turnover number (k_{cat}) and second-order rate constant (k_{cat}/K_m) of the chimera, CelBCelCCA, are both two orders of magnitude less than that of CelB, as shown in Table 7.

Discussion

The eightfold structural symmetry inherent in the arrangement of the beta/alpha units in the TIM barrel fold has encouraged protein engineers and biochemists to modify TIM barrel enzymes, so as to understand the level of plasticity inherent in interactions responsible for maintaining the integrity of the fold [6,11,17], and also to throw light on the evolution of such proteins through gene duplication and fusion [1,11]. Some groups have even attempted to design TIM barrel proteins from first principles, with considerable success [27].

There are three aspects to the correct folding of any engineered, or redesigned, beta/alpha barrel protein: (a) the autonomy of substructure formation by individual beta/alpha units, or by groups of such units, (b) the availability of suitable registers for hydrogen bond formation between residues of designated adjacent beta strands and between the first and the eighth strand for closure of the barrel structure and (c) the plasticity of the hydrogen bonding and hydrophobic interactions comprising the core of the beta/alpha barrel [6]. Given the likely plasticity inherent in the formation of the core, an important characteristic that allows recombination between different TIM barrels is the presence of the active site residues mostly within loops of the β/α-units comprising the C-terminal face of the TIM barrel structure, derived from the sequences following the C-terminal ends of some of the eight strands making up the inner beta barrel of the structure.

To test whether stable and functional chimeras with significant activity can be made by gene-reassortment of half-barrel domains (with half-barrels sourced from unrelated/distantly related organisms), we made two chimeric constructs, CelBCelCCA and CelCCACelB, using the proteins, CelCCA (a mesophile endoglucanase) and CelB (a hyperthermophile beta-glucosidase). Of course, we did partially predispose these chimeras towards success in folding, structure formation and exhibition of endoglucanase and/or β-glucosidase activity, by initially selecting parent enzymes with considerable structural and sequence similarity, and with both enzyme activities, i.e. the endoglucanase and β-glucosidase activities, involving a pair of similarly disposed glutamate residues. Even so, the fact that the chimeras did fold at all, to display any activity, even if these were an order of magnitude (or two orders of magnitude) lower than the activities of the parent enzymes, turns out to be extremely satisfying because it establishes that even when two half-barrel domains are drawn from distantly related or unrelated enzymes, the assembly of such half-barrel structures pre-evolved in different

Table 7. Comparison of different biophysical characteristics of the native and chimeric proteins.

Protein	CelB	CelCCA	CelBCelCCA	CelCCACelB
Oligomeric Status	Tetramer	Monomer	Oligomer	Oligomer
Elution vol. (mL)	10.8	16.2	~ 8	~ 8
Diameter (nm)	11.43	5.79	22.12	8.96, 22.15
Wavelength of maximal emission (λ_{max}, nm)	343	343	333	338
Melting temperature (T_m, °C)	N.D	69.46	85.0	71.2
Molecular weight (kDa)	54.6	43.10	49	51.5
Denaturing concentration (C_m, M)	3.67	2.0	2.6	N.D
Temperature of optimal activity (T_{opt}, °C)	103	51	60	69.5
pH of optimal activity	6.11	6.68	6.4	6.37

organisms can occur. In addition, the fact that one of the engineered chimeras exhibited both activities to a substantial degree, suggests a mimicking of retrograde protein evolution wherein specific enzymes are made to work or behave as generalist enzymes with broader substrate specificity (two enzymes in this case), through the mixing and matching of half-barrel TIM domains from different glycosyl hydrolases. That the chimeras are quite stable and that they show degrees of cooperativity in unfolding with denaturants, and/or upon heating, is also satisfying. Perhaps most intriguing of all, besides the obtaining of reasonable levels of catalytic activity, is the observation that the folding of the N-terminal half-barrel domain determines the overall stability of the chimera, as CelBCelCCA (in which the N-terminal half-barrel is sourced from a hyperthermophile) is more stable than CelCCA (in which the N-terminal half-barrel is sourced from a mesophile). The results support the notion that the N-terminal half-barrel is generally more stable, and consequently, a more important determinant of stability in TIM barrels. Finally, we would like to point out that there is little scope for any of the observed activities being derived from any undetected contaminant enzymes present in the purified protein samples because the enzymes were purified through two stages of purification, i.e. affinity chromatography followed by gel filtration, for all the data presented here. In addition, further confirmation of the observed activities being associated with these enzymes was obtained by subjecting one sample already purified through affinity and gel filtration chromatography to a further round of affinity purification, following which beta-glucosidase activity was reconfirmed.

Materials and methods

Cloning, expression, purification and refolding

The genes encoding CelCCA and CelB were generated or amplified, respectively, from a synthetic gene (GenScript, New Jersey, USA) and from the genomic DNA of *P. furiosus*, by PCR methods. The primers used for amplifying CelCCA were 5′GAAATGGATCCATGTACGACGCGTC3′ (forward) and 5′GCCATTAAGCTTCTAGTTGATCAGACCGAACGCG3′ (reverse), while those for amplifying CelB were 5′AATAATCCATGGCCATGAAGTTCCCTAAAAAC3′ (forward) and 5′CGCACTCTCGAGCTTTCTTGTAACAA3′ (reverse). In the sequences of the primers shown within parentheses in the preceding sentences, the sequences of *Bam*H1 and *Hin*dIII restriction sites are underlined in the forward and the reverse primers used to amplify and clone the gene encoding CelCCA. Likewise, the *Nco*I, and *Xho*I restriction sites are underlined in the forward and reverse primers of the gene encoding CelB. The CelCCA gene was cloned into the pQE30 vector (Qiagen, Maryland, USA) between *Bam*H1 and *Hin*dIII sites and transformed into XL-1 Blue and M15[pRep4] cells, while CelB gene was cloned into the pET28c vector, between *Nco*I and *Xho*I sites and transformed into XL-1 Blue and also BL21Star(DE3)[pLysS] cells. The genes encoding the chimeric proteins (i.e. CelCCACelB and CelBCelCCA) were further created and amplified using splicing by overlap extension PCR (SOE-PCR), with primers designed to incorporate overlaps between the sequence of the 3′ end of N-terminal half of first half-barrel domain with the 5′ end of the C-terminal half-barrel domain. The following primers were used for the SOE-PCR: CelCCACelB SOE fwd (5′GGTAAAAATGCGTCTCGTTCAGAAAAATCCGTGG3′), CelCCACelB SOE rev (5′CCACGGATTTTTCTGAACGAGACGCATTTTTACC3′), CelBCelCCA SOE fwd (5′GCCATAAAAGAGTATTACCTGATGTGCCCG3′), CelBCelCCA SOE fwd (5′CGGGCACATCAGGTAATACTCTTTTATGGC3′). The gene boundaries of the half-barrel domains were demarcated by visualizing the structures of CelCCA and CelB in pyMOL and then by notionally dissecting each TIM barrel at the peptide bond following the last residue of the fourth beta/alpha unit, to generate two half-barrels. Two additional primers for CelB gene containing *Bam*H1 and *Hin*dIII sites were designed, viz. 5′GCAATCGGATCCATGAAGTTCCCAAAAAAC3′ (forward) and 5′GCAATGAAGCTTCTACTTTCTTGTAAC3′ (reverse), which were used for cloning the two chimeric constructs, CelCCACelB and CelBCelCCA. The chimeric genes were cloned into the pQE30 vector between *Bam*H1 and *Hin*dIII sites and transformed into XL-1 Blue cells, as was done with the gene encoding CelCCA. For protein expression, the plasmid carrying the cloned CelB gene was transformed into BL21Star(DE3)[plysS], while plasmids carrying CelCCA, CelBCelCCA and CelCCACelB were transformed into M15[pREP4] cells. Proteins were expressed by inoculating a secondary 1 L culture with primary culture grown overnight at 37 °C, induced with 1 mM isopropyl-thiogalactoside (IPTG) when the OD_{600} of the culture reached 0.6. The culture was grown for 5–6 h postinduction, after which cells were harvested by centrifuging at 11 305 ***g*** for 10 min. The supernatant was discarded and the cell pellet was resuspended in native lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0). To release cytoplasmic proteins, the cell suspension was sonicated at amplitude of 22 with alternating periods of 10 s on and off, for a total of 30 min, after which the suspension was subjected to high speed centrifugation at 30 910 ***g*** for 45 min. The proteins were purified by standard immobilized metal affinity chromatography (IMAC) procedures. The recombinant proteins contain a 6xHis tag (three of the proteins have this tag at the N terminus, while CelB has it at its C terminus) for binding to the Ni-NTA beads. The supernatant obtained after sonication was loaded onto a Ni-NTA column pre-equilibrated with lysis buffer, washed with wash buffer

(50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0), and protein was eluted in elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0). The eluted protein was subjected to dialysis against 20 mM Tris, pH 8.0, to remove imidazole. The protein was further repurified by gel filtration chromatography using a Superdex-200 10/300 GL column in 20 mM Tris, pH 8.0. The above protocol was used for CelCCA and CelB, as these were both produced in soluble form. The chimeric proteins (CelBCelCCA and CelCCACelB) were found to be present in inclusion bodies and purified under denaturing conditions with resuspension, instead, in denaturing lysis buffer (8 M urea, 50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0) prior to sonication. The supernatant obtained after centrifugation was diluted 1 : 1 with native lysis buffer and loaded on to a Ni-NTA column equilibrated with the 1 : 1 diluted buffer containing 4M urea. The column was then washed with wash buffer (2 M urea, 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0). The protein was eluted in elution buffer (0.5 M urea, 50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0). The eluted protein was then subjected to dialysis against 20 mM Tris pH 8.0 containing 5% glycerol. The protein was then subjected to gel filtration in 20 mM Tris, pH 8.0. At all stages of cloning and purification, DNA sequencing and mass-spectrometric confirmation of gene and protein sequence identity were performed.

UV-Vis absorption spectroscopy

Absorption spectra were collected in the range of 200–600 nm using Cary-50 UV-Vis spectrophotometer. The concentrations of proteins were estimated by measuring absorption at 280 nm and calculating concentration per predicted molar extinction coefficients of 87 750, 128 280, 101 800 and 114 230, respectively, for CelCCA, CelB, CelCCACelB and CelBCelCCA.

Size exclusion chromatography

Dialysed proteins were subjected to analytical gel filtration chromatography by allowing protein to be resolved on a Superdex-200 10/300 GL column on an AKTA Purifier-10 chromatographic system (GE Healthcare, Uppsala, Sweden), after equilibration in 20 mM Tris buffer of pH 8.0. A constant flow rate of 0.5 mL·min^{-1} was maintained throughout.

Fluorescence spectroscopy

Emission spectra of proteins were collected on a Cary Eclipse fluorimeter, using excitation at 295 nm, collecting emission spectra in the range of 300–400 nm, and using slit widths of 5 nm for both excitation and emission.

Circular dichroism spectroscopy

Far-UV CD spectra of proteins were collected on a Chirascan spectropolarimeter (Applied Photophysics, Leatherhead, UK) using a cuvette of 0.1 cm path length. Spectra were collected in the range of 200–250 nm. For thermal denaturation and cold renaturation experiments, protein samples were heated between temperatures of 25 and 95 °C (25–90 °C in the case of CelCCACelB) at a constant scan rate of 1 °C·min^{-1} and cooled from 95 to 25 °C at the same scan rate of 1 °C·min^{-1}.

Dynamic light scattering

Dynamic light scattering data were collected on a Zetasizer Nano ZS90 model instrument (Malvern Instruments, Malvern, UK). The gel filtration-eluted samples were further filtered through a 0.22 mm filter, prior to measurements. A total of 100 readings were averaged to generate intensity versus size plots.

Equilibrium denaturation (chemical) experiments

To test the chemical stability of the proteins, the purified proteins were subjected to chemical denaturation by adding different concentrations (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5 and 6.0 M) of guanidinium hydrochloride to the proteins. The samples were incubated overnight at 25 °C, following which far-UV CD spectra were collected. Herein, an equilibrium reaction is assumed to have occurred where the native protein (N) is converted to denatured state (D) through a two-state transition. The reading at 222 nm was noted in all cases and the amount of fraction unfolded was calculated by applying the following formula:

$$\text{Fraction unfolded(fD)} = Y - Y_N / Y_D - Y_N \quad (1)$$

where Y represents the observed value at that particular concentration, Y_N represents the value in native state and Y_D represents the value in denatured state. The fraction folded (fN) can be calculated by putting the value of fD in the following formula:

$$1 = \text{fD} + \text{fN} \quad \text{or} \quad \text{fN} = 1 - \text{fD} \quad (2)$$

These values were plotted against the respective denaturant concentrations so as to obtain a denaturation curve. The midpoint of the transition phase of the sigmoidal unfolding transition reaction was calculated as C_m, or the concentration at which 50% of the population was unfolded. These values were also put in the following formula to calculate the equilibrium constant (K_{eq}) values for the reaction (N↔D) and the following equation was used:

$$K_{eq} = \text{fD}/\text{fN} \quad (3)$$

DNSA assay for endoglucanase/cellulase activity

Endoglucanase activity was checked by the DNSA stopping assay [28]. For this, 2 μM solutions of control protein (CelCCA or CelB) and 10 μM solutions of chimeric proteins (CelCCACelB or CelBCelCCA) were taken in 50 mM citrate buffer of pH 4.5 containing 100 mM NaCl, and the volume was made up to 200 μL, following which 200 μL of 2% CM Cellulose solution was added. Tubes were incubated at different temperatures ranging from 20 °C to 80 °C for 12 h. Thereafter, 600 μL of DNSA reagent was added to each tube. The tubes were boiled for half an hour to see the development of brown colour which was being measured at 540 nm. On similar lines, the endoglucanase activity assay was done for determining pH optima, for which buffers of different pH ranging from pH 2.0 to 10.0 were used, with everything else remaining the same. For calculations of enzyme units, a 1-h assay was performed at 50 °C. In addition, a glucose standard curve was made and the amount of glucose released was calculated by putting the value of absorbance (from assay) into the equation of the linear plot obtained (from standard plot). This value was then put into the following equation to calculate the enzyme units:

$$\text{Enzyme activity } (\text{Units} \cdot \text{mL}^{-1}) = \frac{\text{micromoles of glucose released} \times \text{Vol. of reaction mix}}{\text{Reaction time} \times \text{Volume of diluted enzyme (mL)}} \quad (4)$$

$$\text{Enzyme activity } (\text{Units} \cdot \text{mg}^{-1}) = \frac{\text{Enzyme } (\text{Units} \cdot \text{mL}^{-1})}{\text{Concentration of protein}(\text{mg} \cdot \text{mL}^{-1})} \quad (5)$$

β-glucosidase enzyme activity kinetics

The β-glucosidase activity in native and chimeric enzymes was checked by monitoring the absorbance at 405 nm indicative of the release of p-nitrophenyl upon hydrolysis of β-1,4 glycosidic bond in the substrate, p-nitrophenyl-beta-D-glucopyranoside. For conducting the enzyme kinetics experiments, different concentrations of the substrate, viz. 0.25, 0.5, 1.0, 1.5, 2.0, 2.5 and 3.0 mM in 50 mM acetate buffer, pH 6.0, were incubated with 0.15 μM of CelB and 0.083 μM of CelBCelCCA at 60 °C. The rate of generation of yellow colour corresponding to release of p-nitrophenol was recorded at 405 nm with time (in minutes). The slopes of the absorbance versus time plots at different concentrations were first converted to specific activity (i.e. the amount of the product formed per minute per mg of enzyme) and then finally plotted against the substrate concentration to generate the Michaelis–Menten plot. The plot so obtained was used to calculate the various kinetic parameters, viz. V_{max}, k_{cat}, K_m and k_{cat}/K_m. For CelCCA, different concentrations of the enzyme ranging from 0.12 to 2.0 μM were used to study enzyme kinetics. However, we could not see a substantial formation of product within 1–4 h. However, if the same reaction mixture was incubated overnight, a slight yellow coloured product (p-nitrophenol) was seen to be produced when the enzyme was present, but not when enzyme was absent. This indicated that activity was present but it was too little for us to determine the kinetic parameters for CelCCA's β-glucosidase activity. The amount of CelCCACelB obtained after refolding was also very little and the refolded enzyme displayed poor stability. So, we were not able to do any enzyme kinetics experiments for CelCCACelB. However, the endpoint readings of the reaction were taken after incubating for 2 h at 37 °C.

Acknowledgements

PG, PS and PK acknowledge funding through a grant (MHRD-14-0064) for a centre of excellence in frontier areas of science and technology (COE-FAST) in the area of protein science, design and engineering, at IISER Mohali, from the Ministry of HRD, Government of India.

Author contributions

PS and PK performed the experiments. PG designed the project. PS and PG analysed the results and wrote the manuscript.

References

1 Hocker B, Jurgens C, Wilmanns M & Sterner R (2001) Stability, catalytic versatility and evolution of the (beta alpha)(8)-barrel fold. *Curr Opin Biotechnol* **12**, 376–381.

2 Maiti S, Luthra-Guptasarma M & Guptasarma P (2002) Phenomenological perspectives on the folding of beta/alpha-barrel domains through the modular formation and assembly of smaller structural elements. *IUBMB Life* **54**, 213–221.

3 Selvaraj S & Gromiha MM (1998) Importance of long-range interactions in (alpha/beta) 8 barrel fold. *J Protein Chem* **17**, 691–697.

4 Gunasekaran K, Eyles SJ, Hagler AT & Gierasch LM (2001) Keeping it in the family: folding studies of related proteins. *Curr Opin Struct Biol* **11**, 83–93.

5 Forsyth WR & Matthews CR (2002) Folding mechanism of indole-3-glycerol phosphate synthase from Sulfolobus solfataricus: a test of the conservation of folding mechanisms hypothesis in

(beta(alpha))(8) barrels. *J Mol Biol* **320**, 1119–1133.
6 Shukla A & Guptasarma P (2004) Folding of beta/alpha-unit scrambled forms of S. cerevisiae triosephosphate isomerase: evidence for autonomy of substructure formation and plasticity of hydrophobic and hydrogen bonding interactions in core of (beta/alpha)8-barrel. *Proteins* **55**, 548–557.
7 Fani R, Lio P, Chiarelli I & Bazzicalupo M (1994) The evolution of the histidine biosynthetic genes in prokaryotes: a common ancestor for the hisA and hisF genes. *J Mol Evol* **38**, 489–495.
8 Thoma R, Schwander M, Liebl W, Kirschner K & Sterner R (1998) A histidine gene cluster of the hyperthermophile *Thermotoga maritima*: sequence analysis and evolutionary significance. *Extremophiles* **2**, 379–389.
9 Lang D, Thoma R, Henn-Sax M, Sterner R & Wilmanns M (2000) Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science* **289**, 1546–1550.
10 Hocker B, Claren J & Sterner R (2004) Mimicking enzyme evolution by generating new (betaalpha)8-barrels from (betaalpha)4-half-barrels. *Proc Natl Acad Sci U S A* **101**, 16448–16453.
11 Seitz T, Bocola M, Claren J & Sterner R (2007) Stabilisation of a (betaalpha)8-barrel protein designed from identical half barrels. *J Mol Biol* **372**, 114–129.
12 de Bono S, Riechmann L, Girard E, Williams RL & Winter G (2005) A segment of cold shock protein directs the folding of a combinatorial protein. *Proc Natl Acad Sci U S A* **102**, 1396–1401.
13 Ochoa-Leyva A, Montero-Moran G, Saab-Rincon G, Brieba LG & Soberon X (2013) Alternative splice variants in TIM barrel proteins from human genome correlate with the structural and evolutionary modularity of this versatile protein fold. *PLoS One* **8**, e70582.
14 Riechmann L, Lavenir I, de Bono S & Winter G (2005) Folding and stability of a primitive protein. *J Mol Biol* **348**, 1261–1272.
15 Riechmann L & Winter G (2000) Novel folded protein domains generated by combinatorial shuffling of polypeptide segments. *Proc Natl Acad Sci U S A* **97**, 10068–10073.
16 Goraj K, Renard A & Martial JA (1990) Synthesis, purification and initial structural characterization of octarellin, a de novo polypeptide modelled on the alpha/beta-barrel proteins. *Protein Eng* **3**, 259–266.
17 Houbrechts A, Moreau B, Abagyan R, Mainfroid V, Preaux G, Lamproye A, Poncin A, Goormaghtigh E, Ruysschaert JM, Martial JA *et al.* (1995) Second-generation octarellins: two new de novo (beta/alpha)8 polypeptides designed for investigating the influence of beta-residue packing on the alpha/beta-barrel structure stability. *Protein Eng* **8**, 249–259.
18 Figueroa M, Oliveira N, Lejeune A, Kaufmann KW, Dorr BM, Matagne A, Martial JA, Meiler J & Van de Weerdt C (2013) Octarellin VI: using rosetta to design a putative artificial (beta/alpha)8 protein. *PLoS One* **8**, e71858.
19 Scheerlinck JP, Lasters I, Claessens M, De Maeyer M, Pio F, Delhaise P & Wodak SJ (1992) Recurrent alpha beta loop structures in TIM barrel motifs show a distinct pattern of conserved structural features. *Proteins* **12**, 299–313.
20 Dahiyat BI, Sarisky CA & Mayo SL (1997) De novo protein design: towards fully automated sequence selection. *J Mol Biol* **273**, 789–796.
21 Offredi F, Dubail F, Kischel P, Sarinski K, Stern AS, Van de Weerdt C, Hoch JC, Prosperi C, Francois JM, Mayo SL *et al.* (2003) De novo backbone and sequence design of an idealized alpha/beta-barrel protein: evidence of stable tertiary structure. *J Mol Biol* **325**, 163–174.
22 Huang PS, Feldmeier K, Parmeggiani F, Fernandez Velasco DA, Hocker B & Baker D (2016) De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat Chem Biol* **12**, 29–34.
23 Kallberg M, Wang H, Wang S, Peng J, Wang Z, Lu H & Xu J (2012) Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* **7**, 1511–1522.
24 Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI *et al.* (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res* **43**, D222–D226.
25 Mishra R, Bhat R & Seckler R (2007) Chemical chaperone-mediated protein folding: stabilization of P22 tailspike folding intermediates by glycerol. *Biol Chem* **388**, 797–804.
26 Park C & Marqusee S (2004) Analysis of the stability of multimeric proteins by effective DeltaG and effective m-values. *Protein Sci* **13**, 2553–2558.
27 Nagarajan D, Deka G & Rao M (2015) Design of symmetric TIM barrel proteins from first principles. *BMC Biochem* **16**, 18.
28 Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem* **31**, 426–428.
29 Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320–W324.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:
Fig. S1. NCBI's conserved domain database (CDD) search for (A) CelB which shows maximum similarity with BglB (Accession number: COG2723) and (B) CelCCA which shows maximum similarity with cellulase belonging to glycosyl hydrolase family 5 (Accession number: pfam00150).