

Structural plasticity of a designer protein sheds light on β -propeller protein evolution

Bram Mylemans^{1*} | Ina Laier^{1*} | Kenichi Kamata² |
Satoko Akashi² | Hiroki Noguchi¹ | Jeremy RH Tame²
| Arnout RD Voet¹

¹dpt Chemistry, KU Leuven, Leuven, Vlaams Brabant, 3000, Belgium

²Yokohama City University, Yokohama, Kanagawa, 230-0045, Japan

Correspondence

Arnout RD Voet, dpt Chemistry, KU Leuven, Leuven, Vlaams Brabant, 3000, Belgium
Email: arnout.voet@kuleuven.be

Funding information

Research Foundation Flanders, G0E4717N, G0F9316N, G051917N and GBM-D3229-ASP/17; OpenEye Scientific Software

β -propeller proteins are common in nature, where they are observed to adopt 4- to 10-fold internal rotational pseudo-symmetry. This size diversity can be explained by the evolutionary process of gene duplication and fusion. In this study we investigated a distorted β -propeller protein, an apparent intermediate between two symmetries. From this template, we created a perfectly symmetric 9-bladed β -propeller named Cake, using computational design and ancestral sequence reconstruction. The designed repeat sequence was found to be capable of generating both 8-fold and 9-fold propellers which are highly stable. Cake variants with 2 to 10 identical copies of the repeat sequence were characterised by X-ray crystallography and in solution. They were found to be highly stable, and to self-assemble into 8- or 9-fold symmetrical propellers. These findings show that the β -propeller fold allows sufficient structural plasticity to permit a given blade to assemble different forms, and provide a potential explanation for the wide diversity of repeat numbers observed in natural propeller proteins.

KEYWORDS

Protein design, Protein crystallography, Protein evolution

* Equally contributing authors.

1 | INTRODUCTION

In 1966 Dayhoff suggested, on the basis of internal sequence symmetry in ferredoxin, that protein domains have arisen from smaller peptides [1]. Since then, more evidence has been collected in support of this model of protein evolution. Many more proteins with internal symmetry have been discovered, and similar sequence motifs have been found in distinct and unrelated protein folds, indicating descent from a common ancestor. Although the exact mechanism by which complex protein domains may have evolved is unknown, it is clear that gene duplication and fusion are highly likely to have played a significant role [2, 3, 4]. While DNA recombination events may connect sequences encoding domains with unrelated folds, they are also highly likely to yield genes with tandemly repeated elements. Such repeats occur in 14% of proteins [5]. They can roughly be divided in two categories, repeats that form an open structure and those with a closed structure. In open structures, the repeats are arranged linearly or in a helix with a large pitch; each repeat only interacts with its direct neighbours, so the number of repeats is not constrained. Examples of this group include the Ankyrin and Leucine-Rich repeats. In contrast, the closed structures are arranged around a central axis; repeats share a hydrophobic core and the repeat number is fixed. Two common examples are the β -trefoil and TIM-barrel families [6, 7]. Several classes of repeat protein have been targeted for design, most notably the DARPins (from Ankyrin repeats), Symfoils (from β -trefoils) and a *de novo* TIM-barrel [8, 9, 10]

β -propeller domains share properties with both the open and closed repeat proteins. They have a closed structure, built from 4-stranded anti-parallel β -sheets, known as blades, fixed around a central symmetry axis. However the number of blades is not fixed and they possess a central channel instead of a hydrophobic core [3, 11]. β -propeller domains possess four to ten blades, and display a wide variety of functions, including enzymatic activity, genetic regulation and mediating protein-protein interactions [12]. The relative simplicity of their architecture and large sequence variation make them interesting evolutionary study tools and possible building blocks for protein nanotechnology.

In previous work to demonstrate the feasibility of evolution by duplication-fusion for this family of proteins, the process was reverse-engineered to create a perfectly symmetrical 6-bladed propeller protein called Pizza6 [13]. Proteins containing only two or three copies of the Pizza repeat sequence, Pizza2 and Pizza3, were observed to multimerize to reconstitute the original 6-bladed architecture. Proteins with different numbers of repeats were observed to self-assemble into larger complexes containing multiple 6-bladed propeller units. The 8-bladed propeller Tako8 was designed by a similar procedure, and its crystal structure agreed closely with the expected model, but truncated versions in this case (such as Tako2 and Tako4) failed to assemble stably as a β -propeller [14]. Studies of Pizza and Tako gave no suggestion that either repeat was capable of forming a structure with different symmetry from the template used in the design. This result was also observed in other studies and in other closed repeat domains. Sometimes the difference in the number of repeats and the symmetry of the fold is compensated by a domain swap as was demonstrated by the group of Tawfik. The 5-bladed propeller architecture of tachylectin-2 could be reconstructed by fragments carrying only two blades; one polypeptide was found to contribute one blade to each of two 5-bladed units, so that five polypeptide molecules fold into a single complex [15]. Similarly Lee and Blaber reported a symmetrical trefoil-fold protein, named Symfoil; a double repeat of the Symfoil motif was shown to trimerize into two complete three-bladed trefoil domains [9].

This use of domain-swapping to preserve local symmetry appears to be characteristic of repeat-motif proteins with a fixed repeat number and a hydrophobic core. As noted above however, β -propellers have neither, and thus one might expect that they would show greater structural plasticity, more like linear repeat proteins. This property could allow the removal of one blade without resulting in misfolding. It can be imagined that such malleability may greatly increase the ease of evolving a wide variety of β -propellers with either odd or even symmetry. Structural plasticity has recently been observed in the assembly of fragments containing two to five repeats of a 7-bladed WD40 propeller

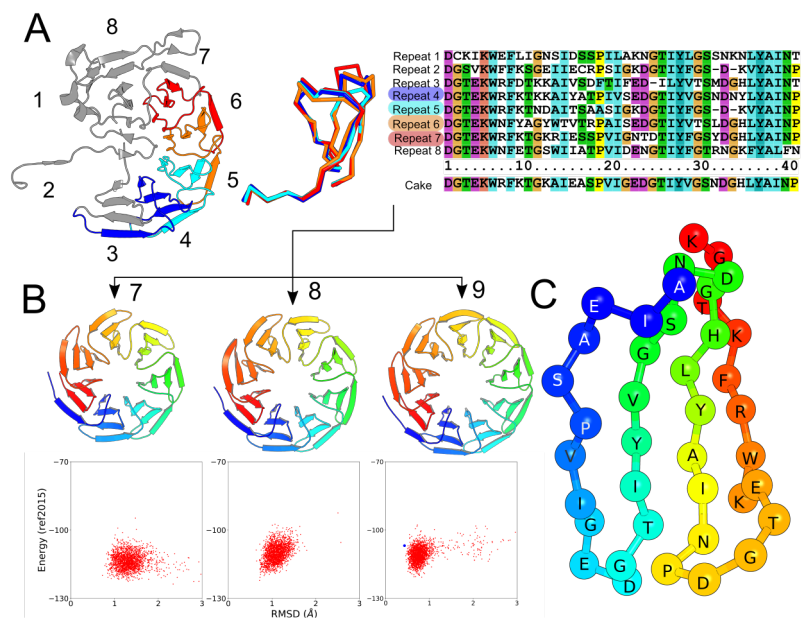


FIGURE 1 Overview of the computational design strategy. The incomplete propeller 3HXJ (A) was used as a template. The first, second, third and eighth repeats showed significant deviations from the desired topology, and were therefore excluded from the design. The remaining four repeats were symmetrically assembled with 7-, 8- or 9-fold symmetry using the SymDock protocol of Rosetta [17] (B). Based on these repeats, a total of 2000 ancestral repeat sequences were calculated using the FastML server [18]. These ancestral sequences were mapped onto the created backbones. The resulting structures were scored with both the Ref2015 and Talaris2014 scoring functions and additionally, the backbone deviation from the template was calculated. Plots of the rmsd against Ref2015 energy score are shown for all three symmetries. For each symmetry, one or two sequences with lowest energy or rmsd were selected for experimental testing. The sequence Cake9 is shown as the blue dot on the plot for models with 9-fold symmetry. This sequence is also shown as a C_{α} trace (C).

[16], but to date it has not been demonstrated in a single polypeptide that forms a complete propeller.

To investigate further the degree to which symmetry is determined by blade sequence or blade number, we applied ancestor reconstruction to a naturally occurring pseudo-symmetric β -propeller with a distorted architecture but high sequence similarity between blades. The resulting symmetric designer protein could self-assemble from fragments and adopt two different architectures depending on repeat number, providing more evidence for structural plasticity as an evolutionary pathway leading to propeller proteins with different symmetries.

2 | RESULTS

2.1 | Computational Design

β -propeller proteins in the PDB database were identified by using the 8-bladed propeller BamB (PDB:3P1L) [19] as a query for structure-based alignment with DALI [20]. Among the results was PDB model 3HXJ, the structure of an incomplete propeller domain, which was determined as part of a structural genomics project. The protein,

from *Methanococcus maripaludis*, remains poorly characterised but may possibly be an oxidoreductase. The sequence identity with the quinoprotein ethanol dehydrogenase from *Pseudomonas aeruginosa* (PDB entry 1FLG) [21] is only 28.5% however, so the true function of the protein remains unproven. RADAR analysis [22] revealed that the protein sequence consists of eight conserved repeats. The structure however is not a complete globular β -propeller, but reminiscent of a cake with a slice taken out. Although the sequence consists of eight highly similar repeats, only six blades are found in the structural model, with the remaining repeats adopting a different conformation.

It thus seems possible that this protein represents an evolutionary intermediate, a recent duplication or deletion having resulted in a protein whose repeat number does not match the symmetry favoured by the sequence. The parent sequence of the blades might be more compatible with a 7- or 9-bladed propeller instead of an 8-bladed one. This structure therefore presents an interesting template for engineering a perfectly symmetrical protein, to discover whether the presumed ancestral sequence preferred to fold with seven, eight or nine blades.

We used a design method described previously [23], based on ancestral sequence reconstruction and symmetrical templates. A schematic overview of the method is given in figure 1. Rosetta symmetric docking (SymDock) [17] was used to create 7-, 8- and 9-fold circular symmetrical backbones from the models of these four blades. The lowest energy template with correct domain folding was retained for each symmetry. All models and corresponding Energy - RMSD plots are shown in figure 2, in the last column a structural alignment between the best model for this symmetry and the template protein 3HXJ is shown. 2000 possible ancestral sequences were mapped onto the templates of different symmetry, yielded energy scores and the structural deviations from the starting structure. Five amino acid sequences were selected for experimental testing based on the energy and rmsd scores, one 7-bladed model, two 8-bladed models and two 9-bladed models. Next, these amino acid sequences were reverse-translated to nucleic acid sequences and codon optimised for expression in *E. coli* (Tab. S1). These five sequences were ordered pre-cloned into the pET28 expression vector to allow production of each protein under the control of IPTG.

2.2 | Protein purification, crystallisation and biophysical analysis

The proteins were expressed and purified using standard methods similar to those employed with Pizza and Tako [13, 14]. The 7-bladed propeller construct failed to express, and both 8-bladed constructs aggregated during purification. The two 9-bladed constructs could be purified, and were subjected to crystallisation trials. Only one of these yielded crystals (that diffracted to 1.43 Å) and was used for further studies. This protein was named Cake9. The structure was solved using molecular replacement with the computational model. In figure 3, the obtained crystal structure is aligned on the computational model, the rmsd over all C α atoms is 1.1 Å.

Genes were constructed to express proteins consisting of different numbers of tandem copies (1 to 10) of the same sequence motif, respectively referred as Cake1 to Cake10. These Cake derivative proteins could all be expressed and purified, but Cake1 did not appear to be monodisperse and was not investigated further. CD spectroscopy shown in figure 4A indicated that each protein was well-folded and highly stable to thermal denaturation. From the thermal denaturation curves in figure 4B Cake7, Cake8, Cake9 and Cake10 were found to be more stable than the smaller proteins such as Cake2 and Cake3, which were the least stable.

With the exception of Cake7, all the proteins crystallised (Tab. S2) and the X-ray structures were solved at high resolution without any conformational outliers. (Tab. S3). The resulting crystal structures are shown in figure 5, the crystals and crystal packing are shown in figure 6. Cake3 trimerized to reconstitute the 9-bladed architecture. Cake10, Cake5 and Cake2 all adopted a 9-bladed architecture with one disordered repeat motif. Cake6 formed a domain-swapped trimer, with two 9-bladed propellers. Cake4 and Cake8 were found to adopt an 8-bladed architecture, with Cake4 forming a dimer. The oligomerization state in solution was analysed by size exclusion chromatography, see

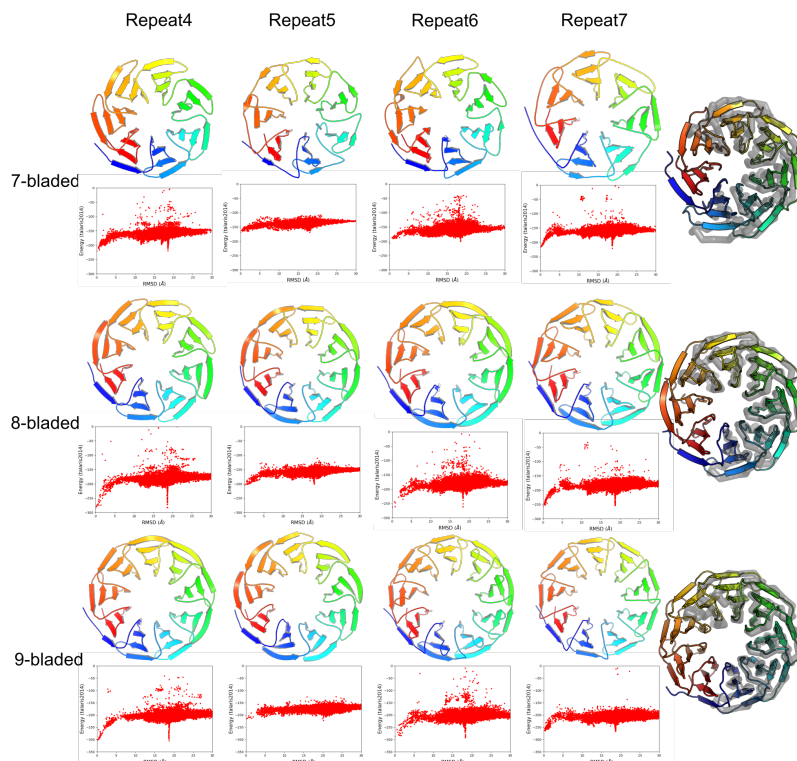


FIGURE 2 Overview of all SymDock templates. Four repeats were used to create a symmetrical template on which ancestral sequences could be mapped. A template was made with each symmetry. The best scoring template is shown as cartoon each time. All other generated structures were compared to this best scoring one to generate an RMSdD Energy plot. The templates based on repeat 4 were chosen for all symmetries because they had the lowest energy and a clear folding funnel. On the right a cartoon of the chosen template is shown on top the natural template protein 3HXJ as C_{α} ribbon

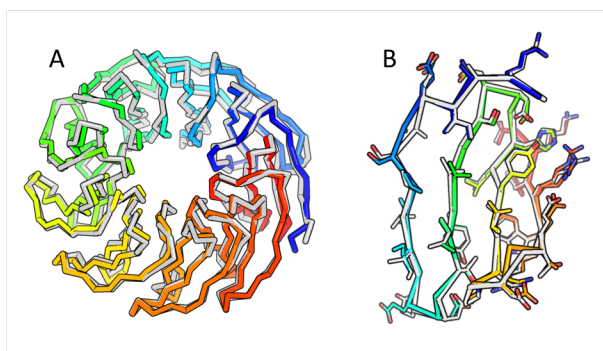


FIGURE 3 Superposition of the designed structure (white) with the crystal structure of Cake9, The rmsd between them is 1.1 Å (A). Superposition of a single blade with the side chains shown (B).

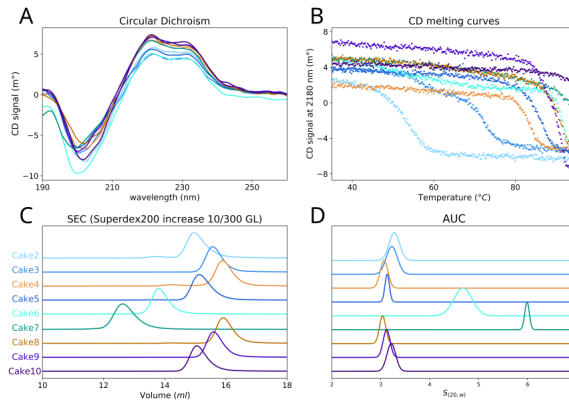


FIGURE 4 Biophysical characterisation of Cake variants. The circular dichroism spectra of all proteins are nearly identical (A). Melting points were determined by following changes in the CD signal at 218 nm (B). All the proteins are stable up to at least 50° C, but the shorter repeats have lower melting points. Cake8 and Cake9 have very similar stability, despite adopting propeller forms with different symmetry. Size exclusion chromatography (SEC) (C) and analytical ultracentrifugation (AUC) (D) data correspond well to the crystal structures. Cake4 and Cake8 form the smallest structures. Cake2, Cake5 and Cake10 proteins are slightly larger than Cake3 and Cake9 because of their one unstructured repeat. Cake6 forms a larger trimeric complex with a total of 18 repeats making two 9-bladed propellers as confirmed by X-ray diffraction. The data for Cake7 suggest it forms a tetramer with a total of 28 repeats, implying it forms three linked 9-bladed propellers.

figure 4C, analytical ultracentrifugation, see figure 4D and nano-electrospray mass spectrometry (Tab. S4).

3 | DISCUSSION

Here we have attempted to reform a 8-fold protein repeat sequence that adopts a distorted propeller structure. This protein could represent an evolutionary intermediate, whose underlying repeat sequence is more compatible with a 7- or 9-bladed propeller. Perfectly symmetrical proteins were computationally designed from this template, in an attempt to retrace its evolutionary pathway and discover the preferred form of the putative ancestor. A 9-bladed protein, Cake9, was successfully produced, whose high-resolution crystal structure closely matched the design model. Although many β -propeller structures have been reported, there is only one 9-bladed propeller known so far, eIF3 (PDB:4NOX) [24]. eIF3 is classified as a WD40 family protein, with a characteristic tryptophan-aspartate interaction in each blade, but Cake9 lacks this feature and there is no significant sequence or structural similarity between the two proteins. While the design of a 9-bladed propeller was successful, all 8-bladed designs aggregated and the 7-bladed design failed to express. This could indicate that the 9-bladed architecture is preferred, even though the natural protein only possesses eight repeats. **A sequence and structural alignment of the successful cake sequence and failed 8-bladed designs is shown in figure 7A. The sequence difference between the successful and aggregating sequences is only minor. An additional charge in the Cake protein might explain why it remained in solution instead of forming aggregates. Only one of the two 9 bladed designs could be crystallised. From the Size exclusion chromatogram shown in figure 7B, we speculate that the unsuccessful 9bestE protein might adopt an 8-fold propeller architecture and is thus more difficult to crystallise.**

Removing one blade from Cake9 gave a perfectly symmetrical monomeric 8-bladed propeller (Cake8) with similar

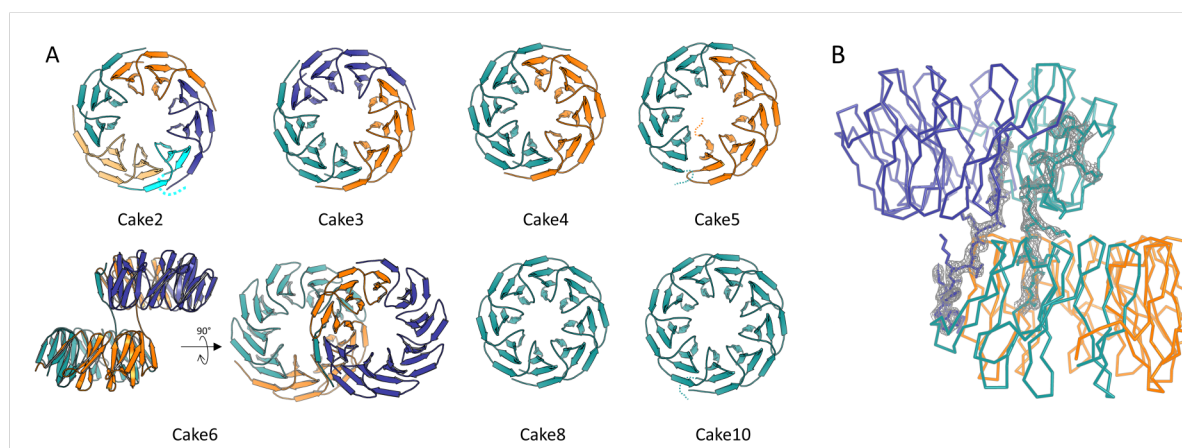


FIGURE 5 A structural overview of all other repeat mutants. Cake2, Cake3 and Cake10 all adopted the single 9-bladed propeller structures, leaving one blade unstructured in Cake2 and Cake10. Cake6 formed two 9-bladed propellers linked by a domain swap. Cake4 and Cake8 folded as eight-bladed propellers (A). The electron density around the connectors of Cake6 is shown as a 2Fo-Fc map at sigma level 1 (B).

stability. This is to our knowledge the first instance of a monomeric protein that adopts a different stable and folded structure if the sequence is extended or reduced. While oligomeric proteins are generally found to have rigid, fixed subunit stoichiometry, it has been observed that the 11-meric TRAP protein is able to form 12-mer rings in the presence of its partner anti-TRAP protein [25]. A similar shift in symmetry is brought about by making tandem copies of the protein [26]. Similar structural plasticity in fragment assembly has also been shown in β -propellers in a study last year by the group of Lupas, which addressed the question of structural diversity in propellers with different numbers of identical blades [16]. Stable proteins were created by expressing tandem copies of a blade, based on the natural 7-bladed propeller Pkwa, as a single polypeptide. 3-, 4-, and 5-fold repeats each gave dimeric proteins, whereas the 2-fold repeat protein gave a tetramer with 8-blades in total. Unfortunately none of these proteins could be crystallised. Different results were obtained with the protein WRAP, which contains 14 nearly identical blades that form two 7-bladed propellers. In this case, 2-, 3-, 4-, and 5-bladed fragments did crystallise. The 2- and 4-bladed polypeptides both formed an 8-bladed complex, whereas the 3-bladed polypeptide produced a trimeric 9-bladed protein. This blade sequence is therefore also capable of adapting to different symmetry, and yields different structures depending on the number of tandem copies in the polypeptide. The 5-bladed fragment produced an unexpected result, with one subunit forming five identical, normal folded blades, and its partner making only four, which were inverted with respect to the first subunit. In this case the plasticity extended beyond a simple expansion of the propeller, and accommodated an excessive number of subdomains by a dramatic rearrangement of the dimer. The multimeric WRAP assemblies were not symmetrically arranged around a central axis and it seems that the interface between fragments also provides some structural flexibility. Although fragments could assemble into 8- and 9-bladed propellers, it is unclear whether monomeric constructs would adopt the same conformations [16].

Previous experiments on the artificial symmetrical propeller proteins Pizza6 and Tako8 showed a notable reluctance to adapt their architectural symmetry if their sequences are cut or extended by one repeat; Tako7 failed to express and Tako9 gave an 8-fold propeller [14]. Since these proteins were designed from pseudo-symmetrical β -propeller templates, the resulting sequence motifs may inherently favour the original symmetry very strongly.

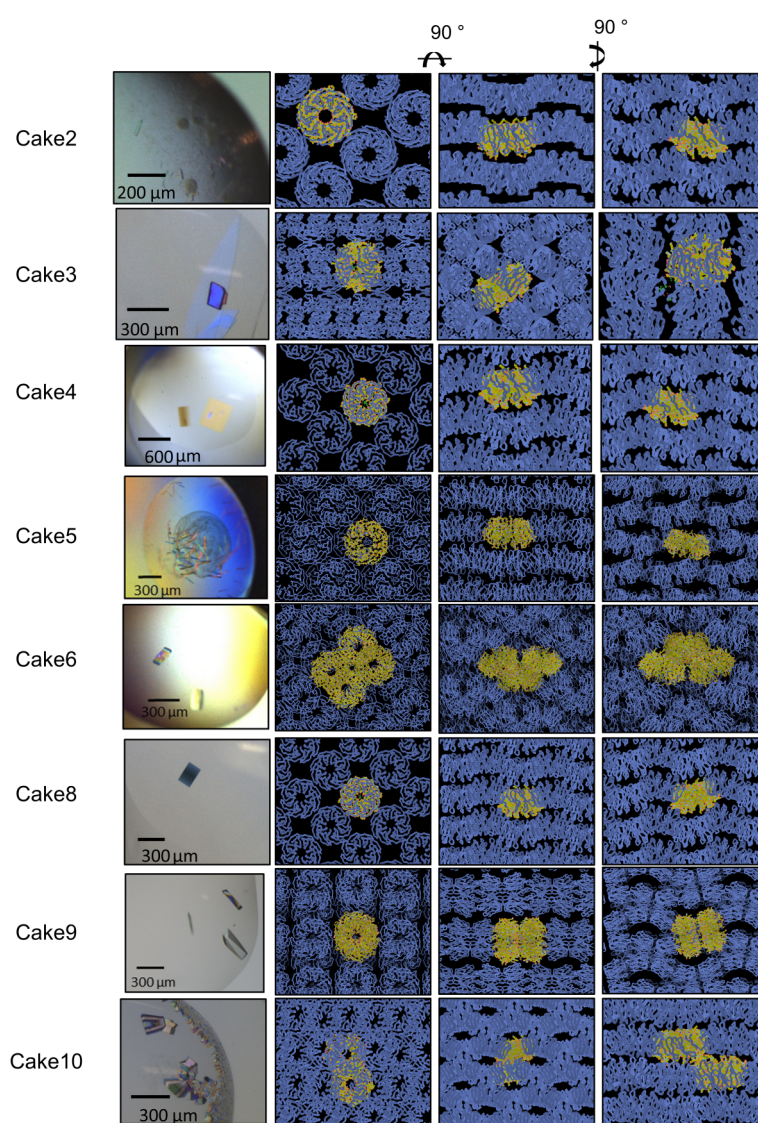


FIGURE 6 Pictures of all the diffracted crystals and the corresponding crystal packing viewed from the three main axes.

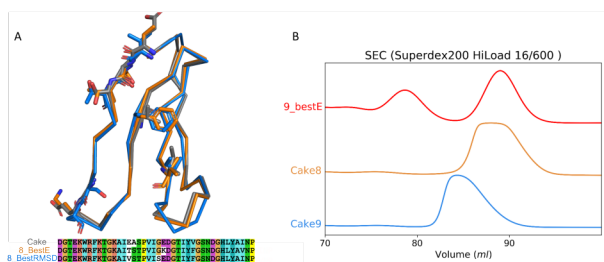


FIGURE 7 Comparison of Cake with the unsuccessful designs. A sequence and structural alignment of Cake to the 8-bladed proteins. Cake is coloured gray, 8bestE orange and 8bestRMSD blue. The amino acids where the proteins differ are shown as sidechains the other parts of the protein as ribbon. The differences between the proteins are minor, most notably is the lack of glutamate on position 15 and the presence of phenylalanine on position 28. The additional charge could make Cake more stable while the larger phenylalanine could hinder core packing (A). Size exclusion chromatogram of 9bestE, Cake8 and Cake9. 9bestE corresponds more with Cake8 than with Cake9 indicating it might favour the 8-bladed propeller (B).

We also studied different repeat fragments from two repeats (Cake2) to ten repeats (Cake10) for their ability to assemble. While Cake4 adopted the 8-bladed propeller form, most fragments favoured the 9-bladed propeller, as can be seen in figure 5. This fold was even preferred if one blade remained unstructured, as was the case for Cake2, Cake5 and Cake10. Cake6 formed a trimer by domain swapping to create a double 9-bladed propeller (fig. S5), in a similar manner to the two interconnected 5-bladed tachylectin-2 complex observed by the Tawfik group previously [27]. From the biophysical data indicating that Cake7 adopted a tetramer, we assume the complex formed a total of three 9-bladed propellers, but this could not be verified by X-ray crystallography. The crystal structures all exhibit a perfectly symmetrical architecture. All proteins had a high thermostability, interestingly the Cake10 variant has the highest resistance. We theorise this could be due to the unstructured blade allowing for the adaptation of a different conformation.

Interestingly the inner ring of either Cake toroid carries uncompensated negative charges, while positive charges are situated around the outer rim. In the case of Tako8 protein, similar separation of charge led to instability at low salt concentration, and inability to assemble from fragments carrying two or four blades, but this was not observed in the Cake protein. One explanation may be the more hydrophobic core of the Cake proteins, compared to the hydrophilic core of Tako8, with conserved hydrogen-bond networks between Trp-Ser-Asp-His residues of the adjacent blades which are more stable in a solute-exposed environment. **The differences between the Cake8 and Tako8 protein are shown in figure 8.**

The different folds of the natural template, PDB:3HXJ and the closest sequence and structural relative found by BLASTP and DALI, BamB (PDB:4XGA) [20] are compared in figure 9, which shows an alignment of the whole protein and alignments of all the individual blades. We also calculated the distance to the central axis and between neighbouring blades, the tilt angle between the blade and the central axis and the dihedral twist angle between the central axis and each blade. The rotational angle was calculated by CE-SYMM, which can identify internal symmetry in proteins [29]. The individual blades of Cake8 and Cake9 are very similar having only 0.2 Å rmsd. They show no significant changes in the parameters except the inter-blade distance and the tilt and twist angles. The small changes in twist and roll angle appear to be caused by the slight deviations observed in the connecting loop region, which allows the Cake motif to compensate for the loss or addition of a single blade. Interestingly the distance parameters, and more notably the rotational angles, of blades 4, 5, 6 and 7 in PDB model 3HXJ lie in-between those of the Cake8

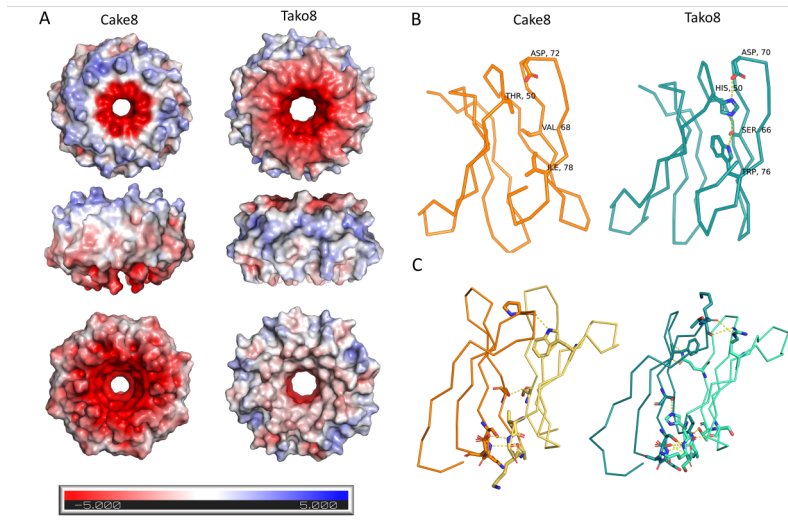


FIGURE 8 Comparison of electrostatics of Cake8 and Tako8 (PDB: 6G6M) calculated with Adaptive Poisson-Boltzmann Solver in PyMOL. In both proteins, the inner channel is strongly negatively charged while the outside contains multiple positive charges. In case of Cake8, the negative charge is most prevalent on one side of the channel (A). In Tako8 a strongly conserved network consisting of Trp-Ser-His-Asp is present, this motive is absent in Cake8 which has a more hydrophobic core (B). In between the adjacent blades of the Cake protein, the hydrogen bonds are formed near the polar exterior. A similar but larger network can be found in Tako8 (C).

TABLE 1 Structural comparison of blade geometry between designed and natural proteins

	Cake8	Cake9	3HXJ	BamB
a	6.6 ± 0.1	8.4 ± 0.1	7.4 ± 0.4	7.0 ± 0.9
b	10.9 ± 0.1	13.0 ± 0.1	11.3 ± 0.3	10.9 ± 0.7
c	5.1 ± 0.1	5.7 ± 0.1	5.4 ± 0.5	5.6 ± 0.6
d	8.3 ± 0.1	8.8 ± 0.2	8.4 ± 0.9	8.4 ± 1.1
e	13.1 ± 0.1	12.9 ± 0.1	13.3 ± 0.9	13.7 ± 1.6
f	14.6 ± 0.1	14.5 ± 0.1	14.4 ± 0.2	14.9 ± 0.6
α	45.0°	40.0°	42.6°	45.0°
β	$11.0^\circ \pm 0.3$	$14.0^\circ \pm 0.4$	$10^\circ \pm 3$	$4^\circ \pm 4$
γ	$28.3^\circ \pm 0.5$	$32.1^\circ \pm 0.4$	$26^\circ \pm 3$	$24^\circ \pm 6$

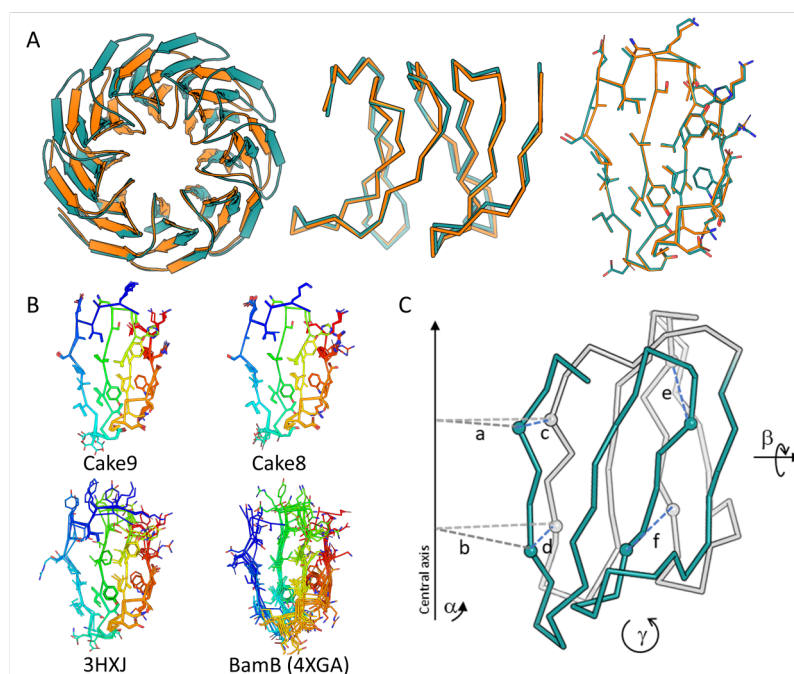


FIGURE 9 Comparison between Cake8, Cake9, the design template and the closest natural relative. Alignments of Cake8 (Orange) and Cake9 (teal), as the full protein, two subsequent repeats and a single repeat with side chains shown (A). This shows that the overall structure is different but the individual blades are almost identical. Alignments of all blades of Cake8, Cake9, 3HXJ used for the design and the closest relative BamB (4XGA) [28] show that the conformations of side-chains pointing into the proteins core are conserved (B). Side-chains facing the solvent are flexible in both proteins. To compare, the propeller structures average parameters were calculated in both designed and natural structures see table 1 (C). (a), (b) the distance between reference atom in inner strand and central axis. (c), (d), (e), (f) distance between two subsequent inner or outer strands respectively. (α) rotational angle as was calculated with CE-Sym [29]. (β) dihedral angle between inner strand and central axis. (γ) tilt of the blade against the central axis.

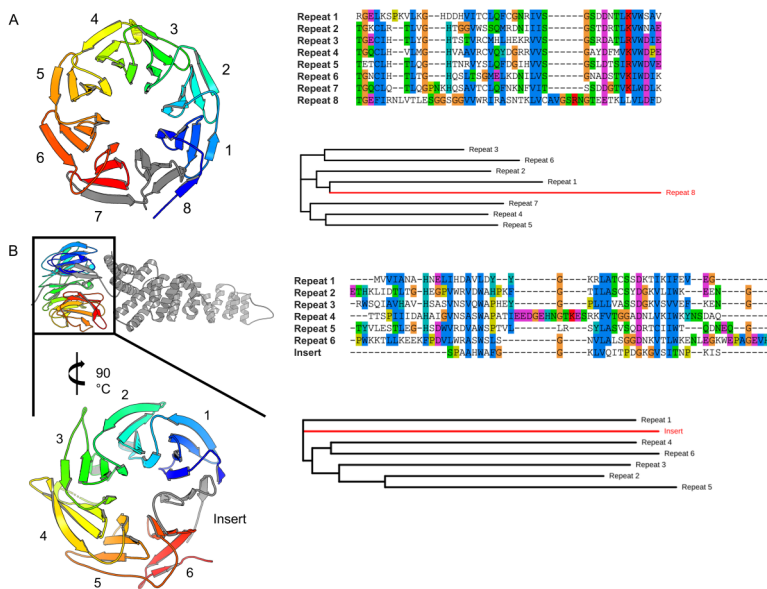


FIGURE 10 Propellers with strange sequence inserts. The structure of the Skp-1-Fbw7 complex (PDB:2OVP) is an eight-bladed propeller. This protein belongs to the WD40 family, but the last repeat does not possess this sequence. A multiple sequence alignment shows that this blade is distinctly different from the other 7 repeats, indicating this blade possibly has a different evolutionary origin (A). The structure of yeast Sec13/31 (PDB: 2PM6) consists of two chains, one comprising a β -propeller with 6 blades while the other chain inserts one blade, creating a full seven-bladed propeller. A sequence alignment and phylogenetic tree shows that the sequence of this inserted blade is not related to the sequences of the others (B)

and Cake9 crystal structures, indicating that the Cake motif has recapitulated the preference of these template blades for a rotation angle between 45° and 40° , and therefore a symmetry between 8- and 9-fold (Tab. 1). BamB appears more similar to Cake8 than Cake9, but the variations between blades are too large to give a conclusive result.

7-bladed β -propellers are the most commonly observed and are theorised to be the most stable [30, 31]. Yet there seems to be no easy route to add or remove a single blade from an ancestral protein, as previously reaffirmed by the Pizza and Tako studies. Additionally, within the β -propeller proteins, there are a number of sequence families that include proteins with different blade number. For example, keap1 and galactose oxidase are respectively 6- and 7-bladed kelch-repeat proteins [32, 33]. One observed way to change rotational symmetry is the insertion of a blade with an unrelated sequence. In the Fbw7-Skp1-Cyclin E Complex (PDB:2OVP), a propeller with seven WD40 repeat motifs is completed by an unrelated sequence that may have evolved to fit into an existing protein, shown in figure 10A. Natural propeller proteins are also known in which a partner protein supplies one blade; this can be observed for example in yeast Sec13/31 (PDB:2PM6), shown in figure 10B. [34].

These examples hint at a widespread role for structural plasticity in the evolution of repeat proteins, but such flexibility is notably absent from known present examples of propeller proteins, whether natural or artificial. Duplication or deletion processes that generate forms with a change in symmetry from the immediate precursor are likely to give proteins that are unstable, and subsequent evolution may well reinforce the new symmetry simply by selecting more stable variants. Protein structures such as PDB:3HXJ may well represent alternative means of stabilising a propeller

with the “wrong” number of repeats, and it is therefore instructive to test experimentally how many the “right” number might have been. At the same time, it might be expected that the derived sequence shows a little more structural flexibility since the template never underwent a prolonged period of evolutionary fixation as a complete propeller. The designed Cake sequence provide more evidence for structural flexibility in β -propellers. Nonetheless, this is a very idealised case in which a single sequence is repeated. In nature the evolutionary pathway would be more complex, with sequences undergoing mutation and selection in between duplication events. A recent study of bacterial outer membrane β -barrels suggested that the number of structural repeats may diversify from that of a parent sequence through more complicated mechanisms than simple duplication [35]. It is unclear whether this would also be the case for β -propellers as a previous bioinformatics study could not identify sequence repeats longer or shorter than a single blade [3].

The self-assembling behaviour of the Cake sequence suggests it may prove to be a useful building block for bionanotechnology, where it can be fused to other units to create large assemblies. Variants of Pizza6 have demonstrated potential as an enzyme and a template for mineralisation [36, 37]. Cake8 and Cake9 have a larger central cavity than Pizza, which may prove suitable as a binding site for ligands or a catalytic site. Engineering such properties into this region of these proteins would still allow them to crystallise readily, creating a stable matrix with chosen symmetry.

4 | CONCLUSION

Starting from an incomplete β -propeller protein, we have designed a novel, symmetrical protein named Cake to study the role of structural plasticity in the evolution of β -propeller proteins. The Cake protein is stable as both an 8- and 9-bladed propeller, showing structural plasticity in monomers for the first time. Proteins consisting of two to ten identical repeats of the Cake motif were found experimentally to form either the 8- or 9-bladed propeller fold. This sequence shows that structural plasticity is possible and might have played a role in the evolution of β -propellers proteins. This structural flexibility might make the Cake protein an interesting building block for bionanotechnology.

5 | MATERIALS AND METHODS

| Computational Protein Design

The selected protein with PDB code 3HXJ has only six complete blades. The intact repeats were isolated and aligned. Repeats with a distorted conformation, or that did not align well, were removed to leave repeats 4, 5, 6 and 7. Rosetta symmetric docking (SymDock) [17] was used to create 7-, 8- and 9-fold circular symmetrical backbones from the models of these four blades. The lowest energy template with correct domain folding was retained for each symmetry, and joined into a single backbone structure with Molecular Operating Environment (MOE) (Chemical Computing Group) (Fig. S1). The sequence alignment of these 4 repeats and the corresponding phylogenetic tree was further used to generate a set of 2000 putative ancestral consensus sequences using FastML [18]. These ancestral sequences were then mapped with suitable symmetry onto the backbones using a pyRosetta script [23], followed by a short energy relaxation with the Fast Relax protocol. For each sequence two energy scoring functions, Talaris2014 and Ref2015, and rmsd deviation to the symmetrical template were calculated. Sequences resulting in a low energy score in both scoring functions and those with a low rmsd deviation from the ideal symmetric backbone were retained. For the 7-bladed models the low-energy conformation had a very high rmsd deviation; only the sequence with the lowest rmsd difference was evaluated further. Two repeat sequences were chosen for each of the 8 and 9-fold symmetrical

designs (Tab. S1). These selected sequences showed sequence similarities with the template protein between 64% and 68%. A schematic overview of the design is given in figure 1.

Protein Expression and Purification

DNA sequences were assembled by reverse-translating the protein sequence and optimising codon usage for *E. coli*. Using NdeI and XhoI restriction sites, synthetic genes were introduced into pET-28b(+) vector, encoding an N-terminal thrombin-cleavable His-tag fused to the desired protein. BL21 (DE3) *E. coli* cells were transformed with the vector, and 1 L expression cultures were grown until an optical density (OD) of 0.6 at 600 nm was reached. IPTG (0.5 mM) was added to induce expression and cells were pelleted. Cells were lysed by suspending and incubating it in 30 mL buffer 50 mM NaH₂PO₄, pH 8, 300 mM NaCl, 10 mM imidazole, 50 ng lysozyme and PMSF (final concentration of 1 mM) on ice, followed by sonication (Branson Sonifier 450, VWR international) and centrifugation. The supernatant containing the soluble protein was filtered (Whatman filter, GE Healthcare) and loaded on nickel nitrilotriacetic acid (Ni-NTA) resin (Qiagen GmbH) equilibrated with 30 mL buffer 50 mM NaH₂PO₄, pH 8, 300 mM NaCl, 10 mM imidazole. The flow-through was collected and the column was rinsed with the same buffer followed by 50 mM NaH₂PO₄, pH 8, 300 mM NaCl, 25 mM imidazole. The protein was eluted with 50 mM NaH₂PO₄, 300 mM NaCl, 300 mM imidazole. The fractions were analysed by SDS-PAGE and those containing the His-tagged protein were collected. The protein was subjected to thrombin while being dialysed overnight against 50 mM NaH₂PO₄, pH 8, 300 mM NaCl using SnakeSkin Dialysis Tubing (Thermo-Fisher Scientific). The protein was reloaded on the Ni-NTA column and the flow-through containing the protein (verified by SDS-PAGE analysis) was collected. The protein solution was concentrated to a final volume of ca. 1.5 mL using Vivaspin 15R (Sartorius AG) and loaded on a HiLoad 16/600 Superdex™ 200 prep grade column (GE Health care) equilibrated with 20 mM HEPES, pH 8, 300 mM NaCl. The fractions corresponding to the main peaks were collected. 1 mg/ml samples of each protein were loaded onto a Superdex™ 200 increase 10/300 (GE Health care) equilibrated with 20 mM HEPES, pH 8, 300 mM NaCl to get analytical data.

Crystallisation

Protein samples were concentrated to ca. 10 mg/mL and dialysed overnight against 20 mM HEPES, pH 8 to remove sodium chloride. Cake3 spontaneously assembled into crystals during this process, however diffraction quality was low. Crystal screening was performed with a Gryphon robot (Art Robinson instruments, Sunnyvale, USA) using sitting-drop 96-well plates (Swissci UVXPO MRC Crystallisation Plate™, Hampton Research) and Screening Suites (Qiagen). Optimisation experiments were set up in 24-well hanging drop crystallisation plates (24-well XRL Plate, Molecular Dimensions) with 250 µL reservoir solution and hanging drops consisting of 1 µL 10 mg/mL protein and 1 µL reservoir solution.

Crystallographic Analysis

Crystals were cryoprotected using the crystallisation liquor with added glycerol (25%, v/v) and flash frozen in liquid nitrogen. X-ray diffraction data were obtained at beamline I-04 of Diamond Light Source (Oxfordshire, United Kingdom) and Proxima 1 at Soleil Synchrotron (Saint-Aubin, France). Diffraction images were processed with XDS [38] and scaling was achieved using AIMLESS [39], part of the CCP4 suite [40]. Molecular replacement using the designed structures as search models was performed with Phaser [41]. Refinement was carried out with phenix.refine [42] and by manual model building with Coot [43]. Cake2, Cake5 and Cake10 were refined with phenix rosetta refine before

getting optimised with the standard phenix protocol [44]. Figures were made with PyMOL [45]. MolProbity was used for validation of the completed structures [46]. All protein structures were deposited in the PDB and assigned the following codes: 6TJB (Cake2), 6TJC (Cake3), 6TJD (Cake4), 6TJE (Cake5), 6TJF (Cake6), 6TJG (Cake8), 6TJH (Cake9), 6TJI (Cake10).

Circular Dichroism

Circular dichroism (CD) spectroscopy was performed on a JASCO J-1500 spectrometer (JASCO). Protein samples (0.05 mg/mL, in 20 mM phosphate, pH 7.6) were analysed using a 2 mm path quartz cuvette at 20° C. The molar ellipticity was measured from 200 nm to 260 nm.

Thermal unfolding experiments were carried out whereby CD spectra of the samples were measured at increasing temperature from 0 to 95° C. First, preliminary measurements were performed in steps of 5° C at wavelengths 200 - 260 nm. This allowed the creation of a heat map which enabled us to choose a wavelength sensitive for secondary structure loss. Subsequently, the unfolding experiment was repeated with the chosen wavelength (218 nm) in steps of 0.2° C. Fitting to calculate melting temperatures was done using equations 1 (for two-state unfolding) and 2 (for three-state unfolding) with $\Delta G = \Delta H_m \left[1 - \frac{T}{T_m} \right]$ [47].

$$Signal = \alpha_N + (\alpha_D + \beta_D T) \left[\frac{\exp \frac{-\Delta G}{RT}}{1 + \exp \frac{-\Delta G}{RT}} \right] \quad (1)$$

$$Signal = \alpha_N + (\alpha_D + \beta_D T) \left[\frac{n_1}{n_1 + n_2} \frac{\exp \frac{-\Delta G_1}{RT}}{1 + \exp \frac{-\Delta G_1}{RT}} + \frac{n_2}{n_1 + n_2} \frac{\exp \frac{-\Delta G_2}{RT}}{1 + \exp \frac{-\Delta G_2}{RT}} \right] \quad (2)$$

Analytical Ultracentrifugation

The sample concentration was estimated as 1.0 mg ml⁻¹ from absorbance at 280 nm. Sedimentation velocity experiments were carried out using an Optima XL-I analytical ultracentrifuge (Beckman-Coulter) using an An-50 Ti rotor. Cells with a standard Epon two-channel center-piece and sapphire windows were used. 400 μL of the sample and 420 μL of the reference solution (50 mM potassium phosphate pH 7.4 and 0.1 M NaCl) were loaded into the cell. The rotor was kept stationary at 293 K in the vacuum chamber for 1 h prior to each run for temperature equilibration. Absorbance at 280 nm scans were collected at 10 min. intervals during sedimentation at 50,000 rpm. The resulting scans were analysed using the continuous distribution *c(s)* analysis module in the program SEDFIT [48]. Frictional ratio (*f*/*f*₀) was allowed to float during fitting. Partial specific volume of the protein, solvent density, and solvent viscosity were calculated from standard tables using the program SEDNTERP [49].

6 | AUTHOR CONTRIBUTIONS

BM designed the proteins. BM, IL and HN purified, crystallised and performed all biophysical experiments. KK performed AUC. JRHT and ARDV conceived and supervised the experiments. All authors contributed to writing the manuscript.

Acknowledgements

We would like to thank Theo Killian for assistance with the designs and Els Deridder for helping with the cloning of the constructs. We thank the beamline scientists at the Diamond Light Source macromolecular beamlines and at Soleil synchrotron for their kind assistance. JRHT thanks OpenEye Scientific Software for financial support. ARDV thanks Research Foundation Flanders for financial support (G0E4717N, G0F9316N and G051917N). BM thanks Research Foundation Flanders for a fellowship (GBM-D3229-ASP/17)

Declaration of interests

The authors declare no competing interests.

references

- [1] Eck RV, Dayhoff MO. Evolution of the structure of ferredoxin based on living relics of primitive amino Acid sequences. *Science* 1966;152(3720):363–366.
- [2] Alva V, Lupas AN. From ancestral peptides to designed proteins. *Current Opinion in Structural Biology* 2018 feb;48:103–109. <https://www.sciencedirect.com/science/article/pii/S0959440X17300945?via=ihub>.
- [3] Chaudhuri I, Södning J, Lupas AN. Evolution of the β -propeller fold. *Proteins: Structure, Function and Genetics* 2008;71(2):795–803.
- [4] Romero Romero ML, Rabin A, Tawfik DS. Functional Proteins from Short Peptides: Dayhoff's Hypothesis Turns 50. *Angewandte Chemie International Edition* 2016;55(52):15966–15971. <https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201609977>.
- [5] Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D. A census of protein repeats¹¹Edited by J. M. Thornton. *Journal of Molecular Biology* 1999;293(1):151 – 160. <http://www.sciencedirect.com/science/article/pii/S0022283699931364>.
- [6] Andrade MA, Perez-Iratxeta C, Ponting CP. Protein repeats: Structures, functions, and evolution. *Journal of Structural Biology* 2001;134(2-3):117–131.
- [7] Vrancken J, Wouters S, Mylemans B, Noguchi H, Tame JRH, Voet ARD. Design of synthetic symmetrical proteins. In: *Synthetic Biology: Volume 2*, vol. 2 The Royal Society of Chemistry; 2018.p. 97–114. <http://dx.doi.org/10.1039/9781782622789-00097>.
- [8] Binz HK, Amstutz P, Kohl A, Stumpp MT, Briand C, Forrer P, et al. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nature Biotechnology* 2004 may;22(5):575–582.
- [9] Lee J, Blaber M. Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. *Proceedings of the National Academy of Sciences* 2011;108(1):126–130. <http://www.pnas.org/cgi/doi/10.1073/pnas.1015032108>.

- [10] Huang PS, Feldmeier K, Parmeggiani F, Velasco DF, Hocker B, Baker D. De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nature Chemical Biology* 2016;12(1):29–34.
- [11] Fülöp V, Jones DT. β Propellers: structural rigidity and functional diversity. *Current Opinion in Structural Biology* 1999;9:715–721.
- [12] Paoli M. Protein folds propelled by diversity. *Progress in Biophysics and Molecular Biology* 2001;76(1-2):103–130.
- [13] Voet ARD, Noguchi H, Addy C, Simoncini D, Terada D, Unzai S, et al. Computational design of a self-assembling symmetrical β -propeller protein. *Proceedings of the National Academy of Sciences* 2014;111(42):15102–15107. <http://www.pnas.org/lookup/doi/10.1073/pnas.1412768111>.
- [14] Noguchi H, Addy C, Simoncini D, Wouters S, Mylemans B, Van Meervelt L, et al. Computational design of symmetrical eight-bladed β -propeller proteins. *IUCr* 2019;6:46–55.
- [15] Yadid I, Tawfik DS. Functional β -propeller lectins by tandem duplications of repetitive units. *Protein Engineering, Design and Selection* 2011;24(1-2):185–195.
- [16] Afanasieva E, Chaudhuri I, Martin J, Hertle E, Ursinus A, Alva V, et al. Structural diversity of oligomeric β -propellers with different numbers of identical blades. *eLife* 2019 oct;8:e49853. <https://doi.org/10.7554/eLife.49853>.
- [17] André I, Bradley P, Wang C, Baker D. Prediction of the structure of symmetrical protein assemblies. *Proceedings of the National Academy of Sciences* 2007;104(45):17656. <http://www.pnas.org/content/104/45/17656.abstract>.
- [18] Ashkenazy H, Penn O, Doron-Faigenboim A, Cohen O, Cannarozzi G, Zomer O, et al. FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Research* 2012 jul;40(W1):W580–W584. <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gks498>.
- [19] Kim KH, Paetzel M. Crystal Structure of Escherichia coli BamB, a Lipoprotein Component of the β -Barrel Assembly Machinery Complex. *Journal of Molecular Biology* 2011 mar;406(5):667–678. <https://www.sciencedirect.com/science/article/pii/S0022283610013276?via=ihub>.
- [20] Holm L. Benchmarking fold detection by DALI Lite v.5. *Bioinformatics* 2019 07; <https://doi.org/10.1093/bioinformatics/btz536>, btz536.
- [21] Keitel T, Diehl A, Knaute T, Stezowski JJ, Höhne W, Görisch H. X-ray structure of the quinoprotein ethanol dehydrogenase from *Pseudomonas aeruginosa*: Basis of substrate specificity. *Journal of Molecular Biology* 2000 apr;297(4):961–974.
- [22] Madeira F, mi Park Y, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* 2019;47(W1):W636–W641.
- [23] Voet ARD, Simoncini D, Tame JRH, Zhang KYJ. Evolution-Inspired Computational Design of Symmetric Proteins. Humana Press, New York, NY; 2017.p. 309–322. http://link.springer.com/10.1007/978-1-4939-6637-0_{_}16.
- [24] Liu Y, Neumann P, Kuhle B, Monecke T, Schell S, Chari A, et al. Translation initiation factor eIF3b contains a nine-bladed β -propeller and interacts with the 40S ribosomal subunit. *Structure* 2014;22(6):923–930.
- [25] Watanabe M, Heddle JG, Kikuchi K, Unzai S, Akashi S, Park SY, et al. The nature of the TRAP-Anti-TRAP complex. *Proc Natl Acad Sci USA* 2009 Feb;106(7):2176–2181.
- [26] Heddle JG, Yokoyama T, Yamashita I, Park SY, Tame JRH. Rounding up: Engineering 12-membered rings from the cyclic 11-mer TRAP. *Structure* 2006 May;14(5):925–933.
- [27] Yadid I, Kirshenbaum N, Sharon M, Dym O, Tawfik DS. Metamorphic proteins mediate evolutionary transitions of structure. *Proceedings of the National Academy of Sciences* 2010;107(16):7287–7292. <https://www.pnas.org/content/107/16/7287>.

- [28] Chen Z, Zhan LH, Hou HF, Gao ZQ, Xu JH, Dong C, et al. Structural basis for the interaction of BamB with the POTRA3–4 domains of BamA. *Acta Crystallographica Section D* 2016 Feb;72(2):236–244. <https://doi.org/10.1107/S2059798315024729>.
- [29] Bliven SE, Lafita A, Rose PW, Capitani G, Prlić A, Bourne PE. Analyzing the symmetrical arrangement of structural repeats in proteins with CE-Symm. *PLOS Computational Biology* 2019 04;15(4):1–18. <https://doi.org/10.1371/journal.pcbi.1006842>.
- [30] Murzin AG. Structural principles for the propeller assembly of β -sheets: The preference for seven-fold symmetry. *Proteins: Structure, Function, and Genetics* 1992 oct;14(2):191–201. <http://doi.wiley.com/10.1002/prot.340140206>.
- [31] Chen CKM, Chan NL, Wang AHJ. The many blades of the β -propeller proteins: conserved but versatile. *Trends in biochemical sciences* 2011;36(10):553–561.
- [32] Ito N, Phillips SEV, Stevenst C, Ogels ZB, Mcphersons MJ, Keen JN, et al. Novel Thioester Bond REvealed by a 1.7A Cyrstal Structure of Galactose Oxidase. *Letters to Nature* 1991;.
- [33] Li X, Zhang D, Hannink M, Beamer LJ. Crystal structure of the Kelch domain of human Keap1. *The Journal of biological chemistry* 2004 dec;279(52):54750–8. <http://www.ncbi.nlm.nih.gov/pubmed/15475350>.
- [34] Fath S, Mancias JD, Bi X, Goldberg J. Structure and organization of coat proteins in the COPII cage. *Cell* 2007;129(7):1325–1336.
- [35] Franklin MW, Nepomnyachyi S, Feehan R, Ben-Tal N, Kolodny R, Slusky JS. Evolutionary pathways of repeat protein topology in bacterial outer membrane proteins. *eLife* 2018 nov;7:e40308. <https://doi.org/10.7554/eLife.40308>.
- [36] Clarke DE, Noguchi H, Gryspeerdt JLA, De Feyter S, Voet AR. Artificial β -propeller protein-based hydrolases. *Chemical Communications* 2019;55(60):8880–8883.
- [37] Voet ARD, Noguchi H, Addy C, Zhang KYJ, Tame JRH. Biom mineralization of a Cadmium Chloride Nanocrystal by a Designed Symmetrical Protein. *Angewandte Chemie - International Edition* 2015;54(34):9857–9860.
- [38] Kabsch W. XDS. *Acta crystallographica Section D, Biological crystallography* 2010 feb;66(Pt 2):125–32. <http://www.ncbi.nlm.nih.gov/pubmed/20124692><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2815665>.
- [39] Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta crystallographica Section D, Biological crystallography* 2013 jul;69(Pt 7):1204–14. <http://www.ncbi.nlm.nih.gov/pubmed/23793146><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3689523>.
- [40] Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 2011 Apr;67(Pt 4):235–242.
- [41] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *Journal of applied crystallography* 2007;40(Pt 4):658–674. <http://www.ncbi.nlm.nih.gov/pubmed/19461840><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2483472>.
- [42] Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica Section D, Biological crystallography* 2010 feb;66(Pt 2):213–21. <http://www.ncbi.nlm.nih.gov/pubmed/20124702><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2815670>.
- [43] Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of *Coot*. *Acta Crystallographica Section D Biological Crystallography* 2010 apr;66(4):486–501. <http://scripts.iucr.org/cgi-bin/paper?S0907444910007493>.
- [44] DiMaio F, Echols N, Headd JJ, Terwilliger TC, Adams PD, Baker D. Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nature methods* 2013;10(11):1102.

- [45] DeLano WL, The PyMOL Molecular Graphics System; 2002. <http://www.pymol.org>.
- [46] Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica Section D, Biological crystallography* 2010 jan;66(Pt 1):12–21. <http://www.ncbi.nlm.nih.gov/pubmed/20057044><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2803126>.
- [47] Santoro MM, Bolen D. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl. alpha.-chymotrypsin using different denaturants. *Biochemistry* 1988;27(21):8063–8068.
- [48] Schuck P, Perugini MA, Gonzales NR, Howlett GJ, Schubert D. Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophys J* 2002 Feb;82(2):1096–1111.
- [49] Laue TM, Shah BD, Ridgeway TM, Pelletier SL. SEDNTERP. In: *Analytical Ultracentrifugation in Biochemistry and Polymer Science* Royal Society of Chemistry; 1992.p. 90–125.