PLOS | ONE

Octarellin VI: Using Rosetta to Design a Putative Artificial $(\beta/\alpha)_8$ Protein

Maximiliano Figueroa[1]◑, Nicolas Oliveira[1]◑, Annabelle Lejeune[1], Kristian W. Kaufmann[2], Brent M. Dorr[2], André Matagne[3], Joseph A. Martial[1], Jens Meiler[2], Cécile Van de Weerdt[1,4]*

[1] GIGA-Research, Molecular Biology and Genetic Engineering Unit, University of Liège, Liège, Belgium, [2] Departments of Chemistry and Pharmacology, Center for Structural Biology, Vanderbilt University, Nashville, Tennessee, United States of America, [3] Laboratoire d'Enzymologie et Repliement des Protéines, Centre for Protein Engineering, University of Liège, Liège, Belgium

Abstract

The computational protein design protocol Rosetta has been applied successfully to a wide variety of protein engineering problems. Here the aim was to test its ability to design *de novo* a protein adopting the TIM-barrel fold, whose formation requires about twice as many residues as in the largest proteins successfully designed *de novo* to date. The designed protein, Octarellin VI, contains 216 residues. Its amino acid composition is similar to that of natural TIM-barrel proteins. When produced and purified, it showed a far-UV circular dichroism spectrum characteristic of folded proteins, with α-helical and β-sheet secondary structure. Its stable tertiary structure was confirmed by both tryptophan fluorescence and circular dichroism in the near UV. It proved heat stable up to 70°C. Dynamic light scattering experiments revealed a unique population of particles averaging 4 nm in diameter, in good agreement with our model. Although these data suggest the successful creation of an artificial α/β protein of more than 200 amino acids, Octarellin VI shows an apparent noncooperative chemical unfolding and low solubility.

Citation: Figueroa M, Oliveira N, Lejeune A, Kaufmann KW, Dorr BM, et al. (2013) Octarellin VI: Using Rosetta to Design a Putative Artificial $(\beta/\alpha)_8$ Protein. PLoS ONE 8(8): e71858. doi:10.1371/journal.pone.0071858

Editor: Alexandre G. de Brevern, UMR-S665, INSERM, Université Paris Diderot, INTS, France

Received July 4, 2012; **Accepted** July 10, 2013; **Published** August 19, 2013

Funding: M.F. is the recipient of postdoctoral fellowships from "Wallonie-Bruxelles International" and the "Belgian Science Policy Office". N.O. is the recipient of a fellowship from the "Fonds pour la Formation à la Recherche dans l'Industrie et l'Agriculture" (FRIA) of the French Community of Belgium. A.M. is supported in part by a grant from the Fonds de la Recherche Fondamentale et Collective (contract number 2.4530.09) and by the Belgian Interuniversity Attraction Poles program initiated by the Belgian Federal Office for Scientific, Technical, and Cultural Affairs (PAI n° P6/19). J.M. and K.K. are supported by the Protein Design Project of the Defense Advanced Research Projects Agency (DARPA grant 04 12). The authors are grateful to the European Space Agency for funding in the context of the Prodex projects C90035 and AO004. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: c.vandeweerdt@ulg.ac.be

◑ These authors contributed equally to this work.

Introduction

The Inverse Protein-folding Problem

The aim of *de novo* protein design, often called the "inverse protein-folding problem", is to find amino acid sequences compatible with a given protein tertiary structure. The primary structure of a protein largely determines its tertiary structure [1,2], and the number of protein sequences compatible with a given fold is limited. Solving the inverse protein-folding problem is therefore a stringent test of our understanding of sequence-structure relationships in proteins. Improving this understanding should help to solve the "protein-folding problem" *per se*: predicting what tertiary structure a given amino acid sequence will adopt. This should ultimately enable us to engineer proteins with custom functions and properties.

Attempts to Model the TIM-barrel Fold

De novo construction of a stable, soluble protein of more than two hundred amino acids is a challenge that remains to be met. Reported successes in designing large artificial proteins involved creating new proteins by assembling, in variable number, multiple copies of a same motif of no more than 40 amino acids long. [3,4]. Attempts to design longer sequences *de novo* have focused on the TIM-barrel fold [5,6,7,8].

The $(\beta/\alpha)_8$ fold, also known as the TIM-barrel fold, is a very widespread protein topology. It is shared by at least 23 superfamilies in the Structural Classification Of Proteins (SCOP) database [9] and is the most common enzyme fold in the Protein Data Bank (PDB) [10]. It is commonly accepted that more than 10% of all enzymes with known structure contain the $(\beta/\alpha)_8$ fold [11,12]. Though more than 76 different sequence families have been listed, they all share a very well defined topology.

Typically, TIM-barrels have between 200 and 250 residues. They can be schematically represented as an eightfold repetition of (βα) units organized in two circular layers of secondary structures. The inner layer consists of eight parallel β-strands, surrounded by an external layer of eight α-helices. The β-strands are paired by a strong hydrogen bond network and form a completely enclosed parallel barrel. The catalytic activity of such proteins is nearly always located on the βα side of the protein, whereas the αβ loops are believed to play a crucial role in stabilizing the structure [13].

Despite the relative ease with which nature creates these $(\beta/\alpha)_8$ barrels, attempts to design artificial TIM-barrels *de novo* have had limited success. Early work, including efforts leading to some of the previous versions of Octarellin, yielded poorly soluble proteins that were hard to characterize and appeared to form molten globule species [5,6,8,14,15,16]. There is one notable exception where computational *de novo* design of an artificial $(\beta/\alpha)_8$ barrel based on an idealized framework yielded a stable protein appearing to adopt a well-defined tertiary structure [7], but the solubility and stability of this protein were low in the long term, making it impossible to characterize its 3D structure by X-ray diffraction.

Today's Powerful Computational Design Methods

The design protocols employed in the above-mentioned studies relied heavily on a combination of chemical intuition and bioinformatic data collected from a limited set of natural sequences. Since then, the number of available crystal structures has increased substantially, and powerful computational methods have emerged, enabling the automated design of sequences folding into a desired topology [17].

The new computational methods use a search function that can rapidly sample the conformational and sequence space and an energy function that can identify minimal energy sequence/conformation pairs [18,19]. The complexity of the conformational search space can be reduced by sampling discrete amino-acid side-chain conformations observed frequently in solved structures [20,21,22]. While the backbone of the protein is usually kept fixed, the side-chain conformations are altered by systematic [23] or random [19] substitutions of rotamers. Recent protocols alternate this side-chain conformational search with an all-atom energy minimization [24,25]. The energy functions used to evaluate the resulting sequences rely on statistical parameters derived from databases of known protein properties [19,20,21,22,26]. These "knowledge-based potentials" increase the accuracy of scoring functions for evaluating the designed sequences.

Using Rosetta to Design an Artificial $(\beta/\alpha)_8$ Barrel

Amongst the programs implementing the new approach, Rosetta has been successfully applied to a wide variety of design problems [27]. Highlight achievements include thermo-stabilizing an enzyme [28], creating a new backbone conformation in a beta turn [29], redesigning the specificity at protein-protein interfaces [30,31], designing novel enzymes based on existing protein scaffolds [32,33,34], and designing an entirely new protein topology [17]. This last result was particularly exciting, as the designed protein, Top7, counts 100 amino acids and is soluble, monomeric, and exceptionally stable. These properties have made it possible to determine a high-resolution crystal structure matching the design model to within 1.2 Å. This success has prompted us to try to push the size limit further. We thus present circular dichroism, dynamic light scattering, and intrinsic fluorescence emission data on Octarellin VI, a 216-amino-acid protein designed with Rosetta to adopt the TIM-barrel fold.

Materials and Methods

Structure Design

To define a protein with a $(\beta/\alpha)_8$ barrel fold, we worked with the RosettaDesign software. The protocol used by RosettaDesign has been explained and detailed previously [17,27,35], and the whole process is summarized in Figure 1. To obtain the desired α/β barrel fold, the objective here was to assemble β-strands (E), α-helices (H), and loops (L) so as to give our backbone an idealized α/β barrel topology: a central sheet of eight parallel strands surrounded by eight helices. A schematic Cα trace consistent with the canonical geometrical features of TIM barrel helix and strand secondary structure was assembled, using as starting point the coordinates of the backbone of our previous design, Octarellin V [7]. For construction of loop regions, six-residue fragments of PDB proteins displaying the secondary structure pattern [E,E/L,L,L,L,L,L/H,H] for βα-loops or [H,H/L,L,L,L,L,L/E,E] for αβ-loops were extracted with the Rosetta loop-building protocol [36]. Those compatible with the geometric coordinates of strands and helices were attached. During the initial design phases, loop positions were set as glycines. A total of 6,000 backbone conformations were constructed with variations in loop conformations.

Each backbone position was classified as being either surface, core, or pore (even when it is known not to be a real pore, we keep this nomenclature for historical reasons) by visual inspection. Surface positions are amino acids belonging to α-helices and loops and are largely exposed to the solvent. Positions projecting the side chain into the space between α-helices and β-strands belong to the core. Pore positions are amino acids belonging to the β-strands and whose side chains project towards the inner barrel. As interactions between side chains of different regions are very limited, the three regions were designed sequentially to reduce the computational demand and thus allow the use of larger rotamer sets (Fig. 2). In the protein core, which was designed first, amino acids were restricted to MAFLIVWYGH. In the pore and surface

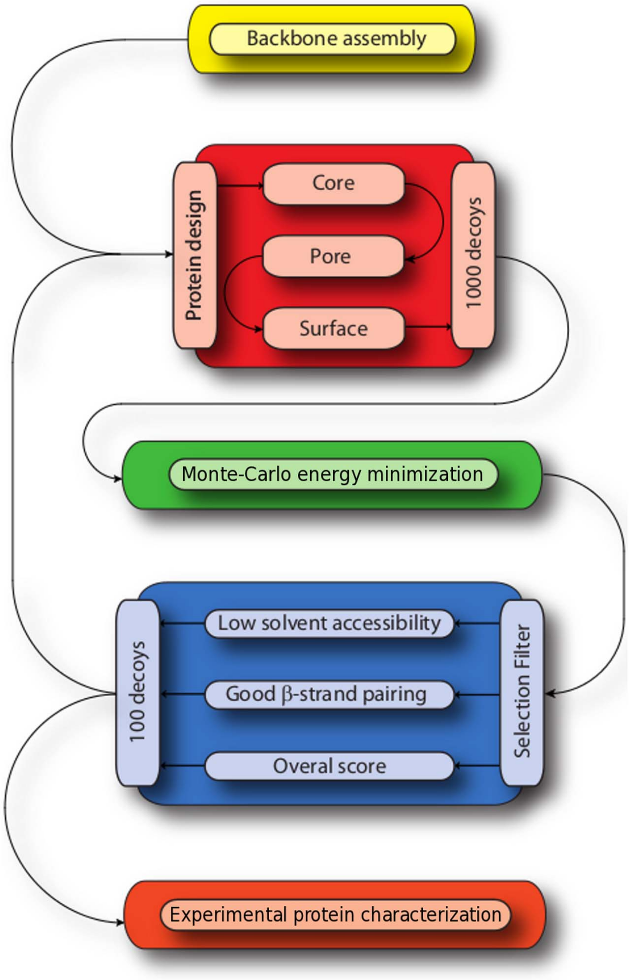

Figure 1. Schematic overview of the design process leading to Octarellin VI. The alternate steps of sequence design, strain removal and filtering procedure are described.
doi:10.1371/journal.pone.0071858.g001

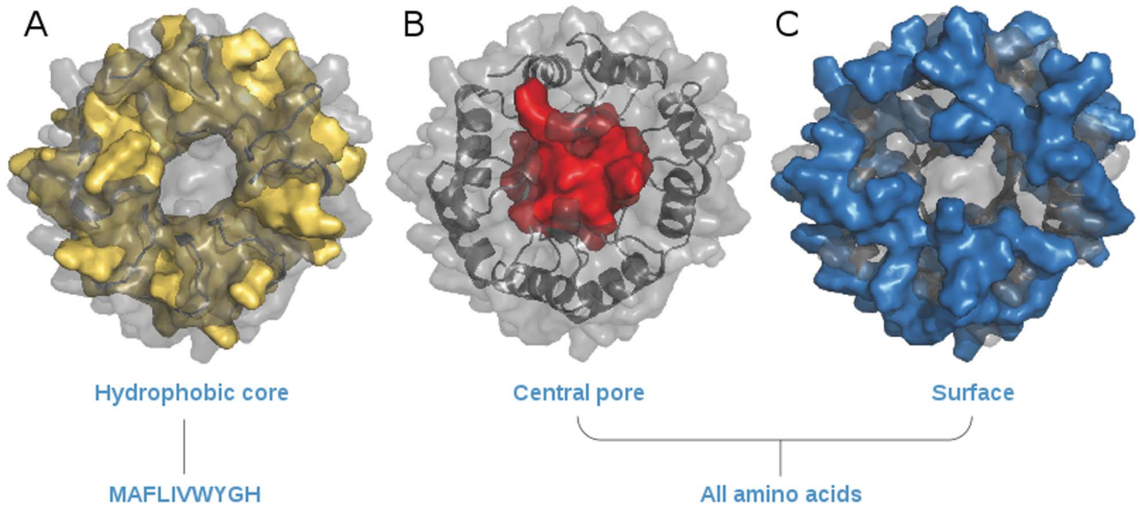

Figure 2. Regions defined by the classification of backbone positions during the hierarchical sequence design procedure. Only amino acids with hydrophobic side chains were used in the core region (A); all residues except cysteine were allowed in the pore (B) and surface (C) regions.

doi:10.1371/journal.pone.0071858.g002

regions, all amino acids except cysteine were allowed. Ten independent design simulations were performed for each of the hundred backbone conformations with the lowest Rosetta energy, generating a total of 1,000 models.

Each of these designed structures was subjected to relaxation with Rosetta's Monte-Carlo energy-minimization protocol. To select models for the next round of design and refinement, a series of filters were applied. The filter criteria applied were that (i) suitable structures should maintain a tight hydrogen-bonding network in the β-barrel of the protein, as evaluated by the Rosetta backbone hydrogen bonding energy; (ii) side chains should be tightly packed so as to exclude solvent from the core, as evaluated by the Rosetta solvent accessible surface area (SASA) measurement; (iii) accepted structures should have minimal Rosetta full-atom energies. After each energy minimization run, the hundred lowest-energy models meeting these criteria were moved forward for further sequence design/backbone optimization. The iterative design process was terminated after five cycles.

Model Selection for Experimental Validation

The best structures from the last round were inspected visually and ranked according to (i) the presence of at least one aromatic residue in the protein core (to facilitate experimental studies) and (ii) the extent to which each protein's amino acid composition, loop geometry, surface hydrophilicity, and predicted secondary structure matched those of natural TIM-barrel proteins.

Further targeted rounds of design were performed to eliminate three hydrophobic patches on the protein surface in the best-ranked design. In these additional design simulations, the residues in the three problematic patches were restricted to ones with small hydrophilic side chains so as to avoid protein aggregation. This phenomenon is not explicitly considered in the Rosetta energy function and was similarly adjusted in previous designs with Rosetta [17]. The final model was called Octarellin VI.

Analysis of the Final 3D Model with External Softwares

To check the accuracy of our 3D model, we performed stereochemical analysis with a Ramachadran plot [37] and energetic analysis with the Anolea [38,39] and ProsaII [40] webservers, using in both cases the default parameters.

Fold Recognition

The sequence of the final model was analyzed with the help of the I-Tasser [41,42], PsiPred [43,44], and 3D-Jury [45] webservers, with the default parameters.

Molecular Dynamics Analysis

To test protein stability and the validity of sequence-structure relationship predictions, a molecular dynamics analysis was performed. Using the software Gromacs [46] and the forcefield OPLS/AA, we first performed an energy minimization by "steepest descent". We then performed a short, 20-ps molecular dynamics simulation for equilibration with the solvent and then 10 full 5-ns molecular dynamics simulations to test the stability of the designed protein and any changes in it. The entire simulation was

Table 1. Rosetta energy values.

Rosetta Energetics	Average value for Octarellin VI	Average value for the control set
Rosetta Energy Units per residue	−2.43	−2.29±0.18
Normalized Solvent Accessible Surface Area	3.80	1.49±1.09
Solvent Accessible Surface Area Probability	0.34	0.46±0.05
Secondary Structure Propensity Energy	−0.62	−0.55±0.1

doi:10.1371/journal.pone.0071858.t001

Table 2. Amino acid composition and properties.

	Value for Octarellin VI	**Average value for the control set**
Amino acid identity	**Percentage (%)**	**Percentage (%)**
Ala	8.8	8.0±2.4
Arg	5.1	4.2±1.5
Asn	7.9	6.1±2.0
Asp	1.9	6.1±1.4
Cys	0.0	1.0±1.0
Gln	7.4	3.4±1.4
Glu	6.0	7.0±3.2
Gly	18.1	7.4±1.7
His	5.1	2.2±1.3
Ile	3.2	6.3±2.1
Leu	8.3	8.4±2.5
Lys	3.7	5.6±2.6
Met	0.5	1.8±1.1
Phe	6.0	3.9±1.0
Pro	0.9	4.3±1.5
Ser	4.2	6.6±1.5
Thr	1.9	4.6±1.4
Trp	5.6	1.7±1.1
Tyr	3.7	3.9±1.5
Val	1.9	7.1±1.8
Side chain nature	**Percentage (%)**	**Percentage (%)**
Aliphatic	41.7	43.5±3.7
Aromatic	20.4	11.8±2.4
Small	31.0	21.9±4.6
Long and flexible	22.7	22.0±4.4
Beta-branched	12.0	20.3±3.3
Charged	16.7	22.8±6.2
Negative	7.9	13.0±3.6
Positive	8.8	9.8±3.4
Polar	26.4	23.9±3.8
Polar charged	43.1	46.7±3.3

doi:10.1371/journal.pone.0071858.t002

done with explicit solvent at 300 K. The values obtained for each trajectory were averaged, the root mean square of the deviation (rmsd) of the backbone being monitored throughout the MD simulation to determine structural convergence. Information about secondary structure, radius of gyration, and the rmsd of the backbone and of each amino acid was extracted from the trajectories.

Table 3. Secondary structure determined by CD.

	Helix (%)	**Strand (%)**	**Turn (%)**	**Unordered (%)**
Octarellin VI	34±3 (45.8)	18±2 (15.3)	19±1 (7.9)	29±2 (31)
***T. maritima* TIM**	36±1 (45.6)	18±1 (15.1)	19±1 (11.1)	27±1 (28.2)

Values in parentheses were obtained from the 3D coordinate files with the DSSP software [47].
doi:10.1371/journal.pone.0071858.t003

Comparison with Natural TIM-barrel Proteins

The final model was also compared with crystallized natural TIM-barrel proteins. Eighteen proteins displaying the $(\beta/\alpha)_8$ fold were selected from the PDB. Each of these structures has a resolution better than 2.2 Å, is known to be a monomer under biological conditions, possesses a chain length of less than 500 residues, and its sequence has less than 70% of identity to that of any other protein in the set. The PDB codes of the eighteen proteins are: 1A53, 1AJ2, 1B54, 1BQC, 1CNV, 1EDG, 1EOK, 1G0C, 1I1W, 1J6O, 1NQ6, 1O1Z, 1PYF, 1UJP, 1VFL, 1WDP, 2CYG, and 7A3H. In addition to energy (Table1) and solvent accessible surface area (SASA) analysis with Rosetta, our synthetic

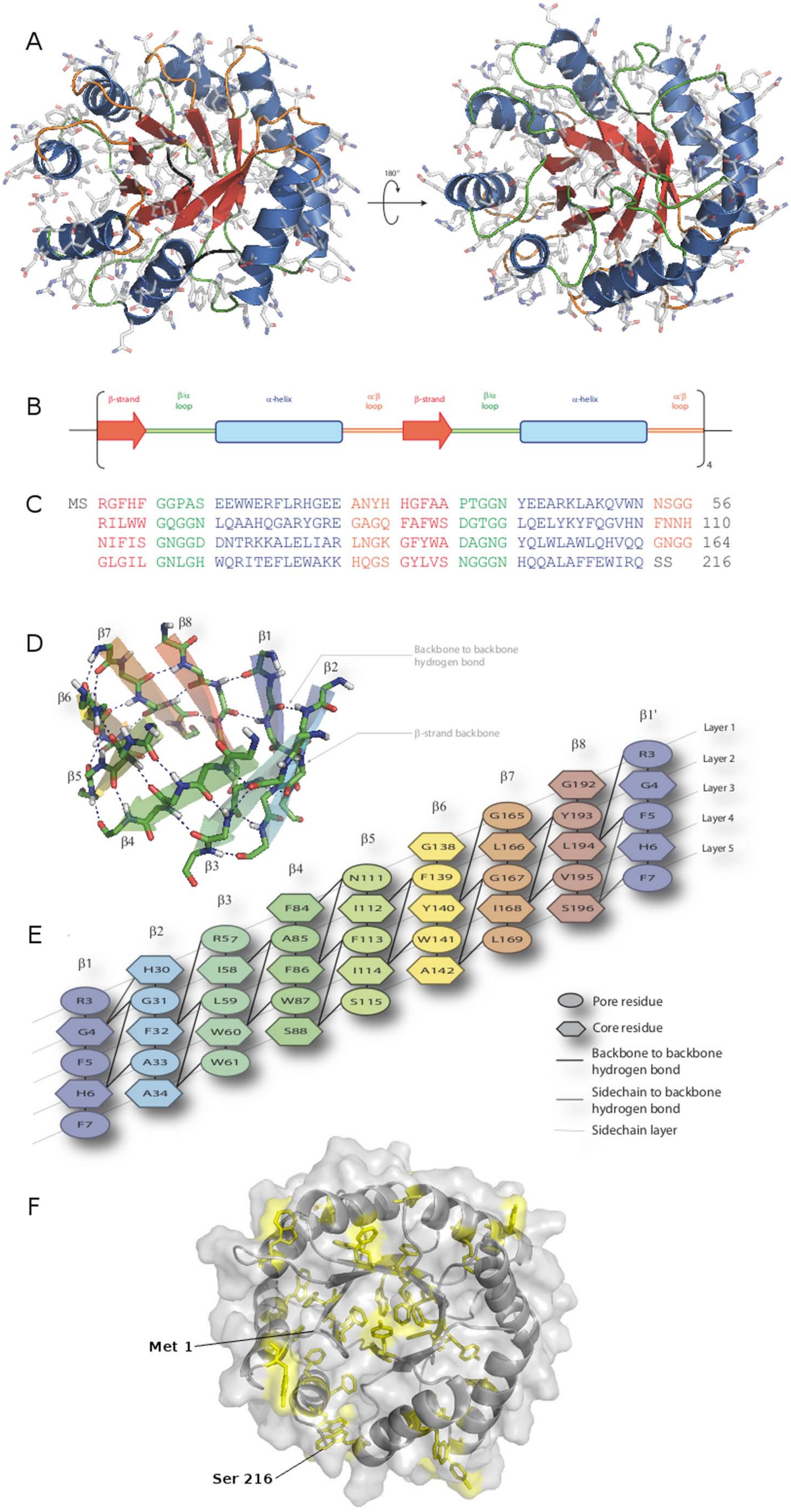
A
180°
B
β-strand
β/α loop
α-helix
α/β loop
β-strand
β/α loop
α-helix
α/β loop
4
C
MS RGFHF GGPAS EEWWERFLRHGEE ANYH HGFAA PTGGN YEEARKLAKQVWN NSGG 56
RILWW GQGGN LQAAHQGARYGRE GAGQ FAFWS DGTGG LQELYKYFQGVHN FNNH 110
NIFIS GNGGD DNTRKKALELIAR LNGK GFYWA DAGNG YQLWLAWLQHVQQ GNGG 164
GLGIL GNLGH WQRITEFLEWAKK HQGS GYLVS NGGGN HQQALAFFEWIRQ SS 216
D
β7
β8
β1
β2
β6
β5
β4
β3
Backbone to backbone hydrogen bond
β-strand backbone
E
β1
β2
β3
β4
β5
β6
β7
β8
β1'
Layer 1
Layer 2
Layer 3
Layer 4
Layer 5
R3
G4
F5
H6
F7
H30
G31
F32
A33
A34
R57
I58
L59
W60
W61
F84
A85
F86
W87
S88
N111
I112
F113
I114
S115
G138
F139
Y140
W141
A142
G165
L166
G167
I168
L169
G192
Y193
L194
V195
S196
R3
G4
F5
H6
F7
Pore residue
Core residue
Backbone to backbone hydrogen bond
Sidechain to backbone hydrogen bond
Sidechain layer
F
Met 1
Ser 216

Figure 3. Schematic representation of the Octarellin VI model. The tertiary (A), secondary (B) and primary (C) structures are shown. (D) Schematic overview of the central β-barrel of the artificial protein model, showing the hydrogen bond network. (E) The nearly ideal pattern expected for a TIM-barrel protein. (F) Aromatic amino acids disposition into the tertiary strucure.
doi:10.1371/journal.pone.0071858.g003

$(\beta/\alpha)_8$ barrel protein was compared with our set of natural TIM-barrel proteins as regards amino acid composition (Table 2) and predicted secondary structure. Agreement in secondary structure prediction (the SS_{score}) was quantified by comparing the DSSP-assigned secondary structure [47] with the probability assigned to that secondary structure type in the three-state prediction by JUFO [48]. The following equation was used to calculate a score:

$$SS_{score} = \frac{\sum -\log\left(\frac{P_{JUFO::DSSP}}{P_{ran}}\right)}{Number \quad of \quad residues}$$

where $P_{JUFO::DSSP}$ is the probability assigned by JUFO to the DSSP-assigned secondary structure and $P_{ran} = 0.33$ is the probability of randomly assigning the correct secondary structure assuming each secondary structure type is equally probable.

Protein Expression and Purification

The gene corresponding to the computationally designed protein Octarellin VI was purchased from BlueHeron Biotechnologies. The gene construct was cloned into the expression plasmid pET-22b (Novagen) and expressed in *E. coli* BL21(DE3) in fusion with a C-terminal hexahistidine tag. Cells transformed with pET22b-Octarellin VI were grown at 37°C in LB containing 100 µg/ml ampicillin. When the culture reached $OD_{600} = 0.6$, production was induced by addition of isopropyl β-D-1-thiogalactopyranoside at 1 mM final concentration. After 4 h, the cells were harvested by centrifugation. Very good Octarellin VI expression was achieved in *E. coli*, but the protein was found in the insoluble fraction of the bacteria. Inclusion bodies were isolated by resuspending the bacterial pellet in 25 mM Tris-HCl pH 8.5, 500 mM NaCl and rupturing the cells by sonication. After centrifugation of the homogenate, the inclusion-body-containing pellet was washed, first with 25 mM Tris-HCl pH 8.5, 500 mM NaCl and 1% Triton X-100, then three times with the same buffer without Triton. Washed inclusion bodies were solubilized in 25 mM Tris-HCl (pH 8.5), 6 M guanidine chloride. All subsequent purification procedures were performed in this buffer. Denatured protein solution was loaded onto an immobilized metal affinity chromatography (IMAC) matrix charged with the Ni^{2+} ion (IMAC Sepharose HP, XK 16/20 column, GE Healthcare). The protein was eluted with an imidazole gradient (0–500 mM). Fractions containing Octarellin VI were pooled and concentrated before size exclusion chromatography (SEC) on an XK 16/70 Sephacryl S-100 column (GE Healthcare).

Refolding

Refolding conditions were determined by following the screening procedure described by Vincentelli and co-workers [49]. The best conditions for Octarellin VI refolding were 1:20 (v/v) dilution in a vigorously stirred solution containing 25 mM Tris-HCl, 500 mM L-arginine, and 100 mM 3-(1-pyridinio)-1-propanesulfonate (NDSB-201) (pH 8.5) followed by incubation at 4°C overnight. Precipitated protein was removed by centrifugation and the refolding solution was concentrated to 1 mg/ml. The concentrated protein solution was dialyzed twice against 10 mM Tris-HCl (pH 8.5). Precipitated protein was again removed by centrifugation and the supernatant filtered with a 0.22-µm filter.

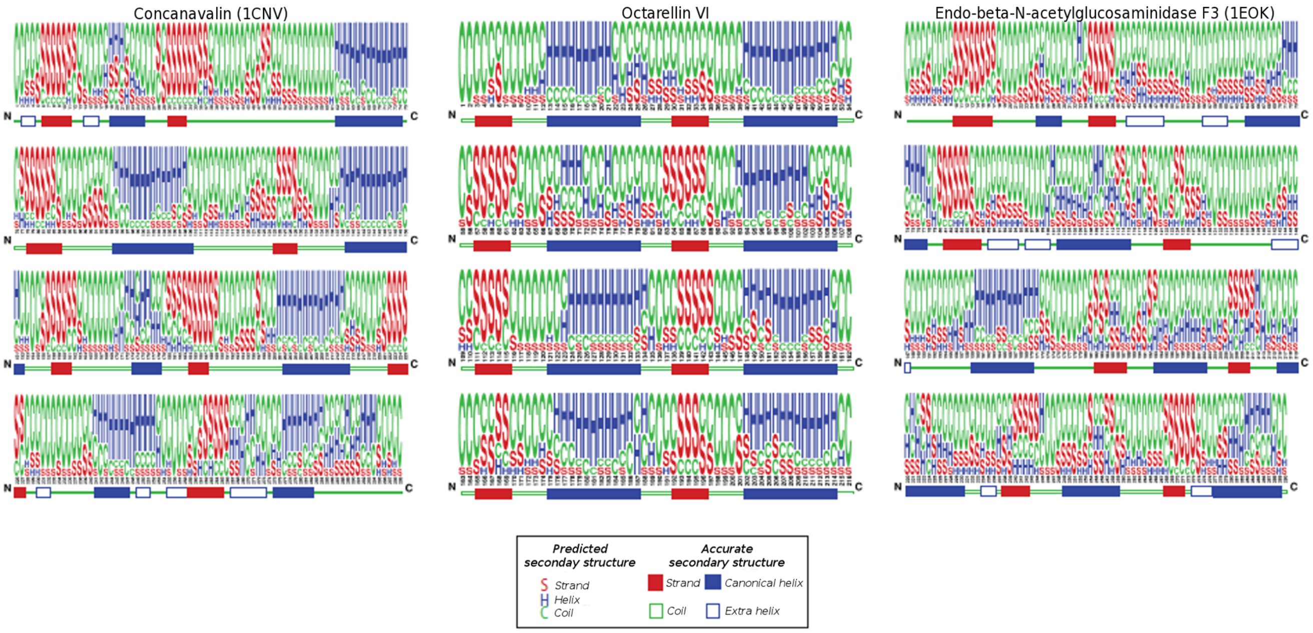

Figure 4. Testing the validity of predicted overall secondary structures. On the left and right, the validity of secondary structures predicted by JUFO is demonstrated for two proteins extracted from our control set (PDB ID 1 cnv and 1 eok). The upper line represents the predicted propensity for each secondary structure type at each position (coil, helix, or strand), while the lower line represents the actual secondary structure elements (based on the 3D crystal structure) listed in the corresponding PDB file. In the middle, the JUFO prediction for Octarellin VI is shown (upper line), as compared to the structure elements based on the model coordinates.
doi:10.1371/journal.pone.0071858.g004

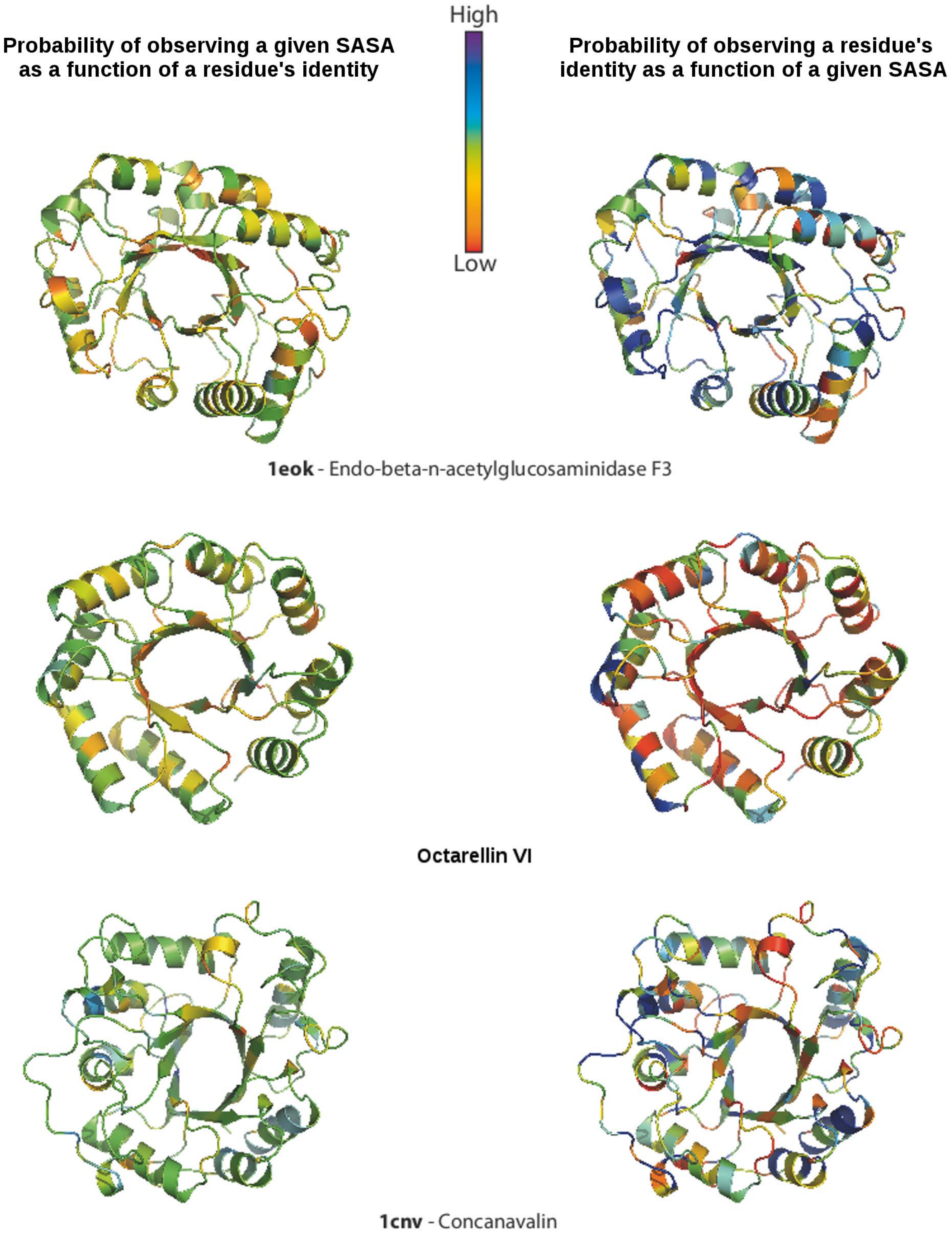
High
Low
Probability of observing a given SASA as a function of a residue's identity
Probability of observing a residue's identity as a function of a given SASA
1eok - Endo-beta-n-acetylglucosaminidase F3
Octarellin VI
1cnv - Concanavalin

Figure 5. Residue packing in Octarellin VI as assessed by SASA analysis. SASA probabilities are shown for each residue of Octarellin VI and of two natural $(\beta/\alpha)_8$ barrels.
doi:10.1371/journal.pone.0071858.g005

Alternatively, when a refolding additive compatible with CD measurements was required, NV10 (Expedeon) was used. In this case, the protein unfolded at 2.6 mg/ml in 6 M urea, 10 mM phosphate buffer, pH 8.0 was refolded by 10-fold dilution in 10 mM phosphate buffer, pH 8.0 containing 1 mg/ml NV10. The refolded protein was extensively dialyzed against 10 mM phosphate buffer, pH 8.0 (to remove urea) and filtered through a 0.45-μm filter. All protein concentrations were determined by

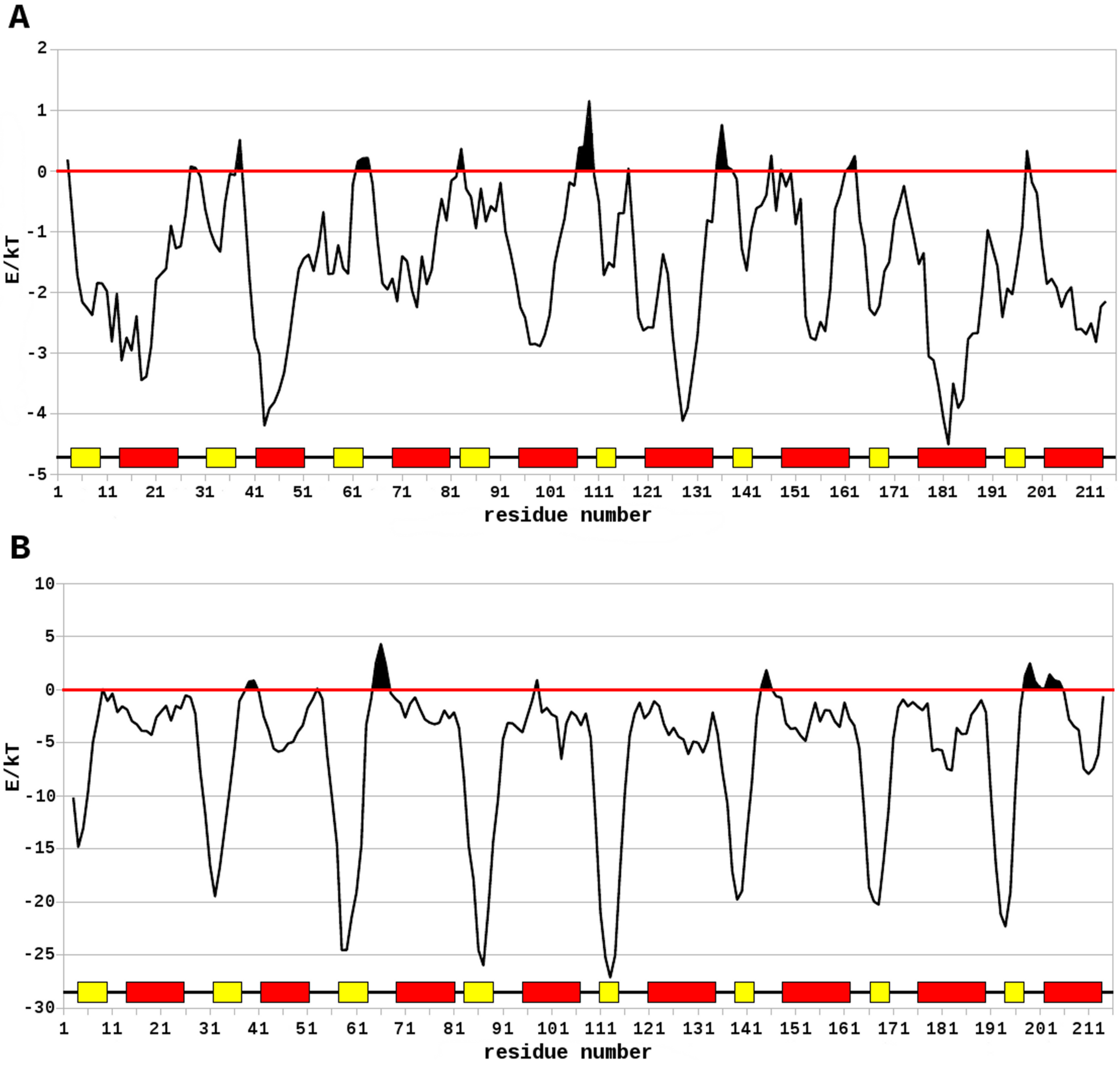

Figure 6. Local energy analysis by Anolea and Prosall. (A) Prosall analysis reveals that most of the protein has low energy, with the exception of some loop regions. (B) Anolea also shows high energy only in the loop regions.
doi:10.1371/journal.pone.0071858.g006

measuring absorbance at 280 nm and using the theoretical values of both the extinction coefficient (78,520 $M^{-1}cm^{-1}$) and molecular mass (MM = 25,504 Da) calculated with the ExPASy Protparam tool (http://www.expasy.org/tools/protparam.html).

Dynamic Light Scattering (DLS)

DLS measurements were performed with a Malvern Zetasizer NanoS instrument fitted with a 633-nm laser and a Peltier cell-holder. Data were recorded with a non-invasive backscatter detection angle of 173° at 25°C. A 45-µl "small-volume" 3-mm-path quartz cell containing the protein at 5 µM in 10 mM Tris-HCl (pH 8.5) or 25 mM Tris-HCl, 2 M L-Arginine (pH 8.5) was used. Eleven 10-s runs were performed and averaged. The resulting measurements were collected, analyzed, and correlated with the help of DTS software (Version 5.03) provided by the manufacturer. Solvent viscosity was measured with an AND SV-10 vibro viscometer. Heat-induced protein denaturation was observed under the same conditions. The temperature was increased from 25°C to 95°C by increments of 1°C. Samples were allowed to equilibrate for two minutes before data acquisition.

Fluorescence Measurements

Fluorescence emission spectra were recorded at 25°C with a Perkin–Elmer LS-50B spectrofluorimeter. The protein concentration was 3 µM in 10 mM Tris-HCl (pH 8.5) and the urea concentration was varied from 0 to 8 M. A stirred cell with a 1-cm pathlength was used. Emission spectra were recorded five times from 300 to 440 nm (excitation at 280 nm) and averaged. The

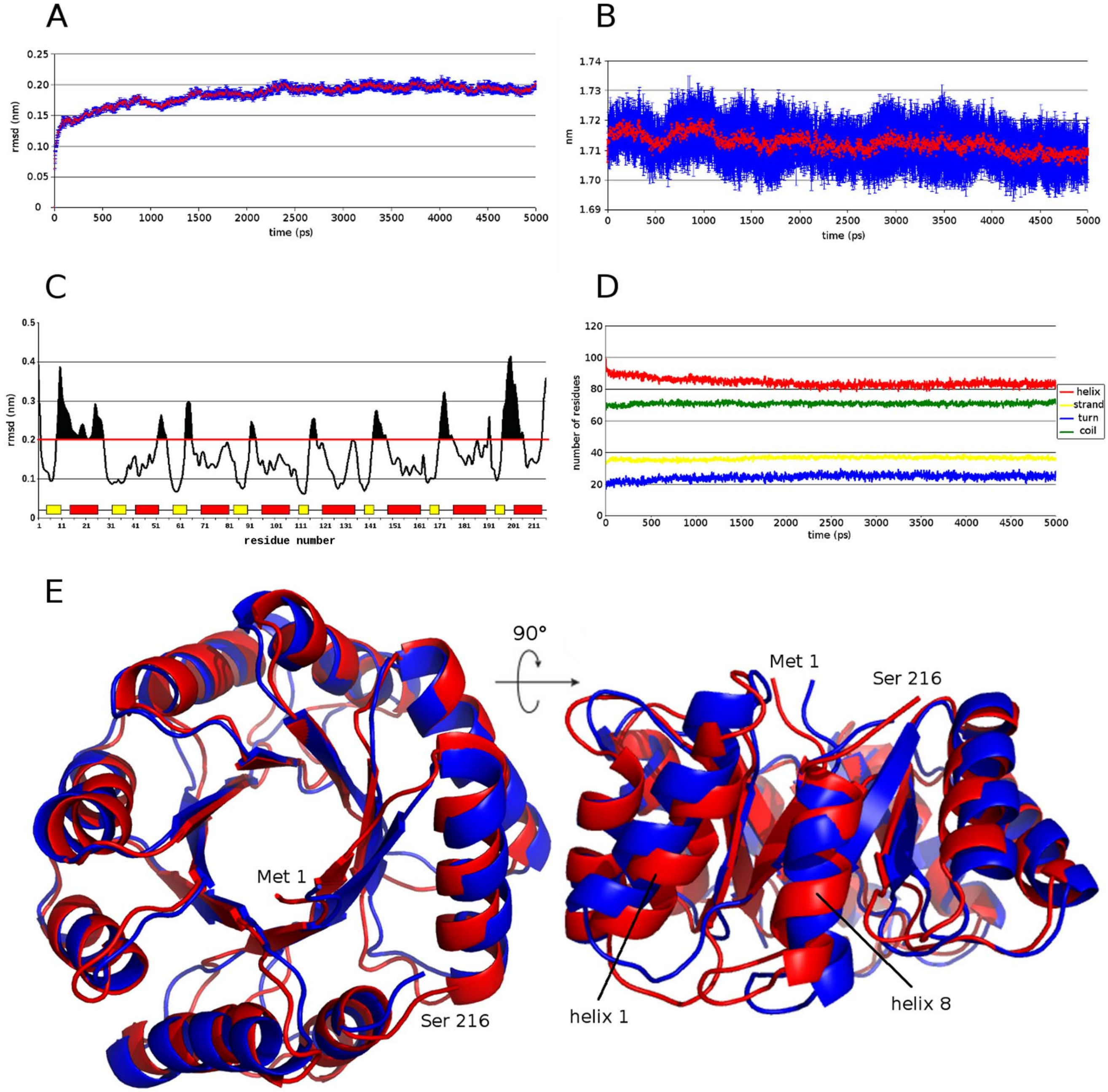

Figure 7. Molecular dynamics simulation analysis of Octarellin VI. (A) Root mean square deviation (rmsd) of the backbone through a 5-ns simulation. (B) Variation of the radius of gyration of Octarellin VI in the MD simulation. (C) Rmsd of each amino acid in the MD simulation. Yellow boxes represent strands and red boxes, helices. A threshold of 2 Å was defined, all rmsd values above this threshold being considered as real movements. The above-threshold area under the curve is represented in black in the latter situation. (D) Evolution of secondary structure contents in the course of the MD simulation. With DSSP software, the secondary structure content was calculated at each frame of the simulation and plotted as number of residues present for each secondary structure elements through the time of simulation. (E) Superposition between the original model of Octarellin VI (red) and the final model after MD simulation (blue). All secondary structure elements are maintained, with movements in loop zones and a rearrangement in helix 1 and 8 without loss of its secondary structure.
doi:10.1371/journal.pone.0071858.g007

excitation and emission slit-widths were 3.2 nm and the scan rate was 100 nm min^{-1}.

Circular Dichroism Measurements

Circular dichroism (CD) measurements were performed at 20°C with a Jasco J-810 spectropolarimeter equipped with a six-cell Peltier holder, in either the far-UV (190–250 nm) or the near-UV (250–310 nm) region, using a protein concentration of 3 or 39 μM and a cell pathlength of 0.1 or 1 cm, respectively. Spectra were acquired at a scan speed of 10 nm min^{-1}, with a 1-nm bandwidth and a 4-s response time. The spectra were measured four times in the presence of NV10 (Expedeon), a synthetic polymer preventing protein aggregation, stabilizing the proteins in solution, and allowing spectroscopic characterization by circular

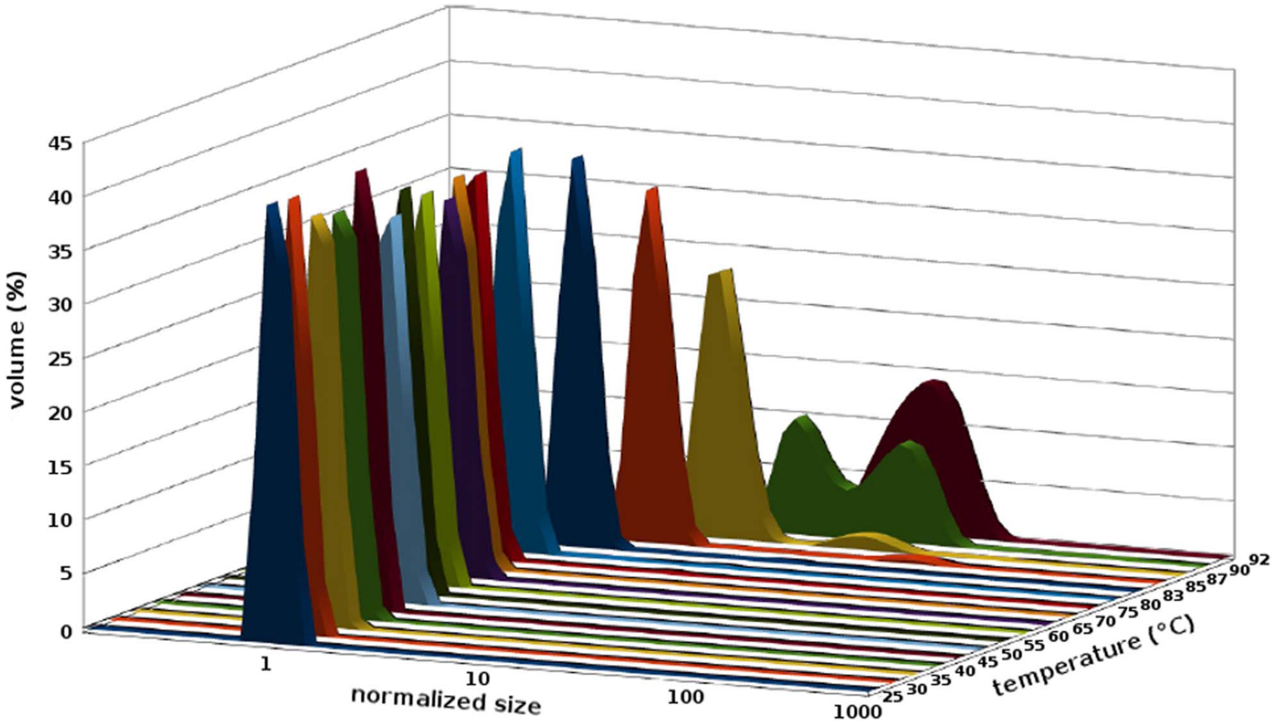

Figure 8. Effect of temperature on the particle size distribution. Dynamic light scattering was used to estimate the hydrodynamic diameter of Octarellin VI (at 5 μM concentration) at different temperatures. The average hydrodynamic diameter at 25 °C is 4.82 nm. Values are normalized and plotted on a logarithmic scale. Each temperature is represented by a different color.
doi:10.1371/journal.pone.0071858.g008

dichroism. The spectra were averaged and corrected by subtraction of the buffer spectrum obtained under identical conditions. Calculation of secondary structures from analysis of the CD data was done with the CONTINLL [50,51], CDSSTR [52,53], and SELCON3 [54,55] algorithms provided by the DichroWeb analysis server [56,57]. Two protein reference databases (4 and 7) were used and the results obtained with the individual algorithms were averaged; the standard deviations between the calculated secondary structures are reported in Table 3. For thermal and chemical unfolding measurements at a fixed wavelength (222 nm), the compound NV10 was not added.

Urea- and Heat-induced Unfolding

For urea-induced unfolding, protein samples were incubated overnight at 25°C in the presence of various concentrations of urea ranging from 0 to 8 M in 10 mM Tris-HCl buffer (pH 8.5). The protein concentration was 3 μM. The denaturant concentration was determined from refractive index measurements [58] performed with a R5000 hand refractometer from Atago. For heat-induced unfolding, the same buffer and protein concentration were used. The protein sample was heated by increasing the temperature monotonically from 25°C to 92°C at the rate of 0.5°C/minute. In chemical and heat unfolding experiments, transition curves were obtained by monitoring, respectively, the shift of the maximum fluorescence emission wavelength (λ_{max}) and the change in CD signal intensity at 222 nm.

Results

Designing an Idealized Artificial TIM-barrel Protein

An idealized $(\beta/\alpha)_8$ backbone was assembled. Sequence design was alternated with energy minimization steps in an iterative process. Models taken from one cycle to the next were selected by application of a filter (see Methods). Finally, after five iterations, targeted rounds of design were performed to eliminate hydrophobic patches and discourage aggregation. In all, more than 5000 different sequences were tested in the whole design process. The final selected model was named Octarellin VI, because it is the sixth Octarellin created in our laboratory. Figure 3 represents the final 3D model, showing a diagram of the different structural elements present in it.

The Designed Protein Structure shows Native-like in Silico Characteristics

The average Rosetta energy per residue of the designed protein (the result of Rosetta's energy function), −2.45 Rosetta energy units per residue, falls within the range of per residue energies observed for a set of 18 crystal structures of TIM barrels (−2.29±0.18 Rosetta energy units per residue, see Table 1). A secondary structure prediction by JUFO [48] identified 7 α-helices and 5 β-strands in the protein, the remaining α-helix and three β-strands being identified at a reduced confidence level (Fig. 4). The overall secondary structure prediction accuracy was comparable to that of predictions performed on a set of 18 natural TIM-barrel crystal structures (−0.62 vs. −0.55±0.12). We further performed a fold recognition analysis of the Octarellin VI sequence, checking the ability of our designed sequence to fold into a TIM-barrel, even though a Blast analysis revealed no similarity between Octarellin VI and any known protein (data not shown). The webservers I-Tasser, PsiPred, and 3D-Jury were used for this analysis. As best template, these webservers identified respectively bacterial luciferase (PDB code 1LUC), dihydrodipicolinate synthase (PDB code 2PUR), and 3D-Jury identified 2-keto-3-

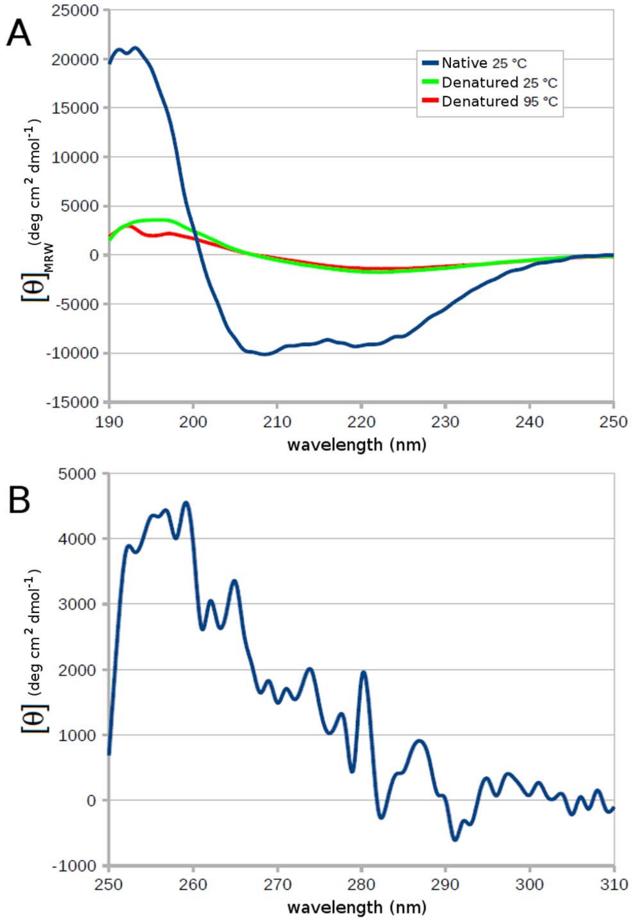

Figure 9. Circular dichroism spectra of the artificial protein. (A) Far-UV spectrum of Octarellin VI recorded at 25°C, after denaturation at 95°C, and after cooling from 95 to 25°C. (B) Near-UV spectrum of Octarellin VI at 25°C.
doi:10.1371/journal.pone.0071858.g009

deoxygluconate aldolase (PDB code 1W37). All three of these proteins have a TIM-barrel fold.

The quality of residue packing was assessed by SASA analysis. On the basis of the overall SASA scores, Octarellin VI appears less tightly packed than the 18 crystal structures (3.80 vs. an average of 1.49±1.09). The comparison appears more favorable, however, when one looks at the overall probability of observing the predicted exposure for a specific amino acid (0.34 vs. 0.46±0.05). Figure 5 shows, residue by residue, the probability of observing the predicted SASA for the amino acid present at each position and the probability of observing the expected residue given the SASA value determined at that position. From these figures, one can see that the solvent accessibility of the designed structure falls within acceptable limits.

In terms of amino acid categories, the amino acid composition of the synthetic TIM is comparable to that of natural $(\beta/\alpha)_8$ barrel proteins (see Table 2), but two categories stand out: first, the percentage of small amino acids is higher than expected (31.0% vs. 21.9% ±4.6%); this is likely mainly due to the fact that the glycine content of the designed protein is higher than the average content observed in our control set (18.1% vs. 7.9% ±1.7%). Second, the aromatic content of the designed protein is higher than expected (20.4% vs. 11.8% ±2.4%) because of our filter forcing the inclusion of aromatics in the designed sequences and our decision to include only nonpolar amino acids in the core.

The designed protein shows good stereochemical features. The Ramachandran plot revealed only 4 residues (1.9%) in a non-allowed region (data not shown): residues Arg 3, Ala 11, Ala 81, and Ala 144, all four present in loop regions. Local energy analyses with the Anolea (Fig. 6B) and ProsaII (Fig. 6A) webservers revealed similar high percentages of residues in the structure having a favorable low energy (92% observed with Anolea). Interestingly, helices showed the lowest local energy levels in the ProsaII analysis, as opposed to strands in the Anolea analysis. In both cases, however, the loop regions showed the highest local energy levels.

Molecular Dynamics Simulations show the Structural Stability of the Designed Protein

Despite its differences in amino acid content as compared to the control group, the Octarellin VI model showed good structural stability in MD simulations (Fig. 7). Ten different MD simulations were performed and the trajectories analyzed. The rmsd of the backbone reached a plateau at 3 ns, indicating no further change in the global structure, and an equilibrated structure (Fig. 7A). The radius of gyration remained constant throughout the simulations, in keeping with the stability suggested by the rmsd of the backbone. The secondary structure content also is proved to be stable: the helix content first decreased slightly, but remained stable after 2.5 ns of simulation. To test local displacements in the structure, a threshold of 2 Å was defined for the rmsd of each residue. According to this criterion, most of the movements in the protein were observed in the loop regions connecting strands with helices (Fig. 7C). Helix one and part of helix eight also showed displacements, but without any loss of structure. All these results suggest that our artificial protein is at a minimum global energy.

Dynamic Light Scattering Indicates a Unique Population with a Hydrodynamic Diameter Close to that Expected for the Designed Protein

To validate our model experimentally, the gene encoding Octarellin VI was expressed in *E. coli* BL21(DE3) as described under "Materials and Methods". As the protein turned out to be completely insoluble in the bacteria, it was necessary to purify it from inclusion bodies and then to refold it. All measurements in this work were done on the refolded protein. We first performed a DLS analysis to measure fluctuations in particle size (hydrodynamic diameter) as a function of temperature in an interval ranging from 25°C to 92°C (Fig. 8)… At temperatures below 73°C, the average hydrodynamic diameter of the particles was found to be fairly constant (4.82±0.21 nm). The molecular weight of the protein, as estimated from these measurements, was 25.3 kDa. This is in excellent agreement with the theoretical molecular weight of 25.5 kDa and further indicates that Octarellin VI is a monomeric protein. Above 74°C, the particle size was found to increase significantly, from approximately 6 nm to more than 200 nm, and the size distribution profile of the protein population was found to shift from a very narrow single peak to several broader peaks (Fig. 8). These results suggest that heating above 74°C causes the protein to aggregate.

Circular Dichroism Reveals a Folded Protein

To check whether the refolded Octarellin VI adopts the predicted secondary structure, we measured CD spectra in the far UV. Two minima were observed close to 222 nm and

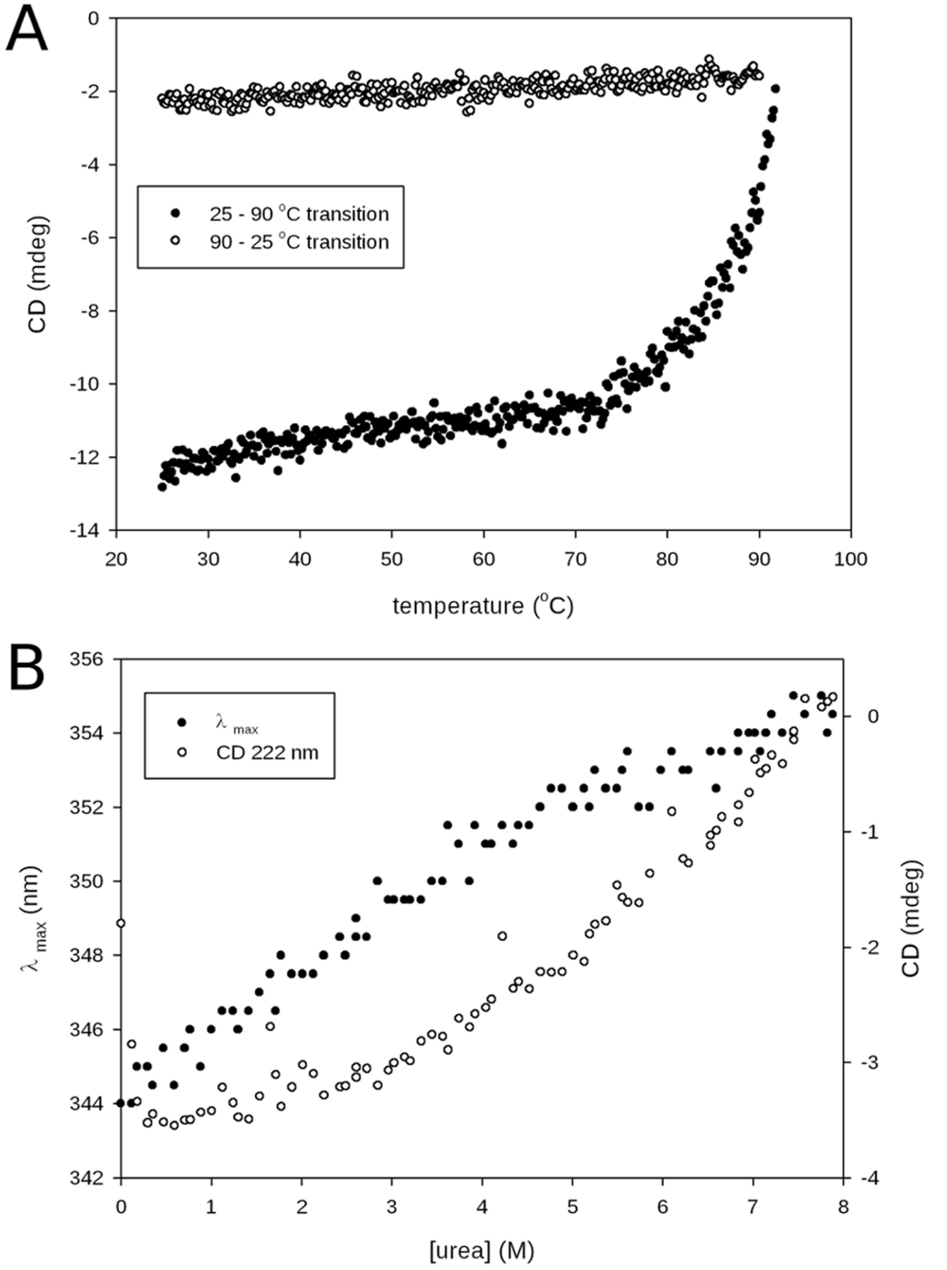

Figure 10. Urea- and heat-induced unfolding of the artificial protein. (A) Heat-induced unfolding monitored by measuring the CD signal at 222 nm; after unfolding, the CD signal at 222 nm was monitored during cooling from 95 to 25°C. (B) Change in the λ_{max} of fluorescence emission (after excitation at 280 nm) and change in the CD signal at 222 nm as a function of the urea concentration. The change in λ_{max} of fluorescene emission shows a shift from 344 nm (folded) to 355 nm (unfolded) as the urea concentration increases, and the same behavior is observed for the CD signal at 222 nm.
doi:10.1371/journal.pone.0071858.g010

208 nm, and the overall spectrum looked typical of that expected of an α/β protein (Fig. 9A). A secondary structure analysis was performed with DichroWeb to estimate the percentage of each type of secondary structure. The spectrum of the protein refolded in the presence of NV10 gave good quality data down to 190 nm and hence, the secondary structure content of the protein was calculated and is given in Table 3. Data are in good agreement with those obtained for *T. Maritima* TIM (PDB code 1B9B), a natural thermostable α/β-barrel protein with 250 amino acids. Furthermore the analysis performed using the Dichroweb server indicated an average content of 3.8 helices per 100 residues, yielding a total value of about 8.2 helix segments in the protein, which is in complete agreement with our 8-helix design.

CD spectra were also obtained in the near-UV region. Absorption bands were observed, indicating that a number of aromatic side chains are held in a rigid environment. This suggests the presence of a tertiary structure (Fig. 9B).

Thermal and Chemical Denaturations Monitored by Circular Dichroism and Tryptophan Fluorescence Reveal an Unfolding Transition

To test the stability of the protein and observe its unfolding, thermal and chemical denaturations were performed.

Heat denaturation was monitored by CD in the far-UV region (at 222 nm). The protein appeared stable up to 70°C (Fig. 10A, shift from 25 to 92°C), but above this temperature, heat-induced unfolding occurred, and this process was irreversible (Fig. 10A, shift from 92°C to 25°C). This result is in agreement with the DLS data (Fig. 8) showing that Octarellin VI remains stable and maintains its secondary structure content even at 70°C. Together, the DLS and CD data suggest that when the protein starts to unfold, stable aggregates appear (Fig. 8 and Fig. 10A).

Chemical denaturation of Octarellin VI (with urea) was monitored by recording tryptophan fluorescence and the CD signal at 222 nm. Increasing the urea concentration caused the wavelength of the emission maximum to shift from 344 nm to 355 nm. This is typical of the transition from a folded protein, where the tryptophans are protected in the core, to an unfolded protein, where the tryptophans are fully exposed to the solvent. The CD signal at 222 nm, which revealed the stability of the protein's secondary structure, also showed a continuous decrease in the signal with the same profile as for the fluorescence assay.

Both techniques (Fig. 10A and 10B) showed a monotonous signal change upon unfolding, instead of a typical sigmoid profile. This suggests a noncooperative transition.

Discussion

Pushing the Size Limit

With the 100-residue protein Top7 [17], Rosetta is the only protein design protocol demonstrated to have yielded *de novo*, without the help of a scaffold protein, a model close to reality. In the wake of this and other successes, we have used Rosetta to design a protein twice as long, intended to adopt the $(\beta/\alpha)_8$ fold. We have thus designed, produced in *E. coli*, and purified the 216-residue protein Octarellin VI.

Our computational analyses of Octarellin VI suggest favorable overall structural energetics and highlight a resemblance to natural $(\beta/\alpha)_8$ barrel proteins as regards amino acid composition (apart from an overabundance of glycine and aromatic residues), predicted energetics, and predicted secondary structure features. Our experimental data are also encouraging: purified Octarellin VI shows a stable tertiary structure with the expected α-helix and β-sheet contents (as suggested by our CD and tryptophan fluorescence data) and high resistance to heat-induced unfolding.

In comparison with our previous work [7], Octarellin VI does not appear to show a big improvement, because it displays the same negative feature, the insolubility. However, the protocol implemented in Rosetta considers all the amino acids, while the proline residues were not allowed in the Octarellin V design. This new protein shows a better thermo stability, with an apparent Tm of 85°C vs 65°C for Octarellin V. Also, *in silico* simulation to test protein stability (Figure 7) shows a correct relationship between primary and tertiary structure in the Octarellin VI model. The same simulation for Octarellin V model shows more movements and changes in the global position of its atoms, leading at the end of the simulation to a structure where the rmsd with the original model is more than 5 Å (data not shown) while maintaining a $(\beta/\alpha)_8$ structure. This data indicates that the new protocol implemented into Rosetta enables to create a protein model where the primary structure has a better relationship with the tertiary structure.

Yet the protein is not soluble enough to allow determining its 3D structure by X-ray analysis, and shows apparently noncooperative unfolding.

Solubility

Historically, attempts to design artificial TIM barrels *de novo* have often produced proteins with low solubility. In the present case, we think this problem is at least partially linked to the design methodology, which seems to produce excessively hydrophobic patches on the protein surface. The observed excess of glycine and aromatic residues might contribute to the problem [59,60] by causing hydrophobic patches to appear, decreasing the proportion of polar residues at the protein surface, and favoring stabilization of intermediates liable to aggregate during the folding process. While the restriction to only polar amino acids to the surface could be a solution to avoid the appearance of hydrophobic patches, this approximation is far from the reality of a natural protein, where some hydrophobic amino acids in the surface are required to stabilize its structure [61].

Moreover, because glycine lacks a side chain, glycine residues increase the conformational space (or perhaps the dynamics, flexibility) of the unfolded polypeptide chain, rendering the unfolded state entropically favorable. This results in stabilization of the unfolded state and hence in a global reduction of the free energy of unfolding [62,63].

The high glycine content is an artefact of the design process. Initially, the loop residues were set as glycines, to be 'mutated' by Rosetta in successive design rounds. For this, Rosetta can search a database of 6-residue loops contained in the PDB. Despite this feature, the initial glycines were not readily removed.

There are several instances where Rosetta users have had to make manual adjustments. The designers of Top7, for instance, had to restrict the protein's twenty-two surface β-sheet positions to polar amino acids [17], and in the recent *de novo* design of a molecular switch, Ambroggio and Kuhlman found it necessary to constrain exploration of the sequence space by using an energy function derived from multisequence alignments of well-conserved members of their design target superfamily [18,64]. We believe that these manual modifications have been necessary because the energy terms in the Rosetta potential only provide an accurate description of solvation effects, without explicitly discouraging aggregation. Yet protein aggregation is a phenomenon that goes beyond solvation, as it includes not just the energetics of the interaction with the solvent but also nonspecific interactions of the protein with itself. A newer version of the Rosetta potential might help to overcome this limitation [65]. Experimentally, furthermore, the choice of buffer can greatly influence the solubility of a designed protein. Understanding such effects might help to improve the design process.

Folding/Unfolding

The relative roughness of the folding free energy landscapes of several $(\beta/\alpha)_8$ barrel superfamilies has been widely explored [6,7,66,67,68,69,70,71,72,73,74,75]. At first glance the TIM-barrel topology appears as a monodomain structure, but many biophysical measurements have highlighted discrepancies between the very complex folding pathways observed and this simple picture. Actually, $(\beta/\alpha)_8$ barrels tend to behave more like multidomain proteins, with sequential folding and unfolding of subdomain folding units [67,76]. Explaining these hierarchical folding patterns [74,77] requires partitioning the unfolded state between off-pathway transient intermediate species with substantial secondary structure and stability [78] and on-pathway equilibrium intermediate species [79].

Our experimental results suggest that while we have succeeded in creating a thermodynamically stable protein, its folding kinetics might differ considerably from that of natural small proteins and might involve multiple pathways and intermediate-state populations. Rosetta optimizes only for thermodynamic stability, without taking pathways and folding kinetics into account. The apparent noncooperative unfolding of Octarellin VI might be due to this fact. With a protein of more than 200 amino acids, the conformational space is larger than with a 100-amino-acid protein, and not taking into count the folding pathway might contribute to the problem. At this point, it is necessary to mention that we performed a 2D-NMR characterization over our artificial protein (data not shown). The result is not what we were expecting, as it shows that our protein is indeed not well folded under the tested experimental conditions. We believe this issue could be due to a wrong folding arising from the renaturation protocol. Clearly the possibility to get a soluble protein will allow a better characterization of the protein. Changing the expression system to yeast or cell lines like HEK cells could be a way to produce soluble proteins. Indeed, while the primary structure of a protein defines its tertiary structure, the environment (*in vivo* or *in vitro*) has a clear influence and impact on the final structure [2].

What Next?

Future attempts to design large proteins will thus need to integrate an adequate amino acid environment potential encompassing both solvation and aggregation energetics. Ideally, they should also incorporate some assessment of potential folding pathways and of the folding kinetics of the designed proteins. This will require learning more about sequence-structure relationships and protein folding pathways. Secondary structure predictions in combination with local energy evaluations might be a good starting point at the present time, but it remains a challenge to perform *de novo* folding simulations with trajectories approaching those observed in nature with a sufficient level of accuracy, and to use this information in the design process. Furthermore, on the basis of proteins such as Octarellin VI, one should be able to create, by directed evolution, variants that are more soluble and whose structure can be determined accurately. With a database of such mutants and their characteristics, it might be possible to deduce rules or parameter changes that could be introduced into protocols such as Rosetta.

Conclusions

We have used the Rosetta computational protein design protocol to design Octarellin VI, a 216-residue artificial protein modeled on the $(\beta/\alpha)_8$ barrel fold. The protein shows evidence of tertiary structure and high resistance to heat-induced unfolding, but low solubility and apparently noncooperative unfolding in the presence of urea. Our results highlight the need to incorporate into design protocols some assessment of potential folding pathways and of the folding kinetics of the designed proteins. Such methods remain to be developed. Secondary structure predictions, *de novo* folding simulations, and directed evolution could be starting points.

Acknowledgments

The authors would like to thank M. Dumoulin and C. Damblon for their helpful comments concerning the spectroscopic measurements. The authors thank the GIGA bioinformatics platform for technical assistance and for providing computing resources.

All molecular graphics were created with PyMOL 0.99rc6.

Author Contributions

Conceived and designed the experiments: MF NO AL KWK BMD AM JAM JM CVW. Performed the experiments: MF NO AL. Analyzed the data: MF NO AL KWK BMD AM JAM JM CVW. Contributed reagents/materials/analysis tools: MF NO AL KWK BMD AM JAM JM CVW. Wrote the paper: MF NO AL AM JM CVW.

References

1. Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181: 223–230.
2. Dobson CM (2003) Protein folding and misfolding. Nature 426: 884–890.
3. Parmeggiani F, Pellarin R, Larsen AP, Varadamsetty G, Stumpp MT, et al. (2008) Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. J Mol Biol 376: 1282–1304.
4. Urvoas A, Guellouz A, Valerio-Lepiniec M, Graille M, Durand D, et al. (2010) Design, production and molecular structure of a new family of artificial alpha-helicoidal repeat proteins (alphaRep) based on thermostable HEAT-like repeats. J Mol Biol 404: 307–327.
5. 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.
6. Houbrechts A, Moreau B, Abagyan R, Mainfroid V, Preaux G, 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.
7. Offredi F, Dubail F, Kischel P, Sarinski K, Stern AS, 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.
8. Tanaka T, Hayashi M, Kimura H, Oobatake M, Nakamura H (1994) *De novo* design and creation of a stable artificial protein. Biophys Chem 50: 47–61.
9. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247: 536–540.
10. Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, et al. (2000) The Protein Data Bank and the challenge of structural genomics. Nat Struct Biol 7 Suppl: 957–959.
11. Wierenga RK (2001) The TIM-barrel fold: a versatile framework for efficient enzymes. FEBS Lett 492: 193–198.
12. Nagano N, Orengo CA, Thornton JM (2002) One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J Mol Biol 321: 741–765.
13. Urfer R, Kirschner K (1992) The importance of surface loops for stabilizing an eightfold beta alpha barrel protein. Protein Sci 1: 31–45.
14. Beauregard M, Goraj K, Goffin V, Heremans K, Goormaghtigh E, et al. (1991) Spectroscopic investigation of structure in octarellin (a *de novo* protein designed to adopt the alpha/beta-barrel packing). Protein Eng 4: 745–749.
15. Tanaka T, Kimura H, Hayashi M, Fujiyoshi Y, Fukuhara K, et al. (1994) Characteristics of a de novo designed protein. Protein Sci 3: 419–427.
16. Tanaka T, Kuroda Y, Kimura H, Kidokoro S, Nakamura H (1994) Cooperative deformation of a *de novo* designed protein. Protein Eng 7: 969–976.
17. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, et al. (2003) Design of a novel globular protein fold with atomic-level accuracy. Science 302: 1364–1368.
18. Ambroggio XI, Kuhlman B (2006) Computational design of a single amino acid sequence that can switch between two distinct protein folds. J Am Chem Soc 128: 1154–1161.
19. Kuhlman B, Baker D (2000) Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci U S A 97: 10383–10388.
20. Dunbrack RL, Jr. (2002) Rotamer libraries in the 21st century. Curr Opin Struct Biol 12: 431–440.
21. Dunbrack RL, Jr., Cohen FE (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci 6: 1661–1681.
22. Dunbrack RL, Jr., Karplus M (1993) Backbone-dependent rotamer library for proteins. Application to side-chain prediction. J Mol Biol 230: 543–574.
23. Voigt CA, Gordon DB, Mayo SL (2000) Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. J Mol Biol 299: 789–803.
24. Schueler-Furman O, Wang C, Bradley P, Misura K, Baker D (2005) Progress in modeling of protein structures and interactions. Science 310: 638–642.
25. Skolnick J (2006) In quest of an empirical potential for protein structure prediction. Curr Opin Struct Biol 16: 166–171.
26. Poole AM, Ranganathan R (2006) Knowledge-based potentials in protein design. Curr Opin Struct Biol 16: 508–513.
27. Das R, Baker D (2008) Macromolecular modeling with rosetta. Annu Rev Biochem 77: 363–382.

28. Korkegian A, Black ME, Baker D, Stoddard BL (2005) Computational thermostabilization of an enzyme. Science 308: 857–860.
29. Kuhlman B, O'Neill JW, Kim DE, Zhang KY, Baker D (2002) Accurate computer-based design of a new backbone conformation in the second turn of protein L. J Mol Biol 315: 471–477.
30. Chevalier BS, Kortemme T, Chadsey MS, Baker D, Monnat RJ, et al. (2002) Design, activity, and structure of a highly specific artificial endonuclease. Mol Cell 10: 895–905.
31. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, et al. (2004) Computational redesign of protein-protein interaction specificity. Nat Struct Mol Biol 11: 371–379.
32. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, et al. (2008) *De novo* computational design of retro-aldol enzymes. Science 319: 1387–1391.
33. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, et al. (2008) Kemp elimination catalysts by computational enzyme design. Nature 453: 190–195.
34. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, et al. (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science 329: 309–313.
35. Butterfoss GL, Kuhlman B (2006) Computer-based design of novel protein structures. Annu Rev Biophys Biomol Struct 35: 49–65.
36. Rohl CA, Strauss CE, Chivian D, Baker D (2004) Modeling structurally variable regions in homologous proteins with rosetta. Proteins 55: 656–677.
37. Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, et al. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins 50: 437–450.
38. Melo F, Devos D, Depiereux E, Feytmans E (1997) ANOLEA: a www server to assess protein structures. Proc Int Conf Intell Syst Mol Biol 5: 187–190.
39. Melo F, Feytmans E (1998) Assessing protein structures with a non-local atomic interaction energy. J Mol Biol 277: 1141–1152.
40. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35: W407–410.
41. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5: 725–738.
42. Zhang Y (2009) I-TASSER: fully automated protein structure prediction in CASP8. Proteins 77 Suppl 9: 100–113.
43. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, et al. (2005) Protein structure prediction servers at University College London. Nucleic Acids Res 33: W36–38.
44. Lobley A, Sadowski MI, Jones DT (2009) pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. Bioinformatics 25: 1761–1767.
45. Ginalski K, Elofsson A, Fischer D, Rychlewski L (2003) 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 19: 1015–1018.
46. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, et al. (2005) GROMACS: fast, flexible, and free. J Comput Chem 26: 1701–1718.
47. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577–2637.
48. Meiler J, Muller M, Zeidler A, Schmaschke F (2001) Generation and evaluation of dimension-reduced amino acid parameter representations by artificial neural networks. Journal Of Molecular Modeling 7: 360–369.
49. Vincentelli R, Canaan S, Campanacci V, Valencia C, Maurin D, et al. (2004) High-throughput automated refolding screening of inclusion bodies. PROTEIN SCIENCE 13: 2782–2792.
50. Provencher SW, Glockner J (1981) Estimation of globular protein secondary structure from circular dichroism. Biochemistry 20: 33–37.
51. van Stokkum IH, Spoelder HJ, Bloemendal M, van Grondelle R, Groen FC (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra. Anal Biochem 191: 110–118.
52. Manavalan P, Johnson WC, Jr. (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. Anal Biochem 167: 76–85.
53. Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem 287: 252–260.
54. Sreerama N, Venyaminov S, Woody R (1999) Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. PROTEIN SCIENCE 8: 370–380.
55. Sreerama N, Woody RW (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. Anal Biochem 209: 32–44.
56. Whitmore L, Wallace B (2008) Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. BIOPOLYMERS 89: 392–400.
57. Whitmore L, Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res 32: W668–673.
58. Pace CN (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. Methods Enzymol 131: 266–280.
59. Idicula-Thomas S, Balaji P (2005) Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli. Protein Science 14: 582–592.
60. Idicula-Thomas S, Balaji PV (2007) Correlation between the structural stability and aggregation propensity of proteins. In Silico Biol 7: 225–237.
61. Van den Burg B, Dijkstra BW, Vriend G, Van der Vinne B, Venema G, et al. (1994) Protein stabilization by hydrophobic interactions at the surface. Eur J Biochem 220: 981–985.
62. Fu H, Grimsley GR, Razvi A, Scholtz JM, Pace CN (2009) Increasing protein stability by improving beta-turns. Proteins 77: 491–498.
63. Matthews BW, Nicholson H, Becktel WJ (1987) Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. Proc Natl Acad Sci U S A 84: 6663–6667.
64. Ambroggio XI, Kuhlman B (2006) Design of protein conformational switches. Curr Opin Struct Biol 16: 525–530.
65. DeLuca S, Dorr B, Meiler J (2011) Design of native-like proteins through an exposure-dependent environment potential. Biochemistry 50: 8521–8528.
66. Akanuma S, Miyagawa H, Kitamura K, Yamagishi A (2005) A detailed unfolding pathway of a (beta/alpha)8-barrel protein as studied by molecular dynamics simulations. Proteins 58: 538–546.
67. Akanuma S, Yamagishi A (2005) Identification and characterization of key substructures involved in the early folding events of a (beta/alpha)8-barrel protein as studied by experimental and computational methods. J Mol Biol 353: 1161–1170.
68. Eder J, Kirschner K (1992) Stable substructures of eightfold beta alpha-barrel proteins: fragment complementation of phosphoribosylanthranilate isomerase. Biochemistry 31: 3617–3625.
69. Forge V, Hoshino M, Kuwata K, Arai M, Kuwajima K, et al. (2000) Is folding of beta-lactoglobulin non-hierarchic? Intermediate with native-like beta-sheet and non-native alpha-helix. J Mol Biol 296: 1039–1051.
70. Gromiha MM, Pujadas G, Magyar C, Selvaraj S, Simon I (2004) Locating the stabilizing residues in (alpha/beta)8 barrel proteins based on hydrophobicity, long-range interactions, and sequence conservation. Proteins 55: 316–329.
71. Hocker B, Beismann-Driemeyer S, Hettwer S, Lustig A, Sterner R (2001) Dissection of a (beta/alpha)8-barrel enzyme into two folded halves. Nat Struct Biol 8: 32–36.
72. Luger K, Hommel U, Herold M, Hofsteenge J, Kirschner K (1989) Correct folding of circularly permuted variants of a beta alpha barrel enzyme *in vivo*. Science 243: 206–210.
73. Scheerlinck JP, Lasters I, Claessens M, De Maeyer M, Pio F, et al. (1992) Recurrent alpha beta loop structures in TIM barrel motifs show a distinct pattern of conserved structural features. Proteins 12: 299–313.
74. Silverman JA, Harbury PB (2002) The equilibrium unfolding pathway of a (beta/alpha)8 barrel. J Mol Biol 324: 1031–1040.
75. Yang X, Vadrevu R, Wu Y, Matthews CR (2007) Long-range side-chain-main-chain interactions play crucial roles in stabilizing the (beta/alpha)8 barrel motif of the alpha subunit of tryptophan synthase. Protein Sci 16: 1398–1409.
76. Gu Z, Zitzewitz JA, Matthews CR (2007) Mapping the structure of folding cores in TIM barrel proteins by hydrogen exchange mass spectrometry: the roles of motif and sequence for the indole-3-glycerol phosphate synthase from Sulfolobus solfataricus. J Mol Biol 368: 582–594.
77. Forsyth WR, Bilsel O, Gu Z, Matthews CR (2007) Topology and sequence in the folding of a TIM barrel protein: global analysis highlights partitioning between transient off-pathway and stable on-pathway folding intermediates in the complex folding mechanism of a (beta/alpha)8 barrel of unknown function from B. subtilis. J Mol Biol 372: 236–253.
78. Bilsel O, Yang L, Zitzewitz JA, Beechem JM, Matthews CR (1999) Time-resolved fluorescence anisotropy study of the refolding reaction of the alpha-subunit of tryptophan synthase reveals nonmonotonic behavior of the rotational correlation time. Biochemistry 38: 4177–4187.
79. 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.