Segregated Protein–Cucurbit[7]uril Crystalline Architectures via Modulatory Peptide Tectons

Kiefer O. Ramberg,[a] Francesca Guagnini,[a] Sylvain Engilberge,[a] Małgorzata A. Wrońska,[a] Martin L. Rennie,[a] Javier Pérez,[b] and Peter B. Crowley*[a]

Abstract: One approach to protein assembly involves water-soluble supramolecular receptors that act like glues. Bionanoarchitectures directed by these scaffolds are often system-specific, with few studies investigating their customization. Herein, the modulation of cucurbituril-mediated protein assemblies through the inclusion of peptide tectons is described. Three peptides of varying length and structural order were N-terminally appended to RSL, a β-propeller building block. Each fusion protein was incorporated into crystalline architectures mediated by cucurbit[7]uril (**Q7**). A trimeric coiled-coil served as a spacer within a **Q7**-directed sheet assembly of RSL, giving rise to a layered material of varying porosity. Within the spacer layers, the coiled-coils were dynamic. This result prompted consideration of intrinsically disordered peptides (IDPs) as modulatory tectons. Similar to the coiled-coil, a mussel adhesion peptide (Mefp) also acted as a spacer between protein–**Q7** sheets. In contrast, the fusion of a nucleoporin peptide (Nup) to RSL did not recapitulate the sheet assembly. Instead, a **Q7**-directed cage was adopted, within which disordered Nup peptides were partially "captured" by **Q7** receptors. IDP capture occurred by macrocycle recognition of an intrapeptide Phe-Gly motif in which the benzyl group was encapsulated by **Q7**. The modularity of these protein–cucurbituril architectures adds a new dimension to macrocycle-mediated protein assembly. Segregated protein crystals, with alternating layers of high and low porosity, could provide a basis for new types of materials.

Introduction

Structural knowledge of natural self-assembling protein systems is being used to engineer next-generation biomaterials with high biocompatibility and biodegradability.[1–3] Assemblies that comprise rigid and flexible subunits, such as the β-sheet polymer networks of spider silk[4] and squid ring teeth proteins,[5] are of particular interest. These materials possess semi-crystalline order and exhibit high tensile strengths and thermoplastic characteristics due to gradations in structural integrity. Few examples of engineered protein-based materials with comparable structural complexity have been reported.[3–6] Here, we describe the manufacture and structure elucidation of segregated crystalline assemblies that contain layers of low and high porosity in which the protein material is rigid or dynamic, respectively.

Engineered protein crystals[7–14] provide a basis for functional biomaterials with applications such as templated catalysis[11] and protein encapsulation.[12–14] Different engineering approaches currently studied include deletion mutations to generate cavities,[7,11] metal-directed protein–protein interfaces,[8] as well as naturally porous (fusion-)protein crystals.[12–14] Ligand-mediated assembly is another route to tuneable protein architectures. Bivalent ligands[15] and anionic calixarenes[9,10] have been used to generate (porous) protein crystals that depend on protein–ligand and ligand–ligand contacts. The commercially available cucurbit[n]urils (**Q**n, $n = 6$, 7, 8), which recognise a range of protein features, are also useful mediators of protein assembly.[16–24] **Q**n are neutral, rigid doughnut structures with a capacity to self-assemble into clusters.[20–22,25] **Q8** mediates protein oligomerization by hosting two N-terminal aromatic residues.[17,18] The smaller, more water-soluble **Q7** accommodates just one N-terminal aromatic residue[16] and also recognizes methylated lysines.[20–22,26] Taking advantage of the latter interaction, we have engineered several biohybrids of **Q7** and the model protein RSL in its methylated form (RSL*).[20–22]

RSL is an exceptionally stable trimeric β-propeller lectin.[10,20–23,27,28] **Q7** complexation of RSL* involves selective binding of the most sterically accessible dimethyllysine (Lys34*), aided by peripheral interactions with Tyr37. Previously, we showed that co-crystallization of RSL* and **Q7** results in two distinct architectures with cage- or sheet-like assemblies mediated by cucurbituril clustering.[20] The RSL*–**Q7** cage

[a] K. O. Ramberg, Dr. F. Guagnini, Dr. S. Engilberge, Dr. M. A. Wrońska, Dr. M. L. Rennie, Prof. P. B. Crowley
School of Chemistry
National University of Ireland Galway
University Road, Galway, H91 TK33 (Ireland)
E-mail: peter.crowley@nuigalway.ie

[b] Dr. J. Pérez
Synchrotron SOLEIL
L'Orme des Merisiers
Saint-Aubin BP 48, 91192 Gif-sur-Yvette Cedex (France)

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202103025

**Chemistry
Europe**
European Chemical
Societies Publishing

assembly (space group *F*432, Figure 1A), comparable in dimension to ferritin,[8] involves tetrameric **Q7** clusters. In addition to the Lys34* site, a distal interaction occurs between **Q7** and the dimethylated N terminus (Ser1*). In the RSL*-**Q7** sheet assembly (space group *C*222$_1$, Figure 1B), trimeric **Q7** clusters coordinate a honeycomb arrangement of RSL*, leading to protein–**Q7** "bilayers". These bilayers can be engineered to include additional **Q7** binding sites[21] or metal-binding sites, which rearrange the order of the bilayers.[22]

In this work, our goal was to engineer new features into the RSL*–**Q7** crystalline architectures. Specifically, we hypothesized that the "bilayer" sheet assembly could accommodate spacers. To determine their impact on RSL*–**Q7** assemblies, peptide tectons were fused to the N terminus of RSL (Table 1, below). A parallel trimeric coiled-coil (PDB ID: 4DZN), designed previously as a modulatory protein building block,[29] was chosen based on geometric matching between its C termini and the N termini of trimeric RSL (Figure S1 in the Supporting Information). Three heptad repeats of 4dzn were fused to RSL via a Gly-Ala linker, resulting in a protein nanoobject with globular and coiled-coil subunits of similar dimensions (~4 nm). RSL fusions with intrinsically disordered peptides were tested also. Features from a mussel adhesion foot protein (Mefp)[30] or a nucleoporin

(Nup)[31] were fused to RSL. Such fusions do not require geometric matching with RSL as the peptide extensions are expected to be disordered and to behave independently, notwithstanding self-association. Mefp and Nup sequences were chosen partly due to their desirability as components in biomaterials,[30,31] but also because they possess potential **Q7** binding sites; namely, the Lys/Tyr residues of Mefp and the Phe-Gly motifs of Nup.

Three RSL fusion proteins were tested (Table 1) and were each found to assimilate into one of the two known **Q7**-mediated architectures (Figure 1). Despite the availability of new potential cucurbituril binding sites, the dimethylated fusion proteins (Mefp–RSL*, Nup–RSL*, and 4dzn–RSL*) conformed to either the cage or sheet **Q7**-mediated assemblies. Cucurbituril capture of a disordered extension occurred within the cage assembly. Reconstitution of the **Q7**-directed sheet assembly resulted in segregated crystalline materials with layers of alternating structural order. These results demonstrate RSL*–**Q7** biohybrids as customizable protein-based crystalline materials. The segregated structures, with regions of high and low porosity, might provide a basis for biohybrid actuators.[3,32]
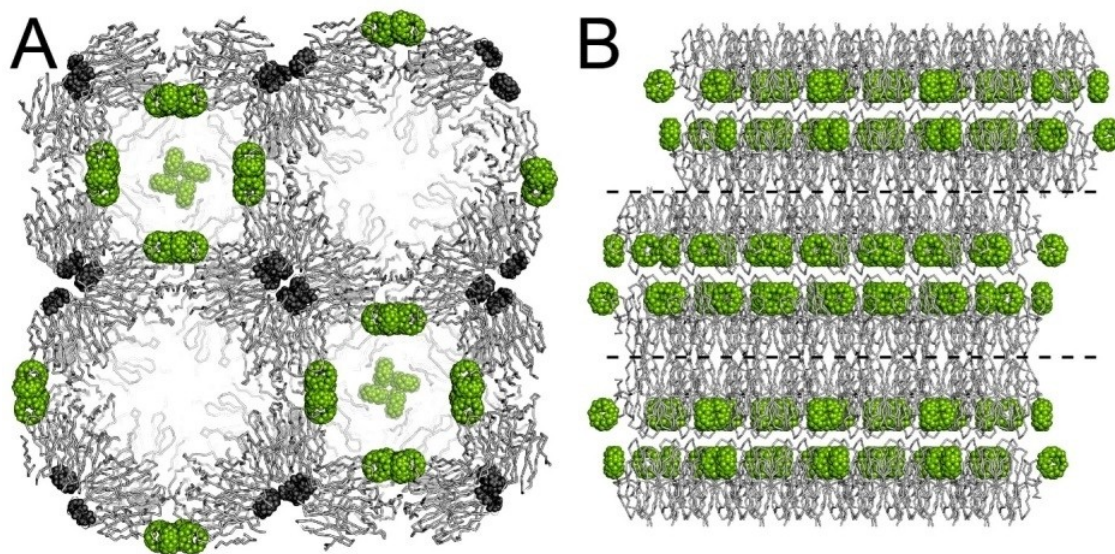

Figure 1. RSL*–**Q7** co-crystals occur as either the A) cage (PDB ID: 6F7X) or B) sheet (PDB ID: 6F7 W) assembly.[20] Dashed lines in (B) denote the RSL*–**Q7** "bilayers" and the location of the N termini. Proteins and **Q7** are shown as grey ribbons and green spheres, respectively. **Q7** at the distal binding site in the cage assembly is shown as dark grey spheres.

Table 1. Properties of RSL and fusion proteins.

Protein	N terminus[a]	Trimer MWt (kDa)			Solubility in H$_2$O
		Predicted	SEC-MALS	SAXS[b]	
RSL	**SSV**	29.2	29.3	28.9	high
Mefp–RSL	SYKGKYYG**SV**	31.8	31.6	32.8	high
Nup–RSL	MFGSTLFGSTLFG**SV**	33.0	33.0	33.4	moderate
4dzn–RSL	SSEIAALKQEIAALKKEIAALKAGA**SV**	36.4	36.3	37.3	moderate

[a] RSL N terminus in bold. 4dzn heptad repeats highlighted grey. Dimethylation sites underlined. [b] MWt estimates based on SAXS data for 0.125 mM samples.

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202103025

Chemistry Europe
European Chemical
Societies Publishing

Results

Fusion protein preparation and characterization

Three fusion proteins (Table 1) were produced in *Escherichia coli* BL21 and purified by mannose-affinity chromatography.[27] The protein purity was confirmed by SDS-PAGE analysis (Figure S2). Mefp–RSL co-eluted with nucleic acids and required treatment with DNase prior to purification by size-exclusion chromatography. Attempts to purify a longer version of Mefp–RSL, with double the Lys/Tyr content were unsuccessful due to nucleic acid complexation (data not shown). Molecular weight determination by ESI-MS was straightforward for Mefp–RSL and 4dzn–RSL (Figure S3 and Table S1). However, Nup–RSL was recalcitrant to this method and yielded no discernible peaks, despite repeated attempts.

Peptide fusion to globular proteins is a common approach to engineered assembly,[33,34] but can result in unintended oligomeric states and altered solubilities.[35,36] Whereas RSL is highly soluble (10 mM in pure water), Nup–RSL and 4dzn–RSL tended to precipitate at >2 and >1 mM, respectively. The precipitation of 4dzn–RSL was overcome by the inclusion of 100 mM NaCl, enabling samples up to 5 mM. Size-exclusion chromatography with multi-angle light scattering (SEC-MALS) was used to determine the molecular weight, oligomeric state and monodisperity of the fusion proteins. Each of the fusions (1 mM) eluted from a Superdex 75, GL 10/300 column earlier than wild-type RSL (Figure 2A). Whereas RSL eluted at 16.9 mL, Mefp–RSL and Nup–RSL eluted at 16.6 mL. The largest fusion protein, 4dzn–RSL, eluted at 16.0 mL. All of the proteins eluted as sharp well-defined peaks indicating high monodisperity. In each case, the derived molecular weight was in agreement with the predicted values of the trimer, confirming that the N-terminal fusions did not alter the oligomeric state (Table 1).

Further size and shape characterization was achieved by small-angle X-ray scattering (SAXS) experiments, performed in batch mode (Figures 2B and S4, Tables S2 and S3). At 0.125 mM protein, the SAXS profiles for Mefp–RSL, Nup–RSL and 4dzn–RSL were similar to that of wild-type RSL[21,27] and distance distribution plots corroborated the sample uniformity suggested by SEC-MALS (Figure S4). The calculated radii of gyration (R_g) for Mefp–RSL (~21 Å), Nup–RSL (~20 Å) and 4dzn–RSL (~22 Å) were significantly higher than that of RSL (~18 Å).[21,27] The maximum particle size (D_{max}) of 4dzn–RSL (8.3 nm) was approximately double that of RSL (4.6 nm), in agreement with the design (Figure S1). The slightly longer R_g and D_{max} values for Mefp–RSL compared to Nup–RSL suggest a more extended conformation of the shorter cationic peptide compared to the longer hydrophobic peptide. SAXS estimates of the molecular weights were in agreement with the predicted values for each construct (Table 1).

At 1 mM protein, the SAXS profiles and derived biophysical parameters for Mefp–RSL and 4dzn–RSL are similar to those obtained with the dilute samples (Figure S4 and Table S3). However, the shape of the scattering profile for Nup–RSL was distinctly different at <0.2 Å, with a sharp increase in intensity at <0.02 Å suggesting aggregation at high concentration.
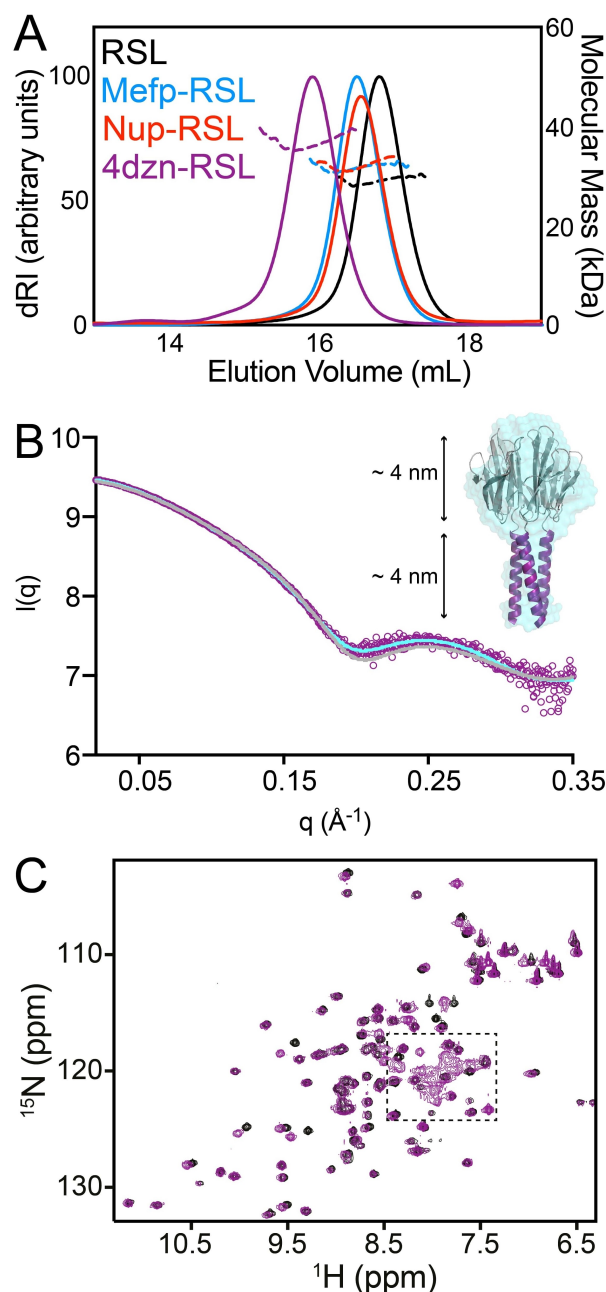

Figure 2. Biophysical characterization of RSL fusion proteins. A) SEC-MALS analysis with dashed and solid lines corresponding to mass average molar masses and differential refractive index (dRI) elution profiles, respectively. B) 4dzn–RSL SAXS profile with the CRYSOL-computed scattering of the trimer in grey and fit of the *ab initio* bead model in cyan. The crystal structure (PDB ID: 6F37, with the RSL and 4dzn subunits in grey and purple) is overlaid with the bead model (cyan). C) Overlaid ^{1}H,^{15}N HSQC spectra of RSL (black contours) and 4dzn–RSL (purple). The dashed box indicates the crowded region that contains resonances of the heptad repeats.

Distance distribution plots further indicated Nup–RSL self-association with a D_{max} of 14.6 nm indicating trimerization of Nup–RSL trimers. Kratky plots for the fusion proteins did not display significantly elevated scattering at higher angles compared to wild-type RSL, consistent with the appendage of dynamic subunits (Figure S5).[37,38] This result might be due to

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202103025

Chemistry Europe
European Chemical
Societies Publishing

the disordered nature of the native RSL N termini, which lack electron density in various crystal structures.[10] Therefore, extension of the termini with additional flexible subunits may not affect significantly the X-ray scattering. Another possibility is that the N-terminal extensions possess some degree of order. Structural order was expected for the coiled-coil of 4dzn–RSL, whereas the Mefp–RSL and Nup–RSL extensions were predicted to be disordered. The close proximity of these extensions afforded by fusion to the RSL N termini (~13 Å separation) might drive their self-assembly. Adjacent Mefp peptides may associate in a manner similar to Mefp biofilms, through extensive cation–π interactions.[30] Self-assembly of nucleoporin peptides is also well established.[31] Modelling of the fusion proteins offered further insight into the structural order of the N-terminal extensions.

Bead models were obtained for RSL and fusion proteins by using a combination of *ab initio* (DAMMIF, Table S4) and hybrid rigid body (CORAL) modelling approaches.[37,38] The N-terminal peptide tails (Table 1) were modelled as 7 (Mefp–RSL) or 12 (Nup–RSL) dummy residues linked to each N terminus of the RSL trimer. The rigid spheroid of the RSL trimer[21,27] was evident in all of the bead models (Figures 2B, S6 and S7). A well-defined cylindrical shape, consistent with a folded helical bundle, was apparent in the 4dzn–RSL model, which was superposable with the X-ray diffraction model (Figure 2B). The Mefp–RSL *ab initio* model exhibited a less defined shape protruding from the well-defined RSL. The hybrid model suggested self-association of the Mefp extensions (Figure S6). The *ab initio* model derived from the 0.125 mM Nup–RSL SAXS data was a compact spheroid akin to RSL with no obvious additional shape that could be attributed to the N-terminal Nup extensions (Figure S7). The Nup–RSL hybrid model fit the SAXS data similar to the *ab initio* model and suggested that the fusions do not extend away from the trimer, but rather collapse across the RSL N-terminal face. This compacted configuration of the Nup extensions was consistent with Nup–RSL displaying a smaller R_g and D_{max} compared to Mefp–RSL, despite Nup comprising more residues that Mefp.

Each of the fusion proteins was characterized also by NMR spectroscopy. The $^1H,^{15}N$ HSQC spectra included well-dispersed peaks that correspond to the spectrum of wild-type RSL, indicating that the N-terminal fusions did not alter the tertiary structure of RSL (Figure 2C and S8). In addition to the RSL peaks, several sharp crosspeaks were observed for Mefp–RSL and

Nup–RSL. At least four new peaks were evident in the Mefp–RSL spectrum, two of which occur in the glycine region consistent with the extra two Gly residues in this construct. Twelve new peaks were evident in the spectrum of Nup–RSL, three of which occur in the glycine region consistent with the three Phe-Gly repeats in this fusion (Table 1). A new peak was observed also in the glycine region of the 4dzn–RSL $^1H,^{15}N$ HSQC spectra, corresponding to the Gly-Ala linker (Figure 2C). As is characteristic for coiled-coils,[39] the resonances of the heptad repeats were crowded in a defined spectral region (^1H ~7.4–8.5 ppm, ^{15}N ~118–124 ppm). Selective ^{15}N Lys-labelling of 4dzn–RSL resulted in a $^1H,^{15}N$ HSQC spectrum with seven crosspeaks accounting for the three lysines of RSL and the four lysines of 4dzn (Figure S9). These spectra confirmed the correct folding of 4dzn–RSL and 4dzn–RSL*, with no significant chemical shift perturbations in the chemically dimethylated form.

Crystal forms and architectures

Co-crystallization of RSL* and **Q7** occurs in simple conditions comprising PEG precipitant, 0.1 M buffer, and up to 0.2 M additives.[20–22] Co-crystals of **Q7** and the methylated fusion proteins were obtained by sitting-drop vapour-diffusion experiments with commercial (Jena JCSG + +) or homemade screens and grew in conditions similar to those previously reported for RSL* and **Q7** (Table 2 and Figure S10). Data collection at SOLEIL synchrotron revealed that two of the fusions adopted the sheet assembly, while one conformed to the cage architecture (Tables S5 and S6).

4dzn–RSL crystallized readily from solutions of PEG 8000 at pH 9.5. The resulting assembly, space group $P2_13$, was densely packed with a solvent content of 49% (Figure 3A). Interestingly, the crystal packing of the 4dzn–RSL nanoobject is dominated by interactions between the globular (RSL) and coiled-coil (4dzn) subunits, with minimal RSL–RSL or 4dzn–4dzn interactions (Table S7). PDB 4DZN is a densely packed crystal in which the coiled-coils form a herringbone weave (Figure S11).[29] Previously, coiled-coil-directed bionanoarchitectures relied on the fusion of a single coiled-coil peptide to globular oligomers.[33] Here, the coiled-coil subunit of 4dzn–RSL is already in its preferred oligomeric state, presenting the opportunity for association with its globular counterpart. Visualization of the 4dzn–RSL assembly by temperature factor, a

Table 2. Crystallization details and structure properties.

Protein	Ligand	Crystallization conditions	[Protein] [mM]	[Q7] [mM])	Res [Å]	Space group	PDB ID
Mefp–RSL*	**Q7**	15% PEG 10000 0.1 M Bis·Tris pH 5.5 0.2 M MgCl₂	1.0	4.0	2.1	*H*3	72PH
Nup–RSL*	**Q7**	25% PEG 3350 0.1 M Bis·Tris pH 5.5 0.2 M MgCl₂	1.5	7.5	1.5	*F*432	72PI
4dzn–RSL*	**Q7**	20% PEG 8000 0.1 M Tris·HCl pH 8.5 0.2 M MgCl₂	1.0	2.0	2.0 / 2.6	*H*3 / *P*1	72PJ / 6S99
4dzn–RSL	–	20% PEG 8000 0.1 M CHES pH 9.5	1.0	–	2.2	$P2_13$	6F37

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202103025

**Chemistry
Europe**
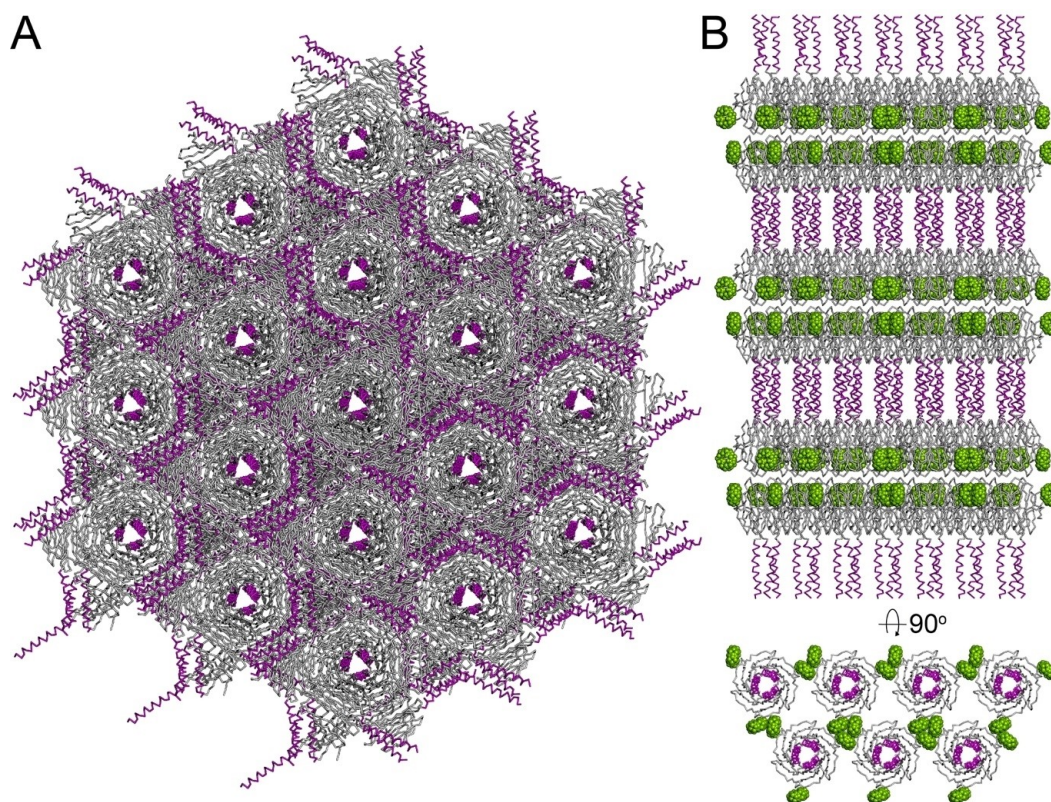European Chemical
Societies Publishing

Figure 3. Crystal packing in A) 4dzn–RSL (PDB ID: 6F37, $P2_13$) and B) 4dzn–RSL* plus **Q7** (PDB ID: 72PJ, $H3$). The RSL and 4dzn subunits are grey and purple, respectively. Structures drawn to scale.

gauge of flexibility/motion,[40] reveals that the RSL and 4dzn subunit dynamics vary widely, yielding distinct regions of high and low rigidity (Figure S12).

Co-crystals of 4dzn–RSL* and **Q7** were obtained in JCSG++ condition D6 at 2–4 equiv. **Q7** and were reproduced manually (Table 2, Figure S10). No crystals were observed below 2 equiv. **Q7**. Co-crystals with a thin plate-like morphology diffracted with high mosaicity and were solved in $P1$ at 2.6 Å resolution, with two 4dzn–RSL* trimers and six **Q7** molecules per asymmetric unit. Co-crystals with a thicker plate morphology had improved diffraction properties and yielded a 2.0 Å resolution structure in $H3$ with two 4dzn–RSL* monomers and two **Q7** molecules per asymmetric unit. Recapitulation of the **Q7**-mediated sheet assembly was evident in both co-crystal structures (Figure 3B). Remarkably, the coiled-coils act as spacers and modulate the assembly of RSL*-**Q7** bilayers, yielding a segregated architecture. The contrast with the packing in 4dzn–RSL crystals is stark. The interactions between the RSL and 4dzn subunits are completely overpowered by the **Q7**-directed assembly.

The 4dzn–RSL assemblies depicted in Figure 3 differ not only in crystalline architecture, but also in helical bundle dynamics. In crystals of 4dzn–RSL alone, the electron density for the helical bundles is well-defined. In contrast, in co-crystals of 4dzn–RSL* and **Q7**, the electron density for the 4dzn bundles is poor (Figure S13) indicating that the coiled-coils are highly dynamic. This result appears to be a consequence of the high porosity and the lack of stabilizing crystal contacts in the spacer layers. Poorly defined electron density was observed also for coiled-coils fused to a homopentameric globular protein within a tubular crystalline assembly.[33]

Considering the flexibility of the bundles, the successful recapitulation of the sheet assembly is surprising. Apparently, rigid tectons are not required. This observation prompted replacement of 4dzn with more dynamic peptides. Mefp- and Nup–RSL fusion proteins were produced to evaluate IDPs as potential modulatory tectons. Although no tenable crystallization leads for pure Mefp–RSL or Nup–RSL (either native or dimethylated) were identified, co-crystallization of both Mefp–RSL* and Nup–RSL* with **Q7** was achieved. Co-crystals of Mefp–RSL* and **Q7** were of a similar triangular plate morphology to co-crystals obtained with 4dzn–RSL* (Figure S10) and also diffracted well, yielding a 2.1 Å resolution dataset which was solved in $H3$ with two Mefp–RSL* monomers and two **Q7** molecules per asymmetric unit. Like 4dzn–RSL*, Mefp–RSL* adopted the **Q7**-directed sheet assembly with the N-terminal extensions functioning as spacers. The peptide spacers within the Mefp–RSL* and **Q7** assembly are even more dynamic. Despite obvious modulation of the sheet assembly, negligible electron density was present for the Mefp extensions and no Mefp residues could be modelled.

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202103025

Chemistry Europe
European Chemical
Societies Publishing

Co-crystals of Nup–RSL* and **Q7** (Figure S10) were identified in JCSG++ condition H11 and were reproduced manually. These co-crystals grew only at high ligand excess (~5 equiv. **Q7**) and were the *F*432 space group with cell parameters (~200 Å³) identical within error to the original structure (PDB ID: 6F7X).[20] In the original RSL*–**Q7** cage structure, the distal **Q7** cavity is empty (Figure 4A). Remarkably, in the Nup–RSL* co-crystal with **Q7**, a Phe-Gly motif of one Nup extension per trimer was captured by the distal cucurbituril through inclusion of a Phe side chain (Figure 4B). No electron density was evident for other residues of the Nup extensions, which remain disordered within the solvent voids of the cage assembly (Figure 4C). **Q7** complexation with N-terminal Phe has been characterised in a variety of protein systems.[16,41] Here, we present **Q7** binding of an intrapeptide Phe. This interaction is facilitated by the placement of the distal **Q7** close to the RSL* N termini within the *F*432 architecture. Notably, the distal **Q7** adopts a single conformation within the co-crystals with Nup–RSL* (Figure 4B). In contrast, alternate conformations of **Q7** were observed at the N-terminal Phe of human insulin.[16] Possibly the **Q7** rim interaction with the dimethylamino of Lys34* combined with the inclusion of the aromatic guest facilitates **Q7** conformation selection within co-crystals with the IDP fusion.

NMR spectroscopy of protein-Q7 interactions

Co-crystals of Nup–RSL* and **Q7** grew only at high ligand excess with no crystals observed at <5 equiv. **Q7**. The requirement for excess **Q7** may be a consequence of an increased number of **Q7** binding sites in Nup–RSL*. NMR-monitored titrations of Nup–RSL* and **Q7** yielded significant chemical shift perturbations for resonances at or near the Lys34* binding site.[20,21] In addition, there was severe broadening of the Nup resonances, even at low equiv. of **Q7** (Figure S14). RSL contains one Phe residue (Phe41). Therefore selective labelling of Nup–RSL* with ¹⁵N Phe resulted in a ¹H,¹⁵N HSQC spectrum with just four resonances. **Q7** titrations gave rise to severe broadening of the three Phe resonances of Nup (Figure S15), suggesting binding

at each of the Phe-Gly motifs. In contrast, titration of **Q7** into selectively ¹⁵N Lys-labelled 4dzn–RSL* indicated negligible interaction between **Q7** and the dimethylysines in the coiled-coil (Figure S16). Including Lys34*, **Q7** appears to bind Nup–RSL* at four sites in solution, compared to just one site in RSL*.[20] *In crystallo*, only the Nup Phe-Gly motif within reach of the distal **Q7** binding site is bound and ordered (Figure 4B). **Q7**-Lys34* binding is in fast exchange with a K_d in the mM range.[20] At 2.2 or 6.0×10^4 M⁻¹, respectively, the **Q7** affinities for an intrapeptide Phe[42] or Lys*[26] are similar, and the NMR data indicate simultaneous binding of the Lys34* and the Phe-Gly sites, even at low equiv. **Q7** (Figure S14). The requirement for excess **Q7** to obtain co-crystals with Nup–RSL* is consistent with binding at the Phe-Gly motifs which acts to limit the **Q7** available for complexation at the Lys34* sites. This competition for Lys34* binding may also explain the observation that Nup–RSL* did not conform to the **Q7**-directed sheet architecture, which requires complete *cucurbiturilation*[21] of all available Lys34* to facilitate the honeycomb-like packing. However, Nup might also be unsuitable as a spacer due to its compaction, as suggested by the SAXS data (Figure S7).

Segregated protein-Q7 architectures

A comparison of the crystal packing within Mefp–RSL*-**Q7** and 4dzn–RSL*-**Q7** by normalized temperature factors[40] revealed striking differences in the dynamics between the protein–**Q7** bilayers and the spacer layers (Figure 5). The bilayers within the fusion-**Q7** co-crystals are rigid at the RSL*–RSL* interfaces, with dynamics increasing in regions close to the spacer layers. The flexibility of the 4dzn bundles is apparent (Figure S13), while the Mefp peptides are completely disordered (Figure 5). A longer N-terminal extension translates to wider separation of the RSL*-**Q7** sheets as indicated by the ~3 and ~2 nm gaps observed in the 4dzn–RSL* and Mefp–RSL* assemblies, respectively (Figure 5). **Q7** co-crystals with 4dzn–RSL*, which is approximately twice the length of RSL*, possess a *c*-axis approximately double that of RSL*–**Q7** co-crystals. With a
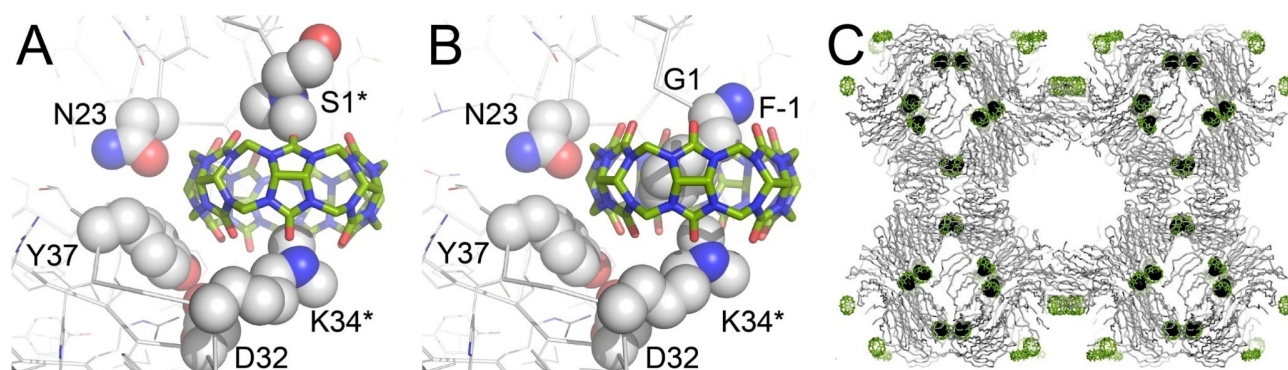

Figure 4. Protein–**Q7** interactions at the distal binding site in the *F*432 assembly of A) RSL*–**Q7** (PDB ID: 6F7X)[20] and B) Nup–RSL*–**Q7**. Protein shown as grey ribbon with key **Q7**-interacting side chains shown as spheres. C) Packing in co-crystals of Nup–RSL* and **Q7**. Phe-1 and cucurbituril are shown as black spheres and green sticks, respectively.

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202103025

Chemistry
Europe
European Chemical
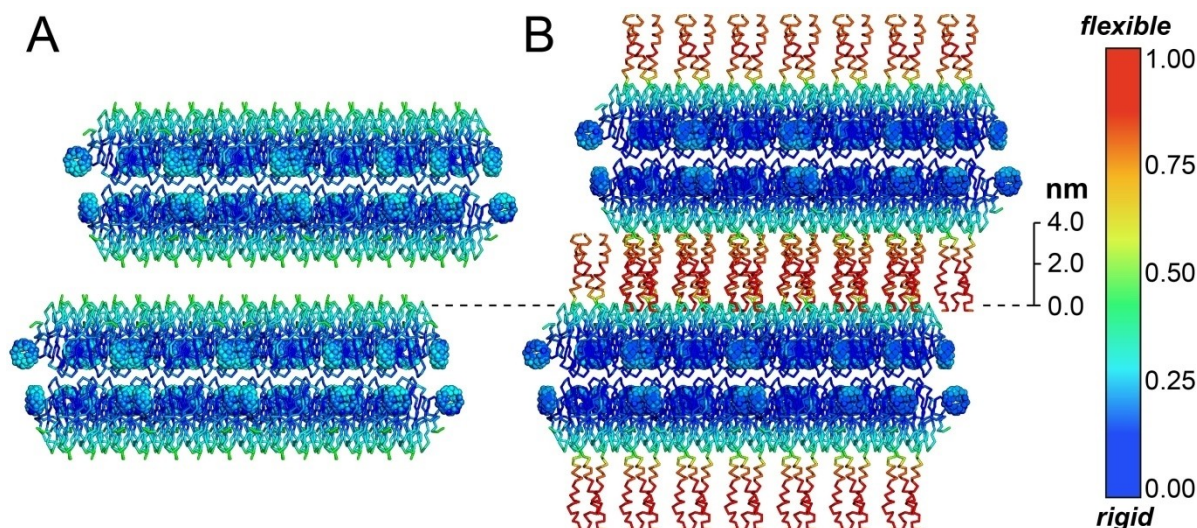Societies Publishing

Figure 5. Crystal packing in co-crystals of **Q7** with A) Mefp–RSL* or B) 4dzn–RSL*. The structures are drawn to scale and coloured by normalized temperature factors (scale bar). Nanometre-scale gaps between the RSL*–**Q7** bilayers are indicated. Proteins and **Q7** are shown as ribbons and spheres, respectively.

shorter extension, co-crystals of Mefp–RSL* and **Q7** display a *c* axis intermediate of the two (Table 3).

In contrast to well-defined RSL*-**Q7** bilayers with a solvent content of 54%, the spacer layers of the fusion proteins are sparsely populated, containing one Mefp or 4dzn portion of ~3 or ~8 kDa, respectively per cucurbiturilated RSL* trimer (~34 kDa). Consequently, the solvent contents of these diffuse regions are approximately 90 and 80% in Mefp–RSL* and 4dzn–RSL*, respectively (Table 3). Porous protein frameworks are being developed as a means of sequestering/co-crystallizing guest molecules of interest. For example, dye and polyamine guests have been captured within porous polyhedron[7] and protein–calixarene frameworks,[9] respectively. Protein cargos of 8, 9, or 32 kDa have been accommodated within porous crystalline assemblies of the endonuclease domain R1EN,[12] the BTB domain from the transcription repressor BCL6,[14] and the bacterial Cry3Aa,[13] respectively. Larger industrial enzymes have been accommodated in crystals of engineered polyhedrin.[11] The RSL*–**Q7** sheet architecture can also be used as a scaffold to accommodate fused protein subunits, as evidenced with Mefp–RSL* and 4dzn–RSL*. Whereas other systems have relied on insertion of fusions into large solvent voids,[12–14] here, the N-terminal fusions are wedged in between RSL*–**Q7** bilayers (Figures 3 and 5). With pores of >10 nm in diameter, the **Q7**-directed assemblies of Mefp–RSL* and 4dzn–RSL* themselves may be applicable to guest capture.

The high porosity and disorder of the spacer layers in the Mefp–RSL* and 4dzn–RSL* co-crystals with **Q7**, raises the question of what holds the crystal together. The Mefp fusion, in particular, is completely disordered yet amenable to crystallization. These segregated crystals comprising alternating dense and sparse layers (Figure 5) bear some resemblance to silk and squid ring teeth proteins that possess crystalline and amorphous domains.[3–5] While the spacer layers may accommodate structural rearrangement, it remains to be tested how the crystals respond to physical deformation.

Conclusions

Protein-based crystalline materials are advancing in sophistication,[6–15] including in vivo construction[6,7,11,13] and industrial application,[11,13] although reliable self-assembling systems are scarce. Tried-and-tested systems such as ferritin[8] and polyhedrin[7,11] are the main focus while ligand-mediated assembly continues to deliver remarkable bionanoarchitectures with the potential for customization.[8–10,15,20–22]

RSL, with its high rigidity and symmetry, has been used extensively in bio-supramolecular chemistry.[10,20–22] Engineering RSL via point mutations or N-terminal extensions is cost effective with high yields obtained from bacterial cell culture with one-step purification (mannose-affinity chromatography).[10,20–23,27,28] Here, we have presented three RSL fusion proteins of varying structural order (Figure 2) and demonstrated their incorporation into cucurbituril-directed assemblies (Figures 3–5). **Q7**-mediated crystalline architectures endured even with significant augmentation of the base protein

Table 3.	Characteristics of protein-**Q7** sheet assemblies.					
Structure	Space group	$a \times b \times c$ [Å]	Pore Ø[a] [nm]	Solvent content [%][b]		
				Overall	RSL*/Q7 bilayer	Spacer layer
RSL*–**Q7**	C222₁	50× 87×147	2.2	54	54	–
Mefp–RSL*-**Q7**	H3	51× 51×276	14.2	60	54	90
4dzn–RSL*-**Q7**	H3	52× 52×323	12.0	58	54	80

[a] Diameter of widest pore calculated in MAP_CHANNELS.[43] [b] Matthews solvent content estimated from total mass (protein plus ligand).

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202103025

Chemistry Europe
European Chemical
Societies Publishing

tecton (RSL). 4dzn and Mefp extensions were packed between RSL*–**Q7** bilayers (Figures 3B and 5). In the case of Nup–RSL*, the distal **Q7** within the cage assembly presented the opportunity for partial capture of the Nup peptide through inclusion of the Phe-1 side chain (Figure 4B). This result complements previous data that indicated that macrocyclic glues facilitate crystallization of proteins even with significant disorder.[44] **Q7** capture of the Nup peptide is a promising development in host–guest chemistry within protein-macrocycle frameworks.[9]

The RSL*–**Q7** sheet assembly is evidently amenable to customization, with previous studies demonstrating Lys or His enrichment as a means of modulating the packing with increased cucurbituril content or Zn-binding sites.[21,22] Here, N-terminal extensions acted as spacers between the rigid RSL*–**Q7** bilayers, giving rise to bipartite or segregated architectures with layers of high flexibility and porosity interspersed with rigid, low-porosity segments (Figure 5 and Table 3). Apparently, the characteristics of the spacer layers are dictated by the length and structural order of the peptide extension (Figure 5). Incorporation of Mefp or 4dzn extensions into the **Q7**-directed sheet architecture resulted in a more than five-fold increase in pore size relative to the original assembly (Table 3). Thus, we have illustrated macrocycle-mediated assembly as a route to dynamic, multicomponent biomaterials with tailorable qualities. In general, the incorporation of the different N-terminal extensions within **Q7**-mediated architectures suggests that these and similar protein-macrocycle architectures could be used to sequester protein cargoes with potential applications including structure elucidation[12,14] and catalysis.[11,13]

Supporting Information

The Supporting Information includes experimental methods, SAXS data and modelling, crystallographic tables, *B*-factor analysis and NMR data.

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

[1] Z. Liu, M. A. Meyers, Z. Zhang, R. O. Ritchie, *Prog. Mater. Sci.* **2017**, *88*, 467–498.
[2] S. L. Kuan, F. R. G. Bergamini, T. Weil, *Chem. Soc. Rev.* **2018**, *47*, 9069–9105.
[3] M. Eder, S. Amini, P. Fratzl, *Science* **2018**, *362*, 543–547.
[4] S. Keten, Z. Xu, B. Ihle, M. J. Buehler, *Nat. Mater.* **2010**, *9*, 359–367.
[5] S. H. Hiew, A. Sánchez-Ferrer, S. Amini, F. Zhou, J. Adamcik, P. Guerette, H. Su, R. Mezzenga, A. Miserez, *Biomacromolecules* **2017**, *18*, 4240–4248.
[6] S. Abe, T. T. Pham, H. Negishi, K. Yamashita, K. Hirata, T. Ueno, *Angew. Chem. Int. Ed.* **2021**, *60*, 12341–12345.
[7] S. Abe, H. Tabe, H. Ijiri, K. Yamashita, K. Hirata, K. Atsumi, T. Shimoi, M. Akai, H. Mori, S. Kitagawa, T. Ueno, *ACS Nano* **2017**, *11*, 2410–2419.
[8] J. B. Bailey, F. A. Tezcan, *J. Am. Chem. Soc.* **2020**, *142*, 17265–17270.
[9] S. Engilberge, M. L. Rennie, E. Dumont, P. B. Crowley, *ACS Nano* **2019**, *13*, 10343–10350.
[10] K. O. Ramberg, S. Engilberge, T. Skorek, P. B. Crowley, *J. Am. Chem. Soc.* **2021**, *143*, 1896–1907.
[11] T. K. Nguyen, S. Abe, M. Kasamatsu, B. Maity, K. Yamashita, K. Hirata, M. Kojima, T. Ueno, *ACS Appl. Nano Mater.* **2021**, *4*, 1672–1681.
[12] N. Maita, *J. Am. Chem. Soc.* **2018**, *140*, 13546–13549.
[13] B. S. Heater, Z. Yang, M. M. Lee, M. K. Chan, *J. Am. Chem. Soc.* **2020**, *142*, 9879–9893.
[14] T. Zacharchenko, S. Wright, *IUCrJ* **2021**, *8*, 154–160.
[15] F. Sakai, G. Yang, M. S. Weiss, Y. Liu, G. Chen, M. Jiang, *Nat. Commun.* **2014**, *5*, 4634.
[16] J. M. Chinai, A. B. Taylor, L. M. Ryno, N. D. Hargreaves, C. A. Morris, P. J. Hart, A. R. Urbach, *J. Am. Chem. Soc.* **2011**, *133*, 8810–8813.
[17] C. Hou, Z. Huang, Y. Fang, J. Liu, *Org. Biomol. Chem.* **2017**, *15*, 4272–4281.
[18] P. J. de Vink, J. M. Briels, T. Schrader, L. G. Milroy, L. Brunsveld, C. Ottmann, *Angew. Chem. Int. Ed.* **2017**, *56*, 8998–9002.
[19] Z. Hirani, H. F. Taylor, E. F. Babcock, A. T. Bockus, C. D. Varnado, C. W. Bielawski, A. R. Urbach, *J. Am. Chem. Soc.* **2018**, *140*, 12263–12269.
[20] F. Guagnini, P. M. Antonik, M. L. Rennie, P. O'Byrne, A. R. Khan, R. Pinalli, E. Dalcanale, P. B. Crowley, *Angew. Chem. Int. Ed.* **2018**, *57*, 7126–7130.
[21] F. Guagnini, S. Engilberge, K. O. Ramberg, J. Perez, P. B. Crowley, *Chem. Commun.* **2020**, *56*, 360–363.
[22] F. Guagnini, S. Engilberge, R. J. Flood, K. O. Ramberg, P. B. Crowley, *Cryst. Growth Des.* **2020**, *20*, 6983–6989.
[23] K. O. Ramberg, S. Engilberge, F. Guagnini, P. B. Crowley, *Org. Biomol. Chem.* **2021**, *19*, 837–844.
[24] W. Cao, X. Qin, Y. Wang, Z. Dai, X. Dai, H. Wang, W. Xuan, Y. Zhang, Y. Liu, T. Liu, *Angew. Chem. Int. Ed.* **2021**, *60*, 11196–11200.
[25] X. Yang, R. Wang, A. Kermagoret, D. Bardelang, *Angew. Chem. Int. Ed.* **2020**, *59*, 21280–21292.
[26] M. A. Gamal-Eldin, D. H. Macartney, *Org. Biomol. Chem.* **2013**, *11*, 488–495.
[27] P. M. Antonik, A. M. Eissa, A. R. Round, N. R. Cameron, P. B. Crowley, *Biomacromolecules* **2016**, *17*, 2719–2725.
[28] J. P. Ribeiro, S. Villringer, D. Goyard, L. Coche-Guerente, M. Höferlin, O. Renaudet, W. Römer, A. Imberty, *Chem. Sci.* **2018**, *9*, 7634–7641.
[29] J. M. Fletcher, A. L. Boyle, M. Bruning, G. J. Bartlett, T. L. Vincent, N. R. Zaccai, C. T. Armstrong, E. H. Bromley, P. J. Booth, R. L. Brady, A. R. Thomson, D. N. Woolfson, *ACS Synth. Biol.* **2012**, *1*, 240–250.
[30] M. A. Gebbie, W. Wei, A. M. Schrader, T. R. Cristiani, H. A. Dobbs, M. Idso, B. F. Chmelka, J. H. Waite, J. N. Israelachvili, *Nat. Chem.* **2017**, *9*, 473–479.
[31] R. L. Schoch, L. E. Kapinos, R. Y. Lim, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 16911–16916.
[32] J. Cai, L. A. Luck, D. Punihaole, J. D. Madura, S. A. Asher, *Chem. Sci.* **2016**, *7*, 4557–4562.
[33] J. F. Ross, G. C. Wildsmith, M. Johnson, D. L. Hurdiss, K. Hollingsworth, R. F. Thompson, M. Mosayebi, C. H. Trinh, E. Paci, A. R. Pearson, M. E. Webb, W. B. Turnbull, *J. Am. Chem. Soc.* **2019**, *141*, 5211–5219.
[34] K. Minamihata, K. Tsukamoto, M. Adachi, R. Shimizu, M. Mishina, R. Kuroki, T. Nagamune, *Chem. Commun.* **2020**, *56*, 3891–3894.
[35] A. S. Cristie-David, A. Sciore, S. Badieyan, J. D. Escheweiler, P. Koldewey, J. C. Bardwell, B. T. Ruotolo, E. N. G. Marsh, *Mol. Syst. Des. Eng.* **2017**, *2*, 140–148.
[36] J. P. Vrancken, J. Aupič, C. Addy, R. Jerala, J. R. Tame, A. R. Voet, *J. Struct. Biol. X* **2020**, *4*, 100027.

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202103025

Chemistry Europe
European Chemical
Societies Publishing

[37] E. V. Shtykova, L. A. Baratova, N. V. Fedorova, V. A. Radyukhin, A. L. Ksenofontov, V. V. Volkov, A. V. Shishkov, A. A. Dolgov, L. A. Shilova, O. V. Batishchev, C. M. Jeffries, D. I. Svergun, *PLoS One* **2013**, *8*, e82431.

[38] A. G. Kikhney, D. I. Svergun, *FEBS Lett.* **2015**, *589*, 2570–2577.

[39] A. K. Sharma, D. J. Friedman, M. R. Pollak, S. L. Alper, *FEBS J.* **2016**, *283*, 1846–1862.

[40] Z. Sun, Q. Liu, G. Qu, Y. Feng, M. T. Reetz, *Chem. Rev.* **2019**, *119*, 1626–1665.

[41] Y. H. Liu, Y. M. Zhang, H. J. Yu, Y. Liu, *Angew. Chem. Int. Ed.* **2021**, *60*, 3870–3880.

[42] L. M. Heitmann, A. B. Taylor, P. J. Hart, A. R. Urbach, *J. Am. Chem. Soc.* **2006**, *128*, 12574–12581.

[43] D. H. Juers, J. Ruffin, *J. Appl. Crystallogr.* **2014**, *47*, 2105–2108.

[44] S. Engilberge, M. L. Rennie, P. B. Crowley, *FEBS Lett.* **2019**, *593*, 2112–2117.