THE PROTEIN SOCIETY

Characterization of a highly flexible self-assembling protein system designed to form nanocages

Dustin P. Patterson,[1] Min Su,[2] Titus M. Franzmann,[3] Aaron Sciore,[1] Georgios Skiniotis,[2,4] and E. Neil G. Marsh[1,4]*

[1]Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109
[2]Life Sciences Institute, University of Michigan, Ann Arbor, Michigan 48109
[3]Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, Michigan 48109
[4]Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109

Received 22 October 2013; Accepted 4 December 2013
DOI: 10.1002/pro.2405
Published online 8 December 2013 proteinscience.org

Abstract: The design of proteins that self-assemble into well-defined, higher order structures is an important goal that has potential applications in synthetic biology, materials science, and medicine. We previously designed a two-component protein system, designated A-(+) and A-(−), in which self-assembly is mediated by complementary electrostatic interactions between two coiled-coil sequences appended to the C-terminus of a homotrimeric enzyme with C_3 symmetry. The coiled-coil sequences are attached through a short, flexible spacer sequence providing the system with a high degree of conformational flexibility. Thus, the primary constraint guiding which structures the system may assemble into is the symmetry of the protein building block. We have now characterized the properties of the self-assembling system as a whole using native gel electrophoresis and analytical ultracentrifugation (AUC) and the properties of individual assemblies using cryo-electron microscopy (EM). We show that upon mixing, A-(+) and A-(−) form only six different complexes in significant concentrations. The three predominant complexes have hydrodynamic properties consistent with the formation of heterodimeric, tetrahedral, and octahedral protein cages. Cryo-EM of size-fractionated material shows that A-(+) and A-(−) form spherical particles with diameters appropriate for tetrahedral or octahedral protein cages. The particles varied in diameter in an almost continuous manner suggesting that their structures are extremely flexible.

Keywords: coiled-coil; self-assembly; protein cages; symmetry; analytical ultracentrifugation; cryo-electron microscopy

Introduction

The assembly of individual protein subunits into higher order (quaternary) structures is a ubiquitous feature of biology and essential for the biological function of many proteins. The remarkable diversity of structural and functional properties exhibited by proteins suggests that assembling them into new quaternary structures should be a promising avenue for the construction of novel, responsive biomaterials.[1,2] Therefore, the design of proteins that self-assemble

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Army Research Office; Grant number: W911NF-11-1-0251 (to E.N.G.M.). Grant sponsor: Chemistry Biology Interface training, NIH; Grant number: T32 GM008597 (to D.P.P.). Grant sponsor: Project pr28ci.

*Correspondence to: Neil Marsh, Department of Chemistry, University of Michigan, Ann Arbor, MI 48109. E-mail: nmarsh@umich.edu

into well-defined, higher-order structures is an important goal that has potential applications in synthetic biology, materials science, and medicine.[3–10]

Multimeric protein assemblies may be broadly classified as either extended (filamentous), for example, collagen, fibrin, actin, and tubulin,[11–14] or closed, cage-like structures of which viral capsids are the best known examples.[15,16] These complex biological structures are often built from repeating units of only one or two proteins and are assembled in a highly symmetrical fashion. Other types of protein cages have also been identified leading to an increased appreciation for the essential and sophisticated roles that such structures play within the cell.[17] For example, ferritin is a well-studied protein microcompartment that stores iron in cells.[18] Enzymes also exploit cage-like assemblies; pyruvate dehydrogenase forms an icosahedral complex that spatially organizes its three component enzyme activities,[19] whereas the hollow barrel-like structure adopted by the proteosome complex[20] sequesters the protease active sites from the cellular milieu and prevents degradation of undamaged proteins.

More recently, several large bacterial microcompartments have been structurally characterized. Encapsulin from *Thermotoga maritima* forms a shell that encloses enzymes involved in oxidative stress response.[21] The carboxysome is found in many photosynthetic bacteria and enhances carbon dioxide fixation by encapsulating the key carbon-fixing enzymes rubisco and carbonic anhydrase.[22] Similarly, *Escherichia coli* cells growing on ethanolamine sequester the enzymes that metabolize ethanolamine in a microcompartment designed to prevent the diffusion of acetaldehyde (a toxic intermediate) into the cell.[23]

It is clear that symmetry plays an important role in specifying the quaternary structures of protein cages. Intrigued by this, we speculated that the application of symmetrical constraints *alone*, that is, without additional explicit constraints on protein orientation or structure, may be sufficient to direct the assembly of proteins into cage-like structures. If true, this could lead to a general strategy for assembling proteins into new supramolecular structures that is independent of the structural details of the protein. To test this idea, we recently constructed a self-assembling protein system[24] based on a trimeric protein, 2-keto-3-deoxy-6-phosphogluconate (KDPG)-aldolase from *T. maritima*,[25] which possesses C_3 symmetry. We engineered two aldolase constructs, A-(+) and A-(−), in which complementary *de novo* designed coiled-coil sequences were fused to their C-termini. The coiled-coil sequences, Helix-(+) and Helix-(−), were designed to form an antiparallel heterodimeric coiled-coil through complementary electrostatic interactions based on designs published previously[26–29] and principles that are now well established. Similarly, we created a system that, when the proteins were mixed together, could self-assemble into multimeric protein complexes, as illustrated in Figure 1.

The C_3 aldolase building block and the coiled-coil domains in A-(+) and A-(−) were connected by a

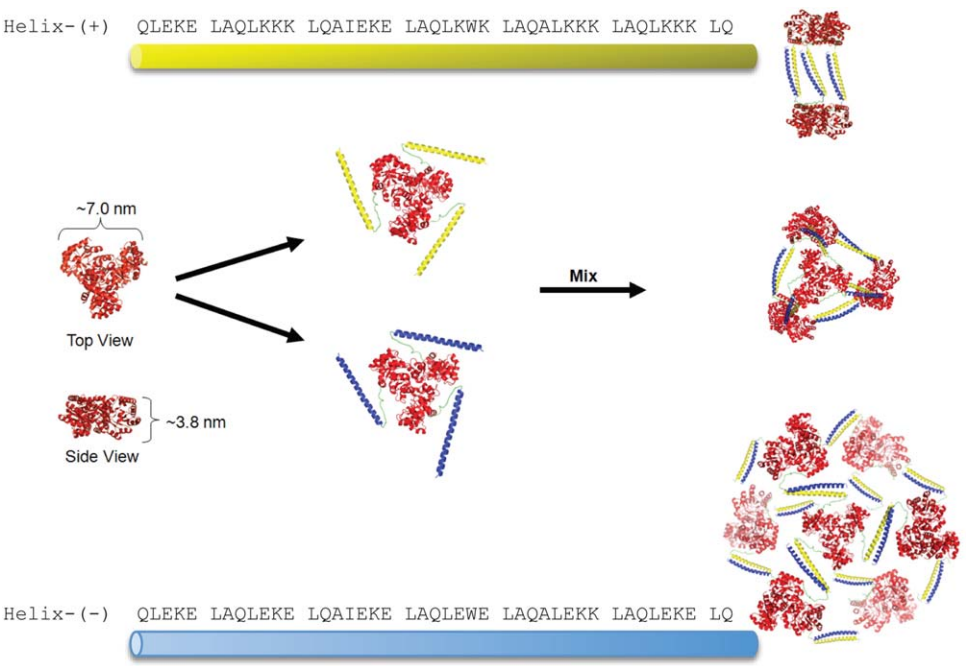

Figure 1. Overview of the design of the A-(+):A-(–) self-assembling system. The structure of KDPG aldolase is shown in red; the yellow and blue helices represent Helix-(+) and Helix-(−), respectively, and are tethered to aldolase through a flexible spacer (green). The three simplest and most symmetrical structures compatible with the design, a back-to-back dimer of A-(+) and A-(–), a tetrahedral cage and an octahedral cage are illustrated on the right, although many other structures are of course possible.

short, flexible spacer sequence, with the intention of allowing the system sufficient conformational freedom to assemble into structures compatible with the symmetry of the system. Thus, we did not intend with this design to produce a uniquely specified protein cage, but rather ask the question: given the ability to self-assemble, what structures do the proteins form? Our initial characterization of this system, using techniques that included size exclusion chromatography, light scattering, and atomic force microscopy, indicated that, despite the inherent flexibility of the assembly strategy, upon incubating together A-(+) and A-(−) assembled into a limited number of globular complexes. We were, however, unable to accurately determine the number of species formed, nor verify the structures of any of the complexes, which has been a common problem faced in the hierarchical assembly field.

In this report, we describe the detailed characterization of this self-assembling system using three complementary techniques: native polyacrylamide gel electrophoresis (native PAGE), sedimentation velocity analytical ultracentrifugation (AUC), and cryo-electron microscopy (cryo-EM). Native gel electrophoresis provides a qualitative analysis of the assemblies formed that is very sensitive to changes in oligomerization state, conformation, and charge. Centrifugation data allow one to determine the molecular weights and frictional ratios of the different species that are formed by the A-(+) and A-(−) complexes and quantify their relative abundances as they exist in solution. Cryo-EM images provide complementary information on the diameters of the protein cages formed and provide insight into their flexibility. The results herein not only provide new insights into these assemblies but also a foundation for future investigations of such systems.

Results

Our previous initial characterization of the A-(+):A-(–) self-assembling system established that grafting coiled-coil domains onto the parent aldolase trimer did not affect the activity or oligomerization state of the individual A-(+) and A-(−) enzymes and that the enzymes retained essentially the same specific activity when mixed and assembled into complexes.[24] Size exclusion chromatography indicated that when mixed, A-(+) and A-(−) assembled into a mixture of species with heterotetrameric and heterodimeric species predominating. (To simplify nomenclature, we use A-(+) and A-(–) to refer to the trimeric proteins, thus an A-(+):A-(–) heterodimer is understood to comprise a dimer formed between the two A-(+) and A-(–) trimers.)

Native polyacrylamide gel electrophoresis

As a sensitive tool to examine protein heterogeneity, we examined the electrophoretic properties of wild-type aldolase (A-WT), A-(−), A(+), and the A-(+) and A-(−) mixture under nondenaturing conditions (Fig. 2). To examine the influence of charge on the electrophoretic properties of the protein, samples were subjected to electrophoresis in the presence or absence of polyanionic charge shift molecule, Ponceau S. A-WT migrated as a single band with apparent M_r of ~146 kDa (as estimated by standard protein markers; the actual M_r for the crystallographically characterized trimer is 69 kDa). The addition of Ponceau S had little effect on the electrophoretic properties of A-WT.

The addition of the charged coil-coil domains to A-WT, unsurprisingly, substantially perturbed the

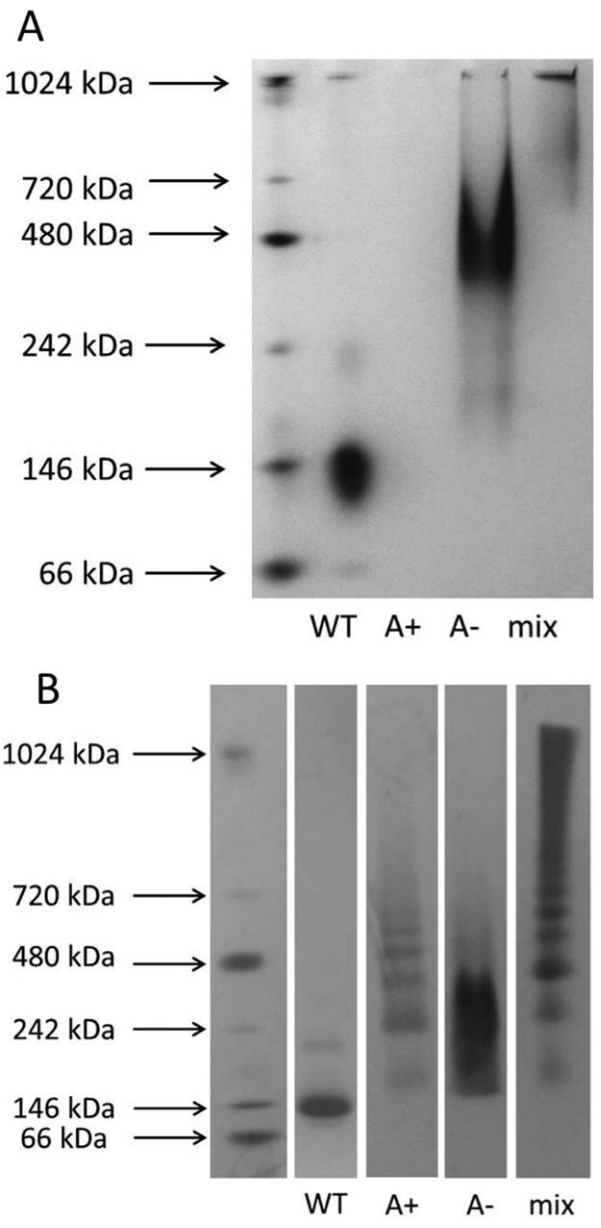

Figure 2. Native gel electrophoresis analysis of complexes formed by A-(−) and A-(+). A: Electrophoresis was performed in the absence of charge neutralizing agents and (B) in the presence of the charge-neutralizing dye Ponceau S. For discussion see the text.

electrophoretic properties of the protein (Fig. 2). In the absence of Ponceau S, A-(+) did not enter the gel, presumably due to the high positive charge associated to the helical linker. In contrast, A-(−) migrating as diffuse band with apparent M_r ~480 kDa (as estimated by standard protein markers). When A-(+) and A-(−) were subjected to electrophoresis in the presence of Ponceau S, A-(+) now migrated as approximately six distinct species with apparent M_r ranging from ~200 to ~600 kDa. Addition of Ponceau S to A-(−) resulted in the protein migrating with a lower apparent M_r of ~150–350 kDa, although the bands were still quite diffuse. These experiments are consistent with the conclusions of previous size-exclusion chromatography and AUC measurements that showed that for A-(−), the coiled-coil linkers cause some self-association. The change in the banding pattern observed in the presence of Ponceau S also indicates that some heterogeneity arises from differently charged species. This may be expected as the coiled-coil linkers contain multiple charged residues. Sodium dodecyl sulfate (SDS)-PAGE (Supporting Information Fig. S1) provided further evidence that the linkers induce electrophoretic heterogeneity into the proteins, as some smearing of the bands for A-(+) and A-(−) is evident even under denaturing conditions.

The complexes formed by a 1:1 mixture of A-(+) and A-(−) were analyzed by native PAGE (Fig. 2). In the absence of Ponceau S most of the sample barely entered the gel, running as a tight band of high apparent M_r, although some material is smeared below this band. Significantly, none of the bands observed for A-(−) were remained, supporting the heteroassociation of the two protein building blocks. In contrast, electrophoresis in the presence of Ponceau S resulted in the sample entering the gel and, although some smearing is evident, a number of distinct species are resolved that migrate with higher apparent M_r than either of the isolated A-(+) or A-(−) components under similar conditions.

Native PAGE provides good evidence that A-(+) and A-(−) associate to form a limited number of discrete species, as opposed to amorphous aggregates. However, the electrophoretic migration of the protein complexes is dependent on molecular weight, shape, and charge (as evidenced by the change in electrophoretic properties in the presence of Ponceau S). Therefore, we sought a more quantitative method that could distinguish between complexes of different shapes, while providing information on their molecular weight.

Analytical ultracentrifugation

We next turned to sedimentation velocity analytical ultracentrifugation, which is uniquely suited to provide quantitative information on macromolecular complexes in solutions.[30,31] A-(+), A-(−), and the assemblies formed by the mixture were subjected to centrifugation at various initial concentrations and rotor speeds and sedimentation traces recorded by monitoring absorbance at 280 nm, as described in the Materials and Methods section. We analyzed the sedimentation velocity experiments with the enhanced van-Holde Weischet method followed by two-dimensional sedimentation spectrum analysis (2-DSA), both of which are well suited to characterize heterogeneous mixtures.[32] The latter is a model-independent analysis approach to fit sedimentation boundaries and permits a reliable determination of the shape and molecular mass distribution of macromolecular mixtures. The sedimentation coefficient (s) is a measure of the velocity at which a particle travels under centrifugal force and is a function of both the molecular mass and shape of the particle. The frictional ratio (f/f_0) is a function of the shape of the particle and is obtained by comparing the experimentally measured translational frictional coefficient, f, with that calculated for a sphere with the same mass and density, f_0. A frictional ratio of close 1 is therefore generally interpreted to imply that the particle is roughly spherical particle, whereas $f/f_0 > 1$ indicates an elongated shape, with globular proteins typically being characterized by $f/f_0 < 5$. From the sedimentation coefficient and frictional ratio, the molecular mass of the sedimenting species can be calculated. The 2-DSA plots generated in these analyses are pseudo-3D plots (contour plots) showing the sedimentation coefficient and frictional ratio, the x and y axes, respectively, with the intensity of the spots indicating the relative concentration of species and probabilities (provided numerically as percentages in Table I).

Table I. *Hydrodynamic Properties of A-(+), A-(−) and the Complexes Formed by Their Assembly*

Species[a]	s_{20w} (S)	f/f_0	M_r (kDa)	Relative abundance (%)
A-(+)				
I	3.8	1.7	91	44
II	5.7	1.6	140	47
III	8.7	1.0	130	7
IV	22.2	1.0	570	2
A-(−)				
I	1.8	1.2	17	12
II	3.3	2.0	85	31
III	5.6	1.8	160	39
IV	8.3	1.0	120	18
A-(+):A-(−)				
I	3.7	1.2	49	10
II	6.8	1.8	220	13
III	9.2	1.1	180	25
IV	12.6	1.5	410	16
V	17.6	1.0	380	17
VI	26.9	1.0	720	10
VII	47.1	1.0	1700	7

[a] The numbering refers to the hydrodynamic species identified in the plots in Figure 2.

We first analyzed the sedimentation of wild-type KDPG aldolase. Sedimentation velocity analysis revealed that the wild-type protein sediments as a single species with 4.75 S ± 0.07 S [Supporting Information Fig. S2(A)]. The sedimentation coefficient was independent of the protein concentration within the tested concentration (15–75 μM). Using 2-DSA, we determined a molecular mass of 70.5 kD ± 3.9 kDa (calculated M_r = 69.9 kDa) and a frictional ratio of 1.19 ± 0.06, which is in agreement with wild-type KDPG aldolase forming compactly folded trimers in solution.

Next, we analyzed the individual A-(+) and A-(–) components. The introduction of the coiled-coil domains results in the formation of higher order assemblies and an increase in heterogeneity of both A-(+) and A-(–) compared with the wild-type protein, evident from both the primary sedimentation traces and the van Holde-Weischet plots [Supporting Information Fig. S2(A,B)]. The species distribution did not change with changes in the protein concentration (data not shown). For A-(+), 2-DSA of the sedimentation traces [(Fig. 3(A)] indicated that the protein exists mainly as two hydrodynamically distinct species characterized by s_{20w} = 3.8 S and f/f_0 = 1.7 (44%) and s_{20w} = 5.7 S and f/f_0 = 1.6 (47%). The 3.8 S species has a molecular mass of ~90 kDa, indicating that attachment of the coiled-coil to the protein does not impair trimerization. The decrease in the S and an increase in the f/f_0 values compared with A-WT are in agreement with an extended state of the trimer resulting from the attachment of the coiled-coil structure to the protein. Two other species with frictional ratios close to 1 and s_{20w} = 8.7 S and 22 S are also present, but their total abundance is less than 10% of the protein. The heterogeneity identified by the 2-DSA analysis of A-(+) is in accord with that observed by electrophoresis in the presence of Ponceau S.

2-DSA of the sedimentation traces for A-(−) [Fig. 3(B)] also identified two major species characterized by s_{20w} = 3.3 S and f_1/f_0 = 2.0 (31%) and s_{20w} = 5.6 S and f_1/f_0 = 1.8 (39%). As found for A-(+), attachment of the coiled-coil does not impair assembly of the trimeric building block and again the extension of the ~85 kDa trimer by the coiled-coil domain is reflected by a decrease in s and increase in f/f_0. We also present two less abundant species that are characterized by s_{20w} = 8.3 S and f_1/f_0 = 1.0 (18%) and s_{20w} = 1.8 S and f_1/f_0 = 1.2 (12%); this latter species is too small to represent A-(−) and most probably arises from a proteolytic fragment as shown in Figure 3 and summarized in Table I. Both A-(+) and A-(−) appear to form extended ~5.7 S particles, suggesting coiled-coiled mediated self-association of the trimeric building units.

We next analyzed the 1:1 mixture of A-(+) and A-(–) that had been allowed to equilibrate overnight

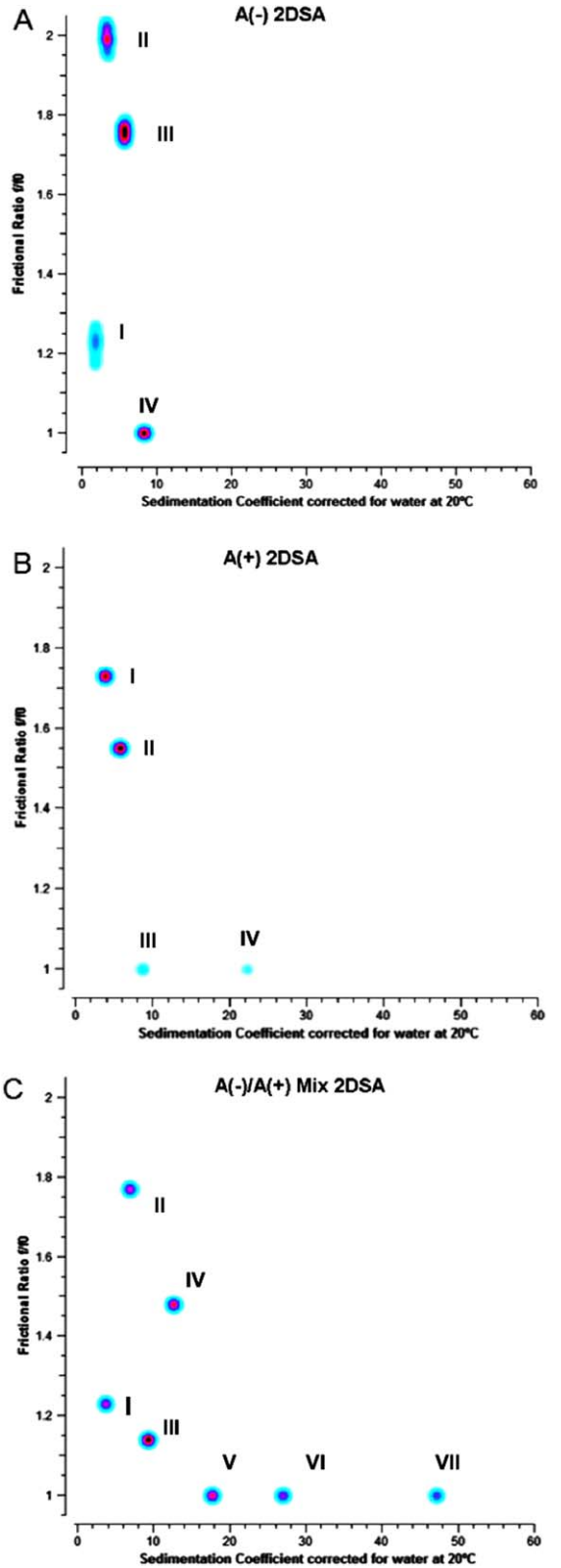

Figure 3. 2-D sedimentation spectral analyses of the complexes formed by A-(−) and A-(+). 2-DSA plots of (A) A-(−) alone, (B) A-(+) alone, and (C) 1:1 mixture of A-(−) and A-(+). The numbers by each species relate them to the hydrodynamic data in Table I.

at 4°C. In this case, the sedimentation profile revealed, as expected, species characterized by larger s_{20w} values, the majority of which have f/f_0 closer to 1 than either A-(+) or A-(–) [Fig. 3(C)]. Remarkably, given both the hydrodynamic heterogeneity of the individual building blocks and the conformational flexibility of the design, only seven species are present in the mixture in significant concentrations. Their hydrodynamic properties and calculated molecular weights are summarized in Table I. It is evident that there are no residual monomers of A-(+) or A-(–) present in the mixture, as species corresponding to their s_{20w} and f/f_0 are not present 2-DSA plot. Furthermore, it appears that species I, $f_1/f_0 = 3.7$, is likely a degradation product or impurity as its calculated M_r is well below that of either A-(+) or A-(–).

Of the other six species, the most abundant species **III** (25%) has the expected hydrodynamic properties to comprise a heterodimer of either A-(+) and A-(–). This is the simplest complex that could be formed and thus might be expected to be the most abundant; however, other larger species are also formed in significant quantities. In particular, two species, **V** and **VI**, with $s_{20w} = 18$ S and 27 S, respectively, are formed in significant concentrations that both possess $f_1/f_0 = 1$. Although an f_1/f_0 of 1 cannot be regarded as definitive that these complexes are spherically symmetric, their molecular weights and friction ratios are consistent with the formation of tetrahedral and octahedral assemblies. Two other relatively abundant species, **II** and **IV**, are present with frictional ratio characteristic of asymmetrical structures. It is unclear what structures they represent, but from their approximate M_r, they could be misassembled dimers and tetramers, respectively. Finally, a very large species, **VII**, $s_{20w} = 48$ S, is present as a minor component. Whereas this might represent an icosahedral complex formed from 10 copies each of A-(+) and A-(–), which would have M_r ~1800 kDa, its large size means that it could equally well comprise nonspecific high-molecular-weight protein aggregates.

Cryo-EM visualization

EM provides a complementary technique for characterizing the complexes formed by A-(+) and A-(–), allowing individual assemblies to be visualized and their structural properties to be investigated through collective analysis of multiple particles. To reduce the complexity of the sample, the ensemble of complexes formed by mixing A-(+) and A-(–) was first subjected to fractionation on a size exclusion column, as described in the Materials and Methods section. Fractions with elution volumes corresponding to those expected for octahedral and tetrahedral complexes of A-(+) and A-(–) were pooled, concentrated and prepared for microscopy.

Initial screening of the size-fractionated complexes formed by A-(+) and A-(–) using negative stain EM revealed that most of the complexes were "collapsed" on the carbon substrate into patches of different size and shape (Supporting Information Fig. S3) and therefore did not appear as cage-like structures. Cryo-EM proved less damaging to the protein complexes allowing them to be imaged in their hydrated state.[33] However, even under the more mild conditions used in the preparation of cryo-EM samples, we observed dissociation of the complexes into smaller fragments [Fig. 4(A)]. This fragmentation is most likely a consequence of the high surface tension generated by blotting the sample to a thin (~100 nm) layer of buffer before vitrification for cryo-EM. Chemically cross-linking the protein complexes using 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) before size-exclusion

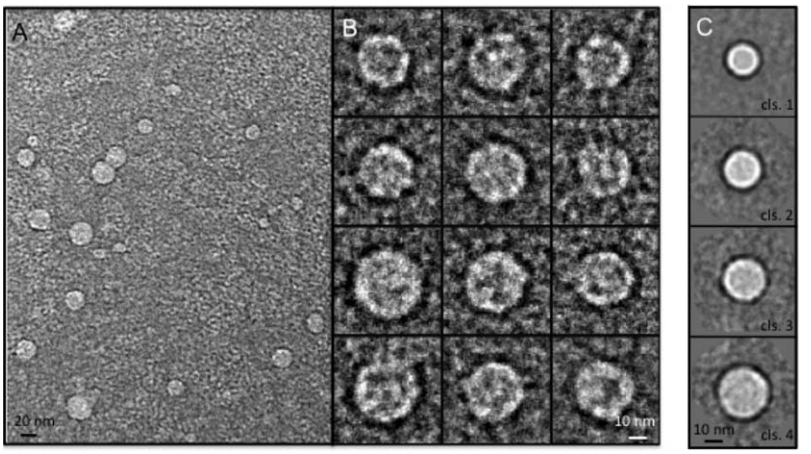

Figure 4. A: Cryo-EM micrograph showing variably sized protein cages embedded in a thin layer of vitrified buffer with fractured proteins dispersed in the background; (B) representative protein cage projections. Note that the contrast of images has been inverted, with protein density projections appearing brighter than their surroundings.

chromatography, as described in the Materials and Methods section partially alleviated this problem. Even so, fragmented material was still evident in the samples and this limited the number of particles that could be imaged and analyzed by single particle averaging methods.

The complexes formed by A-(+) and A-(–) generated roughly circular projections with diameters ranging from ~14 to ~25 nm when imaged by cryo-EM [Fig. 4(B)], consistent with atomic force microscopy image measurements obtained in our previous study.[24] Their sizes are in agreement with those expected for the formation of tetrahedral and octahedral cages by the proteins, which would be expected to predominate in the size-fractionated material. A cage diameter of ~17 nm would be consistent with the formation of tetrameric complexes of A-(+) and A-(–), while diameters from 22 to 26 nm likely represent different conformations of octameric complexes of A-(+) and A-(–). A few particles have an even larger average diameter of ~29 nm, which is incompatible with an octameric complex; these may represent larger oligomeric complexes of A-(+) and A-(–). We also note that we observe a direct correlation between the protein concentration used for vitrification and the number of circular projections observed per experiment, while no particles of this type were observed in control grids that did not contain the proteins of interest.

To analyze the architectural features of the protein complexes further, 376 particle images were interactively excised and subjected to reference-free classification and averaging using EMAN[34] software (Supporting Information Fig. S4). Most of the class averages demonstrate the spherical, closed nature of the protein complexes. Higher densities are seen along the outer edges of the circular projections that are consistent with the formation of hollow, cage-like structures. Furthermore, several two-dimensional (2D) averages of these particles revealed systematic surface features toward the center of the projections, suggesting they represent defined and repetitive architectures that are consistent with the formation of cages (Supporting Information Fig. S4). Interestingly, we observe a large variation in particle diameter. Thus, within the upper and lower bounds for the particles noted above, there appears to be an almost continual range of sizes, rather than discrete subgroups of particles, suggesting that the cages are highly flexible.

Discussion

The primary aim of these studies was to evaluate to what extent a simple and potentially highly generalizable design strategy, based primarily on symmetry constraints, could be used to direct the assembly of higher-order protein structures. The flexibility that we deliberately introduced between the aldolase domain and the coiled-coiled domains of A-(+) and A-(–), in principle allows a very large number of self-assembled complexes to form. A "back-to-back" dimer of A-(+) and A-(–) represents the lowest order complex that could form, whereas tetramers, octamers, and 20-mers (representing tetrahedral, octahedral, and icosahedral complexes, respectively) are the most highly symmetrical structures. However, as we have discussed previously,[24] *any* closed structure comprising an equal number of A-(+) and A-(–) trimers would satisfy the valency rules imposed by the need for each positive helix to associate with a negative helix. Extended one-dimensional and 2D networks of aldolase trimers and nonspecific aggregates could also potentially form.

It is important to note that although our experimental evidence points to the formation of tetrahedrally symmetrical **A**-(+)$_2$**A**-(–)$_2$ tetramers, these could not form, as depicted in Figure 1, if the **A**-(+) and **A**-(–) trimers remain associated as homotrimers because the intended coiled-coil interactions could not be satisfied. However, if the individual **A**-(+) and **A**-(–) subunits equilibrate between trimers, so that heterotrimers (comprising one **A**-(+) and two **A**-(–) subunits or *vice versa*) are formed, then a tetrahedron can readily be constructed. The "shuffling" of subunits between trimer, although hard to detect, may reasonably be expected to occur as they are noncovalently associated, and the monomeric form of the enzyme has been produced by a simple point mutation of the trimer interface.[25]

Using the complementary techniques of native gel electrophoresis, 2-DSA and cryo-EM, to analyze the protein assemblies formed by the A-(+) and A-(–) system, we are able to characterize both the properties of the system as a whole and at the single molecule level. Native gel electrophoresis provides a rapid, simple, and qualitative method to evaluate the heterogeneity of the component proteins and the assemblies they form. An important caveat is that the technique is highly sensitive to the charge state of the protein, as evidenced by the dramatic changes in electrophoretic mobility of A-(+), A-(–), and their complexes when the ion-pairing dye Ponceau S was introduced. This is unsurprising because the coiled-coil domains contain multiple charged residues that form the basis of the heterodimeric association. Thus, some of the heterogeneity observed by electrophoresis may be attributable to polyionic effects rather than changes in oligomerization state. Nevertheless, this technique demonstrates that introduction of the coiled-coil sequences to A-WT results in heterogeneity that is detected by other techniques. It also clearly shows that A-(+) and A-(–) associate and form a discrete set of complexes of higher apparent molecular weight than either of the component proteins and that very large species, characteristic of nonspecific aggregates, are not formed.

2-DSA is a powerful technique that importantly provides information on *all* the species present in the sample over a wide range of molecular weights and, furthermore, is minimally perturbing to the sample.[31,32] This was necessary for our investigation because we wished to obtain, as far as possible, an unbiased assessment of the various protein assemblies formed upon mixing A-(+) and A-(–). 2-DSA has allowed us to successfully characterize and quantify the complete mixture of complexes formed by A-(+) and A-(–) without the need to fractionate the material. Interestingly, both native gel electrophoresis and 2-DSA also identified unexpected heterogeneity in the individual protein building blocks, that was not identified by previous characterization using size exclusion chromatography coupled with multiangle laser light scattering detection,[24] illustrating the ability of these technique to resolve different species under differing conditions.

A priori, one might reasonably suppose that only the lowest order complex comprising the "back-to-back" dimer (Fig. 1) of A-(+) and A-(–) would form, or, on the other hand, given the minimal constraints of the design, that mixing the two proteins would result in an extremely complex and intractable mixture of species. Our results indicate that, in fact, neither of these two extremes prevails. Rather centrifugation identified only six discrete complexes of A-(+) and A-(–) to be present in significant concentrations. About half of the assemblies are accounted for by the three smallest species, as would be expected based simply on symmetry constraints. Symmetry thus appears to play a significant role in guiding which structures are formed.

Cryo-EM has allowed us to directly visualize some of the complexes formed by A-(+) and A-(–). The class averages demonstrate the protein complexes are spherically symmetric and appear hollow, consistent with their intended design. The diameters obtained from the direct cryo-EM imaging are similar to those predicted for the octahedral and tetrahedral protein cages and consistent with data obtained by 2-DSA, discussed above, and from previous characterization by light scattering techniques.[24] Although the variation in the diameter and instability of the particles imaged by Cryo-EM prevented us from obtaining reliable 3D reconstructions of the cages, it is nevertheless informative. Some of the variation may be explained by the samples containing mixtures of tetrameric, octameric and perhaps larger oligomers of A-(+) and A-(–). We consider that the variation in diameter is better explained by the open and intrinsically flexible structures of the protein cages. The flexibility between the aldolase and coiled-coil connecting domains potentially allows for a hinging motion so that the formed cage can expand and contract, as illustrated in Figure 5. Whereas this was not an intentional feature of the design, such flexibility might be useful for some applications.

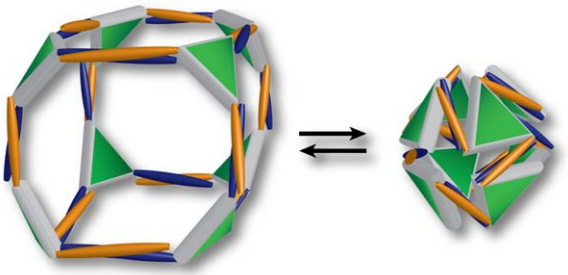

Figure 5. Cartoon illustrating how, for the octahedral protein cage, hinging and rotation of the trimeric protein building block (green) relative to the heterodimeric coil-coil linker domain (blue and gold) can result in expansion and contraction of the structure.

In conclusion, we have designed and characterized a self-assembling protein system with minimal constraints upon the mode of assembly and have demonstrated that a limited number of discrete structures are formed. With the design of the A-(+):A-(–) system, we did not set out to specify a single structure but rather determine to what extent symmetry constraints can be used to guide assembly. We believe there is considerable scope for optimizing this approach to design protein cages that adopt only a single, well-specified structure. For example, the flexibility between the protein building block and the coiled-coil domain could be restricted either by shortening the flexible spacer sequence or adopting a parallel coiled-coil orientation; this would place the protein building blocks closer together so that steric interactions may disfavor smaller complexes. More generally, using simple symmetry principles, it should be possible to assemble various combinations of proteins and coiled-coils of different quaternary structures into a wide variety of protein cages.

The strength of the coiled-coil interaction is another aspect of the design that could also be fine-tuned. We suspect that the coiled-coil interaction in the A-(+):A-(–) system may be too strong, and that some species identified by 2-DSA may represent kinetically trapped complexes. By reducing the number of heptad repeats or introducing destabilizing interactions at core positions, kinetically trapped intermediates could equilibrate to the most thermodynamically stable structure. Once optimized, this design strategy should, in principle, be generalizable.

Materials and Methods

Protein expression and purification

The construction of the genes encoding the A-(+) and A-(–) proteins has been described previously. The proteins were expressed and purified from recombinant *E. coli* strains by standard methods as described previously.[24]

Native polyacrylamide gel electrophoresis

Proteins samples were subjected to electrophoresis under native conditions using precast 3–12% Bis-Tris gels (pH 7.0; Life Technologies). Electrophoresis was conducted at 4°C and at constant voltage, 40 V, in Bis-Tris running buffer pH 7.0. In some samples, the charge-neutralizing agent Ponceau Red S was included in both the loading buffer and running buffer at a concentration of 0.02%.[35]

Analytical ultracentrifugation

Sedimentation velocity analysis was performed using a Beckman analytical ultracentrifuge equipped with an AN50TI rotor. Samples were dialyzed against phosphate-buffered saline, pH 8.0, containing 5% glycerol. Immediately before centrifugation, samples were filtered through 0.1 micron 'anatip' filters (Whatmann). The hydrodynamic behavior of A-(−) and A-(+) was analyzed at several different protein concentrations with initial absorptions of 0.2, 0.4, 0.6, 0.8, and 1.0 at 280 nm. Samples containing a 1:1 molar ratio of **A**-(−) and A-(+) were incubated overnight at 4°C before filtering and centrifugation. Samples were loaded into sector-shaped double channel centerpieces, and allowed to equilibrate at 25°C for 2 h in the nonspinning rotor before sedimentation. The individual A-(−) and A-(+) proteins were sedimented at 25,000 rpm, whereas the mixture was analyzed at both 14,000 and 35,000 rpm to assist in the analysis of multiple species formed. Absorbance data were collected at a wavelength of 280 nm. Sedimentation velocity data were initially analyzed using the enhanced van Holde-Weischet analysis module followed by 2-DSA using the finite element modeling module provided with the Ultrascan software (http://www.ultrascan.uthscsa.edu).[30] Sedimentation profiles were analyzed at a grid resolution of 62,500 using 25 grid repetitions. Confidence levels for statistics were derived from 2-DSA data refinement using a genetic algorithm followed by 30 Monte Carlo simulations. Calculations to analyze analytical ultracentrifugation data were performed on the UltraScan LIMS cluster at the Bioinformatics Core Facility at the University of Texas Health Science Center at San Antonio, the Lonestar cluster at the Texas Advanced Computing Center, and the National Supercomputer HLRB-II at the Leibnitz-Rechenzentrum, Munich, Germany.

EM sample preparation

To stabilize protein cages against fracturing during sample preparation, the subunits of A-(+) were chemically cross-linked between Lys and Glu or Asp residues using 1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) as a coupling agent. (The crystal structure of KDPG aldolase[25] shows that there is one salt bridge form between Lys and Glu at the trimer interface.) A-(+), ~ 1 mg/mL, was dialyzed against 20 mM 2-(N-morpholino)ethanesulfonic acid (MES) buffer, pH 6.8, 150 mM NaCl, 5% glycerol. 15 mL of dialyzed protein was diluted to ~50 mL in MES buffer and EDC (80 mg) dissolved in 1 mL of MES buffer then added to the diluted protein solution and mixed. Then Sulfo-N-hydroxysuccinimide (NHS) (160 mg), dissolved in 1 mL of MES buffer, was added to the protein-EDC solution, mixed, and allowed to incubate at room temperature. After 1 hour, the cross-linking reaction was quenched by addition of solid glycine (375 mg). SDS-PAGE analysis showed that ~ 50 % of the protein was cross-linked as the dimer form with minor amounts of higher order oligomers formed. The cross-linking agents were then removed from the sample by dialysis against the same MES buffer.

To prepare protein complexes, cross-linked A-(+) was concentrated to ~10 mg/mL in the MES buffer described above. A-(−) was dialyzed against 20 mM HEPES buffer, 150 mM NaCl, 5% glycerol, pH 7.5, and concentrated to ~ 10 mg/mL. Each protein component of 0.9 mL was mixed together and allowed to incubate at 4°C overnight. The sample was applied to a column of Sephacryl S400 (2 cm diameter × 90 cm), previously calibrated with appropriate molecular weight standards, and equilibrated with 50 mM Tris, 75 mM NaCl, pH 8.0. The column was eluted at 0.5 mL/min and 1 mL fractions collected. Fractions corresponding to protein complexes with approximate size range 800–400 kDa were pooled, concentrated to ~10 mg/mL, and immediately used to prepare samples for EM visualization.

EM imaging

Protein complex samples were first screened by negative stain EM. The concentrated samples were diluted to ~0.02 mg/mL and fixed on a grid using conventional negative staining procedures.[36] Imaging was performed at room temperature with a Morgagni 268(D) transmission electron microscope (FEI Company) equipped with a tungsten filament operated at an acceleration voltage of 100 kV and a mounted Orius SC200W CCD camera (Gatan Inc.).

For cryo-EM, 3 μL of concentrated sample solution was adsorbed on a glow-discharged Quantifoil grid (R2/2 200 mesh) and vitrified using a Vitrobot (FEI Mark IV). The sample was imaged on a Tecnai TF20 transmission electron microscope (FEI) equipped with a field emission electron gun operated at 200kV. Images were recorded at a magnification of 66,964× on a Gatan US4000 CCD camera, and binned (2 × 2 pixels) resulting in a pixel size of 4.48 Å on the specimen level. All the images were acquired using low-dose procedure to minimize radiation damages to the samples, with a defocus value in the range of 2–4 μm.

2D classifications

A total of 376 particle images representing protein cages were manually excised using EMAN Boxer.[34] The selected particles were subjected to reference-free

alignment and classification into 50 classes using the EMAN refine2d.py routine.

Acknowledgments

We gratefully acknowledge the use of supercomputing facilities at the Lonestar cluster at the Texas Advanced Computing Center and the National Supercomputer HLRB-II at the Leibnitz-Rechenzentrum, Munich, Germany.

References

1. Chockalingam K, Blenner M, Banta S (2007) Design and application of stimulus-responsive peptide systems. Prot Engin Des Sel 20:155–161.
2. Bromley EHC, Channon K, Moutevelis E, Woolfson DN (2008) Peptide and protein building blocks for synthetic biology: from programming biomolecules to self-organized biomolecular systems. ACS Chem Biol 3:38–50.
3. Yeates TO, Padilla JE (2002) Designing supramolecular protein assemblies. Curr Opin Struct Biol 12:464–470.
4. Ringler P, Schulz GE (2003) Self-assembly of proteins into designed networks. Science 302:106–109.
5. Sugimoto K, Kanamaru S, Iwasaki K, Arisaka F, Yamashita I (2006) Construction of a ball-and-spike protein supramolecule. Angew Chem Intl Ed 45:2725–2728.
6. Channon K, Bromley EHC, Woolfson DN (2008) Synthetic biology through biomolecular design and engineering. Curr Opin Struct Biol 18:491–498.
7. Young M, Willits D, Uchida M, Douglas T (2008) Plant viruses as biotemplates for materials and their use in nanotechnology. Ann Rev Phytopathol 46:361–384.
8. Papapostolou D, Howorka S (2009) Engineering and exploiting protein assemblies in synthetic biology. Mol Biosys 5:723–732.
9. King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, Andre I, Gonen T, Yeates TO, Baker D (2012) Computational design of self-assembling protein nanomaterials with atomic level accuracy. Science 336:1171–1174.
10. Fletcher JM, Harniman RL, Barnes FRH, Boyle AL, Collins A, Mantell J, Sharp TH, Antognozzi M, Booth PJ, Linden N, Miles MJ, Sessions RB, Verkade P, Woolfson DN (2013) Self-assembling cages from coiled-coil peptide modules. Science 340:595–599.
11. Mandelkow E, Mandelkow EM (1989) Microtubular structure and tubulin polymerization. Curr Opin Cell Biol 1:5–9.
12. Reisler E, Egelman EH (2007) Actin structure and function: what we still do not understand. J Biol Chem 282:36133–36137.
13. Wolberg AS (2007) Thrombin generation and fibrin clot structure. Blood Rev 21:131–142.
14. Shoulders MD, Raines RT (2009) Collagen structure and stability. Ann Rev Biochem 78:929–958.
15. Rossmann MG, Johnson JE (1989) Icosahedral RNA virus structure. Ann Rev Biochem 58:533–573.
16. Kuhn RJ, Rossmann MG (2005) Structure and assembly of icosahedral enveloped RNA viruses. Virus Struct Assem 64:263–284.
17. Kang S, Douglas T (2010) Some enzymes just need a space of their own. Science 327:42–43.
18. Theil EC (1987) Ferritin - structure, gene regulation, and cellular function in animals, plants, and microorganisms. Ann Rev Biochem 56:289–315.
19. Mattevi A, Obmolova G, Schulze E, Kalk KH, Westphal AH, Dekok A, Hol WGJ (1992) Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex. Science 255:1544–1550.
20. Kim DY, Kim KK (2003) Crystal structure of ClpX molecular chaperone from Helicobacter pylori. J Biol Chem 278:50664–50670.
21. Sutter M, Boehringer D, Gutmann S, Guenther S, Prangishvili D, Loessner MJ, Stetter KO, Weber-Ban E, Ban N (2008) Structural basis of enzyme encapsulation into a bacterial nanocompartment. Nat Struct Mol Biol 15:939–947.
22. Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC, Yeates TO (2008) Atomic level models of the bacterial carboxysome shell. Science 319:1083–1086.
23. Tanaka S, Sawaya MR, Yeates TO (2010) Structure and mechanisms of a protein-based organelle in *Escherichia coli*. Science 327:81–84.
24. Patterson DP, Desai AM, Holl MMB, Marsh ENG (2011) Evaluation of a symmetry-based strategy for assembling protein complexes. RSC Advances 1:1004–1012.
25. Griffiths JS, Wymer NJ, Njolito E, Niranjanakumari S, Fierke CA, Toone EJ (2002) Cloning, isolation and characterization of the *Thermotoga maritima* KDPG aldolase. Bioorg Med Chem 10:545–550.
26. Monera OD, Zhou NE, Kay CM, Hodges RS (1993) Comparison of antiparallel and parallel two-stranded alpha-helical coiled-coils. J Biol Chem 268:19218–19227.
27. Monera OD, Kay CM, Hodges RS (1994) Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in two-stranded alpha-helical coiled-coils. Biochemistry 33:3862–3871.
28. McClain DL, Woods HL, Oakley MG (2001) Design and characterization of a herterdimeric coiled-coil that forms exclusively with an antiparallel relative helix orientation. J Am Chem Soc 123:3151–3152.
29. Oakley MG, Hollenbeck JJ (2001) The design of antiparallel coiled coils. Curr Opin Struct Biol 11:450–457.
30. Demeler B, van Holde KE (2004) Sedimentation velocity analysis of highly heterogeneous systems. Anal Biochem 335:279–288.
31. Demeler B, Brookes E, Nagel-Stegert L (2009) Analysis of heterogeneity in molecular weight and shape by analytical ultracentrifugation using parallel distributed computing. Methods Enzymol 454:87–113.
32. Brookes E, Cao WM, Demeler B (2010) A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur Biophys J 39:405–414.
33. Frank J (2006) Three-dimensional electron microscopy of macromolecular assemblies. New York: Oxford University Press.
34. Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. J Struct Biol 128:82–97.
35. Drab T, Kracmerova J, Ticha I, Hanzlikova E, Ticha M, Ryslava H, Doubnerova V, Manaskova-Postlerova P, Liberda J (2011) Native red electrophoresis - A new method suitable for separation of native proteins. Electrophoresis 32:3597–3599.
36. Ohi M, Li Y, Cheng Y, Walz T (2004) Negative staining and image classification: powerful tools in modern electron microscopy. Biol Proced Online 6:23–34.