Protein-directed self-assembly of a fullerene crystal

Kook-Han Kim1, Dong-Kyun Ko2, Yong-Tae Kim1, Nam Hyeong Kim1, Jaydeep Paul3, Shao-Qing Zhang4,5, Christopher B. Murray6, Rudresh Acharya3, William F. DeGrado4, Yong Ho Kim1,7 & Gevorg Grigoryan8

Learning to engineer self-assembly would enable the precise organization of molecules by design to create matter with tailored properties. Here we demonstrate that proteins can direct the self-assembly of buckminsterfullerene (C60) into ordered superstructures. A previously engineered tetrameric helical bundle binds C60 in solution, rendering it water soluble. Two tetramers associate with one C60, promoting further organization revealed in a 1.67-Å crystal structure. Fullerene groups occupy periodic lattice sites, sandwiched between two Tyr residues from adjacent tetramers. Strikingly, the assembly exhibits high charge conductance, whereas both the protein-alone crystal and amorphous C60 are electrically insulating. The affinity of C60 for its crystal-binding site is estimated to be in the nanomolar range, with lattices of known protein crystals geometrically compatible with incorporating the motif. Taken together, these findings suggest a new means of organizing fullerene molecules into a rich variety of lattices to generate new properties by design.
Programmable self-assembly of molecular building blocks is a highly desirable way of achieving bottom-up control over novel functions and materials. Applications of molecular assemblies are well explored in the literature, ranging from optoelectronic devices to magnetic and photovoltaic devices to chemical and bioanalytical sensing, and medicine. However, it has been a daunting challenge to quantitatively describe and control the driving forces that govern self-assembly, particularly given the broad range of molecular building blocks one would like to organize. In this respect, nature’s self-assembling macromolecules hold considerable promise as standard chassis for encoding precise assembly. By learning to engineer the assembly of these molecules, myriad other molecular building blocks can be co-organized in desired ways through non-covalent or covalent attachment. The protein polymer is a particularly attractive candidate for a standard assembly chassis given its rich chemical diversity, of available assembly geometries, broad ability to engage other molecular moieties, and the possibility of engineered function. Considerable progress has been made in the area of engineering protein assemblies, using either computational or rational approaches, but the problem remains a grand challenge. A major difficulty lies in accounting for the enormous continuum of possible assembly geometries available to proteins to engineer a sequence that predictably prefers just one. General design principles, which provide predictive rules of assembly, are thus of enormous utility in limiting the geometric search space and enabling robust design.

In this work, we demonstrate the first ever high-resolution structure of co-assembly between a protein and buckminsterfullerene (C₆₀), which suggests a simple structural mode for protein–fullerene co-organization. Three separate crystal structures, resolved to 1.67, 1.76 and 2.35 Å, reveal a protein lattice with C₆₀ groups occupying periodic sites wedged between two helical segments, each donating a Tyr residue. A half site of the motif is estimated to have nM-scale affinity for C₆₀ such that binding of fullerene appears to direct the organization of protein units in the co-crystal. The assembly exhibits a nm-spaced helical arrangement of fullerenes along a crystallographic axis, endowing the crystal with electrical conductance properties. We closely investigate the interfacial geometry of the C₆₀-binding motif, finding it to be common among protein crystal lattices. C₆₀ and its derivatives have been previously reported to interact with several proteins, although a high-resolution structure of a protein–C₆₀ has been lacking. Still, prior evidence of interaction indicates that fullerenes and proteins are compatible as materials. This, together with the simple (and naturally recurrent) geometry of the C₆₀-binding motif we discover, suggests that it may be possible to use the structural principles emergent from our study to generate a variety of C₆₀-protein co-assemblies to further explore and exploit the properties of fullerenes.

Results

C₆₀-binding peptide organizes fullerenes. As a candidate for organizing C₆₀, we considered a peptide we had designed in a previous study (sequence in Fig. 1a), which forms an anti-parallel coiled-coil tetramer at μM concentrations (Protein Data Bank, PDB, entry 350R). Two key properties appear to make the peptide suitable for assembling C₆₀. First, the single aromatic residue in its sequence, tyrosine at position 9, is exposed and available for potential C₆₀ binding. Second, the peptide appears to have an exceptionally low barrier to crystallization, forming X-ray diffraction quality crystals within hours. Inter-tetramer contacts within the resulting lattice are not extensive (Supplementary Fig. 1), suggesting that the ease of crystallization may be due to an innately low penalty for freezing out conformational degrees of freedom.

Upon sonication, C₆₀ was readily solubilized in an aqueous solution of the peptide, hereafter referred to as COP (C₆₀-organizing peptide), but not buffer-only solutions. The resulting C₆₀–COP suspensions, stable after centrifugation for at least 3 months (not monitored thereafter), produced characteristic absorbance spectra revealing the presence of both protein and the fullerene (Fig. 1b). Size-exclusion chromatography (SEC) of COP alone was consistent with its tetrameric oligomerization state (black in Fig. 1c). On the other hand, upon the solubilization of C₆₀–COP, an additional peak appeared in the chromatogram, corresponding to a species of molecular mass approximately that of a dimer of tetramers (red in Fig. 1c; Supplementary Fig. 2). This suggests that the solubilization of C₆₀ occurs in a structurally specific manner with a change in oligomerization state of COP.

Despite COP’s propensity to crystallize, attempts to co-crystallize C₆₀ with COP were not met with success. The C₆₀–COP suspensions did form crystals, but these appeared to be devoid of fullerene. We reasoned that this could be due to an insufficient amount of solubilized C₆₀, such that not all binding sites on COP would be saturated and the protein-only species would selectively crystallize. Indeed, a rough estimate based on the C₆₀–COP ultraviolet spectrum (Fig. 1b) and molar absorptivity of C₆₀ at 340 nm taken from water/poly(vinylpyrrolidone) suspension, suggests one molecule of C₆₀ for ~24 COP tetramers (see Methods). To address this issue, we produced solutions of COP with C₆₀ pyrrolidine Tris-acid (C₆₀Sol; Supplementary Fig. 3), a more water-soluble analogue of C₆₀ (solubility 0.002–0.005 mg ml⁻¹ at pH 7.4). The SEC chromatogram of the COP–C₆₀Sol suspension again clearly shows two peaks—one corresponding to COP alone and another with apparent molecular weight corresponding to a dimer of COP tetramers (compare black and blue traces in Fig. 1c; also Supplementary Fig. 2). Further, absorbance at 340 nm (specific to the fullerene) clearly demonstrates that all of C₆₀Sol elutes in the second (octameric) peak, arguing for a specific structure-based association (top plot, blue trace Fig. 1c). These results are further supported by analytical ultracentrifugation (AUC) sedimentation equilibrium experiments at a range of concentrations, showing tetramer–octamer equilibrium for C₆₀Sol–COP solutions (with a dissociation constant of 118 μM), whereas a single-species monomer model is sufficient for COP alone (Supplementary Methods).

Crystals from the resulting suspension grew within 24 h in several conditions, and three separate structures of the C₆₀Sol–COP complex were resolved to 1.67, 1.76 and 2.35 Å, respectively (Fig. 1d–g; Supplementary Figs 5–6; Table 1). To our knowledge, these represent first high-resolution structures of a protein–fullerene complex. As in the protein-only structure, COP forms a canonical tetrameric anti-parallel coiled-coil. Each tetramer presents four tyrosine residues (one per monomer) in exterior positions of the coiled-coil heptad, and each of these engages a C₆₀ moiety. One C₆₀ is wedged between two Tyr residues donated by adjacent tetramers, such that two tetramers are needed to coordinate one C₆₀ (Fig. 1f). This arrangement fits well with the apparent octameric peak in the SEC chromatogram and AUC profiles of C₆₀Sol–COP (Fig. 1c; Supplementary Fig. 4), suggesting that higher-order organization begins already in solution before crystallization. The water-solubilizing Tris-acid side chain of C₆₀Sol is not visible in the electron density map. The group likely points into the solvated inner channel of the crystal and is highly mobile as the C₆₀ core rapidly rotates around its centre. This is consistent with the intended role of the side chain, to increasing the solubility of the fullerene, whereas the C₆₀ core appears responsible for the specific packing arrangement.
The interaction between COP and the fullerene group involves non-polar contacts (Supplementary Fig. 7a), with C₆₀ fitting perfectly into a symmetric hydrophobic cavity created by helices of two adjacent COP tetramers. The dominant contact appears to be the π–π aromatic stacking between C₆₀ and Tyr9, donated by aromatic stacking between C₆₀ and Tyr residues (other aromatic side chains also likely admissible), and further self-assembles into a co-crystalline array with fullerene. Interestingly, the conformation of COP itself is largely unchanged (Fig. 2c,d). This, together with the fact that neither of the crystals simplify the analysis, we concentrated on one half site of the binding interfaces and assembly geometries, arguing that fullerene may have a strong preference for the observed coordination geometry.

Helix–helix motif expected to bind fullerene tightly. We next ask whether the helix(Tyr)–C₆₀–helix(Tyr)-binding mode could serve as a general co-organizer of proteins and fullerene. A necessary (but not sufficient) condition for this is that the motif would need to provide sufficient binding energy to drive assembly into a variety of desired arrangement. So we sought to quantify the affinity of C₆₀ for the identified binding site. Direct equilibrium measurement of C₆₀–protein association is complicated by the exceedingly low solubility of C₆₀ in aqueous solution. Even the C₆₀sol derivative has limited water solubility, hampering binding studies. We thus turned to explicit-solvent molecular dynamics simulations to characterize the strength of COP–C₆₀ association. The observed binding mode is a ternary interaction between two COP tetramers and one fullerene. To simplify the analysis, we concentrated on one half site of the symmetric binding pocket, looking to characterize the affinity of one C₆₀ for one COP. Using the crystal structure as the starting bound configuration, thermodynamics of binding was characterized using a modification of the double-decoupling method in conjunction with the free energy perturbation (FEP) approach (see Methods; Supplementary Fig. 8). A total simulation time of 336 ns permitted accurate monitoring of

Figure 1 | Protein/C₆₀ super-assembly. (a) COP, a stable tetramer in isolation, interacts with C₆₀ moieties by means of a surface-binding site that includes Tyr residues (other aromatic side chains also likely admissible), and further self-assembles into a co-crystalline array with fullerene. (b) Ultraviolet absorption spectra of a C₆₀/COP suspension and COP alone demonstrate that primitive fullerene (green) dissolves in the aqueous phase in the presence of protein. (c) SEC traces of COP alone or in association with C₆₀ or C₆₀sol. Top and bottom plots show absorbances at 340 and 220 nm, respectively. The lower-resolution peaks arising due to the addition of C₆₀ or C₆₀sol are consistent with the molecular weight of a COP octamer (for example, dimer of tetramers; Supplementary Fig. 10). (d) Each COP tetramer in the C₆₀sol–COP crystal is associated with four fullerenes (one per chain), each fullerene being wedged between two adjacent COP tetramers, for an overall stoichiometry of two fullerenes for one COP tetramer. (e) Omit map (2Fo – Fe, contoured at 1.2σ) showing electron density of the C₆₀ group (orange sticks) sandwiched via π–π stacking between Tyr residues from adjacent COPs. (f) Residues involved in C₆₀ coordination are shown with sticks and labeled. (g) Surface representation of the C₆₀ coordination site, colored by relative in vacuo electrostatic potential (red to blue corresponds to negative-to-positive relative potentials).
convergence, with the standard-state free energy of C60–COP
binding estimated at $-9.8 \pm 0.3$ kcal mol$^{-1}$. This corre-
sponds to a dissociation constant in the range of 40–100 nM,
confirming the suspicion that C60 binding provides substantial
energy to drive the assembly of COP units. In fact, the true
energetic contribution of C60 is likely even larger as some
positive cooperativity between the two motif half sites would
be expected due to direct (albeit not extensive) favourable
protein–protein interactions. Interestingly, we find that the $\pi$–$\pi$
stacking between C60 and Tyr9 is not sufficient to explain the
strong interaction, as the affinities of C60 for an isolated
Tyr residue (acetylated and methylamidated at the N- and
C termini, respectively) or a Tyr side-chain analogue
($p$-methylphenol) are estimated to be in the mM range
(Supplementary Fig. 8b,c). Thus, additional aliphatic contacts
in the binding pocket are essential for the collective binding
mode and the high affinity.

Fullerene-binding motif composed of designable elements.
Another necessary property of a generic protein–fullerene
organizing motif is that it must be ‘designable’ in the context of
a multitude of protein lattices—that is, the required geometry
should be easily achievable with natural amino-acid sequences.
Using the structural search engine MASTER$^{33}$, we found that
all of the interfaces involved in the motif are indeed highly
abundant in nature (and are thus necessarily designable),
with emergent sequence preferences in agreement with the
corresponding region of COP (Methods; Supplementary Fig. 7b).
Further, even the entire binding motif, composed of four disjoint
helical segments that account for all contacts with C60 in the
supercrystal, has precedence in PDB lattices. In fact, within a
homology/redundancy-pruned subset of the PDB (13,400
entries), we found 180 unique instances of matching geometries
(below 1.9 Å backbone r.m.s.d., computed over 112 atoms) within
21 unique lattices (Supplementary Fig. 7c). That is, $\sim 0.15\%$ of

### Table 1 | Statistics on data collection and refinement of C60Sol–COP complex.

| Data set:$^*$ | C60Sol–COP-1 | C60Sol–COP-2 | C60Sol–COP-3 |
|--------------|--------------|--------------|--------------|
| Crystallization conditions | 17 mM LiSO$_4$ 85 mM Tris-HCl 25.5% PEG 4,000 pH 8.5 | 0.1 M ADA 1M NH$_4$H$_2$PO$_4$ pH 6.5 | 0.2 M CH$_3$CO$_2$NH$_2$ 0.1 M HOC(COO)Na$_3$ (CH$_3$COO)Na$_2$-2H$_2$O 30% PEG 4,000 pH 5.6 |
| Beam line | 24ID3,NE-CAT | Home source$^+$ | PLS,BL-7A |
| Wavelength (Å) | 0.97919 | 1.54178 | 1.00000 |
| Space group | $P_6_2$ | $P_6_2$ | $P_6_2$ |
| Cell dimensions | $a$, $b$, $c$ (Å) 41.71, 41.71, 66.81 | 41.71, 41.71, 67.23 | 42.15, 42.15, 66.79 |
| $\alpha$, $\beta$, $\gamma$ (°) | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å)$^\ddagger$ | 50.0–2.35 (2.48–2.35) | 50.0–1.76 (1.86–1.76) | 50–1.67 (1.73–1.67) |
| $R_{merge}$ | 0.135 (0.709) | 0.050 (0.309) | 0.066 (0.188) |
| $I/\sigma(I)$ | 15.3 (4.2) | 15.6 (2.4) | 30.9 (10.9) |
| Completeness (%) | 100 (100) | 94.6 (69.1) | 95.9 (99.7) |
| Multiplicity | 13.7 (14.1) | 5.1 (1.8) | 16.3 (11.3) |
| Total/unique reflections | 38460/2799 | 31977/6255 | 336749/7862 |
| Resolution statistics | | | |
| Resolution (Å) | 36.13–2.35 | 24.61–1.76 | 15.0–1.67 |
| Number of reflections | 5421 | 6235 | 7489 |
| Twin fraction ($\xi$), estimated and refined | 0.478 (S(H) plot), 0.447 (Britton plot), 0.5 (refined) |
| $R_{work}$/$R_{free}$ | 0.2019/0.2338 | 0.2027/0.2391 | 0.2181/0.2444 |
| Number of atoms | 512 | 526 | 535 |
| Proteins | 446 | 446 | 446 |
| Ligand/ion | 60 | 60 | 60 |
| Water | 6 | 20 | 29 |
| B-factors ($\AA^2$) | | | |
| Average B-factors ($\AA^2$) | 30.8 | 27.0 | 22.5 |
| Proteins | 30.6 | 26.3 | 21.3 |
| Ligand/ion | 31.8 | 29.5 | 27.5 |
| Waters | 30.6 | 34.4 | 30.5 |
| Bond lengths (Å) | 0.007 | 0.007 | 0.008 |
| Bond angles (°) | 0.808 | 0.828 | 0.893 |
| Ramachandran regions (%) | | | |
| Most favourable | 100 | 100 | 100 |
| Additional allowed | 0.0 | 0.0 | 0.0 |
| Generously allowed | 0.0 | 0.0 | 0.0 |

$^*$r.m.s.d.'s, root mean squared deviations.

$^*$Structures of the same complex determined under different condition in different resolutions.

$^+$Home source, CCMB (Center for Cellular and Molecular Biology), Hyderabad, India.

$^\ddagger$Highest resolution shell is shown in parenthesis.
proteins in the PDB already exhibit backbone geometries similar to the one housing a bound C60 in the co-crystal, suggesting that it may be possible to engineer a variety of fullerene/protein co-assemblies by perturbing sequences of existing parent proteins. Supplementary Fig. 7c shows examples of such putative co-assemblies, where C60 is computationally placed into existing lattices matching the binding motif, giving diverse C60-to-C60 distances and lattice arrangements. Of course, the design of such assemblies will involve not only the placement of a C60-binding motif, but also any appropriate accommodating changes to surrounding amino acids. Further, there is no guarantee that the crystal form will not change upon these perturbations. Still, that our identified motif appears ‘canonical’ in terms of its constituent protein–protein interfaces is encouraging for future design applications.

**Fullerene–protein crystal has emergent electronic properties.** The honeycomb structure of C60Sol–COP is intriguing from the perspective of its electronic properties. Within the helical arrangement of fullerenes, inter-C60 distances appear sufficiently close for potential long-range electronic transfer, especially given the organized nature of the surroundings (Fig. 3a–c). For this reason, we sought to characterize the electrical conductance of the co-crystal. Current–voltage (I–V) characteristic of disordered C60 films showed high electrical resistance of 2.24 × 1011 Ω (Fig. 3d and Supplementary Fig. 9). In addition, COP-alone crystals or crystal buffer similarly showed high resistance, with only 5–10 pA of currents measured with up to 20 V of voltage sweep. On the other hand, C60Sol–COP supercrystals (of similar dimension as protein-alone crystals) exhibited high electrical conductance (1.40 × 10−7 S, corresponding to resistance of 7.14 × 106 Ω) with at least four orders of magnitude higher currents than in any of the controls (Fig. 3d). We speculate that the periodic arrangement of fullerene groups in the co-crystal may facilitate electron wave delocalization over the assembly. This would promote coherent electron transport through the structure with the carrier mobility expected to be several orders of magnitude higher than in disordered systems characterized by hopping transport.36. Inter-fullerene nearest-neighbour distances in the C60Sol–COP supercrystal alternate between 1.2 and 1.7 nm (Fig. 3c). For comparison, strong electron wave delocalization was previously observed when the nearest-neighbour distance approached ~1.5 nm in one-dimensional fullerene wires.37. An alternative explanation of the observed conductive property is that the hexameric channels in the co-crystal may contain unattached/disordered fullerene moieties that are free to diffuse in the channel and can shuttle electrons between ordered in-lattice fullerenes. In either case, as shown in Fig. 3d (yellow dots), destruction of crystalline order (by placement in vacuum) results in very high electrical resistance. In fact, the current measured here is even lower than that of the disordered C60 film. This indicates that the high conductivity of the C60Sol–COP supercrystal is not a trivial property of crystal dimension and/or molecular composition, but rather originates from specific electronic coupling/delocalization in the assembly.

**Discussion**

The aim of programmable self-assembly is to anticipate and harness unique collective properties that arise from precisely organized molecular building blocks. To this end, achieving atomic-level precision is crucial. This work demonstrates the first atomic resolution structures of a fullerene–protein assembly, establishing the feasibility of creating such objects, and further suggests a possible design principle for engineering such
assemblies in general. How robust the discovered C60-binding motif is towards designing novel assemblies will need to be tested through a number of future design studies. However, the straightforward manner in which self-organization arose in our case, the simplicity of the C60-organizing motif in the lattice, together with its high affinity and the ubiquity of associated interfaces in natural protein lattices, are certainly promising with respect to the general applicability of the design principle. Our work also demonstrates the potential utility of exploring C60/protein co-organization, as derived supercrystals already showed synergistic charge conductance properties. Taken together, these results point to an exciting direction of inquiry towards generating protein–fullerene assemblies for the study and design of novel properties.

Methods

Peptide synthesis and purification. Peptides were synthesized by CEM Discover microwave synthesizer using Fmoc chemistry at 100-μmol scales. The Fmoc protecting group was removed by piperidine/dimethylformamide solution (20/80 v/v) at each coupling step reactants were added with the amino acid:HBTU:DIEA:resin ratio of 5:4.9:1:01. Products were cleaved from the H-Rink Amide-ChemMatrix (PCAS, 0.53 mmol g−1 loading) in a cleavage cocktail solution (trifluoroacetic acid (TFA)/triisopropylsilane/deionized water, 95/2.5/2.5 v/v) for 2 h and the remaining solution was vapourized with N2 gas. Peptide was precipitated with cold diethyl ether (90% acetonitrile, 9.9% H2O and 0.1% TFA). Molecular mass of the peptide was confirmed by matrix-assisted laser desorption/ionization-time of flight mass spectrometer in 1 cm Hellma Quartz SUPRASIL (QS) cells. The COP and C60/COP were prepared in a buffer of 20 mM sodium phosphate, 100 mM NaCl and pH 7.5. Ultraviolet–visible spectra of C60/COP and COP were used to roughly estimate the concentration of solubilized fullerene by absorbance at 340 nm (the molar absorptivity of 49,000 M−1 cm−1 was used for C60, ref. 28). The resulting molar concentration of solubilized C60 in the COP/COP solution was 6.22 μM (compared with COP at 585 μM in the same solution).

Size-exclusion chromatography. Size-exclusive gel filtration elution profiles were obtained using a Superdex 75 10/300 GL column with a GE Healthcare fast performance liquid chromatography (FPLC) system (Amersham Pharmacia Biosystems). Peptides (at 200 μM) were prepared in a buffer of 20 mM sodium phosphate, 100 mM NaCl and pH 7.5 at room temperature. 200 μl of each sample was loaded and eluted with the same buffer. The column was equilibrated in 20 mM sodium phosphate, 100 mM NaCl and pH 7.5 with a mobile phase flow rate of 0.5 ml min−1, and absorbance at 220, 280 and 340 nm was recorded. Calibration curves were obtained using the molecular-weight standard kit, MWG70 6,500–66,000 (Supplementary Fig. 10).

Analytical ultracentrifugation. Oligomerization states of COP and C60Sol–COP were determined by equilibrium sedimentation performed at 25 °C using a Beckman XL-1 analytical ultracentrifuge. Both solutions were prepared in a buffer of 25 mM Tris pH 8.0. Equilibrium radial concentration gradients at four different rotor speeds (25, 30, 35 and 40 K rpm) were acquired as absorbance scans at 340 nm for C60Sol with COP and 280 nm for COP peptide alone. Data were globally fit to single-species or two-species models of equilibrium sedimentation by a nonlinear least-squares method using IGOR Pro (Wavemetrics), and the best-fitting model was accepted. Supplementary Figure 4 shows sedimentation equilibrium profiles of C60Sol–COP along with corresponding species distribution plots consistent with a tetramer–octamer equilibrium, whereas COP alone appears as a tight tetramer. This is consistent with results from SEC, shown in Fig. 1e and Supplementary Fig. 2.

Preparation of peptide/fullerene solutions. Samples were prepared with 8 mg ml−1 protein solution (COP) in 25 mM Tris pH 8.0 buffer solution and 1 mg C60 or C60 pyrrolidine Tris-acid (Aldrich). Fullerene powder was mixed with pre-made 0.2 ml of 8 mg ml−1 protein solution in 25 mM Tris pH 8.0. The sample was then tip-sonicated (Qsonica, Q125, 1/8th inch tip) on an ice bath for 5 min to be saturated of fullerene. Ice-bath cooling was to prevent excessive sample heating and destabilization of protein structure. The sonicated samples were warmed up to room temperature and centrifuged at 14,500g for 10 min (Eppendorf, Centrifuge 5430R).

Crystallization, data collection and processing. The first X-ray diffraction quality crystal (C60Sol–COP-1) was obtained by the hanging-drop vapour diffusion technique at 291 K, over a period of 15 days in a 2 μl drop consisting of 1:1 v/v mixture of 1 mM mgml−1 protein solution in 20 mM sodium phosphate/100 mM NaCl pH 7.5 buffer and a reservoir solution of 17 mM lithium sulfate monohydrate. 85 mM Tris-hydrochloride sodium pH 8.5, 2.5% polyethylene glycol (PEG) 4,000, 25% v/v glycerol (Hampton Research sparse matrix). The crystal was flash-frozen, and diffraction data were collected at the 24-ID-E NE-CAT beamline at the
Argonne National Laboratory. Data sets were indexed and integrated with MOSFLM39,40, and scaled using SCALA3 (Collaborative Computational Project, Number 4, 1994)41. Diffraction data were recorded to a maximum resolution of 2.35 Å (Table 1).

Subsequent crystallization attempts were performed with higher concentrations of the COP:C60:C60D solvent mixture, using commercially available sparse-matrix screens from Hampton Research and the hanging-drop vapour-diffusion method at 295 K. Diffraction-quality crystals of (COP)2(C60·C60D·C60)−3 were obtained by mixing equal volumes of the COP−C60−C60D mixture at 8 mg ml−1 in 25 mM Tris pH 8.0 and reservoir solution consisting of 0.2 M ammonium acetate, 0.1 M sodium citrate tribasic dihydrate pH 5.6, 30% w/v polyethylene glycol 4,000. Microcrystals grew in 4 h, with larger oval-shaped crystals appearing in several days (Supplementary Fig. 5). Crystals were cryoprotected using reservoir solution supplemented with an additional 30% (v/v) glycerol and were flash-cooled in liquid nitrogen. Diffraction data, extending to 1.67 Å resolution, were collected at 100 K on beamline 7A equipped with an ADSC Quantum 270 CCD detector at Pohang Accelerator Laboratory (PAL, Pohang, Korea). The COP−C60·C60D complex crystal belonged to space group C2/c, with axial ratios a = b = 42.1 Å, c = 66.7 Å, α = β = 90.0 and γ = 120.0°. Data were processed and scaled using the programs DENZO and SCALEPACK from the HKL-2000 program suite42. The Matthews coefficient43 for COP−C60 or COP−C60D was 2.54 Å3 Da−1 and the estimated solvent content was 51.3%; there were two COP molecules and one COP−C60 in an asymmetric unit.

In addition to the above, diffraction-quality crystals were also obtained in three other conditions (1.5 M ammonium sulfate, 0.1 M Tris pH 8.5, 12% v/v glycerol; 0.1 M HEPES–Na pH 7.5, 0.8 M potassium sodium tartrate tetrahydrate; and 0.1 M N-(2-acetamido)-1-imino-4-hydantoin acid (AIAA) buffer pH 6.5, 1 M ammonium phosphate dibasic), in each case yielding identical unit cell and space group, thus showing the same assembly geometry. Crystals grown under the latter condition diffracted to 1.76 Å at a home source (COP−C60−C60D). Structure solution and refinement. For all the data sets, structure determination was carried out by molecular replacement using the programme PHASER44. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement followed by molecular replacement using the programme PHASER44. The solution model obtained from coordinates of previously solved crystal structure 3S0R as the search probe. The solution model was subjected for rigid body refinement follow...
free-energy change was corrected by $RT \ln C_i^{(2nRT/kT)}$\[3\], where $C_i$ is the standard-state concentration and $k$ is the force constant of the $C_{60}$ restraint in the decoupled state (that is, 10 kкал mol\(^{-1}\) Å\(^{-1}\)).

Transformation 2 is similar to transformation 1, but without protein. In the first step, $C_{60}$ is decoupled from solvent as $C_{60D}$ is coupled. However, the intermediate step again renders the decoupled $C_{60D}$. As with transformation 1, the influence of $C_{60D}$ cancels between the two end states, with the total free-energy difference corresponding to that of decoupling $C_{60}$ from solvent. However, the intermediate step again renders the $C_{60D}$ from the starting state. This intermediate step thus cancels between the end states in the free-energy change of transformation 2.

Measurement of electrical conductance. Current versus voltage curves were obtained using the variable temperature microprobe system from MMR technologies coupled with HP 4145B semiconductor parameter analyser. The samples were deposited on a degenerately doped silicon substrate with 200 nm thermal oxide, which was photolithographically pre-patterned with Au/Cr (45 nm/5 nm) electrodes. The channel length and width were 10 and 6,000 μm, respectively.

References

1. Talapin, D. V. & Murray, C. B. PhSe crystals for n- and p-channel thin film field-effect transistors. Science 316, 86–89 (2007).
2. Kumar, R. J., MacDonald, J. M., Singh, T. R., Washington, J. L. & Holmes, A. B. Hierarchical self-assembly of semiconductor functionalized peptide alpha-helices and optoelectronic properties. J. Am. Chem. Soc. 133, 8564–8573 (2011).
3. Erb, R. M., Son, H. S., Samanta, B., Rotello, V. M. & Yellen, B. B. Magnetic assembly of colloidal superstructures with multipole symmetry. Nature 457, 999–1002 (2009).
4. Koo, J. et al. Acentric 2-D ensembles of D-br-A electron-transfer chromophores via vectorial orientation within amphiphilic n-helix bundle peptides for photovoltaic device applications. Langmuir 28, 3227–3232 (2012).
5. Li, Y. et al. Self-assembly of mesosкопically ordered chiral polydiacetylene/silica nanocomposites. Nature 410, 913–917 (2001).
6. Mathias, P. L., Penney, M. G. & Hegen, M. B. Designing materials to direct stem-cell fate. Nature 462, 433–441 (2009).
7. Zhang, J., Zheng, F. & Grigoryan, G. Design and designability of protein-based assemblies. Curr. Opin. Struct. Biol. 27C, 79–86 (2014).
8. King, N. P. & Lai, Y.-T. Practical approaches to designing novel protein assemblies. Curr. Opin. Struct. Biol. 23, 632–638 (2013).
9. King, N. P. et al. Computational design of self-assembling protein nanomaterials with atomic level accuracy. Science 336, 1171–1174 (2012).
10. King, N. P. et al. Accurate design of co-assembling multi-component protein nanomaterials. Nature 510, 103–108 (2014).
11. Grigoryan, G. et al. Computational design of virus-like protein assemblies on carbon nanotube surfaces. Science 332, 1071–1076 (2011).
12. Gonen, S., DiMaio, F., Gonen, T. & Baker, D. Design of ordered two- and three-dimensional domains mediated by noncovalent protein-protein interfaces. Science 348, 1365–1368 (2015).
13. Lanci, C. J. et al. Computational design of a protein crystal. Proc. Natl Acad. Sci. USA 109, 7304–7309 (2012).
14. Kaltofen, S. et al. Computational de novo design of a self-assembling peptide with predefined structure. J. Mol. Biol. 427, 550–562 (2015).
15. Egelman, E. H. et al. Structural plasticity of helical nanotubes based on coiled-coil assemblies. Structure 23, 280–289 (2015).
16. Xu, C. et al. Rational design of helical nanotubes from self-assembly of coiled-coil locked washers. J. Am. Chem. Soc. 135, 15365–15378 (2013).
17. Brodin, J. D. et al. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. Nat. Chem. 4, 375–382 (2012).
18. Sinclair, J. C., Davies, K. M., Vénien-Bryan, C. & Noble, M. E. Generation of protein lattices by fusing proteins with matching rotational symmetry. Nat. Nanotechnol. 6, 558–562 (2011).
19. Koga, N. et al. Principles for designing ideal protein structures. Nature 491, 222–227 (2012).
20. Friedman, S. H. et al. Inhibition of the HIV-1 protease by fullerene derivatives: model building studies and experimental verification. J. Am. Chem. Soc. 115, 6506–6509 (1993).
21. Chen, B. X., Wilson, S. R., Das, M., Coughlin, D. J. & Erlanger, B. F. Antigenicity of fullerenes: antibodies specific for fullerenes and their characteristics. Proc. Natl Acad. Sci. USA 95, 10809–10813 (1998).
22. Braden, B. C. et al. X-ray crystal structure of an anti-Buckminsterfullerene antibody fab fragment: biomolecular recognition of C(60). Proc. Natl Acad. Sci. USA 97, 12193–12197 (2000).
23. Duarte, F., Magistrato, A. & Bonifazi, D. Interfacing proteins with graphic nanomaterials: from spontaneous attraction to tailored assemblies. Chem. Soc. Rev. 44, 6916–6935 (2015).
24. Calvaresi, M. & Zerbetto, F. Baiting proteins with C60. ACS Nano 4, 2283–2289 (2010).
25. Li, L. et al. The molecular mechanism of fullerene-inhibited aggregation of Alzheimer’s β-amyloid peptide fragment. Nano Lett. 6, 9752–9762 (2014).
26. Georgakilas, V. et al. Supramolecular self-assembled fullerene nanostructures. Proc. Natl Acad. Sci. USA 99, 5075–5080 (2002).
Author contributions

K.-H.K., C.B.M., G.G., R.A., Y.H.K. and W.F.D. conceived and designed the study. K.-H.K. and N.H.K. synthesized the peptide. K.-H.K., N.H.K., J.P. and R.A. performed crystallization and data collection, and solved the crystal structures. G.G. K.-H.K. and N.H.K. characterized the electrical conductance. G.G., K.-H.K., C.B.M., G.G., R.A., Y.H.K. and W.F.D. integrated data and wrote the manuscript.

Additional information

Accession codes: The coordinates for the X-ray structures have been deposited to the Protein Data Bank (PDB) with accession codes 5ETF, 5HKN and SHKR.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Kim, K.-H. et al. Protein-directed self-assembly of a fullerene crystal. Nat. Commun. 7:11429 doi: 10.1038/ncomms11429 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material.

To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/