LETTER

doi:10.1038/nature23320

Proteins evolve on the edge of supramolecular self-assembly

Hector Garcia-Seisdedos[1], Charly Empereur-Mot[1]†, Nadav Elad[2] & Emmanuel D. Levy[1]

The self-association of proteins into symmetric complexes is ubiquitous in all kingdoms of life[1–6]. Symmetric complexes possess unique geometric and functional properties, but their internal symmetry can pose a risk. In sickle-cell disease, the symmetry of haemoglobin exacerbates the effect of a mutation, triggering assembly into harmful fibrils[7]. Here we examine the universality of this mechanism and its relation to protein structure geometry. We introduced point mutations solely designed to increase surface hydrophobicity among 12 distinct symmetric complexes from *Escherichia coli*. Notably, all responded by forming supramolecular assemblies *in vitro*, as well as *in vivo* upon heterologous expression in *Saccharomyces cerevisiae*. Remarkably, in four cases, micrometre-long fibrils formed *in vivo* in response to a single point mutation. Biophysical measurements and electron microscopy revealed that mutants self-assembled in their folded states and so were not amyloid-like. Structural examination of 73 mutants identified supramolecular assembly hot spots predictable by geometry. A subsequent structural analysis of 7,471 symmetric complexes showed that geometric hot spots were buffered chemically by hydrophilic residues, suggesting a mechanism preventing mis-assembly of these regions. Thus, point mutations can frequently trigger folded proteins to self-assemble into higher-order structures. This potential is counterbalanced by negative selection and can be exploited to design nanomaterials in living cells.

The ubiquity of symmetry in biological systems testifies to the unique properties it enables[1–6]. Symmetry allows compact genetic encoding of large protein assemblies such as viral capsids and cytoskeleton tubules and filaments. Symmetric oligomers also allow cooperative, switch-like transitions, as in the conformational changes promoting oxygen capture and release by haemoglobin. Additionally, the repetition of subunits within a protein oligomer introduces multivalence, a central element of polymer and supramolecular chemistry (Fig. 1a). The potential of symmetric oligomers to form supramolecular assemblies has in fact been demonstrated *in vitro* by synthetic design of protein fibres[8], nanotubes[9], cages[10–13], and lattices[14] with remarkable mechanical properties[15]. Whether evolution frequently samples such supramolecular assemblies is, however, unknown.

To trigger a new mode of self-assembly, a new interaction of the protein oligomer with itself must be created. In sickle-cell disease, a new interaction is created by a single glutamate-to-valine mutation[7]. This minor surface modification, repeated on opposite sides of the haemoglobin oligomer owing to its internal symmetry, results in the formation of supramolecular fibres. Is the power of a single point mutation to generate supramolecular assemblies unique to haemoglobin, or does it reflect a general property of protein surfaces? Notably, an analysis of protein surface characteristics has shown that solvent-exposed faces of proteins are, on average, within a few mutations away from resembling protein–protein interaction interfaces[16]. As a result, it can be predicted that random mutations at protein surfaces have the potential to create new interaction sites.

We investigated the capacity of surface mutations to trigger new interactions leading to supramolecular assembly. We used a strategy consisting of an increase in surface hydrophobicity with no regard for other factors such as geometrical or charge complementarity. In contrast to classic protein engineering experiments in which *de novo* properties are selected from libraries containing 10^6–10^9 mutants, we created only a few mutants (fewer than ten) per protein studied. The simplicity of our strategy makes our results uniquely amenable to evolutionary interpretation, as any phenotype we observe should often be sampled during evolution.

We initially examined a peptidase from *E. coli* that assembles into a homo-octamer. We introduced point mutations increasing the hydrophobicity at the head of the ring, reasoning that a new interaction between rings could trigger stacking into fibres (Fig. 1b). To monitor the self-assembly state *in vivo*, we fused a yellow fluorescent protein (YFP) to the subunit forming the octamer (Fig. 1c). To increase the octamer's surface hydrophobicity, we mutated a cluster of solvent-exposed and charged residues (E239/E243/K247) to leucine (Fig. 1d). The wild-type octamer showed homogeneous fluorescence when expressed in yeast cells, but the triple leucine mutant self-assembled into micrometre-long fibres spanning entire cells, as did a triple tyrosine mutant (Fig. 1d). Mutating in turn each amino acid of the triplet to leucine did not show a noticeable effect. Remarkably, however, a point mutant to tyrosine (E239Y) sufficed to trigger the supramolecular

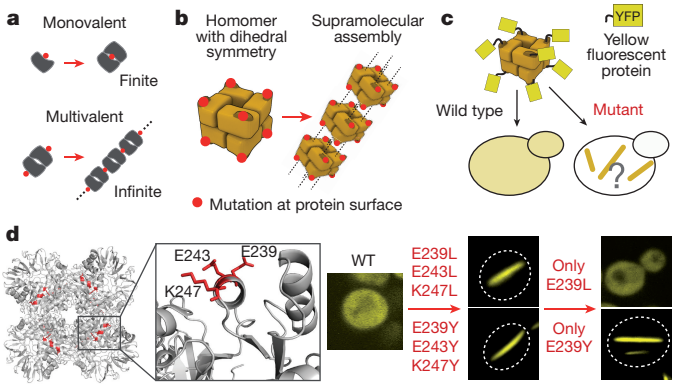


Figure 1 | Supramolecular self-assembly of homomers upon acquisition of a new self-interaction. a, A molecule gaining a single self-interacting patch forms a finite dimer. A self-interacting patch repeated on opposite sides of a symmetric molecule can result in infinite assembly. **b**, A point mutation in a dihedral octamer creates a new self-interacting patch (red), triggering assembly into a fibre. **c**, We fused proteins to YFP to monitor their assembly in *S. cerevisiae*. **d**, The structure of an octameric homomer (Protein Data Bank (PDB) accession 1POK). Expression of the wild-type homomer showed homogeneous fluorescence. Mutation of three charged amino acids to either leucine or tyrosine triggered fibre formation *in vivo*. Mutating in turn each of the three amino acids revealed a point mutant forming fibres (E239Y).

[1]Department of Structural Biology, Weizmann Institute of Science, Rehovot 7610001, Israel. [2]Department of Chemical Research Support, Weizmann Institute of Science, Rehovot 7610001, Israel.
†Present address: Conservatoire National des Arts et Métiers, 75003 Paris, France.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

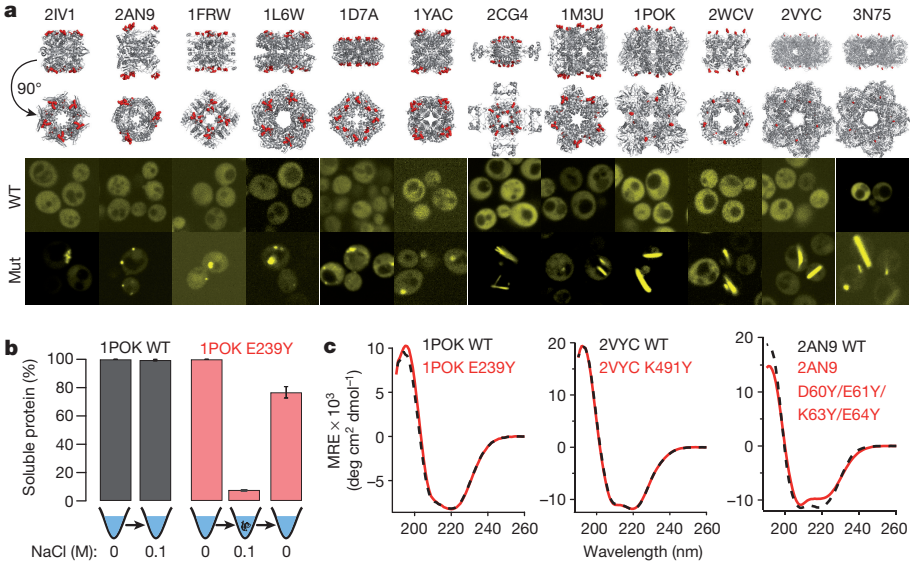


Figure 2 | Increasing surface hydrophobicity triggers supramolecular assembly. a, Structures of 12 dihedral homomers. Corresponding PDB accession numbers are indicated on top. Wild-type proteins gave homogeneous fluorescence upon expression in yeast. Supramolecular assembly of all proteins was triggered by increasing their surface hydrophobicity by point mutations (red). Among the 12 mutants shown, the right-most 6 form fibres and the left-most 6 form foci (Supplementary Table 2). **b**, The supramolecular assembly of the peptidase mutant (E239Y) depended on salt concentration. **c**, We assessed the structural integrity of nine mutants by circular dichroism (Extended Data Fig. 3), of which three are shown. All spectra reflected a folded state, albeit three of them changed relative to the wild type. Spectra are shown as mean residue ellipticity (MRE).

assembly of the enzyme (Extended Data Fig. 1 and Supplementary Video 1).

The fact that we did not rely on any prediction tool for choosing these mutations suggests that the phenotype we observed is naturally frequent and that this enzyme evolves on the edge of supramolecular self-assembly. To assess the generality of this notion, we performed similar experiments for 11 additional homomers (Extended Data Table 1 and Supplementary Table 1), mutating charged surface residues either to leucine or tyrosine. These mutations triggered the formation of fibres in five additional structures, among which three were single point mutants (Fig. 2a and Supplementary Table 2). In the six remaining structures, we observed punctate foci rather than fibres. In all cases, self-assembly took place at concentrations within a physiological range (Extended Data Fig. 2).

Formation of protein fibres is typically associated with misfolding and amyloids[17,18], so we examined the biophysical and structural properties of the mutant proteins. Notably, the peptidase supramolecular assembly was reversible by lowering the ionic strength (Fig. 2b), a property inconsistent with amyloid fibrils. Furthermore, we purified nine mutants exhibiting a super-assembly phenotype and compared their secondary structure contents with those of the wild-type proteins by circular dichroism (Supplementary Table 3). In all cases, the circular dichroism spectra indicated that the mutants were folded (Fig. 2c). Lastly, five mutants selected for such analysis showed high thermostability (melting temperature $T_m > 70\,°C$) like their wild-type counterparts (Extended Data Fig. 3). Together, these biophysical properties indicate that the mutations were not detrimental to protein structure or stability, despite their dramatic impact on the assembly phenotype *in vivo*.

To make the proteins amenable to the biophysical characterization described above, we inhibited supramolecular assembly using a mild detergent (*n*-dodecyl-D-maltoside) and the amino acid L-arginine. To gain structural insights under conditions where self-assembly occurs, we used transmission electron microscopy (TEM) (Fig. 3a and Extended Data Fig. 4). Of the six structures self-assembling into fibres *in vivo*, five also formed fibres *in vitro*. Also, two of the mutants forming punctate foci *in vivo* assembled into fibres *in vitro* (Fig. 3a and Extended Data Table 2). The relatively large size of individual homomers repeated along the fibres allowed us to estimate their size from TEM images (Extended Data Fig. 4). We observed an excellent match between the size of fibre-forming mutants and their wild-type counterparts ($R^2 = 0.94$, $P = 3.6 \times 10^{-4}$; Fig. 3b), indicating that self-assembly occurs

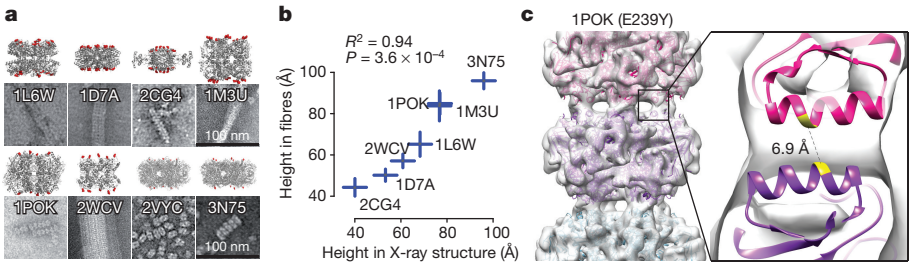


Figure 3 | Supramolecular assembly of homomers occurs in their folded states. a, We examined mutants by negative-stain electron microscopy and observed fibres for five out of six mutants forming fibres *in vivo*. In addition, two mutants forming foci *in vivo* formed fibres *in vitro*. **b**, We measured distances separating adjacent homomers in fibres seen by electron microscopy and compared them with the heights of the homomers measured from the crystallographic structures. Measurements were in good agreement ($R^2 = 0.94$), indicating that fibre-forming mutants maintain the structures of their wild-type counterparts. **c**, Single-particle cryo-electron microscopy reconstruction of the peptidase mutant fibres. Rigid-body fitting of the wild-type peptidase into the map revealed the point of contact between adjacent homomers. The interface involves the α-helix where the mutation was introduced (yellow).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

through the folded state. We also verified that fusion of the YFP tag did not affect the formation of fibres (Extended Data Fig. 4). To capture the molecular details of fibre formation, we obtained a density map of a fibre by cryo-TEM single-particle reconstruction (Electron Microscopy Data Bank accession number EMD-4094; Extended Data Fig. 5). The map had a local resolution of 7–8 Å in the central octamer and at the interface, which included the α-helix where the E239Y mutation was introduced (Fig. 3c). Analysis of side chains at the interface suggested several interactions that may stabilize the assembly (Extended Data Fig. 6).

Altogether, we created 73 mutants, 30 of which triggered supramolecular assemblies *in vivo*, in the form of punctate foci or fibres (Supplementary Table 2). These mutants allowed us to examine properties of amino acids that trigger *de novo* intermolecular interactions upon mutation. This analysis identified a novel structural property: the normal distance to the closest bounding plane of the homomer (henceforth nDp, Fig. 4a and Supplementary Table 5). In contrast to accessible surface area, which describes local solvent exposure, nDp depends on a residue's position on the global quaternary structure. The lower nDp, the closer the amino acid is to the apex of the quaternary structure along a symmetry axis, and the more it enables the interaction of the quaternary structure with a copy of itself. Accordingly, we found that positions where mutations triggered fibre formation exhibited lower median nDp values (2.1 Å; Fig. 4b) than positions with no phenotypic impact (4.7 Å, $P < 0.004$, *t*-test). Positions associated with the formation of foci also showed lower values for nDp, although the difference was not statistically significant.

Thus, by geometry, surface regions with low nDp are hot spots for supramolecular assembly. In civil engineering, structures are strengthened locally, according to the load endured. In a similar manner, the greater potential of geometric hot spots to self-interact might be counterbalanced chemically by the presence of amino acids with low interaction propensity. Such negative design can protect against folding into alternative structures[19,20] and might protect against non-native interactions[21]. We measured the interaction propensity of surface regions by their 'stickiness', defined by the enrichment of amino acids at protein–protein interfaces relative to solvent-exposed surfaces[22] (Extended Data Fig. 7). Analysis of 1,990 dihedral complexes of known structure revealed that geometric hot spots have significantly lower stickiness compared with the rest of the protein surface (Fig. 4c, Supplementary Table 6 and Supplementary Data 1). As a control experiment, we performed the same analyses among 5,481 homomers with cyclic symmetry, which are less likely than dihedral homomers to form fibres. Low values of nDp were not associated with low stickiness in cyclic homomers (Extended Data Fig. 7), suggesting that negative selection acts against infinite assembly rather than against finite dimerization. In another control, we introduced mutations on homomer faces through which the two-fold symmetry axes pass, and did not observe the formation of fibres (Supplementary Table 4).

This analysis of protein structures indicates that mutations inducing fibre formation can act through two mechanisms. First, the residues we mutated (K, D, and E) have the lowest interaction propensities[16]. Thus, their mutation knocks out their protective effect against protein self-association and fibre formation. Second, mutations to hydrophobic residues provide a source for gain in solvent entropy upon binding. Considering that leucine and tyrosine side chains can bury ~100 Å^2, each mutation can yield 600–1,000 Å^2 of surface to bury in homomers with six to ten subunits, which is typical of biological interfaces[23–25]. While we focused our analysis on leucine and tyrosine, we anticipate that mutations to other hydrophobic or aromatic residues will show a similar potential to promote new interfaces.

Besides the mutated region, additional 'passenger' contact regions are expected in the fibre assembly, and these must be energetically neutral or favourable. Here, minimally designed point mutations sufficed to trigger new interactions, even with proteins in nanomolar concentrations (Extended Data Figs 2 and 8). This finding implies that passenger

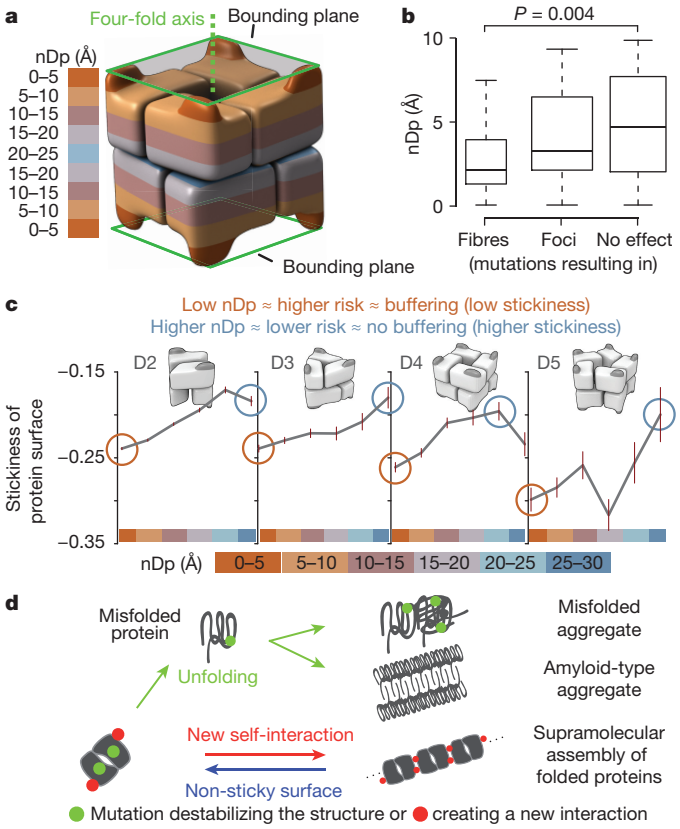


Figure 4 | Self-assembly potential constrains protein surface chemistry. **a**, We define a geometric property of homomers, which is the normal distance to the closest bounding plane (nDp). Bounding planes bound the structure and are orthogonal to symmetry axes. We show two bounding planes defined by the four-fold symmetry axis (green) and corresponding nDp values projected onto the structure as coloured bands. **b**, We computed nDp for all mutated positions and found that mutations triggering fibres showed lower nDp values than mutations with no phenotypic effect. **c**, We analysed 1,990 dihedral homomers of known structure and found that surface patches with high potential to trigger supramolecular assembly upon mutation counterbalance this potential by amino acids with low interaction propensity or 'stickiness'[22]. Grey lines show the median stickiness of surface patches in any given nDp window, and error bars correspond to two standard errors. **d**, Protein aggregation is usually associated with destabilizing mutations and misfolding. We show that aggregation by supramolecular assembly can occur frequently through an alternative pathway, where surface mutations trigger new self-interactions.

contacts have a high probability of showing geometric and chemical complementarity, confirming predictions that symmetry enhances interaction propensity of protein surfaces[1,3,4].

This work demonstrates that protein surfaces are prone to interact by chance and that amplification of such interactions by symmetry can drive supramolecular assembly. This under-appreciated property, in fact, is the basis for protein X-ray crystallography, which requires proteins to self-assemble into crystals. In that respect, it is notable that aggregation-inducing mutations are dogmatically interpreted in the context of misfolding. Our work indicates that a ready pathway to aggregation is the uncontrolled assembly of folded proteins (Fig. 4d). This pathway, which may be further amplified by co-localization[26], will be important to consider in future studies predicting the molecular consequences of mutations, including single nucleotide polymorphisms.

Our observations imply that supramolecular assemblies of folded proteins are frequently sampled by evolution. Consistent with this view, recent works have revealed that specific proteins can self-assemble reversibly into foci[27,28] or fibres[29] in response to changes in cell

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

physiology[30] (Supplementary Text 1). The ease with which they can evolve, as demonstrated here, suggests that many more exist.

Lastly, from a synthetic biology perspective, we have described a simple strategy to program protein self-assembly at length scales of several micrometres *in vivo*. With their composition known and the structure of one fibre characterized, these assemblies pave the way to modelling, engineering and reprogramming protein self-assembly *in vivo*.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 July 2016; accepted 20 June 2017.
Published online 2 August 2017.

1. Lukatsky, D. B., Shakhnovich, B. E., Mintseris, J. & Shakhnovich, E. I. Structural similarity enhances interaction propensity of proteins. *J. Mol. Biol.* **365,** 1596–1606 (2007).
2. Levy, E. D., Boeri Erba, E., Robinson, C. V. & Teichmann, S. A. Assembly reflects evolution of protein complexes. *Nature* **453,** 1262–1265 (2008).
3. André, I., Strauss, C. E., Kaplan, D. B., Bradley, P. & Baker, D. Emergence of symmetry in homooligomeric biological assemblies. *Proc. Natl Acad. Sci. USA* **105,** 16148–16152 (2008).
4. Schulz, G. E. The dominance of symmetry in the evolution of homo-oligomeric proteins. *J. Mol. Biol.* **395,** 834–843 (2010).
5. Ahnert, S. E., Marsh, J. A., Hernández, H., Robinson, C. V. & Teichmann, S. A. Principles of assembly reveal a periodic table of protein complexes. *Science* **350,** aaa2245 (2015).
6. Levy, E. D. & Teichmann, S. Structural, evolutionary, and assembly principles of protein oligomerization. *Prog. Mol. Biol. Transl. Sci.* **117,** 25–51 (2013).
7. Dykes, G. W., Crepeau, R. H. & Edelstein, S. J. Three-dimensional reconstruction of the 14-filament fibers of hemoglobin S. *J. Mol. Biol.* **130,** 451–472 (1979).
8. Grueninger, D. *et al.* Designed protein-protein association. *Science* **319,** 206–209 (2008).
9. Brodin, J. D., Smith, S. J., Carr, J. R. & Tezcan, F. A. Designed, helical protein nanotubes with variable diameters from a single building block. *J. Am. Chem. Soc.* **137,** 10468–10471 (2015).
10. Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl Acad. Sci. USA* **98,** 2217–2221 (2001).
11. Lai, Y. T., Cascio, D. & Yeates, T. O. Structure of a 16-nm cage designed by using protein oligomers. *Science* **336,** 1129 (2012).
12. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336,** 1171–1174 (2012).
13. Fletcher, J. M. *et al.* Self-assembling cages from coiled-coil peptide modules. *Science* **340,** 595–599 (2013).
14. Lanci, C. J. *et al.* Computational design of a protein crystal. *Proc. Natl Acad. Sci. USA* **109,** 7304–7309 (2012).
15. Suzuki, Y. *et al.* Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **533,** 369–373 (2016).
16. Levy, E. D. A simple definition of structural regions in proteins and its use in analyzing interface evolution. *J. Mol. Biol.* **403,** 660–670 (2010).
17. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148,** 1188–1203 (2012).
18. Knowles, T. P., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15,** 384–396 (2014).
19. Hecht, M. H., Richardson, J. S., Richardson, D. C. & Ogden, R. C. De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science* **249,** 884–891 (1990).
20. DeGrado, W. F., Summa, C. M., Pavone, V., Nastri, F. & Lombardi, A. De novo design and structural characterization of proteins and metalloproteins. *Annu. Rev. Biochem.* **68,** 779–819 (1999).
21. Doye, J. P., Louis, A. A. & Vendruscolo, M. Inhibition of protein crystallization by evolutionary negative design. *Phys. Biol.* **1,** 9–13 (2004).
22. Levy, E. D., De, S. & Teichmann, S. A. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc. Natl Acad. Sci. USA* **109,** 20461–20466 (2012).
23. Chothia, C. & Janin, J. Principles of protein–protein recognition. *Nature* **256,** 705–708 (1975).
24. Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl Acad. Sci. USA* **93,** 13–20 (1996).
25. Wodak, S. J. & Janin, J. Structural basis of macromolecular recognition. *Adv. Protein Chem.* **61,** 9–73 (2002).
26. Kuriyan, J. & Eisenberg, D. The origin of protein interactions and allostery in colocalization. *Nature* **450,** 983–990 (2007).
27. Narayanaswamy, R. *et al.* Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proc. Natl Acad. Sci. USA* **106,** 10147–10152 (2009).
28. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483,** 336–340 (2012).
29. Noree, C., Sato, B. K., Broyer, R. M. & Wilhelm, J. E. Identification of novel filament-forming proteins in *Saccharomyces cerevisiae* and *Drosophila melanogaster*. *J. Cell Biol.* **190,** 541–551 (2010).
30. Gsponer, J. & Babu, M. M. Cellular strategies for regulating functional and nonfunctional protein aggregation. *Cell Reports* **2,** 1425–1437 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Wolf and E. Shimoni for help with electron microscopy experiments, and J. Georgeson for setting up the microscope time-lapse. We thank members of the laboratory, D. Fass and A. Horovitz for discussions throughout the realization of this work, H. Weissman for discussions about electron microscopy, and D. Fass for invaluable feedback on the manuscript. This work was supported by the Israel Science Foundation and the I-CORE Program of the Planning and Budgeting Committee (grants 1775/12 and 2179/14), by the Marie Curie Career Integration Grants Program (number 711715), by the Human Frontier Science Program Career Development Award (number CDA00077/2015), and by a research grant from A.-M. Boucher. H.G.S. received support from the Koshland Foundation and a McDonald-Leapman Grant. Electron microscopy studies were supported by the Irving and Cherna Moskowitz Center for Nano and Bio-Nano Imaging. E.D.L. is incumbent of the Recanati Career Development Chair of Cancer Research.

Author Contributions H.G.S. and E.D.L. designed the experiments. C.E.M. and E.D.L. designed the computational analyses of protein structure. H.G.S. performed all experiments with guidance from N.E. for electron microscopy techniques. N.E. and H.G.S. performed the single-particle reconstruction. C.E.M. performed the bioinformatics analyses. H.G.S. and E.D.L. analysed the experimental data. C.E.M. and E.D.L. analysed the computational results. E.D.L. and H.G.S. wrote the manuscript with help from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to E.D.L. (emmanuel.levy@weizmann.ac.il).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Selection of proteins used in this work. We searched the 3DComplex database[31] for *E. coli* homo-oligomers exhibiting dihedral symmetry. The structures chosen were not flagged as being erroneous in the PiQSi database[32]. Structures not annotated in PiQSi underwent manual inspection to discard those with a likely erroneous quaternary state. All dihedral homo-oligomers with higher-order symmetry were retained (D_4, D_5), and two homomers with D_3 symmetry were chosen at random. This process resulted in 16 structures that were initially selected. Details of these structures, including their PDB accession numbers, publication references, number of subunits, symmetries, cellular abundances, and sequence identifiers, are given in Extended Data Table 1.

Cloning procedures and mutagenesis. Genes encoding the 16 homomers were amplified from the strain *E. coli* K12. They were cloned into yeast expression vector p413 GPD (ATCC 87354)[33]. Genes corresponding to the structures of 1POK, 2VYC, 1M3U, and 1D7A were also cloned into the vector p413 CYC1 (ATCC 87378)[33], and the gene encoding the sequence of 1POK E239Y without YFP fusion was cloned into the vector p416 GPD (ATCC 87360)[33]. Molecular cloning was done using the PIPE method[34]. To visualize the homomers *in vivo*, we fused them to a YFP (Venus)[35] separated by a flexible linker of sequence GGGGSGGGGS. Whether the YFP sequence was appended to the N or C terminus was determined by inspection of the structure with the aim of leaving the surface corresponding to the *n*-fold symmetry axis accessible. Upon expression of the 16 proteins in yeast cells, three proteins did not show homogeneous cytosolic localization and were discarded from subsequent analyses (Supplementary Table 1). At this stage, further inspection of all structures revealed that one was erroneously assigned a dihedral symmetry, and was also discarded (Supplementary Table 1). The other 12 homomers underwent site-directed mutagenesis (QuickChange, Agilent). All mutations introduced were either towards leucine or tyrosine. We exclusively targeted residues with high solvent exposure to minimally impact stability: that is, 25% or more relative accessible solvent area, with the majority showing >50% relative accessible solvent area.

Yeast strains and media. Yeast cells were grown in synthetic complete medium lacking histidine (SC-His) and uracil (SC-His-Ura) for co-expression of the YFP-tagged and untagged versions of 1POK E239Y. Plasmids encoding the proteins studied were transformed into *S. cerevisiae* strain BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) using the LiAc method, as described previously[36]. Transformants were inoculated from the Petri plates into a 384-well plate containing SC-His (or SC-His-Ura for the strain co-expressing the YFP-tagged and untagged variants of 1POK E239Y) with 15% glycerol and were stored at −80 °C.

Microscopy. Plate-to-plate transfers of cells were performed using a pintool (FP1 pins, V&P Scientific) operated by a Tecan robot (Evo200 with MCA384 head). Cells were inoculated from glycerol stocks into a polypropylene 384-well plate containing SC-His (or SC-His-Ura for the strain co-expressing the YFP-tagged and untagged variants of 1POK E239Y), and grown until they reached saturation. Cells from the saturated cultures were inoculated into 384-well glass-bottom optical plates (Matrical) filled with media. Cells were grown for at least 6 h until they reached an absorbance of 0.5–1 before being imaged by microscopy. Imaging was done by an automated Olympus microscope X83 with a 60× oil objective (Olympus, plan apo, 1.42 numerical aperture) coupled to a spinning disk confocal scanner (Yokogawa W1). A 488 nm laser (Toptica, 100 mW) was used for fluorescence excitation, and a green LED was used for brightfield imaging. The emission filter-sets used for brightfield and fluorescence images were identical (520/28, Chroma). Images were recorded on a Hamamatsu Flash4 camera. Focus was maintained throughout the imaging experiment by hardware autofocus (Olympus Z-Drift Compensation System).

Time-lapse series were performed in a 96-well glass-bottom optical imaging plate (Matrical), which was coated with concanavalin A as described previously[37]. After coating, a cell culture (absorbance ≈ 0.5–1) was dispensed into the well. Cells were incubated for 15 min, and the liquid was then removed from the plate. Cells adhering to the plate were washed three times with SC-His and, to maximize growth as a monolayer of cells, 200 µl of SC-His supplemented with 7.5 mg ml^{-1} low melting temperature agarose was placed on the sample in the imaging well. Images were taken every 230 s for 10.5 h.

Estimation of intracellular protein concentrations. Intracellular protein concentrations were estimated by calibration against reference solutions containing known concentrations of YFP. Purified YFP was serially diluted in PBS buffer to concentrations ranging from 110 µM to 1 nM. The YFP solutions were transferred to the plate used to image cells. Images of the YFP solutions were analysed using ImageJ[38] by recording their mean fluorescence intensity, which showed that the microscope signal increased linearly with the concentration of fluorescent protein. The equation inferred from linear regression enabled us to convert fluorescence arbitrary units into YFP molarity (1 nM = 0.998 fluorescence arbitrary units − 0.16). We then used the equation so obtained to transform the intracellular fluorescence signal into homo-oligomer concentration. To measure intracellular fluorescence, we used ImageJ[38] to select at least 50 cells per strain on the basis of brightfield pictures. We drew a region of interest for each cell, and subsequently extracted its median fluorescence in the fluorescence channel. For each strain, the median fluorescence across all cells was converted to YFP molarity using the equation obtained above.

Protein expression and purification. Genes encoding the wild-type proteins and mutants were cloned into the vectors pET30b$^+$ and pET29b$^+$, containing an N- or C-terminal 6× His-tag, respectively, and a kanamycin resistance cassette. Vectors were transformed into BL21(DE3) cells. Transformants were grown in lysogeny broth at 37 °C until they reached an absorbance of 0.6. Then, protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside to a final concentration of 0.4 mM, and cells were grown overnight at 25 °C. Cells were harvested by centrifugation and resuspended in lysis buffer: sodium phosphate 20 mM, 500 mM NaCl, and 20 mM imidazole, pH 7.4. Lysozyme and benzonase were added to the lysis buffer at a concentration of 0.2 mg ml^{-1} and 50 units ml^{-1} respectively. Cells were lysed by sonication and centrifuged. From this point, three procedures were used to purify the different proteins, as follows.

1. Wild-type proteins, isolated YFP, and mutants D497L and K491Y of 2VYC were purified from the soluble fraction of the cell lysates. Crude lysates were loaded on His Gravitrap columns (GE Healthcare) and eluted in sodium phosphate 20 mM, 500 mM NaCl, and 500 mM imidazole, pH 7.4.

2. Other mutants were present in the insoluble fraction of the cell lysates. Therefore, pellets were washed twice with lysis buffer 1% Triton X-100 and one more time with lysis buffer. Solubilization of inclusion bodies was performed by resuspending the pellets in Tris 20 mM, 2 M L-arginine, pH 7.5.

3. Purification of 1FRW D170L/D173L/K175L/K176L variant was done in denaturing conditions. After harvesting, cells were resuspended in 7 M urea, sodium phosphate 20 mM, 500 mM NaCl, and 20 mM imidazole, pH 7.4. They were lysed by sonication, loaded into a His gravitrap column, and the elution was done in 7 M urea, sodium phosphate 20 mM, 500 mM NaCl, and 500 mM imidazole, pH 7.4. Refolding was achieved by removing urea by dialysis.

The purity of all the samples was assessed by SDS–polyacrylamide gel electrophoresis and protein concentration was determined using a NanoDrop ND-1000 by absorbance at 280 nm. Purified proteins were dialysed against suitable buffers for subsequent characterization, as described in Supplementary Table 3.

Precipitation assay of 1POK E239Y. Three solutions of the wild-type and mutant (E239Y) isoaspartyl dipeptidase (PDB 1POK) were prepared at a concentration of 1 mg ml^{-1} (measurement 1) in Tris 20 mM, pH 7.5. NaCl was subsequently added to a final concentration of 0.1 M to every sample. Samples were pelleted by centrifugation at 20,000*g* for 20 min, and the protein concentration of the soluble fractions was re-measured using a NanoDrop spectrometer (measurement 2). Pellets were re-dissolved in Tris 20 mM, pH 7.5 without NaCl this time. Re-solubilized pellets were centrifuged for 20 min at 20,000*g* and the protein concentration of the soluble fraction was measured again, corrected for the amount of soluble protein discarded in measurement 2, and compared with the original concentration.

Precipitation assay of 1POK E239Y tagged with YFP. Five serial dilution series of 1POK E239Y–YFP in Tris 20 mM, pH 7.5 were prepared. The concentration of octamers ranged between 6.25 µM and 9 nM. The samples were transferred to a 384-well glass-bottom plate and imaged using the confocal microscope as described in the section on Microscopy. After imaging, 10× PBS was added to four dilution series to reach a final concentration of 1× PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4). The samples were incubated for 2 h at 30 °C. Samples were then centrifuged for 15 min at 2,500*g*, and 10 µl of the supernatant were imaged. Images were analysed with ImageJ, using the median fluorescence intensity of each image to estimate the concentration of soluble protein. The ratio of fluorescence intensity in the supernatant relative to the intensity in the sample where no PBS was added gave the fraction of soluble protein. Resulting values were multiplied by 1.11 to account for the PBS dilution.

Circular dichroism experiments. Circular dichroism experiments were performed on a Chirascan spectrometer (Applied Photophysics). A mild detergent, *n*-dodecyl β-D-maltoside, was used to increase the solubility of some of the proteins assessed. Secondary structure content and thermal stability of wild types/mutants were always compared using the same buffer for each pair (Supplementary Table 3). Protein concentrations used in all the experiments ranged between 0.1 and 0.2 mg ml^{-1}. Thermal denaturation experiments were monitored by following ellipticity at 220 nm from 25 to 90 °C at a heating rate of 1 °C min^{-1}. Since many of the proteins did not exhibit a thermal denaturation transition, we used guanidine hydrochloride to a concentration of 2.5 M at 90 °C to achieve complete denaturation.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

This last measurement was used to infer ellipticity of the unfolded state and served as a reference for normalization.

TEM data acquisition. All proteins studied by electron microscopy were prepared in 20 mM Tris, pH 7.5, except for 1POK E239Y and 1POK E239Y–YFP for which we added 100 mM NaCl, and for 2VYC (wild type, D497L, and K491Y mutants) which were prepared in 100 mM sodium phosphate, pH 7. This last buffer was described to maintain the decameric form of 2VYC[39]. Ten microlitres of the protein samples at about 0.1 mg ml^{-1} concentration were applied to glow-discharged, carbon-coated copper TEM grids (300 mesh, EMS) for 10–20 s. Excess liquid was then blotted, and after a wash with distilled water, the grids were stained with a 2% uranyl acetate solution. Samples were visualized in an FEI Tecnai Spirit or T12 transmission electron microscope, equipped with an FEI Eagle camera or a Gatan ES500W Erlangshen camera respectively.

For single-particle cryo-electron microscopy, 3.5 μl of 1POK E239Y solution (Tris 20 mM, 100 mM NaCl, pH 7.5, 0.2 mg ml^{-1}) was applied to glow-discharged Quantifoil holey carbon grids (R2/1, 300 mesh) coated with a thin layer of carbon. Grids were plunge-frozen in liquid ethane cooled by liquid nitrogen, using a Leica EM-GP plunger (3.5 s blotting time, 95% humidity). Grids were imaged at liquid nitrogen temperature on a FEI Tecnai TF20 electron microscope operated at 200 kV with a Gatan side entry 626 cryo-holder and a condenser aperture of 30 μm. Images were recorded on a K2 Summit direct detector (Gatan) mounted at the end of a GIF Quantum energy filter (Gatan). Images were collected in super-resolution counting mode, at a calibrated magnification of 23,657, yielding a physical pixel size of 2.11 Å. The dose rate was set to ~2 electrons per square ångström per second and a total exposure time of 8 s, fractionated into 40 subframes. Defocus range was 1.2–4 μm. All dose-fractionated images were recorded using an automated low-dose procedure implemented in SerialEM[40].

Single-particle cryo-electron microscopy image processing. Recorded image frames (2–40) were binned by a factor of 2 and subjected to whole-image beam-induced motion correction using MOTIONCORR[41]. Contrast transfer function parameters were estimated using CTFFIND3 (ref. 42). A total of 574 images were selected for processing, from which 38,786 particles (1POK E239Y octamers) were manually picked using EMAN2 e2boxer[43] and extracted into 128 pixel × 128 pixel boxes. Octamers were picked only from within filaments (Extended Data Fig. 5c) so that all extracted boxes contained one octamer in the centre and two adjacent octamers from the filament. Subsequent classifications and refinements were performed using RELION[44]. Particles underwent two-dimensional classification with a round mask of 160 Å diameter (Extended Data Fig. 5d). The vast majority of the particles were included in good classes and retained for subsequent processing. Particles were then classified in three-dimensions into four classes with sphere masks of 160 Å diameter and D_4 symmetry imposed, using the octameric structure solved by X-ray crystallography and filtered to 40 Å as an initial model[45]. Then, particles from the two best three-dimensional classes were combined (17,277 in total) and subjected to three-dimensional refinement with a sphere mask of 136 Å diameter and D_4 symmetry imposed (Extended Data Fig. 5e). Local resolution estimation using ResMap[46] indicated resolution of 7–8 Å in the central octamer and resolution higher than 10 Å in the adjacent octamers (Extended Data Fig. 5f), probably because of flexibility within the filaments. Resolution at the interface-forming helices, which contained the E239Y mutation, was about 8 Å, indicating that this segment is relatively rigid. The global resolution in the final structure, estimated by the gold-standard Fourier shell correlation = 0.143 criterion (Extended Data Fig. 5g), was 10.5 Å.

Cyclic tetramers from the crystallographic structure[45] were fitted in the density map using UCSF Chimera[47]. According to this fit, octamers stacked into a filament along the four-fold symmetry axis with 80.3 Å translation between adjacent octamers. This structure could be interpreted as a left-handed helix with 33.3° rotation between adjacent octamers, or a right-handed helix with 66.7° rotation. The α-helix where the mutation was introduced showed a good fit (Fig. 3c) and was at the main contact point between octamers.

Image analysis. Images showing protein filaments visualized by electron microscopy after negative staining were opened in ImageJ, lines were drawn from one unit to the next along the fibre, and each line was added to the 'region of interest manager'[48]. The distribution of line lengths was then recorded. We used the average of the distribution to infer the distance separating adjacent units, and the standard deviation to estimate measurement errors, which are shown in the plot as vertical bars (Fig. 3b). We compared these distances with the height (h) of the minimum bounding-box containing the crystallographic structure. Because the bounding-box was calculated on the basis of α-carbons, we added a constant value of 5 Å to h to correct for the presence of side chains. We also considered a constant uncertainty of ±5 Å, which is reflected in the figure as horizontal bars.

Computational structural analyses. Proteins of known structure exhibiting either cyclic or dihedral symmetry were retrieved from the 3DComplex database, along with information on symmetry axes[31]. We used a non-redundant set filtered at 80% sequence identity. To avoid biases due to membrane and viral proteins when analysing surface stickiness, we discarded all structures containing one of the following chains of characters in their title or description: '*lipid*', '*transport*', '*rhodopsin*', '*membran*', '*virus*', '*viral*'. A symmetry axis was considered as a unit (1 Å) vector $\boldsymbol{s}$ originating from the centre of mass of the structure. Similarly, the α-carbon of each residue i defined a vector $\boldsymbol{r}_i$ originating from the centre of mass. Two bounding planes orthogonal to a symmetry axis were defined and intersected at the maximal (d_{max}) and minimal (d_{min}) values of the dot product $\boldsymbol{s} \bullet \boldsymbol{r}_i$, considering all residues i. The measure nDp for a given residue i was calculated as the minimal distance to either bounding plane: $nDp_i = \min(d_{max} - \boldsymbol{s} \bullet \boldsymbol{r}_i, \boldsymbol{s} \bullet \boldsymbol{r}_i - d_{min})$. Cyclic symmetries have a single symmetry axis, but dihedral symmetries have multiple axes. For example, D_2 homomers have two two-fold axes. For those, we could not distinguish a unique 'fibre-forming axis', so nDp was computed relative to both axes, and each residue was assigned the lowest of the two values. For higher-order dihedral homomers (D_3, D_4, D_5), nDp was computed in two different ways, which gave similar results, as follows.

A first approach was used in Fig. 4 and Extended Data Fig. 7, where D_3, D_4, and D_5 homomers were processed similarly to D_2 homomers. That is, nDp was computed relative to all symmetry axes, and each residue was assigned the lowest nDp value. Subsequently, we used only those residues for which nDp was lowest relative to the axis along which fibres could grow (three-, four-, and five-fold for D_3, D_4, or D_5 respectively). Extended Data Fig. 7b illustrates nDp values projected onto the structure of 1POK, as well as residues associated with the four-fold symmetry axis used in the analyses. Our rationale for this approach was to exclude potential negative design effects associated with planar assembly along two-fold symmetry axes.

A second approach was used where D_3, D_4, and D_5 homomers were processed as for cyclic homomers. That is, nDp was calculated exclusively according to the three-, four-, and five-fold symmetry axes, without considering the two-fold symmetry axes. Results of this calculation are shown in Extended Data Fig. 7d (green dashed line).

Finally, protein surface stickiness was computed as described earlier[22]. The definition of the amino-acid stickiness scale is also detailed in Extended Data Fig. 7a.

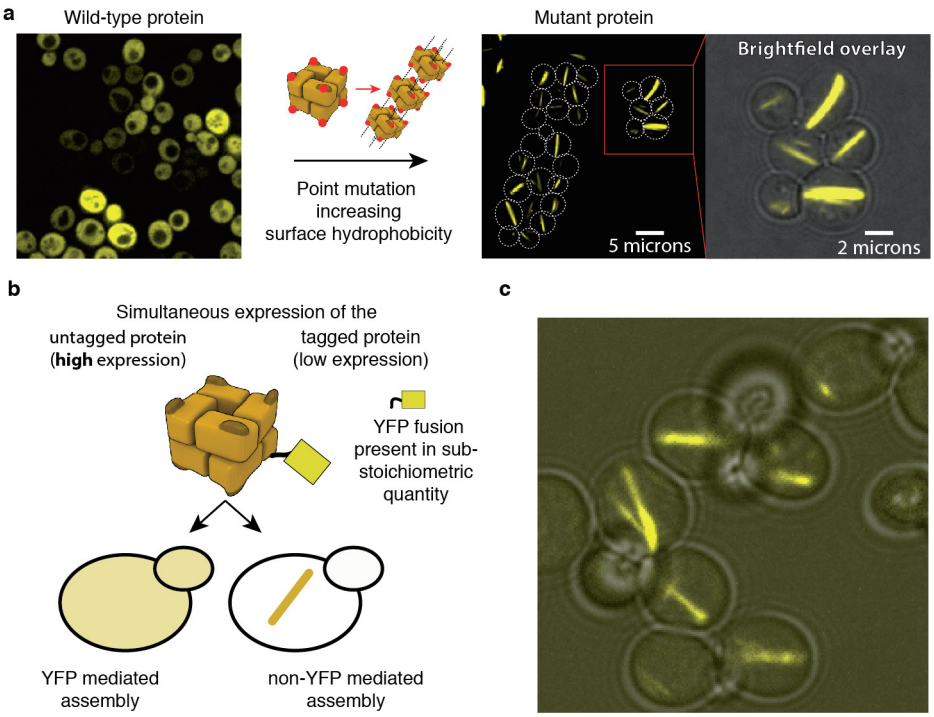
Data and code availability. Structural data used in and resulting from the bioinformatics analyses are available in Supplementary Table 6 and Supplementary Data 1; Supplementary Data 1 is available at Figshare at http://dx.doi.org/10.6084/m9.figshare.5119861. The electron microscopy density map and associated atomic coordinates of the fitted structure have been deposited in the Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB) under accession numbers EMD-4094 and 5LP3, respectively. The R source-code used for analysis and other data are available from the corresponding author upon reasonable request.

31. Levy, E. D., Pereira-Leal, J. B., Chothia, C. & Teichmann, S. A. 3D complex: a structural classification of protein complexes. *PLoS Comput. Biol.* **2,** e155 (2006).
32. Levy, E. D. PiQSi: protein quaternary structure investigation. *Structure* **15,** 1364–1367 (2007).
33. Mumberg, D., Müller, R. & Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156,** 119–122 (1995).
34. Klock, H. E., Koesema, E. J., Knuth, M. W. & Lesley, S. A. Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins* **71,** 982–994 (2008).
35. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20,** 87–90 (2002).
36. Knop, M. *et al.* Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. *Yeast* **15** (10B), 963–972 (1999).
37. Cohen, Y. & Schuldiner, M. Advanced methods for high-throughput microscopy screening of genetically modified yeast libraries. *Methods Mol. Biol.* **781,** 127–159 (2011).
38. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).
39. Boeker, E. A. & Snell, E. E. Arginine decarboxylase from *Escherichia coli*. 2. Dissociation and reassociation of subunits. *J. Biol. Chem.* **243,** 1678–1684 (1968).
40. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152,** 36–51 (2005).
41. Li, X. M. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10,** 584–590 (2013).
42. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142,** 334–347 (2003).
43. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38–46 (2007).
44. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).
45. Jozic, D., Kaiser, J. T., Huber, R., Bode, W. & Maskos, K. X-ray structure of isoaspartyl dipeptidase from *E. coli*: a dinuclear zinc peptidase evolved from amidohydrolases. *J. Mol. Biol.* **332,** 243–256 (2003).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

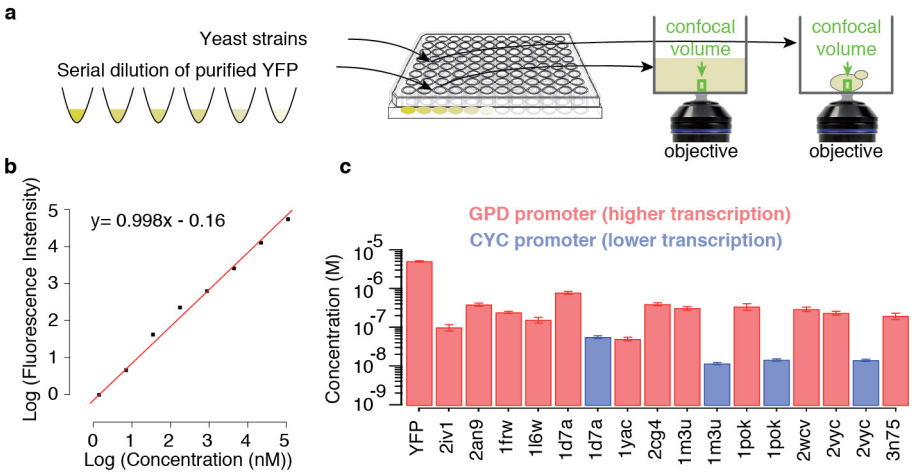
46. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11,** 63– (2014).
47. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
48. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9,** 676–682 (2012).
49. Hible, G. *et al.* Calorimetric and crystallographic analysis of the oligomeric structure of *Escherichia coli* GMP kinase. *J. Mol. Biol.* **352,** 1044–1059 (2005).
50. Hindupur, A. *et al.* The crystal structure of the *E. coli* stress protein YciF. *Protein Sci.* **15,** 2605–2611 (2006).
51. Mathews, I. I., Kappock, T. J., Stubbe, J. & Ealick, S. E. Crystal structure of *Escherichia coli* PurE, an unusual mutase in the purine biosynthetic pathway. *Structure* **7,** 1395–1406 (1999).
52. Lake, M. W., Temple, C. A., Rajagopalan, K. V. & Schindelin, H. The crystal structure of the *Escherichia coli* MobA protein provides insight into molybdopterin guanine dinucleotide biosynthesis. *J. Biol. Chem.* **275,** 40211–40217 (2000).
53. Jaffe, E. K. *et al.* Species-specific inhibition of porphobilinogen synthase by 4-oxosebacic acid. *J. Biol. Chem.* **277,** 19792–19799 (2002).
54. Colovos, C., Cascio, D. & Yeates, T. O. The 1.8 Å crystal structure of the ycaC gene product from *Escherichia coli* reveals an octameric hydrolase of unknown specificity. *Structure* **6,** 1329–1337 (1998).
55. Thaw, P. *et al.* Structural insight into gene transcriptional regulation and effector binding by the Lrp/AsnC family. *Nucleic Acids Res.* **34,** 1439–1449 (2006).
56. Totir, M. *et al.* Macro-to-micro structural proteomics: native source proteins for high-throughput crystallization. *PLoS ONE* **7,** e32498 (2012).
57. Auerbach, G. *et al.* Zinc plays a key role in human and bacterial GTP cyclohydrolase I. *Proc. Natl Acad. Sci. USA* **97,** 13567–13572 (2000).
58. Thorell, S., Schürmann, M., Sprenger, G. A. & Schneider, G. Crystal structure of decameric fructose-6-phosphate aldolase from *Escherichia coli* reveals inter-subunit helix swapping as a structural basis for assembly differences in the transaldolase family. *J. Mol. Biol.* **319,** 161–171 (2002).
59. von Delft, F. *et al.* Structure of *E. coli* ketopantoate hydroxymethyl transferase complexed with ketopantoate and Mg^{2+}, solved by locating 160 selenomethionine sites. *Structure* **11,** 985–996 (2003).
60. Andréll, J. *et al.* Crystal structure of the acid-induced arginine decarboxylase from *Escherichia coli*: reversible decamer assembly controls enzyme activity. *Biochemistry* **48,** 3915–3927 (2009).
61. Lee, K. H. *et al.* Crystal structures and enzyme mechanisms of a dual fucose mutarotase/ribose pyranase. *J. Mol. Biol.* **391,** 178–191 (2009).
62. Kanjee, U. *et al.* Linkage between the bacterial acid stress and stringent responses: the structure of the inducible lysine decarboxylase. *EMBO J.* **30,** 931–944 (2011).
63. Wang, M. *et al.* PaxDb, a database of protein abundance averages across all three domains of life. *Mol. Cell. Proteomics* **11,** 492–500 (2012).
64. Petrovska, I. *et al.* Filament formation by metabolic enzymes is a specific adaptation to an advanced state of cellular starvation. *eLife* (2014). Pages?
65. Landgraf, D., Okumus, B., Chien, P., Baker, T. A. & Paulsson, J. Segregation of molecules at cell division reveals native protein localization. *Nat. Methods* **9,** 480–482 (2012).
66. DeLano, W. L. The PyMOL molecular graphics system (2002).
67. Burley, S. K. & Petsko, G. A. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* **229,** 23–28 (1985).
68. Singh, J. & Thornton, J. The interaction between phenylalanine rings in proteins. *FEBS Lett.* **191,** 1–6 (1985).
69. Dougherty, D. A. Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science* **271,** 163–168 (1996).
70. Thomas, K. A., Smith, G. M., Thomas, T. B. & Feldmann, R. J. Electronic distributions within protein phenylalanine aromatic rings are reflected by the three-dimensional oxygen atom environments. *Proc. Natl Acad. Sci. USA* **79,** 4843–4847 (1982).
71. Claverie, P., Hofnung, M. & Monod, J. Sur certaines implications de l'hypothèse d'équivalence stricte entre les protomères des protéines oligomériques. *C.R. Acad. Sci. III* **266,** 1616–1618 (1968).
72. Lukatsky, D. B., Zeldovich, K. B. & Shakhnovich, E. I. Statistically enhanced self-attraction of random patterns. *Phys. Rev. Lett.* **97,** 178101 (2006).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



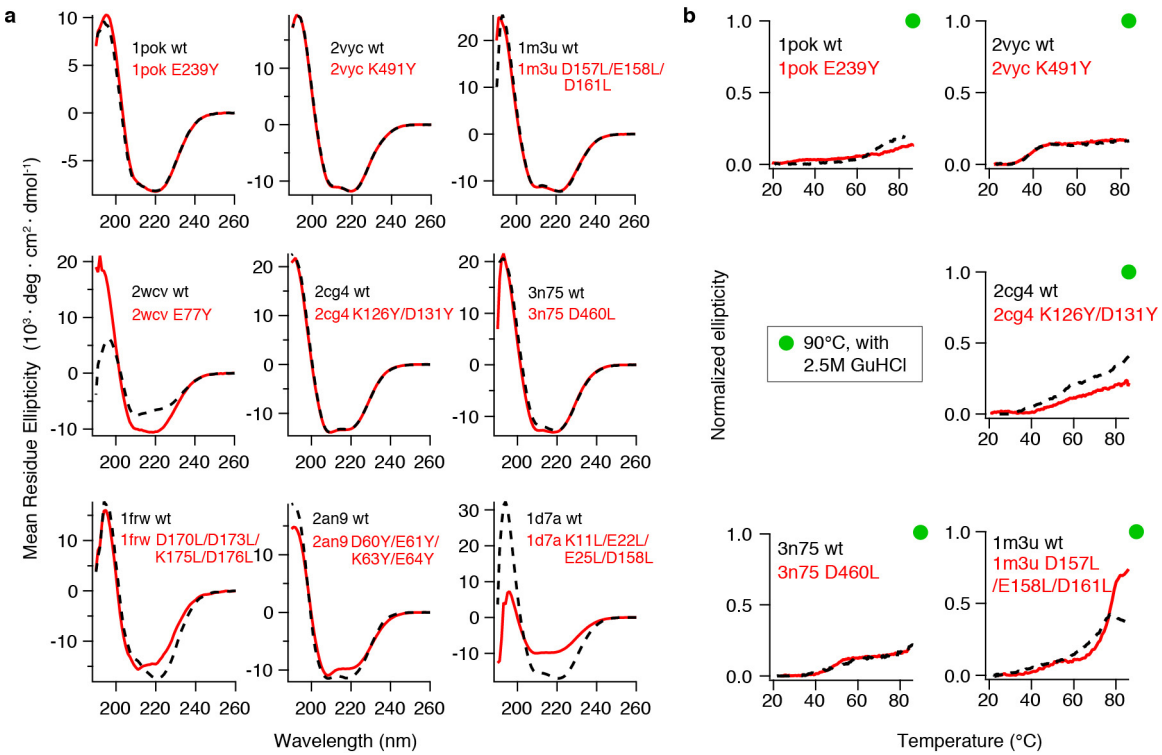
Extended Data Figure 1 | Increasing surface hydrophobicity of a homo-oligomer triggers its supramolecular self-assembly. a, Yeast cells expressing a fluorescently tagged homo-octameric dipeptidase from *E. coli* (PDB accession 1POK). The octamer can be viewed as two rings of four subunits each, stacked tail-to-tail. The localization of the wild-type protein is cytosolic and uniform, but a point mutant (E239Y) triggers head-to-head interactions between octamers and their stacking into micrometre-long fibres. **b**, Assembly is not mediated by interactions between YFP tags. It has been reported that fluorescent proteins can induce protein aggregation through dimerization[65]. In our work, none of the wild-type proteins aggregated despite their fusion to YFP, probably because we used the variant bearing the A206K mutation disrupting a weak tendency to dimerize. In addition, we performed a control experiment consisting of co-expressing an excess of the dipeptidase (PDB accession 1POK) untagged with YFP, together with a sub-stoichiometric quantity of the YFP-tagged subunit. As a result, most octamers will harbour zero or one YFP tag. In this context, the YFP allows monitoring the assembly but does not participate in multivalent interactions. **c**, Fluorescence microscopy revealed that fibres were forming in this context and were about tenfold less fluorescent than the case where all subunits were tagged with YFP. Together, these data indicate that YFP is not mediating fibre assembly *in vivo*.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



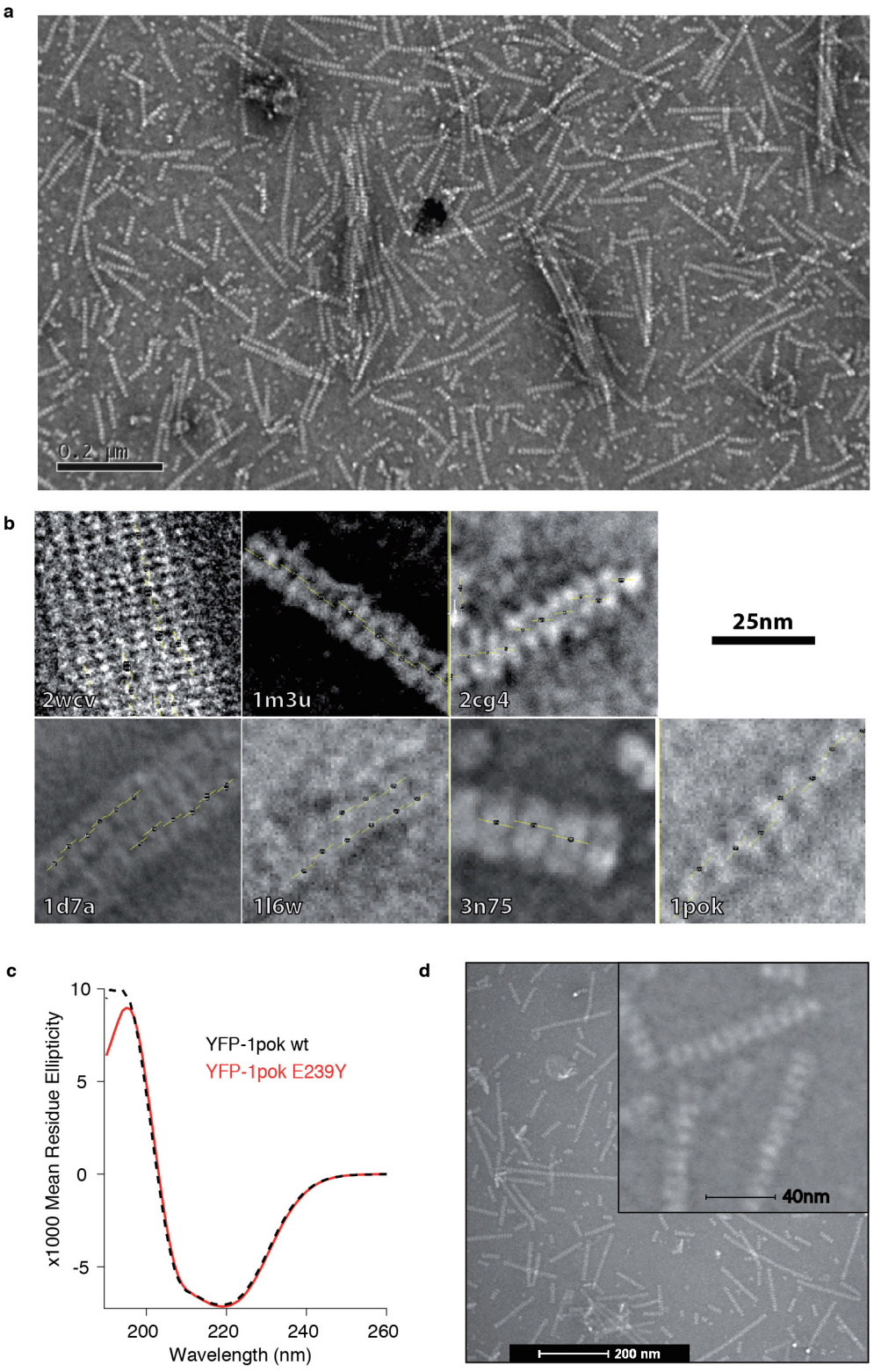
Extended Data Figure 2 | The proteins studied exhibit intracellular concentrations situated within a physiological range. a, Intracellular protein concentrations are estimated against reference solutions containing known concentrations of purified YFP. We transferred solutions of purified YFP into the same plate where we inoculated cells for imaging. This enabled us to relate the fluorescence emitted from YFP solutions to that emitted from cells. **b**, The signal measured by the confocal spinning disk microscope increases linearly with YFP concentration. The equation inferred from linear regression enabled us to convert fluorescence arbitrary units into YFP molarity (1 nM = 0.998 fluorescence arbitrary units − 0.16). **c**, We used the equation so obtained to transform the median intracellular fluorescence signal into homo-oligomer concentrations. Bars show the population median with associated standard error. We initially used the GPD promoter, which gave concentrations in the sub-micromolar range, and subsequently also used a CYC promoter to express four randomly chosen proteins. Expression with the CYC promoter gave concentrations in the range 10–50 nM, at which mutants also underwent supramolecular assembly (images are shown in Supplementary Table 2).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



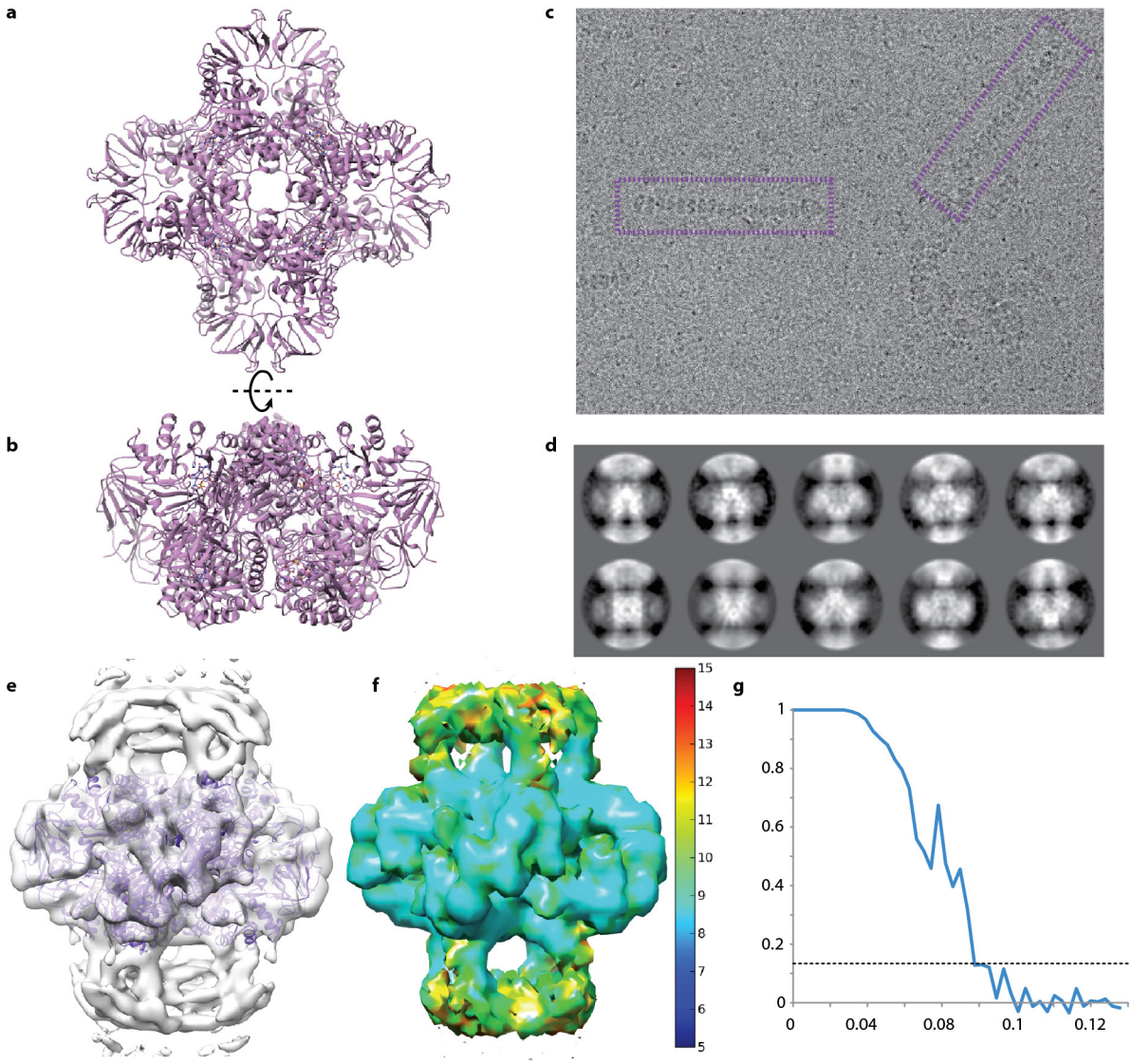
Extended Data Figure 3 | Circular dichroism spectra of all wild-type and mutant pairs studied. a, We examined the secondary structure content for nine protein pairs using far-ultraviolet circular dichroism. The spectra of wild-type proteins are displayed as black dashed lines and those of mutants as continuous red lines. The measurements were taken at 25 °C. Each pair was compared using the same buffer conditions (Supplementary Table 3). Most of the mutants exhibited similar or identical circular dichroism profiles when compared with their wild-type counterpart, indicating that the content in secondary structure was identical or similar between them. Only the mutants 2WCV (E77Y), 1FRW (D170L/D173L/K175L/D176L) and 1D7A (K11L/E22L/E25L/D158L) showed major differences, but all showed a spectrum with negative ellipticity values in the 210–230 nm range, whereas positive values are expected for random coil. These data indicate that mutants retain a folded structure. **b**, Stability measurement curves of five wild-type and mutant pairs forming fibres *in vivo*. Protein stability was assessed for five pairs following the ellipticity at 220 nm from 20 to 85 °C at a heating rate of 1 °C min^{-1}. Wild-type proteins are displayed as black dashed lines and those of mutants as continuous red lines. Each pair was compared using the same buffer conditions (Supplementary Table 3). None of the proteins fully unfolded in the temperature range probed. Thus, we measured the ellipticity of the samples in 2.5 M guanidinium chloride at 90 °C, which was taken as a relative unfolded state (maximal ellipticity, θ_{max}). The ellipticity of the samples at 20 °C was taken as a relative folded state (minimal ellipticity, θ_{min}). We show the normalized ellipticity (θ_{norm}) defined as $\theta_{norm} = (\theta_T - \theta_{min})/(\theta_{max} - \theta_{min})$, where θ_T is the ellipticity measured at temperature T.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



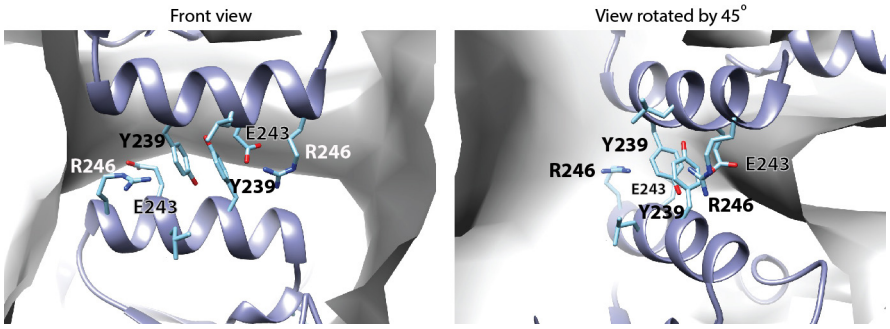
Extended Data Figure 4 | Characterization of protein fibres by electron microscopy. a, Dipeptidase mutant (1POK E239Y) visualized by TEM with negative staining. The protein buffer was Tris 20 mM, 100 mM NaCl, pH 7.5. The protein mutant self-assembles into filaments that tend to bundle together. **b**, Example of electron microscopy images on the basis of which the distance separating adjacent homomers in fibres was measured. Mutants in the images are 1D7A (K11L/E22L/E25L/D158L), 2WCV (E77Y), 1L6W (K97Y/K100Y/E102Y), 1M3U (D157L/E158L/D161L), 3N75 (D460L), 2CG4 (K126Y/D131Y), and 1POK (E239Y). All the samples were in Tris 20 mM, pH 7.5 except 1POK (E239Y) which had in addition 100 mM NaCl. **c**, Fusion of YFP to the dipeptidase mutant (1POK E239Y) does not affect its structure when compared with the wild-type fusion, as seen by circular dichroism. The protein buffer was Tris 20 mM, pH 7.5. **d**, We examined the dipeptidase fibre-forming mutant fused to YFP by TEM with negative straining. The protein buffer was Tris 20 mM, 100 mM NaCl, pH 7.5. The mutant forms filaments similar to those observed without the YFP fusion.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



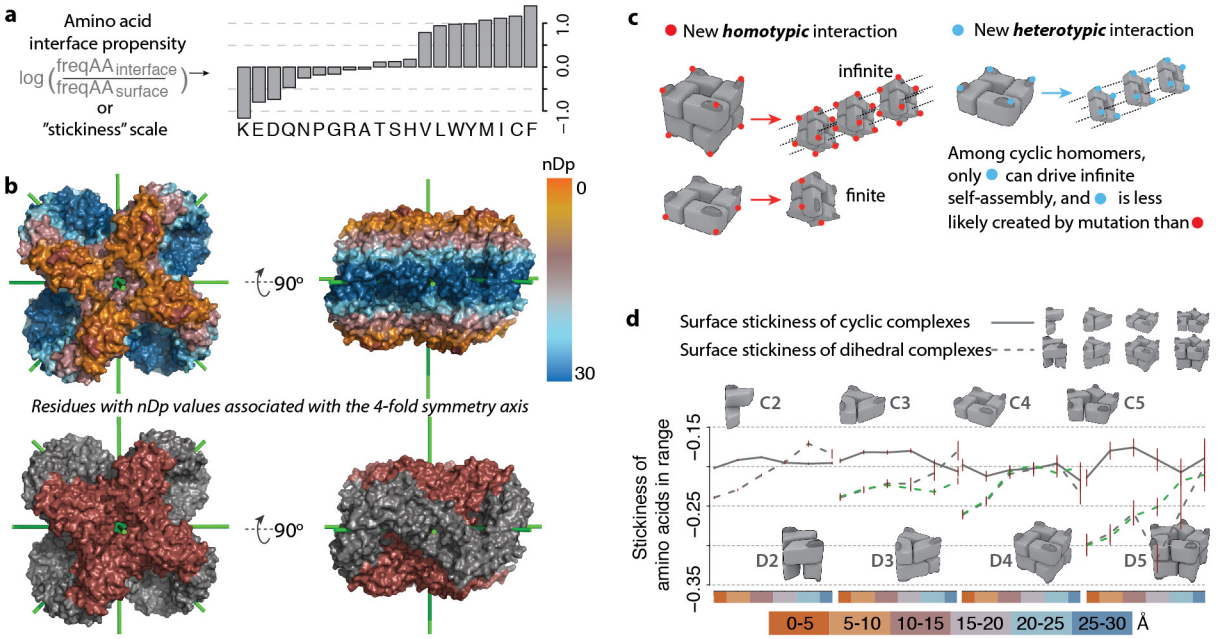
Extended Data Figure 5 | Single-particle three-dimensional reconstruction of the filament formed by the dipeptidase. a, Structure of the wild-type, octameric enzyme (1POK). **b**, Structure is rotated along the horizontal axis to show the orientation seen in filaments. **c**, Representative cryo-electron microscopy image after whole-image motion correction, showing filaments. **d**, Two-dimensional class averages. **e**, Final three-dimensional reconstruction. The wild-type structure solved by X-ray crystallography was fitted as a rigid body inside the map, using separate fits for the two tetrameric rings. **f**, Local resolution map (ResMap) in ångströms. **g**, Fourier shell correlation for the final reconstruction. The horizontal dotted line indicates the Fourier shell correlation = 0.143 criterion.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



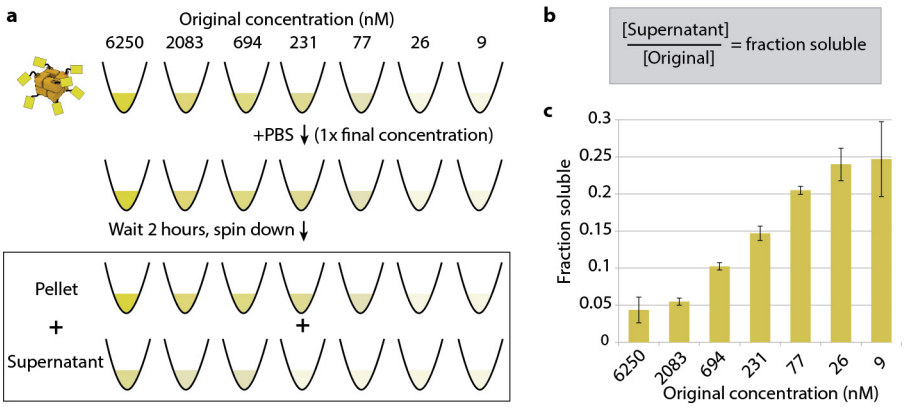
Extended Data Figure 6 | Detailed view of the helix–helix interface driving the formation of the dipeptidase fibre. After automated fitting of the atomic coordinates in the electron microscopy density (Extended Data Fig. 5e), we modelled the tyrosine side chain using PYMOL[66] and chose the most frequent rotamer. We also translated each helix by 1.15 Å in opposite directions to resolve a steric clash created between the two tyrosines. All other side chains remained unchanged. The resulting model suggests at least three possible types of interaction between side chains, which can help explain how the interface is stabilized. First, a 90° rotation of one tyrosine side chain would enable an aromatic–aromatic interaction, between the negatively charged centre of one tyrosine and the positively charged rim of its neighbour[67,68]. Second, the arginine of one octamer (R246) could establish a cation–π interaction with the tyrosine of an adjacent octamer[69]. Finally, the glutamic acid of one octamer (E243) could form an anion–aromatic interaction with the rim of an adjacent tyrosine[70].

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 7 | Stickiness is tuned as a function of nDp in dihedral complexes but not in cyclic complexes. a, The stickiness of an amino acid is defined by the log-ratio of its frequency at protein interfaces relative to protein surfaces. Thus, sticky residues are those enriched at protein interfaces relative to protein surfaces. Stickiness shows strong similarity to hydrophobicity, but also notable differences[22]. In this analysis, we use stickiness as a measure of 'interaction propensity' of surface patches. **b**, The structure of the dipeptidase (PDB 1POK) is coloured according to nDp calculated with respect to the four-fold axis of symmetry. Structural analyses presented in Fig. 4 consider residues associated with symmetry axes along which fibres can grow, meaning that their nDp values are lowest with respect to those axes. This notion is illustrated with the lower structures, where nDp values of red-coloured residues are smaller with respect to the four-fold axis, while nDp values of residues coloured in grey are smaller with respect to two-fold axes. This strategy enables measuring of negative design along three-, four-, or five-fold axes, while eliminating potential confounding effects due to the two-fold axes. **c**, Among cyclic complexes, only heterotypic interactions can trigger the formation of infinite fibres[71], but such interactions are less likely to form by mutation than by homotypic interactions[1,3,4,72]. **d**, We observed that mutations at the 'tip' of dihedral complexes are more likely to trigger the formation of supramolecular assemblies than mutations situated farther from the tip, where nDp is larger. Accordingly, we found that stickiness is tuned according to that distance, with regions at greater risk (orange bins, *x* axis) being associated with lower stickiness (*y* axis). Lines show the median stickiness of surface patches in any given nDp window and dark-red-coloured error bars correspond to two standard errors. The green dashed line is based on an alternative measure where all residues are counted, irrespective of their distance to two-fold symmetry axes (see Methods). Both measures show that stickiness is tuned as a function of nDp in dihedral homomers. Interestingly, however, we do not observe such tuning in cyclic homomers.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 8 | Self-assembly takes place at low concentrations *in vitro*. **a**, We assessed whether self-assembly of the dipeptidase mutant takes place at low concentrations. We made serial dilutions of the dipeptidase fused to YFP and then induced self-assembly by addition of 10× PBS to reach a final concentration of 1× PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4). We then incubated the samples for 2 h at 30 °C, spun them down, and pipetted 10 μl of the supernatant to analyse its protein concentration by fluorescence microscopy. **b**, The concentration of protein in the supernatant relative to the original concentration gave us the fraction of soluble protein. **c**, The fraction of soluble protein was situated between 0.04 and 0.25 at all concentrations studied, indicating that self-assembly does occur at concentrations as low as 9 nM and probably lower. Error bars span two standard errors and were calculated on the basis of four replicates.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Protein structures originally selected for mutation

PDB code	Reference	Number of subunits	Symmetry	Used in the study	Cellular abundance (ppm)	Uniprot ID
2an9	49	6	D3	YES	91.6	P60546
2gs4	50	6	D3	NO	387	P21362
1d7a	51	8	D4	YES	345	P0AG18
1frw	52	8	D4	YES	13.2	P32173
1pok	45	8	D4	YES	36.6	P39377
1l6y	53	8	D4	NO	120	P0ACB2
1yac	54	8	D4	YES	660	P21367
2cg4	55	8	D4	YES	66.9	P0ACI6
3n6q	56	8	D4	NO	290	Q46851
1fbx	57	10	D5	NO	723	P0A6T5
1l6w	58	10	D5	YES	82.8	P78055
1m3u	59	10	D5	YES	721	P31057
2iv1	***	10	D5	YES	0.398	P00816
2vyc	60	10	D5	YES	2.96	P28629
2wcv	61	10	D5	YES	58.3	P0AEN8
3n75	62	10	D5	YES	8.98	P0A9H3

The cellular abundance of the proteins studied[45,49–62] covers a wide range, from 0.4 p.p.m. to over 700 p.p.m. (ref. 63). Four structures out of the 16 were not considered further for reasons detailed in Supplementary Table 1.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | Size and assembly characteristics of the fibres observed *in vitro*

PDB code	Phenotype *in vivo*	Size of filaments *in vitro* (nm)		Assembly type *in vitro*
		Median	Max	
3n75	Fiber	45	162	Individual filaments
2wcv	Fiber	195	451	Large bundles of filaments
1pok	Fiber	213	514	Individual/bundles of filaments
1m3u	Fiber	187	1177	Pairs of twisted filaments
2cg4	Fiber	40	71	Individual filaments
1d7a	Foci	174	715	Individual/pairs of filaments
1l6w	Foci	185	464	Individual filaments
2vyc	Fiber	n/a	n/a	No filament observed *in vitro*

The filaments observed *in vivo* are larger than those observed *in vitro* and tend to bundle together into a single large fibre. Formation of fibres *in vivo* may be favoured owing to macromolecular crowding, as was proposed previously[64].

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.