We show here that computer game players can build high-quality crystal structures. Introduction of a new feature into the computer game Foldit allows players to build and real-space refine structures into electron density maps. To assess the usefulness of this feature, we held a crystallographic model-building competition between trained crystallographers, undergraduate students, Foldit players and automatic model-building algorithms. After removal of disordered residues, a team of Foldit players achieved the most accurate structure. Analysing the target protein of the competition, YPL067C, uncovered a new family of histidine triad proteins apparently involved in the prevention of amyloid toxicity. From this study, we conclude that crystallographers can utilize crowdsourcing to interpret electron density information and to produce structure solutions of the highest quality.
Molecular refinement against high-resolution data is never finished, only abandoned. George Sheldrick's statement on the labour-intensive nature of model building and refining crystal structures reflects the difficulty in producing highly accurate models. As a result, ~85% of deposited protein crystal structures contain discernable errors. Unfortunately, as crystal structures are frequently used as the basis of further studies, inaccurate crystal structures can cause significant harm to the scientific process. Continued improvement of crystal structure accuracy therefore remains an important goal within the biology community.

Recently, in a class assignment, we asked 57 undergraduate students to build the structure of a protein, lectin scytovirin, using only the model-building program Coot and an electron density map downloaded from the electron density server. Students were not given the amino-acid sequence of the protein, but were provided with the position of the N-terminal amino acid. The students were instructed to build the structure of the protein, residue by residue, into the 2Fo−Fo electron density map. Many students expressed appreciation for the puzzle-like quality of the assignment. In addition to learning about protein structure, ~10% of these students improved on the previously published model. One student even generated a structure that ranked in the 100th percentile in both Molprobity clashscore and total score when compared with other structures in its resolution range. These results raised the intriguing possibility that even a relatively small group of amateur model builders could collectively build higher-quality models than a single trained crystallographer. This concept was remarkably reminiscent of ideas recently championed by the online protein-folding computer game Foldit.

On the basis of the success of the undergraduate students in improving a published crystal structure, we hypothesized that crystallographic model building through crowdsourcing might result in more accurate crystal structures than those resulting from traditional model-building methods. Thus, we added electron density features to Foldit to allow players to build directly into density. We then administered a competition to determine if crowdsourcing crystal structures could lead to top-notch structural models. The competition showed that Foldit players could build very high-quality crystal structures into electron density maps, opening up a new method for building and refining crystal structures. In addition, solving the structure of the competition’s test protein unexpectedly led to the discovery a new family of histidine triad (HIT) proteins potentially involved in preventing amyloid fibre formation.

Results

Adding electron density to Foldit. Foldit is a popular video game that crowdsources protein structure prediction, challenging players to discover low-energy protein models by exploring protein conformational space. The newest version of Foldit provides players with electron density maps and the associated protein sequences, and asks players to use experimental data to guide protein folding (Fig. 1; Supplementary Fig. 1). With these new features, players can trim maps around a model and customize features of the electron density map, such as contour level, rendering style and transparency (Supplementary Fig. 1). The standard Foldit score function is supplemented with a fit-to-density term, allowing in-game structure minimization similar to crystallographic real-space refinement. Foldit players are able to view the fit-to-density score for each residue of a model, providing valuable feedback about specific parts of a model that require more attention. As a preliminary test of this feature, we gave an electron density puzzle to the Foldit players that was nearly identical to the lectin scytovirin classroom assignment mentioned above, and found that the Foldit players were also able to improve on the published scytovirin structure (Supplementary Note 1; Supplementary Fig. 2).

The model-building competition. We then held a crystallographic model-building competition to compare the effectiveness of different model-building approaches. Five groups of competitors took part in our model-building competition: (1) 469 Foldit players worldwide, (2) two trained crystallographers, (3) 61 undergraduate students in the University of Michigan class MCD411 (Introduction to Protein Structure and Function) who built the structure as a class assignment, (4) Phenix Autosolve and (5) MR-Rosetta. We chose YPL067C, a yeast protein with no significant sequence similarity to any structure in the Protein Data Bank (PDB) (Supplementary Note 2) as the target for our competition. In addition, YPL067C was chosen for its biological interest, as previous studies suggested that YPL067C is involved in preventing amyloid toxicity. Crystals of this protein diffracted to 1.9 Å resolution (Table 1). We asked all human competitors to build the best possible protein structure that they could given the protein sequence, a secondary structure prediction and an experimentally phased, density refined map of YPL067C. The MCD411 class assignment was conducted similarly to the previous crystallography assignment discussed above, in which undergraduates improved on a published crystal structure. As before, the undergraduates lacked previous model-building experience. In contrast, 54% of the participating Foldit players had attempted to solve early electron density puzzles in Foldit and thus had some experience in Foldit-based model building. Both the trained crystallographers and undergraduates used the model-building and real-space refinement program Coot, whereas the Foldit players used the Foldit version released on 14 October 2015. None of the competitors were given any starting points for building, and thus they needed to establish the relationship between the electron density and the protein’s sequence.

The various groups used different approaches. The students and trained crystallographers worked independently, generally utilizing large aromatic residues to identify the relationship between sequence and electron density features. The best Foldit solutions, in contrast, came from a group of players working collaboratively, with one player serving as the trailblazer who contributed the majority of the moves towards the creation of the model (Fig. 1), and other players providing detailed structural tweaks and refinements (Supplementary Movie 1). Although the Foldit players also used a large aromatic residue to initially anchor the sequence similar to the Coot users, their building process sampled conformational space very widely (Supplementary Movie 1), unlike the Coot users. The Foldit players used many different types of building tools, including constraint changes (band, freeze and cut actions), automated minimization algorithms (global wiggle, local wiggle and shake actions) and tools to modify secondary structure (rebuild and tweak actions) (Supplementary Fig. 3).

To determine which group produced the best structural models, we used an automated refinement procedure on all the structures, and then compared key crystallographic statistics. These statistics consisted of Rfree, r.m.s.d.'s of bonds and angles, the number and severity of steric clashes (represented by Molprobity clashscore), and Ramachandran outliers. Analysing the Rfree values, we quickly realized that the Foldit players were at a distinct disadvantage (Supplementary Fig. 4). Whereas the Coot users, Phenix Autosolve and MR-Rosetta were able to exclude regions that had poor quality electron density, the current version of Foldit required that players model the entire
sequence without gaps, retaining all disordered residues. Because disordered atoms do not substantially contribute to X-ray diffraction, modelling of disordered residues will disagree with diffraction data and inflate $R_{\text{free}}$ values. To correct for this deficiency in Foldit, we pruned the Foldit models afterwards to include only those residues modelled by the top trained crystallographer, so that they contained only the well-ordered regions of the protein. We then re-refined the pruned Foldit structures and compared them with the other models.

The Foldit structures improved considerably after pruning. As a result, the top pruned Foldit structure was the overall highest-quality structure produced in the competition (Fig. 2) as measured by geometry, density fit and steric clashes. In addition to the $R_{\text{free}}$ value of the pruned Foldit structure becoming marginally better than that achieved by the trained crystallographers and containing zero Ramachandran outliers, the Foldit structure had the lowest level of steric clashes (Fig. 2). According to Molprobity\textsuperscript{13}, the top Foldit structure is an exceptional model, ranking in the 100th percentile in both its overall Molprobity score and clashscore of all structures in the PDB of similar resolution ($1.95 \pm 0.25\, \text{Å}$). The superiority of the top Foldit structure can be attributed to better side-chain conformations than those in the top structure produced by the trained crystallographers (Supplementary Fig. 5). Better Foldit scores were associated with lower $R_{\text{free}}$ values (Supplementary Fig. 6), suggesting that the Foldit model-building strategy and its scoring algorithm could be generalizable as a means of producing high-quality structures.

**Discussion**

Here we show that Foldit players can build structural models at least as effectively as trained crystallographers and state-of-the-art automated methods, enabling a novel crowd-powered strategy for solving high-accuracy crystal structures. Combined with the ability to generate molecular replacement solutions in Foldit, and therefore to circumvent the need for experimental phases in some cases\textsuperscript{14}, it is now potentially possible to obtain complete structure solutions using Foldit given only a native crystallography data set. Citizens hold a tremendous reserve of brainpower that remains largely untapped by the scientific community. The new Foldit electron density feature has revealed that non-expert citizen scientists are capable of using structural data to build first-rate models, and we expect that Foldit will be a powerful tool for crowdsourcing many new high-quality structures.

This surprising win by Foldit players suggests that, in at least some cases, this video game can help produce crystallographic models of higher quality than those from trained crystallographers or automated model-building algorithms alone. The difference in accuracy is likely in part the result of different...
primarily uses a real-space refinement system\(^1\) that only respects underlying philosophies behind Coot and Foldit. Whereas Coot moving to more difficult de novo allowing players to practice their model-building skills before could form a base of easier Foldit puzzles for new players, accurately structural models. Editing of nearly complete structures would benefit the entire community of scientists that depend on of questionable quality already in the PDB. These Foldit puzzles automated structure solution algorithms.

Looking forward, we hope that further analysis of electron density or by submitting their crystal refinement problems to Foldit. achieving a similar effect by either having multiple laboratory building process used by the Foldit players could also be a is clearly helpful, as both Phenix Autosolve and MR-Rosetta on benefit from the Foldit force field for real-space refinement. SeMet, selenenomethionine.

Table 1 | Crystallography statistics for HTC1.

| SeMet HTC1 (top pruned Foldit) | Native HTC1 |
|--------------------------------|-------------|
| **Data collection** | | |
| Wavelength (Å) | 0.9876 | 0.97851 |
| Space group | P43212 | P43212 |
| Cell dimensions | | |
| a, b, c (Å) | 63.3, 63.3, 117.8 | 62.5, 62.5, 117.6 |
| α, β, γ (°) | 90, 90, 90 | 90, 90, 90 |
| Resolution (Å) | 50-1.95 (1.98-1.95) | 42.81-1.83 (1.89-1.83) |
| R sărg (%) | 0.077 (0.940) | 0.074 (0.668) |
| I/α1 | 59.5 (1.8) | 12.7 (2.0) |
| Completeness (%) | 99 (89) | 100 (100) |
| Redundancy | 12.4 (7.5) | 7.8 (7.7) |
| Figure of merit | 0.31 | |
| CIF/2 | 0.998 (0.916) | 0.998 (0.873) |
| **Refinement** | | |
| Resolution (Å) | 1.95 | 1.83 |
| No. of reflections | 18,107 | 21,161 |
| Rwork/Rfree | 0.26/0.28 | 0.20/0.25 |
| No. of non-hydrogen atoms | 1,343 | 1,663 |
| Protein | 1,305 | 1,513 |
| Ligand/ion | 0 | 12 |
| Water | 38 | 138 |
| Average B-factors | 53.8 | 48.8 |
| Protein | 53.8 | 49.0 |
| Ligand/ion | 51.9 | |
| Water | 54.3 | 45.7 |
| R.m.s.d's | 0.008 | 0.009 |
| Bond lengths (Å) | 1.0 | 0.89 |

SeMet, selenenomethionine.

underlying philosophies behind Coot and Foldit. Whereas Coot primarily uses a real-space refinement system\(^1\) that only respects local geometry, the Rosetta force field used by Foldit is much more extensive, including additional steric, electrostatic and solvation terms, as well as statistical potentials based on observed backbone torsions and side-chain rotamers. That some of the Foldit models were of higher quality than those of trained crystallographers suggests that expert model builders might also benefit from the Foldit force field for real-space refinement. Human intervention either by crystallographers or Foldit players is clearly helpful, as both Phenix Autosolve and MR-Rosetta on their own produced suboptimal structures. The collaborative building process used by the Foldit players could also be a beneficial strategy for professional crystallographers, who could achieve a similar effect by either having multiple laboratory members take turns working on model building and refinement or by submitting their crystal refinement problems to Foldit. Looking forward, we hope that further analysis of electron density puzzle solutions in Foldit can inform continued improvement of automated structure solution algorithms.

Foldit players might also be tasked with improving structures of questionable quality already in the PDB. These Foldit puzzles would benefit the entire community of scientists that depend on accurate structural models. Editing of nearly complete structures could form a base of easier Foldit puzzles for new players, allowing players to practice their model-building skills before moving to more difficult de novo model-building puzzles. To further improve the capability for Foldit players to aid crystallographers, ongoing development will make it possible for the players to add or remove residues with insufficient electron density, and have these changes accurately reflected in the Foldit score. We envision a future in which professional crystallographers frequently tap the collective model-building expertise of Foldit players for help in the model-building, refinement and validation steps of crystallography.

From an educational perspective, the participation of an undergraduate class in this study explored how crystallographic model building can be used not just to teach students the structures and chemistry of proteins in great depth, but in addition, to teach the scientific process. In our previous study, students built into a high-resolution, fully refined map\(^9\). In the study presented here, students received a lower-resolution, unrefined map and no starting place for building. Dealing with disordered residues, the students were forced to interpret data of varying quality and to decide when the data became too ambiguous to draw firm conclusions. Similar decision-making processes govern the use of scientific data across many disciplines. This assignment thus helped give students a realistic view about the power and limitations of the scientific method. Importantly, the study with which students and Foldit players were able to interpret and understand density maps suggests that scientists other than crystallographers can very readily interpret electron density maps, which will assist them in designing or analysing experiments based on crystal structures.

YPLO67C's structure yielded unexpected insights into its biological function. Despite the lack of sequence homology to any protein in the PDB, a DALI\(^1\) search of YPLO67C (Supplementary Table 1) found that it is structurally similar to members of the widely conserved superfamily of histidine triad (HIT) proteins. These proteins contain three histidine residues with an almost identical spatial organization to that of YPLO67C, as well as a β-sheet core nested inside a similar arrangement of loops and helices (Fig. 3). HIT proteins have been shown to be involved in diverse cellular stress responses, such as DNA damage, oxidative stress and induced apoptosis. However, the specific in vivo activity of HIT proteins remains unclear. Although YPLO67C bears some resemblance to known HIT proteins, it is sequentially and structurally distinct (Supplementary Fig. 7). Its most notable distinguishing structural characteristic is an open channel not found in other HIT proteins (Supplementary Note 2). YPLO67C's characterization makes it the founding member of a new family we are calling HTC (for histidine triad with channel), with YPLO67C being the first member, HTC1. The HTC family contains over 900 members found in a wide variety of eukaryotes and viruses (Supplementary Note 2). As mentioned above, HTC1 null mutants increase the toxicity of amyloid overproduction. We find here that HTC1 is very effective in preventing in vitro amyloid formation of three model proteins, β\(_{1-40}\), α-synuclein and reduced carboxy-methylated α-lactalbumin (RCMaA; Fig. 4). On the basis of docking simulations, HTC1 may bind to unfolded proteins using its conserved channel (Supplementary Note 2; Supplementary Fig. 8). Our crowdsourcing-enabled discovery of a new family of proteins involved in preventing amyloid formation provides insight into a novel physiological role of the ubiquitous HIT proteins.

Methods

Electron density in Foldit. To facilitate work on electron density data in Foldit, new visualizations and tools, along with a tutorial puzzle to introduce them, were developed and distributed to Foldit players in periodic software updates. Electron density maps in Foldit are displayed as a visual guide in the form of an isosurface. Players have control over parameters of the density isosurface, such as the contour level, surface texture, transparency and colour, and can tag regions of the density with notes. After initial testing, it was clear that density visualization alone was insufficient to improve model building by Foldit players. Players simply ignored the density, finding that their existing, familiar strategies were most competitive on Foldit leaderboards. In response, we adapted the Rosetta fit-to-density score term
elec_dens_fast into the Foldit score function\(^6\). This not only provides competitive element RMSD, bond RMSD, clash score, Ramachandran outliers and \(R_{\text{free}}\) for the best models from each group after pruning disordered residues.

**The competition.** Phenix Autosolve\(^7,8\), with model-building disabled, was used to create density-modified maps of selenomethionine (SeMet) YPL067C. To make the map manageable for the Foldit program, the map was masked beyond 5 Å from the initial solution at the start of the competition. This map was given to Foldit players, MCDB411 students and the experienced crystallographers for model building. The individual responsible for model building and refining the initial structure solution of YPL067C, before the contest was initiated, had no contact with any of the competitors.

Sixty-one students in the University of Michigan undergraduate class MCDB411 (Introduction to Protein Structure and Function) were introduced to structural alignment with the top DALI search hit is shown in Supplementary Fig. 7.

**Figure 3 | Overall structure of HTC1.** Structural alignment with the top DALI search hit is shown in Supplementary Fig. 7.
models were clustered into a set of 1,000 such that no two aligned to <1.0 Å Cα r.m.s.d. To this clustered set, we added the 50 best unique models produced by Foldit teams or soloists, as well as any models flagged by Foldit players for special consideration—1,094 Foldit models in total. The puzzle was open to Foldit players for 28 days. Members of the winning team began playing the day after the puzzle opened, and produced the winning structure ~23 days later—4 days before the puzzle closed.

Two trained crystallographers were given the same number of days for model building as the students and Foldit players. They were given specific instructions not to use tools outside of Coot or Molprobity and not to interact with each other during model building. The trained crystallographers spent ~8 and 14 h, respectively, working on the puzzle. The trained crystallographers reported using the following approach to the puzzle, which corresponds well with what the instructor observed with many of the undergraduate students. First, they looked for large density blobs that might correspond to large aromatic side chains such as Trp, Tyr or Phe. Working forwards and backwards from the Trp–Phe–Val–Asn sequence proved particularly useful. Modelling in a few of these large residues led to the assignment of density to sequence location. The direction of the polypeptide chain was reversed on a few occasions, but was fixed by looking at the carbonyl density. The Find Secondary Structure tool in Coot was used, especially for regions where the density was poor. Real Space Refine Zone was used with the refinement weight set to 20 or 10, based on the instructor’s suggestion for building in an unrefined map. Regions where the density was very poor and decisions had to be made about whether to keep trying to build or not proved to be the hardest part of the task. The trained crystallographers reported that at first they did build in these sections of poor electron density. However, when they realized the extent of the guessing involved, they subsequently removed most of the model in these areas. After modelling in the residues, the trained crystallographers used the validation tools in Coot, including Ramachandran plot, Rotamer analysis and Density Fit analysis, which flagged areas with poor geometry. They also ran the structure through MolProbity, which gave similar results to the Coot validation tools.

Finally, the crystallographers fixed problem areas as best as possible with the Coot modelling tools, such as Flip Peptide, Rotamers, Regularize Zone and Real Space Refine Zone. When asked to describe the difficulty level of this assignment, the trained crystallographers rated it as somewhat difficult (on a scale of: very difficult, somewhat difficult, neither easy nor difficult, somewhat easy and very easy).

Phenix Autosolve7,8 was run with default parameters (using phase and build) to produce the Autosolve model. The MR-Rosetta model was obtained by relaxing and rebuilding the Autosolve model in the same electron density map provided to the competition, such as Flip Peptide, Rotamers, Regularize Zone and Real Space Refine Zone. When asked to describe the difficulty level of this assignment, the trained crystallographers rated it as somewhat difficult (on a scale of: very difficult, somewhat difficult, neither easy nor difficult, somewhat easy and very easy).

Phenix Autosolve7,8 was run with default parameters (using phase and build) to produce the Autosolve model. The MR-Rosetta model was obtained by relaxing and rebuilding the Autosolve model in the same electron density map provided to the competition, such as Flip Peptide, Rotamers, Regularize Zone and Real Space Refine Zone. When asked to describe the difficulty level of this assignment, the trained crystallographers rated it as somewhat difficult (on a scale of: very difficult, somewhat difficult, neither easy nor difficult, somewhat easy and very easy).

Biological expression and purification. The gene for YPL067C was amplified from yeast strain Y2HGold (Clontech) and cloned into a pEET28-sumo plasmid using primer 1 (5′-AAATATGGATCCATGCAACAGATTGTCACGATCAC CAG-3′) and primer 2 (5′-AAATATCCGAGTCAGGAAATGGTGCAGA AAC C-3′). pEET28-sumo-ypl067C was transformed into Escherichia coli BL21(DE3) cells.

The expression and purification of SeMet YPL067C was performed with the same protocol except a methionine auxotroph variant of E. coli BL21(DE3) and SeleniumMethionine Medium Complete (Molecular Dimensions) were used. The gene for YPL067C was amplified from Clostridium difficile crystal structure of YPL067C was to a HIT protein of unknown function from Clostridium difficile (PDB: 4EGU), with a Z-score of 4.9. The top 47 hits were all HIT proteins with Z-scores ranging from 4.9 to 4.2 (Z-scores > 2.0 are considered significant). Secondary structure predictions for the competition were generated using PSIPRED22.

Figure 4 | HTCl aggregation inhibition. HTCl prevents amyloid formation of RCMaLa (a) β2,40 (b) and α-synuclein (c), as measured by thioflavin T (THT) fluorescence at 490 nm.

Protein expression and purification. The gene for YPL067C was amplified from yeast strain Y2HGold (Clontech) and cloned into a pEET28-sumo plasmid using primer 1 (5′-AAATATGGATCCATGCAACAGATTGTCACGATCAC CAG-3′) and primer 2 (5′-AAATATCCGAGTCAGGAAATGGTGCAGA AAC C-3′). pEET28-sumo-ypl067C was transformed into Escherichia coli BL21(DE3) cells.

Cells were grown at 37 °C overnight in 100 ml Luria Broth (containing 100 µg ml−1 kanamycin), and 10 ml was inoculated to 1 litre Broth (containing 100 µg ml−1 kanamycin). At early log phase, the temperature was reduced to 20 °C and 0.1 mM isopropyl β-D-1-thiogalactopyranoside was added to induce expression overnight. Cells were collected by centrifugation and resuspended in 100 ml lysis buffer (40 mM Tris, 10 mM sodium phosphate, 400 mM NaCl, 10% glycerol, 10 mM imidazole, pH 8.0) enriched with 1 mg ml−1 D Nasel, 1 mM MgCl2 and two tablets of complete EDTA-free protease inhibitor (Roche). Cells were lysed by two French press cycles at 1,300 s.p.s. and centrifuged at 37,000 g for 30 min at 4 °C. The supernatant was run through a Ni-HisTrap 5 ml column (GE Healthcare) pre-equilibrated with lysis buffer at a rate of 1.5 ml min−1. Following binding, the column was washed with 60 ml lysis buffer. The protein was eluted with 20 ml lysis buffer enriched with 500 mM imidazole. To cleave the sumo-His × 6 tag, 10 µl ULP1 protease (from stock of 50 mg ml−1) was added to the eluted solution. A volume of 10 µl β-mercaptoethanol was added and the solution was dialysed overnight in 40 mM Tris, 10 mM sodium phosphate, 400 mM NaCl, 10% glycerol, pH 8.0. To remove the tag, the solution was run through a Ni-HisTrap 5 ml column (GE Healthcare) pre-equilibrated with dialysis buffer at a rate of 1.5 ml min−1, and the flowthrough was saved and diluted in eight volumes of 20 mM Tris, pH 8.0. The protein was then run through a HiTrap Q HP 5 ml column (GE Healthcare). The flowthrough containing ~95% pure YPL067C as measured by SDS–polyacrylamide gel electrophoresis. Before each experiment, YPL067C was exchanged into appropriate buffer. Expression and purification of SeMet YPL067C was performed with the same protocol except a methionine auxotroph variant of E. coli BL21(DE3) and SeleniumMethionine Medium Complete (Molecular Dimensions) were used.

α-Synuclein was expressed and purified using the protocol described previously23 with minor modifications. In brief, 1% of the overnight grown culture was transferred in fresh media and induced with 0.8 mM isopropyl β-D-1-thiogalactopyranoside for 4 h after the optical density of the culture reached 0.6. The induced cells were pelleted at 4,000 r.p.m. and resuspended in 25 ml lysis buffer (10 mM Tris, 1 mM EDTA, pH 8). The lysed cells were then boiled at 95 °C for 15–20 min and centrifuged at 11,000 r.p.m. for 20 min. The supernatant was thoroughly mixed with 10% streptomycin sulfate (136 µM ml−1) and glacial acetic acid (228 µM ml−1) then centrifuged at 11,000 r.p.m. for 30 min. To the clear supernatant, an equal volume of saturated ammonium sulfate was added, and the solution was incubated at 4 °C for 1 h with intermittent mixing. The precipitated protein was separated by centrifugation at 11,000 r.p.m. for 30 min. The pellet was dissolved in equal volumes of absolute ethanol (chilled) and 100 mM ammonium acetate. Finally, the pellet was washed (twice) optional with absolute ethanol, dried at room temperature and resuspended in 10 mM Tris, pH 7.4. The protein solution was filtered through a 50 kDa cutoff column (AMICON, Millipore) followed by ion-exchange chromatography (Q-sepharose) against a NaCl gradient.

The fractions of pure protein eluted at ~300 mM NaCl were checked on SDS–polyacrylamide gel electrophoresis and the molecular weight was confirmed by mass spectrometry. The pure fractions were pooled and dialysed overnight against buffer (10 mM Tris and 50 mM NaCl, pH 7.4). The concentration of α-synuclein was determined using ε208 = 5,600 M−1 cm−1. The purified α-synuclein was stored at ~80 °C at a concentration of ~100 µM until use.

α-peptide was purchased from AlexisTech AB (Umeå, Sweden) and prepared as previously described24. Aβ40 peptide was dissolved in 10 mM NaOH to a peptide concentration of 1 mg ml−1 and then sonicated for 1 min in an ice bath before dilution in the assay buffer. The preparations were kept on ice.
\( \alpha \)-Lactalbumin (\( \alpha \)LA) from bovine milk (cat: L6010) and porcine citrate synthase (cat: C2620-5KU) were purchased from Sigma Inc. RCMaLA was prepared as previously described\(^{25}\). \( \alpha \)LA (500 \( \mu \)M; freshly prepared in water) was incubated with 1 \( \mu \)M dihydrothreitol in 0.5 M Tris and 1 mM EDTA, pH 7.0, for 10 min, then 3 mM iodoacetic acid (out of 1 M stock solution in water) was added and the solution incubated for another 30 min. \( \alpha \)LA was then dialysed into 50 mM phosphate buffer, pH 7.0, 100 mM KCl, 10 mM MgCl\(_2\).

**Protein crystallization.** Native and SeMet YPL067C crystals were grown at 20°C by vapour diffusion using both sitting (1 \( \mu \)l drops) and hanging drop methods (2 \( \mu \)l drops). Drops were prepared by mixing a 1:1 solution of YPL067C (25 mg ml\(^{-1}\)) and reservoir solution (5.6-8.1% glycerol, 1.6-2.1 M ammonium sulfate and 0.1-0.2 M Tris). Crystals were cryoprotected by gradually supplementing the drop with glycerol up to 25% and were flash-frozen in liquid nitrogen.

**X-ray crystallography.** Data were collected at the Life Sciences Collaborative Access Team (LS-CAT) beamlines at the Argonne National Laboratory’s Advanced Photon Source at 100 K. The data were integrated and scaled using HKL2000. Phases and initial model building of the SeMet derivative were obtained using Phenix AutoSolve\(^{7,8}\). Native YPL067C was solved by molecular replacement with the initial SeMet structure. Iterative refinement and model building were performed using Phenix Refine\(^{17}\) and Coot\(^{1}\). Channel size was analysed using the 3V server\(^{26}\). Data collection and modelling statistics are shown in Table 1, A, a section of the structure in its 2mFo-DFc map shown in Supplementary Fig. 10.

**Fibrillar aggregation assays.** Fibrillar aggregation was monitored by a thioflavin T (ThT) fluorescence assay. ThT is a benzothiazole dye that exhibits enhanced fluorescence specifically binding to amyloid fibrils. For RCMaLA aggregation experiments, solutions containing 100 \( \mu \)M RCMaLA, YPL067C in varying concentrations and 20 \( \mu \)M ThT were prepared in 50 mM potassium phosphate buffer, pH 7.0, 100 mM KCl and 10 mM MgCl\(_2\) (ref. 25). The ThT fluorescence assays with Ap\(_{\alpha}\), peptide were performed with 2.5 \( \mu \)M Ap\(_{\alpha}\), peptide, YPL067C in varying concentrations and 20 \( \mu \)M ThT in PBS, pH 7.4, 1% dimethylsulfoxide. The fibrillar aggregation of \( \alpha \)-synuclein was tested in a solution of 70 \( \mu \)M \( \alpha \)-synuclein, YPL067C in desired concentrations and 20 \( \mu \)M ThT in PBS, pH 7.4. For \( \alpha \)-synuclein assays, four glass beads were added in each well to induce aggregation. ThT fluorescence assays were performed with a final volume of 100 \( \mu \)l of the prepared solution in black 96-microwell plates (costar, UV Plate, 96 well) that were sealed to prevent evaporation. ThT fluorescence was measured in a Synergy HT Multi-Mode Microplate Reader (Biotek) at 37°C with constant medium shaking. Excretion and emission wavelengths were 440 and 490 nm, respectively. All samples were assayed in triplicate and the assay was repeated twice. Incubation of YPL067C with ThT alone produced no fluorescence increase.

**Docking of \( \alpha \)-synuclein and HTC1.** HTC1 was docked against a 200-member NMR ensemble of \( \alpha \)-synuclein\(^{43}\) using ZDOCK 3.0.2 (ref. 28). The top five scoring poses of HTC1 bound to each member of the ensemble were used to generate a contact frequency map of the HTC1: \( \alpha \)-synuclein interaction. To determine the contact map, an interaction was assigned to a given residue pair if their Cz-Cz distance was less than or equal to 4. \( \alpha \)-synuclein, HTC1, were first dialysed into 50 mM potassium phosphate buffer, pH 7.0, 100 mM KCl and 10 mM MgCl\(_2\) and equilibrated to 22°C in the centrifuge for at least 1 h before sedimentation. All samples were spun at 48,000 r.p.m. in a Beckman AN-50 Ti rotor (167341.7g at the centre of the cell), and the sedimentation of the protein was monitored continuously using the interference optics. Data analysis was conducted with SEDFIT (version 14.1)\(^{30}\), using the continuous (c) distribution model. The confidence level for the ME (maximum entropy) regularization was set to 0.7. Buffer density and viscosity were calculated using SEDNTERP (http://sednterp.unh.edu/).

**Data availability.** The final model of HTC1 is deposited in the PDB under the code 5KCI. Other models and raw data are available on request.

**References**

1. Sheldrick, G. M. A short history of SHELX. Acta Crystallogr. A **64**, 112–122 (2008).

2. Touw, W. G., Joosten, R. P. & Friend, G. New biologic insights from better structure models. J Mol Biol **428**, 1375–1393 (2016).

3. Luo, J. H., Warmlander, Skt. S., Graslund, A. & Abrahams, J. P. Non-chaperone proteins can inhibit aggregation and cytotoxicity of Alzheimer amyloid beta peptide. J. Biol. Chem. **289**, 27766–27775 (2014).

4. Kulig, M. & Ecroyd, H. The small heat-shock protein alpha B-crystallin uses fibrillogenesis to control assembly of protein-protein complexes. Proc. Natl. Acad. Sci. **112**, 623–644 (2015).

5. Cooper, S. et al. Protein crystallization: automated crystallization with ZDOCK server: interactive docking prediction of protein-protein complexes. J Mol Biol. **448**, 343–352 (2015).

6. Voss, N. R. & Gerstein, M. 3V: cavity, channel and cleft volume calculator and extractor. Nucleic Acids Res. **38**, W555–W562 (2010).

7. Schwalbe, M. et al. Predictive atomic resolution descriptions of intrinsically disordered hTau40 and alpha-synuclein in solution from NMR and small angle scattering. Structure **22**, 238–249 (2014).

8. Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T. & Weng, Z. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. Bioinformatics **30**, 1771–1773 (2014).

9. Miyazawa, S. & Jerrigan, R. L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation of membrane folding. J. Mol. Biol. **256**, 623–644 (1995).

10. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamell model equation modeling. Biophys. J. **78**, 1606–1619 (2000).

**Acknowledgements**

We thank K. Wan for protein purification, Z. Wawrzak for phased advice, and U. Jakob for reading and editing the manuscript. This work was supported by National Institutes of Health R01GM084106 and R21AI104718.
of Health (NIH) grants GM102829 to J.C.A.B. and GM118651 to M.R.C. D.S.R. was supported by the University of Michigan Chemistry-Biology Interface training program (NIH grant 5T32GM008597) and the University of Michigan Medical Scientist Training Program (NIH grant 5T32GM007863). M.S. was partially supported by a predoctoral fellowship from the Cellular Biotechnology Training Program (T32GM008353). N.M.K. was supported by the University of Michigan Medical School Host Microbiome Initiative. A.S. was supported by the Molecular Biophysics Training Program (GM008270). B.K. acknowledges support by the National Science Foundation Graduate Research Fellowship Program (DGE-1256082). D.B. and J.C.A.B. are Howard Hughes Investigators. This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817). The authors would also like to thank all Foldit players and students who participated in the competition, but chose not to be listed as authors.

Author contributions
The competition was designed by F.K., S.H. and B.K., and was implemented by S.H., B.K., F.K., N.K., D.R., A.S., F.M. and M.S. Foldit design and improvements were carried out by J.F. and S.C. Experiments were designed by J.C.A.B., A.T., T.H., S.H., P.K. and R.M. Experiments were carried out by A.T., P.K., T.H., F.B. and R.M. N.J., and M.R.C. made assistance from all authors. U.M.S. and F.P. participated in the competition.

Foldit players
Ahmet Caglar18, Alan Coral18, Alice Elizabeth Jensen18, Allen Lubow18, Amanda Boitano18, Amy Elizabeth Lisle18, Andrew T. Maxwell18, Barb Failer18, Bartosz Kaszubowski18, Bohdan Hrytsiv18, Brancaccio Vincenzo18, Breno Renan de Melo Cruz18, Brian Joseph McManus18, Bruno Kestemont18, Carl Vardeman18, Casey Comisky18, Catherine Neilson18, Catherine R. Landers18, Christopher Ince18, Daniel Jon Buske18, Daniel Totonjian18, David Marshall Copeland18, David Murray18, Dawid Jagiela18, Dietmar Janz18, Douglas C. Wheeler18, Elie Cali18, Emmanuel Croze18, Farah Rezae18, Floyd Orville Martin18, Gil Beecher18, Guido Alexander de Jong18, Guy Ykman18, Harald Feldmann18, Hugo Paul Perez Chan18, Istvan Kovancz18, Ivan Vasilchenko18, James C. Connellan18, Jami Lynne Borman18, Jane Norrgard18, Jebbie Kanfer18, Jeffrey M. Canfield18, Jesse David Slone18, Jimmy Oh18, Joanne Mitchell18, John Bishop18, John Douglas Kroeger18, Jonas Schinkler18, Joseph McLaughlin18, June M. Brownlee18, Justin Bell18, Karl Willem Fellbaum18, Kathleen Harper18, Kirk J. Abbey18, Lennart E. Isaksson18, Linda Wei18, Lisa N. Cummins18, Lori Anne Miller18, Lyn Bain18, Lynn Carpenter18, Maarten Desnouck18, Manasa G. Sharma18, Marcus Belcastro18, Martin Szew18, Matthew Britton18, Matthias Gaebel18, Max Power18, Michael Cassidy18, Michael Pfützenreuter18, Michele Minett18, Michiel Wesselingh18, Minjune Yi18, Neil Haydn Tormey Cameron18, Nicholas I. Bolibruch18, Noah Benevides18, Norah Kathleen Kerr18, Nova Barlow18, Nykole Krystyne Crevits18, Paul Dunn18, Paolo Sergio Silveira Belo Nascimento Roque18, Peter Riber18, Petri Pikkanen18, Raafay Shehzad18, Randy Viosca18, Robert James Fraser18, Robert Leduc18, Roberto Dominguez Soto18, Roman Madala18, Scott Shinider18, Sharon de Boisblanc18, Slava Butkovich18, Spencer Bliven18, Stephen Hettler18, Stephen Telehany18, Steven A. Schwegmann18, Steven Parkes18, Susan C. Kleinfelter18, Sven Michael Holst18, T.J.A. van der Laan18, Thomas Bausewein18, Vera Simon18, Warwick Pulley18 & William Hull18

18 Rosetta Commons, University of Washington, Seattle, Washington 98195, USA.

University of Michigan students
Alexis Lawton1, Amanda Ruesch1, Anjali Sundar1, Anna-Lisa Lawrence1, Annes Yukyung Kim1, Antara Afrin1, Bhargavi Maheshwer1, Bilal Turfe1, Christian Huebner1, Courtney Elizabeth Killeen1, Dalia Antebi-Lerrman1, Danny Lu1, Derek Wolfe1, Duc Pham1, Elaina Michewicz1, Elizabeth Hull1, Emily Pardington1,
Galal Osama Galal¹, Grace Chen¹, Grace Sun¹, Halie E. Anderson¹, Jane Chang¹, Jeffrey Thomas Hewlett¹, Jennifer Sterbenz¹, Jiho Lim¹, Joshua Morof¹, Junho Lee¹, Juyoung Samuel Inn¹, Kaitlin Hahm¹, Kaitlin Roth¹, Karun Nair¹, Katherine Markin¹, Katie Schramm¹, Kevin Toni Eid¹, Kristina Gam¹, Lisha Murphy¹, Lucy Yuan¹, Lulia Kana¹, Lynn Daboul¹, Mario Karam Shammas¹, Max Chason¹, Moaz Sinan¹, Nicholas Andrew Tooley¹, Nisha Korakavi¹, Patrick Comer¹, Pragya Magur¹, Quresh Savliwala¹, Reid Michael Davison¹, Roshun Rajiv Sankaran¹, Sam Lee¹, Saule Tamkus¹, Shirley Chen¹, Sho Harvey¹, Sin Ye Hwang¹, Sohrab Vatsia¹, Stefan Withrow¹, Tahra K. Luther¹, Taylor Manett¹, Thomas James Johnson¹, Timothy Ryan Brash¹, Wyatt Kuhlman¹, Yeonjung Park¹
Corrigendum: Determining crystal structures through crowdsourcing and coursework

Scott Horowitz, Brian Koepnick, Raoul Martin, Agnes Tymieniecki, Amanda A. Winburn, Seth Cooper, Jeff Flatten, David S. Rogawski, Nicole M. Koropatkin, Tsinatkeab T. Hailu, Neha Jain, Philipp Koldewey, Logan S. Ahlstrom, Matthew R. Chapman, Andrew P. Sikkema, Meredith A. Skiba, Finn P. Maloney, Felix R.M. Beinlich, Foldit Players, University of Michigan students, Zoran Popović, David Baker, Firas Khatib & James C.A. Bardwell

Nature Communications 7:12549 doi: 10.1038/ncomms12549 (2016); Published 16 Sep 2016; Updated 25 Oct 2016

The original version of this Article contained an error in the spelling of a member of the University of Michigan students Consortium, Sam Lee, which was incorrectly given as Sam Lewe. This has now been corrected in both the PDF and HTML versions of the Article.

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/

© The Author(s) 2016
Foldit players
Ahmet Caglar, Alan Coral, Alice Elizabeth Jensen, Allen Lubow, Amanda Boitano, Amy Elizabeth Lisle, Andrew T. Maxwell, Barb Failer, Bartosz Kaszubowski, Bohdan Hrytsiv, Brancaccio Vincenzo, Breno Renan de Melo Cruz, Brian McManus, Bruno Kestemont, Carl Vandeman, Casey Comisky, Catherine Neilson, Catherine R. Landers, Christopher Ince, Daniel Jon Buske, Daniel Totonjian, David Marshall Copeland, David Murray, Dawid Jagieła, Dietmar Janz, Douglas C. Wheeler, Elie Cali, Emmanuel Croze, Farah Rezae, Floyd Orville Martin, Gil Beecher, Guido Alexander de Jong, Guy Ykman, Harald Feldmann, Hugo Paul Perez Chan, Istvan Kovaneicz, Ivan Vasilchenko, James C. Connellan, Jami Lynne Borman, Jane Norrgard, Jazzy Kanfer, Jeffrey M. Canfield, Jesse David Slone, Jimmy Oh, Joanne Mitchell, John Bishop, John Douglas Kroeger, Jonas Schinkler, Joseph McLaughlin, June M. Brownlee, Justin Bell, Karl Willem Fellbaum, Kathleen Harper, Kirk J. Abbey, Lennart E. Isaksson, Linda Wei, Lisa N. Cummins, Lori Anne Miller, Lyn Bain, Lynn Carpenter, Maarten Desnouck, Manasa G. Sharma, Marcus Belcastro, Martin Szew, Matthew Britton, Matthias Gaebel, Max Power, Michael Cassidy, Michael Pfützenreuter, Michele Minett, Michiel Wesselingh, Minjune Yi, Neil Haydn Tormey Cameron, Nicholas I. Bolibruch, Noah Benevides, Norah Kathleen Kerr, Nova Barlow, Nykole Krystyne Crevits, Paul Dunn, Paulo Sergio Silveira Belo Nascimento Roque, Peter Riber, Petri Pikkanen, Raafay Shehzad, Randy Viosca, Robert James Fraser, Robert Leduc, Roman Madala, Scott Shnider, Sharon de Boisblanc, Slava Butkovich, Spencer Bliven, Stephen Hettler, Stephen Telegany, Steven A. Schwengmann, Steven Parkes, Susan C. Kleinfelter, Sven Michael Holst, T.J.A. van der Laan, Thomas Bausewein, Vera Simon, Warwick Pulley & William Hull

University of Michigan students
Alexis Lawton, Amanda Ruesch, Anjali Sundar, Anna-Lisa Lawrence, Annes Yukyung Kim, Antara Afrin, Bhargavi Maheshwer, Bilal Turfe, Christian Huebner, Courtney Elizabeth Killeen, Dalia Antebi-Lerrman, Danny Lan, Derek Wolfe, Duc Pham, Elaina Michewicz, Elizabeth Hull, Emily Pardington, Galal Osama Galal, Grace Chen, Grace Sun, Halie E. Anderson, Jane Chang, Jeffrey Thomas Hewlett, Jennifer Sterbenz, Jiho Lim, Joshua Morof, Junho Lee, Juyoung Samuel Inn, Kaitlin Hahm, Kaitlin Roth, Karun Nair, Katherine Markin, Katie Schramm, Kevin Toni Eid, Kristina Gam, Lisha Murphy, Lucy Yuan, Lulia Kana, Lynn Daboul, Mario Karam Shammas, Max Chason, Moaz Sinan, Nicholas Andrew Tooley, Nisha Korakavi, Patrick Comer, Pragya Magur, Quresh Savliwala, Reid Michael Davison, Roshun Rajiv Sankaran, Sam Lee, Saule Tamkus, Shirley Chen, Sho Harvey, Sin Ye Hwang, Sohrab Vatsia, Stefan Withrow, Tahra K. Luther, Taylor Manett, Thomas James Johnson, Timothy Ryan Brash, Wyatt Kuhlman & Yeonjung Park