the FEBS Journal

The catalytic core of an archaeal 2-oxoacid dehydrogenase multienzyme complex is a 42-mer protein assembly

Nia L. Marrott[1], Jacqueline J. T. Marshall[2], Dmitri I. Svergun[3], Susan J. Crennell[1], David W. Hough[1], Michael J. Danson[1] and Jean M. H. van den Elsen[4]

1 Department of Biology and Biochemistry, Centre for Extremophile Research, University of Bath, UK
2 School of Biochemistry, University of Bristol, Medical Sciences Building, UK
3 EMBL, Hamburg Outstation, Germany
4 Department of Biology and Biochemistry, University of Bath, UK

Keywords
archaea; macromolecular assembly; multienzyme complex; thermophile; X-ray crystallography

Correspondence
Michael J. Danson, Department of Biology and Biochemistry, Centre for Extremophile Research, University of Bath, Bath BA2 7AY, UK
Fax: +44 1225 386779
Tel: +44 1225 386509
E-mail: m.j.danson@bath.ac.uk
Jean M. H. van den Elsen, Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK
Fax: +44 1225 386779
Tel: +44 1225 383639
E-mail: j.m.h.v.elsen@bath.ac.uk

(Received 16 September 2011, revised 6 December 2011, accepted 12 December 2011)

doi:10.1111/j.1742-4658.2011.08461.x

The dihydrolipoyl acyl-transferase (E2) enzyme forms the structural and catalytic core of the tripartite 2-oxoacid dehydrogenase multienzyme complexes of the central metabolic pathways. Although this family of multienzyme complexes shares a common architecture, their E2 cores form homo-trimers that, depending on the source, further associate into either octahedral (24-mer) or icosahedral (60-mer) assemblies, as predicted by the principles of quasi-equivalence. In the crystal structure of the E2 core from *Thermoplasma acidophilum*, a thermophilic archaeon, the homo-trimers assemble into a unique 42-mer oblate spheroid. Analytical equilibrium centrifugation and small-angle X-ray scattering analyses confirm that this catalytically active 1.08 MDa assembly exists as a single species in solution, forming a hollow spheroid with a maximum diameter of 220 Å. In this paper we show that a monodisperse macromolecular assembly, built from identical subunits in non-identical environments, forms an irregular protein shell via non-equivalent interactions. This unusually irregular protein shell, combining cubic and dodecahedral geometrical elements, expands on the concept of quasi-equivalence as a basis for understanding macromolecular assemblies by showing that cubic point group symmetry is not a physical requirement in multienzyme assembly. These results extend our basic knowledge of protein assembly and greatly expand the number of possibilities to manipulate self-assembling biological complexes to be utilized in innovative nanotechnology applications.

Database
The final coordinates of the E2 structure have been deposited in the Protein Data Bank (PDB accession code 3RQC)

Structured digital abstract
- E2 and E2 bind by x-ray crystallography (View interaction)
- E2 and E2 bind by x ray scattering (View interaction)

Abbreviations
CD, catalytic domain; E1, 2-oxoacid decarboxylase; E2, dihydrolipoyl acyl-transferase; E2p, pyruvate dehydrogenase multienzyme complex dihydrolipoyl acyl-transferase; E3, dihydrolipoamide dehydrogenase; OADHC, 2-oxoacid dehydrogenase multienzyme complex; SAXS, small-angle X-ray scattering.

Introduction

The family of 2-oxoacid dehydrogenase multienzyme complexes (OADHCs) functions in the pathways of central metabolism, and these enzymes are responsible for the oxidative decarboxylation of 2-oxoacids to their corresponding acyl-CoAs. Members of the family include the pyruvate dehydrogenase complex, which catalyses the conversion of pyruvate to acetyl-CoA and so links glycolysis and the citric acid cycle; the 2-oxoglutarate dehydrogenase complex, which catalyses the conversion of 2-oxoglutarate to succinyl-CoA within the citric acid cycle; and the branched-chain 2-oxoacid dehydrogenase complex, which oxidatively decarboxylates the branched-chain 2-oxoacids produced by the transamination of valine, leucine and isoleucine. The complexes comprise multiple copies of three component enzymes: 2-oxoacid decarboxylase (E1), dihydrolipoyl acyl-transferase (E2) and dihydrolipoamide dehydrogenase (E3) [1–3].

E2 forms the structural and catalytic core of the complex, and in the well-studied bacterial and eukaryotic complexes multiple E2 polypeptide chains associate into octahedral (24-mer) or icosahedral (60-mer) configurations, depending on the particular complex and the source organism [4]. To achieve this, each E2 polypeptide is composed of three structural domains: a catalytic core domain that assembles into the multimeric core and also contains the acyl-transferase active site, a peripheral subunit binding domain that non-covalently binds E1 and E3, and one to three lipoyl domains in each of which a lipoic acid moiety is covalently attached to a specific lysine residue. The flexible linkers between the three domains enable the lipoic acid to serve as a swinging arm, connecting the active sites of each enzyme and channelling substrate through the complex [3].

Until recently, it was thought that archaea did not possess an OADHC. However, genome sequencing has revealed that the aerobic archaea have genes showing significant identity with genes encoding the bacterial OADHC components [5]. Moreover, we have now expressed those genes from the thermophilic archaeon *Thermoplasma acidophilum* and shown them to be catalytically active and to assemble into a functional OADHC that has whole complex activity with the branched-chain 2-oxoacids and, to a lesser extent, with pyruvate [6,7].

Obtaining X-ray crystal structures of octahedral and icosahedral E2 cores has necessitated the generation of the catalytic cores of these complexes, i.e. E2 molecules without their lipoyl and subunit binding domains and their associated flexible regions [4,8,9]. In the present paper, a similar approach has been taken with the E2 from *T. acidophilum*, and a highly thermostable catalytic core has been produced. Surprisingly, the crystal structure of this core shows it to be a 42-mer assembly, a result confirmed by analytical ultracentrifugation and small-angle X-ray crystallography. To our knowledge, this is the first biological assembly comprising 42 subunits, and it thus represents a truly novel multiprotein assembly.

Results

Characterization of the *T. acidophilum* E2 catalytic core

The segment of the E2 gene encoding the catalytic domain was cloned and expressed in *Escherichia coli* as described in Methods. After purification, the relative molecular mass (M_r) of the recombinant enzyme was confirmed by mass spectrometry; two peaks were observed with M_r values of 25 455 and 25 324, corresponding to those predicted from the gene sequence with and without the N-terminal methionine, respectively. The enzyme was fully active with respect to its transacetylase activity, for which the deleted lipoyl domain is not required, with a k_{cat} = 47 (±4) min^{-1} compared with a k_{cat} = 48 (±3) min^{-1} for the recombinant whole E2 molecule. The E2 catalytic core was highly thermostable, with no loss of catalytic activity after incubation for 30 min at 90 °C, a temperature at which no protein precipitation could be detected by SDS/PAGE (Fig. S1); however, ~ 5% aggregation was observed at 82 °C by dynamic light scattering. On these criteria, therefore, the recombinant E2 core would appear to be an active, folded protein.

Assembly of the E2 catalytic core

To determine if the recombinant E2 core had self-assembled, the protein was subjected to analytical equilibrium centrifugation (Fig. 1) at three different concentrations of protein, each at 3000 and 5000 r.p.m. The data, at each separate concentration and as a combined set, fitted well to a model for a single ideal species with an M_r of 1.08 (±0.02) × 10^6, equating to 42.2 (±0.8) monomers of M_r 25 455. In view of all other 2-oxoacid dehydrogenase complexes from bacteria and eukaryotes possessing E2 cores with either 24 or 60 polypeptide chains, this result for the *Thermoplasma* core was totally unexpected. However, SDS/PAGE analysis after centrifugation confirmed that no proteolysis of the E2 chains had taken place

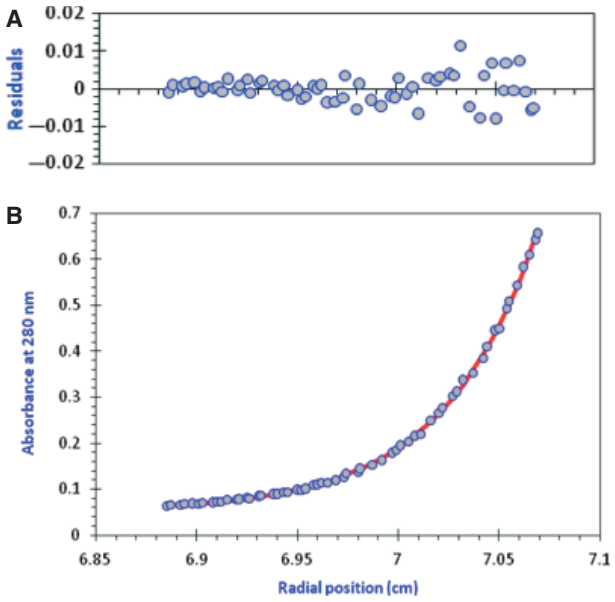

Fig. 1. Sedimentation equilibrium analysis of the assembled E2 catalytic core. The absorbance at 280 nm of the E2 core domain (0.38 mg·mL^{-1}) was recorded at radial positions across a cell after centrifugation for 48 h at 5000 r.p.m. and 20 °C (B; blue circles). Similar data collected at three concentrations of the sample (50, 33 and 15 μM monomer) at two speeds (3000 and 5000 r.p.m.) were fitted simultaneously to a model for a single ideal species (red line). The differences between the data points in this data set and the global fit to the six data sets, the residuals, are shown in (A) and can be seen to be both small and random.

during the centrifugation, and so the 42-mer is not an aberrant result of a degraded E2 core polypeptide. Moreover, we performed analytical equilibrium centrifugation on the *whole* E2 molecule that had been recombinantly expressed and purified in the same way as the truncated core molecule; an M_r of 1.94 ($\pm$0.05) $\times$ 10^6 was obtained, giving 41.8 ($\pm$1.1) monomers of M_r 46 408. Thus it appears that the assembled E2 catalytic core represents the core structure of the native E2 molecule.

Crystal structure of the E2 catalytic core

A complete data set for the *Thermoplasma* E2 catalytic core was processed to 4.0 Å resolution in hexagonal space group H32, with cell parameters $a = b =$ 206.3 Å, c = 438.2 Å, $\alpha = \beta = 90°$, $\gamma = 120°$ (for data collection statistics see Table S1 and for detailed methods see supplementary material Doc. S1). For the structural determination it was assumed that the E2 polypeptides and their trimeric assembly are similar in structure to the dihydrolipoyl transacetylase E2pCD from *Azotobacter vinelandii* (PDB accession code 1EAF) [8,9]. A molecular replacement solution, obtained with the program BALBES [10] using an E2pCD trimer, indicated that the asymmetric unit consisted of six *T. acidophilum* E2 polypeptides, forming two trimers (chains ABC and DEF). Clear evidence for a seventh monomeric E2 molecule (chain G) was observed in the electron density maps that resulted from molecular replacement and it was therefore included in the final model (Fig. 2A; for model refinement statistics see Table S2). This ostensibly monomeric G chain is located at the crystallographic threefold axis and interacts with two symmetry-related E2 G chains, thereby forming a trimer in a fashion that is identical to the other trimers observed in the structure (Fig. 2B). Figure S2 shows the electron density of the crystallographic trimer formed by chain G and its symmetry-related counterparts. As expected from the sequence similarity between the *T. acidophilum* and the *A. vinelandii* E2s, the latter used here as a molecular replacement search model, the E2 core trimers are tightly packed about a threefold symmetry axis. Unexpectedly, analysis of the symmetry-related packing of the *T. acidophilum* E2 core trimers did not yield the cube-shaped octahedral configuration of 24-mer E2 core molecules observed in the 1EAF search model, nor did the structure display the symmetrical dodecahedron-shaped icosahedral geometry observed in the 60-mer dihydrolipoyl acyl-transferase (E2p) cores of pyruvate dehydrogenase multienzyme complexes from *Geobacillus stearothermophilus* and *Enterococcus faecalis* [4]. In contrast with previous structures, the *T. acidophilum* E2 core adopts a novel and unique 42-mer polyhedral sphere with 3_2 symmetry, comprising three regular quadrilateral and six pentameric faces (Figs 3A and S3; stereo images are shown in Fig. S4). This structure exactly accords with the results obtained from our analytical equilibrium centrifugation analysis.

Figure 3B shows that the 42-mer structure forms an oblate spheroid with an equatorial diameter of $\sim$ 210 Å, located on the crystallographic twofold axis, and a polar diameter of $\sim$ 190 Å, located on the crystallographic threefold rotational axis. The oblate spheroid has an aspect ratio (ratio of the polar to equatorial lengths) of 0.9 and an oblateness (f) of 0.09. Similar to the octahedral and icosahedral E2 core structures, the 42-mer *T. acidophilum* E2 core structure contains large windows ($\sim$ 40 Å and $\sim$ 52 Å across the fourfold faces and pentameric faces, respectively). The spheroid is remarkably hollow and filled with solvent, as is indicated by the extraordinarily high solvent content in the crystal (76%; volume : mass ratio (V_m) = 5.03 Å/Da). Such large windows on the fourfold and fivefold faces of the E2 catalytic cores may

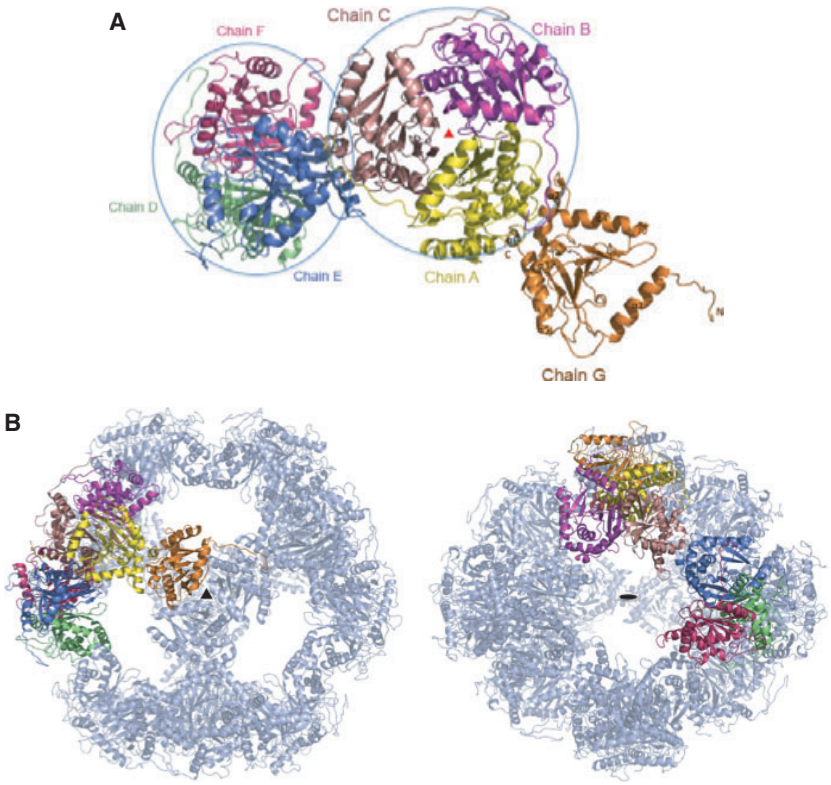

Fig. 2. Crystal structure of the E2 catalytic core. (A) Shown is a ribbon rendering of the seven E2 polypeptides (numbered from A to G) observed in the asymmetric unit. Six of the polypeptides form two tightly packed trimers (circled in blue). Both trimers display a pseudo (non-crystallographic) threefold symmetry, indicated with a red triangle. The apparently monomeric G chain also constitutes a trimer with symmetry-related G chains, related by a crystallographic threefold symmetry. (B) Ribbon representation of E2 core polypeptide chains A–G and their positions with respect to their symmetry-related molecules (rendered in transparent light blue), forming a 42-mer assembly. The left panel presents the assembly viewed down the crystallographic threefold symmetry axis (indicated by a black triangle), and the right panel shows the 42-mer along the crystallographic twofold axis (indicated by a black ellipse). All molecular structure figures were prepared using the program MACPYMOL (http://www.pymol.org).

permit unobstructed passage of the CoA substrate and the acyl-CoA product, because when the catalytic trimers assemble into the E2 core the catalytic residues are located on the inner face of the core.

The 42-mer oblate spheroid observed in the crystal structure of the *T. acidophilum* E2 core is highly unusual. In order to assess whether this atypical amalgamate of cubic and dodecahedral geometries is part of the true biological assembly, we decided to analyse the E2 core assembly in solution, by small-angle X-ray scattering (SAXS).

SAXS analysis of the E2 catalytic core

The X-ray scattering curve for a monodisperse solution of the *T. acidophilum* E2 core is shown in Fig. 4A (for detailed SAXS methods see supplementary material Doc. S1). The curve displays a pronounced subsidiary maximum characteristic of a protein complex with a hollow core, and yields an M_r of 950 000 (±90 000), consistent with the 42-mer structure observed in the crystal. The experimental R_g and the D_{max} were 85 (± 3) and 220 (± 10) Å, respectively, also pointing to a globular shape of the particle. A low resolution *ab initio* model, constructed using the bead modelling program DAMMIF [11,12], neatly fits the experimental data with discrepancy $\chi = 0.95$ (Fig. 4A, curve 2) and superimposes very well with the oblate 42-meric crystal structure as shown in Fig. 4B. Moreover, the scattering curve computed from the crystal structure shows an excellent fit to the experimental data ($\chi = 1.6$; Fig. 4A, curve 3). In contrast, comparisons of the experimental data with those computed for the cubic and the dodecahedral structures show that the *T. acidophilum* E2 core is clearly larger than the 24-mer *A. vinelandii* E2pCD structure ($\chi = 18.6$; Fig. 4A, curve 4) and significantly smaller than the 60-mer *Geobacillus stearothermophilus* E2p structure ($\chi = 13.4$; Fig. 4A, curve 5). Rigid body models constructed using the SAXS data and the E2 crystal structure were compatible with the *ab initio*

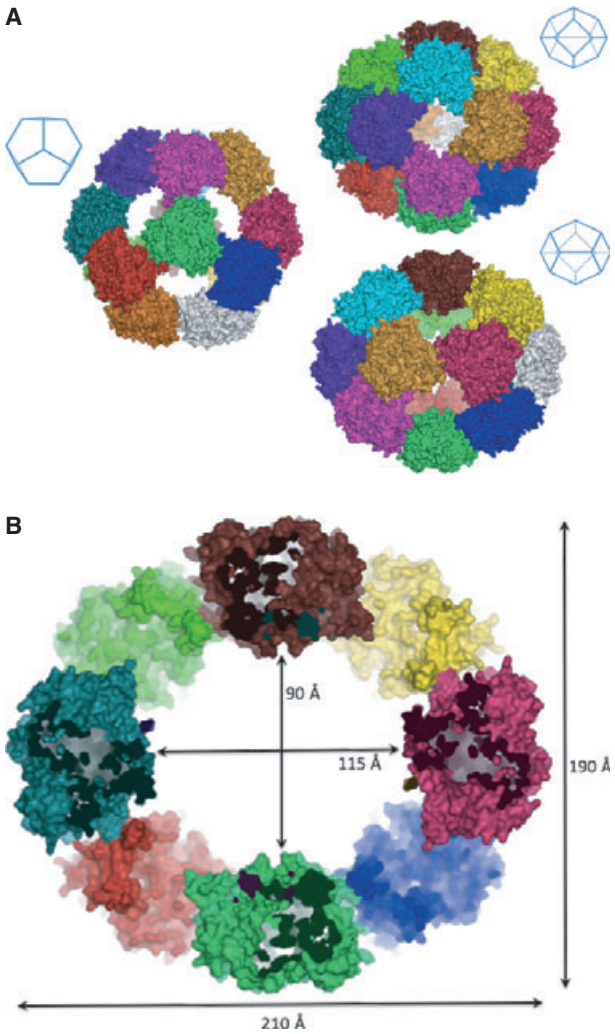

Fig. 3. The crystallographic 42-mer assembly of the *T. acidophilum* E2 catalytic core. (A) Left panel: Surface representations of the complete 42-mer E2 core assembly viewed down the crystallographic threefold axis. Individual E2 trimers are shown in different colours. Right panel: The assembly is also shown from two viewpoints along the plane of twofold symmetry axes, perpendicular to a threefold axis. For clarity, three insets with schematic representations of the polyhedron are shown, corresponding to the individual viewpoints. (B) Cross-section of the E2 assembly depicting the external and internal dimensions of the 42-mer.

SAXS models and with the presented 42-mer assembly. All these results clearly indicate that the 42-mer oblate spheroid structure also occurs in solution and is not an artefact caused by crystal-packing interactions.

Examination of the switch element interactions in the 42-mer E2 core assembly

Even though the side-chain densities are sparse in the 4.0 Å resolution structure presented here, a comparison of the positions of the E2 core protomer main chains can be used to infer in more detail the similarities and differences between the inter-trimeric switch element interactions involving pentameric and cubic trimers. Similar to the previously described inter-trimer interactions of the dodecahedral E2p core of *G. stearothermophilus* [4], the trimers at the pentameric face in the structure of the *T. acidophilum* E2 core include truly equivalent interactions, involving hydrophobic contacts between residues I396 and I397 in the C-terminal 3_{10} helix (helix α6) and the same isoleucine side chains in the corresponding twofold related trimer (Fig. 5B, left panel). As has been described for the dodecahedral E2p core [4], this equivalent trimeric interaction between two pentameric trimers of the 42-mer is further stabilized by main-chain–main-chain interactions between the 3_{10} helix α6 and the N-terminus of helix α4. In addition to these interactions, two ionic interactions can be observed in the *T. acidophilum* E2 structure between R292 and its neighbouring residue D291 in the twofold symmetry related chain (Fig. 5A, left panel). This interaction cross-bridges the N-termini of the α4 helices between the twofold related trimers and is therefore likely to contribute significantly to the stability of the pentameric interactions in the 42-mer. Interestingly, both the aspartate and arginine residues are conserved in the *G. stearothermophilus* E2p sequence (D310 and R311), and in the dodecahedral E2p structure they are in close proximity to their counterpart residues in the twofold related trimers. Conversely, at the pentameric–quadrilateral interfaces in the *T. acidophilum* E2 structure, the distance between R292 and D291 is significantly larger than that at the pentamer–pentamer interfaces (Fig. 5A, left panel), thereby abolishing the possibility for such an ionic interaction. Interestingly, whilst residue D291 is conserved in many E2 sequences, the cross-bridging arginine residue R292 is not present in the cubic *A. vinelandii* E2pCD (Fig. S5).

In addition to the truly equivalent interactions, some of the quasi-equivalent contacts described for the *G. stearothermophilus* E2p, *E. faecalis* E2p and *A. vinelandii* E2pCD structures, involving an anchor residue located at the end of the C-terminal 3_{10} helix (M425 in *G. stearothermophilus*, M537 in *E. faecalis* and L637 in *A. vinelandii*), also appear to be important in stabilizing the α6 helix in the *T. acidophilum* E2 structure. As shown in Fig. 5(B), the C-terminal residue Y398 is secured in a hydrophobic pocket of the opposing trimer like a ball-and-socket joint. This hydrophobic 'socket' is composed of residues M219, L222 and L223 in helix α2; M295 and V296 in helix α4; and amino acids V233 and F238 in helix α3. The differences

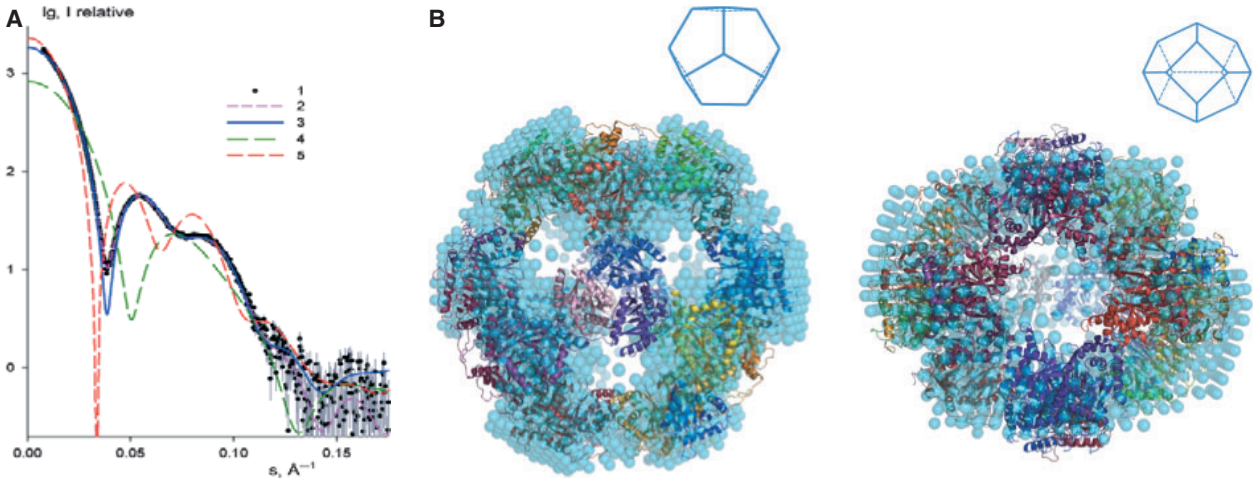

Fig. 4. SAXS analysis of the *T. acidophilum* E2 catalytic core. (A) The X-ray scattering patterns: (1) experimental data of the E2 core (dots with grey error bars), which is characteristic for a hollow sphere; (2) scattering computed from the *ab initio* model; (3) computed scattering from the 42-mer assembly observed in the crystal structure; (4), (5) computed scattering curves from the structures of the cubic 24-mer assembly of dihydrolipoyl transacetylase (E2pCD) from *A. vinelandii* (PDB accession code 1EAF) and from the dodecahedral 60-mer assembly of dihydrolipoyl acetyltransferase (E2p) catalytic core from *G. stearothermophilus* (PDB accession code 1B5S) respectively. (B) Low resolution *ab initio* bead model of the E2 core, represented by cyan transparent spheres, fitted with the crystal structure of the 42-mer E2 core assembly. Two orientations are shown with viewpoints along the crystallographic threefold axis (left panel) and the twofold symmetry axis (right panel). Schematic representations of the 42-mer polyhedron, corresponding to specific viewpoints, are included for clarity.

between these hydrophobic inter-trimer interfaces at the quadrilateral and pentagonal interfaces are very subtle and small rotational movements in the ball-and-socket joint (along and perpendicular to the pseudo twofold axis) will be sufficient to attain the transformation from cubic to dodecahedral symmetry interactions, resulting in the formation of the twofold symmetrical ionic link between D291 of one trimer and R292 of the other, thereby expanding the number of interactions and increasing the corresponding inter-trimer interaction surface from $\sim$ 680 to $\sim$ 780 Å^2.

Discussion

The data presented in this paper, from analytical centrifugation, X-ray crystallography and SAXS, all indicate that the catalytic core of the *T. acidophilum* OADHC assembles into a monodisperse structure comprising 42 identical subunits. At first glance this result appears to conflict with our previously reported sedimentation velocity analysis of the whole E2, suggesting a 24-mer assembly [6]. The discrepancy might be explained by proteolytic degradation to which this molecule is prone, and therefore in the current work the integrity of the E2 samples was checked before and after the equilibrium centrifugation and no sign of degradation was observed. Alternatively, it is known that hydrodynamic and shape parameters can cause complications in the analysis of sedimentation velocity data, whereas this is not the case in the sedimentation equilibrium technique used in the current paper.

The 42-mer structure of the E2 core is unique in that it does not follow strict geometric considerations that have been previously used to explain the cubic and dodecahedral E2 cores of 2-oxoacid dehydrogenase complexes [4,8,9] as well as the architecture of virus capsids [13–17]. Although Casper and Klug [13] argued that molecular structures do not follow exact mathematical concepts, but rather assemble in a manner that satisfies the condition to be in a minimum energy configuration, they assumed that a protein shell built from identical subunits, using the same contacts between subunits over and over again, is the key to the organization of the units into a regular shell. They introduced the term 'quasi-equivalence' to refine this assumption by clarifying that these bonds may be deformed in slightly different ways in different, non-symmetry-related environments. However, they assume that there is a maximum of $\sim$5° of non-equivalence permitted between subunits, which they base on virus models with icosahedral shells.

A more recent study by Izard and co-workers [4] proposes that the assembly of cubic and dodecahedral cores of the pyruvate dehydrogenase complex, like virus particles, is governed by Caspar and Klug's quasi-equivalence principles. They take these mathematical concepts even further by suggesting that the assembly of the cubic and dodecahedral core

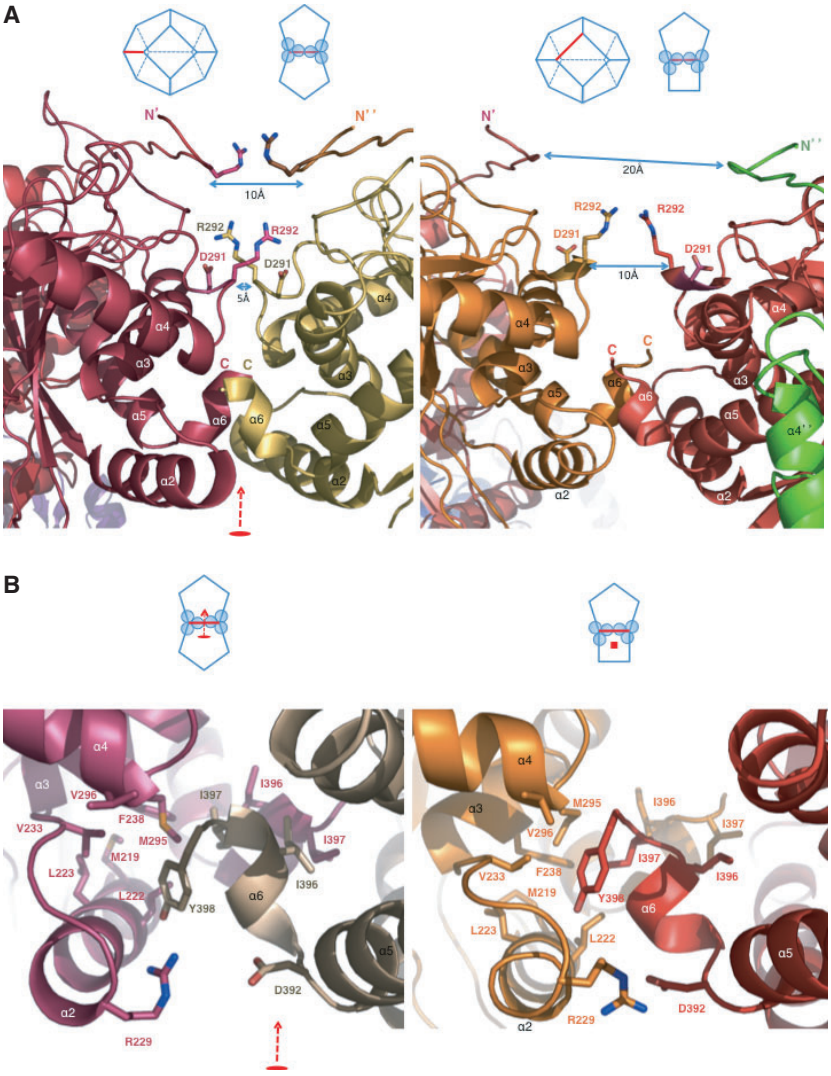

Fig. 5. Inter-trimer interactions in the *T. acidophilum* catalytic E2 core. (A) Left panel: Inter-trimer interactions observed on the pseudo-twofold symmetry axis between two pentameric faces. Right panel: Inter-trimer interactions observed at the interface between pentameric and quadrilateral faces of the 42-mer, located near a pseudo-fourfold symmetry axis. The positions of E2 core helices are indicated. Specific residues that are thought to play a role in the inter-trimer interactions are shown in stick representations and the distances between these residues are indicated. Schematic representations of the polyhedron, corresponding to specific viewpoints, are also included for clarity. (B) Left panel: Close-up view of the hydrophobic inter-trimer interactions at a pentameric interface. Right panel: Close-up view of the inter-trimer interactions at a pentamer–quadrilateral interface of the 42-mer E2 core assembly.

complexes follows precise Euclidean principles of geometry, by showing that the cubic form can be related to that of the dodecahedron by precise expansion and rotation factors [4].

Contrary to Caspar and Klug's quasi-equivalence principles or Izard and co-workers' Euclidean geometry considerations, the *T. acidophilum* E2 core assembles to form a stable and biologically active particle consisting of identical trimer-forming units with a unique oblate spheroid geometry, comprising an amalgam of octahedral and icosahedral geometrical aspects. Although the trimeric building blocks are identical to the cubic and dodecahedral E2 core structures, the arrangement of the trimers into six pentagonal and three quadrilateral elements in the 42-mer is truly extraordinary and surpasses the degrees of non-equivalence between subunits allowed by quasi-equivalence principles.

This is not the first report of a structure that disobeys the Casper and Klug dogma. The nature of the quasi symmetry found in many small viruses with triangulation numbers $T = 3$ or $T = 7$ is generated by structural modifications far greater than those assumed or predicted by Caspar and Klug ([17] and references therein). However, the protein capsids of these viruses may be composed of two, three or four different protein chains. Although there are examples of virus particles that consist of identical subunits, such as Dengue virus [18], their icosahedral lattices are populated with more than 60 subunits (90 in the case of Dengue virus), giving rise to more complex capsids that fall outside the range of principles introduced by Crick and Watson, who predicted that a virus with cubic point group symmetry could only be made up of a multiple of 12 subunits with a maximum of 60 [19]. The 42-mer is unique because it consists of identical

homo-trimers in structurally non-equivalent positions. While it was once thought that identical parts of a macromolecular particle would self-assemble in a regular entity [19], like parts of a toy that could be put together by a child in only one way, it appears that the *T. acidophilum* E2 core has been assembled with two different sets of instructions. A high resolution structure of the 42-mer E2 assembly may provide the clues leading to identification of the specific molecular details that determine the non-equivalent octahedral and icosahedral switch element interactions in this protein assembly.

The structure of the novel 42-mer multienzyme assembly, presented here, can be used to predict novel irregular protein assemblies with combined cubic and dodecahedral elements. Figure S6 shows the possible multimers that can be assembled from trimers with two different non-equivalent contacts, including a 30-mer heptahedral intermediate and a 48-mer decahedral intermediate.

Conclusion

This work shows that assembly of a particle consisting of identical subunits is not limited to the cubic point groups as predicted by the traditional concepts of quasi-equivalence, and that such particles are therefore not restricted to 12, 24 or 60 subunits. This work will enable us to extend our basic knowledge of protein assembly and help us to predict novel irregular protein assemblies, thereby greatly expanding the number of possibilities to manipulate self-assembling biological complexes to function in innovative nanotechnology applications.

Methods

Generation of the E2 catalytic core

The C-terminal catalytic domain of enzyme E2 was identified by sequence alignments with other 2-oxoacid dehydrogenase E2 sequences and was thus judged to comprise amino acids 177–400 of the whole E2 protein. The gene fragment encoding this domain was PCR amplified from the previously cloned [6] whole E2 gene in pGEM-T using Phusion, and cloned into pET-24a vector using *Nde*I and *Xho*I. For expression purposes, Pro177 was changed to M177, and DNA sequencing of the cloned fragment confirmed the fidelity of the PCR amplification.

Protein expression in *E. coli* Rosetta (DE3) grown at 37 °C in LB supplemented with kanamycin (30 $\mu g \cdot mL^{-1}$) and chloramphenicol (34 $\mu g \cdot mL^{-1}$) was induced with 1 mM isopropyl thio-β-D-galactoside. Cells were harvested 20 h after induction and lysed by sonication in 50 mM Tris, pH 8.8. The resultant cell extract was incubated at 55 °C for 15 min before precipitated proteins were removed by centrifugation. DNA was then precipitated with 0.2–0.5% (v/v) polyethylenimine, and the recombinant protein was purified by anion exchange on HiTrap Q (HP), followed by gel filtration on a Superdex 200 column. The relative molecular mass of the recombinant protein was confirmed by mass spectrometry, using a Bruker MICROTOF QII at a resolution of ~ 17 000.

Assay of E2 catalytic activity

The dihydrolipoamide *S*-acetyltransferase activities of the E2 enzyme and of its catalytic core were determined using a coupled assay with phosphotransacetylase from *Bacillus stearothermophilus* [20]. The 1 mL assay contained 0.1 M Tris/HCl (pH 7.6 at 20 °C), 4.0 mM DL-dihydrolipoamide, 0.1 mM CoA, 10 mM acetyl phosphate and 7.5 U phosphotransacetylase; the reaction producing acetyl-CoA was allowed to reach equilibrium at 55 °C before the E2 enzyme was added. The increase in thioester bond concentration due to the E2-catalysed formation of acetyl-lipoamide at 55 °C was followed at 233 nm (molar absorption coefficient $\varepsilon_M = 5.4 \times 10^3\ M^{-1}\ cm^{-1}$) [21].

Dynamic light scattering

The thermostability of the recombinant E2 catalytic core was characterized by dynamic light scattering (Nano-S Zetasizer, Malvern, UK). The protein in 50 mM sodium phosphate, pH 7.0, containing 100 mM NaCl was heated from 25 to 70 °C at 5 °C intervals and with 2 °C intervals between 70 and 82 °C; the sample was left for 2 min to equilibrate before each reading. All readings were taken over a 40-s period in a low-volume, sealed quartz cuvette containing 80 μL of sample.

Analytical equilibrium centrifugation

Three different concentrations of the E2 core in 50 mM Tris/HCl, pH 8.8, 100 mM NaCl (110 μL) plus buffer-only references (120 μL) were loaded into six-sector centre-pieces within centrifuge cells. Protein monomer concentrations of 50, 33 and 15 μM (1.27, 0.83 and 0.38 $mg \cdot mL^{-1}$) were calculated from their absorbance using an ε_M of 14 900 $M^{-1}\ cm^{-1}$ (path length 1.2 cm) [22]. Centrifugation was carried out in an An60Ti rotor in a Beckman XL-A ultracentrifuge at 20 °C. At time intervals, the $A_{280\ nm}$ of the samples was measured at radial positions from 5.8 to 7.2 cm in 0.03-cm steps. Cells were spun at 3000 and 5000 r.p.m. for 48 h each, with scans completed every 6 h after 24 h. Equilibrium was reached by 42 h as confirmed by overlay of scans completed at 42 and 48 h. A final 6-h clearing spin at 40 000 r.p.m., during which time the

protein was removed from the meniscus of the solution, enabled the background absorbance of the buffer in the sample versus the reference to be measured.

At each speed, the data from the final scan were simultaneously fitted in ORIGIN (Beckman Coulter (UK) Ltd, High Wycombe, UK) to the model for a single ideal species using a partial specific volume of 0.7445 $mL \cdot g^{-1}$ and buffer density of 1.00379 $g \cdot mL^{-1}$ [estimated from the protein sequence and the buffer composition, respectively, using SEDNTERP (http://www.jphilo.mailway.com/download.htm)]. Offset values for each data set were floated in this global fit, but they closely matched the absorbance values at the menisci obtained from the clearing spin. Error margins were obtained from a weighted fit to the same data and are from the 95% confidence limits from this fitting routine. The differences between the data points in each data set and the global fit to the six data sets, the residuals, were also calculated.

Crystallization, data collection and structure analysis

X-ray diffraction data were collected at the Diamond Light Source (Oxfordshire, UK) and data were processed using the HKL2000 package [23]. Molecular replacement was carried out with BALBES [10], using the structure of *A. vinelandii* E2pCD (Protein Data Bank accession code 1EAF) as a search model. Model building was done with COOT [24] followed by rounds of refinement using REFMAC5, part of the CCP4i software package [25]. The final values of R_{work} (random errors of the working data set) and R_{free} (random errors computed from a test set of reflections that was not used in the refinement) were 25.3% and 32.8%. The final coordinates of the E2 structure have been deposited to the Protein Data Bank (PDB accession code 3RQC). Detailed crystallography methods are provided in supplementary material Doc. S1.

Small-angle X-ray scattering analysis

Synchrotron radiation X-ray scattering data were collected at the X33 beam line of the EMBL, Hamburg Outstation (DORIS III storage ring at DESY [26]), equipped with an automatic sample changer [27]. The data were processed using standard procedures and extrapolated to zero solute concentration using the program package PRIMUS [28].

The forward scattering $I(0)$ and the radius of gyration (R_g) were computed from the entire scattering patterns using the indirect transform package GNOM [29], which also provided the intraparticle distance distribution function $p(r)$ and the maximum dimension D_{max}. The estimation of excluded volume (V_{ex}) and low resolution *ab initio* models of E2 were obtained using the program DAMMIF [11,12]. Ten DAMMIF runs were performed to check the stability of the solution, and the results were averaged using the program DAMAVER [30] to yield the most probable model. The scattering from the high resolution models was calculated with the program CRYSOL [31]. Rigid body modelling of the E2 core was performed using the program SASREF [32]. Detailed SAXS methods are provided in supplementary material Doc. S1.

Acknowledgements

We are grateful to the UK Biotechnology and Biological Research Council for financial support in the form of a studentship to NLM. Part of the work was supported by a Research Grant from the US Air Force Office of Scientific Research. The analytical ultracentrifugation data were collected with equipment provided by Professor S.E. Halford FRS, University of Bristol, UK, and we thank him for help and valuable discussions. Mass spectrometric analyses were carried out by T.M. Gibson, the BioCentre Facility, University of Reading, UK. Finally, we thank Dr D.J. Crennell for the analysis of the polygon angles. This work was carried out with the support of the Diamond Light Source and access to the DESY-EMBL beam lines was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement 226716.

Author contributions

MJD and DWH planned and supervised the project and, with NLM, designed the experiments. JJTM carried out the ultracentrifugation analyses, and DIS and JMHvdE carried out the SAXS analyses; all other laboratory experiments, including the gene cloning and expression and the purification of the recombinant protein, were conducted by NLM. SJC and JMHvdE collected and processed the X-ray crystallographic data, calculated initial structures and guided NLM in the refinement of those structures. All the authors contributed to the writing of the paper.

Conflict of interest

The authors declare no conflict of interest.

References

1 Perham RN (1991) Domains, motifs, and linkers in 2-oxo acid dehydrogenase multienzyme complexes – a paradigm in the design of a multifunctional protein. *Biochemistry USA* **30**, 8501–8512.

2 Perham RN (2000) Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Ann Rev Biochem* **69**, 961–1004.

3 Perham RN, Jones DD, Chauhan HJ & Howard MJ (2002) Substrate channelling in 2-oxo acid dehydrogenase multienzyme complexes. *Biochem Soc Trans* **30**, 47–51.

4 Izard T, Ævarsson A, Allen MD, Westphal AH, Perham RN, de Kok A & Hol WGJ (1999) Principles of quasi-equivalence and Euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes. *Proc Natl Acad Sci USA* **96**, 1240–1245.

5 Danson MJ, Lamble HJ & Hough DW (2007) Central metabolism. In *Archaea: Molecular and Cell Biology* (Cavicchioli R, ed.), Chapter 12, pp. 260–287. ASM Press, Washington, DC.

6 Heath C, Posner MG, Aass HC, Upadhyay A, Scott DJ, Hough DW & Danson MJ (2007) The 2-oxoacid dehydrogenase multi-enzyme complex of the archaeon *Thermoplasma acidophilum* – recombinant expression, assembly and characterization. *FEBS J* **274**, 5406–5415.

7 Posner MG, Upadhyay A, Bagby S, Hough DW & Danson MJ (2009) A unique lipoylation system in the archaea: lipoylation in *Thermoplasma acidophilum* requires two proteins. *FEBS J* **276**, 4012–4022.

8 Mattevi A, Obmolova G, Schulze E, Kalk KH, Westphal AH, de Kok A & Hol WG (1992) Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex. *Science* **255**, 1544–1550.

9 Mattevi A, Obmolova G, Kalk KH, Westphal AH, de Kok A & Hol WG (1993) Refined crystal structure of the catalytic domain of dihydrolipoyl transacetylase (E2p) from *Azotobacter vinelandii* at 26 Å resolution. *J Mol Biol* **230**, 1183–1199.

10 Long F, Vagin AA, Young P & Murshudov GN (2008) BALBES: a Molecular-Replacement Pipeline. *Acta Crystallogr* **D64**, 125–132.

11 Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* **76**, 2879–2886.

12 Franke DS & Svergun DI (2009) DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J Appl Crystallogr* **42**, 342–346.

13 Caspar DLD & Klug A (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* **27**, 1–24.

14 Caspar DLD (1980) Movement and self-control in protein assemblies Quasi-equivalence revisited. *Biophys J* **32**, 103–138.

15 Caspar DLD (1991) Self-control of assembly. *Curr Biol* **1**, 30–32.

16 Cheng RH, Reddy VS, Olson NH, Fisher AJ, Baker TS & Johnson JE (1994) Functional implications of quasi-equivalence in a $T = 3$ icosahedral animal virus established by cryo-electron microscopy and X-ray crystallography. *Structure* **2**, 271–282.

17 Johnson JE & Speir JA (1997) Quasi-equivalent viruses: a paradigm for protein assemblies. *J Mol Biol* **269**, 665–675.

18 Kuhn RJ, Zhang W, Rossmann MG, Pletnev SV, Corver J, Lenches E, Jones CT, Mukhopadhyay S, Chipman PR, Strauss EG *et al.* (2002) Structure of Dengue virus. *Cell* **108**, 717–725.

19 Crick FH & Watson JD (1956) Structure of small viruses. *Nature* **177**, 473–475.

20 Packman LC, Perham RN & Roberts GCK (1984) Domain structure and ^{1}H-NMR spectroscopy of the pyruvate dehydrogenase complex of *Bacillus stearothermophilus*. *Biochem J* **217**, 219–227.

21 Srere PA & Kosicki GW (1961) The purification of citrate-condensing enzyme. *J Biol Chem* **236**, 2557–2559.

22 Gill SC & von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **182**, 319–326.

23 Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. In *Methods Enzymol Macromolecular Crystallography*, *Part A*, Vol. **276** (Carter CW & Sweet RM, eds), pp. 307–326. Academic Press, New York.

24 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallog* **D60**, 2126–2132.

25 CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallog* **D50**, 760–763.

26 Roessle MW, Klaering R, Ristau U, Robrahn B, Jahn D, Gehrmann T, Konarev P, Round A, Fiedler S, Hermes C *et al.* (2007) Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *J Appl Crystallogr* **40**, s190–s194.

27 Round AR, Franke D, Moritz S, Huchler R, Fritsche M, Malthan D, Klaering R, Svergun DI & Roessle M (2008) Automated sample-changing robot for solution scattering experiments at the EMBL Hamburg SAXS station X33. *J Appl Crystallogr* **41**, 913–917.

28 Konarev PV, Volkov VV, Sokolova AV, Koch MHJ & Svergun DI (2003) PRIMUS – a Windows-PC based system for small-angle scattering data analysis. *J Appl Crystallogr* **36**, 1277–1282.

29 Svergun DI (1992) Determination of the regularization parameter in indirect transform methods using perceptual criteria. *J Appl Crystallogr* **25**, 495–503.

30 Volkov VV & Svergun DI (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. *J Appl Crystallogr* **36**, 860–864.

31 Svergun DI, Barberato C & Koch MHJ (1995) CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* **28**, 768–773.

32 Petoukhov MV & Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* **89**, 1237–1250.

Supporting information

The following supplementary material is available:
Fig. S1. SDS/PAGE analysis of the thermostability of the *T. acidophilum* E2 catalytic core.
Fig. S2. Electron density of the E2 catalytic core structure.
Fig. S3. Schematic representation of the E2 42-mer assembly.
Fig. S4. Stereo images of the 42-mer E2 core.
Fig. S5. Alignment of *T. acidophilum* and bacterial E2 amino acid sequences.
Fig. S6. Schematic representations of possible regular and irregular symmetry forms of E2 core assemblies.
Doc. S1. Supplementary experimental procedures.
Table S1. X-ray data collection statistics.
Table S2. Model refinement statistics.

This supplementary material can be found in the online version of this article.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.