(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 April 2010 (01.04.2010)

PCT

(10) International Publication Number
WO 2010/035009 A1

(51) International Patent Classification:
C07K 14/00 (2006.01) *C07K 14/21* (2006.01)
C07K 14/195 (2006.01) *C07K 14/245* (2006.01)

(21) International Application Number:
PCT/GB2009/002305

(22) International Filing Date:
24 September 2009 (24.09.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0817507.7 24 September 2008 (24.09.2008) GB

(71) Applicant *(for all designated States except US)*: **UCL BUSINESS PLC** [GB/GB]; The Network Building, 97 Tottenham Court Road, London W1T 4TP (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WOOD, Stephen, P.** [GB/GB]; Centre for Amyloidosis and Acute Phase Proteins, Division of Medecine, UCL Medical School, Royal Free Campus, Rowland Hill Street, London NW3 2PF (GB). **MONTGOMERY, Mark, G.** [GB/GB]; Centre for Amyloidosis and Acute Phase Proteins, Division of Medecine, UCL Medical School, Royal Free Campus, Rowland Hill Street, London NW3 2PF (GB).

(74) Agents: **BENTHAM, Andrew** et al.; J.A. Kemp & Co., 14 South Square, Gray's Inn, London WC1R 5JJ (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

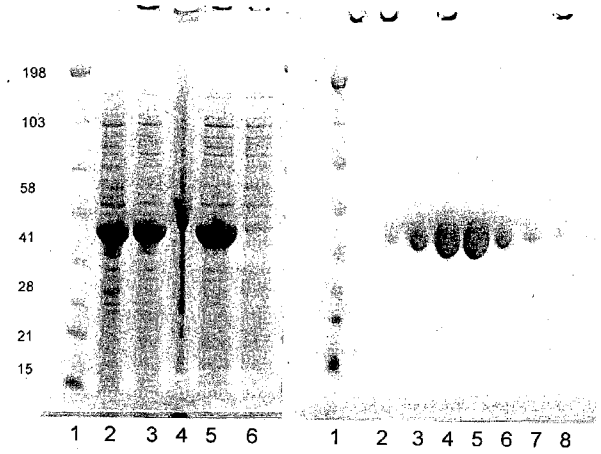
Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) **Title**: PROTEIN CAGES COMPOSED OF TWELVE PENTAMERS

Fig. 1



(57) **Abstract**: This invention concerns protein cage nanostructures that can for example be used to encapsulate chemical processes away from bulk solution and to deliver encapsulated "guest" moieties (e.g. pharmaceutical or imaging agents) through attached targeting domains. The protein monomers from which these nanostructures are self-assembled belong to the FAH fold superfamily. Sixty monomers self-assemble, via twelve pentamers, to form a protein cage, dodecahedral in shape with 532 symmetry and stabilised by coordination by a ligand at the three-fold symmetry axes of three amino acid residues of the same type from different pentamers. A preferred protein is E. *coli* 2-keto-4-pentenoate hydratase (2-hydroxypentedienoic acid dehydratase) (MhpD) (EC4.2.1.80), stabilised by phosphate, sulphate or other negative ions. Protein cages of the invention are unique in that they self-assemble at low pH; however, they can also be modified for use at neutral pH by directed mutation of the amino acid sequence of the monomers.

WO 2010/035009 A1

PROTEIN CAGES COMPOSED OF TWELVE PENTAMERS

FIELD OF THE INVENTION

This invention concerns protein cage nanostructures that can for example be used to encapsulate chemical processes away from bulk solution and to deliver cargoes of encapsulated "guest" materials through attached targeting domains.

BACKGROUND OF THE INVENTION

Natural systems employ a wide range of self-assembling containers in a range of biological processes. For example, virus particles carry and deliver viral genomes to target cells, the protein ferritin sequesters iron for storage, certain heat shock proteins encapsulate protein folding intermediates and chaperone their correct folding. The containers are composed of symmetrical assemblies of protein molecules that come together to define architectures that provide interior cavities that are ideal nanocavities for encapsulation of nanomaterials, other molecules and processes (Neimeyer 2001, Uchida *et al* 2007). There is considerable current interest in exploiting these materials in a wide range of potential applications in materials science, electronics, imaging and drug delivery.

The archetype of this family is ferritin, a multisubunit protein involved in sequestering iron *in vivo* in a 12nm diameter particle. Hydrated ferric oxide is formed following oxidation of ferrous iron entering the protein and nucleation and crystallisation of ferric oxide (Mann and Meldrun 1991). Subsequently a wide range of different protein nanoparticles including viruses, enzymes and heat shock proteins have been examined for their potential applications. They provide a library of materials ranging in size from 9-30nm with variable properties including symmetry, interior cage volume and chemical character, window size, and thermal stability. These are known as protein cages. In addition to the inherent properties of the protein cage arising in nature there is considerable scope through biotechnological methods to modify the particles for particular applications. For instance, polypeptide

chain extensions might be added to the interior or exterior surfaces enabling targeting of particles to particular locations or engineering affinity for particular other molecules within the cage. For a review of protein cages as multifunctional nanoplatforms, see Uchida (2007), and see WO03/096990 for discussion of protein cages in the delivery of medical imaging and therapy.

In general, however, polypeptide structures lack the robustness ideally required for a nanocontainer in materials science as they have evolved to participate in biological processes that mostly take place in mild conditions.

SUMMARY OF THE INVENTION

We have produced a new variety of protein cage nanoparticle. The protein from which we have made this new protein cage is the enzyme 2-keto-4-pentenoate hydratase (2-hydroxypentedienoic acid dehydratase) (MhpD) (EC4.2.1.80) encoded by a gene of the *E. coli* K12 chromosome, whose amino acid sequence is given by Swiss-Prot entry P77608 and shown below as SEQ ID NO: 1.

```
MTKHTLEQLA      ADLRRAAEQG EAIAPLRDLI GIDNAEAAYA   40
IQHINVQHDV      AQGRRVVGRK VGLTHPKVQQ QLGVDQPDFG   80
TLFADMCYGD      NEIIPFSRVL QPRIEAEIAL VLNRDLPATD   120
ITFDELYNAI      EWVLPALEVV GSRIRDWSIQ FVDTVADNAS   160
CGVYVIGGPA      QRPAGLDLKN CAMKMTRNNE EVSSGRGSEC   200
L[G/E*]HPLNAAVW LARKMASLGE PLRTGDIILT GALGPMVAVN   240
AGDRFEAHIE      GIGSVAATFS SAAPKGSLS               269
```

[* Residue 202 may be either G or E. Both versions appear in the literature and both are within the scope of the invention.]

This enzyme forms a particle that is ~20nm in exterior diameter and defines a potential cargo compartment within of 15nm diameter. It is constructed from 60 protein subunits organised with classical 532 symmetry, formed by interaction of 12 stable pentamers. The particle is unique in that it forms at low pH(<4) in the

presence of phosphate ions, sulphate ions and other ions of similar charge and geometry and is readily disassembled at neutral pH. Two ions bind to the cage at each of the three-fold symmetry axes of the particle and coordinate, respectively, three Arginine 15 residues and three Glutamine 19 residues (**bold and underlined** in the protein sequence above) from different pentamers. Gel filtration and multi-angle laser light scattering (MALLS) shows that the particle of 1.8Mda forms rapidly in concentrated phosphate and sulphate solutions at low pH. Other molecules can be trapped within the cage if present during assembly.

2-hydroxypentadienoic acid hydratase (MhpD) is the fourth enzyme in the Mhp – the m-hydroxyphenylpropionic acid catabolic pathway of *Escherichia coli* (Ferrandez *et al*, 1997, Prieto *et al* 1996). It is a divalent cation-dependent hydratase that converts 2-hydroxypenta-2,4-dienoic acid to 2-hydroxy-4-ketopentanoic acid (Fig. 9), which is then converted, by the succeeding MhpFE heterodimer, to the Krebs cycle intermediates pyruvate and acetyl-CoA. The preferential cation for MhpD is manganese, although it has also been shown to be active with magnesium, cobalt and zinc (Pollard & Bugg, 1998). The MhpD catalysed reaction proceeds rapidly, with a Kcat/Km of 11 x 10^6, which is thought to be necessary due to the short half-life of the substrate.

A three-dimensional structure of MhpD has been determined previously to 2.9Å resolution, by a structural genomics consortium (NYSGXRC – see references below), but not reported in the literature (PDB code 1sv6). The quaternary structure of MhpD in this crystal is pentameric and made up of identical 30kDa subunits, each comprising 269 amino acids, whereas the protein cage of the invention is formed from a total of 60 subunits.

Therefore, the invention provides a new and different protein cage crystal form, whose existence was not expected or predictable from the previous crystal form determined by the consortium. Because this is a protein cage, it is useful in ways that were not foreseeable from the previously known crystal form, and it is also advantageous over previously known protein cages formed from other proteins

because of its unique viability at low (<4) pH. Mutation of aspartic acid residues at subunit interfaces near the two-fold axis would remove an electrostatic repulsion destabilising particles at neutral pH and extend the pH range of existence of the protein cage. Similarly, mutation of the arginine residues surrounding the five-fold axis window and removing their mutual repulsion will allow the window to shrink.

Other protein cage nanostructures assembled in the same way – i.e. by the interaction of suitable residues at the three-fold symmetry axes of icosahedral, 532-symetrical structures made up of 60 protein monomers in 12 pentamers, with ligands such as phosphate or sulphate coordinating residues from three different pentamers – can be envisaged.

Accordingly, the invention provides protein cage nanostructure composed of twelve pentamers, each pentamer containing five protein monomers; said pentamers forming a protein cage nanostructure that is dodecahedral in shape with 532 symmetry; said nanostructure being stabilised by coordination, at each of the three-fold symmetry axes, of three amino acid residues of the same type from three different pentamers, by a ligand; each monomer comprising a protein belonging to the FAH fold superfamily whose secondary structure includes an incomplete barrel comprising a strongly twisted β-sheet flanked by α-helices.

The invention also provides a method of producing a protein cage nanostructure of the invention, comprising contacting said monomers with said ligands in solution at a pH at which said monomers will self-assemble to form said nanostructure.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: Coomassie blue stained reducing SDS electrophoresis. Left panel. Lane1 molecular weight (MW) markers, 2 no induction, 3 induction, 4 cell lysis pellet, 5 cell lysis supernatant, 6 untransformed cells. Right panel. Lanes 2-8, Superdex 200 gel filtration fractions.

Figure 2: The entire 20nm MhpD particle viewed down the five-fold symmetry axis.

Figure 3: Removal of five pentamers from the MhpD particle surface shows the hollow centre.

Figure 4: View into the MhpD particle along the five-fold axis showing Arg114 and Arg172 lining the 2nm pore.

Figure 5: View down the three-fold axis of the MhpD particle showing a phosphate ion coordinated by Arg15 sidechains from three N-terminal helices.

Figure 6: View down the two-fold axis of the MhpD particle showing Asp33 and Asp75 side chains hydrogen bonded across a rectangular cavity penetrating the protein shell.

Figure 7: Section of gel filtration profile of MhpD in 1M Na/K phosphate buffer pH3.2 showing MALS estimates of molecular weight at 1.64Mda.

Figure 8: Superdex 200 gel filtration profile of MhpD particles in 1M phosphate buffer pH3.2 assembled in the presence of cytochrome c. Dual wavelength monitoring reveals haem signal (406nm) co-eluting with MhpD.

Figure 9: 2-hydroxypentadienoic acid hydratase (MhpD), the fourth enzyme in the Mhp – the m-hydroxyphenylpropionic acid catabolic pathway of *Escherichia coli* is a divalent cation-dependent hydratase that converts 2-hydroxypenta-2,4-dienoic acid to 2-hydroxy-4-ketopentanoic acid.

DETAILED DESCRIPTION OF THE INVENTION

Protein Cage Nanostructures of the Invention

Monomers

The protein cage nanostructures of the invention are composed of protein monomers. Each individual monomer is a protein molecule belonging to the FAH fold superfamily whose secondary structure includes an incomplete barrel comprising a strongly twisted β-sheet flanked by α-helices. One preferred protein is *E. coli* MhpD (SEQ ID NO: 1). Further preferred proteins are listed in Table 2 and, to the extent that the proteins of Table 3 have RQ at positions corresponding to residues 15 and 19 of SEQ ID NO:1, in Table 3. In addition to the proteins listed in Table 2, some further preferred proteins are MhPD1 from *Pseudomonas putida* W619, HPD hydratase MhpD1 from *P. putida*, MhpD from *Azoarcus sp*, strain BH72, 4-OD from *P. putida*, MhpD2 from *P. putida* or 4-OD from *Sphingomonas wittichi*.

Variants of these sequences can also be used and modifications can be made as long as they do not take away the ability of the protein monomers to self-assemble to form protein cage nanostructures of the invention. Therefore, sequences with at least 70%, at least 80%, at least 90%, at least 95%, at least 98% or at least 99% sequence identity to the sequences of SEQ ID NO: 1 and 2 may be used.

Changes made in this way will often be conservative substitutions. Conservative substitutions may be made, for example according to the following table. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other.

ALIPHATIC	Non-polar	G A P
		I L V
	Polar-uncharged	C S T M
		N Q
	Polar-charged	D E
		K R
AROMATIC		H F W Y

However, changes can also be made in a directed fashion to change the self-assembly characteristics of the monomer, in particular to affect the pH at which the protein cage of the invention forms and/or remains stable.

Modifications to sequences can be made by standard recombinant techniques.

Nanostructures

Monomers of the invention self-assemble to form protein cage nanostructures of the invention. Each protein cage includes 60 monomers. These are organised into pentamers containing five monomers each. Four pentamers, i.e. twenty monomers in total, form a crystal asymmetric unit of the nanostructure. Each asymmetric unit forms a curved surface and three asymmetric units make up the protein cage. The protein cage is dodecahedral in shape with classical 532 symmetry, as exhibited by other large protein structures such as viruses. The invention also provides a unit cell comprising three protein cage nanostructures of the invention.

Within each protein cage nanostructure, all of the monomers are preferably identical.

The protein cage nanostructures of the invention are stabilised by coordination of trios of amino acid residues of the same type from three different pentamers by a single ligand. Each trio of amino acids of the same type is located at the three-fold symmetry axis of the nanostructure, where the ligand coordinates them. The chemical nature of the ligand depends on the chemical nature of the side chains of the three amino acids being coordinated. In general, amino acids with negative polarity will be stabilised by positive ions and amino acids with positive polarity will be stabilised by positive ions.

Therefore Arginine and Lysine residues will typically be stabilised by phosphate, pyrophosphate, sulphate, arsenate, vanadate, chloride, bromide, iodide, or other negative ions. Conversely, Histidine and Aspartic acid residues will typically be stabilised by positively charged ions. In the case of Histidine, preferred ions are Zinc

and Nickel; in the case of Aspartic acid, preferred ions are Calcium, Magnesium and Zinc.

The amino acids that are coordinated by the ligands are generally those that correspond to the arginine residue at position 15 of SEQ ID NO: 1 arising from the N-terminal helices of the monomers and the Glutamine residue at position 19 of SEQ ID NO: 1 from the helices that interact around the three-fold axis.

Arg15 may be replaced by Lysine and Gln19 by Arginine, Lysine or another amino acid with a hydrogen-bonding side-chain.

The ligands coordinating any given protein cage nanostructure may be the same or different. For example, for an MhpD-derived protein cage, they may all be phosphate, they may all be sulphate or they may be a mixture of phosphate and sulphate.

In one embodiment, phosphate ligands are provided by triphosphate moieties, for example provided by a nucleotide such as ATP, GTP, TTP, CTP or UTP (adenosine triphosphate, guanidine triphosphate, thymidine triphosphate, cytosine triphosphate or uracil triphosphate); by a dinucleotide such as Ap3A (Adenosine(5')triphospho(5')adenosine), Gp3G, Tp3T, Cp3C, U3pU or a dinucleotide containing any combination of two different bases selected from A, G, T, C and U; or by sodium triphosphate or another triphosphate.

A unique feature of the MhpD-derived protein cages of the invention is that they self-assemble and are stable, in the presence of appropriate ligands, at low pH. This is useful because the cages can be used in non-physiological conditions. In one embodiment, therefore, it is preferred that the protein cages are capable of self-assembly at less than pH 5.0, preferably less than pH 4.0, preferably less than pH 3.0.

In some embodiments, the protein cages will not remain stable at higher pH, for example, they may be dissembled at pH 7.0.

Nevertheless, cages made from different proteins will have different pH characteristics. Some may be naturally stable at pH 7.0. Also, modifications may be made by directed mutation of the sequence of the protein monomers to change the pH range at which the protein cage will form and remain stable. In particular, modifications may be made such that the protein cage will remain stable at pH 5.0 or greater, pH 6.0 or greater or pH 7.0 or greater. In particular, modifications may be directed towards rendering the cages stable at pH 7.0.

Such modifications may be chosen on the basis of an understanding of the properties and location of the amino acids in the sequence. This will typically be achieved by structural modelling or analysis of actual crystal structures. For example, in MhpD, it appears that Asp33 and Asp75 are located such that two Asp33 and two Asp75 residues from subunits of different pentamers come into hydrogen bonding contact (see Fig.6) but that this arrangement will only be stable below the pKa of Aspartic acid; at neutral pH, side chain repulsion would not be expected to allow formation of the protein cage structure. However, modification of Asp33 and/or Asp75 is expected to reduce this side chain repulsion and extend the viable pH range of MhpD-derived protein cages. In particular, Asp33 and/or Asp75 may be substituted by non-polar amino acids such as Alanine or Glycine or polar neutral amino acids like Asparagine. In other proteins, similar modifications can be contemplated, in particular to replace polar amino acids with non-polar or polar uncharged ones.

Additionally, when ATP or another triphosphate is used to supply phosphate ligands, low-pH assembly is not necessary and assembly can proceed at a pH above 4.0, for example a neutral pH, preferably a pH 6.0 to pH 8.0 or up to 9.0.

Uses of Nanostructures of the Invention

Nanostructures of the invention are expected to have multiple applications in both industrial and healthcare/biological scenarios.

In particular, they may be used to encapsulate guest moieties within the nanostructures. This may be achieved by allowing the guest moieties to become encapsulated during the process of production of the nanostructures. For example, reagents can be encapsulated within the nanostructures so that they may react in close proximity and away from the bulk solution in which the nanostructure is contained. Nevertheless, smaller reaction components may be able to enter through the pores (windows) in the structures. Encapsulaion will be useful in, for example, the production of amyloid fibrils (for example from proteins such as insulin, bet2 microglobulin, transthyretin) whose length is defined by the size of the cargo component of the cage.

Guest moieties may also be incorporated for the purpose of targeting them to particular locations. For example, in an industrial context, binding of the cage may be directed towards a particular metal or semiconductor surface whereas, with cages engineered for higher, e.g. neutral, pH, targeting may be carried out with physiological ligands e.g. to other proteins, to particular tissues or cellular components.

Targeting domains may be attached through extension of the amino acid sequence of the monomers such that the monomers are produced as fusion proteins containing the targeting domains, as long as they do not inhibit formation of the nanostructure. This can be achieved by attachment of the targeting domain at a protruding c-n terminus, typically a c-terminus. Globular N-terminal domains such as GB-1 may tend to inhibit particle assemble but may be tolerated if linked through an extended flexible linker domain.

The inclusion of targeting domains and the encapsulation of guest moieties can be combined to target the protein cage's cargo to particular locations. This will be useful in therapeutic scenarios for the delivery of drugs and/or in imaging scenarios for delivery of imaging reagents.

Methods of Production of Nanostructures of the Invention

Individual monomers and pentamers of the invention can be produced by standard recombinant techniques of cloning and gene expression.

The individual pentamers are then allowed to self-assemble down to protein cages of the invention. This is a simple process in that the pentamers self-assemble at the appropriate pH in the presence of the stabilising ligands. Therefore, for example, *E. coli* MhpD will self-assemble at low pH in the presence of phosphate.

The skilled person will be able to determine the appropriate pH, ligand and other conditions for self-assembly of a given monomer. However, preferred protein concentrations may be from 2 to 10 mg/ml. For unmodified *E. coli* MhpD, preferred pH is from 3.0 to 4.0.

As noted above, however, low pH is not necessary when phosphate ligands are supplied by ATP or another triphosphate. In one embodiment, ATP concentration is ramped through the process to avoid any risk of saturation that could inhibit the formation of the cage. Also, where phosphate ligands are used, the concentration of phosphate can be reduced when the nanocages have formed, as high phosphate concentration facilitates their assembly but is not necessary for their persistence.

Where it is desired to incorporate guest moieties into the protein cages, these can be included in the self-assembly solution and incorporated as the protein cages come together. Smaller components can be allowed to diffuse in through the "windows" in the cage.

EXAMPLES

Protein Expression and Purification

The *mhpD* gene was amplified by polymerase chain reaction (PCR) from the *E. coli* K12 (Nova Blue cells) chromosomal DNA using the following primers

Forward primer (5'-3'): G**GAATTC**GATGACGAAGCATACTCTTGAGC (SEQ ID NO: 3)

Reverse primer (5'-3'): GCCCCAAAAGGAAGTCTGTCA**CTCGAG**CGG (SEQ ID NO: 4)

to include restriction sites for EcoR1 and Xho1 respectively (EcoRI site **bold and underlined** in forward primer, XhoI site underlined in reverse primer), that enabled the gene to be ligated into a modified pET22b (Novagen) expression vector encoding a six-histidine tag preceded by a dipeptide (LE) on the C-terminus of the protein and a 56 residue GB-1 solubility tag at the N-terminus. The GB1 tag is linked via a 3C protease (Precision) site that upon cleavage leaves three extra amino acids (GPL) on the protein N-terminus. These tags and processing enzymes are standard tools in molecular biology. It is unlikely that either of the tags is absolutely required and likely that the full-length protein alone will express efficiently. The 6-His tag provides a convenient affinity purification procedure but the over expression is so good that other methods (salt precipitation, ion exchange chromatography) would suffice. The BG1 tag is normally used as a solubility enhancer but the mature MhpD enzyme is very soluble alone.

Protein was expressed from the construct described above in BL21(DE3) cells at 37°C for 3 hrs after induction with 0.5mM IPTG after the cell density reached OD600nm of 0.7. The cells were harvested in a Beckman J6-MC centrifuge at 4200rpm for 25 minutes and resuspended in Buffer A (50mM Tris pH8, 500mM NaCl, 10mM imidazole). Cell lysis was carried out by sonication and cell debris was removed by centrifugation in a Beckman Avanti J25 centrifuge with a JA25.50 rotor for 60 minutes at 20000rpm. Crude lysate was loaded onto a pre-equilibrated Ni sepharose affinity column (GE Healthcare) at 3mL/min. The GB1 tag was removed by 3C protease cleavage (1:100 enzyme:fusion protein) during buffer recirculation (0.2ml/min) on the column at 4°C overnight. Bound MhpD was then eluted with a 10 to 400mM imidazole gradient. The eluted enzyme was further purified by Superdex 200 gel filtration chromatography. The gel filtration column was pre-equilibrated

with Buffer B (50mM Tris pH8 and 100mM NaCl). Eluted MhpD was concentrated to 4.5mg/ml in 30kDa cut-off Centricon centrifugal ultrafilters. The expression levels and progress of the purification was followed with reducing SDS electrophoresis where the fusion protein showed a coomassie blue stained band at Mr 35,000 and the purified product at Mr 30,000 (see Fig. 1). Typically 80mg of protein showing a single stained band on a heavily loaded electrophoresis gel was gained from a 1 litre culture without any special efforts to optimise production.

Crystallisation and Structure Analysis

Initial screening for crystallisation conditions of MhpD, with and without Mg^{2+} present, was conducted using Molecular Dimensions Screens I & II at 4°C using the hanging-drop vapour diffusion method. Drops consisted of 2ul protein (4-8mg/ml) solution mixed in equal proportions with well solution. After prolonged incubation (4 weeks) flat plate-like crystals grew in a number of conditions (Solutions 10,28, and 41 from the 100-solution set available from Molecular Dimensions Ltd for trial crystallisations - Solution 10 comprises 1.0M ammonium dihydrogen phosphate and 0.1M sodium citrate buffer pH5.6. Solution 28 comprises 1.6M sodium/potassium dihydrogen phosphate and 0.1M sodium Hepes buffer pH7.5. solution 41 contains 0.4M ammonium dihydrogen phosphate in water. In all cases the final pH is low). All of these conditions contained high concentrations of phosphate ions and in spite of the manufacturers description were of low pH. Subsequently the same crystal form was grown using 1M Na/K dihydrogen phosphate buffer pH3.2 or 0.75M ammonium sulphate in 100mM formate buffer pH3.2. Crystals were cryoprotected with glycerol or paraffin oil and flash frozen in a nitrogen stream at 100K. Diffraction data were collected to 2.8Å resolution at the European Synchrotron Radiation Facility, Grenoble (beamline X, wavelength X at 100K). Data were indexed (space group H3/R3 a=b=207.35Å c=545.47Å, $\alpha=\beta=90^{\circ}$, $\gamma=120^{\circ}$) and intensities integrated using MOSFLM (Leslie 2006) and scaled using SCALA from the CCP4 suite (CCP4 1994). Phases were determined by molecular replacement with MOLREP (Vagin & Teplyakov 1997) using the pentameric MhpD structure (PDB code: 1sv6) as a search model. The spacegroup was determined to be H3 from

the translation function during molecular replacement. The model was refined with REFMAC5 (CCP4) and manipulated using COOT (Emsley & Cowtan 2004). Rigid body refinement was carried out with each protomer treated as a separate rigid body. Positional and B-factor refinement were carried out using strict NCS constraints due to the limited resolution. Crystallographic statistics are included in Table 1.

The contents of the asymmetric unit (AU) showed that four pentamers were packed together to produce a curved surface (Fig. 2,3). Three AUs define a single sphere (dodecahedron) of MhpD, with 12 pentamers or 60 monomers per sphere. The complete unit cell contained three spheres, that is: 36 pentamers of MhpD or 180 individual protein subunits.

The electron density indicated a bound ligand at the junction of groups of three pentamers (the 3-fold axis) whose coordination geometry was consistent with bound phosphate ions derived from the crystal mother liquor. One phosphate ion is co-ordinated by three arginine residues (Arg15), each arginine supplied by one monomer (Fig. 5). A second phosphate ion is coordinated by three glutamine residues (Gln19) from the three helices which interact around the three-fold axis. A sulphate ion occupies the Arg15 phosphate position when crystals are grown from ammonium sulphate. When viewed down the 3-fold axis these anions lie one above the other along the axis but are approximately 5Å apart. The phosphate coordinated to the Arg15 residues lies towards the interior of the cage, whereas the phosphate coordinated by the Gln19 residues is towards the exterior of the cage. Interactions between the three *N*-terminal helices of each subunit in this region provide a considerable proportion of the stabilizing interface of the protein particle.

The residues of the *N*-terminal helices that are principally involved in the interactions between pentamers are Lys3, His4, Asp12, Arg15, Glu18 and Gln19 which would all be solvent exposed in the pentameric form of MhpD. The side-chain amine of Lys3 forms a hydrogen bond with the backbone carbonyl of Gln52 (NZ-O, ~2.9Å) of the adjacent subunit. His4 stacks against the guanidinium group of Arg54 of the adjacent subunit, which is also stacked against Phe83 of its own subunit on the other face of the side-chain. The charge on Arg54 would be neutralised by the salt-bridge it forms

with Asp85 of its own subunit (OD1-NE, ~2.8Å; OD2-NH2, ~2.8Å). Arg54 also forms an H-bond with Asp49 of the same subunit (OD2-NH1, ~2.9Å). The Asp12 carboxylate forms H-bonds with the backbone carbonyl of Glu18 (OD1-O, 3.1-3.5Å; OD2-O, 2.8-3.1Å) of the adjacent subunit and the side-chain of Arg15 (OD1-NH2 ~2.6Å) of the same subunit. The side-chain of Arg15 forms H-bonding interactions with the side-chains of Glu18 (NH1-OE1, ~3.2Å; NH2-OE1, ~3.4Å) and Gln19 (NE-OE1, 3.0-3.4Å) of the adjacent subunit, as well as the all important binding of one phosphate ion (see below). Gln19 also binds the other phosphate ion at the three-fold axis. These fairly extensive interactions are those between two adjacent subunits at the three-fold axis and so corresponding interactions are replicated two more times to complete the trimeric arrangement. The tetrahedral phosphate ion coordinated by the three Arg15 residues is arranged with the apical oxygen pointed along the centre of the three-fold axis, towards the exterior of the particle. This oxygen is almost equally bound by all three Arg residues (O-NH1, ~2.9Å). The other three oxygens of the phosphate are positioned such that each one is pointed towards one of the adjacent Arg residues (O-NH1, 2.5-3.0Å). The other phosphate at the 3-fold is also arranged so that the apical oxygen is pointed along the axis, towards the interior of the cage. The three remaining oxygen atoms are pointed toward each adjacent Gln19. The apical oxygen is 3.1-3.5Å from the Gln19 NE2 and the other oxygens are ~2.7Å from the Gln19 NE2. The OE1 atoms of the Gln19 side-chains are also only ~3Å from the phosphate oxygens.

The secondary structure of each protein subunit comprises an incomplete barrel comprising a strongly twisted β-sheet flanked by α-helices, belonging to the FAH fold superfamily of the SCOP database and first defined for the enzyme fumarylacetoacetate hydrolase (FAH). The flanking helices are involved in interactions between subunits within and between the pentameric assemblies of subunits and are absent in some MhpD homologues.

Each spherical protein particle has an external diameter of ~20nm and encloses an internal cavity of 15nm diameter. At the centre of each pentamer there is a 2nm diameter window passing completely through the protein layer, lined with ten basic

residues, (Arg114, Arg172) from each subunit (see Fig. 4). A further, but less pronounced penetration of the protein layer occurs at the two-fold axes of the particle. The opening can be conveniently considered to be rectangular with dimensions of 5 x 1nm, but in fact the long edges are highly corrugated with substantial narrowing towards the axis. The long edges are each defined by termini of three long and three short helices while the short edges derive from corner loops of the β-barrel. The constriction close to the axis brings together two Asp33 and two Asp75 residues from subunits of different pentamers into hydrogen bonding contact(see Fig. 6). This arrangement is only expected to be stable below the pKa of aspartic acid. At neutral pH the sidechain repulsion would not be expected to allow formation of the dodecahedral particle.

The C-terminal region of each subunit, including the 6-His tag, project from the outer surface of the particle and the absence of electron density suggests that they are highly mobile. The protein N-termini are located towards the inner face of the particle.

Gel filtration Chromatography and Laser Light Scattering

10µl of 4mg/ml MhpD was mixed with 90µl of 1M Na/KH_2PO_4 buffer pH3.2 and gel filtered through a 1 x 30cm column of Superdex 200 (GE Healthcare) and the eluate monitored at 280nm. A single symmetrical peak eluted at ~9ml (excluded vol 7.3ml) indicating a single high molecular weight species. In the same conditions, the GB1-MhpD fusion protein showed a small sharp peak at 12 ml and a broad peak spanning the 8-12ml region suggesting a complex mix of larger oligomers. When the experiment was repeated with MhpD at pH4, two peaks emerged, the first at 9ml and the second at 12ml corresponding to molecules with Mr in the region of 2Mda and 150Kda. Gel filtration of this material in buffer B showed a single symmetrical peak eluting at 12ml. When these experiments were repeated and the eluate was passed through a Wyatt Technology multi-angle laser-light scatterer the peak eluting at 8ml showed an Mr of ~1.8Mda and no evidence of polydispersity, corresponding to a particle of the same dimensions as the dodecahedral ball observed in the X-ray

analysis (see Fig. 7). The peak eluting at 12ml showed an Mr of 150,000 corresponding to the known homopentamer of MhpD. When the MhpD particle assembly at pH 3.2 was carried out in the presence of 2mg/ml horse cytochrome c and the gel filtration was monitored at 280nm and 405nm (the λmax for haem in these conditions) peaks at both wavelengths were observed at 9 and 20ml. This result was consistent with some cytochrome c being trapped within the particle during assembly (see Fig. 8). Particles formed in the presence of amyloidogenic proteins that aggregate at low pH (insulin, bet2 microglobulin, transthyretin) limit the extent of the polymerization process that otherwise proceeds to produce very long structures.

Formation of MhpD cages in solution and entrapment

Gel filtration of MhpD on Superdex 200 (10/30) in acidic phosphate buffer produces a symmetrical absorbance peak at 8.5ml, a little after the column void volume, suggesting that a protein particle much like that observed in the crystal is also formed in solution. Dispersal of the concentrated MhpD stock in the low pH buffer and loading took as little as 2 minutes, followed by the 16 minute elution time. This suggests that the particles form rapidly and maintain their stability during the dilution experienced during gel filtration. SEC-MALLS estimates a molecular weight of 1.64 MDa, close to that expected from the crystallographic results (12 pentamers of MhpD would be ~1.8 MDa). The protein cages must first be formed with at least 1 M phosphate, pH3.2 but can then maintain their stability at lower phosphate concentrations. The protein solubility is limited to ~2mg/ml in these conditions. However, once particles are formed the phosphate concentration of the column eluent can be reduced to 0.2M, pH3.2 and still give only one symmetrical peak at 8.5ml indicating complete retention of the caged form. At 0.1M phosphate the ratio of protein cages to pentameric MhpD is approximately 90:10 whilst at 0.05M phosphate the ratio is in the order of 60:40 dodecamers:pentamers. Addition of ~10% glycerol to buffers allows the formation of the particles at higher concentrations of MhpD, possibly by reducing self-association of particles.

Gel filtration of MhpD particles produced in the presence of cytochrome c showed Soret band absorbance from liganded haem at 406nm eluting at the same volume as the protein particle, suggesting entrapment during particle formation. Cytochrome c was shown to be very soluble in the high phosphate low pH buffer and so aggregation onto the outer surface of the MhpD cage during gel filtration is not likely to be the explanation for the co-eluting peaks. Cytochrome c shows no sign of aggregation and gel filters in these conditions as a single low MW species with no sign of excluded volume material. We are also encouraged that we have shown true encapsulation by the fact that cytochrome c is a very basic protein and at pH3.2 MhpD also carries a net positive charge, suggesting strong repulsion even in the presence of phosphate counter-ion screening and little chance of non-specific binding of sufficient affinity to withstand gel filtration. The clustering of cytochrome molecules or other cargoes in solution could lead to greater encapsulation in some cases and no entrapment in others.

BLAST Searching to identify further potential cage-forming proteins

Using the SIB BLAST Network Service (see Altschul et al 1997), BLAST searches were carried out to identify further potential cage-forming proteins. First, all 269 residues of *E.coli* MhpD were used as the query sequence with the default gap penalty, where in the worst match there were still 39 out of 154 residues identical. Remarkably, this enzyme is clearly listed as an FAH and demonstrates that tertiary structure is much more closely conserved than sequence.

Second, the query sequence was cut to the 53 residues of the N-terminal domain that is crucial for phosphate binding and docking of pentamers at the three-fold axis. There were 130 matches with the default gap penalty. Down to match 30, the hits (mainly MhpDs from different E. coli strains) matched expectations for cage forming capability based on the observations reported above. Below match 30, the conservation started to drop and the presence or absence of the critical residues (R15 and Q19) fluctuated in a way not matching the overall likeness.

It is therefore possible to screen available sequences through a number of layers of stringency when trying to decide whether they can form balls.

1. Global homology placing a protein in the FAH family.
2. Presence of an N terminal domain
3. Presence of known MhpD residues that coordinate phosphates (R15,Q19)
4. Presence of related residues that might be equally effective, e.g. K15,T19
5. Inverted combinations of effective residues. Q15,K19
6. Possibly superior combinations e.g. K15,K19 - R15,R19

As an example, some information for the first 30 matches from the 53-residue BLAST is presented in Table 2.

We expect that the Arg15 and Gln19 residues of MhpD (RQ) could be conservatively replaced and still coordinate the phosphate ions to facilitate nanoparticle formation. For instance, Arg15 could be replaced by a Lys residue and Gln19 could be replaced by an Arg or Lys residue, or indeed any residue with a suitable hydrogen-bonding side-chain. In which case, two further proteins of interest from the NTD BLAST search are MhpD2 from *P. putida* and 4-OD from *Sphingomonas wittichi* which have KK and RR in place of the archetypal RQ, respectively. Thus, as the structure described herein demonstrates that the N-terminus is essential for nanoparticle formation it would seem that only a handful of MhpD homologues, in particular those from other strains of E. coli, would in all probability be capable of assembling into this higher oligomeric form. A selection of sequence segments from different MhpD homologues are displayed in Table 3, showing some with very high sequence identity to MhpD but lacking the critical phosphate binding residues and others that retain these residues in the presence of much lower overall sequence similarity.

Cage formation with ATP

Guided by the disposition of the bound phosphate anions on the three-fold axis we investigated whether a triphosphate moiety, such as that found in ATP, could be bound and provide a phosphate spacing consistent with the observed structure.

Aliquots (2µl) of 40mM ATP in 200mM Tris/HCl buffer pH8 were added to 30µl of MhpD (22mg/ml) dissolved in 50mM Tris buffer pH8 containing 100mM NaCl and 1mM EDTA over a period of one hour until the ATP concentration reached 20mM. The sample was then diluted to 200µl with 20mM ATP solution and loaded onto a Superdex 200 gel filtration column previously equilibrated with 50mM Tris/HCl buffer pH8 containing 100mM NaCl, 1mM EDTA and 2mM ATP. The absorbance of eluting materials was monitored by UV absorption at 280nm. The ATP concentration in the eluant was limited to 2mM in view of its high UV absorbance and limitations of the detector. Three peaks eluted from the column, a high molecular weight species at 8.5ml corresponding to the protein cage, a peak of approximately equivalent size at 12ml corresponding to the pentamer of MhpD and a large peak at 20ml corresponding to excess ATP in the sample.

The heavy dilution of the protein and the ATP solution during analysis leads us to believe that following the initial ATP treatment substantially more of the MhpD protein was in the high molecular weight form, but dissociated during the gel filtration. Therefore, we intend to repeat the analysis using 10mM ATP in the eluant but to employ MALLS (multi-angle laser light scattering) and refractive index detection to monitor the protein elution and avoid the complications due to high UV absorption. To investigate further, we also intend to examine formation of cages in the presence of another triphosphate, namely sodium triphosphate at pH8. We will gel filter MhpD solutions in Tris/HCl buffer at pH 8 in the presence of different concentrations of sodium triphosphate, using UV absorption to monitor the different sizes of particles produced.

REFERENCES

PATENT LITERATURE

WO03/096990 (Montana State University/Young & Douglas) – Protein cages for the delivery of medical imaging and therapy

X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS)

New York structural genomix research consortium (NYSGXRC); PDB code 1sv6; "Crystal structure of 2-hydroxypentadienoic acid hydratase from Escherichia coli."; available June 2004 from the PDB data bank. Fedorov, A.A., Fedorov, E.V., Sharp, A., Almo, S.C., Burley, S.K.

JOURNAL PUBLICATIONS

Emsley, P. & Cowtan, K. (2004). COOT: model building tools for molecular graphics. Acta Crystallog. Sect. D, 60, 2126–2132.

CCP4: Collaborative Computing Project Number 4. (1994).The CCP4 suite: programs for protein crystallography. Acta Crystallog. Sect. D, 50, 760–763.

Diaz, E., Ferrandez, A., Prieto, M. A. & Garcia, J. L. (2001). Biodegradation of aromatic compounds by Escherichia coli. Microbiol. Mol. Biol. Rev. 65, 523–569.

Ferrandez A, Garcia JL and Diaz E. 1997. Genetic characterization and expression in heterologous hosts of the 3- (3-hydroxyphenyl)propionate catabolic pathway of Escherichia coli K-12. J. Bacteriol. 179, 2573-2581.

Leslie, A.G. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62**, 48-57.

Mann S and Meldrun FC. (1991) Controlled synthesis of Inorganic Materials Using Supramolecular Assemblies. Adv,Mater. 3, 316-318.

Manjasetty, B. A., Niesen, F. H., Delbruck, H., Gotz, F.,Sievert, V., Bussow, K. *Et al.* (2004). X-ray structure of fumarylacetoacetate hydrolase family member Homo sapiens. J Biol. Chem. 385, 935–942.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallog. Sect. D, 53, 240–255.

Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia,C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536–540.

Niemeyer CM. Nanoparticles, Proteins, and Nucleic Acids: Biotechnology Meets Materials Science (2001) Angew. Chem. Int. Ed, 40, 4128-4158.

Pollard, J. R. & Bugg, T. D. (1998). Purification,characterisation and reaction mechanism of monofunctional 2-hydroxypentadienoic acid hydratase from. Escherichia coli. Eur. J. Biochem. 251, 98-106.

Prieto, M. A., Diaz, E. & Garcia, J. L. (1996). Molecular characterization of the 4-hydroxyphenylacetate catabolic pathway of Escherichia coli W: engineering a mobile aromatic degradative cluster. J. Bacteriol. 178, 111-120.

Uchida M, Klem MT, Allen M, Suci P, Flenniken M, Gillitzer E, Varpness Z, Lars O. Liepold LO, Mark Young M, and Trevor Douglas T. (2007) Biological Containers: Protein cages as Multifunctional Nanoplatforms. Adv. Mater. 19, 1025-1042.

Vagin, A. & Teplyakov, A. (1997) J. Appl. Crystallogr. 30, 1022-1025.

Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. Nucleic Acids Res. 25:3389-3402(1997).

TABLE 1

Processing and Refinement Statistics for MhpD (60mer)

a = b(Å)	207.35
c(Å)	545.47
Space group	H3
Resolution (Å)	2.80
Total Reflections	453517
Unique Reflections	210704
Rmerge (%) (outer shell)	14.0 (52.6)
Completeness %	97.9 (95.7)
I/σ(I) (outer shell)	6.7 (1.8)
Multiplicity	2.2 (2.1)
Rfactor (%)	23.7
Rfree (%)	24.8
Average Bfactor (Å2)	37
Number of protein atoms	39404
Number of phosphates	8
RMSD bonds (Å)	0.016
RMSD angles (°)	1.87

TABLE 2

Results of BLAST Searches to Identify Further Potential Cage-Forming Proteins

(Top 30 matches with 53 residue N-terminal domain, all including Arg and Gln in positions corresponding to Arg15 and Gln19 in SEQ ID NO:1)

```
Query:  53 AA
Query:  MTKHTLEQLAADLRRAAEQGEAIAPLRDLIGIDNAEAAYAIQHINVQHDVAQG
Program: NCBI BLASTP 2.2.17 [Aug-26-2007]
Database: UniProtKB, Release 15.4

List of potentially matching sequences

Db AC/Description/Score/E-value

sp B1LIN5 MHPD_ECOSM 2-keto-4-pentenoate hydratase (EC 4.2.1.80)... 130 2e-29
sp P77608 MHPD_ECOLI 2-keto-4-pentenoate hydratase (EC 4.2.1.80)... 130 2e-29
sp B7NK05 MHPD_ECO7I 2-keto-4-pentenoate hydratase (EC 4.2.1.80)... 130 2e-29
tr B1J0S9 ECOLC 4-oxalocrotonate decarboxylase (EC 4.1.1.77) [Ec... 130 2e-29
sp B5Z2Q5 MHPD_ECO5E 2-keto-4-pentenoate hydratase (EC 4.2.1.80)... 128 1e-28
sp Q8XEC1 MHPD_ECO57 2-keto-4-pentenoate hydratase (EC 4.2.1.80)... 128 1e-28
sp B7L506 MHPD_ECO55 2-keto-4-pentenoate hydratase (EC 4.2.1.80)... 128 1e-28
tr C4W120 ECOLX 4-oxalocrotonate decarboxylase (EC 4.1.1.77) [EC... 128 1e-28
tr C3TMV2 ECOLX 2-keto-4-pentenoate hydratase [ECs0405] [Escheri... 128 1e-28
tr B6ZQL8 ECO57 2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD... 128 1e-28
tr B3XH93 ECOLX 2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD... 128 1e-28
tr B3WTV0 ECOLX 2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD... 128 1e-28
tr B3I9B9 ECOLX 2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD... 128 1e-28
tr B3HJZ9 ECOLX 2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD... 128 1e-28
tr B3C496_ECO57 2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD... 128 1e-28
tr B3BQE6_ECO57 2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD... 128 1e-28
tr B3B1F8 ECO57 2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD... 128 1e-28
tr B3AM61_ECO57 2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD... 128 1e-28
tr B3A819_ECO57 2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD... 128 1e-28
tr B2PTI9_ECO57 2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD... 128 1e-28
tr B2P712_ECO57 2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD... 128 1e-28
tr B2NZ37_ECO57 2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD... 128 1e-28
sp A7ZWZ7 MHPD_ECOHS 2-keto-4-pentenoate hydratase (EC 4.2.1.80)... 125 6e-28
```

sp A7ZI97 MHPD_ECO24 2-keto-4-pentenoate hydratase (EC 4.2.1.80)... 125 6e-28
tr B6HZX6_ECOSE 2-keto-4-pentenoate hydratase [ECSE_0375] [Esche... 125 6e-28
tr B2NAE7 ECOLX 2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD... 125 6e-28
sp B7N8Q7 MHPD_ECOLU 2-keto-4-pentenoate hydratase (EC 4.2.1.80)... 125 1e-27
sp B7M2Z8 MHPD_ECO8A 2-keto-4-pentenoate hydratase (EC 4.2.1.80)... 124 2e-27
sp Q3Z556 MHPD_SHISS 2-keto-4-pentenoate hydratase (EC 4.2.1.80)... 123 3e-27
sp B7MPB7 MHPD_ECO81 2-keto-4-pentenoate hydratase (EC 4.2.1.80)... 122 8e-27

sp B1LIN5 **2-keto-4-pentenoate hydratase (EC 4.2.1.80)** 269 AA
MHPD_ECOSM **(2-hydroxypentadienoic acid hydratase) [mhpD]**
[Escherichia coli (strain SMS-3-5 / SECEC)]

Score = 130 bits (294), Expect = 2e-29
Identities = 53/53 (100%), Positives = 53/53 (100%)

sp P77608 **2-keto-4-pentenoate hydratase (EC 4.2.1.80)** 269 AA
MHPD_ECOLI **(2-hydroxypentadienoic acid hydratase) [mhpD]**
[Escherichia coli (strain K12)]

Score = 130 bits (294), Expect = 2e-29
Identities = 53/53 (100%), Positives = 53/53 (100%)

sp B7NK05 **2-keto-4-pentenoate hydratase (EC 4.2.1.80)** 269 AA
MHPD_ECO7I **(2-hydroxypentadienoic acid hydratase) [mhpD]**
[Escherichia coli O7:K1 (strain IAI39 / ExPEC)]

Score = 130 bits (294), Expect = 2e-29
Identities = 53/53 (100%), Positives = 53/53 (100%)

tr B1J0S9 **4-oxalocrotonate decarboxylase (EC 4.1.1.77) [EcolC_3275]** 269 AA
B1J0S9_ECOLC **[Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks)]**

Score = 130 bits (294), Expect = 2e-29
Identities = 53/53 (100%), Positives = 53/53 (100%)

sp B5Z2Q5 **2-keto-4-pentenoate hydratase (EC 4.2.1.80)** 269 AA

MHPD_ECO5E **(2-hydroxypentadienoic acid hydratase) [mhpD]**

[Escherichia coli O157:H7 (strain EC4115 / EHEC)]

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

sp Q8XEC1 **2-keto-4-pentenoate hydratase (EC 4.2.1.80)** 269 AA

MHPD_ECO57 **(2-hydroxypentadienoic acid hydratase) [mhpD]**

[Escherichia coli O157:H7]

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

sp B7L506 **2-keto-4-pentenoate hydratase (EC 4.2.1.80)** 269 AA

MHPD_ECO55 **(2-hydroxypentadienoic acid hydratase) [mhpD]**

[Escherichia coli (strain 55989 / EAEC)]

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr C4W120 **4-oxalocrotonate decarboxylase (EC 4.1.1.77) [ECBDDRAFT_2222]** 269 AA

C4W120_ECOLX **[Escherichia coli BL21(DE3)]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr C3TMV2 **2-keto-4-pentenoate hydratase** 269 AA

C3TMV2_ECOLX **[ECs0405] [Escherichia coli]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr B6ZQL8 **2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD] [Escherichia** 269 AA

B6ZQL8_ECO57 **coli O157:H7 str. TW14588]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr B3XH93 **2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD] [Escherichia** 269 AA

B3XH93_ECOLX **coli 101-1]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr B3WTV0 **2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD] [Escherichia** 269 AA

B3WTV0_ECOLX **coli B171]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr B3I9B9 **2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD] [Escherichia** 269 AA

B3I9B9_ECOLX **coli E22]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr B3HJZ9 **2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD] [Escherichia** 269 AA

B3HJZ9_ECOLX **coli B7A]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr B3C496 **2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD] [Escherichia** 269 AA

B3C496_ECO57 **coli O157:H7 str. EC508]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr B3BQE6 **2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD] [Escherichia** 269 AA

B3BQE6_ECO57 **coli O157:H7 str. EC869]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr B3B1F8 **2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD] [Escherichia** 269 AA

B3B1F8_ECO57 **coli O157:H7 str. EC4501]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr B3AM61 **2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD] [Escherichia** 269 AA

B3AM61_ECO57 **coli O157:H7 str. EC4486]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr B3A819 **2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD] [Escherichia** 269 AA

B3A819_ECO57 **coli O157:H7 str. EC4401]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr B2PTI9 **2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD] [Escherichia** 269 AA

B2PTI9_ECO57 **coli O157:H7 str. EC4076]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr B2P712 **2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD] [Escherichia** 269 AA

B2P712_ECO57 **coli O157:H7 str. EC4113]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

tr B2NZ37 **2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD] [Escherichia** 269 AA

B2NZ37_ECO57 **coli O157:H7 str. EC4196]**

Score = 128 bits (288), Expect = 1e-28

Identities = 52/53 (98%), Positives = 53/53 (100%)

sp A7ZWZ7 **2-keto-4-pentenoate hydratase (EC 4.2.1.80)** 269 AA

MHPD_ECOHS **(2-hydroxypentadienoic acid hydratase) [mhpD]**

[Escherichia coli O9:H4 (strain HS)]

Score = 125 bits (283), Expect = 6e-28

Identities = 51/53 (96%), Positives = 52/53 (98%)

sp A7ZI97 **2-keto-4-pentenoate hydratase (EC 4.2.1.80)** 269 AA

MHPD_ECO24 **(2-hydroxypentadienoic acid hydratase) [mhpD]**

[Escherichia coli O139:H28 (strain E24377A / ETEC)]

Score = 125 bits (283), Expect = 6e-28

Identities = 51/53 (96%), Positives = 52/53 (98%)

tr B6HZX6 **2-keto-4-pentenoate hydratase [ECSE_0375] [Escherichia coli** 271 AA

B6HZX6_ECOSE **(strain SE11)]**

Score = 125 bits (283), Expect = 6e-28

Identities = 51/53 (96%), Positives = 52/53 (98%)

tr B2NAE7 **2-keto-4-pentenoate hydratase (EC 4.2.1.-) [mhpD] [Escherichia** 269 AA
B2NAE7_ECOLX **coli 53638]**

Score = 125 bits (283), Expect = 6e-28
Identities = 52/53 (98%), Positives = 52/53 (98%)

sp B7N8Q7 **2-keto-4-pentenoate hydratase (EC 4.2.1.80)** 269 AA
MHPD_ECOLU **(2-hydroxypentadienoic acid hydratase) [mhpD]**
[Escherichia coli O17:K52:H18 (strain UMN026 / ExPEC)]

Score = 125 bits (281), Expect = 1e-27
Identities = 51/53 (96%), Positives = 52/53 (98%)

sp B7M2Z8 **2-keto-4-pentenoate hydratase (EC 4.2.1.80)** 269 AA
MHPD_ECO8A **(2-hydroxypentadienoic acid hydratase) [mhpD]**
[Escherichia coli O8 (strain IAI1)]

Score = 124 bits (279), Expect = 2e-27
Identities = 50/53 (94%), Positives = 52/53 (98%)

sp Q3Z556 **2-keto-4-pentenoate hydratase (EC 4.2.1.80)** 269 AA
MHPD_SHISS **(2-hydroxypentadienoic acid hydratase) [mhpD] [Shigella**
sonnei (strain Ss046)]

Score = 123 bits (277), Expect = 3e-27
Identities = 50/53 (94%), Positives = 52/53 (98%)

sp B7MPB7 **2-keto-4-pentenoate hydratase (EC 4.2.1.80)** 269 AA
MHPD_ECO81 **(2-hydroxypentadienoic acid hydratase) [mhpD]**
[Escherichia coli O81 (strain ED1a)]

Score = 122 bits (274), Expect = 8e-27
Identities = 50/53 (94%), Positives = 51/53 (96%)

Lambda	K	H
0.336	0.170	0.565

Gapped

Lambda	K	H
0.299	0.0710	0.270

Matrix: BLOSUM80

Gap Penalties: Existence: 10, Extension: 1

TABLE 3

Partial sequence alignments of MhpD and selected homologues

Protein[a]	*Identity (%)*	*N-terminus (3-fold)*[b]	*Asp33 (2-fold)*	*Asp75 (2-fold)*
MHPD_ECOLI	100	**MTKHTLEQLAADLRRAAEQG**	**GIDNA**	**GVDQP**
MHPD_SHISH	98	**MTKHTLEQLAADLRRAAEQG**	**GIDHA**	**GVDQP**
C1M7G9_9ENTR	91	**MTNRNLEQLAADLRQAEEKG**	**GVDNA**	**GVDQP**
MHPD_KLEP7	81	**SLDALARQLRDAEQSG**	**GVDNA**	**GVNQP**
C1DRI8_AZOVD	71	**VTQQTLEQLAAALRGAEANG**	**GEDNG**	**GVDQP**
B1J6Y5_PSEPW	71	**TLERLAADLRRAQQQG**	**GAENA**	**GVDQP**
MHPD1_PSEPU	67	**ISQETLERLAADLRRAEQQG**	**GAENG**	**GVDQP**
Q397R1_BURS3	51	**MNVDQIQHVADRLRRAESTG**	**AALSI**	**GVDQP**
B1KJL1_SHEWM	41	**IEQLGLELYRALRNQ**	**PQLTL**	**GVHQP**

(a) MHPD_ECOLI = MhpD from *E. coli* K12; MHPD_SHISH = MhpD from *Shigella sonnei* Ss046; C1M7G9_9ENTR = HPD hydratase from *Citrobacter* sp. 30_2; MHPD_KLEP7 = MhpD from *Klebsiella pneumoniae* strain ATCC 700721; C1DRI8_AZOVD = 4-OD from *Azotobacter vinelandii* strain ATCC BAA-1303; B1J6Y5_PSEPW = 4-OD from *Pseudomonas putida* strain W619; MHPD1_PSEPU = MhpD from *Pseudomonas putida*; Q397R1_BURS3 = 4-OD from *Burkholderia* sp. strain 383; B1KJL1_SHEWM = 4-OD from *Shewanella woodii* strain ATCC 51908 (psychrophile).

(b) Identical residues to those in *E. coli* K12 MhpD are shown in bold. Residues presumably essential for nanoparticle formation are underlined.

CLAIMS

1. A protein cage nanostructure composed of twelve pentamers, each pentamer containing five protein monomers;

said pentamers forming a protein cage nanostructure that is dodecahedral in shape with 532 symmetry;

said nanostructure being stabilised by coordination, at each of the three-fold symmetry axes, of three amino acid residues of the same type from three different pentamers, by a ligand;

each monomer comprising a protein belonging to the FAH fold superfamily whose secondary structure includes an incomplete barrel comprising a strongly twisted β-sheet flanked by α-helices.

2. A protein cage nanostructure of claim 1 wherein each of the three-fold symmetry axes is additionally coordinated by a further ligand that coordinates three further amino acids from said three pentamers.

3. A protein cage nanostructure of claim 1 or 2 wherein

(a) said amino acid of claim 1 is Arginine and said ligands are negatively charged ions; or

(b) said amino acid of claim 1 is Lysine and said ligands are negatively charged ions; or

(c) said amino acid of claim 1 is Histidine and said ligands are positively charged ions;

(d) said amino acid of claim 1 is or Aspartic acid or and said ligands are positively charged ions; or

(e) said amino acid of claim 2 is Glutamine, Arginine or Lysine or another amino acid with a hydrogen-bonding side-chain.

4. A protein cage nanostructure of claim 3(a), (b) or (e), wherein said ions are the same or different and are selected from phosphate, pyrophosphate, sulphate, arsenate vanadate, chloride, bromide and iodide ions;

or of claim 3(c) wherein said ions are the same or different and are selected from Zinc and Nickel ions;

or of claim 3(d) wherein said ions are the same or different and are selected from Calcium, Magnesium and Zinc ions.

5. A protein cage nanostructure of claim 4 wherein each monomer is an *E. coli* MhpD protein of SEQ ID NO: 1 or a variant with at least 80% sequence identity to SEQ ID NO: 1 and capable of self-assembly into said protein cage nanostructure via said coordination of said Arginine residues with said negatively charged ions; wherein said Arginine residues are at position 15 of the sequence of SEQ ID NO: 1 or the equivalent position in a variant sequence.

6. A protein cage nanostructure of claim 5 wherein each three-fold axis is additionally coordinated by further negative ions that coordinate Glutamine residues at position 19 of the sequence of SEQ ID NO: 1 or the equivalent position in a variant sequence.

7. A protein cage nanostructure of claim 2(a) or (e), 4, 5 or 6, wherein each of said ligands is a phosphate ion, or wherein each of said ligands is a pyrophosphate ion, or wherein each of said ligands is a sulphate ion.

8. A protein cage nanostructure of claim 7 wherein each of said ligands is a phosphate ion.

9. A protein cage nanostructure of claim 8 wherein said phosphate ions are provided by triphosphate moieties.

10. A protein cage nanostructure of claim 9 wherein said triphosphate moieties are provided by ATP.

11. A protein cage nanostructure of any one of claims 1 to 4 or 7 to 10 wherein each monomer is a molecule of a protein listed in Table 2, or of MhPD1 from *Pseudomonas putida* W619, HPD hydratase MhpD1 from *P. putida*, MhpD from *Azoarcus sp*, strain BH72, 4-OD from *P. putida*, MhpD2 from *P. putida* or 4-OD from *Sphingomonas wittichi*.

12. A protein cage nanostructure of any one of the preceding claims that undergoes of self-assembly at pH 4.0 or less.

13. A protein cage nanostructure of any one of the preceding claims that is disassembled at pH 7.0.

14. A protein cage nanostructure of any claims 1 to 11, wherein the amino acid sequence of the monomers is modified such that the nanostructure is stable at a higher pH than with the unmodified monomers.

15. A protein cage nanostructure of claim 4 wherein:

in SEQ ID NO: 1, one or more of residues Asp33 and Asp75 is replaced with a non-polar amino acid or a polar uncharged amino acid; or

in a variant sequence, one or more of the equivalent amino acids is replaced with a non-polar amino acid or a polar uncharged amino acid;

such that electrostatic repulsion is reduced and the nanostructure is stable at a higher pH than with the unmodified monomers.

16. A protein cage nanostructure of claim 14 or 15 wherein the nanostructure is stable at pH 7.0.

17. A protein cage nanostructure of any one of the preceding claims, wherein each of the monomers is identical.

18. A protein cage nanostructure of any one of the preceding claims, in which is encapsulated one or more guest moieties.

19. A protein cage nanostructure of claim 18, wherein said guest moieties are reagents for a reaction that will take place within the nanostructure or the products of a reaction that has taken place within the nanostructure; or wherein said guest molecules or ions comprise therapeutic or imaging agents.

20. A protein cage nanostructure of any one of the preceding claims which comprises, or is attached to, one or more targeting moieties.

21. A method of producing a protein cage nanostructure of any one of the preceding claims, comprising contacting said monomers with said ligands in solution at a pH at which said monomers will self-assemble to form said nanostructure.

22. A method of claim 21, wherein said pH is less than 4.0.

23. A method of any one of claims 21 to 23, wherein the ligands are phosphate ions provided by ATP or sodium triphosphate and said pH is greater than 4.0, preferably from 6.0 to 8.0.

24. A method of claim 23, wherein the concentration of ATP is gradually increased during the process.

25. A method of any one of claims 21 to 24, wherein the ligands are phosphate ions and the concentration of phosphate is reduced once the nanostructures have formed.

26. A method of any one of claims 21 to 25, wherein one or more guest moieties are included in the solution and encapsulated into the nanostructures; or wherein one or more targeting moieties are attached to said nanostructure.

27. A method of claim 23, wherein (a) said guest moieties are reagents for a reaction that will take place within the nanostructure; or (b) said guest moieties comprise therapeutic or imaging agents.

28. A protein cage nanostructure of claim 19 or a method of claim 27, wherein said reagents are reagents for the synthesis of amyloid fibrils.

Fig. 1

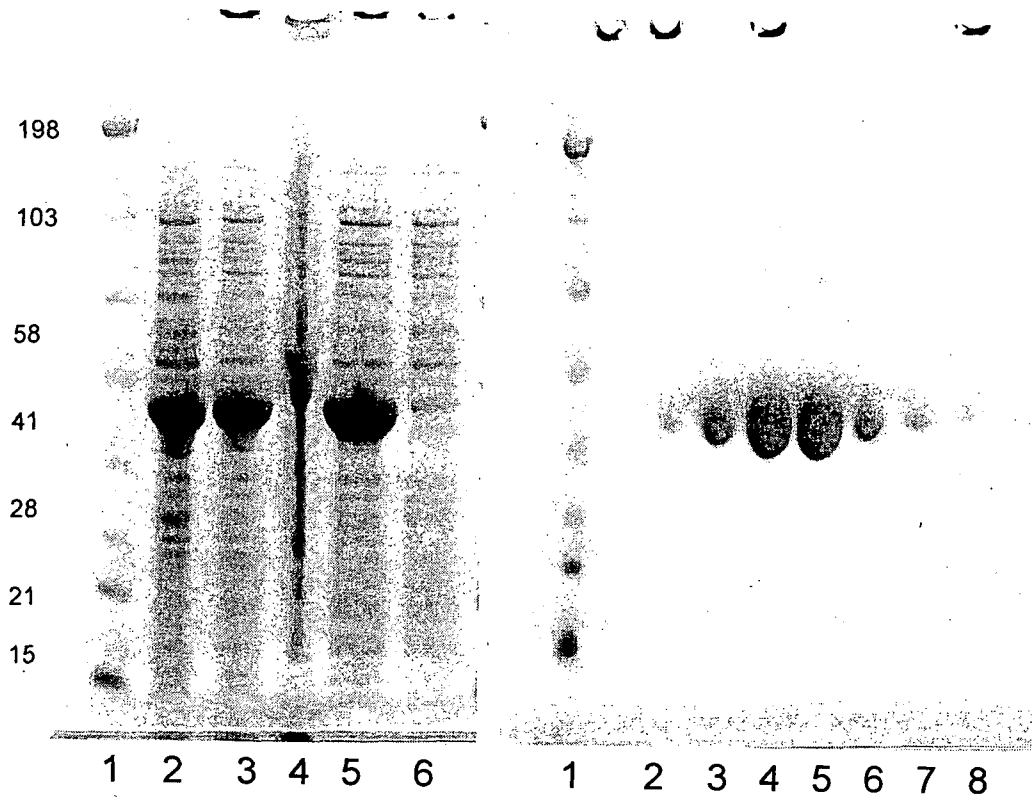


Fig. 2

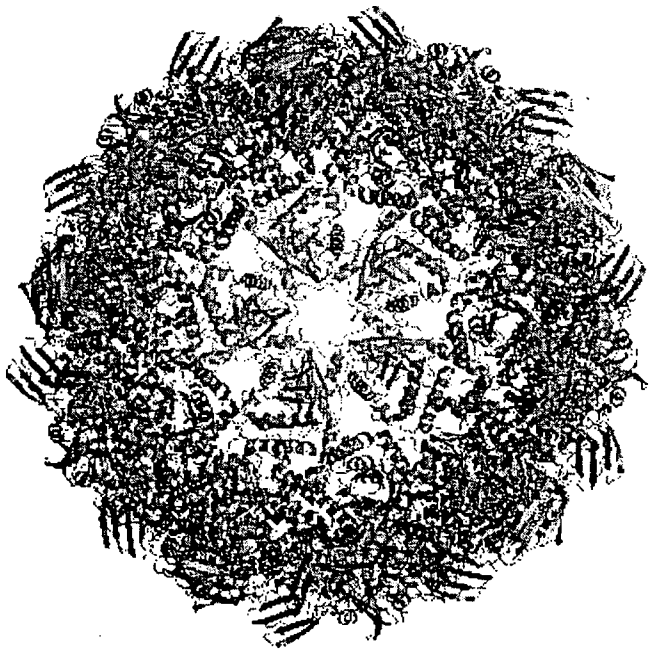


Fig. 3



Fig. 4

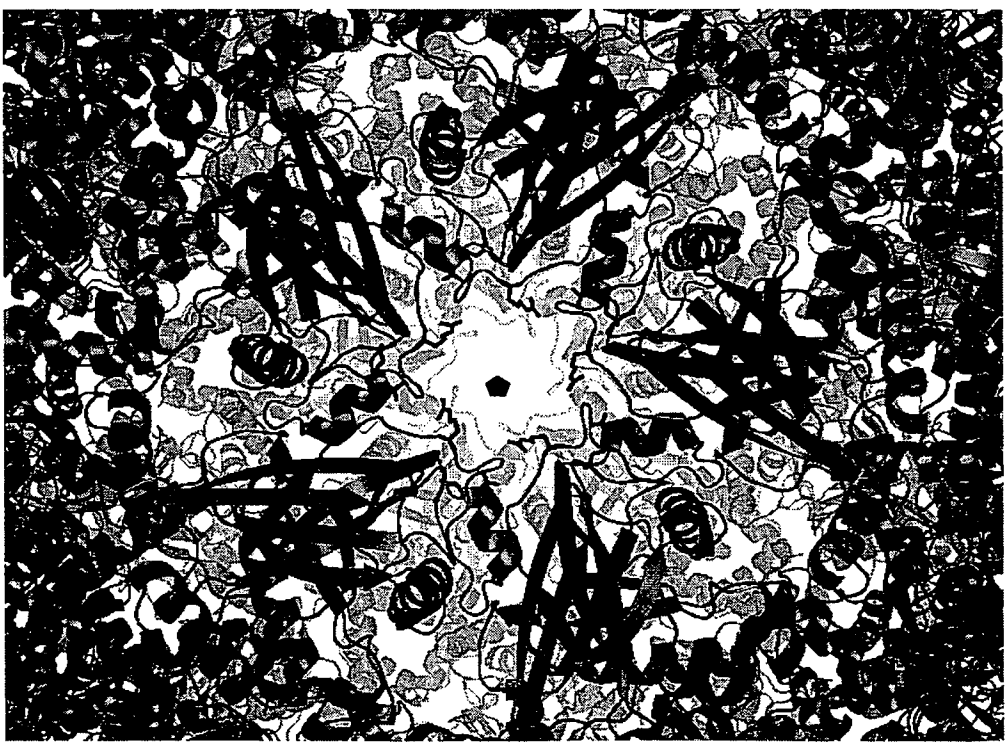


Fig. 5



Fig. 6

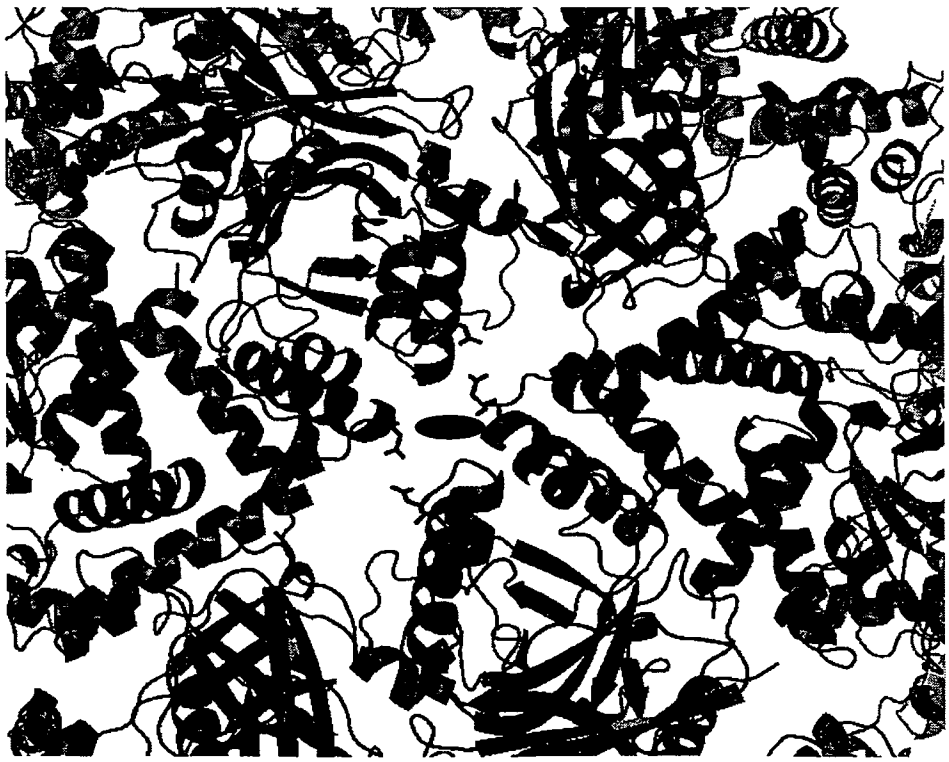


Fig. 7

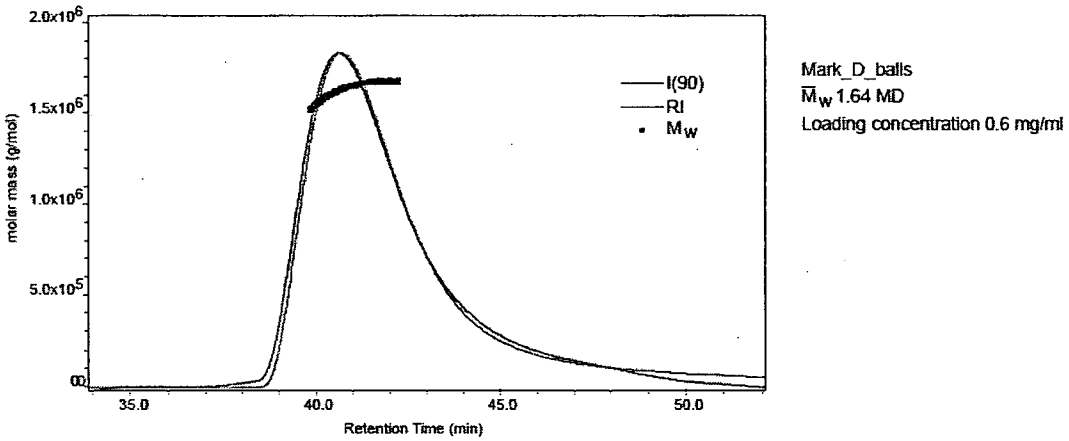

2.0x10^6
1.5x10^6
1.0x10^6
5.0x10^5
00
molar mass (g/mol)
35.0
40.0
45.0
50.0
Retention Time (min)
I(90)
RI
Mw
Mark_D_balls
Mw 1.64 MD
Loading concentration 0.6 mg/ml

Fig. 8

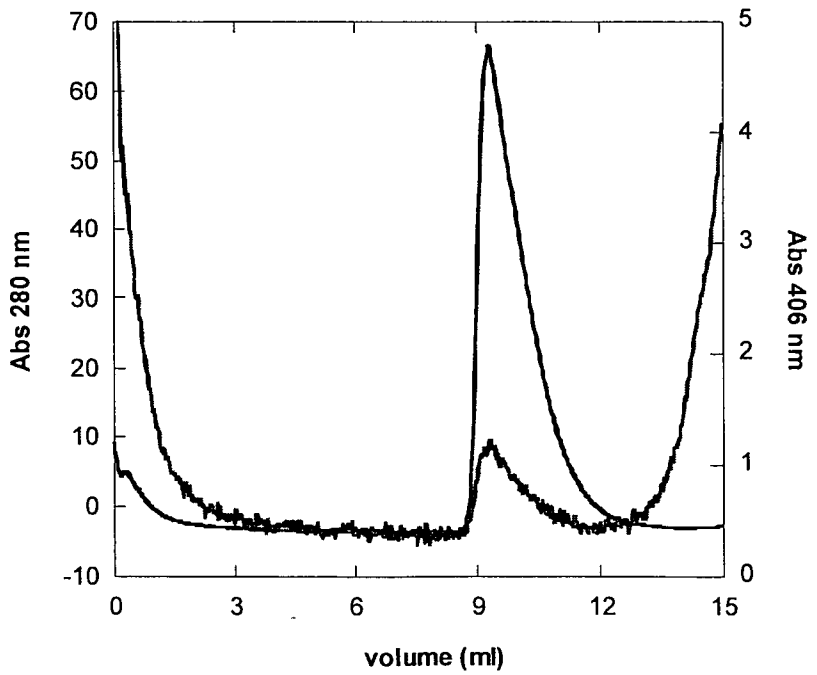
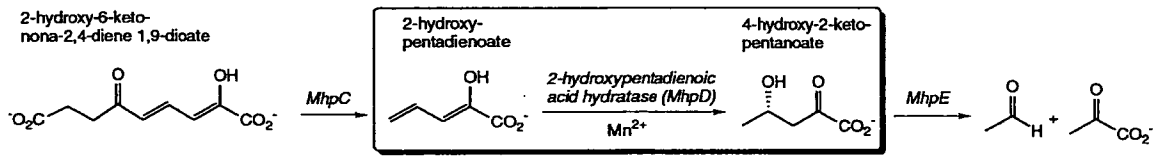

70
60
50
40
30
20
10
0
-10
Abs 280 nm
5
4
3
2
1
0
Abs 406 nm
0
3
6
9
12
15
volume (ml)

Fig. 9


2-hydroxy-6-keto-
nona-2,4-diene 1,9-dioate
MhpC
2-hydroxy-
pentadienoate
2-hydroxypentadienoic
acid hydratase (MhpD)
Mn2+
4-hydroxy-2-keto-
pentanoate
MhpE

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2009/002305

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/00 C07K14/195 C07K14/21 C07K14/245
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, CHEM ABS Data, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2008/105902 A2 (MONTANA STATE UNIVERSITY) 4 September 2008 (2008-09-04) the whole document	1
A	WO 03/096990 A2 (MONTANA STATE UNIVERSITY) 27 November 2003 (2003-11-27) cited in the application the whole document	1-28
A	D E TIMM ET AL.: "Crystal structure and mechanism of a carbon-carbon bond hydrolase" STRUCTURE, vol. 7, no. 9, September 1999 (1999-09), pages 1023-1033, XP002569123 CURRENT BIOLOGY LTD., PHILADELPHIA, PA. ISSN: 0969-2126 cited in the application the whole document	1-28

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 February 2010	09/03/2010

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Masturzo, Pietro

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2009/002305

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008105902 A2	04-09-2008	NONE	
WO 2003096990 A2	27-11-2003	AU 2003239531 A1	02-12-2003