electronic reprint

STRUCTURAL BIOLOGY
COMMUNICATIONS

ISSN: 2053-230X
journals.iucr.org/f

T4 lysozyme-facilitated crystallization of the human molybdenum cofactor-dependent enzyme mARC

Christian Kubitza, Carsten Ginsel, Florian Bittner, Antje Havemeyer, Bernd Clement and Axel J. Scheidig

Acta Cryst. (2018). F**74**, 337–344

IUCr Journals
CRYSTALLOGRAPHY JOURNALS ONLINE

Copyright © International Union of Crystallography

Author(s) of this article may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see http://journals.iucr.org/services/authorrights.html

STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Received 15 February 2018
Accepted 4 May 2018

Edited by N. Sträter, University of Leipzig, Germany

Keywords: T4 lysozyme; fusion protein; crystallization strategy; carrier-driven crystallization.

T4 lysozyme-facilitated crystallization of the human molybdenum cofactor-dependent enzyme mARC

Christian Kubitza,[a] Carsten Ginsel,[b] Florian Bittner,[c] Antje Havemeyer,[b] Bernd Clement[b] and Axel J. Scheidig[a]*

[a]Structural Biology, Zoological Institute, Kiel University, Am Botanischen Garten 1–9, 24118 Kiel, Germany, [b]Pharmaceutical Institute, Kiel University, Gutenbergstrasse 76, 24118 Kiel, Germany, and [c]Julius Kuehn Institute, Federal Research Centre for Cultivated Plants, Erwin-Baur-Strasse 27, 06484 Quedlinburg, Germany. *Correspondence e-mail: axel.scheidig@strubio.uni-kiel.de

The human mitochondrial amidoxime reducing component (hmARC) is a molybdenum cofactor-dependent enzyme that is involved in the reduction of a diverse range of N-hydroxylated compounds of either physiological or xenobiotic origin. In this study, the use of a fusion-protein approach with T4 lysozyme (T4L) to determine the structure of this hitherto noncrystallizable enzyme by X-ray crystallography is described. A set of four different hmARC-T4L fusion proteins were designed. Two of them contained either an N-terminal or a C-terminal T4L moiety fused to hmARC, while the other two contained T4L as an internal fusion partner tethered to the hmARC enzyme between two predicted secondary-structure elements. One of these internal fusion constructs could be expressed and crystallized successfully. The hmARC-T4L crystals diffracted to 1.7 Å resolution using synchrotron radiation and belonged to space group $P2_12_12_1$ with one molecule in the asymmetric unit. Initial attempts to solve the structure by molecular replacement using T4L did not result in electron-density distributions that were sufficient for model building and interpretation of the hmARC moiety. However, this study emphasizes the utility of the T4L fusion-protein approach, which can be used for the crystallization and structure determination of membrane-bound proteins as well as soluble proteins.

1. Introduction

Difficulties in obtaining diffraction-quality crystals of a target protein have always been a bottleneck in structure determination *via* X-ray crystallography and will probably remain a limiting factor in structural biology. In recent years, different tools and strategies have been developed to mediate the crystallization of challenging proteins which could not be crystallized by any conservative approach. These tools include surface-entropy reduction (Cooper *et al.*, 2007; Goldschmidt *et al.*, 2007), lysine methylation (Walter *et al.*, 2006), *in situ* proteolysis (Dong *et al.*, 2007) and metal-mediated crystallization (Laganowsky *et al.*, 2011). All of these manipulations slightly alter the surface characteristics of the target protein and aim to promote the formation of crystal contacts. Another, less subtle, strategy is chaperone-assisted crystallization. This approach relies on specific antibodies (or derivatives thereof) or other high-affinity binding-partner molecules to form stable, more rigid and, ultimately, crystallizable complexes with the target protein (Bukowska & Grütter, 2013).

A related strategy to the use of antibodies is based on the use of fusion proteins to assist crystallization. In principle, this approach has many advantages. The proteins are easily crystallizable on their own and provide additional surface area favourable for crystal lattice formation (Bell *et al.*, 2013), and

© 2018 International Union of Crystallography

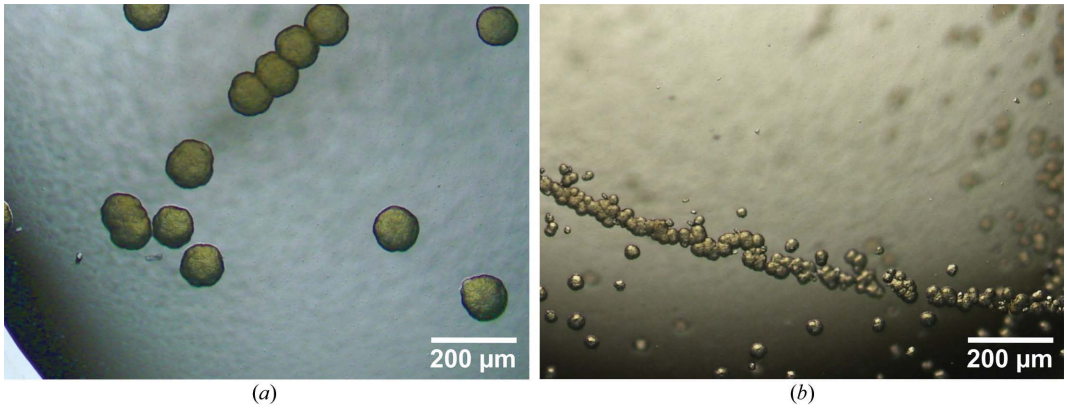
electronic reprint

some fusion partners can also be used as affinity tags for purification. The use of fusion tags has also been shown to enhance the solubility of the target protein and might therefore favour higher expression yields (Stevens, 2000). Ultimately, the three-dimensional structures of the fusion partners can be used as starting models to solve the phase problem by molecular replacement (Niemann *et al.*, 2001). This is of especially high value if no structural homologues of the target protein have been determined. However, there are also some drawbacks to be considered in using this strategy. First of all, the fusion partner might interfere with the three-dimensional arrangement of the target protein. This might result in non-natural altered conformations or even loss of function. Furthermore, the linker sequence between the fusion partners has to be chosen carefully. It needs to provide a certain degree of rigidity to avoid conformational heterogeneity, which would otherwise have negative effects on crystallization (Kobe *et al.*, 2015).

Several protein structures have successfully been determined using a fusion-protein-assisted crystallization approach. The fusion proteins used for this strategy include glutathione *S*-transferase (GST; Kuge *et al.*, 1997), thioredoxin (TRX; Corsini *et al.*, 2008), green fluorescent protein (GFP; Suzuki *et al.*, 2010), barnase (Niemann *et al.*, 2006), an engineered sterile-α motif (SAM; Nauli *et al.*, 2007), maltose-binding protein (MBP; Kobe *et al.*, 1999) and T4 lysozyme (T4L; Rosenbaum *et al.*, 2007). Although the fusion-protein strategy does not seem to be applicable to all target proteins of choice, the number of crystal structures solved by this approach is increasing owing to the fact that intensive research is being conducted in order to improve the method. One of the most thoroughly investigated systems is the MBP-mediated crystallization approach, which is recommended if the target protein starts with an N-terminal α-helix (Jin *et al.*, 2017).

The use of T4L as a crystallization-facilitating fusion partner became popular with the structure determination of the human β_2-adrenergic G-protein-coupled receptor (GPCR; Cherezov *et al.*, 2007), and has since been successfully applied to a number of different GPCRs (see, for example, Doré *et al.*, 2014; Chien *et al.*, 2010; Haga *et al.*, 2012; Miller-Gallacher *et al.*, 2014; Srivastava *et al.*, 2014; Wu *et al.*, 2010). To date, more than 90 crystal structures of T4L fusion proteins have been deposited in the PDB. Although the majority of the deposited protein structures belong to the membrane-protein family of GPCRs, there are a few individual cases of soluble proteins which could also be structurally characterized using this approach (see, for example, Bhabha *et al.*, 2014; Scott *et al.*, 2017; Baumlova *et al.*, 2014). T4L was found to be an optimal fusion partner, since it is a well folded soluble protein which can be crystallized under many conditions. Even more importantly, both termini of this protein are in close proximity to each other and therefore allow internal fusion into a target protein, tethering the T4L at two ends within a loop. This greatly reduces the risk of conformational heterogeneity, since the fusion partners can be quite rigidly linked to each other. However, Zou and coworkers were able to show that even N-terminal T4L fusions can facilitate crystallization, at least in the case of the aforementioned β_2-adrenergic GPCR (Zou *et al.*, 2012). Thorsen and coworkers recently described two ways to maximize T4L rigidity and to improve the utility of the T4L fusion approach (Thorsen *et al.*, 2014). They introduced two disulfide bridges in the interface between the lobes or, in another approach, created a 'minimal T4L' which no longer contains the N-terminal lobe. Several other reported T4L crystal structures contain molecules between two lobes of this protein which are derived from the crystallization buffer and decrease the overall flexibility.

Here, we report the crystallization of a soluble protein which contains unmodified T4L as an internal fusion partner tethered between two strands of a predicted antiparallel β-sheet. The aim was the structure determination of the human mitochondrial amidoxime-reducing component (hmARC). This enzyme was discovered in 2006 as a fourth human molybdenum cofactor (Moco)-containing enzyme in addition to sulfite oxidase, aldehyde oxidase and xanthine oxidoreductase (Havemeyer *et al.*, 2006). In the presence of NADH, mARC proteins exert N-reductive activity towards N-hydroxylated substrates together with the two electron-transport proteins cytochrome b_5 and NADH cytochrome b_5 reductase. This enzyme system is located at the outer



(*a*) (*b*)

Figure 1
Spherulite formation in hmARC1 crystallization setups. (*a*) Spherulites obtained in an initial crystallization trial. (*b*) Spherulites obtained by streak-seeding.

electronic reprint

mitochondrial membrane and is expressed in every tissue studied to date (Ott *et al.*, 2015). mARC enzymes belong to a subfamily of Moco sulfurase C-terminal domain (MOSC)-like proteins which has not yet been structurally characterized. Here, we present our strategy to generate T4L-hmARC fusion proteins for crystallization purposes.

2. Materials and methods

2.1. Preliminary design of soluble hmARC constructs and crystallization trials

Human mARC proteins are physiologically expressed with a predicted N-terminal transmembrane helix, potentially anchoring these proteins to the outer mitochondrial membrane. There are two isoforms, hmARC1 and hmARC2, which have a sequence identity of 65.7% (Wahl *et al.*, 2010). A soluble, N-terminally truncated variant of hmARC1 was designed and expressed for *in vitro* studies and crystallization purposes. In addition to the N-terminal truncation of the first 52 residues, this construct was equipped with a C-terminal His_6 tag, which was used for affinity chromatography. Several crystallization approaches and the setting up of various commercial crystallization screens with different protein concentrations did not result in the formation of protein crystals. At best, spherulites could be obtained (Fig. 1*a*), but neither additive screening nor seeding techniques (Fig. 1*b*) were sufficient to improve these initial conditions to yield diffraction-quality protein crystals.

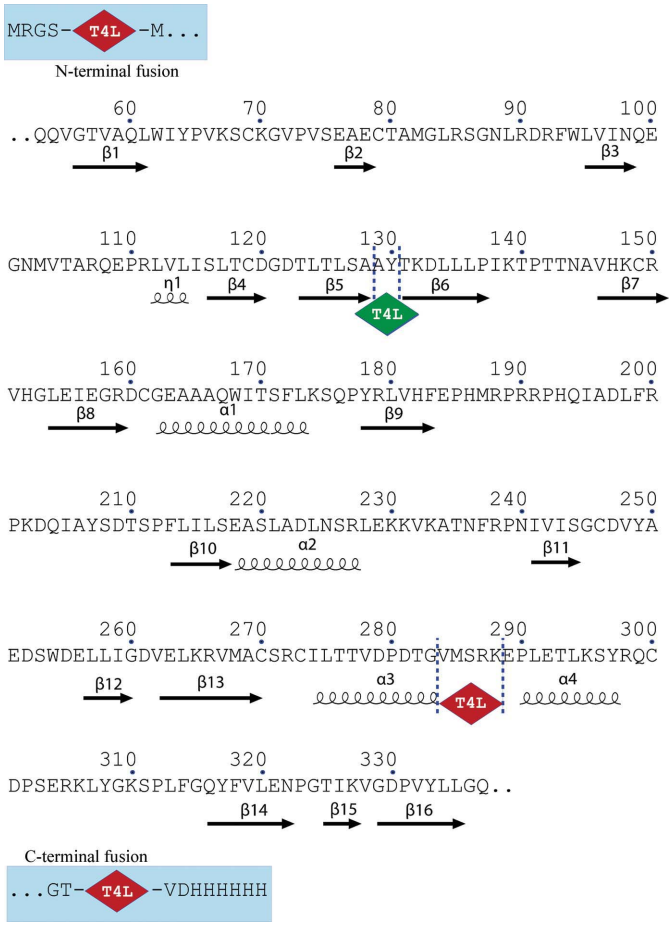


Figure 2
Secondary-structure prediction for hmARC1. T4L insertion sites for the generation of the fusion proteins are indicated by red or green rhombi. Red, constructs resulting in non-expressible or inactive fusion proteins; green, construct resulting in a functional and crystallizable fusion protein.

2.2. Modelling of hmARC1 and *in silico* design of T4L fusion constructs

Based on recent structural publications for challenging proteins, such as GPCRs, we aimed at the rational design of fusion proteins comprised of hmARC1 (33.5 kDa) and phage T4 lysozyme (T4L; 18.3 kDa) in order to facilitate crystallization. Since the N- and C-terminal residues of T4L are in close proximity, this protein is well suited for integration between two secondary-structure elements of the target protein. This enhances the chance of a more rigid linkage between the fusion partners, which might be essential for crystallization. Therefore, secondary-structure prediction and homology modelling (Fig. 2) were carried out for hmARC1 using *MODELLER* within the *HHpred* server (Söding, 2005; Söding *et al.*, 2005) as well as the *Phyre*[2] protein-fold recognition server (Kelley *et al.*, 2015). There are no close structural homologues to the hmARC enzymes, but other family members of the MOSC family, such as YuaD from *Bacillus subtilis* (PDB entry 1oru; Midwest Center for Structural Genomics, unpublished work) and a MOSC_N domain-containing protein of unknown function (PDB entry 2exn; Rossi *et al.*, 2005), share sufficient sequence identity to perform initial secondary- and tertiary-structure predictions. All generated models were analyzed for common features as well as structural differences. The most promising elements were merged into a final model, which was used to determine suitable insertion sites for the T4L fusion partner. It should be noted that this model did not represent the whole hmARC enzyme, but contained a significant portion of unmodelled gaps owing to the lack of homologues. However, some secondary-structure elements were predicted with high confidence. We designed four different hmARC1-T4L fusion proteins, with two of them having T4L fused to the hmARC1 sequence either N- or C-terminally (named N-T4L-hmARC1 and hmARC1-T4L-C, respectively). Additionally, we decided to integrate T4L into a potential three-stranded antiparallel β-sheet, replacing a predicted two-residue β-turn (residues Ala129 and Tyr130) with lysozyme (hmARC1-T4L-β). The fourth construct (hmARC1-T4L-α) had T4L internally fused between two predicted α-helices, replacing residues Val284–Lys288.

2.3. Cloning

The N-terminally truncated hmARC1 construct with a C-terminal His_6 tag (hmARC1_N-del) was present in a modified pQE80L plasmid (without its original 5′ His_6 tag-coding sequence), where it was cloned between BamHI and KpnI restriction sites. To generate the terminal T4L fusion

electronic reprint

Table 1
Oligonucleotide primers used for cloning.

Bold, restriction sites; underlined, complementary to T4L-coding sequence; italic, complementary to the hmARC1-coding sequence at the target site.

N-T4L-hmARC1_fwd	TAGCT**GGATCC**ATGAATATATTTGAAATG
N-T4L-hmARC1_rev	AGCTA**GGATCC**ATACGCGTCCCAAGTGC
hmARC1-T4L-C_fwd	TAGCT**GGTACC**ATGAATATATTTGAAATG
hmARC1-T4L-C_rev	AGCTA**GTCGAC**ATACGCGTCCCAAGTGC
hmARC1-T4L-α_fwd	*GCATTTTAACCACAGTGGACCCAGACACCGGT*ATGAATATATTTGAAATG
hmARC1-T4L-α_rev	*GGCGATAACTCTTCAGTGTTTCCAGCGGTTC*ATACGCGTCCCAAGTGC
hmARC1-T4L-β_fwd	*GCGATGGTGACACCTTGACTCTCAGTGC*AATGAATATATTTGAAATG
hmARC1-T4L-β_rev	*GCGTTTTGATAGGCAGTAGTAGGTCCTTTGT*ATACGCGTCCCAAGTGC

constructs, extension polymerase chain reactions (PCRs) were performed with the T4L-coding sequence (as provided by Addgene plasmid 18110) as a template to introduce BamHI restriction sites (for N-terminal fusion) or KpnI and SalI restriction sites (for C-terminal fusion). Subsequently, the PCR products were digested with the respective restriction enzymes and subcloned into the hmARC1-containing pQE80L vector to generate the N-T4L-hmARC1 and hmARC1-T4L-C constructs. The hmARC1-T4L-α and hmARC1-T4L-β fusion proteins were generated by overlap extension PCR within a restriction-free cloning approach (Bond & Naus, 2012). This allows the insertion of any sequence into any position within a template plasmid, yet this technique is independent of restriction-enzyme recognition sites. Therefore, a pair of hybrid primers were designed which contained complementary sequences to both the T4L insert and the hmARC1-coding sequence at the insertion site on the target plasmid. These primers were used to amplify the T4L-coding sequence from a source vector under high-fidelity conditions in a primary PCR (see Table 1). The resulting product was purified and used as a megaprimer in a secondary PCR with the hmARC1-coding target plasmid as a template. Following the secondary PCR, any template plasmid was degraded using the DpnI restriction enzyme and the purified product was ligated overnight. All fusion constructs were subsequently transformed into competent *Escherichia coli* XL1 Blue and TP1000 cells (Palmer *et al.*, 1996).

2.4. Protein expression and purification

The soluble, C-terminally His_6-tagged hmARC1-T4L fusion proteins were heterologously expressed in *E. coli* TP1000 cells (see Table 2). This protein-expression strain is unable to synthesize the dinucleotide molybdenum cofactor, but rather enriches the mononucleotide Moco as found within mARC proteins (Palmer *et al.*, 1996). A volume of 13 ml from an overnight culture was used to inoculate 2 l LB medium supplemented with 5 m*M* ammonium chloride, 1 m*M* sodium molybdate and 130 µg ml^{-1} ampicillin, resulting in an OD_{600} of approximately 0.02. The bacterial culture was incubated at 310 K and shaken at 90 rev min^{-1} until the OD_{600} reached 0.1. Protein expression was induced with freshly prepared 15 µ*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) solution and the temperature was decreased to 295 K. After 20 h, the cells were harvested at 10 000*g*. Bacterial pellets were resuspended in running buffer (50 m*M* sodium dihydrogen phosphate, 300 m*M* sodium chloride, 25 m*M* sodium molybdate, 10 m*M* imidazole pH 8.0), lysed using an EmulsiFlex-C3 (Avestin, Mannheim, Germany) and subsequently centrifuged at 75 600*g* for 1 h. His_6-tagged recombinant proteins were purified from the resulting crude extract *via* immobilized metal ion affinity chromatography (IMAC) using a HisTrap HP 5 ml column and an ÄKTApurifier FPLC system (both from GE Healthcare, Freiburg, Germany). After loading the crude extract, nonspecifically bound proteins were washed from the nickel column with 20 m*M* imidazole in running buffer. The target protein was eluted from the column by applying a one-step elution gradient to 125 m*M* imidazole in running buffer. Elution fractions were pooled, concentrated and subjected to a 5 ml HiTrap Desalting column (GE Healthcare, Freiburg, Germany) in order to exchange the buffer for a low-salt buffer (50 m*M* Tris–HCl, 5 m*M* sodium chloride pH 7.0) for subsequent ion-exchange chromatography (IEC). The protein was loaded onto a 1 ml HiTrap SP XL column (GE Healthcare, Freiburg, Gremany). Impurities that were still present after IMAC purification were usually washed through the column, while the target protein was bound with high affinity. Fractions of pure fusion protein were eluted *via* a linear gradient over 20 column volumes to a final concentration of 500 m*M* sodium chloride. Elution fractions containing the fusion protein were pooled, concentrated to 14.8 mg ml^{-1} and supplemented with 7%(*v*/*v*) glycerol prior to storage at 193 K.

Table 2
Protein-production information for hmARC1-T4L-β.

UniProt accession No.	Q5VT66 (hmARC1), D9IEF7 (T4 lysozyme)
Source organism	Human (hmARC1), *Enterobacteria phage T4* (T4 lysozyme)
Expression vector	pQE80
Expression host	*E. coli* TP1000
Amino-acid sequence of the construct for crystallization†	*MRGSM*QQVGTVAQLWIYPVKSCKGVPVSEA ECTAMGLRSGNLRDRFWLVINQEGNMVT ARQEPRLVLISLTCDGDTLTLSAMNIFE MLRIDEGLRLKIYKDTEGYYTIGIGHLL TKSPSLNAAKSELDKAIGRNCNGVITKD EAEKLFNQDVDAAVRGILRNAKLKPVYD SLDAVRRCALINMVFQMGETGVAGFTNS LRMLQQKRWDEAAVNLAKSRWYNQTPNR AKRVITTFRTGTWDAYTKDLLLPIKTPT TNAVHKCRVHGLEIEGRDCGEAAAQWIT SFLKSQPYRLVHFEPHMRPRRPHQIADL FRPKDQIAYSDTSPFLILSEASLADLNS RLEKKVKATNFRPNIVISGCDVYAEDSW DELLIGDVELKRVMACSRCILTTVDPDT GVMSRKEPLETLKSYRQCDPSERKLYGK SPLFGQYFVLENPGTIKVGDPVYLLGQ*G TVDHHHHHH*
Molecular weight (kDa)	51.8 (hmARC1-T4L-β), 33.5 (hmARC1 portion), 18.3 (T4L portion)

2.5. HPLC-based activity assay

In order to verify the enzymatic activity of the hmARC1-T4L fusion proteins, an HPLC assay was performed. In concert with the flavin-containing NADH cytochrome b_5

electronic reprint

Table 3
Crystallization.

Method	Sitting drop
Plate type	96-well
Temperature (K)	291
Protein concentration (mg ml^{-1})	14.8
Buffer composition of protein solution	50 m*M* Tris–HCl pH 7.0, 250 m*M* NaCl, 7% glycerol
Composition of reservoir solution	100 m*M* bis-tris propane–HCl pH 6.5, 200 m*M* Na_2MoO_4, 27.5% PEG 3350
Volume and ratio of drop	200 nl protein solution, 1:1 ratio
Volume of reservoir (µl)	50

reductase and the haem-containing cytochrome b_5, hmARC is able to reduce N-hydroxylated structures such as the model substrate benzamidoxime, which is converted to benzamidine (Havemeyer *et al.*, 2006). The incubation mixture consisted of 3.75 µg hmARC1 (or 5.77 µg fusion protein), 37.5 pmol cytochrome b_5 and 3.75 pmol cytochrome b_5 reductase in 20 m*M* MES buffer pH 6.0. The benzamidoxime concentration was set to 3 m*M*. After pre-incubation for 3 min at 310 K in a shaking water bath, the reaction was started under aerobic conditions by the addition of 1 m*M* NADH, yielding a total volume of 150 µl. Incubation was carried out for 15 min at 310 K and was stopped with 150 µl cold methanol, followed by 5 min shaking and 5 min centrifugation at 9500*g*. An isocratic HPLC method was used for the quantification of benzamidine. The samples were separated on a LiChrospher 60 RP-Select B (5 µm) 250-4 column with a RP-Select B 4 × 4 mm guard column at 295 K. The mobile phase consisted of 10 m*M* sodium 1-octanesulfonate, 20% acetonitrile. The flow rate was set to 1 ml min^{-1} and the detection wavelength to 229 nm. The retention times were 7.6 ± 0.2 min for benzamidoxime and 16.4 ± 0.3 min for benzamidine.

2.6. Crystallization of hmARC1-T4L fusion proteins

Initial crystallization experiments with the commercial screens JCSG-*plus* and PACT-*premier* (Molecular Dimensions, Suffolk, England) were carried out using the sitting-drop vapour-diffusion method at 291 K. 200 nl protein solution (14.8 mg ml^{-1} in 50 m*M* Tris–HCl, 250 m*M* NaCl, 7% glycerol pH 7.0) was mixed with an equal amount of precipitant solution using a Mosquito HTS pipetting robot and 96-well sitting-drop plates (both from TTP Labtech, Melbourn, England). Crystals of the hmARC1-T4L-β construct were obtained with a precipitant condition consisting of 100 m*M* bis-tris propane–HCl pH 6.5, 20%(*w*/*v*) PEG 3350 and 200 m*M* of various organic or inorganic salts, such as sodium malonate, lithium sulfate, sodium sulfate or ammonium sulfate. To improve the crystal size and quality, 96-well grid screens were set up in order to fine-tune the initial identified crystallization conditions. Different salts (at a concentration of 200 m*M*), including sodium molybdate, were screened using MES–NaOH or bis-tris propane–HCl buffers at different pH values (5.0–8.0) as well as different concentrations of PEG 3350 or PEG 4000 [15–35%(*w*/*v*)]. Crystals were taken out of the crystallization drop using nylon loops and were flash-cooled by transfer into liquid nitrogen without adding additional cryoprotectant. The most important information on crystallization is summarized in Table 3.

Table 4
Data collection and processing.

Values in parentheses are for the highest resolution shell.

Diffraction source	P14, PETRA III, Hamburg
Wavelength (Å)	0.9789
Temperature (K)	100
Detector	PILATUS 6M
Crystal-to-detector distance (mm)	188.54
Rotation range per image (°)	0.1
Exposure time per image (s)	0.00146
Space group	$P2_12_12_1$
a, *b*, *c* (Å)	61.1, 74.8, 110.7
α, β, γ (°)	90.0, 90.0, 90.0
Resolution range† (Å)	42.52–1.65 (1.68–1.65)
Total No. of reflections†	723415 (18654)
No. of unique reflections†	60665 (2860)
Completeness† (%)	98.4 (95.9)
Multiplicity†	11.9 (6.5)
$\langle I/\sigma(I)\rangle$†	9.0 (0.4)
Mean $\langle I\rangle$ half-set correlation $CC_{1/2}$†	0.99 (0.15)
R_{merge}†	0.21 (3.85)
R_{meas}†	0.22 (4.17)
$R_{p.i.m.}$†	0.06 (1.59)

† Values as provided by *AIMLESS* after data processing, merging and scaling.

2.7. Data collection, processing and phasing

X-ray diffraction data were collected at 100 K on the EMBL MX beamline P14 at PETRA III, DESY, Hamburg, Germany equipped with a PILATUS 6M detector. High-resolution data sets with sufficient quality could only be obtained by combining a double-focused beam (5 × 10 µm) at 100% transmission with helical data collection along the longitudinal axis of the crystals in order to minimize radiation damage. Diffraction data were indexed and integrated using *XDS* (Kabsch, 2010). Space-group determination, data scaling and merging were performed using *AIMLESS* (Evans & Murshudov, 2013) as part of the *CCP*4 suite (Winn *et al.*, 2011), while applying the free *R* flag to 5% of the reflections. Data-collection and processing statistics are summarized in Table 4. Since there were neither sufficient anomalous scattering data derived from potentially bound molybdenum nor any structural homologue of hmARC, phasing approaches were limited to molecular replacement (MR) using T4 lysozyme as a search template within *MOLREP* (Vagin & Teplyakov, 2010).

3. Results and discussion

N-terminally truncated hmARC1 was cloned into a modified pQE80 vector providing a C-terminal His_6 tag. This hmARC1_N-del construct was used as a template to generate fusion proteins comprised of hmARC and T4 lysozyme. In order to reduce conformational heterogeneity of the fusion proteins, T4L was inserted into short loop regions between two predicted α-helices (hmARC1-T4L-α) or between two β-strands of a predicted three-stranded β-sheet (hmARC1-T4L-β). The internal insertion sites were chosen based on secondary- and tertiary-structure predictions performed by

electronic reprint

homology-modelling servers. While proteins with the T4L moiety as an N- or C-terminal fusion partner could not be overexpressed in *E. coli*, fusion proteins containing T4L as an internal fusion partner within the hmARC1 sequence were expressed as soluble proteins with sufficient yields (1–2 mg l^{-1}). Purification *via* affinity and ion-exchange chromatography resulted in protein fractions of high purity.

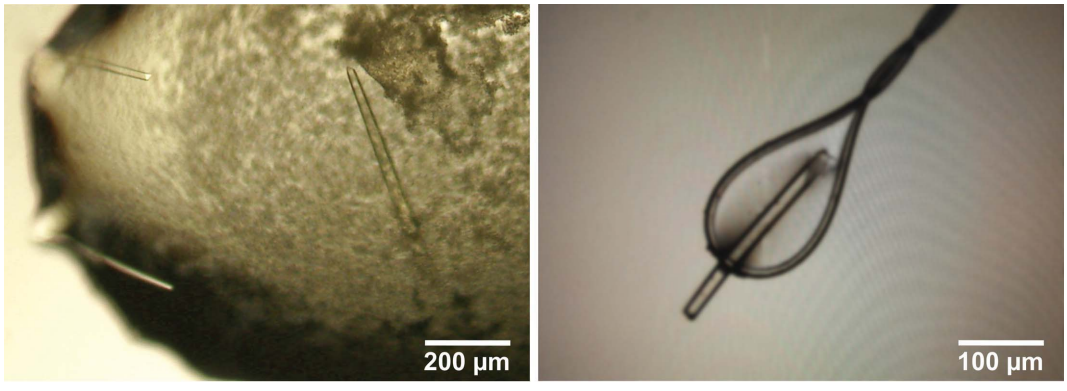
The enzyme activity of the fusion proteins was determined by benzamidoxime turnover within an HPLC-based activity assay. The hmARC1-T4L-α construct did not exert any detectable activity at all. It is most likely that the T4L moiety induced conformational changes within the tertiary structure of hmARC, rendering the enzyme inactive. Likewise, the internal fusion partner might either interfere with the active site of the enzyme or inhibit electron transfer between hmARC and its redox partners. Either way, this construct was discarded from further studies and crystallization trials because it did not reflect the hmARC enzyme under near-physiological conditions. However, the enzyme activity of the hmARC1-T4L-β construct (536 $\pm$ 20 nmol benzamidine per minute per milligram of protein) was comparable to that of soluble hmARC1 (777 $\pm$ 8 nmol benzamidine per minute per milligram of protein). Despite the bulky fusion partner, this enzyme still possesses ∼70% of the specific activity of the N-terminally truncated wild-type enzyme and was therefore regarded as a promising candidate for crystallization and structure determination.

Initial screens identified conditions containing 100 m*M* bis-tris propane adjusted to various pH values, 20%(*w*/*v*) PEG 3350 and various salts at 200 m*M* to be suitable for the crystallization of hmARC1-T4L-β. These initial conditions were refined in grid screens in order to improve crystal size and quality. The optimized crystallization conditions (100 m*M* bis-tris propane pH 6.5, 200 m*M* sodium molybdate, 27.5% PEG 3350) yielded rod-shaped protein crystals with a length of 200–400 µm but with a diameter of only 10–15 µm (Fig. 3). Of note, only precipitating solutions containing bis-tris propane buffer produced protein crystals. Apart from its buffering capabilities, this molecule might serve as a bidentate additive that is necessary to stabilize the two flexible lobes of the T4L moiety. Similar conditions containing MES buffer at the same pH did not result in crystal formation.

Crystals of the hmARC-T4L-β construct diffracted to 2.5–1.7 Å resolution using double-focused synchrotron radiation. These crystals were very susceptible to radiation damage, which hindered the collection of complete data sets using traditional data-collection strategies. However, given the length of the crystals and their homogenous diffraction quality, they could be subjected to a continuous helical data-collection strategy ('4D-scan'), which was achieved on beamline P14 at EMBL Hamburg. Complete data sets of very good quality were obtained.

Owing to the lack of structural homologues of the hmARC1 enzyme, phasing was limited to MR using the T4L moiety as a search template. Indeed, the position of the T4L could be found using *MOLREP* as implemented within the *CCP*4 suite, and convincing electron-density maps could be obtained for this part of the fusion protein using rigid-body as well as restraint refinement steps in *REFMAC*5 (Murshudov *et al.*, 2011; Fig. 4). However, the derived phases and calculated electron-density distributions were not sufficient to build and refine the hmARC1 enzyme. Using the hmARC1 model generated by *MODELLER* or *Phyre*[2] as an MR search template did not result in convincing solutions. Therefore, additional data will be needed in order to overcome the phase problem and solve the crystal structure of hmARC1. Experimental phasing methods, such as single-wavelength or multi-wavelength anomalous dispersion, will have to be applied to selenomethionine- or heavy-atom-derivatized protein crystals and may be combined with the phases already obtained by MR using T4L as a search template. Alternatively, a better hmARC1 homology model (or at least of parts of the enzyme) will have to be generated and used as an MR template in addition to the already fixed T4L position. This way, phases could be obtained and improved in an iterative approach in order to generate electron-density distributions that are sufficient for model building and interpretation of the crystal structure of hmARC1.

This study emphasizes the utility of the T4L fusion-protein approach, which is not only useful for the crystallization and



(*a*) (*b*)

Figure 3
Crystals of hmARC1-T4L-β. (*a*) Crystals grown in 100 m*M* bis-tris propane–HCl pH 6.5, 20%(*w*/*v*) PEG 3350, 200 m*M* sodium molybdate. (*b*) A harvested crystal in a nylon loop.

electronic reprint

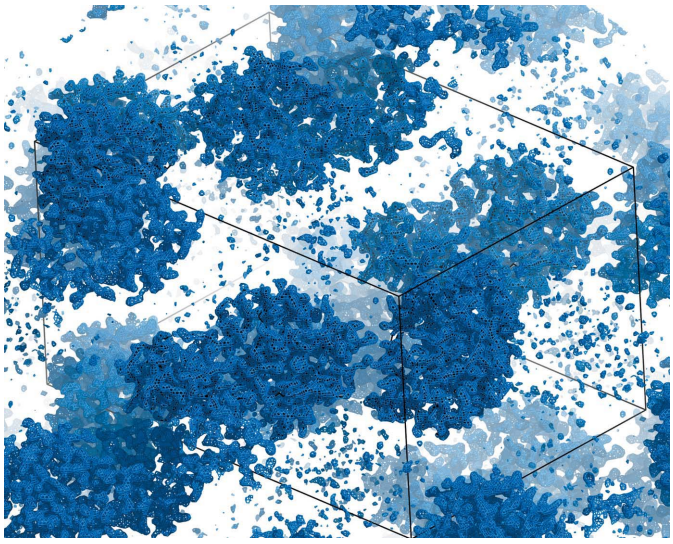


Figure 4
Electron-density map showing T4L molecules, contoured at 2.0σ. Phase information was obtained by molecular replacement using T4L as the template (PDB entry 206l; Blaber *et al.*, 1993). Owing to insufficient phase information, hmARC1 molecules are not yet distinguishable within the unit cell.

structure determination of GPCR membrane proteins but can also be applied to soluble proteins. The ability to use T4L as an internal fusion partner by replacing a small loop of the target protein makes this strategy applicable to a wide range of different proteins. Here, we show that it might be highly advantageous to choose an insertion site within a predicted β-sheet, since this greatly reduces the risk of conformational heterogeneity of the fusion protein. Therefore, we propose that the T4L fusion-protein approach should be investigated with a number of different proteins for which traditional strategies have failed to yield good crystals. We were able to show that even unmodified full-length T4L is suitable as a fusion partner in order to obtain high-quality crystals with good diffraction properties. Furthermore, we propose that the utility of T4L fusion proteins in structural biology has not yet been fully exploited and might contribute to significant advancements in fusion-driven protein crystallization.

Acknowledgements

The plasmid containing the coding sequence for wild-type T4 lysozyme was obtained from Addgene (plasmid 18110) as deposited by Brian Matthews (Eugene, USA). We gratefully acknowledge access to the core facilities of the BiMo/LMB of Kiel University. We thank Felix Helfrich for helpful discussions and support during data collection. Diffraction data were collected on beamline P14 operated by EMBL at the PETRA III storage ring. We are grateful to the beamline staff for providing assistance in using the beamlines.

Funding information

The research leading to these results received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under BioStructX (grant agreement No. 283570). Additionally, we are grateful for access to the HTX crystallization facility by means of a grant from P-Cube and BioStructX at the EMBL Outstation Hamburg. Beamtime at P14 at the EMBL Outstation Hamburg was also funded by a BioStructX grant.

References

Baumlova, A., Chalupska, D., Różycki, B., Jovic, M., Wisniewski, E., Klima, M., Dubankova, A., Kloer, D. P., Nencka, R., Balla, T. & Boura, E. (2014). *EMBO Rep.* **15**, 1085–1092.
Bell, M. R., Engleka, M. J., Malik, A. & Strickler, J. E. (2013). *Protein Sci.* **22**, 1466–1477.
Bhabha, G., Cheng, H.-C., Zhang, N., Moeller, A., Liao, M., Speir, J. A., Cheng, Y. & Vale, R. D. (2014). *Cell*, **159**, 857–868.
Blaber, M., Lindstrom, J. D., Gassner, N., Xu, J., Heinz, D. W. & Matthews, B. W. (1993). *Biochemistry*, **32**, 11363–11373.
Bond, S. R. & Naus, C. C. (2012). *Nucleic Acids Res.* **40**, W209–W213.
Bukowska, M. A. & Grütter, M. G. (2013). *Curr. Opin. Struct. Biol.* **23**, 409–416.
Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Kuhn, P., Weis, W. I., Kobilka, B. K. & Stevens, R. C. (2007). *Science*, **318**, 1258–1265.
Chien, E. Y. T., Liu, W., Zhao, Q., Katritch, V., Han, G. W., Hanson, M. A., Shi, L., Newman, A. H., Javitch, J. A., Cherezov, V. & Stevens, R. C. (2010). *Science*, **330**, 1091–1095.
Cooper, D. R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M. & Derewenda, Z. (2007). *Acta Cryst.* D**63**, 636–645.
Corsini, L., Hothorn, M., Scheffzek, K., Sattler, M. & Stier, G. (2008). *Protein Sci.* **17**, 2070–2079.
Dong, A. *et al.* (2007). *Nature Methods*, **4**, 1019–1021.
Doré, A. S., Okrasa, K., Patel, J. C., Serrano-Vega, M., Bennett, K., Cooke, R. M., Errey, J. C., Jazayeri, A., Khan, S., Tehan, B., Weir, M., Wiggin, G. R. & Marshall, F. H. (2014). *Nature (London)*, **511**, 557–562.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Goldschmidt, L., Cooper, D. R., Derewenda, Z. S. & Eisenberg, D. (2007). *Protein Sci.* **16**, 1569–1576.
Haga, K., Kruse, A. C., Asada, H., Yurugi-Kobayashi, T., Shiroishi, M., Zhang, C., Weis, W. I., Okada, T., Kobilka, B. K., Haga, T. & Kobayashi, T. (2012). *Nature (London)*, **482**, 547–551.
Havemeyer, A., Bittner, F., Wollers, S., Mendel, R., Kunze, T. & Clement, B. (2006). *J. Biol. Chem.* **281**, 34796–34802.
Jin, T., Chuenchor, W., Jiang, J., Cheng, J., Li, Y., Fang, K., Huang, M., Smith, P. & Xiao, T. S. (2017). *Sci. Rep.* **7**, 40991.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. (2015). *Nature Protoc.* **10**, 845–858.
Kobe, B., Center, R. J., Kemp, B. E. & Poumbourios, P. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 4319–4324.
Kobe, B., Ve, T. & Williams, S. J. (2015). *Acta Cryst.* F**71**, 861–869.
Kuge, M., Fujii, Y., Shimizu, T., Hirose, F., Matsukage, A. & Hakoshima, T. (1997). *Protein Sci.* **6**, 1783–1786.
Laganowsky, A., Zhao, M., Soriaga, A. B., Sawaya, M. R., Cascio, D. & Yeates, T. O. (2011). *Protein Sci.* **20**, 1876–1890.
Miller-Gallacher, J. L., Nehmé, R., Warne, T., Edwards, P. C., Schertler, G. F. X., Leslie, A. G. W. & Tate, C. G. (2014). *PLoS One*, **9**, e92727.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Nauli, S., Farr, S., Lee, Y.-J., Kim, H.-Y., Faham, S. & Bowie, J. U. (2007). *Protein Sci.* **16**, 2542–2551.
Niemann, H. H., Knetsch, M. L., Scherer, A., Manstein, D. J. & Kull, F. J. (2001). *EMBO J.* **20**, 5813–5821.

electronic reprint

Niemann, H. H., Schmoldt, H. U., Wentzel, A., Kolmar, H. & Heinz, D. W. (2006). *J. Mol. Biol.* **356**, 1–8.
Ott, G., Havemeyer, A. & Clement, B. (2015). *J. Biol. Inorg. Chem.* **20**, 265–275.
Palmer, T., Santini, C. L., Iobbi-Nivol, C., Eaves, D. J., Boxer, D. H. & Giordano, G. (1996). *Mol. Microbiol.* **20**, 875–884.
Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Yao, X.-J., Weis, W. I., Stevens, R. C. & Kobilka, B. K. (2007). *Science*, **318**, 1266–1273.
Rossi, P., Ramelot, S. S. T., Xiao, R., Ho, C. K., Ma, L.-C., Acton, T. B., Kennedy, M. A. & Montelione, G. T. (2005). *J. Biomol. NMR*, **33**, 197.
Scott, D. C. *et al.* (2017). *Nature Chem. Biol.* **13**, 850–857.
Söding, J. (2005). *Bioinformatics*, **21**, 951–960.
Söding, J., Biegert, A. & Lupas, A. N. (2005). *Nucleic Acids Res.* **33**, W244–W248.
Srivastava, A., Yano, J., Hirozane, Y., Kefala, G., Gruswitz, F., Snell, G., Lane, W., Ivetac, A., Aertgeerts, K., Nguyen, J., Jennings, A. & Okada, K. (2014). *Nature (London)*, **513**, 124–127.
Stevens, R. C. (2000). *Structure*, **8**, R177–R185.
Suzuki, N., Hiraki, M., Yamada, Y., Matsugaki, N., Igarashi, N., Kato, R., Dikic, I., Drew, D., Iwata, S., Wakatsuki, S. & Kawasaki, M. (2010). *Acta Cryst.* D**66**, 1059–1066.
Thorsen, T. S., Matt, R., Weis, W. I. & Kobilka, B. K. (2014). *Structure*, **22**, 1657–1664.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wahl, B., Reichmann, D., Niks, D., Krompholz, N., Havemeyer, A., Clement, B., Messerschmidt, T., Rothkegel, M., Biester, H., Hille, R., Mendel, R. R. & Bittner, F. (2010). *J. Biol. Chem.* **285**, 37847–37859.
Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I. & Grimes, J. M. (2006). *Structure*, **14**, 1617–1622.
Winn, M. D. (2010). *Acta Cryst.* D**67**, 235–242.
Wu, B., Chien, E. Y. T., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F. C., Hamel, D. J., Kuhn, P., Handel, T. M., Cherezov, V. & Stevens, R. C. (2010). *Science*, **330**, 1066–1071.
Zou, Y., Weis, W. I. & Kobilka, B. K. (2012). *PLoS One*, **7**, e46039.

electronic reprint