OPEN ACCESS

University of BRISTOL

Bunnak, W., Winter, A. J., Lazarus, C. M., Crump, M. P., Race, P. R., & Wattana-Amorn, P. (2020). SAXS reveals highly flexible interdomain linkers of tandem acyl carrier protein–thioesterase domains from a fungal non-reducing polyketide synthase. *FEBS Letters*. https://doi.org/10.1002/1873-3468.13954

Peer reviewed version

Link to published version (if available):
10.1002/1873-3468.13954

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Wiley at https://febs.onlinelibrary.wiley.com/doi/full/10.1002/1873-3468.13954 . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/

Solution characterization of tandem acyl carrier protein – thioesterase domains from a fungal non-reducing polyketide synthase

Waraporn Bunnak [1,a], Ashley J. Winter [2,a], Colin M. Lazarus [3], Matthew P. Crump [2], Paul R. Race [4,5] and Pakorn Wattana-Amorn [1,*]

[1]Department of Chemistry, Special Research Unit for Advanced Magnetic Resonance and Center of Excellence for Innovation in Chemistry, Faculty of Science, Kasetsart University, Bangkok, 10900, Thailand

[2]School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, UK

[3]School of Biological Sciences, University of Bristol, Bristol, BS8 1TQ, UK

[4]School of Biochemistry, University of Bristol, University Walk, Bristol, BS8 1TD, U.K.

[5]BrisSynBio Synthetic Biology Research Centre, University of Bristol, Tyndall Avenue, BS8 1TQ, UK

[a]These authors contributed equally to this work.

*To whom correspondence should be addressed: Pakorn Wattana-Amorn, Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand. Telephone: +66(0)2-562-5555 ext 647544; FAX: +66(0)2-579-0658; E-mail: fscipwa@ku.ac.th

Abstract

Menisporopsin A is a fungal bioactive macrocyclic polylactone requiring only reducing (R) and non-reducing (NR) polyketide synthases (PKSs) to guide a series of esterification and cyclolactonization reactions with no structural information pertaining to these PKSs. Here we report the solution characterization of singlet and doublet acyl carrier protein (ACP_2 and ACP_1-ACP_2)-thioesterase (TE) domains from NR-PKS, involved in menisporopsin A biosynthesis. Small angle X-ray scattering (SAXS) studies in combination with homology modelling reveal that these polypeptides adopt a distinctive Beads-on-a-String configuration, characterized by the presence of highly flexible interdomain linkers. These models provide a platform for studying domain organization and interdomain interactions in fungal NR-PKSs, which may be of value in directing the design of functionally optimised polyketide scaffolds.

Keywords

macrocyclic polylactone, thioesterase, acyl carrier protein, fungal non-reducing polyketide synthase, small angle X-ray scattering

Abbreviations

PKS, polyketide synthase; R-PKS, reducing polyketide synthase; NR-PKS, non-reducing polyketide synthase; ACP, acyl carrier protein; TE, thioesterase; KR, ketoreductase; PT, product template; NRPS, non-ribosomal peptide synthetase; PPTase, 4′-phosphopantetheinyl transferase; SEC-SAXS, size exclusion chromatography coupled with small angle X-ray scattering; EOM, ensemble optimisation method

Introduction

Fungal polyketides exhibit a diverse array of chemical structures, which range in complexity from simple aromatic compounds, such as the antibiotic 6-methylsalicylic acid, to the elaborate macrocyclic polylactones. An illustrative example of the latter is menisporopsin A, which possesses a broad spectrum of biological activities including acting as an antimycobacterial, antimalarial and cytotoxic agent [1]. Polyketides are biosynthesized by multidomain megaenzymes termed polyketide synthases (PKSs), which assemble polyketide backbones from simple thioester substrates. Most fungal PKSs contain multiple catalytic domains organized in a single modular architecture, with each domain used repeatedly during multiple iterative cycles of chain extension. For menisporopsin A biosynthesis only two iterative PKSs are required: one reducing (R) (Men1), and one non-reducing (NR) (Men2) (Fig. 1) [2, 3]. These two enzymes are sufficient to perform the multiple esterification and cyclolactonization reactions to produce ascotrichalactone A, an intermediate of menisporopsin A. Ascotrichalactone A lacks reduction of C37 (Fig. 1) suggesting that an additional *trans*-acting ketoreductase (KR) is needed for conversion to menisporopsin A [3]. Esterification and cyclolactonization reactions are believed to be catalyzed by the thioesterase (TE) domain of Men2. This enzyme contains a complement of catalytic domains typical of NR-PKSs albeit with an additional ACP domain (Fig. 1). This is not an uncommon architecture and has been found in other fungal PKSs [4-6].

Although progress has been made in delineating the enzymology of fungal iterative PKSs, much still remains to be learned about the structures and functions of these sophisticated molecular machines. Due to the size and dynamic flexibility of intact PKSs, these systems represent challenging targets for high-resolution structure determination using X-ray crystallography or NMR techniques. In order to overcome those problems, structural

studies of individual domains have proven to be a successful approach, as evidenced by studies of product template (PT) [7], ACP [8] and TE [9] domains from the PksA iterative PKS involved in aflatoxin B_1 biosynthesis in *Aspergillus parasiticus*. However, such studies provide limited insight into domain organization or mechanisms of interdomain interaction. This information can however be accessed by employing appropriate biophysical methods and in particular SAXS, which provides low resolution information on the shape, conformation and assembly state of a protein of interest by X-ray scattering and can characterize large conformationally heterogeneous macromolecules in solution. Parameters extracted from SAXS data (R_g and D_{max}) can enable the generation of *ab initio* shape envelope models, providing low resolution descriptions of polypeptides under investigation.

To date, SAXS methods have been successfully applied to a number of bacterial modular PKSs including the 6-deoxyerythronolide B synthase [10] and the virginiamycin *trans*-AT PKS [11], leveraging the availability of domain organization within these modules. High resolution structures of individual ACP and TE domains of PksA have been reported, but these studies cannot afford insight into interdomain interactions and domain organization that ultimately controls product release. The only studies undertaken to date which report on the conformational dynamics of a thioesterase are those of a thiolation (T)-TE didomain based on a high resolution structure [12] and the crystal structure of an intact non-ribosomal peptide synthetase (NRPS) module [13]. Thus, there is a need to undertake further structural studies of PKSs comprised of conformationally flexible assemblies that cleverly orchestrate protein-protein interactions, especially as applied to fungal iterative systems. Here we report structural models of the ACP_2-TE and ACP_1-ACP_2-TE of Men2 derived from a combination of SAXS analyses, homology modelling and allied biophysical methods.

Materials and methods

Cloning of TE, ACP_2-TE and ACP_1-ACP_2-TE from men2

Three expression vectors containing gene fragments encoding TE and ACP domains of Men2 were constructed. The synthetic gene of Men2 TE domain optimized for *E. coli* expression was cloned into pET28b via *Nde*I/*Xho*I sites (pET28b-men2TE). For ACP_1-ACP_2-TE, the region encoding ACP_1-ACP_2 was amplified from *men2* while the TE domain was amplified from pET28b-men2TE. Primer-introduced overlap between resulting ACP_1-ACP_2 and TE amplicons enabled scar-free joining of these sequences to each other and the *E. coli*-yeast shuttle vector pEYA2eGFP by homologous recombination in yeast. The ACP_1-ACP_2-TE gene was then cloned into pET28a via *Nde*I/*Hind* III sites (pET28a-men2ACP_1ACP_2TE). This was used as a template to amplify ACP_2-TE, which was cloned into pOPINF [14] using the In-Fusion®HD cloning kit (Takara Bio, USA) (pOPINF-men2ACP_2TE). All constructs encode N-terminally hexa-his tagged forms of each of the proteins under investigation. Sequences of all primers used are listed in Table S1.

Expression and purification of Men2 TE, ACP_2-TE and ACP_1-ACP_2-TE

All constructs were introduced into *E. coli* BL21 (DE3) and grown in LB medium supplemented with 50 µg/mL of kanamycin for transformants harbouring pET28b-men2TE and pET28a-men2ACP_1ACP_2TE, and 100 µg/mL of ampicillin for transformants harbouring pOPINF-men2ACP_2TE. Cultures were grown at 37 °C with shaking to an OD_{600nm} of 0.6-0.8 then cooled to 22 °C (18 °C for Men2 TE) and induced with IPTG to a final concentration of 0.25 mM. Cells were harvested 18 hours post induction and resuspended in 50 mM Tris-HCl, 0.1% Triton X-100, 0.5 M NaCl and 10 % glycerol, pH 8. Prior to sonication, lysozyme was added to the cell suspension to a final concentration of 1 mg/mL. Following incubation on ice for 30 min, cells were lysed using a Soniprep 150 with a 10s on/off cycle until the lysate was

no longer viscous. The lysate was then clarified by centrifugation and loaded onto a HisTrap HP column. Proteins were eluted using a linear gradient of 0.02 – 1 M imidazole in 50 mM Tris-HCl, 150 mM NaCl, pH 8. Eluted fractions containing target proteins were pooled and this material subjected to further purification by using a HiLoad™ 16/600 Superdex™ 200 pg column in 50 mM Tris-HCl, 150 mM NaCl, pH 8.0. Molecular weights of all purified proteins were confirmed by electrospray ionization mass spectrometry (ESI-MS) using a Synapt G2-Si (Waters) fitted with a TriVersa NanoMate® (Advion), with the ion source set to positive mode. All spectra were acquired in the 600-3000 m/z range, with data analysed using MassLynx™ 4.1. All protein samples were desalted and concentrated using Jupiter C4 resin 300 Å (Phenomenex) prior to ESI-MS analysis.

Circular Dichroism (CD) Spectroscopy

All proteins, Men2 TE, ACP_2-TE and ACP_1-ACP_2-TE, were buffer-exchanged into 10 mM phosphate, 50 mM NaF, pH 8.0 and diluted to 0.1-0.5 mg/mL. CD spectra were collected using a JASCO J-1500 CD spectrometer. Spectra were recorded at 25 °C using a 0.1 cm path-length cuvette, between 190 - 260 nm, with a data interval of 1 nm. All CD spectra were background corrected and converted to units of mean residue ellipticity [θ] using the formula

$$[\theta] = (100{,}000 \times \theta_{obs}) / (d \times C \times (N - 1))$$

where [θ] = mean residue ellipticity in deg cm^2 $dmol^{-1}$, θ_{obs} = observed ellipticity in mdeg, d = path length in mm, C = concentration in μM and N = number of amino acids.

Calculation of secondary structure contents of all proteins were performed on DichroWeb [15-17] using CONTINLL algorithm [18, 19] and SP175 [20] as a reference data set which gave the highest predicted helix content. BestSel webserver [21, 22] was also used

for calculating secondary structure content and the results were also used to compare with those obtained from DichroWeb.

In vitro conversion of *apo* to *holo*-Men2 ACP_2-TE

For the conversion of *apo* Men2 ACP_2-TE to its *holo* form, purified *apo* Men2 ACP_2-TE was incubated with 1.35 μM 4′-phosphopantetheinyl transferase (PPTase) Sfp, 0.5 mM CoA, 2 mM DTT and 10 mM $MgCl_2$ in 50 mM Tris-HCl buffer (pH 8.0). The reaction was incubated at 37 °C for 20 min prior to ESI-MS analysis.

Homology modelling

Homology models of Men2 ACP_1, ACP_2 and TE were constructed using I-TASSER [23-25] and then assembled using DEMO [26]. The linker region was chosen based on the model obtained from I-TASSER with highest C-score and prediction of secondary structures of amino acid sequence performed on the XtalPred server [27]. The chosen linker regions between each domain were identified as a low complexity region predicted by SEG [28]. For the *holo* form of both Men2 ACP_2-TE and ACP_1-ACP_2-TE, the active site serine of ACP_2 was replaced with a serine-phosphopantetheine moiety (4HH) obtained from PDB:2N6Y in Coot [29].

SAXS data collection and analysis

In-line SEC-SAXS of Men2 ACP_2-TE and ACP_1-ACP_2-TE were collected at B21 Diamond Light Source using an Agilent 1200 HPLC and 2.4 mL Superdex S200 Increase 3.2/300 column (GE Healthcare). Fifty microliters of Men2 ACP_1-ACP_2-TE at 6.9 mg/mL and ACP_2-TE at 15 mg/ml were loaded onto the S200 column in 25 mM Tris, 150 mM NaCl, pH 8.0. Frames were collected at 3 seconds per frame at 25 °C and X-ray scattering was recorded (Pilatus 4M detector) at a fixed camera length of 4.014 m, at 12.4 keV. Angular *q*

range data were collected between 0.0025- 0.34 $Å^{-1}$. SAXS data were analysed and reported by following the 2017 publication guidelines [30]. Data reduction, buffer subtraction and analysis in both the reciprocal and real domain were performed to determine the radius of gyration (R_g) and the maximum particle dimension (D_{max}) using ScÅtter 3.1r [31]. *Ab initio* bead models for each dataset were generated by programs within the ATSAS 2.7.2 package [32] : DAMMIF [33], averaging over 23 independent runs using the program DAMAVER [34], before a single DAMMIN [35] refinement run. *Ab initio* bead models were superimposed to three dimensional structures of proteins using SUPCOMB [36]. The Ensemble Optimisation Method (EOM) [37, 38] was used to model flexible regions of both Men2 ACP_2-TE and ACP_1-ACP_2-TE (Fig. S1) using randomised linkers to generate a large pool of randomly generated models with P1 symmetry. The obtained ensemble of models with various conformations for both samples were selected and quantitatively compared with the experimental scattering profile of each protein. EOM analysis used the individual domains of ACP_1, ACP_2 and the TE domain obtained from homology modelling with the N-terminal His_6 tag and interdomain linkers treated as flexible.

Results and Discussion

ACP domain is a solubility partner of TE domain

Initial attempts to produce soluble Men2 TE by recombinant expression in *E. coli* BL21 (DE3) cells failed to yield soluble protein (Fig. S2) despite exhaustive efforts scouting a range of different expression conditions (induction temperature, induction time, etc.). However, soluble Men2 TE was obtained in *E. coli* ArcticExpress (DE3) cells in the presence of cold-adapted chaperonins [39]. Several purification steps were performed in order to remove TE associated chaperone contaminants resulting in the recovery of a small amount of

purified protein. In contrast, Men2 ACP_2-TE and ACP_1-ACP_2-TE could be readily expressed in *E. coli* BL21 (DE3) cells and purified to homogeneity, indicating that solubility of Men2 TE is significantly enhanced by fusion to at least one ACP domain (Fig. S2). SEC analysis of Men2 ACP_2-TE and ACP_1-ACP_2-TE indicated that both proteins are monomeric in solution, with ~~more aggregation and oligomeric species~~a higher degree of aggregated species (elution volume = 42.8 mL) observed for Men2 ACP_1-ACP_2-TE (Fig. S3). Monomeric fraction of Men2 ACP_1-ACP_2-TE was concentrated for SAXS data collection. Nevertheless, ~~both~~ aggregated~~ion and oligomeric~~ species were still observed on the SEC column used for SAXS data collection indicating the propensity of ~~equilibrium between monomeric, oligomeric and aggregation species of~~ Men2 ACP_1-ACP_2-TE to aggregate into higher-order species.

Purified recombinant Men2 TE, ACP_2-TE and ACP_1-ACP_2-TE were also subjected to CD analysis ~~in an effort~~ to investigate the secondary structure composition of these polypeptides. Notably, Men2 TE, ACP_2-TE and ACP_1-ACP_2-TE possess closely matching spectral profiles (Fig. S4) and are similar to that of PksA TE with ellipticity minimum around 208-209 nm [9] indicating similarity of their secondary structure contents. However, Men2 ACP_1-ACP_2-TE shows more pronounced ellipticity minimum at 208 and around 220 nm consistent with the presence of two α-helical ACP domains. However secondary structure prediction calculators indicated that the obtained CD spectra underrepresented the percentage of secondary structure of all constructs, when compared to the generated homology models (Figure S4).

Holo **conversion of Men2 ACP_2-TE**

Following purification, the molecular masses of both Men2 ACP_2-TE and ACP_1-ACP_2-TE were confirmed by ESI-MS and the predominant species found to be in agreement with predicted values for the *apo* forms of both proteins (Fig. S5). However, evidence of the

presence of *holo* Men2 ACP_2-TE was observed in the purified Men2 ACP_2-TE sample, equating to 33 % of the total material. This is presumably a consequence of the action of the endogenous *E. coli* PPTase showing partial activity with ACP_2. The *apo* form of Men2 ACP_2-TE could be completely converted to *holo* form using Sfp (Fig S5). This PPTase has been previously shown to have promiscuity towards a wide range of ACPs and peptidyl carrier proteins (PCPs) [40, 41]. Similarly, the *holo* form was also detected in purified recombinant ACP_1-ACP_2-TE but with a maximum of one ACP domain modified (Fig. S6). This is expected to be attributed to modification of the ACP_2 domain by *E. coli* PPTase. Men2 ACP_1-ACP_2-TE was also found to be in gluconoylated form for both *apo* and *holo* material, equating to the formation of four individual species in solution with 46% in *apo* and 54% in *holo* forms (Fig. S6). Gluconoylation is often observed in proteins with a Gly-Ser-Ser-$(His)_6$ sequence at the N-terminus following expression in *E. coli* [42].

Primary data analysis

During this study, multiple attempts were made to obtain a monodisperse sample of the Men2 ACP_2-TE and ACP_1-ACP_2-TE. This led us to attempt to first prepare a homogeneous sample consisting of pure *holo* Men2 ACP_2-TE. The purified recombinant *apo/holo* Men2 ACP_2-TE sample could be fully converted to the *holo* form, but the resulting protein was prone to aggregation which precluded its analysis by SAXS. In an effort to circumvent this problem, the post-translationally modified serine (Ser1821) of ACP_2 was mutated to an alanine to enable access to *apo* only ACP_2-TE. This protein was, however, insoluble after recombinant overexpression in *E. coli*. Partial modification of ACPs in PKSs is common [43] and modification to the *holo* form may or may not impact on ACP structure. ACPs often show identical structures between their *apo* and *holo* forms or subtle differences have been reported [44-48]. This often translates to a complete absence or transient interaction between phosphopantetheinyl group and the ACP localised around the modified

serine. With the negligible differences in mind, we sought to further investigate whether samples with both forms present would be amenable to shape analysis by SAXS. The SEC-SAXS was applied on heterogeneous samples of Men2 ACP_2-TE and ACP_1-ACP_2- TE containing both *apo* and *holo* forms. Primary data reduction in reciprocal (Guinier analysis) and real space (*P(r)* analysis) was performed to generate model-free parameters of the particles (R_g and D_{max}) before calculation of *ab initio* shape envelopes and analysis via EOM to investigate the ensemble of conformers that best fit the data. All parameters and statistics for both the independent Guinier and *P(r)* SAXS analyses to determine model-free parameters and subsequent modelling and validation of the data are shown in Table S2.

Initial SEC-SAXS analysis of both Men2 ACP_2-TE and ACP_1-ACP_2-TE removed trace amounts of higher molecular weight species, with both SEC elution peaks containing a consistent R_g value (Fig. S7), which were subsequently processed. Buffer subtraction and data averaging for individual elution peaks yielded a log *I(q)* vs *q* plot for each protein (Fig. S8), which indicated that the shapes of the two proteins were similar, with errors observed from $q > 0.26$ Å^{-1}. This plot was used to determine molecular weight approximations using SAXSMoW 2.0 [49] suggesting that both ACP_2-TE and ACP_1-ACP_2-TE are monomeric in solution (calculated = 54.1/65.4 kDa, expected = 45.2/57.9 kDa). The *a priori* shape classification of the SAXS log *I(q)* vs *q* plot via DATCLASS [50] designated both proteins as extended in solution. Both datasets displayed a linear Guinier fit (Fig. S8B) and are folded particles, determined by their Gaussian peak form in a Dimensionless Kratky plot (Fig. 2A). Features within this plot, such as the considerable increases in intensity at $q*R_g > 10$ imply these particles have a degree of flexibility and are likely to exist as extended particles rather than rigid entities, in line with the DATCLASS classification, with Men2 ACP_1-ACP_2-TE displaying greater flexibility than Men2 ACP_2-TE. The interdomain linkers between the ACPs and TE domains are likely to be dynamic in solution in order for the ACP and TE

domains to interact, thus it is unlikely that these species exist in a solely compact conformation. Model-free parameters (R_g and D_{max}) were determined via a pair-wise distribution function (Table S2), where the *P(r)* vs r profiles for each protein display a smooth curve between 0 < r < 40 Å, then sloping towards the *P(r)* = 0 at r = 0 (Fig. 2B). Here, the D_{max} value is larger for ACP_1-ACP_2-TE than ACP_2-TE (172.5 Å vs 144 Å) highlighting that the addition of ACP_1 has increased the maximum particle diameter, thus extending the shape of the protein rather than causing a conformational compaction. This can be seen in DEBS module 6 -TE integrated with two extra ACP domains for higher production of triketide ketolactone resulting in the 40 Å increase of D_{max} value [51]. Likewise, an increase of the R_g in the real domain (6.11 Å) of ACP_1-ACP_2-TE is consistent with extension rather than compaction of the protein upon addition of ACP_1. The agreement between the calculated *I*(0) values from the Guinier and *P(r)* analyses is excellent, although there is a discrepancy between the R_g values with Guinier approximation resulting in lower values (3.96 and 4.24 Å) respectively for ACP_1-ACP_2-TE and ACP_2-TE, with an error value of 8.8 and 10.6 %.

Ab initio and ensemble models of Men2 ACP_2-TE and ACP_1-ACP_2-TE

Following *P*(r) vs r analysis for both species, *ab initio* models of both Men2 ACP_2-TE and ACP_1-ACP_2-TE were generated by DAMMIF, DAMAVER and DAMMIN analyses revealing an elongated flat model. Examples of ten *ab initio* and final averaged envelopes of Men2 ACP_2-TE and ACP_1-ACP_2-TE are shown in Fig. S9. However, given the expected flexibility and dynamicity of these constructs, *ab initio* models are insufficient to fully capture the range of conformations these proteins may adopt. We therefore extended the *ab initio* modelling with quantitative flexibility analysis using EOM. This approach is suitable for generating models derived from SAXS data of intrinsically disordered/highly flexible

proteins [52]. First, a large pool of models with possible conformations is generated and an ensemble of conformers is selected, so that the resultant computed scattering profile best describes the experimentally observed scattering profile [38]. This approach can also combine multiple pools generated from different EOM iterations to model the mixture of *apo* and *holo* forms of ACP_2-TE and ACP_1-ACP_2-TE.

The homology models of ACP_1, ACP_2 and TE domains were used as starting models, with the N-terminal His_6-tag and interdomain linkers treated as randomized flexible linkers to generate a pool of 10,000 models. In order to gain the possible models derived from SAXS data of our samples containing both *apo* and *holo* species, initial EOM analysis first generated pools of 10,000 models for individual *apo* and *holo* forms of Men2 ACP_2-TE and ACP_1-ACP_2-TE with randomised flexible linkers (Fig. S10-11), before combining pools to model a mixture of *apo* and *holo* for both proteins (Fig. 3) with a ratio of each pool based on mass spectrometry results.

For ACP_2-TE, both *apo* and *holo* EOM profiles adopt a tri-population of discernible conformers with R_g values between 25-30, 37-45 and 48-52 Å (Fig. S10). These sub-population peaks are broadened when modelling a mixture of both forms (Fig 3A), although peaks at 25-30 and 37-45 Å are still observed, with a good fit of the selected ensemble of models to the experimental SAXS profile (χ^2= 1.15). Due to the broad profile of conformers generated, the distributions of the R_g and D_{max} parameters in the selected ensemble structures are both smaller (R_g: 26.61-31.57 Å, D_{max}: 90.35-106.07 Å) and significantly larger (R_g: 37.60-57.92 Å, D_{max}: 108.91-168.33 Å) than the average values of the original random pool of EOM models (R_g: 34.03 Å, D_{max}: 107.60 Å), indicating the presence of both compact and extended structures in solution and the extent of flexibility within the interdomain ACP_2- TE linker.

For ACP_1-ACP_2-TE, the distributions of the overall parameters (R_g and D_{max}) in the EOM selected structures for the mixture of *apo* and *holo* (Fig 3B) indicated a clear population of conformers (R_g: 25-50 Å, D_{max}: 90-145 Å). Models with R_g > 55 Å were also present, indicating the ensemble conformers were on the whole more extended in solution (Average R_g: 46.87 Å, D_{max}: 141.43 Å) when compared to the random pool of conformers (Average Rg: 42.21 Å, Dmax: 132.89 Å). However, several compact conformations are also sampled (Fig 3D). The selected ensemble of *apo/holo* ACP_1-ACP_2-TE conformers fit the experimental data with $\chi^2 = 2.42$.

Based on our EOM results, all models in the ensemble generated from *apo*, *holo* and combined pools for both ACP_2-TE and ACP_1-ACP_2-TE have a similar topology, existing as a Beads-on-a-String configuration, being able to flex between a range of compact and extended variants in solution. The presence of a phosphopantetheine arm in the *holo* form does not appear to bias the topology towards a more compact or extended minima (population), suggesting it does not play a structural role in this system and is required only for the shuffling of intermediates between domains. The sequential nature of the ACPs within ACP_1-ACP_2-TE enables it to sample a larger conformational space, reflected by the greater R_g and D_{max} span when compared to ACP_2-TE. It is a possibility that this extended form is a highly populated conformation naturally and promotes transfer of the polyketide intermediate from other partner domains, before contracting to shuffle the intermediate through ACP_2 towards the TE domain. This can be seen in some of ACP_1-ACP_2-TE models. As this study investigated constructs without the presence of attached intermediates to either ACP domains, the observation that a sub-population of compact conformers are still sampled, hints at a more complex arrangement of domains than a straightforward isolated Beads-on-a-String model.

Here, we show that SAXS techniques, coupled with homology modelling, can be used to gain important structural insights into the molecular organization of multidomain PKS

components. This is of particular significance given the size and flexibility of these proteins, which represent challenging targets for X-ray crystallography and NMR. The elongated shape models of both Men2 ACP_2-TE and ACP_1-ACP_2-TE obtained from SAXS do not reveal any interaction between ACP and TE domains. A productive encounter between ACP and TE may require specific interactions between the TE active site and the polyketide and therefore the presence of the intermediate on the ACP becomes important. In the type I erythromycin (DEBS) PKS, ^{13}C NMR and chemical shift perturbations were used to infer an interaction between the carbonyl of the ACP bound acyl intermediate and TE [53]. Despite this, ACP/TE interactions could not be detected, underlining the challenges in probing these systems. As a highly flexible system without high resolution structure, the Beads-on-a-String models of both Men2 ACP_2-TE and ACP_1-ACP_2-TE could be used as a platform for ACP-TE interaction studies. Shape comparison could be performed upon modification of the ACP domain, such as loading with crosslinker probe to trap the interaction between ACP and TE domains [54-56] or to begin to dissect the influence of alternate linkers and assess their structural bias. Our findings also provide a framework for ACP-TE domain arrangement in Men2 and related systems. This could be of value in directing approaches for the reengineering of this class of megasynthases by domain swapping with other related systems for novel metabolite production or even adding more ACP domains for enhancing production of metabolites.

Conclusions

In this study, recombinant overexpression of the Men2 TE domain from the menisporopsin A NR-PKS has been achieved in *E. coli* and significantly enhanced by production as either a mono or didomain ACP fusion protein. Solution characterization of both Men2 ACP_2-TE and ACP_1-ACP_2-TE using SAXS has served to highlight the degree of

flexibility in these multidomain proteins and provided the first insights into TE and ACP arrangement in a fungal iterative PKS**.**

Acknowledgements

This work was supported by the Royal Society, the Newton Fund (NA150457), the Thailand Research Fund (TRF) (DBG5980009), the Center of Excellence for Innovation in Chemistry (PERCH-CIC), Ministry of Higher Education, Science, Research and Innovation, and BBSRC grants, BB/T001968/1, BB/L01386X/1 and BBS/OS/GC/000004. We also thank the Royal Thai Government for a DPST Fellowship for W.B.

Author contributions

WB and PW constructed the expression vectors; WB performed protein expression, purification, identification and assays; WB, AJW and PRR collected SAXS data; AJW analysed SAXS data; AJW and PW performed ensemble modelling; CML provided tool for cloning; PW, PRR and MPC conceived, designed and supervised the study. All authors contributed to the writing of manuscript.

Conflict of Interests

The authors declare that they have no conflict of interest.

References

1. Chinworrungsee, M., Kittakoop, P., Isaka, M., Maithip, P., Supothina, S. & Thebtaranonth, Y. (2004) Isolation and structure elucidation of a novel antimalarial macrocyclic polylactone, menisporopsin A, from the fungus Menisporopsis theobromae, *J Nat Prod.* **67**, 689-92.
2. Wattana-amorn, P., Juthaphan, P., Sirikamonsil, M., Sriboonlert, A., Simpson, T. J. & Kongkathip, N. (2013) Biosynthetic origins of menisporopsin A, *J Nat Prod.* **76**, 1235-7.
3. Bunnak, W., Wonnapinij, P., Sriboonlert, A., Lazarus, C. M. & Wattana-Amorn, P. (2019) Heterologous biosynthesis of a fungal macrocyclic polylactone requires only two iterative polyketide synthases, *Org Biomol Chem.* **17**, 374-379.

4. Takano, Y., Kubo, Y., Shimizu, K., Mise, K., Okuno, T. & Furusawa, I. (1995) Structural analysis of PKS1, a polyketide synthase gene involved in melanin biosynthesis in Colletotrichum lagenarium, *Mol Gen Genet.* **249**, 162-7.
5. Mayorga, M. E. & Timberlake, W. E. (1992) The developmentally regulated Aspergillus nidulans wA gene encodes a polypeptide homologous to polyketide and fatty acid synthases, *Mol Gen Genet.* **235**, 205-12.
6. Yu, J. H. & Leonard, T. J. (1995) Sterigmatocystin biosynthesis in Aspergillus nidulans requires a novel type I polyketide synthase, *J Bacteriol.* **177**, 4792-800.
7. Crawford, J. M., Korman, T. P., Labonte, J. W., Vagstad, A. L., Hill, E. A., Kamari-Bidkorpeh, O., Tsai, S. C. & Townsend, C. A. (2009) Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization, *Nature.* **461**, 1139-43.
8. Wattana-amorn, P., Williams, C., Ploskon, E., Cox, R. J., Simpson, T. J., Crosby, J. & Crump, M. P. (2010) Solution structure of an acyl carrier protein domain from a fungal type I polyketide synthase, *Biochemistry.* **49**, 2186-93.
9. Korman, T. P., Crawford, J. M., Labonte, J. W., Newman, A. G., Wong, J., Townsend, C. A. & Tsai, S. C. (2010) Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis, *Proc Natl Acad Sci U S A.* **107**, 6246-51.
10. Edwards, A. L., Matsui, T., Weiss, T. M. & Khosla, C. (2014) Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase, *J Mol Biol.* **426**, 2229-45.
11. Davison, J., Dorival, J., Rabeharindranto, H., Mazon, H., Chagot, B., Gruez, A. & Weissman, K. J. (2014) Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module, *Chemical Science.* **5**, 3081-3095.
12. Frueh, D. P., Arthanari, H., Koglin, A., Vosburg, D. A., Bennett, A. E., Walsh, C. T. & Wagner, G. (2008) Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase, *Nature.* **454**, 903-6.
13. Drake, E. J., Miller, B. R., Shi, C., Tarrasch, J. T., Sundlov, J. A., Allen, C. L., Skiniotis, G., Aldrich, C. C. & Gulick, A. M. (2016) Structures of two distinct conformations of holo-non-ribosomal peptide synthetases, *Nature.* **529**, 235-8.
14. Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I. & Owens, R. J. (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications, *Nucleic Acids Res.* **35**, e45.
15. Lobley, A., Whitmore, L. & Wallace, B. A. (2002) DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra, *Bioinformatics.* **18**, 211-2.
16. Whitmore, L. & Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data, *Nucleic Acids Res.* **32**, W668-73.
17. Whitmore, L. & Wallace, B. A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases, *Biopolymers.* **89**, 392-400.
18. Provencher, S. W. & Glockner, J. (1981) Estimation of globular protein secondary structure from circular dichroism, *Biochemistry.* **20**, 33-7.
19. van Stokkum, I. H., Spoelder, H. J., Bloemendal, M., van Grondelle, R. & Groen, F. C. (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra, *Anal Biochem.* **191**, 110-8.
20. Lees, J. G., Miles, A. J., Wien, F. & Wallace, B. A. (2006) A reference database for circular dichroism spectroscopy covering fold and secondary structure space, *Bioinformatics.* **22**, 1955-62.
21. Micsonai, A., Wien, F., Kernya, L., Lee, Y. H., Goto, Y., Refregiers, M. & Kardos, J. (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy, *Proc Natl Acad Sci U S A.* **112**, E3095-103.
22. Micsonai, A., Wien, F., Bulyaki, E., Kun, J., Moussong, E., Lee, Y. H., Goto, Y., Refregiers, M. & Kardos, J. (2018) BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra, *Nucleic Acids Res.* **46**, W315-W322.
23. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction, *BMC Bioinformatics.* **9**, 40.

24. Roy, A., Kucukural, A. & Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction, *Nat Protoc.* **5**, 725-38.
25. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. & Zhang, Y. (2015) The I-TASSER Suite: protein structure and function prediction, *Nat Methods.* **12**, 7-8.
26. Zhou, X., Hu, J., Zhang, C., Zhang, G. & Zhang, Y. (2019) Assembling multidomain protein structures through analogous global structural alignments, *Proc Natl Acad Sci U S A.* **116**, 15930-15938.
27. Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I. A., Lesley, S. A. & Godzik, A. (2007) XtalPred: a web server for prediction of protein crystallizability, *Bioinformatics.* **23**, 3403-5.
28. Wootton, J. C. & Federhen, S. (1993) Statistics of local complexity in amino acid sequences and sequence databases, *Computers & Chemistry.* **17**, 149-163.
29. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr.* **66**, 486-501.
30. Trewhella, J., Duff, A. P., Durand, D., Gabel, F., Guss, J. M., Hendrickson, W. A., Hura, G. L., Jacques, D. A., Kirby, N. M., Kwan, A. H., Perez, J., Pollack, L., Ryan, T. M., Sali, A., Schneidman-Duhovny, D., Schwede, T., Svergun, D. I., Sugiyama, M., Tainer, J. A., Vachette, P., Westbrook, J. & Whitten, A. E. (2017) 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update, *Acta Crystallogr D Struct Biol.* **73**, 710-728.
31. *https://bl1231alslblgov/scatter/*.
32. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V. & Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis, *J Appl Crystallogr.* **45**, 342-350.
33. Franke, D. & Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering, *J Appl Crystallogr.* **42**, 342-346.
34. Volkov, V. V. & Svergun, D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering, *J Appl Cryst.* **36**, 860-864.
35. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, *Biophys J.* **76**, 2879-86.
36. Kozin, M. B. & Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models, *J Appl Cryst.* **34**, 33-41.
37. Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering, *J Am Chem Soc.* **129**, 5656-64.
38. Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. I. (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering, *Iucrj.* **2**, 207-217.
39. Ferrer, M., Chernikova, T. N., Yakimov, M. M., Golyshin, P. N. & Timmis, K. N. (2003) Chaperonins govern growth of Escherichia coli at low temperatures, *Nat Biotechnol.* **21**, 1266-7.
40. Quadri, L. E., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P. & Walsh, C. T. (1998) Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases, *Biochemistry.* **37**, 1585-95.
41. Mootz, H. D., Finking, R. & Marahiel, M. A. (2001) 4'-phosphopantetheine transfer in primary and secondary metabolism of Bacillus subtilis, *J Biol Chem.* **276**, 37289-98.
42. Geoghegan, K. F., Dixon, H. B., Rosner, P. J., Hoth, L. R., Lanzetti, A. J., Borzilleri, K. A., Marr, E. S., Pezzullo, L. H., Martin, L. B., LeMotte, P. K., McColl, A. S., Kamath, A. V. & Stroh, J. G. (1999) Spontaneous alpha-N-6-phosphogluconoylation of a "His tag" in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins, *Anal Biochem.* **267**, 169-84.
43. Crosby, J., Sherman, D. H., Bibb, M. J., Revill, W. P., Hopwood, D. A. & Simpson, T. J. (1995) Polyketide synthase acyl carrier proteins from Streptomyces: expression in Escherichia coli, purification and partial characterisation, *Biochim Biophys Acta.* **1251**, 32-42.
44. Crump, M. P., Crosby, J., Dempsey, C. E., Parkinson, J. A., Murray, M., Hopwood, D. A. & Simpson, T. J. (1997) Solution structure of the actinorhodin polyketide synthase acyl carrier protein from Streptomyces coelicolor A3(2), *Biochemistry.* **36**, 6000-8.

45. Xu, G. Y., Tam, A., Lin, L., Hixon, J., Fritz, C. C. & Powers, R. (2001) Solution structure of B. subtilis acyl carrier protein, *Structure.* **9**, 277-87.
46. Wong, H. C., Liu, G., Zhang, Y. M., Rock, C. O. & Zheng, J. (2002) The solution structure of acyl carrier protein from Mycobacterium tuberculosis, *J Biol Chem.* **277**, 15874-80.
47. Kim, Y., Kovrigin, E. L. & Eletr, Z. (2006) NMR studies of Escherichia coli acyl carrier protein: dynamic and structural differences of the apo- and holo-forms, *Biochem Biophys Res Commun.* **341**, 776-83.
48. Evans, S. E., Williams, C., Arthur, C. J., Burston, S. G., Simpson, T. J., Crosby, J. & Crump, M. P. (2008) An ACP structural switch: conformational differences between the apo and holo forms of the actinorhodin polyketide synthase acyl carrier protein, *Chembiochem.* **9**, 2424-32.
49. Piiadov, V., Ares de Araujo, E., Oliveira Neto, M., Craievich, A. F. & Polikarpov, I. (2019) SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale, *Protein Sci.* **28**, 454-463.
50. Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., Kikhney, A. G., Hajizadeh, N. R., Franklin, J. M., Jeffries, C. M. & Svergun, D. I. (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions, *J Appl Crystallogr.* **50**, 1212-1225.
51. Wang, Z., Bagde, S. R., Zavala, G., Matsui, T., Chen, X. & Kim, C. Y. (2018) De Novo Design and Implementation of a Tandem Acyl Carrier Protein Domain in a Type I Modular Polyketide Synthase, *ACS Chem Biol.* **13**, 3072-3077.
52. Kikhney, A. G. & Svergun, D. I. (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins, *FEBS Lett.* **589**, 2570-7.
53. Tran, L., Broadhurst, R. W., Tosin, M., Cavalli, A. & Weissman, K. J. (2010) Insights into protein-protein and enzyme-substrate interactions in modular polyketide synthases, *Chem Biol.* **17**, 705-16.
54. Nguyen, C., Haushalter, R. W., Lee, D. J., Markwick, P. R. L., Bruegger, J., Caldara-Festin, G., Finzel, K., Jackson, D. R., Ishikawa, F., O'Dowd, B., McCammon, J. A., Opella, S. J., Tsai, S. C. & Burkart, M. D. (2014) Trapping the dynamic acyl carrier protein in fatty acid biosynthesis, *Nature.* **505**, 427-431.
55. Barajas, J. F., Finzel, K., Valentic, T. R., Shakya, G., Gamarra, N., Martinez, D., Meier, J. L., Vagstad, A. L., Newman, A. G., Townsend, C. A., Burkart, M. D. & Tsai, S. C. (2016) Structural and Biochemical Analysis of Protein-Protein Interactions Between the Acyl-Carrier Protein and Product Template Domain, *Angew Chem Int Edit.* **55**, 13005-13009.
56. Thiele, G. A. R., Friedman, C. P., Tsai, K. J. S., Beld, J., Londergan, C. H. & Charkoudian, L. K. (2017) Acyl Carrier Protein Cyanylation Delivers a Ketoacyl Synthase-Carrier Protein Cross-Link, *Biochemistry.* **56**, 2533-2536.

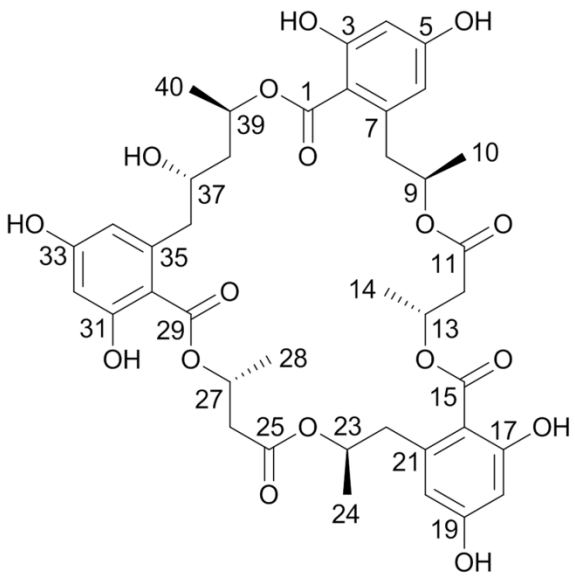
Figure Legends

Fig. 1 Structures of menisporopsin A and ascotrichalactone A and domain organization of Men1 and Men2. The doublet ACP and TE domains of Men2 used in this study are coloured in red and blue, respectively. **(**KS = ketosynthase; MAT = malonyl-CoA:ACP transacylase; ACP = acyl carrier protein; KR = ketoreductase; DH˚ = inactive dehydratase; ER˚ = inactive enoyl reductase; SAT = starter unit:ACP transacylase; PT = product template; TE = thioesterase)

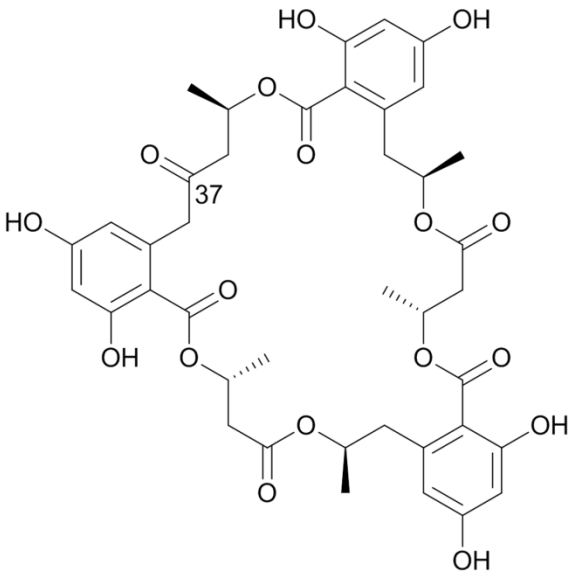
Fig. 2 A) Dimensionless Kratky plots for the data shown in Fig. S8. The peak maximum for both particles, Men2 ACP_2-TE (blue) (1.843, 2.963) and ACP_1-ACP_2-TE (black) (2.009, 3.521), is greater than the values expected for a rigid folded protein (crosshair: 1.1, 1.73), indicating that these particles are likely to be elongated, rather than rigid. A greater deviation from this crosshair and increased intensity at $q^*R_g > 10$ for ACP_1-ACP_2-TE, indicates a more elongated and flexible arrangement in solution when compared to ACP_2-TE. B) *P(r)* vs r profiles for both Men2 ACP_2-TE (blue) and ACP_1-ACP_2-TE (black). The inset displays the fit of this profile to the portion of raw scattering data that is modelled (*q*: 0.00-0.26).

Fig. 3 Ensemble modelling of Men2 A) ACP_2-TE and B) ACP_1-ACP_2-TE using combined *apo* and *holo* pools against the raw experimental scattering with histogram profiles of the R_g and D_{max} distributions of the resultant ensemble models (red) compared to the pool of 10,000 independent models (blue). For each ensemble sub-population, representative models of Men2 C) ACP_2-TE and D) ACP_1-ACP_2-TE are shown, with the linker region between ACP (red) and TE (blue) domains represented as grey spheres. The phosphopantetheinyl group (green) of *holo* ACP_2 is shown in stick representation. Percentages of each model in the ensemble of ACP_2-TE and ACP_1-ACP_2-TE are 9.1 % and 12.5 %, respectively.

Fig. 1



menisporopsin A



ascotrichalactone A

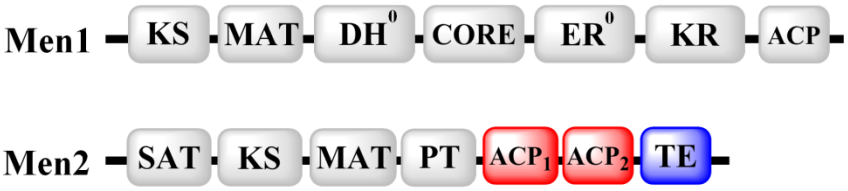


Fig. 2

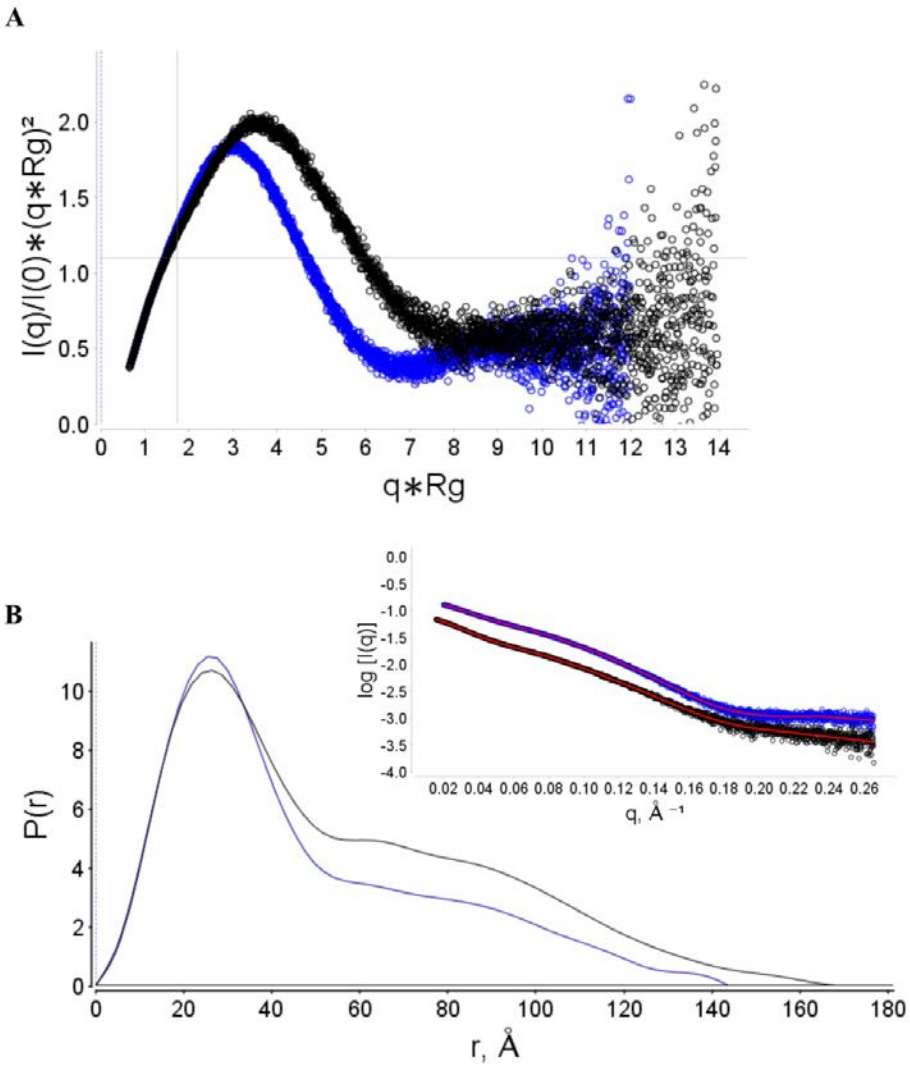
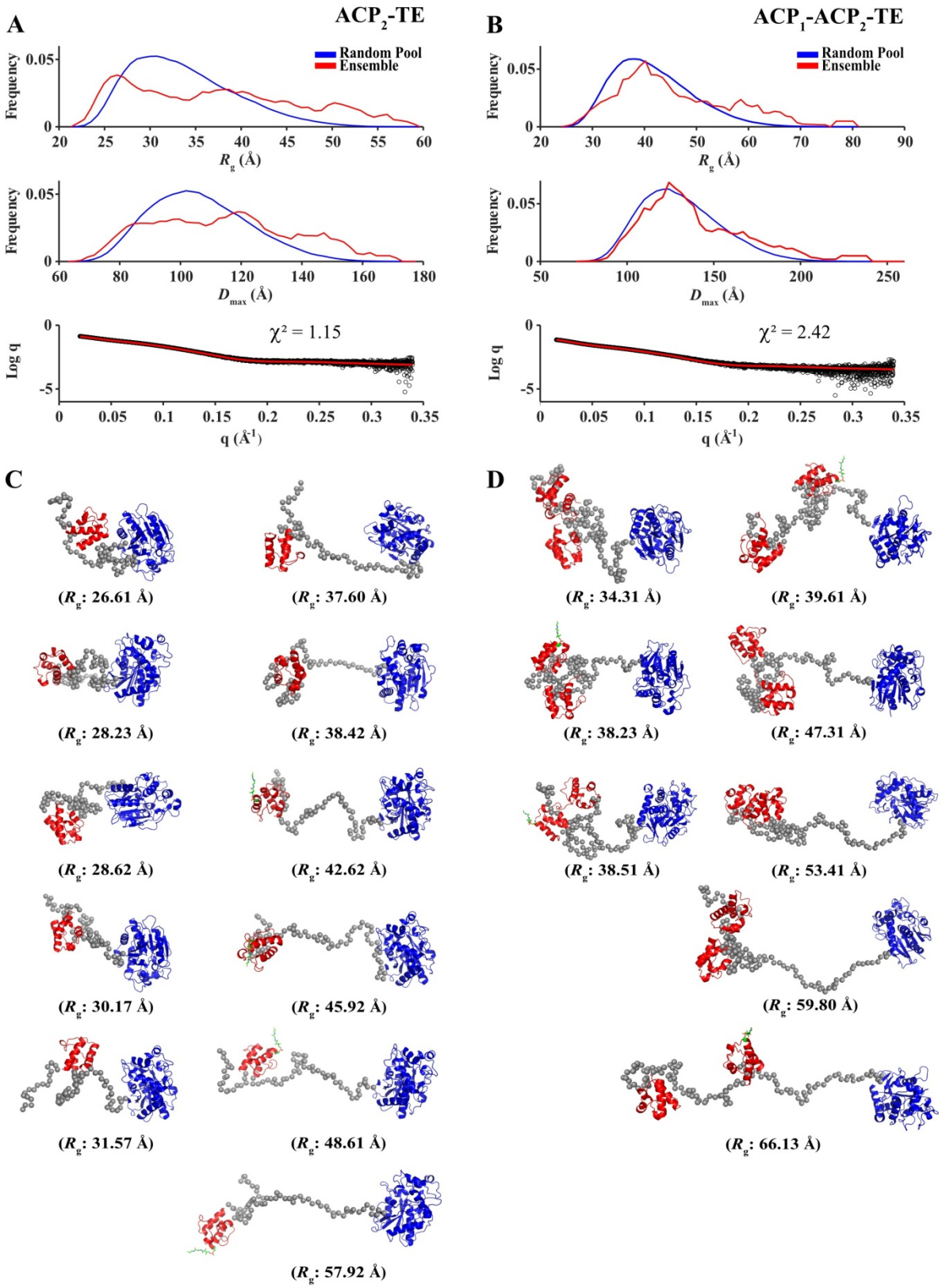


Fig. 3


A
ACP2-TE
Random Pool
Ensemble
Frequency
Rg (Å)
Dmax (Å)
Log q
χ² = 1.15
q (Å-1)
B
ACP1-ACP2-TE
χ² = 2.42
C
(Rg: 26.61 Å)
(Rg: 37.60 Å)
(Rg: 28.23 Å)
(Rg: 38.42 Å)
(Rg: 28.62 Å)
(Rg: 42.62 Å)
(Rg: 30.17 Å)
(Rg: 45.92 Å)
(Rg: 31.57 Å)
(Rg: 48.61 Å)
(Rg: 57.92 Å)
D
(Rg: 34.31 Å)
(Rg: 39.61 Å)
(Rg: 38.23 Å)
(Rg: 47.31 Å)
(Rg: 38.51 Å)
(Rg: 53.41 Å)
(Rg: 59.80 Å)
(Rg: 66.13 Å)