# Mycocerosic Acid Synthase Exemplifies the Architecture of Reducing Polyketide Synthases

Dominik A. Herbst[*,1], Roman P. Jakob[*,1], Franziska Zähringer[2] & Timm Maier[#,1]

[*]Contributed equally

Author affiliation:

[1]Department Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland. [2]Current address: F. Hoffmann-La Roche AG, Grenzacherstrasse 124, 4070 Basel, Switzerland

[#] Corresponding author: timm.maier@unibas.ch

**Polyketide synthases (PKSs) are biosynthetic factories that produce natural products with important biological and pharmacological activities[1-3]. Their exceptional product diversity is encoded in a modular architecture. Modular PKSs (modPKSs) catalyze reactions colinear to the order of modules in an assembly line[3], whereas iterative PKS (iPKSs) use a single module iteratively as exemplified by fungal iPKSs (fiPKSs)[3]. However, in some cases non-colinear iterative action is also observed for modPKSs modules and is controlled by the assembly line environment[4,5]. PKSs feature a structural and functional separation into a condensing and a modifying region as observed for fatty acid synthases (FASs)[6]. Despite PKSs' outstanding relevance, the detailed organization of PKSs with complete fully-reducing modifying regions remains elusive. Here, we report a hybrid crystal structure of *Mycobacterium smegmatis* mycocerosic acid synthase (MAS) based on structures of its condensing and modifying regions. MAS is a fully reducing iPKS, closely related to modPKSs, and the prototype of mycobacterial MAS-like (Msl-)[7,8] PKSs. It is involved in the biosynthesis of $C_{20}$-$C_{28}$ branched-chain fatty acids, which are important virulence factors of mycobacteria[9]. Our structural data reveal a dimeric linker-based organization of the modifying region and visualize dynamics and conformational coupling in PKSs. Based on comparative small angle X-ray scattering (SAXS), the observed modifying region architecture may be common also in modPKSs. The linker based organization provides a rationale for the characteristic variability of PKS modules as a main contributor to product diversity. The comprehensive arechitectural model enables functional dissection and re-engineering of PKSs.**

Each homodimeric PKSs module sequentially elongates acyl-carrier protein (ACP) tethered precursors by the sequential action of an acyltransferase (AT) and a ketosynthase (KS), organized in the essential condensing region (KS-AT). The product can further be sequentially modified by a ketoreductase (ΨKR/KR), a dehydratase (DH), and an enoylreductase (ER)[1,3]. These optional domains form the variable modifying region of PKSs. MAS is a fully-reducing PKS with a complete modifying region (DH-ΨKR-ER-KR). It iteratively elongates linear $C_{12}$-$C_{20}$ starter fatty acids in one to four rounds with methyl-malonyl-CoA extender units[8] to produce mycocerosic acids. These MAS products form the core of phenolic glycolipids and phthiocerol dimycocerosates, key lipids of the mycobacterial cell envelope[8]. The condensing and modifying regions of MAS are centrally connected by non-conserved linkers, which permit large-scale relative motions in related systems[10]. To obtain a high-quality hybrid model, we divided MAS into its condensing and modifying region, and excluded the flexibly tethered ACP (Fig. 1a).

Three constructs of staggered C-terminal length were employed to define the length of the condensing region (see Methods). All variants crystallized under the same condition; structure determination mapped the last ordered residue to Glu887. The structure of the most extended variant (1-892) was refined at 2.3 Å resolution (Extended Data Table 1a). MAS KS-AT comprises an α/β-fold linker domain (LD) connecting AT to KS (Fig. 1b). The monomeric condensing region closely resembles those of other PKSs and FASs[6,11,12], the closest structural homologue at individual domain level is module 5 of the 6-deoxyerythronolide B synthase (DEBS) PKS (Extended Data Table 2a). Compared to previous KS-AT didomain structures, the AT domain is slightly rotated towards the C-terminal post-AT linker.

Isolated MAS KS-AT crystallized as a monomer lacking the canonical KS-based dimerization[6,11,12], but is in monomer-dimer equilibrium in solution with a $K_d$ of 0.4 mM as

determined by analytical ultracentrifugation (AUC). It is the first condensing region crystallized as monomer, but dissociation has been observed for other condensing region fragments in the absence of dimeric partner domains[12-14]. Differences to canonical dimeric KS, as exemplified by DEBS $KS_5$[11] or CurL KS[12], are observed around the dimer interface, presumably due to the absence of stabilizing dimer interactions: The interface-spanning active site tunnel is incomplete and the loop containing the catalytic cysteine (Cys178) is bent outwards by 9 Å into a non-productive conformation, while the active site histidines (His313,349) are at expected positions (Extended Data Fig. 1a-c). Four interface segments of 6-19 amino acid (aa) length are disordered (Fig. 1b), while equivalent regions are ordered in dimeric KS domains.

A single mode of dimerization based on canonical KS organization was identified by automated sequence-based methods (see Methods) and homology-based modeling of dimeric MAS KS-AT restores the active site tunnel and a productive conformation of Cys178 (Extended Data Fig. 1b, c). The KS-AT dimer adopts a linear shape due to the rotation of AT relative to KS (Extended Data Fig. 1d). The C-terminal post-AT linkers of the condensing region, which connect to the modifying region, are proximal to the two-fold dimer axis above the KS active site, as observed in previous condensing region structures[6,11,12].

The DHs connect the modifying region to the post-AT linkers of the condensing region. We solved crystal structures of a MAS DH construct (aa 884-1186), which overlaps in sequence with the crystallized KS-AT, in two crystal forms with a total of six protomers arranged into almost identical dimers (Extended Data Table 1a). The DH protomer is composed of two hot-dog folds connected by a 20 aa hot-dog linker (Fig. 1c). A hydrophobic substrate binding tunnel extends over both hot-dog folds with entrances near the C-terminus and at the distal end of hot-dog fold 2. Active site residues are contributed by both hot-dog folds and are located close to the

C-terminus (Extended Data Fig. 1e). The nearest structural homologs of DH protomers are modPKS DH domains (Extended Data Table 2a). In the DH dimer, the two protomers arrange with their lateral ends bent towards the post-AT linkers with an interdomain angle of 222° (Fig. 1c). The MAS DH dimer is distinct from the V-shaped DH arrangement in FAS[6], which lacks a dimerization interface and is bent into the opposite direction at an angle of 96°. MAS DH rather resembles linear DH dimers of modPKSs with interdomain angles of 167-203°[15-17] and a common mode of dimerization via "handshake" interactions between β-strands of the N-terminal hot-dog folds (Extended Data Fig. 1f-h).

To obtain an authentic representation of the MAS modifying region, we crystallized in presence of $NADP^+$ the complete DH-ΨKR-ER-KR segment, which is dimeric in solution based on AUC. Based on SAXS, ACP deletion is not affecting the overall structure of this region (Extended Data Fig. 2a-c). The crystallographic asymmetric unit reveals a complex packing of nine dimers related by non-crystallographic symmetry (NCS). The corresponding 18 polypeptide chains comprise 20,502 aa (2.2 MDa protein mass), of which 17,680 are modeled. Real-space NCS averaging and NCS-restrained refinement led to a high-quality model ($R_{work}/R_{free}$ = 0.23/0.24) at 3.75 Å resolution (Fig. 2, Extended Data Table 1a, Extended Data Fig. 2d-f). The modifying region dimerizes along an extended interface formed by DH and ER (Extended Data Table 2b); the ΨKR/KR is laterally connected to DH and ER. MAS, as well as most reducing modPKSs, lacks a non-catalytic pseudo-methyltransferase domain (ΨME), which is a characteristic of FASs and fiPKSs. The DH in the modifying region adopts the same dimeric structure as in the isolated form (Extended Data Table 2a), demonstrating the intrinsic nature of DH dimerization and its role in organizing the modifying region. The ER domain is characterized by a large active site tunnel and a well-ordered $NADP^+$ cofactor (Extended Data

Fig. 3a, b). The ER dimerizes via pseudo-continuous β-sheet formation between the nucleotide binding subdomains ($ER_{NB}$) and provides the largest contribution to the modifying region dimer interface. Its closest structural neighbors are the isolated modPKS ERs from *Lyngbya majuscula*[18] and the SpnB ER-ΨKR/KR didomain[19] (Extended Data Table 2a), even though these ERs are monomeric. The dimerization mode of MAS ER closely resembles those of the $ER_{NB}$ subdomain of the PpsC modPKS and the ER of FAS[6,20] (Extended Data Fig. 3c). The split ΨKR/KR resembles modPKSs ΨKR/KR (Extended Data Table 2a)[21]; as in related B-type KR domains[22], a flexible lid region (aa 1948-1960) remains disordered in the absence of ligand, and concomitantly, the nicotinamide moiety of $NADP^+$ is disordered (Extended Data Fig. 3d). The MAS ΨKR exhibits an N-terminal β-α-β-α extension, which is commonly observed in modPKSs, but not in FASs[6,23]; this extension exhibits increased flexibility as indicated by temperature factor distributions (Extended Data Fig. 3e, f).

Previously, modifying region architecture was discussed based on domain interfaces in FAS and PKSs fragments[1]. However, the current analysis of the MAS modifying region reveals a striking absence of stable interfaces between the different domains: The ER dimer rests on a platform formed by the DH dimer, but the interface between the two is small and variable (345-638 $Å^2$) (Extended Data Table 2b, Supplementary Video 1). The ΨKR/KR does not contact its neighboring domains at all and is the region of highest structural variability. Instead, the architecture of the modifying region is based on three linkers interconnecting the ΨKR/KR, DH, and ER domains, which act as spacers as well as interaction partners amongst each other and with catalytic domains (Fig. 3a, b, Extrended Data Fig. 2d): (i) The 27-aa ΨKR-ER linker plays a central organizing role by forming extended interfaces to ΨKR/KR (975±28 $Å^2$) and ER (353±20 $Å^2$). Moreover, it interacts with each of the other two linkers via two double-stranded,

antiparallel β-sheets. The β-sheet formed between the ΨKR-ER and DH-ΨKR linkers (B1 in Fig. 3) is embedded in a surface groove of the ΨKR/KR. It partially extends the Rossman-fold of the KR and is conserved in PKSs (Extended Data Fig. 4). The β-sheet between the ER-KR and ΨKR-ER linker (B2 in Fig. 3) mostly interacts with the ER and establishes a gap between the ER and KR. (ii) The 38-aa DH-ΨKR linker comprises an N-terminal 10-aa α-helix ($\alpha_{D\Psi}$ Fig. 3) followed by the β-strand paired to ΨKR-ER linker and an irregular segment (S1 in Fig. 3a), which wraps around the ΨKR. Helix $\alpha_{D\Psi}$ separates DH and ΨKR/KR; fragments of it are also observed in structures of isolated DH domains from the Curacin PKS (CurH, K, J)[15]. (iii) The 20-aa ER-KR linker consists of a terminal irregular segment (S3 in Fig. 3) and the central β-strand paired to the ΨKR-ER linker. It contacts ER and KR via interfaces of 432±24 $Å^2$ and 547±14 $Å^2$, respectively, and together with the ΨKR-ER linker forms a continuous connection layer between these domains.

To obtain a MAS hybrid model we connected the overlapping modifying and condensing region fragments *in silico* (Fig. 4a). We assume that the condensing region adopts a canonical dimeric state upon tethering to the dimeric modifying domain. The relative orientation of the condensing and modifying regions is not defined by the two structures and was chosen in accordance to intact FAS[6]. As in FAS, the two fragments connect without secondary contacts outside the linking region. Based on multiple modes of motion around the central linkage observed in FAS[10], the selected orientation may represent only one out of an ensemble of states in both multienzymes. Helix formation of the sequence segment linking modifying and condensing domain was observed at the N-terminus of four protomers in the crystallized modifying region under stabilization by crystal contacts. The central connection in the hybrid model consequently was modelled with an α-helix (Extended Data Fig. 5a, b), in contrast to an

irregular linker in FAS. Notably, short helices in equivalent sequence positions are observed in modPKS DEBS $DH_4$[16] and Rif $DH_{10}$[17] as well as in RhiE KS-B[24] (Extended Data Fig. 5c, d), suggesting a more general conservation of helical linkers in modPKSs.

Conformational dynamics are a key component of multi-enzyme action. They have been visualized by EM for FAS[10] and PikAIII[13,25], but not at resolutions required for mechanistic dissection. The crystallographic visualization of 18 instances of the modifying regions now provides an opportunity to analyze conformational variability in MAS. The central DH and ER dimers each behave as rigid bodies, but the ERs move in a screw motion with a translation of up to 8.5 Å and a rotation by 14° on the DH platform (Fig. 4b, Extended Data Fig. 6a, b). The ERs are conformationally coupled to the ΨKR/KRs (Supplementary Video 1): Owing to the tethering of ΨKR/KR to both, the DH and ER, the screw motion of the ERs is transduced into a rotation of the ΨKR/KRs by up to 40° via a pivot in the linkers ( Extended Data Fig. 6c, d). Even larger motions may occur in solution, as indicated by pronounced disorder of some ΨKR/KRs in the crystal. Importantly, conformational coupling via relative DH-ER motions provides crosstalk between the two lateral clefts of MAS. Although a mechanism for reading out active site states remains unknown, this coupling could transmit reaction states across the MAS dimer. Notably, the mobile ACP is tethered to the most flexible catalytic domain (ΨKR/KR), creating a hierarchic network of gradually increasing domain flexibility.

Only one condensing region instance has been visualized here, but it extends the previously observed range of KS-AT conformations[6,11,12] (Extended Data Fig. 6e, f). MAS KS-AT features the most linear conformation, which results in narrowing the gap to the modifying region and shortening of the AT-ACP anchor distances. Variations between condensing regions correspond to a hinge-bending motion of AT around a pivot in LD (Supplementary Video 2, Fig.

4b). Although experimental evidence of flexibility in each system is lacking, normal-mode analysis indicates a conservation of this hinge in all KS-AT didomains. In the EM reconstruction of PikAIII the AT domain is rotated by approx. 90° relative to MAS and remains a clear outlier to the set of KS-AT regions depicted by crystallography[6,11,12], EM[10], and SAXS[14].

The MAS hybrid model is a prototype for Msl-PKS organization[7]. Moreover, our structural data reinforce the sequence-based conclusion that MAS also serves as a paradigm for modPKSs. Despite its iterative mode of action, MAS is clearly assigned phylogenetically to modPKSs (27-35 % sequence identity) rather than fiPKSs (20-22 % id.) or FASs (19 % id.) (Extended Data Fig. 7). Structurally, the closest neighbors of all individual MAS domains are from modPKSs. The absence of a ΨME domain and the presence of a ΨKR β-α-β-α extension distinguish MAS and most modPKSs from FASs and fiPKSs. "Handshake" interactions of isolated dimeric DHs are observed only in modPKSs, but not in FAS. Our structural data reinforce the earlier hypothesis that modPKSs are fundamentally similar to non-colinear iPKSs such as MAS, and presumably evolved by kinetic coupling of modules[26]. Indeed, several modPKS modules act in an iterative mode as part of an assembly line (*e.g.* BorA5[27], AurA[28]). Other modPKS modules can be converted into a non-colinear mode of action by mutation, *e.g.* DEBS module 3[29], or by isolation from their assembly line environment, *e.g.* PikAIII[30].

The analysis of the hybrid MAS structure depicts a unique PKS architecture. It agrees with previous biochemical and structural data on modPKSs fragments, with the exception of the monomeric state of some isolated ER domains[18] or in the domain-swapped crystal structure of the excised ER-ΨKR/KR didomain of the fully-reducing modPKS SpnB[19]. Based on the structure and the monomeric solution state of SpnB-ER-ΨKR/KR, as well as shorter ER-KR linker in modPKSs, a divergent architecture of modPKSs modifying regions based on a dimeric

DH arrangement with laterally positioned monomeric KRs and ERs was proposed[19]. On the contrary, the MAS modifying region retains a central dimeric ER as observed in FAS and in a fragment of the modPKS PpsC ER (PDB: #1PQW) (Extended Data Fig. 3c). Importantly, MAS reveals a dynamic linker-based organization, which (in contrast to FAS) could also accommodate the typical range of ER-KR linker lengths (5-22 aa) observed in modPKSs (Extended Data Fig. 4) by slight adaptions of the ΨKR/KR position.

Models of PKS modifying regions based on SpnB-ER-ΨKR/KR and MAS are clearly distinct on a macromolecular scale and can be experimentally distinguished via SAXS distance distributions. We selected two well-expressed modifying regions from modPKSs bimodules, EryA of *Gamma proteobacterium* HdN1 (GpEryA) and "Pks" (Uniprot: Q3L885) from *Mycobacterium smegmatis* (MsPks) for comparative SAXS analysis. Calculated SAXS curves for SpnB-ER-ΨKR/KR and MAS-like models were compared with experimental SAXS data of MAS, GpEryA, and MsPks. The derived distance distributions closely match those calculated from a MAS-like model, but not those based on SpnB-ER-ΨKR/KR (Extended Data Fig. 8). The SAXS analysis of GpEryA and MsPks clearly supports a wider relevance of the MAS architecture for modPKSs.

Our structural analysis not only provides detailed insights into MAS, a mycobacterial drug target, but also establishes a new paradigm for the organization of PKSs modules. It reveals a unique, dynamic structure of the modifying region based on dimeric DH and ER domains and provides insights into conformational variability and coupling in fully-reducing PKS modifying regions. The linker-based architecture supports modularity of the modifying region by requiring only the adaptation of variable linker regions for evolutionary domain shuffling. It thus rationalizes an important aspect of the outstanding success of the PKS architecture in the

generation of chemical diversity. Our results highlight the relevance of matching linker-, rather than domain-domain interactions in PKS engineering. They contribute to the fundamental understanding of PKS architecture, as well as to the functional dissection and re-engineering of related synthases including relevant drug targets and important producers of bioactive compounds.

## References


1 Weissman, K. J. Uncovering the structures of modular polyketide synthases. *Nat. Prod. Rep.* (2014).

2 Wong, F. T. & Khosla, C. Combinatorial biosynthesis of polyketides--a perspective. *Curr. Opin. Chem. Biol.* **16**, 117-123 (2012).

3 Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed. Engl.* **48**, 4688-4716 (2009).

4 Busch, B. *et al.* Multifactorial control of iteration events in a modular polyketide assembly line. *Angew. Chem. Int. Ed. Engl.* **52**, 5285-5289 (2013).

5 Moss, S. J., Martin, C. J. & Wilkinson, B. Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Nat. Prod. Rep.* **21**, 575-593 (2004).

6 Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315-1322 (2008).

7 Sirakova, T. D., Thirumala, A. K., Dubey, V. S., Sprecher, H. & Kolattukudy, P. E. The Mycobacterium tuberculosis pks2 gene encodes the synthase for the hepta- and octamethyl-branched fatty acids required for sulfolipid synthesis. *J. Biol. Chem.* **276**, 16833-16839 (2001).

8 Chopra, T. & Gokhale, R. S. Polyketide versatility in the biosynthesis of complex mycobacterial cell wall lipids. *Methods Enzymol.* **459**, 259-294 (2009).

9 Cambier, C. J. *et al.* Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. *Nature* **505**, 218-222 (2014).

10 Brignole, E. J., Smith, S. & Asturias, F. J. Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nat. Struct. Mol. Biol.* **16**, 190-197 (2009).

11 Tang, Y., Kim, C. Y., Mathews, II, Cane, D. E. & Khosla, C. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11124-11129 (2006).

12 Whicher, J. R. *et al.* Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem. Biol.* **20**, 1340-1351 (2013).

13 Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510**, 512-517 (2014).

14 Edwards, A. L., Matsui, T., Weiss, T. M. & Khosla, C. Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase. *J. Mol. Biol.* **426**, 2229-2245 (2014).

15 Akey, D. L. *et al.* Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure* **18**, 94-105 (2010).

16 Keatinge-Clay, A. Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* **384**, 941-953 (2008).

17 Gay, D., You, Y. O., Keatinge-Clay, A. & Cane, D. E. Structure and Stereospecificity of the Dehydratase Domain from the Terminal Module of the Rifamycin Polyketide Synthase. *Biochemistry* (2013).
18 Khare, D. *et al.* Structural Basis for Cyclopropanation by a Unique Enoyl-Acyl Carrier Protein Reductase. *Structure* (2015).
19 Zheng, J., Gay, D. C., Demeler, B., White, M. A. & Keatinge-Clay, A. T. Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol.* **8**, 615-621 (2012).
20 Sippel, K. H., Vyas, N. K., Zhang, W., Sankaran, B. & Quiocho, F. A. Crystal structure of the human fatty acid synthase enoyl-acyl carrier protein-reductase domain complexed with triclosan reveals allosteric protein-protein interface inhibition. *J. Biol. Chem.* **289**, 33287-33295 (2014).
21 Keatinge-Clay, A. T. A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.* **14**, 898-908 (2007).
22 Bonnett, S. A. *et al.* Structural and stereochemical analysis of a modular polyketide synthase ketoreductase domain required for the generation of a cis-alkene. *Chem. Biol.* **20**, 772-783 (2013).
23 Hardwicke, M. A. *et al.* A human fatty acid synthase inhibitor binds beta-ketoacyl reductase in the keto-substrate site. *Nat. Chem. Biol.* **10**, 774-779 (2014).
24 Bretschneider, T. *et al.* Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature* **502**, 124-128 (2013).
25 Whicher, J. R. *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560-564 (2014).
26 Sugimoto, Y. *et al.* Freedom and Constraint in Engineered Noncolinear Polyketide Assembly Lines. *Chem. Biol.* (2015).
27 Olano, C. *et al.* Biosynthesis of the angiogenesis inhibitor borrelidin by Streptomyces parvulus Tu4055: cluster analysis and assignment of functions. *Chem. Biol.* **11**, 87-97 (2004).
28 He, J. & Hertweck, C. Iteration as programmed event during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster. *Chem. Biol.* **10**, 1225-1232 (2003).
29 Kapur, S. *et al.* Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4110-4115 (2012).
30 Beck, B. J., Aldrich, C. C., Fecik, R. A., Reynolds, K. A. & Sherman, D. H. Iterative chain elongation by a pikromycin monomodular polyketide synthase. *J. Am. Chem. Soc.* **125**, 4682-4683 (2003).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

## Acknowledgements

We acknowledge Friedrich Widdel and Johannes Zedelius for providing *Gamma proteobacterium* HdN1, Peter Leadlay and Lorena Betancor for providing plasmid pETcoco-2A-L1SL2, and EMBL Heidelberg for providing the pETG-10A vector; John Missimer (Paul-

Scherrer Institute, Villigen, Switzerland (PSI) and Andreas Menzel (PSI) for outstanding support in SAXS data acquisition and raw data processing; Timothy Sharpe (Biophysics Facility, Biozentrum, University of Basel) for analytical ultracentrifugation, Adam Mazur (Research IT, Biozentrum) for SAXS refinement and Martino Bertoni (Swiss Institute of Bioinformatics and Biozenturm) for support of the homology-based assignment of the oligomeric state of MAS KS-AT. Data were collected at beamlines PXI, PXIII and cSAXS of PSI; we acknowledge excellent support from the beamline teams. This work was supported by the Swiss National Science Foundation (SNF) project grants 125357, 138262, 159696 and R'equip grant 14502. D.A.H. acknowledges a fellowship by the Werner-Siemens Foundation.

**Author Contributions**

R.P.J. expressed, purified and crystallized MAS, obtained the crystal structure of the condensing region, collected SAXS data and cloned constructs. F.Z cloned constructs and purified MAS, GpEryA and MsPks. D.A.H. purified MAS, optimized MAS crystallization, determined the structure of the isolated DH domains and the modifying region, collected SAXS data, analyzed the data, performed homology modelling, cloned constructs, and wrote the manuscript. TM designed and guided research, analyzed data, contributed to crystallographic analysis and wrote the manuscript.

**Author Information**

Address for all authors: Department Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel. E-mail addresses: dominik.herbst@unibas.ch, roman.jakob@unibas.ch, franziska.zaehringer@googlemail.com, timm.maier@unibas.ch

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession code 5BP1, 5BP2, 5BP3, 5BP4

Reprints and permissions information is available at www.nature.com/reprints

Correspondence and requests for materials should be addressed to timm.maier@unibas.ch.

The authors declare no competing interests.

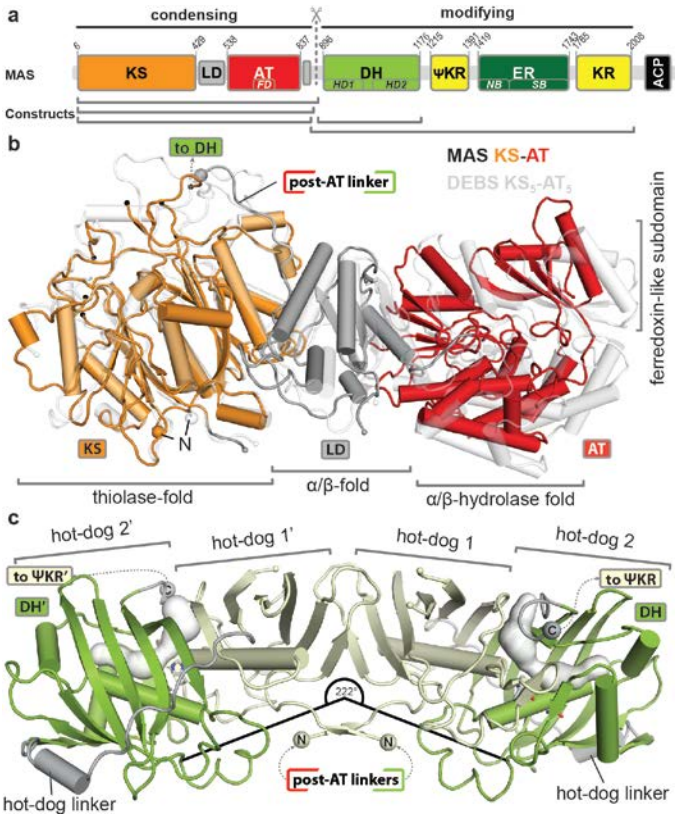



**Figure 1 | Domain organization, condensing region, and dimeric DH domain of MAS. a**, MAS is organized in a condensing (KS: ketosynthase, LD: linker domain, AT: acyltransferase, $AT_{FD}$: ferredoxin-like AT subdomain) and a modifying region (DH: dehydratase, $DH_{HD1/2}$: DH hot-dog fold 1/2, KR: ketoreductase, ΨKR: non-catalytic pseudo-KR domain, ER: enoylreductase, $ER_{NB/SB}$: nucleotide/substrate binding ER subdomain), followed by a flexibly-tethered acyl carrier protein domain (ACP). Crystallized constructs are indicated. **b**, Monomeric condensing region crystal structure. The AT position corresponds to a rotation around a hinge in the LD relative to DEBS $KS_5$-$AT_5$[11] (white). Black spheres indicate ends of disordered segments (aa 47-65, 132-151, 211-220, 277-283). **c**, Crystal structure of the dimeric DH. Each monomer comprises two hot-dog folds connected by a 20 aa hot-dog linker (grey). The DH active site tunnel (white) has two openings and the dimer is bent with an interdomain angle of 222°.

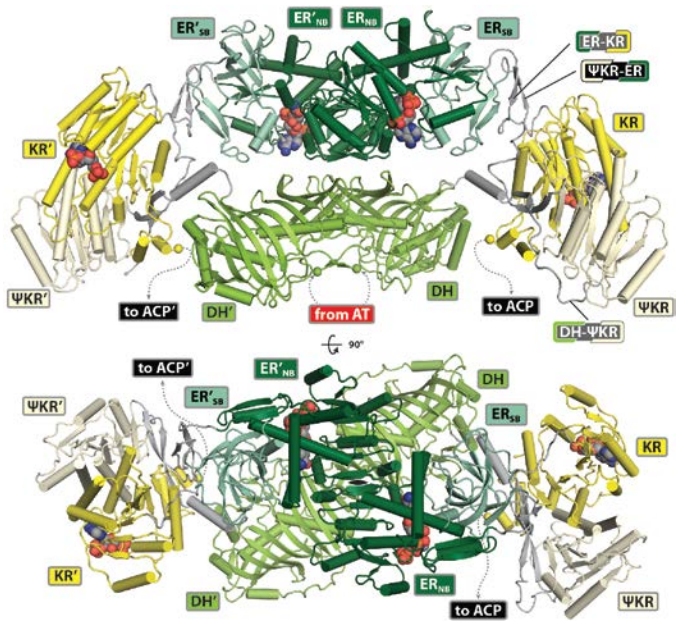



**Figure 2 | Crystal structure of the dimeric MAS modifying region.** The MAS modifying region is organized by the dimerization of the central DH (light green) and ER (darker greens) domains (upper panel: front view, lower panel: top view). The DH dimer reveals virtually the same bent organization as observed in the crystal structures of the isolated DH domains. The ΨKR/KR (yellow) domains are laterally tethered, share no direct interface with any other catalytic domain, and their positioning is the most variable of all domains. Bound cofactors are shown in sphere representation colored by element type. A two fold dimer axis is indicated in the lower panel.

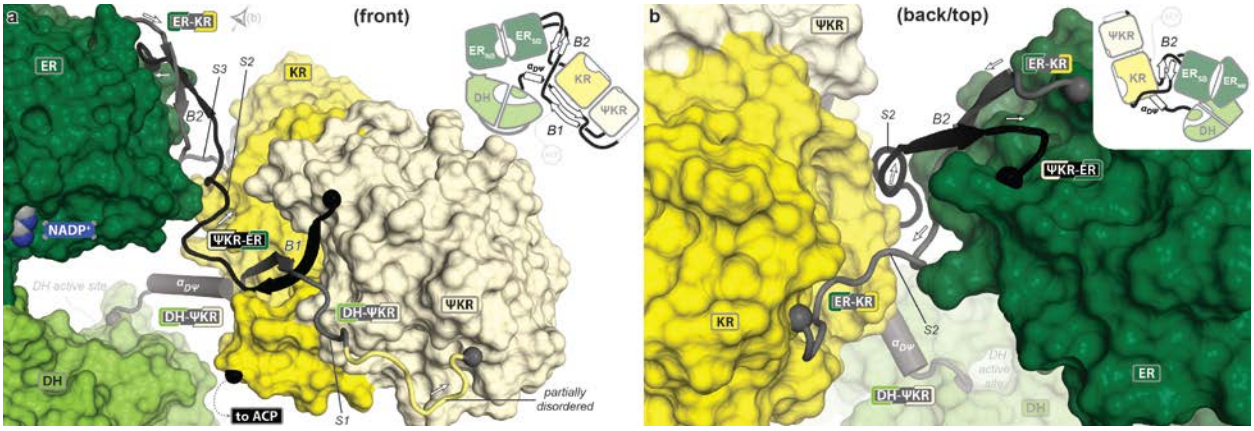


**Figure 3 | Linker-based organization of the MAS modifying region.** The DH and ER lack direct interdomain contacts to the ΨKR/KR domains. The modifying region is established by an interplay of irregular and helical linker segments with two double-stranded antiparallel linker β-sheets (B1, B2), which interact with the ΨKR/KR and ER domain, respectively. **a**, The DH-ΨKR linker (aa 1177-1214) provides helix $\alpha_{D\Psi}$ as a spacer between DH and β-sheet B1 on the surface of the ΨKR/KR. The DH-ΨKR linker continues into segment S1 and ends in a partially disordered loop (light yellow), which was traced only in one chain. The central ΨKR-ER linker (aa 1392-1418) engages in both β-sheets (B1 and B2); the stretch S2 between sheet B1 and B2 adopts two alternate conformations among different chains. **b**, The ER-KR linker (aa 1744-1764) contains an irregular stretch (S3, aa 1753-1764), which is considerably longer than required to bridge the interdomain gap.

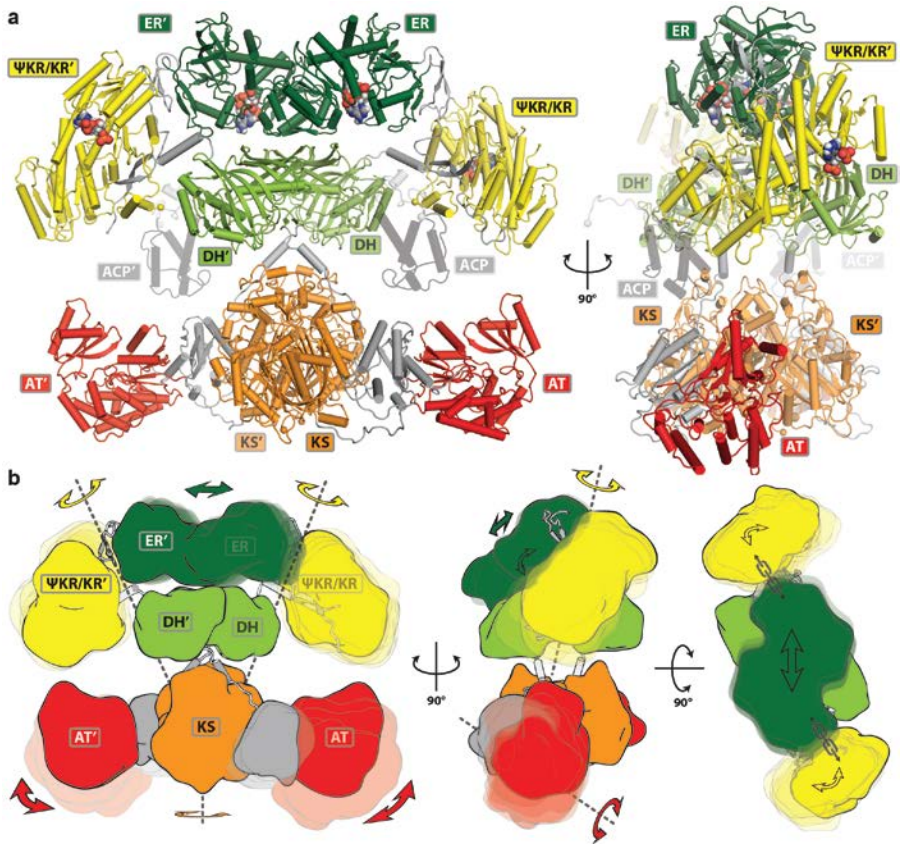



**Figure 4 | Hybrid model of a dynamic MAS dimer. a,** A hybrid MAS model was assembled by linking the condensing and modifying region structures. Central helical linkers connect the two regions without

secondary interactions. The orientation around the linkage is presumably flexibly and has been modeled according to the FAS structure. A homology model of mobile ACP is indicated (transparent grey) in a resting position without domain interactions. **b**, Conformational variability based on a comparison of 18 MAS modifying region chains and five homologous condensing region (Extended Data Fig. 6e, f) structures in combination with normal-mode analysis are shown. The lateral ER dimer motion on the DH platform is coupled to a rotation of both double-tethered ΨKR/KRs by up to 40.4° (Supplementary Video 1). The positions of the AT relative to KS in different condensing region structures correspond to a rotation of up to 43° between the most linear (MAS) and the most bent (human FAS) variant.

## ONLINE METHODS

### Cloning, Expression and Purification

*Mycobacterium smegmatis* (ATCC® 700084) was cultured according to ATCC recommendations. Cells were pelleted and washed with TE buffer. A cell pellet of G*amma proteobacterium HdN1* was provided by Dr. J. Zedelius (Max Planck Institute, Bremen, Germany). Cells were resuspended in lysis buffer (0.1 M Tris pH 8.0, 0.2 M NaCl, 5 mM EDTA, 0.2 mg $ml^{-1}$ lysozyme), incubated for 6 h at 37 °C, subsequently supplemented with 0.5 % SDS and 0.2 mg $ml^{-1}$ proteinase K and incubated at 65 °C for 24 h. DNA was purified by phenol-chloroform extraction and dissolved in TE buffer. The MAS KS-AT constructs (Uniprot: A0R1E8, aa 1-884, 1-887, 1-892), MAS DH (A0R1E8, 884-1186), MAS DH-ΨKR-ER-KR (A0R1E8, 884-2020), "Pks" DH-ΨKR-ER-KR (Q3L885, 2450-3580) were cloned into pNIC28a-Bsa vectors; GpEryA DH-ΨKR-ER-KR (E1VID6, 2420-3575) constructs were cloned into a Gateway® compatible pETG-10A destination vector (provided by EMBL Heidelberg). MAS DH-ΨKR-ER-KR-ACP (A0R1E8, 884-2111) was cloned by codon optimized gene synthesis of ACP (GenScript) and restriction cloning (BsrGI/HindIII) into pNIC28a-Bsa-MAS DH-ΨKR-ER-KR (884-2020). All constructs were designed as N-terminal tobacco etch virus (TEV) protease cleavable hexa-histidine (His6) fusion constructs and co-expressed with *Streptomyces* chaperonins[31] (pETcoco-2A-L1SL2 plasmid) in BL21(DE3) and Rosetta(DE3) pLysS (GpEryA). Cells were cultured in 2xYT media, supplemented with 0.5 % glycerol, NPS (25 mM $(NH4)_2SO_4$, 50 mM $KH_2PO_4$, 50 mM $Na_2HPO_4$), kanamycin (100 µg/ml), chloramphenicol (34 µg/ml), and ampicillin (100 µg/ml). An expression culture (1.5 L) was inoculated (1:20), grown at 37 °C for 2 h, cooled to 20 °C, and induced with isopropyl-β-D-

thiogalactopyranosid (0.1 mM) at an $OD_{600}$ of 1.0. Cells were harvested after 12 h by centrifugation (7,000 x g) and resuspended in lysis buffer (50 mM HEPES pH 7.4, 20 mM imidazole, 0.5 M NaCl, 5 mM $MgCl_2$, 10 % glycerol (v/v), 2.5 mM β-mercaptoethanol), supplemented with protease inhibitors (200 µM phenylmethylsulfonyl fluoride, 20 µM bestatin, 4 µM E64, 2 µM pepstatin A, 20 µM phenantrolin, 2 µM phosphoramidon) as well as DNase, RNase, and lysozyme. Cells were placed on ice and lysed by sonication. The lysate was cleared by centrifugation (100,000 x g, 30 min) and the supernatant was loaded onto a 5 ml Ni-affinity column (GenScript) pre-equilibrated with lysis buffer. Unbound protein was eluted with 4 alternating wash cycles of 5 column volumes (CV) lysis buffer and HisA buffer (50 mM HEPES pH 7.4, 20 mM imidazole, 50 mM NaCl, 5 mM $MgCl_2$, 10 % glycerol (v/v), 2.5 mM β-mercaptoethanol, inhibitors), until a stable baseline (A280) was reached. The sample was eluted with 2 CV HisB buffer (50 mM HEPES pH 7.4, 250 mM imidazole, 50 mM NaCl, 10 % glycerol (v/v), 2.5 mM β-mercapto ethanol, inhibitors) and diluted (1:10) with AIC-A buffer (50 mM Tris-HCl pH 7.4, 20 mM KCl, 10 % (v/v) glycerol, 2.5 mM β-mercaptoethanol). The sample was loaded on a 6.5 ml anion exchange column (PL-SAX 4,000 Å, 10 µm) and washed with 20 CV. The samples were eluted with a stepped gradient to 100 % AIC-B (50 mM Tris-HCl pH 7.4, 1 M NaCl, 10 % (v/v) glycerol, 2.5 mM β-mercaptoethanol). For DH-ΨKR-ER-KR the gradient was held at a conductivity of 15 mS/cm until a stable baseline (A280) was obtained in order to elute *Streptomyces* chaperonins. DH-ΨKR-ER-KR eluted at 17-20 mS/cm. Pure fractions were pooled, supplemented with TEV protease (1 mg protease per 100 mg tagged protein) and incubated for 10 h at 4 °C. Uncleaved protein, as well as the cleaved His6-tag was removed by passing the solution through a 5 ml orthogonal Ni-affinity column (GenScript). The flow-through was pooled, concentrated and subjected to gel permeation chromatography (Superdex 200 16/60, GE

Healthcare) using GPC buffer (20 mM HEPES pH 7.4, 250 mM NaCl, 5 % glycerol (v/v), 5 mM dithiothreitol). Pure fractions were pooled, and monodispersity was monitored by dynamic light scattering at 1 mg $ml^{-1}$. Related purification protocols were applied to MAS DH-ΨKR-ER-KR-ACP, "Pks" DH-ΨKR-ER-KR (HiTrap CaptoQ column), GpEryA DH-ΨKR-ER-KR (both no TEV protease cleavage and orthogonal Ni-affinity column), and MAS DH (no anion exchange chromatography).

**Crystallization**

All crystallization experiments were performed using a robotic setup applying the sitting drop vapor diffusion method.

MAS KS-AT bipyramidal crystals were grown at 4 °C by mixing 0.2 µl of protein in GPC buffer (38 mg $ml^{-1}$) with 0.2 µl reservoir solution (0.1 M MES/imidazole pH 6.5, 0.1 M $MgCl_2$, 0.1 M $CaCl_2$, 12.5 % (v/v) polyethylene glycol 1,000 (v/v), 7.5 % (w/v) polyethylene glycol (PEG) 3,350, 12.5 % 2-methyl-2,4-pentanediol (MPD)). Crystals grew to a final size of 0.8x0.4x0.2 $mm^3$ within one week and were flash frozen in liquid nitrogen.

The MAS DH domain was crystallized in space group $P2_1$ at 18 °C by mixing 0.2 µl of protein in GPC buffer (38 mg/ml) with 0.1 µl reservoir solution (0.1 M bis-Tris pH 6.5, 0.2 M $MgCl_2$, 25 % (v/v) PEG 3,350) and grew to a final size of 0.4x0.2x0.1 $mm^3$ within one week. Crystals in space group $P2_12_12$ appeared after 30 days at 18 °C by mixing 1 µl of protein in GPC buffer (38 mg/ml) with 2 µl reservoir solution (0.25 M di-sodium malonate, 24 % (w/v) PEG 3,350) and grew to a final size of 1x0.4x0.3 $mm^3$. Prior to harvesting all crystals of MAS DH were cryo protected (25 % (v/v) ethylene glycol) and flash frozen in liquid nitrogen.

Needle shaped crystals of MAS DH-ΨKR-ER-KR were obtained by mixing protein solution at 18.4 mg $ml^{-1}$ (GPC buffer, 1.5 mM $NADP^+$) and reservoir solution (0.03 M $MgCl_2$, 0.03 M $CaCl_2$, 20 % ethylene glycol, 10 % PEG 8000, 0.1 M MES/imidazole pH 6.5) at 4 °C. Crystallization was optimized by exchanging PEG 8,000 by PEG 3,350, decreasing the PEG 3,350 concentration to 7-13% (w/v) and by carefully monitored microseeding. Subsequent optimization was performed using automated robotic setup and seeding at 4 °C. Final crystals (1.0x0.3x0.2 $mm^3$) were obtained after mixing 1 µl protein (20.3 mg $ml^{-1}$ in GPC buffer incl. 1.5 mM $NADP^+$) with 1 µl of reservoir solution (5.25 % (w/v) PEG 3,350, 20 % (v/v) ethylene glycol, 0.1 M MES pH 7.0, 52 mM $MgCl_2$, 52 mM $CaCl_2$) and 0.2 µl seed stock. Diffraction properties were optimized by crystal dehydration: Over a period of 4 h crystals were transferred to a dehydration solution (0.05 M MES pH 7.0, 25 % ethylene glycol, 25 % PEG 3,350, 56 mM $MgCl_2$, 56 mM $CaCl_2$, 1.5 mM $NADP^+$) by a step-wise exchange of the drop solution. All crystals were harvested and plunge frozen in liquid nitrogen. Integrity of the protein in final crystals was examined by denaturing polyacrylamide gel electrophoresis.

**Data collection and structure determination**

All data sets were collected at the Swiss Light Source (SLS, Villigen, Switzerland) at a temperature of 100 K. Data sets of DH crystals were collected at beamline X06DA ($P2_1$: $\lambda$= 0.999870 Å, T= 100 K; $P2_12_12$: $\lambda$= 0.97626 Å). All other data sets were collected at beamline X06SA (KS-AT: $\lambda$=0.97940 Å, DH-ΨKR-ER-KR: $\lambda$= 0.97626 Å). Data reduction was performed using XDS[32] and XSCALE[32], datasets were analyzed with phenix.xtriage[33]. All structures were solved with PHASER[34] using molecular replacement.

Crystals of all KS-AT didomain variants of MAS are isomorphic in space group $P4_12_12$. The KS and AT domains of DEBS $KS_5$-$AT_5$[11] were used as molecular replacement templates and initial rebuilding was achieved by BUCCANEER[35]. All three crystal structures were virtually identical except for the identity of the last ordered C-terminal residue. The construct with the most extended C-terminus (1-892) revealed aa 887 as last ordered residue, which is overlapping in sequence with the modifying region. Thus we continued refinement only for crystals of this variant (aa 1-892) with unit cell constants of a= 77.5 Å, b= 77.5 Å, c= 371.2 Å and a solvent content of 56 %. A final model was obtained after iterative cycles of real space model building in COOT[36] and TLS refinement in Phenix[33] and was refined to $R_{work}/R_{free}$ values of 0.21/0.23 at 2.3 Å resolution with excellent geometry (Ramachandran favored/outliers: 97.8 %/0.2 %) (Extended Data Table 1a).

Crystals of the DH domain of MAS belong to space group $P2_1$ (a= 59.7 Å, b= 162.4 Å, c= 66.6 Å, β= 91.4°) and $P2_12_12$ (a= 67.1 Å, b= 162.2 Å, c= 59.5 Å) with a solvent content of 49 % and 51 %, respectively. A molecular replacement search model was based on CurK DH[15]. Initial maps were improved by density modification and NCS averaging with PARROT[37],

followed by automated rebuilding with BUCCANEER[35]. Final models were obtained after iterative cycles of model building in COOT[36], and refinement in BUSTER[38] ($P2_1$) and Phenix[33] ($P2_12_12$), yielding excellent geometry (Ramachandran favored/outliers: $P2_1$=98.2 %/0.0 %; $P2_12_12$=98.2 %/0.2 %) and $R_{work}/R_{free}$ values of 0.18/0.20 ($P2_1$) and 0.15/0.18 ($P2_12_12$) (Extended Data Table 1a).

Crystals of MAS DH-ΨKR-ER-KR in space group P1 (a= 151.4 Å, b= 190.4 Å, c= 270.8 Å, α= 95.6°, β= 91.9°, γ= 103.7°) diffracted to a maximum resolution of 3.75 Å. The asymmetric unit contained 18 protomers in nine dimers with 20,502 amino acids and a molecular mass of 2.2 MDa at 65 % solvent content. Data were collected at four different positions of a single crystal and combined to obtain a complete high-quality dataset. The resolution cutoff was determined by $CC_{1/2}$ criterion[39]. Self-rotation functions revealed non-crystallographic rotational symmetry (NCS) and the native patterson function indicated translational NCS.

Initially, a partial molecular replacement solution was obtained for the ER dimer using the ER domain of porcine FAS (pFAS)[6]. Other known structures of homologous domains did not provide efficient search models. The structure of the isolated MAS DH domain, determined here independently, yielded equivalent solutions in agreement with the pFAS ER based solution. For final structure determination both models where used in subsequent rounds of molecular replacement. Start models for building further regions were generated by homology modelling using Swiss Model[40]. In order to allow unbiased refinement in real and reciprocal space, phenix.reflection_tools[33] was used to define a thin-resolution shell-based test set[41], and test set reflections were excluded from calculating maps, which were used for real-space refinement. Initial refinement cycles included rigid body refinement and restrained refinement. The impact of various low-resolution restraint formulations on refinement were tested carefully. Local NCS is

particularly well-suited for MAS DH-ΨKR-ER-KR intermediate resolution refinement due to the high degree of NCS and the fact that using local NCS restrains avoids any external standard restraints based on assumptions on secondary structure or homologous peptide structures. Thus local structural similarity restraints (LSSR)[42] were combined only with reference model restraints to the authentic DH domain structure using autopruning in BUSTER[38]. After every round of refinement, bias-reduced, solvent flattened and NCS-averaged maps were calculated using DM[43] without applying phase combination. Sharpened NCS-average maps were generated by applying a sharpening B-factor to the structure factor amplitudes prior to averaging. Initially, real-space rigid body fitting of individual secondary structure elements was applied for instances of every domain type (DH, ΨKR, $ER_{NB}$, $ER_{SB}$, KR) followed by symmetry expansion and rigid body fitting for entire domains. Best defined regions of the electron density maps were used for rebuilding of every domain type using Coot[36] and O[44], respectively, symmetry expanded, and recombined into 18 chains. At this point, unambiguous difference electron density indicated the connecting linkers, which were manually built into the maps and refined without symmetry expansion (Extended Data Fig. 2d, e). Later refinement cycles included TLS refinement, using one group per domain and linker, individual B-factor refinement and automated weight factor determination. During rebuilding B-sharpening, NCS average and density modification as well as feature enhanced maps[33] were used. Overall, the use of 18-fold-domain-wise NCS averaging results in highly accurate and unbiased phase determination irrespective of details of the atomic model. The combined use of NCS-averaging and B-factor sharpening led to an exceptional map quality typical for maps at considerable higher resolution; (Extended Data Fig. 2f). Bound $NADP^+$ cofactors were added for final refinement cycles. $NADP^+$ is well ordered in the ER domain, while the nicotinamide moieties are disordered in the KR domains and were not

included in the final model. A total of five KR and four ΨKR domains, which lack stabilization by crystal contacts, were either disordered or present in multiple orientations, and not included in the final model, despite significant positive difference density. The KR domain in chain L shows a significantly more tilted orientation as observed in all other instances of the KR domains, which however agrees with the identified hinge regions. A single model was placed for this domain, which achieved the largest improvement of R-factors and was characterized by the lowest B-factors after refinement, although a secondary alternate conformation might be present. The refinement of the final model (excluding disordered regions (chains): 883-895 (E-R), 1206-1213, 1283-1287, 1948-1960, ΨKR(I/L/O/Q-R) , KR(I/O/Q-R)) was completed with $R_{work}/R_{free}$ values of 0.23/0.24 and very good geometry for the resolution range (Ramachandran favored/outliers: 91.6 % / 1.8 %).

**Analytical Ultracentrifugation**

To determine oligomeric states in solution, sedimentation equilibrium analytical ultracentrifugation experiments were performed for MAS DH-ΨKR-ER-KR and MAS KS-AT. 140 µl columns containing proteins at concentrations of 3.5-4.5 mg $ml^{-1}$ in GPC buffer were subjected to centrifugation at 4,800 and 7,800 rpm at 12 °C, with detection by radial absorbance scanning at 305 nm. At each speed, centrifugation was allowed to proceed until sedimentation equilibrium was attained, as judged by pairwise comparison of scans using the approach to equilibrium function in SEDFIT (https://sedfitsedphat.nibib.nih.gov). Buffer density (1.0277 g/ml) and viscosity (1.5306 cP) were measured at 12 °C using an Anton Paar DMA4500M densitometer and an AMVn viscometer, respectively. Molar extinction coefficients at 305 nm were calculated for each protein from the ratio of observed absorbance at various wavelengths in spectra at different dilutions and calculated molar extinction coefficients. The

partial specific volume for each protein was calculated from sequence in SEDFIT. The radial absorbance scans at equilibrium for the two speeds were globally fitted to the "single species of interacting system" mode in SEDPHAT[45] to determine the apparent molecular mass of the protein in solution. If the obtained molecular mass was intermediate between the value expected for a monomer and a dimer, the data were globally fitted to the monomer-dimer association model in SEDPHAT, with the molecular mass of the monomer fixed to the value calculated from the sequence. In both cases data were fitted using a fixed meniscus position, a floating bottom position, mass conservation constraints, a floating baseline and fitting radially-independent noise components. Confidence intervals on single-species masses or dissociation constants were obtained by the Monte-Carlo method implemented in SEDPHAT.

**Small angle X-ray scattering**

Small angle X-ray scattering (SAXS) data were collected at the beamline X12SA of SLS. Samples were dialyzed into GPC buffer, diluted to concentrations between 3-10 mg $ml^{-1}$ and centrifuged at 13,000 x g and 8 °C until measurement. Glass capillaries (1 mm inner diameter) were mounted on a temperature-controlled holder at 12 °C. Data collection was performed using a Pilatus 2M detector at a distance of 2.14 m and a wavelength of 1.000 Å. Data were collected in eight repetitive scans each including ten 40 ms acquisitions at ten capillary positions yielding a total of 800 frames per buffer and protein, respectively. Frames with artefacts *e.g.* from air bubbles, were identified using SLS/PSI software (SAXS_inspect2) and excluded from the data sets. Radial averages were calculated and exported using beamline software for scattering vectors from 0.005 to 0.7 $Å^{-1}$ defined as $q = 4\pi/\lambda sin\theta$. Scattering curves were averaged using DATAVER[46]; buffer profiles were subtracted using DATOP[46]. Scaling factors and p-values of a Students-T test were analyzed using DATMERGE[46] and DATCMP[46], respectively. Later frames

were affected by increasing radiation damage and were excluded from further processing. Final scattering curves for each sample concentration were thus obtained from 300 individual profiles. The radius of gyration ($R_g$) and zero angle intensity (I(0)) was calculated from the Guinier approximation using AUTORG[46] and is consistent with values obtained from atomic distance distributions p(r) using DATGNOM[46] (Extended Data Table 1b). Scattering profiles at different concentrations were only combined if a noise reduction at medium and high scattering vectors could be obtained.

Modifying regions bear an intrinsic flexibility, which requires a flexible fitting approach in order to sample the full conformational space of the structures. Some approaches for flexible SAXS fitting have been described[47,48], but none was able to refine an individual structure while maintaining two-fold symmetry. Therefore, we combined dynamic elastic network restraints from CNS[49] with SAXS-target refinement and twofold symmetry averaging in XPLOR-NIH[50] for the refinement of individual structures by simulated annealing. SAXS scattering curves of atomic models, fits with experimental data, and distance distributions were calculated using CRYSOL[46] and DATGNOM[46]. All SAXS curves were plotted using Python Matplotlib.

For comparing calculated and experimental SAXS scattering curves, three models for the architecture of modifying regions were generated based on the crystal structure of the domain-swapped SpnB fragment (ER-KR/ΨKR). The first model was obtained according to the original publication[19] by superposing the monomeric ER-KR/ΨKR domain on the KR domain of pFAS[6]. A linear homology model of SpnB DH[40] was placed into the position of pFAS DH and the domain swap in SpnB ER-KR/ΨKR was replaced with the corresponding region from DEBS $KR_1$[51]. The second model was constructed in the same way via a superposition on MAS KR. The relative domain arrangement of SpnB ER-KR/ΨKR was not altered in these two models, only the

domain swap was corrected. The third more generalized modPKS model was constructed in order to verify if shorter ER-KR linkers are in contradiction with the architecture of MAS. As a representative for short ER-KR linkers, the structure of SpnB ER-KR/ΨKR (6 aa) and the corresponding DH homology model were modeled as individual domains on MAS, while the linear DH dimer was maintained. ΨKR-ER linkers could be readily reconnected and regularized, whereas the ER-KR linker required a tilt of the ΨKR/KR domain. The tilt maintained a reasonable distance between the C-terminus of the DH and N-terminus of the ΨKR domain and yielded a linker architecture of a modPKS in agreement with MAS without stable direct interdomain contacts. SAXS curves and distance distributions were calculated of all models and compared to experimental SAXS scattering curves of MAS and two modPKS modifying regions with short ER-KR linkers (GpEryA: 9 aa; MsPks: 8 aa).

**Structure analysis and visualization**

Related structures were identified using PDBeFold[52] and interfaces were analyzed using QtPISA[53]. Transformations and coordinate manipulations were carried out using CCP4[54] tools, MODTRAFO (T. Schirmer, Biozentrum Basel, http://www.biozentrum.unibas.ch) and MOLEMAN[55]. The automated Oligo algorithm[56] as implemented in Swiss Model unambiguously detected and predicted a single mode of dimerization of MAS KS-AT based on sequence homology. Initially, the dimeric form of KS-AT was assembled by least squares fitting of secondary structure elements on DEBS $KS_5$[11]. Then, all residues in a radius of 7.5 Å to the dimer interface were deleted and multi-template homology modelling using modeller 9.15[57] was used to construct a full-length dimeric homology model based on 20 homodimeric PKSs/FASs KS structures and the interface deleted MAS KS-AT structure. Remodeled regions (excluding all crystallographically defined regions beyond the radial cutoff) were geometry minimized using

phenix.geometry_minimization[33]. The position where the post-AT linker becomes disordered was located by crystallization of KS-AT didomains with three different linker lengths (1-884, 1-887, 1-892). Normal mode analyses was carried out using the Bio3D[58] library in "R". Hinge bending analysis was carried out by pre-aligning all structures to a reference substructure using LSQKAB[59], followed by a MODTRAFO (T. Schirmer, Biozentrum Basel, http://www.biozentrum.unibas.ch) analysis of the moving substructure. Principle screw axes were determined by averaging the direction vectors of the screw axes using Python Numpy and locating a central hinge point from the position of all screw axes. Active site distances were calculated using BIOPYTHON[60]. All axes were visualized using PYMOL[61]. Interdomain angles of DH dimers were calculated by pre-aligning all DH dimers to one DH domain of MAS DH, followed by calculating the angle between the first principle component vector of the secondary structure elements of both domains. The angles were visualized using PYMOL[61]. Bias-removal for $F_{obs}$-$F_{calc}$ omit maps was achieved by applying a random perturbation to coordinates (Δ0.2 Å) and B-factors (Δ20 % of the mean overall B-factor) using MOLEMAN2[55] prior to refinement. Figures, movies and active site tunnels were generated using PYMOL[61], LSQMAN[62], and CAVER 3.0[63].

**Sequence analysis**

55 sequences containing fully reducing modifying regions were selected from FASs, fiPKSs, Msl-, one *trans*-AT and 36 modPKSs modules. Structure-based sequence alignments of all PKSs/FASs type I domain structures were generated using PDBefold[52] and used as reference for the alignment of individual domains using ClustalW2[64]. Linkers were aligned without reference, assembled with the individual domain alignments and manually corrected in Geneious v7.1.7[65]. Phylogenetic trees were generated using the neighboring joining algorithm in Geneious v7.1.7[65].

## References


6 Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315-1322 (2008).

11 Tang, Y., Kim, C. Y., Mathews, II, Cane, D. E. & Khosla, C. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11124-11129 (2006).

12 Whicher, J. R. *et al.* Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem. Biol.* **20**, 1340-1351 (2013).

15 Akey, D. L. *et al.* Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure* **18**, 94-105 (2010).

16 Keatinge-Clay, A. Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* **384**, 941-953 (2008).

17 Gay, D., You, Y. O., Keatinge-Clay, A. & Cane, D. E. Structure and Stereospecificity of the Dehydratase Domain from the Terminal Module of the Rifamycin Polyketide Synthase. *Biochemistry* (2013).

18 Khare, D. *et al.* Structural Basis for Cyclopropanation by a Unique Enoyl-Acyl Carrier Protein Reductase. *Structure* (2015).

19 Zheng, J., Gay, D. C., Demeler, B., White, M. A. & Keatinge-Clay, A. T. Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol.* **8**, 615-621 (2012).

21 Keatinge-Clay, A. T. A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.* **14**, 898-908 (2007).

24 Bretschneider, T. *et al.* Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature* **502**, 124-128 (2013).

31 Betancor, L., Fernandez, M. J., Weissman, K. J. & Leadlay, P. F. Improved catalytic activity of a purified multienzyme from a modular polyketide synthase after coexpression with Streptomyces chaperonins in Escherichia coli. *Chembiochem* **9**, 2962-2966 (2008).

32 Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125-132 (2010).

33 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221 (2010).

34 McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658-674 (2007).

35 Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002-1011 (2006).

36 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132 (2004).

37 Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 470-478 (2010).

38 Bricogne G., B. E., Brandl M., Flensburg C., Keller P., Paciorek W., & Roversi P, S. A., Smart O.S., Vonrhein C., Womack T.O. BUSTER version 2.10.2. *Cambridge, United Kingdom: Global Phasing Ltd.* (2011).

39 Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030-1033 (2012).

40 Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381-3385 (2003).

41 Fabiola, F., Korostelev, A. & Chapman, M. S. Bias in cross-validated free R factors: mitigation of the effects of non-crystallographic symmetry. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 227-238 (2006).

42 Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 368-380 (2012).
43 Cowtan, K. An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34-38 (1994).
44 Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved Methods for Building Protein Models in Electron-Density Maps and the Location of Errors in These Models. *Acta Crystallogr. A* **47**, 110-119 (1991).
45 Vistica, J. *et al.* Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Anal. Biochem.* **326**, 234-256 (2004).
46 Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Cryst.* **45**, 342-350 (2012).
47 Pelikan, M., Hura, G. L. & Hammel, M. Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen. Physiol. Biophys.* **28**, 174-189 (2009).
48 Zheng, W. & Tekpinar, M. Accurate flexible fitting of high-resolution protein structures to small-angle x-ray scattering data using a coarse-grained model with implicit hydration shell. *Biophys. J.* **101**, 2981-2991 (2011).
49 Brunger, A. T. *et al.* Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905-921 (1998).
50 Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65-73 (2003).
51 Keatinge-Clay, A. T. & Stroud, R. M. The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases. *Structure* **14**, 737-748 (2006).
52 Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2256-2268 (2004).
53 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774-797 (2007).
54 Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760-763 (1994).
55 Kleywegt, G. J. Validation of protein models from Calpha coordinates alone. *J. Mol. Biol.* **273**, 371-376 (1997).
56 Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252-258 (2014).
57 Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815 (1993).
58 Skjaerven, L., Yao, X. Q., Scarabelli, G. & Grant, B. J. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics* **15**, 399 (2014).
59 Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **32**, 922-923 (1976).
60 Cock, P. J. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423 (2009).
61 Schrodinger, L. *The PyMOL Molecular Graphics System, Version 1.7.0.3* (2010).
62 Kleywegt, G. Use of Non-crystallographic Symmetry in Protein Structure Refinement. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 842-857 (1996).
63 Medek, P. B., P.; Sochor, J. Computation of tunnels in protein molecules using Delaunay triangulation. *Journal of WSCG, University of West Bohemia, Pilsen* **15(1-3)**, 107-114 (2007).

64 Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).
65 Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649 (2012).
66 Pappenberger, G. *et al.* Structure of the human fatty acid synthase KS-MAT didomain as a framework for inhibitor design. *J. Mol. Biol.* **397**, 508-519 (2010).
67 Tang, Y., Chen, A. Y., Kim, C. Y., Cane, D. E. & Khosla, C. Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem. Biol.* **14**, 931-943 (2007).

**a**

| | KS-AT 1-892 | DH 884-1189 | DH 884-1189 | DH-ΨKR-ER-KR 884-2020 |
|---|---|---|---|---|
| **Data collection** | | | | |
| Space group | P $4_12_12$ | $P2_1$ | $P2_12_12$ | $P_1$ |
| Cell dimensions | | | | |
| a, b, c (Å) | 77.53, 77.53, 371.22 | 59.65, 162.40, 66.62 | 67.06, 162.20, 59.49 | 151.38, 190.37, 270.84 |
| $\alpha, \beta, \gamma$ (°) | 90.0, 90.0, 90.0 | 90.0, 91.4, 90.0 | 90.0, 90.0, 90.0 | 95.6, 91.9, 103.7 |
| Resolution (Å) | 92.81 - 2.20 | 66.65 – 1.75 | 47.99 - 1.45 | 78.62 – 3.75 |
| $R_{merge}$ (%)* | 8.7 (135.8) | 3.6 (123.2) | 4.3 (134.5) | 25.0 (315.9) |
| $I/\sigma I$* | 17.98 (2.26) | 12.32 (1.13) | 19.07 (1.36) | 8.95 (1.00) |
| $CC_{1/2}$* | 99.9 (67.7) | 99.9 (46.2) | 100.0 (76.6) | 99.6 (44.9) |
| Completeness (%)* | 99.6 (99.2) | 96.8 (91.6) | 99.3 (97.0) | 99.0 (98.9) |
| Redundancy* | 12.9 (13.2) | 3.9 (3.8) | 6.5 (6.5) | 9.0 (9.3) |
| Unique reflections* | 58,436 (9,008) | 123,118 (8,582) | 114,826 (18,878) | 296,164 (21,835) |
| **Refinement** | | | | |
| Protomers | 1 | 4 | 2 | 18 |
| Resolution (Å) | 54.82 - 2.20 | 66.6 – 1.75 | 47.97 - 1.45 | 78.62 – 3.75 |
| $R_{work}$/ $R_{free}$ | 0.21 / 0.23 | 0.18 / 0.20 | 0.15 / 0.18 | 0.23 / 0.24 |
| No. atoms | 12,279 | 18,419 | 9,282 | 262,498 |
| Protein | 11,984 | 17,273 | 8,725 | 260,724 |
| Ligand/ion | -- | 408 | 60 | 1774 |
| Water | 295 | 738 | 497 | -- |
| B-factors | 81.91 | 57.63 | 35.29 | 171.34 |
| Protein ($Å^2$) | 82.38 | 56.90 | 34.55 | 171.44 |
| Ligand/ion ($Å^2$) | -- | 76.57 | 52.12 | 157.33 |
| Water ($Å^2$) | 63.04 | 64.27 | 46.27 | -- |
| R.m.s deviations | | | | |
| Bond lengths (Å) | 0.003 | 0.010 | 0.011 | 0.008 |
| Bond angles (°) | 0.73 | 1.10 | 1.31 | 0.97 |

**b**

| | MsMAS DH-ΨKR-ER-KR | MsMAS DH-ΨKR-ER-KR-ACP | MsPks DH-ΨKR-ER-KR | GpEryA DH-ΨKR-ER-KR |
|---|---|---|---|---|
| Beamline | X12SA (SLS) | X12SA (SLS) | X12SA (SLS) | X12SA (SLS) |
| Wavelength (Å) | 1.00000 | 1.00000 | 1.00000 | 1.00000 |
| Detector distance (m) | 2.14 | 2.14 | 2.14 | 2.14 |
| q range ($Å^{-1}$) | 0.005-0.707 | 0.005-0.707 | 0.005-0.707 | 0.005-0.707 |
| Capillary diameter (mm) | 0.1 | 0.1 | 0.1 | 0.1 |
| Scan lengths / step size (mm) | 4.5 / 0.5 | 4.5 / 0.5 | 4.5 / 0.5 | 4.5 / 0.5 |
| Positions / acquisitions | 10 / 10 | 10 / 10 | 10 / 10 | 10 / 10 |
| Scan repeats | 8 | 8 | 8 | 8 |
| Exposure time (sec) | 0.04 | 0.04 | 0.04 | 0.04 |
| Concentration (mg $ml^{-1}$) | 3, 6 | 5, 10 | 6 | 3, 6 |
| Temperature (K) | 285 | 285 | 285 | 285 |
| I(0) ($Å^{-1}$) [from P(r)] | 3.00 ± 0.00 | 2.26 ± 0.00 | 2.08 ±0.00 | 2.00 ± 0.00 |
| Rg (Å) [from P(r)] | 58.25 ± 0.09 | 66.90 ± 0.19 | 55.05 ± 0.09 | 56.54 ± 0.09 |
| I(0) ($Å^{-1}$) [from Guinier] | 3.01 ± 0.01 | 2.19 ± 0.01 | 2.1 ± 0.01 | 1.98 ± 0.01 |
| Rg (Å) [from Guinier] | 57.2 ± 0.30 | 62.23 ± 0.63 | 54.7 ± 0.29 | 54.9 ± 0.23 |
| Dmax (Å) | 201 | 250 | 191 | 192 |
| Model fit ($\chi$) | 1.79 | N/A | N/A | N/A |

**Extended Data Table 1 | X-ray data collection and processing table. a,** Crystallographic data collection and refinement statistics. The resolution cutoff was determined by $CC_{1/2}$ criterion (Karplus and

Diederichs, 2012). *, Highest resolution shell is shown in parenthesis. **b**, Small angle X-ray scattering data collection and processing.

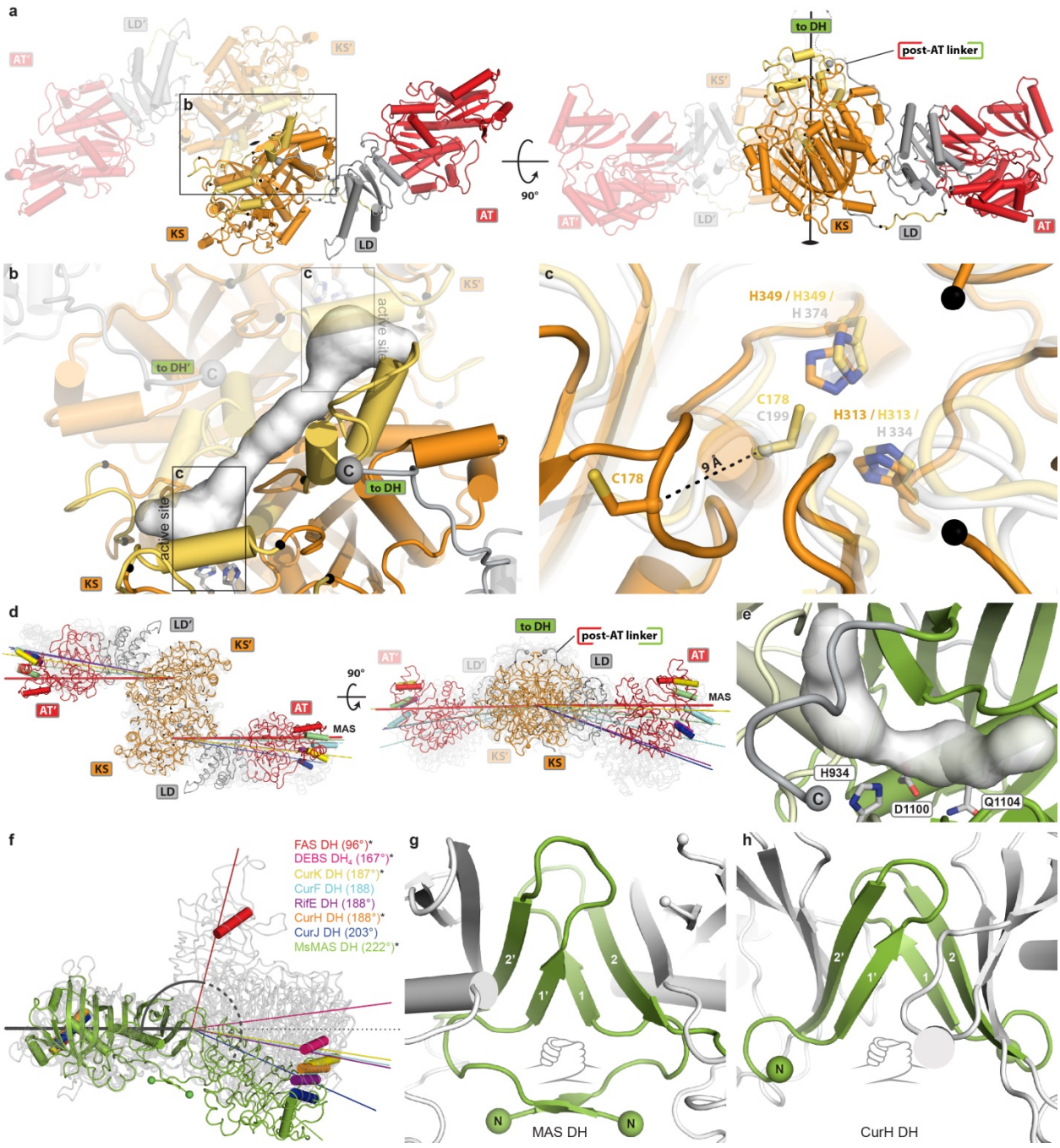
**a**

| Structure 1 | Structure 2 | $C_\alpha$ r.m.s.d. [Å] | Aligned residues |
|---|---|---|---|
| MAS KS | DEBS $KS_5$ | 1.17 | 348 |
| MAS KS | DEBS $KS_3$ | 1.21 | 349 |
| MAS AT | DEBS $KS_5$ | 1.77 | 294 |
| MAS AT | DEBS $KS_3$ | 1.21 | 287 |
| MAS DH | Rif $DH_{10}$* | 1.86 | 249 |
| MAS DH | CurF DH* | 1.61 | 259 |
| MAS DH | CurJ DH* | 1.88 | 252 |
| MAS DH | CurK DH | 1.68 | 257 |
| MAS DH | CurH DH | 1.93 | 258 |
| MAS DH | DEBS $DH_4$ | 1.82 | 244 |
| MAS DH | MAS DH $P2_12_12$ / $P2_1$ | 0.70 | 284 |
| MAS DH $P2_1$ | MAS DH $P2_12_12$ | 0.40 | 284 |
| MAS ER | SpnB ER | 1.98 | 298 |
| MAS $ER_{NB}$ | PpsC $ER_{NB}$† | 1.13 | 170 |
| MAS ER | JamJ ER* | 1.63 | 308 |
| MAS ER | CurF ER* | 1.72 | 309 |
| MAS ER | CurK ER | 2.27 | 306 |
| MAS ER | mFAS ER | 2.15 | 299 |
| MAS ΨKR/KR | Tyl ΨKR/$KR_{10}$* | 1.83 | 380 |

**b**

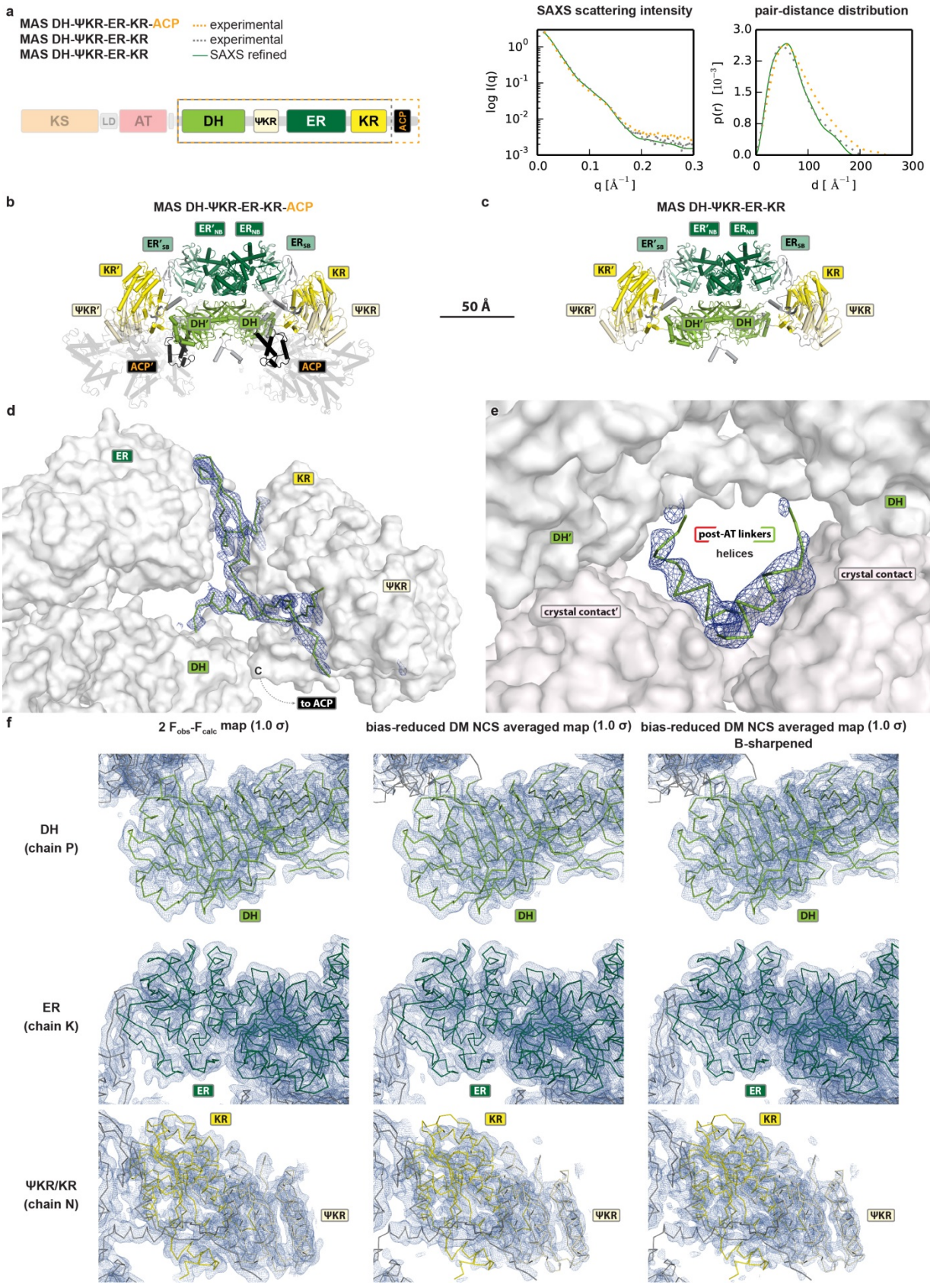
| Interface 1 | Interface 2 | Area [Å$^2$] | SDV [Å$^2$] | Min [Å$^2$] | Max [Å$^2$] |
|---|---|---|---|---|---|
| KS | KS‡ | 1,759 | | | |
| | post-AT linker | 770 | | | |
| LD | KS | 521 | | | |
| | AT | 236 | | | |
| DH ($P2_1$) | DH ($P2_1$) | 979 | 59 | 938 | 1,021 |
| DH ($P2_12_12$) | DH ($P2_12_12$) | 999 | | | |
| DH ($P_1$) | DH ($P_1$) | 961 | 13 | 943 | 978 |
| ER | ER | 1,424 | 20 | 1,397 | 1,449 |
| DH-ΨKR-ER-KR | DH-ΨKR-ER-KR | 2,688 | 112 | 2,547 | 2,904 |
| DH[d] | ER[d] | 532 | 105 | 345 | 638 |
| DH-ΨKR linker | ΨKR/KR | 1,117 | 24 | 1,077 | 1,172 |
| | ΨKR-ER linker | 695 | 9 | 672 | 711 |
| ΨKR-ER linker | ΨKR/KR | 975 | 28 | 926 | 1,027 |
| | ER-KR linker | 471 | 24 | 426 | 517 |
| | ER | 353 | 20 | 330 | 411 |
| ER-KR linker | ER | 432 | 24 | 399 | 480 |
| | KR | 547 | 14 | 527 | 574 |
| ΨKR-ER / ER-KR linkers | KR | 1,111 | 22 | 1,076 | 1,162 |
| | ΨKR/KR | 1,488 | 25 | 1,449 | 1,544 |
| | ER | 683 | 30 | 637 | 735 |

**Extended Data Table 2 | Structural comparison and interface analysis. a,** Cα r.m.s. deviations obtained for structural comparison of MAS domains with their closest structural neighbors. *, not part of a fully-reductive modifying region, †, PDB entry 1pqw (unpublished). **b**, Interfaces in the crystal structures of MAS variants. Standard deviations (SDV) and minima/maxima are given for structures containing more than one interface.‡, by direct superposition of the monomeric KS-AT structure on the DEBS KS5 dimer. In the KS-AT dimer with restored interface (by homology modelling), the total area increases to 2.289 Å$^2$. d, dimer.



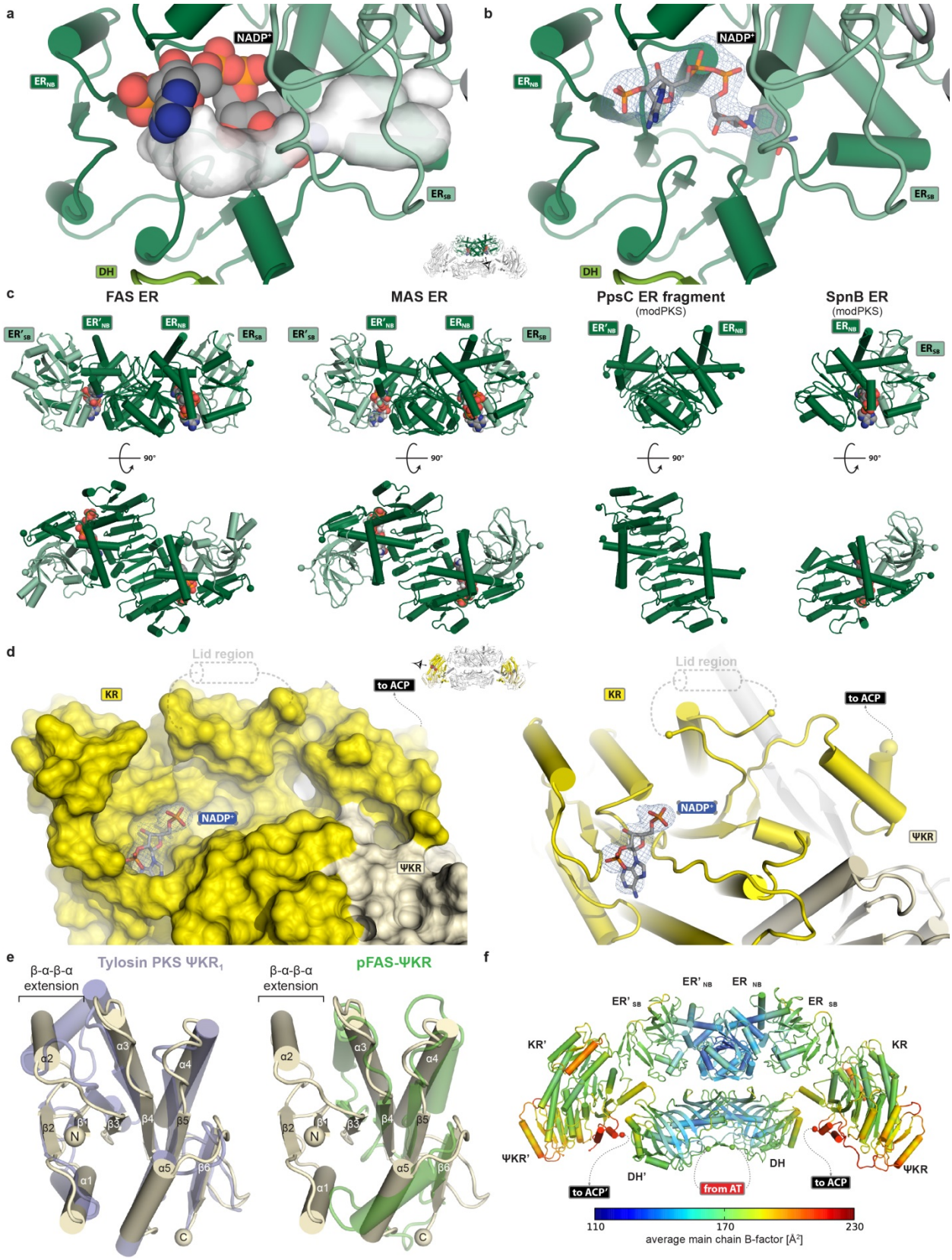

**Extended Data Figure 1 | Reconstruction of the dimeric KS-AT didomain and DH dimer organization. a**, The condensing region dimer was reconstructed by least square fitting on DEBS $KS_5$[11] and multi-template homology modeling of disordered segments and the active site loop (gold). Termini of the remodelled segments are indicated by black spheres. A pseudo-continuous β-sheet is formed across the dimer interface. The post-AT linker terminates close to the dimer axis. **b**, Close-up view on the

reconstructed KS dimer with an active site tunnel spanning both protomers (white), which is enclosed by four remodelled segments (gold). **c,** The active site loop containing the catalytic Cys178 is dislocated in the monomeric (orange) form of MAS KS-AT, whereas the active site His313 and His349 occupy the same position as in the dimeric DEBS $KS_5$-$AT_5$ structure (white-transparent). The canonical conformation of Cys178 observed in dimeric KS domains is restored in the dimeric KS-AT model (gold-transparent) **d**, MAS KS-AT (colored, red line) reveals the most linear overall structure (right panel) of all PKSs/FAS condensing region structures[6,11,12,66,67] (corresponding to Extended Data 6e, f). **e**, The DH active site residues are located at the interface of the two hot-dog folds (light and dark green; active site tunnel in white). **f**, Interdomain angles in DH dimers[6,15-17]. Dimers were superposed onto one protomer (left) of MAS, and the angles between two protomers are compared. For clarity, only MAS DH is shown in green, for other DH domains only one equivalent helix is highlighted in color. The FAS pseudo-dimeric DH domains (red helix) adopt a "V"-shaped structure (interdomain angle: 96°), while PKS DH dimers (various colors) are almost linear (167° - 203°). The MAS DH dimer (green) is bent to the opposite direction relative to FAS, and exhibits the largest interdomain angle (222°) (asterisks indicates DHs that are part of fully-reducing modifying regions). **g**, Dimer interface of MAS DH, and **h,** dimer interface of the isolated DH of the CurH[15] modPKS. Dimerization of MAS and CurH DH are mediated by "handshake" interactions of the N-terminal hot-dog folds. In MAS DH, an N-terminal β-strand extension further contributes to dimerization.


a
MAS DH-ΨKR-ER-KR-ACP
MAS DH-ΨKR-ER-KR
MAS DH-ΨKR-ER-KR
experimental
experimental
SAXS refined
KS
LD
AT
DH
ΨKR
ER
KR
ACP
SAXS scattering intensity
pair-distance distribution
log I(q)
q [Å$^{-1}$]
p(r) [$10^{-3}$]
d [Å$^{-1}$]
b
MAS DH-ΨKR-ER-KR-ACP
c
MAS DH-ΨKR-ER-KR
50 Å
d
e
post-AT linkers
helices
crystal contact
crystal contact'
to ACP
f
2 $F_{obs}$-$F_{calc}$ map (1.0 σ)
bias-reduced DM NCS averaged map (1.0 σ)
bias-reduced DM NCS averaged map (1.0 σ) B-sharpened
DH (chain P)
ER (chain K)
ΨKR/KR (chain N)

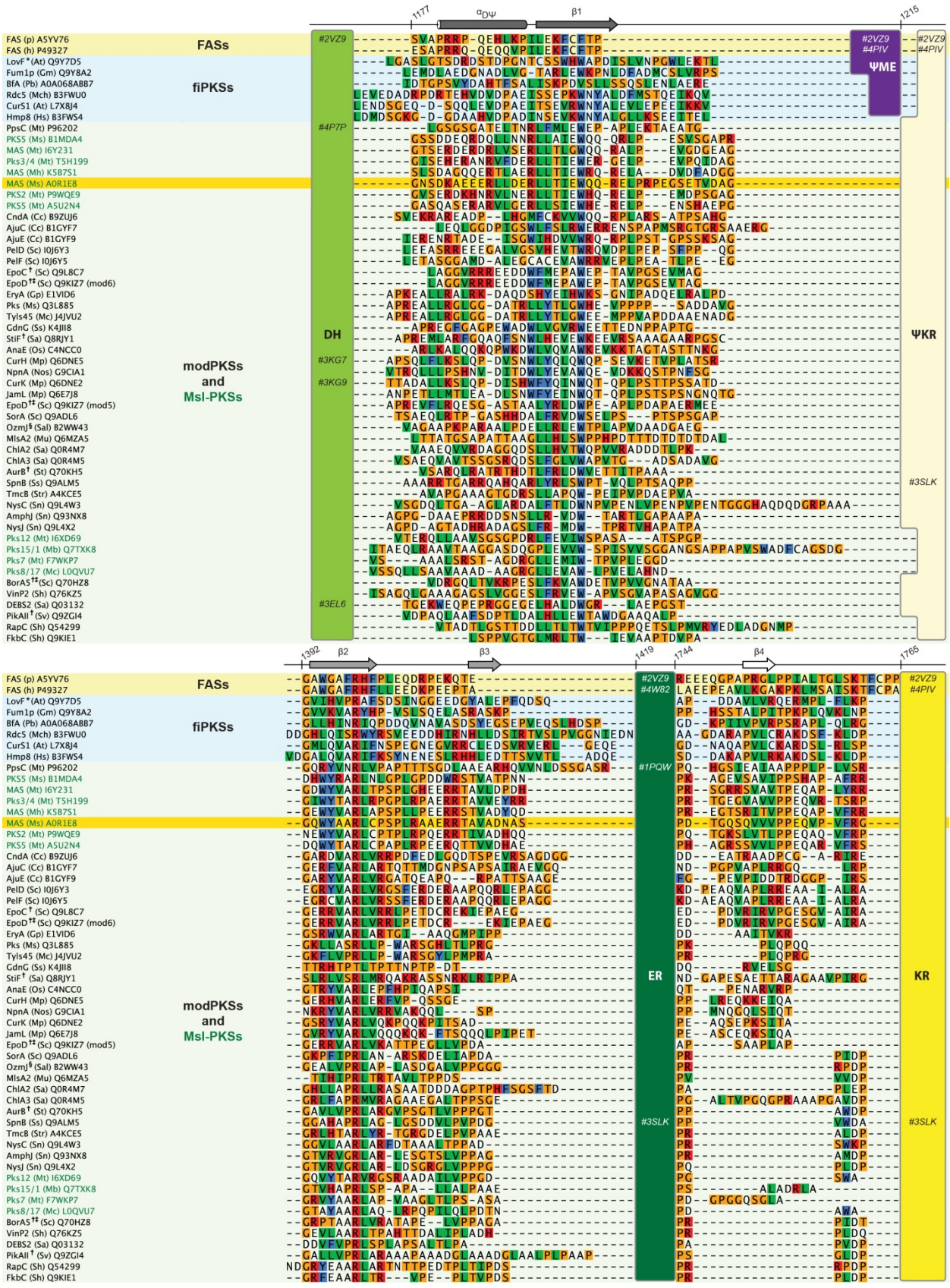
1

**Extended Data Figure 2 | Effect of ACP deletion and electron density maps of the MAS modifying region crystal structure. a**, SAXS experiments reveal conserved scattering profiles for the modifying region with ACP (dotted orange) and without ACP (dotted green), which resemble the scattering curve of the SAXS-refined X-ray structure (green). **b**, **c**, The experimentally determined interatomic distance distributions are in agreement with the maximum extends of the modifying domain with (**b**) and without (**c**) ACP, 250 Å and 201 Å, respectively. In **b** a set of plausible ACP positions is shown (transparent), based on the length of the KR-ACP linker. **d,** Unbiased $F_{obs}$-$F_{calc}$ omit difference map of the modifying region linkers in chain B (contoured at 2.5 σ) is shown. **e**, Unbiased $F_{obs}$-$F_{calc}$ omit difference map of the post-AT linker helices in chain A and B (contoured at 2.5 σ); The helices could be modeled due to stabilizing crystal contacts. **f,** Electron density maps covering the three different domain types as indicated (left: 2 $F_{obs}$-$F_{calc}$ at 1.0 σ, middle: bias-reduced density modified NCS average map at 1.0 σ, right: bias-reduced density modified NCS average map at 1.0 σ, with additional details revealed by applying a B-sharpening factor of -80 $Å^2$).

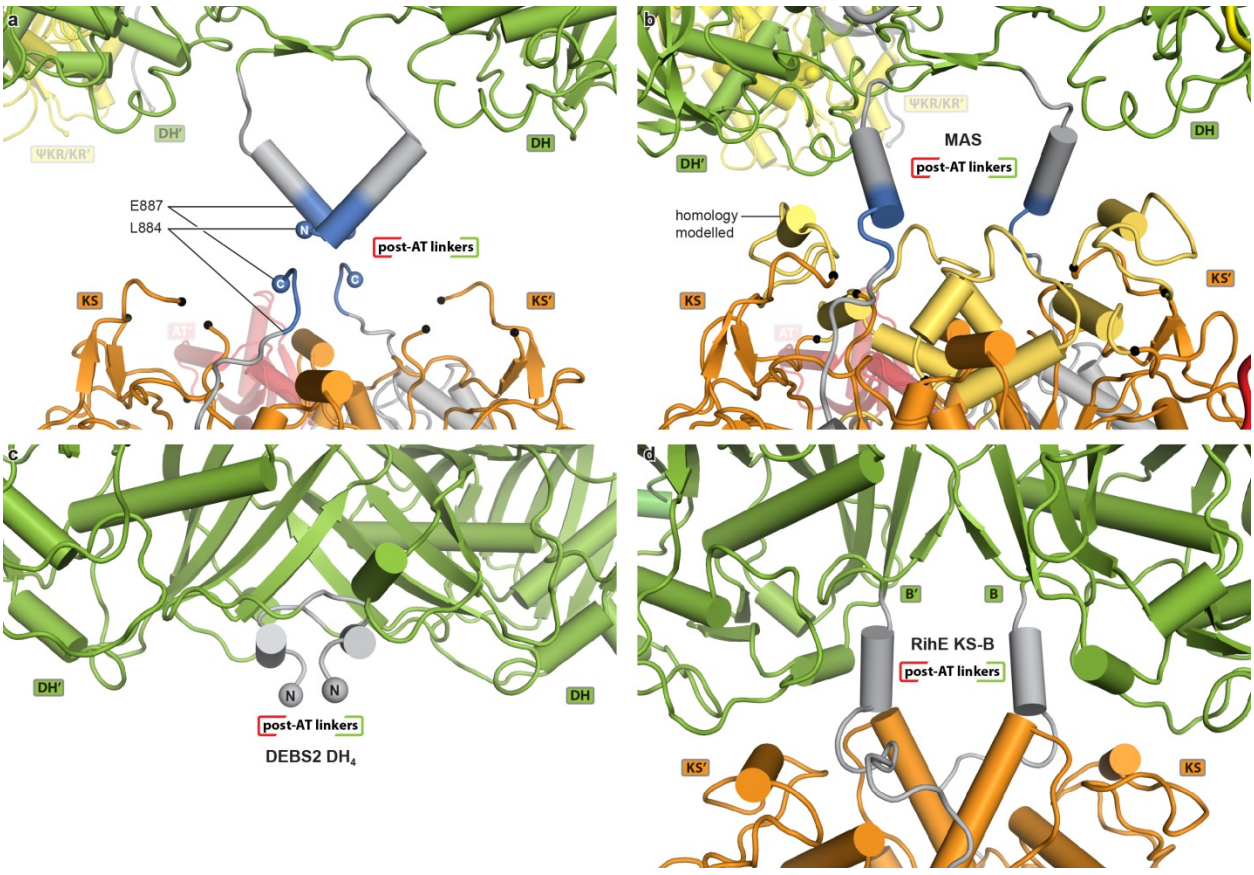

a
NADP+
ER_NB
ER_SB
DH
b
NADP+
ER_NB
ER_SB
DH
c
FAS ER
MAS ER
PpsC ER fragment
(modPKS)
SpnB ER
(modPKS)
ER'_SB
ER'_NB
ER_NB
ER_SB
90°
d
Lid region
KR
to ACP
NADP+
ΨKR
e
Tylosin PKS ΨKR1
pFAS-ΨKR
β-α-β-α extension
α1
α2
α3
α4
α5
β1
β2
β3
β4
β5
β6
N
C
f
ER'_NB
ER_NB
ER'_SB
ER_SB
KR'
KR
ΨKR'
ΨKR
DH'
DH
to ACP'
from AT
to ACP
110
170
230
average main chain B-factor [Å²]

1

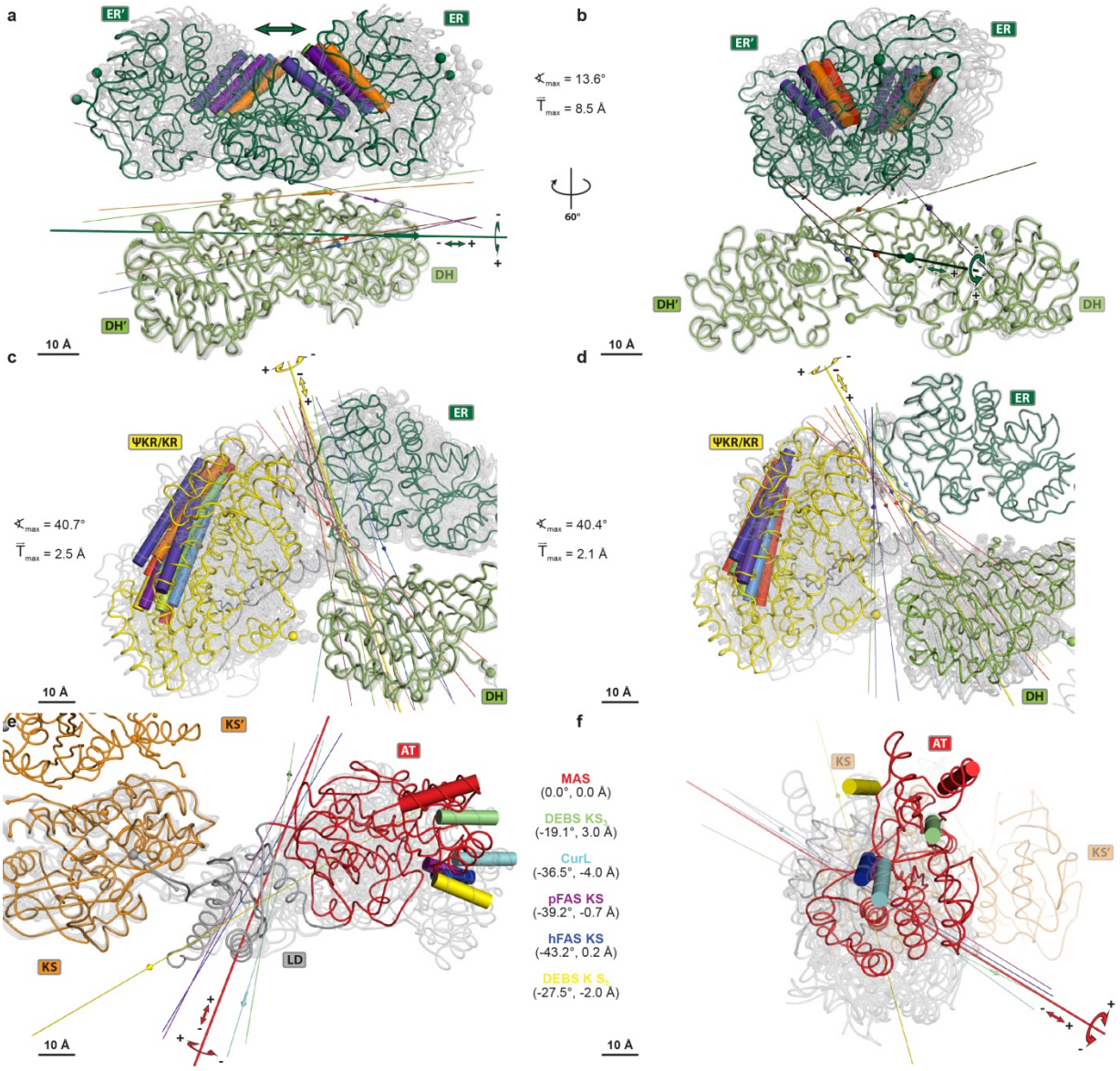
**Extended Data Figure 3 | Active site and structural comparison of the MAS ER and ΨKR/KR domains.** **a**, The MAS ER active site tunnel (white) is lined by an $NADP^+$ cofactor. **b**, An $F_{obs}$-$F_{calc}$ shaked omit map (contoured at 3.0 σ) is shown for the $NADP^+$ cofactor in chain J. **c**, The ER domains of FAS[6], MAS, and the modPKS PpsC dimerize via continuous β-sheet formation between the nucleotide binding subdomains ($ER_{NB}$), whereas the SpnB ER was crystallized as monomer and represents a group of isolated ER domains[18,19]. **d**, The active site of ΨKR/KR locates to an elongated surface groove, which partially extends to the ΨKR domain and is presumably closed upon ligand binding by a disordered lid region (aa 1948-1960). An $F_{obs}$-$F_{calc}$ omit map (contoured at 3.0 σ) is shown for the partially ordered $NADP^+$ cofactor. Left: surface, right: cartoon representation. **e**, MAS (pale yellow) features an N-terminal $\beta_1$-$\alpha_1$-$\beta_2$-$\alpha_2$ extension of the ΨKR Rossmann-fold, which is commonly found in PKSs (violet: Tylosin PKS $\Psi KR_1$[21]), but absent in FASs (green: porcine FAS (pFAS) ΨKR[6]). Secondary structure labels refer to MAS ΨKR. **f**, Average main chain B-factors across all chains reveal distally increasing flexibility with highest B-factors for the ΨKR domain, in particular its β-α-β-α extension, and the C-terminal ACP anchor.


FASs
fiPKSs
modPKSs and MsI-PKSs
DH
ΨME
ΨKR
ER
KR
#2VZ9
#4PIV
#4P7P
#3KG7
#3KG9
#3EL6
#4W82
#1PQW
#3SLK

**Extended Data Figure 4 | Alignment of linker regions of 55 fully reducing modifying regions of PKSs and FASs.** The alignment reveals sequence conservation of the β-sheet B1 (β1 and β2), which is inserted in a surface groove of the ΨKR/KR domain. In MAS, strands β3 and β4 form the second antiparallel β-sheet B2. The ER-KR linker is considerably shorter in a subgroup of modPKSs. Sequence numbers and secondary structure elements correspond to *M. smegmatis* MAS (MAS (Ms) highlighted in orange). All modules are labeled as: Protein name (organism abbr.) Uniprot number. Modules of Msl-PKSs (green text), modPKSs (light green), fiPKSs (blue), and FASs (yellow) are grouped by phylogeny (for details and color coding see Extended Data Fig. 7). PDB identifiers are indicated in the boxes representing the corresponding domains. Amino acids are shown in Clustal colors. ([*], diketide synthase; [†], PKS cluster contains non-colinear iterative modules; [‡], modular non-colinear iPKS module ; [§], *trans*-AT PKS)

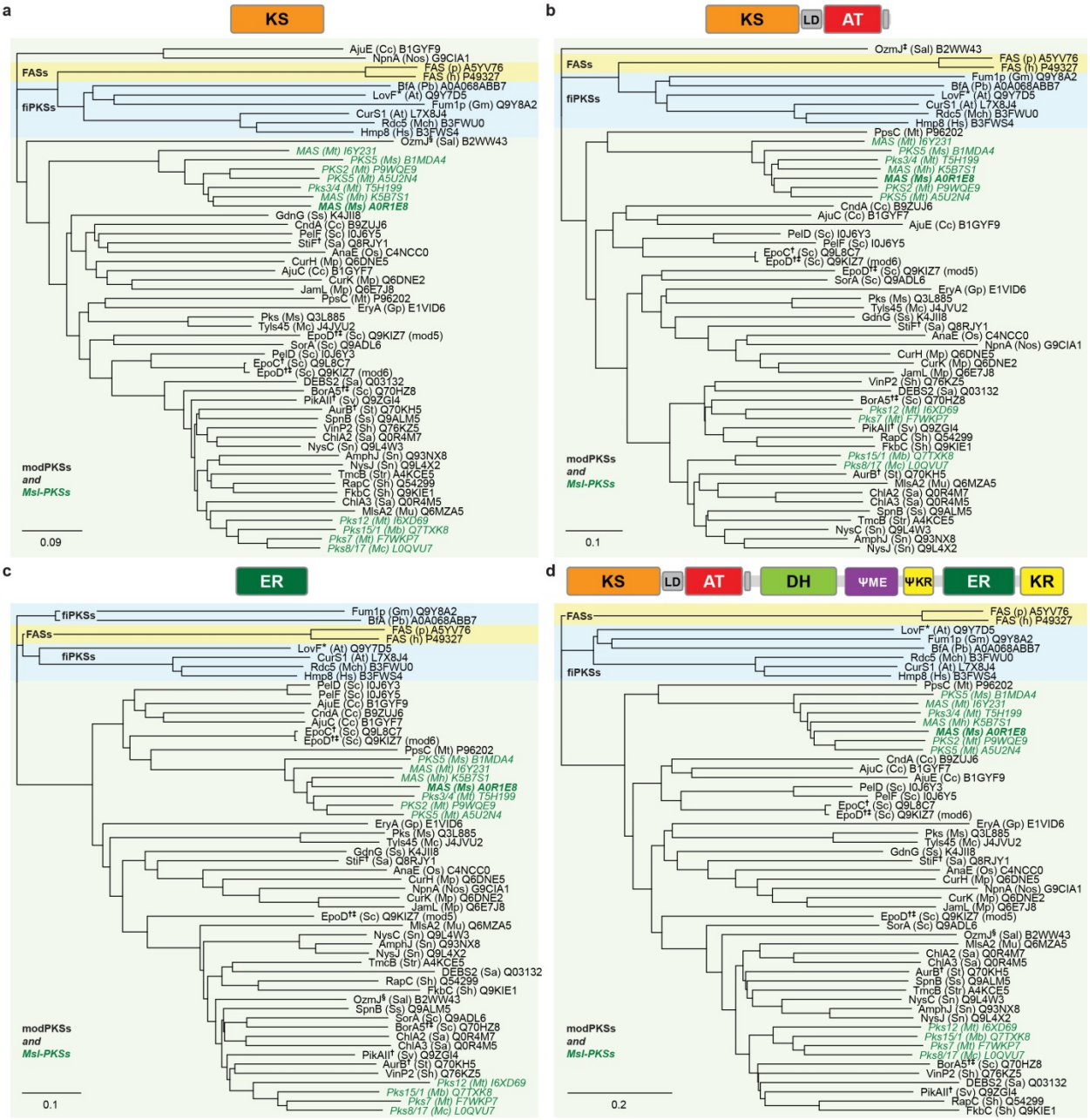


**Extended Data Figure 5 | Helical organization of central linking segments in MAS and modPKSs.** **a,** Assembly of the MAS central linking region from authentic crystal structures of the condensing and modifying regions. The two structures overlap in sequence by four residues (blue). **b**, Hybrid model based on the homology completed KS dimer and reconnected helical linkers. Ends of loops defined by the KS-AT crystal structure are indicated by black spheres. Disordered segments in the dimeric condensing region are reconstructed by multi-template homology modelling (gold); color coding is as in a. **c**, **d,** Helix formation in sequence regions corresponding to central linkers are also observed in the isolated crystal structure of the modPKS DH domain of the fully-reducing DEBS module 4[16] (**c**), $RifDH_{10}$[17] (not shown) and in the crystal structure of the RhiE KS-B didomain[24] (**d**), where a KS domain is connected directly to a DH homologous domain, the B domain.

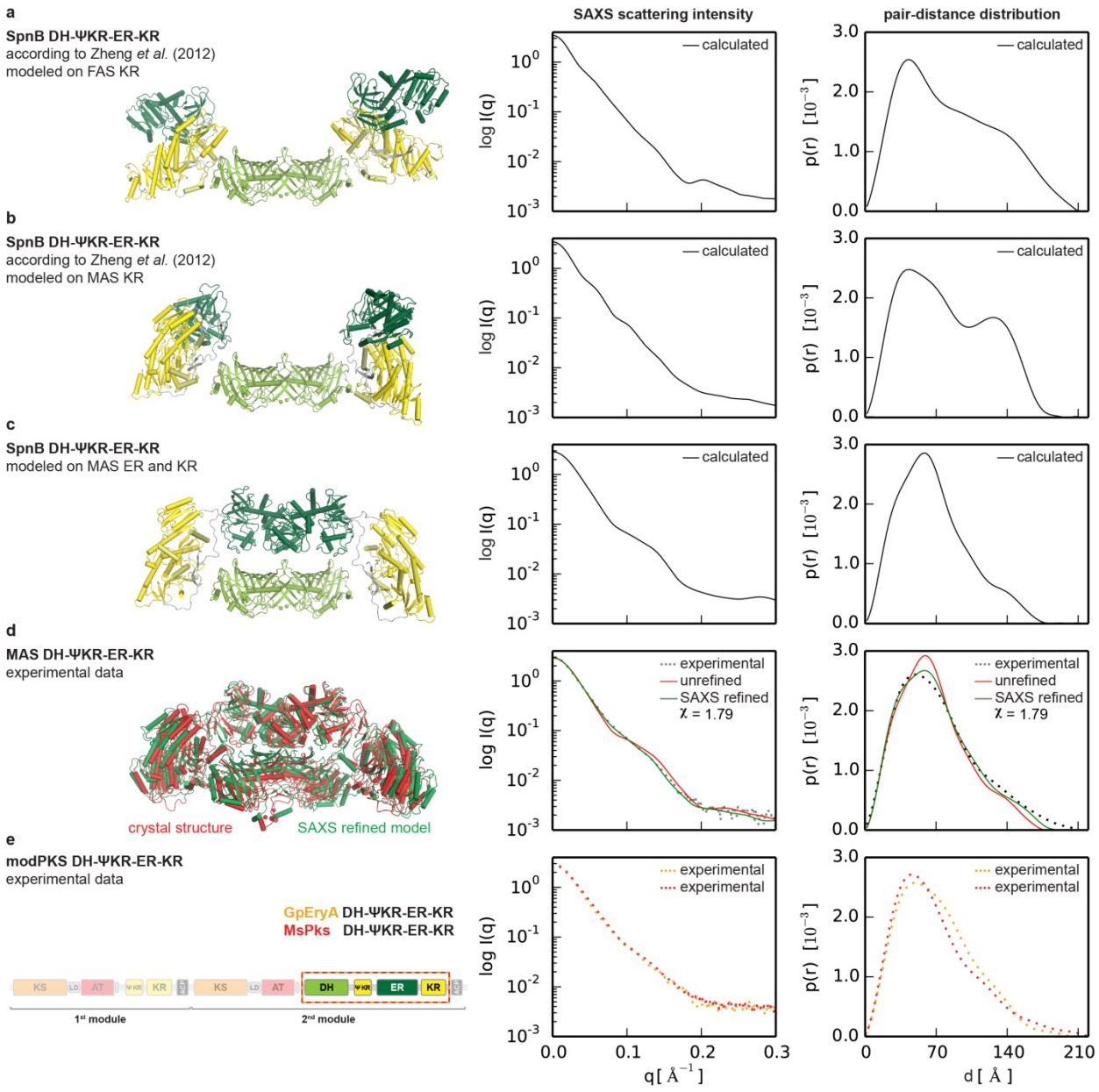



**Extended Data Figure 6 | Analysis of structural variability in the modifying and condensing regions of MAS and related multienzymes. a-d**, Analysis of interdomain conformational variability between the 18 protein chains in the MAS modifying region crystal structure. **a, b**, Variability of ER positioning relative to DH from two perspectives reveals a screw axis motion combining translation of up to 8.5 Å with rotation of up to 13.6°. **c, d**, Variability of ΨKR/KR domain orientation relative to DH (**c**) and ER (**d**), respectively, reveals a hinge located in the interdomain linker region. **e**, **f**, Top and front view of six overlayed KS-AT didomain structures[6,11,12,66,67] as indicated and the derived rotational distance of AT

positioning around a common hinge in the LD. **a-f**, Relative locations of individual structures are highlighted by representative colored helices. Translational components are indicated with an arrow on the rotation axes with signs indicated on the principle axis (thick, colored according to the moving domain). All structures are aligned to a MAS reference domain (colored ribbon). Rotation axes are shown for rotations larger than 6° and arrows are shown for translations larger than 1 Å.



**Extended Data Figure 7 | A comprehensive phylogenetic analysis classifies MAS into the branch of modPKSs.** Phylogenetic tree for 55 fully reducing MASs/PKSs/FASs modules were constructed based on only KS domains (**a**), complete condensing regions (**b**), the ER domain (**c**), or all catalytic domains (**d**). *M. smegmatis* MAS (MAS (Ms), bold, italic) and Msl-PKSs (italic) are more closely related to modPKSs (light green) and distinct to fiPKSs (blue) and animal FASs (yellow). All modules are labeled as: Protein name (organism abbr.) Uniprot number. Units are given as amino acid substitutions per site. Indices correspond to Extended Data Fig. 5.




**Extended Data Figure 8 | SAXS analysis supports a MAS-like organization of PKS modifying regions.** Models (left) of modifying region organization and their respective theoretical and experimental scattering curves as well as pair-distance distributions (right) are shown. **a, b,** As proposed by Zheng *et al.*, the intact SpnB modifying region was modeled based on the domain-swapped SpnB ER-ΨKR/KR structure[19], using either the structure of FAS (**a**) or of the MAS modifying region (**b**) as a guide for positioning KR relative to DH. The SpnB DH structure was generated by homology modelling. **c**, Model

of the intact SpnB modifying region with dimeric DH and ER based on the structure of the intact MAS modifying region. **d,** Crystal structure of MAS before and after fitting to experimental SAXS data. A good fit ($\chi$=1.79) is obtained by fitting SAXS data with a single model corresponding to an average conformation of the MAS structure. **e,** Sequence organization of two authentic modPKS modifying regions of similar ER-KR linker length to SpnB (left), together with experimental SAXS scattering data (right). The data closely match calculated scattering curves for a MAS-like architecture, but disagree with models based on a monomeric ER as suggested for SpnB.

**Corrigendum**

In this letter, we have studied the three-dimensional structure of a protein from *Mycobacterium smegmatis* assigned as mycocerosic acid synthase (MAS) in sequence databases (Uniprot: A0R1E8 ; NCBI: YP_888986.1) to provide, in conclusion, a template structure of MAS-like PKS and a first example for the architecture of reducing polyketide synthases (PKSs). However, we now note that *Etienne et al.*[1] provided a biochemical characterization of a deletion strain of the corresponding gene MSMEG_4727, which indicated a physiological role of the protein in the production of 2, 4,-dimethyl-2-eicosenoic acid, a lipid component of lipooligosaccharides, rather than mycocerosic acids, via a reaction closely related to those of MAS. Until comprehensive characterization at the protein level is available, the studied protein should thus be referred to as a „mycocerosic-acid synthase like-PKS/MAS-like PKS“. We thank the authors for drawing our attention to this publication. Scientific conclusions of our manuscript remain unchanged.

1. Etienne, G., Malaga, W., Laval, F., Lemassu, A., Guilhot, C., Daffé, M. Identification of the polyketide synthase involved in the biosynthesis of the surface-exposed lipooligosaccharides in mycobacteria. *J Bacteriol.* 191(8), 2613-21 (2009).