Mycobacterium tuberculosis FasR senses long fatty acyl-CoA through a tunnel, inducing DNA-dissociation via a transmission spine

Julia Lara[1], Lautaro Diacovich[1,2], Felipe Trajtenberg[3], Nicole Larrieux[3], Emilio L. Malchiodi[4], Marisa M. Fernandez[4], Gabriela Gago[1], Hugo Gramajo[1,*] and Alejandro Buschiazzo[3,5,*]

[1] Laboratory of Physiology and Genetics of Actinomycetes, Instituto de Biología Molecular y Celular de Rosario (IBR-CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, 2000, Argentina
[2] Plataforma Argentina de Biología Estructural y Metabolómica (PLABEM), Rosario, 2000, Argentina
[3] Laboratory of Molecular & Structural Microbiology, Institut Pasteur de Montevideo, Montevideo, 11400, Uruguay
[4] Instituto de Estudios de la Inmunidad Humoral (IDEHU/CONICET-UBA), Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, 1113, Argentina
[5] Integrative Microbiology of Zoonotic Agents, Department of Microbiology, Institut Pasteur, Paris, 75724 Cedex 15, France

* To whom correspondence should be addressed. Alejandro Buschiazzo Tel: +598-2522-0910; Fax: +598-2522-4185; Email: alebus@pasteur.edu.uy; Correspondence may also be addressed to Hugo Gramajo; Email: gramajo@ibr-conicet.gov.ar

ABSTRACT

Mycobacterium tuberculosis is a pathogen with a unique cell envelope including very long fatty acids, implicated in bacterial resistance and host immune modulation. FasR is a two-domain transcriptional activator that belongs to the TetR family of regulators, and plays a central role in mycobacterial long-chain fatty acyl-CoA sensing and lipid biosynthesis regulation. We now disclose crystal structures of *M. tuberculosis* FasR in complex with acyl effector ligands and with DNA, uncovering its sensory and switching mechanisms. A long tunnel traverses the entire effector-binding domain, enabling long fatty acyl effectors to bind. Only when the tunnel is entirely occupied, the protein dimer adopts a rigid configuration, with its DNA-binding domains in an open state that leads to DNA dissociation. Structure-guided point-mutations further support this effector-dependent mechanism. The protein-folding hydrophobic core, connecting the two domains, is completed by the effector ligand into a continuous spine, explaining the allosteric flexible-to-ordered transition. The transmission spine is conserved in all TetR-like transcription factors, offering new opportunities for anti-tuberculosis drug discovery.

INTRODUCTION

The composition and complexity of the mycobacterial cell envelope are two of the most distinctive features of this genus. The complex lipids present in the cell wall of *Mycobacterium tuberculosis* (*Mtb*)

act as major effector molecules that interact with the host, playing key roles in pathogenicity and also providing a barrier against environmental stress, antibiotics, and the host's immune response (1). Understanding the biogenesis of the mycobacterial cell wall will thus provide with relevant insights into the biology of this pathogen, also identifying potential targets for the development of new antimycobacterial compounds.

The outer membrane of *Mtb* comprises very long-chain fatty acids (mycolic acids), found in the inner leaflet covalently bonded to the arabinogalactan–peptidoglycan layer, and also in the outer leaflet as non-covalently associated lipids in the form of trehalose-mono- and di-mycolate (2). Mycolic acids, a hallmark of *Mycobacterium*, are synthesized by way of two fatty acid synthase systems, FAS I and FAS II. The multidomain single protein FAS I catalyses *de novo* biosynthesis of acyl-CoAs in a bimodal fashion rendering C_{16-18} and C_{24-26} derivatives (3). Long chain acyl-CoAs are used as primers by the FAS II multiprotein system, and iteratively condensed with malonyl-acyl carrier protein (malonyl-ACP) leading to very long-chain meromycolyl-ACPs (up to C_{56}). The latter are eventually condensed to FAS I-synthesized C_{24-26} fatty acids (previously activated by the acyl-CoA carboxylase 4 complex (4)) to produce mycolic acids. The long-chain acyl-CoAs generated by FAS I are not only used as precursors of mycolic acids, but also for the biosynthesis of phospholipids, triacylglycerides, diverse polyketides and other complex lipids, relevant for *Mtb* pathogenicity (5,6). A complex regulatory network integrating all these pathways must exist in order to maintain lipid homeostasis. However, and despite the biological relevance of lipid-derived molecules in *Mtb*'s lifecycle, little is known about the environmental signals and the regulation cascades controlling lipid metabolism in this bacterium.

We had previously identified the transcription factor FasR, as a key activator of *fas* and *acpS* gene expression in mycobacteria (7). The *fas* and *acpS* genes form an operon, respectively coding for the FAS I synthase and the 4-phosphopantetheinyl transferase, the latter being an essential enzyme to produce functional ACP, central to all fatty-acid biosynthesis. To activate the *Mtb fas-acpS* operon, FasR binds to three inverted repeats in the promoter region, a binding that is regulated by long-chain acyl-CoAs, products of FAS I (7). More specifically, acyl-CoAs $\geq C_{16}$ disrupt the interaction of *Mtb* FasR with its cognate DNA. Additionally, FasR is essential for *M. smegmatis* viability (7), further highlighting its key role in mycobacterial biology.

FasR sequence reveals homology to members of the TetR family of regulators (TFRs), which are one-component sensory transduction proteins (8), typically dimeric and with each protomer displaying a 2-domain all-helical structure (9). By aligning FasR to orthologous TFRs with known 3D structures (Supplementary Fig. S1A) (10), FasR is expected to comprise a helix-turn-helix DNA-binding domain towards the N-terminus (residues 1-80), and a larger C-terminal domain corresponding to the ligand- or effector-binding domain (residues 82-228). In contrast to the DNA-binding domain, the effector-binding region reveals no sequence homology with other TFR proteins, such sequence diversity being consistent with the large variety of effector signals sensed by different TFRs (11). The prototype of the TFRs is TetR from the Tn10 transposon of *Escherichia coli*, which regulates the expression of the tetracycline efflux pump in Gram-negative bacteria (12). However, TFR proteins are widely distributed among bacteria, and control a broad range of processes, including fatty acid biosynthesis (13).

Interestingly, the vast majority of TFR proteins are transcriptional repressors, with very few exceptions acting as activators (11).

We have now determined the 3D structures of FasR from *Mtb*, in three different states obtained from i- the protein crystallized alone, ii- co-crystallised in complex with the fatty acid C_{20}-acyl-CoA, and iii- in complex with a double-stranded DNA oligonucleotide bearing the specific FasR-binding sequence. The comparison of these crystal structures, together with the functional characterisation of structure-guided FasR point-mutants and molecular dynamics computational simulations, uncovered the molecular mechanisms by which long- (C_{16}-C_{20}) and very long-chain (C_{22}-C_{26}) acyl-CoA molecules are sensed by FasR, as well as the means by which such signal disrupts cognate FasR-DNA binding and hence actuates *fas-acpS* transcriptional activation. World-wide efforts have disclosed hundreds of protein structures from *Mtb* corresponding to potential drug targets, a valuable input for a number of drug discovery projects (14,15). The uncovering of novel structural and mechanistic insights about a key *Mtb* metabolic regulator, contributes with solid molecular bases for target-based drug discovery, certainly one of the sensible strategies to combat tuberculosis.

MATERIAL AND METHODS

Bacterial strains and plasmids

E. coli strain DH5α cells were used for DNA cloning purposes and transformed according to standard methods. For protein expression the *E. coli* BL21 λ(DE3) strain was used instead. The *fasR* gene (*rv3208*) was PCR-amplified from *Mtb* H37Rv genomic DNA using the oligonucleotides F-Rv3208 (5′-CCCTCCATATGGAAAACCTGTACTTCCAGGGTATGAGCGATCTCGCCAAG-3′) to introduce an NdeI site at the translational start codon and encode a fused Tobacco Etch Virus protease digestion site (TEV), and R-Rv3208 (5′-GAATTCCTACGAGCGGGTAAGCG-3′) to introduce an EcoRI site at the end of the ORF. To generate a FasR recombinant protein with a hexa-histidine-TEV-tagged site at the N-terminus, the corresponding PCR product was extracted from the agarose gel, cloned into the pCR BluntII TOPO vector (Invitrogen), digested with NdeI and EcoRI, and finally subcloned into the expression vector pET28a as described by the manufacturer (Merck). The recombinant plasmid (pET28_ *fasR*) was transformed into BL21 λ(DE3) cells, and transformants were selected on LB agar plates containing 50 μg/ml kanamycin. To produce the His-tagged FasR version used in all EMSA experiments (FasRwt), *fasR* was PCR-amplified from genomic DNA of *Mtb* H37Rv using the oligonucleotides F-Rv3208 (5′-CATATGAGCGATCTCGCCAAGACA-3′) to introduce a NdeI site at the translational start codon, and R-Rv3208 (5′-GAATTCCTACGAGCGGGTAAGCGG-3′) to introduce an EcoRI site at the end of the ORF. To generate a *fasR* His-tag fusion gene, the PCR product was cloned into the pCR BluntII TOPO vector, digested with NdeI and EcoRI and subcloned into NdeI/EcoRI-cleaved pET28a. The resulting plasmid (pET28_ $fasR_H$) was transformed into BL21 λ(DE3) cells. The plasmids to produce recombinant $FasR_{\Delta 33}$ and FasR point mutants ($FasR_{LVL}$, $FasR_{L106F}$, $FasR_{L98A}$ and $FasR_{F123A}$) with N-terminal hexa-histidine-TEV-tagged sites, were synthesized (GenScript). NdeI/HindIII restriction sites were engineered in a pUC57-Am plasmid where genes of interest were introduced. Synthetic plasmids were digested with NdeI and HindIII, extracted from the agarose gel, and inserted into the expression vector pET28a, as described by the

manufacturer (Merck). The recombinant plasmids were transformed into BL21 λ(DE3) cells and transformants selected on LB agar plates with 50 μg/ml kanamycin. DNA sequences of all genes were verified by Sanger sequencing.

Expression and purification of proteins

Expressions of all N-terminally hexa-histidine-TEV-tagged or hexa-histidine-tagged proteins used in this work were carried out following isopropyl-β-thiogalactoside (IPTG) induction in BL21 λ(DE3) *E. coli*. Bacteria were grown at 37 C in 500 ml LB broth to 0.6-0.7 absorbance at 600 nm. IPTG was then added to 0.3-0.5 mM and the culture was grown for 12 h at 23 C. Cells were harvested by centrifugation at 2800 *g* at 4ºC, resuspended in 30 ml of lysis buffer (50 mM Tris.HCl pH 8, 150 mM NaCl, 5 mM imidazole, 10 % glycerol, 10mM β-mercaptoethanol) and lysed by sonication. After centrifugation (25000 *g*, 30 min, 4 C), the supernatant was recovered and FasR, $FasR_{\Delta 33}$ or FasR point mutants were separated from whole-cell lysates by Ni-NTA agarose chromatography (Qiagen, Inc). After three washing steps with lysis buffer, His_6-tagged FasR were eluted from the resin with 250 mM imidazole in lysis buffer, dialysed overnight against FasR Buffer 1 (10 mM Tris.HCl pH 8, 300 mM NaCl). In the case of hexa-histidine-TEV-tagged proteins (FasR, $FasR_{\Delta 33}$, $FasR_{LVL}$, $FasR_{L1069F}$, $FasR_{L98A}$ and $FasR_{F123A}$) after three washing steps with lysis buffer, 0.5 mg TEV protease and DTT to 1 mM final concentration were added. The mixture was incubated 2-3 hs at 23 C and 12 hs at 4 C. Proteins were eluted from the resin with 5 mM imidazole in lysis buffer and dialysed overnight against FasR Buffer 2 (10 mM Tris.HCl pH 8, 300 mM NaCl, 5% glycerol). A final size-exclusion chromatography step (S200-Superdex, GE) was performed for all proteins, pre-equilibrating the column in FasR Buffer 2 and eluting isocratically at 0.5 ml/min with high-performance liquid chromatography (Akta Purifier, GE). $FasR_{LVL}$ was the only mutant that showed detectable levels of monomeric form eluting from SEC (all other proteins eluting as dimeric species), in which case the peak corresponding to the dimer was recovered for further functional analyses by EMSA. Protein purity was controlled by Coomassie blue staining after SDS–PAGE on a 15 % polyacrylamide gel. Protein concentrations were determined by UV spectroscopy. Purified proteins were stored at 4 C.

Surface plasmon resonance

Surface plasmon resonance analyses were performed with a Biacore T100 instrument (GE Healthcare). Pure FasR and $FasR_{F123A}$ were dialysed against 10 mM sodium acetate pH 4.5 and coupled to a CM5 sensor chip using the Amine Coupling Kit (GE Healthcare). Micromolar concentrations of C_{20}-CoA were dialysed against 25 mM Tris.HCl pH 8, 150 mM NaCl, 0.005 % Tween 20. Serial two-fold dilutions were made in the same buffer and injected over the chip surface. Dissociation was then carried out by injecting buffer alone. Nonspecific binding was considered by injecting identical analyte concentrations over a control surface with no protein. Experiments were performed at 25 C by triplicate, producing standard deviations of less than 12 %. Data were analysed using Biacore T100 Evaluation software.

Electrophoretic mobility shift assays (EMSAs)

His_6-tagged FasR, and FasR point mutant proteins ($FasR_{LVL}$, $FasR_{L106F}$, $FasR_{L98A}$ and $FasR_{F123A}$) were used to assess the protein binding to P*fas*$_{MT}$ (398 bp) promoter fragment. The promoter DNA fragment for these assays was generated by PCR amplification from *Mtb* genomic DNA with primers N2_Fas1Mt-prom (5′-CATAACGATTTGATAACAAAACTGC-3′) and C_Fas1Mt-prom (5′-CACCCGGTCGTGCTCGTGGATCGTC-3′). N2_Fas1Mt-prom primer was end-labelled with [γ-^{32}P] ATP (3,000 Ci mmol^{-1}) using T4 polynucleotide kinase and the PCR product obtained was purified from agarose gels. Proteins were either pre-incubated or not with acyl-CoAs and then the mixtures incubated with ^{32}P-labelled probe (1000-5000 cpm) in a total volume of 25 µL binding buffer (25 mM Tris.HCl pH 8, 1 mM PMSF, 5 % (v/v) glycerol, 5 mM $MgCl_2$, 150 mM NaCl and 1 µg poly-dIdC) at room temperature for 15 min. DNA-protein complexes were resolved by electrophoresis on a 6% (w/v) non-denaturing polyacrylamide gel in 1X TBE (89 mM Tris Base; 89 mM Boric Acid; 2 mM EDTA), 5 % (v/v) glycerol at 150 V, on ice. Results were visualized and recorded with a Typhoon™ FLA 7000 scanner (GE).

The equilibrium dissociation constant (K_D) is a quantitative measurement to assess the affinity of biological interactions. For the FasR:DNA EMSA binding experiments described in this work we define the K_D as the concentration of FasR for which 50% of the DNA is in complex with the protein. The relationship between K_D and affinity is reciprocal, lower K_Ds correspond to higher affinities. After performing binding reactions in which the protein is titrated, the fraction of DNA bound at each concentration of protein is calculated and the data are adjusted to a binding equation using nonlinear regression (16). Densitometry of DNA bands was performed with GelPro software considering background subtraction. The fraction of bound DNA was plotted as a function of protein concentration and fit to the following binding equation (**Eq. 1**) using Prism software to perform non-linear regression:

Eq 1. Fraction Bound = $B_{max} * [P] / ({}^{app}K_D + [P])$

Knowing [P] = protein concentration, the apparent dissociation constant (${}^{app}K_D$) can be quantified as well as the maximal fraction bound (B_{max}) plateau.

Acyl-CoA effectors trigger FasR:DNA dissociation, and can thus be treated as non-competitive binder inhibitors. Dose-response experiments were performed to assess inhibition activity, using a modified EMSA protocol. A constant and saturating concentration of protein was equilibrated with 0.3 nM labelled DNA and increasing concentrations of acyl-CoAs in equilibration buffer. The EMSA data were collected as above and fit to a sigmoidal dose-response equation in Prism (**Eq. 2**) to determine the acyl-CoA concentration required to displace half of the bound FasR:DNA complex (IC_{50}):

Eq 2. Fraction Bound = $100 / (1 + 10 * (E_f - \log IC_{50}))$

E_f = logarithm of acyl-CoA concentration.

EMSA experiments with $FasR_{wt}$ and selected FasR mutants were performed by triplicate, to express ${}^{app}K_D$ and IC_{50} as average values ± one standard error of the mean.

Apparent inhibition constants (${}^{app}K_i$) were also calculated (17,18), to correct IC_{50} figures taking into account the K_D as well as the concentrations of labelled DNA ([D]) and protein ([P]) according to the Lin and Riggs conversion (**Eq. 3**):

Eq 3. ${}^{app}K_i = 2 * K_D * IC_{50} / (2[P] - [D] - 2K_D)$

Crystallisation and data collection

$FasR_{\Delta 33}$ (5 mg/ml) crystallised at 20 C, mixing 2+2 µl of protein and reservoir solution (0.1 M MES monohydrate pH 6.0, 22 % v/v polyethylene glycol 400) using 1 ml reservoir on a hanging-drop vapour-diffusion setup. Transferred to mother liquor with 20 % (v/v) glycerol as cryoprotectant, crystals were mounted in cryo-loops (Hampton Research), and flash cooled in liquid nitrogen (19). X-ray diffraction data were collected at -163 C at the in-house Protein Crystallography Facility of the Institut Pasteur de Montevideo (Uruguay) with a MicroMax-007 HF rotating Cu anode (Rigaku) and a Mar345 image plate (marXperts).

$FasR_{\Delta 33}$ (5 mg/ml) was crystallised in complex with C_{20}-CoA using a 1:1 molar stoichiometry of the acyl-CoA ligand in the crystallisation drops (2 µl protein:ligand + 2 µl mother liquor 2.2 M NaCl, 0.1 M Na-acetate trihydrate pH 4.7). A hanging-drop vapour-diffusion setup was incubated at 20 C using 1 ml mother liquor as reservoir. Crystals were transferred to mother liquor with 25 % (v/v) glycerol, mounted in cryo-loops and flash cooled in liquid N_2. X-ray diffraction data were collected at -173 C at SOLEIL synchrotron (PROXIMA 1 beamline, France) (20), using a PILATUS 6M detector (Dectris) (21).

FasR-DNA crystals were obtained by co-crystallisation. Double-stranded DNA was generated by co-incubating the two complementary oligonucleotides FwPfas25nt (5'-TACCCGTACGTAGAACTCGCCAGTA-3') and RvPfas25nt (5'-TACTGGCGAGTTCTACGTACGGGTA-3') under standard slow-cooling hybridisation conditions. Double-stranded DNA was mixed in a 1:1 molar stoichiometry with full-length FasR (5 mg/ml) and set for hanging-drop vapour-diffusion crystallisation by mixing 2 µl of protein:DNA with equal volume of mother liquor (28 % (w/v) PEG monomethyl ether 2000, 0.25 M ammonium citrate pH 7.0, 0.1 M imidazole) over 1ml mother liquor as reservoir solution. Crystals were transferred to mother liquor with 35 % (v/v) PEG 400, mounted in cryo-loops and flash cooled in liquid N_2. X-ray diffraction data were collected at -173 C at Diamond Light Source synchrotron (I04-1 beamline, UK), using a PILATUS 6M detector (Dectris).

Bragg diffraction intensities were integrated with XDS (22), and scaled and reduced to amplitudes with Aimless and Ctruncate (23).

Structure determination and refinement

The structure of $FasR_{\Delta 33}$-C_{20}-CoA was solved *ab initio* with Arcimboldo (24,25) which uses Phaser (26) as molecular replacement (MR) engine to place α-helices, and ShelxE (27) for density modification and chain-trace extension. The structure of $FasR_{\Delta 33}$-C_{14} was solved by MR (26) using the refined $FasR_{\Delta 33}$-C_{20}-CoA model as search probe. Buster (28) was used to refine both $FasR_{\Delta 33}$-C_{20}-CoA and $FasR_{\Delta 33}$-C_{14} atomic models, iterating with manual model rebuilding and validation with Coot (29). Final validation was done with MolProbity (30). OMIT maps were calculated for the $FasR_{\Delta 33}$-C_{20}-CoA and $FasR_{\Delta 33}$-C_{14} structures according to the Polder approach (31), using final refined models from which only the acyl-CoA ligands were omitted to calculate structure factors. Final refined maps were also used to calculate real space correlation coefficients with respect to model-derived maps, on a per residue basis (32).

The FasR-DNA complex was solved by MR (26) using a portion (corresponding to the dimeric regulatory domains with no ligands nor HTH domains included) of the refined $FasR_{\Delta 33}$-C_{20}-CoA model as search probe. Initial refinement (33) with this partial model reduced R-factors to <50 %, and produced Fourier difference maps that clearly revealed the presence of both HTH domains and double-stranded DNA. Limited resolution and model incompleteness resulted however in mediocre $2mF_{obs}$-DF_{calc} maps at this point. ShelxE (27) was instrumental at improving electron density continuity, using the unrefined MR solution model and diffraction intensities as inputs, a larger than usual sphere of influence (5 Å) for density modification, the free-lunch option set at 3.85 Å to better handle data incompleteness at the higher resolution shells, and testing different solvent contents (from 0.4 to 0.6). All output maps were visualized superposed, readily allowing for manual main-chain tracing (29) of HTH and DNA portions, only including residues clearly visible in electron density at each cycle. This procedure was cycled iteratively using the progressively more complete protein models, running 5 cycles of ShelxE density modification each time. For the manual model (re)building, recent developments for real-space fitting in very low-resolution maps within Coot (34) proved essential, using Prosmart-generated external restrains, map blurring and optimized Geman-McClure parameters (alpha=0.4 proved best) for optimal movement of DNA within electron density.

The orthorhombic $C222_1$ space group of the FasR-DNA crystals was confirmed by several standard procedures (23,35), notably including integration in the corresponding triclinic group and using molecular replacement as a means of deducing real symmetry. However, the DNA molecule (one double-stranded oligonucleotide per ASU) was found sitting with its long axis approximately perpendicular to one of the orthorhombic 2-fold axes, revealing both alternative 5'→3' orientations to be present in different unit cells of the crystal. This form of static disorder is frequent when a crystal contains two or more very similar molecules (such as complementary strands in DNA with palindromic sequences), which may occupy the same lattice position with low occupancy. Refinement was performed with phenix.refine (33), using external restrains from the high resolution FasR models described above and overall strategies to deal with low resolution data. Intermediate cycles of real space refinement (36) were also instrumental. The occupancies of DNA atoms were reduced to 0.5, avoiding bumping restraints.

Difference Fourier maps very clearly revealed the positions of the two missing DNA-binding domains as well as the double-stranded DNA molecule early on during refinement. However, difference maps did not show any detectable signal at the expected positions of putative ligands bound within the effector-binding tunnel. To confirm this, the acyl moieties from the $FasR_{\Delta 33}$-C_{14} model were superimposed in place within the FasR-DNA effector-binding tunnel, manually regularised to minimise clashes and added into the FasR-DNA model for further restrained refinement. Not only the crystallographic R-factors increased significantly (for both working and free sets of reflections), but difference Fourier maps revealed >5 σ negative peaks on the acyl atoms, indicating they are not actually present in the crystal.

Structural analyses were done with the CCP4 suite (37), PISA (38) and illustrations produced with Pymol (39).

Structural bioinformatics to define the hydrophobic spine

FasR structural homologues were searched with PDBeFold (40), thus retrieving a wide range of sequence similarities. Structural alignment of such hits was performed with T-Coffee Expresso (41). The resulting 337 sequences were filtered keeping only 76 that had < 80% identity. This multiple sequence alignment (MSA) served to generate a profile hidden Markov model using HMMBuild (42). The UNIPROT database was searched with this profile HMM, CD-HIT (43) was used to filter redundant sequences, ultimately producing a list of 2591 sequences, all containing TFR effector-binding and DNA-binding domains. A MSA was calculated with T-Coffee in M-Coffee mode (44), and the resulting alignment allowed the calculation of observed frequencies for each one of the 20 amino acids for each position in the MSA. A score for each MSA position was then generated by multiplying each frequency by the corresponding hydrophobicity index (45) and summing for all 20 amino acids. The hydrophobic spine was defined as the set of positions with a weighted hydrophobicity index (which goes from -4.5 [hydrophilic] to 4.5 [hydrophobic]) equal or larger than 2. Sequence alignment figures were prepared with Espript (46).

Molecular Dynamics simulation of $FasR_{\Delta 33}$ bound to C_{26}-CoA

The $FasR_{\Delta 33}$-C_{26}-CoA complex was built using the $FasR_{\Delta 33}$-C_{20}-CoA (PDB 6O6N) model as template. The bound acyl-CoA was manually extended by six carbons using Pymol (39). C_{26}-CoA was optimized and 10,000 rotamers were generated with RDKit (47). Energy minimization was performed with the Rosetta suite (48) using dimer symmetry constraints, harmonic restraints were used to preserve the ligands positions as observed in the crystal structure, and 10,000 models were generated. The best complex was selected based on Rosetta energy score, with optimal stereochemical geometry and no clashes. The selected model was used as starting structure for classical molecular dynamics simulations using Gromacs 2018_cuda8.0 and GROMOS96 43a1 force field (49). An octahedron box was solvated and charge-balancing counterions were included to neutralize charges (50). Initially, the system was relaxed by energy minimization, and then equilibrated for 200 ps using a reference temperature of 27 C. Simulations were performed for 10 ns with no constraints, recording snapshots every 5 ps for analysis. Visualization of protein models and structural analyses were performed with VMD (51) and Pymol (39).

RESULTS

Three-dimensional structures of FasR

Recombinant FasR eluted from size exclusion chromatography suggesting a dimeric structure (~52 kDa), similar to all TFRs. Attempts to crystallise full-length FasR alone failed. We hypothesised that the 33-amino acid segment at the N-terminus is likely flexible, based on multiple sequence alignments to close orthologues (≥80% identity) from mycobacteria (Supplementary Fig. S1B). A particularly strong sequence variation of the short N-terminal extensions was revealed, also with secondary structure predicted to be absent. We thus generated a truncated form of FasR lacking the first 33 amino acids ($FasR_{\Delta 33}$).

$FasR_{\Delta 33}$ readily crystallised in the absence of added ligands and also in complex with acyl C_{20}-CoA (arachinoyl- or arachidoyl-CoA). Both crystal forms diffracted X-rays at better than 1.7 Å resolution (Table 1), and their structures confirm the dimeric architecture of $FasR_{\Delta 33}$, with each protomer organized in two all-helical domains (Fig. 1A), similar to all TFRs (11).

$FasR_{\Delta 33}$-C_{20}-CoA was solved first using *ab initio* methods (25), facilitated by high enough diffraction resolution and the all-helical nature of the protein. A typical TFR structural architecture was found, with a DNA-binding HTH domain comprising helices α1-α3 from the N-terminus to residue Ser_{82}. And a regulatory effector-binding domain located immediately C-terminal to the HTH, from Lys_{83} to the C-terminus, including helices α4-α9. The latter are roughly organized in two bundles, the α4-α7 core runs along the long axis of the ellipsoid regulatory domain, whereas α8-α9, roughly perpendicular to the core, mediate dimerization by forming a 4-helix bundle with the other protomer's α8'-α9' helices (Fig. 1A). The $FasR_{\Delta 33}$-C_{20}-CoA dimer is strictly symmetric, with the crystallographic 2-fold axis relating one protomer to the other.

One of the most striking features of $FasR_{\Delta 33}$-C_{20}-CoA is a tunnel-like cavity, delimited by helices α4, α5, α7 and α8, with its two openings towards the 'bottom' and the 'top' of the regulatory domain. C_{20}-CoA binds within this tunnel, in an overall parallel orientation with respect to core helices α4, α5 and α7 (Fig. 1A,B). The tunnel is ~28 Å long, with a predominance of hydrophobic residues on its wall pointing their side chains towards the lumen of the tunnel (Fig. 1B). The fatty acid is well defined all along the tunnel (Supplementary Fig. S2A). Towards the upper entrance of the cavity, most of the 4'-phosphopantetheine portion of the CoA cofactor is also observed, with the higher electron density sulphur atom being instrumental in positioning the whole C_{20}-CoA moiety. Likely due to high mobility of the CoA portions exposed to the solvent, electron density is less clear towards the tip of the pantoic group, and eventually becomes indistinguishable from noise in the region corresponding to the 3'-phosphoadenosine diphosphate group, explaining why they were not included in the final refined model.

$FasR_{\Delta 33}$ crystals were also grown in the absence of acyl-CoA with the goal of solving the ligand-free structure. Unexpectedly, additional electron density not corresponding to protein, was clearly visible within the $FasR_{\Delta 33}$ ligand-binding tunnel (Fig. 2A, Supplementary Fig. S2B). The shape of this density is compatible with part of a polyethylene glycol molecule (PEG 400 is a component of the crystallisation mother liquor), or yet also consistent with myristic acid (C_{14}). PEG is less likely, as it bears bridging oxygens along the alkyl chain, which would require satisfying H-bonding interactions in a hydrophobic environment. We hypothesize that C_{14} could have bound during protein expression in *E. coli*, and thus opted to model this fatty acid within the tunnel. Even if the chemical nature of the bound species is not certain, this piece of evidence results in two consequences: $FasR_{\Delta 33}$-C_{14} is not a true apo form of the protein, but it did disclose the structure of FasR with a shorter alkyl chain bound in the effector pocket, as compared to $FasR_{\Delta 33}$- C_{20}-CoA.

In contrast to the strictly symmetric organization of the $FasR_{\Delta 33}$-C_{20}-CoA dimer, $FasR_{\Delta 33}$-C_{14} displays one full dimer per asymmetric unit (Fig. 2B) with each protomer deviating from strict symmetry with respect to the other. A strong 2-fold non-crystallographic operator is nonetheless

relating both protomers, but applying only to the regulatory domains. The HTH domains of $FasR_{\Delta 33}$-C_{14} depart from this relationship, readily observed by superimposing one protomer onto the other (Fig. 2B), which results in the regulatory domains fitting together quite well, while the DNA-binding domains appear rotated by as much as 15º. To further characterise this symmetry deviation, two identical $FasR_{\Delta 33}$-C_{14} dimers were superimposed by maximizing the fit between one HTH domain from distinct protomers on each dimer (0.3Å root mean square deviation [rmsd] considering all atoms of the two superimposed HTHs). This rotation operation resulted in >5.5 Å rmsd between the other pair of HTHs. A similar exercise using the regulatory domains revealed less difference, 0.6 Å on the superimposed pair of domains *vs* 1.5 Å for the other. Departure from intradimer symmetry is thus largely due to substantial flexibility in the region that joins the regulatory and the DNA-binding domains, not the hinge loop covalently linking both domains, but rather the whole region involving the HTH as a rigid domain plus the lower part of the regulatory domain's helices that interact with the HTH (mainly the N-terminal half of α4). Such type of "pendular" flexibility (Fig. 2B,C) is further confirmed by higher atomic displacement parameters and weak electron density in the α6- α7 loop as well as in the N-terminal half of helix α7, features that were not apparent in $FasR_{\Delta 33}$- C_{20}-CoA that displayed more rigidity including in the HTH domains.

The dimerization interface is an extremely well conserved structural feature among the entire TFR (11), always involving a helical bundle constituted by equivalent helices, α8 and α9 from each protomer according to the FasR helix numbering scheme. The dimers of $FasR_{\Delta 33}$-C_{14} and $FasR_{\Delta 33}$-C_{20}-CoA were thus superimposed maximizing the fit of the dimerization helix bundles. In this way, clear differences between both structures were unveiled (Fig. 2C). In $FasR_{\Delta 33}$-C_{20}-CoA, the dimer is opened up, with both protomers separating away from each other, compared with $FasR_{\Delta 33}$-C_{14}. Using the position of conserved Tyr_{77} Cα as a reference (at the centre of helix α3, key in mediating DNA-protein binding), the two HTH domains in $FasR_{\Delta 33}$-C_{14} are ~48 Å apart, *vs* ~57 Å in $FasR_{\Delta 33}$-C_{20}-CoA (Fig. 2C). Taking the above observations together, we hypothesize that binding of a ligand, long enough to fill up the entire hydrophobic tunnel, stabilizes the dimer in a more rigid configuration, fixing an 'open' HTH geometry that is not compatible with DNA-binding. In contrast, when the tunnel is occupied with a too short acyl chain, a greater mobility of the DNA-binding domains is enabled, eventually allowing the HTH helices α3 to accommodate to the required separation distance and bind DNA efficiently (ideal B-DNA has 34 Å separation between major grooves on the same side of the double-helix). This flexibility cannot be described as a simple rigid-body movement of one domain with respect to the other (a hinge-like interdomain flexibility), but rather a pendular movement implicating the DNA-binding domain and part of the effector-binding domains, as if they were acting in a consolidated way.

Structural bases of FasR acyl-CoA-binding and effect on DNA association

To test the structure-derived hypotheses about acyl-binding and its effect on DNA-association, point-mutants were designed to block the entrance of the ligand into the hydrophobic tunnel, selecting residues that appeared to play key roles based on crystallographic evidence. Such 'tunnel blocking' approach was expected to abolish ligand-triggered rigidification of the protein and consequent DNA-

binding hindrance. A single mutant ($FasR_{L106F}$) and a triple mutant ($FasR_{LVL}$) were thus constructed. The former substitutes, Leu_{106} by a phenylalanine, at a critical position that borders the entrance of the tunnel. The triple mutant adds bulky side chains not only on position 106, but also substituting Leu_{185} and Val_{163} by phenylalanines (Fig. 3A). Point-mutations did not affect the dimeric architecture of the single mutant $FasR_{L106F}$ (Supplementary Fig. S3), and while slightly less than 50% of the triple mutant $FasR_{LVL}$ eluted as a monomer, >50% behaved as the wild-type protein.

Electrophoretic mobility shift assays (EMSAs) were performed by pre-incubating FasR, $FasR_{L106F}$ and $FasR_{LVL}$ (its dimeric form) with C_{16}-CoA and C_{20}-CoA, and then incubating these reactions with the ^{32}P-labelled P*fas* probe (P*fas*$_{MT}$). Strongly supporting our hypothesis, the acyl-CoA ligands triggered barely, or no detectable DNA-dissociation in the case of $FasR_{L106F}$ and $FasR_{LVL}$ mutants, while clearly inhibiting DNA-binding of native FasR (Fig. 3B). FasR:DNA dissociation constants ($^{app}K_D$) were not significantly affected by the point mutations (Supplementary Fig. S4A) and although direct monitoring of acyl-CoA binding to $FasR_{L106F}$ and $FasR_{LVL}$ mutants was not feasible due to technical impediments, dose-response analyses by EMSA (Supplementary Fig. S4B) further confirm that the ligands likely do not bind to the tunnel-occluding mutants. While $FasR_{wt}$ displayed nanomolar range half maximal inhibitory concentrations (IC_{50} = 368 nM) and apparent inhibition constants ($^{app}K_i$ = 42 nM) of C20-CoA, the mutants exhibited significantly increased values. Namely, the $FasR_{L106F}$ mutant only showed dissociation of the protein-DNA complex at the highest concentrations of acyl-CoA used (> 4 μM), while $FasR_{LVL}$ remained bound to the DNA probe even at 6 μM of C20-CoA (Supplementary Fig. S4B). Ligand entrance within the tunnel is thus needed in order to trigger the protein conformational change that precludes binding of the regulator to its cognate DNA site.

Signal transmission: allosteric mechanism connecting the ligand-binding pocket with the DNA-binding domain.

The differential shift of the bottom-half of helices α4 and α7, comparing $FasR_{\Delta 33}$-C_{14} and $FasR_{\Delta 33}$-C_{20}-CoA structures (Fig. 2C), strongly suggested that long enough alkyl chains in the effector-binding tunnel are critical in triggering the rigidification rearrangement, which results in arm-opening. To understand the molecular bases for such effect, the protein region around the distal tip of the ligand acyl chains were analysed in detail. The very last carbon atoms of C_{20}-CoA interact with mostly bulky hydrophobic residues towards the end of the tunnel, *i.e.* at the opening of the tunnel that leads to the space separating both protomers in the dimer. Among these hydrophobic residues Leu_{98} (on helix α4), Phe_{123} (on α5) and Phe_{138} (on α6) might play relevant roles (Fig. 3C). The substitution of such voluminous hydrophobic residues by smaller alanine sidechains, could uncouple the HTH mobility-restraining effect from ligand-binding. Two point-mutants were constructed, $FasR_{L98A}$ and $FasR_{F123A}$, which maintained the dimeric architecture of the protein (Supplementary Fig. S2). Both mutants showed significant functional effects, uncoupling ligand-binding and DNA-association (Fig. 3D), with a clearer effect observed in the case of $FasR_{F123A}$. To further dissect the underlying mechanisms, the DNA-binding affinities were first analysed, comparing wild-type *vs* $FasR_{L98A}$ and $FasR_{F123A}$ mutants (Supplementary Fig. S5). Apparent dissociation constants ($^{app}K_D$) were quantitated from electrophoretic mobility shift data, all in the nanomolar range (Table 2). Compared to $FasR_{wt}$, $FasR_{L98A}$ displayed slightly lower affinity for the P*fas*$_{MT}$ probe and $FasR_{F123A}$ higher, but neither one showed

considerable effect. According to the hypothesis that these point mutations would however affect efficient transmission of the signal from effector-binding to DNA-binding domains, we next calculated IC_{50} and $^{app}K_i$ values for the three proteins by evaluating dose-response effects of increasing concentrations of C_{20}-CoA on protein:DNA association (Supplementary Fig. S6). Indeed, both mutants had significantly lower response to the ligand (Table 2), with a particularly pronounced effect observed for $FasR_{F123A}$. The length of the acyl chain was also critical, with IC_{50}s and $^{app}K_i$s all shifted to significantly higher values when C_{16}-CoA was used (Supplementary Fig. S7 and Table 2). These results strongly suggest that residues Leu_{98} and Phe_{123}, and especially the latter, are key to ensure allosteric signal transmission while not influencing DNA-binding. That these mutations uncouple signal transmission but do not alter the protein's affinity for the effector ligand was assessed for the $FasR_{F123A}$ mutant (Supplementary Fig. S8). Surface plasmon resonance showed comparable association kinetics of $FasR_{F123A}$ to C_{20}-CoA as compared to $FasR_{wt}$. In sum, specific residues that are not essential to bind acyl ligands nor DNA, play a key role in transmitting the signal between both domains of the protein once the effector-binding tunnel is fully occupied.

The singular tunnel in FasR is predicted to accommodate very long fatty acyl effectors, triggering allosteric rigidification.

Very long fatty acyl moieties are relevant in the biology of Mycobacteriaceae including *Mtb* (52). Intermediates in the synthesis, and constitutive moieties, of mycolic acids (with acyl components that can reach 60-90 carbons), very long fatty acids are essential components of mycobacterial cell walls. FasR activates the expression of the *fas* gene, which encodes the synthase FAS I implicated in long- and very long-chain fatty acid biosynthesis. The question is how can such long alkyl chains act as effectors of FasR? A fully extended C_{20} acyl chain measures 26.5 Å, within the FasR-C_{20}-CoA tunnel the acyl shows some bending reducing that length to ~20 Å. The latter magnitude is enough for the C_{20} chain to fully occupy the tunnel, its distal tip located immediately beside the bottom opening that leads to the space between protomers in the dimer. We predict that longer acyl chains will accommodate, by extending the additional carbon atoms into the interprotomer space. To provide support for this scenario, classical molecular dynamics trajectories were calculated (Supplementary Video 1) starting from our FasR-C_{20}-CoA structure where the C_{20}-CoA was substituted *in silico* by the ~8 Å longer C_{26}-CoA (cerotic acyl-CoA), a particularly important biosynthesis intermediate synthesized by FAS I to provide for the α-alkyl chain of mycolic acids. After initial energy minimization, 10 ns all-atom trajectories were simulated with explicit solvent, showing that the cerotic acyl chains are stable within the tunnel, their 6-carbon extensions protruding into the interprotomer space and displaying extreme flexibility (Fig. 4). The available volume in the open form of acyl-bound FasR anticipates even longer acyl chains to be readily accommodated, considering that the protein shows very stable behaviour with ~2 Å average rmsd.

The structure of effector-free FasR in complex with DNA confirms the acyl-CoA-triggered allosteric mechanism.

The crystal structure of full-length FasR was eventually solved by co-crystallisation with a 25-bp double stranded oligonucleotide bearing the native FasR-binding sequence motif (Fig. 5). A number of

crystals and cryo-protections methods were tested, consistently producing strongly anisotropic X-ray diffraction data, reaching 3.85 Å resolution in the best direction (Table 1). A form of crystal disorder affected the position of the DNA double helix (details in Methods), occupying equivalent positions in different unit cells while sitting alternatively in both 5'→3' directions according to a crystallographic 2-fold.

The protein that was used to grow the FasR-DNA crystals corresponds to full-length FasR. However, the first ~30-35 amino acids at the N-termini are not visible in electron density, indicating they are not bound to the DNA, likely due to high flexibility. Neighbouring FasR dimers related by crystallographic symmetry, are observed bound to the same DNA fragment on a juxtaposed consecutive manner, at roughly 90º one from the other. Such organization is associated to a pronounced bending of the DNA molecule. The position of the major kink is similar to the one identified in other TFRs, inducing a very similar DNA bending angle as in TetR (53), or yet similar in magnitude but inverted compared to the archaeal FadR (54) due to the significantly shifted positions of the HTH domains relative to the dimeric effector-binding core (Supplementary Fig. S9). The limited resolution of the FasR-DNA structure, and the presence of crystal disorder affecting the occupancy of the DNA molecules, precluded detailed analyses of protein:DNA interactions. However, the structure did provide two accurate pieces of evidence: i- the absence of ligand bound within the tunnel of the effector-binding domain; and, ii- a major conformational rearrangement bringing the HTH DNA-binding domains closer together in the dimer (Supplementary Video 2) such that helices α3 now fit within two successive major grooves on the DNA molecule.

This conformational change hampers effector occupation within the tunnel, the latter seems to be constricted by shifted residues (e.g. Phe_{123}) which move their side chains towards the tunnel's lumen. The $FasR_{\Delta 33}$-C_{14} structure described above proved that acyl-containing compounds from *E. coli* are invariably associated within the effector-binding cavity, even if specific acyl/acyl-CoA molecules are not added during protein purification and crystallisation. That the FasR tunnel in the FasR-DNA complex is free of bound ligands is supported by unequivocal evidence from difference Fourier maps at both early and late stages of refinement (detailed in Methods). The association to DNA thus correlates to expelling ligands from the effector-binding tunnel of FasR, at least those that attach more loosely within the cavity.

DISCUSSION

FasR is a TFR (**T**etR **f**amily of **r**egulators) member that senses long and very long acyl-CoA moieties, subsequently turning off FAS I-mediated fatty acid biosynthesis. *Mtb* is able to synthesize very long acyl-CoA intermediates (e.g. in the way to synthesizing mycolic acids), spanning molecular lengths of 40-50 Å and more. It is known that TFRs possess pockets, sometimes even deeper tunnel-like cavities, which bind to, and enclose the sensed effector ligand. How can FasR deal with the very long effector molecules it senses? It must be stressed that the entire effector-binding domain of FasR measures ~40 Å along its longest axis (substantially less considering inner cavities); hence, the fatty acyl effectors can often be longer than the protein's own physical boundaries. We now answer to this question by revealing a unique hydrophobic tunnel that cuts across the entire effector-binding domain

of FasR (Fig. 1), a tunnel that is conspicuously opened on both ends. This singular solution has evolved to lodge the kind of acylated chains of 20 and more carbons that inhibit FasR binding to its cognate DNA (7).

In addition to the tunnel itself, and in continuity with the bottom opening of it, FasR possesses a cavity delimited by the two protomers. The volume of such cavity (55) in FasR is unusually large (~1800 $Å^3$) compared to many other TFRs: ~600 $Å^3$ (*S. enterica* RamR PDB 3VVY; and *M. tuberculosis* EthR PDB 5NIO), ~530 $Å^3$ (*E. coli* RutR PDB 4XK4), ~340 $Å^3$ (*E. coli* TetR PDB 2XPW), ~135 $Å^3$ (*S. acidocaldarius* FadR PDB 6EL2), or yet 40 $Å^3$ (*P. aeruginosa* DesT PDB 3LSJ). This feature is likely relevant, as it anticipates FasR's ability to accommodate very long fatty acyl chains, such as C_{26}-CoA (synthesized by the *Mtb* FAS I system (56)), while maintaining a stable, open configuration (Fig. 4, Supplementary Video 1). Other fatty acid-sensory TFRs fairly similar to FasR (54,57-59), either do not create true continuous tunnels (58), or engage a different set of residues running in a perpendicular direction as compared to FasR's cavity (54,57,59). EthR is yet another TFR from *Mtb* (60), intensively investigated as a target to develop anti-tuberculosis medicines. However, EthR is substantially different from FasR (22 % sequence identity; ~4.5 Å rmsd after superposition of the effector-binding domains), with a shorter N-terminal extension before the first α-helix, and a tunnel displaying wider regions or bulges (Supplementary Fig. S10). Such bulges seem to correlate with EthR's capacity to bind compounds that include 1 to 3 aromatic or aliphatic rings (61). At difference with FasR, the physiologic molecules sensed by EthR remain unidentified, despite the >70 EthR crystal structures available, most in complex with surrogate ligands. Among these, only one corresponds to a linear chain (hexadecyl octanoate, in PDBs 1U9N and 1U9O), in this way the most similar to FasR effectors. This ligand is positioned in EthR in a similar configuration as the acyl ligands in FasR, but leaving part of EthR's available tunnel volume unoccupied (Supplementary Fig. S10C), suggesting that physiologic effectors are likely to be of larger size, branched and/or containing bulkier cyclic groups.

FasR is thus equipped to binding very long-chain acyl effectors, but how is such binding coupled to inhibiting DNA-association? The type of rearrangements that we have found (Supplementary Video 2) are consistent with the ones observed in a number of other TFRs, in principle conforming to the mechanistic hypothesis that the binding of effector ligand (the signal) induces an HTH-open conformation, eventually inhibiting TFR association to DNA (the output response). Association of TFRs to DNA indeed requires the HTH domains to close in, in order for the α3 helices of the two protomers to fit into two successive major grooves (~34 Å apart) on the same side of the cognate DNA (62). Such a mechanism has been put forward to explain the workings of *E. coli* TetR when binding tetracycline (63), DesT from *Pseudomonas aeruginosa* sensing saturated *vs* unsaturated acyl-CoAs (58), RutR recognizing uracil in *E. coli* (64) or yet the multi-drug binding protein QacR from *Staphylococcus aureus* (65), among many others.

Binding the effector stabilizes an open configuration of FasR, which does not necessarily imply that the effector mechanically triggers a closed to open transition. To ascertain the latter mechanism, the ligand-free structure should exhibit a closed, DNA-binding competent configuration. A ligand-free form of FasR could not be crystallized, but turning our attention to the many other available TFR

structures, those that exhibit no ligand almost exclusively correspond to the open form (9,11), contradicting the predicted outcome. The very few apo structures with a closed configuration have unexplained density within the binding pocket (pdb 2FX0) and/or reveal crystal packings that fortuitously fix HTH domains strongly in place (e.g. the hexagonal form of 2FX0 and the centred monoclinic 1T33). To the best of our knowledge, the closed configuration has only been captured reliably in crystal structures of TFRs in complex with DNA. Additional evidence further challenge a simple open/closure mechanism: *i*- no simple positional shifts of individual residues can explain the mechanical bases of the alleged pendular movement; *ii*- a number of TFR mutants have been identified that either uncouple effector-binding from transcriptional induction (66), or invert the effector's action by triggering a tighter binding to DNA (67,68), in both cases often implicating positions that are not directly involved in effector-binding.

Finite deformations physics theory seems attractive to highlight allosteric regulation pathways by measuring mechanical strain rather than pairwise atomic position deviations (69). Unexpectedly, one of the segments subjected to highest mechanical strain in all TFRs analysed, corresponds to the loop that connects helices α6 and α7 (Supplementary Fig. S11). A triangle defined by α5, α6 and α7, a conserved feature in all TFRs (9), harbours the ligand-binding core cavity (which can in some cases expand into tunnel-like architectures with top, bottom and/or lateral openings). Helix α6, associating to helix α8, is attached to the fixed core, upper-half of the effector-binding domain; but simultaneously, α6's C-terminal tip and the α6-α7 junction also associate to the moving HTH. In sum, α6-α7 loop is bound to fixed and to moving parts, eventually leading to local deformation. Residues that could explain this strain, in contact with helices α6, the α6-α7 loop and the HTH domain, led us to identifying an array of hydrophobic residues, highly conserved among TFRs (Supplementary Fig. S1 and Supplementary Data 1) and configured in three-dimensions as a continuous spine connecting the two domains of FasR. This spine belongs to, and connects the hydrophobic protein-folding cores of both domains, being interrupted by the ligand-binding cavity in all TFRs analysed (Fig. 6 and Supplementary Fig. S12). Only in the ligand-bound condition this hydrophobic spine is completed, by the ligand molecule itself at the effector-binding domain, stabilizing a rigid and open conformation. Such a mechanism predicts a disordered (flexible) to ordered transition of the TFR protein, which is consistent with the evidence we provide for FasR as well as with available evidence from other TFRs (9,11,66-68,70). In particular, fluorescent probes such as 1-anilino-naphthalene-8-sulfonate that bind to partially folded proteins in "molten globule" states (71), have been shown to bind promiscuously to apo TFRs (9). Also, effector-triggered appearance of folding cooperativity between both domains, as well as proteolysis-resistance, have been reported in wild-type but not in allosteric-uncoupled mutants of TetR (70,72). Taken together, the extensive body of evidence lends strong support to the transmission spine mechanism as the most consistent interpretation of the effector-mediated allosteric control of TFRs' DNA-binding function.

When the ligand leaves the site, or if it is too short to fully occupy it, the hydrophobic spine is broken, protein folding is sub-optimal, anticipating a multitude of conformations (with HTH wiggling, illustrated schematically in Fig. 7), including those that are competent for DNA-binding (HTH-closed). The disorder-to-order transition is not fully triggered, resulting in asymmetry and higher flexibility as

observed in the $FasR_{\Delta 33}$-C_{14} complex. This hypothesis also explains why mutating bulky residues that contribute to building and stabilising the spine (*e.g.* Leu_{98} and Phe_{123} in FasR), can uncouple the allosteric effect: the ligand is then insufficient to achieve a complete, compact fold in the mutated TFR (Supplementary Fig. S13). Residues that are not directly involved in effector-binding, but that contribute to building and stabilising the spine, will also be able to exert notable effects on allosteric coupling, upholding reported results (66-68).

A second conformation, the one bound to DNA, is however compactly folded. In this case it is the polynucleotide that pulls on the flexible HTH domains of the dimer, bringing them closer together. Correlated to this HTH movement and again transmitted through the hydrophobic spine, several of the bulky residues that line up the effector-binding cavity walls, occlude the cavity (such as Phe_{123} in FasR-DNA), completing the spine into a "folded protein-like" core. This occlusion of the ligand-binding pocket in other TFRs when bound to DNA has been described in crystal structures of TetR (65), DesT (58) and FadR (59).

This scenario opens up exciting avenues to be explored. How reversible is the binding of effector compounds and what triggers their dislodging from the cavity? Following on our proposal, DNA-binding should likely expel the effector and vice-versa, depending on relative DNA-vs-effector concentrations and affinities. Of note, the assistance of other proteins, in burying the long alkyl moieties when expelled from FasR and other lipid-sensing TFRs, should not be ruled out, as they might be extremely important as part of the thermodynamic equations. The hydrophobic transmission spine hypothesis (Fig. 7) may prove instrumental to design better drugs. Eukaryotic protein kinases (ePKs) are not homologous to TFRs, but do exhibit an analogous hydrophobic spine that regulates the kinase activation switch (73), a mechanism that has been successfully exploited to develop ePK inhibitor-based drugs against human cancer and inflammatory diseases (74,75). A closer example is provided by EthR, a TetR-like repressor that controls the expression of EthA (the enzyme that bioactivates ethionamide, a second-line antituberculosis drug that inhibits mycolic acid biosynthesis). EthR inhibitors have been conceived after comparing EthR crystal structures obtained in complex with larger and smaller ligands (76). In that work, compounds bearing thienyl and piperidinyl pharmacophores were selected as the most potent EthR inhibitors among screening hits. Only the piperidinyl-, and not the thienyl-, interacting pocket engages EthR amino acids that compose the hydrophobic spine (PDBs 3G1O, 3G1M), the transmission mechanism thus explains why the piperidinyl-binding pocket turned out to be the crucial region determining inhibitory activities (76). Structure-guided drug discovery strategies that exploit the allosteric hydrophobic-spine transmission mechanism of TFRs might thus prove successful in developing novel medicines against tuberculosis, including multi- and extensively drug-resistant strains.

ACCESSION NUMBERS

Macromolecular 3D structural data (model coordinates and crystallographic structure factors) presented in this study have been deposited in the wwPDB with accession codes 6O6O, 6O6N and

6O6P. Raw X-ray diffraction data for each one of those structures have been deposited in SBGrid Data Bank with Digital Object Identifiers 10.15785/SBGRID/648, 10.15785/SBGRID/647 and 10.15785/SBGRID/649, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available as a separate PDF file.

ACKNOWLEDGEMENTS

We thank Stanislas Leibler, Michael Mitchell and Pablo Sartori for sharing initial strain analysis scripts; Matias Machado for assistance in molecular dynamics; Frank Lehmann for initial cloning efforts; Sebastian Klinke (Fundación Leloir) and the staff at Proxima 1 beamline (Soleil synchrotron) and at I04-1 beamline (Diamond synchrotron) for assistance with data collection. We acknowledge computational and storage services (TARS cluster) provided by the Institut Pasteur IT Dept (Paris). We thank the CCP4/CeBEM Macromolecular Crystallography School (USP@São Carlos, 2018), especially Isabel Usón and Paul Emsley for helping us respectively with ShelxE and Coot, in dealing with low-resolution density modification and model building.

JL traineeships at Institut Pasteur de Montevideo were funded by the Centro de Biología Estructural del Mercosur CeBEM (www.cebem-lat.org). Support to HG from Agencia Nacional de Promoción Científica y Tecnológica ANPCyT Argentina (grants PICT 2012-0168, PICT 2015-0796 and 2022) and National Institutes of Health NIH USA (grant 1R01AI095183-01) is acknowledged.

REFERENCES

1. Daffé, M. and Draper, P. (1997) In Poole, R. K. (ed.), *Advances in Microbial Physiology*. Academic Press, Vol. 39, pp. 131-203.

2. Bhatt, A., Molle, V., Besra, G.S., Jacobs, W.R., Jr. and Kremer, L. (2007) The *Mycobacterium tuberculosis* FAS-II condensing enzymes: their role in mycolic acid biosynthesis, acid-fastness, pathogenesis and in future drug development. *Mol Microbiol*, **64**, 1442-1454.

3. Marrakchi, H., Laneelle, M.A. and Daffe, M. (2014) Mycolic acids: structures, biosynthesis, and beyond. *Chem Biol*, **21**, 67-85.

4. Bazet Lyonnet, B., Diacovich, L., Gago, G., Spina, L., Bardou, F., Lemassu, A., Quemard, A. and Gramajo, H. (2017) Functional reconstitution of the *Mycobacterium tuberculosis* long-chain acyl-CoA carboxylase from multiple acyl-CoA subunits. *FEBS J*, **284**, 1110-1125.

5. Daniel, J., Deb, C., Dubey, V.S., Sirakova, T.D., Abomoelak, B., Morbidoni, H.R. and Kolattukudy, P.E. (2004) Induction of a novel class of diacylglycerol acyltransferases and

triacylglycerol accumulation in *Mycobacterium tuberculosis* as it goes into a dormancy-like state in culture. *J Bacteriol*, **186**, 5017-5030.

6. Deb, C., Lee, C.M., Dubey, V.S., Daniel, J., Abomoelak, B., Sirakova, T.D., Pawar, S., Rogers, L. and Kolattukudy, P.E. (2009) A novel in vitro multiple-stress dormancy model for *Mycobacterium tuberculosis* generates a lipid-loaded, drug-tolerant, dormant pathogen. *PLoS One*, **4**, e6077.

7. Mondino, S., Gago, G. and Gramajo, H. (2013) Transcriptional regulation of fatty acid biosynthesis in mycobacteria. *Mol Microbiol*, **89**, 372-387.

8. Ulrich, L.E., Koonin, E.V. and Zhulin, I.B. (2005) One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol*, **13**, 52-56.

9. Yu, Z., Reichheld, S.E., Savchenko, A., Parkinson, J. and Davidson, A.R. (2010) A comprehensive analysis of structural and sequence conservation in the TetR family transcriptional regulators. *J Mol Biol*, **400**, 847-864.

10. Rozewicki, J., Li, S., Amada, K.M., Standley, D.M. and Katoh, K. (2019) MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res*, **47**, W5-W10.

11. Cuthbertson, L. and Nodwell, J.R. (2013) The TetR family of regulators. *Microbiol Mol Biol Rev*, **77**, 440-475.

12. Orth, P., Schnappinger, D., Hillen, W., Saenger, W. and Hinrichs, W. (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat Struct Biol*, **7**, 215-219.

13. Zhang, Y.M. and Rock, C.O. (2009) Transcriptional regulation in bacterial membrane lipid synthesis. *J Lipid Res*, **50 Suppl**, S115-119.

14. Chim, N., Habel, J.E., Johnston, J.M., Krieger, I., Miallau, L., Sankaranarayanan, R., Morse, R.P., Bruning, J., Swanson, S., Kim, H. *et al.* (2011) The TB Structural Genomics Consortium: a decade of progress. *Tuberculosis (Edinb)*, **91**, 155-172.

15. Ioerger, T.R. and Sacchettini, J.C. (2009) Structural genomics approach to drug discovery for Mycobacterium tuberculosis. *Curr Opin Microbiol*, **12**, 318-325.

16. Heffler, M.A., Walters, R.D. and Kugel, J.F. (2012) Using electrophoretic mobility shift assays to measure equilibrium dissociation constants: GAL4-p53 binding DNA as a model system. *Biochem Mol Biol Educ*, **40**, 383-387.

17. Clingman, C.C., Deveau, L.M., Hay, S.A., Genga, R.M., Shandilya, S.M., Massi, F. and Ryder, S.P. (2014) Allosteric inhibition of a stem cell RNA-binding protein by an intermediary metabolite. *eLife*, **3**, e02848.

18. Lin, S.Y. and Riggs, A.D. (1972) Lac repressor binding to non-operator DNA: detailed studies and a comparison of eequilibrium and rate competition methods. *J Mol Biol*, **72**, 671-690.

19. Teng, T.-Y. (1990) Mounting of crystals for macromolecular crystallography in a free-standing thin film. *J Appl Crystallogr* **23**, 387-391.

20. Ascone, I., Girard, E., Gourhant, P., Legrand, P., Roudenko, O., Roussier, L. and Thompson, A.W. (2007) Proxima 1, a New Beamline on the Third Generation SR Source SOLEIL Combining PX and Single-Crystal BioXAS. *AIP Conf Proc* **882**, 872-874.

21. Broennimann, C., Eikenberry, E.F., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze-Briese, C., Suzuki, M., Tomizaki, T. *et al.* (2006) The PILATUS 1M detector. *J Synchrotron Radiat*, **13**, 120-130.

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

23. Evans, P.R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr*, **67**, 282-292.

24. Propper, K., Meindl, K., Sammito, M., Dittrich, B., Sheldrick, G.M., Pohl, E. and Uson, I. (2014) Structure solution of DNA-binding proteins and complexes with ARCIMBOLDO libraries. *Acta Crystallogr D Biol Crystallogr*, **70**, 1743-1757.

25. Rodríguez, D., Sammito, M., Meindl, K., de Ilarduya, I.M., Potratz, M., Sheldrick, G.M. and Usón, I. (2012) Practical structure solution with ARCIMBOLDO. *Acta Crystallogr D Biol Crystallogr* **68**, 336-343.

26. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674.

27. Uson, I. and Sheldrick, G.M. (2018) An introduction to experimental phasing of macromolecules illustrated by SHELX; new autotracing features. *Acta Crystallogr D Biol Crystallogr* **74**, 106-116.

28. Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., S., S.O., Vonrhein, C. *et al.* (2017) v2.10.3 Global Phasing Ltd. United Kingdom, Cambridge.

29. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, **66**, 486-501.

30. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B. *et al.* (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci*, **27**, 293-315.

31. Liebschner, D., Afonine, P.V., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C. and Adams, P.D. (2017) Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr D Struct Biol*, **73**, 148-157.

32. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkoczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J. *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol*, **75**, 861-877.

33. Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H. and Adams, P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr*, **68**, 352-367.

34. Brown, A., Long, F., Nicholls, R.A., Toots, J., Emsley, P. and Murshudov, G. (2015) Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr D Biol Crystallogr*, **71**, 136-153.

35. Lebedev, A.A. and Isupov, M.N. (2014) Space-group and origin ambiguity in macromolecular structures with pseudo-symmetry and its treatment with the program Zanuda. *Acta Crystallogr D Biol Crystallogr*, **70**, 2430-2443.

36. Afonine, P.V., Headd, J.J., Terwilliger, T.C. and Adams, P.D. (2013) New tool: phenix.real_space_refine. *Computational Crystallography Newsletter* **4**, 43-44.

37. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*, **67**, 235-242.

38. Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, **372**, 774-797.

39. Schrodinger, LLC. (2015) The PyMOL Molecular Graphics System, Version 2.1.0.

40. Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr*, **60**, 2256-2268.

41. Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*, **302**, 205-217.

42. Finn, R.D., Clements, J., Arndt, W., Miller, B.L., Wheeler, T.J., Schreiber, F., Bateman, A. and Eddy, S.R. (2015) HMMER web server: 2015 update. *Nucleic Acids Res*, **43**, W30-38.

43. Huang, Y., Niu, B., Gao, Y., Fu, L. and Li, W. (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, **26**, 680-682.

44. Wallace, I.M., O'Sullivan, O., Higgins, D.G. and Notredame, C. (2006) M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res*, **34**, 1692-1699.

45. Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, **157**, 105-132.

46. Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*, **42**, W320-W324.

47. Landrum, G. (2006).

48. Das, R. and Baker, D. (2008) Macromolecular modeling with rosetta. *Annu Rev Biochem*, **77**, 363-382.

49. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1-2**, 19-25.

50. Schmit, J.D., Kariyawasam, N.L., Needham, V. and Smith, P.E. (2018) SLTCAP: a simple method for calculating the number of ions needed for MD simulation. *Journal of chemical theory and computation*, **14**, 1823-1827.

51. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *Journal of molecular graphics*, **14**, 33-38, 27-38.

52. Gago, G., Diacovich, L., Arabolaza, A., Tsai, S.C. and Gramajo, H. (2011) Fatty acid biosynthesis in actinomycetes. *FEMS Microbiol Rev*, **35**, 475-497.

53. Le, T.B., Schumacher, M.A., Lawson, D.M., Brennan, R.G. and Buttner, M.J. (2011) The crystal structure of the TetR family transcriptional repressor SimR bound to DNA and the role of a flexible N-terminal extension in minor groove binding. *Nucleic Acids Res*, **39**, 9433-9447.

54. Wang, K., Sybers, D., Maklad, H.R., Lemmens, L., Lewyllie, C., Zhou, X., Schult, F., Brasen, C., Siebers, B., Valegard, K. *et al.* (2019) A TetR-family transcription factor regulates fatty acid metabolism in the archaeal model organism *Sulfolobus acidocaldarius*. *Nat Commun*, **10**, 1542.

55. Tian, W., Chen, C., Lei, X., Zhao, J. and Liang, J. (2018) CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res*, **46**, W363-W367.

56. Brennan, P.J. (2003) Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. *Tuberculosis*, **83**, 91-97.

57. Fujihashi, M., Nakatani, T., Hirooka, K., Matsuoka, H., Fujita, Y. and Miki, K. (2014) Structural characterization of a ligand-bound form of *Bacillus subtilis* FadR involved in the regulation of fatty acid degradation. *Proteins*, **82**, 1301-1310.

58. Miller, D.J., Zhang, Y.M., Subramanian, C., Rock, C.O. and White, S.W. (2010) Structural basis for the transcriptional regulation of membrane lipid homeostasis. *Nat Struct Mol Biol*, **17**, 971-975.

59. Yeo, H.K., Park, Y.W. and Lee, J.Y. (2017) Structural basis of operator sites recognition and effector binding in the TetR family transcription regulator FadR. *Nucleic Acids Res*, **45**, 4244-4254.

60. Frenois, F., Engohang-Ndong, J., Locht, C., Baulard, A.R. and Villeret, V. (2004) Structure of EthR in a ligand bound conformation reveals therapeutic perspectives against tuberculosis. *Mol Cell*, **16**, 301-307.

61. Tanina, A., Wohlkonig, A., Soror, S.H., Flipo, M., Villemagne, B., Prevet, H., Deprez, B., Moune, M., Peree, H., Meyer, F. *et al.* (2019) A comprehensive analysis of the protein-ligand interactions in crystal structures of Mycobacterium tuberculosis EthR. *Biochim Biophys Acta Proteins Proteom*, **1867**, 248-258.

62. Ramos, J.L., Martinez-Bueno, M., Molina-Henares, A.J., Teran, W., Watanabe, K., Zhang, X., Gallegos, M.T., Brennan, R. and Tobes, R. (2005) The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev*, **69**, 326-356.

63. Hinrichs, W., Kisker, C., Duvel, M., Muller, A., Tovar, K., Hillen, W. and Saenger, W. (1994) Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science*, **264**, 418-420.

64. Nguyen Le Minh, P., de Cima, S., Bervoets, I., Maes, D., Rubio, V. and Charlier, D. (2015) Ligand binding specificity of RutR, a member of the TetR family of transcription regulators in *Escherichia coli*. *FEBS Open Bio*, **5**, 76-84.

65. Schumacher, M.A., Miller, M.C., Grkovic, S., Brown, M.H., Skurray, R.A. and Brennan, R.G. (2001) Structural mechanisms of QacR induction and multidrug recognition. *Science*, **294**, 2158-2163.

66. Muller, G., Hecht, B., Helbl, V., Hinrichs, W., Saenger, W. and Hillen, W. (1995) Characterization of non-inducible Tet repressor mutants suggests conformational changes necessary for induction. *Nat Struct Biol*, **2**, 693-703.

67. Kamionka, A., Bogdanska-Urbaniak, J., Scholz, O. and Hillen, W. (2004) Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor. *Nucleic Acids Res*, **32**, 842-847.

68. Scholz, O., Henssler, E.M., Bail, J., Schubert, P., Bogdanska-Urbaniak, J., Sopp, S., Reich, M., Wisshak, S., Kostner, M., Bertram, R. *et al.* (2004) Activity reversal of Tet repressor caused by single amino acid exchanges. *Mol Microbiol*, **53**, 777-789.

69. Mitchell, M.R., Tlusty, T. and Leibler, S. (2016) Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings. *Proc Natl Acad Sci U S A*, **113**, E5847-E5855.

70. Reichheld, S.E., Yu, Z. and Davidson, A.R. (2009) The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor. *Proc Natl Acad Sci U S A*, **106**, 22263-22268.

71. Semisotnov, G.V., Rodionova, N.A., Razgulyaev, O.I., Uversky, V.N., Gripas, A.F. and Gilmanshin, R.I. (1991) Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers*, **31**, 119-128.

72. Reichheld, S.E. and Davidson, A.R. (2006) Two-way interdomain signal transduction in tetracycline repressor. *J Mol Biol*, **361**, 382-389.

73. Taylor, S.S., Meharena, H.S. and Kornev, A.P. (2019) Evolution of a dynamic molecular switch. *IUBMB life*, **71**, 672-684.

74. Hu, J., Ahuja, L.G., Meharena, H.S., Kannan, N., Kornev, A.P., Taylor, S.S. and Shaw, A.S. (2015) Kinase regulation by hydrophobic spine assembly in cancer. *Molecular and cellular biology*, **35**, 264-276.

75. Roskoski, R., Jr. (2019) Properties of FDA-approved small molecule protein kinase inhibitors. *Pharmacological research*, **144**, 19-50.

76. Willand, N., Dirie, B., Carette, X., Bifani, P., Singhal, A., Desroses, M., Leroux, F., Willery, E., Mathys, V., Deprez-Poulain, R. *et al.* (2009) Synthetic EthR inhibitors boost antituberculous activity of ethionamide. *Nat Med*, **15**, 537-544.

77. Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P.C., Timony, M.A., Ottaviano, M. and Sliz, P. (2013) Collaboration gets the most out of software. *eLife*, **2**, e01456.

TABLE AND FIGURES LEGENDS

Table 1. X-ray diffraction data collection and structure refinement statistics.

	$FasR_{\Delta 33}$-C_{14}	$FasR_{\Delta 33}$-C_{20}-CoA	FasR-DNA
Data collection			
X-ray source	In-house rotating Cu anode, Inst. Pasteur Montevideo	Proxima 1 beamline, SOLEIL synchrotron	I04-1 beamline, Diamond Light Source synchrotron
X-ray wavelength (Å)	1.54179	0.97857	0.97948
Space group	P 2_1 2_1 2	P 4_1 2_1 2	C 2 2 2_1
Cell parameters			
a, *b*, *c* (Å)	86.1 100.2 38.5	108.1 108.1 43.6	61.7 174.7 158.2
Resolution (Å)	28.8 - 1.63 (1.66 - 1.63)*	48.2 - 1.7 (1.8 - 1.7)	39.5 - 3.85 (4.31 - 3.85)
Unique reflections	41,807 (1,999)	28,970 (4,563)	6,467 (1,452)
R_{merge}	0.04 (0.47)	0.06 (1.47)	0.16 (1.36)
R_{meas}	0.05 (0.55)	0.07 (1.52)	0.18 (1.58)
$CC_{1/2}$	1.00 (0.80)	1.00 (0.71)	1.00 (0.62)
I / σI	11.3 (1.8)	22.6 (1.6)	8.3 (1.0)
Completeness (%)	98.6 (96.4)	99.8 (98.8)	76.5 (62.3)
Multiplicity	3.5 (3.5)	14.2 (13.9)	8.6 (5.6)
Refinement			
Resolution (Å)	28.78 - 1.63	20.2 - 1.70	39.07 - 3.85
No. of reflections used (No. in the free set)	41,762 (1,692)	27,645 (1,379)	6,431 (647)
R_{work} / R_{free}	0.219/0.237	0.182/0.210	0.287/0.331
Number of atoms:			
Protein	2891	1520	2860
DNA	--	--	943
Ligands	32 (myristate)	36 (34 arachinoyl-CoA;2 Cl^-)	--
Water	205	216	--
B-factors ($Å^2$)			
Wilson plot	29.2	22.8	111.2
Average (overall)	36.2	25.3	148.3
R.m.s. deviations			
Bond lengths (Å)	0.01	0.01	0.01
Bond angles (°)	0.90	0.89	1.22
Ramachandran analysis‡ :			
Favoured (%)	98.4	99.5	97.6

Allowed (%)	1.6	0.5	2.4
Outliers (%)	0.0	0.0	0.0
Real-space correlation coefficients (ligands)	0.83 , 0.88 (myristates A , B)	0.88 (arachinoyl-CoA)	--
PDB ID	6O6O	6O6N	6O6P
Raw diffraction data[¶] (doi)	10.15785/SBGRID/648	10.15785/SBGRID/647	10.15785/SBGRID/649

*Values in parentheses are for highest-resolution shell.

[‡] Calculated by MolProbity (30).

¶ Deposited in the SBGrid Data Bank public database (77).

Table 2. Quantitative parameters of FasR association to DNA, and of its dose-dependent inhibition by effector ligands. Comparison of $FasR_{wt}$ *vs* mutants $FasR_{L98A}$ and $FasR_{F123A}$ that uncouple the allosteric effect.

	$FasR_{wt}$	$FasR_{L98A}$	$FasR_{F123A}$
Binding to DNA (P*fas*$_{MT}$)			
$^{app}K_D$ (nM)	83 ± 11	155 ± 18	39 ± 6
C_{20}-CoA dose-response			
IC_{50} (nM)	368 ± 105	634 ± 100	5269 ± 1067
$^{app}K_i$ (nM)	42	152	571
C_{16}-CoA dose-response			
IC_{50} (nM)	5415 ±1047	ND*	ND
$^{app}K_i$ (nM)	589	ND	ND

$^{app}K_D$ and IC_{50} values are reported as the average of at least three independent experiments ± one standard error of the mean.

* ND = not detectable.

Figure 1. **Crystal structure of $FasR_{\Delta 33}$-C_{20}-CoA. A.** FasR dimer in cartoon representation, the two protomers are distinguished in light blue and pale green. The dimer is strictly symmetric, with one protomer in the asymmetric unit, the crystallographic 2-fold is depicted as a vertical axis in the plane of the figure. Secondary structure elements are labelled, a prime symbol denotes equivalent elements in the other protomer. The two domains are indicated, an N-terminal DNA-binding domain with a helix-turn-helix (HTH) motif, and a C-terminal domain with the co-crystallized effector molecule C_{20}-CoA bound (in spheres coloured by element). **B.** Details of the effector-binding tunnel. Three orthogonal views illustrate the molecular surface of $FasR_{\Delta 33}$.The leftmost cuts along the middle of a protomer, revealing the top tunnel mouth on the right and a large segment of the tunnel itself with the bound acyl moiety (in sticks). A few amino acid sidechains that define the tunnel walls are labelled. The rightmost panel is an open-book perspective, with the footprint of protomer A visible on protomer B's surface (in pale green), the two openings of the tunnel are visible, showing solvent exposure of both tips of the C_{20}-CoA molecule.

Figure 2. **The $FasR_{\Delta 33}$-C_{14} dimer is asymmetric, with larger flexibility and pendular motion of HTHs. A.** SigmaA-weighted mF_{obs}-DF_{calc} difference Fourier map contoured at 3.5 σ (green mesh) calculated with no ligand bound in the effector-binding tunnel during $FasR_{\Delta 33}$-C_{14} refinement. The $FasR_{\Delta 33}$-C_{14} model is shown with grey cartoons, and residues at ≤4 Å from the ligand are depicted with thin sticks. The electron density allowed to model a myristic acid, overlaid within the difference

map in thick sticks coloured by atom. Note polar and charged residues on the top opening, close to the carboxyl head of the ligand. Hydrophobic residues line up the rest of the tunnel's walls. **B.** $FasR_{\Delta 33}$-C_{14} protomers A and B are coloured as in Fig. 1. The dimer is asymmetric, quantitated by superimposing protomer B onto A (red arrow in the top panel), after what the effector-binding domains fit quite well whilst the HTH domains show significant rotation between one another (bottom panel), with max root mean square deviation (rmsd) on helices α2 and α3. **C.** The $FasR_{\Delta 33}$-C_{20}-CoA and $FasR_{\Delta 33}$-C_{14} structures are superposed, resulting in maximal fit on the central 4-helix bundle that mediates dimerization (helices α8-α9 from both protomers). The upper portion of the effector-binding domains end up well aligned, but significant shifts affect the lower part together and the HTH domains (dashed arrows). The distance between the centres of mass of Tyr_{77} on helix α3, differs by more than 10 Å between both structures. The HTH domains move consolidated with the bottom portion of juxtaposed helices (α4 and α7).

Figure 3. **Structure guided mutations in FasR leading to permanent DNA-binding. A**. Selected mutations to occlude the effector-binding tunnel are illustrated. The crystal structure of $FasR_{\Delta 33}$-C_{20}-CoA (left panel) is compared to a model of the triple mutant $FasR_{LVL}$ (right panel). Val_{163} is actually behind the tunnel (red arrow). The substitutions by bulky phenylalanines anticipate steric hindrance of the acyl chain (shown as sticks coloured by element). **B.** Electrophoretic mobility shift assay was performed by incubating the ^{32}P-labelled 398 bp P*fas* promoter region with wild-type FasR and mutant FasR proteins ($FasR_{LVL}$ and $FasR_{L106F}$) in the absence or in the presence of C_{20}-CoA and C_{16}-CoA. Protein–DNA complexes were separated by electrophoresis on a 6 % polyacrylamide gel. **C.** Selected mutations to uncouple allosteric signalling from the ligand-binding pocket to the DNA-binding domain. The bottom mouth of the effector-binding tunnel is located towards the centre of the molecule, revealing the tip of C_{20} (in ball and stick representation, coloured by atom). A few side chains are labelled and shown as orange sticks, chosen among residues bordering the tunnel opening. Two bulky residues (Leu_{98} on helix α4 and Phe_{123} on α5) that were substituted by alanines, are labelled with bold fonts. **D.** Gel shift assays of the wild-type and selected FasR mutants. The DNA probe that corresponds to the 398 bp *Pfas* promoter region ($Pfas_{MT}$) was ^{32}P-labelled and incubated either with FasR wild-type, $FasR_{L98A}$ or $FasR_{F123A}$, in the absence or in the presence of the indicated concentrations of C_{16}-CoA and C_{20}-CoA.

Figure 4. **Longer than C20 acyl chains can be accommodated within the $FasR_{\Delta 33}$ dimer.** A representative 10ns trajectory of all-atom molecular dynamics is illustrated by showing in cartoon representation the simulated model of $FasR_{\Delta 33}$ in complex with C_{26}-CoA every 250 ps (see Supplementary Video 1). Shown as superposed ensemble, the secondary structure elements of both domains are preserved, and the cerotic acyl chain (in sticks) also remains stable within the tunnel. The inset shows a projection of a cut section through protomer's B tunnel, to display the dynamic stability of the dimeric interface and the acyl chain (as opposed to the large flexibility of the nucleotide portion of coenzyme-A). Free volume is available at the inter-protomer space, predicting that even longer acyl moieties should be able to accommodate.

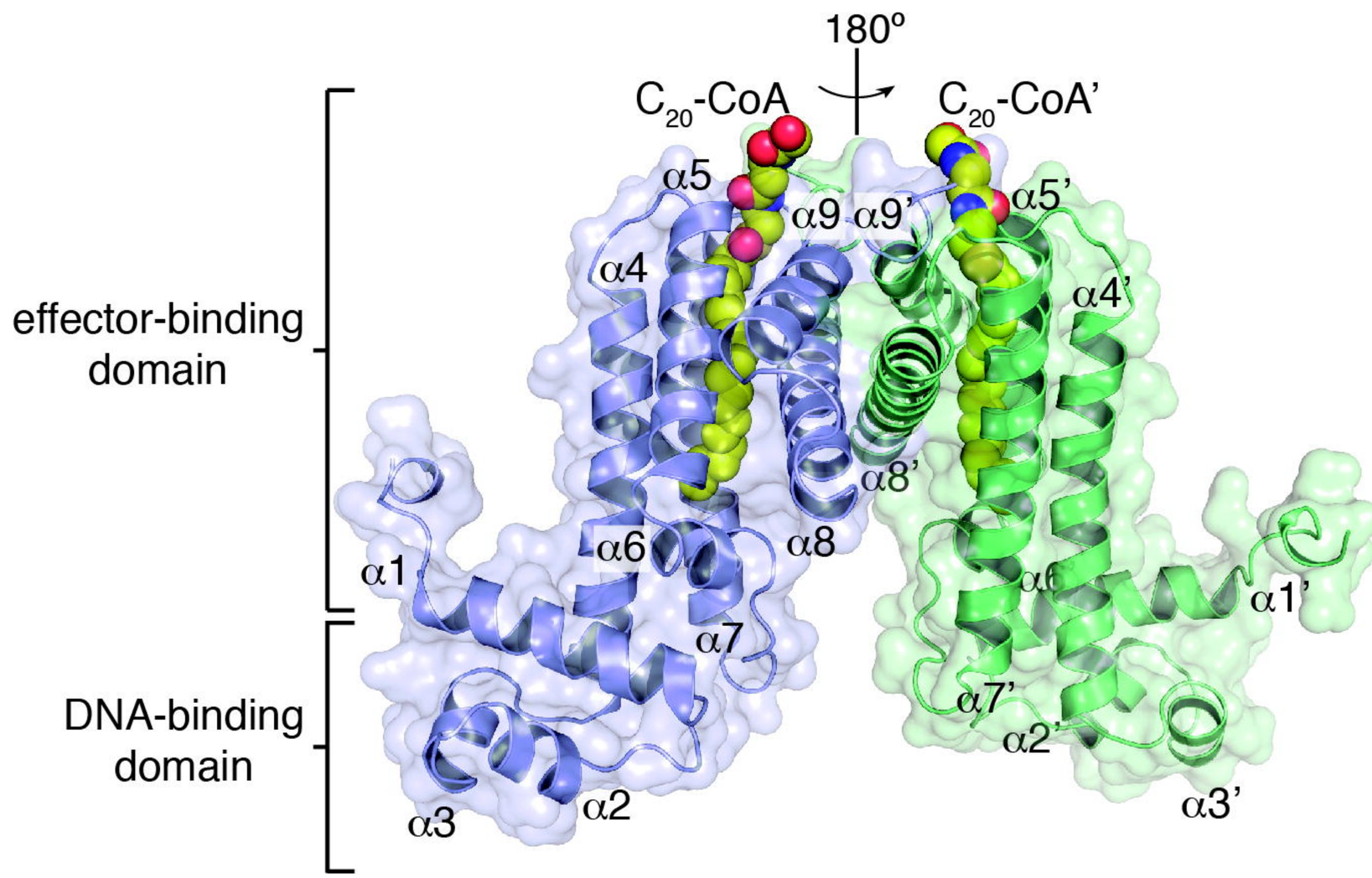
Figure 5. **Crystal structure of full-length FasR in complex with DNA.** Electron density map (sigmaA-weighted $2mF_{obs}$-DF_{calc} Fourier) is shown (magenta mesh), overlaid on the final refined model, showing the protein dimer (protomers in light-blue and pale-green) and the DNA double-helix (orange) in cartoon representation. The map was carved around the atomic model with a border of 3 Å to improve the clarity. Helices α3 are seen fitting within the DNA major groove as expected.

Figure 6. **A TFR-conserved hydrophobic spine connects effector-sensing to DNA-binding output regulation. A.** Hydrophobic spine amino acids in FasR. As detailed in Material and Methods these residues were selected after multiple sequence alignment of >2500 TFRs comprising a large sequence diversity. Thirty-four amino acid positions surpassed a global hydrophobicity cutoff, among which 32 configure a continuous spine in 3D. The colour code depicts residues of the DNA-binding domain in orange, the effector-binding domain in grey and the α6-α7 loop, subjected to strain, in cyan. **B-D.** Distinct TFRs (FasR, RutR, RamR) illustrate the conservation of the hydrophobic spine (the molecular surface of the residues is shown), connecting effector- to DNA-binding domains. Note how the binding of the effector completes the spine in every case. The colour code is identical to panel A. The effector molecules in atom-coloured sticks are labelled, with transparent spheres overlaid (insets show their markedly disparate structures). In FasR and RutR, additional residues are indicated (shown with sticks overlaid by transparent spheres), which are not included in the globally conserved spine, but that contribute to it in individual TFRs, adapting to singularities of the specific effectors. The specific sequence of the spine for each protein is shown at the bottom of panels C and D, numbered according to each protein's sequence (colours respect the domains' scheme). See Supplementary Fig. S12 for additional examples of TFR spines.

Figure 7. **Schematic of the mechanism governing FasR-mediated long acyl-CoA sensing and response.** The model combines information from the three crystal structures presented in this report ($FasR_{\Delta 33}$-C_{14}, $FasR_{\Delta 33}$-C_{20}-CoA and FasR-DNA complexes). A free structure of FasR with no bound effectors nor DNA has not been determined experimentally, and is not currently known whether it builds up to detectable concentrations within the living cell. The dotted arrows reflect plausible equilibria. See Supplementary Video 3 to better grasp the anticipated dynamics.

Fig. 1

A


180°
C20-CoA
C20-CoA'
α5
α9
α9'
α5'
α4
α4'
effector-binding domain
α8'
α6
α8
α1
α6'
α1'
α7
α7'
DNA-binding domain
α2'
α2
α3
α3'

B

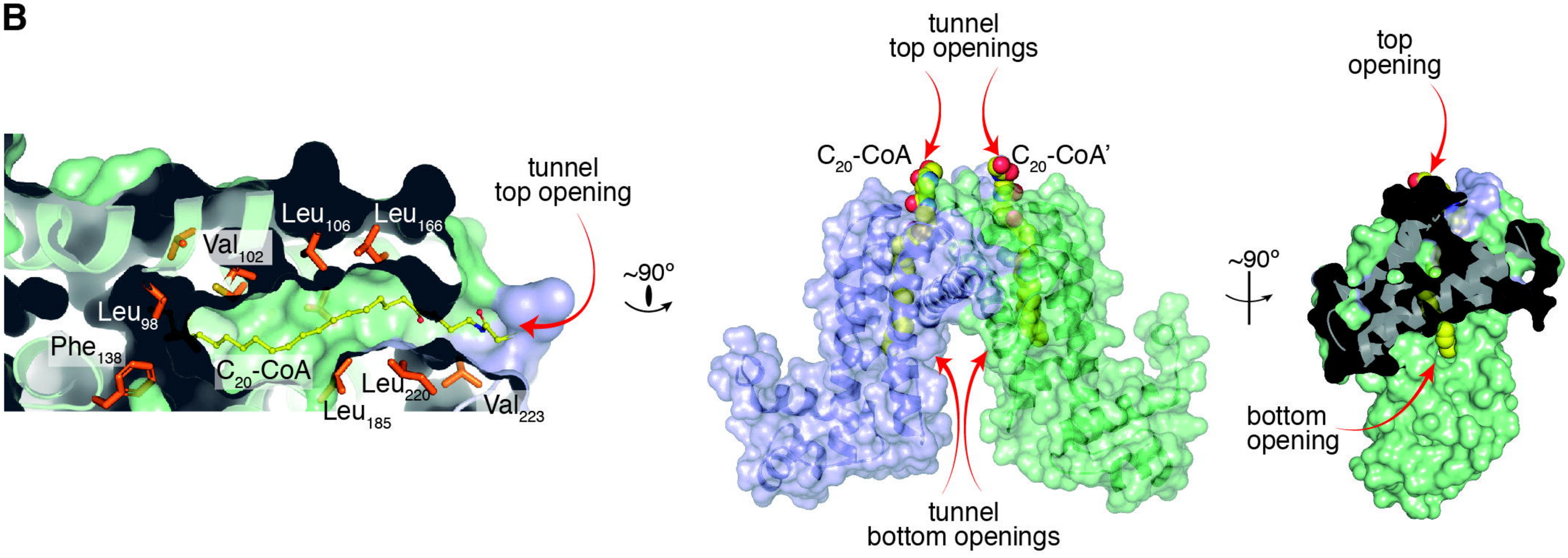

tunnel top opening
Leu106
Leu166
Val102
Leu98
Phe138
C20-CoA
Leu220
Leu185
Val223
~90°
tunnel top openings
C20-CoA
C20-CoA'
tunnel bottom openings
~90°
top opening
bottom opening

Fig. 2

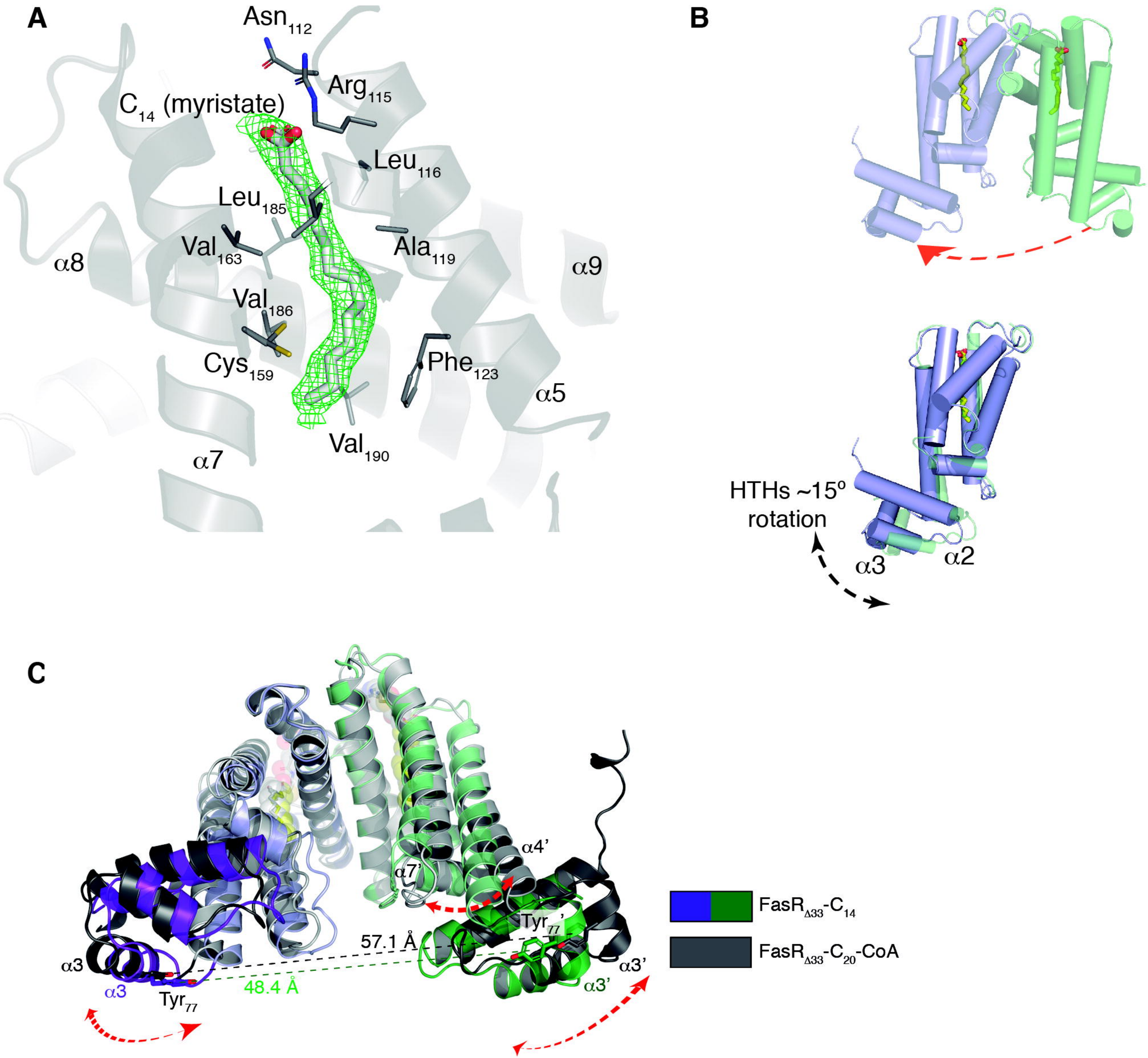

A
Asn112
Arg115
C14 (myristate)
Leu116
Leu185
Val163
Ala119
α8
α9
Val186
Cys159
Phe123
α5
Val190
α7
B
HTHs ~15º
rotation
α3
α2
C
α4'
α7'
Tyr77'
57.1 Å
α3
α3
Tyr77
48.4 Å
α3'
α3'
FasRΔ33-C14
FasRΔ33-C20-CoA

Fig. 3

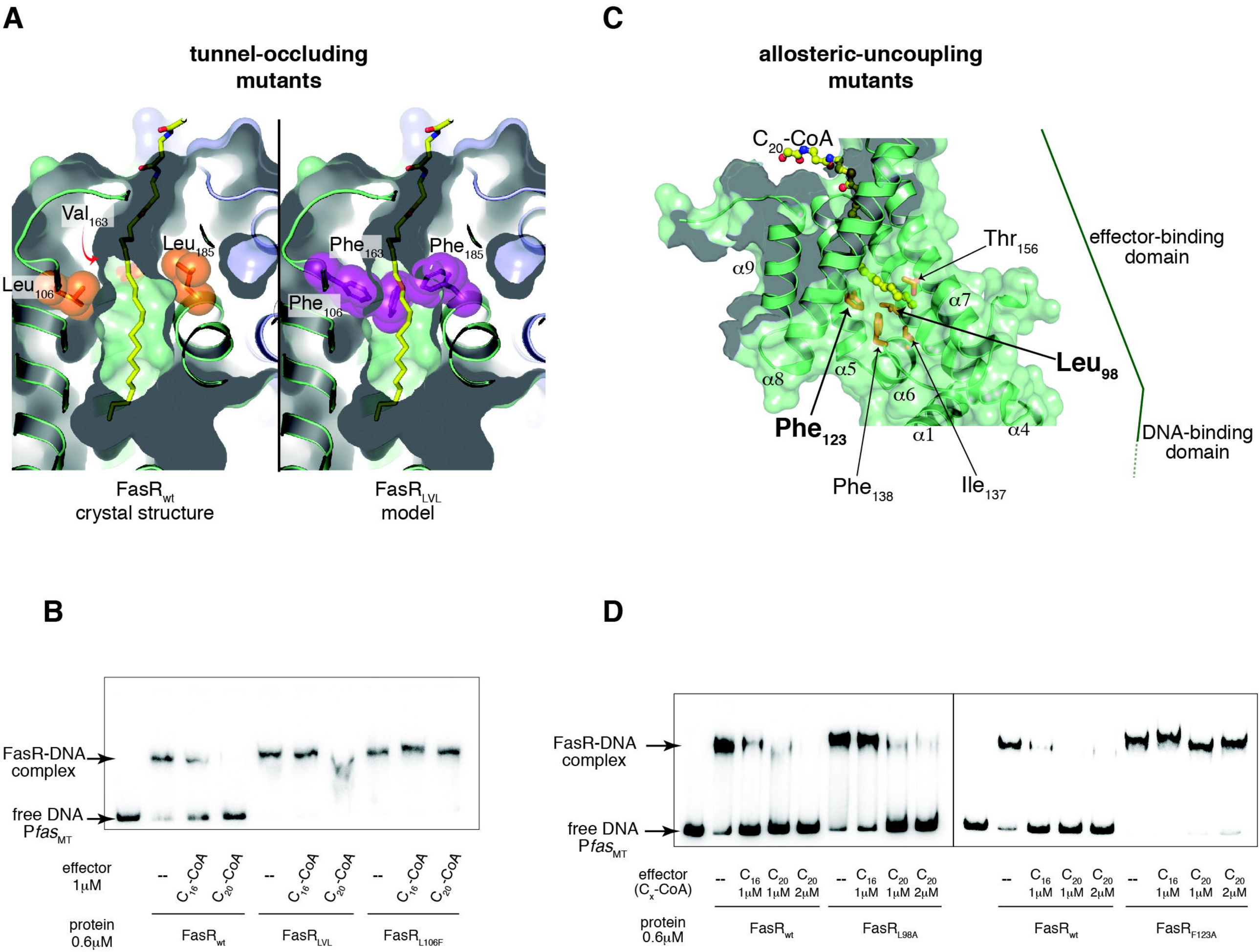

A
tunnel-occluding mutants
Val163
Leu185
Leu106
Phe163
Phe185
Phe106
FasRwt crystal structure
FasRLVL model
C
allosteric-uncoupling mutants
C20-CoA
Thr156
effector-binding domain
α9
α7
Leu98
α8
α5
α6
Phe123
α1
α4
DNA-binding domain
Phe138
Ile137
B
FasR-DNA complex
free DNA PfasMT
effector 1μM
--
C16-CoA
C20-CoA
protein 0.6μM
FasRwt
FasRLVL
FasRL106F
D
FasR-DNA complex
free DNA PfasMT
effector (Cx-CoA)
--
C16 1μM
C20 1μM
C20 2μM
protein 0.6μM
FasRwt
FasRL98A
FasRF123A

Fig. 4

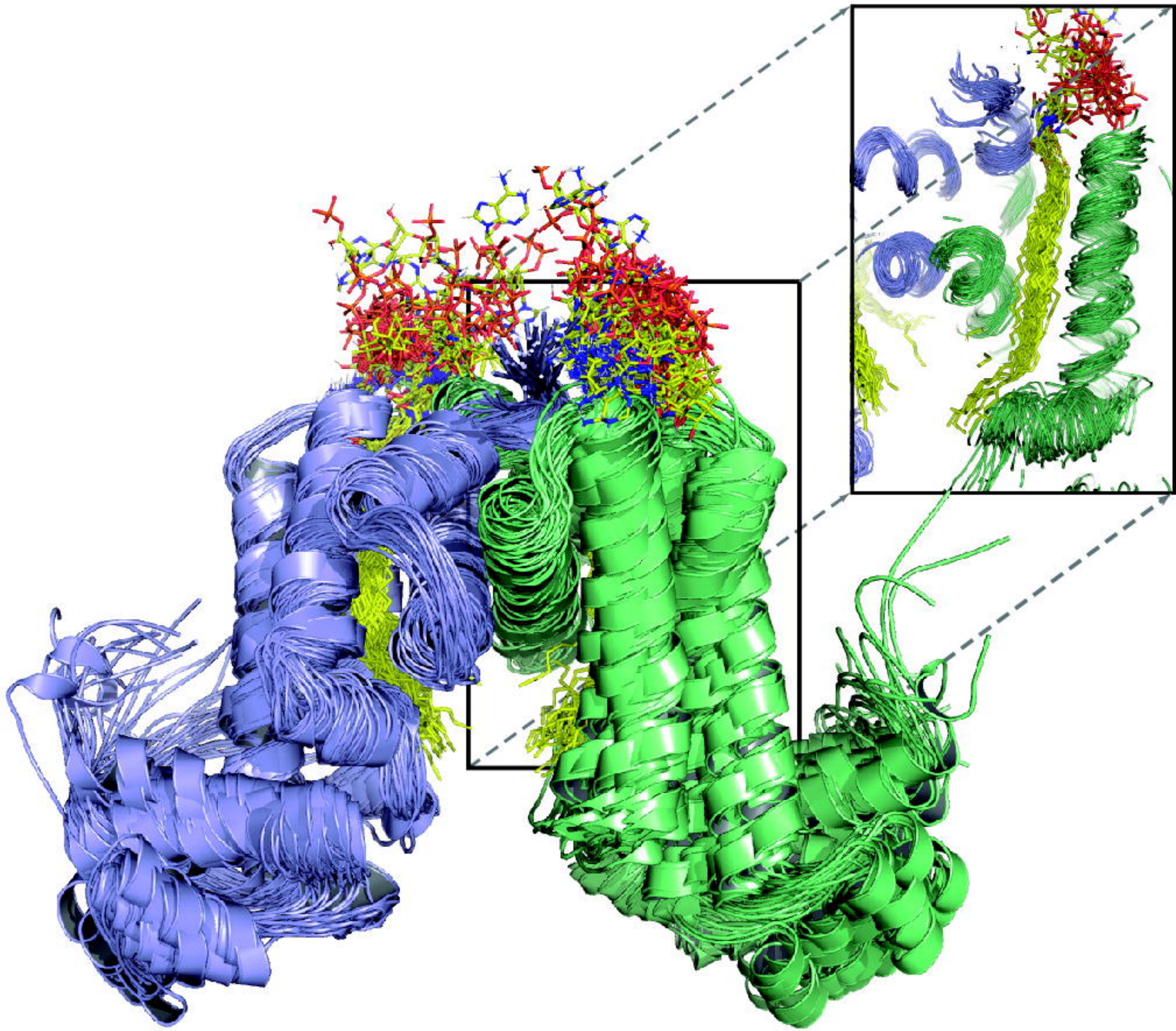


Fig. 5

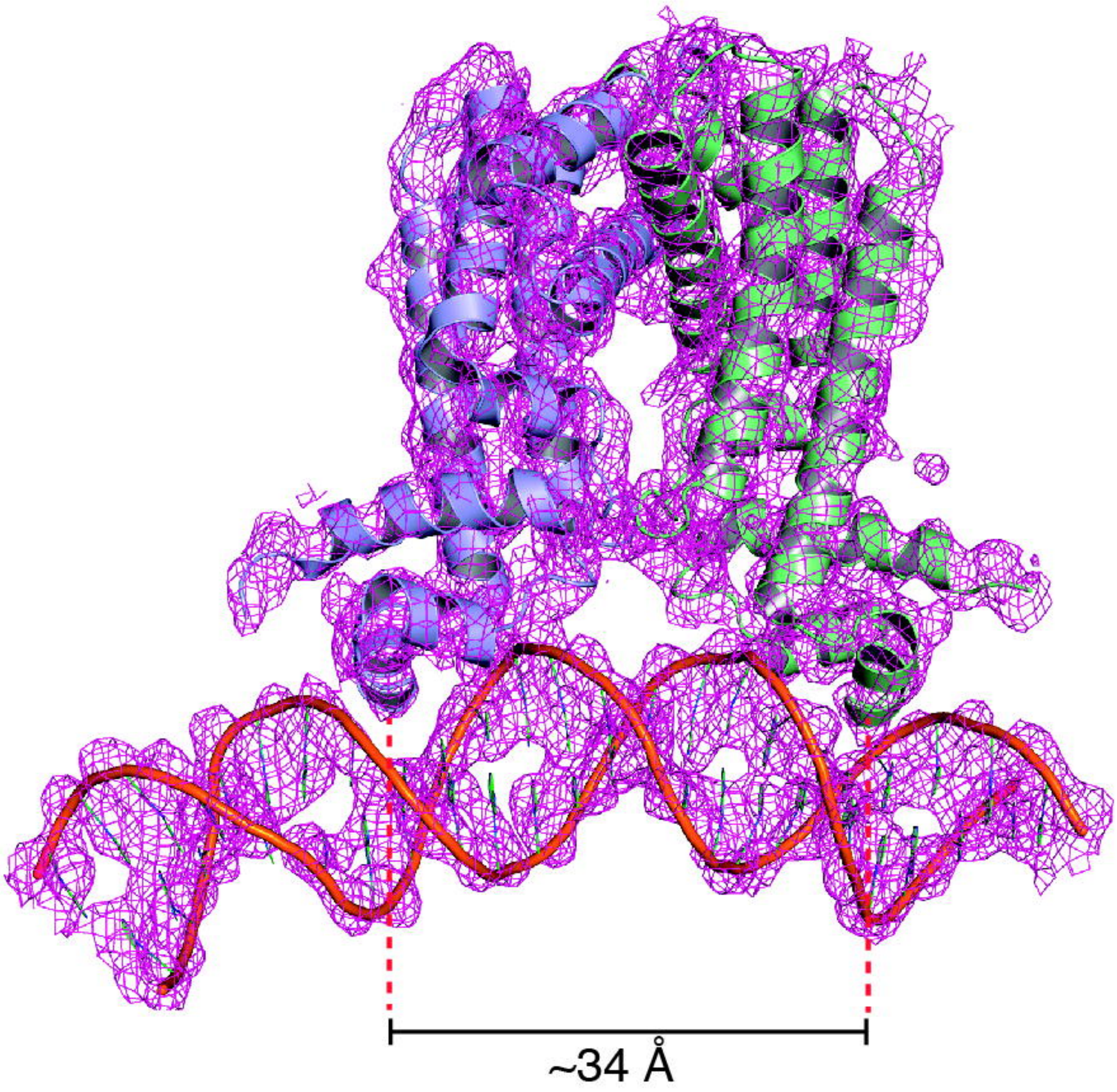


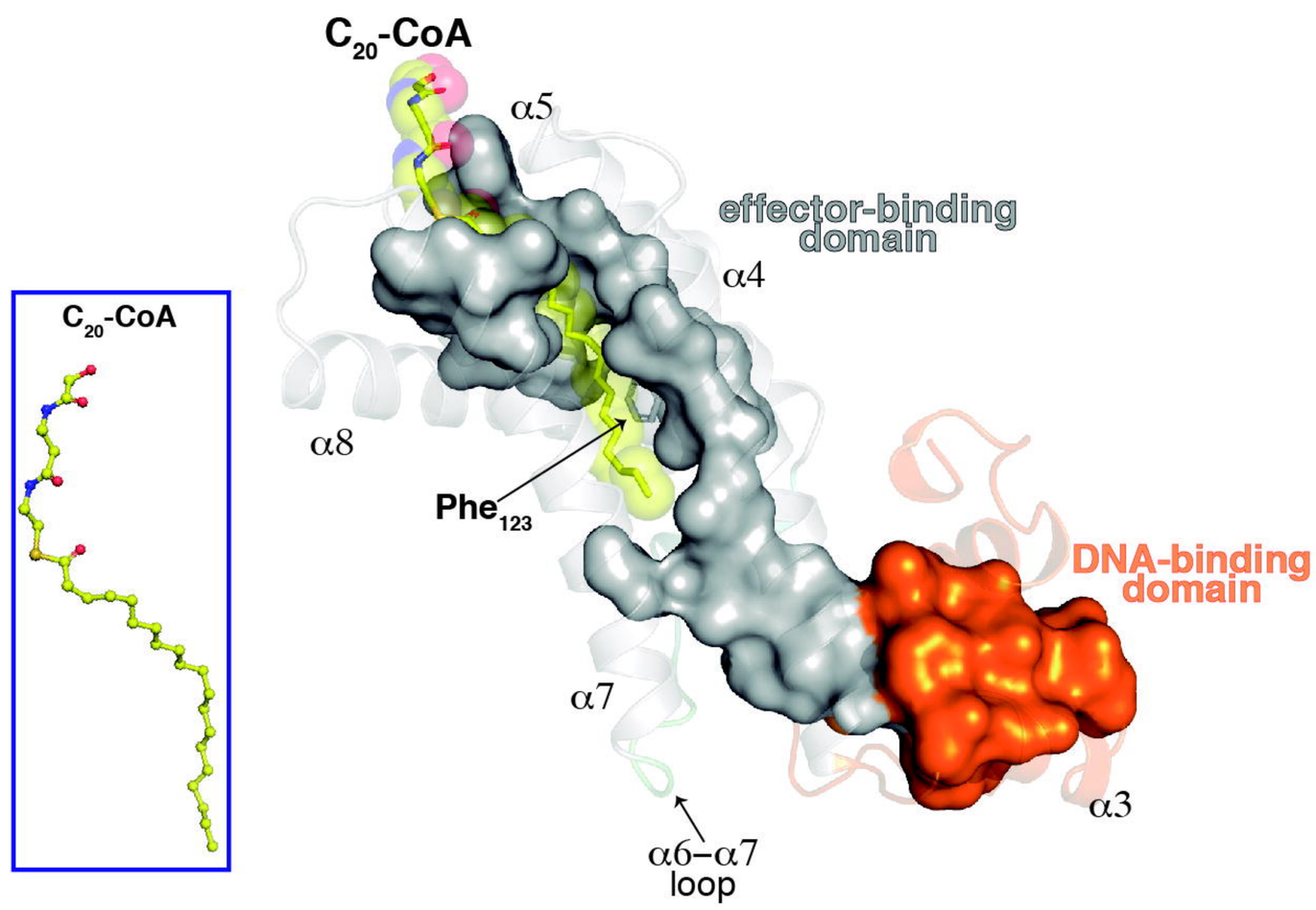
Fig. 6

A

FasR conserved hydrophobic residues conforming a continuous spine

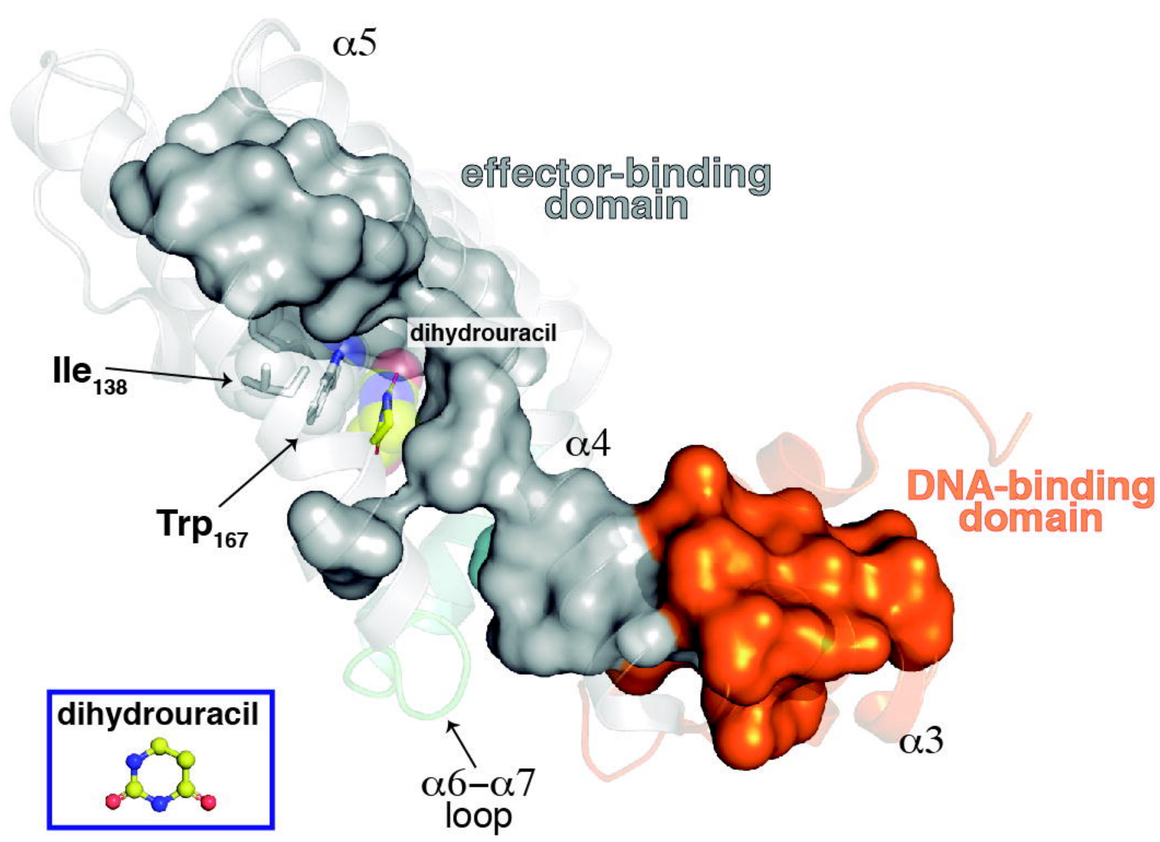
LL-----VF---------M--I---A-V----L---F-----LY--VL---V--L---V-...-L--AV---F--I-...-L-...-V-...-V--LI-...-L---S-...-FAW
44 51 62 65 69 76 80 86 90 95 102 116 124 136 152 163 185 189 217

B



FasR PDB 6O6N
(*Mycobacterium tuberculosis*)

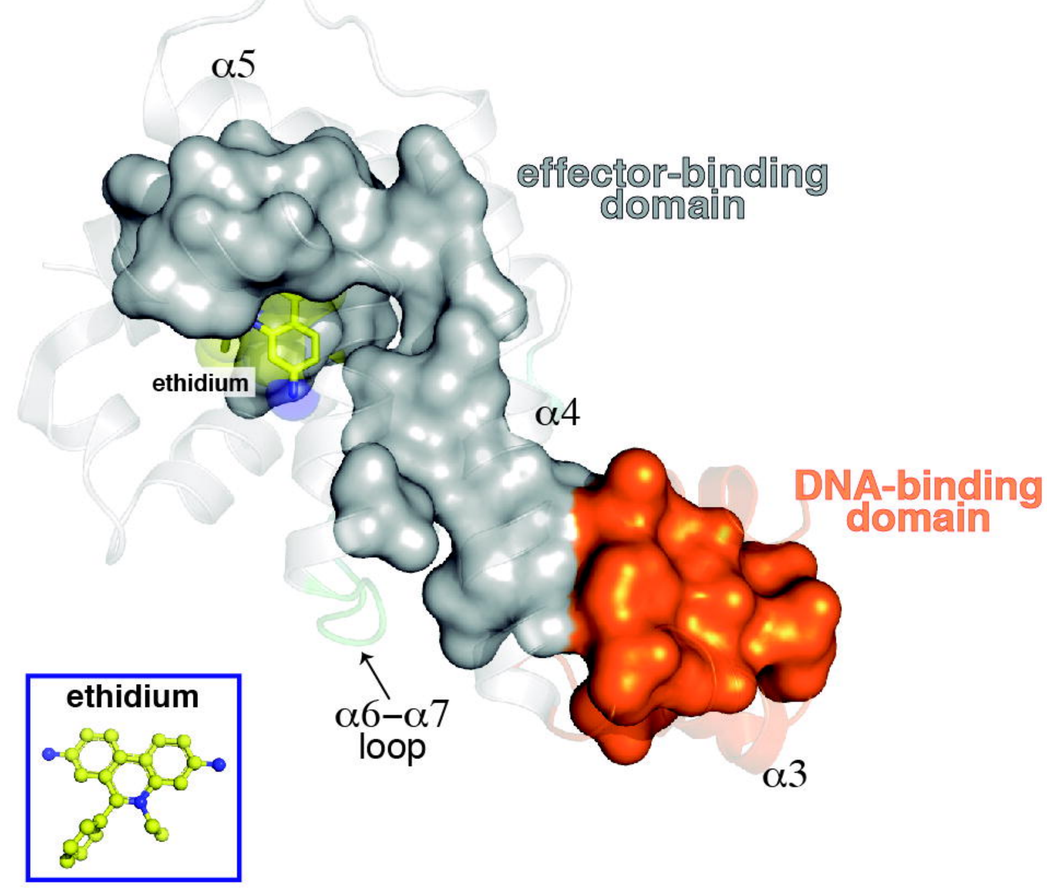
C



RutR PDB 4XK4
(*Escherichia coli*)

IL-----TF---------L--I---A-V----L---F-----LY--VL---L--W---L-...-I--YI---L--S-...-L-...-L-...-K--LI-...-I---T-...-III
23 30 41 48 55 59 65 69 74 81 94 102 114 130 141 166 170 206

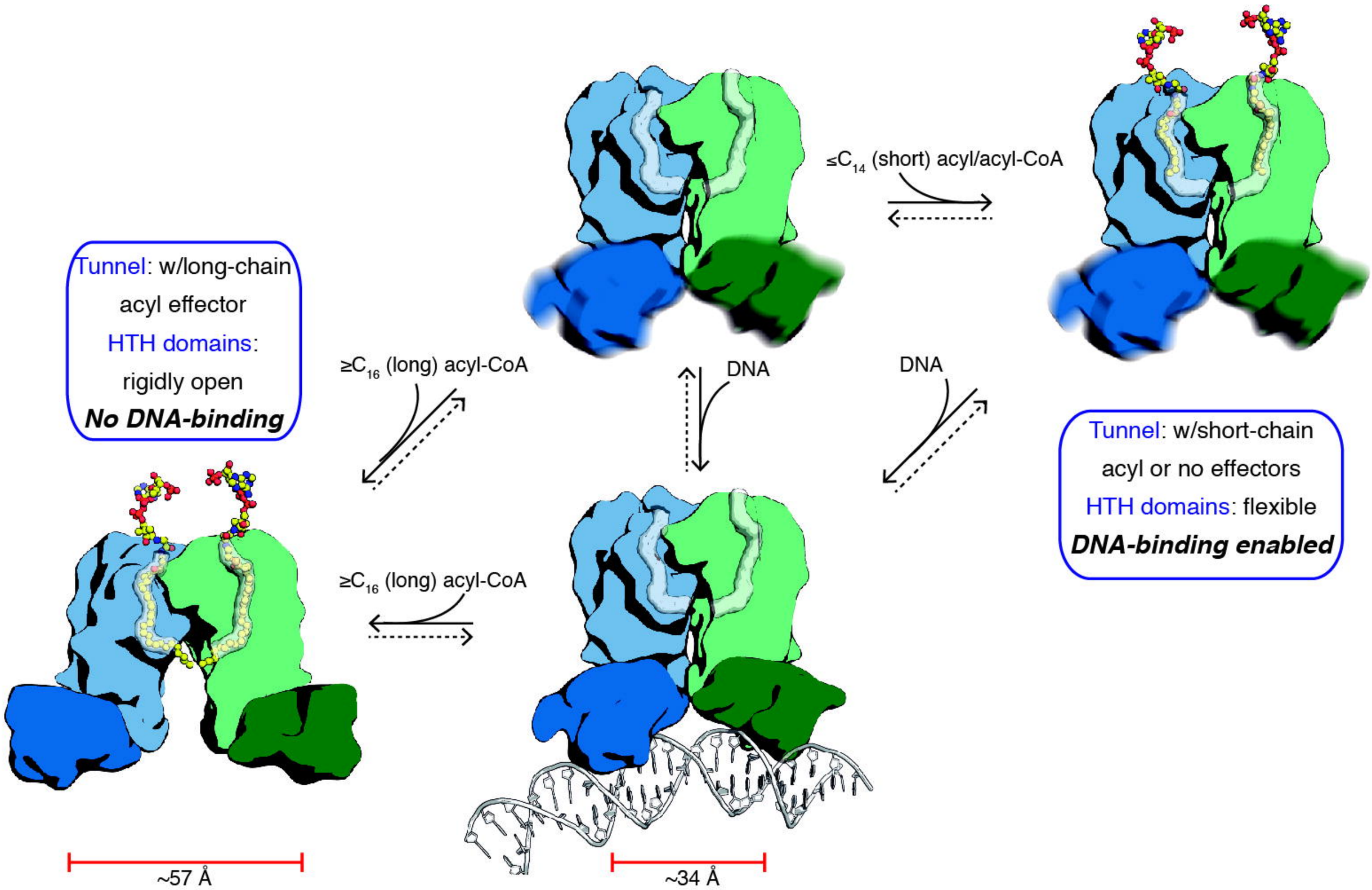
D



RamR PDB 3VVY
(*Salmonella enterica*)

LL-----AI---------T--I---A-V----L---F-----LI--LY---K--L---M-...-T--IW---I--G-...-A-...-T-...-L--LC-...-F---A-...-AMW
13 20 30 37 44 48 54 58 63 70 85 93 105 119 130 155 159 183

Fig. 7


≤C14 (short) acyl/acyl-CoA
Tunnel: w/long-chain acyl effector
HTH domains: rigidly open
No DNA-binding
≥C16 (long) acyl-CoA
DNA
DNA
Tunnel: w/short-chain acyl or no effectors
HTH domains: flexible
DNA-binding enabled
≥C16 (long) acyl-CoA
~57 Å
~34 Å