Transcriptional control of mycobacterial DNA damage response by sigma adaptation

Andreas U. Müller†, Eva Kummert†, Charlotte M. Schilling, Nenad Ban, Eilika Weber-Ban*

Transcriptional activator PafBC is the key regulator of the mycobacterial DNA damage response and controls around 150 genes, including genes involved in the canonical SOS response, through an unknown molecular mechanism. Using a combination of biochemistry and cryo–electron microscopy, we demonstrate that PafBC in the presence of single-stranded DNA activates transcription by reprogramming the canonical −10 and −35 promoter specificity of RNA polymerase associated with the housekeeping sigma subunit. We determine the structure of this transcription initiation complex, revealing a unique mode of promoter recognition, which we term “sigma adaptation.” PafBC inserts between DNA and sigma factor to mediate recognition of hybrid promoters lacking the −35 but featuring the canonical −10 and a PafBC-specific −26 element. Sigma adaptation may constitute a more general mechanism of transcriptional control in mycobacteria.

INTRODUCTION

Cells continuously regulate transcription in response to environmental cues to ensure their survival. In bacteria, a multisubunit RNA polymerase (RNAP) associates with the housekeeping sigma subunit to initiate transcription at specific promoter sequences (1–5). Domains 2 and 4 of the sigma subunit respectively recognize distinct −10 and −35 elements of bacterial promoters (6). Transcription initiation can be modulated by alternative sigma factors or by transcription factors that act as repressors or activators by recognizing additional DNA sequence elements (7). Recently, it was found that PafBC, a member of the widespread family of WYL domain–containing transcription factors, is a key regulator of the DNA damage response in mycobacteria (8, 9). The relevance of DNA damage responses and repair pathways to resistance development makes their understanding particularly important in light of the increasing emergence of drug-resistant Mycobacterium tuberculosis strains (10, 11). The canonical pathway for induction of genes in response to DNA damage, referred to as the SOS response, is present in almost all bacteria (12–14). Expression is achieved by derepression, when RecA senses single-stranded DNA (ssDNA) accumulating under DNA damage and subsequently induces autocatalytic cleavage of the repressor LexA (12, 15). In many bacterial phyla, the SOS response is the only known pathway regulating the DNA damage response. However, in mycobacteria, heterodimeric transcriptional activator PafBC is responsible for the induction of most of the DNA repair genes including the gene for RecA, the regulator of the SOS response (8, 9). Given the occurrence of PafBC in almost every actinobacterial species (16), the PafBC-dependent DNA damage response pathway likely occurs in the entire phylum. A crystal structure of PafBC from Arthrobacter aurescens solved in an inactive conformation in the absence of the promoter revealed that the homologous PaFB and PaFC proteins consist each of an N-terminal winged helix-turn-helix (HTH) domain, followed by a WYL domain (named for a conserved Trp-Tyr-Leu motif) featuring an Sm fold, and a C-terminal extension (WCX) domain with a ferredoxin-like fold (16). PaFB levels remain unchanged upon DNA damage and throughout the stress response, suggesting binding of a response-producing ligand that converts PafBC into an activation-competent form (8, 9). However, it remained unknown how PafBC binds to the promoters under its control and by which mechanism it activates transcription of the target genes. Here, we show that PafBC activates transcription by inserting between DNA and the sigma factor to recognize hybrid promoters featuring the canonical −10 element and the PafBC-specific upstream element located at position −26. Our results explain the characteristic bipartite sequence features of this class of promoters and suggest that “sigma adaptation” may be a common mode of promoter recognition in mycobacteria.

RESULTS AND DISCUSSION

Genome-wide analysis of PafBC DNA binding sites, which identified the 22–nucleotide (nt)–long consensus motif tGTCgg(N10)tA(N3)T for PafBC in vivo, revealed that the motif was not centrally enriched on the protected DNA sequences but exhibited an offset from the chromatin immunoprecipitation sequencing (ChIP-seq) peak center (9), indicating that PafBC could be part of a larger complex bound to the promoter (17). We noticed that the 3′ region of the motif [tA(N3)T] matches the −10 consensus motif of the conserved housekeeping RNAP sigma factor SigA (Fig. 1A) (18), leading us to hypothesize that PafBC interacts with RNAP during transcription initiation.

To test our hypothesis, we used an affinity pulldown assay using RNAP purified from a Mycobacterium smegmatis strain encoding a TwinStrep tag at the C terminus of the β′ subunit. As promoter template, we used a 77-nt-long DNA scaffold based on the promoter region of the recA gene, a member of the PafBC regulon (9). Because mycobacterial RNAP is prone to form unstable open complexes but rapidly binds to DNA scaffolds containing a noncomplementary transcription bubble (19), we designed the DNA scaffold with a mismatched −10 region (fig. S1). We also added the actinobacteria-specific RNAP-binding factor CarD (20, 21), because it stabilizes mycobacterial open complexes (19, 22). Furthermore, as our RNAP isolation contained small amounts of sigma factors other than SigA, we supplemented RNAP with an excess of SigA complexed with RNAP-binding protein RbpA known to stabilize SigA-containing
RNAP holoenzyme (23). Affinity pulldown on a mixture of RNAP, PafBC, and promoter DNA scaffold shows that PafBC is retained with the RNAP complex (Fig. 1B), supporting our hypothesis that PafBC binds to DNA in complex with RNAP. On the basis of the protein folds of the WYL and WCX domains, which are known to frequently occur in RNA binding proteins, we had previously proposed that PafBC requires a nucleic acid ligand for its activation (16). Notably, addition of a 12-nt ssDNA fragment to the reaction mixture substantially increased retention of PafBC (Fig. 1B), suggesting that PafBC, like RecA, senses ssDNA produced under DNA damage. In contrast, using a recA scaffold where the PafBC motif was mutated did not lead to PafBC retention. Likewise, a similarly designed DNA scaffold based on the M. tuberculosis ribosomal RNA promoter rrrAP3 that contains the canonical −10 and −35 consensus motifs of SigA, but not the PafBC motif, could not retain PafBC.

Together, these results establish specific complex formation between PafBC, RNAP, and promoter DNA.

To examine the nature of the complex formed between RNAP, PafBC, and recA promoter DNA and to gain mechanistic insight into PafBC-mediated transcriptional activation, we used single-particle cryo–electron microscopy (cryo-EM). Samples were prepared with purified SigA-RbpA–supplemented holoenzyme, providing recA DNA scaffold, PafBC, ssDNA fragment, and CarD in excess. The three-dimensional (3D) reconstruction calculated on the basis of data collected from two separately prepared samples revealed open complex RNAP containing additional density in the vicinity of domain 4 of SigA, likely corresponding to PafBC (Fig. 2A). To improve the map quality in this region, we carried out local 3D variability analysis resolving conformational heterogeneity (fig. S2). Refining the cluster with the most promising density for PafBC, we obtained a map at 3.2 Å (Fourier shell correlation cutoff at 0.143) from 128,190 particles (RNAP-PafBC global; figs. S2 and S3, A to D), revealing PafBC’s HTH domains and linker regions as clearly defined density at a resolution that permitted tracing of the polypeptide chain (Fig. 2A). The WYL and WCX domains of PafBC that are expected to sense the presence of the added ssDNA were only visible at low resolution upon further local classification and refinement of this region of the map and following subtraction of RNAP density (6.5 Å; 30,575 particles; RNAP-PafBC local; figs. S2 and S3, E to H).

Consequently, while the features of the WYL and WCX domains could be recognized in the maps, they were not assigned to PafB or PafC because of lack of connectivity, and furthermore, we could not assign the density for the ssDNA fragment. Although CarD was added in excess and above the reported \(K_D\) (dissociation constant) of 60 nM (24), it was not observed in the structure, potentially due to a narrower binding interface for CarD between the polymerase and the upstream junction of the transcription bubble compared to CarD-bound complexes (25, 26). The maps reveal that the conformation of the active PafBC heterodimer in the transcription initiation complex is drastically altered compared to the PafBC crystal structure (16). Specifically, the crystal structure shows the HTH domain of PafC buried in the protein core, whereas it is "flipped out" in the initiation complex.

Remarkably, PafBC’s HTH domains are inserted between the promoter DNA and domain 4 of the sigma factor that binds the −35 motif in canonical promoter complexes (Fig. 2, A and B). Thereby, PafBC effectively replaces the contacts of the SigA domain 4 to the −35 motif with contacts to the PafC HTH domain, while the PafB HTH domain provides the specific contact to the promoter DNA in the −23 to −28 region (referred to in the following as −26 element). With its mode of binding, PafBC adapts the RNAP-SigA holoenzyme to recognize hybrid promoters bearing the −26 element of the PafBC consensus in the context of a canonical −10 element. This is consistent with the fact that no additional consensus potentially representing a −35 motif could be identified in PafBC promoters (9). A somewhat related mode of interaction was previously observed for transcription factors produced by phages. To hijack the cellular transcription machinery, Escherichia coli T4 phage proteins AsiA and MotA bridge the phage promoter DNA and Sig70 domain 4 (27). However, they remodel the tertiary structure of domain 4 to enable their binding, a hallmark feature that gave rise to the term "sigma appropriation." In contrast, in the RNAP-PafBC complex, SigA domain 4 retains the same fold as in open complexes on canonical promoters (26, 28). Insertion of PafBC between SigA domain 4 and promoter DNA causes a drastic redirection of the DNA trajectory compared to mycobacterial open complexes where SigA domain 4 establishes contact to the −35 motif (Fig. 2C). In RNAP-SigA open complexes, the DNA bends at positions −18 to −23 by narrowing the minor groove to allow the −35 region to contact SigA domain 4. In contrast, the DNA in the PafBC-RNAP-SigA open complex is straight, leading to a 44° change in the DNA trajectory (Fig. 2C). Formation of a stable initiation complex might be facilitated by the fact that
the energetically unfavorable bending of the DNA (29) does not occur in the PafBC transcription initiation complex. Together, our structure reveals PafBC’s distinct mechanism of transcription activation, which we term sigma adaptation, allowing initiation of the housekeeping RNAP-SigA holoenzyme at promoters lacking the −35 consensus.

By mediating distinct interactions, the two HTH domains of PafB are responsible for specific insertion of the transcription factor between the promoter and SigA domain 4. Winged HTH domains of transcriptional regulators are usually responsible for interaction with the DNA, where helix 3 of the three-helix bundle makes specific interactions with the major groove and the wing contacts the minor groove (30–33). In the heterodimeric PafBC complex, only the PafB HTH domain (HTH-B) makes specific interactions with DNA. Two highly conserved arginines in helix 3 (R46 and R50) mediate the sequence readout in the major groove by hydrogen bonding to two guanine bases—one in the nontemplate strand (G in motif) and one in the template strand (C in motif) (Fig. 3, A to C, and fig. S4A), in addition to phosphate backbone contacts provided by several other conserved residues (K5, R8, S40, K52, R56, and Y76) of HTH-B. Furthermore, the CH$_2$ groups of R46 and E42, together with the CH$_3$ group of A43, provide a hydrophobic environment to accommodate the methyl group of the thymine base of the conserved GTC triplet. Consistent with a role in sequence-specific DNA recognition, complementation of the $M$. smegmatis $\Delta$pafBC strain with a PafBC mutant featuring alanine substitutions at R46/R50 of HTH-B was previously shown to feature the same mitomycin C-sensitive phenotype as the deletion strain (16). The observed interactions rationalize

**Fig. 2.** PafBC adapts the sigma factor to recognize the PafBC-dependent promoter in a transcription initiation complex of mycobacterial RNAP. (A) A chimera of the unsharpened EM map 1 (RNAP-PafBCglobal) and map 2 (RNAP-PafBClocal) is shown with the RNAP subunits, distinct DNA strands (depicted in letter notation above the maps; uncolored bases are not resolved in the EM maps), and PafBC domains colored differently. To create the chimeric map, maps 1 and 2 were superpositioned, and the EM density corresponding to the PafBC HTH domains, DNA, and RNAP was removed from map 2. A schematic representation of the PafBC domain organization is depicted in the middle. (B) Clipping the structural models of RNAP and PafBC vertically reveals the path of the DNA (unclipped, cartoon representation) in the RNAP-PafBC complex. The inset shows the open complex of Mt RNAP on the AP3 promoter (PDB 6EDT) as reference. (C) The upstream element of the PafBC motif is bound by the HTH domain of PafB (HTH-B; dark green), whereas the HTH domain of PafC (HTH-C; light green) contacts domain 4 of SigA (σ$A$; orange). Comparison of the upstream DNA trajectories in RNAP-PafBC and the transcription initiation complex of Mt RNAP on the AP3 promoter (inset; PDB 6EDT) reveals a 44° difference (measured at the intersection of the trajectories). Only DNA, SigA, HTH-B, and HTH-C are displayed. PafBC motif and −10 region of the template (t; salmon) and nontemplate (nt; red) strands are highlighted in yellow.
Fig. 3. PafB’s HTH domain (HTH-B) establishes contact to the −26 element of PafBC-dependent promoters. (A) Hydrogen bonding network at the interface of DNA and HTH-B is shown as black dashed lines. (B) Sequence readout by HTH-B occurs at two guanine bases in the GTC sequence triplet via R46 and R50. (C) Schematic representation of the interactions between PafBC and DNA. Polar and hydrophobic interactions are shown as dashed and solid lines, respectively. (D) Sequence alignment across actinobacterial species of the recA promoter region surrounding the PafBC motif reveals strict conservation of the GTC triplet. Consensus (cons) sequence is given below.

| Species | Accession | Consensus Sequence |
|---------|-----------|--------------------|
| Ace     | GATGGAAMAAAGCTGTGCCGGCGCTTCATAAGCGTCCCCAGC | AAGCGTCCCCAGC |
| Str     | GCTGTTTCTCAGCTGACCTCCTACATTTTGAGCTGCCAGT | AAGCGTCCCCAGC |
| Bsa     | ATCCGAAATTGCTGGATGATAAGCGTCCCCAGC | AAGCGTCCCCAGC |
| Nmu     | TGCGAGAGCAGCTCAGCTCCTAGATTTTGAGCTGCCAGT | AAGCGTCCCCAGC |
| Pfr     | CCCTCAACGTAGCTGACCTCCTACATTTTGAGCTGCCAGT | AAGCGTCCCCAGC |
| Tfu     | CCCTCAACGTAGCTGACCTCCTACATTTTGAGCTGCCAGT | AAGCGTCCCCAGC |
| Mlu     | AGGTTTAAGCATCGCTCCTAGATTTTGAGCTGCCAGT | AAGCGTCCCCAGC |
| Sco     | CCGGACCATTGCTGGATGATAAGCGTCCCCAGC | AAGCGTCCCCAGC |
| Rsa     | CCCTCAACGTAGCTGACCTCCTACATTTTGAGCTGCCAGT | AAGCGTCCCCAGC |
| Sco     | CCGGACCATTGCTGGATGATAAGCGTCCCCAGC | AAGCGTCCCCAGC |
| Ren     | CCGGACCATTGCTGGATGATAAGCGTCCCCAGC | AAGCGTCCCCAGC |
| Rsa     | CCCTCAACGTAGCTGACCTCCTACATTTTGAGCTGCCAGT | AAGCGTCCCCAGC |
| Mtb     | GAGCCGACTTGGTGGCTCCTAGATTTTGAGCTGCCAGT | AAGCGTCCCCAGC |
| Msm     | CCGGACCATTGCTGGATGATAAGCGTCCCCAGC | AAGCGTCCCCAGC |

The conserved GTC triplet of PafBC promoters we identified in our previous study in *M. smegmatis* (9). The conservation of the guanine bases in the motif is almost absolute when comparing sequences among the entire group of actinobacteria and is mirrored by the conservation of R46 and R50 in HTH-B (Fig. 3, D and E).

However, the wing of HTH-B interacts with the holoenzyme instead of binding to the minor groove of the DNA (Fig. 4, A and B). At the tip of the HTH-B wing, W70 inserts into a hydrophobic pocket formed by SigA domain 3 and the β flap of the polymerase (Fig. 4C and fig. S4B). It is possible that this interaction plays a role in slowing the dissociation of PafBC from the complex during promoter escape, when the flap of the polymerase moves away from SigA domain 4 to allow the growing RNA chain to exit the RNAP (34).

The HTH domain of PafC (HTH-C), in contrast to HTH-B, mediates protein–protein interactions with SigA domain 4, which also adopts an HTH fold (Fig. 4, A and B), while only providing a tangential and unspecific DNA backbone contact via a single hydrogen bond from a nonconserved residue (fig. S5). The rather acidic surface of helix 3 of HTH-C (16) also disfavors DNA binding. Consistently, gel shift assays using either PafB or PafC alone showed...
Fig. 4. Atypical binding modes for the winged HTH domains of PafBC in the transcription initiation complex. (A) W70 located at the tip of the wing in HTH-B inserts into a pocket formed by SigA domain 3 (σA3) and the β flap domain instead of binding the minor groove, while HTH-C establishes contact to SigA domain 4 (σA4). Close-ups of the boxed areas are shown in (C) and (D). (B) Schematic representation of the HTH domain binding mode in RNAP-PafBC and the canonical binding mode of winged HTH proteins. (C) Residues of SigA domain 3 and β flap domain form a pocket surrounding W70 of HTH-B. (D) Hydrogen bonding (black dashed lines) and hydrophobic interactions make up the binding interface of HTH-C and SigA domain 4.

Fig. 5. Model of PafBC-mediated transcriptional reprogramming by sigma adaptation under DNA damage. Housekeeping genes featuring the characteristic −10 and −35 motifs can be recognized and transcribed by the RNAP holoenzyme containing SigA (top). PafBC promoters, consisting of the canonical −10 element and the PafBC-specific −26 element, are not recognized by the RNAP-SigA holoenzyme (crossed-out arrow). Under DNA damage, PafBC adapts the sigma factor to recognize the PafBC-specific −26 element (green), leading to transcription of PafBC-dependent genes (bottom).
nonspecific DNA binding for PafB but not PafC (35). Arginines R416 (helix 1) and R442 (helix 3) of SigA domain 4 are in contact with carboxy groups of the HTH-C main chain, while D62 in β1 of HTH-C hydrogen bonds to S446 in helix 3 of SigA domain 4 (Fig. 4D and fig. S4C). R416, R442, and S446 are absolutely conserved in SigA among actinobacteria, and D62 in HTH-C occupies a position where either Asp or Glu is present (fig. S5). SigA R416 and R442 are normally involved in binding of the –35 motif (36–38), whereas in the RNAP-PafBC transcription initiation complex, HTH-C provides the contacts for these residues and therefore, together with the DNA recognition by HTH-B, completes the role of PafBC as an adaptor.

Our structure demonstrates that the two HTH domains, rather than recognizing adjacent sequences in the DNA as do prototypical dimeric HTH-containing transcription factors, play distinct essential roles: HTH-B recognizes a specific promoter DNA sequence, whereas HTH-C recognizes the RNAP holoenzyme. This explains why PafBC occurs as a heterodimer and why both proteins are essential for its function (8).

The structure of the PafBC-activated transcription initiation complex reveals that the canonical RNAP holoenzyme is reprogrammed by PafBC to recognize a bipartite promoter motif composed of the canonical –10 sequence coupled to a PafBC-specific –26 promoter element (Fig. 5). Most of the mycobacterial promoters exhibit the highly conserved –10 sequence characteristic for the housekeeping sigma factors (39, 40). However, motif discovery attempts upstream of the –10 motif could not identify a conserved –35 motif in many of these promoters (28, 39), suggesting that mycobacterial RNAP heavily relies on additional transcription factors that recognize different motifs by altering the canonical –35 interactions mediated by SigA. Hence, transcriptional control by sigma factors that recognize different motifs by altering the canonical −35 inter-

**MATERIALS AND METHODS**

**Preparation of purified PafBC**

*M. smegmatis* PafBC was expressed and purified as previously described (8). Briefly, the *pafBC* transcriptional unit of *M. smegmatis* mc2-155 was cloned into an IPTG (isopropyl-β-D-thiogalactopyranoside)–inducible expression vector encoding an N-terminal His6–TEV (tobacco etch virus) tag. Protein was expressed in *E. coli* Rosetta cells and purified using a Ni-NTA column (GE Healthcare IMAC Sepharose FF resin charged with Ni2+). After TEV protease cleavage and reverse affinity purification of cleaved protein, the sample was subjected to anion exchange chromatography. Fractions containing an equal ratio judged by SDS–polyacrylamide gel electrophoresis (PAGE) analysis with Coomassie staining were pooled and further purified by gel filtration (Superdex 200) in storage buffer [50 mM Hepes–KOH (pH 7.8)/room temperature (RT), 150 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), 5% (v/v) glycerol]. Peak fractions were pooled and concentrated to about 24 mg/ml using centrifugal MWCO (molecular weight cutoff) filters (Amicon). Protein was aliquoted and stored at −20°C until use.

**Generation of a M. smegmatis strain encoding rpoC-TwinStrep**

The genome of *M. smegmatis* SMR5 ΔpafBC (8) was modified by allelic exchange (41) to introduce a TwinStrep tag at the C terminus of the rpoC gene. Fragments of 1.5 kb upstream and downstream of the rpoC stop codon were amplified from purified *M. smegmatis* genomic DNA using primers that would introduce the TwinStrep tag and contained overhangs for subsequent vector assembly (upstream: rpoC-2xS-px-fw and rpoC-2xS-px-rv; downstream: rpoC-2xS-ds-fw and rpoC-2xS-ds-rv; for sequences, see table S2). The fragments were cloned into XmnI–digested pGOAL19 [pGOAL19 was a gift from T. Parish (42); Addgene plasmid no. 20190] by isothermal DNA assembly (43) to generate the suicide vector. Roughly 1.3 μg of suicide vector was ultraviolet-irradiated for 1 min and transformed into *M. smegmatis* SMR5 ΔpafBC by electroporation, and transformants were selected on 7H10 agar containing hygromycin (50 μg/ml) for 3 days at 37°C. Single-crossover (SCO) recombinants were identified by blue-white screening (underlaid 200 μl of 0.4% (w/v) 5-bromo-4-chloro-3-indolyl–β-D-galactopyranoside (X-gal; Sigma-Aldrich) in dimethyl sulfoxide) as blue colonies, and four positive SCO colonies were used to inoculate 5-ml 7H9 liquid culture containing hygromycin (50 μg/ml). SCO cultures were grown to OD600 (optical density at 600 nm) of about 0.7 to 1.8 at 37°C, shaking (1 to 2 days). Cultures were diluted to an OD600 of 0.6, and a dilution series (10−3, 10−2, 10−1) in 7H9 was prepared. Aliquots of 100 μl diluted culture were plated on 7H10 agar supplemented with 2% (w/v) sucrose and incubated at 37°C for 3 to 4 days to perform counterselection. Double-crossover (DCO) colonies were identified by blue-white screening (see above) as white colonies and streaked on plain 7H10 agar and on 7H10 agar containing hygromycin (50 μg/ml). Hygromycin-sensitive DCO colonies, but growing on plain 7H10 agar, were further screened by colony polymerase chain reaction (PCR) to confirm the insertion of the TwinStrep tag. The rpoC-TwinStrep locus of candidate colonies was then amplified by PCR and sequenced to establish sequence integrity. Expression of tagged rpoC was validated by immunoblotting using Streptactin–HRP (horseradish peroxidase) (iba).

**Purification of M. smegmatis RNAP**

*M. smegmatis* SMR5 ΔpafBC rpoC-TwinStrep cells were grown in 9 liters of 7H9 medium at 37°C. At an OD600 of about 1.9, cells were harvested (F9S, 4000g, 4°C, 15 min), and pellets were stored at −20°C until use. Cell pellets were thawed and resuspended in ice-cold buffer A [50 mM Hepes–KOH (pH 7.8)/RT, 1 mM EDTA, 1 mM dithiothreitol (DTT), 5% (v/v) glycerol] supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF) and 1× Complete EDTA-free protease inhibitor (Roche). Cells were lysed by high-pressure shear force (Microfluidizer M110-L, Microfluidics; five passes, 11,000 psi chamber pressure), and insoluble material was removed by centrifugation [SS34, 20,000 rpm (47,810g), 4°C, 30 min]. Polyethyleneimine [10% (v/v) stock solution in 10 mM Hepes–KOH (pH 7.6)/RT] was slowly added to a final concentration of 0.35% (v/v) while stirring. After incubation for 15 min on ice, the precipitate was collected by centrifugation (3000g, 10 min, 4°C) and washed in buffer B [50 mM Hepes–KOH (pH 7.8)/RT, 0.5 M NaCl, 1 mM EDTA, 1 mM DTT, 5% (v/v) glycerol] for a total of three washes. After the last wash, RNAP was eluted by resuspending pellets in buffer C [50 mM Hepes–KOH (pH 7.8)/RT, 1 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% (v/v) glycerol] for a total of three elution steps. Protein was precipitated from the pooled elution fractions by the addition of solid ammonium sulfate [70% (w/v) final, 1-hour stirred incubation at 4°C]. Precipitate was collected by centrifugation [SS34, 20,000 rpm (47,810g), 4°C, 20 min] and resuspended in buffer C supplemented with avidin (0.4 U/ml final; Sigma-Aldrich). After 20-min incubation...
on ice, sample was passed over 4-ml StrepTactin XT 4Flow high-capacity resin (IBA) in a polypropylene column equilibrated in buffer B. Resin was washed five times with one column volume of buffer B, and protein was eluted with buffer B containing 50 mM biotin. Fractions containing RNAP were pooled and diluted in buffer A to achieve about 100 mM NaCl concentration before purification over a Heparin FF16/10 column (GE Healthcare, 20 ml column volume) using a gradient from 0.1 M to 1 M NaCl over 15 column volumes (buffer A/C). Fractions containing all core RNAP subunits were pooled and subjected to gel filtration (Superdex 200, GE Healthcare) in buffer D [20 mM Hepes-KOH (pH 7.8)/RT, 100 mM K glutamate, 10 mM MgCl₂, and 0.5 mM TCEP]. Fractions containing all core RNAP subunits were pooled, and RNAP was concentrated to 4.8 mg/ml using Amicon centrifugal filters (Merck Millipore), aliquoted, frozen in liquid nitrogen, and stored at −80°C until use. The identity of the subunits in the preparation was confirmed by gel-based liquid chromatography–tandem mass spectrometry analysis.

Expression and purification of SigA, RbpA, and CarD
Genes were amplified from purified M. smegmatis mc2-155 genomic DNA (see table S2 for primers), cloned into IPTG-inducible vectors, and expressed in E. coli Rosetta cells. SigA and RbpA were co-expressed (termed SigA-RbpA henceforth) with an N-terminal His₆-thioredoxin-TEV fusion protein from a pET28a vector. Recombinant cells were grown as shaking cultures in LB media containing ampicillin (100 μg/ml; SigA-RbpA) or kanamycin (25 μg/ml; CarD) and chloramphenicol (17 μg/ml; both) at 37°C to OD₆₀₀ of about 0.9 at which IPTG was added to 0.5 mM final concentration. Cultures were further incubated at 30°C for 3.5 hours (SigA-RbpA) or 5 hours (CarD). Cells were harvested by centrifugation [F9S, 7000 rpm (9178g), 10 min, 4°C], and pellets were resuspended in lysis buffer [50 mM Hepes-KOH (pH 7.8)/RT and 0.5 M NaCl] supplemented with 1 mM PMSF and 1× Complete EDTA-free protease inhibitors (Roche). Cells were lysed by high-pressure shear force (Microfluidizer M110-L, Microfluidics; five passes, 11,000-psi chamber pressure), and insoluble material was removed by centrifugation [SS34, 20,000 rpm (47,810g), 4°C, 30 min]. The supernatant was supplemented with deoxyribonuclease I (Roche) at 50 U/ml. After incubation on ice for 15 min, the sample was applied through a 0.45-μm filter onto a Ni-NTA column (GE Healthcare IMAC capacity resin (IBA), in a polypropylene column equilibrated in buffer A (GE Healthcare) eluting with a linear gradient from 0.1 M to 1 M NaCl. Fractions containing SigA-RbpA were pooled and subjected to gel filtration (Superdex 200; GE Healthcare) into buffer [20 mM Hepes-KOH (pH 7.8)/RT, 300 mM NaCl, 0.1 mM EDTA, 0.5 mM TCEP, 5% (v/v) glycerol]. Proteins were concentrated using Amicon centrifugal filters (Merck Millipore) to 13.7 mg/ml (SigA-RbpA) or 25 mg/ml (CarD), aliquoted, frozen in liquid nitrogen, and stored at −20°C until use. Purity and integrity of the protein preparations were validated by electrospray ionization mass spectrometry.

Pulldown assay
Strands of the DNA promoter scaffolds were ordered separately as PAGE-purified, lyophilized oligonucleotides (Microsynth) and resuspended in nuclease-free water, and concentration was determined by absorbance at 260 nm according to (44). To generate DNA scaffolds, strands were annealed in buffer [10 mM Hepes-KOH (pH 7.8)/RT, 50 mM NaCl, and 1 mM EDTA] by heating at 95°C for 10 min and subsequent cooling to RT at 1°C/min. Reactions were prepared using 5× buffer R [1× concentration: 10 mM Hepes-KOH (pH 7.8)/RT, 50 mM KCl, 10 mM MgCl₂, and 1 mM DTT] mixing water, buffer R, DNA (3 μM DNA scaffold and 5 μM ssDNA), and protein components (1.2 μM PafBC, 1 μM RNAP, 5 μM SigA-RbpA, and 5 μM CarD) in a volume of 120 μl. Reactions were incubated at 37°C for 30 min to allow complex formation. Pierce spin columns (Thermo Fisher Scientific), each containing 50 μl Streptactin XT 4Flow high capacity resin (IBA), were equilibrated three times with 300 μl of 1× buffer R (all centrifugation steps at 50g, 30 s, RT). After incubation, the reactions were passed over the columns. Resin was then washed three times with 100 μl of 1× buffer R and then two times with 200 μl of 1× buffer R, and protein was eluted three times with 100 μl of buffer E [50 mM Hepes-KOH (pH 7.8)/RT, 150 mM NaCl, and 50 mM biotin]. Elution fractions were pooled, and protein was precipitated from the pooled fractions using methanol-chloroform precipitation. Samples were subsequently mixed with SDS loading dye, heated at 95°C for 3 to 5 min, and analyzed by SDS-PAGE with Coomassie staining.

Structure determination of RNAP-PafBC by cryo-EM Sample preparation
Purified M. smegmatis RNAP was mixed with fivefold excess of purified SigA-RbpA in sample buffer [20 mM Hepes-KOH (pH 7.8)/RT, 100 mM potassium glutamate, 10 mM MgCl₂, and 1 mM TCEP] and incubated for 30 min at 37°C to form RNAP-SigA holoenzyme before purification by gel filtration (Superose 6 Increase, GE Healthcare) in sample buffer. Peak fractions were pooled and concentrated using Amicon centrifugal filters (Merck Millipore). Oligonucleotides recA-fw and recA-op-rv (see table S2 for sequences) were freshly annealed as described above to yield duplex DNA recA-op. RNAP-PafBC complexes were formed by mixing 1 μM PafBC, 1 μM CarD, 2 μM recA-op, 2 μM ssDNA (CGTCAGTCCAGT), and 0.2 μM RNAP-SigA (added last) in sample buffer. After incubating the reaction for 30 min at 37°C followed by centrifugation at 20,000g for 10 min at 25°C, the supernatant was used for grid preparation.

Grid preparation
Quantifoil Cu 300 R2/2 holey carbon grids were coated with a homemade continuous carbon film and glow-discharged before use. A sample of 5 μl was applied onto the grid, and the sample was allowed to adhere to the support for 30 s at 4°C and 100% humidity before blotting away excess liquid. Subsequently, the sample was...
vitrified in a liquid ethane/propane mixture using a FEI Vitrobot Mark IV instrument.

Data collection

Images were acquired in counting mode at 300 kV on a Titan Krios transmission electron microscope with a total electron dose of 60 e⁻/Å² and a defocus range between 0.3 μm and 4 μm. Dataset 1 (8771 movies) was collected using a Gatan K2 direct electron detector coupled with GIF Quantum LS (×165,000 magnification and 0.84 Å/pixel), whereas dataset 2 (13,154 movies) was recorded using a Gatan K3 direct electron detector equipped with GIF Quantum LS (×105,000 magnification and 0.84 Å/pixel).

Image processing

Movie frames were aligned, summed, and dose-weighted using 5 × 5 patches in MotionCor2 (45, 46). All subsequent steps were carried out in cryoSPARC starting with PATCH CTF estimation (47). The two datasets were initially analyzed separately, and good particles were pooled at a later stage. First, micrographs with an estimated CTF fit resolution lower than 7 Å were discarded from both datasets containing 8453 micrographs in dataset 1 and 12,963 micrographs in dataset 2. In dataset 1, particles were picked using the blob picker in cryoSPARC with a minimum particle diameter of 80 Å and a maximum particle diameter of 250 Å. The normalized cross-correlation (NCC) score was adjusted manually to 0.5 to minimize false-positive picks and still retain the majority of correctly picked particles. A total of 1,730,655 particles were extracted with a box size of 400 × 400 pixels and subjected to unsupervised 2D classification using 40 online EM iterations and 200 classes. A total of 342,185 good particle images were retained and binned two times before joining the images with good particle images from dataset 2. In dataset 2, particles were picked using the blob picker in cryoSPARC with a minimum particle diameter of 100 Å and a maximum diameter of 250 Å. A total of 3,235,525 particle images were extracted at an NCC score of 0.34 with a box size of 400 × 400 pixels and downsampled two times before subjecting them to unsupervised 2D classification using 200 classes and 50 online EM iterations. A total of 931,407 good particle images were retained and merged with the two-time downsampled particle images from dataset 1. Subsequently, any remaining junk images were removed from the particle pool by running heterogeneous refinement against two references, one accounting for RNAP and DNA and the other for noise. Noise and RNAP reference were derived from ab initio 3D reconstruction of the particle images of dataset 1 selected after 2D classification, in which we requested five classes. The noise reference represents one of the five classes, which clearly originated from the alignment of noise images instead of particle images. The RNAP reference was derived after heterogeneous refinement of particle images of three of the five classes that clearly showed features of RNAP and DNA.

Heterogeneous refinement of dataset 1 and 2 images yielded 1,175,953 good particle images that showed partial density for the PafBC dimer in the 3D reconstruction. These particle images were used for local 3D variability analysis with eight modes, filter resolution of 8 Å, and using a mask (mask 1) surrounding the PafBC density (48). Ten particle clusters were requested from the analysis, resulting in 3D reconstructions of RNA and DNA with no, poor, or well-resolved density for PafBC. The most promising class contained 138,441 particle images, which were reextracted unbinned with a box size of 400 × 400 pixels. These particle images were used for heterogeneous refinement against two references, one accounting for the complex of RNAP and PafBC and the other corresponding to noise. A total of 128,190 final particle images were then used for homogeneous refinement yielding map 1 (RNAP-PafBC_global) that contains RNAP, DNA, and well-resolved density for both HTH domains of PafB and PafC, respectively, at an overall resolution of 3.2 Å (fig. S2). However, this reconstruction only provided poor density for the linker region and WYL or WCX domains of PafBC, probably due to their large conformational flexibility. We therefore subjected the 138,441 particle images originating from the first 3D variability approach to another round of local 3D variability analysis using a mask (mask 2) surrounding the PafBC HTH domains and the partially visible density accounting for the WYL and WCX domains more loosely than the mask before. Local 3D variability analysis was carried out using three modes and a filter resolution of 10 Å. Three particle clusters were requested, and the cluster of 30,575 particle images with the most promising density for all domains of PafBC was used for signal subtraction of the RNAP from the particle images combined with subsequent local refinement (maximum alignment resolution restricted to 5 Å). A mask (mask 3) surrounding the HTH domains more tightly but the remaining PafBC domains only very loosely was chosen for signal subtraction to minimize the impact of the rigid RNAP during local refinement. The resulting map 2 (RNAP-PafBC_local) had an overall resolution of 6.5 Å and allowed us to visualize the location of the WCY and WYL domains of PafBC.

Structure determination and modeling

Model building was initiated from homology models of the M. tuberculosis RNAP open promoter complex [Protein Data Bank (PDB) 6E0T; (26)] and PafBC of A. aurescens [PDB 6S9]; (16)]. PDB 6E0T was fitted into map 1 using UCSF Chimera (49) before homology modeling of the M. smegmatis counterparts in Phyre2 to locate the models already close to their final location in the EM map. The models were then manually revised in COOT (50) to account for differences in the amino acid sequence and polypeptide folds in M. smegmatis and to reestablish the two zinc-binding sites in the RNAP β’ subunit. The density of map 1 only allowed for unambiguous modeling of the HTH domains and linker regions of PafBC (residues 2 to 123 for PafC and residues 5 to 128 for PafB). Because of lower local resolution, the side chains of the linker region were stripped to poly-alanine in the final model (residues 82 to 123 for PafC and residues 84 to 128 for PafB). Moreover, the EM density of map 1 did not account for the N-terminal 22 amino acids of RbpA, which were excluded from the final model. While map 2 allowed assigning the approximate location of the WLY and WCX domains, it did not permit to derive the unambiguous orientation of a structural model for these domains, and we therefore decided to not include them in the final model.

To start building the model of the DNA scaffold, the nucleic acid chains of the M. tuberculosis RNAP open promoter complex model (PDB 6E0T) were fitted to map 1. In COOT (50), the model was modified to the sequence of the DNA scaffold used for sample preparation and truncated at the ends as could be accounted for by the EM map. The density allowed us to model 43 bases in each strand including the melted region (of 77 nt present in the scaffold fragments). Subsequently, the location of individual bases was manually revised to match the EM density.

The RNAP-PafBC complex was real space–refined in three macro cycles in PHENIX (51) using default parameters, C-beta restraints, secondary structure restraints, global minimization, and B factor refinement with a weight between experimental data and restraints of 1.0. Ramachandran outliers were then manually fixed in COOT,
and the revised model was subjected to another round of real-space refinement in PHENIX using the previous settings but, in addition, including Ramachandran restraints for three macro cycles. PHENIX model validation statistics are shown in Table S1.

**Figure generation**

Molecular graphics were generated using the PyMOL (Schrödinger), UCSF Chimera, or ChimeraX packages (49, 52).

**Sequence alignments**

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

**SUPPLEMENTARY MATERIALS**

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

**REFERENCES AND NOTES**

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega,

Molecular graphics were generated using the PyMOL (Schrödinger), UCSF Chimera, or ChimeraX packages (49, 52).

Figure generation and the revised model was subjected to another round of real-space refinement in PHENIX using the previous settings but, in addition, including Ramachandran restraints for three macro cycles. PHENIX model validation statistics are shown in Table S1.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Figure generation and the revised model was subjected to another round of real-space refinement in PHENIX using the previous settings but, in addition, including Ramachandran restraints for three macro cycles. PHENIX model validation statistics are shown in Table S1.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.

Protein sequence alignments were performed using Clustal Omega (53). Nucleic acid sequence alignments were done manually.

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.4abl4064

Nucleic acid sequence alignments were done manually.
47. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

48. A. Punjani, D. J. Fleet, 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol.* **213**, 107702 (2021).

49. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

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

51. P. V. Afonine, B. K. Poon, R. J. Read, O. V. Sobolev, T. C. Terwilliger, A. Urzhumtsev, P. D. Adams, Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 531–544 (2018).

52. T. D. Goddard, C. C. Huang, E. C. Meng, E. P. Pettersen, G. S. Couch, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

53. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

54. Y. Z. Tan, P. R. Baldwin, J. H. Davis, J. R. Williamson, C. S. Potter, B. Carragher, D. Lyumkis, Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).

Acknowledgments: We would like to thank A. Scaiola for the support with EM data collection, the ETH scientific center for optical and electron microscopy (ScopeM) for access to electron microscopes, and the ScopeM staff for technical support during cryo-EM data collection. We also acknowledge the Functional Genomics Center Zurich (FGCZ) for support with mass spectrometric analyses. Funding: This work was supported by Swiss National Science Foundation (SNSF) grant 310030_185250 (to E.W.-B.), core funding of ETH Zurich (to N.B.), Roessler Prize (to N.B.), Ernst Jung Prize (to N.B.), and Otto Naegeli Prize for Medical Research (to N.B.). Author contributions: A.U.M. and E.W.-B. conceived the study. A.U.M. expressed and purified proteins and generated the PafBC-bound RNAP initiation complex. C.M.S. and A.U.M. performed biochemical experiments. E.K. and A.U.M. prepared cryo-EM samples and grids. E.K. collected and processed cryo-EM data and determined the cryo-EM structure. A.U.M., E.K., E.W.-B., and N.B. analyzed data and prepared the manuscript. All authors revised the manuscript. Competing interests: The authors declare that they have no competing interests. Data and materials availability: Cryo-EM maps were deposited in the EMDB (EMD-13204 and EMD-13205), and structural model was deposited in the PDB (PDB 7P5X). All other data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 12 July 2021
Accepted 8 October 2021
Published 1 December 2021
10.1126/sciadv.abl4064