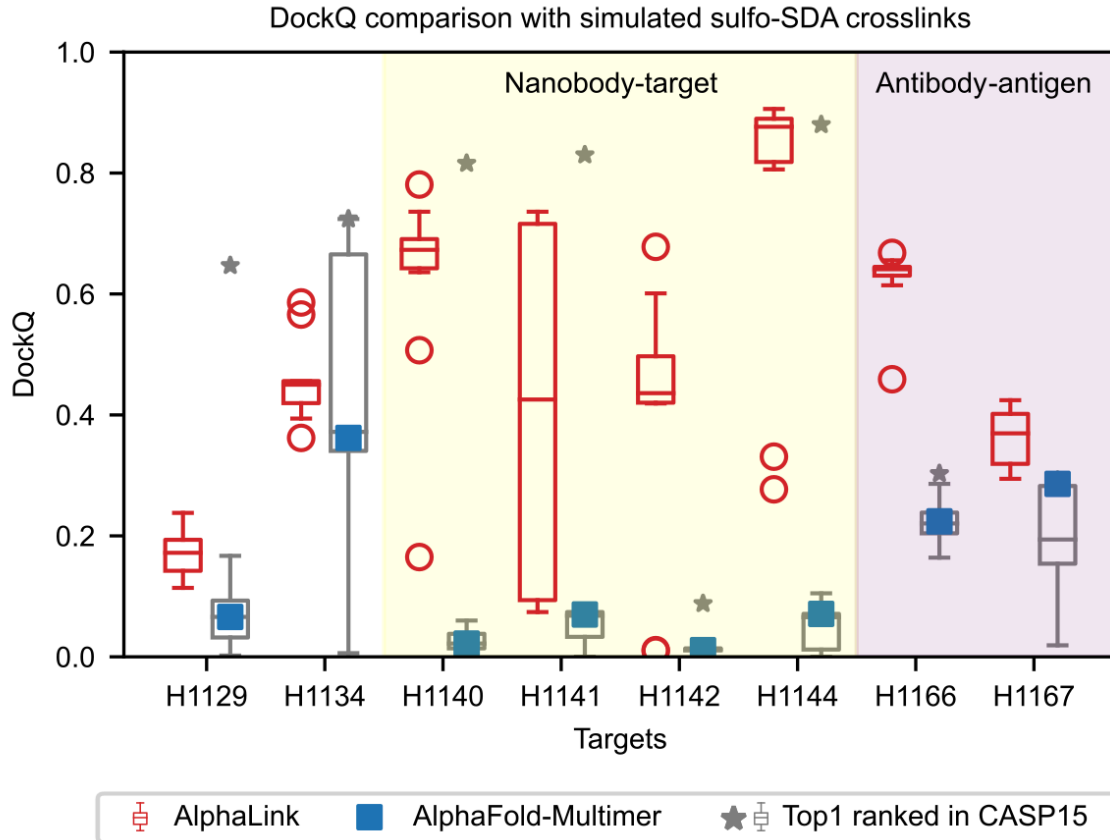
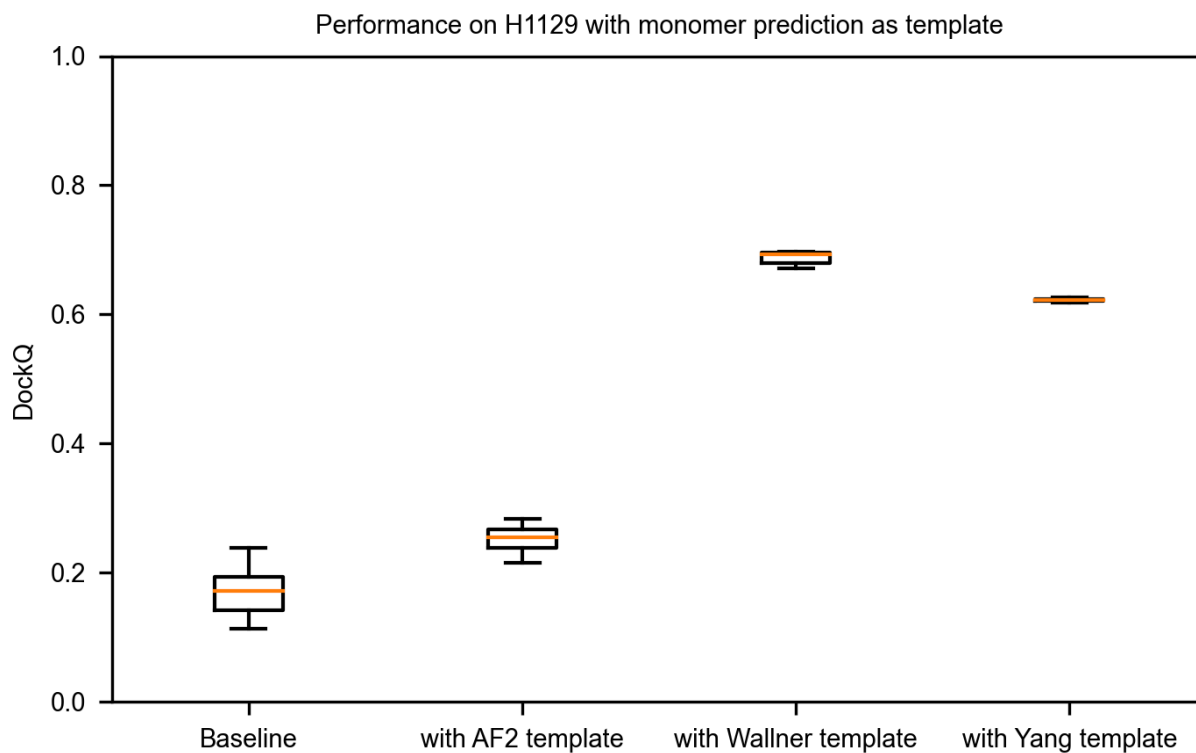


Modelling protein complexes
with crosslinking mass spectrometry and deep learning



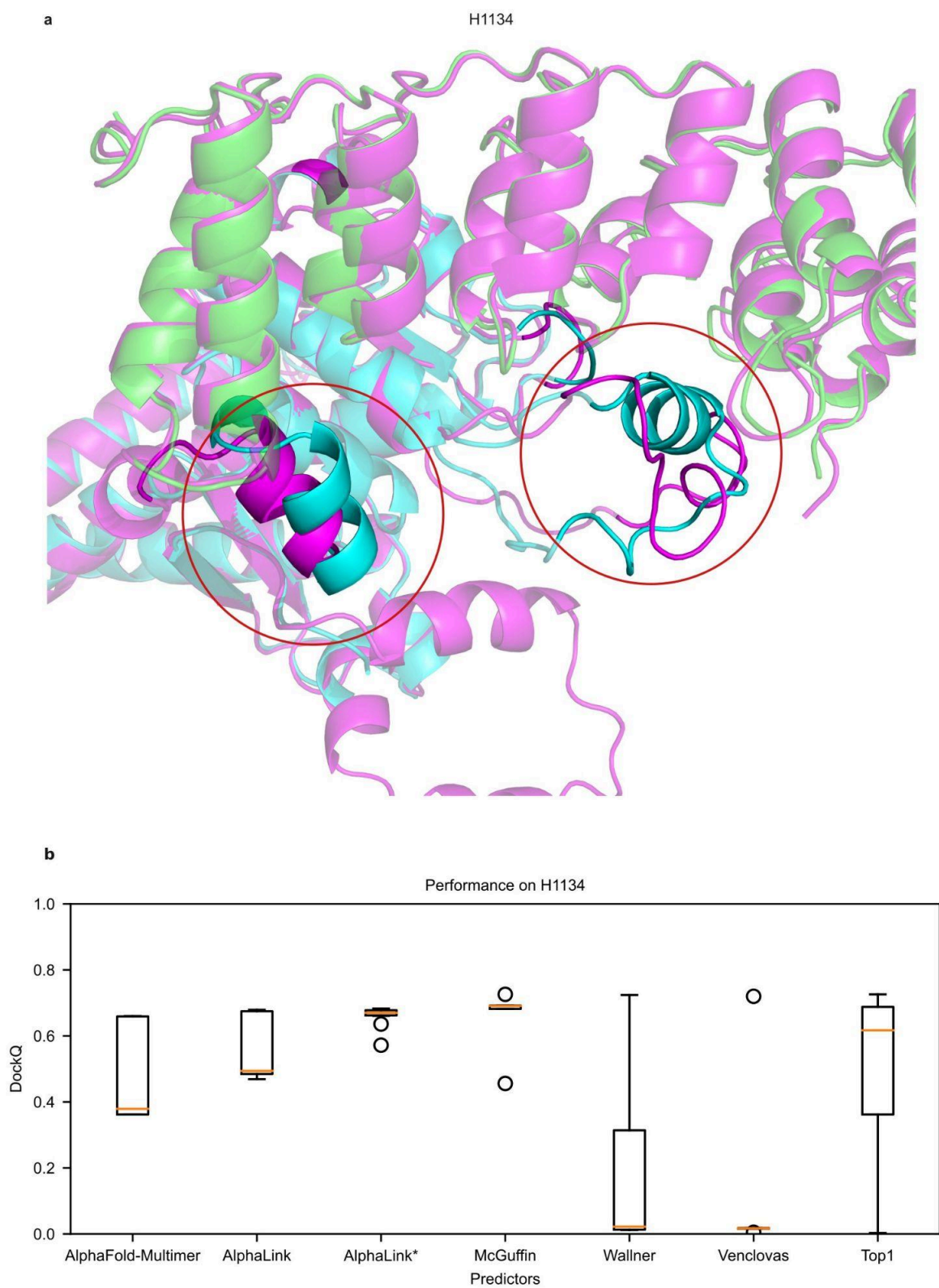
Supplementary Figure 1: DockQ comparison

DockQ comparison on 8 heteromeric CASP15 targets with simulated SDA crosslinks. The AlphaLink boxplot shows the highest model confidence model (N = 10) for 10 randomly sampled crosslink sets. The top ranked predictions (grey) correspond to all submissions that were ranked first by the participants (model₁), the best for each target is highlighted with an asterisk. AlphaFold-Multimer is highlighted in blue (model₁). The yellow shaded area highlights nanobody, and the red shaded area antibody targets. In all boxplots, the line shows the median and the whiskers represent the 1.5x interquartile range.



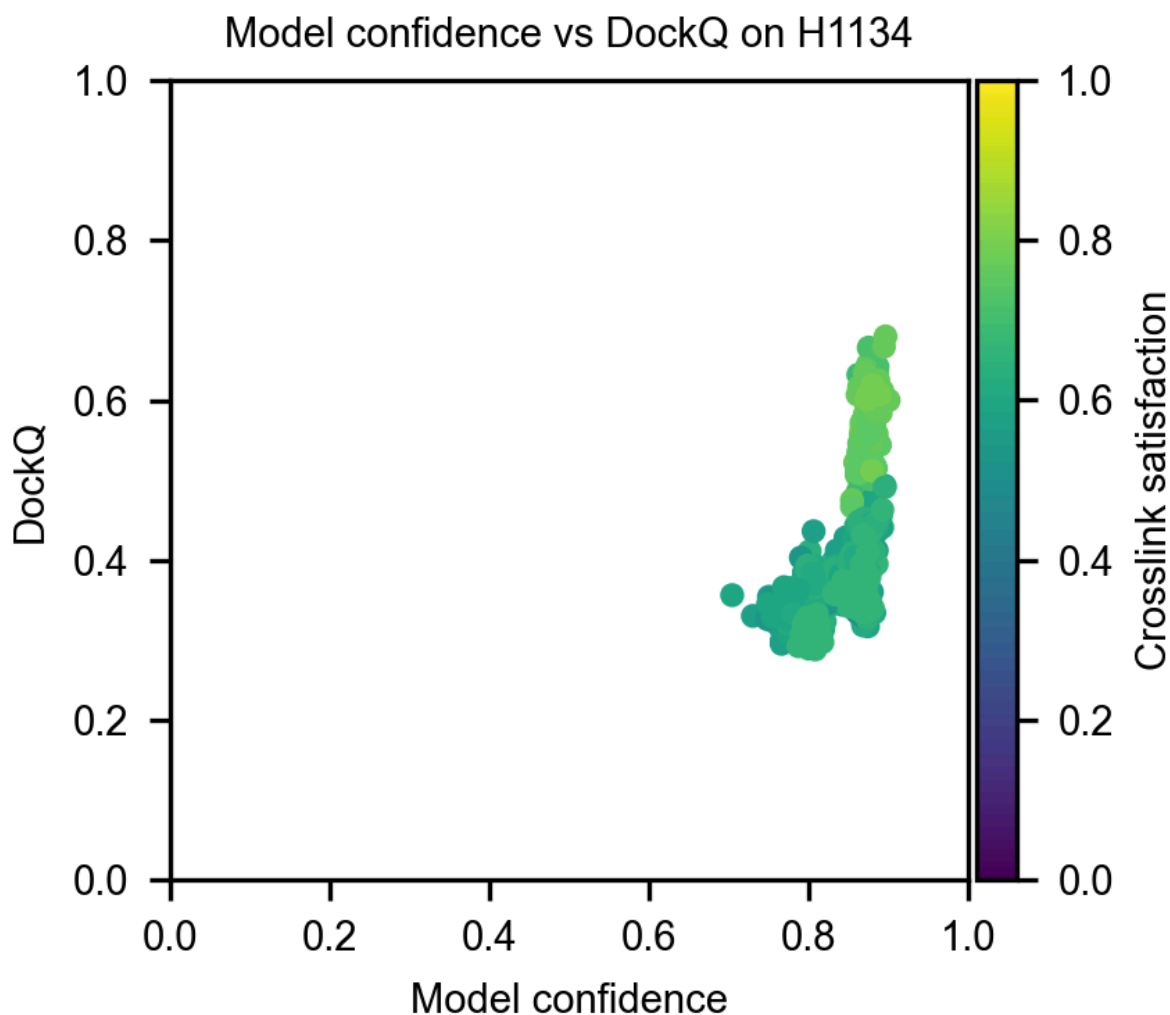
Supplementary Figure 2: Including monomer templates

Performance of AlphaLink with 10 samples and 3 recycling iterations on H1129. Baseline does fold-and-dock, the other runs include the top ranked monomer prediction of chain B from AF2, Wallner, or Yang as a template. In all boxplots, the line shows the median and the whiskers represent the 1.5x interquartile range.



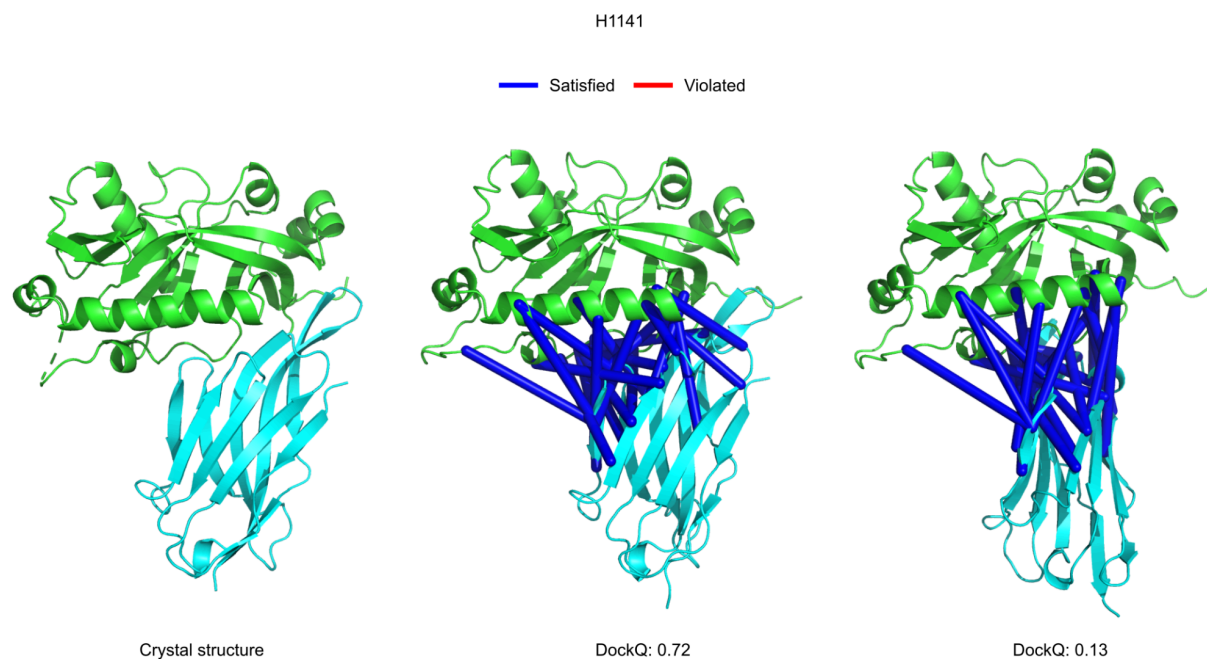
Supplementary Figure 3: Deviations in the interface for the H1134 AlphaLink prediction

a Prediction of AlphaLink (purple) aligned to the native structure (cyan and green). There are few contacts in the interface, the alpha-helical region (red circle, right) is likely very flexible. The biggest deviations are highlighted by the red circle. **b** DockQ scores of all submissions for the top predictors in CASP15 for H1134. Best in CASP15 corresponds to the best DockQ score per group. Most predictors show a large spread in prediction quality. In all boxplots, the line shows the median and the whiskers represent the 1.5x interquartile range. AlphaLink* selects the best prediction first by crosslink satisfaction and second by model confidence.



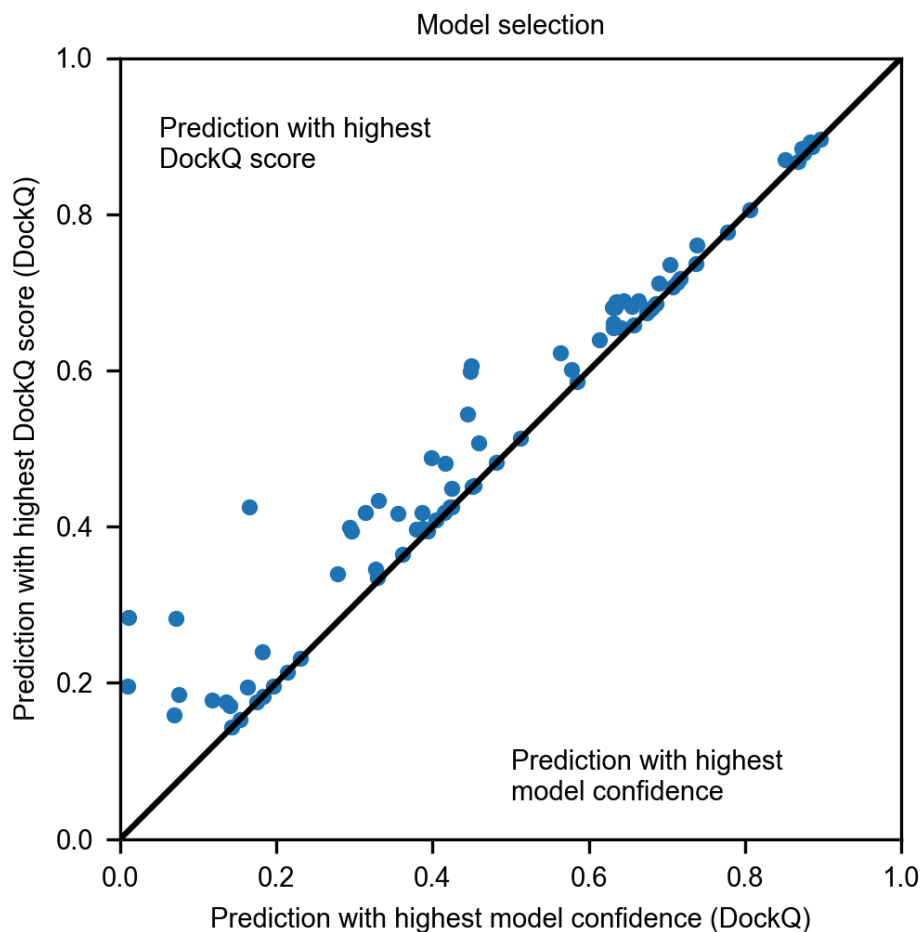
Supplementary Figure 4: AlphaLink model confidence - DockQ correlation on target H1134

Model confidence - DockQ correlation on H1134. We randomly subsampled $N = 10$ crosslink sets and generated $M = 100$ predictions per crosslink set with different seeds. One scatter point is one prediction. The colour gradient is showing the crosslink satisfaction.



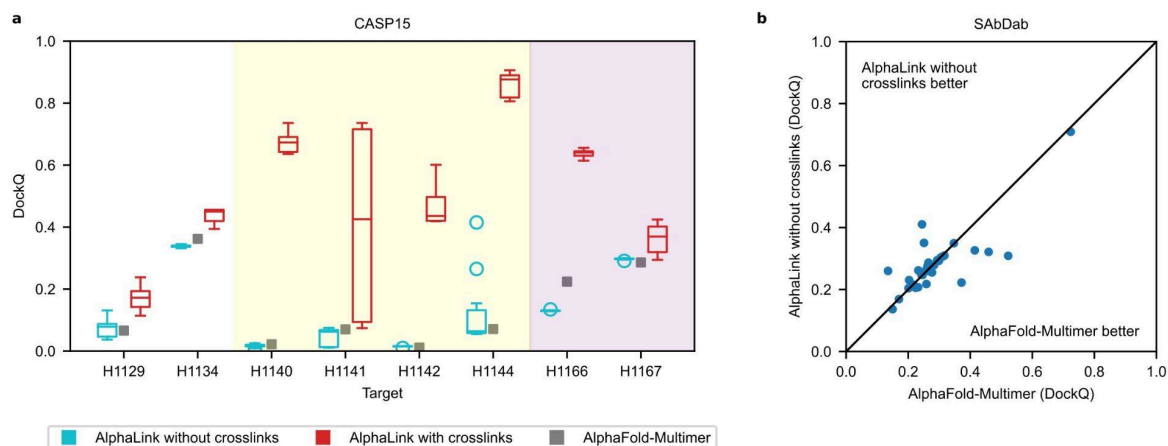
Supplementary Figure 5: Flexibility with SDA

Crystal structure (left) and AlphaLink predictions (middle and right) of H1141. We use the same crosslink set for both predictions. Crosslinks are completely satisfied ($< 25 \text{ \AA}$) (blue). SDA's lower resolution allows predictions with vastly different DockQ scores.



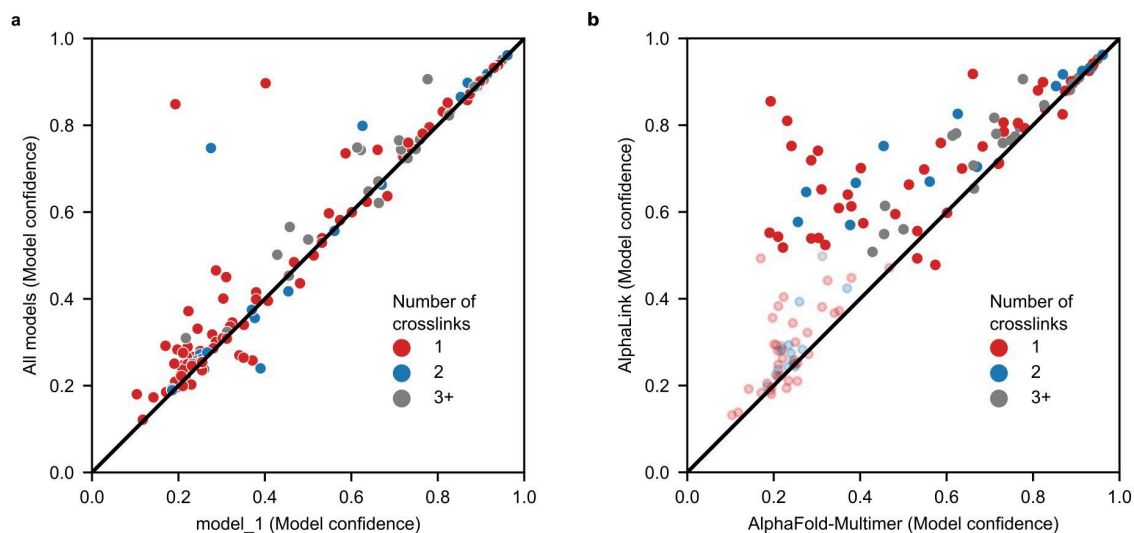
Supplementary Figure 6: Model selection on CASP15 targets

AlphaLink model ranking performance on CASP15 targets (Pearsonr = 0.97, p-value = 3e-51). We plotted the prediction with the highest DockQ score chosen by an oracle against the DockQ score of the prediction with the highest model confidence. We randomly simulated 10 crosslink sets (33 links in the median) and generated 10 predictions per crosslink set with different seeds. One scatter point is one prediction.



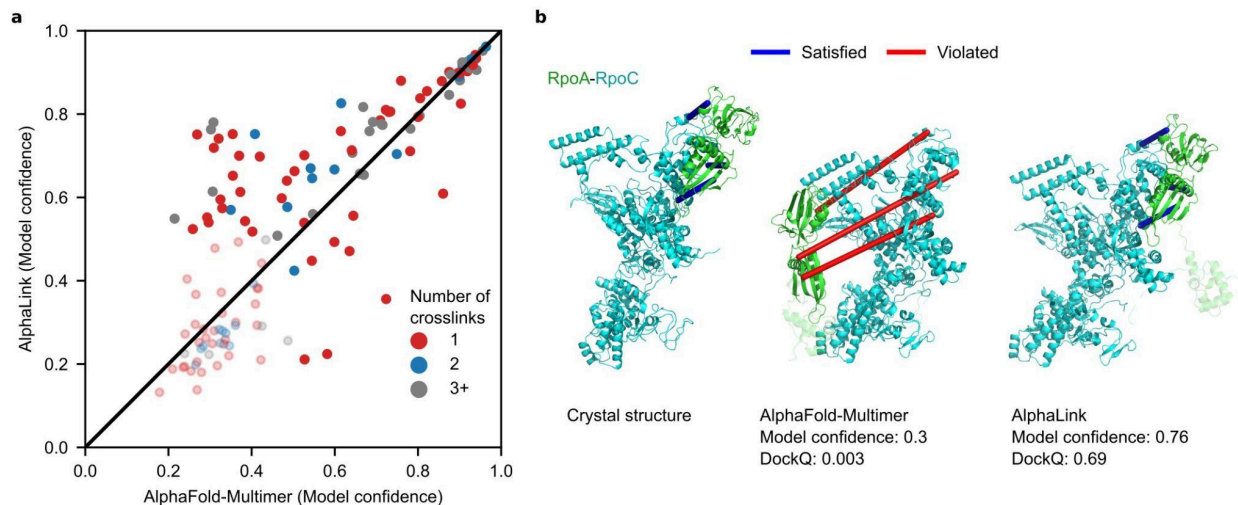
Supplementary Figure 7: AlphaLink performance with and without crosslinks

a DockQ comparison on 8 heteromeric CASP15 targets with simulated SDA crosslinks. The AlphaLink boxplot shows the highest model confidence model ($N = 10$) for 10 randomly sampled crosslink sets. We predict the targets once with crosslinks (red) and once without crosslinks (cyan). AlphaFold-Multimer is shown in grey. The yellow shaded area highlights nanobody, and the red shaded area antibody targets. In all boxplots, the line shows the median and the whiskers represent the 1.5x interquartile range. **b** Comparison of DockQ of AlphaLink without crosslinks and AlphaFold-Multimer on 33 antibody-antigen targets from SAbDab.



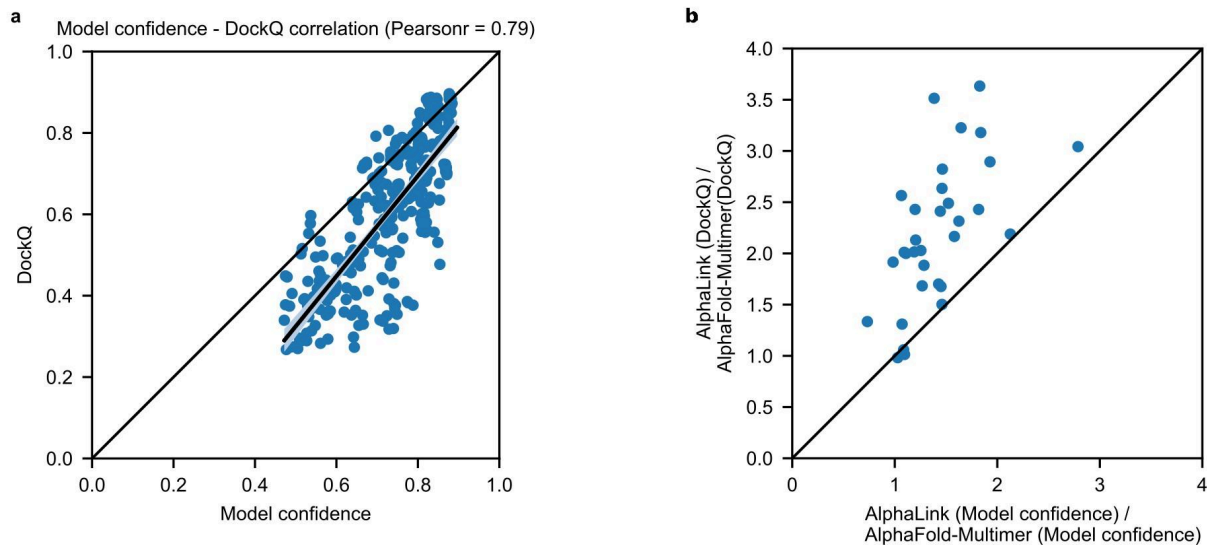
Supplementary Figure 8: Model_1 to all models

a Highest model confidence of AlphaFold-Multimer v2.2 on 134 protein-protein interactions from *Bacillus subtilis* with model_1...model_5 (N = 5) vs predicting the targets only with model_1 (M = 5). Using multiple models improves the model confidence in the mean by 0.03 points. **b** Comparison of model_1 performance for AlphaFold-Multimer v2.2 and AlphaLink v2.2.



Supplementary Figure 9: Comparison with original data from *Bacillus subtilis* paper

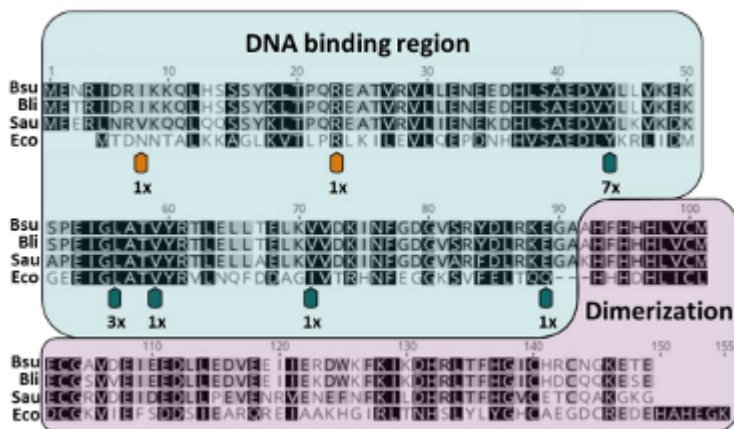
a Highest model confidence of AlphaLink trained on v2.2 weights (N = 5) vs AlphaFold-Multimer trained on v2.1 weights (M = 5) on the *Bacillus subtilis* data on K = 136 dimeric protein-protein interactions. Low confidence predictions for both methods (model confidence < 0.5) have higher transparency. Note that the model confidence might not be directly comparable between different network weights. **b** Comparison of crystal structure (left), AlphaFold-Multimer prediction (middle) and AlphaLink prediction (right) for RpoA-RpoC with annotated crosslinks (blue - satisfied, red - violated).



Supplementary Figure 10: Model confidence - DockQ correlation

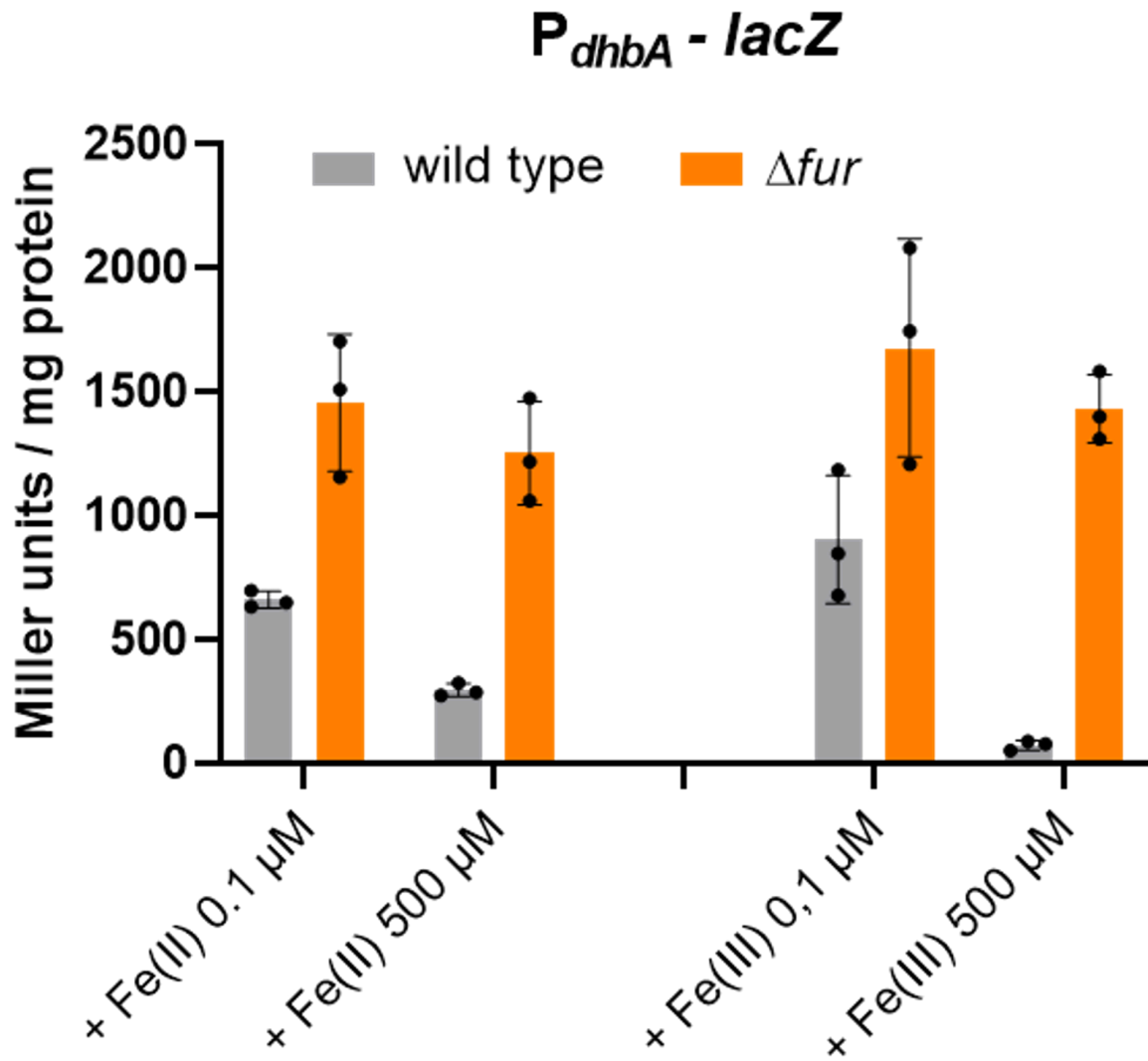
a Comparison of model confidence and DockQ correlation for 33 antibody-antigen targets with 10 crosslink sets per target and five samples each (Pearsonr = 0.79, p-value = 2.9e-72). Best sample is picked by model confidence. The shaded area corresponds to the 95% confidence interval. The line shows the linear fit.

b Plotting the relationship of AlphaLink (DockQ) and AlphaFold-Multimer (DockQ) and the respective model confidences.



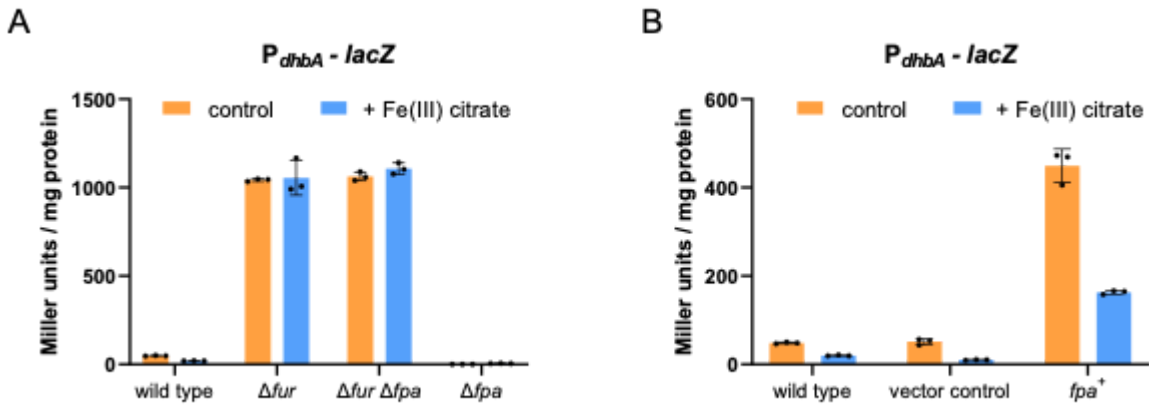
Supplementary Figure 11. Mutations in Fur upon deletion of the *fpa* gene.

Alignment of the sequences of Fur proteins from different bacteria. The DNA-binding and dimerization domains are highlighted in light green and magenta, respectively. The positions of point mutations in the individual suppressor mutants are shown by arrows. The numbers indicate how often an individual amino acid substitution was found. The orange arrows indicate that the corresponding mutant colonies were reddish as the Δfur mutant. Bsu, *B. subtilis*; Bli, *Bacillus licheniformis*; Sau, *S. aureus*; Eco, *E. coli*.



Supplementary Figure 12. Effect of iron on the activity of the *dhbA* promoter.

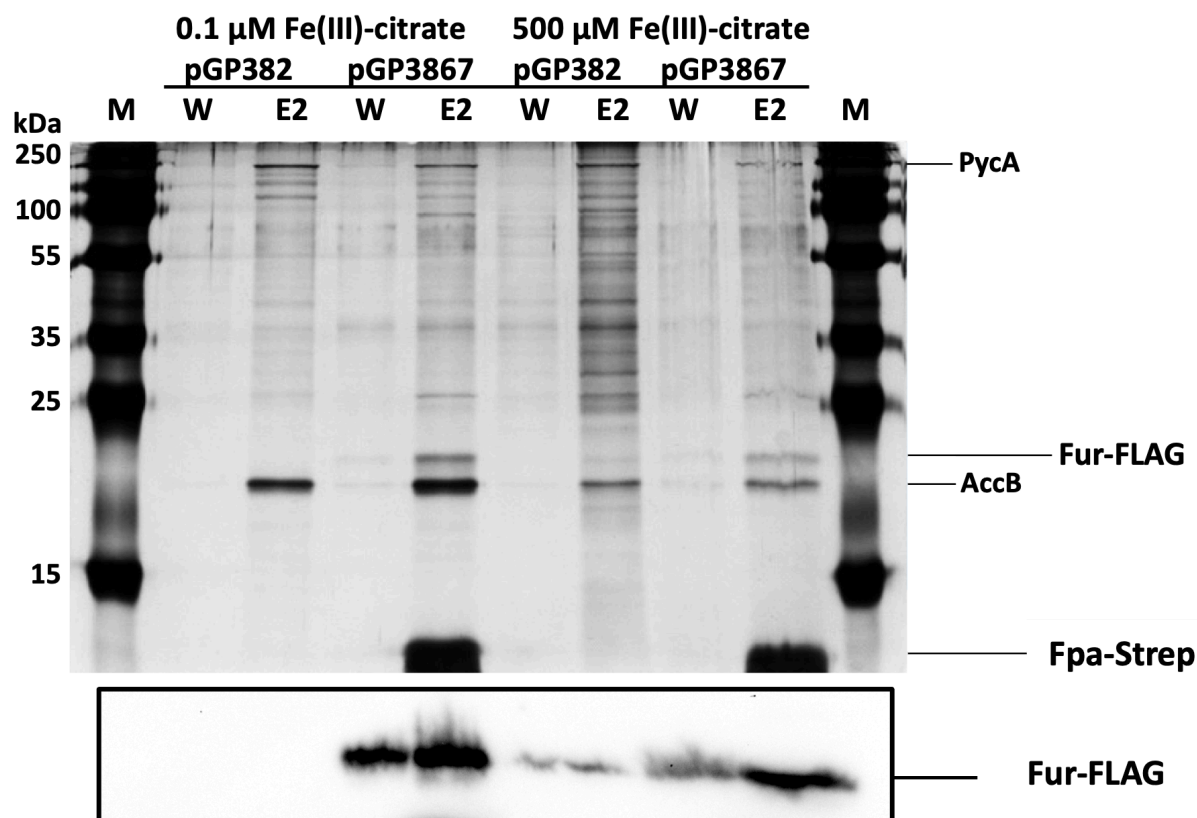
Cultures of a wild type strain (GP3331) and the isogenic Δfur mutant (GP3356) carrying a *dhbA-lacZ* fusion were grown with the indicated iron supplementation, and promoter activities were determined by quantification of β -galactosidase activities. The values are averages of three independent experiments. Standard deviations are shown.



Supplementary Figure 13. The impact of Fpa on *dhbA* promoter activity.

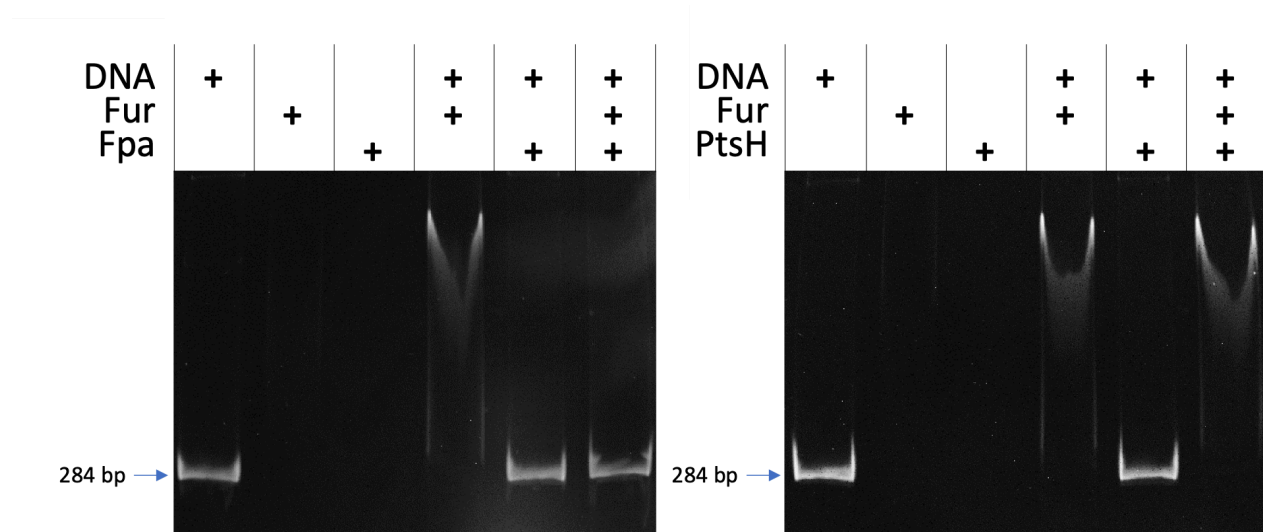
A Effect of *fpa* deletion. Strains carrying a *dhbA-lacZ* fusion were cultivated in LB with or without added ferric citrate, and promoter activities were determined by quantification of β -galactosidase activities. The values are averages of three independent biological replicates. Standard deviations are shown. WT, GP3331; Δfur , GP3356, $\Delta fur \Delta fpa$, GP3361; Δfpa , GP3366.

B Effect of *fpa* overexpression. *B. subtilis* GP3331 without any plasmid (WT) and GP3331 carrying plasmid pGP3897 for *fpa* overexpression or the empty vector pBQ200 were cultivated in LB with or without added ferric citrate, and promoter activities were determined by quantification of β -galactosidase activities. The values are averages of three independent biological replicates. Standard deviations are shown.



Supplementary Figure 14. Physical interaction between Fpa and Fur.

Protein complexes isolated from *B. subtilis* GP3367 (Fur-FLAG) containing either the empty vector pGP382 or pGP3867 (Fpa-Strep). The strains were grown in CSE-Glc minimal medium supplemented with ferric citrate as indicated. The wash and the second elution fractions from each purification were loaded onto the SDS-PAA gel and analyzed by silver staining. The positions of the intrinsically biotinylated proteins PycA and AccB as well as of Fur-FLAG and Fpa-Strep are shown. The lower panel shows a Western blot using antibodies raised against the FLAG tag to detect the Fur-FLAG protein.



Supplementary Figure 15. Fpa is an antagonist of the DNA-binding activity of Fur.

Gel electrophoretic mobility shift assay of Fur binding to *dhbA* promoter fragments. The components added to the assays are shown above the gels.

Supplementary Table 1. *B. subtilis* strains.

Strain	Genotype	Source or Reference
168	<i>trpC2</i>	Laboratory collection
GP879	<i>trpC2</i> $\Delta fur::ermC$	This study
GP3321	<i>trpC2</i> $\Delta fur::ermC$ $\Delta fpa::cat$	This study
GP3324	<i>trpC2</i> $\Delta fpa::cat$	This study
GP3331	<i>trpC2 amyE::</i> (<i>P_{dhbA}</i> - <i>lacZ aphA3</i>)	pGP3594 → 168
GP3356	<i>trpC2</i> $\Delta fur::ermC$ <i>amyE::</i> (<i>P_{dhbA}</i> - <i>lacZ aphA3</i>)	GP879 → GP3331
GP3361	<i>trpC2</i> $\Delta fur::ermC$ $\Delta fpa::cat$ <i>amyE::</i> (<i>P_{dhbA}</i> - <i>lacZ aphA3</i>)	GP3321 → GP3331
GP3366	<i>trpC2</i> $\Delta fpa::cat$ <i>amyE::</i> (<i>P_{dhbA}</i> - <i>lacZ aphA3</i>)	GP3324 → GP3331
GP3367	<i>trpC2 fur</i> -FLAG <i>spc</i>	pGP3899 → 168

Supplementary Table 2. Oligonucleotides used in this study.

Primer	Sequence	Restriction sites
	restriction sites are underlined	
Construction of <i>fur</i> and <i>fpa</i> mutants		
<i>fur</i>-mutant GP879		
HE240	5'-CCTATCACCTCAAATGGTTCGCTGCGTAAGTTTATAGCTGGATGAGTGCAAGTTG	
HE241	5'-GAGGCAGCACACCATCTTTCAGCAC	
HE242	5'-GAAAAAGAAGCGGGGGCATCGATTG	
HE243	5'-CGAGCGCCTACGAGGAATTTGTATCGGGAAATTTAAGATTAAAGATCATAGATTGAC G	
<i>fpa</i>-mutant GP3324		
DR32	5'-GTGAAAAAGGCGAAAAAAGTAACGGACAAGG	
DR33	5'-CCTATCACCTCAAATGGTTCGCTGAAGTATCCCTCCGTTCCCCAGGC	
DR34	5'-CGAGCGCCTACGAGGAATTTGTATCGATGATAAACTCAAACCTATTAACAGTTTGG GTTTTTTTATAACC	
DR35	5'-GTATTTTGGATACTTTCACCGAGCCCC	
FLAG tag constructions		
<i>fur</i>-3xFLAG		

LD35	5'-TTT <u>GGATCC</u> ATGGAAAACCGTATTGATCGTATTAAGAAAC	BamHI
LD38	5'-AAAA <u>AAGCTT</u> TCAGTTTCTTTCCGTTACAGCG	HindIII
Construction of <i>lacZ</i>-fusion promoters		
dhbA-lacZ		
LD01	5'-AAAG <u>AATTC</u> GAGTTTTCAGGAGAAGGACATATTC	EcoRI
LD02	5'-TTT <u>GGATCC</u> AGCTTTTCAGGATTATAATCAACTGCC	BamHI
Expression of His₆-tagged proteins		
His₆-SUMO-Fur		
RB222	5'-AAAG <u>GATCC</u> GCATGGAAAACCGTATTGATCGTATTAAGAAA	BamHI
RB223	5'-TTT <u>CTCGAG</u> CTATTCAGTTTCTTTCCGTTACAGC	XhoI
His₆-Fpa		
FX171	5'-AAAG <u>GATCC</u> TTGGCTTCAGAGATGATAGTTGACCATC	BamHI
FX172	5'-TTT <u>GTCGAC</u> TATTTAGCTGTAAACGCGTCATGCAAAG	Sall
Overexpression of Fpa		
RW114	5'-AAAG <u>GATCC</u> AAAGGAGGAAACAATCTTGGCTTCAGAGATGATAGTTGACC	BamHI
LD23	5'-TTT <u>GTCGAC</u> TATTTAGCTGTAAACGCGTCATGC	Sall
Constructs for two-hybrid interaction screens		

Fur		
LD11	5'-AAAT <u>CTAGAC</u> ATGGAAAACCGTATTGATCGTATTAA	XbaI
LD12	5'-AAAGGTACCGTTCAGTTTCTTTCCGTTACAGCG	KpnI
Fur (K18A R23E Y60E)		
RW848	5'-AAAT <u>CTAGAC</u> ATGGAAAACCGTATTGATCGTATTAAGAAACAACCTGCACTCATCCAG CTATGCACTTACGCCACAGCGTGAAGC	XbaI
LD12	5'-TTT <u>GTCGAC</u> TATTTAGCTGTAAACGCGTCATGCAAAG	Sall
Fpa		
DT49	5'-AAAT <u>CTAGAG</u> TGGCTTCAGAGATGATAGTTGACCATC	XbaI
DT50	5'-TTT <u>GGTACC</u> CGTTTAGCTGTAAACGCGTCATGCAAAGC	KpnI

Supplementary Discussion: The FpaA (YlaN) protein acts as an antirepressor to the ferric uptake regulator Fur in *Bacillus subtilis*

All living organisms require the presence of micronutrients to survive among them iron. As many other essential metal ions, iron is not only required for life, but its accumulation can also have toxic effects.

Since iron is essential but also toxic, intracellular iron levels need to be tightly regulated. This regulation is mediated by Fur, the ferric uptake regulator which is conserved in many archaea and most bacteria. Fur belongs to the ferric uptake regulator family which includes Zur, the major regulator of zinc homeostasis and PerR which responds to peroxide stress. Fur regulates the expression of at about 60 genes in *Bacillus subtilis* (1, 2). Fur almost exclusively represses the expression of its target genes which are mostly iron uptake and siderophore synthesis systems (1). Fur consists of an N-terminal DNA-binding domain and a C-terminal domain which is thought to bind ferrous iron as a cofactor and which mediates dimerization (3, 4). It has long been assumed that Fur binds ferrous iron as a cofactor but only recently it could be shown that Fur actually reversibly binds an iron sulfur cluster in *Escherichia coli* (5).

Even though it could not be shown, it is widely assumed that *B. subtilis* Fur senses the intracellular iron concentration by binding ferrous iron (Fe^{2+}) (6, 7). With decreasing extracellular iron concentrations the Fur regulon is derepressed in three waves in *B. subtilis* (2). First, iron uptake systems for elemental iron, ferric citrate and petrobactin are expressed. Sequentially, the synthesis of the siderophore bacillibactin and uptake systems for bacillibactin and the hydroxamate siderophores to scavenge iron occurs before the iron-sparing response is initiated to inhibit the translation of iron binding proteins.

Despite the absence of direct evidence for the binding of iron to Fur, there has been only little research regarding possible other mechanisms that regulate Fur. However, recent studies revealed that there are proteins which modulate the activity of Fur in different bacteria (8). In *Salmonella enterica*, the EIIA^{Ntr} protein of the non-canonical phosphotransferase system regulates the expression of iron uptake genes via Fur by a direct protein-protein interaction which results in the release of Fur from its DNA binding sites (9). Another but similar mechanism was recently found in uropathogenic *E. coli*, where the proteins YdiV and SlyD cooperatively bind Fur and reduce its DNA binding (10). These results indicate that the regulation of Fur, which was assumed to be exclusively dependent on ferrous iron, actually involves Fur antagonists that might be more common than previously anticipated.

We are interested in the functional characterization of unknown or poorly studied proteins in the model bacterium *B. subtilis*. The YlaN protein is a highly abundant but only poorly studied protein (11, 12). Moreover, the *ylaN* gene is essential under standard growth conditions (13, 14). Both the very high expression and the essentiality suggest that the YlaN protein plays a very important role in the cell. Recently it was shown that the *ylaN* gene becomes dispensable when ferric iron is added to the growth medium (15). This provides a strong indication that the YlaN protein is involved in the control of iron homeostasis. Moreover, recent *in vivo* crosslinking data revealed an intriguing interaction between YlaN and Fur which indicates that YlaN might be another Fur antagonist and thus exert its role in iron homeostasis via Fur (16, 17).

Here, we have studied the role of YlaN in the control of iron homeostasis in *B. subtilis*. We confirm the direct protein-protein interaction between Fur and YlaN, and show that Fur is unable to bind its target DNA in the presence of YlaN. Thus, YlaN acts as Fur protein antagonist which we rename Fpa.

Fpa becomes dispensable in the absence of Fur.

It has been shown that the essential *fpa* (previously *ylaN*) gene becomes dispensable if ferric iron is added to the growth medium (15). Based on the interaction between Fpa and the Fur regulator protein, we hypothesized that Fpa might antagonize Fur to allow the expression of iron uptake systems at low iron concentrations. If this were true, the deletion of the *fpa* gene under standard conditions might be toxic as a result of continued repression of the genes for iron uptake by Fur. To test this idea, we attempted the deletion of *fpa* in the *fur* mutant GP879. As a control, we used the isogenic wild type strain *B. subtilis* 168. In agreement with the published data, *fpa* could not be deleted under standard conditions or if the plates were supplemented with ferrous iron. However, the deletion was possible if the medium was supplemented with ferric iron (Fe^{3+}). In contrast, the *fpa* gene could be deleted under all three conditions in the *fur* mutant strain GP879. These results confirm the conditional essentiality of Fpa depending on the iron supply and they suggest that Fpa might be needed to prevent some harmful activity of Fur under conditions of iron limitation.

Fpa is quasi-essential under standard growth conditions.

When the concept of essentiality was introduced, a gene was regarded essential if it could not be activated under standard growth conditions for the organism, *i. e.* on LB medium supplemented with glucose at 37°C for *B. subtilis* (14). Today, we know that several essential genes can in fact be deleted under standard conditions but that the mutant strains immediately acquire suppressor mutations that help to overcome the growth-limiting problem. Such genes are called quasi-essential as they remain essential under standard conditions in the genomic context of the standard wild type strain.

The possibility to delete *fpa* from a *fur* mutant suggested that *fpa* might also be quasi-essential. Thus, we made use of the *fpa* mutant GP3324 that had been isolated in the presence of ferric iron as described above. The strain was cultivated on complex medium under iron-limiting condition. As expected, we observed the development of few individual colonies that most likely resulted from the acquisition of suppressor mutations. We re-isolated eight independent colonies that had appeared in the absence of added iron or in the presence of ferrous iron each, and sequenced their *fur* alleles as we already knew that *fur* mutants tolerate the inactivation of *fpa*. Of 15 individual mutants, all had single point mutations in the Fur coding sequence that resulted in amino acid substitutions in the Fur protein. These mutations all occurred in the N-terminal part of the protein which is required for DNA binding indicating that the mutated Fur versions are impaired in DNA binding (3). Several of the substitutions were observed in multiple mutants, both from selection without added iron or in the presence of ferrous iron (see Supplementary Fig. 11).

These results demonstrate that *fpa* is a novel quasi-essential gene. Under standard conditions on complex medium, the deletion of *fpa* causes the immediate acquisition of suppressor mutations that interfere with Fur activity. Moreover, these results support the idea that the Fur repressor becomes toxic for the cells in the absence of Fpa and iron, *i.e.* under conditions that require expression of the iron uptake genes that are all members of the Fur regulon.

Fpa affects the activity of the Fur-controlled *dhbA* promoter.

The *dhbACEBF-ybdZ* operon encodes the enzymes for the synthesis of the siderophore bacillibactin (18). It is one of the Fur-controlled operons, and it is induced as a part of the second wave of Fur-controlled genes that are expressed upon iron limitation (2). We used the activity of the *dhbA* promoter as an indicator of Fur activity. For this purpose, we fused the *dhbA* promoter region to a promoterless *lacZ* gene encoding β -galactosidase and integrated this *dhbA-lacZ* fusion into the *B. subtilis* genome of the wild type strain and the isogenic Δfur mutant GP879. The resulting strains, GP3331 and GP3356, respectively, carrying the *dhbA-lacZ* fusion were cultivated in CSE-Glc minimal medium supplemented with 0.1 μ M and 500 μ M of ferrous and ferric iron. As shown in Supplementary Fig. 12, a strong promoter activity was determined at 0.1 μ M iron, irrespective of the redox state of the iron ions. Expression was substantially reduced at the increased iron concentration in the wild type strain. For ferrous iron, we observed an about twofold decrease of expression, whereas a twelvefold decrease was detected in the case of ferric iron. This result corresponds well to the observation that ferrous iron is less bioavailable for the bacteria, so that it causes only a weak repression of *dhbA* expression. In the *fur* mutant GP3356, we observed high β -galactosidase under all tested conditions. The strong repression of *dhbA* expression in response to ferrous iron, and the dependence of repression on a functional Fur protein is in excellent agreement with published data (19). We also tested the expression of the *dhbA-lacZ* fusion in the wild type strain GP3331 in complex medium (LB) in the presence and absence of 500 μ M ferrous or ferric iron. In this case, the activity was already rather low in the absence of added iron (45 unit per mg of protein as compared to 700 ... 900 units in minimal medium in the presence of 0.1 μ M of iron), probably due to the presence of iron in the complex medium which is likely to cause a basal repression. However, the expression was fivefold repressed if one of the iron ions was added (see Supplementary Fig. 13A for Fe^{3+} , data not shown for Fe^{2+}). For further experiments, we used LB medium and ferric iron.

To test the role of Fpa in the Fur-mediated regulation of the *dhbA* promoter, we constructed additional strains carrying the *dhbA-lacZ* fusion in a *fpa* single and *fpa fur* double mutant. The resulting strains were GP3366 and GP3361, respectively. As shown in Supplementary Fig. 13A, expression in the *fur* mutant was strongly enhanced as compared to the wild type background and independent from the iron concentration. The rather weak expression in the wild type strain in the absence of added iron is in good agreement with the conclusion that this medium already contains some iron. The loss of Fpa in addition to Fur in GP3361 had no effect on the *dhbA* promoter activity. Finally, we tested *dhbA* promoter activity in the *fpa* mutant (GP3366). For this strain, only background activities were observed under both conditions. We conclude that Fpa exerts its role via Fur, as already suggested as a result from the suppression analysis. The lack of promoter activity in the *fpa* single mutant is in excellent agreement with the above conclusion that the Fur-controlled genes for iron uptake might be completely repressed in the absence of Fpa thus resulting in the essentiality of Fpa in standard LB medium.

The deletion of *fpa* resulted in complete repression of the *dhbA* promoter, likely due to the inability to release Fur from its target DNA in the absence of Fpa. Given the observed physical interaction between the two proteins, it is tempting to speculate that the overexpression of Fpa might cause the release of Fur from its target sites and thus induction of *dhbA* expression even in the presence of iron. To test this hypothesis, we put the *fpa* gene under the control of the strong *degQ^{hy}* promoter in the expression plasmid pBQ200. The resulting plasmid was pGP3897. This plasmid as well as the empty vector pBQ200 were then introduced into strain GP3331 that harbors the *dhbA-lacZ* fusion. Again, the strains were grown in LB medium in the presence or absence of 500 μ M ferric citrate. As observed before

for the wild type strain, we found a fivefold repression of *dhbA* expression in the presence of the empty vector (see Supplementary Fig. 13B). In contrast, expression was strongly increased in the presence of pGP3897 when Fpa was overproduced. We even observed a substantial expression in this strain if 500 μ M ferric citrate were present. This observation indicates that the overexpression of Fpa counteracts the repressing effect caused by Fur and strongly suggests that Fpa acts as a *bona fide* antagonist of Fur.

Physical interaction between Fur and Fpa. Our previous proteome-wide interaction studies with *B. subtilis* detected an interaction between Fur and Fpa (17). If Fpa acts as an antagonist to Fur, it seems most likely that this activity is achieved by the physical interaction between the two proteins. To confirm this interaction, we decided to verify the interaction *in vivo* in *B. subtilis* by co-precipitation. For this purpose, we constructed a strain that expressed Fur carrying a C-terminal FLAG tag for immunological detection (GP3367). This strain was then transformed either with the empty vector pGP382 (20) or with plasmid pGP3867 for the expression of Fpa fused to a C-terminal Strep-tag for affinity chromatography. Both strains were cultivated in CSE-Glc minimal medium with 0.1 or 500 μ M ferric citrate. The protein extracts were then passed over a StrepTactin column to bind the Fpa-Strep, washed and Strep-tagged proteins with their potential interaction partners were eluted. Two proteins, PycA and AccB, were eluted from the StrepTactin column for both the empty vector control and the strain expressing Fpa-Strep. These proteins contain a biotin cofactor that causes binding to the matrix and are good indicators that the experimental setup was suitable. Upon expression of Fpa-Strep, we observed copurification of a protein of about 20 kDa which corresponds to FLAG-tagged Fur. The identity of the band was confirmed by a Western blot using antibodies directed against the FLAG tag (see Supplementary Fig. 14). The Fur protein was copurified with Fpa both at high and low iron concentrations. Again, the low iron concentration used in this experiment may still be sufficient to allow the interaction between the two proteins.

Fpa prevents the DNA-binding activity of Fur. All our experiments support our initial hypothesis that Fpa can bind Fur and interfere with the repression of target genes by Fur. To get direct evidence for this, we performed DNA binding assays with the *dhbA* promoter region and purified Fur protein. As shown in Supplementary Fig. 15, the DNA fragment was retarded in the presence of Fur, indicating binding of Fur to its target DNA. In contrast, no shift was observed with the purified Fpa protein. The addition of both Fur and Fpa to the DNA did not result in DNA binding. This observation is in excellent agreement with the idea that Fpa might prevent Fur from binding to DNA. To ensure that the effect of Fpa addition to the DNA and Fur is specific, we performed a control experiment in which we used the PtsH protein of the phosphotransferase system as the second protein in the assay. As Fpa, PtsH is a small acidic protein. The results observed with the promoter DNA and Fur were as described above. Similarly, the PtsH protein did not interact with the DNA. However, the presence of PtsH in addition to the promoter fragment and Fur did not prevent the formation of the Fur-DNA complex indicating that PtsH is unable to interfere with the DNA binding activity of Fur (Supplementary Fig. 15). Taken together, our results demonstrate that Fpa specifically interacts with Fur to prevent it from binding to its target DNA sequences and thus to allow the expression of genes that are under negative control by Fur.

Conclusion

The Fpa (YlaN) protein belongs to a small group of so far unknown proteins that are strongly expressed under essentially all conditions in *B. subtilis* (12). Of those about 40 proteins, Fpa is the only that is

essential under standard laboratory growth conditions (13). The data presented in this study identify the so far unknown protein Fpa as a *bona fide* antirepressor of the Fur regulator. The interaction between the two proteins interferes with the binding of Fur to its DNA targets and thus results in the expression of the iron uptake systems which are all subject to Fur-mediated transcription repression.

Fpa perceives the primary signal of the system, the intracellular iron concentration in the form of ferrous iron (21). In this form, the protein does not interact with Fur, and Fur can bind to its DNA target sites to repress the expression of genes for iron uptake systems and to activate expression of an iron exporter gene. In contrast, under conditions of iron limitation, apo-Fpa can bind to Fur, and thus break open the Fur dimer, resulting in release of Fur from its target sites as shown in this work. This interaction allows the expression of Fur-controlled genes for iron uptake systems if iron gets scarce, and is thus a prerequisite for the growth of *B. subtilis* under conditions of iron depletion. This explains why the *fpa* gene is essential for *B. subtilis* under standard conditions. Since iron limitation is the rule rather than the exception for bacteria that live under aerobic conditions, mechanisms that allow the effective induction of genes for iron acquisition are of key importance.

Fur-mediated control of iron homeostasis is widespread in both Gram-negative and Gram-positive bacteria. For long time, it was assumed that the Fur protein directly responds to the presence of ferrous iron (22). However, despite intensive research there is no clear support for this idea in the published data body and the direct sensing of iron by Fur has remained controversial (23, 24). In contrast, there is clear evidence that zinc ions act as cofactor for the regulator of zinc homeostasis Zur, another Fur-type regulator (25). Only more recent studies with Gram-negative bacteria and the data presented here suggest that Fur may be controlled by regulatory protein-protein interactions in many bacteria (8).

An analysis of the phylogenetic distribution of Fpa reveals that the protein is present exclusively in the Bacilli subgroup of the Firmicutes (26). In this class, Fpa is present in most species with the exception of the lactic acid bacteria and few other species. Interestingly, most bacteria that possess Fur family transcription factors, encode multiple, typically three of these proteins, Fur, PerR, and Zur. Those bacteria of the Bacilli class that lack Fpa, do also lack the Fur protein. Most of them have PerR and Zur, with the notable exception of *Lactobacillus acidophilus* and *Streptococcus pneumoniae* that completely lack Fur type regulators. In the genus *Jeotgalibacillus*, one species, *J. malaysiensis* possesses both Fur and Fpa, whereas *J. donkookensis* encodes neither of the two proteins. There are only two bacteria among the Bacilli that seem to possess Fur but not Fpa, *Aneurinibacillus soli* CB4 and *Tumebacillus avium* AR23208. This might result from issues with the genome sequences, or these bacteria have evolved specific strategies to control Fur activity.

The two-faced role of iron as an important player in cellular energy metabolism on one hand and its toxicity on the other make the presence of effective systems to control iron homeostasis critical for bacterial life. As a result of Fur/Fpa-dependent regulation, the genes for iron uptake are not expressed if the metal is already abundant in the cell. The antirepressor activity of Fpa allows expression of iron uptake genes as soon as iron gets limiting. This is a strategic decision of the cell as it determines which protein of the iron homeostasis system will be present or not in the future.

The discovery of the activity of Fpa as an antirepressor to Fur in *B. subtilis* is important for our better understanding of the physiology of this important model organism (27). *B. subtilis* is the model organism for a large group of Gram-positive bacteria, that includes many important pathogens such as *S.*

aureus, *Listeria monocytogenes*, or *Bacillus anthracis* that also possess the Fur/Fpa couple. Iron is the growth-limiting factor for most pathogenic bacteria in the human body. Accordingly, the investigation of the control of iron homeostasis is also very important to better understand the processes of infection and disease and to develop novel treatments. Since Fpa is essential under conditions of iron limitation not only in *B. subtilis* but also in *S. aureus* (21), it might be an attractive novel target for drug development.

Supplementary Methods

Genome sequencing. To identify the mutations in the suppressor mutant strain GP3368 (see Supplementary Table 1), the genomic DNA was subjected to whole-genome sequencing. Concentration and purity of the isolated DNA was first checked with a Nanodrop ND-1000 (PeqLab Erlangen, Germany), and the precise concentration was determined using the Qubit® dsDNA HS Assay Kit as recommended by the manufacturer (Life Technologies GmbH, Darmstadt, Germany). Illumina shotgun libraries were prepared using the Nextera XT DNA Sample Preparation Kit and subsequently sequenced on a MiSeq system with the reagent kit v3 with 600 cycles (Illumina, San Diego, CA, USA) as recommended by the manufacturer. The reads were mapped on the reference genome of *B. subtilis* 168 (GenBank accession number: NC_000964) (28). Mapping of the reads was performed using the Geneious software package (Biomatters Ltd., New Zealand) (29). Frequently occurring hitchhiker mutations (30) and silent mutations were omitted from the screen. The resulting genome sequence was compared to that of our in-house wild type strain. Single nucleotide polymorphisms were considered as significant when the total coverage depth exceeded 25 reads with a variant frequency of $\geq 90\%$. All identified mutations were verified by PCR amplification and Sanger sequencing.

In vivo detection of protein-protein interactions. For the expression of Fpa carrying a C-terminal Strep-tag in *B. subtilis*, we used plasmid pGP3867. This plasmid was obtained by cloning the *fpa* gene between the BamHI and Sall sites of the expression vector pGP382 (20). To add a FLAG tag epitope to the Fur protein, we constructed plasmid pGP3899 by cloning the *fur* gene between the BamHI and HindIII sites of pGP1331 (31). Transformation of the wild type strain 168 with pGP3899 yielded strain GP3367. To verify the interaction between Fur and Fpa *in vivo*, cultures of *B. subtilis* GP3367 (Fur-FLAG) containing pGP3867 (Fpa-Strep), or the empty vector control (pGP382), were cultivated in 500 ml CSE-Glc medium containing the indicated iron source until exponential growth phase was reached ($OD_{600} \sim 0.4-0.6$). The cells were harvested immediately and stored at -20°C . The Strep-tagged protein and its potential interaction partners were then purified from crude extracts using a StrepTactin column (IBA, Göttingen, Germany) and D-desthiobiotin as the eluent. The eluted proteins were separated on an SDS gel and potential interacting partners were analyzed by staining with Colloidal Coomassie and Western blot analysis using antibodies raised against the FLAG-tag.

Supplementary References

1. Baichoo N, Wang T, Ye R, Helmann JD. 2002. Global analysis of the *Bacillus subtilis* fur regulon and the iron starvation stimulon. *Mol Microbiol.* 45:1613–1629.
2. Pi H, Helmann JD. 2017. Sequential induction of Fur-regulated genes in response to iron limitation in *Bacillus subtilis*. *Proc Natl Acad Sci USA* 114:12785–12790.
3. Coy M, Neilands JB. 1991. Structural dynamics and functional domains of the Fur protein. *Biochemistry* 30:8201–8210.
4. Stojiljkovic I, Hantke K. 1995. Functional domains of the *Escherichia coli* ferric uptake regulator protein (Fur). *MGG Mol Gen Genet.* 247:199–205.
5. Fontenot CR, Tasnim H, Valdes KA, Popescu CV, Ding H. 2020. Ferric uptake regulator (Fur) reversibly binds a [2Fe-2S] cluster to sense intracellular iron homeostasis in *Escherichia coli*. *J Biol Chem.* 295:15454–15463.
6. Pinochet-Barros A, Helmann JD. 2017. Redox sensing by Fe²⁺ in bacterial Fur family metalloregulators. *Antioxid Redox Signal.* 29:1858-1871.
7. Helmann JD. 2014. Specificity of metal sensing: iron and manganese homeostasis in *Bacillus subtilis*. *J Biol Chem.* 289:28112–28120.
8. Steingard CH, Helmann JD. 2023. Meddling with metal sensors: Fur-family proteins as signaling hubs. *J Bacteriol.* 205:e00022-23.
9. Choi J, Ryu S. 2019. Regulation of iron uptake by fine-tuning the iron responsiveness of the iron sensor Fur. *Appl Environ Microbiol.* 85:e03026-18.
10. Zhang F, Li B, Dong H, Chen M, Yao S, Li J, Zhang H, Liu X, Wang H, Song N, Zhang K, Du N, Xu S, Gu L. 2020. YdiV regulates *Escherichia coli* ferric uptake by manipulating the DNA-binding ability of Fur in a SlyD-dependent manner. *Nucleic Acids Res.* 48:9571–9588.
11. Nicolas P, Mäder U, Dervyn E, Rochat T, Leduc A, Pigeonneau N, Bidnenko E, Marchadier E, Hoebeke M, Aymerich S, Becher D, Bisicchia P, Botella E, Delumeau O, Doherty G, Denham EL, Fogg MJ, Fromion V, Goelzer A, Hansen A, Härtig E, Harwood CR, Homuth G, Jarmer H, Jules M, Klipp E, Chat LL, Lecointe F, Lewis P, Liebermeister W, March A, Mars RAT, Nannapaneni P, Noone D, Pohl S, Rinn B, Rügheimer F, Sappa PK, Samson F, Schaffer M, Schwikowski B, Steil L, Stülke J, Wiegert T, Devine KM, Wilkinson AJ, Dijk JM van, Hecker M, Völker U, Bessières P, Noirot P. 2012. Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science* 335:1103–1106.
12. Wicke D, Meißner J, Warneke R, Elfmann C, Stülke K. 2023. Understudied proteins and understudied functions in the model bacterium *Bacillus subtilis* – A major challenge in current research. *Mol Microbiol.* 120:8-19.
13. Hunt A, Rawlins JP, Thomaidis HB., Errington J. 2006. Functional analysis of 11 putative essential genes in *Bacillus subtilis*. *Microbiology* 152:2895–2907.
14. Commichau FM, Pietack N, Stülke J. 2013. Essential genes in *Bacillus subtilis*: a re-evaluation after ten years. *Mol Biosyst.* 9:1068-1075.
15. Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, Hawkins JS, Lu CHS, Koo BM, Marta E, Shiver AL, Whitehead EH, Weissman JS, Brown ED, Qi LS, Huang KC, Gross CA. 2016. A

- comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell* 165:1493–1506.
16. de Jong L, Roseboom W, Kramer G. 2021. A composite filter for low FDR of protein-protein interactions detected by in vivo cross-linking. *J Proteomics* 230:103987.
 17. O'Reilly FJ, Graziadei A, Forbrig C, Bremenkamp R, Charles K, Lenz S, Elfmann C, Fischer L, Stülke J, Rappsilber J. 2023. Protein complexes in cells by AI-assisted structural proteomics. *Mol Syst Biol.* 19:e11544.
 18. May JJ, Wendrich TM, Marahiel MA. 2001. The *dhb* operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. *J Biol Chem.* 276:7209–7217.
 19. Bsai N, Helmann JD. 1999. Interaction of *Bacillus subtilis* Fur (ferric uptake repressor) with the *dhb* operator *in vitro* and *in vivo*. *J Bacteriol.* 181:4299–4307.
 20. Herzberg C, Weidinger LAF, Dörrbecker B, Hübner S, Stülke J, Commichau FM. 2007. SPINE: A method for the rapid detection and analysis of protein-protein interactions *in vivo*. *Proteomics.* 7:4032–4035.
 21. Boyd JM, Esquilin-Lebrón K, Campbell CJ, Kaler KR, Norambuena J, Foley ME, Stephens TG, Rios G, Mereddy G, Zheng V, Bovermann H, Kim J, Kulczyk AW, Yan JH, Greco TM, Cristea IM, Carabetta VJ, Beavers WN, Bhattacharya D, Skaar EP, Parker D, Carroll RK, Stemmler TL. 2023. YlaN is an ionic iron binding protein that functions to relieve Fur-mediated repression of gene expression in *Staphylococcus aureus*. *BioRxiv.* <https://doi.org/10.1101/2023.10.03.560778>.
 22. Troxell B, Hassan HM. 2013: Transcriptional regulation by ferric uptake regulator (Fur) in pathogenic bacteria. *Front Cell Infect Microbiol.* 3:59.
 23. Ma Z, Faulkner MJ, Helmann JD. 2012. Origins of specificity and cross-talk in metal ion sensing by *Bacillus subtilis* Fur. *Mol Microbiol.* 86:1144–1155.
 24. Chandrangu P, Rensing C, Helmann JD. 2017. Metal homeostasis and resistance in bacteria. *Nat Rev Microbiol.* 15:338–350.
 25. Ma Z, Gabriel SE, Helmann JD. 2011. Sequential binding and sensing of Zn(II) by *Bacillus subtilis* Zur. *Nucleic Acids Res.* 39:9130–9138.
 26. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. 2021. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* 49:D274–D281.
 27. Stülke J, Grüppen A, Bramkamp M, Pelzer S. 2023. *Bacillus subtilis*, a swiss army knife in science and biotechnology. *J Bacteriol.* 205:e00102–23.
 28. Barbe V, Cruveiller S, Kunst F, Lenoble P, Meurice G, Sekowska A, Vallenet D, Wang T, Moszer I, Médigue C, Danchin A. 2009. From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology* 155:1758–1775.
 29. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.
 30. Reuß DR, Faßhauer P, Mroch PJ, Ul-Haq I, Koo BM, Pöhlein A, Gross CA, Daniel R, Brantl S, Stülke J. 2019. Topoisomerase IV can functionally replace all type 1A topoisomerases in *Bacillus subtilis*.

Nucleic Acids Res. 47:5231–5242.

31. Lehnik-Habrink M, Pförtner H, Rempeters L, Pietack N, Herzberg C, Stülke J. 2010. The RNA degradosome in *Bacillus subtilis*: Identification of Csha as the major RNA helicase in the multiprotein complex. *Mol Microbiol.* 77:958–971.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| ☐ | ☒ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection AlphaFold v2.2 and v2.3 (<https://github.com/deepmind/alphafold>), Uni-Fold v2.1.0 (<https://github.com/dptech-corp/Uni-Fold>), AlphaLink2 v1.0 (<https://github.com/Rappsilber-Laboratory/AlphaLink2>)

Data analysis PyMOL 2.5, ChimeraX 1.7.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The AlphaLink models generated in this study based on experimental crosslinks have been deposited as integrative/hybrid models in PDB-Dev under accession codes PDBDEV_00000221 and PDBDEV_G_1000003. We used previously published data for the experimental validation. Cullin4: <https://www.ebi.ac.uk/pride/archive/projects/PXD020453>

EMD-10611 (core), EMD-10612 (conformational state-1), EMD-10613 (state-2) and EMD-10614 (state-3). Bacillus Subtilis: JPST001796 and PXD035508; JPST001797 and PXD035519; JPST001791 and PXD035362.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Use the terms *sex* (biological attribute) and *gender* (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.

Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.

Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Reporting on race, ethnicity, or other socially relevant groupings

Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).

Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)

Please provide details about how you controlled for confounding variables in your analyses.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

Data exclusions

Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Replication

Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.

Randomization

Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.

Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).

Research sample	State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.
Sampling strategy	Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.
Data collection	Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.
Timing	Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Non-participation	State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.
Randomization	If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.
Research sample	Describe the research sample (e.g. a group of tagged <i>Passer domesticus</i> , all <i>Stenocereus thurberi</i> within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.
Sampling strategy	Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.
Data collection	Describe the data collection procedure, including who recorded the data and how.
Timing and spatial scale	Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Reproducibility	Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.
Randomization	Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.
Blinding	Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	<i>Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.</i>
Validation	<i>Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.</i>

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	<i>State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.</i>
Authentication	<i>Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.</i>
Mycoplasma contamination	<i>Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.</i>
Commonly misidentified lines (See ICLAC register)	<i>Name any commonly misidentified cell lines used in the study and provide a rationale for their use.</i>

Palaeontology and Archaeology

Specimen provenance	<i>Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.</i>
Specimen deposition	<i>Indicate where the specimens have been deposited to permit free access by other researchers.</i>
Dating methods	<i>If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.</i>
☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.	
Ethics oversight	<i>Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.</i>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	<i>For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.</i>
Wild animals	<i>Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.</i>
Reporting on sex	<i>Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify</i>

reasons for lack of sex-based analysis.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.

Study protocol

Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection

Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.

Outcomes

Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
☒	☐ Public health
☒	☐ National security
☒	☐ Crops and/or livestock
☒	☐ Ecosystems
☒	☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
☒	☐ Demonstrate how to render a vaccine ineffective
☒	☐ Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐ Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐ Increase transmissibility of a pathogen
☒	☐ Alter the host range of a pathogen
☒	☐ Enable evasion of diagnostic/detection modalities
☒	☐ Enable the weaponization of a biological agent or toxin
☒	☐ Any other potentially harmful combination of experiments and agents

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.

ChIP-seq

Data deposition

- ☐ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- ☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission

Provide a list of all files available in the database submission.

Genome browser session

(e.g. [UCSC](#))

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.

Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software

Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

Confirm that:

- ☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☐ All plots are contour plots with outliers or pseudocolor plots.
- ☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.

Instrument

Identify the instrument used for data collection, specifying make and model number.

Software

Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.

Cell population abundance

Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.

Gating strategy

Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

- ☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	<i>Indicate task or resting state; event-related or block design.</i>
Design specifications	<i>Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.</i>
Behavioral performance measures	<i>State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).</i>

Acquisition

Imaging type(s)	<i>Specify: functional, structural, diffusion, perfusion.</i>
Field strength	<i>Specify in Tesla</i>
Sequence & imaging parameters	<i>Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.</i>
Area of acquisition	<i>State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.</i>
Diffusion MRI	☐ Used ☐ Not used

Preprocessing

Preprocessing software	<i>Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).</i>
Normalization	<i>If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.</i>
Normalization template	<i>Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.</i>
Noise and artifact removal	<i>Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).</i>
Volume censoring	<i>Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.</i>

Statistical modeling & inference

Model type and settings	<i>Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).</i>
Effect(s) tested	<i>Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.</i>
Specify type of analysis:	☐ Whole brain ☐ ROI-based ☐ Both
Statistic type for inference	<i>Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.</i>
(See Eklund et al. 2016)	
Correction	<i>Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).</i>

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).
Graph analysis	Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).
Multivariate modeling and predictive analysis	Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.

File Name: Supplementary Movie 1

Description: Supplementary movie depicting the conformation change of Fur-Fpa

Modelling protein complexes with crosslinking mass spectrometry and deep learning

Corresponding Author: Professor Juri Rappsilber

Figures originally included in the author's rebuttal have been redacted from this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The study by Stahl, Brock, and Rappsilber describes the inclusion of distance constraints from cross-linking mass spectrometry in the machine learning-based structure prediction tool AlphaFold Multimer. The authors have previously implemented this for single-protein structure prediction by AlphaFold2. With the present additions, the previous tool, AlphaLink, additionally considers crosslinks between protein pairs when choosing the highest-scoring structure prediction for multiprotein complexes. The authors demonstrate improved structure prediction using several multiprotein complexes with experimentally determined structures, using DockQ scores as read-out. They then apply the extended AlphaLink to a previously published whole-cell crosslinking dataset and demonstrate that even single-crosslinks can improve the accuracy of the predicted structures.

The study is clearly described and (at least according to what I can judge from my expertise) conforms to the highest standards of technical accuracy. I fully believe the authors' claims, the issue with the work is rather that it feels very much like an extension of previous work (even if it required changing the base and training the network), and the most significant finding, that is, how distance restraints can improve AF predictions, has already been reported (Stahl et al, Nature Biotech 2023). In this sense, the study would be more suitable for a specialist journal.

To me, the real novelty of the study is its application to whole-cell crosslinking data, which, as the authors note themselves in the abstract, opens the possibility of whole-cell structural investigations that are biased by experimental data. However, the corresponding section in the manuscript is extremely short, and basically only states that even a single crosslink improves model accuracy. Is there a way to extract more biologically relevant information from the improved models, particularly when compared to the group's previous work in this area (O'Reilly et al, Ref 5)? If the authors can show that the improved models translate to novel biological insights, that would greatly elevate the general interest.

Reviewer #2

(Remarks to the Author)

General comments.

The method presented (AlphaLink) is an interesting extension of the authors' previous work on single chain complexes. However, too little data is used for evaluation and the results compared to other computational methods are not satisfactory as AlphaLink is outperformed in most cases (4 vs 3 cases where AlphaLink is better). Another similar method called ColabDock also exists (<https://www.biorxiv.org/content/10.1101/2023.07.04.547599v1>), but was not used in the evaluation.

It is important to consider data overlaps between all different methods that have been integrated into AlphaLink. This is not assessed. The methods section is very short and is lacking important information about fine-tuning (e.g. loss curves?) the data sets used, their overlaps and compositions. In general, a description of exactly what information goes into AlphaLink and how it is provided is lacking.

Significant work is needed to assess the issues with evaluation and explaining exactly what advancements AlphaLink provides compared to other available methods.

Specific comments.

1. The legend in Figure 1a mentions predictions from AlphaFold, but what is meant here is really AlphaFold-multimer(?) which is not the same method. This is an issue throughout the manuscript which makes it difficult to follow.
2. Seven hard targets from CASP15 are evaluated. Four of these are nanobody interactions and two are antibody interactions. Compared to the best performance in CASP15, AlphaLink only outperforms predictions from AlphaFold-multimer in three cases. In four cases (H1129, H1134, H1140, H1141) does AFsample (the best method in CASP15) outperform AlphaLink.
3. The star indicating the top ranked model from CASP15 is missing for H1167 so this may really only be true for two targets (H1142 and H1166). The failure to outperform available computational methods raises questions for the utility of AlphaLink.
4. I assume that the reason that AlphaLink only outperforms purely computational methods in the antibody-antigen cases is because there is very little (and different) coevolution there. AlphaLink may still be useful in these cases, but the data presented is too little to properly assess this.
5. AlphaLink uses simulated crosslinks from XWalk, but no mention of what data this method is trained on and the overlap to the evaluation set is presented.
6. AlphaLink is fine-tuned starting from AFM on a dataset with simulated cross-links, but no information on the composition of this or the overlap to the CASP15 targets or the training set of AFM is provided.
7. To prove the utility of AlphaLink, I suggest extending the analysis significantly. Extract a dataset of antibody-antigen interactions which have not been seen by AFM during training and compare the performance. It is not possible to assess the performance on so few targets, especially since there is much more data available (1000s of structures in SAbDab: <https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabdab>). Other methods for antibody-antigen structure prediction exist as well and these should also be included in the evaluation (e.g. <https://www.nature.com/articles/s41467-023-38063-x>).
8. Figure 1b. This plot is confusing (the axis labels are the same). What are you really plotting here? I suggest plotting the model confidence vs the DockQ after selecting for the highest model confidence and after running a single prediction. This is shown for H1134 in Extended data Figure 2 where one can see that the model confidence correlates poorly with the DockQ (although it is hard to tell since all 100 predictions are plotted on top of each other).
9. Figure 2a. Again, the same axis-labels. I am not sure what this plot is supposed to display, but being "better" is arbitrary. It would be more informative to plot the model confidence vs the number of cross links and analyse the correlation between the two. This together with model confidence vs DockQ answers the questions of if the confidence measure can be trusted and if the cross-links really help.

In addition, plot the model confidence vs the DockQ and colour by the number of cross links.

Reviewer #3

(Remarks to the Author)

The authors have submitted a natural extension of Alphalink to the modeling of heteromeric complexes. They demonstrate that high quality models can be generated in a fraction of the time of a "normal" AlphaFold Multimer modeling event. This claim is established with a very nice simulation/sampling of crosslinking data but also backed up in select cases with real (sparse!) data collected from other studies.

In general I find this a very useful and exciting addition, and a very realistic one. Computational methods will likely (always) benefit from the inclusion of real structural data and crosslinking MS is a natural source for such information.

I have only a few matters to raise:

1. As I say, the simulations are a nice way to illustrate the benefits of Alphalink, but some additional information is needed. If I understand correctly, the authors use a span of ~25 angstroms and random multiple samplings of 10 crosslinks, from a set that samples 10% of sequence (at a 20% FDR). I like the challenge presented by these numbers. However, does the sampling of the crosslinks reflect the bias that we often see in real data? First at the level of residue (K much more preferred than STY) and second at the level of structure (links clustered in one area). Additional discussion around the limitations of the simulation would be valuable.
2. The first paragraph after the introduction strikes a rather discordant note. Any claim of "vastly outperforming" should be made after the data are presented, to avoid biasing the reader.
3. And I contest the very next statement that claims also that the results are similar or better on CASP15 targets than the best performing algorithms that participated in CASP15. Figure 1a shows a mixed bag, with 50% of the CASP15 algorithms outperforming Alphalink. Please adjust the language on this.
4. And more importantly on this topic, what is to be learned about how and why these algorithms outperform Alphalink? I notice in Extended Data Figure 6 that strongly increasing the sampling/recycling helps only a bit (on one example at least).

The authors are perfectly positioned to comment on benefits/weaknesses of approaches.

5.Figures:

- a. 2C. Is there an actual interaction between RpoA and RpoC in the Alphafold model? The interface is obscured by the crosslinks. Can this be made more visible?
- b. 3D. I'm a bit puzzled by this success. Can the authors comment how the crosslinking preserved an open state when Alphafold (which Alphalink is based on) drives a closed form. Near as I can tell, the orientation of the lower hinge looks the same for both Alphafold and Alphalink. Can you prevent an interaction if crosslinks occupy an interface? Would be nice to see some commentary on this and a supporting figure with the crosslinks mapped.
- c. Extended Data figure 1. Better axis labels would be nice.
- d. Extended Data figure 6. What does "more" mean exactly? Better labels here too, please.

Author Rebuttal letter:

Reviewer #1 (Remarks to the Author):

The study by Stahl, Brock, and Rappsilber describes the inclusion of distance constraints from cross-linking mass spectrometry in the machine learning-based structure prediction tool AlphaFold Multimer. The authors have previously implemented this for single-protein structure prediction by AlphaFold2. With the present additions, the previous tool, AlphaLink, additionally considers crosslinks between protein pairs when choosing the highest-scoring structure prediction for multiprotein complexes. The authors demonstrate improved structure prediction using several multiprotein complexes with experimentally determined structures, using DockQ scores as read-out. They then apply the extended AlphaLink to a previously published whole-cell crosslinking dataset and demonstrate that even single-crosslinks can improve the accuracy of the predicted structures.

The study is clearly described and (at least according to what I can judge from my expertise) conforms to the highest standards of technical accuracy. I fully believe the authors's claims, the issue with the work is rather that it feels very much like an extension of previous work (even if it required changing the base and training the network), and the most significant finding, that is, how distance restraints can improve AF predictions, has already been reported (Stahl et al, Nature Biotech 2023). In this sense, the study would be more suitable for a specialist journal.

This work builds on previous work rather than extending it, as it wasn't obvious that the monomer approach would translate. Protein complex prediction is much more difficult because in addition to predicting the individual chains, predicting interactions makes it necessary to search a space that grows exponentially with the size of the complexes. Crosslinks are obviously super helpful here because they can cut down the search space. But the nature of the data is very different. Evolutionary data is generally sparse in the interface and the SDA crosslinks used in this study are lower in resolution (< 25Å CA-CA) compared to the photo-AA crosslinks used in the previous work (< 15Å CA-CA) which are much closer to co-evolutionary data, the main source of information. Because of this, the neural network has to do more heavy lifting here. We added text to point out these challenges.

To me, the real novelty of the study is its application to whole-cell crosslinking data, which, as the authors note themselves in the abstract, opens the possibility of whole-cell structural investigations that are biased by experimental data. However, the corresponding section in the manuscript is extremely short, and basically only states that even a single crosslink improves model accuracy. Is there a way to extract more biologically relevant information from the improved models, particularly when compared to the group's previous work in this area (O'Reilly et al, Ref 5)? If the authors can show that the improved models translate to novel biological insights, that would greatly elevate the general interest.

We respectfully disagree. This work constitutes a surprising and substantial advancement over the current technical state of the field. We would like to raise the reviewer's attention to the fact that a single crosslink from in-cell data can substantially improve model accuracy. Whether in such a proof-of-concept study also novel biology was produced or not is of no relevance to the technical advancement. We feel not alone in this view as Nature Communications published many papers reporting on computational progress without new biology, e.g., <https://www.nature.com/articles/s41467-022-33729-4> and <https://www.nature.com/articles/s41467-023-38063-x>. Also, Nature Methods did so: <https://www.nature.com/articles/s41592-024-02174-0>. Asking us to add biology comes as a surprising decision in this context, also because it is based on the comment of a reviewer that is neither skilled to judge the challenges that we had to overcome in our technical solution nor the significance of our finding for the structure prediction field.

Anyway, we extended the manuscript by adding work on YlaN-Fur to experimentally confirm our structure. The structural model predicted by AlphaLink, which agrees with the crosslinking data, shows how the YlaN-Fur interaction regulates iron homeostasis. YlaN engages the DNA interaction domain of Fur and splits the Fur homodimer, rendering Fur incompetent to bind DNA, repressing the genes necessary for iron uptake. We introduce point mutations in Fur based on our YlaN-Fur complex structure to successfully destroy the predicted interaction and validate our structural and mechanistic model of YlaN-mediated regulation of iron homeostasis in *B. subtilis*.

Reviewer #2 (Remarks to the Author):

General comments.

The method presented (AlphaLink) is an interesting extension of the authors' previous work on single chain complexes. However, too little data is used for evaluation and the results compared to other computational methods are not satisfactory as AlphaLink is outperformed in most cases (4 vs 3 cases where AlphaLink is better). Another similar method called ColabDock also exists (<https://www.biorxiv.org/content/10.1101/2023.07.04.547599v1>), but was not used in the evaluation.

We deliberately kept the parameters (and input data) in Fig. 1a in line with the baseline method AlphaFold-Multimer to show that including crosslinking MS data improves the prediction over the baseline. The other methods in CASP15, e.g., AFSample used as much as 2400x the number of samples in comparison to AlphaLink, in addition to other changes. We slightly increased the parameters now for AlphaLink to show a fairer comparison. The original comparison is moved to the supplement. The results further improved and show much more parity on the cases where AlphaLink was outperformed. Note that the approach taken by AFSample will not scale well for larger complexes since sampling will become prohibitively expensive, e.g., the prediction of a 6000 AA complex takes around 30 hours for a single sample. Being able to reduce the number of samples significantly by focusing the sampling on the interesting regions becomes more and more important with increasing complex sizes. That being said, the strategy taken by AFSample is compatible with AlphaLink and could be adopted. Also note that no single method in CASP15 outperformed all other methods.

We extended the evaluation to 33 challenging antibody-antigen complexes, here AlphaLink also outperforms AlphaFold-Multimer by a large margin.

We have compared our approach with all methods participating in CASP15. ColabDock appeared one month after us on BioRxiv and constitutes work that followed our work.

It is important to consider data overlaps between all different methods that have been integrated into AlphaLink. This is not assessed. The methods section is very short and is lacking important information about fine-tuning (e.g. loss curves?) the data sets used, their overlaps and compositions. In general, a description of exactly what information goes into AlphaLink and how it is provided is lacking.

Indeed, it is important to avoid potential biases and overfitting. We did early stopping on the validation set. We extended the data set description in methods to clarify the cutoff dates and exclude/ indicate potential overlaps. For the fine-tuning, we already show the results of AlphaLink with and without crosslinking MS data to make sure that the observed performance improvements are not due to additional training and using bigger crops (in comparison to AlphaFold-Multimer v2, Extended Data Figure 7). There is no data overlap between the AlphaLink and AlphaFold-Multimer training set and CASP15 targets, the new SAbDab targets, or Cullin4; since the training sets predate the data. Except for a couple of targets, there are no crystal structures for *Bacillus subtilis* complexes, therefore there is only a potential overlap with self-distillation. The targets that have crystal structures might have been in the training set of AlphaFold-Multimer but this is negligible because it would affect both AlphaFold-Multimer and AlphaLink in the same way.

Significant work is needed to assess the issues with evaluation and explaining exactly what advancements AlphaLink provides compared to other available methods.

Specific comments.

1. The legend in Figure 1a mentions predictions from AlphaFold, but what is meant here is really AlphaFold-multimer(?) which is not the same method. This is an issue throughout the manuscript which makes it difficult to follow.

Sorry for the confusion! We had defined in the manuscript that AlphaFold refers to AlphaFold-multimer. To avoid confusion, we explicitly call it now AlphaFold-Multimer or AF-Multimer in short.

2. Seven hard targets from CASP15 are evaluated. Four of these are nanobody interactions and two are antibody interactions. Compared to the best performance in CASP15, AlphaLink only outperforms predictions from AlphaFold-multimer in three cases. In four cases (H1129, H1134, H1140, H1141) does AFSample (the best method in CASP15) outperform AlphaLink. We didn't make sufficiently clear that this experiment is a control to show the improvement over baseline AlphaFold-multimer on the same input data and with the same setting. Even though AFSample uses baseline AlphaFold-multimer (+ dropout at inference time) the settings are vastly different from the standard settings in the baseline.

We can control the comparison with AlphaFold-multimer since the MSAs and templates are available. We show that AlphaLink outperforms AlphaFold-Multimer on all eight hard CASP15 targets (in the median) with the same input features and number of recycling iterations, even though AlphaFold-Multimer uses slightly more samples (25 vs 10) and more models (5 vs 1).

We cannot control for AFSample since the input data is not available and we lack the compute power. AFSample diverged greatly from the standard settings. To be more comparable, we increased the number of samples from 10 to 200 and the recycling iterations from 3 to 20 which is still much lower than the up to 24000 samples for AFSample. We moved the control experiment to the supplement.

Note that we could adopt the AFSample setting and sampling strategy in AlphaLink because AFSample makes only minor changes to standard AlphaFold-Multimer (dropout during inference) and is therefore compatible with AlphaLink. But this approach doesn't scale well. It quickly becomes infeasible to generate large amounts of samples. For example, Cullin4 takes 16 h to predict. Integrating crosslinks on the other hand, allows us to focus on the interesting regions and to drastically reduce the amount of samples required (see H1166 where AFSample failed with 24000 samples).

3. The star indicating the top ranked model from CASP15 is missing for H1167 so this may really only be true for two targets (H1142 and H1166).

The was unfortunately concealed by the AlphaFold-multimer box. We fixed it.

The failure to outperform available computational methods raises questions for the utility of AlphaLink.

In our new data AlphaLink is at the same level or better than the best submission to CASP15. No alternative approach achieved this, as the best prediction is not from the same lab.

4. I assume that the reason that AlphaLink only outperforms purely computational methods in the antibody-antigen cases is because there is very little (and different) coevolution there. AlphaLink may still be useful in these cases, but the data presented is too little to properly assess this.

This is indeed the underlying premise of AlphaLink. Combining experimental data (crosslinking MS) and co-evolutionary information will improve challenging targets, esp., those with little co-evolutionary data. We added a new dataset (Fig. 1b) with more challenging targets to highlight the utility of AlphaLink.

5. AlphaLink uses simulated crosslinks from XWalk, but no mention of what data this method is trained on and the overlap to the evaluation set is presented.

There is no overlap because XWalk is not trained on anything. It simulates crosslinks based on a given PDB.

6. AlphaLink is fine-tuned starting from AFM on a dataset with simulated cross-links, but no information on the composition of this or the overlap to the CASP15 targets or the training set of AFM is provided.

There is no overlap, both training sets (AlphaLink and AlphaFold-Multimer) are from

2021, predating the benchmark data. We added the cutoff dates to the methods section.

7. To prove the utility of AlphaLink, I suggest extending the analysis significantly. Extract a dataset of antibody-antigen interactions which have not been seen by AFM during training and compare the performance. It is not possible to assess the performance on so few targets, especially since there is much more data available (1000s of structures in SAbDab: <https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabdab>).

Unfortunately, most structures in the SAbDab were likely part of the training set of AlphaFold-Multimer and would constitute overlap. We compiled a new data set with 33 antibody-antigen targets from the last two years, which were released after the training set of both AlphaFold-Multimer and AlphaLink. To simplify evaluation, we only used targets which have a single biological assembly.

Other methods for antibody-antigen structure prediction exist as well and these should also be included in the evaluation (e.g. <https://www.nature.com/articles/s41467-023-38063-x>).

The suggested methods are specialised methods that solve a different problem, still e.g., in the case of IgFold, it was outperformed by AlphaFold-Multimer. We (and AlphaFold-Multimer) solve the general problem of predicting interactions. The antibody-antigen targets were interesting to us because they are challenging due to the limited evolutionary information.

8. Figure 1b. This plot is confusing (the axis labels are the same). What are you really plotting here? I suggest plotting the model confidence vs the DockQ after selecting for the highest model confidence and after running a single prediction. This is shown for H1134 in Extended data Figure 2 where one can see that the model confidence correlates poorly with the DockQ (although it is hard to tell since all 100 predictions are plotted on top of each other). Figure 1b compared the DockQ scores of two AlphaLink settings with one another, see labels inside the plot. We removed this plot and replaced it with new data on targets from SAbDab.

It was shown in the AlphaFold-Multimer paper that the ipTM (> 0.6) correlates well with the DockQ score which in turn is close to the model confidence. Extended Data Fig. 6 shows the correlation between the pick with the highest model confidence and the pick with the best DockQ score.

Extended Data Fig. 4 shows an example where the correlation is poor but crosslink satisfaction can help to identify good predictions.

9. Figure 2a. Again, the same axis-labels. I am not sure what this plot is supposed to display, but being 'better' is arbitrary. It would be more informative to plot the model confidence vs the number of cross links and analyse the correlation between the two. This together with model confidence vs DockQ answers the questions of if the confidence measure can be trusted and if the cross-links really help.

In addition, plot the model confidence vs the DockQ and colour by the number of cross links.

Figure 2a compares the model confidence of AlphaFold (x-axis) and AlphaLink (y-axis). The regions are labelled accordingly inside the plot. To make it clearer, we now include the method names directly in the labels. We removed 'better' from the labelling.

Unfortunately, for *Bacillus subtilis* there are no crystal structures for protein complexes except for few cases, which is why we cannot report the DockQ score and have to resort to model confidence. Figure 2c is an example of a crystal structure being available, demonstrating that AlphaLink indeed improves the prediction quality over AlphaFold-Multimer.

We added new plots that show model confidence $\leftrightarrow$ DockQ score correlation on the challenging SAbDab targets (Extended Data Fig. 10a). Here, model confidence is often overestimated. A second plot shows the relationship between the model confidences of AlphaLink / AlphaFold-Multimer and the resulting DockQ scores (Extended Data Fig. 10b). A better model confidence also results in a better DockQ score AlphaLink over AlphaFold-Multimer.

Most of the targets have only one or two crosslinks (see coloring in Figure 2a), it would

be hard to show a correlation here. The impact of the crosslinking data also depends on the evolutionary data and the position of the links, although in general, with an increase in crosslinks you would see an improvement in prediction quality due to the larger coverage.

Reviewer #3 (Remarks to the Author):

The authors have submitted a natural extension of Alphalink to the modeling of heteromeric complexes. They demonstrate that high quality models can be generated in a fraction of the time of a normal AlphaFold Multimer modeling event. This claim is established with a very nice simulation/sampling of crosslinking data but also backed up in select cases with real (sparse!) data collected from other studies.

In general I find this a very useful and exciting addition, and a very realistic one. Computational methods will likely (always) benefit from the inclusion of real structural data and crosslinking MS is a natural source for such information.

I have only a few matters to raise:

1. As I say, the simulations are a nice way to illustrate the benefits of Alphalink, but some additional information is needed. If I understand correctly, the authors use a span of ~25 angstroms and random multiple samplings of 10 crosslinks, from a set that samples 10% of sequence (at a 20% FDR). I like the challenge presented by these numbers. However, does the sampling of the crosslinks reflect the bias that we often see in real data? First at the level of residue (K much more preferred than STY) and second at the level of structure (links clustered in one area). Additional discussion around the limitations of the simulation would be valuable.

Our simulation could indeed be further improved by reflecting these biases, we currently uniformly subsample the crosslink candidates. This doesn't really affect training because for the model it will not matter if the crosslink originates from K or STY, it's more important to see sufficient data to get enough training signal. And we can see that it translates to the real data. However, it may result in a better coverage in testing than what we can normally expect.

2. The first paragraph after the introduction strikes a rather discordant note. Any claim of "vastly outperforming" should be made after the data are presented, to avoid biasing the reader.

We removed the claim.

3. And I contest the very next statement that claims also that the results are similar or better on CASP15 targets than the best performing algorithms that participated in CASP15. Figure 1a shows a mixed bag, with 50% of the CASP15 algorithms outperforming Alphalink. Please adjust the language on this.

Our new data based on more sampling and recycling leads to an improvement that when contrasted to the CASP15 outcome supports our wording: "Integrating simulated SDA crosslinks in the modelling of eight challenging heteromeric CASP15 targets (H1129, H1134, H1140, H1141, H1142, H1144, H1166, H1167) substantially improved the DockQ15 score from 0.14 to 0.62 on average, compared to the AlphaFold-Multimer baseline (Fig. 1a) which matches the average DockQ = 0.62 of the best predictions in CASP15. For comparison, AFSample16, one of the top performing methods in CASP15, averages a DockQ score of 0.56."

4. And more importantly on this topic, what is to be learned about how and why these algorithms outperform Alphalink? I notice in Extended Data Figure 6 that strongly increasing the sampling/recycling helps only a bit (on one example at least). The authors are perfectly positioned to comment on benefits/weaknesses of approaches.

We re-ran the experiment with parameters that are closer to the other groups in CASP15 to have a fairer comparison. Extended Data Figure 7 (now removed since it is contained in the new Fig. 1a) showed the results only for H1134 for most of the other targets there are larger improvements. We added a discussion of the specific targets where other approaches outperform AlphaLink and possible reasons.

5. Figures:

a. 2C. Is there an actual interaction between RpoA and RpoC in the AlphaFold model? The interface is obscured by the crosslinks. Can this be made more visible?

There is no interaction in the AlphaFold model. We rotated the structures to improve visibility.

b. 3D. I am a bit puzzled by this success. Can the authors comment how the crosslinking preserved an open state when AlphaFold (which AlphaLink is based on) drives a closed form. Near as I can tell, the orientation of the lower hinge looks the same for both AlphaFold and AlphaLink. Can you prevent an interaction if crosslinks occupy an interface? Would be nice to see some commentary on this and a supporting figure with the crosslinks mapped.

We would like to stress that understanding AI is a very large and active research field that is not close to our current competencies. While AlphaLink bases on AlphaFold, it differs in two ways. It is a different network as it has been re-trained regarding its refinement step (with crosslinks added). It also uses crosslinks from the open conformation.

We left out SamHD because it doesn't have a crystal structure. It's located close to the hinge and will affect the placement since there are also crosslinks. We include SamHD now in the figure.

c. Extended Data figure 1. Better axis labels would be nice.

Fixed!

d. Extended Data figure 6. What does "more" mean exactly? Better labels here too, please. We removed this plot now because the new results in Fig. 1a with increased recycling/sampling supersede it.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have addressed my comments.

Reviewer #2

(Remarks to the Author)

Reviewer #3

(Remarks to the Author)

The authors have satisfactorily addressed all my concerns. The additional models provided improve the claims that they make.

Author Rebuttal letter:

To say that AFSample doesn't scale is not a very strong argument as the alternative is obtaining crosslinking MS data for a complex of interest which may well be impossible for proteins expressed in low quantities and if substantial overexpression is permitted the proteins may simply interact due to the reason of abundance. Regardless of how long it takes to run AFSample, running a few thousand iterations of predictions will be both faster and cheaper than setting up a lab, buying a MS and expressing the complex of interest.

If ColabDock appeared one month later this is acceptable.

We completely agree that a purely computational strategy holds the advantage over experimentation because its entry barrier is substantially lower. Nonetheless, crosslink studies are being undertaken by an increasing number of labs. Once one has crosslink data, so we argue and show, it is useful to leverage it also for modelling. Crosslinking offers complementary information to evolutionary data, and can thus be leveraged also by an approach like AFSample. Many structural biology labs are willing to invest substantial resources to isolate complexes for cryoEM, all those are very much amenable to crosslinking and then AlphaLink modelling, if size permits. Who knows, maybe even top computational labs like the Wallner lab might find it attractive to align with an experimental lab to invest into acquiring crosslink data should they find interest in specific biological challenges.

To say that there is no overlap between the different datasets due to them being from different dates shows how little the authors understand about the concept of benchmarking which is

evident throughout the presented work. It is not only important to create a useful method but also to demonstrate its performance in a meaningful way. Date cutoffs show no information regarding the overlap between training and test sets. What is important is to be able to predict new complexes, not reiterate already known ones. No proof for the claim that the potential overlaps are negligible are presented. To ensure that the method is meaningful you absolutely need to display the performance on unseen targets (no homology <20% seqid or better TMscore<0.5).

As the reviewer points out, there are different levels of overlap (e.g., targets being part of the training set, homologous targets in the training set, using templates in prediction). We clarified in the text what we are talking about and what possible overlaps are present. Importantly, any overlap improves the AF baseline performance and thus works against us by reducing the possible gains from crosslinks. The goal is to show that integrating crosslinking MS data into AlphaFold-Multimer can improve the prediction quality on challenging targets. In our direct comparison with AlphaFold-Multimer, both methods use the same input data (minus or plus crosslinking MS data). The AlphaLink training set includes some newer targets in comparison to AlphaFold-Multimer v2.2. We showed that the additional training (fine-tuning) of AlphaLink wasn't the reason for the increased performance (Extended Data Figure 7) but it's the addition of crosslinking MS data which substantially improves the prediction quality on many targets over AlphaFold-Multimer (Fig. 1).

The CASP15 targets we benchmark on include unseen targets since they consist of TBM/FM targets, meaning there are only partial templates available for a subunit and / or the assembly. Antibody-antigen targets are challenging in their own right due to the lower evolutionary signal. Running a job for 16h is not very long. How long does it take to get new cross links for a complex of interest? This is the question to be asked to answer how useful the proposed method is.

A couple of days. It's 16h for a single prediction. So in this example, 1000 samples will take almost 2 years to compute on a single A100 GPU. The runtime scales cubically.

We would disagree. This is not the question to ask to determine the 'usefulness'. It is certainly one aspect though and will help with larger complexes where sampling becomes much more expensive. Crosslinking MS data provide additional information and add some degree of experimental validation to the prediction. They are complementary to evolutionary information which for many interesting targets are lacking/ sparse (e.g., antibody-antigen complexes). Fig 1b) demonstrates the usefulness for these targets.

While the primary goal is not to convert computational titans, such as Wallner's group, into crosslink-focused facilities, it is important to demonstrate the significant advantages of utilizing crosslink data in structural biology. As more labs acquire crosslink data, AlphaLink can effectively harness this information to enhance the modelling of complexes. Our work demonstrates that integrating crosslink data, already increasingly common in structural biology, can lead to more accurate and insightful predictions in AlphaLink.

Perhaps you can expand a little on what XWalk does (I am not familiar with the package). As I understand it, it uses structural factors to infer crosslinks. What structures do you input to the package and what information from these structures is used to 'simulate' the crosslinks? If I interpret correctly, you input the native complexes that you later use for evaluation(?). There is therefore obvious bias in this evaluation. You can't input information from the native complex structures and then claim that you predict them and I see little utility in being able to do so. We use XWalk to computationally generate crosslink data for the training and evaluation complexes. Generating experimental data for all these complexes is currently unfeasible. Please realise that we need to get crosslink data somehow for our approach. XWalk identifies residue pairs in the native complex that could be crosslinkable (here with SDA, < 25 Å) based on solvent accessibility and peptide digestion. We select at random a fraction of these pairs and add noise in line with real experiments. This computational crosslink data resembles closely that of real experiments.

Again, cutoff dates are not meaningful here. If you want to prove that your method actually works, you can't do so on structures homologous to those seen by AFM - regardless of when they were released. Please partition the data properly and then do an evaluation. As it is now, it is not possible to assess the utility of your method on new targets which I presumed was the idea.

We find cutoff dates meaningful here because they ensure that we did not directly train on targets from CASP15 or included templates of these targets (same with SAbDab). The evaluation setup is therefore exactly the same as it was in CASP15 which helps to assess the utility of our method, namely improving prediction quality on challenging targets with crosslinking MS data. The CASP15 benchmark targets consist of TBM/ FM targets, meaning only partial templates are available. Repartitioning the data only works if you actually train from scratch, which is both impractical and unfeasible (resources) and would disqualify AlphaFold-Multimer as a comparison point. We would like to stress again that we wanted to demonstrate the additional value offered by adding crosslink information and therefore conduct a comparison. Any advantage for AlphaFold baseline is simply reducing the impact of crosslinks. So, any bias will

work against us and not for us. The positive effects we observe are despite any biases from structure overlap.

This can (and should) be easily checked.

AlphaFold was trained on all targets before the cutoff and may include some homologs after the cutoff date.

Again with the date cutoffs. It seems that you really do not understand the importance of partitioning data for ML. If the data is similar between the two dates, it is impossible to say how well your method generalises.

Please see above.

Redo the data partitioning. This is something you have to consider before training. I can't accept the release of another paper that simply ignores overlaps between train and test sets and instead relies on what day in the week it is (date splits). Too many papers in this space have been published without considering these issues which you and others (editors among them) seem completely unaware of.

Redoing the data partitioning only works if you actually train from scratch which we have not done. As said above, we compare AlphaFold-Multimer with AlphaLink2 with the weights of AlphaFold-Multimer as a starting point for fine-tuning, so essentially AF +/- crosslinks.

How can you claim that the correlation is good when it is not? It is evident that the confidence is high regardless of DockQ score:

[Redacted]

If you plot confidence vs DockQ as I suggest you will see this more clearly. The model confidence is high all over the place and even at DockQ scores as low as 0.1 the confidence is 0.7-0.8.

We stated that it was shown in the AlphaFold-Multimer paper that the ipTM (> 0.6) correlates well with the DockQ score, which in turn is close to the model confidence. This plot actually shows the ranking performance. We fixed the title and removed the colouring by model confidence. The confidence vs DockQ is plotted in Extended Data Figure 10a.

Great that you here finally admit to this. This should be in the main text as a strong limitation. Currently, there is no way to separate accurate from inaccurate predictions making the utility of the predicted models questionable. A reason why methods like AFM and AFsample are useful is that the confidence can be used to select accurate models from many predictions.

Extended Data Figure 6 shows that the model confidence can be used to select accurate models from many predictions. Most of the time, the top ranked prediction is very close to the best prediction, and this gets better with higher model confidence (similar to AlphaFold-Multimer, ipTM > 0.6).

Our direct comparison with AlphaFold-Multimer shows that a higher model confidence for AlphaLink seems to indicate higher model quality (Extended Data Fig. 10b).

Incidentally, AFsample seems to be a poor example. The two plots in the AFsample paper (Fig 1. c,e) show overall a fairly bad correlation between DockQ and ranking_confidence.

If this is the case, please provide support for this statement by calculating and visualising the score differences divided by the number of cross links. You could e.g. plot the score distributions having 1,2,... crosslinks.

Our claim is that adding available crosslinks to AlphaFold-Multimer is generally beneficial. This claim is substantiated by our presented data. That more information leads to better models has been proven in many instances. We have not extended this to crosslinks in our manuscript, however, because of the unfavourable cost/benefit ratio.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>