High-throughput computational discovery of inhibitory protein fragments with AlphaFold

Andrew Savinov[1,*,†], Sebastian Swanson[1,*], Amy E. Keating[1,2,3], Gene-Wei Li[1]

[1]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

[2]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

[3]Koch Center for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA

*equal contributions; [†]to whom correspondence should be addressed (email: savinov@mit.edu)

Abstract

Peptides can bind to specific sites on larger proteins and thereby function as inhibitors and regulatory elements. Peptide fragments of larger proteins are particularly attractive for achieving these functions due to their inherent potential to form native-like binding interactions. Recently developed experimental approaches allow for high-throughput measurement of protein fragment inhibitory activity in living cells. However, it has thus far not been possible to predict *de novo* which of the many possible protein fragments bind their protein targets, let alone act as inhibitors. We have developed a computational method, FragFold, that employs AlphaFold to predict protein fragment binding to full-length protein targets in a high-throughput manner. Applying FragFold to thousands of fragments tiling across diverse proteins revealed peaks of predicted binding along each protein sequence. These predictions were compared with experimentally measured peaks of inhibitory activity in *E. coli*. We establish that our approach is a sensitive predictor of protein fragment function: Evaluating inhibitory fragments derived from known protein-protein interaction interfaces, we found 87% were predicted by FragFold to bind in a native-like mode. Across full protein sequences, 68% of FragFold-predicted binding peaks match experimentally measured inhibitory peaks. This is true even when the underlying inhibitory mechanism is unclear from existing structural data, and we find FragFold is able to predict novel binding modes for inhibitory fragments of unknown structure, explaining previous genetic and biochemical data for these fragments. The success rate of FragFold demonstrates that this computational approach should be broadly applicable for discovering inhibitory protein fragments across proteomes.

Introduction

Peptides have great potential for modulating function in living cells by binding to and regulating target proteins. Indeed, organisms have evolved a large array of polypeptides of ≤100 amino acids, known variously as biological peptides, small proteins, miniproteins, or microproteins (1–4). Evolved miniproteins can be as structurally simple as single alpha helices (5–7) and often control protein function in response to stress conditions (5–7). Synthetic random peptides have also been shown to produce diverse biological activities (11–13), and recently, it has become feasible to design miniproteins and functional peptides, with many exciting developments in this space (14–16).

Peptides that can emulate, modulate, and rewire native protein-protein interactions – and the general features that permit such activities – are therefore of great interest. One class of peptides with an inherent potential for this kind of functionality are protein fragments derived from full-length sequences. Massively parallel measurements of fragment-based inhibition *in vivo* allow a systematic interrogation of such peptides (17–20). This approach revealed that protein fragments can act as dominant negative inhibitors. Savinov *et al.* demonstrated that these fragments titrate parental protein interactions, as well as other principles of fragment-based inhibition, including the role of protein fragment length (20). Notably, generic biophysical properties of protein fragments (such as charge, hydrophobicity, and secondary structure) explained very little of the fragment-to-fragment variation in inhibitory activity. These observations suggest that specific features of fragments and their bespoke binding interfaces largely determine which fragments are potent inhibitors.

A computational approach to predicting protein fragment binding would be a powerful tool to discover inhibitory fragments across proteomes and provide structural models for inhibitory interactions. Such an approach would complement high-throughput measurements of peptide function. Experimental measurements would provide the ground truth, allowing development of a model, and computational prediction would expand the reach of experimental methods by pre-identifying likely inhibitory regions of proteins for screening. Predicting binding using AlphaFold (21) is attractive, as this model was trained using native structural data (22) and coevolutionary sequence information (e.g., ref. (23)) to extract features of natural proteins. AlphaFold has already shown promise for predicting binding modes for peptides engaging proteins (24–27), importantly including peptides not present in the training set (24, 27).

Savinov *et al.* demonstrated an association between AlphaFold-predicted native-like protein-peptide binding and measured inhibitory activity for 6 protein fragments of *Escherichia coli* FtsZ, a structural protein essential for cell division (20). Here we build on that work by applying high-throughput AlphaFold computational predictions to 2,277 tiling protein fragments covering the full sequences of six diverse proteins, mirroring the massively parallel experimental measurements (20). Analysis of the results revealed that computational predictions of protein-fragment binding are in striking agreement with the *in vivo* inhibitory activities of protein fragments that are derived from protein-protein binding interfaces, with the inhibitory fragments predicted to bind in a native-like manner. We also leveraged our approach to predict novel binding modes of fragments from the intrinsically disordered C-terminal tail of FtsZ to several protein binding partners. Excitingly, these new structural models explain biological activity of the FtsZ C-terminus (and its peptide fragments) consistent with genetic and biochemical evidence, providing a demonstration of the

generalizability of our approach to protein interactions of unknown structure.

Overall, we demonstrate the power of AlphaFold to predict protein fragments that inhibit protein interactions in living cells and the complementarity of this computational approach with high-throughput experimental measurements. These results pave the way for proteome-wide discovery of inhibitory protein fragments and provide a promising path to uncover the underlying molecular features of potent peptide inhibitors.

Results

An approach for high-throughput AlphaFold prediction of peptide-protein interactions.

Recently developed methods enable massively parallel measurements of inhibition by protein fragments (17, 18, 20). However, we have lacked complementary, scalable computational approaches to predict which protein fragments are inhibitory and to generate models for their binding modes to aid in the interpretation of the high-throughput data. Noting the reported effectiveness of AlphaFold for predicting protein-protein interactions (28), including peptide-protein binding (24–26), we tackled this problem using the ColabFold implementation of AlphaFold2 (21, 29). In all cases, we used the AlphaFold2 monomer model weights, which were trained exclusively on single polypeptide chain data (21, 24) (in contrast with AlphaFold-Multimer weights that were determined using protein complex structures (30)), to ensure that any memorized native binding interactions would not be present in the training data.

To run AlphaFold2 (hereafter, AlphaFold) on many protein fragments tiling across proteins of interest, we modified the slow multiple-sequence alignment (MSA) generation step. Taking advantage of the fact that the fragments are segments of full-length proteins, we pre-generated full-protein MSAs and then pruned them to generate each fragment MSA, before running AlphaFold. We fed the precalculated fragment MSA and target protein MSA into a fragment + target protein AlphaFold prediction (**Fig. 1**, *Materials and Methods*). The target protein was in some cases the parent protein of the fragment (for homomeric interactions) and in other cases one of its interaction partners (for heteromeric interactions). In contrast to previous approaches that evaluated binding predictions primarily based on AlphaFold confidence metrics (24, 25, 31), we focused on properties of the structural models AlphaFold produced, particularly the number of binding contacts between the protein fragment and protein target (N_{contacts}); we further weighted this number by the interface-predicted template modeling (ipTM) score, a confidence metric for multichain structural models (30). We used weighted N_{contacts} to generate plots of predicted binding as a function of tiling fragment sequence position (**Fig. 1**). We termed this approach FragFold.

FragFold allowed us to rapidly predict binding between all possible protein fragments of a given length and any putative full-length protein binding partner, and score the resulting binding modes. We compared predictions to matched experimental results on the same set of protein fragments from a massively-parallel growth selection (**Fig. 1**, ref. (20)).

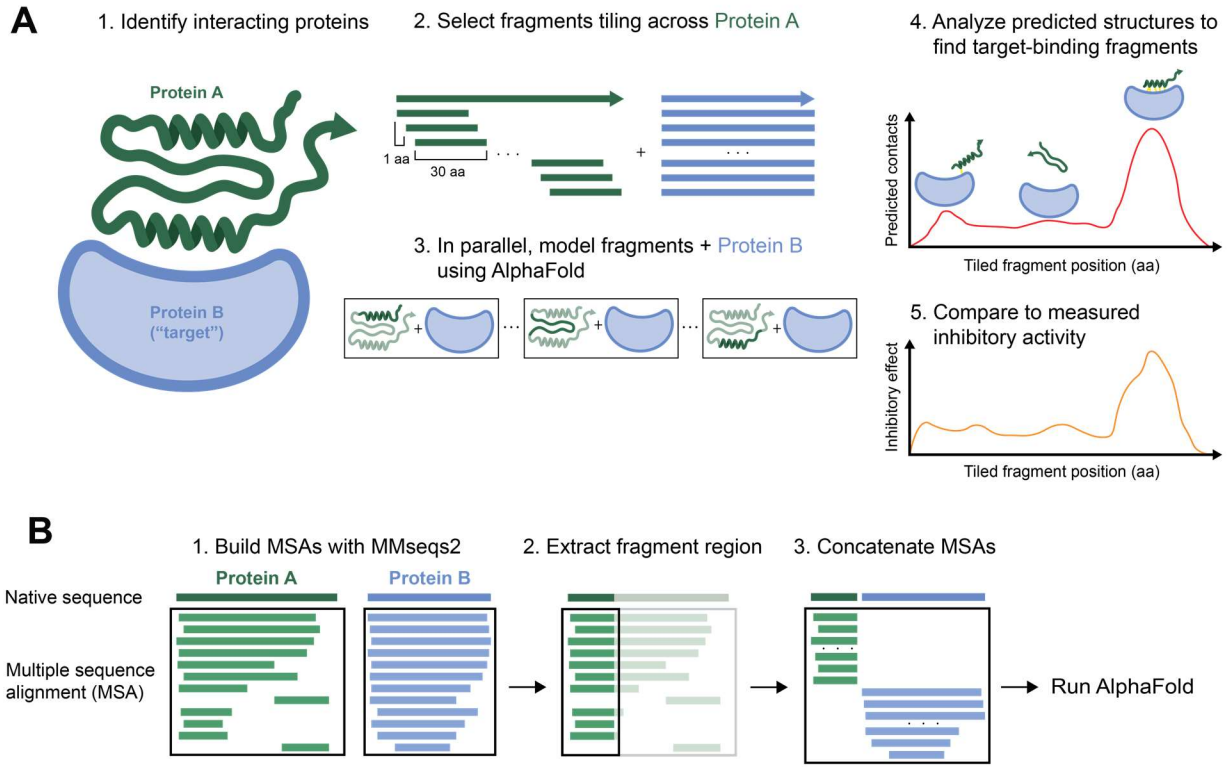

Figure 1. FragFold enables high-throughput prediction of protein fragment binding to protein targets. (**A**) Schematic of the FragFold pipeline. Protein pairs of interest known or suspected to interact are selected and fragmented *in silico*. Each tiling fragment is modeled together with the target interaction partner using AlphaFold2 (see *Materials and Methods*). Predicted structures are analyzed based on a combination of binding contacts and AlphaFold confidence metrics. The results can be compared to experimental measurements of protein fragment inhibitory activity in cells. (**B**) Schematic of the MSA-building scheme employed in FragFold to overcome the computational bottleneck of MSA generation for each fragment + protein pair.

Computational prediction of peptide-protein binding recapitulates inhibitory fragment peaks across the FtsZ protein sequence.

We first applied FragFold to predict protein fragment binding in the context of experimentally mapped inhibitory fragments that originate from structurally resolved protein-protein interaction interfaces. Such complexes provide an ideal test case where the relationship between binding and inhibition is most straightforward. We initially focused on the cell division protein FtsZ because fragments from across the filament polymerization interface act as inhibitors *in vivo* (20). Savinov *et al.* previously found that tiling inhibitory protein fragments form visually evident peaks around maxima in the inhibitory activity; in the case of FtsZ these peaks identified 4 distinct regions of the filament interface (**1**, **1′**, **2**, and **2′**; ref (20)). We first applied FragFold to perform a computational scan for fragments of FtsZ binding to full-length FtsZ, examining all possible 30-amino acid (aa) fragments with a 1-aa offset, as in the experiments (20). For each peak of inhibitory activity across the polymerization interface, there was a corresponding observed peak (local maximum) in weighted N_{contacts} calculated for tiling protein fragments (**Fig. 2*A***). Unweighted N_{contacts} predicted the same inhibitory peaks, whereas the AlphaFold confidence metrics ipTM (30)

and pLDDT (28) alone did not (**Fig. S1**).

Having determined the concordance of predicted binding and measured inhibitory activity for these interface regions, we next analyzed the predicted binding modes of fragments from each

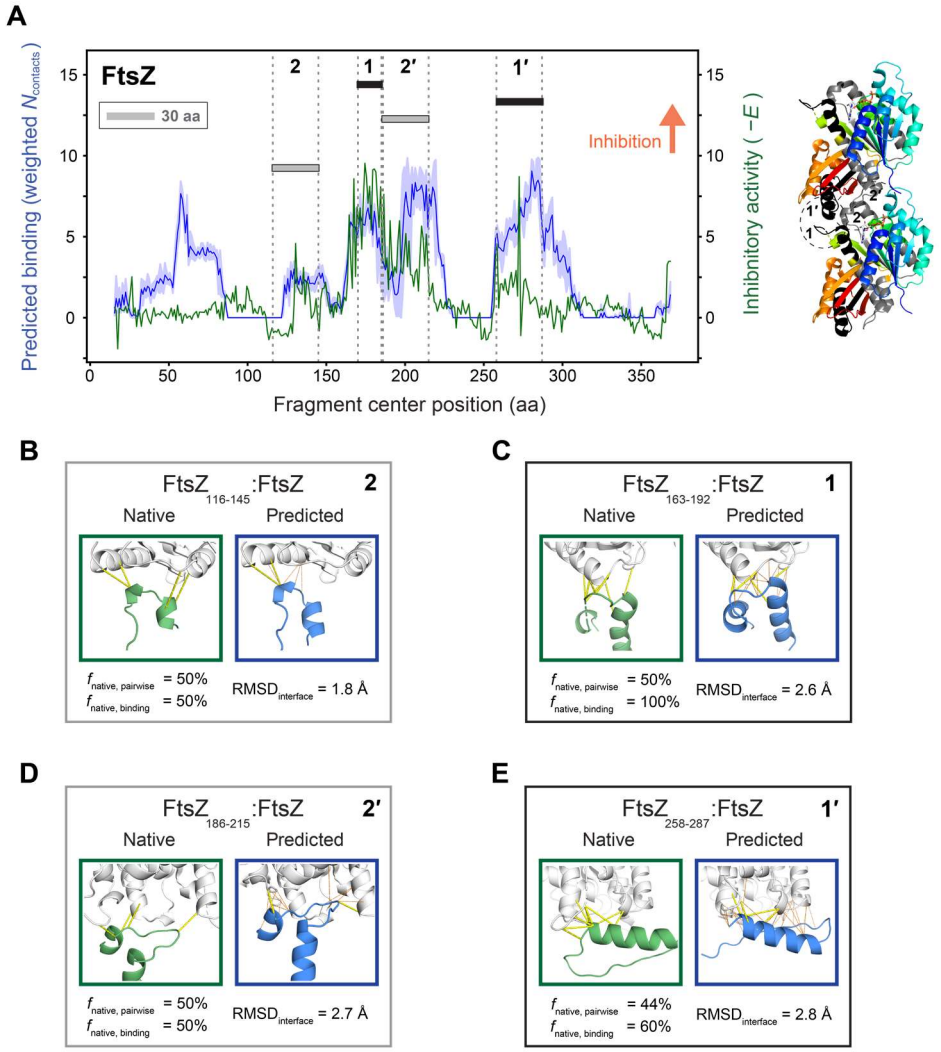

Figure 2. FragFold predicts all measured inhibitory peaks across the FtsZ filament interface, with native-like binding modes. (**A**) *Left*: Predicted target protein binding from FragFold (blue curves) vs. experimentally measured *in vivo* inhibitory activity (green curves, from ref. 20) for 30-aa fragment scans across *E. coli* FtsZ. Darker line and lighter outline for predicted binding data indicate mean and 95% C.I., respectively, across the 5 top-ranked structural models generated by AlphaFold. Filament interface regions **2**, **1**, **1′**, and **2′** are indicated. *Right*: crystallographic structure of two adjacent monomers of the FtsZ filament (PDB ID 6unx, ref. 32). Protein is colored from blue to red from the N- to the C-terminus. Regions corresponding to interface regions are annotated and highlighted in the same black/grey color scheme as in the fragment scan plot (*Left*). aa, amino acids. (**B**) – (**E**) Native, experimentally determined structures from the full FtsZ-FtsZ complex (green outlines; PDB ID 6unx) vs. FragFold-predicted structures of FtsZ fragment + full-length FtsZ (blue outlines) for protein fragments corresponding to each of the inhibitory peaks **2**, **1**, **1′**, and **2′**. Metrics for the extent of native-like binding are shown below. For each inhibitory peak (**B** – **E**), the representative measured inhibitory fragment is indicated, and the corresponding predicted binding fragment from the same structural region is as listed in **Table 1**. Pairwise native residue contacts are shown as dashed yellow lines; additional predicted native binding site contacts are shown as thinner dashed orange lines.

inhibitory peak. Strikingly, the FragFold predictions for sites **1**, **1′**, **2**, and **2′** placed the peptide in a geometry similar to the native binding mode seen in a crystallographic structure of the filament (32) (**Fig. 2*B–E***). To quantify this similarity, we calculated the fraction of native pairwise contacts between fragment and target residues, $f_{\text{native, pairwise}}$; the fraction of target protein residues from the native binding site bound by the fragment, $f_{\text{native, binding}}$; and the root mean-square deviation (RMSD) of the fragment-target interface compared to the native structure, $\text{RMSD}_{\text{interface}}$ (**Fig. 2*B–E***; *Materials and Methods*). All four major peaks yielded predictions with $\text{RMSD}_{\text{interface}} < 3$ Å, $f_{\text{native, binding}} \geq 50\%$, and $f_{\text{native, pairwise}} > 40\%$.

FragFold predicts protein-protein interaction inhibitory fragments across diverse proteins.

We next sought to determine how generally FragFold can predict inhibitory peaks mapping to protein interaction interfaces, and their binding modes, across structurally and functionally diverse proteins. We therefore extended our computational approach to fragment scans of several additional *E. coli* proteins: 50S ribosomal subunit protein L7/L12 (RplL), DNA gyrase subunit A (GyrA), single-stranded DNA binding protein (Ssb), lipopolysaccharide transport protein subunit G (LptG), and the GroEL/ES chaperone subunit GroEL. We previously obtained experimental measurements of *in vivo* inhibitory activity for fragments spanning each of these targets (20), revealing 36 inhibitory fragment peaks. Each of these proteins forms structurally characterized protein-protein interactions (PPIs), which allowed us to compare predicted binding modes to native structural data for complexes.

Across these diverse proteins, we evaluated predicted binding for all inhibitory peaks that we could assign to known PPIs based on experimental structures (14 inhibitory peaks corresponding to 15 PPIs; ref. (20), **Fig. 3**, **Table 1**; green highlighted regions in **Fig. S2**). Overall, we found that ~87% (13/15) of PPI-inhibitory fragment-protein interactions were predicted to bind (based on weighted $N_{\text{contacts}} > 2$) with a native-like binding mode ($f_{\text{native, binding}} \geq 50\%$) (**Fig. 3**, **Table 1**).

For instance, for RplL, we saw that the single inhibitory peak corresponding to the RplL dimerization domain was predicted to bind to RplL in a native-like manner, with $\text{RMSD}_{\text{interface}} = 1.2$ Å and $f_{\text{native, binding}} = 93\%$ (**Fig. 3*A***). The binding of 50S ribosomal subunit L10 (RplJ) to this same fragment of RplL was predicted as well, though with orientational deviation from the native mode ($\text{RMSD}_{\text{interface}} = 4.6$ Å and $f_{\text{native, binding}} = 67\%$; **Table 1**). For GyrA, AlphaFold predicted native-like binding to GyrB for the inhibitory fragment peak at residues 12-41 ($f_{\text{native, binding}} = 89\%$) and native-like binding to GyrA at the C-gate dimerization site (experimental inhibitory peak at residues 371-400; $f_{\text{native, binding}} = 100\%$) (**Fig. 3*B,C***). For LptG, our computational predictions successfully yielded the single inhibitory peak corresponding to a protein-protein interaction with LptF (LptG residues 309-338; $f_{\text{native, binding}} = 100\%$) (**Fig. 3*D***). For Ssb, the lone inhibitory peak was likewise reproduced computationally (residues 52-81; $f_{\text{native, binding}} = 100\%$) (**Fig. 3*E***, **Table 1**). Finally, for GroEL, fragments corresponding to inhibitory peaks at residues 3-32 and 486-515 mapping to GroEL-GroEL polymerization interactions were predicted to bind in native-like modes (each with $f_{\text{native, binding}} = 75\%$; **Fig. 3*F,G***) but $\text{GroEL}_{87\text{-}116}$: GroEL was not ($f_{\text{native, binding}} = 0\%$) (**Table 1**). The GroEL-GroES interface inhibitory peak fragments (GroEL residues 229-258, 246-275) were predicted also, with $f_{\text{native, binding}} = 100\%$ for fragment 246-275, but $f_{\text{native, binding}}$ of only 17% for fragment 229-258. Despite this, the interface RMSD was quite reasonable (3.2 Å) for the 229-258 prediction (**Fig. 3*H,I***). Interface RMSD was also low for the $\text{GroEL}_{87\text{-}116}$: GroEL prediction

(0.7 Å), indicating that in both of these cases where the prediction failed using the weighted $N_{contacts}$ and $f_{native,\ binding}$ criteria, AlphaFold nonetheless came close to predicting a native-like binding mode.

We tested the specificity of FragFold predictions by performing fragment scans against noncognate protein partners (**Fig. S3**). FtsZ fragments tiling the full protein sequence were not predicted to interact with enhanced GFP (eGFP); likewise, fragments of eGFP were not predicted to bind to FtsZ (**Fig. S3*A*–*B***). Repeating this test for GyrA, no GyrA fragments were predicted to bind to FtsZ (**Fig. S3*C***). When scanning FtsZ fragments for binding to GyrA, some binding peaks were predicted, but these were weaker than the specific peaks and not aligned in sequence

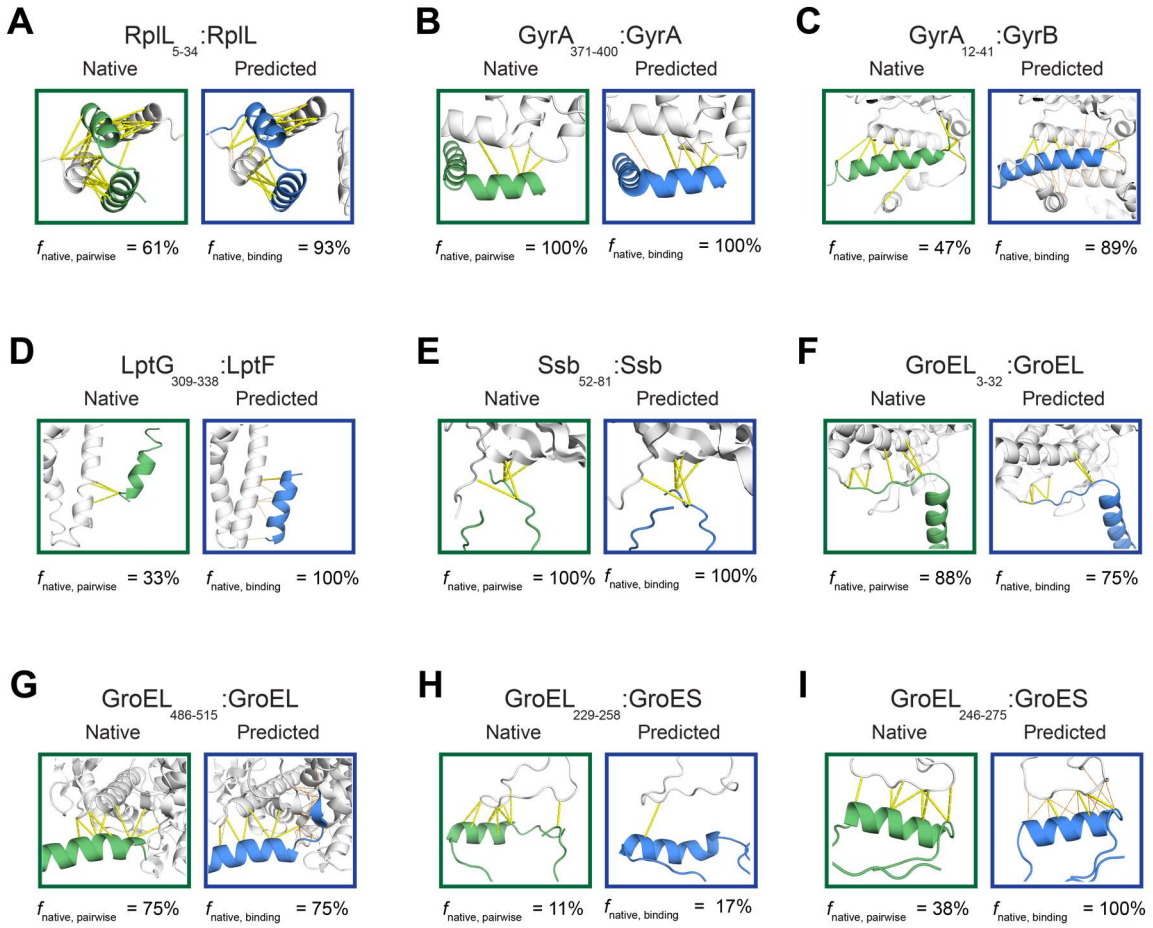

Figure 3. FragFold discovers native-like protein interaction inhibitors across structurally and functionally diverse proteins. (**A**) – (**I**) Native, experimentally determined structures from full-length protein complexes (green outlines) vs. FragFold-predicted structures of protein fragments + full-length target protein (blue outlines) for protein fragments corresponding to each of the indicated inhibitory peaks identified in *in vivo* experiments (ref. 20). Metrics for extent of native-like binding are shown below each structural comparison. For each inhibitory peak (**A** – **I**), the representative measured inhibitory fragment is indicated, and the corresponding predicted binding fragment from the same structural region is as listed in **Table 1**. Pairwise native residue contacts are shown as dashed yellow lines; native binding site contacts are shown as thinner dashed orange lines.

position with the experimentally measured inhibitory peaks (**Fig. S3*D***). We note as well that these noncognate predicted interaction peaks were uniformly narrow compared to typical native interaction-mapped peaks. This suggests that a greater number of overlapping fragments predicted to bind a target is indicative of a *bona fide*, or at any rate stronger, protein-peptide interaction. The structural features of these predicted nonspecific fragment binding modes include parallel alpha-helix and beta-sheet interactions (**Fig. S3*E–H***), suggesting that they are plausible modes of weak nonspecific association. The predicted binding peaks of eGFP fragments to eGFP (**Fig. S3*B***, grey) likely reflect the known propensity of GFP to dimerize, with the peaks centered around residues 157 and 191 (**Fig. S3*B*,*I–J***) including key residues involved in dimerization (33, 34).

Table 1. AlphaFold predicts native-like binding for the majority of inhibitory fragment peaks that correspond to known protein-protein interaction interfaces.

Fragmented protein	Binding partner	Measured inhibitory fragment (aa)	Predicted fragment (aa)	$f_{native, pairwise}$	$f_{native, binding}$	$RMSD_{interface}$ (Å)	$N_{contacts}$	Weighted $N_{contacts}$
FtsZ	FtsZ	116 - 145	118 - 147	50%	**50%**	1.7	6	**2.2**
FtsZ	FtsZ	163 - 192	162 - 191	50%	**100%**	2.6	18	**8.4**
FtsZ	FtsZ	186 - 215	186 - 215	50%	**50%**	2.7	16	**7.0**
FtsZ	FtsZ	258 - 287	271 - 300	44%	**60%**	2.8	27	**10.2**
RplL	RplL	5 - 34	1 - 30	61%	**93%**	1.2	24	**5.8**
RplL	RplJ	5 - 34	5 - 34	0%	**67%**	4.6	14	**2.1**
GyrA	GyrB	12 - 41	12 - 41	47%	**89%**	3.0	39	**16.1**
GyrA	GyrA	371 - 400	371 - 400	100%	**100%**	0.8	7	**2.6**
LptG	LptF	309 - 338	319 - 332	33%	**100%**	— *	7	**2.5**
Ssb	Ssb	52 - 81	53 - 82	100%	**100%**	2.3	5	**3.1**
GroEL	GroEL	3 - 32	3 - 32	88%	**75%**	0.4	7	**3.4**
GroEL	GroEL	486 - 515	493 - 522	75%	**75%**	0.5	24	**12.7**
GroEL	GroEL	87 - 116	87 - 116	0%	0%	0.7	3	1.7
GroEL	GroES	229 - 258	226 - 255	11%	17%	3.2	1	0
GroEL	GroES	246 - 275	246 - 275	38%	**100%**	3.2	16	**5.1**

For each experimental inhibitory peak identified as mapping to a PPI, the best matching fragment from FragFold and its native-like binding characteristics (as well as predicted $N_{contacts}$ and weighted $N_{contacts}$) are listed. Bolded values indicate predictions that meet the criteria for binding (weighted $N_{contacts} > 2$) and native-like interaction mode ($f_{native, binding} \geq 50\%$).

*: $RMSD_{interface}$ is undefined for $LptG_{319-332}$:LptF because only 1 native pairwise residue contact was predicted (though 100% of native binding-site residues were bound by the fragment).

Automated prediction of inhibitory fragments across diverse proteins suggests generalizability across proteomes.

The results from the predictions of inhibitory peaks mapping to known protein-protein interaction interfaces showed that FragFold can discern these types of inhibitory peaks with a high success

rate (**Table 1**). However, not all inhibitory peaks in the experimental data had obvious structural interpretations based on the protein complex structures examined, and more generally the set of known protein interaction structures is incomplete. Furthermore, the relationship between (predicted) protein binding and inhibitory function is not necessarily straightforward, with binding at known interaction interfaces representing the simplest case. We therefore sought to develop an automated approach that could be applied to discover inhibitory protein fragments from computational fragment scans of arbitrary proteins. To do so, we began from our computational results alone and automatically determined predicted binding peaks as clusters of overlapping protein fragments predicted to bind the target protein with a sufficient number of contacts at a common binding site (at least 6 fragments per cluster, $N_{\text{contacts}} \geq 3$, and weighted $N_{\text{contacts}} \geq 3$; *Materials and Methods*). Computationally predicted target-binding fragment clusters were then compared to experimentally determined inhibitory fragment peaks. To make this process as unsupervised as possible, we developed an algorithm to automatedly determine experimental inhibitory peaks from massively parallel fragment scans, which was able to reproduce all previously manually assigned inhibitory peaks (*Materials and Methods*; **Table 1** and ref. (20)). Comparison of the predicted binding clusters with the experimental inhibitory peaks revealed that all 14 previously identified PPI-inhibitory peaks (**Table 1**) were predicted using this automated pipeline (**Fig. S2**, green highlighted regions). In addition to these 14 peaks, 22 additional visually evident inhibitory peaks not mapping to known PPIs were identified (**Fig. S2**, green lines).

We sought to estimate how often a FragFold-predicted protein-binding fragment is inhibitory using an automated approach. We based this assessment purely on the overlap of predicted binding clusters and ground-truth experimental inhibitory peaks across all investigated protein sequences. We optimized the tradeoff between prediction sensitivity (fraction of all inhibitory fragment peaks mapping to a corresponding predicted binding cluster, $f_{\text{inhibitory peaks predicted}}$) and specificity (fraction of all predicted binding clusters (peaks) that map to inhibitory peaks, $f_{\text{predicted binding clusters inhibitory}}$) by adjusting the peak calling criteria (**Fig. S5**; *Materials and Methods*). Using our chosen parameters, we found that $f_{\text{inhibitory peaks predicted}}$ was ~64% (23/36) (**Table 2**, **Fig. S2**). This included all protein interaction-inhibitory peaks in **Table 1**. At this level of sensitivity, predictions also showed good specificity: we found that $f_{\text{predicted binding clusters inhibitory}}$ was ~68% (23/34) (**Table 2**, **Fig. S2**). As a further measure of specificity, we found that $f_{\text{predicted binding clusters inhibitory}}$ (68% over all proteins) was higher than expected from random placement of peaks across all proteins investigated (51%) (**Table S1**). Evaluating the ability of FragFold to predict 36 peaks across all 6 proteins, this difference from the null model was statistically significant (p = 0.048; *Materials and Methods*). Similarly, across all proteins, $f_{\text{inhibitory peaks predicted}}$ (64%) was also significantly larger than expected from random placement (49%; p = 0.035). Together, these results suggest that AlphaFold can successfully predict many protein fragment inhibitory peaks across diverse proteins.

Table 2. **FragFold systematically predicts inhibitory protein fragment peaks.**

Fragmented protein	$f_{\text{inhibitory peaks predicted}}$	$f_{\text{predicted binding clusters inhibitory}}$
All proteins	**64%**	**68%**
FtsZ	86%	71%
GroEL	44%	67%
GyrA	54%	70%
LptG	75%	57%
RplL	100%	100%
Ssb	100%	67%

The fraction of measured inhibitory peaks that matched corresponding predicted binding clusters ($f_{\text{inhibitory peaks predicted}}$) as well as the fraction of predicted binding clusters (peaks) that matched corresponding experimentally determined inhibitory peaks ($f_{\text{predicted binding clusters inhibitory}}$) are listed. These results are presented for all fragmented proteins ("All proteins"), as well as for each protein individually.

AlphaFold predicts previously unknown binding modes of the intrinsically disordered C-terminal tail of FtsZ in agreement with orthogonal functional measurements.

The success of FragFold in predicting inhibitory fragment peaks and native-like binding modes across diverse proteins – despite training on only folding of single-chain polypeptides (21, 24) – prompted us to wonder whether AlphaFold can successfully predict previously unknown fragment binding modes. Success in this task would further support generalizability of FragFold for predicting likely inhibitory protein fragment binders of any protein of interest. To investigate this question, we focused on the intrinsically disordered C-terminal tail of FtsZ, comprised of residues 317-383 (35, 36). This region of the protein is known to be important for regulatory interactions with other divisome proteins including FtsA, MinC, and ZipA in *E. coli* (36), but the intrinsically disordered tail is completely unresolved in FtsZ crystal structures (32). Previously, Savinov *et al.* identified a peak of *in vivo* inhibitory activity corresponding to fragments from the extreme C-terminus of FtsZ (residues 354-383; **Fig. 2**, ref. (20)). Investigating the predicted binding of tiling FtsZ peptide fragments to FtsZ itself, FtsA, MinC, and ZipA, we found that there were clear binding peaks to FtsZ and FtsA mapping to this C-terminal inhibitory peak; there was also a weak signature for predicted binding to ZipA in the same region (**Fig. 4*A***).

We analyzed the predicted structural features of C-terminal FtsZ fragment interactions with divisome proteins. Considering first the predicted binding mode to full-length FtsZ, we found that C-terminal fragments are predicted to bind to the GTPase active site at interaction site **2-2′**, with fragments predicted to both occlude the GTPase binding site and prevent binding to the T7 loop of the adjacent FtsZ monomer in the filament (**Fig. 4*B***). Both features of this binding mode are expected to prevent filament polymerization and hence are consistent with the observed inhibitory activity. Excitingly, this novel structural model agrees with biochemical measurements indicating

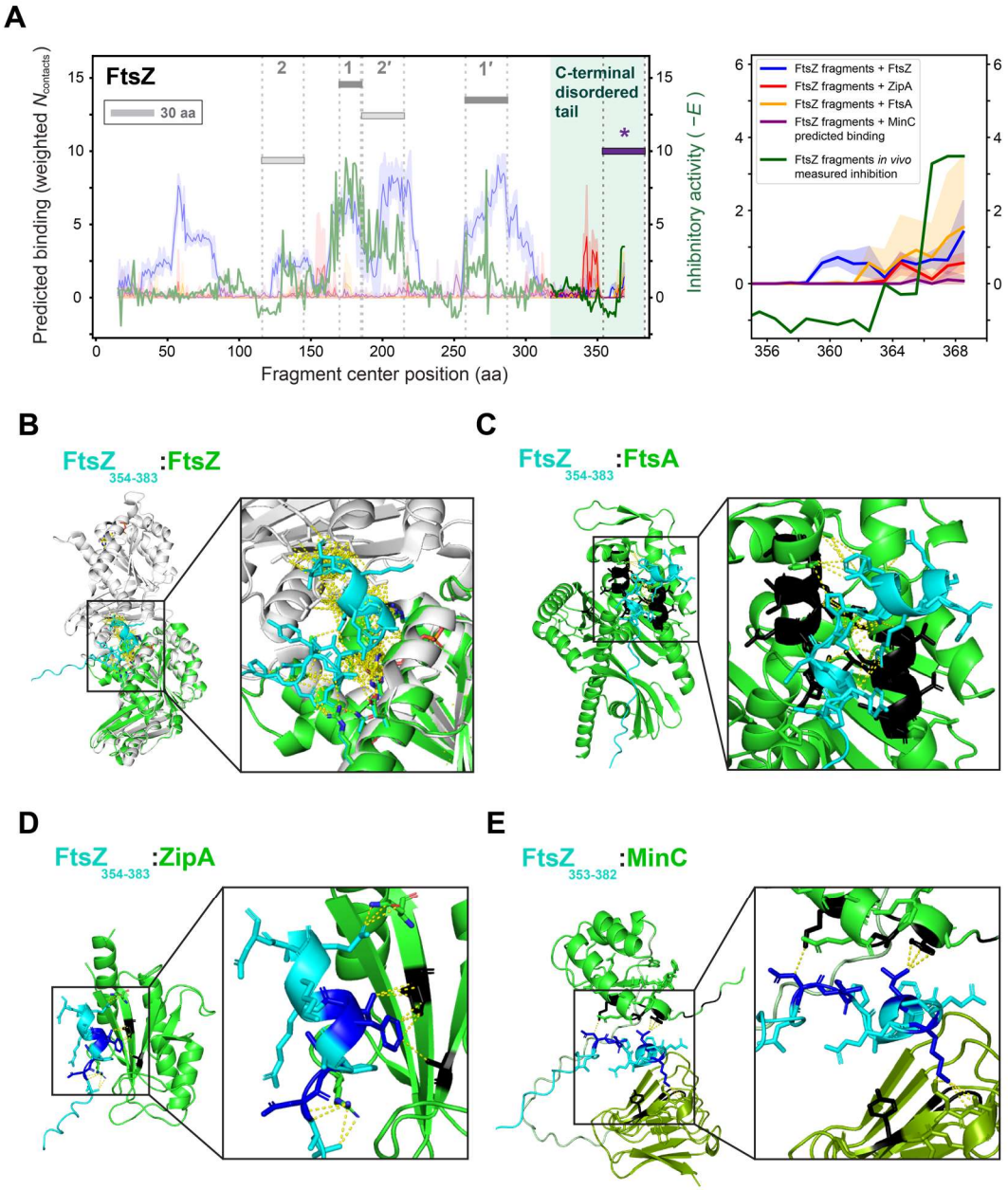

Figure 4. FragFold predicts novel binding modes for the intrinsically disordered FtsZ tail in agreement with genetic and biochemical evidence. (**A**) *Left*: Predicted target protein binding from FragFold (blue, red, tan, and purple curves) vs. experimentally measured *in vivo* inhibitory activity (green curves, from ref. 20) for 30 aa fragment scans across *E. coli* FtsZ with partner proteins FtsZ, ZipA, FtsA, and MinC as indicated in the legend on the *Right*. Darker line and lighter outline for predicted binding data indicate mean and 95% C.I., respectively, across the 5 top-ranked structural models generated by AlphaFold. Filament formation interface regions **2**, **1**, **1′**, and **2′** are indicated. The highlighted region of the plot corresponds to the intrinsically disordered C-terminal tail. The indicated region ***** (purple) corresponds to the peak of fragment inhibitory activity in the C-terminal tail. *Right*: Inset from the plot on the *Left* focusing on the C-terminal inhibitory peak. aa, amino acids. (**B**) – (**E**) AlphaFold-predicted structures of indicated C-terminal FtsZ fragments (cyan) bound to various divisome proteins (green). Contacts between residue pairs are shown as dotted yellow lines. (**B**) Grey: FtsZ dimer structure from PDB ID 6unx (32), shown overlaid with AlphaFold model. (**C**) – (**E**) Black: regions of FtsA, ZipA, or MinC previously identified to play a role in FtsZ interactions (see main text); (**D**) – (**E**) Blue: regions of FtsZ previously shown to play a role in interactions with ZipA or MinC, respectively (main text). (**E**) $MinC^N$ domain shown in bright green, $MinC^C$ domain in olive green; interdomain linker in pale green.

that the C-terminal tail plays an autoregulatory role in FtsZ filament assembly and acts as an auto-

inhibitor of GTPase activity (37, 38). In addition, the binding mode predicted by AlphaFold is consistent with molecular dynamics simulations indicating that the C-terminal end of the C-terminal tail binds in the vicinity of the GTPase active site (38), with our results providing the first detailed molecular model for this interaction.

In the case of FtsA, AlphaFold predicts FtsZ C-terminal fragment binding to subdomain 2B in the vicinity of FtsA residues 234-245 and 294-304 (**Fig. 4*C***). This new molecular model for FtsZ-FtsA binding agrees strikingly well with the results of genetic screens for FtsA mutants disrupting the FtsA-FtsZ interaction (39), which outlined a binding pocket in the same region of the 2B subdomain (**Fig. 4*C***, black). Notably, the predicted binding mode we report here places the FtsZ C-terminal peptide near the ATP-binding site of FtsA, potentially inhibiting nucleotide binding. ATP binding by membrane-associated FtsA is important for FtsA membrane remodeling and Z-ring constriction (40), suggesting that this structural model for the FtsZ tail – FtsA binding mode reflects a key regulatory interaction between these two cell division proteins.

We also considered the predicted binding of C-terminal FtsZ fragments to ZipA (**Fig. 4*D***). In this case an experimental structural model is available (41). Although the predicted binding peak is weak, the predicted binding mode broadly agrees with the experimental structure, with the C-terminal tail peptide of FtsZ binding to the same beta-sheet face of ZipA, but in a different orientation. Furthermore, two of the three most important FtsZ residues for ZipA binding revealed in previous work (Phe377 and Leu378) are also key binding residues in our structure; as in the crystallographic structure, they make contacts with Ala246 and Phe269 of ZipA (**Fig. 4*D***, black and blue residues). Interestingly, Mosyak and colleagues previously found that ZipA binds C-terminal FtsZ fragments weakly (K_D ~20-35 μM, ref. (41)), and the crystallographic peptide-protein interface contains substantial water content, with an elevated B-factor for the FtsZ fragment. Together, these features suggest that various orientations of peptide binding may be plausible at this weak interface, and the structural model produced by AlphaFold represents a reasonable binding mode. We noted as well the predicted ZipA-binding peaks for FtsZ fragments centered around residues 343 and 347 (**Fig. 4*A***). Interestingly, these correspond to predicted binding to the same ZipA interface as fragment 354-383 (**Fig. S4**). This suggests ZipA may act as an anchor for two different regions of the FtsZ intrinsically disordered tail.

Finally, we investigated binding between the FtsZ C-terminus and MinC. Across the top 5 AlphaFold models, the average binding peak was very weak. But a local maximum in weighted $N_{contacts}$ (**Fig. 4*A***, *right*) prompted us to investigate further. The highest-ranked model corresponds to the C-terminal FtsZ fragment 352-383 making 3 residue-residue contacts with MinC (**Fig. 4*E***). The structural model is intriguing, with the FtsZ tail binding in between the two domains of MinC ($MinC^N$ and $MinC^C$). The predicted binding mode is in good agreement with genetic and biochemical analyses of FtsZ and MinC (42–44). All three key binding residues predicted for FtsZ (D373, V378, and M380; **Fig. 4*E*,** blue), as well as their partners on $MinC^N$ (K35 and F41; **Fig. 4*E***, black), were previously shown to be critical for this protein-protein interaction. Although the exact $MinC^C$ residues bound by FtsZ in our model (L203 and D205) were not previously shown to be involved in this interaction, a nearby residue on the same face of $MinC^C$ (Y201) was, suggesting only a small adjustment to this binding mode would be needed to explain the role of this residue as well. We note that the AlphaFold model employs some of the residues shown to have a role in FtsZ binding but not others, which is also consistent with prior measurements: in addition to binding the intrinsically disordered C-terminal tail, MinC also binds elsewhere on FtsZ

as part of its function in filament disassembly (43).

Our results for the example of the FtsZ C-terminal tail demonstrate that AlphaFold can predict novel peptide-protein interactions that inhibit key protein functions in cells. The agreement of these novel structural models with orthogonal experimental results suggests that they capture key features of the interactions in question.

Discussion

We demonstrated in this work the power of high-throughput AlphaFold predictions to discover protein interaction-inhibitory protein fragments and predict their binding modes. The results presented here are based on predictions and measurements for a panel of six diverse essential proteins for which binding partners are known. Yet the approach is expected to generalize to a much larger set of targets. Similar results are anticipated for any protein under conditions in which the relevant protein interactions become important for cell growth. In applying this approach to larger collections of proteins, knowledge of relevant interaction partners that could be the source of inhibitory fragments might come from genetic screens, pull-down experiments, colocalization measurements, or *in vitro* biochemistry, as tabulated in a number of databases (45–47).

An important parameter in evaluating our predicted binding peaks was the peak width – that is, the number of distinct overlapping protein fragments predicted to bind the target protein. In our binding prediction specificity tests, we observed narrower peaks for off-target compared to on-target interactions (**Fig. S3**). In general, binding peaks were more likely to match a measured inhibitory peak when the number of contributing protein fragments was sufficiently large, which we incorporated as a threshold in our automated pipeline for inhibitory fragment discovery (*Materials and Methods*). This filter based on predicted binding peak (cluster) width allowed us to remove false positive binding predictions, contributing to the ~68% success rate in predicting inhibitory fragments systematically. This finding was reminiscent of other recent work showing that the number of overlapping protein sequence regions yielding comparable predictions improved confidence in protein-protein binding predictions using AlphaFold-Multimer (48), and more broadly the benefits of modeling short segments of protein sequence to predict binding interactions (48, 49).

We also found that the number of fragment-target contacts from AlphaFold structural models (N_{contacts}) weighted by the predicted interface TM score (ipTM) (30, 50) was a strong predictor of inhibitory protein fragments. This approach, placing the structural model first and the prediction confidence of AlphaFold second, stands in contrast to other approaches that evaluated predictions purely based on ipTM or pLDDT (24, 48, 51). The demonstrated effectiveness of weighted N_{contacts} in the context of inhibitory protein fragment prediction suggests it may be a useful metric for other applications as well. Our finding of the predictive value of N_{contacts} is in agreement with the recent work of Bryant and colleagues. Similar to our weighting approach, they found that a combination of pLDDT and N_{contacts} was a potent discriminator (52).

The automated overlap analysis of predicted binding clusters (revealing predicted binding peaks) and measured inhibitory peaks had some limitations. First, the floor for successful prediction was high due to the biology of inhibitory protein fragments of this length: the $f_{\text{predicted binding clusters inhibitory}}$ expected from random placement of peaks was 34 – 66% (51% across all proteins) due to the

significant fraction of protein sequence corresponding to inhibitory peaks (20). Second, the maximum success rate is limited by several factors: not all predicted binding modes from AlphaFold represent true positives, not all binding modes are inhibitory, and not all possible cellular protein binding partners of each protein fragment are considered in the predictions. These challenges are exemplified by the predicted binding peak centered around FtsZ residue 60 (**Fig. 2*A***, *left*, blue), which has no corresponding inhibitory activity (**Fig. 2*A***, *left*, green). In this light, the obtained values of $f_{\text{predicted binding clusters inhibitory}}$ (68%) and $f_{\text{inhibitory peaks predicted}}$ (64%) represented successful, but imperfect, predictions. Another limitation is that although predicted binding clusters overlapped with measured inhibitory peaks, the predictions and measurements often exhibited an offset (**Fig. S2**). This indicates that several nearby tiling fragments should be tested for each FragFold-predicted binding peak to identify functional inhibitors. Reducing this offset between the predictions and the experiments represents an area for future improvements and may reflect limitations in the precision of binding predictions.

Our results suggest multiple applications for FragFold. First, the high success rate of our predictions (~68% of predicted binding clusters corresponding to experimentally measured inhibitory peaks) shows that using FragFold as a computational screen for fragment-based inhibition should be fruitful across diverse proteins of interest. Such predictions have direct applications and are highly synergistic with high-throughput experimental measurements of fragment-based inhibition (17–20). The revolution in DNA synthesis capabilities (53, 54), combined with the ongoing revolution in high-throughput sequencing, has enabled these high-throughput experimental approaches (19); nonetheless, limitations persist in library size and cost. In applications with the primary goal of discovering novel inhibitory peptides, e.g., as leads for novel therapeutics, or probes for systematically disrupting PPIs, pre-screening for likely inhibitory protein fragments with FragFold stands to substantially reduce the protein sequence space that needs to be scanned. This will allow for the coverage of significantly larger sets of fragments in a library of given size, greatly accelerating inhibitory fragment discovery and large-scale systematic characterization. FragFold should likewise enable the economical use of smaller libraries for inhibitor characterization. In future work, FragFold has the promise to further synergize with additional experimental modalities, including cell-display approaches and massively-parallel assays of folding and binding interactions (55–59).

Second, FragFold can complement high-throughput measurements by supporting the interpretation of peptide inhibitory mechanisms at scale. We demonstrate this complementarity in this work by combining FragFold structural modeling with high-throughput measurements of inhibitory function *in vivo*, both providing further support for the interpretation of the measured inhibitory effects as driven by dominant-negative competitive inhibition of native interactions (ref. (20); **Table 1**), and predicting novel FtsZ IDR fragment binding modes in agreement with orthogonal functional data. Our findings suggest that AlphaFold has learned generalizable principles of protein fragment binding that can be applied to discover novel interactions. In this, our results add to a growing body of work showing successful AlphaFold prediction of previously unknown protein-protein interactions aligning with experimental measurements. Examples include a phage defense protein complex structure concordant with hydrogen-deuterium exchange mass spectrometry measurements (60), and a Kelch domain pocket – SLiM interaction in agreement with BRET measurements and mutagenesis results (48). Other recent work has similarly leveraged the synergy of AlphaFold predictions with high-throughput experiments,

including the combination of mass spectrometry-based protein interactomics with AlphaFold modeling (51).

In conclusion, we have demonstrated the power of AlphaFold to systematically predict protein fragments that inhibit native protein interactions in living cells. The approach we have developed to scale these predictions across tiling protein fragments, FragFold, promises to be broadly applicable to the discovery of novel inhibitory peptides across proteomes. This is supported by our ability to predict nativelike inhibitory fragments across diverse proteins, as well as previously unknown binding modes of an intrinsically disordered region. FragFold enables both screening for and systematic characterization of inhibitory fragments and provides a natural complement to high-throughput experimental methods for measuring peptide function.

Materials and Methods

Selection of proteins to computationally scan with fragments

Proteins computationally fragmented and investigated with FragFold were selected from among those experimentally measured with a high-throughput approach in previous work (20). We chose to investigate fragments of FtsZ, RplL, GyrA, Ssb, LptG, and GroEL because each forms structurally characterized protein-protein interactions that we might expect AlphaFold to be able to predict (as opposed to interactions of unknown structure, protein-nucleic acid interactions, or interactions with folding intermediates).

Predicting fragment-protein complex structures

We used ColabFold (version 1.5.2) to predict the structure of protein fragments interacting with full-length proteins. We used the AlphaFold2 monomer weights (`model_type = monomer_ptm`) with a residue-index gap between chains (residue-index mode, offset > 32) – which was previously found to allow modeling of multiple polypeptide chains in AlphaFold2 and is natively implemented in ColabFold (29) – to ensure that no information on protein-protein interactions was utilized. No templates were used. Other settings were set to ColabFold defaults. We modified the protocol for multiple-sequence alignment (MSA) generation to maximize the depth of the MSA for the fragments. Specifically, we used MMseqs2 (61) to construct MSAs individually for all full-length proteins from which fragments were derived. The residues in the MSA corresponding to the fragment were then extracted from the full-length MSA and used to construct a fragment MSA. The fragment MSA was concatenated with the full-length MSA of the interacting protein, without pairing. We iterated this process for all fragments of full-length proteins to generate custom input MSAs for ColabFold. The runtime was highly dependent on the total number of residues in the fragment and full-length protein chain. For example, using 4 A100 GPUs it took 16 hours to predict the structures of 354 FtsZ fragments (30 aa) with full-length FtsZ (306 residues) whereas it took 210 hours to predict the structures of 843 GyrA fragments (30 aa) with full-length GyrA (872 residues).

Native interface recovery

We compared experimentally determined structures of interacting full-length proteins (i.e., native) with ColabFold-predicted structures of fragments and full-length proteins (i.e., predicted). These

comparisons were performed for a set of known inhibitory protein fragment peaks from previous experimental measurements that mapped to structurally characterized protein-protein interaction interfaces; these peaks were identified previously (20) or in this work based on inspection of the inhibitory activity profile of high-throughput fragment scans and corresponding experimental structures of protein complexes. For each inhibitory peak, we selected a representative protein fragment possessing the maximal inhibitory activity of all fragments in the peak. The experimental structure for this representative inhibitory fragment – protein interaction was compared to a corresponding AlphaFold-predicted model for a fragment of similar or identical sequence position (offset of 0 – 13 residues, see **Table 1**) corresponding to a local maximum of weighted N_{contacts}. The top AlphaFold model ranked by weighted N_{contacts} was used. After defining all residues corresponding to the fragment of interest in the experimental structure, we removed all other residues from the chain to create a model of the fragment interacting with the full-length protein chain as it does in the context of the full protein-protein interaction. We compared the native and predicted structures using contact recovery and backbone RMSD metrics.

We defined interface residue-residue contacts between the fragment and full-length protein chain based on heavy-atom distances. Specifically, a contact was defined between two residues if they had at least one pair of heavy atoms within 4.0 Å. We defined interface contacts in the native structure and calculated the native contact recovery as the fraction of native contacts observed in the predicted structure.

Peptide root-mean-square deviation (RMSD) was calculated between backbone atoms (N, Ca, C, O) of the native and predicted structures, after optimal superposition of the full-length protein chain using the Bio.PDB Python package. In some cases, residues were missing from the native structure and the number of residues used in the calculation was less than the number of residues in the fragments.

Interface RMSD was calculated between all interface residues in the native structure and corresponding residues in the predicted structure. We designated interface residues using the earlier defined contact definition, with a larger distance cutoff of 8.0 Å. The optimal superposition between the native interface residues and corresponding residues in the predicted structure was applied and the RMSD was calculated. If there were fewer than 5 corresponding residues in the fragment or full-length protein chain, we did not compute the RMSD.

Automated prediction of inhibitory peaks

We filtered and clustered all of the predicted complex structures to predict potential inhibitory peaks. For each set of predicted structures, i.e., ColabFold predictions of fragments of one protein interacting with a full-length protein, we filtered predictions by multiple criteria. Predicted structures were only considered if they had at least 3 interface contacts, 3 ipTM-weighted interface contacts, and an ipTM value of 0.3 or greater. The predicted structures (up to 5 per fragment) were then clustered using Tanimoto distance (62, 63) over interface contacts. Specifically, we performed hierarchical complete-linkage clustering with a maximum Tanimoto distance of 0.4. We then filtered the clusters based on the number of unique fragments in the cluster, requiring at least 6 fragments per cluster, and merged heavily overlapping clusters to obtain the final set of predicted binding clusters. We used the set of all residues in a cluster (from any fragment) to constitute a predicted inhibitory peak. We obtained the final set of parameters

by scanning over potential values for minimum $N_{contacts}$, weighted $N_{contacts}$, ipTM, and number of overlapping fragments per cluster, as well as maximum Tanimoto distance. We manually inspected the predicted clusters and optimized for the resulting $f_{\text{inhibitory peaks predicted}}$, $f_{\text{predicted binding clusters inhibitory}}$, and a number of predicted peaks called similar to the number of experimental peaks measured. The minimum cluster size of 6 was selected based on this optimization of $f_{\text{inhibitory peaks predicted}}$ and $f_{\text{predicted binding clusters inhibitory}}$ (**Fig. S5**). Matching of predicted binding peaks to inhibitory activity peaks (called from the experimental data as described below) was assessed based on peak overlap, with an overlap of at least 2/3 of the fragment length (i.e., 20 aa for 30 aa fragments) required to consider two peaks (widths ranging from 30 − 67 aa for 30 aa fragments) to match.

Calling inhibitory peaks from experimental data

We automatically annotated peaks in the experimental inhibitory data using a simple protocol. Measured *in vivo* inhibitory effects were expressed as the enrichment (E) of protein fragments in a high-throughput massively parallel growth selection experiment, such that negative E values corresponded to inhibitory activity (20). We filtered out fragments that had inhibitory effects less than a gene-specific inhibitory effect cutoff corresponding to an enrichment Z-score of −2.5. The only exception to this universal Z-score cutoff was made for the protein GroEL, in which case we used a cutoff corresponding to a Z-score of −2 due to the apparent lower signal:noise ratio of the inhibitory peaks previously identified (20). After filtering by this Z-score cutoff, only inhibitory fragments remained. To group these fragments into peaks, we considered each fragment in turn, proceeding from the N- to the C-terminus. The N-terminal most inhibitory fragment was considered the starting point of the first peak; for each subsequent inhibitory fragment, the fragment was added to an existing peak if the N-terminal residue of the new fragment was less than 5 aa from the N-terminal residue of the nearest inhibitory fragment in the peak and no more than 25 residues away from the N-terminal residue of the first fragment in the peak. Otherwise, a fragment was considered the starting point for a new peak. This simple algorithm successfully reproduced all inhibitory peaks called by hand here and in previous work (**Table 1** and ref. (20)); comparing manual and automatically-called peaks, we note that in a few cases two nearby peaks were merged, and in other cases peaks were split into nearby sub-peaks.

Peak prediction null model

We developed a null model for simulating peak prediction with randomized peak positions to obtain null expectation values and p-values for the sensitivity ($f_{\text{inhibitory peaks predicted}}$) and specificity ($f_{\text{predicted binding clusters inhibitory}}$) statistics that we measured for our FragFold results. For each protein, we assumed the same number of predicted peaks as calculated by FragFold, and assigned peak positions at random, with starting residues sampled uniformly from the range $[1, L - k)$ (where L and k are the amino acid lengths of the protein and fragments, respectively). Randomly placed peaks that overlapped another randomly placed peak with ≥70% sequence overlap were rejected and resampled. We set the width of each null model peak to the mean peak width from FragFold across all predictions over all proteins. $f_{\text{null, inhibitory peaks predicted}}$ and $f_{\text{null, predicted binding clusters inhibitory}}$ were then calculated for each protein and across all proteins based on overlap with the fixed inhibitory peaks from experimental data (20) just as for the FragFold results. We repeated this 5,000 times for each protein to obtain distributions that we used to calculate null expectation values (means of $f_{\text{null, inhibitory peaks predicted}}$ and $f_{\text{null, predicted binding clusters inhibitory}}$ across all simulation runs), and p-values

across all proteins for the FragFold results compared to this null model.

Acknowledgements

We thank S. Fields, C. Sanchez, R. M. Garner, P. L. Freddolino, H. K. Wayment-Steele, S. Ovchinnikov, and members of the Li lab for helpful discussions. A.S. was supported by NRSA fellowship F32 GM134557 from the National Institutes of Health. This work was supported by awards K99 GM148718 and F32 GM134557 (to A.S.), R35 GM124732 (to. G.W.L.) and R35 GM149227 (to A.K.) from the National Institutes of Health. We acknowledge the MIT SuperCloud for providing computational resources (64).

Author contributions

Research conceptualization and design were performed by A.S. and S.S. Design and implementation of the FragFold computational method was performed by S.S. with feedback from A.S. Data analysis was performed by A.S. and S.S., with suggestions from A.K. and G.W.L. The manuscript was written by A.S. and S.S. with feedback and revisions from A.K. and G.W.L.

Data and code availability

The code used in this paper – and to run FragFold on any protein of interest – is available on GitHub (https://github.com/swanss/FragFold). Data are provided in the main text, the Supplementary Information, and a Source Data file available via figshare (doi: 10.6084/m9.figshare.24841269; URL: https://figshare.com/articles/dataset/Source_Data_for_Savinov_and_Swanson_et_al_2023/24841269).

References

1. G. Storz, Y. I. Wolf, K. S. Ramamurthi, Small Proteins Can No Longer Be Ignored. *Annual Review of Biochemistry* **83**, 753–777 (2014).

2. M. R. Hemm, J. Weaver, G. Storz, Escherichia coli Small Proteome. *EcoSal Plus* **9**, 10.1128/ecosalplus.ESP-0031–2019 (2020).

3. J. Chen, *et al.*, Pervasive functional translation of noncanonical human open reading frames. *Science* **367**, 1140–1146 (2020).

4. Z. Posner, I. Yannuzzi, J. R. Prensner, Shining a light on the dark proteome: Non-canonical open reading frames and their encoded miniproteins as a new frontier in cancer biology. *Protein Science* **32**, e4708 (2023).

5. C.-S. Huang, B. P. Pedersen, D. L. Stokes, Crystal structure of the potassium-importing KdpFABC membrane complex. *Nature* **546**, 681–685 (2017).

6. S. Safarian, *et al.*, Active site rearrangement and structural divergence in prokaryotic respiratory oxidases. *Science* **366**, 100–104 (2019).

7. D. Du, *et al.*, Structure of the AcrAB–TolC multidrug efflux pump. *Nature* **509**, 512–515 (2014).

8. E. C. Hobbs, X. Yin, B. J. Paul, J. L. Astarita, G. Storz, Conserved small protein associates with the multidrug efflux pump AcrB and differentially affects antibiotic resistance. *Proceedings of the National Academy of Sciences* **109**, 16696–16701 (2012).

9. J. E. Martin, L. S. Waters, G. Storz, J. A. Imlay, The Escherichia coli Small Protein MntS and Exporter MntP Optimize the Intracellular Concentration of Manganese. *PLOS Genetics* **11**, e1004977 (2015).

10. A. Zhong, *et al.*, Toxic antiphage defense proteins inhibited by intragenic antitoxin proteins. *Proceedings of the National Academy of Sciences* **120**, e2307382120 (2023).

11. M. Knopp, *et al.*, De Novo Emergence of Peptides That Confer Antibiotic Resistance. *mBio* **10**, 10.1128/mbio.00837-19 (2019).

12. A. M. Babina, *et al.*, Rescue of Escherichia coli auxotrophy by de novo small proteins. *eLife* **12**, e78299 (2023).

13. I. Frumkin, M. T. Laub, Selection of a de novo gene that can promote survival of Escherichia coli by modulating protein homeostasis pathways. *Nat Ecol Evol*, 1–13 (2023).

14. D.-A. Silva, *et al.*, De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 (2019).

15. M. D. T. Torres, S. Sothiselvam, T. K. Lu, C. de la Fuente-Nunez, Peptide Design Principles for Antimicrobial Applications. *Journal of Molecular Biology* **431**, 3547–3567 (2019).

16. A. Chevalier, *et al.*, Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74–79 (2017).

17. M. W. Dorrity, C. Queitsch, S. Fields, High-throughput identification of dominant negative polypeptides in yeast. *Nat Methods* **16**, 413–416 (2019).

18. K. M. Ford, *et al.*, Peptide-tiling screens of cancer drivers reveal oncogenic protein domains and associated peptide inhibitors. *Cell Systems* **12**, 716-732.e7 (2021).

19. A. Savinov, F. P. Roth, Seeds of their own destruction: Dominant-negative peptide screening yields functional insight and therapeutic leads. *Cell Systems* **12**, 691–693 (2021).

20. A. Savinov, A. Fernandez, S. Fields, Mapping functional regions of essential bacterial proteins with dominant-negative protein fragments. *Proceedings of the National Academy of Sciences* **119**, e2200124119 (2022).

21. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

22. H. M. Berman, *et al.*, The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).

23. F. Morcos, *et al.*, Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences* **108**, E1293–E1301 (2011).

24. T. Tsaban, *et al.*, Harnessing protein folding neural networks for peptide–protein docking. *Nat Commun* **13**, 176 (2022).

25. L. Chang, A. Perez, Ranking Peptide Binders by Affinity with AlphaFold. *Angewandte Chemie* **135**, e202213362 (2023).

26. A. Motmaen, *et al.*, Peptide-binding specificity prediction using fine-tuned protein structure prediction networks. *Proceedings of the National Academy of Sciences* **120**, e2216697120 (2023).

27. N. R. Bennett, *et al.*, Improving de novo protein binder design with deep learning. *Nat Commun* **14**, 2625 (2023).

28. K. Tunyasuvunakool, *et al.*, Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).

29. M. Mirdita, *et al.*, ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679–682 (2022).

30. R. Evans, *et al.*, Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.10.04.463034 (2022).

31. I. R. Humphreys, *et al.*, Computed structures of core eukaryotic protein complexes. *Science* **374**, eabm4805 (2021).

32. M. A. Schumacher, T. Ohashi, L. Corbin, H. P. Erickson, High-resolution crystal structures of Escherichia coli FtsZ bound to GDP and GTP. *Acta Cryst F* **76**, 94–102 (2020).

33. F. Yang, L. G. Moss, G. N. Phillips, The molecular structure of green fluorescent protein. *Nat Biotechnol* **14**, 1246–1251 (1996).

34. R. Y. Tsien, The Green Fluorescent Protein. *Annual Review of Biochemistry* **67**, 509–544 (1998).

35. K. A. J. A. Gardner, D. A. Moore, H. P. Erickson, The C-terminal linker of Escherichia coli FtsZ functions as an intrinsically disordered peptide. *Molecular Microbiology* **89**, 264–275 (2013).

36. P. J. Buske, P. A. Levin, A flexible C-terminal linker is required for proper FtsZ assembly in vitro and cytokinetic ring formation in vivo. *Molecular Microbiology* **89**, 249–263 (2013).

37. M. C. Cohan, A. M. P. Eddelbuettel, P. A. Levin, R. V. Pappu, Dissecting the Functional Contributions of the Intrinsically Disordered C-terminal Tail of Bacillus subtilis FtsZ. *Journal of Molecular Biology* **432**, 3205–3221 (2020).

38. M. K. Shinn, *et al.*, Connecting sequence features within the disordered C-terminal linker of Bacillus subtilis FtsZ to functions and bacterial cell division. *Proceedings of the National Academy of Sciences* **119**, e2211178119 (2022).

39. S. Pichoff, J. Lutkenhaus, Identification of a region of FtsA required for interaction with FtsZ. *Molecular Microbiology* **64**, 1129–1138 (2007).

40. J. Conti, M. G. Viola, J. L. Camberg, FtsA reshapes membrane architecture and remodels the Z-ring in Escherichia coli. *Molecular Microbiology* **107**, 558–576 (2018).

41. L. Mosyak, *et al.*, The bacterial cell-division protein ZipA and its interaction with an FtsZ fragment revealed by X-ray crystallography. *The EMBO Journal* **19**, 3179–3191 (2000).

42. B. Shen, J. Lutkenhaus, The conserved C-terminal tail of FtsZ is required for the septal localization and division inhibitory activity of MinCC/MinD. *Molecular Microbiology* **72**, 410–424 (2009).

43. K.-T. Park, A. Dajkovic, M. Wissel, S. Du, J. Lutkenhaus, MinC and FtsZ mutant analysis provides insight into MinC/MinD-mediated Z ring disassembly. *Journal of Biological Chemistry* **293**, 5834–5846 (2018).

44. C. J. LaBreck, J. Conti, M. G. Viola, J. L. Camberg, MinC N- and C-Domain Interactions Modulate FtsZ Assembly, Division Site Selection, and MinD-Dependent Oscillation in Escherichia coli. *Journal of Bacteriology* **201**, 10.1128/jb.00374-18 (2019).

45. N. del Toro, *et al.*, The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Research* **50**, D648–D653 (2022).

46. D. Szklarczyk, *et al.*, The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research* **51**, D638–D646 (2023).

47. R. Oughtred, *et al.*, The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci* **30**, 187–200 (2021).

48. C. Y. Lee, *et al.*, Systematic discovery of protein interaction interfaces using AlphaFold and experimental validation. *bioRxiv*, 2023.08.07.552219 (2023).

49. D. Yu, G. Chojnowski, M. Rosenthal, J. Kosinski, AlphaPulldown—a python package for protein–protein interaction screens using AlphaFold-Multimer. *Bioinformatics* **39**, btac749 (2023).

50. Y. Zhang, J. Skolnick, Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics* **57**, 702–710 (2004).

51. F. J. O'Reilly, *et al.*, Protein complexes in cells by AI-assisted structural proteomics. *Molecular Systems Biology* **19**, e11544 (2023).

52. P. Bryant, G. Pozzati, A. Elofsson, Improved prediction of protein-protein interactions using AlphaFold2. *Nat Commun* **13**, 1265 (2022).

53. S. Kosuri, G. M. Church, Large-scale de novo DNA synthesis: technologies and applications. *Nat Methods* **11**, 499–507 (2014).

54. A. Hoose, R. Vellacott, M. Storch, P. S. Freemont, M. G. Ryadnov, DNA synthesis technologies to close the gene writing gap. *Nat Rev Chem* **7**, 144–161 (2023).

55. H. Q. Nguyen, *et al.*, Quantitative mapping of protein-peptide affinity landscapes using spectrally encoded beads. *eLife* **8**, e40499 (2019).

56. Y. Feng, X. Zhao, A. K. White, K. C. Garcia, P. M. Fordyce, A bead-based method for high-throughput mapping of the sequence- and force-dependence of T cell activation. *Nat Methods* **19**, 1295–1305 (2022).

57. A. Portell, *et al.*, Reprogramming Adeno-Associated Virus Tropism Via Displayed Peptides Tiling Receptor-Ligands. *bioRxiv*, 2022.09.26.509383 (2022).

58. K. Tsuboyama, *et al.*, Mega-scale experimental analysis of protein folding stability in biology and design. *Nature* **620**, 434–444 (2023).

59. C. J. Layton, P. L. McMahon, W. J. Greenleaf, Large-Scale, Quantitative Protein Assays on a High-Throughput DNA Sequencing Chip. *Molecular Cell* **73**, 1075-1082.e4 (2019).

60. T. Zhang, *et al.*, Direct activation of a bacterial innate immune system by a viral capsid protein. *Nature* **612**, 132–140 (2022).

61. M. Mirdita, M. Steinegger, J. Söding, MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **35**, 2856–2858 (2019).

62. T. T. Tanimoto, *An Elementary Mathematical Theory of Classification and Prediction* (International Business Machines Corporation, 1958).

63. A. H. Lipkus, A proof of the triangle inequality for the Tanimoto distance. *Journal of Mathematical Chemistry* **26**, 263–265 (1999).

64. A. Reuther, *et al.*, Interactive Supercomputing on 40,000 Cores for Machine Learning and Data Analysis in *2018 IEEE High Performance Extreme Computing Conference (HPEC)*, (2018), pp. 1–6.

Supplementary Information

Supplementary Figures

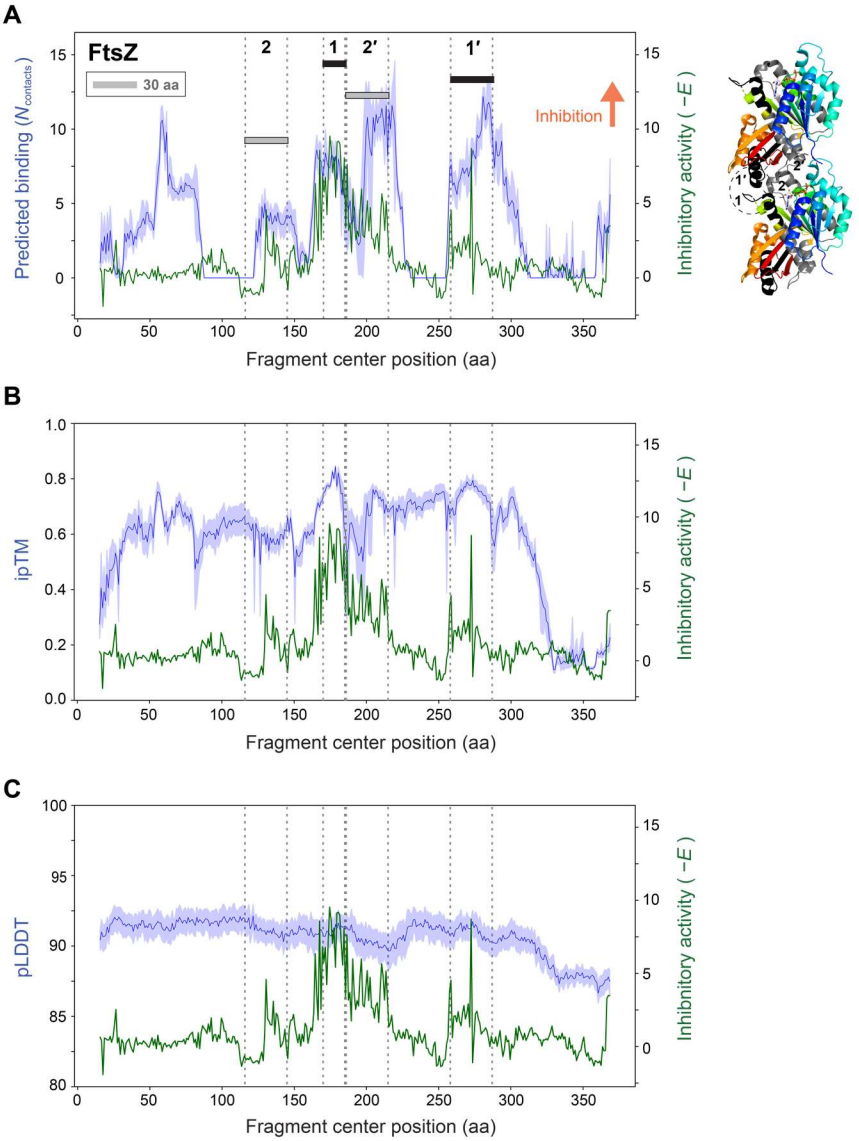

Figure S1. Unweighted $N_{contacts}$ predicts inhibitory fragment peaks; AlphaFold confidence metrics alone do not. (**A**) As in Figure 2*A*, but with $N_{contacts}$ (unweighted) instead of ipTM-weighted $N_{contacts}$ shown. *Left*: Predicted target protein binding from FragFold, unweighted $N_{contacts}$ (blue curves) vs. experimentally measured *in vivo* inhibitory activity (green curves, from ref. 20) for 30 aa fragment scans across *E. coli* FtsZ. Darker line and lighter outline for predicted binding data indicate mean and 95% C.I., respectively, across the 5 top-ranked structural models generated by AlphaFold. Filament formation interface regions **2**, **1**, **1′**, and **2′** are indicated. *Right*: crystallographic structure of two adjacent monomers of the FtsZ filament (PDB ID 6unx, ref. 32). Protein is colored from blue to red from the N- to the C-terminus. Regions corresponding to interface regions are annotated and highlighted in the same color scheme as in the fragment scan plot (*Left*). (**B**) As in (A), but ipTM is shown in blue. (**C**) As in (A), but pLDDT is shown in blue.

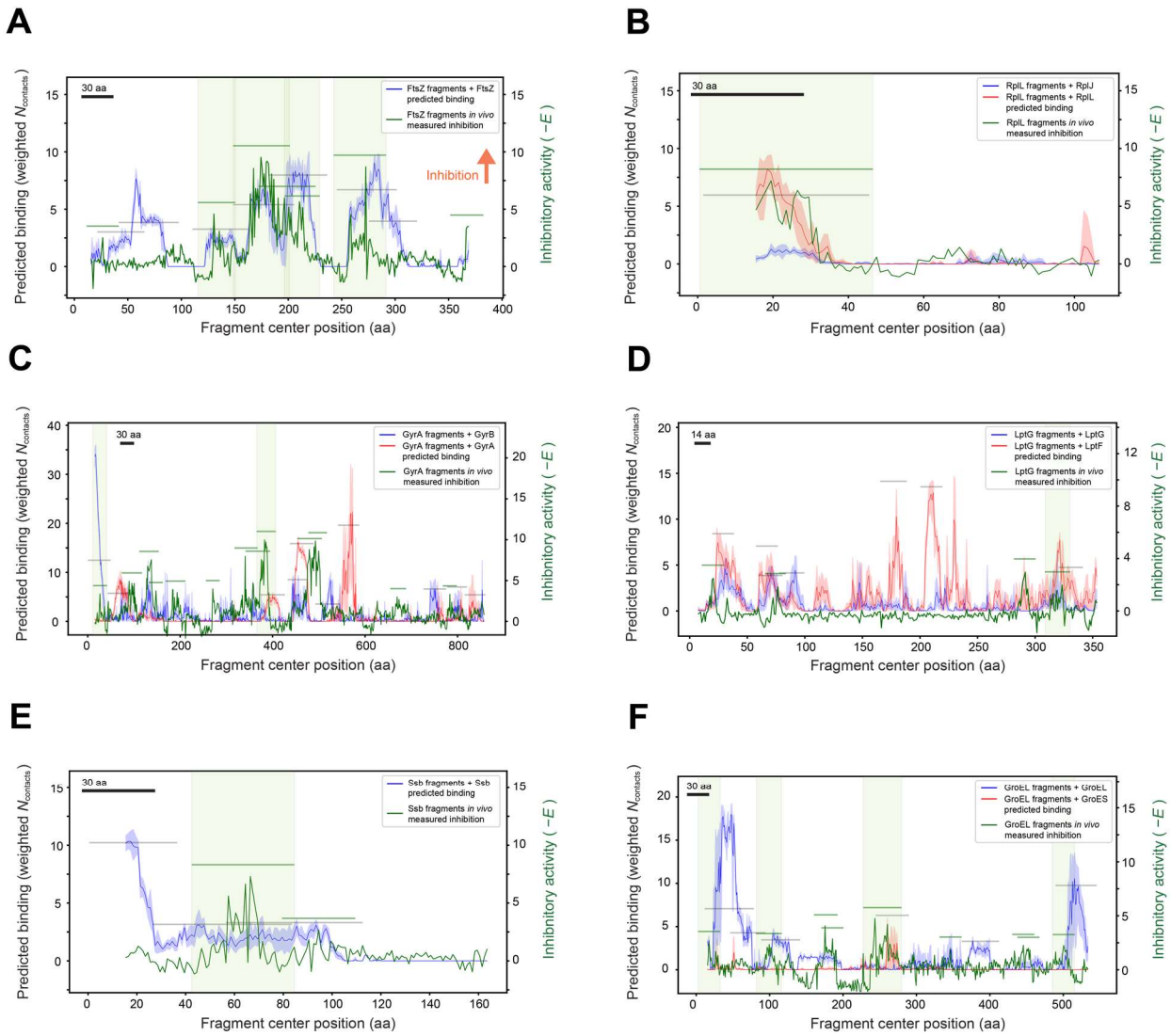

Figure S2. Automated peak discovery systematically reveals matching peaks between experimental inhibitory effects and predicted binding data. Predicted target protein binding (red, blue curves) from FragFold vs. experimentally measured *in vivo* inhibitory activity (green curves, from ref. 20) for fragment scans across 6 diverse *E. coli* proteins. Green highlights indicate peaks mapping to previously identified protein-protein interactions with corresponding inhibitory protein fragments (**Table 1** and ref. 20). Darker line and lighter outline for predicted binding data indicate mean and 95% C.I. across the 5 top-ranked structural models from AlphaFold. Green and grey bars indicate peaks called by the peak-calling algorithm: green bars, inhibitory peaks from experimental data; grey, predicted binding clusters from FragFold results. Fragment scan results are shown for (**A**) FtsZ, (**B**) RplL (L7/L12), (**C**) GyrA, (**D**) LptG, (**E**) Ssb, and (**F**) GroEL. 30 aa fragments were used for all proteins except LptG, which was scanned with 14 aa fragments. aa, amino acids. Fragment sizes are also indicated by the black scale bar in each panel.

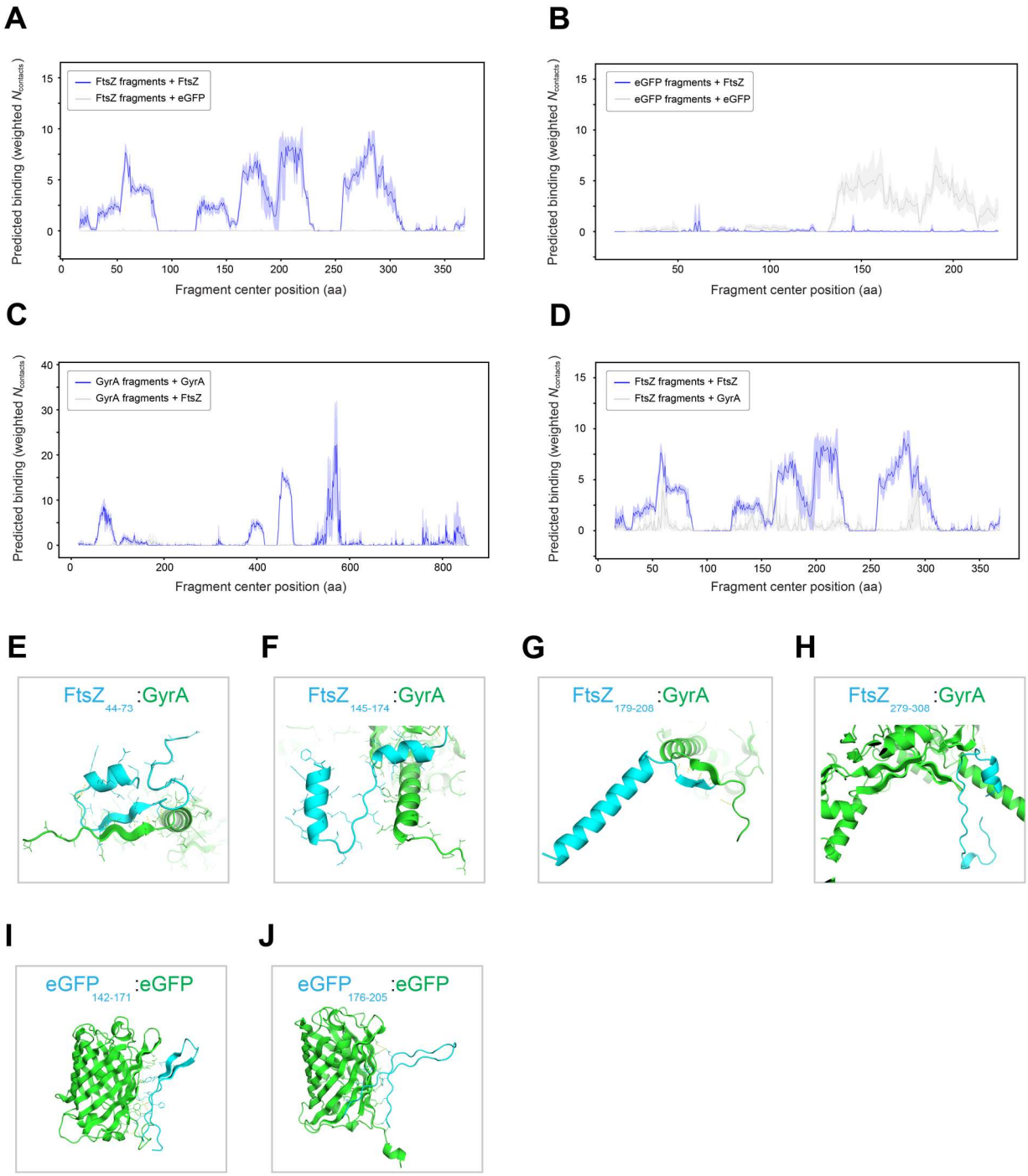

Figure S3. AlphaFold predictions of fragment-protein binding exhibit specificity. (A) – **(D)** Predicted target protein binding from FragFold for 30 aa fragment scans + target protein, depicted as in Figure 2. Colors are as indicated: (**A**) Blue, fragments of FtsZ + FtsZ; grey, fragments of FtsZ + eGFP; (**B**) Blue, fragments of eGFP + FtsZ; grey, fragments of eGFP + eGFP; (**C**) Blue, fragments of GyrA + GyrA; grey, fragments of GyrA + FtsZ; (**D**) Blue, fragments of FtsZ + FtsZ; grey, fragments of FtsZ + GyrA. aa, amino acids. (**E**) – (**H**) AlphaFold structural models of indicated fragments of FtsZ (cyan) + GyrA (green), corresponding to grey peak regions from (D). (**I**) – (**J**) As (E) – (H), but for fragments of eGFP + eGFP, corresponding to grey peak regions from (B).

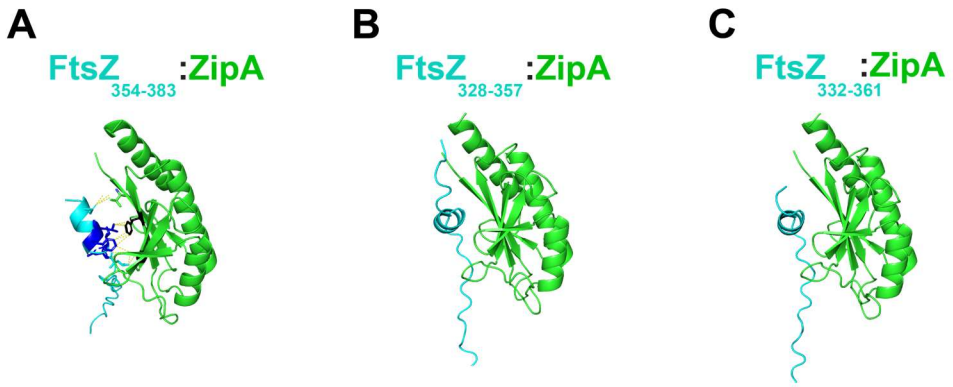

Figure S4. Additional FtsZ C-terminal tail fragments are predicted to bind the same face of ZipA. AlphaFold-predicted structures for indicated fragments of FtsZ (cyan) + ZipA (green). (**A**) Previously identified interacting regions on FtsZ (blue) and ZipA (black) are shown, as in Figure 4*E*, with residue-residue contacts shown as dashed yellow lines. (**B**) – (**C**) AlphaFold models for FtsZ fragments corresponding to FtsZ IDR peaks outside of the extreme C-terminal region (see Figure 4*A*).

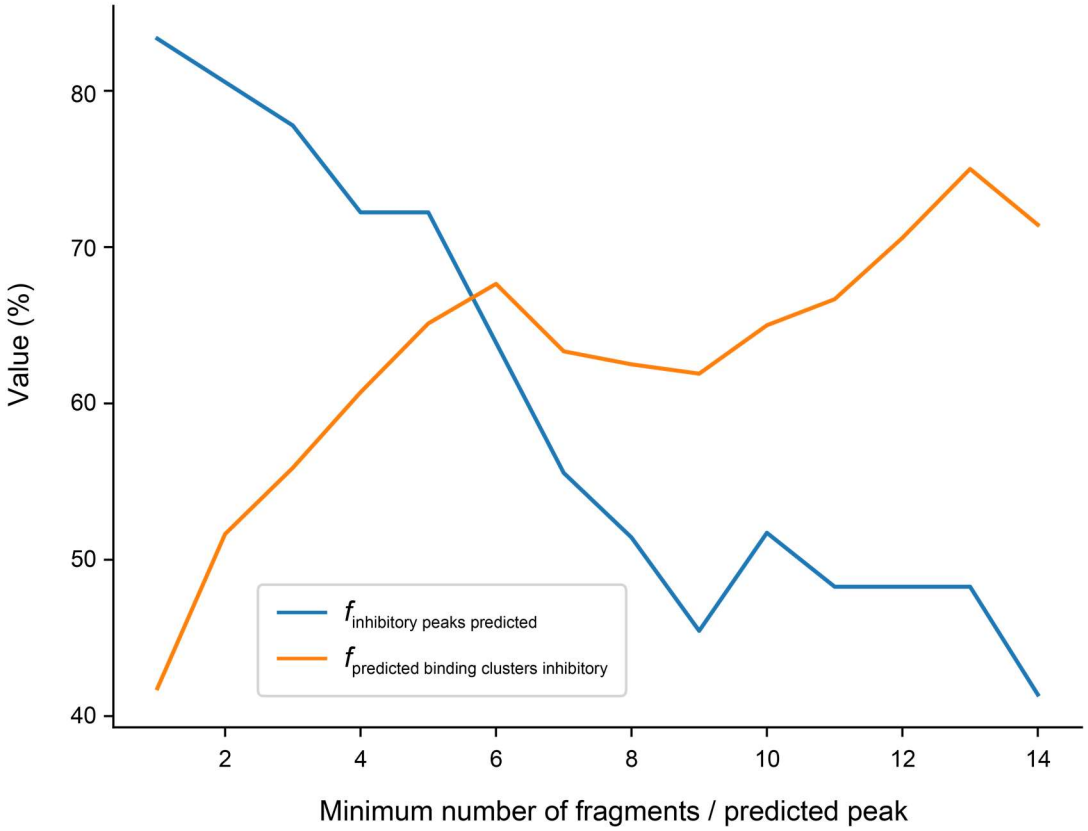

Figure S5. The minimum predicted peak width for automated inhibitory peak prediction was selected based on an optimum of sensitivity and specificity. Sensitivity ($f_{\text{inhibitory peaks predicted}}$) vs. specificity ($f_{\text{predicted binding clusters inhibitory}}$) plot is shown as parametrized by the minimum number of fragments per predicted binding peak (Tanimoto cluster size) accepted by the peak-calling algorithm.

Supplementary Tables

Table S1. FragFold systematic predictions outperform a random model.

Fragmented protein	$f_{\text{inhibitory peaks predicted}}$	$f_{\text{predicted binding clusters inhibitory}}$	$f_{\text{null, inhibitory peaks predicted}}$	$f_{\text{null, predicted binding clusters inhibitory}}$
All proteins	**64%**	**68%**	**49%**	**51%**
FtsZ	86%	71%	69%	66%
GroEL	44%	67%	39%	55%
GyrA	54%	70%	45%	54%
LptG	75%	57%	51%	34%
RplL	100%	100%	34%	34%
Ssb	100%	67%	63%	42%

As in Table 2, The fraction of measured inhibitory peaks that matched corresponding predicted binding peaks ($f_{\text{inhibitory peaks predicted}}$) as well as the fraction of predicted binding peaks that matched corresponding experimentally determined inhibitory peaks ($f_{\text{predicted binding clusters inhibitory}}$) are listed. These results are presented for all fragmented proteins ("All proteins"), as well as for each protein individually. Additionally, the null values of $f_{\text{null, inhibitory peaks predicted}}$ and $f_{\text{null, predicted binding clusters inhibitory}}$ from the model of random peak placement (see *Materials and Methods*) are listed. Overall *p*-values for the extent to which FragFold outperforms the null model in sensitivity and specificity are as follows: $p_{\text{predicted binding clusters inhibitory}} = 0.048$, $p_{\text{inhibitory peaks predicted}} = 0.035$.