

Interaction of the maturation protein of the bacteriophage MS2 and the sex pilus of the *Escherichia coli* F plasmid

Timothy J. Spankie,[‡] Alexe L. Haywood,[‡] Tania Dottorini,[§] Paul A. Barrow,[§]
and Jonathan D. Hirst^{*,‡}

[‡]*School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD,
UK*

[§]*School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington
Campus, Loughborough, LE12 5RD, UK*

E-mail: jonathan.hirst@nottingham.ac.uk

Abstract

One promising strategy to combat antimicrobial resistance is to use bacteriophages that attach to the sex pili produced by transmissible antimicrobial resistance (AMR) plasmids, infect AMR bacteria and select for loss of the AMR plasmids, prolonging the life of existing antimicrobials. The maturation protein of the bacteriophage MS2 attaches to the pili produced by Incompatibility group F plasmid-containing bacteria. This interaction initiates delivery of the viral genetic material into the bacteria. Using protein-protein docking we constructed a model of the F pilus comprising a trimer of subunits binding to the maturation protein. Interactions between the maturation protein and the F pilus were investigated using molecular dynamics simulations. *In silico* alanine scanning and *in silico* single-point mutations were explored, with the longer term aim of increasing the affinity of the maturation protein to other Incompatibility group pili, without reducing the strength of binding to F pilin. We report our computational findings on which residues are required for the maturation protein and F pilin to interact, those which had no effect on the interaction and the mutations which led to a stronger interaction.

Antimicrobial resistance (AMR) is a global issue which threatens the treatment of infectious diseases in humans and animals and its use for growth stimulation in agriculture.¹ The natural phenomenon of evolution conferring resistance has been accelerated by inappropriate use of antimicrobials. This includes prescription of antimicrobials for viral infection (which is ineffective), failure to complete courses of antibiotics, thereby allowing more resistant strains to survive, and the overuse of antimicrobials in farming for short-term benefits. Recent developments in antibiotics have stalled due to challenging regulatory requirements and lowered economic incentives.^{2,3} Even for the few antibiotics released recently, resistance was observed very rapidly.⁴ Attempts to manage AMR are expensive, and new approaches are being advocated including reversing the selective advantage of resistance in order to render microbes susceptible to treatment again.⁵

Despite the morphological simplicity of bacteria, the cell structure is well developed, having unique biological structures and widely known pathogenicity.⁶ Antimicrobial resistance is mainly encoded by semi-autonomous genetic elements called plasmids which may contain genes for antibiotic resistance and other phenotypes. Many plasmids can be transferred between bacteria through a process of genetic information transfer called conjugation.^{7,8} During conjugation adhesive hair-like pili produced by the AMR donor cell can attach to antibiotic sensitive recipient cells which draws the cells close together facilitating replication and ultimately transfer of the AMR plasmid.^{6,8}

Like many bacterial surface structures the pili act as attachment sites for specific bacterial viruses called bacteriophages. Recent studies have shown that application of such bacteriophages to an AMR culture *in vitro*^{9,10} and *in vivo*¹¹ selects for the rare antibiotic sensitive bacterial cells which have lost their plasmid spontaneously and are therefore bacteriophage-resistant and which eventually dominate the culture. Like all viruses, bacteriophages consist of a protein shell that encapsulates either a DNA or RNA genome.^{12,13}

Plasmids can be classified using a system based on the similarity of their replication process¹⁴ which leads to plasmids belonging to the same group being incompatible within an individual bacterial cell.^{15,16} Amongst AMR plasmids the dominant Incompatibility group is F although other groups including I, N and H are also important.¹⁷⁻¹⁹

If a bacteriophage could drive evolution of a bacteria clone towards loss of the antibiotic resistant plasmid, through selecting for plasmid-free bacteria the life of existing antibiotics could be extended. There is a logistical advantage in using a single bacteriophage for this approach and extending the host range of a F-specific bacteriophage to include other Incompatibility group plasmids would be a clear advantage. The expansion of phage host range has the benefits of not being reliant on the lack of control over natural evolution and the additional advantage over difficulties in hybridising multiple existing phages.²⁰

In our study the interaction between the F-specific bacteriophage MS2 maturation protein and the sex pilus produced by the F plasmid *E. coli* is examined. Their structures are shown in Figure 1.²¹⁻²³ The F pilin consists of 64 protein-phospholipid subunits arranged into five helical turns such that each subunit interacts with eight other subunits. MS2 infects *E. coli* via the lytic cycle, beginning with attachment to the sex pilus of the bacterial host. As the pilus retracts, the ssRNA genome is forced from the capsid shell and is delivered into the host. The bacteriophage consists of 89 coat protein dimers arranged in a $T = 3$ icosahedral lattice.²⁴ Three conformations are adopted, denoted A, B and C which form a symmetric CC dimer and an asymmetric AB dimer. In these coat proteins there is a crucial 'FG' loop region (residues His55 to Ser78) which plays a key role in the self-assembly of the capsid.²⁵⁻²⁷ Embedded into the capsid wall is a single copy of the maturation protein which binds to the pilus and protects the ssRNA genome inside. A 50 ns molecular dynamics (MD) simulation found that pores in the capsid wall allow transport of ions and water in both directions.²⁶

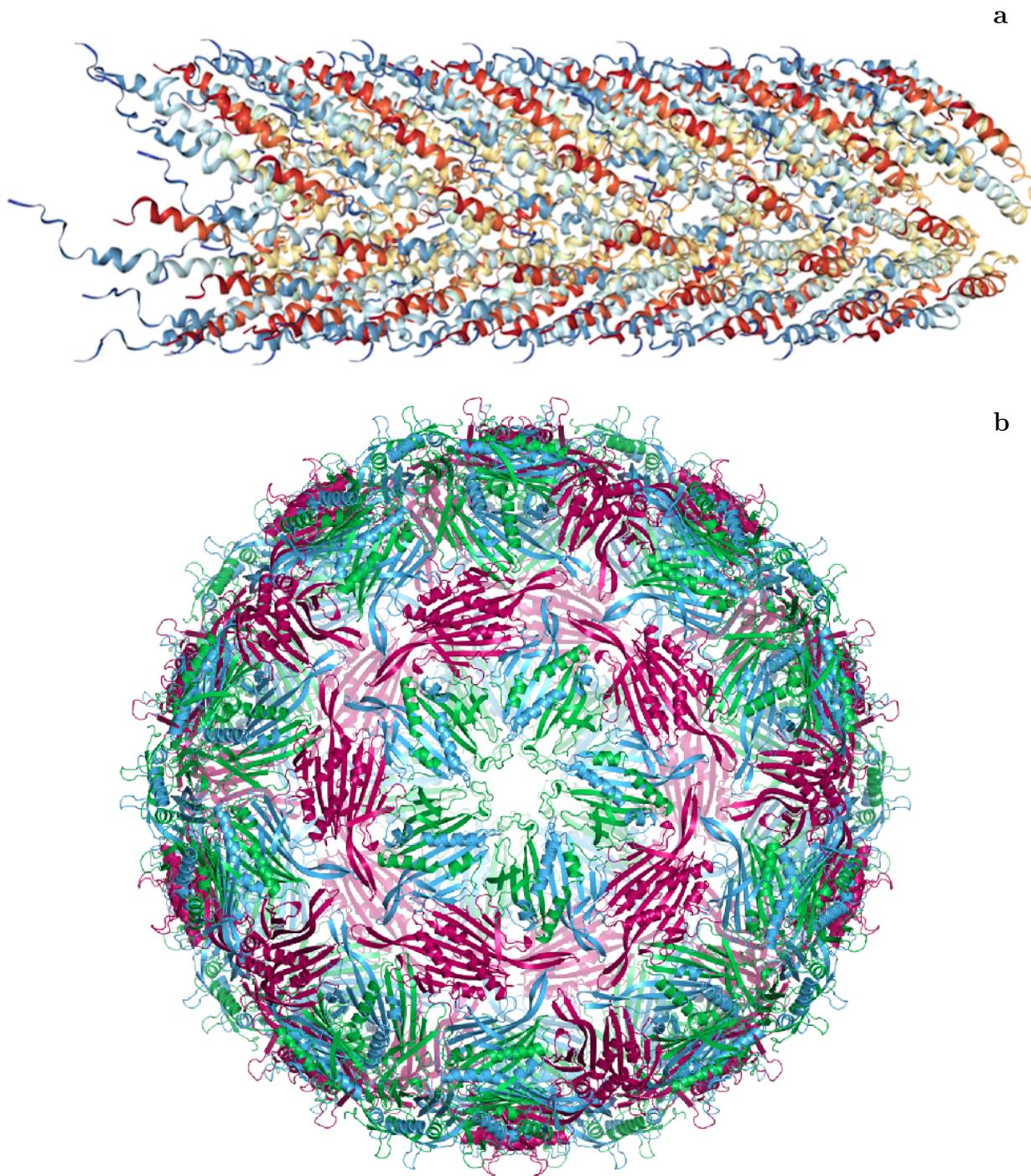


Figure 1: **a.** Structure of the bacterial sex F pilus (PDB ID 5LER).^{21,22} **b.** Bacteriophage MS2, showing the coat protein arranged into an icosahedral shell with triangulation number $T = 3$ (PDB ID 2MS2). Reproduced with permission from Toropova et al.²³

Computational approaches can provide useful and increasingly quantitative insights into

protein-protein interactions.^{28,29} Whilst bacteriophages have been studied extensively, until recently there has been little documentation of their protein-protein interactions beyond the study of Caudovirales.^{30,31} There were several key objectives for our study. A model was generated of the complex between MS2 maturation protein and three sub-units of the F pilin. The initial model was constructed with protein-protein docking. MD simulations were performed to assess the stability of the interaction between MS2 and F pilin in the model. In silico point mutations of the maturation protein were explored to investigate how one might increase the selectivity of the MS2 bacteriophage, so that other Incompatibility group pili could act as targets for attachment. This would allow for increased efficacy of MS2. We also explored which residues are essential to allow the maturation protein to interact with the F pilus. Mutations of residues were selected through prediction of the changes in binding affinity, which suggest how the protein-protein interactions are altered.³²

Methods

A single subunit of the F pilus (Figure 2a) consists of three parallel α -helices; the structure has been determined by Costa et al.²¹ to a 5 Å resolution through electron-counting cryo Electron Microscopy (cryoEM) and helical reconstruction. The subunits each with a phospholipid unit combine to form the tubular F pilus (Figure 1a). The maturation protein (Figure 2b) consists of an α -helix domain and a β -sheet domain and its structure has been determined by Dai et al.²⁴ to a 3.6 Å resolution (PDB code 5TC1) through a combination of cryoEM and asymmetric reconstruction. The α -helix domain (residues Gln140 to Phe225, Asp269 to Pro313 and Phe375 to Arg393) lies inside the capsid structure.²⁴ The β -sheet domain contains a helix-loop-helix motif (residues Tyr86 to Asn138) which is able to interact with the F pilus³³ as the β -sheet domain (residues Met1 to Leu139, Leu226 to Asn268 and Val314 to Val378) is outside the MS2 capsid. The residues in the loops connecting the β -sheet were not located in the experiment. Thus, we rebuilt the structure of the MS2 mat-

uration protein, using homology modelling with 5TC1 used as the template structure. The bacteriophage is limited to where it can bind, as it does not enter the pilus, and consequently the interior face of the pilus is blocked. For our modelling, we focus on a trimer of subunits taken from the structure of Costa et al.²¹ (PDB code 5LFB). Using a trimer (rather than a monomer) means that the model will include interactions between subunits.

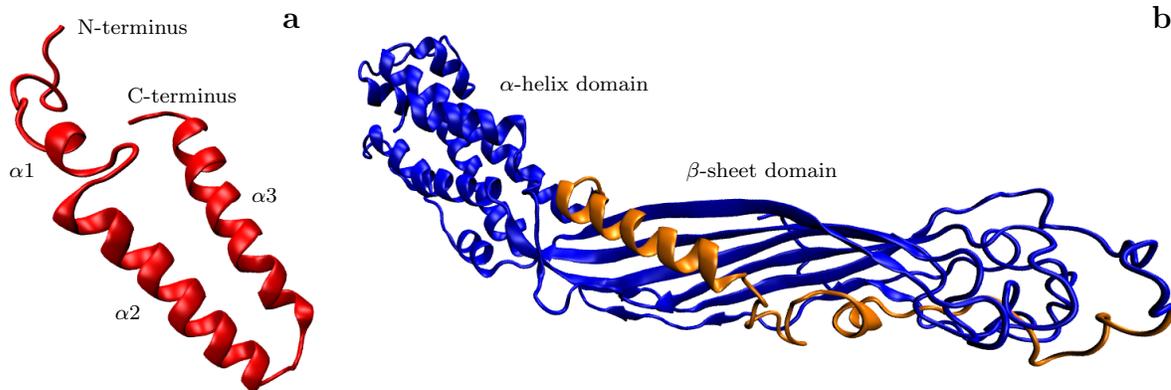


Figure 2: **a.** F Pilin structure. **b.** Maturation protein structure with the helix-loop-helix motif shown in orange.

Protein-protein docking was performed using HADDOCK (High Ambiguity Driven protein-protein DOCKing)^{34,35} and RosettaDock.³⁶ To determine their respective merits, histograms were made for the energies reported by HADDOCK and RosettaDock. For the HADDOCK structures, the 200 top structures were taken from the final stage of docking of four independent runs. Thus, a total of 800 structures were generated using the following rules for selected residues for the F pilin. Run 1 used residues Gln1 to Gln16 and Leu65 of the F pilin as active residues. Run 2 used residues Gln1 to Gln16 and Met60 to Leu65 of the F pilin as active residues. Runs 3 and 4 were the first two runs repeated with additional passively selected residues, those residues within 6.5 Å of active residues. The middle subunit of the trimer was used for selection of active residues. All runs used residues Tyr86 to Asn138 of MS2 as active restraints. Active residues were selected based on a combination of homology modelling and protein threading. Crucially these portions of the proteins are solvent exposed. The restraints were selected on only the middle substructure of the F pilin trimer

in the case of HADDOCK. For RosettaDock 1,000 structures were generated using perturbation by a random direction and magnitude specified by a binomial-distribution around 0.1 Å and 3.0°. As HADDOCK does not consider non-polar hydrogen atoms, the similarities of structures generated by HADDOCK and RosettaDock were analysed using the MaxCluster software.³⁷

The docked complex of the maturation protein and the F pilin trimer, consisting of 9129 atoms was imported into the CHARMM-GUI^{38,39} and solvated in a truncated octahedron of 119,373 TIP3P water molecules,⁴⁰ and neutralised with 342 potassium ions and 360 chloride ions. The box was extended to 10 Å from the edge of the complex and periodic boundary conditions were applied. Bad contacts were removed through short minimisations using steepest descent and the adopted basis Newton-Raphson algorithm. MD simulations were performed using the CHARMM36 forcefield⁴¹ in NAMD 2.12⁴² with a 2 fs time step. A cutoff of 12 Å was used for van der Waals interactions, Particle Mesh Ewald (PME)⁴³ was used for long-range electrostatics and the SHAKE algorithm⁴⁴ was employed for the fixing of bonds including hydrogen. Constant pressure was maintained using the Langevin piston Nosé Hoover method^{45,46} with a target pressure of 1 atm. Langevin dynamics parameters were used to control the temperature, and were applied with a friction coefficient of 5 ps⁻¹ for heating, equilibration and dynamics. System preparation proceeded with a 10,000 step energy minimisation using the conjugate gradient algorithm. Backbone and sidechain atoms were restrained with harmonic restraints with force constants of 10.0 and 5.0 kcal mol⁻¹ Å⁻² respectively. The system was then heated to 298 K by 3 K every 1 ps, giving 500 steps between each velocity reassignment to equilibrate. Then five 500 ps equilibration phases under the NPT ensemble were performed with the force constant restraints reduced to zero by 2.5 kcal mol⁻¹ Å⁻² per phase. A 20 ns production simulation was run in the NPT ensemble.

MD trajectories were analysed using the python package MDTraj⁴⁷ and were visualised

using Visual Molecular Dynamics (VMD).⁴⁸ Root mean square deviation (RMSD) for all non-hydrogen atoms, with the reference structure taken from the start of production dynamics of the complex, were calculated using MDTraj

Protein-protein interactions were calculated by identifying residues within a certain distance from other residues, specifically if their C_β - C_β atoms were within 8 Å of each other. This distance was used so interactions with long side chains such as arginine, lysine, tryptophan and tyrosine would be included. Residues selected by this method were used for calculation of the interfacial RMSD.

Alanine scanning of the MS2 maturation protein in the complex was performed using the Robetta⁴⁹ server, in order to investigate the contribution of individual residues to the protein-protein binding energy. Residues to be mutated were selected by the predicted change in binding free energy of the complex upon mutation of the residue to alanine. Robetta calculates the binding free energy using equation 1.

$$\begin{aligned} \Delta G = & W_{attr}E_{LJattr} + W_{attr}E_{LJrep} + W_{HB(sc-bb)}E_{sc-bb} \\ & + W_{HB(sc-sc)}E_{sc-sc} + W_{sol}G_{sol} + W_{\phi/\psi}E_{\phi/\psi}(aa) + \sum_{aa=1}^{20} n_{aa}E_{aa}^{ref} \end{aligned} \quad (1)$$

where W is the relative weight of the different energy terms,⁵⁰ E_{LJattr} and E_{LJrep} are the attractive and repulsive Lennard-Jones terms, E_{sc-bb} and E_{sc-sc} are orientation-dependent hydrogen bond potentials for sidechain-backbone and sidechain-sidechain respectively, G_{sol} is an implicit solvation term, $E_{\phi/\psi}(aa)$ is an amino acid type dependent backbone torsion angle propensity and E_{aa}^{ref} is an amino acid type dependent reference energy, approximating interactions based on n_{aa} the number of amino acids of a certain type.

Robetta indicated so-called “hotspot” residues, i.e., those residues which upon mutation to alanine changed the free energy of binding by more than 1 kcal/mol. Potentially interesting mutants in the MS2 maturation protein were identified, based on the extent to which the mutation was predicted to stabilise or destabilise the interface. To test these predictions, MD simulations of the complexes with the mutant MS2 maturation protein were performed, using the same protocol described for the complex with the wild-type protein.

Results and Discussion

Comparison of docking models

Our first goal was to establish an initial model of the protein-protein complex using docking. The RosettaDock results showed the best scoring complexes gave scores of 642 REU with a range between 600 and 1,600 Rosetta Energy Units (REU). Restraints cannot be applied to the complex to encourage particular residues to dock in the RosettaDock software; however sidechains are able to move. This means that RosettaDock is predisposed to bind the edge monomers of the trimer rather than the central monomer for steric reasons. The scores of the complexes created by HADDOCK have a spread between -200 to 1,200 HADDOCK Score (HS). The best scoring complex with a single subunit of the F pilin has a score of -7.77 HS. Using a trimer of subunits gives a best score of -198 HS.

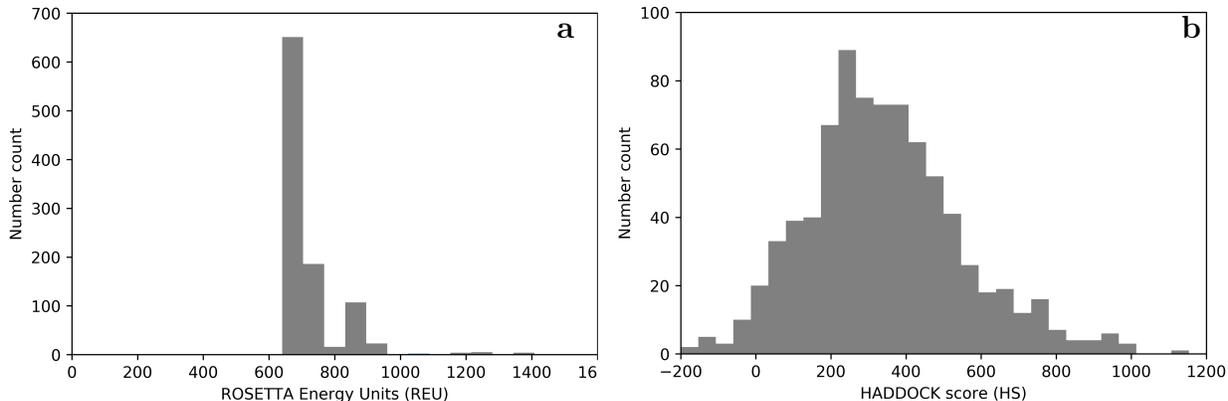


Figure 3: Histograms of the RosettaDock and HADDOCK structures. Panel **a** shows the top scores from the for 1,000 complexes created and scored by RosettaDock. Panel **b** shows the scores for 800 complexes created by and scored by HADDOCK.

As there is no correlation between the HADDOCK and RosettaDock scoring functions, the best scored poses of HADDOCK and RosettaDock were clustered using the MaxCluster algorithm.^{37,51} The structures were aligned with the backbone heavy atoms (C_α , N, O, C_β and C) then hierarchically clustered by maximum linkage into groups whose backbone atoms had a RMSD of less than 1 Å. The clustering of the HADDOCK structures is shown in Table 1. The top ten RosettaDock structures remained as a single cluster even with a cutoff of 0.001 Å, i.e., the RosettaDock algorithm results in identical structures.

Table 1: Cluster analysis with a cutoff of 1 Å. Clusters of only one structure have not been included.

Structure set	N ^o structures	N ^o clusters	Cluster size(s)
HADDOCK top 5%	40	7	8, 5, 4, 4, 3, 3, 2
HADDOCK best 10 [1]	10	2	4, 2
RosettaDock best 10 [2]	10	1	10
[1] + [2]	20	2	4(H) 10(R), 2(H)

There are a number of similar structures in the top 5% of HADDOCK structures. For HADDOCK the structures show several small clusters, where there several distinct structures that appear repeatedly. Four of the top ten HADDOCK structures form a cluster. This cluster at a cutoff of 1 Å matches the cluster that the ten best RosettaDock structures formed, which means both HADDOCK and RosettaDock agree that this structure is the strongest docked

structure.

Two qualitatively different orientations in the protein-protein complex were observed in the docking: one where the pilin is anti-parallel to the β -pleated denoted as orientation A and another where the pilin is parallel to the β -pleated sheets denoted as orientation B. The ratio of orientation A and B is shown in Table 2.

Table 2: Ratios of orientation A to orientation B observed in the docked structures.

Structure set	ratio A:B
All HADDOCK	381:419
HADDOCK best 10	9:1
HADDOCK best	1:0
All RosettaDock	977:23
RosettaDock best 10	10:0
RosettaDock best	1:0

RosettaDock consistently favours orientation A. For HADDOCK, the ratio shifts towards orientation B for all structures, but for the top structures orientation A is clearly favoured. Orientation A is shown in the Figure 4. This concurs with the orientation shown by Meng et al.,⁵² who very recently published cryo-EM data on the complex at resolutions between 5.6 and 7.3 Å. Our model building, simulations and analysis were completed before the study of Meng et al.⁵² was published. Our model of the protein-protein interface is in qualitative agreement with the experimental structure. Although it is beyond the scope of the current study, it would clearly be of interest in future work to investigate the structural differences in detail and to perform some MD simulations using the structure determined by Meng et al.⁵²

We have found concordance between the best results of HADDOCK and RosettaDock. As the variation between the two docking programs is low, the ability of HADDOCK to bind specifically to the middle monomer of the trimer leads us to take the best scoring HADDOCK model forward to study using MD simulations. The model constructed using

the trimer (rather than a monomer) of F pilin subunits is more realistic, despite the fewer contacts, due to including the steric blocking of the pilin from surrounding subunits.

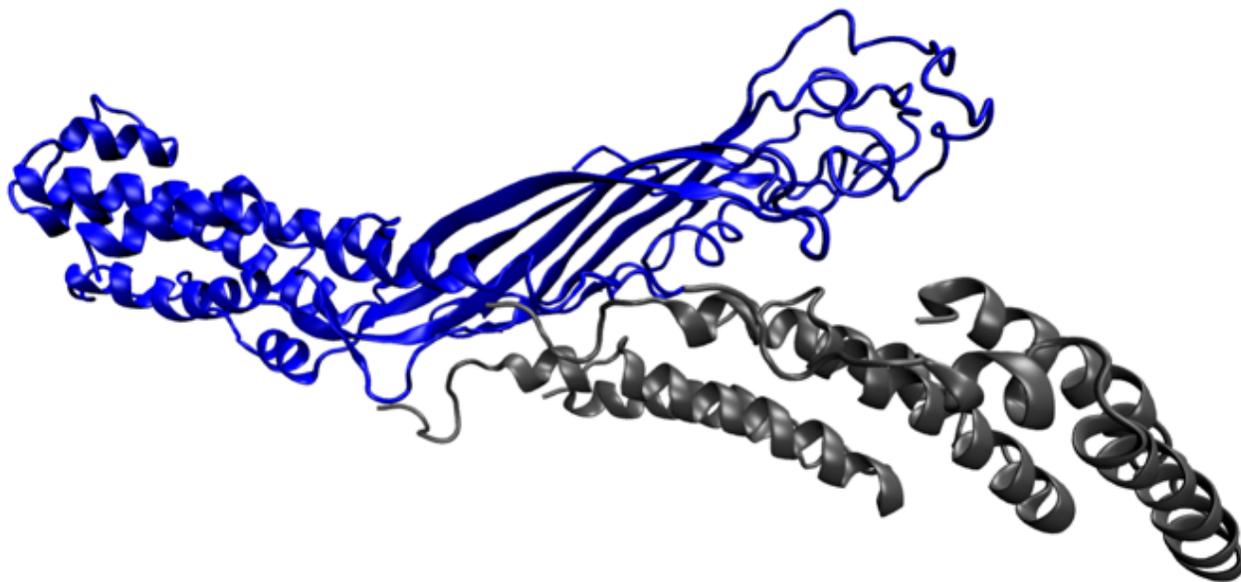


Figure 4: Docked structure of the maturation protein and trimer of F pilin.

Molecular Dynamics Simulations

MD simulations were performed on the best scoring HADDOCK docked complex of the maturation protein and the trimer of F pilin subunits, to explore the interaction in more detail with a more physically realistic model. The contacts present in the docked structure were monitored (Figure 5) to see if they persisted or dissociated throughout five independent 20 ns simulations.

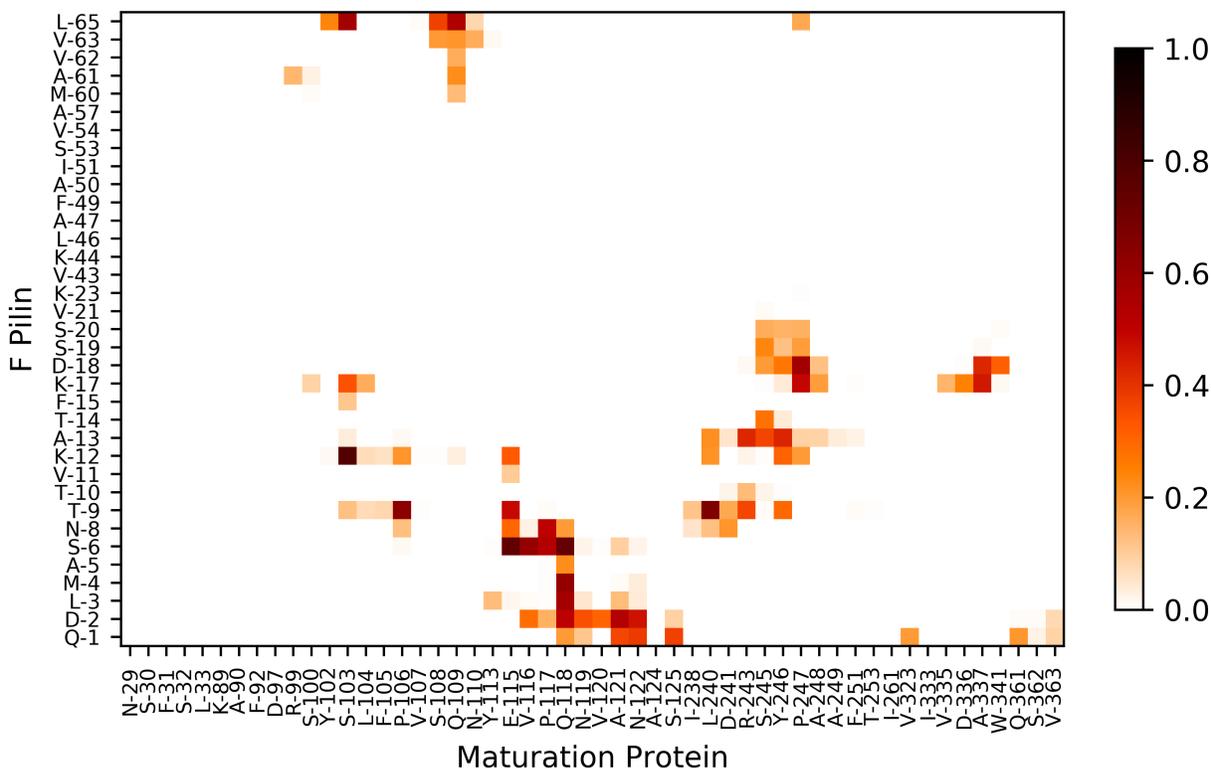


Figure 5: Map of contacts showing the fraction of time that residues between the trimer F pilin and the maturation protein stayed in contact for the average of five independent 20 ns simulations. The fraction of time (between 0 and 1) is shown using the scale in the colour bar. A contact was defined as being present if the C_{β} - C_{β} of two residues were within 8 Å of each other.

Some contacts that were present in the docked structure are lost during the preparation phase of the simulation. Relative to a single subunit of pilin minimal contacts were seen in the region of residues Asn29 to Arg99 of the maturation protein and Val43 to Leu65 of the pilin. Seven contacts between the maturation protein and F pilin are present after equilibration for all repeats of all the F pilin models: Ser103:Lys12 (maturation protein residue:F pilin residue), Val116:Ser6, Ala121:Asp2, Asn122:Asp2, Arg243:Ala13, Pro247:Asp18 and Ala337:Asp18. Just three contacts were present more than 50% of the time for all the simulations: Ser103:Lys12, Gln115:Ser6 and Glu118:Ser6 for 97%, 94% and 93% of the time respectively. These residues are within the same interaction region, but only Ser103-Lys12 is seen for the majority of the time in every trajectory.

The average RMSD for the complex (Figure 6) rises rapidly at the beginning (the RMSD of the individual trajectories are shown in Figure S1 of the Supplementary Information). However, little obvious movement is seen, reaching 2.1 Å after 1 ns, then increasing steadily throughout reaching a final average RMSD of 4.7 Å at 20 ns. The change in RMSD between 10 ns and 20 ns is 1.0 Å which is much smaller than the initial 10 ns of 3.6 Å. The standard deviation also falls to 0.29 Å for the second half of the simulation from 0.59 Å in the first half. The average interfacial RMSD remains lower than that of the complex for the majority of the simulation, but briefly rises above at 14 ns due to one repeat that is significantly above the rest, which results in the large variations shown for the interface. The interface is quite stable for the second half of production dynamics with a change in the RMSD of only 0.6 Å.

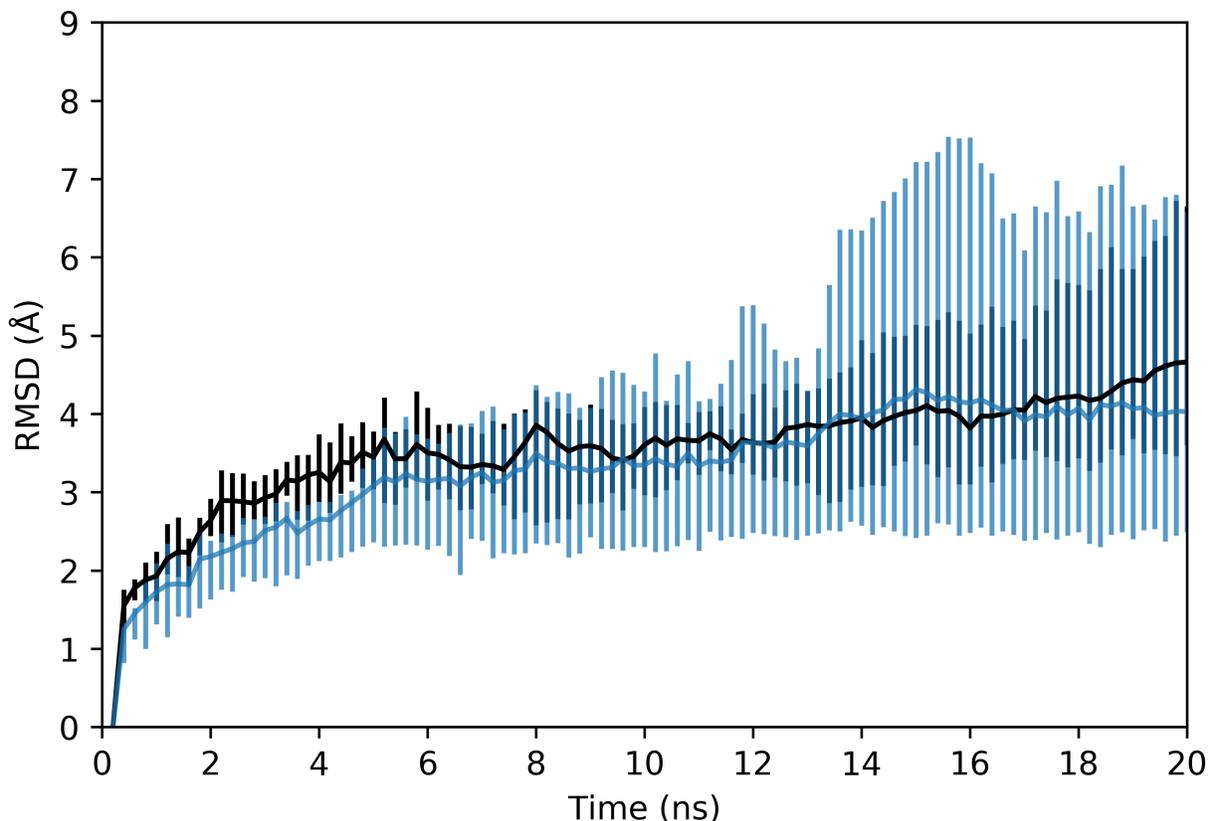


Figure 6: The RMSD of all non-hydrogen atoms for the whole complex (black), and the interface (blue) defined as residues within a C_{β} - C_{β} distance of 8 Å to the other protein at the start of production dynamics, averaged over the five independent simulations of the maturation protein - F pilin complex. Error bars show the maximum and minimum RMSD values at that timestep from all repeats

The individual MD trajectory with the lowest final fraction of native interfacial contacts (Figure 7, where native interfacial contacts are contacts with an C_{β} - C_{β} distance of 8 Å between F pilin and the maturation protein at the start of the simulation) shows a steady drop in the fraction of native interfacial contacts and higher RMSDs for both the complex and the interface. There are three trajectories with a high final native interfacial contact fraction at 0.67, 0.55 and 0.51. Two simulation repeats show weaker interactions, with final fractions of 0.26 and 0.09.

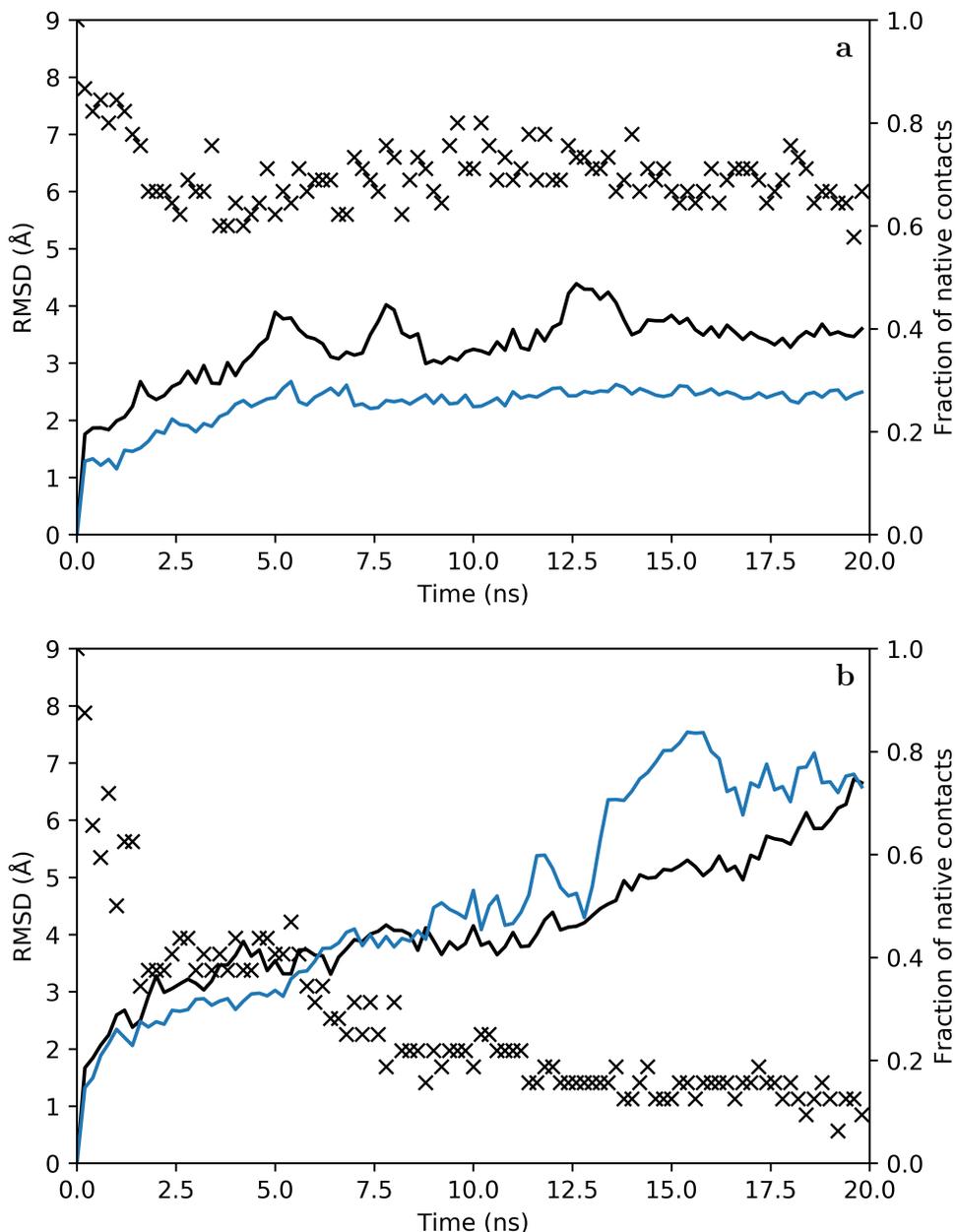


Figure 7: Fraction of native interfacial contacts and the RMSD of all non-hydrogen atoms, for the MS2 - F pilin complex (with three F pilin subunits) for the highest **(a)**, and lowest average percentage of native interfacial contacts **(b)**. The native interfacial contacts and the interface are defined as residues within a C_{β} - C_{β} distance of 8 Å to the other protein at the start of production dynamics.

Key: Complex (black), Interface (blue) and Fraction of native interfacial contacts (crosses).

Point Mutations on MS2 Maturation Protein

To explore which residues of the MS2 maturation protein should be mutated to modulate the binding specificity, alanine scanning was performed on the maturation protein - F pilin trimer complex, using the Robetta server for prediction of Gibbs free energy.⁴⁹ The aim was to find residues that when mutated have no effect, or a stabilising effect on the binding between MS2 and F pilin. Table 3 shows the residues that destabilised and stabilised the complex the most.

Table 3: Change in predicted Gibbs free energy upon mutation to alanine for the five most stabilising residues and five most destabilising residues of the MS2 maturation protein.

Effect	Residue	ΔG (kcal/mol)	Effect	Residue	ΔG (kcal/mol)
Stabilising	Gln118	-1.3	Destabilising	Arg259	7.2
	Trp341	-0.8		Trp364	3.6
	Asn122	-0.6		Ile114	2.6
	Gln361	-0.5		Val360	1.9
	Ser125	-0.5		Val323	1.7

20 ns MD simulations were performed on each of the complexes including one of these mutations, in order to investigate their stability. The variation in the number of persistent native interfacial contacts was quite large. Therefore, the consideration of possible mutations was expanded to further methods included within Robetta, including the Hotspot method. Thus, an extended set of mutants (Table 4) was generated. For each of these mutants, a 20 ns simulation of the protein-protein complex was performed.

Table 4: Calculated characteristics of the protein-protein interface from 20 ns simulations for each mutation. The complex is the docked protein - protein structure between the F pilin trimer and maturation protein of MS2. The n^o residues is the number of residues in the interface, defined as residues within a C_β - C_β distance of 8 Å to the other protein. Native contacts is the fraction of native interfacial contacts with respect to the number present at the start of production dynamics.

Mutation	Maximum RMSD (Å)		n^o residues	Native contacts min
	Complex	Interface		
Trimer (average)	6.5	5.6	50	0.31
Stabilising				
Gln118Ala	6.2	5.1	39	0.28
Trp341Ala	6.4	3.7	39	0.36
Asn122Ala	6.3	8.7	39	0.20
Gln361Ala	4.7	5.5	37	0.08
Ser125Ala	5.3	2.6	40	0.53
Ser17Ala	6.2	4.6	34	0.31
Asp382Ala	5.2	7.1	38	0.10
His191Ala	5.0	4.3	45	0.19
Asp298Ala	4.5	7.0	37	0.20
Ser103Ala	4.6	4.6	50	0.25
Val120Ala	5.3	5.0	43	0.31
Ser362Ala	4.5	5.4	35	0.12
Destabilising				
Arg259Ala	5.6	6.3	43	0.07
Trp364Ala	5.5	4.7	34	0.05
Ile114Ala	6.4	3.1	32	0.42
Val360Ala	5.6	4.6	39	0.26
Val323Ala	4.1	3.3	39	0.29
Cys255Ala	4.1	4.2	38	0.28
Tyr339Ala	4.4	2.3	40	0.50
Asp325Ala	7.2	4.9	34	0.24
Cys101Ala	5.9	3.3	38	0.24
Tyr246Ala	6.6	4.5	44	0.09
Tyr113Ala	6.9	5.5	48	0.18
Tyr102Ala	5.7	2.9	40	0.47
Ser108Ala	4.1	6.8	38	0.05
Gln109Ala	4.9	2.7	37	0.41
Val107Ala	4.9	2.6	42	0.60
Asp336Ala	5.6	5.8	46	0.20

The majority of destabilising residues produced native contact fractions lower than the wild-type models, indicating a reduction in the stability. The stabilising mutations produce similar or higher native interfacial contact fraction than the wild-type model, so would indicate suitability for replacement, as the mutation did not affect the stability of the complex.

There are 12 stabilising mutants and 16 destabilising mutants. For the former the average number of interfacial contacts is 39.7, and for the latter it 39.5. This modest difference is not significant. So the variation in stability largely arises from the strength of the specific interactions that are retained or lost.

MD simulations of the five mutants that lost the fewest native interfacial contacts, independent of their initial designation of stabilising and destabilising from Robetta were repeated a further four times with different initial velocities to improve the sampling and assess the reproducibility and variance. Data analogous that shown in Table 4 is summarised in Table S1 in the Supplementary Information. Ile114Ala showed the lowest native interfacial contact minimum of 0.28 (the average of five independent trajectories minimum values) but this was not significantly different from the wild type (0.31) so mutating this residue may still be worthwhile. Val107Ala was next with an average minimum of 0.36, followed by Ser125Ala at 0.40 and Tyr339Ala at 0.42. The mutation Tyr102Ala retained the greatest number of interfacial contacts with a minimum value of 0.43. This mutant, despite only 36 residues on average in the interface, showed the least fluctuation in the interface (RMSD of 3.0 Å). It would be recommended as the first site to target in engineering specificity for other Incompatibility group pili. As can be seen in Figure 8, the locations of the residues mutated are in the region that is closest to the pilin when docked (shown in Figure 4).

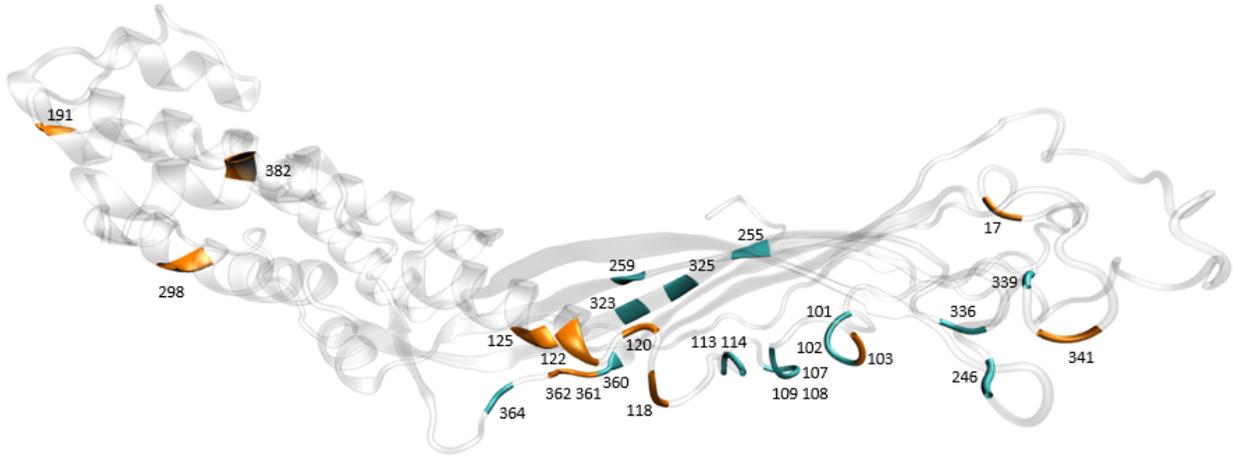


Figure 8: Location of mutated residues on the MS2 maturation protein. Mutations denoted as stabilising in Table 4 are coloured orange, and destabilising are coloured teal.

Of the seven residues that had contacts in all five independent simulations, two appeared in the alanine scanning, Ser103Ala and Asn122Ala. Both were extremely low scoring, suggesting they were required for the pilin and maturation protein interaction. Of the other five residues, two were alanine in the wild type. The other three were simulated twice to see if they also were crucial to the binding. The three residues Val116Ala, Arg243Ala and Pro247Ala all averaged a higher or equal minimum native interfacial contact fraction than the average trimer result of 0.31 at 0.36, 0.31 and 0.36 respectively, which indicates that the residues are not required for successful binding of the pilin. This increases our confidence in the alanine scanning, as these residue contacts were seen in all the independent trajectories of the trimer simulations. Those not selected by Robetta did not affect the stability of the complex.

We have found stabilising and destabilising mutations to alanine for the maturation protein based on an in silico approach. Certain residues such as Ser103 and Asn122 are required for the interaction to remain, and other mutations have little effect such as Val116Ala, Arg243Ala and Pro247Ala. Other mutations, in particular Ser125Ala increased the stability of the complex and therefore make good targets for broadening specificity.

Conclusions

We have developed a model of the maturation protein - F pilin protein-protein complex using three subunits of the F pilus. This represents a more realistic model than using just one F pilin subunit. A similar structure was found through two separate protein-protein docking programs. Contacts made between the maturation protein and F pilin have been identified. Point mutations have been studied *in silico* and identified residues that could be changed to allow other Incompatibility group pili to be susceptible to the MS2 bacteriophage. Other residues have been found to be crucial for binding of F pilin, with a large destabilising effect on the complex if removed.

This work has generated new pathways to be explored. The pilin model could be further extended with more subunits for a more accurate model, albeit at a greater computational cost. Multiple mutations could be made at the same time to identify the combined effect on the binding of F pilin. Other pili docking such as the I plasmid-specific pili, could be attempted to investigate which residues the mutagen sites could be replaced with. This would extend the operational range of the MS2 bacteriophage, currently limited to binding with F plasmid-specific pili. Mutations would need to be optimised to allow docking with the I specific pili without compromising the interaction with F pilin. These mutations can then be taken forward towards experimental studies either by targeted mutagenesis or by synthesising new plasmids *de novo*.

Funding

We thank the University of Nottingham for the use of the High Performance Computing (HPC) service Augusta, and the HPC Midlands Plus for the Tier 2 compute service Athena for access (ESPRC code: EP/P020232/1). The early stages of this work were supported by the EPSRC award “Bridging the Gaps: Systems-level approaches to antimicrobial resistance”

(EP/M027333/1). ALH is supported by EPSRC/NPIF (EP/S515516/1) through a CASE PhD studentship in partnership with GSK.

Supplementary Information

The supplementary information includes time evolution of the molecular dynamics simulations of the protein-protein complex and calculated characteristics of the protein-protein interface from the additional 20 ns simulations. The five 20 ns independent trajectories wild-type simulations are available on Zenodo (DOI:10.5281/zenodo.3739347).

References

- (1) World Health Organisation, *Global action plan on antimicrobial resistance.*; WHO Press, 2015; pp 1–28.
- (2) Ventola, C. L. (2015) The antibiotic resistance crisis: causes and threats. *Pharm. Ther.* *40*, 277–83.
- (3) Piddock, L. J. V. (2012) The crisis of no new antibiotics-what is the way forward? *Lancet Infect. Dis.* *12*, 249–253.
- (4) United States Centers for Disease Control, *Antibiotic resistance threats in the United States, 2013*; CDC, 2013; p 28.
- (5) Baym, M., Stone, L. K., and Kishony, R. (2016) Multidrug evolutionary strategies to reverse antibiotic resistance. *Science* *351*, aad3292.
- (6) Wang, X., Zheng, P., Ma, T., and Song, T. (2018) Small Universal Bacteria and Plasmid Computing Systems. *Molecules* *23*, 1307.

- (7) Michod, R. E., Bernstein, H., and Nedelcu, A. M. (2008) Adaptive value of sex in microbial pathogens. *Infect. Genet. Evol.* 8, 267–285.
- (8) Arutyunov, D., and Frost, L. S. (2013) F conjugation: Back to the beginning. *Plasmid* 70, 18–32.
- (9) Jalasvuori, M., Friman, V.-P., Nieminen, A., Bamford, J. K. H., and Buckling, A. (2011) Bacteriophage selection against a plasmid-encoded sex apparatus leads to the loss of antibiotic resistance plasmids. *Biol. Lett.* 7, 902–905.
- (10) Ojala, V., Laitalainen, J., and Jalasvuori, M. (2013) Fight evolution with evolution: Plasmid-dependent phages with a wide host range prevent the spread of antibiotic resistance. *Evol. Appl.* 6, 925–932.
- (11) Colom, J., Batista, D., Baig, A., Tang, Y., Liu, S., Yuan, F., Belkhiri, A., Marcelino, L., Barbosa, F., Rubio, M., Atterbury, R., Berchieri, A., and Barrow, P. (2019) Sex pilus specific bacteriophage to drive bacterial population towards antibiotic sensitivity. *Sci. Rep.* 9, 12616.
- (12) Brüssow, H., and Hendrix, R. W. (2002) Phage Genomics: Small is beautiful. *Cell* 108, 13–16.
- (13) Lewis, R., and Hill, C. (2020) Overcoming barriers to phage application in food and feed. *Curr. Opin. Biotechnol.* 61, 38–44.
- (14) Carattoli, A., Bertini, A., Villa, L., Falbo, V., Hopkins, K. L., and Threlfall, E. J. (2005) Identification of plasmids by PCR-based replicon typing. *J. Microbiol. Methods* 63, 219–228.
- (15) Coetzee, J. N., Bradley, D. E., Fleming, J., du Toit, L., Hughes, V. M., and Hedges, R. W. (1985) Phage pilH α : A phage which adsorbs to IncHI and IncHII plasmid-coded pili. *J. Gen. Microbiol.* 131, 1115–1121.

- (16) Coetzee, J. N., Bradley, D. E., Lecatsas, G., du Toit, L., and Hedges, R. W. (1985) Bacteriophage D: An IncD group plasmid-specific phage. *J. Gen. Microbiol.* *131*, 3375–3383.
- (17) Lanza, V. F., de Toro, M., Garcillán-Barcia, M. P., Mora, A., Blanco, J., Coque, T. M., and de la Cruz, F. (2014) Plasmid Flux in Escherichia coli ST131 Sublineages, Analyzed by Plasmid Constellation Network (PLACNET), a New Method for Plasmid Reconstruction from Whole Genome Sequences. *PLoS Genet.* *10*, e1004766.
- (18) Johnson, T. J., Danzeisen, J. L., Youmans, B., Case, K., Llop, K., Munoz-Aguayo, J., Flores-Figueroa, C., Aziz, M., Stoesser, N., Sokurenko, E., Price, L. B., and Johnson, J. R. (2016) Separate F-Type Plasmids Have Shaped the Evolution of the H30 Subclone of Escherichia coli Sequence Type 131. *mSphere* *1*, e00121–16.
- (19) Rozwandowicz, M., Brouwer, M. S., Fischer, J., Wagenaar, J. A., Gonzalez-Zorn, B., Guerra, B., Mevius, D. J., and Hordijk, J. (2018) Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *J. Antimicrob. Chemother.* *73*, 1121–1137.
- (20) Yehl, K., Lemire, S., Yang, A. C., Ando, H., Mimee, M., Torres, M. D. T., de la Fuente-Nunez, C., and Lu, T. K. (2019) Engineering Phage Host-Range and Suppressing Bacterial Resistance through Phage Tail Fiber Mutagenesis. *Cell* *179*, 459–469.e9.
- (21) Costa, T. R., Ilangovan, A., Ukleja, M., Redzej, A., Santini, J. J. M., Smith, T. K., Egelman, E. H., and Waksman, G. (2016) Structure of the Bacterial Sex F Pilus Reveals an Assembly of a Stoichiometric Protein-Phospholipid Complex. *Cell* *166*, 1436–1444.e10.
- (22) Sehna, D., Rose, A., Koca, J., Burley, S., and Velankar, S. Mol*: Towards a Common Library and Tools for Web Molecular Graphics. *Work. Mol. Graph. Vis. Anal. Mol. Data.* 2018.
- (23) Toropova, K., Basnak, G., Twarock, R., Stockley, P. G., and Ranson, N. A. (2008) The

- Three-dimensional Structure of Genomic RNA in Bacteriophage MS2: Implications for Assembly. *J. Mol. Biol.* 375, 824–836.
- (24) Dai, X., Li, Z., Lai, M., Shu, S., Du, Y., Zhou, Z. H., and Sun, R. (2017) In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* 541, 112–116.
- (25) Valegård, K., Liljas, L., Fridborg, K., and Unge, T. (1990) The three-dimensional structure of the bacterial virus MS2. *Nature* 345, 36–41.
- (26) Farafonov, V. S., and Nerukh, D. (2019) MS2 bacteriophage capsid studied using all-atom molecular dynamics. *Interface Focus* 9, 20180081.
- (27) Perkett, M. R., Mirijanian, D. T., and Hagan, M. F. (2016) The allosteric switching mechanism in bacteriophage MS2. *J. Chem. Phys.* 145, 035101.
- (28) Siebenmorgen, T., and Zacharias, M. (2019) Computational prediction of protein–protein binding affinities. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1–18.
- (29) Kastritis, P. L., and Bonvin, A. M. J. J. (2012) On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *J. R. Soc. Interface* 10, 20120835.
- (30) Häuser, R., Blasche, S., Dokland, T., Haggård-Ljungquist, E., von Brunn, A., Salas, M., Casjens, S., Molineux, I., and Uetz, P. (2012) Bacteriophage protein-protein interactions. *Adv. Virus Res.* 83, 219–298.
- (31) Lehne, B., and Schlitt, T. (2009) Protein-protein interaction databases: keeping up with growing interactomes. *Hum. Genomics* 3, 291–297.
- (32) Geng, C., Xue, L. C., Roel-Touris, J., and Bonvin, A. M. (2019) Finding the $\Delta\Delta G$ spot: Are predictors of binding affinity changes upon mutations in protein–protein interactions ready for it? *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 9, 1–14.

- (33) Dent, K. C., Thompson, R., Barker, A. M., Hiscox, J. A., Barr, J. N., Stockley, P. G., and Ranson, N. A. (2013) The asymmetric structure of an icosahedral virus bound to its receptor suggests a mechanism for genome release. *Structure* 21, 1225–1234.
- (34) Dominguez, C., Boelens, R., and Bonvin, A. M. (2003) HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* 125, 1731–1737.
- (35) van Zundert, G., Rodrigues, J., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., Melquiond, A., van Dijk, M., de Vries, S., and Bonvin, A. (2016) The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* 428, 720–725.
- (36) Chaudhury, S., Berrondo, M., Weitzner, B. D., Muthu, P., Bergman, H., and Gray, J. J. (2011) Benchmarking and Analysis of Protein Docking Performance in Rosetta v3.2. *PLoS One* 6, e22477.
- (37) Herbert, A. MaxCluster - A tool for Protein Structure Comparison and Clustering. <http://www.sbg.bio.ic.ac.uk/maxcluster/index.html>.
- (38) Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008) CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* 19, 1859–1865.
- (39) Lee, J. et al. (2016) CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* 12, 405–413.
- (40) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935.

- (41) Huang, J., and Mackerell, A. D. (2013) CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* *34*, 2135–2145.
- (42) Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* *26*, 1781–1802.
- (43) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* *98*, 10089–10092.
- (44) Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* *23*, 327–341.
- (45) Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994) Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* *101*, 4177–4189.
- (46) Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, B. R. (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* *103*, 4613–4621.
- (47) McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., Schwantes, C. R., Wang, L.-P., Lane, T. J., and Pande, V. S. (2015) MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* *109*, 1528–1532.
- (48) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph.* *14*, 33–38.
- (49) Kortemme, T., Kim, D. E., and Baker, D. (2004) Computational Alanine Scanning of Protein-Protein Interfaces. *Sci. Signal.* *2004*, pl2.

- (50) Kortemme, T., and Baker, D. (2002) A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 14116–14121.
- (51) De Hoon, M. J. L., Imoto, S., Nolan, J., and Miyano, S. (2004) Open source clustering software. *Bioinformatics* *20*, 1453–1454.
- (52) Meng, R., Jiang, M., Cui, Z., Chang, J. Y., Yang, K., Jakana, J., Yu, X., Wang, Z., Hu, B., and Zhang, J. (2019) Structural basis for the adsorption of a single-stranded RNA bacteriophage. *Nat. Commun.* *10*, 3130.