Characterisation of HilD, the Central Activator of *Salmonella* Pathogenicity, and the Mechanisms Regulating its Activity

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Joe Joiner
aus Carshalton, England,
das Vereinigtes Königreich

Tübingen
2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	28.11.2022
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Dr. Marcus Hartmann
2. Berichterstatter/-in:	Prof. Dr. Andrei Lupas

Abstract

Anti-virulence drugs present an attractive alternative to traditional antibiotics for the treatment of bacterial infections, such as those caused by the major food-borne pathogen *Salmonella enterica*. Expression of virulence factors essential for *Salmonella* infection is tightly controlled by a network of transcription regulators. The AraC-like transcription factor HilD is the main integration point of environmental signals into this regulatory network, and hence is an attractive target for novel anti-virulence drugs. This work investigates the different mechanisms by which HilD is regulated at the protein level to control the expression of virulence genes.

Long chain fatty acids (LCFAs) are highly abundant in varying concentrations throughout the intestine and are utilised as environmental cues by *Salmonella* to coordinate expression of virulence genes. LCFAs bind directly to HilD, in a comparable manner to that reported for the other AraC-like transcription factors ToxT and Rns. HilD can accommodate the binding of a wide range of LCFAs with a chain length of 16 – 24 C-atoms. The binding of LCFAs induces conformational changes in the structure of HilD that disrupt both its dimerisation and DNA-binding ability.

The regulatory protein HilE is structurally homologous to hemolysin-coregulated protein (Hcp), which is a structural component of the bacterial type VI secretion system. We found that unlike other Hcp family members, HilE exists as a monomer in solution. HilE forms a stable 1:1 complex with HilD. Using hydrogen-deuterium exchange mass-spectrometry (HDX-MS), we show that HilE directly disrupts HilD dimerisation by binding to the HilD dimerisation interface, thereby also preventing HilD from binding to target DNA. Our results highlight two distinct mechanisms by which HilD activity is repressed, which can be exploited for the development of new antivirulence leads.

We previously identified the compound C26 as a novel inhibitor of *Salmonella* virulence gene expression. Here it is shown that C26 specifically binds to HilD and consequently inhibits its ability to bind to target DNA. HDX-MS indicated that C26 binds to the same binding pocket of HilD as LCFAs, albeit likely with a different binding mode. Unlike LCFAs, C26 does not affect HilD dimerisation nor its repression by HilE. We performed a structure-activity relationship

analysis of C26 and have so far identified derivatives with 10-fold higher affinity for HilD. The C26 scaffold appears to be an attractive candidate for the development of new antivirulence compounds against *Salmonella* and our results provide a starting point for the optimisation of C26 into a lead compound.

Zusammenfassung

Bei der Behandlung bakterieller Infektionen, z.B. solchen, die durch das wichtigste Lebensmittelpathogen *Salmonella enterica* verursacht werden, repräsentieren antivirulente Medikamente eine attraktive Alternative zu herkömmlichen Antibiotika. Die Expression von für eine *Salmonella*-Infektion essentiellen Virulenzfaktoren wird durch ein Netzwerk von Transkriptionsregulatoren streng kontrolliert. In diesem Netzwerk ist der zu AraC ähnliche Transkriptionsfaktor HilD der hauptsächliche Verarbeitungspunkt eintreffender Umgebungssignale und stellt damit eine attraktive Angriffsmarke für neuartige antivirulente Wirkstoffe dar. In dieser Arbeit werden verschiedene Mechanismen untersucht, durch die HilD auf Proteinebene bei der Kontrolle der Expression von Virulenzgenen reguliert wird.

Im gesamten Darmsystem finden sich in verschiedenen Konzentrationen große Mengen langkettiger Fettsäuren, die für *Salmonella* der Umweltreiz für die Initiierung der Virulenzgen-expression sind. Die Fettsäuren binden direkt an HilD in ähnlicher Weise, wie sie für die AraC-ähnlichen Transkriptionsfaktoren ToxT und Rns beschrieben worden ist. HilD kann eine breite Auswahl an Fettsäuren mit einer Kettenlänge von 16 – 24 C-Atomen binden. Durch die Anlagerung von Fettsäuren an HilD werden konformationelle Änderungen in der HilD-Struktur induziert, durch die sowohl dessen Dimerisierung als auch DNA-Bindungsfähigkeit unterdrückt werden.

Das Regulatorprotein HilE ist strukturell homolog zu Haemolysin-coregulierten Protein (Hcp), einem Strukturbestandteil des bakteriellen Typ-VI-Sekretionssystems. Wir fanden, dass HilE im Gegensatz zu anderen Hcp-Proteinen in Lösung monomer vorliegt. Mit HilD bildet HilE einen stabilen 1:1 Komplex. Mit Wasserstoff-Deuterium-Austausch-Massenspektrometrie (HDX-MS) konnten wir zeigen, dass HilE direkt die HilD-Dimerisierung unterbindet, indem es an dessen Dimerisierungs-Schnittstelle bindet und dabei auch HilD von der Bindung an seine Ziel-DNA abhält. Diese Ergebnisse zeigen zwei unterschiedliche Mechanismen zur Hemmung der HilD-Aktivität auf, die für die Entwicklung neuartiger antivirulenter Leitstrukturen ausgenutzt werden können.

Bereits früher war von uns die Verbindung C26 als Inhibitor der Virulenzgenexpression in *Salmonella* gefunden worden. Hier wird gezeigt, dass C26 spezifisch an HilD bindet und HilD

dann keine DNA mehr anlagern kann. Mittels HDX-MS konnte bestätigt werden, dass C26 in dieselbe Bindungstasche von HilD bindet wie die Fettsäuren, wenngleich wahrscheinlich auch in einer unterschiedlichen Weise, denn durch die C26-Bindung wird weder die HilD-Dimerisierung noch die HilE-induzierte Repression unterdrückt. Wir haben eine Struktur-Aktivitäts-Beziehungsanalyse von C26 durchgeführt und bislang Derivate mit einer 10-fach höheren Affinität zu HilD gefunden. Das C26-Gerüst scheint ein attraktiver Kandidat für die Entwicklung neuer antivirulenter Wirkstoffe gegen *Salmonella* zu sein, und unsere Resultate sind eine gute Basis für die Optimierung von C26 zu einem Medikament.

Table of Contents

Abstract 5

Zusammenfassung 7

Abbreviations 11

Chapter 1. Introduction 12

1.1 Antivirulence Compounds 12

1.2 *Salmonella* Pathogenesis 12

1.3 *Salmonella* Pathogenicity Islands 14

1.4 Regulation of *Salmonella* Pathogenicity 15

1.5 DNA Binding Sites of HilD 18

1.6 Mechanism of HilD-mediated Regulation 19

1.7 Motivations and Project Aims 21

Chapter 2. Materials and Methods 23

2.1 Materials 23

2.2. Molecular Cloning 24

2.3 Protein Expression 27

2.4 Protein Purification 27

2.5 Biophysical Characterisation 31

2.6 Computational Methods 39

Chapter 3. Characterisation of HilD and HilC 40

3.1 The AraC/XylS Family of Proteins 40

3.2 Biophysical Characterisation of HilD and HilC 43

3.3 Structure Prediction of HilD 45

3.4 Attempted Structural Characterisation of HilD 48

3.5 Attempted Structural Characterisation of HilC 52

3.6 Oligomerisation of HilD and HilC 54

3.7 DNA Binding of HilD and HilC 56

Chapter 4. Long Chain Fatty Acids are Small-Molecule Regulators of HilD and HilC 58

4.1 Introduction 58

4.2 Oleic Acid Binding to HilD and HilC 61

4.3 Effect of LCFAs on HilD Dimerisation 63

4.4 Structure-Activity Relationship of LCFAs 64

4.5 Binding Mode of LCFAs to HilD 70

4.6 Discussion 75

Chapter 5. HilE is a Specific Protein Regulator of HilD 78
5.1 Introduction 78
5.2 HilE Does Not form Higher Order Oligomers 79
5.3 Structure Prediction of HilE 81
5.4 Biochemical and Biophysical Characterisation of HilE 84
5.5 HilE Inhibits HilD Dimerisation 85
5.6 HilE and LCFAs Regulate HilD via Mutually Exclusive Mechanisms 86
5.7 HilD and HilE Form a Stable SEC Complex 88
5.8 HilD and HilE Form a 1:1 Complex 89
5.9 Predicted Structure of the HilD-HilE Complex 91
5.10 HDX of the HilD-HilE Complex 93
5.11 Discussion 97
Chapter 6. C26 is a Novel Small-Molecule Inhibitor of HilD 99
6.1 Introduction 99
6.2 Target Validation of C26 101
6.3 Affinity of C26 to HilD 102
6.4 C26 Mechanism of Action 106
6.5 Binding Mode of C26 to HilD 108
6.6 Structure Activity Relationship of C26 116
6.7 Discussion 132
Chapter 7. Concluding Remarks 135
Contributions 137
Acknowledgements 139
References 140
Appendix 153

List of Abbreviations

ANR	AraC Negative Regulator
β-ME	β-mercaptoethanol
c2-HDA	*cis*-2-Hexadecenoic Acid
CD	Circular Dichroism
CoA	Coenzyme A
CryoEM	Cryogenic Electron Microscopy
Cy5	Cyanine-5
Da	Dalton
DBD	DNA-Binding Domain
DMSO	Dimethyl Sulfoxide
EDTA	Ethylenediaminetetraacetic Acid
ELISA	Enzyme-linked Immunosorbent Assay
EMSA	Electrophoretic Mobility Shift Assay
GFP	Green Fluorescent Protein
Hcp	Hemolysin-coregulated Protein
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic Acid
HDX	Hydrogen Deuterium Exchange
HSQC	Heteronuclear Single-quantum Correlation Spectroscopy
HTH	Helix-turn-helix
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
K_d	Dissociation Constant
kDa	Kilodalton
LB	Lysogeny Broth
LCFA	Long-chain Fatty Acid
MALS	Multi-angle Light Scattering
MS	Mass Spectrometry
MST	Microscale Thermophoresis
MWCO	Molecular Weight Cut-off
NanoDSF	Nanoscale Differential Scanning Fluorimetry
NMR	Nuclear Magnetic Resonance
NTD	N-terminal Domain
NTS	Non-typhoidal *Salmonella*
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
P_{hilA}	*hilA* Promoter
RNAP	RNA Polymerase
SAR	Structure-activity Relationship
SCV	*Salmonella*-containing Vacuole
SDS	Sodium Dodecyl Sulfate
SEC	Size Exclusion Chromatography
SPI	*Salmonella* Pathogenicity Island
SUMO	Small Ubiquitin-like Modifier
T1SS	Type One Secretion System
T3SS	Type Three Secretion System
T6SS	Type Six Secretion System
TCA	Tricarboxylic Acid
TCS	Two-component System
TEV	Tobacco Etch Virus
T_m	Melting Temperature
UFA	Unsaturated Fatty Acid
Ulp1	SUMO Protease

Chapter 1. Introduction

1.1 Antivirulence Compounds

The increasing prevalence of antibiotic resistance poses a serious threat to global public health and the treatment of common infectious diseases. Whilst there is an immediate need for new antibacterial agents, insufficient recent progress has been made in the development of novel and potent antibiotics[1–3]. There is also an urgent need to develop alternative therapies to treat illness caused by bacterial pathogens. Common approaches include targeting antibiotic resistance pathways (such as β-lactamase inhibitors) and phage therapy, however both may also be susceptible to the development of resistance[4,5].

One attractive alternative is the use of antivirulence compounds, which target virulence factors essential for bacterial pathogenesis rather than the survival of pathogens, without affecting bacterial growth and viability[6]. Approaches for developing antivirulence agents include inhibiting toxin production and virulence factor secretion, impeding bacterial adhesion to host cells, and disrupting biofilm formation[7]. Use of non-lethal antivirulence agents is thought to lessen selective pressures, delaying the development and spread of resistance. The development of antivirulence agents is of particular relevance for gastrointestinal pathogens. The indiscriminate effect of traditional broad-spectrum antibiotics may cause dysbiosis and lead to the emergence of antibiotic-resistant members of the intestinal microbiota, potentially resulting in endogenous infections[7,8]. The specific mechanisms targeted by antivirulence agents should lessen any undesirable effects on the gut microbiota. Here we sought to identify such compounds against *Salmonella enterica*.

1.2 *Salmonella* Pathogenesis

Salmonella spp. are enteric pathogens and one of the most common causes of food-borne illness. Typically acquired by an animal host through the ingestion of contaminated food or water, *Salmonella spp.* infect an estimated 100 million people worldwide per year[9]. Non-typhoidal *Salmonella* (NTS) infections cause intestinal inflammation and diarrhoea, and although the fatality rate remains low in developed countries, this can rise to 25% in sub-

Saharan Africa[10]. The prevalence of antibiotic resistance amongst NTS infections is increasing and considered a serious threat, with fluoroquinolone-resistant *Salmonella spp.* categorised as high-priority pathogens by the WHO[11,12].

After entering the small intestine, *Salmonella* transverse the intestinal mucous layer and evade being killed by innate immune defences, to gain access to the underlying epithelium. *Salmonella* preferentially enter microfold (M) cells, specialised epithelial cells that ingest antigens from the intestinal lumen through pinocytosis and deliver them to lymphocytes in the underlying submucosa, initiating an immune response. *Salmonella* can exploit this mechanism to cross the epithelial barrier, however invasion of non-phagocytic enterocytes can also occur by use of a number of virulence factors[13,14].

The first step of *Salmonella* pathogenesis consists of adhering to the surface of enterocytes. *Salmonella* then utilise a type III secretion system (T3SS), a needle-like injectosome that enables the direct injection of effector proteins from the bacterial cytoplasm into the host cell[15] (Figure 1.1). These effectors bring about changes in host cell actin cytoskeleton architecture, resulting in the formation of membrane ruffles and subsequent engulfment of *Salmonella* cells by endocytosis, mimicing their uptake by M cells[16]. Other changes induced by T3SS effectors include the disruption of tight junctions, connections between adjacent cells which seal the epithelial monolayer and make it impermeable to ions, water and other nutrients[17]. This provides direct access of *Salmonella* to the basolateral surface.

After internalisation, *Salmonella* cells are contained within large vesicles, called *Salmonella*-containing vacuoles (SCV), in which they can replicate. The SCV migrates to, and fuses with, the basolateral membrane, releasing the contained *Salmonella* cells into the submucosa. Once across the epithelium and in the submucosa, typhoidal *Salmonella* serovars can enter phagocytes and disseminate throughout the bloodstream leading to a life-threatening systemic infection[18]. In contrast, NTS infections result in severe intestinal inflammation. NTS serovars have evolved to utilize inflammation-derived metabolites, and the infection is limited to the intestine[19] .This localised inflammatory response, combined with disruption of the tight junctions results in the initiation of diarrhoea in the mammalian host.

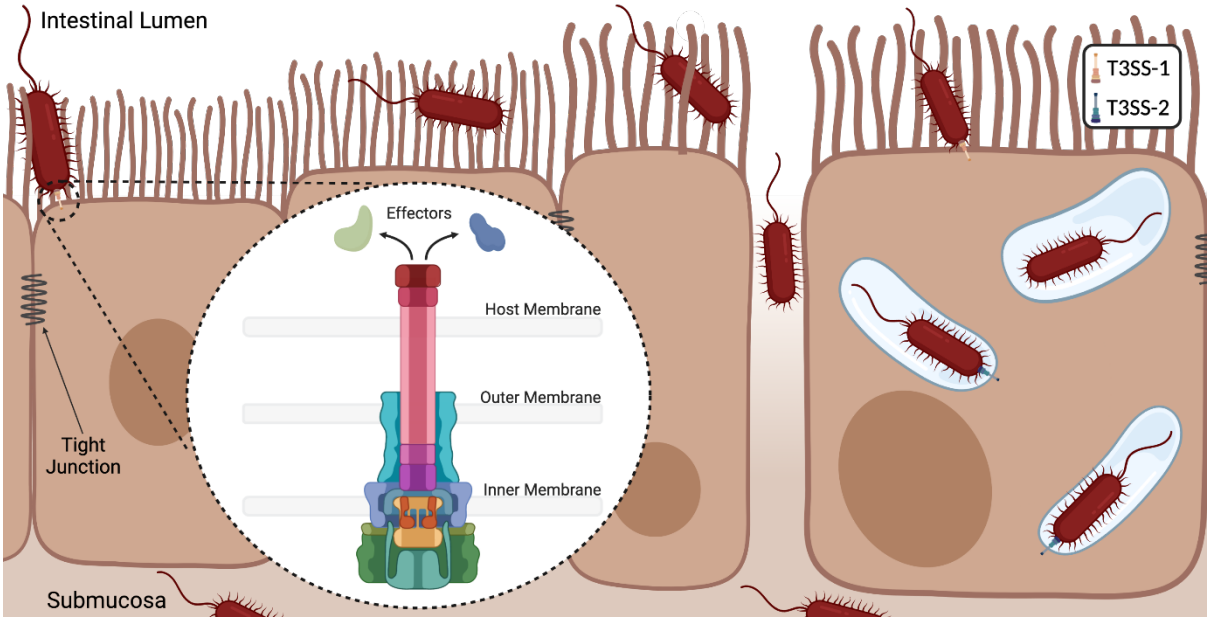


Figure 1.1. Simplified pathogenesis model of *Salmonella*. *Salmonellae* directly inject effector proteins into enterocytes of the intestinal epithelium via T3SS-1, which results in membrane ruffling, disruption of the epithelial brush border and *Salmonella* engulfment. Inside epithelial cells, *Salmonellae* are contained within SCV's. They survive within these compartments via the action of effector proteins excreted through T3SS-2, before gaining access to the submucosa.

1.3 *Salmonella* Pathogenicity Islands

The complex mechanism by which *Salmonellae* invade epithelial cells requires many virulence factors, most of which are located on 5 highly-conserved horizontally-acquired *Salmonella* pathogenicity islands (SPIs)[18,20,21]. *Salmonella* pathogenicity island 1 (SPI-1) is a 40-kb region encoding the genes required for the initial invasion of host cells. These include the numerous effector proteins that cause host cell morphological changes, and the structural components of the type III secretion system (T3SS-1) used to export these effectors into the host cell[22,23].

SPI-2 encodes genes crucial for *Salmonella* replication within phagosomes, enabling replication within the SCV and access to important nutrients for growth, without the need to enter the host-cell cytoplasm and hence trigger host-defence mechanisms. Amongst these are the genes encoding the structural components of another type III secretion system (T3SS-2), which is used to translocate effector proteins across the vacuolar membrane[23,24].

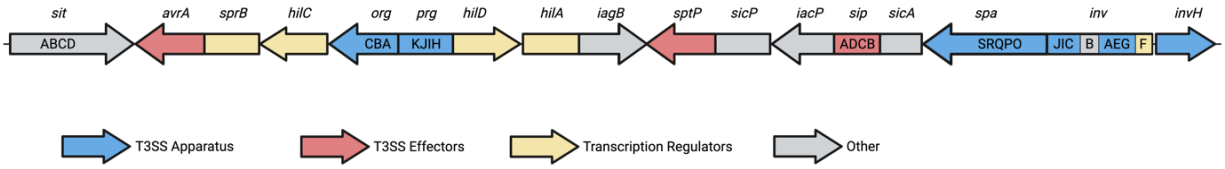


Figure 1.2. Schematic representation of SPI-1. Genes are coloured according to the function of the encoded products.

The other SPIs have not been studied so extensively. SPI-4 encodes the giant non-fimbrial adhesin SiiE, which mediates *Salmonella* adhesion to epithelial cells, and a type I secretion system (T1SS) through which SiiE is excreted[25]. SPI-3 and SPI-5 both contain multiple genes of unknown function, whilst those which have been characterised appear to have no functional relation to one another[18].

1.4 Regulation of *Salmonella* Pathogenicity

To coordinate the sequential expression of different virulence genes according to the stage of the infection process, expression of SPI genes is tightly regulated. This is crucial to ensure efficiency of pathogenesis and prevent uncontrolled expression of genes that may be otherwise detrimental to bacterial fitness. *Salmonellae* encode many genetic regulators to detect and respond to environmental stimuli. These regulators interact with one another, forming a complex regulatory network to active expression of the appropriate virulence factors encoded within the SPIs.

The OmpR/ToxR family transcriptional regulator HilA is the central regulator of SPI-1 and SPI-4, with a deletion of *hilA* shown to be phenotypically equivalent to deletion of the whole SPI-1 locus[26]. HilA directly activates the *prg/org* and *inv/spa* operons of SPI-1, which encode the structural components of T3SS-1. The latter operon includes the transcriptional regulator InvF, which then activates the expression of the *sic/sip* genes encoding several of the effector proteins secreted through T3SS-1[27–30]. HilA also activates the *sii* operon (SPI-4) required for bacterial cell adhesion, whilst repressing SPI-2 under invasion conditions, emphasising its critical role in coordinating gene expression during initial host-cell invasion[31–33].

Expression of *hilA* is in turn controlled by the action of three AraC-like transcription factors: HilD, HilC and RtsA. HilD and HilC are both located on SPI-1, whilst RtsA is encoded on a 15 kb *Salmonella*-specific insert in the *Salmonella* chromosome close to the $tRNA^{PheU}$ gene[34]. These three regulators are close homologs, with particularly high homology in their C-terminal DNA-binding domains (>56% identify and >72% similarity), and bind to overlapping sites within the *hilA* promoter to active expression[34–36]. Each of these three regulators is able to activate not only *hilA*, but also its own promoter and that of the other two regulators, forming a complex feed-forward loop to activate SPI-1 expression (Figure 1.3)[26]. All three regulators can additionally activate *invF* independently of HilA, and interact with small molecule regulators[37–39].

Of the regulatory triad, HilD is the most prominent activator of *hilA*, with HilC and RtsA serving to increase *hilA* transcription and enable *Salmonellae* to adapt invasion ability more rapidly in the changing environments of the gut[39–41]. HilD serves as the main integration point at which environmental signals are fed into the SPI-1 regulatory network, with many regulatory factors shown to affect *hilD* transcription, translation, or HilD activity. Other regulators have also been shown to act directly at the *hilA* promoter or affect all promoters of the feed-forward loop[41]. The whole network functions as a switch to turn on expression of SPI-1, which is triggered once HilD is expressed above a threshold to overcome the effect of negative regulators[42]. One of the most important negative regulators of HilD is HilE, a Hcp-like protein, which specifically represses HilD via a protein-protein interaction. Expression of *hilE* is in turn modulated by many other regulators, providing additional inputs into the control of HilD[43–45]. The protease Lon degrades HilD, moderating its intracellular concentration[46,47].

Salmonella possesses virulence-associated two-component systems (TCSs), enabling them to sense and respond to changing environmental conditions, such as those in the mammalian gut. TCSs generally consist of a membrane-bound histidine kinase that senses a specific environmental stimulus and catalyses phosphoryl transfer to a corresponding response regulator. This response regulator is most commonly a transcription factor, which once activated mediates the cellular response by controlling the expression of target (virulence) genes[48,49].

The TCSs EnvZ/OmpR and BarA/SirA both activate *hilA* expression through HilD. The EnvZ/OmpR TCS consists of the sensor protein EnvZ, which can respond to changes in osmolarity, and the transcription regulator OmpR. Once activated, OmpR induces *hilA* expression, reportedly through interactions with the HilD protein, although the exact mechanism of this is unclear[26,41,50]. The BarA/SirA TCS is responsible for the activation of SPI1 expression by short-chain fatty acids such as acetate and functions by controlling *hilD* translation[51,52]. SirA, the response regulator, activates the expression of the short non-coding RNAs CsrB and CsrC. CsrB/C sequester the RNA-binding protein CsrA, which otherwise binds to the *hilD* transcript, preventing translation[53,54]. TCSs also play an important role in cross-talk between pathogenicity islands; HilD/HilC/RtsA positively regulate SPI-2 through the TCS SsrA/SsrB[55].

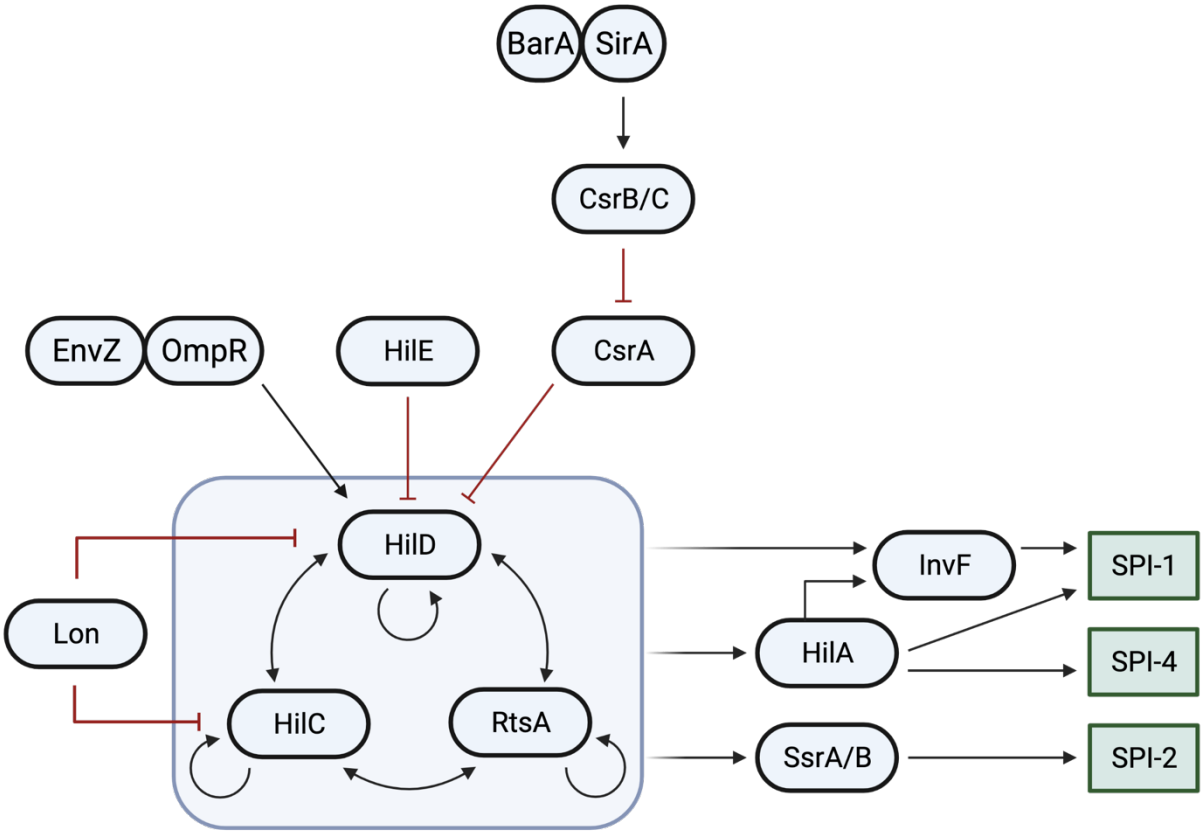


Figure 1.3. Simplified schematic of the regulatory network of SPI-1. Black arrows indicate activation and red lines with blunt ends represent repression.

In addition to the regulators discussed above, HilD is also able to bind a number of small molecules, which prevent it from activating target genes. The long-chain fatty acid (LCFA) oleic acid binds to both HilD and HilC, disrupting their DNA-binding activity[56]. A common

mechanism of repression by LCFAs has been demonstrated for a number of other closely-related AraC-like transcription factors[57–59]. Further studies have shown HilD is capable of binding to a range of different LCFAs, with a preference for binding *cis*-2 unsaturated LCFAs[60,61]. Additional ligands, such as the natural compound myricanol (and its derivatives) have also been observed to bind to HilD and inhibit its function *in vitro*[62].

1.5 DNA Binding Sites of HilD

HilD, HilC and RtsA activate expression by binding to specific sites within the promoter regions of their target genes. The three regulators bind to common sites within the *hilA*, *hilC*, *hilD* and *rtsA* promoters[35,63]. Whilst one common binding site is found in each of the *hilC*, *hilD* and *rtsA* promoters, two separate binding sites are found within the *hilA* promoter, termed A1 and A2. Each site supports the binding of two protein molecules, with HilD/HilC/RtsA able to form both homodimers and all possible combinations of heterodimers to bind to target DNA[64]. The five common binding sites show ~60% similarity, with a consensus binding sequence of CNATTNNT (where N is any nucleotide) repeated twice in each site[63]. The orientation of the consensus sequence varies between binding site, indicating the opposite orientation of dimers.

The formation of multiple retarded electrophoretic species was observed for the binding of HilC at the *hilC* promoter, indicating the presence of two additional binding sites not observed for HilD. Furthermore, RtsA bound to two additional binding sites in fragments of the *hilA* and *hilD* promoters, known to contain only one HilC/HilD site. HilC also binds to all common sites with higher affinity than HilD, showing that these regulators have similar but not identical binding specificities[35]. Although HilD, HilC and RtsA have common binding sites at these four promoters, this may not be the case for all target genes.

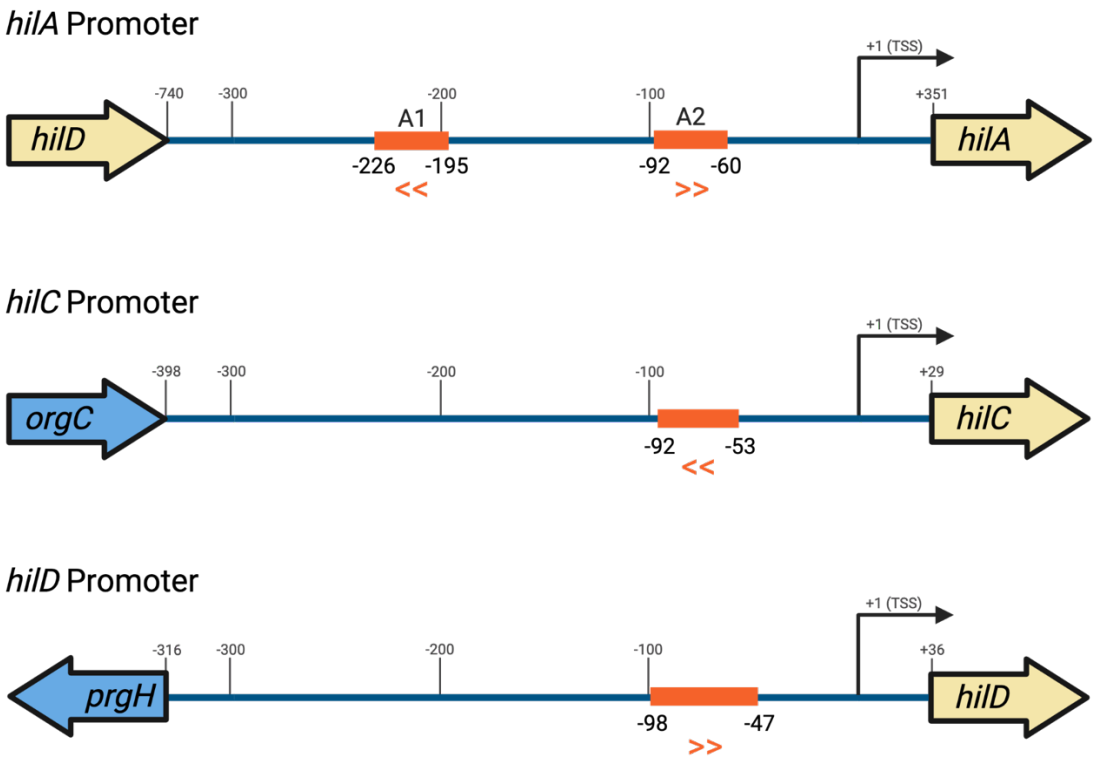


Figure 1.4. Location of the common HilD/HilC binding sites within the *hilA*, *hilC* and *hilD* promoters. Specific binding sites are shown in orange and defined relative to the transcription start site (TSS). The respective orientation of the consensus sequence is indicated by the directions of the arrows beneath each binding site. Genes are coloured as in Figure 1.2.

1.6 Mechanism of HilD-mediated Regulation

The nucleoid-associated protein H-NS is a global transcription regulator, which regulates gene expression by binding to AT-rich sequences, a common feature of horizontally-acquired genes. H-NS has therefore played an important role during the evolution of *Salmonella* pathogenicity by preventing uncontrolled expression of newly acquired DNA that could be otherwise detrimental to bacterial fitness[65–67]. H-NS binds to specific sequences within target promoters, preventing the binding of RNA polymerase (RNAP) and hence inhibiting the translation of target genes. Binding of H-NS can also cause changes in DNA topology. H-NS can form higher-order oligomerisation complexes, bridging together sections of DNA and RNAP at the promoter[68,69]. Relief of H-NS repression is achieved by the antagonistic binding of other proteins, so-called de-repressors. This allows expression of acquired virulence genes only under certain conditions, such as during invasion.

In *Salmonella*, H-NS has been shown to repress several genes within SPI-1 and SPI-2, including *hilA, hilC, hilD* and *rtsA*. At the *hilA*, *hilC* and *rtsA* promoters, another nucleoid-associated protein, Hha, also represses the expression of these genes. Whilst both H-NS and Hha can independently bind to target sequences, they likely also act together, forming a repressive complex to fine-tune repression of these promoters[63]. HilD/HilC/RtsA displace H-NS by binding to specific sites overlapping with regions protected by H-NS[36]. Release of H-NS and subsequent changes in DNA topology allows the binding of RNAP, and hence transcription of target genes[63,70–72]. HilD/HilC/RtsA have also been shown to directly interact with the C-terminal domain of the α-subunit (α-CTD) of RNAP, initiating expression of these genes[72–74]. In summary, HilD, HilC and RtsA act as both activators and de-repressors of target genes.

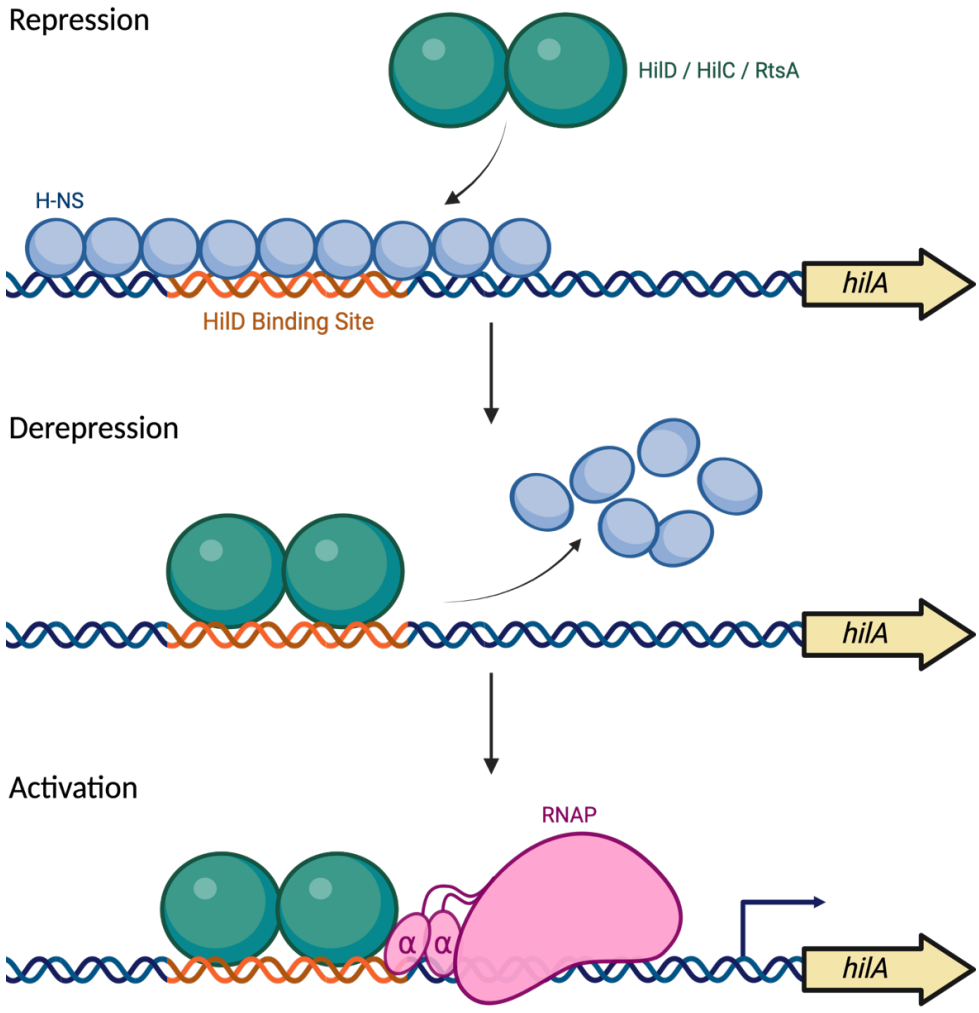


Figure 1.5. Mechanism of HilD activation of target genes. Under normal, non-inducing conditions, the *hilA* promoter is repressed by nucleoid-associated proteins, including H-NS. At conditions suitable for invasion, HilD displaces H-NS by binding to specific sites within the promoter sequence. HilD forms specific contacts with the α-CTD of RNA polymerase (RNAP) to activate *hilA* transcription.

1.7 Motivations and Project Aims

Considering the critical role T3SS-1 plays in *Salmonella* pathogenesis, it presents an attractive target for the development of new anti-virulence compounds. Novel inhibitors have previously been identified which target either the structural components of T3SS-1[75–78], or the regulators controlling its expression[62,79,80]. Small molecules inhibitors of the *Vibrio cholera* transcription regulator ToxT, a close homolog of HilD, were previously identified, showing both activity *in vitro* and reduced intestinal colonisation of *V.Cholerae* in mice[81–83]. Optimisation of higher-potency inhibitors was aided by knowledge of the structure of ToxT[57]. In the case of HilD (and the other regulators of SPI-1), there is currently no such structural information, limiting the development of similar inhibitors against *Salmonella* virulence.

In this thesis, I aimed to biochemically and structurally characterise the AraC-like transcription factors HilD, HilC and RtsA, which form the regulatory triad at the heart of the SPI-1 regulatory network, to provide a basis for the structure-based design of future antipathogenic compounds against *Salmonella*. Although many regulators and small molecules have been identified that act upon HilD at the protein level, there is a lack of information about the underlying functional mechanisms. Consequently, a number of related questions remain unanswered: What are the different mechanisms by which HilD/HilC/RtsA activity can be regulated? Can new inhibitors be designed which mimic such mechanisms? Can different negative regulators act together synergistically? Can these regulatory pathways be exploited as additional mechanisms for the design of new inhibitors? Here I chose to study two of these regulators, long-chain fatty acids (LCFAs) and the protein HilE, and uncover how they negatively regulate HilD activity.

As part of ongoing work to identify new anti-virulence agents against NTS, collaborators in the Wagner lab identified the small molecule Compound 26 (C26) as an inhibitor of effector protein secretion through T3SS-1, leading to a reduction of bacterial invasion of host cells. C26 leads to a downregulation of all SPIs, most likely targeting the regulatory triad of HilD/HilC/RtsA.

Here I performed an *in vitro* target validation of C26, identifying the protein targets of C26 and its specificity for its binding partners. Having identified the target of C26, I aimed to

characterise the interaction between HilD and C26. I sought to obtain a crystal structure of HilD (of particular interest was the HilD-C26 complex) to determine the binding mode of C26 and employed other approaches (NMR, CryoEM and hydrogen-deuterium exchange) when crystallisation proved ultimately unsuccessful. I also aimed to develop assays to determine the affinity of C26 to HilD and explore the effect(s) of C26 on HilD function.

Using these established assays, I next aimed to perform a structure-activity relationship of C26, screening newly synthesised analogs of C26 for their activity against HilD *in vitro*. In combination with *in vivo* assays performed by collaborating groups, we aimed to perform a lead optimisation of C26, identifying structural moieties critical for C26 binding and obtain derivative compounds with improved potency. The structure-activity relationship of C26 will aid the development of C26 into an early lead compound, as a novel anti-virulence agent against *Salmonella* infections.

Chapter 2. Materials and Methods

2.1 Materials

Compounds and chemicals that were purchased specifically for experiments in this study are listed in Table 2.1. C26, and all analogous compounds described in Chapter 6, were synthesised by the group of Mark Brönstrup (HZI Braunschweig). All other chemicals that were used throughout this study (for buffer solutions, crystallisation screening, culture media) were available in the department in which all experiments were carried out. Enzymes used during molecular cloning are listed, along with suppliers, in Table 2.2. DNA oligomers (primers for cloning and DNA fragments for EMSAs) were purchased from Merck.

Table 2.1. List of chemicals used throughout this study.

Chemical	Supplier	Lot Number
BODIPY™FL C16	Invitrogen™	-
cis-2-eicosenoic Acid	Cayman Chemicals	154574-19
cis-2-hexadecenoic Acid	Cayman Chemicals	0554655-6
Elaidic Acid	Cayman Chemicals	0471926-24
Erucic Acid	Cayman Chemicals	0509247-14
Gadoleic Acid	Larodan-Biozol	LX-503
Methyl-oleate	Sigma Aldrich	MKCK7423
Methyl-palmitoleate	Sigma Aldrich	SLCF0841
Myristoleic Acid	Sigma Aldrich	SLCD3021
Nervonic Acid	Cayman Chemicals	0561706-6
Oleic Acid	Sigma Aldrich	0000051240
Palmitic Acid	Sigma Aldrich	SLCD8134
Palmitoleic Acid	Fluka	1157581
SYBR™ Gold Nucleic Acid Stain	Invitrogen™	-
Virstatin	Cayman Chemicals	0498975-3

Table 2.2. List of enzymes used throughout this study.

Enzyme	Supplier
FastDigest BamHI	Thermo
FastDigest DpnI	Thermo
FastDigest EheI	Thermo
T4 Polynucleotide Kinase (T4 PNK)	Thermo
T4 DNA Ligase	Thermo
Q5 High-Fidelity DNA Polymerase	New England Biolabs

2.2. Molecular Cloning

Inserts encoding translational fusions of *hilD*, *hilC*, *rtsA* and *hilE* to the small ubiquitin-related modifier (SUMO) were contained within expression vectors of the pET series. The SUMO-HilD construct, cloned into the pET-24a(+) vector using the NdeI (CATATG) and NotI (GCGGCCGC) restriction sites, was a gift from Samuel Wagner (Universitätsklinikum Tübingen). The *hilC*, *rtsA* and *hilE* SUMO fusion genes were synthesised by Synbio Technologies, inserted into the pET-21a(+) vector via the NdeI and NotI restriction sites.

Expression vectors encoding SUMO protease (Ulp1) and TEV protease, pET28b and pET24, respectively, were available in the department stock collection. Both protease constructs contained an N-terminal His_6-tag, for ease of removal during purification of SUMO fusion proteins.

N-terminal GFP fusion proteins were cloned for each of HilD, HilC and RtsA for use in MST assays. The GFP gene was amplified from a plasmid available in the department stock collection, encoding His-TEV-GFP. The primers GFP_fwd and GFP_rev_BamHI were used for amplification and to introduce BamHI (GGATCC) and EheI (GGCGCC) restriction sites for ligation. The GFP insert was ligated into the expression vectors encoding the SUMO fusion proteins, into which the corresponding restriction sites were also introduced using the primers listed in Table 2.3. In the case of HilD, a TEV protease site was also introduced

Table 2.3. Primers used for the cloning of GFP-fusions of HilD, HilC and RtsA. The TEV protease cleavage site is highlighted in red and introduced restriction sites highlighted in blue.

Primer Name	Sequence
GFP_fwd	GAGAGA**GGCGCC**GGCAAAGTGAGCAAGGGCGAGG
GFP_rev_BamHI	GAGAGA**GGATCC**CTTGTACAGCTCGTCCATGCCGAG
HilD_TEV_fwd	GAGAGA**GGATCC**ACTACT**GAGAATCTTTATTTTCAGGGC**TCATCA ATGGAAAATGTAACCTTTGTAAGTAATAGTCATCAGCG
GFP_HilC_fwd	GAGAGA**GGATCC**GGCGGC GTATTGCCTTCAATGAATAAATCAGTTGAGGCCATTAGC
GFP_RtsA_fwd	GAGAGA**GGATCC**GGCGGC CTAAAAGTATTTAATCCCTCACCTGTCCAGGTGGGGA
SUMO_rev	ACCACCAATCTGTTCTCTGTGAGCC
GFP_HilD_noTEV_fwd	GGCTCATCAATGGAAAATGTAACCTTTGTAAGTAATAGTCATCA
GFP_HilD_noTEV_rev	TCCCTTGTACAGCTCGTCCATGC

between the GFP and HilD genes, although this was subsequently removed by Round-the-Horn PCR[84] using the primers GFP_HilD_noTEV_fwd and GFP_HilD_noTEV_rev.

The His_6-tagged HilD N-terminal domain construct was cloned using Round-the-Horn PCR, using the primer pairs HilD_NTD_fwd/rev and His-NTD_fwd/rev to remove the DNA-binding domain and SUMO fusion tag, respectively. Round-the-Horn PCR was also used to produce all other constructs used throughout this thesis, including the truncated HilD constructs lacking residues T156-I172, and single point mutants (R30A, E102A, N260A, K264A and R267A). The SUMO tag was removed from the HilC and HilE constructs, and an N-terminal TEV cleavage site instead introduced between the His_6-tag and protein N-terminus. Constructs of the HilC N-terminal domain were cloned for use in crystallisation trials, as were different HilE constructs for MST/ SEC-MALS runs. All primers used for Round-the-Horn PCR are listed in Table 2.4.

PCR reactions were performed using Q5 High-Fidelity DNA Polymerase, according to the manufacturer's instructions. Following the PCR reaction, template DNA was digested using FastDigest DpnI and the 5' end of the amplified PCR product was phosphorylated using FastDigest T4 PNK. DNA was ligated by incubation with T4 DNA Ligase for 2 hours at room temperature. The ligated DNA was used to transform electrocompetent *E. coli* TOP10 cells.

Identity of all mutants was confirmed by Sanger sequencing using T7 promoter and terminator primers. Sequencing was performed either *in house* or by Genewiz/ Azenta. For GFP fusion proteins, an additional primer was used for sequencing to check the validity of the central region of the gene, due to the increased sequence length of these constructs.

Table 2.4. Primers used for Round-the-Horn PCR. The TEV protease cleavage site is highlighted in red.

Primer	Sequence
HilD_NTD_fwd	TAAAAGCTTGCGGCCGCACTCGAG
HilD_NTD_rev	CGTTATCTGAGCCGAGCTAAGGATGATC
His-NTD_fwd	GTAAGTAATAGTCATCAGCGTCCTGC
His-NTD_rev	CGAACCATGGTGATGATGGTGATGA
loop_truncated_fwd1	GGC TGTATTGAAAATGAAGAGTTAATTCCTTATTTTCTGC
loop_truncated_fwd2	GGCGGC TGTATTGAAAATGAAGAGTTAATTCCTTATTTTCTGC
loop_truncated_fwd3	GGCTCAGGC TGTATTGAAAATGAAGAGTTAATTCCTTATTTTCTGC
loop_truncated_rev	CGTACAGGAGAACGCCGTTTTC
R30A_fwd	GCACAGCAAATTAAAAGTCAGACTCAGCAG
R30A_rev	GGTATTTGTCAAAAGTGATTTTAATTTCTGTAAG
E102A_fwd	GCA ATACCGACGCAACGACTTG
E102A_rev	GAGTATATCGAAATCCATGTGGCCA
N260A_fwd	CGGCAAGAATGGCACAGGCAGCAAAAC
N260A_rev	ATAAGTAGATGTCGCTAAAGCTGGTACCTTCTTC
K264A_fwd	GCACTTTTACGCATAGGCAACCATAATGTTAATGCTG
K264A_rev	TGCTGCCTGATTCATTCTTGCCGATAAG
R267A_fwd	GCAAAACTTTTAGCAATAGGCAACCATAATG
R267A_rev	TGCCTGATTCATTCTTGCCGATAAGTAGA
His-TEV-HilC_fwd	**GAGAATCTTTATTTTCAGGGC**TCATCA GTATTGCCTTCAATGAATAAATCAGTTGAGGCCATTAG
His-TEV-HilC_rev	CGAACCATGGTGATGATGGTGATGCAT
HilC_NTD_fwd	AAATGAGCGGCCGCACTC
HilC_NTD_rev1	TATTGCTCGCTCAAGGAAATCAAACCCA
HilC_NTD_rev2	TGTCGTAATTTTTATTGCTCGCTCAAGGAAATCAAAC
His-HilE_fwd	GGTGGTGACGCCATCTATTTAAAACTGGA
His-TEV-HilE_fwd	**GAGAATCTTTATTTTCAGGGC**TCATCAGGTGGTGACGCCATCTATTTAAAACTGGA
His-HilE_rev	CGAACCATGGTGATGATGGTGATGCAT
HilE-His_fwd1	GCGGCCGCACTCGAG
HilE-His_rev	TCGCCACAGCGCCTGT
HilE-His_fwd2	GACGCCATCTATTTAAAACTGGA
HilE-His_rev2	CATATGTATATCTCCTTCTTAAAGTTAAACAAA

Table 2.5. Primers used for Sanger sequencing to validate construct identity.

Primer	Sequence
T7_promoter	TAATACGACTCACTATAGGG
T7_terminator	GCTAGTTATTGCTCAGCGG
GFP_seq	CCCAGTCCGCCCTGAGCAA

2.3 Protein Expression

Electrocompetent *E. coli* cells were transformed with the appropriate plasmid, and grown on LB agar plates (supplemented with the appropriate antibiotic) at 37°C. The *E. coli* strains used for expression for each construct, along with the antibiotic resistance genes encoded on the corresponding plasmids, are listed in Table 2.6.

Selected colonies were added to LB medium (50 ml) and cultured overnight at 30°C, 170 rpm. LB media was inoculated 1:100 with incubated overnight culture solution. Cultures were incubated with shaking (37°C, 170 rpm) until an OD_{600nm} within the range 0.6-0.8 was reached (approximately 3 hours), at which point expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were incubated with shaking overnight (~18 hours, 25°C, 150 rpm). Cells were harvested by centrifugation (10 min, 11800 g, 4°C). Supernatant was discarded, and cell pellet stored at -80°C.

Table 2.6. Conditions for proteins expression of constructs used in this study.

Protein	Expression Strain	Antibiotic Resistance	IPTG concentration for induction (mM)
HilD	C41(DE3)	Kan	0.5
HilD_NTD	C41(DE3)	Kan	1.0
GFP-HilD	C41(DE3)	Kan	0.5
HilC	C41(DE3)	Amp	0.5
HilE	LEMO(DE3)	Amp	0.5
SUMO protease (Ulp1)	BL21(DE3)pLysS	Kan	0.5
TEV protease	BL21(DE3)	Amp	1.0

2.4 Protein Purification

Cell pellets were thawed, and cells re-suspended in a lysis buffer (Buffer A), supplemented with DNase and one cOmplete™ EDTA-free protease inhibitor cocktail tablet (Roche #11 873 580 001). Addition of the protease inhibitor tablet was omitted for the purification of Ulp1 and TEV protease. Cells were lysed using a French press (2x, 1000 Psi) and centrifuged (95000 g, 1 hr, 4°C). The resulting supernatant was filtered (0.40 µm) prior to loading to the appropriate chromatography column. An Äkta purification system (Cytiva) was used to perform all purification steps.

Proteins were concentrated using Amicon® Ultra Centrifugal Filters with the appropriate molecular weight cut-off (MWCO). The purity of purified proteins was assessed by SDS-PAGE, with samples loaded to a 15% SDS gel and stained with Coomassie. Protein concentration was determined from UV absorbance at 280 nm, measured using a NanoPhotometer® NP80 (IMPLEN).

2.4.1 HilD

The filtered cell supernatant was loaded to a Ni His-Trap column, equilibrated with Buffer A (50 mM $NaPO_4$, pH 7.0, 300 mM NaCl, 10 mM imidazole). Column was washed with 20% Buffer B (50 mM $NaPO_4$, pH 7.0, 300 mM NaCl, 250 mM imidazole), and the His-SUMO-tagged protein eluted with an increasing gradient of Buffer B. Fractions containing desired protein were combined, supplemented with Ulp1 (600 µl, 0.4 mg ml^{-1}) to cleave the His_6-SUMO tag, and dialyzed overnight at room temperature against 2 L Buffer A.

The dialysed protein was reapplied to the Ni His-Trap column, equilibrated with Buffer A. Column was washed with 25% Buffer B to elute cleaved HilD, and subsequently with 100% Buffer B to elute the cleaved His_6-SUMO tag, Ulp1 and any remaining uncleaved SUMO-HilD.

Flow through and column wash fractions containing the desired protein were combined, concentrated (MWCO: 30 kDa) and loaded to a gel filtration column (Superdex^{TM} 75 26/60), equilibrated with SEC Buffer (50 mM $NaPO_4$, pH 7.0, 200 mM NaCl). Eluted fractions containing purified HilD were concentrated (MWCO: 30 kDa) and stored in aliquots at -80°C.

2.4.2 HilD NTD (His-$HilD_{7-206}$)

The filtered cell supernatant was loaded to a Ni His-Trap column, equilibrated with Buffer A (20 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole). Column was washed with 20% Buffer B (20 mM Tris, pH 8.0, 300 mM NaCl, 250 mM imidazole), and the His-tagged protein eluted with an increasing gradient of Buffer B.

Fractions containing desired protein were combined, concentrated (MWCO: 10 kDa) and loaded to a gel filtration column (Superdex^{TM} 75 26/60), equilibrated with SEC Buffer (20

mM Tris, pH 8.0, 150 mM NaCl). Eluted fractions containing purified protein were concentrated (MWCO: 10 kDa) and stored in aliquots at -80°C.

2.4.3 GFP-HilD

The filtered cell supernatant was loaded to a Ni His-Trap column, equilibrated with Buffer A (20 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole). Column was washed with 20% Buffer B (20 mM Tris, pH 8.0, 300 mM NaCl, 250 mM imidazole), and the His-SUMO-tagged protein eluted with an increasing gradient of Buffer B. Fractions containing desired protein were combined, supplemented with Ulp1 (600 µl, 0.4 mg ml^{-1}) to cleave the His_6-SUMO tag, and dialyzed overnight at room temperature against 2 L Buffer A.

The dialysed protein was reapplied to the Ni His-Trap column, equilibrated with Buffer A. Column was washed with 25% Buffer B to elute cleaved GFP-HilD. Flow through and column wash fractions containing the desired protein were combined, concentrated (MWCO: 50 kDa) and loaded to a gel filtration column (Superdex^{TM} 200 16/60), equilibrated with SEC Buffer (20 mM Tris, pH 8.0, 150 mM NaCl). Eluted fractions containing purified GFP-HilD were concentrated (MWCO: 50 kDa) to a final concentration of 10-15 µM and stored in aliquots at -80°C.

2.4.4 HilC

The filtered cell supernatant was loaded to a Ni His-Trap column, equilibrated with Buffer A (50 mM $NaPO_4$, pH 7.0, 500 mM NaCl, 10 mM imidazole). Column was washed with 20% Buffer B (50 mM $NaPO_4$, pH 7.0, 500 mM NaCl, 250 mM imidazole), and the His-tagged protein eluted with an increasing gradient of Buffer B. Fractions containing desired protein were combined, supplemented with TEV protease (1 mg) and dialyzed overnight at 6°C against 2 L Buffer C (50 mM $NaPO_4$, pH 7.0, 400 mM NaCl), using a dialysis membrane with a MWCO of 12 kDa.

The dialysed protein was reapplied to the Ni His-Trap column, equilibrated with Buffer A. The column was washed with 15% Buffer B to elute cleaved HilC.

Flow through and wash fractions containing purified HilC were combined and dialysed twice against 2L storage buffer (50 mM $NaPO_4$, pH 7.0, 200 mM NaCl) at 6°C. Dialysed protein was concentrated (MWCO: 30 kDa) and stored in aliquots at -80°C.

2.4.5 HilE

The filtered cell supernatant was loaded to a Ni His-Trap column, equilibrated with Buffer A (20 mM Tris, pH 8.0, 300 mM NaCl). Column was washed with 20% Buffer B (20 mM Tris, pH 8.0, 300 mM NaCl, 250 mM imidazole), and the His-tagged protein eluted with an increasing gradient of Buffer B. Fractions containing desired protein were combined, supplemented with 1 mg TEV-protease, and dialyzed overnight at 6°C against 2 L Buffer A (20 mM Tris, pH 8.0, 300 mM NaCl).

The dialysed protein was reapplied to the Ni His-Trap column, equilibrated with Buffer A. Column was washed with 15% Buffer B to elute cleaved HilE. Flow through and column wash fractions containing purified HilE were combined and dialysed twice against 2 L storage buffer (20 mM Tris, pH 8.0, 150 mM NaCl) at 6°C. The dialysed protein was concentrated (MWCO: 10 kDa) and stored in aliquots at -80°C.

2.4.6 SUMO Protease (Ulp1)

The filtered cell supernatant was loaded to a Ni His-Trap column, equilibrated with Buffer A (40 mM HEPES, pH 7.5, 150 mM KCl, 20 mM β-ME). Column was washed with 20% Buffer B (40 mM HEPES, pH 7.5, 150 mM KCl, 20 mM β-ME, 250 mM imidazole), and the His-tagged protein eluted with an increasing gradient of Buffer B.

Fractions containing pure Ulp1 were combined and dialysed overnight against storage buffer (40 mM HEPES, pH 7.5, 100 mM KCl, 10 mM β-ME) at 6°C. The dialysed protein was concentrated (MWCO: 10 kDa) to 1 mg ml^{-1}, diluted 1:1 with 86% glycerol (final concentration: 0.5 mg ml^{-1} Ulp1 in 43% glycerol) and stored at -20°C.

2.4.7 TEV Protease

The filtered cell supernatant was loaded to a Ni His-Trap column, equilibrated with Buffer A (20 mM Tris, pH 8.0, 300 mM NaCl, 0.5 mM EDTA, 1 mM β-ME, 20 mM imidazole). Column

was washed with 20% Buffer B (20 mM Tris, pH 8.0, 300 mM NaCl, 0.5 mM EDTA, 1 mM β-ME, 250 mM imidazole), and the His-tagged protein eluted with an increasing gradient of Buffer B.

Fractions containing pure TEV protease were combined and dialysed overnight against storage buffer (20 mM Tris, pH 8.0, 50 mM NaCl, 0.5 mM EDTA, 1 mM β-ME) at 6°C. The dialysed protein was concentrated (MWCO: 10 kDa) to 12 mg ml^{-1}, diluted 1:1 with 86% glycerol (final concentration: 6 mg ml^{-1} TEV protease in 43% glycerol) and stored in aliquots at -80°C.

2.5 Biophysical Characterisation

2.5.1 Circular Dichroism

Far-UV circular dichroism (CD) spectra were acquired on a JASCO J-810 spectropolarimeter with a Peltier type temperature control system, using 10 µM of purified protein in CD buffer (50 mM $NaPO_4$ pH 7.0, 100 mM NaCl). CD scans were recorded at 20°C, with five whole-spectra accumulations over a wavelength range of 250-190 nm at 100 nm min^{-1} and a data pitch of 0.1 nm. CD melting curves were collected by measuring ellipticity (Θ) at 222 nm over a temperature range of 20-95°C with a temperature ramp of 1°C min^{-1}.

2.5.2 SEC-MALS

SEC-MALS experiments were performed using Superdex^{TM} 75 Increase 10/300 GL column (Cytiva) coupled to a miniDAWN Tristar Laser photometer (Wyatt) and a RI-2031 differential refractometer (JASCO). 50 µl samples of 100 µM protein (50 µM in the case of HilC) were loaded to the SEC column, equilibrated with SEC buffer (50 mM $NaPO_4$ pH 7.0, 200 mM NaCl), and separated using a flow rate of 0.5 ml min^{-1}. Data analysis was carried out with ASTRA v7.3.0.18 software (Wyatt).

2.5.3 NanoDSF

Thermal stability of proteins was also determined using nanoscale differential scanning fluorimetry (NanoDSF). This technique relies on intrinsic protein fluorescence upon excitation, monitored at emission wavelengths 330 and 350nm. Changes in the chemical environment of tryptophan/tyrosine residues upon protein unfolding are reflected in changes in fluorescence. Samples were loaded to standard capillaries (#PR-C002) and runs performed on a Prometheus NT.48 (NanoTemper Technologies).

For protein stability measurements, protein samples were heated from 20-95°C, with a temperature gradient of 0.5°C min^{-1}. Melting temperatures were calculated from changes in the fluorescence ratio (350/330 nm), using PR.Stability Analysis v1.0.3 software (NanoTemper Technologies). For initial screening of binding compounds, 20 µM HilD was incubated with 50 µM of the respective compound for 20 min prior to loading of capillaries. An increase in the calculated melting temperature averaged from five replicate experiments was interpreted as ligand binding.

Binding Affinity (NanoDSF)

For each ligand, a two-fold serial dilution series was prepared in DMSO (12 concentrations, 100 µM highest assay concentration). 5 µM HilD or HilC were incubated with varying concentrations of ligands in assay buffer (50 mM $NaPO_4$ pH 7.0, 200 mM NaCl), and a final DMSO concentration of 1%. Samples were incubated for > 20 min at room temperature and centrifuged for 2 min prior to loading to standard capillaries. An excitation power of 100% was used to obtain fluorescence counts above 3,000 RFU for 330 and 350 nm, and samples heated from 20-80°C with a temperature gradient of 0.5°C min^{-1}. Due to ligand fluorescence, melting temperatures were calculated from changes in fluorescence at 350 nm, using PR.Stability Analysis v1.0.3 and a temperature range of 35-70°C for curve fitting. Data analysis, to determine binding affinities of compounds, was performed using Prism 8.4 (GraphPad). The change in HilD melting temperature, T_m, was fitted as a function of ligand concentration, using equations (1) and (2) to yield apparent affinity ($K_{d,app}$) values.

$$T_m([L]_0) = T_{m,lower} + \left(T_{m,upper} - T_{m,lower}\right) * (1 - \alpha([L]_0)) \quad (1)$$

$$\alpha([L]_0) = \frac{[P]_t - K_d - [L]_t + \sqrt{\left([P]_t + [L]_t + K_{d,app}\right)^2 - (4[P]_t[L]_t)}}{2[P]_t} \quad (2)$$

where $[P]_t$ and $[L]_t$ are the total protein and ligand concentrations, respectively.

For higher affinity compounds, dissociation constants were determined at specific temperatures using an isothermal approach. The fraction unfolded (f_u) of HilD was calculated at specific temperatures close to the melting temperature using the FoldAffinity software[85]. The change in heat capacity (ΔCp) upon unfolding was assumed to be zero. The fraction unfolded was then fitted as a function of ligand concentration using equation (3), implemented in Prism 8.4, to obtain accurate K_d values[86].

$$f_u = \frac{1}{1 + \left(\left(\frac{1}{K_u}\right) * \left(1 + \frac{[L]}{K_d}\right)\right)} \quad (3)$$

where K_d and K_u are the respective equilibrium constants for the ligand binding and protein unfolding reactions. [L], the unbound ligand concentration, is defined as:

$$[L] = \frac{1}{2}\left(\left([L]_t - [P]_t - K_d(1 + K_u)\right) + \sqrt{\left([P]_t - [L]_t + K_d(1 + K_u)\right)^2 + 4[L]_t K_d(1 + K_u)}\right)$$

2.5.4 EMSA

EMSAs were performed using a 62 base pair dsDNA fragment of the *hilA* promoter, encompassing the A1 binding site[36] (A1_62bp). Double stranded DNA fragments were generated by boiling complementary primers together at 95°C for 10 min before annealing, by slowly cooling to room temperature. The forward primer was modified with a 5'-Cy5 fluorescent dye for detection. 600 nM of protein was incubated with 50 nM of labelled DNA in EMSA buffer (20 mM Tris, pH 8.0, 100 mM KCl, 100 µM EDTA, 3% glycerol). Ligands were diluted in DMSO or ethanol and diluted 1:100 into the protein-DNA sample. For specificity

assays, an 80-fold excess of unlabelled specific DNA (A1_62bp) or non-specific DNA (*hilA*_61bp) was added. Samples were incubated at 37°C for 15 min, supplemented with diluted DNA loading dye, and separated on a 1.5 mm thick, 6% TBE gel at 6°C at a constant voltage of 100 V. Gels were imaged using a ChemiDoc™MP imaging system (Bio-Rad Inc). DNA primers for EMSA and MST measurements, including those already modified with a 5'-Cy5 fluorescent dye, were purchased from Merck.

Table 2.7. DNA fragments used for EMSAs. The A1 binding site is highlighted in blue.

DNA Fragment	Sequence
A1_62bp_fwd	GGGAGTAAAGAAAAG**ACGATATCATTATTTTGCAAAAAAATATAAA**AATAAGCGCACCATTA
A1_62bp_rev	TAATGGTGCGCTTATT**TTTATATTTTTTTGCAAAATAATGATATCGT**CTTTTCTTTACTCCC
*hilA*_61bp_fwd	cacagttagttataacaatattattaccaacatgtcagttatttaaagcacaggcataagc
*hilA*_61bp_rev	gcttatgcctgtgctttaaataactgacatgttggtaataatattgttataactaactgtg

2.5.5 MST

All MST measurements were performed on a NanoTemper Monolith NT.115 with a Nano BLUE/RED Detector using MO.Control v1.6. MST runs were performed at 25°C, with an excitation power of 20% and MST power set to medium, unless otherwise stated. Data were analysed using the MO.Affinity Analysis v2.3 software. K_d affinity constants for the direct binding of ligands to GFP-HilD were calculated using the K_d model, and for competitive assays data was fitted using the Hill model to determine IC_{50} values. MST on-times of 1.5 or 2.5 seconds were used for data analysis.

GFP-HilD Binding Assay

MST experiments were performed with 50 nM GFP-HilD and increasing concentrations of the respective ligand in Tris MST assay buffer (20 mM Tris, pH 8.0, 150 mM NaCl, 0.1% Pluronic). For HilD dimerisation assays, an alternative $NaPO_4$ MST assay buffer (50 mM $NaPO_4$, pH 7.0, 200 mM NaCl, 0.1% Pluronic) was used.

For HilD dimerisation and HilE binding, a two-fold serial dilution of proteins was performed in corresponding protein storage buffer. Dilution series of small-molecule ligands were prepared in respective organic buffer (DMSO for C26 compounds, ethanol for fatty acids), and subsequently diluted 1:50 into assay buffer. All ligand samples were then mixed 1:1 with 100

nM GFP-HilD (diluted in assay buffer), incubated together for 10 min at room temperature, centrifuged for 5 min and loaded to standard capillaries (Nanotemper #MO-K022). An excitation power of 40% (lipids, compounds) or 60% (HilE, dimerisation) was used for MST runs in which GFP-HilD was the labelled target. To investigate competitive effects of different ligands, GFP-HilD was pre-incubated with the competing ligand for ≥ 10 min at room temperature, before mixing with the second ligand (of varying concentration).

Competitive MST Assay: BODIPY FL C_{16}

To determine the affinity of the BODIPY FL C_{16} reporter ligand, a two-fold serial dilution of HilD was performed using storage buffer (50 mM $NaPO_4$, pH 7.0, 200 mM NaCl). BODIPY FL C_{16} was diluted first to 20 µM in DMSO, and subsequently 1:50 into $NaPO_4$ MST assay buffer, before mixing 1:1 with HilD resulting in final assay concentrations of 200 nM BODIPY FL C_{16} and 1% DMSO. Samples were incubated for 5 min at room temperature and centrifuged for 3 min before loading to standard capillaries. Affinity of the reporter to HilD was determined from changes in the initial fluorescence.

For competitive runs, HilD (100 µM) was incubated with BODIPY FL C_{16} (400 nM) in $NaPO_4$ MST assay buffer (2% DMSO) for 10 min at room temperature. Ligands were titrated in DMSO, diluted 1:50 into $NaPO_4$ MST assay buffer, mixed 1:1 with the HilD/reporter stock and incubated for a further 10 min at room temperature. Final assay samples contained 200 nM BODIPY FL C_{16}, 50 µM HilD, 1% DMSO, and varying concentrations of the competing ligand. Samples were centrifuged for 5 min and loaded to standard capillaries.

Competitive MST Assay: Cy5-labelled DNA

To determine affinities for the inhibition of HilD/HilC DNA binding, the 5'-Cy5 labelled A1_62bp DNA fragment was used, as described for EMSAs. HilD or HilC (600 nM) were mixed with labelled DNA (50 nM) and varying concentrations of fatty acids in Tris MST assay buffer. Fatty acids were first titrated in ethanol and diluted 1:100 into protein-DNA samples to give a final ethanol concentration of 1%. Samples were incubated at 37°C for 15 min, centrifuged for 5 min and loaded to Premium Capillaries (Nano temper #MO-K025).

2.5.6 HDX

To investigate the impact of ligands on HilD conformation, 217.8 µl of HilD (25 µM) was supplemented with 2.2 µl of either DMSO or selected ligand (oleic acid, C26, C202; all 10 mM in DMSO), giving final HilD and ligand concentrations of 25 and 100 µM, respectively. For experiments probing the HilD/HilE protein complex, samples contained either individual HilD or HilE, or the HilD/HilE complex. The HilD/HilE complex was established prior to HDX-MS, either by purification using a Superdex^{TM} 75 Increase 10/300 GL column (Cytiva) equilibrated in SEC buffer (50 mM $NaPO_4$ pH 7.0, 200 mM NaCl), or by mixing the individually purified proteins in an equimolar ratio. Final assay concentrations of HilD and/or HilE in all experiments were 25 µM. All samples were stored in a cooled tray (1°C) until measurement.

Preparation of the HDX reactions was aided by a two-arm robotic autosampler (LEAP technologies). 7.5 µl of protein sample (see above) was mixed with 67.5 µl of SEC buffer prepared with 99.9% D_2O to initiate the hydrogen exchange reaction. After incubation at 25°C for 10, 30, 100, 1,000 or 10,000 seconds, 55 µl of the HDX reaction was withdrawn and added to 55 µl of pre-dispensed quench buffer (400 mM KH_2PO_4/H_3PO_4, pH 2.22, 2 M guanidine-HCl) kept at 1°C. 95 µl of the resulting mixture was injected into an ACQUITY UPLC M-Class System with HDX Technology (Waters)[87]. Undeuterated protein samples were prepared similarly (incubation for approximately 10 s at 25°C) through 10-fold dilution of protein samples with H_2O-containing SEC buffer. The injected samples were flushed out of the loop (50 µl) with H_2O + 0.1% (v/v) formic acid (100 µl min^{-1}) and guided to a protease column (2 mm x 2 cm) containing proteases immobilized to the bead material, which was kept at 12°C. For each protein state and timepoint, replicates 1-3 (individual HDX reactions) were digested with porcine pepsin, while replicates 4-6 were digested with a column filled with a 1:1 mixture of protease type XVIII from *Rhizopus spp.* and protease type XIII from *Aspergillus saitoi*. In both cases, the resulting peptides were trapped on an AQUITY UPLC BEH C18 1.7 µm 2.1 x 5 mm VanGuard Pre-column (Waters) kept at 0.5°C. After 3 min of digestion and trapping, the trap column was placed in line with an ACQUITY UPLC BEH C18 1.7 µm 1.0 x 100 mm column (Waters), and the peptides eluted at 0.5°C using a gradient of buffers A (H_2O + 0.1% (v/v) formic acid) and B (acetonitrile + 0.1% (v/v) formic acid) at a flow rate of 60 µl min^{-1} as follows: 0-7 min: 95-65% A; 7-8 min: 65-15% A; 8-10 min: 15% A; 10-11 min: 5% A; 11-16 min: 95% A. The eluted proteins were guided to a G2-Si HDMS mass spectrometer with ion mobility

separation (Waters), and peptides ionized with an electrospray ionization source (250°C capillary temperature, spray voltage 3.0 kV) and mass spectra acquired in positive ion mode over a range of 50 to 2000 m/z in HDMSE or HDMS mode for undeuterated and deuterated samples, respectively[88,89]. [Glu1]-Fibrinopeptide B standard (Waters) was employed for lock-mass correction. During separation of the peptide mixtures on the ACQUITY UPLC BEH C18 column, the protease column was washed three times with 80 µl of wash solution (0.5 M guanidine hydrochloride in 4% (v/v) acetonitrile,) and blank injections performed between each sample to reduce peptide carry-over.

Peptide identification and analysis of deuterium incorporation were carried out with ProteinLynx Global SERVER (PLGS, Waters) and DynamX 3.0 softwares (Waters) as described previously[90]. In summary, peptides were identified with PLGS from the undeuterated samples acquired with HDMSE by employing low energy, elevated energy, and intensity thresholds of 300, 100 and 1,000 counts, respectively. Identified ions were matched to peptides with a database containing the amino acid sequence of HilD, HilE, porcine pepsin, and their reversed sequences with the following search parameters: peptide tolerance = automatic; fragment tolerance = automatic; min fragment ion matches per peptide = 1; min fragment ion matches per protein = 7; min peptide matches per protein = 3; maximum hits to return = 20; maximum protein mass = 250,000; primary digest reagent = non-specific; missed cleavages = 0; false discovery rate = 100. Only peptides that were identified in all undeuterated samples and with a minimum intensity of 30,000 counts, a maximum length of 30 amino acids, a minimum number of three products with at least 0.1 product per amino acid, a maximum mass error of 25 ppm and retention time tolerance of 0.5 minutes were considered for further analysis. Deuterium incorporation into peptides was quantified with DynamX 3.0 software (Waters). Hereby, the datasets generated with pepsin digestion (replicates 1-3) or after digestions with proteases type XIII and XVIII (replicates 4-6) were pooled. All spectra were manually inspected and, if necessary, peptides omitted (e.g., in case of low signal-to-noise ratio or presence of overlapping peptides).

2.5.7 Crosslinking

Crosslinking of HilE was performed using glutaraldehyde. Prior to crosslinking, HilE was extensively dialysed into reaction buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM EDTA). For crosslinking, 100 µg of HilE (in a total volume of 100 µl) was treated with 5 µl of a freshly prepared stock of 2.3% glutaraldehyde for 2-10 min at either room temperature or 37°C. The reaction was terminated by the addition of 10 µl of 1 M Tris-HCl, pH 8.0 after the desired incubation time, and samples visualised using SDS-PAGE.

2.5.8 ITC

Binding of HilD to HilE was studied by Isothermal Titration Calorimetry on a Microcal VP-ITC instrument (Malvern) at 30°C. Purified proteins were either dialysed or subjected to size exclusion chromatography to buffer exchange into assay buffer (20 mM HEPES, pH 7.0, 100 mM NaCl) prior to measurements. HilD (30 µM) was added to the instrument's cell and HilE (300 µM) to the syringe for injection. A first injection of 2 µl was followed by 27 injections of 10 µl, with this first injection excluded from subsequent analysis. A buffer reference measurement was performed by injection of HilE into assay buffer only. Data was fitted to a 1:1 binding model using the MICROCAL ORIGIN software provided by the manufacturer to yield binding parameters.

2.6 Computational Methods

2.6.1 Protein Structure Prediction

Unless otherwise stated, protein structures were predicted using AlphaFold 2. The structure of full-length HilD/HilC proteins were obtained from the publicly available protein structure database[91], and other structures predicted using the ColabFold notebook[92]. Molecular docking of small ligands to HilD was performed using Maestro and Glide from the Schrödinger suite. The HilD-HilE complex was model using the AlphaFold Multimer open-source code[93].

2.6.2 Data Visualisation

Protein structures were visualised, and all protein figures generated, using PyMOL v2.4.0[94]. Protein alignment figures were made using ESPript 3.0[95]. All other figures were created with BioRender.com and Adobe Illustrator.

Chapter 3. Characterisation of HilD and HilC

3.1 The AraC/XylS Family of Proteins

HilD, HilC and RtsA are all members of the AraC/XylS family of transcription factors, one of the most abundant families of transcription factors identified in bacteria[96,97]. AraC/XylS proteins are found widespread, and almost exclusively, amongst prokaryotes. A recent study identified AraC/XylS proteins in 80.6% of prokaryotic genomes (in a non-redundant dataset of 1245 genomes), but in only 2.83% (3 out of 106) of archaeal genomes[97,98]. To date, no AraC/XylS proteins have been identified in eukaryotes. Most members of this protein family act as activators of their target genes, however others also act as either repressors or both, under differing conditions.

The AraC/XylS family is defined by a highly conserved DNA-binding domain (DBD) of 100 amino acids[69,96,99]. This domain consists of 7 α-helices, which form two helix-turn-helix (HTH) motifs connected by an α-helix linker, as seen in the structure of the DBD of AraC (Figure 3.1 A)[100]. Each HTH forms direct contacts with DNA, by inserting into adjacent major grooves of the DNA (Figure 3.1 B). The lack of contacts between the two HTH subdomains allows for the rotation of one HTH with respect to the other. The linker helix (α4) imposes orientation and distance restraints on the two subdomains, ensuring both HTH motifs bind concertedly to the same face of the DNA, whilst the increased flexibility is important to allow bending of the DNA. The interaction of the DBD with two adjacent major grooves results in a binding site of 17-21 bases for each protein monomer, dictated by the length of helix α4.

Whilst some AraC/XylS proteins, such as the *E. coli* transcription regulator MarA (Figure 3.1 B), consist of only the DBD, most of the proteins in this family contain at least one additional domain. Commonly referred to as the regulatory, effector-binding or companion domain, these domains have a range of functions, including ligand binding, facilitating protein-protein interactions (such as with the transcriptional machinery), and controlling the DNA-binding activity of the protein. Analysis of 26,540 AraC/XylS proteins has shown that the most common number of domains is two, with 60.4% of proteins having this structure[98]. 34.6% of AraC/XylS proteins consist of only the DBD, whilst 5% have more than two domains. Amongst

the identified two-domain proteins, in which the DBD is connected to the regulatory domain by a flexible linker, the DBD is located at the C-terminus in approximately 70% of cases[98].

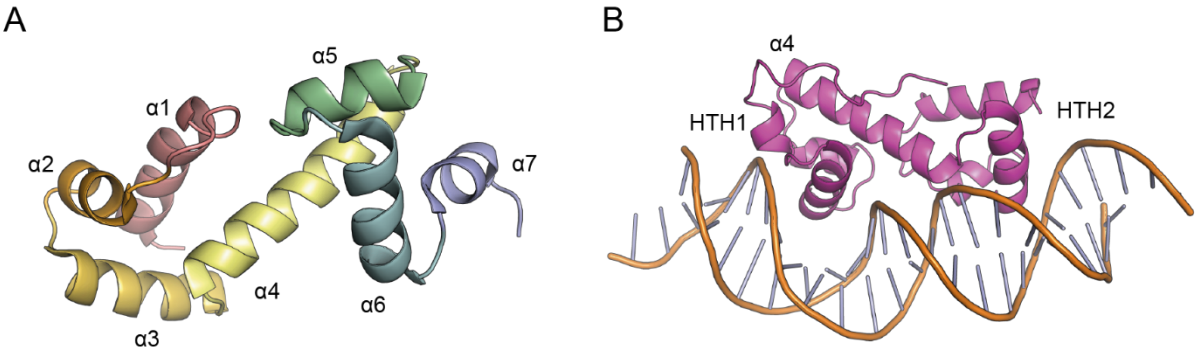


Figure 3.1. The DNA binding domain is highly conserved amongst members of the AraC/XylS protein family. (**A**) Structure of the DNA-binding domain of AraC (PDB: 2K9S), highlighting the 7 α-helices constituting this domain. The DBD is colored in a rainbow-like effect with the N-terminus in blue and the C-terminus in red. Helices α2- α3 and α5-α6 form two HTH motifs, which are connected by helix α4. (**B**) Structure of the transcription regulator MarA bound to DNA (PDB: 1BL0). The two HTH motifs both form direct contacts with target DNA, inserting into the major grooves of the DNA helix.

Although less conserved than the DBD, there is also homology amongst the regulatory domains of closely related AraC/XylS proteins that have similar functions. Analysis of the regulatory domains of AraC/XylS proteins showed that these domains belong to a range of other protein domain families[98]. The most common family of regulatory domains is the regulatory domain of AraC, the transcription regulator that gives its name to this protein family. In AraC, this domain is responsible for dimerisation and binding of the sugar arabinose, which regulates protein activity[101–104]. The AraC ligand-binding domain is closely related to the cupin barrel domain, another of the most common domain families found amongst AraC/XylS regulatory domains. Additionally, a common dimerisation interface has been observed between many AraC/XylS proteins, formed by a conserved helix of the regulatory domain[64].

AraC/XylS proteins regulate genes involved in one of three cellular processes: Metabolism, including carbon and nitrogen source utilisation; stress response, such as to pH changes and DNA damage; and virulence[69,96,98,99]. Some AraC/XylS proteins also regulate many genes with different functions, belonging to multiple of these functional classes. For example, in

enteroaggregative *E. coli*, AggR regulates both virulence genes and those required for lipid metabolism[105]. Alignment of the DBD of AraC/XylS proteins results in the clustering of sequences into groups, according to the three functional groups. The DBD can therefore be used as a template for building a phylogenetic tree, and to predict the regulatory role of newly identified members of this family[99]. The clustering of transcription regulators into groups of common function supports the conclusion that they evolved from a common ancestor[98].

HilD, HilC and RtsA all have a two-domain structure with the DBD at the C-terminus. The N-terminal domain is responsible for dimerisation and ligand binding[61,64]. This mirrors the structure of the *V. cholerae* transcription regulator ToxT, one of the few AraC/XylS proteins for which the full-length structure has been experimentally determined[57,106]. ToxT activates expression of the genes encoding the two main *V. cholerae* virulence factors, the cholera toxin (CT) and the toxin-coregulated pilus (TCP)[107]. ToxT is a close homolog of HilD/HilC/RtsA and has both a high sequence similarity (35.9%) and identity (19.6%) to HilD. Binding of small molecules to a hydrophobic pocket at the interface of the two domains causes changes in overall protein conformation and flexibility, affecting DNA-binding and hence ToxT activity[58].

AraC/XylS transcription factors are often poorly soluble at higher concentrations, which combined with the inter-domain structural flexibility has limited the structural characterisation of the full-length proteins[100,104,108]. However, expression of individual domains has been observed to circumvent these problems. A search for homologs using HHPred[109,110] revealed several AraC/XylS proteins for which only the regulatory domain has been characterised. The regulatory domains of AraC, ExsA, BgaR and RhaR (respective PDB codes: 2ARC, 4ZUA, 6NX3, 5U93) all belong to the AraC-like binding domain (Pfam PF02311) family. The predicted domain organisation of HilD (and HilC, RtsA) can be summarised as an N-terminal AraC-like binding domain connected to a C-terminal DNA-binding domain, akin to the structure of ToxT.

In this chapter, we biochemically and biophysically characterised the AraC/XylS transcription regulators HilD and HilC, which control the expression of virulence-associated genes in *Salmonella*. We employed different structural characterisation techniques to attempt to obtain the first crystal structure of one of these regulators. In the absence of high-resolution experimental data, we predicted the structures of HilD and HilC using computational

methods. We used hydrogen-deuterium exchange mass spectrometry to identify disordered regions of HilD and to support the predicted model. We confirmed both purified proteins to be active *in vitro*, forming homodimers and binding to specific DNA sequences as previously reported, and optimised a number of assays that could be used to probe the binding of different ligands to HilD and HilC.

3.2 Biophysical Characterisation of HilD and HilC

HilD and HilC were recombinantly expressed in *E. coli*. HilD was expressed as a fusion to an N-terminal His_6-SUMO tag, as no expression was detected for constructs containing only an N-terminal His_6-tag. Unlike HilD, HilC could be expressed with an N-terminal His_6-tag, followed by a TEV protease cleavage site for tag removal during purification. Although RtsA could be expressed as a SUMO-fusion protein, we encountered issues with apparent degradation both before and after cleavage of the SUMO-tag, which prevented further analysis. Issues with the solubility of RtsA have also been reported previously[63], and given *in vivo* experiments showed it not to be a target of C26 (discussed further in Chapter 6), no further work on this regulator was pursued.

Both HilD and HilC could be successfully purified to homogeneity (Figure 3.2 A-B), and far-UV circular dichroism spectroscopy (CD) was used to confirm both proteins were folded. Similar CD spectra, indicative of a predominantly alpha helical secondary structure, were obtained for both HilD and HilC, consistent with the prediction that both proteins have a very similar overall fold (Figure 3.2 C-D). The melting temperature of both proteins was determined by monitoring the change in ellipticity (Θ) at 222nm, with increasing temperature (Figure 3.2 E-F). HilD and HilC were found to have respective melting temperatures of 52.8°C and 41.3°C.

Thermostability was also determined using nanoscale differential scanning fluorimetry (NanoDSF). Intrinsic protein fluorescence is monitored at wavelengths 330 and 350 nm, and the melting temperature obtained by fitting the ratio of these two fluorescence values against temperature (Figure 3.2 G-H). This technique has the advantage of lower sample demands compared to CD and no specific buffer requirements, so was used to optimise buffers for protein purification, storage, and structure characterisation experiments. The melting

temperatures of HilD and HilC were determined to be 52.2°C and 42.2°C, respectively, comparable to those calculated from CD. We also used this method to investigate ligand-induced effects on HilD/ HilC stability and subsequently screen different ligands for their binding to HilD.

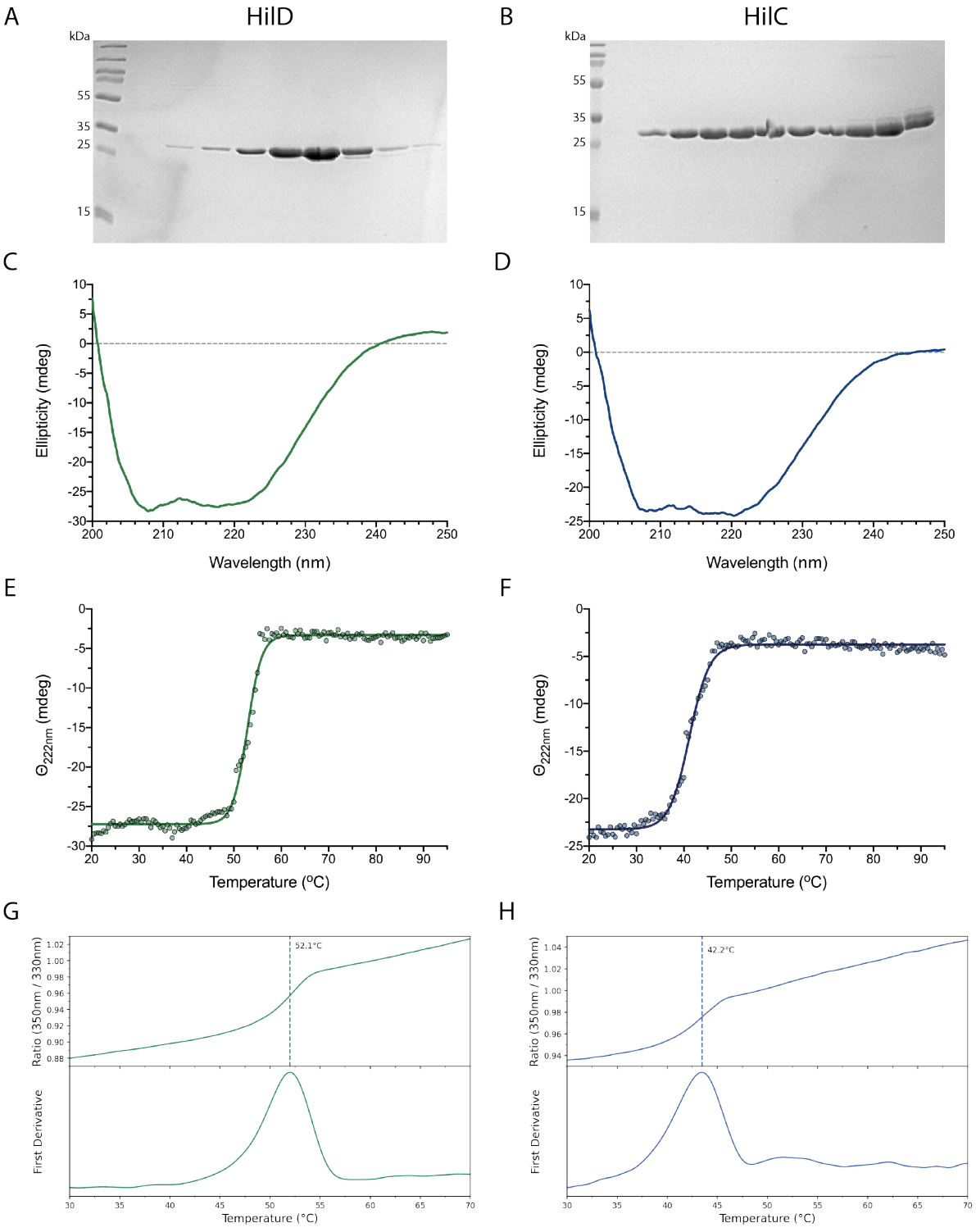


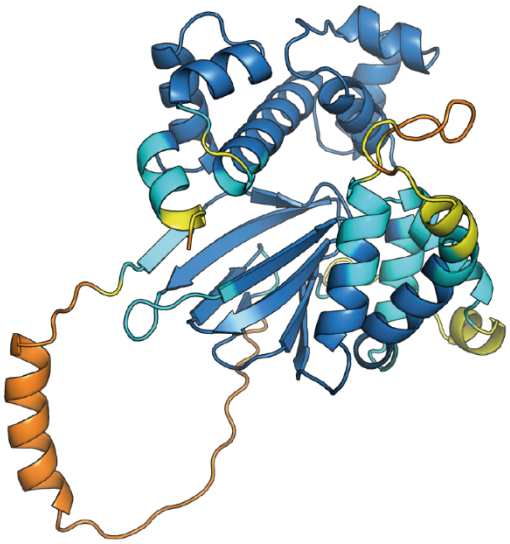
Figure 3.2. Biophysical Characterisation of HilD and HilC. Samples were collected from fractions corresponding to the eluted peak of the desired proteins during the final purification step: (**A**) HilD: Superdex 75 26/60; (**B**) HilC: Ni NTA. Samples were separated on a 15% SDS gel. Single bands of approximately 35 kDa correspond to the expected molecular weight of HilD and HilC monomers. (**C-D**) Far-UV CD spectra of (**C**) HilD and (**D**) HilC were recorded at 20°C using 10 µM of the respective protein, over the wavelength range 195-250 nm. (**E-F**) Thermal stability of (**E**) HilD and (**F**) HilC, as determined using CD. The ellipticity (Θ) was monitored at 222 nm and melting temperatures were calculated using GraphPad Prism. (**G-H**) Thermal stability of (**G**) HilD and (**H**) HilC was also determined using NanoDSF. Top panels show the change in ratio of 330/350 nm with increasing temperature; bottom panels show the first derivative of 330/350 nm, visualising infliction points that indicate protein unfolding and provide an estimate of melting temperature. Melting temperatures were calculated using the NanoTemper PR.Stability Analysis software.

3.3 Structure Prediction of HilD

There has been a considerable recent improvement in the accuracy of protein structure prediction methods, namely with the advent of AlphaFold 2[111]. The predicted structures of HilD, HilC and RtsA concur with the expected domain organisation , and are comparable to other AraC/XylS proteins for which full-length structures have been experimentally determined[57,59,112]. The N-terminus of HilD (residues 1-35) was predicted with very low confidence (pLDDT <50%) in the AlphaFold model, indicative of disorder (Figure 3.3 A). Pairwise sequence alignments showed that HilD has an extended N-terminus compared to ToxT and sequence analysis using Quick2D also predicted these residues to be disordered (Figure 3.3 B)[113–115].

To experimentally validate the predicted disordered regions of HilD, we utilised hydrogen-deuterium exchange mass spectrometry (HDX-MS). HDX-MS relies on the principle that the amide backbone protons of a protein undergo exchange with deuterium in deuterated water (D_2O) on a time scale that can be measured by mass spectrometry. Folded regions of the proteins (contained within the core) are protected from HDX, whilst disordered regions that lack hydrogen-bonding interactions and are hence exposed to the deuterated solvent undergo more rapid hydrogen-to-deuterium exchange[116]. Regions of HilD showing high HDX, with limited further increase in this HDX over the time course of the experiment were defined as disordered (Figure 3.4).

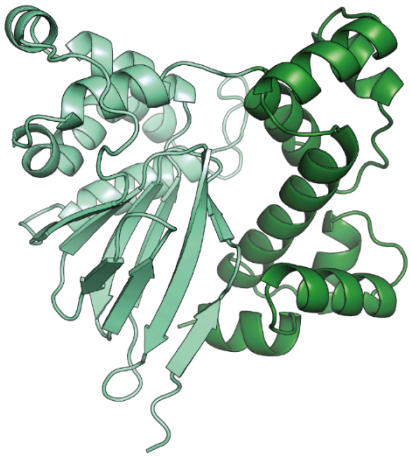
A



B

```
          1         10        20        30        40        50        60
HilD      MENVTFVSNSHQRPAADNLQKLKSLLTNTRQQIKSQTQQVTIKNLYVSSFTLVCFRSGKL
ToxT      ..................MIGKKSFQTNVY..RMSKFDTYIFNNLYINDYKMFWIDSGIA
```

C



D

Figure 3.3. The predicted structure of HilD closely resembles that of *V. cholerae* ToxT. (**A**) AlphaFold model of HilD, coloured according to the per-residue confidence (LDDT): High (>90%, dark blue), good (70-90%, cyan), low (50-70%, yellow), very low (<50%, orange). (**B**) Pairwise sequence alignment (carried out with ClustalΩ[117]) of HilD and ToxT, highlighting the increased length of the N-terminus of HilD with that of ToxT. The full sequence alignment is shown in Figure S2.1 (Appendix 2). (**C**) AlphaFold model of HilD, coloured by domain. The N-terminal regulatory domain is shown in light green, whilst the DNA-binding domain is shown in dark green. Residues 1-35 were removed for clarity. (**D**) Crystal structure of *V. cholerae* ToxT (PDB: 4MLO). The N-terminal domain is coloured in pink and the C-terminal DBD is coloured in purple. Structural alignment of the predicted HilD structure and ToxT gave a root-mean-square deviation (RMSD) of 4.4 Å.

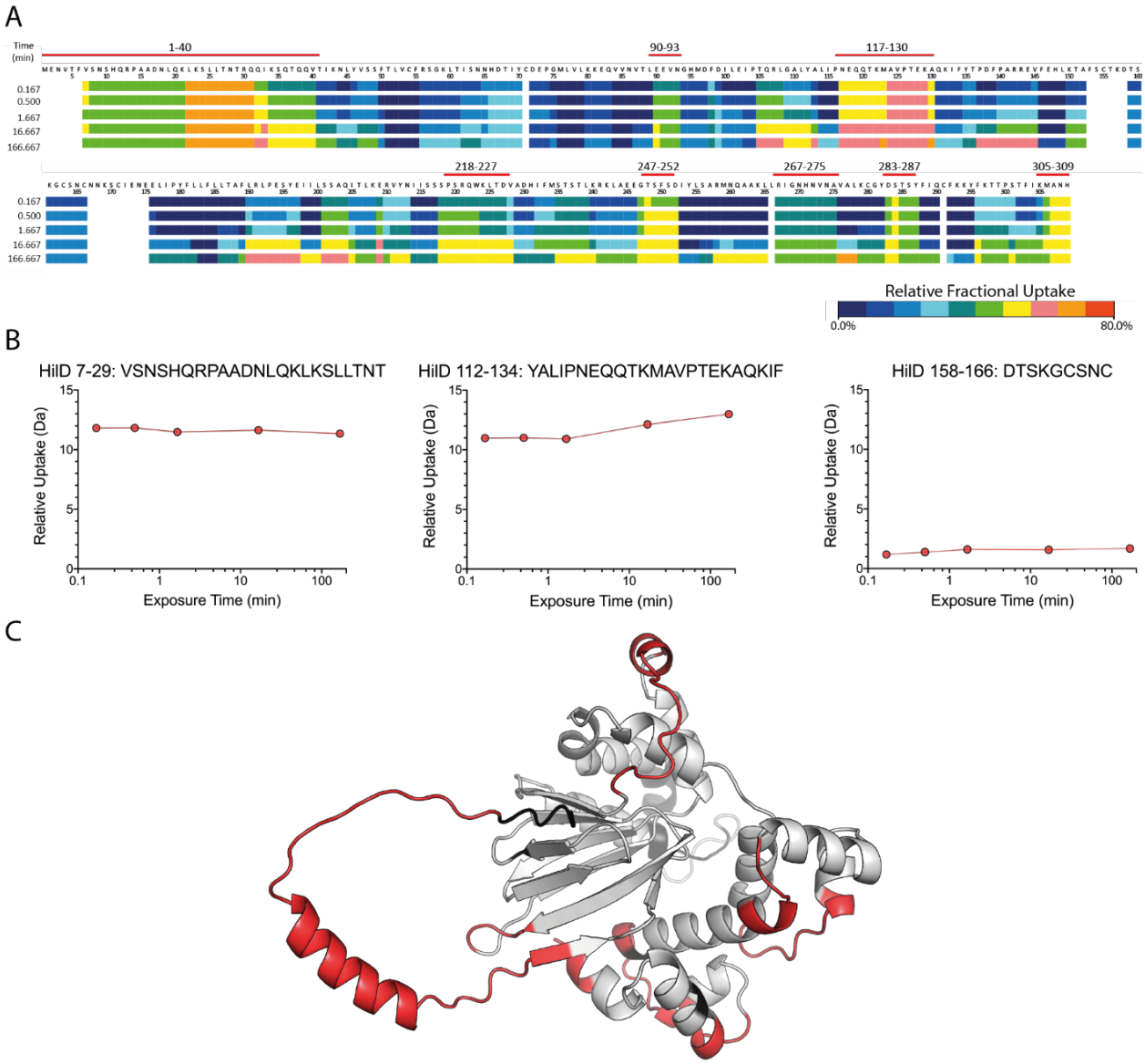


Figure 3.4. Flexible regions of HilD were identified by HDX-MS. (**A**) Relative HDX of HilD. The incorporation of deuterium (D) was evaluated across the entire length of HilD. Following HDX experiments, HilD was digested with porcine pepsin or fungal proteases XIII+XVIII to achieve maximal sequence coverage (94.2%). The D-incorporation of each residue is coloured in rainbow from blue (0%) to red (80%). Maximal uptake was 69.2%. Gaps in the HDX profile correspond to regions with no peptide coverage. The red lines indicate residues that were determined to be disordered. Time lengths of the HDX experiment are indicated on the left of the figure. (**B**) Deuterium uptake charts for three representative peptides of HilD. Peptides encompassing residues 7-29 and 112-134 show high D-uptake and are indicative of disordered regions. The peptide of residues 158-166 shows low relative D-uptake, indicative of a structured region. (**C**) AlphaFold model of HilD, coloured according to the disordered regions defined by HDX-MS. Disordered regions are highlighted in red, whilst those shown in black did not have any sequence coverage.

HDX-MS confirmed that the N-terminus (comprising the first 40 residues) is disordered, with no evidence supporting the existence of the α-helix predicted in this region of the model. Hence, in subsequent computational methods for ligand docking to HilD, and the figures shown throughout this thesis, the first 35 residues of HilD are removed for clarity. Other regions of increased flexibility are the loop (residues 117-130) connecting the cupin barrel to helix 1, and the loops connecting the helices of the DNA-binding domain. In the latter case, increased flexibility is this region important for binding to and dissociation from DNA sequences. Another region of the AlphaFold model predicted with low confidence is the loop formed by residues 155-173. No further information could be obtained from HDX-MS due to lack of peptide coverage of this region, although residues 159-166 do not show increased flexibility.

3.4 Attempted Structural Characterisation of HilD

Initial crystallisation trials of HilD (both apo and in complex with the various ligands discussed in subsequent chapters) failed to yield any crystals. Considering our lack of success in crystallisation of the full length HilD protein, we pursued other approaches for structural characterisation. HilD could be successfully expressed in M9 minimal media for NMR experiments, and whilst the protein could be seen to be well-folded in 1D experiments, issues with precipitation were encountered at the concentrations required for 2D 1H-15N HSQC experiments. The only peaks observed were for a flexible region of around 30-35 residues, likely relating to the unstructured N-terminus, as previously defined by HDX. Initial cryoEM experiments were also performed, with grids prepared for a range of HilD concentrations (1-8 mg ml^{-1}). However, only larger aggregates, and no single particles, were observed.

We next performed limited proteolysis experiments to identify potential proteolytic fragments of HilD. Fragments that are resistant to further proteolysis often form more compact structures and therefore may be more prone to crystallise. Subtilisin digestion yielded a single fragment of 16-20 kDa in size (Figure 3.5).

Mass spectrometry analysis showed this fragment to correspond to the N-terminal domain of HilD. Analysis of detected peptides revealed the likely subtilisin cleavage site to be at residue

L206. This digestion product is smaller than the mass of the intact N-terminal domain (23 kDa), likely due to additional truncation of the flexible N-terminus. However, further analysis was limited by large peptide coverage over the whole HilD sequence.

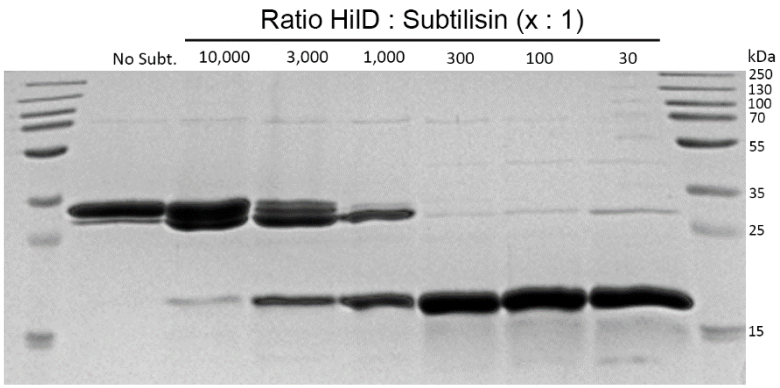


Figure 3.5. Limited proteolysis of HilD yielded a single digestion product, corresponding to the N-terminal domain of HilD. HilD (150 µg) was incubated with increasing molar ratios of subtilisin. A single digestion product of ~18 kDa in size is observed, which is resistant to further proteolysis and remains stable at higher subtilisin concentrations.

Based on these results, we cloned multiple constructs of the N-terminal domain of HilD (NTD), with different purification tags. As observed for other AraC/XylS proteins, the NTD was more soluble than full-length HilD and could be expressed without a SUMO-fusion tag. The first six residues at the N-terminus were replaced by a His_6-tag for purification. The NTD construct was confirmed to be well folded after purification using CD spectroscopy (Figure 3.6 B). Thermal stability was determined from CD and NanoDSF (Figures 3.6 C-D), with calculated melting temperatures of 56.3°C and 56.7°C from each method, respectively.

Although crystallisation trials of the subtilisin digestion product failed to yield any crystals, we did obtain microcrystals for the purified His-tagged NTD construct. Whilst these crystals did diffract X-rays, they did so only to a resolution of >10 Å and we were unsuccessful in further improving either the diffraction resolution or crystal quality.

Sequence alignments with other AraC/XylS transcription factors, for which crystal structures have been published, showed the presence of an insertion present only in HilD (Figure 3.7 A). This was one of the few regions of HilD predicted with low confidence in the AlphaFold model, whilst previous homology models had displayed it as a disordered loop.

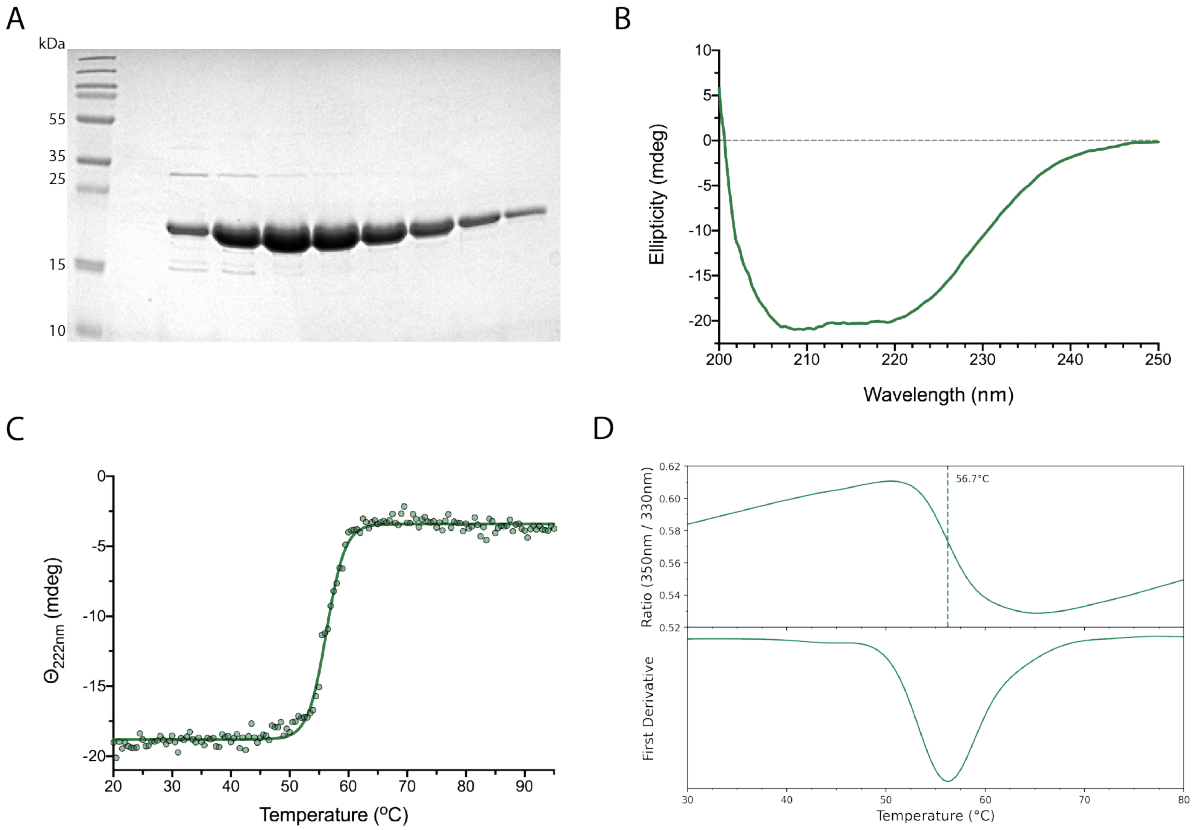


Figure 3.6. Biophysical characterisation of the His_6-$HilD_{7-206}$ (NTD) construct. (**A**) The His_6-NTD construct was purified using a two-step purification protocol. The protein eluted as a single peak from a Superdex 75 column, and samples of eluted fractions corresponding to this peak separated on a 15% SDS gel, showing a pure sample of 23 kDa in size. (**B**) The CD spectrum of the HilD NTD (0.3 mg ml^{-1}) showed the protein to be well folded, with a mixture of helical and β-strand secondary structures. (**C-D**) Thermal stability of the NTD was determined by monitoring the ellipticity at 222nm (CD, **C**) or intrinsic fluorescence (NanoDSF, **D**) at increasing temperatures.

We attempted to optimise our crystallisation constructs by replacing this extended loop with a shorter linker of 1-3 residues. Three different linkers were chosen and cloned into the sequence of the full-length HilD construct. These mutants were successfully expressed and purified, and their stabilities assessed by NanoDSF (Figure 3.7 B). The construct with a single glycine linker in place of this putative loop region was found to have a melting temperature of 51.0 °C. This deletion was also cloned into the NTD construct, and crystallisation screening performed for both full-length and NTD constructs. We again obtained microcrystals for the NTD construct, which could also be optimised to large protein crystals. However, the X-ray diffraction was poor, and despite extensive optimisation of crystallisation conditions, we were unable to obtain a diffraction resolution better than 5Å. Although within the limits for

structural determination, this resolution would not be sufficient to properly validate the AlphaFold model and visualise the binding modes of small molecules.

Subsequent sequence analysis revealed a possible atypical Zn-finger motif within this region. Unlike the homology models that were used when designing these constructs, more recent structure prediction methods (AlphaFold 2 and RoseTTAFold) orientate these cystine residues in such a way they could form a metal-coordination site (Figure 3.7 C). HDX-MS did not identify residues 159-166, at the centre of this region, to be disordered. However, the lack of peptide coverage for residues 153-158 and 167-175 make any further conclusions difficult, and we could not confirm whether a metal ion was bound to HilD in the native state.

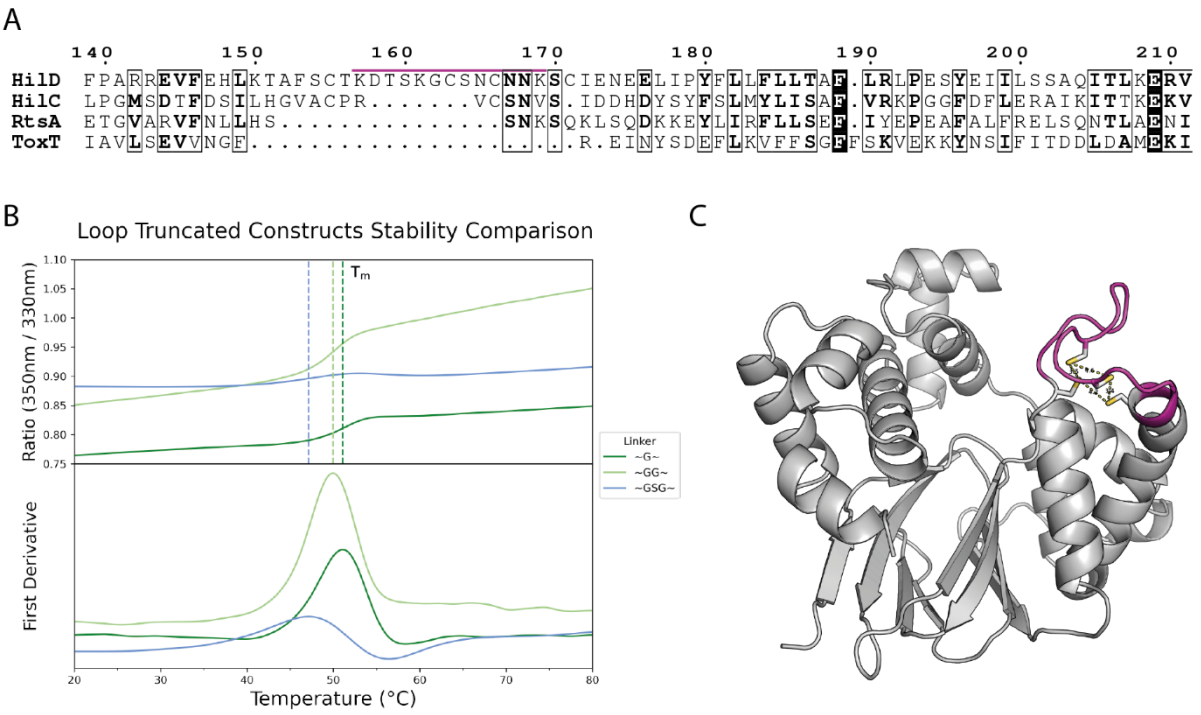


Figure 3.7. HilD contains an elongated loop region compared to other close homologs. (**A**) Multiple sequence alignment of loop regions of HilD, HilC, RtsA and ToxT. Residues that were replaced by a short linker in truncated constructs are denoted by the magenta line. (**B**) The stability of HilD mutants, in which this loop was mutated to a short linker, was determined by NanoDSF. Constructs with a linker of one or two glycine residues both have comparable melting temperature to wild type HilD, whilst unfolding was less clear for the construct with a GSG linker. The dashed line on the upper plot shows the infliction point of unfolding, reflecting differences in the stability of these constructs. (**C**) AlphaFold predicted structure of HilD. The loop region, defined in (A) is coloured in magenta. The four cystine residues that could form a putative Zn-finger motif are shown as sticks, with yellow dashes highlighting the interactions between them.

Following our lack of success in obtaining well-diffracting crystals of HilD, we sought potential close homologs for use in crystallisation trials. We searched for homologs of *S. enterica* HilD in the nonredundant protein database in NCBI, using PSI-BLAST implemented in the MPI Bioinformatics Toolkit[110,118]. To account for the large number of HilD homologs from other *Salmonella* strains, we first removed sequences with identities > 95%. As we sought homologs for the full-length protein, and not just the highly conserved DNA-binding domain, hits were then filtered for sequence coverage of > 80%, yielding 16 sequences. The best hit was the homologous HilD protein from *Enterobacter lignolyticus*, with a sequence identity of 44.3%. A sequence identity of 70-90% would be desirable for suitable homologs with shared ligand binding modes and specificities, and so characterisation of this protein was ultimately not pursued.

3.5 Attempted Structural Characterisation of HilC

Similarly to HilD, HilC activity is also regulated by small molecules, and the two transcription regulators have been shown to bind common ligands[39,56]. However, detailed understanding of these similar, yet non-identical binding specificities, remains unclear in the absence of higher resolution data. HilC is also the closest homolog of HilD, with a sequence identity (33.1%) and similarity (54.1%) to HilD. An experimental structure of HilC would provide a better comparison to the predicted structure of HilD than the structure of ToxT (which has sequence identities of 19.6% and 22.6% to HilD and HilC, respectively).

Crystallisation trials of HilC could not be performed due to insufficient yields and limits of protein solubility and so, as for HilD, we cloned multiple constructs of the N-terminal domain. Whilst the N-terminal domains of HilD and HilC still show high sequence similarity (45.0%) and identity (21.1%), both the unstructured N-terminus and the extended loop are shorter in HilC than in HilD, making this domain a promising target for crystallisation. The C-terminus of this construct was defined by sequence alignment to HilD and ExsA, for which the structure of the N-terminal domain was previously determined[119]. The HilD domain boundary was determined as residue T206, whilst the structure of ExsA (PDB: 4ZUA) contains electron density up to residues L165/ N167 for the respective chains in the unit cell. It has previously

been reported that charged residues (and in particular lysine) at the C-terminus result in higher protein expression compared to hydrophobic residues[120]. Therefore, we designed constructs of the HilC N-terminal domain terminating at residues K190 or K194, and containing a cleavable N-terminal His_6-TEV- purification tag. Whilst both constructs were successfully expressed and purified to homogeneity, no crystals were obtained for constructs either retaining or lacking the N-terminal His_6-tag.

Target validation of the novel antipathogenic compound C26 (Chapter 6) revealed it to bind specifically to HilD, and not HilC. This would hint at notable differences in the putative binding pockets and therefore limit the usefulness of information that could be obtained from the structure of the HilC NTD. Combined with our lack of success in initial crystallisation trials, and the observation that HilC is less thermally stable and soluble than HilD, further construct optimisation and crystallisation screening was not pursued for HilC.

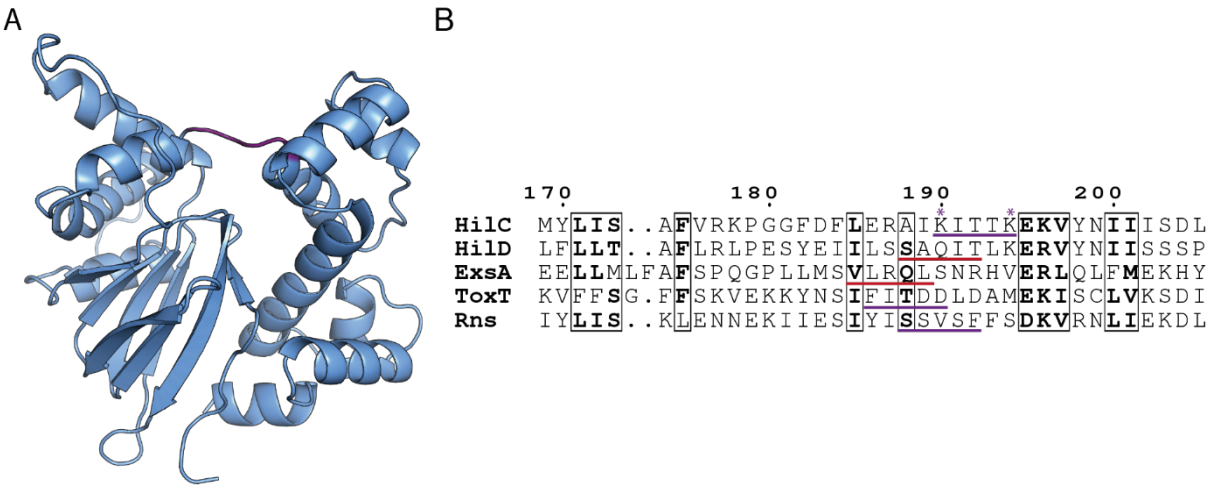


Figure 3.8. Predicted structure and domain boundaries of HilC. (**A**) The structure of HilC, as predicted by AlphaFold. The domain linker region (residues 190-193) is highlighted in purple. The N-terminus (residues 1-26) is removed for clarity. (**B**) Multiple sequence alignment of HilC, HilD, ExsA, ToxT and Rns. The C-terminal residues of the fragments obtained from the proteolysis digestion of HilD (T206) and ExsA (E178)[119] are underlined in red. The domain linker region in the crystal structures of the full-length ToxT and Rns proteins are underlined in purple. The predicted domain linker of HilC is underlined in purple and the residues selected as the C-terminus of the cloned HilC NTD constructs are marked by an asterisk (*).

3.6 Oligomerisation of HilD and HilC

HilD, HilC and RtsA have previously been shown to form both homo- and hetero- dimers with one another[64]. We first used size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) to confirm that both purified proteins exist purely as dimers in solution (Figure 3.9 A). The helix formed by residues 180-192 forms the dimerisation interface of HilD, and SEC-MALS showed that the HilD NTD also exists as a dimer in solution. A lower concentration of HilC (50 μM) was used due to issues with protein solubility and precipitation at higher concentrations.

We then used an orthogonal MST-based assay to quantify the homodimerisation of HilD. HilD was fused to an N-terminal GFP tag for detection in microscale thermophoresis (MST) assays. An N-terminal fusion was chosen due to the predicted flexibility of the HilD N-terminus and to avoid possible interference with the DNA binding activity. GFP-HilD (at a constant concentration of 50 nM) was incubated with increasing concentrations of (unfused) HilD. At lower HilD concentrations, GFP-HilD will exist as (GFP-HilD) homodimers and/or monomers. At higher HilD concentrations, GFP-HilD will exist almost exclusively as heterodimers with unfused HilD. This change in dimerisation state of GFP-HilD is reflected in the thermophoretic behaviour, from which we could calculate an apparent equilibrium dimer dissociation constant, $K_{d,dimer}$, of 3.39 ± 0.77 μM (Figure 3.9 B). The respective K_d for HilC homodimerisation could not be obtained due to a lack of success in purifying a GFP-HilC fusion protein, whilst the lower solubility of HilC prevented us from achieving a protein concentration required to obtain a binding curve for HilD-HilC heterodimerisation. However, changes in the thermophoretic behaviour of GFP-HilD in the absence or presence of HilC demonstrated the formation of a GFP-HilD - HilC heterodimer. This optimised MST assay was subsequently used to investigate the effects of different ligands on HilD homodimerisation.

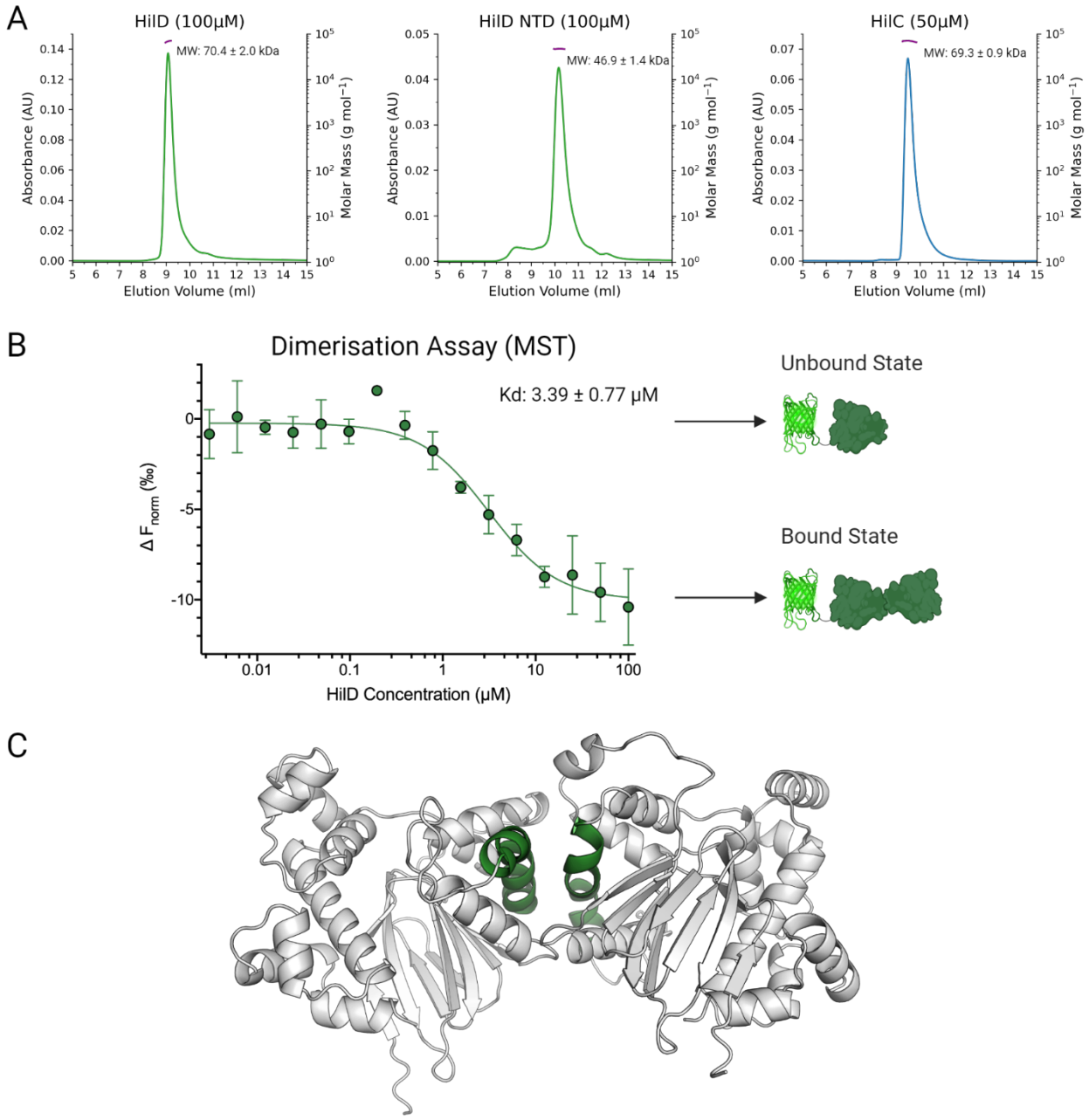
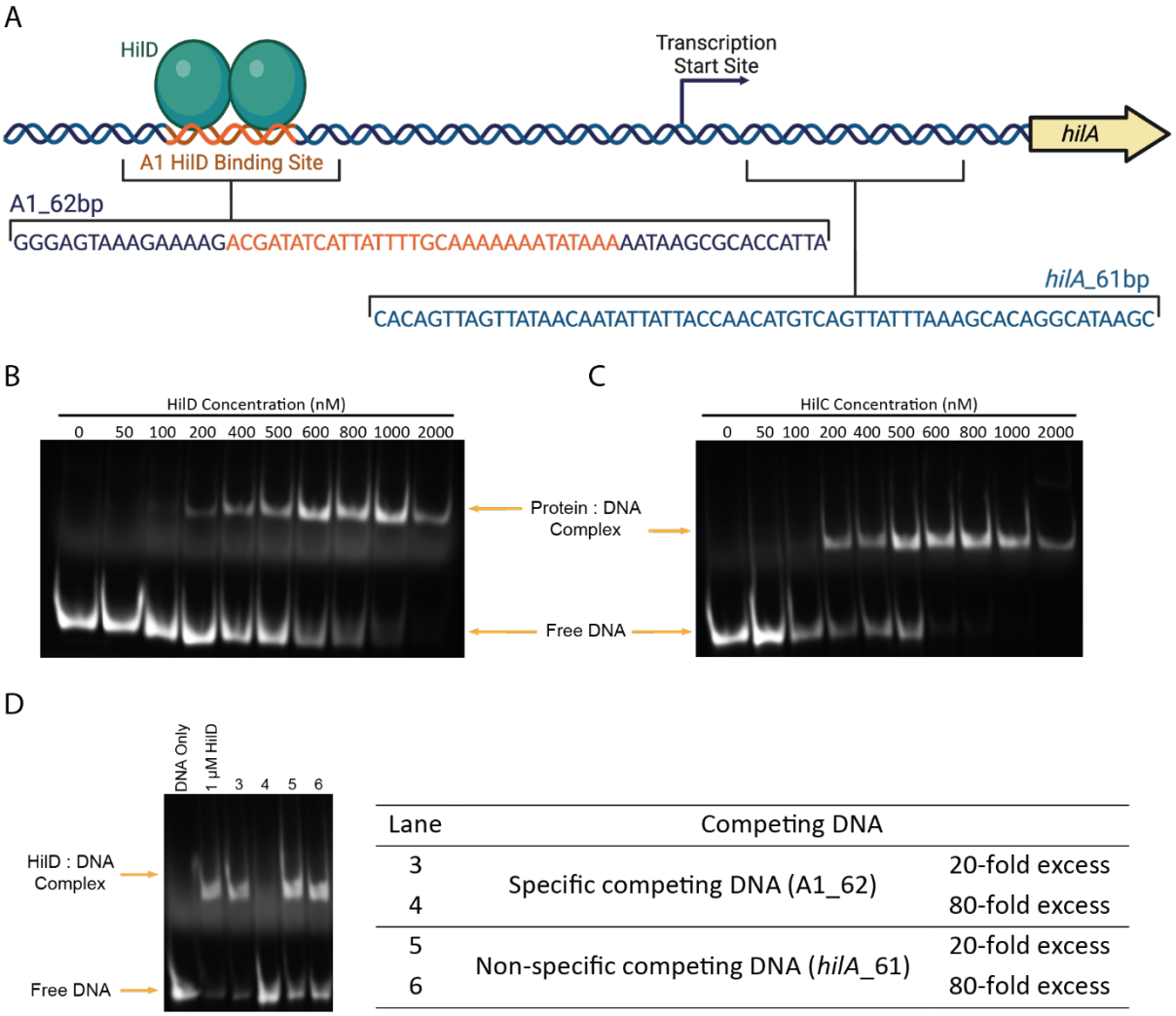


Figure 3.9. HilD and HilC both exist purely as dimers in solution. (**A**) SEC-MALS profiles of HilD, the HilD N-terminal domain (NTD) and HilC. All proteins eluted as a single peak, which light scattering confirmed to correspond to a dimeric species. Calculated molecular weight values correspond to 3 repeat experiments. (**B**) HilD dimerisation measured by MST. Unlabelled HilD protein (3.05 nM to 100 µM) was incubated with 50 nM GFP-HilD. Changes in thermophoresis at an MST on-time of 1.5 seconds were fitted to obtain the $K_{d,dimer}$. (**C**) AlphaFold Model of the HilD homodimer. The dimerisation helix, as reported previously[64], is highlighted in dark green.

3.7 DNA Binding of HilD and HilC

Both HilD and HilC bind to overlapping, specific binding sites within the promoter regions of their target genes. To confirm purified HilD and HilC were active *in vitro*, we performed electrophoretic mobility shift assays (EMSA) with a Cy5-labelled 62bp DNA fragment of the *hilA* promoter, encompassing the A1 binding site (A1_62bp) (Figure 3.10 A). The formation of a stable protein-DNA complex is observed at increasing concentrations of HilD and HilC (Figure 3.10 B-C). 600nM HilD/HilC was sufficient to bind almost all labelled DNA, whilst aggregation of the protein-DNA complex was observed at higher protein concentrations. Therefore, respective DNA/protein concentrations of 50 and 600 nM were used in subsequent EMSAs to study the effects of different ligands on HilD and HilC DNA-binding activity. All further EMSAs shown throughout this thesis used this labelled A1_62bp fragment, unless otherwise stated.

Specificity of this interaction was demonstrated for HilD, as addition of an 80-fold molar excess of specific non-labelled DNA, the same A1_62bp fragment described above but lacking a Cy5 fluorescent tag, inhibited the formation of the observed HilD-DNA complex. In contrast, an 80-fold excess of a non-specific 61bp fragment, comprising a sequence located further downstream in the *hilA* promoter (*hilA*_61bp), had no effect on HilD binding to the A1 site (Figure 3.10 D).



Lane	Competing DNA	
3	Specific competing DNA (A1_62)	20-fold excess
4		80-fold excess
5	Non-specific competing DNA (*hilA*_61)	20-fold excess
6		80-fold excess

Figure 3.10. Purified HilD and HilC bind to specific binding sites within target promoters. (**A**) Schematic of the *hilA* promoter, showing the location and sequences of the DNA fragments used in EMSAs within the promoter region. The specific binding fragment (A1_62bp) encompasses the A1 site, highlighted in orange. The sequence of the non-specific fragment (*hilA*_61bp) is located downstream of the transcription start site. (**B**-**C**) EMSAs showing the DNA-binding activity of HilD and HilC. All lanes contain 50 nM of the Cy5-labelled A1_62bp DNA fragment, with increasing concentrations of HilD/HilC, as labelled. DNA is labelled with a Cy5 fluorophore at the 5' end of the forward strand, with 50 nM DNA found to be optimal for image detection. Aggregation was observed at protein concentrations > 600 nM, indicated by fluorescence in the wells from DNA unable to enter the gel. (**D**) EMSA showing that the DNA-binding of HilD is specific. All lanes contain 50 nM labelled DNA, lanes 2-6 contain 1 µM HilD. Lanes 3-6 contain an excess of non-labelled competing DNA, as annotated.

Chapter 4. Long Chain Fatty Acids are Small-Molecule Regulators of HilD and HilC

4.1 Introduction

Short- and long-chain fatty acids are highly abundant in the mammalian gut, where they are acquired from dietary sources and also produced by both the host (to aid in the digestion of fats) and the intestinal microbiota, as by-products of metabolism[121]. Long-chain fatty acids (LCFAs) can be utilized by prokaryotes as both carbon and energy sources. In *Salmonella enterica*, LCFAs are transported across the outer membrane by the β-barrel FadL and then passively diffuse across the inner membrane. Once in the cytoplasm, LCFAs are converted to their corresponding acyl-CoA form by FadD, preventing diffusion back across the membrane and trapping them in the cytoplasm. Acyl-CoA-LCFAs are then metabolised via the β-oxidation pathway to generate Acetyl-CoA and the shortened fatty acyl-CoA molecule, which re-enters the degradation cycle (Figure 4.1)[122]. Acetyl-CoA is then further metabolised in the tricarboxylic acid (TCA) cycle, to generate both energy and precursors for cellular biosynthetic pathways[122]. LCFAs bind directly to and repress the transcription regulator FadR, countering its repression of the *fad* operon, which encodes the genes essential for fatty acid transport and degradation.

Different fatty acids are present in varying concentrations throughout the intestine and hence, are also utilized by enteric pathogens to sense their intestinal location and coordinate the expression of virulence genes at specific locations where invasion can occur[51]. In *Salmonella*, LCFAs have been shown to down-regulate the expression of *hilA*, and hence the SPI-1 genes required for host-cell invasion. After import by FadL, free LCFAs (rather than the acyl-CoA-LCFA product or subsequent degradation products) directly bind to HilD, HilC and RtsA, inhibiting their DNA-binding ability[56]. Short- and medium-chain fatty acids also affect *hilA* expression, although they do not bind directly to HilD/HilC/RtsA. Acetate and formate both upregulate *hilA* via the BarA/SirA TCS, which activates *hilD* translation[51,52,123]. Butyrate downregulates *hilA* independently of HilD[41,124], propionate indirectly affects the stability of HilD via the propionyl-CoA metabolite[125], whilst the precise mechanism by which medium-chain fatty acids (C6-10) downregulated *hilA* is unknown[126].

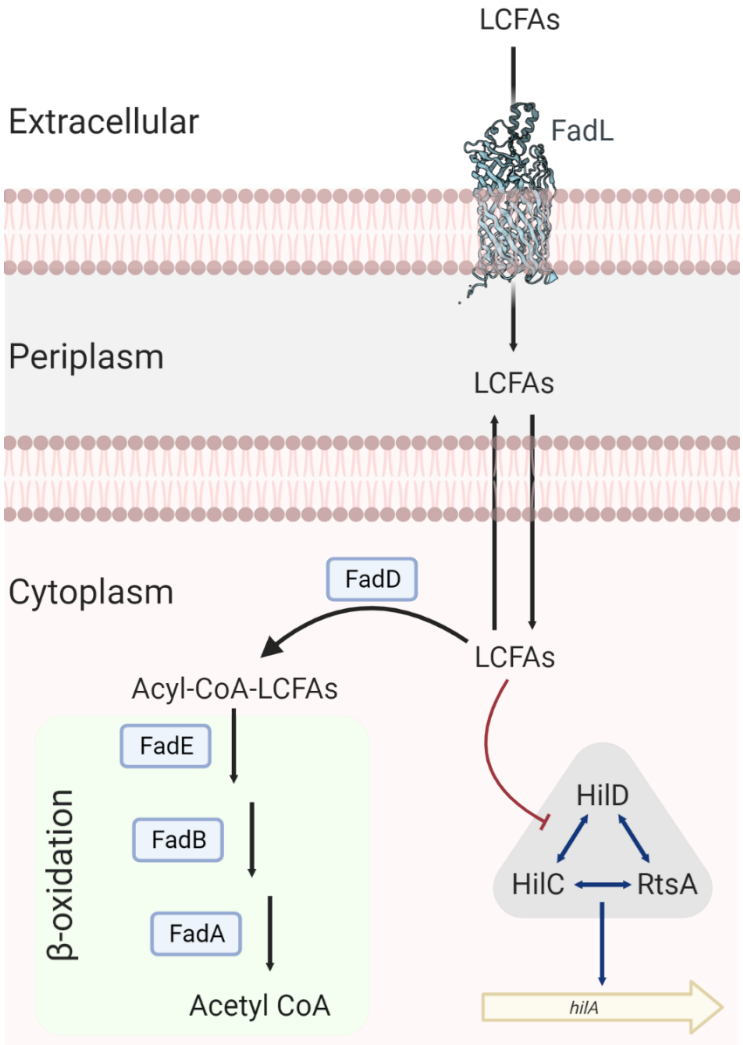


Figure 4.1. Uptake and metabolism of LCFAs in *Salmonella*. LCFAs are imported into *Salmonella* cells by FadL, where they are metabolized via the β-oxidation pathway for use in cell metabolism. Free LCFAs are also able to repress the expression of *hilA* by inhibiting the transcription regulators HilD, HilC and RtsA.

HilD has been shown to be capable of binding to a range of different LCFAs. These include both saturated (myristate and palmitate) and unsaturated (oleate) LCFAs, 14-18 carbon atoms in length[56]. Recently, unsaturated fatty acids (UFAs) containing a *cis*- double bond at the C2 position have been shown to repress *hilA* with higher potency than oleic acid[60,61]. Comparison of *cis*-2-UFAs revealed the repression of *hilA* by UFAs with a range of chain lengths, with the C16 UFA *cis*-2-hexadecenoic acid (c2-HDA) found to have the highest affinity. Whilst also active, the corresponding *trans*-isomer is less a potent repressor, highlighting the *cis*-2 double bond as critical in influencing the affinity of these ligands to HilD. These UFAs destabilise HilD, making it more susceptible to degradation by Lon protease, in addition to

preventing DNA-binding. HilC and RtsA have also been shown to bind a number of UFAs, with a similar preference for *cis*-2-UFAs, albeit with lower affinity compared to HilD (HilD, HilC and RtsA bound c2-HDA with respective affinities of 3.5 μM, 20.5 μM and 22.0 μM, as determined by ELISA[39,61]). Unlike HilD, the stability of HilC and RtsA is not affected by *cis*-2-UFAs, which combined with their lower affinities for UFAs suggests that they may play a crucial role in allowing *Salmonella* to adapt its invasion ability in response to changing environmental conditions.

The mechanism of sensing LCFAs to coordinate invasion appears to be well-conserved amongst enteric pathogens. AraC/XylS transcription regulators from a number of different bacteria have been shown to bind LCFAs, notably ToxT (*V. cholerae*), VirF (*Y. enterocolitica*) and Rns (*E. coli*)[59,121,127]. Structures of ToxT and Rns revealed that fatty acids bind to a conserved pocket at the interface of the two protein domains, although the identity of the bound fatty acid and the specific binding mode differs between the two proteins[57,59].

UFAs have been proposed to regulate ToxT activity via an allosteric mechanism that inhibits dimerisation and DNA-binding[58]. When bound to UFAs, ToxT is constrained to an overall closed, rigid structure. Binding of the UFA palmitoleic acid alters the position, length, and rigidity of the key dimerisation helix of ToxT, confining it to a conformation that prevents dimerisation. Additionally, the recognition helix of the first HTH motif is oriented perpendicular to the corresponding recognition helix of second HTH, preventing it from fitting into the major groove of the DNA. Increased flexibility of ToxT in the apo form allows these helices to adopt conformations that enable dimerisation and subsequent DNA-binding. Although structural details of the fatty acid regulatory mechanism of Rns have not been clarified, the saturated fatty acid decanoic acid was observed to increase thermal stability in a dose-response manner[59]. Along with the conserved binding pocket, this indicates a common mechanism for the regulation of AraC/XylS proteins by fatty acids.

Although several LCFAs have been shown to downregulate *hilA*, with the triad of HilD, HilC and RtsA identified as the molecular targets, until recently there was a lack of biochemical data describing the underlying mechanism. This limited further studies into the potential of these small-molecule regulators as anti-virulence lead compounds targeting the regulators of SPI-1.

Here, we optimised multiple *in vitro* assays to probe the direct binding of small molecules to HilD, HilC and RtsA. We performed a wider and more-comprehensive study of different LCFAs, to identify which chemical properties of these ligands are critical for binding to HilD. To determine the mechanism by which LCFAs bind to and affect HilD activity, we attempted to obtain the structure of HilD bound to these ligands. HDX-MS confirmed that LCFAs bind to a conserved binding pocket within HilD, inhibiting both HilD dimerisation and DNA-binding. A combination of computational and *in vitro* approaches revealed putative binding modes of LCFAs, highlighting specific binding of select LCFAs to HilD. These results support the hypothesis that LCFAs regulate HilD via a conserved mechanism of allosteric inhibition to regulate virulence gene expression.

4.2 Oleic Acid Binding to HilD and HilC

We first confirmed previously published results showing that oleic acid binds directly to HilD and prevents it from binding to target DNA (Figure 4.2 A). HilD or HilC were incubated with the same labelled fragment of the *hilA* promoter previously used to confirm DNA-binding activity *in vitro* (A1_62bp, see Figure 3.10), along with increasing concentrations of oleic acid. HilD and HilC were kept at a constant concentration of 600nM, the highest protein concentration where no aggregates were observed in EMSAs. As shown previously[56], both HilD and HilC are able to bind oleic acid, which subsequently prevents them from binding to target DNA. Oleic acid appears to bind to both HilD and HilC with comparable affinities, although the band corresponding to the HilD-DNA complex is less intense than that of the HilC-DNA complex at equivalent oleic acid concentrations. To determine the affinity of oleic acid to HilD, we performed MST measurements using the N-terminal GFP-fusion of HilD, yielding a K_d of 48 ± 5.27 µM (Figure 4.2 B). This K_d is significantly higher (i.e. lower affinity) than expected when compared to the results obtained from EMSAs. As these gels can be hard to quantify, we sought an orthogonal approach to quantify the concentrations of oleic acid required to inhibit HilD binding to DNA. We performed competitive MST measurements using the same labelled DNA fragment used for EMSAs (Figure 4.2 C-D). This set-up also allowed us to quantitatively compare the binding of oleic acid to HilC and HilD; issues with protein degradation during the purification of a GFP-HilC fusion prevented affinity measurements for

the direct binding of oleic acid to HilC. Comparable EC_{50} values were obtained for both HilD and HilC (9.91 ± 1.75 µM and 7.40 ± 0.69 µM, respectively). The calculated affinity for oleic acid binding to HilD is notably higher than that obtained from the GFP-HilD binding assay, and more aligned with those observed-by-eye from EMSAs. The reason for this discrepancy is unclear, although the trends in affinity observed for the binding of different lipids to HilD are the same in both MST assays (discussed in section 4.4).

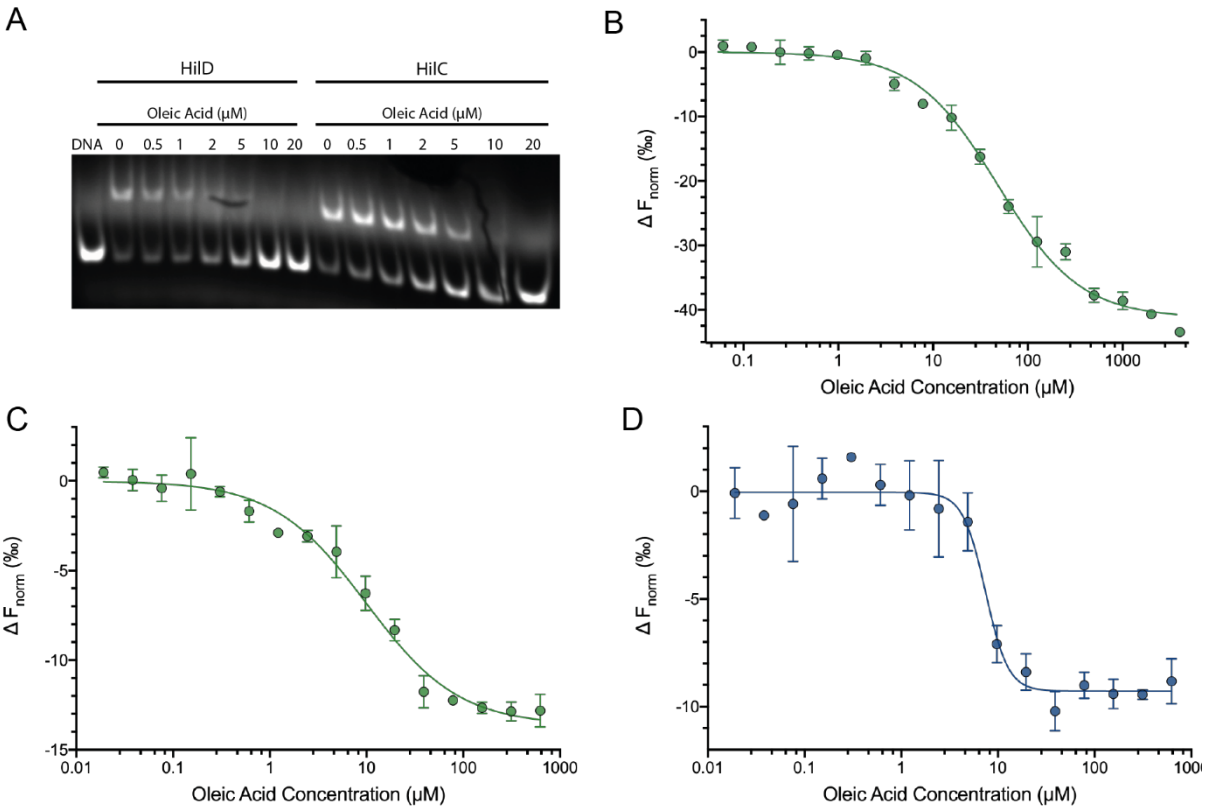


Figure 4.2. Oleic acid binds directly to both HilD and HilC, preventing DNA-binding. (**A**) EMSA showing the effect of oleic acid on the binding of HilD and HilC to the *hilA* promoter. All lanes contain 50 nM DNA and, where indicated, 600 nM of the respective protein. Oleic acid was added to some samples at the specified concentration, with a final ethanol concentration of 1%. (**B**) MST binding curve for oleic acid binding to HilD. 50nM GFP-HilD was incubated with increasing concentrations of oleic acid, and dose-response curves plotted from the change in thermophoresis at an MST on-time of 1.5 seconds. DMSO concentration was kept at a constant final assay concentration of 1%. (**C-D**) Competitive MST measurements quantifying the inhibition of DNA-binding activity of (C) HilD and (D) HilC. HilD or HilC were incubated with 50 nM Cy5-labelled DNA and increasing concentrations of oleic acid. As for EMSAs, labelled DNA and HilD/HilC were kept at constant concentrations of 50 and 600nM, respectively. Affinities in all MST assays were calculated using the NanoTemper MO.Affinity Analysis software, by fitting the data from three repeat experiments.

4.3 Effect of LCFAs on HilD Dimerisation

LCFAs have been shown to disrupt the dimerisation of ToxT, preventing ToxT from binding to target DNA[58,128]. As LCFAs have also been shown to prevent HilD from binding to target DNA, we sought to determine whether this is due to disruption of the HilD dimer, in a similar manner as reported for ToxT.

Due to issues with protein precipitation when lipids were added to HilD at high concentrations, we were unable to confirm the effect of these compounds on HilD dimerisation using SEC-MALS, and hence sought an alternative approach to study the effect of LCFAs on HilD dimerisation. The MST assay previously used to quantify the homodimerisation of HilD (section 3.6) was repeated, in which the GFP-HilD fusion protein was pre-incubated with 100 µM oleic acid (> 2-fold above the previously determined K_d) or 1% DMSO as a control. Incubation with DMSO had no significant effect on the dimerisation of HilD ($K_{d,dimer}$ of 4.52 ± 0.50 µM, compared to 3.39 ± 0.77 µM calculated previously). In contrast, oleic acid completely abolished the formation of heterodimers between GFP-HilD and unfused HilD (Figure 4.3). Therefore, we can conclude that the binding of oleic acid inhibits the homodimerisation of HilD.

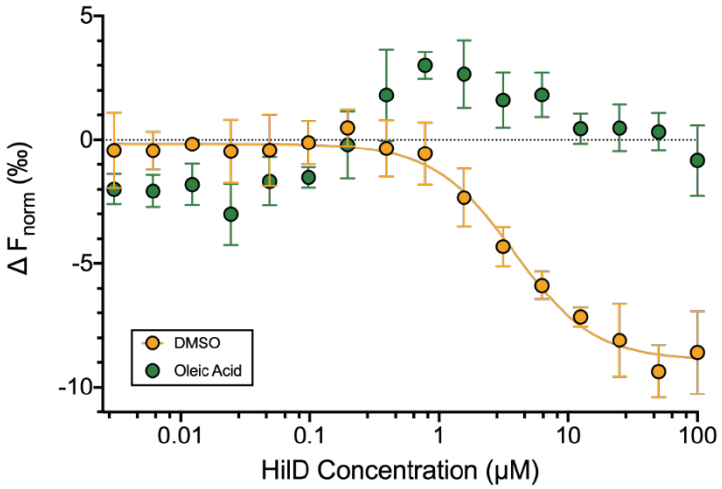


Figure 4.3. Oleic acid disrupts both dimerisation and DNA-binding of HilD. HilD was titrated against a GFP-HilD fusion (50 nM constant concentration), which was incubated with either 1% DMSO (yellow) or 100 µM oleic acid (green). Experiments were carried out in triplicate and the K_d was determined from changes in thermophoresis at an MST on-time of 1.5 seconds.

4.4 Structure-Activity Relationship of LCFAs

As HilD has been shown to bind to several different LCFAs, we performed a structure-activity relationship (SAR) analysis to gain a more comprehensive understanding of the features of these lipids that are critical for binding to HilD. Chain length, saturation and bond geometry have all been previously shown to influence the repressive properties of LCFAs, with respect to HilD binding[60]. Here, LCFAs with specific chemical properties were compared for their ability to bind to HilD.

Binding was investigated using the assays optimized for the binding of oleic acid to HilD. EMSAs were used to check the binding of LCFAs to HilD (and inhibition of HilD DNA-binding activity) and provide an overview of different affinities of these different lipids. Binding affinities were determined using MST, by measuring direct binding to GFP-HilD and, for selected LCFAs, the disruption of the HilD-DNA complex using the competitive MST assay described in section 4.2.

Firstly, we compared the effect of varying the fatty acid chain length on binding affinity for *cis*-9-unsaturated LCFAs with chain lengths of 14-20 carbons (Figure 4.4 A-B). These LCFAs all contain a *cis*- double bond at the 9th carbon from the carboxylic acid head group, with an increasing number of carbon atoms at the saturated end of the fatty acid. LCFAs with chain lengths of 16-20 carbons have affinities in the range of 25-60 µM, with increasing affinity correlating with chain length (K_d values shown in Table 4.1). Oleic Acid (C18) inhibited HilD from binding to a fragment of the *hilA* promoter at all concentrations tested, whilst palmitoleic acid (C16) did so at 50 µM. Myristoleic acid (C14) shows only very weak binding (K_d: 567 µM) and no inhibition of DNA-binding in EMSAs, indicating 16 carbon atoms is the minimum chain length that is required for binding to HilD. As was the case for oleic acid, EC_{50} values of palmitoleic and gadoleic acids calculated from the competitive assay (14.7 and 8.98 µM, respectively) are notably higher affinity than those derived from the direct MST assay, although the same trend of affinity increasing with chain length is observed.

Oleic acid is also an omega-9 fatty acid, defined by the position of the carbon-carbon double bond with respect to the saturated end of the fatty acid (position 9). We compared the affinities for three different omega-9-fatty acids, with an increasing chain length between the

cis- double bond and the carboxylic acid head group (Figure 4.4 C-D). The same overall trend is observed as for *cis*-9-unsaturated LCFAs; increasing chain length results in higher affinity binding to HilD. Whilst erucic acid (C22) showed notably higher affinity (14.71 ± 2.19 μM) to HilD compared to oleic acid (C18), nervonic acid (C24, 17.99 ± 1.84 μM) showed no further increase in affinity. This suggests that chain length of 22 carbons is the maximum length for optimal binding to HilD, and LCFAs of longer length may be protruding from the binding pocket. These combined results imply that the position of this central double bond is not a critical requirement for binding. This is evidenced by the higher affinity of *cis*-2-unsaturated LCFAs, which lack a double bond in the middle of the hydrocarbon chain, for HilD compared to oleic acid[60].

To further test this hypothesis, we investigated the presence and stoichiometry of this double bond. EMSAs showed that elaidic acid, the *trans*-isomer of oleic acid, still inhibited the DNA-binding of HilD, yet interestingly MST measurements did not show any binding for this LCFA (Figure 4.5 A-B). However, binding was observed in the competitive MST assay (Figure 4.5 C), with similar affinity (EC_{50}: 10.09 ± 1.97 μM) as for oleic acid (9.91 ± 1.75 μM). The saturated LCFA palmitic acid (C16) also displayed binding in assays for the inhibition of DNA binding (EMSA, competitive MST), yet no clear binding to GFP-HilD (Figure 4.5 D-F). In the former two assays, binding of palmitic acid was comparable to that of palmitoleic acid (the corresponding *cis*-9-unsaturated LCFA of the same chain length). The low solubility of palmitic acid at higher concentrations in aqueous buffers is a limiting factor for affinity measurements and may explain the lack of observed binding in direct MST measurements. Both palmitate and the 14-carbon saturated LCFA myristate were both previously shown to decrease *hilA* transcription *in vivo*, however further studies were also limited by the low solubility of saturated LCFAs[56].

Taken together these results suggest that whilst the double bond of these LCFAs is not critical for ligand binding, its position and confirmation may affect the binding affinity: a *trans*-conformation at the C9 position may impair binding, whilst lipids in which the double bond is instead in the C2 position have been shown to have markedly higher affinity to HilD[60,61].

The C16 UFA c2-HDA was reported to bind to HilD with higher affinity than the corresponding *trans*-isomer[61]. However, additional runs comparing the affinities of other LCFAs to HilD were not performed, and so we compared the binding of *cis*-2-unsaturated fatty acids with

corresponding *cis*-9-unsaturated fatty acids of the same chain length (Figure 4.6). HilD was inhibited from binding to the *hilA* promoter by similar concentrations of c2-HDA and palmitoleic acid. A competitive MST performed for the C20 UFAs *cis*-2-eicosenoic acid and gadoleic acid yielded almost identical EC_{50} values for the two UFAs (c2-eicosenoic acid: 8.89 ± 3.64 µM; gadoleic acid: 8.98 ± 1.30 µM).

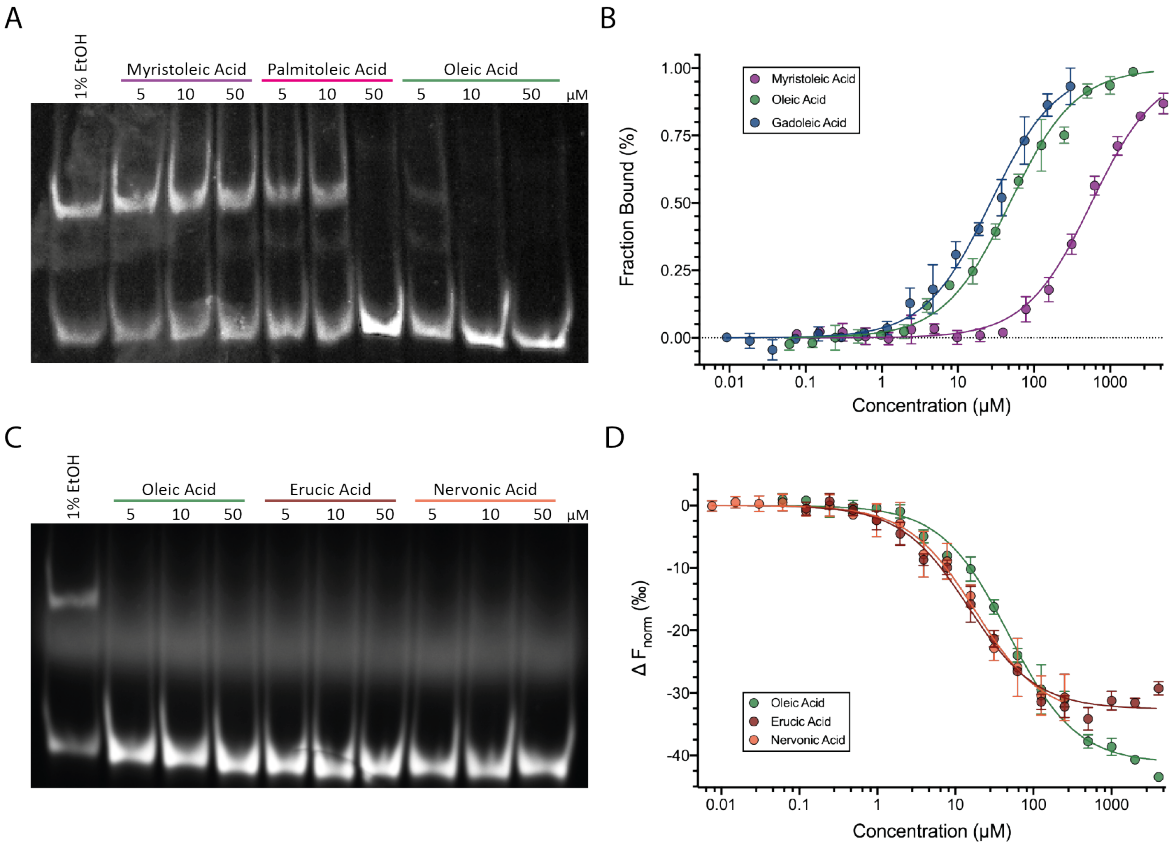


Figure 4.4. Long chain fatty acids with increasing chain length bind to HilD with higher affinities. (**A**) EMSA showing the inhibition of HilD DNA-binding by *cis*-9-UFAs. HilD (600nM) was incubated with DNA (50nM) and fatty acids at the indicated concentrations. Here, a non-labelled form of the A1_62bp DNA fragment was used, and the gel stained using SYBR gold. (**B**) MST binding curves for the binding of *cis*-9-UFAs to GFP-HilD (50nM). The fraction of GFP-HilD in the ligand-bound state is plotted against ligand concentration. (**C**) Omega-9 fatty acids completely abolished HilD-DNA binding at all concentrations tested. HilD was incubated with labelled DNA and the indicated concentration of each LCFA. (**D**) MST binding curves for the binding of omega-9 fatty acids to GFP-HilD. K_d values were calculated from changes in normalised fluorescence (ΔF_{norm}) at an MST on-time of 1.5 seconds with increasing ligand concentrations. K_d values from MST measurements are listed in Table 4.1.

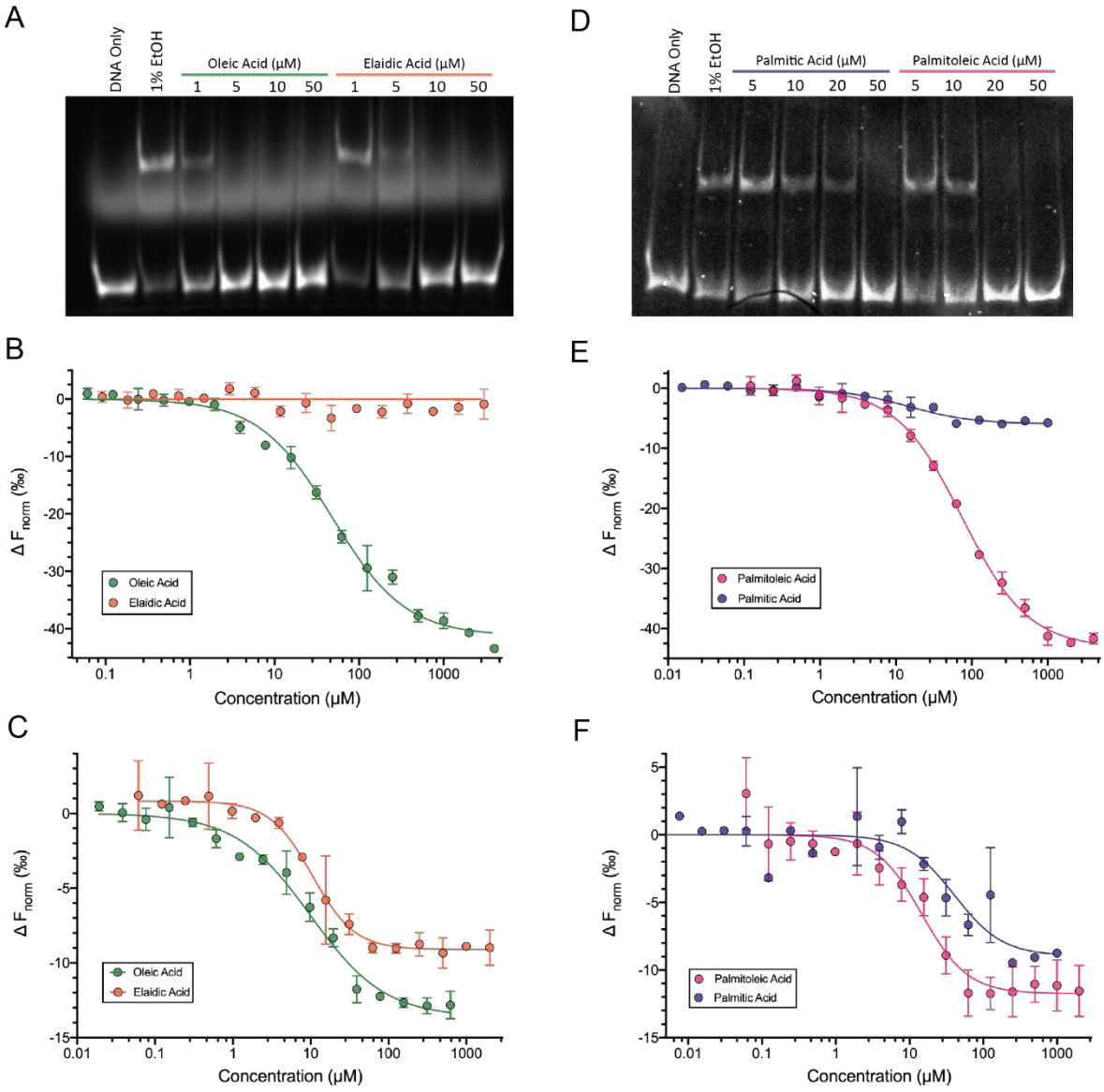


Figure 4.5. The *cis*-9 double bond is not crucial for LCFA binding to HilD. (**A**) EMSA comparing the binding of the *cis*-9 UFA oleic acid with its *trans*-isomer, elaidic acid. (**B-C**) Affinities of oleic and elaidic acid to HilD were calculated for (B) direct LCFA binding to GFP-HilD or (C) using a competitive MST assay. (**D**) The C16 LCFAs palmitic acid (saturated) and palmitoleic (*cis*-9-unsaturated) both inhibited HilD DNA binding in EMSAs. (**E-F**) MST assays quantifying (E) the direct binding of palmitic and palmitoleic acid to GFP-HilD and (F) the inhibition of HilD DNA-binding activity by these LCFAs. EMSAs and competitive MST runs were performed with 50nM labelled DNA, 600nM HilD, and varying concentrations of the respective LCFA (with a final assay concentration of 1% ethanol). Binding constants are displayed in Tables 4.1 and 4.2.

Although *cis*-2-unsaturated fatty acids have been shown to be potent inhibitors of *hilA in vivo*, this effect appears to be less pronounced *in vitro* and is fitting with our other results showing that the position of the double bond is not critical for binding. Whilst LCFAs with a *trans*-2 double bond are reported to bind to HilD with lower affinity than the corresponding *cis*-isomers, we found those lacking a double bond in this position entirely to display similar affinities to *cis*-2-unsaturated LCFAs. LCFAs lacking a double bond at the C2 position would have increased rotation flexibility, enabling them to adopt a conformation mimicking the *cis*-2 double bond. On the other hand, *trans*-2-UFAs are constrained to a single conformation. This restricted conformation may be less favourable for binding to HilD, resulting in the reported lower affinity of these fatty acids.

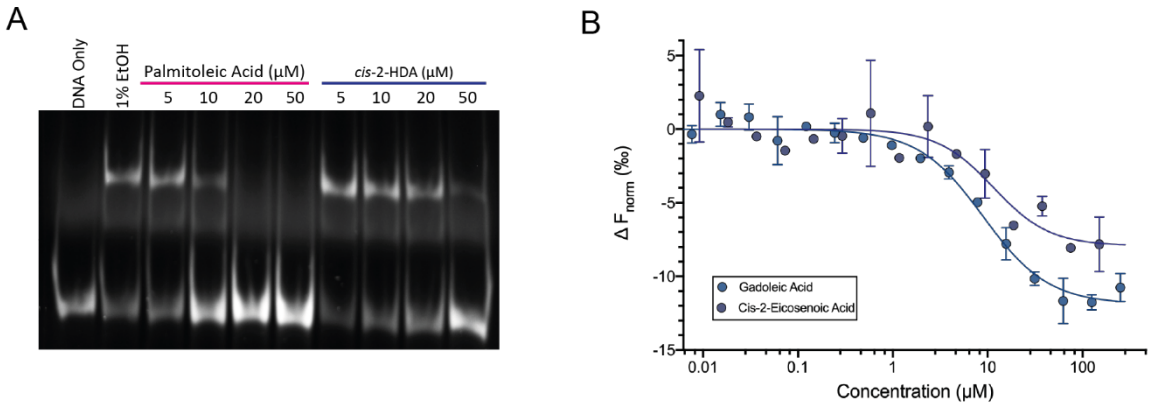


Figure 4.6. *cis*-2-unsaturated fatty acids display comparable affinities with *cis*-9-unsaturated fatty acids of the same chain length. (**A**) EMSA showing inhibition of HilD binding to the *hilA* promoter by c2-HDA or palmitoleic acid. (**B**) Competitive MST measurements, using labelled DNA, yielded comparable binding curves for c2-eicosenoic acid and gadoleic acid. Calculated affinities are shown in Table 4.2.

Our results showing that affinity increases with increasing chain length suggest that the binding of LCFAs is primarily driven by non-specific hydrophobic interactions between the fatty acid chain and the hydrophobic binding pocket of HilD. To investigate whether specific interactions involving the carboxylic acid head group also contribute to LCFA binding (as observed for ToxT and Rns), we compared the binding of oleic and palmitoleic acid with their corresponding methyl esters. Methylation of the carboxyl group was previously shown to reduce the potency of *cis*-2-UFAs in repressing *hilA* expression through HilD[60]. Unlike the corresponding fatty acids, methyl oleate and methyl palmitoleate did not show any binding to HilD (MST), nor did they affect the binding of HilD to DNA (Figure 4.7).

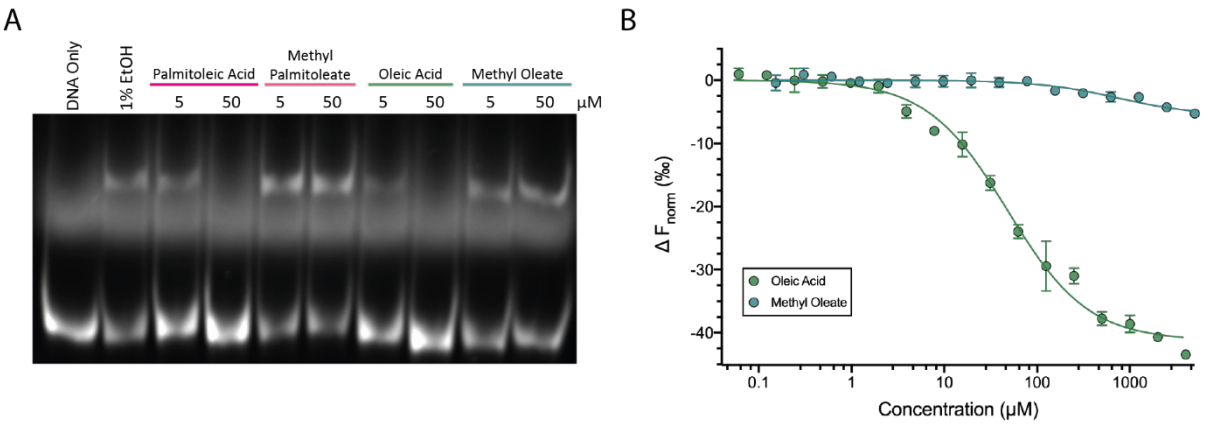


Figure 4.7. Methylation of the carboxyl head group prevents the binding of LCFAs to HilD. (**A**) Increasing concentrations of oleic and palmitoleic acids inhibit HilD from binding to the *hilA* promoter. A band corresponding to the HilD-DNA complex is still observed at the highest tested concentrations of methyl oleate and methyl palmitoleate, indicating a lack of binding to HilD. (**B**) MST binding curves for the binding of oleic acid and methyl oleate to GFP-HilD. K_d values (including those calculated for palmitoleic acid and methyl palmitoleate) were calculated from multiple replicates, as reported in Table 4.1.

In summary, the hydrophobic effect appears to dominate the interaction of LCFAs with HilD, with higher affinity correlated with increasing chain length, up to a maximal length 22 carbons. The existence of a double bond in the centre of the hydrocarbon chain is not critical, however, a double bond that constrains the LCFA in an unfavourable conformation (such as a *trans*- double bond at positions C2 or C9) may be detrimental to binding. Additional specific interactions with the carboxyl head group also exist, which are crucial for binding.

Table 4.1. Calculated affinity values for the binding of LCFAs to HilD, determined by MST. Affinity values were calculated using the K_d Fit model, using the NanoTemper MO.Affinity Analysis software. An MST on-time of 1.5 seconds was used to calculate changes in thermophoresis.

Lipid	Shorthand	K_d (µM)	±	Response Amplitude	No. repeats
Myristoleic Acid	9Z-14:1	566.70	59.46	53.96	4
Palmitoleic Acid	9Z-16:1	58.53	4.86	41.48	4
Oleic Acid	9Z-18:1	48.06	5.27	41.24	3
Gadoleic Acid	9Z-20:1	26.68	3.13	31.10	4
Erucic Acid	13Z-22:1	14.71	2.19	32.60	3
Nervonic Acid	15Z-24:1	17.99	1.84	33.80	3
Elaidic acid	9E-18:1	-	-	-	3
Palmitic Acid	16:0	14.24	3.90	5.96	2
Methyl-palmitoleate	9Z-16:1	126.45	67.4	5.01	2
Methyl-oleate	9Z-18:1	738.54	276	5.61	3

Table 4.2. EC_{50} values for the inhibition of HilD-DNA binding activity by LCFAs, determined from a competitive MST assay. Affinity values were calculated using the Hill Fit model, using the NanoTemper MO.Affinity Analysis software. An MST on-time of 2.5 seconds was used to calculate changes in thermophoresis.

Lipid	Shorthand	EC_{50} (μM)	±	Response Amplitude	No. repeats
Palmitoleic Acid	9Z-16:1	14.71	3.30	11.75	3
Oleic Acid	9Z-18:1	9.91	1.75	13.64	3
Gadoleic Acid	9Z-20:1	8.98	1.30	11.88	3
Erucic Acid	13Z-22:1	7.36	1.15	9.48	3
Elaidic Acid	9E-18:1	10.09	1.97	7.74	3
Palmitic Acid	16:0	31.12	6.36	5.37	2
cis-2-eicosenoic Acid	2Z-20:1	8.89	3.64	6.45	2

4.5 Binding Mode of LCFAs to HilD

4.5.1 Hydrogen-Deuterium Exchange

Crystal structures of the AraC/XylS proteins *V. cholerae* ToxT and *E. coli* Rns have been published, featuring LCFAs bound to a common binding pocket. Given the overall fold of HilD is predicted to be the same as that of these two regulators, it is reasonable to assume that LCFAs bind to the equivalent pocket of HilD.

Structural characterisation attempts were limited by low solubility of LCFAs in aqueous buffers and precipitation of HilD upon incubation with LCFAs. Therefore, we utilised HDX-MS to experimentally validate the LCFA binding pocket in HilD and possible confirmation changes in the overall protein structure upon ligand binding. HDX-MS experiments were performed for both apo HilD and in the presence of oleic acid, and the differences in deuterium (D) uptake compared for both samples. Regions showing decreased D-uptake in the presence of a ligand are indicative of a binding site, with the bound ligand shielding these residues from HDX with the bulk solvent.

Multiple regions of HilD are shown to have lower D-uptake in the presence of oleic acid, with decreased HDX observed across the entire DBD. This protection is most pronounced for the C-terminal end of helix α4 and helices α5-7, which form HTH2. Reduced HDX is also apparent at residues T41-F50, which form beta strand β1 of the NTD and are in close proximity to HTH2. Decreased D-uptake clustered in this area signify this to be the binding site of oleic acid (Figure

4.8 A). No detectable changes in HDX are observed for the other strands forming the barrel structure of the NTD. These residues are shielded from the bulk solvent within the protein core, resulting in very low HDX in both samples that makes detecting differences between them impossible. These results are consistent with the predicted binding pocket based on structural alignment to ToxT and Rns.

Decreased HDX is also observed at helices α1 and α3 of the DBD, which form HTH1. These helices are remote from the ligand binding site and changes in HDX are likely due to secondary effects. Binding of palmitoleic acid to ToxT reduces protein flexibility, whilst the melting temperature of Rns increases in a dose-response manner upon binding of decanoic acid. Thermal stabilisation of HilD was also observed upon incubation with oleic acid (Figure 4.8 B), with calculated HilD melting temperatures of 49.2°C and 51.2°C in the absence and presence of oleic acid, respectively. Although further runs with higher concentrations of oleic acid and other LCFAs were again limited by ligand solubility, this stabilisation effect supports the hypothesis that a conserved dynamic allosteric mechanism, as demonstrated for ToxT, also underpins HilD regulation by LCFAs. Decreased flexibility of HilD upon oleic acid binding could prevent independent rotation of the two HTH motifs with respect to one another, rationalising the changes in HDX behaviour at HTH1, in addition to HTH2, upon the binding of oleic acid.

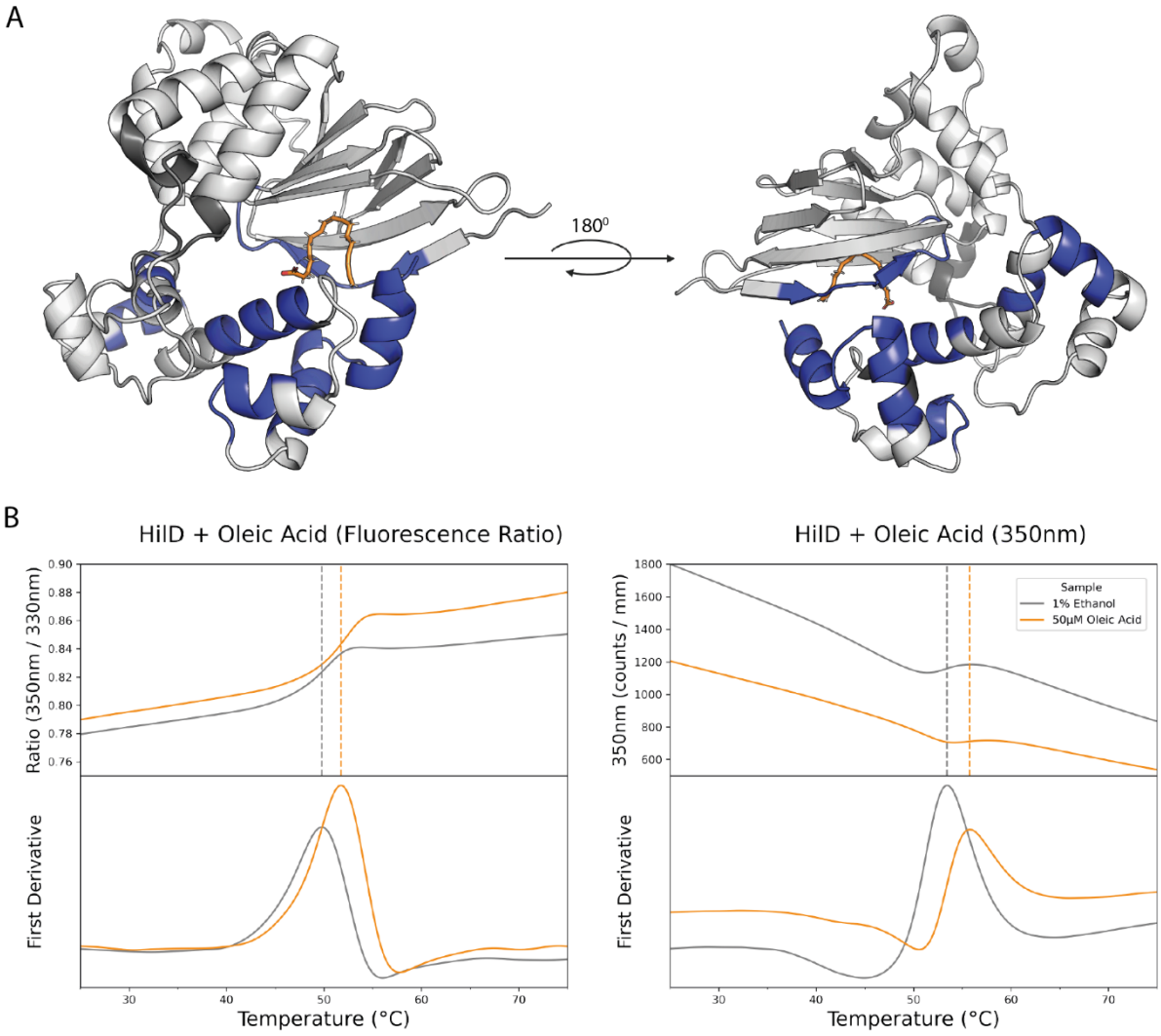


Figure: 4.8. Oleic acid binds to a conserved binding pocket in HilD. (**A**) Oleic acid (shown in orange) was docked into the identified binding pocket of HilD (AlphaFold model). HilD residues are coloured according to changes in HDX upon incubation with oleic acid: White, no change in HDX; blue, decreased HDX; black, no information could be obtained due to a lack of peptide coverage for these residues. (**B**) Oleic acid stabilises HilD. HilD (20 µM) was incubated with 50 µM oleic acid (or 1% ethanol as a control). Protein unfolding was monitored by measuring intrinsic fluorescence at 330 and 350 nm. Unfolding profiles are shown for fluorescence at 350 nm (left) and for the ratio of the two wavelengths (right). In both traces, an increase in melting temperature of HilD upon oleic acid binding is observed.

4.5.2 Computational Docking

Although we identified the LCFA binding pocket of HilD, specific binding residues could not be determined using HDX-MS. We predicted the binding mode of LCFAs to HilD using a computational approach. Computational docking was carried out by Thales Kronenberger (Universitätsklinikum Tübingen) using Glide from the Schrödinger suite. Oleic acid, myristoleic acid and palmitic acid were each docked into the identified binding pocket of the AlphaFold model of HilD.

All three ligands bound to HilD in a bent conformation and formed identical specific interactions with HilD via the carboxyl head group. Specific interactions are formed with residues K264 and R267, which are situated on the DBD at the C-terminal end of helix α4, consistent with our HDX data showing increased protection in this region upon oleic acid binding. Water-mediated interactions are also predicted between the carboxyl group and residue E102. Although no differences in HDX are observed around this residue, it lies within a region exhibiting low D-uptake in all states so no experimental validation can be gained here.

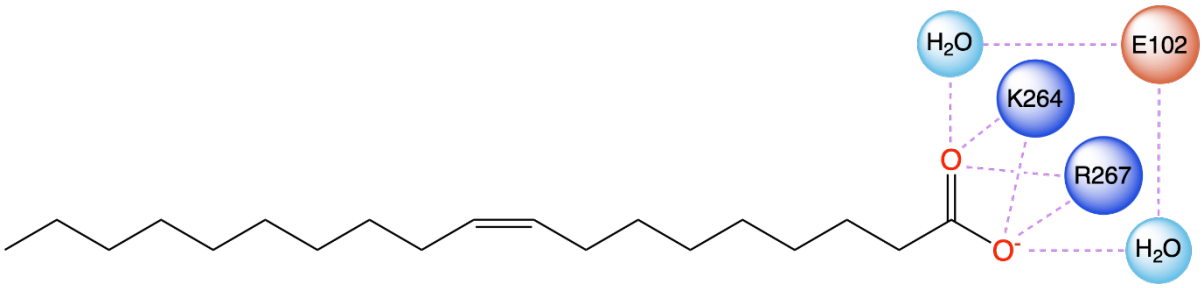
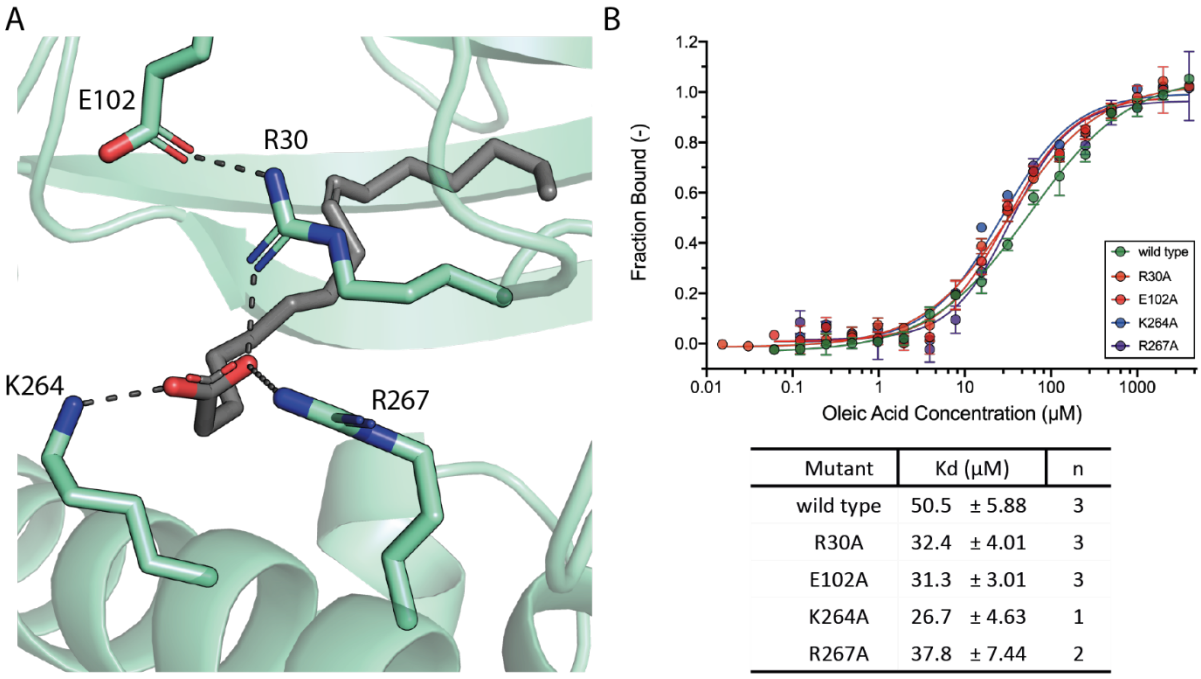


Figure 4.9. Schematic showing specific oleic acid atom interactions with residues of HilD. Displayed interactions occurred in more than 30% of the simulation time in the selected trajectory (0.00-1004.00 nsec). HilD residues are coloured according to their charge (positive: blue; negative: orange).

4.5.3 Point Mutants

Prior to the availability of AlphaFold, we used a homology model based on the crystal structure of ToxT (PDB: 4MLO) to model HilD and ligand binding. Residues R30, E102, K264 and R267 were identified from this model as forming potential interactions with the carboxyl

head group of oleic acid (Figure 4.10 A). Single point mutants to alanine were cloned into the GFP-HilD construct to compare their affinities for oleic acid with the wild type protein. All of these mutants were found to bind oleic acid with almost identical affinities to wild type HilD (Figure 4.10 B).



Mutant	Kd (µM)	n
wild type	50.5 ± 5.88	3
R30A	32.4 ± 4.01	3
E102A	31.3 ± 3.01	3
K264A	26.7 ± 4.63	1
R267A	37.8 ± 7.44	2

Figure 4.10. Selected single point mutants did not affect binding affinity of oleic acid. (**A**) Homology model of HilD (based on the structure of ToxT, 4MLO) with oleic acid docked into the predicted binding pocket. Predicted hydrogen bonds are shown as dashed lines. (**B**) MST binding curves comparing the binding of oleic acid to selected single point mutants of HilD. Calculated affinities are shown in the corresponding table along with the number of replicate MST runs for each mutant.

This result is unsurprising for the R30A mutant, as both the AlphaFold model and HDX results show this residue is located within the unstructured N terminus. However, our latest docked models did predict E102 (water-mediated interaction), K264 and R267 to be involved in specific interactions with the carboxyl head group.

We previously found that the binding of LCFAs to HilD appears to be dominated by the hydrophobic effect (section 4.4). If these residues equally contribute to interactions with the carboxyl head group as predicted, then it is unsurprising that individual point mutations do not show noticeable change in affinity. Further work, involving double and triple point mutants of these identified residues (E102, K264 and R267) is required to confirm the binding mode predicted by computational model.

4.6 Discussion

The regulation of AraC/XylS transcription factors by long chain fatty acids has been demonstrated for several members of this protein family. Crystal structures have been published for the regulators ToxT and Rns in complex with fatty acids. We used a combination of HDX-MS and computational methods to confirm that LCFAs bind to a pocket at the interface of the two domains of HilD, analogous to the fatty acid binding pocket in both ToxT and Rns. In our computational model, the fatty acids are bound within the cupin barrel of the HilD NTD and form specific interactions with residues K264 and R267 of the DBD (Figure 4.11 A). This is comparable to the binding of palmitoleic acid to ToxT, in which residue K230, also located on the DBD helix α4, forms specific interactions with the fatty acid carboxyl head group (Figure 4.11 B). In contrast, in Rns the residues H20 and R75 that form specific interactions with decanoic acid are both located on the NTD (Figure 4.11 C). In all three proteins this binding pocket is predominately hydrophobic, accommodating the hydrocarbon chain of the bound fatty acids. This supports our findings that the binding of LCFAs to HilD is largely non-specific and that HilD, like ToxT, can bind a range of different fatty acids.

Our results also strengthen the hypothesis that regulation of AraC/XylS transcription factors by fatty acids occurs via a common dynamic allosteric mechanism. The binding of oleic acid increases the thermal stability of HilD and predicted interactions of the bound fatty acid with residues on both domains constrain HilD to a more stable, closed confirmation, which disrupts dimerisation and prevents binding to target DNA.

A recent study indicated that different lipids (those containing or lacking a *cis*-2 double bond) may additionally bind to HilD with different binding modes[61]. In our latest predicted model of HilD, several of these residues are distant from the determined pocket, making it unlikely that they are directly involved in fatty acid binding. However, these residues do show decreased deuterium-uptake in HDX-MS measurements upon incubation with oleic acid, making it plausible that mutation of these residues effects the overall protein structure and flexibility, indirectly impacting fatty acid binding to HilD. Further work, including high-resolution structural data, is required to verify these findings.

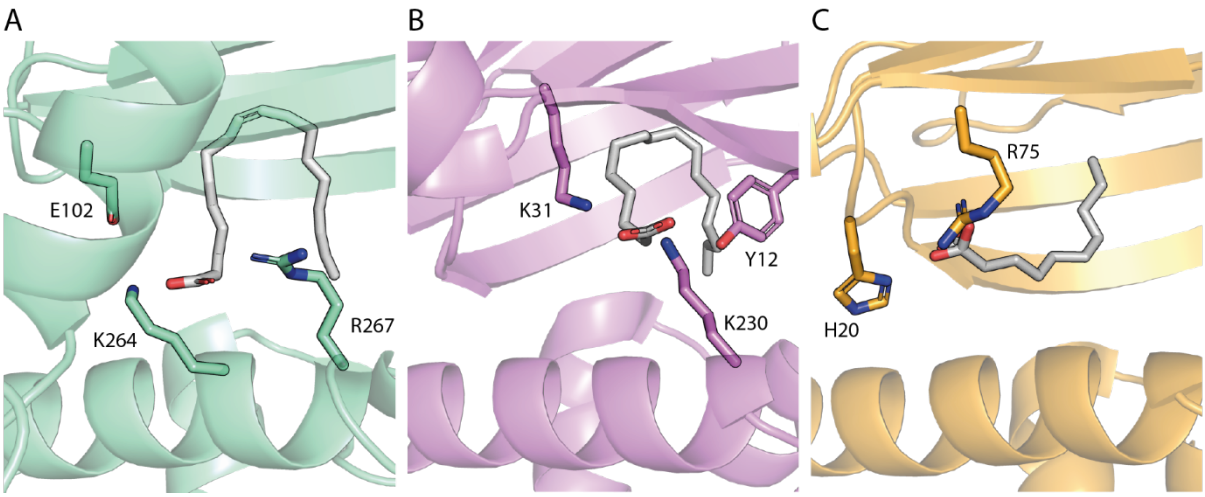


Figure 4.11. Comparison of the binding of fatty acids to AraC/XylS transcription regulators. (**A**) AlphaFold model of HilD, with oleic acid computationally docked into the binding pocket, is shown in green. (**B**) ToxT (PDB: 4MLO) shown in purple, in complex with palmitoleic acid. (**C**) Native Rns (PDB: 6XIU) coloured in orange, bound to decanoic acid. Protein structures were aligned and figures prepared using PyMOL. Side chains of residues that interact directly with the carboxyl group of bound fatty acids are shown as sticks and, along with fatty acids, are coloured by chemical element.

HilD can accommodate the binding of a large range of different fatty acids. The variability of these ligands presents the opportunity for the further optimisation of increasingly potent inhibitors of *Salmonella* virulence. LCFAs have additionally been shown to inhibit both HilC and RtsA activity, and an inhibitor targeting all three regulators simultaneously would be less susceptible to the development of resistance. Our work confirms that the carboxyl head group is vital for binding to HilD, as reported previously[60], as is a chain length of at least 16 C-atoms. Fatty acid saturation and position of any double bond in the hydrocarbon chain are not fundamental for binding, although these properties can be optimised for increased ligand solubility and affinity to HilD. *Cis*-2-hexadenoic acid is the most potent naturally occurring UFA against HilD identified so far[61], providing a scaffold for the development of compounds with increased potency.

Our results suggest that LCFAs bind to HilD in a bent conformation, as observed for fatty acid binding to ToxT. Fatty acid mimetics meeting the binding requirements highlighted above, and constrained to the preferential binding confirmation, present an attractive starting point for the optimisation of HilD inhibitors. Such an approach has previously been demonstrated

for ToxT, yielding binders with 10-fold higher affinity than any naturally-occurring fatty acids[82,83].

Knowledge of the mechanisms by which fatty acids inhibit HilD activity also serves as a reference when investigating the binding of novel antipathogenic compounds structurally unrelated to fatty acids, such as the identified small molecule C26. Using the information that we gained from performing a structure-activity relationship for fatty acid binding to HilD, we can design and optimise probes for use in further assays comparing the binding of different ligands to HilD.

Chapter 5. HilE is a Specific Protein Regulator of HilD

5.1 Introduction

HilE is one of the most important negative regulators of *hilA* and SPI-1 expression, specifically repressing HilD via a protein-protein interaction[43–45]. Like *rtsA*, the *hilE* gene is located outside the five SPIs, on a *Salmonella*-specific insert in the chromosome likely acquired by horizontal gene transfer[43,129]. Expression of *hilE* is in turn modulated by many other regulators, providing additional inputs into the control of HilD. FimZ and LeuO both induce *hilE* expression, whilst the transcription regulator Mlc and the small RNA isrM both repress *hilE*[130–133]. The BarA/SirA TCS activates *hilE* via the same mechanism as for *hilD*, forming an incoherent type-1 feedforward loop (I1-FFL)[134]. HilE has also recently been shown to have additional roles in the regulation of SPI-1, being required for the activation of *hilA* expression by acetate and other nutrients[135].

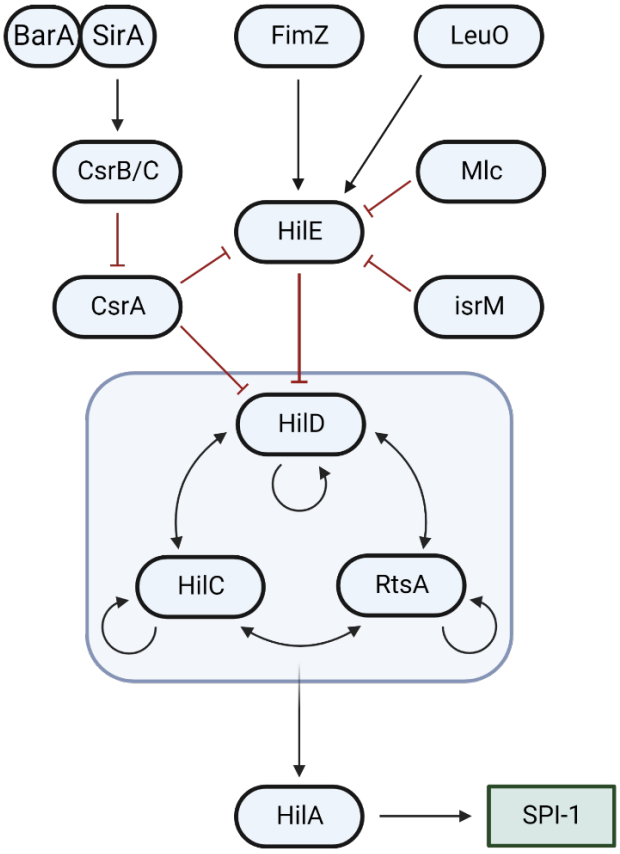


Figure 5.1. HilE is the most important negative regulator of SPI-1. Schematic diagram of the SPI-1 regulatory network, showing the regulators of SPI-1 identified to act via HilE. Black arrows indicate activation and red lines with blunt ends represent repression.

HilE shows high structural similarity to hemolysin-coregulated protein (Hcp), a key structural component of the type VI secretion system (T6SS). The T6SS is reminiscent of an inverted bacteriophage and Hcp proteins are structural homologs of the proteins that form the phage tail tube, with the two systems hypothesised to have evolved from a common ancestor[136–138]. In addition to their structural role, Hcp proteins are also secreted, and act as chaperones to help translocate T6SS effector proteins[139,140]. Hcp proteins form hexameric rings, and it is proposed they are first secreted before assembling into ring structures that stack on top of one another to form the tube through which effector proteins are secreted[141].

As a Hcp-like protein, HilE has been hypothesised to also form hexameric rings. A recent study reported the formation of a large complex with a molecular mass greater than 180 kDa when HilD and HilE were incubated together[44]. This would correspond to complex composed of a HilD dimer bound to two hexameric rings of HilE. However, another study disputed this claim[45], wherein HilE was found to disrupt HilD dimerisation, contradicting the formation of the described 2:12 complex of HilD:HilE.

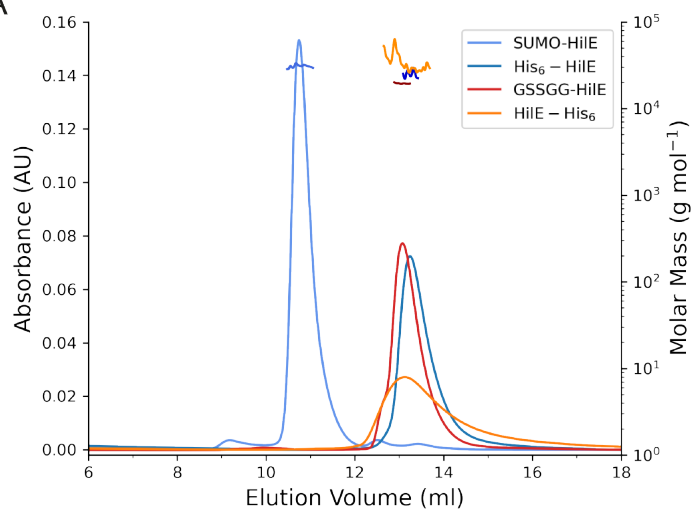
The HilD-HilE interaction has previously been studied using a LexA-based two-hybrid genetic system, so we sought to biochemically validate the existence of this large protein complex by determining its three-dimensional structure. The reported size of this complex made it an attractive target for cryoEM experiments, and hence an alternative approach for determining the structure of HilD, which had thus far proven difficult to uncover. We used a range of biochemical and biophysical approaches to probe this interaction and determine the precise stoichiometry and amino acids mediating the interaction between HilE and HilD. We show that, unlike homologous proteins, HilE does not exist predominantly in a hexameric oligomerisation state. HilE directly disrupts HilD homodimerisation and forms a stable 1:1 complex with HilD, preventing it from binding to DNA.

5.2 HilE Does Not Form Higher Order Oligomers

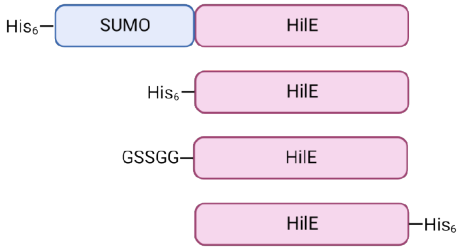
For experimental characterisation, HilE was recombinantly expressed in in *E. coli*. HilE could be successfully expressed and purified to homogeneity with high yields. Gel filtration elution volumes corresponded to the size of a HilE monomer or dimer, rather than an expected

higher-order oligomer. As the reported oligomerisation state of HilE in solution was a hexamer, we cloned and expressed several different constructs to ensure the presence of a His_6-tag at either terminus was not inhibiting ring formation. Purified HilE constructs contained either an N-terminal SUMO-tag, a His_6-tag at either the N- or C- terminus, or an N-terminal His_6-tag followed by a TEV cleavage site for tag removal during purification.

A



C



B

Construct	Molecular Mass (kDa)	
	Theoretical (Monomer)	Measured
SUMO-HilE	28.65	30.4 ± 0.6
His_6-HilE	17.74	25.5 ± 1.8
GSSGG-HilE	16.87	19.4 ± 1.1
HilE-His_6	17.94	46.6 ± 10.8

D

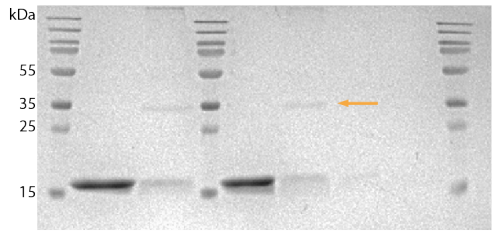


Figure 5.2. SEC-MALS of HilE constructs | (**A**) SEC-MALS elution profile of the four purified HilE constructs. 50 µl of each HilE construct (100 µM) was loaded to a Superdex 75 10/300 increase column. Absorbance at 280 nm was constantly monitored, and molecular mass of the eluted particles was calculated from the static light scattering. (**B**) Theoretical masses of each of the HilE constructs and corresponding calculated molecular weights from light scattering measurements (standard deviation is calculated from three replicate experiments). (**C**) Simplified cartoon of the four purified HilE constructs. The GSSGG- sequence at the N-terminus of the third construct is residual following cleavage of the His-tag by TEV protease. (**D**) HilE-His_6 was cross-linked with glutaraldehyde, with varying incubation times and temperatures. All lanes contained 100 µg HilE and, where indicated, 0.1% glutaraldehyde. Incubation times and temperatures: lane 1: HilE only, room temp.; lane 2: 2 min, room temp.; lane 3: HilE only, 37°C; lane 4: 2 min, 37°C; lane 5: 5 min, 37°C; lane 6: 10 min, 37°C. A band corresponding to the molecular weight of a HilE dimer is highlighted by the yellow arrow.

Negative stain EM was used to visualise any potential hexameric ring structures, although none were observed for any of the purified constructs, over a range of tested concentrations. Oligomerisation state was determined by SEC-MALS (Figure 5.2), showing that all constructs existed as monomeric species in solution. The HilE-His_6 construct was less soluble and prone to precipitation once concentrated, so the molecular weight could not be accurately calculated from the light scattering. Preliminary crosslinking experiments (Figure 5.2 D) of this construct showed a prominent band corresponding to the molecular weight of a monomer and a very faint band at ~35kDa, corresponding to a dimer, supporting that none of our purified HilE constructs form higher-oligomer rings.

5.3 Structure Prediction of HilE

We next performed crystallisation trials for HilE. However, no crystals were obtained for any of the purified constructs, either lacking or retaining the His_6-tag at either terminus. Initial NMR experiments were also attempted, with HilE successfully expressed in M9 minimal media for protein labelling and concentrated to the required concentrations for measurements. However, further progress was limited by protein aggregation and precipitation at these concentrations over the time course of the experiments, and additional buffer optimisation is required for future NMR experiments. As we were so far unsuccessful in obtaining a high-resolution structure of HilE, we instead predicted the structure of HilE computationally.

Paredes-Amaya et al. previously reported that HilE has low sequence similarity but high structural similarity to Hcp proteins[45]. Similarly, all of the top hits for HilE structural homologs identified using HHPred[109] are members of the Hcp protein family. To predict the structure of HilE, we employed three of the best performing structure prediction methods: Alphafold 2[111], RoseTTAFold[142] and tFold (https://drug.ai.tencent.com). All three methods predicted similar folds, consistent with that of Hcp proteins (Figure 5.3 C). The only discrepancy is at the C-terminus, which displays low confidence in each of the models, and is shown as unstructured in the AlphaFold model. Comparison with published structures of Hcp proteins indicates the

C-terminus may form an additional β -strand, as predicted in the tFold model (Figure 5.3 D). Hence the tFold model of HilE is used in subsequent figures throughout this chapter.

It has previously been reported that an extended loop region is crucial for oligomerisation of Hcp proteins into hexameric rings and subsequent stacking during nanotube formation[143]. Single point mutations within this loop region in *Salmonella* Hcp2 prevented the assembly of monomers into hexameric rings, whilst deletion of the entire loop yielded only insoluble protein[144]. This loop can be defined from the structural alignment of Hcp structures and is formed in HilE by residues T34-Y45, which link strands β2 and β3 (Figure 5.3 E). It is notably shorter than the corresponding loop in other Hcp-like proteins and may explain our previous finding that HilE does not appear to form hexameric rings.

The hexameric oligomerisation state is predominant in all species that are associated with nanotube formation. Such a structural role has yet to be identified for HilE within the *Salmonella* T6SS and it is plausible that, unlike other Hcp proteins, it simply does not form hexameric ring structures. Although the ability of HilE to form higher-order oligomers cannot be completely ruled out, it is at least unable do so *in vitro* without additional cofactors. It is possible that HilE evolved from an ancestral Hcp protein, and truncation of the extended loop resulted in a loss of the hexamer oligomerisation state.

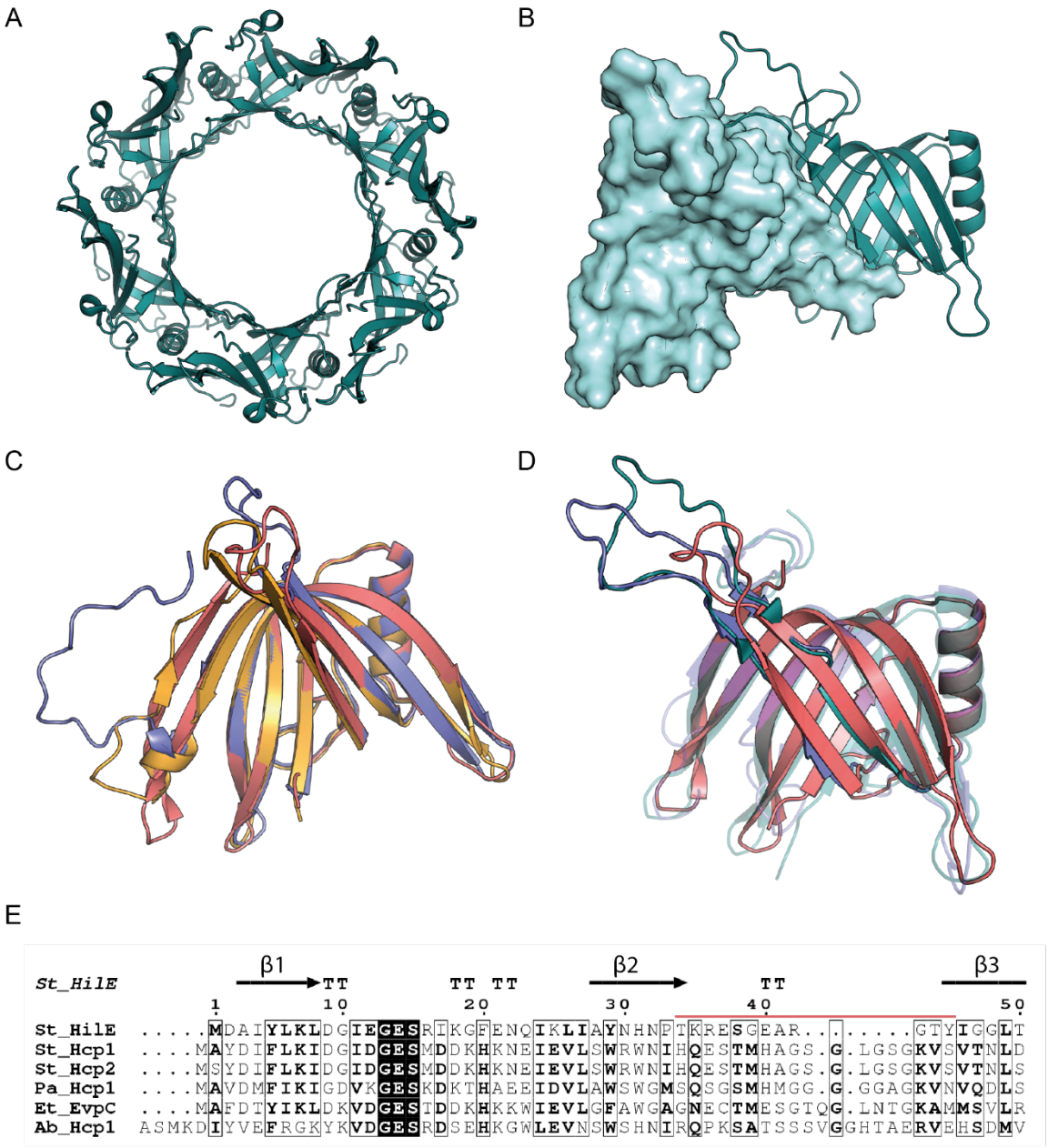
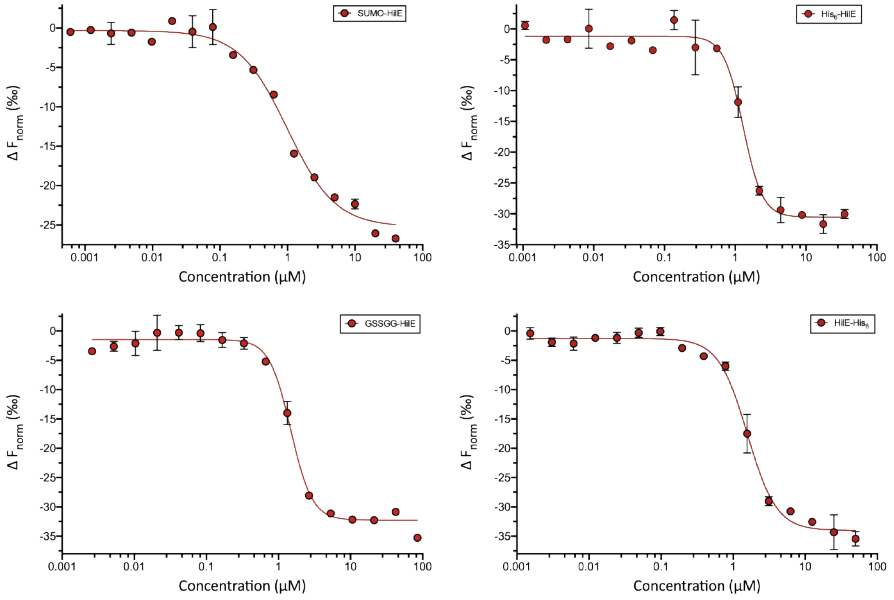


Figure 5.3. HilE shows strong structural similarity to Hcp proteins. (**A**) The hexameric ring formed by *Acinetobacter baumannii* Hcp1 (PDB:4W64). (**B**) The extended loop in Hcp1 forms contacts with neighbouring subunits of the hexamer (shown in surface representation) and is critical for ring formation. (**C**) Predicted structures of HilE from AlphaFold 2 (purple), RoseTTAFold (orange) and tFold (red) were structurally aligned with one another. The C-terminus (shown at the left of the figure) was only modelled as an additional β-strand in the tFold model. (**D**) Structural alignment of HilE (tFold model, red) with Hcp1 from *Acinetobacter baumannii* (teal) and *Pseudomonas aeruginos* (PDB: 1Y12; purple). In HilE, the loop between strands β2 and β3 (top left of the figure) is notably shorter than in other Hcp proteins that form hexameric rings. (**E**) Homologs of HilE were identified using HHPred. Multiple sequence alignment of identified Hcp proteins was performed using ClustalΩ and highlights the shortened length of the loop in HilE (marked in red) compared to other Hcp proteins. β -strands, as predicted in the structure of HilE, are denoted by the arrows above the alignment.

5.4 Biochemical and Biophysical Characterisation of HilE

Given the unexpected oligomerisation state observed for all the HilE constructs, we tested whether each were able to bind HilD *in vitro*, using MST. Surprisingly, all constructs were able to bind to HilD with comparable affinities of around 1 µM (Figure 5.4). This shows that a higher-order oligomerisation state of HilE is not a requirement for binding to HilD, and casts doubt on the previous hypothesis that two hexamers of HilE form a large complex with a HilD dimer[44].



Construct	K_d (µM)	
SUMO-HilE	1.11	± 0.148
His_6-HilE	1.56	± 0.496
GSSGG-HilE	1.68	± 0.471
HilE-His_6	1.88	± 0.396

Figure 5.4. MST shows that HilE binds to HilD with low micromolar affinity. (**A**) Each of the four HilE constructs was titrated against GFP-HilD (50 nM constant concentration), and the thermophoresis was measured. Calculated K_d values were calculated from changes in thermophoresis measured at 2.5 seconds on-time, from two repeat experiments.

For subsequent experiments we used the HilE construct which had an N-terminal His_6-tag followed by a TEV cleavage site. Cleavage of the tag during purification allowed for a simple two-step purification with a higher sample purity than the constructs retaining the His tag, without affecting protein yield. It also avoided issues from the presence of a larger fusion protein (SUMO-HilE construct) or lower protein solubility (HilE-His_6 construct). Following purification to homogeneity, the thermal stability of HilE was determined using NanoDSF, showing a melting temperature of 41.1°C (Figure 5.5).

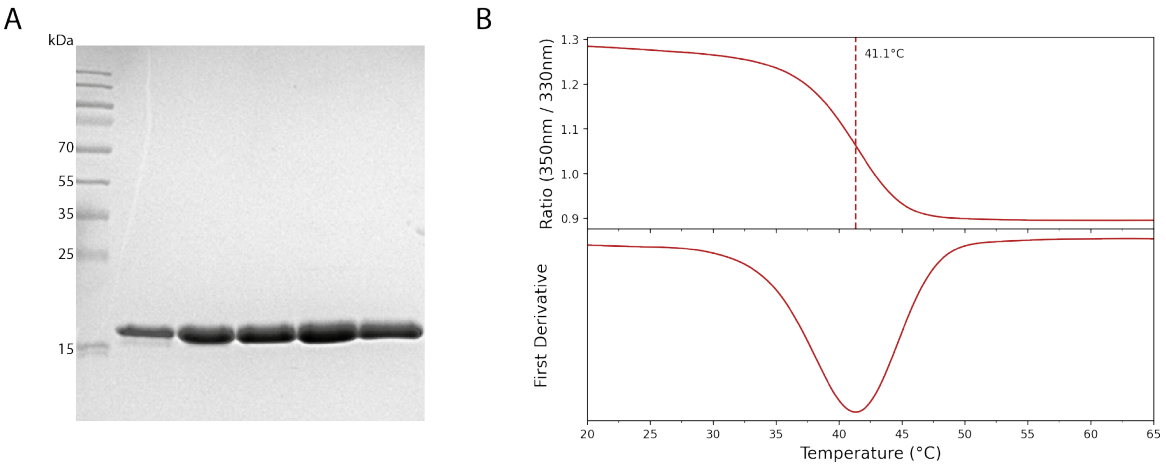


Figure 5.5. Purification and biophysical characterisation of HilE. (**A**) HilE was purified using a two-step purification protocol. Samples of the eluted fractions from the second Ni column were separated on a 15% SDS gel, showing a single band corresponding to the molecular weight of a HilE monomer. (**B**) NanoDSF unfolding profile of HilE (1 mg ml^{-1}). The melting temperature of HilE (41.14 ± 0.09°C) was calculated using NanoTemper PR.Stability Analysis.

5.5 HilE Inhibits HilD Dimerisation.

The inhibition of HilD DNA-binding activity by HilE has previously been reported in multiple studies[44,45] and we reproduced this result using an EMSA to confirm our purified HilE constructs were active and able to bind HilD *in vitro* (Figure 5.6 A). The effect of HilE on HilD dimerisation, however, has been subject to debate. Whereas Grenz et al.[44] found that HilE did not markedly affect the homodimerisation of either HilD or HilC, Paredes-Amaya et al.[45] found HilE to negatively affect HilD dimerisation. We repeated the MST assay previously used to quantify HilD homodimerisation, in the presence of HilE. When GFP-HilD was pre-incubated with excess HilE (10 µM, ~10-fold higher than the calculated K_d), binding of HilD to the GFP-HilD fusion was no longer observed (Figure 5.6 B). This confirms that HilE negatively regulates HilD activity by inhibiting HilD dimerisation and subsequently preventing DNA binding.

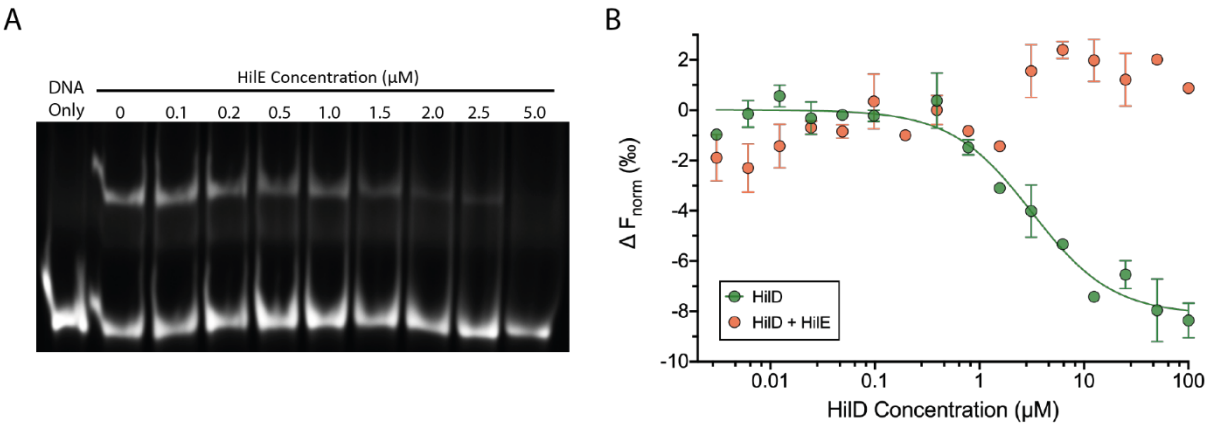


Figure 5.6. HilE effects both HilD dimerisation and DNA binding. (**A**) 600nM HilD was incubated with 50 nM Cy5-labelled *hilA* promoter DNA and increasing concentrations of SUMO-HilE, as indicated. (**B**) HilD was titrated against GFP-HilD (50 nM). To investigate any effect of HilE on dimerisation, GFP-HilD was pre-incubated with 10 μM HilE (orange). Displayed data shows changes in thermophoresis at an MST on-time of 1.5 seconds. Error bars represent standard deviation from three replicate experiments.

5.6 HilE and LCFAs Regulate HilD via Mutually Exclusive Mechanisms

Both HilE and LCFAs negatively regulate HilD activity. HilE was previously shown to not be required for repression of HilD by *cis*-2-unsaturated LCFAs[60], as these compounds bind directly to HilD. We investigated whether competition exists between HilE and LCFAs for HilD, or if a possible additive effect exists between them to repress HilD activity.

We repeated the MST assay used to determine the affinity between HilD and HilE, pre-incubating GFP-HilD with oleic acid (or DMSO as a control), before incubation with varying concentrations of HilE. Whilst the addition of DMSO had no effect on the interaction between HilD and HilE (K_d: 1.70 ± 0.50 μM), oleic acid completely abolished binding (Figure 5.7 A).

A reverse assay setup, in which GFP-HilD was incubated with a constant concentration of HilE (10 μM) and increasing concentrations of oleic acid yielded an apparent K_d of 350 μM for oleic acid binding (Figure 5.7 B). Changes in thermophoresis were only detected at oleic acid concentrations > 40 μM, significantly higher than the HilE assay concentration. The calculated K_d is also significantly higher than that obtained for direct binding of oleic acid to HilD (48 μM), suggesting oleic acid is unable to bind to the HilD-HilE complex. This suggests only one

of these regulators can bind to, and regulate, HilD at a time and that the two regulatory mechanisms exist independently of one another.

Binding of oleic acid (and other LCFAs) to HilD likely leads to conformational changes in the overall protein fold, affecting the HilE-binding region and thus preventing HilE from binding to HilD. HilE was found to have a much higher affinity than any of the tested LCFAs for HilD. This suggests that repression by HilE is the principle regulatory mechanism controlling HilD activity. The binding of LCFAs to HilD then provides an additional level of control over HilD activity under certain environmental conditions encountered by *Salmonellae* in the intestine.

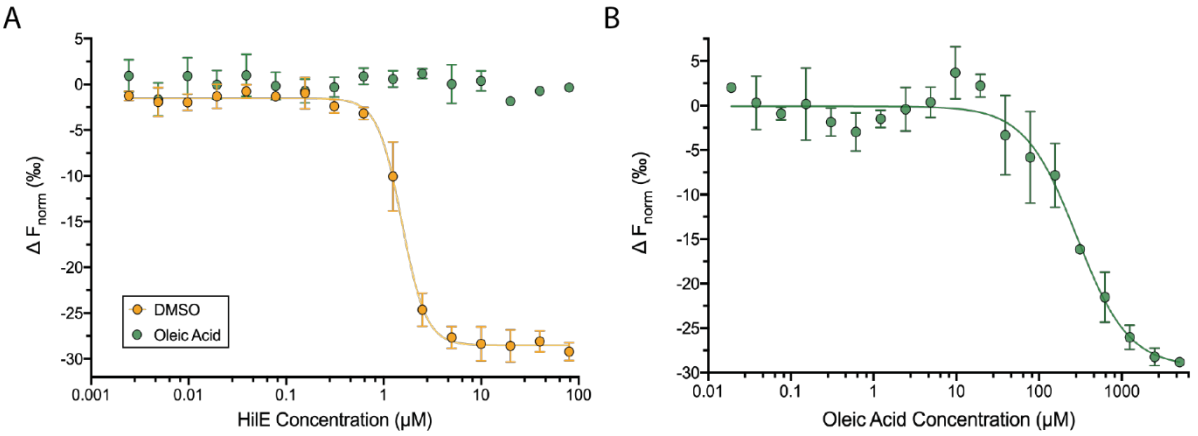


Figure 5.7. HilE and LCFAs bind independently to HilD. (**A**) Oleic Acid prevents the binding of HilD and HilE. GFP-HilD was incubated first with either 1% DMSO (yellow) or 100 μM oleic acid (green), and then increasing concentrations of HilE. (**B**) GFP-HilD (50 nM) was mixed with HilE (10 μM), and the protein complex incubated with varying concentrations of oleic acid. In both assays, binding was determined by monitoring changes in thermophoresis at an MST on-time of (A) 1.5 or (B) 5 seconds, and affinities calculated using MO.Affinity Analysis. Standard deviation was calculated from three replicate experiments.

5.7 HilD and HilE Form a Stable SEC Complex

Given that we were unable to detect the formation of higher-order oligomers for any of our HilE constructs, but found that all constructs were still able to bind to HilD, we next aimed to determine the stoichiometry of the HilD-HilE complex.

We first ran a gel filtration assay to determine whether a stable HilD-HilE complex could be isolated using a size exclusion chromatography (SEC) column. HilD and HilE were mixed in a 1:1 ratio (both proteins at 100µM) and loaded to a Superdex 75 10/300 GL increase column. As seen in Figure 5.8 A, a single elution peak was observed. Eluted fractions were collected (highlighted on the elution trace by the red line) and loaded on an SDS gel, clearly showing this peak corresponds to the HilD-HilE complex with residual amounts of the individual proteins eluting as shoulders either side of the main peak (Figure 5.8 B).

Subsequent HDX experiments performed on the HilD-HilE complex (discussed in section 5.10) showed no differences in HDX profiles for the SEC-purified complex and that constituted by simply mixing HilD and HilE prior to the HDX experiments. This shows simple mixing of the two proteins leads to reasonable complex formation, comparable to a complex obtained from SEC, and so this additional purification step was omitted prior to subsequent crystallisation trials of the HilD-HilE complex.

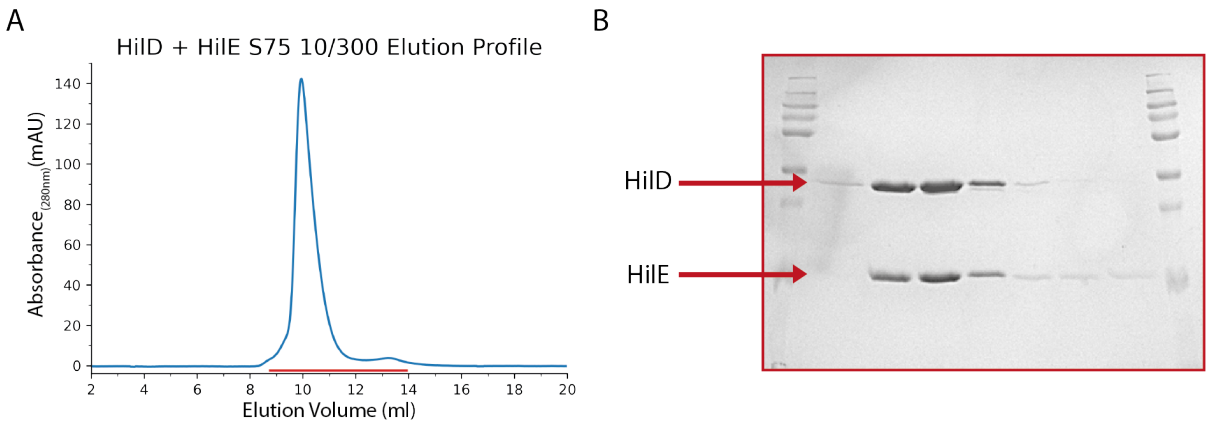
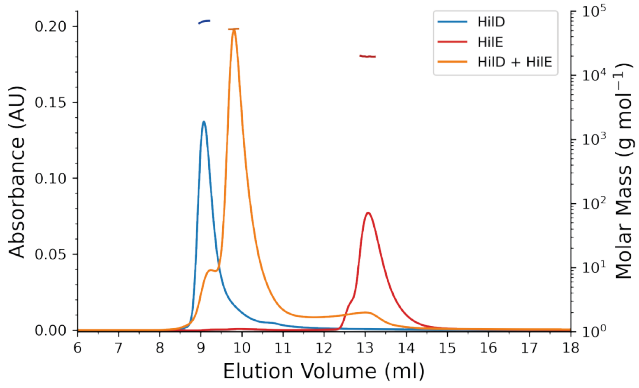


Figure 5.8. HilD and HilE form a stable protein complex. (**A**) Elution profile for the Superdex 75 10/300 increase SEC run. Absorbance was monitored at 280 nm. (**B**) Selected fractions, highlighted by the red line in (A), were loaded to an SDS gel and stained with Coomassie.

5.8 HilD and HilE Form a 1:1 Complex

To determine the stoichiometry of this interaction, we performed SEC-MALS experiments. HilD, HilE, and a 1:1 mixture of the two proteins (concentrations of all proteins at 100 µM) were loaded to a Superdex 75 10/300 GL increase column, and the molecular mass of eluted proteins was calculated from the static light scattering. A clear peak corresponding to the HilD-HilE complex elutes later than that for HilD alone (Figure 5.9 A). The calculated mass of 52.7 ± 0.9 kDa corresponds to a 1:1 complex of the two proteins and supports a model in which one HilE monomer replaces one of the two HilD molecules constituting the homodimer (Figure 5.9 B).

A



B

	Oligomerisation State	Molecular Mass (kDa)	
		Theoretical	SEC-MALS
HilD	Dimer	70.4	70.4 ± 2.0
HilE	Monomer	16.9	19.4 ± 1.1
HilD + HilE	1:1 complex	52.1	52.7 ± 0.9

C

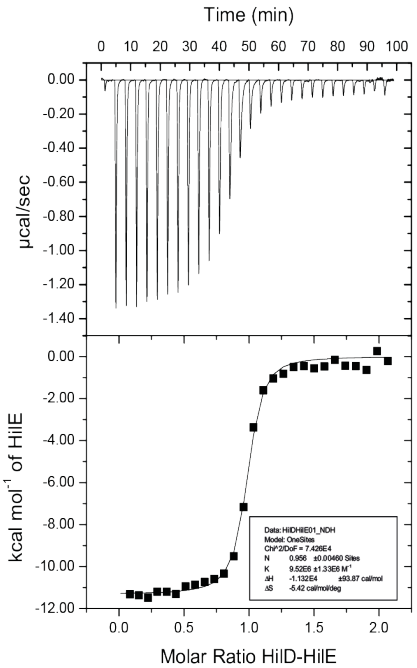
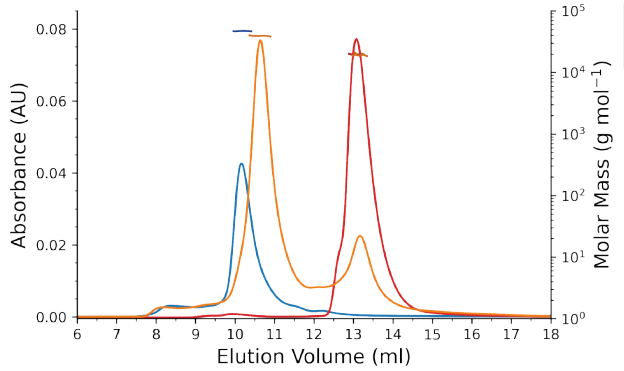


Figure 5.9. HilE binds to HilD with a 1:1 stoichiometry. (**A**) SEC-MALS of HilD and HilE. 50µl of protein sample (100 µM HilD (blue), 100 µM HilE (red) and a 1:1 mixture (both proteins at 100 µM, orange)) was loaded to a Superdex 75 10/300 increase column. (**B**) Calculated molecular weight values for the eluted peaks from each of the SEC-MALS runs, compared to the theoretical masses of the corresponding species. Standard deviation is calculated from three replicate experiments. (**C**) ITC measurement for the binding of HilD and HilE. 300 µM HilE was injected into 1.4 ml of 30 µM HilD, in 10 µl intervals. The fitted binding curve yielded a K_d of 105.9 ± 12.28 nM. ITC data is from a single experimental run.

This result was supported by a preliminary isothermal titration calorimetry (ITC) experiment (Figure 5.9 C), which showed a high-affinity interaction between the full-length proteins (K_d: 105.9 ± 12.28 nM) with a 1:1 stoichiometry. These results support the finding that HilE disrupts HilD dimerisation, as shown in both our MST assay and previously[45]. Disruption of the HilD homodimer, and formation of the HilD-HilE heterodimer, prevents HilD from being able to bind to target DNA and activate *hilA* expression.

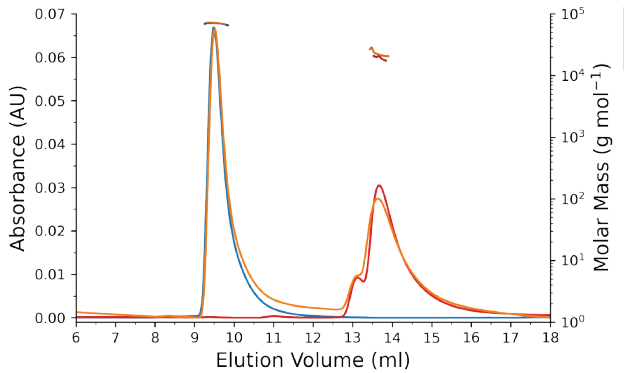
It was previously observed that HilE interacts with the central and C-terminal regions of HilD, encompassing residues 130-309[45]. We repeated the above SEC-MALS experiment, using the His_6-$HilD_{7-206}$ (NTD) construct rather than full-length HilD. As seen in Figure 5.10 A, a stable complex was also formed for the interaction of HilE with the HilD NTD, showing that HilE interacts with the N-terminal domain of HilD, and the DBD is not required for HilE binding. Additional affinity measurements are required to confirm whether the DBD of HilD also contributes to the interaction with HilE. HilC was previously shown not to bind to HilE[41,44] and therefore used a negative control for SEC-MALS experiments. In this instance, no stable complex was formed between HilC and HilE, with the SEC UV trace showing two clear peaks corresponding to the elution of the individual proteins (Figure 5.10 B).

A



		Molecular Mass (kDa)		
		Theoretical	SEC-MALS	
$HilD_{7-206}$ (NTD)	Dimer	47.8	46.9	± 1.4
HilE	Monomer	16.9	19.4	± 1.1
NTD + HilE	Peak 1	40.8	40.6	± 1.1
	Peak2	-	24.1	± 4.5

B



		Molecular Mass (kDa)		
		Theoretical	SEC-MALS	
HilC	Dimer	68.0	69.3	± 0.9
HilE	Monomer	16.9	23.2	± 3.6
HilC + HilE	Peak 1	-	70.4	± 0.6
	Peak2	-	24.2	± 1.6

Figure 5.10. HilE interacts specifically with the N-terminal regulatory domain of HilD. (**A**) SEC-MALS elution traces for the HilD NTD (blue), HilE (red) and a 1:1 mixture of both (orange). A protein concentration of 100 µM was used for all runs. (**B**) SEC-MALS runs for HilC (blue), HilE (red) and a 1:1 mixture of HilC and HilE (orange). For runs involving HilC, all proteins were kept at a concentration of 50 µM, due to the lower solubility and observed precipitation of HilC at higher concentrations. For all runs, 50 µl of each protein sample was loaded to the Superdex 75 10/300 increase column. Calculated molecular weights are the average of three replicate experiments.

5.9 Predicted Structure of the HilD-HilE Complex

Having determined that HilE forms a 1:1 complex with HilD, via binding to the N-terminal domain, we next sought to determine the 3D structure of this complex and the residues involved in this interaction. Crystallisation trials of the HilD-HilE complex proved unsuccessful, so we modelled this complex using AlphaFold Multimer[93].

The predicted structures show HilE directly disrupts HilD dimerisation by displacing a single HilD monomer of the dimer pair. The highest ranked model shows the sole α-helix of HilE (formed by residues 58-68) interacting with the dimerisation helix of HilD (residues 175-192,

as defined by the AlphaFold model of the HilD dimer (Figure 5.11 A) and previously[64]). Lower ranked models varied in their orientation of HilE, but all predicted the dimerisation helix as the binding site on HilD.

In the highest ranked model, the two interacting helices of the corresponding proteins are perpendicular to one another (Figure 5.11 B-C). As for the HilD homodimer, complex formation appears to be largely driven by hydrophobic interactions. However, additional hydrogen-bonding interactions are also predicted between residues Y180 and R191 of HilD with N68 and H106 of HilE. HilD residues Y180, F181 and F184, which all face the HilD-HilE interface, were previously reported to be involved in the interaction with HilE[145].

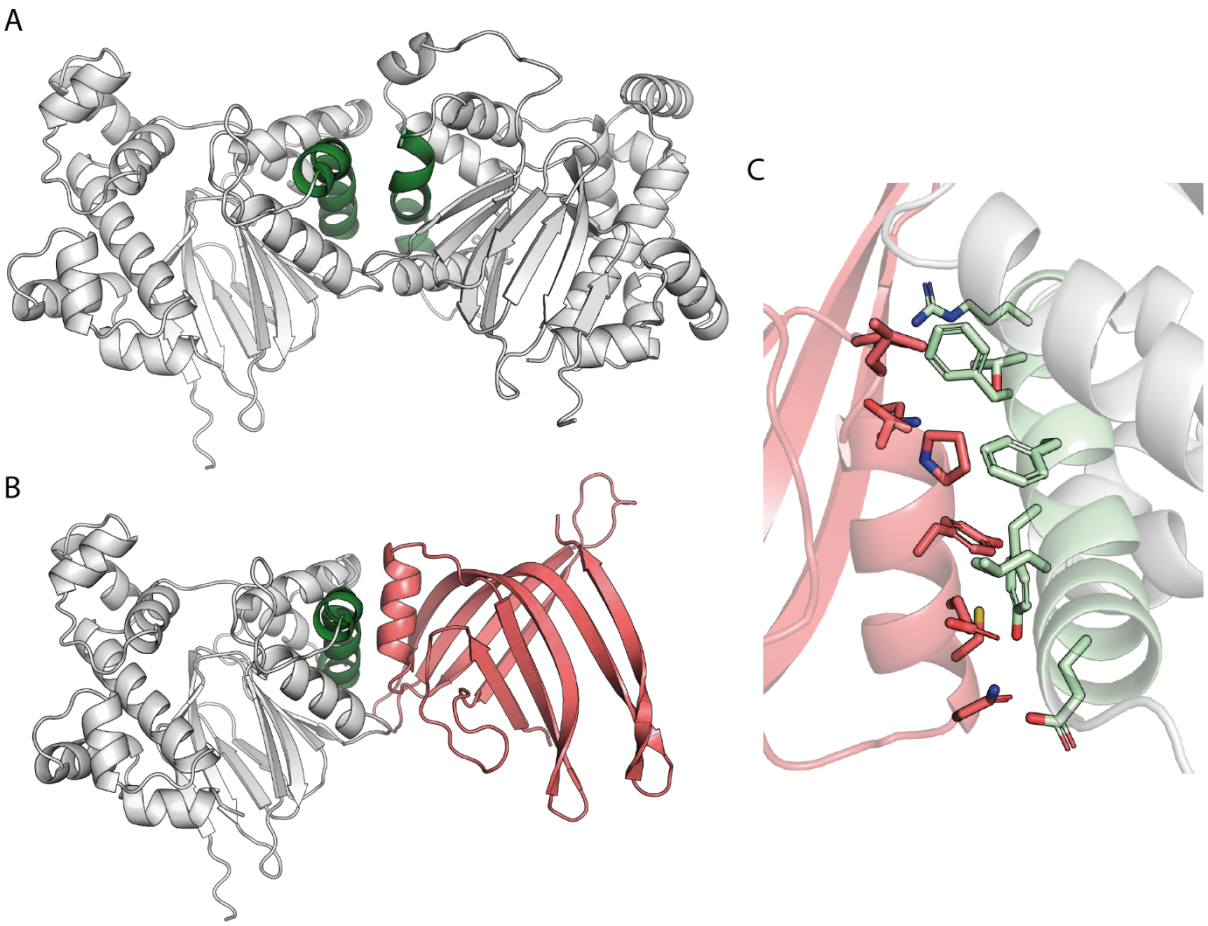


Figure 5.11. HilE is predicted to directly disrupt the HilD homodimer complex. (**A**-**B**) Structures of (**A**) the HilD homodimer and (**B**) the HilD-HilE complex. Both figures show the highest ranked models predicted by AlphaFold. The HilD dimerisation helix is highlighted in dark green. (**C**) Detailed view of the interface of the HilD-HilE complex shown in (B). Side chains of residues forming the interface are shown as sticks.

5.10 HDX of the HilD-HilE Complex

To experimentally validate the HilD-HilE model, we again utilised HDX-MS. HDX-MS experiments were performed for both HilD and HilE alone, and in complex with one another. The HilD-HilE complex was prepared either by purification over a Superdex 75 column or simply mixing the two proteins together in a 1:1 ratio. The two complex samples produced almost identical HDX profiles, confirming that protein mixing leads to reasonable complex formation comparable to purification by SEC.

To determine the HilE binding site on HilD, we compared the HDX profiles of HilD alone and in complex with HilE. Increased HDX (i.e. strong deprotection) at the dimerisation interface of HilD is consistent with the observation that HilE disrupts the dimerisation of HilD (Figure 5.12 A). The strong increase in D-incorporation in this region implicates these helices to be more solvent-exposed in the HilD-HilE complex. This contradicts the binding mode predicted by the AlphaFold model, in which a HilE monomer directly displaces one of the HilD monomers of the dimer pair.

These HDX-MS results imply the binding of HilE to an allosteric site, causing conformational changes that disrupt the HilD dimer and expose the dimerisation interface to the bulk solvent. Binding of HilE to an allosteric site would be expected to show decreased D-incorporation in this region due the formation of a new interface. Strong protection in the presence of HilE is observed at HilD residues T41-F50 and I214-S218, and at the N-terminal end of helix α4 (residues R258-I268) and helices α5-α7 of the DBD, comprising HTH2 and the surrounding regions (Figure 5.12 B). HilE could be envisioned to bind either to helices α6-α7 (the DNA binding site) or in a cleft between the two HilD domains.

However, binding of HilE to one of the HTH motifs contradicts our previous findings that HilE forms a stable complex with the NTD of HilD. The interaction of HilE with the HilD NTD (residues 1-220) was previously also reported using a LexA-based two-hybrid genetic system[45]. These differences in HDX in the DBD may also be due to secondary effects, rather than direct binding. A reduction in HDX at both HTH motifs is indicative of conformational changes across the whole domain upon HilE binding. Decreased HDX was also observed across the entire DBD upon oleic acid binding to HilD, which, like HilE, inhibits HilD from binding to

DNA. It is plausible that the binding of HilE, as predicted in the AlphaFold model, causes conformational changes in the overall structure of HilD. This could result in decreased flexibility of the DBD, preventing independent rotation of the two HTH motifs with respect to one another.

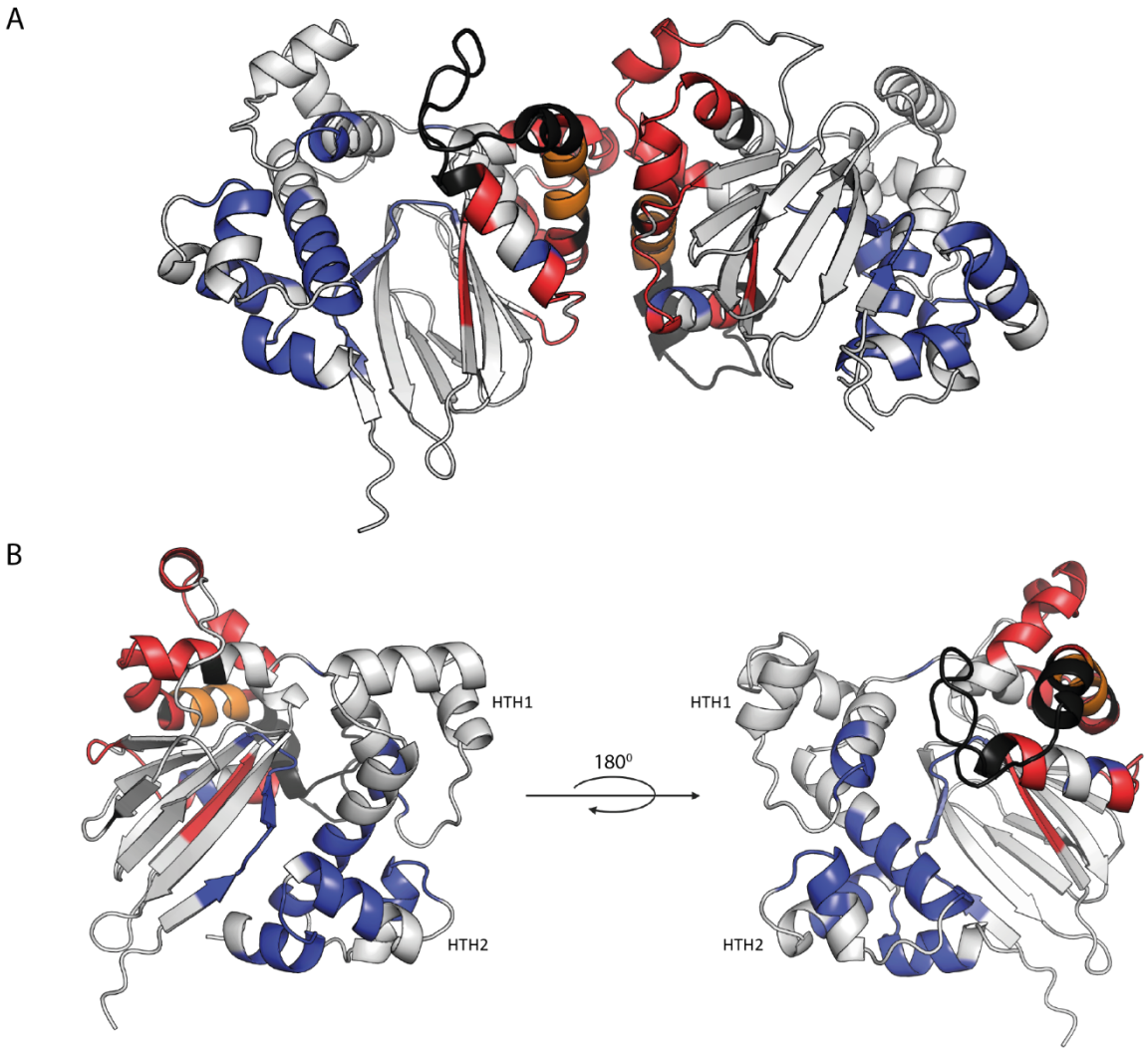


Figure 5.12. Changes in HDX of HilD upon binding of HilE. The predicted structure of HilD is coloured according to differences in HDX when alone and in complex with HilE (for the complex purified by SEC prior to HDX experiments). Residues coloured in blue showed lower D-incorporation in the presence of HilE, whilst those in orange and red show increased D-incorporation. Residues in white showed no detectable changes in D-incorporation, whilst no information could be obtained for those coloured in black due to a lack of sequence coverage following peptide digestion. The top panel (**A**) shows the structure of the HilD homodimer, highlighting changes in HDX at the dimer interface. The bottom panel (**B**) shows the structure of a single HilD monomer constituting the dimer pair, rotated to display the decreased HDX across the DBD.

HDX changes were also investigated for HilE, in the absence and presence of HilD. Whilst strong protection is observed for the helix postulated to form the binding interface with HilE, such HDX changes are also observed across the entire HilE sequence (Figure 5.13). This suggests that upon binding to HilD, HilE undergoes large conformational changes that are transmitted over the β-strands, affecting the entire domain. It is therefore not possible to pinpoint the exact interaction site on HilE.

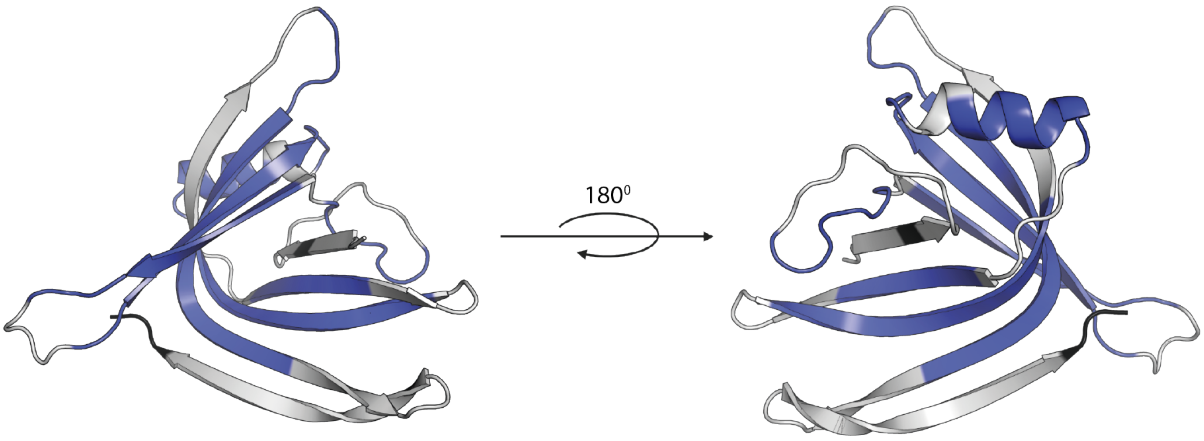


Figure 5.13. Changes in HDX of HilE upon binding of HilD. The predicted structure of HilE is coloured according to differences in HDX when in complex with HilD (for the complex purified by SEC prior to HDX experiments), compared to when alone. Residues are coloured as described for Figure 5.12.

Although HDX-MS shows increased D-incorporation at the HilD dimerisation interface in the presence of HilE, this is less pronounced for residues L177-L183 at the centre of the dimerisation helix (coloured in orange in Figures 5.12 and 5.14). This correlates with the predicted model of the HilD-HilE complex, in which the interacting helix of HilE is perpendicular to the dimerisation helix of HilD. Comparison of the predicted interfaces using PISA[146] revealed the buried surface area in the HilD-HilE complex (652.2 Å^2) to be notably smaller than that formed in the HilD homodimer (923.5 Å^2) (Figure 5.14 C-D). In the HilD homodimer, the dimerisation helices (residues 175-192) are positioned parallel to one another and additional contacts are made between residues A113-Q119 and F139-H148 of the surrounding helices. The smaller interface of the HilD-HilE heterodimer would result in increased exposure of the residues forming these helices to the bulk solvent compared to in the HilD homodimer, rationalising the increased D-incorporation observed for these regions in HDX-MS experiments (Figure 5.14 E).

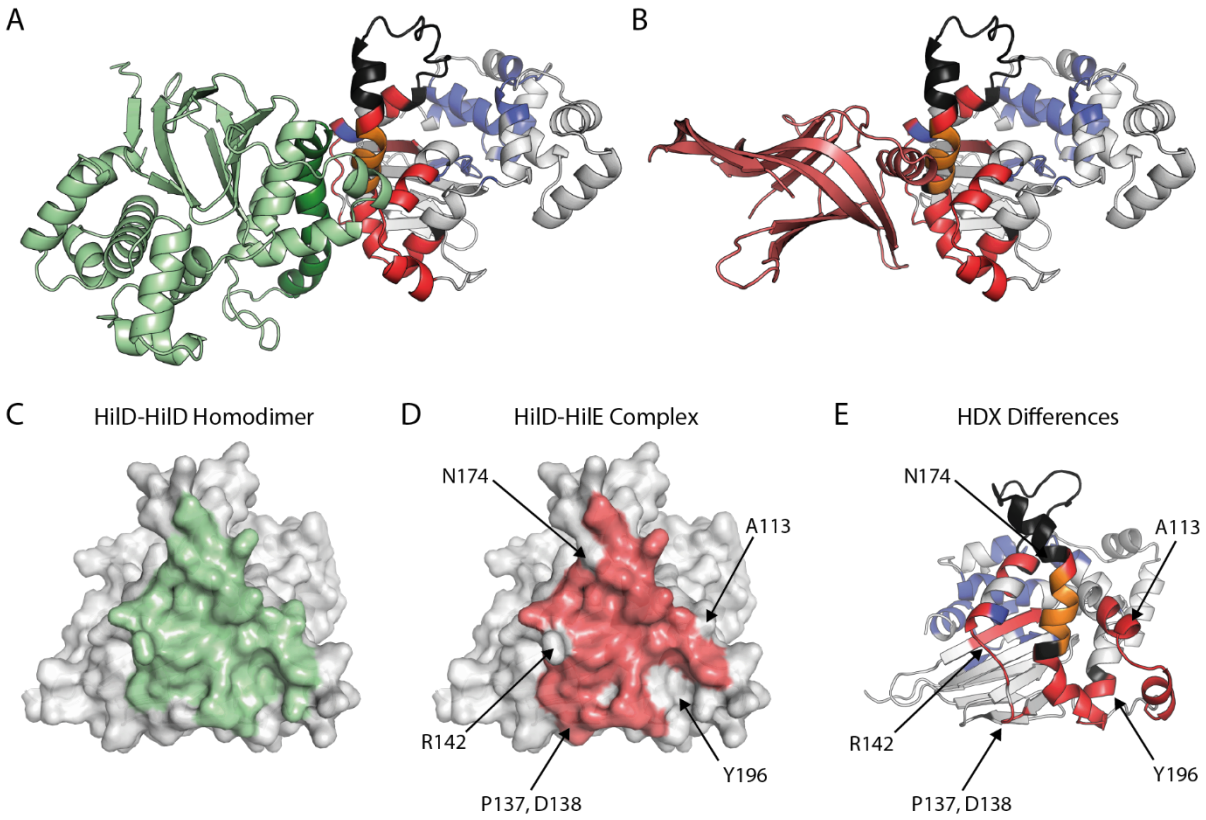


Figure 5.14. The predicted domain interface of the HilD-HilE heterodimer is smaller than that of the HilD homodimer. Predicted structures of (**A**) the HilD homodimer and (**B**) the HilD-HilE heterodimer, with the common HilD monomer coloured by changes in HDX upon HilE binding, as described for Figure 5.12. The second protein of the complex is coloured according to the identity of the protein: HilE, red; HilD, green (dimerisation helix, dark green). (**C**-**D**) Coloured surface representation of HilD, highlighting the residues that form the interface in each complex, as identified using PISA. Interface residues of (**C**) the HilD homodimer and (**D**) the HilD-HilE heterodimer are coloured in green and red, respectively. (**E**) Cartoon representation of HilD, orientated as in (C, D) and coloured by changes in HDX upon incubation with HilE, as described for Figure 5.12. Residues identified as only forming the interface in one of the two complexes are labelled in both (D) and (E).

5.11 Discussion

In addition to being regulated by small-molecule ligands, the activity of AraC/XylS transcription factors may also be modulated through protein-protein interactions. A family of small proteins, the AraC Negative Regulators (ANRs), has recently been identified, which is widespread amongst pathogenetic bacteria species in which virulence genes are regulated by AraC/XylS proteins[147]. These small proteins, comprised of three highly conserved α-helices, act by directly binding to associated AraC/XylS proteins, which prevents both dimerisation and DNA-binding[148]. Other AraC/XylS proteins have also been shown to be regulated by proteins that do not belong to the ANR family. The *P. aeruginosa* transcription regulator ExsA is negatively regulated by ExsD, whilst the central regulator of virulence in *S. enterica*, HilD, is negatively regulated by HilE[43,149]. ExsD and HilE act by directly binding to ExsA and HilD, respectively, disrupting the dimerisation and DNA-binding activity of these AraC/XylS regulators.

HilE displays high structural similarity to Hcp proteins, which are structural components of the bacterial T6SS. However, HilE lacks a sequence shown to be critical for the oligomerisation and nanotube formation of other Hcp proteins and we found that, unlike other characterised Hcp proteins, HilE exists predominately as a monomer in solution. No structural or chaperone functions of HilE have been reported to date. This supports the hypothesis that HilE diverged from an ancestral T6SS or phage Hcp protein, evolving from a structural protein required for virulence to a regulator of such virulence genes.

We showed that HilE forms a 1:1 complex with HilD, to inhibit HilD homodimerisation and prevent binding to DNA. Our results indicate that HilE interacts with the dimerisation helix of HilD, directly replacing one of the HilD monomers constituting the dimer pair. This contrasts to ExsD, which does not bind directly to the dimer interface of ExsA, although the exact binding mode remains unknown[119,150]. Further work, verifying the binding mode of HilE to HilD and the interface residues mediating this interaction, is currently ongoing.

The HilD-HilE interaction is notably high affinity, and our MST assay yielded K_d values of higher affinity than that calculated for the homodimerisation of HilD. This supports a hypothesis in which HilD is bound to HilE under normal, non-invasive conditions. This repressive effect is

only overcome once HilD is expressed above the level of available HilE, by the action of positive regulators under conditions suitable for invasion (i.e. at the intestinal epithelium).

The binding of LCFAs to a conserved pocket in HilD is expected to result in conformational changes in the dimerisation helix, akin to those reported for ToxT. Hence, this helix would be locked in an orientation that makes it incompatible with binding to HilE. HilE is no longer able to bind to HilD in the presence of oleic acid, supporting the hypothesis that the HilD dimerisation helix constitutes the HilE binding site. Our results show that the two mechanisms of HilD repression are independent of each other and mutually exclusive. In conclusion, HilD should only be able to activate the expression of *hilA* when all conditions surpassing these repressive effects are met simultaneously, underlining the level of control over the expression of virulence genes that has evolved to ensure efficiency of *Salmonella* pathogenesis.

Chapter 6. C26 is a Novel Small-Molecule Inhibitor of HilD

6.1 Introduction

In the quest to identity inhibitors of *Salmonella* type III secretion system 1 (T3SS-1), with potential as novel antipathogenic compounds, collaborators in the Wagner lab initially focussed on inhibitors of the T3SS-1 structural protein InvA. InvA (SctV in the unified nomenclature) is the largest of the five proteins forming the export apparatus of T3SS-1, forming a channel through which protein substrates reach the entrance of the export apparatus gate prior to secretion[151,152]. InvA is essential for the assembly and function of T3SS-1 and an *invA* knockout completely abolishes the secretion of effector proteins, a vital step in host-cell invasion.

Potential binding compounds of InvA were identified in an *in-silico* screen and 49 of these compounds were validated in an *in vivo* secretion assay. The secretion of the T3SS-1 effector protein SipA was monitored in the presence of the different hit compounds, using a NanoLuc luminescence-based assay[153]. Compound 26 (C26), a small molecule of less than 400 Da in size, was found to decrease the secretion of SipA in a dose-response manner with low micromolar affinity (Figure 6.1 A-B) whilst exhibiting low toxicity in HeLa cells. Cross-validation by Western blot confirmed that C26 inhibits the secretion of effector proteins (SipA, SipB and InvJ) into the supernatant. Intriguingly, analysis of whole cell extracts revealed C26 additionally reduces the expression of these secreted proteins, in contrast to a *ΔinvA* mutant (Figure 6.1 C), suggesting the target of C26 is one of the transcription regulators of SPI-1 rather than InvA. Subsequent transcriptome analysis by RNA sequencing revealed C26 downregulates each of HilD, HilC and RtsA, along with the other transcription regulators downstream of this regulatory triad in SPI-1 regulatory network and all other SPI-1 genes encoding T3SS-1 structural and effector proteins (Figure 6.1 D). An *in vivo* assay was developed in which the *hilA* promoter (P_{hilA}) was fused to a *sfGFP* reporter gene, enabling direct monitoring of P_{hilA} activation from a fluorescence readout (Figure 6.1 E). C26 results in a dose-dependent reduction in P_{hilA} activation, with an IC_{50} of ~30 µM, confirming that C26 acts upon the regulators forming the HilD/HilC/RtsA feed-forward loop (Figure 6.1 F).

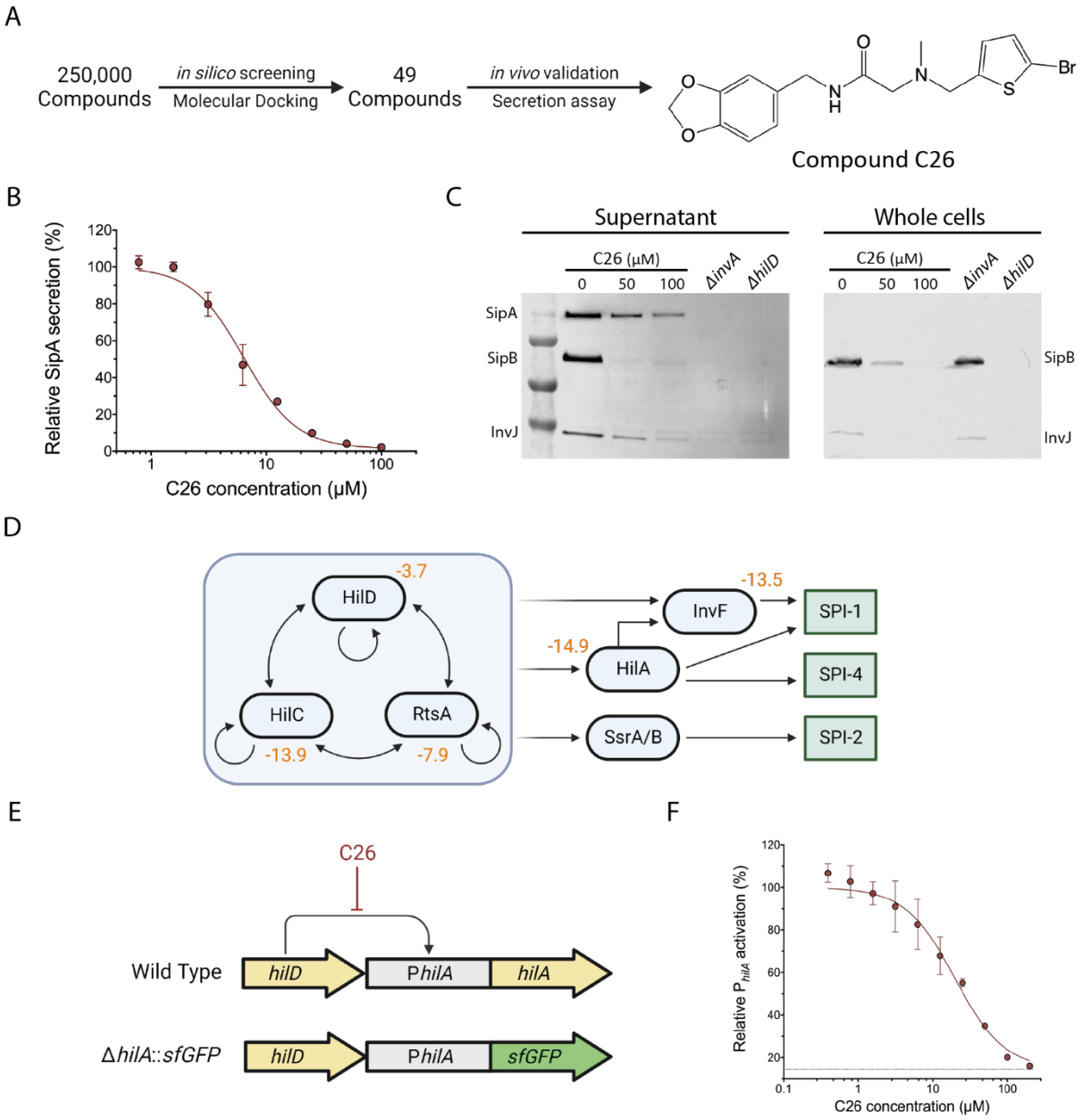


Figure 6.1. C26 inhibits the expression of *Salmonella* virulence genes. (**A**) Initial workflow leading to the identification of the T3SS-1 inhibitor C26. (**B**) Dose-response curve of SipA secretion with increasing concentrations of C26. (**C**) Western blot monitoring the secretion (into the supernatant) and expression (whole cells) of T3SS-1 effector proteins, in the presence of C26. *ΔinvA* and *ΔhilD* knockout mutants are shown as controls. (**D**) Simplified model of the core SPI regulatory network. The fold- changes in expression levels, as determined RNA sequencing, upon treatment of *Salmonella* with C26 (100μM) under SPI-1-inducing conditions, are indicated in orange. (**E**) Activation of the *hilA* promoter was monitored by replacing the *hilA* gene with a *sfGFP* reporter gene. (**F**) Dose-response curve of changes in P_{hilA} activation with increasing concentrations of C26.

As the dominant regulator of the regulatory triad, HilD was proposed as the most likely target of C26. We sought to test this hypothesis and determine whether C26 interacts specifically with HilD (as is the case for the negative regulator HilE) or additionally with HilC and RtsA (as shown for long chain fatty acids). We employed a range of biophysical techniques to characterise the C26-target interaction and determine the mechanism-of-action of C26 leading to the abolishment of secretion of effector proteins.

Having identified the target of C26, I aimed to characterise the interaction between HilD and C26. We initially sought to obtain a crystal structure of HilD in complex with C26, and subsequently employed HDX-MS in combination with computational methods to uncover the binding mode of C26 to HilD. In combination with *in vivo* assays performed by collaborators in the Wagner lab, a lead optimisation of C26 was performed, identifying structural moieties critical for binding and C26 analogs with higher target affinity (K_d < 5 µM).

6.2 Target Validation of C26

I investigated the binding of C26 to both HilD and HilC. *In vitro* binding to RtsA could not be investigated due to low protein stability, although *in vivo* target validation of C26 nonetheless showed C26 not to bind to RtsA.

Nano differential scanning fluorometry (NanoDSF) was previously used to determine the thermal stability of HilD and HilC. Protein unfolding was monitored in the presence of increasing concentrations of C26. Here, protein melting temperatures were determined from changes in fluorescence at 350 nm, rather than the 330/350 nm fluorescence ratio, due to significant intrinsic fluorescence of C26 at 330 nm, which masked protein unfolding and prevented data evaluation (Figure 6.2 A-B). C26 increases the melting temperature (T_m) of HilD in a dose-response manner yet has no significant effect on the thermal stability of HilC (Figure 6.2 C-D). This result indicated that C26 interacts specifically with HilD.

Other regulators of SPI-1 that modulate HilD activity at the protein level do so by preventing HilD from binding to DNA and activating target genes. We investigated whether C26 impairs the DNA-binding activity of HilD and HilC using EMSAs. Increasing concentrations of C26 resulted in a decreased intensity of the HilD-DNA complex band yet had no effect on HilC DNA

binding (Figure 6.2 E). At the highest concentration of C26, a significant band corresponding to the HilD-DNA complex is still observed, which could be due to low affinity of C26 to HilD. To confirm this observed result is due to specific binding of C26 and not an artefact, we repeated this assay with two higher-affinity compounds, C202 and C206, which were identified during a structure-activity relationship (SAR) analysis of C26 (section 6.6). These compounds differ from C26 in the identity of the substituent on the benzyl group, containing a para- halogen atom (Cl and F, respectively) in place of the dioxolane functional group. Here, we obtained a clearer result (Figure 6.2 F), confirming that this group of compounds disrupt the DNA-binding activity of HilD, but not of HilC. Taken together, these results show that C26 and its derivative compounds bind specifically to HilD, but not to HilC.

This finding was supported by *in vivo* target validation experiments, using the P_{hilA} reporter assay described previously. The effect of C26 on P_{hilA} activation was investigated in knockout mutants, lacking one or more of the three regulators. All strains lacking *hilD* showed no change in P_{hilA} activation in the presence of C26. In contrast, P_{hilA} activation was reduced in each of the *ΔhilC*, *ΔrtsA*, *ΔhilCΔrtsA* mutants, akin to the observed effect for the wild type. This confirms that HilC and RtsA are not targets of C26, which interacts specifically with HilD.

6.3 Affinity of C26 to HilD

Having identified HilD as the target of C26, we next sought to determine the affinity of this interaction. We first utilised the MST assay optimised to calculate the affinities of LCFAs to HilD. However, binding above the minimum confidence threshold was not observed for C26, and a dose-response curve could not be fitted (Figure 6.3 A). When this MST assay was repeated with higher affinity analogs, full binding curves could be obtained (Figure 6.3 B). However, even for these compounds the fitted binding curves were less convincing than those attained for LCFAs, with significantly lower response amplitudes and reduced signal-to-noise ratios alongside issues with reproducibility (Figure 6.3 C). One possible explanation for this is that the binding of C26 (and derivative compounds) does not cause a large conformational change in the overall structure of HilD, as was shown to be the case for the binding of LCFAs.

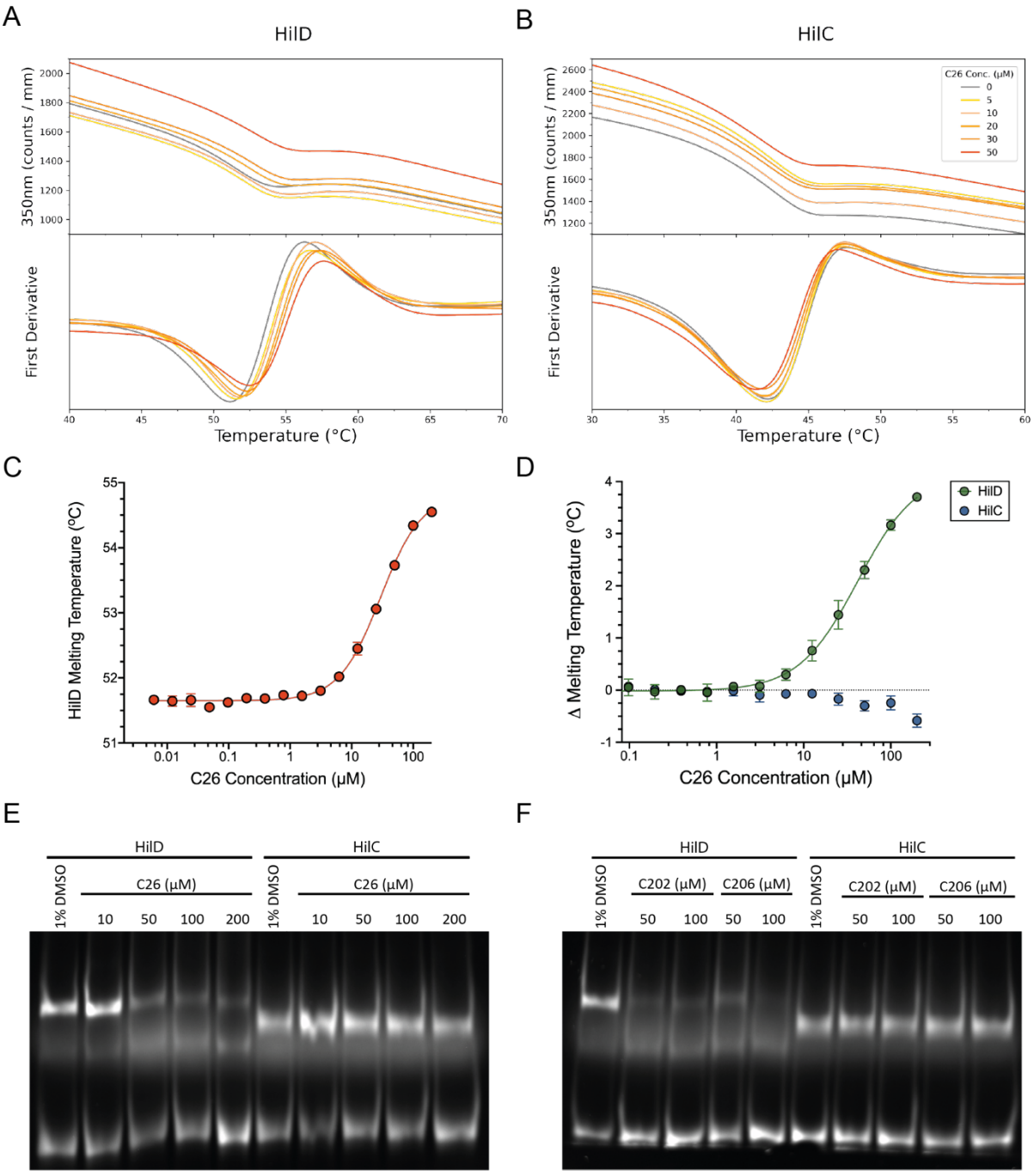


Figure 6.2. C26 interacts specifically with HilD. (**A-B**) NanoDSF melting curves of HilD and HilC (10 µM) in the presence of increasing C26 concentrations. Intrinsic protein fluorescence at 350 nm was monitored with increasing temperature, with a temperature gradient of 0.5°C min^{-1}. (**C**) The melting temperature of HilD, calculated from the fluorescence at 350nm, increases in a dose-response manner upon incubation with C26. (**D**) Changes in the calculated melting temperature of HilD and HilC, upon incubation with increasing concentrations of C26. (**E-F**) EMSAs showing binding of HilD or HilC to a fragment (A1_62bp) of the *hilA* promoter, when incubated with inhibitor compounds. 600 nM HilD or HilC was incubated with 50 nM DNA and increasing concentrations of (**E**) C26 or (**F**) the higher affinity inhibitors C202 and C206.

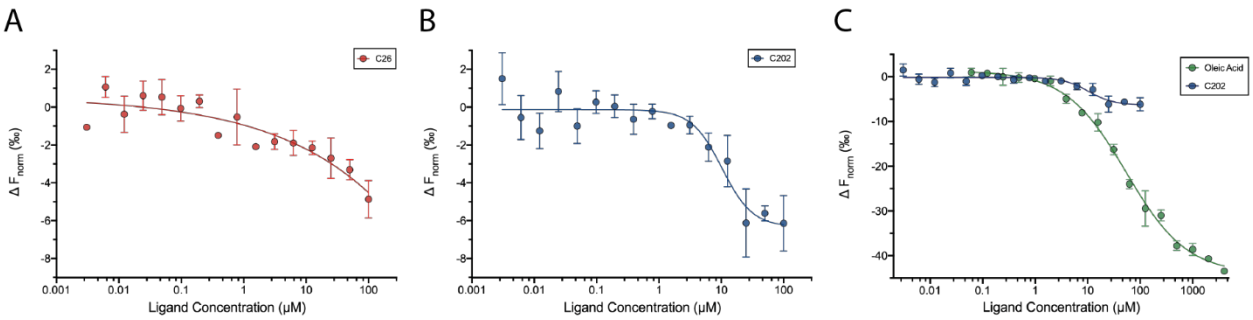


Figure 6.3. MST dose-response curves for the binding of (**A**) C26 and (**B**) C202 to GFP-HilD. (**C**) Overlay of the MST binding curves for C202 and oleic acid, highlighting the large difference in response amplitude for the binding of the two ligands. An MST on-time of 1.5 seconds was used to plot the change in normalised fluorescence (ΔF_{norm}) for all ligands. Error bars show the standard deviation from three replicate experiments.

Since these results were unconvincing and not suitable for a high throughput affinity-determination assay, we sought an alternative approach to determine binding affinities of different compounds to HilD. C26 stabilises HilD in a dose-response manner, and by fitting the changes in HilD melting temperature (T_m) as a function of ligand concentration, the binding affinity of this interaction can be determined[154]. However, this approach does not take into account that K_d is temperature-dependent and defined at specific temperatures. Hence, this method of K_d determination is thermodynamically incorrect and only provides an approximate affinity value, defined as the apparent K_d ($K_{d,app}$)[85,86].

An alternative method to calculate binding affinities from DSF data is the isothermal approach. The fraction of protein that is folded/unfolded is evaluated at a specific temperature close to the T_m, for a range of ligand concentrations. The fraction unfolded is fitted as a function of ligand concentration, dependent on both K_u, and K_d, the equilibrium constants for the protein unfolding and ligand binding reactions, respectively. Binding affinities calculated using the isothermal method have previously been reported to be in closer agreement with the corresponding values obtained from other biophysical methods[85].

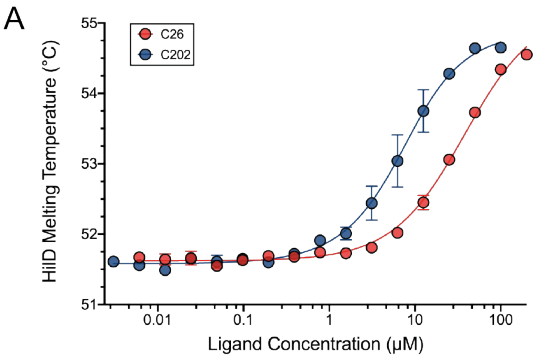
A potential limitation of the isothermal approach is that the K_d can only be determined at temperatures close to the T_m of the protein, at which significant fractions of protein exist in both the folded and unfolded states. The T_m of HilD is 52.2°C, therefore accurate K_d values can only be determined at temperatures > 50°C. This is significantly higher than physiological

conditions, and hence in the case of HilD this assay may yield considerably higher K_d values than would be determined at lower temperatures using other methods.

NanoDSF relies on intrinsic protein fluorescence at 330 and 350 nm to monitor unfolding. HilD only contains one tryptophan residue, and a HilD concentration of at least 5 µM was required to accurately monitor protein unfolding and determine melting temperatures. This is considerably higher than the protein concentration required for other methods and means this assay would be unsuitable for affinity determination of ligands with significantly high affinity. Nevertheless, this assay provides a reproducible and reasonably high throughput approach to calculate the affinities of different C26 derivatives to HilD and has the additional advantages of relatively low sample quantity demands and no requirement for protein labelling.

An additional complication of this NanoDSF-based affinity assay is the significant intrinsic fluorescence of C26 and several of the synthesised derivative compounds at 330 nm. Therefore, single wavelength data at 350 nm, which yielded clear unfolding profiles, was used to calculate melting temperatures. Due to the low solubility of the ligands in aqueous solvents, we could only achieve maximal ligand concentrations of 100 µM in assays whilst limiting the final DMSO concentration to 1%. Consequently, in the case of C26 and other low-affinity compounds, full saturation in binding curves could not be achieved, making data unsuitable for isothermal analysis. Here, we calculated $K_{d,app}$ values to compare the relative affinities of all compounds. For higher affinity compounds, we determined accurate K_d values at temperatures in the range of 50-56°C. Analysis was best at temperatures 50 and 52°C; at higher temperatures almost all HilD is unfolded, regardless of the ligand concentration.

Affinity of C26 was determined by temperature shift analysis, yielding a $K_{d,app}$ of 35.56 µM in agreement with *in vivo* IC_{50} values in the range 30-38 µM for C26 inhibition of P_{hilA} activation. The higher affinity compound C202 was determined to bind to HilD with a $K_{d,app}$ of 4.92 µM, whilst isothermal analysis yielded K_d values in the low micromolar range at 50-56°C (Figure 6.4). Both methods gave affinity values for C202 comparable to the K_d determined at 25°C by MST (13.8 µM, 1.5 seconds on-time) and the *in vivo* IC_{50} of 6.35 µM.



B

Apparent K_d

Compound	$K_{d,\ app}$ (μM)	± (95% CI)	R^2
C26	35.36	5.73	0.994
C202	4.92	1.34	0.991

C

Isothermal K_d (C202)

Temperature (°C)	K_d (μM)	± (95% CI)	R^2
50	1.58	0.79	0.962
52	2.43	0.82	0.965
54	4.79	2.36	0.901
56	9.49	5.60	0.825

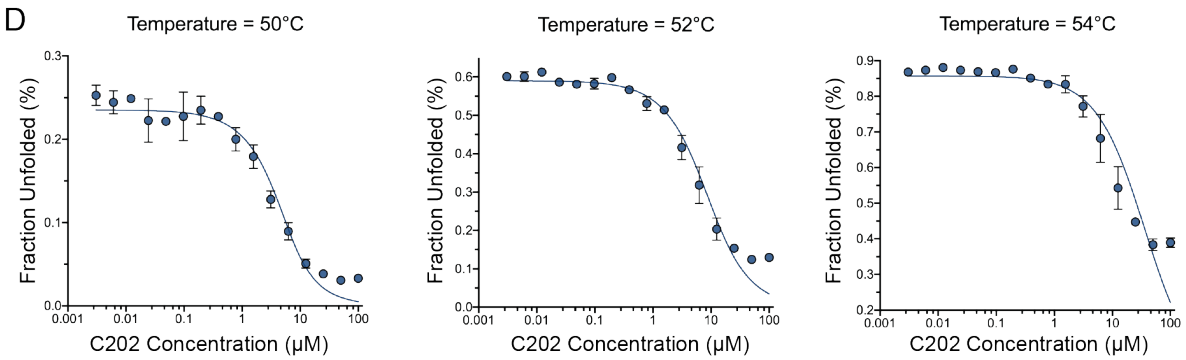


Figure 6.4. Binding affinities of compounds can be determined using a NanoDSF thermal shift assay. (**A**-**B**) Analysis of the change in HilD melting temperature with increasing ligand concentration yields $K_{d,app}$ values. (**C**-**D**) Isothermal analysis of NanoDSF data with $\Delta C_p = 0$, for the binding of C202 to HilD at selected temperatures. 5 μM of HilD was used for all runs, and melting temperatures calculated from changes in intrinsic fluorescence at 350 nm. Affinities were calculated from two independent measurements for each ligand concentration.

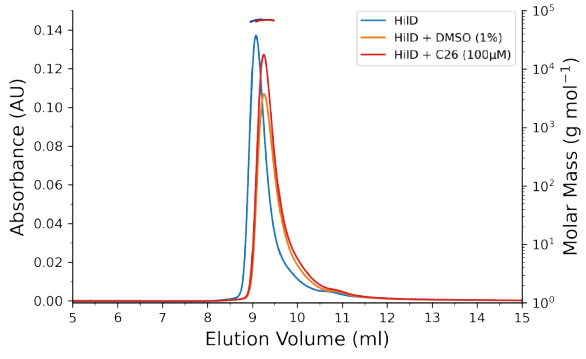
6.4 C26 Mechanism of Action

C26 was shown to downregulate *hilA* expression by preventing HilD from binding to the *hilA* promoter. There are several plausible mechanisms by which the binding of C26 to HilD could inhibit its ability to activate target genes. C26 may directly prevent HilD from binding to DNA, for example by binding to the DBD, or disrupt HilD dimerisation and prevent DNA-binding via an allosteric mechanism analogous to that described for LCFAs. C26 could destabilise HilD *in vivo*, making HilD more susceptible to degradation by Lon protease, as previously demonstrated for *cis*-2-UFAs[61]. C26 may interact with HilD via other negative regulators, such as HilE; if C26 were to bind to the HilD-HilE binding interface, it could act as a molecular glue, preventing dissociation of the HilD-HilE complex and trapping HilD in an inactive state. Given

our previous findings that DNA is not required for C26-binding *in vitro* and that C26 disrupts the formation of the HilD-DNA complex, an alternative mechanism whereby C26 regulates HilD once bound to DNA can be excluded.

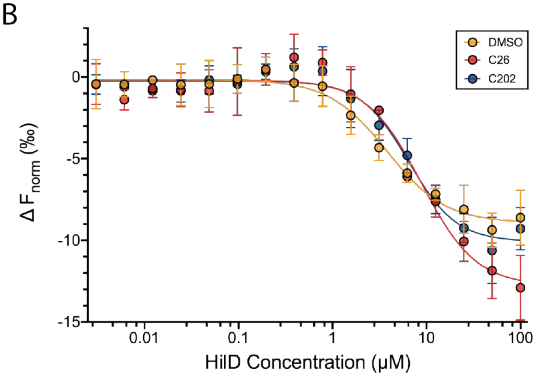
Both long chain fatty acids and HilE disrupt the homodimerisation of HilD. We investigated whether C26 also effects HilD dimerisation by first performing SEC-MALS experiments. Incubation of HilD with C26 (or a DMSO control) had no significant effect on oligomerisation, with HilD eluting as a single peak corresponding to a dimer (Figure 6.5 A). In an orthogonal quantitative approach, we repeated the MST dimerisation assay previously used to study the effect of other regulators on HilD dimerisation, incubating HilD with either C26 or C202. Unlike oleic acid and HilE, which both clearly disrupt the dimerisation of HilD, neither C26 nor C202 had any significant effect on the affinity of the HilD dimerisation interaction (Figure 6.5 B). The calculated $K_{d,dimer}$ values show no significant difference to that obtained for the DMSO control, and the high similarity between the C26 and C202 runs suggest this not a result mediated by the affinity of these ligands to HilD. Together these data agree that, unlike other negative regulators of HilD, these antivirulence compounds do not affect HilD dimerisation, despite inhibiting its DNA-binding activity.

A



Construct	Molecular Mass (kDa)	
	Theoretical (Dimer)	Measured
HilD	70.4	70.4 ± 2.0
HilD + DMSO	70.4	71.4 ± 2.8
HilD + C26	70.4	69.8 ± 1.8

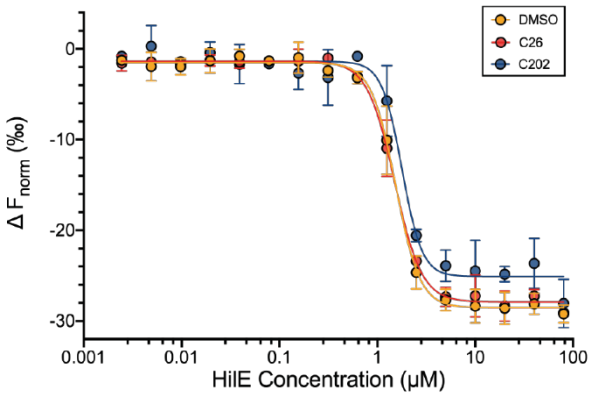
B



Ligand	$K_{d,dimer}$ (µM)
DMSO	4.52 ± 0.50
C26	12.50 ± 1.03
C202	8.10 ± 0.81

Figure 6.5. C26 does not affect the homodimerisation of HilD. (**A**) SEC-MALS profiles of HilD, in the absence and presence of C26. HilD (100 µM) was incubated with C26 (100 µM), with a final DMSO concentration of 1%. (**B**) MST binding curves for the homodimerisation of HilD. GFP-HilD (50 nM) was incubated with DMSO (1%, yellow), C26 (100 µM, red) or C202 (100 µM, blue), and increasing concentrations of HilD. Affinities were calculated from changes in thermophoresis at an MST on-time of 1.5 seconds. Standard deviation in both (A) and (B) was calculated from three replicate experiments.

We next investigated whether a possible mechanism of action of C26 involves HilE and stabilising the HilD-HilE interaction to repress HilD activity *in vivo*. The MST assay for the binding of HilD to HilE was repeated, and HilD was incubated with C26/C202/DMSO prior to incubation with HilE (Figure 6.6). HilD bound to HilE with similar affinity in the presence of C26 or DMSO (1.71 and 1.70 µM, respectively), demonstrating that C26 does not affect the negative regulation of HilD by HilE. This result also suggests that the binding of C26 does not result in large conformational changes in HilD structure, as HilD can still bind to HilE whilst in complex with C26. In contrast, oleic acid results in large changes in HilD conformation and flexibility, which impends subsequent binding to HilE. The difference in effect of C26 and oleic acid on HilD dimerisation and HilE binding indicates that the two ligands may bind to HilD with different binding modes.



Ligand	K_d (µM)	
DMSO	1.70	± 0.50
C26	1.71	± 0.50
C202	2.09	± 0.70

Figure 6.6. C26 does not affect the interaction between HilD and HilE. GFP-HilD (50 nM) was incubated with DMSO (1%, yellow), C26 (100 µM, red) or C202 (100 µM, blue), and subsequently increasing concentrations of HilE. An MST on-time of 1.5 seconds was used to plot changes in thermophoresis and calculate binding affinities. Calculated K_d values for the interaction between HilD and HilE, were determined from three repeat experiments.

6.5 Binding Mode of C26 to HilD

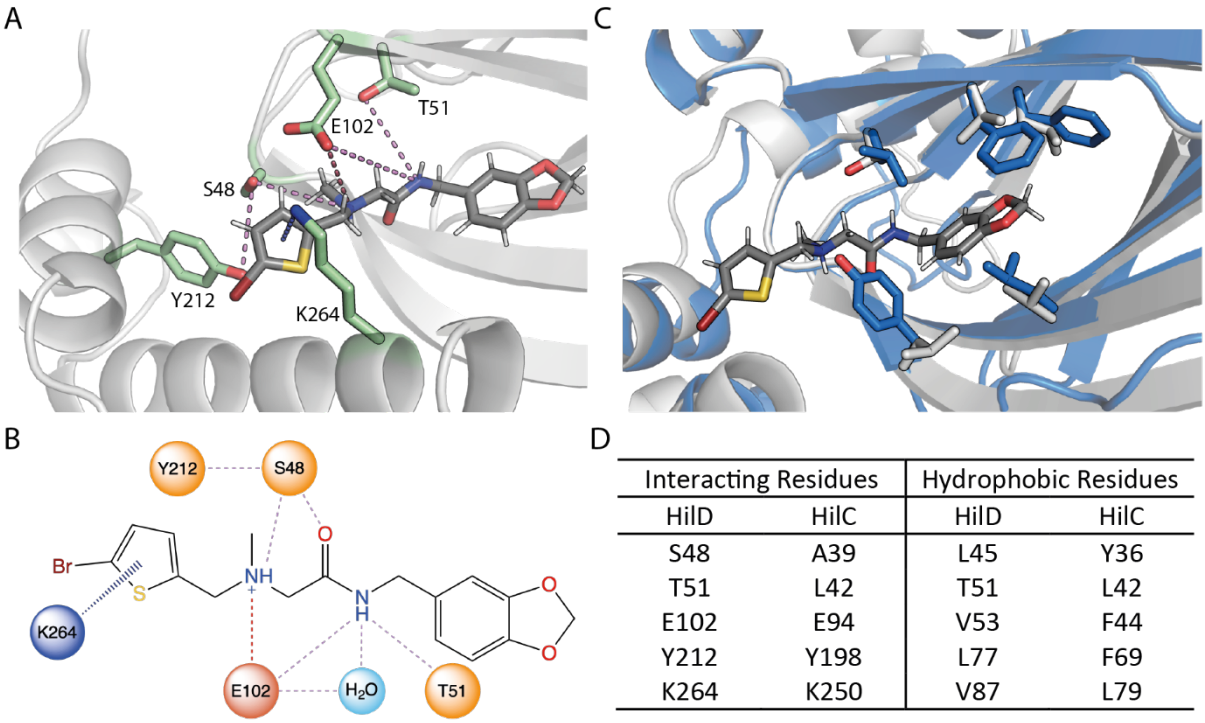
6.5.1 Computation Model

We next sought to determine the binding mode of C26 to HilD. Structural characterisation of HilD proved challenging and co-crystallisation trials with both C26 and other higher affinity analogs did not lead to any further success. The AlphaFold model was used for molecular

docking of C26 (and selected analogs) to HilD and performed by Thales Kronenberger, as described for oleic acid. Docking was performed for both protonated and unprotonated forms of C26 and compound C202. C26 was protonated at the tertiary amine of the backbone, as would occur under physiological conditions and those used in *in vitro* assays, with docking performed for each of the resultant stereoisomers.

In all models, C26 is predicted to bind to the same pocket as oleic acid. This pocket is at the interface of the NTD and DBD, and C26 is predicted to form specific interactions with residues of both domains (Figure 6.7 A-B). C26 would act as a cross linker between the two domains, resulting in a more closed conformation of HilD and rationalising the observed stabilisation effect of C26 on HilD. This is comparable to the effect seen for the binding of oleic acid to HilD. C26 adopts two main conformations in simulations: a linear conformation, protruding from the barrel, and a bent conformation nested within the barrel. The linear conformation results in more interactions between C26 (particularly of the thiophene group) and surrounding residues of the HilD DNA-binding domain.

The benzodioxol moiety is largely buried within the cupin barrel of the N-terminal domain. Beta strands β1, β2, β6 and β7 form the opening of the pocket into which C26 is bound. Hydrophobic residues situated on these beta strands line the inside of the pocket and form hydrophobic interactions with the benzodioxol moiety. When C26 is protonated, residue E102 is predicted to form hydrogen bonding interactions with both amines of the C26 backbone. Direct hydrogen bonds are formed between E102 and the tertiary amine, whilst both direct and water-mediated hydrogen bonds are formed with the secondary amide. Residues S48 and T51 are additionally predicted to form hydrogen bonds with the C26 backbone at certain orientations of C26 during the time course of the simulation: S48 with the tertiary amine and amide carbonyl; T51 with the amide NH group. Residue K264, situated on helix α4 of the DBD is predicted to form a cation-π interaction with the thiophene group of C26. This interaction was predicted in all simulations and would be critical in the binding of C26 between the two HilD domains. Additionally, halogen bonding interactions are predicted between the thiophene Br substituent and residues of the DBD helix α4. Whilst not predicted to directly interact with the thiophene moiety, residues Y212 and S216, situated on DBD helix α1, are both involved in a hydrogen bonding network involving S48 and water molecules, to further stabilise the C26-bound conformation.



D

Interacting Residues		Hydrophobic Residues	
HilD	HilC	HilD	HilC
S48	A39	L45	Y36
T51	L42	T51	L42
E102	E94	V53	F44
Y212	Y198	L77	F69
K264	K250	V87	L79

Figure 6.7. Predicted binding mode of C26 to HilD. (**A**) Computational model of HilD in complex with C26, highlighting the potential binding pocket of C26. Residues predicted to specifically interact with C26 are shown as sticks and highlighted in green. (**B**) Schematic showing predicted specific interactions between C26 atom and residues of HilD. Displayed interactions occurred in more than 15% of the simulation time in the selected trajectory (0.00-1004.00 nsec). HilD residues are coloured according to their charge (positive: blue; negative: orange; polar: yellow). In both (A) and (B), contacts are coloured by the nature of the interaction (pink: hydrogen bond; red: salt bridge; blue: cation-π). (**C**) Comparison of the predicted binding pockets of HilC and HilD. The AlphaFold structure of apo HilC (blue) is overlain with the docked AlphaFold model of HilD in complex with C26 (white, with C26 coloured by chemical element). Non-conserved hydrophobic residues lining the inside of the binding pocket are shown as sticks. (**D**) Selected residues of HilD identified as specifically interacting with C26 (left) or lining the hydrophobic binding pocket (right) are listed alongside the corresponding residues of HilC, identified from a structural alignment of the two proteins.

Interestingly, most of these residues are spatially conserved between HilD and HilC despite C26 interacting specifically with HilD. To compare the respective binding pockets of the two proteins, we performed a structural alignment of HilD and HilC (using HilD residues Q38-T105 and HilC residues N28-Y97) with PyMOL. In HilC, several of the hydrophobic residues lining the inside of the binding pocket contain larger side chains than the corresponding aligned residues in HilD (Figure 6.7 C-D). This would result in a smaller pocket in HilC that would likely

be inaccessible to C26. The linear, flexible structure of long chain fatty acids, compared to the bulkier and rigid C26 structure, would enable them to also bind to the HilC binding pocket, rationalising the difference in binding specificity of the two groups of ligands.

6.5.2 Hydrogen-deuterium Exchange (HDX)

We used HDX-MS to experimentally validate this model of the HilD-C26 complex and compare the binding of C26 with that of oleic acid. For HDX experiments, HilD was incubated with a four-times excess of C26 or the higher affinity compound C202, as performed previously for oleic acid.

C26 and C202 resulted in almost identical HDX profiles, with the same changes in HDX behaviour (increased or decreased D-incorporation) observed at the same residues for both samples. C202 showed slightly stronger changes in D-incorporation than C26 for some regions of HilD (residues 259-265, 302-305), likely due it's higher affinity for HilD (Figure 6.8 A). Full HDX difference maps, showing changes in HDX across the entire HilD sequence, are displayed in Figure S3.2 (Appendix 3).

These results are consistent with our computational model and confirm that C26 binds to the same pocket on HilD as oleic acid, with most changes in HDX clustered in the same region for both ligands (Figure 6.8 B-C). Residues T41-F50 (strand β1) and A257-L265 (DBD helix α4) display lower D-uptake in the presence of C26/C202 (Figure 6.8 A), mirroring the effect seen for oleic acid. Decreased D-uptake is also observed for the DBD helices α1 (I214-S218) and α5 (V276-Y282), which form HTH1 and HTH2, respectively. However, overall changes in HDX are less pronounced on both HTHs for C26/C202 binding than for oleic acid. This is particularly noticeable for HTH1, which is further from the identified binding site. Whilst this could indicate C26 only causes significant conformational changes in HTH2 upon binding, it may also be due to a lower affinity, and hence incomplete saturation of HilD, of C26 compared to oleic acid.

Interestingly, there are also multiple regions of increased D-uptake (i.e. decreased protection) upon binding of C26/C202. Increased D-uptake is observed at residues D99-P104 (β7 of the N-terminal barrel structure), I253-S256 (at the N-terminus of helix α4 of the DBD) and S301-

M306 (DBD helix α7). These changes are likely due to conformational changes in HilD structure upon ligand binding, and the increased HDX at helix α7 indicates a re-arrangement of HTH2 that results in the observed disruption of DNA binding.

Whilst HDX profiles show that C26/C202 bind to the same pocket of HilD as oleic acid, as predicted by our model, the opposite HDX behaviour in certain regions (particularly at HTH2) strongly indicates that the two groups of ligands bind to this pocket with different binding modes. This disparity in HDX behaviour is unlikely due to differences in affinity, as is seen in the differences between the HDX difference profiles for C26 and C202 (Figure S3.2). Whilst the relative D-uptake is generally higher for C202 than C26, owing to its higher affinity, the changes in HDX behaviour (increased/decreased D-uptake) compared to apo HilD are identical for both ligands. More detailed elucidation of the respective binding modes is impeded by the consistently low HDX for the residues forming the cupin barrel structure.

6.5.3 Competitive Assay

In addition to HDX-MS experiments, we also performed a competitive MST assay to determine whether competition exists between C26 and LCFAs for binding to HilD. For this assay, we required a fluorescent fatty acid reporter ligand for detection in MST assays. The SAR analysis of LCFAs showed a minimum chain length of 16 C-atoms is required for binding to HilD, with affinity increasing with chain length. We purchased the BODIPY™FL C16 (Invitrogen™) fatty acid, the longest commercially available fluorescent fatty acid we could find. This fatty acid contains the BODIPY fluorophore at the end of the fatty acid chain, rather than in the centre, and is therefore more resemblant of the tested LCFAs which showed binding to HilD.

To determine the affinity of this reporter ligand to HilD, we performed an MST binding experiment, varying the concentration of HilD. Titration of HilD resulted in specific changes in fluorescence of the reporter, from which we could calculate a K_d of ~75µM. Interestingly, the binding of HilD resulted in an inversion in the thermophoretic behaviour of the reporter ligand. A similar change in thermophoretic behaviour was previously reported for other BODIPY-containing compounds, highlighting the advantages of these compounds as reporter ligands for competitive MST measurements[155]. When free in solution, BODIPY FL C16 shows

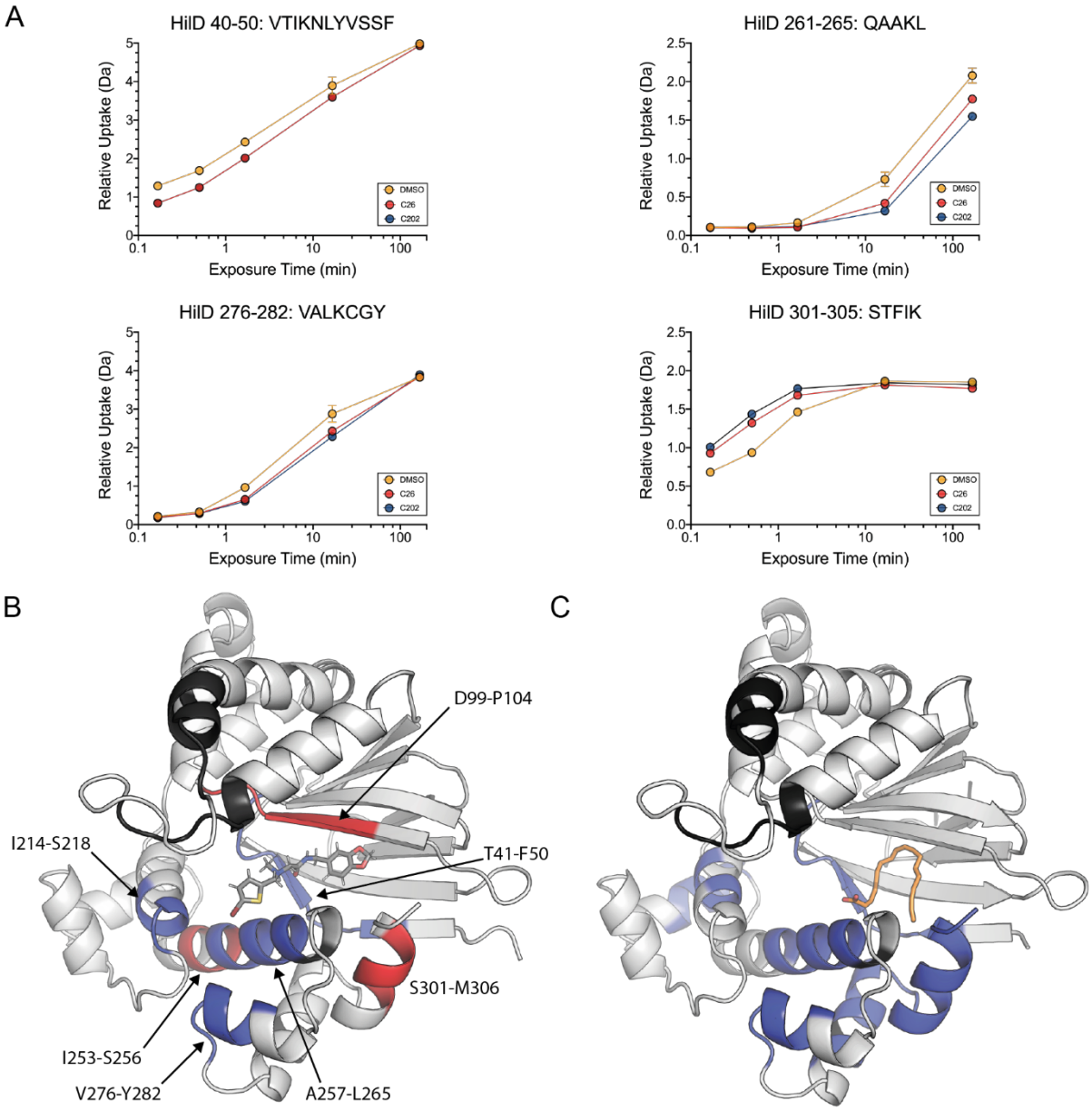


Figure 6.8. C26 binds to the same binding pocket of HilD as oleic acid. (**A**) Hydrogen-deuterium exchange profiles of representative HilD peptides, highlighting changes in deuterium uptake in specific regions of HilD. HilD samples were incubated with 1% DMSO (yellow), C26 (red) or C202 (blue). Data represent the mean and standard deviation of results from three replicate experiments. (**B**-**C**) Model of the HilD-C26 and HilD-oleic acid complexes, with HilD coloured according to changes in HDX upon incubation with the respective ligand, relative to a DMSO control sample. HilD is coloured as follows: white, no change in HDX; blue, decreased HDX; red, increased HDX; black, no information could be obtained due to a lack of peptide coverage in this region of the protein sequence.

negative thermophoretic behaviour, diffusing from the cold to hot region where the detector is located (hence resulting in an increase in detected fluorescence). However, once bound to HilD, the reporter-HilD complex instead displays positive thermophoretic behaviour, diffusing away from the detector (Figure 6.9 A). This clear change in behaviour of the BODIPY FL C16 ligand in the bound/unbound state allowed us to perform qualitative competition experiments despite its low affinity to HilD.

Competitive measurements were performed for C26, the higher affinity compound C203, and the ToxT inhibitor virstatin for comparison. MST runs were performed with 200 nM BODIPY FL C16 and 50 µM HilD, within the range of the calculated K_d to ensure sufficient reporter saturation, and varying concentrations of the competing ligand. At lower competing ligand concentrations HilD remains bound to the reporter, which hence shows positive thermophoretic behaviour as described above. With increasing competitor concentration, the reporter ligand is outcompeted and instead exhibits negative thermophoresis (Figure 6.9 B). Fitting a simple dose-response curve to the change in normalised fluorescence yielded apparent EC_{50} values of 1.28, 1.32 and 488.9 µM for C26, C203 and virstatin, respectively. Accurate affinity values could not be calculated due to C26 and C203 having significantly higher affinity than BODIPY FL C16 for HilD. However, the distinct changes in thermophoresis of the reporter ligand clearly show a competitive effect between the two sets of ligands, supporting the conclusion that C26 binds to the same pocket of HilD as LCFAs.

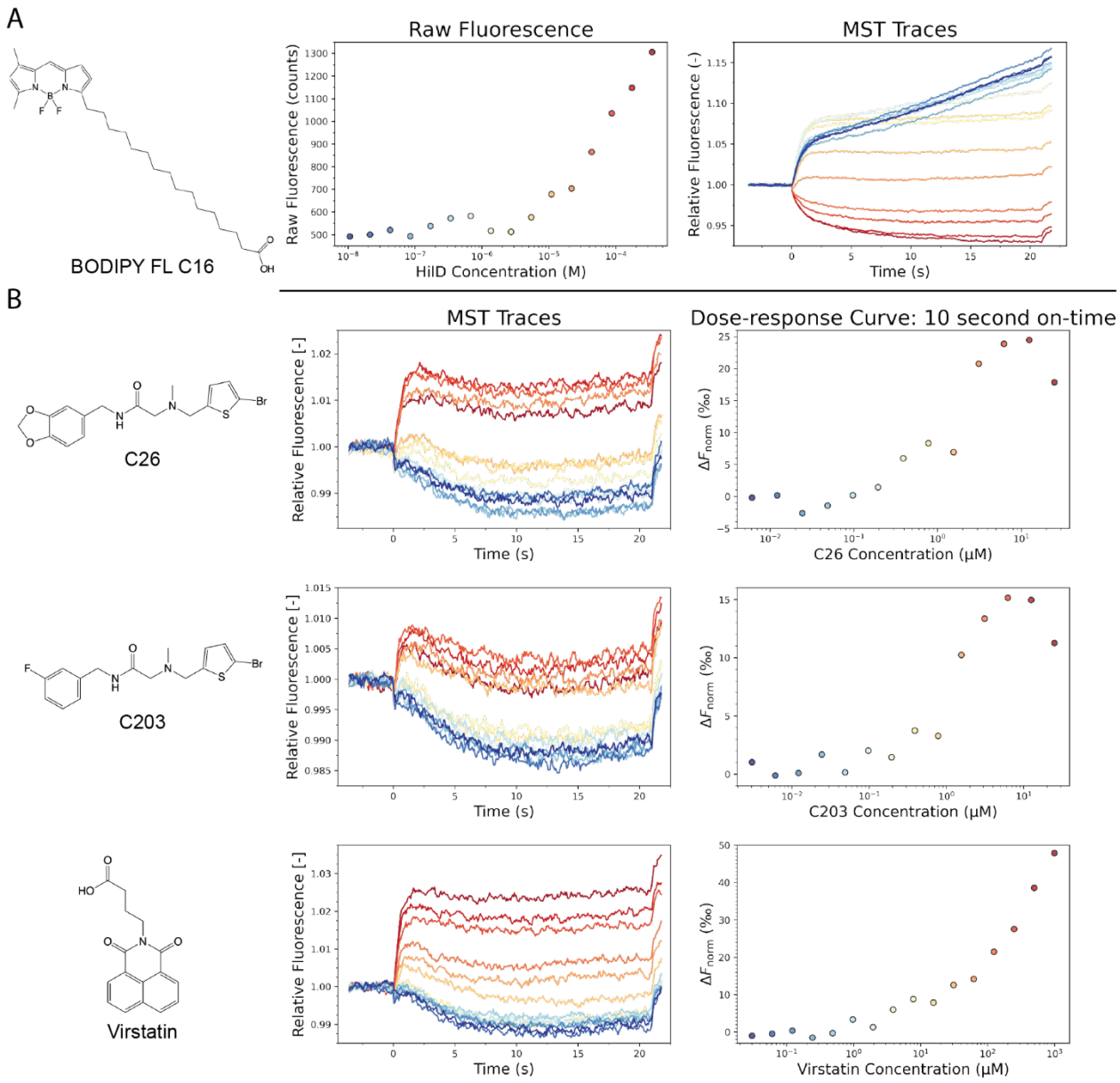


Figure 6.9. C26 and LCFAs compete for binding to HilD. (**A**) Affinity of the BODIPY FL C16 reporter was calculated from changes in ligand fluorescence with increasing HilD concentration (left). Binding of HilD resulted in a reversal in the thermophoretic behaviour of BODIPY FL C16, seen in the corresponding MST traces (right). (**B**) MST traces (left) for the out-competition of the BODIPY FL C16 reporter by C26, C203 and virstatin. MST traces are coloured according to concentration of the competing ligand, as shown in the corresponding dose-response curves (right). Dose-response curves were plotted from changes in normalised fluorescence (ΔF_{norm}) at an MST on-time of 10 seconds and display representative data from one experimental replicate. Two replicates were performed for each competing ligand to calculate apparent EC_{50} values.

6.6 Structure Activity Relationship of C26

After characterizing the interaction between C26 and HilD, we next performed a structure-activity relationship (SAR) of C26. We sought to identify which chemical moieties are critical for C26 binding to HilD, and those that present opportunities for further hit optimization. In this first round of SAR analysis, > 110 analogs of C26 were tested, focusing on individual changes in each chemical moiety in turn.

We investigated the importance of the benzyl and thiophene moieties, varying the identity of the aromatic group at each end of the C26 molecule, as well as the substituents on each aromatic group. We additionally probed alterations to the central region of C26: the importance of the tertiary amine and the amide bond, varying the length of the alkyl linkers between each functional group, substitution of the tertiary amine, and the introduction of branching or cyclic moieties into the backbone.

C26 analogs were tested for binding to HilD *in vitro* using the EMSA and NanoDSF assays established for characterising the HilD-C26 interaction to orthogonally screen different compounds. The NanoDSF assay provided a relatively high-throughput method, from which comparative affinities of the different compounds could be determined. K_d values were calculated from NanoDSF data using the isothermal method, with $\Delta C_p = 0$ at 2°C temperature intervals between 50 and 56°C. Throughout this chapter, the reported K_d's were determined at 52°C, which gave the best fit and signal-to-noise ratio for most of the tested compounds. K_d values calculated at other temperatures are displayed in Table S4 (Appendix). Combined with the *in vivo* P_{hilA} activation assay, we can also estimate potentiation efficiency to gain a comprehensive overview of different activity parameters for all newly synthesised compounds.

6.6.1 Benzodioxol

We first investigated changes in the benzodioxol moiety, and the importance of the dioxolane functional group. Compounds C178 and C182 differ from C26 only in the substituents of the methylene bridge of the dioxolane ring. C178 contains two methyl (CH_3) groups at the

dioxolane linker, and hence has increased steric bulk at this position. This compound showed comparable affinity to C26 in both EMSAs and the NanoDSF assay. In contrast, compound C182, which contains a difluoromethylene (-CF_2-) dioxolane linker, shows a clear inhibition of HilD activity in EMSAs, indicative of considerably higher affinity compared with that of C26. To verify the increased potency of this compound, it was synthesised twice, with both synthesis batches showing high affinity for HilD (Figure 6.10 A). The NanoDSF assay supported the increased affinity of C182 compared to C26, with the melting temperature analysis yielding an apparent K_d of 3.44 µM for batch 2 (Figure 6.10 C). The higher affinity of C182 enabled us to also calculate the true K_d of this compound, with values of 1.08 and 1.58 µM at 50 and 52°C, respectively (Figure 6.10 D). The increased affinity of this compound was also

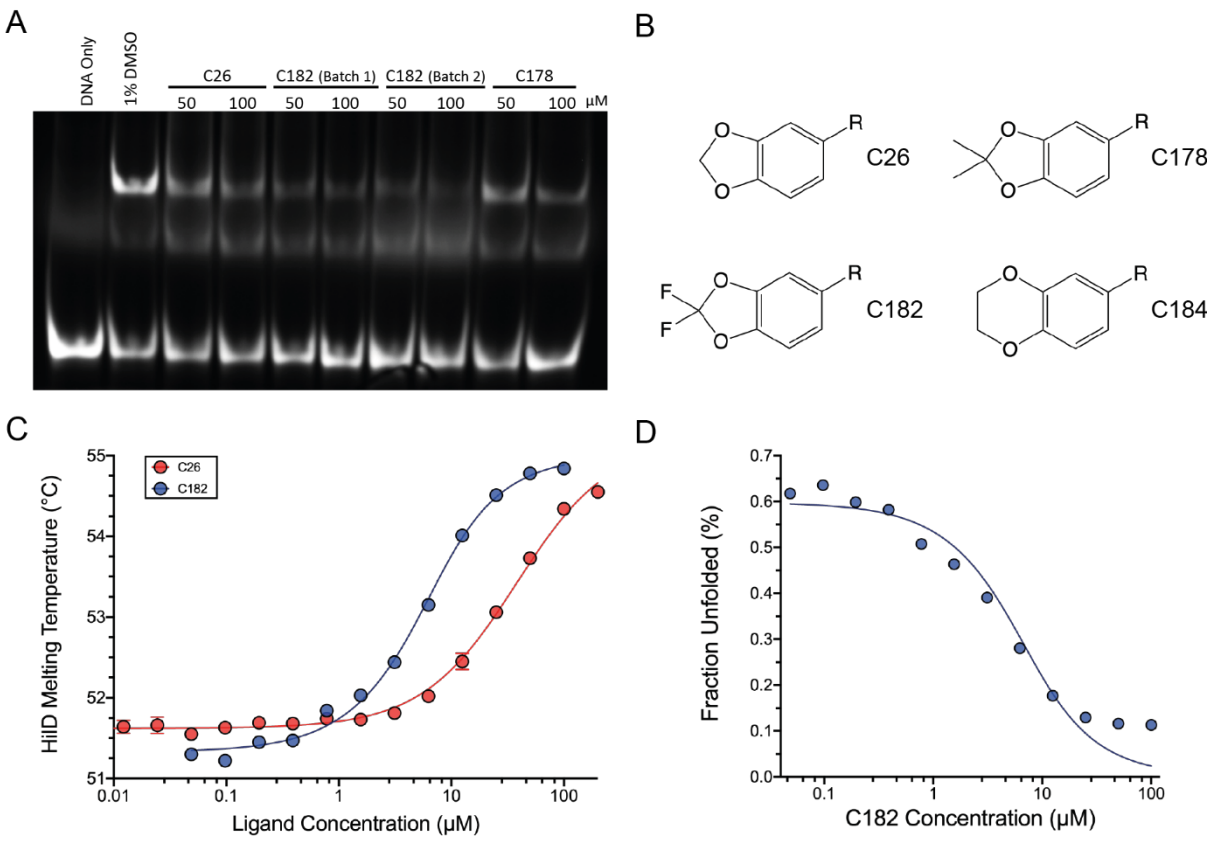
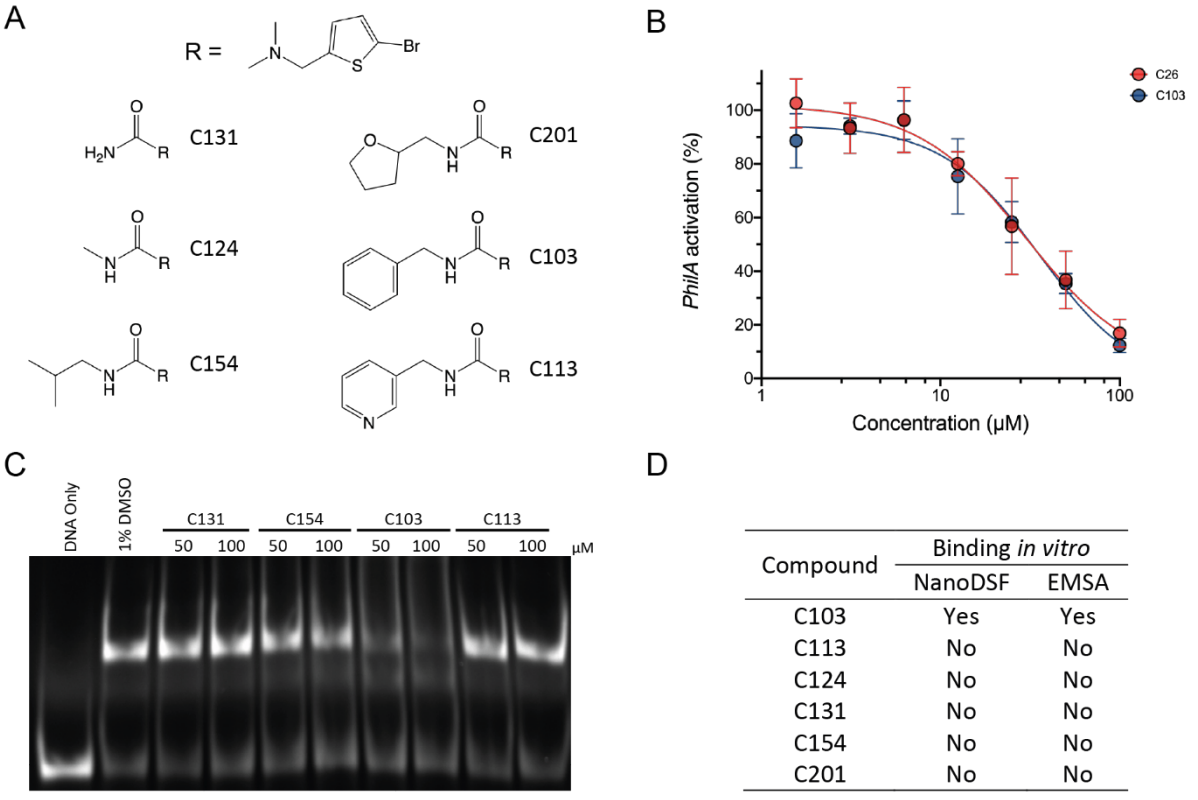


Figure 6.10. Substitution of the dioxolane of C26 affects binding affinity to HilD. (**A**) EMSA for the inhibition of the HilD-DNA complex by different compounds. The two indicated batches of C182 correspond to two separate chemical syntheses of this compound. (**B**) Simplified chemical structures of compounds with variations in the dioxolane linker. (**C**) Changes in the melting temperature of HilD with increasing concentrations of C26/ C182. (**D**) Isothermal analysis with $\Delta C_p = 0$, for C182 binding to HilD at 52°C.

confirmed *in vivo*, with an IC_{50} of 4.83 µM for this synthesis batch of C182. The corresponding benzodioxan compound (C184), in which the phenyl substituents form a 6-membered, rather than 5-membered, ring showed no activity either in EMSAs nor *in vivo*, although stabilisation of HilD was observed in the NanoDSF assay suggesting residual, low affinity binding to HilD.

We next compared the binding of C26 analogs completely lacking the dioxolane functional group (Figure 6.11 A). Compound C103 inhibited HilD DNA-binding in EMSAs and stabilised HilD. Although the low binding affinity of C103 prevented the determination of accurate K_d values from the NanoDSF assay, C103 showed comparable inhibition of P_{hilA} activation to C26 (Figure 6.11 B) and the calculated IC_{50} of 35.7 µM is within the range of *in vivo* affinities determined for different synthesis batches of C26 (30-38 µM).



Compound	Binding *in vitro*	
	NanoDSF	EMSA
C103	Yes	Yes
C113	No	No
C124	No	No
C131	No	No
C154	No	No
C201	No	No

Figure 6.11. The dioxolane functional group is not important for C26 binding. (**A**) Chemical structures of compounds with varying substituents at the C26 amide. (**B**) Comparison of the effect of increasing concentrations of C26 or C103 on P_{hilA} activation. (**C**) EMSA showing the inhibition of HilD DNA-binding by compounds with variations in the substituent of the secondary amide. (**D**) Summary of screening of ligands for binding to HilD. For the NanoDSF screen, 20 µM HilD was incubated with 50 µM of the respective compound, and a shift in the T_m of HilD was interpreted as binding.

This demonstrated the dioxolane group is not required for binding to HilD, and we then tested compounds also lacking the benzyl functional group. Compounds containing either a primary amide or assorted smaller aliphatic and aromatic groups in place of the benzyl group did not display any binding to HilD. Replacement of benzene with pyridine also impeded binding, highlighting the importance of the benzyl group at this position for activity.

Having identified that the benzyl group, but not the dioxolane, is critical for binding to HilD, we investigated the relationship between substitution of the aromatic ring and ligand binding affinity. Both the identify and position of aromatic substituents were varied, along with comparison of mono- vs di-substituted compounds.

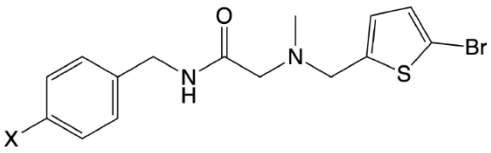
We first compared the binding of several compounds with substituents in the para- position of the benzene ring, relative to the rest of the molecule (Figure 6.12 A). Compounds containing a halogen atom at this position displayed markedly higher affinity binding than C26, showing complete inhibition of HilD binding to the *hilA* promoter (Figure 6.12 B) and affinities < 10 μM, as determined by NanoDSF. The higher affinity of these compounds was verified by *in vivo* P_{hilA} activation experiments. Having identified these compounds as higher affinity binders, we also utilised them in the previously discussed mechanism of action and binding mode experiments for C26. Compounds containing methyl (CH_3), methoxy (OCH_3) or trifluoro-methoxy (OCF_3) substituents all also bound to HilD, as did C189, which has an alkyne-containing substituent. However, compounds with larger functional groups (t-butyl, phenyl and morpholine) did not display activity in any of our binding assays, showing that substitution of compounds at this position is limited to only small and linear functional groups and suggests there is only a very narrow opening in this region of the binding pocket.

Interestingly, whilst both C110 (para-OCH_3) and C251 (para-OCF_3) bound to HilD, only C251 inhibited P_{hilA} activation *in vivo*. Collaborators in the Brönstrup group previously found that the fluorinated C182 showed vastly increased (ca. 5-fold) cellular uptake compared to C26, suggesting that fluorination may increase cellular uptake of these compounds.

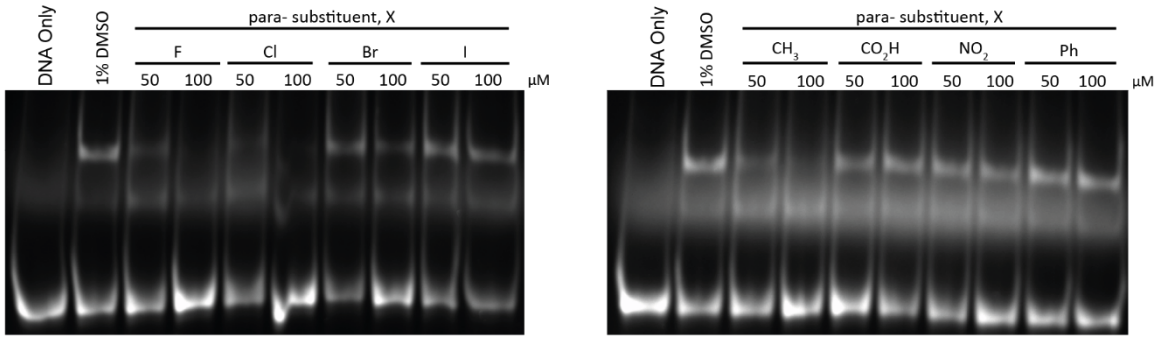
Compounds with other para-substituents (CO_2H, NO_2, SO_2CH_3) do not show binding to HilD. Each of these groups are electron-withdrawing, making the aromatic ring less electron-rich and hence less reactive. Halogen and trifluoromethyl (CF_3) substituents are also electron withdrawing, whilst methyl (CH_3) and methoxy (OCH_3) substituents are both electron

donating, suggesting that there is no correlation between the electron density of the benzyl aromatic ring and ligand binding to HilD. This is in consistent with our model in which the benzyl moiety is bound within the cupin barrel, surrounded by predominantly hydrophobic residues, and forms no specific interactions with residues of HilD.

A



B

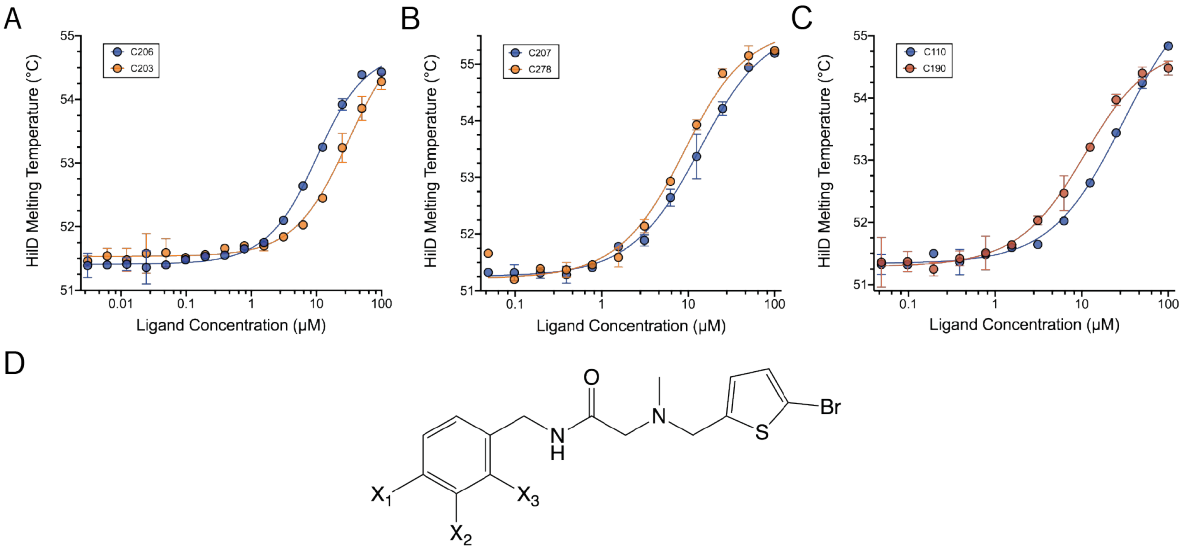


C

Compound	Para-substituent	NanoDSF		EMSA (Binding?)	*in vivo* IC_{50} (µM)
		$K_{d,app}$ (µM)	K_d, 52°C (µM)		
C103	H	> 100	-	b	35.7
C206	F	7.62	3.66	b	13.1, 14.4
C202	Cl	4.92	2.43	b	8.81, 6.35
C205	Br	6.89	4.50	b	10.6
C204	I	2.47	5.49	b	7.68
C207	CH_3	10.98	2.69	b	29.9
C247	CF_3	-	-	-	13.2
C249	$C(CH_3)_3$	n.b.	n.b.	n.b.	n.b.
C110	OCH_3	27.52	6.16	b	>100
C251	OCF_3	3.70	6.71	-	7.44
C189	OC≡CH	12.55	10.37	b	>100
C111	Phenyl	n.b.	n.b.	n.b.	n.b.
C162	Morpholine	n.b.	n.b.	-	n.b.
C248	SO_2CH_3	n.b.	n.b.	n.b.	n.b.
C191	NO_2	-	-	n.b.	n.b.
C232	CO_2H	n.b.	n.b.	n.b.	n.b.

Figure 6.12. Effect of para-substituents of the benzene ring on compound affinity. (**A**) General structure of tested ligands, which differ in the identity of substituent X. (**B**) Representative EMSAs showing the inhibition of HilD DNA binding by selective compounds with a para- substituent. (**C**) Overview of binding for compounds with varying para-substituents. Compounds were screened for binding in EMSAs (b: binder, n.b: non-binder), and affinity determined by NanoDSF. *In vivo* IC_{50} values were determined from the P_{hilA} activation assay.

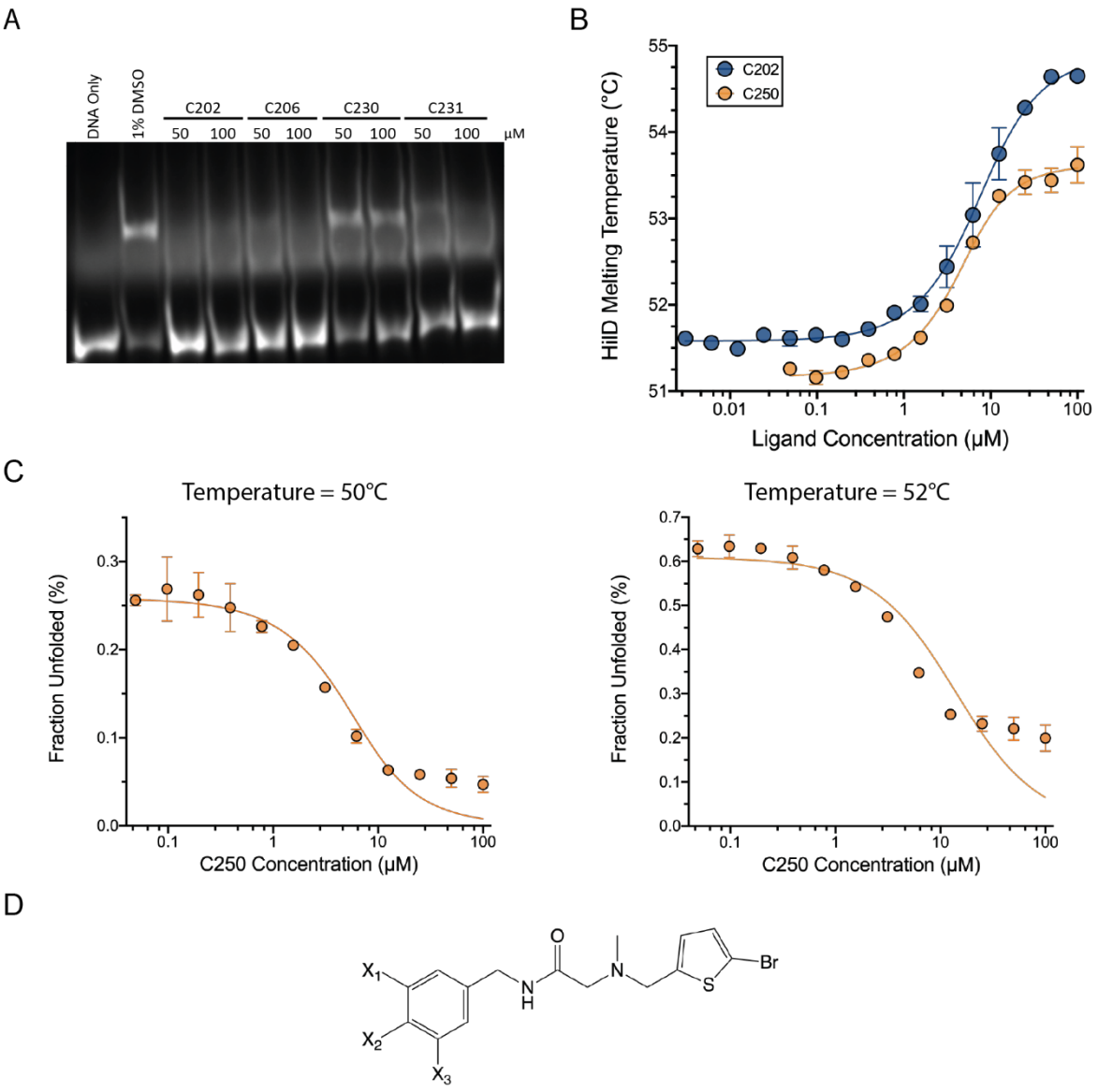
We selected three of the above compounds with para-substitutions (C206, C207, C110) to investigate the effect of altering the position of phenyl substituents on binding to HilD. Changing the position of the substituents to either the meta- (F, CH_3) or ortho- (OCH_3) positions did not significantly affect the binding affinity of these compounds, compared to the corresponding para-substituted compounds (Figure 6.13). For comparison, we also tested the binding of compound C192, containing a meta- NO_2 substituent, which, like the para-NO_2 substituted C191 did not bind to HilD. This shows that whilst the identity of aromatic substituents is critical for binding, the position of substituents the phenyl ring does not notably impact binding affinity.



Compound	X_1	X_2	X_3	NanoDSF		EMSA	*in vivo*
				$K_{d,app}$ (μM)	K_d, 52°C (μM)	(Binding?)	IC_{50} (μM)
C206	F	H	H	7.62	3.66	b	13.1, 14.4
C203	H	F	H	28.55	9.13	b	16.5
C207	CH_3	H	H	10.98	2.69	b	29.9
C278	H	CH_3	H	6.34	2.49	-	7.21
C110	OCH_3	H	H	27.52	6.16	b	> 100
C190	H	H	OCH_3	8.20	3.65	b	> 100
C191	NO_2	H	H	-	-	n.b	> 100
C192	H	NO_2	H	-	-	n.b	> 100

Figure 6.13. Position of the substituents on the ring does not significantly affect ligand binding. Changes in the melting temperature of HilD, as determined using NanoDSF, with increasing concentrations of compounds with the following phenyl substituents: (**A**) para- vs meta- F, (**B**) para- vs meta- CH_3, (**C**) para- vs ortho- OCH_3. (**D**) General chemical structure and summary of binding activity for these compounds, highlighting the importance of the position of substituents on binding affinity.

The highest affinity compounds identified so far contained halogen substituents at the para- or meta- positions of the aromatic ring. We compared the binding of compounds containing two halogen substitutions on the phenyl ring with the corresponding mono-substituted compounds. The 3,4-dichloro (C230) and 3,4-difluoro (C231) compounds bound to HilD, although neither displayed increased binding affinity (Figure 6.14 A, D). C231 showed comparable binding affinity to the 4-fluoro C206 (5.58 and 3.66 µM, respectively), whilst the affinity of C230 was notably lower than that of C202 (with respective K_d values of 15.5 and 2.43 µM determined at 52°C). On the other hand, the 3,5-dichloro compound C250, with K_d values of 2.22 and 4.48 µM at 50 and 52°C, respectively, is one of the highest affinity compounds we identified during this SAR analysis, with high affinity both *in vitro* and *in vivo* (Figure 6.14 B-D). This suggests there is potential for the further improvement of binding affinity, by optimising the substituents of compounds containing > 1 substituent on the phenyl ring.



Compound	X_1	X_2	X_3	NanoDSF		EMSA	*in vivo*
				$K_{d,app}$ (µM)	K_d, 52°C (µM)	(Binding?)	IC_{50} (µM)
C230	Cl	Cl	H	9.50	15.50	b	8.13
C231	F	F	H	5.30	5.58	b	45.7
C250	Cl	H	Cl	1.97	4.48	b	3.76

Figure 6.14. Binding of compounds with multiple phenyl substituents. (**A**) EMSA comparing the binding of the para-substituted compounds C202 and C206 with the 3,4-di-substituted compounds C230 and C231. (**B**) Changes in the melting temperature of HilD with increasing concentrations of C202 (para-Cl substituent, blue) or C250 (3,5-dichloro, yellow). (**C**) Isothermal analysis with $\Delta C_p = 0$, for binding of C250 to HilD at 50 and 52°C. (**D**) General chemical structure and determined activity of compounds with multiple halogen substituents.

6.6.2 Linker

The central backbone of C26, connecting the two terminal aromatic groups, is six atoms in length and contains both a secondary amide and tertiary amine. We first investigated variations in the alkyl linker of the benzyl group, which we previously found to be crucial for binding to HilD. Compounds either lacking (C129) or containing a longer (C195) alkyl linker between the benzodioxol and the amide nitrogen do not exhibit any binding to HilD. Compounds C238 and C239, which contain a phenyl and benzyl ether, respectively, did not show any inhibition of HilD DNA-binding in EMSAs (Figure 6.15 A-B). Compound C238 did stabilise HilD in the NanoDSF assay and, although affinity was too low to be determined using this assay, also showed *in vivo* activity with comparable affinity to C103 (IC_{50}: C103, 35.7 µM; C238, 28.1 µM). Both C103 and C238 contain a linker between the phenyl and amide groups one atom in length, indicating this to be the required linker length at this position.

We next tested whether branching at the alkyl of the benzyl group could be tolerated. Compounds C276 and C277 both contain a single methyl substituent at this position and correspond to the two stereoisomers formed by the introduction of a chiral centre at this carbon. Interestingly, C276 (the S-isomer) binds to HilD with higher affinity ($K_{d,app}$: 19.11 µM; K_d, 52°C: 6.73 µM) than C26, whereas C277 (the R-isomer) does not exhibit any binding at all (Figure 6.15 C-D). Specificity of binding for only the S-isomer was also confirmed for corresponding compounds containing a para-Cl substituent instead of the dioxolane (akin to C202) or lacking any aromatic substituents (as in C103).

In the absence of a high-resolution structure, this result provides an attractive starting point for the further optimisation of analogs with potential higher affinity. In our model, the benzodioxol moiety is buried within the hydrophobic barrel of the NTD. If there is significant empty space within this pocket, the introduction of branched alkyl substituents could occupy this space and potentially exclude water from the hydrophobic pocket, driving ligand binding. Determining whether the HilD binding pocket can accommodate ligands with larger aliphatic groups (trifluoromethyl (CF_3), ethyl (CH_2CH_3), tert-butyl ($C(CH_3)_3$)) at this position is a current focus of our investigation.

The amide bond at the centre of the C26 backbone is critical for binding to HilD, highlighted by the lack of activity of compound C274, which lacks the amide carbonyl (Figure 6.16 A). C26 contains a single methylene group between the backbone amide and tertiary amine. Compounds C228 and C177 contain two and three methylene groups, respectively, at this position. C228 inhibited HilD DNA-binding akin to C26, whereas C177 did not show binding (Figure 6.16 B). Compounds containing cyclic moieties in the backbone, including C194, C199 and C236, do not bind to HilD (Figure 6.16 B-D). This can likely be attributed to the increased steric bulk and rigidity of these ligands, which prevents them from fitting into the HilD binding pocket. The critical importance of the carbonyl group, combined with the limited flexibility in the length of the alkyl linker between the two amino groups, supports our model in which specific interactions are formed between both the amide and tertiary amine groups with HilD residues. Residue E102 is predicted to form hydrogen-bonding interactions with both backbone amino groups. A chain length greater than three carbon atoms between the two amino groups would break at least one of these hydrogen bonds and be disadvantageous to binding.

Compounds C219, C225 and C186 all contain alkyl chain branching between the amide and amine groups (Figure 6.16 C). Both C219 and C225 bind to HilD, with comparable affinities, with K_d's of 7.24 and 6.55 µM, respectively, at 52°C. Like for the branched compound C276, these compounds have higher affinity than C26, although there is no preference for one of the two stereoisomers. C186 also binds to HilD, albeit with significantly lower affinity (both $K_{d,app}$ and IC_{50} >100 µM), possibly due to the increased hydrophilicity of this group. Compounds containing extended hydrophobic groups at this position may result in further improvements in affinity, although this has yet to be determined.

We also probed replacing with the methyl group of the tertiary amine with other functional groups. Whilst C123, containing an ethyl group appears to show some residual activity in EMSAs, no binding could be detected in any other assays nor for any other compounds with larger R-groups (Figure 6.16 D). This shows that increased steric hinderance is not tolerated at this position. Increasingly large substituents may also hinder hydrogen bond interactions between the amine nitrogen and surrounding residues.

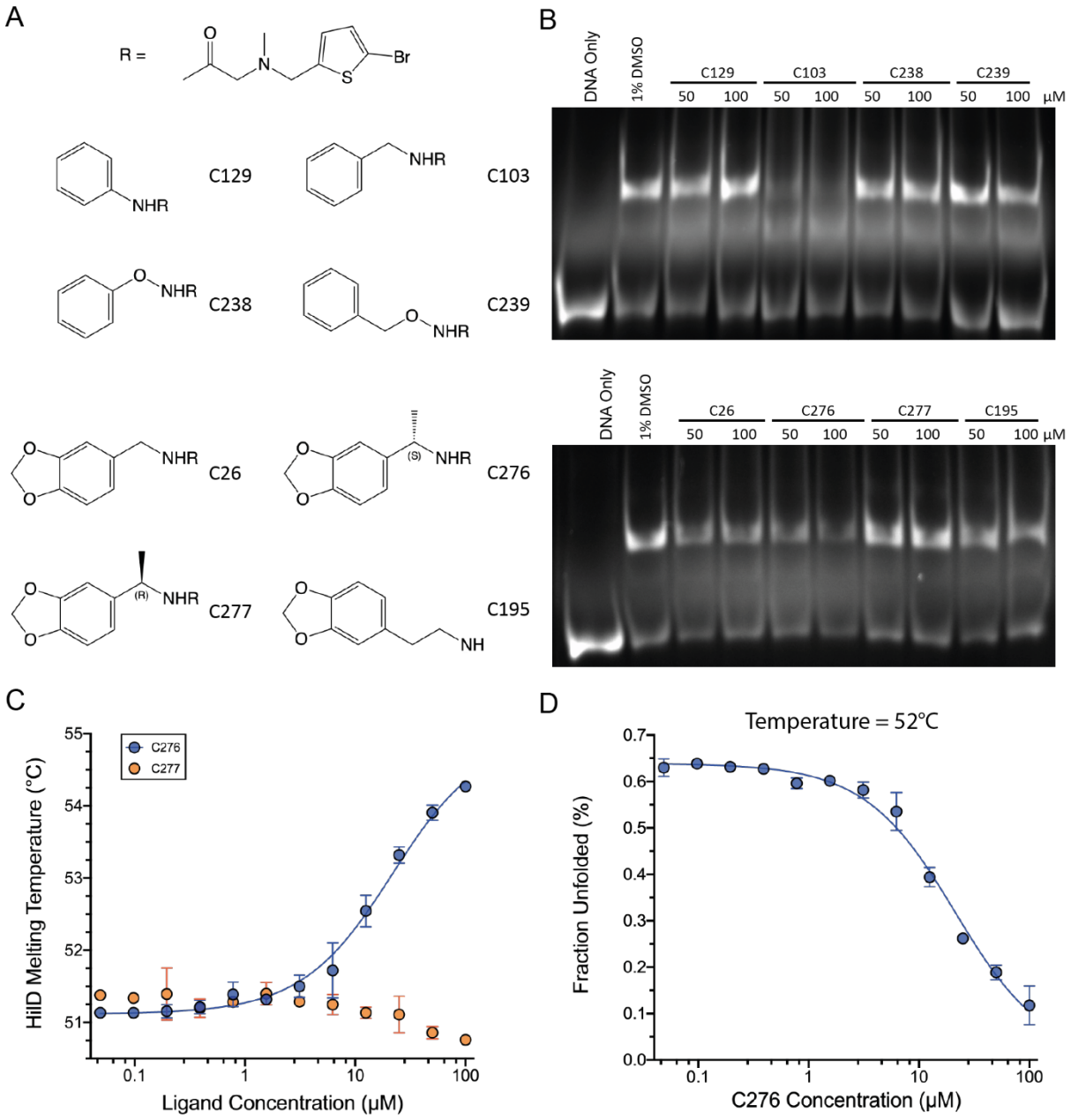


Figure 6.15. Variations in the benzyl group linker dramatically affect binding. (**A**) Chemical structures of the compounds with variations in the linker between the phenyl and amide groups. (**B**) EMSAs showing binding of the compounds show in (A) to HilD. (**C**) C276 increases the T_m of HilD in a dose-response manner, whilst C277 shows not significant effect on calculated T_m. (**D**) Isothermal analysis with $\Delta C_p = 0$, for C276 binding to HilD at 52°C.

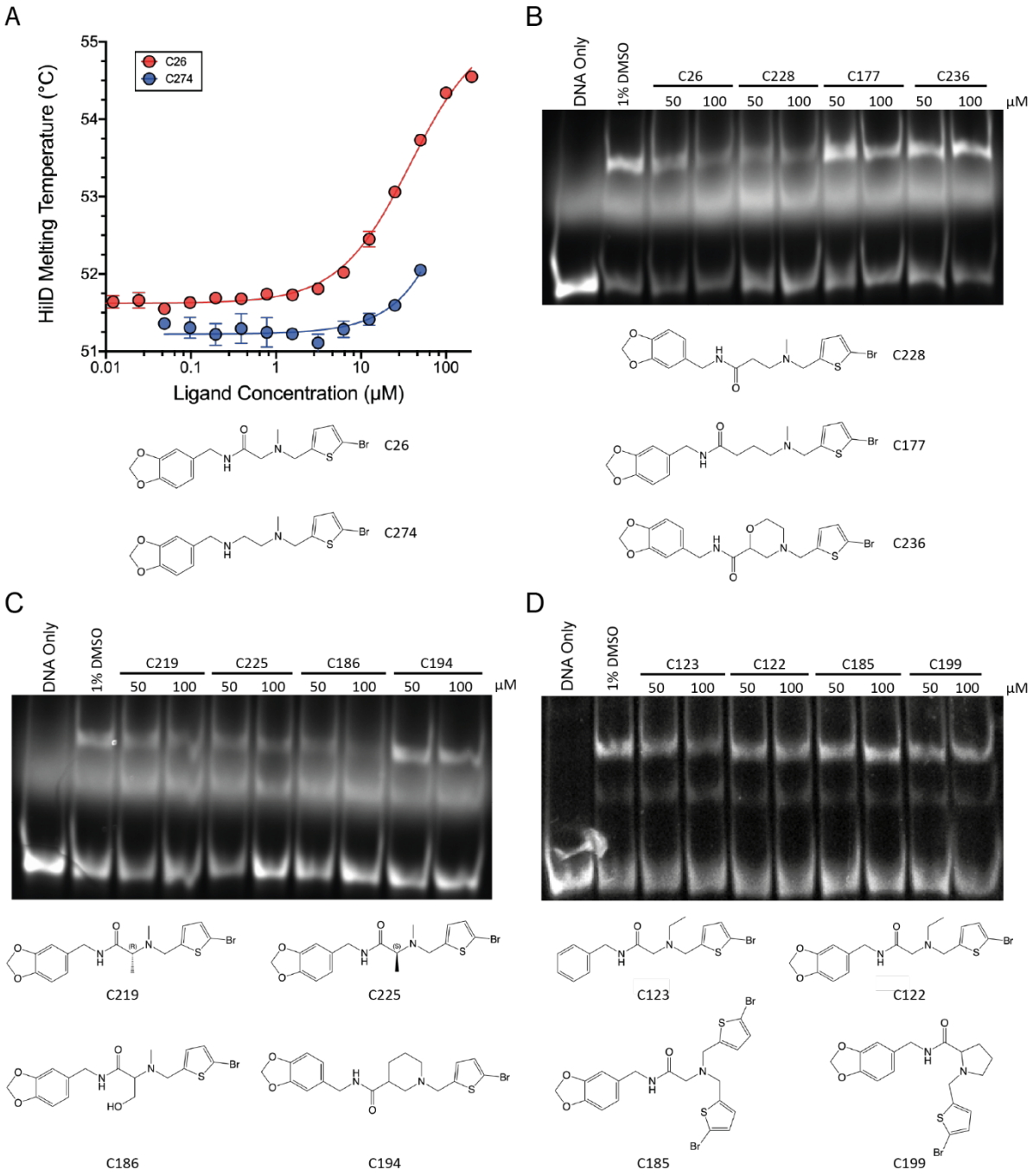


Figure 6.16. Variations in the central region of C26 are mostly detrimental to activity. (**A**) Comparison of the changes in the melting temperature of HilD with increasing concentrations of C26 or C274. (**B-D**) EMSAs and chemical structures showing the binding to HilD of compounds with variations in the central backbone of the molecule.

6.6.3 Thiophene

C26 contains a thiophene moiety with a Br substituent at the C5 position (Figure 6.17 A). Replacement of this Br substituent by Cl had no significant effect on ligand binding. However, C104, which contains an unsubstituted thiophene, did not bind to HilD in EMSAs (Figure 6.17 B). Interestingly, the presence of larger substituents was not detrimental to binding, with C235 containing a phenyl substituent and binding to HilD with high affinity (K_d: 7.84 µM at 52°C). However, compounds with methyl, nitrile or methyl sulfide (SCH_3) substituents at this position only showed very low affinity binding (>100 µM), indicating only limited optimisation of this substituent can be achieved. As C235 also did not show any activity *in vivo*, it appears that a halogen substituent is optimal for compound activity.

The position of the halogen substituent at the C5 position is essential for binding and was also demonstrated for the corresponding compounds lacking the dioxolane group (Figure 6.17 C). Compound C115, which contains a Br at position C4, rather than at position C5 as in C26/ C103, did not bind to HilD. Compounds with a 4,5-disubstituted thiophene, with Br at C5 and either a Br or Me group at C4, also did not bind to HilD nor inhibit P_{hilA} activation *in vivo*.

We tested several compounds in which the thiophene group was replaced by other 5-membered heterocycles, with a substituent at the C5 position akin to C26. Surprisingly, we found that replacement of thiophene by furan resulted in almost complete loss of binding (Figure 6.18 A-C). Some residual activity was observed for compound C210 (overall structure the same as C26), but with greatly reduced affinity (>100 µM) both *in vivo* and *in vitro* (Figure 6.18 B). A similar result was observed for corresponding 1,3-thiazole compounds, with the nature of the substituent on the heterocycle having no effect on the lack of binding (Figure 6.18 D, F). A structural isomer of the thiazole-containing C220, in which the positions of the R_1 and Br groups was swapped, also did not bind to HilD.

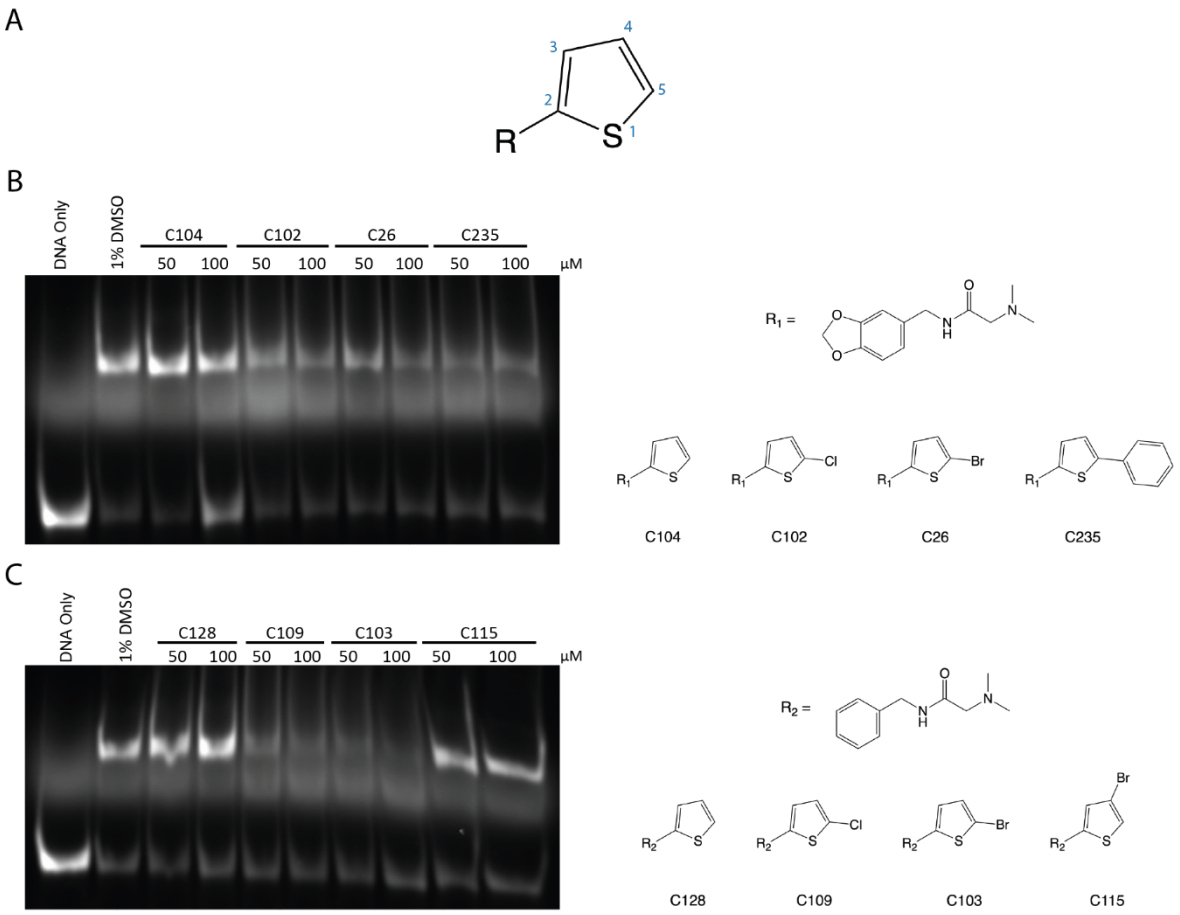


Figure 6.17. Substitution at the C5 position of thiophene is required for ligand binding. (**A**) The position of substituents on the thiophene moiety is defined by the numbering of atoms constituting the ring. Numbering begins at the sulfur heteroatom and the R group, denoting the rest of the compound structure, defined as position C2. (**B**-**C**) EMSAs and chemical structures of compounds differing in the identity of substituents on the thiophene moiety.

Replacement of the thiophene moiety of C26 with para-bromobenzene resulted in low-affinity binding (IC_{50}: 69.8 µM) in the *in vivo* screen. Compound C107 (which also lacks the dioxolane group) showed low affinity binding in both NanoDSF ($K_{d,app}$: 36.3 µM) and *in vivo* (IC_{50}: 49.3 µM). Although this appeared to open a wide range of possibilities for further optimisation, further alterations to this structure resulted in a loss of binding activity. We altered the position of the Br atom on the phenyl ring (to ortho or meta) in compounds either containing or lacking the dioxolane group (Figure 6.18 E-F), and additionally tested the equivalent compounds containing meta-F, meta-Cl, para-Cl substituents rather than Br. Derivative compounds of C107, containing additional phenyl substituents in addition to para-Br, or substituted pyridine moieties (in place of thiophene/ phenyl) also did not bind to HilD.

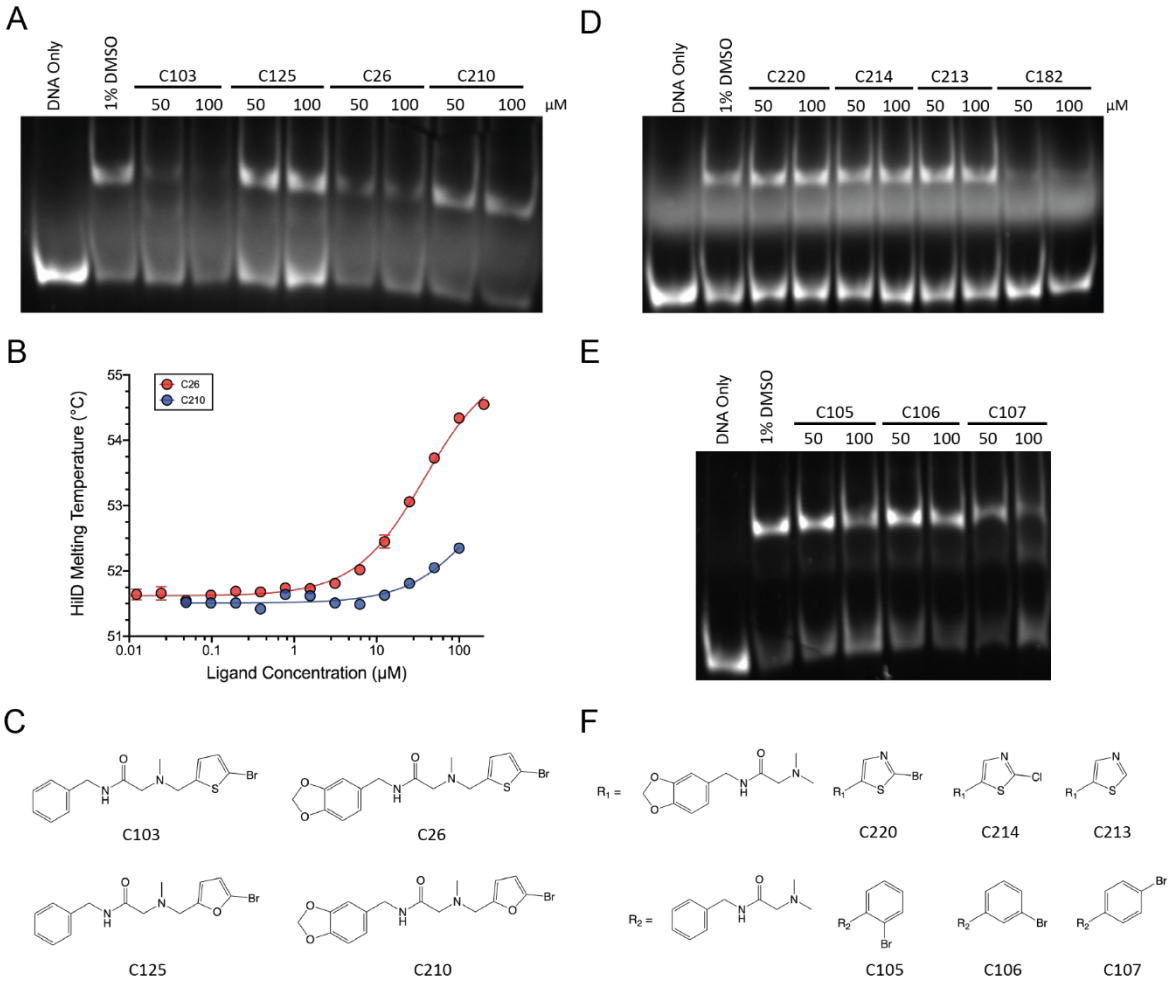


Figure 6.18. Compounds containing other substituted aryls in place of thiophene do not bind to HilD. (**A**-**B**) Comparison of activity of compounds containing either a thiophene or furan moiety, using (**A**) EMSA and (**B**) NanoDSF. (**C**) Chemical structures of compounds shown in (A, B). (**D**-**E**) EMSAs showing the binding of compounds containing either a (**D**) 1,3-thiazole or (**E**) bromobenzene moiety in place of thiophene. Compound C182 was run as a positive control, showing clear inhibition of HilD DNA-binding. (**F**) Chemical structures of the representative tested compounds shown in (D, E).

The strong preference for thiophene over other heterocycles can be explained by our model of the HilD-C26 complex, in which the thiophene moiety participates in a cation-π interaction with K264 (Figure 6.19). The strength of cation-π interactions depends on the availability of the π-electrons at the centre of the aromatic ring, rather than the reactivity of the aromatic group towards electrophiles, and can visualised from the electrostatic potential surfaces of molecules[156,157]. The strong electronegativity of nitrogen and oxygen distort the π-electrons away from the centre of the ring, hence weakening any cation-π interactions. This effect is less pronounced for thiophene, due to the lower electronegativity of sulfur compared to

nitrogen or oxygen, which contains higher π-electron density at the centre of the ring[158]. Hence, thiophene is a stronger cation-π donor than furan, thiazole or pyridine, although it is a much weaker cation-π donor than benzene[156,157]. Compound C107, which contains a benzene ring in place of thiophene, does display binding to HilD. The lower affinity of C107 is likely due to the increased steric restraints from the presence of a 6-membered rather than 5-membered ring at this position of the molecule. This is supported by the fact that the corresponding compounds containing additional aromatic substituents no longer bind to HilD.

The presence of electron-withdrawing halogen substituents on the thiophene moiety may be expected to weaken any potential cation-π interactions. However, substitution at the C5 position of thiophene is critical for ligand binding. This suggests the halogen substituent is involved in specific interactions with HilD residues. In our model, the Br substituent is predicted to halogen-bonding interactions with Q261 and the backbone atoms of other residues of the DBD helix α4 (Figure 6.19). This is supported by the binding of C107, containing a para-bromobenzene moiety, in which the Br substituent would be in a similar orientation to that in C26 and hence able to partake in halogen-bonding. On the other hand, compounds that contain a Br substituent at other positions on the thiophene/benzene ring do not show any activity. Compounds containing fluorobenzene or chlorobenzene in place of thiophene also do not bind to HilD. The higher electronegativity of these elements compared to Br would result in a greater distortion of the π-system, weakening the cation-π interaction. In these cases, the halogen-bonding interactions cannot compensate for the reduced strength of the cation-π interaction.

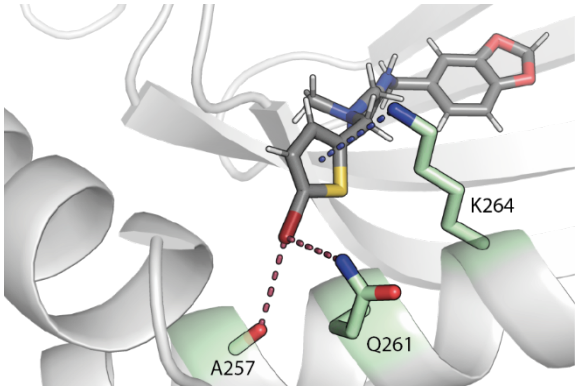


Figure 6.19. The thiophene moiety of C26 is predicted to form specific interactions with residues of HilD. Computational model of HilD in complex with C26. Residues predicted to interact with the thiophene moiety are shown as sticks and highlighted in green. Predicted contacts are coloured by the nature of the interaction (red: halogen bond; blue: cation-π).

6.7 Discussion

The transcription regulators of virulence genes are a key target for novel antipathogenic compounds. As the most important regulator of *Salmonella* virulence, HilD is an attractive target for the development of such inhibitors. Long chain fatty acids are the most well-studied small-molecule regulators of HilD, and a group of LCFAs were recently identified as high-affinity inhibitors of HilD[60].

The small molecule C26 was recently discovered as a novel inhibitor of *Salmonella* pathogenicity. We show here that C26 acts specifically upon HilD to repress the expression of *hilA* and virulence factors. Unlike fatty acid-based HilD inhibitors, C26 does not additionally interact with HilC (the closest homolog of HilD) or RtsA. The high specificity of C26 for HilD would reduce the likelihood of undesirable off-target effects. High target specificity is a key advantage of antipathogenic compounds compared to traditional broad-spectrum antibiotics, the use of which can be detrimental to the intestinal gut microbiota. Antivirulence compounds are also less prone to the development of resistance than traditional antibiotics, given they do not affect the survival of the organism. However, specific binding to HilD, and not to HilC or RtsA, may result in C26 being more susceptible to resistance development than an inhibitor targeting multiple of these regulators simultaneously. Characterisation of potential C26 resistance mechanisms, including the frequency of resistance development and the molecular basis of resistance using HilD point mutants, is a focus of our ongoing work.

The binding pocket of C26 was predicted computationally and validated using HDX-MS. C26 binds to the same pocket of HilD as natural LCFA regulators, albeit it appears with a different binding mode. The benzodioxol moiety of C26 is predicted to be buried within the cupin barrel of the HilD NTD, whilst the backbone and thiophene moieties form specific electrostatic interactions with residues of both HilD domains. This is comparable to the binding of oleic acid and results in a more stable, closed protein conformation, with C26 shown to stabilise HilD in a dose-response manner. Biophysical assays utilising HilD point mutants and further structure characterisation attempts are currently ongoing to determine the specific residues involved in the HilD-C26 interaction.

C26 and oleic acid prevent HilD from binding to DNA and both ligands give rise to HDX differences in both HTH motifs. However, the differences in HDX behaviour suggest C26 and oleic acid may cause different conformational changes in the overall structure of HilD, and indeed both ligands have different effects on HilD activity. C26 neither disrupts HilD dimerisation nor the interaction between HilD and HilE. Whether C26 has additional effects on HilD *in vivo*, such as increasing HilD degradation by Lon protease, as observed for LCFAs, remains to be determined. Development of an *in vitro* transcription assay would be advantageous to probe the effects of C26 at specific stages of the HilD activation mechanism, including the displacement of the nucleoid-associated protein H-NS and interactions with RNA polymerase.

The initial structure-activity relationship analysis of C26 revealed the benzyl moiety to be the most promising for further optimisation of higher affinity HilD inhibitors (Figure 6.20). The dioxolane functional group is not required for binding to HilD and the presence of other aromatic substituents greatly improved affinity. Branching at the CH_2 of the benzyl group also led to a stereoselective improvement in binding affinity. On the other hand, variations in the thiophene moiety or compound backbone were detrimental to binding, showing both to be critical for binding. The highest affinity compounds identified so far have a K_d < 5 µM and an IC_{50} ≤ 10 µM. This is comparable to the K_d of 14.7 µM determined for erucic acid, the highest affinity LCFA of those tested for binding to HilD, and shows the C26 scaffold presents an attractive alternative to fatty acid mimics for the design of HilD inhibitors. All these higher affinity compounds contain variations in the benzyl moiety and further modifications to this moiety are a key focus of our future optimisation efforts. Further SAR analysis will now take a more rational design approach, optimising and/ or combining properties of the highest affinity compounds, with the aim to identify inhibitors with an affinity of < 250 nM.

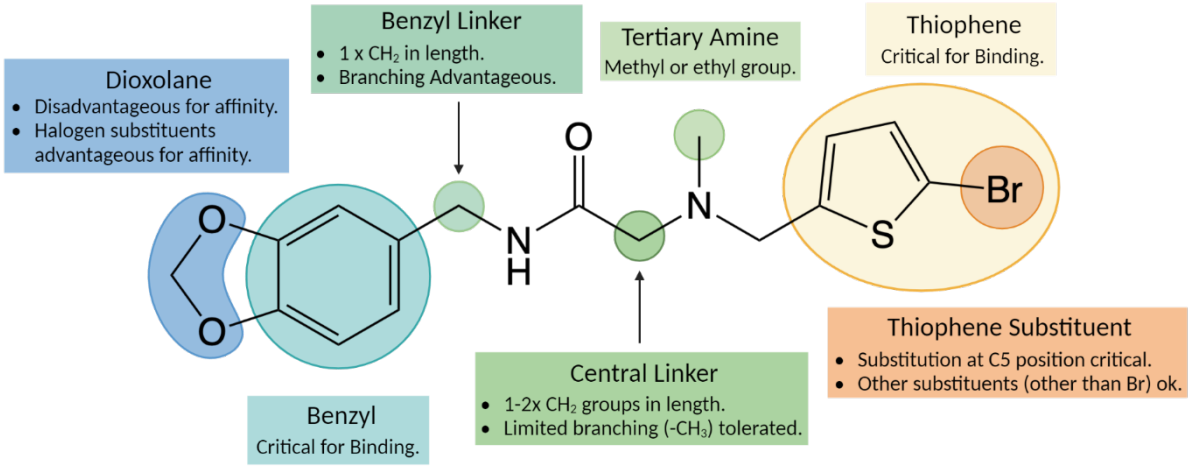


Figure 6.20. Summary of the C26 SAR Analysis.

So far, our SAR studies have been limited to sequentially probing the individual functional groups of C26 structure to decipher their importance to binding in the hope of identifying higher-affinity homologs. A high-resolution structure of HilD, either in the apo-form or bound to C26, would allow us to employ a structure-based approach to compound optimisation and is an ongoing focus of our current investigations of HilD and C26. Characterising the binding pocket of HilD would additionally provide a template for *in silico* design of other virulence inhibitors distinct from C26. This would be important should further optimisation of inhibitors derived from the C26 motif prove unsuccessful.

Chapter 7. Concluding Remarks

As a key component of *Salmonella* pathogenesis, the SPI-1 encoded T3SS-1 and the regulators controlling its expression are prime targets for novel antivirulence compounds. Inhibitors of pathogenesis-relevant transcription regulators have previously been discovered and optimised, aided by structural knowledge of these proteins[82,83]. In the work presented throughout this thesis, we have evaluated the different mechanisms by which HilD is regulated at the protein level. Knowledge of these regulatory mechanisms of HilD will aid the understanding of the different approaches by which putative antipathogenic compounds can be designed to target HilD.

HilD is capable of binding a range of long chain fatty acids that are present throughout the mammalian gut. *Salmonella spp.* have adapted to recognise these compounds and use them to sense their intestinal location and coordinate the expression of virulence genes. Here we show that the binding of LCFAs to HilD is comparable to that shown for other AraC/XylS transcription factors from other enteric pathogens. LCFAs bind to a conserved pocket in HilD, ToxT and Rns, and our results indicate that LCFA regulation of HilD likely occurs via a mechanism of allosteric regulation, as has been described for ToxT[58]. Through a brief SAR study, we identified properties of LCFAs that are preferential for binding to HilD, which highlights different LCFA binding specificities for each of these three regulators. This suggests that different enteric pathogens have evolved in parallel to detect fatty acids present in the intestine as environmental cues to ensure the efficiency of pathogenesis. Naturally-occurring LCFAs that have a higher affinity to HilD were recently identified[60], and fatty acid mimics are attractive candidates for the design of novel HilD inhibitors.

The Hcp-like protein HilE specifically represses HilD via a protein-protein interaction to prevent activation of virulence genes under normal conditions. We biochemically characterised the HilD-HilE interaction, revealing that HilE binds to HilD with a 1:1 stoichiometry, rather than forming a large protein complex as had previously been hypothesised. Although a high-resolution structure is required to decipher the exact binding mode of this interaction, our results highlight a previously unreported mechanism of regulation amongst AraC/XylS proteins. The apparent novelty of this mechanism presents an attractive prospect for highly specific HilD binders. In addition to small molecules targeting

the HilE-HilD complex rather than HilD alone, peptide-based inhibitors could potentially be designed that bind analogously to HilE to disrupt HilD dimerisation and activity.

Compound C26 was recently identified as a novel inhibitor of T3SS-1, inhibiting the expression of both T3SS-1 structural components and secreted effectors. Here we show that C26 acts by binding specifically to HilD to inhibit the expression of *hilA* and all virulence genes of SPI-1. Although seemingly binding to the same pocket of HilD as LCFAs, C26 appears to do so with a different mode and exhibits a different mechanism-of-action to these natural regulators. The preliminary SAR analysis of C26 revealed multiple molecular modifications that resulted in more potent binders, providing confidence in the potential of C26 to be developed into a lead compound. Future work will seek to extend this study to uncover more potent inhibitors, with optimised lead compounds selected for an *in vivo* proof-of-concept of HilD inhibitors in a mouse infection model. The interaction of C26 with HilD highlights the large chemical diversity of ligands capable of binding to HilD and the potential for further distinct inhibitors of HilD to be identified in the future.

Contributions

Throughout this thesis, the personal pronoun 'we' is used to collectively indicate me and my scientific collaborators. I performed all experiments described throughout this work, wrote this manuscript, and prepared the figures contained within it, unless otherwise stated. This work was performed under the supervision of Marcus Hartmann. Together, we regularly discussed the progress of each of the described projects, analysed results and designed further experiments.

General Methods

Preliminary expression tests of HilD were performed by Iwan Grin (Universitätsklinikum Tübingen), who provided the SUMO-HilD construct. I performed crystallisation screening experiments with support from Reinhard Albrecht, who mounted resultant crystals and collected diffraction data at the beamline. HDX experiments (including subsequent data analysis) were performed by Wieland Steinchen (Philipps-Universität Marburg), who also provided the HDX difference plots (Appendix 3) and contributed to writing of section 2.5.6 (HDX Methods) of this thesis. Mass Spectrometry experiments, to verify the identity of purified proteins and proteolysis digestion products, were performed by Johannes Madlung (Universität Tübingen). Computational modelling of HilD and molecular docking of ligands was performed by Thales Kronenberger (Universitätsklinikum Tübingen).

Chapter 5

I initiated this project and designed all experiments described throughout this chapter. ITC measurements were performed by Reinhard Albrecht and Nick Mozer, who together also performed crystallisation trials of HilE and the HilD-HilE complex. Modelling of the HilD-HilE complex was performed by Hadeer Elhabashy.

Chapter 6

The work described throughout this chapter was carried out as part of a wider consortium, involving the groups of Samuel Wagner (Universitätsklinikum Tübingen), Antti Poso (Universitätsklinikum Tübingen) and Mark Brönstrup (HZI Braunschweig), and the project initiated as a collaboration between these groups. Iwan Grin and Abdelhakim Boudrioua (Universitätsklinikum Tübingen) performed all *in vivo* assays: secretion assays (including

western blot analysis), RNA sequencing and the *in vivo* P_{hilA} activation assay. *In silico* screening for T3SS-1 inhibitors and molecular docking of synthesised ligands to HilD was performed by Thales Kronenberger and Antti Poso (Universitätsklinikum Tübingen). Synthesis of C26 analogues used in the SAR study was performed by Simon Kalverkamp and Vadim Korotkov (HZI Braunschweig). I performed all other experiments described throughout this chapter with regular feedback from all other members of this consortium.

Acknowledgements

First and foremost, I thank my supervisor, Marcus Hartmann. For entrusting me to work on this project. For the continued supervision and guidance throughout the past four years, but also the freedom to pursue my own ideas and grow as a scientist over the course of my PhD. And lastly, for critically reading and providing valuable feedback on this thesis.

I am also grateful to Andrei Lupas for providing me the opportunity to be a part of his department and for the scientific support and feedback throughout my PhD. Next, I would next like to acknowledge my second supervisor Samuel Wagner, thank you for the continued support and encouragement.

I also acknowledge all the other collaborators throughout my PhD, especially Abdelhakim Boudrioua, Iwan Grin and Thales Kronenberger, for the regular discussions, feedback and encouragement. I am also grateful to Thilo Stehle for being part of my initial TAC committee, and to Dirk Schwarzer and John Weir for agreeing to be part of my examination committee.

I am incredibly grateful to Reinhard Albrecht for teaching me the ways of the lab during the first weeks and months of my PhD, the continued support throughout the following four years, and for proof-reading and translating my thesis. I also thank all other past and present members of department who have helped me throughout my time here in Tübingen. In particular, Mikel Martinez Goikoetxea, Yimin Hu, Hadeer Elhabashy and Tran Hoang for both the emotional and scientific support during the last four years. I would also like to specifically mention the following people who regularly offered their advice and guidance in the lab, teaching me everything I needed to successfully complete my PhD: Anja Rau, Christopher Heim, Felipe Merino, Carolina Sánchez Rico, Kerstin Bär and Lorena Vöhringer.

And finally, a big thank you to the family and friends who have sustained me over years. In particular, I have to mention Anasuya, Sophie and Xi, for proof-reading and providing helpful comments on this thesis, in addition to the unconditional support and encouragement throughout my PhD.

References

1. Theuretzbacher, U. *et al.* Critical analysis of antibacterial agents in clinical development. *Nature Reviews Microbiology* **18**, 286–298 (2020).
2. Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet. Infect. Dis.* **18**, 318–327 (2018).
3. Projan, S. J. & Shlaes, D. M. Antibacterial drug discovery: is it all downhill from here? *Clin. Microbiol. Infect.* **10 Suppl 4**, 18–22 (2004).
4. Allen, H. K., Trachsel, J., Looft, T. & Casey, T. A. Finding alternatives to antibiotics. *Ann. N. Y. Acad. Sci.* **1323**, 91–100 (2014).
5. Tse, B. N. *et al.* Challenges and Opportunities of Nontraditional Approaches to Treating Bacterial Infections. *Clin. Infect. Dis.* **65**, 495–500 (2017).
6. Rex, J. H., Fernandez Lynch, H., Cohen, I. G., Darrow, J. J. & Outterson, K. Designing development programs for non-traditional antibacterial agents. *Nat. Commun.* **10**, 3416 (2019).
7. Theuretzbacher, U. & Piddock, L. J. V. Non-traditional Antibacterial Therapeutic Options and Challenges. *Cell Host Microbe* **26**, 61–72 (2019).
8. Dickey, S. W., Cheung, G. Y. C. & Otto, M. Different drugs for bad bugs: antivirulence strategies in the age of antibiotic resistance. *Nat. Rev. Drug Discov.* **16**, 457–471 (2017).
9. Stanaway, J. D. *et al.* The global burden of non-typhoidal salmonella invasive disease: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Infect. Dis.* **19**, 1312–1324 (2019).
10. Feasey, N. A., Dougan, G., Kingsley, R. A., Heyderman, R. S. & Gordon, M. A. Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. *Lancet (London, England)* **379**, 2489–2499 (2012).
11. V T Nair, D., Venkitanarayanan, K. & Kollanoor Johny, A. Antibiotic-Resistant Salmonella in the Food Supply and the Potential Role of Antibiotic Alternatives for Control. *Foods (Basel, Switzerland)* **7**, (2018).
12. Medalla, F. *et al.* Increased Incidence of Antimicrobial-Resistant Nontyphoidal *Salmonella* Infections, United States, 2004–2016. *Emerg. Infect. Dis.* **27**, 1662–1672

(2021).

13. Kobayashi, N., Takahashi, D., Takano, S., Kimura, S. & Hase, K. The Roles of Peyer's Patches and Microfold Cells in the Gut Immune System: Relevance to Autoimmune Diseases. *Front. Immunol.* **10**, 2345 (2019).
14. Jones, B. D., Ghori, N. & Falkow, S. Salmonella typhimurium initiates murine infection by penetrating and destroying the specialized epithelial M cells of the Peyer's patches. *J. Exp. Med.* **180**, 15–23 (1994).
15. Galán, J. E., Lara-Tejero, M., Marlovits, T. C. & Wagner, S. Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells. *Annu. Rev. Microbiol.* **68**, 415–38 (2014).
16. Francis, C. L., Ryan, T. A., Jones, B. D., Smith, S. J. & Falkow, S. Ruffles induced by Salmonella and other stimuli direct macropinocytosis of bacteria. *Nature* **364**, 639–42 (1993).
17. Jepson, M. A., Collares-Buzato, C. B., Clark, M. A., Hirst, B. H. & Simmons, N. L. Rapid disruption of epithelial barrier function by Salmonella typhimurium is associated with structural modification of intercellular junctions. *Infect. Immun.* **63**, 356–9 (1995).
18. Fàbrega, A. & Vila, J. Salmonella enterica serovar Typhimurium skills to succeed in the host: Virulence and regulation. *Clin. Microbiol. Rev.* **26**, 308–341 (2013).
19. Gal-Mor, O., Boyle, E. C. & Grassl, G. A. Same species, different diseases: how and why typhoidal and non-typhoidal Salmonella enterica serovars differ. *Front. Microbiol.* **5**, 391 (2014).
20. Ilyas, B., Tsai, C. N. & Coombes, B. K. Evolution of Salmonella-host cell interactions through a dynamic bacterial genome. *Front. Cell. Infect. Microbiol.* **7**, (2017).
21. Lerminiaux, N. A., MacKenzie, K. D. & Cameron, A. D. S. Salmonella Pathogenicity Island 1 (SPI-1): The Evolution and Stabilization of a Core Genomic Type Three Secretion System. *Microorganisms* **8**, (2020).
22. Haraga, A., Ohlson, M. B. & Miller, S. I. Salmonellae interplay with host cells. *Nat. Rev. Microbiol.* **6**, 53–66 (2008).
23. Hansen-Wester, I. & Hensel, M. Salmonella pathogenicity islands encoding type III secretion systems. *Microbes and Infection* **3**, (2001).
24. Kuhle, V. & Hensel, M. Cellular microbiology of intracellular Salmonella enterica: functions of the type III secretion system encoded by Salmonella pathogenicity island

2. *Cell. Mol. Life Sci.* **61**, 2812–26 (2004).

25. Wagner, C., Barlag, B., Gerlach, R. G., Deiwick, J. & Hensel, M. The Salmonella enterica giant adhesin SiiE binds to polarized epithelial cells in a lectin-like manner. *Cell. Microbiol.* **16**, 962–75 (2014).

26. Ellermeier, C. D., Ellermeier, J. R. & Slauch, J. M. HilD, HilC and RtsA constitute a feed forward loop that controls expression of the SPI1 type three secretion system regulator hilA in Salmonella enterica serovar Typhimurium. *Mol. Microbiol.* **57**, 691–705 (2005).

27. Bajaj, V., Hwang, C. & Lee, C. A. hilA is a novel ompR/toxR family member that activates the expression of Salmonella typhimurium invasion genes. *Mol. Microbiol.* **18**, 715–27 (1995).

28. Darwin, K. H. & Miller, V. L. InvF is required for expression of genes encoding proteins secreted by the SPI1 type III secretion apparatus in Salmonella typhimurium. *J. Bacteriol.* **181**, 4949–54 (1999).

29. Eichelberg, K. & Galán, J. E. Differential regulation of Salmonella typhimurium type III secreted proteins by pathogenicity island 1 (SPI-1)-encoded transcriptional activators InvF and hilA. *Infect. Immun.* **67**, 4099–105 (1999).

30. Lostroh, C. P. & Lee, C. A. The Salmonella pathogenicity island-1 type III secretion system. *Microbes and Infection* **3**, 1281–1291 (2001).

31. Gerlach, R. G. *et al.* Cooperation of Salmonella pathogenicity islands 1 and 4 is required to breach epithelial barriers. *Cell. Microbiol.* **10**, (2008).

32. Thijs, I. M. V *et al.* Delineation of the Salmonella enterica serovar Typhimurium HilA regulon through genome-wide location and transcript analysis. *J. Bacteriol.* **189**, 4587–96 (2007).

33. Main-Hester, K. L., Colpitts, K. M., Thomas, G. A., Fang, F. C. & Libby, S. J. Coordinate regulation of Salmonella pathogenicity island 1 (SPI1) and SPI4 in Salmonella enterica serovar Typhimurium. *Infect. Immun.* **76**, 1024–35 (2008).

34. Ellermeier, C. D. & Slauch, J. M. RtsA and RtsB coordinately regulate expression of the invasion and flagellar genes in Salmonella enterica serovar Typhimurium. *J. Bacteriol.* **185**, 5096–108 (2003).

35. Olekhnovich, I. N. & Kadner, R. J. DNA-binding activities of the HilC and HilD virulence regulatory proteins of Salmonella enterica serovar Typhimurium. *J. Bacteriol.* **184**,

4148–4160 (2002).

36. Olekhnovich, I. N. & Kadner, R. J. Crucial roles of both flanking sequences in silencing of the hilA promoter in Salmonella enterica. *J. Mol. Biol.* **357**, 373–386 (2006).
37. Rakeman, J. L., Bonifield, H. R. & Miller, S. I. A HilA-independent pathway to Salmonella typhimurium invasion gene transcription. *J. Bacteriol.* **181**, 3096–104 (1999).
38. Akbar, S., Schechter, L. M., Lostroh, C. P. & Lee, C. A. AraC/XylS family members, HilD and HilC, directly activate virulence gene expression independently of HilA in Salmonella typhimurium. *Mol. Microbiol.* **47**, 715–728 (2003).
39. Chowdhury, R., Pavinski Bitar, P. D., Adams, M. C., Chappie, J. S. & Altier, C. AraC-type regulators HilC and RtsA are directly controlled by an intestinal fatty acid to regulate Salmonella invasion. *Mol. Microbiol.* **116**, 1464–1475 (2021).
40. Ellermeier, J. R. & Slauch, J. M. Adaptation to the host environment: regulation of the SPI1 type III secretion system in Salmonella enterica serovar Typhimurium. *Curr. Opin. Microbiol.* **10**, 24–29 (2007).
41. Golubeva, Y. A., Sadik, A. Y., Ellermeier, J. R. & Slauch, J. M. Integrating global regulatory input into the Salmonella pathogenicity Island 1 type III secretion system. *Genetics* **190**, 79–90 (2012).
42. Saini, S., Ellermeier, J. R., Slauch, J. M. & Rao, C. V. The role of coupled positive feedback in the expression of the SPI1 type three secretion system in Salmonella. *PLoS Pathog.* **6**, 1–16 (2010).
43. Baxter, M. A., Fahlen, T. F., Wilson, R. L. & Jones, B. D. HilE interacts with HilD and negatively regulates hilA transcription and expression of the Salmonella enterica serovar Typhimurium invasive phenotype. *Infect. Immun.* **71**, 1295–1305 (2003).
44. Grenz, J. R., Chubiz, J. E. C., Thaprawat, P. & Slauch, J. M. HilE regulates HilD by blocking DNA binding in Salmonella enterica serovar Typhimurium. *J. Bacteriol.* **200**, 1–13 (2018).
45. Paredes-Amaya, C. C., Valdés-García, G., Juárez-González, V. R., Rudiño-Piñera, E. & Bustamante, V. H. The Hcp-like protein HilE inhibits homodimerization and DNA binding of the virulence-associated transcriptional regulator HilD in Salmonella. *J. Biol. Chem.* **293**, 6578–6592 (2018).
46. Boddicker, J. D. & Jones, B. D. Lon protease activity causes down-regulation of

Salmonella pathogenicity island 1 invasion gene expression after infection of epithelial cells. *Infect. Immun.* **72**, 2002–13 (2004).

47. Takaya, A., Kubota, Y., Isogai, E. & Yamamoto, T. Degradation of the HilC and HilD regulator proteins by ATP-dependent Lon protease leads to downregulation of Salmonella pathogenicity island 1 gene expression. *Mol. Microbiol.* **55**, 839–52 (2005).
48. Stock, A. M., Robinson, V. L. & Goudreau, P. N. Two-component signal transduction. *Annu. Rev. Biochem.* **69**, 183–215 (2000).
49. Gao, R. & Stock, A. M. Biological insights from structures of two-component proteins. *Annu. Rev. Microbiol.* **63**, 133–54 (2009).
50. Barrett, J. F. & Hoch, J. A. Two-component signal transduction as a target for microbial anti-infective therapy. *Antimicrob. Agents Chemother.* **42**, 1529–36 (1998).
51. Lawhon, S. D., Maurer, R., Suyemoto, M. & Altier, C. Intestinal short-chain fatty acids alter Salmonella typhimurium invasion gene expression and virulence through BarA/SirA. *Mol. Microbiol.* **46**, 1451–64 (2002).
52. Martínez, L. C. *et al.* Integration of a complex regulatory cascade involving the SirA/BarA and Csr global regulatory systems that controls expression of the Salmonella SPI-1 and SPI-2 virulence regulons through HilD. *Mol. Microbiol.* **80**, 1637–1656 (2011).
53. Hung, C.-C. *et al.* Salmonella invasion is controlled through the secondary structure of the hilD transcript. *PLoS Pathog.* **15**, e1007700 (2019).
54. Chavez, R. G., Alvarez, A. F., Romeo, T. & Georgellis, D. The physiological stimulus for the BarA sensor kinase. *J. Bacteriol.* **192**, 2009–12 (2010).
55. Bustamante, V. H. *et al.* HilD-mediated transcriptional cross-talk between SPI-1 and SPI-2. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14591–14596 (2008).
56. Golubeva, Y. A., Ellermeier, J. R., Chubiz, J. E. C. & Slauch, J. M. Intestinal long-chain fatty acids act as a direct signal to modulate expression of the Salmonella pathogenicity island 1 type III secretion system. *MBio* **7**, 1–9 (2016).
57. Lowden, M. J. *et al.* Structure of Vibrio cholerae ToxT reveals a mechanism for fatty acid regulation of virulence genes. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2860–2865 (2010).
58. Cruite, J. T. *et al.* Structural basis for virulence regulation in Vibrio cholerae by unsaturated fatty acid components of bile. *Commun. Biol.* **2**, 1–9 (2019).

59. Midgett, C. R., Talbot, K. M., Day, J. L., Munson, G. P. & Kull, F. J. Structure of the master regulator Rns reveals an inhibitor of enterotoxigenic Escherichia coli virulence regulons. *Sci. Rep.* **11**, 1–13 (2021).

60. Bosire, E. M. *et al.* Diffusible Signal Factors Act through AraC-Type Transcriptional Regulators as Chemical Cues To Repress Virulence of Enteric Pathogens. *Infect. Immun.* **88**, 1–14 (2020).

61. Chowdhury, R., Bitar, P. D. P., Keresztes, I., Condo, A. M. & Altier, C. A diffusible signal factor of the intestine dictates Salmonella invasion through its direct control of the virulence activator HilD. *PLoS Pathog.* **17**, 1–20 (2021).

62. Wu, Y., Yang, X., Zhang, D. & Lu, C. Myricanol Inhibits the Type III Secretion System of Salmonella enterica Serovar Typhimurium by Interfering With the DNA-Binding Activity of HilD. *Front. Microbiol.* **11**, 1–11 (2020).

63. Olekhnovich, I. N. & Kadner, R. J. Role of nucleoid-associated proteins Hha and H-NS in expression of Salmonella enterica activators HilD, HilC, and RtsA required for cell invasion. *J. Bacteriol.* **189**, 6882–6890 (2007).

64. Narm, K. E., Kalafatis, M. & Slauch, J. M. HilD, HilC, and RtsA form homodimers and heterodimers to regulate expression of the salmonella pathogenicity island i type iii secretion system. *J. Bacteriol.* **202**, 1–13 (2020).

65. Dorman, C. J. H-NS, the genome sentinel. *Nat. Rev. Microbiol.* **5**, 157–61 (2007).

66. Navarre, W. W. *et al.* Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. *Science* **313**, 236–8 (2006).

67. Lucchini, S. *et al.* H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog.* **2**, e81 (2006).

68. Dame, R. T., Wyman, C., Wurm, R., Wagner, R. & Goosen, N. Structural basis for H-NS-mediated trapping of RNA polymerase in the open initiation complex at the rrnB P1. *J. Biol. Chem.* **277**, 2146–50 (2002).

69. Egan, S. M. Growing Repertoire of AraC / XylS Activators GUEST COMMENTARY Growing Repertoire of AraC / XylS Activators. *J. Bacteriol.* **184**, 5529–5532 (2002).

70. Martínez, L. C., Banda, M. M., Fernández-Mora, M., Santana, F. J. & Bustamante, V. H. HilD induces expression of Salmonella pathogenicity Island 2 genes by displacing the global negative regulator H-NS from ssrAB. *J. Bacteriol.* **196**, 3746–3755 (2014).

71. Banda, M. M., Zavala-Alvarado, C., Pérez-Morales, D. & Bustamante, V. H. SlyA and

HilD counteract H-NS-mediated repression on the ssrAB virulence operon of Salmonella enterica serovar typhimurium and thus promote its activation by OmpR. *J. Bacteriol.* **201**, 1–14 (2019).

72. Palmer, A. D., Kim, K. & Slauch, J. M. PhoP-Mediated Repression of the SPI1 Type 3 Secretion System in Salmonella enterica Serovar Typhimurium. *J. Bacteriol.* **201**, (2019).
73. Boddicker, J. D., Knosp, B. M. & Jones, B. D. Transcription of the Salmonella invasion gene activator, hilA, requires HilD activation in the absence of negative regulators. *J. Bacteriol.* **185**, 525–533 (2003).
74. Olekhnovich, I. N. & Kadner, R. J. Contribution of the RpoA C-Terminal Domain to Stimulation of the Salmonella enterica hilA Promoter by HilC and HilD. *J. Bacteriol.* **186**, 3249–3253 (2004).
75. Hudson, D. L. *et al.* Inhibition of type III secretion in Salmonella enterica serovar Typhimurium by small-molecule inhibitors. *Antimicrob. Agents Chemother.* **51**, 2631–5 (2007).
76. Felise, H. B. *et al.* An inhibitor of gram-negative bacterial virulence protein secretion. *Cell Host Microbe* **4**, 325–36 (2008).
77. Aiello, D. *et al.* Discovery and characterization of inhibitors of Pseudomonas aeruginosa type III secretion. *Antimicrob. Agents Chemother.* **54**, 1988–99 (2010).
78. Martinez-Argudo, I. *et al.* Isolation of Salmonella mutants resistant to the inhibitory effect of Salicylidene acylhydrazides on flagella-mediated motility. *PLoS One* **8**, e52179 (2013).
79. Li, J. *et al.* Cytosporone B, an inhibitor of the type III secretion system of Salmonella enterica serovar Typhimurium. *Antimicrob. Agents Chemother.* **57**, 2191–8 (2013).
80. Tsai, C. N. *et al.* Targeting Two-Component Systems Uncovers a Small-Molecule Inhibitor of Salmonella Virulence. *Cell Chem. Biol.* **27**, 793-805.e7 (2020).
81. Shakhnovich, E. A., Hung, D. T., Pierson, E., Lee, K. & Mekalanos, J. J. Virstatin inhibits dimerization of the transcriptional activator ToxT. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2372–2377 (2007).
82. Woodbrey, A. K. *et al.* A new class of inhibitors of the AraC family virulence regulator Vibrio cholerae ToxT. *Sci. Rep.* **7**, 1–11 (2017).
83. Markham, L. E., Tolbert, J. D., Kull, F. J., Midgett, C. R. & Micalizio, G. C. An

Enantiodefined Conformationally Constrained Fatty Acid Mimetic and Potent Inhibitor of ToxT. *ACS Med. Chem. Lett.* **12**, 1493–1497 (2021).

84. Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G. & Galas, D. J. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res.* **17**, 6545–6551 (1989).
85. Niebling, S. *et al.* FoldAffinity: binding affinities from nDSF experiments. *Sci. Rep.* **11**, 9572 (2021).
86. Bai, N., Roder, H., Dickson, A. & Karanicolas, J. Isothermal Analysis of ThermoFluor Data can readily provide Quantitative Binding Affinities. *Sci. Rep.* **9**, 2650 (2019).
87. Wales, T. E., Fadgen, K. E., Gerhardt, G. C. & Engen, J. R. High-speed and high-resolution UPLC separation at zero degrees Celsius. *Anal. Chem.* **80**, 6815–20 (2008).
88. Geromanos, S. J. *et al.* The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9**, 1683–1695 (2009).
89. Li, G.-Z. *et al.* Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**, 1696–1719 (2009).
90. Osorio-Valeriano, M. *et al.* ParB-type DNA Segregation Proteins Are CTP-Dependent Molecular Switches. *Cell* **179**, 1512-1524.e15 (2019).
91. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
92. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
93. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034 (2021). doi:10.1101/2021.10.04.463034
94. The PyMOL Molecular Graphics System, Version 2.4.0 Schrodinger, LLC.
95. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
96. Gallegos, M. T., Schleif, R., Bairoch, A., Hofmann, K. & Ramos, J. L. Arac/XylS family of transcriptional regulators. *Microbiol. Mol. Biol. Rev.* **61**, 393–410 (1997).
97. Perez-Rueda, E. *et al.* Abundance, diversity and domain architecture variability in

prokaryotic DNA-binding transcription factors. *PLoS One* **13**, e0195332 (2018).

98. Cortés-Avalos, D. *et al.* An update of the unceasingly growing and diverse AraC/XylS family of transcriptional activators. *FEMS Microbiol. Rev.* **45**, (2021).
99. Ibarra, J. A., Pérez-Rueda, E., Segovia, L. & Puente, J. L. The DNA-binding domain as a functional indicator: the case of the AraC/XylS family of transcription factors. *Genetica* **133**, 65–76 (2008).
100. Rodgers, M. E. & Schleif, R. Solution structure of the DNA binding domain of AraC protein. *Proteins Struct. Funct. Bioinforma.* **77**, 202–208 (2009).
101. Bustos, S. A. & Schleif, R. F. Functional domains of the AraC protein. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 5638–5642 (1993).
102. Soisson, S. M., MacDougall-Shackleton, B., Schleif, R. & Wolberger, C. Structural basis for ligand-regulated oligomerization of AraC. *Science (80-.).* **276**, 421–425 (1997).
103. Schleif, R. AraC protein: a love-hate relationship. *Bioessays* **25**, 274–82 (2003).
104. Schleif, R. AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action. *FEMS Microbiol. Rev.* **34**, 779–796 (2010).
105. Belmont-Monroy, L. *et al.* Characterization of a novel AraC/XylS-regulated family of N-acyltransferases in pathogens of the order Enterobacterales. *PLOS Pathog.* **16**, e1008776 (2020).
106. Li, J., Wehmeyer, G., Lovell, S., Battaile, K. P. & Egan, S. M. 1.65 Å resolution structure of the AraC-family transcriptional activator ToxT from Vibrio cholerae. *Acta Crystallogr. Sect. Struct. Biol. Commun.* **72**, 726–731 (2016).
107. DiRita, V. J., Parsot, C., Jander, G. & Mekalanos, J. J. Regulatory cascade controls virulence in Vibrio cholerae. *Proc. Natl. Acad. Sci.* **88**, 5403–5407 (1991).
108. Egan, S. M. & Schleif, R. F. DNA-dependent renaturation of an insoluble DNA binding protein. Identification of the RhaS binding site at rhaBAD. *J. Mol. Biol.* **243**, 821–9 (1994).
109. Söding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–60 (2005).
110. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
111. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

112. Schäper, S. *et al.* AraC-like transcriptional activator CuxR binds c-di-GMP by a PilZ-like mechanism to regulate extracellular polysaccharide production. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E4822–E4831 (2017).
113. Alva, V., Nam, S.-Z., Söding, J. & Lupas, A. N. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res.* **44**, W410-5 (2016).
114. Gabler, F. *et al.* Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr. Protoc. Bioinforma.* **72**, (2020).
115. Jones, D. T. & Cozzetto, D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31**, 857–863 (2015).
116. Konermann, L., Pan, J. & Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**, 1224–1234 (2011).
117. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
118. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
119. Shrestha, M., Xiao, Y., Robinson, H. & Schubot, F. D. Structural analysis of the regulatory domain of ExsA, a key transcriptional regulator of the type three secretion system in Pseudomonas aeruginosa. *PLoS One* **10**, 1–17 (2015).
120. Weber, M. *et al.* Impact of C-terminal amino acid composition on protein expression in bacteria. *Mol. Syst. Biol.* **16**, (2020).
121. Chatterjee, A., Dutta, P. K. & Chowdhury, R. Effect of Fatty Acids and Cholesterol Present in Bile on Expression of Virulence Factors and Motility of *Vibrio cholerae*. *Infect. Immun.* **75**, 1946–1953 (2007).
122. Clark, D. P. & Cronan, J. E. Two-Carbon Compounds and Fatty Acids as Carbon Sources. *EcoSal Plus* (2005). doi:10.1128/ecosalplus.3.4.4
123. Huang, Y., Suyemoto, M., Garner, C. D., Cicconi, K. M. & Altier, C. Formate Acts as a Diffusible Signal To Induce *Salmonella* Invasion. *J. Bacteriol.* **190**, 4233–4241 (2008).
124. Gantois, I. *et al.* Butyrate specifically down-regulates Salmonella pathogenicity island 1 gene expression. *Appl. Environ. Microbiol.* **72**, 946–949 (2006).
125. Hung, C.-C. *et al.* The intestinal fatty acid propionate inhibits Salmonella invasion through the post-translational control of HilD. *Mol. Microbiol.* **87**, 1045–60 (2013).

126. Van Immerseel, F. *et al.* Medium-Chain Fatty Acids Decrease Colonization and Invasion through *hilA* Suppression Shortly after Infection of Chickens with *Salmonella enterica* Serovar Enteritidis. *Appl. Environ. Microbiol.* **70**, 3582–3587 (2004).
127. Day, J., Kovacikova, G., Taylor, R. K. & Kull, F. J. Unsaturated Fatty Acid Regulation of AraC/XylS Transcription Factors. *Biophys. J.* **106**, 497a (2014).
128. Childers, B. M. *et al.* N-terminal Residues of the Vibrio cholerae Virulence Regulatory Protein ToxT Involved in Dimerization and Modulation by Fatty Acids. *J. Biol. Chem.* **286**, 28644–28655 (2011).
129. Fahlen, T. F., Mathur, N. & Jones, B. D. Identification and characterization of mutants with increased expression of hilA, the invasion gene transcriptional activator of Salmonella typhimurium. *FEMS Immunol. Med. Microbiol.* **28**, 25–35 (2000).
130. Baxter, M. A. & Jones, B. D. The fimYZ genes regulate Salmonella enterica serovar Typhimurium invasion in addition to type 1 fimbrial expression and bacterial motility. *Infect. Immun.* **73**, 1377–1385 (2005).
131. Espinosa, E. & Casadesús, J. Regulation of Salmonella enterica pathogenicity island 1 (SPI-1) by the LysR-type regulator LeuO. *Mol. Microbiol.* **91**, 1057–69 (2014).
132. Lim, S. *et al.* Mlc regulation of Salmonella pathogenicity island I gene expression via hilE repression. *Nucleic Acids Res.* **35**, 1822–32 (2007).
133. Gong, H. *et al.* A Salmonella Small Non-Coding RNA Facilitates Bacterial Invasion and Intracellular Replication by Modulating the Expression of Virulence Factors. *PLoS Pathog.* **7**, e1002120 (2011).
134. Pérez-Morales, D. *et al.* An incoherent feedforward loop formed by SirA/BarA, HilE and HilD is involved in controlling the growth cost of virulence factor expression by Salmonella Typhimurium. *PLoS Pathog.* **17**, 1–26 (2021).
135. Hamed, S., Shawky, R. M., Emara, M., Slauch, J. M. & Rao, C. V. HilE is required for synergistic activation of SPI-1 gene expression in Salmonella enterica serovar Typhimurium. *BMC Microbiol.* **21**, 49 (2021).
136. Pell, L. G., Kanelis, V., Donaldson, L. W., Lynne Howell, P. & Davidson, A. R. The phage λ major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc. Natl. Acad. Sci.* **106**, 4160–4165 (2009).
137. Leiman, P. G. *et al.* Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc. Natl. Acad. Sci.* **106**, 4154–4159

(2009).

138. Ho, B. T., Dong, T. G. & Mekalanos, J. J. A View to a Kill: The Bacterial Type VI Secretion System. *Cell Host Microbe* **15**, 9–21 (2014).

139. Pukatzki, S. *et al.* Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc. Natl. Acad. Sci.* **103**, 1528–1533 (2006).

140. Silverman, J. M. *et al.* Haemolysin Coregulated Protein Is an Exported Receptor and Chaperone of Type VI Secretion Substrates. *Mol. Cell* **51**, 584–593 (2013).

141. Mougous, J. D. *et al.* A Virulence Locus of *Pseudomonas aeruginosa* Encodes a Protein Secretion Apparatus. *Science (80-.).* **312**, 1526–1530 (2006).

142. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science (80-.).* **373**, 871–876 (2021).

143. Lim, Y. T. *et al.* Extended Loop Region of Hcp1 is Critical for the Assembly and Function of Type VI Secretion System in Burkholderia pseudomallei. *Sci. Rep.* **5**, 8235 (2015).

144. Lin, Q.-P., Gao, Z.-Q., Geng, Z., Zhang, H. & Dong, Y.-H. Crystal structure of the putative cytoplasmic protein STM0279 (Hcp2) from *Salmonella typhimurium*. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **73**, 463–468 (2017).

145. Paredes-Amaya, C. C. Caracterización de la proteína HilD : un regulador central de la virulencia de Salmonella enterica. (UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO (UNAM), 2019).

146. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).

147. Santiago, A. E. *et al.* A Large Family of Antivirulence Regulators Modulates the Effects of Transcriptional Activators in Gram-negative Pathogenic Bacteria. *PLoS Pathog.* **10**, e1004153 (2014).

148. Santiago, A. E. *et al.* A large family of anti-activators accompanying XylS/AraC family regulatory proteins. *Mol. Microbiol.* **101**, 314–332 (2016).

149. McCaw, M. L., Lykken, G. L., Singh, P. K. & Yahr, T. L. ExsD is a negative regulator of the Pseudomonas aeruginosa type III secretion regulon. *Mol. Microbiol.* **46**, 1123–33 (2002).

150. Shrestha, M., Bernhards, R. C., Fu, Y., Ryan, K. & Schubot, F. D. Backbone Interactions

Between Transcriptional Activator ExsA and Anti-Activator ExsD Facilitate Regulation of the Type III Secretion System in Pseudomonas aeruginosa. *Sci. Rep.* **10**, 9881 (2020).

151. Butan, C., Lara-Tejero, M., Li, W., Liu, J. & Galán, J. E. High-resolution view of the type III secretion export apparatus in situ reveals membrane remodeling and a secretion pathway. *Proc. Natl. Acad. Sci.* **116**, 24786–24795 (2019).
152. Erhardt, M. *et al.* Mechanism of type-III protein secretion: Regulation of FlhA conformation by a functionally critical charged-residue cluster. *Mol. Microbiol.* **104**, 234–249 (2017).
153. Westerhausen, S. *et al.* A NanoLuc luciferase-based assay enabling the real-time analysis of protein secretion and injection by bacterial type III secretion systems. *Mol. Microbiol.* **113**, 1240–1254 (2020).
154. Vivoli, M., Novak, H. R., Littlechild, J. A. & Harmer, N. J. Determination of Protein-ligand Interactions Using Differential Scanning Fluorimetry. *J. Vis. Exp.* (2014). doi:10.3791/51809
155. Maiwald, S., Heim, C., Hernandez Alvarez, B. & Hartmann, M. D. Sweet and Blind Spots in E3 Ligase Ligand Space Revealed by a Thermophoresis-Based Assay. *ACS Med. Chem. Lett.* **12**, 74–81 (2021).
156. Mecozzi, S., West, A. P. & Dougherty, D. A. Cation-pi interactions in aromatics of biological and medicinal interest: electrostatic potential surfaces as a useful qualitative guide. *Proc. Natl. Acad. Sci.* **93**, 10566–10571 (1996).
157. Ma, J. C. & Dougherty, D. A. The Cation−π Interaction. *Chem. Rev.* **97**, 1303–1324 (1997).
158. Kumar, S. & Das, A. Effect of acceptor heteroatoms on π-hydrogen bonding interactions: A study of indole···thiophene heterodimer in a supersonic jet. *J. Chem. Phys.* **137**, 094309 (2012).

Appendices

Appendix 1. Protein Sequences

Protein sequences were obtained from the Uniprot database, translated from the genome of *Salmonella Typhimurium* str. SL1344. The GFP-HilD sequence corresponds to the final purified protein product used in MST assays.

HilD

MENVTFVSNSHQRPAADNLQKLKSLLTNTRQQIKSQTQQVTIKNLYVSSFTLVCFRSGKLTISNNHDTIYCDEPGMLVLKKEQVVNVTL
EEVNGHMDFDILEIPTQRLGALYALIPNEQQTKMAVPTEKAQKIFYTPDFPARREVFEHLKTAFSCTKDTSKGCSNCNNKSCIENEELIPY
FLLFLLTAFLRLPESYEIILSSAQITLKERVYNIISSSPSRQWKLTDVADHIFMSTSTLKRKLAEEGTSFSDIYLSARMNQAAKLLRIGNHNVN
AVALKCGYDSTSYFIQCFKKYFKTTPSTFIKMANH

HilC

MVLPSMNKSVEAISNNHLQQPNKFPLINGLADVRDYYVANCLLFKLNKGSLRIENEFGEFIERSAPCLFLLEKDQTITLSMSEIEGHIDFS
SLEVSYDLMQKFYKVFYSTRNYNDRELSLKTKPKYFFHADLLPGMSDTFDSILHGVACPRVCSNVSIDDHDYSYFSLMYLISAFVRKPGG
FDFLERAIKITTKEKVYNIIISDLTRKWSQAEVAGKLFMSVSSLKRKLAAEEVSFSKIYLDARMNQAIKLLRMGAGNISQVATMCGYDTPS
YFIAIFKRHFKITPLSFMRTMNH

RtsA

MLKVFNPSPVQVGSIECLQSAQNWQRKSLSLQGLNLLQSVLIKLTTGKISITTSSGEYITASGPMLIFLAKDQTIHITMEETHEQLNYNLIE
LDSASIKNAYNFFLYEHADFSAPLTKPTTKHLLAPIETGVARVFNLLHSSNKSQKLSQDKKEYLIRFLLSEFIYEPEAFALFRELSQNTLAENI
YNIIISDISRKWALKDISDSLYMSCSTLKRKLKQEHTSFSEVYLNARMNKATKLLRNSEYNITRVAYMCGYDSASYFTCVFKKHFKTTPSEF
LAFLSSSRHQYVN

HilE

MDAIYLKLDGIEGESRIKGFENQIKLIAYNHNPTKRESGEARGTYIGGLTLTKPVDLATPGLYEHYCNGKTVKEGVLTLCRRDKGAMLPFI
IYTLTNVRISRMSNHGDAEGSATETVDLVYSHIRWDIPALASKSKTRRPLHRQALWR

GFP-HilD

AGKVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFGYGLQCFARYPDHMKQHDFFK
SAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDG
SVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGSSMENVTFVSNSHQRPAADN
LQKLKSLLTNTRQQIKSQTQQVTIKNLYVSSFTLVCFRSGKLTISNNHDTIYCDEPGMLVLKKEQVVNVTLEEVNGHMDFDILEIPTQRL
GALYALIPNEQQTKMAVPTEKAQKIFYTPDFPARREVFEHLKTAFSCTKDTSKGCSNCNNKSCIENEELIPYFLLFLLTAFLRLPESYEIILSS
AQITLKERVYNIISSSPSRQWKLTDVADHIFMSTSTLKRKLAEEGTSFSDIYLSARMNQAAKLLRIGNHNVNAVALKCGYDSTSYFIQCFK
KYFKTTPSTFIKMANH

Appendix 2. Protein Alignments

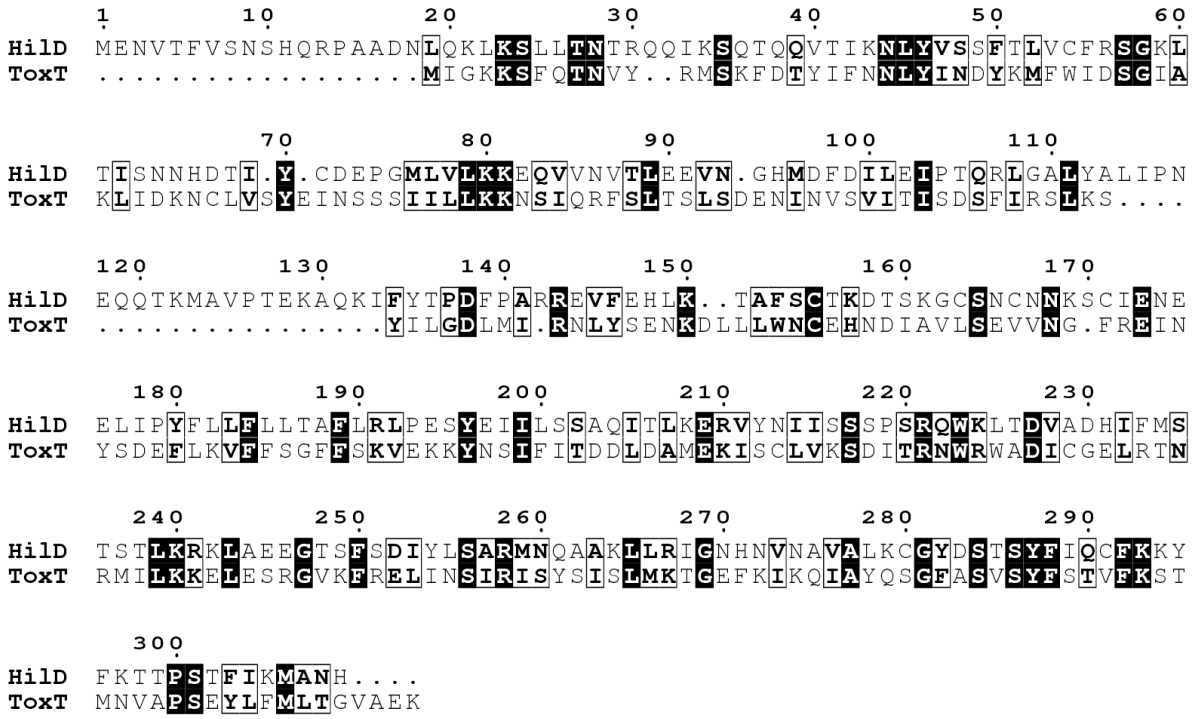


Figure S2.1. Pairwise sequence alignment of HilD and ToxT, performed with Clustal Ω.

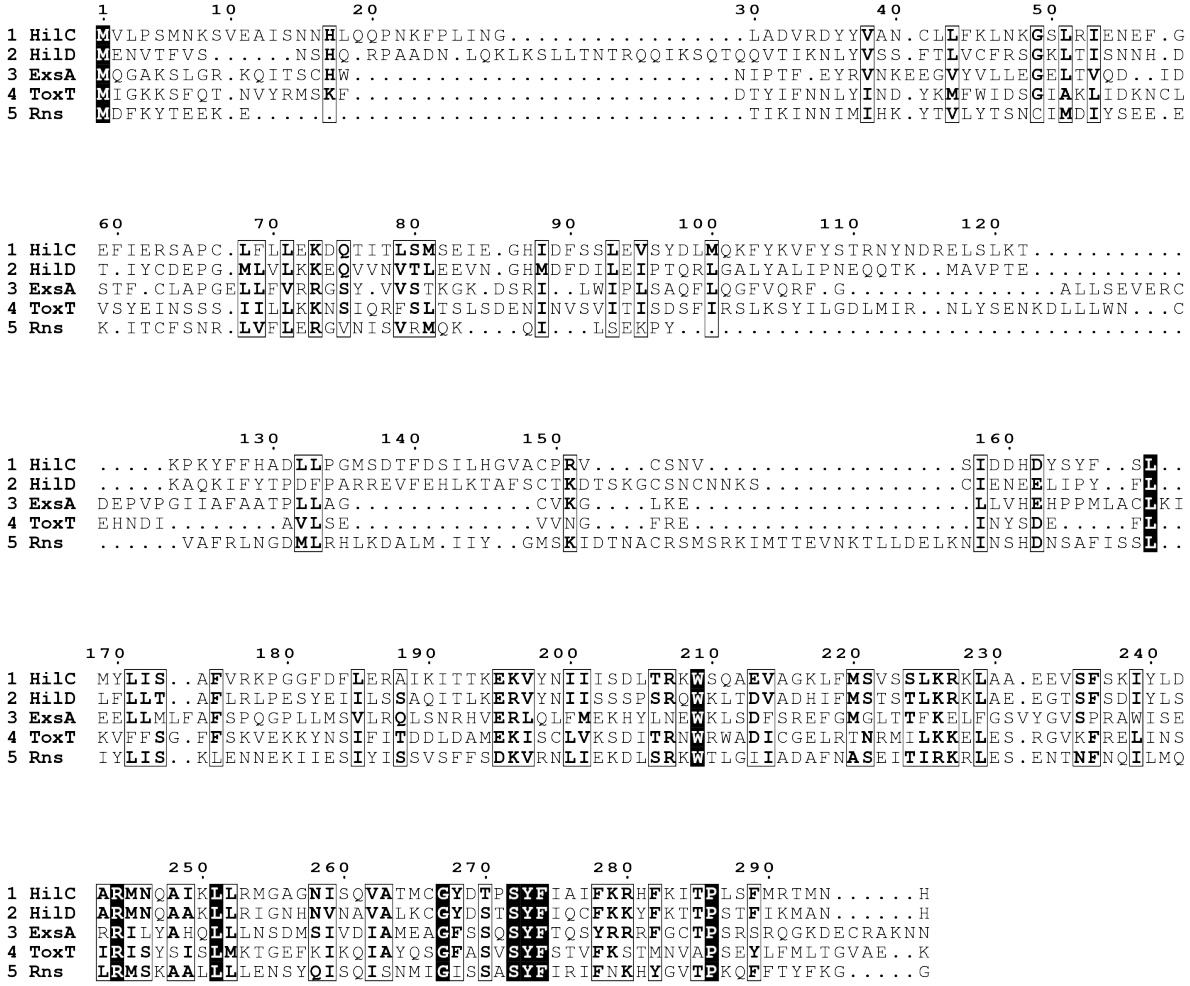


Figure S2.2. Multiple sequence alignment of AraC/XylS proteins, performed with T-Coffee.

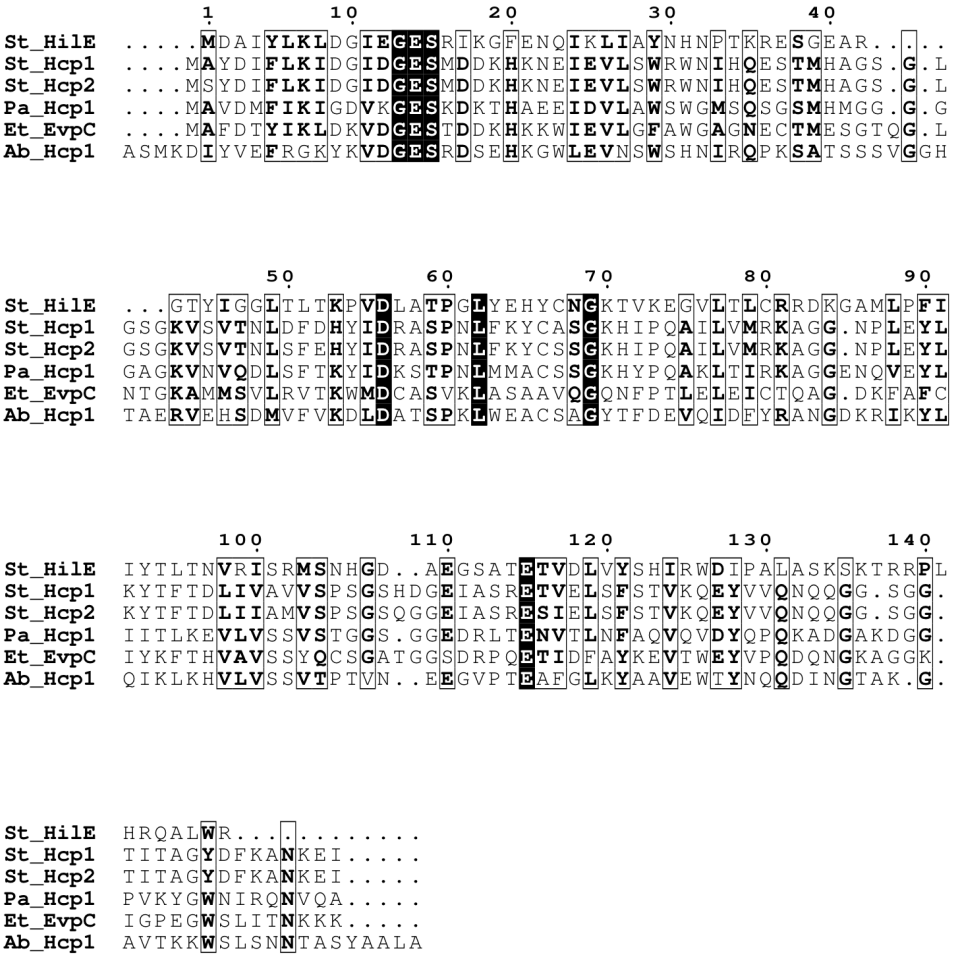


Figure S2.3. Multiple sequence alignment of Hcp-like proteins, performed with Clustal Ω.

Appendix 3. HDX Difference Maps

Heat plots showing the maximal differences in relative deuterium uptake of proteins in the presence of different ligands, compared to when alone, across the entire protein sequence. Gaps in the heat map correspond to regions of the protein sequence with no peptide coverage. Time lengths of the HDX experiment are indicated on the left of figures.

Figure S3.1. Difference in HDX of HilD in the presence of oleic acid.

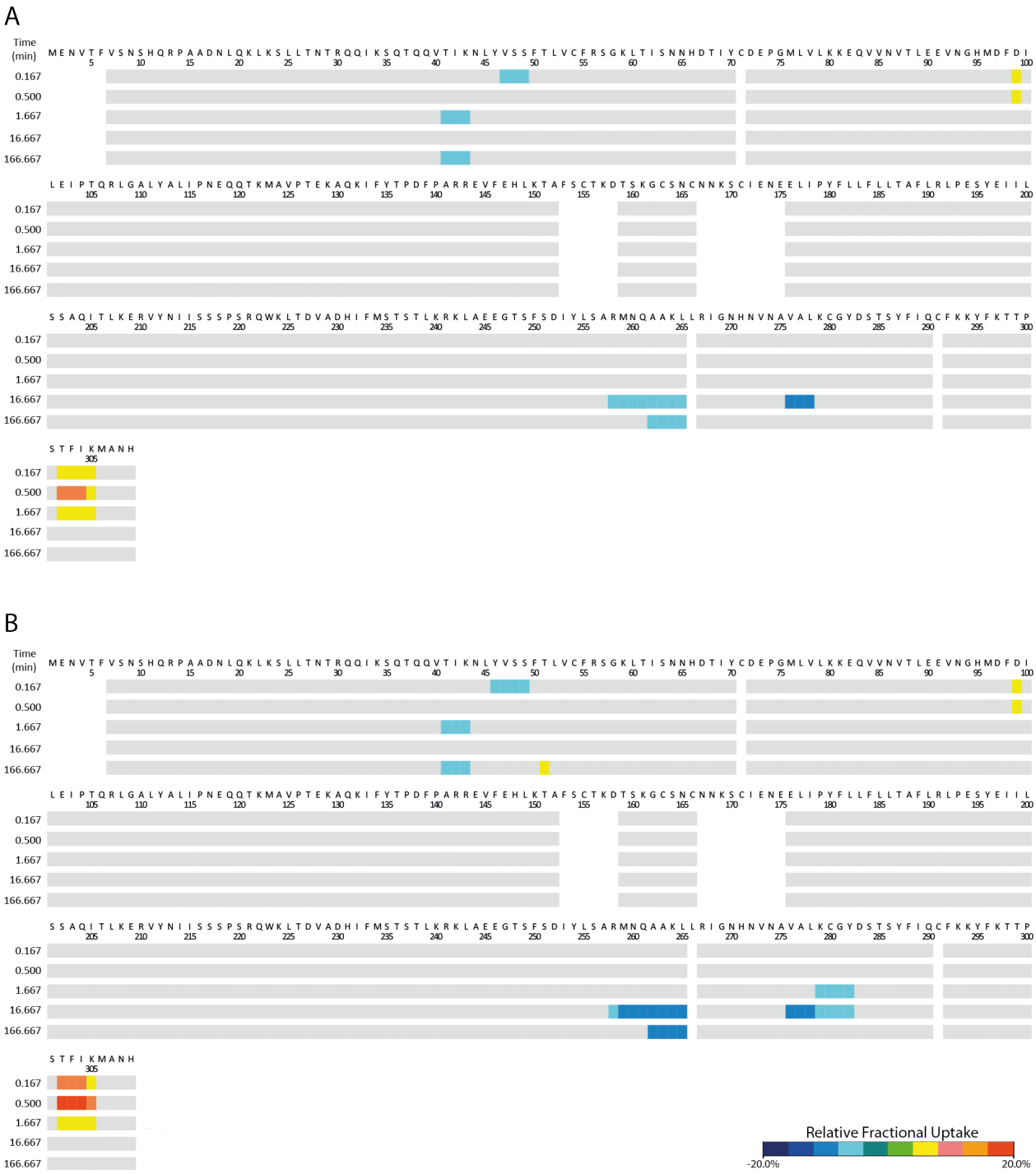


Figure S3.2. Difference in HDX of HilD in the presence of (**A**) C26 or (**B**) C202.

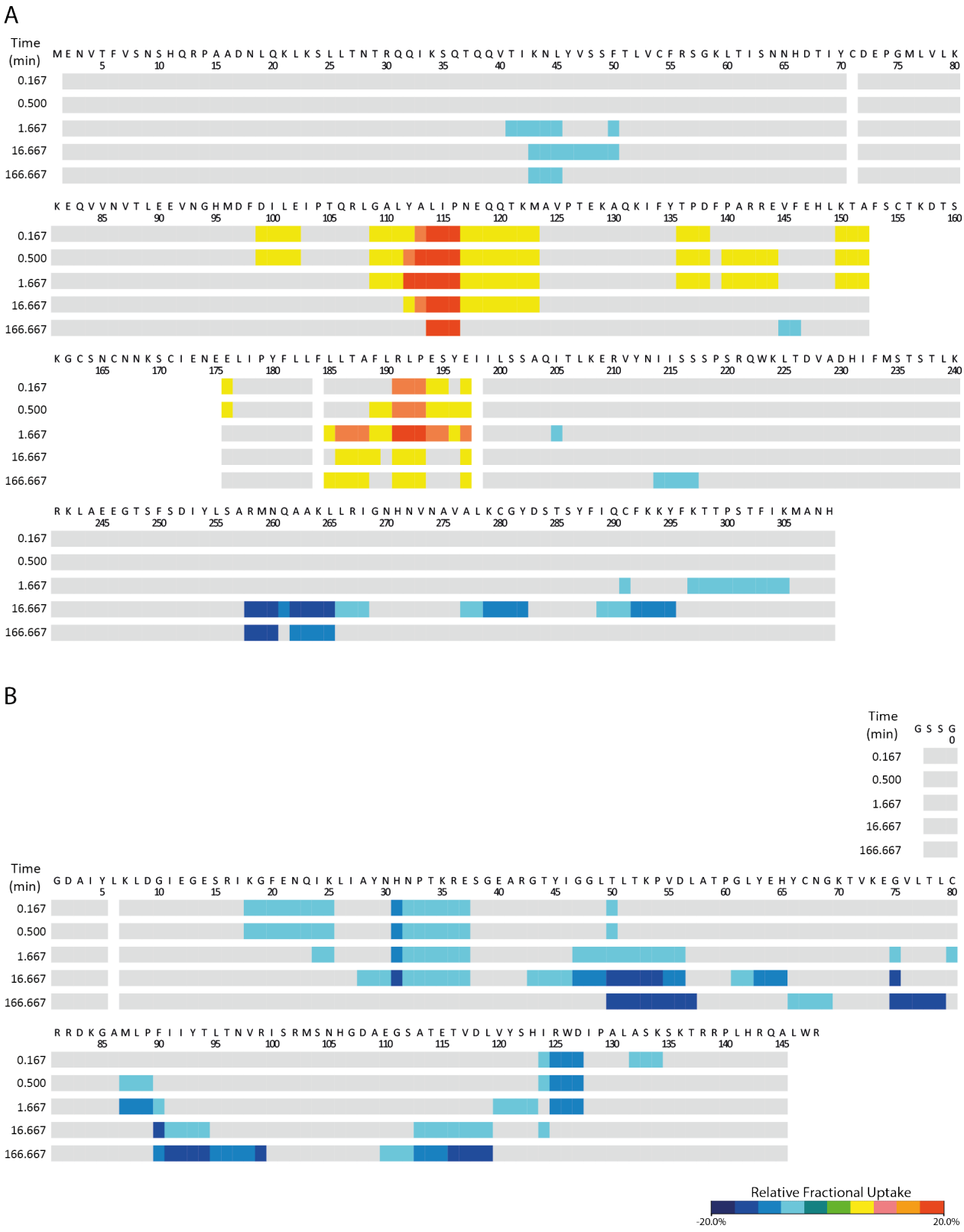


Figure S3.3. Changes in HDX of HilD and HilE when in complex with one another, compared to when alone. (**A**) Differences in HDX of HilD in the presence of HilE. (**B**) Differences in HDX of HilE in the presence of HilD.

Appendix 4. K_d Values (NanoDSF Assay)

Table S4. Calculated K_d values for ligand binding to HilD, determined from the NanoDSF thermal shift assay. K_d's were calculated using the isothermal method at specific temperatures in the range of 50 to 56°C.

Compound	$K_{d,app}$ (μM)	Isothermal K_d (μM)			
		50°C	52°C	54°C	56°C
C26	35.36	11.40	12.24	16.51	24.47
C102	9.66	18.41	20.30	21.11	19.60
C103	-	124.50	96.39	-	-
C107	36.27	11.23	12.88	13.85	13.49
C109	25.65	7.143	6.15	6.42	6.26
C110	27.52	6.29	6.16	5.66	5.19
C178	8.83	6.45	7.40	12.11	17.87
C182	3.44	1.08	1.58	3.29	8.38
C184	11.59	8.65	10.12	15.20	18.43
C187	3.66	9.65	16.07	28.84	35.88
C189	12.55	8.64	10.37	14.84	19.08
C190	8.20	3.36	3.65	5.03	8.04
C202	4.92	1.58	2.43	4.79	9.49
C203	28.55	8.56	9.13	10.10	11.75
C204	2.47	3.43	5.49	15.42	25.01
C205	6.89	4.06	4.50	8.02	14.03
C206	7.62	2.99	3.66	5.64	7.13
C207	10.98	2.56	2.69	3.11	4.47
C219	17.13	7.15	7.24	9.383	12.79
C225	16.41	5.41	6.55	9.03	12.47
C230	9.50	11.07	15.50	23.81	27.60
C231	5.23	3.50	5.58	10.45	13.55
C235	2.52	5.84	7.84	19.55	27.89
C238	-	62.01	49.08	-	-
C245	6.59	3.65	4.48	6.96	11.87
C249	-	123.10	125.10	-	-
C250	1.97	2.23	4.48	16.54	29.35
C251	3.70	4.11	6.71	15.16	21.86
C254	-	67.41	57.11	-	-
C272	4.42	3.99	4.86	9.203	17.98
C276	19.11	6.36	6.73	9.46	19.51
C278	6.34	-	1.49	1.81	4.49