



STRUCTURAL BIOLOGY
COMMUNICATIONS

Volume 73 (2017)

Supporting information for article:

The crystal structure of a *Haemophilus* surface fibril domain, PD1

Jack Wright, Maren Thomsen, Robert Kolodziejczyk, Joshua Ridley, Jessica Sinclair, Glenn Carrington, Birendra Singh, Kristian Reisbeck and Adrian Goldman

Table S1 Construct generation information for HsfPD1-GCN4

<i>PD1-GCN4</i>	
Source organism	<i>Haemophilus influenzae</i>
DNA source	pET16b- <i>hsf</i> ^{1–2414}
Forward primer	GCGCGTATTAAAAACTGATTGTGGACTTGTAGTGGAGATAAA
Reverse primer	CAATTATCTTCAATCTGTTCATGTCTAGTTAACCCATCAGCAC
Cloning vector	pIBA-GCN4tri-His ₆
Expression vector	pIBA-GCN4tri-His ₆
Expression host	<i>Escherichia coli</i> BL21*
Complete amino acid sequence	MKQIEDKIEEILSKIYHIENEIARIKKLIVDFVSGDKDTSVTVESKDNGKRTE VKIGAKTSVIKDHNGKLFTGKELKDANNNGVTVTETDGKDEGNGLVTAKA VIDAVNKAGWRVKTTGANGQNDDFATVASGTNVTFADGNGTTAEVTKAN DGSITVKYNVKVADGLKLDMKQIEDKIEEILSKIYHIENEIARIKKLIKHHHH HH

Hia161-1098	ITFALAKDLGVKTATVSDTLTIGGGAAAGATTTPKVNVTSTTDGLKFAKDAAAGANGDTTV
Hsf1484-2413	ITFALANDLSVKSATVSDKLSLG-----TNGNKVNITSDTKGKLNFAKDSKTGD-DANI *****:***.*:*****.***:*****: . . *:..
Hia161-1098	HLNGIGSTLTDLVGS PATHIDGGDQSTHY--TRAASIKDVLNAGWNIKGVKAGSTTGQS
Hsf1484-2413	HLNGIASTLTDLLNSGATTNLGGNGITDNEKKRAASVKDVLNAGWNVRGVKPASANNQV *****.*****:..* ** * *: * . ***:*****:*****:..***..*:..*
Hia161-1098	ENVDFVHTYDTVEFLSADTETTTVTD SKENGKRTEVKIGAKTSVIKEKDGLKFTGKANK
Hsf1484-2413	ENIDFVATYDTVDVSGDKDTTSVTVESKDNGKRTEVKIGAKTSVIKDHNGLFTGKELK ***:*** ****:*.**.***:***:***:*****:*****:*****:***** *
Hia161-1098	ETNKVDGANATE--DADEGKGLVTAKDVIDAVNKTAGWRIKTTDANGQNGDFATVASGTNV
Hsf1484-2413	DANN-NGVTVTE TDGKDEGNGLVTAKAVIDAVNKAGWRVKTGANGQNDDFATVASGTNV ::*: :*...** . ***:***** *****:***:***.*****.*****:*****
Hia161-1098	TFASGNGTTATVTNGTDG-ITVKYDAKVG DGLKLDGDKIAADTTALTVDGKNANNPKGK
Hsf1484-2413	TFADGNGTTAEVTKANDGSITVKYNVKVA DGLKLDGDKIVADTTVLTVDGKV TAPNNG- ***.***** **:..** *****:..**.*****:****.****.*** *** : :*
Hia161-1098	VADVASTDEKLVTAKGLVTALNSLSWTTAAEADGGTLD-GNASEQEVKAGDKVTFKAG
Hsf1484-2413	-----DGKKFVDASGLADALNKL SWTATAGKEGTGEVDPANSAGQEVKAGDKVTFKAG * * :* *.*. ***.*****:..: . * :* .*: *****:*****
Hia161-1098	KNLVKQEGANFTYSLQDALTGLTSITLGTGNNGA---KTEINKDGLTITPANG---AGA
Hsf1484-2413	DNLKIKQSGKDFTYSLKKELKDLTSVEFKDANGGTGSESTKITKDGLTITPANGAGAAGA .***:***.* :*****:.. *..***: : .*.*: .*:*****:***** ***
Hia161-1098	NNANTISVTKDGISAGGQSVKNVVSGLKKFGDANFDPLTSSADNLTKQNDDAYKGLTNLD
Hsf1484-2413	NTANTISVTKDGISAGNKAVTNVVSGLKKFGDGH-TLANGTVADFEKHYDNAYKDLTNLD *.*****:*****:..:*****:*****:.. . . . : *: *:*****:*****
Hia161-1098	EKGTDKQT P VADNTAATVGDLRGLGWVISADKTTGG-STEYHDQVRNAMEVKFKSGNGI
Hsf1484-2413	EKGADNN-PTVADNTAATVGDLRGLGWVISADKTTGEPNQEYNAQVRNAMEVKFKSGNGI ***:***: * .*****:*****:***** . ***: *****:*****
Hia161-1098	NVSGKTVN G REITFELAKGEVVKSNEFTVKETNGKETSLVKVGDKYYSKEDIDLTTGQP
Hsf1484-2413	NVSGKTLN GTRV ITFELAKGEVVKSNEFTVKNADGSETNLVKVGDMYYSKEDIDPATSKP *****:*** * *****:*****:*****:..:..***.***** *****:***** :..:*
Hia161-1098	KLKDGN TVAAKYQDKGGKVVSVTDN-TEATITNKGS YVTGNQVADAIKSGFELGLADE
Hsf1484-2413	--MTGKT--EKYKVENGKVV SANGSKTEVTLTNKGSGYVTGNQVADAIKSGFELGLADA *: * **: :.*****.... **.*:*****:*****:*****
Hia161-1098	ADAKAAFDD--KTKALSAGTTEI VNAHDKVRFANGLNTKVSAATVESTDANGDKVTTTFV
Hsf1484-2413	AEAEKAFAESA KDKQLSKDKAETVNAHDKVRFANGLNTKVSAATVESTDANGDKVTTTFV

Figure S1 MUSCLE sequence alignment of shared regions in Hia (Hia161-1098) and Hsf (Hsf1484-2413), demonstrating 72% sequence identity (identical residues, red/asterisk; strongly similar residues, colon; weakly similar residues, full stop).

N-TERMINUS	---	DFVSGDKD	TTSVT	VESKDNGKR	TEVKIGA				
C-TERMINUS		TNVT	FADGNGT	TAEVT	---	KANDGSIT	VKYNV		
		*	..*	*:	**	*	.	**	..

Figure S2 MUSCLE sequence alignment of 29 N- and C-terminal residues of HsfPD1 demonstrating 31% sequence identity (annotation as in Supplementary Fig. S1).

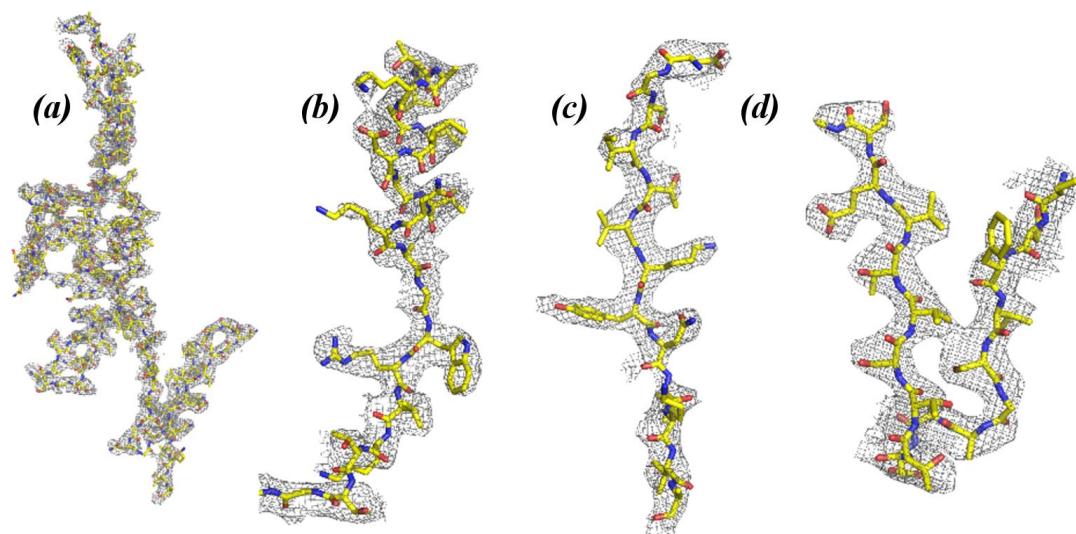


Figure S3 Electron density maps of HsfPD1 after refinement. Representative presentations of $2F_o - F_c$ electron density maps of (a) overall chain D and various regions within chain D: (b) amino acids Val1677 – Ala1698, (c) C-terminus (Asp1730 – Ala1743) and (d) N-terminus (Met1607 – Ser1623). All maps are contoured at 1σ .

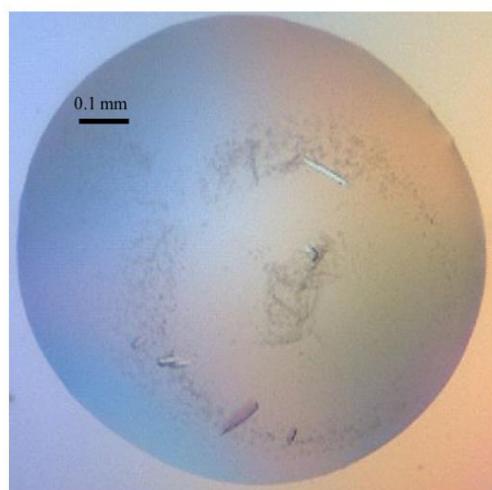


Figure S4 Crystals of PD1.