

Crystal structure of the *Haemophilus influenzae* Hap adhesin reveals an intercellular oligomerization mechanism for bacterial aggregation

Guoyu Meng^{1,*}, Nicole Spahich², Roma Kenjale^{2,#}, Gabriel Waksman³ and Joseph W. St. Geme, III^{2,*}

¹State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Rui-Jin Hospital affiliated to Shanghai JiaoTong University School of Medicine, 197 Ruijin Er Road, Shanghai 200025, P. R. of China

²Department of Pediatrics and Department of Molecular Genetics & Microbiology, Duke University Medical Center, Children's Health Center Rm T901, Durham, NC 27710, USA

³Institute of Structural and Molecular Biology at UCL/Birkbeck, Malet Street, London, WE1C 7HX, UK

[#]Current address is Department of Biochemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514

FIGURE LEGENDS

Supplementary Figure 1. A) The σ_A -weighted $2Fo-Fc$ electron densities of a cross-section of C-terminal β -helix, i.e. residues 801-825, in the self-associating autotransporter (SAAT) domain is countered at 1σ . B) Sequence of *H. influenzae* Hap_S and the secondary structure assignment. The strictly conserved catalytic triad, residues His98, Asp140 and Ser243, are highlighted with yellow boxes. The residues that separate the functional domains of Hap_S are colored with red boxes and the residues in SD1-4 subdomains are colored with green boxes. C) Relative location of the serine protease domain. The surface is colored according to the electrostatic surface potential (negative charges $-4K_B T$ in red and positive charges $+4K_B T$ in blue, with linear interpolation in between). The helical core is shown in a ribbon diagram and colored in grey. The subdomains protruding from the helical spine are labeled and colored in green. The surface diagrams were prepared using programs APBS (Baker *et al*, 2001) and Pymol.

Supplementary Figure 2. Structural comparison between Hap, Hbp and IgA1. A, B) Structural superimposition of the β -spine from *H. influenzae* Hap_S (red) and *E. coli* Hbp (blue) and *H. influenzae* IgA1 protease (green). A major architectural variation, SD2 sub-domain, is labelled. C, D) Structural superimposition between the upper parts of Hap_S (red), Hbp (blue) and IgA1 protease (green). The serine protease domain from Hap_S is colored in magenta. SD1-4 sub-domains are labelled.

Supplementary Figure 3. Intra-molecular interactions between serine protease domain and SD1 sub-domain (A), β -spine (B), SD3 sub-domain (C), SD4 sub-domain (D). Residues involved in hydrogen bonding network are shown in stick and labelled with different colors, i.e. black for those residues in serine protease domain and red for the rest of the structure. The residues that are strictly conserved among Hap, Hbp and IgA1 are underlined in the figure. Hydrogen bonds are shown in dashed lines.

Supplementary Figure 4. The conserved inner core of SAAT domain. A, B) Side and top views of the conserved hydrophobic residues in the inner core of β -helix between residues 732-801. The residues are shown in stick representation. Blue and red colors are used to highlight the repetitive nature in this region. C) Sequence alignment of different turns of β -helix between residues 732-801, suggest a (I/V)XLXXXXX(A/F)X(V/L) sequence motif. These four positions are indicated above the sequence alignment. The orientations of the side-chains of the Ile, Leu, Ala, and Val residues in the β -helix turns of residues 732-742 and residues 772-782 are nearly identical. The same applies to the Val, Leu, Phe, and Leu residues in the β -helix turns of residues 752-762 and 791-801. This repetitive nature is highlighted with green and black arrows on the side of sequence alignment.

Supplementary Figure 5. Overall protein architecture is important for SAAT-type assembly. Sub-domain 2 might play an important structural role in shaping autotransporters into different functional entities such as proteases and self-associating cell-linkers (Figure 8BE). The recent structure of an Hbp mutant (Pdb code: 3AK5),

(Nishimura et al, 2010) shows that deletion of Hbp sub-domain 2 can promote a Hap_S-like packing, suggesting that the overall architecture of the protein is important for the function of the SAATs. A) Hbp mutants packing in a *trans* configuration. B) Hbp mutants packing in a Hap_S-Hap_S-like assembly. Despite Hbp mutant can mimic a SAAT-type binding, but due to the lack of self-complementary surface, the Hbp-Hbp packing is much loose with inter-molecular distance >10Å.

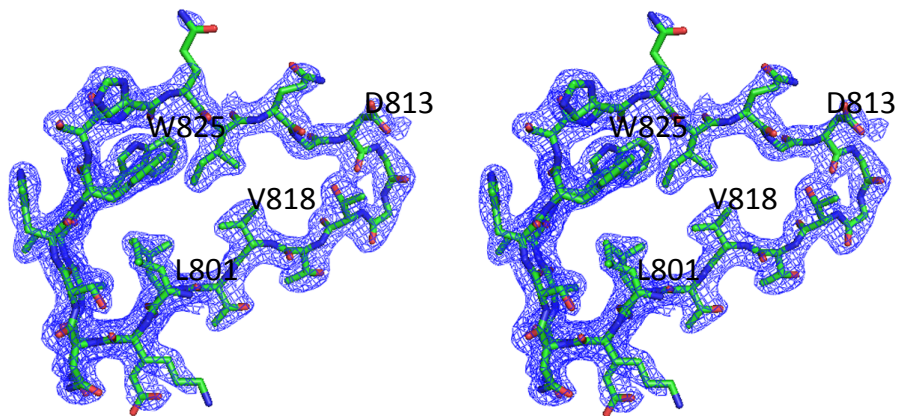
Supplementary Figure 6. Sequence and structural comparison between *H. influenzae* Hap_S and other self-associating autotransporters. A) Sequence alignment of the SAAT domains from different self-associating autotransporters, including IcsA from *Shigella flexneri* (gi 34101173), AIDA-I from *Escherichia coli* O157:H7 str. Sakai (gi 15830650), hypothetical protein CV_0837 from *Chromobacterium violaceum* ATCC 12472 (gi 34496292), YfaL from *Escherichia coli* (gi 2506696), YapG from *Yersinia pestis* (gi 10945162), hypothetical protein SMC02406 from *Sinorhizobium meliloti* 1021 (gi 15964814), AIDA from *Burkholderia cepacia* R1808 (gi 46323824), Antigen 43 from *Escherichia coli* (gi 2506898) and AIDA from *Burkholderia cepacia* R18194 (gi 46317444). Red and blue represent strictly and relatively conserved residues, respectively (Marchler-Bauer et al, 2009). B) Structural superimposition of the crystal structure of *H. influenzae* Hap_S (green) and the homology model of *E. coli* AIDA-I (black). The AIDA-I homology model was generated using the standard protocol implemented in SwissModel (<http://swissmodel.expasy.org>).

Supplementary Figure 7. Hap-mediated interbacterial interaction is enhanced by the presence of secretory leukocyte peptidase inhibitor (SLPI). For the purpose of illustration, only two Hap_S molecules are shown in different surfaces colored in magenta and by electrostatic potential. The SLPIs are shown in space-filled representations and colored in cyan. The dashed line is used to represent the linking polypeptide, i.e. residues 977-1036, between Hap_S and membrane embedded Hap_β. OM stands for outer membrane. The distance of two adjacent C α atoms in a β -strand conformation is about 2.3 Å. Therefore, if loop 977-1036 is fully stretched, the predicted distance between Hap_S and Hap_β can be ~ 135 Å, leaving plenty of space of membrane-bound Hap to interact with each other.

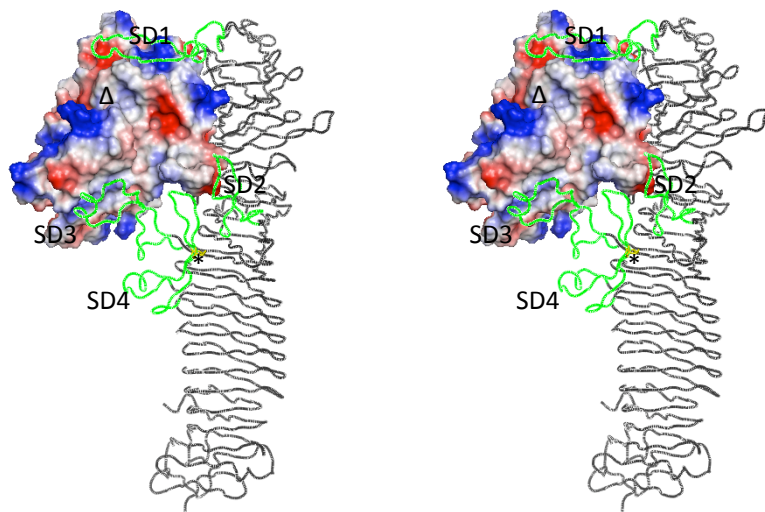
Supplementary Figure 8. Hap_S-Hap_S in *cis* configuration derived from the self-association is agreeable with published Hbp_β:Hbp_β (a Hap_β homolog) packing. The SAAT domains are bracketed, and the inter-molecular distance between the self-adhesive domains in this configuration is >14 Å, implying the *cis* packing is driven by oligomerization *in trans*. The Hbp_β structures (PDB code: 3AEH), (Tajima et al, 2010), which are in scale with Hap_S structures, are used to judge whether the oligomerization of Hap_S presented in this manuscript will leave enough packing space for the membrane anchoring domain, Hap_β.

Supplementary Figure 1 Meng et al, 2011

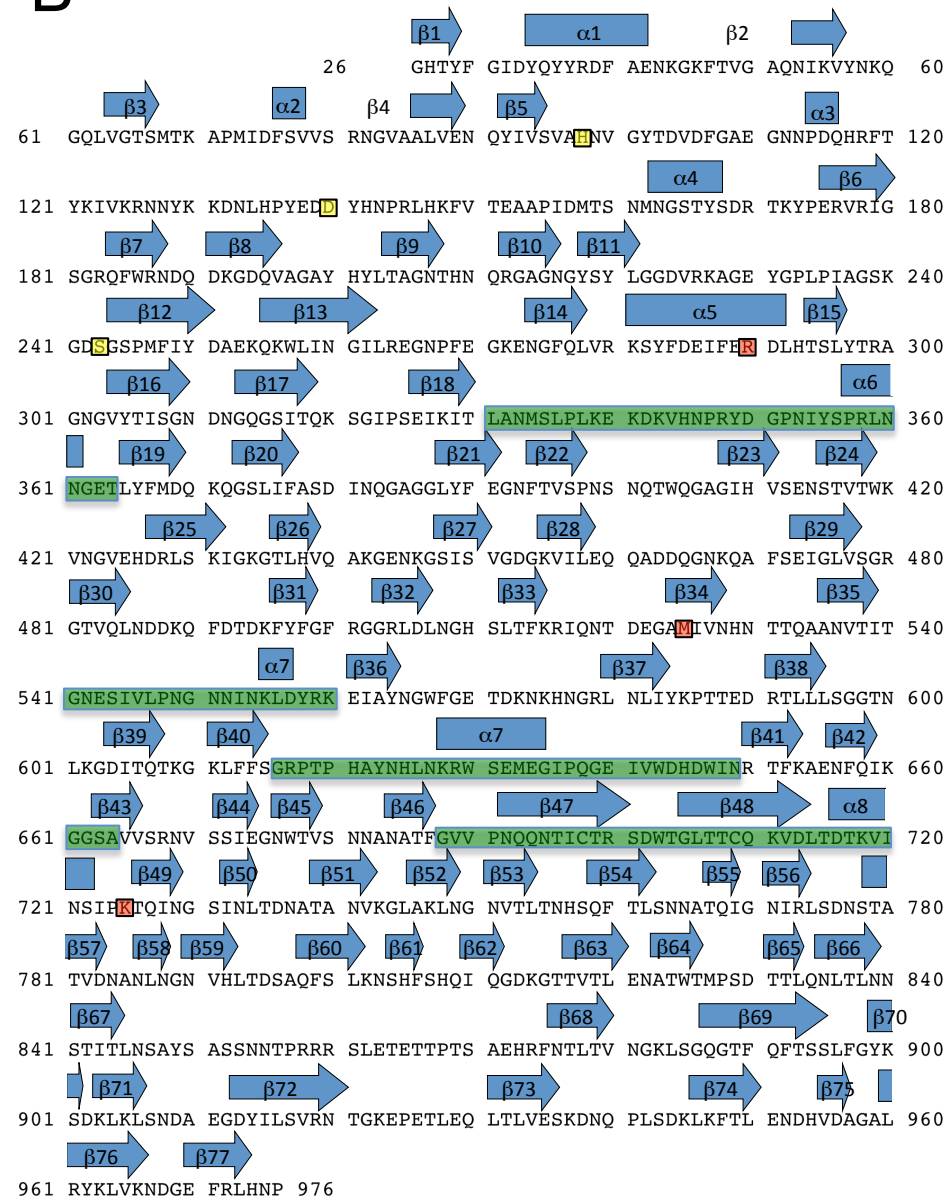
A



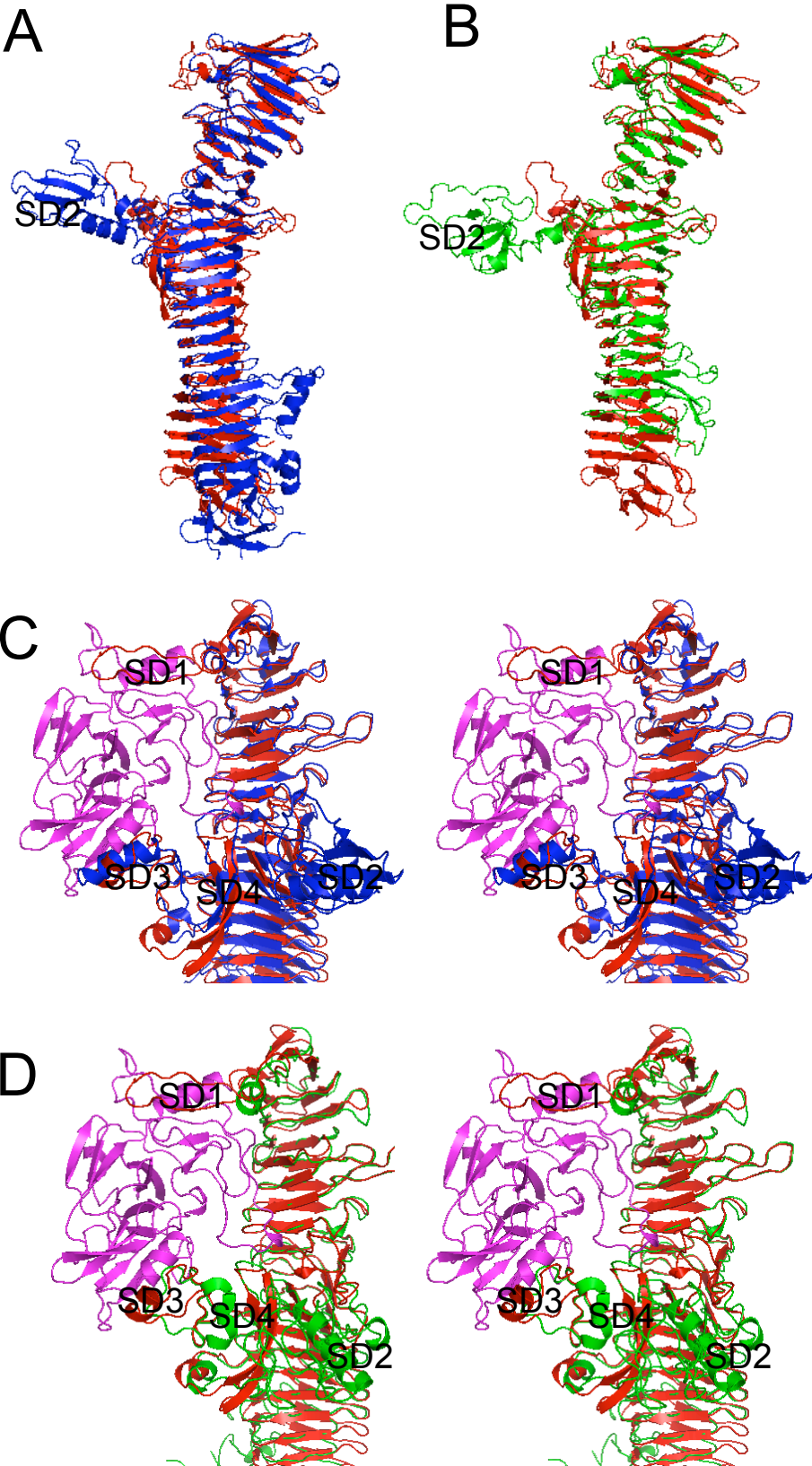
C



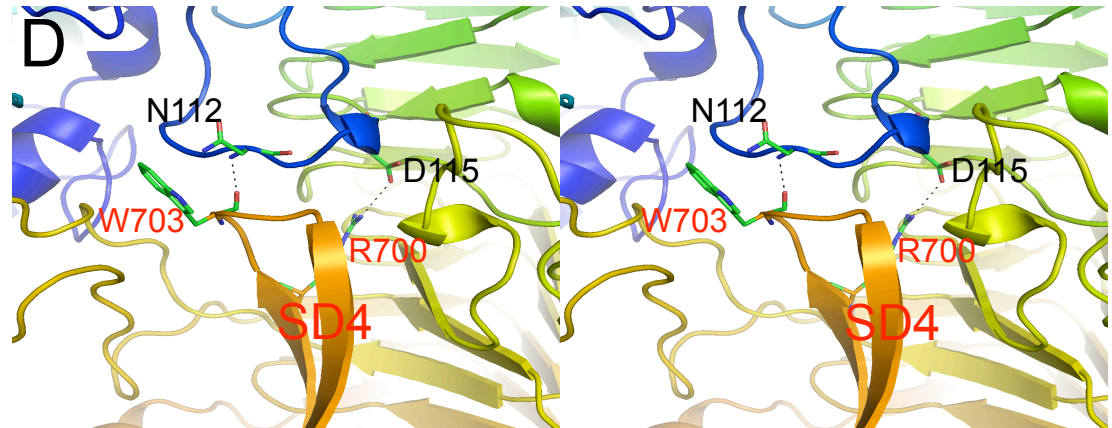
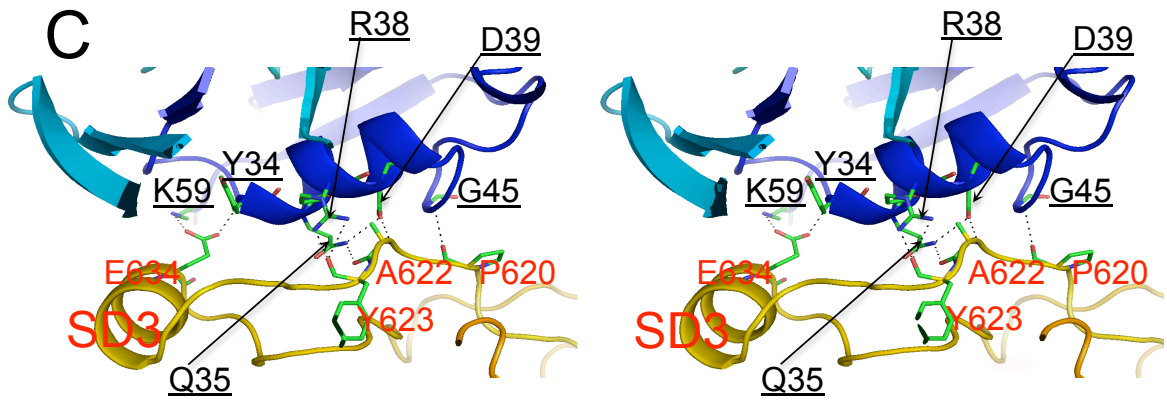
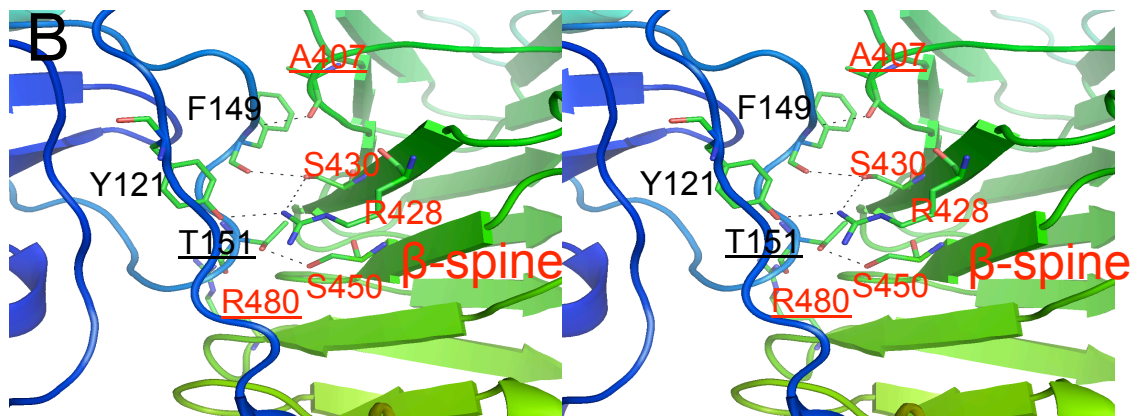
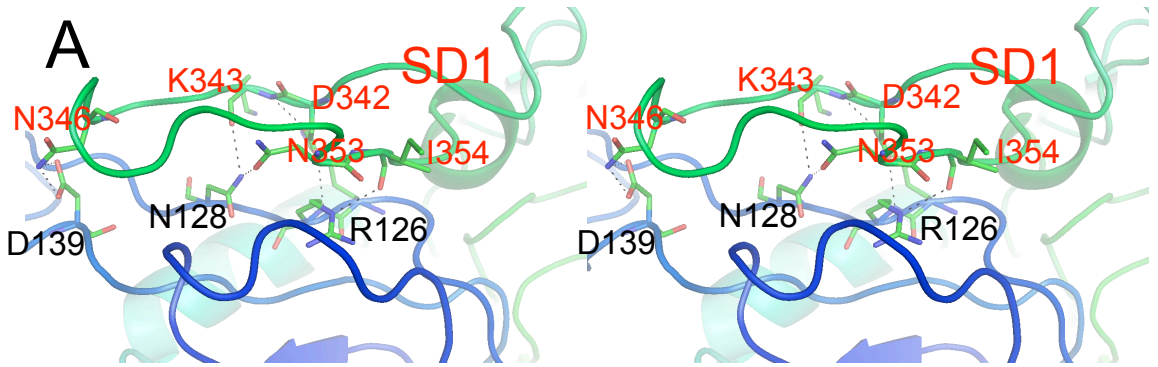
B



Supplementary Figure 2 Meng et al, 2011

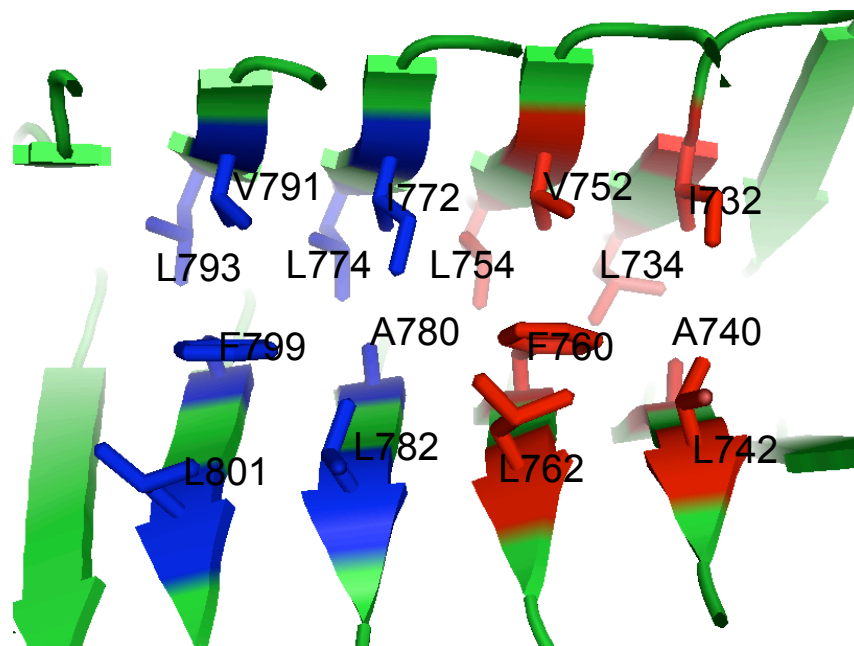


Supplementary Figure 3 Meng et al, 2011

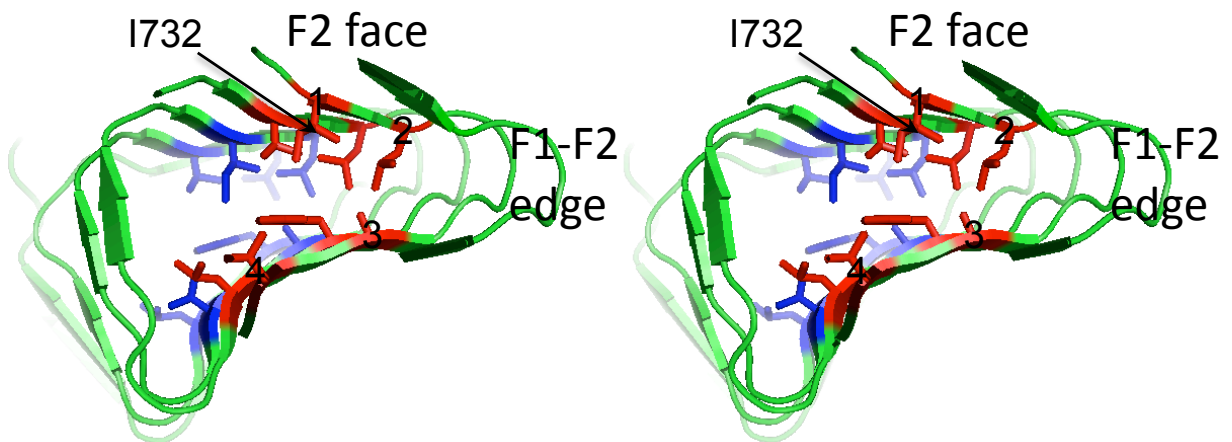


Supplementary Figure 4 Meng et al, 2011

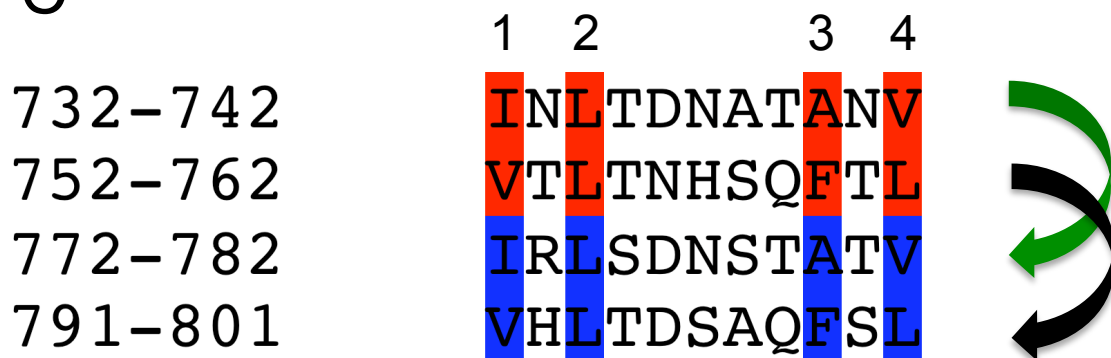
A



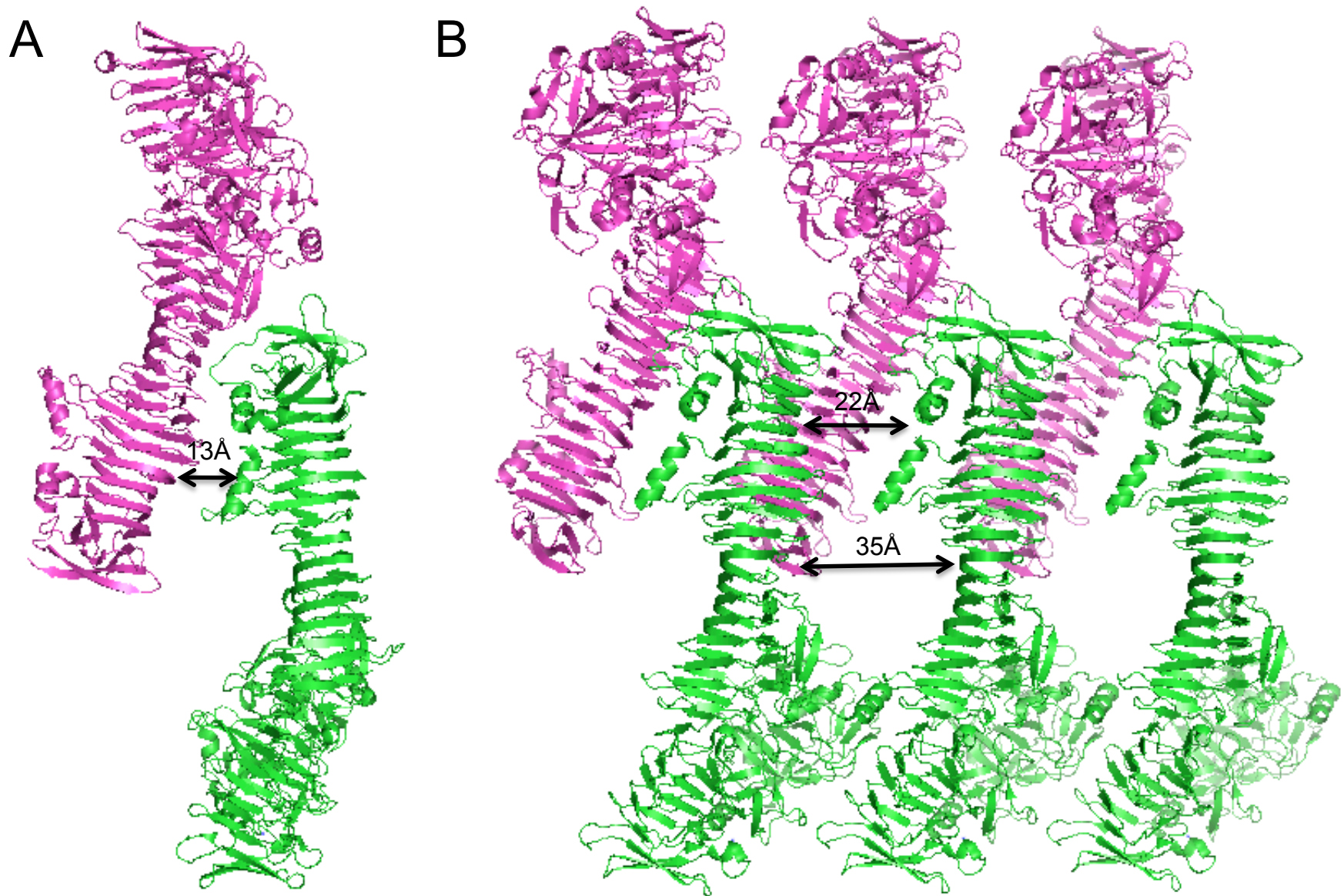
B



C



Supplementary Figure 5 Meng et al, 2011

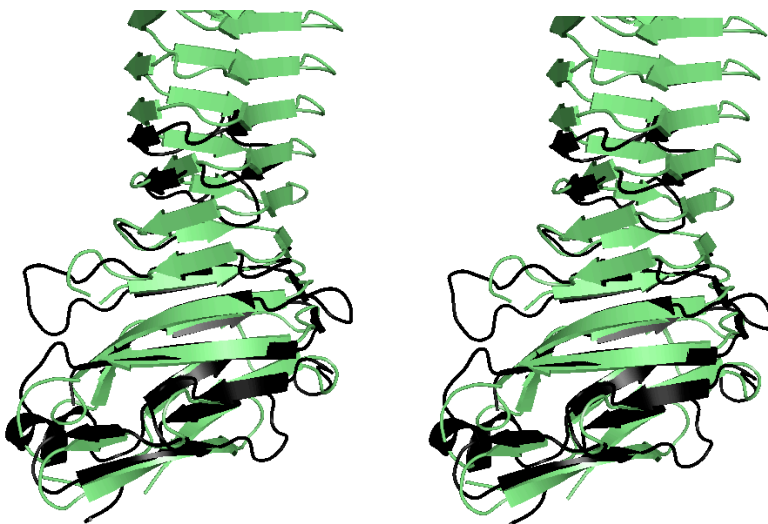


Supplementary Figure 6 Meng et al, 2011

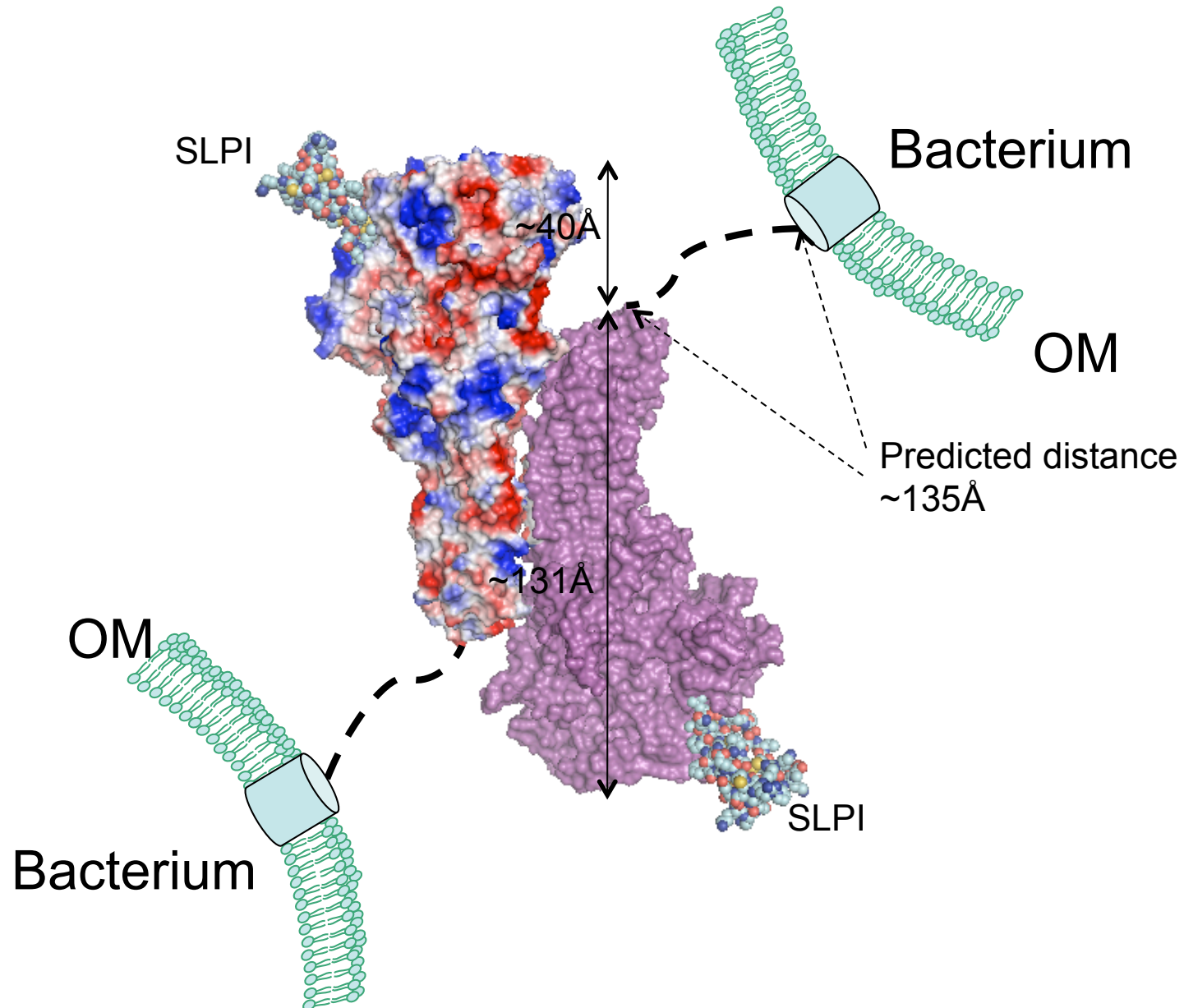
A

		10	20	30	40	50	60	70	80					
gi 34101173	557	GTVEa	---MTRTa	--GVIVnk	--gaTLNFsg	---mNOTV	--NTLLNsg	-----tVLINn	-INAPflp	--dpVIVTG	612			
gi 15830650	423	GSFTvn	--aGQQA	g---NTTVgh	--rgTLTLA	----aGGSLS	--GRTQLskgas	---mvlngD	VVstGDIVn	-----aGEIRF	484			
gi 34496292	601	NYVAg	---VQNVava	QINGv	----fNNAd	----gVLRV	--LAVNapg	-----sVNNA	--GAWQt	ggsglpGNGVV	656			
gi 2506696	718	QDIQsi	--dAISSg	--TIDIsd	--gtVLRLt	gg--dtSVALn	--ASLFNg	gdgtlvnat	gdgVTLTge	LNTNl	-----eTDSL	784		
gi 10945162	296	GETHIn	--agNLKLa	--NTHFfg	---sPISGnp	--ntRLILek	STLDTtvq	----gssVFID	khsiWNm	-----lGDSNI	357			
gi 15964814	722	ANIVnw	--eIVTLda	adVTIld	---gALEAg	sepetGLFLtn	GSVLNgsdaf	--alagnMAID	gtSQFIa	-----fGEDIA	790			
gi 46323824	403	VRLTld	--tGSHAY	--gDIVNhp	addgSPTPal	---tDVALS	--NASTWtgat	--avrtLSL	sdSRWvt	-----aDSSVG	469			
gi 2506898	517	GTLTl	---nDSTVt	--tDVIaqr	--gtALKL	---gSTVLn	--GAIDPtn	-----vtLASG	--ATWNip	d--natVQSVV	575			
gi 46317444	551	GLGgp	---aTMTGg	--gMVRIe	---pGATFg	---gYGSVa	--GDVVnyg	-----tvSVANal	SGLAsgh	--tgnLEIKG	609			
.....*.....*.....*.....*.....*.....*.....*.....*.....														
gi 34101173	613	n	-----MTLEKn	--GHVILn	nnsss	-----nvvgq	TYVQk	-----GNWHGK	g--GILSL	--GAVLGNDn	--SKTDRLE	665		
gi 15830650	485	dn	-----gtTPNAAL	--SRAVaks	ns	-----pvtfHKL	Tt	-----TNLTGQ	g--GTINM	--RVRLDGS	---NASDQLV	538		
gi 34496292	657	qqgl	--tgvSNFN	ns--GTITL	qsadigaag	avagkLLVI	tggqtagas	ggGTYTSN	g--GQLRL	--DAVLNQG	gaaSLNDMLV	732		
gi 2506696	785	ylsnvtvn	GNLNTs	GAVSLqngv	-----agdT	TLTVn	-----GDYTTG	--GTLLE	--DSELNG	dd--SVSDQLV	842			
gi 10945162	358	h	-----hLDILDs	--GRHDL	nnpg	-----ktgnQLI	IIn	-----GDYFSD	n--GTLIF	--HSQLAG	dd--SVTDHIL	410		
gi 15964814	791	ngvy	--sisGNLV	na--GTVDM	qddg	-----tgdVLSV	g	-----GNYTGA	enSTVYI	--DTYLGDD	a--SPTDHIT	848		
gi 46323824	470	s	-----VALNDs	--TIAFA	epvar	---alatp	RTLvv	-----tGDYAAH	n--GKLV	--HTTLQDD	a--SPTDRLV	524		
gi 2506898	576		-----DDLSha	--GQIHft	strt	---gkfv	pATLkv	-----KNLNGQ	n--GTISL	--RVRPDMA	g--NNADR	628		
gi 46317444	610	dl	---tnAGLLQ	l--GGSGV	g	-----nLTTI	i	-----gGYSQ	Sg--GTAV	--NTYLAD	g--AASDRLI	659		
.....*.....*.....*.....*.....*.....*.....*.....*.....														
gi 34101173	666	It	---GHASG	ITYVAVT	NEGGSG	dd---kTLEGV	QIISTds	-----sDKNAF	IQK	g-----RIVAGS	YDYRLKQ	722		
gi 15830650	539	In	---gQATG	KTWLAFT	NVGNs	nlgv--atTGG	IRVVDA	ngq--attEEG	AFALS	r-----PLQAG	AFNYTLNR	602		
gi 34496292	733	Vdrta	VSGG	GATRVSVN	NVGGLG	gytgng	PSDGI	ELIKVl	da--gysAANAF	SLAa	-----PLGAG	AFQYSLRQ	799	
gi 2506696	843	Mn	---GNTAG	NTTVV	VNSITG	Ige---pTSTG	IKVVDF	aadptq	fgNNAQF	SLAg	-----gYVNM	GAYDYTLVE	906	
gi 10945162	411	Ik	---GNTGG	HNTVR	VINV	NEGEn---kTDSG	IQLEIV	rg---iSDG	EFSQV	g-----RITAG	AYEYRLGR	467		
gi 15964814	849	Vg	---GDTAG	STAI	SIGNAG	GPSa---gTVAG	IRVVG	vag---aSSG	TFVLAN	anseiket	geaAITRG	AYAYALRQ	916	
gi 46323824	525	Id	---gGHA	AGDTG	I	VVKRAG	GTga---gTTVG	IPIVET	rng--gttDVS	AFTL	dagsdgy--ragfg	TL	SAGGYDYMLER	595
gi 2506898	629	Id	---gGRAT	GKTI	LNLV	NAGNS	agql--atSGK	GIVVEA	ing--attEEG	AFVQ	gn-----RLQAG	AFNYSLNR	692	
gi 46317444	660	Vg	---aggID	GSSMLK	VTVN	VGPG	gaq---tTGDG	IEVVQV	tng--atsSAG	AFSLS	g-----gTVS	GAGYSYFLAK	722	
.....*.....*.....*.....*.....*.....*.....*.....*.....														
gi 34101173	723	gtvsglnt	NKWY	LTSQMD	NQ	742								
gi 15830650	603	ds	-----dEDWY	LRS	ENAYR	617								
gi 34496292	800	ad	-----gQNWY	LQTTDL	VLP	814								
gi 2506696	907	d	-----nNDWY	LRSQ	EVTP	920								
gi 10945162	468	gkd	---elsKNWY	LSSD	ITDY	485								
gi 15964814	917	vd	-----nDDWY	LQSTL	AED	931								
gi 46323824	596	ggrg	--grtDDWY	LVSAA	QPQ	614								
gi 2506898	693	ds	-----dESWY	LRS	ENAYR	707								
gi 46317444	723	ggaatgtg	DNWY	LRNTV	PPK	742								

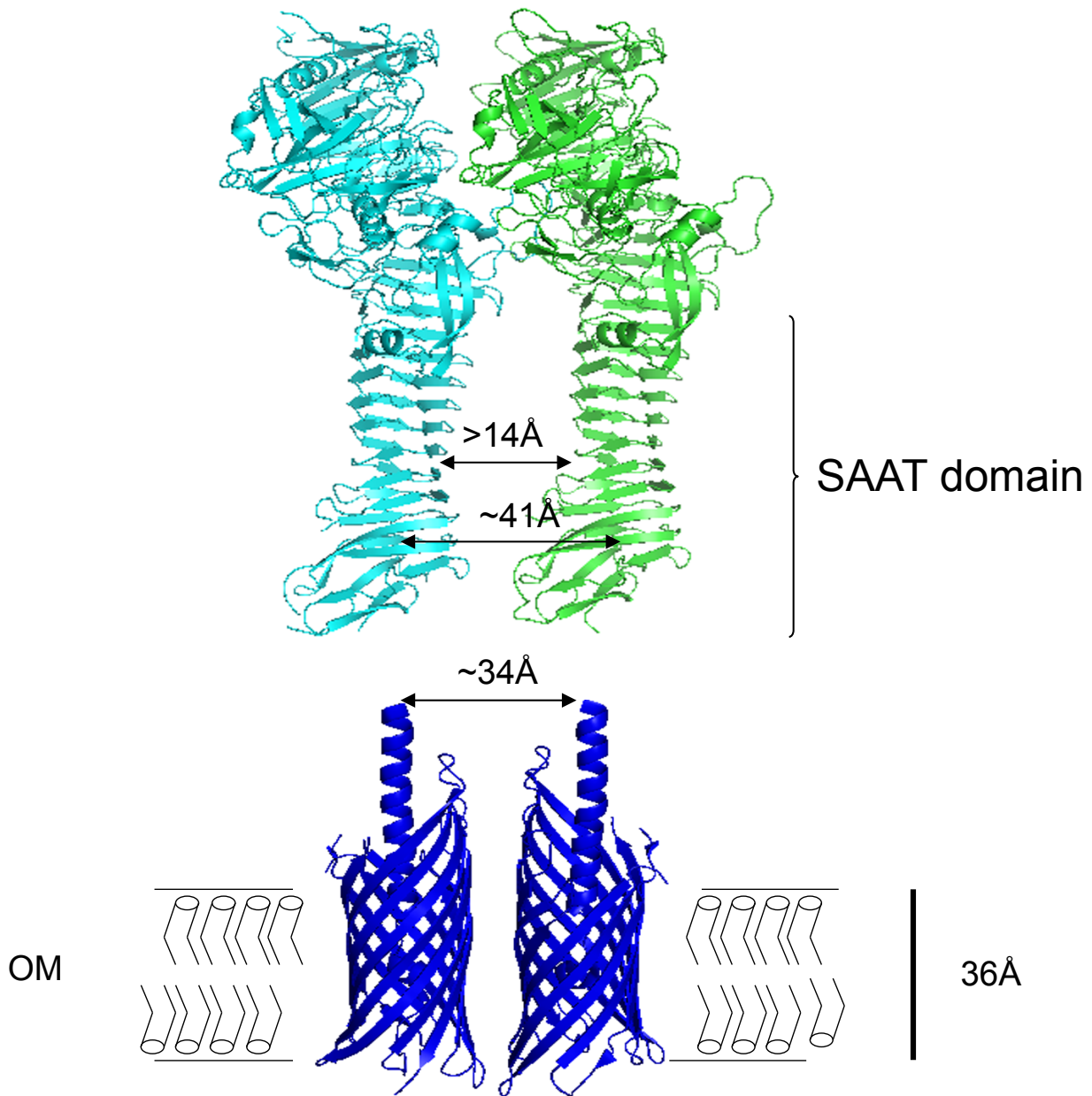
B



Supplementary Figure 7 Meng et al, 2011



Supplementary Figure 8 Meng et al, 2011



Supplementary Table 1. Dynamic light scattering assay of purified Haps.

Temperature (°C)	Rd* (nM)	Polydispersity (%)	Baseline	SOS
4	986.1	33.2	0.097	1.029
10	1447.7	33.6	1.000	0.981
16	1509.7	42.2	0.999	1.708
22	1717.3	48.8	0.999	3.253
37	1902.6	56.1	1.003	8.722

*Hydrodynamic radius (Rd) was calculated from 200 independent measurements.

REFERENCES

- Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**(18): 10037-10041
- Tajima N, Kawai F, Park SY, Tame JR (2010) A Novel Intein-Like Autoproteolytic Mechanism in Autotransporter Proteins. *J Mol Biol* **402**(4): 645-656