Supplementary Materials

Additional Details on Model Size & Structure, Residue alignment

Although models were predominately right-handed, the handedness of the CsgA protein is not definitively known. The RaptorX structure takes a beta helical structure with a more triangular core, and is right handed.

Tian-LH had both Ser, and Gln (residues 49, 72, 94, 117, 139) inward facing and aligned, but the Asn and other Gln row had residues N54, N77 and Q83 outward facing. Tian-RH had the Ser row inward facing and aligned, both Gln repeats were inward facing and aligned except Q60 and Q72, and the Asn repeat all outward facing except N54. The RaptorX structure had all Ser facing inward and aligned, the Gln (residues 49, 72, 94, 117) and Asn outward facing and aligned except Q139 and N144, and the second Gln row inward facing and aligned except for Q150. The other models from FALCON@home and Quark were not thoroughly investigated, as none formed beta-helical structures.

	Residue Pair		RAPTOR	CsgA-map	Tian-LH	Tian -RH	RobA-3
Val	115	135	6.97	8.40	8.30	8.46	10.94
Val	115	137	4.54	4.75	4.63	4.69	6.50
Val	115	140	12.77	9.62	11.49	9.98	9.10
Val	135	137	6.38	6.58	6.45	6.49	6.47
Val	137	140	9.82	9.12	10.21	9.08	8.96
Val AVERAGE			8.10	7.69	8.22	7.74	8.39
Leu	45	57	11.53	9.58	14.98	12.07	11.30
Leu	45	59	10.56	8.57	13.21	10.03	9.62
Leu	45	68	4.72	4.75	4.89	5.00	6.26
Leu	57	59	6.46	6.56	6.58	6.34	7.13
Leu	59	68	9.36	9.12	13.33	10.02	7.64
Leu	68	92	8.23	8.02	8.31	8.20	8.14
Leu AVERAGE			8.48	7.77	10.21	8.61	8.35
Phe	97	118	8.95	9.51	8.05	6.92	7.90
Phe	118	142	13.98	9.16	8.17	7.94	9.46
Phe AVERAGE			11.46	9.33	8.11	7.43	8.68

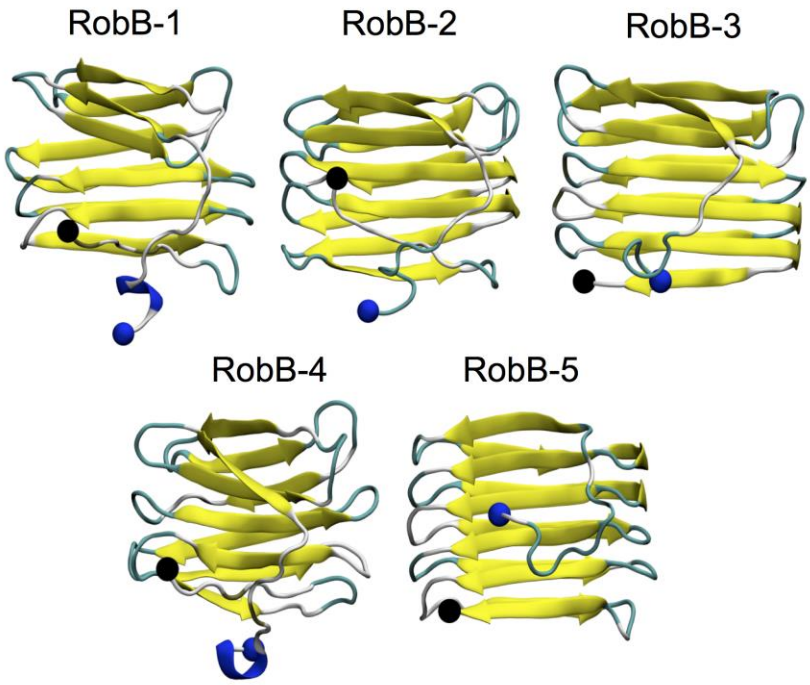
Intermolecular carbonyl distances were calculated for selected models for the residues

Val, Leu and Phe. These are measured to compare with ssNMR data[1]. Indeed, intermolecular averages for Val, Leu and Phe are all above 7 Å, as these pairs include not only neighboring residues in adjacent strands (Val 95 to Val 117), residues separated by a third residue in the same strand (Val 115 to Val 117), but also residues on opposite beta sheets (Leu 25 to Leu 39) and residues in adjacent strands, but offset (Phe 77 to Phe 98). For the distances measured, the CsgA-map model has Val and Leu intermolecular distances closest to those measured in experiment (~ 7 Å).

CsgB

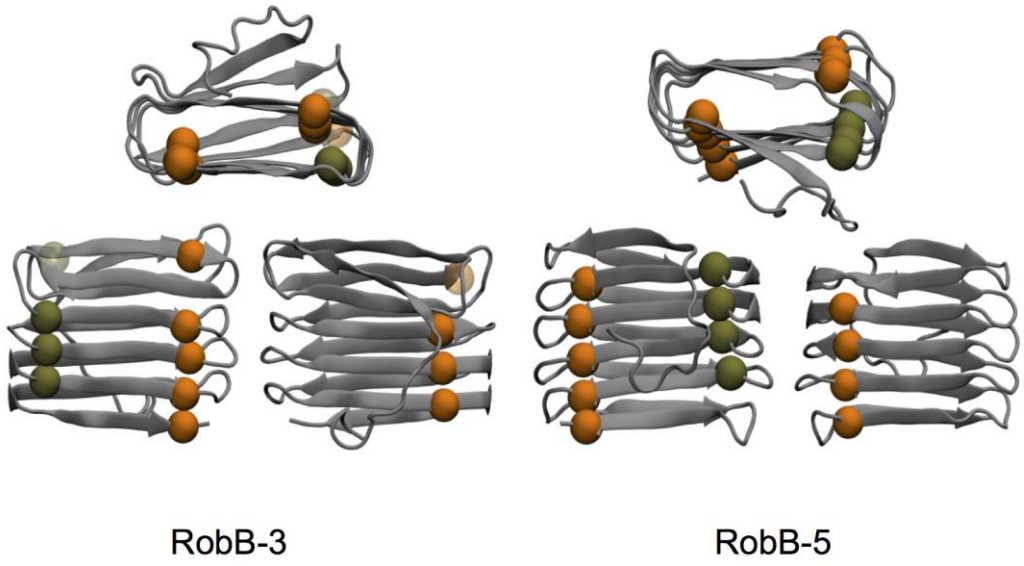
Models RobB-1 and RobB-3 have 3 helical repeats and beta-meander like strands near the N terminus. Model RobB-2 has a helical central portion, and meander-like strands at either terminus. Model RobB-4 is less ordered and contains beta-bends and beta-meander regions. Model RobB-5 is beta-helical, with an unstructured region near the N-terminus. Model RobB-3 is left-handed, while model RobB-5 is right handed. Each Robetta model has an unstructured region near the N terminus.

S1 Fig. CsgB models produced using Robetta server.



Of all models, model 5 is the only fully beta-helical model. Other models have partial helical portions combined with meander motifs.

S2 Fig. Residue alignment for CsgB models.



For highlighted residues, Gln is shown in orange, and Asn in green. In RobB-3, the first residue in each row (near the N-Terminus) is separated from the rest by a beta-strand, and do not fully align. For RobB-5, all three repeats are inward facing and aligned.

Equilibration Analysis

The model with the highest percentage of beta-sheet secondary structure was CsgA-map, followed by Quark-7, Falcon-7, Quark-2, and Quark-3. We do note that the FALCON@home and Quark models tended to have a larger amount of secondary structure because of the beta-strands in terminal regions, whereas Robetta models generally have an unstructured N terminus. These areas are not necessarily assumed to be beta-structured.

Internal protein hydrogen bonds were calculated, and we interpret higher numbers of hydrogen bonds to indicate stability. A loss of hydrogen bonds during the simulation may correspond to a loss of secondary structure, often associated with unfolding and increased Rg or iRMSD. The models with the highest number of hydrogen bonds at the end of equilibration are CsgA-map, Quark-7, Quark-2, and RM-5.

For SASA, the models with the lowest final values include CsgA-map, Tian-RH, and RobA-5. Although Tian-RH has a low final SASA value, it began at a low value and increased significantly over the course of the simulation.

S1 Table. Metric Summary Table.

	Last ns average			
	iRMSD (Å)	Total Beta Structure (%)	H Bonds (#)	SASA (Å^2)
Tian-LH	4.31	31.84	26.03	8125.50
Tian-RH	4.87	44.86	24.11	7577.09
Falcon-1	5.92	31.57	20.66	9444.45
Falcon-2	7.13	32.25	17.35	9971.03
Falcon-3	4.05	31.90	19.11	8683.79
Falcon-4	5.36	36.59	20.32	9600.67
Falcon-5	5.49	37.73	23.22	8943.49
Falcon-6	4.10	31.71	18.97	9919.74
Falcon-7	4.47	48.24	26.34	8509.63
Falcon-8	3.47	38.47	19.80	9352.60
Falcon-9	4.97	29.89	15.53	9942.90
Quark-1	3.66	44.14	23.26	8243.18
Quark-2	3.74	47.08	27.00	8180.83
Quark-3	6.42	46.88	24.07	9502.97
Quark-4	4.42	39.51	26.09	8400.41
Quark-5	7.20	46.58	22.59	9164.75
Quark-6	5.11	45.55	23.62	9433.48
Quark-7	3.35	48.41	27.27	8698.34
Quark-8	3.74	39.42	24.49	8006.76
Quark-9	9.38	43.15	25.36	9095.17
Quark-10	3.07	31.21	21.61	8413.07
Raptor	4.42	26.88	14.32	8747.03
RobA-1	2.93	32.29	21.65	8071.31
RobA-2	3.76	24.85	20.36	8111.09
RobA-3	2.06	32.91	22.41	8004.46
RobA-4	2.55	28.80	20.65	8143.95
RobA-5	2.30	35.79	21.66	7897.41
CsgA-map	1.76	50.16	30.04	7372.40

The iRMSD, total percentage of beta structure, number of hydrogen bonds, and solvent accessible surface area are presented for each model. Values indicate the mean value calculated over the last nanosecond of a 10 ns equilibration run. The lowest values for

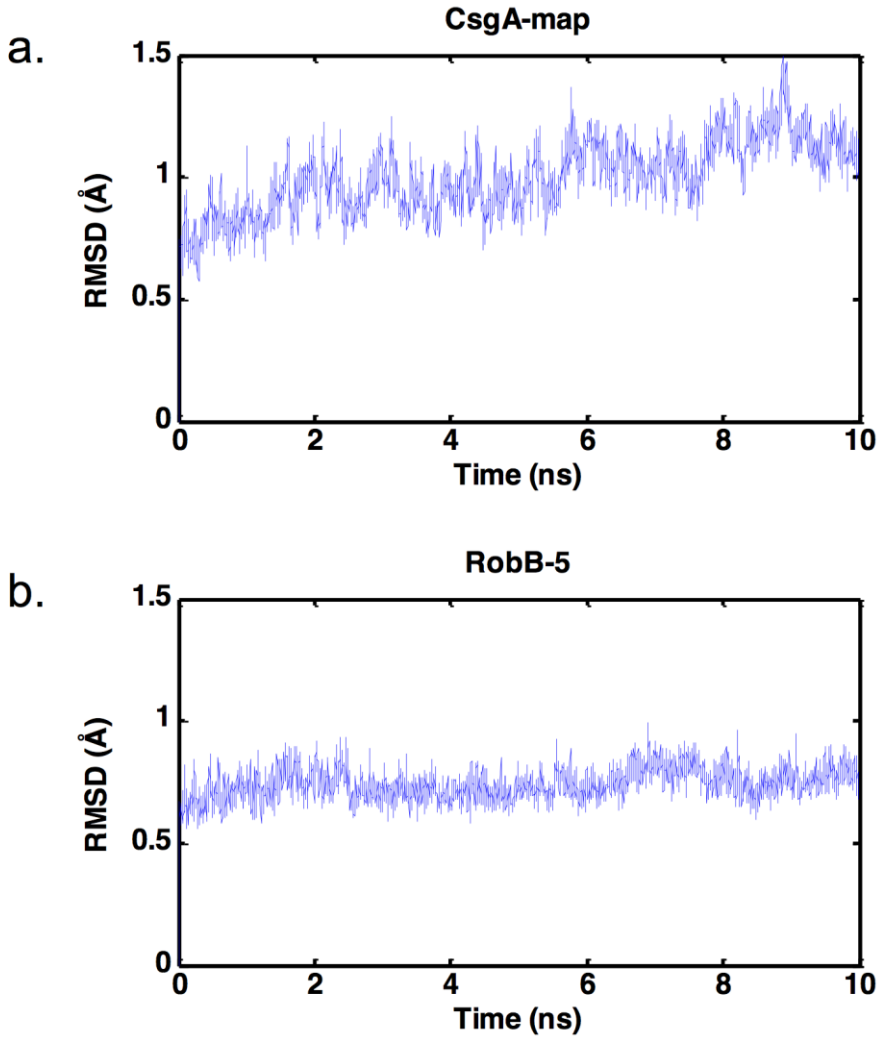
iRMSD and SASA are highlighted, as well as the highest values for beta sheet content and hydrogen bonds.

S2 Table. Metric Summary for CsgB.

	Last ns average			
	RMSD (Å)	Total Beta Structure (%)	H Bonds (#)	SASA ($Å^2$)
RobB-1	2.18	48.69	33.56	8331.23
RobB-2	2.84	49.55	34.45	7887.75
RobB-3	1.80	51.37	35.36	7632.91
RobB-4	3.20	42.86	26.55	8632.18
RobB-5	1.16	54.72	41.49	7690.53

For iRMSD, RobB-5 had the lowest value. For total beta-sheet content and number of hydrogen bonds, RobB-5 had the highest. RobB-5 also had the second lowest amount of solvent-accessible surface area, after RobB-3. The lowest values for iRMSD and SASA are highlighted, as well as the highest values for beta sheet content and hydrogen bonds.

S3 Fig. iRMSD plots for refined CsgA (A) and CsgB (B) models.



Model RobB-5 maintained an iRMSD below 1 Å for the entirety of the simulation, and both models displayed improved stability over the initial models used.

Creation of CsgA-map

The CsgA sequence was threaded onto the structure of CsgB model RobB-5. This CsgB model was the only structure to fit all four criteria outlined in the Introduction. The sequence was threaded such that conserved residues would align on the inside of CsgA, flanking the turns in a way similar to AgfA. The mapping was done using MODELLER[2], and the sequences are shown below.

```
CsgB:
-AGYDLANSEYNFAVNELSKSSFN
QAAIIGQAGTNNSAQLRQGGSK-
LLAVVAQEGSSNRAKIDQTGDY
NLAYIDQAGSANDASISQGAYG
-NTAMIIQKGSGNKANITQYGTQ
KTAIVVQRQSQMAIRVTQR--*

CsgA:
--GVVPQYGGGGNHGGGGNNSGPN
SELNIYQYGGGNSALALQTDARN
SDLTITQHGGGNGADVGQGSDD
SSIDLTQRGFGNSATLDQWNGKN
SEMTVKQFGGGNGAAVDQTASN
SSVNVTQVGFGNNATAHQY-*
```

References

1. F. Shewmaker, R. P. McGlinchey, K. R. Thurber, P. McPhie, F. Dyda, R. Tycko and R. B. Wickner, *Journal of Biological Chemistry*, 2009, **284**, 25065-25076.
2. B. Webb and A. Sali, *Protein Structure Prediction*, 2014, 1-15.