Supporting Information

The structure of the *Clostridium thermocellum* RsgI9 ectodomain provides insight into the mechanism of biomass sensing

Brendan J. Mahoney¹⁻², Allen Takayesu¹, Anqi Zhou¹, Duilio Cascio², Robert T. Clubb¹⁻³*

¹Department of Chemistry and Biochemistry, ²UCLA-DOE Institute of Genomics and

Proteomics, and the ³Molecular Biology Institute, University of California, Los Angeles, 611

Charles E. Young Drive East, Los Angeles, CA 90095, USA.

*To whom correspondence should be addressed:

Prof. Robert T. Clubb, Department of Chemistry and Biochemistry, University of California, Los Angeles, 602 Boyer Hall, Los Angeles, CA 90095.

		Rsgi N	19 ins	
Acetivibrio_thermocelius_ATCC_27405 Acetivibrio_thermocelius_BC1 Acetivibrio_cellulolyticus Acetivibrio_mesophilus Acetivibrio_straminisolvens_JCM_21531	1 MKITGVIVRIHKDRAIIRTDDN 1 MKITGVIVRIHKDRAIIRTDDN 1 MKITGVIVGIONDKAIVLTDEK 1 MKITGIIVEIODDKAIIOTDDE 1	LLAVKRHINDMMVGQIVSFDAN LLAVKRHINDMMVGQIVSFDAN FLAIKKRDGMFIOQTVKFGKE LIAVKKRDGMVGKTISIDES LAVKKRDGMVVGKTISIDES	EVHKVESKKYKYAASGKRIEKVOKTPKIKNF Evhkveskkykyaasgkriekvoktpkiknf Dickskskyvik <mark>yspskriekteesekykni</mark> Dyvkperkvykyapsskriekt <mark>ps</mark> ingikgf Dyvkeer <mark>kvykyapsskriektssaskik</mark> gf	75 75 75 75 43
Acetivibrio_thermocellus_ATCC_27405 Acetivibrio_thermocellus_BC1 Acetivibrio_cellulolyticus Acetivibrio_mesophilus Acetivibrio_straminisolvens_JCM_21531	76 SRINNIKEFSRVD 76 SRINNIKEFSRVD 76 TRLINIKNFSRINSTKNEN VS 76 SRISNIKGFSRVNI 44 SRVSNIKGFSRVNI	DIKNFSRVAATKETSODSPO DIKNFSRVAATKETSODSPO DIKNFSRVAATKETSODSPO DIKNFSRISSKESSOE VIKNFSRISGKESSEKSSO	ESKVENFSRVVDFSRV	24 24 50 14 92
		ТМ	RsgI_CRE	
Acetivibrio_thermocellus_ATCC_27405 Acetivibrio_thermocellus_BC1 Acetivibrio_cellulo/ticus Acetivibrio_mesophilus Acetivibrio_straminisolvens_JCM_21531	125 WNF SRVSN	NFSRIISNIKNFSRIASIAAAFV NFSRIISNIKNFSRIASIAAAFV NFSRIITRISGFSRVVGIAAAFV NFSRIISGIKNFSRIASVAAAFV NFSRISGIKNFSRIASVAAAFV	LIFLF <mark>GRN</mark> VMLNNS <mark>SDSEYAYVS</mark> VDVNPSVE1 LIFLFGRNVMLNNS <u>SGSEYAYVS</u> VDVNPSVE1 MVFMFSHVALLNKGASNEYACIGIDINPSIE1 LIFLF <mark>GRNVMLNNGSRSEYAYVSVDINPSIE1 LIFLF<mark>GRNVMLNNGSRSEYAYVSVDINPSIE</mark>1</mark>	92 92 25 82 60
Acetivibrio_thermocellus_ATCC_27405 Acetivibrio_thermocellus_BC1 Acetivibrio_cellulolyticus Acetivibrio_mesophilus Acetivibrio_straminisolvens_JCM_21531	193 FTINSKHKVIVTSAINQDASEV 193 FTINSKHKVIVTSAINQDASEV 226 IVIDNEYKVIETVÄKNSDAESV 183 FVINSKHKVTDTSAINQDALEV 181 FAVNSKHKVTDTSAINQDALEL	LDGLELKEKDLKSALVMVLEKA LDGLELKEKDLKSALVMVLEKA LNGLKLEGIPLKSALSISEIANKA LEGLELIGKDLKSAMALVVEKA LESLNVEEKDLKSAMVTVLEKA	SLGY ISD. DKNYVLVSMALNDKNKK TR 2 SLGY ISD. DKNYVLVSMALNDKNKK TR 2 SCGY FGD. SKNYVVVSMALNNISKEYNTDK 2 LIGY ISDDDKNYVLVSMALNDEGKNSKNTE 2 SLGY ISDDDTNYVLVSMALNDKNKNTE 2	63 99 57 32
Acetivibrio_thermocellus_ATCC_27405 Acetivibrio_thermocellus_BC1 Acetivibrio_cellulolyticus Acetivibrio_mesophilus Acetivibrio_straminisolvens_JCM_21531	264 DKREEK I DELKET I EOG I EALD 264 DKREEK I DELKET I EOG I EALD 300 SVEEK LNNLKOELEENS I AI EO 258 EKKEK LNELKES I GKOMEELD 233 AKRDEKLDELKEN I EOGMEEAG	IDTIVHRTVTVDLEERMALEN IDTIVHRTVTVDLEERMALEN STNIIGTVLLTPEEMEAAKEK IDSVTVKTVVVALEEREMALEN IDTIIHKTVTVSFEEREMALEN	LISMORYYLYLEAKEKOMDITIDEVKSSKIS JISMORYYLYLEAKEKOMDITIDEVKSSKIS JISMORYYLYLEAKEKOVEISINEAKEBSLE JISMORYYLYLEAIEKOVDITIDEAKSSKIA JUSMORYYLYLEAIEKOVDITIDEXKSSKIA JUSMORYYLYLEAKEKOIDVTIDEVKSAKLS	38 38 74 32 07
	Pro	o-rich linker	S1C	
Acetiviibrio_thermocellus_ATCC_27405 Acetiviibrio_thermocellus_BC1 Acetiviibrio_cellula/pitcus Acetiviibrio_cellula/pitcus Acetiviibrio_mesophilus Acetiviibrio_straminisolvens_JCM_21531	339 DLIEKIEDNTELAPTPTP. 339 DLIEKIEDNTELAPTPTP. 375 ELVGNIEVAEEŠVVOTPTP 333 ELIERIEGOSEPTPMPSI.TPT 306 ELIVKIEDDSEPVPTPSITPTPI	VPPETPEPTPTASEATPS VPPETPEPTPTASEATPS .ITEVPTPTLTPSPVASPT .KSAEPSPTPVPTPTASEMIPS .KSPEPSPTPVPTPSPSETAPT	USPVESKSPEAVPELGSEIEILGESVVLVT 4 ISPVESKSPEAVPELGSMEIEILGESVVLVT 4 ISPVESKSPEAVPELGSMEIEILGKSIVTVV 4 ISPVESKSPEVSPGLSAMEITTFSESVVSVT 4 ISPVASKSPESGPELGAMEIAALGESVVLIT ISPVASKSPESGPELGAMEIAALGESVVLIT	07 07 43 05 82
		*	*	
Acetivibrio_thermocellus_ATCC_27405 Acetivibrio_thermocellus_BC1 Acetivibrio_cellulolyticus Acetivibrio_mesophilus Acetivibrio_straminisolvens_JCM_21531	408 AYDENRKVVSOGSGFAVGTGLF. 408 AYDENRKVVSOGSGFAVGTGLF. 444 AYNKNKKAILHGSGFCVAPGLF. 406 VFDONKSIISGGSGFCVGTGLF. 383 AYDENKKVISOGSGFSVGAGLF.	ATNYHLVKDGVVVKITAGDGKV ATNYHLVKDGVVVKITAGDGKV VTNYHVIENAVCLKITTNDGRT VTNYRLVEDGVVINITEKDGKV ATNYHLVKDGVVI <mark>KIT</mark> ANDGKV	YDVDG IVK YDKAKDLALLKTRVETGVNPLKL 4 YDVDG IVK YDKAKDLALLKTTVETGVNPLKL 4 YNVEG IVK MDKKTDLAILK TVKDPGVKALET YNVDG IVK YDKARDLVVLK TTVDTGMKPLK I 4 YDVDG IVK YDKTRDLAIIK ATVDTGVKPLK I 4	82 182 18 80
			*	
Acetivibrio_thermocellus_ATCC_27405 Acetivibrio_thermocellus_BC1 Acetivibrio_cellulolyticus Acetivibrio_mesophilus Acetivibrio_straminisolvens_JCM_21531	483 GTKKSLTKGSRIVAIGKANGAK 483 GTKKSLTKGSRIVAIGMANGAK 519 DTKPVLDKGDRIAAIGSONGNK 481 GSKASLAKGDWIVAIGKSKGLT 458 GSKTSLVKGGKIVAIGKSTGLK	NTVTKCSIKSLAVDGLTDAIEL ITVTKGSIKSLAVDGLTDAIEL (AVSEGSIVGFKKRASIDLIGI ITVSEGSIKAFKTDKIEL ITVSEGKINGLKIEGFADAIDL	AS IS KESTGGP VFDM CNVVG I TAYG I SKO AS IS KESTGGP VFDM GNVVG I TAYG I SKO VPLL KONSGGP L FDMEONV I G I ASYGVSDN 5 AP I SSESSGGP VLDR CNVVG I TSYG I SNK 5 AS IS KESSGGP VFDM GNVVAVATYGVSKO	57 57 93 51 32
			CTD	
Acetivibrio_thermocellus_ATCC_27405 Acetivibrio_thermocellus_BC1 Acetivibrio_cellulolyticus Acetivibrio_mesophilus Acetivibrio_straminisolvens_JCM_21531	558 NVNAV I PADYVADWYKELSKHSI 558 NVNAV I PADYVADWYKELSKHSI 594 S LNLA I PIDYVSGWYKELSKHSI 552 NVNSA I PADY I VDWYKELAF 533 GVNAA I PADYVADWYKELAFYSI	GNIRIVEKTUVFDSDFEFNFV GNIRIVEKTUVFDSDFEFNFV KNTKVIRTUSFDDNYDLNIV NNIKVIRKTLAFDSEFELNFV NNIKIVEKTUVFDSDFELNFA	YKIIRALENEDAATYFGCMTDELYKDETRK 6 YKIIRALENEDAATYFGCMTDELYKDETRK 6 YFNIVNALEKEDMESYFSNMTEELYSEETKN 6 YKIIRALENEDINAYFNCMIDELYTNETKD 6 YKIIRALENEDVTAYFECMTDELYKDETKE 6	32 32 68 23 07
Acetivibrio thermocellus ATCC 27405	633 NLEVLETTYDLAYNIES INVVS	SEEQARVSYVYT I NKEAGPNE		07

Figure S1. Protein sequence alignment of RsgI9 sequences in *Clostridia.* Alignment of the amino acid sequences of RsgI9 in *Clostidium thermocellum (Acetivibrio thermocellus)* and related species using Clustal Omega¹. The domains are boxed and labeled in the same colors, and the catalytic triad residues in the S1C peptidase domain are indicated by the asterisks.



Figure S2. Comparison of RsgI9's CTD with a structurally similar NTF2-like protein. (A) Surface representation of RsgI9 C-terminal domain (CTD) (pink), and (B) a NTF2-like structural homolog shown in the same orientation (cyan) (PDB accession code 3ROB). The structures are similar and have a DALI Z-score = 13.7 and PDBeFold Z-score = 7.8. (C) Cartoon figure showing the extended β -sheet found in other NTF2-like proteins (red) that is missing in RsgI9 (pink). RsgI9 secondary structural elements are labeled. The positioning of the structures is the same in each panel.



Figure S3. 2-D NMR spectrum of RsgI9S1C-CTD. A 2-D TROSY-HSQC NMR spectrum of ¹⁵N-labeled RsgI9^{S1C-CTD} was acquired prior to ¹⁵N-TRACT correlation time determination. The spectrum indicates a well-folded protein with backbone amide signals well-dispersed.



Figure S4. Rsgl9's catalytic triad resembles inactive form of homolog DegS. (A) Alignment of the catalytic triad residues found in Rsgl9 with those in the active (magenta, PDB: 4RQZ) and inactive (yellow, PDB: 4RQY) forms of the *E. coli* DegS protease. Rsgl9's configuration more closely resembles the inactive form, as the distance between the His Nɛ2 and the Ser/Thr hydroxyl group is too far to make a productive hydrogen bond. (B) Alignment of the oxyanion hole residues adjacent to the nucleophilic Ser/Thr, with the same structures and coloring as in panel A. Both Rsgl9 and the inactive form of DegS presumably fail to stabilize the tetrahedral intermediate during proteolysis due to a deformation relative to the active conformation. Notably, the amide N_{*i*-2} is flipped out while the carbonyl C_{*i*-3} points inward.



Figure S5. Fit of SAXS data and *ab initio* reconstructions. (A) Fit of experimental scattering curve to that predicted for the Rsgl9^{S1C-CTD} crystal structure by the FoXS server (χ^2 value of 1.23). (B) *Ab initio* bead model and electron density reconstructions of Rsgl9^{S1C-CTD}, using DAMMIF/N and DENSS, respectively. The maximum dimension of the model (84 Å) is consistent with the D_{max} observed from the data (82 Å). (C) *Ab initio* models for intact Rsgl9^{ecto}. The length of the model is in agreement with the value of D_{max} derived from the SAXS data (160 Å).



Figure S6. *C. thermocellum* **Rsgl architecture vs. DISOPRED3 disorder predictions.** The primary sequences of the nine Rsgl proteins from *Clostridium thermocellum* were analyzed using DISOPRED3². Values above 0.5 are predicted to be disordered, while those below are likely structured. The domain architecture of each Rsgl is shown below each plot based on previous structures, UniProtKB annotation, and this work³⁻⁶. A good agreement is seen between annotated domains and predicted structured regions.

REFERENCES

- 1. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*. 2011;7(1):539.
- 2. Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*. 2015;31(6):857-863.
- 3. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-Prot. In: Humana Press; 2007:89-112.
- 4. Yaniv O, Fichman G, Borovok I, et al. Fine-structural variance of family 3 carbohydratebinding modules as extracellular biomass-sensing components of Clostridium thermocellum anti-sigmal factors. *Acta Crystallogr D Biol Crystallogr.* 2014;70(Pt 2):522-534.
- 5. Grinberg IR, Yaniv O, Ora LO, et al. Distinctive ligand-binding specificities of tandem PA14 biomass-sensory elements from Clostridium thermocellum and Clostridium clariflavum. *Proteins: Structure, Function, and Bioinformatics.* 2019;87(11):917-930.
- 6. Bahari L, Gilad Y, Borovok I, et al. Glycoside hydrolases as components of putative carbohydrate biosensor proteins in Clostridium thermocellum. *Journal of Industrial Microbiology & Biotechnology*. 2011;38(7):825-832.