Crystal Structure of the Full-Length ATPase GspE from the *Vibrio vulnificus* Type II Secretion System in Complex with the Cytoplasmic Domain of GspL

Supplementary Material

Connie Lu[1], Konstantin V. Korotkov[1,#] and Wim G. J. Hol[*]

Department of Biochemistry and Biomolecular Structure Center, University of Washington,
Seattle, Washington 98195

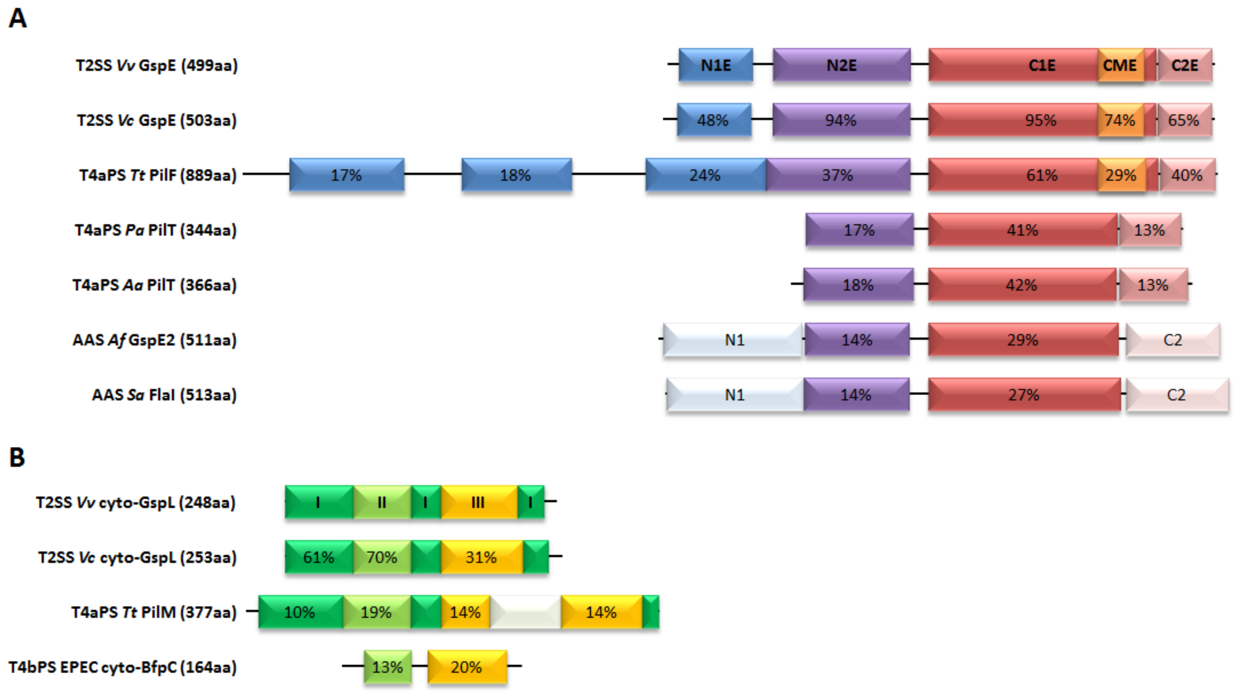
[1]These authors contributed equally to this study.

*Correspondence to: Wim G. J. Hol, Department of Biochemistry and Biomolecular Structure Center, University of Washington, Seattle, Washington 98195. Phone: (206) 685-7044, Fax: (206) 685-7002, E-mail: wghol@u.washington.edu

[#] Present address: Department of Molecular & Cellular Biochemistry, University of Kentucky, Lexington, Kentucky 40536.

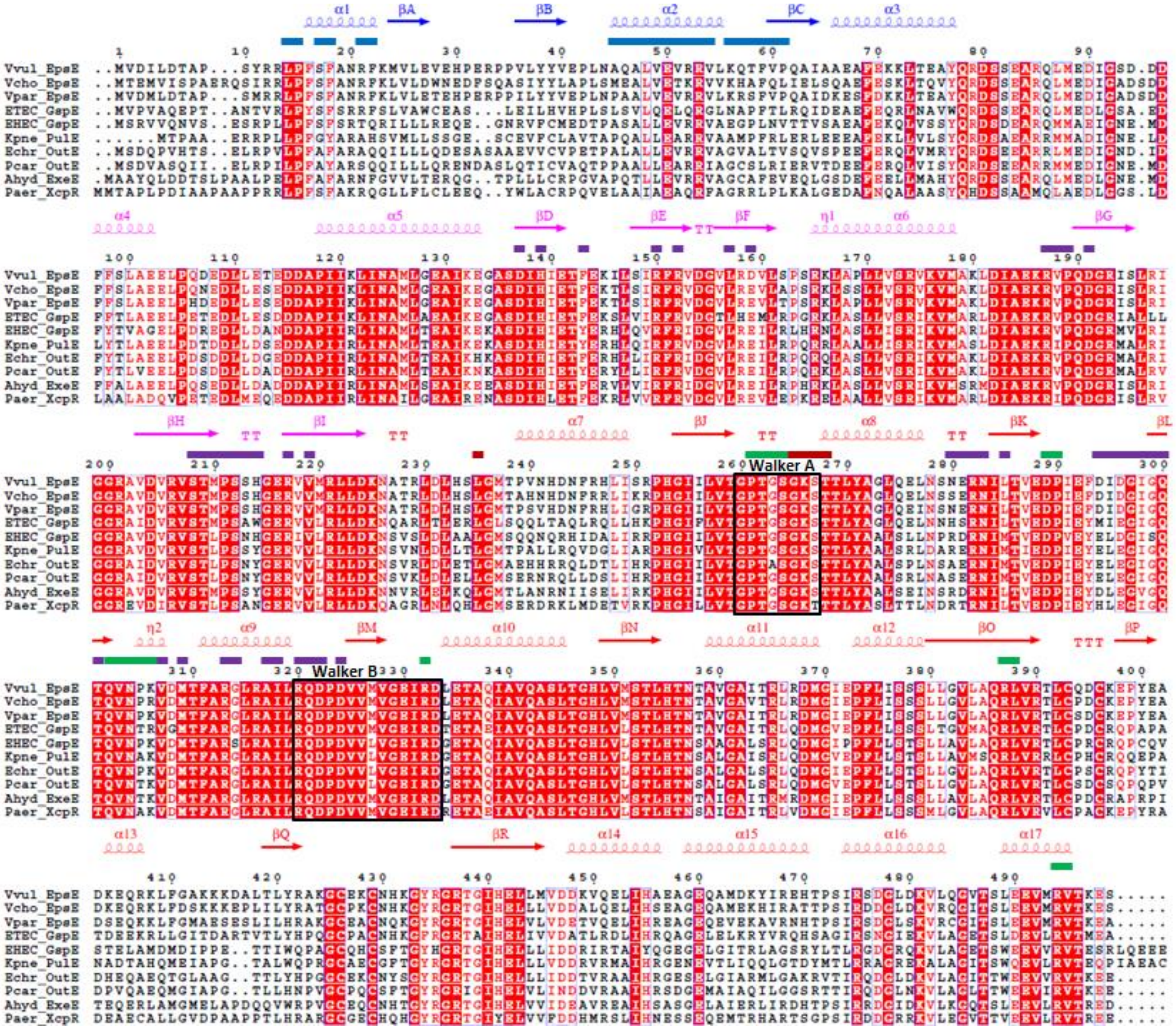
Supplement

Supplementary Figure S1. Domain structures of GspE and cytoplasmic domain of GspL and homologs



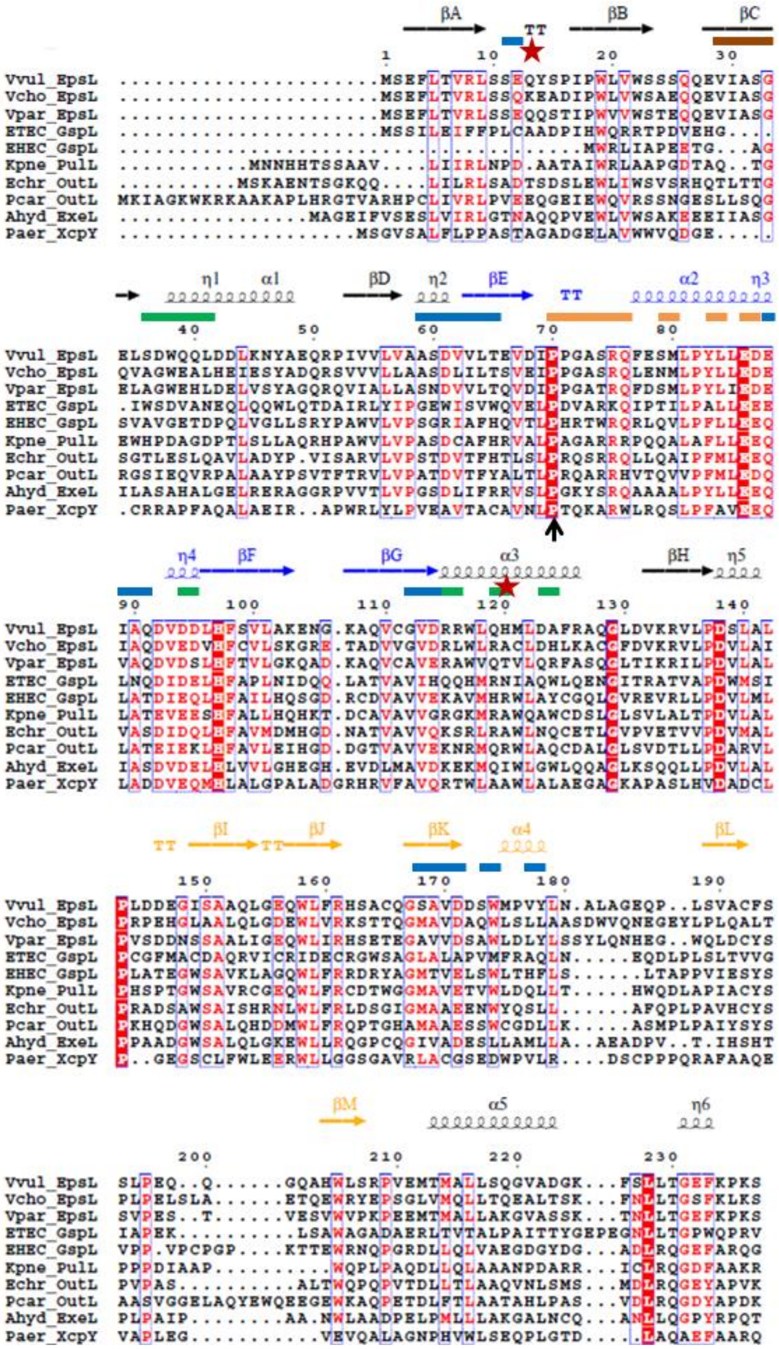
Domain bar diagrams of GspE and cyto-GspL homologs. For domains that are homologous in structure to GspE, the percentage sequence identity is given compared to *Vv* GspE. **(A)** GspE and homologous ATPases with electron microscopy reconstructions or crystal structures. *Vv* GspE: *Vibrio vulnificus* GspE. *Vc* GspE: *Vibrio cholerae* GspE. *Tt* PilF: *Thermus thermophilus* assembly ATPase PilF. *Pa* PilT: *Pseudomonas aeruginosa* retraction ATPase PilT. *Aa* PilT: *Aquifex aeolicus* retraction ATPase PilT. *Af* GspE2: one of three related ATPases in *Archaeoglobus fulgidus*; *Sa* FlaI: *Sulfolobus acidocaldarius* FlaI. **(B)** cyto-GspL and homologs with known crystal structures. The white bar in T4aPS *Tt* PilM indicates sub-domain 2B that is missing in T2SS cyto-GspL. *Vv* cyto-GspL: *Vibrio vulnificus* cytoplasmic domain of GspL. *Vc* cyto-GspL: *Vibrio cholerae* cytoplasmic domain of GspL. *Tt* PilM: *Thermus thermophilus* PilM. EPEC cyto-BfpC: enteropathogenic *Escherichia coli* cytoplasmic domain of BfpC.

Supplementary Figure S2. Family sequence alignment of GspE



Representative aligned sequences from a broad family of T2SS GspE. Shown are *Vibrio vulnificus, Vibrio cholerae, Vibrio parahaemolyticus,* enterotoxigenic *Escherichia coli,* enterohaemorrhagic *Escherichia coli, Klebsiella pneumonia, Erwinia chrysanthemi, Pectobacterium carotovorum, Aeromonas hydrophila, Pseudomonas aeruginosa* GspE homologs. The secondary structure elements of *Vibrio vulnificus* GspE are annotated at the top with N1E in blue, N2E in pink and CTE in red. The bar above the alignment indicates key residues in the interactions of N1E•cytoL (blue), cytoL•CTE (green), CTE•N2E' (purple), and AMPPNP•CTE (red). Walker A and Walker B motifs are shown in black boxes.

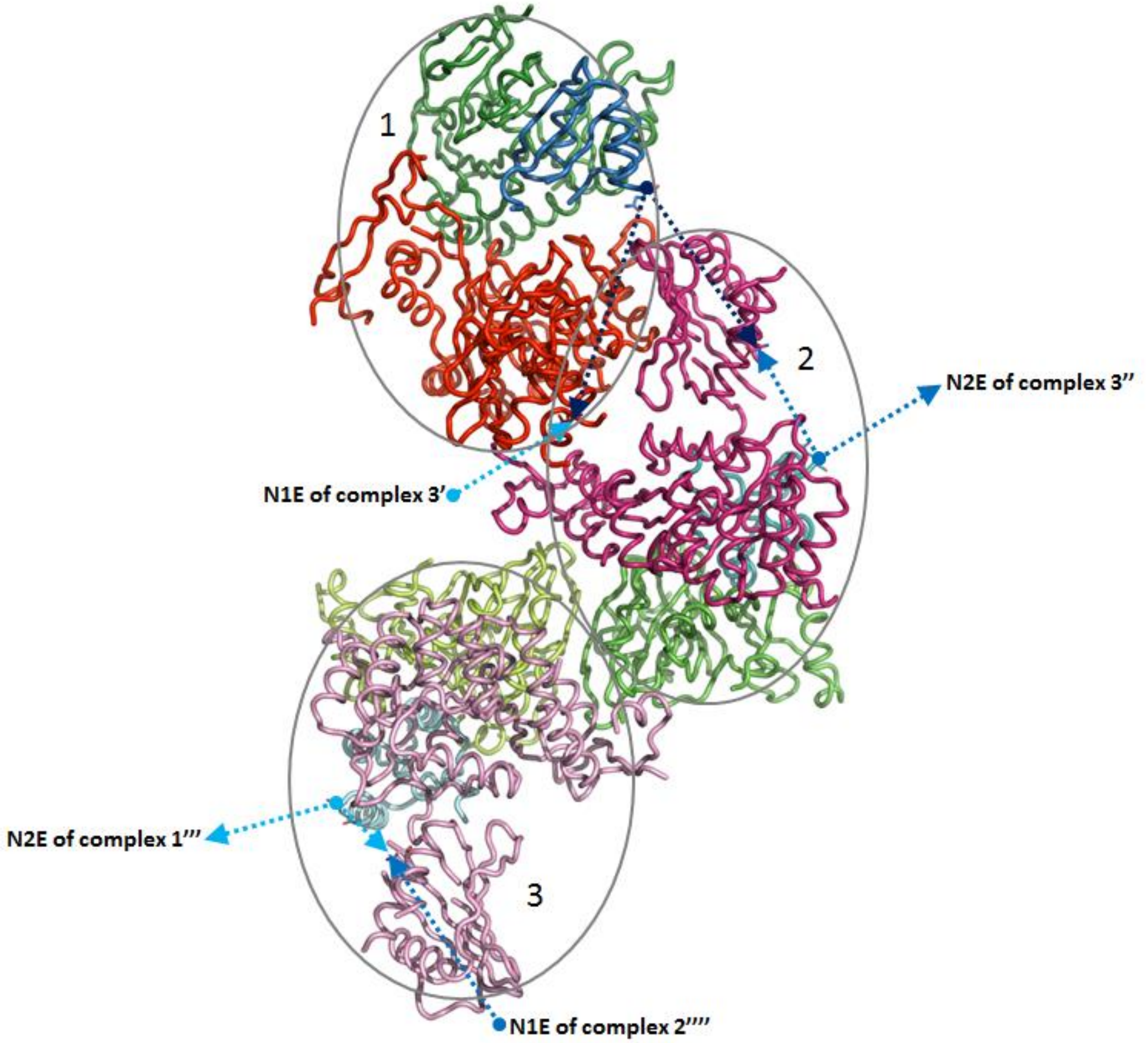
Supplementary Figure S3. Family sequence alignment of the cytoplasmic domain of GspL



```
             1         10        20        30
Vvul_EpsL  ......................MSEFLTVRLSSEQYSPIPWLVWSSSQQEVIASG
Vcho_EpsL  ......................MSEFLTVRLSSQKEADIPWLVWSAEQQEVIASG
Vpar_EpsL  ......................MSEFLTVRLSSEQQSTIPWVVWSTEQQEVIASG
ETEC_GspL  ......................MSSILEIFFPLCAADPIHWQRRTPDVEHG....
EHEC_GspL  .......................................MWRLIAPEETG...AG
Kpne_PulL  ...........MNNHHTSSAAV....LIIRLNPD.AATAIWRLAAPGDTAQ..TG
Echr_OutL  ..........MSKAENTSGKQQ....LILRLSADTSDSLEWLIWSVSRHQTLTTG
Pcar_OutL  MKIAGKWKRKAAKAPLHRGTVARHPCLIVRLPVEEQGEIEWQVRSSNGESLLSQG
Ahyd_ExeL  ................MAGEIFVSESLVIRLGTNAQQPVEWLVWSAKEEEIIASG
Paer_XcpY  ....................MSGVSALFLPPASTAGADGELAVWWVQDGE.....

             40        50        60        70        80
Vvul_EpsL  ELSDWQQLDDLKNYAEQRPIVVLVAASDVVLTEVDIPPGASRQFESMLPYLLEDE
Vcho_EpsL  QVAGWEALHEIESYADQRSVVVLLAASDLILTSVEIPPGASRQLENMLPYLLEDE
Vpar_EpsL  ELAGWEHLDELVSYAGQRQVIALLASNDVVLTQVDIPPGATRQFDSMLPYLIEDE
ETEC_GspL  .IWSDVANEQLQQWLQTDAIRLYIPGEWISVWQVELPDVARKQIPTILPALLEEE
EHEC_GspL  SVAVGETDPQLVGLLSRYPAWVLVPSGRIAFHQVTLPHRTWRQRLQVLPFLLEEQ
Kpne_PulL  EWHPDAGDPTLSLLAQRHPAWVLVPASDCAFHRVALPAGARRRPQQALAFLLEEQ
Echr_OutL  SGTLESLQAVLADYP.VISARVLVPSTDVTFHTLSLPRQSRRQLLQAIPFMLEEQ
Pcar_OutL  RGSIEQVRPALAAYPSVTFTRVLVPATDVTFYALTLPRQARRHVTQVVPFMLEDQ
Ahyd_ExeL  ILASAHALGELRERAGGRPVVTLVPGSDLIFRRVSLPGKYSRQAAAALPYLLEEQ
Paer_XcpY  .CRRAPFAQALAEIR..APWRLYLPVEAVTACAVNLPTQKARWLRQSLPFAVEEQ

             90        100       110       120       130       140
Vvul_EpsL  IAQDVDDLHFSVLAKENG.KAQVCGVDRRWLQHMLDAFRAQGLDVKRVLPDSLAL
Vcho_EpsL  IAQDVEDVHFCVLSKGRE.TADVVGVDRLWLRACLDHLKACGFDVKRVLPDVLAI
Vpar_EpsL  VAQDVDSLHFTVLGKQAD.KAQVCAVERAWVQTVLQRFASQGLTIKRILPDVLAL
ETEC_GspL  LNQDIDELHFAPLNIDQQ.LATVAVIHQQHMRNIAQWLQENGITRATVAPDWMSI
EHEC_GspL  LATDIEQLHFAILHQSGD.RCDVAVVEKAVMHRWLAYCGQLGVREVRLLPDVLML
Kpne_PulL  LATEVEESHFALLHQHKT.DCAVAVVGRGKMRAWQAWCDSLGLSVLALTPDVLAL
Echr_OutL  VASDIDQLHFAVMDMHGD.NATVAVVQKSRLRAWLNQCETLGVPVETVVPDVMAL
Pcar_OutL  LATEIEKLHFAVLEIHGD.DGTVAVVEKNRMQRWLAQCDALGLSVDTLLPDARVL
Ahyd_ExeL  IASDVDELHLVVLGHEGH.EVDLMAVDKEKMQIWLGWLQQACLKSQQLLPDVLAL
Paer_XcpY  LADDVEQMHLALGPALADGRHRVFAVQRTWLAAWLALAEGACKAPASLHVDADCL

                 150       160       170       180       190
Vvul_EpsL  PLDDEGISAAQLGEQWLFRHSACQGSAVDDSWMPVYLN.ALAGEQP..LSVACFS
Vcho_EpsL  PRPEHGLAALQLGDEWLVRKSTTQGMAVDAQWLSLLAASDWVQNEGEYLPLQALT
Vpar_EpsL  PVSDDNSSAALIGEQWLIRHSETEGAVVDSAWLDLYLSSYLQNHEG..WQLDCYS
ETEC_GspL  PCGFMACDAQRVICRIDECRGWSAGLALAPVMFRAQLN.....EQDLPLSLTVVG
EHEC_GspL  PLATEGWSAVKLAGQWLFRRDRYAGMTVELSWLTHFLS......LTAPPVIESYS
Kpne_PulL  PHSPTGWSAVRCGEQWLFRCDTWGGMAVETVWLDQLLT.....HWQDLAPIACYS
Echr_OutL  PRADSAWSAISHRNLWLFRLDSGIGMAAEENWYQSLL......AFQPLPAVHCYS
Pcar_OutL  PKHQDGWSALQHDDMWLFRQPTGHAMAAESSWCGDLLK.....ASMPLPAIYSYS
Ahyd_ExeL  PPAADGWSALQLGKEWLLRQGPCQGIVADESLLAMLLA..AEADPV..T.IHSHT
Paer_XcpY  P..GEGSCLFWLEERWLLGGSGAVRLACGSEDWPVLR....DSCPPPQRAFAAQE

                 200       210       220       230
Vvul_EpsL  SLPEQ..Q......GQAHWLSRPVEMTMALLSQGVADGK...FSLLTGEFKPKS
Vcho_EpsL  PLPELSLA......ETQEWRYEPSGLVMQLLTQEALTSK...FNLLTGSFKLKS
Vpar_EpsL  SVPES..T......VESVWVPKPEEMTMALLAKGVASSK...TNLLTGEFKPKS
ETEC_GspL  IAPEK..........LSAWAGADAERLTVTALPAITTYGEPEGNLLTGPWQPRV
EHEC_GspL  VPP.VPCPGP....KTTEWRNQPGRDLLQLVAEGDGYDG...ADLRQGEFARQG
Kpne_PulL  PPPDIAAP..........WQPLPAQDLLQLAAANPDARR...ICLRQGDFAAKR
Echr_OutL  PVPAS.........ALTWQPQPVTDLLTLAAQVNLSMS...MDLRQGEYAPVK
Pcar_OutL  AASVGGELAQYEWQEEGEWKAQPETDLFTLAATAHLPAS...VDLRQGDYAPDK
Ahyd_ExeL  PLPAIP........AA.NWLAADPELPMLLLAKGALNCQ...ANLLQGPYRPQT
Paer_XcpY  VAPLEG...........VEVQALAGNPHVWLSEQPLGTD.....LAQAEFAARQ
```

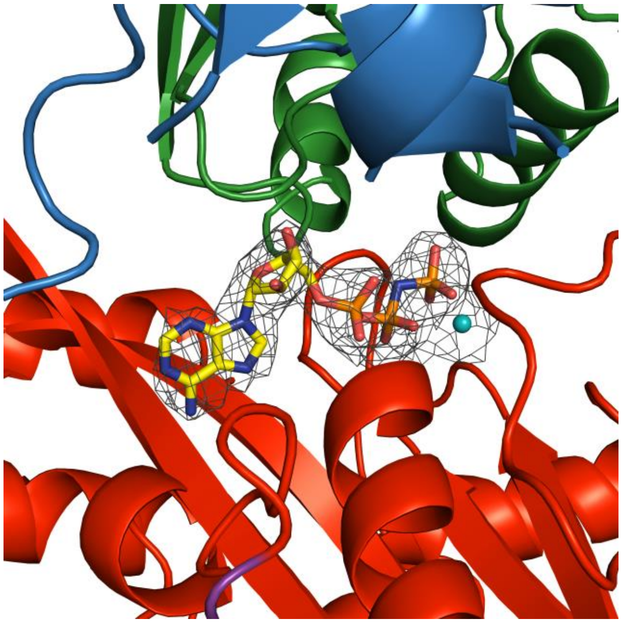
Aligned T2SS cyto-GspL sequences from representative species. Shown are *Vibrio vulnificus, Vibrio cholerae, Vibrio parahaemolyticus,* enterotoxigenic *Escherichia coli,* enterohaemorrhagic *Escherichia coli, Klebsiella pneumonia, Erwinia chrysanthemi, Pectobacterium carotovorum, Aeromonas hydrophila, Pseudomonas aeruginosa* cyto-GspL homologs. The secondary structure elements of *Vibrio vulnificus* cyto-GspL are annotated at the top, with subdomain I in black, II in blue and III in yellow. The bar above the alignment indicates key residues in the interactions of β-interface (brown), α-interface (orange), AMPPNP•CTE (red star), cytoL•CTE (green) and N1E•cytoL (blue). Black arrows indicate complete or highly conserved residues in the α-interface.

Supplementary Figure S4. The asymmetric unit of *V. vulnificus* GspE•cyto-GspL crystal



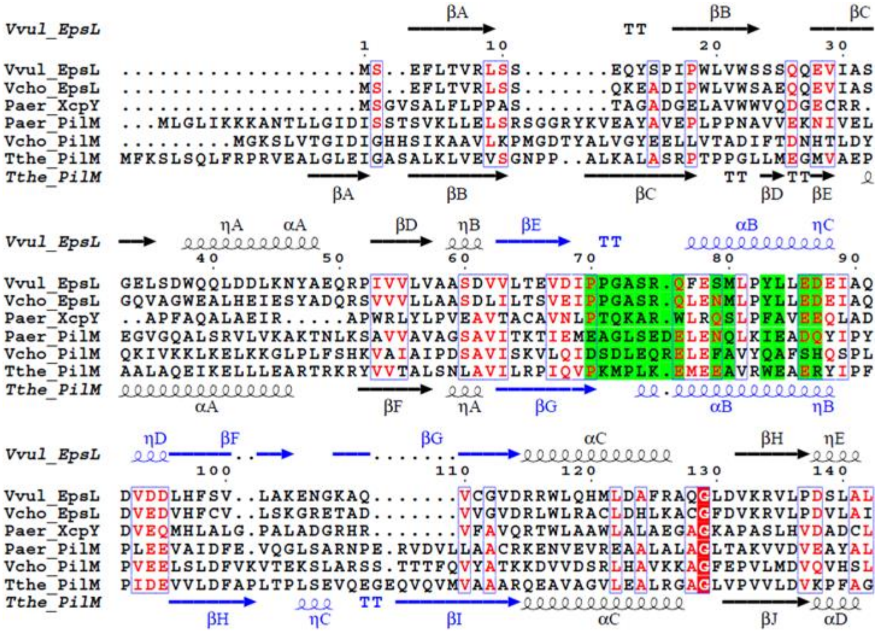
Three GspE•cyto-GspL complexes in the asymmetric unit. N1E of GspE in shades of blue, N2E-CTE of GspE in shades of red, and cyto-GspL in shades of green. Each grey ellipsoid indicates one GspE•cyto-GspL complex. Dashed-blue line indicates possible connections of N1E and N2E, with distances ranging from 33 to 41 Å. Primers denote complexes in neighboring asymmetric units (that are not shown).

Supplementary Figure S5. Electron density of nucleotide in the *V. vulnificus* GspE•cyto-GspL structure



A (2Fobs-Fcalc) difference electron density at the 1.5 sigma level. Proteins are shown in ribbon with CTE in red, N1E in blue, cyto-GspL in green. AMPPNP is shown in sticks with carbon in yellow, nitrogen in blue, oxygen in red, phosphorous in orange. Mg is shown in cyan sphere. The phases were calculated without including nucleotide coordinates.

Supplementary Figure S6. Sequence alignment of T2SS cyto-GspL homologs



```
Vvul_EpsL  ...................MS..EFLTVRLSS.......EQYSPIPWLVWSSSQQEVIAS
Vcho_EpsL  ...................MS..EFLTVRLSS.......QKEADIPWLVWSAEQQEVIAS
Paer_XcpY  ...................MSGVSALFLPPAS.......TAGADGELAVWWVQDGECRR.
Paer_PilM  ...MLGLIKKKANTLLGIDISSTSVKLLELSRSGGRYKVEAYAVEPLPPNAVVEKNIVEL
Vcho_PilM  .........MGKSLVTGIDIGHHSIKAAVLKPMGDTYALVGYEELLVTADIFTDNHTLDY
Tthe_PilM  MFKSLSQLFRPRVEALGLEIGASALKLVEVSGNPP..ALKALASRPTPPGLLMEGMVAEP

Vvul_EpsL  GELSDWQQLDDLKNYAEQRPIVVLVAASDVVLTEVDIPPGASR.QFESMLPYLLEDEIAQ
Vcho_EpsL  GQVAGWEALHEIESYADQRSVVVLLAASDLILTSVEIPPGASR.QLENMLPYLLEDEIAQ
Paer_XcpY  ..APFAQALAEIR.....APWRLYLPVEAVTACAVNLPTQKAR.WLRQSLPFAVEEQLAD
Paer_PilM  EGVGQALSRVLVKAKTNLKSAVVAVAGSAVITKTIEMEAGLSEDELENQLKIEADQYIPY
Vcho_PilM  QKIVKKLKELKKGLPLFSHKVAIAIPDSAVISKVLQIDSDLEQRELEFAVYQAFSHQSPL
Tthe_PilM  AALAQEIKELLLEARTRKRYVVTALSNLAVILRPIQVPKMPLK.EMEEAVRWEAERYIPF

Vvul_EpsL  DVDDLHFSV..LAKENGKAQ.......VCGVDRRWLQHMLDAFRAQGLDVKRVLPDSLAL
Vcho_EpsL  DVEDVHFCV..LSKGRETAD.......VVGVDRLWLRACLDHLKACGFDVKRVLPDVLAI
Paer_XcpY  DVEQMHLALG.PALADGRHR.......VFAVQRTWLAAWLALAEGAGKAPASLHVDADCL
Paer_PilM  PLEEVAIDFE.VQGLSARNPE.RVDVLLAACRKENVEVREAALALAGLTAKVVDVEAYAL
Vcho_PilM  PVEELSLDFVKVTEKSLARSS.TTTFQVYATKKDVVDSRLHAVKKAGFEPVLMDVQVHSL
Tthe_PilM  PIDEVVLDFAPLTPLSEVQEGEQVQVMVAAARQEAVAGVLEALRGAGLVPVVLDVKPFAG
```

Representative aligned sequences from T2SS cyto-GspL and T4aPS PilM subdomains I and II. Shown are sequences of T2SS cyto-GspL from *Vibrio vulnificus, Vibrio cholerae* and *Pseudomonas aeruginosa,* and T4aPS PilM from *Vibrio cholerae, Pseudomonas aeruginosa* and *Thermus thermophilus.* The secondary structure elements of *Vibrio vulnificus* cyto-GspL and *Thermus thermophilus* PilM are annotated at the top and bottom, respectively, with subdomain I in black and subdomain II in blue. The α-interface residues are highlighted in green (bottom).