# Non-cellulosomal Cohesin- and Dockerin-like Modules in the Three Domains of Life

Ayelet Peer[a], Steven P. Smith[b], Edward A. Bayer[c,]*, Raphael Lamed[a] and Ilya Borovok[a]

[a]Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Ramat Aviv 69978 Israel
[b]Department of Biochemistry, Queen's University Kingston Ontario Canada K7L 3N6
[c]Department of Biological Sciences, Weizmann Institute of Science, Rehovot 76100 Israel

*Corresponding author: Edward A. Bayer
Tel: (+972) -8-934-2373
Fax: (+972)-8-946-8256
Email: ed.bayer@weizmann.ac.il

## Supplementary Table S1.
## Compendium of cohesins and dockerins in the three domains of life.

In order to discover new putative cohesin/dockerin-containing proteins, we used sequences of **all** the classical cohesin and dockerin modules from *C. thermocellum*, *C. cellulovorans*, *C. cellulolyticum*, *B. cellulosolvens* and *Acetivibrio cellulolyticus* as well as cohesins and dockerins recently discovered in rumen bacteria, *Ruminococcus albus* and *R. flavefaciens* as BlastP queries for the main NCBI Blast server against all non-redundant protein sequences deposited in GenBank/EMBL/DDBJ databases. We also performed extensive searches using the TblastN algorithm through all publicly available microbial genome databases including those attached to the NCBI BLAST server for bacterial genomes (**http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?**), as well as several additional microbial genome databases – Microbial Genomics at the DOE Joint Genome Institute (**http://genome.jgi-psf.org/mic_home.html**), the Rumenomics database at TIGR/JCVI (**http://tigrblast.tigr.org/rumenomics/index.cgi**) and Bacterial Genomes at the Sanger Centre (**http://www.sanger.ac.uk/Projects/Microbes/**). Once a putative cohesin or dockerin-encoding gene product was identified, gene-walking techniques were employed to analyze and locate possible cellulosome-like gene clusters.

The date of the final search of the database, used for construction of this table, was **September 23, 2008.**

# ARCHAEA

## **Euryarchaeota** (No cohesins or dockerins are currently predicted in Crenarchaeota)

| Gene | Protein size (no. residues) | Putative cohesin | Putative dockerin | Annotation and comments[1] |
|---|---|---|---|---|

*Archaeoglobus fulgidus* **DSM 4304 / ATCC 49558 (Archaeoglobi; Archaeoglobales)**

| | | | | |
|---|---|---|---|---|
| **AF2375 (operon)** | 506 | 279..418 | 436..488 | **No obvious signal peptide (SP). TMD (69..91) + Cohesin + Dockerin**<br>**Cohesin AF2375**<br>TTIIAGSAEAPQGSDIQVPVKIENADKVGSINLILSYPNVLEVEDV<br>LQGSLTQNSLFDYNVEGNQIKVGIADSNGISGDGSLFYVKFRVTGN<br>EKAEQAENVKGKLRGLGQQLSEITLRNSHALTLQGIEIYDIDGNSV<br>KV<br>**Dockerin AF2375**[1]<br>G**D**V**N**G**D**GEI**N**SL**D**ALLALQMSIG  KVEPNPV<br>A**D**M**D**G**D**GKV**L**AK**D**ATEIMKMATD |
| **AF2376 (operon)** | 190 | 33..156 | - | **SP, Hypothetical protein, +Cohesin, a 'strong' TMD (167..185)**<br>**Cohesin AF2376**<br>MVVKIPDTSGAVGGTVEVPVEVENAQNLGSMDIVIVYDPTILKVQN<br>VGKGELNKGLLSSNTGEGMVAISLADSKGINGKGSVAVITFQVLKA<br>GSTDLTIQSVKAYDVNTHVDIPVKADNGKFEA |

[1] In dockerin sequences, calcium-coordinating amino acid residues in positions 1, 3, 5, 9 and 12 are highlighted (those fitting the consensus are in blue, atypical residues – in gray; an asterisks follow the C-terminal residue in some proteins.

***Haloarcula marismortui* / *Halobacterium marismortui* ATCC 43049 (Halobacteria; Halobacteriales)**

| | | | | |
|---|---|---|---|---|
| *phoD* or rrnAC0273 (chromosome I) | 741 | - | 688..741 | **Alkaline phosphatase D**, EC 3.1.3.1; a putative dockerin is similar to that in Mhun_3046, *Methanospirillum hungatei*,<br>**Putative dockerin:**<br>E**D**V**N**G**N**GRV**D**YD**D**IQLLFDNFDD DSVALNKTA<br>Y**D**F**N**E**N**GKL**E**YD**D**IVTLYSEVN***** |
| rrnAC0766 (chromosome I) | 892 | - | 839..892 | Putative **cell surface protein,**<br>**Putative dockerin:**<br>E**D**L**N**G**N**GRL**D**YE**D**VQVLFSNMDS DSVQLNTGA<br>Y**D**F**N**E**N**GKL**D**FA**D**VTALYEEVN***** |
| **rrnAC2380; operon?** (chromosome I) | 615 | - | 559-613 | putative **serine/threonine phosphatase** (PF00149; Metallophos)<br>**Putative dockerin:**<br>G**D**A**D**G**D**GDV**D**TD**D**VDRIQRSVAG EDSDIDRDA<br>A**D**I**D**G**D**GDV**D**IG**D**AVSARNLAGD |
| **rrnAC2381; operon?** (chromosome I) | 442 | 42-175 | - | **Hypothetical protein; Cohesin domain predicted in both, Pfam (PF00963) and InterPro (IPR002102).**<br>**Putative cohesin:**<br>PEASGDALSVAPGETVTVQVWANATAVRGYQTNVTFDPTIAEIDSV<br>AGSDDFDSPVTNVNNEAGWVAFNQLRSRETSDPVLAEITLTVPADA<br>AGSTQLAFIEADTKLSNSDGETVAPAAFNSIELRVNNGSTTP |
| rrnB0167 (chromosome II) | 1562 | - | 126..180 | **Cbp**, **calcium-binding protein**-like: approx. 8 TSP_3 like domains;<br>**PF00404; dockerin_1**<br>E**D**V**N**G**D**GAV**N**IV**D**VDALSRHLES TAAGANWSA<br>Y**D**Y**T**G**D**NRT**D**VG**D**IQWLFVATRS |

***Halobacterium salinarium* NRC-1 / *Halobacterium halobium* NRC-1 (Halobacteria; Halobacteriales)**

| | | | | |
|---|---|---|---|---|
| VNG1951G | 585 | - | 532..585 | **Subtilisin homolog or Sub;**<br>**PF00404; Dockerin_1**<br>E**D**V**N**G**D**GVV**D**MA**D**VELLYQLALR NAVPESTTA<br>F**D**V**T**G**D**GKF**D**MH**D**VQALTRQVE***** |

***Haloquadratum walsbyi*** **DSM 16790 (Halobacteria; Halobacteriales)**

| | | | | |
|---|---|---|---|---|
| **HQ1205A, operon?** | **1626** | 1285-1420 | 1571-1626 | **Probable cell surface adhesin,**<br>**Putative cohesin:**<br>VEGVELQSGNTTTLATNFTLREDTTADLTATVEFSNETAIENTDVT<br>FSGSGGVPSSNDASLPVTQTTNSSGIVQFEIPASDTNDAYTITADS<br>ANFSNASTGVSPAPGDSVSRDFSLEEAQAQEPSVSVNFTDQTV<br>**Putative dockerin:**<br>E**D**V**N**G**D**NAT**N**FD**D**VVALAFVLTE  KDDLTQQQRDA<br>F**D**I**N**D**D**GVV**N**FD**D**VVRLAFPGQ* |
| **HQ1208A, operon?** | **619** | - | 558-619 | **Subtilisin-like serine protease Sub,**<br>**Putative dockerin:**<br>E**D**V**N**G**D**NGY**T**FD**D**VIALAFADTE  RLTNQQRDA<br>L**D**F**D**H**D**GNI**D**FD**D**VVELAFER* |
| **HQ1210A, operon?** | **125** | | 71-125 | **Hypothetical protein,**<br>**Putative dockerin:**<br>E**D**V**N**G**D**AES**T**FD**D**VIALAFVISQ  TGQLTEQQRDA<br>L**N**I**D**G**E**DDV**D**FD**N**VIALTFNI* |
| **HQ3467A, operon?** | **2079** | - | 1952-2006;<br>2027-2079 | **Probable cell surface glycoprotein Hmu3**<br>**Putative dockerin:**<br>E**D**I**N**G**D**GEF**N**FD**D**VVALAFNTGT  DPIQQNSDL<br>F**D**I**N**N**D**GVV**N**FD**D**VVALAFESPS<br>**Putative dockerin:**<br>E**D**I**N**G**D**GEL**N**FD**D**VVALAFNIES  DAVQQNTDQ<br>F**D**F**D**N**D**GNV**D**FD**D**VVELAFET* |
| **HQ3469A, operon?** | **506** | - | 443-506 | **Probable cell surface glycoprotein,**<br>**Putative dockerin:**<br>E**D**I**N**G**D**GKA**N**FD**D**VVTFAFKITT  NAVQENADL<br>F**D**F**N**D**D**NDV**N**FD**D**VIELAFTLPT |

Bolhuis *et al*., 2006.

*Methanococcoides burtonii* **DSM 6242 (Methanomicrobia; Methanosarcinales)**

| | | | | |
|---|---|---|---|---|
| **Mbur_0125** | **418** | 35-167 | - | **Hypothetical protein precursor;**<br>**Cohesin domain, Pfam (PF00963) and InterPro (IPR002102):**<br>VSISPPDQLISIGSNVTVIIYVDPDVPVSGAQFDLDFDGDILSVIS<br>ISEGDIFTNGVSTLFSGGTIDNLEGNIMNVYGALFGKNEVIKSGTF<br>ATIVFETKSSGSSSLQLSNVVVSNSIGAAVPIAVINGIVSINGD |
| **Mbur_0728*,**<br>**operon with**<br>**Mbur_0729?**<br>*** see its ortholog,**<br>**MM_2098** | **1126** | 846-979 | 1072-1123 | **Putative Ig precursor,**<br>**Cohesin, Pfam (PF00963) and InterPro (IPR002102):**<br>VVMSSSAQIVSPGQPFSIGISIDPSEPVSGAQLDLLFDEQLVSAEG<br>TIEGDLFNLNGASTLFNSGTIGDSGVITDIY**C**SIIGAAPVSSKGTM<br>ATIAFTAGNTAGVAEFSLSDVVISNARSEGTPYTVTDTMVVI<br>**Putative dockerin:**<br>W**D**I**N**R**D**CAV**N**IL**D**ITLVSQKIGT   DGAGTD<br>Q**D**V**N**Q**D**GVV**N**IQ**D**LTLVAQHFGE |
| **Mbur_0729,**<br>**operon with**<br>**Mbur_0728?** | **226** | 53-178 | - | **SP, annotated as "Cellulosome anchoring protein, precursor";**<br>**Cohesin, Pfam (PF00963) and InterPro (IPR002102):**<br>FTVSSGEVLNVAPGETFNVDVFVSPDVDIAGMQFDLLFDSSKFQID<br>SVTEGNLFKQSGMGTFFVAASVTPGLLDNTYG**C**ILGNAAVLTPANF<br>ATITITANEHASGRSAFILKDVIVSDPKGNAVNV |

*Methanococcus aeolicus* **Nankai-3 / ATCC BAA-1280 (Methanococci; Methanococcales), isolated from deep marine sediment from the Nankai Trough off the coast of Japan**

| | | | | |
|---|---|---|---|---|
| **Maeo_0111** | **421** | - | 24-73 | **Periplasmic binding protein precursor,**<br>**Putative N-terminal dockerin:**<br>G**D**V**N**E**D**GSI**N**II**D**VVYLFKNRNL   DYEV<br>G**D**I**N**C**D**NSI**N**II**D**VVYLFKHYDQ |
| **Maeo_0115** | **703** | 25-155 | | **SP, PKD domain containing protein precursor,**<br>**Pfam PF00963; InterPro IPR002102; Cohesin:**<br>NITVELTPNNTNTIVGDTINLDLFVKNIPNDTKCSGFETTLSYDSN<br>LLNLTDIKLSDISNNADLKEADLSTGSIKIAWFSNEPSGNITIASL<br>SFKTLNEATSPPATISLSSKTAVSNINGVGYNNITIINSNITIN |

| | | | | |
|---|---|---|---|---|
| **Maeo_0117** | **182** | 22-150 | | **SP, hypothetical protein,**<br>**Pfam PF00963; InterPro IPR002102; Cohesin:**<br>ASNVEVNCPSEVNVGDTFDIVITIKNNESLSGFECQVNVPDNLKII<br>NVLENDDIRKKASNYYKNDFSDRNGIISFMLINEPMQSDFYLATIK<br>VKVLKDSSNTTIHITPKAADKDANEIKMNNINLKLNV |
| **Maeo_0118** | **454** | - | 36-85 | **SP, ABC-type Fe3+-hydroxamate transport system periplasmic component-like protein precursor,**<br>**Putative N-terminal dockerin:**<br>G**D**V**N**A**D**GSI**N**II**D**VVYLFKNRNL  DYED<br>G**D**I**N**A**D**NSI**N**II**D**VVYLFKHYDQ |
| **Maeo_0119** | **976** | 23-160 | - | **SP, APHP domain protein, Pfam PF07705; CARDB (x2); PF04473; DUF553;**<br>**Pfam PF00963; InterPro IPR002102; Cohesin:**<br>NVSVELTPSNVNAIVGDTISLDLVVKNIPNDTKCSGFDTTLSYDSN<br>LLNLTDIKLSDISNNANLKRANLSTGFISLSWSSDEPSGNITIASL<br>SFKALNEATSPTTISLTQTIVSDIEPKGYKITLNNSNIIITKPKAD |

***Methanococcus vannielii*** **SB / ATCC 35089 / DSM 1224** (+ two other strains) **(Methanococci; Methanococcales)**

| | | | | |
|---|---|---|---|---|
| **Mevan_0131** | **411** | - | 14-63 | **Periplasmic binding protein, precursor**<br>**Putative N-terminal dockerin:**<br>G**D**V**N**G**D**GNI**N**IA**D**VVYLFKNRNV  PIDV<br>G**D**L**N**C**D**GNV**N**VA**D**VVYLFRNYDK |
| **Mevan_0132** | **695** | 22-145 | | **SP, Putative uncharacterized protein, PKD domain,**<br>**Pfam PF00963; InterPro IPR002102; Cohesin:**<br>SVVLNSSSDSFFIGESFDLELNLLNIPSDKKCGGFETRISYDKTQL<br>QLSSITLSKDLNSDIGDVSLSNGRISIVWFSTPPTGDINIAKIKFN<br>VLKEGNSTIELLGTTVSDSNGFTYNNMTLSNLKVKLSPLGYTVT |
| **Mevan_0135** | **754** | 18-145 | | **SP, Putative uncharacterized protein,**<br>**Pfam PF00963; InterPro IPR002102; Cohesin:**<br>IELSPNVINIQKGNNFTVDVLLRDIPEIAGYDSKLNYDKTILKLNS<br>ISLSTDVNSASLKDTNLNNDGSVSLLWMGNSVTGNLTVATFTFEAI<br>NEGDTKISLKNAALSDIEGQGINNFDTLDSEVLVFE |

***Methanococcus voltae*** **A3 (Methanococci; Methanococcales)**

| | | | | |
|---|---|---|---|---|
| **MvolDRAFT_0511, operon?** | **453** | - | 26-82 | **Putative iron transport system substrate-binding protein, precursor,**<br>**Putative dockerin:**<br>G**N**Y**D**Y**R**LGD**V**NC**D**GKISVSDVVY LFNNRNMDIED<br>G**D**V**N**A**D**NKI**S**VS**D**VVYLFNYYDK |
| **MvolDRAFT_0512, operon?** | **1103** | 36-170 | - | **SP, APHP domain protein precursor,**<br>**Putative cohesin:**<br>TNDNIVVSLAPQTQQVNNGDQFVLELWVNNSNITISAFESNFVYDS<br>SNVNLTAIQLSGIANTAGLKTVDVSDNSISLAWMGGGITESNVNIA<br>NLTFETFNVSSNEIRLRNLDITNQGGISTMPTDEHILDASVEV |
| **MvolDRAFT_0578** | **789** | 49-186 | - | **Hypothetical protein.**<br>**Pfam PF00963; Cohesin; InterPro IPR002102; Cohesin:**<br>NYSMTLSPTVKSAGVGDSFELTLWGNTPKVSAIQSKFNYDTAMLEL<br>TDIELGTIADTASSKRAEVSSSLITMLWFDPSTAPEGYYKIATLSF<br>KVLKGGNTSIVPRNYEFSDNTGVSITPVINKANISVSEGMKAELII |
| **MvolDRAFT_0734** | **451** | - | 25-81 | **Putative iron transport system substrate-binding protein, precursor,**<br>**Putative dockerin:**<br>G**N**Y**D**Y**R**LGD**I**NE**D**GKISVSDVVY LFNNRDLDIEDADVN<br>A**D**V**N**A**D**NKI**S**VS**D**VVYLFNYYDE |

***Methanoculleus marisnigri*** **JR1 / ATCC 35101 / DSM 1498 (Methanomicrobia; Methanomicrobiales)**

| | | | | |
|---|---|---|---|---|
| **Memar_0673** | **857** | - | 800-852 | **2 CARDB (PF07705), Aminotrans_II_pyridoxalP_BS (IPR001917),**<br>**Putative dockerin:**<br>G**D**L**N**G**D**GNV**D**WA**D**VTIAAGMAQK TAPSDPA<br>A**D**V**N**G**D**GTV**D**WK**D**VALLADFFFG |
| **Memar_1713** | **421** | 191-320 | - | **Beta-lactamase-like**<br>**Putative cohesin:**<br>SYLFTGDAGTPAEESMAEAGLDLDADVLKVGHHASRYASSAEFLSS |

| | | | | |
|---|---|---|---|---|
| | | | | VSPAISVIEVGEGNDYGHPHEEAVERLEATGSRIYRTDLDGTVIVA<br>TDGTALTVAAGGAPAATVTAGATTAAATATPTATATPA |
| **Memar_2164** | **512** | 41-170 | - | **Peptidase M20:Peptidase M28, Precursor,**<br>**Putative cohesin:**<br>NSLTEIIGVLTDYQRVPGFDDGPAADYIAGRLEEDGYDVQQEVFAV<br>ETDAGPATTQNVIGIKKGSGPGIVVVCAHYDVYGPDCPGADDNAAG<br>VAVMLEVARALRTESLDRSVYFIAFSGEEVGLQGSADW |

***Methanosaeta thermophila*** **PT /** ***Methanothrix thermophila*** **PT[a] (Methanomicrobia; Methanosarcinales)**

| | | | | |
|---|---|---|---|---|
| **Mthe_0972** | **429** | - | 32-84 | **Periplasmic binding protein precursor**;<br>**N-terminal dockerin:**<br>G**N**A**N**M**D**DRI**D**EN**D**LEFIRMVIQG  EKSATEL<br>C**D**A**D**H**N**GVV**D**AS**D**ITQVERIING |
| **Mthe_1591** | **377** | - | 27-79 | **Periplasmic binding protein precursor**;<br>**N-terminal dockerin:**<br>S**D**A**N**C**D**GKI**D**FE**D**IDQINALIKD  EAKFLIV<br>V**D**S**V**G**R**NVT**I**KL**P**INKLIALGSY |
| **Mthe_1596; operon?** | **339** | - | 33-85 | **Periplasmic binding protein precursor**;<br>**N-terminal dockerin:**<br>G**N**A**N**M**D**DAI**N**EE**D**VAYIQGILNN  EDKVTEF<br>A**D**A**N**Y**D**GKI**D**SN**D**IDQIEAIIHN |
| **Mthe_1597; operon?** | **409** | - | 28-80 | **Periplasmic binding protein precursor**;<br>**N-terminal dockerin:**<br>G**N**A**N**M**D**DTI**D**ES**D**IAYTRGIIDG  KEKNTEF<br>A**D**A**N**Y**D**GKV**D**SR**D**IEQIQHIIQG |

[a] **Cellulose-fermenting methanogenic archaeon**.

***Methanosarcina acetivorans*** **C2A / ATCC 35395 (Methanomicrobia; Methanosarcinales)**

| | | | | |
|---|---|---|---|---|
| **MA_0850** | **2275** | - | 2223..2275 | **Cell surface protein: 5 FG-GAP (PF01839), 1 NosD (PF05048), 10 PKD (PF00801), 5 PQQ (PF01011),** |

| | | | | |
|---|---|---|---|---|
| | | | | **Putative dockerin:**<br>`E`**D**`L`**T**`G`**N**`GEF`**S**`FV`**D**`IVAYFHNMDW   IEENMPVEY`<br>`F`**D**`F`**N**`G`**N**`GRI`**D**`FD`**D**`VVDMFAMI`* |
| **MA_0851** | **2566** | - | 2505..2566 | **Cell surface protein: 1 Big_2 (PF02368), 1 CARDB (PF07705), 9 PKD (PF00801), 7 SBBP (PF06739).**<br>**Putative dockerin:**<br>`E`**D**`L`**T**`G`**N**`GEF`**S**`FV`**D**`IVAYFHNMDW   IEENMPVEY`<br>`F`**D**`F`**N**`G`**N**`GRI`**D**`FD`**D**`VVDMFAMI`* |
| **MA_1510** | **2115** | - | 2063..2115 | **Hypothetical protein: 7 LVIVD (PF08309), 2 LGFP (PF08310), 1 PKD (PF00801).**<br>**Putative dockerin:**<br>`E`**D**`I`**N**`G`**N**`GRL`**G**`FN`**D**`LVAYFANLDW   IEENVPLEF`<br>`F`**D**`Y`**N**`K`**N**`GLI`**D**`FD`**D**`VVTLFDML`* |
| **MA_1513** | **584** | - | 532..584 | **Peptidase_C1** (**PF00112**); strong ortholog of **Mhun_0648**<br>`E`**D**`V`**N**`G`**N**`GIL`**D**`FD`**D**`VVAYYDNMGW   IEENALVAF`<br>`F`**D**`F`**S**`N`**N**`GLI`**D**`FD`**D**`VVKLYDRL`* |
| **MA_1753** | **1149** | - | 1088..1149 | **Cell surface protein: 1 CARDB (PF07705), 1 NosD (PF05048), 2 PKD (PF00801).**<br>**Putative dockerin:**<br>`E`**D**`L`**T**`G`**N**`GEF`**S**`FV`**D**`IVAYFHNMDW   IEENMPVEY`<br>`F`**D**`F`**N**`G`**N**`GRI`**D**`FD`**D**`VVDMFAMI`* |
| **MA_1756** | **1188** | - | 1136-1188 | **Cell surface protein: 1 CARDB (PF07705), 1 NosD (PF05048), 3 PKD (PF00801).**<br>**Putative dockerin:**<br>`E`**D**`L`**T**`G`**N**`GEF`**S**`FV`**D**`IVAYFHNMDW   IEENMPVEY`<br>`F`**D**`F`**N**`G`**N**`GRI`**D**`FD`**D**`VVDMFAMI`* |
| **MA_1761** | **1075** | - | 1016..1075 | **Surface antigen gene**: **1 CARDB (PF07705), 5 PKD (PF00801).**<br>**Putative dockerin:**<br>`E`**D**`L`**T**`G`**N**`GEF`**S**`FV`**D**`VVAYFHNMDW   IEANMPVEY`<br>`F`**D**`F`**N**`G`**N**`GRI`**D**`FD`**D**`VVDMFAMI`* |
| **MA_1762** | **3988** | - | 3927..3988 | **Cell surface protein: 2 Big_2 (PF02368), CARDB (PF07705), 18 PKD (PF00801), 6 PQQ (PF01011).** |

| | | | | |
|---|---|---|---|---|
| | | | | **Putative dockerin:**<br>E**D**L**T**G**N**GEF**S**FV**D**IVAYFHNMDW   IEENMPVEY<br>F**D**F**N**G**N**GRI**D**FD**D**VVDMFAMI* |
| **MA_1766** | **1747** | - | 1686..1747 | **Cell surface protein: CARDB (PF07705), 4 PKD (PF00801.**<br>E**D**L**T**G**N**GEF**S**FV**D**VVAYFHNMDW   IEENMPVEY<br>F**D**F**N**G**N**GRI**D**FD**D**VVDMFAMI* |
| **MA_1850, operon?** | | - | 43-95 | **Iron(III) ABC transporter, solute-binding protein,**<br>**Putative dockerin:**<br>G**N**A**N**E**D**GAI**D**MK**D**VEYTESIILG   SGNRTQF<br>A**D**A**T**G**D**N**S**I**D**ML**D**VTKIELIKLG |
| **MA_1851, operon?** | **247** | 79-207 | - | **Hypothetical protein [precursor?],**<br>**Putative cohesin:**<br>ATVGLSTPELVKESTVYATVNVENAENLDAGQFDLEFNPAVLKVVG<br>IENGSIGKTEIPVQWRSIDSGTVRVIFNLEGVTGVSGSGQLASVGF<br>EVIGDGESELFISDGLLGDTEAKSIDTDWGVSGASSV |
| **MA_2793** | **1857** | - | 1796..1857 | **Cell surface protein: 1 CARDB (PF07705), 5 PKD (PF00801).**<br>**Putative dockerin:**<br>E**D**L**T**G**N**GEF**S**FV**D**IVAYFHNMDW   IEENMPVEY<br>F**D**F**N**G**N**GRI**D**FD**D**VVDMFGMI* |
| **MA_3698** | **2016** | - | 1964-2016 | **Cell surface protein: 9 PKD (PF00801), 5 PQQ (PF01011), 3 Reg_prop (PF07494).**<br>**Putative dockerin:**<br>E**D**L**T**G**N**GEF**S**FV**D**IVAYFHNMDW   IEENMPVEY<br>F**D**F**N**G**N**GRI**D**FD**D**VVDMFAMI* |
| **MA_3699** | **2523** | - | 2462..2523 | **Cell surface protein: 1 CARDB (PF07705), 3 FG-GAP (PF01839), 8 PKD (PF00801), 5 PQQ (PF01011).**<br>**Putative dockerin:**<br>E**D**L**T**G**N**GEF**S**FV**D**IVAYFHNMDW   IEENMPVEY<br>F**D**F**N**G**N**GRI**D**FD**D**VVDMFAMI* |
| **MA_3700** | **1881** | - | 1829..1881 | **Cell surface protein: 3 CARDB (PF07705), 4 FG-GAP (PF01839), 6 PKD (PF00801),**<br>**Putative dockerin:** |

| | | | | |
|---|---|---|---|---|
| | | | | EDLTGNGEFSFVDIVAYFHNMDW IEENMPVEY<br>FDFNGNGRIDFDDVVDMFAMI* |
| **MA_4285** | **1743** | - | 1691-1743 | **Cell surface protein: 5 PKD (PF00801), 2 PQQ (PF01011).**<br>**Putative dockerin:**<br>EDLTGNGEFSFVDIVAYFHNMDW IEENMPVEY<br>FDFNGNGRIDFDDVVDMFAMI* |
| **MA_4287** | **2165** | - | 2113-2165 | **Cell surface protein: 1 Reg_prop (PF07494), 6 PKD (PF00801), 4 PQQ (PF01011).**<br>**Putative dockerin:**<br>EDLTGNGEFSFVDIVAYFHNMDW IEENMPVEY<br>FDFNGNGRIDFDDVVDMFAMI* |
| **MA_4289** | **1734** | - | 1682-1734 | **Cell surface protein: 3 PKD (PF00801).**<br>**Putative dockerin:**<br>EDLTGKGEFSFVDIVAYFHNMDW IEANMPVEY<br>FDFNGNGRIDFDDVVDMFAMI* |
| **MA_4291** | **1898** | 275-410 | 1837..1898 | **Cell surface protein: 2 NosD (PF05048), 3 PKD (PF00801), 2 PQQ (PF01011).**<br>**Putative cohesin (similar to 'eukaryotic' cohesins):**<br>TITESPSELFFEPSDTVMAGPGDTAEIELLADELPEGFSTYVLNIS<br>VTDPEIARISGVSFPAWADTTSSTEIPAASVEIKASDLNQQVEAGA<br>SNVTLATIGLESLGYGTSNLTVSVESFKDDSENPVTLRTMEAGL<br><br>**Putative dockerin:**<br>EDLTGNGEFSFVDIVAYFHNMDW IEENMPVEY<br>FDFNGNGRIDFDDVVDMFGMI* |
| **MA_4292** | **1995** | - | 1943-1995 | **Cell surface protein: 6 PKD (PF00801).**<br>**Putative dockerin:**<br>EDLTGNGEFSFVDIVAYFHNMDW IEENMPVEY<br>FDFNGNGRIDFDDVVDMFAMI* |
| **MA_4297** | **1817** | - | 1756..1817 | **Cell surface protein: 6 PKD (PF00801, 2 PQQ (PF01011).**<br>**Putative dockerin:**<br>EDLTGNGEFSFVDVVAYFHNMDW IEENMPVEY |

| | | | | |
|---|---|---|---|---|
| | | | | FDFNGNGRIDFDDVVDMFAMI* |
| **MA_4305** | **1948** | - | 1896-1948 | **Cell surface protein: 2 NosD (PF05048), 4 PKD (PF00801), 2 PQQ (PF01011).**<br>**Putative dockerin:**<br>EDLTGNGEFSFVDIVAYFHNMDW IEENMPVEY<br>FDFNGNGRIDFDDVVDMFAMI* |
| **MA_4307** | **76!** | - | 15..76 | **EF hand family protein:**<br>**Putative dockerin:**<br>EDLTGNGEFSFVDIVAYFHNMDW IEENMPVEY<br>FDFNGNGRIDFDDVVRMFAMI* |
| **MA_4309** | **1262** | - | 1210-1262 | **Cell surface protein: 1 NosD (PF05048), 4 PKD (PF00801).**<br>**Putative dockerin:**<br>EDLTGNGEFSFVDIVAYFHNMDW IEENMPVEY<br>FDFNGNERIDFDDVVRMFAMI* |
| **MA_4312** | **2029** | - | 1977-2029 | **Cell surface protein: 2 NosD (PF05048), 5 PKD (PF00801), 2 PQQ (PF01011).**<br>**Putative dockerin:**<br>EDLTGNGEFSFVDIVAYFHNMDW IEENMPVEY<br>FDFNGNGRIDFDDVVDMFAMI* |
| **MA_4315** | **1632** | - | 1571..1632 | **Cell surface protein: 1 NosD (PF05048), 8 PKD (PF00801), 2 PQQ (PF01011).**<br>**Putative dockerin:**<br>EDLTGNGEFSFVDIVAYFHNMDW IEENMPVEY<br>FDFNGNGRIDFDDVVRMFGMI* |
| **MA_4588** | **2951** | - | 2899-2951 | **Cell surface protein: 2 Big_2 (PF02368), 2 CARDB (PF07705), 12 PKD (PF00801).**<br>**Putative dockerin:**<br>EDLTGNGEFSFVDIVAYFHNMDW IEENMPVEY<br>FDFNGNGRIDFDDVVDMFGMI* |
| **MA_4589** | **1357** | - | 1305-1357 | **Cell surface protein: 1 CARDB (PF07705), 5 PKD (PF00801), 5 Reg_prop (PF07494).**<br>**Putative dockerin:** |

| | | | | |
|---|---|---|---|---|
| | | | | E**D**L**T**G**N**GEF**S**FV**D**IVAYFHNMDW IEENMPVEY<br>F**D**F**N**G**N**GRI**D**FD**D**VVDMFAMI***** |

### *Methanosarcina mazei* (*frisia*) Go1 / ATCC BAA-159 / DSM 3647 (Methanomicrobia; Methanosarcinales)

| | | | | |
|---|---|---|---|---|
| **MM_2097,<br>operon?** | **187** | 33-166 | - | **Hypothetical protein;<br>Cohesin, Pfam (PF00963) and InterPro (IPR002102):**<br>VTLQPASKNVYAGENITLNVVVTPDTQISGIQFDLEYDGSLFQIAG<br>VTEGNLFSQSGQNTLFSPGQIENGKLTNVYGCILGSSSVSTPETFA<br>TVTLKANTGVQGTSQFYLRNVTVSSSASKPVDVRLVNTGITISEPE |
| **MM_2098*<br>operon?<br><br>* see its ortholog, Mbur_0728** | **1132** | 757-891 | 1076-1129 | **Conserved hypothetical Ig protein; PF05345, He_PIG and SM00736, CADG.<br>Putative cohesin:**<br>IAITPEYQIVKPKQSFGLDITVDPSAPIIGTQLDFEFNSSMASANS<br>VIEGDLFKQKGASTFFNEGDINSSEGTIKHIYGLIIGTSNVSSSGT<br>FATVNLTAGNKTGMTEFSLSNVLISDINSKSVPYTVTNATVLI<br><br>**Putative dockerin, PF00404, Dockerin_1;**<br>Y**D**V**N**E**D**GVV**N**IL**D**ITNVSREYET TVSKPYPR<br>Y**D**T**N**Q**D**GEI**N**IL**D**LTLVGHHFGE |

### *Methanospirillum hungatei* JF-1 / DSM 864 (Methanomicrobia; Methanomicrobiales)

| | | | | |
|---|---|---|---|---|
| **Mhun_0185** | **426** | - | 37..89 | **Periplasmic binding protein** [Precursor] (**N-terminal dockerin**), **PF00404; Dockerin_1**<br>G**N**A**N**S**D**MNI**D**QQ**D**IDLINEIAAG KTASTPL<br>A**D**A**N**Q**D**GKV**D**SA**D**AEQVQKIIDG |
| **Mhun_0648** | **547** | - | 495..547 | **Peptidase_C1** (**PF00112**) [**Precursor**]; strong ortholog of **MA_1513,<br>Putative dockerin:**<br>E**D**V**N**G**N**GRA**D**YS**D**VVTFFKQMDW ISKNEPVAA<br>F**D**F**S**G**N**NRI**D**YT**D**VVKLFQSI***** |
| **Mhun_0894** | **472** | - | 414..468 | Hypothetical protein [**Precursor**], |

| | | | | |
|---|---|---|---|---|
| | | | | **Putative dockerin:**<br>EDLNGNGRLDMHDLVLLFQNFRW  IGESNLSLR<br>IDFNKNGRADYADIVTIFNSMNQ |
| **Mhun_0967** | **720** | - | 668..719 | **Periplasmic copper-binding protein,**<br>**Putative dockerin:**<br>TDLNQNGREDLQDVVKYMNMVSS  GDTSSL<br>YDFSGDGKINLNDVVSLFHIIIK |
| **Mhun_1456** | **594** | - | 542..594 | **Peptidase_S8 (PF00082) [Precursor], 1 PKD (PF00801),**<br>**Putative dockerin:**<br>EDINGNGWLEYEDPKLLFDQILF  AMKEEPIGQ<br>FDFDKSGFIGFGDVVKLYQMV* |
| **Mhun_2497** | **1531** | - | 1477-1531 | **Hypothetical protein: 5 PKD (PF00801),**<br>**Putative dockerin:**<br>DDFDGNGVVNTADVTVFFNTWVS  GGFTGMPVPP<br>FDYNHNNRIDTDDIVEYFNLIW* |
| **Mhun_3046** | **528** | - | 474..528 | **Hypothetical protein, C-terminal dock (~rrnAC0273),**<br>**Putative dockerin:**<br>EDMNGNGKTDLQDPTVFFKYFSW  LQSQGYVSA<br>FDFNENGALDLSDVQALFTEIQS* |

*Natronomonas pharaonis* **DSM 2160 / ATCC 35678 (Halobacteria; Halobacteriales)**

| | | | | |
|---|---|---|---|---|
| **NP3026A** | **229** | 31-171 | - | **Hypothetical protein; IPR008965; Carb_bd;**<br>**Cohesin, Pfam (PF00963) and InterPro (IPR002102):**<br>VYFEPDETEATAGETVDIDAEIRVISVYDNEAVESFEYTVEYDPEL<br>LSVDDIEEGPWLSGGNETTVAFETDIDEEAGAVTVSQARQPPAGGV<br>IGEGTTATITFAVAADAPATNTTVAYTDASAQMLEYPLPVIETDAS<br>IEI |

*Pyrococcus abyssi* **Orsay / GE5 (Thermococci; Thermococcales)**

| | | | | |
|---|---|---|---|---|
| **PAB2366** or<br>**PYRAB10440** | **495** | - | 441..495 | **Alkaline phosphatase IV, PhoA like,**<br>**Putative dockerin:**<br>GDVTGDYRVDERDAYATLMLLLG  DLVDTELENI |

| | | | | |
|---|---|---|---|---|
| | | | | A**D**M**D**N**N**GII**D**LL**D**VMAILQASS* |

***Pyrococcus furiosus* Vc1 / ATCC 43587 (Thermococci; Thermococcales)**

| | | | | |
|---|---|---|---|---|
| **PF1004** | **494** | - | 441..494 | **Alkaline phosphatase IV,**<br>**Putative dockerin:**<br>G**D**I**T**G**D**FKV**D**EQ**D**AYVTLMMLLG ERVDTELERK<br>V**D**M**N**N**N**GII**E**LG**D**VLLILQES* |
| **PF0341** | **322** | 95-235 | - | **Hypothetical protein,**<br>**Putative cohesin:**<br>RIEKEIEKLKENIDELDRRTLVVSNLEKFLPRLSEIEEKLYSYPLE<br>VAESMEKRILESLEKKVENIVNSKLKEIESMTPGVIKDIMEKYEKA<br>IRENLELKSKLEEREKIIKKLKEQLKGFEEGAKKVQEIEKKIGEYE<br>KVA |

***Pyrococcus kodakaraensis* KOD1 / *Thermococcus kodakaraensis* (Thermococci; Thermococcales)**

| | | | | |
|---|---|---|---|---|
| **TK1840** | **499** | 20-140 | - | **Cobalt-activating carboxypeptidase, M32 family,**<br>**Putative cohesin:**<br>IWAIGHAQSVLGWDMEVNMPKEGILERSVAQGELSVLSQEFLLKPE<br>FVELVEKAKGIEGLNEYERGVVRVLDRSIRISKSFPPEFLREMSEV<br>TSQATKAWEEAKRSDDFSKFEPWLDRIID |

***Thermoplasma volcanium* ATCC 51530 (Thermoplasmata; Thermoplasmatales)**

| | | | | |
|---|---|---|---|---|
| **TV1521** | **1459** | 1320-1459 | - | IPR000209; **Pept_S8_S53**<br>**Putative cohesin [similar to cohesins of the cellulosomal scaffoldin ScaA** (*Acetivibrio cellulolyticus*)]:<br>VMNLPADEYYPTVNIIAPVTGSSQTGSLNITFTYTGFDDTVILYVN<br>GSSGNVITANVTGKNLYTITEGKGSSSLPAGNYTVTVLVISKDGQV<br>AKQSANFDLKAPVVPPVQSSIIIEYLVIG |

**Uncultured archaeon GZfos1D1**

| | | | | |
|---|---|---|---|---|
| **GZ1D1_26** | **858** | - | 798-854 | **Chitosanase-glucanase like, 3xCMB32,** |

| | | | | |
|---|---|---|---|---|
| | | | | **Putative dockerin:**<br>GDLNCDGEITPADAVIALQIAAG SRLCDDAALAA<br>ADVNDDGRVTSLDALMILQAAAH |
| **C4_0035** | **735** | | 680-731 | **Conserved hypothetical protein**<br>**Putative dockerin:**<br>GDLNSDGTLTPADAAIALRIAAT GAHNDA<br>ADVSGDGCVTSLDALMILQAAAD |

**Uncultured archaeon GZfos17F1**

| | | | | |
|---|---|---|---|---|
| **GZ1D1_26** | **619** | - | 564-615 | **FOG PKD domain protein,**<br>**Putative dockerin:**<br>GDLNHDGQITSADALIALAIAAS GAHHPA<br>ADVSDDNRVTSLDALMIMQAAEG |

**Uncultured archaeon GZfos31B6**

| | | | | |
|---|---|---|---|---|
| **GZ31B6_25** | **81** | - | 17-73 | **Unknown protein,**<br>**Putative dockerin, IPR016134; Cellulos_enz_dockerin_1:**<br>GDINDGGRADFQDAAFLAKHMLG IPGFDNITEGA<br>ADVDGNGLIDIADVMYLARSLLG |

# BACTERIA

## Firmicutes, Mollicutes

| Gene | Protein Size (no. residues) | Putative cohesin | Putative dockerin | Annotation and comments |
|---|---|---|---|---|
| *Acholeplasma laidlawii* PG-8A (Mollicutes; Acholeplasmatales) | | | | |
| **ACL_1119** | **5552** | - | - | **SP, uncharacterized protein precursor,**<br>**IPR016134; Cellulos_enz_dockerin_1:**<br>`G`**D**`I`**N**`G`**D**`GRV`**N**`VM`**D**`QTTLINHLNR   KQPILLSNLLA`<br>`A`**D**`T`**T**`G`**D**`GKI`**N**`VL`**D**`QTALINHLNS` |
| *Eubacterium dolichum* DSM 3991 | | | | |
| **EUBDOL_01701** | **2314** | 1265-1394 | 1204-1266 | **SP, GH31 family protein; 3xCBM32 (F5_F8_type_C), 1 FNIII, FIVAR; a sortase cleavage site,**<br>**PF00963; Cohesin; IPR002102; Cohesin**<br>`NGGVKEKTNASGVLSAVPSKQNVKAGETFTVDIYGSGLKNVYAFDFE`<br>`LSFDKGALSYSDDDFVVKTTAKTRNQKVYEYLHHEETDEKDRVYATI`<br>`TNLGNQTSLNGDMKLATVTLKANKELDLSNLQLSHG`<br><br>**Putative ‘internal’ dockerin**<br>`G`**D**`F`**E**`A`**N**`GII`**N**`DN`**D**`EVWLDSHNPE   AAGARKGGAFYESDANK`<br>`F`**D**`L`**N**`E`**N**`GIF`**D**`AY`**D**`IAVLTTQLNG` |
| **EUBDOL_01704** | **2026** | 1241-1379 | | **SP, GH31 family protein; 2xCBM32 (F5_F8_type_C), PF02368; Big_2; PF07554; FIVAR; 2.**<br>**PF00963; Cohesin; IPR002102; Cohesin** |

| | | | | |
|---|---|---|---|---|
| | | | | FNLDGGTKKEGSVSGNALLLPSATSVNAGDTFTIDVYAEDVKNLNAI<br>GQVINYDPSKLEFQKAEQSALIAQMEDLTVNKTYDSDNTAYVNLAYV<br>NRGNKPVYEGSGILATITMKAKTNLTNVAEAIELNKVTLIGPNHN |

# Firmicutes, Bacillales (A+T rich genomes)

| **Gene** | **Protein Size (no. residues)** | **Putative cohesin** | **Putative dockerin** | **Annotation and comments** |
|---|---|---|---|---|

***Bacillus cereus* ATCC 10987**

| | | | | |
|---|---|---|---|---|
| **BCE_5031,**<br>See also **BC4869** in strain ATCC 14579 | **1407** | 1060-1180 | 1322-1374 | **Subtilase family domain protein** (PF00082; **Peptidase_S8**), or **Minor extracellular protease VPR**; **Cohesin (IPR002102)**<br>**Putative cohesin BCE_5031:**<br>NYWFIPQGAQYAKTSYDKKEVYKGDEFTVTLNAKHVKQFVAGEFNVK<br>FLEKNFKFANAKLNPAFEKLLSEKGVTAKVNEPKLEEGSVTVGGAID<br>DKNFAGLDGDFPFIDVTFKVENDEFYE<br><br>**Dockerin BCE_5031:**<br>G**D**V**N**G**D**KVI**D**IK**D**AEIIA**N**NYGK  KGLSVKD<br>G**D**L**N**K**D**GIV**D**EK**D**IRFVEKNFLK |

***Bacillus cereus* ATCC 14579 or DSM 31**

| | | | | |
|---|---|---|---|---|
| **BC_4869** | **1407** | 1060-1180 | 1322-1374 | **Subtilase family domain protein** (PF00082; **Peptidase_S8**), or **Minor extracellular protease VPR**; **Cohesin (IPR002102)**<br>**Putative cohesin BC_4869:**<br>NYWFIPQGAQYAKTSYDKKE**L**YK**N**DEFTVTLNAKHVKQFVAG**Q**FNV**N**<br>FLEKNFKFANAK**F**NPAFEKLLS**Q**KGVTAKVNEPKLE**A**GSVTVGGAID |

| | | | | |
|---|---|---|---|---|
| | | | | DKNFAGLDGDFPFIDVTFKVENDEFYE<br><br>**Dockerin BC_4869:**<br>G**D**V**N**G**D**KVI**D**IK**D**AEIIASNYRK   KGLTVKD<br>G**D**L**N**K**D**GIV**D**EK**D**IRFVEKNFLK |

*Bacillus cereus* G9241

| | | | | |
|---|---|---|---|---|
| **BCE_G9241_4981** | **1407** | 1060-1180 | 1322-1374 | **Subtilase family domain protein** (PF00082; **Peptidase_S8**), or **Minor extracellular protease VPR**; **Cohesin (IPR002102)**<br>**Putative cohesin BCE_G9241_4981**<br>NYWFIPQGAQYAKTSYNKKEVYKGDEFTVTLNAKHVKQFVAGEFNVK<br>FLEKNFKFANAKLNPTFEKLLSEKGVTAKVNEPKLEAGSVTVGGAID<br>DKNFAGLDGDFPFIDVTFKVENDEFYE<br><br>**Dockerin BCE_G9241_4981**<br>G**D**V**N**G**D**KVI**D**IK**D**AEIIA**S**NYGK   KGLSVKD<br>G**D**L**N**K**D**GIV**D**EK**D**IRFVEKNFLK |
| **BCE_G9241_0617** | **1413** | 1062-1187 | 1328-1380 | **Reticulocyte binding protein (Peptidase_S8), SP**<br>**Cohesin BCE_G9241_0617**<br>DYVFVKEGTEYAVPSYNKDKIKLGEKITLTLNLNNVKQLMTGTFEIP<br>YYKQLFKFADVKPNPALTEYVKQHGLNLKLEDPKINEEGAWENKVKV<br>GASLEGKEFKGLDGDTPFLDVTFEMTNDEYFN<br><br>**Dockerin BCE_G9241_0617**<br>G**D**V**N**G**D**KMV**D**IQ**D**ARIAALSYGK   GKVSVKD<br>G**D**I**N**Q**D**GVV**D**ET**D**IRFIEKNFLK |

*Bacillus thuringiensis* **strain Al Hakam, a relative to** ***B. anthracis***

| | | | | |
|---|---|---|---|---|
| **BALH_0575** | **1448** | 1097-1222 | 1363-1415 | **Minor extracellular protease VPR**, **Peptidase Vpr. Serine peptidase. MEROPS family S08A** (**Peptidase_S8**); **Cohesin (IPR002102)**<br>**Putative Cohesin**  (see a similar domain in the ***B. cereus*** **G9241** |

| | | | | |
|---|---|---|---|---|
| | | | | protein **BCE_G9241_0617**)<br>DYVFVKEGTEYAVPSYDKDKVKLGEKITLTLNLNNVKQLMSGTFEIP<br>YSKQLFKFVDVKPNPALAEYAKQHGLNIKLEDPVINEEGNWENKVKV<br>GASLEGTEFKGLDGDTPFVDVTFETTSDEYFN<br><br>**Putative dockerin** (see in ***B. cereus*** **G9241**):<br>G**D**V**N**G**D**KMV**D**IQ**D**ARIAALSYGK   GKVFVKD<br>G**D**I**N**Q**D**GVV**D**ET**D**IRFIEKNFLK |
| **BALH_4434** | **1407** | 1060-1180 | 1322..1374 | **Subtilase family domain protein** (PF00082; **Peptidase_S8**), or **Minor extracellular protease VPR**; **Cohesin (IPR002102)**<br>**Putative cohesin (95% identity to BCE_5031):**<br>NYWFIQQGAQYAKASYDKKELYKGDEFTVTLNAKHVKQFVAGEFNVK<br>FLEKNFKFANVKLNPAFEKLLSEKGVTAKVNEPKLEEGSVTVGGTID<br>DKNFAGLDGDFPFIDVTFKVENDEFYE<br><br>**Putative dockerin:**<br>G**D**V**N**G**D**KVI**D**IK**D**AEIIASNYGK   KGLSVKD<br>G**D**L**N**K**D**GIV**D**EK**D**IRFVEKNFLK |

***Bacillus* sp. B14905**

| | | | | |
|---|---|---|---|---|
| **BB14905_18325** | 134 | - | 48-100 | **Microbial serine proteinase without predictable signal peptide**.<br>**Dockerin_BB14905_18325**<br>G**D**V**N**G**D**NVI**D**IK**D**AFFLKNHINK   PGKGPEK<br>A**D**I**N**K**D**GAI**D**VK**D**MKIIVENYLR |

***Bacillus* sp. GL1 (there is no genome project for this bacterium)**

| | | | | |
|---|---|---|---|---|
| **No name, accession No AB006853** | 2475 | 2100-2245 | - | **SP, gellan lyase precursor, FNIII,**<br>**Putative cohesin (similar to *Paenibacillus* cohesins):**<br>GGTLAASAEQLQPGQQLELTVGVSDASRFTGADILVHYDPQALTFAT<br>ELYEGVRMLKAEAIASLQANYQVAAAMAEQPGTIKILLFTAGAGQPL<br>SGTLPLFKLRASVKDDAQTGVSTAVSLSDFELTFEGEDSVWPDTTRA |

| | | | | |
|---|---|---|---|---|
| | | | | AVSLQ |

***Oceanobacillus iheyensis*** **HTE831 / DSM 14371**

| | | | | |
|---|---|---|---|---|
| **OB0236** | **1358** | 1033-1162 | 1276..1325 | **SP, putative serine proteinase (PF00082; Peptidase_S8),**<br>**Putative cohesin:**<br>TYYFIKEDLPYGYFESNASSLKTGEEFTATLVLENVENLKEGNWSIS<br>NFTESFELVDAKANDSLTEDENPIVNTELQGSTLNITMNTDKPLSGK<br>VKAVDLNIKVSDSSFQVLGEMEPTLEYTDSDSTKHT<br>**Putative dockerin:**<br>G**D**V**N**Q**D**DVI**D**IE**D**AILIKEHWNS   SHRN<br>T**D**I**N**F**D**GTV**D**SE**D**INFVVSNYLT |
| **OB0731;<br>operon?** | **94** | - | 15..64 | **SP, hypothetical protein** (Incomplete? Frame shift?), **PF00404;**<br>**Putative dockerin:**<br>G**D**I**N**Q**D**SVI**D**II**D**ALLMVKYWGS   DKQA<br>A**D**F**N**F**D**GTG**D**KK**D**FELLAANFLK |
| **OB0732;<br>operon?** | **263** | 28-147 | - | **SP? hypothetical protein OB0732,**<br>**Putative cohesin:**<br>MHNFVKEGTPYSTVSSNKKEITVDEEVEFTLTAHNVEMLQEANFQLN<br>YDKQQLQLVDVKLHEDVESYGDPTITYSEESISNTRNNISISTAIND<br>IEQEVTGDIPIAVATFTGKEADFGRN |

***Paenibacillus*** **sp. JDR-2 (Bacillales, Paenibacillaceae)**

| | | | | |
|---|---|---|---|---|
| **Pjdr2DRAFT_0219,<br>putative operon** | **348** | 28-160 | - | **S-layer domain protein (PF00395; SLH; x3),**<br>**Putative Cohesin domain (~Pjdr2DRAFT_4326):**<br>ASYEISADQTSVGVNKALTVTVKASELKNVYGYEIVLSYDTTKLSLQ<br>SANNLLPKNNAAEQNGFAISPIKSGSKVTFAYTKVGSATGSEGSKAL<br>AAFTFGGLAAGKASVKIESVKLVHTDLTSEKQTPDKSV |
| **Pjdr2DRAFT_0220,<br>putative operon** | **1503** | - | 1445-1502 | **SP, putative GH3 family protein, CBM6 (CBM_IV), Ig family protein,**<br>**PF00404; Dockerin_1:** |

| | | | | |
|---|---|---|---|---|
| | | | | G**D**V**N**G**D**NRI**S**VL**D**LAQVSVNYGK TSQSADWSSIKR<br>M**D**L**N**N**D**GVI**D**IS**D**LVIIAQKILQ |
| **Pjdr2DRAFT_0859, putative operon** | **865** | ~661-808 | 809-865 | **SP, putative Xylan 1,4-beta-xylosidase, GH3 family protein; PF00963; Cohesin; IPR002102; Cohesin, 27%ID / 56%SIM to CPF_1442**:<br>GFGLSYEANVSDISATLTGETSVTAGKTFNLNFGLDKVAQNVYAQEL<br>TLNFDPAKVDFVDATSLKAGLEIVSKEVKAPGKVRIIFASLGAGNAI<br>NSDGDLLRLQLKAKSNAQIASTDVKLTNIILADGNGVETPLAEASKS<br>INIGYIK<br>**Putative Dockerin:**<br>G**D**I**N**G**D**GKV**S**IG**D**LGIVAAAFGK  TDKSPDWADYKA<br>A**D**V**K**T**D**GKI**D**IF**D**LSTVAQLIE***** |
| **Pjdr2DRAFT_0860, putative operon** | **575** | 31-160 | - | **S-layer domain protein (PF00395; SLH; x3)**<br>**PF00963; Cohesin; IPR002102; Cohesin:**<br>SFTLNVTSTKVELGNELLITVTGNDLRDVYGYEFSLSFDQTMLEYKG<br>AQSGISSGFSIKPMLEGNQLRFASTKLHKVAGESGALTLATLTFKTI<br>RQGTATLQLKSVQTVDSQLTGHMYLPDTKVSATITS |
| **Pjdr2DRAFT_1008, putative operon** | **336** | 35-170 | - | **S-layer domain protein (PF00395; SLH; x3)**<br>**Putative Cohesin domain (~Pjdr2DRAFT_4326):**<br>VELLISASNSKVELRVMGHNLDDLYAYDFVVHYNNQKLKFLSTTSSF<br>GGFAVDPLKKDNTVRIARTKVGAVDGLNGDAELATLTFERIASGASE<br>ISLTNVQVVNSKLDLKALSDAKTNVPNNVTFSDIKGHWAESA |
| **Pjdr2DRAFT_1009, putative operon** | **1224** | 853-1002 | - | **SP, putative GH3 family protein, IPR011490; FIVAR_sugar_bd, PF00963; Cohesin; IPR002102; Cohesin:**<br>EEKAAVELKGPATAISGQSIDLSVNVNDVASEFNTMSVVVNYDPAKL<br>EFATNTDSEGLVSLAEQALESTNPDLQILGSGVKPDKGQILLLLSTT<br>GDLIKPDGQLLVLHAKVKSDQTGSVHTFLSNFTVSANGQAQQTLDTT<br>AAVSDIALR |
| **Pjdr2DRAFT_1202** | **336** | 30-155 | - | **SP, S-layer domain protein (PF00395; SLH; x3),**<br>**Putative cohesin domain (~Pjdr2DRAFT_4326):**<br>NATFDLKLSYSGKDVKLSIMGHGLTDMYAYDFELSYDEKILSFSKVE<br>TSIKGFSVDPILGANTIRIAHTKMGETGGEKGDVELSAITFKRIGSG |

| | | | | |
|---|---|---|---|---|
| | | | | TTSVKLNTMKLVDSKLDMAAFEPKVTATVID |
| **Pjdr2DRAFT_1556, putative operon** | **1152** | - | 135-191 | **Putative uncharacterized protein precursor,**<br>**Putative dockerin:**<br>GDTTGDGLLTSADVLYIYKVINS KLPITDQEKNR<br>LDVNRDGVITTADATVLMNSYVG |
| **Pjdr2DRAFT_1557, putative operon** | **611** | 37-177 | - | **SP, S-layer homology (SLH) domain, x3; pfam00395 PF00963; Cohesin; IPR002102; Cohesin (similar to cohesins in Cellulosomal anchoring scaffoldin ScaB (*Bacteroides cellulosolvens*):**<br>GVPVQLKIGDVSVNTGGTVDVPVSIKQPERGIGSYNVQLNYDPKALE<br>IISVKPKYGDANTETCSESQEGCFVSNFDQENGWLRVIWVDTTGGDR<br>LINEAKQLFTIQVKAKPGSESAKQTLTVDAEDPASLTFTDGDMHTLP |
| **Pjdr2DRAFT_1583, putative operon** | **543** | 30-166 | - | **SP, S-layer homology (SLH) domain, pfam00395; SGG-linker, Putative cohesin domain (~Pjdr2DRAFT_4326):**<br>PPSFSMTSDKSALNSGDEVTVTVKGSKLQDVYAYELQVYYDSDLLTF<br>KTGSEKTALVGFSTPAKVVADGGGESHLVFAHTKSGSSPGDSGDLDL<br>ASFTFTAAANGKADVEVRKIRLVDSKLASSTIDSPYGTSIQIG |
| **Pjdr2DRAFT_1584, putative operon** | **725** | 40-179 | 180-235 | **SP, putative TWO-enzymatic domain protein, PF01425; Amidase (247..716),**<br>**N-terminal PF00963; Cohesin; IPR002102; Cohesin:**<br>EQPQVQEPKALVQLSGPALTSVGNSFDVTVSLTGATGDATAGKFIFE<br>YDPTQFEYLSASSANSAVSIAGVNAVSGKVTIVTTDATGSLTTAGSL<br>LTLTFKANAAADLSGISGTVELGIDDGSTVQTPEESLAVQITGAAE<br>**N-terminal PF00404; Dockerin_1:**<br>GDLNNNSVIDVGDLSVLIKYFGI ASDDPNWTKVAN<br>GDFNKNNTIDVQDLASLGKKVLY |
| **Pjdr2DRAFT_1620** | **248** | - | 189-246 | **PFAM: cellulosome protein dockerin type I, SP**<br>**Putative dockerin:**<br>TDLFKDGKLTSADLSVAIDHYRA SSTDTDWANAKL<br>ADVNFDGKVDLTDFTIIILHIFQ |
| **Pjdr2DRAFT_2233, putative operon** | **1745** | 1458-1610 | - | **SP, putative GH6 family protein, IPR011490; FIVAR_sugar_bd, putative Cohesin domain (~Pjdr2DRAFT_1584,** |

| | | | | |
|---|---|---|---|---|
| | | | | **Pjdr2DRAFT_4327):**<br>GSGLSGSARIAISGQEGALLSGPASVVSGEAFKLNFGLNNLTSDVYA<br>IDLSVRYDPAKVEYVSGISLVDGFSIVLDQEGASGEGEANIRFLAAG<br>AGQAISNTDNLLELSFKAKNVDATVELNAITVEKVIISDGITETPVS<br>STSHNVVIEVTD |
| **Pjdr2DRAFT_2234, putative operon** | **590** | 43-170 | - | **SP, S-layer homology (SLH) domain, pfam00395; putative Cohesin domain (~Pjdr2DRAFT_4326):**<br>SYTLSVSNNEPVAGREVRVQVIGQELSDVYATELQATFDTDHLRFKS<br>ASSDRFGYAVAPAVKGSQIVLAFTKVGPVSGESGTVVLAEMVFEATA<br>AGSGTVELKKVKTVDSAMNVADQDANAIANLKVV |
| **Pjdr2DRAFT_4026** | **629** | 17-153 | - | **S-layer domain protein (PF00395; SLH; x3), PF00963; Cohesin; IPR002102; Cohesin:**<br>TSAPRAEANAPALPAVTAAPSVQGELVNVTIKGSGVGHLYGVQLQIA<br>YDATALEFVSAKPRYFEYASQRQVNTDAYDLLDTAGAPVGFQLVTAE<br>DGNIDYLASQVKQGATVSDNSAVLDLQFKALHAGTATVSVVKS |
| **Pjdr2DRAFT_4326, putative operon** | **343** | 22-162 | - | **S-layer domain protein (PF00395; SLH; x3), PF00963; Cohesin; IPR002102; Cohesin:**<br>SDTYAAETPSFTITADKTTPQAGGEVVVTVHGDHLQALYGYEINLSY<br>SSKKWQFVSAASALSGGFAVPPIVKDDTLTFAYTKTGSTTAGVSGSS<br>NLATVKFKSIEAGSSDIKLTRIKTVDKSLKANEYTPNVTVYIQPSSI |
| **Pjdr2DRAFT_4327, putative operon** | **313** | 34-176 | 255-312 | **SP, uncharacterized protein, EF-hand, calcium binding motif PF00963; Cohesin; IPR002102; Cohesin:**<br>RSSAAAGVLSGPNQIHSGAAFDVKYGLAGVSSGNVLAEDVTVTYDPS<br>KMDFVSVASLDEEKYLVVGQQPDAAAGKLRFLGVRLGDAQTNPNGEL<br>VTLKFKAKRNAGAGIANISITNLVIADEEAHETSLDGSTLGVQINVI<br>DR<br>**Putative dockerin:**<br>G**D**S**N**G**D**NNL**T**VG**D**LAIVAKAYGA  KSTDPNWNSVKS<br>Y**D**R**N**N**D**GKI**D**IE**D**LVWLAVKILN |
| **Pjdr2DRAFT_4830** | **696** | - | 613-668 | **IPR002048; EF_hand_Ca_bd, Putative dockerin:**<br>T**D**L**D**G**N**GQV**N**WQ**D**VIEFLRAYGT  KRGQLHYDPR |

| | | | | |
|---|---|---|---|---|
| | | | | A**D**F**D**A**D**GVI**G**KK**D**LIAFLRAGHF |
| **Pjdr2DRAFT_5030** | **761** | 180-320 | - | **SP, S-layer homology (SLH) domain; pfam00395,**<br>**Putative cohesin domain (CipA and ScaB like):**<br>KKAEVTIGSKAGEPGTLVQVPVKLSDATTGVGSYGVALRFDKDALEV<br>NRVDGPPQSGFASNFDNEAGIVQAGWADLSGGDNPILAGQDLFTVTF<br>KIKNEAQLGDKAVNVADPENIEQFTLTDADAEEMVKSLHAGNVNVYR |
| **Pjdr2DRAFT_5031** | **448** | - | 389-445 | **SP, uncharacterized protein precursor,**<br>**Putative dockerin:**<br>G**D**A**N**G**D**GYV**T**PA**D**ALLANKYIQG   KITLTAEQKLA<br>L**D**M**D**N**D**NDV**D**AV**D**AQLILNVYLG |

# Firmicutes, Lactobacillales (A+T rich genomes)

| **Gene** | **Protein Size** (no. residues) | **Putative cohesin** | **Putative dockerin** | **Annotation and comments** |
|---|---|---|---|---|

***Enterococcus faecalis*** **V583 / ATCC 700802**

| | | | | |
|---|---|---|---|---|
| **EF_1824** | **1866** | ~1206-1330 | 1134-1192 | **SP, GH31 family protein; 1 CBM32, 1 FNIII, 5-6 FIVAR; a sortase cleavage site**,<br>**Putative cohesin-like domain:**<br>LSLAVVNENGKDTYLPGDTLTFILKGQDLKNINALSTKMSFDSSKFE<br>LVGQPATTNNTQQMENYSKYRKHSNDVENLYLVLSNQGNKQLLNGSM<br>DLVTFKVKVKETTRVKRATTVEQPLQFDMSQ<br><br>**An 'internal' dockerin**<br>G**D**I**T**N**D**GTI**D**EN**D**AMSYRNYTGL   ESVDSDFNGYVEK<br>G**D**L**N**K**N**GVI**D**AY**D**ISYVLRQLDG |

***Lactobacillus casei*** **ATCC 334**

| | | | | |
|---|---|---|---|---|
| **LSEI_2769** | **1771** | 1210-1332 | 1138..1197 | **SP, GH31 family protein, 1 CBM32, 1 FNIII, 5-6 FIVAR, Putative cohesin-like domain**<br>SGTLALQSDKTTYAPGETITLNLIGKDLTNVNSLFARLPLTNPNVDL<br>IQVEPTAATAGMVNFSKVRTHSNSSRDLYLIFANEGEQKRLTGNEKL<br>ATITLKAKTQLTKADLSFALTDPMLTNQWG<br><br>**Dockerin,** strongly similar to **CPF_1301 dockerin** (*C. perfringens* ATCC 13124)<br>G**D**I**T**N**D**GAI**T**GD**D**QMSLVNYAGL TANIDSDFNGYVQN<br>G**D**L**N**Q**N**GVI**D**AF**D**INYVMTKLGK |

*Lactobacillus johnsonii* **NCC 533**

| | | | | |
|---|---|---|---|---|
| **LJ_0382** | 3619 | 1342-1465 | - | **SP, hypothetical protein, 4 MucBP (MUCin-Binding Protein), LPxTG_anchor,**<br>**Cohesin-like domain_(29%ID / 40% SIM to Mbur_0729)**<br>PSTSTDADKVSLSKTITRTITVEGLPTAVEGTTQTVTFTRTAVVDEV<br>TGKVIGYVDPSDTSQTITDGDNAWTSVNNTWSAFTPKNIPLGYSIKS<br>ITDANGNYANVTSTDGKLTGVAQTTVA |

# Firmicutes, Clostridia

| Gene | Protein Size (no. residues) | Putative cohesin | Putative dockerin | Annotation and comments |
|---|---|---|---|---|

*Clostridium bolteae* **ATCC BAA-613 (isolated from human feces)**

| | | | | |
|---|---|---|---|---|
| **CLOBOL_00465** | **582** | 34-160 | - | **SP, uncharacterized protein, IPR001444; Flag_bb_rod PF00963; Cohesin; IPR002102; Cohesin:**<br>ANARIAFSDPSATVGGKIKVNMKITSSDNLANADVMLSYDSNILEFV |

| | | | | |
|---|---|---|---|---|
| | | | | DGTNADGGAGAVRVHGDAGTPNTGTLVFTLNFNAIAAGTSKIEVTSQ<br>EIYDSNSQIVTVNQQGNSTVTVSALQSASKDAT |

***Clostridium butyricum* 5521**

| | | | | |
|---|---|---|---|---|
| **CBY_1438** | 1989 | - | 280-334 | **SP, hypothetical exported protein,**<br>**Putative 'internal' dockerin (Pfam: PF00036, efhand; InterPro: IPR016134; Cellulos_enz_dockerin_1):**<br>G**D**I**N**G**D**SII**D**ME**D**YKLVFDKIGI   KNTKTDLKK<br>Y**D**L**N**R**D**GKI**D**VA**D**LSVVYENIGK   NQEEAKVE |

***Clostridium difficile* strain 630** (see also the orthologous domains in **strain QCD-32g58**)

| | | | | |
|---|---|---|---|---|
| **CD2793 or SlpA** | **719** | **~241-407** | - | **Cell surface protein or S-layer protein precursor SlpA** (Calabi *et al*., 2001; Eidhin *et al*., 2006 *etc*); it was shown that SlpA is a processing protein – there are two form of SlpA, a low-molecular weight peptide (**residues 25-345**) and a high-molecular weight (**residues 346-719**). **THREE** CW_binding_2/putative cell wall binding repeat 2 (Pfam: **PF04122**).<br>**A 'questionable' cohesin, has a severe divergence in strains**:<br>GGSTPSAVAVSGFVTKDDTDLAKSGTINVRVINAKEESIDIDASSYT<br>SAENLAKRYVFDPDEISEAYKAIVALQNDGIESNLVQLVNGKYQVIF<br>YPEGKRLETKSANDTIASQDTPAKVVIKANKLKDLKDYVDDLKTYNN<br>TYSNVVTVAGEDRIETAIELSSKYYN |
| **CD2797** | **1987** | - | **182-231 ..279** | **Putative exported protein** (SP: **1..24 aa**): **CBM32, 2 FNIII**<br>**Possible dockerin (Pfam: PF00036, efhand; InterPro: IPR016134; Cellulos_enz_dockerin_1), two first 'repeats' are strongly similar to several *C. thermocellum* 'enzymatic' dockerins, see also a NagH dockerin in *C. perfringens*:**<br>G**D**V**N**G**D**GKV**D**KS**D**YNKVLENIDS  NKRE<br>F**D**L**N**R**D**GKV**D**IV**D**LDYVQKNLGQ  SEDAKSLESIVSTNPIVDTSKVELK<br>G**N**S**D**V**D**ING**N**VE**N**LFSGENNGVT<br>**See identical amino acid sequences in:**<br>>1987-aa ORF **CdifQCD-6_10159** in ***C. difficile* QCD-63q42**; |

| | | | | |
|---|---|---|---|---|
| | | | | >1987-aa ORF **CdifQCD_07039** in ***C. difficile*** **QCD-37x79**;<br>>1987-aa ORF **CdifQCD-7_05756** in ***C. difficile*** **QCD-76w55**;<br>>1987-aa ORF **CdifQCD-_18531** in ***C. difficile*** **QCD-97b34**;<br>>**381-aa** ORF **CdifQ_04003264** in ***C. difficile*** **strain QCD-32g58.** |

***Clostridium difficile*** **strain QCD-32g58** (an **incomplete genome**)

| | | | | |
|---|---|---|---|---|
| **CdifQ_04003264** | 381 | - | **2..49<br>..99** | **Exported protein** (a putative **mutant ORF**, represents 'internal' sequence of the **1987 aa CD2797**); **binary toxin positive**; tcdC 18 bp deletion; PFGA<br>**Possible dockerin (Pfam: PF00036, EF hand; InterPro: IPR016134; Cellulos_enz_dockerin_1), two first 'repeats' are strongly similar to several *C. thermocellum* 'enzymatic' dockerins, see also a NagH dockerin in *C. perfringens*:**<br>G**D**V**N**G**D**GKV**D**KS**D**YNKVLENIDS NKRE<br>F**D**L**N**R**D**GKV**D**IV**D**LDYVQKNLGQ SEDAKSLESIVSTNPIVDTSKVELK<br>G**N**S**D**V**D**ING**N**VE**N**LFSGENNGVT |

***Clostridium leptum*** **DSM 753**

| | | | | |
|---|---|---|---|---|
| **CLOLEP_03013, putative operon** | **1051** | - | 989-1043 | **SP, uncharacterized protein, PF02368; Big_2, Pfam PF00404; Dockerin_1**;<br>G**D**I**D**N**N**GTI**N**LK**D**VTQLQKYLAG WDVEVATKY<br>C**D**V**N**A**D**GKI**N**LK**D**VTHLQKYLAG |
| **CLOLEP_0301, putative operon** | **455** | 28-170 | - | **SP, uncharacterized protein,**<br>**Pfam motif PF00963; Cohesin,** ~ ScaB cohesins in *Bacteroides cellulosolvens*:<br>VGNDEPAFIVEEVSGYAGDEVTVNINVNNNPGICAAGVTVSYDGNTL<br>ELKSVTNGSVLGGNIFGPVSNNPIKLTWSTTENSTANGVLATLTFTV<br>KENADSGKSAITLTYDPDEVYNIDMTNVAFAVENGSVTINCNHVAGA<br>WE |

***Clostridium perfringens*** **strain ATCC 13124**

| | | | | |
|---|---|---|---|---|
| ***nagH*** or<br>**CPF_0184** | **1627** | **-** | 1578-1627 | **SP, Mu toxin** or **NagH, hyaluronidase / hyaluronoglucosaminidase** (EC 3.2.1.35) = **GH84, 2-3 x CBM32, 3 proposed N-acetyl glucosamine-binding modules (FIVAR?),**<br>**Dockerin_CPF_0184 interacts with Cohesin_CPF_1442:**<br>G**D**F**N**G**N**GKI**D**IG**D**LAMVSKNFGS   TNTS<br>L**D**L**N**K**D**GSI**D**EY**E**ISFINHRILN |
| **CPF_0259** | **845** | 641-781 | - | **SP, GH3 family enzyme** (**β-N-acetylhexosaminidase?**), contains a putative **β-glican binding domain**,<br><br>**Cohesin_CPF_0259 (18.3%ID / 49.3%SIM to cohesin CPF_1442)**<br>HGITNINSLGNVNISMNKKINLGDNFQVKFNLSDFNEIVAGKYRAKIKFQ<br>GEKLEFINGKVEFSGDLQTNIIDNNTLEVLINLEEGSLKANEVNFILEFK<br>AIDKAELTSIEITSSELIDEKGR**C**FKQKYVISEFSIEDNKE |
| ***nanJ***<br>or **CPF_0532** | **1173** | 939-1081 | - | **NanJ**, putative **exo-α-sialidase (GH33), CBM32 (or CBM40?), 3 BNR (Asp-box repeat)** and **1 FNIII**, **SP**<br><br>**Cohesin_CPF_0532 (34%ID / 56%SIM to cohesin CPF_1442)**<br>QIGELKTTVGNSTIKVNDEVQVGSAFEAILGIEGLNGDTEVYSAEYLFEY<br>NAEAFILNEITSFNDSLFVKSKEVEPGKVRILVASLGNEIEKDSDLVKVN<br>LTPKISSELEVLGLTTALVGAGDGNTHDLELSSKEVKINEEAS |
| **CPF_0859** | **2095** | ~1880-2022 | - | **SP, probable alpha-N-acetylglucosaminidase, GH89** (similar to the *Homo sapiens* NagLU), **5 (6?) CBM32, 2 FIVAR** (Uncharacterised Sugar-binding Domain), **FNIII**,<br><br>**Cohesin_CPF_0859 (20%ID / 38%SIM to cohesin CPF_1442**)<br>VYIPNAIEELKKALDKAKEVLNNEGANEEDINKAISDLTEAEANFELSKV<br>PNKDELNSEIERAYDVLEKLKGFEHLEGIKIKLEDAILEAERVKNDENSL<br>KVDVDKVLGELKIVIEEALKAIDSSEVIV**C**PVRDFKATNISKK<br><br>**see similar domains in:**<br>*Clostridium spiroforme* DSM 1552, *Clostridium leptum* DSM 753, *Clostridium ramosum* DSM 1402. |

| | | | | |
|---|---|---|---|---|
| **CPF_1301**. | **1965** | 1710-1873 | 1105-1162 | **SP, α-glucosidase, GH31** domain**, 2xCBM32, 2 FNIII (CBM37?), FIVAR, Big_2 (Bacterial Ig-like domain, group 2),**<br><br>**An 'internal' Dockerin_CPF_1301 (55% ID to the *Enterococcus faecalis* dock EF_1824)**<br>G**D**W**N**N**S**GSI**E**EG**D**LVFLENYAGL  TTADSDWGYVSM<br>A**D**L**N**N**N**GLI**D**AY**D**ISYVSSKLEG<br><br>**Possible Cohesin_CPF_1301 ( 20%ID / 37%SIM to CPF_1442)**:<br>VTEYKIETSLDNENWTEVRTGNLEVNEAGNALANRGYNPIRFNTINAQYL<br>RFTALKTLGDTNNKYASAAELVFYGKEGKVSAESITLEKTELKLNVNESE<br>QLKAVLNPIESNDTITWTSSDESIAKVDENGVVTGIGKGEALITATIPNG<br>KSATSKVIVEDSVS<br>**See similar domains in:**<br>*Anaerostipes caccae, Bacteroides spp., Clostridium acetobutylicum, Clostridium beijerinckii, Clostridium botulinum, Clostridium butyricum, Clostridium ramosum, Clostridium spiroforme, Desulfitobacterium hafniense, Eubacterium dolichum etc.* |
| ***nagJ***<br>or **CPF_1442** | **1001** | **768..909** | - | **SP, Hyaluronidase NagJ** or **Hexosaminidase B (GH84/GH20b), CBM32, FNIII, Cohesin (G3DSA:2.60.40.680),**<br>i) A 31...624 aa domain has been crystallized (PDB: 2CBJ, release Jan 5, 2006).<br>ii) 768..909 aa; <u>cohesin like</u> domain has been crystallized (PDB: 2JH2, release Dec 14, 2007). Chitayat et al. J. Mol. Biol (2008)<br>**Cohesin_CPF_1442 (interacts with Dockerin_CPF_0184)**<br><br>KLKENAEVTGSVSLEALEEVQVGENLEVGVGIDELVNAEAFAYDFTLNYD<br>ENAFEYVEAISDDGVFVNAKKIEDGKVRVLVSSLTGEPLPAKEVLAKVVL<br>RAEAKAEGSNLSVTNSSVGDGEGLVHEIAGTEKTVNIIEGTS |
| ***lacZ***<br>or **CPF_1474** | **1355** | **-** | 1299-1354 | **SP, Beta-galactosidase (GH2)**, **2 Big_4** (Bacterial Ig-like domain, group 4), |

| | | | | |
|---|---|---|---|---|
| | | | | **Dockerin_CPF_1474** (similar to that in the GH_48 glycoside hydrolase of *Bacteroides cellulosolvens*)<br>A**D**I**N**E**D**GSI**N**VG**D**LSIVSKYQGE  IISGNALSEK<br>S**D**I**N**K**D**GVV**D**KA**D**IQIVMGKILG |
| ***nagK*** or<br>**CPF_1487** | **1172** | 876-1016 | - | **SP, Hyaluronidase NagK or Nag84A (GH84), CBM32, FIVAR, Gram_pos_anchor (PF00746), SP**<br><br>**Putative Cohesin_CPF_1487 (18.5%ID / 37.3%SIM to CPF_1442)**:<br>RIINNTDKDVTFNLNKLEAEYSNNDVNFDIRPTAEAKFEPKNLIDGKLNT<br>AFKPLESAPKSGQLTYRISDKTDIKKFTIVQNPNTISNAIVSVRNENGWK<br>EIGSLGKSFNEFNTEDFENVFEIKVEWDGFAPTIYEIGLST |
| **CPF_1614** | **1471** | 1236-1379 | - | **SP, Putative Glycosyl hydrolase domain (~GH20), 3xCBM32 (one of CBM32's has been crystallized** [Ficko-Blean & Boraston, 2006]**), FNIII**,<br><br>**Cohesin_CPF_1614 (35%ID / 58%SIM to CPF_1442)**:<br>VLEVPQENVGQSFLTLPESLKVNENFDVFLGANEIKEDLSAYAAEFTLKY<br>NQEVFDLVEVASGIDGVLVNYKEVEPGVVKILVASLEKTIENDTNLVKVT<br>LTPKSESEREIFEVISANLGDGENGKVCELGLTSGEVAVGKGSV |
| **CPF_2129** | **1479** | - | 1430-1479 | **SP, putative Glycosyl hydrolase, GH95** domain, **CBM51, a novel putative carbohydrate binding module (NPCBM/NEW2) domain is found at the N-terminus of GH_98 family proteins**, **~FNIII motif**,<br><br>**Dockerin_CPF_2129 (46%ID / 84%SIM to dockerin CPF_2130)**:<br>A**D**F**N**K**N**GEI**D**LG**D**LSMVSKYFGS  NNSE<br>F**D**L**D**G**D**GLV**G**EY**E**INFVSSELLK |
| **CPF_2130** | **932** | - | 883-932 | **SP, α-L-fucosidase domain** (**GH29**), **CMB32, calx-beta domain**,<br><br>**Dockerin_CPF_2130 (46%ID / 84%SIM to dockerin CPF_2129)**:<br>T**D**L**N**K**D**GKV**D**IG**D**LSLASKHYGE  NIKE<br>Y**D**I**N**G**D**GVI**D**EF**E**LNSITNDILK |
| | | | | |
| ***nagL* or CPE1523** | **1127** | 885-1031 | - | **SP**, **Hyaluronidase NagL (GH84)**, **CBM32**, **FNIII**, |

| | | | | |
|---|---|---|---|---|
| (strain 13); there is **NO such gene in ATCC 13124** | | | | **putative Cohesin_CPE1523 (30%ID / 53%SIM to *nagK*** or **CPF_1487)**:<br>RYIKIQSNGHNYGCWVEIYEIEFNKTAPEFDESAVNLASGTLEGNFNAFY<br>DEDLSTAYEAKSVKDGDSLIYKMSRVTNISELEILQDKNRISEAKVSVKD<br>LGGNWTEIGHLNAQINKLKVDNNITEVKFEFEGSKPAPKIYEIIARE |

NOTE: See orthologous cohesin and dockerin domains in *C. perfringens* strain 13, *C. perfringens* strain SM101, *C. perfringens* strain NCTC 8239, *C. perfringens* CPE strain F4969, *C. perfringens* B strain ATCC 3626, *C. perfringens* C strain JGS1495, *C. perfringens* D strain JGS1721, *C. perfringens* E strain JGS1987.

***Clostridium phytofermentans* ISDg (ATCC 700394)**, a soil bacterium that **degrades cellulose anaerobically**

| | | | | |
|---|---|---|---|---|
| **Cphy_3465** | **416** | 34-170 | - | **SP (1..32), extracellular uncharacterized protein,**<br>**Putative cohesin:**<br>ATIKIEVPNDEVGKGDEFKISIVIASDSILGDFEGYLSYNAEIIEFM<br>SGPSCVSGGNGMLRFMDTNASSSDGERRYVLRFKALEVGNCQLLMTG<br>NPIAYEFESGDPMSVSAYSGSVTIRAKEAASKNTNLAILKVSP |

***Clostridium ramosum* DSM 1402 (isolated from human adult as well as infant feces)**

| | | | | |
|---|---|---|---|---|
| **CLORAM_00251** | **1739** | - | 179-231 | **SP, uncharacterized protein,**<br>**Possible dockerin (Pfam: PF00036, efhand; InterPro: IPR016134; Cellulos_enz_dockerin_1): 179..231**<br>G**D**F**N**K**D**GKI**D**SK**D**KKILSTAIFN  NDSSNSL<br>L**D**I**D**D**N**NKI**D**LL**D**LHTFTYAYSD |

***Clostridium scindens* ATCC 35704**

| | | | | |
|---|---|---|---|---|
| **CLOSCI_00154** | **502** | 30-161 | - | **SP, uncharacterized protein,**<br>**PF00963; Cohesin; IPR002102; Cohesin** (very strong orthologs in *Dorea longicatena* and *D. formicigenerans*)**:**<br>AELRFTDPSTTVGAEVEVTAKLTSSSNLQSLDATLTYDTSMLKFISG<br>DSVSGGDGTLNISGSGSGTTLEYKITFQALAEGNAKVEVSDASGTDA<br>SGGALQITKGSSSVTIGPGDPTLIQNEDEGTAAPTGDG |

***Clostridium spiroforme* DSM 1552**

| | | | | |
|---|---|---|---|---|
| **CLOSPI_02309** | **2169** | - | 1134-1191 | **SP, PF01055; GH31; PF02368; Big_2; 2 x CBM32 (PF00754; F5_F8_type_C; 2), PF07554; FIVAR; 4; PF00041; fn3. Putative dockerin (a ’C. perfringens /Bacteroides type’):**<br>G**D**V**N**N**S**GKV**D**NN**D**LTFYENYVGL KPVDNDWEYSTL<br>G**N**I**D**N**N**QII**D**AY**D**ISFVSRMLGD |

*Clostridium* sp. **L2-50**

| | | | | |
|---|---|---|---|---|
| **CLOL250_01290, putative operon** | **677** | - | 614-670 | **SP, COG4193 Beta-N-acetylglucosaminidase, PF08239; SH3_3 Pfam PF00404; Dockerin_1;** strongly similar to several dockerins in *Clostridium cellulolyticum* H10:<br>G**D**V**T**G**D**GVI**N**AL**D**LLKIQKHIIG ASSLKGAFLQA<br>A**N**I**K**R**S**GGL**S**AL**D**LLKVQKFLMG AAKISQK* |
| **CLOL250_01291, putative operon** | **794** | 40-180 | - | **SP, extracellular uncharacterized protein, PF00963; Cohesin; IPR002102; Cohesin:**<br>TQFSISLSSSSVSKGGSVTVTVSVSCSEALGAYSYCLSYDSSVLEYT<br>SGDGYGGGGTISCAGYGDGNIKSASNSFTFSAIASGSSYVGTGSSDV<br>YTWGEESCSVSNAGATITVTAAGSGNNGGDSTTQAATTQEQGSTEAG |
| **CLOL250_01591** | **835** | 221-359 | ND | **SP, uncharacterized protein, PF00149; Metallophos; putative Cohesin**, similar to scaffolding protein CipA cohesins (*Clostridium josui*)<br>TLTFETTVSVPVYTQKNISWTVGKDASEINVTWYADVDGTGTLLVAK<br>NSEVSGNEMPADAKSFTANETASNKSGYYNYQTTATGLSADTTYAYQ<br>LVNGETKSEIRSFTTGGTGAFSFAAAGDPQIGASGSSVNDTDGWE |

***Coprococcus eutactus*** **ATCC 27759, isolated from human feces**

| | | | | |
|---|---|---|---|---|
| **COPEUT_01027** | **474** | 33-170 | - | **SP, extracellular uncharacterized protein, PF00963; Cohesin; IPR002102; Cohesin; stongly similar to cohesins in cellulosomal scaffoldin CipV (*Acetivibrio cellulolyticus*):**<br>SVTFNVNVPSLSVGVGDTVEITVGIDSNEMNISSYSYEVHYDSSLLE<br>SISGGTKLEDGIISYSSNGESGSSMTATFKFKTLASGTASIETTATE<br>IISADGESIEMTPAYNSINISDSAAPASEVASDSPLSDTETYDS |

***Dorea formicigenerans* ATCC 27755**

| | | | | |
|---|---|---|---|---|
| **DORFOR_00931** | **566** | 30-160 | - | **SP, uncharacterized protein,**<br>**PF00963; Cohesin; IPR002102; Cohesin:**<br>GRISFADPSTAVGDMVDVKCVLKSSSGSLGSSSVTLSYDASALKFNS<br>GDGVTGGDGTLTYSGDGGSSEVSFTMTFQALKEGSTEITVASQDVKS<br>SSGSEVKLTEGKSTVTIAAGDPSKIVDDTQAAEGADT |

***Dorea longicatena* DSM 13814**

| | | | | |
|---|---|---|---|---|
| **DORLON_01694** | **566** | 30-163 | - | **SP, uncharacterized protein,**<br>**PF00963; Cohesin; IPR002102; Cohesin:**<br>AELRFADPSTTVGAEVDVKTKLTSASSMQSMEATLTYDKSELRFISG<br>DNATGKDGTIKISWTGSGTTAEFNLKFQALKEGTSNIEVSSATGTAE<br>DGTNLEVTQGSSAVTIGAGDPSLIKEETSTEDTTTSTADG |

***Epulopiscium* sp. 'N.t. morphotype B'**

| | | | | |
|---|---|---|---|---|
| **Epulo_00050** | **6868** | - | 161-205 | **SP, hypothetical protein, Pfam: PF00036, efhand; InterPro: IPR016134; Cellulos_enz_dockerin_1,**<br>**putative 'internal' dockerin**<br>G**D**V**D**G**S**GSV**D**QA**D**RENIVQGLTT   DANKKE<br>L**D**V**T**H**D**GKL**D**IR**D**LAFAFWGSND |

***Eubacterium ventriosum* ATCC 27560 (Clostridia; Clostridiales; Eubacteriaceae)**

| | | | | |
|---|---|---|---|---|
| **EUBVEN_02653** | **2004** | - | 1943-1998 | **SP, IPR015500; Peptidase_S8_subtilisin-rel; PF02368; Big_2; x2, PF00404; Dockerin_1**<br>G**D**V**N**M**D**GKV**T**AV**D**AMLALKLALL   DNPTDAILGL<br>A**D**V**N**G**D**GKI**T**AV**D**AMRILQYATG |

***Ruminococcus gnavus* ATCC 29149 (Clostridia, Clostridiales, Lachnospiraceae)**

| | | | | |
|---|---|---|---|---|
| **RUMGNA_03512** | **516** | 31-168 | - | **SP, uncharacterized protein,**<br>**Possible Cohesin like domain**<br>EGTFQFSDPTAKVGEDVTVKAKISTGGEAIGDGFVTVTYDKAALEFV |

| | | | | |
|---|---|---|---|---|
| | | | | SGTNATGGDGTVKLEATGDGTVSELEYTMVFKALKEGATKLEVSDYT<br>SYLYSDETLNLTTGDSTVTVEAGDGTSSESSAGTTAVATGTGSV |

### *Ruminococcus torques* ATCC 27756 (Clostridia, Clostridiales, Lachnospiraceae)

| | | | | |
|---|---|---|---|---|
| **RUMTOR_00254**<br>or **ORF966**<br>**(Cont52)** | **966** | | 905-961 | **SP, extracellular uncharacterized protein,**<br>**Putative dockerin:**<br>G**D**V**N**E**D**GKV**E**IS**D**LRLILRSVCK   KVELTEQQKLA<br>A**D**V**T**E**D**GKV**E**IG**D**LRKVLRYICH |
| **RUMTOR_00843** | **514** | 29-150 | - | **SP, extracellular uncharacterized protein,**<br>**Putative Cohesin:**<br>DGVLMFSDPSTKVGENVSVDFVVQSSGGTIGSVDVTMSYDPKALEF<br>VSGDGFTADGAGTLTYKGTGGGTELRTAVVFRALTVANTQISVSSS<br>SATISSGDTLNLNNGSSAISIAAADDGTTS |
| **RUMTOR_01757**<br>or **ORF542**<br>**(Cont9.1)** | **542** | - | 481-537 | **uncharacterized protein, ~strongly similar to ORF668,**<br>**Putative dockerin:**<br>G**D**V**N**L**N**GKV**E**IG**D**VREALRSICK   KTQLTETQKQA<br>G**D**I**N**K**N**GTV**D**IE**D**LRTILRVVCG |
| **RUMTOR_01812** | **275** | 27-170 | - | **SP, extracellular uncharacterized protein,**<br>**Pfam motif PF00963; Cohesin:**<br>DVYTEFHVDNQKGKAGDVVTVPVEFNTGQEVGGFQISIYYDKEVLE<br>FQSLEAGDLIEKAGGGIFDYNHIEESSEIIVVYVVPDTVKNEGSIV<br>DLKFKLKKDCTEKLPIGMKPDEVVDNTESSNPITGEVSGVDKEFQE<br>KVVGDL |
| **RUMTOR_02725**<br>or **ORF545**<br>**(Cont29.1)** | **545** | - | 484-540 | **SP, extracellular uncharacterized protein, new CBM?**<br>**Putative dockerin:**<br>G**D**V**D**L**D**GNV**A**IG**D**VRVVLRSICQ   KIELNEMQELA<br>A**D**V**E**K**D**DKV**D**IK**D**LRKILRYVCG |
| **RUMTOR_02789** | **2102** | 1248-1390 | - | **SP**, putative **glycoside hydrolase GH31**, **2x CBM32, FNIII**, **SP**<br>**Pfam motif PF00963; Cohesin:**<br>DSVSGEAMLLASSDSVKKGETFTIDVYAQNVAGLNAFGQVINYDNS<br>KVQYVSCEVSPLLAQMENLTANKEYDDRAYVNLAFANRGDKELYAG<br>SDVLATITMKAKENISTTDAKVIDLSKVTLIGPNYSTIESKVDTNI |

| | | | | |
|---|---|---|---|---|
| | | | | EIPDV |
| **ORF629** | **629** | | | **SP, extracellular uncharacterized protein,**<br>G**D**V**D**A**N**SKV**D**VG**D**VRTALRYICK KTNLTETQMKA<br>G**D**V**T**G**D**EKV**T**IE**D**LRKILRYVCK |
| **ORF641 (Cont46)** | **641** | | 580-636 | **"LRR repeats", ~ORF1030**<br>**Putative dockerin:**<br>G**D**S**D**S**D**GKV**T**IS**D**VRTTLRYVCQ KVELDEEQKLA<br>A**D**V**E**K**D**GVI**N**IK**D**LRKVLRFVCN |
| **ORF668 (Cont9.1)** | | | | **~strongly similar to RUMTOR_01757** or **ORF542**<br>G**D**V**N**L**S**GKV**E**IG**D**VREALRSICK KTELTALQKQA<br>A**D**V**N**E**N**GNV**D**IE**D**LRKILRFVCG |
| **ORF1030 (Cont46)** | **1030** | | 969-1025 | **"LRR repeats"**<br>**Putative dockerin:**<br>G**D**V**D**A**D**RKV**S**IG**D**VRMTLRSICK KAELNGTQKLA<br>A**D**V**E**K**D**GTV**D**IK**D**LRKILRYVCG |

# Deinococcus-Thermus group

| **Gene** | **Protein Size** (no. residues) | **Putative cohesin** | **Putative dockerin** | **Annotation and comments** |
|---|---|---|---|---|

*Deinococcus radiodurans* R1

| | | | | |
|---|---|---|---|---|
| **DR1405/frame shift, then DR1404** | 73 | - | 22-72 | **Hypothetical protein (a putative partial ORF because of a frame shift), an ortholog of Dgeo_1646,**<br>**Putative dockerin:**<br>A**D**L**N**N**D**GVV**D**LA**D**LAVLMENFGK TGALP<br>G**D**L**N**G**D**RRV**D**DA**D**LKLFSRQYTL |

***Deinococcus geothermalis* DSM 11300**

| | | | | |
|---|---|---|---|---|
| **Dgeo_1646** | **178** | - | 128-178 | **SP, uncharacterized protein precursor (an ortholog of DR1405),**<br>**Putative dockerin:**<br>A**D**L**N**G**D**GVI**D**LA**D**LALLMENYGK TGAVL<br>G**D**L**N**Q**D**RKV**D**DA**D**VRLFSAQYQP***** |

# Bacteroidetes

| **Gene** | **Protein Size** (no. residues) | **Putative cohesin** | **Putative dockerin** | **Annotation and comments** |
|---|---|---|---|---|

***Bacteroides caccae* ATCC 43185 (Bacteroidetes (class); Bacteroidales)**

| | | | | |
|---|---|---|---|---|
| **BACCAC_01242** | **1289** | 1158-1289 | 1093-1151 | **SP**, putative **glycoside hydrolase GH31**, **CBM32, FNIII**,<br>**Putative dockerin:**<br>G**D**I**N**V**D**GKI**D**GN**D**LTSYTNYTGL RKCDSDFEGYISN<br>G**D**I**N**K**N**GLI**D**AY**D**ISVVATQLEG<br><br>**Possible cohesin:**<br>TEKVDGTIEISTSKQVYDKDEIIEIQVKGVDLRAVNAFSFALPYNQ<br>QDYEFVGVELQHMNAMENLTYDRLHSNGSKALYPTFVNVGNKKTLE<br>GTVDLFLLKFKAKRQVKFGLKIIDGLLVDKRLNTRRSGVQ |
| **BACCAC_01386** | **1293** | 1165-1293 | 1100-1158 | **SP**, putative **glycoside hydrolase GH31**, **CBM32, FNIII**,<br>**Putative dockerin:**<br>G**D**I**N**N**D**GKI**D**RN**D**LTSYMNYTGL RRGDSDYEGYISK<br>G**D**I**N**M**N**DLI**D**AY**D**ISVVATQLEG<br><br>**Possible cohesin:**<br>TLKVSGSLSISTPKRLYQKDEIVEIRVKGNDLKAVNALSFALPYDQ |

| | | | | |
|---|---|---|---|---|
| | | | | NDFEFVGVEPLNMKAMENLTYDRLHTNGVKSLYPTFVNIGKQEALN<br>GSEELFVLKLKAKRKVKFELTLKDGILVDKELRMHQF |

***Bacteroides fragilis* NCTC 9343**
**See also corresponding orthologs in *Bacteroides fragilis* YCH46**

| | | | | |
|---|---|---|---|---|
| **BF0607** | **1011** | 560-695 | - | **SP, Multidrug efflux membrane fusion protein, Cohesin-like**<br>INTRGAEGVTYEYIRDYTEDINDLVDSIVPDAESVTARVSSGSGNV<br>RITLKDMKDRDYTQMDVAEKLSAAVQKKTMARSFVQQSSSFGGRRG<br>GMPVQYVLQATNIEKLQEVLPKFMAKVYENPVFQMADVDLKFSK |
| **BF1059 (BF1144 in YCH46)** | **1288** | 1165-1288 | 1095-1153 | **SP, putative glycoside hydrolase GH31**, **CBM32, FNIII**, **Dockerin_BF1059**; **67% identity** to the *Enterococcus faecalis* dockerin (**EF_1824**)<br>G**D**I**N**N**D**GKI**D**NN**D**LTSYTNYTGL  RKGDSDYEGYISV<br>G**D**I**D**Q**N**GLI**D**AY**D**ISVVATQLED<br><br>**Cohesin_BF1059**<br>GTIEISTAKRNYSKGDVVEVLVKGVNLRSVNALSFALPYNQQDYEF<br>VGVEPLNLKAMENLTYDRLHTNGTKALYPTFVNLGAKEALEGTNDL<br>FILKLKAKRAVKFDLKAIDGVLVDKNLNTRKF |

***Bacteroides intestinalis* DSM 17393**

| | | | | |
|---|---|---|---|---|
| **BACINT_03859** | **1284** | 1161-1284 | 1091-1149 | **SP, putative glycoside hydrolase GH31**, **CBM32 (FA58C), Putative dockerin:**<br>G**D**I**N**N**D**KWV**D**EN**D**LTSYINYTGL  RRGDADFEGYISN<br>G**D**L**N**R**N**GLI**D**AY**D**ISVVATQLDG<br><br>**Putative cohesin:**<br>GKLDISTPKQTYAKDELVEVLVKGTGLRFVNALSFALAYDPADYEF<br>VGIQPLGMKNMENLTNDRLHTNGTKSLYPTFVNIGEHESLEGESNL<br>FVIKLKAKRKVSFKLTATDGYLVDKLLNVHTF |

***Bacteroides thetaiotaomicron*** **VPI-5482 / ATCC 29148**

| | | | | |
|---|---|---|---|---|
| **BT_3169** | **1294** | 1171-1294 | 1101-1159 | **Alpha-xylosidase GH31, CBM32, FNIII,**<br>**Putative dockerin:**<br>G**D**I**N**N**D**GKI**D**RN**D**LTSYMNYTGL RRGDSDFEGYISK<br>G**D**I**N**M**N**DLI**D**AY**D**ISVVATQLDG<br><br>**Putative cohesin:**<br>GSLSISTPKKQYQKDEIVEIRVKGNDLRSVNALSFALPYDQSDYEF<br>VGVEPLNMKAMENLTYDRLHTNGVKSLYPTFVNMGKQEALEGSEEL<br>FILKLKAKRKVKFDLNLKDGILVDKQLRMHSF |

***Flavobacterium johnsoniae*** **UW101 / ATCC 17061 / DSM 2064 (Bacteroidetes; Flavobacteria)**

| | | | | |
|---|---|---|---|---|
| **Fjoh_0601** | **1857** | 731-866 | - | **SP, hypothetical protein precursor,**<br>**Possible cohesin:**<br>GENGGIGYDSKMYKLDKISLYSKPEYLAQGDNATPIKEANFEYNYS<br>LCKGIDNSINNEGKLTLSKVYFTYKNSNMGKYTPYVFNYGFNPDYD<br>MKAYDVWGNYKKVKEGVNEGTLAQLSNAEFPFVDQSNRTQADEY |
| **Fjoh_3247** | **2324** | - | 2051-2106 | **28 Leucine-rich repeats (PF00560; LRR_1); precursor,**<br>**Putative dockerin:**<br>G**D**S**N**L**D**GIV**N**IL**D**IQSTLNYTLS QKPKFFNYGA<br>A**D**I**N**K**D**KNL**N**VL**D**IISLINKIQS |

***Microscilla marina*** **ATCC 23134 (Bacteroidetes; Sphingobacteria; *Sphingobacteriales; Flexibacteraceae*)**

| | | | | |
|---|---|---|---|---|
| **M23134_05751** | **849** | 307-437,<br>628-766 | - | **Cohesin domain protein; pfam00963; Cohesin domain (similar to the 'last' Cohesin (ScaB-5) of the FD-1 ScaB scaffoldin)**<br>**Cohesin_1**<br>GLIVADHIEAMAGNTVKVPVRAKDLSALAGVQFDINIPSDKASFVG<br>LTNINSKLTDFGTDNYNEVSAGKVRVLWANASMQEQTFQNNEVLFE<br>IELNIPGNATTGDKFFLTFDGSIVVDQEAKTLKVGHKTG<br>**Cohesin_2**<br>VELNTPAVEVINGDVIRIPVAVGANYNQIAGFQGSLEFDPKVLQYQ<br>GVEAAALSINASQHFNTANARQGVISFLYDHPQGEANSFENGQILY |

| | | | | |
|---|---|---|---|---|
| | | | | YLTFKAIGNAGQSAAIKLTSSQTKATAYSQAFKGTSGLILSAGEVK<br>L |
| **M23134_07263** | **899** | 357-473,<br>678-796 | - | **Similar to Strictosidine synthase; pfam00963; Cohesin domain (similar to the 'last' Cohesin (ScaB-5) of the FD-1 ScaB scaffoldin.)**<br>**Cohesin_1**<br>GLIVADHIEAMAGNTVKVPVRAKDLSALAGIQFDINIPSDKASFVG<br>LTNINSKLTDFGTDNYNEVSAGKVRVLWANATMQEQTFQNNEVLFE<br>LELNIPANATTGDKFFLTFDSSIVV<br>**Cohesin_2**<br>VELNTPAVEVINGDVIRVPVVVGANYNQIAGFQGSLEFDPKVLQYQ<br>SVEAAALSINASQHFNAANVRQGVISFLYDHPQGESNSFENGQVLY<br>YLTFKAIGNAGQTAAIKLTSSQTKATA |

***Parabacteroides distasonis* ATCC 8503 (Bacteroidetes; Bacteroidetes (class); Bacteroidales; Porphyromonadaceae)**

| | | | | |
|---|---|---|---|---|
| **BDI_2066** | **1511** | - | 1235-1293 | **SP, PF00560; LRR_1; x2,**<br>**Putative dockerin:**<br>G**D**A**N**F**D**MLV**N**VQ**D**VTATVSYILR DKDNMIPNFGFAE<br>A**D**V**N**Y**N**NNV**E**IA**D**VIGIVNLIHG |

***Parabacteroides merdae* ATCC 43184 (Bacteroidetes; Bacteroidetes (class); Bacteroidales; Porphyromonadaceae)**

| | | | | |
|---|---|---|---|---|
| **PARMER_02202** | **1288** | 1159-1288 | 1096-1153 | **SP, GH31 family protein; 1 CBM32, 1 FNIII,**<br>**An internal dockerin, ‘upstream’ of a cohesin**<br>G**D**I**N**N**D**KKI**D**TN**D**LTSYMNYTGL RRGDSDFDYVSA<br>G**D**I**N**R**N**GLI**D**AY**D**ISVVATQLED<br><br>**Possible cohesin like domain**<br>TDRVAGTIFLSTPKQTYNAGETVEITVKGDSVKAVNALSFALPYDQQ<br>DYDFVGIEPANLGTMENLTYDRLHTSGQKALYPTFVNLGDKQVLEGS<br>EDLFTIKLKTKRKVTFNLKAVDGILVDKNLNMQKF |

***Robiginitalea biformata* HTCC2501* (Bacteroidetes; Flavobacteria)**

| | | | | |
|---|---|---|---|---|
| **RB2501_12909** | **3144** | - | 3083..3138 | **ring canal kelch motif protein (COG3055),**<br>**Putative dockerin:**<br>`G`**D**`I`**D**`G`**N**`NVV`**N**`ID`**D**`FTLLAATFAL GEGDPGYDEG`<br>`S`**D**`L`**N**`G`**D**`GVV`**N**`ID`**D**`FTLLAGNFTV` |

# Planctomycetes

| **Gene** | **Protein Size (no. residues)** | **Putative cohesin** | **Putative dockerin** | **Annotation and comments** |
|---|---|---|---|---|

***Blastopirellula marina*** **DSM 3645**

| | | | | |
|---|---|---|---|---|
| **DSM3645_12601** | **474** | **-** | 251-308 | **Alkaline phosphatase,**<br>**Putative dockerin:**<br>`Y`**D**`L`**N**`D`**D**`QSI`**N**`MI`**D**`LSMMIRALGN STSAPNGNAAWY`<br>`A`**D**`V`**N**`K`**D**`YVV`**S**`LI`**D**`MSYMIRNLGK` |
| **DSM3645_23860** | **2830** | **-** | 2423-2480 | **Fat protein-possibly involved in cell-cell attachment**; cadherin repeat domains,<br>**Putative dockerin:**<br>`Y`**D**`L`**D**`D`**D**`GTI`**S**`LI`**D**`LSQLVRRIGS TSIAANDALTAA`<br>`T`**D**`F`**D**`N`**D**`GRT`**S**`IV`**D**`LSQLIRNIGI` |
| **DSM3645_23875** | **1140** | **-** | 890-943 | **COG0652 Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family,**<br>**Putative dockerin:**<br>`Y`**D**`L`**N**`D`**D**`GQV`**D**`LT`**D**`VVIFRDDYLG NTSADATP`<br>`G`**D**`F`**D**`L`**N**`GVV`**N**`TY`**D**`FSLLIRELGE` |
| **DSM3645_23920** | **799** | 120-245 | 24-91 | **probable surface-associated protein CshA precursor,**<br>**Putative N-terminal dockerin:**<br>`L`**D**`V`**N**`D`**D**`QKV`**D**`NA`**D**`LEIMIAQIQQ GNRSFMQTSQTAGGLEGEQVLF` |

| | | | | |
|---|---|---|---|---|
| | | | | L**D**V**N**N**D**GVF**N**PL**D**ITRMLESLIE<br><br>**Putative cohesin:**<br>SFDGNNDPDAGPAISINVEQSVFTLVVTVRNTSALTGGDLDTPDIT<br>GAFLDIGFDPLVVQPNEPGTLGEYFGPGTVQIGVTSTGLIKNAGGT<br>SAISNGNALPGGWTADQVLYTIDFTPTSTGTFTL |
| **DSM3645_23930** | **622** | **-** | 418-473 | **Alkaline phosphatase,**<br>**Putative dockerin:**<br>Y**D**L**D**D**D**GSI**S**LI**D**LTQVIRKIGS PVTTVNNLYD<br>M**D**F**N**R**N**GSV**D**LV**D**ISLLIRNIGI |
| **DSM3645_23935** | **919** | **-** | 699-763 | **COG0840 Methyl-accepting chemotaxis protein,**<br>**Putative dockerin:**<br>S**D**I**N**D**D**GQV**D**ST**D**FEMVLAELLD PIYSDAHQFGGDHHSPAPF<br>M**D**A**N**D**D**GRF**N**AL**D**VIRVLEAIFV |
| **DSM3645_23945** | **4718** | 2390-2512 | **-** | **COG3210 Large exoproteins involved in heme utilization or adhesion,**<br>**Possible cohesin:**<br>GSPNDGPNNDTNEQRAERDSGTDVRAVGQIRTDFDAVTPGSSTIIV<br>NPVSTFDADGVTVISNYDDGDTFVVDGDGLSTTTGDRITVELNLGF<br>DLSFNALSRQTTETTFQGTTLTRRDGGVFPA |
| **DSM3645_23955** | **1506** | **-** | 1299-1363 | **COG1404 Subtilisin-like serine proteases,**<br>**Putative dockerin:**<br>G**D**V**N**N**D**GFI**T**SH**D**LAIVLREIER FGSRALPQLGTPRAGFAKF<br>I**D**V**N**G**D**NSF**D**PR**D**IVSILEKMVQ |

***Planctomyces maris*** **DSM 8797**

| | | | | |
|---|---|---|---|---|
| **PM8797T_05885** | **5570** | - | 5154-5211 | **PS00142; ZINC_PROTEASE,**<br>**Putative dockerin:**<br>F**D**L**N**D**D**GTI**N**YR**D**LITFTSYYDT SVVGAESGLAWS<br>L**D**F**D**K**S**GKI**N**YR**D**LIFLATNYNR |
| **PM8797T_11741** | **1559** | - | 1316-1373 | **PS00142; ZINC_PROTEASE,**<br>**Putative dockerin:** |

| | | | | |
|---|---|---|---|---|
| | | | | F**D**L**N**D**D**ERI**N**FR**D**LQIFMNAYGS QAATQNNNLSQF<br>I**D**Y**N**Q**N**GRV**D**LN**D**LILLAMKYGT |
| **PM8797T_11756** | **1653** | - | 1407-1464 | **PS00142; ZINC_PROTEASE,**<br>**Putative dockerin:**<br>F**D**L**N**D**D**DLV**N**YR**D**LIQFIRVFGA IPGLSHASHAWA<br>A**D**Y**N**Q**N**GRV**D**FP**D**LLAMVYNFGK |
| **PM8797T_11761** | **1540** | - | 1292-1349 | **PS00142; ZINC_PROTEASE,**<br>**Putative dockerin:**<br>Y**D**L**N**D**D**DKI**D**YR**D**LLLFINVFNS DPRETNSGYAWF<br>A**D**L**D**Q**N**HNV**N**YR**D**LISFVSNYGR |
| **DSM3645_12601** | **474** | - | 251-308 | **Hypothetical protein,**<br>**PF00404; Dockerin_1:**<br>Y**D**L**N**D**D**QSI**N**MI**D**LS**MM**IRALGN STSAPNGNAAWY<br>A**D**V**N**K**D**YVV**S**LI**DM**SY**M**IRNLGK |
| **PM8797T_13947** | **7840** | | 7583-7640 | **PS00142; ZINC_PROTEASE,**<br>**Putative dockerin:**<br>F**D**F**N**D**D**DKI**N**YR**D**LILLAGDYSD IPSESDSQNSWI<br>S**D**L**N**Q**D**NKV**N**YR**D**LILLVGNYGK |
| **PM8797T_21028** | **5844** | - | 5435-5492 | **PS00142; ZINC_PROTEASE,**<br>**Putative dockerin:**<br>Y**D**I**D**D**N**NAI**D**YR**D**LIFFTSAYNQ NVFNATSPYASA<br>L**D**F**D**K**S**GKV**D**YR**D**LIALAGNYGK |
| **PM8797T_21833** | **1565** | - | 1309-1366 | **PS00142; ZINC_PROTEASE, PS00136; SUBTILASE_ASP,**<br>**Putative dockerin:**<br>Y**D**L**N**D**D**GAI**N**YR**D**LIAFVGVYNK SPGSPDADLAWA<br>A**D**F**D**R**S**GKV**D**YR**D**LILMVSNYGK |
| **PM8797T_24191** | **2475** | - | 2046-2103 | **PS00142; ZINC_PROTEASE,**<br>**Putative dockerin:**<br>Y**D**V**N**D**DD**TI**N**YR**D**LIILASVYGE NVLDTNSPYVWA<br>L**D**A**D**K**SG**TV**N**YK**D**LNFFASNYGV |
| **PM8797T_27382** | **1133** | | 874-931 | **PS00142; ZINC_PROTEASE,**<br>**Putative dockerin:** |

| | | | | |
|---|---|---|---|---|
| | | | | F**D**L**N**D**D**DII**N**YR**D**LILFVGLYNT VPSESDSSFAWF<br>S**D**F**N**Q**D**DRI**N**YR**D**LISLVGNYNK |
| **PM8797T_27894** | **960** | | 714-771 | **PS00142; ZINC_PROTEASE,**<br>**Putative dockerin:**<br>F**D**L**N**D**D**DKI**D**YR**D**LIQQISMYNT RPSESNSEYAWF<br>A**D**Y**D**Q**N**DRI**D**YR**D**LVALVKNYGK |
| **PM8797T_31253** | **1421** | - | 1175-1232 | **PS00142; ZINC_PROTEASE,**<br>**Putative dockerin:**<br>F**D**L**N**D**D**DRI**N**YR**D**LILLVGLYNV IPSESDSEYAWF<br>A**D**F**D**Q**S**DRI**D**YR**D**LIALVSNYGK |

***Rhodopirellula baltica*** **SH 1**

| | | | | |
|---|---|---|---|---|
| **RB633** | **831** | **-** | 96-173 | **PF07595; Planc_extracel, Peroxidase,**<br>**Putative dockerin:**<br>E**D**V**N**E**D**GQV**S**AI**D**ALAIINQMNS<br>DGVSGFSSQLQGADGAGSEGQRVNQSRPSRGRM<br>T**D**V**N**N**D**GRD**T**AL**D**ALMVINRLNR |
| **RB844** | **1543** | **-** | 95-155 | **PF07595; Planc_extracel, probable surface-associated protein CshA**; **7 repeatable domains (~85 aa each)**<br>**Putative dockerin:**<br>Y**D**V**D**N**N**KYV**T**AL**D**ALMVINHLNS HGEGEPLSGSDEFTGF<br>V**D**V**T**G**D**NRV**T**AL**D**ALSVINALNR |
| **RB886** | **2009** | 195-328 | - | **PF07595; Planc_extracel, fat protein-possibly involved in cell-cell attachment, 7 BNR** (PF02012), **14 Cadherins**;<br>**PF00963, Cohesin; IPR002102, Cohesin:**<br>VQLPVGGSGTELPAGLPIQLSNADGVTSVTMTIQYDPALLDISGVQLGADAPTGSQVEANLDTPGVATITFFSLEAMDAGQADIIDLIATIPEDAPYGSTGSLTISSLDVNAGAMTASADDAIQVVAFPGDVNA |
| **RB893** | **419** | 39-174 | - | **PF07595; Planc_extracel, probable fibrinogen-binding protein homolog**, **Planc_extracel** (PF07595), **Carb_bd** (IPR008965),<br>**PF00963, Cohesin; IPR002102, Cohesin:**<br>EGRRLLAAVNIPDDLTSAPDAIVSVPVNIDSATGVRAAEIRLSYDTS |

| | | | | |
|---|---|---|---|---|
| | | | | VLDLDADDIDFGSIWGAGSDTQVTANVDDAAGTVVIFVSASSALTDV<br>SGSLVELPFSVASNAVVDSTTVLDLTQVTLNEGQISVDPAPV |
| **RB897** | **225** | 28-160 | - | **PF07595; Planc_extracel, Hypothetical protein, Putative Cohesin, strongly similar to Cohesin_RB893**<br>EGRHLMASVSLDLIGPVPMQESVASIRVEDAPALRAAEIRFEFDPQF<br>VQIEKEDIRPGAIWNNKAALIANVDEESGTVVAYVFSTQPILGSDGN<br>LLDIEIDQSRNAACLTPAKLDLQHVRLNEGTIELESEPV |
| **RB1934** | **1826** | **-** | 1691-1743 | **PF07595; Planc_extracel, Alkaline phosphatase (EC 3.1.3.1),**<br>**Putative dockerin:**<br>T**D**V**N**N**D**GQV**S**AL**D**ALLVINALNQ   PNTQMSRAF<br>L**D**V**N**E**D**GAT**T**AI**D**ALRVINALGD |
| **RB2401** | **706** | **-** | 78-136 | **PF07595; Planc_extracel, Vanadium chloroperoxidase (EC 1.11.1.10)**,<br>**Putative N-terminal dockerin:**<br>E**D**V**N**E**D**DQV**T**AL**D**ALTIINRMNR   EQAGESAGDVRRGQM<br>T**D**V**N**N**D**GRN**S**AL**D**ALMVINRLNR |
| **RB2442** | **1050** | **-** | 870-928 | **PF07595; Planc_extracel,** Hypothetical protein,<br>**Putative dockerin:**<br>H**D**V**N**G**D**GDV**S**SL**D**ALTIINYLNT   YGPGPVGFGDPGFG<br>Y**D**V**N**G**D**GFV**T**AL**D**ALLVINYLNH |
| **RB3075** | **888** | **-** | 745-806 | **PF07595; Planc_extracel,** Hypothetical protein,<br>**Putative dockerin:**<br>E**D**V**N**G**D**GVV**T**AL**D**ALQVINDLNA   NGDRTLTFAETAGFK<br>I**D**V**T**N**D**GNV**S**AL**D**ALRVINKLNE |
| **RB3077** | **6007** | **-** | 5821-5884 | **PF07595; Planc_extracel,** Hypothetical protein,<br>**Putative dockerin:**<br>H**D**V**N**N**S**GAI**T**PL**D**ALLIINAINA   ADGDINLANIPAGITLPQY<br>P**D**V**N**G**S**GTV**S**AL**D**ALNVINRLNE |
| **RB4375** | **3056** | **-** | 2904-2961 | **PF07595; Planc_extracel, Probable extracellular nuclease**,<br>**Putative dockerin:**<br>T**D**T**S**A**D**GET**T**AL**D**ALMVINELNR   LNDAGGEPLLGSF<br>V**D**V**N**R**D**ASV**T**AL**D**ALLVINHLNR |

| | | | | |
|---|---|---|---|---|
| **RB5524** | **7223** | **-** | 7065-7139 | **PF07595; Planc_extracel, Hypothetical protein, peptidase,**<br>**Putative dockerin:**<br>S**D**V**N**R**N**GEV**T**AL**D**ALLIINRLNE<br>VSGSSSGENIPVTDADFGIGTNNGVNEQFY<br>Y**D**Q**S**G**D**GFI**S**SL**D**ALRVINEINA |
| **RB6025** | **1094** | **-** | 48-112 | **PF07595; Planc_extracel,** Hypothetical protein,<br>**Putative N-terminal dockerin:**<br>S**D**V**N**G**D**GEV**T**AL**D**ALVVVNGLNR   PDSSILKTHVPPILGGEGIH<br>Y**D**V**N**G**D**GEV**S**AI**D**SLRVINRLNQ |
| **RB6459** | **2028** | **-** | 1857-1920 | **PF07595; Planc_extracel, Probable aggregation factor core protein MAFp3, isoform E,**<br>**Putative Dockerin:**<br>E**D**V**N**A**N**GEV**T**AL**D**ALTVINRLNS   GESNELSPPNDPNELIGRF<br>Y**D**V**S**G**D**GRV**S**AL**D**ALKVINYLNG |
| **RB7321** | **6157** | **-** | 5953-6024 | **PF07595; Planc_extracel,** Hypothetical protein, **3 PPC (PF04151),**<br>**Putative dockerin:**<br>F**D**V**T**A**D**GFV**T**AI**D**ALRIINLLNR<br>DDLRDPNDPNAELTIDRIPTSPPDY<br>Y**D**V**N**G**D**GRV**S**AL**D**ALQVINELNS |
| **RB7341** | **7538** | **-** | 7351-7430 | **PF07595; Planc_extracel, Similar to surface-associated protein CshA,**<br>**Putative Dockerin:**<br>A**D**V**N**D**D**GEV**S**AL**D**ALLIINHLNR<br>FSSFPAGYTPTVDEDLVVTQGDNGRPSASQPGRY<br>Y**D**V**S**G**D**LRI**S**SL**D**ALQVINELNR |
| **RB9376** | **1827** | **-** | 1724-1786 | **PF07595; Planc_extracel, Probable G-cadherin-possibly involved in cell-cell-attachment,**<br>**Putative dockerin:**<br>Y**D**V**D**R**N**QSV**A**PL**D**ALRIINYLEA   NGPGRLEPGNMQSFPGY<br>L**D**V**S**G**N**GEV**T**SL**D**ALMVINEINR |
| **RB9950** | **736** | - | 555-617 | **PF07595; Planc_extracel, Conserved hypothetical protein**<br>**Putative dockerin:**<br>S**D**V**T**G**D**GEV**S**AL**D**ALRVINALNA   AGASEINIGNEALEQGNY |

| | | | | |
|---|---|---|---|---|
| | | | | YDVTGDGKVTSLDALRVINTLSG |
| **RB10413** | **1352** | **-** | 1248-1310 | **PF07595; Planc_extracel, Hypothetical protein, peptidase S8,**<br>**Putative dockerin:**<br>MDSNRDGSISALDALIIINFLSH SSMSEPIDVGVSNMALT<br>FDANEDGFVTARDALVVINYLTK |
| **RB10416** | **1564** | **-** | 73-132 | **PF07595; Planc_extracel, Similar to xylanase GH10 (EC 3.2.1.8),**<br>**Putative N-terminal dockerin:**<br>LDVNRDGTITELDALQVINVLQA DHTALGAGKSSVEAR<br>HDVNQDGRVSAVDALWVINFVND |
| **RB10423** | **2079** | - | 1987-2048 | **PF07595; Planc_extracel, probable cycloinulo-oligosaccharide fructanotransferase, PKD,**<br>**Putative dockerin:**<br>EDTNADGTVTASDALRVINYMNN ESIAKAEGEQINTGWSR<br>LDVNGDGSVTALDALLVINKLNQ |
| **RB10159** | **709** | - | 581-649 | **PF07595; Planc_extracel,**<br>**Putative dockerin:**<br>FDVTGSENVQSLDALRIINELND RVHSDPVTGVLRDPILVDPYPRNF<br>YDVNRDGKATALDALRVINYLNL |
| **RB11131** | **779** | **-** | 111-165 | **PF07595; Planc_extracel, Peroxinectin, An_peroxidase (PF03098),**<br>**Putative dockerin:**<br>EDVNDDGRVSAMDALLIINAMTR EVSHDEGMF<br>TDVNGDGRRSALDALRVINRIER |
| **RB11133** | **635** | **-** | 297-355 | **PF07595; Planc_extracel, Matrix metalloproteinase 1,**<br>**Putative dockerin:**<br>TDVNEDDRTTALDALMVINRLNQ TSEGESDDVEMDGL<br>CDVNGDGSVTSLDALMVINAMND |
| **RB12697** | **1703** | **-** | 1559-1620 | **PF07595; Planc_extracel,** Hypothetical protein,<br>**Putative Dockerin:**<br>PDVNGDNLVTAIDALLVINELSR LDAAAAGEPGSANTPLFY<br>HDVNGDRRVSALDALRVINYLND |
| **RB12720** | **1901** | **-** | 1743-1803 | **PF07595; Planc_extracel, Probable mannan-binding protein MBP (Lectin),** |

| | | | | |
|---|---|---|---|---|
| | | | | **Putative dockerin:**<br>`HDVDRSGSVSAIDALLVINHLNS EQAQTGTWLSLGGLY`<br>`FDVNGDSRFSPLDALQIINQLRF` |

## Verrucomicrobia

| **Gene** | **Protein Size (no. residues)** | **Putative cohesin** | **Putative dockerin** | **Annotation and comments** |
|---|---|---|---|---|

***Akkermansia muciniphila* ATCC BAA-835 (Verrucomicrobia; Verrucomicrobiae)**

| | | | | |
|---|---|---|---|---|
| **AmucDRAFT_1120** | **1311** | - | 1090-1148 | **SP, GH31 family protein, CBM32, FNIII**<br>**Putative dockerin:**<br>`GDINNDGKIDRNDLTSYINYTGL RKGDSDFEGYISN`<br>`GDINKNGLIDAYDISVVATQLED` |

## Cyanobacteria

| **Gene** | **Protein Size (no. residues)** | **Putative cohesin** | **Putative dockerin** | **Annotation and comments** |
|---|---|---|---|---|

***Anabaena* sp. strain PCC 7120 (Cyanobacteria; Chroococcales)**

| | | | | |
|---|---|---|---|---|
| **all0274** | **826** | 650-775 | - | **Hypothetical protein;** |

| | | | | |
|---|---|---|---|---|
| | | | | **Putative cohesin (31% ID to GMET_1859):**<br>ILPLGVNLQSQAQGELIDLRGVTQLVKADFVVNRDAAFNNFIGFYQ<br>VTDENGGIDTNGDGSADILPGQSGYTQAAIRGRVPGIDLVVNNRGT<br>ATYTGTFQPGSLFAPFIIINSRPERILDNNPNND |

***Gloeobacter violaceus*** **PCC 7421 (Cyanobacteria, Gloeobacteria; a cyanobacterium that lacks thylakoids)**

| | | | | |
|---|---|---|---|---|
| **gll0037** | **398** | - | 34-94 | **Hypothetical protein**;<br>**PF00404; Dockerin_1:**<br>G**D**I**D**G**D**GRV**D**AG**D**IQAIDAYLEG AVVLQDSQIAS<br>A**D**A**D**G**D**GKI**T**RR**D**RDLLKRRVDG |
| **glr0058 operon?** | **364** | - | 29-85 | **Hypothetical protein**;<br>**PF00404; Dockerin_1:**<br>G**D**V**N**S**D**GRV**D**DL**D**IRIVEQYLRG DLLLQNDQIKA<br>A**D**A**D**G**D**GKV**T**RN**D**RDLLKRRLEG |
| **glr0059 operon?** | **363** | - | 25-81 | **Hypothetical protein**;<br>**PF00404; Dockerin_1**<br>G**D**V**N**G**D**GRV**D**AG**D**IRLIDEYLGG TRLLQDEQIKA<br>A**D**A**D**G**D**GQI**T**AT**D**REMLDRRVSG |
| **glr0060 operon** | **377** | - | 23-79 | **Hypothetical protein**;<br>**PF00404; Dockerin_1:**<br>G**D**V**N**G**D**GRV**D**AG**D**IRLIDEYLEG SLLLQDEQIKA<br>A**D**A**D**G**D**GKI**T**AT**D**REMLDRRLSG |
| **gll1450** | **569** | - | 79-148 | **Hypothetical protein, nine FG-GAP (PF01839),**<br>**Putative dockerin:**<br>G**D**V**D**G**D**GDL**D**LV**T**ASGQRVLVLL NTGNGTLANTRFFRNGDSPSSAVL<br>G**D**V**D**G**D**GDL**D**IA**T**ANSGSVYQVT |
| **gll2606** | **361** | - | 28-84 | **Hypothetical protein,**<br>**Putative dockerin:**<br>G**D**A**N**G**D**GRF**D**GK**D**IETIDAYLAG TALLQDEQIRA<br>A**D**G**D**G**D**GRI**T**GA**D**REKLAGRLMG |
| **glr2935** | **346** | - | 76-132 | **Hypothetical protein,**<br>**Putative dockerin:** |

| | | | | |
|---|---|---|---|---|
| | | | | GDIDGNGKTEGADLTLLEQYLDG SGLLVEEQSRR<br>ADITGDGQVDGKDANALRRQLGV |

***Synechococcus* sp. JA-2-3B'a (2-13) or Cyanobacterium Yellowstone B-Prime (Cyanobacteria; Chroococcales)**

| | | | | |
|---|---|---|---|---|
| **CYB_1054** | **369** | - | 35..91 | **Hypothetical protein,**<br>**Putative dockerin:**<br>GDLDGDGRLTVQDLNLLGAYLRG ERGLTDDQIRA<br>ADVDQDGRITEADWQVLQQRIQA |

***Synechococcus* sp. (strain JA-3-3Ab) or Cyanobacterium Yellowstone A-Prime (Cyanobacteria; Chroococcales)**

| | | | | |
|---|---|---|---|---|
| **CYA_2212** | **368** | - | 35..91 | **Hypothetical protein,**<br>**Putative dockerin:**<br>GDLDGDGRLTVQDLNLLGAYLRG ERELTDDQIRA<br>ANVDQDGRITEADWQVLQQRIQA |

# Acidobacteria (a novel phylum)

| **Gene** | **Protein Size (no. residues)** | **Putative cohesin** | **Putative dockerin** | **Annotation and comments** |
|---|---|---|---|---|

**Acidobacteria bacterium strain Ellin345 (Acidobacteriales)**

| | | | | |
|---|---|---|---|---|
| **Acid345_1997** | **764** | 630..764 | - | **SP, Type II and III secretion system protein; Secretin (PF00263), Cohesin, Pfam (PF00963) and InterPro (IPR002102):**<br>VLSFDPATVDVPVGSTVAVNVAIAGGQNVYSVPVQIGYDPKTLSLV<br>NVSNGNFLSRDGQVVALVHRDDPITGQVQVTATRPPGAAGVSGEGQ<br>VFTLTFMAKAPGASQLSIHQAVLRDTAQQKIPATGSQAMITVH |
| **Acid345_4018** | **1275** | - | 223-293 | **SP, Integrin-like protein precursor, five FG-GAP (PF01839), Putative dockerin:** |

| | | | | |
|---|---|---|---|---|
| | | | | G**D**F**N**G**D**GKL**D**FV**V**VSNLANKAQV FLGDGLGSFTQGATLAADSPNAVIA<br>A**D**L**N**G**D**GKL**D**IA**I**LNLETIAIYL |

***Solibacter usitatus*** **Ellin6076 (Solibacteres)**

| | | | | |
|---|---|---|---|---|
| **Acid_1060** | **872** | 734-866 | - | **type II and III secretion system protein. TPR domain, Secretin_C,**<br>**Cohesin, Pfam (PF00963) and InterPro (IPR002102):**<br>VGNATVRFLPPQVETNPQGMMTIALIIENATDVASAPIQVAFDNKV<br>VKLNDVGRGDFFSGDGQIPVFTKNIQNDAGAAAVNLNRLPGTPGSS<br>GSGVLATFIFQGVAKGTSTVTVPNLTVRNAQGQVVFSGSPQ |

# Eubacteria, Betaproteobacteria

| **Gene** | **Protein Size** (no. residues) | **Putative cohesin** | **Putative dockerin** | **Annotation and comments** |
|---|---|---|---|---|

***Ralstonia*** **(formerly *Pseudomonas*) *solanacearum* GMI1000 (*Burkholderiales*)**

| | | | | |
|---|---|---|---|---|
| **RSp0143, megaplasmid Rsp** | **754** | 623-754 | - | **Putative GSPD-related protein;**<br>**Cohesin domain predicted in both, Pfam (PF00963) and InterPro (IPR002102).**<br>SALTIQGPSSVPAGSTFDVLVNVQPERPFTSLPLTLQYNPAQLQII<br>SVTPGSLVSANEVFANNIDAAKGLVYMTMNAKQTQGKDGAGNVVRL<br>TFKALAAGETSIHMQDQMLITPSGMASLKPTAPLVMTVQP |
| **RSp0474, megaplasmid Rsp** | **698** | 564-698 | - | **Putative general secretion pathway GSPD-related protein;**<br>**Cohesin domain predicted in both, Pfam (PF00963) and InterPro (IPR002102).**<br>IALTMNGPSSAKAGEAFWVSLNLRTDTALRSMPIQFAYDKERFAFD<br>GVEAGSLFSGQKTAPNLAKTDLPAVGRVMVTLSAADNAPLTGQGEW |

| | | | | |
|---|---|---|---|---|
| | | | | AKLRFRAKAAGTGAISMATATPVGMTAAPQAPQMPAPVSISVK |

# Eubacteria, Deltaproteobacteria

| **Gene** | **Protein Size** (no. residues) | **Putative cohesin** | **Putative dockerin** | **Annotation and comments** |
|---|---|---|---|---|

***Bdellovibrio bacteriovorus*** **HD100 (Bdellovibrionales), a bacterial predator**

| | | | | |
|---|---|---|---|---|
| **Bd0726** | **1959** | 1100-1230 | - | **Cell wall surface anchor family protein precursor,**<br>**Cohesin-like motif:**<br>PAFSSATSTHTLNIPMAATASVTAGLISKAEYDSFAAKQAAGNYVTA<br>LTGDVTATGPGSVAATIAANAVTSAKIANGTIVGADLDFTGVNTATT<br>SFAMKDSTGKFYNFACATTSHVPTWTATGFACQAPGA |

***Desulfatibacillum alkenivorans*** **AK-01 (Desulfobacterales; Desulfobacteraceae)**

| | | | | |
|---|---|---|---|---|
| **DalkDRAFT_0326** | **1262** | - | 1073-1128 | **SP, conserved hypothetical protein precursor,**<br>**Putative dockerin:**<br>G**D**I**N**G**D**GGV**G**LS**D**AAAAIQIAAR  TSKAYTPGPL<br>S**D**C**D**G**D**GSL**G**LA**D**ALIVMNHEAG |
| **DalkDRAFT_1144** | **812** | - | 239-298 | **SP, conserved hypothetical protein precursor,**<br>**Putative dockerin:**<br>G**D**L**D**D**N**GLV**Q**AK**D**ANIALQILGG  LSPADLVNDISDPA<br>A**D**T**N**G**D**GKF**N**LM**D**ALYILRSVHG |
| **DalkDRAFT_2072** | **2130** | - | 2074-2129 | **SP, Regulator of chromosome condensation, RCC1 precursor,**<br>**Putative dockerin:**<br>G**D**I**D**H**N**NSV**D**LA**D**AVIALQIASG  NVPGEPVFAD<br>A**D**V**N**N**D**NRI**G**TA**E**ACFALQKAAG |
| **DalkDRAFT_227** | **3502** | - | 3444-3500 | **SP, conserved hypothetical protein precursor, PF02368; Big_2;** |

| | | | | |
|---|---|---|---|---|
| **5, operon?** | | | | **PF07603; DUF1566; x6,**<br>**Putative dockerin:**<br>GDLDGNGRIDVSDAIAVLQIMGG EPPSLQDIPCL<br>VDTNGDQHIDLQDAVFIMHQIAN |
| **DalkDRAFT_2276, operon?** | **1389** | - | 1336-1389 | **SP, conserved hypothetical protein precursor, IPR003343; Big_2,**<br>**Putative dockerin:**<br>GDLDGDGLLTIKDMILGLRILCG LEDESGFPQG<br>ADVNEDGALGLEEIVFLLNNM* |
| **DalkDRAFT_3256** | **142** | - | 82-141 | **SP, conserved hypothetical protein precursor,**<br>**Putative dockerin:**<br>GDLNGDCRVSLEDAALALQTAAG MKPSVFRADYTSSG<br>ADPDGAPPMSLADIVWILQAVAE |
| **DalkDRAFT_3431** | **517** | - | 461-515 | **SP, conserved hypothetical protein precursor,**<br>**Putative dockerin:**<br>GDTDQDGVLTLKDALFCLQVQAG ARGVGNFPL<br>YDMDGDGRTGVEEAIIILKKLAG |
| **DalkDRAFT_3730** | **844** | - | 789-843 | **SP, conserved hypothetical protein precursor,**<br>**Putative Dockerin:**<br>GDLNGDSNINLTDAVLVLKILAG RETGALNPN<br>ADVNADSAFGLEEAIYILQTQAG |
| **DalkDRAFT_4144** | **2243** | - | 2189-2243 | **SP, IPR011042; 6-blade_b-propeller_TolB-like, precursor,**<br>**Putative dockerin:**<br>GDLNGDSLVTLEDAILAVQVVSG YDPASGINLS<br>GDANNDGAADMRDALFVLQQLQ* |
| **DalkDRAFT_4743** | **662** | - | 609-662 | **SP, conserved hypothetical protein precursor, IPR006626; PbH1,**<br>**Putative dockerin:**<br>GDVNADGMLSMEDAILAVQDLVD KPYAGKERY<br>ADVDGDGAVSLTEALFVLQSLQ* |
| **DalkDRAFT_4952** | **1225** | - | 1170-1225 | **SP, hypothetical protein precursor,**<br>**Putative dockerin:**<br>GDVNHDGLVAMEDLILAIQATAG LETAAPVYGN<br>ADVNGDGVLGLTDSLYIMREVLQ |

| | | | | |
|---|---|---|---|---|
| **DalkDRAFT_4957** | **463** | - | 406-461 | **SP, hypothetical protein precursor,**<br>**Putative dockerin:**<br>G**D**V**N**D**S**GAV**N**LT**D**AVLALQALAG   METAGEVCDA<br>A**D**F**N**G**D**GMI**G**LA**E**VMYIFQVIRA |
| **DalkDRAFT_5272** | **4045** | - | 61-116 | **SP, PKD domain containing protein, 6-blade_b-propeller_TolB-like (IPR011042),**<br>**Putative dockerin:**<br>G**D**I**N**G**S**GRV**D**GY**D**LISFGLANGS   SSGDANWNPQ<br>A**D**M**D**G**N**GII**D**AA**D**LAMLSTHFGR |

***Geobacter bemidjiensis*** **Bem (Desulfuromonadales)**

| | | | | |
|---|---|---|---|---|
| **GbemDRAFT_2750** | **694** | - | 626-690 | **Fibronectin, type III domain protein precursor,**<br>**Putative dockerin:**<br>G**D**L**S**G**D**NSV**G**VD**D**AQLAMQMGVG   MKKPDAGQLQRGDLAPMVG<br>G**I**P**Q**P**D**GVI**D**TG**D**ALLILGIVTG |
| **GbemDRAFT_3187** | **868** | 734-868 | - | **TPR repeat:type II and III secretion system protein:NolW-like [Precursor],**<br>**Cohesin domain, Pfam (PF00963) and InterPro (IPR002102):**<br>TDSLIALSLPAKVKLNDQFSVQVNGSGMRDLYKAVFVLSYDPKLLDA<br>ISQSEGNLLKQPGKPSTFQAFADKKKGEIWMSGMREEPTGTANGILA<br>NISFKAIGTGSAAVSVNNTNFSKKTGEDIPVTAFKSVVEVQ |
| **GbemDRAFT_3732** | **441** | - | 381-432 | **SP, conserved hypothetical protein,**<br>**Putative dockerin:**<br>V**D**L**D**G**S**GKV**T**EA**D**FEAYLKPVKG   EAAIKR<br>R**D**L**N**G**D**GRL**D**YR**D**DYIFTANYVA |

***Geobacter lovleyi*** **SZ (Desulfuromonadales)**

| | | | | |
|---|---|---|---|---|
| GlovDRAFT_1807 | 940 | 335-460 | ??? | KEGG: **gsu:GSU0968 hypothetical protein,**<br>**Putative cohesin:**<br>FSIRISTSGLTIGQTYSAEVRTTTTSGTVSAPFSLTITGVASIKFYT<br>TDPGTGAKTYHTTLPTVTKQGTVSVGIEAQDSTGQILTTPINISSSN<br>PSILGVYPIYFYGSSIFALNSGSANLIADTLD |

| | | | | |
|---|---|---|---|---|
| **GlovDRAFT _1819** | **795** | 661-795 | - | **TPR repeat:type II and III secretion system protein:NolW-like [Precursor],**<br>**Cohesin domain, Pfam (PF00963) and InterPro (IPR002102):**<br>NTLTINAPSSTVPGQSFKVLVSAEGVSDMKGGSFSLNYTPALLELLS<br>TAQGTLLGKDNGPDAFTTTDDPVKGNVTVTLQQADTATGATGSGTLA<br>GFIFRSKGKGTANLTIGQTRILSATGQFLPLKATSKSITIQ |

***Geobacter metallireducens*** **strain GS-15 / ATCC 53774 / DSM 7210 (Desulfuromonadales)**

| | | | | |
|---|---|---|---|---|
| **Gmet_0556** | **899** | - | 831-895 | **SP, NHL repeat domain protein** precursor: **3 NHL (PF01436),**<br>**Putative dockerin:**<br>G**D**L**N**G**D**GAV**T**VH**D**VILALRMFTR QMVPTANQLAHGDIGPLYQ<br>G**K**A**K**P**N**GVV**D**LV**D**AILIMRKALG |
| **Gmet_0933** | **2636** | - | 2588-2635 | **SP, hypothetical protein. SP, cadherins,**<br>**Putative dockerin:** |
| **Gmet_1635** | **665** | 211-408 | - | **TPR repeat protein** |
| **Gmet_1851** | **551** | - | 491-541 | **Pentapeptide repeat protein,**<br>**Putative dockerin:**<br>V**D**V**D**K**N**GVT**D**EG**D**FSLFLSKNGA AKVPA<br>Y**D**L**N**G**D**GKH**D**YI**D**DYIYTANYLT |
| **Gmet_1859** | **861** | 728-861 | - | **SP, type II and III secretion system protein. TPR, secretin_C,**<br>**Cohesin domain, Pfam (PF00963) and InterPro (IPR002102):**<br>GSLNISAPAAVNLGSQFKVEIRVSDVKGLLKAPFVLIYDPIFIDYVG<br>AVEGTFLNRDGKPTSFNAQVDKAKGRVTITLARTGAEGIDGAGSLLT<br>ATFTAKNKGPASLGLQSVNFTDQANKPIDVVPYNTVVEVK |
| **Gmet_2512** | **188** | 21-160 | - | **protein of unknown function DUF1555,**<br>**Putative cohesin:**<br>EAFSLSLEPPSQTIAVGDTATFALRVTDLADPILFYTLDVLFDASVL<br>AFQGAIFTNALGNPSDSMVDTSLVDTGLLGLVGTSATGSLTGDFDLA<br>YLTFSGITAGTSDLIIGSNTFFGIDPNTGGDIPLNVDSLFDSRATV |

***Geobacter sulfurreducens*** **PCA (Desulfuromonadales); metal reduction in subsurface environments**

| | | | | |
|---|---|---|---|---|
| **GSU0754** | **2192** | 794-940 | - | **Fibronectin type III domain protein, 5 Chlam_PMP (PF02415); 3 FNIII (PF00041)**<br>**Putative cohesin:**<br>NSAGTSSASAASNSVTPTPGPAVISVAVPANGSYGTGSSLDFTVTW<br>DSTATITGTPRIALTIGGTTVYADYASSPTATTSLFRYVVTAGKND<br>TDGITIGALTLNGGSIRNSSGVDATLTLNSVASTTGVLVDTTIPTV<br>SSVTVPGNG |
| **GSU1770** | **452** | - | 392-442 | **Hypothetical protein,**<br>**Putative dockerin:**<br>V**N**V**D**G**K**GRV**T**EA**D**FLLFLARKSV   GPARP<br>F**D**L**N**G**D**GKV**D**YL**D**DYIYTANYLT |
| **GSU1778** | **822** | 688-822 | - | **SP, type II secretion system protein, putative, TRP domain, Cohesin domain, Pfam (PF00963) and InterPro (IPR002102):**<br>GSLNIAAPAGVDLGGQFKVEVKVTDVKGLAKAPFTLLYDPIFIEYV<br>GAAEGNFLNRDGKPTIFNALADKAAGRVVITMDRSAAGEGVDGSGT<br>LLSATFKAKNKGPASLGLQNVKFVDQANRPLDIIPYNTVVEVK |
| **GSU2073** | **1779** | - | 1731-1779 | **EF hand domain/PKD domain protein. Cadherin repeat domain, Dystroglycan-type cadherin-like domains,**<br>**Putative dockerin:**<br>F**D**A**N**G**N**GVV**T**PG**D**FMQILRRGEA   GSIL<br>P**D**S**N**G**N**GRI**D**GH**D**IRTYLGEML***** |
| **GSU2657** | **840** | - | 785-835 | **Spore coat protein A like, a putative multicopper oxidase,**<br>**Putative dockerin:**<br>P**D**V**N**A**D**GKL**A**VT**D**ALDLLKKLKS   PLLAGAP<br>Y**D**L**T**G**D**GTL**D**VR**D**VLALLRTIVF |
| **GSU2746** | **359** | - | 243-292 | **Hypothetical protein,** |
| **GSU2746** | **888** | - | 820-884 | **NHL repeat domain protein: 4 NHL (PF01436),**<br>**Putative dockerin:**<br>G**D**V**N**S**D**GTV**T**IA**D**VLRVLRIAVG   AVAPTADDYAAADVGPLVN<br>G**I**V**R**P**D**GKI**D**LQ**D**AILILRKSIG |

| | | | | |
|---|---|---|---|---|
| **GSU2914** | **898** | - | 830-894 | **NHL repeat domain protein: 4 NHL (PF01436),**<br>**Putative dockerin:**<br>G**D**V**D**G**D**GRI**T**VA**D**ALVALRIFTG QLSPSASHLASGDIGPLYQ<br>G**K**S**R**P**N**GVI**D**LV**D**ALLILRKALG |

***Geobacter uraniireducens*** **(*Geobacter uraniumreducens*) strain Rf4 (Desulfuromonadales)**

| Gene | Protein size | Putative cohesin/s | Putative dockerin/s | Annotation and comments |
|---|---|---|---|---|
| **Gura_0496** | **937** | - | 868-932 | **SP, NHL repeat containing protein precursor,**<br>**Putative dockerin:**<br>G**D**I**N**G**D**GPV**N**IL**D**ARYALQCVVG LATPTAAEILRGDIGPLLN<br>G**K**A**N**P**D**GKI**D**LV**D**AMLILRKALG |
| **Gura_0747** | **1241** | - | 1173-1237 | **SP, Hypothetical protein precursor,**<br>**Putative dockerin:**<br>G**D**I**N**S**D**GKV**D**IS**D**ALMALQMAIG LSASTPTQLTAGDVAPLVN<br>G**K**S**A**P**D**GTI**D**IA**D**AMLILEKAVG |
| **Gura_0756** | **2042** | - | 1974-2038 | **Polymorphic membrane protein** in *Chlamydia* **precursor**; **4 FNIII,**<br>**Putative dockerin:**<br>G**D**I**N**G**D**GKV**D**IA**D**ALLALQMAVG LITPTSAQLATGDVAPLSN<br>G**K**P**A**P**D**GAI**D**IA**D**AMLILEKAVG |
| **Gura_0767** | **1804** | - | 1736-1800 | **SP, uncharacterized protein precursor,**<br>**Putative dockerin:**<br>G**D**I**N**G**D**GKV**D**IS**D**ALLALQMAVG LATPSPAQLAAGDVAPLVN<br>D**K**P**S**S**D**GVI**D**IA**D**AMLILEKAVG |
| **Gura_1161** | **1902** | - | 1834-1898 | **Hypothetical protein,**<br>**Putative dockerin:**<br>G**D**L**N**G**D**RNV**D**IL**D**ALTALQISTG LVAQTDIALINGDVAPLVN<br>G**K**S**V**P**D**GRI**D**VG**D**AVIILKVVVN |
| **Gura_1836** | **906** | - | 837-902 | **SP, NHL repeat domain protein precursor,**<br>**Putative dockerin:**<br>G**D**I**N**S**S**sGKP**D**LV**D**ALKAMRIAMG FDVATSDELLRGDVAPLIS<br>G**V**P**T**P**D**-GKI**D**TG**D**VLVILRKVVG |

| | | | | |
|---|---|---|---|---|
| **Gura_1852** | **184** | 27-166 | - | **SP, uncharacterized protein precursor,**<br>**IPR002102; Cohesin:**<br>TKSATATVTVREVSALSIDQSGPGVFTVEAIKLANVAAIDLTITYDA<br>LSLANPQVTPGAIISGAFMVTNTSVPGRVRCALINTQPISGAGTLMT<br>ISFDLLGASSGRILSMSASFLDNQGRPLATTALIAAPSGTIDMPPG |
| **Gura_1929** | 796 | 662-795 | - | **TPR repeats / Secretin_C,**<br>**Cohesin, Pfam (PF00963) and InterPro (IPR002102).**<br>ITLNVVAPASVKVNEQFSVGIKVADANNLLNAPLTLVYDPLFLDFVG<br>LAEGGFMKQDGKATLFRPTIDKNTGQIAVTLNRVGDVGGISGSGTLL<br>TATFKAKNQGPASIGFLDVNFTAPGGKPLEAIPYNAVVDIK |
| **Gura_2327** | **482** | - | 413-478 | **Fibronectin, type III precursor,**<br>**Putative dockerin:**<br>GDANGDGTVAIADALMVLRAALD PSLQTASIMSTGDVWPLDPS<br>SRPQGDGVIDINDARLILQRAVG |
| **Gura_3430** | **913** | - | 845-909 | **NHL repeat domain protein** precursor: **5 NHL (PF01436),**<br>**Putative dockerin:**<br>GDLDGNGKVDIVDALRALQISAG LLTQTDNDLLHGDVAPLID<br>GKPSPDGVIDVGDALVILRKAVG |

***Geobacter* sp. strain FRC-32 (Desulfuromonadales)**

| | | | | |
|---|---|---|---|---|
| **GeobDRAFT _1296** | **948** | - | 879-943 | **SP, NHL repeat precursor,**<br>**Putative dockerin:**<br>GDLNADSLVNTVDVQLALDFAYG RKVPTQEQLLQGDIGPLLN<br>GFANPNGVIDLSDAILIRKKAVD |
| **GeobDRAFT _2054** | **192** | - | 123-188 | **Fibronectin, type III [Precursor],**<br>**Putative dockerin:**<br>GDGNGDGVVNIADALVILQASLK PALQTQTMKAAGDVWPLDAN<br>GKPKGDGVINLNDARLILQRAVG |
| **GeobDRAFT _3085** | **805** | 670-805 | - | **SP, TPR repeat:type II and III secretion system protein:NolW-like,**<br>**Cohesin domain, Pfam (PF00963) and InterPro (IPR002102):**<br>KVTVNLAAPASVELNSQFSVEIKVANATNLINAPFTFAYDPIFLELI |

| | | | | |
|---|---|---|---|---|
| | | | | RLTEGPFMKKDGKPTTFRPTIDKVTGQVAISLSRPVEVGGVTGSGTLAIATFRAKNQGPASMGFLDTNLTAVGGKPVDALSYSAVVEVK |

***Myxococcus xanthus*** **DK 1622 (Myxococcales)**

| | | | | |
|---|---|---|---|---|
| **MXAN_1733** | **883** | 675-805 | - | **Cysteine-rich repeat protein, 1 PKD**<br>**Cohesin_like (?)**<br>SRGDALYISVSRGSTGVSVLQNFARGANDFSVTDVDQNNAFFDVIVDKISGDVFSVSDDPSFRIRSSSDQGTTFGPQSSPPGQAFFSDWTGSNGFIYVTGSFGDDNVDVIPMSAPGTSTQVPGLPTDIGPA |

***Pelobacter carbinolicus*** **DSM 2380 / Gra Bd 1 (Desulfuromonadales)**

| | | | | |
|---|---|---|---|---|
| **Pcar_0136** | **779** | 650-779 | - | **Putative type II secretion system protein,**<br>**Cohesin domain, Pfam (PF00963) and InterPro (IPR002102):**<br>FLLTAPPTIAVGETFEVLIDITDASDLYSASMNLSCNDNQMEVMEIIQGGFFEKEAHFSSRRDPENRGYVIDFRLPESQVGASGDRNLAKLFCKAINTGDVLLRVAGQNLRDSQGRMIAIPAQEIIINIE |

***Pelobacter propionicus*** **DSM 2379 (Desulfuromonadales)**

| | | | | |
|---|---|---|---|---|
| **PproDRAFT_1704** | **376** | 242-376 | - | **Type II and III secretion system protein,**<br>**Cohesin domain, Pfam (PF00963) and InterPro (IPR002102):**<br>VMLQVGAPPSVASGEQFTLEVKLGPVNDLVKAPFVLGFDPLLVEFVAVNEGEFLKRDGVPTVFGNSVSPAAGSVTVNLGREEGRAGISGEGTLASLLFRAKQKGAAAFSLRDAGFSRSGGGQLELSTSGATVEIQ |

***Stigmatella aurantiaca*** **DW4/3-1 (Myxococcales)**

| | | | | |
|---|---|---|---|---|
| **STIAU_3961** | **1271** | 401-530 | - | **F5/8 type C domain (CBM32) protein,**<br>**Putative Cohesin:**<br>TYTLEKMILGFGRLPPTPAELQVELSVDDIHYHTVFNGTVPPGHAPVLTWSTAGGQGDAKYVRVTFPRTAAPHNVQVLSRLELYGLPVASPAILPGEAPEGSLPVGTLDVPFGGYLSAAIYDSQGKLVRT |

***Syntrophobacter fumaroxidans*** **strain DSM 10017 / MPOB (Syntrophobacterales)**

| | | | | |
|---|---|---|---|---|
| **Sfum_2918, operon?** | **873** | 31-160 | - | **SP, hypothetical protein**<br>**Putative Cohesin:**<br>CFSFPDLRDRPGARLSMWLQASNMDAALAVEATVTYPADKLTAVSVA<br>KTGFSGDMTLVSNLNTPGEIRIAMAGAGGISGNGELFEITFDATSEA<br>GAHAAVQLQGIIVNEVAHPDCITHGSITTTTPAVSI |
| **Sfum_2919, operon?** | **1876** | 1441-1564 | - | **SP, Peptidase S8 and S53, subtilisin, kexin, sedolisin precursor, cell surface receptor IPT/TIG domain**<br>**F00963; Cohesin; IPR002102; Cohesin** (“**eukaryotic type**”, see cohesins in choanoflagellate *Monosiga brevicollis*):<br>ASLSLPNVGGGQYAIVQVPINASVQGLAAADLKVTFNNGVLHARGVS<br>NGSLTPGWSVAAHTGTPGEVRISMASPGGTVAGSGVLVNIEFEVVGT<br>PGGGSALHPVDVSLNDGAIPVQTADGSFTVDAVYGI |

# Eubacteria, Gammaproteobacteria

| **Gene** | **Protein Size (no. residues)** | **Putative cohesin** | **Putative dockerin** | **Annotation and comments** |
|---|---|---|---|---|

*Alteromonas agarilytica* GJ1B (Alteromonadales)

| | | | | |
|---|---|---|---|---|
| ***agaA*** | **1429** | - | 425-480 | **SP, alpha-agarase precursor, 3xCBM6 (PF03422), 5xTSP_3 (PF02412),**<br>**Putative dockerin:**<br>A**D**A**D**N**D**GIE**D**SQ**D**FCPNTPAGEA VNNSG**C**GASQ<br>L**D**A**D**N**D**GVT**N**NI**D**QCPNTPAGTQ |

*Alteromonas macleodii* 'Deep ecotype' (Alteromonadales)

| | | | | |
|---|---|---|---|---|
| **MADE_02381** | **1346** | - | 1289-1341 | **Extracellular ribonuclease/nuclease fusion protein**<br>**Putative dockerin:** |

| | | | | |
|---|---|---|---|---|
| | | | | G**D**F**D**G**D**NDV**D**YN**D**IMAFYQLFLQ GQATAES<br>Y**D**F**N**G**D**GQI**S**FF**D**IQALMAMCTR |

*Colwellia psychrerythraea* **34H (Alteromonadales)**

| | | | | |
|---|---|---|---|---|
| **CPS_2634** | **1310** | - | 1252-1304 | **Extracellular ribonuclease/nuclease fusion protein,**<br>**Dockerin:**<br>G**D**F**D**G**D**ADI**D**RN**D**VTQFSMMVRR GETGDIQ<br>H**D**F**N**N**D**SVV**N**SL**D**VRALMSLCTR |

***Hahella chejuensis*** **KCTC 2396 (Oceanospirillales)**

| | | | | |
|---|---|---|---|---|
| **HCH_01320**<br>**operon?** | **224** | 38-196 | **-** | **SP, hypothetical protein**<br>**Putative cohesin:**<br>VITVLPSTSTVLPGQLFTVDISISGLAPGGAPSLSTFDLNLSFAPG<br>AVAIDTTDADANGVIDSVALDSDSQLDVLGLGLNPMSAELLAPGEL<br>NLFDLSLDIPADLDAHQSNSFLLASLSFTALSPGFSYFSVVVNALG<br>DSLGNPIAATVNNGGVNVVSE |
| **HCH_01321**<br>**operon?** | **1145** | - | 1086-1139 | **SP, hypothetical protein,**<br>**Putative dockerin:**<br>G**D**I**D**G**D**GDV**D**RD**D**LDEVISARNT PASGSDDP<br>R**D**L**D**G**D**GMI**T**SL**D**ARQLRLLCTR |
| **HCH_02113** | **429** | - | 219-268 | **hypothetical protein** |
| **HCH_06114** | **1361** | | 100-157 | **RTX toxins and related Ca2+-binding protein,**<br>**Putative dockerin:** |
| **HCH_06552** | **835** | - | 784-837 | **SP, Predicted extracellular nuclease,**<br>**Endonuclease/Exonuclease/phosphatase family,**<br>**Putative dockerin:**<br>G**D**F**D**G**D**GDV**D**GR**D**LLMLLKQLFR SVNDDNRQ<br>Y**D**L**N**E**D**GRI**S**FL**D**VFAFLRLL* |
| **HCH_06681**<br>**operon?** | **187** | 20-181 | **-** | **hypothetical protein,**<br>**Cohesin like:** |

| | | | | |
|---|---|---|---|---|
| **HCH_06682<br>operon?** | **885** | - | 828-881 | **SP, putative predicted phosphohydrolase precursor, FNIII, x1, Putative dockerin:**<br>G**D**L**N**G**D**GVA**D**GA**D**RDLLRSMFGK CNGDAGYQ<br>A**D**A**D**Y**D**GDG**C**IG**R**QDYRLWYAMY |

*Legionella pneumophila* strain Philadelphia 1 (Legionellales)

| | | | | |
|---|---|---|---|---|
| **lpg0898** | **243** | 90-210 | - | **Hypothetical protein [intracellularly induced gene]<br>Putative cohesin:**<br>VSAPTLSSKKKTEANTTLVGSTGIMQKLLGNIGVNSEQSIEVLSKH<br>FNGAVERFTQDVSKAANEAVCEKIKSIQPEDVVKGINSLIGNTLKV<br>DSESGDTTLSFKLESAVQILGVILAAMGM |

*Marinobacter aquaeolei* ATCC 700491 (DSM 11845 / VT8) (Alteromonadales)

| | | | | |
|---|---|---|---|---|
| **Maqu_0358** | **826** | - | 773-825 | **SP; putative extracellular nuclease; Exo_endo_phos (PF03372), PF00404; Dockerin_1:**<br>A**D**L**N**G**D**GRV**N**GR**D**LSRLVLAMLF GKATAPA<br>Y**D**I**D**G**D**GRI**N**HY**D**LIALIRVLTS |

*Pseudoalteromonas atlantica* T6c (Alteromonadales)

| | | | | |
|---|---|---|---|---|
| **Patl_1033<br>(in operon)** | **1503** | - | 1446-1498 | **SP, Endonuclease I,<br>Putative dockerin:**<br>G**D**Y**D**G**D**GDV**D**MM**D**IRALTRAIQL RQTIDNS<br>F**D**F**N**E**D**GQV**S**YT**D**VRLLQRMCTR |
| **Patl_1034<br>(in operon)** | **183** | 21-162 | **-** | **SP; PEP motif putative anchor-like,<br>Cohesin, Pfam (PF00963), InterPro (IPR002102) [~Patl_2342]:**<br>LDNVNPSVGDTVSININLTGVTEDFSAIFTGLNYDSNLFAYVDGSA<br>SSDFPPFDPFTASGLDVDTTGPGLVSINLYEDFFAPVVYGAGDYLV<br>ASFDLLAIGAGQSSFSLVASSLFAPFPGDETALDVSASATSIAVPA<br>PGAMAFMGLAVLGL |
| **Patl_1035<br>(in operon)** | **1474** | - | 1417-1466 | **SP, peptidase S8 and S53, subtilisin, kexin, sedolisin (or proteinase inhibitor?).** |

| | | | | |
|---|---|---|---|---|
| | | | | **Putative dockerin:**<br>G**D**W**D**G**D**GDV**D**MM**D**IRALTRAIQL  RQDIDMS<br>F**D**F**N**D**D**GQV**T**YT**D**VRLLQRMCT |
| **Patl_1501**<br>**(in operon)** | **1376** | - | 1316-1367 | **SP, PKD domain containing protein, peptidase-like,**<br>**Putative dockerin:**<br>C**D**V**D**G**N**DFV**D**SL**D**IRAIARLRNA  TASEGDV<br>A**D**W**D**Q**N**GVI**N**VL**D**ARGCALECD |
| **Patl_1502**<br>**(in operon)** | **203** | 35-176 | **-** | **SP, Hypothetical protein**<br>**Putative cohesin:**<br>FNNTQSNVNETVNIGDTIIFELWFSGLENVDVGGFEFLLGFNDAVS<br>SLNSAVGNASLDEFDIFDLAGNSNHVDAYAVSFTADLSTQANEFMF<br>ATLSFMASGAGVSNLTFNNLIVSDLNALAMDISVFDAQITVVDDTN<br>NMPVPTPTSWGIFLIGVLAL |
| **Patl_2341**<br>**(in operon)** | **1171** | - | 1114-1165 | **SP**, **GH5 Agarase precursor (EC 3.2.1.81)**; **Bacterial Ig-like domain 2,**<br>**Putative dockerin:**<br>G**D**Y**D**L**D**GDV**D**IN**D**IQSLIVAIHM  RQSIDLS<br>F**D**M**N**S**D**GTV**N**LL**D**TRLLMNACT |
| **Patl_2342**<br>**(in operon)** | **175** | 25-170 | **-** | **SP, hypothetical protein precursor,**<br>**Putative cohesin (~Patl_1034):**<br>VSSPELNIGEALSVTVSATNFDP**FS**TLGFEIEFDTDLFAFAPNSVG<br>GDLANTGAFIFEVSAQTYGVAMSFIDFEDFTQPNFSAFTFNLTAIA<br>SGSTDFTLANIQSFMPIDSAALNASAQVNSVPAAGTFSIFLFTAFC |
| **Patl_2757**<br>**(in operon)** | **623** | - | 565-615 | **SP, KEGG: lxx:Lxx11970 hemagglutinin/hemolysin-related protein,**<br>**Putative dockerin:**<br>C**D**I**D**G**D**MNV**D**MD**D**LMMIRAARNM  TALPGDL<br>R**D**N**D**A**N**GVI**N**IL**D**FRQCAVQCT |
| **Patl_2758**<br>**(in operon)** | **195** | 25-165 | - | **SP, PEP motif putative anchor-like,**<br>**Putative cohesin:**<br>AVITMTPSSQTIELGETANVDVSISDFATDQYLGAYDFEVAFNDSI<br>LSVSNIVFGTELGFSFQDEFSFGANLHAVLESSLEDAQYLVDNQAS |

| | | | | |
|---|---|---|---|---|
| | | | | EFLLFSIEFTAVGFGTSSVAFDSSLLGDQDGNEITDVDFNSARITVNNP |

*Shewanella baltica* OS195 (Alteromonadales)

| | | | | |
|---|---|---|---|---|
| **Sbal195_2031** | **2402** | - | 600-668 | **SP, YD repeat protein precursor, FN3 etc, Putative dockerin**<br>G**D**F**D**G**D**GSL**D**FP**E**YNTNAEGEYT KGNSNRLDSCILDTSLQMFHCVS<br>P**D**I**D**N**D**GRT**D**RL**S**VADDETLVVN |

*Shewanella woodyi* ATCC 51908 (Alteromonadales)

| | | | | |
|---|---|---|---|---|
| **Swoo_0198** | **3333** | - | 1517-1585 | **SP, YD repeat protein precursor, putative peptidase; 3xFN3 etc, Putative dockerin**<br>G**D**I**N**G**D**GTL**D**WP**N**VLASAEMELI SEHQKPVATCHYTYLSASNTCIA<br>A**D**I**D**A**D**GKV**D**LH**Q**VSAGTLHFYL |

*Shewanella* sp. MR-7 (Alteromonadales)

| | | | | |
|---|---|---|---|---|
| **Shewmr7_1124** | **2401** | - | 596-666 | **SP, YD repeat protein precursor, six RHS_repeats (PF05593), Putative dockerin:**<br>G**D**A**N**G**D**GST**D**FP**G**YDSNAEQDML GQNGFTHEKACYNDYVSDGWICVA<br>F**D**A**D**H**D**GKV**D**PF**K**SINNILHIQL |
| **Shewmr7_1180** | **2437** | - | 596-666 | **SP, YD repeat protein precursor, six RHS_repeats (PF05593), Putative dockerin:**<br>G**D**A**N**G**D**GST**D**FP**G**YDSNAEQDML GQNGFTHEKACYNDYVSDGWICVA<br>F**D**A**D**H**D**GKV**D**PF**K**SINNILHIQL |
| **Shewmr7_3898** | **2283** | - | 473-543 | **SP, YD repeat protein precursor, six RHS_repeats (PF05593), Putative dockerin:**<br>G**D**A**N**G**D**GST**D**FP**G**YDSNAEQDML GQNGFTHEKACYNDYVSDGWICVA<br>F**D**A**D**H**D**GKV**D**PF**K**SINNILHIQL |

# Candidate division WWE1

| **Gene** | **Protein Size** (no. residues) | **Putative cohesin** | **Putative dockerin** | **Annotation and comments** |
|---|---|---|---|---|
| **Candidatus Cloacamonas** | | | | |
| **CLOAM1811** | **959** | - | 718-786 | **SP, putative zinc-carboxypeptidase D (Metallocarboxypeptidase D), EC 3.4.17.22,**<br>**Putative dockerin**<br>A**D**V**D**G**N**GTI**D**AF**D**SYLLCKYIYV PDYSLPVQTGIPEEVPVPVLTAS<br>Y**N**N**N**L**N**LNF**S**NI**D**ELKSLTVSVA |

# EUKARYOTA

## Fungi/Metazoa group, Choanoflagellida

| **Gene** | **Protein Size** (no. residues) | **Putative cohesin** | **Putative dockerin** | **Annotation and comments** |
|---|---|---|---|---|
| **Choanoflagellate *Monosiga brevicollis* MX1 / ATCC 50154** | | | | |
| **MBCDH16**<br>**ORF Name: 11672** | **5741** | 3905-4040 | 4083-4145 | **IPR000152; Asx_hydroxyl_S,**<br>**Putative cohesin:**<br>GDIDLGNSATPAISILGPGEVAWISVFANVSGVMMRGFDLTVTSAS<br>SSTVSLEGAVLGRDVTGQFMTAQQTASRVRFGAVVAAVTGGRLEIA<br>RVQLRAGSSASAAATTISVTVNDLFDQDLAIIGAPTPRLAIAGS<br><br>**Putative dockerin ("Planctomycetes" type):**<br>G**D**V**N**G**D**CVF**T**LR**D**VAFVNDYAAD VLAGIPPPNVTAEVYAE<br>M**D**V**D**R**D**NTI**T**VG**D**SALLADILYG |
| **MBCDH18**<br>**ORF Name: 12200** | **10,110** | 6959-7101,<br>8732-8876 | 7190-7256,<br>8944-9013 | **PF00028; Cadherin; x 58; IPR008162; Pyrophosphatase,**<br>**1st putative cohesin:**<br>NDADLGNLNGIPLPPTAVGAQLNVRVALNTGSVAFGAFTVNVYYDE<br>TKLQAIDDGFILGNDWEDGPFSARTNNPPGEYQVTGVAGGSSDRGT<br>NYHLVTLPFTVIGSGLTQLTGVIETLNRADSATAVPIGTPNRQMVA<br>GVVNV<br><br>**1st putative dockerin ("Planctomycetes" type):**<br>G**D**A**D**G**N**CVF**D**IA**D**VAYTLQYIVE ALVDFTGSFGAAFDDITERQIKD<br>M**D**A**D**R**N**GVV**D**GT**D**AFYMVKALVL |

| | | | | |
|---|---|---|---|---|
| | | | | **2nd putative cohesin:**<br>GDVDLGRQGNFPLADVLVGESFDVPVYINTGTLPLGAIDLQIVYSS<br>AHLEIPTTSAGVAVRRGRDWPGGSVYAVVDPPGTLSFGGVTAVDST<br>AKGELLEIAVVTFRALTATAGSIPIQGHVVTMASNDLQGTFIGSAE<br>ERDFVAG<br><br>**2nd putative dockerin (“Planctomycetes“ type):**<br>G**D**T**N**G**D**CVF**D**IR**D**VGYLQTYLAE RQLDFARAQGQILNASLIPVQLSN<br>M**D**A**N**L**D**SVI**T**VA**D**AEFLARVNFG |
| **ORF Name: 23706** | **3677** | 2722-2859 | 2887-2947 | **PF00595; PDZ; 1; PF00020; TNFR_c6; 14.**<br>**Pfam PF00963; Cohesin; IPR002102; Cohesin:**<br>GDIDLGASSGAALGEDWGLGGSLTIPVRVNAGLTGLAGLDLTIEYD<br>PAVLAIQAVEPGSDWVGNSFIFTANDPPGQLRFGGTTTTNDGIVEV<br>AVITFSIMGTETDPALHGYATSVSDAERRGIGSPTPRAFVAGDLRV<br><br>**Putative Dockerin (“Planctomycetes“ type):**<br>G**D**A**D**A**D**CVV**G**VN**D**VLYLQDYLVQ AQVGPTAAPPLEPGQ<br>L**D**V**D**L**S**GDV**T**IL**D**AVHLLRVVFG |
| **MBCDH8**<br>**ORF Name: 27264** | **8735** | 3504-3645,<br>5289-5430 | 3732-3803,<br>5517-5588 | **PF00028; Cadherin; 21 (to 25?); PF00963; Cohesin; x2.**<br>**Pfam PF00963; Cohesin; IPR002102; Cohesin 1:**<br>FDVDLGAATGLPLSARAIGDTFTVPVRVNAGSSTLASIDLSVYYDP<br>SRLEAVSVAPGSAWPSSAQLTATLDDPPGVVLFGGALDTTGLQGTK<br>EIAVLTFRVLAGATVGQRLALTGQVTTMADTDGVLMGAENRAFVAG<br>SVEM<br><br>**1st putative Dockerin (“Planctomycetes“ type):**<br>G**D**A**D**G**N**CLF**D**VR**D**VRYTQDYIVY ASLGFATPAGAAFRADMLSSGEQEAN<br>M**D**A**D**L**N**GVI**D**GR**D**SSFLARVNFR<br>**Pfam PF00963; Cohesin; IPR002102; Cohesin 2:**<br>FDVDLGAATGLPLSARAIGDTFTVPVRVNAGSSTLASIDLSVYYDP<br>SRLEAVSVAPGSAWPSSAQLTATLDDPPGVVLFGGALDTTGLQGTK<br>EIAVLTFRVLAGATVGQRLALTGQVTTMADTDGVLMGAENRAFVAG<br>SVEM |

| | | | | |
|---|---|---|---|---|
| | | | | **2nd putative dockerin (“Planctomycetes“ type):**<br>G**D**A**D**G**N**CLF**D**VR**D**VRYTQDYIVY ASLGFATPAGAAFRADMLSSGEQEAN<br>M**D**A**D**L**N**GVI**D**GR**D**SSFLARVNFR |
| **ORF Name: 27653** | **5165** | 2577-2714,<br>4160..4297 | 2751-2822,<br>4334-4405 | **PF00023; Ank; x3,**<br>**1st putative cohesin:**<br>GDVDLGADAGLPVPASQSVGTSFTLGVYVNTGGASAGAYAFEVLFD<br>PEALAATEAVASVDGIFSARFDENRVVFGGTVTAGAVQGTRVQVAS<br>LTFEVVGSGTTTVSGVVTTLSQSSVDAADIGLETPRSFVAGRIEIQ<br><br>**1st putative dockerin (“Planctomycetes“ type):**<br>G**D**V**N**G**D**GIL**S**VR**D**TQFLLQYYNN RFSNFVGAQGAQILAALSGNNYSQGE<br>L**D**A**D**Q**N**GEI**N**PR**D**AIYVNQVVFG<br><br>**2nd putative cohesin:**<br>GDVDLGADAGLPVPASQSVGTSFTLGVYVNTGGASAGAYAFEVLFD<br>PEALAATEAVASVDGIFSARFDENRVVFGGTVTAGAVQGTRVQVAS<br>LTFEVVGSGTTTVSGVVTTLSQSSVDAADIGLETPRSFVAGRIEIQ<br><br>**2nd putative dockerin (“Planctomycetes“ type):**<br>G**D**V**N**G**D**GIL**S**VR**D**TQFLLQYYNN RFSNFVGAQGAQILAALSGNNYSQGE<br>L**D**A**D**Q**N**GEI**N**PR**D**AIYVNQVVFG |
| **ORF Name: 27688** | **2775** | 1683-1819 | 1853-1914 | **IPR006212; Furin_repeat;**<br>**Putative cohesin:**<br>GLADIGAQSGIPLPTLSVGDVVDVPLLLHFGAVAVEGAAASLSIFY<br>NATAWTVLDVVAAPAFAGIVAANGRSTGVVNATAILQGSTGVSGLQ<br>PVLLLTLRAEVAGDAALSVRVLSFTNQDQAPIGDAVLPALAPAAA<br><br>**Putative dockerin (“Planctomycetes“ type):**<br>G**D**A**D**G**N**CCL**T**PL**D**VLYQLDYIVL AELGTPLAVEAFQLAG<br>L**D**A**D**L**D**GVV**A**LS**D**ALLTLRAQAG |
| **MBCDH11** | **5417** | 4275-4410 | 4445-4508 | **PF00028; Cadherin; x2,** |

| | | | | |
|---|---|---|---|---|
| **ORF Name: 28599** | | | | **Putative cohesin, a ‘very weak’ prediction:**<br>HDADLGTANGLPLGQPQVGQSQVADLRINVGTESLVEARLRVTFSE<br>HVRVLSVSRAAATVDGTFTTVVGSTGEVTIIAAPASSATGVVSLGG<br>LTLAMDSAGDASVLVSILRLETSAGVVVPSSAPVDAVAASFDFT<br>**Putative dockerin (“Planctomycetes“ type):**<br>G**D**V**D**G**N**CIF**D**AS**D**ITYIQTVALA TLLDPAFPDTLLQNQQAN<br>L**D**A**D**G**D**GSF**D**MS**D**ADYLLRVYSR |
| **ORF Name: 28603** | **4527** | 3462-3598 | 3640-3710 | **PF01302; CAP_GLY; PF00051; Kringle;**<br>**Putative cohesin:**<br>GDVDAGLADIEGSSVIEAQPGDTVQLAFSANVGAQDLGAFALSVQV<br>PVGMAVVRSVDSLYAEGRFSSNVGESSVSVAGLLEGAATGSAVPLF<br>TITLQLLAVPTNPLNVDIQTLKDKSVPAATIGATTPRAAVAATPV<br>**Putative dockerin (“Planctomycetes“ type):**<br>F**D**V**D**G**DC**VF**D**VA**D**AQLTLFYAVE ALTEFASPRGQQILQLLANNSNAGRG<br>L**D**A**N**G**D**GMV**T**IR**D**VAAQLNALAG |
| **ORF Name: 29619** | **4310** | 3281-3417 | 3461-3520 | **Poly Thr linkers;**<br>**Pfam PF00963; Cohesin; IPR002102; Cohesin:**<br>LDADLGAITGLALGTLNVGQEIEVPVRIHMGARALRAMQIAITYDA<br>TALSPVRVLKGSNWATGQMLANLDGYAGRVEVGGTLTTTAGVTLPE<br>VVRVVFEVLDTAPATSYIGGTLIRLYAGVNLDEVVTDAPIIAGHL<br>**Putative Dockerin (“Planctomycetes“ type):**<br>G**D**Y**D**G**D**CSV**T**IK**D**VATGLQTLLA SDSSRIVVTSATSH<br>M**D**L**D**H**D**GYL**T**VA**D**VDYLSKAVFG |
| **MBCDH21**<br>**ORF Name: 30335** | **10056** | 6453-6587 | 6653-6719 | **PF00028; Cadherin; 40; PF00053; Laminin_EGF; 4 etc**<br>**Putative cohesin:**<br>GDVDLGDVTGHYLQNLRKQQNKVVPVWVAVGGASLGNLTITLTYNS<br>QRMQVVGTAIGPGFQGTIAVPSTGRLDLDLQVLNPMNVDASAAGLL<br>RVAEITVFPTSTRDITFGGTVQRATSTEGAPLGSPTPRTIVAG |

| | | | | |
|---|---|---|---|---|
| | | | | **Putative dockerin ("Planctomycetes" type):**<br>G**D**V**N**G**D**CRK**S**DA**D**VDLLALYLEH ASYNFRLPESPAVLASVYEPE<br>L**D**V**D**R**D**GRV**T**LA**D**LNALRDLAVQ |
| **ORF Name: 39394** | **>1224** | 176-310 | 350-421 | **Predicted protein [Fragment]; TSS-poly linker (~150 aa)**<br>**Putative cohesin:**<br>GDVDLGPISGSLTLQVPIGPSALTLPVRVNTNGADVGAFDLRLTWT<br>AGSLSIGGVTKSYAEGEIAFNVDGRELSVSGVVQSTLQGATVELFQ<br>VTLSQDTPAAQLAILSGQVDSLKDKDVPAQDIGSSTPRTFVAG<br>**Putative dockerin ("Planctomycetes" type):**<br>L**D**A**D**G**T**CTF**D**AG**D**SYFTLLYTIT APSNFVSATGQQVAAVLAANAELEGR<br>L**D**A**D**R**N**GVV**T**GT**D**LLFQLQILVG |

# Fungi/Metazoa group, Metazoa, Placozoa

***Trichoplax adhaerens*** **Grell-BS-1999**

| | | | | |
|---|---|---|---|---|
| **TRIADDRAFT_62730** | **271** | - | 38-91 | **Hypothetical protein, fragment**<br>**Putative dockerin:**<br>G**D**S**N**G**D**CSF**N**IM**D**AAYIKIESSK ALQPLNVS<br>L**D**Y**D**H**N**QVV**D**YA**D**ALFLVKVFYN |

# Alveolata; Ciliophora; Intramacronucleata

**Ciliate** ***Tetrahymena thermophila*** **SB210, a model eukaryote**

| | | | | |
|---|---|---|---|---|
| **TTHERM_00388560** | **1029** | - | 919..985 | **Hypothetical protein,**<br>**Putative dockerin:**<br>Y**D**V**N**G**D**GFL**D**KN**E**VARMFEVAGR LRGISFSIDQLKNNTEEFFKL<br>A**D**Q**N**N**D**GQI**S**YS**E**FRIAIQKRPE |