

C. watsonii FG-GAP propeller motifs

Extended motifs

>Q4BZF5		
20	SGISVSGIGDFNGDNISDLVIGSHQADPNGTS-SGQAHVV	FG----- 59
>Q4BYZ3		
306	SGRSVSSAGDVNGDGFSDILIGTYHADPNGNSRAGESYVV	FG----- 347
>Q4BY77		
2	-----SGAGDVNGDGFDDLIIGARSADPNGIGQAGESYVV	FGKAGVFSPSFDLSSLNGSNGFVLNGIDSVDF 68
69	SGVSVSGAGDVNGDGFDDLIIGANGADPNGNEQAGESYVV	FG----- 110
>Q4C0A4		
16	---SMSGAGDINGDGFDDILIGASGADPNGNRSAGESYVI	FGSASGFSGSLDLSSLDGSNGFVLNGIDEFDG 84
85	SGSSVSGAGDINGDGFDDIIIGAI FADPNGNESAGESYVI	FGS----- 127
>Q4C0A5		
2	-----SGAGDINGDGFDDIIIGARTADPNGNERAGESYVV	FGAVGGFSGSLDLSSLDGNGFVINGIDEYDF 68
69	SGESVSGAGDINGDGFDDILIGAANA-PNGIYDAGESYVV	FGSASGFSASLDLSSLDGSNGFVINGIDEFDF 139
140	SGESVSGAGDINGDGFDDILIGASSADPNGNYDAGESYVV	FGAASGFSASLELSSLDGSNGFVLNGIDEDDY 211
212	SGSSVSGAGDINGDGFDDILIGASSADPNGNERAGDSYVV	FGSASGFSASLELSSLDGSNGFVLNGIDEDDF 283
284	SGNSVSGAGDINGDGFDDIIIGALGASPNGNRSAGESYVV	FGAVGGFTASLELSSLDGSNGFVLNGIDQGDN 355
356	LGVSVSGAGDINGDGFDDIIIGALGASPNGNRLAGESYVI	FGS----- 398
>Q4BWI7		
75	FGRSVSNAGDVNGDGIDDLIIIGAEGADPNGNSYAGETYVV	FGSES-FGDTLNLSSLNGSNGFVINGIDARDY 145
146	SGRSVSNAGDINGDGIDDIIIGAF LADRNGNSLLGQTYVV	FGSES-FDSTLNLSSLDGSNGFVISGIDARDY 216
217	SGRSVSNAGDVNGDGIDDFIIGAYGAEPNGNRFAGETYVI	FGSES-FGNTLNLSSLDGNGFVINGIDEFDN 287
288	SGFSVSNAGDINGDGFDDLIIGAF LAQPNGNIDAGETYVV	FGSSS-FEDTLNLSSLDGSNGFVINGIDEFDN 358
359	SGFSVSNAGDINGDGFDDLIIGAFFADPNGNEDAGETYVI	FGSES-FGSTLNLSSLNGNGFVINGIDAGDN 429
430	SGFSVSNAGDVNGDGIDDLIIIGADPAEANGNSYAGQTYVI	FGGQNNNSNSL-LS----- 482

Frankia sp. strain EAN1pec, WD40 propeller motifs

>A8LD20
---SLS----GSPLRGHTDS----VRSVAFSPDGRTLASPSQDRTV-----RLWD
>A8L0Z3
VTSSTNSRRTLGAFLAGHTST----VRAVAFSPDGRILASASDDEPV-----RLWD
VTDPGDARPLDASLTGHSGW----VHSVAFSPDGHTLASAGDDHTV-----RLWN
VTDPANAHPLGAPLTGHTST----VWAVAFSPDGRILASAGNDET-----TLWD
VADPAQARPLDVISETR-----AVRSVAFSPDGRILASAGDDGTA-----SLWN
VADPTNPRPLGTPLAGHTNT----VWVAFSPNGHTLASAGDDHTV-----RLWN
VTDPANAHPLGAPLTGHTST----VRSVAFSSDSRTLASGSDDHTV-----RLWD
VIDPANAHPRGASLTGHSSW----VRSVAFAPDGRTLASGSDDHTM-----RLWD
>A8L0X3
---SLS----GSPLRDHTDS----VRSVAFSRDGRTLASASQDGTA-----RLWD
IAERTSQP-LTGRIA-----VWSVAFSPDKHTLASANGDSTV-----QLWD
VAEGTLPH-PVASLPGHSDA----VGSVAFSPDGRTLASASDDHTV-----RLWD
VATGTT----THTLTGHTGP----VNSVAFSRDGRTLASASDDHTV-----RLWD
VAEGTL----LRTLPGHTEP----VMSVAFSPDRRTLASASQDNTV-----RLWD
VAARTAPR-LVGSLSDHHTW----VMSVAFSPDGRTLASASQDRTV-----RLWD
VAARTT----THTLTGHTGP----VFSVAFSLDGRTLASASDDNTV-----RLWD
>A8LFK4
PAAQAL----GQPLTDHTDW----VASVAFSPDGHTLASSSKDTTV-----RLWD
ITDRTRPHPLGQPLAGHTLG----VMSVAFSPDGNTLASSSRDTTI-----RLWD
ITDRTRPHPLGQPLTGHTDA----VTSVAFFSDGRTLASSSRDTTI-----RLWD
ITDRTRRPLGSPLSGHSDW----VTSLALTMGRTLASSSLDSTV-----RLWN
MADRSHQPQIPLPLTGHTGG----VNSVAFSLDSRTLASSGRDTTI-----RLWD
VTDRSTPRLGAPITGHANT----VGPLTFSQDGDTLVSGSYDDTV-----RIWD
VTDRSHPRLLGLPLTGHTDW----IWSVALSPDGQTLASGSKDNTI-----RLWA
>A8L1U6
VTDRTPTPLEKTLYGH-TG---WVRSVVFSPDGRTLASGSDDDTV-----RFWD
MSDPTDPTPLGKPLTGH-TDP---IAAVAYSPDGRTLATSSTDRTV-----RFWD
MSDRADPTPLGKPLTRHSTDPTSVAAMAYS PDGRTLATGGLDNTL-----RLWD
VTDRTNPTPLGPPLTGH-TKT---VAFLAYSPDGRTLATGSYDRTV-----RLWD
VSDRTSPAPLGPPLTGHTNP----LMSLAYSPDGRTLATSASDYST-----RFWD
VSDRTRPTPLGEPIFLGDSPSVWVVGLAYSPDGHTLATISGDRLV-----RFWD
>A8L579
IGRRAARPAGPPLAGHGGW----VLSADFS PDGMIMATSAKDGEV-----LLWD
VAEPAAPRLVGRTPDGPDDG----VTSVVFSPDGHTLAGTSWDGSL-----RLWD
VTRPGAPHLGLRPLRAHAGP----VRSAAFATDGRMLATGSDDRSV-----QLWD
MADRGAPRPLGGPVTGATSF----VTSVAFAPDDHLLVAAGYDRDV-----LFWD
ITNRRAPRRLAR-IVAGPTS----ALVAAFAPDGRTLATGGVDGVI-----RLWD
VTVPDRPTALGRPLAGHDNR----VWALAFAPDGRTLASGGFDNVV-----RLWD
TSDRTRARPRGRPLAGHADW----VMSVRFAPHGRILASTGKDQTI-----RLWS
>A8LHK5
-SRNVEVTTRVALTG-HERD----VTSAAFSPDGRLLATT SKDGT-----RLWD
---TTTGRTVGRLSGRKISA----VHGCAFSPDGDLLATTGSDKTA-----RIWE
---IATERLALTLAGHKGP----VYGCAFSPDGRLLATVSTDRTV-----KLWG
---VSTGTNIATLTG-HRGS----VYGCAFSPDGRLLVTAGAEST-----LLWD
---VTIGETITSLAG-HTNF----ANGCSFSPDGLLLATT SNDGT-----RLTD
---TPTGTTTTLTPGS-----AQSCAFSPDGVLLATASTDDTA-----RLWD
---VATGTAVATLTG-HSST----VMACAFAPYGLLLATTSTDKTA-----RLWD
>A8L9U4
-LRTSYSAATATVLTGHTGS----VLGLGISPDGRTIATSGADNVA-----RLWD
VSDRTRPRQLSTIDAAGAW----VLDAAFSPDGKLLATVGYDRSA-----RLWD
IGDRTRPKQLSSMLAHNGY----VLDAAFSPDGRMLATSGYDNTA-----RLWD
ITDPRQPHELAVLDRHTSW----VNEVAFSPDGKLLATASADHTA-----RLWD
IANPRQPRPLAA-ITHTDF----VWTVAFSPDGRRLATGAYDGLV-----KIWD

ITDPTRPGATASFRADEKW-----VFDVAYSPDGRTLATAGWDTSV-----HLWD
>A8KX81
LAREPGVRPLGRPLADQAGW----ALSTAFSPDGKVMASSSRKGGV-----WLWN
LAVPEDPHPIGRPLTGHNTNR----VWSLAFSPDGGTLASSGFDNSV-----RLWD
ITDSGASKPAGRALTDDSGP----IWSVAFSADGRTLASGSDDTTV-----RLWD
MADPATPVRIDPALTGPRDG----VTSLAFSPDGSLLAGGSWDGSI-----WLWD
MADRGAPKRLAQ-SLSTPAT----TYVAAFSPNGRLLATGSTDGLV-----RLWD
MTNRARPWFQFVR-LSSDMEF----VTSVAFSADNRLLLVAAGFSRTI-----AIWD
>A8L1R5
GLFAADVDPVVAPLGSR-----VLSAALRPDGRLLAAGTEAGTI-----ELWD
LTDLAHPVHAGTISGVGD-----WVYSVAFNPGGNLLAAGVGDGAV-----RLWN
VTDPARAGALATIAFHRD----RVRSVAFAPDGGTLASGGDDGQV-----GLWA
VTDPSHPQRRSATDGAVA----GIRSLAFSPRGLLALAGNDGSV-----RLWN
VADPARPATSSTLRGTGR----TVQSVAFSADSSTLAAGGIDGSV-----HTWR
VDGPGSVVDLSTPGGVGG----VTSVGFSPGAILVSASEDETV-----RLTD
ISAPADPVLLTDLRGHTKA----VSAAMFVPGGRTVVSASGDGSV-----RLWT
>A8L653
VRDPAKPSRISDVPGPAGG-----TAELLFSPDGRTLALGGLNSPPEPAFQARVETWD
VTDLSRPRRLASLIAHRSS-----VRDLTFSPDGRTLVSAA-ERSV-----KLWD
VTDPRRLRLVSELPEFPGG-----VWEVRFSPDGRTLAAGGANPFA-----TLWD
-VTRMDAPRQIADLPGHSAS----VTSVAFSPDGTQLATGSNDNTV-----RIWD
VTEHDSPTLIEKLARSAGSEAG--IEEILYTRDGEKLVGVI FTVPA-----VVWD
>A8L516
ITDPRRPFMITAFEGPAGG-----VFSLRFSPDGRTLVLGGLGSSPAPAFSASLETWD
VRDPRHPARLASTVAHRGS-----VRDIQFSEDGHVMASVA-DRTV-----QLWD
VSDPRRIAARATLPDFPGA-----ALRAAFAPDGRTLAVGGGGPYA-----TLWN
ITDPARPTRTASLPGHVSE-----VNTVTFSPDGRTLVTGSGDNSV-----RVWD
VRRPSNPRLVERLSRSAGTNSG--IGTVAFSRDGSTLVGVVFTEPG-----ALWD

Single-motif proteins homologous to the *Frankia* series

<u>Genbank ID</u>	<u>Length</u>	<u>Organism</u>
297156412	73	Actinobacteria; <i>Streptomyces bingchenggensis</i> BCW-1
357265231	83	Cyanobacteria; <i>Crocospaera watsonii</i> WH 0003
357265846	99	Cyanobacteria; <i>Crocospaera watsonii</i> WH 0003
493214827	58	Cyanobacteria; <i>Nodularia spumigena</i>
497319093	71	Cyanobacteria; <i>Synechocystis</i> sp. PCC 7509
499925710	68	Actinobacteria; <i>Frankia alni</i>
518321537	72	Cyanobacteria; <i>Calothrix</i> sp. PCC 7103
518334279	105	Cyanobacteria; <i>Pleurocapsa</i> sp. PCC 7319
605047033	147	Cyanobacteria; <i>Richelia intracellularis</i>
651290729	77	Actinobacteria; <i>Actinokineospora inagensis</i>
658299017	107	Cyanobacteria; <i>Planktothrix agardhii</i> NIVA-CYA 126/8
664265184	79	Actinobacteria; <i>Streptomyces</i> sp. NRRL S-337
737391083	98	Cyanobacteria; <i>Dolichospermum circinale</i>
739354603	164	Cyanobacteria; <i>Richelia intracellularis</i>
746136515	167	Cyanobacteria; <i>Aphanizomenon flos-aquae</i>
750668271	67	Cyanobacteria; <i>Prochloron didemni</i>
750668722	67	Cyanobacteria; <i>Prochloron didemni</i>
752584810	84	Actinobacteria; <i>Frankia</i> sp. EAN1pec
752686059	66	Chloroflexi; <i>Roseiflexus</i> sp. RS-1
808671128	93	Actinobacteria; <i>Mycobacterium conceptionense</i>