

logo

NTDB id 1410 SGL RS13950 WP 010873519.1	MSVLTIGDKRLGAALLDAGLLTDEELQRALERHREVGGSLAEVLVDMGLLSERRIAQTIEDRFGIPLVELHRVEIPPKVK	0
NTDB id 1028 TT RS08215 WP 011173991.1	MSVLTIGDKRLGAALLDAGLLTDEELQRALERHREVGGSLAEVLVDMGLLSERRIAQTIEDRFGIPLVELHRVEIPPKVK	80
NTDB id 1168 A1552VC RS11070 WP 000957200.1		0
NTDB id 1293 VP RS12240 WP 005479695.1		0
NTDB id 1401 DSB67 RS12665 WP 010643259.1		0
NTDB id 1110 NGFG RS09235 WP 003689817.1		0
NTDB id 1251 GC085 RS07725 WP 011946523.1		0
NTDB id 252323 AYP RS01780 WP 001274990.1		0
NTDB id 1058 ABD1 RS01615 WP 001274986.1		0
NTDB id 1018 ACIAD RS01685 WP 004920473.1		0
consensus		

logo

NTDB id 1410 SGL RS13950 WP 010873519.1	ALLPAEKAKELKAIPFALDEEAGVVRVAFLNPLDTLSLEEVEDLTGLVVEPYQTTKSAFLYALAKHYPELGLPVPPPPSG	0
NTDB id 1028 TT RS08215 WP 011173991.1	ALLPAEKAKELKAIPFALDEEAGVVRVAFLNPLDTLSLEEVEDLTGLVVEPYQTTKSAFLYALAKHYPELGLPVPPPPSG	160
NTDB id 1168 A1552VC RS11070 WP 000957200.1		0
NTDB id 1293 VP RS12240 WP 005479695.1		0
NTDB id 1401 DSB67 RS12665 WP 010643259.1		0
NTDB id 1110 NGFG RS09235 WP 003689817.1		0
NTDB id 1251 GC085 RS07725 WP 011946523.1		0
NTDB id 252323 AYP RS01780 WP 001274990.1		0
NTDB id 1058 ABD1 RS01615 WP 001274986.1		0
NTDB id 1018 ACIAD RS01685 WP 004920473.1		0
consensus		

logo

NTDB id 1410 SGL RS13950 WP 010873519.1	EGQKDLKLGELLQKGWISREALEEALVEQEKTGDLLGRILVRKGLPEEALYRALAEQKGLEFLESTEGIVDPDPSAALLL	0
NTDB id 1028 TT RS08215 WP 011173991.1	EGQKDLKLGELLQKGWISREALEEALVEQEKTGDLLGRILVRKGLPEEALYRALAEQKGLEFLESTEGIVDPDPSAALLL	240
NTDB id 1168 A1552VC RS11070 WP 000957200.1		0
NTDB id 1293 VP RS12240 WP 005479695.1		0
NTDB id 1401 DSB67 RS12665 WP 010643259.1		0
NTDB id 1110 NGFG RS09235 WP 003689817.1		0
NTDB id 1251 GC085 RS07725 WP 011946523.1		0
NTDB id 252323 AYP RS01780 WP 001274990.1		0
NTDB id 1058 ABD1 RS01615 WP 001274986.1		0
NTDB id 1018 ACIAD RS01685 WP 004920473.1		0
consensus		

NTDB id 1410 SGL RS13950 WP 010873519.1
NTDB id 1028 TT RS08215 WP 011173991.1
NTDB id 1168 A1552VC RS11070 WP 000957200.1
NTDB id 1293 VP RS12240 WP 005479695.1
NTDB id 1401 DSB67 RS12665 WP 010643259.1
NTDB id 1110 NGFG RS09235 WP 003689817.1
NTDB id 1251 GC085 RS07725 WP 011946523.1
NTDB id 252323 AYP RS01780 WP 001274990.1
NTDB id 1058 ABD1 RS01615 WP 001274986.1
NTDB id 1018 ACIAD RS01685 WP 004920473.1
consensus

[illegible]

```

NTDB id 1410 SGL RS13950 WP 010873519.1
NTDB id 1028 TT RS08215 WP 011173991.1
NTDB id 1168 A1552VC RS11070 WP 000957200.1
NTDB id 1293 VP RS12240 WP 005479695.1
NTDB id 1401 DSB67 RS12665 WP 010643259.1
NTDB id 1110 NGBF RS09235 WP 003689817.1
NTDB id 1251 GC085 RS07725 WP 011946523.1
NTDB id 252323 AYP RS01780 WP 001274990.1
NTDB id 1058 ABD1 RS01615 WP 001274986.1
NTDB id 1018 ACIAD RS01685 WP 004920473.1
consensus

```

MSA of the protein sequences of the 1000 randomly selected sequences. The sequences are aligned using the ClustalW algorithm. The alignment is shown in the figure, with the sequences numbered 1 to 1000. The sequences are color-coded by their amino acid type: A (green), C (blue), D (red), E (yellow), F (purple), G (cyan), H (magenta), I (brown), K (pink), L (grey), M (dark green), N (dark blue), P (dark red), Q (orange), R (red), S (yellow), T (grey), V (dark green), W (dark blue), Y (purple). The alignment shows a high degree of conservation, with many residues being identical across the sequences. The sequences are also color-coded by their position in the protein: 1-100 (green), 101-200 (blue), 201-300 (red), 301-400 (yellow), 401-500 (purple), 501-600 (cyan), 601-700 (magenta), 701-800 (brown), 801-900 (pink), 901-1000 (grey).

```

NTDB id 1410 SGL RS13950 WP 010873519.1
NTDB id 1028 TT RS08215 WP 011173991.1
NTDB id 1168 A1552VC RS11070 WP 000957200.1
NTDB id 1293 VP RS12240 WP 005479695.1
NTDB id 1401 DSB67 RS12665 WP 010643259.1
NTDB id 1110 NGF6 RS09235 WP 003689817.1
NTDB id 1251 GC085 RS07725 WP 011946523.1
NTDB id 252323 AYP RS01780 WP 001274990.1
NTDB id 1058 ABD1 RS01615 WP 001274986.1
NTDB id 1018 ACIAD RS01685 WP 004920473.1
consensus

```

[illegible]

logo

NTDB id 1410 SGL RS13950 WP 010873519.1	QDDFNQLIEQYVAIREEWFAAQEKRDQERALDKLNDLTDIVGSIDMNLG	EDKEDVDNALDSNDANQAPVINLVNKILAKA	244
NTDB id 1028 TT RS08215 WP 011173991.1	EAAITTKLIERFYGKAELSEIAKEF.....	AKKQAEIEVPSPLELDESAAQKFVKQVIREA	518
NTDB id 1168 A1552VC RS11070 WP 000957200.1	HRALQAAIRRLYGRSISQGAANQGK..EI.SQDELANL.....	VKVSDDQLQSTIEDLSQDDSPVSRFINQVLLDA	193
NTDB id 1293 VP RS12240 WP 005479695.1	FRELSTAIRRLYGRSLSHEKSGLK..EI.NQEELASL.....	VDVGADEIDNIEDLSQDESVPVSRYINQILLDA	193
NTDB id 1401 DSB67 RS12665 WP 010643259.1	FRELTAAIRRLYGRSLGQEKSGLK..EI.NQDELAGL.....	VDIGEDEIENIEDLSQDESVPVSRYINQILLDA	193
NTDB id 1110 NGFG RS09235 WP 003689817.1	DDQLAGLLDWVGSFSTSL..LQEL.....	GEG.....QEEEEHSHTLYIDNEEAEDGPVPRFIHKHTLSDA	183
NTDB id 1251 GC085 RS07725 WP 011946523.1	TDKLSALIDNLLTAKESQGLSEYVEDSG.DLEGLEIS.....	AD.DEDQDIDTATSVTDDAPIVKFVNKILLDA	202
NTDB id 252323 AYP RS01780 WP 001274990.1	HHKLEKVLGQHFAEESSFDFNDE...EF.D...LDVN.....	LDGPTTQEDEEEAPQGDEAPIVKYINKLLIDA	197
NTDB id 1058 ABD1 RS01615 WP 001274986.1	HHKLEKVLGQHFAEESSFDFNDE...EF.D...LDVN.....	LDGPTAQEDEEEAPQGDEAPIVKYINKLLIDA	197
NTDB id 1018 ACIAD RS01685 WP 004920473.1	HDKLERLLSEHFVEETHFNFDTE...EL.D...LDVE.....	VDPHTTDDDDDEDKDKLDEAPIVKYINKLLIDA	197
consensus	* ** **** * *	** ***** *	*****!!!!!!

logo

NTDB id 1410 SGL RS13950 WP 010873519.1	LQEGTSDIHVEPQEETLRIRFRKDGVLNQAFDPLPRKITPAVVARFKIMADMDIAERRSPQDGKIRRIYQ..GRKVDFRV	322
NTDB id 1028 TT RS08215 WP 011173991.1	FLQDASDIHIEPRQNDVQVRLRIDGALRPYSTL.PKGALNAVISVVKIMGGLNIAEKRLPQDGRVRYR..EGAIDVDLRL	595
NTDB id 1168 A1552VC RS11070 WP 000957200.1	VRKGASDIHFEPYENQYRIRLRCDGILVETQQP.ASHLSRRLAARIKILSKLDIAERRLPQDGRIKLRLSR.DTAIDMRV	271
NTDB id 1293 VP RS12240 WP 005479695.1	VRKGASDIHFEPYEKMYRVRLRCDGILVETQQP.PNHLSRRLSARIKILSKLDIAERRLPQDGRIKLKLNQ.DTAIDMRV	271
NTDB id 1401 DSB67 RS12665 WP 010643259.1	VRKGASDIHFEPYENMYRVRLRCDGILVETQQP.PSHLSRRLSARIKILSKLDIAERRLPQDGRIKLKLNQ.DTAIDMRV	271
NTDB id 1110 NGFG RS09235 WP 003689817.1	LRSGASDIHFEFYEHNARIRFRVDGQLREVVQP.PIAVRGQLASRIKVM SRLDISEKRIPQDGRMQLTFQKGGKPVDFRV	262
NTDB id 1251 GC085 RS07725 WP 011946523.1	IRQGASDIHFEPYEREYRIRYRQDGLHEVATP.PASLSRITARIKVM SNLDISERRIPQDGGFKMKISK.SRAIDFRV	280
NTDB id 252323 AYP RS01780 WP 001274990.1	IRMGASDLHFEPYEKSYRVRYRVDGMLRQIANP.PLQLANRLASRLKVM SQMDISEKRVPPQDGRIKLKLSK.SKAIDFRV	275
NTDB id 1058 ABD1 RS01615 WP 001274986.1	IRMGASDLHFEPYEKSYRVRYRVDGVLRLQIANP.PLQLANRLASRLKVM SQMDISEKRVPPQDGRIKLKLSK.SKAIDFRV	275
NTDB id 1018 ACIAD RS01685 WP 004920473.1	IRMSASDLHFEPYEKSYRVRYRVDGVLRLIATP.PLQLATRLASRLKVM SQMDISEKRVPPQDGRIKLKMSK.SKIDFRV	275
consensus	** !!!*!*!*!* **!*!*!*!*!* ** *	** ** *!*!* **!* !!!*!!!!!!* ***!*!* *

logo

NTDB id 1410 SGL RS13950 WP 010873519.1	NTLPSRYGEKVCLRILDNSATQLGLDFLITNPETLQTVRELAARPYGLMLVTGPTGSGKSTTLYSVLAERNSPEVNINTA	402
NTDB id 1028 TT RS08215 WP 011173991.1	STLPTVYGEKAVMRLLKKASDIPEIEDLGFAPGVFERFKEVTSKPYGIFLITGPTGSGKSFTTFSILKRIATPDKNTQTI	675
NTDB id 1168 A1552VC RS11070 WP 000957200.1	STLPTLWGEKIVLRLLDSSAANLIDIDKLGYNPQQKQLYLEALNRKPQGMILMTGPTGSGKTVSLYTGLRILNTSQINISTA	351
NTDB id 1293 VP RS12240 WP 005479695.1	STLPTLWGEKIVLRLLDSSASLIDIDKLGYSSEQKQLYLEALRRPQGMILMTGPTGSGKTVSLYTGLNILNKPEINISTA	351
NTDB id 1401 DSB67 RS12665 WP 010643259.1	STLPTLWGEKIVLRLLDSSAALDINKLGYSDSQKQLYLDALRRPQGMILMTGPTGSGKTVSLYTGLSILNKPEINISTA	351
NTDB id 1110 NGFG RS09235 WP 003689817.1	STLPTLWGEKIVMRILNSDAASLNIDQLGFEPFQKKLLLEATHRPYGMVLVTGPTGSGKTVSLYTCLNILNTESVNIATA	342
NTDB id 1251 GC085 RS07725 WP 011946523.1	STCPTSAGEKIVMRVLDSGAAKLGIEALGFNPVQRTNFKLAIQRPQGMILVTGPTGSGKTVTLYTALNILNTIEVNISTA	360
NTDB id 252323 AYP RS01780 WP 001274990.1	NSLPTLWGEKIVLRILDPSSAMLGIDALGYEEDQKALFMEALDKPQGMILLITGPTGSGKTVSLYTGLNILNTESSNISTA	355
NTDB id 1058 ABD1 RS01615 WP 001274986.1	NSLPTLWGEKIVLRILDPSSAMLGIDALGYEEDQKALFMEALDKPQGMILLITGPTGSGKTVSLYTGLNILNTESSNISTA	355
NTDB id 1018 ACIAD RS01685 WP 004920473.1	NSLPTLWGEKIVLRILDPASAMLGIDALGYEPEKQALFMEALNKPQGMILLITGPTGSGKTVSLYTGLNILNTEHANISTA	355
consensus	***!***!!*!***!*!***** **** *!*** ** ***** *!***!!*!*****!***** *!***!*	

logo

NTDB id 1410 SGL RS13950 WP 010873519.1
NTDB id 1028 TT RS08215 WP 011173991.1
NTDB id 1168 A1552VC RS11070 WP 000957200.1
NTDB id 1293 VP RS12240 WP 005479695.1
NTDB id 1401 DSB67 RS12665 WP 010643259.1
NTDB id 1110 NGFG RS09235 WP 003689817.1
NTDB id 1251 GC085 RS07725 WP 011946523.1
NTDB id 252323 AYP RS01780 WP 001274990.1
NTDB id 1058 ABD1 RS01615 WP 001274986.1
NTDB id 1018 ACIAD RS01685 WP 004920473.1
consensus



482
755
431
431
431
422
440
435
435
435

logo

NTDB id 1410 SGL RS13950 WP 010873519.1
NTDB id 1028 TT RS08215 WP 011173991.1
NTDB id 1168 A1552VC RS11070 WP 000957200.1
NTDB id 1293 VP RS12240 WP 005479695.1
NTDB id 1401 DSB67 RS12665 WP 010643259.1
NTDB id 1110 NGFG RS09235 WP 003689817.1
NTDB id 1251 GC085 RS07725 WP 011946523.1
NTDB id 252323 AYP RS01780 WP 001274990.1
NTDB id 1058 ABD1 RS01615 WP 001274986.1
NTDB id 1018 ACIAD RS01685 WP 004920473.1
consensus



561
819
493
492
492
488
505
501
501
501

logo

NTDB id 1410 SGL RS13950 WP 010873519.1
NTDB id 1028 TT RS08215 WP 011173991.1
NTDB id 1168 A1552VC RS11070 WP 000957200.1
NTDB id 1293 VP RS12240 WP 005479695.1
NTDB id 1401 DSB67 RS12665 WP 010643259.1
NTDB id 1110 NGFG RS09235 WP 003689817.1
NTDB id 1251 GC085 RS07725 WP 011946523.1
NTDB id 252323 AYP RS01780 WP 001274990.1
NTDB id 1058 ABD1 RS01615 WP 001274986.1
NTDB id 1018 ACIAD RS01685 WP 004920473.1
consensus



641
889
562
561
561
558
574
570
570
570

logo

NTDB id 1410 SGL RS13950 WP 010873519.1	KRKSALTCRTCAAELQQEWLDCPYCMTSRFS	672
NTDB id 1028 TT RS08215 WP 011173991.1		889
NTDB id 1168 A1552VC RS11070 WP 000957200.1		562
NTDB id 1293 VP RS12240 WP 005479695.1		561
NTDB id 1401 DSB67 RS12665 WP 010643259.1		561
NTDB id 1110 NGFG RS09235 WP 003689817.1		558
NTDB id 1251 GC085 RS07725 WP 011946523.1		574
NTDB id 252323 AYP RS01780 WP 001274990.1		570
NTDB id 1058 ABD1 RS01615 WP 001274986.1		570
NTDB id 1018 ACIAD RS01685 WP 004920473.1		570
consensus		

-  non conserved
-  similar
-  ≥ 50% conserved