

DeepCoil – a fast and accurate prediction of coiled-coil domains in protein sequences

Supplementary Table 1. Statistics on the coiled coil oligomeric states, orientation, and presence of non-canonical interactions for the sequences used in the training and test sets.

Category	Training set ^(a)	Test set #1 ^(b)	Test set #2 (Li et al.) ^(b, c)
Number of sequences	10,438	1,193	518
Number of residues	2,569,729	259,070	123,678
Number of coiled-coil residues	77,780	7,862	14,385
% of non-canonical residues	5.6%	5.0%	3.1%
Number of coiled-coil regions	4,140	417	745
% of parallel dimers	21.3%	25.9%	18.1%
% of antiparallel dimers	54.6%	55.6%	65.9%
% of trimers	17.3%	14.3%	11.4%
% of tetramers	6.2%	3.8%	4.3%

^(a) Maximal pairwise sequence identity in the training set is 50%; ^(b) Maximal pairwise sequence identity in the test sets is 30% and none of the test sets sequences shows more than 30% sequence identity to the training set sequences; ^(c) A test set derived from the test set defined in a study of (Li *et al.*, 2016).

Supplementary Table 2. Per-residue classification performance scores (area under the ROC curve) for the categories defined in Supplementary Table 1 assessed with the aid of test set #1 (defined in this study).

Method	Parallel dimers	Antiparallel dimers	Trimers	Tetramers	Non-canonical
DeepCoil_PSSM	0.968	0.958	0.946	0.985	0.942
DeepCoil_SEQ	0.954	0.920	0.897	0.963	0.900
PCOILS_28 ^(a)	0.924	0.830	0.825	0.931	0.863
PCOILS_21 ^(a)	0.928	0.832	0.842	0.939	0.870
COILS_28 ^(a)	0.885	0.766	0.747	0.880	0.785
Marcoil	0.862	0.766	0.777	0.900	0.817
PCOILS_14 ^(a)	0.926	0.814	0.833	0.915	0.864
COILS_21 ^(a)	0.908	0.803	0.801	0.908	0.843
CCHMM_PROF ^(b)	–	–	–	–	–
COILS_14 ^(a)	0.912	0.799	0.790	0.906	0.837
Multicoil2	0.829	0.702	0.615	0.795	0.707

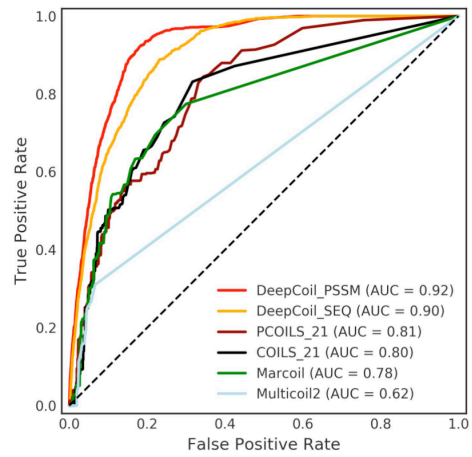
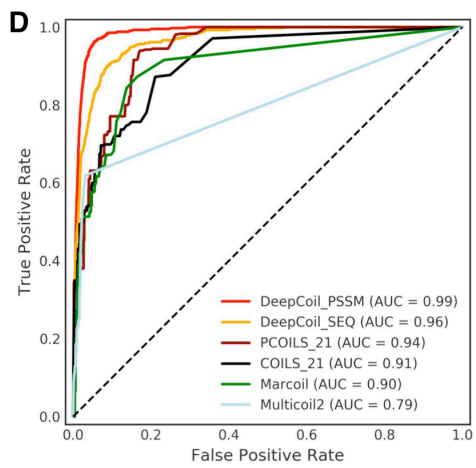
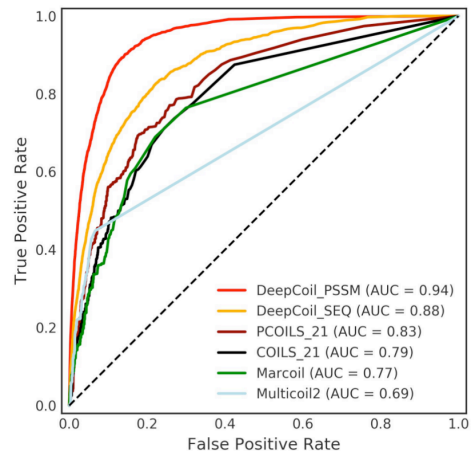
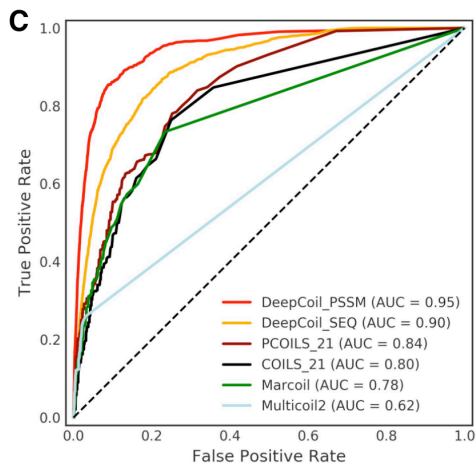
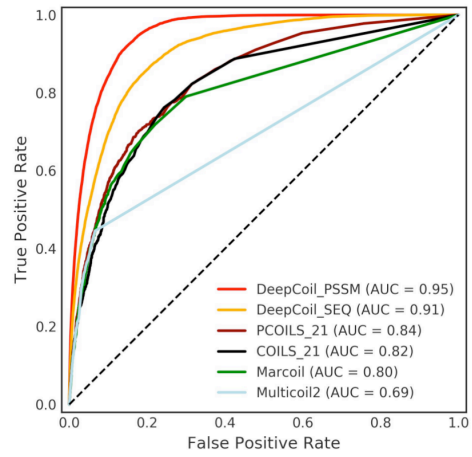
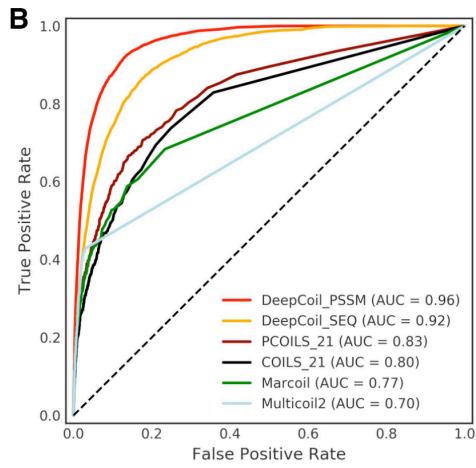
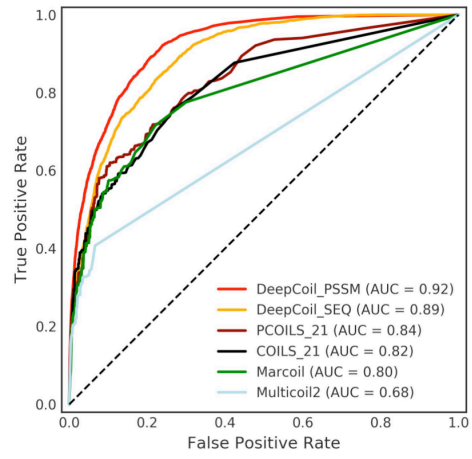
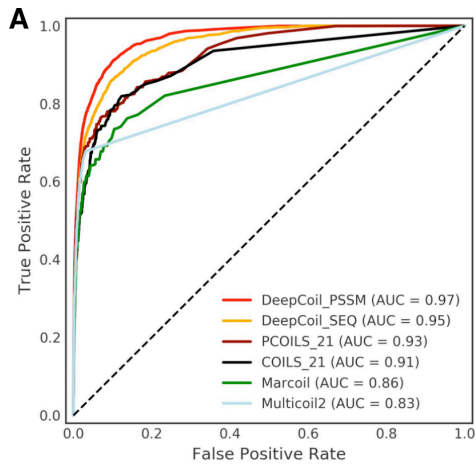
^(a) suffix refers to the length of the scanning window; ^(b) CCHMM_PROF does not return per-residue probabilities.

Supplementary Table 3. Per-residue classification performance scores (area under the ROC curve) for the categories defined in Supplementary Table 1 assessed with the aid of test set #2 (defined based on the study of Li et al.).

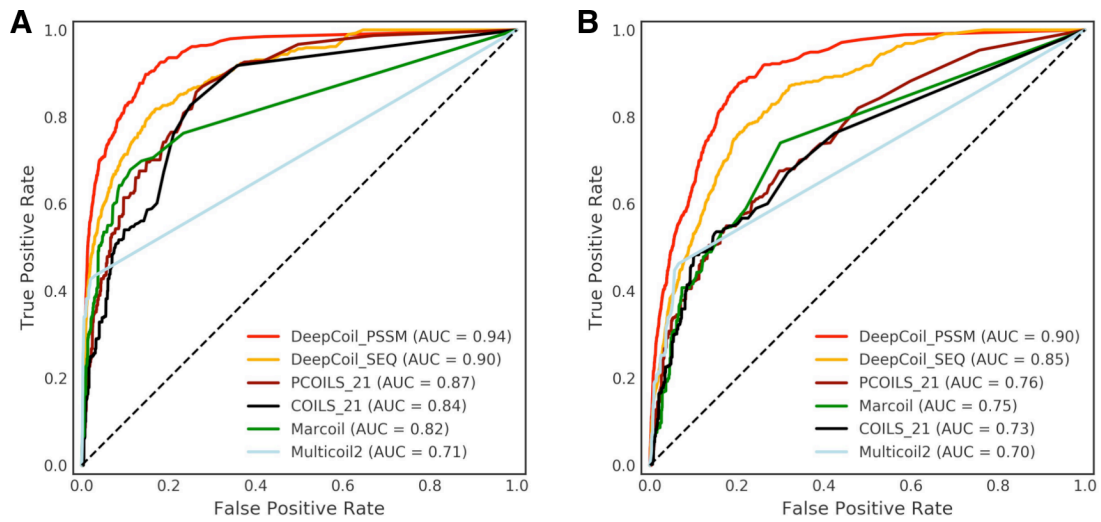
Method	Parallel dimers	Antiparallel dimers	Trimers	Tetramers	Non-canonical
DeepCoil_PSSM	0.921	0.951	0.943	0.925	0.900
DeepCoil_SEQ	0.893	0.906	0.879	0.900	0.849
PCOILS_28 ^(a)	0.827	0.837	0.836	0.817	0.765
PCOILS_21 ^(a)	0.835	0.837	0.827	0.811	0.755
COILS_28 ^(a)	0.765	0.791	0.783	0.765	0.724
Marcoil	0.797	0.799	0.770	0.777	0.749
PCOILS_14 ^(a)	0.830	0.819	0.795	0.800	0.731
COILS_21 ^(a)	0.819	0.818	0.795	0.798	0.732
CCHMM_PROF ^(b)	–	–	–	–	–
COILS_14 ^(a)	0.819	0.808	0.776	0.792	0.747
Multicoil2	0.676	0.693	0.694	0.621	0.701

^(a) suffix refers to the length of the scanning window; ^(b) CCHMM_PROF does not return per-residue probabilities.

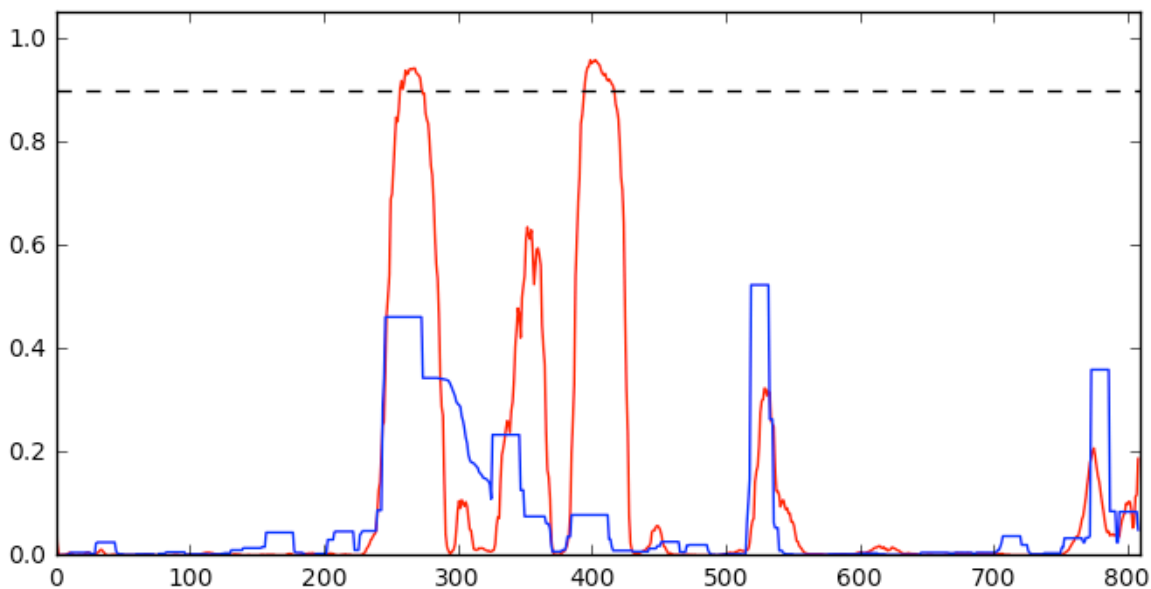
Supplementary Figure 1. Performance of per-residue coiled coil predictions for **(A)** parallel dimers, **(B)** antiparallel dimers, **(C)** trimers, **(D)** tetramers assessed with the aid of test set #1 (defined in this study; left column) and test set #2 (defined based on the work of (Li *et al.*, 2016); right column). For clarity, some of the methods were omitted. For all AUC scores see Supplementary Tables 2 and 3.



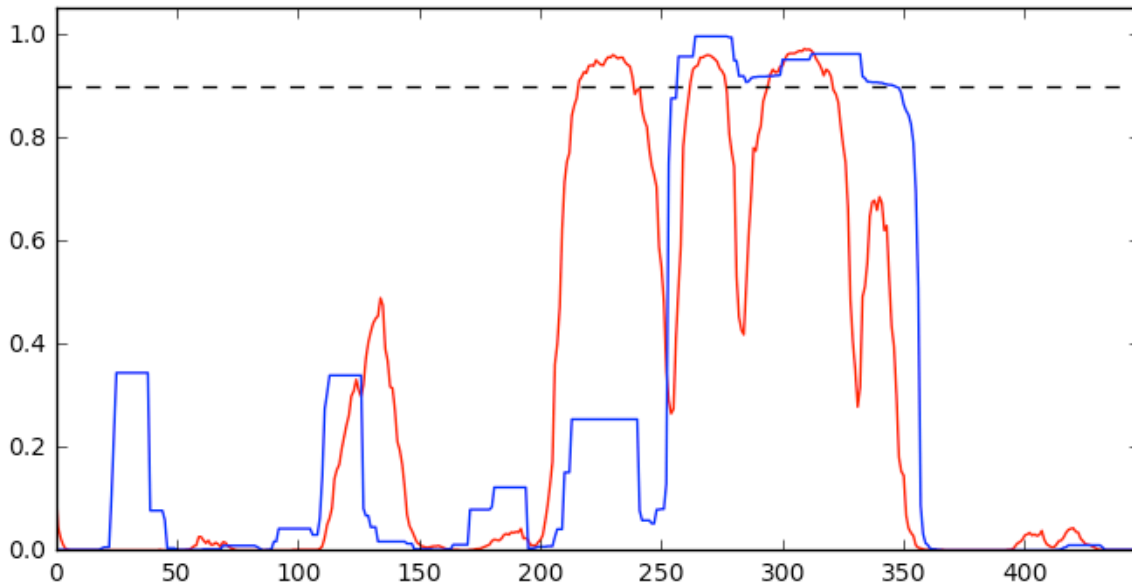
Supplementary Figure 2. Performance of per-residue predictions of non-canonical coiled coils with the aid of **(A)** test set #1 (defined in this study) and **(B)** test set #2 (defined based on the work of (Li *et al.*, 2016)). For clarity, some of the methods were omitted. For all AUC scores see Supplementary Tables 2 and 3.



Supplementary Data 1. New coiled-coil regions identified in human proteins using DeepCoil_PSSM. Plots indicate per-residue scores provided by DeepCoil_PSSM (red) and the maximal score provided by all other methods (blue). The dashed horizontal line indicates the probability threshold used for DeepCoil_PSSM. Sequences are annotated with the following labels: **NH** – no homologous structure found for the region predicted with DeepCoil; **H+** – homologous structure found and it contains a coiled coil in the prediction region; **H-** – homologous structure found, but it does not contain a coiled coil in the predicted region.

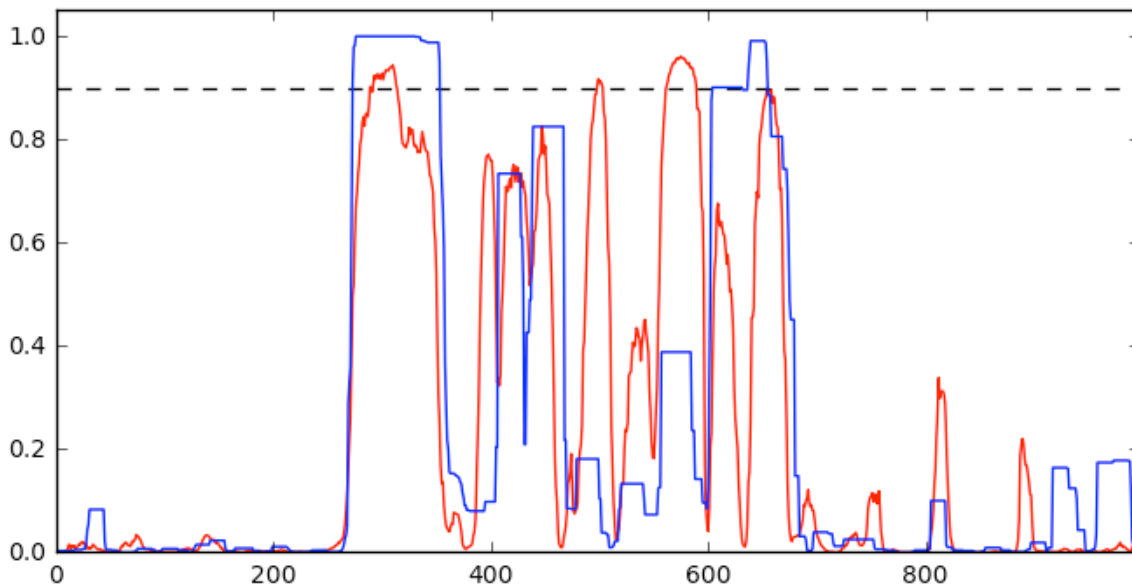


```
>NP_001273685.1 H+
MLRFPTCFPSFRVVGKQLPQEIIFLVWSPKRDIALANTAGEVLLHRLASFHRVWSFPPNENTGKEVTCCLAWRPDGKLL
AFALADTKKIVLCDVEKPELSHSFSVEAPVSCMHWMEVTVESVLTSTFYNAEDESNNLLPKLPTLPKNYSNTSKIIFSEEN
SDEIIKLLGDVRLNILVLGGSSGFIELYAYGMFKIARVTGIAGTCLALCLSSDLKSLSVVTEVSTNGASEVSYFQLETNL
LYSFLPEVTRMARKFTHISALLQYINLSLTCMCEAWEEILMQMDSRLTKFVQEKNTTTSVQDEFMHLLLWGKASAELOTL
LMNQLTVKGLKKGQSIESSYSSIQKLVISHLQSGSESLYHLSELKGMASWKQKYEPLGLDAAGIEEAITAVGSFILKA
NELLQVIDSSMKNFKAFFRWLYVAMLRMTEDHVLPELNKVMQDKDITFVAEFLTEHFNEAPDLYNRKGKGFNVERVGQYL
KDEDDDLVSPNTEGNQWYDFLQNSSHLKESPLLFPYYPBRKSLHFVKRRMENIIDQCLQKPADVIGKSMNQAICIPLYRD
TRSEDSTRRLFKFPFLWNNKTSNLHYLLFTILEDLSLYKMCILRRHTDISQSVSNGLIAIKFGSFTYATTEKVRRSIYSCL
DAQFYDDETVTVLKDVTGREGDRLLVQLPLSLVYNSEDSAEYQFTGTYSTRLDEQCSAIPTRTMHFEEKHWRLLESMDKA
QYVAGNGFRKVCVLSNLRHVRVFEMDIDDEWELDESSDEEEEEASNPKVVKIKEEVLSESEEAENQQAGAAALAPEIVIKV
EKLDPELDS
```



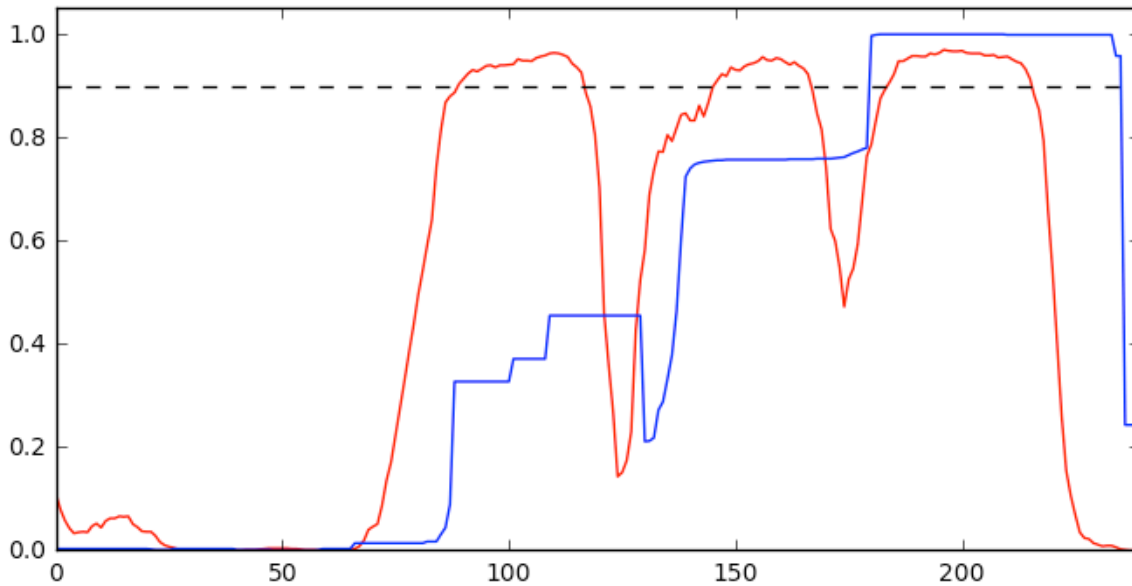
>XP_024304185.1 H+

MET~~Y~~A~~E~~V~~G~~K~~E~~G~~K~~P~~S~~C~~A~~S~~V~~D~~L~~Q~~G~~D~~S~~S~~L~~Q~~V~~E~~I~~S~~D~~A~~V~~S~~E~~R~~D~~K~~V~~K~~F~~T~~V~~Q~~T~~K~~S~~C~~L~~P~~H~~F~~A~~Q~~T~~E~~F~~S~~V~~V~~R~~Q~~H~~E~~E~~F~~I~~W~~L~~H~~D~~A~~Y~~V~~E~~N~~E~~E~~Y~~
 A~~G~~L~~I~~I~~P~~P~~A~~P~~P~~R~~P~~D~~F~~E~~A~~S~~R~~E~~K~~L~~Q~~K~~L~~G~~E~~G~~D~~S~~S~~V~~T~~R~~E~~E~~F~~A~~K~~M~~K~~Q~~E~~L~~E~~A~~E~~Y~~L~~A~~I~~F~~K~~K~~T~~V~~A~~M~~H~~E~~V~~F~~L~~Q~~R~~L~~A~~A~~H~~P~~T~~L~~R~~R~~D~~H~~N~~F~~F~~V~~F~~
 L~~E~~Y~~G~~Q~~D~~L~~S~~V~~R~~G~~K~~N~~R~~K~~E~~L~~L~~G~~G~~F~~L~~R~~N~~I~~V~~K~~S~~A~~D~~E~~A~~L~~I~~T~~G~~M~~S~~G~~L~~K~~E~~V~~D~~D~~F~~F~~E~~H~~E~~R~~T~~F~~L~~L~~E~~**Y**H**T**R**I**R**D**A**C**L**R**A**D**R**V**M**R**A**H**K**C**L**A**D
 D~~Y~~I~~P~~I~~S~~A~~A~~L~~S~~S~~L~~G**T**Q**E**V**N**Q**L**R**T**S**F**L**K**L**A**E**L**F**E**R**L**R**K**L**E**G**R**V**A**S**D**E**D**L**K**L**S**D**M**L**R**Y**M**R**D**S**Q**A**A**K**D**L**L**Y**R**R**L**R**A**L**A**D**Y**E**N**A
NK**A**L**D**K**A**R**T**R**N**R**E**V**R**P**A**E**S**H**Q**Q**L**C**C**Q**R**F**E**R**L**S**D**S**A**K**Q**G**E**P**A**A**P**S**P**S**L**G**P**H**A**L**P**C**P**D**P**H**A**S**S**P**L**Q**S**S**W**T**S**S**P**A**G**S**P**L**F**E**R**
 I~~S~~L~~S~~W~~Q~~S~~W~~S~~S~~N~~T~~P~~S~~S~~K~~V~~H~~Q~~A~~H~~F~~L~~P~~R~~Q~~Y~~D~~Q~~I~~L~~E~~T~~P~~R~~R~~A~~P~~L~~Q~~H~~P~~Q~~P~~S~~L~~Q



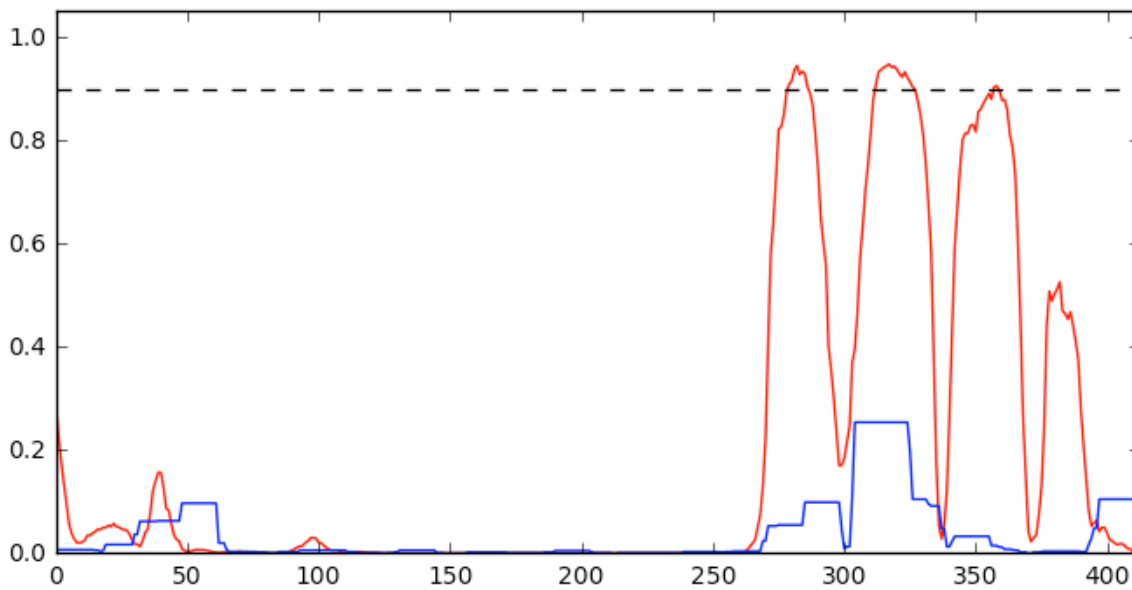
>NP_001247433.1 NH

M~~E~~D~~T~~H~~Y~~D~~K~~V~~E~~D~~V~~V~~G~~S~~H~~I~~E~~D~~A~~V~~T~~F~~W~~A~~Q~~S~~I~~N~~R~~N~~K~~D~~I~~M~~K~~I~~G~~C~~S~~L~~S~~E~~V~~C~~P~~Q~~A~~S~~S~~V~~L~~G~~N~~L~~D~~P~~N~~K~~I~~Y~~G~~G~~L~~F~~S~~E~~D~~Q~~C~~W~~Y~~R~~C~~K~~V~~L~~K~~I
 I~~S~~V~~E~~K~~C~~L~~V~~R~~Y~~I~~D~~Y~~G~~N~~T~~E~~I~~L~~N~~R~~S~~D~~I~~V~~E~~I~~P~~L~~E~~L~~Q~~F~~S~~S~~V~~A~~K~~K~~Y~~K~~L~~W~~G~~L~~H~~I~~P~~S~~D~~Q~~E~~V~~T~~Q~~F~~D~~Q~~G~~T~~T~~F~~L~~G~~S~~L~~I~~F~~E~~K~~E~~I~~K~~M~~R~~I~~K~~A~~T~~S~~
 E~~D~~G~~T~~V~~I~~A~~Q~~A~~E~~Y~~G~~S~~V~~D~~I~~G~~E~~E~~V~~L~~K~~K~~G~~F~~A~~E~~K~~C~~R~~L~~A~~S~~R~~T~~D~~I~~C~~E~~E~~K~~L~~D~~P~~G~~Q~~L~~V~~L~~R~~N~~L~~K~~S~~P~~I~~P~~L~~W~~G~~H~~R~~S~~N~~Q~~S~~T~~F~~S~~R~~P~~K~~G~~H~~L~~S~~E~~K~~M
 T~~L~~D~~L~~K~~D~~E~~N~~D~~A~~G~~N~~L~~I~~T~~F~~P~~K~~E~~S~~L~~A~~V~~G~~D~~F~~N~~L~~G~~S~~N~~V~~**S**L**E**K**I**K**Q**D**Q**K**L**I**E**N**E**K**L**K**T**E**K**D**A**L**L**E**S**Y**K**A**L**E**L**K**V**E**Q**I**A**Q**E**L**Q**E**K**A
AA**V**D**L**T**N**H**L**E**Y**T**L**K**T**Y**I**D**T**R**M**K**N**L**A**A**K**M**E**I**L**K**E**M**R**H~~V~~D~~I~~S~~V~~R~~F~~G~~K~~D~~L~~S~~D~~A~~I~~Q~~V~~L~~D~~E~~G~~C~~F~~T~~T~~P~~A~~S~~L~~N~~G~~L~~E~~I~~I~~W~~A~~E~~Y~~S~~L~~A~~Q~~E
 N~~I~~K~~T~~C~~E~~Y**V**S**E**G**N**L**I**A**Q**R**N**E**M**Q**Q**K**L**Y**M**S**V**E**D**F~~I~~L~~E~~V~~D~~E**S**S**L**N**K**R**L**K**T**L**Q**D**L**S~~V~~S~~L~~E~~A~~V~~Y~~G~~Q~~A**E**K**A**N**S**D**E**I~~L~~K~~K~~F~~Y~~D~~W~~K
 D~~K~~R~~E~~E~~F~~T~~S~~V~~R~~S~~E~~T~~D~~A~~S~~L~~H~~R~~L~~V~~A~~W~~F~~Q~~R~~T~~L~~K~~V~~F~~D~~L~~S~~V~~E~~G~~L~~I~~S~~E~~D~~A~~M~~N~~D~~I~~E~~I~~L~~E~~K~~T~~E~~S~~S~~V~~C~~K~~E~~L~~I~~A~~L~~V~~D~~Q~~G~~D~~A~~D~~K~~E~~I~~I~~S~~N
 T~~Y~~**S**Q**V**L**Q**K**I**H**S**E**R**L**I**A**T**V**Q**A**K**Y**K**D**S**I**E**F**K**Q~~L~~I~~E~~Y~~L~~N~~K~~S~~P~~S**V**D**H**L**L**S**I**K**K**T**L**K**S**L**K**A**L**L**R**W**K**L**V**E**K**S~~N~~L**E**S**D**D**P**D**G**S**Q**
IE**K**I**K**E**I**T**Q**L**R**N**N**V**F**Q**E**I~~Y~~H~~E~~R~~E~~E~~Y~~E~~M~~L~~T~~S~~L~~A~~Q~~K~~W~~F~~P~~E~~L~~P~~L~~H~~P~~E~~I~~G~~L~~L~~K~~Y~~M~~N~~S~~G~~L~~L~~T~~M~~S~~L~~R~~D~~L~~L~~D~~A~~E~~P~~M~~K~~E~~L~~S~~S~~K~~R
 P~~L~~V~~R~~S~~E~~V~~N~~G~~Q~~I~~I~~L~~L~~K~~G~~Y~~S~~V~~D~~V~~D~~T~~E~~A~~K~~V~~I~~E~~R~~A~~A~~T~~Y~~H~~R~~A~~W~~R~~E~~A~~E~~G~~S~~G~~L~~L~~P~~L~~I~~F~~L~~F~~L~~C~~K~~S~~D~~P~~M~~A~~Y~~L~~M~~V~~P~~Y~~P~~R~~A~~N~~L~~N~~A~~V~~Q~~A~~N~~
 M~~P~~L~~N~~S~~E~~E~~T~~L~~K~~V~~M~~K~~G~~V~~A~~Q~~L~~H~~T~~L~~H~~K~~A~~D~~I~~H~~G~~S~~L~~H~~Q~~N~~N~~V~~F~~A~~L~~N~~R~~E~~Q~~I~~V~~G~~D~~F~~D~~F~~T~~K~~S~~V~~S~~Q~~R~~A~~S~~V~~N~~M~~M~~V~~G~~D~~L~~S~~L~~M~~S~~P~~E~~L~~K~~M~~G~~K
 P~~A~~S~~P~~G~~S~~D~~L~~Y~~A~~G~~C~~L**L**L**L**W~~S~~V~~Q~~N~~Q~~E~~F~~E~~I~~N~~K~~D~~G~~I~~P~~K~~V~~D~~Q~~F~~H~~L~~D~~D~~K~~V~~K~~S~~L~~L~~C~~S~~L~~I~~C~~Y~~R~~S~~S~~M~~T~~A~~E~~Q~~V~~L~~N~~A~~E~~C~~F~~L~~M~~P~~K~~E~~Q~~S~~V~~P~~N~~P
 E~~K~~D~~T~~E~~Y~~T~~L~~Y~~K~~K~~E~~E~~E~~I~~K~~T~~E~~N~~L~~D~~K~~C~~M~~E~~K~~T~~R~~N~~G~~E~~A~~N~~F~~D~~C~~



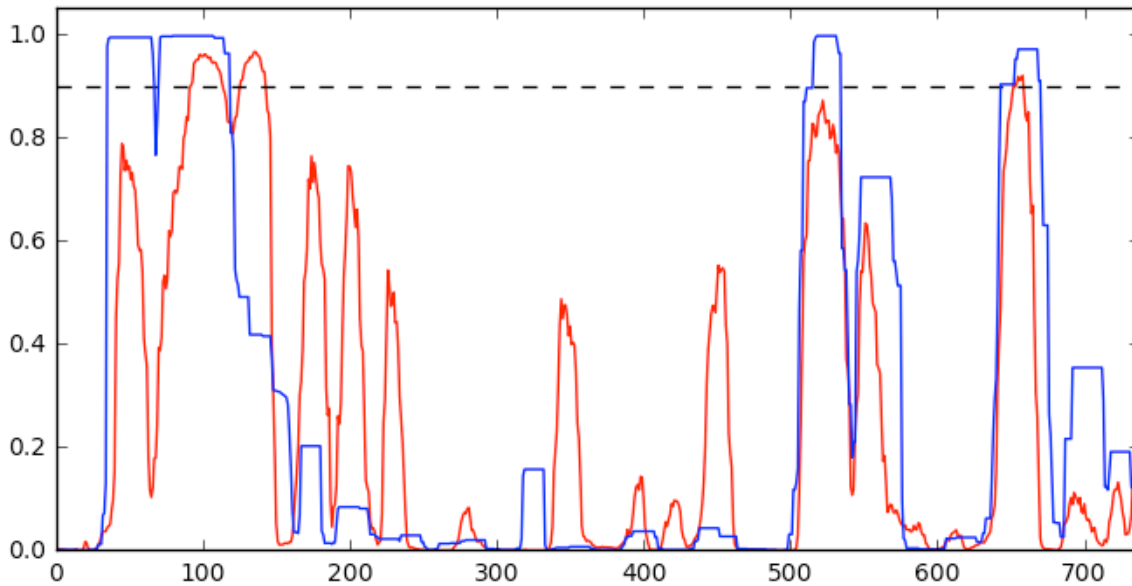
>NP_063940.1 H+

MAALKSWLSRSVTSFFRYRQCLCVPVVFANFKKRCFSELIRPWHKTVTIGFGVTLCAVPIAQKSEPHSLSSSEALMRRVSL
 VTDSTSTFLSQTYYALIEAITEYTKAVYTLTSLYRQYTSLLGKMNSEEEDEVWQVIIGARAEMTSKHQEYLKLETTWMTA
 VGLSEMAEEAAYQTGADQASITARNHIQLVKLQVEEVHQLSRKAETKLAEAQIEELRQKTQEEGEERAESQEQAYLRED



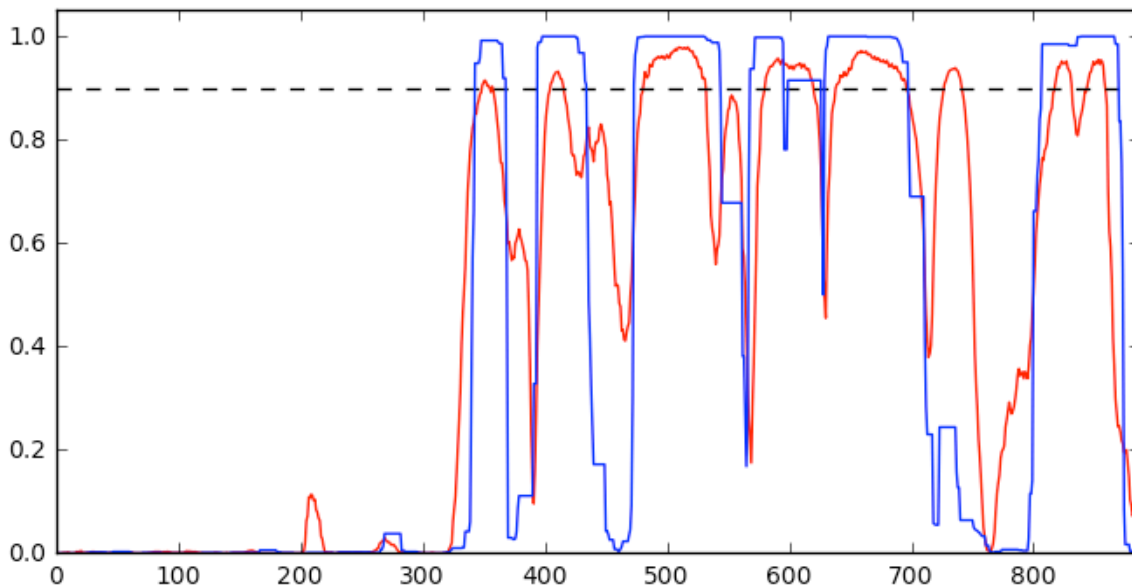
>NP_000008.1 H+

MAAALLARASGPARRALCPRAWRQLHTIYQSVELPETHQMLLQTCRDFAEKELFPIAAQVDKEHLFPAAQVKKMGGLGLL
 AMDVPEELGGAGLDYLAYAIAMEEISRGCASTGVIMSVNNSLYLGPILKFGSKEQKQAWVTPFTSGDKIGCFALSEPGNG
 SDAGAASTTARAEGDSWVLNGTKAWITNAWEASA AVVFASTDRALQNKGISAF LVPMPPTGLTLGKKEDKLGIRGSSTAN
 LIFEDCRIPKDSILGEPGMGFKIAMQTLDMGRIGIASQALGIAQTALDCAVNYAENRMAFGAPLTKLQVIQF**KLADMALA**
LESARLLTWRAAMLKDNKKPFIKEAAMAKLAASEAATAISHQAIQILGGMGYVTEMPAERHYRDARITEIYEGTSEIQRL
 VIAGHLLRSYRS



>NP_001138508.1 H+

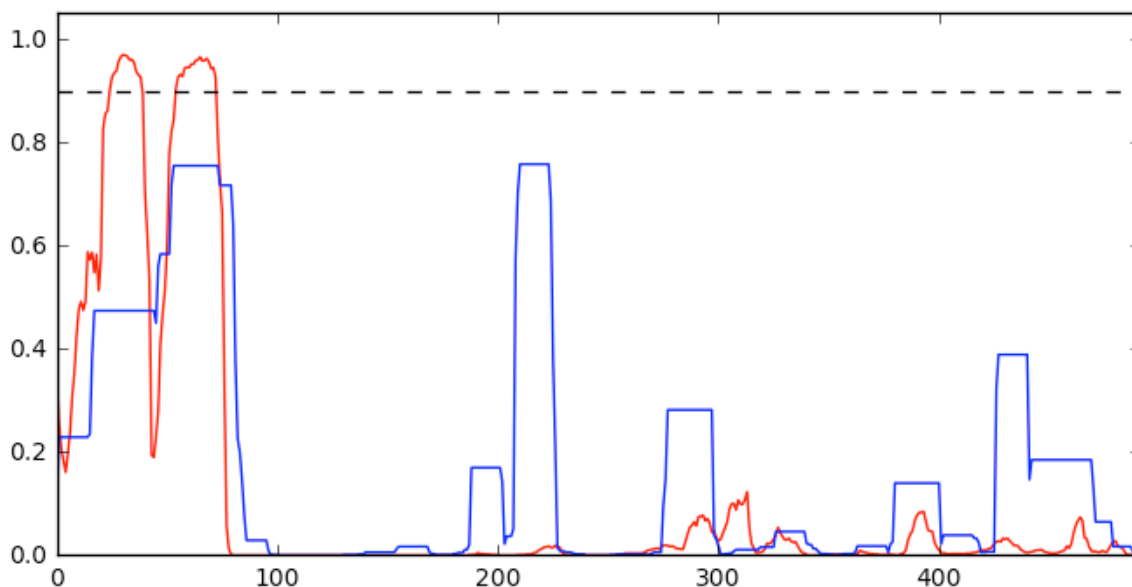
MEKSRMNLPGKPD TLCFDKDEFMKEDFDVDHFVSD **CRKRVQLEELRDDLELYKLLKTAMVELINKDYADFVNLSTNLVG**
MDKALNQLSVPLGQLREEVLSLRSSVSEGIRAVDERMSKQEDIRKK**KMCVLRLLIQVIRSV**EKIEKILNSQSSKETSAL
 SSPLLTGQILERIATEFNLQFHAVQSKGMPLLDKVRPRIAGITAMLQQSLEGLLLEGLQTSVDVIIRHCLR
 TYATIDKTRDAEALVGQVLVKPYIDEVIEEQFVESH
 PNLQVMYNKLLFVPHHCRLRLREVTGGAISS
 SEKNTVPGYDFLVNSVWPQIVQGLEEKLP
 SLFNPGNPDAFHEKYTISMDFVRRLERQCG
 SQASVKRLRAHPAYHSFNKKWNLVYFQIR
 FREIAGSLEAALTVDVLEDAPAESPYCL
 LASHRTWSSLRRCWSDMFLPLLVHRLWRL
 TLQILARYSVFVNELSLRPISNESPK
 EIKKPLVTSKEPSITQGNTEQSGSPSETK
 PVS**ISRTQLVYVVADLDKLEQLPELLEI**
 IKPKLEMIGFKN**FSSISALED**SQSSFS
ACVPSLSKIIQDLSDSCFGFLKSALE
 VPRLYRRTNKEVPTTASSYVDSALKPL
 FQLQSGHKDKLKQAI IQQWLEGLTSE
 ST**HKYYETVSDV**LN**SVK**MEESLRLKQ
 ARKTPANPVGPGSGMSDDDKIRLQLAL
 DVEYLGEQIQKLGQLQASDIKSFSA
 LAELVAAAKDQATAEQP



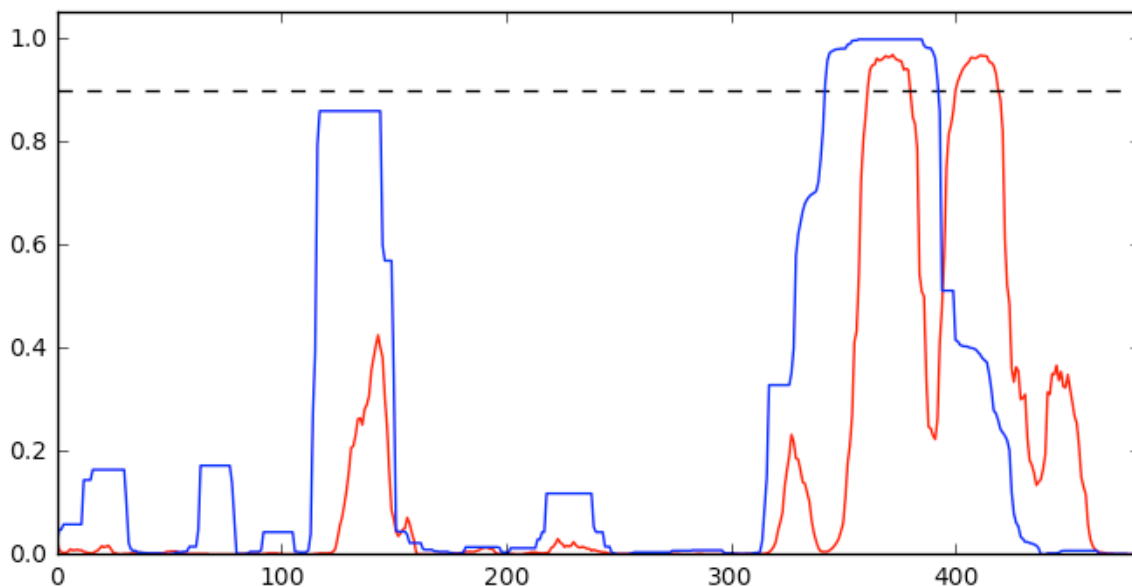
>NP_001158009.1 NH

MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFR
 RRRRLARRPGYMRSSSTGPGIGFLSPAVGTLFR
 FPGGVSGEESHSESRARQCGLDSRGLLVRSPV
 SKSAAAPT VTSVRGTS AHFGIQLRGGTRL
 PDRLSWPCGPGSAGWQOEFAMDSSETLDASW
 EAACSDGARRVRAAGSLPSAELSSNSCSPGCG
 PEVPTPPGSHSAFTSSFSFIRLSLGSAGERGE
 AEGCPPSREAESHCQSPQEMGAKAASLDGPH
 EDPRCLSRPFSLLATRV SADLAQAARNSSRP
 ERDMHSLPDMDPGSSSLDPSLAGCGGDGSSGS
 GDAHSWDTLLRKWEPVLRDCL**LRNRRQMEVI**
SLRLKLQKLQEDAVENDDYDKVETGFHYV
 GQAGLELLTSSNPP**PASASQS**
AGITAET**LQOR**LE**LE**QEKISLHFQLPSRQ
 PALSSFLGHLAAQVQAALRRGATQQASGDDT
 HTPLRMEPRLLE**PTAQDSL**
HVSITRRDWLLQEKQQLQKEIEALQARMFV
LEAKDQQLRREIEEQEQQLQWQCDL**TPLV**
QLSLGLQEVSKALQDTLA
SAGQIPFHAEP**PETIRSLQERIKSLNLSL**
KEITTKVCMSEKFCSTLRKKVNDI**ETQL**
PALLEAKMHAISGNHFWTAKDLT

EEIRSLTSEREGLEGLLSKLLVLSSRNVKLGSVKEDYNRLRREVEHQETAYETSVKENTMKYMETLKNKLCSCCKPLLK
 KVWEADLEACRLLIQSLQLEARGSLSVEDERQMDDEGAAPPPIPPRLHSEDKRKTPLKVLEEWKTHLIPSLHCAGGEQK
 EESYILSAELGEKCEDIGKLLYLEDLQHTAIHSHDEDLIQSLRRELQMVKETLQAMILQLQPAKEAGEREAAASCMTAG
 VHEAQA

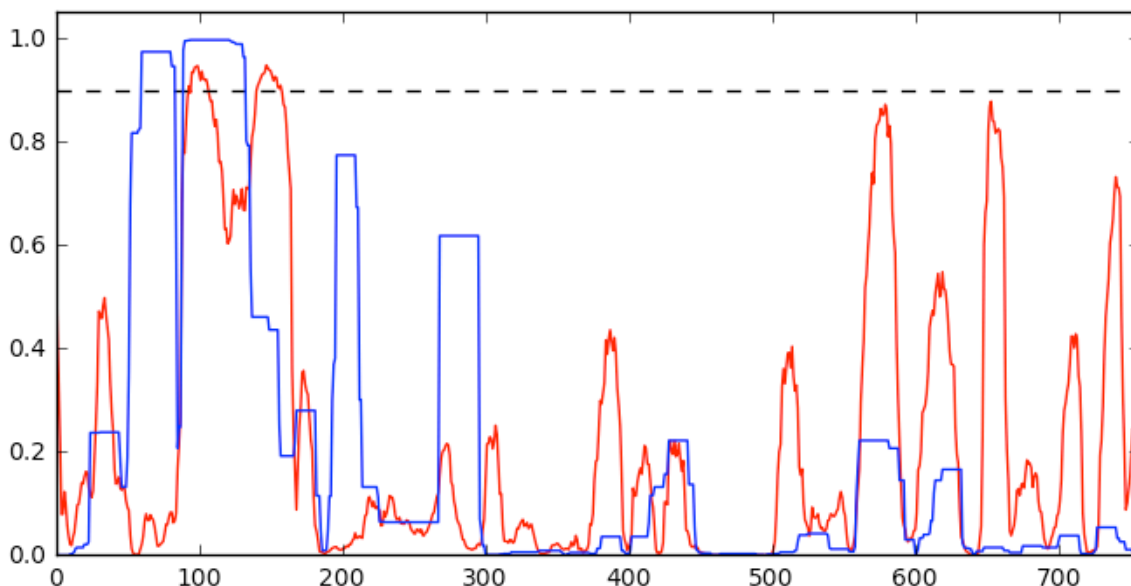


>NP_008975.1 H+
 MSLLMISENVKLAREYALLGNYDSAMVYYQGVLDQMNKYLYSVKDTYLQOKWQVWQEIINVEAKHVKDIMKTLESFKLDS
 TPLKAAQHDLPASEGVEVWSPVPVERRPSPGPRKRQSSQYSDPKSHGNRPSTTVRVHRSSAQNVHNDRGKAVRCREKKEQ
 NKGREEKNKSPAAVTEPETNKFSTGYDKDLVEALERDIISQNPVNRWDDIADLVEAKKLLKEAVVLPMMPEFFKGIRR
 PWKGVLMVGPPTGKTLAKAVATECKTTFNVSSTLTSKYRGESEKLVLLFEMARFYSPATIFIDEIDSICSRGTS
 EEHEASRRVKAELLVQMDGVGGTSENDDPSKMVMVLAATNFPWDIDEALRRRLEKRIYIPLPSAKGREELLRISLRELEL
 ADDVDLASIAENMEGYSGADITNVCRDASLMAMRRRIEGLTPEEIRNLSKEEMHMPTTMEDEFEMALKKVSXSVSAADIER
 YEKWIFEFGSC

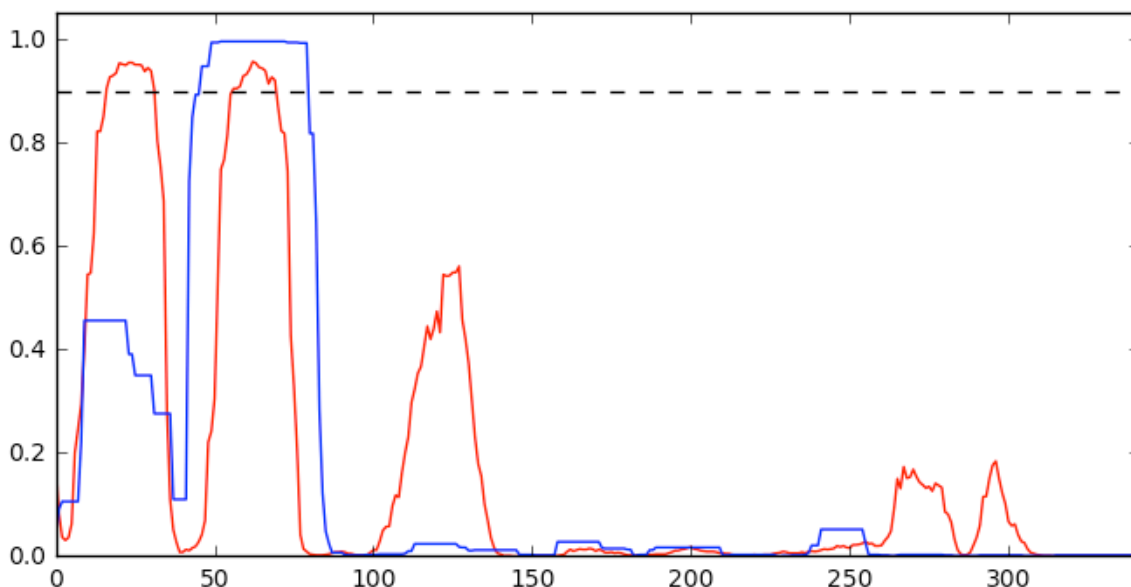


>XP_016881021.1 NH
 MNTEAEQQLLHHARNGNAEEVRQLLETMARNEVIADINCKGRSKSNLGTPLHLACYFGHRQVVQDLLKAGAEVNVLNDM
 GDTPLHRAAFTGRKELVMLLLEYNADTTIVNGSGQTAKEVTHAEERMSLEAVERTQQRKLEELLAAAREGKTTELTAL
 LNRPNPPDVNCSQDLGNTPLHCAAYRAHKQCALKLLRSGADPNLKNKNDQKPLDLAQAEMKHILVGNKVIYKALKRYEG
 PLWSSRFFGWRLFWVVLHGVLSWYRKQPDVHNIYRQGCKHLTQAVCTVKSTDSCLFFIKCFDDTIHGFRVPKNSLQQ

SREDWLEAI **EEHSAYSTHYCSQDQLTDEEEEDTVSAADLKKSLEKAQSCQQR**LDREISNFLKMIKECDMAKEMLP SFLQK
VEVSEASRETCVALTDCLNLFTKQEGVAVTFRCNKEKPVSFHVLFSRKY LKRNLC TTQGEKGSKTGLWHGVLEERNPDG
 KQ

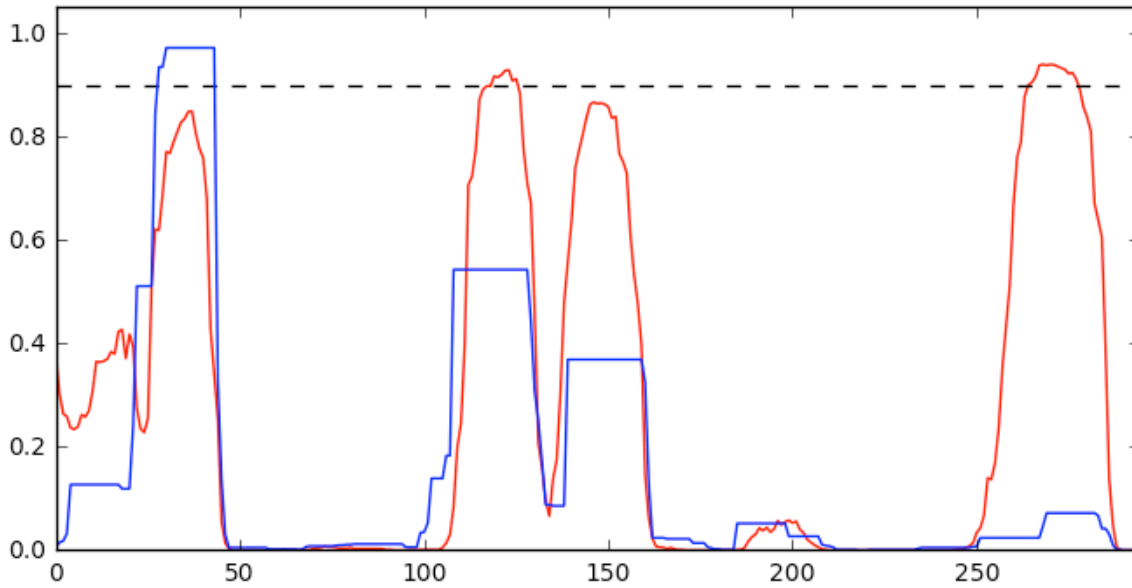


>XP_016871837.1 H-
 MCCVGGAGSCSDTGKLGACRAMLEEETDQTYENVLAEIQSFELPVEATLRSV**YDDQPNAHKKFMEKLDACIRNHDKEIEK**
MCNFHHQG**FVDAITELLKVRTDAEKLK**VQVTD TNRRFQDAGKEVIVHTEDI IRCRIQQRNI **TTVVEK LQLCLPVLEMYSK**
 LKEQMSAKRYYSALKTMEQLENVYFPWVSQYRFCQL**MIENLPKLREDIKEIS**MSDLKDFLESIRKHSDKIGETAMKQAQH
 QKTF SVSLQKQNKMKFGKNMYINRDRI**PEERNETVLKHSLEEEDENEEIILTVQDL**VDFSPVYRCLHIYSVLGDEETFEN
 YYRKQRKKQARLVLPQSNMHETVDGYRRYFTQIVGFFVVEDHILHVTQGLVTRAYTDELWNMALSKI IAVLRAHSSYCT
 DPDFLVELEKNLTVI FADTLQYGFVNRFLDLLFEIRDQYNETLLKKWAGVFRDIF EEDNYSPIPVNNEEYKIVISKFP
 FQDPDLEKQSFPPKFPMSQSVPHIYIQVKEFIYASLKFSESLHRSSTEIDDMLRKSTNLLLTRTLSSCLLNLIRKPHIGL
 TELVQIIINTHLEQACKYLEDFITNITNISQETVHTTRLYGLSTFKDARHAAEGE IYTKLNQKIDEFVQLADYDWTMSE
 PDGRASGYLMDLINFLRSIFQVFTHLPGKVAQTACMSACQHLSTSLMQMLLDSELKQISMGAVQQFNLDVIQCELFASSE
 VVPGFQGD TLQLAFIDLRQECLFALKQPILRKILPW

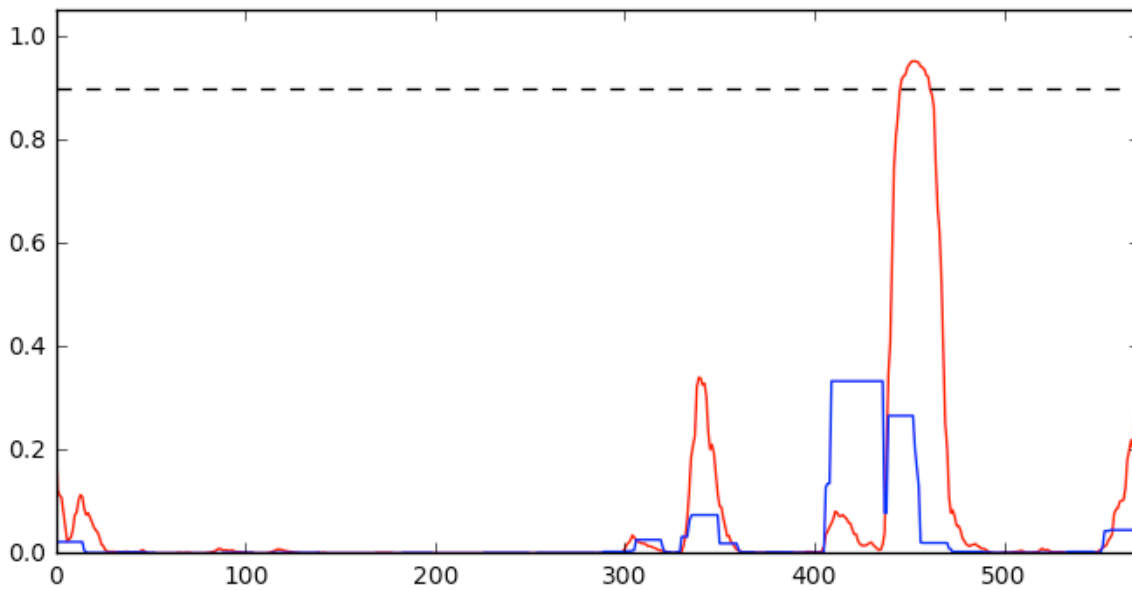


>XP_016870818.1 NH
 MADVINSVNLEAFSQ**AISAIQALRSSVSRVF**DCLKDGM RNK**ETLEGREKAFIAHFQDNLH SVNRDLNELERL**SNLVGKP
 SENHPLHNSGLLSLDPVQDKTPLYSQLQAYKWSNKLQYHAGLASGLLNQOQLKRSANQMGVSAKR RPKAQPTTLVLPPQ

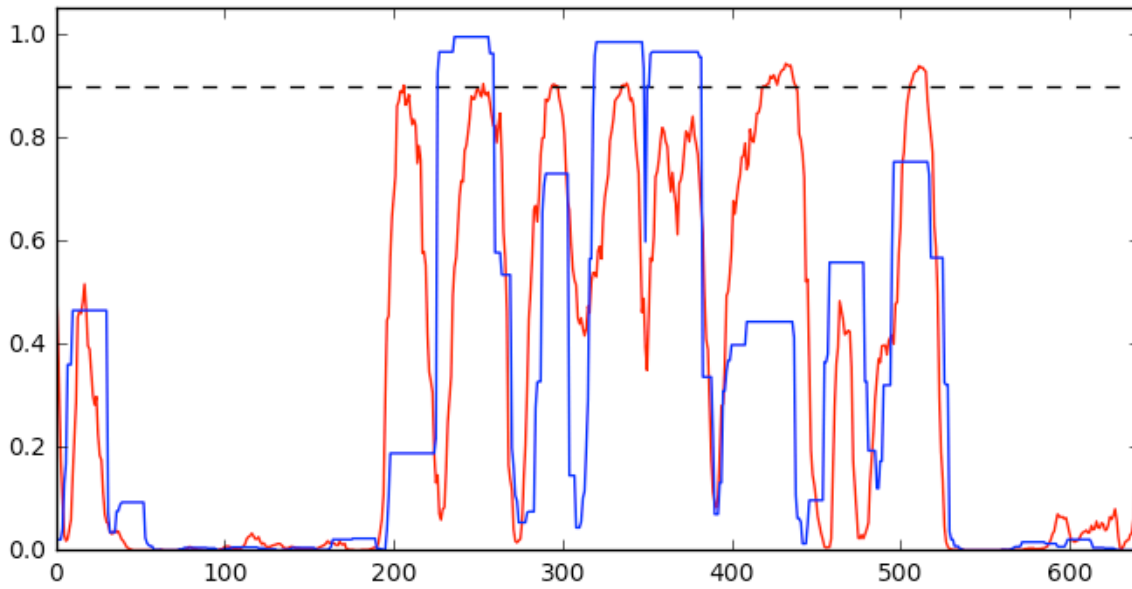
YVDDVISRIDRMFPMSIHLSRPNGTSAMLLVTLGKVLKVIIVMRSFLIDRTIVKGYNENVYTEDGKFFLCVKRSFDVLV
 FALRTDVSVLKQLVICKLDIWSKSNYQVFQKVTDHATTALLHYQLPQMPDVVVRSFMTWLRSYIKLQAPCQRCGKFLQD
 GLPPTWRDFRTLEAFHDTCRQ



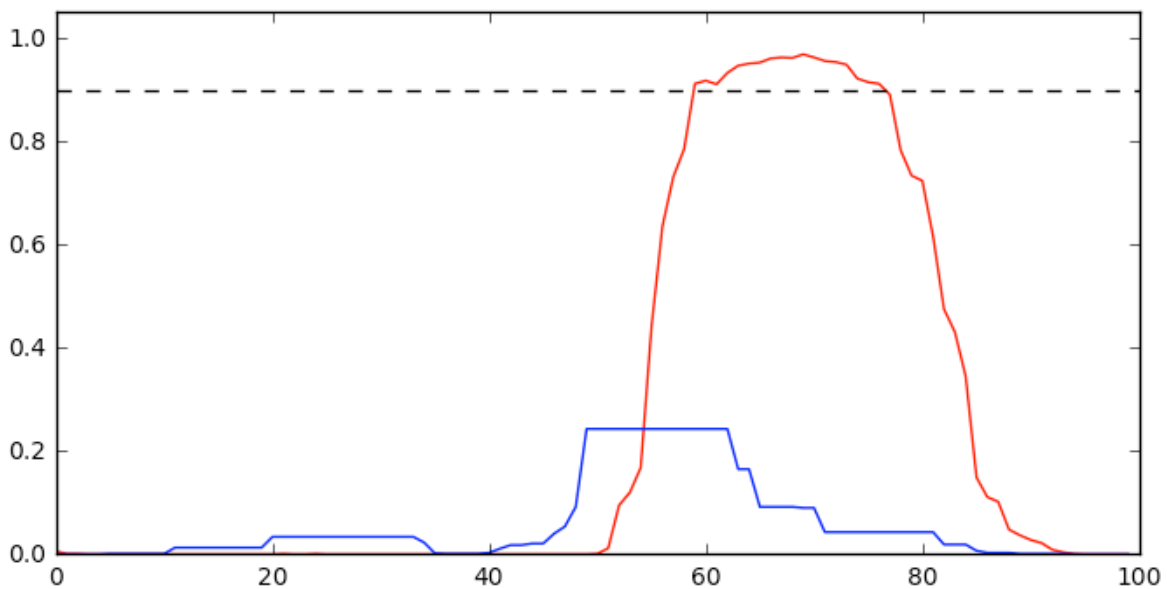
>XP_006721010.1 H+
 MATVMAATAAERAVLEEEFRWLLHDEVHAVLKQLQDILKEASLRFTLPGSGTEGPAKQENFILGSCGTDQVKGVLTLOGD
 ALSQADVNLKMPRNNQLLHFAFREDKQWKLQIQDARNHVSQAIYLLTSRDQSYQFKTGAEVLKLMDAVMLQLTRARNRL
 TTPATLTLPEIAASGLTRMFAPALPSDLLVNVYINLNKLCCLTVYQLHALQPNSTKNFRPAGGAVLHSPGAMLDPLTAHSE
 WGSQRLEVSHVHKVECVIPWLNDALVYFTVSLQLCQQLKDKISVFSSYWSYRPF



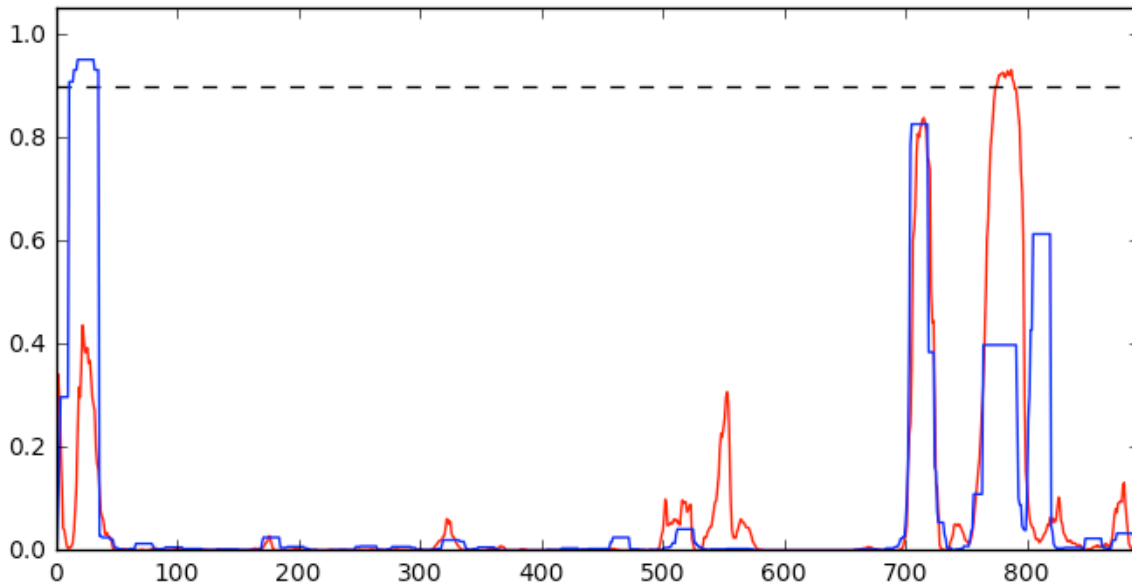
>XP_016868700.1 H-
 MDSLCKHLKPVQMLLPQVAVKTEPMSSSETASTTADGSLNFFSGSAIGSSSFSPRPTHQFSPQQIYPSNRYPHILPTP
 SSQTMAAAYGQTQFTTGMQQATAYATYPQPGQPYGISSYGALWAGIKTEGGLSQSQSPGQTGFLSYGTSFSTPQPGQAPYS
 YQMGGSSFTTSSGIYTGNNSLTNSSGFNSSQQDYPSYPSFGQGQYAQYINSSPYPAHYMTSSNTSPTTPSTNATYQLQEP
 PSGITSQAVTDPTAEYSTIHPSTPIKDSDSLRRGSDGKSRGRRRNNNPSPPPDSLRFVFIWDLDETIIVFHSLLT
 GSYANRYGRDPPTS VSLGLRMEEMI FNLADTHLFFNDLEECDQVHIDVSSDDNGQDLSTYNFGTDGFPAAATSANLCLA
 TGVRRGGVDWMRKLAFRYRRVKEIYNTYKNNVGGLLGPAKREAWLQLRAEIEALD SWLTLAL KALS LIHSRTNCVNILVT
 TTQLIPALAKVLLYGLGIVFPIENIYSATKIGKESCFERIIQRFGRKVVYVVIIGDGVVEEQGAKKHAMPFWRISSHSDLM
 ALHHALELEYL



>XP_016861433.2 H+
 MLSCLKEEMPPQELTRRLATVITHVDEIMQQEVRPLMAVEIIEQLHRQFAILSGGRGEDGAPIITFPFESGFKHIPDEDF
 LNVMTYLTSIPSVEAASIGFIVVIDRRRDKWSSVKASLTRIAVAFPGNLQLIFILRPSRFIQRTFTDIGIKYYRNEFKTK
 VPIIMVNSVSDLHGIDKSQLTRELGGTLEYRHGQWVNHRTAIENFALTLKTTAQMLQTFGSCLATA**AELPRSMLEDLL**
MSHTRQRDKLQDELKLLGKQGTTLLSCIQEPATKCPNSKLNLNQLENV**TTMERLLVQL**DETEKAFSHFWSEHHLKLN**QCL**
QLQHFEDFCKAKLALDN**LLEEQAFTGIGDSVMHVEQILKEHKKLEEK**SQEPLEKAQLLALVGDQLIQSHHYAADAIRP
 RCVELRHLCDDFINGNKKKWD**ILGKSLEFHRQLDKVSQW**CEAGIYLLASQAVDKCQSR**EGVDIALNDIATFLGTVKEY**PL
 LSPKEFYNEFELLT**LDAKAKAQVLQRLDDVQEIF**HKRQVSLMKLAQTRPVQPVAPHPESPKWVSSKTSQPSTSV
 LARPLRTSEEPYTETELNSRGKEDDETKFEVKSEEIFESHHERGNPELEQQARLGDLSPRRPGVVAQVCNPNTFQHFGR
 L

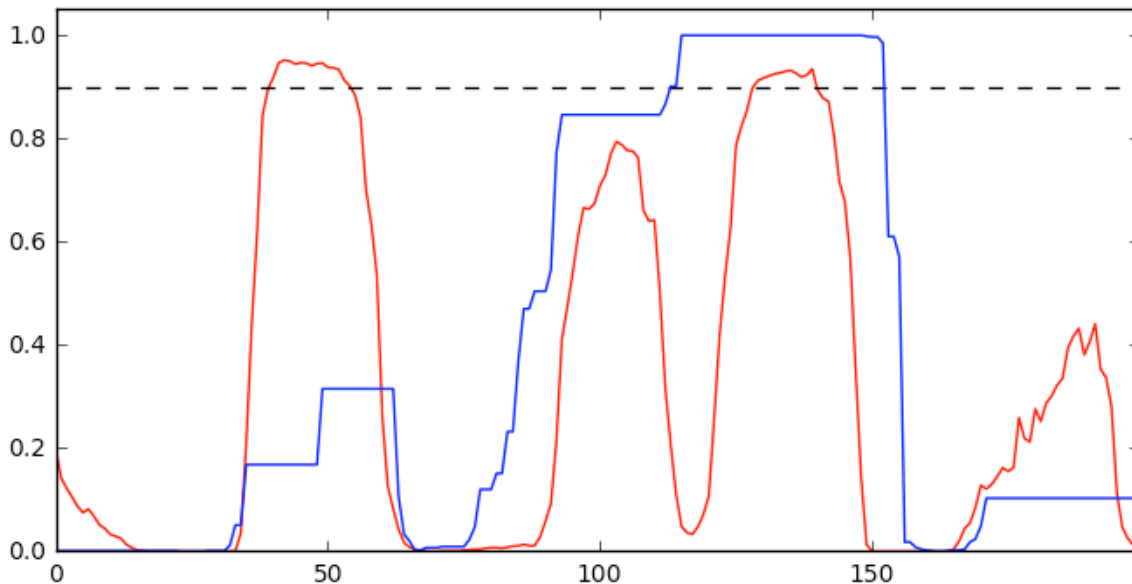


>XP_016858202.1 H+
 MPPKFKRHLNDDDDVTGSVKSERNNLEDDSDDEEDFFLRGSPGPRFGPRNDKIKHVQN**QVDEVIDVMQENITKKAYR**IMQ
 QLLATDPNNFEGKCGGVDAK



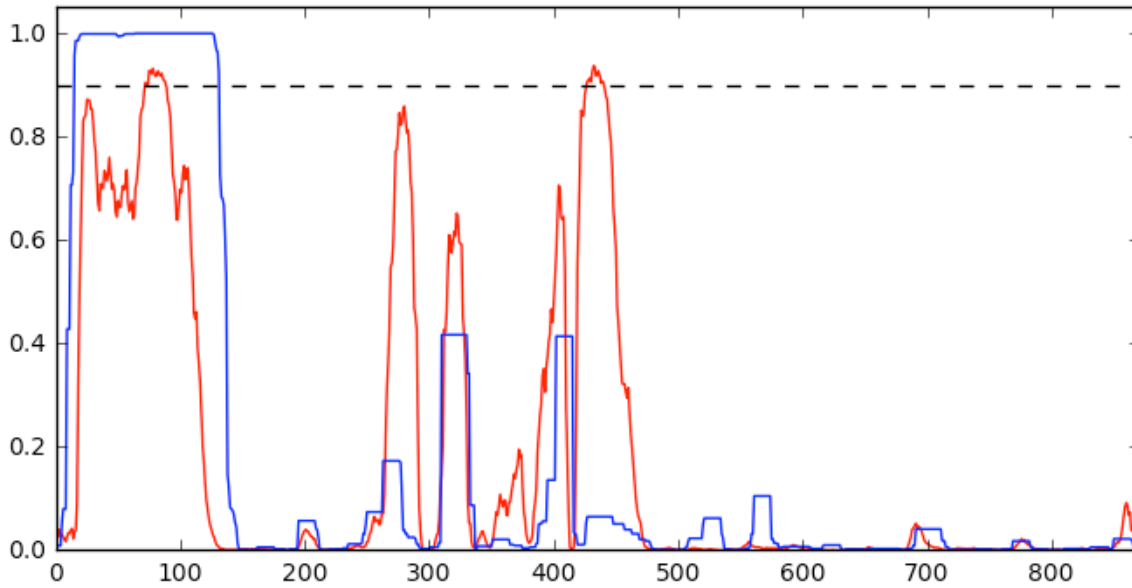
>XP_011512467.1 H+

MPSLLGAAMP**ASTSAALQEAL**ENAGRLIDRQLQEDRMPDLSSELLMVSAPNNPTVSGMSDMDYPLQGPGLLSVPNLPE
 ISSIRRVPLPPELVEQFGHMQCNCMMGVFPPIISRAWLTIDSDIFMWNYEDGGDLAYFDGLSETILAVGLVKPKAGIFQPH
 VRHLLVLATPVDIVILGLSYANLQTGSGVLNDSLGGMQLLPDPLYSLPTDNTYLLTITSTDNGRIFLAGKDGCLYEVAY
 QAEAGWFSQRCRKINHSKSSLSFLVPSLLQFTFSEDDPILQIAIDNSRNILYTRSEKGVIQVYDLGQDQGMSRVASVSQ
 NAIVSAAGNIARTIDRSVFKPIVQIAVIENSESLDCQLLAVTHAGVRLYFSTCPFRQPLARPNTLTLVHVRLLPPGFSASS
 TVEKPSKVHRALYSKGIILMAASENEDNDILWCVNHDTFFPQKPMMETQMTAGVDGHSWALSIDELKVDKIITPLNKDH
 IPTDSPVVVQQHMLPPKKFVLLSAQGLMFHKLRPVDQLRHLLVSNVGGDGEIEERFFKLHQEDQACATCLILACSTAA
 CDREVSAWATRAFFRYGGEAQMRFPPTLPPPSNVGPILGSPVYSSSPVPSGSPYPNPSFLGTPSHGIQPPAMSTPVCALG
 NPATQATNMSCVTGPEIVYSGKHNGICIFYSRIMGNIWDASLVVERIFKSGNREITAIESSVPC**QLLESVLQELKGLQEF**
 LDRNSQFAGGPLGNPNTTAKVQQRLIGFMRPENGNPQQMQQELQRKFHEAQLSEKIALALWKLCEHQFTI**IVAELQKEL**
QEQLKITTFKDLVIRDKELTGALIASLINCYIRDNAAVDGISLHLQDICPLLYSTDDAICSKVEKINKSVSHGLPAAQ



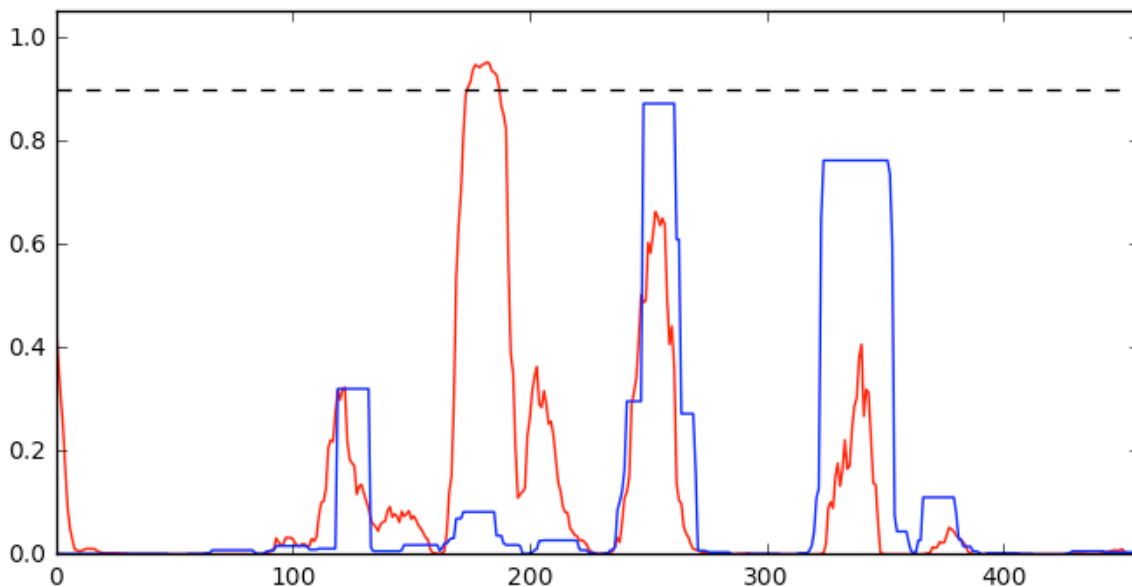
>NP_000632.1 H+

MNCVCRLVLVLSLWPD TAVAPGPPPGPPRVSPDPRAELD**STVLLTRSL**LADTRQLAAQLRDKFPADGDHNLDSLPTLAM
 SAGALGAL**QLPGVLR**LRADLLSYLRHVQWLR**RAGGS**SLK**TEPELGT**LQARLD**RLRL**QLLMSRLALPQPPDPFPAPP
 LAPPSSAWGGIRAAHAILGGLHLLTDWAVRGLLLLLKTRL



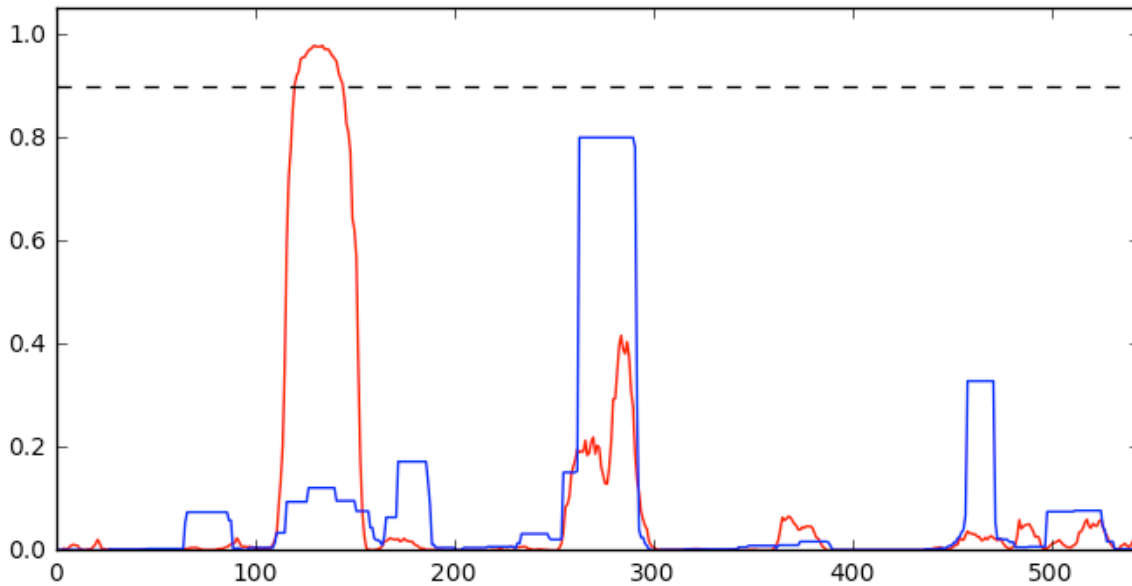
>NP_001138923.2 H+

MEKTKTKQGENE**HMPVNNPSTQIYQLQALASELKTGFTEAMQELSRIQHGEYALEEKVKSCRCSMEEKVTEMKNSLNYFK**
EELSNAMSMIQAITSKQEEMQOKIEQLQOQEKRRRESRKVKAKKTQKEEHSSQAGPAQAQGSPPFRSINIPEPVLPSDFTNL
 LPSQAYEKAQESRSVHVGDSNVKGMMPGVNPTTPEAEENLKSCLSADIQSKGHLPSGMWRQPKDGKEWGEEYVTKDHPD
 KLKEAGQGRHSSLENVLCETSLAAKRQTVALELLESERKYVINISLILKIKATFQGS DGKRN SKERSLFPGLRYLVQQH
 LDLHALQERVLKWPQGVLDLFLKLTNDENNFLDYVAYLRDLPECISLVHVVLKEGDEEIKSDIYTLFFHIVQRIP
 EYLIHLQNVLKFTEQEHPDYLLLVCV**QRLRVFISHYTL**LLFQCNEDLLIQKRKLLKSSMAKLYKGLASQCANAGQDASP
 TAGPEAVRDTGIHSEELLQPYPSAPSSGPAITHLMPPVKKSQQQQSLMESMQPGKPSDWELEGRKHERPESLLAPTQFCA
 AEQDVKALAGPLQAIPEMDFESSPAEPLGNVERSLRAPAELLPDARGFVPAAYEEFEYGGEIFALPAPYDEEPPQAPALF
 ENCSPASSESLDICFLRPVSFAMEAERPEHPLQPLPKSATSPAGSSSAYKLEAAAQAHGKAKPLSRSLKEFPAPPADG
 VAPRLYSTRSSSGGRAPIKAERAAQAHGPAAAVAARGASRTFFPQQRSQSEKQTYLEVRREMHLEDTRFCPKEERESE
 QTSFSDQNPRQDQKGGFRSSFRKLFKKKSSGSEYREKTNENPSMDPSPTKQDFFRNRLALANDLDQGTAV

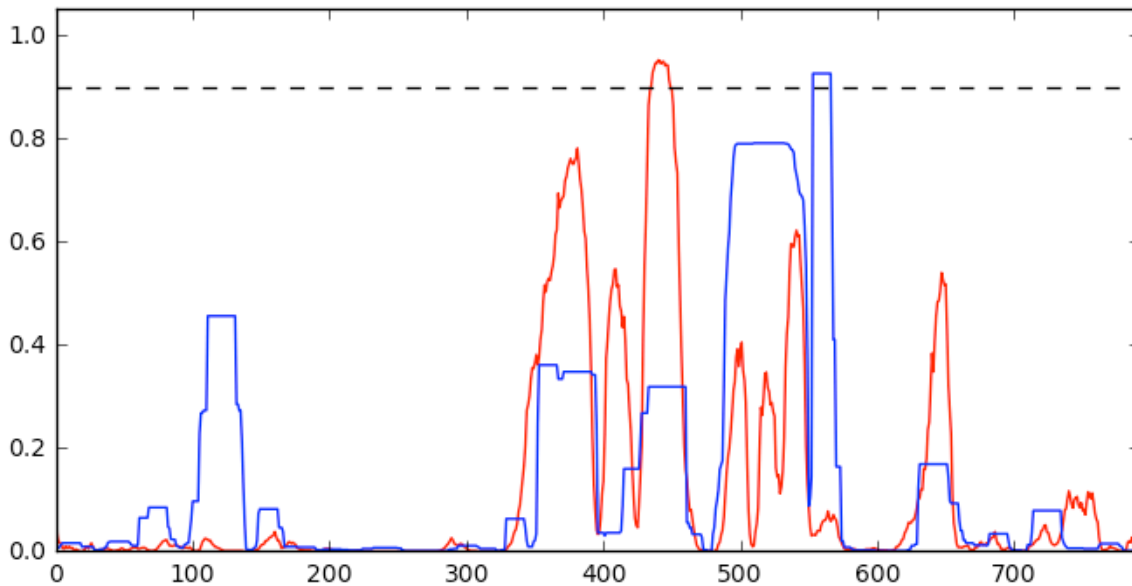


>NP_001159631.1 H+

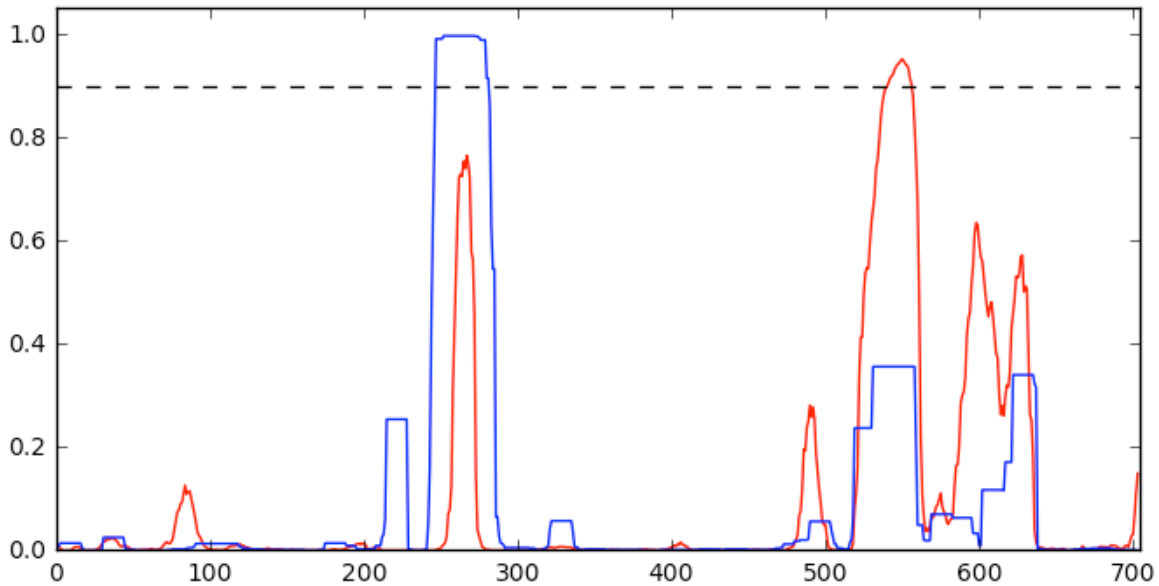
MAALRALCGFRGVAAQVLRPGAGVRLPIQPSRGVRQWQPDVEWAQQFGGAVMYPKETAHWKPPPWNVDVDPKDTIVKNI
 TLNFGPQHAAHGVLRLVMELSGEMVRKCDPHIGLLHRGTEKLI EYKTYLQALPYFDRLDYVSMCNEQAYS LAVEKLLN
 IRPPPRAQWIRVLF**GEITRLLNHIMAVT**THALDLGAMTPFFWLFEEEREKMF EYERVSGARMHAAYIRPGGVHQDLPLGL
 MDDIYQFS**KNFSLRLDELELLTN**NRIWRNRTIDIGVVTAAEALNYGFSGVMRLRGSGIQWDLRKTQPYDVYDQVEFDVPV
 GSR**GDCYDRYLCRVEEMRQSLRI**IAQCLNKMPPEIKVDDAKVSPPKRAEMKTSMESLIHHFKLYTEGYQVPPGATYTAI
 EAPKGEFGVYLVSDGSSRPYRCKIKAPGFAHLAGLDKMSKGHMLADVVAIIIGTRPIV



>NP_001230177.1 H-
 MRDNTSPISVILVSSGSRGNKLLFRYPFQRSQEHSPASQTSKPRSRYAASNTGDHADEQDGDSTRFSVDVILATILATKSEMC
 GQKFELKIDNRFVGHPTLLQHALGQANADPSVINCLHNL**SRRIATVLQHEERRCQYLTREAKLILALQ**DEVSAMADGNE
 GPQSPFHILPKCKLARDLKEAYDSLCTSGVRLHINSWLEVSFCLPHKIHYAASSLIPPEAIERSLKAIRPYHALLLS
 DEKSLGELPIDCSPALVRVIKT**TSAVKNLQQLAQDADLALLQVFQLAA**HLVYWGKAI I IYPLCENNVYMLSPNASVCLY
 SPLAEQFSHQFP SHDLPSVLAKFSLPVSLSEFRNPLAPAVQETQLIQMVVWMLQRLLIQLHTYVCLMASPSEEEPRPRE
 DDVPTARVGGRLSTPNALSFGSPTSSDDMTLTSPSMDNSSAELLPSGDSPLNQRMTENLLASLSEHERAAILSVPAAQ
 NPEDLRMFARLLHYFRGRHHLEEIMYNENTRRSOLLMLFDKFRSVLVVTTTHEDPVI AVFQALLP

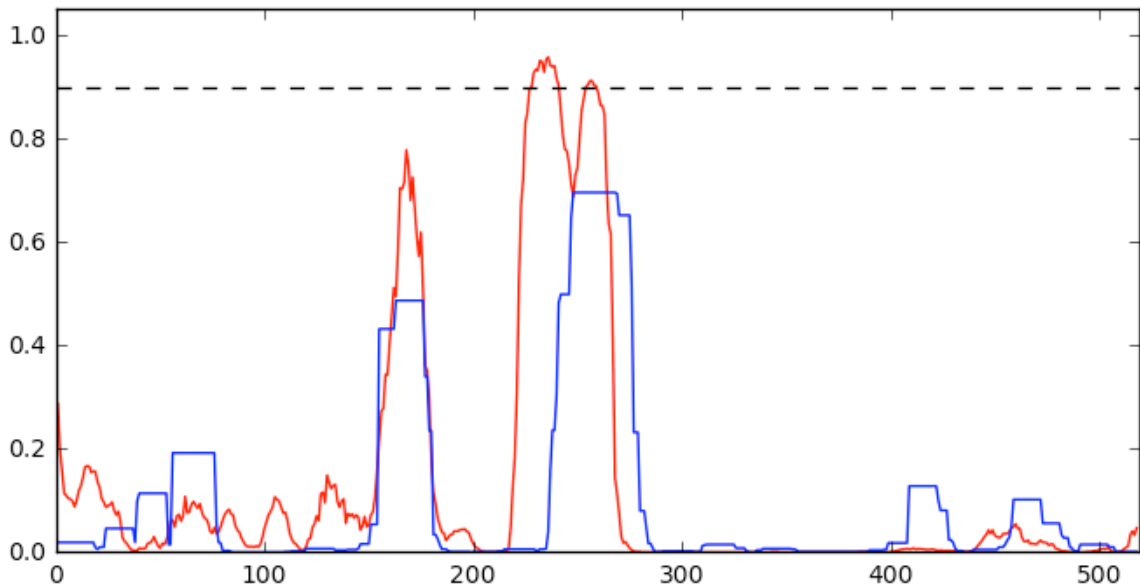


>NP_001269784.1 NH
 MSNQQEKYEAQNI VNSTEESDDAFDVTI PVPSEEPQESDQTEEHESGIEQFSESHAIHVEEQSDQSFSSLEPDNEQLME
 EVISPRQVS YTPQHHEKQYAMQRPNDDSLAF LDKIKSVKESLQESVEDSLATVKVVLIPVQGEI V I PFKVDTILKYLKDH
 FSHLLGI PHSVLQIRYSGKILKNNETLVQHGVKQEI VQVEI FSTNPDLYPVRRIDGLTDV SQI I T V T V Q T G L D Q Y Q V P
 VEIVKSDFHK PFLGGFRHKV TGV EYHNAGTQTVPKRI PERLSIFCRDTQT V FQKKNLQQT T N T T S T Q M T N I G V Y V S N M T D
 KLVTPGKYFSAAEYHAQRLKAVI V I Q T Y Y R Q W H A K I F V E N L R R Q K S L R L E W E T Q Q E L R K I R E K E E W I K L D Y H R R H N P K T N
 E D F E F L Y N A L E F W R Q E E L T R I N Q S F T G A E R K A A L C E L L E K E T Q I I A S I G R H R Y I A Y M A N Q E A A I Q A F L D K C S A P K I W R T P
 N G K T I E M D T Q **FTIRARELQNIYKCIMLNISQDERLDVLLTLKHTVKEHECKLTQEILELIDREVDLMMRGVKHHNLEGL**
RKRIATLFFHYIKTPLFNPEVAKYLKVPQDPLKFKYKIIYFCHSCQLYLPSTEFVSSTSRRIYRCRNCINLQNEAQKRES
 FLKYKCLLQQLLYTEADYEDDSKIAFLMQLQDIQYL TENI WASQSVLSACDNLSDLVMVRWNKSLEWSPWNCILLTKDEA
 AAHLKLT SIEEGYERSFIHKIKHKHILAKNYFSQVPVLASFILDDGEIDEIRWKYHSDTTPKIIESQRPPH



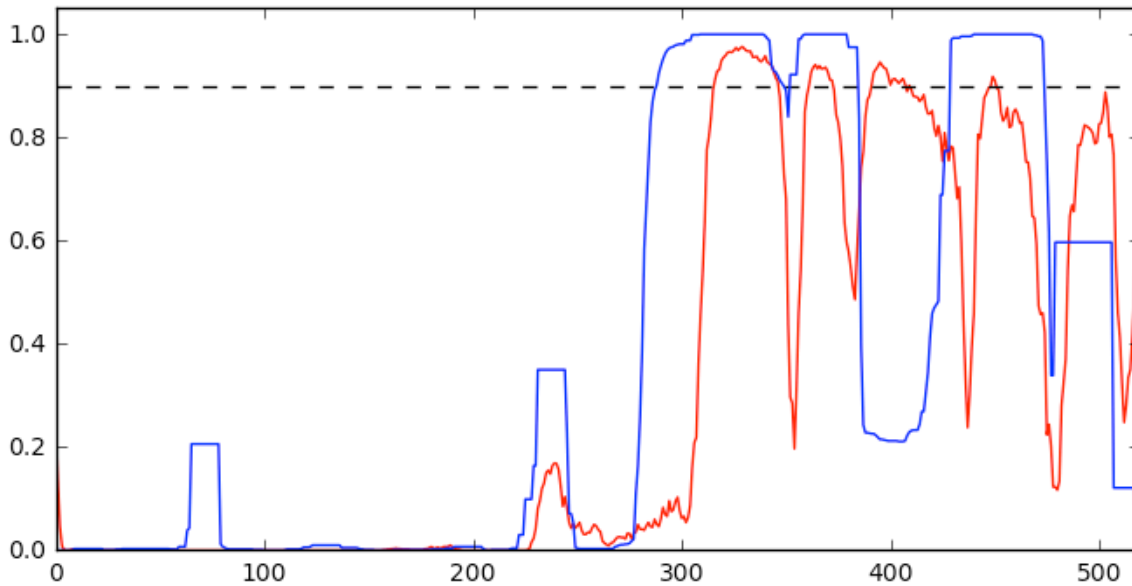
>NP_001334966.1 NH

MKLMYEHNLQRTACNMTYGSFSGVKGKRDLCIQSMDGMLMVFEQESYAFGRFLPGFLLPGPLAYSSRTDSFLTIVSSCQV
 ESYKYQVLA FATDADKRQETEQQKLGSGKRLVVDWTLNIGEALDICI VSNQSSASSVFLGERNFFCLKDNGQIRFMKK
 LDWSPSCFLPYCSVSEGTINTLIGNHNNMLHIYQDVTWKWATQLPHIPVAVRVGCLHDLKGVIVTLSDDGHLQCSYLGTD
 PSLFQ**APNVQSRELNYDEL DVEMKELQKI IKDVNKS**QGVWPMTEREDDLNVSVVVSPNFDSVSQATDVEVGTDLVPSVTV
 KVTLQNRVILQKAKLSVYVQPPLELTCQDQFTFEMTPDLTRTVSFSVYLKRSYTPSELEGNVVSYSRPTDRNPDGIPRV
 IQCKFRLPLKLI CLPGQPSKTASHKITIDTNKSPVLLSLFPGFASQDDDDQVNVMGFHFLLGGARITVLASKTSQRYRIQ
 SEQFEDLWLI TNELILRLQEYFEKQGVKDFACFSGSIPLQEYFELIDHDFELRINGEKLE**ELLSERAVQFRAIQRRLLA**
 RFKDKTPAPLQHLDTLLDGTYKQGWEE TVDAAISHLLKTCLSKSSKEQALNLSQLNIPKDTSQLKKHITLLCDRLSKGG
 RLCLSTDAAPQTMVMPGGCTTIPESDLEERSVEQDSTELFTNHRHLTAETPRPEVSPLOGVSE



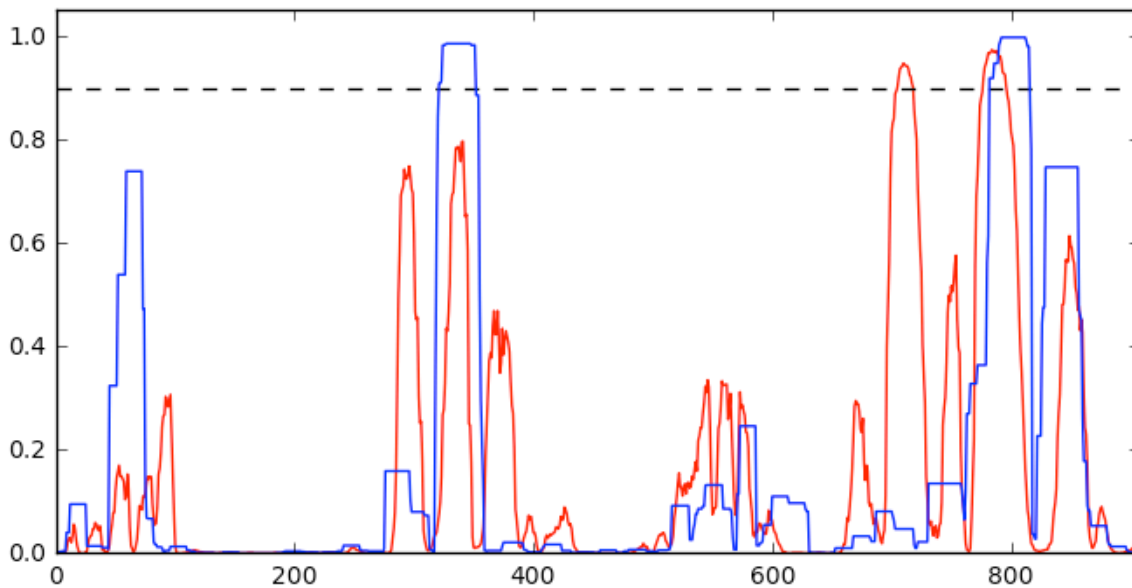
>NP_001340522.1 H+

MLIVLNSTLKSFDQFTCNLLYVSWRKDLTEHLHRLYFRGRAYYTLNVLRDDIDNPDQIRISQDVERFCRQLSSMASKLIIS
 PFTLVYYTYQCFQSTGWLGPVSI FGYFILGTVVNKTLMGPIVMKLVHQEKLEGDFRFRKHMQIRVNAEPAAFYRAGHVEHM
 RTDRRLQRLQOTQRELMSELWLYIGINTFDYLGSI LSYVVI AIPISGVYGDLSPAELSTLVSKNAF**VCIYLI SCFTQL**
IDLSTTLSDVAGYTHRIGQLRETTLLDMSLKSQDCEILGESEWGLDTPPGWPAEAPADTAFLLRVSI SAPSSDKPLIKDL
 SLKISEGQSLITGNTGTGKTSLLRVLGGLWTSTRGSVQMLTDFGPHGVFLFPQKPFFTDGLTREQVIYPLKEVYPDSGS
 ADDERILRFLELAGLSNLVARTEGLDQQVDWNWYDVLSPGEMQRLSFARLFYLPKYAVLDEATSALTEEVESELYRIGQ
 QLGMTFISVGHRSLEKFSLSVLKLCGGGRWELMRIKVE



>NP_003091.2 H+

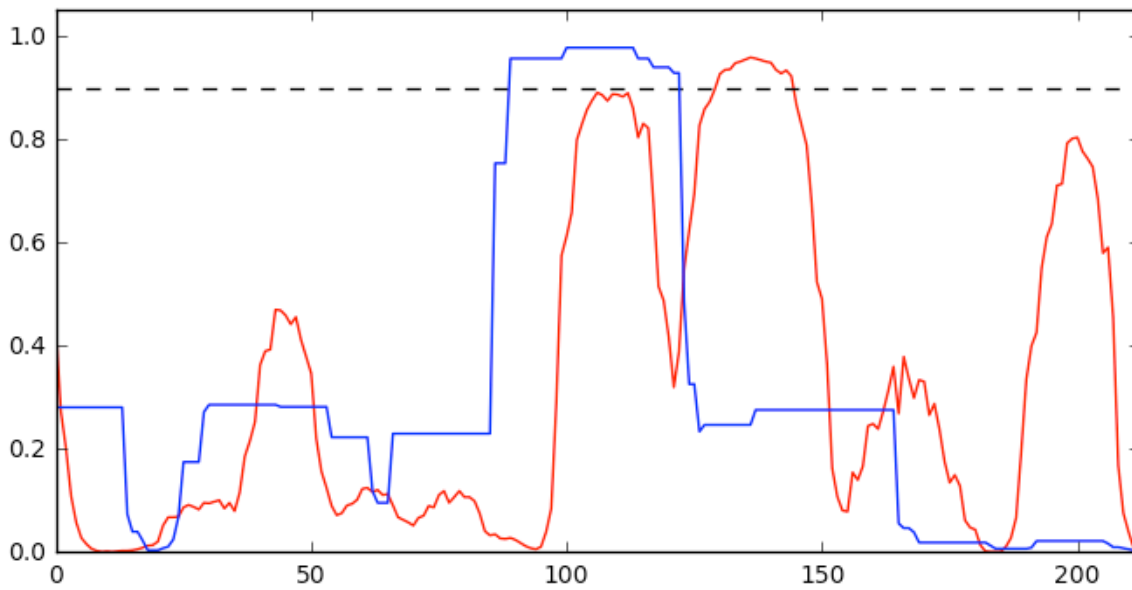
MAAEREPPPLGDGKPTDFEDLEDGEDLFTSTVSTLESSPSSPEPASLPAEDI SANSNGPKPTEVVLDDDDREDLFAEATEE
 VSLDSPEREPILSSEPSPAVTPVTPPTTLIAPRIESKMSAPVIFDRSREEIEEEEANGDIFDIEIGVSDPEKVGDMNAYM
 AYRVTTKTSLSMFSKSEFSVKRRFSDFLGLHLSKASKYLHVGYIVPPAPEKSIVGMTKVKGKEDSSSTEFVEKRRAALE
 RYLQRTVKHPTLLQDPDLRQFLESSELPRAVNTQALSGAGIL**RMVNKAADAVNKMTIKMNESDAWFEEKQQQFENLDQQL**
RKLHVSVEALVCHRKELSANTAFAKSAAMLGNSDHTALSRALSQLAEEVEEKIDQLHQEQAFADFYMFSELLSDYIRLI
AAVKGVFDHRMKCWQKWEDAQITLL**LKKREAEAKMMVANKPKDIQQAKNEIREWEAKVQQGERDFEQISKTIRKEVGRFEK**
ERVKDFKTVI IKYLESLVQTOQQQLIKYWEAFLPEAKAIA



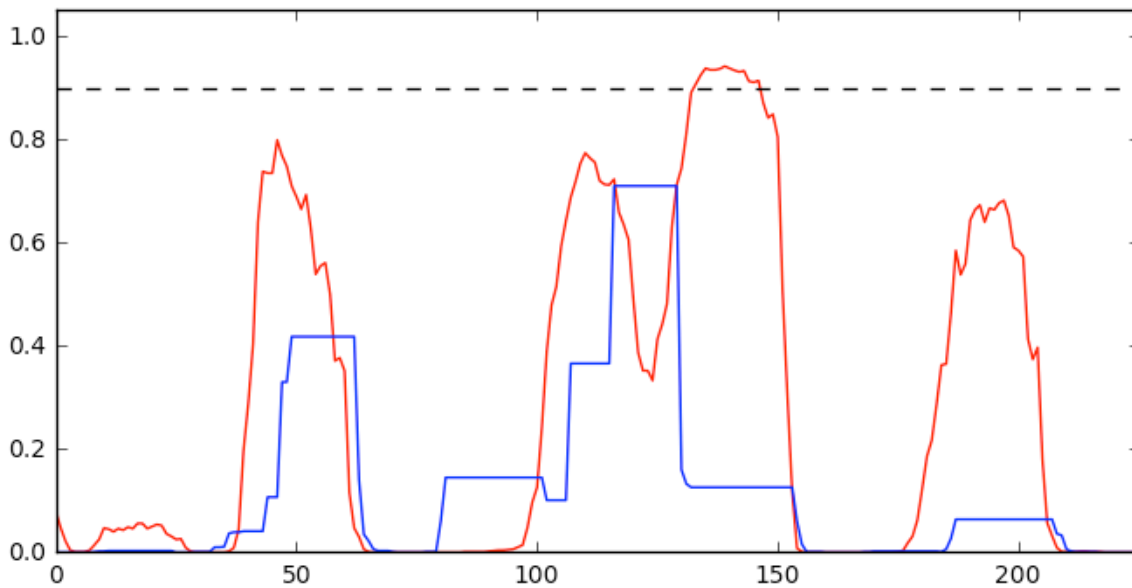
>NP_006313.1 H+

MATPDQKSPNVLLQNLCCRILGRSEADVAQQFQYAVRVIGSNFAPTVERDEF**LVAEKIKKELIRQPREADAAL**FSELHRK
 LHSQGVLKNKWSIYLILLSLSEDPRRQPSKVSSYATLFAQALPRDAHSTPYYYARPQTLPLSYQDRSAQSAQSSGSGVSS
 GISSIGLCAISGPAPAPQSLLPGQSNQAPGVGDCLRQQLGSR LAWLTANQPSSQATTSGKVPSAVSRNMTRSRRREGDTG
 GTMEITEAALVRDILYVFQIDGKNIKMNNTEENCYKVEGKANLSRSLRDTAVRSELGWLHNKIRRYTDQRSLDRSFGLV
GQSFCAALHQELREYYRLLSVLHSQQLEDDQGVNLGLESSLTLRLLVWTDPKIRLKTALALVDHCQGRKGGELASAV
 HAYTKTGDPYMRSLVQHILSLVSHPVLSFLYRWIYDGELEDTYHEFFVASDPTVKTDRLWHDKYTLRKSMI PSFMTMDQS
 RKVLLIGK SINFLHQVCHDQTPTTKMIAVTKSAESPQDAADLFTDLENAFQGKIDAAYFETSKYLLDVLNKKYSLLDHMQ
 AMRYYLLGQGFIRHLMDLLKPELVRPATTLYQHNLTGILETAVRATNAQFDSPEILRRLDVRLLLEVSPGDTGWDVFSL
 DYHVDGPIATVFTRECM SHYLRVFNFLWRAKRMEYILTDIRKGHMCNAKLLRNMPFSGVLHQCH**ILASEMVHF IHQMQY**
 YITFEVLECSWDELWNKVQQAQDL DHI IAAHEVFLDTII SRCLLSDSRALLNQLRA**VFDQI IELQNAQDAIYRAALEEL**

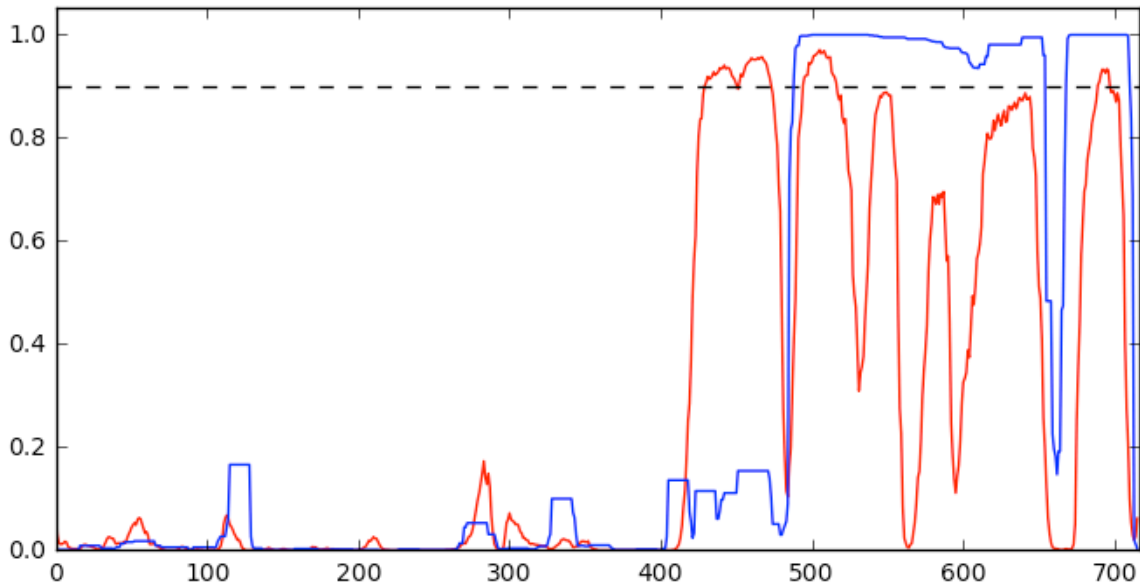
QRRLQFEEKKQREIEGQWGVTAEEEEENKRIGEFKESIPKMCSQLRILTHFYQGI VQQLVLLTTSSDES LRFLSFRL
DFNEHYKAREPRLRVSLGTRGRSSHT



>NP_009148.2 H+
MRMTMEEMKNEAETTSMVSMPLYAVMYPVFNELERVNLSAAQTLRAAFIKA EKENPGLTQDIIMKILEKKSVEVNFTESL
LRMAADDVEEYMIERPEPEFQDLNEKARALKQILSKI PDEINDRVRFLQT IKDIASAIKELLDTVNNVFKKYQYQNRAL
EHQKKEFVKYSKSFSDLKTYFKDGKAINV FVSANRLIHQTNLILQTFKTVA

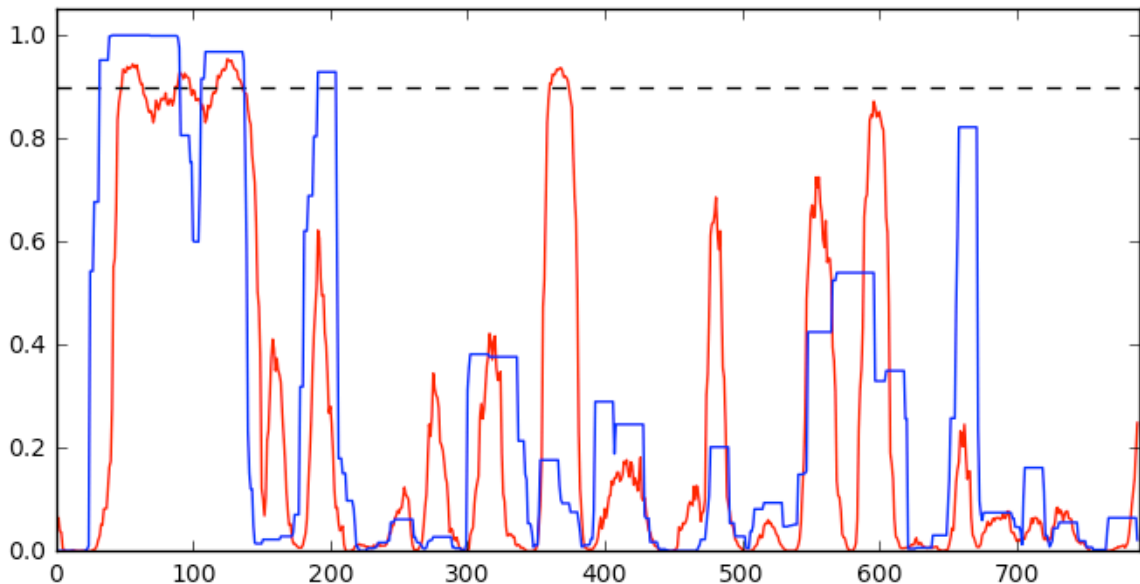


>NP_037378.1 H+
MDLRAGDSWGLACLCTVLWHLPAVPALNRTGDPGPGPSIQKTYDLTRYLEHQLRSLAGTYLNYLGPPFNEPDFNPPRLG
AETLPRATVDLEVWRS LNDKLRLTQNYEAYSHLLCYLRGLNRQAATAELRSLAHFCTSLQGLLGS IAGVMAALGYPLPQ
PLPGTEPTWTPGPAHSDFLQKMDDFWLLKELQTWLWRS AKDFNRLKKKMQPPAAAVTLHLGAHG F



>NP_054721.1 H+

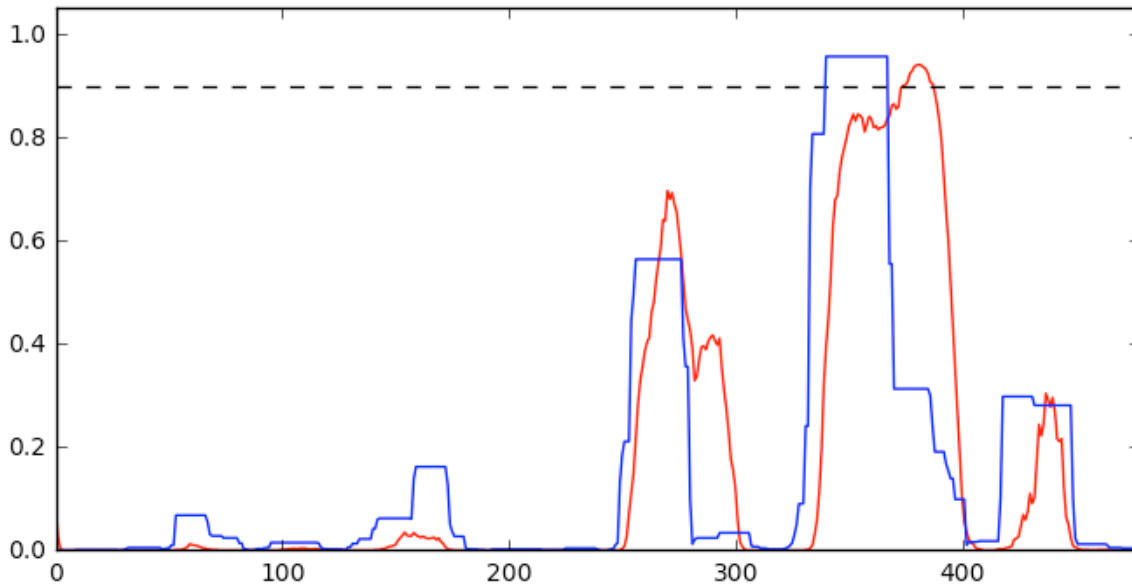
MQSTANYLWHTDDLLGQGATASVYKARNKKSSELVAVKVFNTTSYLRPREVQVREFEVLRLKLNHQNIVKLFAVEETGGSR
 QKVLVMEYCSSGSLLSVLESPENAFGLPEDEFVLRVLCVAGMNLRENGIVHRDIKPGNIMRLVGEEGQSIYKLTDFGA
 ARELDDDEKFSVYGTTEYLHPDMYERAVLRKPKQKAFGVTVDLWSIGVTLYHAATGSLPFIPIFGGPRNKEIMYRITTE
 KPAGAIAGAQRRENGPLEWSYTLPIITCQLSLGLQSQLVPILANILEVEQAKCWGFDQFFAETSDILQRVVVHVFSLSQAV
 LHIIYIHAHNTIAIFQEAVHKQTSVAPRHQEYLFEGHLCVLEPSVSAQHIAHTTASSPLTLFSTAIPKGLAFRDPALDVP
 KFVPKVDLQADYNTAKGVLGAGYQALRLA**RALLDQELMFRGLHWVMEVLQATCRRTLEVARTSLLYLSSSLGTERFSSV**
AGTPEIQELKAAEELRSRLRTLAEVLSRCSQNTITETQESLSSLNRELVKSRDQVHEDRSIQIQCCLDKMNFIYKQFKKS
RMRPGLGYNEEQIHKLDKVNFSHLAKRLLQVFQEECVQKYQASLVTHGKMRVHVHETRNHLRLVGCVAACNTEAQGVQE
SLSKLLEELSHQLLQDRAKGAQASPPPAPYPSPTRKDLLLHMQELCEGMKLLASDLLDNNRIIERLNRVPAPPDV



>NP_056201.2 H+

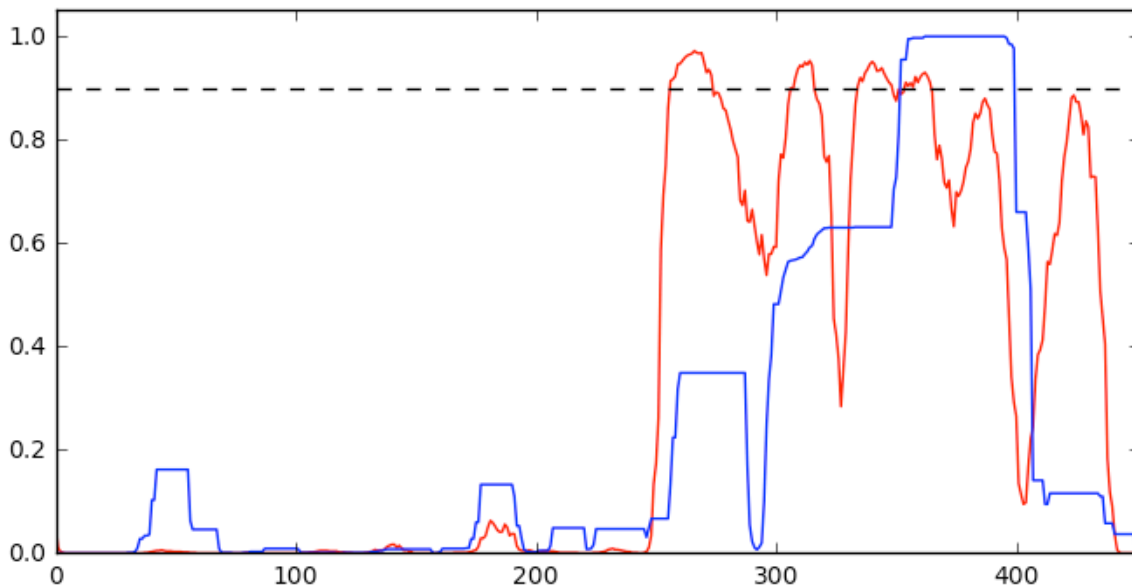
MGTKMADLDSPPKLSGVQQPSEGVGGRCSEISAEILRSLTELQELEAVYERLCGEEKVVERELDALLEQQNTIESKMVT
LHRMGPNLQLIEGDAKQLAGMITFTCNLAENVSSKVRQLDLAKNRLYQAIQRADDILDLKFCMDGVQ TALRSEDYEQAAA
 HTHRYLCLDKSVIELSRQKESGSMIDANLKLQEAQRLKAIVAEKFAIATKEGDLQPVERFFKIFPLLGLHEEGLRKFS
 EYLCKQVASKAEENLLMVLGTDMSDRRAAVIFADTLTLLFEGIARIVETHQPIVETYYGPRLYTLIKYLQVECDRQVEK
 VVDKFIKQRDYHQQFRHVQNNLMRNSTTEKIEPRELDPIL**TEVTLMNARSELYL**RFLKKRISDFEVGDSMASEEVKQEH
 QKCLDKLLNCLLSCTMQELIGLYVTMEEYFMRETVNKAVALDITYEKGQLTSSMVDDVIFYIVKKCIGRALSSSIDCLCA
 MINLATTELESDFRDVLCNKLRMGFPATTFQDIQRGVTSAVNIMHSSSQGKFDTKGIESTDEAKMSFLVTLNNVEVCSE
 NISTLK**KTLESDC TKLFSQIGGEQAQAKFDSCLSDLA**AVSNKFRDLLQEGLTELNSTAIPQVQVPWINSFFSVSHNIEE

EEFNDYEANDPWVQQFI LNLEQQMAEFKASLSPVIYDSLTLGLMTSLVAVELEKVVVKSTFNRLGGLQFDKELRSLIAYLT
TVTTWTIRDKFARLSQMATILNLERVTEILDYWGPNSGPLTWRLTPAEVRQVLALRIDFRSEDIKRLRL



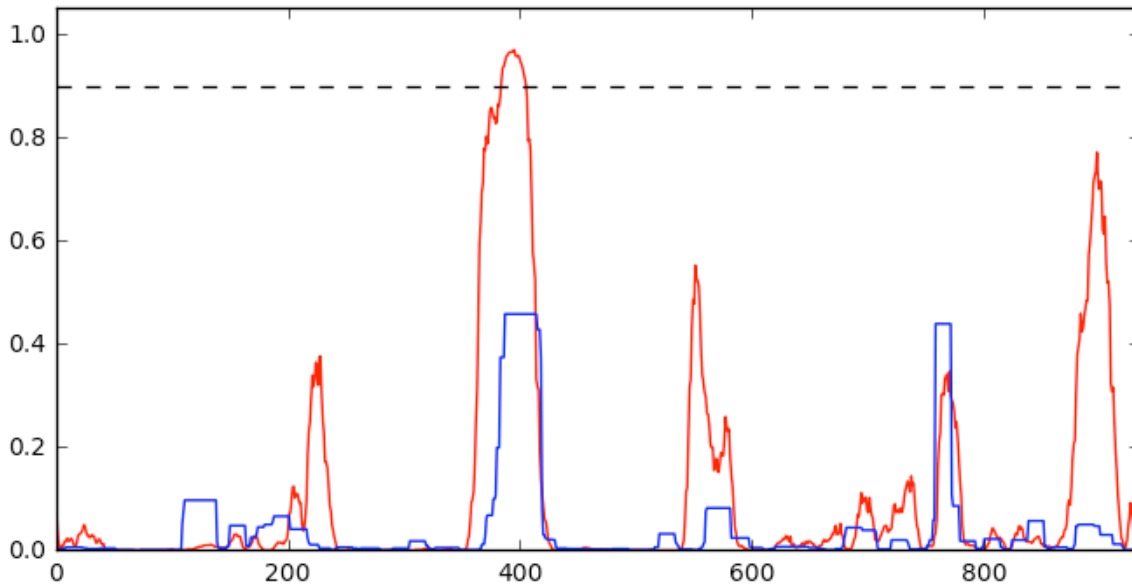
>NP_056525.2 H-

MAAGGSGVGGKRSSKSDADSGFLGLRPTSVDPALRRRRRGPRNKKRGWRRLAQEPLGLEVDQFLEDVRLQERTSGLLSE
APNEKLFVDTGSKEKGLTKKRTKVQKSSLLKPLRVDLILENTSKVPAPKDVLAHQVPNAKLRKEQLWEKLAKQGE
LPREVRRAQARLLNPSATRAKPGPQDIVERPFYDLWASDNPLDRPLVGQDEFFLEQTKKKGVKRPARLHTKPSQAPAVEV
APAGASYNPSFEDHQT LLSAAHEVELQRQKEAEKLERQLALPATEQAATQESTFQELCEGLLEESDGEPEPGQEGPEAG
DAEVCPTPARLAT TEKKTEQQRREKAVHRLRVQQAALRAARLRHQELFRLRGIKAQVALRLAELARRQRRRQARREAEA
DKPRLGRLKYQAPDIDVQLSSELTDSLRTLKPEGNILRDRFKSFQRRNMIEPREKAKFKRKYKVKLVKRAFREIQL



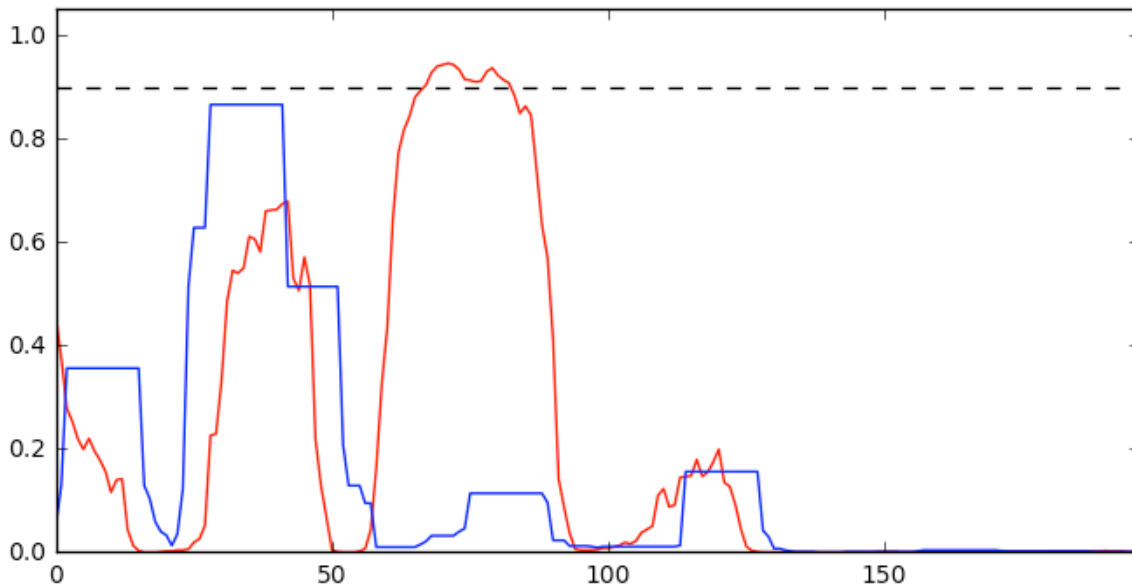
>NP_057060.2 H+

MEGERRASQAPSSGLPAGGANGESPGGGAPFPSSGSSALLQAEVLDLDEDEDDLEVFVSKDASLMDMNSFSPMMPTSPLS
MINQIKFEDEPDLKDLFITVDEPESHVTTIETFITYRIITKTSRGEFDSSEFEVRRRYQDFLWLKKGKLEEAHPTLIIPPL
PEKFIVKGMVERFNDDFIETRRKALHKFLNRIADHPTLTFNEDFKIFLTAQAWELSSHKKQGPGLLSRMGQTVRAVASSM
RGVKNRPEEFMEMNFI EILFSQKINLIDKISQRIYKEEREYFDEMKEYGPIHILWSASEEDLVDTLKDVA SCIDRCCKAT
EKRM SGLSEALLPVVHEYVLYSEM LGMV MKRRDQIQAE L DSKVEVLT YKKADTDLLPEEIGKLEDKVECANNALKA DWER
WKQNMQNDIKLAFTDMAENIHYEQCLATWESFLTSQTNLHLEEASEDKP



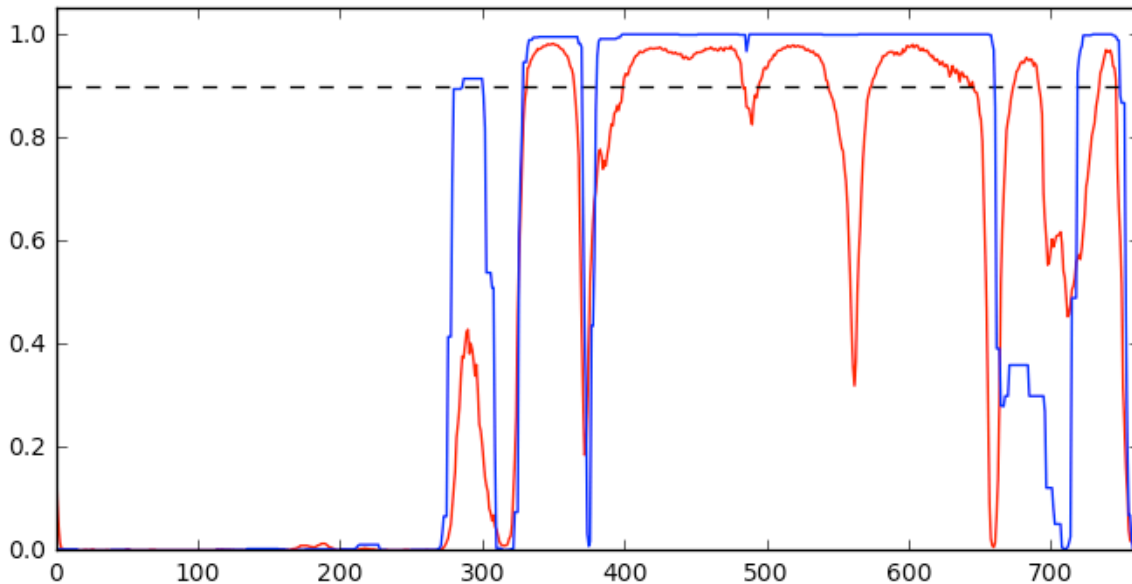
>NP_058649.3 H+

MAASRLDFGEVETFLDRHPELFEYLMRKKGQEMVEKWLQRHSQGQALGPRPSLAGTSSLAHSTCRGSSVGGGTGPNG
 SAHSQPLPGGDCGGVPLSPSWAGGSRGDGNLQRRASQKELRKSFARSKAIHVNRTYDEQVTSRAQEPLSSVRRRALLRK
 ASSLPPTTAHILSALLESRVNLPYPPTAIDYKCHLKKHNERQFFLELVKDISNDLDTLSYKILIFVCLMVDADRCSL
 FLVEGAAAGKKTLSKFFDVHAGTPLLPCSSSTENSNEVQVPWGKGIIGYVGEHGETVNI PDAYQDRRFNDEIDKLTGYKT
 KSLLCMPIRSSDGEIIGVAQAINKIPEGAPFTEDDEKVMQMYLPFCGIAISNAQLFAASRKEYE**RSRALLEVNDLFEEQ**
TDLEKIVKKIMHRAQTLLKCERC SVLLEDIESPVVKFTKSFELMSPKCSADAEN SFKESMEKSSYSDWLINNSIAELVA
 STGLPVNISDAYQDPRFDAEADQISGFHIRSVLCVPIWNSNHQIIGVAQVLNRLDGKPFDDADQRLFEAFVIFCGLGINN
 TIMYDQVKKSWAKQSVALDVLSYHATCSKAEVDKFKAAANIPLVSELAIDDIHFDDFSLDVDAMITAAALRMFEMELGMVQKF
 KIDYETLCRWLLTVRKNYRMVLYHNWRHAFNVCQLMFAMLT TAGFQDILTEVEILAVIVGCLCHDLDRGTNNAFQAKSG
 SALAQLYGT SATLEHHHFNHAVMILQSEGHNIFANLSSKEYSDLMQLLKQSILATDLTLYFERRTEFFELVSKGEYDWN
 KNHRDIFRSMMLTACDLGAVTKPWEISRQVAELVTSEFFEQGDRELERLEKLTPSAIFDRNRKDELPRLQLEWIDSICMPL
 YQALVKVNVKLPMLDSVATNRSKWEELHQKRLLASTASSSPASVMVAKEDRN



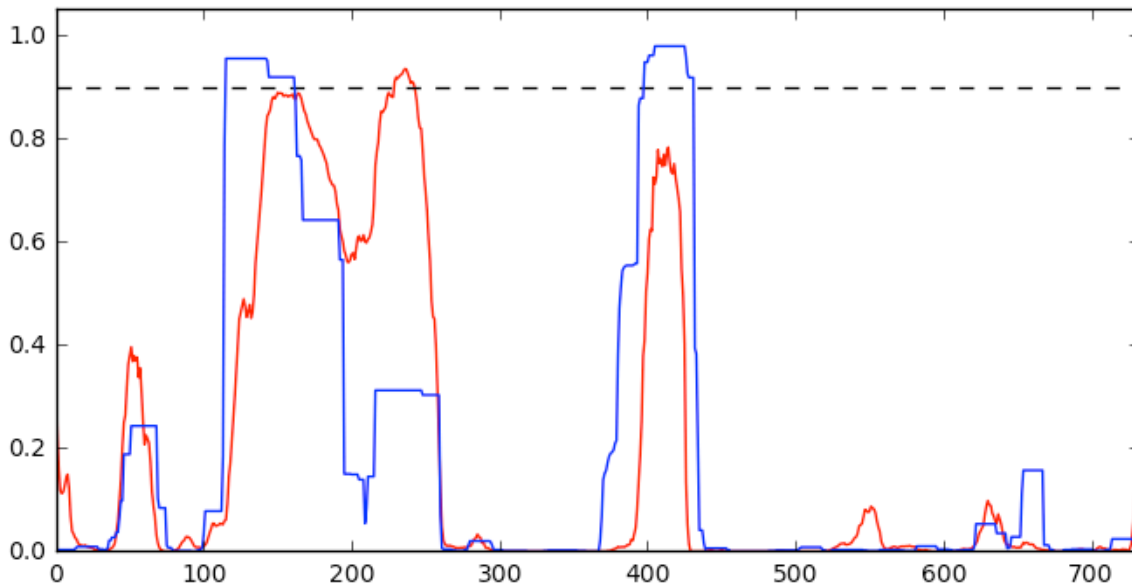
>NP_066545.3 H+

MFCEKAMELIRELHRAPEGQLPAF**NEDGLRQVLEEMKALYEQN**QSDVNEAKSGGRSDLIPTIKFRHC**SLLRNRRCTVAYL**
YDRLLRIRALRWEYGSVLPNALRFHMAAEEWENNYKRSLATYMRSLGGDEGLDITQDMKPPKSLYIEVRCLKDYGEFE
 VDDGTSVLLKKN SQHFLPRWKCEQLIRQGVLEHILS



>NP_073615.2 NH

MAMQEKYPTEGISHVTSPSSDVIQKSSLGTEWQTPVISEPFRSRFRSRCSSVADSGDTAIGTSCSDIAEDFCSSSGSPPF
 QPIKSHVTIPTAHVMPSTLGTSPAKPNSTPVGPPSSKPLPSGLAESVGMTRNGDLGAMKHSPGLSRDLMYFSGATGENGI
 EQSWFPAVGHERQEEARKFDIPSMESTLNQSAMMETLYSDPHHRVRFHNPRTSTSKELYRVLPEAKKAPGSGAVFERNGP
 HSNSSGVLPLGLQPAPGLSKPLPSQVWQPSPTWHPREQ**SCELSTCRQOLELIRLQMEQMLQNGAICHHPAAFGPSLPI**
LEPAQWISILNSNEHLLKEKELLIDKQRKHISOLEQKVRESELQVHSALLGRPAPFGDVCLLRLQELQRENTFLRAQFAQ
KTEALSREKIDLEKKSASEVEVQLIRESLKVALQKHSEEVKQEEVVKGRDKHINNLKCKQKQKQREKQORIE
RYLADLPTLEDHQKQSQQLKDESELKSTELQEKVTELESLEETQAIQREKEIQLESRLQREAEFSSAGHSLQDKQSV
EGEGPEVEMESWQKRYDSLQKIVEKQKQKMDQLRSQVQSLEQVAQEEGTSQALREEAQRDSALQQLRTAVKELSVQ
DLIEKNLTLQEHLRQAQPGSPSPDTAQLALELHQELASCLQDLQAVCSIVTQRAQGHDPNLSLLLGIHSAQHPETQLDL
QKPDVIKRLKEEVQQLRRDIEDLRTTMSDRYAQDMGENCVTQ



>NP_683698.2 NH

MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELLHLQQQQALQAAQQLLQQQTSGLKSPKSSD
 KQRPLQVPVSVAMMTPQVITPQQMQQILQQQVLS**PQQLQALLQQQAVMLQDFLDSGLENFRAALEKNQQLQEFYKQ**
EQHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHPGKQAKEQQQQQQQQQLAAQQLVFQQQLLQ
MQQLQQQHLLSLQRQLISIPPGQAALPVQSLPQAGLSPAIEIQLWKEVTGVHSMEDNGIKHGGLDLTTNNSSTSSN
 TSKASPPITHHSIVNGQSSVLSARRDSSSHEETGASHTLYGHGVCKWPGCESICEDFGQFLK**HLNNEHALDDRSTAQCRV**
QMQVVQOLEIQLSKERERLQAMMTHLHMRPSEPKPSPKPLNLVSSVTMSKNMLETSPQSLPQTPTTPTAPVTPITQGPSV
 ITPASVPNVGAIARRRHSKYNI PMSSEIAPNYEFYKNADVRPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFR
 NAATWKNVAVRHNLHLKCFVRVENVKGAVWTVDEVEYQKRRSQKITGSPTLVKNIPTSLGYGAALNASLQAAALAESSLPL

LSNPGLINNASSGLLQAVHEDLNGSLDHIDSNNGNSSPGCSPQPHIHSIHVKEEPVIAEDEDCPMSLVTTANHSPELEDDR
EIEEEPLSEDL

Bibliography

Li,C. *et al.* (2016) Critical evaluation of in silico methods for prediction of coiled-coil domains in proteins. *Brief. Bioinform.*, **17**, 270–82.