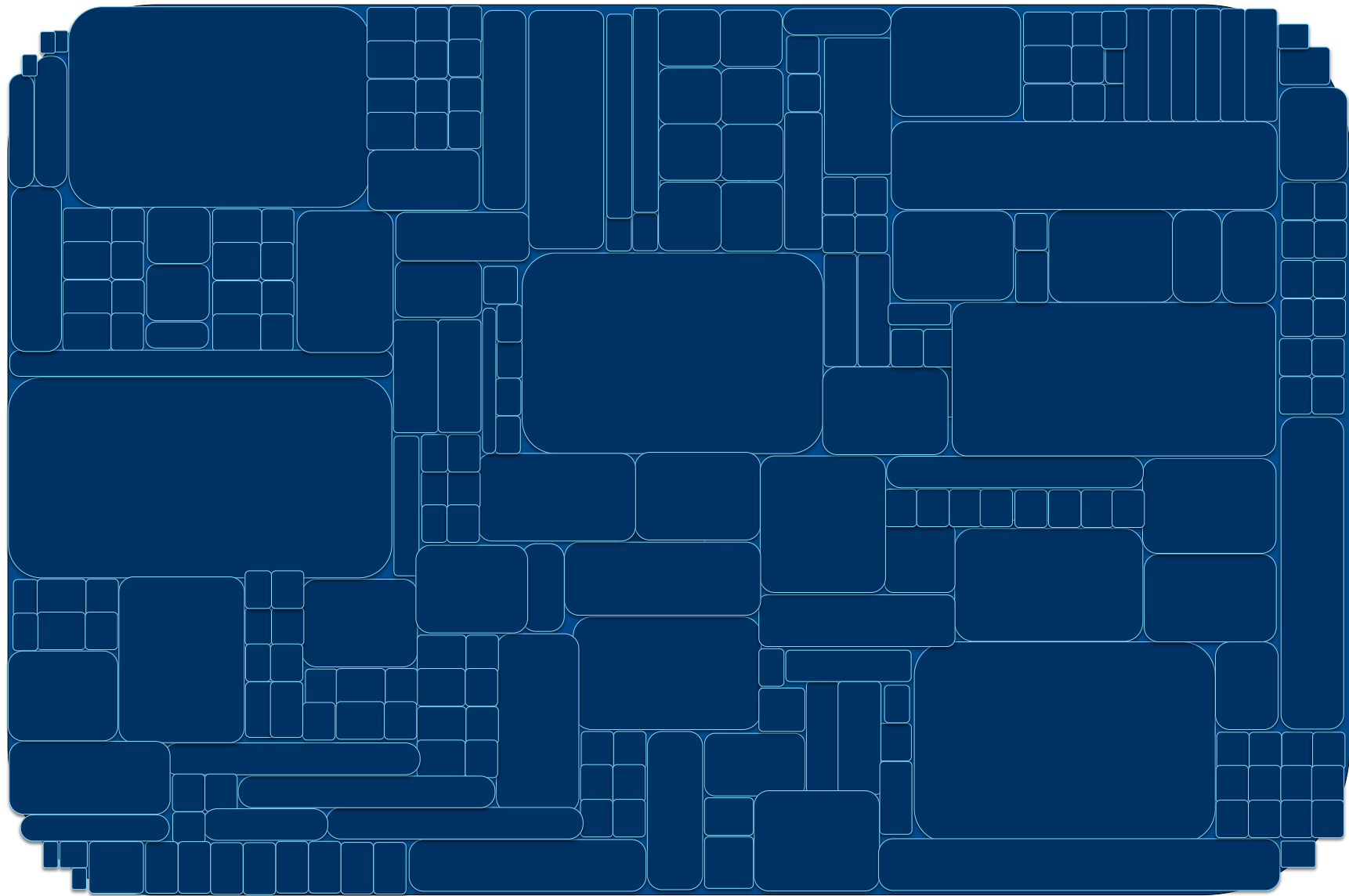
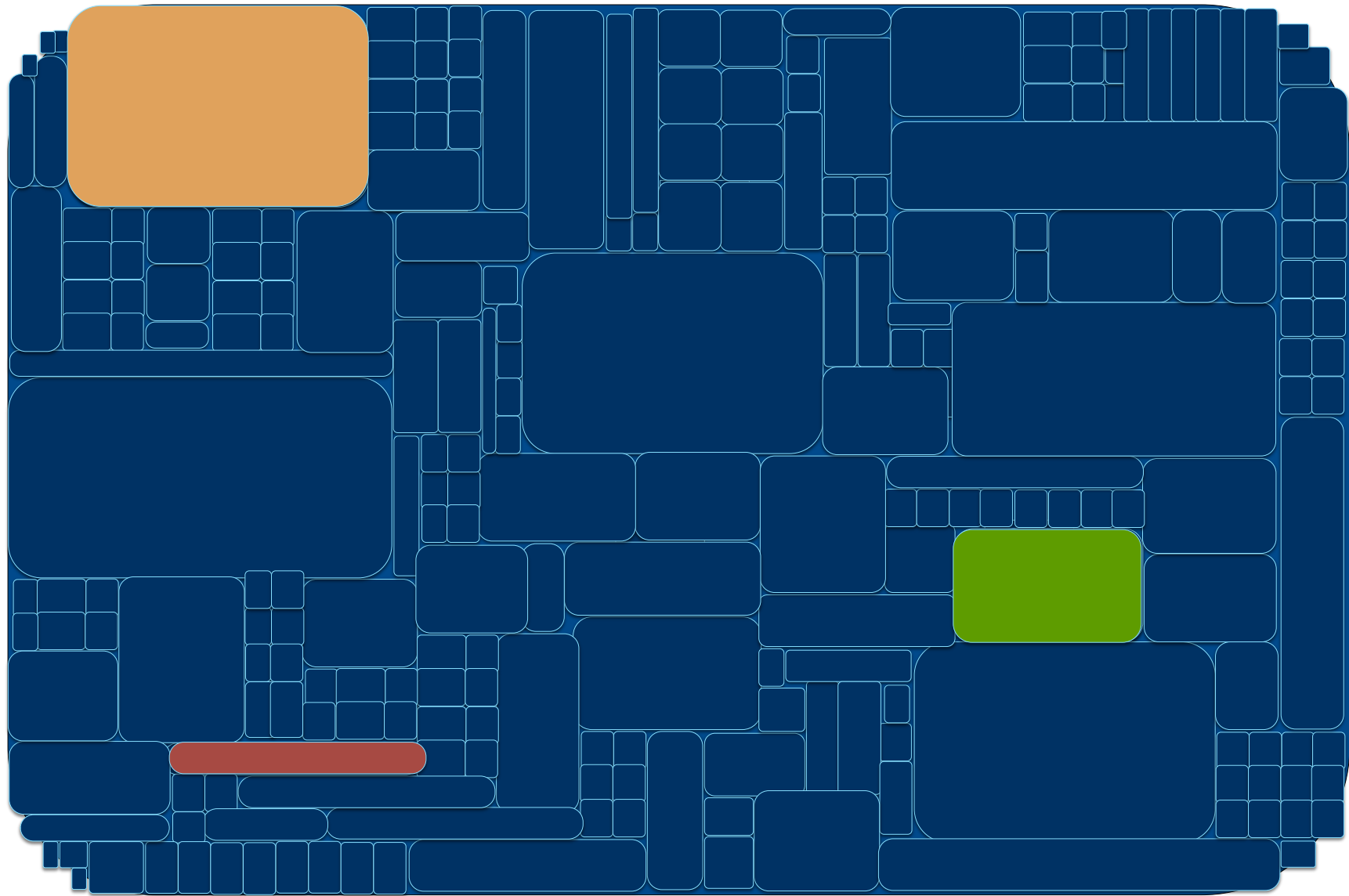


Protein family databases

Protein families

- Members will be structurally similar
- Members may share aspects of function
- The whole set of members may reveal elements of protein and organism evolution (phylogenies)





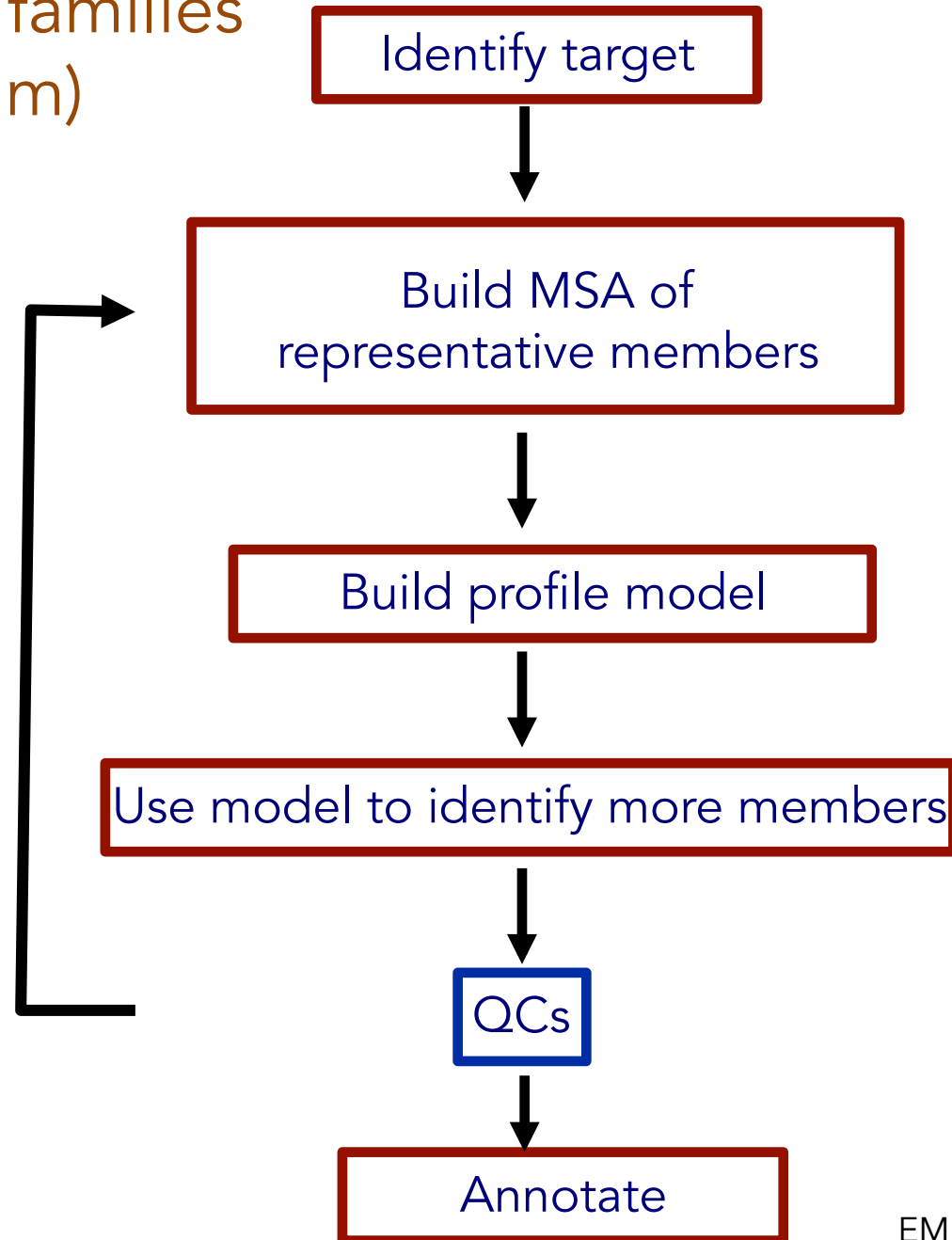
Ion transport

cNMP

L-zip

Building families (Pfam)

Marco Punta



Human: 1 MGLSDGEWQLVLNVWGKVEADIPGHGQEV LIRLFKGH PETLEKFDKFKHLKSEDEMKASE 60

Human: 61 DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH 120

Human: 121 PGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154

Family power

```

Human: 1  MGLSDGEWQLVLNVWGKVEADIPGHGQEV LIRLFKGHPETLEKFDKFKHLKSEDEMKASE 60
          MGLSDGEWQLVLNVWGKVEAD  GHGQEVLI LFK HPETL KFDKFK LKSE  MK SE
Mouse: 1  MGLSDGEWQLVLNVWGKVEADLAGHGQEV LIGLFKTHPETLDKFDKFKNLKSEEDMKGSE 60

Human: 61  DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH 120
          DLKKHG TVLTALG ILKKKG H AEI PLAQSHATKHKIPVKYLEFISE II VL  H
Mouse: 61  DLKKHGCTVLTALGTILKKKGQHAAEIQPLAQSHATKHKIPVKYLEFISEIIIEVLKKRH 120

Human: 121 PGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154
          GDFGADAQGAM KALELFR D A  YKELGFQG
Mouse: 121 SGDFGADAQGAMSKALELFRNDIAAKYKELGFQG 154

```

Family power

Human: 1 MGLSDGEWQLVLNVWGKVEADIPGHGQEVLI R L F K G H P E T L E K F D K F H L K S E D E M K A S E 60
MGLSDGEWQLVLNVWGKVEAD GHGQEVLI L F K H P E T L K F D K F L K S E M K S E
Mouse: 1 MGLSDGEWQLVLNVWGKVEADLAGHGQEVLI G L F K T H P E T L D K F D K F N L K S E E D M K G S E 60

Human: 61 DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH 120
DLKKHG TVLTALG ILKKKG H AEI PLAQSHATKHKIPVKYLEFISE II VL H
Mouse: 61 DLKKHGCTVLTALGTILKKKGQHAAEIQPLAQSHATKHKIPVKYLEFISEIIIEVLKKRH 120

Human: 121 PGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154
GDFGADAQGAM KALELFR D A YKELGFQG
Mouse: 121 SGDFGADAQGAMSKALELFRNDIAAKYKELGFQG 154

BLOSUM62 matrix

| | | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ala | 4 | | | | | | | | | | | | | | | | | | | |
| Arg | -1 | 5 | | | | | | | | | | | | | | | | | | |
| Asn | -2 | 0 | 6 | | | | | | | | | | | | | | | | | |
| Asp | -2 | -2 | 1 | 6 | | | | | | | | | | | | | | | | |
| Cys | 0 | -3 | -3 | -3 | 9 | | | | | | | | | | | | | | | |
| Gln | -1 | 1 | 0 | 0 | -3 | 5 | | | | | | | | | | | | | | |
| Glu | -1 | 0 | 0 | 2 | -4 | 2 | 5 | | | | | | | | | | | | | |
| Gly | 0 | -2 | 0 | -1 | -3 | -2 | -2 | 6 | | | | | | | | | | | | |
| His | -2 | 0 | 1 | -1 | -3 | 0 | 0 | -2 | 8 | | | | | | | | | | | |
| Ile | -1 | -3 | -3 | -3 | -1 | -3 | -3 | -4 | -3 | 4 | | | | | | | | | | |
| Leu | -1 | -2 | -3 | -4 | -1 | -2 | -3 | -4 | -3 | 2 | 4 | | | | | | | | | |
| Lys | -1 | 2 | 0 | -1 | -3 | 1 | 1 | -2 | -1 | -3 | -2 | 5 | | | | | | | | |
| Met | -1 | -1 | -2 | -3 | -1 | 0 | -2 | -3 | -2 | 1 | 2 | -1 | 5 | | | | | | | |
| Phe | -2 | -3 | -3 | -3 | -2 | -3 | -3 | -3 | -1 | 0 | 0 | -3 | 0 | 6 | | | | | | |
| Pro | -1 | -2 | -2 | -1 | -3 | -1 | -1 | -2 | -2 | -3 | -3 | -1 | -2 | -4 | 7 | | | | | |
| Ser | 1 | -1 | 1 | 0 | -1 | 0 | 0 | 0 | -1 | -2 | -2 | 0 | -1 | -2 | -1 | 4 | | | | |
| Thr | 0 | -1 | 0 | -1 | -1 | -1 | -1 | -2 | -2 | -1 | -1 | -1 | -1 | -2 | -1 | 1 | 5 | | | |
| Trp | -3 | -3 | -4 | -4 | -2 | -2 | -3 | -2 | -2 | -3 | -2 | -3 | -1 | 1 | -4 | -3 | -2 | 11 | | |
| Tyr | -2 | -2 | -2 | -3 | -2 | -1 | -2 | -3 | 2 | -1 | -1 | -2 | -1 | 3 | -3 | -2 | -2 | 2 | 7 | |
| Val | 0 | -3 | -3 | -3 | -1 | -2 | -2 | -3 | -3 | 3 | 1 | -2 | 1 | -1 | -2 | -2 | 0 | -3 | -1 | 4 |
| | Ala | Arg | Asn | Asp | Cys | Gln | Glu | Gly | His | Ile | Leu | Lys | Met | Phe | Pro | Ser | Thr | Trp | Tyr | Val |

Family power

```

10      20      30      40      50      60      70      80      90      100     110     120
HBAZ_CAPHI/7-107  ERTIILSLWSK-IST-----QADVIGTETLERLFSCYPQAKTYFPHF--DLH-----SGSAQLRAHGSKVVAAGDAVKS-----D-NVTSALSKLSELHAYVL---RVDPVNFKFLSHCL
HBA3_PLEWA/7-107  EKALVVGLCGK-ISG-----HCDALGGEALDRLFASFGQTRTYFPHF--DLS-----PGSADVKRHGGKVLSAIGEAAKHI-----D-SMDQALSKLSDLHAYNL---RVDPGNFQLLSHCI
HBA_CATCL/6-107   DKADVKAIAWAK-ISP-----RADEIGAEALGRMLTVYPQTKTYFAHW--ADLS-----PGSGPVKHGKKVIMGAIGDAVTKF-----D-DLLGGLASLSELHASKL---RVDPNFKILANCI
HBB_HETPO/7-106   ELHEITTTWKS-I-----DKHSLGAKALARMFIVYPWTTRYFPHL-KEFT-----ACSYGVKEHAKKVTGALGVAVTHL-----G-DVKSQFTDLSKKHAEEL---HVDVESFKLLAKCF
HBB_SQUAC/7-107   EKALVNAVWTK-T-----DHQAVVAKALERLFVVPWTKTYFVKFNGKFH-----ASDSTVQTHAGKVVSALTVAYNHI-----D-DVKPHFVELSKKHYEEL---HVDPENFKLLANCL
HBB1_CYGMA/8-112  ELTIINDIFSH-L-----DYDDIGPKALSRLIIVYPWQRFHFGF--GNLYNAEAIIGNANVAAHGIKVLHGLDRGLKNM-----D-NIVDAYAELSTLHSEKL---HVDPDNFKLLSDCI
HBB1_XENBO/7-111  DRQLINSTWGK-V-----CAKTIGKEALGRLLWPTYWQRFHFGF--GNLSADAVFHNEAAHGEKVVTSIGEAIKHM-----D-DIKGYAQLSKYHSETL---HVDPCNFKRFGGCL
HBB_LITCT/1-105   GGSDVSAFLAK-V-----DKRAVGGEALARLLIIVYPWQRFHFGF--GNLGSADAISHNSKVLAHGQRVLDSIEEGLKHP-----Z-BLKAYYAKLSERHSGEL---HVDPANFYRLGNVL
HBB_LEPPA/7-111   EKQYIVSVFSK-I-----DVDHVGANTLERVLIIVFPWTKRYFPHF--GDLSSPGAIKHNNKVSAGHRKVLAAIIECTRHF-----G-NIKGHLANLSHLHSEKL---HVDPHNFRVLGQCL
HBB2_XENLA/8-112  EKAAITSVWQK-V-----NVEHDGHDALGRLLIIVYPWQRFHFGF--GNLSNAAVAGNAKVQAAGKKVLSAVGNAISHI-----D-SVKSSLQQLSKIATHEL---FVDPENFKRFGGVL
HBB_ALLMI/7-111   ERKFIVDLWAK-V-----DVAQCGADALSRMLIIVYPWKRRYFPHF--GKMCNAHDIHNSKVQEHGKKVLSFGEAVKHL-----D-NIKGHFANLSKLHCEKF---HVDPENFKLLGDI
HBB0_MOUSE/8-112 EKAAITSIWDK-V-----DLEKVGGETLGRLLIIVYPWQRFHFGF--GNLSSAQAIMGNPRIKAHGKKVLTSLGLAVKNM-----D-NLKETFAHLSLHCDKL---HADPENFKLLGNML
HBBN_AMMLE/2-106  BKALITGFWSK-V-----KVBZVGAZALGRLLVVPWTZRFFHFGF--GBLSSABAVMBBAKVKAHGKKVLSFSBGLKHL-----B-BLKGAFASLSZLHCBKL---HVBPBZBFRLLGBVL
HBA_LEPPA/7-108   DEVLKEAWGL-L-H-----QIPNAGGEALARMFSCYPGTKSYFPHFGHDFS-----ANNEKVKHHGKKVVDAIGQGQVHL-----H-DLSSCLHTLSEKHAREL---MVDPCNFQYLIEAI
HBA1_TORMA/6-107  NKKAIKNLLQK-IHS-----QTEVLGAELARLFECHPQTKSYFPHF--SGFS-----ANDKRKVKHGHVGLVLAALVDTNKHL-----D-DLPHHLNKLAEKHGKGL---LVDPHNFKLLSDCI
HBA_SQUAC/6-107   DKTAIKHLTGS-LRT-----NAEAWGAESLARMFATTPSTKTYFPHF--TDFS-----ANGKRKVAHGKKVLSAVADATDHL-----D-NVAGHLDPALVHGTTL---CVDPHNFKLLTQCI
HBA_HETPO/13-114  DRALAALSKEV-LAQ-----NAEAFGAELARMFATTPSTKTYFPHF--KDFT-----AAAPSIKAHGAKVVTALAKACDHL-----D-DLKTHLHKLATFHGSEL---KVDPANFYQLSYCL
GLB1_TYLHE/7-110  QRIVKVKQWAAQ-VYSV---GESRTDFAIDVFNNFRTNPDRS--LFRVVMGDNV-----YSEPKAHMVRVVFAGFDILISVL---DDKPVLDQALAHYAFAHKQFG---TIP--FKAFGQTM
GLB4_LUMTE/11-120  DRREIRHIWDD-VWSSS--FTDRRVAIVRAVFDLFFKHYPTSKALFERVKIDEP-----ESGEFKSHLVRVANGDLLINLL---DDTLVLQSHLGHADQHIQRK---GVTKEYFRGIGEAF
GLB3_TYLHE/8-117  DRHEVLDNWKG-IWSAE--FTGRRVAIGQAIQELFALDPNAKGVFGRVND-K-----PSEADWKAHVIRVINGLDLAVNLL---EDPKALQEELKHLARQHRERS---GVKAVYFDEMEKAL
GLB4_TYLHE/8-117  DRREVQALWRS-IWSAE--DTGRRTLIGRLLFEELFEIDGATKGLFKKLVNDDT-----HSPPEFAHVLRRVNGLDLTIGVL---GDSD-TLNSLIDHLAEQHKARA---GFKTVYFKEFGKAL
GLB2_TYLHE/9-115  QRLKVKQWAAQ-VYSV---GHERVELGIALWKSMAFDNDARDLFFKLVHGDDT-----HSPAFEAHMAVRVNGLDLDRVSSLL---TDEPVLNAQLEHLRQHIKLG---ITGHMFNLMRTGL
GLB2_LUMTE/8-114  EGLKVKSEWGR-AYGS---GHDRFAFSQAIWRATFAQVPSERSLFFKLVHGDDT-----SHPAFIAHAERVVGLGLDIAISTL---DQPATLKEELDHLQVQHEGRK---IPDNYFDAFKTAI
GLB_TUBTU/6-112  QRFKVKHQWAE-AFGT---SHHRLDFGLKLWNSIFRDAPEIRGLFKRVGDD-N-----AYSAEFEAHAERVVGLGLDMTISLL---DDQAAFDAQLAHLKSQHAERN---IKADYYGVFVNEL
GLB3_LAMSP/7-113  QRLKVKRQWAE-AYGS---GNDREEFGHFVWTHVFKDAPSARDLFFKRVGDDNI-----HTPAFRAHATRVLGLDMCJALL---DDEGVNLTQLAHLASQHSRG---VSAAQYDVVEHSV
GLB_PAREP/8-117   QDILLKELGPH-V-DT---PAHIVETGLGAYHALFTAHQYIIHFRL--EG-HTIENVMQSEGIKKHYARTLTEAIVHMLKEI---SNDAEVKKIAAQYGKDHTSRK---VTKDEFMSGEPIF
Q21978_CAEL/165-283  SCEVVASWRL-VESRSSAAETSACFGLFVQRFVFSKIPMLRPLFLGL--SESDVDLDPDNHPPVRRHARLFTSLHISVKNV---DELEAQVAPTVEKYGERHYRPDITPHMTEENVRVFCQAI
GLB_PSED/21-134   TRELCKMSLEH-AKVG--SKEAKQDGDIDLYKHMFEHYPMKKYFKHR--ENYTPADVQKDPFFIKQGGNILLACHVLCATY---DDRETFDAYVGEELMARHERDHV---KIPNDVWNHFWEHF
GLB_ASCSU/21-134  TRELCKMSLEH-AKVD--SNEARQDGDIDLYKHMFEHYPPLRKYFKHR--EEYTAEDVQNDPFFAKQGGKILLACHVLCATY---DDRETFNAYTRELLDRHARDHV---HMPPEVWTFDFWKL
GLB_C_NIPBR/21-135  DVK--KHTVES-MKAVP-VGRDKAQNGIDFYKFFTHHKDLRKFFKGA--ENFGADDVQKSRFEKQGTALLAVHVLAVNY---DNQAVFHGFVRELMNRHEKRGVDPKLWKI--FFDDVWVVP
GLB_C_CAEL/10-119  DLC-VKLEGR-MVGT--AQNI--ENGNAFYRYFFTNFPDLRVYFKGA--EKYTADDVKKSERFDKQGRILLACHLLANVY---TNEEVFKGYVRETINRRIYK---MDPALWMAFFTFV
GLB2_NIPBR/16-114  PISKAQ-----AQ-----VGKDFYKFFTNHPDLRKYFKGA--ENFTADDVQKSRFEKLGSGLLSVHILANTF---DNEDVFRAFRETIDRHVGRG---LDPALWKAFFSVW
GLB_TRICQ/30-132  DVPVPLGSTPEK-L-----ENGREFYKYFFTNHQDLRKYFKGA--ETFTADDLAKSDRFKKLGNQLLLSVHLAADTY---DNEMIFRAFVRDIDRHVDRG---LDPKLWKEFWSIY
Q20638_CAEL/74-184  EKELLRRTWSD-EFD-----NLYELGSAIYCYIFDHNPNCKQLFPF--ISKYQGDWEKSEKFRSQALKFVQTLAQVVKNIYHMERTESFLYMGVQKHVKFADRG---FKHEYWDIFQDAM
Q19601_CAEL/105-215  ERILLLEQSWRK-TRK-----TGADHIGSKIFFMVLTAPQDIAKIFGL--EKIPTGRLLKYPDRFRQHALVYTKLDFVIRNL---DYPGKLEVYFENLGRHVAMQG--RGFEPGYWETFAECM
Q18311_CAEL/32-140  TKKLVIQEWPR-VLA-----QCPELFTIWHKSASTRSTIKLAFGL--IAE-N--ESPMQNAALGLSSTIQAFFYKLIITYE-L-NDDQVREACEQLGARHVDFIS--RGFNSHFWDFILVCM

```

Family power

| | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 | 110 | 120 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|------------------|------|---------|------|-------|----------|-------|--------|-------|--------|------|-------|-------|-------------|---------|--------|--------|-------|--------|-------|-------|--------|-------|-------|-------|-----|-------|-------|-------|-------|-----|-------|-----|-------|-------|-----|-------|-----|---|-------|-----|-------|-------|-----|----|---|---|-------|-----|-----|---|-------|-----|-----|-------|-------|-----|---|---|---|---|---|---|---|---|---|---|---|
| HBAZ_CAPH1/7-107 | ERTI | LSLWSK | -IST | ----- | QADVIGET | LERL | FS | CA | QAKT | EPHF | -DLH | ----- | SGSAQLRAHG | KVVAAY | GDAVKS | I | ----- | D-NVTS | ALS | SKLS | ELHAY | VL | ----- | RVD | PNV | EKFL | LSHCL | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| HBA3_PLEWA/7-107 | EKAL | VVGLGK | -ISG | ----- | HCDA | LGGA | LDRL | FA | SGTRT | FSHF | -DLS | ----- | PGSADWRKHG | GKVL | SA | GEAAKH | I | ----- | D-SMD | QALS | SKLS | LDHAY | NL | ----- | RVD | PGN | QFLSH | CHCI | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| HBA_CATCL/6-107 | DKAD | VKIAWA | -ISP | ----- | RADE | IGAE | ALGR | MLTV | PQTKT | FANW | -ADLS | ----- | PGSGPVKHG | KKVIMGA | IGDAV | TKF | ----- | D-DLL | GGLAS | LS | ELHASK | L | ----- | RVD | PSN | FKI | LANCI | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| HBB_HETPO/7-106 | ELHE | ITTTWKS | -I | ----- | DKHS | LGA | KALARM | FIV | PWTTTR | FGNL | -KEFT | ----- | ACSYGVEKHAK | VT | GALGVA | AVTHL | ----- | G-DVK | SQFT | DL | SKKHA | EEL | ----- | HVD | VS | EKLL | LAKCF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| HBB_SQUAC/7-107 | EKAL | VNAVMT | -K | ----- | DHQAVV | KALER | LFFV | VV | PWT | TKT | FKYK | -KGFH | ----- | ASDSTVQ | TAGK | VVSAL | VAYNH | ----- | D-DVK | PHF | LVLS | SKKH | EEL | ----- | HYD | PEN | FKL | LANCL | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| HBB1_CYGMA/8-112 | ELTI | IINDIF | -SH | ----- | DYDD | IGPK | ALS | RCLIV | PWT | QR | FSGF | -GNLY | NAEAI | IGNAN | VAAHG | IKVLH | GLDR | GL | KL | KN | ----- | D-NIV | D | AYAE | LS | TH | SEK | L | ----- | HYD | PDN | EK | LLSD | GC | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| HBB1_XENB/07-111 | DRQL | INSTWG | -V | ----- | CAKT | IGKE | ALGR | LLWTV | PWT | QRY | FSF | -GNLS | ADAV | FHN | EA | VAHGE | KVUTS | IG | EAKH | ----- | D-DIK | G | Y | Y | A | LS | KH | SE | T | L | ----- | HYD | PN | FKR | FG | CC | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| HBB_LITCT/1-105 | GGSD | VSAFLA | -V | ----- | DKRA | VGE | ALGR | LLWTV | PWT | QR | FSF | -GNLS | ADAI | S | NSK | VLA | HG | QV | LD | S | IG | E | L | K | H | P | ----- | Z-B | L | K | A | F | A | S | Y | A | K | L | S | E | R | H | S | G | E | L | ----- | HYD | PAN | F | Y | R | L | G | N | V | | | | | | | | | | | |
| HBB_LEPPA/7-111 | EKQY | IVSVFS | -K | ----- | DVDH | VGAT | L | ERVL | FIV | PWT | QR | FSF | -GNLS | SPGA | IKHN | NKVS | AHGR | KV | LA | I | E | C | T | R | H | F | ----- | G-N | I | K | G | H | L | AN | S | L | H | S | E | K | L | ----- | HYD | PN | F | N | R | V | L | G | Q | L | | | | | | | | | | | | | | | |
| HBB2_XENLA/8-112 | EKAA | ISVWQK | -V | ----- | NVEH | DG | HAL | GR | LLIV | PWT | QR | FSF | -GNLS | SA | AVAG | NKQA | HG | KKV | LS | AV | NA | I | S | H | I | ----- | D-SV | K | S | L | Q | L | S | K | I | H | A | T | E | L | ----- | F | V | D | P | N | F | K | R | G | G | V | | | | | | | | | | | | | | | |
| HBB_ALLMI/7-111 | ERKF | IVDLWA | -K | ----- | DVAQ | CGAD | ALS | RMLIV | PWK | R | FEHF | -GK | M | NAH | D | LHNS | KVQ | E | H | G | K | V | L | S | F | G | E | A | V | K | H | L | ----- | D-N | I | K | G | H | F | A | N | L | S | K | L | H | C | E | K | L | ----- | HYD | PEN | F | K | L | G | D | I | | | | | | | | |
| HBB0_MOUSE/8-112 | EKAA | ISWDK | -V | ----- | DLEK | VGG | E | T | LGR | LLIV | PWT | QR | FSF | -GNLS | SA | QA | I | MGN | P | RI | KA | H | G | K | V | L | S | F | G | E | A | V | K | H | L | ----- | D-N | L | K | E | T | F | A | H | L | S | E | L | H | C | E | K | L | ----- | HAD | PEN | F | K | L | G | N | M | | | | | |
| HBBN_AMMLE/2-106 | BKAL | ITGFWS | -K | ----- | DVBZ | VGA | Z | ALGR | LLV | PWT | ZR | FSF | -GL | SS | ABA | VMB | BAK | V | KA | H | G | K | V | L | S | F | G | E | B | L | K | H | L | ----- | B-B | L | K | A | F | A | S | L | S | Z | L | H | C | B | K | L | ----- | H | B | P | Z | B | F | R | L | G | B | V | | | | | |
| HBA_LEPPA/7-108 | DEV | LKEAWGL | -L | ----- | QIPN | AGGE | ALARM | FS | C | P | G | T | K | S | FP | H | F | H | D | F | ----- | -ANNE | K | V | K | H | G | K | V | D | A | I | G | Q | V | Q | H | L | ----- | H-D | L | S | S | C | H | L | T | S | E | K | A | R | E | L | ----- | M | V | D | P | N | F | Q | F | L | E | A | I |

Family power

Human: 1 MGLSDGEWQLVLNVWGKVEADIPGHGQEV LIRLFKGHPETLEKFDKFKHLKSEDEMKASE 60
MGLSDGEWQLVLNVWGKVEAD GHGQEVLI LFK HPETL KFDKFK LKSE MK SE

Mouse: 1 MGLSDGEWQLVLNVWGKVEADLAGHGQEV LIGLFKTHPETLDKFDKFKNLKSEEDMKGSE 60

Human: 61 DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH 120
DLKKHG TVLTALG ILKKKG H AEI PLAQSHATKHKIPVKYLEFISE II VL H

Mouse: 61 DLKKHGCTVLTALGTILKKKGQHAAEIQPLAQSHATKHKIPVKYLEFISEIIIEVLKRRH 120

Human: 121 PGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154
GDFGADAQGAM KALELFR D A YKELGFQG

Mouse: 121 SGDFGADAQGAMSKALELFRNDIAAKYKELGFQG 154

Family power

```

10      20      30      40      50      60      70      80      90      100     110     120
HBAZ_CAPHI/7-107 ERTIILSLWSK-I-ST-----QADVIGTETLERLFCYPAQAKTYEPHF--DLH-----SGSAQLRAHGSKVVAAGDAVKS I-----D-NVTSALSKLSELHAYVL---RVDPVNFKFLSHCL
HBA3_PLEWA/7-107 EKALVVGLCGK-ISG-----HCDALGGEALDRLEAFSGQTRTYPSHF--DLS-----PGSADVRRHGKVLSAIGEAAKHI-----D-SMDQALSKLSDLHAYNL---RVDPGNFQLLSHCI
HBA_CATCL/6-107 DKADVKI AWAK-ISP-----RADEIGAEALGRMLTVYPQTQTYFANW-ADLS-----PGSGPVKHKKKVIMGAIGDAVTKF-----D-DLLGGLASLSELHASKL---RVDPSPNFKILANCI
HBB_HETPO/7-106 ELHEITTTWKS-I-----DKHSLGAKALARMFIVYPWTTRYFANL-KEFT-----ACSYGVKEHAKKVTGALGVAVTHL-----G-DVKSQFTDLSKKHAEEL---HVDVESFKLLAKCF
HBB_SQUAC/8-107 EKALVNVAWTK-T-----DHQAVVAKALRFLFVVPWTQTYFVKFNGKFH-----ASDSTVQTHAGKVVSAITVAYNHI-----D-DVKPHFVELSKKHYEEL---HVDPENFKLLANCL
HBB1_CYGMA/8-112 ELTIINDIFSH-L-----DYDDIGPKALSRLIIVYPWTQRYHSCGF--GNLYNAEAIIGNANVAAHGIKVLHGLDRGLKNM-----D-NIVDAYAELSTLHSEKL---HVDPDNFKLLSDCI
HBB1_XENBO/7-111 DRQLINSTWGK-V-----CAKTIGKEALGRLLWTPWTQRYPSFF--GNLSADAVFHNEAAHGEKVVTSIGEAIKHM-----D-DIKGYAQLSKYHSETL---HVDPCNFKRFGGCL
HBB_LITCT/1-105 GGSDDVSFAFLAK-V-----DKRAVGGAEALARLLIIVYPWTQRYPSFF--GNLGSADAISHNSKVLAHQQRVLDSIEEGLKHP-----Z-BLKAYYAKLSERHSGEL---HVDPANFYRLGNVL
HBB_LEPPA/7-111 EKQYIVSVFSK-I-----DVDHVGANTLERVLIVYPWTQRYPSFF--GDLSSPGAIKHNKVSAGHGRKVLAAIECTRH-----G-NIKGHLANLSHLHSEKL---HVDPHNFRVLGQCL
HBB2_XENLA/8-112 EKAAITSVWQK-V-----NVEHDGHDALGRLLIIVYPWTQRYPSFF--GNLSNSAAVAGNAKVQAHGKKVLSAVGNAISHI-----D-SVKSSLQQLSKIATL---HVDPENFKRFGGVL
HBB_ALLMI/7-111 ERKFIIVDLWAK-V-----DVAQCGADALSRLIIVYPWKRRYFENF--GKMCNAHDI LHNKSKVQEHGKKVLSFGEAVKHL-----D-NIKGHFANLSKLHCEKF---HVDPENFKLLGDI I
HBB0_MOUSE/8-112 EKAAITSIWDK-V-----DLEKVGGETLGRLLIIVYPWTQRYPSFF--GNLSSAQAIMGNPRIKAHGGKVLTSGLAVKNM-----D-NLKETFAHLSLHCDKL---HADPENFKLLGNML
HBBN_AMMLE/2-106 BKALITGFWSK-V-----KVBZVGAZALGRLLVVPWTZRFPFNF--GBLSSABAVMBBAKVKAHGGKVLBSFSBGLKHL-----B-BLKGAFASSLZLHCBKL---HVBFPZBFRLLGBVL
HBA_LEPPA/7-108 DEVLIKEAWGL-L-H-----QIPNAGGEALARMFSCYPGTQSYFPHFCHDFS-----ANNEKVKHHGKKVVDAIGQGVQHL-----H-DLSSCLHTLSEKHAREL---MVDPCNFQYLI EAI
HBA1_TORMA/6-107 NKKAIKNLLQK-IHS-----QTEVLGAELARLFECHPQTQSYFPRKF--SGFS-----ANDKRVKHHGALVLKALVDTNKHL-----D-DLPHHLNKLAEKHGKGL---LVDPHNFKLLSDCI
HBA_SQUAC/6-107 DKTAKHLLTGS-LRT-----NAAEAWGAESLARMFATTSTKTYSKRF--TDFS-----ANGKRVKAHGGKVLNVAADATDHL-----D-NVAGHLDPLAVLHGTTL---CVDPHNFPLLTQCI
HBA_HETPO/13-114 DRAELAALSKEV-LAQ-----NAAEAFGAELARMFVYATKSYFKDY--KDFT-----AAAPSIKAHGAHVVTALAKACDHL-----D-DLKTHLHLKATFHGSEL---KVDPANFYLSYCL
GLB1_TYLHE/7-110 QRIKVKQQAQ-VYSV---GESRTDFAIDVFNNFRTNDRS--FNRVNGDNV-----YSPFEKAHMRVRFAGFDILISVL---DDKPVLDQALAHYAFAFKQFG---TIP--FKAFGQTM
GLB4_LUMTE/11-120 DRREIRHIWDD-VWSSS--FTDRRVAIVRAVFDLFLKHYPTSKALFERVNI DEP-----ESGEFKSHLVRVANGLDLLINLL---DDTLVLQSHLGHADQHIQRK---GVTKEYFRGIGEAF
GLB3_TYLHE/8-117 DRHEVLNDNWKG-IWSAE--FTGRRVAIGQAI FQELFAENAKGVFGRVNVD-K-----PSEADWKAHVIRVINGLDLAVNLL---EDPKALQEEELKHLARQHRERS---GVKAVYFDEMEKAL
GLB4_TYLHE/8-117 DRREVQALWRS-IWSAE--DTGRRTLIGRLLFEELEFIDGATKGLFKRVNVDOT-----HSPPEFAHVLRVVNGLDITLIGVL---GDSO-TLNSLIDHLAEQHKARA---GFKTVYFKEFGKAL
GLB2_TYLHE/9-115 QRLKVKQQAQ-AYGV---GHERVELGIALWKSMEFAQDNDAARDLFRVHGEDV-----HSPAFAEHMARVFNGLDRVSSLL---TDEPVLAQLEHLRQHIKLG---ITGHMFNLMRTGL
GLB2_LUMTE/8-114 EGLKVKSEWGR-AYGS---GHDRFAFSQAIWRATFAQVPESRSLFRVHGDDT-----SHPAFIAHAERVVLGGLDIAISTL---DQPATLKEELDHLQVQHEGRK---IPDNYFDAFKTAI
GLB_TUBTU/6-112 QRFKVKHQWAE-AFGT---SHRLDFGLKLWNSIFRDAPEIRGLFRVVGDN-N-----AYSAEFAHAERVVLGGLDMTISLL---DDQAAFDQAHLKLSQHAERN---IKADYYGVFVNEL
GLB3_LAMSP/7-113 QRLKVKRQWAE-AYGS---GNDREEFGHFIMTHVFKDAPSARDLFRVVGDN-I-----HTPAFRAHATRVLGGGLDMCJALL---DDEGVLNTQLAHLASQSSRG---VSAAQYDVVEHSV
GLB_PAREP/8-117 QDILLKELGPH-V-DT---PAHIVETGLGAYHALFTAHQYIIHFSRL-EG-HTIENVMQSEGIKHYARTLTEAIVHMLKEI---SNDAEVKKIAAQYGKDHTRSK---VTKDEFMSSGEPIF
Q21978_CAEL/165-283 SCEVWADSWRL-VESRSSAAETSACFGLFVFQRFVFSKIPMLRPLFGL-SESDDVDFLDPNHPVRRHARLFTSILHISVKNV---DELEAQVAPTVEKYGERHYRPDITPHMT EENVRVFCQAI
GLB_PSED/21-134 TRELCMKSLH-AKVGTSKEAKQDGDIDLYKHMFEHY PAMKKYFKNR--ENYTPADVQKDPFFIKQGNILLACHVLCATY---DDRETFDAYVGE LMARHERDHV---KIPNDVWNHFWEHF
GLB_ASCU/21-134 TRELCMKSLH-AKVDT--SNEARQDGDIDLYKHMFEHY PPLRKYFKNR--EETAEADVQNDPFFAKQGGKILLACHVLCATY---DDRETFNAYTRELLDRHARDHV---HMPPEVWTD FWKLF
GLB_C_NIPBR/21-135 DVK--KHTVES-MKAVP-VGRDQAQNGIDFYKFFTHHKDLRKFKA--ENFGADDVQKSRFEKQGTALLAVHVLANVY---DNQAVFHGFVRELMNRHEKRGVDPKLWKIIFDDVWVPF
GLB_C_CAEL/10-119 DLC-VKSL EGR-MVGT E--AQNI-ENGNAFYRYFTNFPLRLVYFKGA--EKYTADDVKKSERFDKQQRILLACHLLANVY---TNEEVFKGYVRETINRRIYK---MDPALWMAFFTVF
GLB2_NIPBR/16-114 PISKAQQ-----AQ-----VGKDFYKFFETNHPDLRKYFKGA--ENFTADDVQKSRFEKQGLSGLLSVHILANTF---DNEDVFRAFCRETIDRHVGRG---LDPALWKA FWSVW
GLB_H_TRICO/30-132 DVVPLGSTPEK-L-----ENGREFYKYFTNHQDLRKYFKGA--ETFTADDIAKSDRFFKLLGNQLLLSVHLAADTY---DNEMIFRAFVRDIDRHVDRG---LDPKLWKEFWSIY
Q20638_CAEL/74-184 EKELLRTWSD-EFD-----NLYELGSAIYCYIFDHNPNCKQLFRF--ISKYQGD EWKESKERSQALKFVQTLAQVVKNIYHMERTESFLYMGVQKHVKFADRG---FKHEYWDIFQDAM
Q19601_CAEL/105-215 ERILLQSWRK-TRK-----TGADHIGSKIFFMVLTAAQDIIKATFGL-L-EKIPTGRLLKYDPRFRQHALVYTKTLDVFIRNL---DYPGKLEVYFENLGKRVAMQGRGFEPGYWETFAECM
Q18311_CAEL/32-140 TKKLV IQEWPR-VLA-----QCPELFTFIWHKSA TRSTSIKLAFGL-I-AE-N--ESPMQNAALGLSSTIQAFYKLLITYE-L-NDDQVREACQLGARHVDFIS-RGFNSHFWDFILVCM

```

seq> AWRTWEAPIFLKRYSTLPGGRAS...

Family power

```

      10      20      30      40      50      60      70      80      90     100     110     120
HBAZ_CAPHI/7-107 ERTIILSLWSK-IST-----QADVIGTETLERLFCYPAQAKTYEPHF--DLH-----SGSAQLRAHGSKVVAAGDAVKS-----D-NVTSALSKLSELHAYVL---RVDPVNFKFLSHCL
HBA3_PLEWA/7-107 EKALVVGLCGK-ISG-----HCDALGGEALDRLEAFSGQTRTYFSHF--DLS-----PGSADVRRHGKVLSAIGEAAKHI-----D-SMDQALSKLSDLHAYNL---RVDPGNFQLLSHCI
HBA_CATCL/6-107 DKADVKI AWAK-ISP-----RADEIGAEALGRMLTVYPQTQTYFANW-ADLS-----PGSGPVKHKKKVIMGAIGDAVTKF-----D-DLLGGLASLSELHASKL---RVDPSPNFKILANCI
HBB_HETPO/7-106 ELHEITTTWKS-I-----DKHSLGAKALARMFIVYPWTTRYFANL-KEFT-----ACSYGVKEHAKKVTGALGVAVTHL-----G-DVKSQFTDLSKKHAEEL---HVDVESFKLLAKCF
HBB_SQUAC/8-112 EKALVNVAWTK-T-----DHQAVVAKALERLFVVPWTQTYFVKFNGKFH-----ASDSTVQTHAGKVVSAITVAYNHI-----D-DVKPHFVELSKKHYEEL---HVDPENFKLLANCL
HBB1_CYGMA/8-112 ELTIINDIFSH-L-----DYDDIGPKALSRLIIVYPWTQRYFSGF--GNLYNAEAIIGNANVAAHGIKVLHGLDRGLKNM-----D-NIVDAYAELSTLHSEKL---HVDPDNFKLLSDCI
HBB1_XENBO/7-111 DRQLINSTWGK-V-----CAKTIGKEALGRLLWTPWTQRYFSFF--GNLSADAVFHNEAAHGEKVVTSIGEAIKHM-----D-DIKGYAQLSKYHSETL---HVDPCNFKRFGGCL
HBB1_UTCT/1-105 GGSDDVSFAFLAK-V-----DKRAVGGAEALARLLIIVYPWTQRYFSFF--GNLGSADAISHNSKVLAHGQRVLDSIEEGLKHP-----Z-BLKAYYAKLSERHSGEL---HVDPANFYRLGNVL
HBB1_LEPPA/7-111 EKQYIVSVFSK-I-----DVDHVGANTLERVLIVYPWTQRYFNSF--GDLSSPGAIKHNKVSAGHGRKVLAAIECTRH-----G-NIKGHLANLSHLHSEKL---HVDPHNFRVLGQCL
HBB2_XENLA/8-112 EKAAITSVWQK-V-----NVEHDGHDALGRLLIIVYPWTQRYFSNF--GNLSNSAAVAGNAKVQAHGKKVLSAVGNAISHI-----D-SVKSSLQQLSKIHATEL---HVDPENFKRFGGVL
HBB1_ALLMI/7-111 ERKFIVDLWAK-V-----DVAQCGADALSRLIIVYPWKRRYFENF--GKMCNAHDI LHNKSKVQEHGKKVLSFGEAVKHL-----D-NIKGHFANLSKLHCEKF---HVDPENFKLLGDI I
HBB0_MOUSE/8-112 EKAAITSIWDK-V-----DLEKVGGETLGRLLIIVYPWTQRYFQKF--GNLSSAQAIMGNPRIKAHGKKVLTSLGLAVKNM-----D-NLKETFAHLSLHCDKL---HADPENFKLLGNML
HBBN_AMMLE/2-106 BKALITGFWSK-V-----KVBZVGAZALGRLLVVPWTZRFPNF--GBLSSABAVMBBBAKVKAHGKKVLSFSBGLKHL-----B-BLKGAFAFSLSZLHCBKL---HVBFPZBFRLLGBVL
HBA1_LEPPA/7-108 DEVLIKEAWGL-L-H-----QIPNAGGEALARMFSCYPGTQKSYFPHFCHDFS-----ANNEKVKHHGKKVVDAIGQGVQHL-----H-DLSSCLHTLSEKHAREL---MVDPCNFQYLI EAI
HBA1_TORMA/6-107 NKKAIKNLLQK-IHS-----QTEVLGAELARLIECHPQTQKSYFQKF--SGFS-----ANDKRVKHHGALVLKALVDTNKHL-----D-DLPHHLNKLAEKHGKGL---LVDPHNFKLLSDCI
HBA_SQUAC/6-107 DKTAKHLLTGS-LRT-----NAEAWGAESLARMFATTPTSTKTYSKFF--TDFS-----ANGKRVKAHGKKVLSAVADATDHL-----D-NVAGHLDPLAVLHGTTL---CVDPHNFPLLTQCI
HBA_HETPO/13-114 DRAELAALSKEV-LAQ-----NAEAFGAELARMFVYATKSYFQDY--KDFT-----AAAPSIKAHGAHVVTALAKACDHL-----D-DLKTHLHLKATFHGSEL---KVDPANFYLSYCL
GLB1_TYLHE/7-110 QRIKVKQQAQ-VYSV-----GESRTDFAIDVFNNFFRTNDRS--FNRVNGDNV-----YSPFEKAHMRVRFAGFDILISVL---DDKPVLDQALAHYAFAFKQFG---TIP--FKAFGQTM
GLB4_LUMTE/11-120 DRREIRHIWDD-VWSSS--FTDRRVAIVRAVFDDLFKHYPTSKALFERVNDIDEP-----ESGEFKSHLVRVANGLDLLINLL---DDTLVLQSHLGHADQHIQRK---GVTKEYFRGIGEAF
GLB3_TYLHE/8-117 DRHEVLNWKKG-IWSAE--FTGRRVAIGQAI FQELFAENAKGVFERVNDV-K-----PSEADWKAHVIRVINGLDLAVNLL---EDPKALQEELKHLARQHRERS---GVKAVYFDEMEKAL
GLB4_TYLHE/8-117 DRREVQALWRS-IWSAE--DTGRRTLIGRLLFEELEIDGATKGLFERVNDVDT-----HSPFEFAHVLRVVNGLDITLIGVL---GDSD-TLNSLIDHLEAQHAKARA---GFKTVYFKEFGKAL
GLB2_TYLHE/9-115 QRLKVKQQAQ-AYGVS---GHERVELGIALWKSMEFQDNDARDLFRVHGEDV-----HSPAFAEHMARVFNGLDRVSSLL---TDEPVLSAQLEHLRQHIKLG---ITGHMFNLMRTGL
GLB2_LUMTE/8-114 EGLKVKSEWGR-AYGS---GHDRFAFSQAIWRATFAQVPESRSFRVHGGDT-----SHPAFIAHAERVVLGGLDIAISTL---DQPATLKEELDHLQVQHEGRK---IPDNYFDAFKTAI
GLB_TUBTU/6-112 QRFKVKHQWAE-AFGT---SHHRLDFGLKLWNSIFRDAPEIRGLFRVVDG-N-----AYSAEFAHAERVVLGGLDMTISLL---DDQAAFDQAHLKLSQHAERN---IKADYYGVFVNEL
GLB3_LAMSP/7-113 QRLKVKRQWAE-AYGS---GNDREFFGFHFIWTHVFKDAPSARDLFRVVDGNI-----HTPAFRAHATRVLGGGLDMCJALL---DDEGVNLTQLAHLASQSSRG---VSAAQYDVVEHSV
GLB_PAREP/8-117 QDILLKELGPH-V-DT---PAHIVETGLGAYHALFTAHQYIIHFSRL-EG-HTIENVMQSEGIKHYARTLTAIVHMLKEI---SNDAEVKKIAAQYGKDHTRSK---VTKDEFMSSGEPIF
Q21978_CAEL/165-283 SCEVWADSWRL-VESRSSAAETSACFGLFVFQRFVFSKIPMLRPLFGL-SESDDVDFLDPNHPVRRHARLFTSILHISVKNV---DELEAQVAPTVEKYGERHYRPDITPHMT EENVRVFCQAI
GLB_PSED/21-134 TRELCMKSLH-AKVGTSKEAKQDGDIDLYKHMFEHY PAMKKYFKNR--ENYTPADVQKDPFFIKQGNILLACHVLCATY---DDRETFDAYVGE LMARHERDHV---KIPNDVWNHFWEHF
GLB_ASCU/21-134 TRELCMKSLH-AKVDT--SNEARQDGDIDLYKHMFEHY PPLRKYFKNR--EETVAEDVQNDPFFAKQGGKILLACHVLCATY---DDRETFNAYTRELLDRHARDHV---HMPEVWTDVFWKLF
GLB_C_NIPBR/21-135 DVK--KHTVES-MKAVP-VGRDKAQNGIDFYKFFTHHKDLRKFKA--ENFGADDVQKSRFEKQGTALLAVHVLANVY---DNQAVFHGFVRELMNRHEKRGVDPKLWKIIFDDVWVPF
GLB_C_CAEL/10-119 DLC-VKSL EGR-MVGT E--AQNI-ENGNAFYRYFTNFPLRLVYFKGA--EKYTAADVKKSERFDKQQRILLACHLLANVY---TNEEVFKGYVR ETINRRIYK---MDPALWMAFFTTF
GLB2_NIPBR/16-114 PISKAQQ-----AQ-----VGKDFYKFFETNHPDLRKYFKGA--ENFTADDVQKSRFEKQGLSGLLSVHILANTF---DNEDVFRAFCRETIDRHVGRG---LDPALWKA FWSVW
GLB_H_TRICO/30-132 DVVPLGSTPEK-L-----ENGREFYKYFTNHDQLRKYFKGA--ETFTADDIAKSDRFFKLGNQLLLSVHLAADTY---DNEMIFRAFVRDTIDRHVDRG---LDPKLWKEFWSIY
Q20638_CAEL/74-184 EKELLRTWSD-EFD-----NLYELGSAIYCYIFDHNPNCKQLFRF--ISKYQGD EWKESKERSQALKFVQTLAQVVKNIYHMERTESFLYMGVQKHVKFADRG---FKHEYWDIFQDAM
Q19601_CAEL/105-215 ERILLQSWRK-TRK-----TGADHIGSKIFFMVLTAAQDIAKAFGL--EKIPTGRLLKYDPRFRQHALVYTKTLDVFNRL---DYPGKLEVYFENLGKRVAMQGRGFEPGYWETFAECM
Q18311_CAEL/32-140 TKKLV IQEWPR-VLA-----QCPELFT EIWHKSATRSTSIKLAFF--I-AE-N--ESPMQNAALGLSSTIQAFFYKLIITYE-L-NDDQVREACQLGARHVDFIS-RGFNSHFWDIFLVCM

```

seq> AWRTWEAPIFLKRYSTLPGGRAS...

Family power

score(i, aa)

Sequence logo visualization showing amino acid conservation across 120 positions. The y-axis represents the log-odds of an amino acid being at a position relative to the background frequency. The x-axis shows positions 1 to 120. A red box highlights positions 45 and 46, which show high conservation for Aspartate (D) and Asparagine (N). A blue arrow points from the 'score(i, aa)' text to the sequence logo.

Sequence alignment (positions 1-120):

```

HBAZ_CAPHI/7-107  ERTIILSLWSK-I-ST-----QADVIGTETLERLFCYDQAKTYEPHF--DLH-----SGSAQLRAHGSKVVAAGDAVKS I-----D-NVTSALSCLKSELHAYVL---RVDPVNFKFLSHCL
HBA3_PLEWA/7-107  EKALVVLGCGK-ISG-----HCDALGGEALDRLEAFSGQTRTYFSHF--DLS-----PGSADVVRHGKVLSAIGEAAKHI-----D-SMDQALSCLKSDLHAYNL---RVDPGNFQLLSHCI
HBA_CATCL/6-107   DKADVKI AWAK-ISP-----RADEIGAEALGRMLTVYPQTQTYFANW--ADLS-----PGSGPVKHGKKVIMGAIGDAVTKF-----D-DLLGGLASLSELHASKL---RVDPSPNFKILANCI
HBB_HETPO/7-106   ELHEITTTWKS-I-----DKHSLGAKALARMFIVYPWTTRYFANL--KEFT-----ACSYGVKEHAKKVTGALGVAVTHL-----G-DVKSQFTDLSKKHAEEL---HVDVESFKLLAKCF
HBB_SQUAC/7-107   EKALVNAVWTK-T-----DHQAVVAKALERLFVVPWTQTYFVKFNGKFFH-----ASDSTVQTHAGKVVSAITVAYNHI-----D-DVKPHFVELSKKHYEEL---HVDPENFKLLANCL
HBB1_CYGMA/8-112  ELTIINDIFSH-L-----DYDDIGPKALSRLIIVYPWTQRYFSGF--GNLYNAEAIIGNANVAAHGIKVLHGLDRGLKNM-----D-NIVDAYAELSTLHSEKL---HVDPDNFKLLSDCI
HBB1_XENBO/7-111  DRQLINSTWKG-V-----CAKTIGKEALGRLLWTPWTQRYFSFF--GNLSADAVFHNEAAHGEKVVTSIGEAIKHM-----D-DIKGYAQLSKYHSETL---HVDPCNFKRFGGCL
HBB1_UTCT/1-105   GGSDDVSFAFLAK-V-----DKRAVGGAEALARLLIIVYPWTQRYFSFF--GNLGSADAISHNSKVLAHGQRVLDSIEEGLKHP-----Z-BLKAYYAKLSERHSGEL---HVDPANFYRLGNVL
HBB1_LEPPA/7-111  EKQYIVSVFSK-I-----DVDHVGANTLERVLIVYPWTQRYFNSF--GDLSSPGAIKHNKVSAAHGRKVLAAIIECTRH-----G-NIKGHLANLSHLHSEKL---HVDPHNFRVLGQCL
HBB2_XENLA/8-112  EKAAITSVWQK-V-----NVEHDGHDALGRLLIIVYPWTQRYFSNF--GNLSNSAAVAGNAKVQAHGKKVLSAVGNAISHI-----D-SVKSSLQQLSKIHATEL---HVDPENFKRFGGVL
HBB1_ALLMI/7-111  ERKFIVDLWAK-V-----DVAQCGADALSRLIIVYPWKRRYFEHF--GKMCNAHDI LHNKSKVQEHGKKVLSFGEAVKHL-----D-NIKGHFANLSKLHCEKF---HVDPENFKLLGDI I
HBB0_MOUSE/8-112 EKAAITSIWDK-V-----DLEKVGGETLGRLLIIVYPWTQRYFQKF--GNLSSAQAIMGNPRIKAHGKKVLTSLGLAVKNM-----D-NLKETFAHLSLHCDKL---HADPENFKLLGNML
HBBN_AMMLE/2-106  BKALITGFWSK-V-----KVBZVGAZALGRLLVVPWTZRFFNF--GBLSSABAVMBBAKVKAHGKKVLSFSBGLKHL-----B-BLKGAFAFSLSZLHCBKL---HVBFPZBFRLLGBVL
HBA1_LEPPA/7-108  DEVLIKEAWGL-L-H-----QIPNAGGEALARMFSCYPTKTSYFPHFCHDFS-----ANNEKVKHHGKKVVDAIGQGVQHL-----H-DLSSCLHTLSEKHAREL---MVDPCNFQYLI EAI
HBA1_TORMA/6-107  NKKAINKLLQK-IHS-----QTEVLGAELARLFECHPQTQTSYFPRF--SGFS-----ANDKRVKHHGALVLKALVDTNKHL-----D-DLPHHLNKLAEKHGKGL---LVDPHNFKLLSDCI
HBA_SQUAC/6-107   DKTAKHLLTGS-LRT-----NAAEAWGAESLARMFATTPTSTKTYSKFF--TDFS-----ANGKRKVAHGKKVLSAVADATDHL-----D-NVAGHLDPLAVLHGTTL---CVDPHNFPLLTQCI
HBA_HETPO/13-114  DRAELAALSKEV-LAQ-----NAAEAFGAELARMFVYAATKSYFKDY--KDFT-----AAAPSIKAHGAHVVTALAKACDHL-----D-DLKTHLHLKATFHGSEL---KVDPANFYLSYCL
GLB1_TYLHE/7-110  QRIKVKQQAQ-VYSV---GESRTDFAIDVFNNFFRTNDRS--FNRVNGDNV-----YSPFEKAHMRVRFAGFDILISVL---DDKPVLDQALAHYAFAFKQFG---TIP--FKAFGQTM
GLB4_LUMTE/11-120 DRREIRHIWDD-VWSSS--FTDRRVAIVRAVFDDLFKHYPTSKAFERVNIDEP-----ESGEFKSHLVRVANGLDLLINLL---DDTLVLQSHLGHADQHIQRK---GVTKEYFRGIGEAF
GLB3_TYLHE/8-117  DRHEVLNDWKG-IWSAE--FTGRRVAIGQAI FQELFADNAKGYFGRVND-K-----PSEADWKAHVIRVINGLDLAVNLL---EDPKALQEELKHLARQHRERS---GVKAVYFDEMEKAL
GLB4_TYLHE/8-117  DRREVQALWRS-IWSAE--DTGRRTLIGRLLFEELFEIDGATKGLFRVNVDDT-----HSPPEFAHVLRVVNGLDTLIGVL---GDSD-TLNSLIDHLAEQHKARA---GFKTVYFKEFGKAL
GLB2_TYLHE/9-115  QRLKVKQQAQ-AYGV---GHERVELGIALWKSMAQPDNDARDLFRVHGEDV-----HSPAFAEHMARVFNGLDRVISL---TDEPVLSAQLEHLRQHIKLG---ITGHMFNLMRTGL
GLB2_LUMTE/8-114  EGLKVKSEWGR-AYGS---GHDRFAFSQAIWRATFAQVPESRSFRVHGGDT-----SHPAFIAHAERVVLGGLDIAISTL---DQPATLKEELDHLQVQHEGRK---IPDNYFDAFKTAI
GLB_TUBTU/6-112   QRFKVKHQWAE-AFGT---SHRRLDFGLKLWNSIFRDAPEIRGLFRVVDG-N-----AYSAEFAHAERVVLGGLDMTISLL---DDQAAFDQAHLKLSQHAERN---IKADYYGVFVNEL
GLB3_LAMSP/7-113  QRLKVKRQWAE-AYGS---GNDREFFGFHIWTHVFKDAPSARDLFRVVDGNI-----HTPAFRAHATRVLGGGLDMCJALL---DDEGVLNTQLAHLASQSSRG---VSAAQYDVVEHSV
GLB_PAREP/8-117   QDILLKELGPH-V-DT---PAHIVETGLGAYHALFTAHQYIIHFSRL--EG-HTIENVMQSEGIKHYARTLTEAIVHMLKEI---SNDAEVKKIAAQYGKDHTRSK---VTKDEFMSSGEPIF
Q21978_CAEL/165-283 SCEVWADSWRL-VESRSSAAETSACFGLFVFQRFVFSKIMPLRPLFGL--SESDDVDFLDPDNHVPVRRHARLFTSILHISVKNV---DELEAQVAPTVEKYGERHVRPDITPHMT EENVRVFCQAI
GLB_PSED/21-134   TRELCKMSLEH-AKVG---SKEAKQDGDIDLYKHMFEHYPMAMKKYFKNR--ENYTPADVQKDPFFIKQGNILLACHVLCATY---DDRETFDAYVGEELMARHERDHV---KIPNDVWNHFWEHF
GLB_ASCU/21-134   TRELCKMSLEH-AKVD---SNEARQDGDIDLYKHMFEHYPLRKYFKNR--EETAEADVQNDPFFAKQGGKILLACHVLCATY---DDRETFNAYTRELLDRHARDHV---HMPPEVWTDVWKL F
GLB_C_NIPBR/21-135 DVK--KHTVES-MKAVP-VGRDQAQNGIDFYKFFTHHKDLRKFKA--ENFGADDVQKSRFEKQGTALLAVHVLANVY---DNQAVFHGFVRELMNRHEKRGVDPKLWKIIFDDVWVPF
GLB_C_CAEL/10-119 DLC-VKSL EGR-MVGT E--AQNI--ENGNAFYRYFTNFPLRLVFKGA--EKYTADDVKKSERFDKQQRILLACHLLANVY---TNEEVFKGYVRETINRRIYK---MDPALWMAFFTFVF
GLB2_NIPBR/16-114 PISKAQQ-----AQ-----VGKDFYKFFETNHPDLRKYFKGA--ENFTADDVQKSRFEKQGLSGLLSVHILANTF---DNEDVFRAFCRETIDRHVGRG---LDPALWKAFFSVW
GLB_C_TRICO/30-132 DVVPLGSTPEK-L-----ENGREFYKYFTNHQDLRKYFKGA--ETFTADDIAKSDRFFKLLGNQLLLSVHLAADTY---DNEMIFRAFVRDIDRHVDRG---LDPKLWKEFWSIY
Q20638_CAEL/74-184 EKELLRTWSD-EFD-----NLYELGSAIYCYIFDHNPNCKQLFRF--ISKYQGD EWKESKFRSQA LKFVQT LAQVVKNIYHMERTESFLYMGVQKHVKFADRG---FKHEYWDIFQDAM
Q19601_CAEL/105-215 ERIILLEQSWRK-TRK-----TGADHIGSKIFFMVLTAAQDIAKAFGL--EKIPTGRLLKYDPRFRQHALVYTKTLDVFI RNLL---DYPGKLEVYFENLGKRHVAMQG--RGFEPGYWETFAECM
Q18311_CAEL/32-140 TKKLV IQEWPR-VLA-----QCPELFT EIWHKSATRSTSIKLAFF--I-AE-N--ESPMQNAALGLSSTIQAFYKLLITYE-L-NDDQVREACEQLGARHVDFIS--RGFNSHFWDIFLVCM
  
```

seq> AWRTWEAPIFLKRYSTLPGGRAS...

Sequence-profile alignments

- Position specific substitution matrices
- profile-hidden Markov models

Functions, organisms, structures

Marco Punta

Prokaryotes



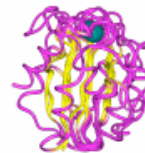
~4400 families

Signalling, extracellular and
chromatin-associated proteins



~1000 domains

Superfamily 1.75
HMM library and genome assignments server



Structural domains from SCOP

Gene3D
Structural domains from CATH

No limits, domains

Marco Punta

The logo for Pfam, consisting of the word "Pfam" in a bold, blue, sans-serif font.

~14000 families

The logo for ProDom, featuring the word "ProDom" in a bold, black, sans-serif font. The letters "o" and "D" are highlighted in green. Above the text is a horizontal bar with colored segments (blue, green, orange, grey). Behind the text is a faint, semi-transparent image of a protein structure or sequence alignment.

No limits, full-length proteins

Marco Punta

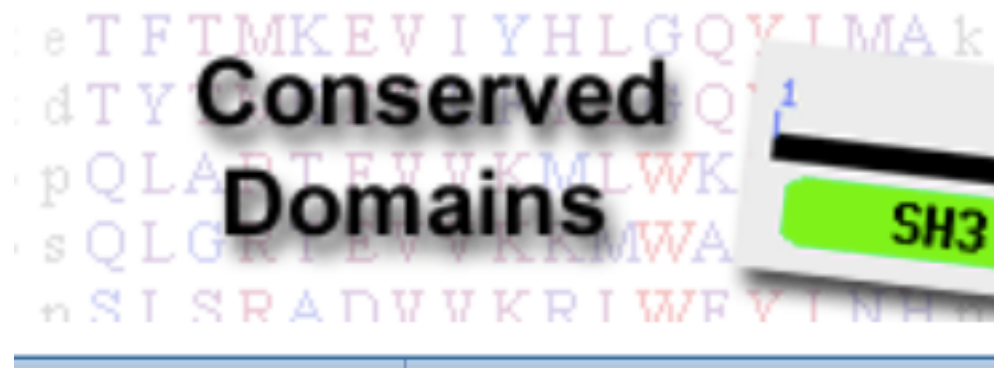


>7000 families, >50000 subfamilies



~2000 families

CDD



Uses RPS-BLAST

Integration

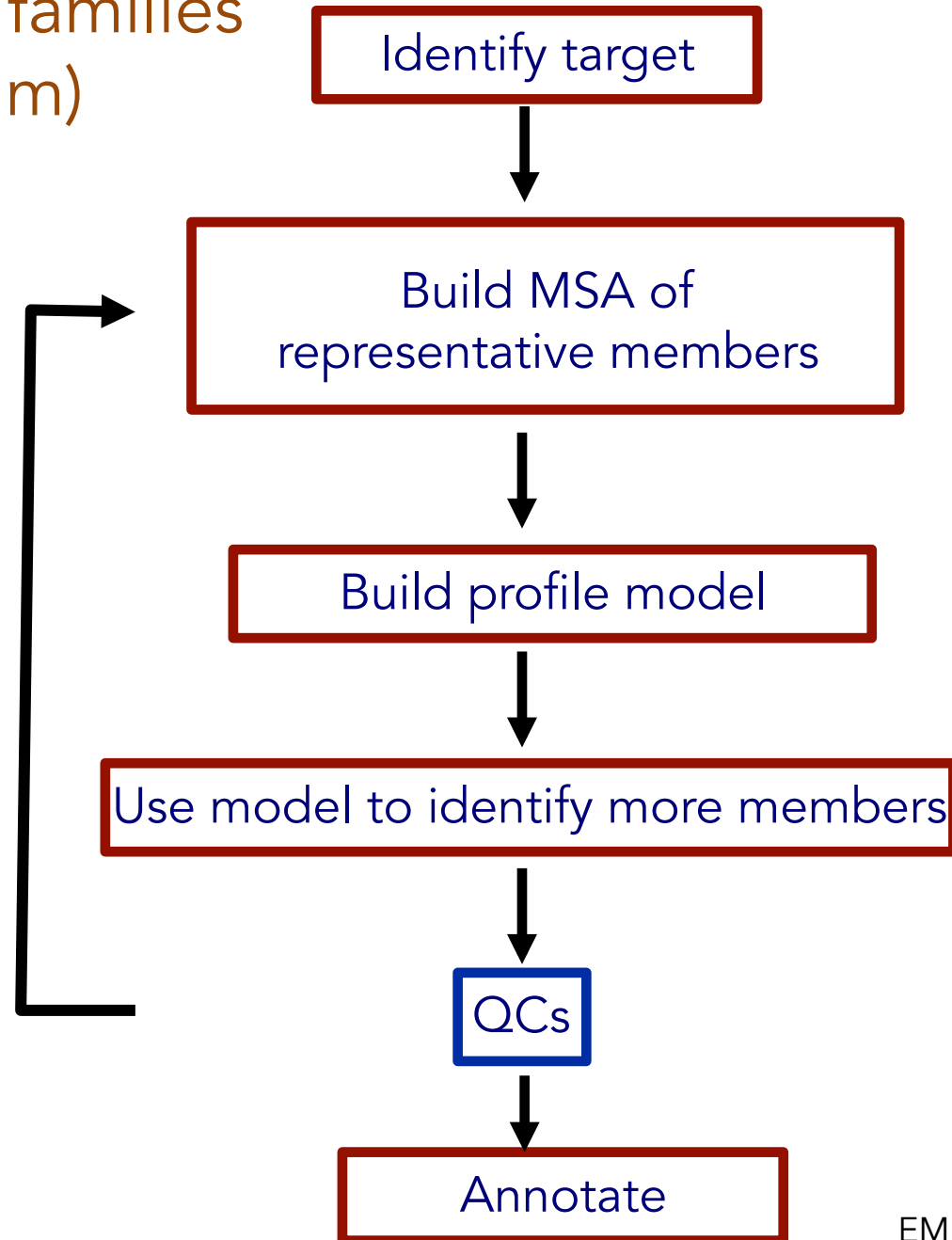
Marco Punta



Building a new protein family

Building families (Pfam)

Marco Punta



Is Pfam a database of globular domain families?

Is Pfam a database of globular domain families?

Pfam entries are classified in one of six ways:

Family:

A collection of related protein regions

Domain:

A structural unit

Repeat:

A short unit which is unstable in isolation but forms a stable structure when multiple copies are present

Motifs:

A short unit found outside globular domains

Coiled-Coil:

Regions that predominantly contain coiled-coil motifs, regions that typically contain alpha-helices that are coiled together in bundles of 2-7.

Disordered:

Regions that are conserved, yet are either shown or predicted to contain bias sequence composition and/or are intrinsically disordered (non-globular).

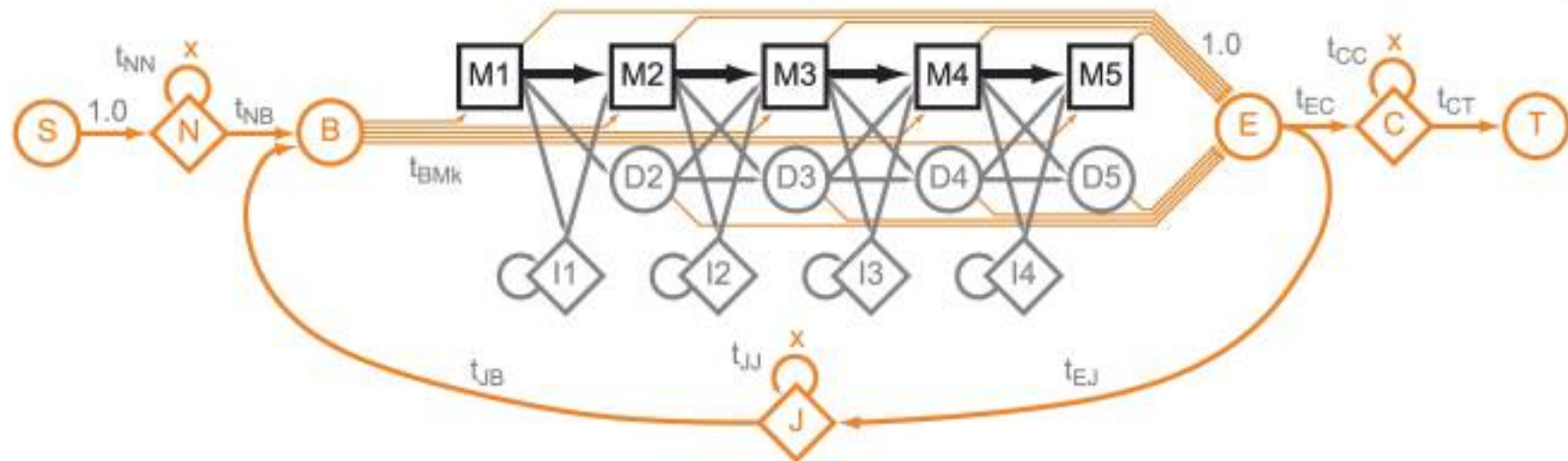
Team Exercise

Building a new Pfam family

- I. (HMMER) search team
- II. Annotation team

HMMER

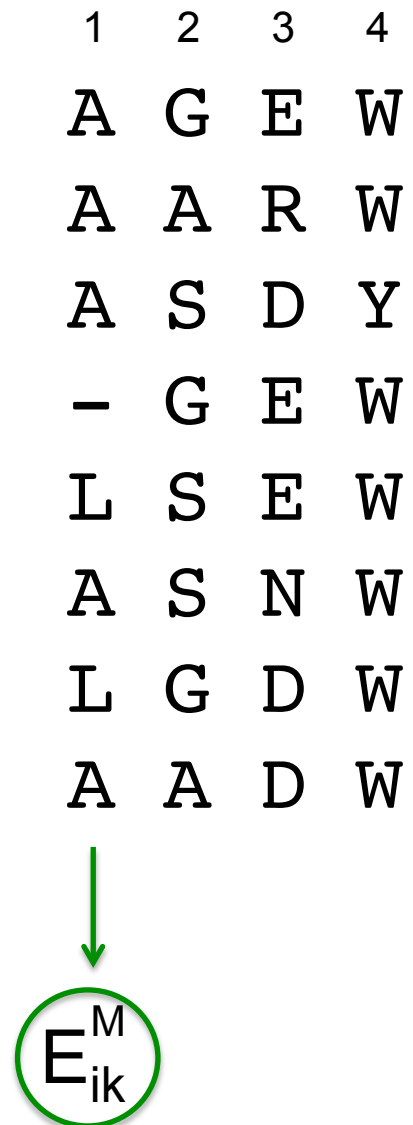
Profile-HMMs [Eddy Bioinformatics 1998]



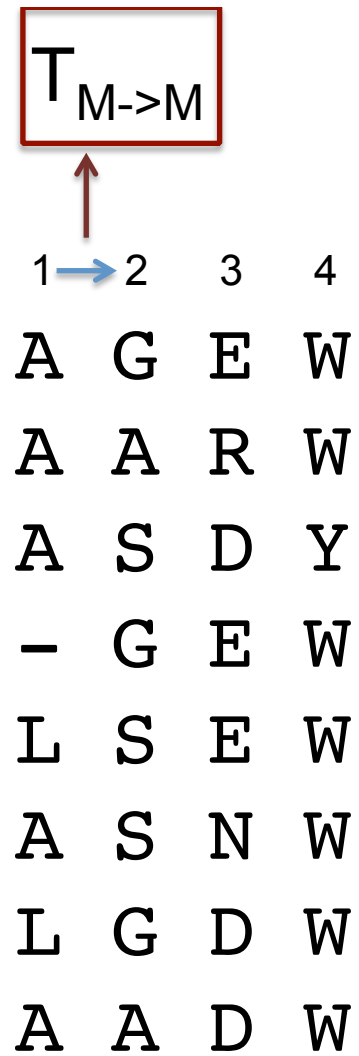
Eddy PLOS CB 2011

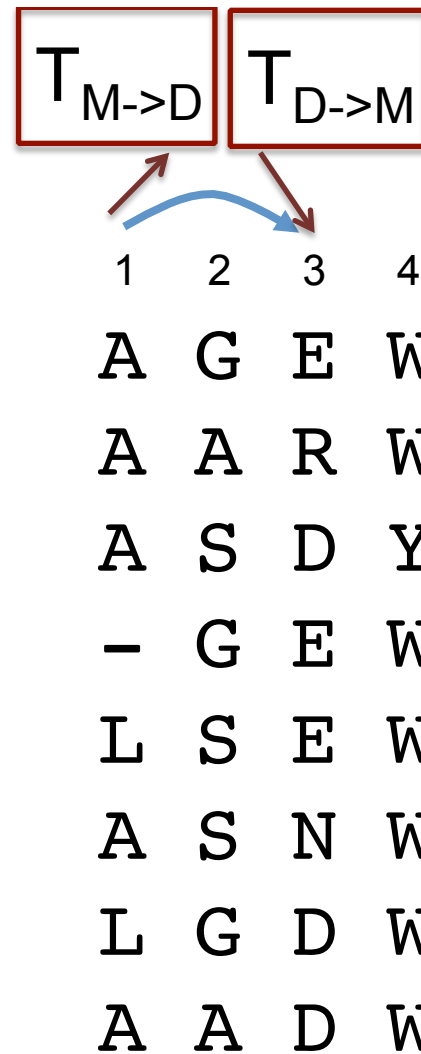
A G E W
A A R W
A S D Y
- G E W
L S E W
A S N W
L G D W
A A D W

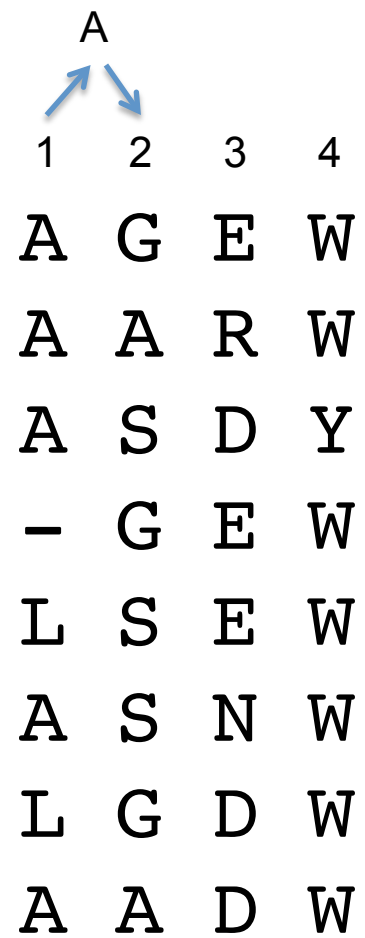
| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| A | G | E | W |
| A | A | R | W |
| A | S | D | Y |
| – | G | E | W |
| L | S | E | W |
| A | S | N | W |
| L | G | D | W |
| A | A | D | W |



| 1 → | 2 | 3 | 4 |
|-----|---|---|---|
| A | G | E | W |
| A | A | R | W |
| A | S | D | Y |
| – | G | E | W |
| L | S | E | W |
| A | S | N | W |
| L | G | D | W |
| A | A | D | W |







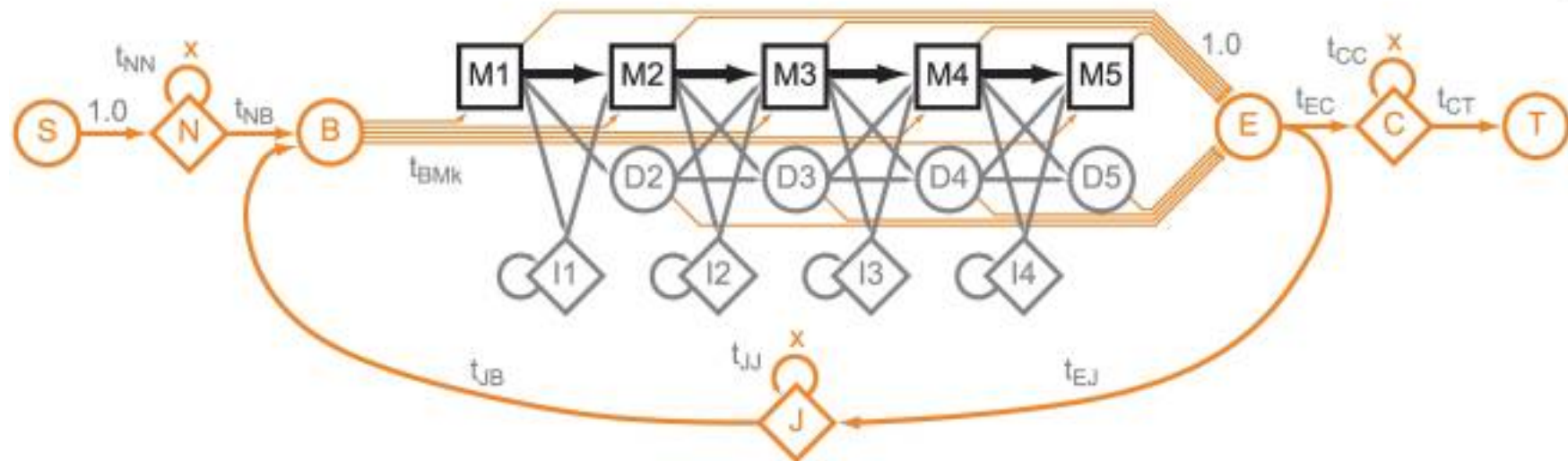


Marco Punta

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| A | G | E | W |
| A | A | R | W |
| A | S | D | Y |
| - | G | E | W |
| L | S | E | W |
| A | S | N | W |
| L | G | D | W |
| A | A | D | W |

HMMER

Profile-HMMs [Eddy Bioinformatics 1998]



Eddy PLOS CB 2011



protein sequence vs protein sequence database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Paste in your sequence or use the [example](#)

Submit

Reset

▼ Sequence Database

Frequently used databases

☐ Reference Proteomes ☐ UniProtKB ☐ SwissProt ☐ PDB

Representative Sets (UniProt)

☐ rp75 ☐ rp55 ☐ rp35 ☐ rp15

Other databases

☐ Ensembl (Plants) ☐ QfO

► Restrict by Taxonomy

<http://www.ebi.ac.uk/Tools/hmmer/search/phmmer>
Finn et al. NAR 2015

► Restrict by Taxonomy ?

▼ Cut-Offs ?

☒ E-value ☐ Bit score

Significance E-values: Sequence Hit

Report E-values: Sequence Hit

Advanced

Comments or questions on the site? Send a mail to hmmer-help@ebi.ac.uk
European Bioinformatics Institute

 Follow @hmm3r

<http://www.ebi.ac.uk/Tools/hmmer/search/phmmer>
Finn et al. NAR 2015

▼ **Customize Results** ?

Select Visible Columns ?

- | | |
|---|---|
| <input type="checkbox"/> Row Count | <input type="checkbox"/> Identical Seqs |
| <input type="checkbox"/> Secondary Accessions and Ids | <input type="checkbox"/> Number of Hits |
| <input checked="" type="checkbox"/> Description | <input type="checkbox"/> Number of Significant Hits |
| <input checked="" type="checkbox"/> Species | <input type="checkbox"/> Bit Score |
| <input type="checkbox"/> Kingdom | <input type="checkbox"/> Hit Positions |
| <input checked="" type="checkbox"/> Known Structure | |

Rows Per Page ?

- ☐ 50
☒ 100
☐ 250
☐ 1000
☐ 2500
☐ All

▼ **Gap Penalties** ?

open extend

Substitution scoring matrix:

▼ **Filters** ?

- ☐ Turn off bias composition filter

<http://www.ebi.ac.uk/Tools/hmmer/search/phmmer>
Finn et al. NAR 2015



protein sequence vs protein sequence database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Paste in your sequence or use the [example](#)

```
>sp|P0A334|KCSA_STRLI pH-gated potassium channel KcsA OS=Streptomyces lividans GN=kcsA PE=1 SV=1
MPPMLSGLLARLVKLLGRHGSALHWRAAGAATVLLVIVLLAGSYLAVLAERGAPGAQLI
TYPRALWWSVETATTVGYGDLVPVTLWGRLVAVVMVAGITSFGLVTAALATWVFGREQE
RRGHFVRHSEKAAEEAYTRTTRALHERFDRLERMLDDNRR
```

Submit

Reset

▼ Sequence Database

Frequently used databases

☒ Reference Proteomes ☐ UniProtKB ☐ SwissProt ☐ PDB

Representative Sets (UniProt)

☐ rp75 ☐ rp55 ☐ rp35 ☐ rp15

Other databases

☐ Ensembl (Plants) ☐ QfO

► Restrict by Taxonomy

<http://www.ebi.ac.uk/Tools/hmmer/search/phmmmer>
Finn et al. NAR 2015



protein sequence vs profile-HMM database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Paste in your sequence or use the [example](#)

```
>sp|P0A334|KCSA_STRLI pH-gated potassium channel KcsA OS=Streptomyces lividans GN=kcsA PE=1 SV=1
MPPMLSGLLARLVKLLGRHGSALHWRAAGAAATVLLVIVLLAGSYLAVLAERGAPGAQLI
TYPRALWWSVETATTGYGDLYPVTWGRVAVVVMVAGITSFGLVTAALATWFGREQE
RRGHFVRHSEKAAEEAYTRTTRALHERFDRLERMLDDNRR
```

Submit

Reset

▼ HMM Database

Protein Families

☒ Pfam ☐ TIGRFAM ☐ Gene3D ☐ Superfamily ☐ PIRSF
([select all](#))([clear all](#))

▼ Cut-Offs

<http://www.ebi.ac.uk/Tools/hmmer/search/phmmer>
Finn et al. NAR 2015



protein alignment/profile-HMM vs protein sequence database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Paste in your alignment/hmm or use the [example](#) 

```
>L5MIQ2_MYODS/799-861 [subseq from] L5MIQ2_MYODS
S--AA-PNNWADFSSTWRWGGG-GR---NGVG-----STS-H-QAGTPRVRSAFPAT
A-TG-SSPSPVLGQGEKVEGL
>M7B574_CHEMY/761-834 [subseq from] M7B574_CHEMY
S--TAATNNWADFSSTWPTS-T-SEKPETDNWDAWAA-----Q-----PSLTVP-SAGQVRQRSAFTPAT
V-TG-SSPSPVLGQGEKVEGL
>G3PBG4_GASAC/811-890 [subseq from] G3PBG4_GASAC
S--L--NNWADFSSTWRWGGG-GR---NGVG-----STS-H-QAGTPRVRSAFPAT
```

Submit

[Reset](#)

▼ [Sequence Database](#)

Frequently used databases

☒ Reference Proteomes ☐ UniProtKB ☐ SwissProt ☐ PDB

Representative Sets (UniProt)

☐ rp75 ☐ rp55 ☐ rp35 ☐ rp15

Other databases

☐ Ensembl (Plants) ☐ QfO

<http://www.ebi.ac.uk/Tools/hmmer/search/phmmer>
Finn et al. NAR 2015



iterative search vs protein sequence database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Paste in your sequence ([example](#)), HMM ([example](#)) or multiple sequence alignment ([example](#))

Submit

Reset

▼ Sequence Database

Frequently used databases

☒ Reference Proteomes ☐ UniProtKB ☐ SwissProt ☐ PDB

Representative Sets (UniProt)

☐ rp75 ☐ rp55 ☐ rp35 ☐ rp15

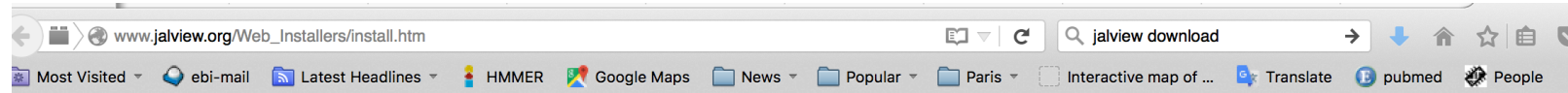
Other databases

☐ Ensembl (Plants) ☐ QfO

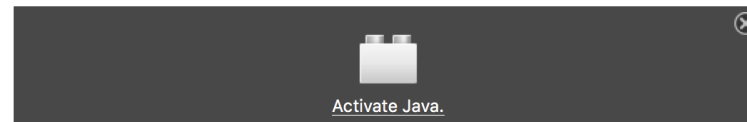
► Restrict by Taxonomy

<http://www.ebi.ac.uk/Tools/hmmer/search/phmmer>
Finn et al. NAR 2015

http://www.jalview.org/Web_Installers/install.htm



Jalview

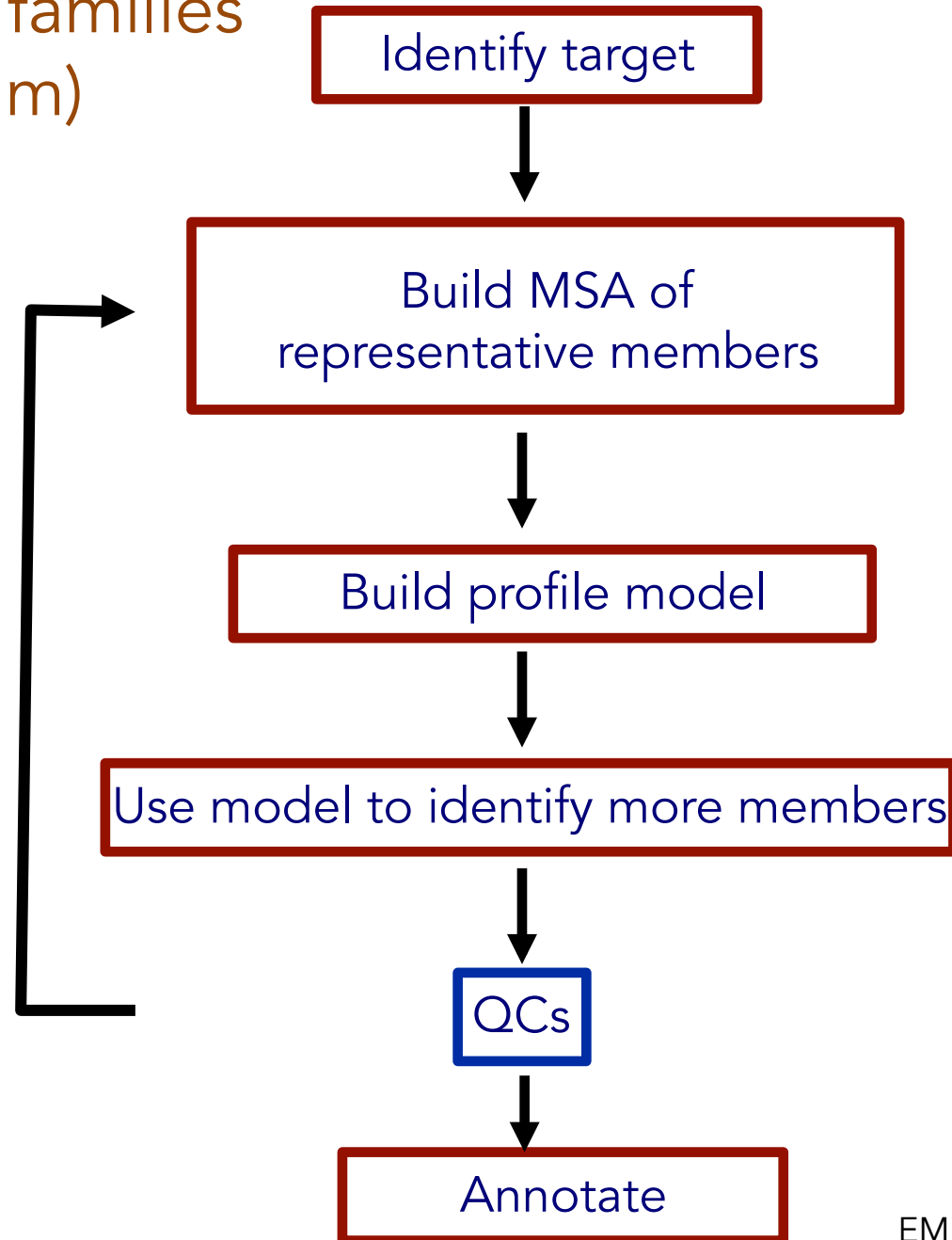


Available Installers

| Platform | includes Java VM | without Java VM | Instructions |
|-------------|-----------------------------------|----------------------------------|----------------------|
| > X MacOSX | Download (92.3M) | Download (34.3M) | View |
| Windows | Download (67.5M) | Download (34.4M) | View |
| IBM AIX | | Download (36.9M) | View |
| HP HPUX | Download (102.1M) | Download (36.9M) | View |
| Linux | Download (82.2M) | Download (36.9M) | View |
| Sun Solaris | Download (86.4M) | Download (36.9M) | View |

Building families (Pfam)

Marco Punta



<http://www.ebi.ac.uk/Tools/hmmer/>



QUICK SEARCH

Paste in your sequence or use the [example](#) ?

search against

☒ Reference Proteomes ☐ UniProtKB ☐ SwissProt ☐ Pfam

Submit

Reset

[Alternative Search Options](#)

The [HMMER web server](#): fast
This site has been designed to provide
coupled with **intuitive and**

Quicks



● Blog News

August, 2015

hmmmer.org is updating

hmmmer.org is moving off of Janelia tonight, into the great cloud. You may see some flakiness as DNS nameservers update.

◆ Download HMMER

Get the latest version

v3.1b2


■ Recent

[HMMER we](#)
R.D. Finn, J.
F. Schreibe
Nucleic Ac
12, 1122-112



protein sequence vs profile-HMM database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Paste in your sequence or use the [example](#) 

```
HEAIGSGDLDLRSAFRRTSLAGAGRRRTSDSHEDAGTLDFSSLLKKRD  
SFRRDSKLEAPAEEDVWEILROAPPSEYERIAFOHGVTDLRGMLKRL  
KGMKQDEKK
```

Submit

Reset

▼ HMM Database

Protein Families

☒ Pfam ☒ TIGRFAM ☒ Gene3D ☒ Superfamily ☒ PIRSF
([select all](#))

<http://www.ebi.ac.uk/interpro/>



InterPro
Protein sequence analysis & classification

Examples: IPR020405, kinase, P51587, PF02932, GO:0007165

[Home](#)[Search](#)[Release notes](#)[Download](#)[About InterPro](#)[Help](#)[Contact](#)

InterPro: protein sequence analysis & classification

InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. We combine protein signatures from a number of member databases into a single searchable resource, capitalising on their individual strengths to produce a powerful integrated database and diagnostic tool. [Read more about InterPro](#)

Analyse your protein sequence

v.54

InterPro 54.0
15th October 2015

Features include:

- An update to PANTHER (10.0).
- Integration of 2138 new methods from the CATH-Gene3D (2), Pfam (1319), SUPERFAMILY (5), ProDom (7) and PANTHER (805) databases.

[Read more](#)

IDA
Domain architecture search

Interproscan 5

[Learn more >>](#)

Documentation

[About InterPro](#): core concepts, update frequency, how to cite, team and consortium members.

Protein focus



[The sweetest thing](#)

Our sugar consumption has reached to such a level

Publications



[The InterPro protein families database: the classification resource after 15 years](#)

Our latest paper describing

<http://www.ebi.ac.uk/interpro/>



Examples: IPR020405, kinase, P51587, PF02932, GO:0007165

[Home](#) [Search](#) [Release notes](#) [Download](#) [About InterPro](#) [Help](#) [Contact](#)

InterPro: protein sequence analysis & classification

InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. We combine protein signatures from a number of member databases into a single searchable resource, capitalising on their individual strengths to produce a powerful integrated database and diagnostic tool. [Read more about InterPro](#)

Analyse your protein sequence

```
HEAIGSGDLRLSAFRRTSLAGAGRTSDSHEDAGTLDFFSLLKKRDSFRDSEKLEAPA
EEDVWEILRQAPPSEYERIAFHQGVTDLRGMLKRLKGMKQDEKK
```

|


Documentation

[About InterPro](#): core concepts,

Protein focus

 [The sweetest](#)

Publications

 [The InterPro protein](#)

v.55 **InterPro 55.0**
10th December 2015

Features include:


- An update to HAMAP (201511.02).
- Integration of 399 new methods from the Pfam (273), SUPERFAMILY (23), PANTHER (48), CATH-Gene3D (11), HAMAP (43) and ProDom (1) databases.

| [Read more](#)

IDA 
Domain architecture search

<http://www.ebi.ac.uk/interpro/>

EMBL-EBI
Services
Research
Training
About us



Home
Search
Results
Software
Help
About

phmmer
hmmsearch
hmmsearch
jackhmmer

protein sequence vs profile-HMM database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Paste in your sequence or use the [example](#)

HEAIGSGDLRLRSFAFRRTSLAGAGRRTSDSHEDAGTLDFFSLLKKRDSFRDSEKLEAPAEEDVWEILRQAPPSEYERIAFOHGVTDLRGMLKRLKGMKQDEKK

Submit
Reset

▼ HMM Database

Protein Families


☒ Pfam
☒ TIGRFAM
☒ Gene3D
☒ Superfamily
☒ PIRSF

(select all)(clear all)

▼ Cut-Offs

EMBL-EBI

Services | Research | Training | About us



[Home](#)
[Search](#)
[Results](#)
[Software](#)
[Help](#)
[About](#)

[phmmer](#)
[hmmscan](#)
[hmmsearch](#)
[jackhmmer](#)

protein sequence vs protein sequence database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Paste in your sequence or use the [example](#)

>protein region

HEAIGSGDLDLRSAFRRTSLAGAGRTSDSHEDAGTLDFFSLLKKRDSFRDRSKLEAPAEEDVWEILRQAPPSEYERIAFQHGVTDLRGMLKRLKGMKQDEKK

Submit

Reset

▼ Sequence Database

Frequently used databases

☒ Reference Proteomes
 ☐ UniProtKB
 ☐ SwissProt
 ☐ PDB

Representative Sets (UniProt)

☐ rp75
 ☐ rp55
 ☐ rp35
 ☐ rp15

Other databases

☐ Ensembl (Plants)
 ☐ QfO

EMBL-EBI

Services
Research
Training
About us

Home
Search
Results
Software
Help
About

Score
Taxonomy
Domain
Download

PHMMER Results

Search Again

Sequence Matches and Features

Pfam

103

disorder

103

hit coverage

hit similarity

✓ disorder

✓ coiled-coil

✓ tm & signal peptide

[Show hit details](#)

Distribution of Significant Hits

■ Bacteria

■ Eukaryota

■ Archaea

■ Viruses

■ Unclassified Sequences

■ Other Sequences

« First
« Previous
Page 1
of 5
Next »
Last »

Significant Query Matches (376) in uniprotrefprot (v.2016-03-02)

Customize

| | Target | Description | Species | E-value |
|---|------------------------------|--|---------------------------------------|---------|
| > | Q3UIK0_MOUSE | Myosin-binding protein C, cardiac-type | Mus musculus | 2.5e-60 |
| > | E9Q9T8_MOUSE | Myosin-binding protein C, cardiac-type | Mus musculus | 5.9e-58 |
| > | Q3TF37_MOUSE | Myosin-binding protein C, cardiac-type | Mus musculus | 2.0e-57 |
| > | MYPC_RAT | Myosin-binding protein C, cardiac-type | Rattus norvegicus | 3.8e-56 |
| > | M3XYE3_MUSPF | Uncharacterized protein | Mustela putorius furo | 1.8e-54 |
| > | MYPC3_MOUSE | Myosin-binding protein C, cardiac-type | Mus musculus | 5.8e-54 |

| Home | Search | Results | Software | Help | About | |
|-----------------------|--------------|--|----------|------|-------|---|
| > | G3PS86_GASAC | Uncharacterized protein (Fragment) | | | | Gasterosteus aculeatus 3.0e-22 |
| > | G3VYD1_SARHA | Uncharacterized protein (Fragment) | | | | Sarcophilus harrisi 4.1e-22 |
| > | F6PYQ5_ORNAN | Uncharacterized protein | | | | Ornithorhynchus anatinus 5.9e-22 |
| > | F6PYS3_ORNAN | Uncharacterized protein | | | | Ornithorhynchus anatinus 6.7e-22 |
| > | H2V4W3_TAKRU | Uncharacterized protein | | | | Takifugu rubripes 9.1e-22 |
| > | G3XAI7_HUMAN | Myosin-binding protein C, slow-type | | | | Homo sapiens 1.0e-21 |
| > | F1Q615_DANRE | Uncharacterized protein | | | | Danio rerio 1.3e-21 |
| > | M7B6S6_CHEMY | Myosin-binding protein C, slow-type (Fragment) | | | | Chelonia mydas 1.4e-21 |
| > | F6Y6C7_HORSE | Uncharacterized protein | | | | Equus caballus 1.4e-21 |
| > | F6Y6F3_HORSE | Uncharacterized protein | | | | Equus caballus 1.4e-21 |
| > | G1R3C3_NOMLE | Uncharacterized protein | | | | Nomascus leucogenys 2.0e-21 |
| > | G3V1V7_HUMAN | Myosin binding protein C, slow type, isoform CRA_e | | | | Homo sapiens 2.1e-21 |
| > | MYPC1_HUMAN | Isoform 5 of Myosin-binding protein C, slow-type | | | | Homo sapiens 2.2e-21 |
| > | MYPC1_HUMAN | Isoform 3 of Myosin-binding protein C, slow-type | | | | Homo sapiens 2.2e-21 |
| > | MYPC1_HUMAN | Myosin-binding protein C, slow-type | | | | Homo sapiens 2.3e-21 |
| > | MYPC1_HUMAN | Isoform 2 of Myosin-binding protein C, slow-type | | | | Homo sapiens 2.3e-21 |
| > | MYPC1_HUMAN | Isoform 10 of Myosin-binding protein C, slow-type | | | | Homo sapiens 2.3e-21 |
| > | F8VZY0_HUMAN | Myosin-binding protein C, slow-type | | | | Homo sapiens 2.4e-21 |
| > | MYPC1_HUMAN | Isoform 8 of Myosin-binding protein C, slow-type | | | | Homo sapiens 2.4e-21 |
| > | MYPC1_HUMAN | Isoform 6 of Myosin-binding protein C, slow-type | | | | Homo sapiens 2.4e-21 |
| > | MYPC1_HUMAN | Isoform 4 of Myosin-binding protein C, slow-type | | | | Homo sapiens 2.4e-21 |
| > | MYPC1_HUMAN | Isoform 7 of Myosin-binding protein C, slow-type | | | | Homo sapiens 2.5e-21 |
| > | G3RPA3_GORGO | Uncharacterized protein | | | | Gorilla gorilla gorilla 2.7e-21 |
| > | F7IK85_CALJA | Uncharacterized protein | | | | Callithrix jacchus 2.9e-21 |
| > | H2LNL0_ORYLA | Uncharacterized protein | | | | Oryzias latipes 2.9e-21 |
| > | H2Q6Q1_PANTR | Uncharacterized protein | | | | Pan troglodytes 3.7e-21 |
| > | F7CWG3_CALJA | Uncharacterized protein | | | | Callithrix jacchus 3.8e-21 |
| > | HOW0Z3_CAVPO | Uncharacterized protein | | | | Cavia porcellus 5.2e-21 |
| > | HOVIL5_CAVPO | Uncharacterized protein | | | | Cavia porcellus 5.4e-21 |
| (show all) alignments | | | | | | Your search took:0.97 secs showing rows 1 - 100 of 439 |
| Search Details | | | | | | « First « Previous Page 1 of 5 Next » Last » |

| | | | |
|--------------------|--|-------------------------------|--------|
| > W5LBC4_ASTMX | Uncharacterized protein | Astyanax mexicanus | 0.0085 |
| > L9KLA3_TUPCH | Titin | Tupaia chinensis | 0.0088 |
| > A0A0P7USR2_9TELE | Titin-like (Fragment) | Scleropages formosus | 0.0091 |
| > E9Q8K5_MOUSE | Titin | Mus musculus | 0.0092 |
| > E9Q8N1_MOUSE | Titin | Mus musculus | 0.0092 |
| > TITIN_MOUSE | Titin | Mus musculus | 0.0092 |
| > E4XET7_OIKDI | Uncharacterized protein | Oikopleura dioica | 0.0094 |
| > I3KJU9_ORENI | Uncharacterized protein | Oreochromis niloticus | 0.0098 |
| > G3HAC6_CRIGR | Titin | Cricetulus griseus | 0.01 |
| > A0A0Q3ZH98_ALLMI | Hemicentin 1 isoform F | Alligator mississippiensis | 0.011 |
| > A0A0Q3ZKQ5_ALLMI | Hemicentin 1 isoform C | Alligator mississippiensis | 0.011 |
| > A0A0Q3ZJ46_ALLMI | Hemicentin 1 isoform G | Alligator mississippiensis | 0.011 |
| > A0A0Q3ZJ19_ALLMI | Hemicentin 1 isoform B | Alligator mississippiensis | 0.011 |
| > A0A0Q3ZKX4_ALLMI | Hemicentin 1 isoform A | Alligator mississippiensis | 0.011 |
| > A0A0Q3ZD94_ALLMI | Hemicentin 1 isoform E | Alligator mississippiensis | 0.011 |
| > A0A0Q3ZJN0_ALLMI | Hemicentin 1 isoform D | Alligator mississippiensis | 0.011 |
| > F1R7N8_DANRE | Uncharacterized protein | Danio rerio | 0.011 |
| > M7BUV6_CHEMY | Myosin-binding protein C, cardiac-type | Chelonia mydas | 0.011 |
| > W5Q9D3_SHEEP | Uncharacterized protein (Fragment) | Ovis aries | 0.012 |
| > E1C7I7_CHICK | Uncharacterized protein | Gallus gallus | 0.014 |
| > G1U9S3_RABIT | Uncharacterized protein | Oryctolagus cuniculus | 0.014 |
| > W4Z4T9_STRPU | Uncharacterized protein | Strongylocentrotus purpuratus | 0.017 |
| > I3KLJ9_ORENI | Uncharacterized protein | Oreochromis niloticus | 0.018 |
| > M3XTP2_MUSPF | Uncharacterized protein | Mustela putorius furo | 0.018 |
| > A0A087X7Y5_POEFO | Uncharacterized protein | Poecilia formosa | 0.023 |
| > G3WA50_SARHA | Uncharacterized protein | Sarcophilus harrisii | 0.023 |
| > W5MH34_LEPOC | Uncharacterized protein | Lepisosteus oculatus | 0.028 |
| > F1PV45_CANLF | Uncharacterized protein | Canis lupus familiaris | 0.028 |
| > H2P803_PONAB | Uncharacterized protein | Pongo abelii | 0.031 |
| > H2QJ24_PANTR | Uncharacterized protein | Pan troglodytes | 0.031 |
| > TITIN_HUMAN | Isoform 5 of Titin | Homo sapiens | 0.031 |
| > TITIN_HUMAN | Isoform 11 of Titin | Homo sapiens | 0.032 |

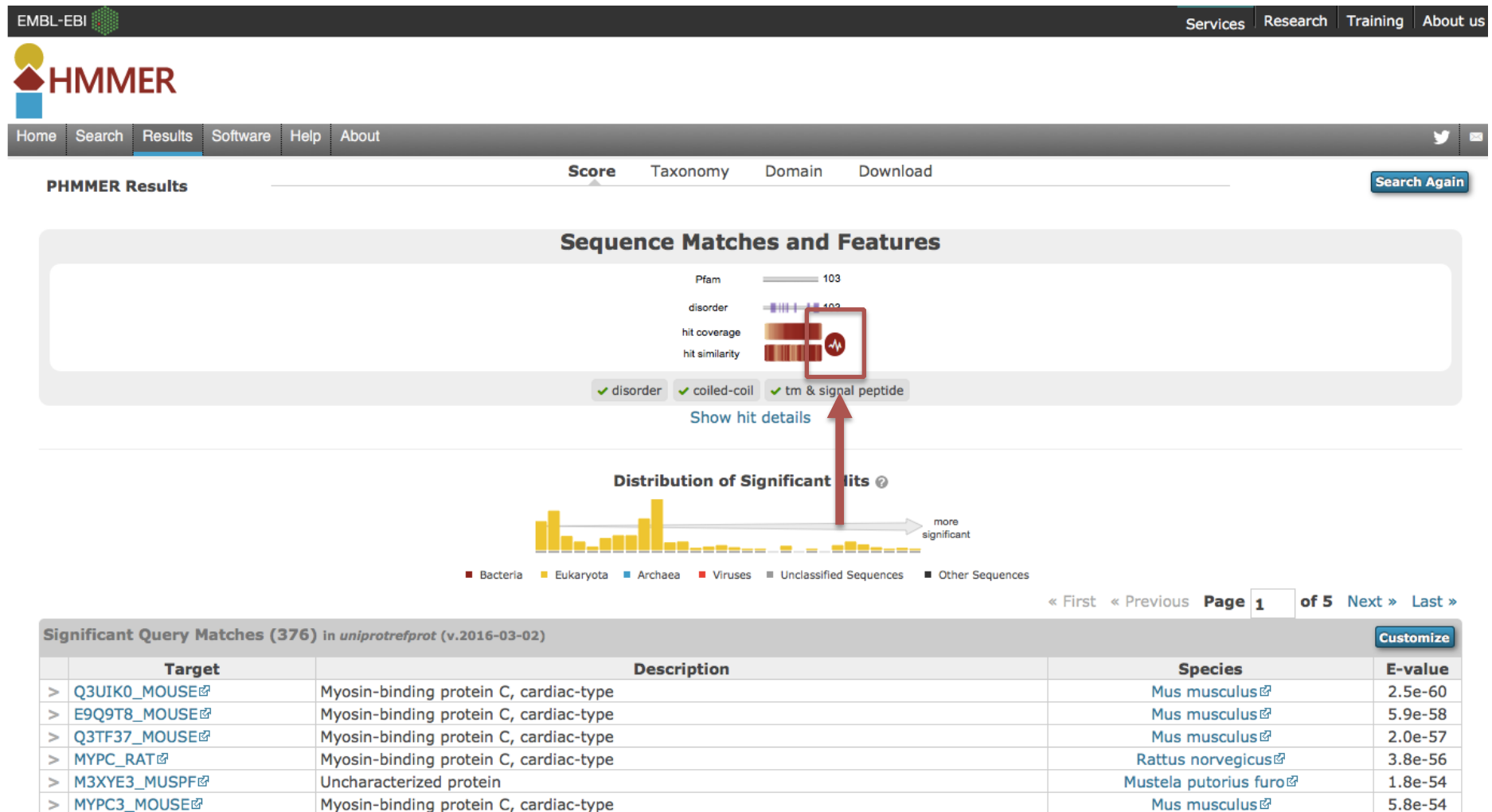
(show all) alignments

Your search took: 0.97 secs

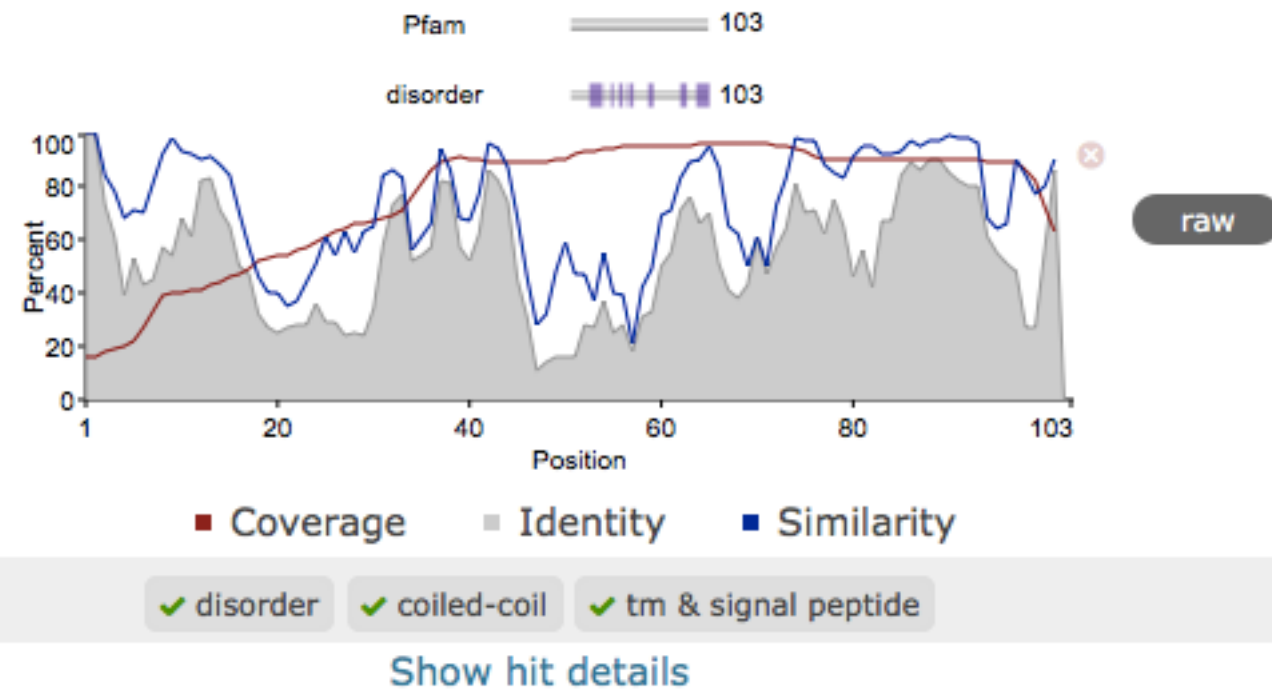
showing rows 301 - 400 of 439

Search Details

« First « Previous Page 4 of 5 Next » Last »



Sequence Matches and Features



EMBL-EBI

[Services](#)
[Research](#)
[Training](#)
[About us](#)

[Home](#)
[Search](#)
[Results](#)
[Software](#)
[Help](#)
[About](#)

[Score](#)
[Taxonomy](#)
[Domain](#)
[Download](#)

[Search Again](#)

PHMMER Results

Sequence Matches and Features

Pfam

103

disorder

103

hit coverage

hit similarity

✓ disorder

✓ coiled-coil

✓ tm & signal peptide

[Show hit details](#)

Distribution of Significant Hits

■ Bacteria

■ Eukaryota

■ Archaea

■ Viruses

■ Unclassified Sequences

■ Other Sequences

« First

« Previous

Page 1

of 5

Next »

Last »

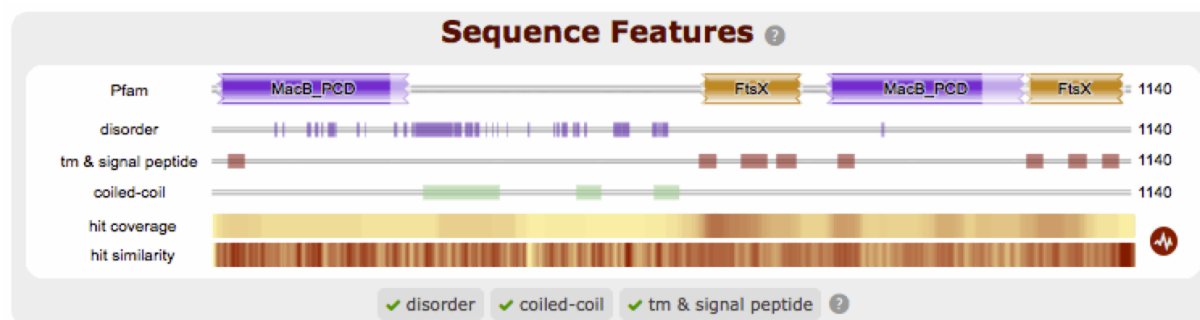
Significant Query Matches (376) in uniprotrefprot (v.2016-03-02)

[Customize](#)

| Target | Description | Species | E-value |
|----------------|--|---------------------------------------|---------|
| > Q3UIK0_MOUSE | Myosin-binding protein C, cardiac-type | Mus musculus | 2.5e-60 |
| > E9Q9T8_MOUSE | Myosin-binding protein C, cardiac-type | Mus musculus | 5.9e-58 |
| > Q3TF37_MOUSE | Myosin-binding protein C, cardiac-type | Mus musculus | 2.0e-57 |
| > MYPC_RAT | Myosin-binding protein C, cardiac-type | Rattus norvegicus | 3.8e-56 |
| > M3XYE3_MUSPF | Uncharacterized protein | Mustela putorius furo | 1.8e-54 |
| > MYPC3_MOUSE | Myosin-binding protein C, cardiac-type | Mus musculus | 5.8e-54 |

Other Sequence Features

When a sequence is searched using hmmscan, phmmer or jackhmmer, the query sequence is also searched with three additional methods to identify sequence features, namely regions of disorder, signal peptides, transmembranes and coiled-coils.



If a search returns no results, then the graphic is not displayed. To make it clear when a search has been run, we have added small indicators at the bottom of the sequence features section. When a search has successfully completed it will be shown with a small green tick (✓) next to it.

Annotation team: results?



Overview

Similar proteins

Structures

Filter view on

Entry type

- ☒ **F** Family
- ☒ **D** Domains
- ☒ **R** Repeats
- ☒ **S** Site

Status

- ☒ **?** Unintegrated

Colour by

[help](#)

- ☒ domain relationship
- ☐ source database

P Protein

Submitted

Length 103 amino acids

Protein family membership

None predicted.

Domains and repeats

None predicted.

GO term prediction

Biological Process

None predicted.

Molecular Function


None predicted.

Cellular Component

None predicted.

EMBL-EBI

[Services](#)
[Research](#)
[Training](#)
[About us](#)



[Home](#)
[Search](#)
[Results](#)
[Software](#)
[Help](#)
[About](#)

[Score](#)
[Download](#)

[Search Again](#)

HMMSCAN Results

Sequence Matches and Features

Pfam

103

disorder

103

✓ disorder

✓ coiled-coil

✓ tm & signal peptide

No hits were found for your query.

TIGRFAM Matches

[Advanced](#)

| Family | | Description | Start | End | Domain E-values | |
|------------------------------------|-----------|-------------|-------|-----|-----------------|-------|
| Id | Accession | | | | Ind. | Cond. |
| No hits were found for your query. | | | | | | |
| Your search took: 0.05 seconds | | | | | | |

Gene3D Matches

[Advanced](#)

| Family | | Description | Region | Start | End | Domain E-values | |
|------------------------------------|-----------|-------------|--------|-------|-----|-----------------|-------|
| Id | Accession | | | | | Ind. | Cond. |
| No hits were found for your query. | | | | | | | |
| Your search took: 0.06 seconds | | | | | | | |

| Superfamily Matches | | | | | | | | | | | | | | |
|------------------------------------|-------------|---------|-----------|-------------|---------|--------|-------------|-----|-----------|-----|-------|-----|--------|-----------|
| Superfamily | | | Family | | | Region | Model Match | | | | | | | |
| | | | | | | | Start | End | Alignment | | Model | | | Bit Score |
| Accession | Description | E-value | Accession | Description | E-value | | | | Start | End | Start | End | Length | |
| No hits were found for your query. | | | | | | | | | | | | | | |
| Your search took: 0.03 seconds | | | | | | | | | | | | | | |

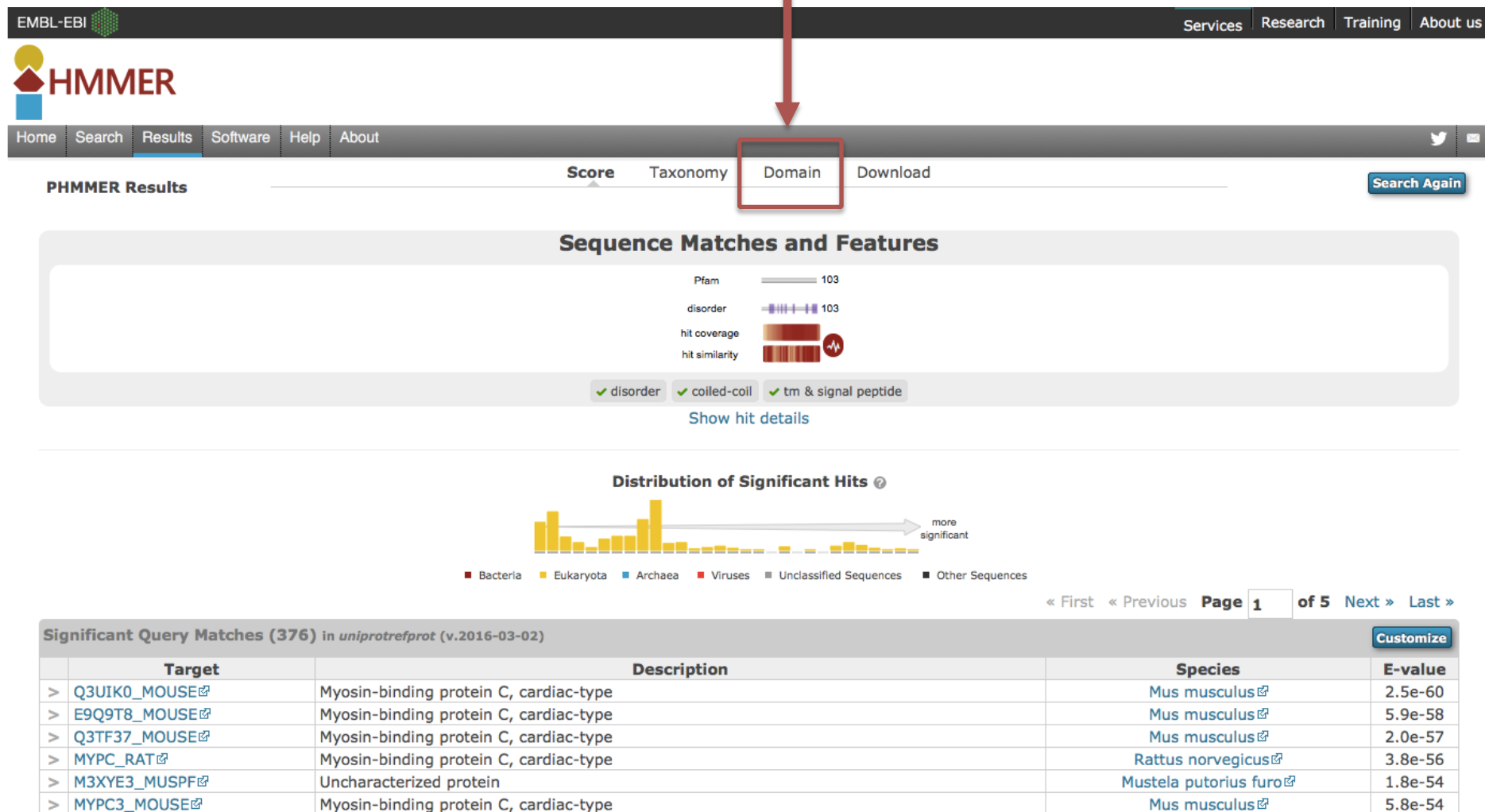
* These hmmscan results have been modified by Superfamily post-processing and family assignment code.

| PIRSF Matches | | | | | | Advanced |
|------------------------------------|-------------|---------|-----------|-------------|---------|----------|
| Family | | | Subfamily | | | Region |
| Accession | Description | E-value | Accession | Description | E-value | |
| No hits were found for your query. | | | | | | |
| Your search took: 0.06 seconds | | | | | | |

* These hmmscan results have been modified by PIRSF post-processing and family assignment code.

[Download](#) your results in various formats.

[Search Details](#)





PHMMER Results

Search Again

Sequence Matches and Features

Pfam 103

disorder 103

✓ disorder ✓ coiled-coil ✓ tm & signal peptide

Show hit details

Jump to the exact match for your query architecture

Domain Architectures

« First « Previous Page 1 of 2 Next » Last »

119
SEQUENCES
Show All

with domain architecture: **I-set, I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set**, example:Q3TF37_MOUSE

View Scores

Sequence Features I-set I-set I-set I-set I-set fn3 fn3 I-set fn3 I-set 1113

45
SEQUENCES
Show All

with domain architecture: **I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set**, example:F6ZHP7_HORSE

View Scores

Sequence Features I-set I-set I-set I-set fn3 fn3 I-set fn3 I-set 063

41
SEQUENCES
Show All

with domain architecture: **I-set, I-set, I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set**, example:Q3UIK0_MOUSE

View Scores

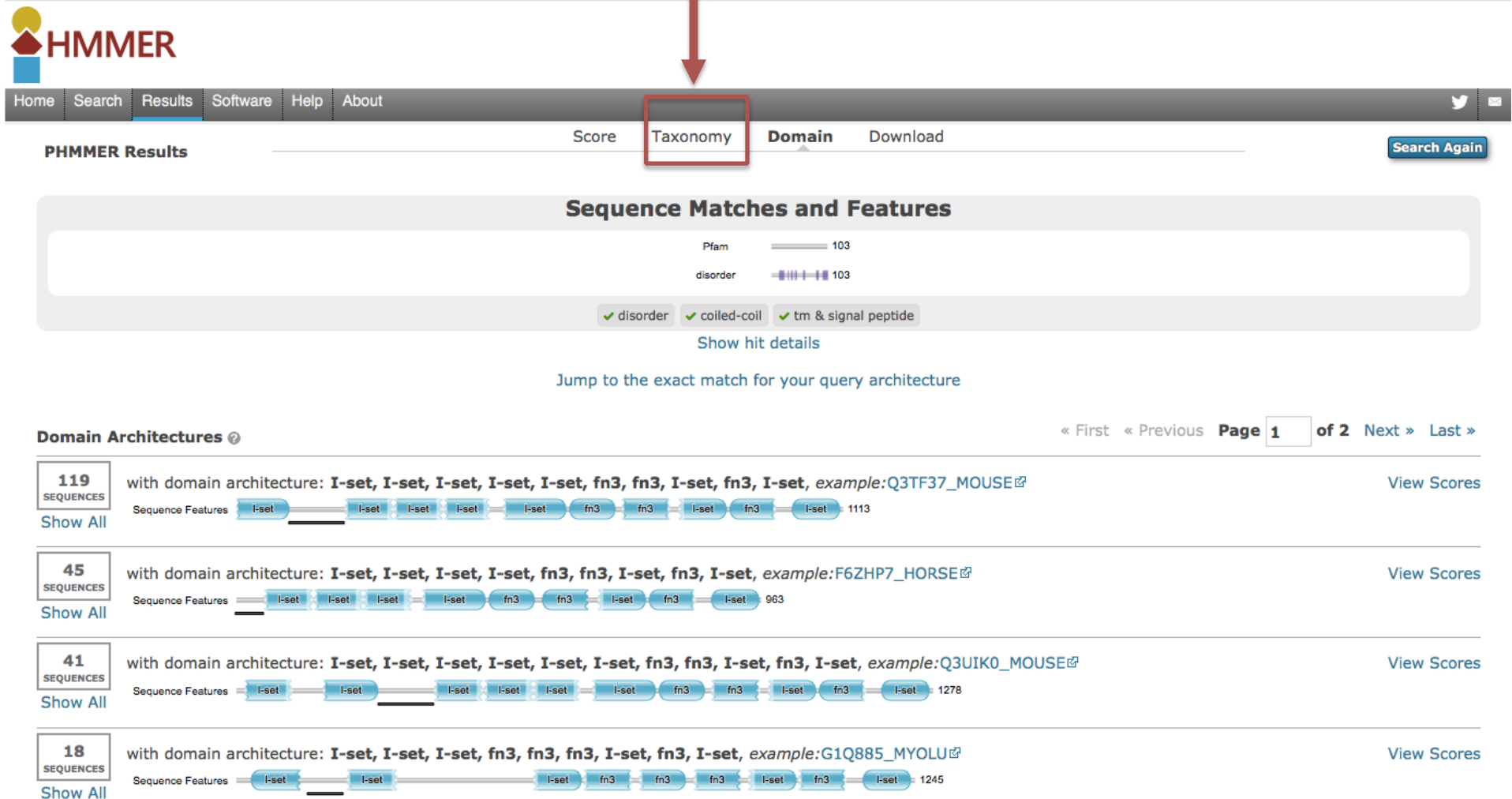
Sequence Features I-set I-set I-set I-set I-set I-set I-set fn3 fn3 I-set fn3 I-set 1278

18
SEQUENCES
Show All

with domain architecture: **I-set, I-set, I-set, fn3, fn3, fn3, I-set, fn3, I-set**, example:G1Q885_MYOLU

View Scores

Sequence Features I-set I-set I-set fn3 fn3 fn3 I-set fn3 I-set 1245



HHMER

Home Search **Results** Software Help About

Score **Taxonomy** Domain Download

PHMMER Results [Search Again](#)

Sequence Matches and Features

Pfam 103
disorder 103

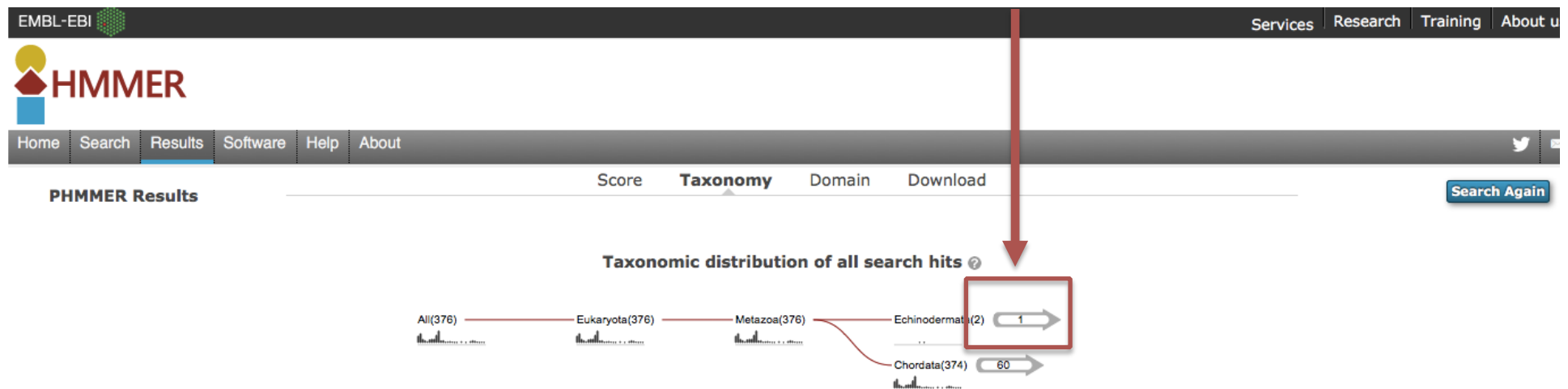
✓ disorder ✓ coiled-coil ✓ tm & signal peptide

[Show hit details](#)


[Jump to the exact match for your query architecture](#)


Domain Architectures [?](#) [« First](#) [« Previous](#) **Page 1** **of 2** [Next »](#) [Last »](#)



| | | |
|--|--|-----------------------------|
| 119 SEQUENCES Show All | with domain architecture: I-set, I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set , example: Q3TF37_MOUSE | View Scores |
| 45 SEQUENCES Show All | with domain architecture: I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set , example: F6ZHP7_HORSE | View Scores |
| 41 SEQUENCES Show All | with domain architecture: I-set, I-set, I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set , example: Q3UIK0_MOUSE | View Scores |
| 18 SEQUENCES Show All | with domain architecture: I-set, I-set, I-set, fn3, fn3, fn3, I-set, fn3, I-set , example: G1Q885_MYOLU | View Scores |




| Species Distribution | | |
|--|-------|----------------------|
| Species | Count | View |
| Homo sapiens | 25 | Show |
| Danio rerio | 22 | Show |
| Takifugu rubripes | 22 | Show |
| Callithrix jacchus | 18 | Show |
| Mus musculus | 15 | Show |
| Gasterosteus aculeatus | 11 | Show |
| Oreochromis niloticus | 11 | Show |
| Canis lupus familiaris | 10 | Show |
| Macaca mulatta | 10 | Show |
| Poecilia formosa | 9 | Show |
| Alligator mississippiensis | 9 | Show |
| Astyanax mexicanus | 9 | Show |
| Gorilla gorilla gorilla | 8 | Show |
| Tetraodon nigroviridis | 8 | Show |

EMBL-EBI  [Services](#) [Research](#) [Training](#) [About us](#)

 **HMMER**


[Home](#) [Search](#) [Results](#) [Software](#) [Help](#) [About](#)  

PHMMER Results [Score](#) **[Taxonomy](#)** [Domain](#) [Download](#) [Search Again](#)

Taxonomic distribution of all search hits 

All (376) / Eukaryota (376) / Metazoa (376) /


[back](#) [Echinodermata\(2\)](#) [Echinoidea\(2\)](#) [Echinoidea\(2\)](#) [Strongylocentrotidae\(2\)](#) [1](#)

| Species Distribution | | |
|---|-------|--|
| Species | Count | View |
| Strongylocentrotus purpuratus  | 2 | Show Show Scores For All |

[Search Details](#)

EMBL-EBI

Services | Research | Training | About us



[Home](#) | [Search](#) | [Results](#) | [Software](#) | [Help](#) | [About](#)

[Score](#) | [Taxonomy](#) | [Domain](#) | [Download](#)

[Search Again](#)

PHMMER Results

Sequence Matches and Features

Pfam

103

disorder

103

✓ disorder

✓ coiled-coil

✓ tm & signal peptide

[Show hit details](#)

Your results have been filtered [Cancel](#)

[All Results](#) → [Echinodermata](#)

Query Matches (2) in uniprotrefprot (v.2016-03-02) [Customize](#)

| | Target | Description | Species | E-value |
|---|------------------------------|-------------------------|---|---------|
| > | W4Z1N0_STRPU | Uncharacterized protein | Strongylocentrotus purpuratus | 1.6e-29 |
| > | W4Y2S3_STRPU | Uncharacterized protein | Strongylocentrotus purpuratus | 3.4e-25 |

(show all) alignments

Your search took:0.97 secs
showing rows 1 - 2 of 2

[Search Details](#)

EMBL-EBI

Services Research Training About us

HMMER

Home Search Results Software Help About

Score Taxonomy **Domain** Download

PHMMER Results [Search Again](#)

Sequence Matches and Features

Pfam 103
disorder 103

✓ disorder ✓ coiled-coil ✓ tm & signal peptide

[Show hit details](#)

Your results have been filtered [Cancel](#)

[All Results](#) → [Echinodermata](#)

Query Matches (2) in uniprotrefprot (v.2016-03-02) [Customize](#)

| | Target | Description | Species | E-value |
|---|------------------------------|-------------------------|---|---------|
| > | W4Z1N0_STRPU | Uncharacterized protein | Strongylocentrotus purpuratus | 1.6e-29 |
| > | W4Y2S3_STRPU | Uncharacterized protein | Strongylocentrotus purpuratus | 3.4e-25 |

(show all) alignments

Your search took: 0.97 secs
showing rows 1 - 2 of 2

[Search Details](#)

EMBL-EBI

Services
Research
Training
About us

Home
Search
Results
Software
Help
About

Score
Taxonomy
Domain
Download

Search Again

PHMMER Results

Your results have been filtered
Cancel

All Results → Echinodermata

Sequence Matches and Features

Pfam 103

disorder 103

✓ disorder

✓ coiled-coil

✓ tm & signal peptide

Show hit details

There were no exact architecture matches.

Domain Architectures

1 SEQUENCE

with domain architecture: **I-set, I-set, I-set, fn3, I-set, I-set**, example:W4Y2S3_STRPU

View Scores

Sequence Features

I-set I-set I-set fn3 I-set I-set

1 SEQUENCE

with domain architecture: **I-set, I-set, I-set, I-set, I-set, I-set**, example:W4Z1N0_STRPU

View Scores

Sequence Features

I-set I-set I-set I-set I-set I-set

EMBL-EBI

Services Research Training About us

HMMER

Home Search Results Software Help About

PHMMER Results

Score Taxonomy **Domain** Download

Search Again

Your results have been filtered [Cancel](#)

All Results → Echinodermata

Sequence Matches and Features

Pfam 103

disorder 103

✓ disorder ✓ coiled-coil ✓ tm & signal peptide

[Show hit details](#)

There were no exact architecture matches.

Domain Architectures

1 SEQUENCE with domain architecture: **I-set, I-set, I-set, fn3, I-set, I-set**, example: [W4Y2S3_STRPU](#) [View Scores](#)

Sequence Features: I-set I-set I-set fn3 I-set I-set

2672

1 SEQUENCE with domain architecture: **I-set, I-set, I-set, I-set, I-set, I-set**, example: [W4Z1N0_STRPU](#) [View Scores](#)

Sequence Features: I-set I-set I-set I-set I-set I-set

3373



Home Search Results Software Help About

Score Taxonomy Domain **Download**

PHMMER Results

- **Job:** 11F56758-277F-11E6-9D89-A9DDD26C98AD.1
- **Started:** 2016-05-31 23:29:00
- **Algorithm:** phmmer
- **HMMER Options:** -E 1 --domE 1 --incE 0.01 --incdomE 0.03 --mx BLOSUM62 --pextend 0.4 --popen 0.02 --seqdb uniprotrefprot

▼ Format

Text

A plain text file containing the hit alignments and scores.



Tab Delimited

A tab delimited text file containing the hit information. No alignments.



XML

An XML file formatted for machine parsing of the data.



JSON

All the results information encoded as a single json string.



FASTA

Download the significant hits from your search as a gzipped FASTA file.



Full length FASTA

A gzipped file containing the full length sequences for significant search hits.



Aligned FASTA

A gzipped file containing aligned significant search hits in FASTA format.



STOCKHOLM

Download an alignment of significant hits as a gzipped STOCKHOLM file.




Annotation team: annotate!

[phmmmer](#) [hmmscan](#) [hmmsearch](#) [jackhmmer](#)

protein sequence vs protein sequence database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Paste in your sequence or use the [example](#) 

HEAIGSGDLRLRSAPFRRTSLAGAGRRRTSDSHEDAGTLDFFSSLKKRDSFRRDSKLEAPAEEDVWEILROAPPSEYERIAFOHGVTDLRGMLKRLKGMKODEKK

Submit

[Reset](#)

▼ [Sequence Database](#)


Frequently used databases

☐ Reference Proteomes ☐ UniProtKB ☒ SwissProt ☐ PDB

Representative Sets (UniProt)

☐ rp75 ☐ rp55 ☐ rp35 ☐ rp15

Other databases

☐ Ensembl Genomes  ☐ QfO


► [Restrict by Taxonomy](#) 



[phmmer](#) [hmmscan](#) [hmmsearch](#) [jackhmmer](#)

protein sequence vs protein sequence database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Paste in your sequence or use the [example](#) 

HEAIGSGDLDLRSAFRRTSLAGAGRRTSDSHEDAGTLDFFSLLKKRDSFRRDSKLEAPAEEDVWEILROAPPSEYERIAFOHGVTDLRGMLKRLKGMKODEKK

Submit

Reset

▼ Sequence Database


Frequently used databases

☐ Reference Proteomes ☐ UniProtKB ☐ SwissProt ☒ PDB

Representative Sets (UniProt)

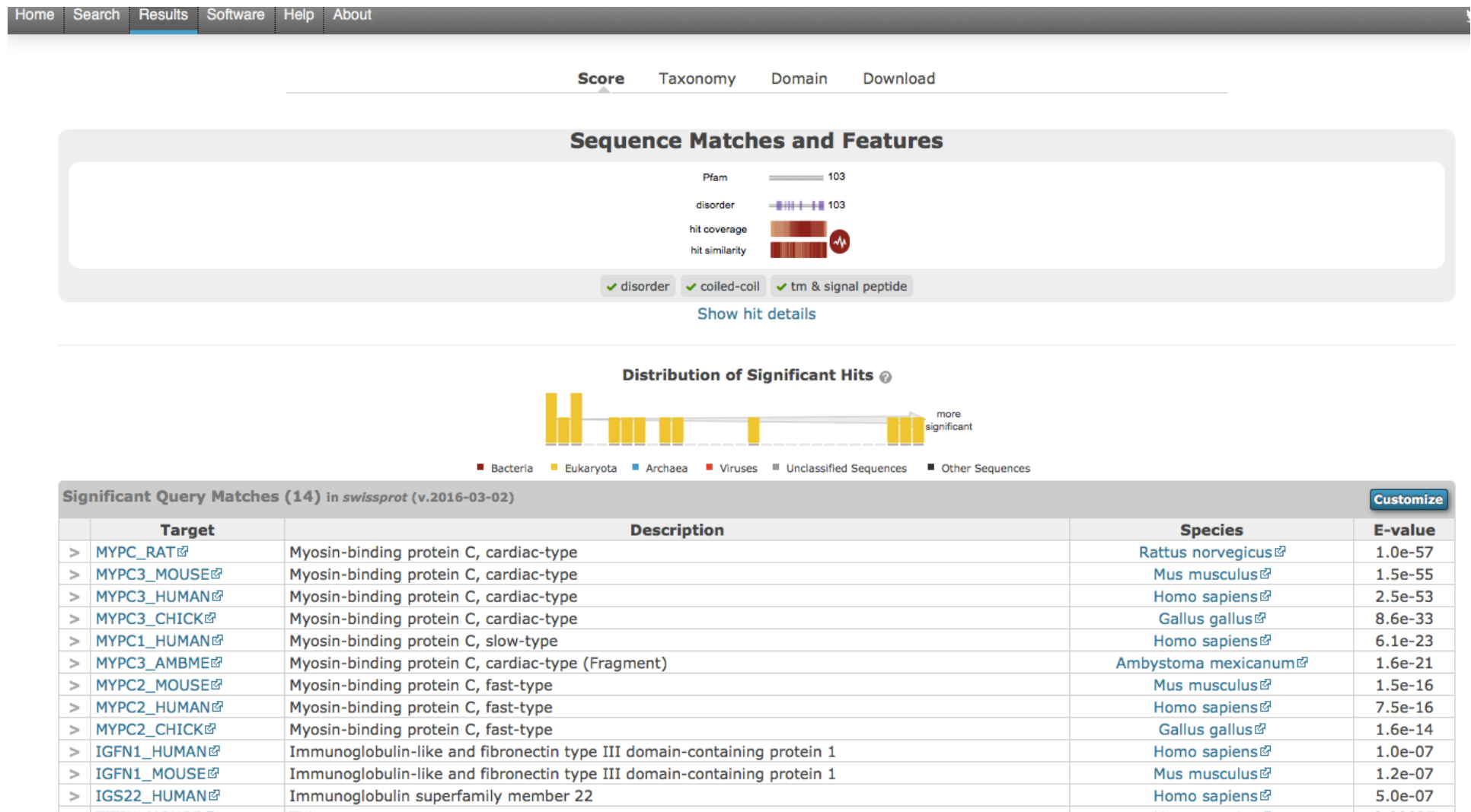
☐ rp75 ☐ rp55 ☐ rp35 ☐ rp15


Other databases


☐ Ensembl Genomes  ☐ QfO

► [Restrict by Taxonomy](#) 








UniProtKB ▾
Advanced ▾
 Search

BLAST
Align
Retrieve/ID mapping
Help
Contact






UniProtKB - P56741 (MYPC_RAT)
 Basket ▾




Display

Entry
Feature viewer
Feature table

None

☒ Function
☒ Names & Taxonomy
☒ Subcell. location
☐ Pathol./Biotech
☒ PTM / Processing
☒ Expression
☒ Interaction
☒ Structure


 BLAST
 Align
 Format
 Add to basket
 History

 Feedback
 Help video
 Other tutorials and video

Protein | **Myosin-binding protein C, cardiac-type**

Gene | **Mybpc3**

Organism | *Rattus norvegicus (Rat)*


Status |  Reviewed - Annotation score: ●●●●○ - Experimental evidence at protein levelⁱ


Functionⁱ

Thick filament-associated protein located in the crossbridge region of vertebrate striated muscle a bands. In vitro it binds MHC, F-actin and native thin filaments, and modifies the activity of actin-activated myosin ATPase. It may modulate muscle contraction or may play a more structural role.


Sites

| Feature key | Position(s) | Length | Description | Graphical view | Feature identifier | Actions |
|-------------|-------------|--------|-------------|----------------|--------------------|---------|
| | | | | | | |



UniProtKB ▾
Advanced ▾
 Search

BLAST
Align
Retrieve/ID mapping
Help
Contact






UniProtKB - P56741 (MYPC_RAT)
 Basket




Display

Entry
Feature viewer
Feature table

None

☒ Function
☒ Names & Taxonomy
☒ Subcell. location
☐ Pathol./Biotech
☒ PTM / Processing
☒ Expression
☒ Interaction
☒ Structure


 BLAST
 Align
 Format
 Add to basket
 History

 Feedback
 Help video
 Other tutorials and video

Protein | **Myosin-binding protein C, cardiac-type**

Gene | **Mybpc3**

Organism | *Rattus norvegicus (Rat)*


Status |  **Reviewed** - Annotation score: ●●●●○ - Experimental evidence at protein levelⁱ


Function

Thick filament-associated protein located in the crossbridge region of vertebrate striated muscle a bands. In vitro it binds MHC, F-actin and native thin filaments, and modifies the activity of actin-activated myosin ATPase. It may modulate muscle contraction or may play a more structural role.


Sites

| Feature key | Position(s) | Length | Description | Graphical view | Feature identifier | Actions |
|-------------|-------------|--------|-------------|----------------|--------------------|---------|
| | | | | | | |



UniProtKB ▾
Advanced ▾
 Search

BLAST Align Retrieve/ID mapping
Help Contact

UniProtKB - P56741 (MYPC_RAT)
 Basket

Display

Entry
Feature viewer
Feature table

None

☒ Function
☒ Names & Taxonomy
☒ Subcell. location
☐ Pathol./Biotech
☒ PTM / Processing
☒ Expression
☒ Interaction
☒ Structure



BLAST
Align
Format
Add to basket
History

Feedback
Help video
Other tutorials and video

Protein | **Myosin-binding protein C, cardiac-type**

Gene | **Mybpc3**

Organism | *Rattus norvegicus (Rat)*

Status |  Reviewed - Annotation score:  - Experimental evidence at protein levelⁱ

Functionⁱ

Thick filament-associated protein located in the crossbridge region of vertebrate striated muscle a bands. In vitro it binds MHC, F-actin and native thin filaments, and modifies the activity of actin-activated myosin ATPase. It may modulate muscle contraction or may play a more structural role.

Sites

| Feature key | Position(s) | Length | Description | Graphical view | Feature identifier | Actions |
|-------------|-------------|--------|-------------|----------------|--------------------|---------|
| | | | | | | |

Display

Entry

Feature viewer

Feature table

None

☒ Function

☒ Names & Taxonomy

☒ Subcell. location

☐ Pathol./Biotech

☒ PTM / Processing

☒ Expression

☒ Interaction

☒ Structure

☒ Family & Domains

☒ Sequence

☒ Cross-references

☒ Publications

☒ Entry information

| | | | | |
|----------------------------|-----------|--------|---------------|--|
| Metal binding ⁱ | 212 – 212 | 1 Zinc | By similarity | |
| Metal binding ⁱ | 214 – 214 | 1 Zinc | By similarity | |
| Metal binding ⁱ | 227 – 227 | 1 Zinc | By similarity | |
| Metal binding ⁱ | 229 – 229 | 1 Zinc | By similarity | |

GO - Molecular functionⁱ

- metal ion binding Source: UniProtKB-KW
- structural constituent of muscle Source: Ensembl

GO - Biological processⁱ

- cardiac muscle contraction Source: BHF-UCL
- cell adhesion Source: UniProtKB-KW
- regulation of striated muscle contraction Source: BHF-UCL
- ventricular cardiac muscle tissue morphogenesis Source: Ensembl

Complete GO annotation...

Keywords - Molecular functionⁱ

Muscle protein

Keywords - Biological processⁱ

Cell adhesion

Keywords - Ligandⁱ

Actin-binding, Metal-binding, Zinc

Display

Entry

Feature viewer

Feature table

None

☒ Function

☒ Names & Taxonomy

☒ Subcell. location

☐ Pathol./Biotech

☒ PTM / Processing

☒ Expression

☒ Interaction

☒ Structure

☒ Family & Domains

☒ Sequence

☒ Cross-references

☒ Publications

☒ Entry information

| | | | | |
|----------------------------|-----------|--------|---------------|--|
| Metal binding ⁱ | 212 – 212 | 1 Zinc | By similarity | |
| Metal binding ⁱ | 214 – 214 | 1 Zinc | By similarity | |
| Metal binding ⁱ | 227 – 227 | 1 Zinc | By similarity | |
| Metal binding ⁱ | 229 – 229 | 1 Zinc | By similarity | |

GO - Molecular functionⁱ

- metal ion binding Source: UniProtKB-KW
- structural constituent of muscle Source: Ensembl

GO - Biological processⁱ

- cardiac muscle contraction Source: BHF-UCL
- cell adhesion Source: UniProt
- regulation of striated muscle contraction
- ventricular cardiac muscle contraction

Inferred from expression patternⁱ

PubMed 8799143

Complete GO annotation...

Keywords - Molecular functionⁱ

Muscle protein

Keywords - Biological processⁱ

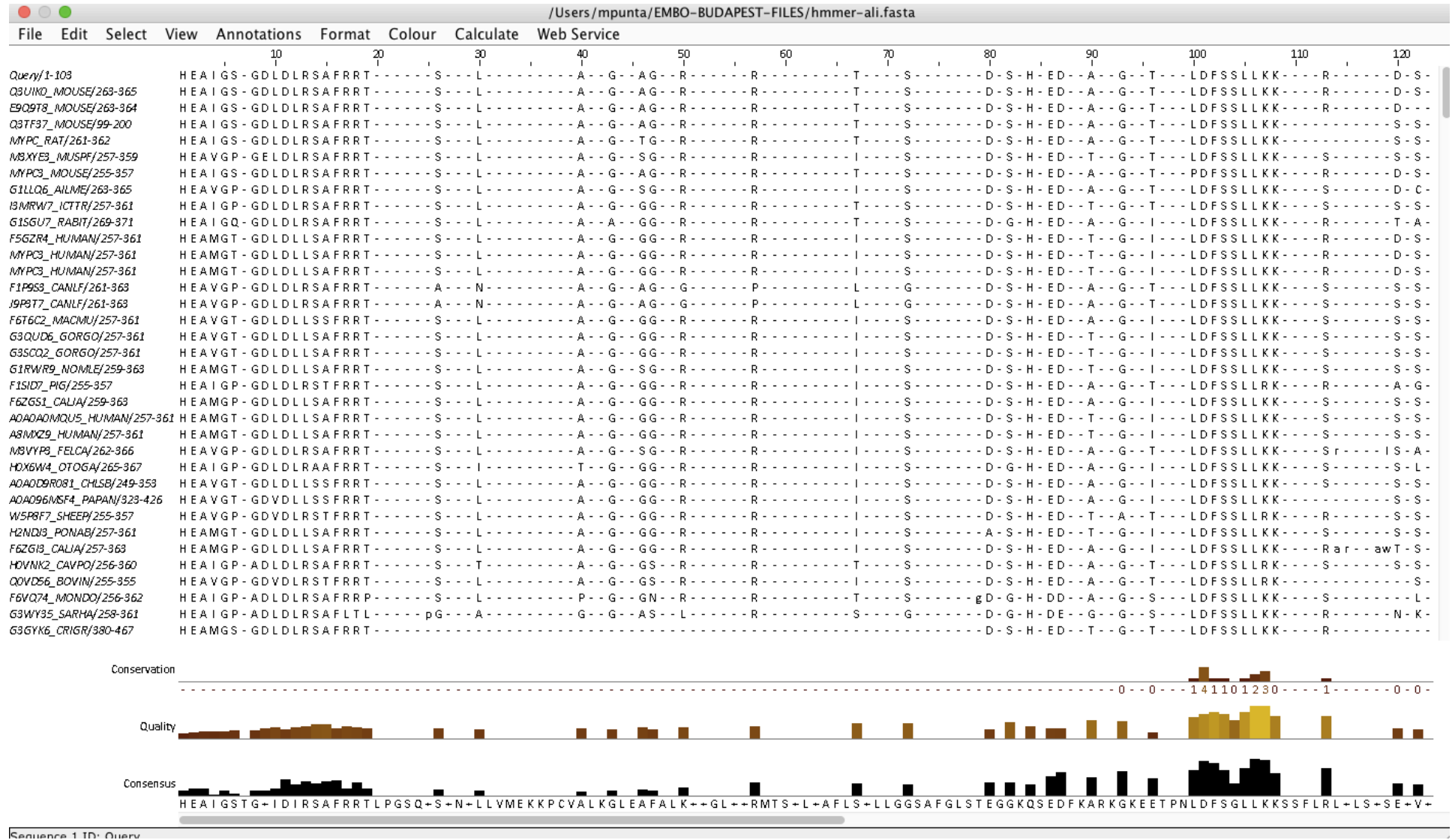
Cell adhesion

Keywords - Ligandⁱ

Actin binding Metal binding Zinc

Start Jalview

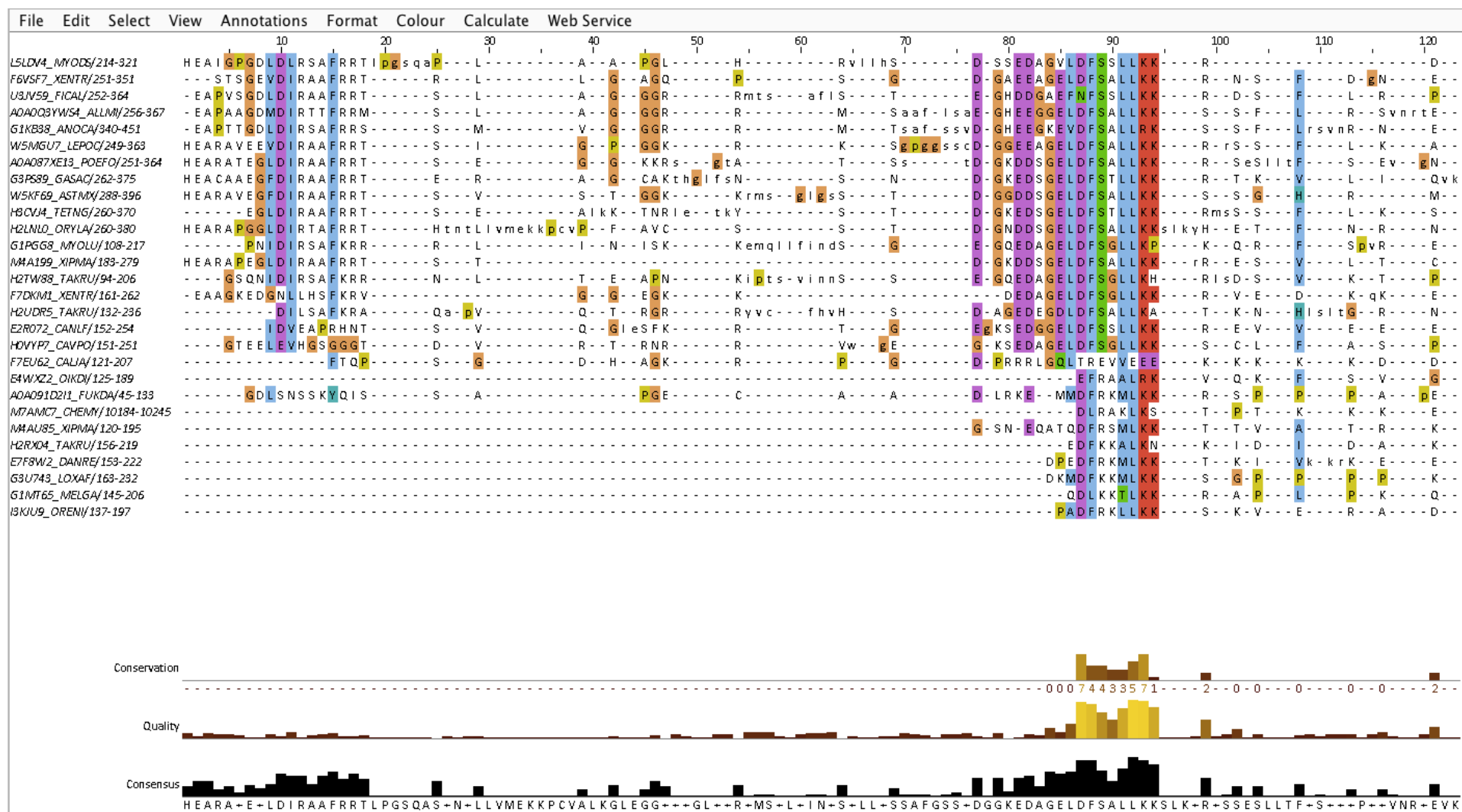
1. File -> Input alignment -> From file



1. Edit -> Remove redundancy
2. Select 90% and Remove

1. Edit -> Remove redundancy
2. Select 90% and Remove
3. Edit -> Remove empty columns

1. Edit -> Remove redundancy
2. Select 90% and Remove
3. Edit -> Remove empty columns
4. Colour -> Clustalx



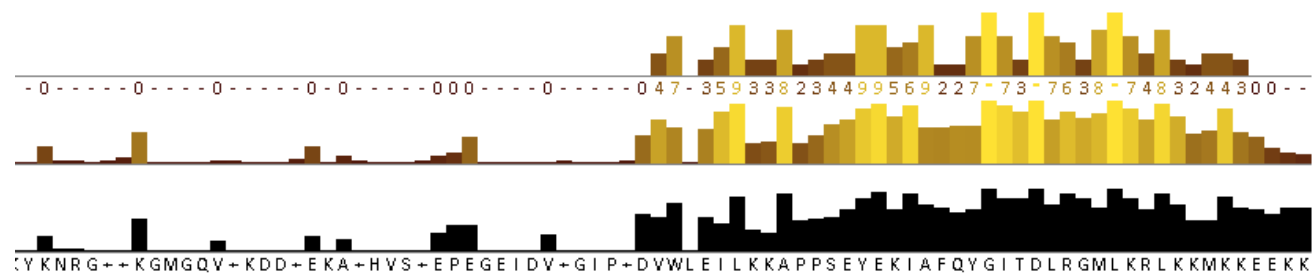
/mpunta/EMBO-BUDAPEST-FILES/hmmer-ali.fasta

vice

```

      130      140      150      160      170      180      190      200
- S - - - K - - - L - - - E - A - - - P A E - - - E - - - D V W - E I L R Q A P P S E Y E R I A F Q H G V T D L R G M L K R L K G M K H D E K K
- P - - - K - - - Q - - - G - A - - - E P D - - - V - - - D V W - D I L K K A P P S E Y E K I A F Q Y G I T D L R G L L K R L K K M K K E E K K
- S n r g e g K - - - S - - - E - S - - - Q S D - - - V - - - D V W - D I L R K A P P T E Y E K I A F Q Y G I T D L R G M L K R L K R I K K E E K K
- A - - - K - - - T - - - E - S - - - Q P S - - - V - - - D V W - E I L R K A P P S E Y E K I A F Q Y G I T D L R G L L K R L K R M K K E E K K
- T - - - K - - - T - - - E - T - - - R P E - - - V - - - D V W - E I L R N A P P S E Y E K I A F Q Y G I T D L R G M L K R L K R I K K E E K K
- S n r - a v Q - - - V - - - N - A - - - Q P E - - - I - - - D V W - D I L R K A P P T E Y E K I A F Q Y G I T D L R G M L K R L K K M K K E E K K
- H - - - K - - - S - - - N - S - - - H P D - - - V - - - D V W - N I L S H A P P S E Y E K I A F Q H G I T D L R G M L K R L K K M K K E E K K
: y S - - - H - - - M r - - - n K - A - - - E P D - - - M - - - D V W - N I L S H A P P S E Y E K I A F Q H G I T D L R G M L K R L K K M K K E E K K
- V - - - Q - - - V - - - S - S - - - E P D - - - V - - - D V W - E I L S H A P A S E Y E K I A F Q Y G I T D L R G M L K R L K K M K K E E K K
- S - - - s R - - - V - - - R - A v h v s s E P E - - - V - - - D V W - E I L S K A P S S E Y E K I A F Q H G I T D L R G M L K R L K K M K K E E K K
- R - - - A - - - V - - - Q - V - - - H S E - - - P d - - - v D V W - S I L S K A P P S A F E K I A F Q Y G I T D L R G M L K R L K K M K K D E K K
- V - - - K - - - Q - - - Q - Q d - - - e E P E - - - I - - - D V W - E L L K N A N P N E Y E K I A F Q Y G I T D L R G M L K R L K R M R R V E K K
- S - - - Q - - - K - - - K - T - - - T A E - - - K - - - K K L I L K M G H A P P S E F E K I A F Q Y G I T D L R G M L K R L K K M K K E E K K
- N - - - K g m g q F - - - D - D - - - T P E - - - V - - - D V W - E I L K K A R P D E Y E K I A F T Y G I T D L R G L L R R M K K I P K E E K K
- Q - - - K - - - K k k d d d D - C - - - G I P - - - P - - - D V W - E L L K N A K P S D F E R I A F E H G I T D L R G M L K R L K K V K K E V K K
- K - - - K - - - P - - - Q k E - - - E P E - - - I - - - D V W - E L L K S A H P S E Y E K I A F Q Y G I T D L R G M L K R L K K M K V V E - -
- K - - - K - - - K - - - K - K - - - D D D - - - D I g i p p E I W - E L L K G A K K S E Y E K I A F Q Y G I T D L R G M L K R L K K A K V E V K K
- Y - - - D - - - D - - - D - L - - - G I P - - - P - - - E I W - E L L K G A K K S E Y E K I A F Q Y G I T D L R G M L K R L K K A K V E V K K
- D - - - D - - - L - - - G - I - - - P P E - - - I W - E L L K G A K K S E Y E K I A F Q Y G I T D L R G M L K R L K K A K V E V K K
- K - - - K - - - V - A - - - D A K - - - E - - - D I W - A L L K S A N P R E Y D R I A F Y W G I K D L R K L L K K L A N A K K N N K K
- K - - - K - - - I - - - E - - - S - - - E - - - Q V W - Q L L M N A D R K D Y E K I C L K Y G I V D F R G M L R K L Q E M K K E Q - -
- E - - - E - - - E - - - E - K - - - P I - - - D I M - E L L K N V D P K E Y E K Y A R M Y G I T D F R G L L Q A F E L L K Q S Q - -
- K - - - Q - - - L - - - P - Q - - - Q K E g e i d P - - - K L L - E L L S A P K K D Y E R I C L E F G I T D F R W F L K K L K Q I K K E R - -
- E - - - E - - - N - - - K - M - - - E A D - - - E - - - K F F - E V L M S A E K K D Y E S I C I Q Y G V T D F R G M L K K L N E K K I E - -
- P - - - K - - - K - - - E - G - - - E I D - - - P - - - K F W - D V M L N A K K S D Y E R I C H E F G I T D Y R W M L K Q L N L K K K - E K -
- K - - - K - - - Q - - - k K - V - - - I D E - - - K - - - E M L - E I L S K V P K K D F E R V C M V Y G F T D F W G L L K K L K E M K K Y E E K
- K - - - E - - - V - - - D - - - K - - - E - - - A I F - Q L L L H A D K K D Y E R I C I K Y G I S D F R G M L R A L Q D L R K D T - -
- G - - - - - - - - - - - E S D - - - E - - - R F W - D V M L K A D R N D Y E R I C S E F G V K D L H S I L K K L D E K K E - -

```



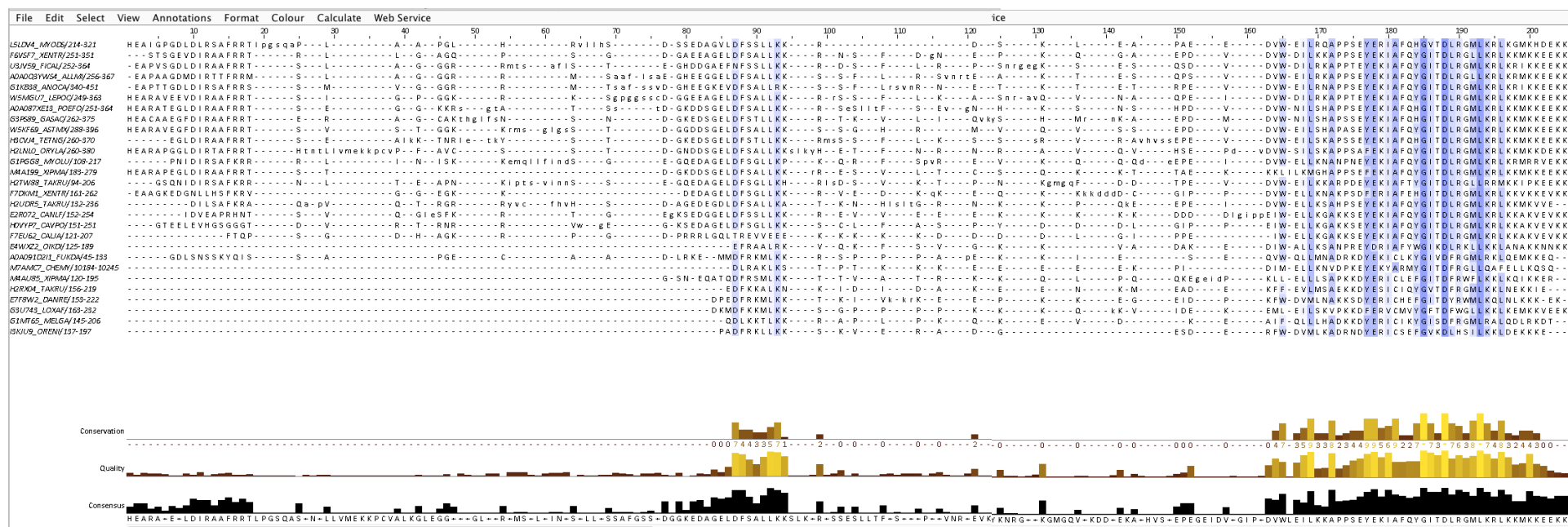
1.

| |
|--------------------------|
| Colour -> BLOSUM62 Score |
|--------------------------|

 /

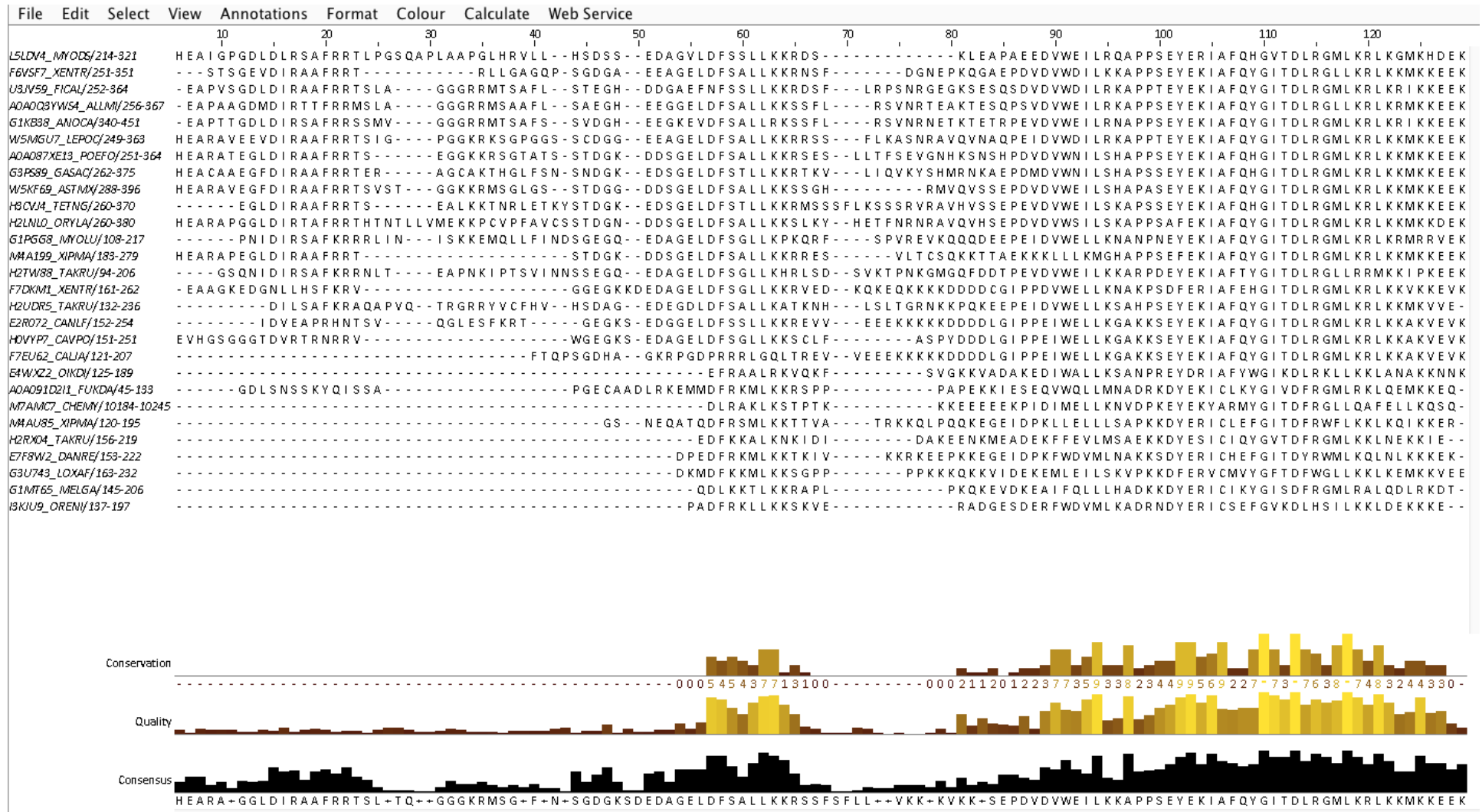
| |
|-------------------------------|
| Colour -> Percentage Identity |
|-------------------------------|

| 1. | Colour -> BLOSUM62 Score | / | Colour -> Percentage Identity |
|----|--------------------------|---|-------------------------------|
|----|--------------------------|---|-------------------------------|

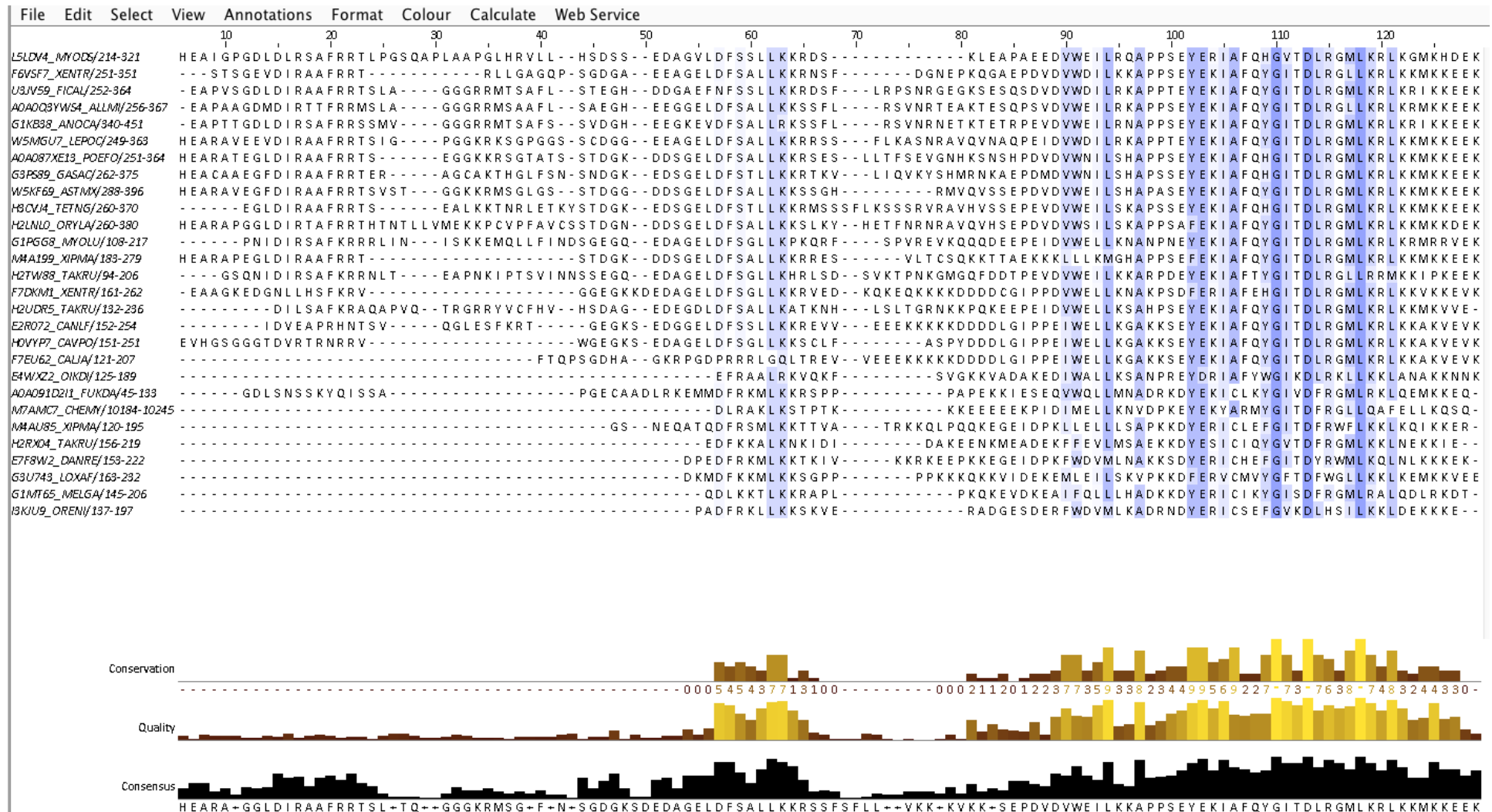


1. Web Service -> Alignment-> Run Muscle with Default

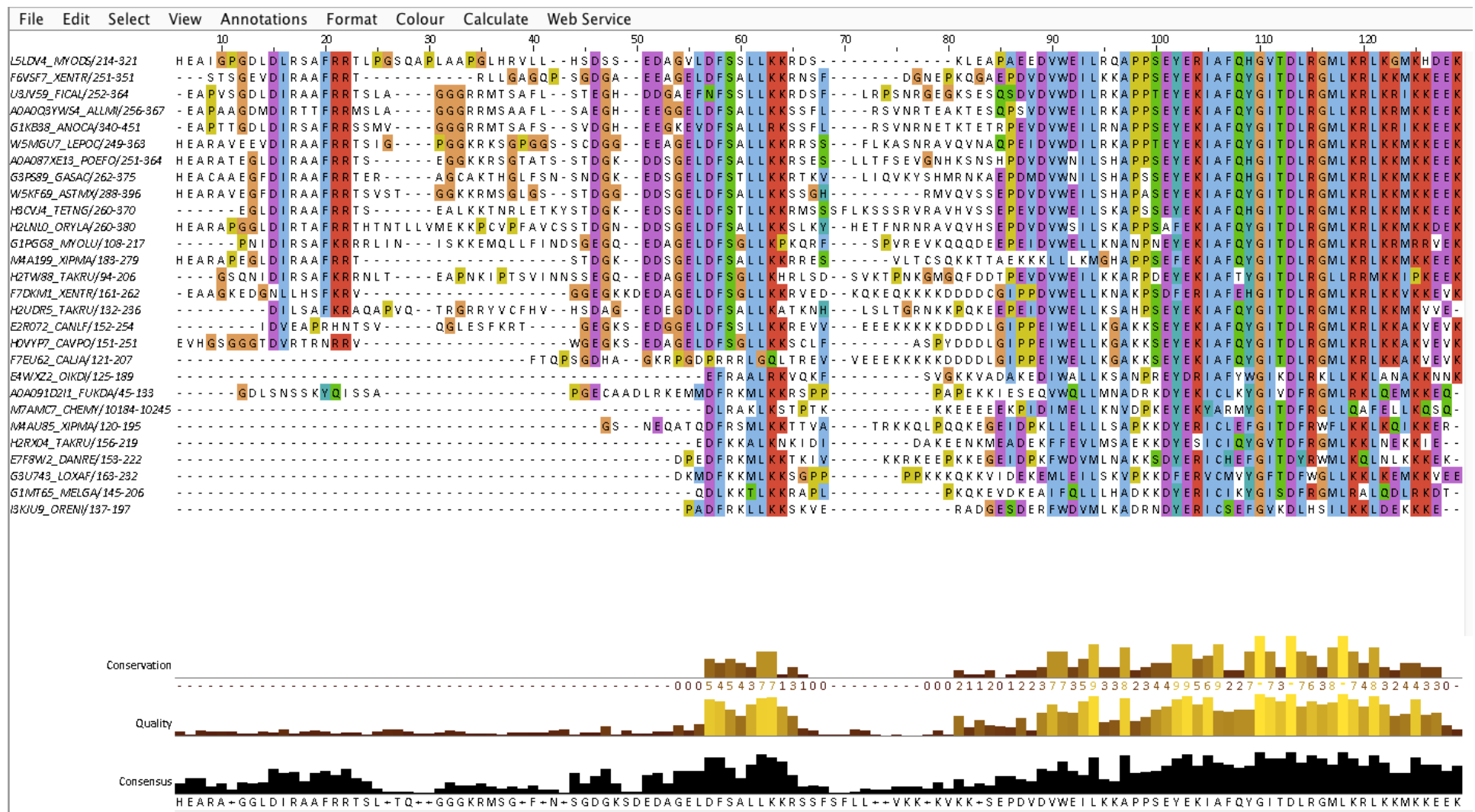
1. Web Service -> Alignment -> Run Muscle with Default



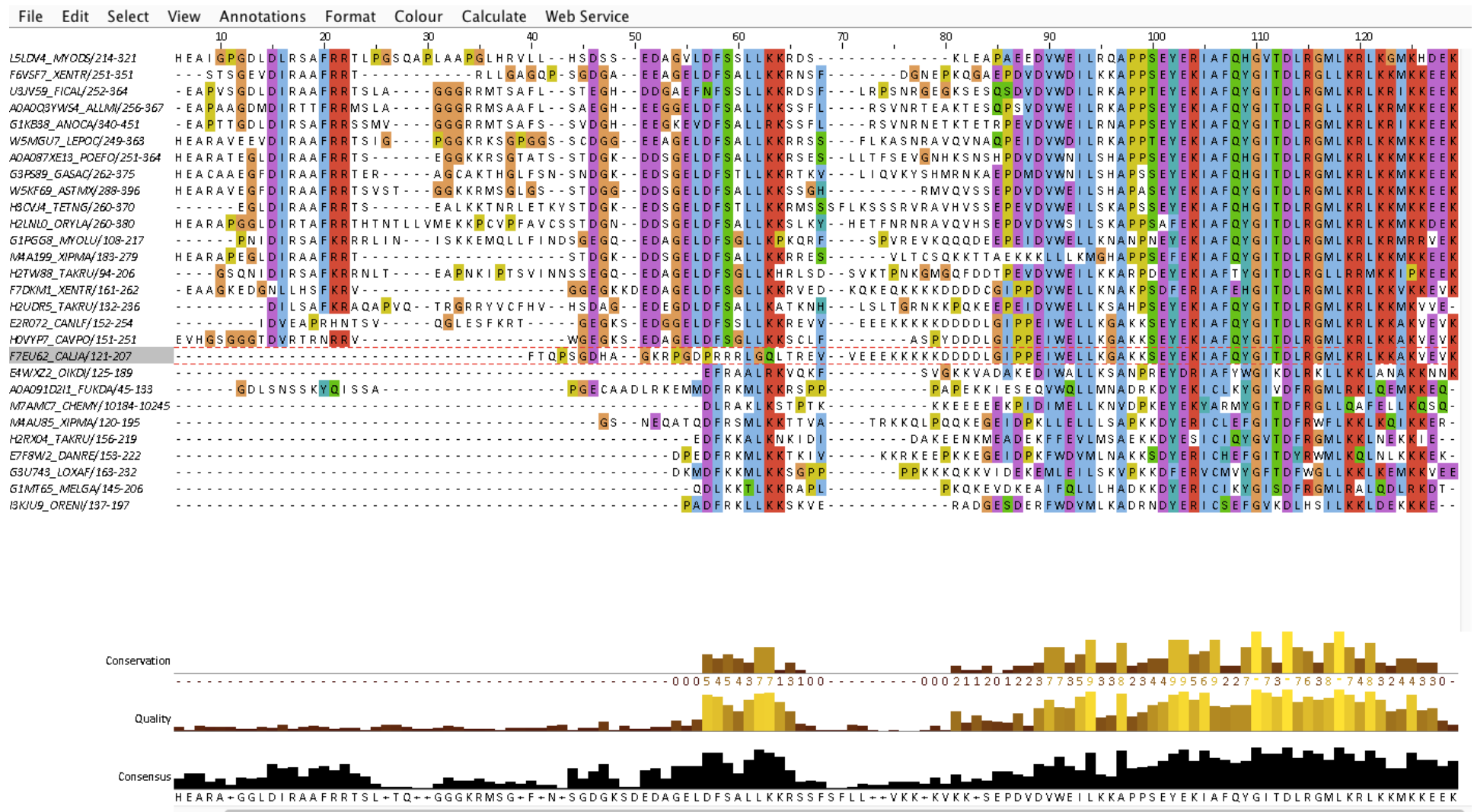
1. Colour -> BLOSUM62 Score



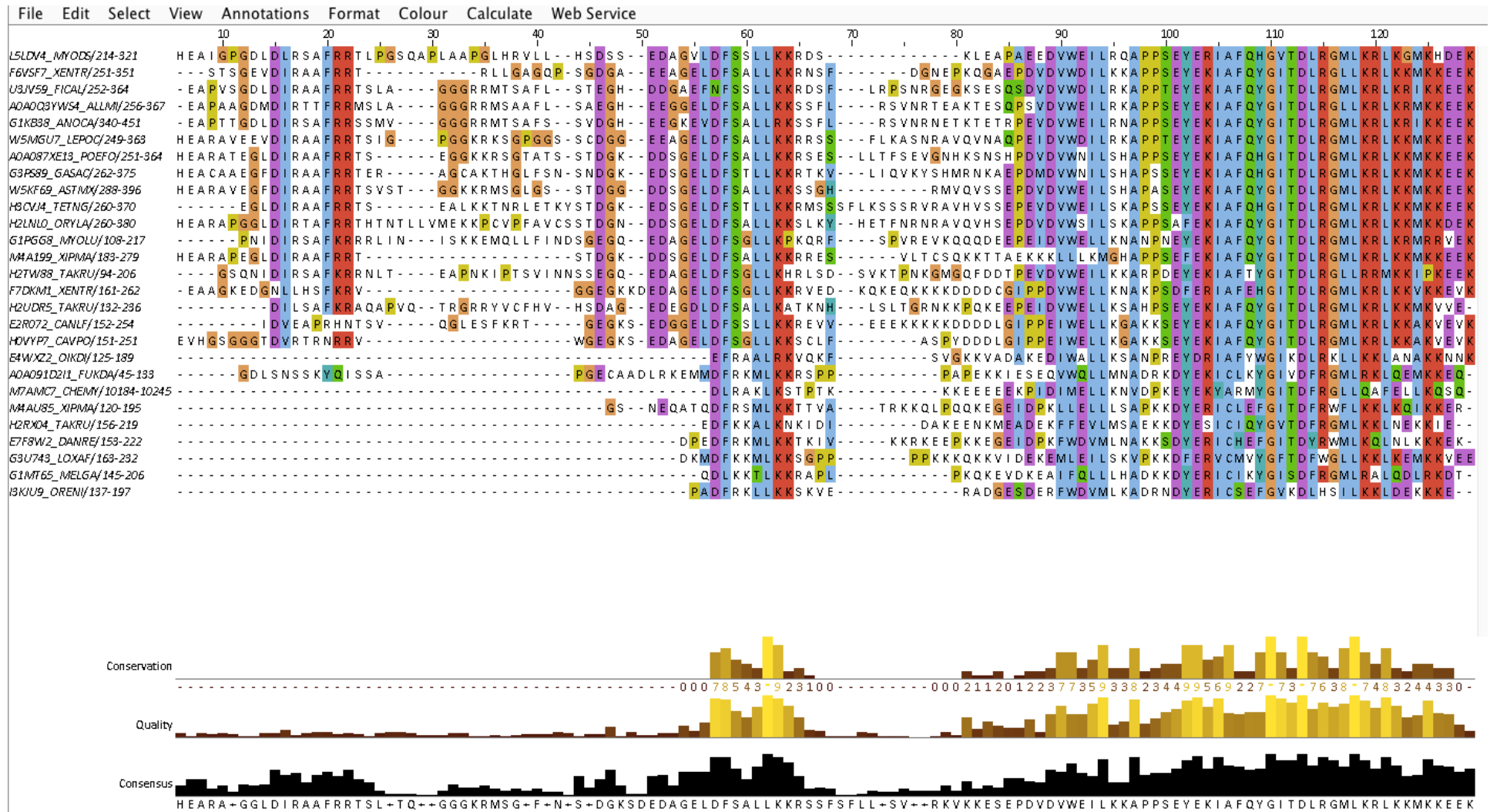
1. Colour -> Clustalx



1. Click on F7EU62_CALJA



1. Edit -> Delete

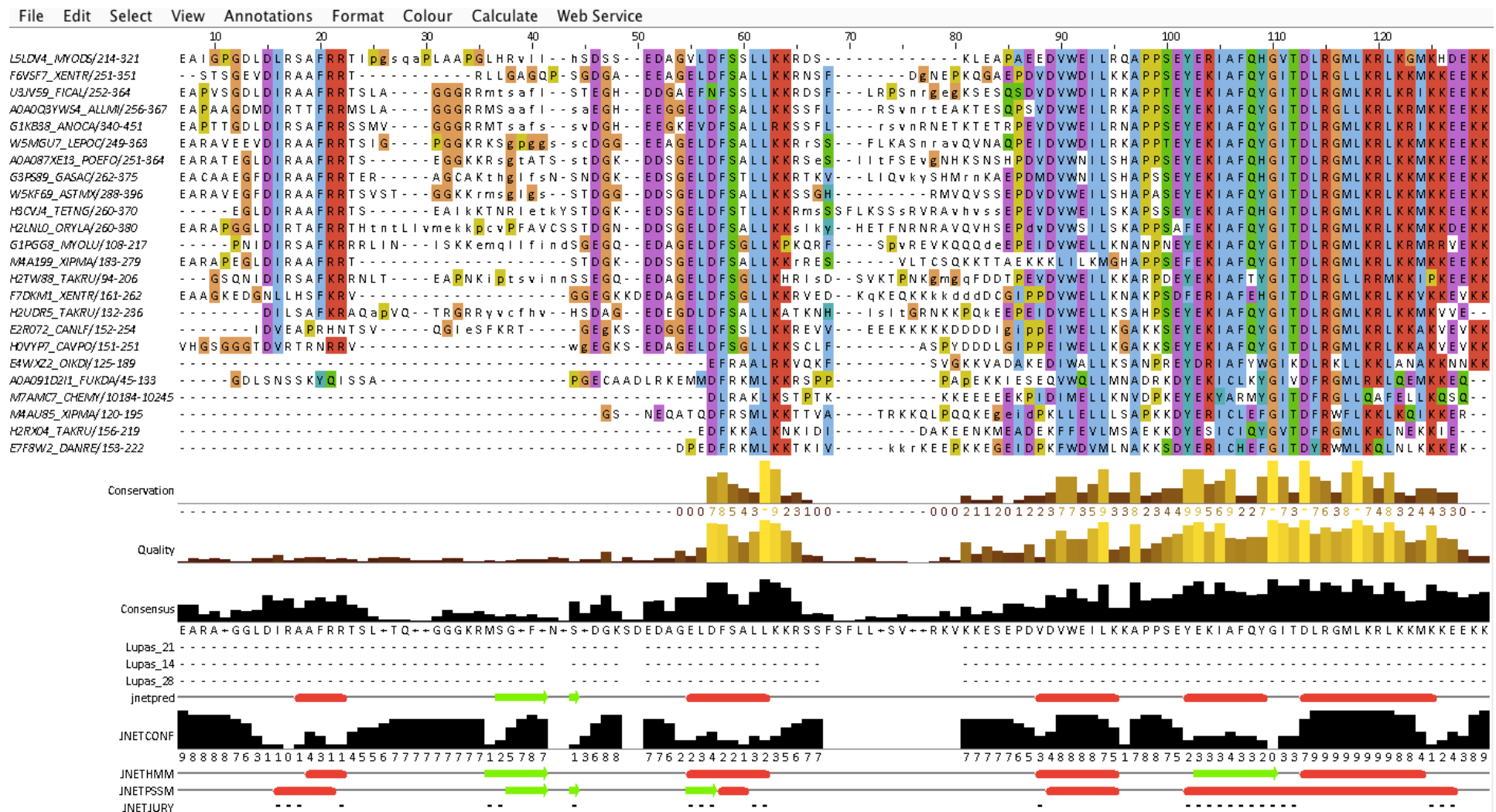


1. Select -> Select all

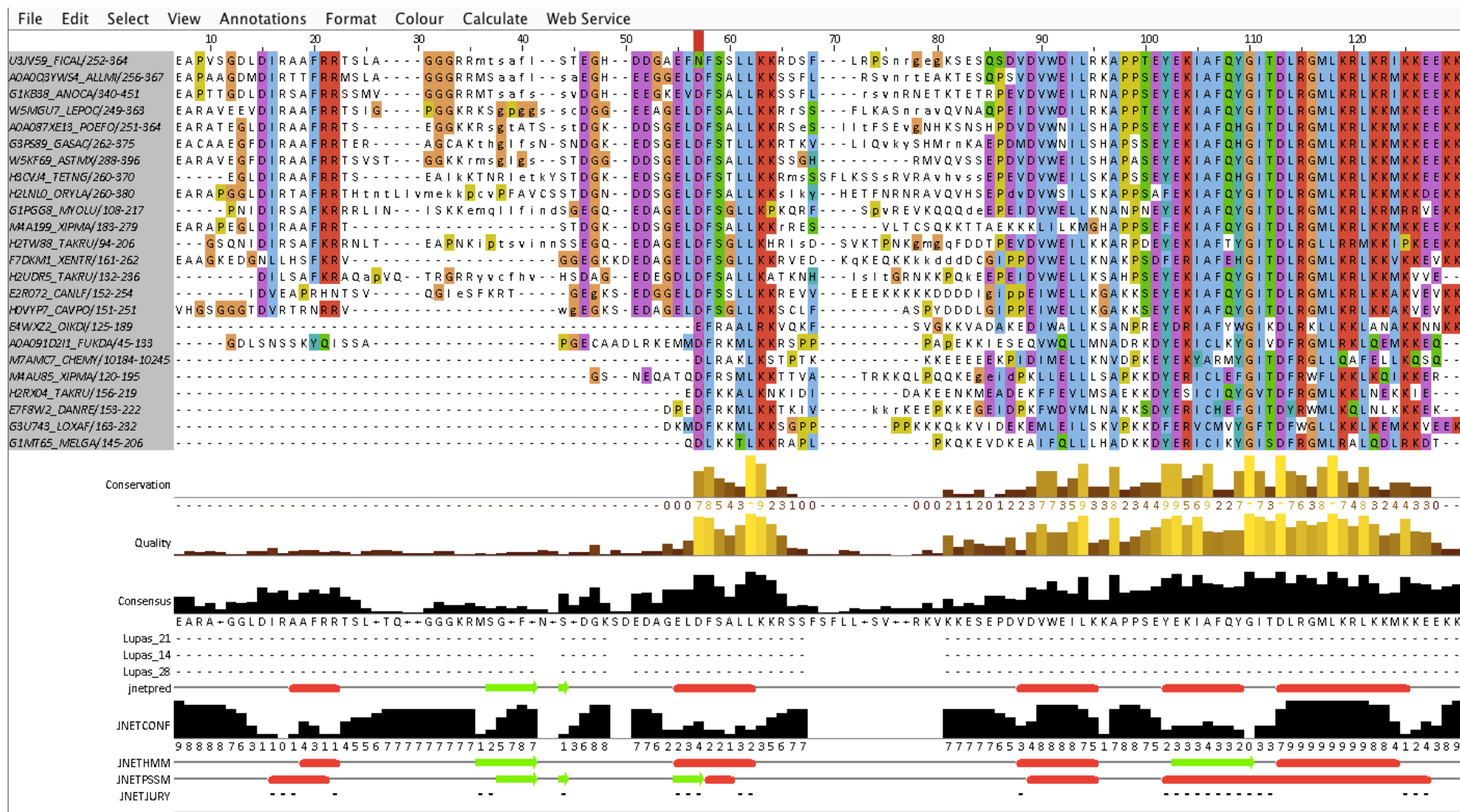
2. Web Service -> Secondary Structure Prediction-> JNet Secondary Structure Prediction

1. Select -> Select all

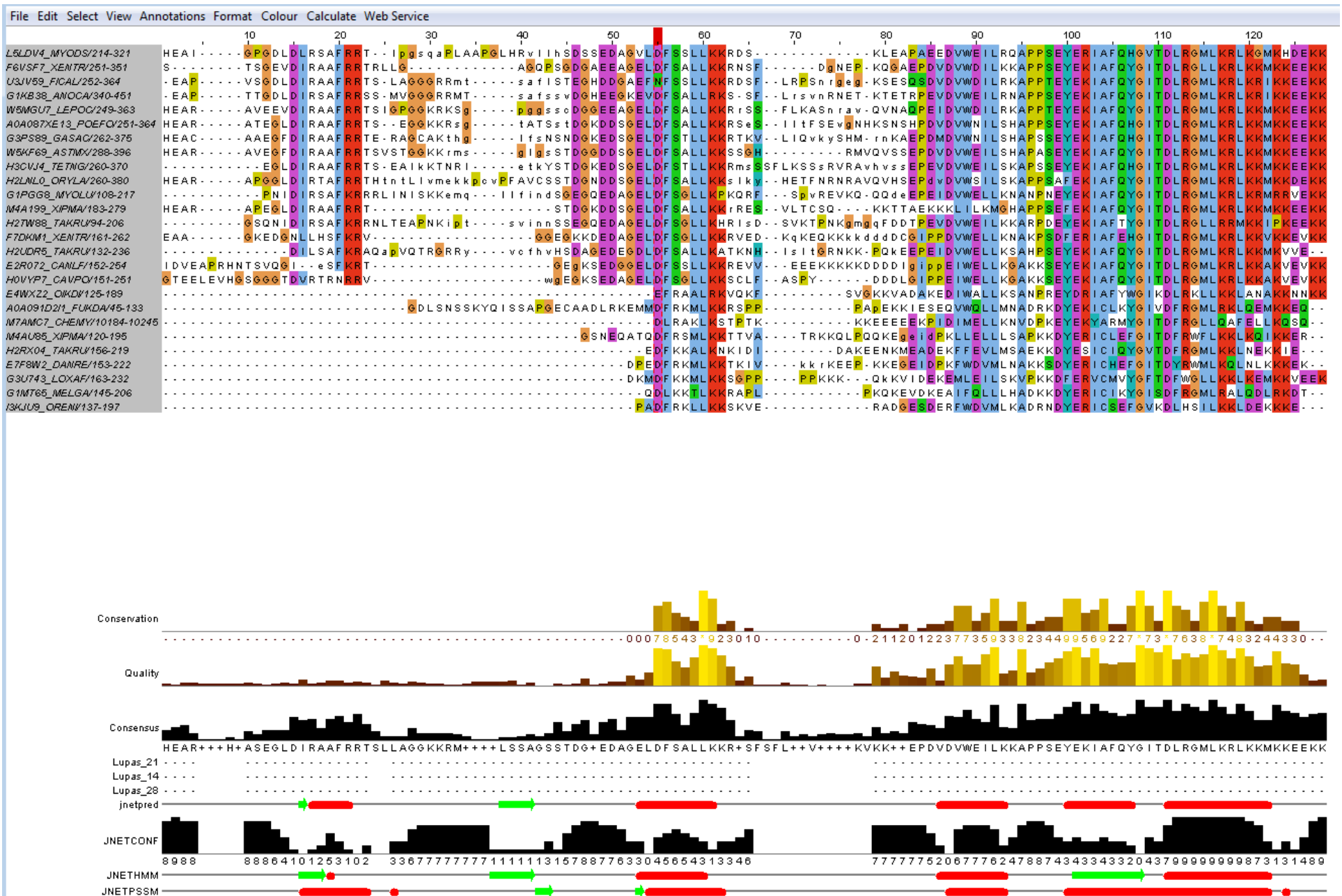
2. Web Service -> Secondary Structure Prediction-> JNet Secondary Structure Prediction



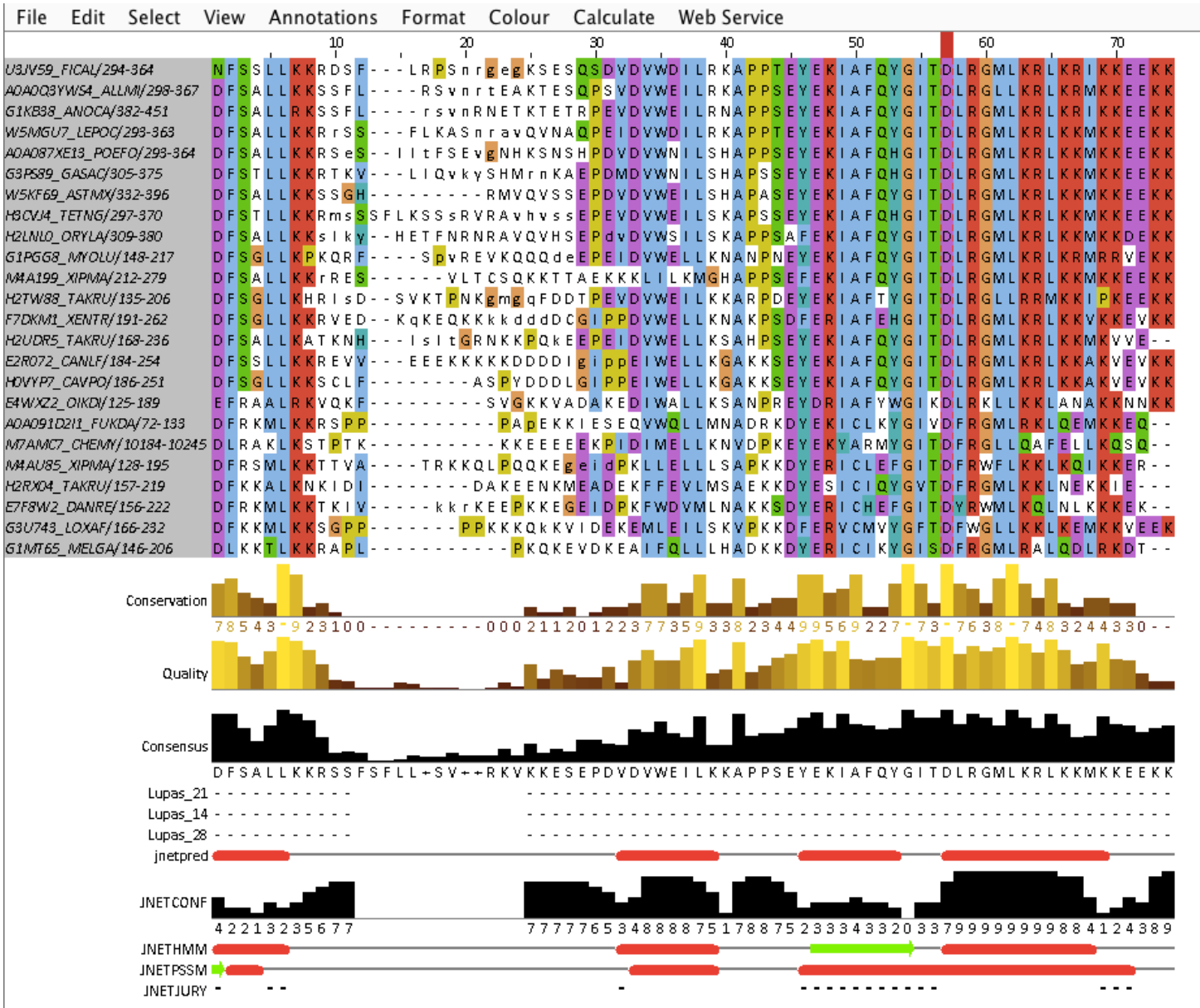
Click on top of column corresponding to alignment position 57



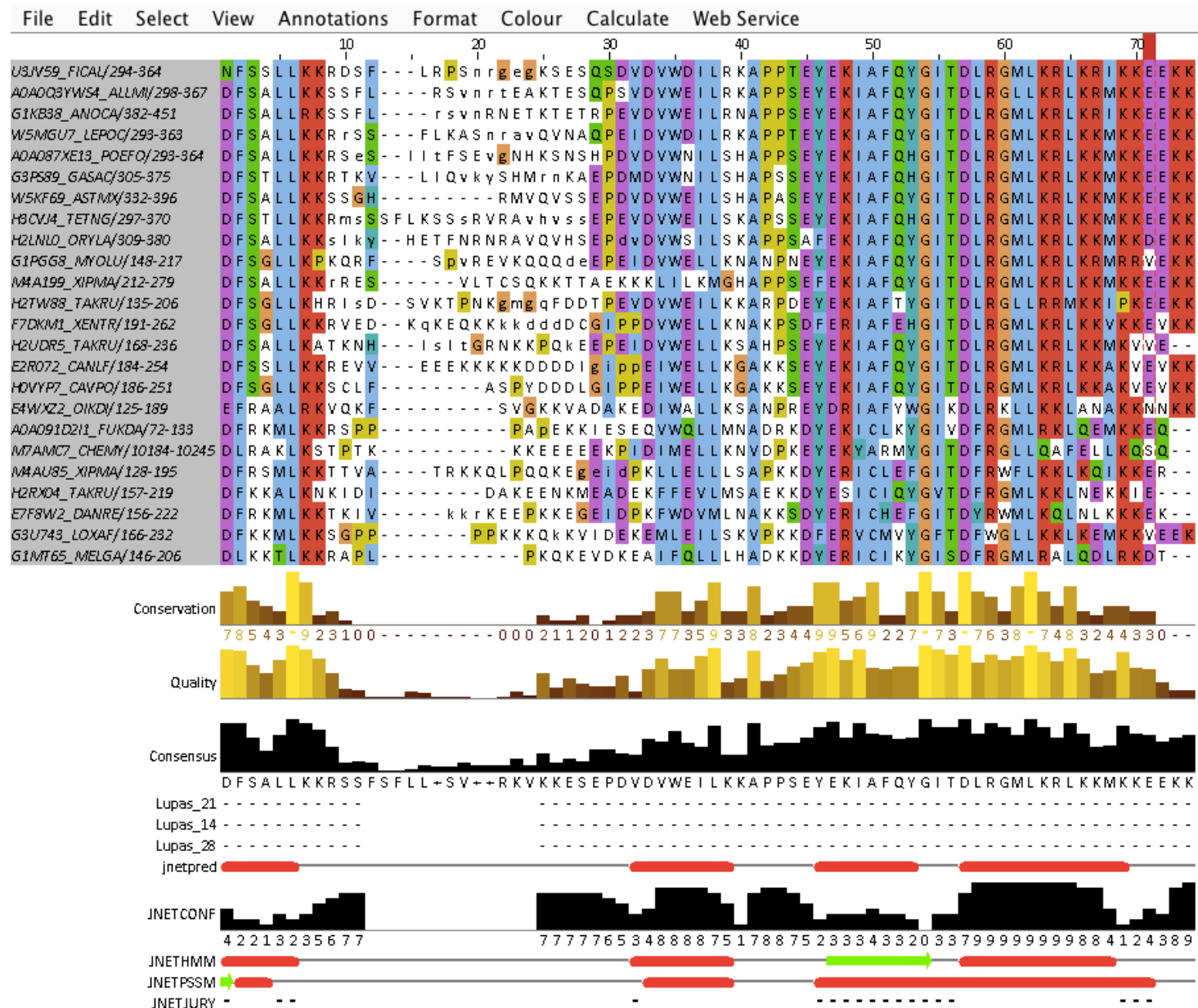
1. Edit -> Remove left



Edit -> Remove left

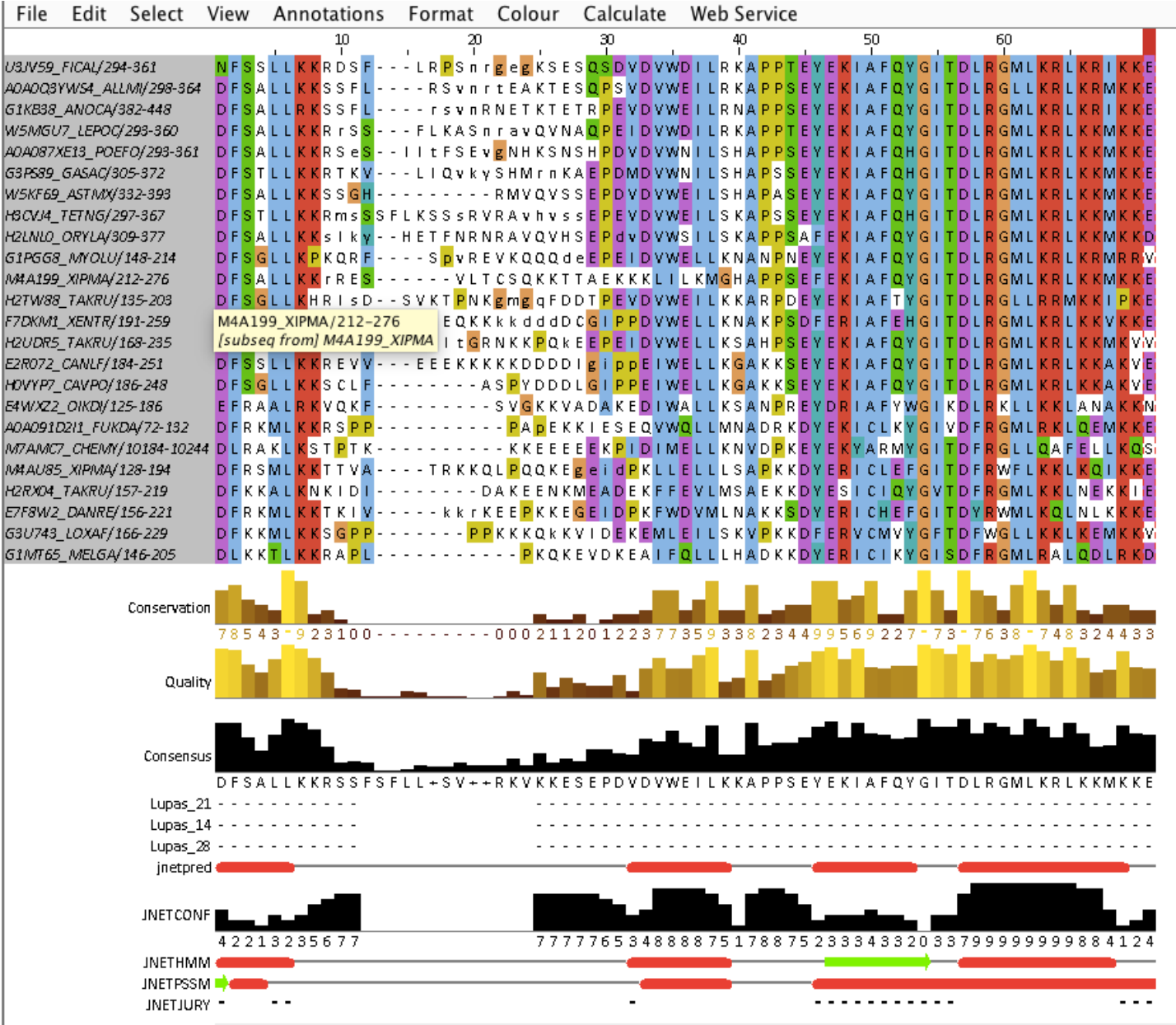


Click on top of column corresponding to alignment position 71

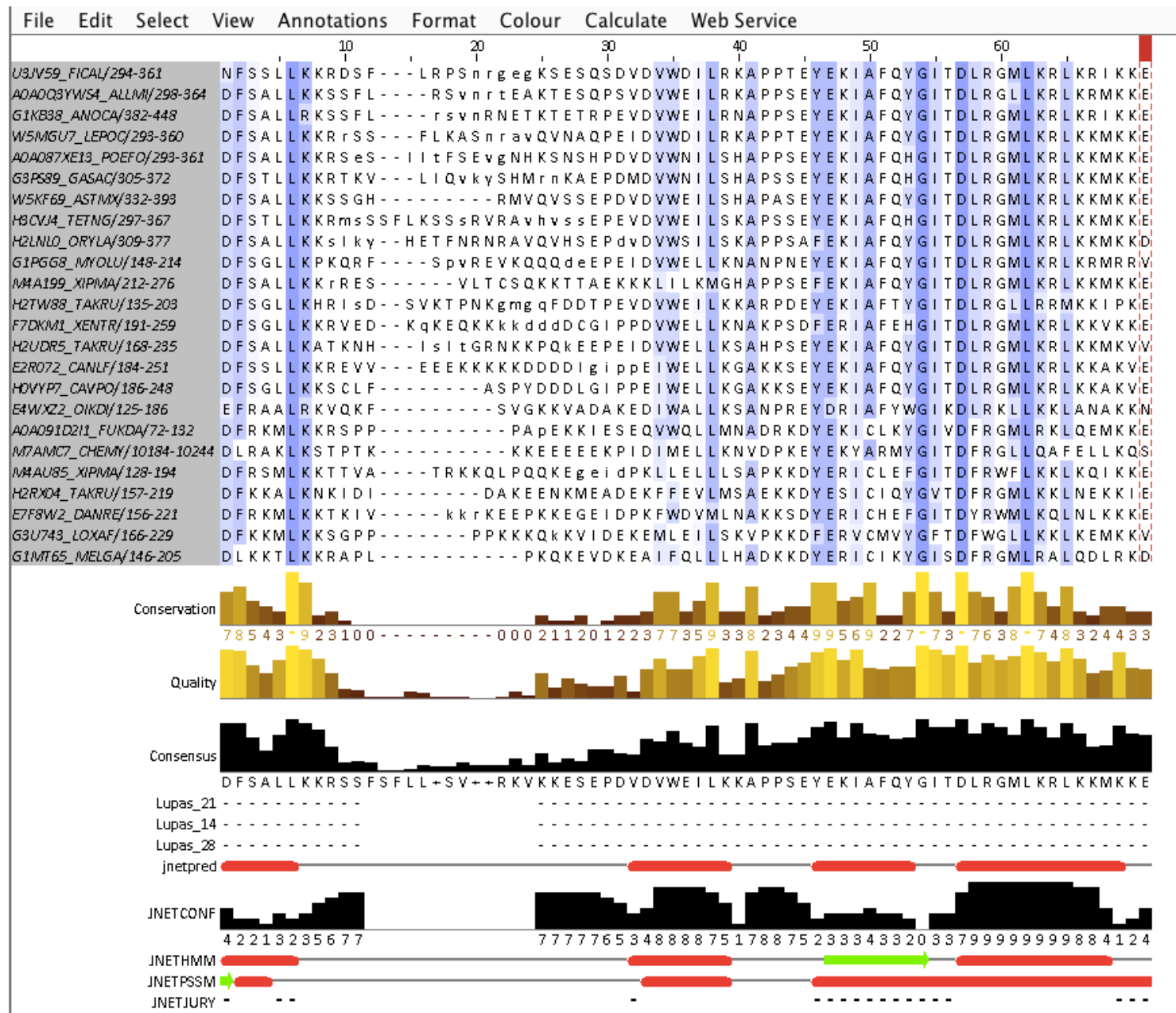


1. Edit -> Remove right

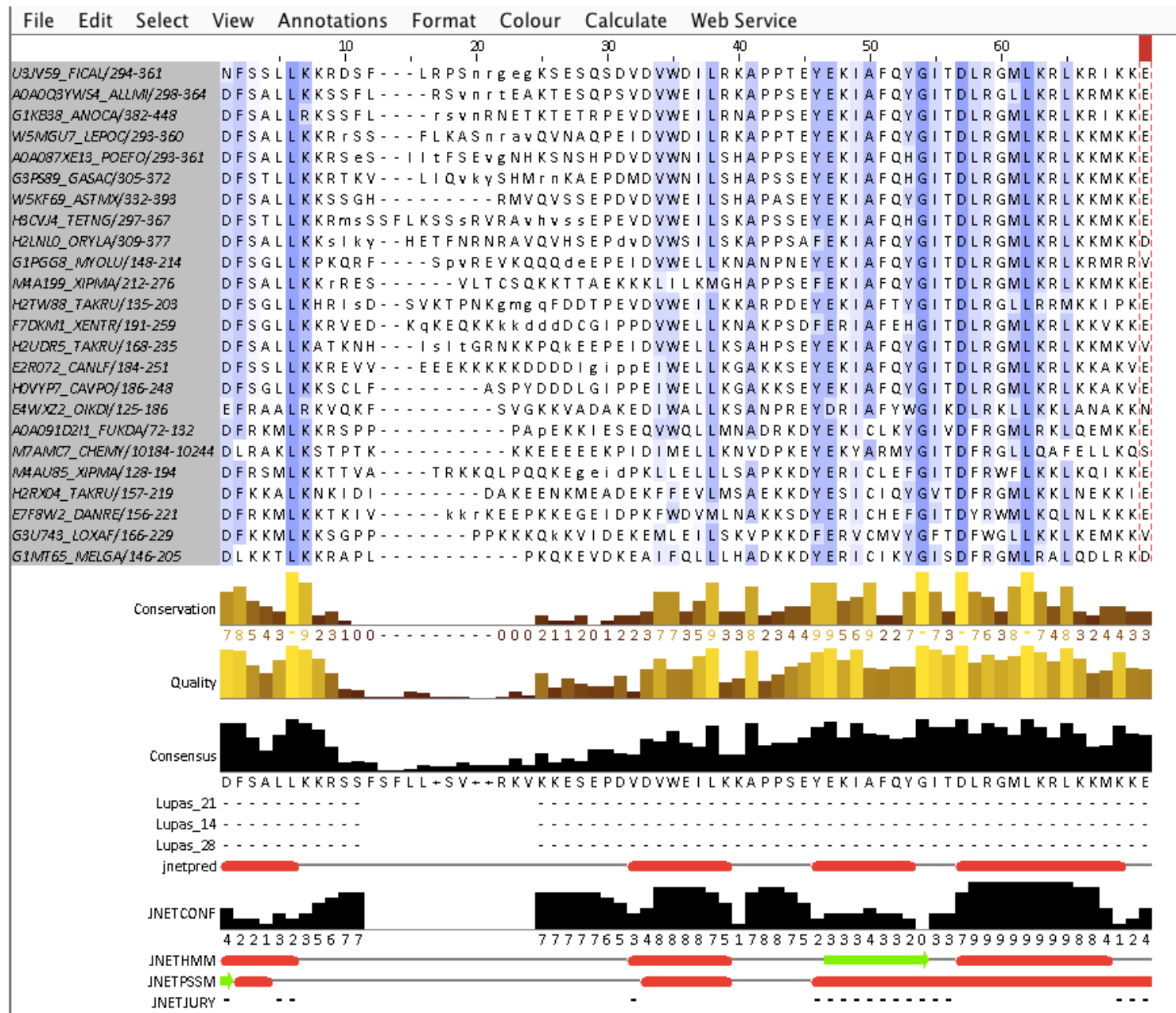
Edit -> Remove right



1. Colour -> BLOSUM62 Score / Colour -> Percentage Identity



1. File -> Save as



Annotation team: results?

 Basket ▼








[illegible]

Display












None

| | |
|-------------------------------------|-------------------|
| <input checked="" type="checkbox"/> | Function |
| <input checked="" type="checkbox"/> | Names & Taxonomy |
| <input checked="" type="checkbox"/> | Subcell. location |
| <input checked="" type="checkbox"/> | Pathol./Biotech |
| <input checked="" type="checkbox"/> | PTM / Processing |
| <input checked="" type="checkbox"/> | Expression |
| <input checked="" type="checkbox"/> | Interaction |
| <input checked="" type="checkbox"/> | Structure |
| <input checked="" type="checkbox"/> | Family & Domains |
| <input checked="" type="checkbox"/> | Sequences (2) |
| <input checked="" type="checkbox"/> | Cross-references |
| <input checked="" type="checkbox"/> | Publications |
| <input checked="" type="checkbox"/> | Entry information |
| <input checked="" type="checkbox"/> | Miscellaneous |
| <input checked="" type="checkbox"/> | Similar proteins |

GO - Molecular functionⁱ

- ATPase activator activity  Source: BHF-UCL
- identical protein binding  Source: IntAct ▼
- metal ion binding  Source: UniProtKB-KW
- myosin binding  Source: BHF-UCL ▼
- myosin heavy chain binding  Source: BHF-UCL ▼
- structural constituent of muscle  Source: BHF-UCL ▼
- titin binding  Source: BHF-UCL ▼

GO - Biological processⁱ

- cardiac muscle contraction  Source: BHF-UCL
- cell adhesion  Source: UniProtKB-KW
- heart morphogenesis  Source: BHF-UCL ▼
- muscle filament sliding  Source: Reactome
- myosin filament assembly  Source: Ensembl
- positive regulation of ATPase activity  Source: BHF-UCL
- regulation of heart rate  Source: Ensembl
- regulation of muscle filament sliding  Source: BHF-UCL
- regulation of striated muscle contraction  Source: BHF-UCL
- sarcomere organization  Source: Ensembl
- ventricular cardiac muscle tissue morphogenesis  Source: HGNC ▼








Complete GO annotation...

Display

None

- ☒ Function
- ☒ Names & Taxonomy
- ☒ Subcell. location
- ☒ Pathol./Biotech
- ☒ PTM / Processing
- ☒ Expression
- ☒ Interaction
- ☒ Structure
- ☒ Family & Domains
- ☒ Sequences (2)
- ☒ Cross-references
- ☒ Publications
- ☒ Entry information
- ☒ Miscellaneous
- ☒ Similar proteins












GO - Molecular functionⁱ

- ATPase activator activity  Source: BHF-UCL
- identical protein binding  Source: IntAct
- metal ion binding  Source: UniProtKB-KW
- myosin binding  Source: BHF-UCL
- myosin heavy chain binding  Source: BHF-UCL
- structural constituent of ribosome  Source: UniProtKB-KW
- titin binding  Source: BHF-UCL

Inferred from direct assayⁱ

PubMed 10024460

GO - Biological process

- cardiac muscle contraction  Source: BHF-UCL
- cell adhesion  Source: UniProtKB-KW
- heart morphogenesis  Source: BHF-UCL
- muscle filament sliding  Source: Reactome
- myosin filament assembly  Source: Ensembl
- positive regulation of ATPase activity  Source: BHF-UCL
- regulation of heart rate  Source: Ensembl
- regulation of muscle filament sliding  Source: BHF-UCL
- regulation of striated muscle contraction  Source: BHF-UCL
- sarcomere organization  Source: Ensembl
- ventricular cardiac muscle tissue morphogenesis  Source: HGNC

Complete GO annotation...

Literature citation

Map to

 Format

Mapped (2)

Mutations in beta-myosin S2 that cause familial hypertrophic cardiomyopathy (FHC) abolish the interaction with the regulatory domain of myosin-binding protein-C.

Gruen M., Gautel M.

The myosin filaments of striated muscle contain a family of enigmatic myosin-binding proteins (MyBP), MyBP-C and MyBP-H. These modular proteins of the intracellular immunoglobulin superfamily contain unique domains near their N termini. The N-terminal domain of cardiac MyBP-C, the MyBP-C motif, contains additional phosphorylation sites and may regulate contraction in a phosphorylation dependent way. In contrast to the C terminus, which binds to the light meromyosin portion of the myosin rod, the interactions of this domain are unknown. We demonstrate that fragments of MyBP-C containing the MyBP-C motif localise to the sarcomeric A-band in cardiomyocytes and isolated myofibrils, without affecting sarcomere structure. The binding site for the MyBP-C motif resides in the N-terminal 126 residues of the S2 segment of the myosin rod. In this region, several mutations in beta-myosin are associated with FHC; however, their molecular implications remained unclear. We show that two representative FHC mutations in beta-myosin S2, R870H and E924K, drastically reduce MyBP-C binding (K_d approximately 60 μM for R870H compared with a K_d of approximately 5 μM for the wild-type) down to undetectable levels (E924K). These mutations do not affect the coiled-coil structure of myosin. We suggest that the regulatory function of MyBP-C is mediated by the interaction with S2 and that mutations in beta-myosin S2 may act by altering the interactions with MyBP-C.



Download PDF



Export



Advanced search



Journal of Molecular Biology

Volume 286, Issue 3, 26 February 1999, Pages 933–949



Regular article

Mutations in β -myosin S2 that cause familial hypertrophic cardiomyopathy (FHC) abolish the interaction with the regulatory domain of myosin-binding protein-C¹

Mathias Gruen¹, Mathias Gautel^{1, 2}, [+ Show more](#)

doi:10.1006/jmbi.1998.2522

[Get rights and content](#)

Abstract

The myosin filaments of striated muscle contain a family of enigmatic myosin-binding

Previous investigations suggested that the cAPK-mediated phosphorylation of c-MyBP-C is associated with structural changes of the myosin crossbridges (Weinberg & Winegrad, 1996). Our observation that the N-terminal phosphorylation domain of MyBP-C binds to thick filaments points to an interaction with myosin itself. We mapped this interaction to the ≈ 100 residue MyBP-C motif. This domain is highly conserved between species and isoforms, with only two exceptions: the cardiac-specific phosphorylation loop insertion and an isoform-specific region rich in charged residues (Figure 1(c)). Both cardiac and skeletal C1C2 fragments show similar binding behaviour in the cosedimentation assay and in ITC. We therefore conclude that this interaction is a general feature of the regulatory domain of MyBP-C. This finding is further supported by

Gruen and Gautel JMB (1999)

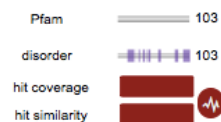


PHMMER Results

Score Taxonomy Domain Download

Search Again

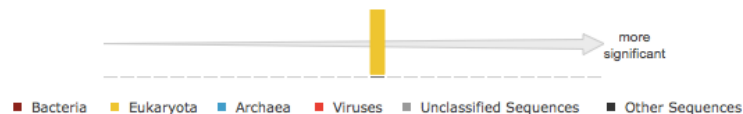
Sequence Matches and Features



✓ disorder ✓ coiled-coil ✓ tm & signal peptide

Show hit details

Distribution of Significant Hits



Significant Query Matches (1) in *pdb* (v.2016-03-02) Customize

| | Target | Description | Species | E-value |
|---|-----------------|----------------|--------------|---------|
| > | 2lhu_A PDB RCSB | Mybpc3 protein | Mus musculus | 6.3e-64 |

(show all) alignments

Your search took:0.09 secs
showing rows 1 - 1 of 1

Search Details

hit coverage 
hit similarity 

✓ disorder ✓ coiled-coil ✓ tm & signal peptide

[Show hit details](#)

Distribution of Significant Hits



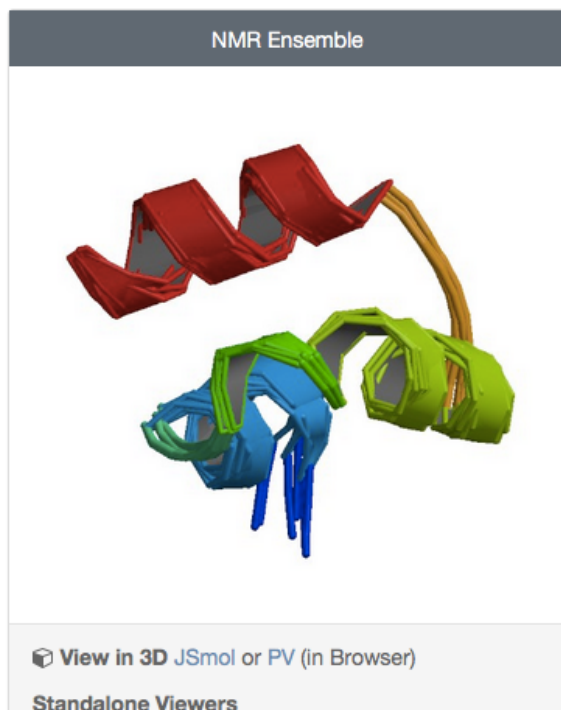
■ Bacteria ■ Eukaryota ■ Archaea ■ Viruses ■ Unclassified Sequences ■ Other Sequences

Significant Query Matches (1) in *pdb* (v.2016-03-02)

[Customize](#)

| | | Target | | | | Description | | | Species | | E-value | |
|--|-------|---|-----|------------------|-----|----------------|----------|--------------------|------------------------------|-----------|---------|---------|
| V | | 2lhu_A PDB RCSB | | | | Mybpc3 protein | | | Mus musculus | | 6.3e-64 | |
| Query | start | Target Envelope | | Target Alignment | | Bias | Accuracy | % Identity (count) | % Similarity (count) | Bit Score | E-value | |
| | | start | end | start | end | | | | | | Ind. | Cond. |
| | 1 | 103 | 22 | 124 | 22 | 1.99 | 1.00 | 100.0 (103) | 100.0 (103) | 216.0 | 7.4e-64 | 2.6e-69 |
| Query 1 heaigsgldlrslafrrtslagagrtsdshedagtl dfssllkkrdsfrdskleapaeedvweilrqappseyeriaf 80 heaigsgldlrslafrrtslagagrtsdshedagtl dfssllkkrdsfrdskleapaeedvweilrqappseyeriaf Target 22 HEAIGSGDLRLRSFRRSLAGAGRRSDSHEDAGTLDFSSLLKKRDSFRD SKLEAPAEEDVWEILRQAPPSEYERIAF 101 PP g***** | | | | | | | | | | | | |
| Query 81 qhgvtldlrgmlkrlkgmkqdekk 103 qhgvtldlrgmlkrlkgmkqdekk Target 102 QHGVTDLRGMLKRLKGMKQDEKK 124 PP *****98 | | | | | | | | | | | | |

Structure Summary 3D View Annotations Sequence Sequence Similarity Structure Similarity Experiment



2LHU

Structural Insight into the Unique Cardiac Myosin Binding Protein-C Motif: A Partially Folded Domain

DOI: [10.2210/pdb2lhu/pdb](https://doi.org/10.2210/pdb2lhu/pdb) BMRB: 17867

Classification: [STRUCTURAL PROTEIN](#)

Deposited: 2011-08-18 Released: 2012-01-11

Deposition author(s): [Howarth, J.W.](#), [Rosevear, P.R.](#), [Ramisetty, S.](#), [Nolan, K.](#), [Sadayappan, S.](#)

Organism: [Mus musculus](#)

Expression System: Escherichia coli

Structural Biology Knowledgebase: 2LHU (21 models >6 annotations) [SBKB.org](#)

Experimental Data Snapshot

Method: SOLUTION NMR

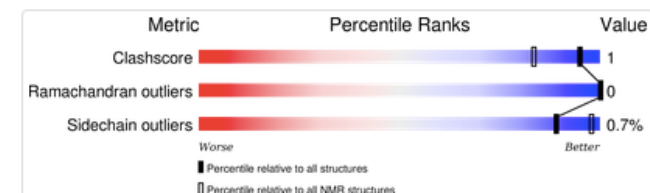
Conformers Calculated: 5000

Conformers Submitted: 20

Selection Criteria: Structures with the Lowest Energy

wwPDB Validation

[Full Report](#)



Mouse over an annotation to see more details. Click on any annotation to enable Jmol.

| Annotations | Details |
|--|--------------------------------------|
| Secondary Structure: DSSP [hide] [reference] | 18% helical (4 helices; 23 residues) |

Sequence Chain View

DSSP

PDB **MGSSHHHHHSSGLVPRGSHMHEAIGSGDLDLRSAFRRTSLAGAGRRTSDSHEDAGTLDF**
PDB

DSSP

PDB **SSLLKKRDSFRRDSKLEAPAEEDVWEILRQAPPSEYERIAFQHGVTDLRGMLKRLKGMKQ**
PDB

315 320 330 340

DSSP

PDB **DEKK**
PDB

DSSP Legend



T: turn



empty: no secondary structure assigned



G: 3/10-helix



S: bend



H: alpha helix

Mouse over an annotation to see more details. Click on any annotation to enable Jmol.

| Annotations | Details |
|--|--------------------------------------|
| Secondary Structure: DSSP [hide] [reference] | 18% helical (4 helices; 23 residues) |

Sequence Chain View

DSSP

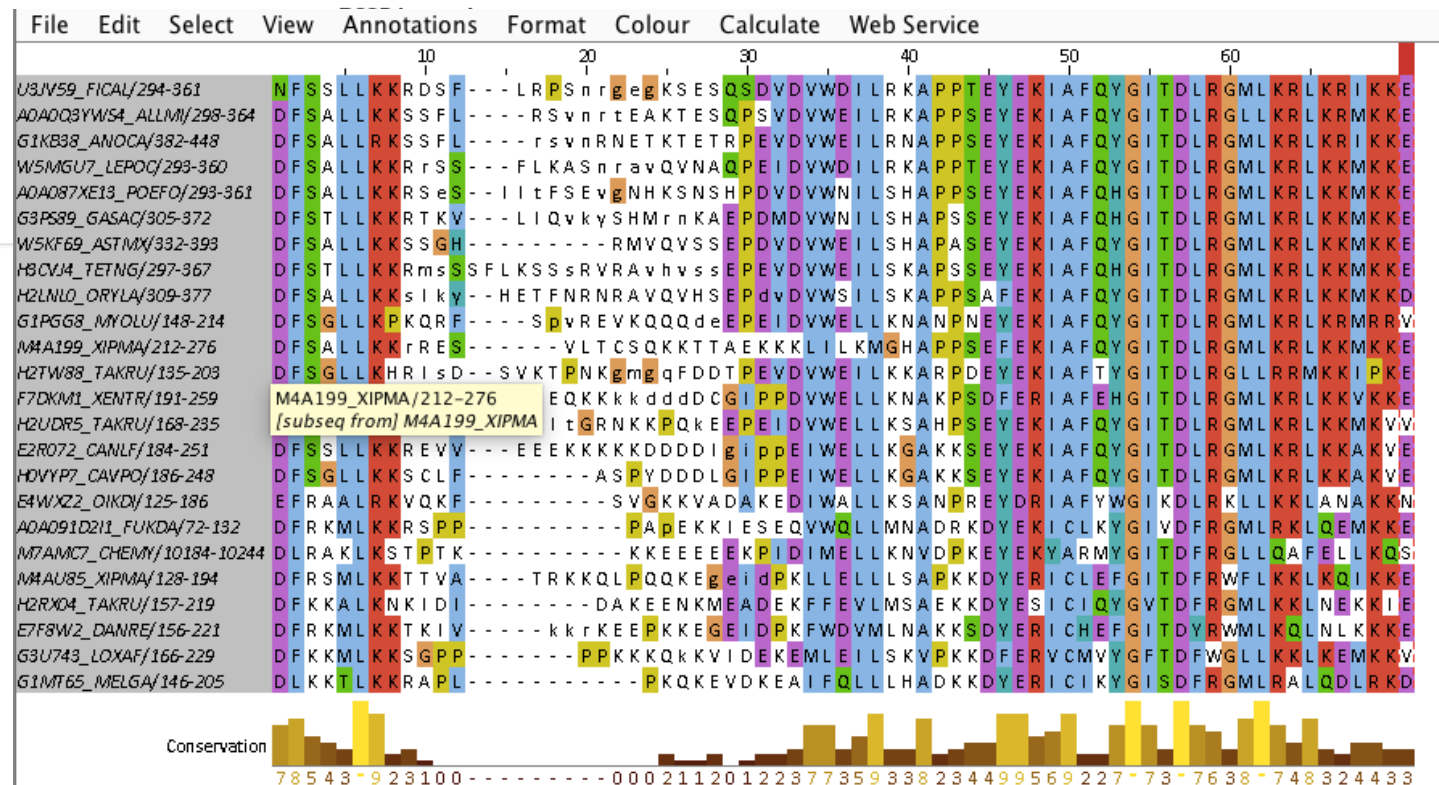
PDB MGSSHHHHHHSSGLVPRGSHMHEAIGSGDLDLRSFRRTSLAGAGRRTSDSHEDAGTLDFF
PDB

DSSP

PDB SSLKKRDSFRRDSKLEAPAEEDVWEILRQAPPSEYERIAFQHGVTDLRGMLKRLKGMKQ
PDB

DSSP

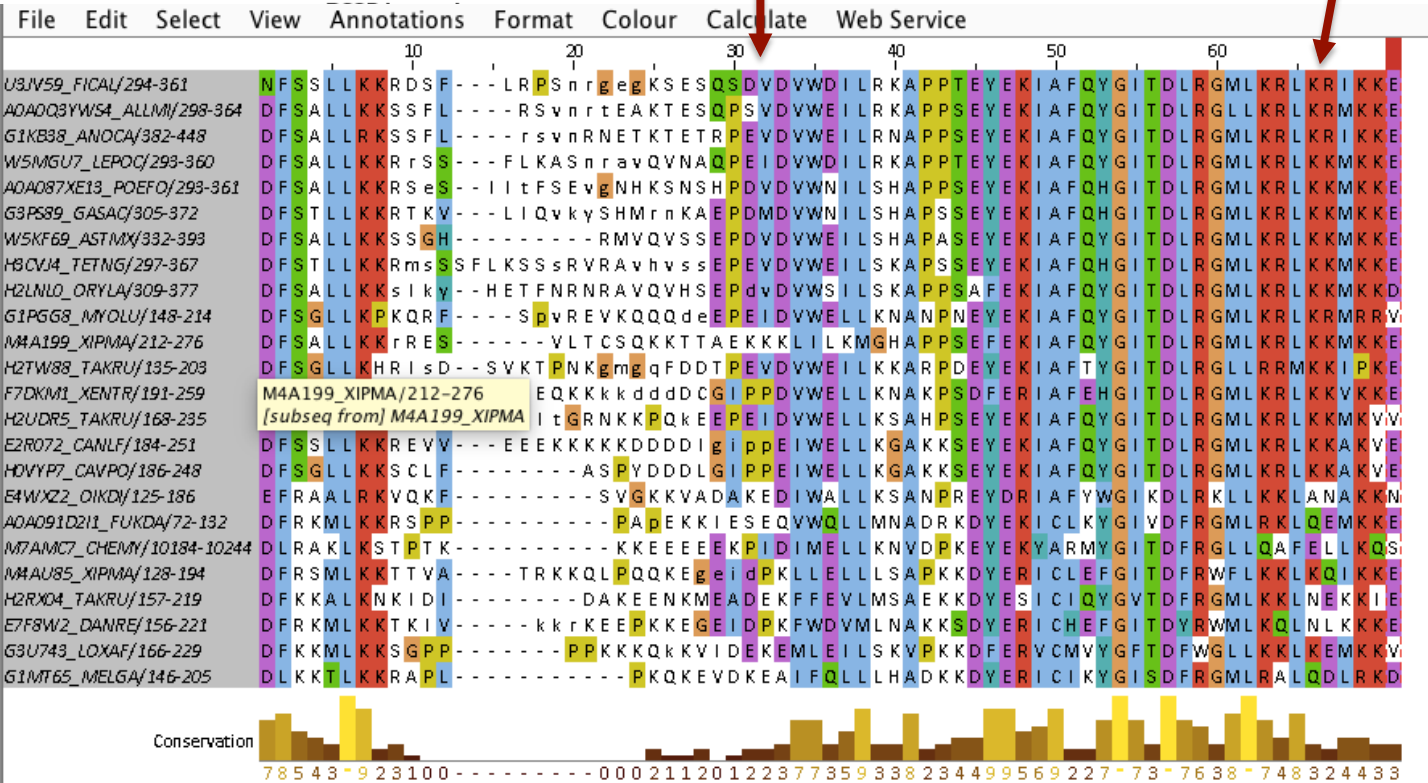
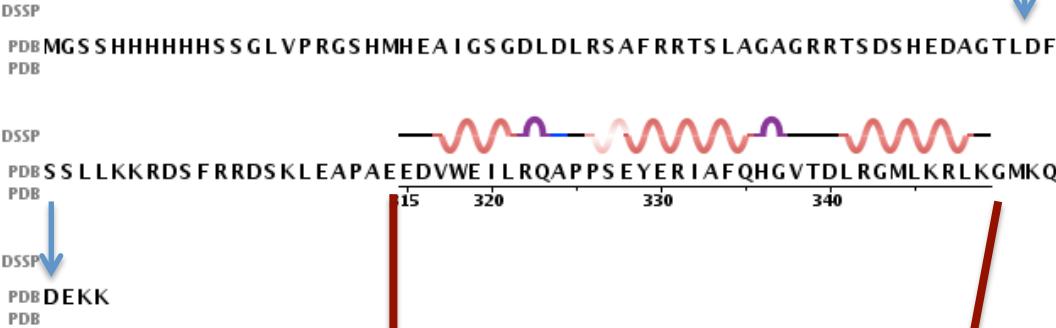
PDB DEKK
PDB

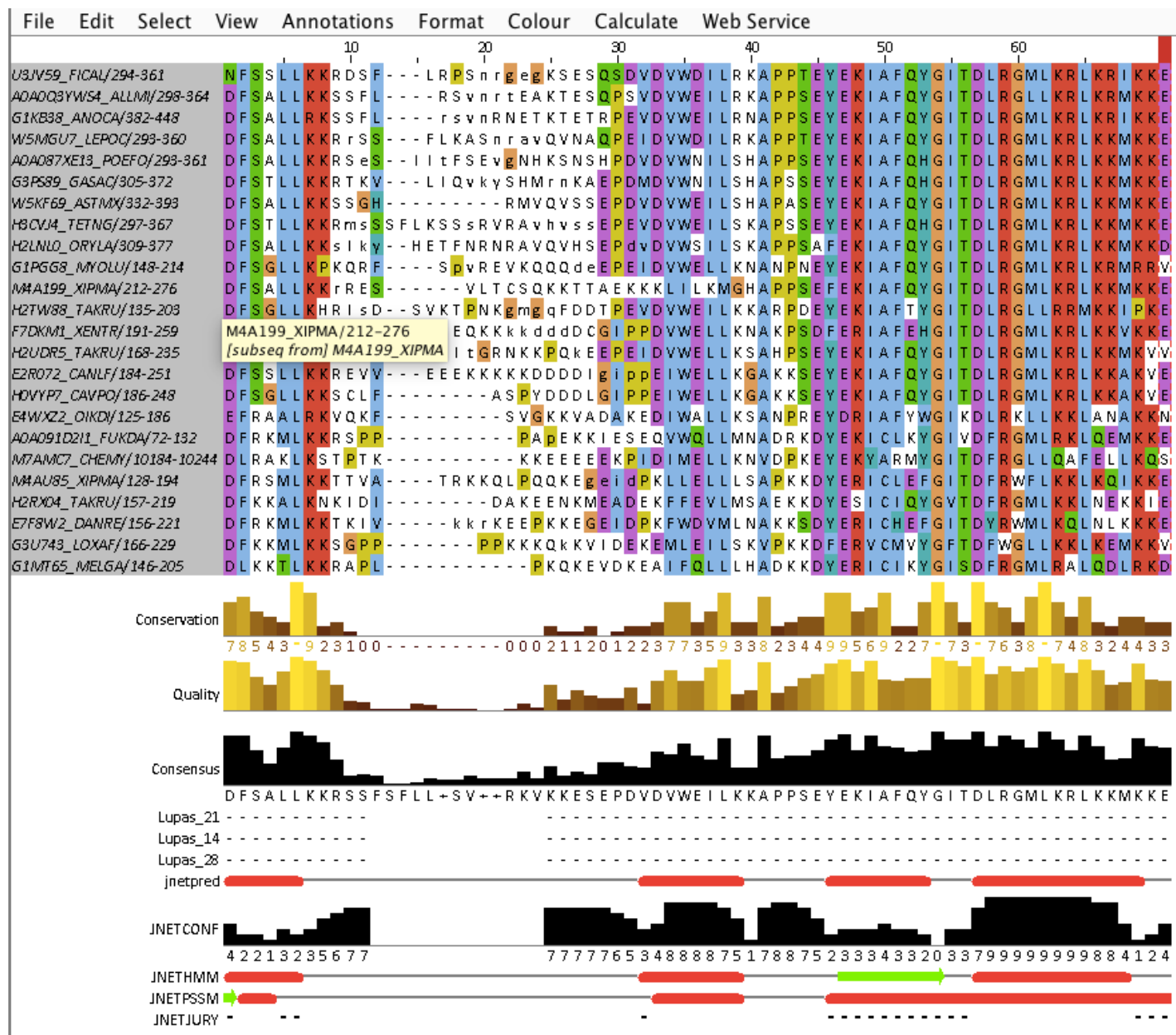


Mouse over an annotation to see more details. Click on any annotation to enable Jmol.

| Annotations | Details |
|--|--------------------------------------|
| Secondary Structure: DSSP [hide] [reference] | 18% helical (4 helices; 23 residues) |

Sequence Chain View





Structural Insight into Unique Cardiac Myosin-binding Protein-C Motif

A PARTIALLY FOLDED DOMAIN*

Jack W. Howarth[‡], Srinivas Ramiseti[‡], Kristof Nolan[‡],
Sakthivel Sadayappan[§] and Paul R. Rosevear^{‡,1}

+ Author Affiliations

↩¹ To whom correspondence should be addressed: Dept. of Molecular Genetics, Biochemistry, and Microbiology, University of Cincinnati College of Medicine, 231 Albert Sabin Way, Cincinnati, OH 45267. Tel.: 513-558-3370; E-mail: paul.rosevear@uc.edu.

Capsule

Background: Cardiac myosin-binding protein-C is a sarcomeric assembly protein necessary for the regulation of sarcomere structure and function.

Results: The cMyBP-C motif is composed of two subdomains, a largely disordered N-terminal portion and a more ordered C-terminal subdomain.

Conclusion: The C-terminal subdomain is capable of forming a three-helix bundle.

Significance: The three-helix bundle may provide a platform for actin binding.

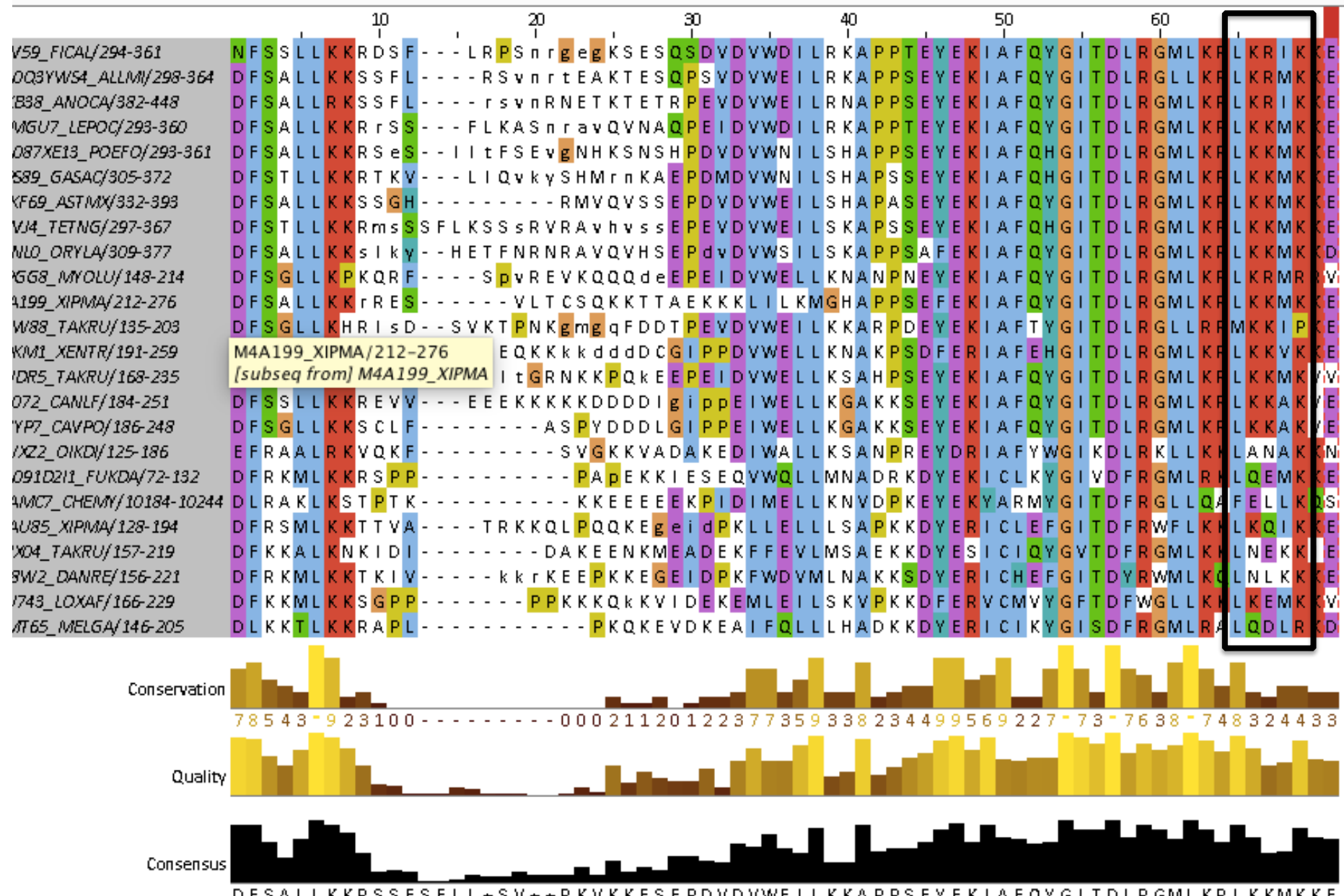
Abstract

The structural role of the unique myosin-binding motif (m-domain) of cardiac myosin-binding protein-C remains unclear. Functionally, the m-domain is thought to directly interact with myosin, whereas phosphorylation of the m-domain has been shown to modulate interactions between myosin and actin. Here we utilized NMR to analyze the structure and dynamics of the m-domain in solution. Our studies reveal that the m-domain is composed of two subdomains, a largely disordered N-terminal portion containing three known phosphorylation sites and a more ordered and folded C-terminal portion. Chemical shift analyses, $d_{NN}(i, i +$

In summary, we have shown that the N-terminal portion of the m-domain, containing phosphorylation sites Ser-273, Ser-282, and Ser-302, is conformationally flexible, containing two transiently formed helices that span residues 263–267 for helix 1 and residues 292–298 for helix 2. Bioinformatics analyses predict that introduction of negative charge at Ser-273, mimicking phosphorylation, will extend and stabilize helix 1. In contrast, $^{15}\text{N}\{^1\text{H}\}$ NOE and chemical shift analyses show that the C-terminal portion of the m-domain folds into a stable three-helix bundle. A known actin-binding motif, LK(R/K)XK, is positioned in the third helix ($\alpha 3$), similar to that found in villin and related proteins. These results suggest that the m-domain in cMyBP-C may alter actomyosin interactions in the heart through interactions with actin. Current studies are aimed at the role of the N-terminal portion of the m-domain in modulating actomyosin interactions via protein phosphorylation.

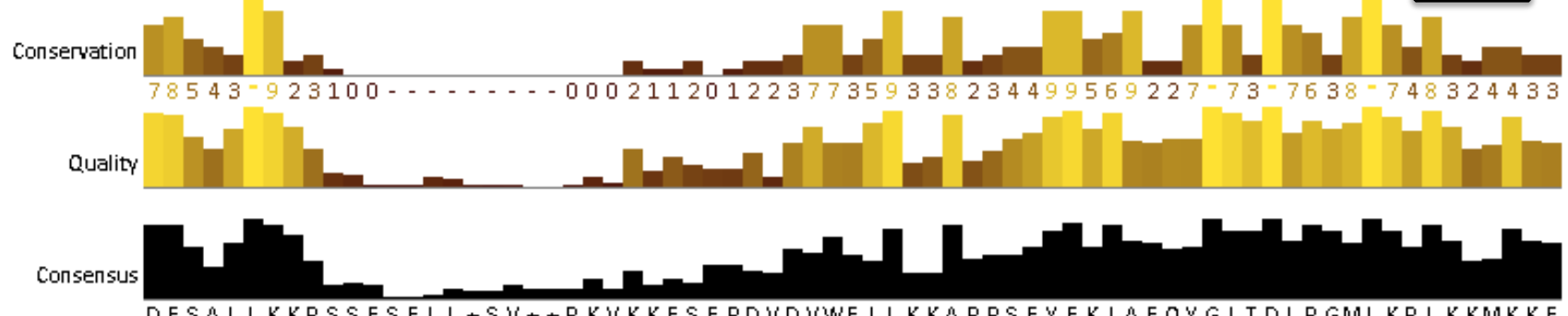
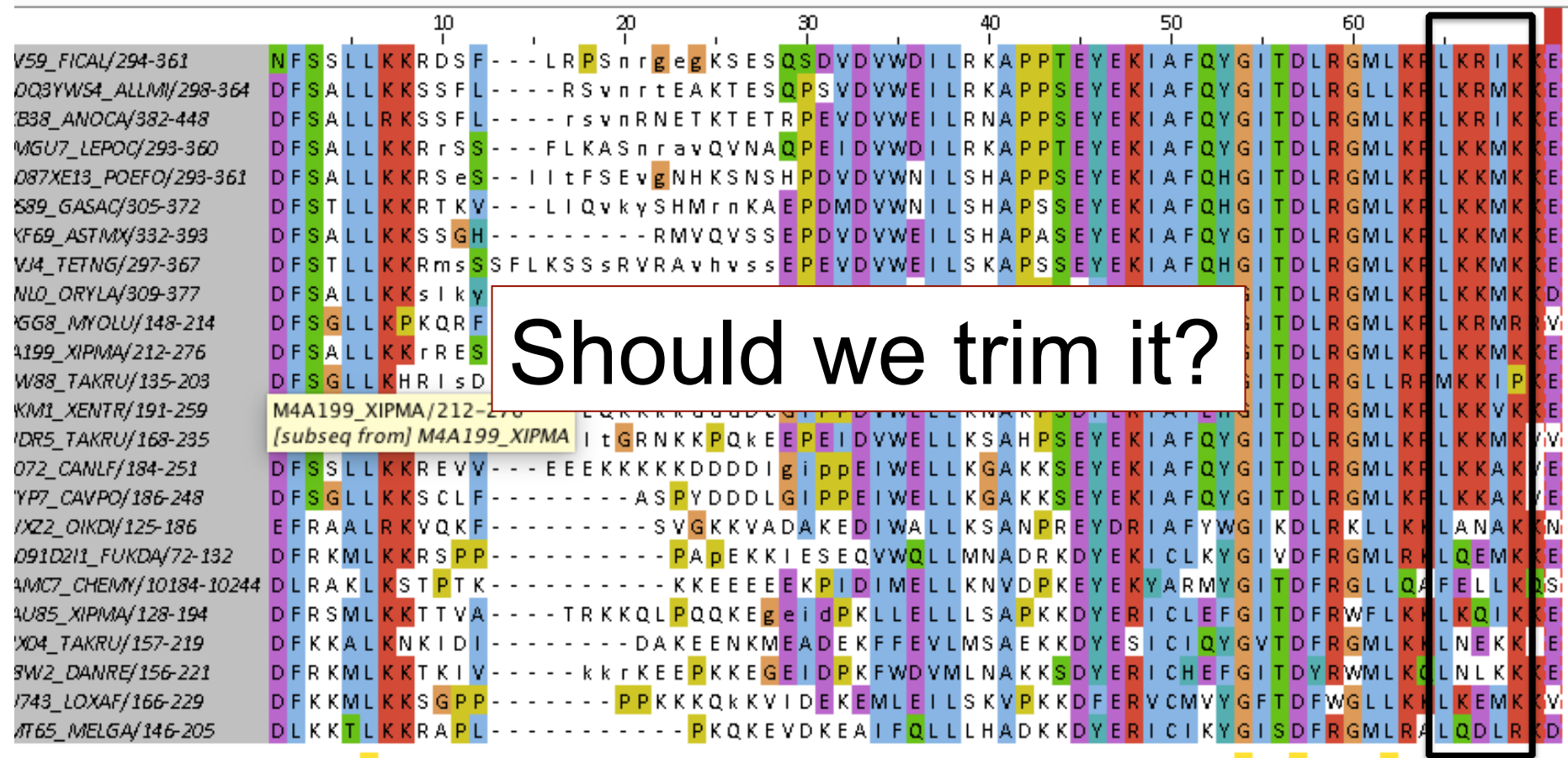
LK(R/K)XK

File Edit Select View Annotations Format Colour Calculate Web Service



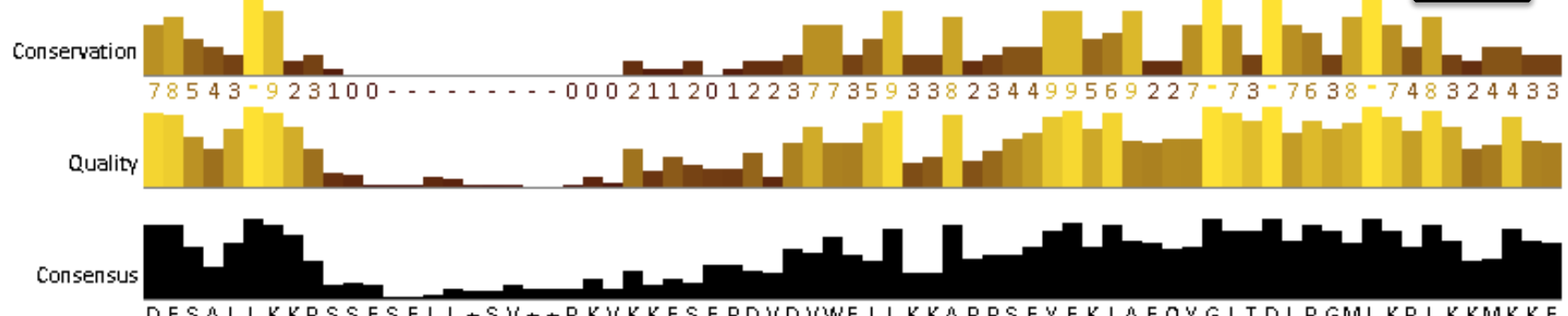
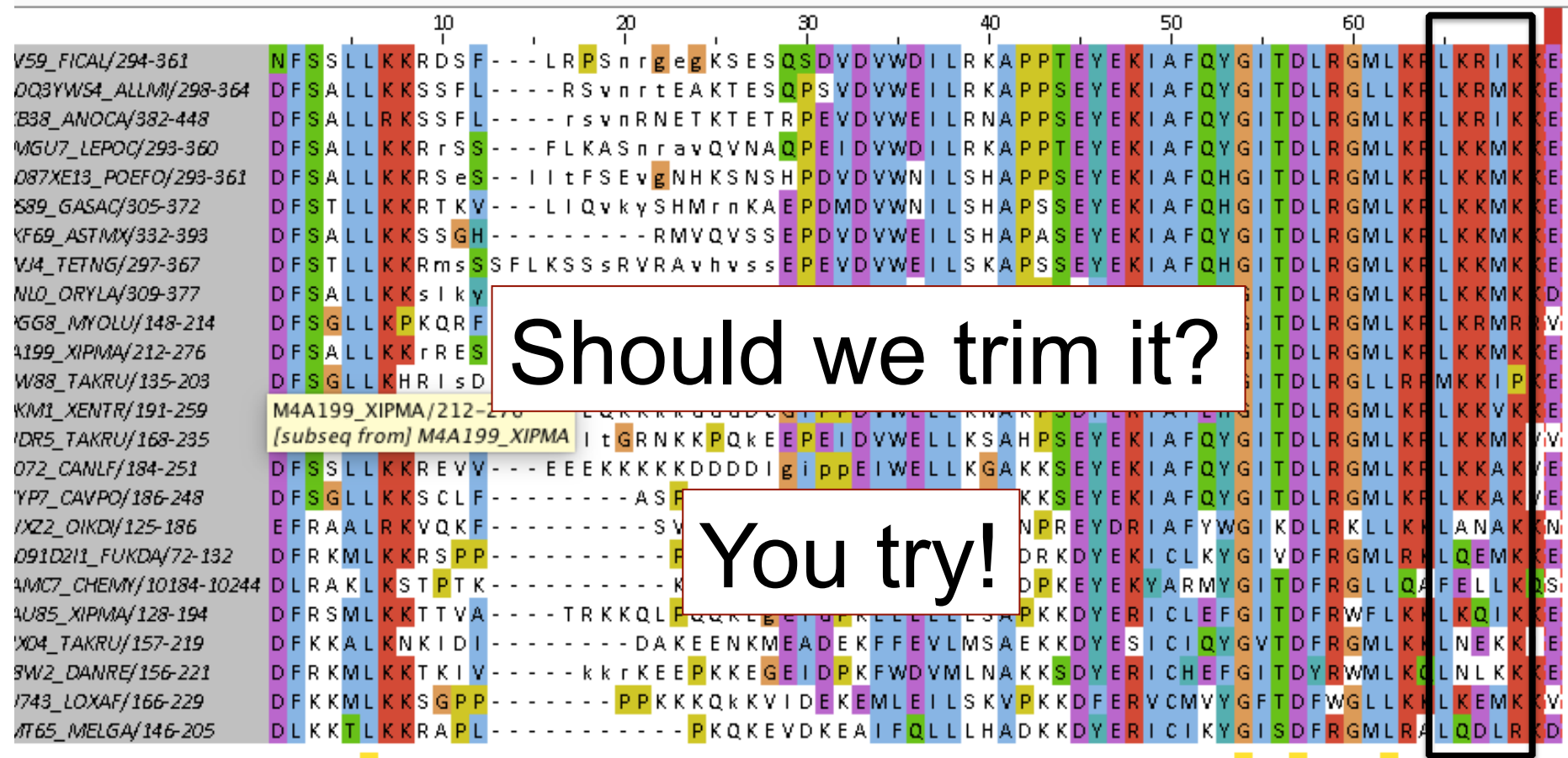
LK(R/K)XK

File Edit Select View Annotations Format Colour Calculate Web Service



LK(R/K)XK

File Edit Select View Annotations Format Colour Calculate Web Service



Search Team



[phmmer](#) [hmmscan](#) **[hmmsearch](#)** [jackhmmer](#)

protein alignment/profile-HMM vs protein sequence database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Upload a file

[Browse...](#)

jalview-ali.fasta

Submit

[Reset](#)

▼ Sequence Database ?

Frequently used databases

☒ Reference Proteomes ☐ UniProtKB ☐ SwissProt ☐ PDB

Representative Sets (UniProt)

☐ rp75 ☐ rp55 ☐ rp35 ☐ rp15

Other databases

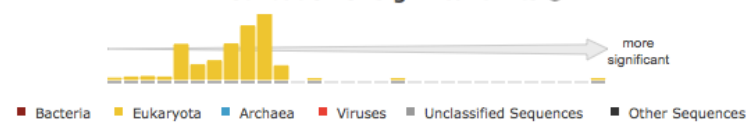
☐ Ensembl Genomes ? ☐ QfO

► [Restrict by Taxonomy](#) ?

▼ Cut-Offs ?



Distribution of Significant Hits



Significant Query Matches (437) in uniprotrefprot (v.2016-03-02)

Customize

| Target | Description | Species | E-value |
|--------------------|---|---|---------|
| > W4Y2S3_STRPU | Uncharacterized protein | Strongylocentrotus purpuratus | 1.4e-69 |
| > W4Z1N0_STRPU | Uncharacterized protein | Strongylocentrotus purpuratus | 3.0e-41 |
| > M7AMC7_CHEMY | Titin (Fragment) | Chelonia mydas | 4.4e-32 |
| > E7EZ79_DANRE | Uncharacterized protein | Danio rerio | 1.0e-27 |
| > F6VSF7_XENTR | Uncharacterized protein (Fragment) | Xenopus tropicalis | 1.2e-27 |
| > W5MGU7_LEPOC | Uncharacterized protein | Lepisosteus oculatus | 1.8e-27 |
| > F1QVS8_DANRE | Uncharacterized protein | Danio rerio | 2.5e-27 |
| > A0A087XE13_POEFO | Uncharacterized protein | Poecilia formosa | 3.1e-27 |
| > G1NE13_MELGA | Uncharacterized protein | Meleagris gallopavo | 5.8e-27 |
| > H3BI37_LATCH | Uncharacterized protein | Latimeria chalumnae | 7.1e-27 |
| > W5KF69_ASTMX | Uncharacterized protein | Astyanax mexicanus | 7.2e-27 |
| > A0A0Q3Z721_ALLMI | Spleen focus forming virus (SFFV) proviral integration oncogene isoform B | Alligator mississippiensis | 7.6e-27 |
| > I3J4A1_ORENI | Uncharacterized protein | Oreochromis niloticus | 7.7e-27 |
| > A0A0Q3YWS4_ALLMI | Spleen focus forming virus (SFFV) proviral integration oncogene isoform A | Alligator mississippiensis | 7.7e-27 |
| > MYPC3_CHICK | Isoform Type II of Myosin-binding protein C, cardiac-type | Gallus gallus | 7.7e-27 |
| > MYPC3_CHICK | Myosin-binding protein C, cardiac-type | Gallus gallus | 7.9e-27 |
| > A0A0Q3Z7H8_ALLMI | Spleen focus forming virus (SFFV) proviral integration oncogene isoform D | Alligator mississippiensis | 8.5e-27 |

| | | | |
|----------------|--|-------------------------|---------|
| > G5BBJ0_HETGA | Myosin-binding protein C, slow-type | Heterocephalus glaber | 2.1e-24 |
| > M3YKB3_MUSPF | Uncharacterized protein | Mustela putorius furo | 2.1e-24 |
| > G3QUD6_GORGO | Uncharacterized protein | Gorilla gorilla gorilla | 2.4e-24 |
| > F6ZGS1_CALJA | Uncharacterized protein | Callithrix jacchus | 2.4e-24 |
| > F1PWA8_CANLF | Uncharacterized protein (Fragment) | Canis lupus familiaris | 2.4e-24 |
| > W5Q0I1_SHEEP | Uncharacterized protein (Fragment) | Ovis aries | 2.5e-24 |
| > Q3UIK0_MOUSE | Myosin-binding protein C, cardiac-type | Mus musculus | 2.6e-24 |

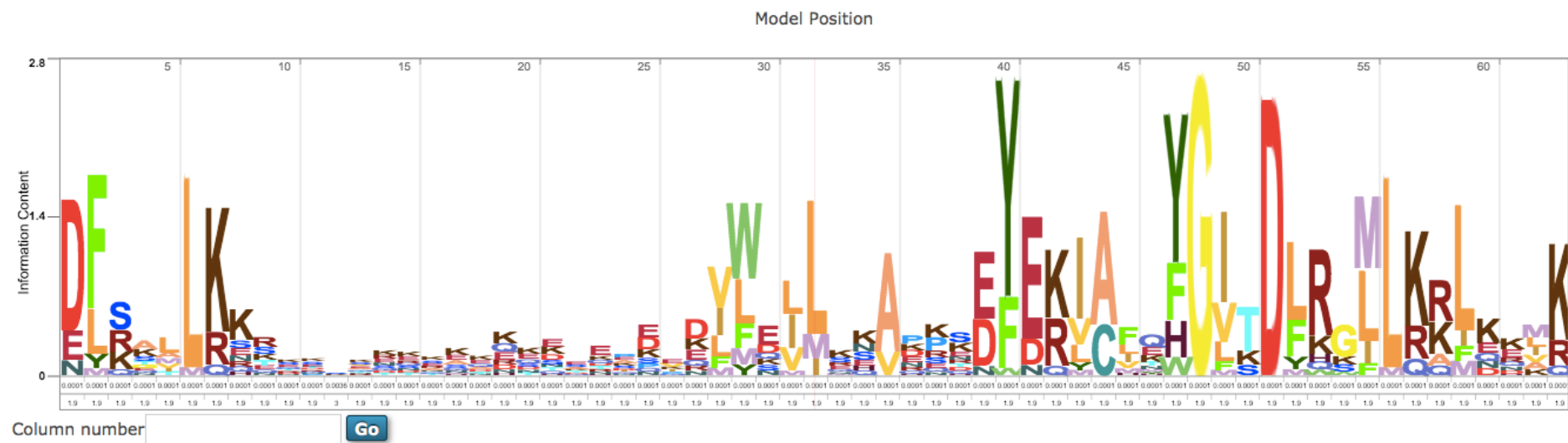
(show all) alignments

Your search took: 0.82 secs

showing rows 1 - 100 of 453

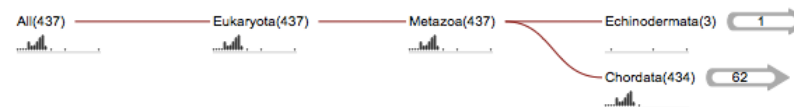
Search Details

« First « Previous **Page 1** of 5 Next » Last »





Taxonomic distribution of all search hits ?



| Species Distribution | | | |
|--|-------|----------------------|--|
| Species | Count | View | |
| Homo sapiens | 35 | Show | |
| Takifugu rubripes | 25 | Show | |
| Danio rerio | 23 | Show | |
| Callithrix jacchus | 19 | Show | |
| Alligator mississippiensis | 16 | Show | |
| Mus musculus | 15 | Show | |
| Poecilia formosa | 12 | Show | |
| Oreochromis niloticus | 12 | Show | |
| Gasterosteus aculeatus | 11 | Show | |
| Canis lupus familiaris | 11 | Show | |
| Sarcophilus harrisii | 10 | Show | |
| Macaca mulatta | 10 | Show | |
| Gorilla gorilla gorilla | 9 | Show | |



HMMSEARCH Results

[Search Again](#)Domain Architectures [?](#)« First « Previous **Page 1** of 3 Next » Last »

| | | |
|---|---|-----------------------------|
| 119 SEQUENCES Show All | with domain architecture: I-set, I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set , <i>example:F6VSF7_XENTR</i> ↗ Sequence Features  1264 | View Scores |
| 46 SEQUENCES Show All | with domain architecture: I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set , <i>example:H2UDR5_TAKRU</i> ↗ Sequence Features  1152 | View Scores |
| 41 SEQUENCES Show All | with domain architecture: I-set, I-set, I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set , <i>example:E7EZ79_DANRE</i> ↗ Sequence Features  1266 | View Scores |
| 18 SEQUENCES Show All | with domain architecture: I-set, I-set, I-set, fn3, fn3, fn3, I-set, fn3, I-set , <i>example:I3MSM7 ICTTR</i> ↗ Sequence Features  1253 | View Scores |
| 11 SEQUENCES Show All | with domain architecture: I-set, I-set, I-set , <i>example:F7CWG3_CALJA</i> ↗ Sequence Features  467 | View Scores |
| 11 SEQUENCES Show All | with domain architecture: I-set, I-set, I-set, I-set, I-set, fn3, fn3, fn3, I-set, fn3, I-set , <i>example:E2REQ2_CANLF</i> ↗ Sequence Features  1324 | View Scores |
| 9 | with domain architecture: I-set, I-set, I-set, I-set, fn3, fn3, fn3, I-set, fn3, I-set , <i>example:H3AVM7_LATCH</i> ↗ | View Scores |



HMMSEARCH Results

Score

Search Again

Domain Architectures ?

119
SEQUENCES
Show Allwith domain architecture: **I-set, I-set, I-set, I-set, I-set, fn3,**

Sequence Features

46
SEQUENCES
Show Allwith domain architecture: **I-set, I-set, I-set, I-set, fn3, fn3, I**

Sequence Features

41
SEQUENCES
Show Allwith domain architecture: **I-set, I-set, I-set, I-set, I-set, I-se**

Sequence Features

18
SEQUENCES
Show Allwith domain architecture: **I-set, I-set, I-set, fn3, fn3, fn3, I-s**

Sequence Features

11
SEQUENCES
Show Allwith domain architecture: **I-set, I-set, I-set, example:F7CWG3**

Sequence Features

11
SEQUENCES
Show Allwith domain architecture: **I-set, I-set, I-set, I-set, I-set, fn3,**

Sequence Features

9

with domain architecture: **I-set, I-set, I-set, I-set, fn3, fn3, fn3, I-set, fn3, I-set, example:H3AVM7_LATCH**

Search Again

Perform a new search

- with new input
- with these results

Previous Page 1 of 5 Next » Last »

Customize

Species

E-value

3 Next » Last »

View Scores

View Scores

View Scores

View Scores

View Scores

View Scores

View Scores

Thank you!!