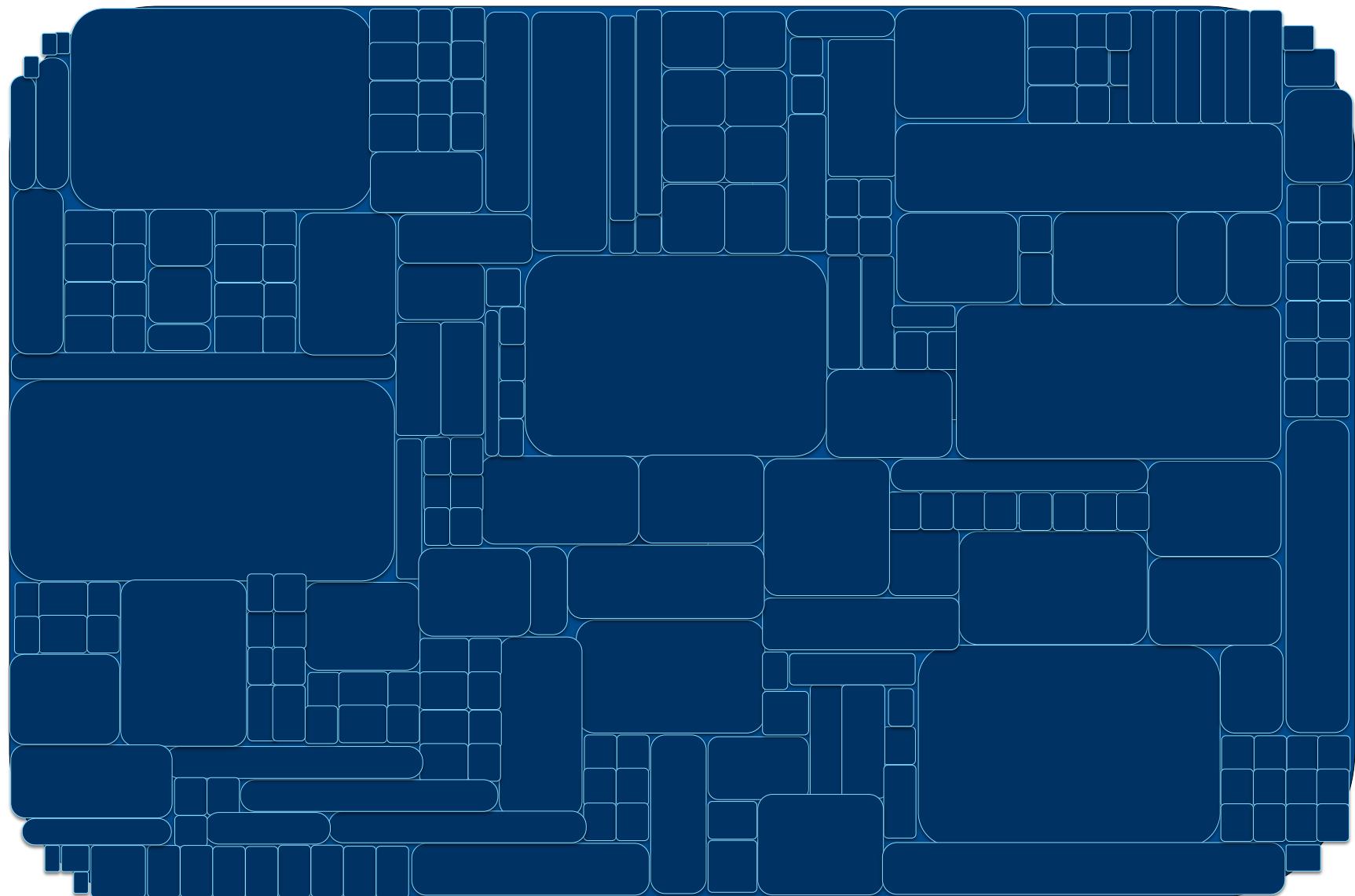


Protein family databases

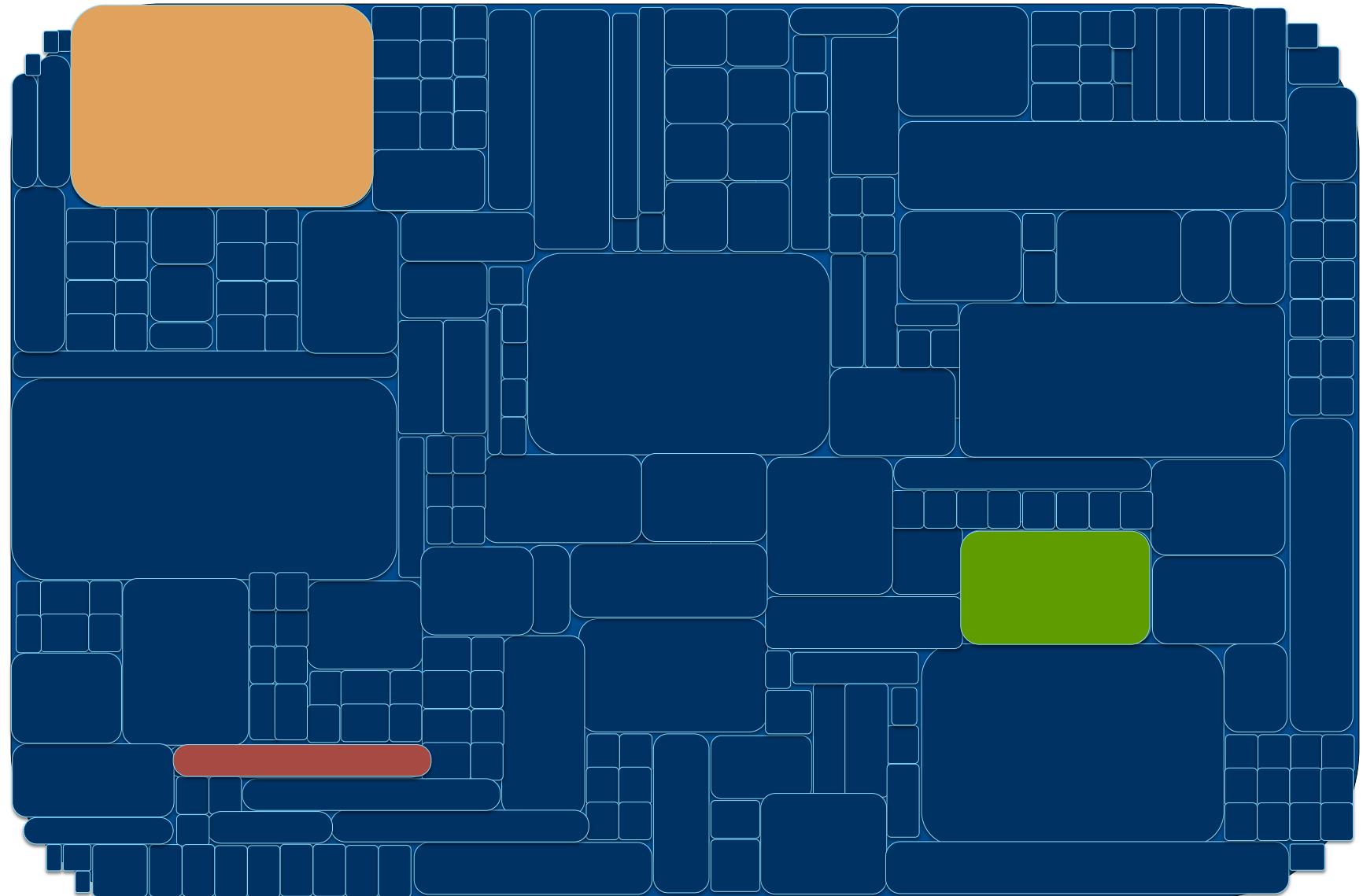
Protein families

- 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)

Marco Punta



EMBO Workshop, Budapest, 2016



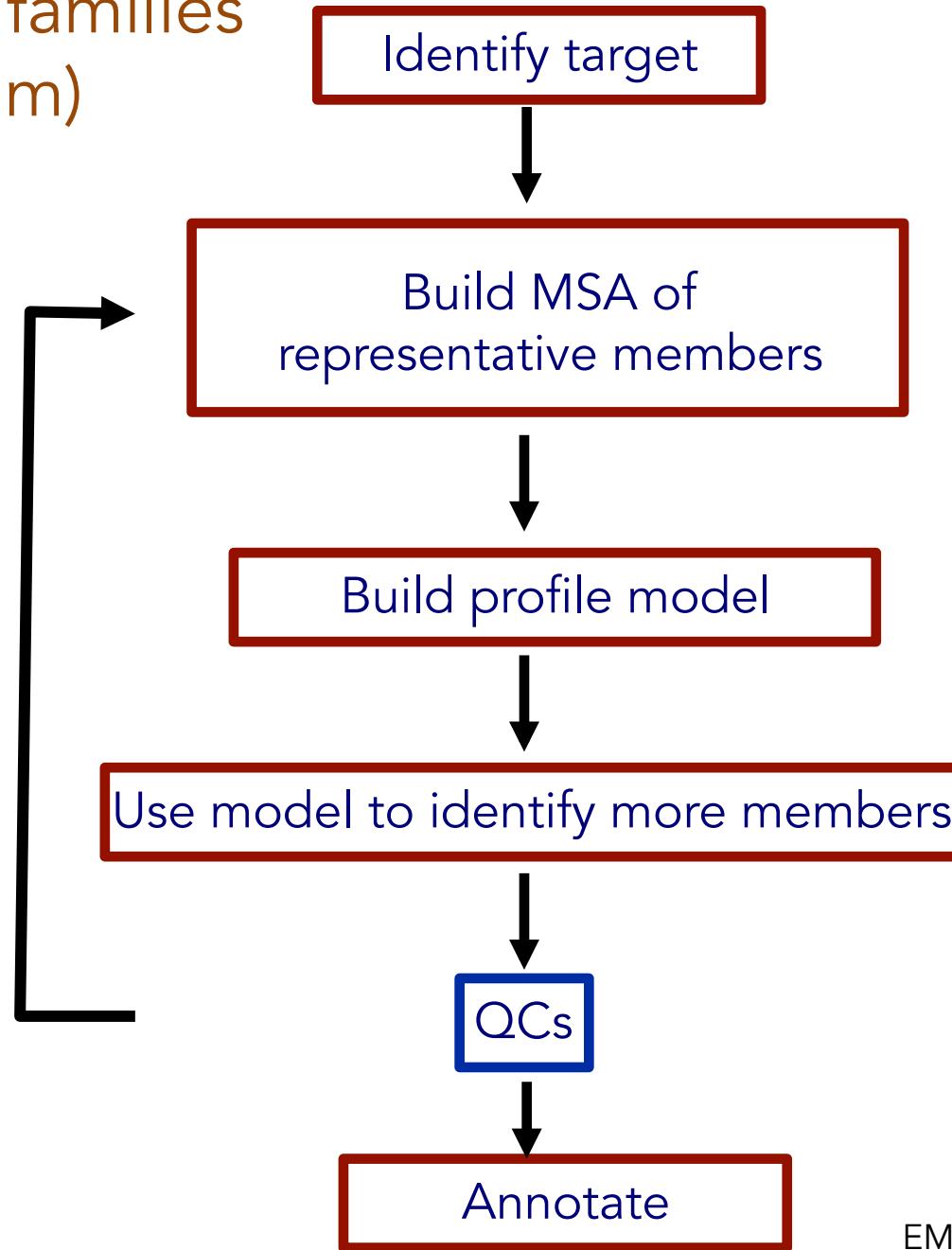
Ion transport

cNMP

L-zip

Building families (Pfam)

Marco Punta



Human: 1 MGLSDGEWQLVLNWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFHLKSEDEMKA 60

Human: 61 DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH 120

Human: 121 PGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154

Family power

Human: 1 MGLSDGEWQLVLNWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKA 60
MGLSDGEWQLVLNWGKVEAD GHGQEVL I LFK HPETL KFDKFK LKSE MK SE

Mouse: 1 MGLSDGEWQLVLNWGKVEADLAGHGQEVLIGLFKTHPETLDKFDKFKNLKSEEDMKG 60

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

Mouse: 61 DLKKHGCTVLTALGTILKKKGQHAAEIQPLAQSHATKHKIPVKYLEFISEIIIIEVLKKRH 120

Human: 121 PGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154
GDFGADAQGAM KALELFR D A YKELGFQG

Mouse: 121 SGDFGADAQGAMSKALELFRNDIAAKYKELGFQG 154

Family power

Human: 1 MGLSDGEWQLVLNWGKVEADIPGHGQEVLIRLFKGHPETLEKFIDIKFKHLKSEDEMKGASE 60
MGLSDGEWQLVLNWGKVEAD GHGQEVLIGHGQEVLI LFK HPETL KFDIKFKHLKSEDEMKGASE

Mouse: 1 MGLSDGEWQLVLNWGKVEADLAGHGQEVLI GLFKTHPETLDKFIDIKFKHLKSEEDMKGSE 60



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

Mouse: 61 DLKKHGCTVLTALGTILKKKGQHAAEIQPLAQSHATKHKIPVKYLEFISEIIIIEVLKKRH 120

Human: 121 PGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154
GDFGADAQGAM KALELFR D A YKELGFQG

Mouse: 121 SGDFGADAQGAMSKALELFRNDIAAKYKELGFQG 154

BLOSUM62 matrix

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
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

Family power

	10	20	30	40	50	60	70	80	90	100	110	120
HBAZ_CAPI/7-107	ERT I ILS LWSK-IST-----	QADVIGET T LER L FSCYPQAKT Y FPF F-	-DLH-----	S GS AQL RAHG SKVVA AVGDAVKS I-----	D-NVT SALS KLS E L HAY V-----	RVDPVNFKFLSHC I						
HBA3_PLEWA/7-107	EKA LVV GLCGK- IS G-----	HCD ALGGEA LDRL FAS FGQT RTY FSH F-----	-DLS-----	PGS ADVK R HG KKVL S AIGEA AKH I-----	D-SMD QALS KLS D L HAY NL-----	RVDPGNFQLLSHC I						
HBA4_CATCL/6-107	DKADVKIAWAK- ISP-----	RADE I GAEAL GRML T VY P QTKT Y FAHW-----	-ADLS-----	PGS GPV K HG KK VIMGA I GD AVT K F-----	D-DLLGG LAS L S E L HASKL-----	RVDPNSNFKILANC I						
HBB_HETPO/7-106	ELHE I TTT WKS- I-----	DKH S LGAKA LARMF I VY PWTTR Y F GNL-----	-KEFT-----	ACSYGVKEHAKKV T GALGVAVTH L-----	G-DVKS QFT DLSKKHAEEL-----	HVDV E SFKLLAKC F						
HBB_SQUAC/7-107	EKA LVNAV WTK-T-----	DHQ A VVAKA LER L F VVY PWT K T Y F V K F GKF H-----	-ASD ST VOT HAGK VVSALT VAY NHI-----	D-DVKP HF V ELSKKH Y EEL-----	HVDPENF KLLANCL							
HBB1_CYGMA/8-112	ELTI I INDIF SH-L-----	DYDDI GPKA LS RCL I VY PWT Q R Y F S F-----	GNLY NAEAI I GNAN VAAH GIKV LHGL RGL KN M-----	D-NIVDAY A ELS T LHS EKL-----	HVDP DNFKL LSDCI							
HBB1_XENBO/7-111	DROL IN STW GK-V-----	CAKT I GKEA L GRLL WTY PWT Q R Y F S F-----	GNL NS A DAV FHNE VAAH GEK VTS I GEA I KHM-----	D-DIKGY Y AQLSKY HS ET L-----	HVDPCNFKRFGGC L							
HBB_LITCT/1-105	GGSDVSAFLAK-V-----	DKRAVGGEA LAR L I VY PWT Q R Y F S F-----	GNL GS ADA I HNSK VLA H GQR V L D S-----	E E E GLKHP-----	Z-BLKAY YAKLS E RHSG E L-----	HVDPANFYRLGNV L						
HBB_LEPPA/7-111	EKOY I VS VFS K-I-----	DVDHV GANT LER L V I V FPWT K R Y F S F-----	GD LSS PG A I KHN NK VSA HGRK VLA A I E CTRHF-----	G-NIKGH LAN L S H L S EKL-----	HVDPHNFRVLGQCL							
HBB2_XENLA/8-112	EKAIA TS V WQK V-----	NVEHDG H D A L G R L L V Y PWT Q R Y F S F-----	GNL NS A A VAGNA Q V A H G K K V L S A V G N A I S H I-----	D-SVKSS L QQLSKSI HATEL-----	FVDPENF KRFGGV L							
HBB_ALLMI/7-111	ERKF I VDLWAK-V-----	DVAQCGA DALS RML I VY PWK R R Y F E H F-----	KGMC NAHD I L HNSK V Q E H G K K V L S F G E A V K H L-----	D-NIKGH FAN S K L H C E K F-----	HVDPENF KLLGD I							
HBB0_MOUSE/8-112	EKAIA TS I WDK V-----	DLEK VGG E T L G R L L I V Y PWT Q R F F D K F-----	GNL SS QA I MG N P R I K A H G K K V L T S L G L A V K N M-----	D-NL KET FAHLS E L H C D K L-----	HADPENF KLLGNM L							
HBBN_AMMLE/2-106	BKAL IT GFWS K-V-----	KVB ZV GAZ A L GRLL WTY PWT Z R F F D F-----	GBL SS A B A V M B B A V K V A H G K K V L B S F S B G L K H L-----	B-BLK GAF A S L Z L H C B K L-----	HVB PZB F R L L G B V L							
HBAZ_ LEPPA/7-108	DEV L I KEAW GL-L-H-----	QIPNAGGEA LARM FCS C PGT K S-----	PF H F GHD F S-----	ANNEK V KHHG K K V V D A I G Q G V Q H L-----	H-DLSSC LHT L S E K H A R E L-----	MVDP CNF Q Y L I E I						
HBA1_TORMA/6-107	NKKAI KNL L QK I IHS-----	QTE VL GAE A L A R L F E C H P Q T K S-----	PF H F GHD F S-----	ANND K R V K H H G A L V L K A V D T N K H L-----	D-DLPHH L N K L A E K H G K G L-----	LVDPHNFKLFDI						
HBA_SQUAC/6-107	DKTA I KHL T GS-L R T-----	-NAE A W G A E S L A R M F A T T P S T K T Y F S K F-----	TDF S-----	ANG R V K A H G G K V L N A V A D A T D H L-----	D-NVAGHLDPLA V L H G T T L-----	CVDPHNFP LLT Q C I						
HBA_HETPO/13-114	DRA E A L A L S K V -L A Q-----	-NAE A F G A E A L A R M F T V Y A A T K S-----	KF D Y-----	AAAPS I K A H G A K V T A L A K C D H L-----	D-DLK T H L K L A T F H G S E L-----	KVDPANFQYLSYCL						
GLB1_TYLHE/7-110	Q R I KV K Q Q W A Q-V Y S V-----	GESRT D F A I D V F N N F R T N P D R S-----	L F N R V N G D N V-----	Y S P E F K A H M R V F A G F D I L I S V-----	DDK P V L D Q A L A H Y A A F H K Q F G-----	T I P - F K A F Q G T M						
GLB4_LUMTE/11-120	DR RE I R H I W D D-V W S S-----	FT DRR V A I V R A F F D L F K H Y P T S K A L F E R V I D E P-----	-----	E S G E F K S H L V L R V A N G L D L I N L L-----	DDT L V L Q S H L G H L A D Q H I Q R K-----	GVT KEY FRG I G E A F						
GLB3_TYLHE/8-117	DRH E V L D N W K G -I W S A E-----	FT G R R V A I Q A I F Q E L F A L D P N A K G-----	F E V Y V D D-K-----	PSEADW KAH V I R V I N G L D L A V N L L-----	EDPK A L Q H K L A R Q H R E R S-----	GVK A V Y F D E M E K A L						
GLB4_TYLHE/8-117	DR RE V Q L A L W R S -I W S A E-----	FT G R R T L I G R L L F E E L F E I D G A T K G-----	L F K F V N V V D D T-----	PSEADW KAH V I R V I N G L D L A V N L L-----	EDPK A L Q H K L A R Q H R E R S-----	GVK A V Y F D E M E K A L						
GLB2_TYLHE/9-115	Q R L K V K Q Q W A K -A Y G V-----	GHER V E L G I A L W K S M F A Q D N D A R D-----	L F K F V N V G E D V-----	HSPA F E A H M A R V F N G L D R V I S S L-----	T D E P V L N A Q L E H L R Q Q H I K L G-----	I T G H M F N L M R T G L						
GLB2_LUMTE/8-114	EGL K V K S E W G R -A Y G S-----	GHD REA F S Q A I W R A T F A Q V P E S R S-----	L F K F V N V G E D D T-----	S H P A F I A H A E R V L G G L D I A S T L-----	D Q P A T L K E E L D H L Q V Q H E G R K-----	I P D N Y F D A F K T A I						
GLB_TUBTU/6-112	Q R F K V K H Q W A E -A F G T-----	S H H R L D F G L K L W N S I F R D A P E I R G L F-----	V G D N-----	A Y S A F E F A H A E R V L G G L D M T I S L L-----	D D Q Q A F D A Q L A H L K S Q H A E R N-----	I K A D Y Y G V F V N E L						
GLB3_LAMSP/7-113	Q R L K V K R Q W A E -A Y G S-----	G N D R E E F G H F I W T H V F K D A P S A R D-----	L F K F V N V G D N I-----	H T P A F R A H A T R V L G G L D M C I A L L-----	D D E G Y L N T Q L A H L A S Q H S S R G-----	V S A A Q Y D V V E H S V						
GLB_PAREP/8-117	Q D I L L K E L G P H -V-D T-----	-PAH I V E T G L G A Y H A L F T A H P Q Y I I H F R-----	-----	EG-H T I E N V M Q S E G I K H Y A R T L T E A I V H M L K E I-----	S N D A E V K K I A A Q Y G K D H T S R K-----	V T K D E F M S G E P I F						
Q21978_CAEEL/165-283	S C E V V A D S W R L -V E S R S S A E T S A C F G L F V F Q R V F S K I P M L R P L F-----	-----	-----	S E S D D V F D L P D N H P V R R H A R L F T S I L H I S V K N V-----	D E L E A Q V A P T V F K Y G E R H R P D I T P H M T E E N V R V F C A Q I	I						
GLB_PSED/C/21-134	T R E L C M K S L E H -A K V G T-----	-S K E A K Q D G I D LY K H M F E H Y P A M K K-----	F F K H R-----	E N Y T P A D V Q K D P F F I K Q G Q N I I L L A C H V L C A T Y-----	D D R E T F D A Y V G E L M A R H E R D H V-----	K I P N D V W N H F W E H F						
GLB_ACSVU/21-134	T R E L C M K S L E H -A K V D T-----	-S N E A R Q D G I D LY K H M F E H Y P P L R K Y-----	F F K H R-----	E E Y T A E D V Q N D P F F A K Q G Q K I I L L A C H V L C A T Y-----	D D R E T F D A Y V G E L M A R H E R D H V-----	H M P P E V W T D F W K L F						
GLBC_NIPBR/21-135	D V K -K H T V E S-M K A V P-----	W G R D K A Q N G I D F Y K F F T H H K D L R K F F K G A-----	-----	E N F G A D D V Q K S K R F E K Q G T A L L A V H V L A N V Y-----	D N Q A V F H G F V R E L M N R R H E K R G V D P K L W K I F F D D V W V P F							
GLBH_CAEEL/10-119	D L C -V K S L E G R -M V G T E-----	-A Q N I -E N G N A F Y R Y F F T N F P D L R V Y-----	F F K G A-----	E K Y T A D D V K K S E R F D K Q G Q R I I L L A C H L L A N V Y-----	T N E E V F K G Y V R E T I N R H R I Y K-----	M D P A L W M A F F T V F						
GLB2_NIPBR/16-114	P I S K A Q Q -----A Q-----	-V G K D F Y K F F F T N H P D L R K Y-----	F F K G A-----	E N F T A D D V Q K S D R F E K L G S G L L L S V H I I L A N T F-----	D N E D V F R A F C R E T I D R H V G R G-----	L D P A L W K A F W S V W						
GLBH_TRICO/30-132	D V V P L G S T P E K L-----	-E N G R E F Y K Y F F T N H Q D L R K Y-----	F F K G A-----	E T F T A D D I A K S D R F K K L G N Q L L L S V H I A A D T Y-----	D N E M I F R A F V R D T I D R H V D R G-----	L D P K L W K E F W S I Y						
Q20638_CAEEL/74-184	E K E L L R R T W S D-E F D-----	-N L Y E L G S A I Y C Y I F D H N P N C K Q L F P-----	F F K G A-----	I S K Y Q G D E W K S E F R S Q A L K F V Q T L A Q V V K N I Y H M E R T E S F L Y M V Q G K H V K F A D R G-----	F K H E Y W D I F Q D A M							
Q19601_CAEEL/105-215	E R I L L E Q S W R K -T R K-----	-T G A D H I G S K I F F M V L T A Q P D I K A I F G-----	L-----	E K I P T G R L K Y D P F R Q H A L V Y T K T L D F V I R N L-----	D Y P G K L E V Y F E N L G K R H V A M Q G-----	R G F E P G Y W E T F A E C M						
Q18311_CAEEL/32-140	T K K L V I Q E W P R -V L A-----	-Q C P E L F T E I W H K S A T R S T S I K L A F-----	I-----	A E -N-----	E S P M Q N A A F L G L S S T I Q A F F Y K L I I T Y E -L-----	N D D Q V R E A C E Q L G A R H V D F I S-----	R G F N S H F W D I F L V C M					

Family power

HBAZ_CAPH/7-107
HBA3_PLEWA/7-107
HBA4_CATCL/6-107
HBB4_HETPO/7-106
HBB5_SQUAC/7-107
HBB1_CYGMG/8-112
HBB1_XENBO/7-111
HBB2_LTCT/1-105
HBB2_LEPPA/7-111
HBB2_XENLA/8-112
HBB2_ALLMI/7-111
HBB2_MOUSE/8-112
HBBN_AMMLE/2-106
HBA4_LEPPA/7-108
HBA1_TORMA/6-107
HBA5_SQUAC/6-107
HBA4_HETPO/13-114
GBL1_TYLHE/7-110
GLB2_LUMTE/11-120
GLB3_TYLHE/8-117
GLB4_TYLHE/8-117
GLB2_TYLHE/9-115
GLB2_LUMTE/8-114
GLB2_TUBTU/6-112
GLB3_LAMSP/7-113
GLB4_PAREP/8-117
Q21978_CAEFL/165-283
GLB5_PSEDC/21-134
GLB6_ASCU/21-134
GLB7_NIPPR/21-135
GLB8_CAEFL/10-119
GLB2_NIPPR/16-114
GLB7_TRICO/30-132
Q20838_CAEFL/74-184
Q19601_CAEFL/105-215
Q18331_CAEFL/32-140

10 20 30 40 50 60 70 80 90 100 110 120

ERT I L S LWSK - IST -- QADVIGT ET LER L F SCY PQAKTY F PPNF - D LH - SGSQAQLRAHGSKVVAAGDAVKS I - D - N VTSALSKL SEL HAYVL -- RVDPVNFKFLSHCL
 EKA L VV GLCGK - ISG - HCDALGEALDRFL FAS FGQT RY F PNF - D LS - PGSDAVKRHGGKVL S A GEAHKH - D - SMDQALSKL SDL HAYNL -- RVDPGNFSKLHSCH
 DKADMVKIAWAK - ISP -- RADEI GAE ALRMLTVY PQT KTY F ANW ADLS - PGSGPVKHGKKVIMGAIDAVTKF - D - DLLGGLASL SEL HASKL -- RVDPSNFKLHANC
 ELHE ITT WKS I - - D KHSLGAKALARMFIVY PWTT RY F GNL I KEFT - ACSYGVKEAHAKVTGALGVATHL - G - DVKS QFT DLSK K HAEEL -- HVDESFKLLAKCF
 EKA L UNAWWT K T - - D HQAQVAKALERL F VVY PWT KTY F VFKM GKF H - - ASDSTVQTHAGKVMUSALT VAYNHI - D - DVKPHFVELSKK HYE L -- HVDPENFKLLANC
 ELTI INDIFSH L - - DDYD D GPKALRS RL CIVY PWT QRY F HFS G - GNLYNAEAIIGNANAAVHAGIKVHLGDLRGKLN M - D - NIVDAYAELSTLHS EKL -- HVDPDNFKL LSCDI
 DRQL IN STWKG V - - CAKT I GKE ALGR LLW TY PWT QRY F S F - GNL NSADAVFHNEA VAHAGK VVTS I GEA KHM - D - DIKGYA QALSKY HSE TL -- HVDP CNFKR FGGCL
 GGS D VSAFLAK M - - DKRAVGEGALR LLV I Y PWT QRY F S F - GNL GSADA I SHNS KVLA H GQR VLDS I EEGL KHP - - Z - BLKAYA QALSKR HS GEL -- HVDPANF YRLGNV
 EKQY IVS VFSK I - - D VDHVGANT LERLV LIV PWT KRY F S F - GDL S SPGAI KHN NKVSAH GRKV LAII ECTR HF - - G - NIKGH LANL SHL HS EKL -- HVDPHN F RL GQCL
 EKAA I TS VVWK V - - NVEHDGHD ALGR LLIVY PWT QRY F S F - GNL NS S AAVAGNAK VQAHG K KVLS A VGNAL SHI - D - SVKSS LQQL SKI HATEL -- FVDPENFKR FGGV
 ERK F VDLWAK M - - DVAQCGAD ALS RML I VY PWT KRY F S F - GKM CNAHD IL HNS KVQE HGKKV LAS F GEAV KHL - D - NIKGH FANL SKL CEKF -- HVDPENFKL LGGI
 EKAA I TS IWDK V - - D LKV GGET L GRLL I VY PWT QRF F S F - GNL S QA IMGP NR I KAHG K VVTL S GLAVK NM - D - NLKTFE FAHLS ELC DCKL -- HADPENFKL LGNM
 BKAL I TGFWSK V - - KV BZV GAZ ALGR LLV VY PWT ZR F F S F - GBL S SAB AVMB BAK V KAHG K KV LB S FB G LKHL - B - BLKG AFAS L SZ L HCBKL -- HVBPZB F RL LGBV
 DEV L KEAWGL L - H - QIPNAGGE ALRM SC PGTK S Y PFM F HDFS - - ANNEKVHHG KVV D A I QGVQVHL - H - DLSS CLHT L S EK HAREL -- MVDP CNF QY LIEA
 NKK A I KNLL QK I - HS - QT EVLG AE ALR L F ECH P QTK S Y PFM F HDFS - - AND KRV KHHG K V ALV DT N KHL - D - DLPH HLN K LAEK H GKG L -- LVDPHN FK L F SCDI
 DKA I KHL T GS - LRT - - NAE AWGAE S LARM F ATT P ST KTY F S F - T DFS - - ANG KRV KAHG K KV LN VANA VAT ADT H L - D - NVAGH LDPLA V L H GTTL -- CVDP HF PPL TQC
 DRA E AL ALS KV LAQ - - NAE AFGAE S LARM F ATT P ST KTY F S F - KDF T - - AAAPS I KAHG K VV T ALAKA C DHL - - D - DLK TH H K L A T F HG S E L -- KVDPANF QY LSY
 QRI K VQK QWQ AY VSY V - - GES RT D FAID VF N FRT N PDRS - - F R V N GDN V - - Y S P F EKAH MVR V FAG F D I L S V L - - DDK PVL DQLA QH YAAF HK QFG -- T I P - FKA G QTM
 DR RE I RH I WWD - VWS S S - FT DRR R V A I V R A F D D L K H Y P T S K A - F E V H I D E P - - E S G E F K S H L V R V A N G L D L I N L L - - DDT L V L Q S H L G H L A D Q H I Q R K -- GWT KEY FRG I GEF
 DR HE VLD NWKG - IWSAE - FT GRR R V A I G Q A I F Q E L F A L D P N A K G Y F R V N V D - - P S EADW KAH V I R V I N G L D L A V N L L - - EDP K A L Q E L K H L A R Q H E R S - - GV KAVY F DEME KAL
 DR RE V QAL WRS - IWSAE - DT G R R T L G R L I F F E L I D G T A K G L - F R V N V W D D T - - H S P E F A H V L R V V N G L D T I G V L - - G D S D T L N L S I D H L A E Q H K A R A - - G F K T V Y F K E F G K A L
 QRL K V K QWQ WA K - AY GV - - G H E R V E G L A W K S M F A Q D N D A R D - F R V H G E D V - - H S P A F E A H M A R V F N G L D R V I S S L - - T D E P V L N A Q L E H R Q O H K I L G - - I T C H M F N L M R T G L
 EGL K V K S E W G R - AY GS - - GHD REA F S Q A I W R A T F A Q V P E S R S - - F R V H G D D T - - S H P A F I A H A E R V L G G L D I A S T L - - D Q P A T K I E L D H L Q V Q H E G R K -- I P D N Y F D A F K T A I
 QR F K V K H Q W A E - AY GT - - S H R L D F G L L W N S I F R D A P E I R G L - F R V D G D - N - - A Y S A F E F A H A E R V L G G L D M T I S L L - - D D Q A F D Q A L Q H L K S Q A H E R N - - I K A D Y G V F V N E L
 QRL K V K R QW A E - AY GS - - GND REE F G H F I W T H V F K D A P S A R D - F R V R G D N I - - H T P A E R A H A T R V L G G L D M C I A L L - - D D E G V L N T Q L A H L A S Q H S S R G - - V S A A Q Y D V V E H S V
 QD I L K E L G P H - M D - - P A H I V E T G L G A Y H A L F T A H P Q Y I I H F S L R L - E G - H T I E N V M Q S E G I K H Y A R T L T E A I V H M L K E I - - S N D A E V K K I A A Q Y G K D H T S R K - - V T K D E F M S G E P I F
 S C E V W A D S W R L - M E S R S S A E T S A C F G L F F Q R V F S K I P M L R P I - F G L - E S D S D V F D L P D N H P V R R H A L F T S I L H I S V K N V - - D E L A E Q V A P T V F K Y G E R H Y R P D I T P H M E T E V N V R F C Q A I
 T R E L C M K S L E H - A K V G T - - S K E A K Q D G D I L Y K H M F E H Y P A M K K Y F K H R - - E N Y T P A D V Q K D P F F I L Q Q G N I L L A C H V L C A T Y - - D D R E T F D A Y V C E L M A R H E R D H V - - K I P N D V W H F W E H F
 T R E L C M K S L E H - A K V D T - - S N E A R Q D G D I L Y K H M F E H Y P A M K K Y F P P R K Y F K H R - - E E Y T A E D V Q N D P F F K A Q Q K Q K I L L A C H V L C A T Y - - D D R E T F N A Y T E R E L L D R H A R D H V - - H M P E F V W T D F W K L F
 D V K - - K H T V E S - M K V A P - - W G R D K A Q N G I D F Y K F F F T H K H D L R K F K G A - - E N F G A D D V Q K S R K F E Q K Q G T A L L L A V H V L A N V Y - - D N Q A V F H G F V R E L M N R H E K R G V D P L K W I K F I F D D V W V P
 D L C - M K S L E G R - M V G T E - - A Q N I - - E N G N A F Y R Y F F T N F P D L R V Y F K G A - - E K Y T A D D V K K S E R F D K Q Q G R I L L A C H L L A N V Y - - T N E E V F K G Y V R E T I N R H R I Y K - - M D P A L W M A F F T V F
 P I S K A Q Q - - - A Q - - - V G K D F K Y F F F T H N P D L R K Y F K G A - - E N F T A D D V Q K S D R F E K L G S G L L S V H I L A N T F - - D N E D V F R A F C R E T I D R H V G R G - - L D P L W K A F F W S V W
 D V V P L G S T P E K L - - - E N G R E F Y K F F T H N Q D L R K Y F K G A - - E T F T D A D I A K S R F E K L G S G L L S V H I L A N T F - - D N E M I F R A F V R D T I D R H V D R G - - L D P L K W K F E W S V
 Q20838_CAEFL/74-184 E K E L L R R T W S D - E F D - - N L Y E L G S A I Y C Y I F D H P N P N Q C K Q P P F F I L S K Y Q D E W K E S K E F R S Q A L K V Q T L A Q V V K N I Y H M E R T E S F L Y M V Q K H V K F D A R G - - F K H E Y W D I F Q D A M
 Q19601_CAEFL/105-215 E R I L L E Q S W K R - T R K - - T G A D H I G S V C F M V L T A Q P D I K A F G F L - - E K I P T G R L K Y D P R F Q H A L V Y T K T L D F V I R N L - - D P G K L E V Y F E P G Y W E T F A E C M
 Q18331_CAEFL/32-140 T K K L M Q S W K R - M V A - - Q C P C L E F T E I W H K S A T R S T S I K L A F G F L - - A E - N - E S P M Q N A F F L G L S S T Q I Q A F F Y K L I Y T E L - N D D Q V R E A C Q L G A R H V D F I S R G F N S H F W D I F L V C M

Family power

Human: 1 MGLSDGEWQLVLNWGKVEADIPGHGQEVLIRLFKGHPETLEKFIDIKFKHLKSEDEMKGASE 60
MGLSDGEWQLVLNWGKVEAD GHGQEVLIGHGQEVLI LFK HPETL KFDIKFKHLKSEDEMKGASE

Mouse: 1 MGLSDGEWQLVLNWGKVEADLAGHGQEVLI GLFKTHPETLDKFIDIKFKHLKSEEDMKGSE 60



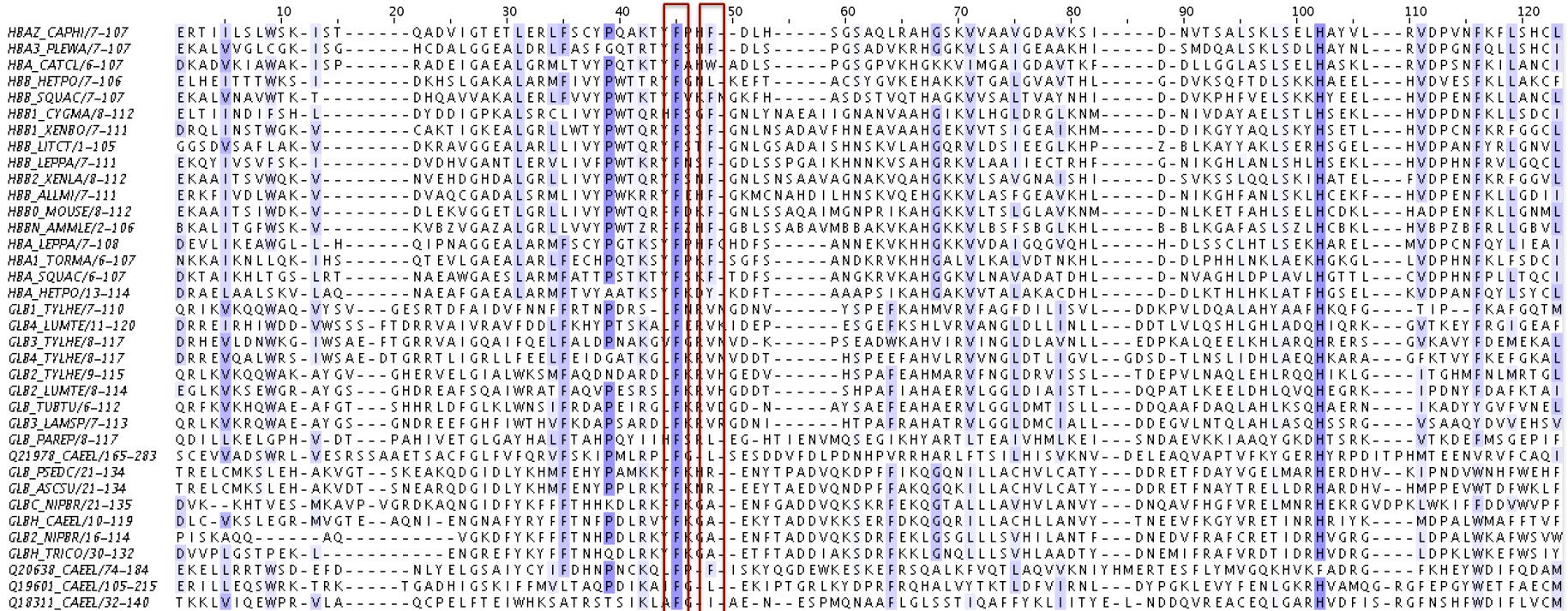
Human: 61 DLKKHGATVLTALGGILKKKGHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH 120
DLKKHG TVLTALG ILKKKG H AEI PLAQSHATKHKIPVKYLEFISE II VL H

Mouse: 61 DLKKHGCTVLTALGTILKKKGQHAAEIQPLAQSHATKHKIPVKYLEFISEIIIEVLKKRH 120

Human: 121 PGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154
GDFGADAQGAM KALELFR D A YKELGFQG

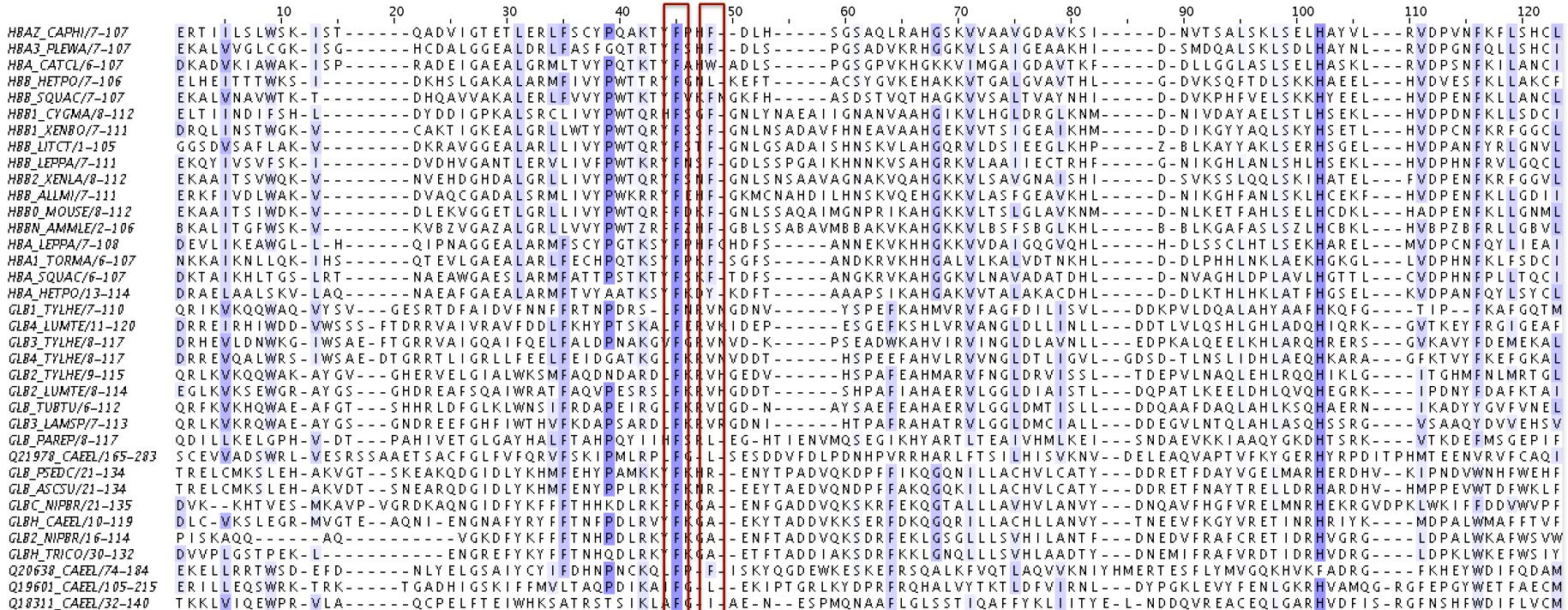
Mouse: 121 SGDFGADAQGAMSKALELFRNDIAAKYKELGFQG 154

Family power



seq> AWRTWEAPIFLKRYSTLPGGRAS...

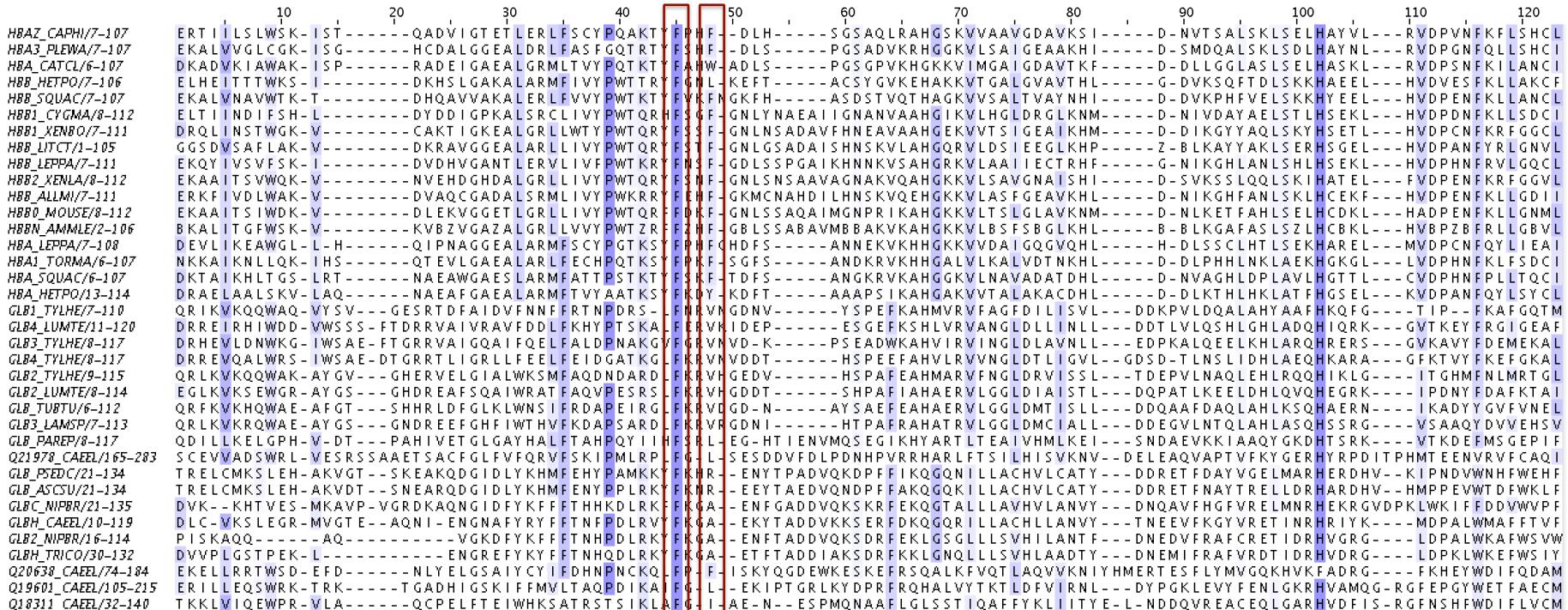
Family power



seq> AWRTWEAPIFLKRYSTLPGGRAS...

Family power

score(i, aa)



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



Structural domains from SCOP

Gene3D
Structural domains from CATH

No limits, domains

Marco Punta



No limits, full-length proteins

Marco Punta

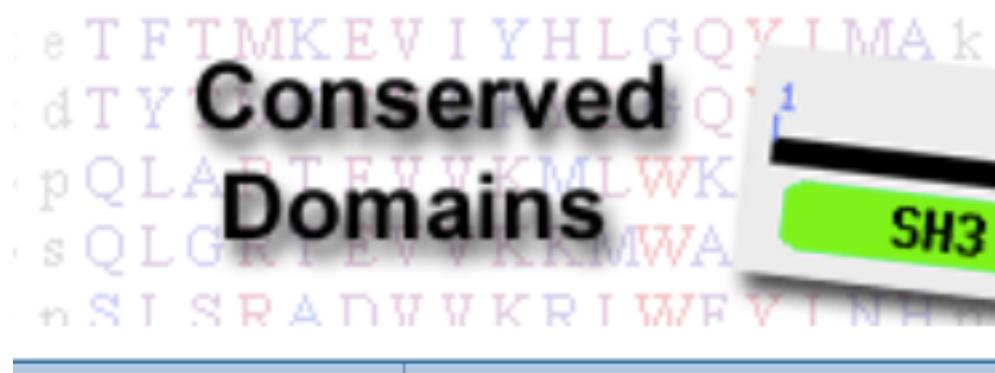


>7000 families, >50000 subfamilies



~2000 families

CDD



Uses RPS-BLAST

Integration

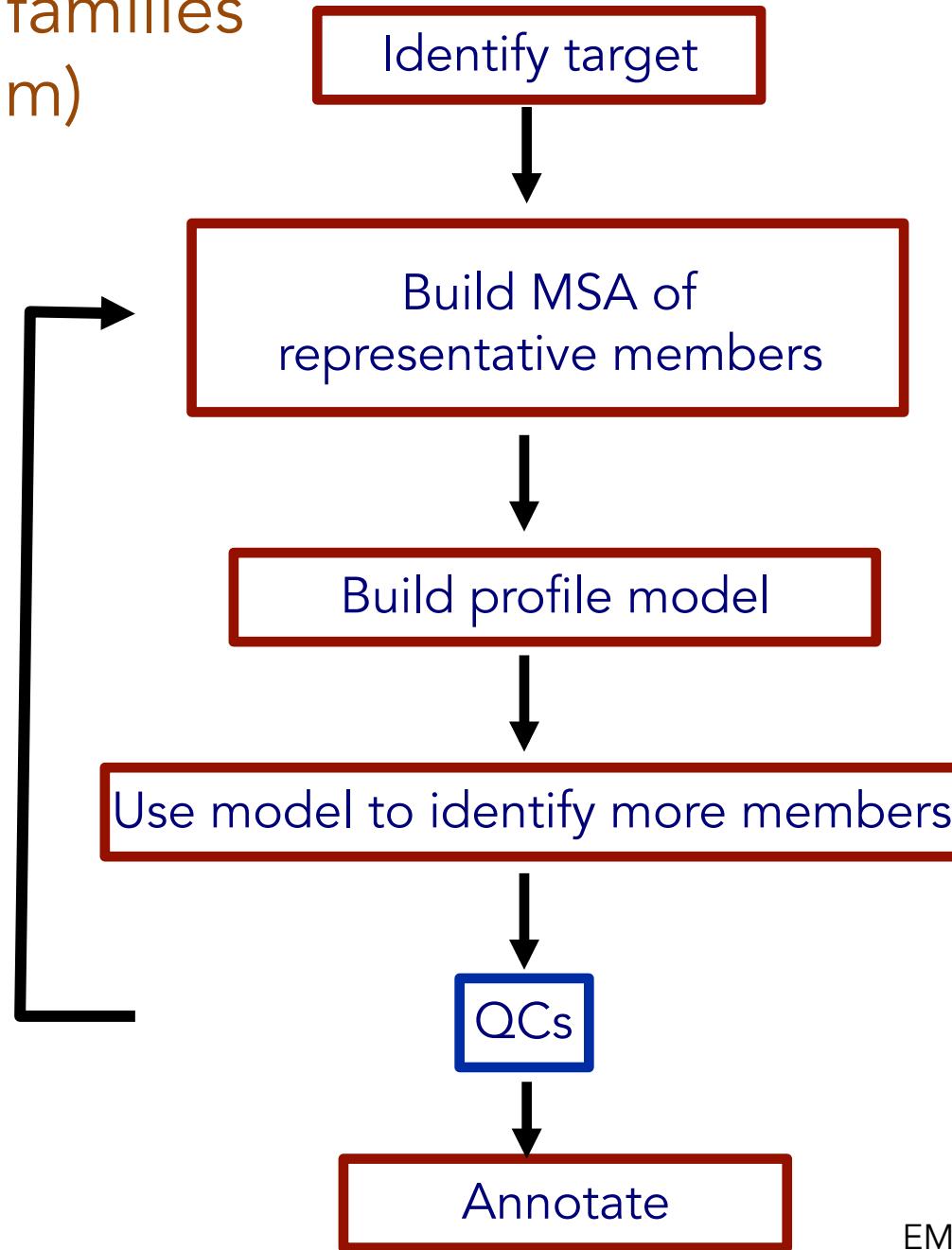
Marco Punta



Building a new protein family

Building families (Pfam)

Marco Punta



Is Pfam a database of globular domain families?

Marco Punta

<http://pfam.xfam.org>

EMBO Workshop, Budapest, 2016

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).

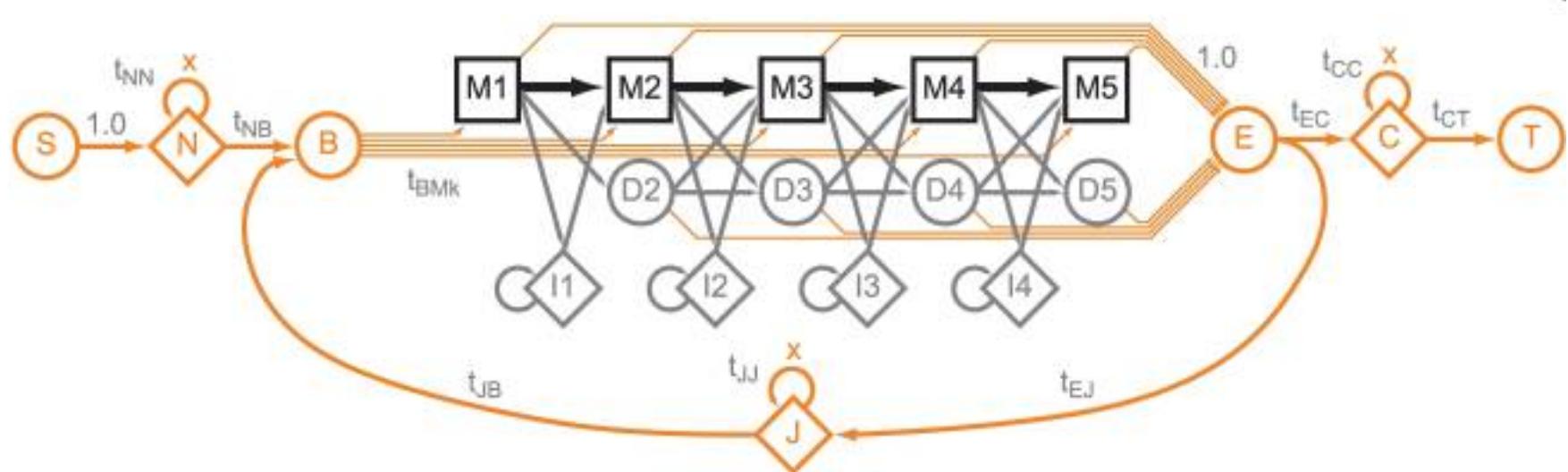
Team Exercise

Building a new Pfam family

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

HMMER

Profile-HMMs [Eddy Bioinformatics 1998]



Eddy PLOS CB 2011

EMBO Workshop, Budapest, 2016

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

↓

E_{ik}^M

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

T_{M->M}



1 → 2 3 4

A G E W

A A R W

A S D Y

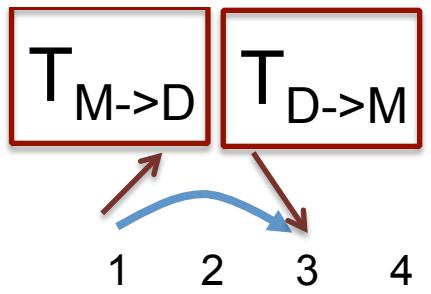
- G E W

L S E W

A S N W

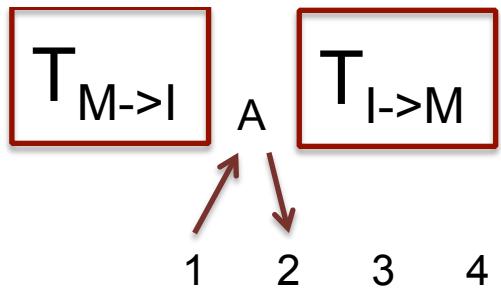
L G D W

A A D W



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

	A		
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

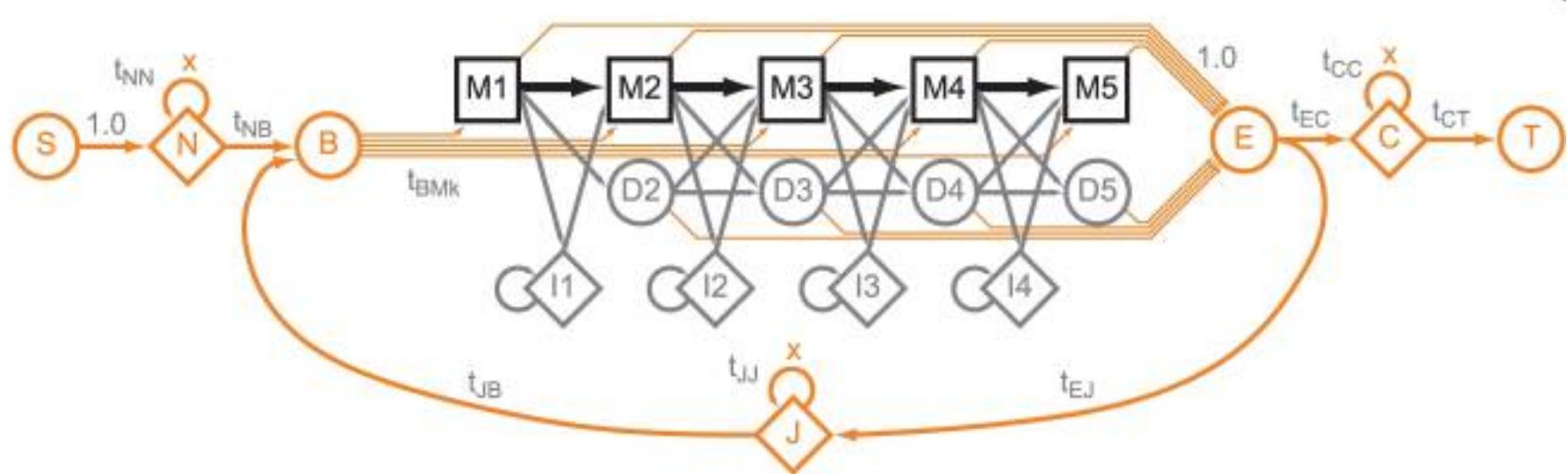


A	G	E	W
A	A	R	W
A	S	D	Y
-	G	E	W
L	S	E	W
A	S	N	W
L	G	D	W
A	A	D	W

Marco Punta

HMMER

Profile-HMMs [Eddy Bioinformatics 1998]



Eddy PLOS CB 2011

EMBO Workshop, Budapest, 2016

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](#) [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

► 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 @hmmer3r

<http://www.ebi.ac.uk/Tools/hmmer/search/phmm3r>
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/phmmers>
Finn et al. NAR 2015

EMBL-EBI

HMMER

Home Search Results Software Help About

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 ⓘ

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

[Home](#) | [Search](#) | [Results](#) | [Software](#) | [Help](#) | [About](#)phmmmer **hmmscan** hmmsearch jackhmmer

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
MPPMLSGLLARLVKLLLGRHGSALHWRAAGAAATVLLVIVLLAGSYLAVLAERGAPGAQLI
TYPRALWWWSVETATTVGYGDLYPTLWGRVLAVVVVMVAGITSGLVTAALATWFVGREQE
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/phmmmer>
Finn et al. NAR 2015

Home | **Search** | Results | Software | Help | Aboutphmmer hmmsearch **hmmsearch** jackhmmer

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-TATTNNWADFSSTWPTS-T-SEKPETDNWDAWAA----Q-----PSLTVP-SAGQVRQRSAFTPAT
V-TG-SSPSPVLGQGEKVEGL
>G3PBG4_GASAC/811-890 [subseq from] G3PBG4_GASAC
```

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

[Home](#) | [Search](#) | [Results](#) | [Software](#) | [Help](#) | [About](#)phmmmer hmmscan hmmsearch **jackhmmer**

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** **▼ Sequence Database**

Frequently used databases

- Reference Proteomes UniProtKB SwissProt PDB

Representative Sets (UniProt)

- rp75 rp55 rp35 rp15

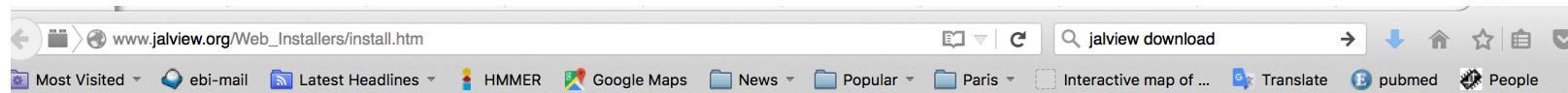
Other databases

- Ensembl (Plants) QfO

► Restrict by Taxonomy

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

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



Jalview

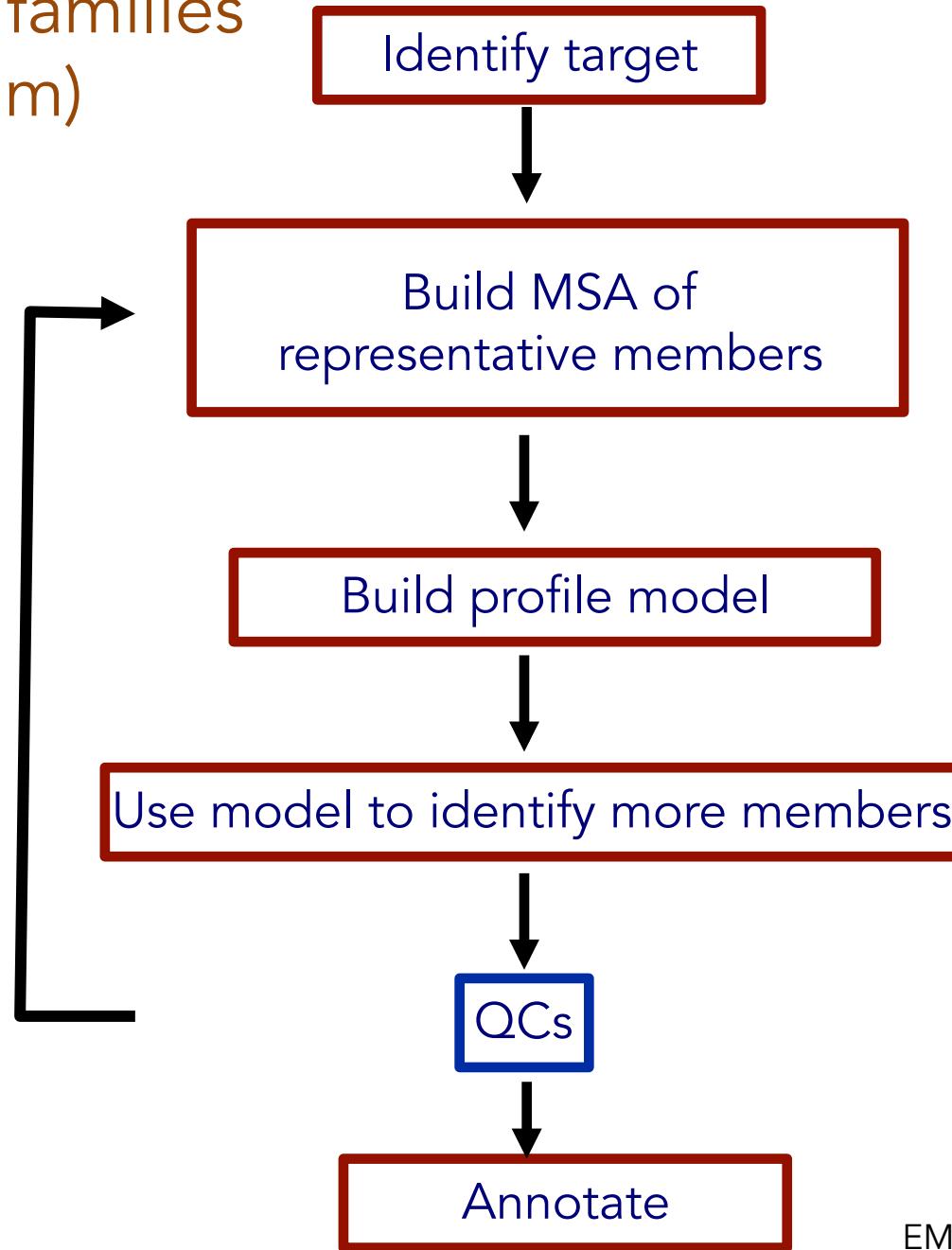


Available Installers

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

Building families (Pfam)

Marco Punta



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



Home | **Search** | Results | Software | Help | About



QUICK SEARCH

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

search against

- Reference Proteomes
- UniProtKB
- SwissProt
- Pfam

Submit

Reset

[Alternative Search Options](#)

Blog News

hmmer.org is updating

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

August, 2015

Download HMMER

Get the latest version

v3.1b2

Recent

[HMMER web server](#)

R.D. Finn, .

F. Schreiber

[Nucleic Acids Res.](#)

DOI: 10.1093/nar/gkv730



phmmmer

hmmscan

hmmsearch

jackhmmer

protein sequence vs profile-HMM database

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

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

```
HEAIGSGDLDLRSAFRRRTSLAGAGRRTSDSHEDAGTLDFSSLLKKRD  
SFRRDSDKLEAPAEEDVWEILRQAPPSEYERIAFQHGVTLRGMLKRL  
KGMKQDEKK
```

Submit

Reset

▼ HMM Database [?](#)

Protein Families

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

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



Search InterPro...

Search

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

[Search](#) | [Clear](#) Example protein sequence

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

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.

[Download](#) | [Read more](#)

Search >>



Learn more >>

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

EMBL-EBI

Services Research Training About us

InterPro

Protein sequence analysis & classification

Search InterPro... Examples: IPR020405, kinase, P51587, PF02932, GO:0007165 Search

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

```
HEAIGSGDLDLRSAFRRRTSLAGAGRRTSDSHEDAGTLDFSSLLKKRDSFRRDSKLEAPA  
EEDVWEILRQAPPSEYERIAFQHGVTDLRGMLKRLKGMIKQDEKK
```

Search | Clear Example protein sequence

Documentation Protein focus Publications

About InterPro: core concepts, The sweetest Nucleic Acids Research 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.

[Download](#) | [Read more](#)

IDA Domain architecture search [Search >>](#)

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

The screenshot shows the HMMER web interface. At the top, there's a dark header bar with the EMBL-EBI logo and links for Services, Research, Training, and About. Below the header is the HMMER logo and a navigation bar with Home, Search (which is selected), Results, Software, Help, and About. A sub-navigation bar below the main one includes phmmmer, **hmmscan**, hmmsearch, and jackhmmer. On the right side of the header are social media icons for Twitter and LinkedIn.

protein sequence vs profile-HMM database

Paste a Sequence | Upload a File | Accession Search

Paste in your sequence or use the example ⓘ

```
HEAIGSGDDLRSAFRRTSLAGAGRRTSDSHEDAGTLDFSSLLKKRDSFRRDSKLEAPAEEDVWEILRQAPPSEYERIAFOHGVTDLRGMLKRLKGKMKODEKK
```

Submit

▼ HMM Database ⓘ

Protein Families

Pfam TIGRFAM Gene3D Superfamily PIRSF
(select all)(clear all)

▼ Cut-Offs ⓘ

The screenshot shows the HMMER search interface. At the top, there's a dark header bar with the EMBL-EBI logo and links for Services, Research, Training, and About us. Below the header is the HMMER logo and a navigation bar with Home, Search (which is selected), Results, Software, Help, and About. On the far right of the navigation bar are social media icons for Twitter and GitHub. Underneath the navigation bar is a horizontal menu with four items: phmmmer, hmmscan, hmmsearch, and jackhmmer. The main content area has a title "protein sequence vs protein sequence database". Below it are links to Paste a Sequence, Upload a File, and Accession Search. A text input field contains a protein sequence starting with >protein region HEAIGSGDLDLRSAFRRTSLAGAGRRTSDSHEDAGTLDFSSLKKRDSFRRDSKLEAPAEEDVWEILRQAPPSEYERIAFQHGVTLRGMLKRLKGKQDEKK. Below the input field are two buttons: "Submit" and "Reset". Under the "Sequence Database" section, there are three boxes: "Frequently used databases" with radio buttons for Reference Proteomes (selected), UniProtKB, SwissProt, and PDB; "Representative Sets (UniProt)" with radio buttons for rp75, rp55, rp35, and rp15; and "Other databases" with radio buttons for Ensembl (Plants) and QfO.

EMBL-EBI 

HMMER

Services | Research | Training | About us

Home | Search | Results | Software | Help | About

Score Taxonomy Domain Download

PHMMER Results **Search Again**

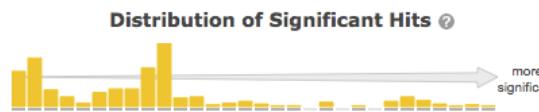
Sequence Matches and Features

Pfam 103
disorder 103
hit coverage
hit similarity

disorder coiled-coil tm & signal peptide

Show hit details

Distribution of Significant Hits



more significant

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	Gasterosteus aculeatus	3.0e-22				
> G3VYD1_SARHA	Sarcophilus harrisii	4.1e-22				
> F6PYQ5_ORNAN	Ornithorhynchus anatinus	5.9e-22				
> F6PYQ5_ORNAN	Ornithorhynchus anatinus	6.7e-22				
> H2V4W3_TAKRU	Takifugu rubripes	9.1e-22				
> G3XAI7_HUMAN	Homo sapiens	1.0e-21				
> F1Q615_DANRE	Danio rerio	1.3e-21				
> M7B6S6_CHEMY	Chelonia mydas	1.4e-21				
> F6Y6C7_HORSE	Equus caballus	1.4e-21				
> F6Y6F3_HORSE	Equus caballus	1.4e-21				
> G1R3C3_NOMLE	Nomascus leucogenys	2.0e-21				
> G3V1V7_HUMAN	Homo sapiens	2.1e-21				
> MYPC1_HUMAN	Homo sapiens	2.2e-21				
> MYPC1_HUMAN	Homo sapiens	2.2e-21				
> MYPC1_HUMAN	Homo sapiens	2.3e-21				
> MYPC1_HUMAN	Homo sapiens	2.3e-21				
> MYPC1_HUMAN	Homo sapiens	2.3e-21				
> MYPC1_HUMAN	Homo sapiens	2.3e-21				
> F8VZY0_HUMAN	Homo sapiens	2.4e-21				
> MYPC1_HUMAN	Homo sapiens	2.4e-21				
> MYPC1_HUMAN	Homo sapiens	2.4e-21				
> MYPC1_HUMAN	Homo sapiens	2.4e-21				
> G3RPA3_GORGO	Gorilla gorilla gorilla	2.7e-21				
> F7IK85_CALJA	Callithrix jacchus	2.9e-21				
> H2LN10_ORYLA	Oryzias latipes	2.9e-21				
> H2Q6Q1_PANTR	Pan troglodytes	3.7e-21				
> F7CWG3_CALJA	Callithrix jacchus	3.8e-21				
> HOW0Z3_CAVPO	Cavia porcellus	5.2e-21				
> HOVIL5_CAVPO	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
> A0A0Q3ZXK4_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 »

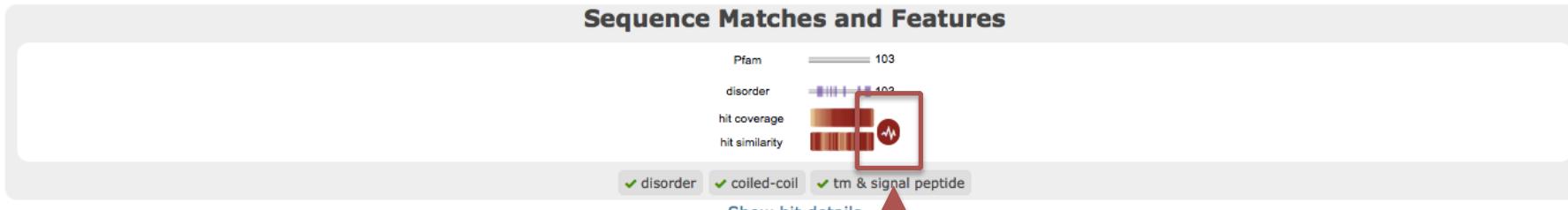
EMBL-EBI 

HMMER

Home | Search | Results | Software | Help | About | Services | Research | Training | About us |  | 

PHMMER Results | Score | Taxonomy | Domain | Download | **Search Again**

Sequence Matches and Features

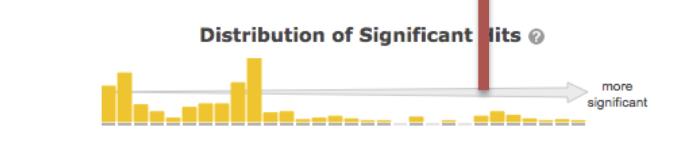


Pfam 103
disorder 103
hit coverage
hit similarity

✓ disorder ✓ coiled-coil ✓ tm & signal peptide

Show hit details

Distribution of Significant Hits

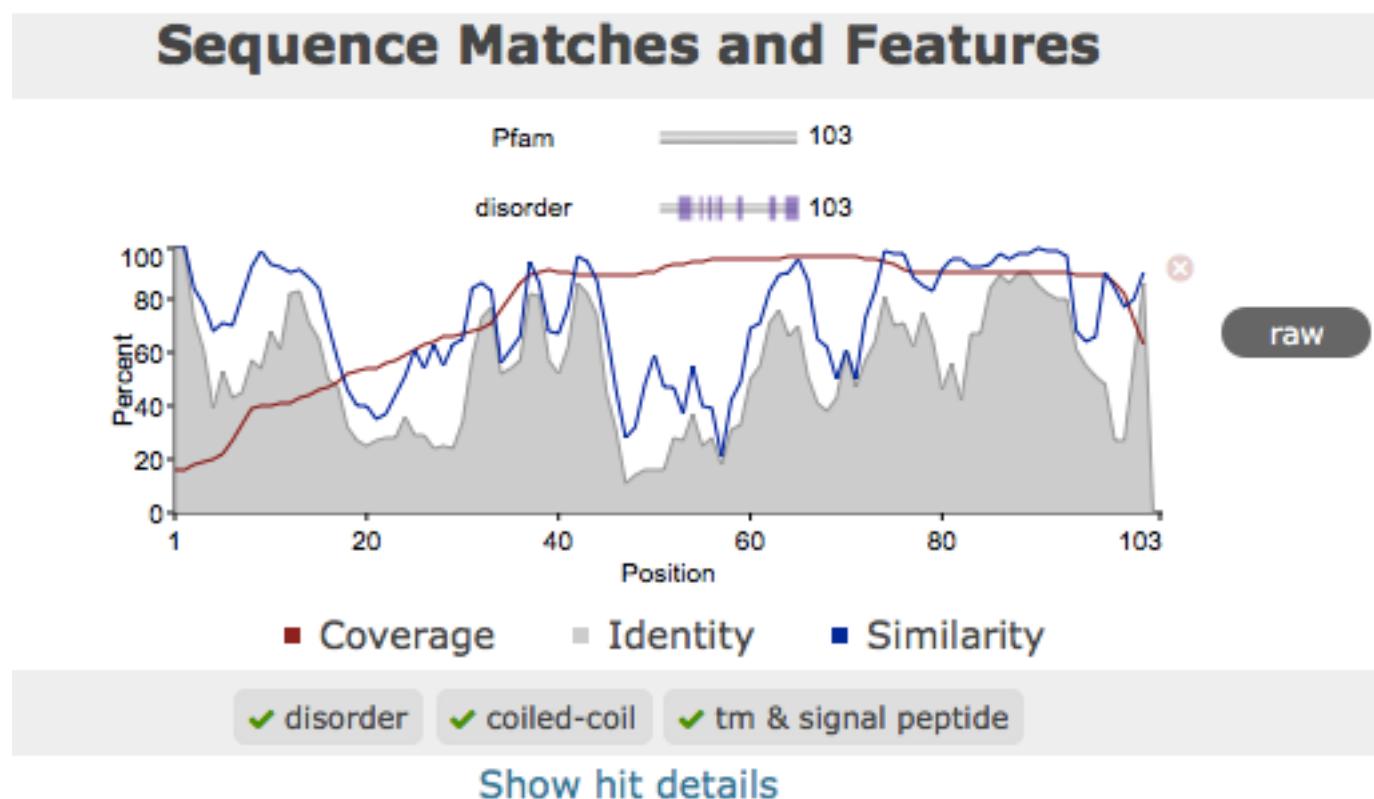


Bacteria Eukaryota Archaea Viruses Unclassified Sequences Other Sequences

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

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

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



EMBL-EBI HMMER

Home | Search | Results | Software | Help | About | Services | Research | Training | About us

Score Taxonomy Domain Download

PHMMER Results

Sequence Matches and Features

Pfam 103
disorder 103
hit coverage
bit similarity

✓ disorder ✓ coiled-coil ✓ tm & signal peptide

Show hit details

Distribution of Significant Hits

more significant

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

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

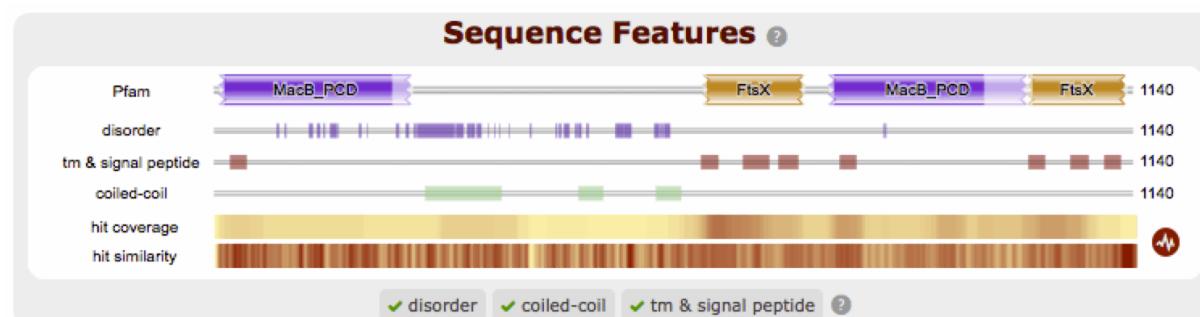
Significant Query Matches (376) in uniprotrefprot (v.2016-03-02)

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?



Search InterPro...

Examples: IPR020405, kinase, P51587, PF02932, GO:000

[Home](#) | [Search](#) | [Release notes](#) | [Download](#) | [About InterPro](#) | [Help](#) | [Contact](#)**Overview**[Similar proteins](#)
[Structures](#)**Filter view on****Entry type**

-  Family
-  Domains
-  Repeats
-  Site

Status

-  Unintegrated

Colour by

- 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

HMMER

Home | Search | Results | Software | Help | About

Score Download

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

Family 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

Family 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								
Accession	Description	E-value	Accession	Description	E-value		Start	End	Alignment		Model		Bit Score	Domain E-values	
Ind	Cond						Start	End	Start	End	Length	Ind	Cond		
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](#)

The screenshot shows the HMMER web interface. At the top, there is a navigation bar with links for EMBL-EBI, Services, Research, Training, About us, Home, Search, Results (which is the active tab), Software, Help, and About. A red arrow points down to the 'Domain' tab in the navigation bar, which is also highlighted with a red box.

The main content area is titled 'PHMMER Results'. It displays 'Sequence Matches and Features' with a summary table:

	Pfam	disorder	hit coverage	hit similarity
	103	103		

Below this are filters: disorder, coiled-coil, tm & signal peptide, and a 'Show hit details' button.

The next section is 'Distribution of Significant Hits' with a histogram showing the distribution of significant hits across different taxonomic groups. The legend includes: Bacteria (blue), Eukaryota (yellow), Archaea (green), Viruses (red), Unclassified Sequences (grey), and Other Sequences (black). An arrow indicates the direction of increasing significance.

At the bottom, there are navigation links: « First, « Previous, Page 1 of 5, Next », Last ».

The final section is 'Significant Query Matches (376) in uniprotrefprot (v.2016-03-02)'. It contains a table with columns: Target, Description, Species, and E-value. The table lists several entries, mostly from Mus musculus, with one entry from Rattus norvegicus and Mustela putorius furo.

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

HMMER

Home | Search | Results | Software | Help | About

Score Taxonomy Domain Download

PHMMER Results

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

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

Domain Architectures

119 SEQUENCES 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
Show All

Sequence Features  1113

45 SEQUENCES with domain architecture: **I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set, example:F6ZHP7_HORSE** View Scores
Show All

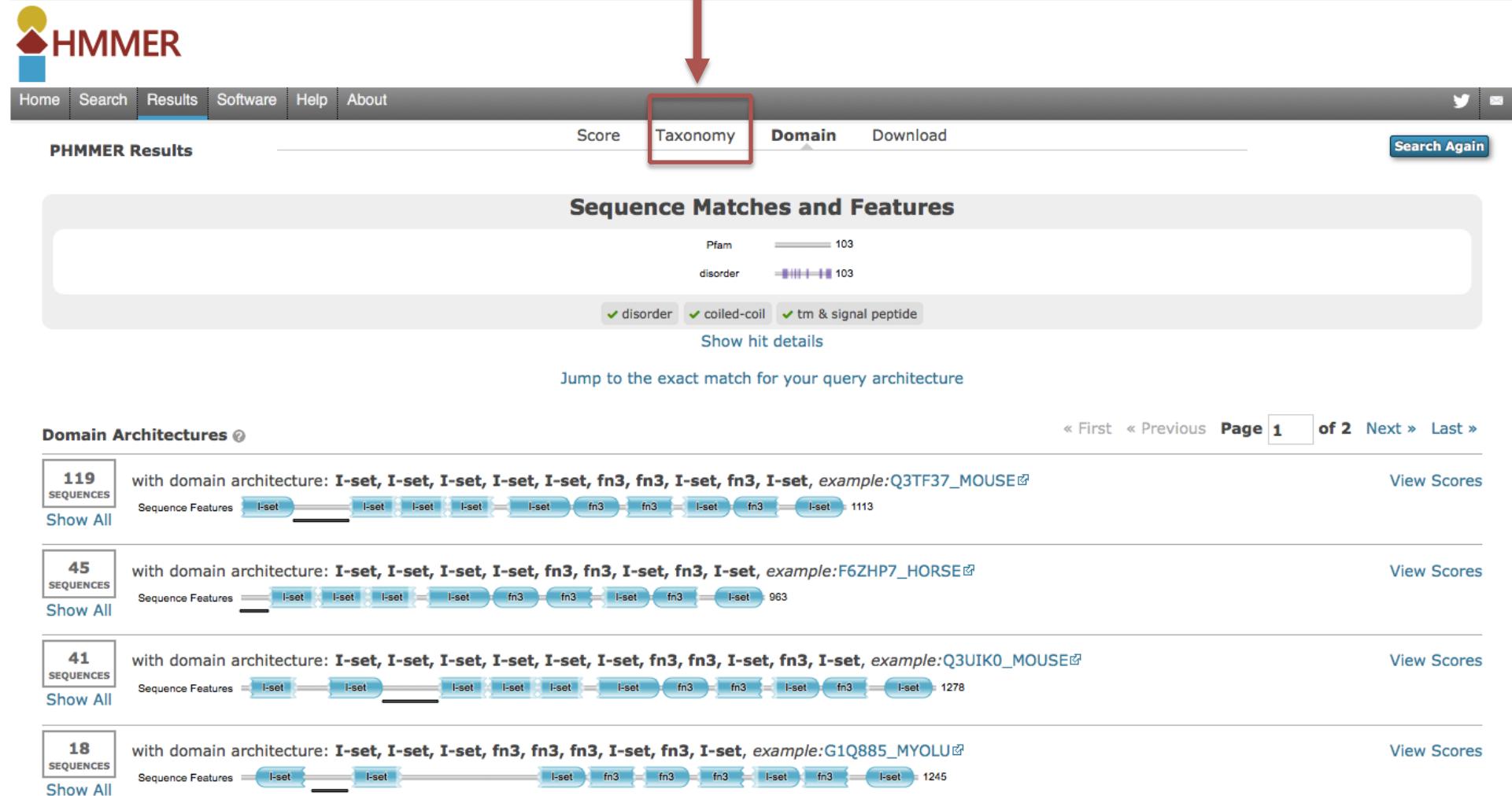
Sequence Features  963

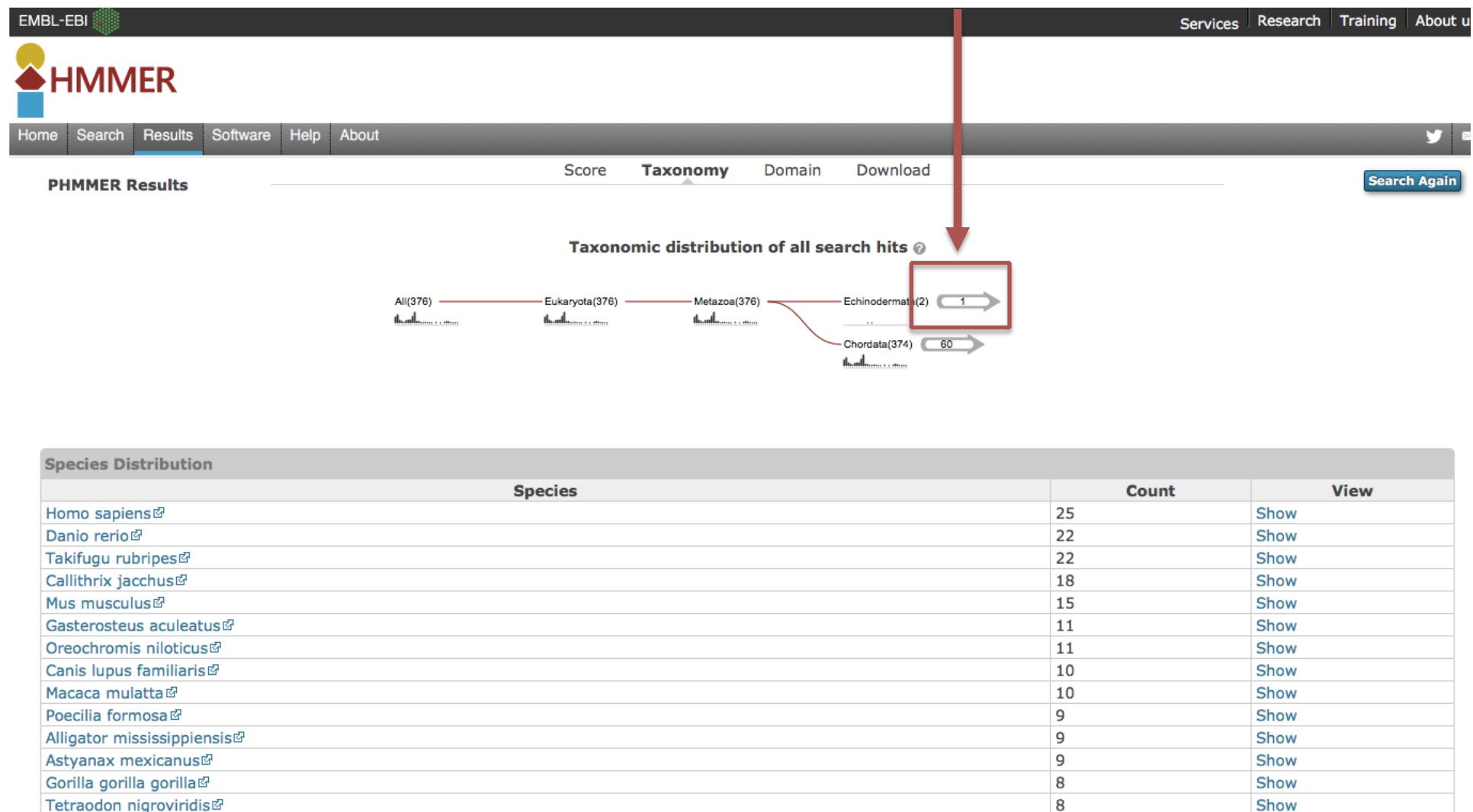
41 SEQUENCES with domain architecture: **I-set, I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set, example:Q3UIK0_MOUSE** View Scores
Show All

Sequence Features  1278

18 SEQUENCES with domain architecture: **I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set, example:G1Q885_MYOLU** View Scores
Show All

Sequence Features  1245





EMBL-EBI

Services | Research | Training | About us

HMMER

Home | Search | Results | Software | Help | About

Score | **Taxonomy** | Domain | Download

PHMMER Results | Search Again

Taxonomic distribution of all search hits

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

back Echinodermata(2) — Echinoidea(2) — Echinida(2) — Strongylocentrotidae(2) 1

Species Distribution

Species	Count	View
Strongylocentrotus purpuratus	2	Show Scores For All

Search Details

EMBL-EBI

Services Research Training About us

HMMER

Home Search Results Software Help About

Score Taxonomy Domain Download

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)

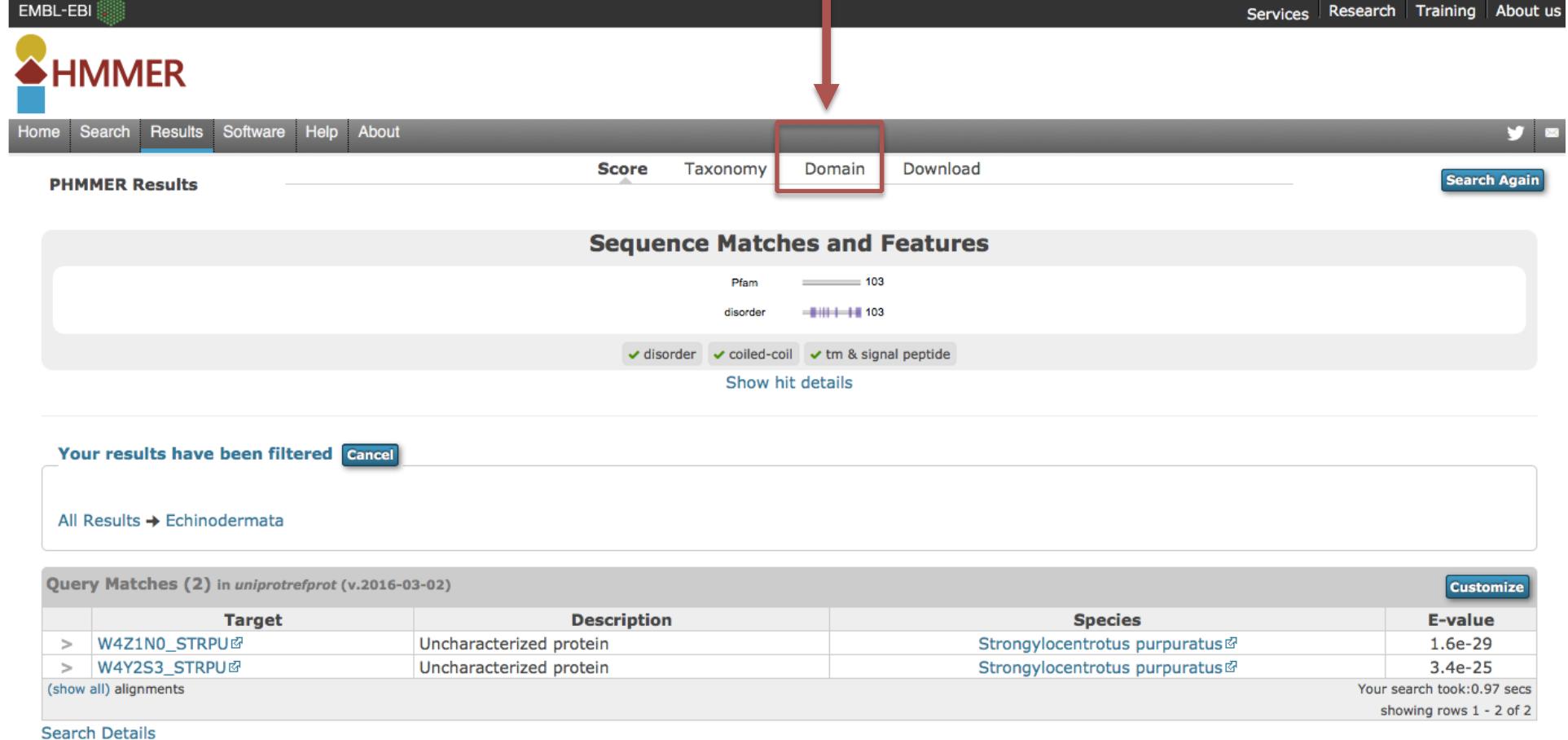
	Target	Description	Species	E-value
>	W4Z1N0_STRPU	Uncharacterized protein	Strongylocentrotus purpuratus	1.6e-29
>	W4Y2S3_STRPU	Uncharacterized protein	Strongylocentrotus purpuratus	3.4e-25

(show all) alignments

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

Customize

Search Details



The screenshot shows the HMMER web interface. At the top, there's a dark header with the EMBL-EBI logo, a search bar, and links for Services, Research, Training, and About us. Below the header is the HMMER logo and a navigation bar with Home, Search, Results (which is selected), Software, Help, and About. The main content area has tabs for Score, Taxonomy, Domain (which is highlighted with a red box and has a red arrow pointing to it from above), and Download. A 'PHMMER Results' section follows, containing a 'Sequence Matches and Features' summary. It shows Pfam and disorder annotations with their respective logos and lengths (103). Below this are checkboxes for disorder, coiled-coil, and tm & signal peptide, all of which are checked. A 'Show hit details' link is present. A message 'Your results have been filtered' with a 'Cancel' button is displayed. Under 'All Results → Echinodermata', there's a table titled 'Query Matches (2) in uniprotrefprot (v.2016-03-02)'. The table has columns for Target, Description, Species, and E-value. Two entries are shown: W4Z1N0_STRPU and W4Y2S3_STRPU, both described as Uncharacterized protein and belonging to Strongylocentrotus purpuratus. The E-values are 1.6e-29 and 3.4e-25 respectively. A 'Customize' button is at the top right of the table. A note at the bottom right says 'Your search took: 0.97 secs showing rows 1 - 2 of 2'. At the very bottom left is a 'Search Details' link.

Query Matches (2) in uniprotrefprot (v.2016-03-02)			
	Target	Description	Species
>	W4Z1N0_STRPU	Uncharacterized protein	Strongylocentrotus purpuratus
>	W4Y2S3_STRPU	Uncharacterized protein	Strongylocentrotus purpuratus

EMBL-EBI

Services Research Training About us

HMMER

Home Search Results Software Help About

Score Taxonomy Domain Download

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 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 I-set 3373

EMBL-EBI

Services Research Training About us

HMMER

Home Search Results Software Help About

PHMMER Results Score Taxonomy Domain Download

Your results have been filtered Cancel

All Results → Echinodermata

Search Again

Sequence Matches and Features

Pfam — 103
disorder — 103

✓ disorder ✓ coiled-coil ✓ tm & signal peptide

Show hit details

There were no exact architecture matches.

PHMMER Results

- **Job:** 11F56758-277F-11E6-9D89-A9DDD26C98AD.1
- **Started:** 2016-05-31 23:29:00
- **Algorithm:** phmmr
- **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 formated 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!

EMBL-EBI

HMMER

Home | Search | Results | Software | Help | About

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 ?

```
HEAIGSGDLDLRSAFRRTSLAGAGRRTSDSHEDAGTLDFSSLLKKRDSFRRDSKLEAPAEEDVWEILRQAPPSEYERIAFOHGVTDLRGMLKRLKGKODEKK
```

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 ?



EMBL-EBI

HMMER

Home Search Results Software Help About

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 ⓘ

```
HEAIGSGDDLRSAFRRTSLAGAGRRTSDSHEDAGTLDFSSLLKKRDSFRRDSKLEAPAEEDVWEILRQAPPSEYERIAFQHGVTDLRGMLKRLKGKODEKK
```

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 ⓘ



Home | Search | Results | Software | Help | About

Score Taxonomy Domain Download

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

more significant

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

Significant Query Matches (14) in swissprot (v.2016-03-02)

Target	Description	Species	E-value
> MYPC_RAT	Myosin-binding protein C, cardiac-type	Rattus norvegicus	1.0e-57
> MYPC3_MOUSE	Myosin-binding protein C, cardiac-type	Mus musculus	1.5e-55
> MYPC3_HUMAN	Myosin-binding protein C, cardiac-type	Homo sapiens	2.5e-53
> MYPC3_CHICK	Myosin-binding protein C, cardiac-type	Gallus gallus	8.6e-33
> MYPC1_HUMAN	Myosin-binding protein C, slow-type	Homo sapiens	6.1e-23
> MYPC3_AMBME	Myosin-binding protein C, cardiac-type (Fragment)	Ambystoma mexicanum	1.6e-21
> MYPC2_MOUSE	Myosin-binding protein C, fast-type	Mus musculus	1.5e-16
> MYPC2_HUMAN	Myosin-binding protein C, fast-type	Homo sapiens	7.5e-16
> MYPC2_CHICK	Myosin-binding protein C, fast-type	Gallus gallus	1.6e-14
> IGFN1_HUMAN	Immunoglobulin-like and fibronectin type III domain-containing protein 1	Homo sapiens	1.0e-07
> IGFN1_MOUSE	Immunoglobulin-like and fibronectin type III domain-containing protein 1	Mus musculus	1.2e-07
> IGS22_HUMAN	Immunoglobulin superfamily member 22	Homo sapiens	5.0e-07

UniProt

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

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
:	11.00	11.00				

UniProt

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

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
:	110	110				

UniProt

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

Protein | Myosin-binding protein C, cardiac-type

Gene | Mybpc3

Organism | Rattus norvegicus (Rat)

Status | Reviewed - Annotation score: 5/5 - 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
:	110	110		110		

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: UniProtKB-KW
- regulation of striated muscle contraction Source: UniProtKB-KW
- ventricular cardiac muscle contraction Source: UniProtKB-KW

Inferred from expression patternⁱ

PubMed 8799143

Complete GO annotation...

Keywords - Molecular functionⁱ

Muscle protein

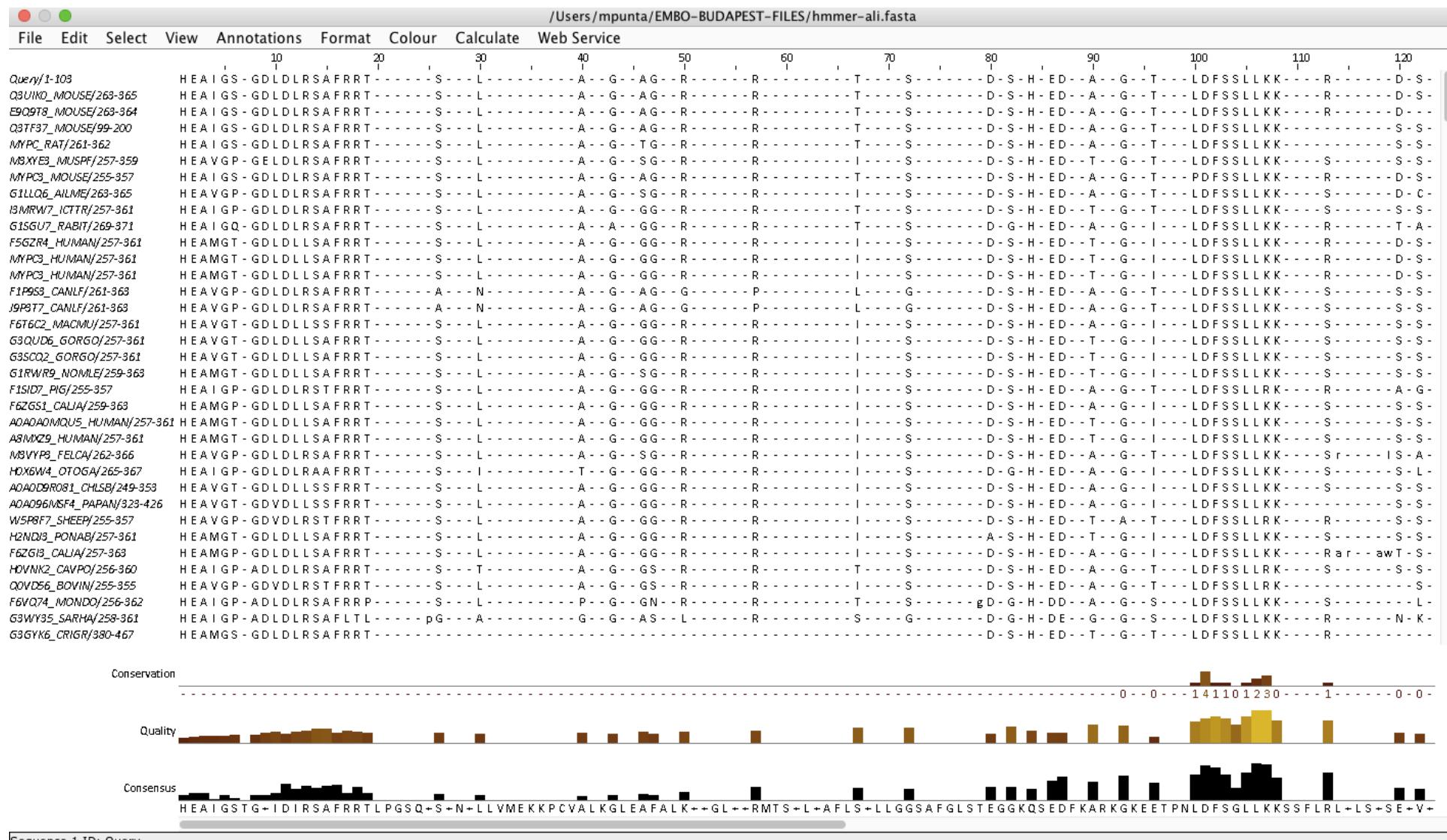
Keywords - Biological processⁱ

Cell adhesion

Keywords - Ligandⁱ

Start Jalview

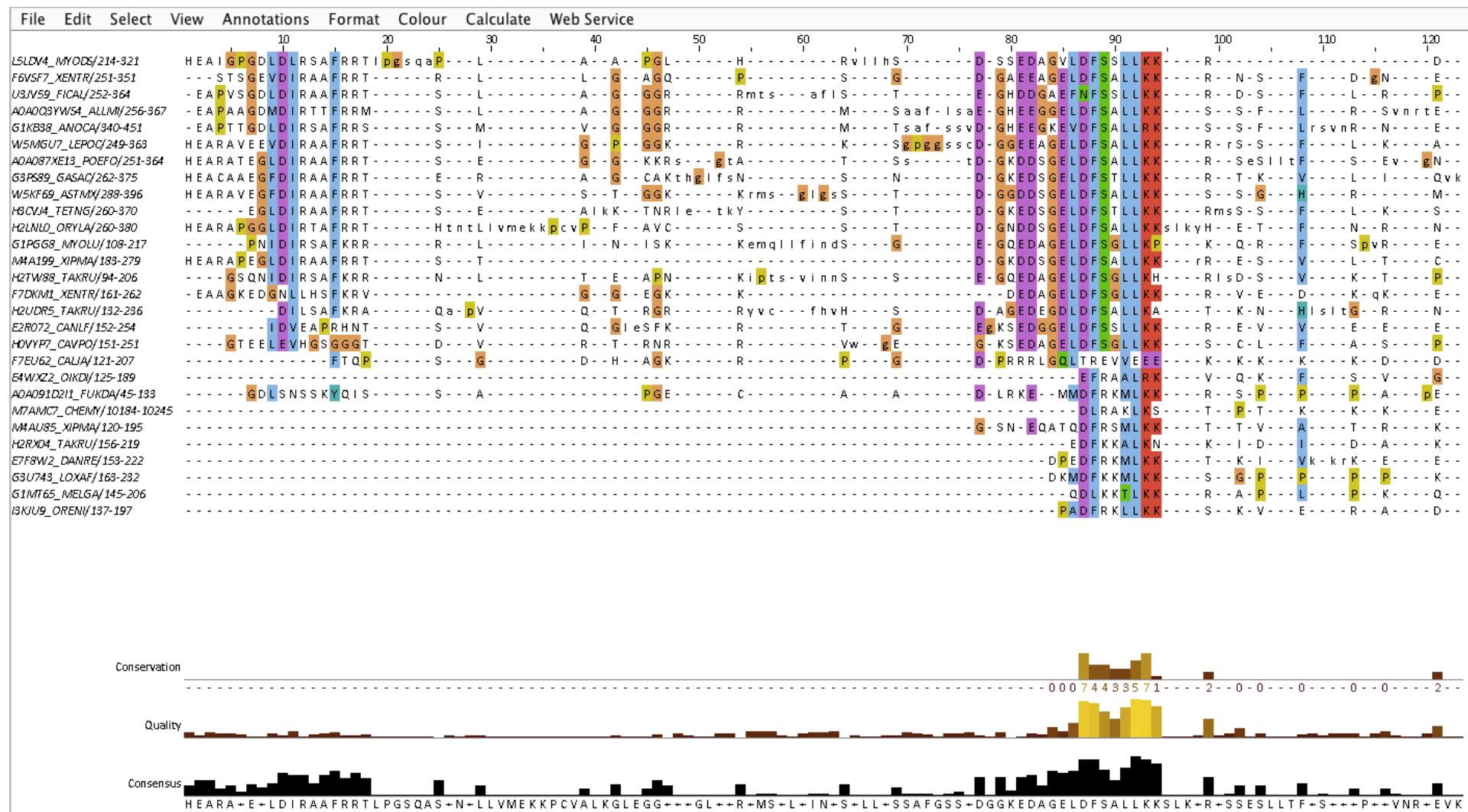
1. File -> Input alignment -> From file



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

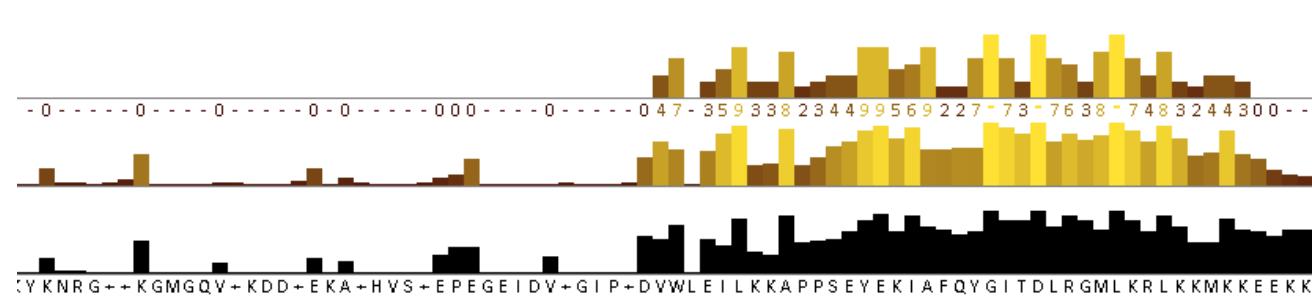
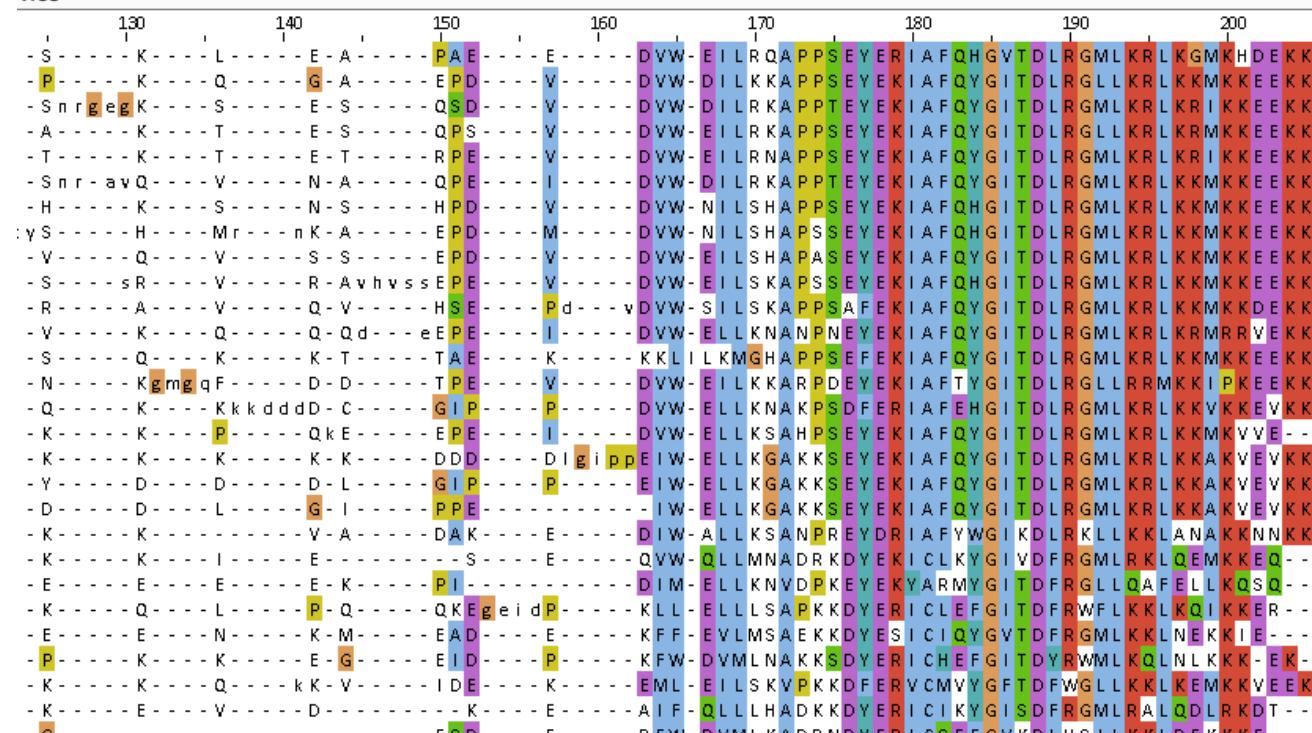
1. Edit -> Remove redundancy
2. Select 90% and Remove
3. Edit -> Remove empty columns

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



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

1. Colour -> BLOSUM62 Score

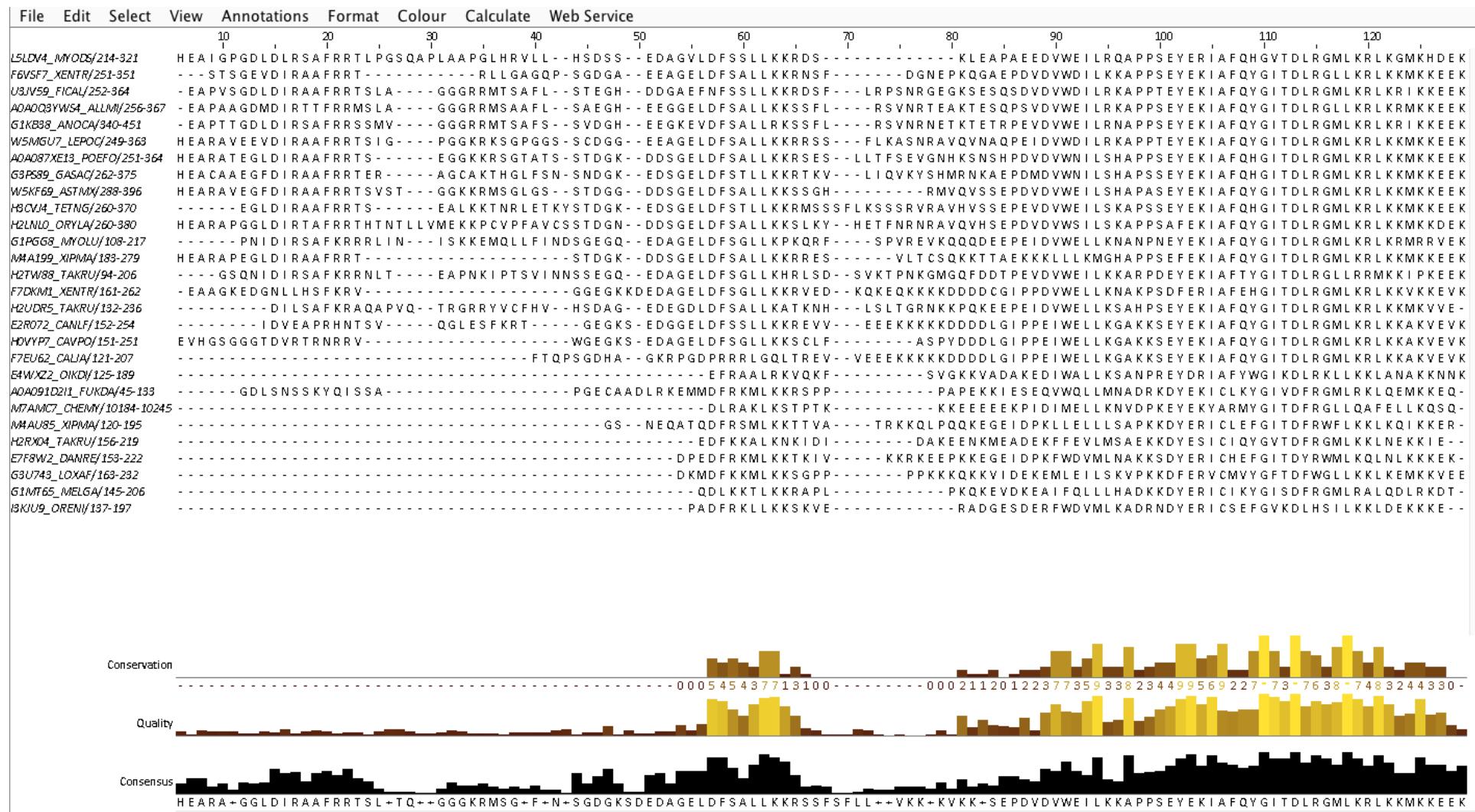
Colour -> Percentage Identity

Marco Punta

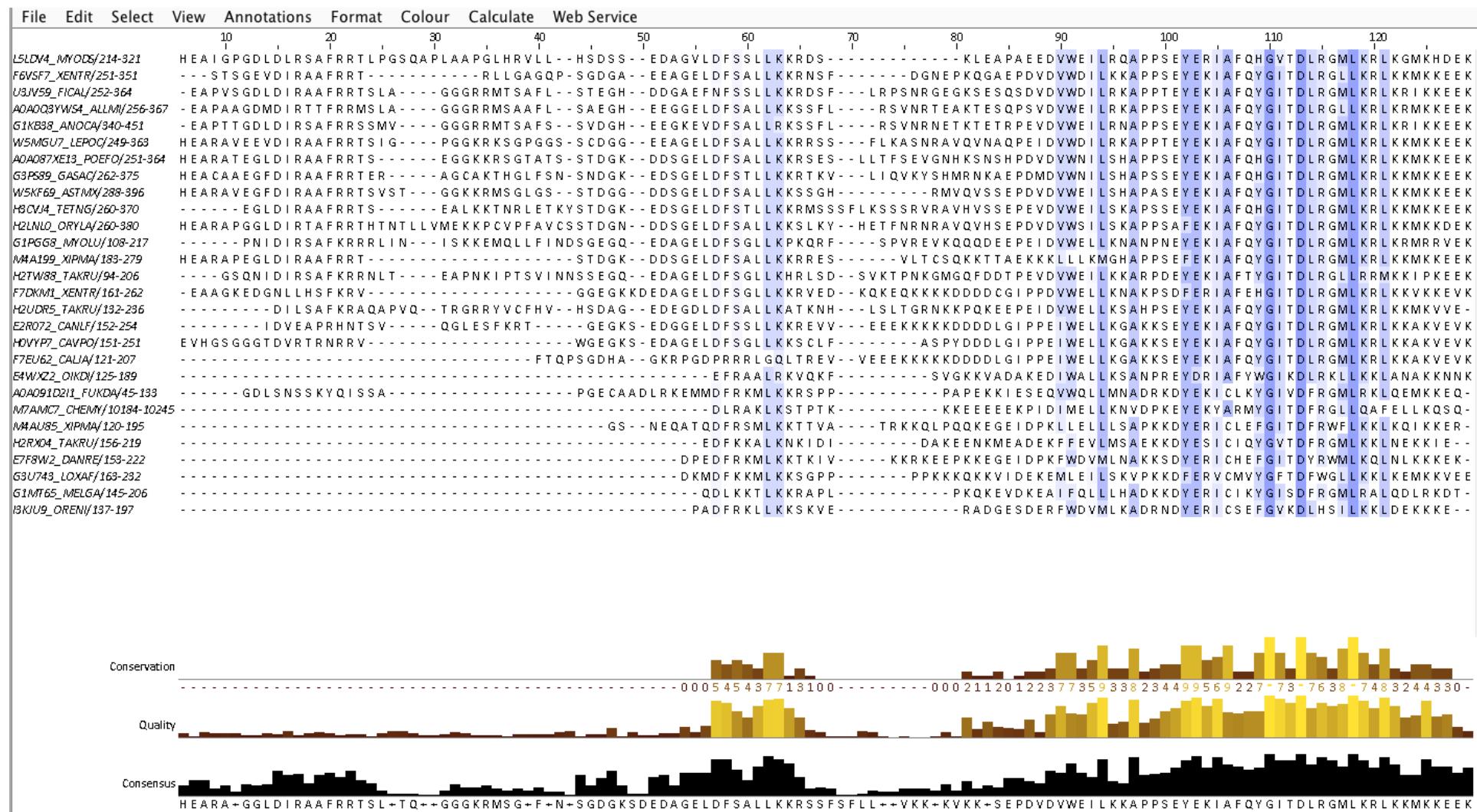
EMBO Workshop, Budapest, 2016

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

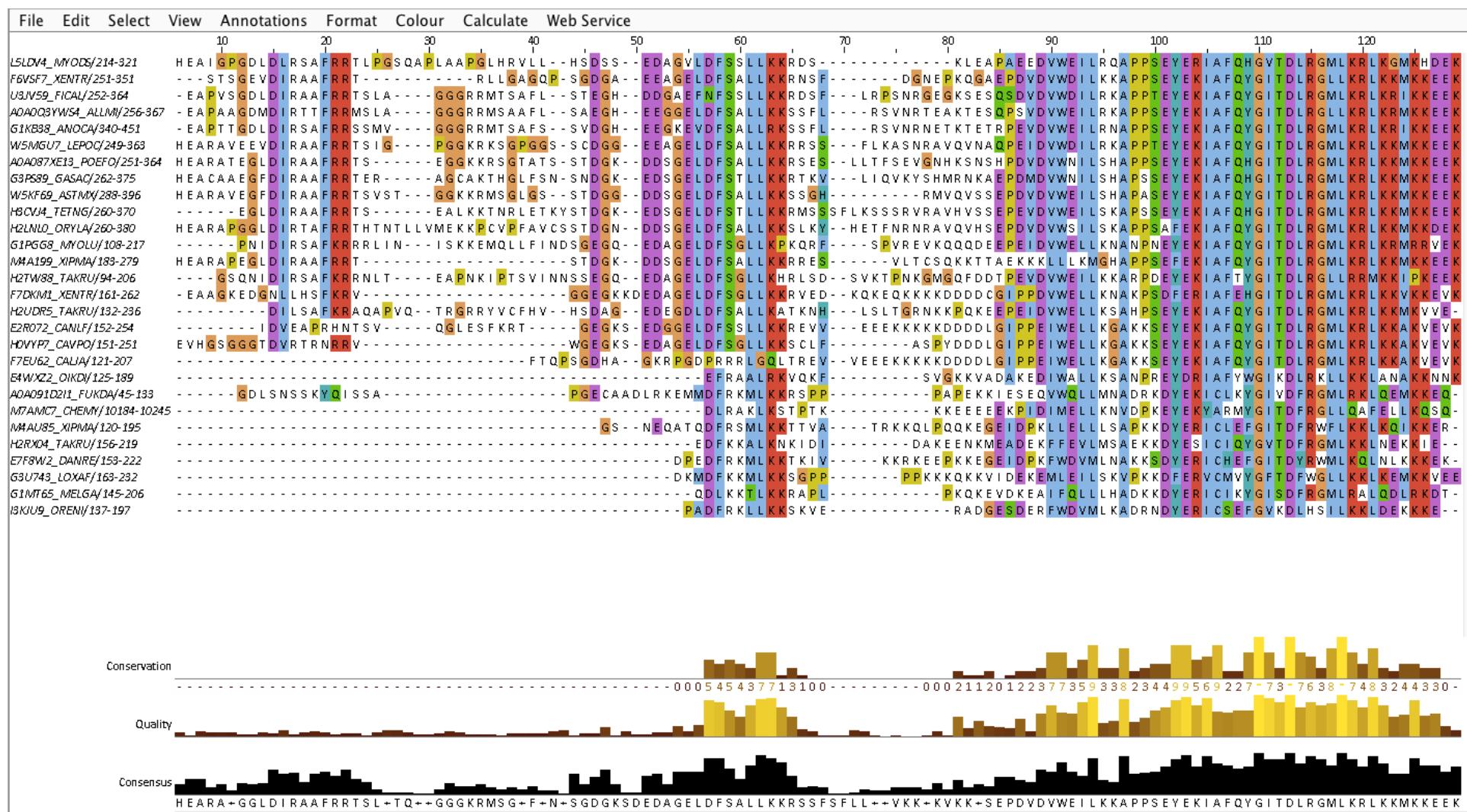
1. Web Service -> Alignment-> Run Muscle with Default



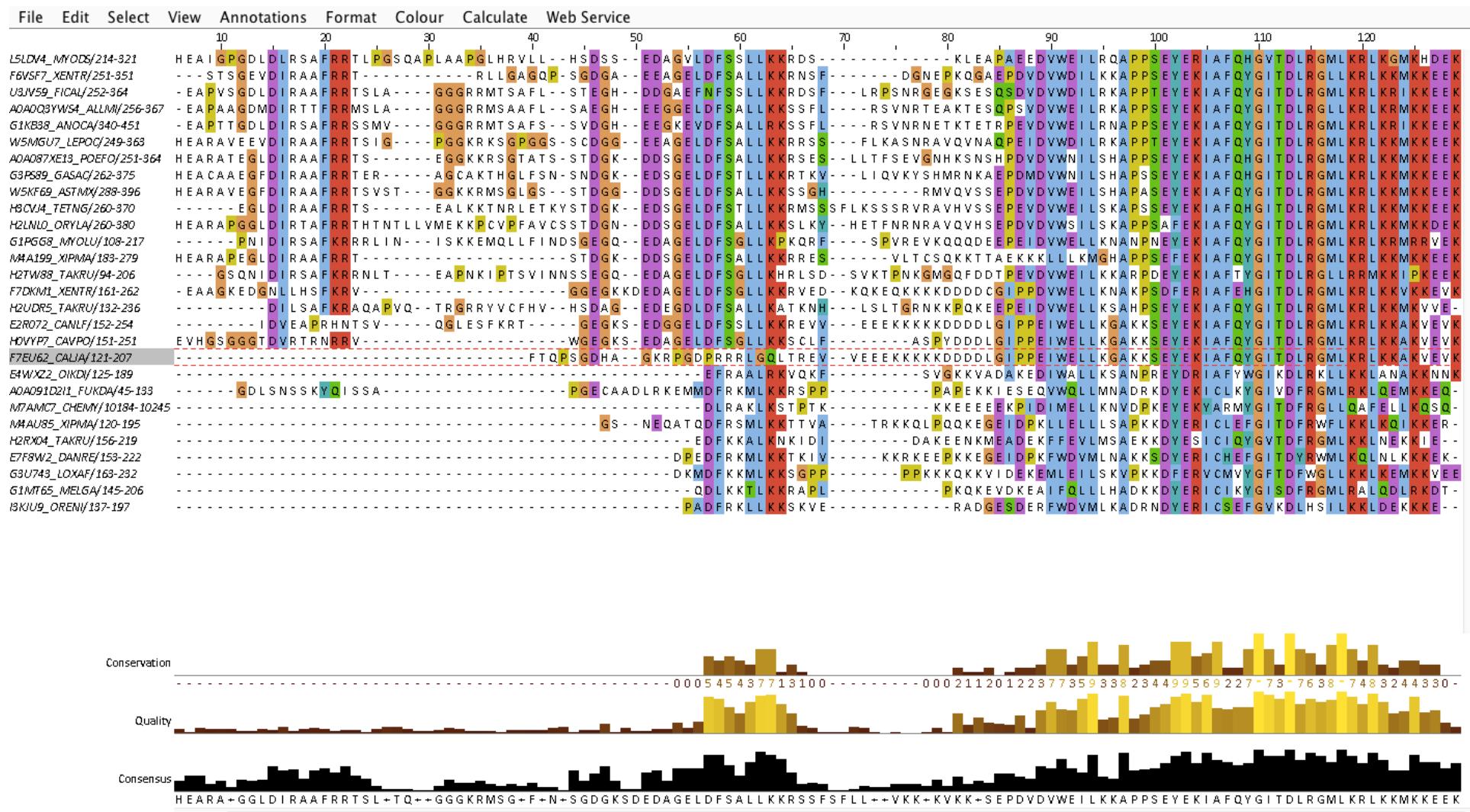
1. Colour -> BLOSUM62 Score



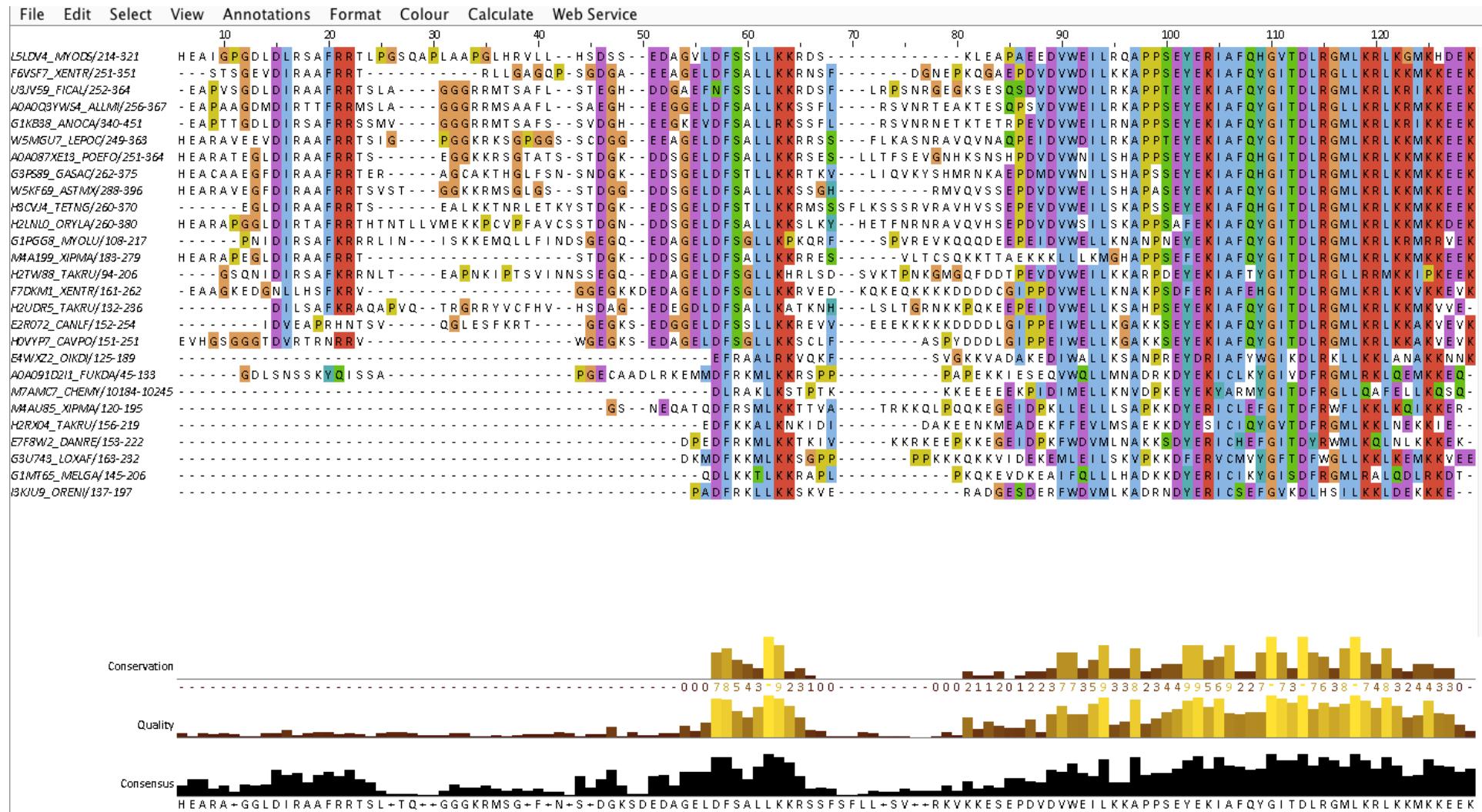
1. Colour -> Clustalx



1. Click on F7EU62_CALJA



1. Edit -> Delete

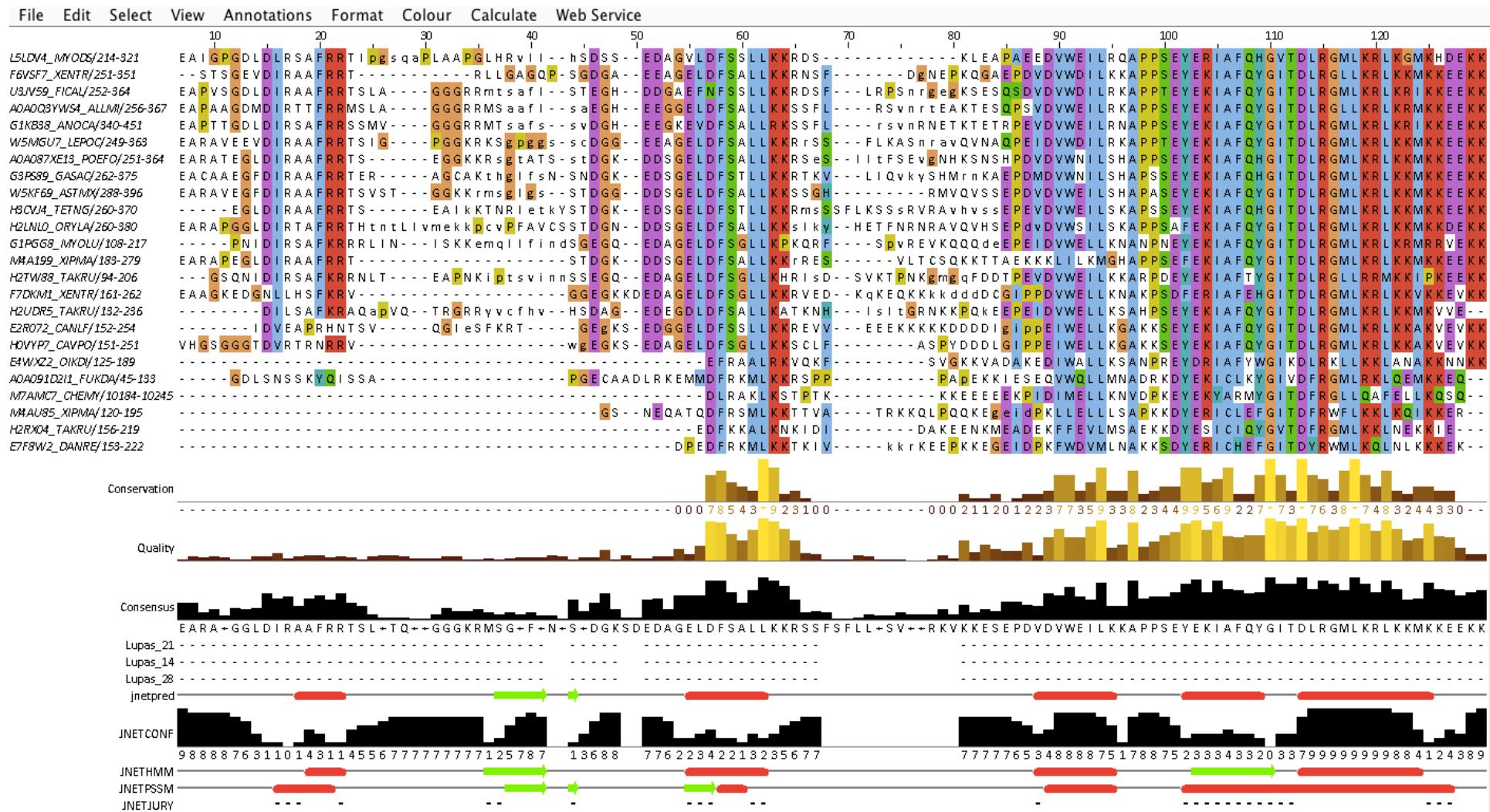


1. Select -> Select all

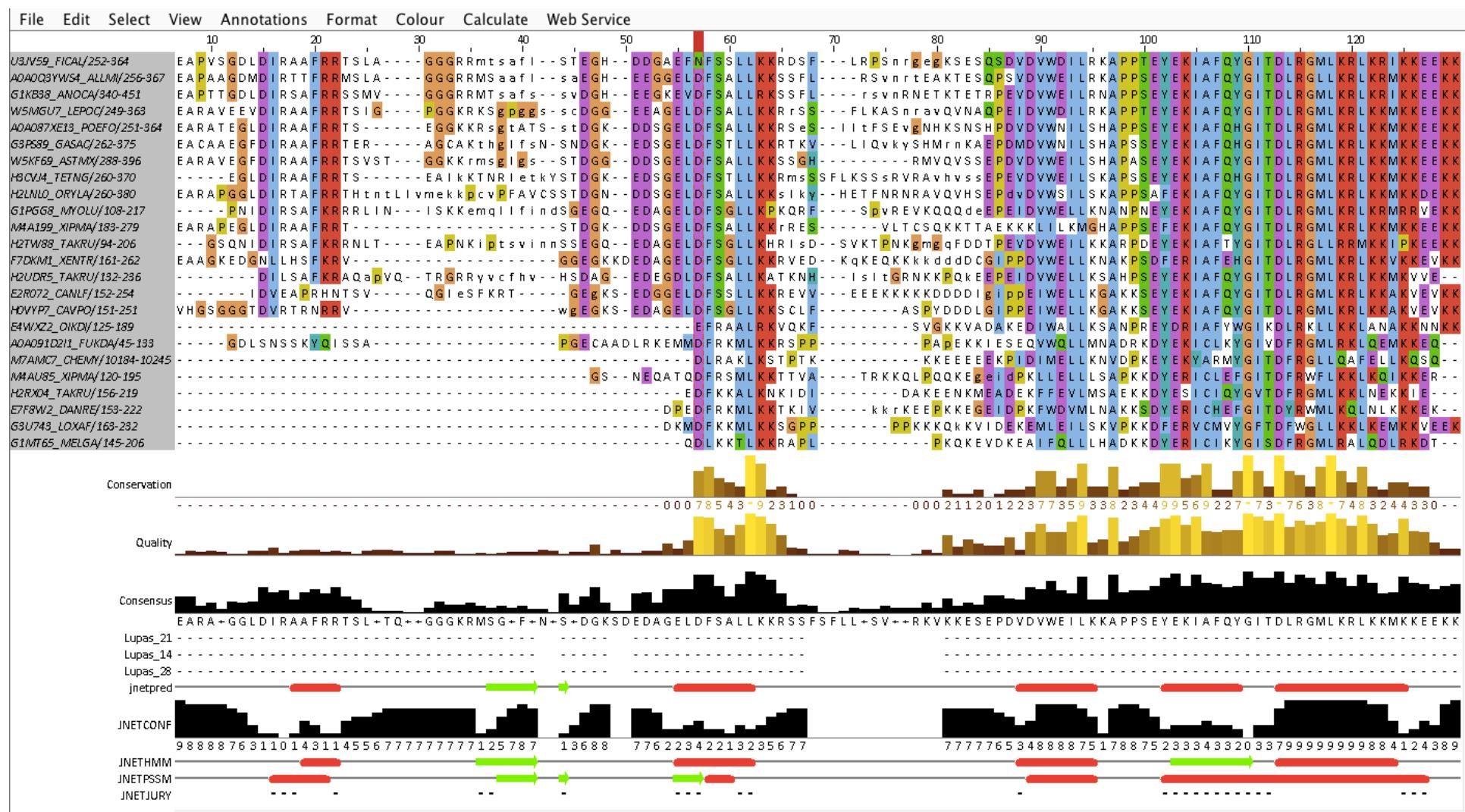
2. Web Service -> Secondary Structure Prediction-> JNet Secondary Structure Prediction

1. Select -> Select all

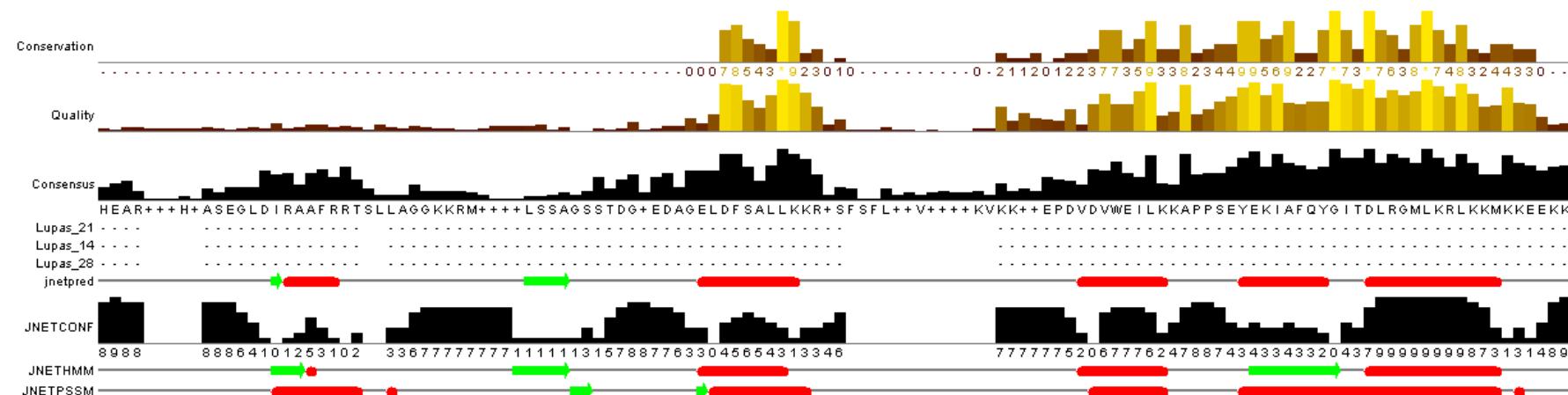
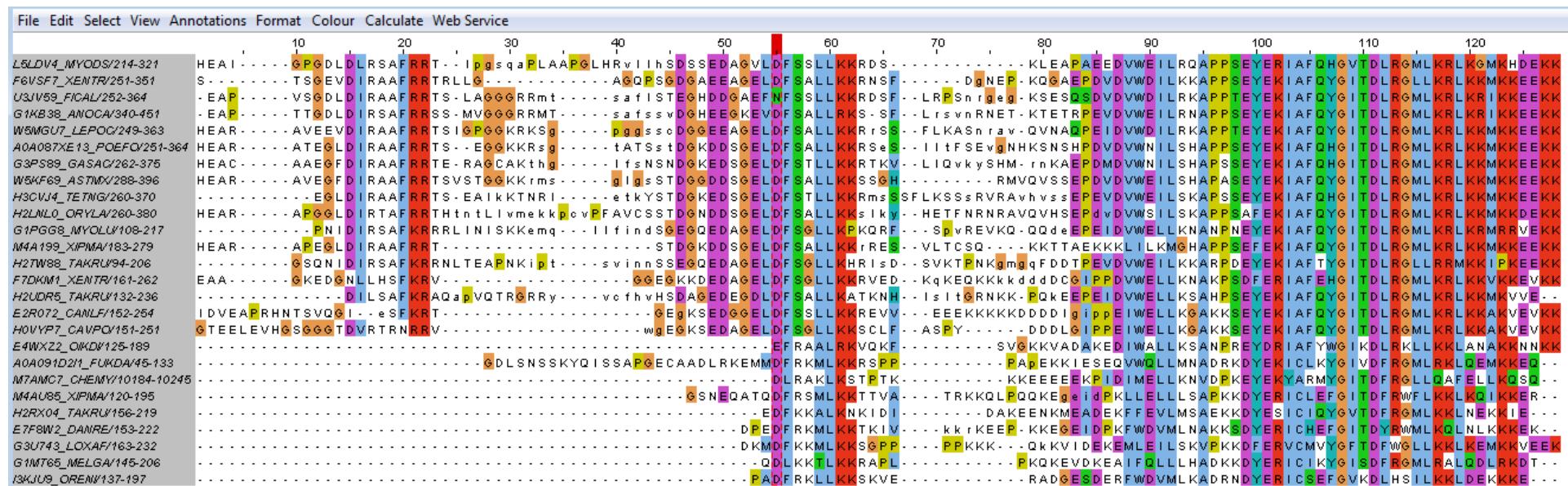
2. Web Service -> Secondary Structure Prediction-> JNet Secondary Structure Prediction



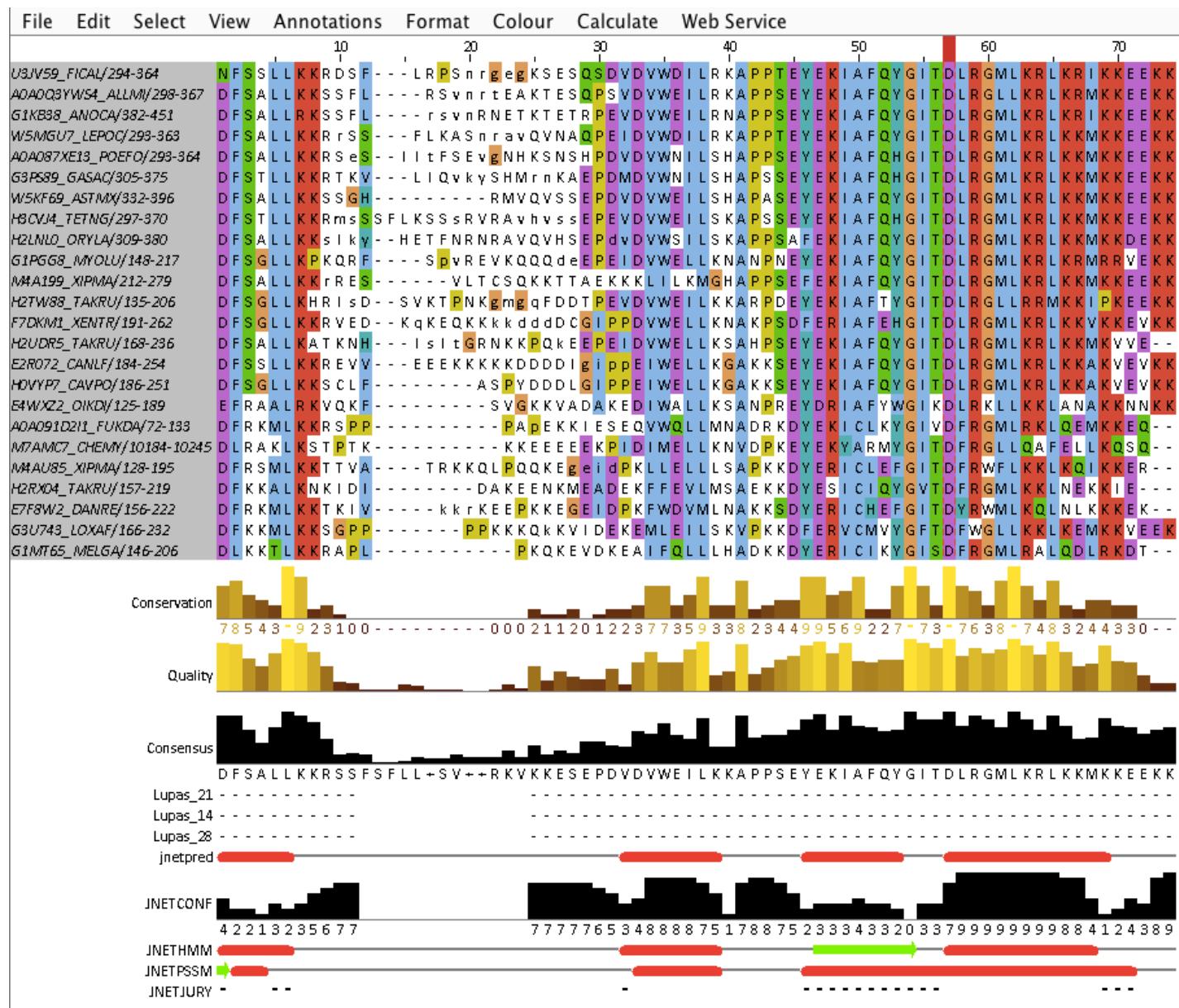
Click on top of column corresponding to alignment position 57



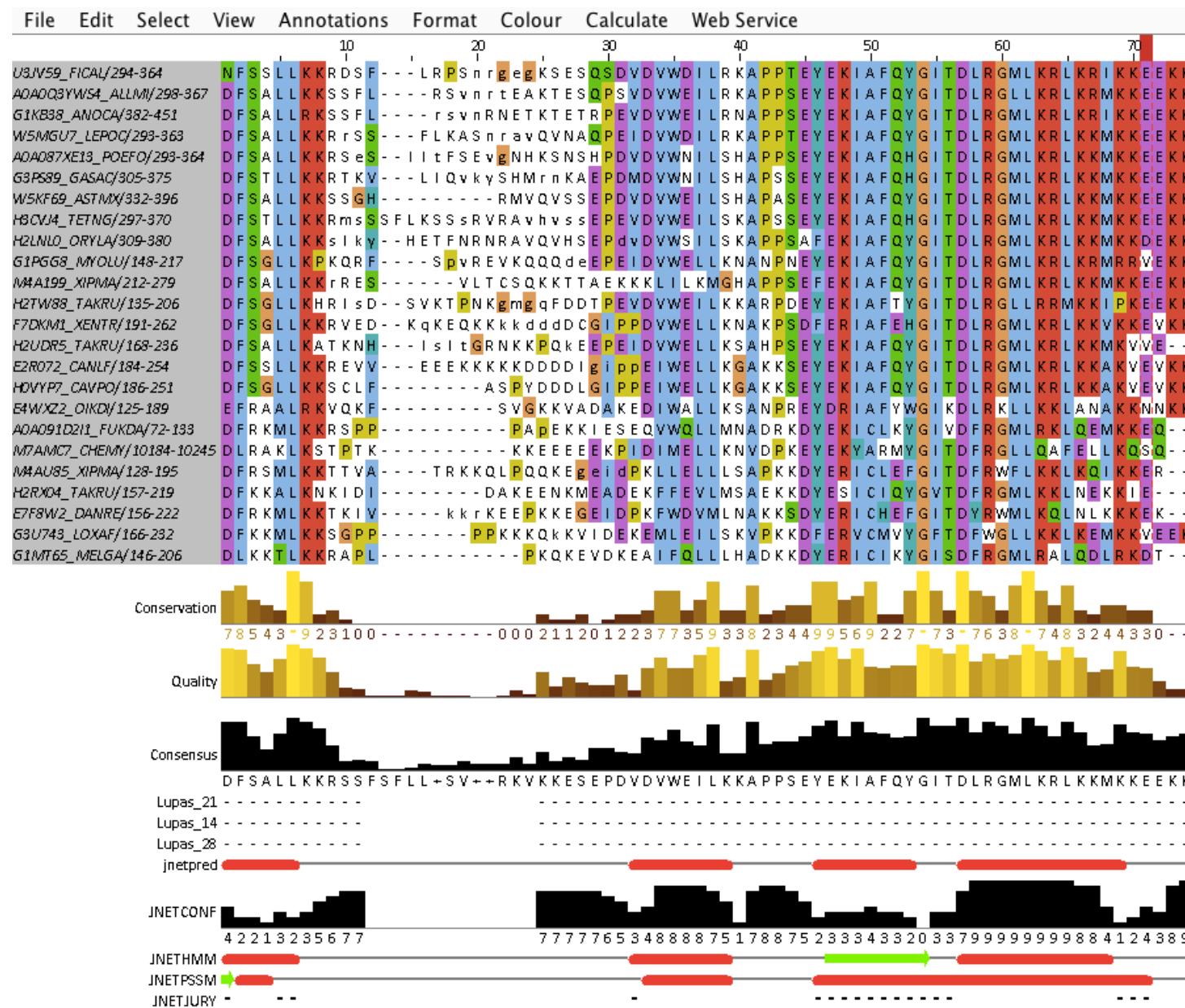
1. Edit -> Remove left



1. Edit -> Remove left

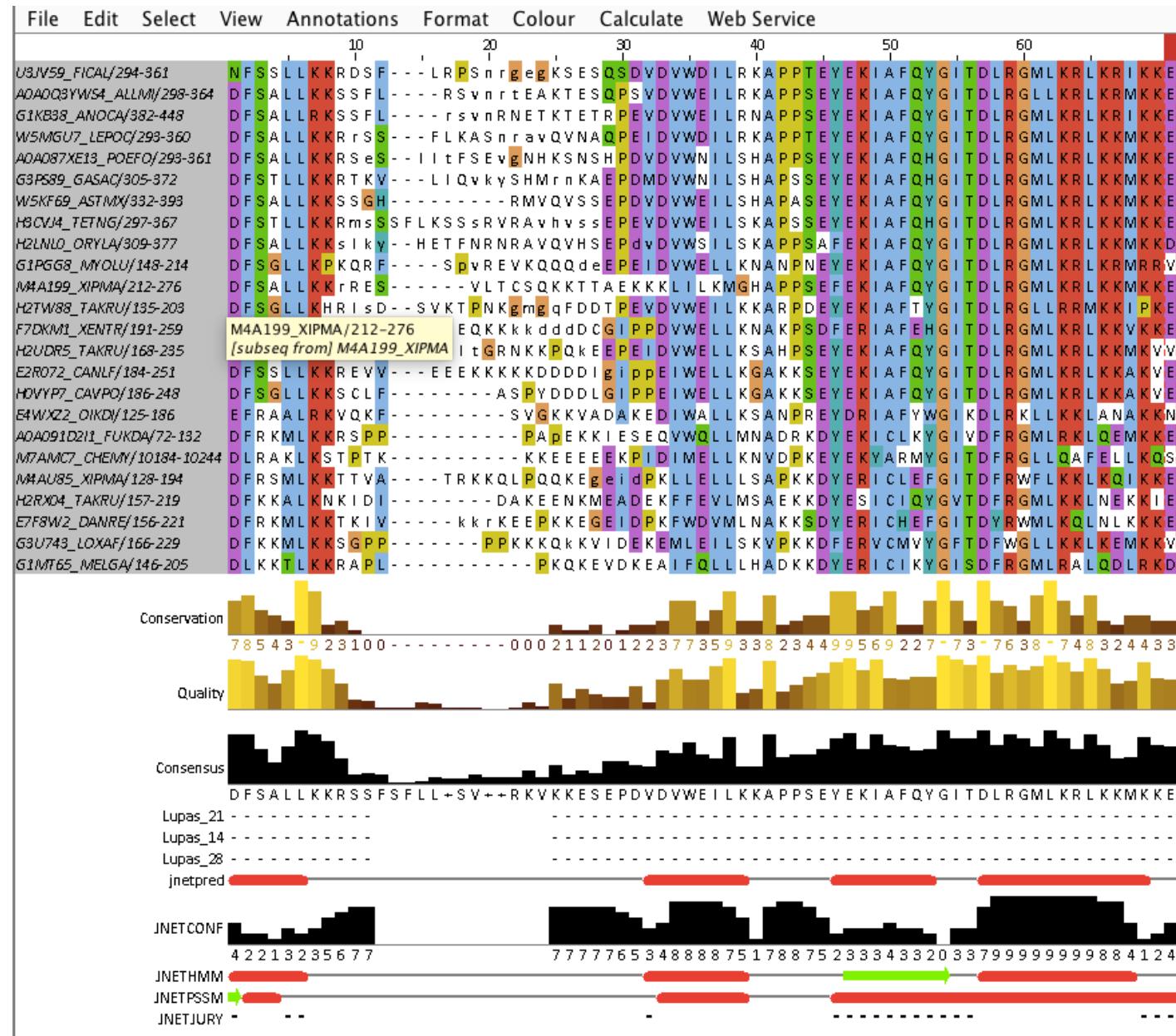


Click on top of column corresponding to alignment position 71

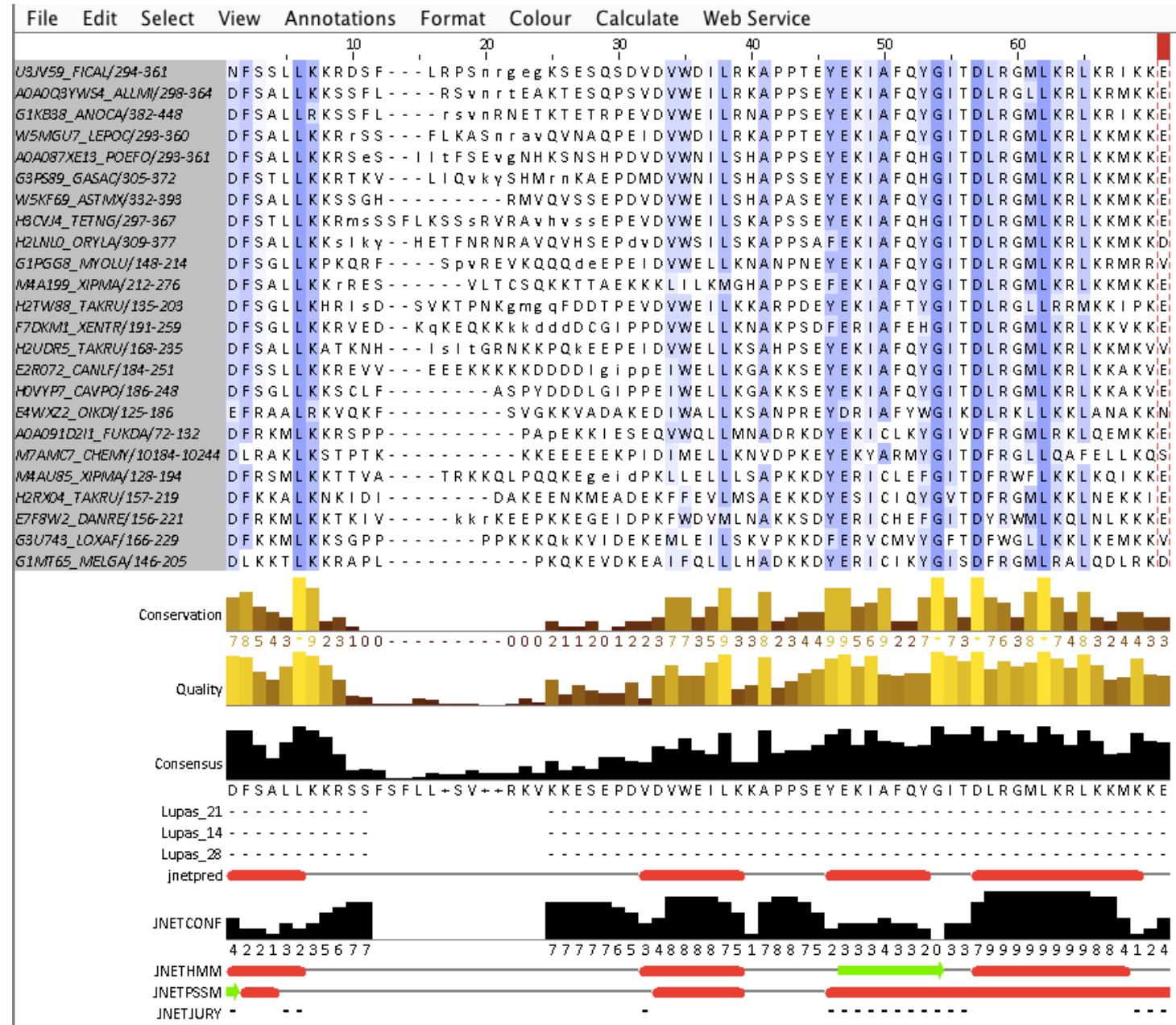


1. Edit -> Remove right

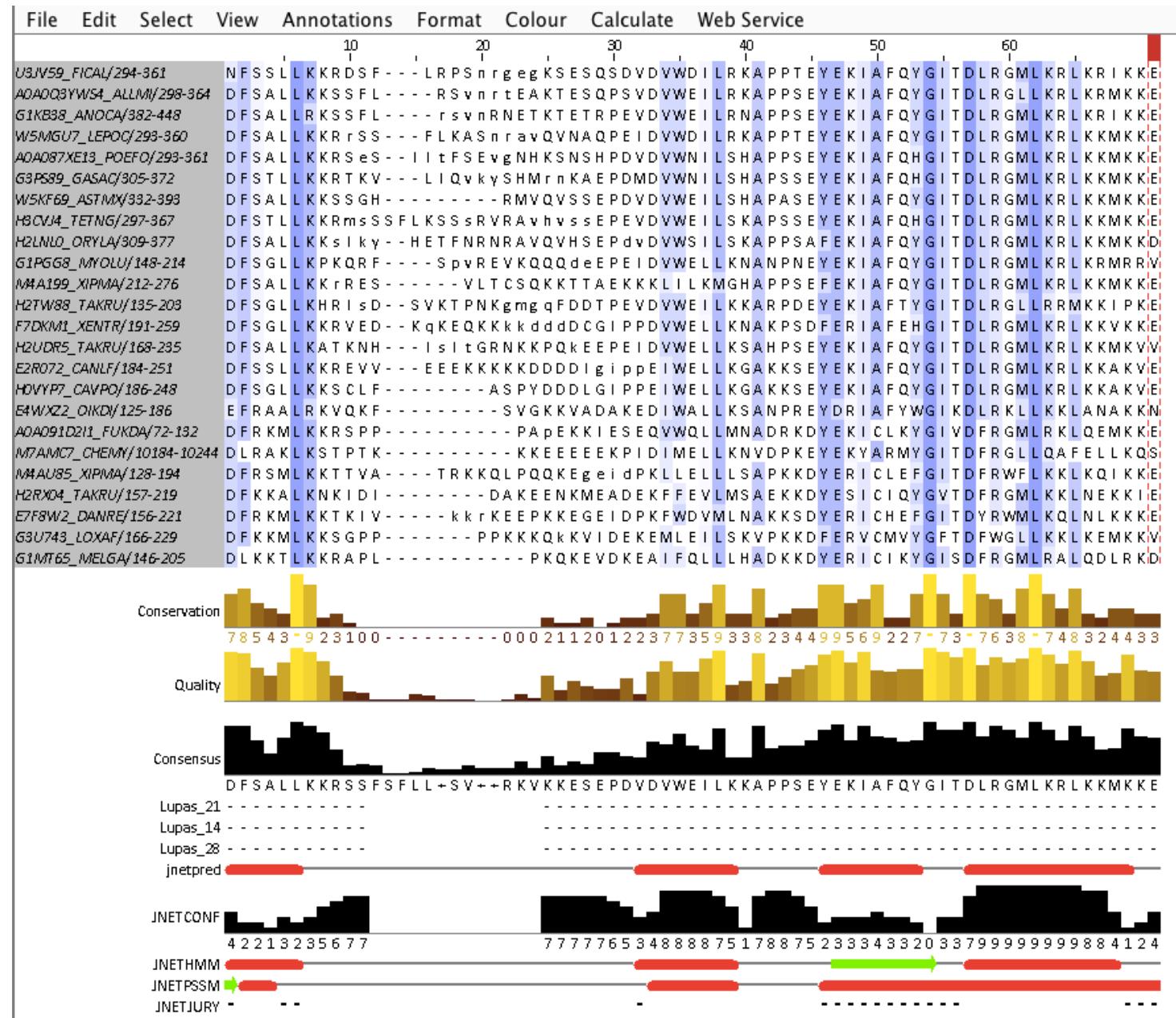
1. Edit -> Remove right



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



1. File -> Save as



Annotation team: results?

UniProt

UniProtKB ▾ Advanced ▾ Search

BLAST Align Retrieve/ID mapping Help Contact

UniProtKB - Q14896 (MYPC3_HUMAN)

Basket ▾

Protein Myosin-binding protein C, cardiac-type

Gene MYBPC3

Organism Homo sapiens (Human)

Sequence features View only features (sites, domains, PTMs ...)

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

Display None

BLAST Align Format Add to basket History Feedback Help video Other tutorials and videos

Function

Names & Taxonomy

Subcell. location

Pathol./Biotech

PTM / Processing

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

				Graphical	Feature

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 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 c Inferred from direct assayⁱ
- structural const
- titin binding  S 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...

The UniProt header features a blue navigation bar at the top. On the left is the UniProt logo. To the right is a search bar with a dropdown menu labeled "Literature citations ▾". Further right are "Advanced" and "Search" buttons. Below the search bar are links for "BLAST", "Align", "Retrieve/ID mapping", "Help", and "Contact".

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 microM for R870H compared with a K_d of approximately 5 microM 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 ▾

Search ScienceDirect



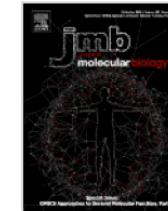
Advanced search



ELSEVIER

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},

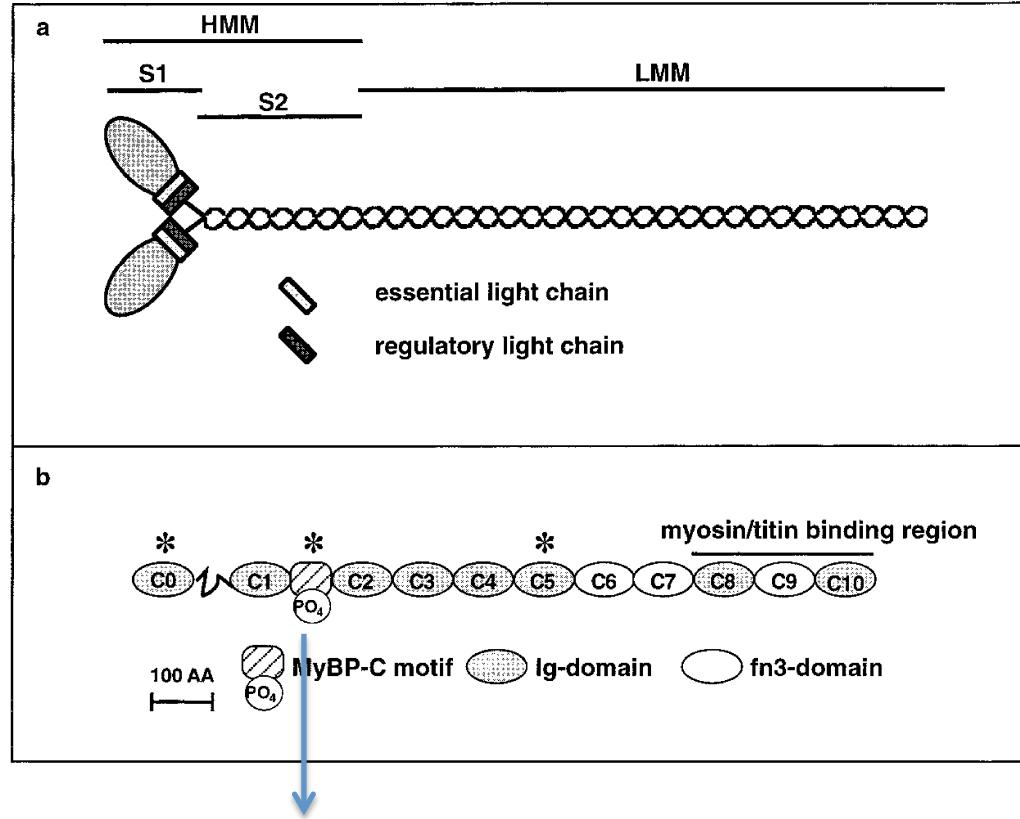
Show more

doi:10.1006/jmbi.1998.2522

[Get rights and content](#)

Abstract

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

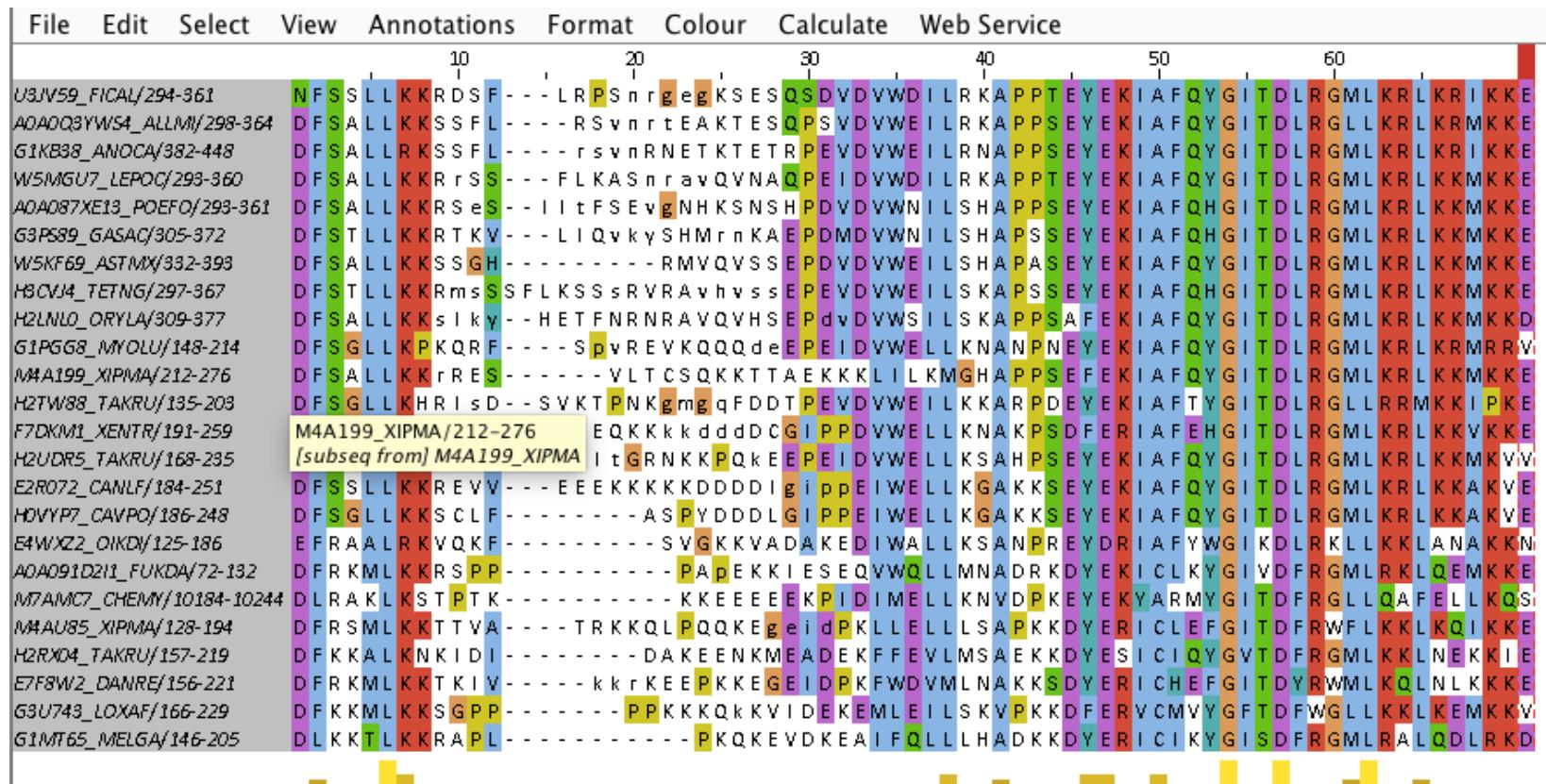


C	*
Ggc	1 VNEAPVS - GEMDIRA AFRRTS LAGGGRRMT SAFL STEGLEES GELNFS A 48
Hsc	1 VHEAMG - TGDL DLLSAF RRTS LAGGGRRIS - - - DSHEDTGILD FSS 42
Ggf	1 VEAPRQ - DASGQSLESFKRTS - - - EKKSDTAGELDFSG 34
Hsf	1 VEAPR - - SSEGNVLQAFKRTG - - - EGKDDTAGELDFSG 33
Hss	1 VHESTGTPNIDIRSAFKRSG - - - EGQEDA GELDFSG 34
Ggc	49 LLKKRDSFLRTANRGDGKSDSQP - DV DVWEILRK APPSEYEKI AAFQY G 95
Hsc	43 LLKKRDSFRTPRDSKLEAPAE - - - EDVWEILRQ APPSEYERI AAFQY G 86
Ggf	35 LLKKREVVEEEKKKKKKDDDDLGIPPEIWELLKG AKKSEYEKI AAFQY G 82
Hsf	34 LLKKREVQVEEKKKKKDEDDQF - - PPEIWELLKGVTKKSEYERI AAFQY G 80
Hss	35 LLKRRREVVKQQEEEPQV - - - DVWEILKN AKPSEYEKI AAFQY G 72
Ggc	96 ITDLRGMLKRLKRIKKEEKKST
Hsc	87 VTDLRGMLKRLKGMRDDEKKST
Ggf	83 ITDLRGMLKRLKKAKVEVKKSA
Hsf	81 ITDLRGMLKRLKKVHVEPKKSE
Hss	73 ITDLRGMLKRLKRMRREKKSA
Gruen and Gautel JMB (1999)	

Gruen and Gautel JMB (1999)

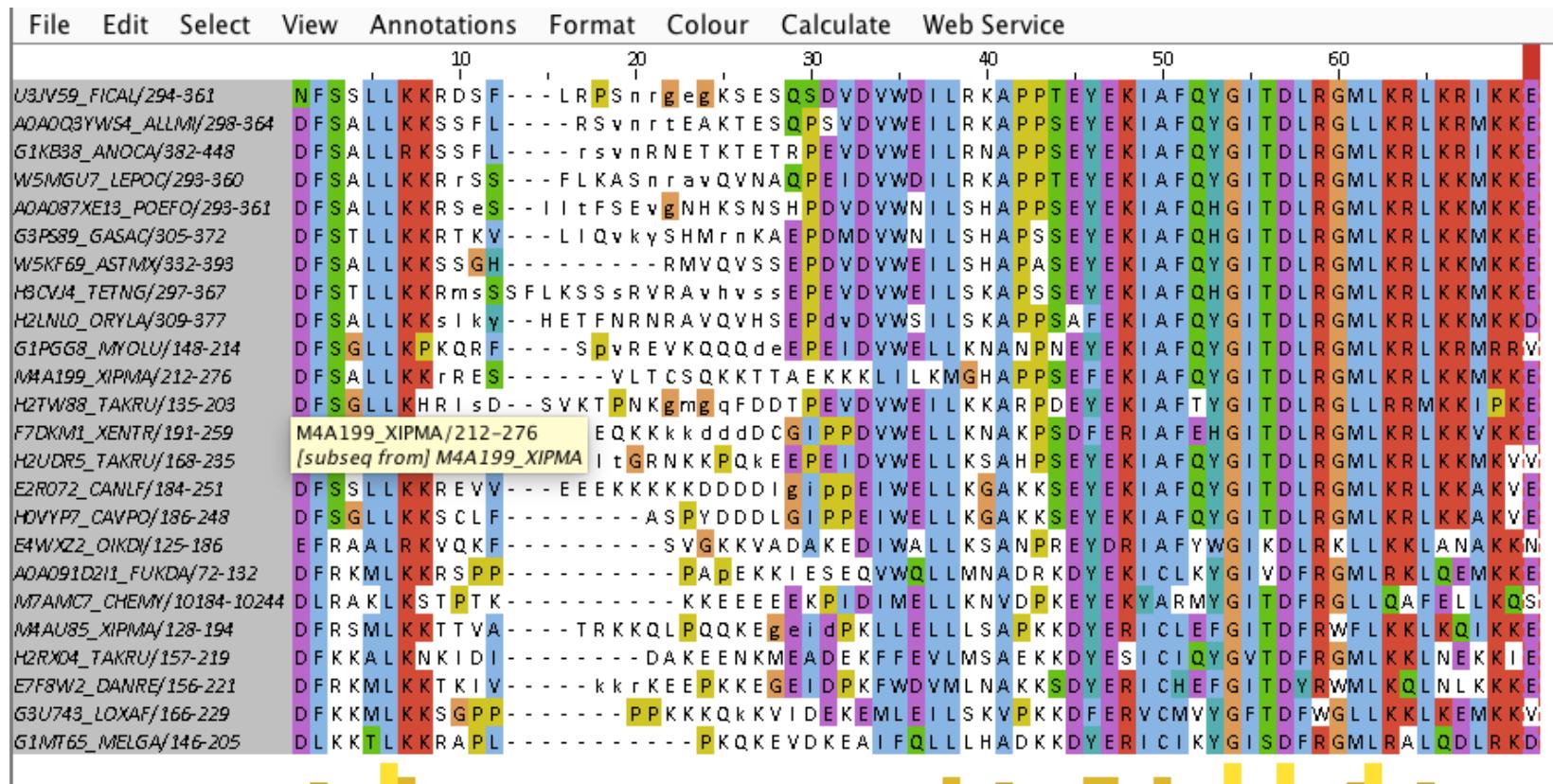
*

Ggc	1	VNEAPVS - GEMDI R A A F R RTSLAGGGRRMTSAFLSTEGLEESGELNFS	A	48			
Hsc	1	VHEAMG - TGDL DLLS A F R RTSLAGGGRRIS	- - - DSHEDT	GILD FSS	42		
Ggf	1	VEAPRQ - DASGQSLESF K KRTS	- - - EKKSDTA	GELDFSG	34		
Hsf	1	VEAPR - SSEG NVLQAFKRTG	- - - EGKDDTA	GELDFSG	33		
Hss	1	VHESTGTTPNIDIRS A F K RSG	- - - EGQEDA	GELDFSG	34		
Ggc	49	LLKKRDSFLRTANRGDGKSDSQP - DVDVWEILRK - APPSEYEKI	AFQY	YG	95		
Hsc	43	LLKKRDSFRTPRDSKLEAPAE - - - EDVWEILRQ - APPSEYE	R	IAFQY	YG	86	
Ggf	35	LLKKREVVEEEKKKKKKDDDDLGIPPEI	WELLKG - AKKSEYE	KI	IAFQY	YG	82
Hsf	34	LLKKREVQVEEKKKKKDEDDQF - - PPEI	WELLKGVTKKSEYE	R	IAFQY	YG	80
Hss	35	LLKRREVVKQQEEEPQV - - - DVWELLKN - AKPSEYE	KI	IAFQY	YG	72	
Ggc	96	ITDLRGMLKRLKRIKKEEKKST			117		
Hsc	87	VTDLRGMLKRLKGMRDDEKKST			108		
Ggf	83	ITDLRGMLKRLKAKVEVKKS	A		104		
Hsf	81	ITDLRGMLKRLKVKHVEPKKSE			102		
Hss	73	ITDLRGMLKRLKRMRREREKKSA			94		



*

Ggc	1	VNEAPVS - GEMDI RAAF RRTSL AAGG GRRMT SAFL S T E G L E E S GEL N FSA	48	
Hsc	1	VHEAMG - TG DLD DLLS AF RRTSL AAGG GRRIS	- - - DSHEDT GIL DFSS	42
Ggf	1	VEAPRQ - DAS QQSLES F KRTS	- - - EKKSDTA GEL DFSG	34
Hsf	1	VEAPR - SSEGNVLQAF KRTG	- - - EGKDDTA GEL DFSG	33
Hss	1	VH EST GTTPN IDIRSA F KRSG	- - - EGQEDA GEL DFSG	34
Ggc	49	LLKKR D SFL RTANRGDGKS DSQP - DV DV WE I LRK - APP SEYE KIAF QY G	95	
Hsc	43	LLKKR D SF RTPR DSKL E APAE - - - EDVWE I LRQ - APP SEYE RIAF QY G	86	
Ggf	35	LLKKR EV VEE EKKKKKK DDDDLGIPPEI WELLKG - AKKSEYE KIAF QY G	82	
Hsf	34	LLKKR EV QVEE EKKKKK DEDDQF - - PPEI WELLKG VTKKSEYE RIAF QY G	80	
Hss	35	LLKKR REV KQQ EEEPQV - - - DV WELL KN - AKP SEYE KIAF QY G	72	
Ggc	96	IT DLR GML KRL KRI KKE EKKST	117	
Hsc	87	VT DLR GML KRL KGM RR DE EKKST	108	
Ggf	83	IT DLR GML KRL KAK VE YKKSA	104	
Hsf	81	IT DLR GML KRL KKV HVE P KKSE	102	
Hss	73	IT DLR GML KRL KRM R RE EKKSA	94	



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)

EMBL-EBI

HMMER

Home | Search | Results | Software | Help | About

Services | Research | Training | About us

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

more significant

Bacteria Eukaryota Archaea Viruses Unclassified Sequences Other Sequences

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

	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

Customize

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																																																																																
V	Target					Description			Species			E-value																																																																														
	Query		Target Envelope		Target Alignment		Bias	Accuracy	% Identity (count)	% Similarity (count)	Bit Score	Ind.	Cond.																																																																													
start	end	start	end	start	end																																																																																					
1	103	22	124	22	124	1.99	1.00	100.0 (103)	100.0 (103)	216.0	7.4e-64	2.6e-69																																																																														
Description: Mybpc3 protein Species: <i>Mus musculus</i> E-value: 6.3e-64 <table border="1"> <thead> <tr> <th>Query</th> <th>Target</th> <th colspan="2">Target Envelope</th> <th colspan="2">Target Alignment</th> <th>Bias</th> <th>Accuracy</th> <th>% Identity (count)</th> <th>% Similarity (count)</th> <th>Bit Score</th> <th>Ind.</th> <th>Cond.</th> </tr> <tr> <th>start</th> <th>end</th> <th>start</th> <th>end</th> <th>start</th> <th>end</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> </tr> </thead> <tbody> <tr> <td>1</td> <td>80</td> <td>1</td> <td>103</td> <td>1</td> <td>103</td> <td>heai...gldl...rsafrrt...slagagr...rt...sdshedag...tldf...ssllk...rdsfr...dskleap...eedv...eilrg...appsey...iaf</td> <td>1.99</td> <td>1.00</td> <td>100.0 (103)</td> <td>100.0 (103)</td> <td>216.0</td> <td>7.4e-64</td> <td>2.6e-69</td> </tr> <tr> <td>Target</td> <td>101</td> <td>22</td> <td>HEAIGSGDLDLRSAFRRRTSLAGAGRRTSDSHEDAGTLDFSSLLKKRDSFRRDSKLEAPAEEDVWEILRQAPPSEYERIAF</td> <td>PP</td> <td>9*****</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Query</td> <td>103</td> <td>81</td> <td>qhgvtndlrgmlkrlkgmkgdekk</td> <td>PP</td> <td>*****98</td> <td>qhgvtndlrgmlkrlkgmkgdekk</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Target</td> <td>124</td> <td>102</td> <td>QHGVTDLRGMLKRLKGGMQDEKK</td> <td>PP</td> <td>*****98</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>													Query	Target	Target Envelope		Target Alignment		Bias	Accuracy	% Identity (count)	% Similarity (count)	Bit Score	Ind.	Cond.	start	end	start	end	start	end							1	80	1	103	1	103	heai...gldl...rsafrrt...slagagr...rt...sdshedag...tldf...ssllk...rdsfr...dskleap...eedv...eilrg...appsey...iaf	1.99	1.00	100.0 (103)	100.0 (103)	216.0	7.4e-64	2.6e-69	Target	101	22	HEAIGSGDLDLRSAFRRRTSLAGAGRRTSDSHEDAGTLDFSSLLKKRDSFRRDSKLEAPAEEDVWEILRQAPPSEYERIAF	PP	9*****								Query	103	81	qhgvtndlrgmlkrlkgmkgdekk	PP	*****98	qhgvtndlrgmlkrlkgmkgdekk							Target	124	102	QHGVTDLRGMLKRLKGGMQDEKK	PP	*****98							
Query	Target	Target Envelope		Target Alignment		Bias	Accuracy	% Identity (count)	% Similarity (count)	Bit Score	Ind.	Cond.																																																																														
start	end	start	end	start	end																																																																																					
1	80	1	103	1	103	heai...gldl...rsafrrt...slagagr...rt...sdshedag...tldf...ssllk...rdsfr...dskleap...eedv...eilrg...appsey...iaf	1.99	1.00	100.0 (103)	100.0 (103)	216.0	7.4e-64	2.6e-69																																																																													
Target	101	22	HEAIGSGDLDLRSAFRRRTSLAGAGRRTSDSHEDAGTLDFSSLLKKRDSFRRDSKLEAPAEEDVWEILRQAPPSEYERIAF	PP	9*****																																																																																					
Query	103	81	qhgvtndlrgmlkrlkgmkgdekk	PP	*****98	qhgvtndlrgmlkrlkgmkgdekk																																																																																				
Target	124	102	QHGVTDLRGMLKRLKGGMQDEKK	PP	*****98																																																																																					

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB An Information Portal to 119303 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands Go

Advanced Search | Browse by Annotations

PDB-101 Worldwide Protein Data Bank EMDDataBank StructuralBiology Knowledgebase Worldwide Protein Data Bank Foundation

F T Y A I O

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

NMR Ensemble

Display Files Download Files

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.](#), [Ramisetti, 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

Metric	Percentile Ranks	Value
Clashscore	Worse	1
Ramachandran outliers	Worse	0
Sidechain outliers	Worse	0.7%

Full Report

View in 3D JSmol or PV (in Browser)

Standalone Viewers

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 SSSLKKRDSFRRDSKLEAPAEEDVWEILRQAPPSEYERIAFQHGVTDLRGMLKRLKGKQ
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 MGSSHHHHHSSGLVPRGSHMHEAIGSGDLDLRSAFRRTSLAGAGRRTSDSHEDAGTLDF
PDB

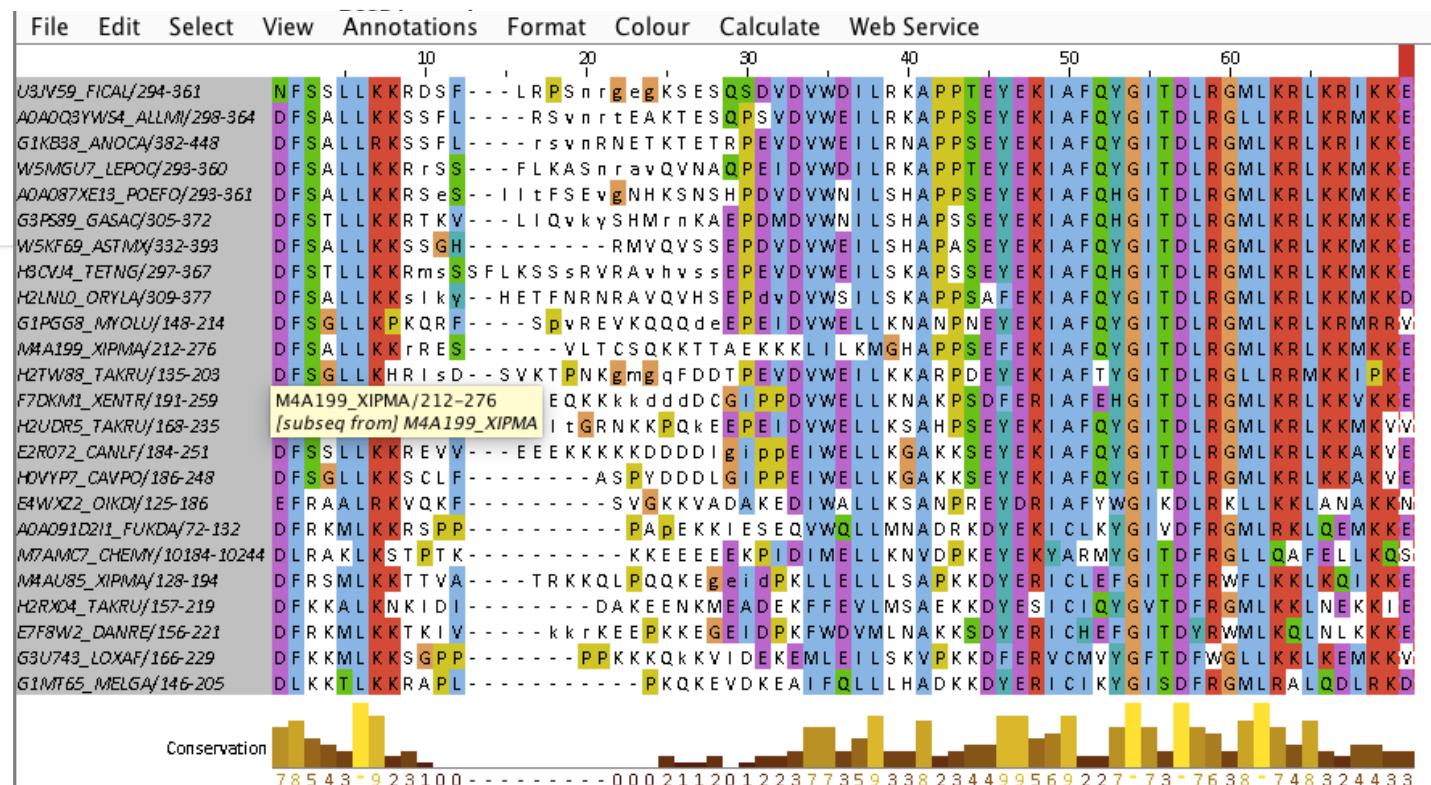


DSSP

PDB S S L L K K R D S F R R D 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 Q
PDB

315 320 330 340

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

DSSP

PDB MGSSHHHHHSSGLVPRGSHMHEAIGSGDLDLRSAFRRTSLAGAGRRTSDSHEDAGTLDF
PDB



DSSP

PDB S S L L K K R D S F R R D 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 Q
PDB

DSSP

PDB DEKK
PDB

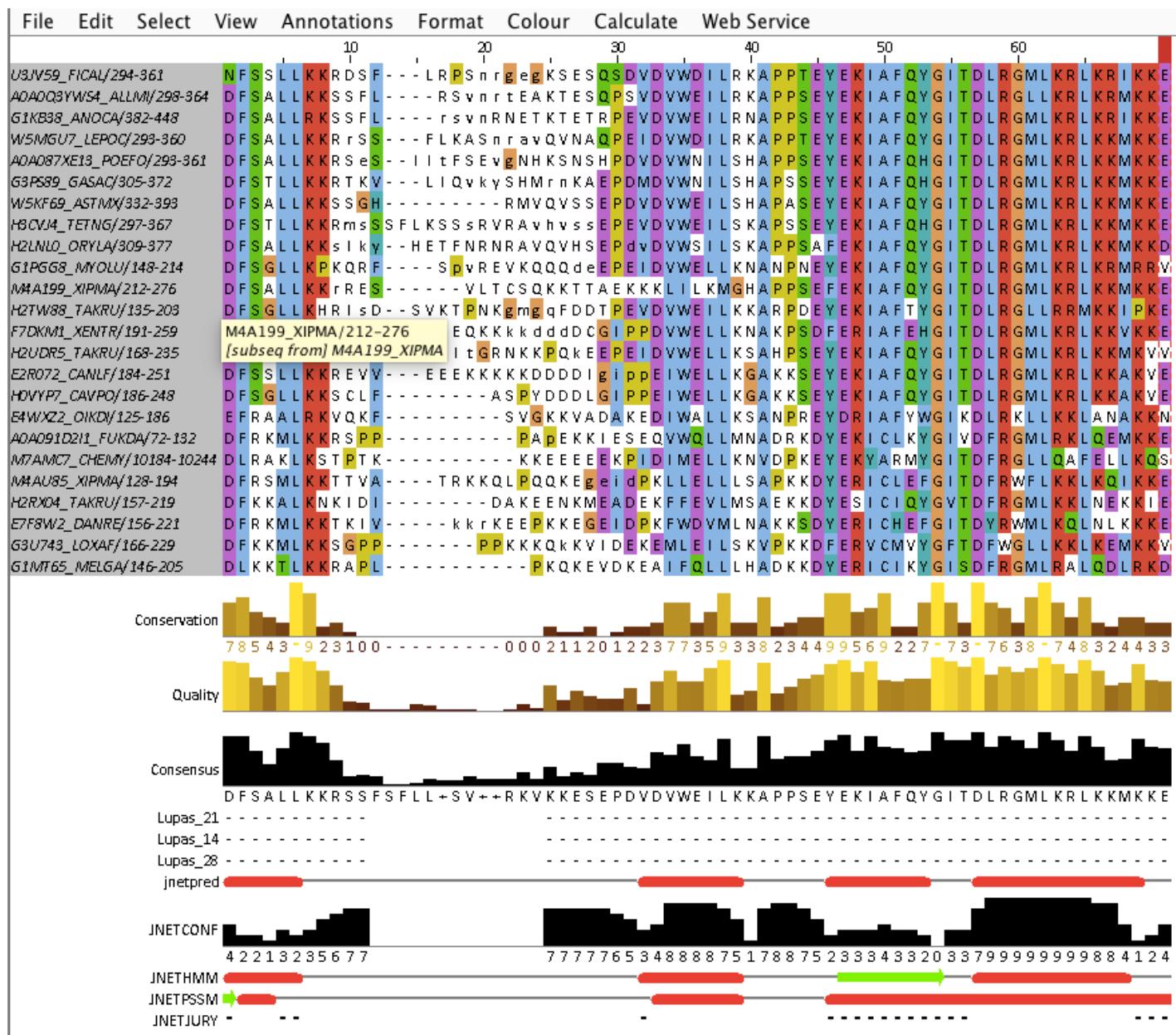
DSSP

PDB

File	Edit	Select	View	Annotations	Format	Colour	Calculate	Web Service	10	20	30	40	50	60	70	80	90			
U3JV59_FICAL/294-361				N F S S L L K K R D S F - - - L R P 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																
ADA0Q3YW54_ALLMI/298-364				D F S A L L K K S S F L - - - R S v n r t E 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 M L K R L K R M K K E																
G1K888_ANOCA/382-448				D F S A L L K K S S F L - - - r s v n R N E 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																
W5MGU7_LEPOC/298-360				D F S A L L K K R r S S - - - F L K A 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																
ADA087XE13_POEFO/298-361				D F S A L L K K R S S e s - - - l t F S E v g n 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																
G3PS89_GASAC/305-372				D F S T L L K K R T K V - - - L I Q v k y S H M r n K A E P D M D V W N I l S H 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																
W5KF69_ASTMX/332-393				D F S A L L K K S S G H - - - - - R M 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																
H3CVJ4_TETNG/297-367				D F S T L L K K R m s s F L K S 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																
H2LNID_ORYLA/309-377				D F S A L L K K s i k y - - - H E T F N R N 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																
G1PG68_MYOLU/148-214				D F S G L L K P K Q R F - - - S p v R E 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																
M4A199_XIPMA/212-276				D F S A L L K K r E S - - - V L T C S Q K K 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																
H2TW88_TAKRU/135-203				D F S G L L K H R I s d - - - S V K T P 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																
F7DKM1_XENTR/191-259				M4A199_XIPMA/212-276 E Q K K K k 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																
H2UDRS_TAKRU/169-235				[subseq from] M4A199_XIPMA I t G R N 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																
E2R072_CANLF/184-251				D F S S S L L K K R E V V - - - E E E 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																
HOVYP7_CAVPO/186-248				D F S G L L K K S C L F - - - - - A S P 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																
E4WXZ2_OIKDI/125-186				E F R A L K R V Q K F - - - - - S V G 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																
ADA091D211_FUKDA/72-132				D F R K M L K K R S P P - - - - - P A P E K K I E S Q V W O l l M 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																
M7AMC7_CHEMY/10184-10244				D L R A K L K S T P T K - - - - - K K E E E E K P I D 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																
M4AU85_XIPMA/128-194				D F R S M L K K T T V A - - - - - T R K K Q L P Q Q K E g i d P K L L E l 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																
H2RX04_TAKRU/157-219				D F K K A L K N K I D I - - - - - D A K 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																
E7F8W2_DANRE/156-221				D F R K M L K K T K I V - - - - - k k r K E 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 I C H E F G I T D Y R W M L K Q L N L K K E																
G3U743_LOXAF/166-229				D F K K M L K K S G P P - - - - - P P 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 V																
G1MT65_MELGA/146-205				D L K K T L K K R A P L - - - - - P K Q K E V D K E A I F Q L L H A D K K D Y E R I C I C I K Y G I S D F R G M L R A L Q D L R K D																

Conservation





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

A PARTIALLY FOLDED DOMAIN*

Jack W. Howarth[‡], Srinivas Ramisetty[‡], 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

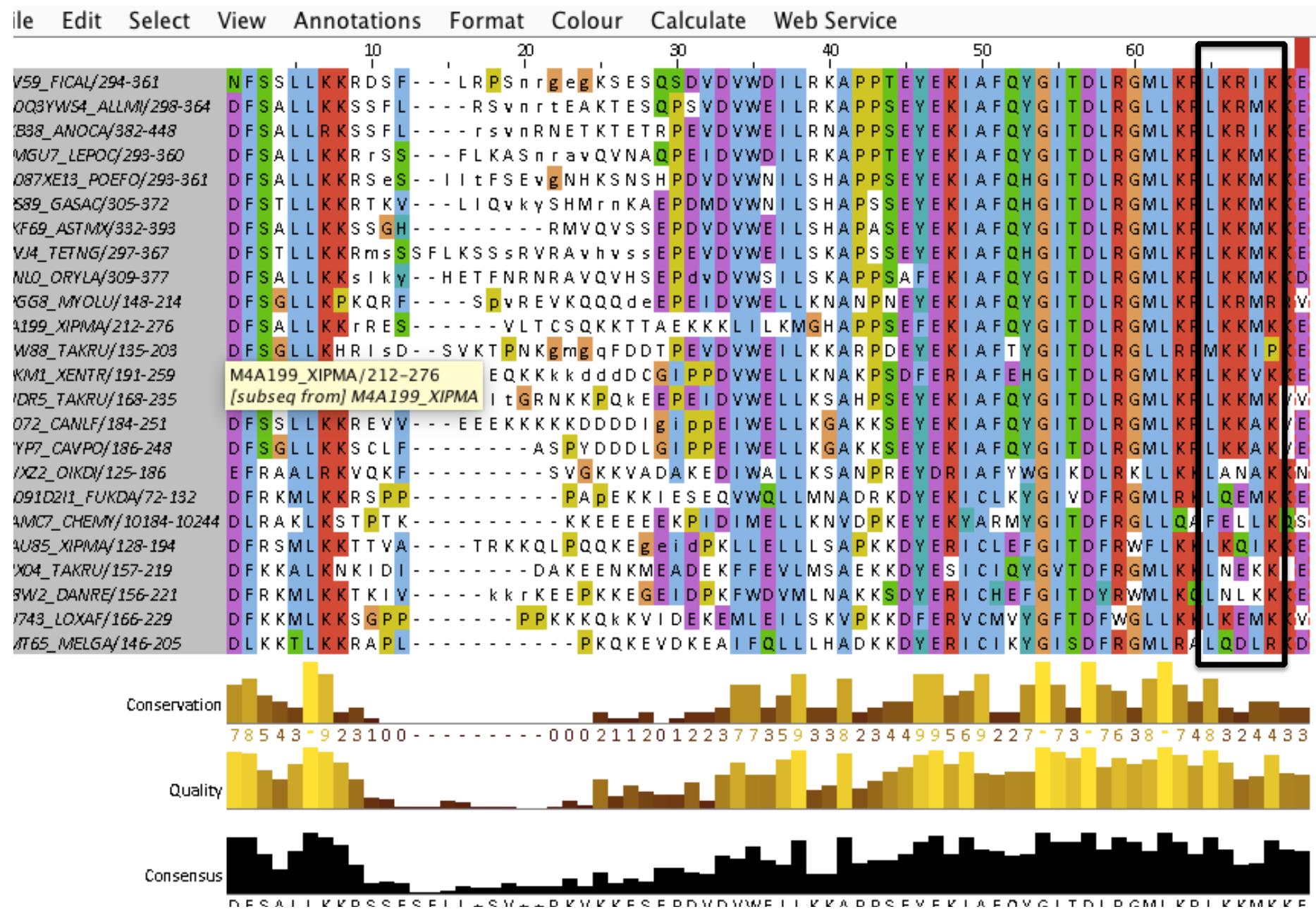


The structural role of the unique myosin-binding motif (m-domain) of cardiac myosin-binding protein-C remains unclear. Functionally, the m-domain is thought to directly interact with myosin, whereas phosphorylation of the m-domain has been shown to modulate interactions between myosin and actin. Here we utilized

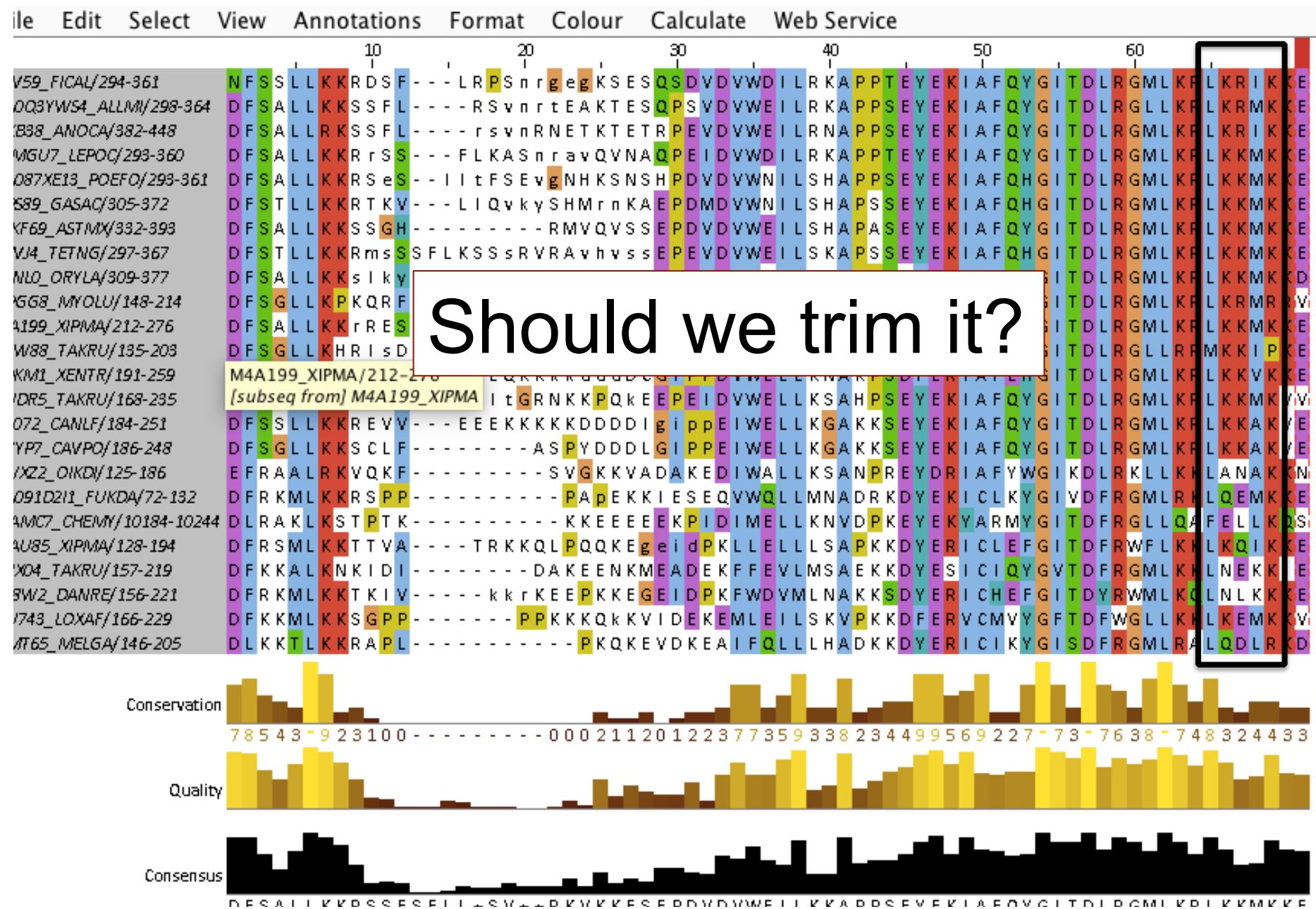
NMR to analyze the structure and dynamics of the m-domain in solution. Our studies reveal that the m-domain is composed of two subdomains, a largely disordered N-terminal portion containing three known phosphorylation sites and a more ordered and folded C-terminal portion. Chemical shift analyses, $d_{NN}(i, i +$

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, $^{15}\text{N}\{\text{H}\}$ 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 (α 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.

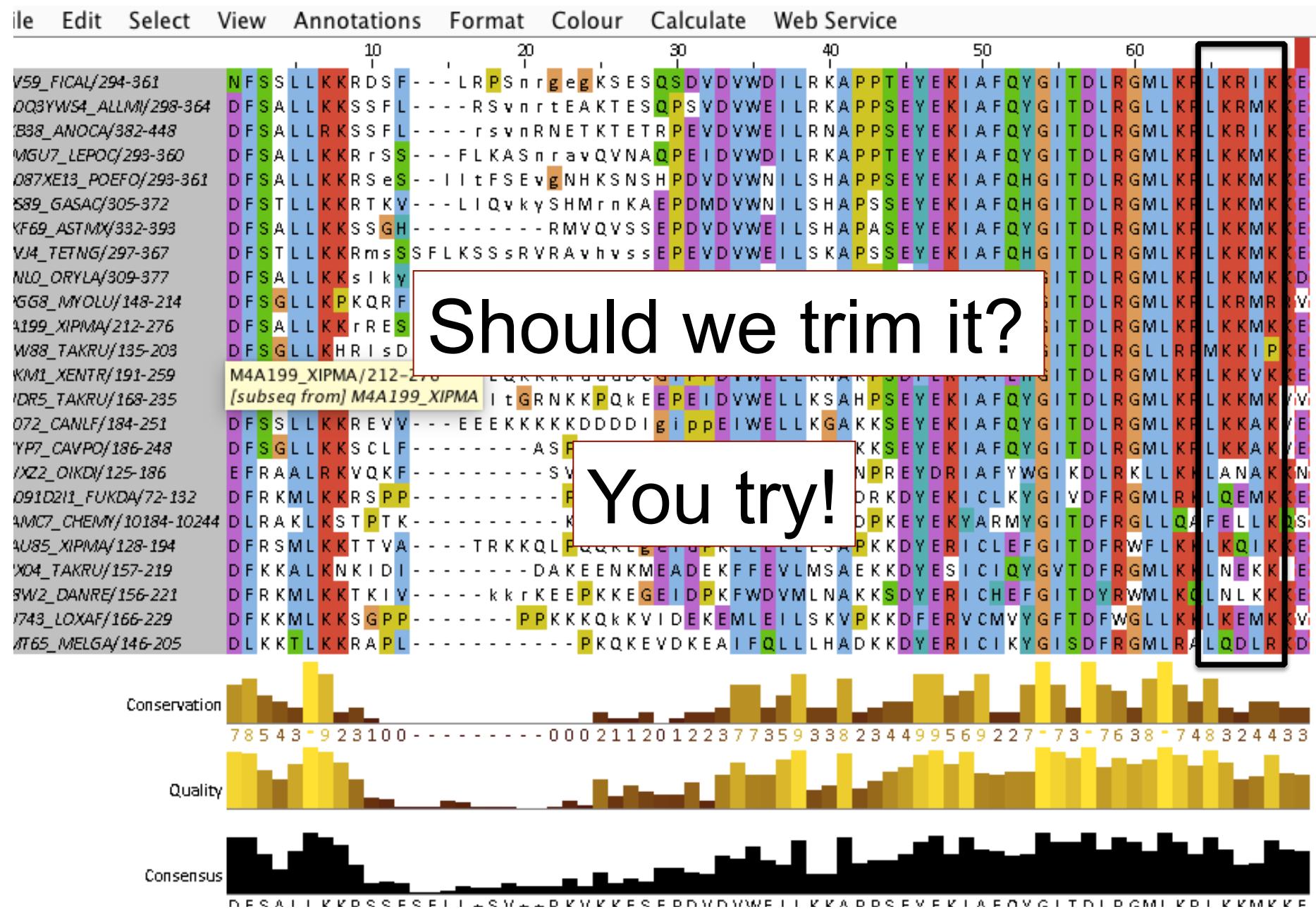
LK(R/K)XK



LK(R/K)XK



LK(R/K)XK



Search Team

Marco Punta

EMBO Workshop, Budapest, 2016

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

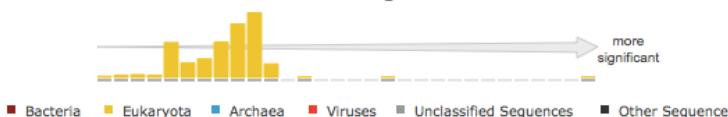
 rp75 rp55 rp35 rp15

Other databases

 Ensembl Genomes QfO**► Restrict by Taxonomy** **▼ Cut-Offs**



Distribution of Significant Hits


[« First](#) [« Previous](#) **Page 1 of 5** [Next »](#) [Last »](#)

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

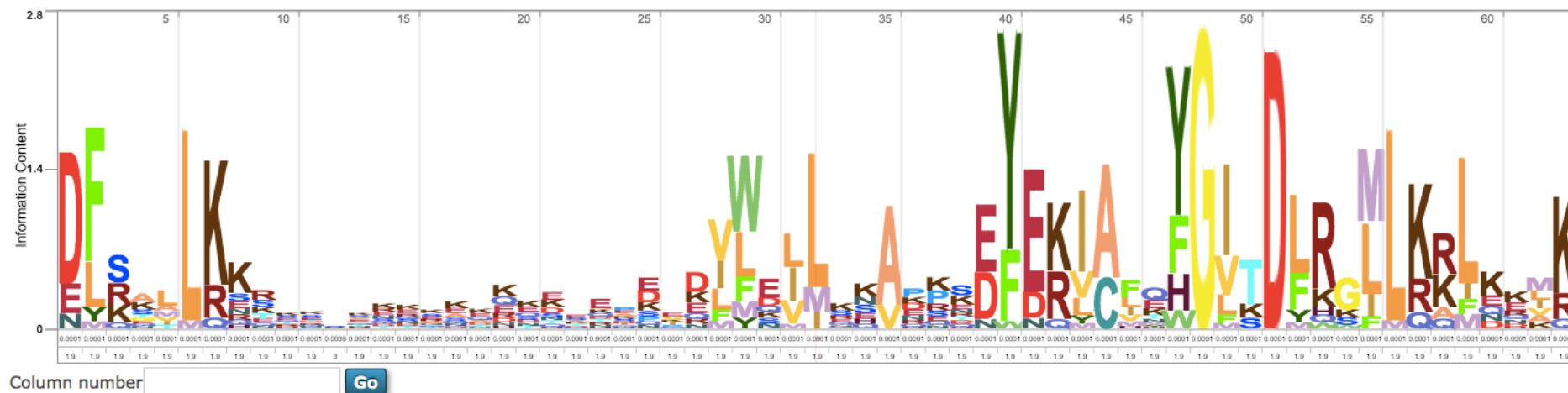
(show all) alignments

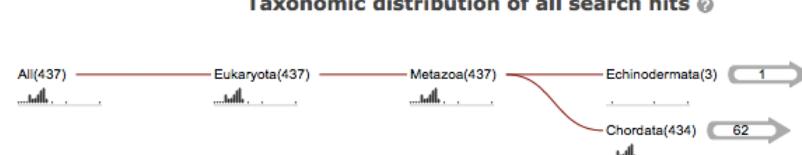
Your search took: 0.82 secs
showing rows 1 - 100 of 453

Search Details

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

Model Position



[Home](#) | [Search](#) | **Results** | [Software](#) | [Help](#) | [About](#)[Search Again](#)**HMMSEARCH Results**[Score](#) **Taxonomy** [Domain](#) [Download](#)**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



Home | Search | Results | Software | Help | About



Score Taxonomy Domain Download

Search Again

HMMSEARCH Results

Domain Architectures ①« First « Previous Page 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, example:[F6VSF7_XENTR](#)[View Scores](#)

Sequence Features

46
SEQUENCES
[Show All](#)with domain architecture: I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set, example:[H2UDR5_TAKRU](#)[View Scores](#)

Sequence Features

41
SEQUENCES
[Show All](#)with domain architecture: I-set, I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set, example:[E7EZ79_DANRE](#)[View Scores](#)

Sequence Features

18
SEQUENCES
[Show All](#)with domain architecture: I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set, example:[I3MSM7_ICTRR](#)[View Scores](#)

Sequence Features

11
SEQUENCES
[Show All](#)with domain architecture: I-set, I-set, I-set, example:[F7CWG3_CALJA](#)[View Scores](#)

Sequence Features

11
SEQUENCES
[Show All](#)with domain architecture: I-set, I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set, example:[E2REQ2_CANLF](#)[View Scores](#)

Sequence Features

9
[Show All](#)with domain architecture: I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set, example:[H3AVM7_LATCH](#)[View Scores](#)

[Home](#) | [Search](#) | [Results](#) | [Software](#) | [Help](#) | [About](#)| [Email](#)[Search Again](#)[3 Next »](#) [Last »](#)[View Scores](#)[View Scores](#)[View Scores](#)[View Scores](#)[View Scores](#)[View Scores](#)[View Scores](#)

Score

HMMSEARCH Results**Domain Architectures** **119
SEQUENCES**with domain architecture: **I-set, I-set, I-set, I-set, I-set, fn3,**

Sequence Features

[Show All](#)**46
SEQUENCES**with domain architecture: **I-set, I-set, I-set, I-set, fn3, fn3, I**

Sequence Features

[Show All](#)**41
SEQUENCES**with domain architecture: **I-set, I-set, I-set, I-set, I-set, I-set**

Sequence Features

[Show All](#)**18
SEQUENCES**with domain architecture: **I-set, I-set, I-set, fn3, fn3, fn3, I-**

Sequence Features

[Show All](#)**11
SEQUENCES**with domain architecture: **I-set, I-set, I-set, example:F7CWG3**

Sequence Features

[Show All](#)**11
SEQUENCES**with domain architecture: **I-set, I-set, I-set, I-set, I-set, fn3,**

Sequence Features

[Show All](#)**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**

Thank you!!